VALENCIA 24-25 November 2011 ISVA International Seminar on Virtual Acoustics

RECORDING AND REPRODUCING CONCERT HALL ACOUSTICS FOR SUBJECTIVE EVALUATION

Reference PACS: 43.55.Mc, 43.55.Gx, 43.38.Md

Lokki, Tapio

Aalto University School of Science, Dept. of Media Technology
P.O.Box 15400, FI-00076 AALTO, Finland
Tapio.Lokki@aalto.fi

ABSTRACT

Listen to the acoustics of an existing concert hall in the laboratory environment requires capturing and reproducing spatial sound as authentically as possible. Based on our recent research, the best quality can be achieved by measuring spatial impulse responses with a large number of source positions, process the responses with spatial impulse response rendering (SIRR), and convolve these processed responses with anechoic symphony orchestra recordings. This paper explains the recording and reproduction processes in detail and gives references to publications in which the presented methods are used in subjective evaluation of concert hall acoustics.

1. INTRODUCTION AND SOUND SOURCES

The subjective evaluation of concert hall acoustics can be done in-situ by listening the real symphony orchestra or later in the laboratory if the music in the hall is captured and rendered properly. Both techniques are widely used and they have some pros and cons, as discussed, e.g., in [1]. In this paper, we are discussing on the techniques with which the spatial sound in the real hall can be captured and transferred to the laboratory for subjective listening tests. The presented techniques aim at simultaneous comparison of multiple halls, or different seats in one or more halls. The key point is that a virtual symphony orchestra is applied to guarantee exactly the same performance in each hall. Such simultaneous listening is essential to careful listening to the differences between acoustics of the concert halls as discussed in [1,2,3].

The most typical sound source in a concert hall is a symphony orchestra that has dozens of musicians with instruments of varying directivity [4]. Therefore, for auralization studies it is not very convincing to use only a few source positions on stage as is recommended for measuring objective parameters in the ISO standard 3382-1:2009 [5]. In addition, no real instrument has similar directivity as the omnidirectional dodecahedron that fulfils the standard. Naturally, it would be desirable to record a real orchestra, but even professional orchestra cannot play exactly the same way in different halls. Therefore, a controllable orchestra, which can be calibrated is needed for concert hall acoustics evaluation studies.

In our solution, we use 24 source positions, with in total of 34 active two-way loudspeakers, to simulate the symphony orchestra. The loudspeakers are covering an area equivalent to the area that a symphony orchestra would occupy [6]. The number of loudspeakers and the number of individual channels are a compromise between practicality and accuracy. Naturally we could use one loudspeaker at the location of each individual player, but then the recording of one hall would not be possible in one day.

All strings are reproduced with channels 1-14 and with 23 loudspeakers, see Fig. 1. Thus, most of the string channels use two loudspeakers to better match the total directivity of a channel to the directivity of a violin or a viola. Channels 15-18 are for woodwinds, one for each instrument group. Channels 19 and 20 represent 4 French horns each and channels 21-24 are for trumpets, trombones, tuba, and timpani as well as other percussion instruments.

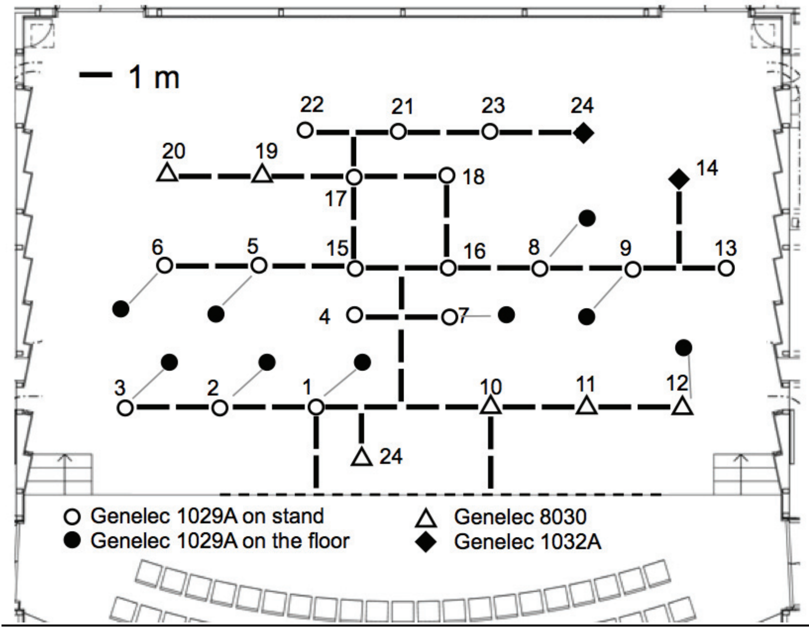

Figure 1: The layout of the loudspeaker orchestra [6]. The distribution of the instruments to the loudspeaker channels is explained in [2].

2. RECORDING THE SPATIAL SOUND

The loudspeaker orchestra can be used to reproduce anechoic music [7] in-situ in the concert hall, or it can be used for impulse response measurements. The music or impulse responses should be captured with multiple microphones so that the sound field in the hall could be reproduced as authentically as possible in the laboratory. Traditionally, the dummy heads are used for binaural recordings and then these recordings are listened to with high quality headphones. However, in our opinion the quality of binaural technology is not good enough for accurate evaluation of concert hall acoustics. There are at least three major issues that hinder the use of binaural technology:

1. A dummy head is static, thus subjects in the laboratory cannot move their heads when listening carefully to the recordings. The lack of head movements, in addition to lack of visual cues, makes it harder to concentrate to the orchestra and in our opinion this is one of the reasons why dummy head recordings sound often noisy. Our brains do not have movement and visual cues that help in concentrating to the orchestra, thus suppressing other sound events, as is the case in-situ in a concert hall.
2. The headphone compensation is very cumbersome at high frequencies above 5 kHz. Even though the compensation is done very carefully, a small change in headphone position after compensation measurement can make large error to binaural signals as the wavelength of sound at high frequencies is so small. Below 5 kHz, the compensation can be done properly, but unfortunately there is important information in the music also at high frequencies [8, 9], and these frequencies should be reproduced correctly.
3. The externalization of binaural recordings is not working well enough for all subjects. The dummy head is not the same size and shape as the heads of subjects, thus there is a HRTF mismatch between the heads. That, in addition to lack of head movements, might hinder the proper externalization.

Mainly due to these three problems, we have been using loudspeakers in the spatial sound reproduction. Multichannel 3D sound reproduction system allows a listener to move and rotate his head, the loudspeakers reproduce the whole audible frequency range with flat spectrum and there are no individual problems due to HRTF mismatch.

2.1 Recording music for spatial sound reproduction

Recording spatial sound in a concert hall so that it can be properly reproduced with a 3D loudspeaker array is not a trivial task. The most well known method is Ambisonics [10]. Currently only first order microphones are commercially available. The B-format consists of one omni and three figure-of-eight microphone signals. The B-format recording can be reproduced with a 2D or 3D loudspeaker array and Ambisonics processing treats equally sounds arriving from all directions. When Ambisonics is listened to anechoic or dry listening conditions it reproduces the 3D sound field quite nicely. However, the spatial image is not as accurate as it could be and in some cases Ambisonics can produce undesired phase errors. However, current research on 2nd order microphones and related signal processing will make spatial sound more accurate in the near future.

Another spatial sound coding technology is Directional Audio Coding (DirAC) [11, 12]. It is based on the idea that the original sound field in a concert hall is not required in the laboratory. If all binaural cues needed by human spatial hearing are reproduced, listeners would perceive the sound as in the recording space. Therefore, DIrAC applies the intensity and diffuseness analysis in time-frequency domain to distribute the B-format recording to the 3D loudspeaker setup. Figure 2 shows how we have processed signals captured with 3D intensity probe (GRAS Type 50 VI-1). First the anechoic music emitted by the loudspeaker orchestra is recorded with six high quality omnidirectional microphones. They consist of three pairs in x, y, and z directions and recordings in each receiver position are done twice with 25mm and 100mm spacers. Two spacers are needed to obtain good figure-of-eight signals from the probe at wide frequency range [13]. Then B-format signals are coded with DirAC for the 16-channel 3D reproduction system, consisting of 8 loudspeakers at ear level and 4 at elevation of 45 degrees and 4 at elevation of -45 degrees. Finally, only the usable frequency range from both recordings is used.

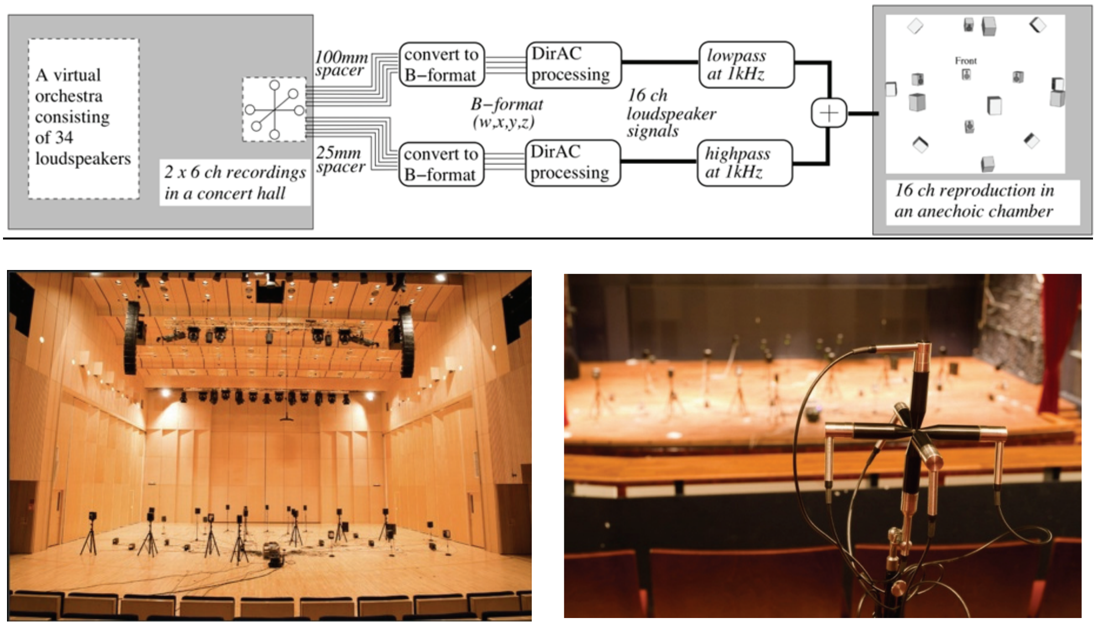

Figure 2: The recording of music emitted by the loudspeaker orchestra in a concert hall. The six microphone signals are converted to B-format, which is furthermore processed with DirAC for 16-channel loudspeaker reproduction [2].

The DirAC processing has some drawbacks that might affect to the sound quality. First of all the conversion from six omnidirectional microphones to B-format increases the noise level at low frequencies, which might be audible when orchestra is playing in *piano*. In addition, the transients of the recorded signals might be slightly blurred due to the smoothing and windowing that are necessary to avoid audible artefacts in the final signals. Although, these problems can lower the sound quality, the processing is equal to all recordings, thus the differences between the studied concert halls remain, if the processing artefacts are not overriding them.

2.2 Capturing impulse responses for spatial sound reproduction

Another option for spatial sound capturing is to measure a spatial impulse response in a concert hall, distribute it for each reproduction loudspeaker and finally convolve the loudspeaker responses with anechoic music. Again, we are using for measurements a six-channel GRAS vector intensity probe (Type 50 VI-1) and the impulse responses are measured with swept-sine technique [14] to obtain high signal to noise ratio. Each loudspeaker on the stage was calibrated in each hall by measuring 85 dBA at 1 m distance when the loudspeaker emitted bandpass (200 - 1000 Hz) white noise. All six omnidirectional microphones were calibrated with the B&K 4231 calibrator.

To subdivide the B-format impulse response to a 3D loudspeaker array we use spatial impulse response rendering (SIRR) algorithm [15,16], as illustrated with one real measured example in Fig. 3. It divides a spatial impulse response in time-frequency domain into individual impulse responses, one for each reproduction channel. In our concert hall measurements, one measured spatial impulse response was distributed for the 14 channel spatial sound reproduction system, consisted of eight loudspeakers at ear level at 45 degree intervals, four loudspeakers horizontally equispaced at 55 degree elevation above the ear level, and two loudspeakers 40 degree below ear level at azimuth angles -22 and 22 degrees. The processing of one measurement from one loudspeaker orchestra channel is illustrated in Fig. 3. With the full loudspeaker orchestra, the SIRR processing produced 672 impulse responses (24 source channels X 14 reproduction channels X 2 frequency ranges, crossover at 1 kHz) for convolution with the anechoic music [7].

The signals of individual instruments were convolved with the SIRR processed responses of the loudspeaker orchestra channels as presented earlier [2]. Each string instrument recording was individually processed with time varying delays, pitch shifting, amplitude modulation, and varying the recording microphone. When these copies were reproduced from spatially separated loudspeakers, a natural and convincing string section sound was achieved [17]. This processing method produced high quality samples that can be simultaneously compared in the laboratory environment.

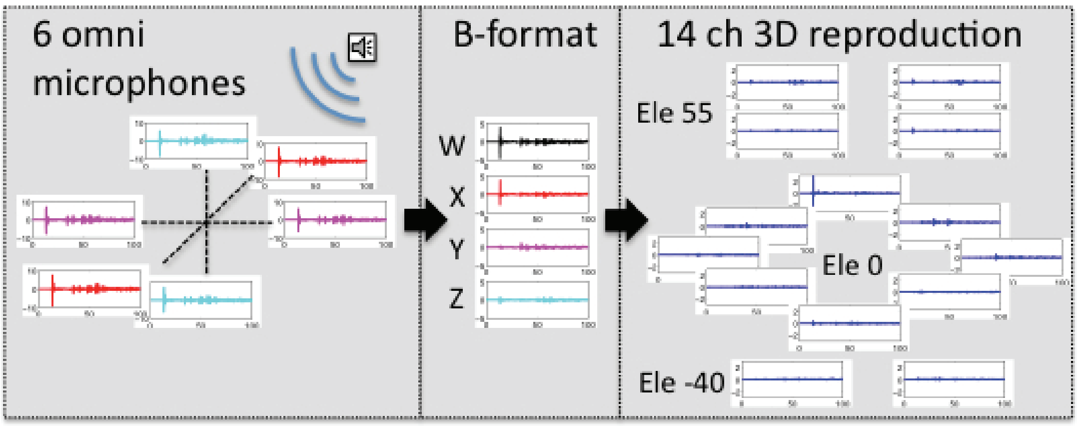

Figure 3: The processing of one spatial impulse response captured with six omnidirectional microphones with the SIRR method [15, 16]. The computed response for each reproduction channel is convolved with corresponding anechoic recording for subjective evaluation.

3. CONSIDERATIONS REGARDING THE LISTENING SPACE IN THE LABORATORY

When listening to reproduced sound in the laboratory, the listening room should be acoustically designed. The best possible solution is an anechoic room, which has 3D multichannel setup, as in Fig. 4 on the left. This room has 16 loudspeakers as explained in Section 2.1. However, anechoic rooms are very expensive and they are not always available. Therefore more practical rooms are needed and we have solved this problem by converting one normal office room to a listening space. The room is acoustically treated with a lot of absorptive materials, such as a soft sofa, mineral wool and some absorptive wedges that were left over in the renovation of an anechoic room. The room is not anechoic, but reverberation time is less than 0.2 seconds at mid frequencies and the room does not have pronounced modes at low frequencies. In addition, 14 reproduction loudspeakers (see Section 2.2) are at 1.0 m distance from the head of the listener, thus the reverberation of the room is not affecting to the listening experience.

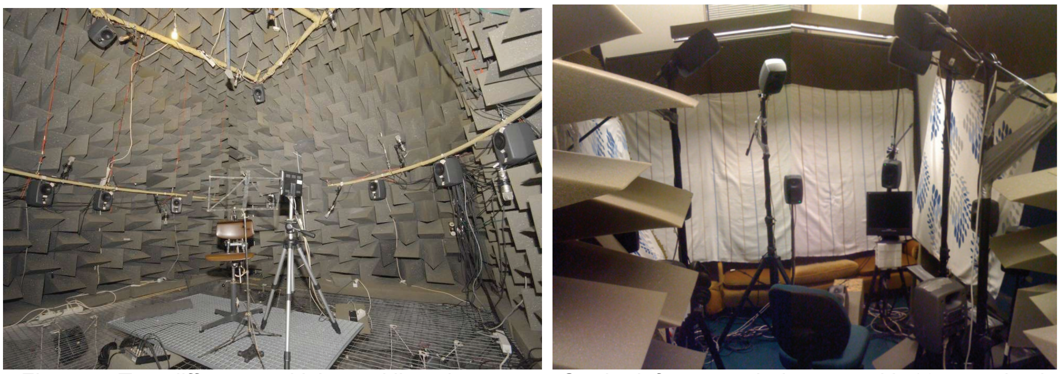

Figure 4: Two different multichannel listening rooms. On the left an anechoic room with 16 reproduction channels. On the right an acoustically treated normal office room with 14 reproduction channels.

There is no optimal number of reproduction loudspeakers. The number and locations of loudspeakers depend on the spatial sound rendering algorithm and they should be optimal for the particular algorithm. The SIRR method renders the sound with defined direction (based on sound intensity analysis) with vector base amplitude panning (VBAP) [18] and the diffuse sound is distributed to all loudspeakers. Therefore, the loudspeaker locations should be optimized for VBAP, keeping in mind the properties of human spatial hearing. In a concert hall the sound reaches the listener from the upper hemisphere and possible in front below the ear level. Therefore, our current loudspeaker setup has 8 loudspeakers at ear level (45 degree intervals), four at elevation of 55 degrees (90 degree intervals) and two loudspeakers 40 degree below ear level at azimuth angles -22 and 22 degrees.

4. CONCLUSIONS

This paper discusses the recording of acoustics of concert halls for listening to the sound later in the laboratory environment. Such methodology is needed to perform subjective evaluations of concert hall acoustics. The presentation is mainly describing the technology that we have been using in our recent studies. Although many other techniques can also be used, our choices are based on the careful thinking of the main problem; how to capture and reproduce the sound of a symphony orchestra. Here, an overview of the applied methods is presented and more detailed reasoning why these techniques are used can be found from the original articles.

5. ACKNOWLEDGMENTS

The research leading to these results has received funding from the Academy of Finland, project nos. [218238 and 140786] and the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement no. [203636].

6. REFERENCES

[1] T. Lokki, H. Vertanen, A. Kuusinen, J. Pätynen, and S. Tervo, “Auditorium acoustics assessment with sensory evaluation methods,” In *the International Symposium on Room Acoustics (ISRA2010)*, Melbourne, Australia, August 29-31 2010.

[2] T. Lokki, J. Pätynen, A. Kuusinen, H. Vertanen, and S. Tervo, “Concert hall acoustics assessment with individually elicited attributes”, *Journal of the Acoustical Society of America*, 130, 835–849 (2011).

[3] T. Lokki, J. Pätynen, and N. Zacharov, “Concert hall acoustics assessment with sensory evaluation – tools and practices,” in *The IOA International Conference on Auditorium Acoustics*, Dublin, Ireland, 2011, pp. 35–44.

[4] J. Pätynen, and T. Lokki, “Directivities of symphony orchestra instruments,” *Acta Acustica united with Acustica*, 96, 138–167 (2010).

[5] ISO 3382-1:2009, “Acoustics – measurement of room acoustic parameters – part 1: Performance spaces,” (International Standards Organization, Geneva, 2009).

[6] J. Pätynen, S. Tervo, and T. Lokki, “A loudspeaker orchestra for concert hall studies,” in *The Seventh International Conference On Auditorium Acoustics*, 45–52 (Institut of Acoustics, Oslo, Norway, 2008), also published in *Acoustics Bulletin,* 34(6), 32–37 (2009).

[7] J. Pätynen, V. Pulkki, and T. Lokki, “Anechoic recording system for symphony orchestra”, *Acta Acustica united with Acustica*, 94, 856–865 (2008).

[8] T. Lokki, J. Pätynen, S. Tervo, S. Siltanen, and L. Savioja, “Engaging concert hall acoustics is made up of temporal envelope preserving reflections,” Journal of the Acoustical *Society of America,* 129, EL223–EL228 (2011).

[9] T. Lokki, J. Pätynen, “Lateral reflections are favorable in concert halls due to binaural loudness,” *Journal of the Acoustical Society of America,* 130, EL345-EL351 (2011).

[10] M. Gerzon, “Periphony: With-height sound reproduction,” *Journal of the Audio Engineering Society*, 21, 2–10 (1973).

[11] V. Pulkki, “Spatial sound reproduction with directional audio coding,” *Journal of the Audio Engineering Society,* 55, 503–516 (2007).

[12] J. Vilkamo, T. Lokki, and V. Pulkki, “Directional audio coding: Virtual microphone based synthesis and subjective evaluation,” *Journal of the Audio Engineering Society*, 57, 709–724 (2009).

[13] J. Merimaa, “*Analysis, synthesis, and perception of spatial sound – binaural localizaion modeling and multichannel loudspeaker reproduction*”, Ph.D. thesis, Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, report 77 (2006).

[14] A. Farina, “Simultaneous measurement of impulse response and distortion with a swept-sine technique,” presented at the 108th Convention of the Audio Engineering Society, *Journal of the Audio Engineering Society* (Abstracts), vol. 48, p. 350 (2000 Apr.), preprint 5093.

[15] J. Merimaa and V. Pulkki, “Spatial impulse response rendering I: Analysis and synthesis”, *Journal of the Audio Engineering Society*, 53, 1115–1127 (2005).

[16] V. Pulkki and J. Merimaa, “Spatial impulse response rendering II: Reproduction of diffuse sound and listening tests”, *Journal of the Audio Engineering Society*, 54, 3–20 (2006).

[17] J. Pätynen, S. Tervo, and T. Lokki, “Simulation of the violin section sound based on the analysis of orchestra performance”, in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA),* (New Paltz, New York, USA) (2011), pp. 173–176.

[18] V. Pulkki, “Virtual sound source positioning using vector base amplitude panning,” *Journal of the Audio Engineering Society*, 45, 456–466 (1997).