ICSV13 - Vienna

The Thirteenth International Congress on Sound and Vibration

Vienna, Austria, July 2-6, 2006

LOW–LATENCY VIRTUAL ACOUSTICS FOR LIVE MUSIC PERFORMANCE AND RECORDING

Wieslaw Woszczyk

Centre for Interdisciplinary Research in Music, Media and Technology (CIRMMT)

Graduate Program in Sound Recording

Schulich School of Music of McGill University

555 Sherbrooke Street West, Montreal, QC, Canada H3A 1E3

Email: wieslaw@music.mcgill.ca

Abstract

A system for generating virtual room acoustics with low-latency response and immersive surround sound with height is described. The system has been primarily developed for shared reality applications in distance music teaching and collaboration, where low-latency is required for successful interactive communication between distant participants. It is also used in a sophisticated recording and performance studio where artists can rehearse, perform and be recorded in virtual acoustics of rooms measured in other locations. The method of reconstructive acoustics uses specifically captured multichannel high-resolution impulse responses (IR) of the original concert halls, in multiples of six. To ensure a detailed characterization of the diffuse field of a room, the excitation of the room is done with a distributed sound source whose power response is flat across almost all frequencies within the audible range. High acoustic power output of the source used to excite the room and slow-swept sine wave stimulus ensure wide dynamic range of the captured impulse responses approaching 120 dB. The room response is recorded in four layers of 6 microphone channels, each layer at a different elevation, all later convolved with the direct signal and projected around and above the performer using four layers of surround loudspeakers. The convolution process uses dedicated signal processing hardware capable of 6 real-time convolutions of 15-seconds-long IRs with 10 ms latency. Time-variant impulse responses can be rendered using overlapping convolutions. The goal is to provide the performer with the most representative interpretation of the room that can best approximate the subjective impression of the original space. Examples will be given describing virtual acoustic spaces developed to make recordings of Haydn's fortepiano and harpsichord sonatas for the upcoming SA-CD release.

INTRODUCTION

Recording music on location in historical rooms is typically a difficult and compromising task because most such spaces have been built in downtown areas at a time when environmental and industrial noise were not an issue. Consequently, these spaces have poor isolation from high background noise making them often unusable for recording. An improvement over this situation is possible when music performance is recorded in a quiet recording studio where stage and room acoustics of desired space are reconstructed electronically and adjusted to suit the needs of the performers, the music, and the recording.

Recently, with the advent of fast computers and signal processing devices, and with the introduction of high-resolution multichannel recording and digital sound field synthesis, it becomes feasible to capture a very detailed acoustic signature of the room and to recreate the complete acoustic response of that room in another space [1],[3],[5]. Acoustic properties of rooms can be auditioned and archived for posterity, thus preserving and protecting our cultural heritage from decay or loss [2]. Performing artists can practice their music in virtual rooms in preparation for concerts or recordings. These captured rooms can be used in music teaching, recording for film, records or television, in videoconferencing, and in simulations of acoustical reality in theme parks and games. Moreover, they can be the subject of studies in performance practice, history of art and architecture, development of period and modern instruments, musicology, and composition.

VIRTUAL ROOM ACOUSTICS FOR MUSIC PERFORMANCE AND RECORDING

This paper presents a comprehensive methodology for capturing the musical "fingerprint" of rooms and acoustic spaces for their subsequent use in music performance and recording (Figure 1). By recording the acoustic response of the room to an audio test signal, we capture the unique spatial, temporal, and dynamic characteristic of that acoustic space and register its distinct sonic character activated between the sound-source located in that space and the microphone-array set up to capture it. The quality and layout of the source and the receiver in the room determines the measured and perceived quality of the room sample.

Perceived Acoustic Presence of the Architectural Space

Most methods of room measurement use acoustic excitation of the room with a single loudspeaker located in one unique point of the room, and recording of the acoustic response of the room captured with a single microphone placed elsewhere in the room. The impulse response captured this way represents a unique contribution of the room between these two points but the single response is not representative of how this room may sound to a human listener.

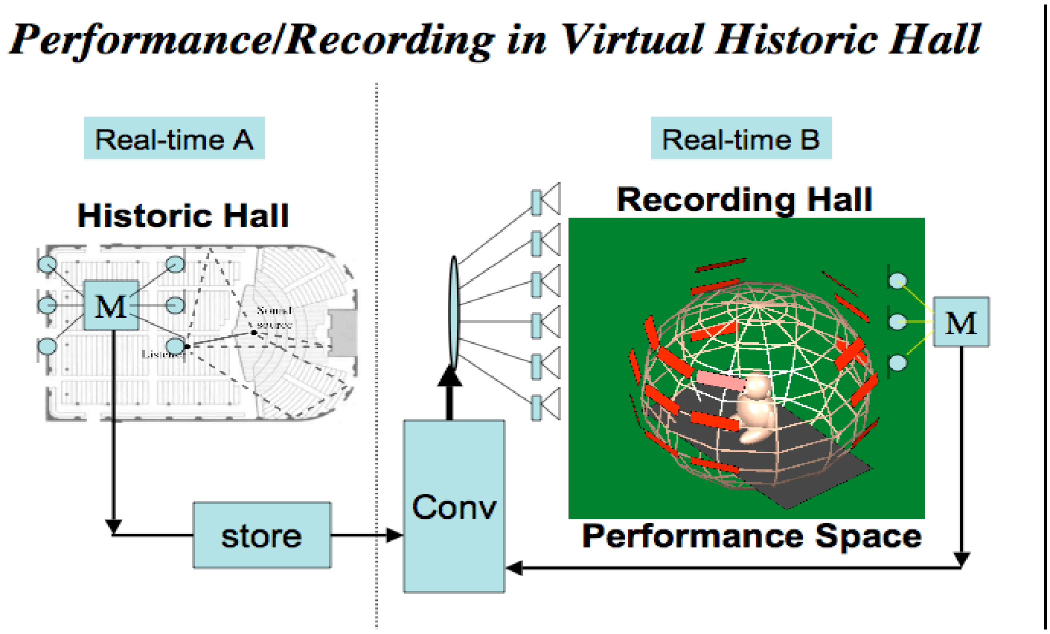

Figure 1 – Schematic diagram of virtual reconstruction of room acoustics for music performance and recording in a studio using multiple impulse responses captured in a historic hall, and fast digital convolution with low latency.

Of interest to an acoustician, a musician or a recording producer, is how a given room responds to a musical signal radiated by a musical instrument or a group of instruments. Musical source is usually a complex multidirectional radiator and often consists of several sources distributed over an area with varying distances to the floor, and frequently moving in space during the production of the sound. A violin, voice or a clarinet will not usually stay fixed to a firm position but move ever so slightly during performance.

We should, therefore, make a distinction between the room response recorded for the purpose of measurement (for example, to verify the effect of acoustical treatment applied to one of its surfaces), and the response captured to reconstruct the correct subjective impression of the room in the mind of a listener in a virtual room. The method of capture and reproduction of room response related to the later case, that of subjective reconstruction of virtual room needed for music performance and recording, is described below.

ROOM RECORDING METHOD

Room's acoustical properties should be made clear by a strategic distribution of properly chosen sound sources, and by using a microphone system able to capture and articulate the perceived characteristics of the room in reproduction over loudspeakers.

Sound Stimulus

Impulse stimulus of short-duration, and high-energy content cannot be used to trigger the room response without overloading the loudspeaker system [4]. Large rooms require forceful excitation using substantial acoustic power mainly at low frequencies where rooms' ambient noise is usually high. A slow logarithmic frequency sweep of *Pinguin* HDIR Creator satisfies acoustic power requirements and provides the extra benefit of isolating harmonic distortion from the impulse response in post-processing.

The Sound Source

The source employed to trigger the room response consists of a set of independent radiators spread over an area of the room to represent a distributed musical source (for example a grand piano, or a small chamber group consisting of independent instruments). The individual radiators are placed at varied heights between the floor and the elevation of five feet above the floor. All sources together cover an extended audio bandwidth from 10Hz to 40kHz, have high acoustic output and low distortion, and flat diffuse-field response. It is essential that the acoustic power radiated into the room at slow sine-sweep rates is well balanced, therefore all radiators have flat power response in the dominant frequency range of their contribution to the source signal. Two custom woofers (made by Tymphany™) have 10" cone drivers each housed in a separate 1 cubic foot closed-box enclosure and powered by a 1000W amplifier. The output of the woofers contributes substantially up to 200Hz, and above that two tetrahedral-spheres with 12 cone loudspeakers (50mm dia.) in each cover the frequency range from 200Hz to 2000kHz. Two ribbon bi-directional loudspeakers in custom frame enclosures cover the range from 1000Hz to 10kHz (Figure 2), and two ribbon super-tweeters cover the range above up to 50kHz. All these loudspeakers are powered with 100W amplifiers.

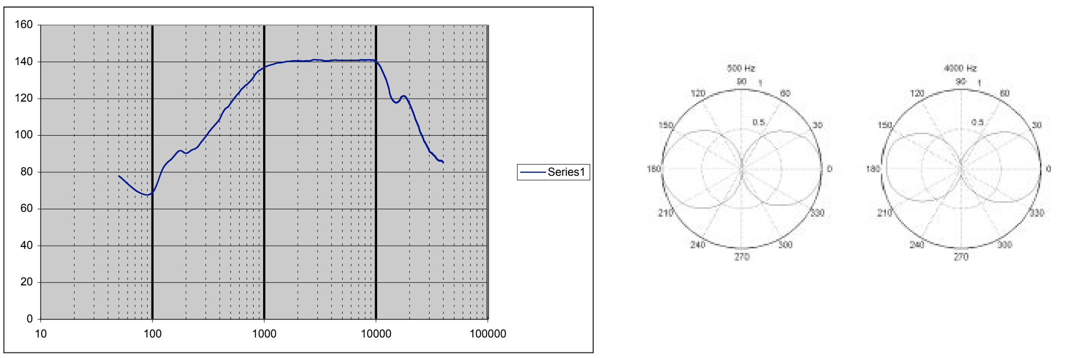

Figure 2 –Free-field frequency response curve and polar radiation pattern of the flat ribbon loudspeaker used to trigger room response in mid-band (1kHz-10kHz).
The loudspeaker has flat free-field and flat diffuse-field response.

Low-frequency sine-sweeps are best radiated in some rooms by woofers having ultra-low distortion and extended low-frequency response. These loudspeakers (Velodyne™ Digital Drive) employ a proprietary sealed digital accelerometer-based system used to measure, compare and control cone movements. The motion feedback-loop technique substantially reduces distortion in woofers that require large cone excursions and much power to drive them in order to reproduce low frequencies.

Source Layout

The individual radiators are placed as a group over a small area of the room floor or stage as if they were a group of chamber music performers. The bass drivers mark the widest spread of the overall source from left to right. The spherical and bi-directional drivers are distributed within the spread in depth, azimuth, and elevation, to represent a complex multidirectional sound source. The flat ribbon drivers are directed towards the more distant areas of the room near the ceiling.

This layout of the radiators promotes directional diversity of the sound source and improves room's spatial response to the excitation, plus helps to compensate for the limited spatial selectivity of the microphone system. Such spatial pre-emphasis is allowed when capturing rooms alone, without the actual musical source and the performers being present. Room's spatial image can be made narrower, if needed, in virtual room reconstruction, but widening a narrow image is not as easy to achieve.

The Microphone System

For each room, several 6 (or 8) -channel impulse responses are recorded, each at a different position in the room, different height above the floor, and location to the source. Because our virtual acoustics laboratory uses 24-channel loudspeaker system to project sound from all directions around the listener (from around, above, and below), there are typically four 6-channel impulse responses recorded for each room, for a single location of the source and the microphone in the room. Thus, we capture 24 channels of spatially unique impulse responses for each source-to-receiver situation in the room to reproduce an immersive virtual room.

Microflown Probes

To achieve greater definition of the low frequency room response in recording we use Microflown™ probes (Figure 3), true figure-of-eight-pattern velocity microphones having extended response down to the lowest audible frequencies, low noise and high output. Unlike pressure-gradient microphones, velocity probes do not measure acoustic pressure at two points to derive a pressure gradient. When particle velocity is present, acoustical particle velocity sensors measure the temperature difference of the two closely separated and heated platinum wire resistors, and quantify particle velocity from the temperature measurement. Their cosine directional selectivity can be used to change the ratio between early reflections and the diffuse sound since only the 1/3 of the power in the diffuse sound field is measured with the particle velocity probe. This feature can correct for any unfavourable room characteristics such as excessive room mode build up, uneven spectral balance between the modes, excessive loudness of late arriving reflections, room or machine noise, etc., and thus contribute to a more acceptable presentation of the room in a virtual environment.

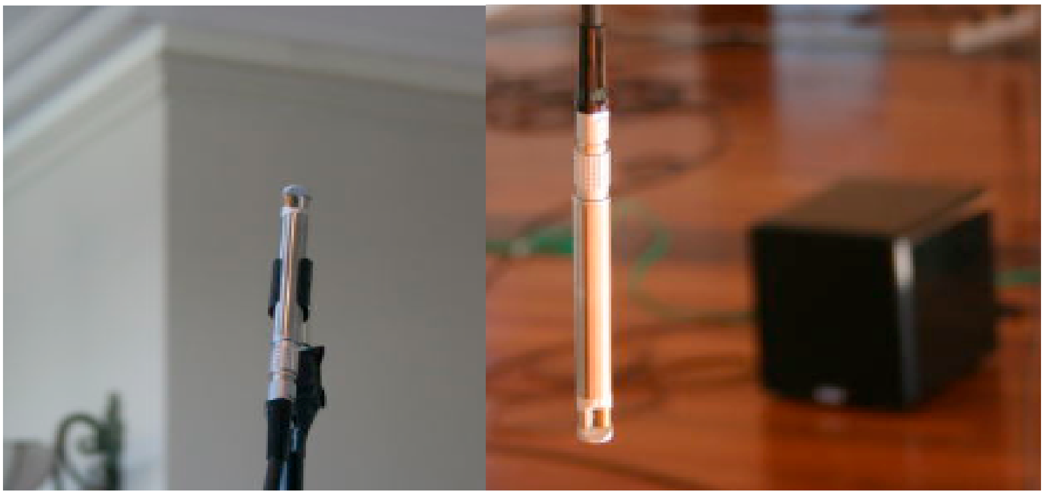

Figure 3 – Microflown particle velocity probe used to capture selective response of the room at low frequencies (foreground image). Digital Drive Velodyne woofer used to excite the room with low-distortion sine-sweep signal at low frequencies (background, right side).

ROOM REPRODUCTION METHOD

The auditory display system is configured (see Figure 4) as spherical loudspeaker array consisting of 6 low-frequency drivers (ranging from 25 to 300 Hz) and 96 mid- and high-frequency drivers (ranging from 300 to 30,000 Hz). The lower-frequency drivers are placed at standard locations for the 6 main speakers in surround sound reproduction (the speaker angles in degrees relative to the median plane are 0°, ±30°, ±110°, and 180°). The upper-frequency drivers are dipole radiating, full-range electro-dynamic ribbon transducers placed 4 units wide in 24 panels (two loudspeaker per audio channel, two channels per panel) in 24 locations on the surface of an imaginary sphere of 4-meter diameter. Besides 6 locations at extreme high elevation, the spatial organization of the upper-frequency drivers is defined by 3 planes at elevation angles of –15°, +25°, and +45° degrees relative to the horizontal plane. Within each plane of differing elevation angle, 6 speakers are placed at azimuth angles matching those of the 6 lower-frequency drivers (again, 0°, ±30°, ±110°, 180° relative to the median plane). The perceived benefits of the height channels contribute to the augmented sense of immersive presence within the virtual acoustic environment.

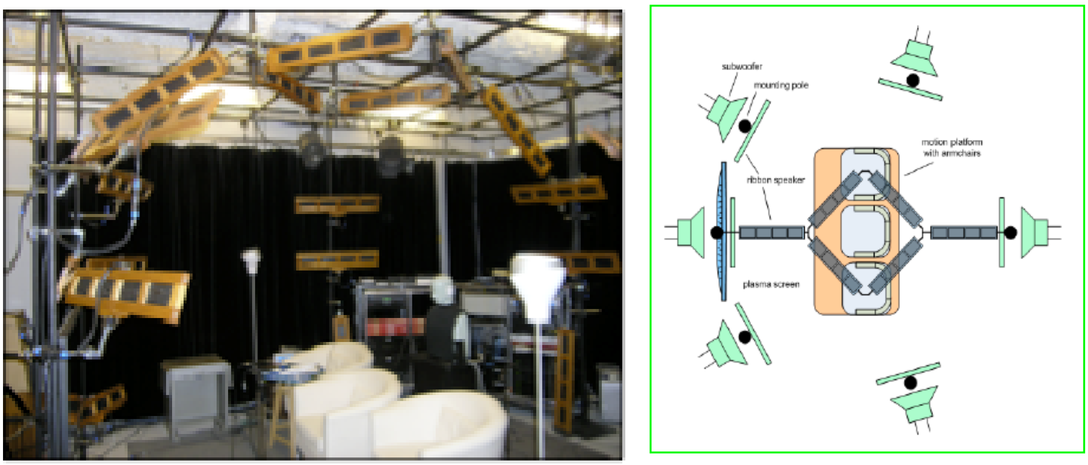

Figure 4 – Immersive acoustic environment for performance and recording using 24 channels of audio reproduction in 4 levels of elevation and 6 directions around the center.

Convolution Engine

When considering an optimal reproduction of room response for a musician performing in a virtual room, latency needs to be reduced as much as possible because any delay inhibits the interaction between the musician and the environment. To ensure sufficient processing power for real-time applications, the virtual room system employs massive digital audio processor designed and fabricated by Weiss Engineering in Switzerland. Its efficient FFT segmentation permits to achieve real-time convolution of 6 impulse responses, each 15 seconds long, at 96kHz sampling frequency with 24 bit word-length resolution. Each input uses two Sharc DSP engines with 16 Mega Words of fast dedicated onboard memory. Floating point 32 bit architecture ensures sufficient dynamic range for most demanding applications and post-processing. A host internal 32bit processor provides the control of DSP array and interface to external controllers via Ethernet, MIDI, RS232 or Flash Card. The hardware allows for maximum of 8 input and output channels, however multiple units can be synchronized and used in parallel, with linked control using a common GUI interface, allowing for a large number of convolution channels simultaneously. A high-speed data interchange buss links all DSP modules allowing for quick uploading of impulse responses for processing. Four interface cards (AES/EBU, MADI, Analog, Firewire) provide connectivity to other digital and analog equipment.

CONCLUSION: HAYDN PROJECT

Reconstruction of room acoustics for music performance and recording has to go beyond simple measurement of a standard impulse response, and must lead towards the perceptual reconstruction of the room for the intended musical purpose. Music communication (by means of performance or recording) requires an insightful and convincing interpretation of the score presented within an appropriate acoustical context. An example here is the Haydn Project undertaken at McGill University in Montreal where Haydn's solo sonatas for pianoforte and harpsichord will be performed and recorded using virtual acoustics of rooms existing in Austria and Hungary where Haydn had actually composed and performed the music (Figure 5).


Figure 5 – Esterházy Castle in Fertód (Hungary) where Haydn composed and performed his music for the Prince. Microphone and loudspeaker arrays in magnificent rooms.

ACKNOWLEDGMENTS

The help of the following persons is graciously acknowledged: Ronald Aarts, Tom Beghin, Hans-Elias de Bree, Ieronim Catanescu, Chris Chafe, Trudy Erickson, Martha de Francisco, Charles Gagnon, Ruchi Goel, Michael and Eben Grace, Bruce Hall, Ken Kantor, Ralph Kessler, Minoru Kobayashi, Dr. Wolfgang Kuzmits, Jean Langlais, Gabriela Liber, Harald Mayer, Chris Price, Márton Rácz, Derk Reefman, Dr. Walter Reicher, Daniel Weiss, Gerhard Winkler, Jorg Wuttke. This research is supported by Valorisation-Recherche Québec of the Government of Québec and by Social Sciences and Humanities Research Council of Canada.

REFERENCES

[1] Farina, A., and Tronchin, L., "Advanced Techniques for measuring and reproducing spatial sound properties of auditoria", International Symposium on Room Acoustics: Design and Science 2004. Awaji Yumebutai International Conference Center, Hyogo, Japan. April 11-13. 2004.

[2] Gerzon, M., "Recording Concert Hall Acoustics for Posterity", J. Audio Eng. Soc., Vol. 23, No.7, p.569, (1975).

[3] Hulsebos, E., "Auralization using Wave Field Synthesis", Ph.D. Dissertation, Delft University of Technology, October 4, 2004. The Netherlands.

[4] Stan, G-B., Embrechts, J-J., and Archambeau, D., "Comparison of Different Impulse Response Measurement Techniques", J. Audio Eng. Soc., Vol.50, No.4, 2002 April, pp.249-262.

[5] Ueno, K., Tachibana, H., and Kanamori, T., "Experimental study on stage acoustics for ensemble performance in orchestra", the Proceedings of the 18th International Congress on Acoustics, Kyoto, April 4-9, 2004, Japan.