19th INTERNATIONAL CONGRESS ON ACOUSTICS
MADRID, 2-7 SEPTEMBER 2007

INTERACTIVE MULTI-CHANNEL AURALIZATION WITH CAMERA-BASED TRACKING

PACS: 43.55.Ka

Kajastila, Raine[1]; Lokki, Tapio[1]; Savioja, Lauri[1]
[1]Helsinki University of Technology, Telecommunications Software and Multimedia Laboratory; P.O.Box 5400, FI-02015 TKK, FINLAND; {raine.kajastila | tapio.lokki | lauri.savioja}@tkk.fi

ABSTRACT
A multi-channel auralization system, in which a user can interact with a virtual acoustic environment without carrying any equipment, is presented. The user, who is standing in a real room equipped with several microphones and loudspeakers and a camera-based tracking system, can hear the surrounding virtual geometry. Microphones are used to record sounds from various directions and surrounding loudspeakers reproduce the auralization of a 3-D model. The early reflections from the virtual 3-D model are computed with a beam tracing method. The sound emitted by the user is processed with appropriate transmission and reflection losses and reproduced in the spatially correct location. The late reverberation is implemented with a statistical time-variant algorithm and reproduced from all the loudspeakers. The time-variance is needed to minimize the effect of feedback which is an evident problem of the system. In addition, the late reverberation is reproduced such that possible non-diffuse reverberation conditions are simulated by applying different gain values in different reproduction directions. Finally, some perceptual evaluation results and application scenarios of the implemented system are presented.

INTRODUCTION
Auralization is often utilized to render the room acoustic prediction results audible with either binaural or multi-channel reproduction techniques. The sound source in a model is typically a static point source and the stimulus signal is anechoically recorded speech or music. However, in some cases it would be nice if the listener her/himself could be the sound source. In addition, a moving sound source is sometimes a needed feature. In this paper we present such a system in which the user does not need any extra equipment for the interaction with the virtual space. The user can turn freely and shout to any direction and consequently he/she hears the response of the virtual, modeled space. In addition, moving the listener in a virtual model is implemented intuitively by body movements.

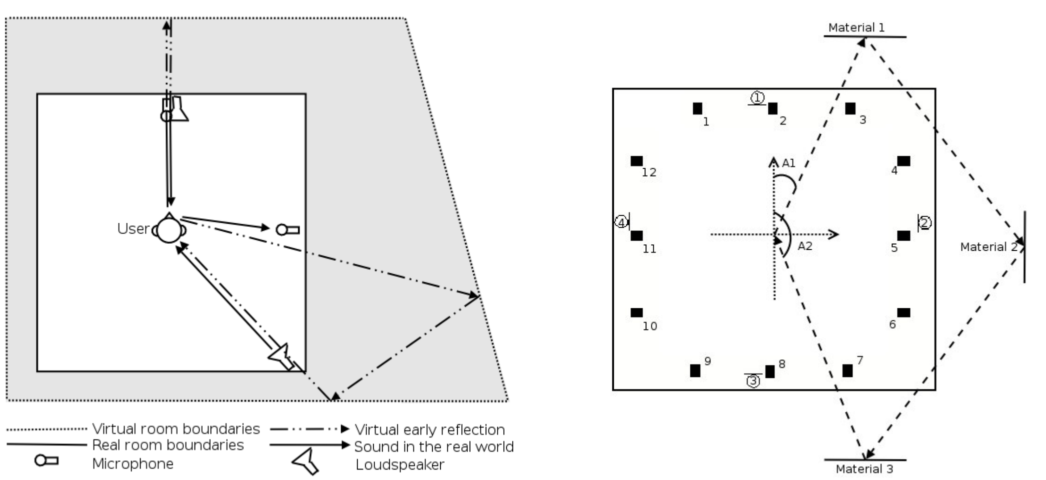


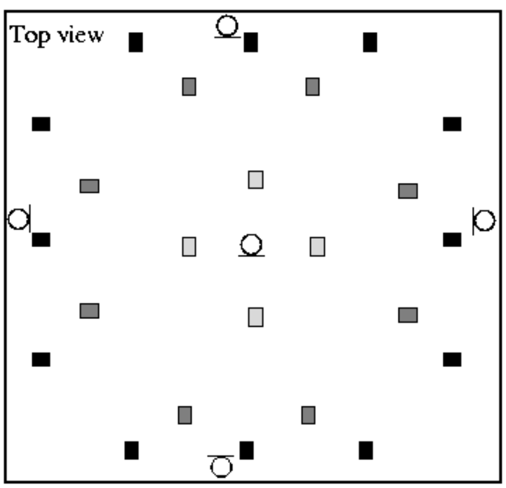
Figure 1: General concept of auralizing early reflection paths in an interactive environment (left). Two-dimensional example of one auralized reflection path (right).

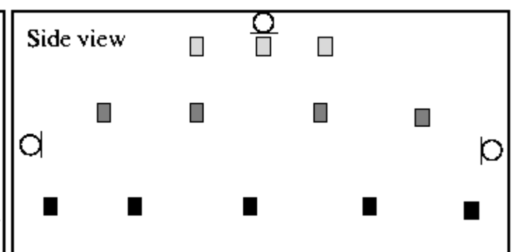
HARDWARE FRAMEWORK OF THE SYSTEM

Figure 1 explains the idea of the implemented system. The user, who is standing in the middle of a room, can hear the surrounding virtual geometry. Microphones are used to record the sound emitted by the user, and surrounding loudspeakers are used to reproduce the virtual acoustics.

The auralization system is built in the Odeion hall at the Helsinki University of Technology. It is a multipurpose facility (volume 790 m^3, dimensions 12x11x6m) designed to be used both as a lecture hall and a performance theatre. It is also an experimental space for virtual reality technology and applications. The walls and the ceiling are designed to be very absorbing while the floor is concrete. The wall structure consists of a 50 mm mineral wool mounted about half a meter in front of the concrete wall, imitating a suspended ceiling structure with an air gap. The table in Fig. 2 points out that the reverberation time of the Odeion is short compared to the rooms with the same volume and the computed absorption coefficients are high.

The hardware of the implemented system consists of microphones, loudspeakers and computers. The applied five microphones are low-noise condenser microphones (Shure KSM109) having the cardioid directivity pattern. In spite of the used damping materials, the wall structures are not totally sound absorbing at low frequencies. Therefore, the null of the cardioid is pointing to the wall to minimize low-frequency boost. The loudspeakers (Genelec 1029A) are mounted so that twelve are at ear level, eight are about 40 degrees elevated and four are almost on the ceiling as illustrated in Fig. 2. The loudspeakers are approximately on the surface of a hemisphere, and they are virtually positioned using delays exactly equally distant from the center of the room. In addition, the loudspeaker gains are adjusted to obtain equal sound pressure level to the center of the room from each loudspeaker. A more detailed description of the Odeion and the equipment has been published earlier [1].





Band	125	250	500	1k	2k	4k	8k
RT	0.52	0.38	0.34	0.25	0.23	0.21	0.18
Abs.	0.80	0.87	0.92	0.99	0.99	0.99	0.99

Figure 2: The positions of the loudspeakers and the microphones. Input signals from the 5 microphones are routed to one computer dedicated to signal processing which also controls the output signals to the 24 loudspeakers. Table on the right shows measured reverberation times of the test space when auralization system is turned off. Top row: octave band [Hz]. Middle row: measured reverberation times [s]. Bottom row: Theoretical absorption coefficients, computed with the Sabine equation from the measured RTs.

IMPLEMENTED INTERACTIVE AURALIZATION SYSTEM

The presented system consists of a chain of tools, developed in the Uni-verse project [2], as outlined in Fig. 3. Three-dimensional information of the virtual geometry and the user location in the virtual world is stored in a Verse server [3], which allows several applications to share the same 3-D geometry. The geometry is loaded to the server as Vector Markup Language (VML) file and the user’s location is updated with software connected to a tracking camera. Thus, the acoustic simulation software gets the geometry as well as the listener and source information from the Verse server and calculates the reflection tree using beam-tracing. Reflection tree is forwarded to the audio renderer that controls the microphones and loudspeakers.

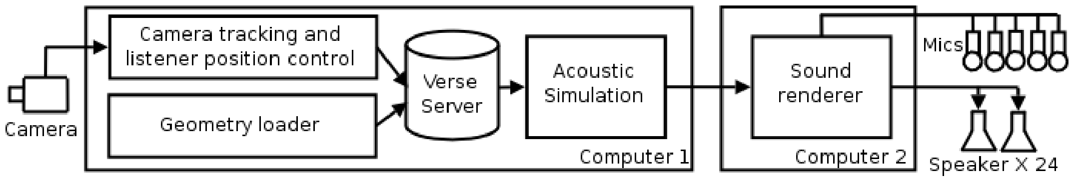


Figure 3: The main building blocks to create interactive auralization with multi-channel reproduction. In addition, the distribution of computational load between two computers is presented. The two Linux computers are interconnected using an Ethernet network.

The first computer is used to handle the user input via a cheap web camera and to run the acoustic simulation in the desired geometry. Another computer is dedicated for sound rendering (i.e. auralization). This computer is equipped with an RME DIGI9652 sound card offering 3 ADAT digital I/O. The five microphones are connected to an 8-channel AD converter which feeds the controlling PC via one ADAT interface. The output from the sound card is routed to three 8-channel DA converters that feed the 24 active loudspeakers directly.

MOVING THE USER WITH CAMERA-BASED TRACKING

The position of the listener is tracked with a camera-based system. A webcam is mounted to the edge of the ceiling of the Odeion and the captured frames are processed to find movement in successive frames. The center of gravity of movement is used to pinpoint the user position, as illustrated in Fig. 4. This simple camera tracking is implemented with already existing modules of PureData [4] and it allows the tracking of the user in about 3m x 3m area. The tracking area can be extended, e.g., with a wide angle lens.

When using the system, the user stands in the middle section of the room and he/she sees a 3-D geometry on a screen. The geometry is displayed from a bird's eye view, as illustrated in Fig. 4. When the user moves he/she can see how the user position changes in the projected image. The system has two modes for changing the listener's position in the virtual space. In the first mode the user's location is exaggeratedly mapped to the virtual location. In other words when the user moves one meter in some direction the virtual listener moves, e.g., 5 meters in the virtual space. In the second mode camera tracking is used to find out if the user leaves the center of the room. The virtual listener starts moving to this particular direction where the user has moved from the center. The distance from the center defines the speed that the virtual listener moves. To stop the virtual listener user has to return to the center.

The drawback of the first mode is that the virtual listener cannot move in large area. When virtual area is large, even the small movement of the user moves the virtual listener a long distance. It is hard to place the virtual listener to a desired point. In addition, with the first mode the user can easily end up far away from the sweet spot of the listening room, which is not good for optimal sound reproduction. In informal tests, the second mode is found to be more practical. The user can always listen to the auralization of his/her own voice in the sweet spot. Naturally, the second mode also enables the moving in virtual geometry of any size.

For visualization purposes, the ceiling of a virtual geometry has to be transparent from outside. This is possible by using only one sided polygons in the ceiling. Transparent ceiling allows the user to view a larger part of the geometry a same time. When using only the camera tracking, the moving is restricted into one horizontal level. No method for moving the listener up or down has yet been developed, although it would be easy to implement motion tracking, where, for example, squatting down and jumping moves the virtual listener up and down. This would enable moving also in a multi-storey geometry. Naturally, using many cameras from different directions would give more freedom in user interface design and all moving metaphors currently applied in virtual reality applications could be used [5].

ROOM ACOUSTIC MODELING AND AURALIZATION

The modeling of room acoustics is done in real time with a well-know concept in which early reflections and late reverberation are simulated separately [6]. The early reflection paths are searched with a beam-tracing algorithm, the implementation of which has been presented earlier [2]. The resulting early reflection parameters, such as distance, incoming angle, and surface materials, are sent to another computer for the auralization process which is an adapted version of the one introduced by Kajastila et. al. [2]. The sound rendering process is depicted in Fig. 5.

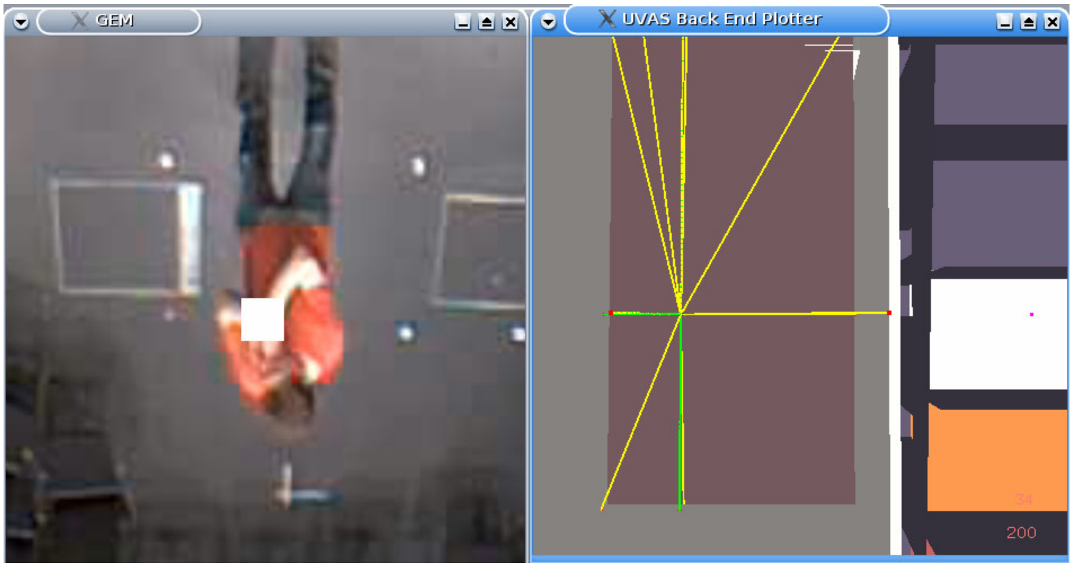


Figure 4: On the left, one captured frame applied in tracking the user's position. The white square is the center of gravity of movement. The right figure presents an example geometry and a few reflection paths used in auralization.

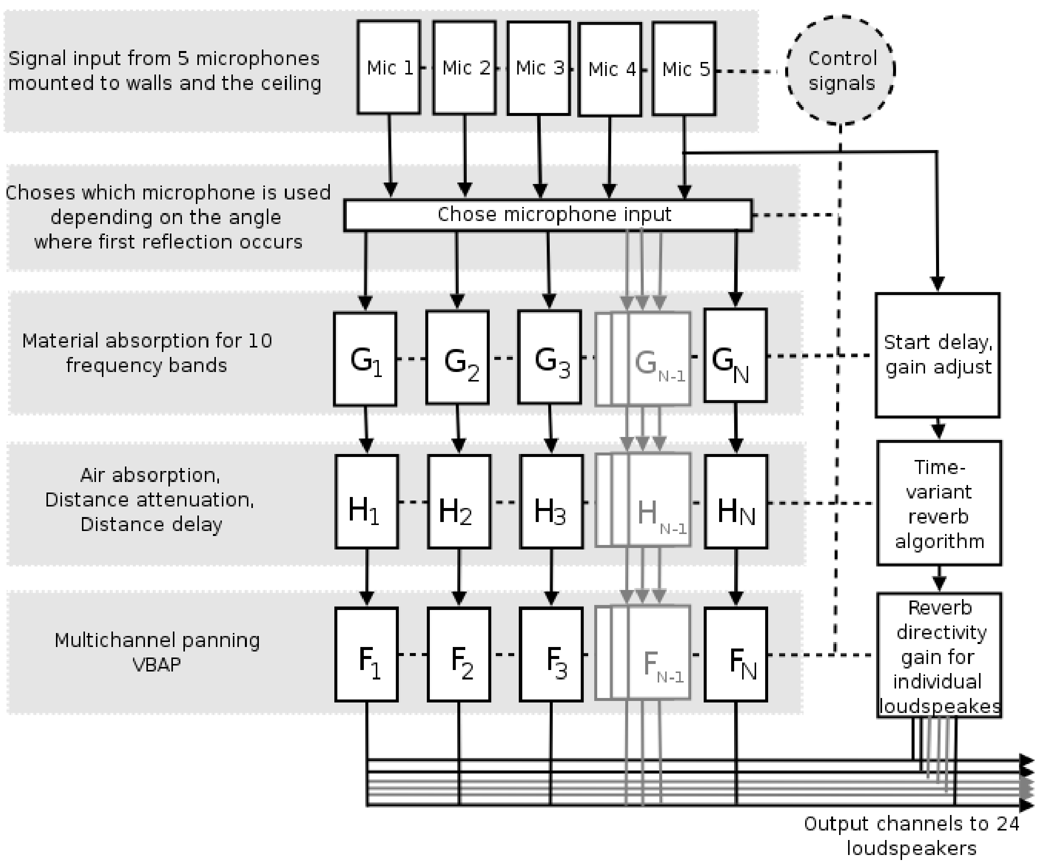


Figure 5: The block diagram of the implemented sound renderer.

Sound pick-up with microphones

As depicted in Fig. 1 a novel method is implemented to record sound from the same position where the auralization is listened to. Thus, the listener can be a sound source by him/herself and he/she can evaluate the surrounding virtual space by speaking, singing, clapping hands or making any noise. In other words, the acoustics simulation module calculates the early reflection paths that leave and come back to the user's position. The system tries also to reproduce the directivity of the source. Even though the user can turn freely in the room, there are always microphones surrounding him/her. To implement a simple model for sound source directivity the azimuth and the elevation angle of each reflection is applied to specify the input microphone (e.g. A1 in Fig. 1). The single microphone pick-up for each reflection is also important to prevent any additional coloration of sound. If multiple microphones would be applied for one reflection, the different propagation delays from the user to the microphones could produce comb filtering to the signal.

Rendering of early reflections

The length of each early reflection path is used to calculate air absorption, distance attenuation and delay. Since the sound has already traveled few meters before reaching the microphone, the early reflection path length is shortened accordingly. Material absorption data for 10 octave bands is used to implement material absorption in the boundaries of the virtual geometry. The material absorption filters are implemented by dividing each microphone input signal into 10 frequency bands and applying one gain value to each band. This implementation, originally introduced by Deille et. al. [7], is computationally efficient when many reflections are to be rendered. Note that only early reflections are produced because the user him/herself is the sound source, thus no direct sound is needed to be simulated.

The reproduction direction of the processed sound is defined by the azimuth and the elevation angles (A2 in Fig. 1) of the last reflection occurred before reaching the virtual listener. Loudspeaker setup using Vector Base Amplitude Panning (VBAP) [8] is applied to position the sound to the correct spatial direction. For example, in Fig. 1 the loudspeakers 7 and 8 are used to reproduce simulated third order reflection.

Late reverberation

To minimize the feedback and avoid instability of the system a time-variant late reverberation algorithm introduced by Lokki et. al. [9] is applied. Input signal for the reverberation is taken only from the ceiling microphone to avoid comb filtering effect which can easily occur if multiple microphone signals are fed to the reverberation algorithm.

Changing parameters automatically allows adaptation of reverberation in different rooms that the geometry might contain. One important parameter is the start delay, which specifies the point in time when the reverberation is played after the early reflections. The changing lengths of the early reflection paths are kept in a table and an average value of the longest ones are used to adjust the start delay. The average value can also be used to set the reverberation gain to match the level of the early reflections.

Traditionally, the good reverberation algorithms generate diffuse reverberation that surrounds the user equally from all directions. However, in the real world there are many places where the late reverberation is not diffuse at all. It is easy to hear that in a narrow corridor or in a corner of a bigger hall the reverberation has a "dominating incoming direction", or at least reverberation is not heard from all directions. The area of such non-surrounding reverberation has not been widely studied in auralization research, since usually the late reverberation is assumed to be diffuse.

In the presented implementation non-diffuse reverberation is simulated by adjusting gains of individual loudspeakers. Indeed, the reverberation is still diffuse by nature, but the louder gains in some directions make the perceived sound field non-diffuse. The non-ideally diffuse reverberation algorithm is implemented in the sound renderer, and since it receives only the reflection trees, the shape of the surrounding space is approximated from directions and lengths of incoming reflections. This novel algorithm works as follows.

First, the early reflections are grouped according to their incoming angle. The longest reflection length that is coming from a certain loudspeaker within a certain boundary area is stored. This is done for each 24 loudspeakers, producing one distance value for each loudspeaker. The average length of all stored values is used as reference when setting the loudspeaker gains. The gain is set between 0.2 and 1.0 on linear scale. If the stored length is close to the average length, the loudspeaker gain is set to 0.6. Therefore, in the middle of a normal sized room all the gains are equal and the reverberation is same in all directions. In a narrow corridor model, the closest walls produce low gains to those directions and the end of the corridor will have a higher gain. Similar gain adjustment happens when the virtual listening position is in a corner of a room.

With informal listening of the system, it can be stated that the implemented ad-hoc non-diffuse reverberation algorithm gives, e.g., the feeling of standing close to a wall. However, maybe a more sophisticated method would be to calculate a 3-D distance function and weight the outgoing reverberation signals according to it. In addition, the 24 gain values should be more dependent on each other, so that the total radiated sound energy would always remain the same in any position in the geometry.

DISCUSSION AND APPLICATIONS

Only virtual rooms that are bigger than the listening room itself can be rendered "physically correct". This is caused by the fact that the sound has to travel first to the microphones and then back from the loudspeakers to the user. In addition, the latency from the signal processing makes virtual walls to be more distant than the walls of the real room. However, the presented system is not constructed for physically correct auralization, it is rather to be applied in installation art, live performances, and computer games. In these applications the "physically correct" sound rendering is often a secondary goal. It is more important that the system creates a perceptually relevant sound effect.

References:

[1] T. Lokki, R. Kajastila, T. Takala: Virtual acoustic spaces with multiple reverberation enhancement systems. The AES 30th International Conference, Saariselkä, 15-17 March 2007.

[2] R. Kajastila, S. Siltanen, P. Lunden, T. Lokki, L. Savioja: A distributed real-time virtual acoustic rendering system for dynamic geometries. In AES 122th Convention, Wien, Austria, May 5-8 2007. Paper nr. 7160.

[3] Verse, 2007. http://verse.blender.org/cms/Verse.571.0.html.

[4] M. Puckette: Pure Data, 2007. http://www.crca.ucsd.edu/ msp/software.html

[5] W.R. Sherman and A.B. Craig: Understanding Virtual Reality, Morgan Kaufman, 2003

[6] D.R. Begault: 3-D Sound for Virtual Reality and Multimedia, Cambridge, MA: Academic Press Professional, 1994.

[7] O. Deille, J. Maillard, N. No, K. Bouatouch, J. Martin: Real time acoustic rendering of complex environments including diffraction and curved surfaces. In AES 120th Convention, Paris, France, May 20-23 2006, Paper nr. 6743.

[8] V. Pulkki: Spatial sound generation and perception by amplitude panning tecniques. PhD thesis, Helsinki University of Technology, 2001.

[9] T. Lokki, J. Hiipakka: A time-variant reverberation algorithm for reverberation enhancement systems. In COST-G6 Conference on Digital Audio Effects (DAFX-01), pp. 28-32, Limerick, Ireland, Dec. 6-8 2001.