A Real Time and Interactive Music Spatialization System

Leandro Ferrari Thomaz[1], Regis Rossi Alves Faria[1]

[1]Laboratório de Sistemas Integráveis (LSI) – Escola Politécnica – Universidade de São Paulo (USP)
Av. Prof. Luciano Gualberto, travessa 3 nº 380 – 05508-900 – São Paulo – SP – Brazil
{lfthomaz, regis}@lsi.usp.br

Abstract. *This work involves the investigation on techniques for musical spatialization, and the development of a real time spatialization system, based on a multichannel auralization engine (AUDIENCE). The constructed system uses the Ambisonics technique, which allows the recreation of a three-dimensional sound field. The system reproduces spatial music through an arrangement of loudspeakers around the listener where sound sources can assume some free positions, and also be put into motion. Results of experiments with a musical piece show that the system can be used by composers, conductors, and producers for the musical spatialization task.*

1. Introduction

Spatial localization relates to the human ability of judging the direction and the distance of a sound source. Thus, we can understand sound spatialization as the act of positioning a sound source with the intention of stimulating this capacity in a listener.

Having this property to call the attention for certain events, the natural course would be using sound spatialization in music. But this would be delayed by some centuries until it was consciously realized by the composers. By the 20th century, some composers would consider space as a new parameter to explore in music and, more than fifty years later, systematically use it in compositions called space or spatial pieces (TROCHIMCZYK, 2001).

The motivation of this research is to create a computational tool to assist composers, conductors and producers in the task of spatialization of sound sources. With a system assisted by computational tools, the user can create experiments with virtual space positioning (a positioning example can be seen in Figure 1).

The system is capable of reproducing spatial pieces in the form suggested by the composer, allowing the user to manipulate and create new forms of spatial presentation of the same musical piece. It can also be used for live presentations. This work results from an application of the first implementation of the AUDIENCE spatial sound production, transmission and reproduction architecture (FARIA, 2005; FARIA; ZUFFO, 2006).

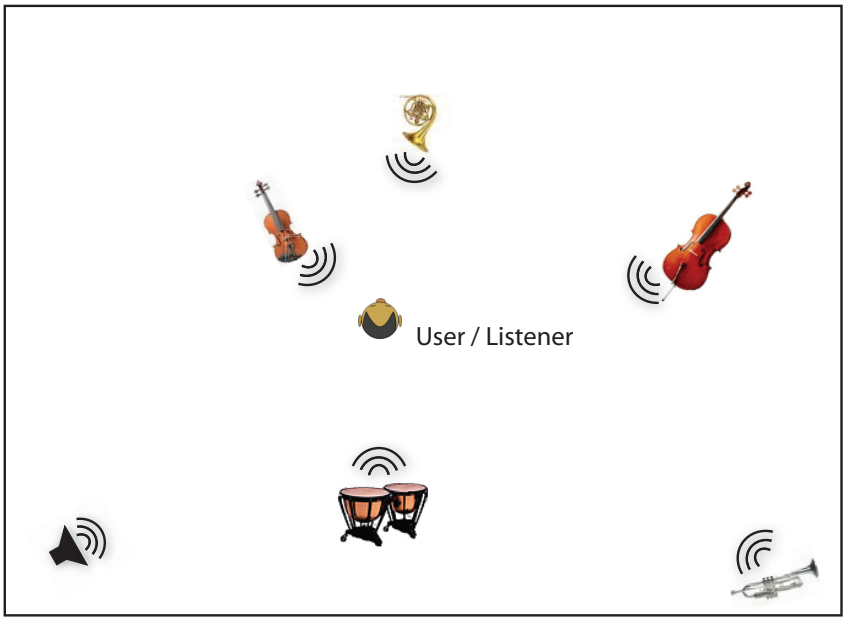

Figure 1. Positioning of virtual sound sources surrounding the user

In this paper, we describe the implementation of the sound spatialization system and the experiments made with it for the production and execution of a spatial musical piece. Further details about the historical use of spatialization, other sound spatialization systems and this implementation can be obtained in THOMAZ (2007).

2. Sound Spatialization System

A sound spatialization system is understood to be a complete chain of processes used to spatialize sound sources, whether a musical piece or a movie with surround sound. The spatial reproduction, by means of a configuration of some loudspeakers or headphones, is just one of the tasks.

In the AUDIENCE system the input of a sound spatialization system is comprehended by the description of the acoustic scene, i.e., which sounds will be spatialized and how they will behave spatially throughout the execution, and of the description of the temporal behavior of sounds in the scene, that can be a score, recorded sounds or other instruction that produces sound.

From the acoustic scene description, the next block is responsible for positioning the sound sources, either through the positioning of the proper musicians, the use of a mixer to position a sound in a recording channel or spatialization software.

Next, the sound can be encoded by means of some auralization technique, so as to organize temporal and spatial sound information. The decoding block does the inverse work and reconstructs the original information. Within these blocks, algorithms of sound signal compression can be used.

A transmission block is responsible for taking the sound information of the place where it was produced and bring it to the place where it will be reproduced. This can be done using analog or digital medias (e.g. magnetic tapes, DVDs, CDs) or digital transmission through networks and protocols of communication (e.g. computer networks).

Finally, the reproduction block materializes in sound waves the spatialized sound information, as it was described in the input of the system. It can be made, for example, by means of a musician performance or using loudspeakers.

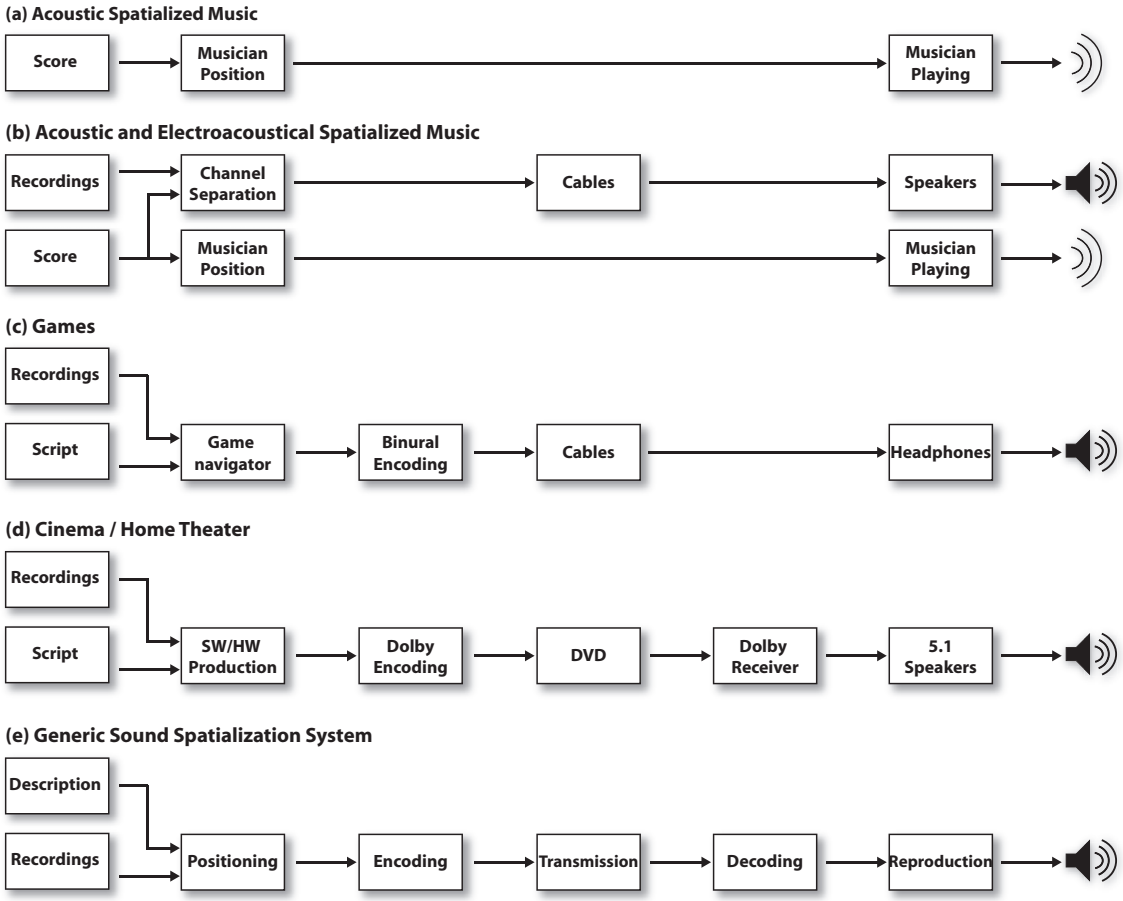

Figure 2. Generic Sound Spatialization System and examples

Figure 1 shows some examples of sound spatialization systems and a generalization. As it can be seen, it is not necessary for a system to present all blocks. Frequently, a sound spatialization system is implemented by a subgroup of these blocks, and their use depends on the necessities of the spatial sound scene described by the composer/conductor/producer.

3. Ambisonics Auralization Technique

Auralization techniques have the goal of recreating three-dimensional sound fields by means of physical and mathematical modeling (FARIA, 2005). The Ambisonics auralization technique was chosen for this work due to its easy implementation, low computational cost, and its *sweet spot* capability to serve a number of users. This technique also satisfies the need for a realistic three-dimensional sound field recreation, and allows the use of flexible configurations of loudspeakers, which is a requirement for the free spatialization of musical scenes. Next, we give a brief description of the Ambisonics technique. Further details can be found in Gerzon (1973), Bamford (1995)

and Daniel (2001). Gerzon (1973) developed a system capable of recording and reproducing sound fields. The system was based in physical/mathematical approach and it was capable of reproducing sound fields with height information (periphony).

Ambisonics coding can also be made by means of synthesis, where a monophonic signal can be (virtually) positioned in space through manipulations with gains in each channel. The virtual sources, through encoding equations that receive the rotational and elevation angles, are located on a unitary reference sphere. Perception of distance must be applied to the signal through attenuations and delays. The storage and transmission of Ambisonics signals depend mainly on the order of the system, which defines the number of channels. An interesting property of the encoding equations is the ability to manipulate the sound field through simple trigonometric operations, with low computational cost, as described by Malham and Myatt (1995).

Decoding of Ambisonics channels is made through matrices of gains calculated for a particular speaker's configuration. As the calculation of these gains can turn extremely complicated depending on the regularity of loudspeakers positioning, they are previously calculated for the most common configurations (FURSE, 2006). Moreover, shelf filters are used to make psychoacoustics adjustments.

4. Using the AUDIENCE Spatial Sound Engine

The AUDIENCE spatial sound engine, proposed by Faria (2005) is the framework of the present real time interactive spatialization system implementation. It is based on a modular approach with four functional layers.

This division allows the use of different techniques and tools in the implementation of the functions executed in each layer (e.g. programs, plug-ins, and external equipment), keeping the communication between them through predefined interfaces. With this approach, it is not necessary to remake the entire system when a substitution of tools and techniques used in one layer is needed. This guarantees independence and interoperability between layers, and flexibility in the choice of algorithms for third part tools. Next, we briefly present the present implemention abstraction of the four layers following the AUDIENCE architecture.

The first layer is responsible for the description of the acoustic scene, informing to the next layer sound sources attributes such as the position of objects, position of the listener and parameters of the environment (e.g. acoustic parameters). These data can be supplied through user interaction directly on the software or by another system that uses the AUDIENCE engine, such as a virtual reality navigator.

Acoustic simulation has a very important role in the production of sound realism; therefore the spatial perception and its quality is directly associated to this. This simulation is done in the second layer and calculates the sound propagation from the sources to the listener, considering the environment (surrounding) where they are immersed. Several techniques such as geometric models (e.g. ray tracing, radiosity, source-image), ondulatory models (e.g. finite elements) or statistical models can be used and implemented.

In the third layer Ambisonics or other formats can be used for both temporal and spatial sound encoding, and to generate coded items for transmission. Here compression

or streaming can occur, depending on the way of transmission or on the used media. In the present implementation, the transmission is accomplished using digital audio patching within the same engine instantiation.

The last layer of the engine decodes the signal generated by the encoder and reproduces the final sound field through loudspeakers or headphones. It is sensitive to the equipment available in the system, number and disposal of loudspeakers. Depending on the choice of codification format and the availability of loudspeakers, a great variety of reproduction configurations can be considered. The mixing from the various sound sources is also made in this layer.

Further information on the AUDIENCE engine can be found in FARIA (2005) and FARIA (2006).

5. Sound Spatialization System Implementation

Implementation of the sound spatialization system was made over the Pure Data (Pd) programming platform. Pd, developed by Miller Puckette (2006), is a graphical environment for programming musical and audio applications, similar to MAX/MSP.

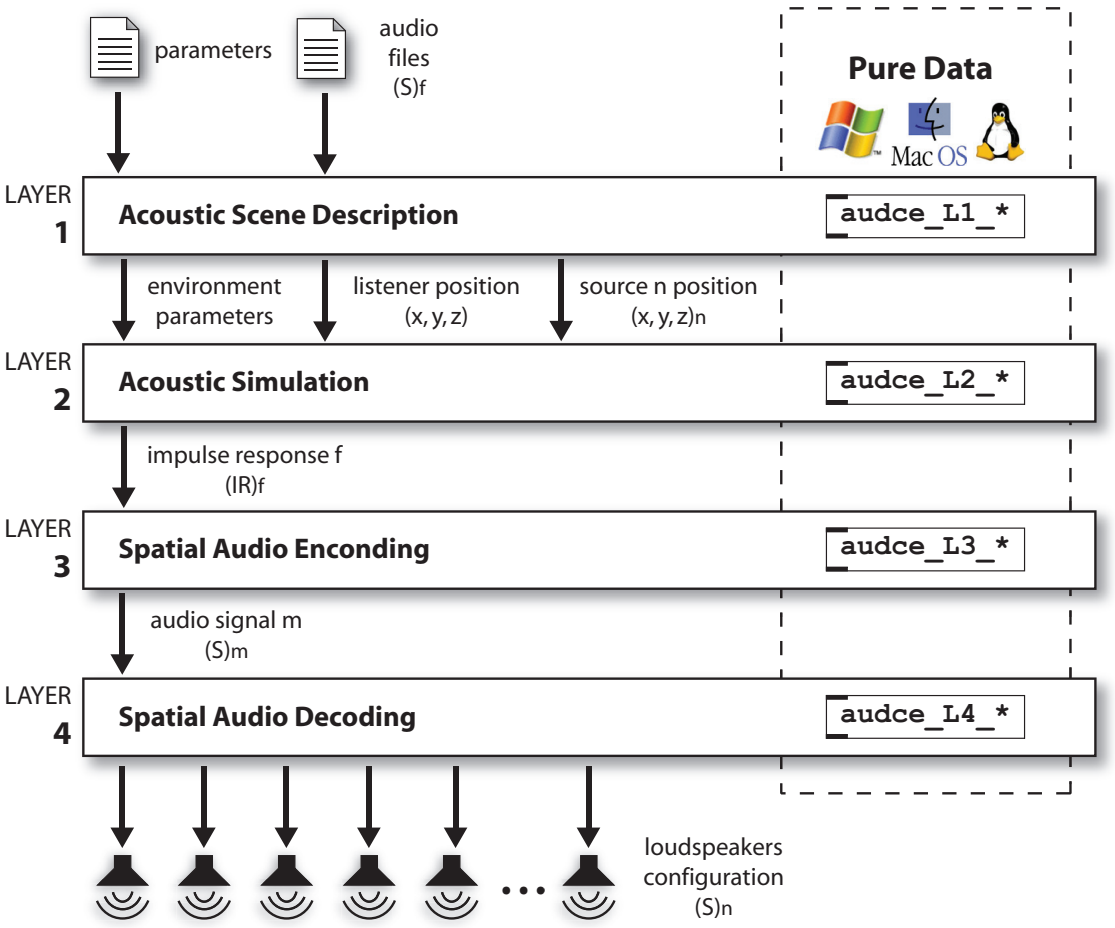

Figure 6. AUDIENCE engine implementation

Pd is an open and flexible tool, with low latency processing of audio. Source code is open, and it can be compiled in diverse platforms such as Windows, MacOS and Linux. Pd allows new blocks to be created by the user, called externals, which were a valuable tool to prototype AUDIENCE layers functionalities. Programming of these blocks is made with C language, and must follow some rules so that they can be used in a Pd patch.

Figure 6 illustrates an adaptation, specific for this project, of AUDIENCE layers. The index *f* represents the number of sound sources, *m* the number of signals of audio codified and *n* the number of loudspeakers. Strict aspects of synchronization between the layers were relaxed, since Pd itself is responsible for the synchronization of the audio flow.

A set of Pd blocks was developed to satisfy the description of each layer functionality, the AUDIENCE system conventions, the desired interface between layers, and the application specific requirements,(see Table 1).

Blocks were divided in six groups, identified by its nomenclature, in which `Ln` indicates the layer (`n` being this layer) that the block implements, `gui` indicates blocks related with the graphical interface, and `aux` indicates the auxiliary blocks.

Table 1. Pd blocks developed, under AUDIENCE framework

Group		Blocks	Pd Type	Description
Layer 1	`audce_L1_*`	`audce_L1_pd_gui`	*patch*	Control of scene input parameters based on a graphical interface
Layer 2	`audce_L2_*`	`audce_L2_allen`	*external*	Based Acoustic simulator in in the method of virtual sources of Allen (1979)
Layer 3	`audce_L3_*`	`audce_L3_amb_3rd~`	*external*	3rd order Ambisonics Encoder based on the source position (using Ambisonics equations in a free-field environment)
		`audce_L3_amb_ir~`	*patch*	Ambisonics Encoder based in impulse responses generated by acoustic simulation
Layer 4	`audce_L4_*`	`audce_L4_amb_3rd~`	*external*	3rd order Ambisonics Decoder
Graphical Interface	`audce_gui_*`	`audce_gui_grid`	*patch*	Sound sources and listener positioning
		`audce_gui_srcfile`	*patch*	Audio source file gui tool
		`audce_gui_volume`	*patch*	Volume controller
Auxiliary	`audce_aux_*`	`audce_aux_convolver~`	*external*	FFTW library based convolution block
		`audce_aux_mixer`	*patch*	Signals mixer
		`audce_aux_move`	*patch*	Automatically moves a sound source
		`audce_aux_srcfile`	*patch*	Control de audio source file reproduction

6. Experiment and Results

6.1. Experiment Methodology

This experiment included the recording of the sound sources in studio, the construction of a Pd patch using the blocks described in the previously section, and based in the musical piece's requirements, and the reproduction of the musical piece. From this reproduction, validation tests have been carried through to evaluate the quality of the spatialization.

6.2. Musical Piece Selection

The selected piece was *The Unanswered Question*, by American composer Charles Ives, due to its wealth of timbres (with one string quartet, two flutes, oboe, clarinet and trumpet), and for the division in three groups of the instruments, facilitating the rehearsal and recording. This piece was composed in 1906, and presented unusual elements, a mark of the compositions of Charles Ives. The piece lasts about five minutes, and was called by Ives a "cosmic drama" (IVES, 1953).

The piece is constituted of *three layers*, each one with its own style, texture, instrumentation and music. The string quartet executes a choral music in the first layer. A trumpet periodically enunciates a question and, after each one, the woodwind quartet tries to formulate a reply. Ives suggests, in the text that precedes the score, that one must separate physically the instruments in three groups: *string quartet*, which must be outside the stage or far from the other instruments; *trumpet*, in the middle of the audience; and *woodwind quartet*, at the stage. Thus, we have three distinct sound source groups.

Based on Ives's suggestions, innumerable possibilities of spatial arrangements can apply. One can experiment beyond the positions suggested by Ives, considering each instrument a sound source, and not as a group. Thus, nine sound sources can be arranged in space, creating new effects, different from the ones imagined by the composer.

6.3. Musical Piece Revision and Production

Edition and revision of the score had been made based in two editions. Both presented small errors and differences, needing wide revision and confection of a new edition.

Although, for the execution of the piece, the instruments that compose each group could have been recorded in a joint form, we preferred to capture each instrument separately, so that it was possible for each one of them to be a sound source in the system, and not a group. Due to the necessity of having each group of musicians executing the piece in a joint form (mainly for synchronism tuning) they were isolated inside the studio through bays of absorbent material (fibreglass), leaving the musicians only with the vision of the conductor.

For each instrument, recording was made with two microphones, a dynamic one and a condenser one, to promote isolation features in mixing the capture of each sound source. Thus, for the mixing and final mastering of the sources both signals were used, taking into account the advantages of good isolation of the dynamic microphone and the wider frequency response of the condenser microphone.

6.4. Experiment Pd Patch

Each application has a specific patch that is constructed in accordance with the following requirements: *number of sound sources, movement of the sources* (if some source needs automatic movement) and *level of sound immersion* (influences the Ambisonics order and the configuration of loudspeakers). Figure 7 shows the patch used in tests with nine sound sources (shown around the listener in the central rectangle).

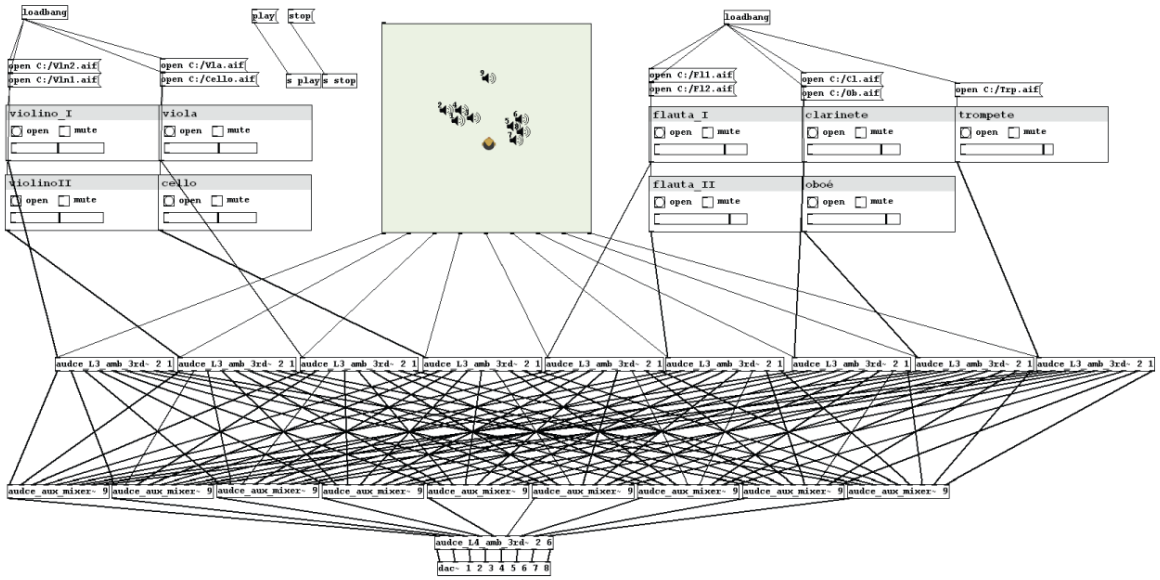

Figure 7. Patch *for The Unanswered Question* spatialization with nine sound sources

6.5. Tests with Users

The goal of the validation tests is to estimate a level of sound immersion achieved by the system through simple usage and subjective tests.

Table 2. Validation tests

#	Tests	Evaluated parameters	Description
1	Blind source localization	Presence Directionality Source distance Room size	Identification of sound sources only by hearing, without access to the interface, to evaluate the quality of the sound field recreation. Executed two times, with three and nine sound sources with predetermined positions.
2	Interactive source localization	Usability Applicability	Interactive use of the system by the user, to evaluate the applicability of the system, in which the user could modify the positions of sound sources.

Experiment was made with musicians and non-musicians. Each person was directed sit to a chair located in the center of the arrangement of loudspeakers and submitted individually to three executions of the musical piece, first in a blind hearing

session, then in an interactive hearing one, where one could manipulate the position(s) of sources in real time. Later they answered a questionnaire to evaluate the subjective item of the experience. Table 2 contains a characterization of the carried tests.

The questionnaire had questions based on attributes we intend to evaluate, and each one of these questions has an explanation of the parameter for better evaluation. Answers are based in a scale from one to five, as a quality indication of the measured parameter. In accordance with Berg (2003), this type of evaluation, where numerical scales are associated with verbal anchors, are commonly used in tests of this nature with good results.

6.6. Results

The following results refers to the system without the use of the second layer acoustic simulator, and with a second order Ambisonics system. Audition was accomplished using a quadraphonic square shaped configuration.

In the tests, answers of four users of the system had been evaluated (three musicians and one not). Table 3 summarizes the answers of the questionnaires filled by the four users (the correct source identification column indicates how many sources had been correctly pointed by the user in the two test cases of table 2).

Table 3. Results of validation tests

User	Musician	Correct source indentification (3 and 9 sources)	Source Localization	Source Distance	System Usability
1	Yes	3 / 6	4	4	5
2	Yes	3 / 6	4	3	4
3	Yes	3 / 7	4	4	5
4	No	3 / 5	3	4	5

During the experiment, users tended to fix the vision in a loudspeaker, as an aid to perceive the localization of the sound. Therefore, some of the answers had pointed the position of the loudspeaker next to the virtual sound source. To reduce this visual guidance and to improve the immersive experience, the users were instructed to close their eyes during the execution of test 1.

Due to the polyphonic nature of the piece texture (the instruments do not present easily independent melodic lines amongst them) the individual localization of the nine instruments were degraded during test 1. One revealed difficult to identify the strings position, maybe due to the choral style of its composition In the same way, the relatively fast interventions of the woodwind quartet (around five seconds), increased difficulty in locating the instruments. On the other hand, when the user could effectively move the instruments with the graphical interface, the identification was significantly facilitated (test 2).

The answers to the questionnaires, based on the two tests, scored an average of four points for source localization and distance. Maximum score was obtained for easiness of use.

7. Conclusions and Future Works

The system had a steady behavior on a commodity Pentium IV-based computer, and ran without performance problems, that is, without perceivable delays in the reproduction of sound, even when nine sounds sources were auralized simultaneously. However, we expect some decrease in performance when using both the acoustic simulator and the convolution of impulse responses. For this, the use of more powerful computers or a computer cluster may be necessary.

Results drawn from individual tests of the acoustic simulation block indicated the necessity of improvements, especially in the calibration of scales and values for the Ambisonic signals.

The lack of familiarization with the instruments and its timbral differences degraded the localization of the instruments, mainly between the strings, the two flutes and oboe with clarinete. Therefore, for source localization without access to the graphical interface (test 1), the acquired ability was necessary to distinguish timbres and to hear similar instruments executing different (melody) lines.

The results obtained during test 2, when users interacted with the system, showed that directionally and distance of the sources, based on second order Ambisonics encoding, were similar to a real sound field.

Through the experiments, mainly test 2, we conclude that the system is promising for using in musical applications and that it can be a low cost and valuable tool for composers, conductors and producers to use spatialization. The easiness in patch construction and use of the system corroborates this conclusion.

These results supply cues for further improvements. Some future works and possible extensions of the system can be enumerated:

1. advanced tests, with more users and based on standard subjective tests methodologies;
2. development of new blocks for the system of sound spatialization, such as the new acoustic scene descriptors (e.g. using X3D), user interfaces (e.g. joystick, head-tracking), acoustic simulators (e.g. finite ray tracing for arbitrary polyhedrons (BORISH, 1984)) and new encoders/decoders (e.g. WFS, binaural with HRTF, and MPEG Surround). Part of these are addressed in the scope of the OpenAUDIENCE[1] project;
3. experiments with three-dimensional configurations of loudspeakers, already implemented in the decoding block for cubic and dodecahedrical speakers configuration;
4. the use of high order Ambisonic encoders/decoders (DANIEL; MOREAU, 2003) to improve the sound field reconstruction;
5. Ambisonics decoding for any speaker configuration, not restricted to regular geometry dispositions (LEE; SUNG, 2003), allowing the adaptation of the

[1] Available at: http://openaudience.incubadora.fapesp.br/portal. Last access: 3 jun 2007.

system in places where the positioning of loudspeakers are restricted (e.g. CAVEs, where the projection cannot be blocked by a speaker).

6. automatic calibration of the system, to normalize the intensity and equalization of each speaker.

Acknowledgements

The authors want to express their gratitude to LAMI laboratory of the School of Arts and Communication of the University of São Paulo (ECA-USP), for kindly providing the studio facility for sound recording experiments, and to prof. José Augusto Mannis, for his valuable technical and artistic assistance in recording sessions. We also thank the CAVERNA Digital laboratory of the Polytechnic School and all the music undergraduate students who have collaborated in the recording.

References

Allen, J. B. and Berkley, D. A. (1979) "Image method for efficiently simulating small-room acoustics". Journal of the Acoustical Society of America, v.65, n.4, p. 943-950.

Bamford, J. S. (1995) "An Analysis of Ambisonic Sound System of First and Second Order". 270 p. Dissertação (Mestrado) – University of Waterloo. Waterloo, Canada.

Berg, J., Rumsey, F. (2003) "Systematic Evaluation Of Perceived Spatial Quality". In: AES International Conference On Multichannel Audio, 24., p.184-198.

Borish, J. (1984) "Extension of the image model to arbitrary polyhedral". Journal of the Acoustical Society of America, v.75, n.6, p. 1827-1836.

Daniel, J. (2001) "Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimedia". 319 p. PhD Thesis – l'Université Paris. Paris.

Daniel, J. and Moreau, S. (2004) "Further Study of Sound Field Coding with Higher Order Ambisonics". In: Proceedings of the 116th AES Convention, Berlin.

Faria, R. R. A.; Zuffo, J. A. (2006) "An Auralization Engine Adapting A 3-D Image Source Acoustic Model To Ambisonics Coder For Immersive Virtual Reality". In: Aes International Conference, 28., Piteå.

Faria, R. R. A. et al. (2005) "AUDIENCE - Audio Immersion Experiences in the CAVERNA Digital". In: Proceedings of the 10th Brazilian Symposium on Computer Music, Belo Horizonte. p. 106-117.

Furse, R. (2006) "First and Second Order Ambisonic Decoding Equations". Available at: http://www.muse.demon.co.uk/ref/speakers.html. Last access: 3 jun 2007.

Gerzon, M. (1973) "Periphony: With-Height Sound Reproduction", Journal of the Audio Engineering Society, v. 21, n. 1, p. 2-10.

Ives, C. E. (1953) "The Unanswered Question", New York: Southern Music Publishing Co. Inc.

Lee, S.; Sung, K. (2003) “Generalized Encoding and Decoding Functions for a Cylindrical Ambisonic Sound System”. IEEE Signal Processing Letters, v. 10, n. 1, p. 21-23.

Malham, D. and Myatt, A. (1995) “3-D Sound Spatialization using Ambisonic Techniques. Computer Music Journal, v. 19, n. 4, p. 58-70, inverno 1995.

Oppenheim, A.; Schafer, R. (1998) “Discrete-Time Signal Processing”. 2 Ed. Prentice Hall.

Puckette, M. (2007) “Pd Documentation”. Available at: http://crca.ucsd.edu/~msp/Pd_documentation/. Last access: 3 jun 2007.

Thomaz, L. F. (2007) “Aplicação à música de um sistema de espacialização sonora baseado em Ambisonics”, Dissertação (Mestrado), EPUSP, São Paulo.

Trochimczyk, M. (2001) “From Circles to Nets: On the Signification of Spatial Sound Imagery in New Music”. Computer Music Journal, v. 25, n. 4, p. 39-56.