Consiglio Nazionale delle Ricerche
ISTI – Istituto di Scienza e Tecnologie della Informazione

EXPLOITATION OF THE AURALIZATION PARADIGM IN A VIRTUAL REALITY ENVIRONMENT APPLICATION

Bertini Graziano, Carreras Francesco, Di Salvo Vincenzo Giovanni

B4 – 16

July 2003

PISA

Consiglio Nazionale delle Ricerche
ISTI – Istituto di Scienza e Tecnologie della Informazione
Area della ricerca di PISA-S.CATALDO

EXPOITATION OF THE AURALIZATION PARADIGM IN A VIRTUAL REALITY ENVIRONMENT APPLICATION

Bertini Graziano(*), Carreras Francesco(*), Di Salvo Vincenzo Giovanni(**)

(*) ISTI-CNR
(**) MOSART Project Young Researcher

INDEX

July 2003

Abstract

The aim of this paper is to show preliminary results of a study aimed at dealing with the specific problem of assuring real-time sound rendering in a Virtual Reality application concerning the virtual tour of the inside of the Baptistery of Pisa. It is well known that the impulse response of the Baptistery of Pisa is characterized by a 12 sec impulse response time, which raises serious computing difficulties for meeting real-time sound rendering constraints. In usual situations, like those of concert halls or historical theatres, the response time is from about 1 to 3 sec.

The process of auralization of virtual reality environments, which is becoming an integral component in advanced applications of this kind, has, in recent times, attracted much research interest and is part of the wider domain of design of interior spaces with expected sound response behaviour.

This article reports on the results of trial proofs performed with the *Huron* multi-DSP Sound Workstation, that is produced by the Australian LAKE Company, leader in the methodologies of the sound rendering at worldwide level.

A preliminary investigation of the problem was developed in cooperation between the Norwegian University of Science and Technology of Trondheim and the National Research Council of Pisa, in Italy, in the framework of a stage at NTNU within the European Project “Mosart” (2003).

* * *

1. Introduction.

The high quality of immersive visual rendering effects obtainable by the technology for the realization of virtual reality environments requires that also the audio components create realistic effects. The *presence effect* in Sound Rendering and the *auralization paradigm* (sound directionality, sound spazialization, etc…) are, by now, an integral component of advanced virtual reality applications.

Commercial solutions in the market, such as those provided by the *Huron* workstation of the Australian LAKE Company [1], whose costs are targeted around tens of thousands of euros, meet many of the most common requests. These computers are provided with multi-DSP cards and a high-level software environment to manage the convolutional tools in an optimal way.

However the actual challenge deals with the search for cheaper solutions for specific applications, at the same time assuring satisfactory quality performances.

In this article the analysis of the reverberation effect inside the Virtual Representation of the Italian Baptistery of Pisa is discussed. The analysis was executed on the Sound Workstation LAKE Huron, with the purpose of evaluating state-of-the-art technological solutions and, at the same time, to investigate new methodologies of audio signals processing which lend to be implemented on lower cost commercial multi-DSP platforms with acceptable performance levels. In fact, this investigation aims at producing results to be used in the framework of the european DHX project, where the ISTI research unit has the task of providing a low cost VR solution.

In this preliminary stage of the reasearch some promising directions will be investigated and evaluated.

2. Preliminary Theoretical Issues.

2.1. Auralization: An Overview.

Auralisation is the method of processing an anechoic sound signal (speech or music) with the purpose of adding information related to the reverberation properties of an acoustic space. [2]

To better understand what "information related to the reverberation of an acoustic space" means, let the reader think of the theatres of the Ancient Greece. The theatre, as a building, was born really there. The first theatres consisted of wood benches, iron-horse disposed, and built in steps on the slope of a hill. In front of the steps, a flat space was the stage for the representations. This arrangement of the scenic space, that positioned the audience on levels of different highth, was dictated by the requirement of making thousands of spectators fully partecipant, that is making them able to see the scene but, at the same time, to hear the actors' voices. The Ancient Greeks had guessed that closed structures of this kind could provide the solution. Such geometries were ***reverberating***, adding a loud tail to the sound and the voice, able to confer it sense of surround and fullness [3].

The art of theatre design developed through the centuries up to our time, as the result of a continuos search for perfection in sound rendering within enclosed spaces.

In the following of this paragraph the basic concepts of the auralization process will be briefly recalled.

At present a branch of the physics, the Acoustics, deals with carefully studying the acoustic behaviour or the "loud answer" of an environment. Techniques to design spaces with pre-determined acoustic response have recently been developed [4,5,6,7]. The same domain of knowlegdge can be applied to the simulation of the acoustical effects in virtual reality applications.

The acoustic qualities of a real world space depend how the sound energy near the listner decays in the time.

In a listening space, such as a room, a concert hall or a cathedral, the sound waves, arriving to our ears, consist of direct sound from the sound source, the early reflections and the late reflections (fig. 1). The first few reflections off the walls correspond to the early reflections. Continuing to bounce off the walls, the spatial density of the waves increases, arriving to the listener from all directions. Their energy, however, decreases exponentially, because every environment acts as a dispersive system. These reflections correspond to the late reflections. The Reverberation phenomenon describes the trend of the reflections on the walls.

2.2. The Impulse Response of an Environment.

It is not easy to characterize operationally the reverberating characteristics of an environment.

The Reverb is strongly affected by a number of random parameters.

For instance the roughness of the walls is one factor which affects the randomness of the reverber effect of a room.

The covering of the walls (marble, velvet, wood, etc.) and the temperature's gradient inside the room, due to the presence of the audience, are responsible of a slowly spatially and temporally not-stationary quantity.

The human ear is able to perceive the variability which characterizes these parameters.

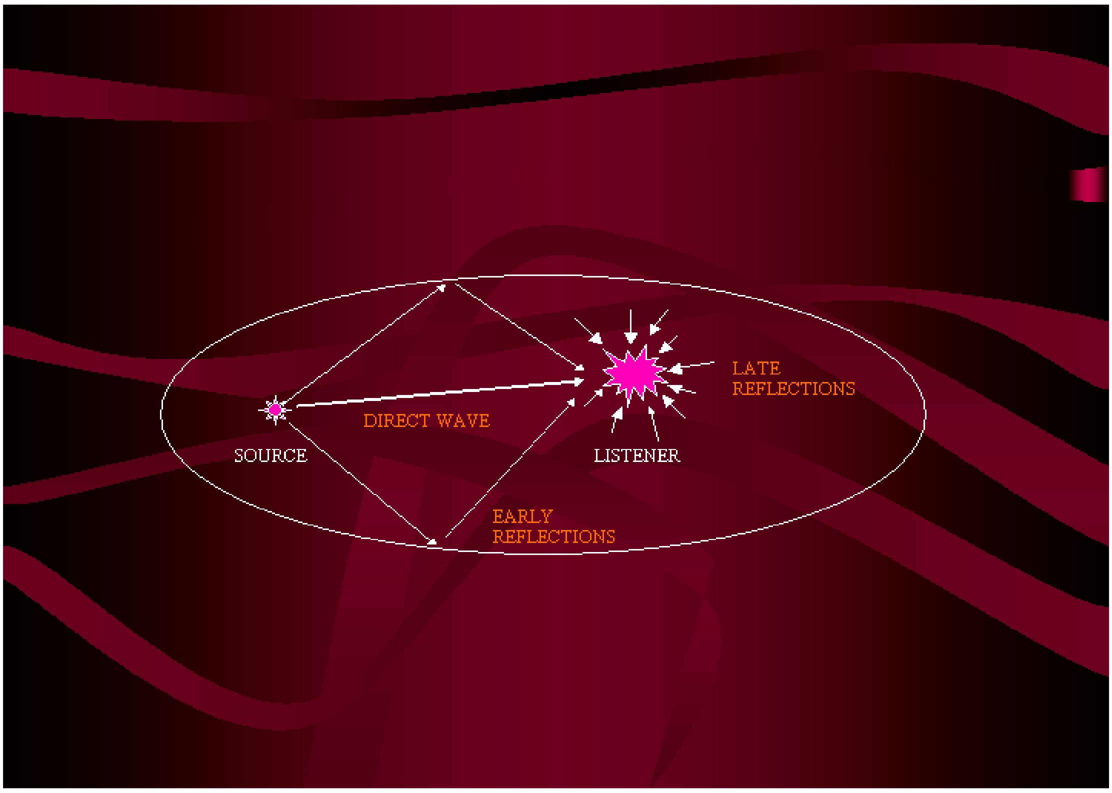

Fig. 1 – Direct wave, Early Reflection and Late Reflection in an acoustic space

Therefore the design of the acoustic qualities of a concert space becomes an inescapable element of the architectural one [3].

In the case of the Virtual Reality applications a not close consistency of the synthesized parameters with those of the real model can make the listening quality unnaturally clean.

The possibility of having an objective *acoustic descriptor* becomes, therefore, a fundamental requirement to carefully describe the acoustic qualities of a room [8].

The istantaneous acoustic internal field of an environment is the result of how the acoustic sources match themselves to the environment response.

In light of this, the problem can be faced, in a convenient way, taking into consideration the answer of the system to an impulsive loud sound.

In linear and stationary systems, the answer to an impulse input takes the name of Impulse Response. The same condition holds also for linear and (slowly) non-stationary systems, like a concert hall. This takes the name of Pressure-Level Impulse Response (PLIR) of the environment and in the following it is expressed with h(t).

The Reverberation Time is measured like that part of impulse response of the environment whose amplitude is decayed of 60 dB (fig. 2).

The measure of the PLIR is the fulcrum of the experimental investigations on the resposnse of listening environments. Its knowledge allows one to elaborate an arbitrary dry signal x(t), adding information related to the listening's characteristics of the room (*Auralization Paradigm)*.

The auralizated signal can be computed by the following equation:

$$\blacktriangleright \quad \mathbf{y(n) = \sum_{m=0}^{M} h(m)\, x(n\text{-}m)} \qquad \mathbf{(1)}$$

where h(m) is the PLIR and x(t) is an arbitrary input signal.

It can be noted that this equation describes the characteristics of sound transmission of a room, for a given configuration of the variables of the Auralization Paradigm: it depends on the position of the receiver (listener) regards the geometry of the environment, the directionality of the source, etc.

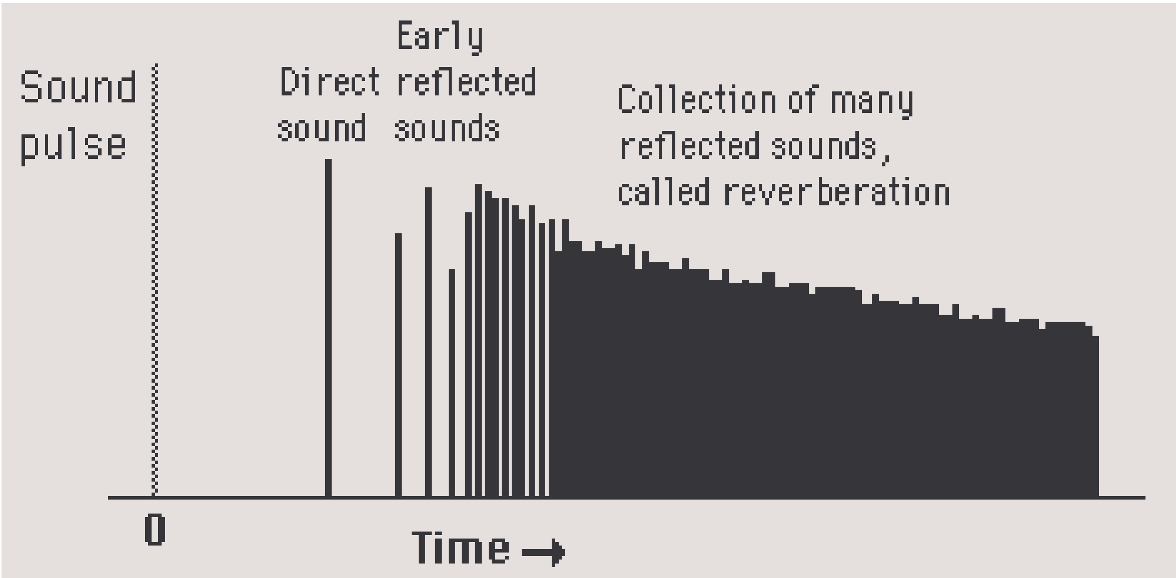

Fig. 2 – Impulse responce of a listening space

By Virtual-Sound-Reality Environment we mean a space where the reverberation effect of a real world acoustic space is simulated [9]. The listener, inside it, will perceive the sound as if he was in the real-world environment (see Fig. 3).

In auralization of sound environments the aim is usually to produce a perceptually satisfying audible result to a listener at one position at a time in the simulated, virtual sound environment.

To achieve the best results, the reproduction room should be anechoic, and the original non-processed acoustic signal dry.

The steps that are needed for performing such auralization can be presented as follows (see Fig. 4):

1. *Model definition* comprises of providing the data and control parameters for describing the properties and dynamic events of the modeled environment (consisting of the source, medium, and the receiver models) [11]. The model definition is a preliminary stage to the actual auralization process that consists of the two following steps.

2. *Modeling* comprises of the actual (nowadays mostly computational) acoustic simulation of the environment, which is carried out according to the source–medium–receiver model. The simulation may contain dynamic (time-varying) events and user interaction with the sound environment, through interactive modification of the defined control parameters. These control mechanisms can be defined as fixed features of the auralization paradigm, in which case the interaction is always similar (e.g., listener movement in the defined environment). Another option is to make the interaction a part of the model definition in which case it can more flexibly be adjusted to the contents and the purpose of the application.

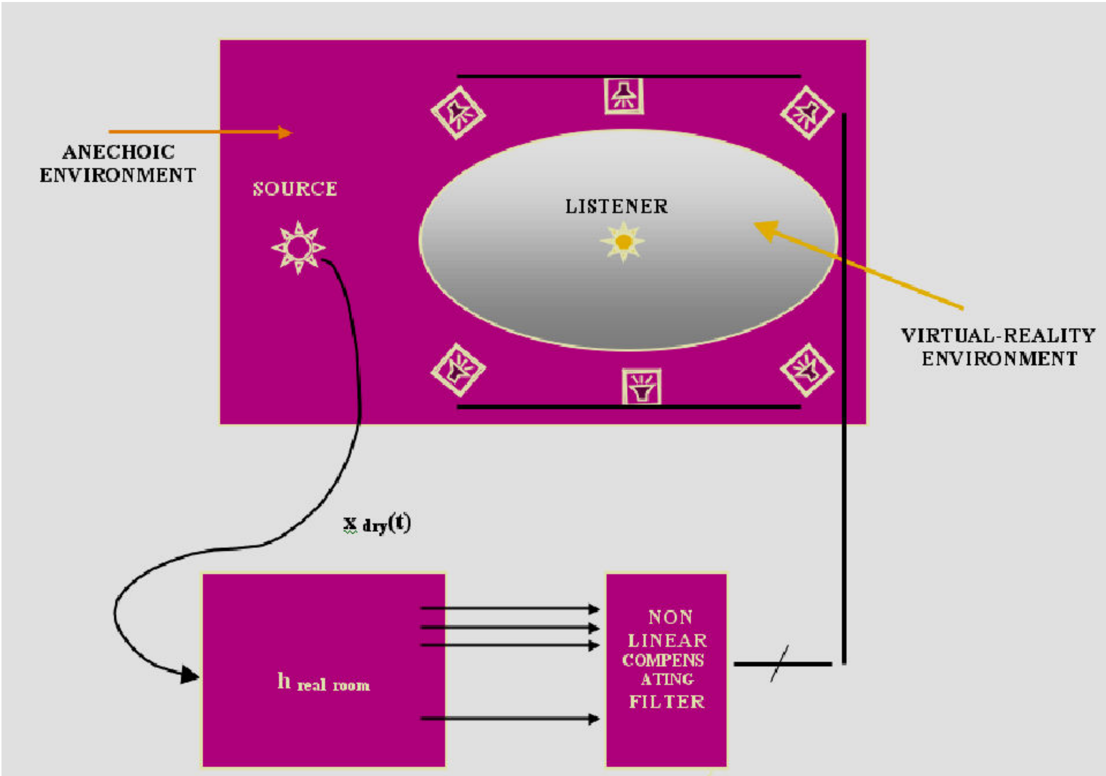

Fig. 3 – A solution for realizing a virtual-reality environment

3. *Reproduction* consists of rendering the digitally simulated sound field to an acoustic, audible sound field, where the user can hear the virtual sound sources at their defined positions, and the effect of the defined room acoustic model.
The required stages of auralization are applied to the three components of a virtual sound environment described previously (the source, medium, and the receiver). The two latter stages of auralization (modeling and reproduction) together form the rendering of a sound environment. In the following sections, the above three stages of auralization has been realized using the potentiality of the Lake Huron Sound Workstation.

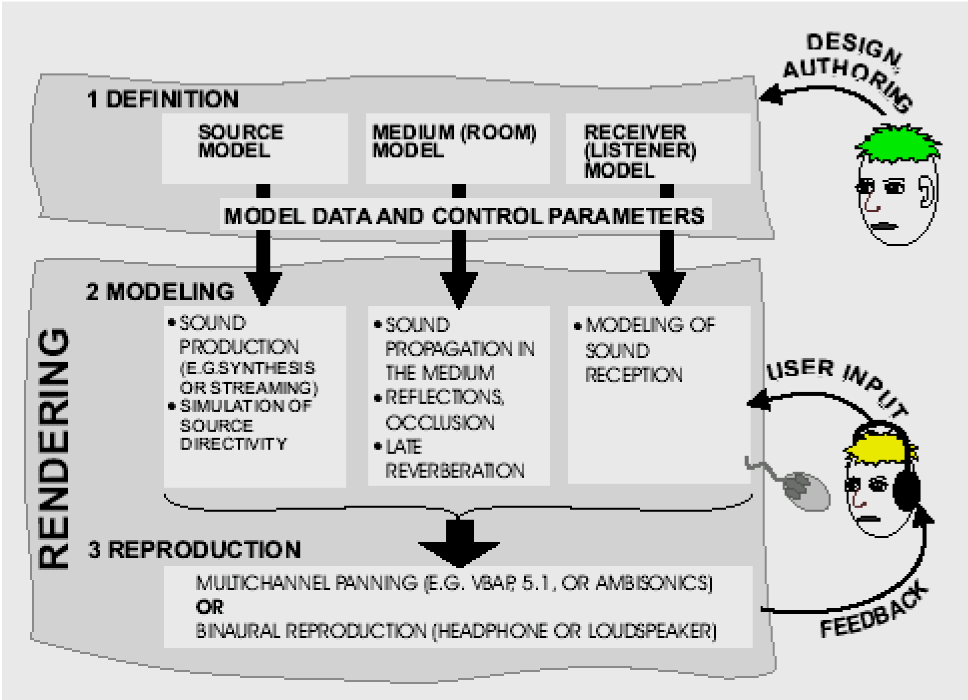

Figure 4: The three stages of auralization. At model definition, a parametric representation of the sound environment is created. These parameters are used at the modeling stage for producing the appropriate spatial effect to sound. Rendering of the sound environment includes both the modeling and reproduction that are needed to reproduce the modeled sound field to the listener. This stage may involve user interaction that leads to control parameter changes.

3. The Lake Huron Solution.

In this section the advantages offered by the convolutional algorithms of Lake Huron Sound Workstation are briefly discussed with regard to the impulse response of the Baptistery of Pisa, shown in fig. 5. As it is possible to see that it is quite long-lasting: 12 sec, that is 524288 samples at a Fs = 44100 Hz, 16 bits quantized.

Purpose of this preliminary analysis is to test the response for well defined situations in order to gather technical data to be used for comparison for the evaluation of alternative algorithms, computationally cheaper but, depending on the application, with acceptable levels of performances.

The Lake Huron Workstation is a PC- and Windows-based system equipped with multi-DSP sound cards. Every card has 4 DSP processors. The low-latency Convolver Tool provides up to 278,244 tap FIR filters, so it can compute in real-time quite long-lasting impulse responses (up to 12 sec at 48 KHz).

To realize the auralization of the Baptistery of Pisa the Equation (1) has been used. Equation (1) is a *FIR filtering equation,* so one can say that a FIR filter is a convolver. Depending on the application and hardware, a FIR digital filtering operation can be organized to operate on a *block-by-block* basis or a *sample-by-sample* basis.

In block processing methods the data are collected and processed in blocks.

In sample processing methods the data are processed one at a time. Such methods are used primarily in *real time applications* that require the continuous processing of the incoming input and are executed with the help of DSP platforms.

The reverberation effect was obtained with the simultaneous use of both block-by-block and sample-by-sample methods.

The block-by-block method, that is not implemented in the Lake software, was used to divide the impulse response in two blocks. After this, each of the two blocks was convolved using a bank of FIR filters, so realizing a sample-by-sample elaboration. This algorithm was implemented on the Lake Huron Workstation as shown in fig. 6.

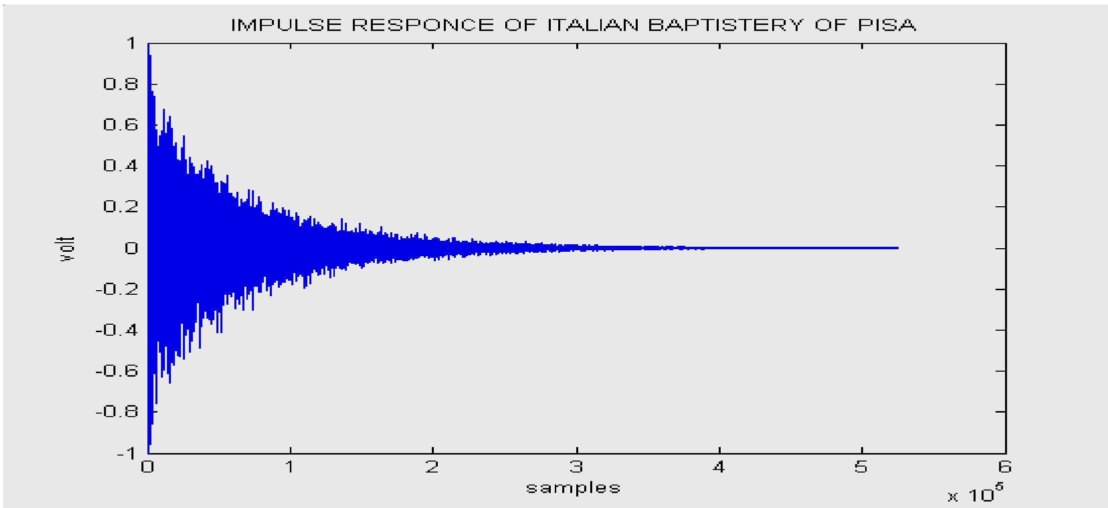

Fig. 5 – Impulse response of Italian Baptistery of Pisa

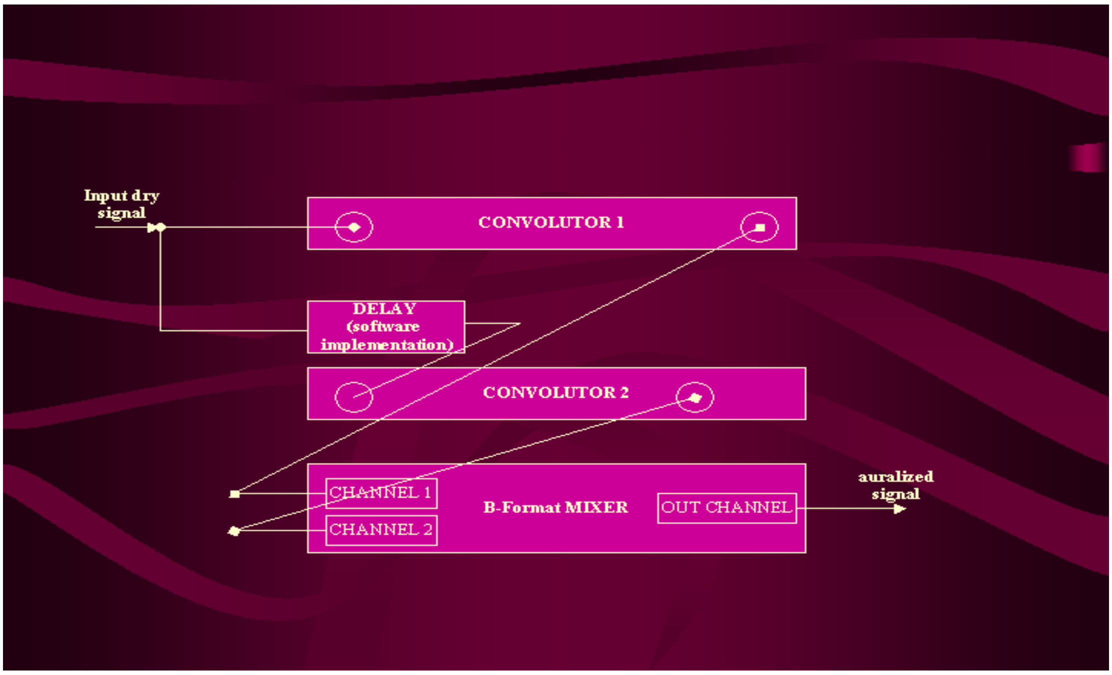

Fig. 6 – This picture shows the implementation of the Auralization Algorithm to realize the virtual reverberation of the Baptistery of Pisa (Lake Huron environment).

The obtained results are, undoubtedly, praiseworthy, because has been possible obtain a faithful realization of the parameters which compose the paradigm of the Auralization (presence effect, directionality, spatialization, etc.). For this reason they will be considered as comparison unity for the future developments.

4. Future developments

A future objective is to obtain a reduction of the computing time of the convolution algorithm, having in mind the use of cheap equipment for reproducing real-time sound effects.

The future developments will move in some different directions of analysis of the problem.

- Because in applications, for which the useful band is limited, as in the case of speech where a band of about 5 KHz is required, the use of the Time Decimate Techniques will be investigated. An off-line analysis, using a Personal Computer platform, showed that there is no appreciable sound rendering quality degradation if one uses this Technique. Tests have been done with decimation factors of ½ (22050 Hz) and ¼ (11025 Hz). The only factor to keep present is the Reproduction Data Rate, which must have the same the frequency as the (decimated) sampling.

A consequent application could be the use of DSP multiprocessor platforms to realize the techniques of convolution in real-time on the decimate signals.

- Another solution, whose development is in under study, regards the possibility of computing the convolutional algorithms in the mp3 coded domain, with the purpose of resolving the sum of convolution in real-time so to have the possibility to consider the entire audio frequencies range and a cheaper band-occupation cost.
- The possibility to obtain the perception of a sound, in function of the position of the listener, as regards the relative source, can be also realized by means of the Ray-Tracing Technique [12,13]. This technique allows to recreate the real environment by CAD tools. It is possible to gather good results when is not required a faithful rebuilding of the acoustic space. Using the technique of the Ray-Tracing, the filtering of the sounds coming from several directions with SHRTF Filter (SHRTF: Simplified Haed Related Transfer Function) [9], allows a good reduction of the calculation's times, regards those necessary to perform the convolution operations in real-time.

A proposal study about the sound source localization in virtual environment has been developed with a DSP low-cost system in [14].

References.

[1] – http://www.lake.com/
[2] A. Farina, A. Bellini, E. Armelloni - Non-Linear Convolution: A New Approach For The Auralization Of Distorting Systems - Proceedings of the 110th AES Convention, 2001 May 12-15 Amsterdam, The Netherlands
[3] - M. Barron, "Spatial impression as measured in concert and opera auditoria", Atti del *16° Congresso Internazionale di Acustica (ICA)*, Seattle (USA), 20-26 Giugno 1998 , pp. 5-6.
[4] R.Pompoli, *Progetto Acustico delle Sale da Spettacolo*. Dispensa del Dipartimento di Ingegneria, Università di Ferrara, A.A 1997/98.
[5] M. Sakurai, Y.Korenaga, Y. Ando, "A sound simulation system for seat selection", Atti del congresso *Music & Concert Hall Acoustics*, Tokio (Japan), Maggio 1995, pp. 51-59.
[6]- www.ramsete.com
[7]- CATT-Acoustic. http://www.netg.se/~catt/
[8] – L'Acustica dei Teatri Storici: Un bene Culturale. Atti del Convegno Bicentenario del Teatro Comunale di Ferrara. 4 Novembre 1998.
[9] M. Kleiner, B.-I. Dalenbäck, P. Svensson, "Auralization - An overview," J. Audio Eng. Soc. 41, 861-875 (1993).
[10] G. Eckel, *A Spatial Display for the Cyberstage*, ICAD98
[11] - G.M. Naylor, "Odeon - Another Hibrid Room Acoustical Model", *Applied Acoustic* Vol. 38 no.s 2-4 1993, p.131-143.
[12] – Un Modello Computazionale per la Simulazione Acustica Basato su Tecniche di Auralizzazione. Tesi di Laurea. A.A. 1995-96. Candidati: Lalli L, Jacovone M. Relatori: Bertini G., Marani M., Taraballa L. Controrelatore: Maestrini P.
[13] – Farina A. – Pyramid Tracing vs. Ray Tracing for the Simulation of the Sound Propagation in Large System. Proc. Of the Conf. On Comput. Acoustic and its Envionmental Application. Southampton. 1995
[14] – Bertini G., Jacovone M., Lalli L., Marani M. - Studio della Realizzabilità di un Modello per la Localizzazione di Sorgenti Sonore su un Sistema DSP Low-Cost. Associazine Italiana di Acustica. Atti del XXV Convegno Nazionale. Vol.2. Perugina, 21-23 maggio 1997.

Interesting articles are:

- R.Lazzarin, M. Strada, *Elementi di Acustica Tecnica*, CLEUP, Padova, 1983.
- L.L. Beranek, *Music Acoustic & Architecture*, John Wiley & Sons inc., New York - London, 1962.
- L.Tronchin, "Qualità Acustica dei Teatri: Analisi Sperimentale e Confronto con Valutazioni Soggettive in Alcuni Esempi Italiani", *Rivista Italiana di Acustica*, Luglio-Settembre 1994, pag. 51 - 64.
- Y. Ando, "Calculation of subjective preference at each seat of a concert hall", *Journal of the Acoustical Society of America*, vol. 74 (3), Settembre 1983, pag. 873-887.
- H.A. Marshall, "A note on the inportance of room cross section in concert halls", *Journal of Sound and Vibration,* vol. 5 - 1 1967, p.100 - 112.

Acknoledgments

Several people have contributed with efforts and comments in the discussion to this presentation. Mentioned should be: Peter Svensson and Bard Stofringsdal of Norwegian University of Science and Technology of Trondheim and the people of Creative Virtual System Labs of ISTI-CNR of Pisa.

Bertini Graziano is senior researcher by the Istituto di Scienza e Tecnologie del CNR in Pisa. His research interests concern the audio signals numeric processing in the speech and musical domains with special attention to DSP based applications

Carreras Francesco is senior researcher by the Istituto di Scienza e Tecnologie della Informazione del CNR in Pisa. His main research interests are in the fields of cognitive musicology, sound applications in virtual reality.

Di Salvo Vincenzo Giovanni graduated in Telocoomuncations Engineering at the University of Pisa. He is currently active in the research group of the Signals and Images Labs of the Istituto di Scienza e Tecnologie della Informazione del CNR in Pisa. His interests are addressed to Methodologies of Analysis ans Synthesis of Telecommunications Signals.