DIVA Virtual Audio Reality System

Jyri Huopaniemi, *Helsinki University of Technology* (*HUT*)
Lauri Savioja (*HUT*)
Tapio Takala (*HUT*)
More Author Information

Abstract: We have created a model of real-time audio virtual reality. This model system includes model-based sound synthesizers, geometric room-acoustics modeling, binaural auralization for headphone or loudspeaker listening, and high-quality animation. This project aims to create a virtual musical event that is highly authentic in both audio and video. To reach this goal, we innovated this system with a real-time image-source algorithm for arbitrarily shaped rooms; shorter HRTF filter approximations for more efficient auralization; and a network-based distributed implementation of the audio-processing software and hardware.

Introduction

Audio virtual reality applies to fields ranging from computer music to room acoustics to multimedia and training simulation. Because of computational constraints, many systems of audio virtual reality are simple and faintly resemble the physical reality. As a solution to this problem, we present a distributed expandable virtual audio reality system that can accurately yet efficiently model room acoustics and spatial hearing in real time.

The DIVA system consists of two parallel data streams: the audio stream and the visual stream. These streams share some common control and synchronization mechanisms as well as information sources. The output of the streams is what the user, when using the system, will hear and see. Figure 1 features the DIVA system components and the overall information-flow graph. The upper half of the figure shows the audio stream while the lower half shows the visual stream. The system components fall into three categories:

- static data repositories
  - user HRTFs
  - the input MIDI file
  - room geometry
  - musician and instrument models
- information processing units
  - sound synthesis
  - auralization
  - room acoustic modeling
  - MIDI-to-movement mapper
  - animation control
- real-time connections
  - audio and video signal
  - viewer orientation data
  - information about visible image sources.

Based on the user's real-time input, which controls the actions of a virtual viewer (using a GUI or, in the future, a head-tracking system), the system produces the visual and audio output.

This paper discusses the follwing components of the project: real-time binaural modeling of room acoustics; digital signal processing (DSP) aspects of room acoustics and head-related transfer function (HRTF) implementation; and the distributed implementation of the system. For a more detailed presentation of this project, please see (Takala et al., 1996).

Real-time Binaural Room Acoustics Model

Computational performance issues are important to making real-time applications (Kleiner et al., 1993); therefore, few alternative algorithms are available for simulating room acoustics. Of these algorithms, the image-source method best models low-order reflections. Based on geometrical room acoustics, this method is explained thoroughly in many articles (Allen and Berkley, 1979;, Borish, 1984). The algorithm implemented in this software is thus traditional. For auralization, simulations must be binaural. Because the directions of incoming sounds are calculated easily from the positions of the image sources, the image-source method suits real-time binaural processing

Real-Time Communication

In our application, the room-acoustics calculation module communicates with two other processes. The graphical user interface supplies the input, which represents the listener's movements. The model then calculates the position and orientation of each image source. The model finally passes the following parameters about each image source to the auralization unit:

- the distance from the listener,
- the relative azimuth angle to the listener,
- the relative elevation angle to the listener,
- two filter coefficients which describe the material properties in reflections.

The number of image sources that can be calculated depends on the computing capacity available. In our real-time solution, 20-30 visible image sources pass to the auralization unit. The model, moreover, tracks the previously calculated situation and compares it with new inputs.

Updating the Image Sources

In the updating process, the system must respond immediately to any changes in the environment. To do so, the system gradually refines the calculation. First, the system calculates only the direct sound and passes its parameters to the auralization unit. If no other changes cue to be processed, the system calculates first-order reflections, then second order, and so on. In a changing environment, three different possibilities may cause recalculations: movement of the sound source, movement of the listener, and turning of the listener or the sound source.

If the sound source moves, all image-source positions must be recalculated. (The same applies also to the situation when something in the environment, such as a wall, moves). When only the listener moves, the visibilities of all image sources must be validated. Because the locations of the image sources never vary, they never cause recalculation. If the listener or the sound source turns without moving, moreover, there are no position changes and no recalculations necessary. Only the azimuth and elevation angles might change and thus need recalculation.

Auralization Issues

Real-time auralization seeks to preserve the acoustical characteristics of the modeled space and meet computational requirements. This goal constrains the accuracy and quality of the final auditory illusion. To transcend constraints, our aurilization strategy involves the following steps: 1) model the first reflections with an image-source model of the room, 2) use accurate HRTF processing for the direct sound, 3) apply simplified directional filtering for the first reflections, and 4) create a recursive reverberation filter to model late reverberation.

Processing the Real-Time Room-Impulse Response

Real-time modeling of the full room-impulse response exceeds the calculation capacity of modern computers. To solve this problem, we need hybrid systems that exhibit the behavior of room impulse responses in a computationally efficient manner.

We ultimately used a recursive digital filter structure based on earlier reverberator designs (Schroeder, 1962) (Moorer, 1979). Computationally efficient and yet accurate (Huopaniemi, et al. 1994), this structure combines the implemented image-source method and late reverberation generation. The early reverberation filter is a tapped delay line with lowpass filtered outputs designed to fit the early reflection data of a real concert hall while the recursive late reverberation filter is based on comb and allpass filters.

HRTF Filter Design

In a static case, locating sound sources is achieved primarily with three cues (Blauert, 1983): the interaural time difference (ITD), the interaural level difference (ILD), and the frequency-dependent filtering due to the pinnae, the head, and the torso of the listener. The head-related transfer function (HRTF) represents a free-field transfer function from a fixed point in a space to a point in the test person's ear canal. Often, computational constraints require approximating HRTF impulse responsea process that conventional digital filter design techniques can handle.

In most cases, the measured HRTFs must be preprocessed to account for the effects of the loudspeaker, microphone, and headphones (for binaural reproduction) that were used in the measurement. Obtaining a generalized set of filters requires further equalization. Some equalization methods include free-field equalization and diffuse-field equalization. Smoothing the responses before applying the filter design also helps obtain a generalized set of filters.

Minimum-Phase Reconstruction

An attractive solution to the limits of HRTF modeling is reconstructing data-reduced minimum-phase versions of the modeled HRTF impulse responses. In this way, a mixed-phase impulse response can take minimum-phase form and not affect the amplitude response. The use of minimum-phase systems in binaural simulation thus boasts many attractions, namely: the shortest possible filter lengths for a specific amplitude response; simple structure of filter implementation; and better performance of minimum-phase filters in dynamic interpolation. According to Kistler and Wightman (1992), moreover, minimum-phase reconstruction has no perceptual consequences.

With minimum-phase reconstructed HRTFs, the model can separate and estimate the ITD of the filter pair and insert the delay as a separate delay line to one of the filters in the simulation stage. The delay error that results from rounding the ITD to the nearest unit-delay multiple can be avoided using fractional delay filtering (see Laakso et al., 1996 for a comprehensive review on this subject).

FIR and IIR Filter Implementations

Digital filter approximations (FIR and IIR) of HRTFs have been studied to some extent in the literature over the past decade. Filter design using auditory criteria has been proposed by few authors, however. There are two alternatives to a non-linear frequency scale approach: weighting the error criteria and frequency warping. Frequency warping results from resampling the magnitude spectrum on a warped frequency scale (Smith, 1983; Jot et al., 1995). In practice, bilinear conformal mapping with first-order allpass filters implements warping. The resulting filters have better low-frequency resolution but lower high-frequency accuracy, a tolerable trade-off according to the psychoacoustic theory. Filters may be implemented either in the warped domain or, through dewarping, in the normal domain(Jot et al., 1995).

Research in the HUT Acoustics Laboratory has taken a similar approach of non-uniform frequency resolution. Specifically, this research applies warped FIR (WFIR) and IIR (WIIR) structures to HRTF approximation (for details, see Huopaniemi, J. and Karjalainen, M., 1996a and b).

Real-Time Auralization

The auralization system obtains the following input parameters which are fed into the computation:

- direct sound and image-source parameters
- HRTF data for the direct sound and directional filters (minimum-phase WFIR or WIIR implementation stored at 10-degree azimuth and elevation intervals)
- "dry" audio input from a physical model or an external audio source.

The output of the auralization unit is at present directed to headphone listening (diffuse-field equalized headphone, e.g., AKG K240DF), but software that converts the output to a transaural or multispeaker format also has been used.

System Implementation

We used a distributed implementation on an Ethernet-based network to improve the system's computational power and flexibility. Currently we use one Silicon Graphics workstation for real-time visualization and the graphical user interface (GUI) and another for image-source calculations. We also have used a Texas Instruments TMS320C40-based signal-processor system that performs direction- and frequency-dependent filtering as well as the ITD for each image source, the recursive reverberation filtering, and the HRTF processing. We intended for the Ethernet-based system to use the multiprocessor system as a remote-controlled signal-processing system. In the transfer process, the audio-source signal or control-parameter block is transmitted through the network to the signal-processing system, which receives the data, processes the audio signal, and returns the stereophonic audio result to the workstation in real time (Figure 2).

Summary

We have developed a software and hardware system for producing virtual audio-visual performances in real-time. We designed this system for virtual reality simulations that require high-quality audio rendering such as modeling and animating concert hall acoustics. Using this system, the listener can move freely in the virtual concert hall, where a virtual musician plays a virtual instrument. The system computes early reflections in the concert hall binaurally with the image-source method. For late reverberation, we implemented a recursive filter structure that consists of comb and allpass filters. The system performs aurilization through the interaural time difference (ITD) and head-related transfer functions (HRTF). With this system, we created a high-quality auralized and animated concert performance in which a virtual musician plays a virtual flute in a virtual concert hall (computer model of the Sigyn Hall in Turku, Finland). A demonstration of this performance appeared at ICAD'96. The following still pictures are samples from the DIVA demonstration.

- Figure 3. Virtual flutist in the virtual Sigyn Hall.
- Figure 4. Another virtual flutist in the virtual Sigyn Hall.
- Figure 5. Picture of the original Sigyn Hall, located in Turku, Finland.

References

Allen, J. and Berkley, D. (1979). Image method for efficiently simulating small-room acoustics. *Journal of Acoustics Society of America* (JASA), 65(4), 943-950.

Borish, J. (1984). Extension of the image model to arbitrary polyhedra. *JASA*, 75(6), 1827-1836.

Blauert, J. (1983). *Spatial Hearing*. Cambridge, MA: M.I.T. Press.

Huopaniemi, J., Karjalainen, M., Välimäki, V., and Huotilainen, T. (1994). Virtual instruments in virtual rooms - a real-time binaural room simulation environment for physical models of musical instruments. In *Proceeding from the 1994 International Computer Music Conference* (pp. 455-462) Aarhus, Denmark.

Huopaniemi, J. and Karjalainen, M. (1996a). HRTF filter design based on auditory criteria. In *Proceedings from the Nordic Acoustical Meeting (NAM'96)* (pp.323-330) Helsinki, Finland.

Huopaniemi, J. and Karjalainen, J. (1996b) Comparison of digital filter design methods for 3-D sound. In *Procedings from the IEEE Nordic Signal Processing Symposium (NORSIG'96)* (131-134). Espoo, Finland

Huopaniemi, J. and Karjalainen, M. (1996c) Review of digital filter design and implementation methods for 3-D sound. To be presented at the *102nd AES Convention*, 1997. Munich, Germany.

Jot, J.M., Larcher, V., & Warusfel, O. (1995). Digital signal processing issues in the context of binaural and transaural stereophony. Presented at the *98th AES Convention*. Preprint 3980 (E-2). Paris, France, 1995.

Kistler, D. & Wightman, F. (1992). A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. *JASA*, 91(3), 1637-1647.

Kleiner, M., Dalenbäck, B. I., and Svensson, P. (1993). Auralization--an overview. *Journal of the Audio English Society*, 41(11), 861-875.

Laakso, T.I. Välimäki, V., Karjalainen, M., & Laine, U.K. (1996) Splitting the unit delay - tools for fractional delay filter design. *IEEE Signal Processing Magazine*, 13(1), 30-60.

Moorer, James A. (1979). About this reverberation business. *Computer Music Journal*, 3(2), 13-28.

Schroeder, M. (1962). Natural sounding artificial reverberation. *JASA*, 10(3), 1962.

Smith, J. (1983) *Techniques for digital filter design and system identification with application to the violin*. Ph.D. dissertation, CCRMA. Department of Music, Stanford University.

Takala, T., Hänninen, R. Välimäki, V., Savioja, L., Huopaniemi, J., Huotilainen, T., & M. Karjalainen. (1996) An integrated system for virtual audio reality.

Presented at the *100th Audio Engineering Society (AES) Convention*. Preprint no. 4229 (M-4). Copenhagen, Denmark.

Author Information

DIVA Group

Helsinki University of Technology
Laboratory of Acoustics and Audio Signal Processing
Otakaari 5A, FIN-02150 Espoo, Finland

Helsinki University of Technology
Laboratory of Telecommunications Software and Multimedia
Otakaari 1, FIN-02150 Espoo, Finland

Jyri.Huopaniemi@hut.fi, Lauri.Savioja@hut.fi, Tapio.Takala@hut..fi

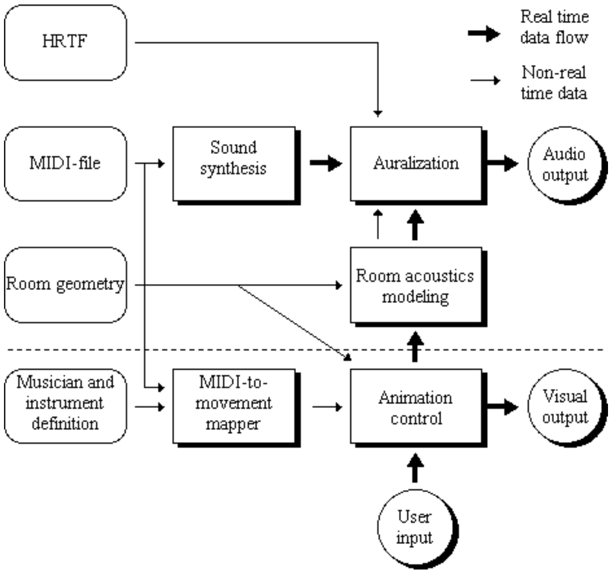

Figure 1: DIVA System Information flowgraph

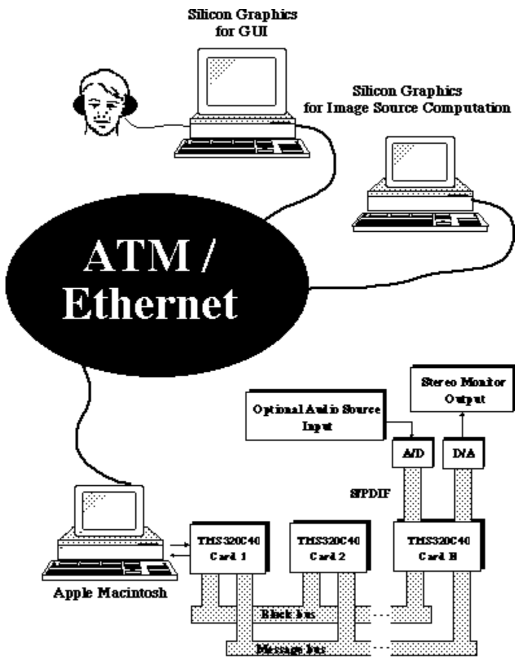

Figure 2: DIVA System Implementation