15th International Congress on Acoustics (ICA'95), Trondheim, Norway, June 26-30, 1995

Title of the paper:

Physical Models of Musical Instruments in Real-Time Binaural Room Simulation

Authors:

Jyri Huopaniemi[1], Matti Karjalainen[1,2], and Vesa Välimäki[1,3]

Affiliations:

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

[2]Stanford University, CCRMA, Stanford, CA 94305, USA

[3]CARTES, Ahertajankuja 4, FIN-02100 Espoo, Finland

SUMMARY

We introduce methods for designing a binaural listening environment using computer-generated sounds. The system is capable of creating a natural and realistic spatial simulation in real time. This environment has been developed by combining high-quality model-based sound synthesis of plucked string and woodwind instruments with binaural room simulation using a multiprocessor DSP programming environment. A new method for approximating head-related transfer functions (HRTF) using cepstral smoothing is proposed. The system and the techniques presented here are intended for applications of virtual reality, multimedia, and music technology.

MODEL-BASED SOUND SYNTHESIS

Physical modeling of musical instrument families such as plucked string and woodwind instruments has evolved to a new sound synthesis technique, called *model-based sound synthesis* [1]. The goal in this synthesis method is to model the acoustic-mechanical principles found in the instrument and to apply modern digital signal processing (DSP) techniques that allow for real-time processing. We have applied physical modeling techniques to plucked string and woodwind instruments. An overview of the modeling techniques can be found in [2] and [3].

The directivity patterns of musical instruments can also be taken into account in model-based sound synthesis. In a companion paper [4] we have considered three methods suitable for real-time processing. The method that we call *directional filtering* uses low-order digital filters that simulate the directivity of the instrument model. This set of filters is applied according to the direction of interest at the output of the instrument model.

Modeling of the directivity patterns of musical instruments is attractive both from the research and application point of view. Virtual reality environments such as the system described in this paper are capable of simulating moving and rotating sound sources and producing real-time 3D audio output for headphone or loudspeaker reproduction.

REAL-TIME MODELING OF ROOM ACOUSTICS

The goal of real-time processing places several constraints on the amount of detail that can be achieved in the simulation of room acoustics. We have concentrated on applying the *image-source* method for approximating the early reflections in a rectangular room. The image-source method is based on the theory of using phantom images of the desired source position in a bounded space. The Mth-order reflections of a sound source can be calculated by adding $N = L(L-1)^{M-1}$ images of the room with L boundaries and calculating the length of the path between the two points.

The frequency-dependent attenuation of the sound radiating from the source and its images is caused by three major factors: 1) absorption of the boundary materials, 2) the frequency-dependent air absorption, and 3) the attenuation due to the distance between the source and the listener. In DSP terms frequency-dependent filtering and attenuation can be accomplished with low-order digital filters. The image-source method that we have implemented assumes the modeled room to be rectangular and all the reflections to be specular.

An artificial reverberation model for approximating late reverberation in a room has been proposed in [2]. The fundamental idea is 1) to use the direct sound and image-source method for accurate modeling of the first reflections, and 2) to use outputs from the direct and imaged sources as inputs to the late reverberation filter. The late reverberation filter is a recursive structure consisting of comb filters and an allpass section. The combined image-source and late reverberation model simulates the impulse response of a rectangular room to the extent needed in realistic room simulation.

BINAURAL AURALIZATION

Auralization refers to processing of sounds in such a way that a three-dimensional illusion of a sound space in a room is created [5]. Modeling of human spatial hearing involves localization cues such as the *interaural time difference* (ITD) and the *interaural amplitude difference* (IAD). Frequency-dependent coloration due to reflections and diffraction from the torso and pinnae are modeled by measuring *head-related transfer functions* (HRTF). An HRTF represents a free-field transfer function from a fixed point in a space to the entrance of the test person's ear canal. In the real-time environment design we have used HRTFs based on measurements carried out on human subjects [2] and a dummy head [6].

Filter Design from Measured HRTF Data

It is necessary from the computational point of view to reduce the order of the HRTF filters when designing real-time simulation. Several methods for approximating measured HRTF data have been presented in the literature over the past few years. The methods have mainly concentrated on the design of finite-impulse response (FIR) filters, which can exactly preserve the phase behavior. This implies that the impulse response of, e.g., an HRTF measurement can be directly applied as an FIR filter. Suggested techniques for approximating HRTF magnitude have included the use of sliding third-octave window averaging [7] and principal components analysis (PCA) [8]. We have studied a method for reducing the order of the FIR filters using *cepstral smoothing*, which has been widely used in speech processing (see, e.g., [9]). The cepstrum $c(n)$ of the measured HRTF impulse response $h(n)$ has been defined by:

$$c(n) = \mathbf{F}^{-1}\left\{\log\left|\mathbf{F}\{h(n)\}\right|\right\} = \mathbf{F}^{-1}\left\{\log\left|H(\omega)\right|\right\} \tag{1}$$

where $\mathbf{F}$ and $\mathbf{F}^{-1}$ denote the operation of discrete Fourier transform (DFT) and inverse DFT, respectively, and $H(\omega)$ is the frequency response of $h(n)$. The cepstrum is multiplied by a window function (e.g., a Hanning window) to remove those peaks that correspond to rapid deviations in the frequency response.

$$\hat{c}(n) = c(n)w(n), \quad \text{where } w(n) = 0.5 - 0.5\cos\bigl(2\pi n/(2M+1)\bigr), \quad -M \le n \le M \tag{2}$$

Thereafter, the smoothed magnitude response is obtained by applying the DFT to the windowed cepstrum:

$$\hat{C}(\omega) = \mathbf{F}\{\hat{c}(n)\} \tag{3}$$

The smoothed HRTF impulse response $g(n)$ is then calculated by canceling the logarithmic operation and by adding the original phase function $\phi(\omega)$:

$$g(n) = \mathbf{F}^{-1}\left\{10^{\hat{C}(\omega)} e^{-j\phi(\omega)}\right\} \tag{4}$$

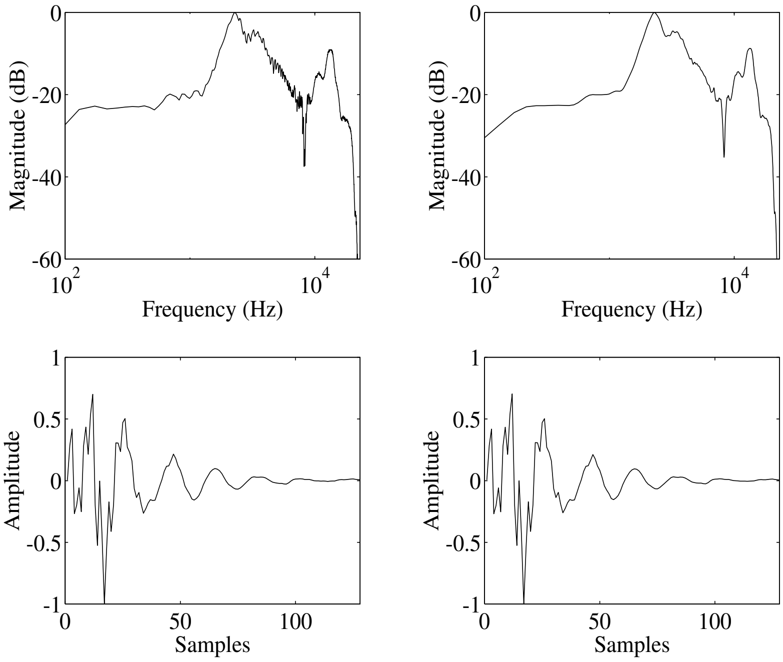

Fig. 1 *Magnitude response (upper left) and the impulse response (lower left) of the measured HRTF at 0° azimuth and 0° elevation, and the corresponding smoothed magnitude response (upper right) and the smoothed and windowed impulse response (lower right).*

Shown in Fig. 1 are the magnitude spectrum and the impulse response of an original HRTF measurement (of a dummy head [6]) and the corresponding cepstral smoothed magnitude and impulse response. A rectangular window was used for truncating the impulse response. It can be clearly seen that the overall shape of the magnitude response is well preserved, but the ripple resulting from noise and anomalies in the measurement are removed. The phase behavior of the HRTF shows nearly a minimum-phase characteristic, in the example case of order 128 some 9 zeros were located outside the unit circle. However, the original phase function of the HRTF is retained as seen in Eq. (4) to ensure the best possible result in the simulation.

OVERVIEW OF THE SIMULATION SYSTEM

The designed simulation system is currently functional in the Apple Macintosh platform. The hardware is based on a multiprocessing environment consisting of Texas Instruments TMS320C40 signal processors [2]. An overview of the complete system is illustrated in Fig. 2. The current system contains two C40 processors, one being dedicated to the room and auralization modeling and the other for running the physical models. The room simulation system is operated by a mouse-controlled user interface and the physical models are controlled via MIDI. The binaural audio output is directed to headphone listening. The current sampling rate used is 22.05 kHz. We are experimenting on an 8x8-matrix AD/DA converter which enables processing for multiple loudspeaker auralization.

APPLICATIONS OF VIRTUAL AUDIO REALITY

The aim of this project has been to develop an environment that would serve both as an expandable simulation system for model-based sound synthesis and as a stand-alone room simulation system for real-time auralization. The combination of three-dimensional sound and model-based sound synthesis has attractions to several hardware and software applications: 1) virtual reality environments, 2) computer games, 3) computer sound cards, and 4) stand-alone synthesizers. The computational efficiency is a strong constraint and compromises have to be made to achieve real-time processing in affordable hardware environments. Nevertheless, virtual acoustic environments containing physical models of musical instruments as sound sources are a step towards virtual audio reality.

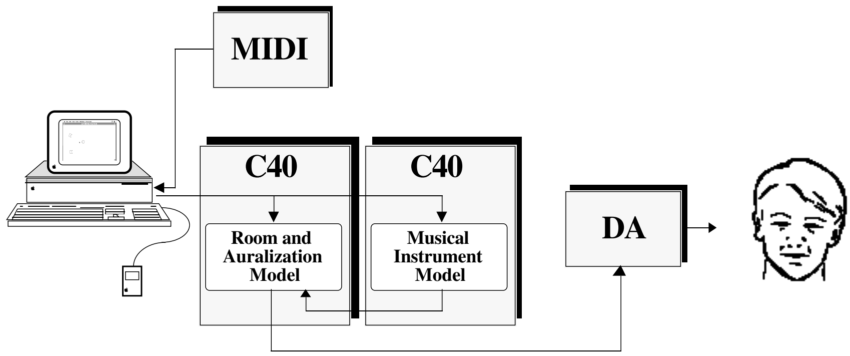

Fig. 2. *Overview of the designed real-time simulation environment. The system consists of two TMS320C40 signal processors, an Apple Macintosh host computer, and a MIDI interface for the control of the physical models.*

CONCLUSIONS

A real-time room simulating environment for model-based sound synthesizers has been proposed. The system produces real-time virtual three-dimensional audio output for stereo headphone listening. Radiation directivity of the musical instruments has been incorporated by applying direction-dependent low-order digital filters. Room simulation is performed with a combined image-source and late reverberation model. Binaural processing is carried out using digital filter approximations of measured HRTFs. The HRTF approximation is achieved by applying cepstral smoothing and windowing the smoothed impulse responses.

ACKNOWLEDGMENTS

The authors would like to thank Prof. Unto K. Laine for his valuable comments on the use of cepstral smoothing in the reduction of the HRTF filter order. Mr. Tommi Huotilainen is acknowledged for the design and implementation of the hardware for the simulation system. This study was financed by the Academy of Finland.

REFERENCES

[1] J. O. Smith, “Physical modeling using digital waveguides,” *Computer Music Journal*, vol. 16, no. 4, pp. 74–87, Winter 1992.
[2] J. Huopaniemi, M. Karjalainen, V. Välimäki, and T. Huotilainen, “Virtual instruments in virtual rooms—A real-time binaural room simulation environment for physical models of musical instruments,” in *Proc. Int. Computer Music Conf. (ICMC'94)*, Aarhus, Denmark, pp. 455–462, Sept. 12–17, 1994.
[3] V. Välimäki, J. Huopaniemi, M. Karjalainen, and Z. Jánosy, “Physical modeling of plucked string instruments with application to real-time sound synthesis,” presented at the *98th AES Convention*, Paris, France, Feb. 25–28, 1995.
[4] M. Karjalainen, J. Huopaniemi, and V. Välimäki, “Direction-dependent physical modeling of musical instruments,” in *Proc. 15th Int. Congress on Acoust.* (this proceedings), Trondheim, Norway, June 24–28, 1995.
[5] M. Kleiner, B.-I. Dalenbäck, and P. Svensson. “Auralization – an overview,” *J. Audio Eng. Soc*, vol. 41, no. 11, pp. 861–875, 1993.
[6] B. Gardner and K. Martin, “HRTF measurements of a KEMAR dummy-head microphone,” *MIT Media Lab Perceptual Computing*, Technical Report #280, May 1994.
[7] J. Köring and A. Schmidt, “Simplifying cancellation of cross-talk for playback of head-related recordings in a two-speaker system,” *Acustica*, vol. 79, pp. 221–232, 1993.
[8] D. J. Kistler, “A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction,” *J. Acoust. Soc. Am.*, vol. 91, pp. 1637–1647, 1992.
[9] A. V. Oppenheim and R. W. Schafer, “Homomorphic analysis of speech,” *IEEE Trans. Audio Electroacoust.*, vol. AU–16, pp. 221–226, June 1968.