

References

- [1] M.Kliner, B.Dalenack & P.Svensson, "Auralization - An Overview", J.Audio Eng.Soc., Vol. 41, No. 11, 1993
- [2] J.Blauret, "Spatial Hearing: The Psychophysics of Human Sound Localization", MIT Press, Cambridge, MA, 1983
- [3] N.Sakamoto, T.Gotoh & Y.Kimura, "On 'Out-of-Head Localization' in Headphone Listening", Proceedings of the AES 52nd Convention, 1975
- [4] Electronic Musician, October 1992, Vol. 8, No. 10, pp. 38
- [5] S.H.Nielsen, "Auditory Distance Perception in Different Rooms", J. Audio Eng. Soc., Vol. 41, No. 10, pp. 755-769, 1993
- [6] J.A.Moorer, "About this Reverberation Business", Computer Music Journal, Vol. 3, No. 2, pp. 13-28, 1979
- [7] J.Jot, "Digital Delay Networks for Designing Artificial Reverberators", Audio Eng. Soc. Preprint, No. 3030 (E-2), 1991
- [8] F.L.Wightman, & D.J.Kistler, "Headphone Simulation of Free-Field Listening: Stimulus Synthesis", J. Acoust. Soc., Vol. 85, pp. 858-867, 1989
- [9] F.L.Wightman, & D.J.Kistler, "Headphone Simulation of Free-Field Listening: Psycho-physical Validation", J. Acoust. Soc., Vol. 85, pp. 868-878, 1989
- [10] D.R.Begault, "Perceptual Effects of Synthetic Reverberation on Three-Dimensional Audio Systems", J. Audio Eng. Soc., Vol. 40, No. 11, 1992
- [11] D.R.Begault, "Challenges to the Successful Implementation of 3-D Sound", J. Audio Eng. Soc., Vol. 39, No. 11, 1991
- [12] G.R.Starks, "Virtual Reality Audio Applications", M.Mus. Dissertation, University of Glasgow, 1993
- [13] M.Cohen, N.Koizumi & S.Aoki, "Design and Control of Shared Conferencing Environments for Audio Telecommunication", Proceedings of the 2nd ISMIR Conference, 1992
- [14] W.R.Sherman, "Integrating Virtual Environments into the Dataflow Paradigm", Fourth Eurographics Workshop on VISC, 1993
- [15] C.Scaletti & A.B.Craig, "Using Sound to Extract Meaning from Complex Data", Extracting Meaning from Complex Data: Processing, Display, Interaction II, SPIE, Vol. 1459, 1991
- [16] G.S.Kendall, W.L.Martens, & M.D.Wilde, "A Spatial Sound Processor For Loudspeaker and Headphone Reproduction", Proceedings of the AES 8th International Conference, 1988
- [17] E.M.Wenzel, P.K.Sione, S.S.Fisher, & S.H.Foster, "A System for Three-Dimensional Acoustic Visualisation", in *Virtual Environment Workstation*, in Visualisation '90, IEEE Computer Society Press, pp. 329-337, 1990

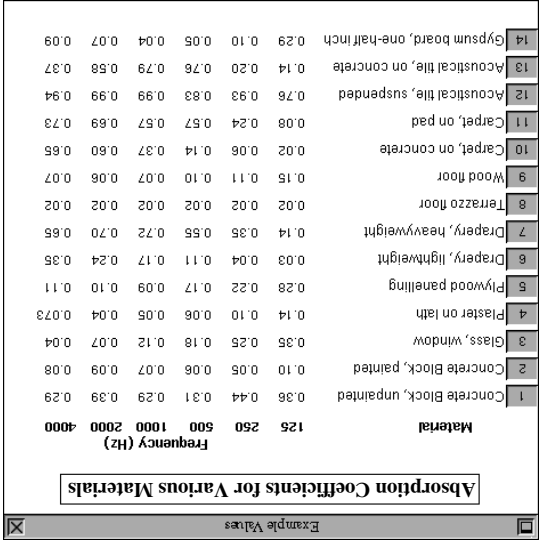


Figure 3(c) Example Absorption Coefficients

Once all data has been input, the user simply has to click on a single button to execute the processing: the application creates the relevant `Sound_orchestra` and `score` files and executes them. The resultant sound can then be monitored and the environment data edited if the result is not satisfactory.

Since the nature of `Csound` allows any number of instruments in an orchestra the user is not limited to the manipulation of a small number of sources. The Object-Oriented nature of `NeXTSTEP` facilitates the use of a large number of manipulation windows, allowing the user to enter positional data for a large number of sources. Although it is desirable to create an 'environment template', which is common to all sources, the user is not constrained to doing this. Positional templates can also be created which can be fitted within a particular time scale. Basic sound editing is also incorporated, allowing simple cut/copy/paste operations within, or between sound files, sample rate conversion between 8013, 22050 and 44100Hz and format conversion between Linear, Mu Law and Compressed, mono and stereo formats.

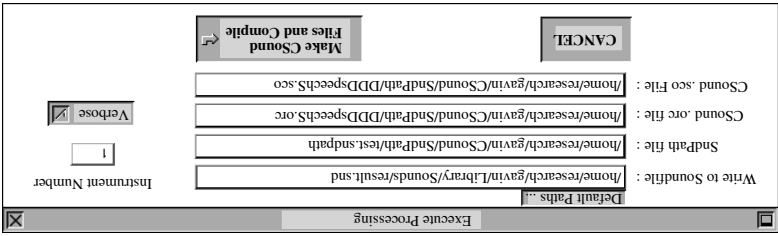


Figure 3(d) Process Execution Window

3 Applications

Applications of 3-D Stereo are abundant [1], [12], [13]. Music production has benefited from its current state, computer generated music is extremely lacking in spatial information. Although many synthesizers currently support 'onboard effects', they usually constitute of chorus and reverburation: the directional properties of the instruments are ignored. The inclusion of more specific directional properties may result in more pleasing effects.

Immersive VR systems are usually based around a helmet-mounted display (HMD) and 'data' glove. 3-D visual cues are related using a binocular stereoscopic viewing system and audio cues given over headphones. The position of the head and glove can be monitored using magnetic sensors above the operator, or micro-switches inside the apparatus. This allows the state of the visual and audio images to be updated as the operator moves, giving the impression of a 3-D environment.

Desktop VR systems also allow the creation and visualisation of 3-D environments, only on a 2D computer screen. In most cases, the operator moves around using a 'Spaceball' (allowing 6 degrees of freedom). As with the immersive system, a high degree of interaction can be achieved.

Recreation of three dimensional sound fields in VR is essential if the creation of any virtual 'world' is to be realistic. The benefits of VR graphics in areas such as data visualisation are obvious and are being developed at many sites. The additional benefits of 3-D sound are also beginning to be appreciated in a number of fields [14], [15], [16]. Development of effective methods with which to impart spatial information are therefore extremely important, particularly in areas such as the design of auditory icons.

The auditory system gives additional bandwidth to communicate information, particularly when there is a large amount of visual information. The use of sound in, for example, aircraft cockpit and EVA has proved extremely successful [17]. More diverse applications, such as visually and aurally mapping blood velocity to supplement diagnosis are beginning to appear. Research has also highlighted the importance of creating 'appropriate' audio icons for the sonification of data [15].

4 Conclusion

The development of graphical desktop VR systems appears to be sufficient to implement realistic 3-D graphics representations. It follows that acoustic manipulation can be approached in a similar manner, although ultimately an immersive application may be required to realise its full potential. This application has not implemented sophisticated 3-D graphical manipulation, but it is easy to envisage how external hardware (e.g. a VR glove or 'Spaceball') could be incorporated into the design of the interface.

It is hoped that, once complete, this application will benefit 'acoustic designers' and composers alike. Future developments include modification of the software to allow real-time manipulation. While this will require extensive software development and more powerful hardware (processing currently runs at ~8:1), parallel DSP configurations (e.g. Ariel's Quint Processor and Ircam's ISPW) are commercially available. Current audio results are effective: although many of the problems detailed earlier have been encountered, their solution is in hand.

The application of artificial approximations of the environment in conjunction with HRTFs should result in a convincing 'virtual' sound space. In this implementation filters are created which simulate the effects of air and wall absorption. These are convolved (in the frequency domain) with the appropriate HRTF for each direct/reflected signal. Global reverberation is then added using a combination of comb and all-pass filters.

2.1 Csound

Csound is an industry standard C based computer music language which allows users to manipulate signals with an extremely high degree of precision. Processing centres around two files: an 'orchestra' file and a 'score' file. The orchestra file contains definitions of how sounds ('instruments') are to be created or manipulated, and its associated score file contains the specific data pertaining to each event.

In this application, routines have been incorporated into the structure of Csound which enable the user to accurately position an arbitrary number of sound sources in a defined space by the addition of a simple function call in their Csound programs. The orchestra function call is of the form

```
asigl, asig2  sndspace  asound, "pathfile", xr, yr, zr, xu, yu, zu, xh, yh, zh, xs, ys, zs, w1, w2, w3, w4, w5, w6
```

where

asigl, asig2 are the (stereo) output signals from the calculations,
sndspace is the name of the function,

asound is the input sound,

"pathfile" is an (optional) file containing all positional data,

xr, yr, zy are the dimensions of the room (in meters)

xu, yu, zu are the users coordinates,

xh, yh, zh is the where the listeners head is pointing,

xs, ys, zs are the source coordinates,

w1, . . . , w6 are the absorption coefficients of the walls.

Since the spatial processing routines incorporated into Csound are written entirely in C, they are easily portable across a number of platforms. Clearly entering this data manually is tedious, especially if parameters are to be varied (e.g. if the sound moves across the sound stage). To solve this problem, a graphical interface has been implemented to allow the user to enter all relevant data in a more intuitive manner.

2.2 Graphical Interface

The main problem in the implementation of the interface was allowing the flexibility of control offered by Csound without compromising simplicity of operation. NeXT workstations allow easy manipulation of sample data and NeXTSTEP provides excellent interface development tools. Figure 3(a) shows the two main control windows with which the user manipulates positional data.

3

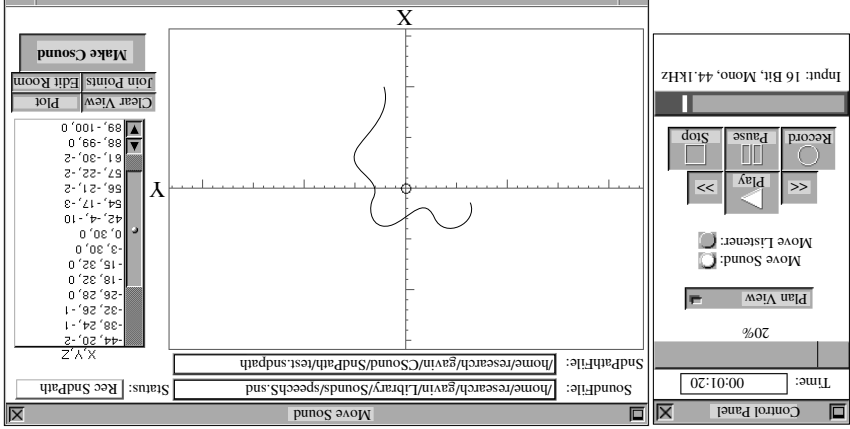
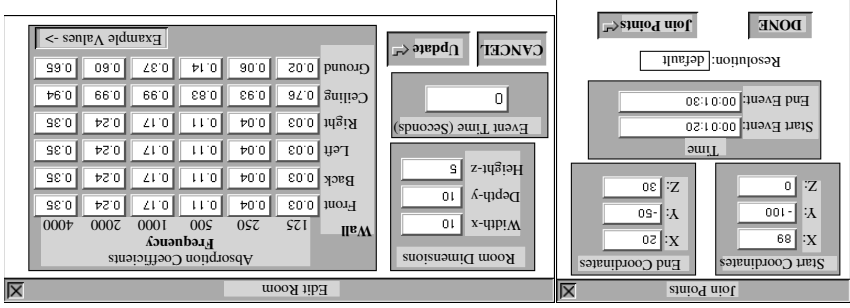


Figure 3(a) Main Control Windows

The Control Panel allows the user to initiate playback of any sampled sound, select the viewpoint for drawing and select the object to move (i.e. the Listener or the Source). Since the user is working with a 2D display, standard elevation, end elevation and plan views are used for plotting. As the sound is playing, the positional data entered using the mouse is recorded.

Subsidiary control windows are depicted in Figure 3(b). Points can be joined linearly over a certain time period. The 'Edit Room' window allows the (rectangular) room dimensions and the absorption coefficients of its walls to be varied with respect to time. Example values are also given which can be easily used as template data.

Figure 3(b) Subsidiary Control Windows



4

A 3-D Stereo Processing Tool

Gavin R Starks, Ken N Linton

Department of Music, 14 University Gardens,
University of Glasgow, Scotland. G12 8QH

e-mail: gavin@music.gla.ac.uk, knl@music.gla.ac.uk

Demands for spatial processing are increasing not only in music production, but in the rapidly growing field of virtual reality. A graphical interface has been constructed which allows users to easily manipulate a large number of sources. This approach aims to meet the need for high precision, multi-instrument spatialization without compromising simplicity of operation.

Introduction

Physical, mathematical and psychoacoustic theories allowing the accurate simulation of acoustic spaces are well established [1]. Research in the field of three-dimensional audio has focused mainly on commercial music processing, data visualisation and virtual reality (VR) systems.

3-D stereo allows the recreation of convincing ambience and directional cues through headphones. Directional cues are obtained using digital filters which simulate the filtering performed by the shape of a listeners head and ears, referred to as the head-related transfer function (HRTF) [2]. Ambience cues are achieved through the use of artificial reverberation [3].

Processing implementations do not yet readily involve the user within the context that he or she may desire. Many commercial systems often allow a small number of inputs and are usually aimed at manipulating sound in a post-processing manner [4]. The development of flexible graphical interfaces, similar to those used in the design of graphical VR environments is required to facilitate 3-D audio processing.

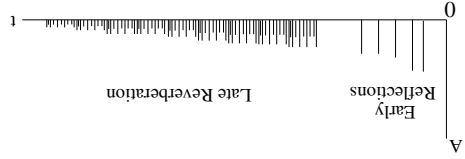
1 Basic Psychoacoustic Theory

Clearly the environment in which a sound emanates is a crucial factor when determining the resultant sound field heard by the listener. Physically accurate modelling of acoustic spaces yields computationally heavy algorithms. It is therefore desirable to obtain approximations which give satisfying distance and ambience cues and are more easily computable [5].

1.1 Acoustic Modelling

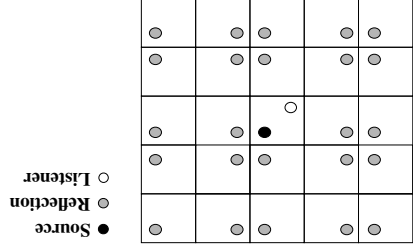
It is common practice to separate acoustic modelling into two areas: the 'early reflections' and the 'late reverberation' (see Figure 1). Early echoes are quite distinct from the more isotropic form of the reverberant tail and both areas can be modelled accordingly.

Figure 1 Early Reflections and Reverberation



Early reflections can be calculated by considering them as sources in 'virtual rooms' surrounding the listening room (see Figure 2). Evaluating the distance and direction of these virtual sources from successive shells results in an accurate simulation of the early signal.

Figure 2 Significant Early Reflections



1.2 Reverberation

Due to the computational load required to accurately model the dense pattern of echoes which form the reverberation process, algorithms based on recursive digital delay networks are used which attempt to simulate the natural environment. The main problem in this form of simulation is that, even if a large number of time-varying delay lengths are used, it is difficult to eliminate undesirable resonances [6], [7].

1.3 Head-Related Transfer Function

In binaural simulation, the two most obvious parameters to consider are the *inter-aural time delay* (ITD) and *inter-aural intensity difference* (IID). Consideration of these cues, however, results in the localisation of sounds in only one dimension: between the ears.

The human auditory system derives directional information by analysing the filtering performed by the torso, head, pinna and ear canal. Placing minute probe microphones in a listener's ear canals allows their individual frequency characteristics to be calculated. Measurements are made at a large number of angles around the head and the data from the resultant HRTF pairs applied directly as FIR digital filters [8], [9].

Application of this form of binaural sound processing is clearly subjective, HRTFs vary from person to person, thus reducing the effect for other listeners. Front-back ambiguity and in-head localization are also major problems, but they can be reduced by tracking the position of the listener's head [10], [11].