PinnaWorks: a NeXT Application for Three-Dimensional Sound Processing in Real-Time

Gary S. Kendall
School of Music, Northwestern University
Evanston, IL 60208 U.S.A.
(708) 491-3178
gary@music.nwu.edu

1. INTRODUCTION

PinnaWorks is a NeXT computer application for producing three-dimensional sound images. It was written in Objective-C and developed with the NeXT Interface Builder, version 2.0.1 PinnaWorks produces stereo audio output which can be reproduced either by headphones or by loudspeakers. PinnaWorks combines the audio input material with data from a set of directional cues. Each directional cue is captured as the coefficients of a pair of FIR filters, one for each output channel. Directional cues are typically derived from measurements of head related transfer functions (HRTFs) which capture the directionally dependent acoustic properties of the head, torso, and pinna, although synthetic cues derived by other means are also used. Because natural HRTFs are somewhat easier to grasp than synthetic cues, the following discussion will be illustrated with HRTFs measured at Northwestern University using the Kemar dummy head.2 The program's signal flow is represented in Figure 1. The FIR filters are implemented on the NeXT computer's on-board DSP chip, the Motorola 56000, while the user interface, the control information, and the files of directional cues are implemented on the NeXT's main processor. The discussion of the PinnaWorks that follows will focus on the user interface starting with its three main panels: the Control Panel, the Analysis Panel, and the Input/Output Panel.

2. CONTROL PANEL

The Control Panel is shown in Figure 2a.; its upper region is demarcated by a box labeled "Direction". The current spatial direction is expressed as an angular pair, azimuth and elevation, which the user specifies by manipulating the sliders and text forms. In Figure 2a, both the azimuth and elevation angles are set at 0-degrees, (0,0); this specifies a direction directly in front of the listener. In Figure 3a, azimuth and elevation are set to 90- and 32-degrees respectively, (90,32), indicating a direction to the left side with a slight elevation. The middle region of the panel, the area within the box labeled "HRTF Filters", reports information back to the user about the directional filters which implement the current spatial direction specified above. The azimuth and elevation angles of the closest corresponding measured HRTF are reported, both for the "ipsilateral" ear and the "contralateral" ear, these are respectively the closer and farther ear from the sound source. Obviously, the filter of directional cues is limited to a discrete number of spatial directions. At present, PinnaWorks' files of HRTFs quantify three-dimensional space by 10-degrees resolution in azimuth and elevation. In Figure 3a., the specified elevation of 32-degrees is shown to be rounded down to 30-degrees for the usual HRTF. This region of the panel also contains sliders and text forms for examining or altering interaural differences (interaural time difference, 'ITD', and interaural intensity difference, 'IID') and the output gain of the filters. Adjustments of these parameters have been found to be very helpful in

1 The conversion of the program to version 3.0 is still in process during the writing of this report. The present discussion will be limited to the 2.0 version.
optimizing imagery for specific reproduction settings and for individual listeners. Adjustments made by the user can be saved in customized HRTF sets.

The lowest region of the Control Panel is contained within a box labeled "HRTF sets". The "HRTF sets" are the external files which contain collections of HRTF measurements, modified HRTFs, or synthetic directional cues. Text forms are provided for as many as five files which the application can have open simultaneously. One set is always selected to be the "active" set by the user with the "Selector" radio buttons on the right side. The buttons enable the user to switch quickly between HRTF sets. Individual sets have been created with other programs for headphone and loudspeaker reproduction. Example sets include the direct measurements from the Kernar mannequin illustrated here, Kernar with either headphone or loudspeaker equalization, and the Atal-Schroeder cross-talk cancellation technique for loudspeaker reproduction. Users must "open", "save" and "close" these files from this region of the panel.

3. HRTF ANALYSIS PANEL

The "HRTF Analysis" panel provides a graphic representation of the characteristics of the current directional cues. Once again, the representation is organized by ipsilateral and contralateral ears. The upper region of the panel is contained within a box labeled "Time Domain". Here the HRTF impulse responses are represented along with text fields that report the peak amplitude and the arrival time of the peak amplitude both in msec and in samples. In Figure 2b, the sound direction is (0,0) indicating a sound source location which is directly in front of the listener and equally distant from each ear. In this case, the directional information for the ipsilateral and contralateral ears is identical. The resulting impulse responses are identical and there is no interaural intensity nor time difference. In Figure 3b, the sound direction is (90,30) indicating a sound source much closer to the left ear. The differences between the ipsilateral and contralateral impulse responses are quite striking. The ipsilateral ear signal is more intense and arrives much earlier in time than the contralateral ear signal.

The bottom region of the "HRTF Analysis" panel is contained within a box labeled "Frequency Domain". Here are presented spectral analyses of HRTFs. The (0,0) sound source direction illustrated in Figure 2b shows no interaural spectral difference. Even so, the spectral profile is quite complicated and includes multiple peaks and notches. The (90,30) sound direction illustrated in Figure 3b reveals not only interaural time and intensity differences, but spectral differences as well.

4. INPUT/OUTPUT PANEL

The Input/Output Panel has "Input" controls on the left and "Output" controls on the right as shown in Figure 4. Between are the "DSP" controls. The PinnaWorks user can select a real-time audio input source from a soundfile or from the A r t e l digital microphone. The input processed by PinnaWorks must be monophonic and the output is always stereo. With a stereo input signal, the user must select a mono-to-stereo conversion—the left channel alone, the right channel alone, or a left/right mix. The user can also select an audio output destination of either stereo D/A signal or a NeXT soundfile. Tape deck styled controls are provided for previewing or reviewing input and output soundfiles. Whenever either input or output soundfiles are being used, the DSP Start button automatically syncs and runs the tape deck controls. Ideally the sampling rate should be switchable, but with current NeXT hardware real-time processing DSP is limited to

---

3 All potential sound source locations that are equidistant from both ears lie on a plane called the median plane. Generally speaking, it is more difficult for listeners to judge the direction of sound sources on the median plane than elsewhere.

4 On the median plane, there are no interaural time or intensity differences, but there a variety of spectral profiles which depend on the elevation angle.
22K. In the case that both input and the output soundfiles are at 44K, the processing proceeds by dropping out of real-time.\(^5\)

5. MAIN MENU

The main menu provides buttons for "Info", access to the "Input Soundfile" and the "Output Soundfile" using standard NeXT UO panels, "Help", and "Quit". PinnaWorks "Help" facility not only explains the program's options, but also attempts to teach the user about the underlying basis of directional hearing cues and the practical application of these cues in audio reproduction.

6. REAL-TIME CONTROL

Besides the real-time audio output described above, PinnaWorks provides support for two types of the Polhemus position sensing system which is used either to track the position and orientation of the listener's head during headphone listening or to control the location of the three-dimensional sound image. In this configuration, the system is appropriate for virtual reality applications.

7. CONCLUSION

There are several limitations with the current implementation of PinnaWorks which will remain (even after the conversion to NeXT Interface Builder, version 3.0). Most important among these are the restriction of real-time audio to the 22K sampling rate and the long time latency of the Polhemus when used for head tracking. These are limitations imposed by the structure of the NeXT DSP implementation and can only be addressed by changes in system architecture. Future development will focus on providing more effective and specialized sets of directional cues.

8. REFERENCES


9. FIGURES

Figure 1. Signal flow overview.

---

\(^5\) This constraint is imposed because of the large number of FIR filter taps that the DSP is required to execute for real-time.
Figure 2. Control and Analysis Panels for (0,0).

Figure 3. Control and Analysis Panels for (90,32).

Figure 4. Input/Output Panel.