MEASUREMENT OF DISTANCE PERCEPTION USING VIRTUAL AUDIO

William D’Angelo, Mark Ericson 1
Eric Scarborough, Steven Rogers, Philip Amburn, and Dennis Ruck 2

1 Armstrong Laboratory
2 Air Force Institute of Technology
Wright-Patterson Air Force Base, OH

ABSTRACT

Relative auditory distance perception of a direct path signal by itself, with synthetic reflection(s), and with reverberation, presented over headphones, was measured using a 2AFC task. The stimulus, either a 500 millisecond pink noise burst or a two second phrase of male speech, was presented twice, first at a reference distance and second at an incremental distance from the reference. Reference distances included five, fourteen, and twenty-two feet, and the incremental distances included multiples of .25, .5, and 1 foot, respectively. Stimulus pairs at each of the three distances were presented from four directions: front, back, left, and right. For each stimulus pair, the task of three subjects was to indicate which of the two sounds appeared closer to himself/herself. From histograms for each condition, the just noticeable difference (JND) was calculated by determining the minimum interval at which the subjects could respond correctly seventy-five percent of the time. The results of all locations indicated that the JNDS for the direct path signal were seven percent, JNDS for the reflections were around six percent, and JNDS for the reverberant sounds were about twelve percent. All of the sounds appeared out-of-head, however the sounds presented from behind were perceived to be in front. Side sounds appeared to be biased towards the front when presented at the farther distances. Virtual audio cues were shown to be effective in creating the perception of auditory distance over headphones.

INTRODUCTION

Most humans can perceive, separate, and locate sound sources in three dimensions (3-D) and correlate the apparent source location with the visual and vestibular senses. In nature, this ability serves as a low resolution early-warning system and is useful in pointing the high resolution visual system to a location of interest. A large body of experimental data on auditory localization capabilities has been collected over the last century and continues to grow (Blauert, 1983; Yost and Gourevitch, 1987; and Durlach, 1991).

Recently, the synthesis of the cues necessary to give directionality to a sound source heard over headphones has become feasible (McKinley, 1988). Virtual audio is the creation of the illusion of a sound source existing at a location in 3-D space, either over an array of loudspeakers or headphones, by encoding a signal.
with various localization cues.

Virtual audio may help a pilot to perform a variety of functions in the military cockpit. The visual workload may be reduced by providing a spatial auditory beacon for navigation waypoints. The current audio alarm used by the radar warning receiver could be more effective if it led the pilot's eyes to the area of danger (Folds, 1990). The intelligibility of speech has been shown to increase by spatially separating the communications channels (Ericson and McKinley, 1992). Formation flying at night or in bad weather may be aided by attaching audio "running lights" to wingmen. Virtual audio may also provide a means for pilots to preserve their spatial orientation in addition to existing visual cues.

Background

The process of encoding localization cues can be framed in terms of a transmitter-medium-receiver relationship, in this case sound generation-auralization-head related characteristics. Each sound source has its own spectral and temporal qualities. Properties of the source such as the bandwidth, pitch, degree of modulation, intensity and directionality play an important role in localization. Familiarity with the sound source and a priori information affect the listener's perception. For a familiar source, the level of the sound as it reaches the ears can give an indication of distance (Moore, 1989).

The effects that the acoustic environment imposes on the sound source before it is received by the ears are called auralization cues. Reflections and reverberation occur when a sound wave bounces off structures such as the ground or walls. These structures absorb and pass sound differently depending on the frequencies of the sound and the material of the structures. At distances greater than 100 feet, the atmosphere begins to absorb high frequency sounds (Coleman, 1963). An inverse relationship exists between the ratio of direct path to reflected energy and the source distance (von Bekesy, 1960). These auralization cues are the primary contribution to the perception of distance (Moore, 1989).

The final factor in localization is the transformation of the sound source due to an individual's pinnae, head and torso. The fact that a person has two ears separated by the width of the head produces two cues: a difference in the time of arrival and in the spectral content of the soundwave reaching each ear. The unique shape of each ear and the mass of the head and shoulders alter the frequency spectrum that reaches the eardrum in a characteristic way for each person. These two cues combined are called the head related transfer function (HRTF). The cues included in one's HRTFs enable the localization of sounds in azimuth and elevation. The time information dominates the perception of azimuth, while the spectral cues are important for elevation perception (Wightman and Kistler, 1989).

While ample experimental data exists for human performance in azimuth and elevation, data for distance performance is scarce. This situation is reflected in the state of virtual audio technology in that the performance in azimuth is very successful, the perception of elevation is good and improves with training, while the use of distance cues has not been fully successful. Therefore, the need exists for research into distance performance. Existing and current virtual audio technology can be very useful to meet this need.

OBJECTIVE AND PURPOSE
The objective of this work was to study the effectiveness of using intensity scaling, reflections, and reverberation to create the perception of distance over headphones. The purpose was to incorporate distance cues into virtual audio synthetic environments.

METHOD

Equipment

A distance perception experiment was conducted at the Air Force Institute of Technology (AFIT). In order to study synthetic audio and visual environments, AFIT has combined two 3-D Audio Generators with a Silicon Graphics, Inc. (SGI) Personal Iris 4D/35 and an accompanying SGI audio processor board (Scarborough, 1992). This setup provided a powerful testbed with flexibility and control of both graphics and digital audio (Figure 1).

The Bioacoustics and Biocommunications Branch (AL/CFBA) of the Armstrong Laboratory, United States Air Force, has developed the 3-D Audio display generator which provides virtual acoustic capabilities. In the procedure developed by AL/CFBA, the combined frequency transformations of the pinna and head of a mannequin are measured for 272 points in 3-D space. These transformations were recorded in the center of a 14 foot diameter geodesic sphere of loud speakers in an anechoic chamber. The HRTFs are stored in memory and used to encode the sound source, separately for each ear, to make the sound appear to come from a particular location.

The use of a pair of two-channel 3-D Audio Generators provided four independent channels for simulating four spatially separate acoustic sources. The SGI audio processor board allowed the coordination (time delays and intensity scaling) of the four sources to produce one source consisting of a direct path and up to three reflections. Reverberation was provided with a Realistic (Cat. No. 32-1110B) electronic reverberation box. The delay between echoes was set at 15 ms. Two echoes were produced with a successive 75% attenuation for each. The stimulus was presented to the subject over Sennheiser HD520 headphones.

In software, a virtual room was created. It was 60 feet in length and 12 feet wide, with the subject centered (Figure 2). Virtual sound sources could then be placed along the center of the room, lengthwise, from 3 to 30 feet on either side of the subject. With the source in any location, the subject could hear the direct path sound source alone or accompanied by any combination of a floor reflection or a right or left wall reflection. The software would calculate the appropriate time delays and intensity scaling of the reflections using an acoustic ray tracing program to determine path length and the inverse ratio law to simulate absorption. The SGI main processor would pass these factors to the audio processing board.

Stimuli

When the subject pressed the appropriate mouse button, the audio processor board would play the sampled stimulus over the four 3-D Audio channels, modifying them as instructed by the main processor. The stimulus consisted of a digitized, 500 millisecond, pink noise burst. The noise burst was bandlimited between 100 Hz and 10 kHz with a 50 ms onset and offset time. Also, the digitized, two second phrase "fear of lawyers" was used to test the effects of sound source familiarity. The subject's response was entered using the mouse. Data files were used to set...
the initial conditions, control the experiment trials, and to store the subject’s response data.

The sound source locations were divided into two sets: the primary and the secondary sources. The primary sources were set at ±5, 14, and 22 feet from the subject. Each primary source had eight secondary sources associated with it: four in front and four behind. The minimum interval used was one quarter foot, yielding secondary sources for the primary source at 5 feet at 4, 4.25, 4.5, 4.75, 5.25, 5.5, 5.75, and 6 feet. The interval for the 14 foot source was one half foot and the interval was one foot for the source at 22 feet.

Experimental Design

The experiment consisted of a two alternative, forced choice task in which the subject responded with which of the two sounds appeared closer. In each trial, the subject heard a primary sound source followed immediately by a secondary sound source. The subject was not allowed to repeat the trial presentation. A section consisted of 48 trials corresponding to six primary source locations, each with eight secondary source locations. In each section, the 48 source pair locations were randomized.

The eight sections corresponded to four reflection conditions for two head-orientation angles. The reflection conditions consisted of (1) a direct path signal, (2) a direct path and a floor reflection, (3) a direct path with reflections from walls on either side of the subject, and (4) a direct path with both floor and wall reflections. In half of the sections, the subject faced forward in the virtual room, while in the other half, the subject faced to the right. The order of the eight sections was randomized for each subject. Each subject was trained for at least one hour (eight sections for each of the noise and speech stimuli) but with intervals twice as large as those used in the actual experiment. The subjects were presented the eight sections with the noise stimulus, the speech stimulus, and the noise stimulus with reverberation.

Subjects

The three subjects tested included a female and two males. One of the males, the experimenter, had prior experience with the distance stimuli; the other two subjects had no prior experience in distance perception experiments. All subjects were in their early twenties and had normal hearing sensitivity and function.

RESULTS

Data was collected for the three subjects and percent just noticeable difference (%JND) was calculated for different experimental conditions. From histograms for each condition, the %JND was calculated by first determining the minimum interval at which the subjects could respond correctly (the sound they perceived to be closer was actually designed to be closer) seventy-five percent of the time. The length of this interval was then divided by the distance from the subject to the primary source. For example, if the primary source was at 5 feet, while for a certain set of conditions the subjects reached seventy-five percent correct at the second interval (one half foot), then the %JND would be 10%. In several cases, all using reverberation without reflections, the percent correct never reached seventy-five percent indicating the need for more or larger intervals.

The first relationship studied was the effect of reflections and reverberation in general. For those sections without reflections the %JND was 7%, while for those
sections with reflections, a 6% JND was calculated. With the limited data set, the addition of reflections did not affect the judgement. When reverberation was included and when the subjects were able to reach seventy-five percent correct, a 12% JND was calculated, indicating a two-fold increase in the difficulty of discrimination.

When the data was broken down for the three different source distances (Table 1), the effect of reflections was found to depend on the source location. For the five foot source, the %JND without reflections was 9% (.45 ft.), while with reflections added, the %JND dropped to 5% (.25 ft.). With the source at fourteen feet, the opposite relationship was found. The %JND without reflections was 5% (.7 ft.), while the addition of reflections increased the %JND to 8% (1.12 ft.). The reflections had no effect on the discrimination at twenty-two feet. No difference was found for the results using the noise versus speech stimuli in terms of the effects of reflections.

As mentioned previously, the case of reverberation without reflections did not yield seventy-five percent correct responses for the five and fourteen foot source locations. For the source at twenty-two feet, the %JND was 15% (3.3 ft.). When the reflections were included with reverberation, a 12% JND (2.64 ft.) was calculated.

While being trained, the subjects were questioned about their impression of sound source location, absolute distance, and the discrimination task. All subjects reported externalization of the sound source, more so with the speech stimulus than the noise burst. The reflections and reverberation were reported to add volume or spaciousness to the sound source.

The stimuli were never perceived to be coming from the rear of the subject. Sounds presented in the median plane sometimes appeared to be biased slightly to either side. For the sources presented on the sides, at the farther distances, the positions were perceived to wrap around towards the front of the subject. Elevation was at the horizon or ear level.

All subjects reported hearing a close, medium and far source region. The subjects consistently heard the nearest sounds at an arms length from the head. The absolute judgement of the further distances was more difficult with estimates ranging from ten to one hundred feet.

**DISCUSSION**

Although the data set is small, some interesting trends emerged. As the reflections are configured presently, they do not contribute greatly to the discrimination task. The reflections are important in that they do impact the perception of the sound source by giving it volume. In one case, the reflections may aid in the discrimination task. When the sound source is close to the head, the distance discrimination seems to be achieved by an intensity judgement alone. The energy of the reflections may increase the intensity difference heard between sources that are near to each other, easing the discrimination.

The addition of reverberation, while adding color to the sound, has a negative effect on the subjects' discrimination capability. The task appears particularly difficult under the contradictory condition of having reverberation without reflections. The reverberation segment of the experiment should be repeated with larger intervals so that seventy-five percent correct performance is reached.

For the noise burst stimuli, a
pitch shift was heard when the distance was changed. This cue destroyed the perception of a distance change as externalization was lost. Subjects reported attempting to use pitch shift as a means of discrimination but were advised that pitch shift was an unreliable cue and to attempt to visualize the source at a distance. With training and concentration this unwanted percept could be overcome. For close source locations, the intensity alone was most often used as the means of discrimination.

This experiment demonstrated the functionality of an audio distance simulator using a virtual environment. Because of the processing power available, data can be collected without the need for extensive physical hardware. Also, with the flexibility of the system, room dynamics can be simulated with modifications to the software only. Without the need for mobile physical sources, the length of time for an experiment is much less.

Further experimentation is planned to include collecting more data with the present setup, studying the role of differential Doppler shifts, the role of visual cues, the effect of training, and of measuring absolute distance judgement.

REFERENCES


TABLE 1: RESULTS OF DISTANCE DISCRIMINATION TASK

<table>
<thead>
<tr>
<th>Source Location:</th>
<th>5 FT</th>
<th>14 FT</th>
<th>22 FT</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition</td>
<td>JND</td>
<td>JND</td>
<td>JND</td>
<td>JND</td>
</tr>
<tr>
<td>Without Reflections</td>
<td>9%</td>
<td>5%</td>
<td>7%</td>
<td>7%</td>
</tr>
<tr>
<td>With Reflections</td>
<td>5%</td>
<td>8%</td>
<td>6%</td>
<td>6%</td>
</tr>
<tr>
<td>Reverberation Alone</td>
<td>-</td>
<td>-</td>
<td>15%</td>
<td>13%</td>
</tr>
<tr>
<td>Reverb and Reflections</td>
<td>15%</td>
<td>14%</td>
<td>7%</td>
<td>12%</td>
</tr>
</tbody>
</table>