M.I.T Media Laboratory Perceptual Computing Section Technical Report No. 342
July 6, 1995

Transaural 3-D audio

William G. Gardner
MIT Media Lab
20 Ames Street, Room E15-401B
Cambridge, MA 02139
email: billg@media.mit.edu

Abstract

An audio system has been constructed that renders three-dimensional sound using conventional stereo loudspeakers. This paper discusses the implementation and performance of the system.

1 Introduction

This paper describes a recently developed 3-D audio system that can position sounds at arbitrary azimuths and elevations around a listener's head. The system uses stereo loudspeakers arranged conventionally (say at ±30 degrees with respect to the listener). The system works by reconstructing the acoustic pressures at the listener's ears that would occur with a free-field sound source at the desired location. This is accomplished by combining a *binaural spatializer* with a *transaural audio system*. The spatializer convolves the source sound with the direction dependent filters that simulate the transmission of sound from free-field to the two ears. The resulting binaural output of the spatializer is suitable for listening over headphones. In order to present the audio over loudspeakers, the output of the spatializer is fed to a transaural audio system which delivers binaural signals to the ears using stereo speakers. This system filters the binaural signals so that the crosstalk leakage from each speaker to the opposite ear is canceled.

This technology is applicable to situations where a single listener is in a reasonably constrained position facing stereo speakers. Common examples of where the known listener position can be exploited include a person using a computer workstation, playing a video game, driving a car, playing a console electric piano, or making a video-phone call. The 3-D audio technology has obvious applications to multimedia presentation, entertainment, immersive simulation, and teleconferencing.

The technology used in this system is not new, and the performance is far from ideal. This paper will discuss the implementation and performance of the system, and will suggest ways to greatly improve system performance by adapting to the listener's head position and orientation.

2 Principles of binaural spatial synthesis

Humans have a remarkable ability to localize sounds in three dimensions, despite having only two ears. Consider a sound source to the right of a listener. Sounds from the source arrive at the right ear first, and a short time later reach the left ear. The amplitude of the left ear sound will be attenuated, particularly at high frequencies, due to head shadowing. The predominant auditory cues for determining whether a sound is coming from the left or right directions are the interaural time difference (ITD) and the interaural intensity difference (IID). However, the auditory system can also determine whether sounds are in front of or behind the listener, and can estimate the elevation of sound sources. This is possible because the incident sound waves interact with the torso, head, and external ear (pinna) prior to arriving at the inner ear. The various reflections and diffractions cause a spectral modification of the sound which depends on the direction of incidence. The human auditory system is able to use the directional dependent filtering to discriminate front from rear directions and also determine the elevation of sounds. The interaural differences are particularly important cues, because they don't depend on the details of the source spectrum. Human auditory localization has been studied extensively [2, 19].

The directional dependent filtering to each ear can be expressed as a frequency response, called a head-related transfer function (HRTF)[1], and thus a pair of HRTFs describes how sound from one location reaches the two ears. Depending on the application, a head response may be measured at the entrance of the ear canal or at the eardrum. The ear canal response itself has been shown to be invariant of source position [11]. HRTFs are usually measured using human subjects or dummy-head microphones, and consist of response pairs, for the left and right ears, corresponding to a large number of source positions surrounding the head.

For source locations close to the head, the spherical curvature of the incident sound waves cause the HRTFs to change qualitatively as a function of distance. At moderate distances from the head, where the incident waves can be considered planar, the HRTFs do not change as a function of distance, except for the expected inverse relationship between amplitude and distance. At extreme distances, atmospheric effects including dispersion and lowpass filtering can audibly alter the incident spectrum, but these affect both ear responses equally. Consequently, the auditory cue for source distance at moderate distances simply depends on the loudness of the sound. If the sound is familiar, then prior knowledge of the expected loudness can be compared to the actual loudness to roughly determine distance. The presence of reverberation can enhance our distance percep-

[1] The time domain equivalent of an HRTF is called a head-related impulse response (HRIR) and is obtained via the inverse Fourier transform of an HRTF. In this paper, we will use the term HRTF to refer to both the time and frequency domain representation.

tion. In general, the level of background reverberation is roughly constant throughout a room, whereas the level of the initial sound decreases as a function of distance from the source. Thus, it is possible to compare the loudness of the initial sound with the subsequent reverberation to estimate the source distance.

A binaural spatializer simulates the auditory experience of one or more sound sources arbitrarily located around a listener. The basic idea is to reproduce the acoustical signals at the two ears that would occur in a normal listening situation. This is accomplished by convolving each source signal with the pair of HRTFs that correspond to the direction of the source. The resulting binaural signal is presented to the listener over headphones. A schematic diagram of a single source system is shown in figure 1. The direction of the source (θ = azimuth, ϕ = elevation) determines which pair of HRTFs to use, and the distance (r) determines the gain. Figure 2 shows a multiple source spatializer that adds a constant level of reverberation to enhance distance perception.

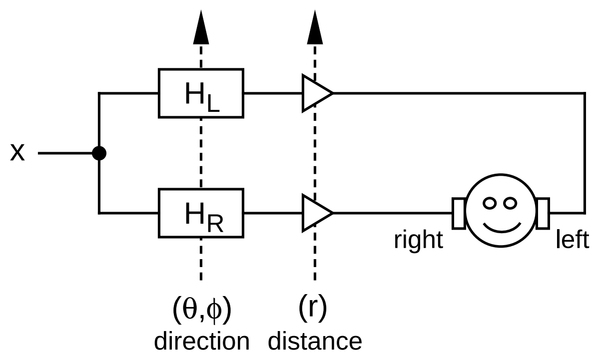

Figure 1: Single source binaural spatializer.

The simplest implementation of a binaural spatializer uses the measured HRTFs directly as finite impulse response (FIR) filters. Because the head response persists for several milliseconds, HRTFs can be more than 100 samples long at typical audio sampling rates. The interaural delay can be included in the filter responses directly as leading zero coefficients, or can be factored out in an effort to shorten the filter lengths. It is also possible to use mimimum phase filters derived from the HRTFs, since these will in general be shorter than the original HRTFs. This is somewhat risky because the resulting interaural phase may be completely distorted. It would appear, however, that interaural amplitudes as a function of frequency encode more useful directional information than interaural phase [8]. Because interaural differences are of paramount importance for localization, any attempt to parameterize pairs of HRTFs to simplify implementation should focus on the interaural differences as an error metric.

There are several problems common to headphone spatializers:

- The HRTFs used for sythesis are often a generic set and not the specific HRTFs of the listener. This can cause localization performance to suffer [16, 17], particularly in regards to front-back discrimination, elevation perception, and externalization. When the listener's own head responses are used, their localization performance is comparable to natural listening [18].
- The auditory scene created moves with the head. This can be fixed by dynamically tracking the orientation of the head and updating the HRTFs appropriately. Localization performance and realism should both improve when dynamic cues are added [16].
- The auditory images created are not perceived as being external to the head, but rather are localized at the head or inside the head. Externalization can be improved by using the listener's own head responses, adding reverberation, and adding dynamic cues [4].
- Frontal sounds are localized between the ears or on top ot the head, rather than in front of the listener. Because we are used to seeing sound sources that are in front of the head, it is difficult to convince the perceptual system that a sound is coming from the front without a corresponding visual cue [16]. However, when using the listener's own HRTF's, frontal imaging with headphones can be excellent.

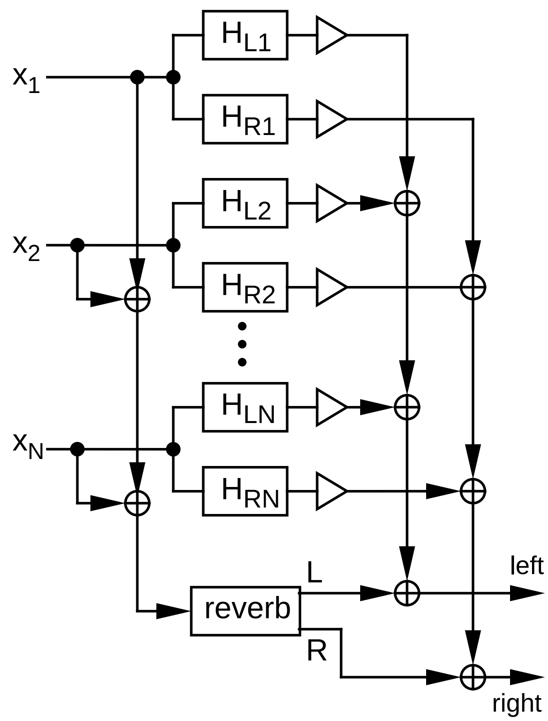

Figure 2: Multiple source binaural spatializer with reverberation.

3 Principles of transaural audio

Transaural audio is a method used to deliver binaural signals to the ears of a listener using stereo loudspeakers. The basic idea is to filter the binaural signal such that the subsequent stereo presentation produces the binaural signal at the ears of the listener. The technique was first put into practice by Schroeder and Atal [14, 13] and later refined by Cooper and Bauck [3], who referred to it as "transaural audio". The stereo listening situation is shown in figure 3, where $\hat{x_L}$ and $\hat{x_R}$ are the signals sent to the speakers, and y_L and y_R are the signals at the listener's ears. The system can be fully described by the vector equation:

$$\mathbf{y} = \mathbf{H}\hat{\mathbf{x}} \tag{1}$$

where:

$$\mathbf{y} = \begin{bmatrix} y_L \\ y_R \end{bmatrix}, \mathbf{H} = \begin{bmatrix} H_{LL} & H_{RL} \\ H_{LR} & H_{RR} \end{bmatrix}, \hat{\mathbf{x}} = \begin{bmatrix} \hat{x}_L \\ \hat{x}_R \end{bmatrix} \tag{2}$$

and H_{XY} is the transfer function from speaker X to ear Y. The frequency variable has been omitted.

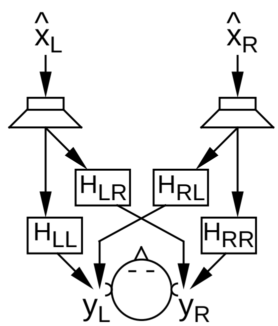

Figure 3: Transfer functions from speakers to ears in stereo arrangement.

If $\mathbf{x}$ is the binaural signal we wish to deliver to the ears, then we must invert the system transfer matrix $\mathbf{H}$ such that $\hat{\mathbf{x}} = \mathbf{H}^{-1}\mathbf{x}$. The inverse matrix is:

$$\mathbf{H}^{-1} = \frac{1}{H_{LL}H_{RR} - H_{LR}H_{RL}} \begin{bmatrix} H_{RR} & -H_{RL} \\ -H_{LR} & H_{LL} \end{bmatrix} \tag{3}$$

This leads to the general transaural filter shown in figure 4. This is often called a crosstalk cancellation filter, because it eliminates the crosstalk between channels. When the listening situation is symmetric, the inverse filter can be specified in terms of the ipsilateral ($H_i = H_{LL} = H_{RR}$) and contralateral ($H_c = H_{LR} = H_{RL}$) responses:

$$\mathbf{H}^{-1} = \frac{1}{H_i^2 - H_c^2} \begin{bmatrix} H_i & -H_c \\ -H_c & H_i \end{bmatrix} \tag{4}$$

Cooper and Bauck proposed using a "shuffler" implementation of the transaural filter [3], which involves forming the sum and difference of x_L and x_R, filtering these signals, and then undoing the sum and difference operation. The sum and difference operation is accomplished by the unitary matrix $\mathbf{D}$ below, called a shuffler matrix or MS matrix:

$$\mathbf{D} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \tag{5}$$

It is easy to show that the shuffler matrix $\mathbf{D}$ diagonalizes the matrix $\mathbf{H}^{-1}$ via a similarity transformation:

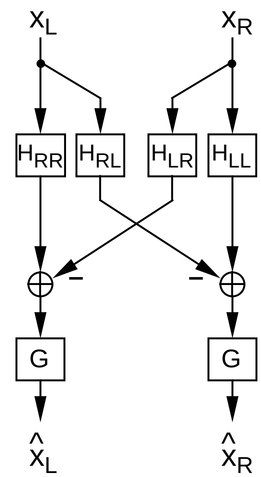

Figure 4: General transaural filter, where $G = 1/(H_{LL}H_{RR} - H_{LR}H_{RL})$.

$$\mathbf{D}^{-1}\mathbf{H}^{-1}\mathbf{D} = \begin{bmatrix} \frac{1}{H_i + H_c} & 0 \\ 0 & \frac{1}{H_i - H_c} \end{bmatrix} \tag{6}$$

Thus, in shuffler form, the transaural filters are the inverses of the sum and the difference of H_i and H_c. Note that $\mathbf{D}$ is its own inverse. This leads to the transaural filter shown in figure 5. The $1/\sqrt{2}$ normalizing gains can be commuted to a single gain of 1/2 for each channel, or can be ignored.

In practice, the transaural filters are often based on a simplified head model. For instance, the responses H_i and H_c may be derived from a spherical head model [3]. This can lead to a simple implementation of the transaural filter, and one that is not specific to a particular listener. At high frequencies, where pinna response becomes important (> 8 kHz), the head effectively acts as a baffle, and therefore the crosstalk between channels is small. Furthermore, the variation in head response for different people is greatest at high frequencies [11]. Consequently, there is little point in modeling pinna response when constructing a transaural filter.

The transaural system is fragile in the sense that it only works for a single listener in a known position and orientation. The system is tolerant of small deviations in head position, but large changes cause the transaural filtering to fail to deliver the proper binaural signal to the listener, often resulting in the perception of sounds originating at the loudspeakers.

4 Implementation of binaural spatializer

Our implementation of the binaural spatializer is quite straightforward. The HRTFs were measured using a KEMAR (Knowles Electronics Mannequin for Acoustics Research), which is a high quality dummy-head microphone. The HRTFs were measured in 10 degree elevation increments from -40 to +90 degrees [6]. In the horizontal plane (0 degrees elevation), measurements were made every 5

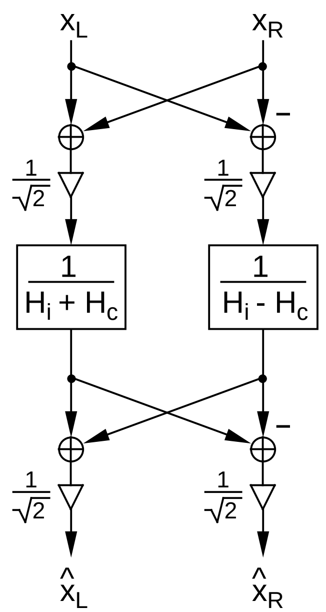

Figure 5: Shuffler implementation of transaural filter for symmetric listening arrangement.

degrees of azimuth. In total, 710 directions were measured. The sampling density was chosen to be roughly in accordance with the localization resolution of humans. The HRTFs were measured at a 44.1 kHz sampling rate.

The raw HRTF measurements contained not only the desired acoustical response of the dummy head, but also the response of the measurement system, including the speaker, microphones, and associated electronics. In addition, the measured HRTFs contained the response of the KEMAR ear canals. This is undesirable, because the final presentation of spatialized audio to a listener will involve the listener's own ear canals, and thus a double ear canal resonance will be heard. One way to eliminate all factors which do not vary as a function of direction is to equalize the HRTFs to a diffuse-field reference [7]. This is accomplished by first forming the diffuse-field average of all the HRTFs:

$$|H_{DF}|^2 = \frac{1}{N} \sum_{i,k} |H_{\theta_i,\phi_k}|^2 \qquad (7)$$

where H_{θ_i,ϕ_k} is the measured HRTF for azimuth θ_i and elevation ϕ_k. $|H_{DF}|^2$ is therefore the power spectrum which would result from a spatially diffuse soundfield of white noise excitation. This formulation assumes uniform spatial sampling around the head. The HRTFs are equalized using a minimum phase filter whose magnitude is the inverse of $|H_{DF}|$. Thus, the diffuse-field average of the equalized HRTFs is flat[2]. Figure 6 shows the diffuse-field average of the HRTFs. It is dominated by the ear canal resonance at 2-3 kHz. The low-frequency dropoff is a result of the poor low-frequency response of the measurement speaker. The inverse equalizing filter was gain limited to prevent excessive noise amplification at extreme frequencies.

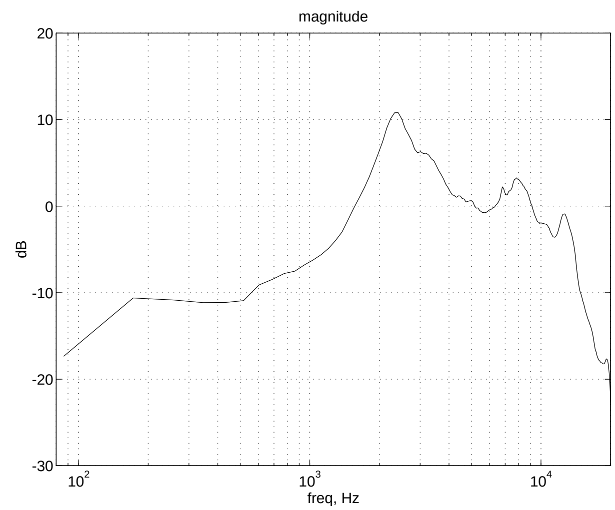

Figure 6: Diffuse-field average of KEMAR HRTFs.

It was desired to build a realtime spatializer on a Silicon Graphics Indigo workstation. After an initial attempt at an implementation running at 44.1 kHz sampling rate, it was decided to reduce the sampling rate to 32 kHz. This required resampling all the HRTFs to 32 kHz. The equalized and resampled HRTFs were cropped to 128 points (4 msec) which was more than sufficient to capture the entire head response including interaural delays.

The spatializer simply convolves an monophonic input signal with a pair of HRTFs to produce a stereophonic (binaural) output. The HRTFs that are closest to the desired azimuth and elevation are used. For efficiency, the convolution is accomplished using an overlap-save block convolver [12] based on the fast Fourier transform (FFT). Because the impulse response is 128 points long, the convolution is performed in 128-point blocks, using a 256-point real FFT [5]. The forward transforms of all HRTFs are pre-computed. For each 128-point block of input samples (every 4 msec), the forward transform of the samples is calculated, and then two spectral multiplies and two inverse FFTs are calculated to form the two 128-point blocks of output samples. In addition to the convolution, a gain multiplication is performed to control apparent distance. The propagation delay inherent in the block convolution algorithm is not noticable, especially compared to the much larger blocksizes of the input and output audio buffers. The total audio propagation delay from input to output is approximately 40 msec, which is not a problem for our application.

It is essential that the position of the source can be changed smoothly without introducing clicks into the output. This is easily accomplished as follows. Every 12 blocks (48 msec) the new source position is sampled and a new set of HRTFs is selected. The input block is convolved with both the previous HRTFs and the new HRTFs, and the

[2] The equalization would normally be done separately for each ear to adjust for differences between left and right channels. However, our HRTFs were derived from a single ear measured for the full 360 degrees azimuth [6].

two results are crossfaded using a linear crossfade. This assumes reasonable correlation between the two pairs of HRTFs. Subsequent blocks are processed using the new HRTFs until the next position is sampled. The sampling rate of position updates is about 20 Hz, which is quite adequate for slow moving sources.

5 Performance of binaural spatializer

The binaural spatializer (single source, 32 kHz sampling rate) runs in realtime on an SGI Indigo workstation. The software is written in C and C++, and the implementation is very simple. Source position is controlled using a MIDI (Musical Instrument Digital Interface) controller that has a set of sliders. Three sliders are assigned to control azimuth, elevation, and distance (gain). A constant amount of reverberation can be mixed into the final output using an external reverberator as shown in figure 2.

The spatializer was evaluated using headphones (AKG-K240, which are diffuse-field equalized [15, 9, 10]). The input sound, usually music of various sorts, was taken from one channel of a compact disc recording. The spatializer worked quite well for lateral and rear directions for all listeners. As expected, some listeners had problems with front-back reversals. Elevation control off the medial plane was also good, though this varied considerably among listeners. For the author, elevation worked extremely well below and above the horizontal plane for non-medial azimuths. All listeners experienced poor externalization of frontal sounds. At zero degrees elevation, as the source was panned across the front, the perception was always of the source moving through the head between the ears, or sometimes over the top of the head. Externalization was far better at lateral and rear azimuths. Adding reverberation did improve the realism of the distance control, but did not fix the problem of frontal externalization. The author has listened to a spatializer based on his own HRTFs and the frontal imaging (even without reverb) is remarkable. Clearly a problem of using non-individualized HRTFs with headphones is the difficulty of externalizing frontal sources.

In order to increase the number of sources, or to add integral reverberation, the performance of the spatializer would need to be improved. Several things could be done:

- Reduce the filter size to 96 or 64 points by simple cropping (rectangular windowing).
- Reduce the filter size by factoring out the interaural delay and implementing this separately from the convolution.
- Reduce the filter size by using minimum phase filters.
- Model the HRTFs using infinite impulse response (IIR) filters.

Many of these strategies are discussed in [7]. To obtain the best price/performance ratio, commercial spatializers attempt to be as efficient as possible, and usually run on dedicated DSPs. Consequently, the filters are modeled as efficiently as possible and the algorithms are hand-coded. In addition, there are usually serious memory constraints which prevent having a large database of HRTFs, and thus parameterization and interpolation of HRTFs is an important issue. Lack of memory is not a problem in our implementation.

6 Implementation of transaural filter

Our transaural filter is based on a simplified head model for sources at standard stereo speaker locations. The ipsilateral response is taken to be unity and the contralateral response is modeled as a delay, attenuation, and a lowpass filter. Hence

$$H_i(z) = 1, H_c(z) = gz^{-m} H_{LP}(z) \qquad (8)$$
$$H_{LP}(z) = \frac{1-g}{1-gz^{-1}}$$

where $g < 1$ is a broadband interaural gain, m is the ITD in samples, and $H_{LP}(z)$ is a one-pole, DC-normalized, lowpass filter that models the frequency dependent head shadowing. For a horizontal source at 30 degrees azimuth, typical contralateral parameters might be g = 0.85 (-1.5 dB broadband attenuation), m = 7 (ITD of 0.2 msec at 32 kHz) and a lowpass filter cutoff of 1000 Hz (the frequency where head shadowing becomes significant). These parameters were not in fact calculated but were established through a calibration procedure discussed later.

Using this simplified head model the transaural filter in shuffler form is given by:

$$\mathbf{H}^{-1}(z) = \mathbf{D} \begin{bmatrix} \frac{1}{1+gz^{-m}H_{LP}(z)} & 0 \\ 0 & \frac{1}{1-gz^{-m}H_{LP}(z)} \end{bmatrix} \mathbf{D} \qquad (9)$$

This filter structure is efficiently implemented using only two delays and two lowpass filters.

The transaural filter is calibrated as follows. A standard stereo listening setup was constructed with speakers at ± 30 degrees with respect to the listener. Several stereo test signals are sent through the transaural filter and presented to the listener. The signals include stereo uncorrelated pink noise, left only pink noise and right only pink noise, and commercial binaural recordings made with dummy head microphones. During playback, the listener can continuously adjust the three transaural parameters (g, m, and the lowpass cutoff frequency) using a MIDI controller. The calibration procedure involves adjusting the parameters such that single sided noises are located as close as possible to their corresponding ears and the stereo noise is maximally enveloping. The interaural delay parameter has the most effect of steering the signal and changing the timbre, provided the gain parameter is sufficiently close to 1. The lowpass cuttoff has the most subtle effect. Interestingly, while it is possible to steer the single sided noise close to the corresponding ear, this often has the effect of moving the opposite sided noise closer to its corresponding speaker. Consequently a compromise has to be reached. In general, the final parameters one obtains via the calibration procedure agree with the physics of the situation. However, inexperienced listeners are confused by the procedure and are unable to correctly calibrate the system.

Listening to the binaural recordings through the transaural system is an enjoyable experience. The speakers vanish and are replaced by an immersive auditory scene. Sounds can be heard from all directions except the rear. The system is tolerant of head motion, particularly front-back translation, less so for left to right translation, and least tolerant of turning side to side. After turning beyond about 30 degrees, the scene breaks down and sounds are localized as originating from the speakers.

7 Performance of combined system

The binaural spatializer and transaural filter were combined into a single program. This runs in realtime on an SGI Indigo workstation.

Listening to the output of the binaural spatializer via the transaural system is considerably different than listening over headphones. Overall, the spatializer performance is much improved by using transaural presentation. This is primarily because the frontal imaging is excellent using speakers, and all directions are well externalized. The drawback of transaural presentation is the difficulty in reproducing extreme rear directions. As the sound is panned from the front to the rear, it often suddenly flips back to a frontal direction as the illusion breaks down. Most listeners can easily steer the sound to about 120 degrees azimuth before the front-back flip occurs. It is easier to move the sound to the rear with the eyes closed.

Elevation performance with transaural presentation parallels that of headphone presentation: if the cues work with headphones, they will work with transuaral speakers. Because the sounds are more externalized with the speakers, changing either the azimuth or elevation induces more apparent motion than with headphone presentation. Many listeners reported that changing the elevation also caused the azimuth to change. For instance, starting the sound directly to the right and moving it up often causes the sound to move left towards center before it reaches overhead.

All the performance evaluation discussed is completely informal. It would be useful to devise an efficient procedure for evaluating the performance of such systems, one that does not require lengthy training sessions or experimentation.

8 Conclusions

We have discussed a single source transaural spatializer that runs in realtime on an SGI Indigo workstation. Despite a straightforward implementation, the informal performance results are quite good.

We are currently working to improve this basic system by adding dynamic head tracking. The tracking information, consisting of the head orientation and position, will be used in two ways:

1. The direction and amplitude of sources will be calculated relative to the current head orientation and position, such that the auditory scene remains fixed as the listener's head moves, rather than moving with the head. This should improve localization performance by adding dynamic cues and should also improve the naturalness of the presentation.

2. Head orientation and position will be used to update the transaural filter coefficients, such that the proper binaural signals are delivered to the ears as the listener moves. In practice, using only two speakers, the transaural filter will only be effective for a range of listener orientations and positions. However, we believe that a dyanamic transaural system will be far more powerful than the current static transaural filters. It may also be possible to add more speakers, and choose the combination of speakers that leverage the situation the best (most likely those that are on opposite sides of the listener's head).

We intend to use a visual head tracking system being developed in the Vision and Modeling group at the Media Lab. The system uses two cameras to identify the 3-D position of the head and hands of a subject seated in front of the cameras [1]. A system to track position and orientation of a head can easily be constructed by requiring the subject to wear a cap marked with both a red dot and a green dot. By tracking the dots with stereo cameras, both orientation and position can be recovered. Eventually the tracking may be possible without wearing a special cap, by recognizing facial features such as the eyes and nose. Ultimately we imagine a visual system capable of reconstructing the 3-D structure of the listeners head, and using this information to synthesize individual HRTFs for the listener, though this is not currently feasible.

References

[1] Ali Azarbayejani, Thad Starner, Bradley Horowitz, and Alex Pentland. Visually controlled graphics. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 15(6):602–605, June 1993. (Special Section on 3-D Modeling in Image Analysis and Synthesis).

[2] Jens Blauert. *Spatial Hearing*. MIT Press, Cambridge, MA, 1983.

[3] Duane H. Cooper and Jerald L. Bauck. "Prospects for Transaural Recording". *J. Audio Eng. Soc.*, 37(1/2):3–19, 1989.

[4] N. I. Durlach, A. Rigopulos, X. D. Pang, W. S. Woods, A. Kulkarni, H. S. Colburn, and E. M. Wenzel. "On the Externalization of Auditory Images". *Presence*, 1(2):251–257, 1992.

[5] W. G. Gardner. "Efficient convolution without input-output delay". *J. Audio Eng. Soc.*, 43(3):127–136, 1995.

[6] W. G. Gardner and K. D. Martin. "HRTF measurements of a KEMAR". *J. Acoust. Soc. Am.*, 97(6):3907–3908, 1995.

[7] J. M. Jot, Veronique Larcher, and Olivier Warusfel. "Digital signal processing issues in the context of binaural and transaural stereophony". In *Proc. Audio Eng. Soc. Conv.*, 1995.

[8] Keith D. Martin. A computational model of spatial hearing. Master's thesis, MIT Dept. of Elec. Eng., 1995.

[9] Henrik Moller, Dorte Hammershoi, Clemen Boje Jensen, and Michael Fris Sorensen. "Transfer Characteristics of Headphones Measured on Human Ears". *J. Audio Eng. Soc.*, 43(4):203–217, 1995.

[10] Henrik Moller, Clemen Boje Jensen, Dorte Hammershoi, and Michael Fris Sorensen. "Design Criteria for Headphones". *J. Audio Eng. Soc.*, 43(4):218–232, 1995.

[11] Henrik Moller, Michael Fris Sorensen, Dorte Hammershoi, and Clemen Boje Jensen. "Head-Related Transfer Functions of Human Subjects". *J. Audio Eng. Soc.*, 43(5):300–321, 1995.

[12] Alan V. Oppenheim and Ronald W. Schafer. *Discrete Time Signal Processing*. Prentice Hall, Englewood Cliffs, NJ, 1989.

[13] M. R. Schroeder. "Digital simulation of sound transmission in reverberant spaces". *J. Acoust. Soc. Am.*, 47(2):424–431, 1970.

[14] M. R. Schroeder and B. S. Atal. "Computer simulation of sound transmission in rooms". *IEEE Conv. Record*, 7:150–155, 1963.

[15] G. Theile. "On the Standardization of the Frequency Response of High-Quality Studio Headphones". *J. Audio Eng. Soc.*, 34:956–969, 1986.

[16] E. M. Wenzel. "Localization in Virtual Acoustic Displays". *Presence*, 1(1):80–107, 1992.

[17] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman. "Localization using nonindividualized head-related transfer functions". *J. Acoust. Soc. Am.*, 94(1):111–123, 1993.

[18] F. L. Wightman and D. J. Kistler. "Headphone simulation of free-field listening". *J. Acoust. Soc. Am.*, 85:858–878, 1989.

[19] William A. Yost and George Gourevitch, editors. *Directional Hearing*. Springer-Verlag, New York, NY, 1987.