

Corey I. Cheng* and
Gregory H. Wakefield†

*Dolby Laboratories
100 Potrero Avenue
San Francisco, California 94103-4813, USA
cnc@dolby.com

<http://www.eecs.umich.edu/~coreyc>
*This work was completed at the University
of Michigan, Ann Arbor.*

†University of Michigan
Department of Electrical Engineering
and Computer Science

1101 Beal Ave.
Ann Arbor, Michigan 48109, USA
ghw@umich.edu
<http://www.eecs.umich.edu/~ghw>

Moving Sound Source Synthesis for Binaural Electroacoustic Music Using Interpolated Head-Related Transfer Functions (HRTFs)

Recent advances in computational power, acoustic measuring techniques, and hearing technology have made sound spatialization and moving sound source synthesis both popular and widely accessible sound-sculpting tools. In particular, much attention has been given to the use of Head-Related Transfer Functions (HRTFs), filters that mimic the directionally dependent filtering of the human external ear. HRTFs have been used in headphone- and loudspeaker-based spatialization systems to simulate the spectral cues responsible for directional hearing.

This article presents some techniques on how to compose with interpolated HRTFs to synthesize moving sound sources for binaural electroacoustic music intended for headphone listening. In this sense, we intend the article to describe some pragmatic compositional techniques and “rules of thumb” that may serve as a link between spatially based musical ideas and the scientific and technological realities of current binaural, headphone-based sound reproduction systems, such as those described in Kendall (1995).

We first review duplex theory, a simple model that explains directional hearing in the azimuth (left-right) direction. Next, we introduce HRTFs and show how a novel, spatially based visualization strategy for HRTF data employing Spatial Frequency Response Surfaces (SFRSs) provides important insight about the structure of HRTF data. We

then define interpolated HRTFs, describe why they are important for synthesizing moving sound sources for headphone listening, and describe a new method for computing interpolated HRTFs based on observations made from the SFRS-based visualization of HRTF data.

Using a MATLAB-based implementation of a simple spatialization system that uses interpolated HRTFs to synthesize moving sound sources, we describe the capabilities and limitations of HRTF-based moving sound source synthesis with some observations made from informal listening experiments. Based on these observations, we suggest that current technical restrictions define certain compositional problems that a composer might choose to solve, avoid, or exploit to produce musical results. We isolate left-right, front-back, and up-down spatial trajectories, demonstrate some techniques useful for emphasizing these trajectories, and show how these techniques can be used to express certain spatially based musical ideas.

There are several binaural sound examples on the accompanying compact disc which demonstrate many of the spatialization techniques discussed here. These binaural examples have been specifically processed to be listened to over a good pair of headphones. Nonetheless, some of these effects are more successful than others, and we realize that not all listeners may be able to immediately hear the intended spatial effects. Because spatialization effects can be delicate and may vary somewhat from person to person, we suggest listening to each

sound example in a quiet environment several times over headphones, closing one's eyes to better concentrate on the sound. Sound examples 1–9 on the accompanying CD contain demonstrations that isolate certain spatialization techniques, while sound examples 10–16 contain spatialized excerpts from *Fishbowl*, a short piece of binaural electro-acoustic music composed to exploit these spatialization techniques. Sound example 17 contains the entire composition, *Fishbowl*.

Background

We begin our background discussion with an overview of duplex theory.

Duplex Theory

Duplex theory is a perceptual model for estimating a sound's spatial location using two binaural cues: interaural time differences and interaural intensity differences (Rayleigh 1907). An *interaural time difference* (ITD) is defined as the difference in arrival times of a sound's leading wavefront at the left and right ears. Similarly, an *interaural intensity difference* (IID) is defined as the amplitude difference generated by a sound in the free field between the left and right ears. In general, a sound is perceived to be closer to the ear at which the first wavefront arrives, where a larger ITD translates to a larger lateral displacement. However, at frequencies above about 1500 Hz, the wavelength of sound becomes comparable to the diameter of the head, the head starts to shadow the ear farther away from the sound, and ITD cues can become ambiguous. At these higher frequencies, the IID generated by head shadowing becomes important to perceptual decoding of azimuth angle. Loosely speaking, perceived azimuth varies approximately linearly with the logarithm of the IID (Blauert 1983).

Although the simplicity and success of duplex theory are attractive, the theory only explains the perception of sounds in the azimuth or "left-right" direction. If one attempts to apply duplex theory to estimate a sound's location in the free field where

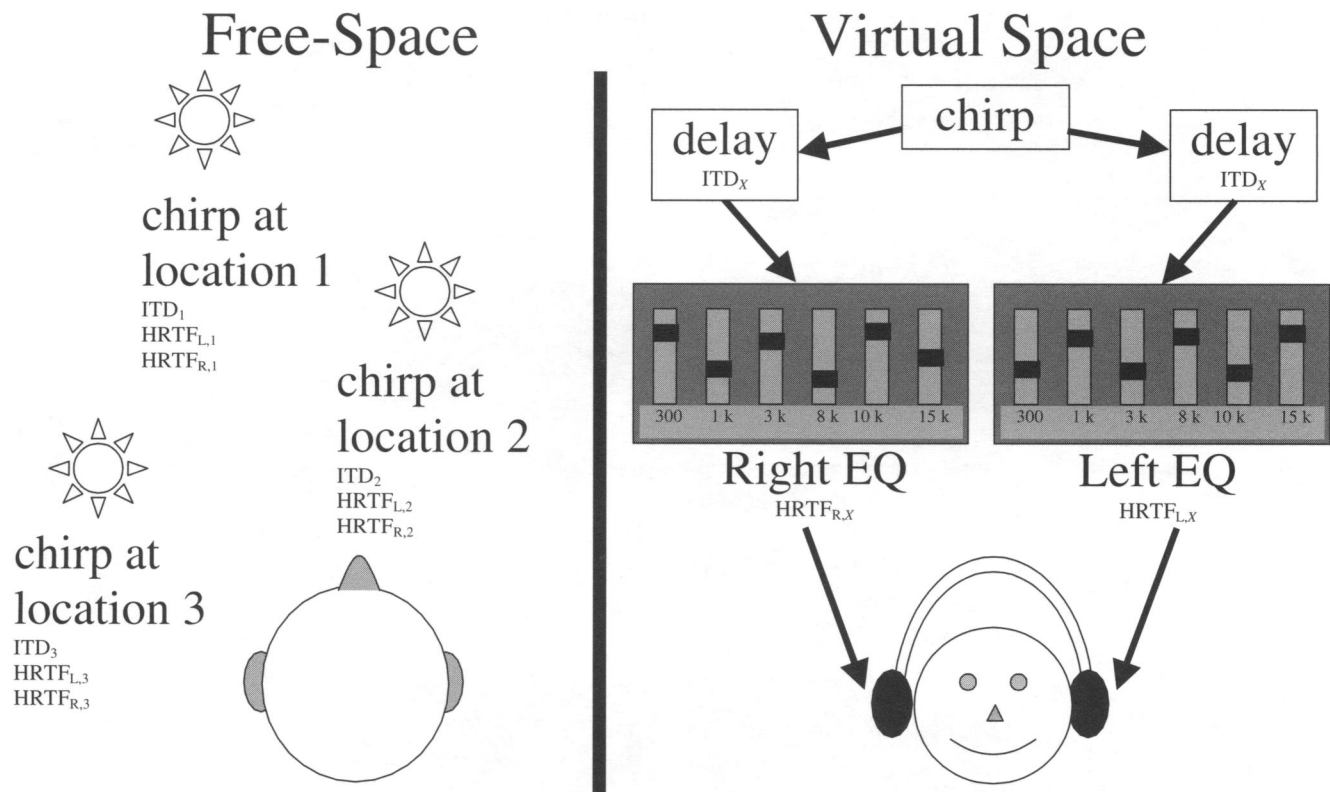
location becomes a function of azimuth, elevation, and distance, ITDs and IIDs do not specify a unique spatial location, because there are an infinite number of locations along curves of equal distance from the head having the same ITD and IID. The problem is acute for locations in the median plane, which separates the ears and runs vertically through the head. At all points in this plane, ITD and IID are ideally zero, so that interaural difference information is at a minimum.

Head-Related Transfer Functions (HRTFs)

Because listeners can differentiate among the locations of sounds in the free field, it is widely thought that the auditory system relies on ITD, IID, and other spectral cues to determine spatial location. These spectral cues are called Head-Related Transfer Functions (HRTFs) and summarize the direction-dependent acoustic filtering a sound undergoes owing to interactions with the head, torso, and outer ear (or pinna). Intuitively, HRTFs are simply filters that mimic the acoustic filtering of the head, torso, and pinna and operate on free-field sounds much as graphic equalizers operate on recorded sounds, as is shown in Figure 1. Different HRTFs corresponding to different spatial locations can be described by different equalizer settings. HRTFs are useful because they can be used to filter a monaural sound into a binaural, stereo sound which will sound as though it originates from a prescribed spatial location.

Formally, a single HRTF is defined as a particular subject's left or right far-field frequency response as measured from a specific point in the free field to a specific point in the ear canal. Because the overall spectral contours of an individual's HRTFs do not change significantly as a function of distance (Duda and Martens 1998), both left- and right-ear HRTFs are empirically measured from humans or mannequins for several azimuths (locations along the "left-right" direction) and elevations (locations along the "up-down" direction) at a fixed radius from the head. Thus, spatial location is designated by an ordered pair of angles (azimuth θ , elevation ϕ), where $(0^\circ, 0^\circ)$ corresponds to the location di-

Figure 1. Analogy of HRTFs to graphic equalizers. HRTFs filter sounds much like equalizers do, where different equalizer settings correspond to different HRTFs measured at different spatial locations.



rectly in front of a listener. Similarly, $(-90^\circ, 0)$ and $(+90^\circ, 0)$ correspond to locations directly opposite the left and right ears, respectively; $(0^\circ, -45^\circ)$ and $(0^\circ, +45^\circ)$ correspond to locations in front-below and in front-above the listener, respectively. In addition to ordered pairs of angles, spatial location may also be described with the general terms *ipsilateral* and *contralateral*, meaning “on the same side of the head as the sound source” and “on the opposite side of the head as the sound source,” respectively. For example, the right ear is the contralateral ear for sounds originating on the left side of the head.

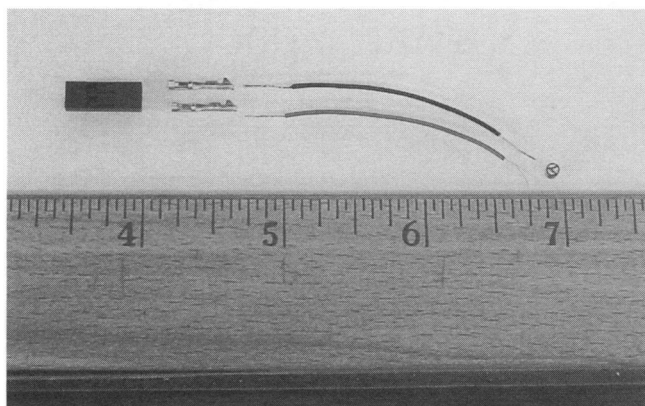
Technically, HRTFs are commonly specified as minimum phase filters. Note that an HRTF subsumes both ITD and IID information: time delays are encoded into the filter’s phase spectrum, and IID information is related to the overall power of the filter. However, HRTFs have been found empirically to be minimum phase systems (Kulkarni et al. 1995, 1999), which allows us to simplify the Fi-

nite Impulse Response Filter (FIR) description of HRTFs in two important ways. First, the minimum phase assumption allows us to uniquely specify an HRTF’s phase by its magnitude response alone, because the log magnitude frequency response and the phase response of a minimum phase causal system form a Hilbert transform pair (Oppenheim and Schaffer 1989). Second, the minimum phase assumption allows us to separate ITD information from the FIR specification of HRTFs. Because minimum phase filters have the minimum group delay property and minimum energy delay property, most of an HRTF’s energy occurs at the beginning of its impulse response, so that the left and right ear minimum phase HRTFs both have approximately zero delay. Thus, complete characterization of the auditory cues associated with a single spatial location involves the measurement of three quantities: left and right ear HRTF magnitude responses and the ITD.

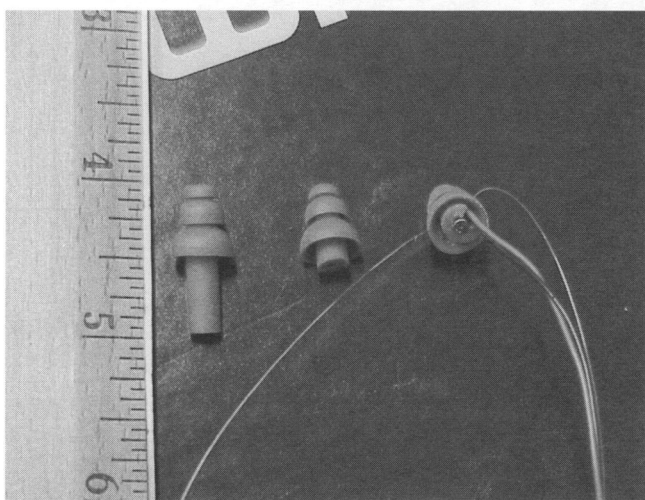
There are several different methods for measur-

Figure 2. HRTF measurement apparatus at the Naval Submarine Medical Research Laboratory (NSMRL). (a) Microphone

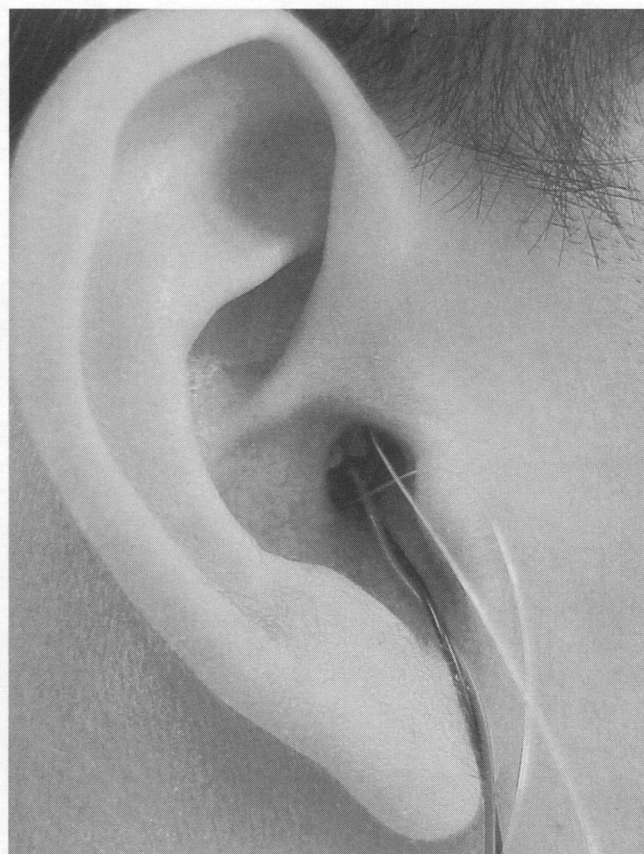
assembly. (b) Earplug assembly. (c) Insertion of earplug assembly into subject's ear.



(a)



(b)



(c)

ing HRTFs. A common technique used to empirically measure a specific subject's HRTFs is to insert small microphones partially into a subject's ears, and then to perform a simple form of system identification by playing a known-spectrum stimulus through a loudspeaker placed at a specified azimuth, elevation, and distance from the subject's head (Wightman and Kistler 1989a).

Figure 2 shows the HRTF measurement apparatus used and co-constructed by the authors and members of the auditory department at the Naval Submarine Medical Research Laboratory (NSMRL). HRTF measurement can involve intricate equipment setups, including specially designed speaker arrays and anechoic chambers, and thus the measurement process can be complicated, involve sig-

nificant subject time, and be prone to error (Cheng and Wakefield 2001a). To manage the complexity of the process, HRTFs are usually not measured at all spatial locations, but instead are measured only at a few hundred spatial locations. The sparseness of the spatial sampling pattern directly influences the computation of interpolated HRTFs for other, non-measured spatial locations, and this subject is discussed in depth below.

Because measured HRTF data sets appear to contain some patterns, there has been considerable interest in finding both analytical and numerical methods that explain certain features of HRTF data. The earliest physical models for HRTFs were derived by calculating the pressure on a rigid sphere due to an incident plane wave (Rayleigh 1945). There have also been attempts to model HRTFs by analyzing physical models (Shaw and Teranishi 1968; Shaw 1974). Other work in spatial

Figure 2. (continued) (d)
Subject sitting in anechoic
chamber with headtracker
and inserted earplug as-
sembly.



hearing has attempted to find low-order, perceptually acceptable parameterizations of HRTFs. Some of these methods include singular value decomposition and principal components analysis (Kistler and Wightman 1992; Blommer 1996), pole-zero models (Blommer 1996; Jenison 1995), “beamforming” models of the external ear (Chen et al. 1992), and boundary element models of the ear based on concatenated ducts (Speyer 1999).

Spatial Frequency Response Surfaces (SFRSs) and Interpolated HRTFs

Although a stationary sound occupies only one point in virtual space, a moving sound passes through many points in virtual space. Thus, in order to synthesize a moving sound source traversing

an arbitrary spatial trajectory using HRTFs, HRTFs corresponding to every possible spatial location are required. However, as discussed above, HRTFs are usually only measured at a prescribed set of spatial locations. Therefore, HRTF-based moving sound source synthesis requires the calculation of “interpolated HRTFs,” HRTFs corresponding to arbitrary spatial locations computed from existing, previously measured HRTFs.

Producing perceptually acceptable, interpolated HRTFs is a difficult problem. To create a reasonable interpolation strategy that preserves important structures in HRTF data, we must first examine the structure of HRTF data by asking more elementary questions. What features does a typical HRTF exhibit? What is the most intuitive way of examining and comparing these large, irregularly sampled HRTF data sets?

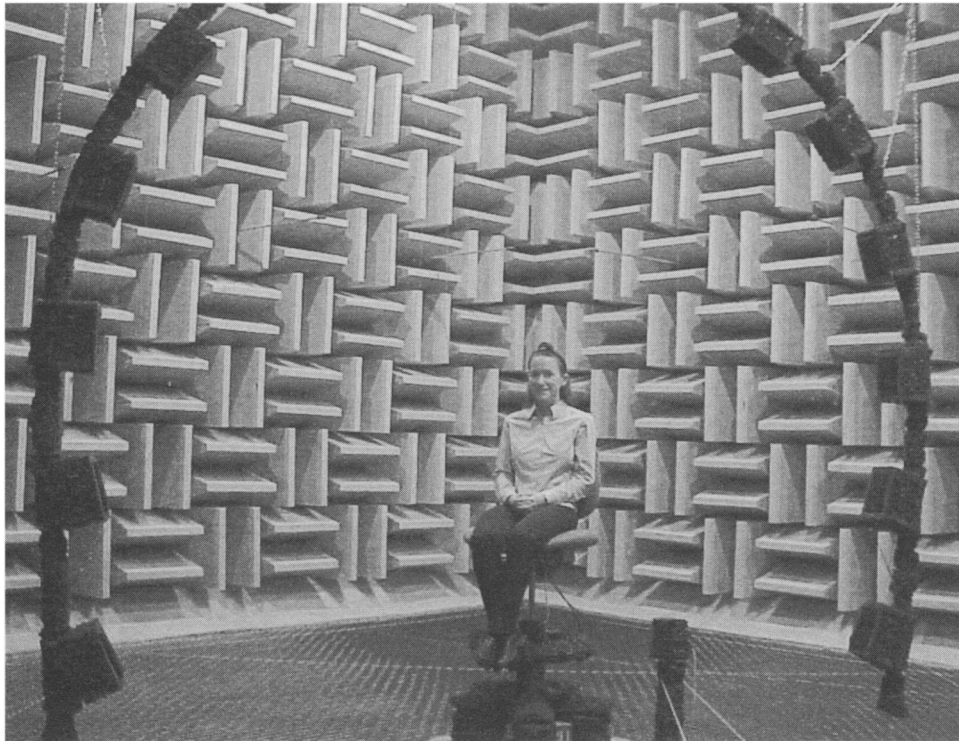
This section shows how a novel HRTF visualization strategy highlights important structures in HRTF data and suggests a way to compute interpolated HRTFs required for moving sound source synthesis.

Spatial Frequency Response Surfaces (SFRSs)

Many previous studies have attempted to visualize HRTF data by examining how certain macroscopic properties of HRTF sets, such as peaks and notches in particular locations of the magnitude frequency responses, associate or systematically vary with the perception of azimuth and elevation (Blauert 1983; Kendall 1995; Kistler and Wightman 1992; Shaw 1974). Typically, one attempts to see how these features change when the azimuth changes while elevation is held constant, or vice versa. For example, Figure 3 shows left- and right-ear HRTFs for several locations along the horizontal plane for which elevation is zero.

Visualization using this technique highlights certain HRTF features, as one can see several patterns in how the peaks and notches slowly change with varying azimuth. For example, ipsilateral HRTFs in Figure 3 seem to be smoother than contralateral HRTFs, especially at high frequencies. Diffraction effects can also be seen on the contralateral side of

Figure 2. (continued)
(e) Subject situated in
middle of speaker array.
Ruler units are in inches.



the head. Specifically, the head can have an amplifying effect on sounds originating from certain locations, even when these locations are completely blocked or “shadowed” by the head. Described as “bright spots” by Shaw (1974) because they appear as local maxima, these diffraction effects can be seen by noting the relatively large peaks at low frequencies at azimuths $+90^\circ$ for the left and -90° for the right HRTFs, respectively, in Figure 3.

Although this method for visualizing HRTF data in the frequency domain is intuitive, it is cumbersome when used to compare different sets of HRTFs, as only a single “slice” of HRTFs sharing the same azimuth or elevation can be compared at a time. Alternatively, there are other ways of visualizing different cross-sections of HRTF data, including additional frequency-domain methods (Carlile and Pralong 1994), cluster analysis (Martens 2000), and time-domain comparisons (Duda and Martens 1998).

In this article, we focus on spatial representations of HRTFs which plot the magnitude response

of all HRTFs in a data set for a fixed frequency as a function of azimuth and elevation. Specifically, one color plot is constructed for every frequency bin in the HRTF left or right magnitude response, where magnitude is plotted as a color height against azimuth and elevation. Because the spatial sampling pattern used during HRTF measurement is irregular, triangulation and linear interpolation is used to construct a surface that approximates a continuous surface. These graphs are called Spatial Frequency Response Surfaces (SFRSs), and they indicate how much energy the right and left ears receive as a function of spatial location. Figure 4 shows the construction of SFRSs from HRTFs.

SFRSs provide an immediate, visual, and very compact presentation of HRTF data. The most striking quality of SFRSs is the location and apparent motion of the “hot spots,” or well-defined local peaks, seen in SFRSs as a function of frequency. Figure 5 shows SFRSs constructed for several frequencies, and Table 1 summarizes the main structures found in each of these SFRSs. In addition, one

Table 1. Summary of cross-subject similarity in SFRS's from 1-13 kHz

<i>Frequencies</i>	<i>Figure</i>	<i>Description of corresponding SFRS's</i>
1-600 Hz	5a,b	Roughly equal power is received from all directions. No salient features present.
.6–1 kHz	5c,d	Head shadowing can be seen, as the ipsilateral ear receives more energy than the contralateral ear. Diffraction effects due to the head can be seen on the contralateral side of the head, near ± 100 degrees.
1–2 kHz	5e,f	Head shadowing becomes more prominent; diffraction effects are clearly seen on the contralateral side of the head. Two to three distinct peaks at ipsilateral azimuth 100, elevations $+30$ and -30 are starting to form. Attenuating effects of the torso can be seen on the contralateral side, lower elevations.
2–2.5 kHz	5g,h	Three peaks in the surface can be seen at ipsilateral azimuth 70–80, elevations -30 , 10 , and 50 . Diffraction effects can still be seen on the contralateral side near contralateral azimuth 100. Torso effects can be seen.
2.5–4 kHz	5i,j	The three peaks on the ipsilateral side have moved closer to the median plane and slightly higher in elevation. A fourth peak is starting to form beneath the other three, at ipsilateral azimuth 40, elevation -50 . Diffraction effects are starting to lessen, as the contralateral peak at contralateral azimuth 100 is beginning to fade. There are some nulls in the ipsilateral side, and torso effects can still be seen.
4–5 kHz	5k,l	The three peaks on the ipsilateral side have become one, large peak centered near ipsilateral azimuth 50, elevation 0. Diffraction effects are nearly gone, but torso effects can still be seen in lower elevations.
5–6 kHz	5m,n	The large ipsilateral “hotspot” has moved farther away from the median plane, and upwards in elevation. The spot is now at ipsilateral azimuth 75, elevation 20. Torso effects can still be seen.
6–8 kHz	5o,p	The single 5–6 kHz peak has become two smaller peaks at ipsilateral azimuth 75, elevations -40 , $+40$. Torso effects can still be seen.
8–10kHz	5q,r	The two ipsilateral “hotspots” are still present, but the lower elevation peak is more prominent than the higher elevation peak. A third hotspot is beginning to form on the median plane at azimuth 0, elevation -30 . Torso effects can still be seen.
10–13kHz	5s,t	Four hotspots are now apparent, one on the median plane at azimuth 0 elevation -20 , and the other three at ipsilateral azimuth 100, elevations -40 , 0 , $+40$. Torso effects can still be seen.

can readily see general, cross-subject trends in these plots which are more difficult to identify by analyzing raw magnitude frequency responses alone. Specifically, initial observations reveal that SFRSs are relatively smooth surfaces that change slowly as a function of frequency, and that different subjects' SFRSs are very similar to each other for a specific frequency. Figure 6 shows a comparison of three subjects' SFRSs for 2.4 kHz and 9.7 kHz.

Interpolated HRTFs

Because the goal of synthesizing virtual auditory space is to be able to place a sound at an arbitrary lo-

cation in virtual space, computing interpolated HRTFs is of central importance for synthesizing both stationary and moving sounds. We define an interpolated HRTF to be an HRTF corresponding to an arbitrary spatial location which is computed from a known, finite set of HRTFs. A widely used time-domain HRTF interpolation method constructs a time-domain, interpolated impulse response from a weighted average of other time-domain HRTFs corresponding to locations near the desired spatial location. A similar interpolation algorithm can be performed in the frequency domain, where an interpolated HRTF magnitude response is constructed from a weighted average of other HRTF magnitude responses corresponding to locations near the desired

Figure 3. Magnitude frequency response visualization of several HRTFs in the horizontal plane (0° elevation, θ° azimuth). Although

some patterns are evident in the raw HRTF magnitude responses as azimuth changes, cross-comparison of HRTF data is difficult.

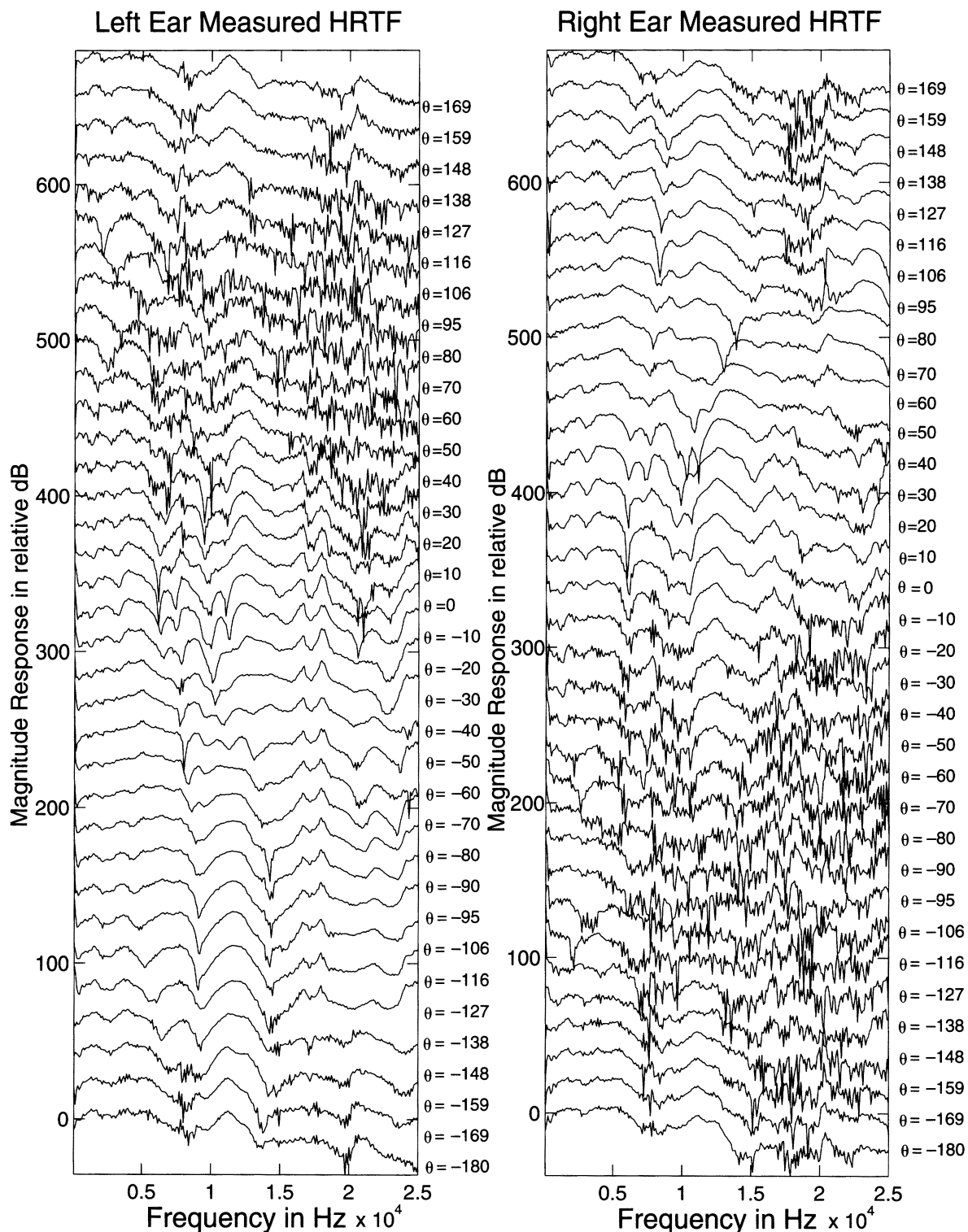
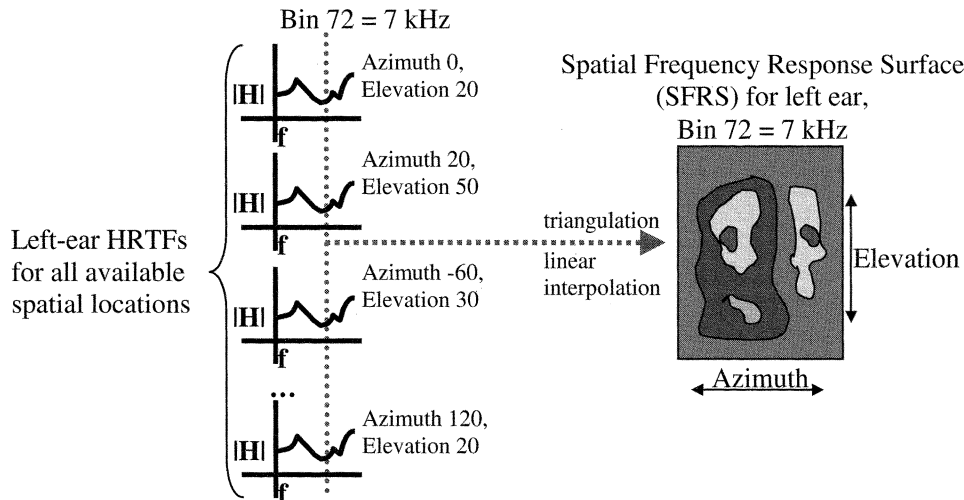


Figure 4. Construction of SFRSs from existing HRTF data. SFRSs are constructed from HRTF data for a particular frequency. Triangulation and linear interpolation of existing data are

used to create a continuous surface. Intuitively, this SFRS indicates how much energy the left ear receives as a function of a sound's spatial location.



spatial location. A good summary and comparison of some of these HRTF interpolation methods can be found in Hartung et al. (1999).

Our method for computing interpolated HRTFs is derived from observations made from SFRSs. Examination of SFRSs in Figures 5 and 6 suggests that for a fixed frequency, HRTF data contain similar, cross-subject features which are continuous and change relatively slowly as a function of spatial location. Because the construction of SFRSs relies on linear interpolation and the resulting SFRSs seem to preserve these features, it is reasonable to assume that SFRSs suggest an alternative HRTF interpolation method that exploits the interpolation used to create the SFRSs themselves.

The present algorithm constructs an interpolated HRTF magnitude response for a desired spatial location one frequency at a time, according to a weighted average of values taken directly from each SFRS. Specifically, the algorithm first performs a triangulation of the azimuth-elevation coordinate system to create a grid for the available, irregularly spaced HRTF data. The vertices of the triangulation are the locations at which HRTFs are known. In order to minimize the effect of the irregularity in spatial sampling, interpolated locations are taken only from where the triangulation is most uniform. For each SFRS, a plane is constructed using the three magnitude response values associated with the three vertices of the triangle enclosing the desired spatial location. The interpo-

lated value for that surface is taken as the value of the plane evaluated at the desired spatial location. This process is repeated for each SFRS, and each interpolated value is placed into the appropriate frequency bin of the interpolated magnitude response. Figure 7 describes the interpolation algorithm and shows how it relates to the interpolation used to create SFRSs.

This interpolation algorithm is an immediate, quantitative application of SFRSs. In addition, this interpolation algorithm allows different frequency components of different HRTFs to contribute to the interpolated HRTF to varying degrees, unlike simpler spatial methods (Middlebrooks 1992). No internal parameters, such as model order, need be predetermined, as is the case with pole-zero interpolation (Blommer and Wakefield 1995). The present interpolation algorithm has been shown to produce perceptually convincing HRTFs corresponding to arbitrary spatial positions, and further details on the algorithm's performance can be found in Cheng and Wakefield (1999a). All of the interpolated HRTFs generated for this article and accompanying sound examples were generated by our algorithm described herein.

Composing in Space with Moving Sound Sources

There are several well-known techniques that can be used to spatialize moving sound sources. Simple

Figure 5. (a)–(l) Subject Corey Cheng's SFRSs for several frequencies.

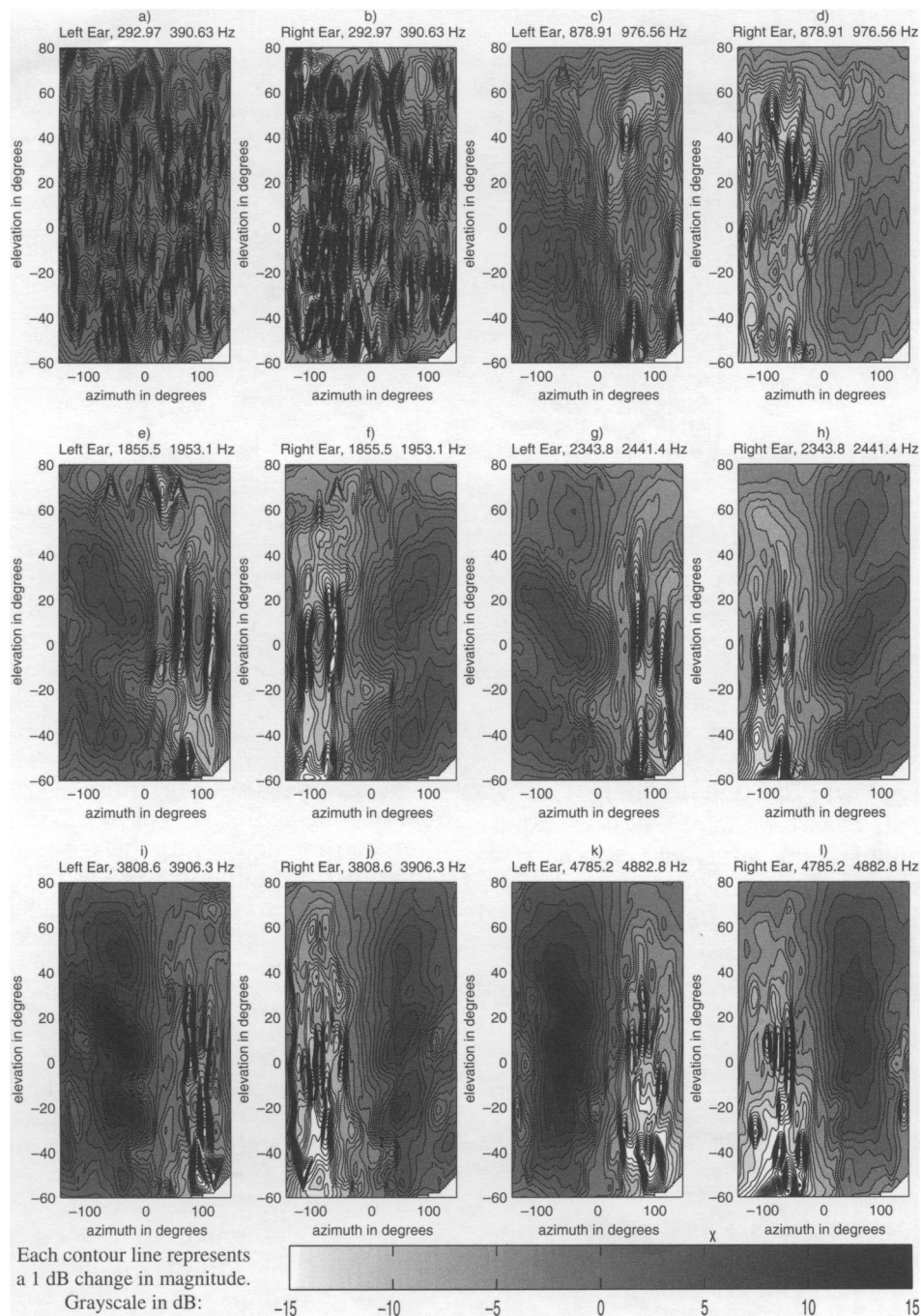
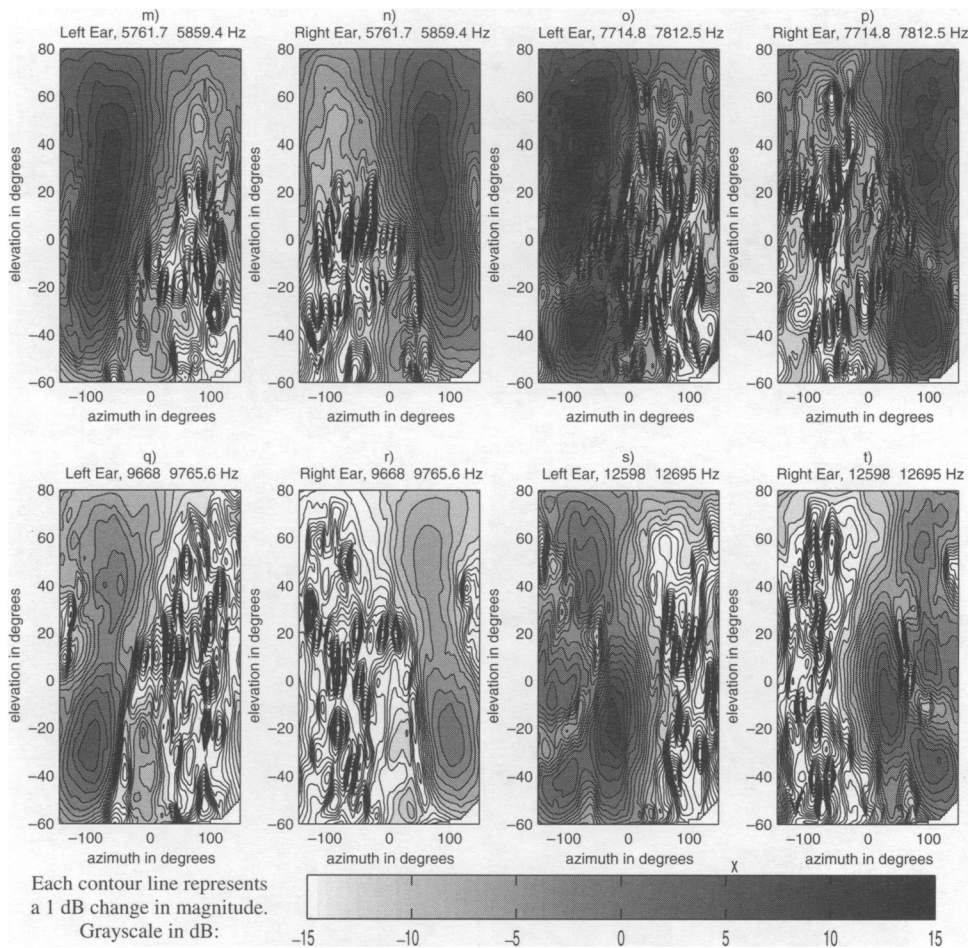


Figure 5. (continued)
 (m)–(t) Subject Corey
 Cheng's SFRs for several
 frequencies. See Table 1
 for details.



panning methods are still very effective for many types of left-right motions, and dynamic reverberation manipulation produces effective illusions of varying depth. More complex, pitch-based processing, such as Doppler shifting, can also produce compelling examples of moving sounds. Other methods for producing moving sound sources, such as ambisonics (Malham and Myatt 1995) and interaural cross-talk cancellation (Cooper and Bauck 1989), rely on systems designed for loudspeaker reproduction of spatial sound. An excellent production-quality software package that incorporates these and other spatialization techniques is IRCAM's Spatialisateur ("Spat") program.

However, our method for synthesizing moving

sound sources relies only on the use of interpolated HRTFs and ITDs as described in previous sections of this article, and thus the present results are intended for headphone listening only. In this sense, our current research attempts to discover how well we can spatialize sounds using these two parameters alone, and thus to isolate the strengths and weaknesses of HRTF-based moving sound synthesis. This section documents a MATLAB-based spatialization tool and graphical user interface (GUI) designed to synthesize moving sound sources with interpolated HRTFs and ITDs. Then, we outline the strengths and weaknesses of the spatialization tool in creating convincing spatialized sounds having different spatial trajectories. Finally, we discuss compositional techniques designed to take advan-

**Figure 6. Cross-comparison
of SFRs for three subjects
at 2.4 kHz and 9.7 kHz.**

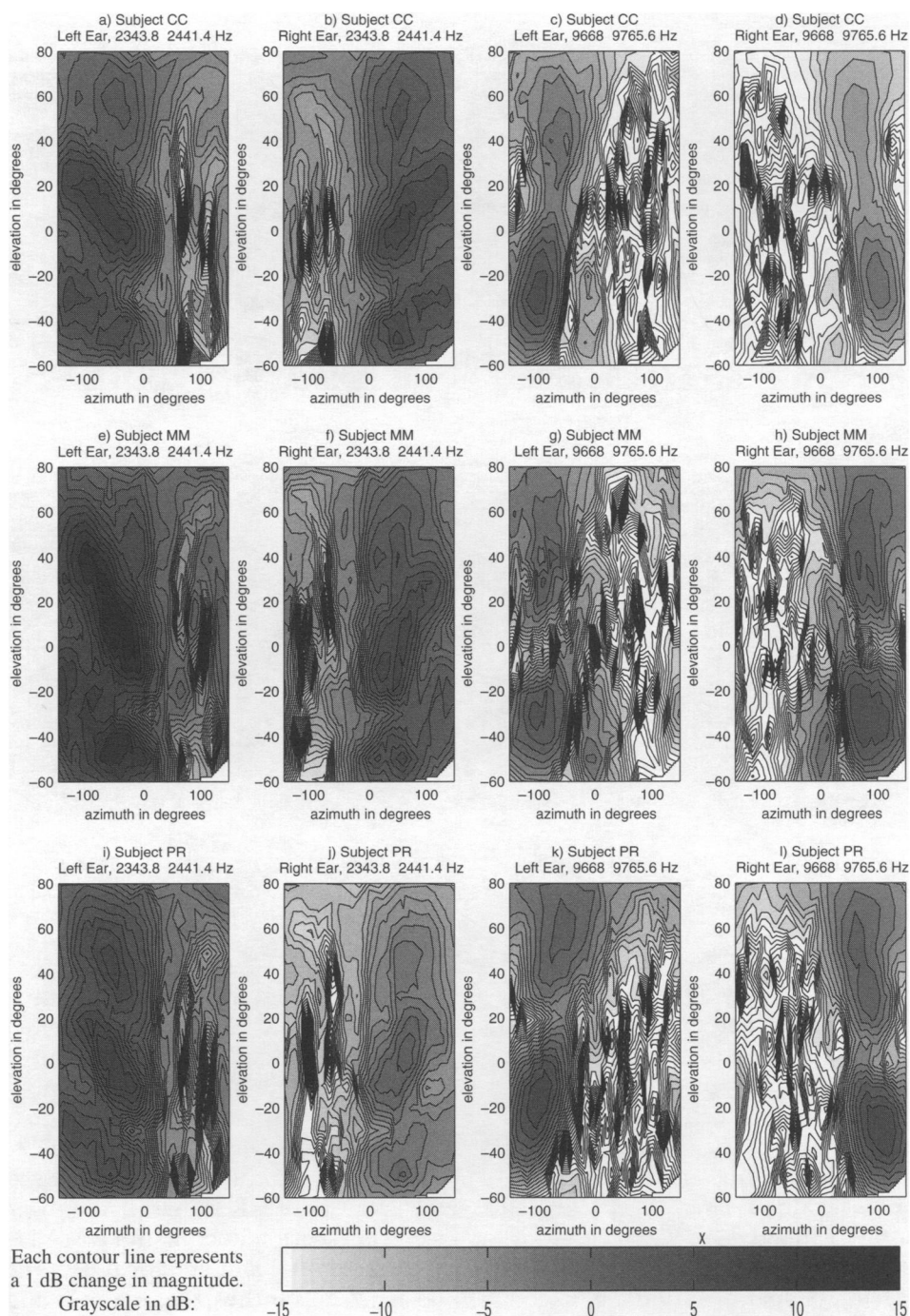
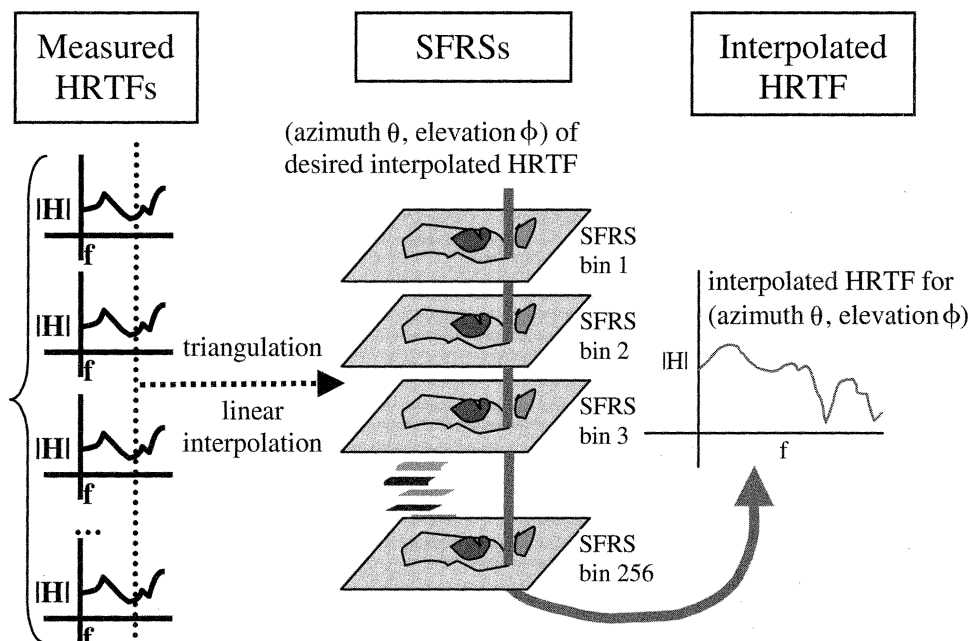


Figure 7. Computing interpolated HRTFs from SFRSs. This HRTF interpolation algorithm exploits the interpolation used to create SFRSs.



tage of the strengths of HRTF-based moving sound source synthesis.

Implementation of a MATLAB-Based Spatialization GUI

Interpolated HRTFs can be used not only to place a sound at an arbitrary point in space, but they can also be used to produce a moving sound following an arbitrary trajectory through space. In order to synthesize moving sounds for headphone playback, interpolated HRTFs and interpolated ITDs corresponding to slowly varying spatial locations along the desired spatial trajectory are dynamically updated during the processing of a monaural sound source, as depicted in Figure 8.

To make the process of experimenting with interpolated HRTFs and spatial trajectories more intuitive and user-friendly, we have developed a simple MATLAB-based user interface which implements a non-real-time version of the spatialization system shown in Figure 8. All of the accompanying HRTF-based sound examples were produced using this interface. Following the paradigm of volume and panning envelope editors present in many mul-

titrack mixing programs, this interface allows the user to manipulate a sound's azimuth and elevation trajectories independently with piecewise linear envelopes that can be saved and applied to other sound files of different lengths. Sound designers can use the interface to process sounds with different sets of interpolated HRTFs to customize spatialized sounds to a particular individual. The interface also shows the source waveform to be spatialized so that sound designers can synchronize azimuth and elevation trajectories with important, visually identifiable cues in the source sound, such as sharp transients, clicks, or onsets of sounds. Finally, the interface provides rudimentary playback capabilities for both source and target (spatialized) sounds, so sound designers can evaluate input sounds, output sounds, and spatial trajectories simultaneously. Figure 9 shows a snapshot of this interface.

Although the implementation of the system depicted in Figure 8 is relatively straightforward, there are some details about its use of interpolation and time-varying filtering which are important to discuss. First, note that it is undesirable for either the delay lines or the filters to change so abruptly that their combined processing will produce clicks,

Figure 8. A simple headphone-based spatialization system which relies on ITDs and HRTFs only. Changing ITDs and HRTFs during processing can produce the illusion of moving sound sources.

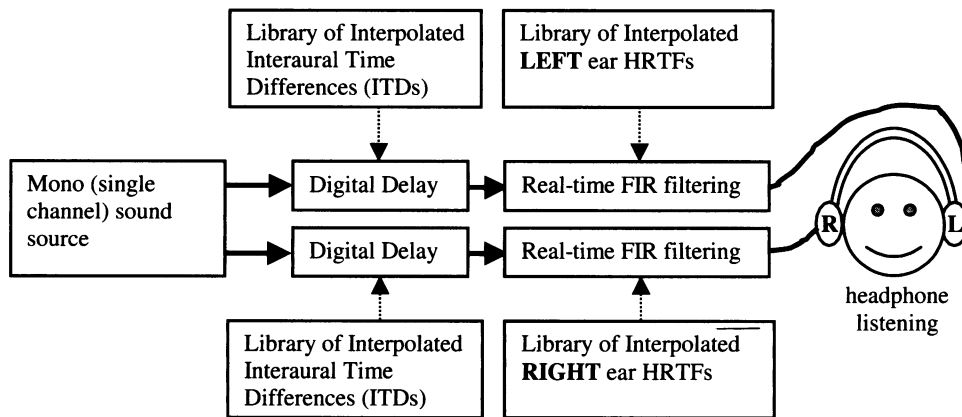


Figure 8.

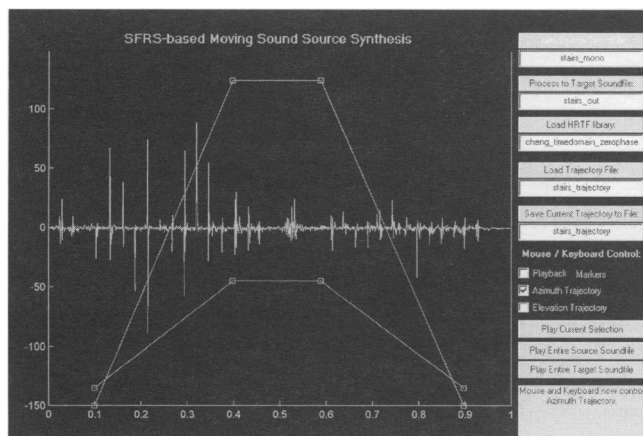


Figure 9.

pops, or clipping in the output signal. The delay associated with the left and right channels in Figure 8 corresponds to the simulated ITD between the two ears and may change abruptly for two reasons. First, the intended spatial trajectory may change too quickly, causing the ITD to change too quickly. Alternatively, an interpolated ITD computed for a specific spatial location may correspond to a non-integral sample delay, which is not realizable using a conventional, integral-tapped delay line. In this case, the non-integral delay must be rounded to the nearest integer, and abrupt changes in delay times for closely neighboring spatial locations could result from this type of ITD quantization error. Finally, abrupt changes in the left/right HRTF filters themselves could produce a disruption in the output.

Figure 9. MATLAB-based GUI constructed to facilitate the design and processing of spatial trajectories. This GUI shows the trajectory used to produce Sound example 7, a demonstration of footsteps climbing up

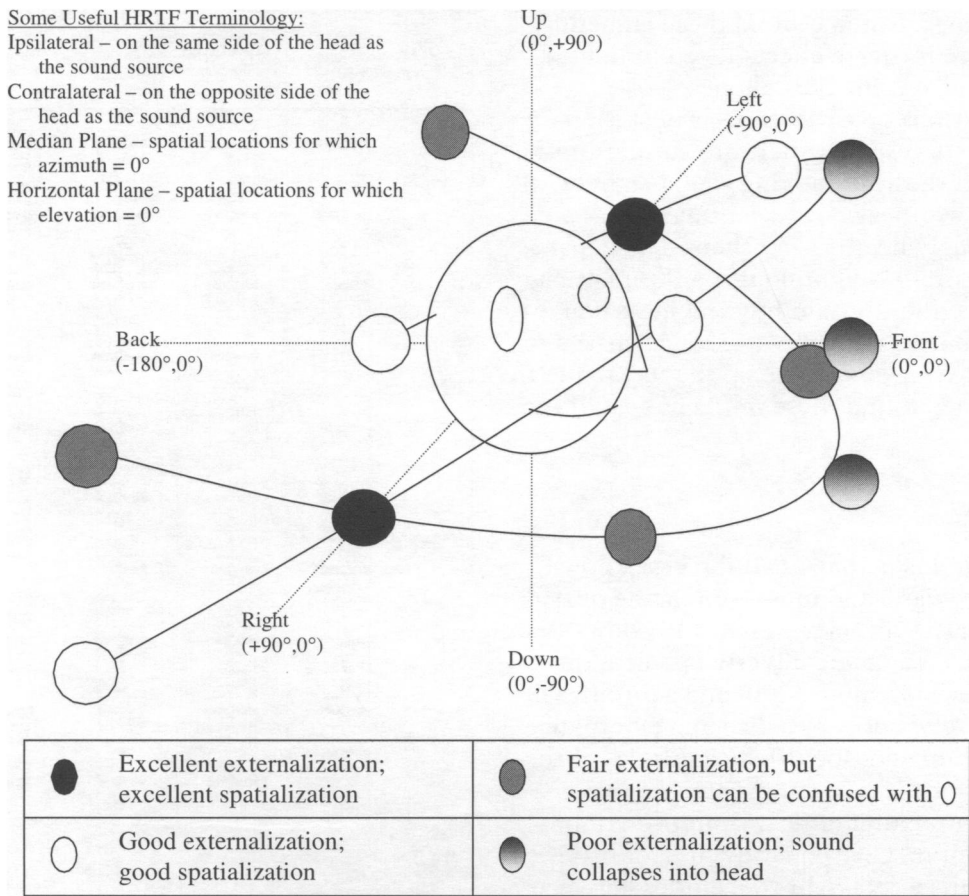
and down a flight of stairs. Note how both elevation and azimuth increase to accentuate the desired change in elevation by exploiting other well-spatialized areas (see Figures 10 and 13).

To alleviate some of these problems, we have incorporated the following informal “rules of thumb” to our sound processing. To ensure that sound sources do not move too quickly, we have limited the speed of directional change to roughly 30–40 degrees per second in both the azimuth and elevation directions. Although the maximum rate of change possible without producing disruption in output seems to be highly dependent on the source signal, we have found 30–40 degrees per second to be a safe estimate for many sounds. Second, to alleviate ITD quantization problems, we employ interpolated delay lines, a well-known structure used in the implementation of some physical models (Rocchesso 2000). It is important to note that interpolated delay lines attenuate the high-frequency components of sounds, and therefore the system’s overall intended spatialization effects may be altered due to the interpolated delay lines’ spectral coloration. Nonetheless, informal listening seems to suggest that the use of interpolated delay lines does not appreciably change the spatialization of most sounds.

Strengths and Weaknesses of HRTF-Based Moving Sound Synthesis

Although the theory of using HRTFs to synthesize spatial audio is simple, there are many limitations

Figure 10. Informal comparison of spatialization and externalization properties for various areas in virtual space. Spatialization and externalization of stationary, HRTF-processed sounds is more convincing in some areas than others.



and problems that occur in practice. While some of these restrictions place limits on the “spatial tessitura” available to composers, other restrictions can be viewed as interesting compositional problems that can be exploited to produce musical results.

Many limitations of moving sound source synthesis are closely related to some well-known limitations of stationary sound source synthesis, in which the synthesized sounds do not move through space. For example, the simple HRTF-based spatialization algorithm shown in Figure 8 may not produce the intended spatialization effects, even for broadband stationary sounds: listeners often report that there is a lack of “presence” in spatially-synthesized sounds, and that sounds spatialized near the median plane (0° azimuth) sound as though they are inside the head (Griesinger 1999). Signals processed to sound as though they originate

from in front of or above a listener actually sound like they originate from in back of or below the listener (the so-called “front-back” and “up-down” confusions) (Wightman and Kistler 1989b). Synthesis of sounds with non-zero elevations is difficult. Also, because every individual has a unique set of HRTFs, a subject listening to a spatialized sound generated from a “generalized” set of HRTFs may not perceive the sound in the intended spatial location (Pralong and Carlile 1996; Wenzel et al 1993).

Figure 10 summarizes some informal observations that show how stationary sounds are more convincingly spatialized or externalized at some spatial locations rather than others. Composers might choose to limit themselves to these spatial areas if they wish to render spatial ideas that can be easily heard by a wide range of listeners over wide variety of headphones. Alternatively, compos-

Figure 11. Intended orientations of the listener and spatial locations of a tennis ball in Sound examples 1–5, sounds from a virtual tennis game.

ers might attempt to step beyond these guidelines to attempt more intricate effects fitted to individual listeners and specific headphones.

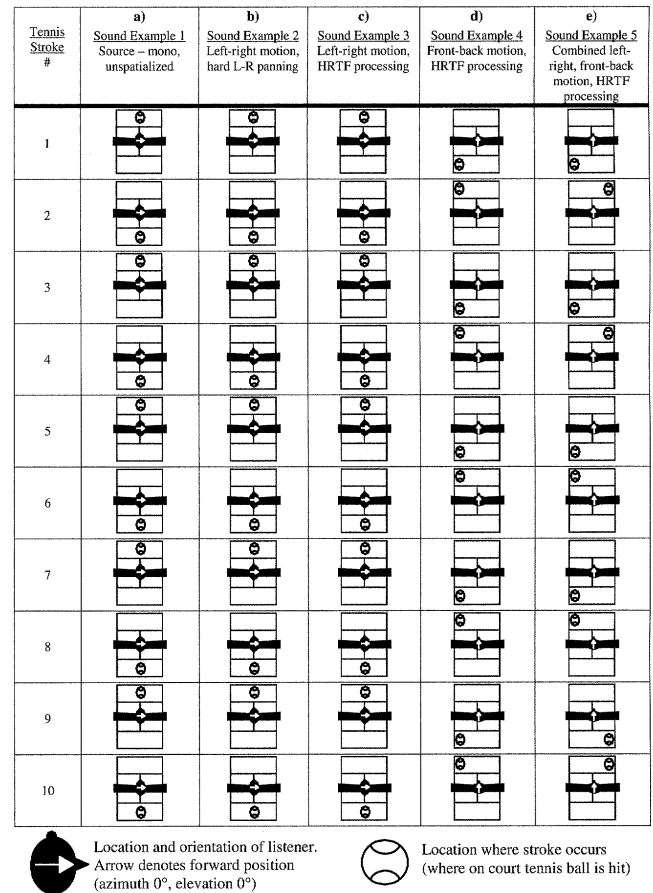
These limitations on the synthesis of stationary sound sources suggest that there are similar limitations on the synthesis of moving sound sources, and that some synthesized spatial trajectories are more convincingly heard than others. Thus, just as a skilled orchestrator might double a flute with a piccolo to strengthen certain musical ideas contained in the top melodic line, a skilled “spatial orchestrator” might need to develop the technique to work within and around these inherent spatial limitations.

Left-Right Motion

Left-right motion is perhaps still the best known, most straightforward, and most convincing of spatial trajectories to synthesize. Figure 10 shows that spatial locations which are directly opposite the left and right ears (azimuth -90° and azimuth $+90^\circ$, respectively) both spatialize and externalize extremely well. In addition, the magnitudes of the ITDs and IIDs for these locations are at a maximum, so that interaural difference information is also at a maximum. Consequently, it is of little surprise that direct, left-right trajectories between these two locations are also synthesized extremely well.

Sound examples 1–5 contain different spatial renderings of the same monaural sound file, a virtual tennis game between two players on the opposite sides of the tennis net. Figure 11 shows the intended spatial locations of the players, as well as the intended spatial location and orientation of the listener for each of these sound examples.

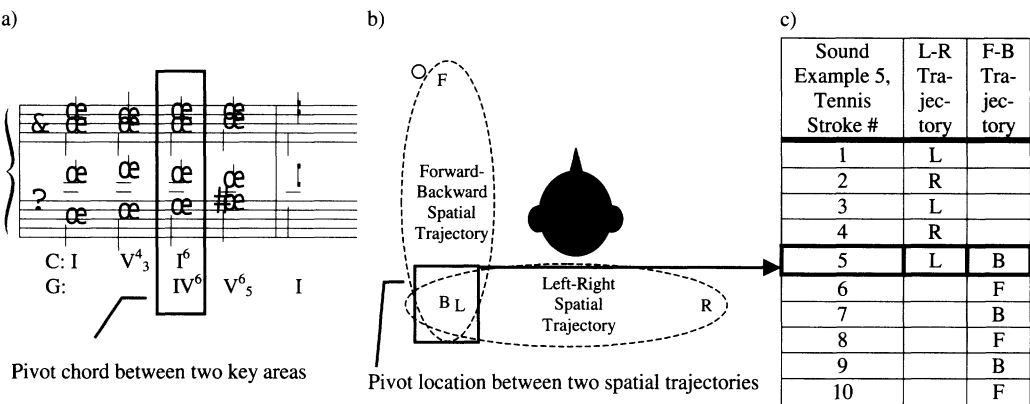
Sound example 1 presents the monaural source for the tennis examples, and Sound example 2 illustrates left-right motion of the tennis ball using “hard” panning. Note that the panning sounds somewhat imbalanced, or “uncomfortable,” as the sound shifts from left to right. Sound example 3 achieves left-right motion of the tennis ball using HRTF-based moving sound source synthesis. There is less of an uncomfortable feeling when listening to Sound example 3, and the externalization of the



sound is much better. It is interesting to note that these two different methods for realizing left-right motion have varying amounts of perceived tension, and it might be interesting to intermix these two techniques in the same composition for musical purposes. This possibility is explored further in the discussion of compositional technique.

The fact that externalization is best accomplished at locations directly opposite the ears also has musical implications. HRTF-based moving sound source synthesis is not designed to render changes in the sound source’s distance from the listener, since most HRTFs are specified by azimuth and elevation angles only. However, if it is important to maintain good spatialization and externalization for a moving sound source simultaneously, then the spatial palette becomes fairly limited. Composers may instead choose to achieve an exag-

Figure 12. Analogy of pivot locations to pivot chords. The perceptual ambiguity of some front/back virtual spatial locations can be exploited as pivot locations to shift a listener's attention from one spatial trajectory to another. Depending on the musical context, this technique can change the listener's perceived orientation in virtual space. (a) Example of a pivot chord. (b) Example of a pivot location. (c) Example of pivot location in the context of the intended spatialization for Sound example 5, a virtual tennis game. See Figure 11e for details.



generated change in externalization by intentionally and quickly moving a sound source from a location with good externalization (directly opposite the ears at $\pm 90^\circ$ azimuth) to a location with poor localization (directly in front of the listener at 0° azimuth).

Front-Back Motion

One of the advantages HRTF-based spatialization algorithms have over simpler methods is the ability to render sources in front of and behind a listener. However, synthesizing these types of sounds is still very difficult, as front-back confusions occur often, especially for sounds generated from generic, non-individualized sets of HRTFs and played back over commonly available headphones.

In Sound example 4, the tennis ball travels repeatedly from left and in front of the listener to left and in back of the listener. Note that this effect is more subtle than left-right motion and might be more difficult to hear. For this reason, we have found it helpful to keep changes in the angular speed of the sound source fairly low when trying to achieve front-back motion. In addition, another helpful "rule of thumb" is to try to keep the trajectory's elevation constant and near 0° azimuth when trying to achieve front-back motion.

The delicacy of front-back motion has important musical consequences. Rendering a sound that passes either extremely slowly or extremely quickly from front to back is difficult using HRTF-based spatialization techniques alone. This is because sounds that travel too quickly from front to back can sound like they travel from back to front,

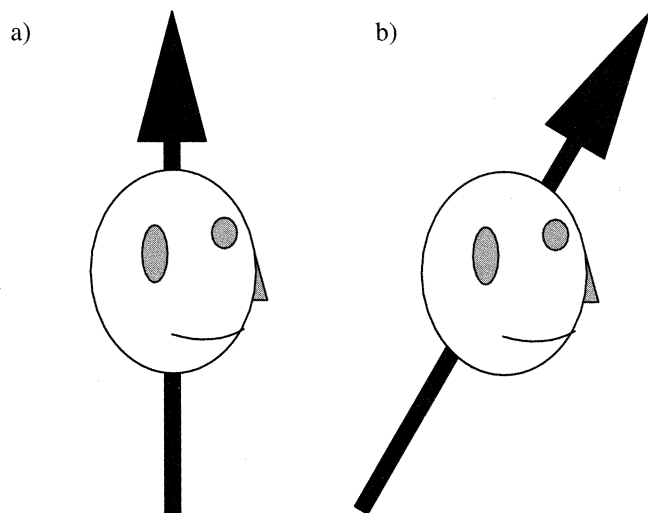
and sounds that travel too slowly from back to front to back can lose their overall front-back spatialization effect.

However, this limitation may be of musical interest to composers who wish to exploit the spatial ambiguity between front and back locations over an extended period of time. For example, a composer may wish to write in a "minimalist" style in which recurring patterns of sounds have spatial trajectories that change slowly over time. By gradually swapping front locations with rear locations in these repeating spatial trajectories, a composer may be able to create a piece in which a listener suddenly perceives a large change in the spatial character of the sound from front to back after an extended period of listening.

The rendering of an ambiguous spatial location can also be used as a transition point, or "pivot location," that connects two different spatial ideas. Specifically, a front or rear spatial location can serve as a common spatial location shared by two distinct spatial trajectories. In this sense, a front or rear spatial location can function much like a "pivot chord" does during the modulation between two different tonal key areas: whereas a pivot chord is a chord common to both the source and destination key areas, a front or rear location may serve as a "pivot location"—which is a location common between two spatial trajectories. Figure 12 shows the analogy between "pivot chords" and "pivot locations," and illustrates the idea of linking two different spatial trajectories.

Sound example 5, another spatialization of the tennis example, attempts to mix two distinct spa-

Figure 13. (a) Intended trajectory for Sound example 6, footsteps walking up and down stairs. (b) Implemented trajectory for Sound example 6. The implemented trajectory passes through locations at lower elevations behind the listener, as these areas tend to spatialize well according to



tial ideas: left-right motion and front-back motion. The intent here is to “trick” the listener into initially hearing one spatial trajectory, and then to shift the listener’s perceived spatial orientation and attention to another spatial trajectory through the use a “pivot location” common to both trajectories. In Sound example 5, the listener hears left-right motion for the first five strokes of the tennis ball, and is most likely convinced that the two tennis players are on the listener’s left and right. However, strokes 5–8 all occur on the left side of the listener, such that there is a subtle front-back motion of the tennis ball for these strokes. Therefore, the listener realizes that, in order for the tennis players to be on different sides of the net, one player must be in front of the listener, and the other player must be in back of the listener.

Thus, for the first five strokes, one of the players is in back of and to the left of the listener, and the other player is in front of and to the right of the listener. Because the front-back spatialization of the first five strokes is more subtle than the left-right spatialization of the first five hits, most listeners will not hear the front-back trajectory of the tennis ball and will instead hear only the tennis ball’s more dominant left-right trajectory. It is only on the fifth stroke, which occurs in a rear-left “pivot location,” that the listener may realize that the first five strokes might have contained a front-back component to spatialization all along.

Figure 10. The implemented trajectory for Sound example 6 can be seen in the screenshot contained in Figure 9.

Up-Down Motion

Using interpolated HRTFs also enables a composer to move sounds from positions above the listener (positive elevations) to positions below the listener (negative elevations). However, this effect is more difficult to produce than left-right motion, as both the discernment of up-down and front-back trajectories depends heavily on the particular set of HRTFs used during processing. For example, it has been shown that the number of “front-back” and “up-down” confusions increases when listening to spatialized sounds produced from non-individualized HRTFs (Wenzel et al. 1993).

In order to orchestrate up-down trajectories so that they sound more convincing to a wider range of listeners, we have found it useful to modify up-down spatial trajectories having a single azimuth so that they traverse other spatial locations which are more easily spatialized. For example, Figure 13a is a sketch of a spatial trajectory intended to pass from lower elevations to higher elevations directly across from the left and right ears (azimuths -90° and $+90^\circ$, respectively). However, according to Figure 10, it might be useful to slightly skew the “straight,” up-down trajectory in Figure 13a to take advantage of better spatialized areas behind/below and in front of/above the listener. Thus, to realize the intended spatial trajectory in Figure 13a, we actually implement the spatial trajectory in Figure 13b, which traverses several different azimuths to exploit these spatial “sweet spots.”

Sound example 6 contains a monaural, unspatialized recording of footsteps climbing up and down a flight of stairs. Sound example 7 contains a spatialized version of the footsteps that makes use of the modified spatial trajectory sketched in Figure 13b to accentuate the intended up-down trajectory. Figure 9 shows the exact spatial trajectory programmed into the spatialization GUI used to produce this sound example. Note that as elevation increases from -40° to $+40^\circ$, the azimuth trajectory increases from -135° to -45° as well. In this manner, the elevation change is reinforced with an azimuth change that takes advantage of the good spatialization properties behind and in front of the

listener as depicted in Figure 10. Thus, although the overall sonic effect of Sound example 7 is a change in elevation, the actual trajectory used to accomplish this effect incorporates front-back motion as well.

Dependence of Spatial Trajectory Effectiveness on Sound Source

Although sound spatialization techniques ideally work with any sound, certain sounds are better suited to spatialization than others. For example, short, broadband bursts of sound a few hundred milliseconds long produced with granular synthesis techniques tend to spatialize well (Roads 2000). Listener expectations and familiarity with sound sources can also add to or detract from certain spatial trajectories.

The successful rendering of spatial trajectories may depend heavily on the spectral content of the sound source. For example, spatialization of stationary broadband sounds, such as white noise, is generally more convincing than that of narrowband sounds and sine waves. This observation could have musical implications for the binaural spatialization of electroacoustic music which relies heavily on speech. Because most of the energy of ordinary speech lies below 4 kHz, spatialization of speech might be difficult unless accompanied by more broadband components. Thus, “breathy” speech such as whispering, or speech in which fricative sounds such as /f/ and /s/ are exaggerated, may be good candidates for convincing spatialization.

Sound example 8 contains a short example of monaural, unspatialized male and female voices. Sound example 9 contains a spatialized version of the voices whose spatial trajectory moves from right and in front of the listener to right and in back of the listener. Because neither the male nor the female voice contains many breathy, fricative, or otherwise broadband sounds, note how the speech does not spatialize very well, as the spatial trajectory does not sound very clear.

On the other hand, the theory of directional bands states that some narrowband sounds are indeed localized very well and are associated with preferred, perceptual directions in space. For example, one study has shown that listeners tend to as-

sociate narrowband sounds at approximately 6 kHz and 8 kHz with higher and lower elevations, respectively (Middlebrooks 1992). Furthermore, as seen in Figure 5, the dominant “hot spots” or local maxima contained in the SFRSs near these frequencies occur in higher and lower elevations, respectively, and thus some “hot spots” might serve as perceptual markers for certain locations in space (Cheng and Wakefield 1999b). Thus, another possible musical implication is to see if certain narrowband noises can be emphasized to suggest different spatial locations in a piece of electroacoustic music. For example, heavy emphasis of 6 kHz narrowband noise during a certain section of a piece might suggest higher elevations, whereas heavy emphasis of 8 kHz narrowband noise during a certain section of a piece might suggest lower elevations.

Compositional Technique for Moving Sound Sources

How can these binaural spatialization techniques for moving sound sources be incorporated into an effective electroacoustic composition? On the one hand, a significant danger in including any specialized effects such as moving sound sources into an electroacoustic composition is that the effect becomes clichéd through overuse or facile presentation. Certain spatialization effects such as the “ping-pong” effect might become too recognizable and thus too distinct to be incorporated into a larger, more flexible musical vocabulary. On the other hand, there is an equal danger that more complex spatialization effects are too difficult to implement, require too many specialized playback setups, and produce results that are too subtle to sustain an extended, sonic discourse of musical ideas.

To examine how some of these techniques may be put to creative use, one of the authors (Corey Cheng) composed *Fishbowl*, a short piece of binaural electroacoustic tape music composed primarily with processed water sounds that incorporates some of the spatialization techniques described above. *Fishbowl* explores how the interplay between tension and release might be expressed in spatial terms, in a manner analogous to the tradi-

tional use of consonance and dissonance in harmony. The following discussion and associated sound examples show how some of the spatialization techniques discussed in the previous sections have been used to develop these and other ideas in *Fishbowl*.

In previous sections, we have synthesized moving sound sources solely from the use of interpolated HRTFs. However, requiring the synthesis of moving sound sources to rely only on interpolated HRTFs is a severe restriction in practice, as much more musical results can be obtained with the concurrent use of other well-known tools. For example, adding reverberation to a sound before spatialization can greatly enhance the perception of depth. Sound example 10 presents a reverberated, monaural source sound, and Sound example 11 presents a spatialized version of Sound example 10 which moves from left to right; this spatialized version is used in *Fishbowl*. Note how the reverberation enhances the already good externalization properties of locations directly opposite the left and right ears to produce a wider range of motion.

Moving sounds can also be contrasted simultaneously with stationary sounds to produce more complicated examples that introduce the notion of spatial tension and release. Sound example 12 is an excerpt from *Fishbowl* that contains a mix of two sounds: a sound that moves from left to right, and a short, static “blip” sound that occurs when the first sound reaches the center point of its trajectory (0° azimuth). Note how the static sound in the center provides a reference point against which to hear the other moving sound.

However, care must be taken not to present a listener with too many differently spatialized sound objects at once. For example, up-down motion is a fairly delicate effect, and if an intended change in elevation is to be heard clearly, the background musical texture should be fairly light to not interfere with the spatialized sound. Sound example 13, an excerpt from *Fishbowl*, contains the reverberated, spatialized sound of soda being poured into a glass. The intended spatial trajectory of this sound example is for the soda to move from lower to higher elevations on the listener’s right side, as is depicted in Figure 13a. To accentuate the intended down-up motion of this example, several “composi-

tional tricks” have been used. First, the intended down-up trajectory was modified to include other well spatialized areas, as sketched in Figure 13b. Second, the down-up trajectory was chosen to match the spectral content of the sound: as the glass is slowly filled with soda, the sound changes from having more low frequency content to having more high frequency content. Because informal observations suggest that low and high pitched sounds are commonly associated with lower and higher elevations, respectively, the matching of low and high spectral content to low and high spatial positions helps to create the illusion of down-up motion. Finally, when incorporating the soda sound into the overall context of *Fishbowl* near the end of the piece, care is taken to lighten the overall texture of the background.

As previously discussed, panning and HRTF-based spatialization are two different methods for realizing left-right motion. Panning is a method for spatialization that simply changes the relative levels of the same monaural sound presented simultaneously to both the left and right ears. Although effective in many applications, panning can create an unbalanced feeling, or the impression of a “vacuum” in one ear for listeners wearing headphones. HRTF-based spatialization does not suffer from these imbalances, and in general also provides for better externalization. Mixing panned and HRTF-based spatialization can therefore provide an interesting method for presenting tension and release in spatial terms. Sound example 14 is an excerpt from *Fishbowl* containing two sounds that “chase” each other from left to right: a first sound spatialized along a left-right trajectory with panning and a second sound spatialized along the same trajectory with HRTFs. Note how the slight “vacuum” feeling created in the left ear by the first, hard-panned sound creates an expectation for another sound to fill the “vacuum” created in the left ear. Accordingly, the second, HRTF-spatialized sound then enters on the left and quickly “chases” and eventually overtakes the first sound on the right side of the listener.

The idea of contrasting sounds spatialized with hard panning and HRTFs can be expanded even further. For example, cross-fading sounds spatialized with panning and HRTFs can create an exaggerated

Figure 14. The “slingshot” technique. Cross-fading between differently spatialized realizations of the same monaural sound can produce illusions of exag-

gerated acceleration around the head. The “slingshot” technique exploits different externalization properties of HRTF and panning methods for left-right spatialization.

Figure 15. Volumetric (3D) visualization of HRTFs. 3D objects generated from 2D data may reveal other important HRTF structures. Are SFRs “slices” of a larger object?

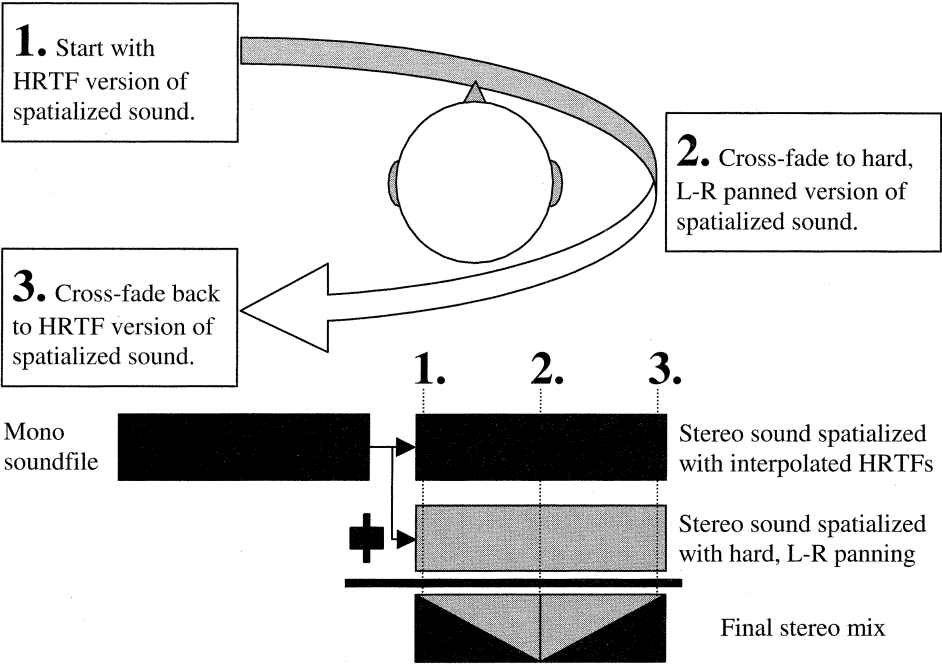


Figure 14.

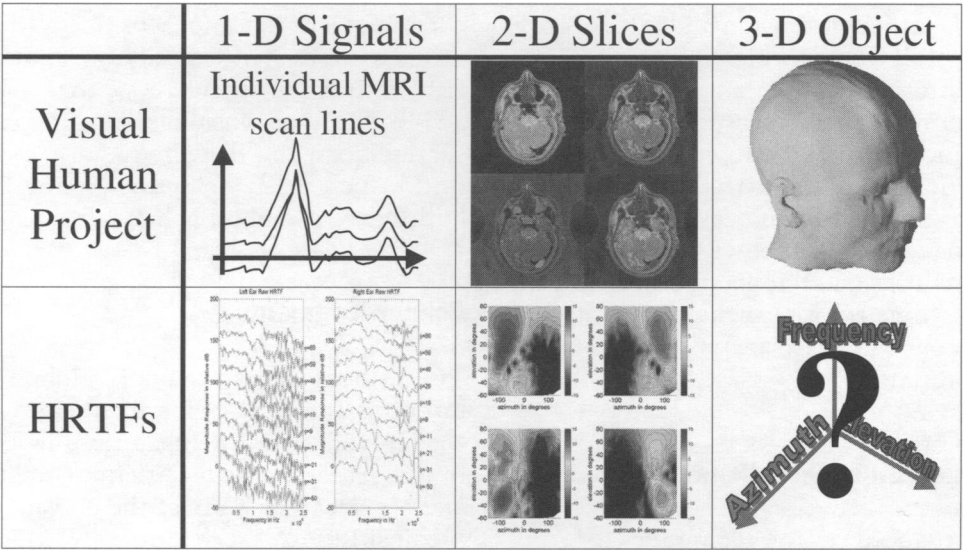


Figure 15.

impression of highly accelerated motion around the listener's head. This "slingshot" effect exploits the different externalization properties of each spatialization method. Because the panned sound sounds closer to the listener and the HRTF-spatialized sound sounds farther away from the listener, cross-fading between these two versions of the same source sound can produce the impression of increasing acceleration. Figure 14 shows how the same sound spatialized these two different ways can be cross-faded to produce a sound that originates from a distant spatial location to the left of a listener, approaches and accelerates quickly around the listener's right side, and then returns to a distant location on the left.

Sound example 15, another excerpt from *Fishbowl*, demonstrates the "slingshot" effect. Note how the spatial focus of the excerpt moves from the center to the near right, and then quickly to the far left, producing the illusion that the sounds are accelerating quickly from near the listener to far away. The sounds at the near right of the listener are produced with panning, and the sounds at the far left of the listener are produced with HRTF-based spatialization. Compare this to Sound example 16, an excerpt from *Fishbowl* that attempts to develop this same spatial trajectory over a longer period of time with different sounds.

Finally, Sound example 17 contains the entire piece *Fishbowl*. Although tension and release are developed in spatial terms throughout the piece, *Fishbowl* also attempts to integrate these spatial ideas with other non-spatial ideas into a larger musical context. For example, the manipulation of spatial trajectories combined with the extramusical ideas implied by some of the processed water sounds helps to paint the fishbowl's creatures with different personalities and temperaments. Also, the final "splash" of tone color near the end of the piece is an attempt to release spatial tension by using pitched material.

Conclusions and Future Directions

This article introduced several techniques for composing with moving sound sources generated from

interpolated HRTFs. Using SFRSs to visualize HRTF data suggests a new signal processing strategy for computing interpolated HRTFs, a technique that is useful for the synthesis of moving sound sources for binaural electroacoustic music. In the future, new visualization approaches may uncover more structure in HRTFs that could suggest other signal processing and compositional strategies for handling the synthesis of moving sound sources. Investigations into even more compact, three-dimensional, volumetric representations of HRTFs similar to those already used to visualize the human anatomy are already underway. Figure 15 shows the motivation behind these studies, and the interested reader can find preliminary results in Cheng and Wakefield (2000).

Admittedly, it is sometimes easy for the authors, as scientists trained in inquiry for inquiry's sake, to forget about music for music's sake. Science and technology may be art forms in themselves, but we must not forget that in the ideal case, electroacoustic music should not have to exist simply to express the intricate technology behind it. Indeed, with reference to the current subject of spatialization, others have cautioned that "[i]t is important to remember that space, and spatialization as a sound parameter, can and should be used compositionally in computer music" (Pope 1995). With this in mind, this article has presented several practical, concrete techniques for working with moving sound sources synthesized from interpolated HRTFs. It is our sincere hope that these techniques add to a growing vocabulary for the expression of spatial ideas, and that the science and engineering from which these techniques are motivated and built prove useful for achieving artistic goals.

Acknowledgments

The authors wish to thank Dr. John C. Middlebrooks at the Kresge Hearing Research Institute of the University of Michigan for providing data used in this research. We also thank Dr. Thomas Buell and Ms. Heather Kelly at the Naval Submarine Medical Research Laboratory (NSMRL) in Groton, Connecticut, for their work on developing the

HRTF measurement apparatus. We also thank Abhijit Kulkarni and Bill Rabinowitz at Bose Corporation for their help and guidance in the initial design of the HRTF measurement apparatus. Finally, we thank the editors of *Computer Music Journal* and the organizing committee of the Sound in Space 2000 Symposium at the University of California at Santa Barbara for encouraging us to share our work. Some of the material concerning SFRSs and interpolated HRTFs was previously discussed in Cheng and Wakefield (1999b) and Cheng and Wakefield (2001b). This work was supported by a grant from the Office of Naval Research in conjunction with NSMRL.

References

- Blauert, J. 1983. *Spatial Hearing*. Cambridge, Massachusetts: MIT Press.
- Blommer, A., and G. Wakefield. 1995. "A Comparison of Head Related Transfer Function Interpolation Methods." In *1995 IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*. Piscataway, New Jersey: IEEE Press, pp. 88–91.
- Blommer, M. A. 1996. Pole-Zero Modeling and Principal Component Analysis of Head-Related Transfer Functions. Ph.D. diss., University of Michigan, Department of Electrical Engineering and Computer Science, Systems Division.
- Carlile, S., and D. Pralong. 1994. "The Location-Dependent Nature of Perceptually Salient Features of the Human Head-Related Transfer Functions." *Journal of the Acoustical Society of America* 95(6):3445–3459.
- Chen, J., et al. 1992. "External Ear Transfer Function Modeling: A Beamforming Approach." *Journal of the Acoustical Society of America* 92(4):1933–1944.
- Cheng, C. I., and G. H. Wakefield. 1999a. "Spatial Frequency Response Surfaces: An Alternative Visualization Tool for Head-Related Transfer Functions (HRTF's)." In *Proceedings of the 1999 International Conference on Acoustics, Speech, and Signal Processing (ICASSP99)*. Piscataway, New Jersey: IEEE Press, pp. 961–964.
- Cheng, C. I., and G. H. Wakefield. 1999b. "Spatial Frequency Response Surfaces (SFRS's): An Alternative Visualization and Interpolation Technique for Head-Related Transfer Functions (HRTF's)." In *Proceedings of the 16th Audio Engineering Society (AES) International Conference on Spatial Sound Reproduction*. New York: Audio Engineering Society, pp. 147–159.
- Cheng, C. I., and G. H. Wakefield. 2000. "A Tool for Volumetric Visualization and Sonification of Head-Related Transfer Functions (HRTF's)." In *Proceedings of the International Conference on Auditory Display*. International Community for Auditory Display, pp. 135–140.
- Cheng, C. I., and G. H. Wakefield. 2001a. "Error Analysis of HRTF's Measured With Complementary (Golay) Codes." *Abstracts of the Journal of the Acoustical Society of America*. New York: Acoustical Society of America, p. 2419.
- Cheng, C. I., and G. H. Wakefield. 2001b. "Introduction to Head-Related Transfer Functions (HRTF's): Representations of HRTF's in Time, Frequency, and Space." *Journal of the Audio Engineering Society* 49(4):231–249.
- Cooper, D. H., and J. L. Bauck. 1989. "Prospects for Transaural Recording." *Journal of the Audio Engineering Society* 37(1/2):3–19.
- Duda, R. O., and W. M. Martens. 1998. "Range Dependence of the Response of a Spherical Head Model." *Journal of the Acoustical Society of America* 104(5):3048–3058.
- Griesinger, D. 1999. "Objective Measures of Spatioussness and Envelopment." In *Proceedings of the 16th Audio Engineering Society (AES) International Conference on Spatial Sound Reproduction*. New York: Audio Engineering Society, pp. 27–41.
- Hartung, K., et al. 1999. "Comparison of Different Methods for the Interpolation of Head-Related Transfer Functions." In *Proceedings of the 16th Audio Engineering Society (AES) International Conference on Spatial Sound Reproduction*. New York: Audio Engineering Society, pp. 319–329.
- Jenison, R. L. 1995. "A Spherical Basis Function Neural Network for Pole-Zero Modeling of Head-Related Transfer Functions." In *1995 IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*. Piscataway, New Jersey: IEEE Press, 92–95.
- Kendall, G. S. 1995. "A 3-D Sound Primer: Directional Hearing and Stereo Reproduction." *Computer Music Journal* 19(4):23–46.
- Kistler, D. J., and F. L. Wightman. 1992. "A Model of Head-Related Transfer Functions Based on Principal Components Analysis and Minimum-Phase Reconstruction." *Journal of the Acoustical Society of America* 91(3):1637–1647.
- Kulkarni, A., et al. 1995. "On the Minimum-Phase Approximation of Head-Related Transfer Functions." In

- 1995 *IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*. Piscataway, New Jersey: IEEE Press, pp. 84–87.
- Kulkarni, A., et al. 1999. "Sensitivity of Human Subjects to Head-Related Transfer-Function Phase Spectra." *Journal of the Acoustical Society of America* 105(5):2821–2840.
- Malham, D. G., and A. Myatt. 1995. "3-D Sound Spatialization Using Ambisonic Techniques." *Computer Music Journal* 19(4):58–70.
- Martens, W. 2000. "Cluster Analysis of the Head-Related Transfer Function." Available online at <http://www.svl.u-aizu.ac.jp/~wlm/cluster.html>.
- Middlebrooks, J. C. 1992. "Narrow-Band Sound Localization related to External Ear Acoustics." *Journal of the Acoustical Society of America* 92(5):2607–2624.
- Oppenheim, A. V., and R. W. Schaffer. 1989. *Discrete-Time Signal Processing*. Englewood Cliffs, New Jersey: Prentice Hall.
- Pope, S. T. 1995. "About This Issue." *Computer Music Journal* 19(4):1.
- Pralong, D., and S. Carlile. 1996. "The Role of Individualized Headphone Calibration for the Generation of High Fidelity Virtual Auditory Space." *Journal of the Acoustical Society of America* 100(6):3785–3793.
- Rayleigh, L. 1907. "On Our Perception of Sound Direction." *Philosophical Magazine* 13:214–323.
- Rayleigh, L. 1945. *The Theory of Sound*. New York: Dover Publications.
- Roads, C. 2000. Personal communication with the authors. March 2000.
- Rocchesso, D. 2000. "Fractionally Addressed Delay Lines." *IEEE Transactions on Speech and Audio Processing* 8(6):717–727.
- Shaw, E. A. G. 1974. "The External Ear." *Handbook of Sensory Physiology V/1: Auditory System, Anatomy Physiology(Ear)*. New York: Springer-Verlag.
- Shaw, E. A. G., and R. Teranishi. 1968. "Sound Pressure Generated in an External-Ear Replica and Real Human Ears by a Nearby Point Source." *Journal of the Acoustical Society of America* 44(1):240–249.
- Speyer, G. 1999. A Boundary Element Model for Predicting the Head-Related Transfer Function. M.S. thesis, Department of Electrical Engineering-Systems, Tel Aviv University, Israel.
- Wenzel, E. M., et al. 1993. "Localization using Nonindividualized Head-Related Transfer Functions." *Journal of the Acoustical Society of America* 94(1):111–123.
- Wightman, F. L., and D. J. Kistler. 1989a. "Headphone Simulation of Free-Field Listening. I: Stimulus Synthesis." *Journal of the Acoustical Society of America* 85(2):858–867.
- Wightman, F. L., and D. J. Kistler. 1989b. "Headphone Simulation of Free-Field Listening. II: Psychophysical Validation." *Journal of the Acoustical Society of America* 85(2):868–878.