The Composition of Auditory Space: Recent Developments in Headphone Music

Durand R. Begault

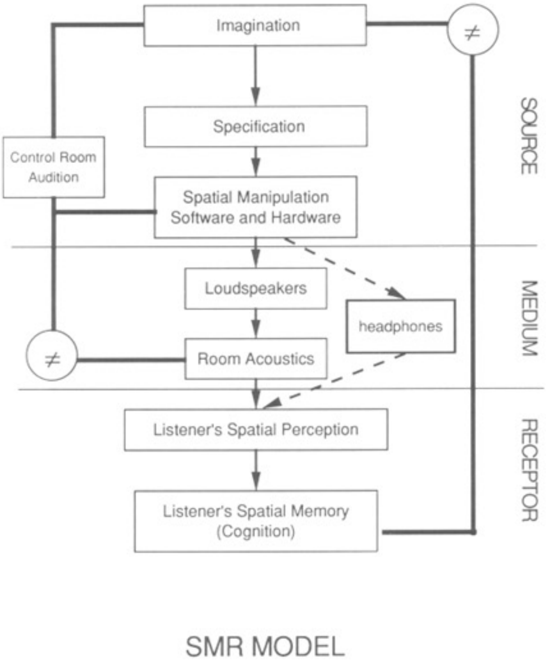
ABSTRACT

Research into the psychoacoustics of spatial hearing and into computer-based technologies has brought about an exciting potential for the development of 'spatial music': a compositional approach to the musical organization of sound that considers the position of sound sources and the character of the environmental context to be as musically important as melody, harmony or orchestration. Space as a musical parameter is overviewed, the potential areas and concurrent limitations of spatial music composition are described, and the likely causes of perceptual mismatch between the composer and the listener are reviewed. Headphone music as a solution to the mismatch problem is proposed, and a description of the spatial signal-processing technique developed by the author is given. The compositional considerations used in two computer music headphone compositions, *Revelations* by the author and *Begault Meadow* by Gordon Mumma, conclude the discussion.

There is no such thing as nonspatial hearing; all musical experience has an inherent spatial component, even if it is not noticed by the listener. The listener must occupy a location different from that of the sound source, and musicians must occupy locations different from each other if there is more than one performer. The compositional manipulation of the spatial aspect of music was as inevitable as the manipulation of pitch, timbre or duration, but little attention has been given to it heretofore.

Using psychoacoustically based digital signal-processing techniques, composers today are able to access cues to spatial hearing in the composition of headphone music. Simply put, the advantage of using headphones over loudspeakers or live performers is that composers can more easily and assuredly convey their musical-spatial intentions. Research in the development of binaural mixing consoles allows composers greater control of the perceived azimuth, or angle of incidence, of a sound source. Using the same techniques, synthesized reflected sound can be added to the original sound to convey the illusion of distance and of the environmental context of the sounds.

Fig. 1. SMR (Source-Medium-Receptor) Model. Each box represents a stage of non-linear transformation that occurs between the composer's imagined spatial gesture and what the listener actually hears. The *source* includes the composer's imagined spatial gesture, the means of specifying the gesture to spatial manipulation software and hardware, and the software and hardware itself. The *medium* refers to loudspeakers and room acoustics; note that, although feedback occurs for the composer at the source level by audition in the control room, the control room's acoustics are not equal (≠) to the acoustics of the medium. The *receptor* refers to the listener of spatial music; the listener's perceptual and cognitive mechanisms may further alter the composer's intended effect.



THE FIFTH ELEMENT

Music, particularly new music, often is analyzed in a restrictive way in terms of elements that are assumed to be separable. The assumption is based on basic physical descriptions of sound waves. The four so-called separable musical elements (with their corresponding psychological descriptions) are frequency (pitch), spectral content (timbre), intensity (loudness) and duration (perceived duration).

This analysis applies most readily to the description of the smallest element of music, a single note. From a psychophysical standpoint, however, these elements cannot be discussed separately, because variation in any one of them can affect several psychoacoustic parameters. For instance, intensity can affect pitch and timbre, spectral content can affect loudness, and duration can affect timbre and loudness.

This four-element description of music is compositionally incomplete, because it excludes large-scale musical considerations like the horizontal temporal organization of events or the vertical variation in event density. More importantly, it excludes a widely used but largely unrecognized

Durand R. Begault (researcher, composer), Aerospace Human Factors Research Division, NASA-Ames Research Center, Mail Stop 239-3 FLM, Moffett Field, CA 94035, U.S.A.

Manuscript solicited by Larry Polansky.

Received 30 September 1987.

Pergamon Press plc. Printed in Great Britain.
0024-094X/90 $3.00+0.00

physical parameter in music: the localization of the sound and its environmental context. The physical description of sound does not depend on the location of the listener or on the listener's perceptual mechanism in forming an idea of the location of events or of the environmental context. Yet these sensations are present in normal musical audition, and throughout history they have been manipulated by composers. Space can therefore be called the fifth element of a musical sound.

When the spatial element of sound is unchanging, spatial hearing is not regarded by the listener as an important compositional or expressive attribute of the music. By contrast, a musical composition that involves any sort of compositional control over the apparent spatial location of sound is termed *spatial music*. In spatial music, the spatial parameter is either dynamic (undergoing change) or static (and calling attention to itself through the use of an unusual distribution of performers or loudspeakers). In such music the spatial element is a compositional parameter subject to manipulation and/or organization beyond the typical spatial distribution of sound sources used in normal performance practice.

Before the discussion of spatial music, it is important to note two other ways that musical space is sometimes described. The first way is from the perspective of precompositional structure. This usually involves the description of the large-scale transformations of a group of musical elements in a composition; a simple example would be a two-dimensional graph of pitch against time. The second way that space is used in a musical context is by philosophers of music. Both Susanne Langer (in *Feeling and Form* [1]) and Victor Zuckerkandl (in *Sound and Symbol* [2]) have used the metaphor of space to describe the nontemporal nature of the musical experience. Both of these approaches have little to do with spatial music per se for the simple reason that they are not descriptions of spatial *hearing*. These are conceptual, rather than perceivable, uses of space, used as metaphorical tools for discussing philosophical relationships or compositional organization.

Spatial hearing is far more basic to the human experience than these sorts of spatial conceptualizations. The ability of human beings to localize sound is often cited in psychoacoustic literature as an important factor in early human survival. The musical awareness of spatiality is demonstrated in a long history of works where musical gestures were passed between two or more locations; an example is the *polychoral* tradition that flourished in Italy and Germany between 1515 and 1650. A different approach to spatial sound can be found beginning around the late eighteenth and early nineteenth centuries, when composers began to develop the resources of the large orchestra and to write program music, using instruments to create the illusion of spatial environments.

Stereo recording techniques developed in the popular music industry have allowed electronic-music composers to give movement to sound sources and to create virtual environments at the mixing board. The disposition of sounds in space is created long after the original recording and independently of it. The technique of mixing a multitrack recording to stereo is inseparable from creating what is essentially an auditory illusion. Most music available on commercial recordings has sounds that can be considered as placed into various spatial relationships, as a result of the amplitude scaling, filtering and processing of each individual track of the recording.

Digital audio technology has expanded the sophistication of amplitude variation, the principal means used at the analog mixing board to create spatial effects; moreover, it has allowed other parameters to be utilized as cues to the spatial hearing mechanism. One of the first attempts at utilizing several cues for localization simultaneously in a software synthesis environment was described in John Chowning's "Simulation of Moving Sound Sources" [3]. In this program, cues for the azimuthal location of a sound source were created by amplitude panning between the four speakers of a quadraphonic playback setup, and distance cues were created by controlling the ratio of reverberant to direct sound as well as the amplitude of the sound source. A composer could manipulate the spatial parameter of sound by drawing sound paths (or sound trajectories) with interactive software.

The current trend in signal-processing software is to apply the psychoacoustic knowledge of spatial hearing even more thoroughly within a software context; important work in this area has been done by Moore [4] and Kendall and Martens [5], among others. The REFL program described below [6] is one such program that attempts to create convincing auditory spatial illusions by utilizing psychoacoustic cues.

Composers have only begun to explore the potential for computer-based spatial manipulation in loudspeaker and headphone music. This potential includes the ability of a composer to give the listener the sense of sound movement, where the trajectory of the sound source is important [7]. However, there are both 'hard' and 'soft' limits to what a composer can do, that are discussed below.

PSYCHOACOUSTICS, COMPOSITIONAL SPECIFICATION AND THE PROBLEM OF THE MISMATCH

Composers are faced with the problem of accurately conveying their musical-spatial concepts to listeners. Consider the entire chain of communication involved for the transmission of spatially manipulated sound. This transmission of the composer's intention to the listener can be viewed in terms of a *source-medium-receptor* (SMR) model (Fig. 1). Each step of this model represents a particular nonlinear transformation of the original compositional intent. The problem is to avoid an undesired perceptual mismatch between the composer's intent and the listener's perception.

The *source* includes the composer's spatial conception for a sound and its conversion into specifications for a particular spatial-manipulation computer program that ultimately determines the waveforms supplied to the loudspeakers. The chain of events can be sequentially envisioned as follows: the composer's imagination (for example, for a single sound event, the spatial conception of a sound source and a listener within a particular environmental context; or, on a macro level, a pattern of sound-movement trajectories); the composer's means of specifying this spatial conception to the computer (either by manipulating the controls of real-time audio-processing devices such as mixers and reverberators or by specifying values to the parameters of a software program); the computer's interpretation of this specification into parameters for acoustic modification of the sound, and the hardware or software interpretation of the user's specification (for instance, the actual parameters used by a commercial reverberator when one selects a room called a 'large hall'); and, finally, the nonlinear signal ramifications (such as phase and harmonic distortion) inherent to the amplification system that supplies the loudspeakers.

The *medium* involves the effects of loudspeakers and room acoustics, which can greatly modify the spatially manipulated sound before it reaches a listener. Although a composer may carefully audition the work when composing in a studio control room, the fact that the room's acoustics and the loudspeakers differ in the context of concert presentation invariably manifests unexpected or uncontrolled perceptual results. This is due to the fact that different types of loudspeakers have particular frequency responses and will sound different depending on the characteristics of the room they are heard in. Even a variation of the number of people in the audience in a given performance space can result in different modifications of the sound once it leaves the loudspeaker. Also, in a concert presentation, each audience member will have a unique orientation to the position of the speakers, which can radically affect the intended spatial image. The effects of loudspeakers and room acoustics can be bypassed by using headphone playback, the drawback being that a composer must either create solely for a tape playback context or supply a headphone playback system for each member of the audience.

The *receptor* is the listener who experiences these sound waves in some manner. The experience of the receptor includes the *immediate perceptual recognition* of the spatial aspects of the sound, as given by cues based on monaural and binaural differences in intensity, spectra and time delay, and the higher-level *cognition* of spatial manipulation experienced by a listener (based on memory, association, expectation and patterning).

The chain of events illustrated in Fig. 1 shows how the listener's cognition of the musical-spatial gesture will not be equivalent to the composer's compositional intention. This is partly a result of the number of translations that the idea must undergo and partly because composers have difficulty in predicting the limitations of their tools. There is also the argument that the listener needs to be prepared by being familiarized with the particular grammar of spatial syntax or assisted by means of program notes or visual media.

The responsibility for the mismatch problem in the production of musical-spatial sound is not directly attributable to any discipline. Psychoacousticians have experimental goals that isolate parameters of localization in order to examine specific responses of human subjects. Research in concert-hall acoustics tends to emphasize the subjective quality for a given musical application, usually symphonic music. Composers are perhaps the greatest bearers of responsibility for the mismatch, in their assumption that spatial gestures drawn on a computer screen or described verbally can actually be perceived.

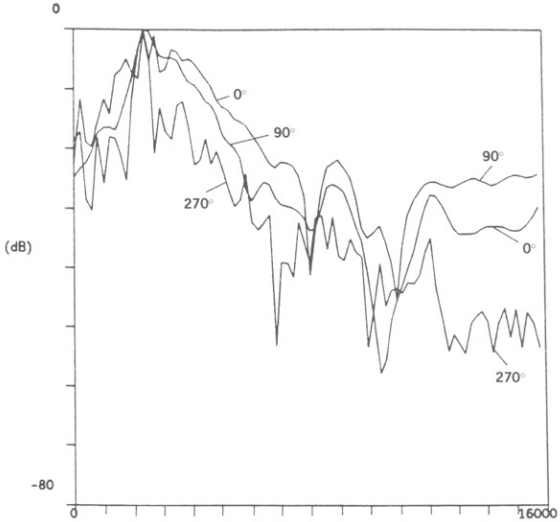


Fig. 2. Spectra of HRTF for one ear of a single subject, for a source at 0°, 90° and 270° azimuth, 0° elevation (adapted from measurements performed by F. Wightman and D. Kistler of the University of Wisconsin-Madison). These spectral changes have been shown to be significant perceptually for auditory localization.

HEADPHONE MUSIC AND COMPOSITIONAL CONTROL

Writing music for headphone listening frees the composer somewhat from the problems of the mismatch described above. In particular, the effects of the medium are bypassed; headphone listening is more consistent with the composer's experience because room acoustics are eliminated. There are considerable differences between models and brands of headphones, but these differences are less noticeable than the potential differences between loudspeaker environments. In other words, rooms vary more in their effect on the mismatch than do transducers.

The headphone music experience is qualitatively different from a typical concert situation. Undesirable sounds, such as the coughing, rustling of programs and emergency-vehicle sirens that are often heard at concerts, are almost completely masked. And while at a concert one shares in the community space of an audience, one enjoys a direct personal relationship with sound when listening with headphones; the sociological and cultural accessories of music are eliminated.

There is an advantage to composing headphone music because of the proliferation of portable cassette and compact disk (CD) players that are intended to be used with headphones. Their popularity with music listeners is easily explained: they are lightweight, the sound quality is generally very good (especially with coherent systems, where the entire playback chain stems from a single design), and they are especially suited to a society where privacy and control of one's personal sonic space is increasingly rare. Many New Age music composers and sound sculptors have borrowed heavily from the environmental sound recording tradition and offer headphone music as a type of therapy. The relative isolation that headphone listening affords is difficult to achieve in either concert-hall or home-loudspeaker environments.

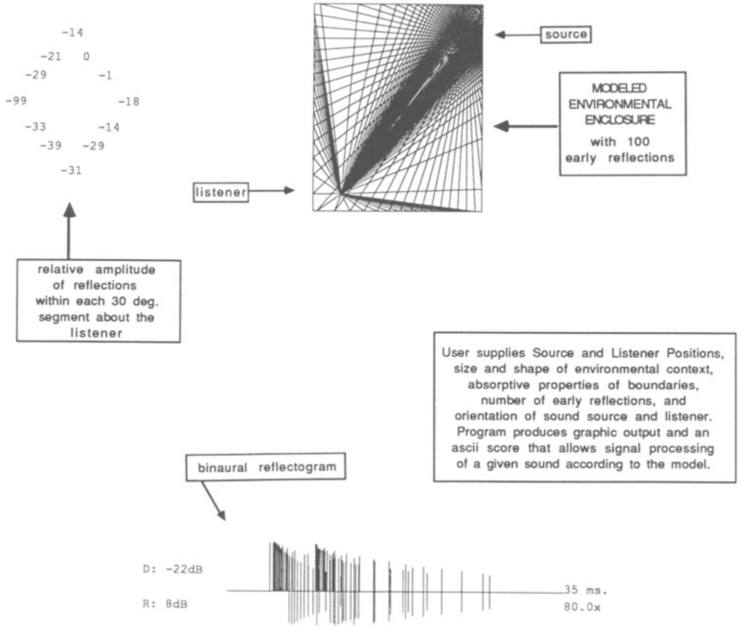
Headphones offer greater control for the researcher or composer because the signal at each ear can be predicted more accurately. In particular, determination of thresholds can be quite accurate. Psychoacoustic research usually makes the following distinctions between types of headphone presentation. In a *diotic* presentation, a single signal is played to both ears; in a *dichotic* presentation, two different signals are fed separately to each ear; and a *binaural* presentation approximates normal hearing with two ears—specifically, it is a dichotic presentation in which the content of one of the two signals is to some degree present in the other. The latter case represents the situation of spatial hearing most accurately and is most useful to the composer.

Simultaneous use of diotic and binaural presentation can be exploited for compositional ends in headphone music. Perceptually, a diotic signal is heard within the head, while binaural signals are externalized to either side of the head. In the computer music work *Revelations* described below, both kinds of presentation were used simultaneously to organize and differentiate several layers of text.

MODELING THE HEAD-RELATED TRANSFER FUNCTION FOR SPATIAL SIMULATION

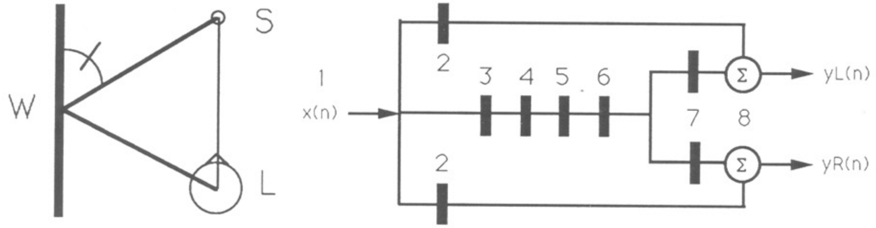
I have designed a digital signal-processing algorithm called REFL for creating spatialized versions of a digital sound file according to an arbitrary model [8]. This algorithm allows compositional specification of a model that includes the position of a listener and of the sound source within a variable environmental context. The interesting aspect of the algorithm is the incorporation of filters that simulate the head-related transfer function (HRTF). In essence, these filters create spatial listening cues for a listener by modifying an input sound in the same way that the outer ears (or *pinnae*) and the head would modify a sound in an actual environmental context. Simply put, the filtering effect changes as a function of the angle of incidence of the sound source; listeners may interpret the resulting changes in spectra and delay as changes in the spatial position of the source. Figure 2 shows the differences in spectra at one ear for a particular listener when the sound source is at 0°, 90° and 270° azimuth.

These changes in spatial cues can be rendered by use of *Kunstkopf* (or 'dummy head') recordings in which omnidirectional microphones are placed in the outer ear of either the head of a mannequin or an actual listener. (The mannequin head is usually preferred because it is immobile; head movement by a real listener during recording can cause shifts in spatial imagery during playback). When one listens through headphones to recordings made in this way, the spatial imagery is recreated to a greater degree than with normal stereo. Convincing auditory spatial illusions of azimuth are easy to achieve, but there is a persistent problem of front-image distortion; for



User supplies Source and Listener Positions, size and shape of environmental context, absorptive properties of boundaries, number of early reflections, and orientation of sound source and listener. Program produces graphic output and an ascii score that allows signal processing of a given sound according to the model.

Fig. 3. Graphic output of the REFL program. This program can be used compositionally for spatial signal-processing of a digital sound recording by supplying information about sound source and listener positions and orientations within an environmental context.



Figs. 4a, 4b. Signal-processing scheme used in the REFL program. Fig. 4a (left): simplified illustration of source S, listener L and wall W, with one direct path of sound (source-listener) and one reflected path of sound (source-wall-listener). Fig. 4b (right): circuit diagram showing how a one-channel input *X(n)* is spatialized into a two-channel binaural signal, *yL(n)* and *yR(n)*, corresponding to the source-wall-listener illustration in Fig. 4a. Refer to the text for an explanation of the numbers in the circuit diagram.

reasons still under investigation, sounds coming from in front of the listener are spatially perceived by many listeners to be coming from behind.

Despite the relatively simple technology and perceptual convincingness of *Kunstkopf* recordings, a serious problem for composers is that the spatial imagery is unalterable once the recording is made. Although one could conceive of moving a sound source around a dummy head to create spatial imagery, composers usually require greater control and ease of spatial variation. Hence, HRTF filtering software and hardware were developed to do essentially what a binaural recording does but within the domain of digital signal processing. According to Blauert [9], this technique was first implemented in West Germany in the 1970s. It has subsequently been implemented in various ways by several universities and corporations.

Besides implementing the HRTF filtering technique, the REFL program also allows the user to specify the shape, number of boundaries and frequency-dependent absorption coefficients of a modeled environmental context. The boundaries can be considered the walls of an enclosure, such as a concert hall or room; the sides of a canyon wall; or any sound-reflecting obstacle. The position and orientation of the sound source and the listener are specified within the enclosure. A ray-tracing method is used to sample the indirect sound field of the source as it reaches the listener, in terms of a given number of discrete sound reflections.

Figure 3 shows the graphic output of the REFL program. At the upper right is a two-dimensional graph of the ray-tracing method's output, showing the pattern of reflected sound from the source to the listener within the specified environmental context. The arrival time and relative amplitudes of the reflected sound at each ear are shown in the 'binaural reflectogram' at the bottom. At upper left is a summary of the relative amplitude in decibels (dB) of the reflected sound in each of 12 directions on the horizontal plane over the total period of time modeled by the ray-tracing method.

The direct sound and the reflected sound are spatialized based on their time of arrival and angle of incidence to the listener using HRTF filtering. As a result, the program is both listener-based in its implementation of important binaural cues and environmentally based in its implementation of sound reflections.

The program calculates the angle of incidence of the reflection to the wall and the angle of incidence to the listener, as given by the results of the ray-tracing algorithm. Presently 12 spatialization filters are used to group any of the potential incoming angles of sound reflections on the horizontal plane into 30° increments. The angle of incidence of the reflection to the listener is synthesized by an HRTF filter within 15° of the calculated value.

A diagram illustrating the signal-processing scheme used for the REFL program is shown in Fig. 4. At the left of the figure (Fig. 4a) is a simplified illustration of the model on which the signal-processing scheme is based, involving a listener L, a sound source S and a wall W. At the right of the figure (Fig. 4b) is a diagram of the signal-processing operation, which is read from left to right. The numbers in this circuit diagram show each stage of the signal-processing operation, corresponding to the numbers in the following description.

1. A digital recording *X(n)* is split into three paths. The center path involves the processing of reflected sound, based on the ray-tracing method. This corresponds to the two lines in Fig. 4a that span between S and W and between W and L. The upper and lower paths in Fig. 4b represent the processing of the direct sound that arrives to the two ears of the listener, corresponding to the line that spans between S and L.

2. The direct sound is spatialized by two HRTF filters, corresponding to the angle of incidence of the source to the left and right ears. An attenuation is also applied, based on the distance between the source and the listener.

3. Steps 3–7 refer to the processing of the reflected sound. At step 3, the distance the reflected sound must travel between S and W results in a delay and an amplitude loss.

4. A digital filter whose transfer function corresponds to the frequency-dependent characteristics of a specified surface is used to process the reflected sound. One can choose from a variety of surfaces with the REFL program, including typical wall surfaces (e.g. plaster, wood or velour curtains) and environmental contexts (e.g. granite, thick shrubbery or water).

5. The angle of incidence of the reflected sound to the surface is analyzed and then attenuated according to an adaptation of a physical model of specular reflection (specifically, the Fresnel-Kirchoff extension of Huygen's law). Essentially, for frequencies above approximately 500 Hz, the reflection amplitude is decreased as the angle of incidence of the reflection to the surface becomes smaller.

6. The distance the sound must travel between W and L corresponds to an additional delay and amplitude loss.

7. The angle of incidence of the reflected sound from the surface to each of the listener's ears is calculated. This angle is used to determine which pair

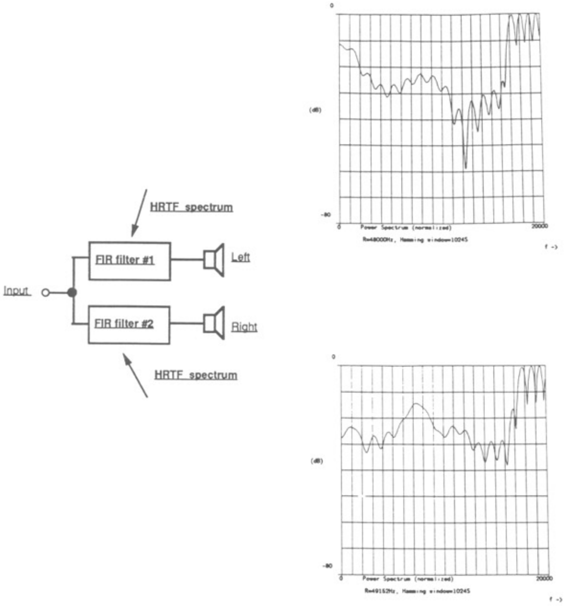


Fig. 5. Application of two HRTF spectra (top right and bottom right) to two finite impulse response (FIR) filters to derive a binaural signal from a single input. The spectra are used to produce a sound source located opposite the right ear.

of HRTF filters will be used to spatialize the reflected sound.

8. The output of the left ear direct-sound HRTF filter (step 2) is summed with the output of the left ear reflected-sound HRTF filter (step 7) and then is passed to the left channel output $yL(n)$. The same operation is performed to feed output $yR(n)$ for the right ear.

The frequency response and group-delay characteristics of HRTF filters used in the REFL program are based on data given by Blauert that represent the average values for 25 subjects [10]. Using a filter design program, I was able to design a set of finite impulse response (FIR) filters that closely approximated Blauert's data [11]. An advantage to using the filter design program was that the number of multiplications used in the computation could be reduced significantly; this translates into a faster-running algorithm. (I am currently using this approach for eventual real-time application of these filters.)

Figure 5 shows how two of these FIR filters are used to process a single channel of input into a binaural signal. The HRTF spectra shown here are from the filter design program and correspond to the frequency response modification needed to create the illusion of a sound positioned directly opposite the right ear. (A frequency-specific time delay is also inherent in these filters, but is not shown in the illustration.)

When one listens over headphones, the perceptual effect of these HRTF filters is quite dramatic; this allows a composer a level of control over spatial imagery that is impossible to attain using standard stereophonic techniques. However, for most listeners there is still a mismatch between compositional specification and human perception. Some persons are unable to externalize HRTF-filtered sounds over headphones, particularly those sounds filtered to be perceived as coming from directly in front of the listener (0° azimuth). There are also frequent reversals of 0° HRTF-filtered sound; many people hear these sounds as coming from behind them at 180° instead. This has been mitigated in some three-dimensional sound demonstration tapes by using examples that take advantage of the listener's cognitive 'rearrangement' of spatial sound. For example, it is quite easy to imagine an HRTF-filtered sound of lighting a cigarette or drinking a glass of water at 0° azimuth, since this corresponds to the location of the mouth.

Figure 6 shows the 'bow tie' pattern that is frequently reported by listeners when a composer intends to create a circular pattern of 12 sounds that have a constant radius from the center of the head [12]. Notice that the ability to localize left and right positions is more successful than front and rear positions. For many listeners the sounds also seem higher vertically as they move toward the front.

In spite of these perceptual mismatches, this HRTF signal-processing technique is still a valuable one to the composer in that 12 spatially distinct auditory positions in a three-dimensional virtual space can be attained. Eliminating these mismatches is a current area of my research at NASA-Ames research center; other investigators are researching this as well. At present, a substantial amount of psychoacoustic investigation is required to completely eliminate ambiguity between a composer's intention and the perceived result. However, if the composer is aware of this ambiguity and takes pains to remain aware of progress in the domains of psychoacoustics and signal processing, then headphone music using HRTF filtering should remain an increasingly fecund area for spatial music composition.

COMPOSITION OF SPACE: TWO EXAMPLES

Revelations

The REFL program was applied to a segment of my composition *Revelations*. This headphone composition uses the spatial manipulation of text fragments, and the juxtaposition of these fragments, as its principal type of compositional grammar. Instrumental and computer sounds function as accompaniment to the text. The work was realized on a DEC Vax 11-780 computer at the Computer Audio Research Laboratory (CARL) at the University of California, San Diego.

The text in the part of the composition described below is based in part on a segment from a poem by John Giorno, "We Got Here Yesterday, We're

Here Now, and I Can't Wait to Leave Tomorrow" [13].

you are bored
you are bored
you are bored and restless
you can't think of anything to do
to do
to do
to do
to do
you can't think of anything to do

Spatialized piano and violin sounds (trills, plucked strings and upward glissandi) were used to accompany the voice. A digital recording was made of a reading of the text, and then it was divided into separate 'soundfiles' ranging in duration from 1 to 8 sec. Except for the soundfile consisting of the words 'to do', the entire reading of the text in the example discussed here was delivered diotically, resulting in an inside-the-head spatial perception. This caused the spatialized words 'to do' and instrumental sounds to be differentiated in the overall musical-spatial texture; these sounds surround and move outside the listener, while the remainder of the text is heard within the head. The sounds were spatialized by the REFL program and then were mixed with the diotic sounds using Cmusic software [14].

In this example, there were three types of spatial gestures or 'trajectories' of sound movement associated with certain sounds. The intention was to compose spatial gestures that were procedurally related in their organization (but not necessarily cognitively related to the listener). By maintaining one aspect of spatial movement with all gestures, but varying other aspects, a compositional approach to space can be made that is analogous to melodic 'theme' and 'variation'. The spatial aspect that remained constant for all gestures was 'circularity', defined here as moving completely around the listener, according to some pattern of movement. The spatial aspects that were variable included the kind of pattern used to attain circularity, the overall time required to complete the pattern, and the direction of circularity (clockwise or counterclockwise).

The basic spatial 'theme' for this example was an 8-sec counterclockwise circle, used for the piano and violin sounds. To form this trajectory, a series of 12 attacks at successive 30° positions was created, with about 0.5 sec between each attack. The violin and piano were spatially separated by 180° before the beginning of the trajectory so that one sound followed the other at the opposite position within the perceived space [15]. The piano and violin sounds each had a duration of about 1.5 sec; as a result, the decay of each previous attack overlapped with the attack of the next sound. The perceived result was that of sound movement rather than that of a sound switching between 12 discrete points.

For spatial variation, two types of trajectories were used for the words 'to do'. In the first trajectory, positions shifted in 0.5-sec intervals, moving clockwise 30° and then 60° around the listener (see Fig. 7a). This trajectory was meant to be a variation on the circular trajectory discussed above; it differs in the use of an opposite (clockwise) direction and use of 60° 'leaps'. Also, the movement effect discussed previously did not result, because the sound was approximately 0.5 sec in duration, with no overlap between attacks.

In the second trajectory, positions shifted in 0.25-sec intervals to the 180° opposite position and then back 150°: 0–180–30–210–60–240, until all 12 positions were articulated (see Fig. 7b). This resulted in a rapid see-saw effect, chosen for its alternation from side to side while the circular identity of the other trajectories is maintained.

Begault Meadow

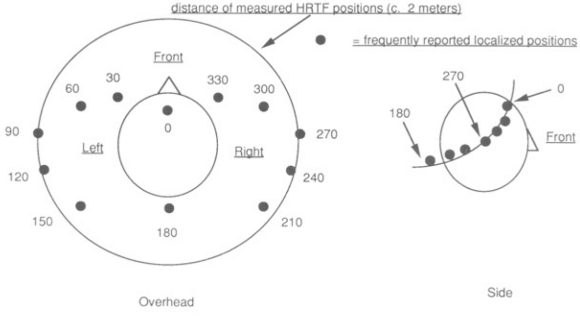
A composition for headphone listening by Gordon Mumma has been completed in sketch form at the University of California, San Diego, Computer Audio Research Laboratory using the HRTF filters and the REFL program described above. Titled *Begault Meadow,* it uses only a single, short sound, a 50-msec pulse of relatively broad-band noise with a mildly tapered attack and decay.

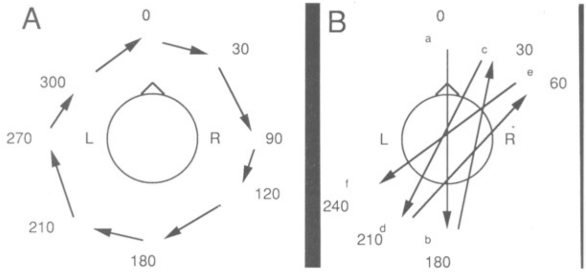
Mumma writes:

> The musical syntax is achieved entirely with the spatial placements of that single sound; there are no pitch, timbre, or loudness attributes except as they result from spatial location of the single sound. The poetic impetus for the music derives from Abbott's classic book *Flatland;* to support the 'meadow' part of the title, the acoustical characteristics of an outdoor environment are synthesized. The duration of *Begault Meadow* is projected at between 20 and 30 minutes. The spatial choreography of *Begault Meadow* varies from one section to another of the piece. In some sections the sounds clearly inhabit specific azimuth regions around the listener; in other sections the location of sounds is more ambiguous. The play between specificity and ambiguity is largely a function of compositional choices [16].

Since few of us experience a 50-msec burst of white noise in a meadow, one does not hear the work and then immediately recall such an environment. If one wanted to create such associations for the listener, it would be best to use cognitive cues, as is done with sound design for film; for example, cows, wind through the grass and insect sounds. Even then, we still might not be sure it was a meadow that we were in, using only our ears. The meadow of this work is a virtual, artificial creation that is in certain ways unique; it must be learned through repeated experience (listening). Because of the implementation of significant physical parameters of the meadow into the REFL program, the

Fig. 6. 'Bow tie' pattern showing the perceptual result of HRTF filtering, as reported informally by many listeners. When one intends to create a circle of sounds equidistant from the center of the head, a distorted pattern is usually heard. Not shown is the fact that the 0° azimuth position is often 'reversed' (i.e. heard as if it sounded at 180°).





Figs. 7a, 7b. Two variations on a circular spatial trajectory 'theme' used in the author's composition *Revelations.* The numbers represent the HRTF-filtered positions, in degrees, of the speech fragment 'to do' from a digitally recorded poem. (a) Clockwise movement—leaps between positions alternate between 30 and 60° at 0.5-sec intervals. (b) See-saw effect—leaps between clockwise opposite positions (ordering shown by letters a-f) at 0.25-sec intervals.

sound source's articulation of the environment results in timbral changes to the sound source as a function of its location. One consequently has a more realistic sense of spatial extent and of being surrounded by the sound source; the environment is differentiated, as in the real world [17].

Variations in the spectral content and temporal pattern of reflected sounds were explored for their usefulness as environmental cues. This was done with the REFL program using the environmental model shown in Fig. 1. Forty-eight reflected sounds were spatialized according to angle of incidence to the listener and the effect of wall surfaces and distances. The landscape of the meadow to be simulated was differentiated by modeling a grassy meadow to the rear, a steeply rising hill to the right front, and a lake to the left front. The simulation was achieved through specification of frequency-dependent absorption for each modeled 'wall' of the REFL program. The left-front 'wall' extended into the distance to attenuate the energy arriving from the area of the 'lake'. Overall, the pattern of sound reflections was varied both spectrally and temporally as a function of the location of the sound source in the 'environment'.

In collaboration with Mumma, I designed a compositional algorithm to produce a computer music 'note statement list' that uses 'compositionally weighted' random distributions of attack times and spatial positions [18]. Unlike the music for *Revelations, Begault Meadow* uses as many as 600 note statements over a single minute. The algorithm allows the composer to specify the range of azimuth and distance information to be used for a particular section of the work from precomposed files. Events can be distributed over time according to (1) specified minimum and maximum times between sounds, (2) a 'seed' for the distribution of events and (3) a 'divisor' factor to compress or expand the time distribution of events. The composer edits each resulting note list and then layers it with or appends it to other note lists before running a program in order to hear and evaluate the result.

CONCLUSION

The above section shows but two examples of the potential application of the HRTF signal-processing technique to the composition of spatial music. Undoubtedly, the match between compositional specification and spatial hearing will improve with further developments in audio signal-processing techniques and with additional psychoacoustic studies. Headphone music continues to be a largely unexplored territory of the infrequently visited universe of spatial music, and we should expect our 'mind's aural eye' to be further surprised and challenged in the future.

References and Notes

1. Susanne Langer, *Feeling and Form* (New York: Macmillan, 1977).

2. Victor Zuckerkandl, *Sound and Symbol: Music and the External World* (Princeton, NJ: Princeton University Press, 1956).

3. John M. Chowning, "The Simulation of Moving Sound Sources", Audio Engineering Society 38th Convention Reprint No. 726, 1970.

4. F. Richard Moore, "A General Model for the Spatial Processing of Sounds", *Computer Music Journal* **7**, No. 3, 6–15 (1983).

5. Gary S. Kendall and William L. Martens, *Simulating the Cues of Spatial Hearing in Natural Environments*, Technical Report (Evanston, IL: Northwestern University Computer Music, 1984).

6. I developed and then implemented the REFL spatialization software at the Computer Audio Research Laboratory (CARL) at the University of California, San Diego.

7. Roger Reynolds, "Thoughts on Sound Movement and Meaning", *Perspectives of New Music* **26**, No. 2, 181–190 (1979).

8. Durand R. Begault, "Control of Auditory Distance", Ph.D. diss., University of California, San Diego, 1987.

9. Jens Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization* (Cambridge, MA: MIT Press, 1983).

10. Blauert [9].

11. This algorithm was an adaptation by F. Richard Moore of the EQFIR program distributed as part of the IEEE's *Programs for Digital Signal Processing* (New York: IEEE Press, 1979).

12. See Ref. [8].

13. John Giorno, "We Got Here Yesterday, We're Here Now, and I Can't Wait to Leave Tomorrow" (1981), in John Giorno, *Grasping at Emptiness* (New York: Kulchur Foundation, 1985).

14. Cmusic is a software synthesis program written by F. Richard Moore that is a part of the software distribution package of the Computer Audio Research Laboratory at the University of California, San Diego.

15. When listening to two 180°-opposed circularly moving sounds, listeners seem to focus attention in two ways: most listeners could choose either hearing the two streams of spatial information simultaneously as a unit or hearing one stream only. When listening to both streams, one has the impression of being at the pivot of an axis with a sound source at either end.

16. Gordon Mumma, "Audioearotica", in *Center for Music Experiment and Related Research Annual Report 1986–7*, John D. Lauer, ed. (La Jolla, CA: Univ. of California, San Diego, Center for Music Experiment, 1986) pp. 16–17.

17. Reflected sound can both convey information about the environment and cause coloration of the sound source. An example is how the sound of one's voice changes in different rooms of a house.

18. The computer music software used for interpreting these note statements was also Cmusic.