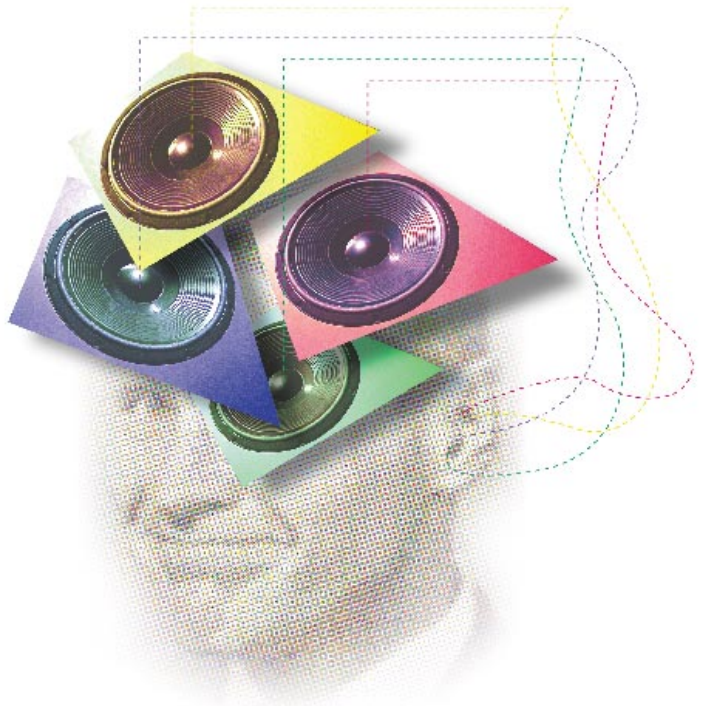YOU MAY PLAN TO INCLUDE ONLY TWO SPEAKERS, JUST A SET OF HEADPHONES, OR SUPPORT FOR ONE- AND TWO-CHANNEL SOUND SOURCES IN YOUR NEXT DESIGN. NOT TO WORRY. PSYCHOACOUSTICS, PROCESSORS, AND PLENTY OF MEMORY CAN STILL GIVE YOUR CUSTOMERS AN IMMERSIVE AUDIO EXPERIENCE. "HEAR" IS HOW.

Decoding and virtualization
bring surround sound to the masses

PEOPLE COMMONLY USE THE TERMS "stereo," "two channel," and "two speaker" to mean the same thing, but they're committing an etymological error. "Stereo" derives from the Greek word *stereos*, meaning "solid." In other words, a stereo recording-and-listening environment is one that reproduces the 3-D atmosphere of the original performance. Bob Dylan's *Live 1966* ideally should sound just like it did when first performed in Manchester, England, and Miles Davis' *Kind of Blue* should transport you to 1959 and Columbia's 30th Street Studio, New York, even if you're listening to them in your living room in Sacramento, CA, in 2001. And, when the Tyrannosaurus Rex in the movie *Jurassic Park* growls on-screen, you should hear it over your shoulder and 20 feet above you, and the hair on the nape of your neck should rise.

As early as the 1930s, Bell Labs researchers JC Steinberg and WB Snow determined that you need at least three transducers to realistically reproduce an audio source (**Figure 1a** and **b**). Their research didn't comprehend *additional* speakers necessary to replicate the acoustics of the listening environment. Monophonic radios, phonographs, and tape players, eventually supplanted by their two-channel variants and by audio CDs, existed for reasons of the economics and technology limitations of the time, not because they delivered realism. Some of you may be familiar with the ill-fated quadraphonic (four-channel) systems that briefly appeared in the early 1970s. The theory behind quadraphonic audio was fairly solid, but the implementations weren't. Manufacturers' proprietary systems and corresponding media were incompatible with each other, and compelling content was lacking. Deluded audiophiles like to point to the failure of quadraphonic sound as justifica-

tion for two-channel-only audio (**Reference 1**).

The fact is surround sound in movie theaters has thrilled consumers for decades. Leopold Stokowski and the Philadelphia Orchestra performed the classical music in Disney's (www.disney.com) 1940 film *Fantasia* in Fantasound, and it is but one early example of surround sound. Now-ubiquitous Dolby Surround first appeared in the mid-1970s, Dolby Digital debuted with the movie *Batman Returns* in June 1992, and *Jurassic Park* followed in 1993 in DTS (Digital Theater Systems) surround. Dolby Surround-encoded television broadcasts and videotapes are now commonplace and, along with Dolby Digital- and DTS-aware DVD and audio CD players, have brought surround music, movies, and other programs into living rooms and automobiles. DVD-Audio and SACD (Super Audio Compact Disc), assuming they succeed in the market (which is *not* a foregone conclusion), will accelerate this awareness, as will immersive gaming and other computer-based audio environments (**references 2 and 3**). Teleconferencing, another po-

AT A GLANCE

▷ Armed with an auditory system understanding, a powerful-enough processor, and sufficient RAM and ROM, you can give your customers an immersive audio experience no matter what the sound source is.

▷ Artificially created and artfully extracted reflections and reverberations can both approximate and alter the ambiance of the original recording environment.

▷ Low- and high-frequency enhancement compensate for bass-deficient speakers and lossy compression artifacts.

▷ Interaural time and intensity adjustments and head-related transfer-curve data enable you to place phantom sound sources all around a listener.

▷ Speakers and headphones require different audio-processing schemes.

tential surround-sound application, enables listeners to differentiate individual speakers, talking at the same time at the other end of the line, within a group.

Another common argument of the audiophile Luddites against audio with more than two channels is the baffling statement that they need only two speakers because they have only two ears. In reality, the human auditory system, working with other sensory faculties, such as vision, can accurately locate a sound source in 3-D space using both absolute and differential time-of-arrival, intensity, and frequency cues along with training in how the head, shoulders, and ears modify incoming sounds. Even if the sound source is directly in front of you, the echoes and reverberations of the recording environment will alter it unless the recording occurs in an anechoic chamber. Because the end listening environment doesn't have the same acoustics, you need audio processing to re-create a semblance of the original.

Now that listeners have enjoyed surround sound with portions of their television programming, movies, and music collections, they'd like to extend immersiveness to all of their multimedia experiences, regardless of the sources' characteristics—that is, monophonic or

HAVEN'T HEARD ENOUGH?

If this article has whetted your appetite to learn more about surround sound, I encourage you to check out **references 1** to **10** in the main article. *Surround Sound Professional* magazine (www.surroundpro.com), which audio pioneer Tomlinson Holman heads, is also an excellent source of news and information and has particular appeal to recording engineers. The magazine sponsors a conference in Beverly Hills, CA, each December, as well as seminars at the Consumer Electronics Show (www.cesweb.org) and other forums.

Speaking of conferences, the biannual Audio Engineering Society (www.aes.org) convention is an impressive group of some of the finest minds in audio, and the AES also sponsors periodic conferences on audio topics. An AES member-

ship is worthwhile if only for the 10-issue-per-year *Journal of the Audio Engineering Society*.

Most of the vendors listed in the "For more information..." **sidebar** have Web sites flush with white papers, application notes, other documentation, presentations, and sound samples. Each thinks its own approach is the best, but, by surveying a number of them, you gain a breadth of knowledge on technology and product alternatives.

And speaking of Web sites, I list some that have been particularly useful to me in my research. Fire up the Google (www.google.com) search engine using keywords "surround sound" to find hundreds of others.

● Floyd Toole, long-time and well-known audio researcher, has several detailed white

papers on the Harman Web site (www.harman.com).

● David Griesinger, principal scientist at Lexicon, a division of Harman, provides a large number of his AES and other papers and presentations for free downloading at www.world.std.com/~griesngr. You can even see a picture of him sketching **Figure 3**!

● Well-known audio engineer Bobby Owsinksi runs Surround Associates, and the company's Web site at www.surroundassociates.com contains several helpful articles.

● The Ambisonics FAQ is at www.stanford.edu/~mleese/Ambisonic/faq_latest.html. Also, check out www.ambisonics.net for more information on this set of techniques, developed in the 1970s and intended for the recording, studio processing, and reproduction of the com-

plete sound field that occurs during an audio performance.

● Bob Stuart from Meridian Audio, the developer of the MLP compression scheme used in DVD-Audio and one of the key definers of the DVD-Audio standard, maintains the Acoustic Renaissance for Audio Web site at www.meridian-audio.com/ara, which contains a number of interesting papers.

● The Ambiophonics Web site at www.ambiophonics.org discusses how to optimize the acoustics of typical home listening room environments.

● The 3D Audio Immersion site at www.3dai.net keeps up to date on the latest news in the world of surround sound.

Finally, watch for my Jan 10, 2002 cover story, which will describe and compare product alternatives, including complete encoding/decoding systems.

two-channel—or the listeners' settings, such as in an automobile, an airplane, an office, a conference room, a living room, or on the jogging track. But, because of financial and aesthetic resistance to buying and installing center, rear, and sub-woofer speakers—the so-called SAF

(spousal-acceptance factor)—many of them would like to gain an approximation of the true surround-sound experience using their current two-speaker or headphone configurations. Fortunately, just as with lossy audio compression, an in-depth understanding of both the

strengths and the shortcomings of the auditory system, along with some DSP horsepower and memory, can credibly accomplish these seemingly divergent objectives (**references 4** and **5**).

If you're starting with a single-channel—that is, monophonic—audio
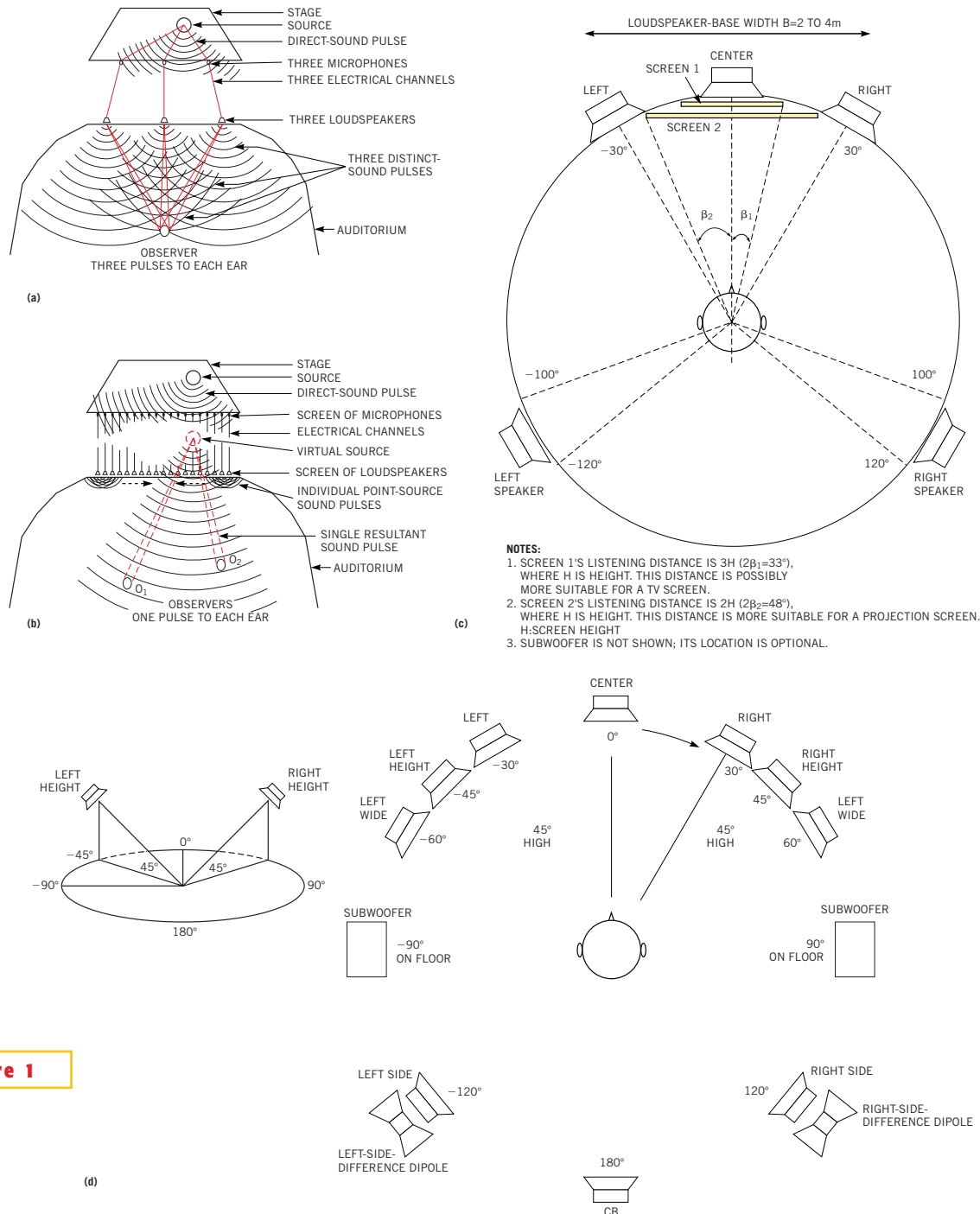
Bell Labs research in the 1930s suggested that you need at least three speakers (a, courtesy Focal Press) to credibly reproduce a sound source (b, courtesy Focal Press). The 5.1-speaker configuration is common in today's home-theater settings (c, courtesy Focal Press), whereas Tomlinson Holman's 10.2-speaker proposal suggests one possible future arrangement (d, courtesy Tomlinson Holman Labs).

source, it might seem impossible at first glance to create a two-channel variant—never mind a full-blown surround-sound representation—of it (**Reference 6**). Recall, though, that people generally regard low-frequency sounds as nondirectional; that is, the human auditory system can perceive them no matter what location in the listening area they come from. Therefore, you can apply a lowpass filter to the source and direct frequencies of less than 100 Hz or so only to those channels whose transducers will likely be able to reproduce the frequencies, such as a distinct 0.1 subwoofer channel.

Next, delay the highpass-filter output by an adjustable amount of time if you want to allow for user customization and then add it to the nondelayed signal to create one output channel and subtract it from the nondelayed signal to create a second channel (**Figure 2**). The results are true and complementary comb filters, and, when you present the two channels through two speakers, the frequencies split evenly between them. Running both channels through the same speaker cancels out the delay and re-creates the original monophonic signal. Advanced technologies from some of the companies that the **sidebar** "For more information..." lists employ more elaborate filtering techniques to transform monophonic audio. Patent searches and conference papers reveal some clues on these advanced techniques, but you must sign a nondisclosure agreement to get all the details (see **sidebar** "Haven't heard enough?"). In contrast, some simple and cheap pseudo-two-channel algorithms simply subdivide the audio into multifrequency bins, allocating the bins among the various channels.

Now, you've got a two-channel audio clip, either as an original recording, which, let's assume, contains no matrix-encoded center and mono-surround information, or one resulting from the earlier mono-to-two-channel conversion. Next, you might want to first alter the "sweet spot," or region within which a person listening to the audio can hear both channels. For a computer user sitting before a display, the sweet spot can be narrow and shallow, which often results in more accurately perceived sound positioning and which is particularly appealing with 3-D games. A home-theater setting, in which listeners may be in 10-

Figure 2
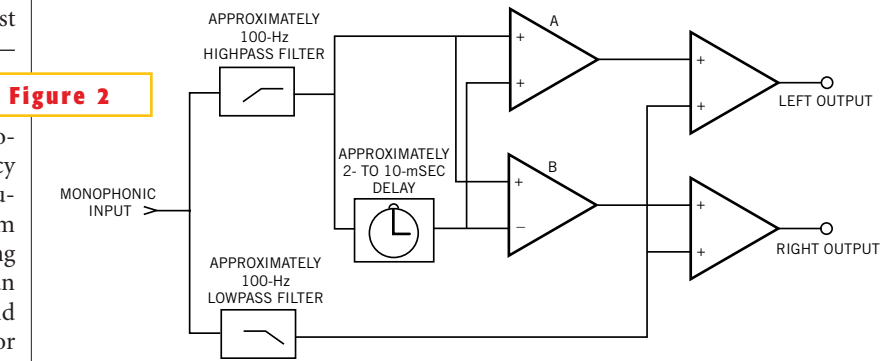
Frequency subdivision, followed by time delay and mixing, is one means of converting monophonic audio into a two-channel representation that retains monophonic-speaker backward compatibility (courtesy *DV Magazine*).

seat rows or in which the listeners are milling about instead of remaining in one location, requires a wide, deep sweet spot. A similar situation exists in automobiles, in which neither the driver nor any of the passengers is in an ideal listening location. Bigger sweet spots result in a more immersive surround-sound experience at the possible expense of reduced sound-source-positioning precision.

The sweet-spot characteristics also depend on the anticipated placement of the two front speakers and on the geometrical relationship between their spacing and the listener's location. If the speakers are several yards apart, as with most audio/video receiver setups, the sweet spot is naturally wider; however, the center-perceived location of audio material that they share is indistinct. Adjacent transducer placement, such as with speakers on either side of a television tube or computer display, creates a narrow sweet spot but a well-defined center location for dialogue and other common material. To enhance the center-channel characteristics, you determine what material the front left and right channels share and then emphasize it using HRTF (head-related-transfer-function) transformations. Ideally, center-channel information transmits through a dedicated speaker, because a "phantom center," which you create by coupling the left and right front speakers, exhibits timbre that differs from the real thing.

In addition to emphasizing the shared information, you might also want to broaden the audio image created by data in one channel and not in the other. One quick and dirty means of accomplishing this goal involves inverting the phase of

one of the two channels; the "stereo-wide" button in low-end consumer-electronics gear frequently activates this normally undesirable technique. This technique obliterates center imaging of shared-channel content. A more elaborate technique that less destructively scatters the sound-source-directional cues involves first calculating the channel $(A-B)$ and $(B-A)$ information and then employing frequency-dependent time, spectrum, and overall intensity alteration to create the perception that the sound is originating beyond the physical boundary of each speaker.

Next, how do you create listening-room acoustic effects for the rear surround speakers? First, you must differentiate among early reflections, echo, and late reverberations (**Figure 3**). These three phenomena result from sound reflecting off objects before entering listeners' ears, but the auditory system perceives them in different ways. *Early reflections*—those lagging behind the original sound by as much as 30 msec—enable the ear and brain both to locate the sound source and to perceive the room dimensions. Their amplitude depends on the reflectivity of objects the sound waves bounce off before entering your ear, whereas their delay is a function of room width, depth, and height and of the presence of reflective objects within a room.

We perceive direct reflections—those that bounce off only one or only a few objects before entering the ear—as echoes beyond 30 msec; hence, these reflections tend to degrade the sound. Conversely, sounds that have been reflected many times with attenuation at each reflection point come at a listener from all directions. Many of these low-amplitude,

diffuse late reflections, or *reverberations*, simultaneously arrive at the listener. A certain amount of reverberation is generally desirable. Think, for example, how much richer your voice sounds when you sing in the shower. Conversely, in rooms with little reverberation, the sound you hear is often unpleasant.

You create the perception of reflections and reverberations by employing RAM-based delay lines, along with signal processing, to modify the audio as a real-life reflection would. Different delays, intensities, and spectral transformations can create the illusion of a cavernous concert hall or an intimate jazz club. Artificial ambiance seems appealing in theory, but reality is often underwhelming, especially if a listener exaggerates the effect. An acoustical model that sounds good with a symphony orchestra, for example, might sound horrible with a solo pianist or vocalist. Short cuts in memory and processing power to reduce system cost and power consumption leave the resulting reverberation sounding artificial. Artificial ambiance also clashes with other ambiance already in the audio.

An alternative approach is more processing-intensive but more authentic in its results. It involves analyzing and extracting this existing ambiance and sending the reflections and reverberations to the rear channels. Audio engineers capture this ambiance during the original recording by using binaural, hypercardioid, or omnidirectional microphones, or they can add the ambiance to the audio during mixing—a high-tech version of singing in the shower. Directing reflections and reverberations to the rear speakers can be an effective arrangement in an automobile. The traditional auto audio configuration replicates the right and left audio channels in both the front and the rear speakers. This so-called dual-stereo configuration significantly diminishes the listeners' appreciation: The speakers not only are in poor locations but also bombard your ears with destructive crosstalk from both the front and the rear.

After creating additional audio channels, you might also want to enhance the perceived low frequencies of the audio to compensate for anticipated bass-deficient speakers or the high frequencies to counterbalance the effects of lossy compression. Harmonics play a part in both
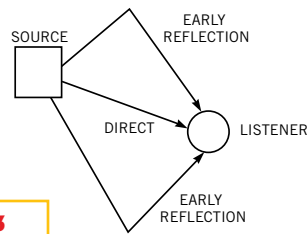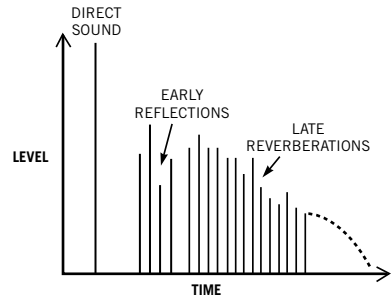

Figure 3

The auditory system perceives early reflections, echo, and late reverberations in different ways (courtesy Focal Press).

operations. Mix and play back 100- and 150-Hz tones in an audio-editing program, and you'll also hear what sounds like a 50-Hz tone. At the high end of the frequency spectrum, Kenwood's (www.kenwood.com) Supreme technology interpolates high-order fundamental tones from lower-order harmonics that have survived lossy encoding. Thomson Multimedia's (www.thomsonmultimedia.com) MP3pro compression scheme takes similar advantage of harmonics to shift high frequencies beneath harm's way during encoding, subsequently restoring them during decoding.

At this point, you've created front left and right channels, a center channel, one or more rear channels, and maybe even a subwoofer channel. Today's most common ideal reproduction setting is a six-speaker setup like the one that the ITU (International Telecommunication Union) defines (**Figure 1c**). Aesthetic considerations can drive subwoofer placement because the channel's sound isn't directional. However, if you place the subwoofer against a wall, especially in the corner of a room, you will perceive its sound as the loudest. For music reproduction, all other speakers should have full frequency response, and the surround-sound speakers should be directly radiating. Conversely, for movies, the center channel often carries only dialogue, and the surround-sound speakers find most use in reproducing special effects. Such home theaters often employ bipole or dipole surround-sound speakers for immersiveness. Unfortunately, they also often trade off speaker frequency response and other characteristics to reduce cost.

TWO-CHANNEL VIRTUALIZATION

What if having more than two speakers isn't an option? In this case, you need

to create the illusion of more speakers than actually exist. Recall that, in the ITU configuration, each ear of each listener perceives sounds that originate from all six speakers. Interaural time differences play a key role in locating sound sources of frequencies of 1 kHz and lower (**Figure 4**). Conversely, interaural intensity differences are the primary means by which the auditory system locates the source of sounds higher than 1 kHz. When a sound source is close to a listener, the spherical outward radiation of sound emanating from an off-center source and the resulting level difference between the ears of the listener are additional factors in determining location. The unique shape of each person's head and shoulders is an appreciable barrier to and spectral modifier of sound waves, and the spacing between ears, shape of each ear, and shape of the auditory canal leading to the eardrum are also key identifiers of a sound's direction.
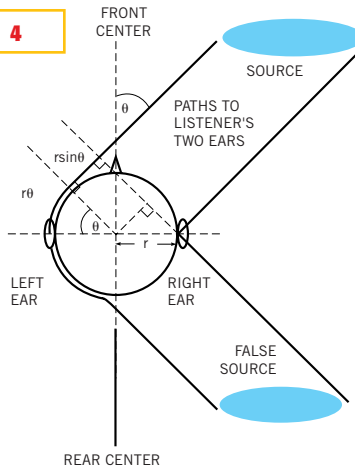
HRTF curves summarize these listener-specific phenomena. By transforming a sound's timing and frequency spectrum using the HRTF data, you can theoretically "place" a sound anywhere in space around a listener's head using only two speakers. Some algorithm developers claim that an audio engineer can employ basic amplitude-based rather than the more elaborate HRTF-based alternatives to give the illusion of a sound moving from front to back. Some algorithm developers claim that if an audio engineer has employed basic amplitude-based effects rather than more elaborate HRTF-based techniques to give the illusion of a sound moving from front to back, an HRTF-aware virtual-surround-sound system can deliver a more realistic result than a front- and back-speaker combination. Because our ears are on the sides of our heads, we're particularly sensitive

to shortcomings in "phantom" midlateral speakers, which HRTF transforms account for, particularly if those transforms also adjust for interaural crosstalk effects.

Conversely, when a front-placed speaker pair creates virtual surround, it's difficult to spin a psychoacoustically altered sound completely around to the back of your head and equally challenging to communicate sound source height cues. If you're playing a video game, your head is stable and centered, and you also have visual "suggestions" in the form of on-screen objects to aid in locating sound. However, in the absence of visual assistance, you may experience front-versus-back location confusion. In real life, you'd turn your head to help find the sound. But, because the sound's location is virtual and the HRTF-transformation effect depends on your head's orientation to the real speakers, this reflex reaction makes the problem *worse*.

Ideally, HRTFs should be customizable for each listener, because we all have different head shapes, ear spacings and contours, and concha openness and depth. Multilistener environments obviate the effectiveness of such customization, however. Some HRTF algorithms also adapt their responses to plane sound sources, to nearby sources where spherical effects are important, and to listeners who are not in the sweet spot. (For example, when a train speeds

Figure 4

FRONT CENTER

SOURCE

PATHS TO LISTENER'S TWO EARS

θ

rsinθ

rθ

θ

r

LEFT EAR

RIGHT EAR

FALSE SOURCE

REAR CENTER

Frequency-dependent interaural time and intensity differences enable the ears and brain to determine a sound source's location, but lacking collaborating visual or head reorientation cues, these differences may result in front-versus-back confusion (courtesy Focal Press).

by, it does not serve as a point source.)

When audio engineers mix music and movie soundtracks, they assume that users will play the end results on traditionally placed speakers. As noted, even in a simple two-speaker configuration, each ear senses not only the output of its corresponding front speaker, but also a spectrally modified and time-delayed version of the opposite speaker. However, this acoustic crosstalk doesn't occur

with headphones, in which each channel is isolated to only a single ear. Uncompensated audio, intended for speakers but played back through headphones, produces an annoying in-the-head effect, compared with a compensated alternative, which mixes each channel with a time-delayed and spectral-modified version of the opposite.

HRTF transformations for virtualization with more than two speakers over headphones take a different form from their two-speaker equivalents. Because the speakers are on the ears and follow the head regardless of its orientation, sweet-spot issues disappear, more realistically achieving full 360° sound placement and convincing vertical positioning. However, the direct coupling of the transducer to the ear effectively removes outer-ear HRTF effects, which can make it difficult to achieve consistent results from user to user. Headphone-virtualization algorithms are also sometimes complicated because they can't use the acoustics in the listening room to help recreate the original environment. On the other hand, if the listening-room acoustics are detrimental (echo-conducive, for example), you should probably begin with a "clean slate."

To achieve more realistic reproduction of sound-source height (think of a rocket taking off) and to improve the realism of front-to-back and back-side-to-side sound source movement, audio

pioneer Tomlinson Holman advocates a 10.2-channel system (**Figure 1d**). His approach builds on the 5.1-channel ITU standard with the addition of dual side—that is, midlateral—channels, a rear center channel, dual subwoofers, and two height channels. The height channels target reproduction over speakers ±45° horizontally and 45° vertically away from the listener. As an interim step to 10.2 channels, several 6.1-channel approaches are gaining popularity. Dolby Digital Surround EX, which Dolby Labs developed with Holman, matrix-encodes a rear center channel, as does DTS-ES Matrix. DTS-ES Discrete employs a distinct rear center channel; Digital Theater Systems designed its bit-stream format to be backward-compatible with respect to additional audio channels; higher precision sample sizes; and higher sampling frequencies, such as the upcoming 24-bit, 96-kHz DTS. And THX Ultra2 technology extracts seven full-range channels and a subwoofer channel from 5.1-channel source material.□

REFERENCES

1. Guttenberg, Steve, and Brent Butterworth, "Stereo vs 5.1: Is more more....or less?," *Stereophile*, August 2001, pg 49.

2. Dipert, Brian, "'Bassless' buzz impairs advanced audio's image," *EDN*, May 24, 2001, pg 20.

3. Dipert, Brian, "Security scheme doesn't hold water(marking)," *EDN*, Dec 21, 2000, pg 35.

4. Dipert, Brian, "Digital audio gets an audition: part two, lossy compression," *EDN*, Jan 18, 2001, pg 87.

5. Dipert, Brian, "Digital audio breaks the sound barrier," *EDN*, July 20, 2000, pg 71.

6. Rose, Jay, "Stereotypes," *DV Magazine*, September 2001, pg 98.

7. Layton, Leonard, "Surround sound takes to the air," *IEEE Spectrum*, August 2001, pg 56.

8. Kraemer, Alan, "Two speakers are better than 5.1," *IEEE Spectrum*, May 2001.

9. Rumsey, Francis, *Spatial Audio*, ISBN 0-240-51623-0, Focal Press, Woburn, MA, 2001.

10. Coulter, Doug, *Digital Audio Processing*, ISBN 0-879-30-566-5, CMP Media, Lawrence, KS, 2000.

AUTHOR'S BIOGRAPHY

Technical Editor Brian Dipert experiences surround sound every time the four dogs lying around him in his office sense that the mailman's at the front door. You can reach Brian at 1-916-454-5242, fax 1-916-454-5101, bdipert@pacbell.net, and see his four-legged children's pictures at www. bdipert.com.

EARS-ON ANALYSIS

After reading this article, you may get a sense that I'm a surround-sound enthusiast. Well, you're right. I'm surrounded by surround sound in my home, office, and car.

My living room unfortunately and typically has less-than-ideal acoustics. It *does* have a 5.1-channel speaker setup, comprising Dahlquist front speakers, a KLH (www.klhaudio.com) center and direct radiating rear surround set, and an AudioSource (www.audiosource.net) SW Fifteen subwoofer. Driving the speakers is a Technics (www.panasonic.com) SA-DX1050 audio/video receiver, supporting Dolby Digital and DTS decoding, and connected to a Toshiba (www.toshiba.com) SD-2108 DVD player (with Spatializer Labs' (www.spatializer.com) N-2-2 virtualization) and Mitsubishi (www.mitsubishi-tv.com) SR-HD5 HDTV receiver via Belkin (www.belkin.com) coaxial and optical cabling.

The four desktop PCs in my house contain Analog Devices' (www.analog.com), now-defunct Aureal, Creative Labs (www.creative.com), and Zoltrix (www.zoltrix.com) audio subsystems, each providing virtual-surround capabilities. My Labtec (www.labtec.com) LCS-2414 speakers include a "3-D-stereo" adjustment knob, and I've installed a variety of audio-enhancement plug-ins for MusicMatch Jukebox, Real Player, Real Jukebox, WinAmp, and the Windows Media Player. Many of the PCs' DVD software packages support Dolby Headphone, all can virtualize Dolby Digital and Dolby Stereo, and InterVideo's (www.intervideo.com) WinDVD 3.0 even decodes DTS. QSound's (www.qsound.com) UltraQ and SRS Labs' (www.srslabs.com) WOW Thing deliver audio enhancement in hardware.

In my car, I have a 4.1-channel speaker system. I have no desire to cut into the dashboard; therefore, I have no center speaker. I also have a Sherwood (www.sherwoodusa.com) XCM-7370 AM/FM radio and CD player, X-DTS80 audio decoder, and an AX-6275 multichannel amplifier. The X-DTS80 decodes DTS audio CDs, and it also supports SRS Labs' CircleSurround algorithm for one- and two-channel audio sources. It *almost* makes me wish I had a long commute to work.