3D Audio and Acoustic Environment Modeling

William G. Gardner, Ph.D.

Wave Arts, Inc.
99 Massachusetts Avenue, Suite 7
Arlington, MA 02474
billg@wavearts.com
http://www.wavearts.com

March 15, 1999

1. INTRODUCTION

Recently there has been a proliferation of 3D audio technologies intended for desktop computers. Many sound cards, multimedia speakers, video games, audio software, and CD-ROMs are marketed as having some sort of 3D capability. In addition, a new technology called acoustic environment modeling has emerged which combines basic 3D technology with reverberation and other effects in order to simulate natural acoustic scenes.

This paper describes the 3D audio and acoustic environment modeling technology developed by Wave Arts, Inc. Wave Arts technology is the result of extensive research and development by Bill Gardner, a graduate of the MIT Media Lab. Dr. Gardner's research at the Media Lab focussed on the key technologies of 3D audio: binaural synthesis [8,11,13], crosstalk cancellation [9,11-13], and reverberation algorithms [4,5,7,10]. The Wave Arts 3D technology has recently been incorporated into the InMotion 3D Audio Producer software, developed jointly with Human Machine Interfaces, Inc.

This paper both gives a tutorial on 3D audio, and also describes particular implementation details of the Wave Arts 3D technology.

2. WHAT IS 3D AUDIO?

A 3D audio system has the ability to position sounds all around a listener. The sounds are actually created by the loudspeakers (or headphones), but the listener's perception is that the sounds come from arbitrary points in space. This is similar to stereo panning in conventional stereo systems: sounds can be panned to locations between the two loudspeakers, creating virtual or "phantom" images of the sound where there is no loudspeaker. However, conventional stereo systems generally cannot position sounds to the sides or rear of the listener, nor above or below the listener. A 3D audio system attempts to do just that.

A lot of commercial audio products are described as having 3D capability, but in fact there is great disparity between the various technologies in use. Unfortunately, many of the weakest products are marketed with the most exaggerated claims. For example, a number of stereo multimedia speakers are marketed as having "3D" technology. These speakers incorporate a simple circuit that has the effect of widening the perceived soundfield of a stereo recording. That is, the sound images that would normally extend to the locations of the left and right speakers are widened to extend beyond the speakers. These systems should more properly be called stereo enhancement or "widening" systems. They have no ability to position individual sounds around a listener, nor do they have the ability to position sounds behind, above, or below the listener. We use the term 3D audio to describe a much more sophisticated system than can ideally position sounds anywhere around a listener.

Even within the field of what we would consider to be true 3D technology, there is a wide range of technologies in use, with corresponding variation in the performance and cost of the products. Compounding this is the fact that even the best technologies are subject to unavoidable limitations that guarantee that the performance of 3D audio will always fall a bit short of the marketing claims. This has created some confusion in the marketplace. Nevertheless, 3D technology is rapidly improving, thanks in part to the increasing availability of inexpensive computational power.

2.1. How does 3D Audio Work?

To answer how 3D audio systems work, it is useful to start by considering how humans can localize sounds using only two ears. A sound generated in space creates a sound wave that propagates to the ears of the listener. When the sound is to the left of the listener, the sound reaches the left ear before the right ear, and thus the right ear signal is delayed with respect to the left ear signal. In addition, the right ear signal will be attenuated because of "shadowing" by the head. Both ear signals are also subject to a complicated filtering process caused by acoustical interaction with the torso, head, and in particular, the pinna (external ear). The various folds in the pinna modify the frequency content of the signals, reinforcing some frequencies and attenuating others, in a manner that depends on the direction of the incident sound. Thus an ear acts like a complicated tone control that is direction dependent. We unconsciously use the time delay, amplitude difference, and tonal information at each ear to determine the location of the sound. These indicators are called sound localization "cues". Sound localization by human listeners has been studied extensively [2,14].

The transformation of sound from a point in space to the ear canal can be measured accurately; the measurements are called head-related transfer functions (HRTFs). The measurements are usually made by inserting miniature microphones into the ear canals of a human subject or a manikin. A measurement signal is played by a loudspeaker and recorded by the microphones. The recorded signals are then processed by a computer to derive a pair of HRTFs (for the left and right ears) corresponding to the sound source location. This process is diagrammed in figure 1. Each HRTF, typically consisting of several hundred numbers, describes the time delay, amplitude, and tonal transformation for the particular sound source location to the left or right ear of the subject. The measurement procedure is repeated for many locations of the sound source relative to the head, resulting in a database of hundreds of HRTFs that describe the sound transformation characteristics of a particular head.

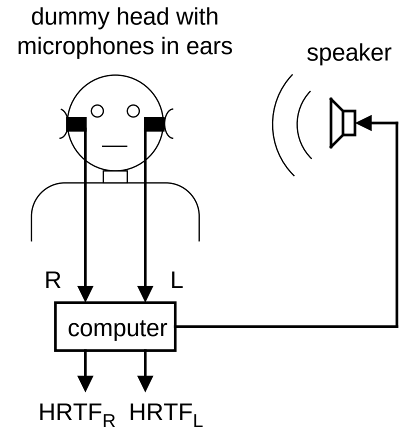

Figure 1. Measurement of HRTFs.

A 3D audio system works by mimicking the process of natural hearing, essentially reproducing the sound localization cues at the ears of the listener. This is most easily done by using a pair of measured HRTFs as a specification for a pair of digital audio filters (equalizers). When a sound signal is processed by the digital filters and listened to over headphones, the sound localization cues for each ear are reproduced, and the listener should perceive the sound at the location specified by the HRTFs. This process is called binaural synthesis (binaural signals are defined as the signals at the ears of a listener). The binaural synthesis process is diagrammed in figure 2.

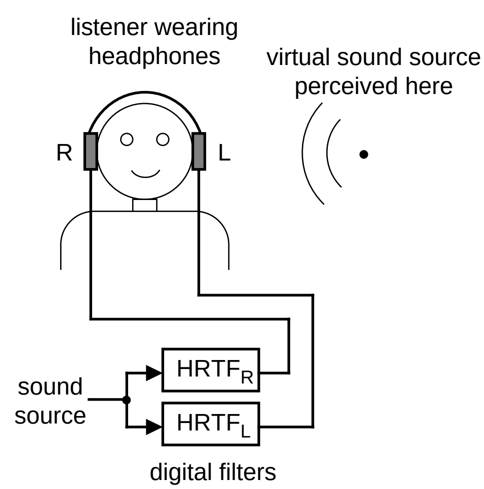

Figure 2. Binaural synthesis using HRTFs.

Binaural synthesis works extremely well when the listener's own HRTFs are used to synthesize the localization cues [18-19]. However, measuring HRTFs is a complicated procedure, so 3D audio systems typically use a single set of HRTFs previously measured from a particular human or manikin subject.

Localization performance generally suffers when a listener listens to directional cues synthesized from HRTFs measured from a different head [17], called non-individualized HRTFs. Human heads are all different sizes and shapes, and there is also great variation in the size and shape of individual pinna. This means that every individual has a different set of directional cues. The greatest differences are in the tonal transformations at high frequencies caused by the pinna. It is clear we become accustomed to localizing with our own ears, and thus our localization abilities are diminished when listening through another person's ears. Our uniqueness as individuals is the source of the greatest limitation of 3D technology.

The use of non-individualized HRTFs results in two particular kinds of localization errors commonly seen with 3D audio systems: front/back confusions and elevation errors [17]. A front/back confusion results when the listener perceives the sound to be in the front when it should be in back, and vice-versa. When 3D audio is reproduced over frontal loudspeakers, back to front confusions tend to be common, which simply means that some listeners may not be able to perceive sounds as being in the rear. In practice, this means that when panning a sound from the front, around to the side, and to the rear, the result will be perceived as a sound panning to the side and then back to the front.

Elevation errors are also common with 3D audio systems. In practice, when a sound is moved from directly to the right to directly overhead, this may be perceived as though the sound is moving from the right to directly in front. This is a typical manifestation of elevation errors, commonly observed when using loudspeakers. Elevation performance is much better when using headphones than when using loudspeakers because the high frequency cues are more faithfully reproduced.

The HRTFs used by Wave Arts 3D were measured from a Knowles Electronic Manikin for Acoustic Research (KEMAR) [8]. The measurements were made in MIT's anechoic chamber. The KEMAR is an anthropomorphic manikin whose dimensions were designed to equal those of a median human. The pinna used were molded from human pinna. In total, 710 measurements were made at different locations around the KEMAR. When synthesizing a location that is not in the measured set, HRTFs from four adjacent locations are interpolated.

2.2. How does 3D audio work over loudspeakers?

When reproducing localization cues to a listener, it is important that the left and right audio channels remain separated, that is, the left ear signal should go to the listener's left ear only, and the right ear signal should go to the listener's right ear only. This is easy to achieve when the listener is using headphones. When using loudspeakers, however, there is significant "crosstalk" between each speaker and the opposite ear of the listener. A large portion of the left speaker signal will go to the right ear of the listener, and similarly a large portion of the right speaker signal will go to the left ear of the listener. In figure 3, the crosstalk paths are labeled A_{LR} and A_{RL}. The crosstalk severely degrades localization performance and must be eliminated.

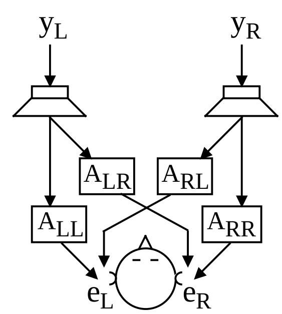

Figure 3. Direct and crosstalk transmission paths from loudspeakers to the ears of a listener.

Fortunately, it is possible to build an elaborate digital filter, called a "crosstalk canceller," that eliminates crosstalk [11-13]. The crosstalk canceller adds a cancellation signal to each of the two channels of audio, such that when the listener is properly positioned between the loudspeakers, the crosstalk is acoustically cancelled at the listener's ears. The listener must be centered between the two loudspeakers in order for the crosstalk to be cancelled. In 3D audio parlance, the listener must be in the "sweet spot" to get the full 3D effect. Provided the listener is centered between the loudspeakers, crosstalk cancellation is relatively insensitive to front-back motions of the listener, however, crosstalk cancellation is degraded when the listener is off-center or not facing forward.

Loudspeaker 3D audio systems are extremely effective in desktop computing environments. This is because there is usually only a single listener (the computer user) who is almost always centered between the speakers and facing forward towards the monitor. Thus, the primary user gets the full 3D effect because the crosstalk is properly cancelled. In typical 3D audio applications, like video gaming, friends may gather around to watch. In this case, the best 3D audio effects are heard by others when they are also centered with respect to the loudspeakers. Off-center listeners may not get the full effect, but they still hear a high quality stereo program with some spatial enhancements.

Many crosstalk cancellers are based on a highly simplified model of crosstalk, for example modeling crosstalk as a simple delay and attenuation process, or a delay and a lowpass filter (book, trans 3D audio). Other crosstalk cancellers have been based on a spherical head model [3]. The crosstalk canceller used by Wave Arts 3D audio is based on actual HRTF measurements and thus accurately models the crosstalk that occurs with human listeners. For typical human listeners, the Wave Arts crosstalk canceller improves channel separation by about 20 dB in the 100 Hz to 6 kHz range. This may seem like a modest improvement, but in fact it is quite good. Even a small improvement in channel separation leads to a large improvement in localization performance. As with binaural synthesis, crosstalk cancellation performance is ultimately limited by the variation in the size and shape of human heads.

3. ACOUSTIC ENVIRONMENT MODELING

Acoustic environment modeling refers to combining 3D spatial location cues with distance, motion, and ambience cues, to create a complete simulation of an acoustic scene. By simulating the acoustical interactions that occur in the natural world, we can achieve stunningly realistic recreations, above and beyond that possible with just 3D positional control [1]. The Wave Arts Acoustic Environment Modeling system combines Wave Arts 3D with accurate simulations of the following acoustic phenomena: reverberation, distance cues, Doppler motion effect, air absorption, and object occlusion. These phenomena are described in the following sections.

3.1. Reverberation

When an object in a room produces a sound, a soundwave expands outward from the source reaching walls and other objects where sound energy is both absorbed and reflected. Technically speaking, all reflected energy is called reverberation. Assuming a direct path exists between the source and the listener, the listener will first hear the direct sound, followed by reflections off nearby surfaces, called early reflections. After a few tenths of a second, the number of reflected waves becomes very large, and the resulting reverberation is characterized by a dense collection of soundwaves travelling in all directions, called diffuse reverberation. The time required for the reverberation to decay 60 dB below the initial level is defined as the reverberation time. Generally, reverberation in a small room decays much faster than reverberation in a large room, because in a small room the soundwaves collide with walls much more frequently, and thus are absorbed more quickly, than in a large room.

Reverberation is an important acoustic phenomena. There is at most one direct path from the source to the listener, whereas there may be millions of indirect paths, particularly in a room where a sound can bounce around hundreds of times before being absorbed. Thus, in typical listening situations, most of the energy we hear from a sound source is actually reflected energy.

The perception of reverberation depends on the type of reverberation and the type of sound. In small room with fast decaying reverberation, the reverberation imparts a tonal quality to the sound that is readily identified as a small room signature. In a larger room, the reverberation can create a background ambience that is easily distinguished from the foreground sound, and this is readily identified as a characteristic of large spaces. In this manner, reverberation imparts useful spatial information about the size of the surrounding space.

Reverberation that contains a lot of high frequency energy in the decay is associated with rooms that have hard, reflective walls, which do not readily absorb high frequencies. Similarly, reverberation that is dull sounding is associated with rooms that contain soft materials, such as plush carpets and drapes, which readily absorb high frequencies. In this manner, reverberation imparts useful information about the composition of the surrounding space.

Reverberation is also important for establishing distance cues. In a reverberant space, when the distance between the source and the listener is increased, the level of the direct sound decreases considerably, but the level of reverberation does not decrease much. Thus, the level of direct to reverberant sound can be used as a distance cue, with dry (non-reverberant) sounds perceived as being close, and reverberant sounds perceived as being distant.

Simulating reverberation is essential for establishing the spatial context of an auditory scene. Reverberation gives information about the size and character of the surrounding space, it is very useful for correctly perceiving distances, and it adds greatly to the realism of the simulation.

3.2. Reverberation algorithm

In acoustic environment modeling systems, reverberation is often simulated by considering a simple geometrical model of the simulated space. Based on the positions of the source, listener, and the reflective surfaces (walls, floor, ceiling), it is easy to use a ray tracing procedure to calculate the time and direction of all early reflections. Each reflection can then be rendered using (1) a delay line to delay the sound according to the total travel time along the reflected path, (2) an attenuation or filter to approximate the transmission and reflection losses, and (3) a binaural synthesizer to properly spatialize the reflection [10,16]. This method is theoretically justified by acoustics, but it is computationally expensive. Furthermore, it is doubtful that the early portion of the reverberation needs to be modeled so accurately.

The early reflection model does not address the late portion of the reverberation, which contains millions of reflections travelling in all directions. Alternative methods must be used to generate the late reverberation. Late reverberation is usually generated using recursive filters (filters that have feedback elements) such as comb and allpass filters. Other recursive filter topologies have been proposed for rendering reverberation [10], including allpass feedback loops, feedback delay networks [15], and waveguide reverberators. The challenge with reverberation algorithm design is to produce a natural sounding reverberation without excessive coloration in the late decay.

The approach we have taken is to render a generic reverberation which provides both a natural pattern of early reflections and a natural late reverberation. The reverberator is implemented using an allpass feedback loop topology [4,5,10]. The character of the reverberation is controlled by several independent parameters, which include the reverberation time, room size, and damping. The independence of the room size and reverberation time is important to achieve a varied number of reverberation effects. The room size parameter alters both the pattern of early reflections and the character of the late reverberation to simulate various room sizes. This parameter is particularly effective. The damping frequency parameter controls the absorption of high frequencies in the late reverberation; high damping frequencies result in a bright sounding room, low damping frequencies result in a warm sounding room.

The reverberator is specifically designed to process binaural input signals, so that the early reflections will be localized near the sound source, whereas the late reverberation is spatially diffuse. This has proven to be an efficient way to generate realistic early reflections. The resulting reverberator has many properties that make it effective for use in acoustic environment modeling:

- Localization of early reflections depends on location of source.
- Early reflections and late reverberation depend on room size.
- Independent control of reverberation time, room size, and damping frequency.
- Natural colorless decay.
- Spatially diffuse late reverberation.

3.3. Distance Cues

The principal cue for distance is the loudness of the sound. A sound source will be louder when it is closer to the listener than when it is farther away. However, this cue is often ambiguous because the listener doesn't know a priori how loud the source is. Thus, a moderately loud crashing sound could be perceived as a quiet, close crash, or a distant, loud crash.

Another important cue for distance is the relative loudness of reverberation. When sound is produced in a reverberant space, the associated reverberation may often be perceived as a background ambience, separate from the foreground sound [6]. The loudness of the reverberation relative to the loudness of the foreground sound is an important distance cue. The reason for this is due to the acoustics of reverberant spaces. The foreground sound consists largely of the sound that propagates directly from the sound source to the listener, this so-called direct sound decreases in amplitude as the distance to the listener increases. For every doubling of distance, the amplitude of the direct sound decreases by a factor of one half, or 6 dB. The amplitude of the reverberation, on the hand, does not decrease considerably with increasing distance. The ratio of the direct to reverberant amplitude is greater with nearby objects than it is with distant objects. Thus, distant objects sound more reverberant than close objects.

This relationship is diagrammed in figure 4. The direct sound amplitude drops 6 dB for each doubling of distance (equal to 20 dB drop for a factor of 10 increase in distance). The reverberation amplitude shown below drops at half this slope, or 3 dB per doubling of distance (equal to 10 dB drop for a factor of 10 increase in distance). In most reverberant spaces, the reverberation does not actually drop this fast with increasing distance. However, for the purposes of creating an effective sounding scene, it is often necessary to tweak the parameters to get the desired effect. In particular, when synthesizing virtual acoustic scenes, it can sound unnatural if the reverberation doesn't attenuate sufficiently with increasing distance. It also becomes difficult to localize the sound source if there is too much reverberation.

The relationship between direct and reverberant sound shown in figure 4 is the default distance model used by Wave Arts Acoustic Environment Modeling. For very close distances, the reverberation is 20 dB below the direct sound, equal to a 10% reverb mix. For increasing distances, the ratio of direct sound to reverberation decreases, and at 100 feet the reverberation is louder than the direct sound. This model is not physically accurate, but produces good sounding results.

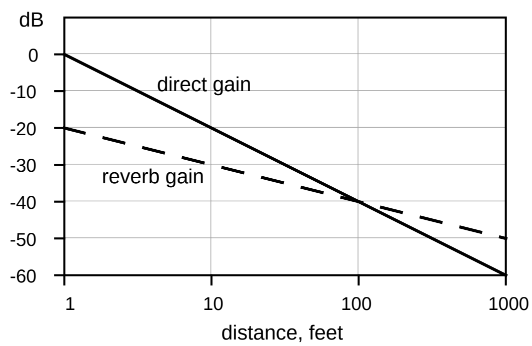

Figure 4. Default distance model.

3.4. Doppler motion effect

The Doppler motion effect is commonly heard in nature as a pitch change when a speeding object passes a listener. When the object is approaching the listener, the pitch is higher than the resting pitch of the object. This is because in the time it takes the object to emit one waveform the object has moved closer to the listener, and thus the emitted wavelength is shorter than normal. Similarly, when the object is retreating from the listener, the pitch is lower than the resting pitch, because the emitted wavelengths are longer than normal.

Simulating the Doppler effect is important for generating realistic motion effects. The Doppler motion effect is particularly easy to simulate using a variable delay line. The amount of delay is proportional to the distance between the listener and the sound object. Thus, the delay line effectively simulates the propagation of sound through the air. When the distance changes, so does the length of the delay, and the pitch also changes as it would in nature. Care must be taken that to change the delays smoothly and continuously to avoid distortion and clicks.

3.5. Air absorption

When sound propagates through air, some sound energy is absorbed in the air itself. The amount of energy loss depends on the frequency of the sound and atmospheric conditions. High frequencies are more readily absorbed than low frequencies, so the high frequencies are reduced with increasing distance. For example, at 100 meters distance, 20 degrees Celsius, and 20% humidity, a 4 kHz tone will be attenuated by about 7.4 dB [1]. However, the attenuation would be less than 1 dB for distances less than 10 meters. The effect can be simulated by a lowpass filter whose cutoff frequency depends on the distance to the source.

3.6. Object Occlusion

When a sound source is behind an occluding object, the direct path sound must diffract (bend) around the occluding object to reach the listener. Low frequencies with wavelengths larger than the size of the occluding object will not be affected much by the occluding object. High frequencies with wavelengths smaller than the size of the occluding object will be shadowed by the object, and will be greatly attenuated. Thus, the effect of an occluding object can be simulated by a lowpass filter whose cutoff frequency depends on the size of the occluding object. Simulating object occlusion is important to achieve realism in film/video soundtracks where sound emitting objects are visibly moving behind occluding objects.

4. SIGNAL ROUTING

The Wave Arts Acoustic Environmental Modeling system is implemented using the signal routing shown in the figure below. The signal routing is conceptually similar to the routing seen in multichannel mixing consoles: input signals are individually processed, mixed to a set of shared signal busses, and then the bus signals are processed and output. In the figure, the input signals shown at the top represent the individual object sounds that are to be spatially processed to create the scene. The input signals are monophonic.

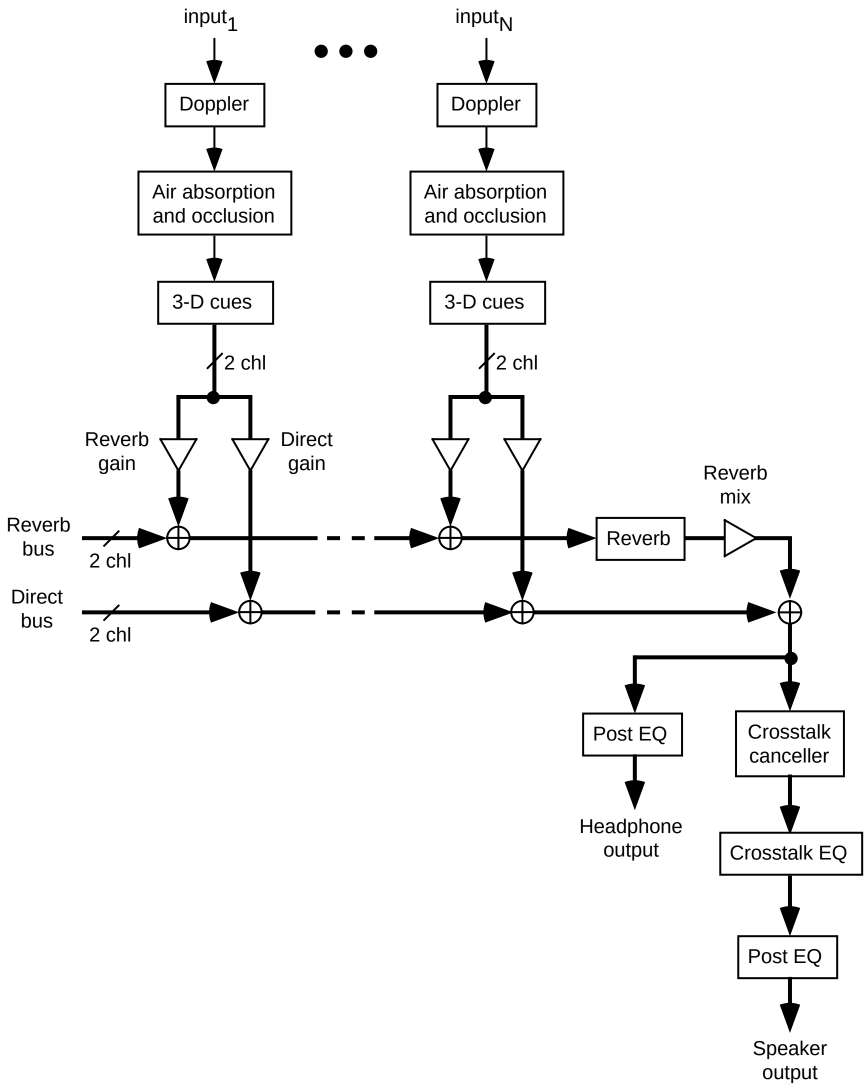

Figure 5. Signal routing used in Wave Arts Acoustic Environment Modeling.

Each input signal is processed through the Doppler effect, then the air absorption and occlusion effect, and then the 3D spatial effect, labeled “3D cues” in the figure. The Doppler effect and air absorption effect are controlled by the distance between the sound object and the listener. The occlusion effect is controlled by the position of the sound object, which determines the degree to which the sound object is occluded, and the dimensions of the occluding objects. The 3D spatial effect is controlled by the position of the sound object relative to the listener. The 3D spatial effect creates stereophonic (two channel) output. The figure does not show the individual left and right channels output from the 3D spatial processor; instead the stereo signals are drawn with a thick line and labeled “2 chl.”

The output from the 3D spatial processor is split into two stereo signals, which are mixed to the “reverb bus” and the “direct bus,” each of which is a stereo bus. The amount of sound mixed to each bus depends on the “reverb gain” and “direct gain” mixing gains. These gains are controlled by the distance from the sound to the listener according to the current distance model. Typically, the distance model parameters are set up so that the direct to reverberant ratio increases as the sound object distance decreases.

The reverb bus contains a mix of all sounds that are to be sent to the reverberator. These are processed by the reverberator and the result is mixed with the direct bus. The reverb mix gain determines the overall level of reverberation in the scene. The reverberator is controlled by the scene environment parameters, which include the reverb time, room size, damping, etc.

For playback over loudspeakers, the direct bus must be further processed by the crosstalk canceller. The crosstalk canceller is controlled by the speaker angle parameter. The output of the crosstalk canceller is processed by the crosstalk equalization stage, and this signal is further processed by a set of tone controls labeled "Post EQ" and the result is output to the speakers.

4.1. Virtual speakers

The Wave Arts 3D Audio system provides a special type of sound object called a virtual speaker. Virtual speakers are intended to simplify the management of virtual surround processing. Using virtual speakers it is easy to convert a conventional stereo sound into an immersive 3D sound. This can be done by assigning the left channel input to a virtual left speaker positioned far to the left of the actual left speaker, and similarly assigning the right channel input to a virtual right speaker positioned far to the right of the actual right speaker. Another application is to convert 5.1 surround into 3D stereo by setting up virtual left and right surround speakers as well as virtual left, right, and center speakers.

A virtual speaker is like a stationary sound object; it is fixed in space and assigned a sound. Unlike a stationary sound object, however, a virtual speaker is not subject to environmental effects. Thus, the Doppler effect, air absorption, object occlusion, and reverberation have no effect on a virtual speaker. Only the angle and distance of the virtual speaker with respect to the listener is important; these are used to synthesize the 3D location and amplitude of the virtual speaker.

Instead of environmental effects, virtual speakers implement a variable delay line, gain control, and a user-defined filter to permit customization of the signal that feeds each virtual speaker. For example, using the bandpass filters, one can easily set up virtual speakers that reproduce only certain frequency ranges of the input sound; these virtual speakers can then be positioned anywhere around the listener. This makes it easy to create pseudo-surround mixes from conventional stereo inputs.

The direct bus output is suitable for listening to over headphones. The headphone output is simply the direct bus processed through a set of tone controls labeled "Post EQ".

The processing for a virtual speaker is shown in the figure below. The input sound is processed through a user-adjustable variable delay and a user-adjustable filter. The filter may be a bandpass, lowpass, highpass, or notch (bandstop) filter. The delayed and filtered signal is then passed to the 3D spatial effect to position the virtual speaker. The output of the 3D effect is summed to the direct bus through a gain that depends on the distance between the virtual speaker and the listener according to the current distance model.

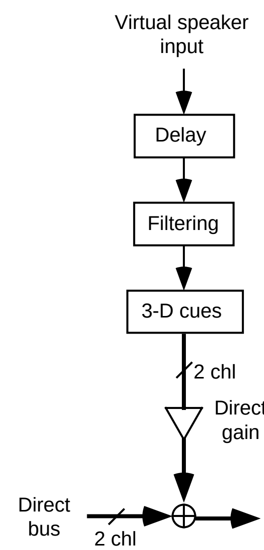

Figure 6. Virtual speaker input processing.

4.2. Channel-assigned virtual speakers

Virtual speakers can also be assigned directly to the output channels. In this case, the virtual speaker's location is not synthesized by applying 3D spatial cues; instead, the input sound is routed directly to the stereo output channels. These virtual speakers are useful for mixing original stereo program material into a 3D scene.

Virtual speaker can be directly assigned to the left, right, or center channels. .A virtual speaker that is directly assigned to the left, right, or center channels is implemented differently than in figure 6. Instead of being processed with the 3D spatial effect, the delayed and filtered signal is summed directly to the appropriate output channel. In the case of a center assignment, the signal is summed equally to the left and right channels with a gain of sqrt(2)/2.

5. SUMMARY

This paper has given a tutorial on 3D audio technology for headphones and loudspeakers, and has discussed the additional effects needed to model acoustic environments. Specific implementation details of the Wave Arts 3D Audio and Acoustic Environment Modeling technology have been described. With the rapid advances in computational power, and the effectiveness of desktop 3D loudspeaker audio, we expect that this technology will be used increasingly in desktop multimedia applications.

6. REFERENCES

[1] Begault, D. R. (1994). *3-D Sound for Virtual Reality and Multimedia*, Academic Press, Cambridge, MA.

[2] Blauert, J. (1983). *Spatial Hearing*, MIT Press, Cambridge, MA.

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

[4] Gardner, W. G. (1992). "The Virtual Acoustic Room", Master's thesis, Dept. of Media Arts and Sciences, MIT.

[5] Gardner, W. G. (1992). "A realtime multichannel room simulator", *J. Acoust. Soc. Am*., 92(4), pp. 2395. http://sound.media.mit.edu/papers.html

[6] Gardner, W. G., and D. Griesinger (1994). "Reverberation level matching experiments", *Proceedings of the Sabine Centennial Symposium*, Cambridge, MA.

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

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

[9] Gardner, W. G. (1995). "Transaural 3-D audio", M.I.T. Media Lab Perceptual Computing Section. http://sound.media.mit.edu/papers.html.

[10] Gardner, W. G. (1998). "Reverberation Algorithms", in *Applications of Digital Signal Processing to Audio and Acoustics,* Kahrs, M., and K. Brandenburg, Ed., Kluwer Academic, Norwell, MA.

[11] Gardner, W. G. (1997). "3-D Audio Using Loudspeakers", Ph.D. thesis, Dept. of Media Arts and Sciences, MIT.

[12] Gardner, W. G. (1997). "Head Tracked 3-D Audio Using Loudspeakers", *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY.

[13] Gardner, W. G. (1998). *3-D Audio Using Loudspeakers*, Kluwer Academic, Norwell, MA.

[14] Gilkey, R. H., and T. R. Anderson, Ed (1997). *Binaural and Spatial Hearing in Real and Virtual Environments*, Lawrence Erlbaum Associates, Mahwah, NJ.

[15] Jot, J. M., and A. Chaigne (1991). "Digital delay networks for designing artificial reverberators", *Proc. Audio Eng. Soc. Conv.*, Preprint 3030.

[16] Kendall, G. S., W. L. Martens, and M. D. Wilde (1990). "A spatial sound processor for loudspeaker and headphone reproduction", *Proc. Audio Eng. Soc. 8th Int. Conf.*

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

[18] Wightman, F. L., and D. J. Kistler (1989). "Headphone simulation of free-field listening. I: Stimulus synthesis", *J. Acoust. Soc. Am.*, 85(2), pp. 858-867.

[19] Wightman, F. L., and D. J. Kistler (1989). "Headphone simulation of free-field listening. II: Psychophysical validation", *J. Acoust. Soc. Am.*, 85(2), pp. 868-878.