11 Sound Spatialization and Reverberation

Modeling Sound Spaces

Conclusion

The art of sound spatialization has assumed a similar position today as the art of orchestration had in the nineteenth century. To deploy space is to choreograph sound: positioning sources and animating movement. Immersing sound in reverberation, we bathe listeners in its lush ambience.

Sound spatialization has two aspects: the virtual and the physical. In the virtual reality of the studio, composers spatialize sounds by imposing delays, filters, panning, and reverberation—lending the illusion of sounds emerging from imaginary environments. Sometimes these virtual spaces take on characteristics that would be impossible to realize architecturally, such as a continuously changing echo pattern. In the physical world of concert halls, sounds can be projected over a multichannel sound system from a variety of positions: around, above, below, and within the audience.

Sound architecture or spatialization has evolved into an important aspect of composition. A trend toward "cinematic" use of space is seen in compositions that feature dramatic appositions between sounds that are closely miked and those that are distantly reverberated. Some composers use microphone techniques and spatial processing in a manner similar to the cinematic use of camera angle, lens perspective (width), and depth of field. Jean-Claude Risset's *Sud* (1985, Wergo 2013-50) comes to mind.

This chapter opens with a glimpse at the projection of sound in three-dimensional space. The second part describes the art of digital reverberation, an area of spatialization that will continue to be refined in years to come. The final section extends the discussion of the first two sections by surveying research that attempts to model specific spatial environments. A recommended prerequisite to this chapter is a familiarity with the concepts of filtering introduced in chapters 5 and 10.

Sound Spatialization

The movement of sound through space creates dramatic effects and can serve as an important structural element in composition. Composers can articulate the voices in a contrapuntal texture by giving each one a unique spatial location. The virtual and physical *sound stage* around the audience can be treated as a landscape, with its background and foreground, and fixed and moving sources. This sound stage can be fixed in playback, or controlled by gestures in concert (Harada et al. 1992).

Digital simulations of moving sound sources pose special problems. In many concerts the audience is surrounded by a number of loudspeakers. How does one create the illusion of a sound traveling about the hall, moving away from or toward the listener as it goes? In listening situations with only two loudspeakers or with headphones, the illusion of sounds moving freely in space is even more difficult.

The most popular spatial illusions are horizontal *panning*—lateral sound movement from speaker to speaker—and *reverberating*—adding a dense and diffuse pattern of echoes to a sound to situate it in a larger space. Vertical panning (up and down and overhead) can also create striking effects in electronic music. (See Gerzon 1973 for a discussion of "sound-with-height" recording and playback.)

Spatialization in Music: Background

Von welcher Seite, mit wievielen Lautspechern zugleich, ob mit Links- oder Rechtsdrehung, teilweise starr und teilweise beweglich die Klänge und Klanggruppen in den Raum gestrahlt werden: das alles ist für das Verständnis dieses Werkes massgeblich. (From which side, with how many loudspeakers, whether with rotation to left or right, whether motionless or moving, how the sounds and sound groups should be projected into space: all this is decisive for the understanding of the work.) (Karlheinz Stockhausen 1958, describing his composition *Gesang der Jünglinge* [*Song of the Youths*])

Spatial techniques in music composition are not new. In the sixteenth century, composers associated with the Basilica San Marco in Venice (notably Adrian Willaert and his pupil Andrea Gabrieli) employed spatial antiphony in their compositions for two or more choirs. In these works, an initial verse was heard from one side of a hall, and a response verse came from another side. This arrangement was facilitated by two facing organs in the basilica. W. A. Mozart wrote compositions for two spatially separated orchestras (K. 239 and K. 286), and Hector Berlioz and Gustav Mahler wrote compositions for multiple orchestras and choruses, some of which were offstage. After these experiments, however, there is little documentation of spatial techniques in composition until the electronic era.

The invention of the loudspeaker could be compared to the invention of the light bulb. Suddenly it was possible to project sonic energy in spaces small and large, at any angle or intensity. But the use of loudspeakers—in movie theaters, stadiums, railroad stations, and home radios—remained for the most part plain and functional. Only with the dawn of the post-World War 2 era were the aesthetic possibilities of sound projection via loudspeakers exploited in electronic music.

Examples of Spatialization in Electronic Music

A number of famous examples of spatial projection in electronic and computer music deserve mention here.

- Karlheinz Stockhausen's *Gesang der Jünglinge* was projected in a 1956 concert over five groups of loudspeakers in the auditorium of the West German Radio (Stockhausen 1961). His opus *Kontakte,* realized in 1960, was the first electronic music composition performed from a four-channel tape, using the Telefunken T9 tape recorder (Stockhausen 1968).
- In 1958 Edgard Varèse's classic tape music composition *Poème Electronique* and Iannis Xenakis's *Concret PH* were projected over 425 loudspeakers through an eleven-channel sound system installed on the curved walls of the Philips pavillion, designed by Xenakis and Le Corbusier at the Brussels World's Fair.
- Stockhausen played his electronic music over loudspeakers distributed on the interior surface of the geodesic dome of the German pavilion at EXPO 70 in Osaka (Stockhausen 1971a).
- At the same exposition, Iannis Xenakis performed his twelve-channel electroacoustic composition *Hibiki Hana Ma* in the Japanese Steel Pavilion on a system of 800 loudspeakers distributed around the audience, over their heads, and under their seats (Matossian 1986). A twelve-channel sound projection system animated his sound-and-light spectacle *Polytope de Cluny* projected on the interior of the ancient Cluny Museum in Paris (Xenakis 1992).
- Composer Salvatore Martirano built a complex digital apparatus called the Sal-Mar Construction to control a custom analog synthesizer and distribute the sound to 250 thin loudspeakers suspended at various heights from the ceilings of concert halls (Martirano 1971).
- The idea of projecting sound over an orchestra of dozens of loudspeakers on stage was realized in the Gmebaphone, conceived by the Groupe de Musique Expérimentale de Bourges, and first heard in concert in 1973 (Clozier 1993).
- The first concert of the Acousmonium—an assemblage of dozens of "sound projectors" conceived by the Groupe de Recherches Musicales (figure 11.1)—took place at the Espace Cardin, Paris, in 1974 (Bayle 1989, 1993).
- The steel frame used in the mid-1980s performances of Pierre Boulez's *Répons* held loudspeakers suspended over the heads of the audience. Spatial control was implemented using Di Giugno's 4X synthesizer (Asta et al. 1980; Boulez and Gerzso 1988).
- In 1987 researchers at Luciano Berio's Tempo Reale studio in Florence developed a computer-based sound distribution system called Trails that

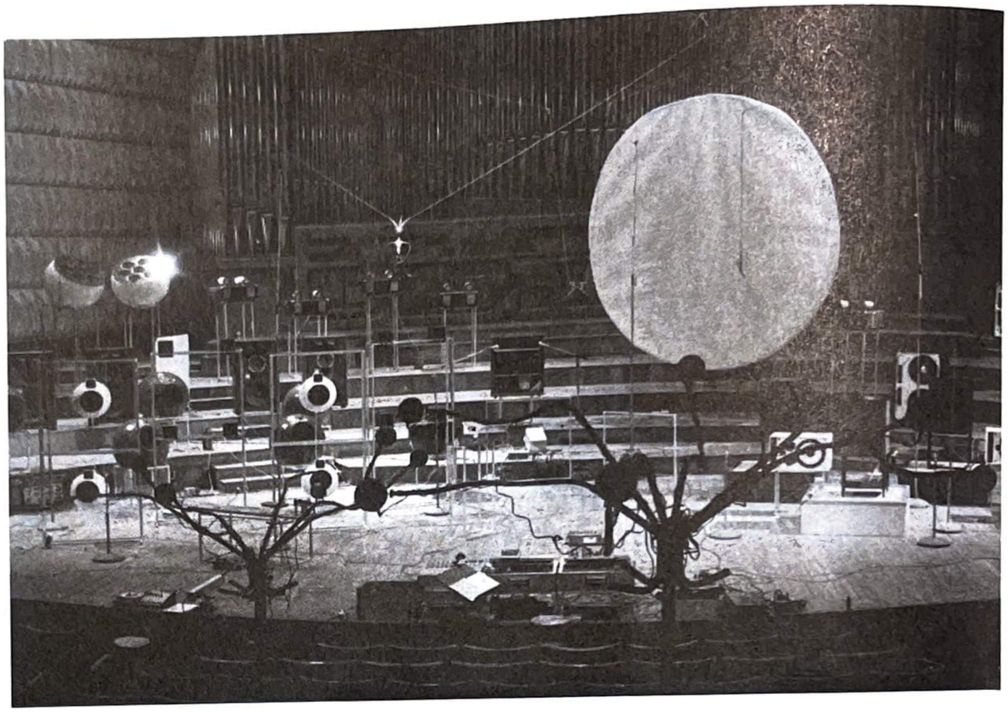


Figure 11.1 The Acousmonium—a multichannel spatializer designed by the Groupe de Recherches Musicales (GRM)—installed in Olivier Messiaen concert hall, Maison de Radio France, Paris, in 1980. Projecting sound over 80 loudspeakers played through a 48-channel mixer, the Acousmonium achieves a complexity of sound image rivaling that of an orchestra. It lets a composer "reorchestrate" an electronic composition for Acousmonium spatial performance. (Photograph by L. Ruszka and supplied courtesy of F. Bayle and the Groupe de Recherches Musicales.)

could distribute sound to up to 32 audio channels, combining preprogrammed and real-time spatial patterns (Bernardini and Otto 1989).

Many other sound spatialization systems have been developed, including Edward Kobrin's sixteen-channel HYBRID IV (Kobrin 1977) (figure 11.2), the SSSP sound distribution system (Federkow, Buxton, and Smith 1978), the AUDIUM installation (Loy 1985b), Hans Peter Haller's Halaphon used by P. Boulez and L. Nono (Haller 1980), the computer-controlled Sinfonie developed by the GRAME studio in Lyon, and the all-digital spatializer implemented by Marina Bosi (1990) at Stanford University.

Enhancing Spatial Projection in Performance

Even ad hoc concerts of electroacoustic music without elaborate sound projection systems can take steps to enhance the spatial qualities of the performance. Figure 11.3 illustrates a few standard configurations.

Figure 11.2 Edward Kobrin's HYBRID IV studio set up in Berlin, 1977, featuring a computer-controlled 16-channel spatialization system. The loudspeakers are mounted on the walls.

1. When possible, use at least a *quadraphonic* sound projection system (four channels of amplification with four loudspeaker systems), located around the audience (figure 11.3b).
2. When two-channel recordings are played on the quadraphonic system, route two channels to the front and two channels to the back with the left-right configuration in the back channels reversed. This way, when a sound pans from left to right in the front, it also pans from right to left in the back, increasing the sense of spatial animation.
3. To add even more spatial articulation, situate the loudspeakers at opposite corners in an elevated position. This is called *periphony* or "sound with height" playback (Gerzon 1973). In this scheme, when a sound pans from left to right, it also pans vertically (figure 11.3c).
4. When amplified instruments or vocalists are employed, give each performer their own amplifier and loudspeaker unit, along with effects (such as equalization) that articulate that particular instrument. To root each instrument on the sound stage and mitigate the "disembodied performer" syndrome, the loudspeaker should be near the performer (Morrill 1981b). In the disembodied performer syndrome, the sound of an instrument is

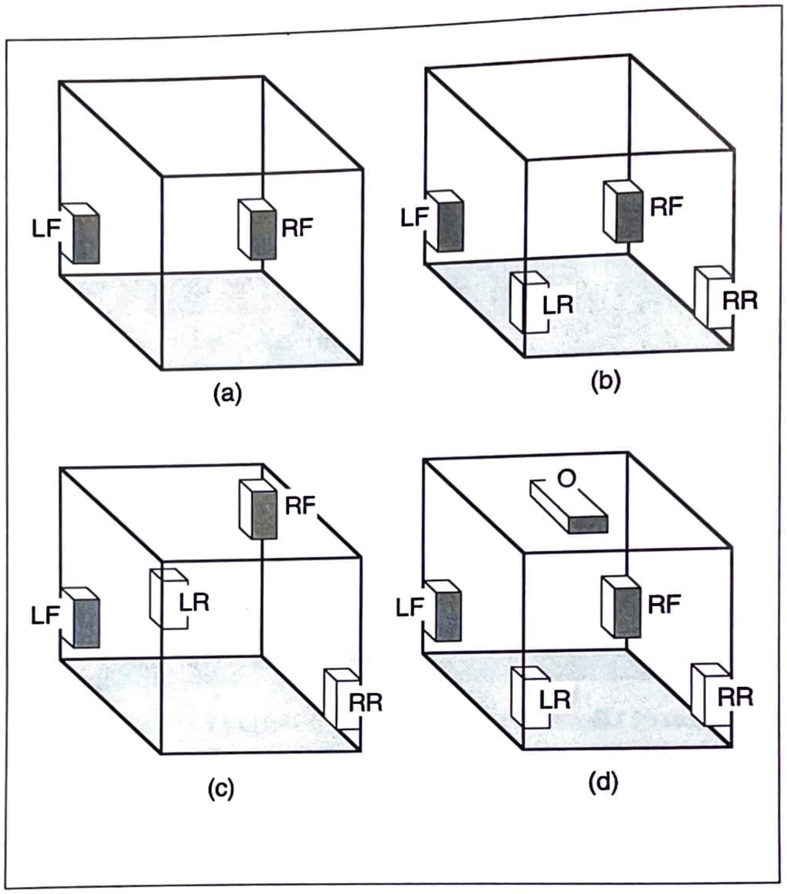


Figure 11.3 Selected loudspeaker configurations for spatialization of electronic and computer music. (*a*) Basic stereo, LF = left front, RF = right front. (*b*) Quadraphonic, RR = right rear, LR = left rear. (*c*) Quadraphonic periphony. The right front and left rear loudspeakers are mounted above ear level, so that when sound pans from horizontally it also pans vertically. (*d*) Five-speaker configuration with vertical loudspeaker projecting downward.

fed to a general sound reinforcement system that is far from the performer. Since listeners's image of the source of a sound in dominated by the first sound to reach their ears (this is the *precedence effect;* Durlach and Colburn 1978), any global amplification of a performer playing an acoustic instrument should be delayed by 5 to 40 ms to allow the local amplifier to make the first impression as to the source (Vidolin 1993). (Sometimes, of course, the composer wants to project the sound of an instrument around a hall, or to merge it with a prerecorded source; this is another case.)

5. A different approach is to assemble an "orchestra" of different loudspeakers onstage (the Gmebaphone/Acousmonium approach). This

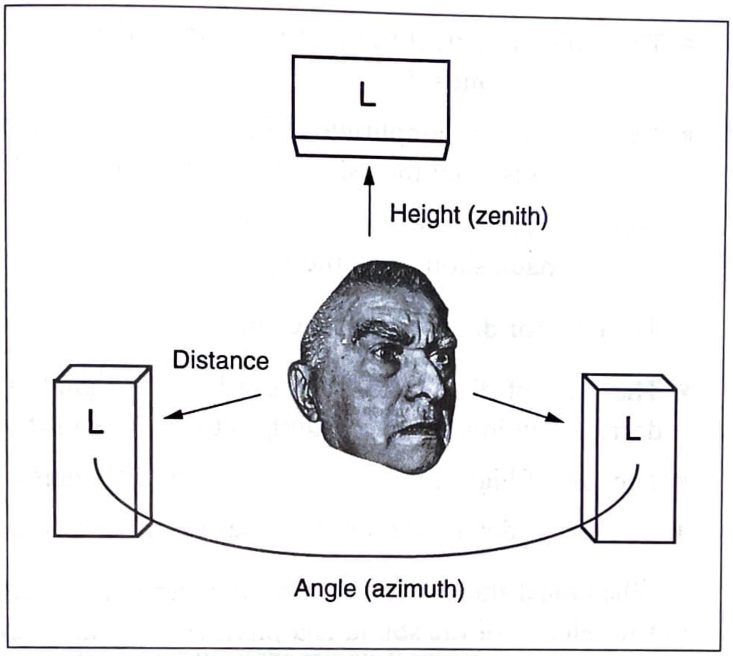


Figure 11.4 The attentive listener can localize a sound source from cues of its horizontal angle, height, and distance. L = loudspeaker.

creates a spatial source multiplicity and sonic diversity usually associated with an orchestra of acoustic instruments.

Precise control of spatial illusions requires knowledge of the *theory of localization*—how human beings perceive a sound's direction, the subject of the next section.

Localization Cues

Before delving into techniques of sound spatialization, it is important to understand basic principles of how listeners pinpoint the locale from which a sound emanates. This subject, an extensively mined area of psychoacoustics, is called *sound localization*. Localization is dependent on cues for three dimensions (figure 11.4; localization):

Azimuth or horizontal angle

Distance (for static sounds) or *velocity* (for moving sounds)

Zenith (altitude) or vertical angle

To determine the azimuth of a sound, listeners use three cues:

- The different arrival times of a sound to the two ears when the sound is coming from one side
- The difference in amplitude of high-frequency sounds heard by two ears, which results from the "shadow effect" of the head
- Spectral cues provided by asymmetrical reflections of sound off the outer ears (pinnae), shoulders, and upper torso

The cues for distance are threefold:

- The ratio of direct signal to reverberated signal, when the direct signal decreases in intensity according to the square of the distance
- The loss of high-frequency components with increasing distance
- The loss of detail (absence of softer sounds) with increasing distance

When the distance between the sound and the listener is changing, the cue to the velocity of the sound is a pitch change called the *Doppler shift effect* (explained later).

The main cue for zenith is a change in the spectrum caused by sound reflections off the pinnae and shoulders.

Simulating the Azimuth Cue

Listeners can easily localize an intense high-frequency sound coming from a particular direction at ear level. Logically enough, for a sound source to be positioned directly at a loudspeaker position, all of the signal should come from that loudspeaker. As the source pans from one loudspeaker to another, the amplitude in the direction of the target loudspeaker increases, and the amplitude in the direction of the original source loudspeaker decreases.

In performances where a number of loudspeakers are placed equidistantly in a circle around the audience, an algorithm for spatial position needs only to calculate the amplitudes of two adjacent loudspeakers at a time, regardless of the total number of loudspeakers. To position a sound source at a precise point *P* between the two loudspeakers *A* and *B*, first find the angle (θ) of the source measured from the middlepoint between *A* and *B* (figure 11.5).

Many different panning curves are possible, each of which lends a slightly different spatial impression of sound movement. We discuss two panning curves next: linear and constant power. For a symmetrical pan these curves assume that a listener sits in the exact center between the two loudspeakers.

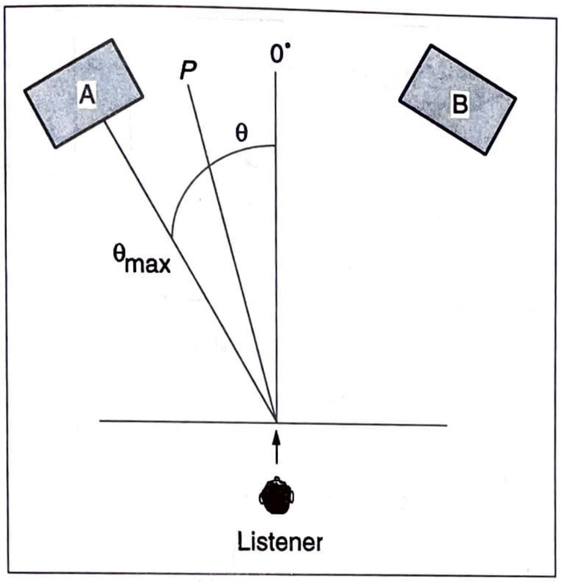


Figure 11.5. To position a sound source at a point *P* between the two loudspeakers A and B, ascertain the angle (θ) of the source measured from the middlepoint between A and B. In the middle θ equals 0 degrees. The angle θ_{max} is the maximum angle, typically plus or minus 45 degrees. Use the formulas given in the text to derive the amplitude of the signals sent to the two loudspeakers.

When the listener sits off center there is an azimuth offset in the sound image. For efficiency the curves can be computed in advance, requiring only a table-lookup operation using the index θ.

Linear Panning

The most simple formula for positioning is a simple linear relation:

$$A_{amp} = \theta/\theta_{max}$$

$$B_{amp} = 1 - (\theta\text{–}\theta_{max})$$

The problem with this type of pan is that it creates a "hole in the middle" effect since the ears tend to hear the signal as being stronger in the endpoints (the loudspeakers) than in the middle (figure 11.6). This is due to the *law of sound intensity*, which states that the perceived loudness of a sound is proportional to its intensity. The intensity of the sound can be given as follows:

$$I = \sqrt{A^2_{amp} + B^2_{amp}}.$$

In the middle of the pan (i.e., where $\theta = 0$), $A_{amp} = B_{amp} = 0.5$), this becomes

$$\sqrt{0.5^2 + 0.5^2} = \sqrt{0.25 + 0.25} = \sqrt{0.5} = 0.707.$$

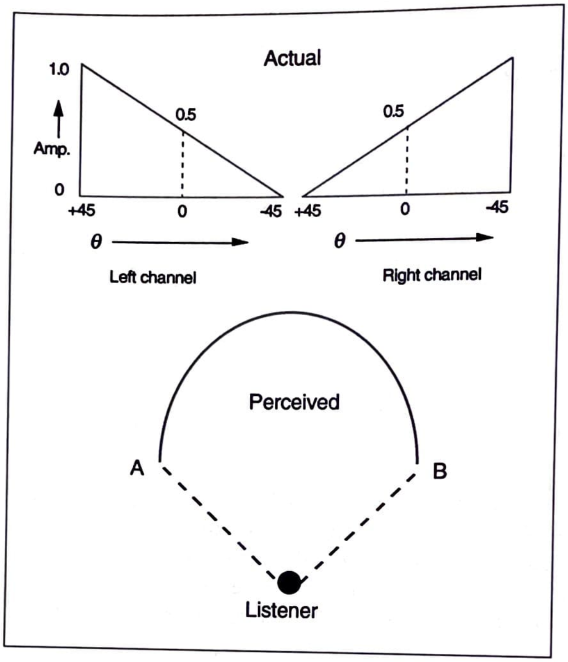


Figure 11.6 A linear panning curve is perceived as receding in the middle due to a diminution of intensity. The amplitude curves for each channel are shown at the top; the perceived trajectory is shown at the bottom.

Thus the intensity drops to 0.707 in the middle, from a starting point of 1 at the side. This is a difference of 3 dB. To the ear, whose sensitivity is more correlated to intensity than to amplitude, the sound appears to be fainter in the center, as if it has moved away from the listener.

Constant Power Panning

A constant power pan uses sinusoidal curves to control the amplitude emanating from the two loudspeakers (Reveillon 1994). This creates the impression of a pan with a more stable loudness:

$$A_{amp} = \frac{\sqrt{2}}{2} \times [\cos(\theta) + \sin(\theta)]$$

$$B_{amp} = \frac{\sqrt{2}}{2} \times [\cos(\theta) - \sin(\theta)].$$

In the middle of this pan, $A_{amp} = B_{amp} = 0.707$, thus:

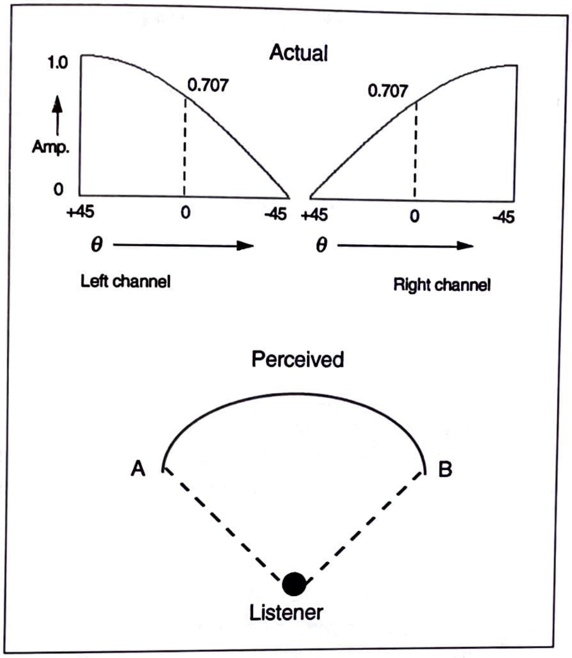


Figure 11.7 A constant-power panning curve maintains the perceived distance and intensity in the middle. The amplitude curves for each channel are shown at the top; the perceived trajectory is shown at the bottom.

$$I = \sqrt{0.707^2 + 0.707^2} = \sqrt{0.5 + 0.5} = \sqrt{1} = 1$$

and a constant intensity is preserved.

Figure 11.7 shows the constant intensity pan. The perceived pan is seen as rotating between the two loudspeakers at a constant distance from the listener.

Reflections

As sound pans from loudspeaker to loudspeaker in a concert hall, reflections in the hall provide more cues to source position. (At certain positions in some halls they may confuse the sense of directionality, but this is a degenerate case.) Thus to enhance the localization effect, the composer can add small delays to the signal coming from the "nondirect" channels (i.e., the channels from which the main source is not being projected). These delays simulate the reflections of a hall; they tell the ear that the source direction is elsewhere. In the ideal, the reflection pattern should change as the sound pans.

Table 11.1 Distance traveled by sound waves per unit of time

Time (in ms)	Total distance (in m)	Wavelength frequency (in Hz)
1.0	0.34	1000
3.4	1	340
6.8	2	168
34	10	34
68	20	16.8
100	34	10
340	100	3.4
680	200	1.68
1000	340	1

Note: The corresponding wavelength is also shown. To calculate the delay time of a reflection, use the total distance from the source to the reflecting surface to the listener. The speed of sound is assumed to be about 340 m/sec.

To impart an idea of the relationship between the delay time and the perceived distance of a sound, consider table 11.1. This shows the distance sound travels per unit time. The third column in table 11.1 is added for the sake of the curious, showing the wavelength corresponding to a given distance. As the third row down shows, for example, an acoustical tone at 166 Hz (about an E) takes shape in two meters of air.

Simulating Distance Cues

To make a sound recede into the distance, one can lower its amplitude, apply a lowpass filter, add echoes, or blend in reverberation. The first two

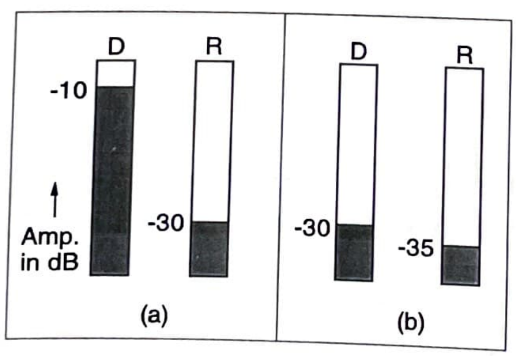


Figure 11.8 Level indicators for simulating a sound that moves away from the listener. D = direct, R = reverberated. (*a*) Close sound in which the direct sound is much higher in amplitude than the reverberated sound. (*b*) Distant sound. The overall amplitude is lower, and the ratio of the direct to the reverberated sound has narrowed.

cues model what happens outdoors in a large open field, where we sense the distance of a sound by its intensity and the filtering effect of air absorption on high frequencies.

Echo and reverberation cues model what happens in an enclosed space such as a concert hall. To simulate a specific distance within a room, the simplest method is to keep the level of reverberation constant and scale the direct signal to be inversely proportional to the desired distance (figure 11.8). An extension of this technique is to scale the reverberant signal as well, according to a function that decreases less rapidly than the direct signal. As the source moves away, the total sound emanating from the source diminishes.

Local and Global Reverberation

Another distance cue is the relationship of *local* reverberation to *global* reverberation, which can be demonstrated with a multiple loudspeaker system. Global reverberation is distributed equally among all loudspeakers, while local reverberation feeds into adjacent pairs of loudspeakers. Thus, a sound might have a short and weak global reverberation but have a long and strong local reverberation coming from one pair of loudspeakers in a multispeaker setup. This would simulate the case of an opening into a large space between the two loudspeakers.

A distinction between local and global reverberation helps overcome a masking effect that occurs at distances where the amplitudes of the direct and global reverberant signals are equal. This masking eliminates the azimuth cue. One way to negate this effect is to split the reverberation into local and global components and make local reverberation increase with distance according to the relation:

$$local\_reverberation \cong 1 - (1/distance).$$

As the distance increases, this relation tends toward 1. Thus, when the source is close to the listener, the reverberation is distributed equally well in all channels. As the source moves away, the reverberant signal concentrates in the direction of the source.

The Velocity Cue or Doppler Shift Effect

Basic localization cues for static sounds can be extended to the simulation of moving sound sources. This is accomplished through a cue to the velocity of the sound source, namely *Doppler shift*, first described by the astronomer

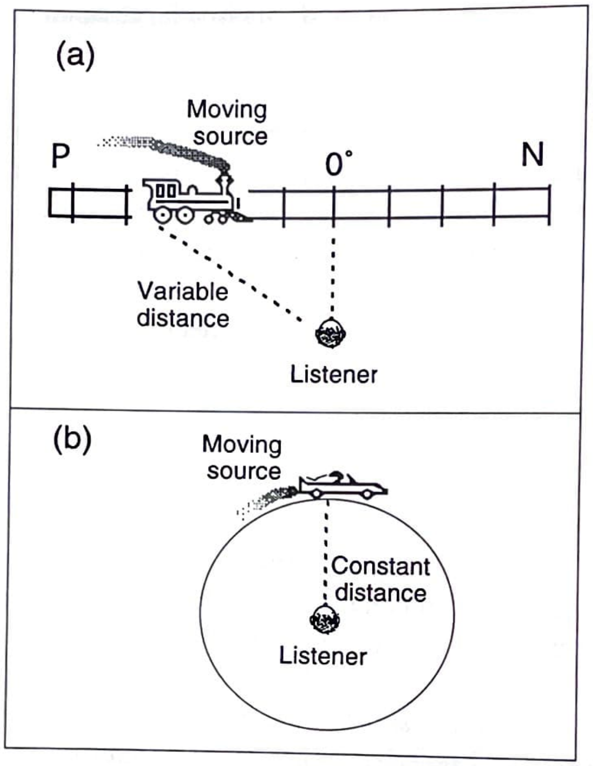


Figure 11.9 A sound moving toward the listener has positive (P) radial velocity. Sound moving away has negative (N) radial velocity. (*b*) Sound moving in a circle is always the same distance away from the listener and so has zero radial velocity.

C. Doppler (1842). The first simulations of Doppler shift to computer music were carried out by John Chowning (1971).

Doppler shift is a change in pitch that results when the source and the listener are moving relative to each other. A common example is heard when standing next to a train track as a train approaches at high speed and then passes. As the train moves closer, the wavefronts of the sound reach us more quickly, causing the pitch to be raised. When the train passes we hear the pitch shift downward.

A Doppler shift is a cue to the *radial velocity* of a source relative to a listener. Radial movement is motion with respect to a center—in this case, a listener (figure 11.9a). Radial velocity is different from *angular velocity*. For a sound to have angular velocity, it must move in a circle around a listener (figure 11.9b). In this case the distance from the source to the listener is constant (i.e., the radial velocity is zero), so there is no Doppler effect. If the position of the listener remains fixed, the Doppler shift effect can be expressed as follows:

$$new\_pitch = original\_pitch \times [vsound/(vsound - vsource)]$$

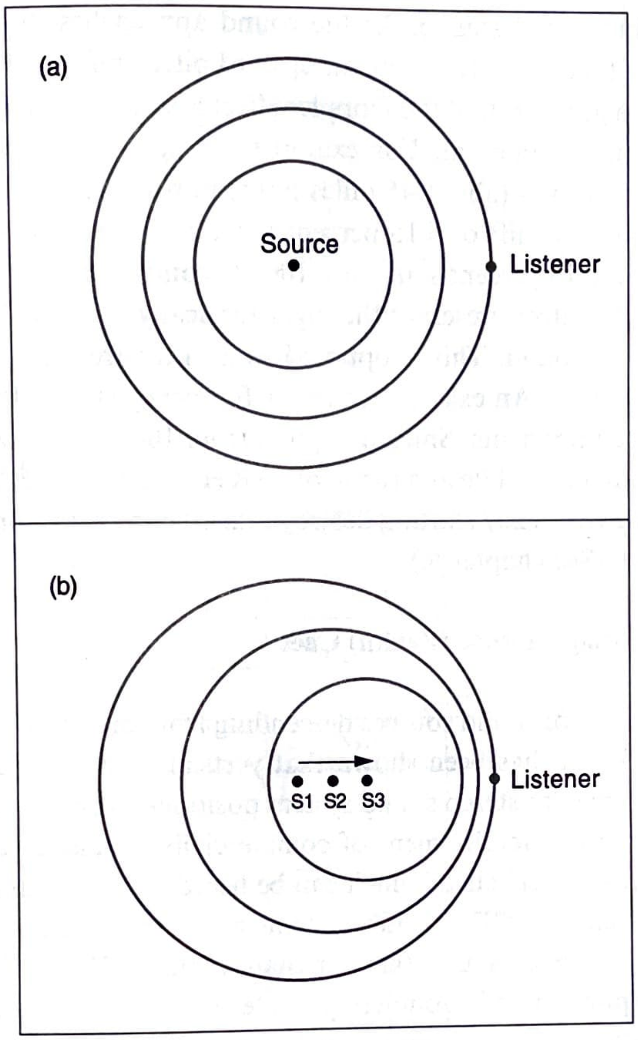


Figure 11.10 Doppler shift wavefront patterns. (*a*) Static sound, wavefronts arrive at constant intervals so there is no pitch change. (*b*) S1, S2, and S3 represent successive positions of a moving sound source. Upward pitch shift.

where *original_pitch* is the original pitch of the sound source, *vsound* is the velocity of sound (~344 meters/second or 1100 feet/second), and *vsource* is the velocity of the source relative to the listener. If *vsource* is positive, the sound is moving closer to the listener, and the pitch shift is upward. If it is negative, the pitch shifts downward.

The pitch change that occurs in Doppler shifting can be explained as a shrinking of the interval between wavefronts as the source moves closer to the listener. Figure 11.10a depicts a static sound emitting wavefronts at a constant rate or pitch. Figure 11.10b depicts a sound source moving toward the listener. The dots S1, S2, and S3 represent successive positions of a moving sound source. As the sound approaches, the wavefronts become closer together, producing an upward pitch shift.

At a given instant the Doppler effect is shifting all frequencies by the same logarithmic interval. For example, an approaching sound moving at 20 meters/second (about 45 miles/hour) raises by about a minor second (6.15 percent). A shift of 6.15 percent for a component at 10 KHz is is 615 Hz, while a 6.15 percent shift for a 100 Hz component is only 6.15 Hz. Thus the Doppler effect preserves the logarithmically scaled interharmonic relations within a sound. This is opposed to a linear frequency shift that occurs in modulation. An example of linear frequency shift is the addition of 50 Hz to all components. Shifting a pitch from 100 to 150 Hz constitutes a major fifth interval, while at a range of 10 KHz, a 50 Hz shift is barely perceptible. Linear frequency shifting destroys existing interharmonic relationships in a sound. (See chapter 6.)

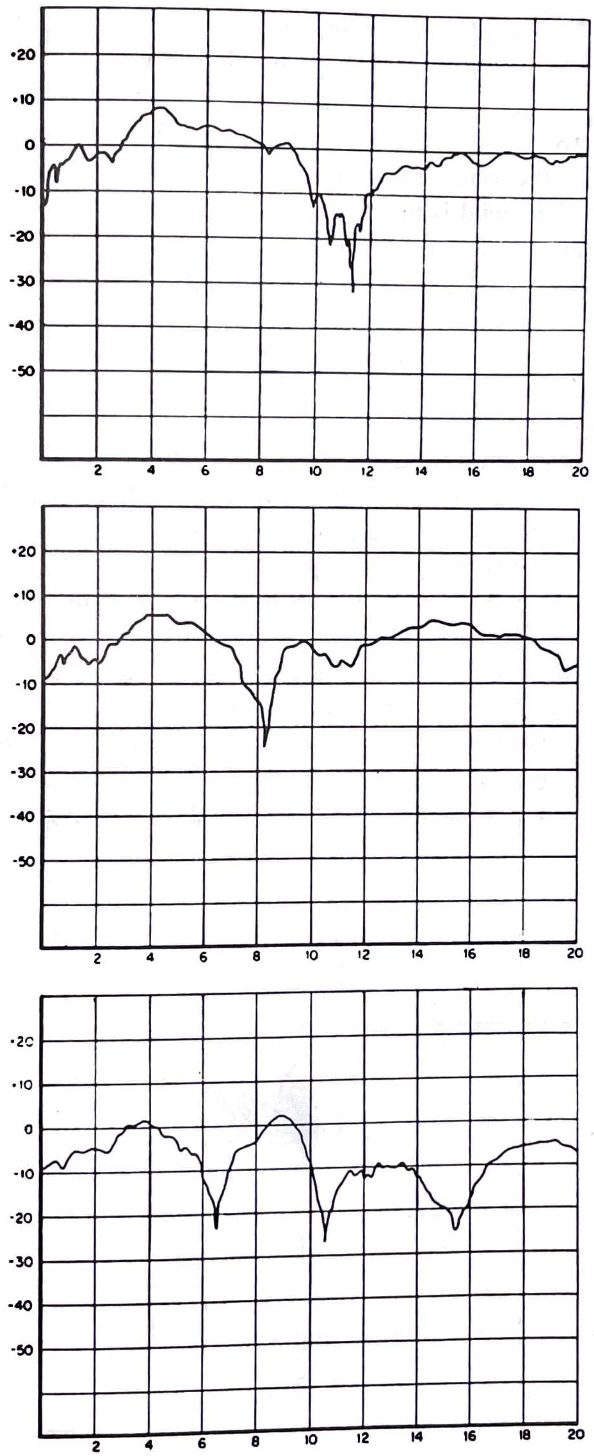
Simulating Altitude (Zenith) Cues

The effect of sound sources descending from on high can be dramatic. Since the 1970s it has been shown that vertical sound illusions can be achieved with a regular stereo sound system positioned at ear level. This research has inspired the development of commercially available vertical spatialization systems, the effects of which can be heard in numerous recordings.

In general, "3D sound" systems are based on research that shows that high-frequency sound (greater than about 6 KHz) reflecting off the outer ears (pinnae) and shoulders provides a critical cue to vertical localization. The surfaces of the pinnae and shoulders act as reflectors, creating short time delays that are manifested in the spectrum as a comb filter effect (Bloom 1977; Rodgers 1981; Kendall and Martens 1984; Kendall, Martens, and Decker 1989).

Zenith cues can be simulated electronically, giving the impression that a sound is emanating from high places. This is done by filtering the input signal, imposing the change in spectrum caused by reflections off the head and shoulders. The filters are set according to the position of the source that one is trying to simulate. This frequency response of this filtration is

Figure 11.11 HRTF spectra for sounds heard at 90 degrees (straight into left ear) at various altitudes. (Top) 15 degrees above ear level. (Middle) Ear level. (Bottom) Below ear level. (After Rodgers 1981; published courtesy of the Audio Engineering Society.)


+20
+10
0
-10
-20
-30
-40
-50
2
4
6
8
10
12
14
16
18
20

called the *head-related transfer function* (HRTF) (Begault 1991). Figure 11.11 plots typical HRTFs for sounds above, at, and below ear level.

In practice, vertical spatial effects are greatly enhanced if the sound is projected in an environment with both front and rear loudspeakers. By panning the sound from front to back or vice versa and applying the HRTF effect, the sound is heard is going over the head of the listener as it pans. Like all spatial effects, vertical panning works best on broadband impulsive sounds rather than low-frequency sounds with smooth envelopes.

Problems with Vertical Sound Illusions

As figure 11.12 shows, a problem with projecting sound in a simulated vertical plane is the variation in HRTFs for different people (Begault 1991; Kendall, Martens, and Decker 1989). When the wrong HRTF is used for a particular person, the vertical panning effect weakens. For a home listening situation, where the filtering is performed in real time on playback, one solution to this problem is to provide several different HRTFs and test signals so that individuals can tune their system to match the response of their ears beforehand.

The robustness of vertical illusions is dependent on the quality of the loudspeakers used and the proximity of the listener to the loudspeakers. Listening to small nearfield monitors, for example, one must remain within their direct sound path, or the vertical illusion falls apart. Thus in a concert

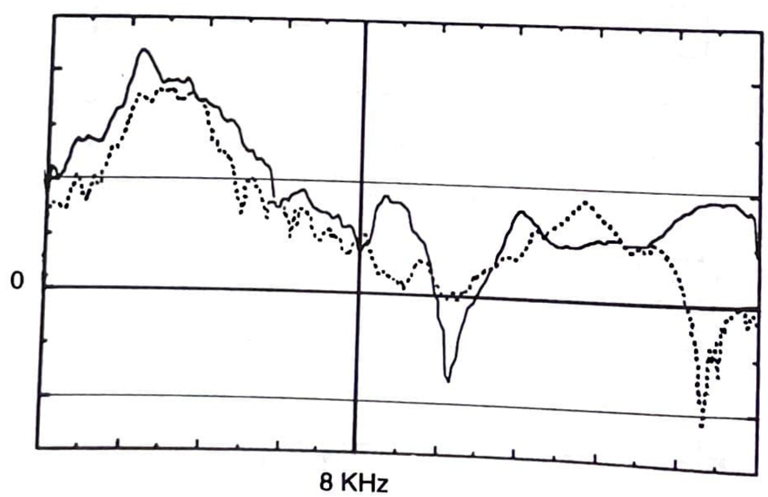


Figure 11.12 HRTF spectra for two different persons. Left ear, source at ear level. The frequency plot goes from 1 to 18 KHz. The vertical line indicates the 8 KHz mark. The differences between the two HRTFs above 8 KHz are striking. The horizontal lines indicate 20 dB differences.

situation it is more practical to suspend actual loudspeakers over the heads of the audience (see figure 11.3d) than to rely on the more fragile illusion of virtual sources.

Binaural Sound

In psychoacoustics research, *binaural* originally referred to a unique listening situation in which subjects are placed in an anechoic chamber with their heads held still by a mechanical restraint and probe tubes inserted into their ear canals. These conditions are designed to analyze a variety of auditory mechanisms in a controlled environment (Durlach and Colburn 1978; Colburn and Durlach 1978; Buser and Imbert 1992). Due to the difficulty of such experiments, many investigations employ headphones. In other experiments, a dummy head with microphones in its ears substitutes for the human subject.

An outgrowth of this research is a genre of *binaural recordings,* made with two microphones in a dummy head or a similar construction, that are meant to be heard through headphones. This genre has been particularly popular in radio productions and has led to the availability of binaural recording systems, including computer-based mixing consoles with horizontal and vertical panning controls.

One of the results of binaural research has been the realization that is possible to create an illusion of a sound source at a specific position in a binaural field through filtering alone. By "binaural field" we refer to the space perceived through headphones, including above and behind the head. These techniques employ the HRTFs discussed earlier. See Blauert (1983), Durlach and Colburn (1978), and Begault (1991) for details.

Sound Radiation

We conclude the discussion of localization with a note on sound raditation. Every sound-producing mechanism has a characteristic *radiation pattern.* This three-dimensional pattern describes the amplitude of sound projected by the device in all directions. In traditional acoustical instruments, the radiation pattern is frequency-dependent (Fletcher and Rossing 1991). That is, it changes depending on the frequency being radiated. Radiation pattern is one clue to the identity and locale of the source.

Loudspeaker systems exhibit their own radition patterns, characterized by the technical specification called *dispersion pattern.* The dispersion pattern of a front-projecting loudspeaker indicates the width and height of the region in which the loudspeaker maintains a linear frequency response.

To the extent that listeners can detect the difference between a real violin and playback of a violin recording has been blamed on their different radiation patterns. Thus one line of acoustics research over the years has concentrated on modeling the radiation patterns of instruments, projecting them on spherical multiloudspeaker setups (Bloch et al. 1992). Such systems, under computer control, could also be used for compositional purposes, to give each voice in a piece its own radiation pattern, for example.

Rotating Loudspeakers

The radiation of sound emitted by a spinning loudspeaker creates a striking spatial effect. The physical rotation of a loudspeaker enlivens even dull, stable sounds, animating them with time-varying qualities.

Rotating Loudspeakers: Background

The original rotating loudspeaker mechanism was the Leslie Tone Cabinet, which routed an incoming signal into two separate rotating mechanisms: a spinning horn for high frequencies and a rotating baffle (blocking and unblocking a stationary woofer) for low frequencies. A remote control for motor speed let musicians adjust the speed of rotation. The resonant horn of the Leslie Tone Cabinet makes it immediately identifiable.

The Leslie Tone Cabinet was designed to enrich the static sound emitted by electric organs such as the famed Hammond B3, with which it was often coupled. But musicians and recording engineers discovered that any sound could be enriched this way, including voice and electric guitar.

In the 1950s, engineers working Hermann Scherchen's Experimental Studio Gravesano in Switzerland developed a spherical loudspeaker (figure 11.13) that rotated both horizontally and vertically (Loescher 1959, 1960). Their goal was to reduce the "directional soundbeam" characteristics of normal loudspeakers. According to one of the designers:

> *A double rotation in the horizontal and vertical plane results in inclined rotational planes of the single speakers and gives best results. The sound field becomes practically homogenous, reproduction takes on an astonishing fullness and smoothness, and the harshness of normal reproduction is completely gone.* (Loescher 1959)

K. Stockhausen manually rotated a loudspeaker affixed to a turntable to create the spinning sounds in his compositions *Kontakte* (1960) and *Hymnen* (1967) (figure 11.14). Later, engineers at the West German Radio

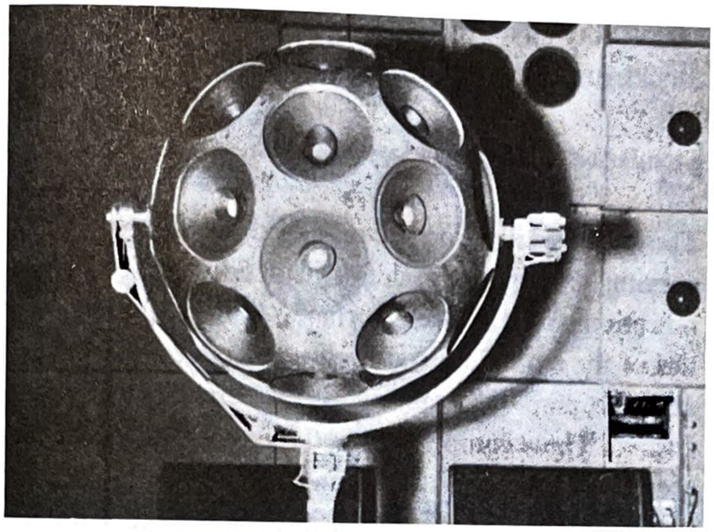


Figure 11.13 Rotating spherical loudspeaker constructed in 1959 at the Experimental Studio Gravesano.

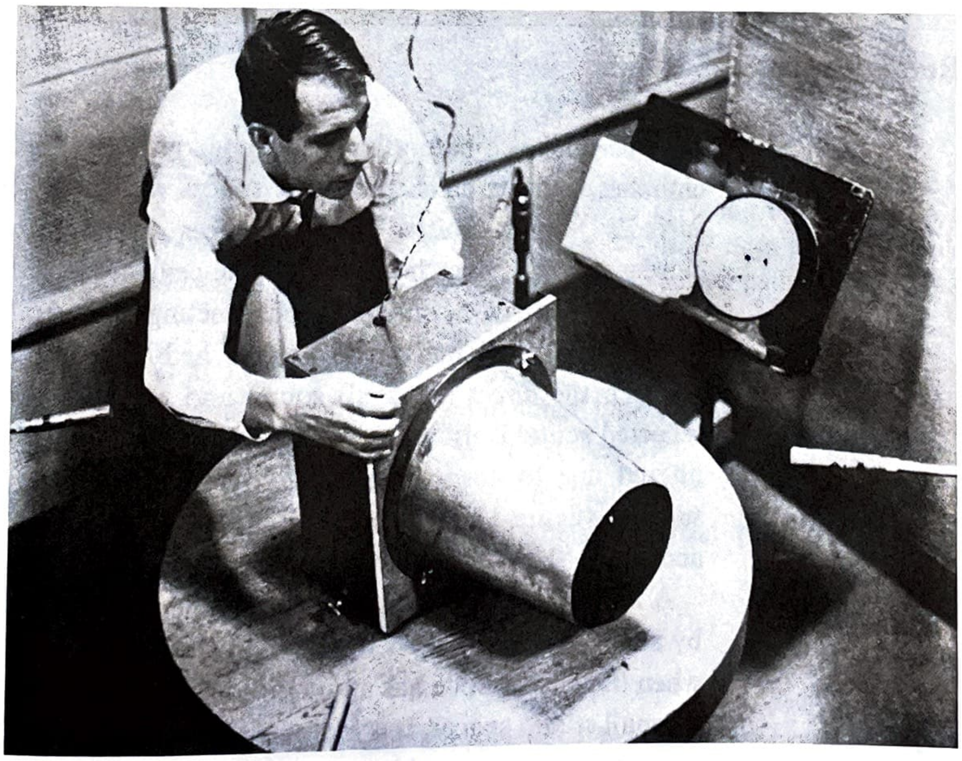


Figure 11.14 K. Stockhausen with rotating loudspeaker mechanism (1960). Four microphones are positioned around the loudspeaker turntable, which was manipulated by hand. A later version was controlled by a motorized mechanism. (Photograph copyright WDR, Cologne.)

(WDR) built a motorized sound rotation system for concert performance of Stockhausen's works (Morawska-Büngler 1988).

Simulation of Rotating Loudspeakers

The effects of rotation are manifold, involving Doppler shift vibrato, time-varying filtering, phase shifts, distortions caused by air turbulence, and echo reflections from adjacent surfaces—not to mention the transfer characteristics of the amplifiers and loudspeakers used. The Leslie Tone Cabinet, for example, employed vacuum tube electronics with "overdrive" distortion if desired. These complicated and interacting acoustical and electronic effects are difficult to simulate convincingly using digital signal processing. Nonetheless, a number of synthesizers and effects units offer programs that simulate rotating loudspeakers. Such programs should improve as more sophisticated algorithms are developed.

Reverberation

Reverberation is a naturally occurring acoustical effect. We hear it in large churches, concert halls, and other spaces with high ceilings and reflective surfaces. Sounds emitted in these spaces are reinforced by thousands of closely spaced echoes bouncing off the ceiling, walls, and floors. Many of these echoes arrive at our ears after reflecting off several surfaces, so we hear them after the original sound has reached our ears. The ear distinguishes between the *direct* (original) sound and the *reflected* sound because the reflected sound is usually lower in amplitude, slightly delayed, and lowpass filtered due to absorption of high frequencies by the air and reflecting surfaces (figure 11.15). The myriad echoes fuse in our ear into a lingering acoustical "halo" following the original sound.

A microphone recording of an instrument in a concert hall is surrounded by an envelope of reverberation from the hall. This is particularly the case when the microphone has an omnidirectional pattern. For recordings made in small studio spaces, it is often desirable to add reverberation, since without it a voice or ensemble sounds "dry," lacks "space" or "depth."

Certain synthesized sounds have little or no intrinsic spaciousness. These acoustically "dead" signals can be enhanced by spatial panning, echoes, and reverberation processing.

But space is not merely a cosmetic appliqué for sounds. Spatial depth can be used to isolate foreground and background elements in a compositional architecture. Further, reverberation is not a monolithic effect; there are

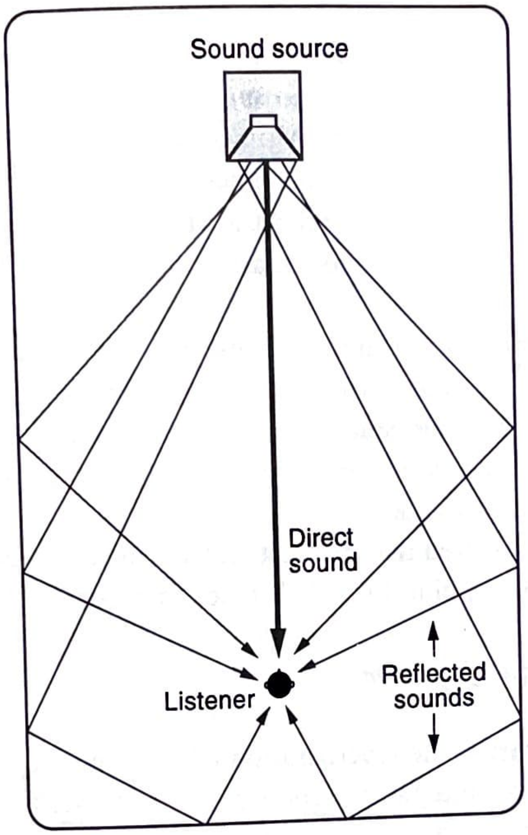


Figure 11.15 Reverberation is caused by reflections of sound off surfaces in a space. The dark line is the path of direct sound; all other lines represent sonic reflections that arrive later than the original due to their longer paths.

many colors and qualities of reverberation—as many as there are natural spaces and synthetic reverberators. No single type of reverberation (natural or synthetic) is ideal for all music. Most electronic reverberation units simulate several types of reverberation. Some attempt (often crudely) to simulate known concert halls, while others create bizarre spatial images that would be impossible to duplicate in a real hall.

Properties of Reverberation

Glorious-sounding salons and concert halls have been constructed since antiquity, but their basic acoustical properties were not well understood from a scientific standpoint until the late nineteenth century. The pioneering work on the analysis of reverberant spaces was carried out by Wallace Sabine (1868–1919), who advised in the construction (starting from an existing structure) of Boston's acclaimed Symphony Hall in 1900. Symphony Hall was the first performance space designed according to rigorous, scientifically derived principles of acoustics. Sabine observed that a room's reverberation is dependent on its volume, geometry, and the reflectivity of its surfaces (Sabine 1922). It is no surprise that large rooms with reflective surfaces have long reverberation times, and small rooms with absorptive surfaces have short reverberation times. Smooth, hard surfaces like glass, chrome, and marble tend to reflect all frequencies well, while absorptive surfaces like heavy curtains, foam, and thick carpeting tend to absorb high frequencies.

The geometry of the room surfaces determines the angle of sound reflections. Walls that are not parallel scatter the wavefronts in complicated dispersion patterns, and small irregularities such as plaster trimmings, indentations, columns, and statues tend to diffuse the reflections, creating a richer, denser reverberation effect.

Sabine also observed that humidity affects the reverberation time in large halls, since up to a point, humid air tends to absorb high frequencies.

Impulse Response of a Room

One way to measure the reverberation of a room is to trigger a very short burst (an *impulse*) and plot the room's response over time. This plot, when corrected for the spectrum of the burst, shows the *impulse response* of the room. As mentioned in chapter 10, circuits also exhibit an impulse response, making the impulse response measurement an often-used tool both in circuit design and concert hall design. Natural reverberation typically has an

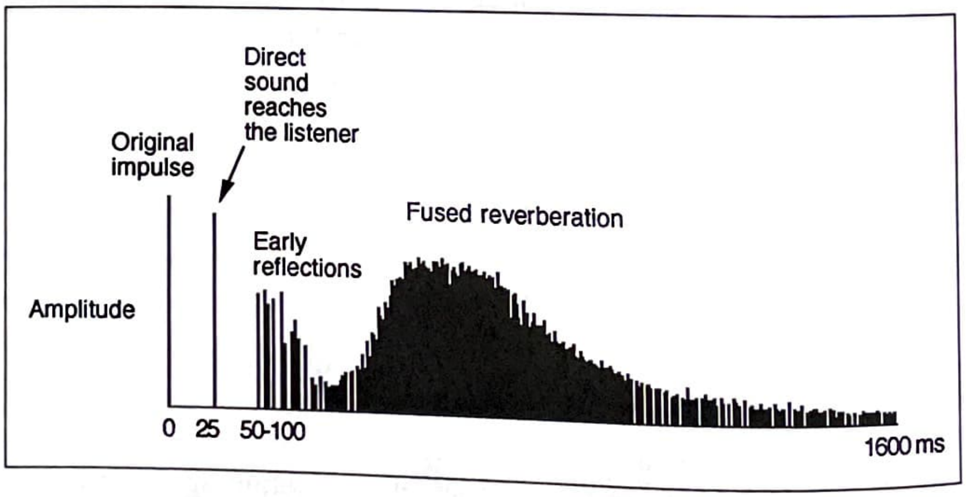


Figure 11.16 Impulse response envelope of a reverberant hall. The components of reverberation are shown as the *predelay* (shown as a 25-ms delay before the direct sound reaches the listener), the early reflections, and the fused reverberation.

impulse response envelope similar to that shown in figure 11.16. The buildup of reverberation follows a quasi-exponential curve that reaches a peak within a half-second and decays more or less slowly.

In general, an irregular time interval between peaks is desirable in a concert hall. Regularly spaced peaks indicate "ringing"—resonant frequencies in the hall—which can be annoying.

Reverberation Time

Another important measurement of reverberation is *reverberation time* or RT60. The term RT60 refers to the time it takes the reverberation to decay 60 dB from its peak amplitude (1/1000 of its peak energy). Typical RT60 times for concert halls are from 1.5 to 3 seconds. The RT60 point of the plot in figure 11.17 is 2.5 seconds.

Artificial Reverberation: Background

The earliest attempts at artificial reverberation of recordings transmitted the sound through an *acoustic echo chamber,* then mixed the reverberated signal with the original. Some large recording studios still allocate a separate room as an echo chamber. They place a loudspeaker at one end of a reflective room and put a high-quality microphone at the other end. The sound to be reverberated is played over the loudspeaker and picked up by the microphone (figure 11.18). An echo chambers offers a unique acoustical ambience created by a specific room, loudspeaker, and microphone. When all these conditions are sympathetic, the quality of reverberation may be excellent. A drawback to the echo chamber approach (besides the practicalities of

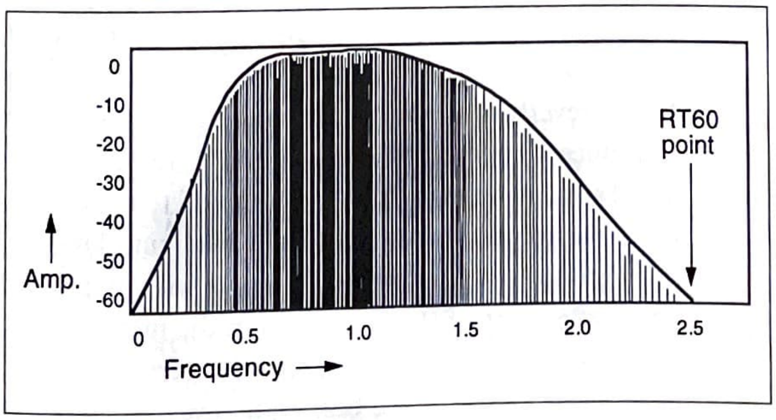


Figure 11.17 Reverberation time is measured as the point at which the reverberation decays to -60 dB of its peak level.

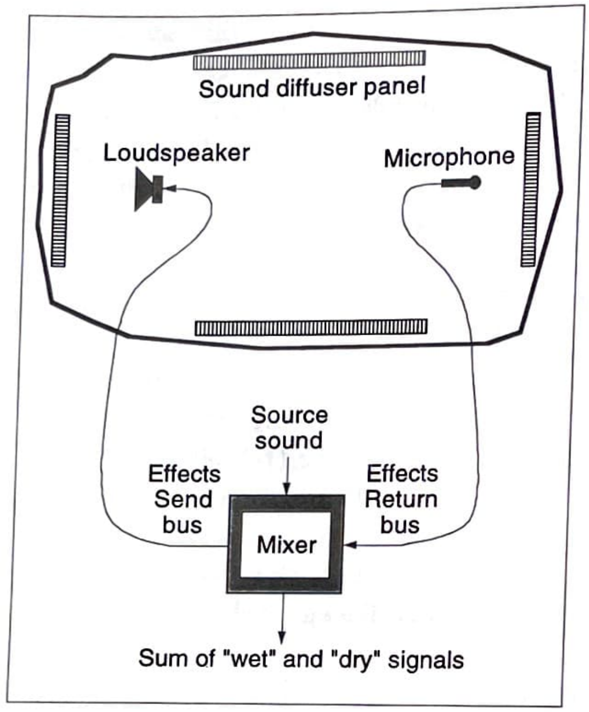


Figure 11.18 To create an acoustic ambience effect, sound can be fed into an echo chamber via a loudspeaker. The reflected, indirect sound is picked up by a microphone at the other end of the room. Ideally, the room is irregularly shaped. To maximize and randomize the reflections, the room should be fitted with *sound diffuser panels*. Sound diffuser panels contain many recesses spaced at different distances. As sound waves strike them they are reflected back at different delay times, depending on which recess they hit. This diffusion effect tends to eliminate *standing waves* (resonant frequencies in the room) caused by parallel walls.

constructing such a space) is that the reverberation cannot be varied tremendously.

The more usual way of adding reverberation is with a *reverberation unit* or *reverberator*. Before digital reverberators were introduced in the mid-1970s, reverberators were electromechanical contraptions containing two transducers (input and output) and a reverberating medium like a long metal spring or a metal plate. The sound to be reverberated was transmitted from the input transducer to the medium. The medium transmitted the sound to the output transducer mixed with myriad echoes caused by vibrations/reflections of the signal within the medium. The result was amplified and mixed with the original signal to create a rather "colored" artificial reverberation effect. The best plate reverberators produced relatively clean and diffuse reverberation, but they were limited to an RT60 of only a few seconds and a fixed reverberation pattern.

Digital Reverberation Algorithms

Digital reverberators use time delays, filters, and mixing to achieve the illusion of sound scattering within a room. From a signal-processing standpoint, a reverberator is a filter with an impulse response that resembles the impulse response of a room. Manfred Schroeder of the Bell Telephone Laboratories (1961, 1962, 1970) was the first to implement an artificial reverberation algorithm on a digital computer. His reverberation programs soaked up hours of computation time on the behemoth mainframe computers of the epoch. Modern reverberation units are compact and run in real time. Control knobs and buttons on their front panels let musicians dial up a variety of effects. Most reverberators can be controlled via MIDI (see chapter 21).

Parts of Reverberation

The effect of reverberation can be broken into three parts, shown earlier in figure 11.16.

- *Direct (unreflected) sound* travels in a straight path and is the first sound to arrive at the listener's ears
- *Discrete early reflections* hit the listener just after the direct sound
- *Fused reverberation* contains thousands of closely spaced echoes but takes some time to build up and then fade away

Commercial reverberation units usually provide controls that let one manipulate these parts more or less independently. On these units, the balance between the reverberated and direct sound is sometimes called the *wet/dry* ratio (the reverberated sound is said to be "wet"), and the delay just before the early reflections is called the *predelay*.

Effective simulation of natural reverberation requires high *echo density*. Some early digital reverberators produced as few as 30 echoes per second, while in actual concert halls, an echo density of more than 1000 echoes per second is typical. Many reverberators today provide a control that lets users adjust the echo density to suit the desired effect, from discrete echoes to a dense, fused reverberation pattern.

The discrete early reflections of a concert hall can be simulated by means of a *tapped delay line*. This is simply a delay unit that can be "tapped" at several points to put out several versions of the input signal, each delayed by a different amount. (See chapter 10 for an explanation of tapped delay lines.)

The lush sound of fused reverberation requires a greater echo density than a tapped delay line can efficiently provide. Many different algorithms for fused reverberation exist, but they all usually involve a variation on M. R. Schroeder's original algorithms, described next.

Unit Reverberators

Schroeder called the building blocks *unit reverberators*, of which there are two forms: *recursive comb filters* and *allpass filters*, both of which were introduced in chapter 10.

Recursive Comb Filters

As explained in chapter 10, a recursive or *infinite impulse response* (IIR) comb filter contains a feedback loop in which an input signal is delayed by D samples and multiplied by an amplitude or gain factor g, and then routed back to be added to the latest input signal (figure 11.19a).

When the delay D is small (i.e., less than about 10 ms) the comb filter's effect is primarily a spectral one. That is, it creates peaks and dips in the frequency response of the input signal. When D is larger than about 10 ms, it creates a series of decaying echoes, as shown in figure 11.19b. The echoes decay exponentially, so for the maximum number of echoes (the longest decay time), g is set to nearly 1.0. The time it takes for the output of the comb filter to decay by 60 dB is specified by the following formula (Moore 1990):

$$decay\_time = (60/-loopGain) \times loopDelay$$

where *loopGain* is the gain g expressed in decibels $= 20 \times \log_{10}(g)$, and *loopDelay* is the delay D expressed in seconds $= D/R$, where R is the sampling rate. Thus if $g = 0.7$, then $loopGain = -3$ dB.

Allpass Filters

Allpass filters transmit all frequencies of steady-state signals equally well (see chapter 10). But they "color" sharp transient signals by introducing frequency-dependent delays. When the delay time is long enough (between 5 and 100 ms), the allpass filter shown in figure 11.20a has an impulse response as shown in figure 11.20b: a series of exponentially decaying echo pulses, like a comb filter with a long delay. The uniform spacing between the pulses suggests that when a short, transient sound is applied, the filter rings with a period equal to the delay time of the filter. This explains why allpass

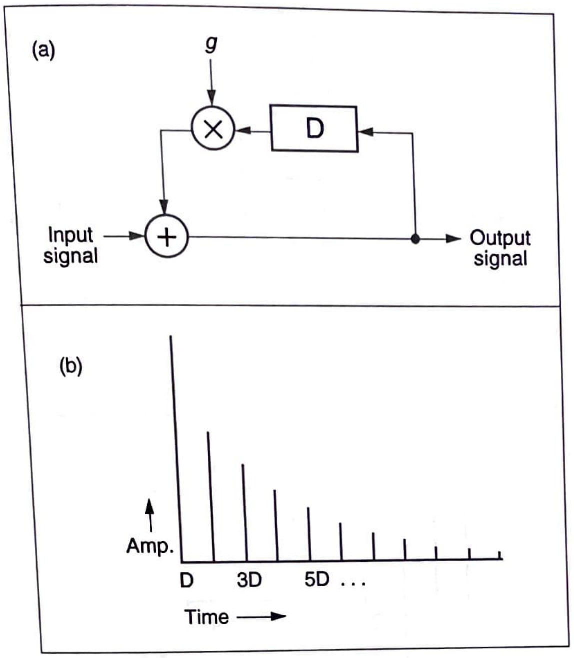


Figure 11.19 A recursive comb filter for reverberation. (*a*) Circuit of comb filter with coefficients D (number of samples to delay) and g (amount of feedback). (*b*) Impulse response, as a series of echoes.

filters are not "colorless" when they treat sounds with sharp attack and decay transients.

Reverberation Patches

We have established that both recursive comb and allpass filters can generate a series of decaying echoes. For lush reverberation, it is necessary to interconnect a number of unit reverberators to create sufficient echo density so that the echoes fuse. When unit reverberators are connected in parallel, their echoes add together. When they are connected in series, each echo generated by one unit triggers a series of echoes in the next unit, creating a much greater echo density. The number of echoes produced in series is the product of the number of echoes produced by each unit.

In Schroeder's designs, comb filters are interconnected in parallel to minimize spectral anomalies. For example, a frequency that passes through one comb filter might be attenuated by another. Allpass filters are usually connected in series. Because of the phase distortion they introduce, connecting

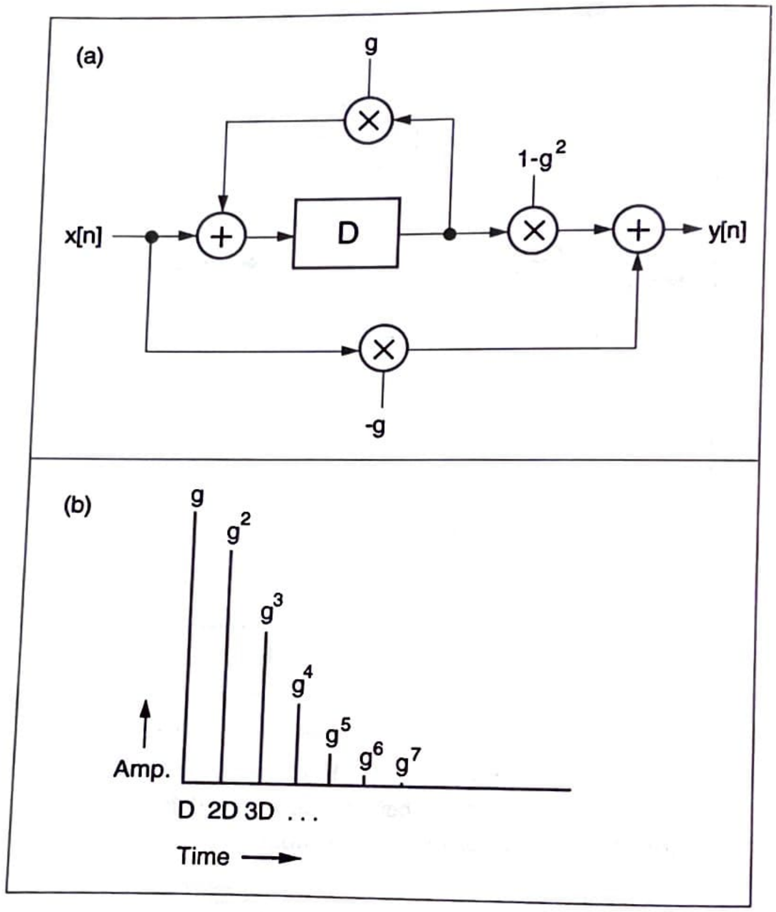


Figure 11.20 A first-order allpass network. (*a*) By adding -*g* times the input into the output of the delay, a comb filter is changed into an allpass filter. (*b*) The impulse response of an allpass filter is an exponentially decaying series of echo pulses. This makes the impulse filter useful as a building block of reverberators.

allpass filters in parallel can result in a nonuniform amplitude response due to phase cancellation effects.

Figure 11.21 shows two reverberators proposed by Schroeder. In figure 11.21a the parallel comb filters initiate a train of echoes that are summed and fed to two allpass filters in series. In figure 11.21b five allpass filters cause the echo density to be multiplied by each unit. If each allpass generates just four audible echoes, the end result is 1024 echoes at the output of allpass number 5.

The characteristic sound of a digital reverberation system of this type is dependent on the choice of the delay times *D* (these determine the spacing of the echoes) and amplitude factors *g* (these determine the decay or rever-

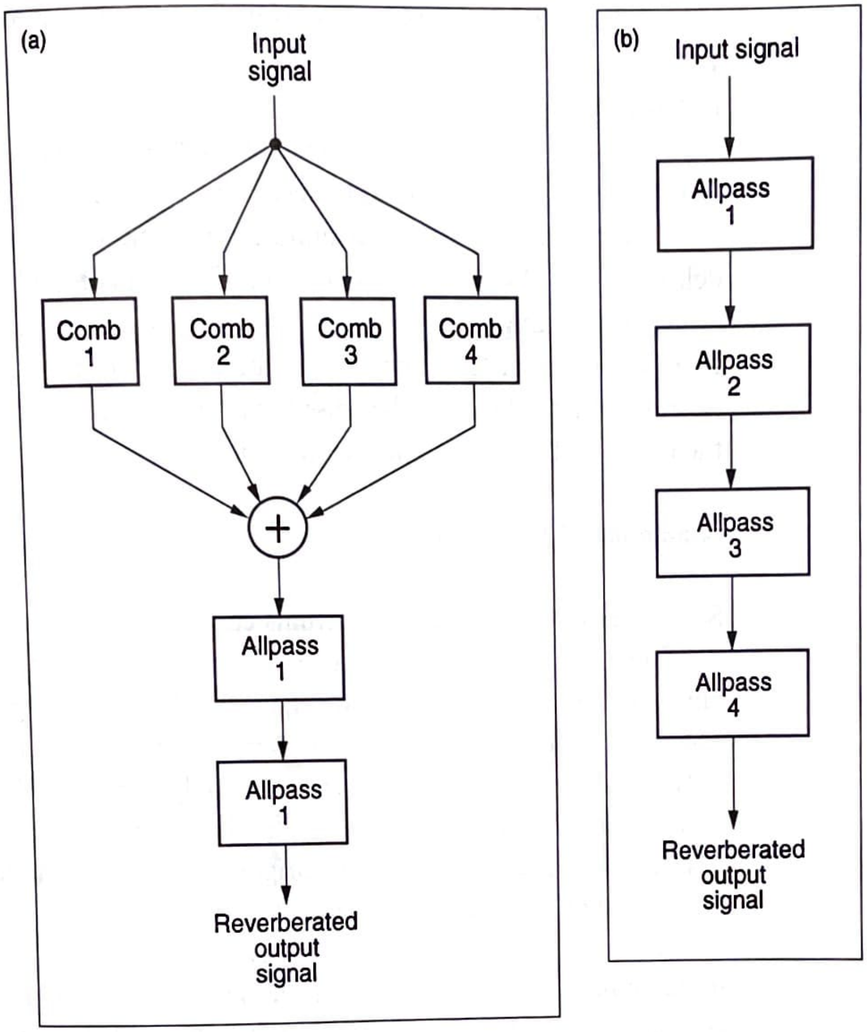


Figure 11.21 Schroeder's original reverberator designs. (*a*) Parallel comb filters fed into two allpass filter stages. (*b*) Five allpass filter stages in series.

beration time) for each of the unit reverberators inside it. The delay time is also called the *loop time*.

For natural-sounding reverberation it is important to choose delay times that are relatively prime to one another (i.e., that have no common divisor) (Moorer 1977, 1979c). Why is this? Consider two comb filters, where the delay time of the first is 10 ms and that of the second is 12.5 ms. The length of their delay lines are 800 samples and 1000 samples, respectively, at a sampling rate of 40 KHz. Because the lengths of both delay lines are divisible by 200, a reverberator built from these two units does not have a smooth decay. At multiples of 200 ms, the echoes coincide to increase the amplitude at that point, causing a sensation of discrete echoes or regular "bumps" in the decay. When the delay times are adjusted to 10.025 and 24.925 ms, the length of their delay lines are 799 and 997, respectively. Now the first coincidence of echoes does not occur until (799 × 997)/40 KHz = 19.91 seconds. (See Moorer 1979c for a discussion of how to tune these parameters.)

As might be expected, shorter delay times correlate with the sound of smaller spaces. For a large concert hall, the reverberator in figure 11.21a uses comb filter delay times around 50 ms with a ratio of longest: shortest delay of 1.7:1. For a small tiled room effect the comb filter delay times can be set in the range of 10 ms. The allpass filters have relatively short loop times of 5 msec or less. The reverberation time of the allpass filters must be short (less than 100 msec) because their purpose is to increase the density of the overall reverberation, not its duration.

Simulation of Early Reflections

Schroeder's reverberation algorithms can be characterized as *tapped recirculating delay* (TRD) models. As explained earlier, the reverberator is usually partitioned into comb and allpass sections, which generate sufficient echo density to create a reasonable simulation of *global reverberation.* The TRD model is efficient, but it simulates only generic global reverberation, and not the detailed acoustic properties of an actual performance space.

In 1970 Schroeder extended his original reverberator algorithms to incorporate a *multitap delay line* to simulate the early reflections that are heard in a hall before the outset of the fused reverberant sound. (See chapter 10 for more on multitap delay lines.) This design, which has been adopted in most commercial reverberators, is shown in figure 11.22. Thus to simulate a particular concert hall, a straightforward way to improve the basic TRD model is to graft the measured early reflection response of the hall onto the generic global reverberator (Moorer 1979c). A further extension is to lowpass filter the global reverberation according to the measured sound absorption characteristics of the hall.

Another important consideration in reverberation design is that the sounds presented in each ear should be *mutually incoherent.* That is, the reverberation algorithm should be slightly different (*decorrelated*) for each channel of processing.

Fictional Reverberation Effects

The goals of the electronic music composer extend beyond the simulation of natural reverberant spaces. A reverberator can conjur up many unusual "fictional" spatial effects that are not meant to be realistic. A common example is "gated" reverberation that explodes quickly in echo density,

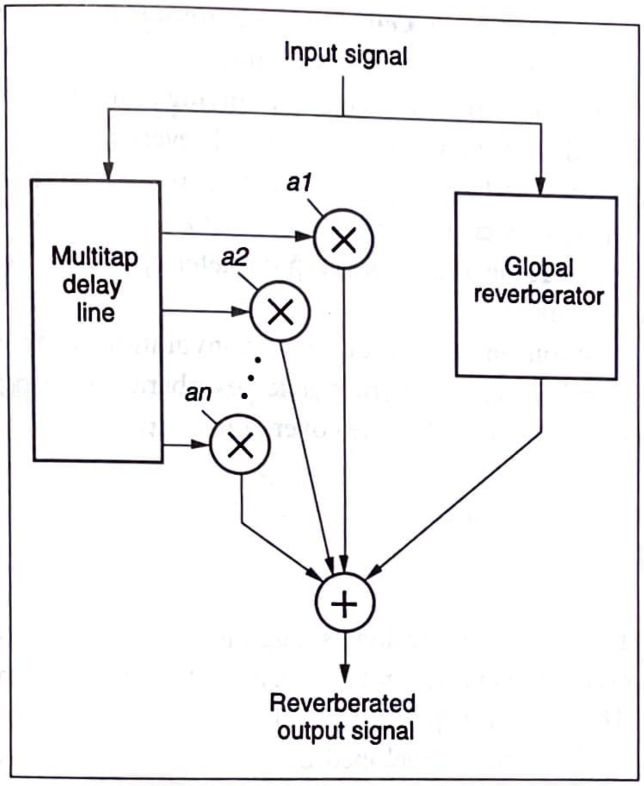


Figure 11.22 In Schroeder's later designs, a multitap delay line simulated the early reflections of sound in a concert hall.

Table 11.2 Typical parameters of reverberators

Parameter	Description
Type of reverberation	Choice between "Hall," "Chamber," "Plate," or "Gated"
Size	Sets the delay times within the unit reverberators
Predelay	Controls the onset time of the effect
Input delay	Causes the effect to precede the cause (the wet sound precedes the dry sound)
Reverberation time	Sets the decay time
Diffusion	Determines the echo density
Mix	Ratio of input sound to reverberated sound at the output of the device
Highpass filter	Reverberates only the upper octaves of the sound, creating a "sizzling" reverberation effect
Lowpass filter	Reverberates only the lower octaves of the sound, creating a "muffled" reverberation effect

then cuts off suddenly. Gated reverberation was used on snare drums in the 1980s and quickly became a pop music cliché. Other effects include a "sizzling" reverberation, obtained by applying a highpass filter to the reverberated sound, and its opposite, a muffled reverberation obtained by applying a steep lowpass filter. By manipulating the parameters of a reverberator, one can create weird combinations such as tiny rooms with long reverberation times. Table 11.2 lists the parameters provided on many commercial reverberators.

The section on reverberation by convolution, later in this chapter, presents another type of nonrealistic reverberation using the asynchronous granular synthesis technique covered in chapter 5.

Modeling Sound Spaces

The study of reverberation is ongoing. The algorithms described in the earlier section on reverberation are a starting point for the designs discussed here. This section explains several approaches to more realistic reverberation that have been developed in recent years. These include extensions to the basic Schroeder algorithms, geometric models, reverberation via convolution, waveguide reverberation, and multiple-stream reverberation.

Several of these techniques represent a *physical modeling* approach to reverberation. (See chapter 7 for an introduction to the theory of physical modeling in the context of sound synthesis.) These mathematically intensive methods model the diffusion of acoustical waves in actual spaces. Besides creating more realistic models, they offer the possibility of simulating imaginary spaces. In this category we include rooms whose characteristics and geometry can change over time—such as an elastic concert hall that "expands" and "shrinks" over the course of a phrase—or impossible spaces such as a closet with a long reverberation time. Thus the goal of these techniques is not always realistic reverberation, but rather a dramatic spatial transform.

Extensions to Schroeder Reverberation Algorithms

In the standard Schroeder reverberation algorithms, the allpass filters generate a series of echoes with an exponential decay. An extension to the Schroeder model is to substitute an *oscillatory allpass* filter for the regular allpass filter in the Schroeder design. In this case, the impulse response of the allpass filter is a pulse train with an amplitude of a damped sinusoid

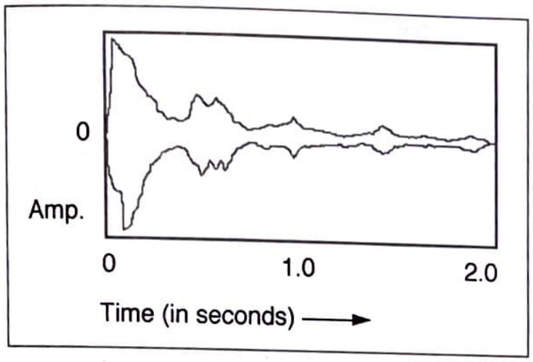


Figure 11.23 The impulse response of an oscillatory allpass unit reverberator.

(figure 11.23). This models the case of a "good sounding" room with a slightly undulating reverberation pattern (Chowning et al. 1974; Moorer 1979c).

Geometric Modeling of Sound Spaces

An alternative to the TRD approach is to build a physical model of a room's geometry using a computer-aided design (CAD) system. The loudspeakers that project the sound constitute an "acoustic window" into the simulated room surrounding them.

In F. R. Moore's design (1983), each sound source becomes a vector with an adjustable position, directionality, magnitude, and dispersion. Starting from the projection of a source vector into the room, the computer traces the path of sound reflections (Moore 1983). In a complete geometric model, the reverberation algorithm would model the reflection patterns of hundreds of simulated sound rays. Depending on the detail of the model, this approach can be computationally expensive. For efficiency's sake, Moore used a geometric approach to model only the early reflections of a simulated room. He used a standard Schroeder TDR model for global reverberation.

A problem of a too-simple geometrical approach to reverberation has been pointed out by Moorer (1979). Such an approach fails to take into account the *diffusion* (scattering) of sound rays that occurs in real halls. Diffusion occurs since no surface is 100 percent smooth and reflective, meaning that sound waves scatter and their energy is partially absorbed at each point of reflection. Thus a number of methods try to improve on the ray-tracing model by explicitly modeling sound diffusion. These may insert a stochastic scattering function at each point of reflection. The *waveguide network* reverberation, discussed later, is another attempt to model explicitly sound diffusion.

Reverberation via Convolution

An accurate but computationally intensive means of simulating the reverberation in a given space is to convolve the impulse response of the space with the signal to be reverberated. (See chapter 10 and Smith 1985a for more on convolution.) One can think of reverberation as a type of filter, where the length (in samples) of the impulse response corresponds to the reverberation time (in samples) of the simulated hall. The impulse response of a room is gathered by recording the room's response to a very brief explosive sound. This set of samples is then convolved with the signal to be reverberated.

Chapter 10 distinguishes between *direct* and *fast* convolution. Direct convolution is not practical for reverberation because of the enormous amount of computation it entails. For example, at a sampling rate of 48 KHz and an impulse response length of three seconds, each sample of each channel of the input signal must be multiplied and summed 48,000 × 3 times. For a second of input sound this translates into the following:

144,000 Multiply/adds per sample (impulse response)	×	48,000 Samples per second	=	6,912,000,000. Multiply/adds per second per channel

Thus, to reverberate one second of stereo sound by convolution would require 13.824 billion multiply/adds. Calculating this in real time requires a level of performance that is currently associated with expensive supercomputers. On a signal-processing engine benchmarked at 100 million multiply/adds per second in a practical application, such as a plug-in board for a personal computer, this calculation would take about two minutes and eighteen seconds to compute, a factor of 138:1 as compared to real time.

Thus the only practical reverberation by convolution uses fast convolution, taking advantage of speedups offered by the fast Fourier transform (FFT). See chapter 10 for details on fast convolution; Appendix A explains the FFT.

Granular Reverberation

The rolling of thunder has been attributed to echoes among the clouds; and if it is to be considered that a cloud is a collection of particles of water . . . and therefore each capable of reflecting sound, there is no reason why very [loud] sounds should not be reverberated . . . from a cloud. (Sir John Herschel, quoted in Tyndall 1875)

This section describes a reverberation effect that can be achieved by convolving an arbitrary input sound with a cloud of sonic grains.

It is well known that clouds in the atmosphere contribute a reverberation effect. The nineteenth-century French acoustical scientists Arago, Mathieu, and Prony, in their experiments on the velocity of sound, observed that under a perfectly clear sky the explosions of cannons were always singular and sharp. Whereas, when the sky was overcast or a large cloud filled part of the sky, cannon shots were frequently accompanied by a long continuous "roll," similar to thunder (Tyndall 1875). (See Uman 1984 for an analysis of the acoustics of thunder.)

Provided that one understands convolution, it is not a great surprise to learn that the convolution of a sound with a cloud of sound particles creates a scattershot "time-splattered" effect akin to atmospheric reverberation. Time-splattering begins with a more-or-less dense cloud of sound grains generated by the technique of *asynchronous granular synthesis* (AGS), described in chapter 5. AGS scatters grains statistically within a region defined on the time/frequency plane. In convolution, this mass of grains can be thought of as the impulse response of a cumulus cloud enclosure. The virtual "reflection" contributed by each grain splatters the input sound in time; that is, it adds multiple irregularly spaced delays. If each grain was a single-sample pulse, then the echoes would be faithful copies of the original input. Since each grain may contain hundreds of samples, however, each echo is locally time-smeared.

Time-splattering effects can be divided into two basic categories, which depend mainly on the attack of the input sound. If the input begins with a sharp attack, each grain generates an echo of that attack. If the cloud of grains is not continuous, these echoes are irregularly spaced in time. If the input has a smooth attack, however, the time-splattering itself is smoothed into a kind of strange colored reverberation (figure 11.24). The "color" of the reverberation and the echoes is determined by the spectrum of the grains, which is a factor of the duration, envelope, and waveform of each grain. See chapter 5 for more details on grain parameters.

Waveguide Reverberation

A *waveguide* is a computational model of a medium in which waves travel. Physicists have long used *waveguide networks* to describe the behavior of waves in resonant spaces (Crawford 1968). The waveguide network approach to reverberation is built out of a set of bidirectional delay lines (Smith 1985c, 1987a, b; Garnett and Mont-Reynaud 1988; chapter 7 presents more on waveguides in the context of sound synthesis). Each delay line

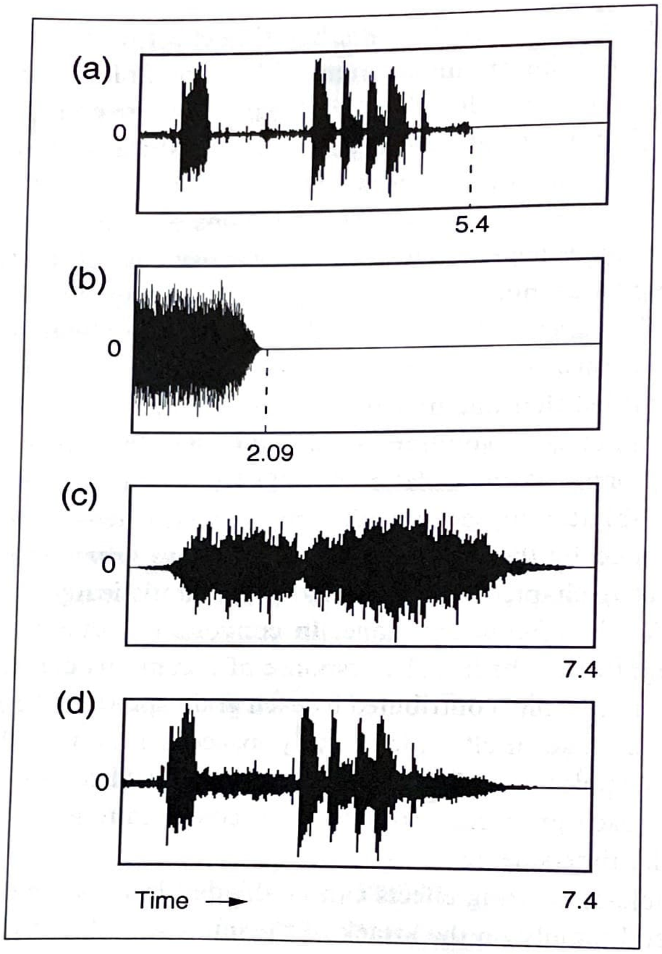


Figure 11.24 Reverberation by granular convolution. (*a*) Speech input: "Moi, Alpha Soixante." (*b*) Granular impulse response, consisting of 1000 9-ms sinusoidal grains centered at 14,000 Hz, with a bandwidth of 5000 Hz. (*c*) Convolution of (*a*) and (*b*). (*d*) Mixture of (*a*) and (*c*) in a proportion of 5:1, creating reverberation around the speech.

contains a wave propagating in the one direction and reflecting back to the center junction when it reaches the end of the line. By connecting a number of waveguides together into a network, one can build a model of acoustical media, such as the reflection pattern of a concert hall.

In waveguide reverberation, the lengths of the individual waveguide delay lines are different from one another to simulate the different echo times within a hall. At the junction of multiple waveguides the energy is scattered among them, causing a diffusion effect that is typical of fused reverberant sound (figure 11.25). In a *closed network*, once a signal is introduced it recirculates freely throughout the network without loss of energy. To obtain

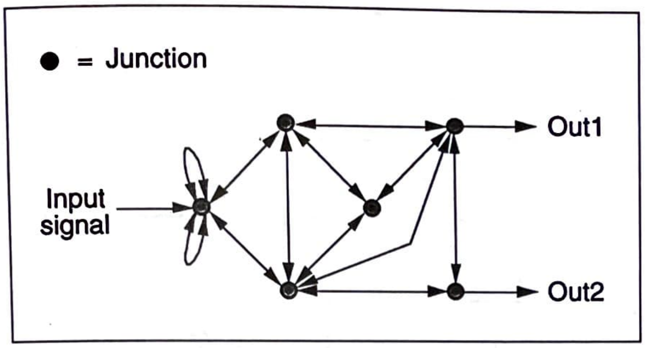


Figure 11.25 A three-port waveguide network with six nodes. This waveguide propagates energy out of the outputs, meaning that it is an open network that eventually loses energy, as a reverberant hall does.

a reverberation effect one must introduce small losses of amplitude energy within the network to achieve the desired reverberation time. Signal inputs and outputs can be chosen anywhere in the network.

Waveguide networks make efficient reverberation models. A network with N junctions requires N multiplies and $2N - 1$ additions to generate an output sample. The number of junctions N depends on the system being modeled. A model of a resonating box might require eight intersections, while a model of a complex room's reverberation response might take hundreds of junctions, since any place where a signal might scatter requires a junction.

The structure of a waveguide network ensures that there is never any numerical overflow or oscillation within the network. Moreover, the important property of diffusive scattering of sound rays (Moorer 1979), which is poorly handled by a simple geometric model, is simulated well by a waveguide network. A "moving walls" effect can be obtained by smoothly varying the delay line lengths.

Multiple-stream Reverberation

Multiple-stream reverberation can be viewed as a compromise between detailed but computationally intensive approaches to reverberation (such as geometrical modeling or reverberation via convolution) and the efficient but global TRD model. Multiple-stream reverberation splits the reverberated signal into several *streams,* each of which models the local reverberation emanating from a small spatial region of the virtual room. Each stream is implemented with a TRD network (comb and allpass filters) tuned for that region of the room.

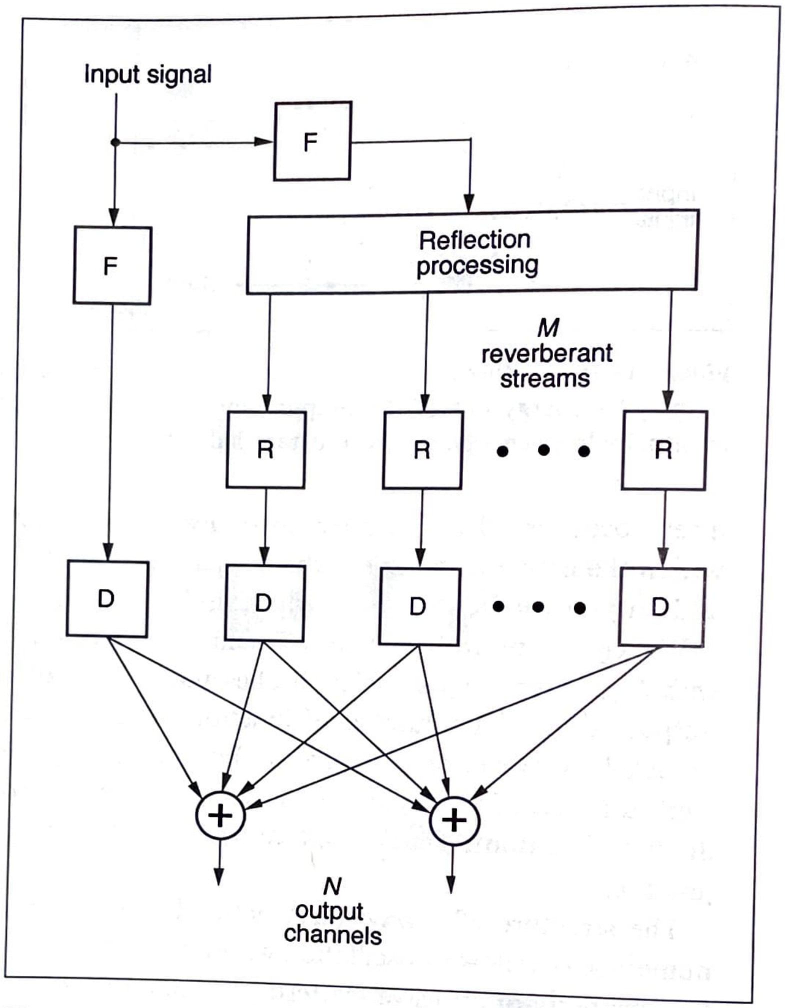


Figure 11.26 Simplified view of a "spatial reverberator" after Kendall, Martens, and Decker (1989). This system models a space by summing the contributions of *M* local reverberators, which ultimately generate *N* output channels. F is a prefilter that imposes spectrum changes caused by distance and air absorption. R is a local reverberant stream, modeling the reverberation in a subspace of the total room. D is a directionalizer that filters the sound according to its position in the virtual space. The implemented system has two independent reflection processors, and some cross-feeding in the reverberant streams.

The "spatial reverberator" system, developed at Northwestern University in the 1980s, takes the multiple-stream approach and combines it with two other processes: (1) a model of room reflections, and (2) localization cues caused by the reflection of sound off of the pinnae, shoulders, and upper torso (Kendall and Martens 1984; Kendall et al. 1986; Kendall, Martens, and Decker 1989). First- and second-order reflections determine the delay times of each independent reverberation stream. Then, after reverberating each stream separately, a "directionalizer" filters each stream to impose additional cues as to its position in a virtual three-dimensional space (figure 11.26).

The user of the system can specify the characteristics of a virtual space in acoustical terms such as room dimensions, sound location, listener location, sound absorption of the walls, and so on. To simulate a room's reverberation pattern, each of the main directions of reverberation is processed as a separate stream, with up to eighteen streams in one implementation (Kendall, Martens, and Decker 1989). As figure 11.26 shows, the number of reverberation streams is independent of the number of output channels used ultimately to project the sound.

The concept of separate reverberation streams was also present in quadraphonic reverberation research carried out at MIT in the early 1980s (Stautner and Puckette 1982). In this work the loudspeaker outputs were spatially responsive to the input channel of the source. For example, a direct sound emanating from a left front loudspeaker would be heard to reverberate from two adjacent loudspeakers and finally from the opposite right rear loudspeaker.

Conclusion

Sound spatialization by electronic means pervades musical production, through microphone techniques, signal processing, and sound projection via loudspeakers or headphones. As psychoacoustical knowledge about spatial perception has increased, systems for spatialization have grown more sophisticated. Many studios own several spatial effects processors. As always, there is a wide range in the audio quality reflected in the devices available at a given time.

Simulations of natural effects such as loudspeaker rotation and reverberation remain approximations to the real world. What happens to sound when a loudspeaker rotates is not entirely understood. The best concert halls have a lush and enveloping quality of reverberation that exceeds the best synthetic reverberation. Although this sensuous quality is best appreciated in the hall, the presence of fine natural reverberation transmits even on recordings played through loudspeakers.

A main advantage of synthetic spatial processors is flexibility—many different types of reverberation and spatial processing. They can also implement supernatural effects that would be impossible in the real world, as well as such features as the ability to switch from one spatial effect to another in synchrony with a musical line.

As a final word, a comment about stereo audio media is in order. Stereophonic recording and playback was a great advance when it was introduced in the 1930s. But the continued dominance of two-channel media for broadcast and sale make it difficult and impractical to realize more sophisticated spatial processing. A true multiple-channel audio medium distributed on a mass scale would greatly stimulate further advances in musical spatialization.