Chapter 6

Abstract Spatialization Backend Interface

The abstract spatialization backend interface can be thought of as an abstraction layer between spatialization resource management and spatialization backends. This concept is also known as a hardware abstraction layer, but the backend is not necessary a piece of hardware, so that we do not want to use this term.

The spatialization device interface is an abstract interface to spatialization devices (e.g., Acoustetron and PSFC [Amano *et al.*, 1996]). This interface ensures that the resource manager (and an application) does not need to be changed if a new device is to be supported. Only a new spatialization device driver, derived from a template, needs to be developed.

The interface informs the resource manager about number of spatialization channels, number of non-spatialization channels, spatialization channel identifiers, non-spatialization channels, Doppler shift support, and volume spatial resolution (e.g., minimum audible angles).

In the other direction the resource manager informs the spatialization device about soundscape and system changes.

A second abstract interface exists for audio channels. A channel can get audio data from a MIDI synthesizer, a sound file player, or a port (e.g., mic). If available, a channel (player device) returns to the spatialization resource manager via the interface frequency and volume in normalized form (i.e., float between 0. and 1.). If a source is set inactive by the resource manager, then the related channel can also be stopped.

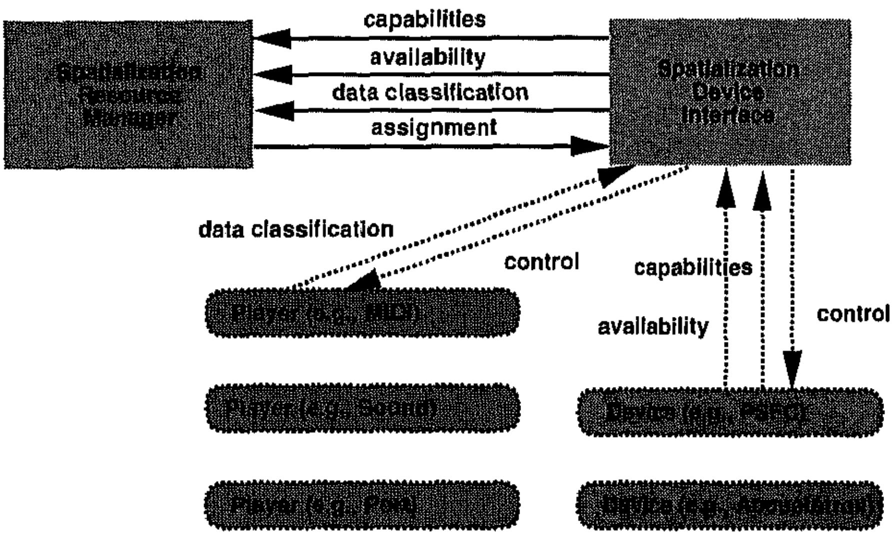

Figure 6.1: Spatialization device interface

6.1 Sound spatialization backends

The developed resource management was tested with different spatialization backends. Backends differ in their way to spatialize sound, the provided interface and realism. The developed algorithm for spatial sound resource management tries to cover all kind of spatialization backends. As representatives for different classes of spatialization backends the following sections introduce the PSFC, a loudspeaker array and the Acoustetron II, a HRTF-based system. The common features are shown and a method to unify the interfaces for the resource management.

6.1.1 Pioneer Sound Field Controller

The PSFC, or Pioneer Sound Field Controller [Amano *et al.*, 1996] [Amano *et al.*, 1998], is a DSP-driven hemispherical loudspeaker array, installed in the Synthetic World Zone at the University of Aizu Multimedia Center. The PSFC system features realtime configuration of an entire sound field, including sound direction, virtual distance, and simulated environment (virtual room characteristics: reverberation level, room size and liveness) for each of two sources. It can also configure a dry (DSP-less) switching matrix for direct directionalization. The PSFC speaker dome is about 10 m in diameter, accommodating about fifty simultaneous users and allowing about twenty users at once to comfortably stand or sit near its sweet spot. Collocated with a large screen rear-projection stereographic display, the PSFC is intended for advanced multimedia and virtual reality applications. The screen and some of the speakers are shown in Figure 6.2.

The PSFC can configure both the context and the content of a virtual sound field: the context is the ambiance or presence — the room size, liveness, reflection pattern, and overall level; the content is the source direction — azimuth, elevation, and suggested distance. The direct sound moves by amplitude panning; the reflected sound moves by rotating the impulse response; the reverberant sound is orientation-independent.

The hemispherical speaker array's audio presentation is complemented by the wide-screen visual presentation. The spatially immersive environment provides a natural group experience, wide visual field-of-view, comfort (no fatigue-inducing or cumbersome head-mounted display). Such a system can be described as "roomware," software for a room, putting the users *inside* the computer [Brooks, 1997]. This notion is also related to the idea of an "immobot," an immobile robot that concentrates on attending and servicing the needs of collocated human users, rather than the traditional robotic tasks of exploration and manipulation of an external environment.

The PSFC's omnidirectional sound field complements the exaggerated visual display's inadequacies for immersion.

Direct sound directionalization

Source direction is set by programming azimuth θ and elevation ϕ for each of two independent source channels, which can move around with an update cycle of less than 100 ms. The transfer functions for are interpolated at the audio sampling rate (and not at the slower control rate), avoiding reconstruction (aliasing) problems. The signal is reproduced by three loudspeakers surrounding the sound image, the level of each loudspeaker determined by the distance between the projected image source and each loudspeaker.

Early reflection

Early reflection patterns were calculated via an image source model from architectural drawings, modeling walls, ceiling and floor with characteristic reflection coefficients to obtain intensity, delay, and direction parameters [Meyer *et al.*, 1965] [Kendall and Martens, 1988]. The FIR filters, room-related impulse responses, were generated by simulating transmission through the solid angle subtended by the respective speakers. Rather than a data-intensive, time-domain FIR filter (specifying amplitudes at a sampling rate) or a frequency-domain FIR filter (multiplying the FFTs of the sources

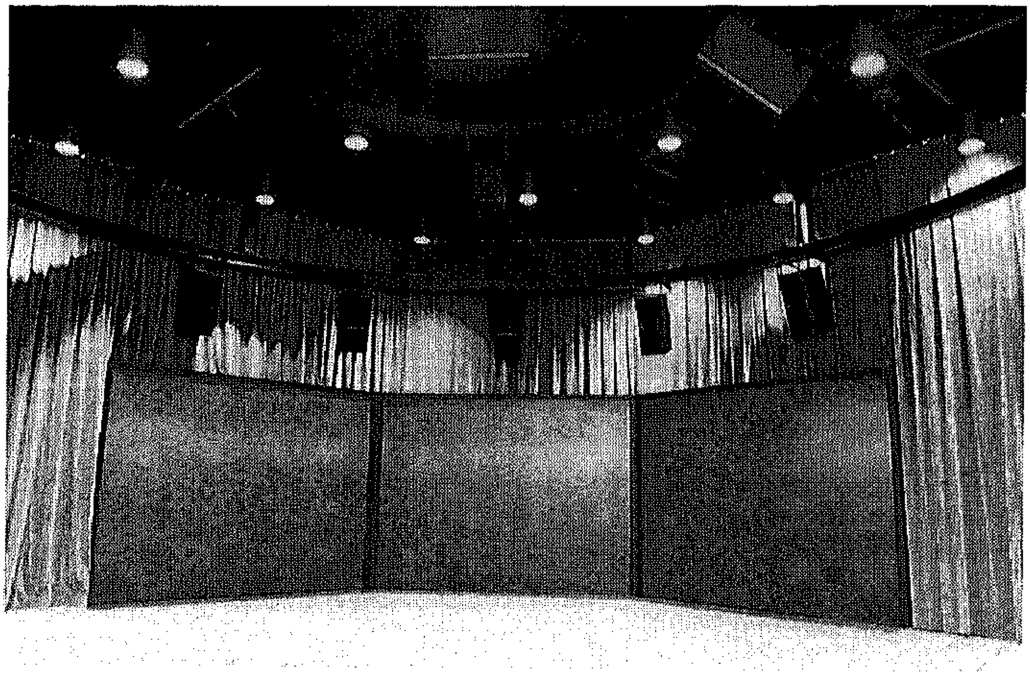

Figure 6.2: Multimedia Center: Virtual Reality Zone

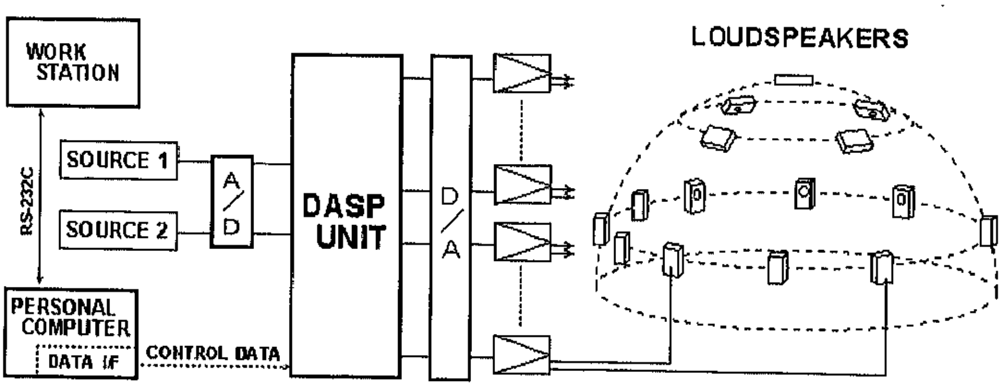

Figure 6.3: Multimedia Center: Pioneer Sound Field Controller

and transfer functions), early reflection modeling is implemented as a table of discrete (potentially sparse) delay time/normalized level pairs, the reflection being generated via a delay network. As implemented in the PSFC, early reflections are series of pure impulses, instead of the results of convolution with a realistic wall response. To simplify the taps, reflection patterns were elided (pruned) whenever the simulated amplitude fell below a certain threshold or when the reflection closely followed another stronger (masking) one. For each of the two source signals, up to seven reflections are reproduced from each of the twelve loudspeakers, for a total of 168. To correct the greater distance between the lower ring of speakers and the center of the room compared to the upper ring, delay is added to the signals bound for upper speakers.

Presence and ambiance: Room size, liveness, and reverberation The perceived sound quality of the simulated sound field is an interaction of programmed room size, liveness, and reverberation parameters.

Room size commands contribute to the perceived auditory spaciousness of the simulated environment. The apparent distance of reflective walls can be changed by varying the initial time gap (a.k.a. predelay) separating the arrival time of the direct sound from that of the first early reflection between 0–500 ms for each source [Tohyama *et al.*, 1995].

Overall volume is controlled by level, but control of the simulated environment's liveness (i.e., how reflective the walls, floor, and ceiling are) is accomplished by scaling the decay rate of simulated early reflections. The room liveness parameter is a kind of time constant, which contributes to the perceptual response characteristic termed "definition" [Rasch and Plomp, 1984]. Definition can be predicted from the ratio of the sound energy (sum of squares of coefficients) of the first 50 ms of the impulse response (including direct sound) to the total sound energy of the impulse response. As illustrated by Figure 6.4, the liveness also implicitly determines the level of the following reverberation, since the reverberation level of each liveness value was prescaled to ensure smooth continuity. Changes in room size and liveness can be effected within a latency of 300 ms.

The reverberation level parameter contributes both to how spacious the simulated environment seems, and indirectly, to how distant the sources seems. As discussed above, the reverberation level is scaled down if the liveness is reduced. But for a given liveness, the apparent distance of the sound source depends on the relationship between the level of the direct sound and the level of indirect sound. For a given virtual room simulation, varying the gain on a source relative to a fixed level of reverberant energy provides a strong cue to source distance. The distance so-cued is called the

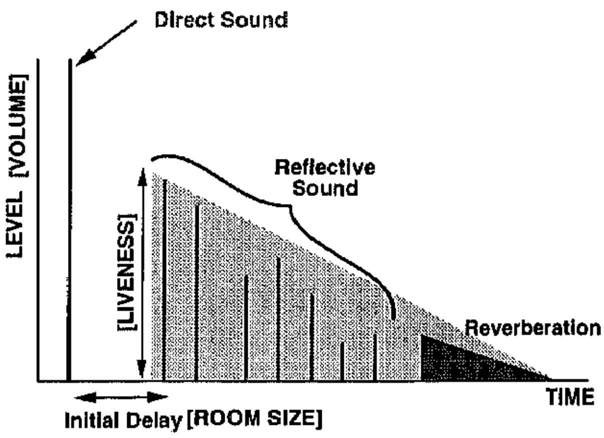

Figure 6.4: Early reflections and reverberation

"indirect-to-direct ratio" (IDR). In the PSFC, the IDR is controlled by simultaneously scaling the output level (which includes direct, reflective, and reverberant sound) and inversely adjusting the liveness (which includes only indirect sound). For example, low gain and high liveness suggest a distant source, while high gain and low liveness suggest a close source.

Exploiting the limited resolution of human hearing, the reverberation patterns are exponentially decayed noise filters, representing a Poisson distribution with time-increasing average (decreasing amplitude and increasing density), independent of source position.

6.1.2 Acoustetron II: an HRTF-based system

The Acoustetron II, developed by Chrystal River Engineering, is based on research at NASA Ames Research center [Foster *et al.*, 1991] [Begault, 1994, p. 205–208] and has its own API [CRE, 1994]. Sound sources are spatialized using head-related transfer functions for headphones. The HRTFs in the system are measured for a constant distance. The number of available spatialization channels vary depending on number of installed convolution boards (a typical configuration has 8 channels.) The system does not provide reverberation for distance cues or room impression. First-order reflection can be configured for "shoe box" (rectangular prism) environments, where each reflection takes up one spatialization channel. Doppler shift is supported.

6.1.3 MIDI: simple spatialization

Simple spatialization can be performed within a MIDI synthesizer or tone generator. Stereo panning and intensity control can give approximate spatialization adequate for many applications. More advanced spatialization would use pitch shift to implement the Doppler effect and mimicking interaural time delays. Control over reverberation would allow better distance cues and room effects. Further, occluder effects can be implemented through simple filtering. Experiments of using pitch shift for interaural time delays using two MIDI channels show problems regarding unpredictable time delays in the execution of pitch shift commands [Ishikawa, 1999]. Implementing MIDI spatialization on the level of a synthesizer might solve those problems.

6.2 Specification

Table 6.1 lists abstract interface functions for spatialization backends. Not all functions are listed, but only those related to the resource management process. Details are documented in the manual of the Sound Spatialization Framework [Herder, 1998].

Besides initialization and termination of a connection to a spatialization backend, primary functions are for updating source and sink data. The framerate for this updating process can be limited not to overburden the spatialization backend. Listener movements are passed via the `updateSink` function using sink location and orientation as parameters. Sound source changes like location, direction, and intensity are passed by the function `updateSource`.

6.3 Comparing the backend interfaces and unification

Table 6.2 shows which interface functions are used to implement the abstract interface for several backends.

Neither the PSFC nor the MIDI backend have functions for sink positioning, but the functionality can be achieved by translation of all sources.

6.4 Discussion of the features

Table 6.3 lists the different features of spatialization backends used within the prototype of the framework. The table does not show the quality of

function	parameters	description
`init`	`numberOfHeads`	Initialize the spatialization module and return the number of available mixels (spatialization channels).
`close`	`fromHead`, `toHead`	Shutdown the spatialization module. This can be limited to a range of of sinks using the parameters `fromHead` and `toHead`.
`updateSource`	`sourceIndex`, `location`, `direction`, `intensity`	Update a sound source. If the `sourceIndex` doesn't exist yet, create a new sound source. The source gain level is given using the parameter `intensity`, which is a normalized (0 ...1).
`updateSink`	`sinkIndex`, `location`, `orientation`	Update a sound sink. If the `sinkIndex` dosn't exist, a new sound sink is created. The location is given in world coordinates.
`removeSink`	`sinkIndex`	Removes a sound sink from the registered sinks.

Table 6.1: Abstract spatialization backend interface

function	PSFC	MIDI	Acoustetron II
initialize	`CL_PSFCinit`	`MD_RESETALL-CONTROLLERS`	`cre_init`
update source	`CL_Coordinates`	`MD_PAN` `MD_CHANNELVOLUME`	`cre_locate_source` `cre_amplify_source`
update sink	`CL_Coordinates`	`MD_PAN` `MD_CHANNELVOLUME`	`cre_locate_head` `cre_update_audio`

Table 6.2: Comparison between spatialization backend interfaces

immersion which can be achieved with different backends.

feature	PSFC	MIDI	Acoustetron II
Doppler effect	no	no	yes
mixels	2	16-32	8/12
output channels	14	2	2
distance cue	yes	yes	yes
left/right localization	yes	yes	yes
front/back localization	yes	no	yes
up/down localization	yes	no	yes
participants	20	limited by audio system	1

Table 6.3: Comparison between spatialization backend features

Bibliography

[Amano *et al.*, 1996] Katsumi Amano, Fumio Matsushita, Hirofumi Yanagawa, Michael Cohen, Jens Herder, Yoshiharu Koba, and Mikio Tohyama. PSFC: the Pioneer Sound Field Control System at the University of Aizu Multimedia Center. In RO-MAN *'96 - 5th* IEEE *International Workshop on Robot and Human Communication*. IEEE, November 1996.

[Amano *et al.*, 1998] Katsumi Amano, Fumio Matsushita, Hirofumi Yanagawa, Michael Cohen, Jens Herder, William Martens, Yoshiharu Koba, and Mikio Tohyama. A Virtual Reality Sound System Using Room-Related Transfer Functions Delivered Through a Multispeaker Array: the PSFC at the University of Aizu Multimedia Center. *TVRSJ:* **T**rans. *of the* **V***irtual* **R***eality* **S***ociety of* **J***apan*, 3(1):1–12, March 1998. ISSN 1342-4386.

[Begault, 1994] Durand R. Begault. *3-D Sound for Virtual Reality and Multimedia*. Academic Press, 1994. ISBN 0-12-084735-3.

[Brooks, 1997] Rodney A. Brooks. The Intelligent Room Project. In Jonathon P. Marsh, Chrystopher L. Nehaniv, and Barbara Gorayska, editors, *CT'97 - Second International Cognitive Technology Conference*, pages 271–278. IEEE, IEEE press, August 1997. Aizu-Wakamatsu, Japan, August 25-28, 1997.

[CRE, 1994] Crystal River Engineering, Inc. *CRE_TRON Library Reference Manual*, August 1994. Revision B.

[Foster *et al.*, 1991] Scott H. Foster, Elizabeth M. Wenzel, and R. Michael Taylor. Real-time synthesis of complex acoustic environments. In *Proc.* (IEEE) *ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, October 1991. Summary.

[Herder, 1998] Jens Herder. Sound Spatialization Framework. Web site, University of Aizu, Japan, 1998. `http://www-ci.u-aizu.ac.jp/SF/`.

[Ishikawa, 1999] Kimitaka Ishikawa. Using a MIDI module as a sound spatialization backend. Master's thesis, University of Aizu, 1999.

[Kendall and Martens, 1988] Gary Kendall and William L. Martens. Spatial reverberation. U.S. Patent 4,731,848; European patent specification 0 207 084 B1, 1988.

[Meyer *et al.*, 1965] E. Meyer, W. Bugtorf, and P. Damaske. An apparatus for electroacoustical simulation of sound fields: Subjective auditory effects at the transition between coherence and incoherence. *Acustica*, 15:339–344, 1965.

[Rasch and Plomp, 1984] R. A. Rasch and R. Plomp. The listener and the acoustic environment. In D. Deutsch, editor, *The Psychology of Music*, pages 135–147. Academic Press, 1984. ISBN 0-12-213560-1 or 0-12-213562-8.

[Tohyama *et al.*, 1995] Mikio Tohyama, Hideo Suzuki, and Yoichi Ando. *The Nature and Technology of Acoustic Space.* Academic Press, London, 1995. ISBN 0-12-692590-9.