Chapter 1

Introduction

Sound spatialization is the processing of an audio stream in a virtual environment so that the audio stream is perceived from a specified location with source attributes like radiation pattern relative to listener location and ambiance. Included in this process is all the necessary audio rendering, like filtering by the media, reverberation and reflections. Sound spatialization resource management [Herder and Cohen, 1997] is the process of controlling a spatialization backend, which performs the actual audio rendering. Good resource management can be achieved only if the application-driven requirements are known and the audio rendering process is well understood. The context of sound spatialization resource management, the main theme of this thesis, is shown Figure 1.1. The upper layer includes multimedia or virtual reality applications which request audio rendering from an audio rendering system. This system has a well-defined sound spatialization application programmer interface and allows one to abstract the application from the resource management and vice versa. The resource management controls the audio rendering using the abstract sound spatialization backend interface.

Ease for application programmers Without a resource management system like that provided with the toolkit [Herder, 1998] developed by the author, application programmers must anticipate a lot of different configurations, with a consequent burden in programming spatial audio sources. During development of a systems which use spatial audio, the required spatialization resources and available spatialization resources are hard to predict (e.g., the number of participants in a chatspace application might vary, or the available spatialization resources on a certain computer system might be limited). Resource management eases the implementation by taking over the resource management processes, gradually scaling down the resource demand, depending on resource availability, in a perceptually optimal form.

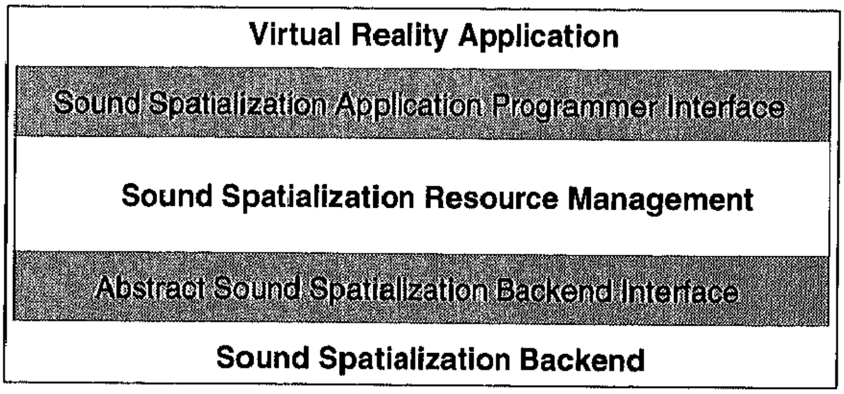


Figure 1.1: Resource management context

This has been also recognized by industry, which introduces resource management using priorities defined on the level of the application programmer interface. Early audio spatialization devices supported only a few of spatialization channels under the assumption that a listener could track only a small number of simultaneous moving sound events. User feedback and testing of applications show that application programmers have difficulties using sound spatialization channels effectively for best experience. Taking that into account, resource management, introduced on the level of the spatialization device and allowing more virtual channels, improves the spatial impression [Creative Technology Ltd., 1998].

1.1 Related research

Sound (source) prioritization for realtime scheduling can be performed by using a heuristic (estimate) for what a listener would attend [Fouad *et al.*, 1997]. This rating uses three factors: the listener's gaze (based on orientation response to an aural stimuli), the intensity of the sound (as a masking criteria), and the age of the sound (for modeling the adaptation response of the human aural system).

The number of calculated early reflections using an image source model can be dynamically adjusted to available processing power [Huopaniemi *et al.*, 1996]. As long as time is left within a processing cycle, higher-order reflections are calculated. Ranking is done from direct sound to higher orders, but not within a class (i.e., reflections of same order).

Common image sources (i.e., reflection) for several sound sources can be calculated in the case that the sound sources are close to each other [Savioja *et* *al.*, 1997, p. 44]. Such a representative image source is calculated by averaging multiple sound source into one sound source.

Predictors for human perception have been developed and applied in the field of global illumination. A perceptually-based visible difference predictor can be used as a criteria in an algorithm for adaptive mesh subdivision in image synthesis applications [Martens and Myszkowski, 1998].

A computational model for spatial hearing [Martin, 1995] might be used as a predictor for spatialization resource management. Such models have high computational costs.

Perceptual coding is used in different domains like audio (e.g., Dolby NR) and vision (e.g., MPEG and JPEG). Mainly masking effects are used to determine compressible information. In the auditory domain, masking is where a signal overwhelms another signal, which might be undesired noise, in a nearby frequency. The concepts in this thesis use also masking, but in the spatial domain. The main interest is to reduce the computational costs for spatialization and not reproduction (including transmission) of audio signals.

Sound spatialization is not surround sound (e.g., AC-3 Standard [Todd *et al.*, 1994]). In a surround sound system, listener and sound sources are usually at predetermined locations. In a sound spatialization system those can be freely placed depending on user or other data input. A surround sound system can be used to display auditory output from a sound spatialization system.

In general, localization task performance using only an auditory display is worse than using a visual display [Barfield *et al.*, 1997].

1.2 Applications of spatial sound

Spatial sound applications vary widely and are introduced in the following sections. The sections provide the reader with the context and helps to define requirements for a sound spatialization resource management framework.

1.2.1 Auditory feedback

Non-visual interfaces for visually disabled users employ spatial sound as a display and feedback device [Aritsuka and Hataoka, 1997]. Hierarchical graphical user interfaces can be extended by an auditory user interface applying auditory icons to convey user interface objects [Meinard, 1997], activated by moving a pointer (e.g., mouse-driven cursor) over certain objects. In the same way, an action applied to an object (e.g., pressing a button) activates an auditory icon giving information about the object and the applied action.

Such user interfaces address the needs of blind users but also benefit other users. Spatial sound can be used to convey the location of objects, including pointing indicators. Audio messages (earcons) [Blattner *et al.*, 1992, p. 89] can be used to report events to the user without disturbing a visual task like reading a text. An earcon coupled with spatial location can convey spatial location of an event or tighten the association of an earcon and a graphical object.

Acoustic cues can enhance the level of immersion provided by haptic applications [Ruspini and Khatib, 1998]. The objects in contact produce a sound depending on their materials and performed action (e.g., pushing, scratching, ...). A haptic interface provides sensation at a specific location. This sensation should be coupled with sound at same location for greater realism. Those sounds must be synchronized and controlled with the haptic events.

1.2.2 Teleconferencing

A sound spatialization system enhances teleconferencing applications [Aoki *et al.*, 1994]. Three primary benefits [Barfield *et al.*, 1997] of auditory spatial information displays are identified as:

1. Relieving processing demands on the visual modality,
2. Enhancing spatial awareness, and
3. Directing attention to important spatial events.

Speech intelligibility degrades as target and distracting source become closer [Hawley *et al.*, 1996]. On the other hand, judgment of the number of concurrent voices does not improve with spatialization [Kashino and Hirahara, 1996].

The well-known "cocktail party effect" [Arons, 1992] can be used for enhancing speech recognition. Talker identification can be eased by matching visual and acoustical location or each talker has an own unique acoustical spatial location. Auditory cues like talking at close range can suggest communication modalities like `confide` as described in the following sections.

A Chatspace is a VR environment which enables people to meet and communicate in cyberspace. People are represented as avatars, graphical objects (Figure 1.2) with behavior controlled by each participant. Such an environment will be dramatically improved using voice spatialization.

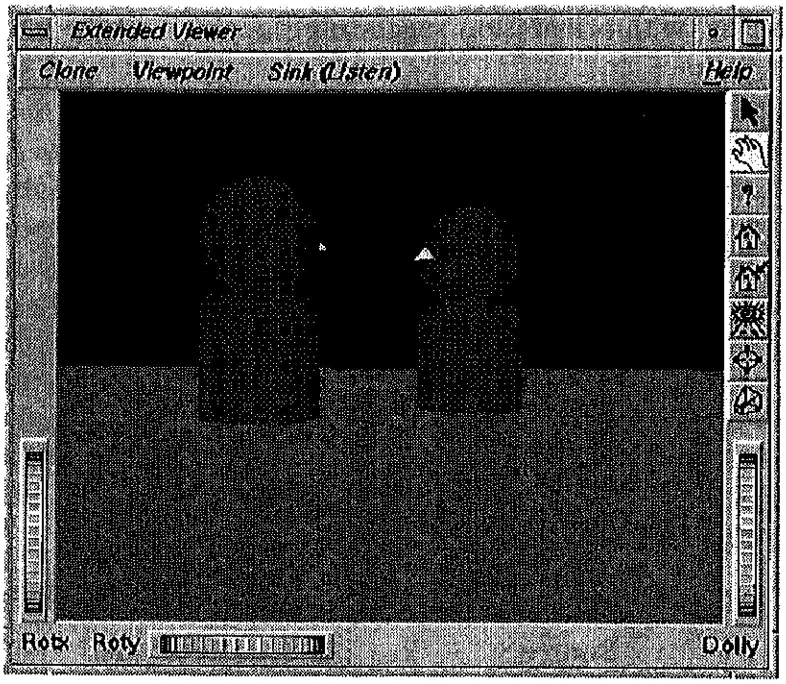


Figure 1.2: Chatspace: people meet and communicate in cyberspace

Symbolic representations of exclude and include for sound sources and sinks

Figurative suggestions of `mute/solo` & `cue` and `deafen/confide` & `harken` are described for exclude and include for sound sources and sinks [Cohen and Herder, 1998]. Those selection functions for the communication channels have direct influence on the resource management. On a logical level, sound source and sinks (receivers, generalization of listener and microphones) are resources assigned to users.

Shared virtual environments, e.g., chatspaces, require generalized control of user-dependent media streams. Traditional audio mixing idioms for enabling and disabling various sources employ `mute` and `solo` functions, which, along with `cue`, selectively disable or focus on respective channels. Exocentric interfaces which explicitly model not only spatial audio sources, but also location, orientation, directivity, and multiplicity of sinks, motivate the generalization of `mute/solo` & `cue` to exclude and include, manifested for sinks as `deafen/confide` & `harken`, a narrowing of stimuli by explicitly blocking out and/or concentrating on selected entities. This section introduces figurative representations of these functions, virtual hands to be clasped over avatars' ears and mouths, with orientation suggesting the nature of the blocking.

Virtual Mixing Non-immersive perspectives in virtual environments enable fluid paradigms of perception, especially in the context of frames-of-reference for conferencing and musical audition [Cohen, 1995] [Cohen, 1998]. Traditional mixing idioms for enabling and disabling various sources employ `mute` and `solo` functions, which, along with `cue`, selectively disable or focus on respective channels. Exocentric interfaces which explicitly model not only spatial audio sources, but also location, orientation, directivity, and multiplicity of sinks, described by Table 1.1, motivate the generalization of `mute/solo` & `cue` commands to exclude and include, manifested for sinks as `deafen/confide` & `harken`, a narrowing of stimuli by explicitly blocking out and/or concentrating on selected entities [Cohen, 1997] [Cohen and Koizumi, 1998]. (`harken` is used to describe focusing on one's own sink.)

	Role	
	Source	Sink
Function	radiation	reception
Level	amplification	sensitivity
Direction	output	input
Instance	speaker (human or loud-)	listener (human or dummy-head)
Organ	mouth	ear

Table 1.1: $^{\mathrm{s}}\mathrm{OU}^{\mathrm{rce}}_{\mathrm{Tput}}$ and $^{\mathrm{s}}\mathrm{IN}^{\mathrm{k}}_{\mathrm{put}}$

Exclude and Include Audio Functions A source can be disabled with `mute`; its complement `solo` disables all non-`solo`ed sources. The semantics of `mute` and `solo` can be described in predicate calculus notation:

$$\text{active}(\text{source}_x) = \neg\text{mute}(\text{source}_x) \;\wedge\; (\exists\, y\ \text{solo}(\text{source}_y) \;\Rightarrow\; \text{solo}(\text{source}_x)) \tag{1.1}$$

As sinks are duals of sources, the semantics of `deafen` and `confide` (& `harken`) are analogous:

$$\text{active}(\text{sink}_x) = \neg\text{deafen}(\text{sink}_x) \;\wedge\; (\exists\, y\ \text{confide}(\text{sink}_y) \;\Rightarrow\; \text{confide}(\text{sink}_x)) \tag{1.2}$$

These two predicates can be described by a generalized representation, using "exclude" to stand for `mute` and `deafen` and "include" to stand for `solo` and `confide` (& `harken`):

$$\text{active}(x) = \neg\text{exclude}(x) \ \wedge \ (\exists\ y\ \text{include}(y) \ \Rightarrow \ \text{include}(x)) \qquad (1.3)$$

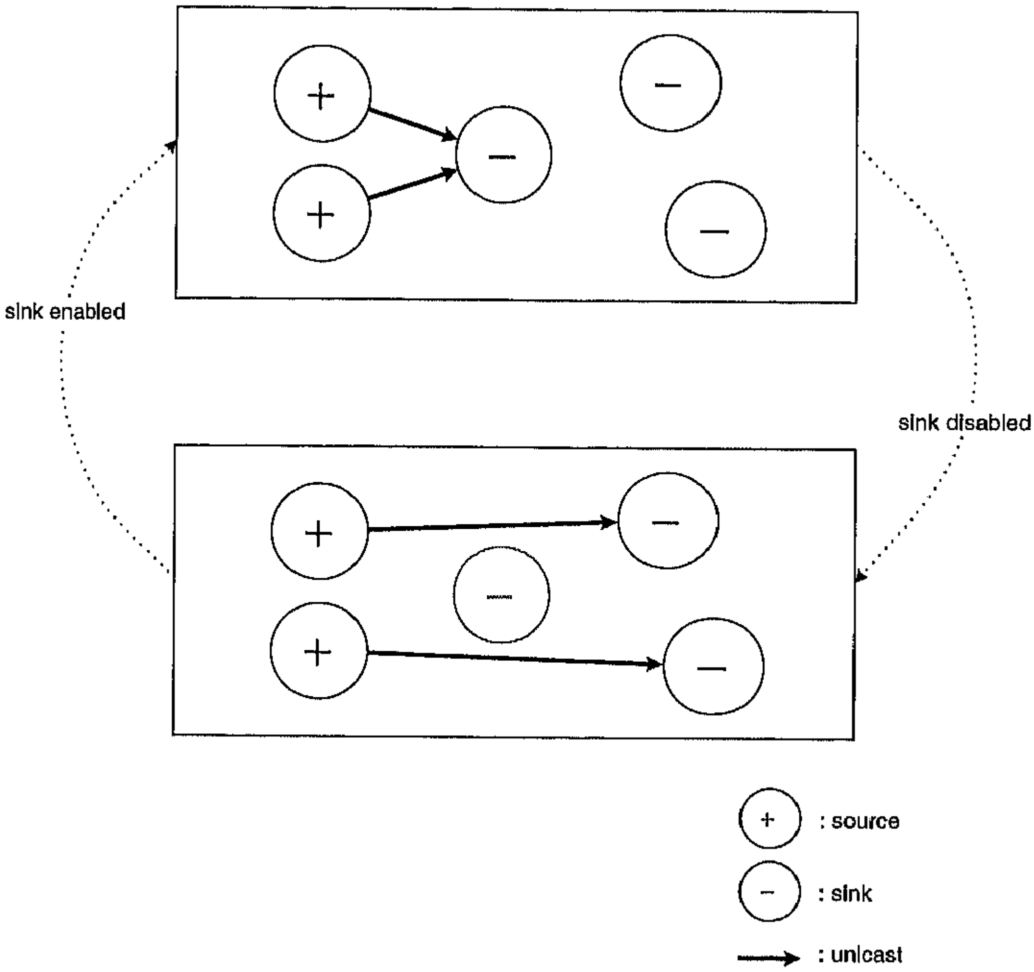


Figure 1.3: Unicast source → sink transmissions: orphaned sources adopted by sinks

Such functions can be applied not only to other users' sinks for privacy, but also to one's own sinks for selective presence [Herder and Cohen, 1996b]. Multiple sinks are useful in both groupware, where a common environment implies social inhibitions to rearranging shared sources like musical voices or conferees, and individual sessions in which spatial arrangement of sources, like the configuration of a concert orchestra, has mnemonic value. As shown by Figure 1.3, an "autofocus" mode can be used to adjust source→sink mappings, depending on sink activation.

User (human pilot)	Delegate (representative, projected presence)
human body	avatar
carbon community	electronic community
RL (real life)	virtual life
	synthespian (synthetic thespian)
motion capture & human actor	vactor (virtual actor)

Table 1.2: User and delegate

Figurative Avatars

Representation of Exclude Audio Functions A human user can be represented in virtual space by one or more avatars, as suggested by Table 1.2. A figurative avatar in virtual space is naturally humanoid, including especially a head, since it not only embodies a center of consciousness, but also important communication organs: ears, mouth, and eyes. Exclude and include source and sink operations can be visually represented by iconic attributes which can distinguish between operations reflexive, invoked by a user associated with a respective icon, and transitive, invoked by another user in the shared environment. Distributed users might typically share spatial aspects of a groupware environment, with attributes like `mute`dness or `deaf`endness determined and displayed on a per-user basis.

For example, as shown in Table 1.3, a source representing a human teleconferee denotes `mute`edness with an iconic hand clasped over its mouth, oriented differently (thumb up or down) depending on whether the source was `mute`d by its owner (or one of its owners) or another, unassociated user. (In the former case, all the users in the space would observe the `mute`, but in the latter, only users disabling the remote source would typically see the `mute`.) An audio muffler could be wrapped around an iconic head to denote its deafness, but to distinguish between self-imposed deafness, invoked by a user whose attention is directed elsewhere, and distally imposed, invoked by a user desiring privacy, hands clasped over the ears should be oriented differently depending on the agent of deafness. As such attributes are orthogonal, simultaneously applied filters could be represented by interpenetrated virtual models.

Representation of Include Audio Functions Include functions (`solo` and `confide`) manifest visually as the respective complementary exclude functions applied to the complement of the appropriate selection. `solo` is implemented as a straight-forward extension of `mute`, effectively muting the

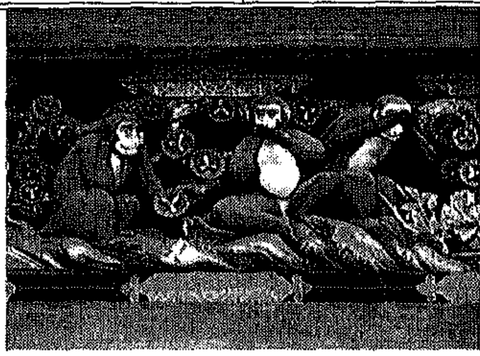
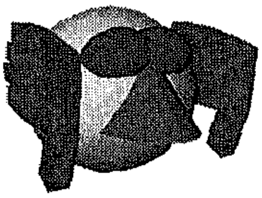
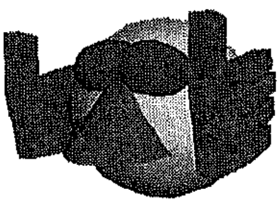
	action	
	deafen (muffle)	`mute` (muzzle)
object	sink	source
own		
	 (thumbs down)	 (thumb up)
other	  (thumbs up)	  (thumb down)

Table 1.3: Deafening and muting one's own and others' avatars

complement of the `soloed` selection, as `confide` (& `harken`) `deafens` the complement of the selection. These pairs of fields are maintained separately, however, anticipating a visual idiom that distinguishes the explicit and implicit operations. For example, a visual spotlight could be used to denote `soloed` sources or `confided` sinks.

avatar	own	other
mode	reflexive	transitive
sink	`deafen` `harken`	`deafen` `confide`
source	`mute` `solo`	`mute` `solo`

Table 1.4: Reflexive and transitive audio exclude and include operations

Groupware Actions along the main diagonal of Tables 1.3 and 1.4, i.e., applied to one's own sinks or to others' sources, manifest locally (in the respective user's spatialization process and soundscape), while actions along the secondary diagonal, i.e., applied to others' sinks or to one's own sources, manifest remotely (in other users' spatialization processes and soundscapes). Such medial attributes do not propagate to distal users' soundscapes; a normal user can `mute` a personally undesired source, but can't prevent its disturbing others.

1.2.3 Sonification

Sonification, auditory visualization, is the computer-based transformation of numeric data into human-interpretable acoustical information [Astheimer, 1995, p. 16–17].

Displays are used to monitor systems. Auditory displays can substitute for or extend graphical displays. A human listener is quite sensitive to changes in audio streams, which makes auditory displays suitable for monitoring tasks over a long period as long as the normal state does not require too much concentration. Monitoring complex processes necessitates multi-dimensional displays. An auditory system for sonification [Kramer, 1994, p. 187] can add several dimensions. Compared to a monoral audio system a sound spatialization system can provide at least three additional dimensions.

Visualization systems using data flow paradigm for modeling and user interface are extended by auditory interfaces for providing sonification [Astheimer, 1995, p. 103–126]. Such systems are easy to handle and inherit all benefits of state of the art visualization systems.

1.2.4 Spatial music — the Helical Keyboard project

Inspired by the cyclical nature of octaves and helical structure of a scale, a model of a piano-style keyboard (see Figure 1.4) was prepared for the helical keyboard project [Herder and Cohen, 1996a] [Herder and Cohen, 1999], which was then geometrically warped into a helicoidal configuration, one octave/revolution, pitch mapped to height. It can be driven by MIDI events, realtime or sequenced, which stream is both synthesized and spatialized by a spatial sound backend. The sound of the respective notes is spatialized with respect to sinks, avatars of the human user, by default in the tube of the helix.

The helical keyboard system is designed to allow, for instance, separate audition of harmony and melody, commonly played by the left and right hands, respectively, on a normal keyboard. Perhaps the most exotic feature of the helical keyboard is the ability to fork one's presence, replicating subject instead of object by installing multiple sinks at arbitrary places around a virtual scene so that, for example, harmony and melody can be separately spatialized, using two heads to normalize the octave; such a technique effectively doubles the helix from the perspective of a single listener. Rather than a symmetric arrangement of the individual helices, we perceptually superimpose them in-phase, coextensively, so that corresponding notes in different octaves are at the same azimuth.

The inherently limited number of sound spatialization resources motivated development of a spatialization resource manager, described in this thesis. To meet the need for better perceiving the large space of harmony and melody, the unique feature of multiple sinks in Section 2.2.3 was introduced. The Helical Keyboard project explores multiple acoustic presence, introduced in [Cohen, 1995] [Cohen and Koizumi, 1995], for three-dimensional space, including manipulations and different visualizations.

1.2.5 Virtual acoustics

Virtual acoustics can be divided into several fields depending on scale and application. In (virtual) room acoustics [Dalenbäck *et al.*, 1996] or architectural acoustics, the goal is to simulate the acoustic of rooms for improving or designing rooms, mainly concert halls. In this thesis, such simulations are

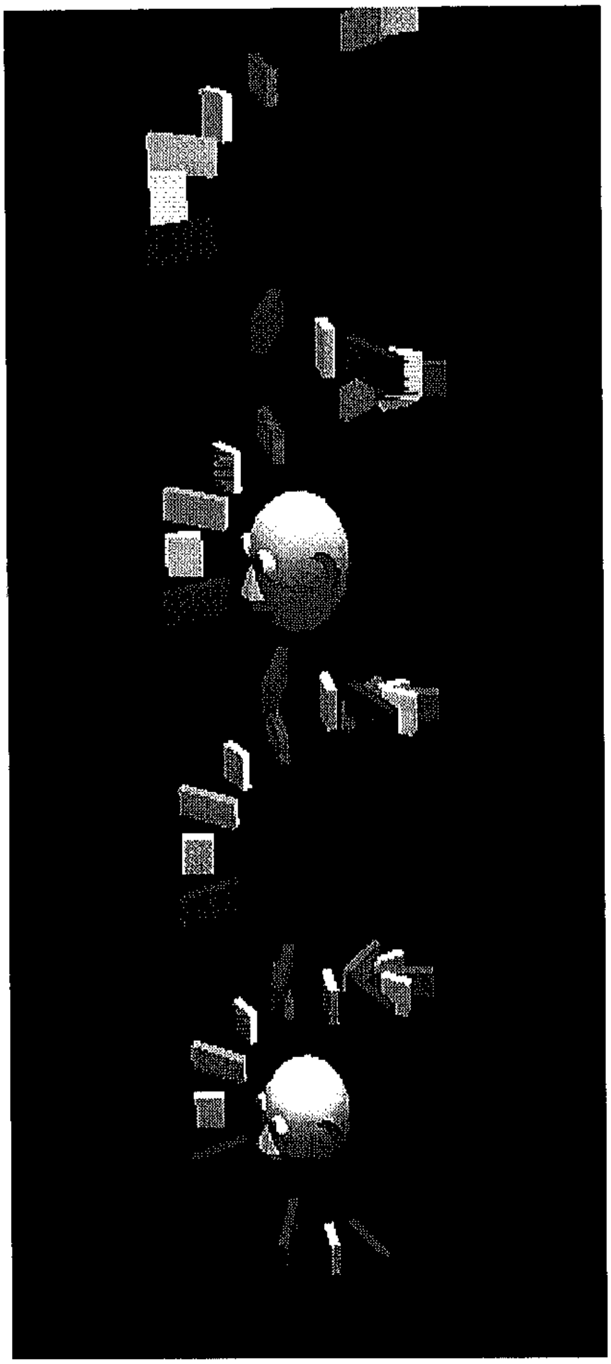


Figure 1.4: Helical Keyboard: an application

handled only briefly, because the involved complexity does not allow real-time applications on currently available systems without neglecting important acoustic properties. Virtual acoustics includes modeling of sound sources at the level of musical instruments like shown in virtual concerts [Cohen and Koizumi, 1995] where sound sources can be arranged. Virtual acoustics goes even further by modeling the physical properties of musical instruments and radiation patterns. Modeling of radiation patterns can be done physically using a set of arranged and adjusted loudspeakers [Caussé *et al.*, 1992] or virtually using source radiation transfer functions [Cook and Trueman, 1998]. It is also reasonable to split the modeling of an instrument into several sound sources (called "elementary sources" in [Karjalainen *et al.*, 1995]), each with its own radiation patterns. The sound of a musical instrument (e.g., clarinet) interacts with the environment (e.g., floor) and depends on the orientation of the instrument (as in Figure 1.5). The overall radiation pattern of a clarinet depends on which tone holes are open.

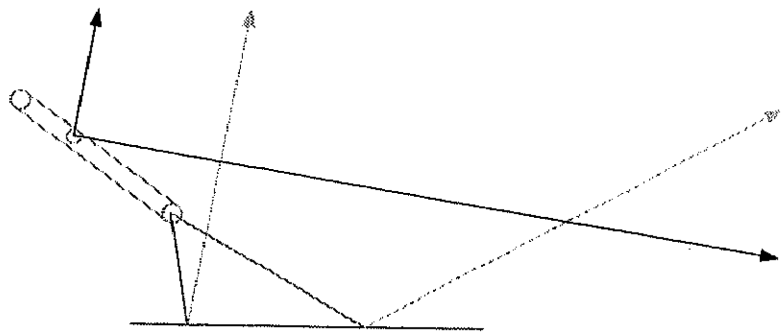


Figure 1.5: Multiple sources can be used to model complex radiation pattern

Physical modeling for sound synthesis is resource demanding. Managing and controlling sound synthesis can be achieved by monitoring resource availability and in case of over load requiring the involved synthesis modules to reduce demand by reducing frequency resolution [Goudeseune and Hamman, 1998].

1.3 Resource management requirements

Sound spatialization quality requirements for the resource management vary depending on the application. This means the resource manager needs to be configurable or adaptable for different applications. The same flexibility is required for the spatialization module (hard or software).

A teleconferencing application might place most sound sources into the horizontal plane. Applications for blind users would give higher priorities to sound sources which are involved with a user. Elementary sound sources of an virtual acoustics application can escalate the number of necessary spatialization resources. The time intervals between which resource assignments take place vary also widely between applications. Musical applications like the Helical Keyboard (or simulation of a clarinet using elementary sound sources) switch resource allocation with each note played. Teleconferencing applications would switch when new participants enter a shared space or the talker changes.

The spatialization might be done in hardware or on a separate system. Then the application requested CPU resources are not effected. Available systems (e.g., Intel RSX [Intel, Inc., 1997]) allow spatialization running on the same system as the application, enabled by simplification of the spatialization algorithms and general hardware improvements. In such a case an application might demand more CPU resources and the spatialization manager must reduce the number of spatialization channels to free CPU cycles. This process has to be done gracefully, minimizing acoustic artifacts that will irritate the user. Similar research is conducted in the field of multimedia (e.g., MPEG [Moser, 1996]).

The spatialization module might be in a server environment, providing service to more than one application. Also, with the introduction of multiple sinks the available spatialization channels vary. The spatialization resource manager must respond dynamically to such resource constraints.

If the computational costs for spatialization management are high, the win in better resource use might be lost or even worse.

The requirements may be summarized as:

- application adaptable;
- spatialization module/device flexible/independent;
- dynamic resource allocation/control;
- low computation costs compared to spatialization and application.

1.4 Outline

This thesis has the following structure. After the introduction of sound spatialization resource management in this chapter, sound spatialization management based on simple geometry is introduced in Chapter 2. The fundamental background is introduced in Chapter 3, starting with sound perception as a criteria for resource allocation, and also describing room simulation and reverberation, including early reflections, and the simulation and management of occlusion. An additional approach to reducing the necessary spatialization channels is optimizing through clustering, presented in Chapter 4. How applications can access the framework through the sound spatialization application programmer interface is presented in Chapter 5. The framework becomes portable by using an abstract spatialization backend interface, described in Chapter 6. The testing and calibration of spatialization backends is discussed in Chapter 7. The framework contains tools for resource allocation monitoring and visualization of the management process, presented in Chapter 8. How to use the framework for creating multimedia content is presented in Chapter 9. Finally, Chapter 10 concludes and extrapolates future trends.

Bibliography

[Aoki *et al.*, 1994] Shigeaki Aoki, Michael Cohen, and Nobuo Koizumi. Design and control of shared conferencing environments for audio telecommunication. *Presence: Teleoperators and Virtual Environments*, 3(1):60–72, 1994.

[Aritsuka and Hataoka, 1997] Toshijuki Aritsuka and Nobuo Hataoka. Gui representation system using spatial sound for visually disabled. In *ASVA'97 — Int. Symposium on Simulation, Visualization and Auralization for Acoustic Research and Education*, pages 415–420, Tokyo, Japan, April 1997. The Acoustical Society of Japan (ASJ).

[Arons, 1992] Barry Arons. A review of the cocktail party effect. *Journal of the American Voice I/O Society*, 12:35–50, July 1992.

[Astheimer, 1995] Peter Astheimer. *Sonifikation numerischer Daten für Visualisierung und Virtuelle Realität.* PhD thesis, Technische Hochschule Darmstadt, April 1995. (In German), ISBN 3-8265-1021-6.

[Barfield *et al.*, 1997] Woodrow Barfield, Michael Cohen, and Craig Rosenberg. Visual, Auditory, and Combined Visual-Auditory Displays for Enhanced Situational Awareness. *The International Journal of Aviation Psychology*, 7(2):123–138, 1997.

[Blattner *et al.*, 1992] Meera M. Blattner, Robert M. Greenberg, and Minao Kamegai. Listening to turbulence: An example of scientific audiolization. In *Multimedia Interface Design*, chapter 6, pages 87–102. ACM Press, 1992.

[Caussé *et al.*, 1992] René Caussé, J. F. Bresciani, and O. Warusfel. Radiation of musical instruments and control of reproduction with loudspeakers. In *International Symposium on Musical Acoustics*, Tokyo, 1992.

[Cohen and Herder, 1998] Michael Cohen and Jens Herder. Symbolic representations of exclude and include for audio sources and sinks: Figurative suggestions of mute/solo & cue and deafen/confide & harken. In *Virtual Environments 98*, pages 95/1–4, Stuttgart, June 1998.

[Cohen and Koizumi, 1995] Michael Cohen and Nobuo Koizumi. Audio Windows for Virtual Concerts II: Sonic Cubism. In Susumu Tachi, editor, *Video Proc. ICAT/VRST: Int. Conf. Artificial Reality and Tele-Existence/Conf. on Virtual Reality Software and Technology*, page 254, Makuhari, Chiba; Japan, November 1995. ACM-SIGCHI (TBD), SICE (Society of Instrument and Control Engineers), JTTAS (Japan Technology Transfer Association), and NIKKEI (Nihon Keizai Shimbun, Inc.).

[Cohen and Koizumi, 1998] Michael Cohen and Nobuo Koizumi. Virtual gain for audio windows. *Presence: Teleoperators and Virtual Environments*, 7(1):53–66, February 1998. ISSN 1054-7460.

[Cohen, 1995] Michael Cohen. Besides immersion: Overlaid points of view and frames of reference; using audio windows to analyze audio scenes. In Susumu Tachi, editor, *Proc. ICAT/VRST: Int. Conf. Artificial Reality and Tele-Existence/Conf. on Virtual Reality Software and Technology*, pages 29–38, Makuhari, Chiba; Japan, November 1995. ACM-SIGCHI (TBD), SICE (Society of Instrument and Control Engineers), JTTAS (Japan Technology Transfer Association), and NIKKEI (Nihon Keizai Shimbun, Inc.).

[Cohen, 1997] Michael Cohen. Exclude and include for audio sources and sinks: Analogs of mute/solo & cue are deafen/confide & harken. In *ICAD: Proc. Int. Conf. Auditory Display*, pages 19–28, Palo Alto, CA, November 1997.

[Cohen, 1998] Michael Cohen. Quantity of presence: Beyond person, number, and pronouns. In Tosiyasu L. Kunii and A. Luciani, editors, *Cyberworlds*, chapter 19, pages 267–286. Springer-Verlag, 1998. ISBN 4-431-70207-5.

[Cook and Trueman, 1998] Perry R. Cook and Dan Trueman. A database of measured musical instrument body radiation impulse responses, and computer applications for exploring and utilizing the measured filter functions. In *International Symposium on Musical Acoustics*, Leavenworth, Washington State, USA, June 1998. Catgut Acoustical Society and Acoustical Society of America.

[Creative Technology Ltd., 1998] Creative Technology Ltd. Why did creative opt to support only 8 accelerated 3D streams? White paper, 1998. `http://www.sblive.com/livenews/3dstreams.html`.

[Dalenbäck *et al.*, 1996] Bengt-Inge Dalenbäck, Mendel Kleiner, and Peter Svensson. Auralization, virtually everywhere. In *the 100th Convention of the AES*, Copenhagen, May 1996. Preprint 4228 (M-3).

[Fouad *et al.*, 1997] Hesham Fouad, James K. Hahn, and James A. Ballas. Perceptually based scheduling algorithms for real-time synthesis of complex sonic environments. In Steven P. Frysinger and Gregory Kramer, editors, ICAD*'97 — Int. Conf. on Auditory Display*, pages 77–81, Palo Alto, CA; USA, November 1997.

[Goudeseune and Hamman, 1998] Camille Goudeseune and Michael Hamman. A real-time audio scheduler for pentium pcs. In *Proceedings of the 1998 International Computer Music Conference (ICMC 98)*, pages 158–162, Ann Arbor, October 1998.

[Hawley *et al.*, 1996] Monica L. Hawley, Ruth Y. Litovsky, Leah B. Dunton, Jennifer K. Jones, and H. Steven Colburn. Benefit of binaural hearing in a multi-source environment: Speech intelligibility. *J. Acous. Soc. Amer.*, 99(4):2596, April 1996. 131st Meeting: Acoustical Society of America.

[Herder and Cohen, 1996a] Jens Herder and Michael Cohen. Design of a Helical Keyboard. In Steven P. Frysinger and Gregory Kramer, editors, ICAD*'96 — Int. Conf. on Auditory Display*, Palo Alto, CA; USA, November 1996.

[Herder and Cohen, 1996b] Jens Herder and Michael Cohen. Project report: Design of a helical keyboard. In *ICAD: Proc. Int. Conf. Auditory Display*, Palo Alto, CA, November 1996.

[Herder and Cohen, 1997] Jens Herder and Michael Cohen. Sound Spatialization Resource Management in Virtual Reality Environments. In *ASVA'97 — Int. Symp. on Simulation, Visualization and Auralization for*

Acoustic Research and Education, pages 407–414, Tokyo, Japan, April 1997. The Acoustical Society of Japan (ASJ).

[Herder and Cohen, 1999] Jens Herder and Michael Cohen. The Helical Keyboard — Another Perspective for Virtual Reality and Music. *IJVR - International Journal for Virtual Reality*, 1999. In press.

[Herder, 1998] Jens Herder. Sound spatialization framework: An audio toolkit for virtual environments. *Journal of the 3D-Forum Society, Japan*, 12(9):17–22, September 1998.

[Huopaniemi *et al.*, 1996] Jyri Huopaniemi, Lauri Savioja, and Tapio Takala. Diva virtual audio relative system. In Steven P. Frysinger and Gregory Kramer, editors, ICAD *'96 — Int. Conf. on Auditory Display*, pages 111–116, Palo Alto, CA; USA, November 1996.

[Intel, Inc., 1997] Intel, Inc. Intel Realistic Sound Experience (3D RSX). White paper, 1997. `http://developer.intel.com/ial/rsx/WPAPER.HTM`.

[Karjalainen *et al.*, 1995] Matti Karjalainen, Jyri Huopaniemi, and Vesa Välimaäki. Direction-dependent physical modeling of musical instruments. In *15th International Congress on Acoustics (ICA'95)*, Trondheim, Norway, June 1995.

[Kashino and Hirahara, 1996] Makio Kashino and Tatsuya Hirahara. One, two, many — judging the number of concurrent talkers. *J. Acous. Soc. Amer.*, 99(4):2596, April 1996. 131st Meeting: Acoustical Society of America.

[Kramer, 1994] Gregory Kramer. Some organizing principles for representing data with sound. In Gregory Kramer, editor, *Auditory Display*, volume XVIII of *SFI Studies in the Sciences of Complexity*, pages 185–221. Addison-Wesley, 1994.

[Martens and Myszkowski, 1998] William L. Martens and Karol Myszkowski. Psychophisical validation of the visible differences predictor for global illumination applications. In C. M. Wittenbrink & A. Varshney, editor, *Late Breaking Hot Topics Proceedings (IEEE Visualization '98)*, pages 49–52. IEEE Computer Society, October 1998.

[Martin, 1995] Keith Dana Martin. A computational model of spatial hearing. Master's thesis, Massachusetts Institute of Technology, June 1995.

[Meinard, 1997] Elizabeth D. Meinard. Creating auditory interfaces from graphical interfaces. In *ASVA'97 — Int. Symposium on Simulation, Visualization and Auralization for Acoustic Research and Education*, pages 123–126, Tokyo, April 1997. The Acoustical Society of Japan (ASJ).

[Moser, 1996] Martin Moser. Declarative Scheduling for Optimally Graceful QoS Degradation. In IEEE *Multimedia Systems '96*, Hiroshima; Japan, June 1996. IEEE.

[Ruspini and Khatib, 1998] Diego Ruspini and Oussama Khatib. Acoustic cues for haptic rendering systems. In *Proceedings of the third PHANToM User Group Workshop*, Dedham, MA, USA, October 1998.

[Savioja *et al.*, 1997] Lauri Savioja, Jyri Huopaniemi, Tapio Lokki, and Riitta Väänänen. Virtual Environment Simulation — Advances in the Diva Project. In ICAD*'97 — Int. Conf. on Auditory Display*, pages 43–46, Palo Alto, CA; USA, 1997.

[Todd *et al.*, 1994] Craig C. Todd, Grant A. Davidson, Mark F. Davis, Louis D. Fiedler, Brian D. Link, and Steve Vernon. AC-3: Flexible perceptual coding for audio transmission and storage. In *Proc. 96th Convention of the AES*, February 1994. Preprint 3796.