

The Helical Keyboard: Perspectives for Spatial Auditory Displays and Visual Music

Jens Herder
Department of Media
FH Düsseldorf, University of Applied Sciences
Josef-Gockeln-Str. 9
40474 Düsseldorf; Germany
voice: [+49](211)4351-800
fax: [+49](211)4351-803
e-mail: herder@fh-duesseldorf.de

Michael Cohen
Spatial Media Group
University of Aizu
Aizu-Wakamatsu
Fukushima-ken 965-8580; Japan
voice: [+81](242)37-2537
fax: [+81](242)37-2549
e-mail: mcohen@u-aizu.ac.jp

Abstract

Auditory displays with the ability to dynamically spatialize virtual sound sources under realtime conditions enable advanced applications for art and music. The listener can be deeply immersed while interacting and participating in the experience. We review some of those applications while focusing on the Helical Keyboard project and discussing the required technology.

Inspired by the cyclical nature of octaves and helical structure of a scale, a model of a piano-style keyboard was prepared, which was then geometrically warped into a helicoidal configuration, one octave/revolution, pitch mapped to height and chroma. It can be driven by MIDI events, realtime or sequenced, which stream is both synthesized and spatialized by a spatial sound display. 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. Alternative coloring schemes can be applied, including a color map compatible with chromastereoptic eyewear. The graphical display animates polygons, interpolating between the notes of a chord across the tube of the helix. Recognition of simple chords allows directionalization of all the notes of a major triad from the position of its musical root.

The 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 interface 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, they are perceptually superimposed in-phase, co-

extensively, so that corresponding notes in different registers are at the same azimuth.

Keywords: spatial auditory displays, virtual reality, computer music, spatialization, interaction, visual music.

1 Spatial Symphonies, Theatre, and Installations

Both space and audience have a “compile-time” and a “run-time” influence on the creation of music. Composers have written pieces targeting specific performance situations. While recordings allow a larger audience to appreciate certain music, even with the latest technology, recordings are not comparable to a live concert: a recording cannot reproduce the interaction between audience, space, and performers. Auditory displays with the ability to place virtual sound sources into space under realtime conditions allow interaction between audience and space. We cite some examples of spatially arranged music elements in Section 1.1. We consider some new applications in the field of art using spatial auditory display in Section 1.2, and review virtual concert applications in Section 1.3. Beginning with Section 2, we present the Helical Keyboard project, including system description and implementation.

1.1 Using spatially arranged musical elements in composition

Composers take advantage of spatial imagery and try to immerse the audience. Specially arranged choirs in a church, for example, might straddle the congregation to highlight antiphonal hymns. To some well-known examples:

- In 1829–1830, Hector Berlioz wrote the “Symphonie Fantastique” [Berlioz, 1830], in which spatial arrangements are an important part of the music. In the “Scene in the Country,” two shepherds pipe in the distance to each other: in concert, pipe musicians are separated from the main orchestra, creating wonderful spacious imagery [New York Philharmonic Orchestra and Mehta, 1980].
- Karlheinz Stockhausen [Stockhausen, 1956] used a multichannel recording and playback system in his “Gesang der Jünglinge,” to create spatial movements.
- Andrew Lloyd Webber’s musical “Cats” [Webbers, 1981] plays at specially arranged theaters (e.g., Hamburg) in which the stage surrounds the seats and performers move freely close to and behind the audience.
- The Audium¹ is a specially-constructed theater featuring sound dancing around its 169 speakers in an intimate (50 seats) space [Loy, 1985] in San Francisco.
- “The House on the Rock”² features a collection of room-sized musical robots.

¹ | www.audium.org—

² | www.houseontherock.com/the_attraction.htm—

Such theaters and automata are charming— taking “theatre-in-the-round” and turning it inside-out, and enhancing the ability to “hear out” instruments, but impractical for anything besides special-purpose venues and LBE (location-based entertainment). A much better review of the history of music space and reverberation in general is given by Barry Blesser [Blesser, 2001].

While each listener has individual experiences, his or her interaction or participating in the performance is usually limited. Art installations, reviewed in the next section, go beyond such restrictions.

1.2 Interactive installations

Virtual environments are influencing art [Dolinsky, 1997] [Brown, 1997] and new technology is used for various installations.

- Feedback and complete isolation are explored in an installation by Seiko Mikami [Mikami, 1997]. The heartbeat of the visitor is spatially auralized and visualized, in the context of an anechoic chamber providing total isolation. The sound wanders around the listener and depends on her/his reaction.
- Distance plays an important role in an installation by Akitsugu Maebayashi [Maebayashi, 1997]. Distance between people is visualized and auralized via a personal spatial auditory display. Participants wander around a virtual (controlled) environment, wearing an HMD (head-mounted display), location tracker, and stereo earphones.
- Concrete location of sound is not always desired. In an interactive installation by Yunchul Kim [Yunchul, 2001], the sound moves and is influenced by the visitors’ movements (i.e., position), but the sound seems not come from a specific location: instead an omnipresent floating soundscape is created. A video-based person tracking controls a realtime composition using self-organizing maps.

1.3 Spatial emancipation of the audience

With spatial auditory displays allowing realtime composition of the experience, an audience can emancipate itself from prefabricated (spatially composed) content. Virtual concerts are possible with multiple audio windows [Cohen, 1994] [Cohen and Koizumi, 1995], in which a listener can rearrange (virtual) sound sources and sinks. Virtual concerts are also presentable in virtual environments [Chodura and Kaup, 1999], allowing a participant to take active role by contributing a source. With numerous tracks available on DVD and tagging of voices, and consumer spatial auditory display becoming available, virtual concerts can become another common listening experience [Herder, 2000].

2 The Helical Keyboard

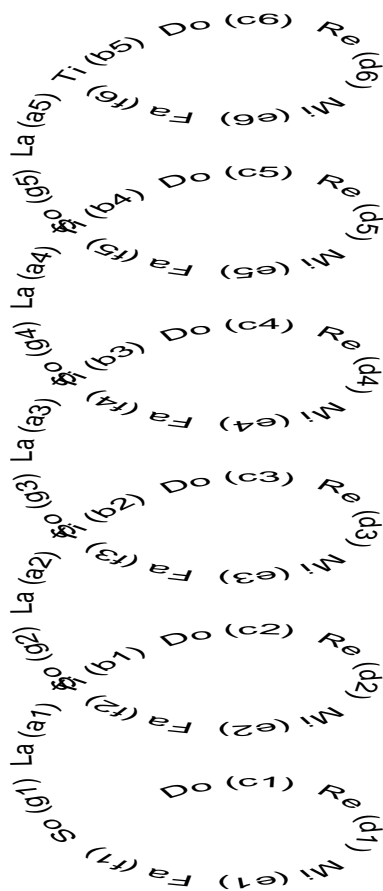


Figure 1: Helical structure of a (C major or A minor) scale

A Helical Keyboard can visualize and auralize the cyclical nature of octaves and helical structure of a scale, as described by [Shepard, 1983] [Shepard, 1984]. Shepard describes the perception of musical pitch as representable by a 5-dimensional manifold, the first two dimensions of which form a helix:

1. the linear dimension (simplex), called tone height: pitch gets higher as fundamental frequency increases, and
2. the circular dimension (circumplex), called tone chroma: pitch goes around in cycles (one revolution per octave) as fundamental frequency increases.

Figure 1 illustrates the logical structure of such a left-handed helix. Pitch is mapped to elevation (simplex) and azimuth (circumplex), exploiting the sense of both pitch height and circular dimension of pitch class (which implies perceptual

equivalence of the same tones in different octaves). Note-specific sound spatialization enables a unique style of musical appreciation and voice discrimination.

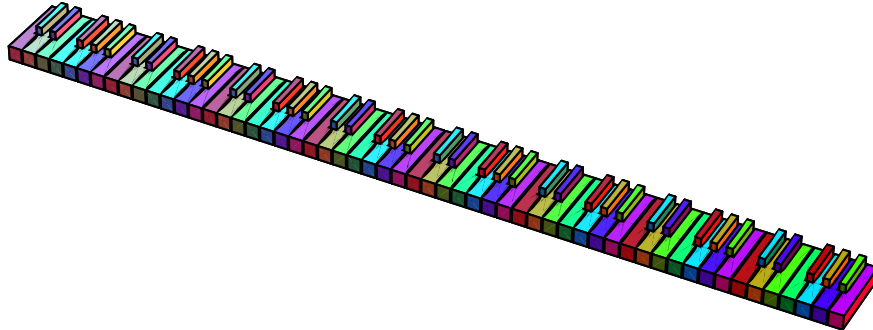


Figure 2: Color wheel unrolled across normal keyboard

If a traditional keyboard has its black and white keys replaced by hues indicated by a color wheel rolled across the length of the keyboard, it would look like Figure 2. This same keyboard, wrapped into a helix and viewed from above, is shown in Figure 3, emphasizing octave equivalence classes.

Benefits from using virtual reality technology Not only are new applications made possible in virtual reality as cost factors are reduced, but also virtual reality is influencing “real” art [Brown, 1997] [Dolinsky, 1997]. The objectives of the Helical Keyboard project were several, including pedagogical (illustrating musical concepts), artistic (towards visual music), and technical (developing an extensible infrastructure capable of supporting such applications). The main engineering challenge was the effective coupling of synthesis, animation, and spatialization of musical sounds. To this end, a software framework was developed which could easily support this and similar applications, including a modular structure capable of receiving input streams asynchronously (from sequenced MIDI files) or synchronously (from realtime MIDI controllers), and driving heterogeneous spatialization backends (including DSP-driven headphones, transaural speakers, and speaker arrays).

The idea of the helical structure of music, has been traditionally demonstrated using sound, a static diagram, accompanied by a verbal (spoken or written) description. If visualization were presented in form of a video using animation techniques, interactivity (choice of music, viewpoint, dynamic spatialization [with a tracker], etc.) would be sacrificed. Spatialization of the sound could also have been realized directly using a physical structure of a helix with 88 loudspeakers and control equipment plus a walkway for users to move around. In both cases, using video

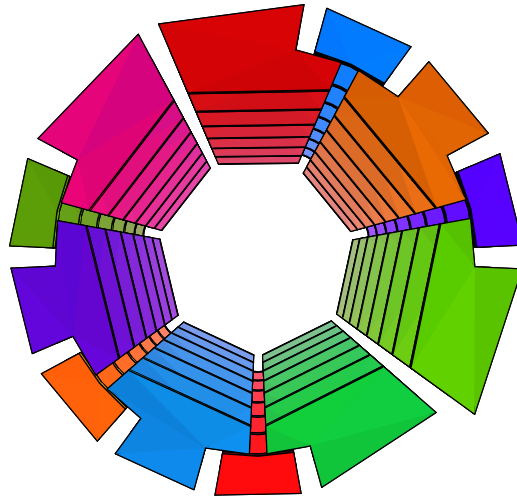


Figure 3: Discontinuous spiraling color wheel (view from above helicoid)

or physical installation, flexibility would be lost, and experiments like changing the dimensions, introducing multiple listener positions, changing the scale (to pentatonic, for instance, using only five notes/octave), etc., become time-consuming and prohibitively expensive. Virtual reality technology enables experiments which heretofore could not have been practically designed.

In the course of this research, we developed and applied several new techniques. The inherently limited number of sound spatialization resources motivated development of a spatialization resource manager, described in Section 5.3. To meet the need for better perceiving the large space of harmony and melody, we introduce the unique feature of multiple sinks, described in Section 5.4. The Helical Keyboard project explores multiple acoustic presence, introduced in [Cohen, 1995] [Cohen and Koizumi, 1995], including manipulations and different visualizations. In Section 6.2, a very affordable technique to get a three dimensional (visual) impression is described. Group demonstrations encouraged development of a realtime camera control system, which automates the animation. The lack of authoring and manipulation tools for three dimensional soundscapes encouraged development of a soundscape deformer, described in Section 7.

While other researchers (including [Pressing, 1997]) are investigating virtual reality, visual music, and spatial sound, our project is unique in combining these separate threads.

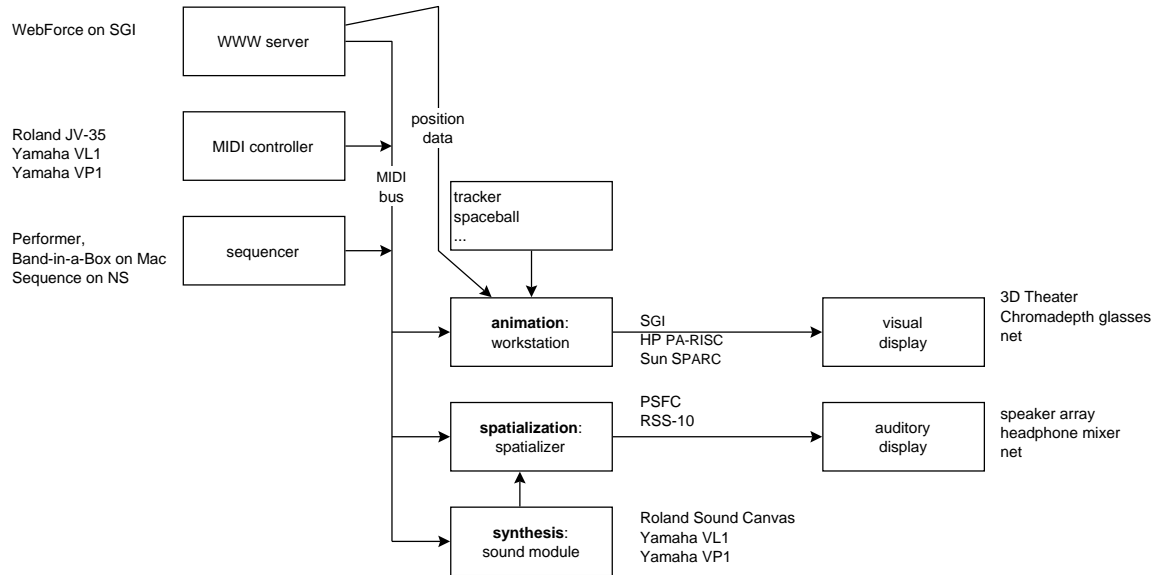


Figure 4: Global architecture: animation, spatialization, and synthesis

3 System architecture

A data stream, like MIDI from a sequencer or generated by a realtime controller, can be used thrice simultaneously (as shown in Figure 4):

graphical animation, visualized by a high-performance virtual reality workstation, including SGI machines like the Onyx Reality engine. Visual options include dynamic viewpoint and optional stereographic output,³ for viewing with LCD⁴-shuttered, polarized, anaglyphic, or chromastereoptic eyewear;

musical rendering, interpreted by a software (JavaSound or QuickTime for Java), hardware (sound module, sampler, or tone generator) synthesizer, and performed through a set of conceptually monophonic channels;

spatialization, used to directionalize the audio stream so that the respective notes revolve about the listener. The helical keyboard uses a variety of spatial sound backends, both DSP convolution engines for headphone presentation and the Pioneer Sound Field Controller (PSFC) speaker array (in our University-Business Innovation Center⁵), driven by derivatives of audio windowing systems,⁶ directional mixing consoles.

³www.u-aizu.ac.jp/~mcohen/spatial-media/stereograms.html

⁴liquid crystal display: featuring active switching between left and right images, synchronized via infrared signals with interlaced frame display

⁵www.mmc-aizu.pref.fukushima.jp/

⁶www.u-aizu.ac.jp/~mcohen/spatial-media/audio-windows/

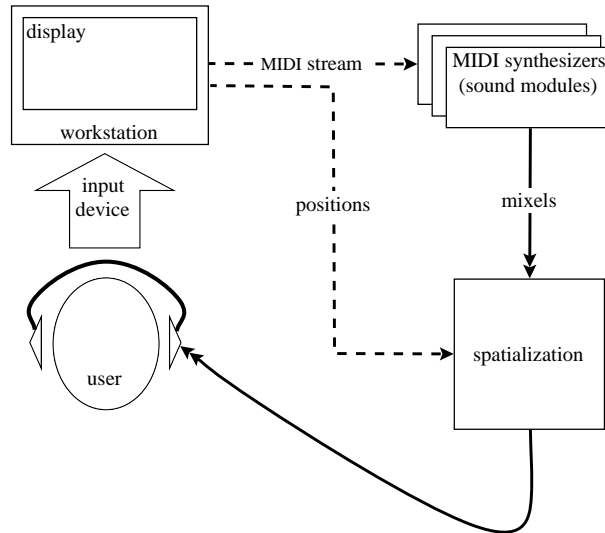


Figure 5: System configuration. (Dashed lines indicate digital data, and solid lines indicate analog signals.)

A schematic of a typical system configuration with the main components is shown in Figure 5.

4 Implementation

The Helical Keyboard [?] [Kanno and Cohen, 2001] is defined algorithmically (and not as a traditional CAD model) [Cohen, 2001]. The geometry and warping of the Helical Keyboard were modeled in Mathematica [Wolfram, 1999] and then exported to VRML 1.0 [Bell et al., 1995]. For simplicity and to minimize download time by exploiting repetition in a keyboard, five different style key styles were defined, so that the keyboard is an 88-key collection of instances of these five kinds of keys, appropriately translated, rotated, colored, and named. The helix has left-handed chirality, so that notes ascend as they cross in front of a naturally oriented sink in the tube of the helix. Our prototype was deployed on an SGI Indigo 2 Extreme, connected to an Acoustetron II from Aureal/Crystal River Engineering and Roland SC-55 Sound Modules. The Open Inventor graphics toolkit [Wernecke, 1994] was expanded for classes (nodes) to support spatial sound extensions, which follow the VRML 2.0 specification [Bell et al., 1996], which was used as an application programmer interface. Open Inventor 2.x is a superset of the VRML 1.0 standard, which does not support sound or dynamic behavior of objects.

Table 1: $^s\text{OU}_{\text{Tput}}^{\text{rce}}$ and $^s\text{IN}_{\text{put}}^{\text{k}}$

	Role	
	Source	Sink
Function	radiation	reception
Level	amplification	sensitivity
Direction	OUTput	INput
Instance	speaker (human or loud-)	listener (human or dummy-head)

5 Sound Spatialization

5.1 Sources and sinks

As outlined by Table 1, since the word ‘speaker’ is ambiguously overloaded, meaning both ‘loudspeaker’ and ‘talker,’ this paper uses “source” to mean both, a logical sound emitter. Similarly and symmetrically, “sink” is used to describe a virtual listener, a logical sound receiver and delegate of a human listener in a virtual room.

5.2 Mixels

In the context of a concert, individually recording (or mic-ing or synthesizing) individual instruments, presenting each of the channels to a spatial sound backend, and mixing them at audition time, rather than in “post-production” (as tracks and sub-groups), allows the instruments to be rearranged by the listener [Starks and Linton, 1994]. Atomic sources can be called “mixels”— acronymic for ‘[sound] **m**ixing **e**lements,’ in analogy to hogels (holographic elements in holography) pixels (picture elements), taxels (tactile elements), texels (texture elements), and voxels (volumetric elements, a.k.a. boxels)— as they form the raster across which a soundscape is projected, defining the granularity of control and degree of spatial polyphony. In the Helical Keyboard, each key is a potential sound source, which must be multiplexed through the mixels provided by the spatial sound processor, as mediated by the resource manager described in the next section.

5.3 Spatialization resource manager

Any practical system has only limited resources, including spatialization channels (mixels), MIDI/audio channels, and processing power. A sound spatialization resource manager [Herder and Cohen, 1997] controls audio channels and optimizes fidelity (presence) under given conditions. For that, a priority scheme based on psychoacoustics is needed. Parameters for spatialization priority might include intensity calculated from volume and distance, orientation for non-uniform radiation patterns, occluding objects, frequency spectra (low frequencies are harder to localize), expected activity, and others. Objects which are spatially or perceptually close together (depending on distance and direction) can be mixed together as a

clustered mixel. Extra sources that can not be spatialized separately can be mixed as ambient sources. Important for resource management is resource assignment, i.e., minimizing swap operations, which makes it desirable to look-ahead and predict upcoming events in a scene. Prediction is achieved by monitoring objects' position, speed, and past evaluation values (priorities, probabilities, ...).

5.4 Multiple sinks

Multiple sinks allow forked presence in auditory space. This is being like being in more than one place at once, a concept familiar from teleconferencing and studio recording. Two techniques have been suggested for disambiguating the paradoxes of multiple presence. One method is to “autofocus,” logically partitioning the sources across collocated sinks, in which case the requested number of mixels does not exceed the total number of sources. A second method simply adds the overlaid impressions, crossing all sources and sinks, and requiring that product of mixels.

Using the first method is like overlaying multiple soundscapes (e.g., the harmony and melody part of the keyboard). This allows the azimuth of the sound location to be perceived as note attribute even though a single space would not allow it. Because of perspective, the cylindrical helix would seem conical (i.e., if the sink were towards the bottom of the helix, a note played high above would seem to come from near the upper pole, azimuthal discriminability having been lost). As a virtual head clones, perception forks, and if one of multiple heads moves up the tube of the helix, azimuthal acuity increases, since the perceived direction of respective notes move away from the upper pole closer to the soundscape horizon.

Instead of building a MIDI driver for a convolution engine, our sound manager parses positions in the stream, integrates these updates with its own model of the distribution of sources and sinks in a given room, determines the best sink for each source, and calculates the gain and selects the spatialization function, which parameters are finally passed to the backend (e.g., convolution engine).

Analytical cubism, as developed by Braque or Picasso, presents multiple visual perspectives on a scene simultaneously. An exocentric visual perspective of forked presence via multiple sinks is shown in Figure 6, perceiving normalized melody and harmony with separate sinks, superimposing the top and bottom halves of the soundscape.

6 Graphical presentation

Besides a totally immersive (hereafter called “egocentric”) view, the virtual reality environment we developed also provides an exocentric viewing mode. These two modes can be thought of in analogy to “world” and “model” navigation/manipulation modes in VRML browsers [Bell et al., 1995, p. 73] [Barrilleaux, 2001, p. 127–9].

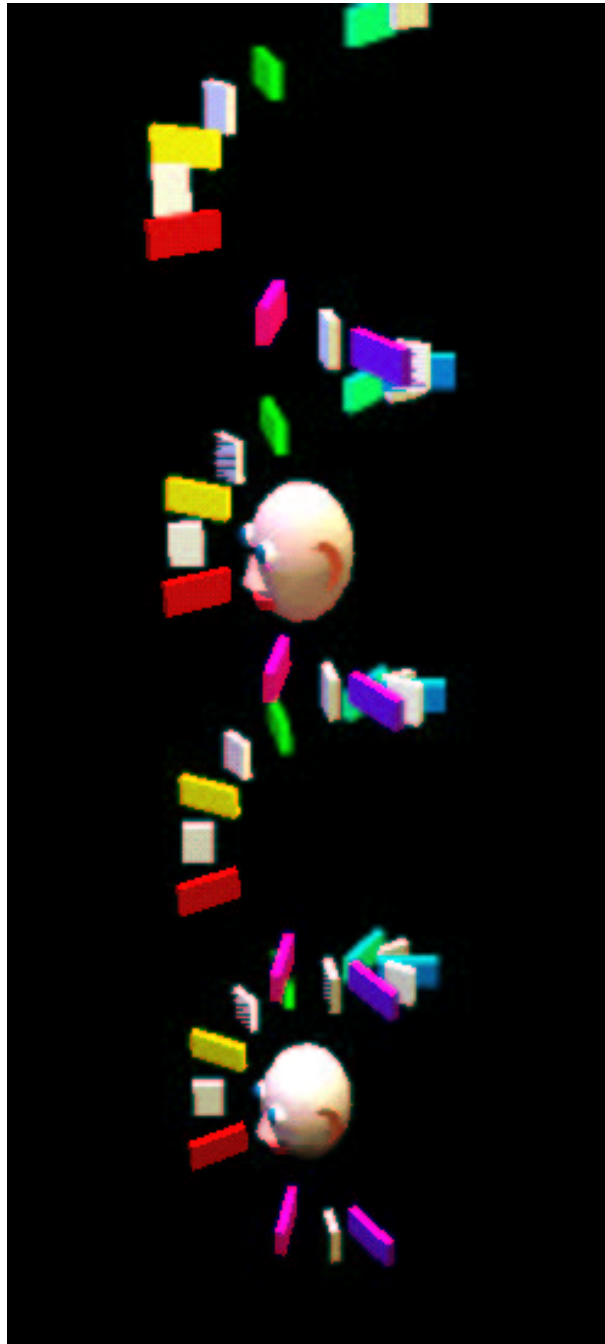


Figure 6: Exocentric visual perspective of forked presence via multiple sinks

Table 2: Points of View

point of view	person	intimacy	object	distance	mode	perspective	navigation
exocentric vicariousness, empathy	3 rd 2 nd	public social, multipersonal	other familiar	distal medial	transitive imperative	objective	model
telepresence, autoempathy immersive egocentric	1 st	remote self personal	self	proximal	reflexive	subjective	world

Our system allows multiple views simultaneously, including runtime switching between ego- and exocentric modes.

In an egocentric mode, a sound sink is attached to an active camera controlled by a manipulator, with which a user interacts directly. If the camera location and/or orientation change, the sound sink position is changed identically. In an exocentric mode, camera and sound sink are decoupled; changes of the camera do not affect the sound sink, but another egocentric view (in a separate window) might. In such a mode, the sound sink is visualized as the head of an avatar, which is invisible when associated with an egocentric mode (i.e., one’s own sink is not visible).

The application of blurred egocentric and exocentric views is informed by their deployment in games like flight and driving simulators, as in Figure 7, allowing users to seamlessly switch between perspectives. Such perspectives can be classified further as in Table 2, which relates them to social intimacy.

All sound sinks can be controlled by a global tabular numeric panel (shown in Figure 8), which is part of the sound spatialization manager. Attributes like `attached`, `active`, `position`, and `orientation` can be freely modified. The `attached` toggle controls the coupling of user interaction with the current viewpoint to the sink. This allows one to separate camera motion from listener position and disable coupled control of visual and acoustical instantiation. If the `active` flag is enabled, equivalent to a “cue” toggle on a mixer, the respective sink contributes to the soundscape. The `location` and `orientation` fields control the sink separately from the camera control.

6.1 Camera control

Navigation in virtual reality environments is difficult, especially for inexperienced users [He et al., 1996]. Automatic steering could help to overcome this problem and can contribute to animation in virtual reality without predefined camera control. Further motivation for automatic camera control includes VR (virtual reality) systems in which demonstrations run without requiring user interaction (see Figure 13). Commercial applications appear in VR games (e.g., replay of a sequence with good view). An event model provides a framework for automatic camera con-



(a) Drive mode (1st person view)



(b) Follow mode (2nd person view)

Figure 7: Sega Virtua Racing

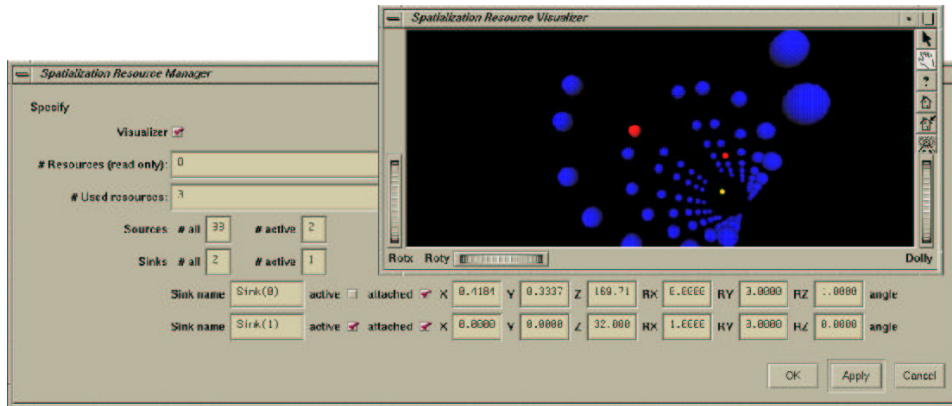


Figure 8: Spatialization resource manager control panel

trol. Objects (i.e., keys) in a scene can produce events (e.g., `note on`, etc.). By monitoring the event system, an automatic camera can decide to focus on or follow an active object. A viewer presents the following options to the user:

predefined cameras Defines scene positions which can be individually selected.

smooth motion Instead of skipping discretely to a new viewpoint, moves the camera continuously to the next viewpoint via spatial interpolation. Acceleration is bounded to softly depart and arrive.

shuffle Randomly selects predefined cameras in the scene.

point at Changes the camera orientation to automatically point at active objects.

focus on Changes the camera orientation, using a “dolly” to point at an active object and regard it.

Different options may be applied together. For example, a combined mode could include `shuffle`, `smooth motion`, and `point at`, causing the camera to move in a smooth animation between predefined camera positions while pointing at active keys. (Imagine a mother walking through a supermarket with her child in a trailer: The child looks at points of interest but cannot change his/her position, which is controlled by the mother.) Two processes can even compete against each other (e.g., `focus on` and `shuffle`).

6.2 Depth perception using color coding

Stereo computer graphics are used to create depth perception from two-dimensional media like a screen; methods and costs vary [McAllister, 1993]. Chromastereoscopy [Steenblik, 1993] uses color cues to provide the illusion of depth when the scene is viewed through inexpensive prismatic eyewear.⁷

Figure 9 shows a Helical Keyboard rendered using such a chromastereoptic scheme. (For a 3D impression, special eyewear is necessary.) One obvious disadvantage of such a method is the loss of any other color coding dimensions, like those employed by Figure 3.

6.3 Chord-chords and Chord-kites

Closely-spaced MIDI note events are not monophonic. When multiple notes are played simultaneously, the Helical Keyboard application detects the note numbers of MIDI events, then draws a “chord chord” or “chord kite” [Nakashima and Cohen, 2001], as illustrated by Figure 10.

⁷ | www.chromatek.com—

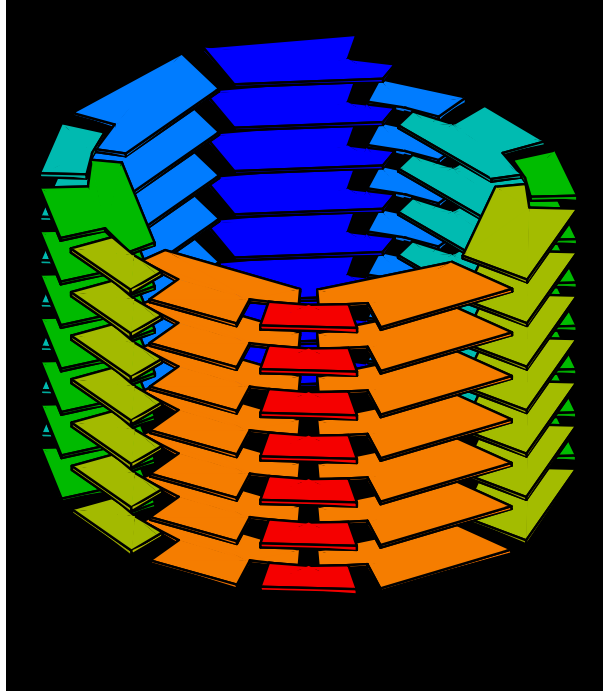


Figure 9: Chromastereoptic rendering

6.3.1 Chord Recognition

There are many kinds of chords, but the Helical Keyboard application can recognize only major triads (in root position, first inversion, or second inversion) but no minor chords, 7ths, etc. The algorithm is as follows:

1. Every key object is tagged by its corresponding MIDI note number. (For example, E₄ is 52, G₄ is 55, and C₅ is 60.) Let S1 be the set of note numbers played.
2. Let $x = \min(S1)$.
3. Reduce every element of S to its modulo 12 value using x as the 0 point; replace every value y in S with $\text{modulo}(y-x,12)$, so that x goes to 0, and other values range from 0 to 11. Call the new set formed S2.
4. If S2 contains more or less than three elements, there is no triad; stop.
5. Otherwise S2=(0, m, n) such that $n > m$. If m=4 and n=7, we have a major triad in root position. If m=3 and n=8, we have a major triad in first inversion. If m=5 and n=9, we have a major triad in second inversion. If any other combination, there is not a major triad.

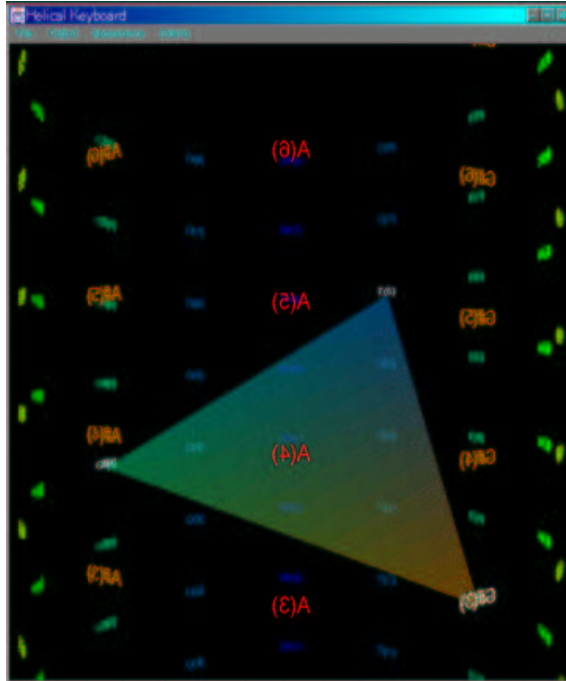


Figure 10: Helical keyboard drawing a triangular “chord-kite,” interpolating $G\#_3$, $C\#_4$, and E_6 .

Applying the above algorithm, if E_4 , G_4 , and C_5 are played, x of this set is 52, and the set S_2 is (0, 3, 8). So the chord is characterized as “major triad in first inversion,” and directionalized in its entirety from the root C_5 location.

7 Soundscape manipulation

A major aspect of the design of the Helical Keyboard was the soundscape. All keys were intended to be differentiable by location. The helix itself became geometrically a very tall object. If the sound sink is placed in the center, then keys playing in the far upper part or lower part could not be well differentiated by azimuth, and also the volume for them was too low. As a solution to these problems we developed two techniques. Multiple sinks (described in Section 5.4) allow a user to have presence in multiple places simultaneously by overlaying the soundscape. Another approach separates the scaling of the audio and visual representations, motivating the development of a soundscape deformer, a 3D widget that controls the mapping between scene space \rightarrow soundscape [Herder, 1998]. The scene space can be shifted around, inducing a translation of the soundscape. In that regard the soundscape deformer can be seen as a generalization of stereo panning for 3D.

The soundscape deformer, shown in Figure 11, provides a visual representation of the mapping. A sphere in the center represents the case in which the

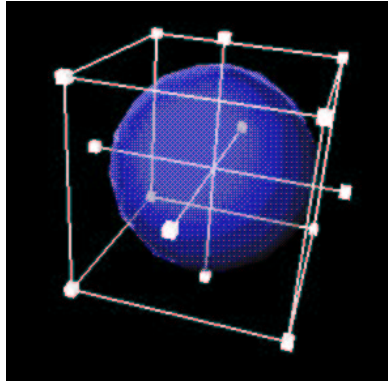


Figure 11: Default (no distortion)

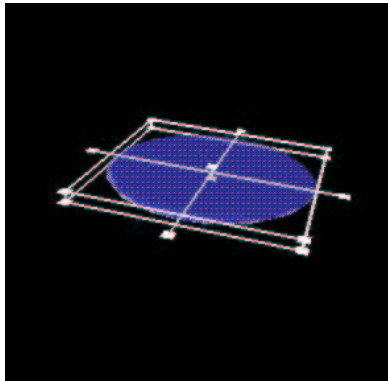


Figure 12: Reducing height, compressing vertical space

scene space is directly mapped to the soundscape. Figure 12 shows the soundscape reduced in height, flattening the spatial audio position of all keys to a plane (sometimes called “pantophonic” reproduction [Everest and Streicher, 1998, Section 13.15–18], in contrast with “periphonic” reproduction that also has vertical directional information).

8 Spatially Immersive Deployment

The PSFC, or **Pioneer Sound Field Control System** [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 about half the speakers are pictured in Figure 13, while the system layout of the PSFC is illustrated by Figure 14.

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 range. 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 wide-screen visual presentation. The spatially immersive environment provides a natural group experience, broad 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 navigation and exploration of an environment.

The PSFC’s omnidirectional sound field complements the exaggerated visual display’s inadequacies for immersion, since flying into the helix from the outside envelopes the listener as it overflows the visual system’s field-of-view. The Helical Keyboard’s multiple sinks are natural idioms for the PSFC, compensating for the fact that the speakers span the ceiling but not the floor; multiple sinks can be used to normalize the register (e.g., one sink for harmony and one for melody, as shown in Figure 6), so that all sound images are in the upper hemisphere.

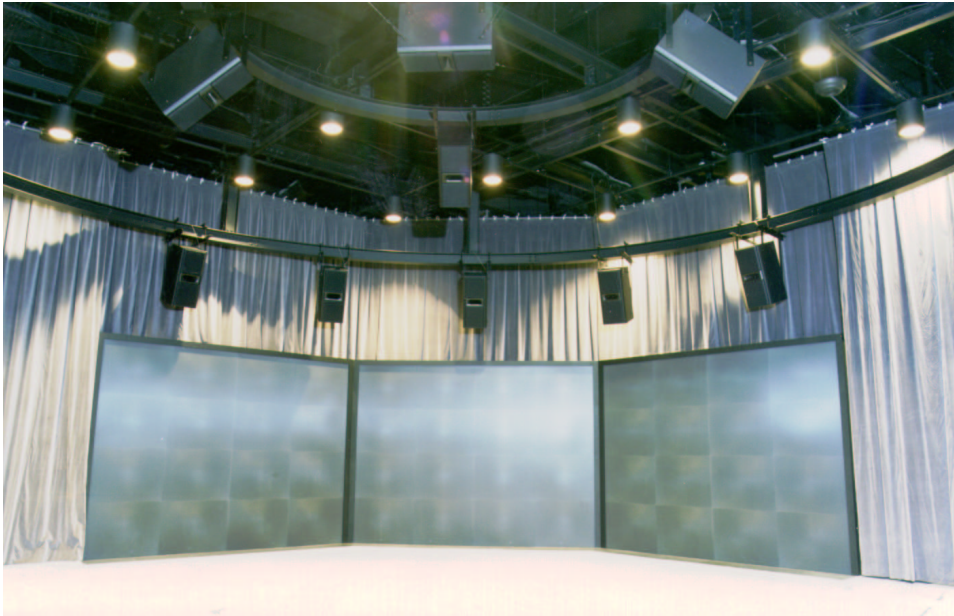


Figure 13: University-Business Innovation Multimedia Center: Virtual reality zone (3D Theater)

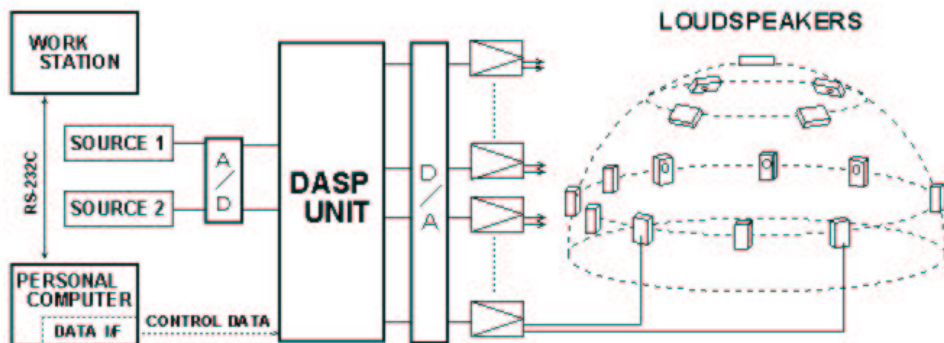


Figure 14: Multimedia Center: Pioneer Sound Field Controller System

9 Future research

The developed sound spatialization resource manager does not adjust for the different resolutions of spatialization backends like the PSFC, for which measurements and tests are necessary. To make flexible the integration of applications and the spatialization resource manager across heterogeneous backends, an abstract sound spatialization interface is under development, to allow seamless switching between different backends. Primary research will be focused on the different sound models employed by various backends.

In our current prototype, the user can interact and play with a recorded MIDI stream only via pointing devices. New input devices are under investigation, along with the possibility of allowing multiple users.

Further work will be directed towards intelligent camera control. The current functions are hard-coded and do not encourage changes or programming by example. A rule base which uses knowledge from movie making could enhance the experience of the user.

10 Conclusion

We have built a virtual reality environment for a Helical Keyboard which gives new insights into music through a novel way of perceiving it. A note has a spatial interpretation in the form of sound (i.e., tone height and pitch chroma), soundscape (i.e., location), visual space (i.e., animated key), and color (i.e., logical order).

VR enables spaces with arbitrary psychophysical characteristics. The soundscape can be decoupled from the visual space, have separate characteristics, and use a different coordinate system via mapping. The creation of such unique experiences motivates development of flexible and generalized spatial sound tools for VR environments as well as advanced spatial sound authoring. What you hear is not necessarily what you see.

Acknowledgments

The Helical Keyboard project was funded in part by the Fukushima Prefectural Foundation for the Advancement of Science and Education. The authors thank Minefumi Hirose, Soushi Hiruta, Kimitaka Ishikawa, Toshifumi Kanno, Kazuhisa Nakashima, and Taku Suzuki for their contributions as part of their Senior and Masters projects at the University of Aizu.

References

- [Amano et al., 1998] Amano, K., Matsushita, F., Yanagawa, H., Cohen, M., Herder, J., Martens, W., Koba, Y., and Tohyama, M. (1998). 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: Trans. of the Virtual Reality Society of Japan*, 3(1). ISSN 1342-4386.
- [Barrilleaux, 2001] Barrilleaux, J. (2001). *3D User Interfaces with Java 3D*. Manning Publications. ISBN 1-88477-790-2.
- [Begault, 1994] Begault, D. R. (1994). *3-D Sound for Virtual Reality and Multimedia*. Academic Press. ISBN 0-12-084735-3.
- [Bell et al., 1996] Bell, G., Carey, R., and Marrin, C. (1996). The Virtual Reality Modeling Language, Version 2.0 Specification, ISO/IEC CD 14772. <http://www.vrml.org/VRML2.0.old/>.
- [Bell et al., 1995] Bell, G., Parisi, A., and Pesce, M. (1995). The Virtual Reality Modeling Language, Version 1.0 Specification. www.vrml.org/Specifications/VRML1.0/.
- [Berlioz, 1830] Berlioz, H. (1830). *Symphonique Fantastique*.
- [Blessner, 2001] Blessner, B. (2001). An interdisciplinary synthesis of reverberation viewpoints. *Journal of the Audio Engineering Society*, 49(10):867–903.
- [Brooks, 1997] Brooks, R. A. (1997). The Intelligent Room Project. In Marsh, J. P., Nehaniv, C. L., and Gorayska, B., editors, *CT'97 - Second Int. Cognitive Technology Conf.*, pages 271–278. IEEE, IEEE press. Aizu-Wakamatsu, Japan, August 25-28, 1997.
- [Brown, 1997] Brown, S. (1997). Real Art and Virtual Reality. *Computer Graphics*, 31(4):36–39. FOCUS: The State of Fine Art.
- [Chodura and Kaup, 1999] Chodura, H. and Kaup, A. (1999). A Virtual Environment for Interactive Music Reproduction. In Mudur, S., Shikhare, D., Encarnacao, J., and Rossignac, J., editors, *Proceedings of IFIP TC/WG5.10 and CSI International Conference on Visual Computing 1999*, pages 95–100. International Federation for Information Processing (IFIP). www.igd.fhg.de/igd-a9/research/audio/index.html.
- [Cohen, 1994] Cohen, M. (1994). Conferences, concerts, and cocktail parties: Besides immersion. In *JMACS: Proc. Japan Music And Computer Science Society Meeting*, pages 17–26, Musashino, Tokyo.
- [Cohen, 1995] Cohen, M. (1995). Besides immersion: Overlaid points of view and frames of reference; using audio windows to analyze audio scenes. In Tachi, S., editor, *Proc. ICAT/VRST: Int. Conf. Artificial Reality and Tele-Existence/Conf. on Virtual Reality Software and Technology*, pages 29–38, Makuhari, Chiba; Japan. ACM-SIGCHI (TBD), SICE (Society of Instrument and Control Engineers), JTTAS (Japan Technology Transfer Association), and NIKKEI (Nihon Keizai Shimbun, Inc.).

- [Cohen, 2001] Cohen, M. (2001). Helical warping and chromastereoptic visualization of a piano-style keyboard. In Tazawa, Y., editor, *Proc. IMS: 4th Int. Mathematica Symp.*, pages 223–230, Tokyo Denki Daigaku, Chiba. ISBN 4-501-73020-X.
- [Cohen and Koizumi, 1995] Cohen, M. and Koizumi, N. (1995). Audio Windows for Virtual Concerts II: Sonic Cubism. In Tachi, S., editor, *Video Proc. ICAT/VRST: Int. Conf. Artificial Reality and Tele-Existence/Conf. on Virtual Reality Software and Technology*, page 254, Makuhari, Chiba; Japan. ACM-SIGCHI (TBD), SICE (Society of Instrument and Control Engineers), JTAS (Japan Technology Transfer Association), and NIKKEI (Nihon Keizai Shimbun, Inc.).
- [Dolinsky, 1997] Dolinsky, M. (1997). Creating Art Through Virtual Environments. *Computer Graphics*, 31(4):34–35. FOCUS: The State of Fine Art.
- [Everest and Streicher, 1998] Everest, F. A. and Streicher, R. (1998). *The New Stereo Soundbook*. Audio Engineering Associates, Pasadena, CA, second edition. ISBN 0-9665162-0-6.
- [He et al., 1996] He, L., Cohen, M. F., and Salesin, D. H. (1996). The virtual cinematographer: A paradigm for automatic real-time camera control and directing. In Rushmeier, H., editor, *Computer Graphics (Proc. SIGGRAPH'96)*, pages 217–224. ACM SIGGRAPH, Addison-Wesley. New Orleans, Louisiana, 4-9 August 1996.
- [Herder, 1998] Herder, J. (1998). Tools and Widgets for Spatial Sound Authoring. *Computer Networks & ISDN Systems*, 30(20-21):1933–1940.
- [Herder, 2000] Herder, J. (2000). Interactive Sound Spatialization - a Primer. *MM News, University of Aizu Multimedia Center*, 8:8–12. (Japanese).
- [Herder and Cohen, 1997] Herder, J. and Cohen, M. (1997). 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. The Acoustical Society of Japan (ASJ).
- [Kanno and Cohen, 2001] Kanno, T. and Cohen, M. (2001). A Helical Keyboard Client. In *Proc. CIT: 2nd Int. Conf. on Computer and Information Technology*, pages 163–165, Shanghai. ISSN 1007-6417.
- [Loy, 1985] Loy, G. (1985). About Audium— A Conversation with Stanley Shaff. *Computer Music Journal*, 9(2):41–48. ISSN 0148-0267.
- [Maebayashi, 1997] Maebayashi, A. (1997). Audible distance. Installation at the NTT InterCommunication Center (ICC), Hatsudai, Tokyo. www.ntticc.or.jp/Collection/Icc/AUDIBLE/index.html.

- [McAllister, 1993] McAllister, D. F., editor (1993). *Stereo Computer Graphics and Other True 3D Technologies*. Princeton University Press. ISBN 0-691-08741-5.
- [Mikami, 1997] Mikami, S. (1997). World, membrane and the dismembered body. Installation at the NTT InterCommunication Center (ICC), Hattori, Tokyo. www.ntticc.or.jp/Collection/Icc/World/index.html.
- [Nakashima and Cohen, 2001] Nakashima, K. and Cohen, M. (2001). Animated extensions to a helical keyboard client: Chord-chords, chord-kites, and intelligent spatialization. In *Proc. HC2001: 4th Int. Conf. on Human and Computer*, pages 39–41, Aizu-Wakamatsu, Japan.
- [New York Philharmonic Orchestra and Mehta, 1980] New York Philharmonic Orchestra and Mehta, Z. (1980). Hector Berlioz: *Symphonie fantastique*, op. 14. CD, The Decca Record Co. LTD. MCPS 400046-2.
- [Pressing, 1997] Pressing, J. (1997). Sound Perspectives on Performed Sound and Music in Virtual Environments. *Presence*, 6(4):482–503.
- [Rumsey, 2001] Rumsey, F. (2001). *Spatial Audio*. Focal Press. ISBN 0-240-51623-0.
- [Shepard, 1983] Shepard, R. N. (1983). Demonstrations of circular components of pitch. *J. Aud. Eng. Soc.*, 31(9):641–649.
- [Shepard, 1984] Shepard, R. N. (1984). Structural representations of musical pitch. In Deutsch, D., editor, *The Psychology of Music*, pages 343–390. Academic Press. ISBN 0-12-213560-1 or 0-12-213562-8.
- [Starks and Linton, 1994] Starks, G. R. and Linton, K. N. (1994). A 3-D Stereo Processing Tool. In *AES: Audio Engineering Society Conv.*, Amsterdam. 3830 (P9.1).
- [Steenblik, 1993] Steenblik, R. A. (1993). Chromastereoscopy. In McAllister, D. F., editor, *Stereo Computer Graphics and Other True 3D Technologies*, pages 183–195. Princeton University Press. ISBN 0-691-08741-5.
- [Stockhausen, 1956] Stockhausen, K. (1955/1956). *Gesang der Jünglinge (Song of the Youths)*. www.stockhausen.org/tape_loops.html.
- [Webbers, 1981] Webbers, A. L. (1981). *Cats*. Musical. Played in Hamburg in a specially build theater from 1986 to 2001.
- [Wernecke, 1994] Wernecke, J. (1994). *The Inventor Mentor*. Addison-Wesley. ISBN 0-201-62495-8.

- [Wolfram, 1999] Wolfram, S. (1999). *The Mathematica Book*. Wolfram Media/Cambridge University Press, fourth edition. ISBN 1-57955-004-5 and 0-521-64314-7.
- [Yunchul, 2001] Yunchul, K. (2001). *Implex*. Installation at Altitude 2001, Academy of Media Arts, Cologne.

Jens Herder



Figure 15: Jens Herder

Jens Herder got appointed in 2000 as professor at the FH Düsseldorf, University of Applied Sciences. Currently, he chairs the Department of Media. At the Virtual Sets and Virtual Environments Laboratory, he experiments with interaction processes for new media applications. After studying computer science at the Technical University of Darmstadt, he worked as a software engineer for Symbolics Systemhaus GmbH. In 1993-2000, he taught computer graphics, algorithms, and programming at the new University of Aizu. He did research on image processing, user interface design, sound spatialization, and virtual reality. In 1999, he got his Ph.D. in Engineering from the University of Tsukuba.

Michael Cohen



Figure 16: Michael Cohen

Michael Cohen (Sc.B. EE, Brown University; M.S. CS University of Washington (Seattle); Ph.D. EECS, Northwestern University) is an Assoc. Prof. in the Human Interface Lab. at the University of Aizu (Aizu-Wakamatsu, Japan), teaching Acoustic Modeling, and Computer Music, and Information Theory. He has interests in telecommunications semiotics, groupware and CSCW (computer-supported collaborative work), audio windows and stereotelephonics, digital typography and electronic publishing, hypermedia, and mixed and virtual reality. He has worked at the Air Force Geophysics Lab (Hanscom Field, Massachusetts), Weizmann Institute (Rehovot, Israel), Teradyne (Boston, Massachusetts), BBN (Cambridge, Massachusetts and Stuttgart, Germany), Bellcore (Morristown and Red Bank, New Jersey), the Audio Media Research Group at the NTT Human Interface Labs (Musashino and Yokosuka, Japan), and the Human Interface Technology Lab. at the University of Washington (Seattle). Cohen is a member of the ACM, IEEE Computer Society, and the Virtual Reality Society of Japan. He is the author or coauthor of four book chapters and many papers, and has two patents.