

Networked Speaker Array Streaming Back to Client: the World's Most Expensive Sound Spatializer?

Michael Cohen, Takuya Azumi, Masahiro Sasaki,
Osamu Takeichi, Sō Yamaoka, & Yoshiki Yatsuyanagi

Spatial Media Group, University of Aizu; Aizu-Wakamatsu 965-8580; Japan

{mcohen, s1080008, s1070098, s1100134,
m5051135, s1070216}@u-aizu.ac.jp

Abstract

We have integrated the spatial sound capability of our University-Business Innovation Center's 3D Theater¹ with our collaborative virtual environment suite,² building a sound spatializer based on the actual room, which can combine artificially directionalized sources naturally with ambient sounds and transmit to remote users, either unicast or broadcast through a reflector to designated users running multimodal groupware, so they can experience telepresence and dynamic control of artificially spatialized sources.

A "reality room" equipped with two sets of sound spatialization speaker arrays is networked to allow remote control. An interface crafted with Java3D presents a 3D model of that space, along with widgets to steer its three mixels of spatial audio, through a perspective display configurable for stereographic or mixed-mode ("security camera") styles. A binaural dummy head suspended from the middle of the ceiling picks up the soundscape, comprising sources spatialized by hemispherical and equatorial speaker arrays as well as ambient sounds (like the voices of people in the room). The stereo pair is captured, sampled, digitized, compressed and encoded by an MPEG-4 broadcaster, and streamed by a network streaming server back to the original remote client, which feedback closes the control/display loop.

Key words: CVE (collaborative virtual environments), groupware, CSCW (computer-supported collaborative work), mixed audio reality, telepresence, immobot, roomware.

0. Introduction

0.1 CVE: Multimodal Groupware Architecture

We have designed and implemented an architecture and framework [KCNH01] to support a collaborative virtual environment (CVE), allowing distributed users to share multimodal virtual spaces [CM01]. Our CVE architecture is based upon a client/server model, and its main transaction shares the state of virtual objects and users (avatars) by replicated unicast of position parameters (translation and rotation) to client peers in a session. Our system supports heterogeneous interfaces, including graphical (WIMP/GUI), auditory (displayed through headphones [including stereotelephonic], nearphones, stereo or quadraphonic speakers,

and speaker arrays), and eventually haptic (force-feedback display) modalities.

0.2 Mixed or Augmented Audio Reality/Virtuality

Telepresence deploys surrogates to sense on behalf of a human master: a pilot, controlling from afar, wears effectors corresponding to the puppet's sensors, like *bunraku* with feedback. Spatial sound systems [Beg94] [ES98] [Rum01] control virtual audio sources through a combination of intensity panning/fading, time delays, and frequency-dependent DSP filtering ("earprints," or head-related transfer functions, commonly implemented as dynamically chosen filters in a convolution engine).

"Augmented" or "mixed reality" (also known as, or related to, annotated, enhanced, hybrid, mediated, and virtualized reality systems) [CM92] [MK94] [BC01] is used to describe hybrid presentations that overlay computer-generated information on top of "real" scenes. Our group is exploring broadband mixed reality/virtuality systems which blur sampled and synthesized data, mixing actual and artificial information, especially realtime media streams.

Augmented audio reality [CAK93] [YCHY01] applies this notion to sonic effects, superimposing computer-generated sounds on top of more directly acquired audio signals. (One common example of augmented audio reality is sound reinforcement, as in a public address system.)

By juxtaposing and mixing 'real' and 'synthetic' audio transmissions, we are exploring the relationship between acoustic telepresence and VR presentations: telepresence manifests as the actual configuration of sources in a sound field, as perceivable by a dummy-head; VR is the perception yielded by HRTF filtering of virtual sources with respect to virtual sinks.

0.3 Spatially Immersive Displays: media spaces

The most direct way of implementing spatial audio and increasing spatial resolution is by simply distributing real sources in space, as in antiphonal music. The Audium³ is a specially-constructed theater featuring sound dancing around its 169 speakers in an intimate (50 seats) space [Loy85] in San Francisco. "The House on the Rock"⁴ features a collection of room-sized musical robots. Such theaters and automata are charming—taking "theatre-in-the-round" and turning it inside-out, and enhancing the ability to "hear out" instruments, as suggested by the decomposition in Table 2— but impractical for anything besides special-purpose venues and LBE [Coh02]. We work-around such

¹www.mmc-aizu.pref.fukushima.jp/pub/system.file/system.html

²www.u-aizu.ac.jp/~mcohen/spatial-media/groupware/clients.html

³www.audium.org

⁴www.houseontherock.com/the_attraction.htm



Fig. 1: Synthetic World Zone [3D Theater] at the University of Aizu University-Business Innovation Center, featuring PSFC and RSS-10 Systems. (Our disembodied dummy head is shrouded in a burlap bag [to avoid startling visitors], visible at the top of the picture.)

role	literature	television	theatre	music	MPEG-4
source	author	teleplay writer	writer	composer	capture function
presentation	editor	producer, director, actor		arranger	broadcaster
distribution: delivery	publisher	broadcaster		performer	streaming server
consumption: sink (terminal)	reader	viewer	audience	audience	client

Table 1: Producer→Distributor→Consumer: Interactive multimedia presentations blur the distinction between composer/arranger, performer, and listener, as hypermedia blurs the distinction between author, publisher, and reader.

impracticality, demonstrating telepresence to such a venue, namely the “3D Theater” at our own “University-Business Innovation Center (UBIC).”

1. Implementation

1.1 Spatial Audio via Speaker Array Driven by PSFC or Coupled Roland RSS-10s

A “reality room” equipped with two sets of sound spatialization speaker arrays is networked to allow remote control. Our system steers a total of three mixels,⁵ two through through an older “PSFC,” and one through a new RSS-10 system, both described following.

We previously developed a networked driver for the PSFC (Pioneer Sound Field Controller⁶) [AMY⁺98] [HMC01] in the University of Aizu University-Business Innovation Center’s 3D Theater,⁷ shown in Fig. 1, controlling spatialization of audio sources through a DSP-driven hemispherical speaker array.

We have also newly developed a controller to drive a collocated eight-channel equatorial speaker array sound spatialization system based on four ganged Roland RSS-10 Sound Space Processors [Ro197], as shown in Fig. 2, integrated via Roland’s C-language “Staff” API wrapped with JNI (Java Native Interface). The driver is deployed on PCs running W2K Pro or Server, using JRE 1.3.x.

Control signals from a computer to the RSS-10s are distributed as initial serial MIDI interface, and by a regular MIDI daisy chain between the modules. (We had previously developed a Helical Keyboard [NC01b] [KC01] [HC02] to visualize the helical structure of a musical scale, animated in realtime by a GUI or MIDI controllers or events (sequencers). If the Helical Keyboard client is running, it uses the USB MIDI port.)

Even though the driver software is designed primarily for display, reading events off the network for directionalization or spatialization via the speaker array, a graphical debugger is integrated for stand-alone operation. The visual debugging display is a dynamic map, shown in Fig. 3, displaying a plan view (overhead, or bird’s or かみさんの知覚 Gods’ eye), an orthographic 2.5D projection of the space.

The RSS-10 modules are configured for “fly mode,” allowing dynamic rearrangement without artifacts. The driver has three modes:

Panning Disabled Since the RSS-10 is based upon a 2 channel audio system, interference across the speaker array degrades horizontal localizability.

Gain adjustment Yielding a richer spatialization, if more ambiguous directionalization, the RSS-10’s native spatialization algorithm can be used to drive all the speakers by embedded transfer functions. The algorithm used is like

⁵We call source→sink spatialization channels mixels, acronymic for sound **mixing elements** — in analogy to dexels (**depth elements**), hogels (**holographic elements**), pixels (**picture elements**), taxels (**tactile array elements**), texels (**texture elements**), and voxels (**volumetric elements**, a.k.a. boxels), since they are like a raster across which a soundscape is projected — define the granularity of control and degree of directional or spatial polyphony.

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

⁷www.mmc-aizu.pref.fukushima.jp/mmc/system/sys7.html

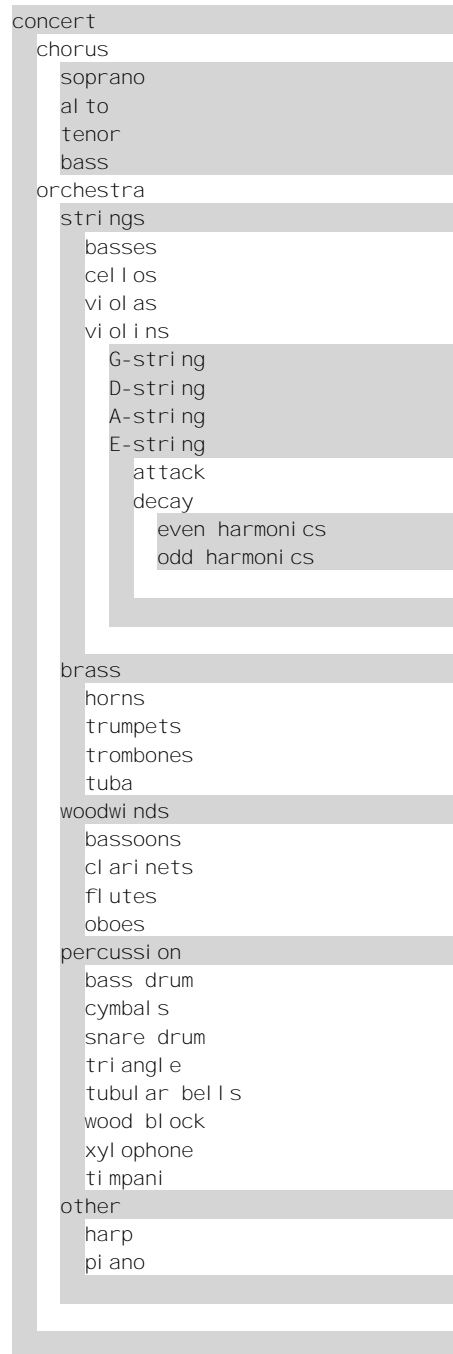


Table 2: Concert decomposition: Individually presenting instruments allow them to be “heard-out” by a listener. With an appropriate interface, one could separate, for instance, an orchestral cluster into separate instruments, differentiating through concert → orchestra → section → instrument and actually analyze an instrument’s sound. Such super-decomposition might allow, for example, a user to listen to spatially separate strings of a violin.

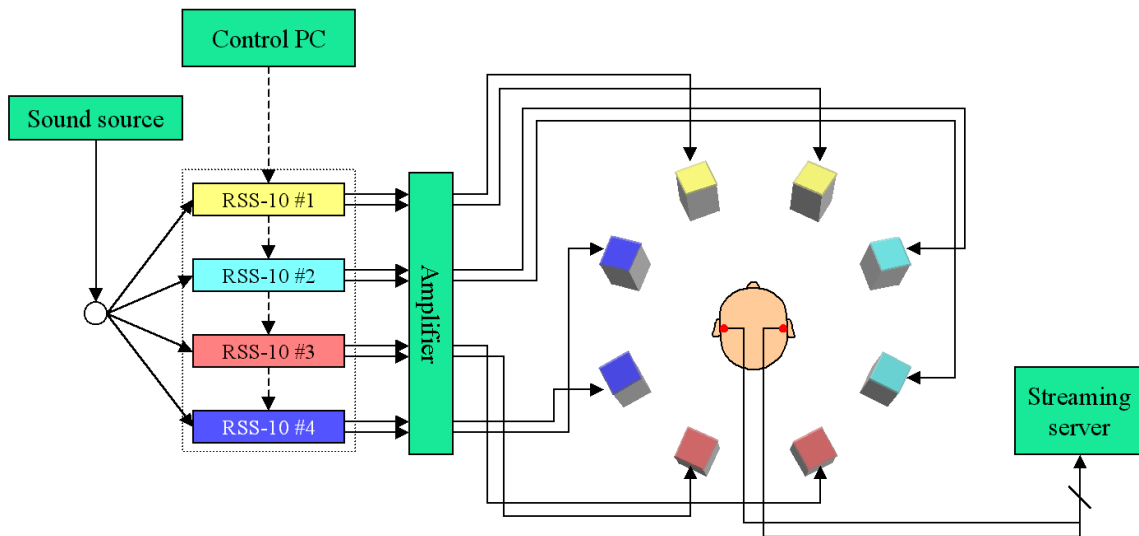


Fig. 2: System schematic

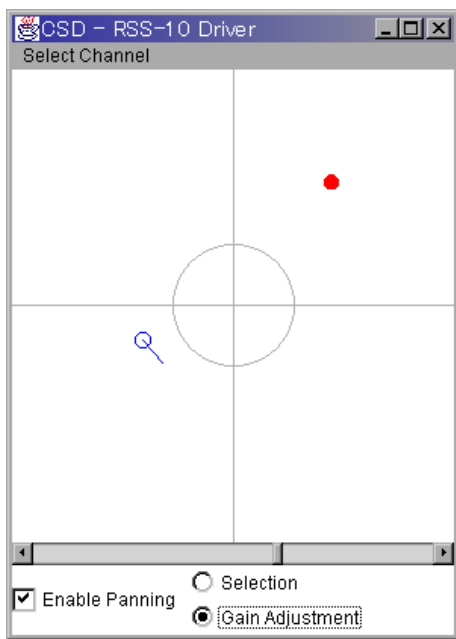


Fig. 3: 2.5D Plan Debugger: orthographic projection of soundscape: displaying and controlling iconic planar translation and rotation

vector base amplitude panning [Pul97], but currently does not support z axis positioning (horizontal plane only).

Selection For simplest debugging, but poor spatialization, selection mode causes the RSS-10s to act as a switching matrix deployed as a demultiplexer, choosing one of the subsystem’s four pairs of speakers.

Doppler shifting is supported by the RSS-10s, but disabled in our interface. Room effects (size and absorption characteristics) are programmable, but we currently don’t change the default configuration.

1.2 Java3D Widgets

Java3D⁸ [SRD00] [Pal01] [Bar01] [WG02] (hereafter “J3D”) is used to craft a perspective user interface for our system, configurable for stereographic or mixed-mode (“security camera”) styles, as shown in Fig. 4. Models of the room, speaker arrays, and dummy head were created in Discreet 3DStudio Max,⁹ exported to Newtek LightWave¹⁰ via .dxf, and exported from Lightwave as an .lwo file, loaded at runtime, and composed with scenegraph-programmed movable objects. These widgets, colored spheres representing the location of three separately-locatable sources, circulate around the dummy head in the middle of the space, which we dub “Threater,” it being a 3D 3-mixel theater. Our interface allows display and control of the three mixels, driving spatial audio through stereoscopic exocentric and/or egocentric perspective displays and soundscapes. Picking and dragging gestures rearrange the soundscape via direct WYSIWYG manipulation of the sources (simple planar translation, since sources are omnidirectional).

1.3 Sources

Currently our sound sources are pre-prepared, played either from an anechoic CD (like “Impact” [Jap86], “Music for Archimedes” [Ban92], or “Sound material for RSS #1”) or a hard disk recorder. (We use Fostex D824mkII 8-track digital recorders.) With three mixels, we could separately steer, for example, a *koto*, *shakuhachi*, and *shamisen* trio. *Karaoke* records, with separate singing voice and orchestra channels, also make satisfying, if non-anechoic, sources.

Source signals can be distributed by recording the same track 4× in parallel on the removable hard disk recorder or by simple fan-out cable constructed by one of the authors. Despite the impedance mismatch between source (the recorder specifi-

⁸java.sun.com/products/java-media/3D/

⁹www.discreet.com/products/3dsmax

¹⁰www.newtek.com

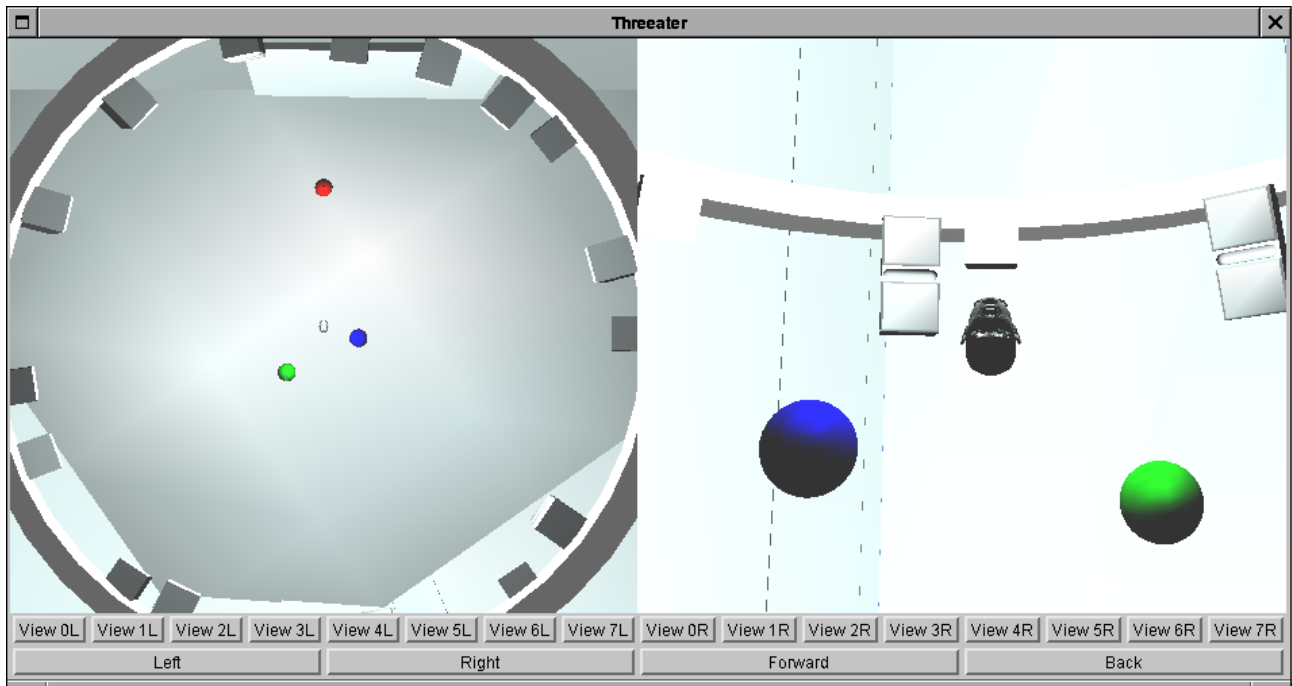


Fig. 4: Threater Perspective Interface

cations recommend a 10 k Ω load) and quadruple parallel load (each RSS-10 unit has 10 k Ω input impedance, for ganged combination of 2.5 k Ω), no problems (distortion, etc.) have been noted.

1.4 Capture



Fig. 5: "Mimi": disembodied, impaled, inverted, and suspended

A binaural dummy head, suspended from the middle of the ceiling as pictured in Fig. 5, picks up the soundscape, comprising sources spatialized by the hemispherical and equatorial speaker arrays as well as ambient sounds (the voices of people in the room). Embedded in the mannequin's ear canals are a pair of

(battery-powered Sony EMC-T140 electret condenser) microphones, driving cables that run across the ceiling down to the broadcasting computer. Our androgynous dummy head (shown again in Fig. 7), taking the French girls' name "Mimi";¹¹ is mounted upside-down for aesthetic reasons (the top of her head being more pleasant to look up at than her severed neck), on about the same plane as the speakers, but as pinna elevation effects are negligible (Mimi being an inexpensive mannequin, rather than an acoustically rigorous dummy head like [BK]), such inversion does not further impair localization.

1.5 MPEG-4

MPEG-4¹², included in QuickTime 6, is an important standard for object-oriented multimedia, including audio streaming and spatialization [Mik98] [OK01]. It is designed for broadcast (streaming) as well as interactive and conversational applications. We use MPEG-4 to transmit to remote users, either unicast to a designated user running multimodal groupware or broadcast through a reflector, who can experience telepresence and dynamic control of artificially spatialized sources.

The stereo pair from the dummy head is captured, sampled, digitized, compressed by an MPEG-4 broadcaster, and encoded and streamed by a network streaming server back to a remote client running the graphical controller, which feedback closes the control/display loop.

For a client we use the Apple Macintosh QuickTime Client

¹¹Because of the coincidence that Alexander Graham Bell was born on March 3 (1847), Anne Sullivan started teaching Helen Keller on March 3 (1887), the resemblance of the figure '3' to a (left) pinna, and some felicitous word-play ("3" may be pronounced /mi/ in Japanese [as in "Do-Re-Mi"], so "三日月" can be read /mimi/, homonymic with the Japanese word for 'ear'), March 3rd is designated "Ear Day" in Japan, promoting aural hygiene and hearing safety.

¹²www.csel.t.it/mpeg

(v. 6), but probably the Windows Media Architecture Media Player¹³ or the Real Networks Real One Player¹⁴ could also be integrated.

2. Experience

More satisfying as a concept than as an actual tool, the streamed sound is from our spatializer is somewhat muddy (due to room effects and compression), as well as delayed (by about 2 s in informal tests). Other engines (J3D, DirectX, Sound-Blaster, a locally deployed stand-alone RSS-10, our own S^6 [CS00] or its quadrasonic extension [FCB⁺02], post-production effects like SoundSaVR,¹⁵ ...) are less expensive and sound better.

3. Future Research

3.1 Video Pickup and Streaming

MPEG-4 can stream video as well as audio, and we are experimenting with using a “mixel-cam” to provide a live feed, especially of the carbon-based audio sources! Installation of a Firewire-equipped computer will make such capture even faster.

We are also investigating the use of USB microphones, more convenient than active battery-powered mics we use now (since the USB bus supplies power) and less noisy, since long cables are not a liability for digital signals.

3.2 Mobile Telephone Interface

We have designed and implemented a mobile telephone interface [NC01a], fully integrated into our multimodal CVE, so that the mixels can also be controlled by a MIDlet running on a DoCoJa mobile phone (“*keitai*”: 携帯 電話). Programmed with “DoCoJa” J2ME (Java 2 Platform, Micro Edition¹⁶) a dynamic map application runs on an (NTT DoCoMo) *iappli* mobile phone, as shown in Fig. 6. Featuring selectable icons with one rotational and two translational degrees of freedom, like the 2.5D map shown in Fig. 3, the interface is used to control avatars in a chatspace. The Sony *iappli* models feature a thumb jog wheel shuttle, which can be used as a continuous controller to manipulate such icons in a teleconference. The user interface is further extended with musical and vibrational cues. We hope to eventually deploy full teleconferencing with spatial audio via such a mobile phone¹⁷ for full CTI (computer-telephone integration), but unfortunately voice communication is currently disabled during such *iappli* sessions, so a second phone must be used for teleconferencing.

Emerging mobile telephones will feature high-fidelity stereo audio capability, as suggested by Fig. 7, so we hope to experiment with MPEG-4 streaming directly to a mobile phone. JPhone supports MPEG-4 streaming, and FoMA (NTT’s 3rd generation “Freedom of Mobile Access” service) supports MPEG-4 streaming through DoCoMo’s “iMotion” service from WME (Windows Media Encoder), as well as multiaccess features (simultaneous voice and data connections).



Fig. 7: “Poor person’s mobile stereotelephony:” two mobile phones deployed as a microphone array attached to a dummy head simultaneously calling a dual voice line (like that provided by ISDN) realizes wireless (if still low-fidelity) binaural telepresence.

4. Conclusion

We are now evaluating the usefulness and potential of our per-versely complicated system. Test applications could include multimedia chat spaces (social) and conferencing (business), as well as VR-style gaming and music. An imrobot is an immobile robot: instead of moving around to assist its masters, an imrobot is part of the architecture, concentrating on attending and servicing the needs of collocated human users, rather than the traditional robotic tasks of navigation and exploration of an environment. In that sense, our spatial audio speaker arrays system can be considered an imrobot: instead of serving tea, it serves spatial audio! Since the rearrangement of a dynamic map is used to dynamically select transfer functions, a user may specify the virtual location of sound sources, throwing the audio into perceptual space, using exocentric graphical control to drive an egocentric auditory display. Because the UBIC cost about ¥100,000,000 (≈ \$1,000,000), we speculate that our round-about spatializer might be “the world’s most expensive.”

Acknowledgments

Thanks to Kazuo Kunieda; 國枝 和雄 at NEC Human Media Research Laboratories for encouraging our RSS-10 development, to Keiji Takanami; 高波 慶二 in the University-Business Innovation Center for technical support, and to “*あいつ・ジャパン*”¹⁸ for strategic consulting.

¹³www.microsoft.com/windows/windowsmedia/

¹⁴www.realone.com

¹⁵www.smgvr.com

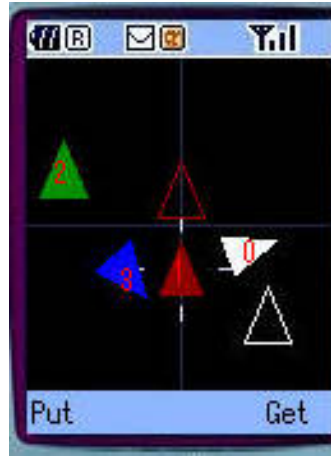
¹⁶java.sun.com/j2me/

¹⁷java.sun.com/products/jtapi/

¹⁸www.aizu.com



(a) iJade emulator showing “i-Con” iappli MIDlet.)



(b) Close-up of the screen, showing ghost icons used to distinguish local and remote session state.

Fig. 6: In the naming style of “applets” and “servlets,” a MIDP application (for *mobile information device profile*) is called a MIDlet [Knu01]. (Programming by Yutaka Nagashima; 長島 豊 and extended by Makoto Kawaguchi; 川口 誠.)

References

- [AMY⁺98] 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: Trans. Virtual Reality Society of Japan*, 3(1):1–12, March 1998. ISSN 1344-011X; www.u-aizu.ac.jp/~mcohen/welcome/publications/PSFC.ps.
- [Ban92] Bang & Olufsen. *Music for Archimedes*, 1992. CD B&O 101.
- [Bar01] Jon Barrilleaux. *3D User Interfaces with Java 3D*. Manning Publications, 2001. ISBN 1-88477-790-2.
- [BC01] Woodrow Barfield and Thomas Caudell, editors. *Fundamentals of Wearable Computers and Augmented Reality*. Lawrence Erlbaum Assoc., 2001. ISBN 0-8058-2902-4.
- [Beg94] Durand R. Begault. *3-D Sound for Virtual Reality and Multimedia*. Academic Press, 1994. ISBN 0-12-084735-3.
- [BK] Brüel and Kjør. Head and torso simulator, type 4128.
- [CAK93] Michael Cohen, Shigeaki Aoki, and Nobuo Koizumi. Augmented audio reality: Telepresence/VR hybrid acoustic environments. In *Proc. Ro-Man: 2nd IEEE Int. Workshop on Robot and Human Communication*, pages 361–364, Tokyo, November 1993. ISBN 0-7803-1407-7.
- [CM92] T. P. Caudell and David W. Mizell. Augmented reality: An application of heads-up display technology to manual manufacturing processes. In *Proc. Hawaii Int. Conf. on Systems Sciences*. IEEE, January 1992.
- [CM01] Michael Cohen and William L. Martens. Spatial Media Research at the University of Aizu. *JVRSJ: J. Virtual Reality Society of Japan*, 6(2):52–57, September 2001. ISSN 1342-6680; www.u-aizu.ac.jp/~mcohen/welcome/publications/spatial-media.ps.
- [Coh02] Michael Cohen. A survey of emerging exotic auditory interfaces. In Ryohei Nakatsu and Hideki Kawahara, editors, *Proc. ICAD: Int. Conf. on Auditory Display*, pages 1–16, Kyoto, July 2002. www.u-aizu.ac.jp/~mcohen/publications/exotic.ps.
- [CS00] Michael Cohen and Kenta Sasa. An interface for a soundscape-stabilized spiral-spring swivel-seat. In *Proc. WESTPRAC VII: 7th Western Pacific*

- Regional Acoustics Conf.*, pages 321–324, Kumamoto, Japan, October 2000. ISBN 4-9980886-1-0 and 4-9980886-3-7.
- [ES98] F. Alton Everest and Ron Streicher. *The New Stereo Soundbook*. Audio Engineering Associates, Pasadena, CA, second edition, 1998. ISBN 0-9665162-0-6.
- [FCB⁺02] Owen Newton Fernando, Michael Cohen, Noor Alamshah Bolhassan, Dishna Wanasinghe, and Toshifumi Kanno. Mobile Control in Cyberspace of Image-based & Computer Graphic Scenes and Spatial Audio Using Stereo QTVR and Java3D. In *Proc. HC2002: 5th Int. Conf. on Human and Computer*, Aizu-Wakamatsu, Japan, September 2002.
- [HC02] Jens Herder and Michael Cohen. The helical keyboard: Perspectives for spatial auditory displays and visual music. *J. of New Music Research*, December 2002.
- [HMC01] Kuniaki Honno, William L. Martens, and Michael Cohen. Psychophysically-derived control of source range for the Pioneer Sound Field Controller. In *Proc. AES: Audio Engineering Society Conv. (110th Convention)*, Amsterdam, May 2001. Preprint #5313; www.u-aizu.ac.jp/~wlm/papers/aes110_98.ps.
- [Jap86] Japan Audio Society. Impact II CD. 1-14-34 Zingumae, Shibuya-ku; Tokyo 150, October 1986.
- [KC01] Toshifumi Kanno and Michael Cohen. A Helical Keyboard Client. In *Proc. CIT: 2nd Int. Conf. on Computer and Information Technology*, pages 163–165, Shanghai, September 2001. ISSN 1007-6417.
- [KCNH01] Toshifumi Kanno, Michael Cohen, Yutaka Nagashima, and Tomohisa Hoshino. Mobile control of multimodal groupware in a distributed virtual environment. In *Proc. ICAT: Int. Conf. on Artificial Reality and Tele-Existence*, pages 147–154, Tokyo, December 2001. ISSN 1345-1278; sklab-www.pi.titech.ac.jp/~hase/ICATPHP/upload/39_camera.pdf.
- [Knu01] Jonathan Knudsen. *Wireless Java: Developing with Java 2, Micro Edition*. Apress, 2001. ISBN 1-893115-50-X.
- [Loy85] Gareth Loy. About Audium— A Conversation with Stanley Shaff. *Computer Music Journal*, 9(2):41–48, 1985. ISSN 0148-0267.
- [Mik98] Norizaku Miki; 三木 弼一. *MPEG 4 のすべて*. 1998. ISBN 4-7693-1167-2.
- [MK94] Paul Milgram and Fumio Kishino. A taxonomy of mixed reality visual displays. *IEICE Trans. Inf. Sys.*, E77-D(12):1321–1329, December 1994.
- [NC01a] Yutaka Nagashima and Michael Cohen. Distributed virtual environment interface for a mobile phone. In *Proc. HC2001: 4th Int. Conf. on Human and Computer*, pages 43–46, Aizu-Wakamatsu, Japan, September 2001.
- [NC01b] Kazuhisa Nakashima and Michael Cohen. 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, September 2001.
- [OK01] Sakae Ōkubo and Masahisa Kawashima. *H.323/MPEG-4 教科書*. MCR (Multimedia Communication Research), 2001. 大久保 榮 and 川島 正久; ISBN 4-901280-03-1.
- [Pal01] Ian Palmer. *Essential Java 3D fast: developing 3D graphics applications*. Springer, 2001. ISBN 1-85233-394-4.
- [Pul97] Ville Pulkki. Virtual source positioning using vector base amplitude panning. *J. Aud. Eng. Soc.*, 45(6):456–466, June 1997.
- [Rol97] Roland Corp., 〒 530 絨紹冚緋俱 1-4-16. *How To Make The True 3D Sound by RSS-10*, January 1997.
- [Rum01] Francis Rumsey. *Spatial Audio*. Focal Press, 2001. ISBN 0-240-51623-0.
- [SRD00] Henry Sowizral, Kevin Rushforth, and Michael Deering. *The Java 3D API Specification*. Addison-Wesley, second edition, 2000. ISBN 0-201-71041-2.
- [TSA95] Mikio Tohyama, Hideo Suzuki, and Yoichi Ando. *The Nature and Technology of Acoustic Space*. Academic Press, 1995. ISBN 0-12-692590-9.
- [WG02] Aaron E. Walsh and Doug Gehringer. *Java 3D API jump-start*. Prentice-Hall, 2002. ISBN 0-13-034076-6.
- [YCHY01] Yasuhiro Yamazaki, Michael Cohen, Jie Huang, and Tomohide Yanagi. Augmented audio reality: compositing mobile telerobotic and virtual spatial audio. In *Proc. ICAT: Int. Conf. on Artificial Reality and Tele-Existence*, pages 213–216, Tokyo, December 2001. ISSN 1345-1278; sklab-www.pi.titech.ac.jp/~hase/ICATPHP/upload/41_camera.pdf.