A SURVEY OF EMERGING AND EXOTIC AUDITORY INTERFACES

Michael Cohen

Spatial Media Group, Human Interface Lab.
University of Aizu 965-8580; Japan
mcohen@u-aizu.ac.jp

ABSTRACT

Anticipating some emerging audio devices and features, this paper surveys trends in mobile telephony (especially regarding mobile internet in Japan), wearable/intimate multimedia computing, handheld/nomadic/portable interfaces, and embedded systems like multimedia furniture and spatially immersive displays, gleaned from recent press releases and advertisements, popular media, and publications by industrial and academic laboratories, especially the author's own research group. Representative instances are cited, and conferencing narrowcasting selection functions are reviewed.

Keywords: audio interaction, CVEs (**c**ollaborative **v**irtual **e**nvironments), embedded systems, handheld/mobile/portable interfaces, integration of mobile devices and telecommunication, mobile information device, mobile internet, multimodal interaction, novel user interfaces, pervasive Java, telematics, telerobotics, ubicomp (**ubi**quitous **comp**uting) (a.k.a. ambient, calm, pervasive) technology, wearable/intimate multimedia computing.

0. INTRODUCTION

The oft-mentioned digital "3C convergence" generally refers to the confluence of **c**ommunication devices, **c**omputing, and **c**onsumer electronics (or, alternatively, **c**ontent). As summarized by Table 1, such integration enables ubicomp (**ubi**quitous **comp**uting),[1] [2] the smooth interaction of devices at different scales. The author's research group is investigating interfaces related to robotics, **s**patially **i**mmersive **d**isplays, information furniture, and mobile computing.

1. PORTABLE AUDITORY INTERFACES

1.1. Wireless Computer-Enhanced Telephony

The dream motivating wireless technology is "anytime, anywhere" communications. In Japan, where cell phones out-sell wired, mobile phones are abbreviated "*keitai*" (meaning 'mobile': 携帯 電話). Voice-only *keitai*s are considered the first generation of mobile phones, "1G." As revenue for voice communication steadily decreases, operators are compelled to start new services to develop new markets and increase revenue from data communication. Wireless computing offers unique challenges because of the interesting form factors (weight, size, interface), noise (less robust network), limited bandwidth, and social potential.

[1] For reader convenience, this paper includes relevant URLs for supplemental WWW browsing, also collected at www.u-aizu.ac.jp/~mcohen/welcome/publications/exotic.html

[2] www.viktoria.se/ubicomp/

smart spaces and entertaining (aware) environments [1]
cooperative buildings [2]
roomware (software for rooms) [3] and reactive rooms
media spaces
immobots (**immo**bile r**obots**)
spatially **i**mmersive **d**isplays
information furniture
networked appliances [4]
handheld/mobile/nomadic/portable/wireless
wearable/intimate computing [5] [6] [7]
computational clothing (smart clothes)

Table 1: *Saturated: distributed & pervasive, continuous & networked, transparent or invisible— Spatial hierarchy of ubicomp or ambient intimacy.*

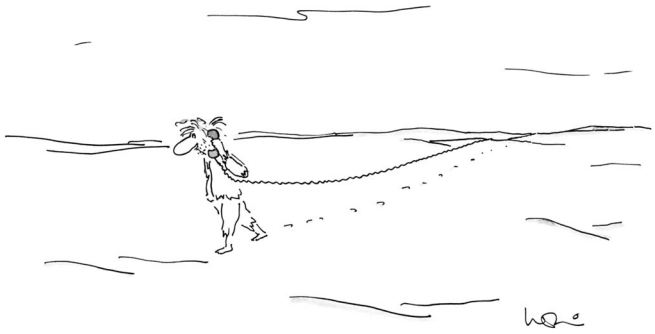


Figure 1: *Non-wireless telephony.* *(©2002 The New Yorker Collection from* cartoonbank.com*. All rights reserved.)*

1.1.1. Mobile Internet: "2G"

Japan has more than 2000 mobile services, and the three operators that offer internet-enabled phones have (as of Spring, 2002) more than 70 million subscribers, anticipating a saturated diffusion of 60% or 80 million (contrastable to the 5 million users of personal computers). Mobile internet services are concerned with personalization and offering wireless internet-delivered convenience,[3] on demand in both time and space [8].

Japan's younger citizens, especially, have become members of the *Oyayubi-sedai* (thumb generation) or *Oyayubi-zoku* (thumb tribe), intensely using their mobile phones for e·mail, "buddy" tex-

[3] apro.techno.net.au/apb203.htm

Parent company	**Subsidiary**	**Internet (2G)**	**Java capability (2.5G)**	**Broadband (3G)**
NTT	DoCoMo	i-mode	iαppli	FoMA (W-CDMA: 64 kbps upstream, 384 kbps downstream)
KDDI	au (H˜)	EZ-web	EZplus	(cdma2000-1×: 144 kbps downstream)
Japan Telecom	J-Phone	J-Sky	Java Appli	(launched in June, 2002)

Table 2: *Japanese mobile telephone services.*

ting features like SMSs (**s**hort **m**essage **s**ervices) and IM (**i**nstant **m**essaging), web surfing, and even conversing. DoCoMo "i-mode" (not to be confused with a Unix inode) is NTT's mobile internet service,[4] rivaled by "J-Sky" by J-Phone (the mobile communications arm of Japan Telecom [which is owned by British Vodaphone]) and "EZ-web" by au (a subsidiary of KDDI [which was formed by the merger of KDD, DDI, and IDO]), as summarized by Table 2. Such services are used to check ski conditions, movie times, and restaurant addresses; download screen backgrounds (cartoon characters or pictures of idols, athletes, musicians, etc.); as well as for online shopping and banking, navigation, weather reports, and so on. From this Autumn, Japanese gamblers will be able to wager on horse races using i-mode.

AT&T in the U.S. recently launched its own version of i-mode, called "mMode."[5] Even in Europe, i-mode is challenging WAP (**w**ireless **a**pplication **p**rotocol, a.k.a. "**w**rong **a**pproach to **p**ortability"), because of WAP's unfamiliar language (WML), slow operating speeds, difficult user interface, and ability to duplicate only parts of existing web sites [9, p. 456–7]. DoCoMo has tied up with Sony Computer Entertainment to develop games that can be played both on i-mode phones and on Playstation consoles, and Natsume makes a (Keitai Denjyu) "Telefang" interface allowing Nintendo[6] GameBoys to network for online gaming.

Any telephone keypad is awkward for text input, cumbersome as it is to thumb-type up to eight key-presses for mixed-case characters. Mobile browsers will be naturally extended by voice interfaces, using VXML (for VoiceXML[7]) to create audio dialogs that feature mixed-initiative conversations: digitized audio, recognition of touchtone key and spoken input (via ASR: **a**utomated **s**peech **r**ecognition), recording of spoken input, and synthesized speech (via TTS: **t**ext-**t**o-**s**peech) [10]. Speaking of (and with) which, AT&T has developed a set of algorithms, called "Natural Voices,"[8] which represent a vast improvement over the familiar "drunken Scandanavian robot" intonations.

1.1.2. Mobile Telephonic Computing: "2.5G"

In the U.S., the FCC-mandated "E-911" (for "Enhanced 911") initiative requires carriers to ensure that emergency services can pinpoint the location of distressed callers using mobile phones. Steve Wozniak has launched "Wheels of Zeus,"[9] to design mobile products for the consumer electronics market exploiting advances in GPS (**g**lobal **p**ositioning **s**ystems), advanced processors, and two-way wireless communication systems.

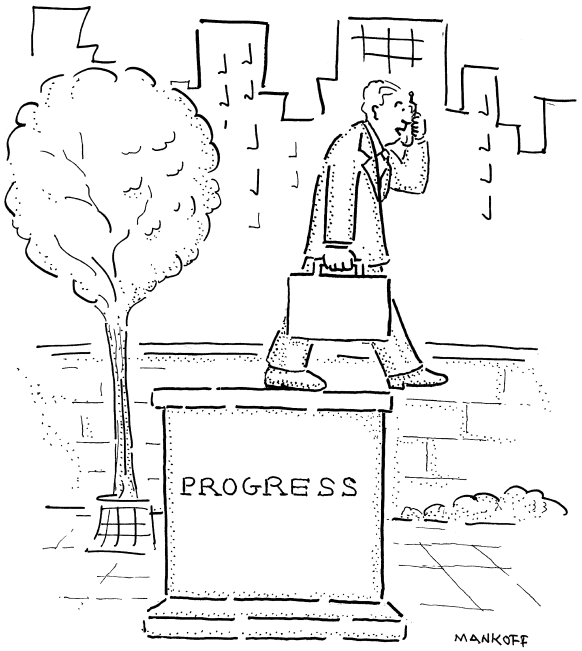


Figure 2: *Preoccupied by his conversation, this everyman is in danger of striding off the pedestal of progress. (©2002 The New Yorker Collection from* cartoonbank.com*. All rights reserved.)*

J2ME[10] is Sun's Java 2 microedition engine,[11] offering a cross-platform programming language for environments including mobile phones, PDAs (**p**ersonal **d**igital **a**ssistants), and on-board car computers. Such capability enables location-based information systems: "LBE" (**l**ocation-**b**ased **e**ntertainment) and "LBS" (**l**ocation-**b**ased **s**ervices).[12] NTT DoCoMo offers a Java-based service called "iαppli" [11] [12], and J-Phone and Au serve "Java Appli" and "EZplus," respectively. Microsoft makes a wide-screen "Stinger" combination phone and web browser, while Pocket PCs, like those made by Compaq and Hewlett-Packard, connect to the internet via modem cards, and the RIM BlackBerry[13] features an embedded radio-modem.

Our own research group has designed and implemented a mobile telephone interface [13] for use in CVEs (**c**ollaborative **v**irtual **e**nvironments) [14] [15] [16]. Anticipating ubicomp networked appliances[14] and information spaces, we are integrating various multimodal (auditory, visual, haptic) I/O devices into a virtual re-

[4] www.nttdocomo.com
[5] www.attws.com/mmode/
[6] www.nintendo.com
[7] www.w3.org/Voice/
[8] www.naturalvoices.att.com
[9] www.wOz.com
[10] java.sun.com/j2me/
[11] java.sun.com/products/cldc/
[12] www.atip.org/public/atip.reports.02/atip02.022.pdf
[13] www.blackberry.net
[14] computer.org/pervasive/, www.pervasive2002.org

Figure 3: *NTT DoCoMo i-mode iαppli iJade emulator running "ι·Con" application. The quasi-realtime synchronization with* CVE *server motivates the use of "ghost icons" to distinguish local and session states of avatars. (Developed by Yutaka Nagashima;* 長島 豊.*)*

ality groupware suite. Programmed with J2ME, one of our client applications runs on a mobile phone, as illustrated by Figure 3. The interface is used to control avatars in a chatspace via a 2.5D dynamic map (one rotational and two translational degrees of freedom for each icon). The Sony 503-series iαppli models feature a thumb jog shuttle, which can be used as a continuous controller to manipulate such icons. The "ι·Con" user interface is further extended with musical and vibrational cues, to signal mode changes and successful transmission/reception (which feedback is important in wireless communication, as it is much less deterministic than terrestrial lines). Other developers are reportedly working a web-based fishing game in which the phone vibrates when a fish takes the bait and is ready to be reeled in with the jog shuttle.

A direct, if naïve, CVE integration would be inappropriate for i-mode clients, as the i-mode protocol supports only client pull (and no server push). Our workaround is to deploy a thin mobile client connected to a heavier network proxy client, initiating data exchanges from the mobile handset regardless of whether the request is to send or receive updates. Data communication between the ι·Con iαppli mobile phone interface and server is made via a servent (**serv**er/cli**ent** hybrid) as HTTP↔TCP intermediary. (Our middleware servelet was developed by strategic partner "Eyes, Japan"[15] in conjunction with "GClue."[16]) We use this interface to control multimodal internet client applications, including spatial audio and panoramic browsing [18]. Such capability recalls Nokia's mission to build a "remote control for your life." We hope to eventually develop integrated teleconferencing with spatial audio via such a mobile phone[17] with full CTI (**c**omputer-**t**elephone **i**ntegration) [19], but unfortunately voice communication is currently disabled during such iαppli sessions, so a second phone must be used for an audio duplex channel.

1.1.3. Mobile Broadband: "3G"

DoCoMo i-mode and its rivals are considered "2.5 generation" devices. The next wave of iαppli hardware (the 504 series) features a half-again processor speed increase, faster download speed (from 9.6 kbps → 28.2 kbps), 10% increase in screen size (up to 2.2′′), tripling of allowed program size (to 30 kbytes), a camera, and PCM and stereo output, further tightening the integration of music functionality like MP3 players. 3G (3^{rd} generation) mobile telephone service was begun in Tokyo and Osaka last year, and in the U.S. in February (as Verizon "Express Network"). DoCoMo's 3G service, named "FoMA" (for "**F**reedom **o**f **M**obile **m**ultimedia **A**ccess"), provides videoconferencing; multiaccess capabilities, which let users receive voice calls while sending or receiving data; and enhanced i-mode service,[18] including personal calendar functions, multiplayer games, and m·commerce (for mobile commerce).[19] The service's W-CDMA (**w**ideband **c**ode-**d**ivision **m**ultiple **a**ccess) spread-spectrum communication with integrated MPEG-4[20] enables a broadband data rate of 384 kbps, fast enough for video downstreaming. (MPEG-4 is bundled in QuickTime 6.) Nissan Motor Co. and NTT Docomo are developing a "telematics" service based on 3G mobile communications technology, giving drivers access to news and weather as well as information on nearby resturants and entertainment facilities through onboard devices such as naviation systems. Taxi passengers in Singapore can already pay their fares via mobile phone.[21]

1.1.4. "ABC": "4G"

The catchphrase for 4^{th} generation mobile is "**a**lways **b**est **c**onnected." Anticipated features include wireless technology integration (linking global systems with local, like IEEE 802.11,[22] a.k.a. "Wi-Fi" ["Airport"[23] on Macintosh computers], and Bluetooth[24]), SDR (**s**oftware-**d**efined **r**adio) [20], and advanced multimedia mobile communications (IPv6, high-resolution video transmission, digitial broadcasting, security, etc.) including 3D VR interfaces. Speed is expected

[15] www.aizu.com
[16] www.gclue.com
[17] java.sun.com/products/jtapi/
[18] www.atip.org/ATIP/public/atip.reports.99/atip99.017r.html
[19] www.mobiletransaction.org
[20] www.tnt.uni-hannover.de/project/mpeg/audio
[21] www.telecab.com.sg and www.telemoneyworld.com
[22] grouper.ieee.org/groups/802/11/
[23] www.apple.com/airport
[24] www.bluetooth.com

to reach 50–100 Mbps by 2010, interpolating a 30 Mbps "3.5G" system in 2005.

1.2. Hi-Fi Telephony

Figure 4: *Convergence: Handspring VisorPhone* = PDA ∪ *telephone, integrating palmtop computing with telephony.*

The convergence described in the introduction, like that embodied by the telephonic PDA (**p**ersonal **d**igital **a**ssistant) shown in Figure 4, suggests high-bandwidth telephony, capable of high-fidelity multichannel audio. Broadband mobile phones, like those supporting FoMA, are already used to enjoy music videos, which will further motivate **hi**gh-**fi**delity audio.

Rapidly adopted technology has caused several new words to enter the vernacular. For example, "prosumer" catches the sense of a class of product and user reconciling the amateur/professional dichotomy, reflecting a more discriminating user and more affordable high-performance equipment, as cycles and bandwidth flirt with human sensitivity.

1.2.1. "Stereotelephony"

"Stereotelephony" means putting multichannel audio into the network, using stereo effects in telephones. For example, three associates with access to two lines each (like an ISDN service) can call each other cyclically, each holding the calling and called handsets

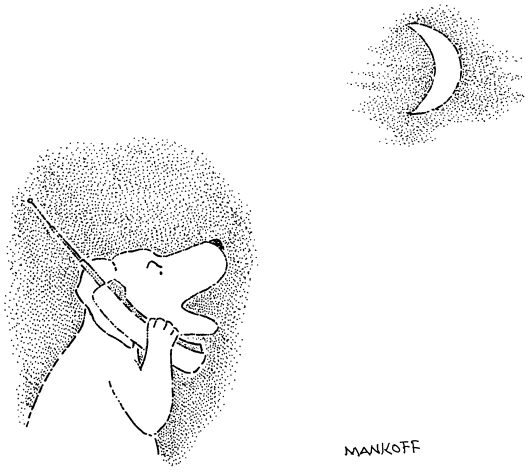


Figure 5: *Telehowl. Canine acoustic sensitivity (up to around 30 kHz, half-again the range of humans) requires extended voice bandwidth.* *(©2002 The New Yorker Collection from* cartoonbank.com*. All rights reserved.)*

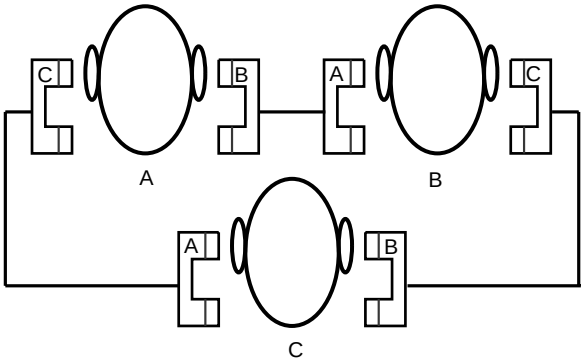


Figure 6: *Stereotelephonics and 3-way cyclic conferencing*

to different ears, as in Figure 6. Another example, illustrated by Figure 7, exploits the combination in a mobile phone of a transport mechanism with wireless capture: a dummy head equipped with (upside-down) *keitai*s enables mobile stereotelephonic telepresence.

1.2.2. "Chakumero" Ringing Melodies

A *chakumero* (チャクメロ, from 着信 (ちゃくしん) [meaning 'arrival'] plus "**melo**dy") is an incoming call tune. One may subscribe to *chakumero* services or download them from special kiosks. One can take SMFs (**s**tandard **m**IDI **f**iles), including those self-composed through DTM (**d**esk-**t**op **m**usic) systems (like Band-in-a-Box[25]), and convert them to a personal ringing melody.[26] Figure 8 imagines a contact database with name-specific musical cues. Java-implemented *karaoke* services for mobile phones[27] combine such synthesis capability with

[25] www.pgmusic.com/bandbox.htm
[26] coolweb.kakiko.com/kana2/i/eng.htm
[27] www3.dkkaraoke.co.jp

Figure 7: *"Poor person's mobile stereotelephony": a pair of inverted mobile phones, deployed as a microphone array attached to a dummy head, simultaneously calling a dual voice line (like an* ISDN *service) realizes wireless (if low-fidelity) binaural telepresence.*

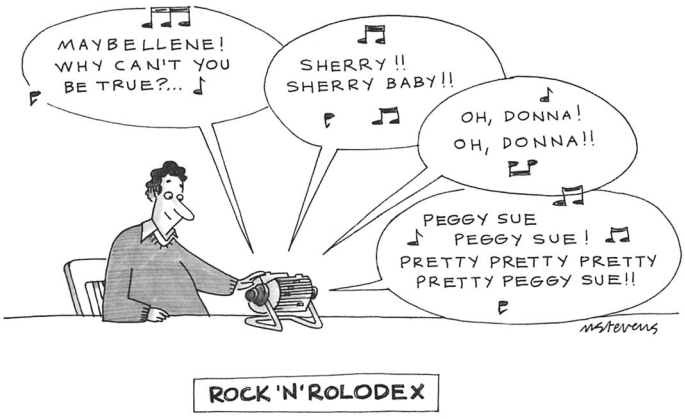


Figure 8: *Old-fashioned chakumeros. (©2002 The New Yorker Collection from* cartoonbank.com*. All rights reserved.)*

animation of the lyrics.

*1.2.3. "*FPS*" and "*MMORPG*"*

High-speed, multiplayer games have yielded a new interpretation of "fps": not "**f**rames **p**er **s**econd," but "**f**irst-**p**erson **s**hooter," as in so-called "twitch" games like "Doom" or "Quake." Typically less violent than FPS counterparts, RPGs (**r**ole-**p**laying **g**ames) depend on coherent stories, rich graphical environments, and interaction with other players. In MMORPGs (**m**assively-**m**ultiplayer **o**nline RPGs) [21]— fantasy games like Sony's "EverQuest," Cyro's "Mankind," Origin Systems' "Ultima Online," and Microsoft's "Asheron's Call"— players create characters (avatars) to explore persisitent universes that exist across sessions, and "massive" means on the order of thousands of users per server. Such games increasingly feature audio, including both locally-generated sound effects and distally transmitted voice channels. Advanced floor control in chatspaces and conferences spawned by such coteries is needed, like that suggested by Table 3.

1.3. Wearable Audio Interfaces

Figure 9: *ElekTex soft cell phone (*www.elektex.com*), with touch controls integrated into the fabric. (©ElekSen. Photo by Marcus Rose.)*

The idea of nomadic computing[28] has gone beyond hand-helds and palmtops to include wearable[29] [22] and intimate systems,[30]

[28] www.acm.org/sigmobile/mobicom/2002/

[29] www.media.mit.edu/wearables, iswc.gatech.edu

[30] www.usatoday.com/life/cyber/wireless/2001-06-26-bonus-wearable-computers.htm

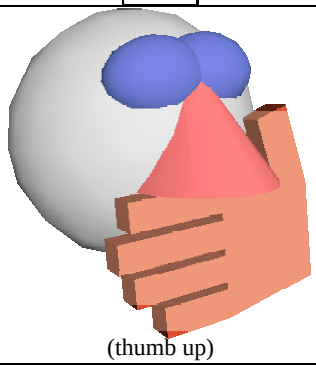
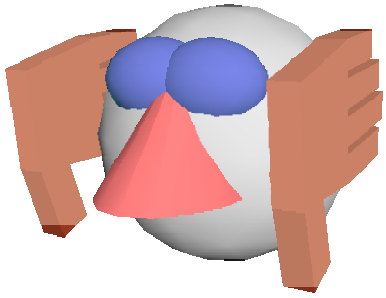
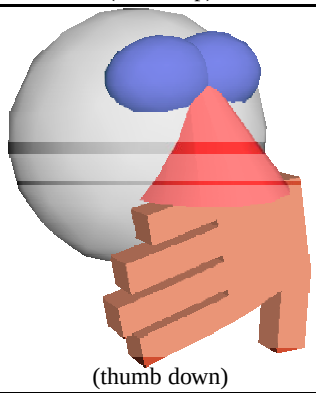
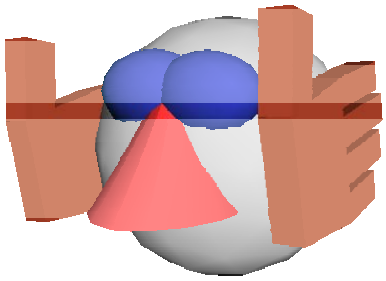
	Source	Sink
Function	radiation	reception
Level	amplification	sensitivity
Direction	OUTput	INput
Instance	speaker	listener
Transducer	loudspeaker	microphone or dummy-head
Organ	mouth	ear
Tool	megaphone	ear trumpet
Include	`solo` (`select`) or `cue`	`attend`: `confide` and `harken`
Attenuate	muzzle	muffle
Exclude	`mute`	`deafen`
own *reflexive*	 (thumb up)	 (thumbs back)
other *transitive*	 (thumb down)	 (thumbs up)

Table 3: Roles of ${}^{\mathrm{s}}\mathrm{OU}^{\mathrm{rce}}_{\mathrm{Tput}}$ and ${}^{\mathrm{s}}\mathrm{IN}^{\mathrm{k}}_{\mathrm{put}}$

including eartop [23] [24], wrist-top (like that shown in Figure 9), and collar-top.[31] Casio, Panasonic, Toshiba, Sony, and others are releasing miniature MP3 players, worn on the wrist, built into headphones, and incorporated into *keitai*s. Telecom equipment maker Motorola and watch maker Swatch have developed a wristwatch equipped with a phone, as is NTT DoCoMo.[32] Researchers at IBM Japan have developed a Linux watch "WatchPad" [25] [26] (cousin to the Java ring), with a speaker and a microphone for speech interfaces, and Nokia is working on Linux-based cell phone (although whether Linux handhelds can compete against Palm PDAs and Microsoft Pocket PCs is debatable[33]). Mitsubishi[34] recently announced the "Eye Vision" wearable display, featuring a LCD positioned in front of and below one eye (like the "Private Eye" display[35]) integrated with a stereophonic headset. The Vocollect[36] "Talkman" wearable computer uses speech recognition and synthesis to allow hands-free inventory manipulation.

Charmed[37] makes a wearable computer and an infrared badge, ViA[38] makes general-purpose wearable computers especially for hands-free applications, and Hitachi just released the "Wearable Internet Appliance (WIA-100NB)" [27]. Starlab, a Belgian re-

[31] `www.nytimes.com/2000/11/30/technology/30NEXT.html`

[32] `www.docomo-tohoku.co.jp/product/pb/pokebell/syopb_stn-ziks.html`

[33] `www.allnetdevices.com/industry/reality/2001/04/16/linux_and.html`

[34] `www.mitsubishielectric.com`

[35] `www.ndirect.co.uk/~vr-systems/priveye1.htm`

[36] `www.vocollect.com`

[37] `www.charmed.com`

[38] `www.flexipc.com`

(a) *Communicator: Early versions of the communicator were compact handheld units with a flip-up antenna grid, like the "clamshell" design of some contemporary mobile phones.*

(b) *Next Generation Combadge (**Com**munications **Badge**): Starfleet briefly used wrist communicators, but newer units were incorporated into the Starfleet insignia worn on each crew member's uniform, including a dermal sensor used to restrict usage to authorized individuals.*

Figure 11: *In many ways, we have already arrived at technology anticipated by Star Trek communicators and combadges. Communicators serve as transceivers between planet surfaces and orbiting spacecraft, and amoung members of a Starfleet landing party, and also as a means for a transporter system to determine the exact coordinates of a crew member for teleporting.*

search group, is working to develop intelligent clothing called "i-Wear"[39] containing computing and communications technology. Software and hardware developers, biologists, and fashion designers are working on a project funded by a consortium of clothing manufacturers, including Adidas, Levi-Strauss Europe,[40] and Samsonite. A multilayered approach to integrating technology into clothing, or a FAN (**f**abric-**a**rea **n**etwork), permits wireless networked communication between layers. "Soft technology" has come to mean smart interactive fabrics. Scientists at the University of Stuttgart have developed a synthetic fiber that can generate electricity from solar power,[41] potentially important in wearable computers and portable devices requiring power supplies. What could be more intimate than an article of clothing? How about a digit in an orifice?! Figure 14 illustrates utilizing sound transmission through bones: 骨伝導(こつでんどう).[42]

[39] www.starlab.org/bits/intell_clothing/project.html

[40] www.levis-icd.com

[41] www.newscientist.com/news/news.jsp?id=ns9999618

[42] www.wired.com/news/technology/0,1282,39447,00.html

2. NONPORTABLE AUDITORY INTERFACES

Personal audio interfaces— including handheld, intimate, mobile, nomadic, portable, and wearable— represent one end of a spectrum, the other end of which is marked by public spaces, which together delimit a continuum of opportunities for useful interfaces, as outlined by Table 4.

2.1. 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 Audium[44] is a specially-constructed theater featuring sound dancing around its 169 speakers in an intimate (50 seats) space [32] in San Francisco. "The House on the Rock"[45] features a collection of room-sized musical robots, one of which is shown in Figure 15. 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 5— but impractical for anything besides special-purpose venues and LBE. Fully articulated spatial audio [34] [35] [36] allows dynamic (runtime), arbitrary

[44] www.audium.org

[45] www.houseontherock.com/the_attraction.htm

(a) *Mooring* (b) MP3 *player* (c) *Remote Control*

Figure 12: *Phillips and Levi-Strauss have teamed up to develop jackets with built-in electronic equipment, including an* MP3 *player, headset, mobile phone handset, and remote control. The system is equipped with a so-called "*Personal Area Network*," or* PAN*, an electronic bus woven into jacket that allows transport of data, power, and control signals within the garment.*

(a) Vest (b) HMD *(***h***ead-***m***ounted or -worn* **d***isplay*

Figure 13: *Xybernaut's patented MA4 system (*`www.xybernaut.com`*) features a computer (2 pounds) worn on utility belt alongside a battery. The head-worn display incorporates earphones, a microphone, and a screen about 4 cm wide. Costing about $2000, it is being marketed to companies supporting maintenance workers, who need to access manuals, maps, references, etc. while using their hands for repair.*

Proxemic Context	architecture	Display: audio	Display: visual
intimate	headset, wearable computers	*eartop* headphones, ear buds	*eyetop* HWDs (**h**ead-**w**orn **d**isplays) HMDs (**h**ead-**m**ounted **d**isplays)
personal	chair	nearphones	*laptop* display, *desktop* monitor
interpersonal	couch or bench	transaural speakers (ex: SDP (**s**tereo **dip**ole) [28])	HDTV
multipersonal	automobiles, **s**patially **i**mmersive **d**isplays (ex: Cave,™ Cabin)	surround sound (ex: Ambisonics)	projection
social	clubs, theaters	speaker array (ex: VBAP [29])	large-screen displays (ex: IMAX[43])
public	stadia, concert arenas	**p**ublic **a**ddress	(ex: Jumbotron)

Table 4: *Audio and visual displays along a private↔public continuum.*

Position

Static **Dynamic**

Location	**Scalar**
lateral displacement	abscissa x
frontal displacement	ordinate y
height	altitude z

Translation; 基本動作		**Along Axis**	**Perpendicular to Plane**
sway; 左右動作	left↔right	x	sagittal
surge; 前後動作	back (aft) ↖↘ forth (fore)	y	frontal
heave; 上下動作	up ↕ down	z	horizontal

Orientation or Attitude	
elevation or tilt	ϕ
(roll)	ψ
azimuth	θ

Rotation; 基本ゆれ		**About Axis**	**In Plane**
pitch; 縦ゆれ	climb/dive	x	sagittal
roll; 横ゆれ	left/right	y	frontal
yaw; 偏 or へん ゆれ	CW/CCW	z	horizontal

Table 6: *Physically spatial dimensions: taxonomy of positional degrees of freedom.*

placement and movement of multiple sources in soundscapes, including those spatial dimensions presented by Table 6— as well as control of extra dimensions [37] like apparent extent (including ASW: **a**uditory **s**ource **w**idth), orientation, directivity, and environmental characteristics, outlined by Table 7— for true cyberspatial capabilities.

2.2. Information Furniture and Robots

The Inada "Medical Chair H.9"[46] is a synæsthetic seat, synchronizing *shiatsu* massage with music. Our own research group's "Internet Chair" [38] [39] [40] includes a swivel chair with transaural "nearphones" (for "near earphones"), directionalizing audio using dynamically-selected transfer functions determined by chair rotation, as shown in Figure 16. (Our prototypes are being developed with Mechtec.[47]) With researchers from Yamagata University, our group is developing this concept of information furniture [41] [42], integrating a servomotor with the chair to allow active control as well as passive sensing. Further, we are working on integrating it with a hearing telerobot, shown in Figure 17, equipped with four microphones arrayed on its spherical head [43], streaming via JMF (Java Media Foundation[48]). Commercially, Sanyo has unveiled prototypes of robot guard dogs that include built-in speakers and cameras that can link to their masters' mobile phones,[49] and Sony's newest robotic dog, the Aibo ERS-311B, includes Bluetooth communication and speech synthesis.[50]

2.3. Ultrasonics

Ultrasonic-based display systems, which create audible signals through propagation distortion, non-linear effects on air of ultrasonic sig-

[46] www.inada-chair.com

[47] www.mechtec.co.jp

[48] java.sun.com/products/java-media/jmf/

[49] www.sanyo.co.jp/koho/hypertext4/0203news-j/0325-1.html

[50] www.sony.jp/CorporateCruise/Press/200205/02-0516/

Figure 15: *In the setting of a baroque music chamber, a robotic musical ensemble performs in the Blue Room at "House on the Rock," an amazingly unique museum outside Madison, Wisconsin (about four hours from Chicago). The intensity of which the Blue Room is capable necessitates the acoustical canopy and full-carpeting, lest the rococo mirrors and candelabrum shatter.*

Figure 10: *Dieceland Technology disposable mobile phone (*`www.dtcproducts.com`*): This 2″ × 3″ (the size of a thick credit card) disposable ($10) mobile phone is made by printing cell-phone circuitry onto a paper substrate, which is then laminated.*

Figure 14: *NTT DoCoMo "Whisper Wearable Terminal." By sticking her finger in her ear, a user can hear sound conducted through arm, hand, and finger bones while speaking into a microphone on the back of the wristband, snapping fingers to signal "connect" and "disconnect" commands. The manufacturer hopes to release a commercial model in 2005. Presumably a stereotelephonic system could be realized by sticking fingers in both ears.*

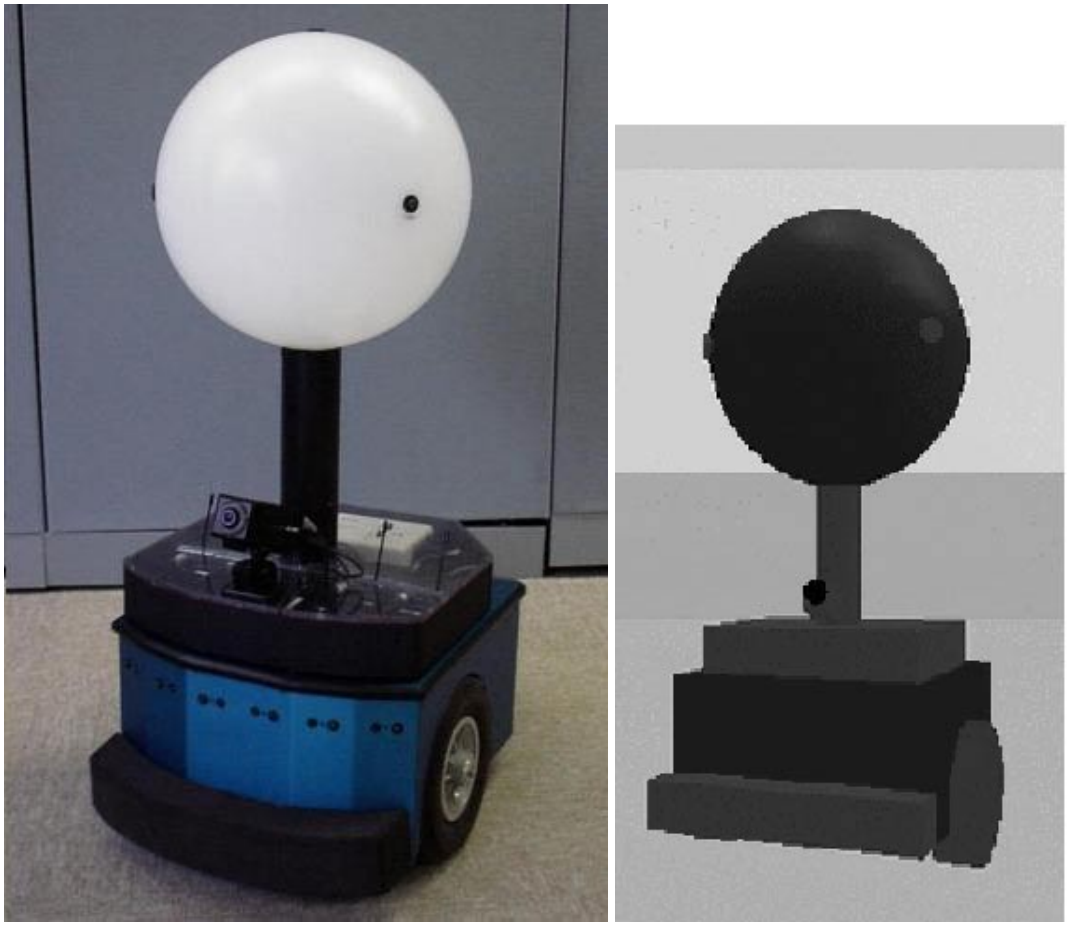


Figure 17: *"Hero" Labo-3 mobile* **he***aring tele***ro***bot, operable in piloted and autonomous (automatic obstacle avoidance) modes, with cyclopean (one-eyed) camera and quadraphonic tetrahedral microphone array. (Deployed in lab of J. Huang;* 黄 捷*, and programmed by Yasuhiro Yamazaki;* 山崎 泰宏*. Java3D model by Shūnō Kazuki;* 収納 和樹*. Streaming media functionality programmed by Yanagi Tomohide;* 柳 智英*.)*

nals (nominally above around 20 kHz, but in current practice around 40–60 kHz), have been researched for decades (as "parametric acoustic arrays"), but heretofore no practical systems were available. Recently released products, including Holosonic Research Lab's Audio Spotlight[51] and American Technology's "Hypersonic Sound System,"[52] feature ultrasonic sound beams, personalizable through their focus and controllable dispersion (reportedly as low as 3°), including the ability to direct reflections off walls, etc. If technical issues regarding such systems' lower-frequency response (since there is a 12 dB/octave high pass slope, a direct result of the way that ultrasound transfers energy into the audible range) and concerns about health hazards (as the inaudible sounds are very intense, in the range of 140-150 dB SPL!) can be resolved, ultrasonic-based audio displays could be as flexible as analogous light-based visual displays.

3. CONCLUSION

It's ironic that the participle "wired," formerly flattering, is now pejorative, connoting a cumbersome tether. The emergence of mobile internet,[53] global roaming, software-defined radio, wearable computer interfaces, ubiquitous (or "calm" or "ambient") computing, spatially immersive displays, and information furniture offer opportunities for innovative design and advanced applications, both creative and "re-creative" [44]. Enriched connectivity will foster use of the `deafen`/`attend`: `harken` or `confide` sink selection functions [45], like those outlined in Table 8, illustrated by Table 3, and formalized by Figure 21. The feedback between co-potentiating hardware companies (like those making mobile phones) and software content providers, the composition of infrastructure and applications, will continue to blur the distinction between designers & engineers, art & science, and invention & discovery.

[51] `www.media.mit.edu/~pompei/spotlight`
[52] `www.atcsd.com/tl_hss.html`
[53] `www.mobilehci.org`

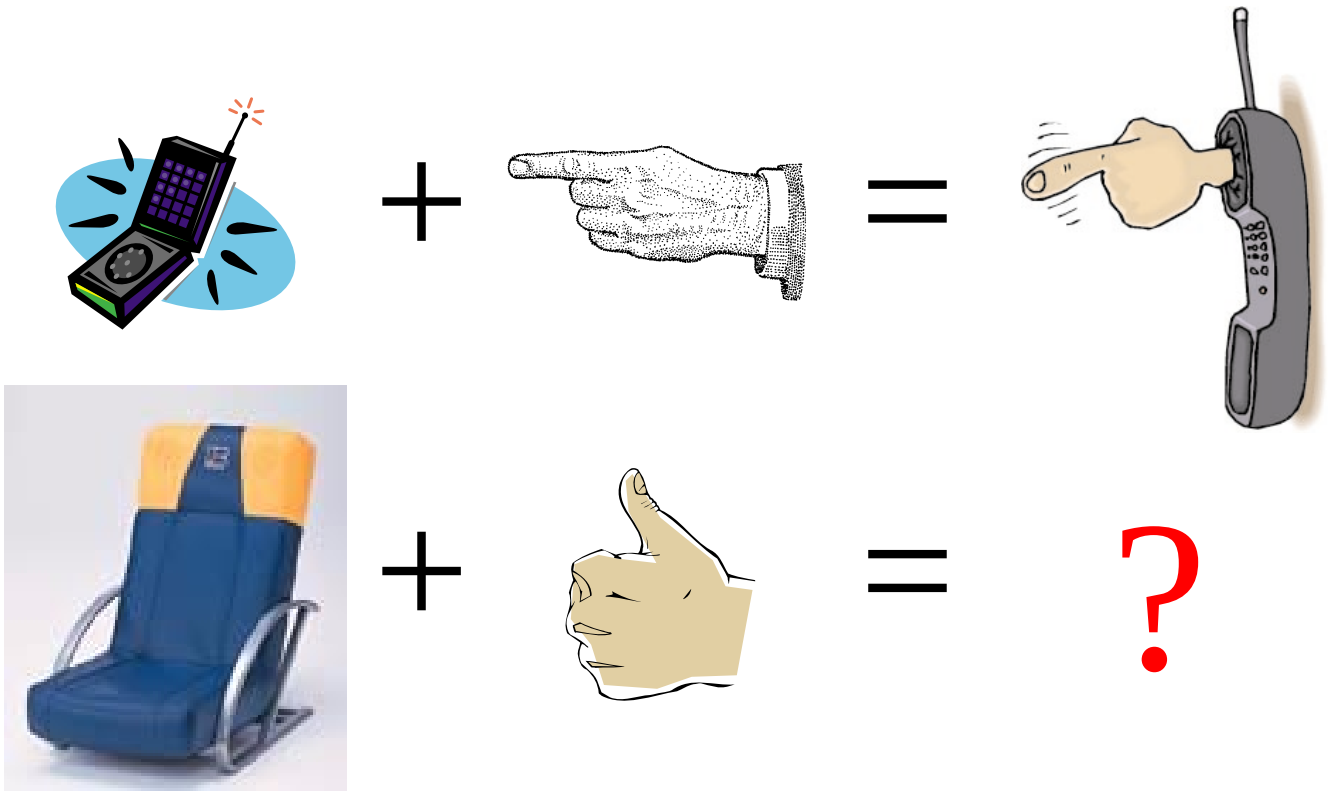


Figure 18: *From a bit in the office to a digit in the orifice. Some gaming stations and* LBE *cockpits feature* LFEs *(**l**ow-**f**requency **e**ffects), delivered via BodySonic chairs or similar equipment with subwoofers in the seat for visceral sensation. (Such chairs can be found, for instance, at Narita Airport, Terminal 2 in the* 3^{rd} *floor Refresh Lounge.) How might finger-conducted sound, like that illustrated by Figure 14, be posteriorly ("bunaurally") delivered?*

The general expression of inclusive selection is

$$\texttt{active}(\mathrm{x}) \;=\; \neg\texttt{exclude}(\mathrm{x}) \;\wedge\; (\exists\, \mathrm{y}\; \texttt{include}(\mathrm{y}) \;\Rightarrow\; \texttt{include}(\mathrm{x})). \tag{1}$$

So, for `mute` and `solo` (or `select`), the relation is

$$\texttt{active}(\mathrm{source_x}) \;=\; \neg\texttt{mute}(\mathrm{source_x}) \;\wedge\; (\exists\, \mathrm{y}\; \texttt{solo}(\mathrm{source_y}) \;\Rightarrow\; \texttt{solo}(\mathrm{source_x})), \tag{2a}$$

`mute` explicitly turning off a source, and `solo` disabling the collocated (same room/window) complement of the selection (in the spirit of "anything not mandatory is forbidden"). For `deafen` and `attend`, the relation is

$$\texttt{active}(\mathrm{sink_x}) \;=\; \neg\texttt{deafen}(\mathrm{sink_x}) \;\wedge\; (\exists\, \mathrm{y}\; \texttt{attend}(\mathrm{sink_y}) \;\Rightarrow\; \texttt{attend}(\mathrm{sink_x})). \tag{2b}$$

Figure 21: *Formalization of narrowcasting and selection functions in predicate calculus notation, where '¬' means "not," '∧' means conjunction (logical "and"), '∃' means "there exists," and '⇒' means "implies." The suite of inclusion and exclusion narrowcast commands for sources and sinks are like analogs of burning and dodging (shading) in photographic processing. The duality between source and sink operations is tight, and the semantics are identical: a mixel is inclusively enabled by default unless, a) it it explicitly excluded (with* $\overbrace{\texttt{mute}}^{\text{source}}$ *or* $\overbrace{\texttt{deafen}}^{\text{sink}}$*), or, b) peers are explicitly included (with* $\overbrace{\texttt{solo}\ [\textit{or}\ \texttt{select}]}^{\text{sources}}$ || $\overbrace{\texttt{attend} : \texttt{confide}\ \textit{or}\ \texttt{harken}}^{\text{sinks}}$*) when the respective icon is not.*

	General		Visual		Audio	
	Source	Sink	Source	Sink	Source	Sink
Own, Self (*reflexive*)	hide/ manifest, appear, project	ignore/ `attend`, notice	hide, conceal/ expose	avert/ see	`mute` (stifle)/ `cue` (play, speak)	`deafen`/ `harken` (heed)
Other (*transitive*)	hide/ expose	block/ invite, select	hide/ expose	mask, blind/ ?	`mute`/ `solo` (`select`)	`deafen`/ `confide`

Table 8: `Exclude/ Include` taxonomy: enable/disable for one's own and others' representatives

	Peripheral	Network
Wired	USB IEEE 1394: Firewire	Ethernet
Wireless	Bluetooth	WLAN: IEEE 802.11 (Wi-Fi, Airport) cdma2000, W-CDMA, etc.

Table 9: {*Wired, Wireless*} × {*Peripheral, Network*}.

4. REFERENCES

[1] Jeremy R. Cooperstock, Sidney S. Fels, William Buxton, and Kenneth C. Smith, "Reactive environments: Throwing away your keyboard and mouse," *Communications of the* ACM, vol. 40, no. 9, pp. 65–73, Sept. 1997.

[2] Norbert A. Streitz, Jane Siegel, Volker Hartkopf, and Shin'ichi Konomi, Eds., *Cooperative Buildings: Integrating Information, Organizations, and Architecture (CoBuild'99: Proc. of the Second Int. Workshop)*, Pittsburgh, Oct. 1999. Springer (LNCS 1670), ISBN 3-540-66596-X.

[3] Rodney A. Brooks, "The intelligent room project," in *Proc. CT: Int. Conf. on* **C***ognitive* **T***echnology*, Aizu-Wakamatsu, Japan, Aug. 1997, IEEE, pp. 271–278, ISBN 0-8186-8084-9.

[4] Ted Lewis, "Information appliances: Gadget netopia," *(*IEEE*) Computer*, pp. 59–68, Jan. 1998, ISSN 0018-9162.

[5] Steve Mann, "Smart clothing: The wearable computer and wearcam," *Personal Technologies*, vol. 1, no. 1, pp. 21–27, 1997, ISSN 0949-2054 and ISSN 1433-3066.

[6] Mark Billinghurst and Thad Starner, "Wearable devices: New ways to manage information," *IEEE Computer*, vol. 32, no. 1, pp. 57–64, Jan. 1999.

[7] Woodrow Barfield and Thomas Caudell, Eds., *Fundamendals of Wearable Computers and Augmented Reality*, Lawrence Erlbaum Assoc., 2001, ISBN 0-8058-2902-4.

[8] Jeffrey Lee Funk, *The Mobile Internet: How Japan Dialed Up and the West Disconnected*, ISI Publications, 2001, ISBN 9-6277-6269-5.

[9] Lillian Goleniewski, *Telecommunications Essentials*, Addison Wesley Professional, 2001, ISBN 0-201-76032-0.

[10] Savitha Srinivasan and Eric Brown, "Is speech recognition becoming mainstream?," *(*IEEE*) Computer*, pp. 38–41, Apr. 2002.

[11] ASCII editing group, *iMode Java Programming*, ASCII, 2001, ISBN 4-7561-3727-X.

[12] Yukinori Yamazaki, *How to make an iAppli*, SOFTBANK, 2001, ISBN 4-7973-1573-3.

[13] Yutaka Nagashima and Michael Cohen, "Distributed virtual environment interface for a mobile phone," *3D Forum: J. of Three Dimensional Images*, vol. 15, no. 4, pp. 102–106, Dec. 2001, ISSN 1342-2189.

[14] Toshifumi Kanno, Michael Cohen, Yutaka Nagashima, and Tomohisa Hoshino, "Mobile control of multimodal groupware in a distributed virtual environment," in *Proc. ICAT:* **I***nt.* **C***onf.* **A***rtificial Reality and* **T***ele-Existence*, Tokyo, Dec. 2001, pp. 147–154, ISSN 1345-1278; `sklab-www.pi.titech.ac.jp/~hase/ICATPHP/upload/39_camera.pdf`.

[15] Michael Cohen and William L. Martens, "Spatial Media Research at the University of Aizu," *JVRSJ:* **J***.* **V***irtual* **R***eality* **S***ociety of* **J***apan*, vol. 6, no. 2, pp. 52–57, Sept. 2001, ISSN 1342-6680; `www.u-aizu.ac.jp/~mcohen/welcome/publications/spatial-media.ps`.

[16] William L. Martens and Michael Cohen, "Virtual Acoustic Research at the University of Aizu," *JVRSJ:* **J***.* **V***irtual* **R***eality* **S***ociety of* **J***apan*, vol. 6, no. 3, pp. 261–269, Dec. 2001, ISSN 1342-6680.

[17] National Research Council, "The auditory channel," in *Virtual Reality— Science and Technological Challenges*, Nathaniel I. Durlach and Anne S. Mavor, Eds., chapter 3, pp. 134–160. National Academy Press, 1995, ISBN 0-309-05135-5.

[18] Noor Alamshah Bolhassan, William L. Martens, and Michael Cohen, "VR_4U_2C: A Multiuser Multiperspective Panoramic Browser Using QuickTime VR and Java Featuring Multimonitor and Stereographic Display," in *Proc. ICAT:* **I***nt.* **C***onf.* **A***rtificial Reality and* **T***ele-Existence*, Tokyo, Dec. 2001, pp. 161–168.

[19] Spencer Roberts, *Essential* JTAPI*: Java telephony*, Prentice Hall, 1999, ISBN 0-13-080360-X.

[20] Benny Bing and Nikil Jayant, "A cellphone for all standards," *(*IEEE*) Spectrum*, vol. 39, no. 5, pp. 34–39, May 2002.

- concert
  - chorus
    - soprano
    - alto
    - tenor
    - bass
  - orchestra
    - strings
      - basses
      - cellos
      - violas
      - violins
        - G-string
        - D-string
        - A-string
        - E-string
          - attack
          - decay
            - even harmonics
            - odd harmonics
    - brass
      - horns
      - trumpets
      - trombones
      - tuba
    - woodwinds
      - bassoons
      - clarinets
      - flutes
      - oboes
    - percussion
      - bass drum
      - cymbals
      - snare drum
      - triangle
      - tubular bells
      - wood block
      - xylophone
      - timpani
    - other
      - harp
      - piano

Table 5: *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.*

Figure 16: *Internet Chair with servomotor rotation: a pivot (swivel, rotating) chair deployed as an* I/O *device, an information appliance. The input modality is orientation tracking, which can dynamically select transfer functions used to spatialize audio in a rotation-invariant soundscape. In groupware situations, like teleconferencing or chat spaces, such orientation tracking can also be used to twist iconic representations of a seated user, avatars in a virtual world, enabling social situation awareness via coupled visual displays, soundscape-stabilized virtual source locations, and direction-dependent projection of non-omnidirectional sources. As an audio output modality, transaural speakers (without crosstalk), nearphones in the headrest, can present unencumbered binaural sound with soundscape stabilization for multichannel sound image localization. As a haptic output modality, servomotors will render kinesthetic force-feedback, turning each* $\mathrm{{}^{s}_{c}ha_{i}r^{e}}$ *(pronounced "schaire," for "share-chair") under networked control, distributing torque across the internet to direct the attention seated subjects (with adjustable insistence/forcefulness), orienting seated users (like a "dark ride" amusement park attraction), or subtly nudging them in a particular direction. (Java3D model by Daisuke Kaneko;* 金子 大輔*.)*

- frequency content
  - pitch and register: tone, melody, harmony
  - waveshape (sawtooth, square, triangle, ...)
  - timbre, filtears [33], vibrato, and equalization
- dynamics
  - intensity/volume/loudness
  - envelope: **a**ttack, **d**ecay, **s**ustain, **r**elease (volume shape)
- timing
  - duration
  - tempo
  - repetition rate
  - duty cycle
  - rhythm and cadence
  - syncopation
- spatial position: location and orientation
  - direction: azimuth, elevation
  - distance/range
  - directivity: attitude and focus
- ambience: environmental spatial impression
  - presence
  - resonance, reverberance
  - spaciousness
  - envelopment
- representationalism: literal, everyday ("auditory icons") ↔ abstract ("earcons")

Table 7: *Dimensions of sound.*

Figure 19: *Lavatory Laboratory.*

[21] Elizabeth Kolbert, "Pimps and Dragons: How an online world survived a social breakdown," *The New Yorker*, pp. 88–98, May 28 2001.

[22] Thad E. Starner, "Wearable computers: No longer science fiction," *Pervasive Computing*, vol. 1, no. 1, pp. 86–88, January–March 2002.

[23] Michael Cohen, Jens Herder, and William L. Martens, "Cyberspatial Audio Technology," **J**. **A***cous.* **S***oc.* **J***ap. (English)*, vol. 20, no. 6, pp. 389–395, Nov. 1999, ISSN 0388-2861; `www.u-aizu.ac.jp/~mcohen/welcome/publications/JASJ-reviewE.ps`.

[24] Michael Cohen, Jens Herder, and William L. Martens, "Cyberspatial Audio Technology," **J**. **A***cous.* **S***oc.* **J***ap.*, vol. 55, no. 10, pp. 730–731, Oct. 1999, (In Japanese). ISSN 0388-2861; `www.u-aizu.ac.jp/~mcohen/welcome/publications/JASJ-reviewJ.ps`.

[25] Chandra Narayanaswami, Noboru Kamijoh, Mandayam Raghunath, Tadanobu Inoue, Thomas Cipolla, Jim Sanford, Eugene Schlig, Sreekrishnan, Venkiteswaran, Dinakar Guniguntala, Vishal Kulkarni, and Kazuhiko Yamazaki, "IBM's Linux Watch: The Challenge of Miniaturization," *(*IEEE*) Computer*, vol. 35, no. 1, pp. 33–41, Jan. 2002.

[26] Noboru Kamijo, Tadanobu Inoue, Kohichiroh Kishimoto, and Ken Tamagawa, "Research for Wearable Computing: WatchPad 1.5: The Challenge of Miniaturization," *JVRSJ:* **J**. **V***irtual* **R***eality* **S***ociety of* **J***apan*, vol. 7, no. 1, pp. 78–80, 3 2002, (In Japanese).

[27] Yoshinobu Ohinata, "Wearable Computer Release," *JVRSJ:* **J**. **V***irtual* **R***eality* **S***ociety of* **J***apan*, vol. 7, no. 1, pp. 80–81, 3 2002, (In Japanese).

[28] Ole Kirkeby, Philip A. Nelson, and Hareo Hamada, "The stereo dipole — binaural sound reproduction using two closely spaced loudspeakers," in *Proc. AES:* **A***udio* **E***ngineering* **S***ociety Conv.*, Munich, Mar. 1997, Preprint 4463 (I6).

[29] Ville Pulkki, "Virtual source positioning using vector base amplitude panning," *J.* **A***ud.* **E***ng.* **S***oc.*, vol. 45, no. 6, pp. 456–466, June 1997.

[30] Katsumi Amano, Fumio Matsushita, Hirofumi Yanagawa, Michael Cohen, Jens Herder, William Martens, Yoshiharu Koba, and Mikio Tohyama, "A Virtual Reality Sound System Using Room-Related Transfer Functions Delivered Through a Multispeaker Array: the PSFC at the University of Aizu Multimedia Center," *TVRSJ:* **T***rans.*

Figure 20: *The price of privacy.* *(©2002 The New Yorker Collection from* cartoonbank.com*. All rights reserved.)*

V*irtual* **R***eality* **S***ociety of* **J***apan*, vol. 3, no. 1, pp. 1–12, Mar. 1998, ISSN 1344-011X; www.u-aizu.ac.jp/~mcohen/welcome/publications/PSFC.ps.

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

[32] Gareth Loy, "About Audium— A Conversation with Stanley Shaff," *Computer Music Journal*, vol. 9, no. 2, pp. 41–48, 1985, ISSN 0148-0267.

[33] Michael Cohen, "Throwing, pitching, and catching sound: Audio windowing models and modes," *IJMMS: the Journal of Person-Computer Interaction*, vol. 39, no. 2, pp. 269–304, Aug. 1993, ISSN 0020-7373; www.u-aizu.ac.jp/~mcohen/welcome/publications/tpc.ps.

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

[35] Simon Carlile, Ed., *Virtual Auditory Space: Generation and Applications*, Springer, 1996, ISBN 1-57059-341-8, 0-412-10481-4, and 3-40-60887-7.

[36] Fancis Rumsey, *Spatial Audio*, Focal Press, 2001, ISBN 0-240-51623-0.

[37] Michael Cohen and Elizabeth M. Wenzel, "The design of multidimensional sound interfaces," in *Virtual Environments and Advanced Interface Design*, Woodrow Barfield and Thomas A. Furness III, Eds., chapter 8, pp. 291–346. Oxford University Press, 1995, ISBN 0-19-507555-2.

[38] Michael Cohen, "The Internet Chair," in *Proc. ICAT:* **I***nt.* **C***onf.* **A***rtificial Reality and* **T***ele-Existence*, Tokyo, Dec. 1999, VRSJ, pp. 29–36, www.u-aizu.ac.jp/~mcohen/welcome/publications/ic3.ps.

[39] Nobuo Koizumi, Michael Cohen, and Shigeaki Aoki, "Japanese patent #3042731: Voice reproduction system," Mar. 2000.

[40] Michael Cohen, "The Internet Chair," *IJHCI: Int. J. of Human-Computer Interaction*, vol. 15, no. 2, 2003, www.u-aizu.ac.jp/~mcohen/welcome/publications/ic4.ps. In press.

[41] Hong Z. Tan, Ifung Lu, and Alex Pentland, "The chair as a novel haptic interface," in *Proc. Workshop on Perceptual User Interfaces*, M. Turk, Ed., Banff, Alberta, Canada, Oct. 1997, pp. 56–57, dynamo.ecn.purdue.edu/~hongtan/Hongtan-pub/Abs/6.PUI97.abs.html.

[42] Hong Z. Tan, "Haptic interfaces," *Communications of the* ACM, vol. 43, no. 3, pp. 40–41, Mar. 2000.

[43] Yasuhiro Yamazaki, Michael Cohen, Jie Huang, and Tomohide Yanagi, "Augmented audio reality: compositing mobile telerobotic and virtual spatial audio," in *Proc. ICAT:* **I***nt.* **C***onf.* **A***rtificial Reality and* **T***ele-Existence*, Tokyo, Dec. 2001, pp. 213–216, ISSN 1345-1278; sklab-www.pi.titech.ac.jp/~hase/ICATPHP/upload/41_camera.pdf.

[44] F. Alton Everest and Ron Streicher, *The New Stereo Soundbook*, Audio Engineering Associates, Pasadena, CA, second edition, 1998, ISBN 0-9665162-0-6.

[45] Michael Cohen, "Exclude and include for audio sources and sinks: Analogs of mute & solo are deafen & attend," *Presence: Teleoperators and Virtual Environments*, vol. 9, no. 1, pp. 84–96, Feb. 2000, ISSN 1054-7460; www.u-aizu.ac.jp/~mcohen/welcome/publications/ie1.pdf.