Audio Engineering Society
Convention Paper 5819
Presented at the 114th Convention
2003 March 22–25 Amsterdam, The Netherlands

Emerging and Exotic Auditory Interfaces

Michael Cohen

[1] *Spatial Media Group, Human Interface Lab.; University of Aizu; Aizu-Wakamatsu 965-8580; Japan*

Correspondence should be addressed to Michael Cohen (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, popular media, and publications by industrial and academic laboratories, and the author's own research group. Such extended and enriched audio interfaces, especially coupled with position tracking systems, encourage multipresence, the inhabiting by sources and sinks of multiple spaces simultaneously, allowing, for instance, a user to monitor several aligned spaces at once (conferences, entertainment, navigation, warnings, etc.). Representative instances are cited, and groupware selection functions and their predicate calculus notation 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.

INTRODUCTION

The digital "4C ['foresee'] convergence" is the confluence of **c**ommunication devices, **c**omputing, and **c**onsumer electronics, and **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. This survey outlines emerging auditory control and display interfaces. The author's research group is investigating interfaces related to robotics, **s**patially **i**mmersive **d**isplays, information furniture, and mobile computing.

PORTABLE AUDITORY INTERFACES

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 'portable [device]' or 'mobile [phone]'). 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.

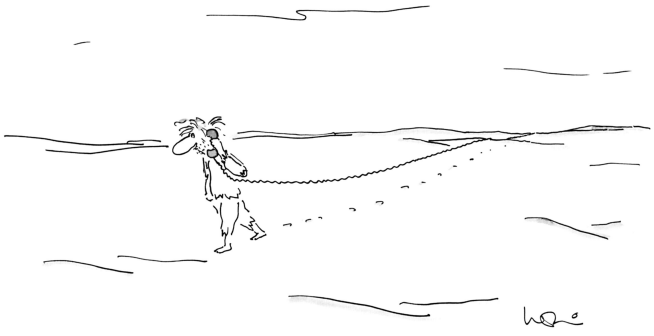


Fig. 1: Non-wireless telephony. (©2003 The New Yorker Collection from cartoonbank.com. All rights reserved.)

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

[2] `www.viktoria.se/ubicomp/`

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 Japan's 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" texting features like SMSs (**s**hort **m**essage **s**ervices) and IM (**i**nstant **m**essaging), web surfing, and even conversing. NTT DoCoMo "i-mode" [9] (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, wager on horse races, and so on.

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 [10, 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.

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

[4] `www.nttdocomo.com`

[5] `www.attws.com/mmode`

[6] `www.nintendo.com`

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.

Parent Company	**Subsidiary**	(2G) **Internet**	(2.5G) **Java Capability**	(3G) **Broadband**
NTT	DoCoMo	i-mode	iαppli	FoMA— W-CDMA: 64 kbps upstream, 384 kbps down
KDDI	au (H˜)	EZ-web	EZplus	cdma2000-1×: 144 kbps downstream
Japan Telecom	J-Phone	J-Sky	Java Appli	128 kbps upstream, 384 kbps down

Table 2: Japanese mobile telephone services.

Any telephone keypad is awkward for text input, cumbersome as it is to thumb-type up to eight keypresses for mixed-case characters. Mobile browsers will be naturally extended by voice interfaces, using VXML (for VoiceXML[7]) [11] 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) [12]. 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.

MMS (**m**obile **m**ultimedia **s**ervices) are extended by technology like 3DMe's,[9] which can analyse voice mail prosody to extract intended mood, animating emotional characters with lib synchronization.

[7] www.w3.org/Voice
[8] www.naturalvoices.att.com
[9] wwww.3dme.com

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.[10] Steve Wozniak has launched "Wheels of Zeus,"[11] 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.

J2ME[12] is Sun's Java2 microedition engine,[13] 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

[10] www.popsci.com/popsci/science/article/0,12543,266052-2,0=0.html
[11] www.w0z.com
[12] java.sun.com/j2me
[13] java.sun.com/products/cldc, www.javamobiles.com, wwww.microjava.com/devices, wireless.java.sun.com/devices

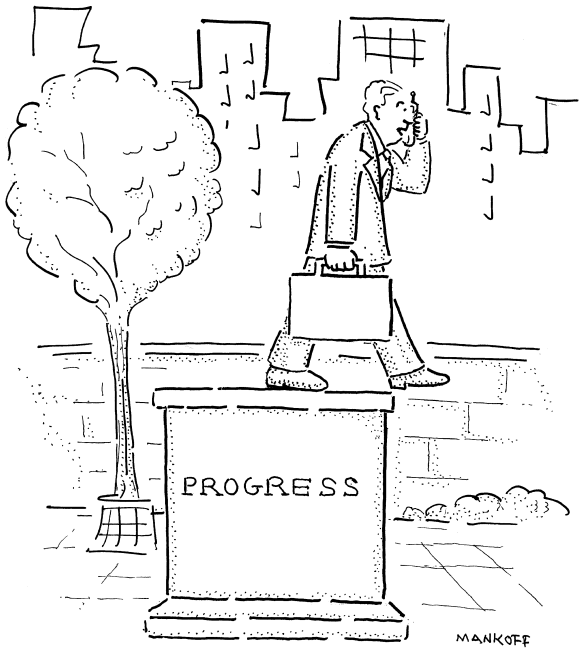


Fig. 2: Preoccupied by his conversation, this everyman is in danger of striding off the pedestal of progress. (©2003 The New Yorker Collection from cartoonbank.com. All rights reserved.)

services).[14] DoCoMo offers a Java-based service called "iαppli" [13] [14], 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, the RIM BlackBerry[15] features an embedded radio-modem, and the Palm Tungsten[16] is designed for wireless voice and data services.

Our own research group has designed and implemented a mobile telephone interface [15] for use in CVEs (**c**ollaborative **v**irtual **e**nvironments) [16] [17] [18]. Anticipating ubicomp networked appliances[17] and information spaces, we are integrating various multimodal (auditory, visual, haptic)

[14] www.atip.org/public/atip.reports.02/atip02.022.pdf
[15] www.blackberry.net
[16] www.palm.com/products/handhelds/tungsten-w
[17] computer.org/pervasive

Fig. 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 and Makoto Kawaguchi.)

I/O devices into a virtual reality groupware suite. Programmed with J2ME, one of our client applications runs on a mobile phone, as illustrated by Figure 3. The "ι·Con" 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 iαppli models feature a thumb jog shuttle, which can be used as a continuous controller to manipulate such icons. The interface is further extended with musical and vi-

brational 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"[18] in conjunction with "GClue."[19]) We use this interface to control multimodal internet client applications, including spatial audio and panoramic browsing [20, 21]. Such capability recalls the oft-aspired mission to build a "remote control for your life." We hope to eventually develop integrated teleconferencing with spatial audio via such a mobile phone[20] with full CTI (**c**omputer-**t**elephone **i**ntegration) [22], but unfortunately voice communication is currently disabled during such iαppli sessions, so a second phone must be used for an audio duplex channel.

Mobile Broadband: "3G"

DoCoMo i-mode and its rivals are considered "2.5 generation" devices. The newest wave of iαppli hardware 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 (3rd generation) mobile telephone service was begun in Japan 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, including personal calendar functions, multiplayer games, and e· & m·commerce (for electronic & mobile commerce).[21] The service's W-CDMA (**w**ideband **c**ode-**d**ivision **m**ultiple **a**ccess) spread-spectrum communication with integrated MPEG-4[22] has a broadband data rate of 384 kbps, fast enough for video downstreaming, enabling applications that use mobile phones as digital TV terminals, like DoCoMo's "i-motion" service that offers video clips from various providers. (MPEG-4 is also bundled in QuickTime 6.) KDDI plans to initiate a higher-speed data transmission service, cdma2000-1×-EV-DO (for **ev**olution **d**ata **o**nly), in late 2003, with a top download speed of 2.4 Mbps. Nissan Motor Co. and 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.[23]

"ABC": "4G"

The catchphrase for 4th generation mobile is "**a**lways **b**est **c**onnected." Anticipated features include wireless technology integration (linking global systems with local, like IEEE 802.11,[24] a.k.a. "Wi-Fi" ["Airport"[25] on Macintosh computers], and Bluetooth[26]), SDR (**s**oftware-**d**efined **r**adio) [23], and advanced multimedia mobile communications (IPv6, high-resolution video transmission, digitial broadcasting, security, etc.) including 3D VR interfaces. Speed is expected to start at 20 Mpbs, reaching 50–100 Mbps by 2010, interpolating a 30 Mbps "3.5G" system in 2005.

Hi-Fi Telephony

[18] www.aizu.com
[19] www.gclue.com
[20] java.sun.com/products/jtapi
[21] www.mobiletransaction.org
[22] www.tnt.uni-hannover.de/project/mpeg/audio
[23] www.telecab.com.sg and www.telemoneyworld.com
[24] grouper.ieee.org/groups/802/11
[25] www.apple.com/airport
[26] www.bluetooth.com

Fig. 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.

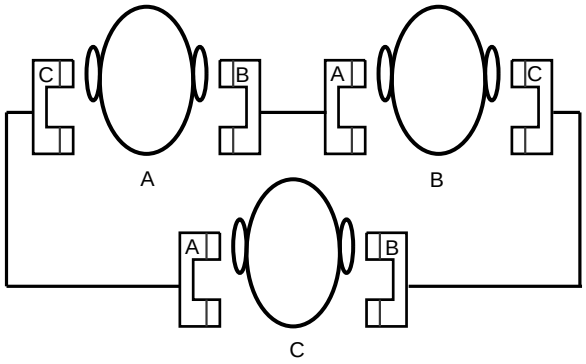


Fig. 5: Stereotelephonics and 3-way cyclic conferencing

"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 to different ears, as in Figure 5.

Another example, illustrated by Figure 6, exploits the combination in a mobile phone of a transport mechanism with wireless capture: a dummy head equipped with (upside-down) handy phones enables mobile stereotelephonic telepresence.

"*Chakumero*" Ringing Melodies

A *chakumero* (from the Japanese for '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, and convert them to a personal ringing melody.[27] Figure 7 imagines a contact database with name-specific musical cues. Java-implemented *karaoke* services for mobile phones[28] combine such synthesis capability with animation of the lyrics.

[27] coolweb.kakiko.com/kana2/i/eng.htm
[28] www3.dkkaraoke.co.jp

Fig. 6: "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.

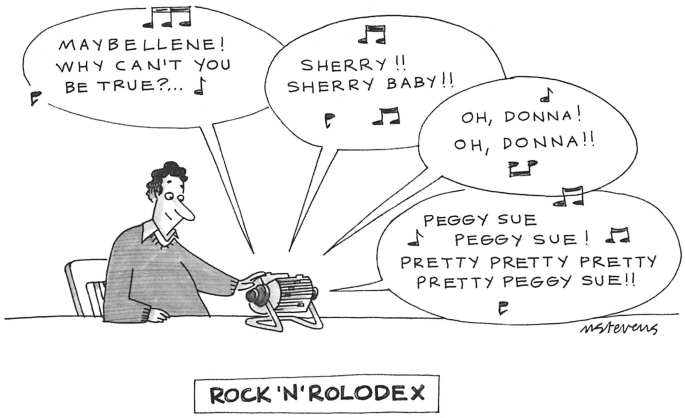


Fig. 7: Old-fashioned *chakumeros*. (©2003 The New Yorker Collection from cartoonbank.com. All rights reserved.)

"FPS" and "MMORPG"

Classic role-playing games and enviroments grew out of "Dungeons and Dragons," extended to computers as MUDs (**m**ulti**u**ser **d**ungeons and **d**omains) and MUSEs (**m**ulti**u**ser **s**imulated **e**nvironments). Recent, 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) [24]— fantasy games like Sony's "EverQuest,"[29] Cyro's "Mankind,"[30] Origin Systems' "Ultima Online,"[31] and Microsoft's "Asheron's Call"[32]— 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 chat-spaces and conferences spawned by such coteries is needed, like that suggested by Table 3.

Wearable Audio Interfaces

The idea of nomadic computing[35] has gone beyond hand-helds and palmtops to include wearable[36] [25] and intimate systems, including eartop [26] [27], wrist-top (like that shown in Figure 8), and collar-top form factors.

Casio, Panasonic, Toshiba, Sony, and others are releasing miniature MP3 players, worn on the wrist, built into headphones, and incorporated into mobile phones. "Dick Tracy"-style wrist watches are already on the market. Telecom equipment maker

[29] everquest.station.sony.com
[30] www.mankind.com
[31] www.uo.com
[32] www.microsoft.com/games/zone/asheronscall
[34] www.elektex.com
[35] www.acm.org/sigmobile/mobicom/2002/
[36] www.media.mit.edu/wearables, iswc.gatech.edu

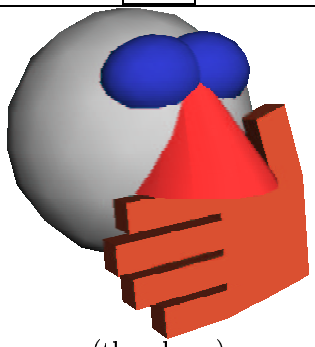
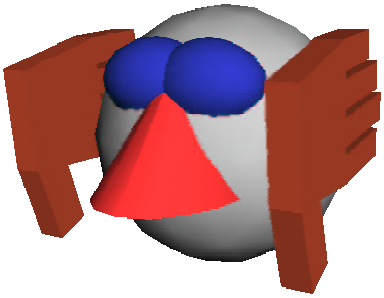
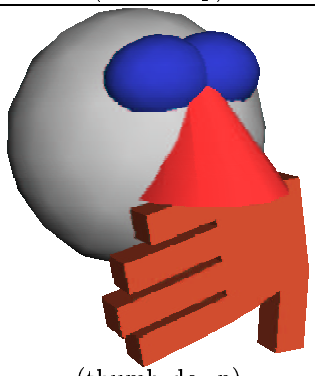
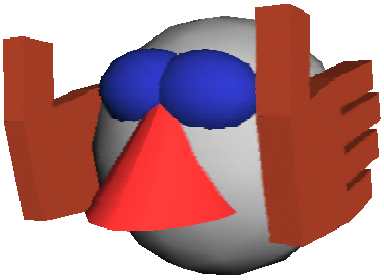
	Source	Sink
Function	radiation	reception
Level	amplification	sensitivity
Direction	OUTput	INput
Instance	speaker	listener
Transducer	loudspeaker	microphone or dummy-head
Organ	mouth	ear
Tool	megaphone	ear trumpet
Include	`solo` (`select`) or `cue`	`attend`: `confide` and `harken`
Attenuate	muzzle	muffle
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 ${}^{\mathrm{s}}\mathrm{IN}^{\mathrm{k}}_{\mathrm{put}}$

Motorola and watch maker Swatch have developed a wristwatch equipped with a phone, as is DoCoMo.[37]

Researchers at IBM Japan have developed a Linux watch "WatchPad" [28] [29], with a speaker and a microphone for speech interfaces along with a Bluetooth Personal Area Network interface, and Citizen has licensed it.[38] Nokia is working on Linux-based cell phone. Mitsubishi[39] recently announced the "Eye Vision" wearable display, featuring a LCD positioned in front of and below one eye (like the "Private Eye" display[40]) integrated with a stereo-

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

[38] `www.extremetech.com/article2/0,3973,82714,00.asp`

[39] `www.mitsubishielectric.com`

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

Fig. 8: ElekTex soft cell phone,[34] with touch controls integrated into the fabric. (©ElekSen. Photo by Marcus Rose.)

Fig. 9: Dieceland Technology disposable mobile phone.[43] 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.

phonic headset. The Vocollect[41] "Talkman" wearable computer uses speech recognition and synthesis to allow hands-free inventory manipulation.

Charmed[44] makes a wearable computer and an infrared badge, ViA[45] makes general-purpose wearable computers especially for hands-free applications, and Hitachi just released the "Wearable Internet Appliance (WIA-100NB)" [30]. Starlab, a Belgian research group, is working to develop intelligent clothing called "i-Wear"[46] 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,[47] 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. Sensatex[48] has developed "SmarTextile" technology to weave sensing, monitoring, and information processing devices into fabric. Scientists at the University of Stuttgart have developed a synthetic fiber that can generate electricity from solar power,[49] potentially important in wearable computers and portable devices requiring power supplies. Researchers at the University of Southern California and Virginia Tech have developed e·textiles, fabric woven with conductive wires and a cluster of seven button-size microphones that can be used to localize the faint sound of remote objects like approaching

[41] www.vocollect.com
[43] www.dtcproducts.com
[44] www.charmed.com
[45] www.flexipc.com
[46] www.starlab.org/bits/intell_clothing/project.html
[47] www.levis-icd.com
[48] www.sensatex.com
[49] www.newscientist.com/news/news.jsp?id=ns9999618

(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.

Fig. 10: 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 among 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. Occasional malfunctions cuase the transporter to act as a replicator, cloning the surprised teleportee. In contemporary GUIs, the `cut/paste` convention is used as a transporter, and `copy/paste` is used as a replicator.

vehicles.[50]

What could be more intimate than an article of clothing? How about a digit in an orifice? Figure 13 illustrates utilizing sound transmission through bones.[53] An even more extreme example is "in-body" technology, like the "molar cellular" concept,[54] developed by the Royal College of Art in London: a microvibration module connected to wrieless receiver implanted into one's tooth during dental surgery carrys sound from tooth to ear via bone conduction.

[50] `www.wired.com/news/gizmos/0,1452,55764,00.html`

[52] `www.xybernaut.com`

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

[54] `www.sciencemuseum.org.uk/corporate_commercial/press/ShowPressRelease.asp?Show=134`

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

Fig. 11: 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

Fig. 12: Xybernaut's patented MA4 system[52] 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.

Fig. 13: 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.

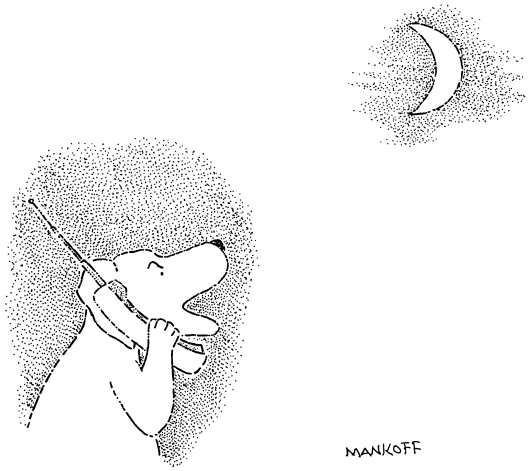


Fig. 14: Telehowl. Canine acoustic sensitivity (up to around 30 kHz, half-again the range of humans) requires extended voice bandwidth. (©2003 The New Yorker Collection from cartoonbank.com. All rights reserved.)

Audio Telemetry and Interspecies Communication

Brookstone[55] markets a "Grill Alert," which telemetric thermomoter notifies a moderately distal human via LCD and audio display when a barbeque's coals are hot.

A similar architecture structures Takara Toys' "Bowlingual"[56] which applies *kansei* engineering (basically differential semantic factor analysis enabling emotional associations) towards inter-species interpretation. An ASR engine is trained by canine utterances, classified by animal behaviorists into six emotional states (frustration, anger, happiness, sadness, entreaty, and 'self-expression'). An 8 cm collar-mounted microphone captures a pet's vocalizations, transmitting them to a hand-held base station, where the samples are vector-quantified and compared to the (16 Mbit memory) training templates. Interpretation of matches are visually displayed as cartoons of dog faces. Even though transliterations of dogs' barks vary widely both orthographically and phonetically from language-to-language and country-to-country, the characterizations are reportedly valid across different breeds and countries.

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.

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[61] is a specially-constructed theater featuring sound dancing around its 169 speak-

[55] brookstone.com
[56] www.takaratoys.co.jp/bowlingual
[57] www.jabra.com/products/mobile.htm
[58] www.elumens.com/products/visiondome.html
[59] www.ambisonics.net
[60] www.imax.com
[61] www.audium.org

Proxemic Context	architecture	Display: audio	Display: visual
intimate	headset, wearable computers	*eartop* headphones, ear buds[57]	*eyetop* HWDs (**h**ead-**w**orn **d**isplays), HMDs (**h**ead-**m**ounted **d**isplays)
individual	chair	nearphones	*laptop* display, *desktop* monitor
interpersonal	couch or bench	transaural speakers (ex: SDP (**s**tereo **di**p**o**le) [31])	HDTV, "fishtank VR," VisionDome,[58] workbench
multipersonal	automobiles, **s**patially **i**mmersive **d**isplays (ex: Cave, Cabin)	surround sound (ex: Ambisonics[59])	projection
social	clubs, theaters	speaker array (ex: VBAP [32], PSFC [33])	large-screen displays (ex: IMAX[60])
public	stadia, concert arenas	**p**ublic **a**ddress	(ex: Jumbotron)

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

ers in an intimate (50 seats) space [35] in San Francisco. "The House on the Rock"[62] 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 [36] [37] [38] allows dynamic (runtime), arbitrary placement and movement of multiple sources in soundscapes, including those spatial dimensions presented by Table 6— as well as control of extra dimensions [39] like apparent extent (including ASW: **a**uditory **s**ource **w**idth), orientation, directivity, and environmental characteristics, for true cyberspatial capabilities.

Information Furniture and Robots

The Inada "Medical Chair H.9"[63] is a synæsthetic seat, synchronizing *shiatsu* massage with music. Our own research group's "Internet Chair" [40, 41] integrates 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.[64])

[62] www.houseontherock.com/the_attraction.htm
[63] www.inada-chair.com
[64] www.mechtec.co.jp

Commercially, Sanyo has unveiled prototypes of robot guard dogs that include built-in speakers and cameras that can link to their masters' mobile phones,[65] and Sony's newest robotic dog, the Aibo ERS-311B, includes Bluetooth communication and speech synthesis.[66] The Fujitsu Maron-1 ambulatory home-use robot[67] can be controlled remotely via an iαppli-enabled mobile phone, to run appliances (via infrared remote control) or act as a watchdog (via visual intruder detection), including the ability to capture images and transmit them to the phone so owners can visually check on home conditions.

Ultrasonics

Ultrasonic-based display systems, which create audible signals through propagation distortion, nonlinear effects on air of ultrasonic signals (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[68] and American Technology's "Hypersonic Sound System,"[69] convert an audio source

[65] www.sanyo.co.jp/koho/hypertext4/0203news-j/0325-1.html
[66] www.sony.jp/CorporateCruise/Press/200205/02-0516/
[67] pr.fujitsu.com/en/news/2002/10/7.html
[68] www.media.mit.edu/~pompei/spotlight
[69] www.atcsd.com/tl_hss.html

Fig. 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.

into an ultrasonic signal, which is then amplified and emitted into the air by ultrasonic ampliers and transducers. The highly directionalized ultrasonic sound beams, steerable through their focus and controllable dispersion (reportedly as low as 3°), including the ability to direct reflections off walls, etc., decompose into three bands as they disperse through air one of which is audible. 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 can be as flexible as analogous light-based visual displays. Such technology is also being considered for military purposes, a so-called "sonic bullet" or "sonic cannon": a recording of a high-pitched baby's wailing, played backwards and aimed as a focused beam, is nonlethal but reportedly painful enough to cause targets (enemy soldiers) to flee.

Cryptosteganaphonics

Steganography is the embedding of one data channel in another. Cryptosteganography is the embedding of hidden data, encrypting information into a carrier, like a photograph ("digital watermarking"). Generics' Intrasonics[70] has developed a technique for piggy-backing an audio channel on a carrier as sounds that would be psychoaoustically masked by the carrier, at least to the human auditory system. These low data-rate signals are spread over the spectrum and disguised into a soundtrack, either prerecorded or in realtime. During *forte* passages, for example, the hidden signals could be stronger and still remain inaudible to humans, but not to decoding software embeddable in terminals. This

[70] www.thecommslab.com/tech_casestudy_intrasonics.asp

Position

Static

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

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

Dynamic

Translation		Along Axis	Perpendicular to Plane
sway	left↔right	x	sagittal (median)
surge	back (aft): retreat ↘ forth (fore): advance	y	frontal
heave	up: ascend ↕ down: descend	z	horizontal

Rotation		About Axis	In Plane
pitch (tumble, flip)	climb/dive	x	sagittal (median)
roll (flop)	left/right	y	frontal
yaw (whirl, twist)	CW/CCW	z	horizontal

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

technology could allow transmission of ads, URLs, and other information through interactive TV, PA systems, etc, but would not be robust against perceptually lossless but entropically lossy compression techniques like MP3.

CONCLUSION

It's ironic that the participle "wired," formerly flattering, is now pejorative, connoting a cumbersome tether. The emergence of mobile internet,[72] global roaming, software-defined radio, wearable computer interfaces, ubiquitous (or "calm" or "ambient") computing,[73] spatially immersive displays, and information furniture offer opportunities for innovative design and advanced applications, both creative and "re-creative" [42]. Enriched connectivity and multipresence will foster use of the `deafen/attend`: `harken` or `confide` sink selection functions [43], like those outlined in Table 7, illustrated by Table 3, and formalized by Figure 19. The feedback between co-potentiating hardware companies (like those making mobile phones) and software content providers— driven by "aftermarket" sales of applications and services, the composition of infrastructure and applications— will continue to blur the distinction between designers & engineers, art & science, and invention & discovery.

[71] `www.irda.org`
[72] `www.mobilehci.org`
[73] `www.personal-ubicomp.com`

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, 40(9):65–73, September 1997.

[2] Norbert A. Streitz, Jane Siegel, Volker Hartkopf, and Shin'ichi Konomi, editors. *Cooperative Buildings: Integrating Information, Organizations, and Architecture (CoBuild'99: Proc. of the Second Int. Workshop)*, Pittsburgh, October 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*, pages 271–278, Aizu-Wakamatsu, Japan, August 1997. IEEE. ISBN 0-8186-8084-9.

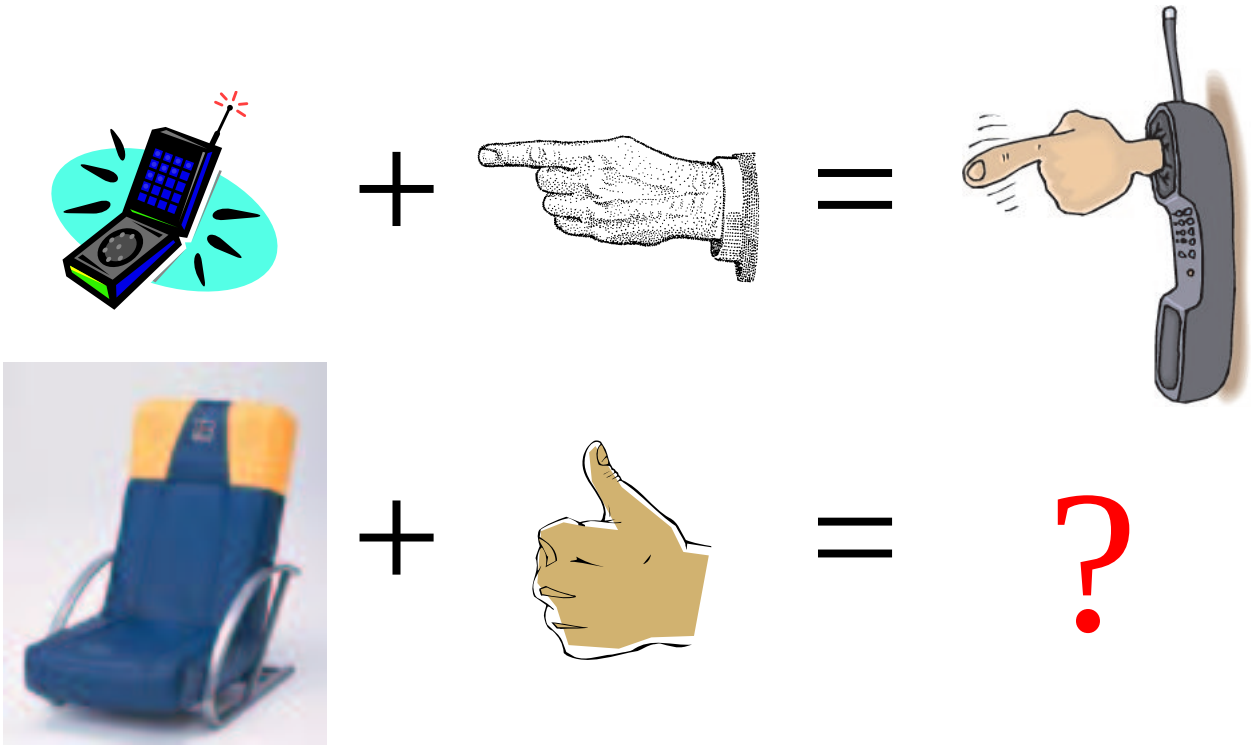


Fig. 17: From a bit in the office to a digit in the orifice. Some gaming stations and LBE cockpits feature LFEs (**l**ow-**f**requency **e**ffects), delivered via BodySonic chairs or similar equipment with subwoofers in the seat for visceral sensation. How might finger-conducted sound, like that illustrated by Figure 13, be posteriorly ("bunaurally") delivered?

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

Table 7: `Exclude`/`Include` taxonomy: disable/enable for one's own and others' representatives.

	Peripheral	Network
Wired	USB IEEE 1394: Firewire	Ethernet
Wireless	IrDA: infrared[71] Bluetooth	WLAN: IEEE 802.11 (Wi-Fi, Airport) cdma2000, W-CDMA, etc.

Table 8: {Wired, Wireless} × {Peripheral, Network}.

The general expression of inclusive selection is

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

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

$$\texttt{active}(\text{source}_x) \;=\; \neg\texttt{mute}(\text{source}_x) \;\wedge\; (\exists\, y\ \texttt{solo}(\text{source}_y) \;\Rightarrow\; \texttt{solo}(\text{source}_x)), \tag{2a}$$

`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

$$\texttt{active}(\text{sink}_x) \;=\; \neg\texttt{deafen}(\text{sink}_x) \;\wedge\; (\exists\, y\ \texttt{attend}(\text{sink}_y) \;\Rightarrow\; \texttt{attend}(\text{sink}_x)). \tag{2b}$$

Fig. 19: 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 $\overbrace{\texttt{solo}\ [\textit{or}\ \texttt{select}]}^{\text{sources}}$ or $\overbrace{\texttt{attend}:\texttt{confide}\ \textit{or}\ \texttt{harken}}^{\text{sinks}}$) when the respective icon is not.

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

[5] Steve Mann. Smart clothing: The wearable computer and wearcam. *Personal Technologies*, 1(1):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*, 32(1):57–64, January 1999.

[7] Woodrow Barfield and Thomas Caudell, editors. *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] John R. Vacca. *I-Mode Crash Course*. McGraw-Hill, 2002. ISBN 0-07-138187-2.

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

[11] James A. Larson. *Voice XML: Introduction to Developing Speech Applications*. Prentice Hall, 2003. ISBN 0-13-009262-2.

[12] Savitha Srinivasan and Eric Brown. Is speech recognition becoming mainstream? *(IEEE) Computer*, pages 38–41, April 2002.

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

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

[15] Yutaka Nagashima and Michael Cohen. Distributed virtual environment interface for a mobile phone. *3D Forum: J. of Three Dimensional Images*, 15(4):102–106, December 2001. ISSN 1342-2189.

[16] Toshifumi Kanno, Michael Cohen, Yutaka Nagashima, and Tomohisa Hoshino. Mobile control of multimodal groupware in a distributed virtual environment. In *Proc. ICAT: Int. Conf.*

- 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 can 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. 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, 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 ${}^{s}_{c}\mathrm{ha}_{i}\mathrm{r}^{e}$ (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.)

Fig. 18: The price of privacy. (©2003 The New Yorker Collection from cartoonbank.com. All rights reserved.)

Fig. 20: Social mute. (©2003 The New Yorker Collection from cartoonbank.com. All rights reserved.)

on **A***rtificial Reality and* **T***ele-Existence*, pages 147–154, Tokyo, December 2001. ISSN 1345-1278; sklab-www.pi.titech.ac.jp/~hase/ICATPHP/upload/39_camera.pdf.

[17] Michael Cohen and William L. Martens. Spatial Media Research at the University of Aizu. *JVRSJ:* **J***. Virtual* **R***eality* **S***ociety of* **J***apan*, 6(2):52–57, September 2001. ISSN 1342-6680; www.u-aizu.ac.jp/~mcohen/welcome/publications/spatial-media.ps.

[18] William L. Martens and Michael Cohen. Virtual Acoustic Research at the University of Aizu. *JVRSJ:* **J***. Virtual* **R***eality* **S***ociety of* **J***apan*, 6(3):261–269, December 2001. ISSN 1342-6680.

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

[20] 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. on* **A***rtificial Reality and* **T***ele-Existence*, pages 161–168, Tokyo, December 2001. sklab-www.pi.titech.ac.jp/~hase/ICATPHP/upload/26_camera.pdf.

[21] Noor Alamshah Bolhassan, William L. Martens, and Michael Cohen. Beyond Flat Panning and Zooming: Dolly-Enhanced SQTVR. In Shietung Peng, Vladimir V. Savchenko, and Shuichi Yukita, editors, *Proc. CW'02: First Int. Symp. on Cyber Worlds: Theory and Practice*, pages 545–552,

Tokyo: Hosei University, November 2002. ISBN 0-7695-1862-1.

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

[23] Benny Bing and Nikil Jayant. A cellphone for all standards. *(*IEEE*) Spectrum*, 39(5):34–39, May 2002.

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

[25] Thad E. Starner. Wearable computers: No longer science fiction. *Pervasive Computing*, 1(1):86–88, January–March 2002.

[26] Michael Cohen, Jens Herder, and William L. Martens. Cyberspatial Audio Technology. **J**. **A***cous.* **S***oc.* **J***ap. (English)*, 20(6):389–395, November 1999. ISSN 0388-2861; `www.u-aizu.ac.jp/~mcohen/welcome/publications/JASJ-reviewE.ps`.

[27] Michael Cohen, Jens Herder, and William L. Martens. Cyberspatial Audio Technology. **J**. **A***cous.* **S***oc.* **J***ap.*, 55(10):730–731, October 1999. (In Japanese). ISSN 0388-2861; `www.u-aizu.ac.jp/~mcohen/welcome/publications/JASJ-reviewJ.ps`.

[28] 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*, 35(1):33–41, January 2002.

[29] 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*, 7(1):78–80, 3 2002. (In Japanese).

[30] Yoshinobu Ohinata. Wearable Computer Release. *JVRSJ:* **J**. **V***irtual* **R***eality* **S***ociety of* **J***apan*, 7(1):80–81, 3 2002. (In Japanese).

[31] 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, March 1997. Preprint 4463 (I6).

[32] Ville Pulkki. Virtual source positioning using vector base amplitude panning. *J.* **A***ud.* **E***ng.* **S***oc.*, 45(6):456–466, June 1997.

[33] 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.* **V***irtual* **R***eality* **S***ociety of* **J***apan*, 3(1):1–12, March 1998. ISSN 1344-011X; `www.u-aizu.ac.jp/~mcohen/welcome/publications/PSFC.ps`.

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

[35] Gareth Loy. About Audium— A Conversation with Stanley Shaff. *Computer Music Journal*, 9(2):41–48, 1985. ISSN 0148-0267.

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

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

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

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

[40] Nobuo Koizumi, Michael Cohen, and Shigeaki Aoki. Japanese patent #3042731: Voice reproduction system, March 2000. www6.ipdl.jpo.go.jp/Tokujitu/PAJlayout.ipdl?N0000=62&N0120=01&N2001=2&N%3001=H05-037994.

[41] Michael Cohen. The Internet Chair. *IJHCI: Int. J. of Human-Computer Interaction*, 15(2), 2003. www.u-aizu.ac.jp/~mcohen/welcome/publications/ic4.ps.

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

[43] Michael Cohen. Exclude and include for audio sources and sinks: Analogs of mute & solo are deafen & attend. *Presence: Teleoperators and Virtual Environments*, 9(1):84–96, February 2000. ISSN 1054-7460; www.u-aizu.ac.jp/~mcohen/welcome/publications/ie1.pdf.