ICAT 2003
December 3-5, Tokyo, Japan

Audio Narrowcasting for Multipresent Avatars on Workstations and Mobile Phones

Owen Noel Newton Fernando, Kazuya Adachi, Uresh Chanaka Duminduwardena, Makoto Kawaguchi, and Michael Cohen

Spatial Media Group; University of Aizu
Aizu-Wakamatsu, Fukushima-ken 965-8580; Japan
{*d8052101,s1080006,m5062101,m5071102,mcohen*}*@u-aizu.ac.jp*

Abstract

Our group is exploring interactive multi- and hypermedia, especially applied to virtual and mixed reality groupware systems. The apparent paradoxes of multipresence, having avatars in multiple places or spaces simultaneously, are resolvable by an "anycast" or "autofocus" feature to project overlaid soundscapes and simulate the precedence effect to consolidate the audio display. Our goal is to develop user interfaces to control source→sink transmission in synchronous groupware (like teleconferences, chatspaces, virtual concerts, etc.). We have developed two interfaces for narrowcasting (selection) functions in **c**ollaborative **v**irtual **e**nvironments (CVEs): for a workstation-style WIMP (**w**indows/**i**con/**m**enu/**p**ointer) and GUI (**g**raphical **u**ser **i**nterface), and for a networked mobile device, a 2.5G-generation mobile phone. The interfaces are integrated with other CVE clients, interoperating with a heterogeneous groupware suite, including stereographic panoramic browsers and spatial audio backends and speaker arrays. The narrowcasting operations comprise an idiom for selective attention, presence, and privacy— an infrastructure for rich conferencing capability.

Keywords: audibility permissions and protocols, chatspace, CSCW (**c**omputer-**s**upported **c**ollaborative **w**ork), graphical binaural directional mixing console, groupware, massively multiplayer online role-playing games (MMORPG), mobile computing, soundscape superposition, spatial sound, teleconferencing, virtual concerts.

1. Introduction

Our group is researching CVEs, **c**ollaborative **v**irtual **e**nvironments: realtime interactive interfaces and applications for teleëxistence and artificial reality groupware [1] [2] Anticipating ubicomp networked appliances and information spaces [3], we are integrating various multimodal (auditory, visual, haptic) I/O devices into a virtual reality groupware suite. Such environments are characterized, in contrast to general hypermedia systems, by the explicit notion of the position (location and orientation) of the perspective presented to respective users, and often such vantage points are modeled by the standpoints and directions of icons in a virtual space. These icons might be more or less symbolic (abstract) or figurative (literal), but are representatives of human users, and are therefore "avatars" (after the Hindu notion of a earthly manifestation of a diety). Avatars reify embodied virtuality, treating abstract presence as a user interface object.

Non-immersive perspectives in virtual environments enable flexible paradigms of perception, especially in the context of frames-of-reference for conferencing and musical audition. Traditional mixing idioms for enabling and disabling various audio sources employ `mute` and `solo` functions which selectively disable or focus on respective channels.[1] Previous research [4] defined sinks as symmetric duals of audio sources in virtual spaces, along with symmetric analogs of source select and mute attributes. Exocentric interfaces which explicitly model not only sources, but also sinks, motivate the generalization of `mute` & `select` (or `cue` or `solo`) to `exclude` and `include`, manifested for sinks as `deafen` & `attend` (`confide` and `harken`), as shown in Figure 1.

Such functions which filter stimuli by explicitly blocking out and/or concentrating on selected entities can be applied not only to other users' sinks for privacy, but also to one's own sinks for selective attendance or presence. These narrowcasting commands control superposition of soundscapes. In the awareness parlance of [5] [6] [7], an aura delimited by a graphical window is like a room, sink attributes affect "focus," and source attributes affect "nimbus."

A unique feature of our system is the ability of a single human pilot to delegate multiple avatars simultaneously, increasing the *quantity* of presence. Multiple sources are useful, for instance, in directing one's remarks to specific groups, decreasing the granularity of audibility control. Multiple sinks are useful in situations in which a common environment implies social inhibitions to rearranging shared sources like musical voices or conferees, as well as individual sessions in which spatial arrangement of sources, like the configuration of a concert orchestra, has mnemonic value.

[1] On many interfaces, "mute" and "solo/select" are abbreviated simply 'M' and 'S' (not to be confused with "master/slave," "mid/side" [as in coincident microphone techniques], etc.).

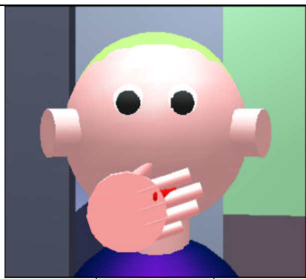
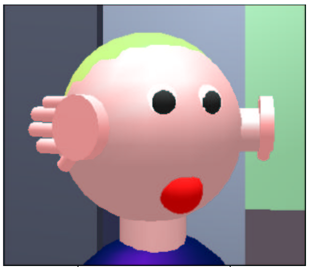
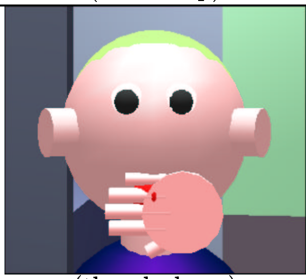
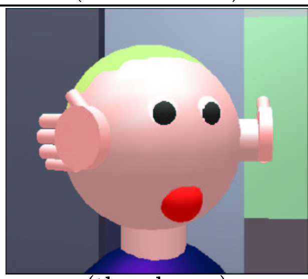
	Source	Sink
Function	radiation/transmission	reception
Level	amplification	sensitivity
Direction	OUTput	INput
Instance	speaker	listener
Transducer	loudspeaker	microphone or dummy-head
Organ	mouth	ear
Tool	megaphone	ear trumpet
Exclude	`mute`	`deafen`
Inhibit in ı·Con	$\overset{-}{\triangle}$	$-\triangle-$
own in Multiplicity *reflexive*	 (thumb up)	 (thumbs back)
other in J3D *transitive*	 (thumb down)	 (thumbs up)
Attenuate	muzzle	muffle
Include	`select` (solo or cue)	`attend`: `confide` and `harken`
Assert in ı·Con	$\overset{+}{\triangle}$	$+\triangle+$
target in Multiplicity *explicit*	 (megaphone)	 (ear trumpets)
others in Multiplicity *implicit*	 (translucent hand)	 (translucent hands)

Table 1: Roles of $^{s}OU^{rce}_{Tput}$ and $^{s}IN^{k}_{put}$: An arbitrary number of avatars can be instantiated at start-up time, and associated with the respective user at runtime. Iconic attributes of narrowcasting functions extend the figurative avatars to illustrate the invoked filter.

The general expression of inclusive selection is

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

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

$$\mathtt{active}(\mathrm{source}_{\mathrm{x}}) \quad = \quad \neg\mathtt{mute}(\mathrm{source}_{\mathrm{x}}) \ \wedge \ (\exists \ \mathrm{y} \ \mathtt{select}(\mathrm{source}_{\mathrm{y}}) \ \Rightarrow \ \mathtt{select}(\mathrm{source}_{\mathrm{x}})), \tag{2a}$$

`mute` explicitly turning off a source, and `select` 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

$$\mathtt{active}(\mathrm{sink}_{\mathrm{x}}) \quad = \quad \neg\mathtt{deafen}(\mathrm{sink}_{\mathrm{x}}) \ \wedge \ (\exists \ \mathrm{y} \ \mathtt{attend}(\mathrm{sink}_{\mathrm{y}}) \ \Rightarrow \ \mathtt{attend}(\mathrm{sink}_{\mathrm{x}})). \tag{2b}$$

Fig. 1: 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 $\overbrace{\mathtt{mute}}^{\text{source}} \overset{\text{or}}{||} \overbrace{\mathtt{deafen}}^{\text{sink}}$), or, b) peers are explicitly included (with $\overbrace{\mathtt{select}\ [\mathtt{solo}]}^{\text{sources}} \overset{\text{or}}{||} \overbrace{\mathtt{attend}: \mathtt{confide}\ \text{or}\ \mathtt{harken}}^{\text{sinks}}$) when the respective icon is not. Narrowcasting attributes are not mutually exclusive, and the dimensions are orthogonal Because a source or a sink is active by default, invoking `exclude` and `include` operations simultaneously on an object results in its being disabled. For instance, a sink might be first `attend`ed, perhaps as a member of some non-singleton subset of a space's sinks, then later `deafen`ed, so that both attributes are simultaneously applied. (As audibility is assumed to be a revocable privilege, such a seemingly conflicted attribute state disables the respective sink, whose attention would be restored upon resetting its `deafen` flag.) Symmetrically, a source might be `select`ed and then `mute`d, akin to making a "short list" but relegated to backup.

2. Implementation

The apparent paradoxes of multipresence, having avatars in multiple places or spaces simultaneously, are resolvable by an "anycast" or "autofocus" feature, simulating the precedence effect [8] projecting overlaid soundscapes to unify a display in an "audio windowing" system, modernizing graphical binaural spatial mixing. Our goal is to develop user interfaces to control source→sink transmission in synchronous groupware (like teleconferences, chatspaces, [M]MORPGs ([massively] multiplayer online role-playing games), virtual concerts, etc.). Narrowcasting operations comprise an idiom for selective attention or presence, an infrastructure for rich conferencing capability. We have developed two compatible and interoperable exocentric interfaces for narrowcasting audio functions in **c**ollaborative **v**irtual **e**nvironments (CVEs), using figurative and iconic avatars, respectively, described in the following subsections:

- **"Multiplicity"** Java3D (J3D) is used to deploy audio windowing systems on workstations— as shown in Figs. 2, 10(b), and 11(b)— featuring 3D perspectives and spatial audio.
- **"ı·Con"** Java 2 Microedition (J2ME) is used to deploy audio windowing systems on *keitais* (via DoCoJa iαppli) and for a networked mobile device, a 2.5-Generation mobile phone, as shown in Figs. 4, 5, 10(a), and 11(a).

All the controls from these interfaces are multicast to all the other (perhaps heterogeneous) clients in a session, synchronizing state, including narrowcasting attributes.

2.1 "Multiplicity": Multipresence through Java3D on a Workstation

We have implemented a narrowcasting interface [9] using Java3D[3] [10] [11] [12]. An arbitrary number of avatars can be instantiated at start-up time, and associated with the respective user at runtime. Iconic attributes of narrowcasting functions, summarized by Table 1, extend the figurative avatars to illustrate the invoked filter.

`Select` and `attend` avatar attributes are denoted by characteristic features. For example, a megaphone appears in front of `select`ed avatars' mouths, and ear trumpets straddle `attend`ed avatars' ears. If any avatars have been `select`ed, non-`select`ed avatars are implicitly `mute`d, and in the dual case that `attend`ed avatars exist, non-`attend`ed avatars are implicitly `deafen`ed. These implicit effects are represented by translucent hands, implicit `mute` represented by a translucent hand clapped over the mouth, and implicit `deafen` represented by translucent hands clapped over the ears, as shown in Figure 11(b). An animated arrow flying from source → sink indicates the autofocus (anycast) determination of the best sink (if any) for each source, strobed in Figure 2, a dynamic representation of the process illustrated by Figure 3.

[2] `www.zentek.com`
[3] `java.sun.com/products/java-media/3D/`

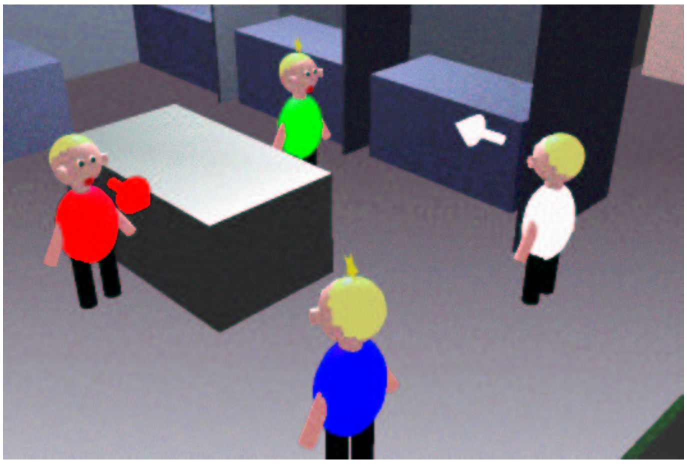


Fig. 2: "Multiplicity" J3D interface for spatial audio and autofocus (anycast) visualization of multipresence CVE: For each speaker, talker, or musical mixel in a teleconference, chatspace, or concert, the application discovers which of the possibly several designated avatar delegates (represented by stars over their heads) is most sensitive (visualized by arrows which fly from source → best sink), directionalizes the sources accordingly, and composites the overlaid soundscapes for display to each user via headphone, nearphones, stereo speakers, or speaker array. "Phantom sources" are used to logically separate listening and viewing positions, allowing the interface a fluid perspective.

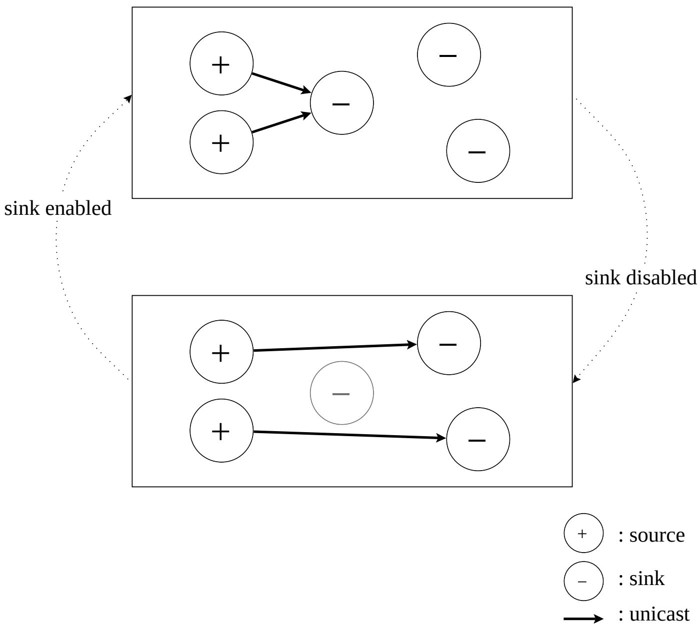


Fig. 3: Anycast source → sink transmissions: if an `attending` sink is `deafen`ed (or peers `confide`d in), remaining sinks adopt orphaned, unicasting, sources (like "discovered check" in chess.)

Fig. 4: NTT DoCoMo i-mode iαppli (iJade emulator[2]) running "ı·Con." (Originally developed by Yutaka Nagashima.)

2.2 "ı·Con": iαppli (DoJa) Mobile Device Dynamic Map

We have designed and implemented a mobile telephone interface [13] [14] for use in CVEs [15]. Programmed with J2ME (Java 2, Micro-Edition[4]) [16] [19] [17] [18] [20] [21], our application runs on an (NTT DoCoMo[5]) iαppli mobile phone, as illustrated by Figure 4. Featuring selectable icons with one rotational and two translational degrees of freedom, the "ı·Con" 2.5D dynamic map interface is used to control position, sensitivity, and audibility of avatars in a groupware session. Its isosceles triangle icons are representatives of symbolic heads in an orthographic projection; its narrowcasting operations are shown in Figure 12 and Table 2.2. The interface is further extended with musical and vibrational cues, to signal mode changes and successful transmission/reception (which feedback is

[4] `java.sun.com/j2me`
[5] `www.nttdocomo.com`

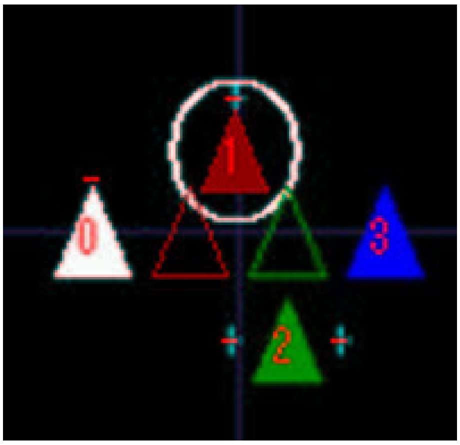
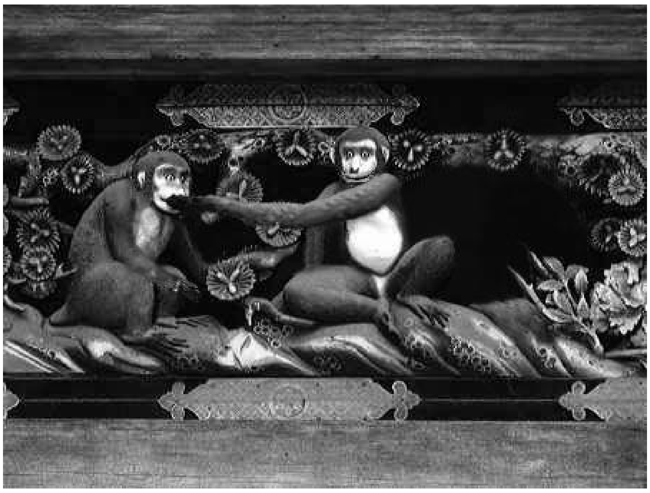


Fig. 5: ‘ı·Con” J2ME dynamic map for (NTT DoCoMo) iαppli mobile phone: Featuring selectable icons in a “2.5D” application, the interface can be used to control position, sensitivity, and audibility of avatars in a groupware session. Quasi-realtime synchronization with a CVE server motivates the use of “ghost icons,” shown as outlines, to distinguish local and session states of avatars. The teleconferencing selection attributes’ graphical displays are triply encoded— by position (before the “mouth” for `mute` and `select`, straddling the “ears” for `deafen` and `attend`), symbol (‘+’ for assert & ‘−’ for inhibit, as shown in Table 1), and color (green for assert & red for inhibit). The attributes are not mutually exclusive, and the encoding dimensions are orthogonal (coloring, for example, the cross bar of a plus sign red even while the vertical bar is green. In this example, #0 is `mute`d; #1 is `mute`d and `solo`ed and selected for rotation; and #2 is simultaneously `attend`ed and `deafen`ed.

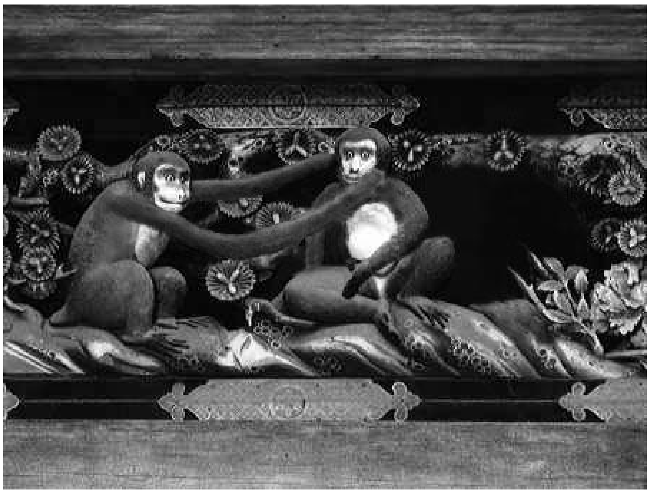
important in wireless communication, as it is much less deterministic than terrestrial systems).

Current user interfaces for mobile phones cannot strictly be characterized as “GUI”s since, in its usual interpretation, the acronym (for “**g**raphical **u**ser **i**nterface”) connotes a “WIMP” idiom (being itself acronymic for “**w**indow, **i**con, **m**enu, **p**ointer”), and the mobile phone lacks a windowing system, menus, and a cursor-style pointer. A better association might be what is sometimes called a “SUI,” for “**s**olid **u**ser **i**nterface,” as modern mobile phones feature unique interface conventions, including vibration, thumb-favored text input, and, on some models, a jog shuttle.

Ongoing complimentary research in our group is exploring techniques for multiwindowing on mobile devices, which capability will require and amplify the multipresence capable selection features described here, multiple avatars associated with a single human user distributed across multiple spaces. Anticipated windowed virtual reality mobile phone interfaces will allow teleport (`cut`/`paste`) and cloning (`copy`/`paste`) operations. For instance, a user could instantiate several avatars in possibly multi-



(a) Disabling source: `mute` by *Iwasenaisaru*.



(b) Disabling sink: `deafen` by *Kikasenaisaru*.

Fig. 6: Distal `exclude`. (Illustrations prepared by Hiroki Sato.)

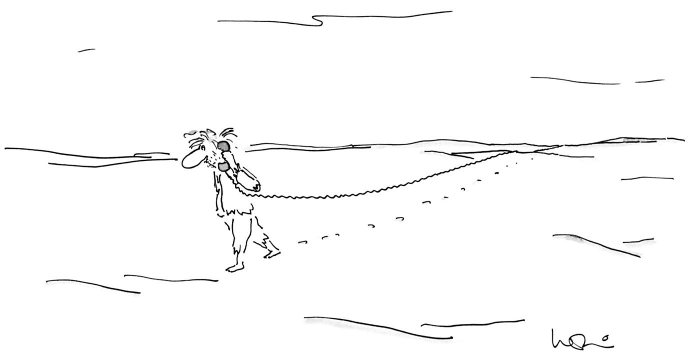
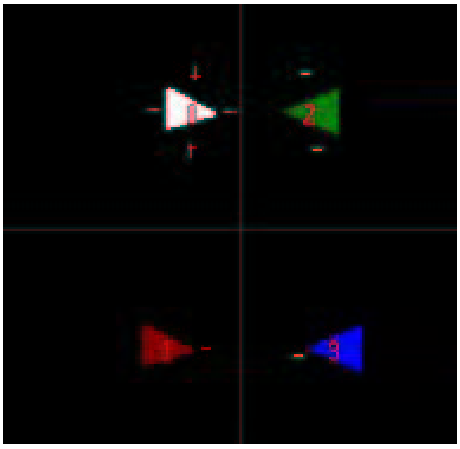
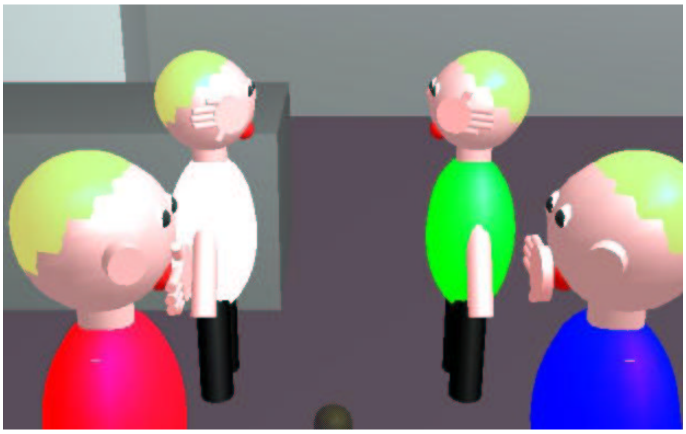


Fig. 7: Non-wireless telephony. (©2003 The New Yorker Collection from `cartoonbank.com`. All rights reserved.)

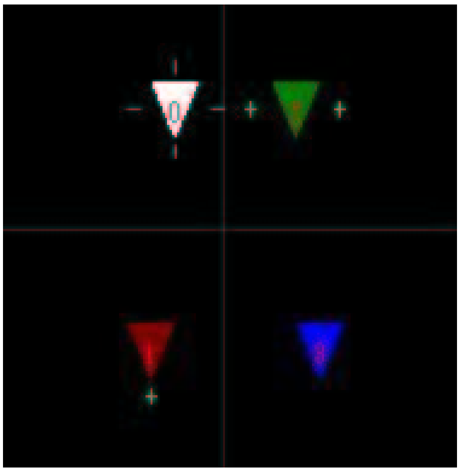


(a) The minus signs straddling the upper right icon indicate that it's `deafen`ed, and the minus sign before the lower right icon indicates that it's `muted`.

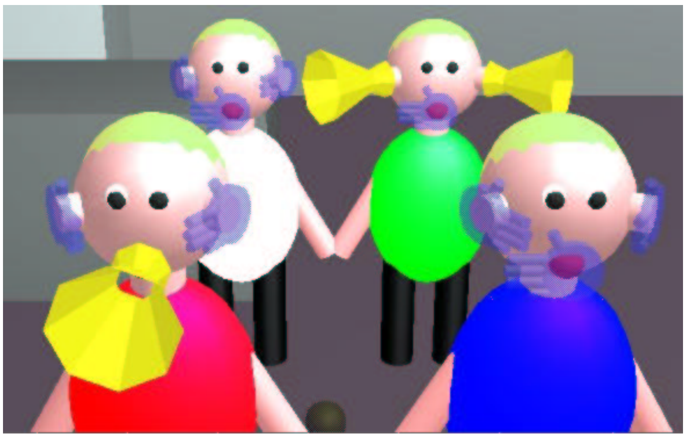


(b) An opaque hand before the mouth of the front right figure indicates that it's `muted`, while hands clapped over the ears of the rear right figure indicate that it's `deafen`ed.

Fig. 10: Synchronized narrowcasting control on mobile device (left) and workstation (right): The two interfaces are coextensive, spanning the same virtual space. In this example, avatar #2 is `deafen`ed, and avatar #3 is `muted`.



(a) Plus signs straddling the upper right icon indicate its `attend`ance, and the plus sign in front of the lower left icon indicates its `select`ion (`solo`).



(b) Megaphones in front of the front left figure indicates its `select`ion (`solo`), so the other figures are implicitly `muted`, indicated by translucent hands in front of their mouths. Ear trumpets straddling the right rear figure indicate its `attend`ance, the other figures implicitly `deafen`ed, as indicated by translucent hands clapped over their ears.

Fig. 11: Avatar #1 is `select`ed, so its complement (comprising all the other avatars) is `mute`d, and avatar #2 avatar is `attend`ed, so its complement is implicitly `deafen`ed.

Fig. 8: The price of privacy. (©2003 The New Yorker Collection from cartoonbank.com. All rights reserved.)

"With your kind permission, I've taken the liberty of putting Marvin on 'mute.'"

Fig. 9: Social mute. (©2003 The New Yorker Collection from cartoonbank.com. All rights reserved.)

```
phrase := selectionToggle ||
  operationToggle || exit
selectionToggle := channelNumber + '#'
operationToggle := attribute + '*'
attribute := (<attend> || <deafen> ||
  <mute> || <select>) || <sink>
exit := '*' + '*'
```

Fig. 12: Postfix grammar for keypad entry: Operands are chosen by toggling avatars tagged with session-unique IDs into/out of the selection set, upon which operations to change position or attributes are subsequently invoked.

attend	ABC 2
deafen	DEF 3
mute	MNO 6
select (solo)	PQRS 7
sink/self	GHI 3

Table 2: Mnemonic initials of conferencing selection operations on the alphanumeric keypad used to toggle selection set attributes.

ple spaces, using the selection functions described here to multiplex and mix their soundscapes.

3. Conclusion

The basic goal of this research is to develop idioms for privacy and selective attention, narrowcasting for groupware applications, whether the interface is via workstation or a nomadic device like a mobile phone. A multipresence scenario using these idioms encourages users to install avatar representatives of themselves in several places and spaces at once. For instance, one might "fork presence" in virtual rooms corresponding to home (chatspace), school (teleconference), and music (virtual concert). Activity or information in a space might cause the user to focus on that particular soundscape, using these narrowcasting functions. Being anywhere is better than being everwhere, since it is selective; multipresence is distilled ubiquity, narrowcasting-enabled audition (for sinks) or address (for sources) of multiple objects of regard. This research can be considered an extension of presence technology [22], and anticipates deployment of such narrowcasting protocols into the internet infrastructure (routers, etc.) itself.

References

1. Noor Alamshah Bolhassan, Michael Cohen, Owen Newton Fernando, Tomoya Kamada, William L. Martens, Hiroki Osaka, and Takuzou Yoshikawa. "Just Look At Yourself!": Stereographic Exocentric Visualization and Emulation of Stereographic Panoramic Dollying. In *Proc. ICAT: **I**nt. **C**onf. on **A**rtificial Reality and **T**ele-Existence*,

pages 146–153, University of Tokyo, December 2002. vrsj.t.u-tokyo.ac.jp/ic-at/02146.pdf.

2. Michael Cohen, Takuya Azumi, Yoshiki Yatsuyanagi, Masahiro Sasaki, Sō Yamaoka, and Osamu Takeichi. Networked Speaker Array Streaming Back to Client: the World's Most Expensive Sound Spatializer? In *Proc. ICAT:* **I***nt.* **C***onf. on* **A***rtificial Reality and* **T***ele-Existence*, pages 162–169, Tokyo, December 2002. vrsj.t.u-tokyo.ac.jp/ic-at/papers/02162.pdf.

3. Tadashi Okoshi, Shirou Wakayama, Yousuke Sugita, Takeshi Iwamoto, Jin Nakazawa, Tomohiro Nagata, Daichi Furusaka, Masayuki Iwai, Akihiko Kusumoto, Noriyuki Harshima, Jun'ichi Yura, Nobuhiko Nishio, Yoshito Tobe, Yasushi Ikeda, and Hideyuki Tokuda. Smart space laboratory project: Toward the next generation computing environment. In *Proc. IEEE Third Int. Workshop on Networked Appliances*, Singapore, March 2001.

4. 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.

5. Chris Greenhalgh and Steven Benford. Massive: A collaborative virtual environment for teleconferencing. *ACM Trans. on Computer-Human Interaction*, 2(3):239–261, September 1995.

6. Steve Benford, John Bowers, Len Fahlén, Chris Greenhalgh, John Mariani, and Tom Rodde. Networked virtual reality and cooperative work. *Presence: Teleoperators and Virtual Environments*, 4(4):364–386, 1995. ISSN 1054-7460.

7. Steve Benford, Chris Greenhalgh, Gail Reynard, Chris Brown, and Boriana Koleva. Understanding and Constructing Shared Spaces with Mixed-Reality Boundaries. *ACM Trans. on Computer-Human Interaction*, 5(3):185–223, September 1998.

8. Jens Blauert. *Spatial Hearing: The Psychophysics of Human Sound Localization.* MIT Press, revised edition, 1997. ISBN 0-262-02413-6.

9. Owen Noel Newton Fernando, Kazuya Adachi, and Michael Cohen. Phantom sources for separation of listening and viewing positions for multipresent avatars in narrowcasting collaborative virtual environments. In *Proc. MNSA: Int. Workshop on* **M***ultimedia* **N***etwork* **S***ystems and* **A***pplications*, Tokyo, March 2004.

10. Henry Sowizral, Kevin Rushforth, and Michael Deering. *The Java 3D API Specification.* Addison-Wesley, second edition, 2000. ISBN 0-201-71041-2.

11. Jon Barrilleaux. *3D User Interfaces with Java 3D.* Manning Publications, 2001. ISBN 1-88477-790-2.

12. Aaron E. Walsh and Doug Gehringer. *Java 3D API jump-start.* Prentice-Hall, 2002. ISBN 0-13-034076-6.

13. Yutaka Nagashima and Michael Cohen. Distributed virtual environment interface for a mobile phone. *3D Forum: J. of Three Dimensional Images*, 15(4):102–106, 12 2001. ISSN 1342-2189.

14. Michael Cohen and Makoto Kawaguchi. Narrowcasting Operations for Mobile Phone CVE Chatspace Avatars. In Eoin Brazil and Barbara Shinn-Cunningham, editors, *Proc. ICAD:* **I***nt.* **C***onf. on* **A***uditory* **D***isplay*, pages 136–139, Boston, July 2003. www.icad.org/websiteV2.0/Conferences/ICAD2003/paper/33Cohen.pdf.

15. Toshifumi Kanno, Michael Cohen, Yutaka Nagashima, and Tomohisa Hoshino. Mobile control of multimodal groupware in a distributed virtual environment. In Susumu Tachi, Michitaka Hirose, Ryohei Nakatsu, and Haruo Takemura, editors, *Proc. ICAT:* **I***nt.* **C***onf. on* **A***rtificial Reality and* **T***ele-Existence*, pages 147–154, Tokyo: University of Tokyo, December 2001. ISSN 1345-1278; sklab-www.pi.titech.ac.jp/~hase/ICATPHP/upload/39_camera.pdf; vrsj.t.u-tokyo.ac.jp/ic-at/papers/01147.pdf.

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

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

18. Kim Topley. *J2ME in a Nutshell.* O'Reilly, 2002. ISBN 0-596-00253-X.

19. Jonathan Knudsen. *Wireless Java: Developing with Java 2, Micro Edition.* Apress, 2001. ISBN 1-893115-50-X.

20. John R. Vacca. *I-Mode Crash Course.* McGraw-Hill, 2002. ISBN 0-07-138187-2.

21. Mikko Kontio. *Mobile Java with J2ME.* IT Press, 2003. ISBN 951-826-554-2.

22. Steven J. Vaughan-Nichols. Presence technology: More than just instant messaging. *(IEEE) Computer*, 36(10):11–13, October 2003. ISSN 0018-9162.