NARROWCASTING OPERATIONS FOR MOBILE PHONE CVE CHATSPACE AVATARS

Michael Cohen and Makoto Kawaguchi

Spatial Media Group, Human Interface Lab.
University of Aizu
Aizu-Wakamatsu 965-8580
Japan
{mcohen, m5071102}@u-aizu.ac.jp

ABSTRACT

We have developed an interface for narrowcasting (selection) functions for a networked mobile device deployed in a **c**ollaborative **v**irtual **e**nvironment (CVE). Featuring a variable number of icons in a "2.5D" application, the interface can be used to control motion, sensitivity, and audibility of avatars in a teleconference or chatspace. The interface is integrated with other CVE clients through a "servent" (**serv**er/cli**ent** hybrid) HTTP↔TCP/IP gateway, and interoperates with a heterogeneous groupware suite to interact with other clients, including stereographic panoramic browsers and spatial audio backends and speaker arrays. Novel features include mnemonic conferencing selection function keypad operations, multiply encoded graphical display of such non-mutually exclusive attributes, and explicit multipresence features.

Keywords and Phrases: audibility permissions and protocols, CSCW (**c**omputer-**s**upported **c**ollaborative **w**ork), graphical (binaural directional) mixing console, groupware, mobile computing, narrowcasting functions, soundscape superposition, solid user interface, spatial sound, teleconferencing.

1. INTRODUCTION

Our group is working on CVEs, **c**ollaborative **v**irtual **e**nvironments: realtime interactive interfaces and applications for teleëxistence and artificial reality groupware [1] [2] [3] [4]. Anticipating ubicomp networked appliances[1] and information spaces [5], 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 a respective user. Often, such a vantage point is modeled by the standpoint and direction of an icon in the virtual space. This icon might be more or less symbolic or figurative (literal), but it is a representative of a human user or users, "avatars" (after the Hindu notion of a earthly manifestation of a diety). Avatars can be said to reify iconic presence.

We have designed and implemented a mobile telephone interface [6] for use in CVEs [7]. Programmed with J2ME[2] (Java 2, micro edition) [8] [9], our dynamic map application runs on an (NTT DoCoMo) iαppli mobile phone, as illustrated by Figure 1. Featuring selectable icons with one rotational and two translational degrees of freedom, the "$\imath\cdot$Con" 2.5D dynamic map interface is used to control avatars in a chatspace. Some 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 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 systems).

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, along with `cue`, selectively disable or focus on respective channels.[3] Previous research [10] defined sinks as symmetric duals of audio sources in virtual spaces, along with symmetric analogs of source solo and mute attributes. Exocentric interfaces which explicitly model not only sources, but also sinks, motivate the generalization of `mute` & `solo` (or `cue`) to `exclude` and `include`, manifested for sinks as `deafen` & `attend` (`confide` and `harken`), as shown in Figure 2.

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. Multiple sinks are useful in groupware, where 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. These narrowcasting commands control superposition of soundscapes. In the awareness parlance of [11] [12] [13], an aura delimited by a graphical window is like a room, sink attributes affect "focus," and source attributes affect "nimbus."

attend	ABC 2
deafen	DEF 3
mute	MNO 6
solo	PQRS 7
si**n**k/self	GHI 3

Table 1: *Mnemonic initials of conferencing selection operations on the alphanumeric keypad used to toggle selection set attributes*

[1] computer.org/pervasive
[2] java.sun.com/j2me
[3] On many interfaces, "mute" and "solo" are abbreviated simply 'M' and 'S' (not to be confused with "master/slave," "mid/side" [as in coincident microphone techniques], "masochism/sadism," or "Micro/soft").

The general expression of inclusive selection is

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

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

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

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

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

Figure 2: *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.*

	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
Include	solo (select) or cue	attend : confide and harken
Assert	$\overset{+}{\triangle}$	$+\triangle+$
Attenuate	muzzle	muffle
Exclude	mute	deafen
Inhibit	$\overset{-}{\triangle}$	$-\triangle-$

Table 2: *Roles of* ${}^{\mathrm{s}}\mathrm{OU}^{\mathrm{rce}}_{\mathrm{Tput}}$ *and* ${}^{\mathrm{s}}\mathrm{IN}^{\mathrm{k}}_{\mathrm{put}}$

2. CONFERENCE NARROWCASTING (SELECTION) FUNCTIONS IMPLEMENTED ON MOBILE DEVICE

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 has come to be called a "SUI," for "**s**olid **u**ser **i**nterface," as a modern mobile phone features unique interface conventions, like vibration, thumb-favored text input, and, on some models, a jog shuttle.

In full-screen GUIs, a new selection resets the selection set, unless it is explicitly extended (typically by chorded shift+ or command+ <click>s, which toggle contiguous or picked objects' membership in the selection set). For our application's mouse- and cursor-less interface, a simple postfix grammar, shown in Figure 4, was developed for keypad entry, used to toggle avatars into and out of the selection set as a prelude for invoking some operation on them (motion: rotation or translation; attribute: set or reset).

The selection functions are invoked by the key corresponding to the attribute initial, as shown in Table 1. The teleconferencing selection attributes' graphical displays are triply encoded— by position (before the "mouth" for mute and solo, straddling the "ears" for deafen and attend), symbol ('+' for assert & '–' for inhibit, as shown in Table 2), 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, as shown in Figure 3). For instance, a sink might be first attended, perhaps as a member of some non-singleton subset of a space's sinks, then later deafened, 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 soloed then muted, akin to making a "short list" but relegated to backup.

Figure 1: *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. (Originally developed by Yutaka Nagashima.)*

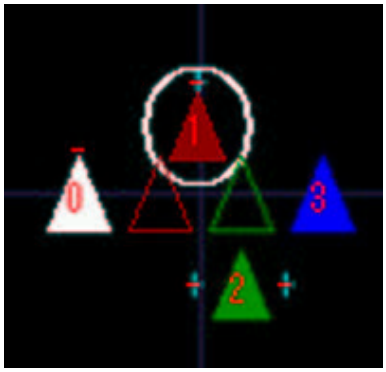

Figure 3: *Screen shot illustrating multiply encoded icons: #0 is* `muted`*; #1 is* `muted` *and* `solo`*ed and selected for rotation; and #2 is* `attend`*ed and* `deafen`*ed.*

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

Figure 4: *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 may be subsequently invoked.*

3. COMPLEMENTARY RESEARCH

We use this interface to control multimodal groupware, including spatial audio applications integrated with panoramic stereographic browsing [14] 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[4] with full CTI (**c**omputer-**t**elephone **i**ntegration) [15], but unfortunately voice communication is currently disabled during such iαppli sessions, so a second phone must be used for an audio duplex channel. Using our mobile networked narrowcasting interface, users will be able to control the spatialized audio (and other realtime media streams) of inevitable multiparty chatspaces, using the cocktail party effect as well as selection to make useful sense of the cacophonies, as imagined by Figure 5.

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 might instantiate several avatars in spaces corresponding to music, intercom at home, and conferences at work or school, using the selection functions described here to multiplex and mix such soundscapes.

[4] `java.sun.com/products/jtapi`

4. 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*, Tokyo, Dec. 2002, pp. 146–153.

[2] Michael Cohen, Takuya Azumi, Yoshiki Yatsuyanagi, Masahiro Sasaki, S¯o 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*, Tokyo, Dec. 2002, pp. 162–169.

[3] 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,

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

Figure 5: *Social mute. (©2003 The New Yorker Collection from* `cartoonbank.com`*. All rights reserved.)*

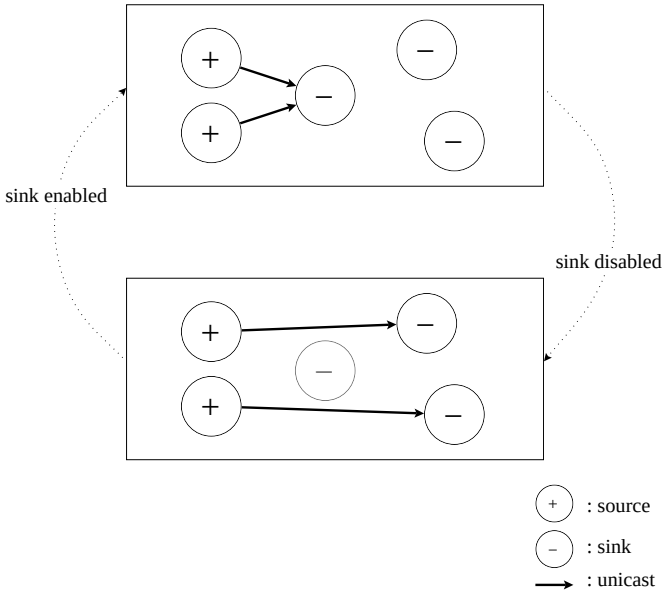

Figure 6: *Unicast source → sink transmissions: if an attending sink is* `deafen`*ed (or peers* `confided` *in), remaining sinks adopt orphaned sources (like "discovered check" in chess.)*

ISSN 1342-6680; `www.u-aizu.ac.jp/~mcohen/welcome/publications/spatial-media.ps`.

[4] 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`.

[5] 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 Workshop on Networked Appliances*, Mar. 2001.

[6] 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.

[7] 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. on* **A***rtificial Reality and* **T***ele-Existence*, Susumu Tachi, Michitaka Hirose, Ryohei Nakatsu, and Haruo Takemura, Eds., Tokyo, 12 2001, pp. 147–154, ISSN 1345-1278; `sklab-www.pi.titech.ac.jp/~hase/ICATPHP/upload/39_camera.pdf`.

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

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

[10] 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`.

[11] Chris Greenhalgh and Steven Benford, "Massive: A collaboratie virtual environment for teleconferencing," *ACM Transactions on Computer-Human Interaction*, vol. 2, no. 3, pp. 239–261, Sept. 1995.

[12] Steve Benford, John Bowers, Len Fahlén, Chris Greenhalgh, John Mariani, and Tom Rodde, "Networked virtual reality and cooperative work," *Presence: Teleoperators and Virtual Environments*, vol. 4, no. 4, pp. 364–386, 1995, ISSN 1054-7460.

[13] Steve Benford, Chris Greenhalgh, Gail Reynard, Chris Brown, and Boriana Koleva, "Understanding and Constructing Shared Spaces with Mixed-Reality Boundaries," *ACM Transactions on Computer-Human Interaction*, vol. 5, no. 3, pp. 185–223, Sept. 1998.

[14] 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*, Tokyo, Dec. 2001, pp. 161–168, `sklab-www.pi.titech.ac.jp/~hase/ICATPHP/upload/26_camera.pdf`.

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