PAPER Special Section on Artificial Reality and Telexistence

Audio Narrowcasting and Privacy for Multipresent Avatars on Workstations and Mobile Phones

Owen Noel Newton FERNANDO†, **Kazuya ADACHI**†, **Uresh DUMINDUWARDENA**†, **Makoto KAWAGUCHI**†, *and* **Michael COHEN**†, ***Nonmembers***

SUMMARY Our group is exploring interactive multi- and hypermedia, especially applied to virtual and mixed reality multimodal groupware systems. We are researching user interfaces to control source→sink transmissions in synchronous groupware (like teleconferences, chatspaces, virtual concerts, etc.). We have developed two interfaces for privacy visualization of narrowcasting (selection) functions in **c**ollaborative **v**irtual **e**nvironments (CVEs): for a workstation WIMP (**w**indows/**i**con/**m**enu/**p**ointer) GUI (**g**raphical **u**ser **i**nterface), and for networked mobile devices, 2.5- and 3rd-generation mobile phones. The interfaces are integrated with other CVE clients, interoperating with a heterogeneous multimodal groupware suite, including stereographic panoramic browsers and spatial audio backends & speaker arrays. The narrowcasting operations comprise an idiom for selective attention, presence, and privacy— an infrastructure for rich conferencing capability.
key words: *audibility permissions, chatspaces, groupware, mobile computing, multipresence, multiuser interface, narrowcasting (selection) functions, soundscape superposition, spatial sound, teleconferencing*

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." Avatars reify embodied virtuality, treating abstract presence as a user interface object.

1.1 Points of View

A classic example of an exocentric display is a map. If someone allows themself an imagined out-of-body experience, flying above the landscape to see the world the way it is portrayed in the map, then the map has become an egocentric display. (This is especially easy to accept if the map is replaced by or superimposed upon an aerial photograph of the same area and an avatar of the subject remains embedded in the original space.) One can slide back and forth along a spectrum between egocentric and exocentric impressions or perspectives.

We use the word "egocentric" (centered on the self) to describe displays logically centered on the avatar or position associated with a given user (or users). Such perspective include both first- and second-person metaphors, sometimes called "tethered" or "yoked" perspectives. We reserve the neologism "endocentric" (centered within) to denote strictly first-person displays, with no display of, for instance, the user's avatar's head. "Exocentric" (centered on the outside) describes a third-person perspective— for instance, the view of a camera attached to the ceiling. These frames-of-reference are illustrated by Fig. 1.

1.2 "FPS" and "MMORPG"

Classic role-playing games and environments grew out of "Dungeons and Dragons," extended to computers as MUDs (**multiu**ser **d**ungeons and **d**omains) and MUSEs (**multiu**ser **s**imulated **e**nvironments). Typically less violent than "**f**irst-**p**erson **s**hooter" 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) [4]— fantasy games like Microsoft's "Asheron's Call,"*, Sony's "EverQuest,"** ArenaNet's "Guild Wars,"*** Cyro's "Mankind,"**** Origin Systems' "Ultima Online,"***** and in Korea, NCsoft's "Lineage: The Blood Pledge"******— players create characters (avatars) to explore persistent universes that exist across sessions, and "massive" means on the order of thousands of users or more per server. Such large-scale social spaces are sometimes called "metaverses." Along with their mobile- and PDA-

†This research was done by the Spatial Media Group at the University of Aizu, Aizu-Wakamatsu, Fukushima-ken, Japan.

*`www.microsoft.com/games/zone/asheronscall`
**`everquest.station.sony.com`
***`www.guildwards.com`
****`www.mankind.com`
*****`www.uo.com`
******`www.lineage.com`

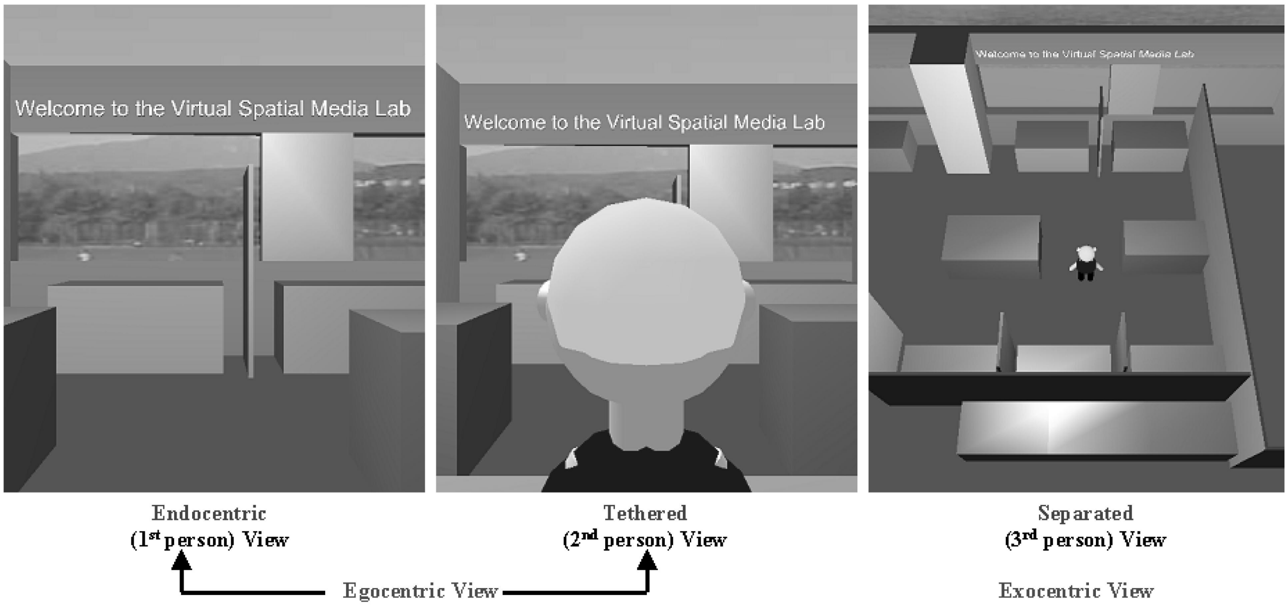


Fig. 1 Points of View

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 `select` (`solo` or `cue`), the relation is

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

`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

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

Fig. 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: an object is inclusively enabled by default unless, a) it explicitly excluded (with $\overbrace{\texttt{mute}}^{\text{source}} \overset{\text{or}}{\|} \overbrace{\texttt{deafen}}^{\text{sink}}$), or, b) peers are explicitly included (with $\overbrace{\texttt{select}\ [\texttt{solo}]}^{\text{sources}} \overset{\text{or}}{\|} \overbrace{\texttt{attend} : \texttt{confide} \text{ or } \texttt{harken}}^{\text{sinks}}$) when the respective avatar 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.

platformed younger siblings, they increasingly feature audio, including both locally-generated sound effects and distally transmitted voice channels, and require broadband network service. Advanced floor control in chat-spaces and conferences spawned by such coteries and "smart mobs" is needed. On a logical level, sources and sinks (receivers, generalizations of listeners and microphones) are resources assigned to users [5]. Shared virtual environments— e.g. chatspaces, online role-playing games, virtual concerts— require generalized control of user-dependent media streams.

Fig. 3 The price of privacy. (© The New Yorker Collection 1996 Sam Gross from `cartoonbank.com`. All rights reserved.)

1.3 Formalization of Narrowcasting and Privacy

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.† Previous research [6] defined sinks as duals of sources in virtual spaces, logical media stream receivers, along with symmetric analogs of source select and mute attributes. 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 elaborated by Fig. 2.

Such functions which filter stimuli by explicitly blocking out and/or concentrating on selected entities [7] can be applied not only to other users' sinks for privacy, but also to one's own sinks for selective attendance or presence. "Privacy" has two interpretations, as suggested by Fig. 3. The first association is that of avoiding "leaks" of confidential information, protecting secrets. But a second interpretation means "freedom from disturbance," in the sense of not being bothered by irrelevance or interruption. Our distributed interface features narrowcasting operations that manage privacy [8] in both senses, by filtering duplex information flow through an articulated conferencing model we call "audio windows" [9]–[11], in analogy to graphical windows.

1.4 Related Research

Benford et al. [12]–[17] derive a model for awareness and interaction in virtual environment. Their "spatial model of interaction" mediates interaction based on the physical properties of space. Thus, the abilities to see and to hear are affected by distance, direction, and possible obstruction. The key awareness abstractions in the spatial model are "aura," "focus," and "nimbus":

Aura The portion of space for which interaction is enabled and allowed.

Focus The more an object is within one's focus, the more aware one is of it.

Nimbus The more a subject is within one's nimbus, the more aware it is of one.

Audio windowing narrowcasting commands control superposition of soundscapes. Using the awareness parlance of Benford et al., an aura delimited by a graphical window is like a room, sink attributes affect focus, and source attributes affect nimbus.

The configuration developed by Benford et al. is represented in Table 1 in an original taxonomy. There are many ways of mapping these scenarios into equivalent configurations supported by our own group's narrowcasting idioms. A direct analogy between nimbus and source "visibility" (audibility, etc.) and between focus and sink attention allows the equivalence illustrated by Table 2.

For example, if a source is `muted`, either by its owner or the other participant, its nimbus excludes the other avatar. For instance, one might hold their hand over the mouthpiece (microphone) or push a "Hold" button (like that in chatspaces like Yahoo! Messenger††) of a phone handset to block the transmission, or use a

†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.).

††`messenger.yahoo.com`

Table 1 Modes of mutual awareness (after Benford et al.). Circles depict the nimbus projected by an object, and arrows depict the direction of the subject's focus. Because of symmetry, the relation is essentially reducible to a triangular matrix, with analogous transposition reflected across the main diagonal.

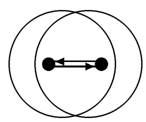
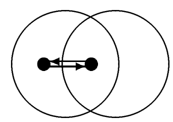
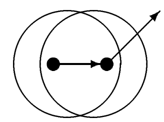
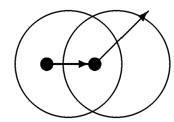
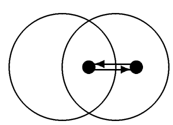
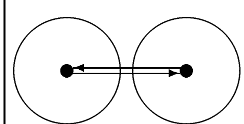
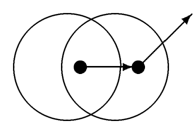
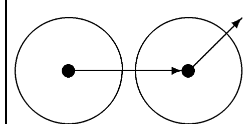
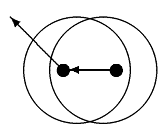
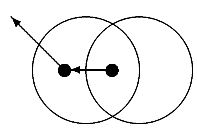
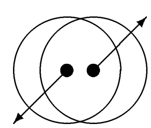
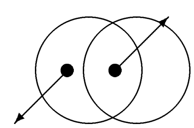
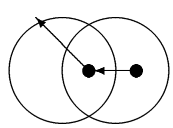
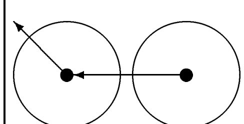
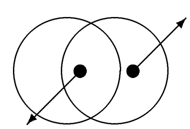
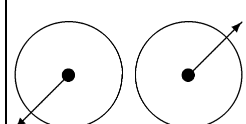
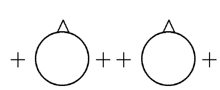
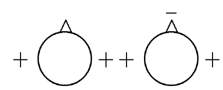
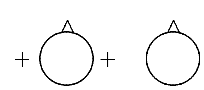
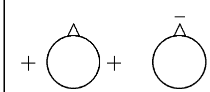
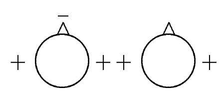
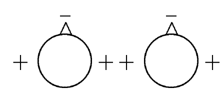
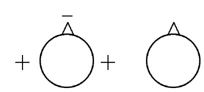
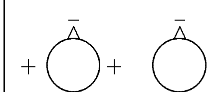
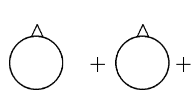
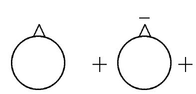
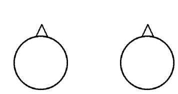
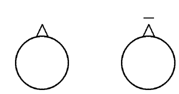
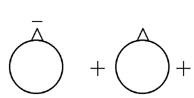
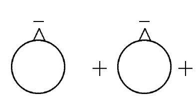
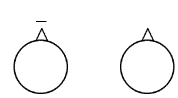
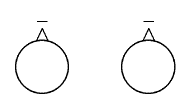
		B focused on A		B not focused on A	
		$A \in$ B's nimbus	$A \notin$ B's nimbus	$A \in$ B's nimbus	$A \notin$ B's nimbus
A focused on B	$B \in$ A's nimbus	 10. fully reciprocal mutual awareness	 9. withdrawal	 7. monitoring	 5. ignoring
A focused on B	$B \notin$ A's nimbus	 (9.)	 8. mutual minimal awareness	 6. eavesdropping	 4. minimal asymmetrical awareness
A not focused on B	$B \in$ A's nimbus	 (7.)	 (6.)	 3. mutual overhearing	 2. overhearing / distraction
A not focused on B	$B \notin$ A's nimbus	 (5.)	 (4.)	 (2.)	 1. no mutual awareness

Table 2 Corresponding narrowcasting modes for mutual awareness (Cohen and Fernando). The +s at the ears, straddling the iconic heads, denote explicitly enabled sinks, and −s denote disabled sources.

		B attended		B not attended	
		B not muted	B muted	B not muted	B muted
A attended	A not muted	 10. fully reciprocal mutual awareness	 9. withdrawal	 7. monitoring	 5. ignoring
	A muted	 (9.)	 8. mutual minimal awareness	 6. eavesdropping	 4. minimal asymmetrical awareness
A not attended	A not muted	 (7.)	 (6.)	 3. mutual overhearing	 2. overhearing / distraction
	A muted	 (5.)	 (4.)	 (2.)	 1. no mutual awareness

"sneeze" button to freeze a video stream.

One might choose to deliberately block one's access to information, if the source were boring or less interesting or important than another source competing for one's attention, or for social reasons. Japan, for instance, has a "*shoji* (sliding rice-paper partition) culture," that offers privacy more symbolic than actual: family members in small apartments afford each other virtual privacy by "choosing not to notice."

It should be stressed that this analogy is only one of several. Since our narrowcasting interface is designed for more than two participants, there are differences between, for instance, `solo`ing a source and `mute`ing its compliment. Tables 1 & 2 show a coarse projection of a much more complicated space:

Self vs. Other Table 2 makes no distinction between narrowcasting attributes invoked by one's self vs. by another. That is, capability can be determined by combined narrowcasting attributes, independent of agent, but socially such distinctions are very important. A can't hear B if B is muted by A or by B himself, but there is a big difference socially, especially in presence of a third actor C, who could hear sources muted transitively (by others) but not reflexively (by themselves).

Bipolar attributes The aura and narrowcasting attributes are interpolated by no continuum; the binary effects are all-or-nothing, but a more natural interface would allow intermediate partial values, fractional distributions of attention. We are developing "`muzzle`" and "`muffle`" narrowcasting commands, which are like partial `mute` for sources and partial `deafen` for sinks, respectively.

Physical position There are no source orientation effects [18]. The (virtual) physical position of the avatars is important, especially for non-omnidirectional projection. In the Benford model, attention via focus is a binary attribute. Even though the focus arrows representing focus show various directions, the classification scheme is "hard," considering only whether or not the sink (the subject) is directly "facing" the source (object of regard), not the degree of such attention.

Multimodal interfaces Multimodal interfaces complicate such considerations even more, since a particular avatar might be "on [some]one's radar" visually, for instance, but not auditorilly.

Also not considered yet are higher-order considerations, like visibility of applied attributes. Our interface is parsimonious: all attributes invoked by any participants are manifest to the affected avatars' displays and interfaces. Such "perfect information" (from game study, in which all actors have access to all information) begs the question: if A `mutes` B, should B always be aware of it? Depending on the conditions, such transparency could be appropriate or not. A parent might insist upon the ability to override a teenager's petulant 'ignore' command: "How dare you mute me‽" Such issues are subtle and sociological, and are the subject of ongoing consideration.

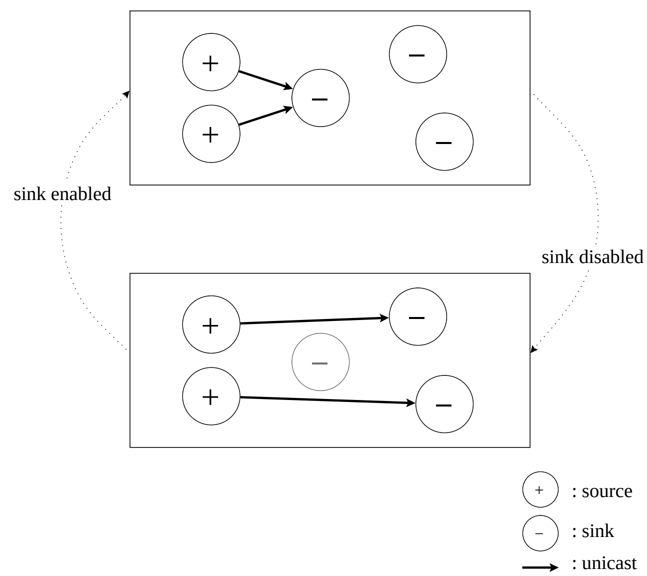


Fig. 4 Autofocused source → sink transmissions: If an intercepting auto-focused sink is `deafen`ed (or peers `attend`ed), remaining sinks adopt orphaned, anycasting, sources.

1.5 Multipresence, Anycast, and Autofocus

Humans are indivisible, so one cannot physically be in two or more places at the same time. A unique feature of our system is the ability of a human pilot to delegate multiple avatars simultaneously, increasing *quantity* of presence [19]; such multipresence enables us to overcome some fundamental constraints of this human condition. Our virtual environment interfaces encourage multipresence [20], by supporting sources and sinks in multiple places simultaneously— allowing, for example, a user to monitor several spaces at once. Multiple sources are useful, for example, 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 configuration of sources, like the arrangement of a concert orchestra, has mnemonic value.

"Anycast" is transmission between a single sender and one of possibly several receivers on a network. The term exists in contradistinction to "multicast," transmission between a single sender and multiple receivers, and "unicast," transmission between a single sender and a single receiver. An anycasting service uses some criteria to choose a "best" or single destination from a set of candidates. We apply the same idea, finding the best sink (the one for which the source is loudest— a function of mutual distance, orientation, directivity, sensitivity, and amplification) for each source in our virtual environment using an "autofocus" technique, illustrated by Fig. 4.

The apparent paradoxes of multipresence, having avatars in multiple places or spaces simultaneously, are resolvable by this autofocus feature, which uses reciprocity, logical exchangability of source and sink, to project overlaid soundscapes and simulate the precedence effect to consolidate the audio display. If the sinks are distributed across separate conference rooms, each source is localized only with respect to the sink in the same space. If multiple sinks share a single conference room, the autofocus algorithm is employed, by anticipating "the rule of the first wavefront" [21]–[23] [24, Part III], the tendency to perceive multiple identical simultaneous sources from different locations as a single fused source. Rather than adding and averaging the contribution of each source [25], [26] to possibly multiple sinks, our system localizes each source only with respect to its respective best sink.

1.6 Pasteboard Operations

We are developing pasteboard operations for workstation- and mobile-based interfaces. The pasteboard operations can be used for teleporting (`cut/paste`) and cloning (`copy/paste`) avatars in collaborative virtual environments. This operations allow teleported and cloned avatars to convey narrowcasting attributes in and across multiple spaces. Dynamic deletion and creation of avatars will enable teleporting and cloning in distributed applications.

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 each user, distributed across multiple spaces. For instance, a user might instantiate several avatars in spaces corresponding to music (virtual concert), intercoms at home, and conferences at school, using selection functions described here to multiplex and mix such soundscapes.

1.7 Proofs-of-Concept

We have designed and implemented an architecture and framework [27]–[31] to support a **c**ollaborative **v**irtual **e**nvironment (CVE) [32], allowing distributed users to share multimodal virtual worlds [33]. Our CVE architecture, sketched in Fig. 5, is based upon a **c**lient/**s**erver (C/S) model, and its main transaction shares the state of virtual objects and users (avatars) by effective multicast via replicated-unicast of position (translation, rotation) parameters and narrowcasting attributes to client peers in a session. The mobile interface is integrated with our CVE through a "servent" (**serv**er/**client** hybrid) HTTP↔TCP/IP gateway.

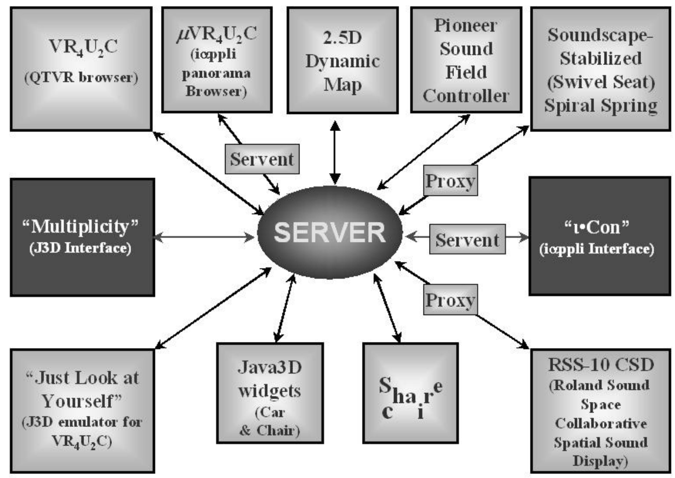


Fig. 5 CVE Architecture: groupware suite. Multiplicity and $\imath$·Con exchange conferencing permisions attributes, including **`mute`** & **`solo`** for sources and **`attend`** & **`deafen`** for sinks.

The client/server architecture enables multimodal communication, platform independence, and easy network connectivity, as components are built with Java (and JMF [**J**ava **M**edia **F**ramework] [34], QuickTime for Java [35], [36], its multimedia API rival Java3D [37]–[42], Java 2 Micro-edition [43]–[49], and Swing [50]).

We have developed two compatible and interoperable interfaces for narrowcasting cyberspatial audio functions in CVEs, using figurative and iconic avatars, respectively described in the following sections:

"Multiplicity" Java3D (J3D) is used to deploy audio windowing systems on workstations— as shown in Figs. 7 & 8, 12(a), and 13(a)— featuring 3D perspectives and spatial audio.

"$\imath$·Con" Java 2 Micro-edition (J2ME) is used to deploy audio windowing systems on networked mobile devices, 2.5- and 3rd-generation mobile phones, as shown in Figs. 9, 10, 12(b), and 13(b).

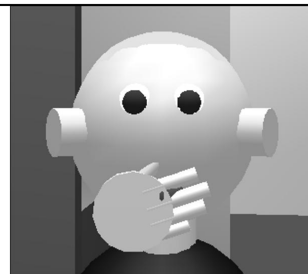
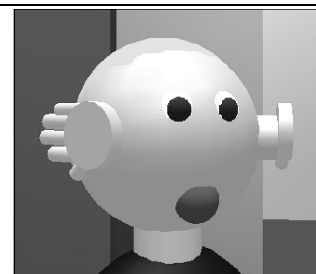
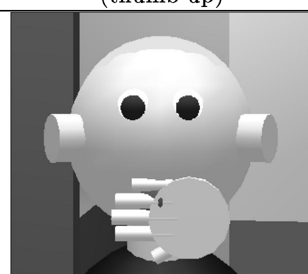
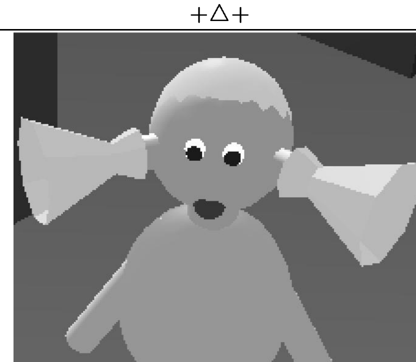
All the controls from these interfaces are multicast to all the other (generally heterogeneous) clients in a session, synchronizing state, including narrowcasting attributes.

2. "Multiplicity": Java3D Workstation-Platformed Multiperspective Interface

2.1 Implementation

We have implemented a workstation-based narrowcast-

Table 3 Roles of ${}^{s}\mathrm{OU}^{\mathrm{rce}}_{\mathrm{Tput}}$ and ${}^{s}\mathrm{IN}^{\mathrm{k}}_{\mathrm{put}}$: An arbitrary number of avatars can be instantiated and associated with users at runtime. Iconic and figurative attributes of narrowcasting functions extend avatars to denote the invoked filters.

	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 $\imath$·Con	$\bar{\triangle}$	$-\triangle-$
Inhibit Self in Multiplicity *reflexive*	 (thumb up)	 (thumbs back)
Inhibit Other in Multiplicity *transitive*	 (thumb down)	 (thumbs up)
Include	`select` (`solo` or `cue`)	`attend`: `confide` and `harken`
Assert in $\imath$·Con	$\overset{+}{\triangle}$	$+\triangle+$
Assert Target in Multiplicity *explicit*	 (megaphone)	 (ear trumpets)
Assertion side-effect in Multiplicity *implicit*	 (translucent hand)	 (translucent hands)

ing interface, "Multiplicity" [51], using Java3D.[†] An arbitrary number of avatars can be instantiated and dynamically associated with respective users at runtime. Attributes of narrowcasting functions, summarized by Table 3, extend the figurative avatars to denote the invoked filters.

Fig. 6 "I copy and paste but nothing happens!"

A simple teleconferencing configuration typically consists of several avatars, representing distributed users, moving around a shared conference space. Each of these icons represents a source, the voice of the associated user, as well as sink, that user's ears. Our system allows each user to designate multiple avatars as "self," effectively increasing anyone's attendance in a conference, as humorously suggested by Fig. 6. Such a feature might be used to pay close attention to multiple sources, even if those sources are not repositionable; just as in ordinary settings, social conventions might inhibit dragging someone else around a shared space.

Each avatar, in general in our environment, can act as both sink and source. Every source in this environment can identify and be associated with a best sink by an autofocus algorithm. Various kinds of "selection" are used in our virtual environment. A singleton selection (multiple avatars not simultaneously selectable) is used to determine the target for locally generated reposition commands and the standpoint and orientation for endo- and ego-centric visual perspectives (View 0 and View 1 buttons in Fig. 7). Further, the `select` and `self` commands are used, along with the other narrowcasting attributes, to resolve privacy operations. When an avatar is selected (as in the leftmost column of the panel in Fig. 8), the best sink from among all `self`-designated avatars for each source does not change, but the other sources conceptually spatialized by other `self`-designated avatars are displayed around the selected avatar according to the displacement from the respective best sinks, via the "phantom sources" feature described in the next section.

[†] `java.sun.com/products/java-media/3D/`

2.2 Phantom Sources

In the absence of multipresence, an egocentric display (from the viewpoint of a particular avatar) naturally and correctly spatializes sound sources. A problem arises when a CVE is extended by multipresence, allowing a single human to designate multiple avatars as associated with the user: exocentric camera positions will not have avatar-centric audio displays. A Java3D listening point is implicitly associated with the virtual camera position (viewpoint). Our technique is to transparently relocate sources to compensate for the selected viewpoint. The multipresence extension, motivated by desire to reduce granularity of control and refined by an autofocus function, works-around the Java3D assumption of coupling of listening and viewing points.

Phantom sources are used to control superposition of soundscapes relative to a selected viewpoint. Relative displacement from sources→sinks can be used to display phantom sources from alternate locations, exocentrically visibly and endocentrically auditorilly. Logical separation of viewpoint and listening point is used to overcome Java3D assumptions and make the interface more fluid. Phantom sources manifest visually, displayed relative to a avatar (whether self or not), and sonically, spatial sound invisibly offset relative to the current viewpoint (virtual camera). An extra feature of the phantom source displacement is the accommodation of a rotatable speaker axis (including sagittal or median plane arrangement) for both panning and fading [52].

Diegetic sound, originally considered in the context of analysis of movies, means sound presented as originating from within a virtual space; non-diegetic sound comes from a source logically outside of the virtual space. The distinction between diegetic and non-diegetic sound depends upon understanding of the conventions of viewing and listening. Phantom sources realize pseudo-diegetic sound, since the audio display is compiled from multipresence sinks' soundscapes in an intuitive but necessarily unnatural way. The sonically displayed phantom source positions depend on viewpoint. Our interface has built-in camera positions, whose settings can be adjusted. Sonic positions of phantom sources are derived relative to the camera in the scene graph, thereby inheriting all the panning, tilting, and dollying that camera might have performed. The selected avatar visually collects all the soundscapes as the union of all sources' best sinks, while the independent viewpoint is the frame-of-reference for the auditorilly superimposed soundscape.

Fig. 7 Java3D narrowcasting interface with phantom sources: Avatars in green (second from the right) and blue (rear) with stars overhead act as sinks for the user and avatars in red (front left) and white (rightmost) are sources. Though the best sink for both sources is the green avatar in this particular configuration, the best sink depends in general on the position of the sinks and sources and the narrowcasting attributes. Each source's arrows fly to its respective best sink. This space's sources are heard before and behind the green avatar. Visual phantom sources are drawn silently displaced relative to the selected avatar (blue in this example), while auditory phantom sources are invisibly displaced relative to the virtual camera. (The red cube and white sphere straddling the blue avatar represent the red and white sources, reflecting the displacement from the green sink.) The sound heard by the avatars is independent of the viewpoint (virtual camera position).

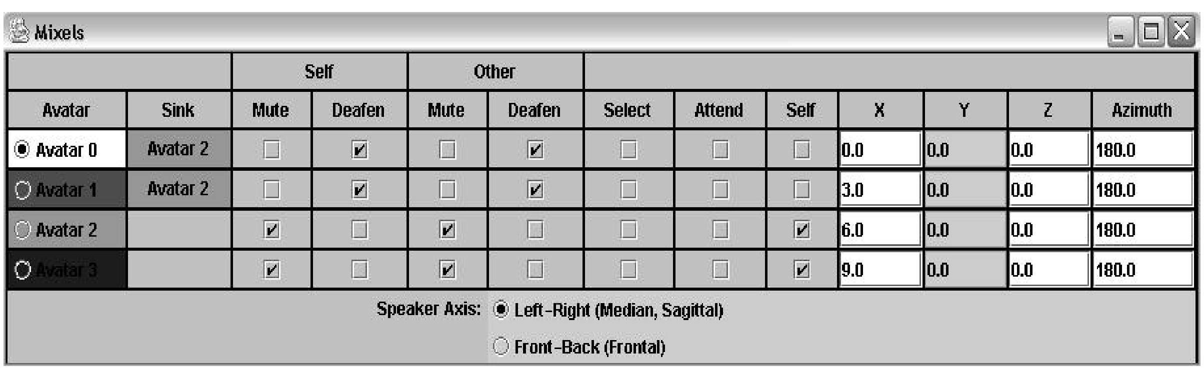


Fig. 8 The Mixels panel can display and control states of each avatar, including selection for motion commands and perspective, autofocused sinks, narrowcasting attributes, designation as self, and position.

3. "ı·Con": (iαppli DoJa) Mobile Device-Platformed Dynamic Map

3.1 Implementation

We have designed and implemented a mobile telephone interface [53], [54] for use in CVEs. Programmed with J2ME†, our application runs on (NTT DoCoMo†† — iαppli) mobile phones, as illustrated by Fig. 9. 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 representations of symbolic heads in an orthographic projection, including narrowcasting attributes shown in Table 3, which operations are shown in Table 4. The interface is further extended with musical and vibrational cues, to signal mode changes and successful transmission/reception (which feedback is especially important in wireless communication).

Fig. 9 NTT DoCoMo i-mode MIDlet (iappliTool for DoJa) running "ı·Con."

Current user interfaces for mobile phones cannot be strictly characterized as "GUI"s since, in its usual interpretation, the acronym connotes a "WIMP" idiom (being itself acronymic for "**w**indow, **i**con, **m**enu, **p**ointer"), but mobile phone interfaces currently lack a proper windowing system. 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.

Table 4 Mnemonic initials of narrowcasting operations on the alphanumeric keypad used to toggle selection set attributes.

`attend`	ABC 2
`deafen`	DEF 3
`mute`	MNO 6
`select (solo)`	PQRS 7
`sink/self`	GHI 4

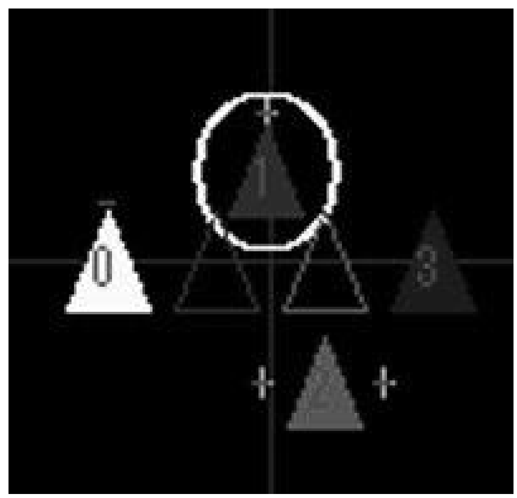


Fig. 10 "ı·Con" J2ME dynamic map for (NTT DoCoMo) iαppli mobile phone, featuring controllable icons in a "2.5D" application. Quasi-realtime synchronization with a CVE server motivates the use of "ghost icons," shown as outlines, to distinguish local and session states of avatars. Workstation commands are synchronized with the mobile interface upon explicit or automatic transmission. (When a fixed-rate tarrif is applied, rather than a per packet charge, our J2ME application will abandon the explicit synchronization requirement, and simply transmit reposition and narrowcasting commands as they arise. Because NTT DoCoMo can only invoke client pull, though, and not server push, explicit read will still be needed.) In this example, #0 is `muted`; #1 is simultaneously `mute`d and `select`ed and also selected for rotation (as indicated by its "halo"); and #2 is simultaneously `attend`ed and `deafen`ed.

In our ı·Con application, narrowcasting 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 3), and color (green for assert & red and yellow for inhibit). The attributes

† `java.sun.com/j2me`
†† `www.nttdocomo.com`, `doja-developer.net`

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 Fig. 10).

3.2 Spatial Audio

Currently mobile applications can not control spatial audio on the same handset, except for very particular special cases, like playback of MLD (compressed MIDI) files, which can be rendered in stereo. As new generations of mobile devices supporting rich audio emerge, spatial audio will be fully deployed, the stereo output of the phone driving a binaural display, through ear- or head-phones, perhaps via a wireless link like a Bluetooth earbud pair.

As mobile phones evolve, spatial audio is becoming ever more important. Dimagic's† "DVX Mobile Surround Technology" features a stereo surround effect that makes it sound as if there are five channels even when only twin speakers are used, employing the "stereo dipole" system [55]. Sonaptic†† has been working with NTT DoCoMo to set standards for 3D audio on mobile phones, and Sonaptic's audio processing algorithms, licensed to Yamaha and Rohm, mimic that 3D encoding, giving the impression that sound is coming from around a listener when in fact it is coming from a single source. Vodafone uses Panasonic's††† "Hyper surround system." Various semiconductor manufacturers are making chips with such features to incorporate into mobile phones. Rohm's†††† sound-source LSI enables 3D positional audio technology for mobile phones, creating the impression of listening to 3D sound using the cellular phone stereo speaker playback. The TI OMAP media processor is used by Panasonic's phones for Vodafone, and the Yamaha MA-7 chip uses the Sonaptic HRTF technology for games as well as the DVX stereo speaker technology for music.

We plan to explore integrated teleconferencing with spatial audio [56], [57] via such a mobile phone with full CTI (**c**omputer-**t**elephone **i**ntegration) [58], but unfortunately voice communication is currently disabled during iαppli sessions. 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 [59], [60] as well as narrowcasting to make useful sense of the cacophonies, as imagined by Fig. 11.

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

Fig. 11 Social mute. (© The New Yorker Collection 2002 William Hamilton from `cartoonbank.com`. All rights reserved.)

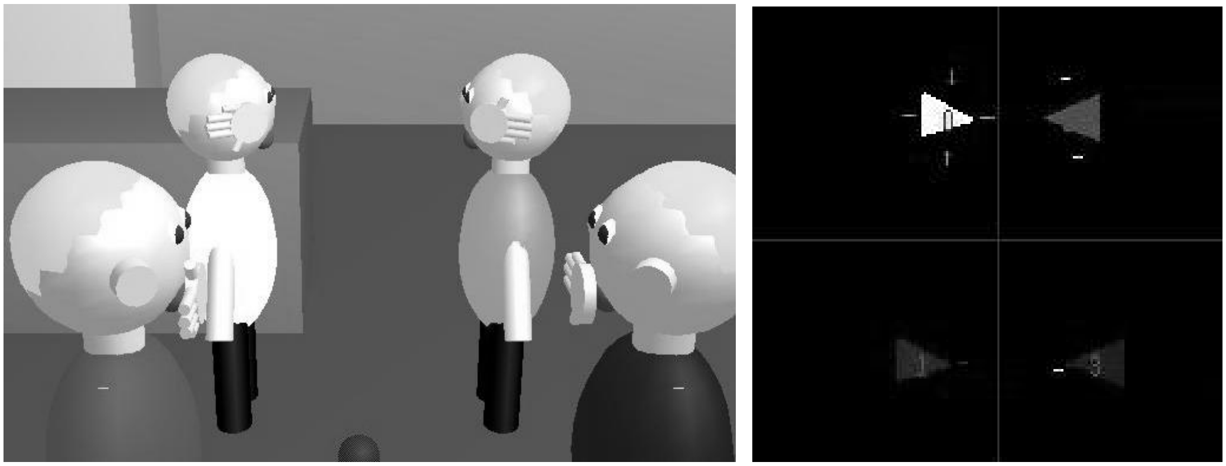
4. Multipresence Scenarios

Our interfaces allow each user to designate multiple sinks, effectively increasing their attendance or "point of presence" in virtual situations. For example, in a chatspace application, a user might choose to designate two avatars as "self"— one near an avatar corresponding to the user's mate; another, perhaps on the other side of the virtual room, near an avatar corresponding to the user's colleague. Each of these avatars enjoys a "local" perspective, a situation awareness encompassing where the respective conversationalists are relative to the (no longer unique) self-associated avatars, as manifested visually and auditorilly.

In a virtual rock concert, for continued example, a listener might want to pay close attention to both the drum and rhythm guitar, avoiding rearranging the instruments around a singleton sink to maintain consistency with other participants. An active listener could fork her presence as `self`-designated avatars, locating one avatar near the drum, and another near the guitar. To focus on only the previously-described chatspace, the listener might `attend` her `self`-designated avatars in it, so her other sinks in the virtual concert are implicitly deafened.

While the low-resolution screen on which our ı·Con mobile interface is deployed supports only a zoomable exocentric display, Multiplicity supports both endocentric and exocentric visual perspectives. The user might want to look out the eyes (endocentric) of an arbitrary avatar, through a view tethered (egocentric) to an avatar, or even through a camera not associated

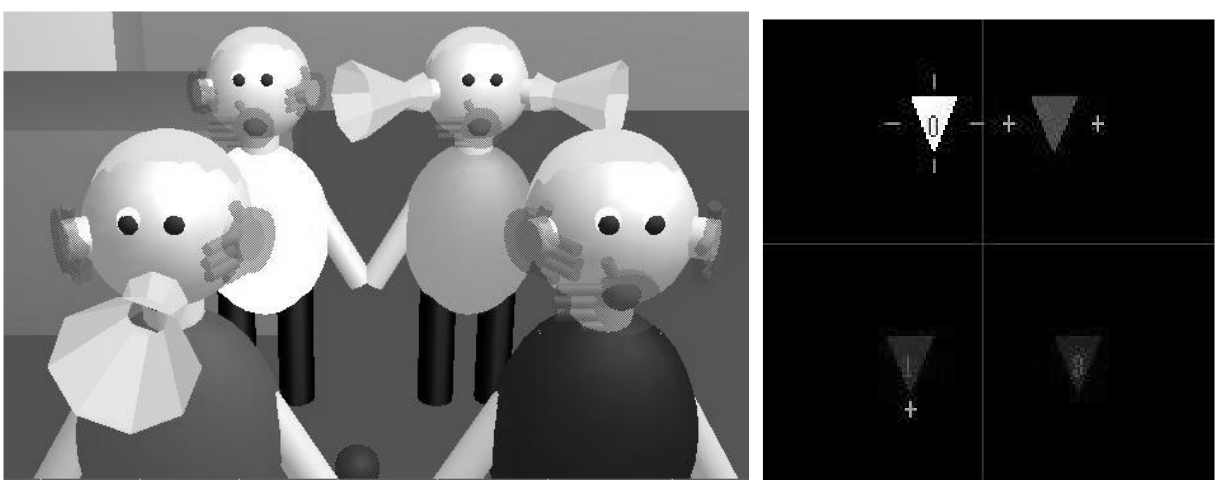
† `www.dimagic.co.jp`
†† `www.sonaptic.com`
††† `panasonic.co.jp`
†††† `www.rohm.co.jp`



(a) Mutedness is indicated by an opaque hand clapped before the mouth, oriented differently depending on whether the source was `muted` by its owner or another (thumb up or down, respectively). In the same way, hands wrap over the ears differently to distinguish deafness for self and other (thumbs down or up). The front left figure is self-`muted`, while the front right figure is other-`muted`. The rear left figure is self-`deafen`ed, while the rear right figure is other-`deafen`ed.

(b) Minus signs (red and yellow) straddling the upper left and right icons indicate that they are `deafen`ed, by self and other, respectively. The minus signs (red and yellow) before the lower icons indicate that they are `muted`, by self and other, respectively.

Fig. 12 Synchronized narrowcasting control on a workstation (left) and mobile device (right): The two interfaces are coextensive, spanning the same virtual space. In this example, avatars #0 and #2 in the rear are `deafen`ed, by self and other, respectively, while avatars #1 and #3 in the front are `muted`, by self and other, respectively.



(a) Megaphones in front of the front left figure indicates its `selection` (`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.

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

Fig. 13 Avatar #1 in the front left is `select`ed, so its complement (comprising all the other avatars) is `muted`, and avatar #2 in the back right is `attend`ed, so its complement is implicitly `deafen`ed.

with any avatar (exocentric), while preserving the cubist (multiperspective) egocentric soundscape. Phantom sources enable this functionality. For instance, one might watch the baby while listening to one's wife, or watch an output while listening to a base.

"Sorry, I can't—I have to be everywhere."

Fig. 14 Divine ubiquity. (© The New Yorker Collection 2003 Bruce Eric Kaplan from `cartoonbank.com`. All rights reserved.)

5. Conclusion

The basic goal of this research is to develop idioms for selective attention, privacy, and presence: narrowcasting for groupware applications, whether the interface is via workstation or a nomadic device like a mobile phone. We have deployed a practical multiplatform implementation of multipresence-enabled narrowcasting functions, including autofocus determination for both workstations and mobile devices. The workstation application features a multiperspective interface, including logical separation of eyes and ears (virtual camera and stereo microphones), exploiting the "phantom source" feature we developed. The mobile interface features equivalent exocentric narrowcasting commands, displayed and controlled in a manner appropriate for the unique form factor of the contemporary mobile phone. The platform-agnostic deployment of the audio narrowcasting idioms— including `deafen`, `mute`, `solo`, and `attend`— encourages the modernization of office- and mobile-based conferencing, leveraging our session integration across coextensive spaces and anticipating multipresence enabled by higher bandwidth and more durable mobile connectivity.

Normally, what one sees is tightly aligned with what one hears, since the eyes and ears are "concentric," locked together as they are in one's head (behind the nose and above the tongue), but users can fork themselves through designation of multiple avatars, compositing phantom sources via the superposition of multiple sinks' soundscapes. 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 [61]. As suggested, by Fig. 14, being anywhere is better than being everywhere, 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 [62], and anticipates deployment of such narrowcasting protocols into session protocols like SIP/SIMPLE† [63], [64] or the internet infrastructure (routers, etc.) itself [65].

References

[1] N.A. Bolhassan, M. Cohen, O.N.N. Fernando, T. Kamada, W.L. Martens, H. Osaka, and T. Yoshikawa, ""Just Look At Yourself!": Stereographic Exocentric Visualization and Emulation of Stereographic Panoramic Dollying," Proc. ICAT: **I**nt. **C**onf. on **A**rtificial Reality and **T**ele-Existence, University of Tokyo, pp.146–153, Dec. 2002. `vrsj.t.u-tokyo.ac.jp/ic-at/ICAT2003/papers/02146.pdf`.

[2] M. Cohen, T. Azumi, Y. Yatsuyanagi, M. Sasaki, S. Yamaoka, and O. Takeichi, "Networked Speaker Array Streaming Back to Client: the World's Most Expensive Sound Spatializer?," Proc. ICAT: **I**nt. **C**onf. on **A**rtificial Reality and **T**ele-Existence, Tokyo, pp.162–169, Dec. 2002. `vrsj.t.u-tokyo.ac.jp/ic-at/ICAT2003/papers/02162.pdf`.

[3] T. Okoshi, S. Wakayama, Y. Sugita, T. Iwamoto, J. Nakazawa, T. Nagata, D. Furusaka, M. Iwai, A. Kusumoto, N. Harshima, J. Yura, N. Nishio, Y. Tobe, Y. Ikeda, and H. Tokuda, "Smart space laboratory project: Toward the next generation computing environment," IWNA: Proc. IEEE Third **I**nt. **W**kshp. on **N**etworked **A**ppliances, Singapore, March 2001.

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

[5] M. Naef, O. Staadt, and M. Gross, "Spatialized audio rendering for immersive virtual environments," Proc. ACM Symp. on **V**irtual **R**eality **S**oftware and **T**echnology, Hong Kong, pp.65–72, ACM Press, 2002. ISBN 1-58113-530-0.

[6] M. 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`.

[7] O.N.N. Fernando, K. Adachi, U.C. Duminduwardena, M. Kawaguchi, and M. Cohen, "Audio Narrowcasting for Multipresent Avatars on Workstations and Mobile Phones," Proc. ICAT: Thirteenth **I**nt. **C**onf. on **A**rtificial Reality and **T**elexistence, Tokyo, pp.106–113, Dec. 2003. ISSN 1345-1278, `http://vrsj.t.u-tokyo.ac.jp/ic-at/ICAT2003/php/program.php`.

[8] M.S. Ackerman, "Privacy in pervasive environments: next generation labeling protocols," **P**ersonal **U**biquitous **C**omput., vol.8, no.6, pp.430–439, 2004.

† `www.cs.columbia.edu/sip/drafts_presence.html`

[9] M. Cohen and L.F. Ludwig, "Multidimensional audio window management," IJMMS: the J. of Person-Computer Interaction, vol.34, no.3, pp.319–336, March 1991. Special Issue on Computer Supported Cooperative Work and Groupware. ISSN 0020-7373.

[10] M. Cohen and L.F. Ludwig, "Multidimensional audio window management," in Computer Supported Cooperative Work and Groupware, ed. S. Greenberg, ch. 10, pp.193–210, Academic Press, London, 1991. ISBN 0-12-299220-2.

[11] M. Cohen and N. Koizumi, "Virtual gain for audio windows," Presence: Teleoperators and Virtual Environments, vol.7, no.1, pp.53–66, Feb. 1998. ISSN 1054-7460.

[12] S. Benford, J. Bowers, L.E. Fahlén, J.A. Mariani, and T. Rodden, "Supporting Cooperative Work in Virtual Environments," Comput. J., vol.37, no.8, pp.653–668, 1994.

[13] C. Greenhalgh and S. Benford, "Massive: A collaborative virtual environment for teleconferencing," ACM Trans. on Computer-Human Interaction, vol.2, no.3, pp.239–261, Sept. 1995.

[14] S. Benford, J. Bowers, L. Fahlén, C. Greenhalgh, J. Mariani, and T. Rodde, "Networked virtual reality and cooperative work," Presence: Teleoperators and Virtual Environments, vol.4, no.4, pp.364–386, 1995. ISSN 1054-7460.

[15] S. Benford, C. Greenhalgh, and D. Lloyd, "Crowded collaborative virtual environments," CHI'97: Proc. SIGCHI Conf. on Human Factors in Computing systems, Atlanta, pp.59–66, ACM Press, 1997.

[16] C. Greenhalgh and S. Benford, "Boundaries, awareness and interaction in collaborative virtual environments," WET-ICE '97: Proc. 6th Int. **W**kshp. on **E**nabling **T**echnologies: **I**nfrastructure for **C**ollaborative **E**nterprises, pp.193–198, IEEE Computer Society, 1997. ISBN 0-8186-7967-0.

[17] S. Benford, C. Greenhalgh, G. Reynard, C. Brown, and B. Koleva, "Understanding and Constructing Shared Spaces with Mixed-Reality Boundaries," ACM Trans. on Computer-Human Interaction, vol.5, no.3, pp.185–223, Sept. 1998.

[18] M. Garau, M. Slater, D.P. Pertaub, and S. Razzaque, "The responses of people to virtual humans in an immersive virtual environment," Presence: Teleoperators and Virtual Environments, vol.14, no.1, pp.104–116, 2005.

[19] M. Cohen, "Quantity of presence: Beyond person, number, and pronouns," in Cyberworlds, ed. T.L. Kunii and A. Luciani, ch. 19, pp.289–308, Springer-Verlag, Tokyo, 1998. ISBN 4-431-70207-5; `www.u-aizu.ac.jp/~mcohen/welcome/publications/bi1.pdf`.

[20] H. Christein and P. Schulthess, "A General Purpose Model for Presence Awareness," Proc. Forth Int. Conf. on Distributed Communities on the Web, Sydney, 2002.

[21] H. Wallach, E.B. Newman, and M.R. Rosenzweig, "The precedence effect in sound localization," American J. of Psychology, vol.57, pp.315–336, 1949.

[22] H. Haas, "The influence of a single echo on the audibility of speech," J. **A**ud. **E**ng. **S**oc., vol.20, pp.146–159, 1972.

[23] J. Blauert, Spatial Hearing: The Psychophysics of Human Sound Localization, revised ed., MIT Press, 1997. ISBN 0-262-02413-6.

[24] R.H. Gilkey and T.R. Anderson, eds., Binaural and Spatial Hearing in Real and Virtual Environments, Lawrence Erlbaum Associates, Mahway, NJ, 1997. ISBN 0-8058-1654-2.

[25] S. DiPaola and D. Collins, "A 3D Virtual Environment for Social Telepresence," Proc. Western Computer Graphics Symp., Vernon, BC, Canada, March 2002.

[26] N. Tsingos, "Perceptual audio rendering of complex virtual environments," Proc. of SIGGRAPH, San Diego, pp.249 – 258, Aug. 2004.

[27] T. Kanno and M. Cohen, "An architecture for collaborative virtual environments," Proc. HC2001: 4th Int. Conf. on **H**uman and **C**omputer, Aizu-Wakamatsu, Japan, pp.31–38, Sept. 2001.

[28] T. Kanno, M. Cohen, Y. Nagashima, and T. Hoshino, "Mobile control of multimodal groupware in a distributed virtual environment," Proc. ICAT: **I**nt. **C**onf. on **A**rtificial Reality and **T**ele-Existence, ed. S. Tachi, M. Hirose, R. Nakatsu, and H. Takemura, Tokyo: University of Tokyo, pp.147–154, Dec. 2001. ISSN 1345-1278; `sklab-www.pi.titech.ac.jp/~hase/ICATPHP/upload/39_camera.pdf`; `vrsj.t.u-tokyo.ac.jp/ic-at/ICAT2003/papers/01147.pdf`.

[29] T. Mikuriya, M. Shimizu, and M. Cohen, "A collaborative virtual environment featuring multimodal information controlled by a dynamic map," 3D Forum: J. of Three Dimensional Images, vol.15, no.1, pp.133–136, 3 2001. ISSN 1342-2189.

[30] T. Kanno and M. Cohen, "An architecture for collaborative virtual environments," 3D Forum: J. of Three Dimensional Images, vol.16, no.1, pp.166–174, March 2002. ISSN 1342-2189.

[31] O.N.N. Fernando, M. Cohen, N.A. Bolhassan, D. Wanasinghe, and T. Kanno, "Mobile Control in Cyberspace of Image-based & Computer Graphic Scenes and Spatial Audio Using Stereo QTVR and Java3D," 3D Forum: J. of Three Dimensional Images, vol.16, no.4, pp.101–106, 12 2002. ISSN 1342-2189.

[32] E.F. Churchill, D.N. Snowdon, and A.J. Munro, eds., Collaborative Virtual Environments: Digital Places and Spaces for Interaction, Computer Supported Cooperative Work (CSCW), Springer, 2001. ISBN 1-85233-244-1.

[33] M. Cohen and W.L. Martens, "Spatial Media Research at the University of Aizu," JVRSJ: **J**. of the **V**irtual **R**eality **S**ociety of **J**apan, vol.6, no.2, pp.52–57, Sept. 2001. ISSN 1342-6680.

[34] R. Gordon and S. Talley, Essential JMF — Java Media Framework, Prentice Hall, 1999. ISBN 0130801046.

[35] T. Maremaa and W. Stewart, QuickTime for Java: a Programmer's Guide to Building Multimedia Applications with Java, 2nd ed., Morgan Kaufmann, 2005. ISBN 155860927X.

[36] C. Adamson, QuickTime for Java: a Developer's Notebook, O'Reilly, 2005. ISBN 0-596-00822-8.

[37] H. Sowizral, K. Rushforth, and M. Deering, The Java 3D API Specification, second ed., Addison-Wesley, 2000. ISBN 0-201-71041-2.

[38] I. Palmer, Essential Java 3D fast: developing 3D graphics applications, Springer, 2001. ISBN 1-85233-394-4.

[39] J. Barrilleaux, 3D User Interfaces with Java 3D, Manning Publications, 2001. ISBN 1-88477-790-2.

[40] A.E. Walsh and D. Gehringer, Java 3D API jump-start, Prentice-Hall, 2002. ISBN 0-13-034076-6.

[41] A. Ota, Yasasiku wakaru Java 3D, Software Technology, Gijutsu Hyoron Sya, 2003. ISBN 4-7741-1669-6.

[42] T. Hirouchi, Java 3D Graphics, Serendhippu, 2004. ISBN 4-7978-2055-1.

[43] J. Knudsen, Wireless Java: Developing with Java 2, Micro Edition, Apress, 2001. ISBN 1-893115-50-X.

[44] Y. Feng and J. Zhu, Wireless Java Programming with J2ME, Sams, 2001. ISBN 0-67232135-1.

[45] ASCII Editing Group, iMode Java Programming, ASCII, 2001. ISBN 4-7561-3727-X.

[46] Y. Yamazaki, How to make an iAppli, SOFTBANK, 2001. ISBN 4-7973-1573-3.

[47] K. Topley, J2ME in a Nutshell, O'Reilly, 2002. ISBN 0-596-00253-X.

[48] J.R. Vacca, I-Mode Crash Course, McGraw-Hill, 2002. ISBN 0-07-138187-2.

[49] M. Kontio, Mobile Java with J2ME, IT Press, 2003.

ISBN 951-826-554-2.
[50] K. Walrath and M. Campione, The JFC Swing Tutorial, Addison-Wesley, 1999. ISBN 0-201-43321-4.
[51] O.N.N. Fernando, K. Adachi, and M. Cohen, "Phantom sources for separation of listening and viewing positions for multipresent avatars in narrowcasting collaborative virtual environments," Proc. MNSA: Int. Wkshp. on **M**ultimedia **N**etwork **S**ystems and **A**pplications (in conjunction with ICDCS: 24th **I**nt. **C**onf. on **D**istributed **C**omputing **S**ystems), Hachiōji, Tokyo, pp.170–175, March 2004. ISBN 0-7695-2087-1, `ieeexplore.ieee.org/iel5/9027/28651/01284027.pdf`.
[52] O.N.N. Fernando, "Java3D-based Development of a Spatial Audio Multipresence Application Featuring Narrowcasting Functions and Rotatable Loudspeaker Axis (Pairwise-Selected Speakers) for a Collaborative Virtual Environment," Master's thesis, University of Aizu, 2003.
[53] Y. Nagashima and M. Cohen, "Distributed virtual environment interface for a mobile phone," 3D Forum: J. of Three Dimensional Images, vol.15, no.4, pp.102–106, 12 2001. ISSN 1342-2189.
[54] O.N.N. Fernando, M. Cohen, U.C. Dumindawardana, and M. Kawaguchi, "Duplex narrowcasting operations for multipresent groupware avatars on mobile devices," IJWMC: **I**nt. **J**. of **W**ireless and **M**obile **C**omputing (Special Issue on Mobile Multimedia Systems and Applications), vol.1, no.5, 2005. ISSN 1741-1084, 1741-1092.
[55] O. Kirkeby, P.A. Nelson, and H. Hamada, "The stereo dipole — binaural sound reproduction using two closely spaced loudspeakers," AES: **A**udio **E**ngineering **S**ociety Conv., Munich, March 1997. Preprint 4463 (I6).
[56] Y. Yasuda, T. Ohya, D. McGrath, and P. Flanagan, "3-D Audio Communications Services for Future Mobile Networks," AES: **A**udio **E**ngineering **S**ociety Conv. (23rd Int. Conf.), Helsingør, Copenhagen, Denmark, May 2003. 4–2.
[57] Y. Yasuda, T. Ohya, D. McGrath, and P. Flanagan, "Reality Speech/Audio Communications Technologies," NTT Technical Review, vol.1, no.8, pp.95–100, Nov. 2004.
[58] S. Roberts, Essential JTAPI: Java telephony, Prentice Hall, 1999. ISBN 0-13-080360-X.
[59] M. Kobayashi and C. Schmandt, "Dynamic soundscape: mapping time to space for audio browsing," Proc. SIGCHI: Conf. on Human Factors in Computing Systems, Atlanta, pp.194–201, ACM Press, 1997.
[60] P.M. Aoki, M. Romaine, M.H. Szymanski, J.D. Thornton, D. Wilson, and A. Woodruff, "The mad hatter's cocktail party: a social mobile audio space supporting multiple simultaneous conversations," CHI: Proc. of the Conf. on Human Factors in Computing Systems, Ft. Lauderdale, Florida, pp.425–432, ACM Press, 2003.
[61] A.S. Pentland, "Socially aware computation and communication," Computer, vol.38, no.3, pp.33–40, 2005.
[62] S.J. Vaughan-Nichols, "Presence technology: More than just instant messaging," (IEEE) Computer, vol.36, no.10, pp.11–13, Oct. 2003. ISSN 0018-9162.
[63] L. Dang, C. Jennings, and D. Kelly, Practical VoIP Using VOCAL, O'Reilly Media, Inc., Sebastopol, 2002. ISBN 0596000782.
[64] A.B. Johnston, SIP: Understanding the Session Initiation Protocol, Artech House, London, 2004. ISBN 1580531687.
[65] M. Radenkovic, C. Greenhalgh, and S. Benford, "Deployment issues for multi-user audio support in CVEs," Proc. ACM Symp. on **V**irtual **R**eality **S**oftware and **T**echnology, Hong Kong, pp.179–185, ACM Press, 2002.

Owen Noel Newton Fernando A Systems Analyst from '98–'01 at the People's Bank (Colombo, Sri Lanka), Fernando got his M.Sc. in 2003 from the University of Aizu, where he is currently a Ph.D. candidate, with research interests in mobile computing, entertainment computing, mixed and virtual reality, narrowcasting commands with spatial sound, and conference selection functions for multipresence. He is a member of the ACM and IEEE.

Kazuya Adachi Currently a M.Sc. candidate at the University of Aizu (Aizu-Wakamatsu, Japan), Adachi has research interests in driving simulators, non-omnidirectional sound sources for workstations and mobile phones, and entertainment computing.

Uresh Duminduwardena From 2000–'02, Dumiduwardena was a Software Engineer at Millenium IT (Colombo, Sri Lanka). He got his M.Sc. from the University of Aizu (Aizu-Wakamatsu, Japan) in 2005, with research interests in mobile computing and haptic devices. He is currently a Senior Software Engineer at Dialog Telekom (Colombo, Sri Lanka).

Makato Kawaguchi Currently at the Japan Research Institute, Kawaguchi received his M.Sc. from the University of Aizu (Aizu-Wakamatsu, Japan) in 2005, with research interests in mobile computing and narrowcasting functions for mobile phones.

Michael Cohen Besides mobile computing, Michael Cohen's research interests include telecommunication semiotics, spatial audio, stereography, mixed and virtual reality, and telepresence. Cohen received a Ph.D. in EECS from Northwestern University (Evanston, Illinois) in 1991. He is a Prof. at the University of Aizu (Aizu-Wakamatsu, Japan). Cohen is the author or coauthor of two patents, four book chapters, and over forty journal articles. He is a member of the ACM, IEEE Computer Society, 3D-Forum, TUG (TeX Users Group), and VRSJ (Virtual Reality Society of Japan).