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 collaborative virtual environments (CVEs): for a workstation-style WIMP (windows/-frame/menu/pointer) and GUI (graphical user interface), and for a networked mobile device, a 2.5G-generation mobile phone. The interfaces are integrated with other CVE clients, interfacing 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 (computer-supported collaborative work), 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, collaborative virtual environments: realtime interactive interfaces and applications for telepresence 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 deity). 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. 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 parable 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.

1On 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.).
<table>
<thead>
<tr>
<th>Source</th>
<th>Sink</th>
</tr>
</thead>
<tbody>
<tr>
<td>Function radiation/transmission</td>
<td>reception</td>
</tr>
<tr>
<td>Level amplification</td>
<td>sensitivity</td>
</tr>
<tr>
<td>Direction output</td>
<td>Input</td>
</tr>
<tr>
<td>Instance speaker</td>
<td>listener</td>
</tr>
<tr>
<td>Transducer loudspeaker</td>
<td>microphone or dummy-head</td>
</tr>
<tr>
<td>Organ mouth</td>
<td>ear</td>
</tr>
<tr>
<td>Tool megaphone</td>
<td>ear trumpet</td>
</tr>
<tr>
<td>Exclude mute</td>
<td>deafen</td>
</tr>
<tr>
<td>Inhibit in (\uparrow\Con)</td>
<td>(\uparrow\Delta)</td>
</tr>
<tr>
<td>own in Multiplicity reflexive</td>
<td>(thumb up)</td>
</tr>
<tr>
<td>other in J3D transitive</td>
<td>(thumbs back)</td>
</tr>
<tr>
<td>Attenuate muzzle</td>
<td>muzzle</td>
</tr>
<tr>
<td>Include select (solo or cue)</td>
<td>attend confide and harken</td>
</tr>
<tr>
<td>Assert in (\uparrow\Con)</td>
<td>(\uparrow\Delta)</td>
</tr>
<tr>
<td>target in Multiplicity explicit</td>
<td>(megaphone)</td>
</tr>
<tr>
<td>others in Multiplicity implicit</td>
<td>(ear trumpets)</td>
</tr>
<tr>
<td>(translucent hand)</td>
<td>(translucent hands)</td>
</tr>
</tbody>
</table>

Table 1: Roles of \(\text{OU}_{\text{Tput}}\) and \(\text{IN}_{\text{put}}\): 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
\[
\text{active}(x) = \neg \text{exclude}(x) \land (\exists y \text{ include}(y) \Rightarrow \text{include}(x)).
\] 
(1)

So, for \text{mute} and \text{select} (solo), the relation is
\[
\text{active}(\text{source}_a) = \neg \text{mute}(\text{source}_a) \land (\exists y \text{ select}(\text{source}_y) \Rightarrow \text{select}(\text{source}_y)),
\] 
(2a)

\text{mute} explicitly turning off a source, and \text{select} disabling the collocated (same room/window) complement of the selection (in the spirit of “anything not mandatory is forbidden”). For \text{deafen} and \text{attend}, the relation is
\[
\text{active}(\text{sink}_a) = \neg \text{deafen}(\text{sink}_a) \land (\exists y \text{ attend}(\text{sink}_y) \Rightarrow \text{attend}(\text{sink}_y)).
\] 
(2b)

Fig. 1: Formalization of narrowcasting and selection functions in predicate calculus notation, where ‘\neg’ means “not,” ‘\land’ means conjunction (logical “and”), ‘\exists’ means “there exists,” and ‘\Rightarrow’ 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 explicitly excluded (with \text{mute} || \text{deafen}), or b) peers are explicitly included (with \text{select} [solo] || \text{attend: confide or harken}) 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 \text{exclude} and \text{include} operations simultaneously on an object results in its being disabled. For instance, a sink might be first \text{attended}, perhaps as a member of some non-singleton subset of a space’s sinks, then later \text{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 \text{deafen} flag.) Symmetrically, a source might be \text{selected} and then \text{muted}, 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, \text{MORPGE}s [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 collaborative virtual environments (CVEs), using figurative and iconic avatars, respectively, described in the following subections:

“\text{Multiplexity}” Java3D (3D) 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.

“\text{Con}” Java2 Microedition (J2ME) is used to deploy audio windowing systems on \text{A3TA}s (via DoCada ioppli) 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 “\text{Multiplexity}”: Multipresence through Java3D on a Workstation

We have implemented a narrowcasting interface [9] using Java3D\footnote{www.zentek.com} [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 \text{attend} avatar attributes are denoted by characteristic features. For example, a megaphone appears in front of \text{selected} avatars’ mouths, and ear trumpets straddle \text{attended} avatars’ ears. If any avatars have been selected, non-selected avatars are implicitly \text{muted}, and in the dual case that \text{attended} avatars exist, non-\text{attended} avatars are implicitly \text{deafened}. These implicit effects are represented by translucent hands, implicit \text{mute} represented by a translucent hand clapped over the mouth, and implicit \text{deafen} represented by translucent hands clapped over the ears, as shown in Figure 11(b). An animated arrow flying from source \rightarrow 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.

\text{J2ME}\footnote{java.sun.com/products/java-media/3D/}
Fig. 2: “Multiplicity” 3D interface for spatial audio and auto-focus (anycast) visualization of multipresence CVE. For each speaker, talker, or musical nise in a teleconference, chat-space, 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), directionizes the sources accordingly, and composites the overlaid soundscapes for display to each user via headphone, earphones, 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 deafened (or peers confided in), remaining sinks adopt orphaned, unicasting, sources (like “discovered check” in chess.)

Fig. 4: NTT DoCoMo i-mode iappli (iJade emulator) running “i-Con.” (Originally developed by Yutaka Nagashima.)

2.2 “i-Con”: iappli (DoJa) Mobile Device Dynamic Map

We have designed and implemented a mobile telephone interface [13] [14] for use in CVEs [15]. Programmed with i2ME (Java 2, Micro-Edition) [16] [19] [17] [18] [20] [21], our application runs on an (NTT DoCoMo) iappli mobile phone, as illustrated by Figure 4. Featuring selectable icons with one rotational and two translational degrees of freedom, the “i-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
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 "graphical user interface") connotes a "WIMP" idiom (being itself acronymic for "window, icon, menu, pointer"), and the mobile phone lacks a windowing system, menus, and a cursor-style pointer. A better association might be what is sometimes called a "sub" for "solid user interface," 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-
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 deafened, and avatar #3 is muted.

Fig. 11: Avatar #1 is selected, so its complement (comprising all the other avatars) is muted, and avatar #2 avatar is attended, so its complement is implicitly deafened.
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.

<table>
<thead>
<tr>
<th>Operation</th>
<th>Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>attend</td>
<td>A</td>
</tr>
<tr>
<td>deafen</td>
<td>B</td>
</tr>
<tr>
<td>mute</td>
<td>C</td>
</tr>
<tr>
<td>select (solo)</td>
<td>D</td>
</tr>
<tr>
<td>sink/self</td>
<td>E</td>
</tr>
</tbody>
</table>

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

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 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 [22], and anticipates deployment of such narrowcasting protocols into the internet infrastructure (routers, etc.) itself.

References


