Michael Cohen

mcohen@u-aizu.ac.jp

http://www.u-aizu.ac.jp/
∼mcohen

Spatial Media Group

Human Interface Lab.

University of Aizu 965-8580

Japan

Exclude and Include for Audio Sources and Sinks:
Analogs of *Mute & Solo* Are *Deafen & Attend*

Abstract

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, that, along with *cue,* selectively disable or focus on respective channels. Exocentric interfaces which explicitly model not only sources but also sinks, motivate the generalization of *mute* and *solo* (or *cue*) to `exclude` and `include,` manifested for sinks as *deafen* and *attend* (*confide* and *harken*). Such functions, which narrow 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. A taxonomy of modal narrowcasting functions is proposed, and an audibility protocol is described, comprising `revoke, renounce, grant,` and `claim` methods, invocable by these narrowcasting commands to control superposition of soundscapes.

1   Introduction

An exocentric model in which a user is represented by an icon (avatar, synthespian, vactor, and so on) in the context of a virtual space (as suggested by Table 1) is useful in spatial sound systems; virtual environments with audio can be thought of as graphical mixing consoles. As outlined by Table 2, since the word *speaker* is ambiguously overloaded, meaning both *loudspeaker* and *talker,* this paper uses *source* to mean both, a logical sound emitter. Similarly and symmetrically, *sink* is used to describe a virtual listener, a logical sound receiver. Icons embodying sources and sinks may wander around virtual spaces, like minglers at a cocktail party, or upon the stage during a concert, hovering over the shoulder of favorite musicians. For example, if a sink rotates (exocentrically visually), the apparent sonic location of the source revolves (egocentrically acoustically) accordingly.

Most discussions of presence in virtual environments are about its quality—degrees of resolution and interactivity (Held & Durlach, 1992; Sheridan, 1992; Sheridan, 1997). This paper assumes elaboration of its quantity (Cohen, 1995,

1998; Cohen & Koizumi, 1998). One's perceptual focus need not be unique or singular. Split or shared perception can be thought of as violating the ''one [sensory] sink to a customer'' allocation that is inherent to immersive systems; in an exocentric paradigm, each user may have an arbitrary number of dedicated virtual sensor instances, and the mapping between sinks and humans may be one to many, many to one, or many to many.

The case of many sinks designated by a single user, explained in more detail later, describes situations in which one has various simultaneous telepresences (like talking on a phone while monitoring an intercom while listening to music). Illustrating a one-to-many mapping of sinks to users (as in broadcast media like TV or radio which effectively employ a single delegate of a collective audience), Cohen and Koizumi (1991) allowed two users to synchronously adjust the position of multiple sources and a single shared sink in a virtual concert, as if they were simultaneously conductor and (singleton) audience. More prosaically, a normal conference call just sums the signals from everyone participating, so that they can be said to share a sink. An example of a many-to-many sink:user mapping is a virtual concert in which the audience shares a distribution of sinks: each user may attend the same soundscape, but multiple sinks can be used to decrease the granularity of audition. Such presentation styles blur the distinctions between composer, conductor/performer, and audience, as hypertext blurs the distinctions between author, publisher, and reader. The extension into audio of spatial sound and the flexible perspective models of virtual reality—catalyzed by the convergence of telecommunication (including telephony), computing, and electronics (including audio, television, and video)—motivate extensions to traditional idioms for sound mixing in musical and conferencing applications (Cohen, 1997).

2   *Deafen & Attend* (*Confide* and *Harken*)

Traditional mixing idioms for selectively activating multiple sources employ *mute* and *solo* functions, which, along with *cue,* disable or focus on respective channels.

Table I. *User and Delegate—Projected Presence*

Human pilot	Representative (projected presence)
carbon community	avatar
RL (real life)	electronic puppet
meatspace	synthespian (synthetic thespian)
motion capture	vactor (virtual actor)

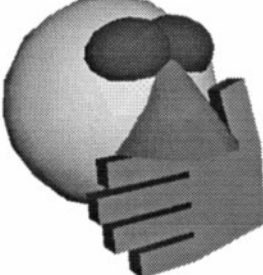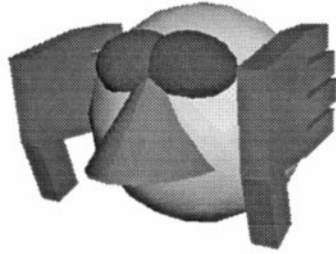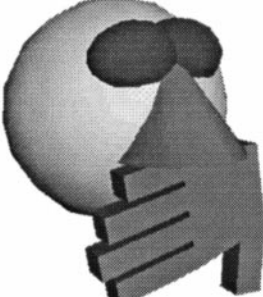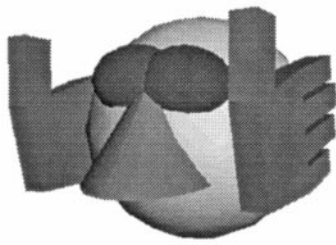An avatar is the reification of an icon in a virtual environment.

Sometimes, just the initial letters *s* and *m* are used (with no lascivious association intended), the *s* standing for *select* as well as *solo.* That *mute* blocks the output of a source, it goes without saying. Exocentric interfaces, which explicitly model not only sources but also location, orientation, directivity, and multiplicity of sinks, motivate the generalization of *mute/solo* and *cue* to `exclude` and `include`, manifested for sinks as *deafen/confide* and *harken,* a narrowing of stimuli by explicitly blocking out and/or concentrating on selected entities (Cohen, 1999). *Deafen* disables sinks; *confide* and *harken* focus on them by disabling others. These extensions can be described in the context of application to three situations, presented in the following sections.

2.1 *Deafen/Confide* Invoked on Other Users' Sinks for Privacy

A simple conferencing configuration typically consists of various icons representing distributed users, moving around shared spaces. These icons each represent a source (the voice of the associated user) as well as a sink (that user's ears). Source attributes *mute* and *solo* or *cue,* settable by each user for each source, are used to focus on some channels exclusively, or selectively still them. *Solo* picks out selected channels for aural scrutiny; *mute* blocks the selection out of the mix. If privacy is desired, confidentialities can be shared in a separate, acoustically isolated (if still virtual), space.

Anisotropic (direction-dependent) sound radiation

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

	Source	Sink
Function	radiation	reception
Level	amplification	sensitivity
Direction	OUTput	INput
Instance	speaker	listener
Transducer	loudspeaker	microphone or dummy-head
Organ	mouth	ear
Tool	megaphone	ear trumpet
Include	solo (select) or cue	attend: confide and harken
Attenuate	muzzle	muffle
Exclude	mute	deafen
own *reflexive*	(thumb up)	(thumbs back)
other *transitive*	(thumb down)	(thumbs up)

patterns (Cohen & Koizumi, 1998; CRE, 1994), like that shown in Figure 1, can be used to define projection of sources and thereby control audibility. Such "nearest-neighbors" or proximity-based techniques of spatial partitioning (Viegas & Donath, 1999) are useful, as in normal conversations, for situations in which one doesn't mind others noticing (as third-party witnesses to a first-to-second person address), but multilobed radiation patterns become impractical for situations in which confidants are spatially distributed, and impossible when sound must "skip over" unaddressed intervening sinks, like knights in chess. Just as a megaphone directionally projects sound or an ear trumpet collects it, a multimegaphone or ears trumpet, like that imagined by Figure 2, represents a generalization, capable of projecting sound fields to multiple arbitrary locations.

Alternatively and more practically, *deafen/confide* functions, sink analogs of *mute/solo,* can be used to narrow multicasts to representatives of other humans in a
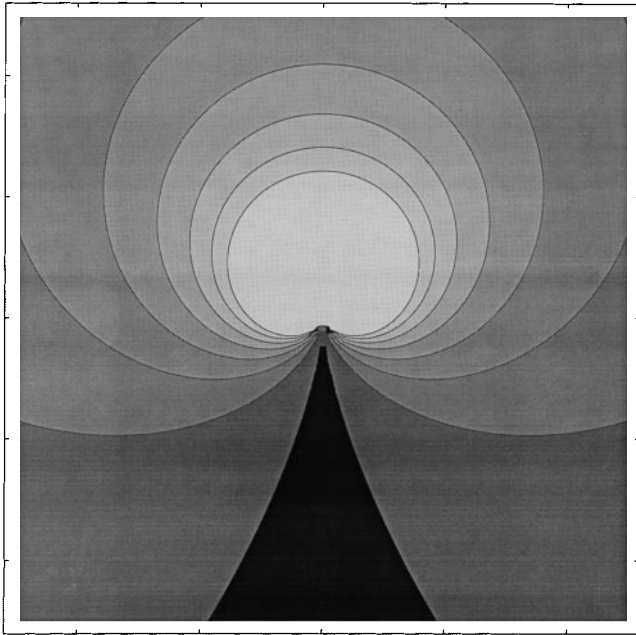
Figure 1. *Contour plot showing projection of radiation pattern combining distance and directional effects. (Generated in Mathematica.)*

Figure 2. *Speaker or microphone as Hydra—multimegaphone (sources) or ears trumpet (sinks).*

groupware environment.[1] Rather than `exclude/include` attributes that näively assume symmetry in a "gossip circle" by linking *mute* to *deafen* and *solo* to *attend,* such differentiated functions allow tighter control. One might want to monitor an ongoing conference while focusing remarks, or share with a gallery whose cacophony one has no use for.

2.2 *Deafen/Harken* Invoked on One's Own Sinks

2.2.1 *Deafen/Harken* (Invoked on One's Own Sinks) Across Several Spaces for Selective Attendance. Designation of multiple sinks across several spaces effectively increases one's attendance. A user may simply fork themselves, leaving one clone hither while installing another yon, compositing soundscapes via the superposition of multiple sinks' presence (Begault, 1994, pp. 213–216). Such a multisink presence, en-

Figure 3. *Superposition of soundscapes (reproduced from Begault with permission).*

abling multiple receivers in different locations, explicitly overlays multiple audio displays, allowing a conferee to leave a pair of ears in one conversation, while sending other pairs to side caucuses.

Audio entities, unlike visual, do not in general occlude, although masking can be thought of as audio occlusion (Bregman, 1990; Blauert, 1997; Cohen & Wen-

---

1. The "Cone of Silence," used by Agent 86 and the Chief in Mel Brooks' TV show *Get Smart,* was intended to acoustically seclude two spies, so that they could exchange secret information without anyone eavesdropping.
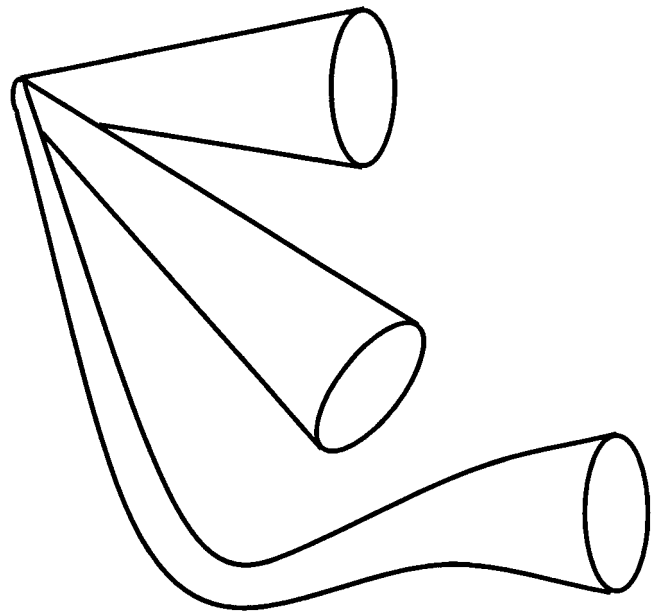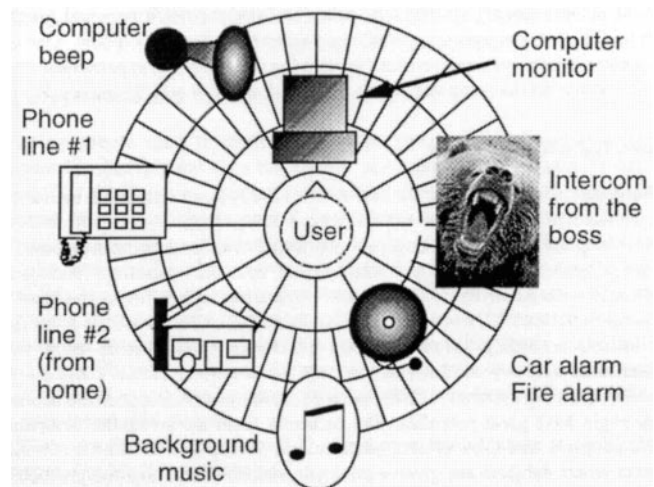
zel, 1995). Combination of soundscapes can be done directly, monaurally or stereophonically, as in a mixer, as shown by Figure 3. In particular, stereo sources—real (or mic'd on a dummy head) or artificial (binaurally spatialized)—may be simply added (Cohen et al., 1993). The overlaid existence so enabled suggests the name given to this effect—*sonic cubism,* presenting multiple simultaneous acoustic perspectives collapsed into a single soundscape (comparable to the way visual cubism collapses several viewpoints of a 3D scene onto a 2D surface). Being anywhere is better than being everywhere, since it is selective. A multisink presence is distilled ubiquity, regarding multiple objects at once. Distinguishing between other users' sinks and one's own (as the two sets might be distinct, identical, or partly overlapping) motivates choice of a special word to describe focusing on a subset of (possibly many of) one's own locations. In the case of a user represented by multiple sinks, *harken* recalls a transitive form of *hark,* a reflexive *confide,* denoting a sense of listening attentively or closely via one's designated sinks' ears.

2.2.2 *Deafen/Harken* **(Invoked on One's Own Sinks) in a Single Space for Selective Multipresence.** The designation of multiple sinks can be used to sharpen the granularity of control within a single space, as separate sinks can monitor individual sources via selective amplification, even if those sources are not repositionable; just as in ordinary settings, social conventions might inhibit dragging someone else around a shared space. One could pay close attention to particular instruments in a concert without rearranging the ensemble, which would disturb the soundscape perceived by icons representing other users in the common model. A useful analogy is a ''Rashōmon simulcast,'' after the eponymous Akira Kurosawa film based on the stories of Ryūnosuke Akutagawa (Akutagawa, 1952), which contrasted multiple perspectives of a single incident. Table 3 presents a taxonomy of points of view, sweeping the continuum from egocentric through exocentric user experiences (Laurel, 1986).

Imagine, for example, that a concert attendee wanted to pay special attention to a drum and rhythm guitar, while preserving the configuration of the instruments.

Besides tradition and mnemonics, one reason for not just rearranging the instruments around a singleton sink is to maintain consistency with other listeners, distributed in time and space (both physical and virtual). One could replicate, and be literally besides oneself. In Figure 4, one avatar is located inside the drum, while another doppelgänger is near the rhythm guitar.

The apparent paradoxes of one's being in multiple places simultaneously (Firesign Theatre, 1968) can be resolved by partitioning the sources across the sinks. If the sinks are distributed across separate virtual rooms, each source is spatialized with respect to the sink in the same room. In the case of autothronging—or multiple sinks designated by a single user in the same space—an autofocus mode can be employed by anticipating level-difference localization, the tendency to perceive multiple identical sources in different locations as a single fused source. This is related to the precedence effect, or ''rule of the first wavefront'' (Wallach et al., 1949; Haas, 1972; Blauert, 1996; Gilkey & Anderson, 1997). Rather than adding or averaging the contribution of each source to one's multiple sinks, each source can be spatialized with respect to only the best (loudest, as a function of distance and mutual gain, including focus and orientation) sink, as shown in Figure 5.

3 Applications

These themes of multiple, selective presence have been explored by two proofs-of-concept.

3.1 Helical Keyboard

The Helical Keyboard (Herder & Cohen, 1996) is a virtual, piano-style keyboard wrapped through a left-handed helix (Shepard, 1984), so that chroma (note within an octave) maps to azimuth and pitch height maps to elevation. The model was generated algorithmically with Mathematica (Wolfram, 1996), exported into VRML (Hartman & Wernecke, 1996), and imported into Open Inventor (Wernecke, 1994), where it is animated by a MIDI streamer.

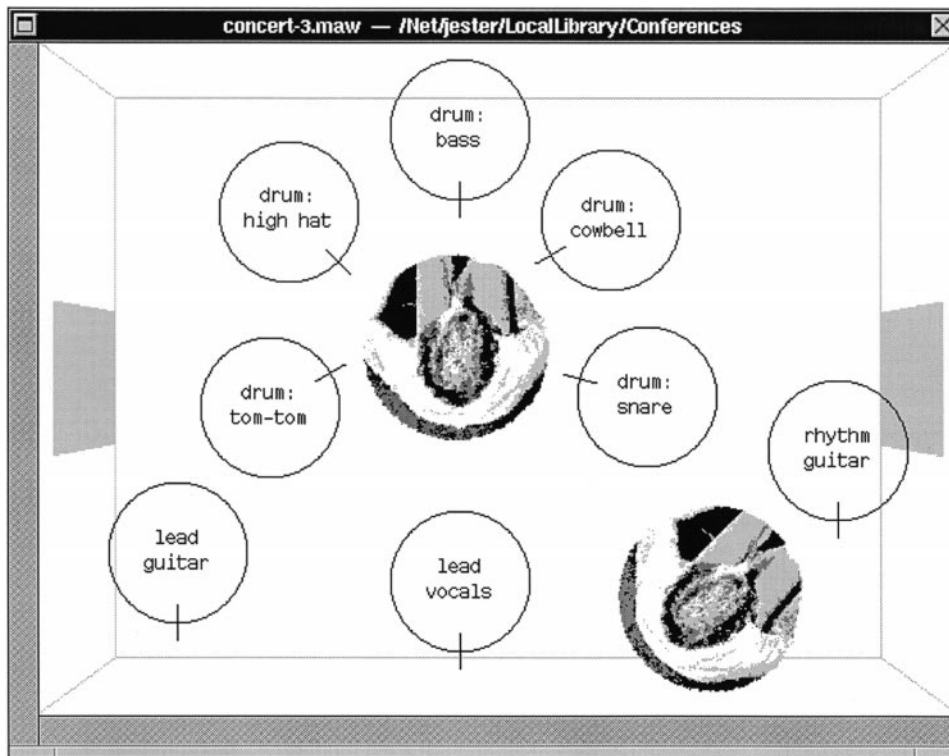Designed to allow separate audition, for instance, of

Figure 4. *Virtual concert: multiple sinks (generalized multifocus audio fish-eye).*

harmony and melody, the model is meant to be experienced in a hemispherical speaker array (Amano et al., 1998). A single sink inside the helix near its base might easily determine the azimuth of the harmony, but the melodic notes would all seem to come from the upper pole. As shown by Figure 6 (see p. 94), multiple sinks can normalize the octave, and can be selectively disabled by an `active` toggle.

3.2 MAW: Multidimensional Audio Windows

MAW is an application for manipulating sound sources and sinks in 2D virtual rooms, capable of driving a heterogeneous backend (Cohen & Ludwig, 1991a, 1991b; Cohen, 1993). The graphical representation of MAW's virtual rooms is an orthographic plan view. Figure 7 shows a snapshot of such a representation as part of a typical session (mixing the top-down metaphor used

in Figure 4 with frontal snapshots), using multiple sinks as foci of a generalized fisheye (Furnas, 1986) audio ''lens.''

MAW supports multiple, independent, simultaneous conferences and concerts; a source is inaudible to a sink in a different virtual room. The *cut/paste* idiom is used as a wormhole (teleporter), so a subcaucus may be spawned simply by cutting a coterie out of one room and pasting it (''beaming down'') into another. Users wanting to monitor simultaneous conferences need only fork themselves with, for example, *copy/paste,* installing (multiply designated) replicant sinks in each room of interest. A mixels panel,[2] shown in

---

2. *Mixels,* acronymic for sound mixing elements,—in analogy to *dexels* (depth elements), *hogels* (holographic elements), *pixels* (picture elements), *taxels* (tactile array elements), *texels* (texture elements), and *voxels* (volumetric elements, a.k.a. *boxels*), since they are like a raster across which a soundscape is projected—define the granularity of control and degree of directional or spatial polyphony.
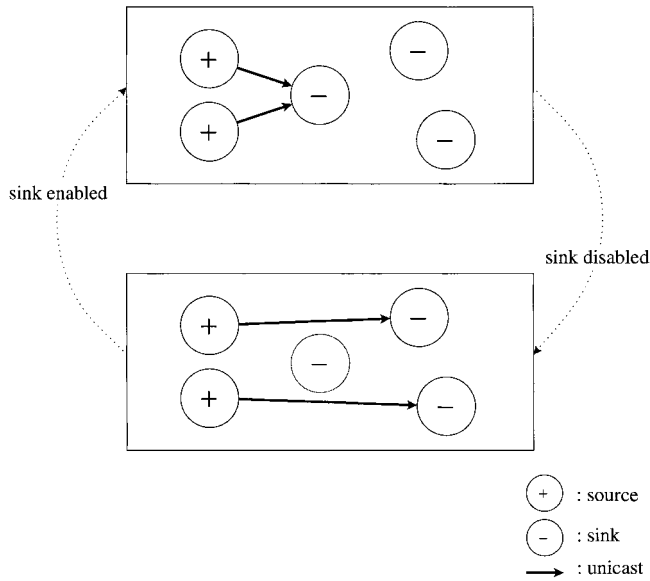
Figure 5. *Unicast source → sink transmissions. If an attending sink is deafened (or peers confided in), remaining sinks adopt orphaned sources.*

Figure 8, can be used to activate or deactivate sources and sinks with *solo, mute, attend* (*confide* or *harken*), and *deafen.*

4   Formalization

The suite of inclusion and exclusion narrowcast commands for sources and sinks are like analogs of burning and dodging (shading) in photographic processing. The analogy between source and sink operations is close, and the semantics are identical: an icon is enabled by default unless it is explicitly excluded (with *mute* and *deafen*), or peers are explicitly included (with *solo* or *cue* and *attend: confide* or *harken*) when the respective icon is not. Because a source or a sink is active by default, invoking `exclude` and `include` operations simultaneously on an object results in its being disabled. In predicate calculus notation,

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

So, for *mute* and *solo,* the relation is

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

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/attend,* the relation is

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

5   Figurative Representation

Distinguishing between operations involving one's own and others' representatives, Table 4 proposes a taxonomy of narrowcasting functions by juxtaposing disabling and enabling operations, generally as well as for visual and audio modalities. The famous ''hear/speak/see no evil'' monkeys, pictured in Figure 9, are examples of reflexive ''[sink] *deafen/*[source] *mute/*[sink] avert,'' which audio relations fill the top-right sextant of Table 4.[3] Traditional mixing console functions *mute/solo* and *cue,* operating as they do on sources that are metaphorically remote, correspond to transitive ''[source] *mute*'' (illustrated by Figure 10a), which, along with transitive ''[sink] *deafen*'' (illustrated by Figure 10b), fills the bottom-right sextant.

A figurative avatar in virtual space is humanoid, and especially includes a head, which embodies not only a center of consciousness, but also the ears, mouth, and eyes. `Exclude` and `include` source and sink properties can be visually represented by iconic attributes which can distinguish between operations reflexive (invoked by a user associated with a respective icon) and transitive (invoked by another user in the shared environment) (Cohen & Herder, 1998). Distributed users might typically share spatial aspects of a groupware environment, but attributes like *mute*dness or *deafen*edness are determined and displayed on a per-user basis (Chen et al., 1999).

---

3. ''Blind,'' the dual of the visual ''see no evil'' avert operation, corresponds to a video conference ''sneeze button,'' which blocks or freezes transmission.
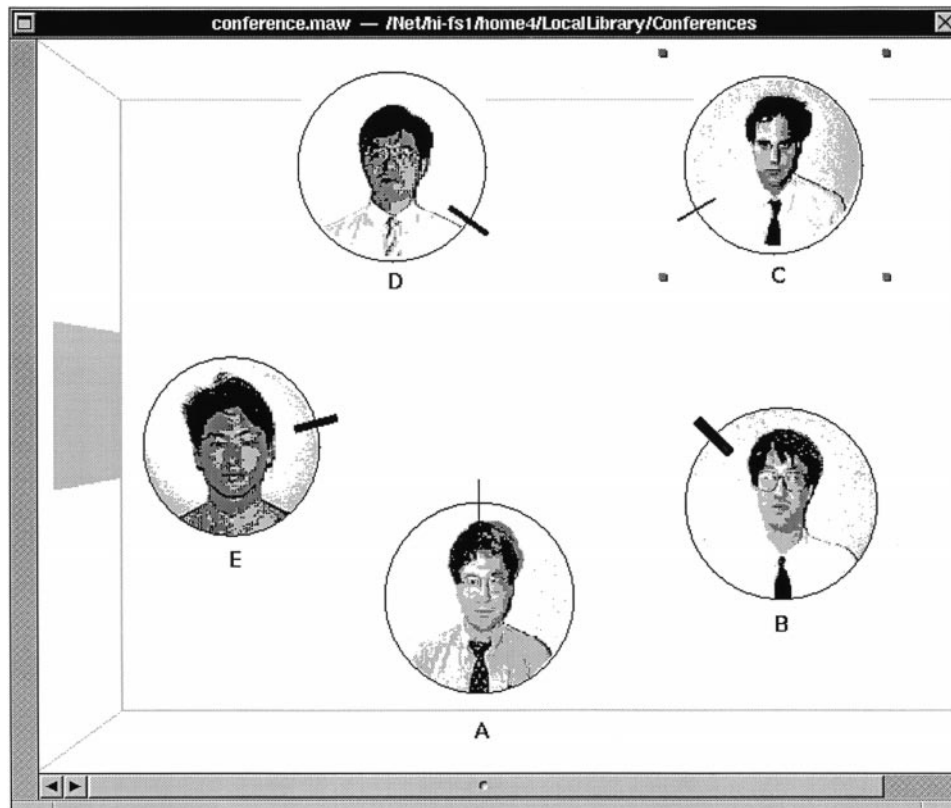
Figure 7. *MAW conference.*

channel	room	gain	sink	solo	mute	attend	deafen	x	y	azimuth	focus	size
sink 0	concert	active		☐	☑	☐	☐	409	344	180	0	132
sink 1	concert	inactive		☐	☑	☐	☑	179	353	90	0	110
sink 2	concert	active		☐	☑	☐	☐	299	154	0	0	132
source 3	concert	0.218	0	☑	☐	☐	☐	291	324	180	0	110
source 4	concert	0.218	0	☑	☐	☐	☐	440	275	-120	0	110
source 5	concert	0.187	2	☑	☐	☐	☐	151	231	135	0	110
sink 6	conference	active		☐	☐	☑	☐	529	199	-30	0.8	167
source 7	conference	0.189	6	☐	☐	☑	☐	310	122	0	0	166
source 8	conference	0.203	6	☐	☐	☐	☑	516	423	-120	0.2	155
source 9	conference			☐	☑	☐	☑	136	227	75	0.6	155
source 10	conference	0.192	6	☐	☐	☐	☐	249	422	125	0.4	165

Azimuth range
◯ 0 – 360
◯ (–180) – 180

Azimuth units
degrees ▾

Figure 8. *Mixels panel: sinks and sources across multiple spaces. The `solo` and `confide` columns employ the familiar "radio buttons" idiom, in which the selection is presumed to be a singleton. Asserting an attribute for one object (by checking it in the respective column at the designated row) resets it for any others, unless the selection set is explicitly extended (by holding down a shift key while asserting the property).*

Table 3. *Points of View*

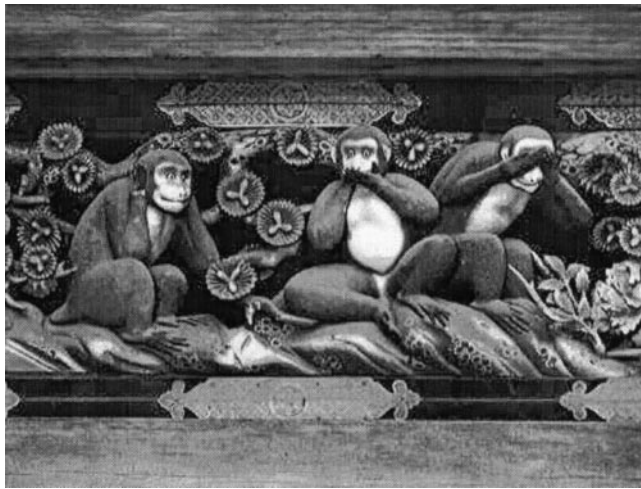| Point of view	Person	Intimacy	Object	Distance	Mode	Perspective
exocentric	3rd	public	other	distal	transitive	objective
vicariousness, empathy	2nd	social, multipersonal	familiar	medial	imperative	
telepresence, autoempathy		remote self				
immersive	1st	personal	self	proximal	reflexive	subjective
egocentric						

Figure 9. *Monkeys at Tōshōgū Shrine (in Nikkō, Japan): Kikazaru ("hear no evil"), Iwazaru ("speak no evil"), and Mizaru ("see no evil").*

For example, a source representing a human teleconferee denotes *mute*dness with an iconic hand clapped over its mouth, oriented differently (thumb up or thumb down) depending on whether the source was *mute*d by its owner (or one of its owners) or another unassociated user. (In the former case, all the users in the space would observe the *mute,* but, in the latter, only the user disabling the remote source would see the *mute.*) An audio muffler might be wrapped around an iconic head to denote its deafness, but to distinguish between self-imposed deafness (invoked by one whose attention is directed elsewhere) and distally imposed (invoked by another desiring selective privacy), hands clasped over the ears can be oriented differently depending on the agent of deafness. These cases are illustrated by Table 2.

Such disabling qualities are not mutually exclusive, and, indeed, the orthogonal iconic attributes can be superimposed, albeit confusingly, as in Figure 11, an "omnigrope" extreme case. Simultaneously applied filters can be represented by interpenetrated virtual models, nonverbal communication being used to symbolize access permissions.

6 Future Research

6.1 Continuum of Audibility

Besides the dichotomous on/off of the `include/exclude` functions, we plan to fuzzify the audibility continuum by programming functions that focus source-to-sink transmissions without blocking them from others in the same space, a "sharing" meant to denote a non-private aside. (For instance, some home listening consoles have a so-called "mute" function that reduces the volume by approximately 20 dB instead of cutting it altogether.) A "casual confide" function could eventually be combined with an obtrusive mode (Mershon, 1997; Martens, 1997), invoking source-side, near-field transfer functions for whispering and *sotto voce* effects. A notion of "virtual social distance" (Michelitsch et al., 1998) can be used to scale the quality of audio narrowcasts, including representation as a synthetic murmur.

6.2 Audibility Protocol

Modeling sources and sinks as software objects, an audibility protocol describes transitions between states in which the respective methods are appropriate, as shown by Figure 12.

Table 4. *Exclude/Include Taxonomy: Enable/Disable for One's Own and Others' Representatives*

	General		Visual		Audio	
	Source	Sink	Source	Sink	Source	Sink
Own, Self (*reflexive*)	hide/ manifest, appear, project	ignore/ `attend`, notice	hide, conceal/ expose	avert/ see	`mute` (stifle)/ `cue` (play, speak)	`deafen`/ `harken` (heed)
Other (*transitive*)	hide/ expose	block/ invite, select	hide/ expose	mask, blind/ ?	`mute`/ `solo`	`deafen`/ `confide`
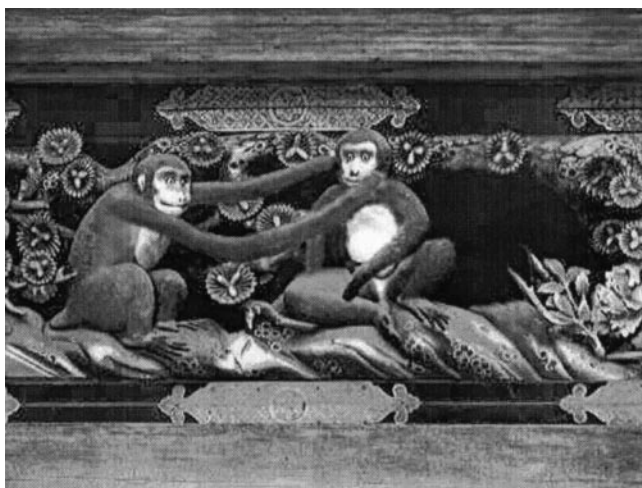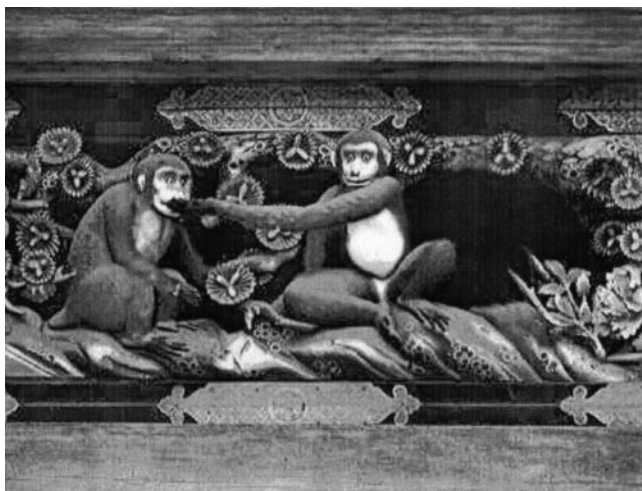
Figure 10. *Distal* `exclude`.

Because of the asymmetry of both *mute/ solo* and *deafen/ attend* (that is, audibility is assumed for collocated icons), audibility of a source with respect to a sink should be treated as a revocable privilege and a forsakable right. For example, audibility would be `granted` by a source upon one's entering a space, acknowledgeable by the respective sink by `claiming` that attribute. A source wishing to `exclude` certain sinks from audibility would invoke, directly via *deafen* or indirectly via *attend,* a `revoke` method, duly acknowledgeable by each disabled sink's `renounce`. Further policy extensions will relax the symmetry of such a protocol, including the ability to force audibility by overriding a source's *mute* or sink's *deafen* (which a parent might invoke when telechiding a distracted child: "How dare you attenuate my voice?!").

Groupware implementation of *deafen/ attend* should eventually be done nondistally, as privacy concerns become relevant. Rather than distributing all the source streams, as our prototypes do, and expecting the software to ignore private transmissions for others, a properly secure implementation would restrict distribution earlier. Full articulation of groupware extensions for privacy requires multicasting of listening permissions.

Such protocols might ultimately be deployed directly on the internet, dynamically loaded into multicasting routers. Active networks allow programmable network infrastructure. An application-specific audibility protocol like that described here could dynamically reconfigure the routers' policy, for security, privacy, and to reduce network traffic.

7 Conclusion

The protocols and methods defined and suggested by this research enable narrow- and multicasting idioms for selective privacy and attendance, scalability and LOD (level of detail), and side- and back-channels. Usually one thinks of one's perspective as residing in a single place—namely, behind one's eyes, between one's ears,
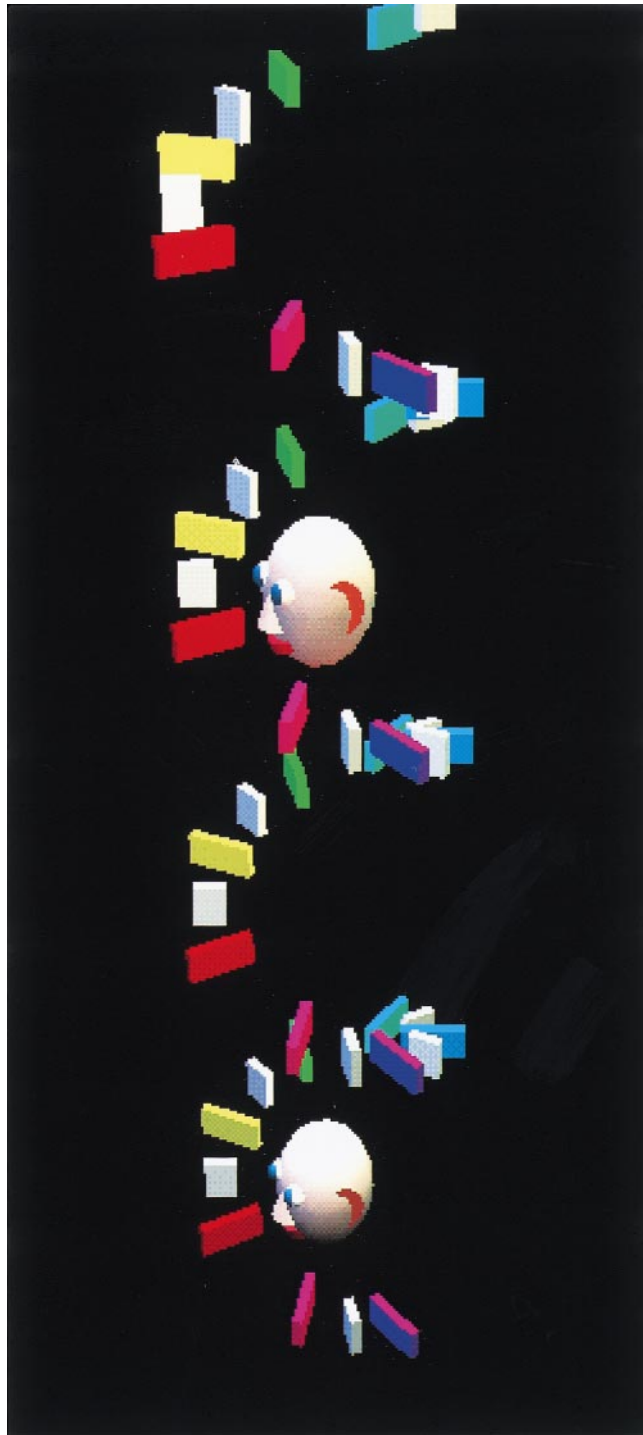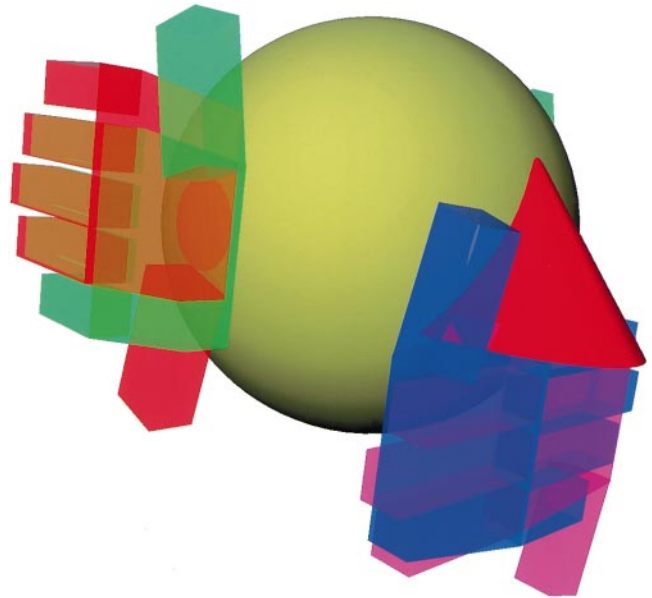
Figure 6. *Multipresence via multiple sinks.*

Figure 11. *Figurative avatar omnigroping interdigitation. A source representing a human teleconferee denotes mutedness with an iconic hand clapped over its mouth, oriented differently (thumb up or thumb down) depending on whether the source was muted by its owner (or one of its owners) or another user. To distinguish between deafness self-imposed (invoked by a user whose attention is directed elsewhere) versus distally imposed (invoked by a user desiring selective privacy), hands clasped over the ears orient differently depending on the agent of deafness. Being both virtual and conceptually orthogonal, these various hands interpenetrate.*
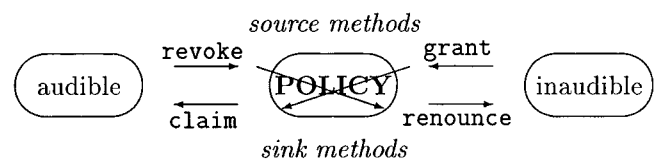

Figure 12. *Audibility protocol.*

and so forth—but telepresence enables such points of attendance to be distributed and nonsingular, by replicating subject instead of object. In Figure 13, separate soundscapes corresponding to music (top left), telephony (top right), mobile vehicular communication (bottom left), and a workstation (bottom right) are combined into a single percept, one's perceptual spaces
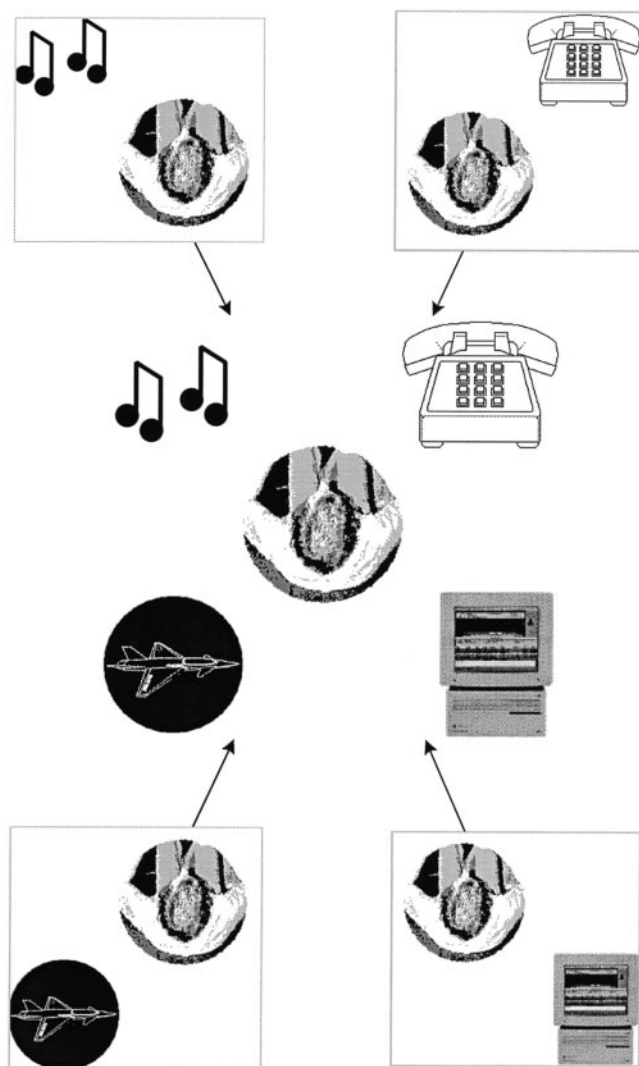
Figure 13. *Soundscape superposition: overlaid mutiple soundscapes.*

being naturally coextensive, the center of one's consciousness being singular.

Acknowledgments

References

Akutagawa, R. (1952). *Rashomon and Other Stories.* Charles E. Tuttle Company, Inc. ISBN 0-8048-1457-0.

Amano, K., Matsushita, F., Yanagawa, H., Cohen, M., Herder, J., Martens, W., Koba, Y., & Tohyama, M. (1998). A virtual reality sound system using room-related transfer functions delivered through a multispeaker array: The PSFC at the University of Aizu Multimedia Center. *TVRSJ: Trans. of the Virtual Reality Society of Japan, 3*(1), 1–12. ISSN 1344-011x.

Begault, D. R. (1994). *3-D Sound for Virtual Reality and Multimedia.* Academic Press. ISBN 0-12-084735-3.

Blauert, J. (1996). *Spatial Hearing: The Psychophysics of Human Sound Localization* (2nd ed.). MIT Press. ISBN 0-262-02413-6.

———. (1997). Acoustical simulation and auralization for VR and other applications. *Proc. ASVA: Int. Symposium on Simulation, Visualization and Auralization for Acoustic Research and Education* (pp. 261–268). Tokyo.

Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound.* MIT Press. ISBN 0-262-02297-4.

Chen, C., Thomas, L., Cole, J., & Chennawasin, C. (1999). Representing the semantics of virtual spaces. *IEEE MultiMedia,* April–June, 54–63.

Cohen, M. (1993). Integrating graphical and audio windows. *Presence: Teleoperators and Virtual Environments, 1*(4), 468–481. ISSN 1054-7460.

———. (1995). Besides immersion: Overlaid points of view and frames of reference; using audio windows to analyze audio scenes. In S. Tachi (Ed.), *Proc. ICAT/VRST: Int. Conf. Artificial Reality and Tele-Existence/Conf. on Virtual Reality Software and Technology* (pp. 29–38). Makuhari, Chiba, Japan.

———. (1997). Exclude and include for audio sources and sinks: Analogs of mute∕solo & cue are deafen∕confide & harken. *Proc. ICAD: Int. Conf. Auditory Display* (19–28), Palo Alto, CA.

———. (1998). Quantity of presence: Beyond person, number, and pronouns. In T. L. Kunii, & A. Luciani (Eds.), *Cyberworlds* (pp. 289–308). Springer-Verlag. ISBN 4-431-70207-5.

———. (1999). Chat space models. *Proc. Joint Meeting of the 137th Regular Meeting of the Acoustical Society of America and the 2nd Convention of the European Acoustics Association:*

Forum Acusticum, 2 (pp. 1099). Berlin. Signal Processing for Teleconferencing and Smart Microphones, 2pSPa7.

Cohen, M., Aoki, S., & Koizumi, N. (1993). Augmented audio reality: Telepresence/VR hybrid acoustic environments. *Proc. Ro-Man: 2nd IEEE Int. Workshop on Robot and Human Communication* (pp. 361–364). Tokyo. ISBN 0-7803-1407-7.

Cohen, M., & Herder, J. (1998). Symbolic representations of exclude and include for audio sources and sinks. In M. Göbel, J. Landauer, U. Lang, & M. Wapler (Eds.), *Proc. VE: Virtual Environments* (pp. 235–242). Stuttgart: IEEE, Springer-Verlag Wien. ISSN 0946-2767; ISBN 3-211-83233-5.

Cohen, M., & Koizumi, N. (1991). Audio window. *Den Gaku.* Tokyo Contemporary Music Festival: Music for Computer.

———. (1998). Virtual gain for audio windows. *Presence: Teleoperators and Virtual Environments, 7*(1), 53–66. ISSN 1054-7460.

Cohen, M., & Ludwig, L. F. (1991a). Multidimensional audio window management. *IJMMS: the Journal of Person-Computer Interaction, 34*(3), 319–336. Special Issue on Computer Supported Cooperative Work and Groupware. ISSN 0020-7373.

———. (1991b). Multidimensional audio window management. In S. Greenberg (Ed.), *Computer Supported Cooperative Work and Groupware* (pp. 193–210). London: Academic Press. ISBN 0-12-299220-2.

Cohen, M., & Wenzel, E. M. (1995). The design of multidimensional sound interfaces. In W. Barfield & T. A. Furness III (Eds.), *Virtual Environments and Advanced Interface Design*, Chapter 8 (pp. 291–346). Oxford University Press. ISBN 0-19-507555-2.

CRE (1994). *CRE_TRON Library Reference Manual.* Crystal River Engineering, Inc., Revision B.

Firesign Theatre (1968). *How can you be in two places at once when you're not anywhere at all.* Lp Columbia 9884; Cd Mobile Fidelity 834.

Furnas, G. W. (1986). Generalized fisheye views. *Proc. CHI: ACM Conf. on Computer-Human Interaction* (pp. 16–23). Boston.

Gilkey, R. H., & Anderson, T. R. (Eds.). (1997). *Binaural and Spatial Hearing in Real and Virtual Environments.* Mahway, NJ: Lawrence Erlbaum and Associates. ISBN 0-8058-1654-2.

Haas, H. (1972). The influence of a single echo on the audibility of speech. *J. Aud. Eng. Soc., 20,* 146–159.

Hartman, J., & Wernecke, J. (1996). *The VRML 2.0 Handbook.* Reading, MA: Addison-Wesley Developers Press. ISBN 0-201-47944-3.

Held, R. M., & Durlach, N. I. (1992). Telepresence. *Presence: Teleoperators and Virtual Environments, 1*(1), 109–112. ISSN 1054-7460.

Herder, J., & Cohen, M. (1996). Project report: Design of a helical keyboard. *Proc. ICAD: Int. Conf. Auditory Display* (pp. 139–142). www.santafe.edu/~icad/ICAD96/proc96/herder.htm. Palo Alto, CA.

Laurel, B. (1986). Interface as mimesis. In D. A. Norman & S. W. Draper (Eds.), *User Centered System Design.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Martens, W. L. (1997). Acoustics and perception of sound sources at close range. In preparation. http://www.u-aizu.ac.jp/~wlm/research/close_range.

Mershon, D. H. (1997). Phenomenal geometry and the measurement of perceived auditory distance. In Gilkey & Anderson (1997), chapter 13 (pp. 257–274).

Michelitsch, G., Welling, G., & Ott, M. (1998). The role of virtual distance in the design of communication services. In T. Kamae (Ed.), *IWNA: Proc. Int. Workshop on Networked Appliances,* Kyoto 55-2.

Shepard, R. N. (1984). Structural representations of musical pitch. In D. Deutsch (Ed.), *The Psychology of Music* (pp. 343–390). Academic Press. ISBN 0-12-213560-1 or ISBN 0-12-213562-8.

Sheridan, T. B. (1992). Musings on telepresence and virtual presence. *Presence: Teleoperators and Virtual Environments, 1*(1), 120–125. ISSN 1054-7460.

———. (1997). Further musings on the psychophysics of presence. *Presence: Teleoperators and Virtual Environments, 5*(2), 241–246. ISSN 1054-7460.

Viegas, F. B., & Donath, J. S. (1999). Chat circles. *Proc. CHI: ACM Conf. on Computer-Human Interaction*, Pittsburgh.

Wallach, H., Newman, E. B., & Rosenzweig, M. R. (1949). The precedence effect in sound localization. *American Journal of Psychology, 57,* 315–336.

Wernecke, J. (1994). *The Inventor Mentor.* Addison-Wesley Developers Press. ISBN 0-201-62495-8.

Wolfram, S. (1996). *The Mathematica Book* (3rd ed.). Wolfram Media/Cambridge University Press. ISBN 0-9650532-0-2.