Quantity of Presence:
Beyond Person, Number, and Pronouns

Michael **Cohen**
Spatial Media Group
Human Interface Lab
University of Aizu 965-8580
Japan
voice: [+81](242)37-2537
fax: [+81](242)37-2549
e-mail: mcohen@u-aizu.ac.jp
www: http://www.u-aizu.ac.jp/~mcohen

Abstract

Alternative non-immersive perspectives enable new paradigms of perception, especially in the context of frames-of-reference for musical audition and groupware. MAW, acronymic for **m**ultidimensional **a**udio **w**indows, is an application for manipulating sound sources and sinks in virtual rooms, featuring an exocentric graphical interface driving an egocentric audio backend. Listening to sound presented in such a spatial fashion is as different from conventional stereo mixes as sculpture is from painting. Schizophrenic virtual existence suggests sonic (analytic) cubism, presenting multiple acoustic perspectives simultaneously. Clusters can be used to hierarchically organize mixels, [sound] **mix**ing **el**ements. New interaction modalities are enabled by this sort of perceptual aggression and liquid perspective. In particular, virtual concerts may be "broken down" by individuals and groups.
Keywords and Phrases: binaural directional mixing console, CSCW (**c**omputer-**s**upported **c**ollaborative **w**ork), **f**rames **o**f **r**eference, groupware, mixel ([sound] **mix**ing **el**ement), **p**oints **o**f **v**iew, sonic (analytical) cubism, sound localization, spatial sound.

0 Introduction

"Traditional" immersive VR systems feature a HMD (**h**ead-**m**ounted **d**isplay) that tracks the user's position, adjusting visual and audio displays accordingly. Because of the intimate coupling between control and display in such a system, there is a sense of framelessness, of being inside the projected world. This intimacy is not without its cost, however, as it implies a strict mapping between each user and the respective displays. To enable potentially useful modalities like omniscient views and shared or overlaid displays, different **c**ontrol/**d**isplay conventions are needed that relax the mapping between user and presence, applied, for instance, to desktop or 'fishtank' VR systems. This chapter explores the philosophical distinction between egocentricism and exocentricism, especially as blurred by emerging technologies.

1 Duality and Synthesis of Self/Other: Beyond Person

In any kind of display, there is a constant tension between the realism of the presence and one's unwillingness to suspend disbelief. As the realism of the presentation increases, one becomes increasingly, if subconsciously, willing to accept immersion, enabling an egocentric impression. Exocentricism, in contrast, is an awareness that the display derives from a perspective different from where the user imagines themself to be. The egocentric nature of a display is not an inherent quality of the presentation, but a subjective willingness of the user to project their perceptual center to the point-of-view of the display. A few examples demonstrate:

- A good movie or book is absorbing partly to the extent that the attendee or reader projects themself into the story or scene. Immersed in a compelling situation, the subject loses their identity (empathy and vicariousness are projected egocentricism), only to be brought back to an awareness of their actual place by a crunch of popcorn or jangle of a telephone, reasserting an exocentric perspective.

channel, was unable to perceive a single object; he couldn't (let himself) ignore the fact that the headphones were actually playing separate sounds to each ear. For him, the egocentric display was hobbled, reduced to its exocentric shadow by an overzealous self-consciousness.

- A classic example of an exocentric display is a map. If someone allows themself an imagined out-of-body (but not out-of-mind) 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.) One can slide back and forth along a spectrum between egocentric and exocentric impressions or perspectives.

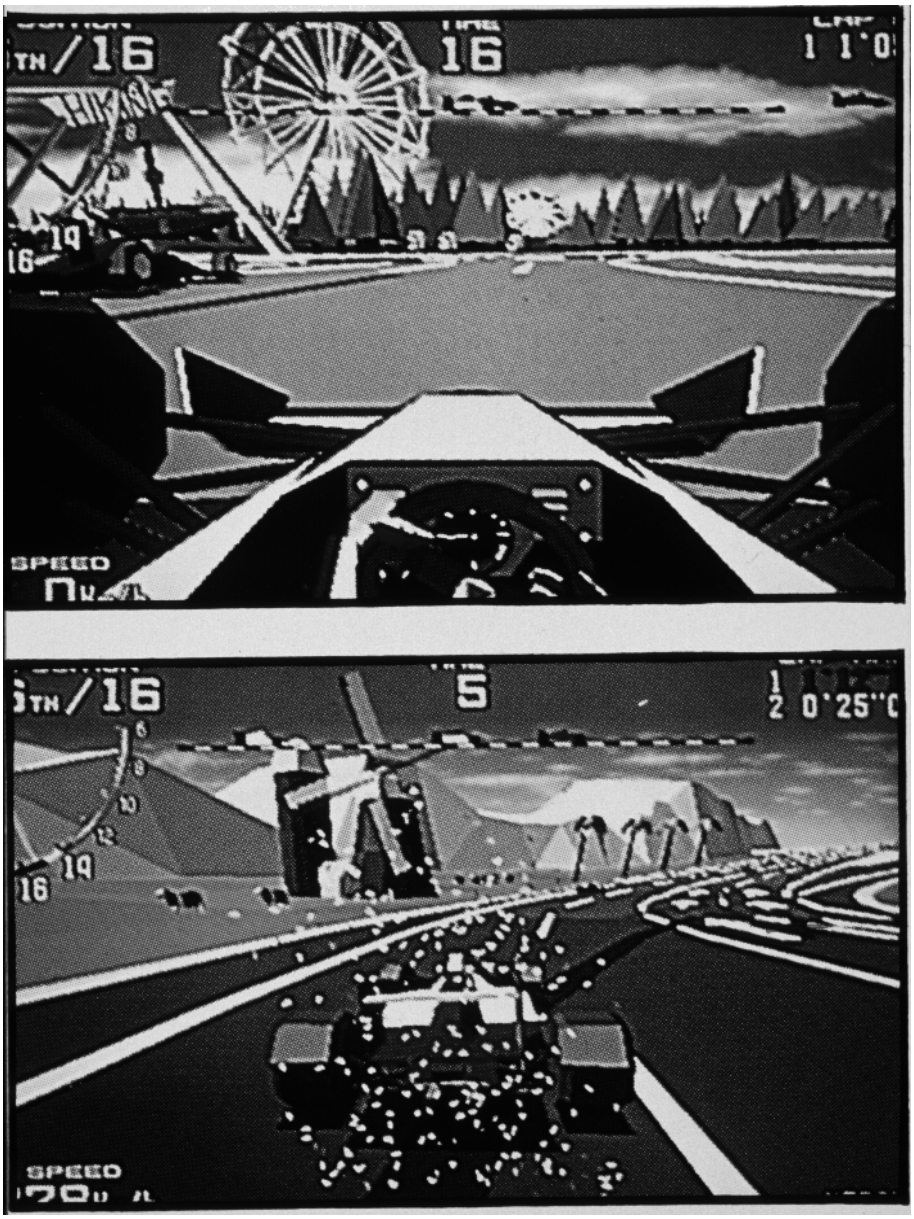

Figure 1: Sega **V**irtua **R**acing

- A networked Formula 1 racing simulator arcade game, Sega's "**V**irtua **R**acing," allows each driver to switch between four perspective modes:

  cockpit (Figure 1 top), in which the visual presentation is as if the user were inside the car, including the dashboard, top of the steering wheel (including driver's hands), and rearview mirrors;

  follow (Figure 1 bottom), in which the driver's perspective is just behind and above the vehicle, tracking synchronously;

'forward' from the driver's point-of-view; and

fly, in which the monitor tracks the car as if from a blimp, clearly showing one's own car in the context of the field.

Even though the simulator's 'radio buttons' select a predetermined degree of immersion, drivers may switch modes during a race, and the visual display slides seamlessly between them, by zooming, focusing, and soaring the virtual camera through the **c**omputer **g**raphic raceway. Further blurring the sampled/synthesized distinction, monitors for spectators show live video of the drivers, panning shots of the lead car, static shots of strategic curves, and instant replays of crashes [Coh94].

For conversational groupware systems, the notion of egocentric and exocentric frames of reference can be reconciled with grammatical person. In sliding from an immersive (subjective) perspective to an "exmersive" (objective) perspective, the user transforms from a 1st person to a 3rd person. If all participants are represented by separate icons, a user could adjust another's virtual position as easily as her own, blurring the self/other distinction. Reflexive and imperative operations are thereby cast as special cases of transitive commands. By projecting the metaphorical world onto an external and egalitarian medium, the 1st and 2nd persons have become special cases of the 3rd.

2 Shared and Split Perception: Beyond Number

Most discussions of virtual presence are about its quality– degree of individual resolution and interactivity [HD92] [She92a]; here its *quantity* is elaborated. Once it is admitted that any display can be egocentric, given appropriately imaginative users, the issue of multiple simultaneous or overlaid egocentric perspectives, or multifocal virtual presence, can be addressed. One's perceptual center need not be unique or singular, just as the effects of one's actions need not be limited to a single place.

These split or shared perceptions can be thought of as violating the "one [sensory] sink to a customer" rule inherent to immersive systems; each user may have an arbitrary number of dedicated virtual sensor instances, and the mapping between sinks and users may be one→many, many→one, or many→many.

Imagine this experiment: A user is connected to a hand position sensor, which drives, via telerobotics, a pair of identical manipulators, playing separate instruments — a harpsichord and a grand piano, in arbitrarily different locations. (This experiment is easily simulated by using a MIDI configuration, say, to fork-drive multiple voices.) The user can be said to have a presence in multiple places.

Now imagine the dual of this multiple effector situation, multiple sensory locations. This notion is related to the idea of multiple cooperating agents in a telepresence environment [She92b]. Different modalities can superimpose separate channels in different ways, outlined later.

The opposite situation, multiple users sharing a single sensor instance, can also be useful: "This is interesting; share it with me..." Mass broadcast media like radio and TV employ this one→many mode (made explicit by first-person movies like "84 Charlie Mopic"). Of course they lack the control of VR systems, but interactive television (suggested by zapping movies whose simultaneous parallel broadcast of multiple characters' stories allows viewers to follow alternate threads) and call-in shows blur this distinction.

2.1 Video

There are several ways of presenting multiple video channels simultaneously. Distributed camera systems can present multiple views at once. Visual superposition is achieved by tiling strategic perspectives, like security monitors, or by embedding a view in a less important section ("picture in picture"). It is difficult in general to use translucency to overlay opaque scenes, except in special circumstances. Split-screen television and cinematographic techniques are common. Montage offers a time-domain multiplexed worldview, as one's perceptual center flitters from place to place, which may or may not correspond to a character's location. Music videos, for example, often composite or crossfade visual scenes. Analytical cubism, as developed by Braque and Picasso, presents multiple visual perspectives on a scene simultaneously.

HUDs (**h**ead-**u**p **d**isplays) are used in airplanes to throw navigation, tracking, and status information onto the windshield. Half-silvered mirrors can be used to view translucent images. Clearboard (Figure 3) [Ish92] [IKG93] uses superimposed translucent viewing planes for teleconferencing with video of the conferees plus a shared whiteboard; different focal distances can help distinguish the layers. [OTT92]

Figure 2: RCAST Telerobot

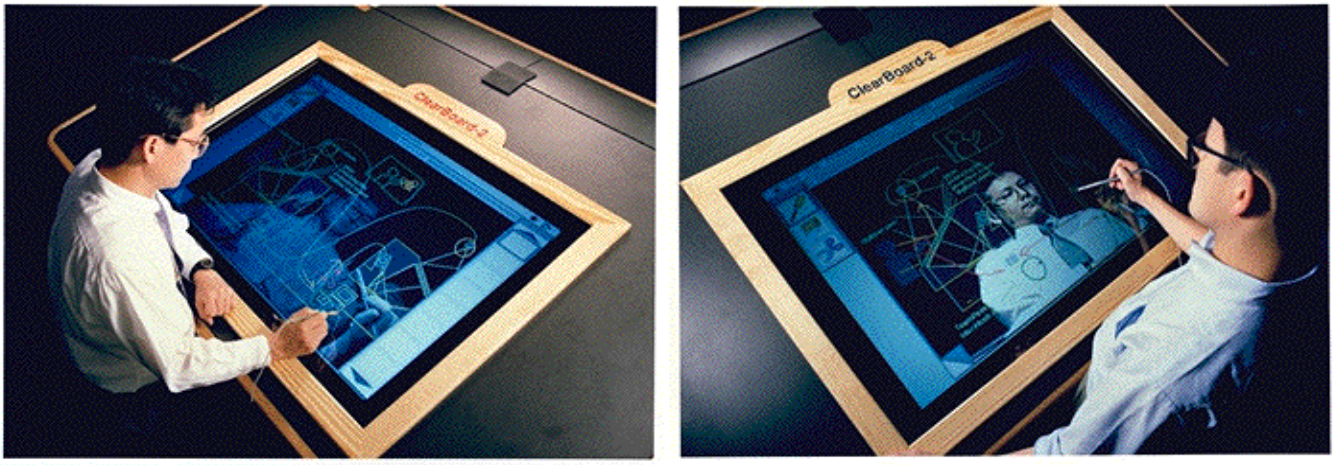

Figure 3: Clearboard-2: shared multilayer drawing tool

virtual image as a (non-occluding) wireframe. "Mirror-type" VR systems like Mandala [WV93] (Figure 4) can combine CG and (chroma-key captured) sampled signals, overlaid on arbitrary background graphics.

Figure 4: Vivid Mandala

Visual "augmented reality" describes the superposition of computer-generated imagery on top of a see-through display [CM92] [WMG93]. The dual of augmented reality is augmented virtual reality, manifested in the video domain by texture mapping camera-captured images on polygon models, as suggested by Figure 5 [HY92].

Presenting different signals presented to separate eyes (of which using computer graphics to simulate stereopsis is a special case) is also possible. While future generations of users might be able to mentally integrate or perceptually multiplex separate scenes presented to each eye, binocular views, augmented with status information tucked into a corner of a display (as in Private Eye [Bec92] or ScopeHand [SK92]), seems like the most we can expect for the near future.

2.2 Audio

Video is not the only modality in which multiple displays may be superimposed. For example, multiple tactile or temperature channels can be simultaneously experienced, by presenting them to different hands.[1] Similarly, dichotic experience involves simultaneous presentation of separate audio scenes to each ear. More directly, an arbitrary number of audio channels may be simply added and played diotically, the same composite signal at each ear. Audio entities, unlike visual, do not occlude (although masking can be thought of as audio occlusion). It is usually straightforward to overlay sonic landscapes, monaurally or stereophonically, as in a mixer. In particular, stereo sources– real (or mic'd via a dummy head) or artificial

[1] This recalls the adaptation parlor trick of immersing opposite hands in baths of hot and cold water, then plunging them together into tepid, to consequent cognitive confusion.

Figure 5: Hirose Lab Virtual Dome

(binaurally spatialized)– may be simply added.

Using such a scheme, distributed microphone systems can superimpose auditory scenes. Musical recording can be thought of as presenting sound as if the listeners had their ears near all the respective instruments, even though the tracks might have been laid down in separate (acoustically isolated) rooms and at different times.

One could share or swap ears with another user, and listen to oneself as a distal source. This is also not terribly exotic: singers often amplify their voice, and musicians want to be able to monitor a live performance from the perspective of the audience, the same way people look in a mirror.

Augmented reality in the audio domain can superimpose computer synthesized sounds upon natural, using some non-exclusive sound presentation like loudspeakers or open-ear headphones [CAK93]. For instance, the author has perceptually thrown a ringing sound to a location occupied by a muted telephone, recalling [Nai91]'s visual analog of projecting a picture of a room on the same space after it was painted white. **P**ublic **a**ddress, or sound reinforcement, systems are a common example of augmented audio reality.

This kind of superposition potential is manifested in MAW (acronymic for **m**ultidimensional **a**udio **w**indows), an audio windowing system with a visual map and auditory display: an interface for manipulating iconic sound sources and sinks in virtual rooms, deployed as a binaural directional mixing console. MAW is suitable for synchronous applications like teleconferences or concerts, as well as asynchronous applications like voicemail and hypermedia [Coh93a] [Coh93b] [CK94], which can be thought of as equivalent (because of spatial data models) to cyberspace [ZPF+94], as diagramed by Figure 6.

MAW's main view is a dynamic map of iconic sources and sinks in a virtual room. A source is a sound emitter; a sink is a sound receptor, a delegate of a human listener in a virtual room. (In a teleconference, an icon might represent both a source and a sink.)

Source→sink directionalization can be performed by a DSP (**d**igital **s**ignal **p**rocessing) module which convolves the digitized input streams with HRTFs (**h**ead-**r**elated **t**ransfer **f**unctions) that capture directional effects [Wen92]. This spatialization enables auditory localization, the identification of the location of a source, which can be used for "the cocktail party effect." The use of such effects might be used in a concert to 'hear out' an instrument, virtually and perceptually pulling it out from the mix, or for sub-caucusing

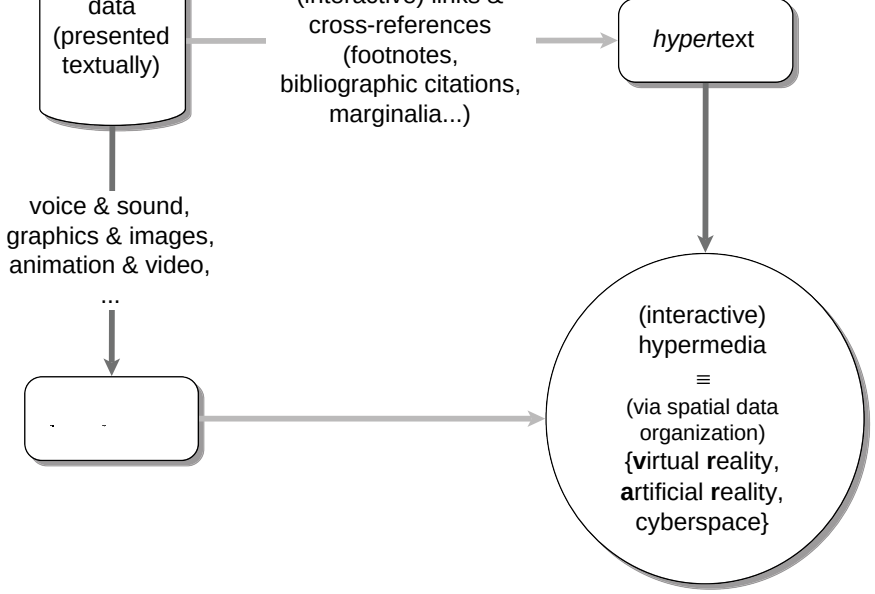

Figure 6: Hypermedia: *hyper*text × multi*media*

	Icon	
	Source:	Sink:
Function	radiator	receiver
Direction	output	input
Instance	speaker (human or loud-)	listener (human or dummy-head)
Include	solo	confide
Exclude	mute/conceal	deafen/blind

Table 1: ${}^{s}OU^{rce}_{Tput}$ and ${}^{s}IN^{k}_{put}$

in a teleconference. Listening to sound presented in this spatial fashion is as different from conventional stereo mixes as sculpture is from painting.

Audio window icons may move around each other and the virtual room. For example, if a sink rotates (exocentrically visually, perhaps driven by a chair tracker [CK92]), the apparent sonic location of the source revolves (egocentrically acoustically) accordingly. The sinks and sources may wander around, like minglers at a cocktail party, or upon the stage during a concert, hovering over the shoulder of a favorite musician. Background music may be brought into the perceptual foreground.

2.3 Shared Perspective: Sink Fusion

Illustrating a one→many mapping of sinks to users (as in broadcasts), [CK91] 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 director and (singleton) audience. (For graphical displays, such inter-user consistency is called "[relaxed] common view," since the various users might zoom or scroll their room windows differently.) This style presentation blurs the distinction between composer, performer, and listener, as hypermedia blurs the distinction between author, publisher, and reader.

2.4 Split Perspective: Sink Fission

Some systems support multiple visual windows, each featuring a different perspective on a scene. In flight simulators, for example, these might be used to display (egocentric) views out cockpit windows, and/or views from a completely different location— high above the airplane, for example, looking down (exocentrically): a virtual 'out-of-body' experience. Since audition is omnidirectional, perhaps audio windows

superimposable. MAW further generalizes multipoint audio perspective by allowing users to fork their presence, as explained below:

2.4.1 Schizophrenia

A simple configuration typically consists of several icons, representing distributed users, moving around a shared space. Each icon represents a source, the voice of the associated user, as well as a sink, that user's ears.

MAW's graphical windows correspond to virtual rooms. Using the `cut`/`paste` idiom as a transporter or 'wormhole,' one may leave a room and beam down into others. Such a control mechanism can be used to focus selectively on various sources. If several rooms were interesting, it would get tiresome to have to bounce back and forth.

Allowing users to designate multiple sinks effectively increases their attendance in conference. A user may simply fork themself (with `copy`/`paste`, for instance), leaving one clone hither while installing another yon, overlaying soundscapes via the superposition of multiple sinks' presence. Such a 'schizophrenic' mode, enabling replicated sinks in same or different conference rooms, explicitly overlays multiple audio displays, allowing a teleconferee to leave a pair of ears in one conversation, while sending other pairs to side caucuses.

This feature can be used to sharpen the granularity of control, 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 multiple instruments in a concert without rearranging the ensemble, which would disturb the soundscape of the icons that personify other users in the shared model.

2.4.2 Autofocus

The apparent paradoxes of one's being in multiple places simultaneously can be resolved by partitioning the sources across the sinks. If the sinks are distributed in separate conference rooms, each source is directionalized only with respect to the sink in the same room. In the case of autothronging, or multiple sinks sharing a single conference room, 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" [Bla83].) Rather than adding or averaging the contribution of each source to the multiple sinks, MAW localizes each source only with respect to the best (loudest, as a function of distance and mutual gain, including focus and orientation) sink.

Figure 7 illustrates this behavior for a top-down view of a conference (top row) with two sinks, represented by icons (distinguished by shared rings), and two different sources, represented by a square and a triangle. In the absence of room acoustics, multiple sinks perceiving a single source is equivalent, via "reciprocity" or symmetry, to a single sink perceiving multiple identical sources. Therefore the exampled scene can be decomposed source-wise into two additive scenes (second row), each single sink combining the parent sinks' (shared) perceptions of the respective sources. These configurations reduce (third row), via 'autofocus' level difference anticipation, to the respective sinks and only the loudest source. The loudest source is typically the closest, since the respective pairs of sources are identical, the chorus of phantom sources being a manifestation of the multiple sinks. Finally (bottom row), the additive scenes are recombined, yielding the overall simplified percept.

Say, for example, that a listener wanted to pay special attention to an ensemble's 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 (both physical and virtual) space. Using MAW, the user could fork themself, as in Figure 8, locating one instance inside the drum, and the other doppelgänger near the rhythm guitar.

2.4.3 Sonic Cubism

The experience of being in multiple places simultaneously, like all virtual situations, may define its own rules. A psychophysical interpretation, as elaborated above, however, is important as an interface strategy,

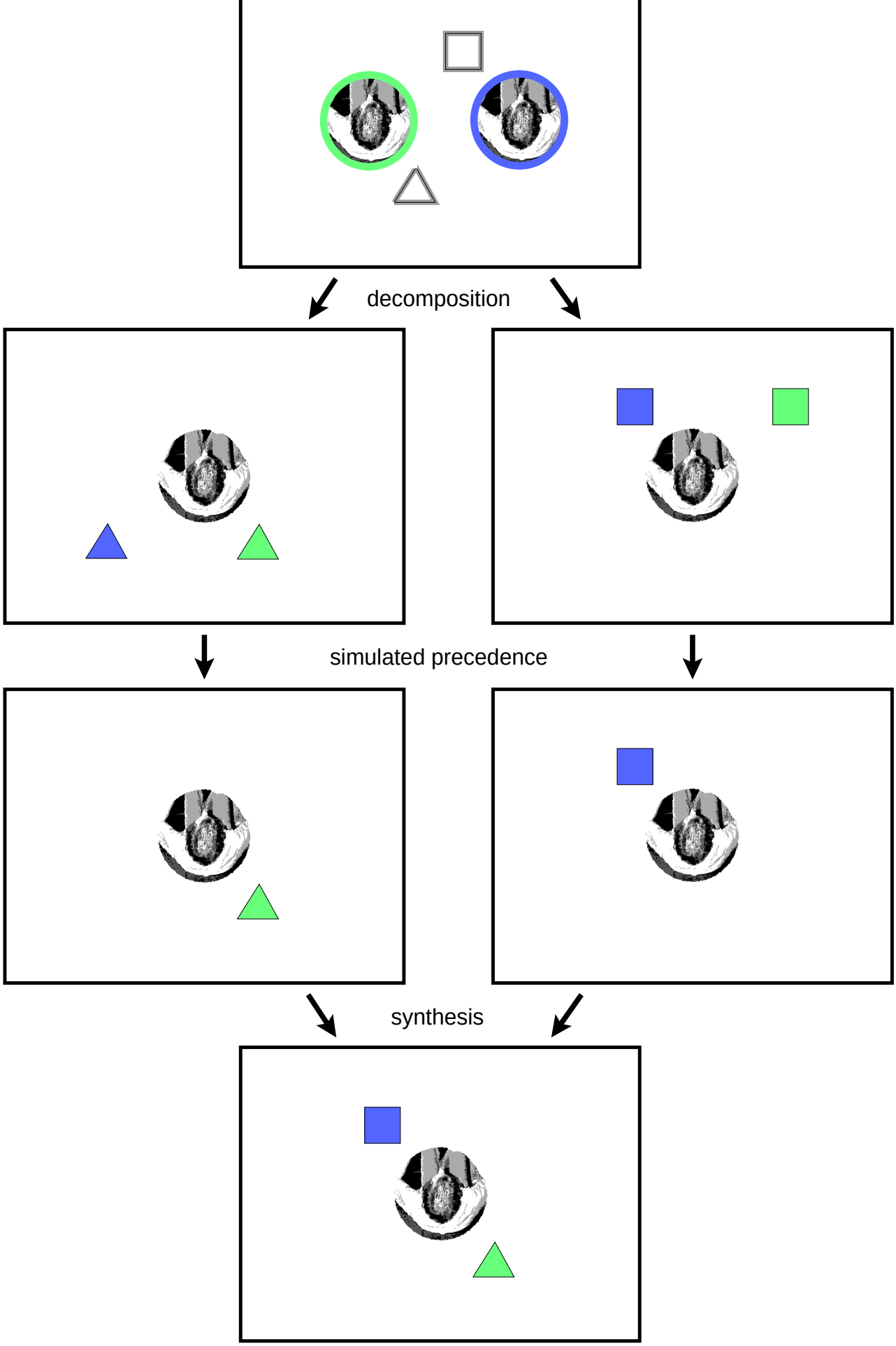

Figure 7: Sonic cubism: schizophrenic mode with autofocus

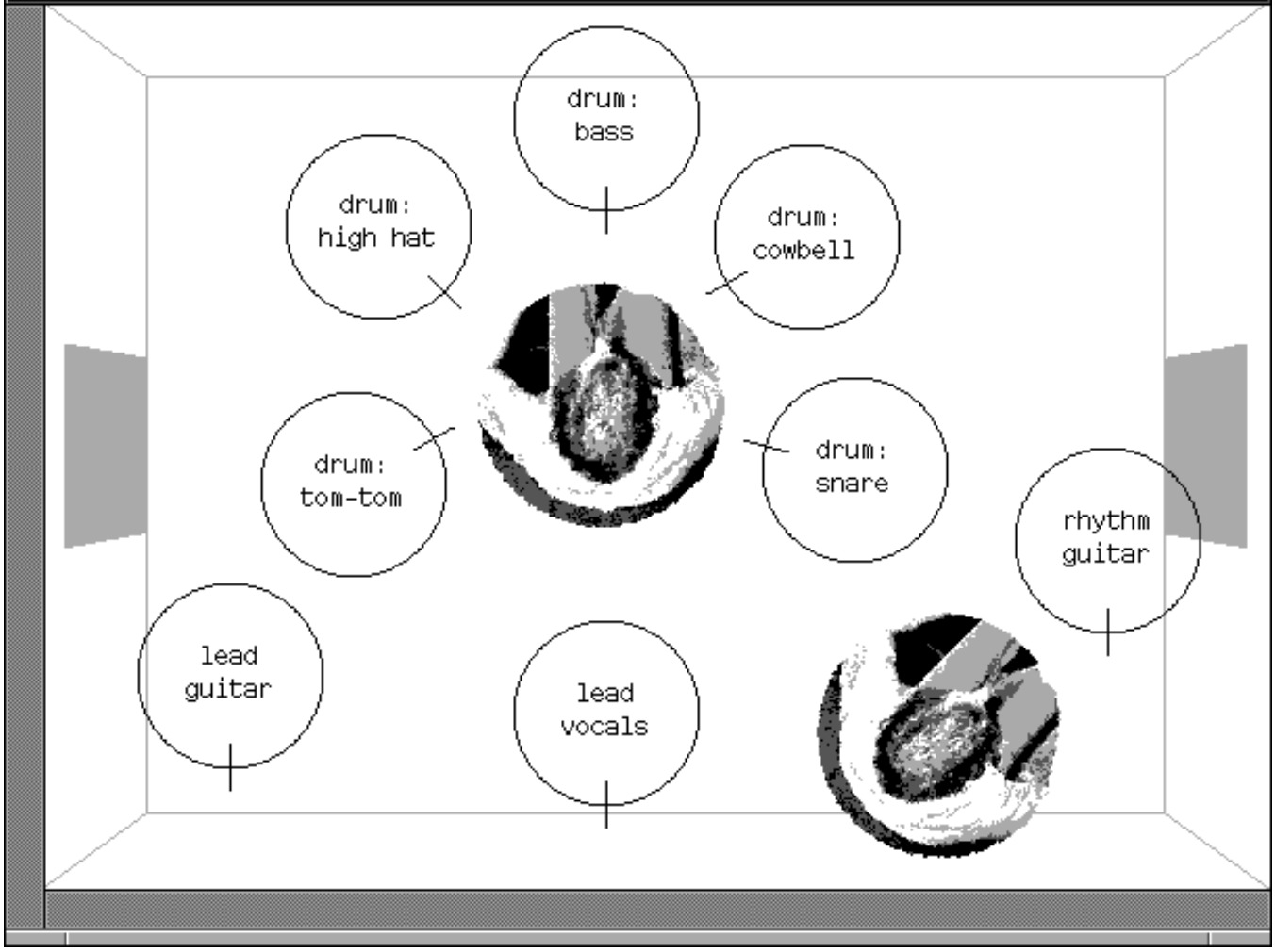

Figure 8: Virtual concert: multiple sinks and exploded clusters (generalized multi-focus audio fish-eye)

making the system behavior consistent with users' intuitions, artificial but accessible. Other schemes are possible, like adding or averaging source→sink transmissions, or disambiguating fancifully by focusing sinks on more distant sources. The overlaid existence suggests the name given to this effect: sonic (analytic) cubism, presenting multiple simultaneous acoustic perspectives. Being anywhere is better than being everywhere, since it is selective; MAW's schizophrenic mode is distilled ubiquity: (groupware-enabled) audition of multiple objects of regard.

2.5 Non-atomic Sinks and Sources: Clusters

Clusters are hierarchically collapsed groups of objects [SZG+96]. MAW features such a cluster utility for organizing spatial sound objects. By bundling multiple channels together (like the drums in Figure 8), a composite timbre is obtained. Clusters have two main purposes:

conservation of spatializer resources Postulating a resource manager, like a switching matrix on either side of the spatial sound processor [CL91] along with dynamic allocation of spatializer channels, a source cluster feature organizes separate input streams that share a single spatializing channel. One application might involve zooming effects. Distant sources would not be displayed; but as it approaches, a cluster would appear as a single point; only to disassociate and distribute spatially as it gets closer. Such variable **l**evel **o**f **d**etail ("LOD") allows navigation in arbitrarily large spaces, assuming bounded density of point sources. Alternatively, with limited spatializing resources, a user might chose to group a subset of the (less important or less pleasant) channels together, piling them in a corner or closet.

logical organization of hierarchical structure In the context of a concert, individually recording (or mic-ing or synthesizing) individual instruments, presenting each of the channels to a binaural direc-

```
concert
  chorus
    soprano
    alto
    tenor
    bass
  orchestra
    strings
      basses
      cellos
      violas
      violins
        G-string
        D-string
        A-string
        E-string
          attack
          decay
            even harmonics
            odd harmonics
    brass
      horns
      trumpets
      trombones
      tuba
    woodwinds
      bassoons
      clarinets
      flutes
      oboes
    percussion
      bass drum
      cymbals
      snare drum
      triangle
      tubular bells
      wood block
      xylophone
      timpani
    other
      harp
      piano
```

Table 2: Concert decomposition

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						

Table 3: Points of View

tional mixing console like MAW, and mixing them at audition time, rather than in "post-production" (as tracks and subgroups), allows the instruments to be rearranged by the listener [SL94]. One could grab an orchestral cluster, for instance (shown as part of the concert in Table 2), explode it to separate different instruments, and drag one of those instruments across the room. Successive differentiation can go right through concert → orchestra → section → instrument and actually break down the instrument itself. Such a superdecomposition aspect of the cluster feature could allow, for example, a user to listen to spatially separate strings of a guitar (assuming a hexaphonic pickup for performance, or decoupled tracks for digital synthesis), or different components of each string's sound. Even more radical decompositions than the partitioning suggested by Table 2 are possible, enabled by advanced workstation musical capability [JB89] and such techniques as physically-based modeling [Yam94]. A generalized approach, ultimately fractal, assumes limitless levels of zooming or analysis.

Atomic sources, the leaves of the tree in Table 2, are called "mixels,"— acronymic for '[sound] **mix**ing **el**ements,' in analogy to pixels, taxels (tactile elements), texels (texture elements), or voxels (volumentric elements, a.k.a. boxels)— since they form the raster across which a soundscape is projected, defining the granularity of control and degree of spatial polyphony. While eventually such decompositions might be dynamically performed, using some equivalent of subtractive synthesis, the current audio window system requires anticipation of the atomization, assuming *a priori* assembly of the finest-grained mixels.

Unclustering can be likened to viewing the sources through a generalized fish-eye lens [Fur86] [SB94] [RPH+95], which spatially warps the perception of the localized sources to enlarge an area of focus and shrink everything else. That is, when the user indicates a direction of special interest, the sources in that direction effectively approach the user and recede from each other in perspective.

3 Grammatical Blur: Beyond Pronouns

An example of a many→many sink:user mapping is a virtual concert in which the audience shares a distribution of sinks: each user hears the same thing, but multiple sinks are used to increase the granularity of audition [CK93] [CK95].

Grammatical constructions like the taxonomy in Table 3 could not anticipate exotic forms of reference, like shared, multiple or reciprocal existence. In an exocentric system, all icons in a dynamic map are potential sensation sinks, and designations associated with pronouns become very fluid. For example, say I choose to think of "my location" in a shared virtual environment as where my voice or instrument comes from, as perceived by some other users. For the purposes of a teleconference or concert, it is philosophical whether the various iconic sinks are thought of as

- multiple manifestations of a singleton ("I" [or lowercase 'i'], or perhaps the Rastafarian "I and I" [DOP]),
- a plural deployment of self ("we," inclusive or exclusive, editorial or royal, ...),
- another user's position ("you" or "thee," singular or plural ["y'all" or "ye"], "he" or "she"),
- a many-eared eavesdropper ("it"), or
- an army of dedicated robot listeners ("they").

Questions about whether or not non-immersive systems are pure 'virtual reality' are really besides the point; what's important is that they enable a computer-enhanced view of the world that is useful and interesting. Such "deconstructions of the body," not in a literary sense, but in a literal sense, as in interfaces developed by [Kru91], relax sink↔user mappings. The extension of an exocentric perspective beyond a multimedia interface is a (possibly multiple or shared) projection of the user into the virtual world. Discussions about workstation-oriented "desktop-" or "fishtank-VR" usually involve issues like cost, constraints on movement, ergonomic engineering (sensor lag, update rates, display resolution), "simulator sickness" [HW92], and user recalibration, but philosophical differences are deeper.

We generally think of our centers of consciousness and perception as residing together, in a single place inside the head attached to our body. But by sidestepping subjectivity of the 1st person, non-immersive systems can augment (instead of simulate) reality. For some applications, an exocentric presentation is more convenient than an egocentric or immersive one. To get a global perspective, for instance, a map is more useful than an immersive display. Down-scaling enables a quicker overview than possible with an immersive world, and humans are quite good at conceptualizing 3-space from projections.

It is important to note that the advantages of non-immersion are not limited to 2D "gods' eye" views. The argument that a map is like a omniscient perspective on an immersive world fails, since the location of the subject, usually thought of as unique, is not above the terrain, but in it. Participatory and experiential $\not\Rightarrow$ inclusive [Rob92] [PBBW95]!

Explicitly distinguishing the domain of the ([virtual] conference, concert, cocktail party) inhabitants from the observing point of view has benefits not afforded by even aerial perspective in immersive systems:

- A user perceives, and can manipulate, themself in the context of the colloquia.
- A user can perceive everyone else in the conference at once. In a groupware environment, others can run but they can't hide. There is no possibility, for instance, of the immersive trick of one user hiding inside another's head. Users might not want to (have to) turn around to see who is approaching from behind.
- Exocentric interfaces allow multipoint audio perspectives. It is hard to imagine how multiple instances of self might be implemented effectively in an immersive system.

Audio window metaphors apply to full 3D graphical interfaces and earprints (HRTFs) as well. Rather than encase the user in a HMD and glove/wand configuration, we can augment the telephone and stereo, using the computer as a map. Using such a full spatial model, music can be we spatialized according to a helical structure of scale [She84] [She83]. The harmony and melody of a song can be perceived by separate sinks, using the audio cubism idiom to normalize the octave, as suggested by Figure 9 [HC96].

Such schizophrenic modes can be thought of as forking reality, rather than cloning self. The perception of telepresence is auto-empathy, imagining how oneself would feel elsewhere. New interaction modalities are enabled by this sort of perceptual aggression and liquid perspective, as style catches up with technology.

Acknowledgments

This research has been supported in part by grants from NTT Human Interface Laboratories and the Fukushima Prefectural Foundation for the Advancement of Science and Education.

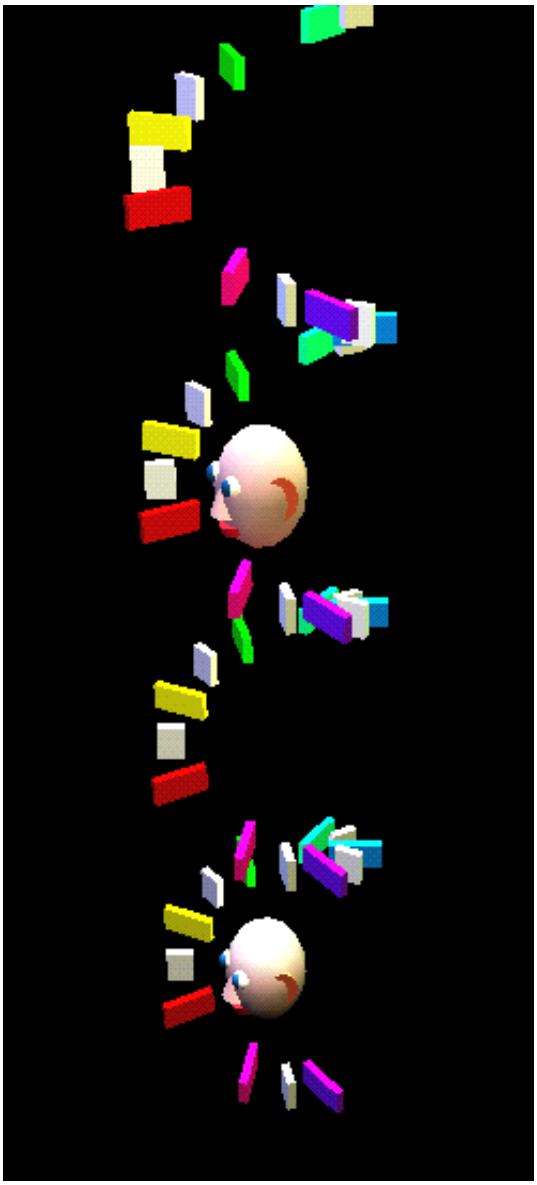

Figure 9: Octave normalized by separate sinks for harmony and melody

[Bec92] Allen Becker. High resolution virtual displays. In *Proc. ICAT:* **I***nt.* **C***onf.* **A***rtificial Reality and* **T***ele-Existence*, pages 27–34, 1992.

[Bla83] Jens Blauert. *Spatial Hearing: The Psychophysics of Human Sound Localization.* MIT Press, 1983. ISBN 0-262-02190-0.

[Bre90] Albert S. Bregman. *Auditory Scene Analysis: The Perceptual Organization of Sound.* MIT Press, 1990. ISBN 0-262-02297-4.

[CAK93] Michael Cohen, Shigeaki Aoki, and Nobuo Koizumi. Augmented audio reality: Telepresence/VR hybrid acoustic environments. In *Proc. Ro-Man:* 2^{nd} IEEE *Int. Workshop on* **Ro***bot and Hu***man** *Communication*, pages 361–364, Tokyo, November 1993. ISBN 0-7803-1407-7.

[CK91] Michael Cohen and Nobuo Koizumi. Audio window. In *Den Gaku.* Tokyo Contemporary Music Festival: Music for Computer, December 1991.

[CK92] Michael Cohen and Nobuo Koizumi. Iconic control for audio windows. In *Proc. Eighth Symp. on Human Interface*, pages 333–340, Kawasaki, Japan, October 1992. SICE (**S**ociety of **I**nstrument and **C**ontrol **E**ngineers). 1411.

[CK93] Michael Cohen and Nobuo Koizumi. Audio windows for virtual concerts. In *Proc. JMACS:* **J***apan* **M***usic* **A***nd* **C***omputer Science* **S***ociety Meeting*, pages 27–32, Tokyo, February 1993. No. 47.

[CK94] Michael Cohen and Nobuo Koizumi. Putting spatial sound into voicemail. In *NR94: Proc. 1st International Workshop on* **N***etworked* **R***eality in TeleCommunication*, Tokyo, May 1994. IEEE COMSOC, IEICE. Session 1-2.

[CK95] Michael Cohen and Nobuo Koizumi. Audio Windows for Virtual Concerts II: Sonic Cubism. In Susumu Tachi, editor, *Video Proc. ICAT/VRST:* **I***nt.* **C***onf.* **A***rtificial Reality and* **T***ele-Existence/Conf. on* **V***irtual* **R***eality* **S***oftware and* **T***echnology*, page 254, Makuhari, Chiba; Japan, November 1995. ACM-SIGCHI (TBD), SICE (**S**ociety of **I**nstrument and **C**ontrol **E**ngineers), JTTAS (**J**apan **T**echnology **T**ransfer **A**ssociation), and NIKKEI (**Ni**hon **Kei**zai Shimbun, Inc.).

[CL91] Michael Cohen and Lester F. Ludwig. Multidimensional audio window management. *IJMMS: the Journal of Person-Computer Interaction*, 34(3):319–336, March 1991. Special Issue on Computer Supported Cooperative Work and Groupware. ISSN 0020-7373.

[CM92] T. P. Caudell and David W. Mizell. Augmented reality: An application of heads-up display technology to manual manufacturing processes. In *Proc.* **H***awaii* **I***nt.* **C***onf. on* **S***ystems* **S***ciences.* IEEE, January 1992.

[Coh93a] Michael Cohen. Integrating graphical and audio windows. *Presence: Teleoperators and Virtual Environments*, 1(4):468–481, Fall 1993. ISSN 1054-7460.

[Coh93b] Michael Cohen. Throwing, pitching, and catching sound: Audio windowing models and modes. *IJMMS: the Journal of Person-Computer Interaction*, 39(2):269–304, August 1993. ISSN 0020-7373.

[Coh94] Michael Cohen. Cybertokyo: a survey of public VRtractions. *Presence: Teleoperators and Virtual Environments (Special Issue on Pacific Rim Research)*, 3(1):87–93, Winter 1994. ISSN 1054-7460.

[Coh95] Michael Cohen. Besides immersion: Overlaid points of view and frames of reference; using audio windows to analyze audio scenes. In Susumu Tachi, editor, *Proc. ICAT/VRST:* **I***nt.* **C***onf.* **A***rtificial Reality and* **T***ele-Existence/Conf. on* **V***irtual* **R***eality* **S***oftware and* **T***echnology*, Makuhari, Chiba, Japan, November 1995.

http://www.willamette.edu/~tjones/languages/rasta-lang.html.

[Fur86] George W. Furnas. Generalized fisheye views. In *Proc. CHI:* ACM *Conf. on* **C***omputer-***H***uman* **I***nteraction*, pages 16–23, Boston, April 1986.

[HC96] Jens Herder and Michael Cohen. Project report: Design of a helical keyboard. In *Proc. ICAD:* **I***nt.* **C***onf.* **A***uditory* **D***isplay*, pages 139–142, Palo Alto, CA, November 1996. www.santafe.edu/~icad/ICAD96/proc96/herder.htm.

[HD92] Richard M. Held and Nathaniel I. Durlach. Telepresence. *Presence: Teleoperators and Virtual Environments*, 1(1):109–112, 1992. ISSN 1054-7460.

[HW92] Lawrence J. Hettinger and Robert B. Welch. Presence: Teleoperators and virtual environments, Summer 1992. ISSN 1054-7460.

[HY92] Michitaka Hirose and Kensuke Yokoyama. VR Application for Transmission of Synthetic Sensation. In Susumu Tachi, editor, *Proc. ICAT:* **I***nt.* **C***onf.* **A***rtificial Reality and* **T***ele-Existence*, pages 145–154, Tokyo, July 1992.

[IKG93] Hiroshi Ishii, Minoru Kobayashi, and Jonathan Grudin. Integration of inter-personal space and shared workspace: Clearboard design and experiments. *TOIS:* ACM **T***rans. on* **I***nformation* **S***ystems (Special Issue on* CSCW *'92)*, 11(4):349–375, July 1993.

[Ish92] Hiroshi Ishii. Translucent multiuser interface for realtime collaboration. *IEICE Trans. on Fundamentals of Electronics, Communications and Computer Sciences* (Special Section on Fundamentals of Next Generation Human Interface), E75-A(2):122–131, February 1992. 0916-8508.

[JB89] David Jaffe and Lee Boynton. An Overview of the Sound and Music Kits for the NeXT Computer. *Computer Music Journal*, 13(2):48–55, Summer 1989.

[Kru91] Myron W. Krueger. *Artificial Reality II.* Addison-Wesley, Reading, MA, 1991. 0-201-52260-8.

[Nai91] Michael Naimark. Elements of realspace imaging: A proposed taxonomy. In *Proc. SPIE/SPSE Electronic Imaging Conf.*, San Jose, CA, 1991. Vol. 1457.

[OTT92] Eimei Oyama, Naoki Tsunemoto, and Susumu Tachi. Remote manipulation using virtual environment. In *Proc. ISMCR:* **I***nt.* **S***ymp. on* **M***easurement and* **C***ontrol in* **R***obotics*, pages 311–318, Tsukuba Science City, Japan, November 1992. SICE (**S**ociety of **I**nstrument and **C**ontrol **E**ngineers).

[PBBW95] Randy Pausch, Tommy Burnette, Dan Brockway, and Michael E. Weiblen. Locomotion in virtual worlds via flight into hand-held miniatures. In SIGGRAPH *Proc.*, LA, CA, July 1995.

[Rob92] Warren Robinett. Synthetic experience: A proposed taxonomy. *Presence: Teleoperators and Virtual Environments*, 1(2):229–247, 1992. ISSN 1054-7460.

[RPH+95] Ramana Rao, Jan O. Pedersen, Marti A. Hearst, Jock D. Mackinlay, Stuart K. Card, Larry Masinter, Per-Kristian Halvorsen, and George G. Robertson. Rich interaction in the digital library. *Communications of the* ACM, 38(4):29, April 1995.

[SB94] Manojit Sarkar and Marc H. Brown. Graphical fisheye views. *Communications of the* ACM, 37(12):73–84, December 1994.

[She83] Roger N. Shepard. Demonstrations of circular components of pitch. *J.* **A***ud.* **E***ng.* **S***oc.*, 31(9):641–649, September 1983.

[She84] Roger N. Shepard. Structural representations of musical pitch. In D. Deutsch, editor, *The Psychology of Music*, pages 343–390. Academic Press, 1984. ISBN 0-12-213560-1 or 0-12-213562-8.

[She92a] Thomas B. Sheridan. Musings on telepresence and virtual presence. *Presence: Teleoperators and Virtual Environments*, 1(1):120–125, 1992. ISSN 1054-7460.

Symp. on **M**easurement and **C**ontrol in **R**obotics, pages 19–29, Tsukuba Science City, Japan, November 1992. SICE (**S**ociety of **I**nstrument and **C**ontrol **E**ngineers).

[SK92] Gen Suzuki and Takashi Kouno. Virtual collaborative workspace. In *Proc. ICAT:* **I***nt.* **C***onf.* **A***rtificial Reality and* **T***ele-Existence*, pages 119–125, 1992.

[SL94] Gavin R. Starks and Ken N. Linton. A 3-D Stereo Processing Tool. In *Proc. AES:* **A***udio* **E***ngineering* **S***ociety Conv.*, Amsterdam, February 1994. 3830 (P9.1).

[Sta93] Julie Stanfel. Mandala: Virtual cities. In SIGGRAPH *Computer Graphics Visual Proc.*, page 208, Anaheim, CA, August 1993.

[SZG+96] Doug Schaffer, Zhengping Zuo, Saul Greenberg, John Dill, Shelli Dubs, Mark Roseman, and Lyn Bartram. Navigating hierarchically clustered networks through fisheye and full-zoom methods. *TOIS:* ACM **T***rans. on* **I***nformation* **S***ystems*, January 1996.

[TAM+91] Susumu Tachi, Hirohiko Arai, Taro Maeda, Eimei Oyama, Naoki Tsunemoto, and Yasuyuki Inoue. Tele-existence in real world and virtual world. In *ICAR: Proc. Fifth Int. Conf. on Advanced Robotics*, pages 193–198, Pisa, Italy, June 1991.

[Wen92] Elizabeth M. Wenzel. Localization in virtual acoustic displays. *Presence: Teleoperators and Virtual Environments*, 1(1):80–107, 1992. ISSN 1054-7460.

[WMG93] Pierre Wellner, Wendy Mackay, and Rich Gold. Communications of the ACM, July 1993.

[WV93] Susan Wyshynski and Vincent John Vincent. Full-body unencumbered immersion in virtual worlds. In Alan Wexelblat, editor, *Virtual Reality: Applications and Explorations*, chapter 6, pages 123–144. Academic Press, 1993. ISBN 0-12-745046-7.

[Yam94] Yamaha Corp. *VP1 Virtual Acoustic Synthesizer*, 1994.

[ZPF+94] Michael J. Zyda, David R. Pratt, John S. Falby, Chuck Lombardo, and Kristen M. Kelleher. The software required for the computer generation of virtual environments. *Presence: Teleoperators and Virtual Environments*, 2(2):130–140, 1994. ISSN 1054-7460.