

Besides Immersion: Overlaid Points of View and Frames of Reference — Using Audio Windows to Analyze Audio Scenes

MICHAEL COHEN

Human Interface Lab, University of Aizu

Aizu-Wakamatsu; Fukushima-ken 965-80; 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 multidimensional audio windows) 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] mixing elements. 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. [Note: Some of the ideas presented in the paper are animated by a two-part video in the ICAT/VRST95 video proceedings: “Audio Windows for Virtual Concerts I/II (Sonic Cubism)”]

Keywords and Phrases: binaural directional mixing console, CSCW (computer-supported collaborative work), groupware, mixel ([sound] mixing element), sonic (analytical) cubism, sound localization, spatial sound.

0 Introduction

“Traditional” immersive VR systems feature a HMD (head-mounted display) 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 control/display conventions are needed that relax the mapping between user and presence, especially as applied to desktop or ‘fishtank’ VR systems. This paper explores the philosophical distinction between egocentricism and exocentricism, especially as blurred by emerging technolo-

gies.

1 Non-duality 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 themselves 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 themselves 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.
- A subject in a spatial sound experiment, presented with a stereo signal simulating a directionalized 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 themselves 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.
- A networked racing simulator arcade game 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;

float, in which the camera position is well above the car, still orienting 'up' on the display with '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.

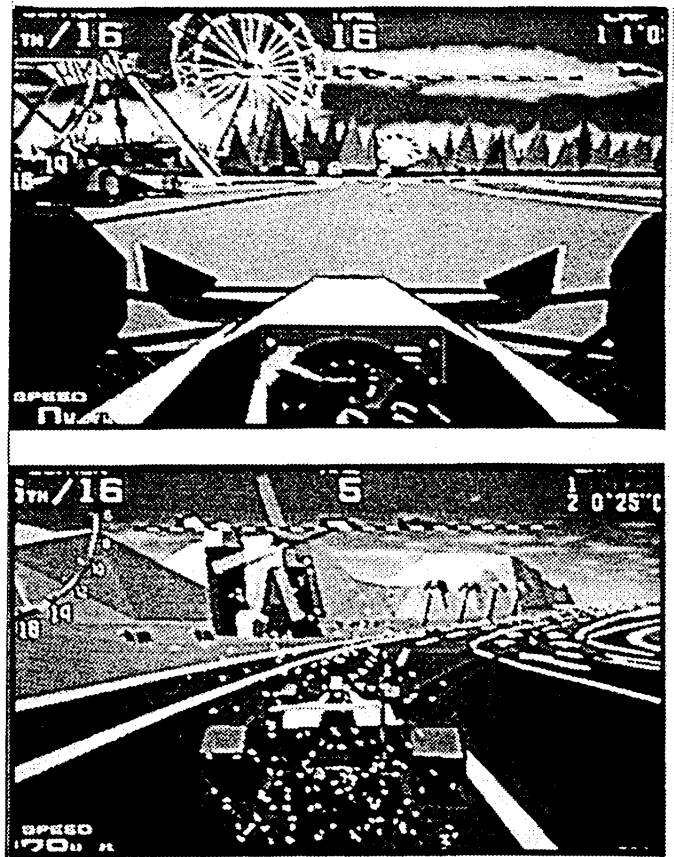


Figure 1: Sega Virtua Racing

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 computer graphic raceway. Further blurring the sampled/synthesized distinction, separate monitors for spectators can show live video of the drivers, panning shots of the lead car, static shots of strategic curves, and instant replays of crashes [Cohen, 1994].

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 [Held and Durlach, 1993] [Sheridan, 1993]; 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.



Figure 2: Hirose Lab Virtual Dome

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 [Sheridan, 1992]. 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

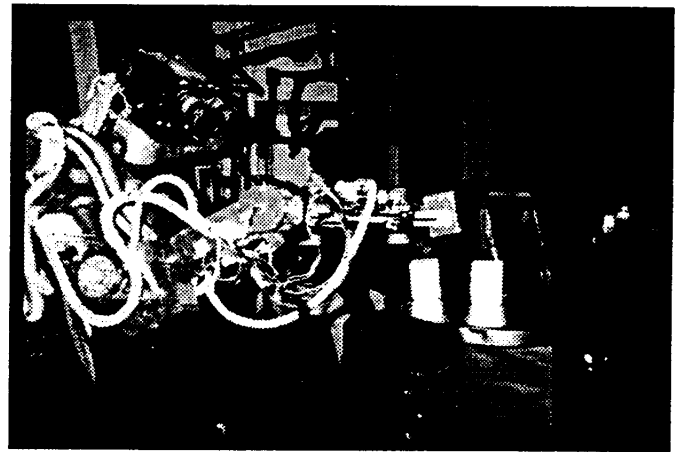


Figure 3: RCAST Telerobot

is interesting; share it with me...” Mass broadcast media like radio and TV employ this one→many mode. Of course they lack the control of VR systems, but interactive television (suggested by the zapping movie “Murderous Intentions,” whose simultaneous parallel broadcast of two characters’ stories allowed 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 non-overlappingly 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 (head-up displays) are used in airplanes to throw navigation, tracking, and status information onto the windshield. Half-silvered mirrors can be used to image translucent images. Clearboard [Ishii, 1992] [Ishii et al., 1993] uses superimposed translucent viewing planes for teleconferencing with video of the conferees plus a shared whiteboard; different focal distances can help distinguish the layers. [Oyama et al., 1992] superimposed real and virtual images by using a

virtual image as a mask for a real, or by rendering a virtual image as a (non-occluding) wireframe. "Mirror-type" VR systems like Mandala [Wyshynski and Vincent, 1993] (Figure 4) can combine CG and (chroma-key captured) sampled signals, overlaid on arbitrary background graphics.

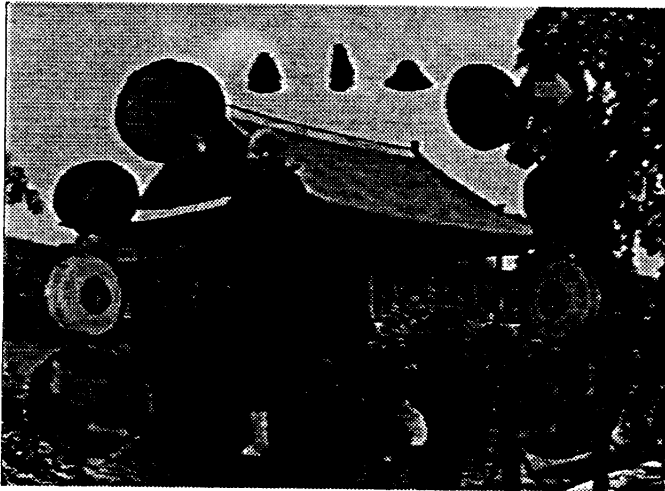


Figure 4: Vivid Mandala

"Augmented reality" describes the superposition of computer-generated imagery on top of a see-through display [Caudell and Mizell, 1992] [Wellner et al., 1993]. The dual of augmented reality is augmented virtual reality, which might be manifested in the video domain by compositing camera-captured images on otherwise synthetic CG buffers.

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* [Becker, 1992] or *ScopeHand* [Suzuki and Kouno, 1992]), 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 presented, by presenting them to different hands.¹ Similarly, dichotic presentation

¹This 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.

involves simultaneous presentation of separate audio scenes to each ear. More directly, an arbitrary number of audio channels may be simply added and presented 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 (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 [Cohen et al., 1993]. For instance, the author has perceptually thrown a ringing sound to a location occupied by a muted telephone, recalling [Naimark, 1991]'s visual analog of projecting a picture of a room on the same space after it was painted white. Public address, 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. Implemented as a NextStep front-end, MAW is suitable for synchronous applications like teleconferences or concerts, as well as asynchronous applications like voicemail and hypermedia [Cohen, 1993] [Cohen and Koizumi, 1994], which can be thought of as equivalent (because of spatial data models) to cyberspace [Zyda et al., 1994], as illustrated by Figure 5.

MAW's main view is a dynamic map of iconic sources and sinks in a virtual room. A source

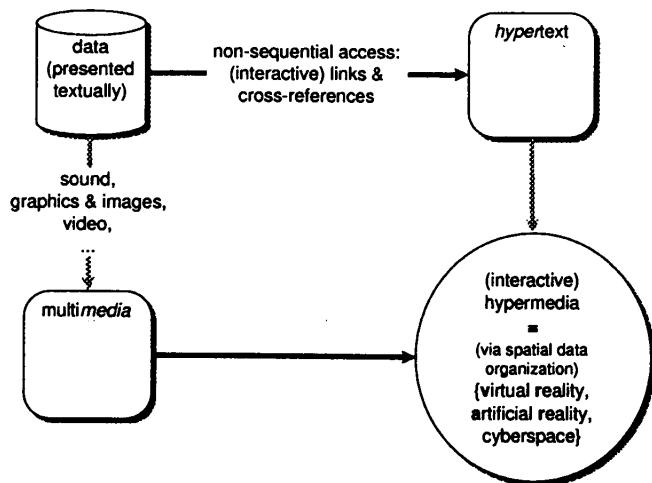


Figure 5: Hypermedia: *hypertext* × *multimedia*

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:
output	input
speaker (human or loud-)	listener (human or dummy-head)
radiator	receiver

Table 1: ${}^sOU_{Tput}^{rce}$ and ${}^sIN_{put}^{nk}$

Source→sink directionalization can be performed by a DSP (digital signal processing) module which convolves the digitized input streams with HRTFs (head-related transfer functions) that capture directional effects [Wenzel, 1992]. 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 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 the sink rotates (exocentrically visually), 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), [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 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 can be thought of as implicitly providing this multiperspective capability, audio sources being inherently superimposable. MAW further generalized 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.

MAW allows users to designate multiple sinks, effectively increasing their attendance in conference. A user may simply fork themselves (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, en-

abling 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 are 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, 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" [Blauert, 1983].) 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 6 illustrates this behavior for a top-down view of a conference (top row) with two sinks, represented by identical icons, 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 exemplified scene can be decomposed source-wise into two additive scenes (second row), each single sink combining the parent sinks' perceptions of the respective sources. These configurations reduce (third row), via the 'autofocus' level difference anticipation, to the respective sinks and only the loudest source. The loudest source is actually the closest, since the respective pairs of sources are identical, the chorus of phantom sources being a manifestation of the multiple sinks. Finally

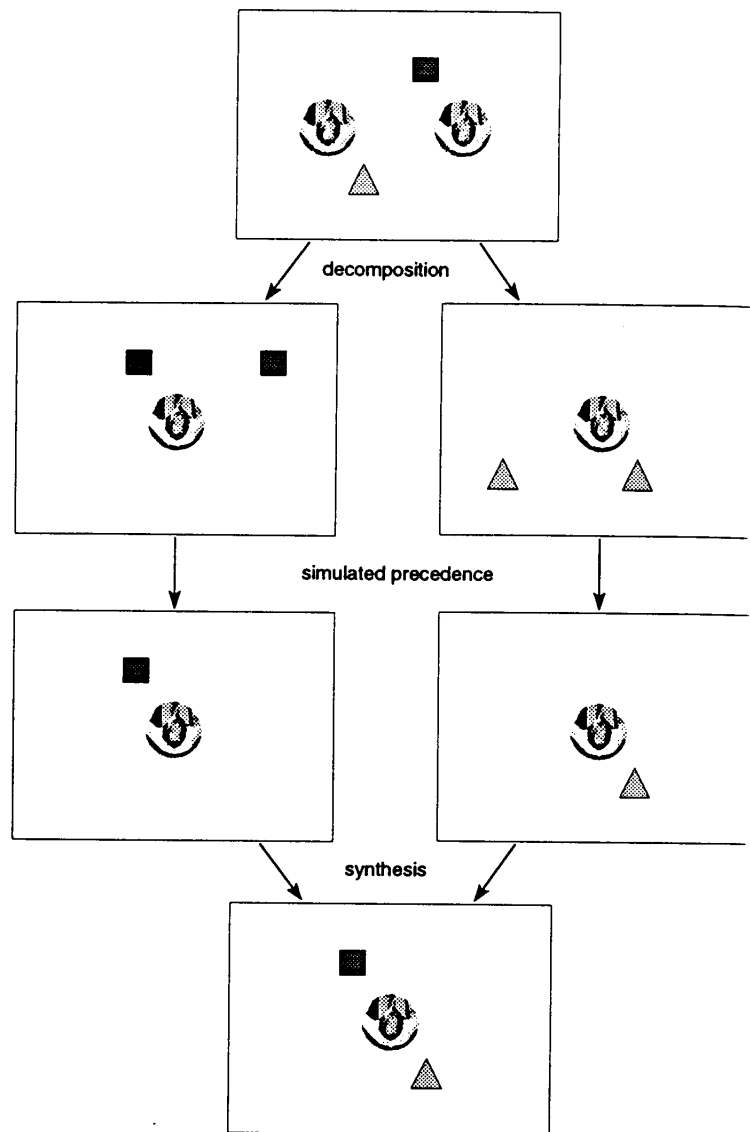


Figure 6: Sonic cubism: schizophrenic mode with autofocus

(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 would fork themselves, as in Figure 7, 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 interpreta-

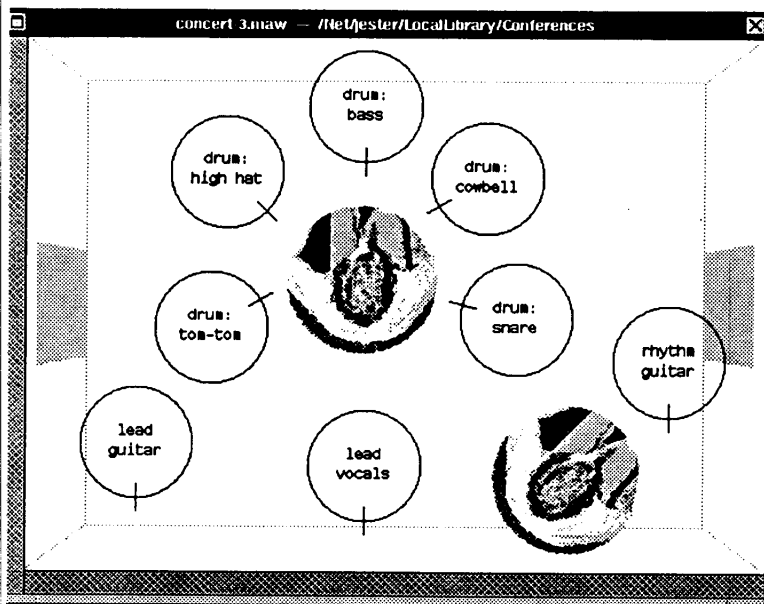


Figure 7: Virtual concert: multiple sinks and exploded clusters (generalized multi-focus audio fish-eye)

tion, as elaborated above, however, is important as an interface strategy, making the system behavior consistent with users' intuitions, artificial but accessible. (A different implementation might choose a more fanciful disambiguation scheme.) 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.

2.4.4 Non-atomic Sinks and Sources: Clusters

Clusters are hierarchically collapsed groups of objects [Schaffer et al., 1996]. MAW features such a cluster utility for organizing spatial sound objects. By bundling multiple channels together (like the different drums in Figure 7), a composite timbre is obtained. Clusters have two main purposes:

conservation of spatializer resources

Postulating a switching matrix on either side of the spatial sound processor [Cohen and Ludwig, 1991], 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 sin-

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

gle point; only to disassociate and distribute spatially as it gets closer. This variable level of detail ("LOD") allows navigation in arbitrarily large space, assuming a limited 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, stacking 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 directional 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 [Starks and Linton, 1994]. One could grab onto an orchestral cluster, for instance (shown as part of the concert in Table 2), explode it to separate the different instruments, grab one of those instruments and move it across the room. This successive differentiation could 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, the 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 [Jaffe and Boynton, 1989] and such techniques as physically-based modeling [Yamaha, 1994]. A generalized approach, ultimately fractal, assumes that there would always be levels of zooming or analysis.

The atomic sources, the leaves of the tree in Table 2, are called "mixels," acronymic for '[sound] mixing elements,' in analogy to pixels, taxels (tactile elements) or voxels (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 partitioned, using some equivalent of subtractive synthesis, the current audio window system requires

anticipation of the atomization, assuming a *a priori* assembly of the finest-grained mixels.

Unclustering can be likened to viewing the sources through a generalized fish eye lens [Furnas, 1986] [Sarkar and Brown, 1994], 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 [Cohen and Koizumi, 1993].

Grammatical constructions could not anticipate exotic forms of reference, like shared, multiple or reciprocal existence. In an exocentric VR system, all the icons in the 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 quite irrelevant (if perhaps philosophical) whether the various iconic sinks are thought of as

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

4 Conclusion

The audio window metaphor applies to full 3D graphical interfaces and 3D 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, we are

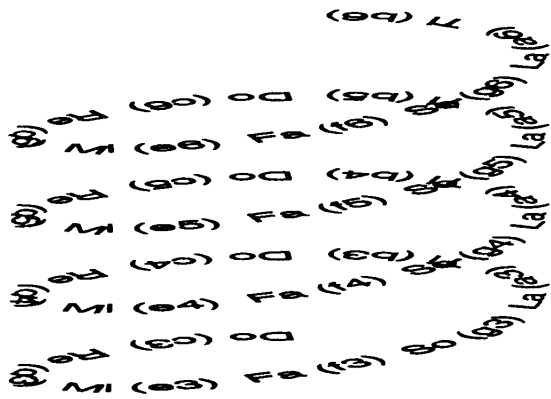


Figure 8: Helical structure of scale

working on spatializing music according to a helical structure of scale [Shepard, 1982] [Shepard, 1983], as shown in Figure 8. The harmony and melody will be perceived by separate sinks, using the audio cubism idiom to normalize the octave, as shown by Figure 9.

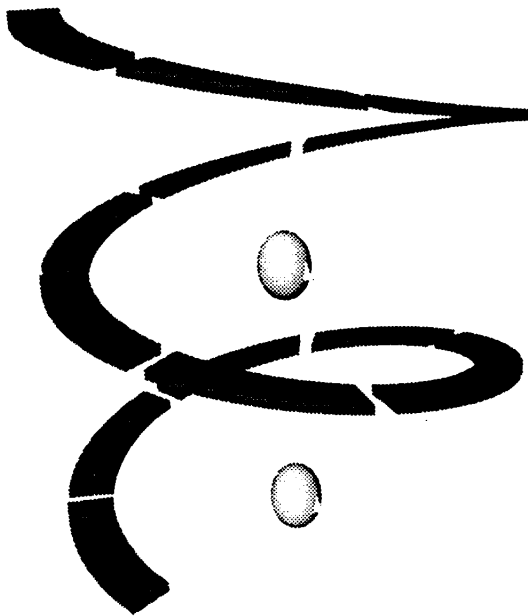


Figure 9: Octave normalized by separate sinks for harmony and melody

MAW's schizophrenic mode can be thought of as forking reality, rather than cloning self. The perception of telepresence, especially forked, 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.

Acknowledgements

The helical keyboard is being developed as a student research project led by Jens Herder. The audio windows research is supported in part by a grant from the NTT Human Interface Labs.

References

- Becker, A. [1992] High resolution virtual displays. In *ICAT: Proc. Int. Conf. Artificial Reality and Tele-Existence*, pages 27-34.
- Blauert, J. [1983] *Spatial Hearing: The Psychophysics of Human Sound Localization*. MIT Press. ISBN 0-262-02190-0.
- Bregman, A. S. [1990] *Auditory Scene Analysis: The Perceptual Organization of Sound*. MIT Press. ISBN 0-262-02297-4.
- Caudell, T. P. and Mizell, D. W. [1992] Augmented reality: An application of heads-up display technology to manual manufacturing processes. In *HICSS: Proc. Hawaii Int. Conf. on System Sciences*. IEEE.
- Cohen, M. [1993] Integrating graphical and audio windows. *Presence: Teleoperators and Virtual Environments*, 1(4):468-481.
- Cohen, M. [1994] Cybertokyo: a survey of public VRtractions. *Presence: Teleoperators and Virtual Environments (Special Issue on Pacific Rim Research)*, 3(1):87-93.
- Cohen, M., Aoki, S., and Koizumi, N. [1993] Augmented audio reality: Telepresence/VR hybrid acoustic environments. In *Ro-Man: Proc. 2nd IEEE International Workshop on Robot and Human Communication*, pages 361-364, Tokyo. ISBN 0-7803-1407-7.
- Cohen, M. and Koizumi, N. [1991] Audio window. In *Den Gaku*. Tokyo Contemporary Music Festival: Music for Computer.
- Cohen, M. and Koizumi, N. [1993] Audio windows for virtual concerts. In *JMACS: Proc. Japan Music And Computer Science Society Meeting*, pages 27-32, Tokyo. No. 47.
- Cohen, M. and Koizumi, N. [1994] Putting spatial sound into voicemail. In *NR94: Proc. 1st International Workshop on Networked Reality in TeleCommunication*, Tokyo. IEEE COMSOC, IE-ICE.
- Cohen, M. and Ludwig, L. F. [1991] Multidimensional audio window management. *IJMMS: the Journal*

- of *Person-Computer Interaction*, 34(3):319-336. Special Issue on Computer Supported Cooperative Work and Groupware. 0020-7373.
- Furnas, G. W. [1986] Generalized fisheye views. In *CHI: Proc. ACM Conf. on Computer-Human Interaction*, pages 16-23, Boston.
- Held, R. M. and Durlach, N. I. [1993] Telepresence. *Presence: Teleoperators and Virtual Environments*, 1(1):109-112.
- Hirose, M. and Yokoyama, K. [1992] VR Application for Transmission of Synthetic Sensation. In Tachi, S., editor, *ICAT: Proc. Int. Conf. Artificial Reality and Tele-Existence*, pages 145-154, Tokyo.
- Ishii, H. [1992] 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. 0916-8508.
- Ishii, H., Kobayashi, M., and Grudin, J. [1993] Integration of inter-personal space and shared workspace: Clearboard design and experiments. *TOIS: ACM Trans. on Information Systems* (Special Issue on CSCW '92), 11(4):349-375.
- Jaffe, D. and Boynton, L. [1989] An Overview of the Sound and Music Kits for the NeXT Computer. *Computer Music Journal*, 13(2):48-55.
- Naimark, M. [1991] Elements of realspace imaging: A proposed taxonomy. In *Proc. SPIE/SPSE Electronic Imaging Conf.*, San Jose, CA. Vol. 1457.
- Oyama, E., Tsunemoto, N., and Tachi, S. [1992] Remote manipulation using virtual environment. In *ISMCR: Proc. Int. Symp. on Measurement and Control in Robotics*, pages 311-318, Tsukuba Science City, Japan. Society of Instrument and Control Engineers.
- Sarkar, M. and Brown, M. H. [1994] Graphical fisheye views. *Communications of the ACM*, 37(12):73-84.
- Schaffer, D., Zuo, Z., Greenberg, S., Dill, J., Dubs, S., Roseman, M., and Bartram, L. [1996] Navigating hierarchically clustered networks through fisheye and full-zoom methods. *TOIS: ACM Trans. on Information Systems*.
- Shepard, R. N. [1982] *The Psychology of Music*, chapter Structural Representations of Musical Pitch, pages 343-390. Academic Press.
- Shepard, R. N. [1983] Demonstrations of circular components of pitch. *J. Aud. Eng. Soc.*, 31(9):641-649.
- Sheridan, T. B. [1992] Progress and challenge in human telerobot control. In *ISMCR: Proc. Int. Symp. on Measurement and Control in Robotics*, pages 19-29, Tsukuba Science City, Japan. Society of Instrument and Control Engineers.
- Sheridan, T. B. [1993] Musings on telepresence and virtual presence. *Presence: Teleoperators and Virtual Environments*, 1(1):120-125.
- Stanfel, J. [1993] Mandala: Virtual cities. In *SIGGRAPH Computer Graphics Visual Proc.*, page 208, Anaheim, CA.
- Starks, G. R. and Linton, K. N. [1994] A 3-D Stereo Processing Tool. In *AES: Proc. Audio Engineering Society Conv.*, Amsterdam. 3830 (P9.1).
- Suzuki, G. and Kouno, T. [1992] Virtual collaborative workspace. In *ICAT: Proc. Int. Conf. Artificial Reality and Tele-Existence*, pages 119-125.
- Tachi, S., Arai, H., Maeda, T., Oyama, E., Tsunemoto, N., and Inoue, Y. [1991] Teleexistence in real world and virtual world. In *ICAR: Proc. Fifth Int. Conf. on Advanced Robotics*, pages 193-198, Pisa, Italy.
- Wellner, P., Mackay, W., and Gold, R. [1993] Communications of the ACM.
- Wenzel, E. M. [1992] Localization in virtual acoustic displays. *Presence: Teleoperators and Virtual Environments*, 1(1):80-107.
- Wyshynski, S. and Vincent, V. J. [1993] Full-body unencumbered immersion in virtual worlds. In Wexelblat, A., editor, *Virtual Reality: Applications and Explorations*, chapter 6, pages 123-144. Academic Press. ISBN 0-12-745046-7.
- Yamaha [1994] *VP1 Virtual Acoustic Synthesizer*. Yamaha Corp.
- Zyda, M. J., Pratt, D. R., Falby, J. S., Lombardo, C., and Kelleher, K. M. [1994] The software required for the computer generation of virtual environments. *Presence: Teleoperators and Virtual Environments*, 2(2):130-140.