Conferences, Concerts, and Cocktail Parties: Besides Immersion

公園マイケル
会津大学, ヒューマン インタフェース 講座
電話: (0242)37-2537; ファクス: (0242)37-2549
電子メール: mcohen@u-aizu.ac.jp

コンサート、および、電話会議などの音響受聴の場面で、カクテルパーティー効果の実現など、単なる音量調整以上の高度な知覚的制御が容易にできれば新しい有効な音響システムとなる。こうしたシステムとして Maw を提案して検討を続けている。MAW とは multidimensional audio windows の略であり、仮想空間中で、リスナーが音源をアイコンによって操作するためのインタフェースである。MAW は NeXT 上で動作し、ボイスメールや、ハイパーメディアだけでなく、上記のコンサート、電話会議などの場面にリアルタイムメディアとして利用できる。この空間音響を考慮して表現された音は、従来のステレオミックスに比べると、まるで彫刻と絵ほどの違いがある。また、アイコンにより、仮想空間の中でリスナー、音源とも移動、回転などの操作が行なえる。さらに、コピー／ペースト機能を用いて音源の重ね合わせ、リスナーのクローンの重ね合わせなどが可能となる。これにより、同じ人間が異なった場所でそれぞれ違った音源をモニターでき、微妙な知覚的な制御が可能となる。これをここでは Pixel をもじって、mixel (sound mixing element) と呼び、これらの実現方法と効果について述べる。

Conferences, Concerts, and Cocktail Parties:
Besides Immersion

Michael Cohen
Human Interface Lab
University of Aizu 965-80
voice: [+81](242)37-2537; fax: [+81](242)37-2549
internet: McOhen@u-aizu.ac.jp (NeXTmail capable)

Alternative non-immersive perspectives enable new paradigms of perception, especially in the context of frames-of-reference for musical audition or groupware. MAW, acronymic for multidimensional audio windows, is an interface for manipulating iconic sound sources and sinks in virtual rooms. Listening to sound presented in this spatial fashion is as different from conventional stereo mixes as sculpture is from painting. A schizophrenic existence suggests sonic (analytic) cubism, presenting multiple acoustic perspectives simultaneously. Clusters can be used to hierarchically group related mixels (acronymic for '[sound] mixing elements') together.

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 universe. 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. These musings explore the philosophical distinction between egocentricism and exocentricism, especially as blurred by emerging technologies.

1 The 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 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 user loses their identity (empathy is 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.

- One (unusual) 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 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.

- A networked Formula 1 racing simulator arcade game, Sega's "Virtua Racing," allows each driver to switch between four modes:

  cockpit, in which the visual presentation is as if the user were inside the F1 car, including the dashboard, top of the steering wheel, and rearview mirrors;

  follow, in which the user'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.

  Even though the simulator's 'radio buttons' select a degree of immersion discretely, drivers may switch mode 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 show live video of the drivers, panning shots of the lead car, static shots of strategic curves, and instant replays of crashes [CC93] [Coh94].

For conversational groupware systems, the notion of egocentric and exocentric frames of reference can be reconciled with the grammatical person. In sliding from an immersive (subjective) perspective to an "exmersive" (objective) perspective, the user transforms from a 1^{st} person to a 3^{rd} person. Since all the 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 cast as special cases of transitive commands. By projecting the metaphorical universe onto an external and egalitarian medium, the 1^{st} and 2^{nd} persons have become special cases of the 3^{rd}.

2 Shared and split perception: beyond number

Most discussions of virtual presence are about its quality, individual resolution and interactivity; here we discuss scaling its quantity. Once we admit 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 users and sinks 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 faked by using a MIDI configuration 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 coöperating agents in a telepresence environment [She92]. 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. Of course they lack the control of VR systems, but interactive tv (like the zapping movie "Murderous Intentions") and call-in radios blur this distinction.

2.1 Video

There are several ways of presenting multiple video channels simultaneously. Distributed camera systems can present, via tiled or composited displays, multiple views simultaneously. 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. Music videos often crossfade visual scenes. Analytical cubism presents multiple visual perspectives on a scene simultaneously.

HUDs (head-up displays) are used in airplanes to throw naviagtion, tracking, and onto the windshield. Half-silvered mirrors can be used to image translucent images. Clearboard [Ish92] [IKG93] uses superimposed translucent viewing planes for teleconferencing with video of the conferees plus a shared whiteboard; different focal distances help distinguish the layers. [OTT92] superimposes real and virtual images by using the virtual image as a mask for the real, or by rendering the virtual image as a wireframe. "Mirror-type" VR systems like Mandala [Sta93] can superimpose (chromakey captured) sampled signals on arbitrary background graphics.

"Augmented reality" [CM92] [WMG93] is sometimes used to describe the superposition of computer-generated imagery on top of a see-through display. 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 between 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 presented, by presenting them to different hands.[1] Similarly, dichotic presentation 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, in general, occlude. Masking can be thought of as audio occlusion, but in typical voice and music appli-

[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.

cations masking is not a primary concern. 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 [CAK93].

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 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 might superimpose computer synthesized sounds, using some nonexclusive sound presentation like loudspeakers or openear headphones. For example, the author has thrown a ringing sound to a location occupied by a muted telephone. (This recalls [Nai91a, Nai91b]'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 instance of augmented audio reality.

This kind of superposition potential is manifested in MAW, acronymic for multidimensional audio windows, an audio windowing system with a visual map and auditory display: an interface for manipulating iconic sound sources and sinks in virtual rooms. Implemented as a NeXT-based frontend, MAW is suitable for synchronous applications like teleconferences or concerts, as well as asynchronous applications like voicemail and hypermedia, which can be thought of as equivalent to cyberspace [CL91] [Coh93a] [Coh93b] [CK94].

Its main view is a top-down 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 the human listener in the virtual room. (In a teleconference, an icon might be both a source and a sink.)

Source→sink spatialization is performed by a DSP (digital signal processing) module which convolves the digitized input streams with HRTFs (head-related transfer functions) that capture directional effects. 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.

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

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

The 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.

2.2.1 Shared perspective: fusion

Illustrating the many→one mapping of users to sinks, [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 conductor and (singleton) audience. This style presentation blurs the distinction between composer, performer, and audience. For graphical displays, this inter-user consistency is called "[relaxed] common view."

2.2.2 Non-atomic sources: cluster

MAW features a cluster utility. Clusters are hierarchically collapsed groups of objects [SZD+93], in this case spatial sound objects. By bundling multiple channels together, 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, along with dynamic allocation of spatializer channels, a 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. This focus 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 For example, in the context of a concert, individually recording (or mic-ing or synthesizing) the individual instruments, presenting each of the channels to MAW, and mixing them at audition time, rather than in "post-production" (as tracks and subgroups), allow the instruments to be rearranged by the listener. With the appropriate interface, one could grab onto an orchestral cluster, for instance (shown as part of the concert in Table 2), shake 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. This super decomposition aspect of the cluster feature could allow, for example, the user to listen to spatially separate strings of a violin, or different components of each string's sound. A generalized approach, ultimately fractal, assumes that there are always levels of zooming, or analysis, to be exploited.

We call the atomic sources "mixels," acronymic for '[sound] mixing elements,' in analogy to pixels or voxels, since they form the raster across which a soundscape is projected, and define the granularity of control.

Unclustering can be likened to viewing the sources through a generalized fish eye lens [Fur86], 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. While the other objects do not get pushed into the background, the idea is the same: to effect an external rearrangement of sources that complements an internal reordering.

Originally conceived as a tool for organizing sources, the cluster works for sinks as well. An example of a many→many 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 the audition [CK93].

- 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

2.2.3 Split perspective: 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 (biasedly) omnidirectional, perhaps audio windows can be thought of as implicitly providing this multiperspective capability, audio sources being inherently superimposable.

A simple teleconferencing configuration typically consists of several icons, representing the distributed users, moving around a shared conference space. These icons each represent a source, the voice of the associated user, as well as a sink, that user's ears.

Schizophrenia The 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. Therefore, the user can simply fork themself (with **copy**), leaving one clone hither while installing another yon, overlaying soundscapes via the superposition of multiple sinks' presence.

MAW allows users to designate multiple sinks, effectively increasing their attendance in the conference. Such a 'schizophrenic' mode allows multiple sinks in the same or different conference rooms, explicitly overlaying 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 the 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 (relaxed common view) model.

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 localized 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 is 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 [sometimes called Haas] effect.) 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 1 illustrates this behavior for a conference (top row) with two sinks, represented by top-down 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 exampled 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 (bottom row), the additive scenes are recombined, yielding the overall simplified percept.

Say, for example, that a listener wanted to pay special attention to the 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 themself, as in Figure 2, locating one instance inside the drum, and the other doppelgänger near the rhythm guitar.

Sonic cubism The situation of being in multiple places simultaneously has no physical correspondence in the real world, and therefore may define its own rules. A psychophysical interpretation, 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.

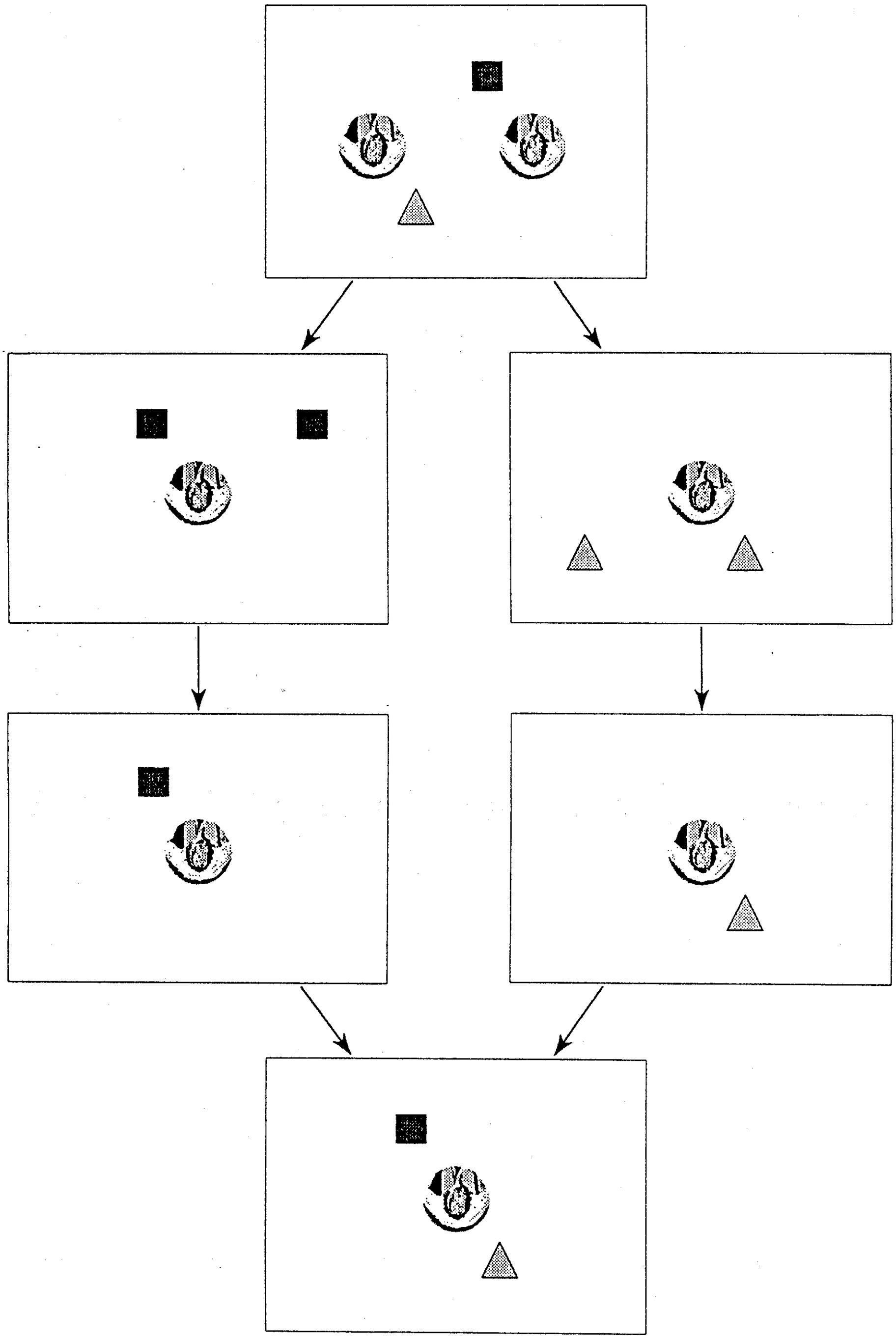

Figure 1: Sonic cubism: schizophrenic mode with autofocus

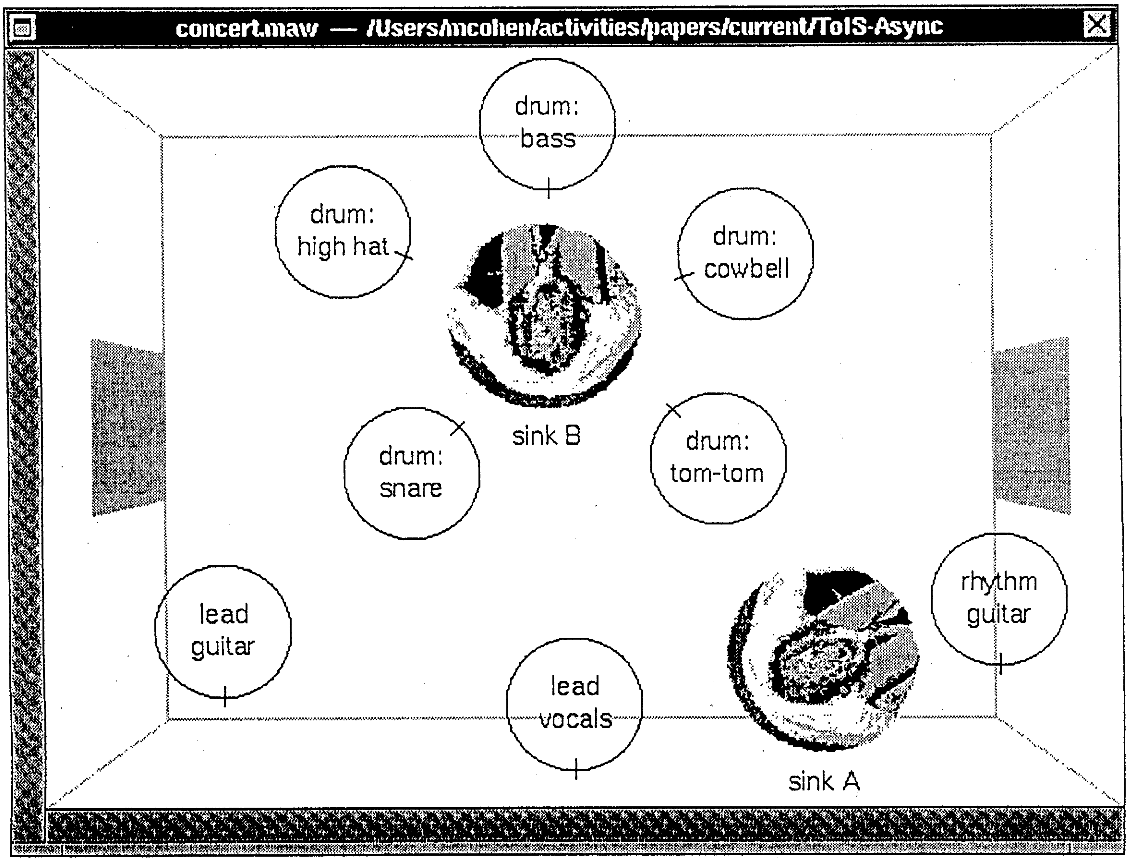

Figure 2: Virtual concert: multiple sinks and exploded clusters (generalized multi-focus audio fish-eye)

3 Grammatical blur: beyond pronouns

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 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 (and 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

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 [Kru82, Kru91], relax user↔sink mappings.

The difference between an exocentric perspective VR system and a multimedia interface is a (possibly multiple or shared) projection of the user into the virtual world. The arguments usually presented in favor of such workstation-oriented presentations, "desktop VR," involve issues like cost, fettered movement, engineering (sensor lag, update rates, display resolution), avoidance of "simulator sickness" [HW92], and user recalibration [Int92], but the 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, immersive, one. Truth constructed from multiple perspectives captures the essence of hypermedia.

To get a global perspective, for instance, a map is more useful than an immersive display. The scaling enables a quicker overview than possible with an immersive world, which is sort of a 1:1 map. Humans are quite good at conceptualizing 3-space from 2D projections. Sometimes maps are simply more convenient than dioramas.

It is important to note that the advantages of non-immersion 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 user, usually thought of as unique, is not above the terrain, but in it. Participatory and experiential $\not\Rightarrow$ inclusive!

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:

- The user perceives themself in the context of the colloquia.
- The user can perceive everyone else in the conference at once. In a multiple participant environment, the 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. In a teleconference, a user seated at a desk might not want to (have to) turn around to see who is approaching from behind.
- It is hard to imagine how multiple instances of self might be implemented effectively in an immersive system.

Rather than encase the user in a HMD and glove configuration, we augment their telephone and stereo, using the computer as a map. MAW's schizophrenic mode can be though of as forking reality, rather than cloning self. The perception of telepresence, especially forked, is auto-empathy, imagining how one would feel. New interaction modalities are enabled by this sort of perceptual aggression and liquid perspective.

References

[Bec92] Allen Becker. High resolution virtual displays. In *Proc. Second Int. Conf. Artificial Reality and Tele-Existence*, pages 27–34, Reflection Technology, Inc.; 230 Second Av., Waltham, MA 02154, 1992.

[CAK93] Michael Cohen, Shigeaki Aoki, and Nobuo Koizumi. Augmented audio reality: Telepresence/VR hybrid acoustic environments. In *Ro-Man: Proc. 2nd* IEEE *International Workshop on Robot and Human Communication*, Tokyo, November 1993.

[CC93] Michael Cohen and Mikiko Cohen. Technotokyo. *Tokyo Journal*, pages 32–33, March 1993.

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

[CK93] Michael Cohen and Nobuo Koizumi. Audio windows for virtual concerts. In *JMACS: Proc. Japan Music And Computer Science Society Meeting*, pages 27–32, Tokyo, February 1993. No. 47.

[CK94] Michael Cohen and Nobuo Koizumi. Putting spatial sound into voicemail. In Review by *1st International Workshop on Networked Reality in TeleCommunication*, Tokyo, May 1994. IEEE COMSOC, IEICE.

[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. 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. Hawaii Int. Conf. on Systems Sciences*. IEEE, January 1992.

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

[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.

[Coh94] Michael Cohen. Cybertokyo. *Presence: Teleoperators and Virtual Environments (Special Issue on Pacific Rim Research)*, 2(4), 1994.

[Fur86] George W. Furnas. Generalized fisheye views. In *CHI: Proc.* ACM *Conf. on Computer-Human Interaction*, pages 16–23, Boston, April 1986.

[HW92] Lawrence J. Hettinger and Robert B. Welch. Presence: Teleoperators and virtual environments, Summer 1992. 1(3).

[IKG93] Hiroshi Ishii, Minoru Kobayashi, and Jonathan Grudin. Integration of interpersonal space and shared workspace: Clearboard design and experiments. ACM *Transactions on Information Systems, Special Issue on* CSCW *'92*, 11(4), July 1993.

[Int92] Dimension International. Superscape virtual realities. Zephyr One, Calleva Park; Aldermaston, Berkshire, England RG7 4QZ, 1992.

[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.

[Kru82] Myron W. Krueger. *Artificial Reality*. Addison-Wesley, Reading, MA, 1982.

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

[Nai91a] Michael Naimark. Elements of realspace imaging. Technical report, Apple Multimedia Lab, 1991.

[Nai91b] 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 *ISMCR: Proc. Int. Symp. on Measurement and Control in Robotics*, pages 311–318, Tsukuba Science City, Japan, November 1992. Society of Instrument & Control Engineers.

[She92] Thomas B. Sheridan. 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, November 1992. Society of Instrument & Control Engineers.

[SK92] Gen Suzuki and Takashi Kouno. Virtual collaborative workspace. In *Proc. Second Int. Conf. Artificial Reality and Tele-Existence*, pages 119–125, 1992.

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

[SZD+93] D. Schaffer, Z. Zuo, J. Dill, L. Bartram, S. Dubs, S. Greenberg, and D. Roseman. An experiment comparing variable zoom to standard methods. In *Proc. Graphics Interface*, 1993.

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

[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), 1994.