Editors: Bernd Froehlich and Mark Livingston

Spatial Interaction in a Multiuser Immersive Instrument

Graham Wakefield
Korea Advanced Institute for Science and Technology

Tobias Höllerer, JoAnn Kuchera-Morin, Charles Roberts, and Matthew Wright
University of California, Santa Barbara

The appropriate support of spatial interaction is a perennial challenge in all kinds of VR environments. However, the results can be especially rewarding when you're interacting alongside other users in a surround-view and surround-sound immersive environment, such as the AlloSphere at the University of California, Santa Barbara (www.allosphere.ucsb.edu). The AlloSphere, conceived by JoAnn Kuchera-Morin, is a large scientific and artistic instrument for immersive human-centered visualization, sonification (using nonspeech audio to present information), and interactive data manipulation.[1] We allude to both the scientific and musical senses of "instrument." Like a microscope or telescope, the AlloSphere makes new realms accessible to human perception. Like the musical instruments in an orchestra or ensemble, it aims to facilitate multiuser parametric control of complex information.

A Distinctive Virtual Environment

The AlloSphere's metal screen comprises two 10-meter-diameter hemispheres and a connecting cylindrical section forming a three-story capsule enclosing more than 700 m^3 (see Figure 1). A 2-m-wide bridge supports up to 30 users. The super-black (12 percent reflectivity) screen limits secondary reflections; its fine perforations allow free placement of loudspeakers and tracking cameras behind it. Sound-absorbing materials and special heating and air conditioning provide low noise and dry acoustics. A 13-computer cluster renders 41.5 million pixels across 26 active stereo projectors, with 55 independent loudspeakers, using a software infrastructure based on AlloSystem, our C++ library. The instrument also incorporates a variety of wired, wireless, mobile, tablet, and laptop devices for interaction and display.

Like CAVEs (cave automatic virtual environments), planetariums, and head-mounted displays, the AlloSphere aims to mimic real-world environments, with minimal artifacts. However, the combination of three characteristics distinguishes it in particular.

First, the AlloSphere's seamless curved surround screen enables a nearly complete field of regard that avoids sharp distortion due to room corners. Projection coverage below the foot level affords bird's-eye aerial views, and seamless projection above the head supports "worm's-eye" views, with corresponding psychological effects.

Second, although most immersive environments focus on visual data presentation, the AlloSphere equally emphasizes visual and auditory surround display.[2] This emphasis aims to leverage a combined audiovisual representation's benefits for information exploration, including synergy, redundancy, disambiguation, and increased bandwidth of information transfer.[3]

Finally, the AlloSphere supports groups of participants, enabling collaboration in a shared surround-view environment. The 10-m diameter provides a large sweet spot for spatialized audio and stereoscopic or nonstereoscopic projection. Projects in the AlloSphere typically don't use head tracking, and participants are either unencumbered or minimally encumbered with wireless stereo glasses.

These design differences derive directly from the goals of supporting collaborative scientific data exploration and empowering human perception and action. Although these differences certainly pose implementation challenges, they bring op-

 Published by the IEEE Computer Society

portunities to investigate new spatial-interaction approaches and methods.

Spatial-Interaction Building Blocks

Each AlloSphere project uses its own data, audiovisual representations, metaphors, and subset of the instrument's affordances. However, many projects share the following spatial-interaction building blocks.

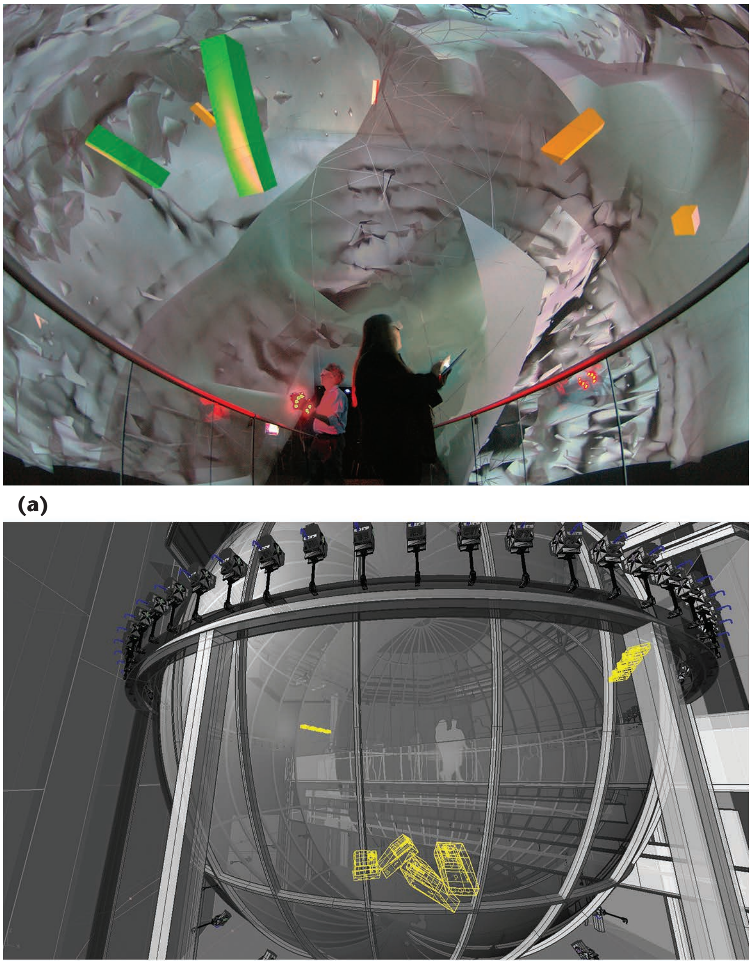
Multimodal Surround Display

AlloSphere applications incorporate surround graphics (stereoscopic or not), surround sound, and surround interaction.

Surround stereoscopy. AlloSphere applications can run in either windowed or full-surround mode. The former allows for the convenient display of legacy desktop content on two large quasi-rectangular display areas on either side of the bridge (see Figure 2). However, here we focus on full-surround display, which particularly challenges application design.

Although the AlloSphere can effectively display seamless nonstereoscopic surround-view content, most projects employ the instrument's stereo projection capabilities. Supporting surround-view stereoscopic projection for multiple non-head-tracked users (as is typical in the AlloSphere) is nontrivial. We've employed two methods, both related to the omnistereo approach demonstrated in cylindrical environments[4] and using calibration data providing each pixel's 3D position and blend factor. These methods correctly model stereoscopic parallax horizontally, tapering out stereoscopy toward the vertical apex, where the parallax axis could be unresolvably ambiguous.

Our first method uses object-order rendering through a standard multipass technique. Off-screen, it renders the scene into per-eye cubemap textures, assuming that the viewer is at the AlloSphere's center and is oriented toward the current



(a) (b)

Figure 1. The AlloSphere. (a) A fish-eye photograph of users interacting with immersive surround-view content (part of the Allobrain project). (b) An architectural rendering (by Gustavo Rincon). The screen comprises two hemispheres and a connecting cylindrical section forming a three-story capsule; a bridge supports up to 30 users.

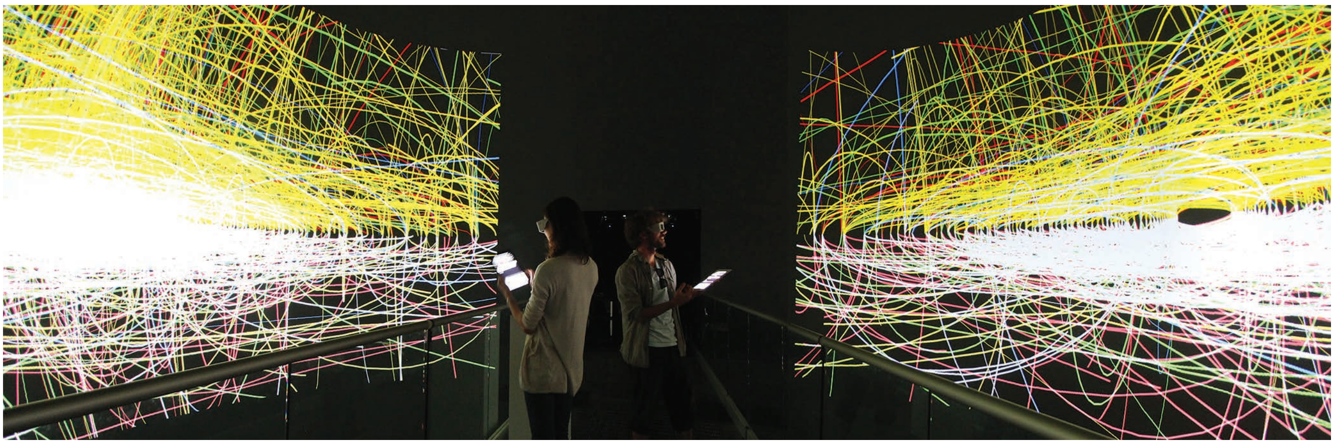
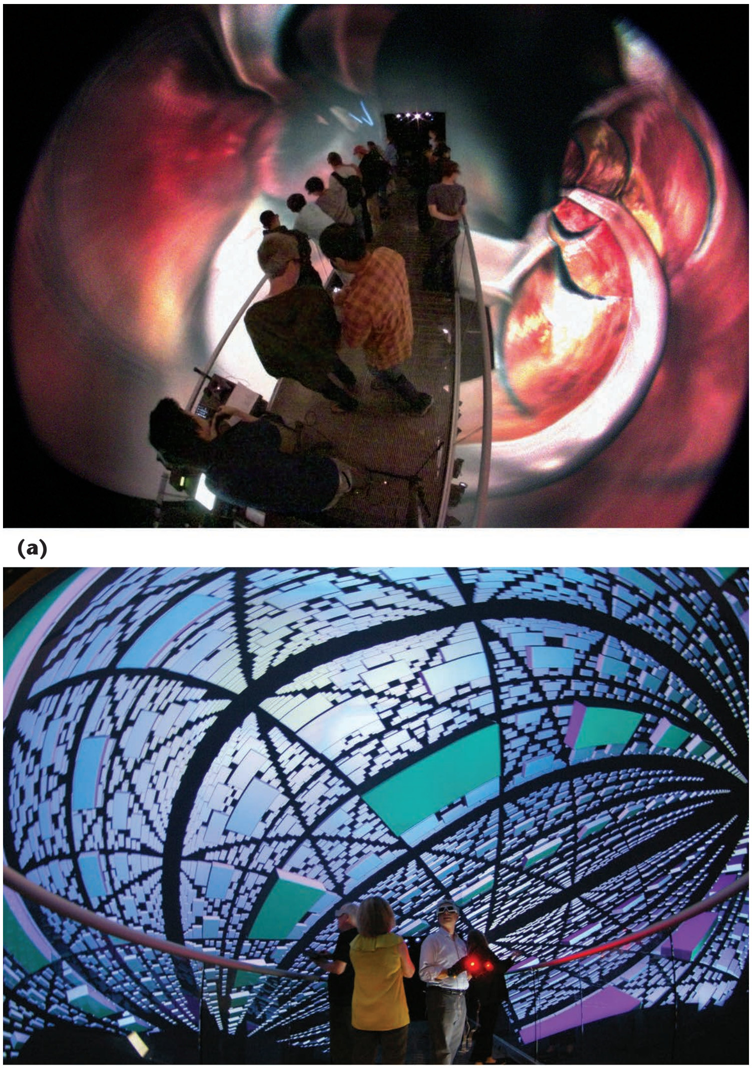


Figure 2. Users querying the Cloud Bridge audiovisual display via tablets. Users explore a dataset derived from the Seattle Public Library's checkout history.



(a)

(b)

Figure 3. Surround stereoscopy. (a) A fish-eye photograph of multiple users viewing a human-anatomy dataset, which the AlloSphere projected using object-order rendering. (b) A fish-eye photograph of multiple users viewing a ray-cast implicit surface, which the AlloSphere projected using image-order rendering.

vertex. It then warps the cubemaps to the screen using pixel shaders (see Figure 3a). We're investigating a more efficient single-pass object-order technique using direct vertex displacement, alleviating the need for off-screen rendering.

The second method uses image-order rendering, such as ray casting or ray tracing. To determine the per-pixel ray direction, it uses the calibration data, current view orientation, and ensuing stereo parallax (see Figure 3b). Current hardware is on the cusp of attaining sufficiently high frame rates with this method for reasonably high image resolutions.

Surround sound. The AlloSphere supports full-surround 3D spatialized audio with controlled acoustics. Sonification lends itself to process monitoring and data exploration.[5] Augmenting visualization with sonification can overcome occlusion and cognitive-overload problems because humans can more readily perceive and process audio in parallel. Audio can also display finer temporal structures over a wider range of frequencies than visual display. So, many AlloSphere applications use low-level parameter mapping to algorithmic sound synthesis to fluidly reflect subtle changes in an ongoing simulation.

Spatializing audio in full-surround 3D places sounds at specific directions in an unlimited field of "view." This is particularly conducive to exploratory orientation and localization. For example, sonic cues outside the visual field of regard can strongly direct attention to points of interest. In addition, our spatial-audio implementations can convey distance by amplitude and high-frequency attenuation, relative motion via Doppler shift, and environmental properties through reverberation. Most AlloSphere projects spatialize audio using third-order Ambisonics with common distance cues. We've found that higher-order Ambisonics provides a good compromise between vector base amplitude panning, which is efficient and simple, and wave field synthesis, which has a large sweet spot but requires high speaker density.

Surround interaction. Our open-source Device Server software configures spatial-navigation and interactive-control interfaces for AlloSphere applications.[6] It provides high-level abstractions for networked interactivity. It also provides a unified interface to low-level drivers for a variety of input sensing, including three-degree-of-freedom and six-degree-of-freedom controllers, motion tracking, depth camera sensing, and biometrics. A library for hand-based gesture recognition supports gestural control of surround information (see Figure 4).[7]

Shared and Individual Control

An important aspect of our interaction research is enabling researchers to use their personal mobile devices as interfaces to AlloSphere content. By exploiting such devices, we enable large-group interaction without the associated costs and complexities of maintaining a library of duplicate devices. Some shared-display applications require dedicated iOS and Android mobile apps to implement particular features. However, many applications provide browser-based interfaces, extending support to laptops and requiring no software installation by the user.[8]

Handheld-device displays' interaction capabilities differ from those of immersive displays. Combining the two display types involves decisions on how to present content on each type. For example, large-scale displays pose problems when presenting textual data because of the inherent tradeoff between the size requirements of legible text viewed at a distance and the subsequent occlusion of 3D data. So, many AlloSphere projects use personal displays for text-oriented interaction and the shared display for 3D interaction.

We also employ tracked phones and tablets as interactive magic lenses providing personalized annotated views into the AlloSphere content they're oriented toward. Even stereoscopic binocular views of AlloSphere simulations are possible, using the University of Southern California MxR Lab's VR2GO technology (http://projects.ict.usc.edu/mxr/diy/vr2go).

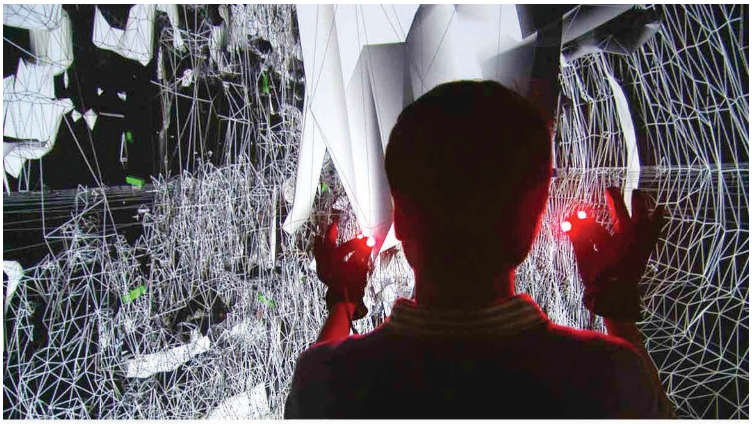
Supporting multiple types of display devices introduces a challenge of awareness: how to clearly identify relations between data on personal displays and the shared display. Supporting multiple users raises a corresponding challenge: when and how to distinguishably display different users' locations and activities in the shared dataset.

Spatialized Agents

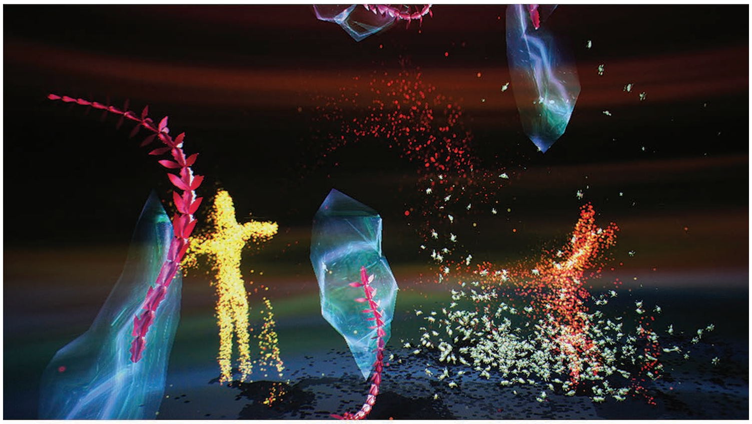
Many projects in the AlloSphere use multiagent systems. Although these systems don't yet appear to be a widely used spatial technique in scientific visualization, many multiagent systems have been visualized and sonified (including spatialized audio) in the arts.[9] As a form of spatially distributed AI, multiagent systems offer a computer-assisted means to distribute data exploration roles and to address scenes that could be difficult to analyze and monitor unassisted.

You can view a multiagent visualization as an extension of particle visualization: a system of multiple mobile entities responding to local conditions. However, particles typically follow a relatively simple global algorithm that often prioritizes efficiency in large numbers. In contrast, agents are designed toward greater autonomy and intelligence, with more complex, individual, and stateful operations.[10] Moreover, by detecting and responding to their environment—including each other—populations of agents can support cooperative macrobehavior. The canonical example is flocking.

Agents provide a spatial metaphor for information display[11] that's well suited to surround sonification. They provide an ideal vehicle to extend arbitrary sounds with spatial trajectories of direction, distance, and velocity using the techniques we described earlier. The sounds of agents behind a user's head or occluded by other visual elements can always be audible and localizable.



(a)



(b)

Figure 4. Environments with multiagent interaction. (a) In Allobrain, users employ gestures through wireless tracked gloves to navigate or to summon agents, using Ritesh Lala's Quintilian system.[7] (b) In Time of Doubles, populations of artificial-life agents consume the participants' particle cloud representations.

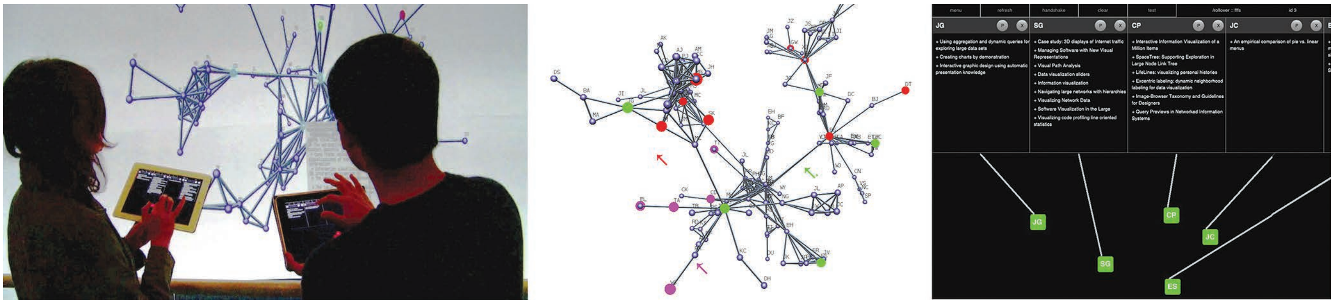
Applications

Since 2007, the AlloSphere Research Group has developed many immersive representations of nanotechnological, neurological, cosmological, and other multidimensional information systems. Here we present four of these applications in terms of the building blocks we just discussed.

Graph Browser and Cloud Bridge

These two projects investigate the collaborative benefits of providing personal viewing and interaction in conjunction with the shared display.

In Graph Browser (developed by Basak Alper, Charles Roberts, and Tobias Höllerer), each user controls a cursor in the shared display by touching his or her tablet. When the cursor rolls over an annotated feature in the structural data, the corresponding textual annotations appear on the tablet (see Figure 5). The interplay between the personal and shared displays offers a useful cognitive abstraction to separate content types (here, textual versus structural). More important,



(a) (b) (c)

Figure 5. In Graph Browser, a project for collaborative graph exploration, tablets provide personal views and search and annotation tools. (a) Two users interacting with the data through their tablets. (b) The shared display. (c) The display on a tablet. Multiple users can explore the data simultaneously without cluttering the shared display with query results.

multiple users can explore the data simultaneously without cluttering the shared display with query results. Nevertheless, when users find an item they deem relevant to the group, they can push the text to the shared display.

Color and spatial correspondences provide an awareness of other users' activities and the relation between the tablets and the shared display. The shared display identifies each user's cursor with a unique color, which the tablet also employs. In the shared display, regions with queries are marked by the appropriate user colors. So, users can focus on areas of the structural data that haven't been explored or, conversely, collaboratively explore a particular area. The tablet coarsely represents the spatial layout of regions a user has visited, situating users so that they can quickly transition between their tablet and the shared display. Similarly, the tablet's touch interaction uses absolute rather than relative mapping from the tablet to the shared display.

Cloud Bridge (developed by Qian Liu, Yoon Chung Han, Matthew Wright, and George Legrady) applies similar techniques. Multiple users explore a dataset derived from the Seattle Public Library's checkout history (see Figure 2). The shared display presents a broad view of this database as a chronological spiral repeating over one-week periods.

Users pose queries by entering keywords on their tablets, which display the textual results. Cloud Bridge visualizes the results on the shared display as arcs from check-out to check-in points in the time spiral. It sonifies the results by frequency modulation synthesis, mapping checkout dates, times of day, and loan duration to parameters of fundamental frequency, harmonicity, and duration, respectively. Cloud Bridge displays more data sonically than visually.

Like Graph Browser, Cloud Bridge uses color to provide contextual awareness of each user's queries. However, it also introduces user identification to the shared auditory display; each user has a unique amplitude envelope curve.

Allobrain and Time of Doubles

Allobrain (developed by Marcos Novak, Graham Wakefield, John Thompson, Lance Putnam, Dan Overholt, and Wesley Smith) constitutes our first experiments toward agent-based computer-assisted data exploration. It presents a virtual world consisting of isosurfaces of brain blood density drawn from fMRI (functional magnetic resonance imaging) data (see Figures 1a and 4a).

A small population of autonomous agents traverses the data space, driven partly by an asynchronous random walk. Each agent continuously emits a spatialized sound whose pitch reflects the locally measured blood density. Allobrain spatializes these "songs" according to the agent trajectories, creating a complex, multipath soundscape. If a user hears an interesting feature (possibly outside the field of view), he or she can summon the agent into near-field view to report its findings. Although only one user at a time can navigate the dataset, up to 12 users can simultaneously interact with the agents through diverse devices.

Surround spatial audio is essential to the interaction design. Each agent's song is unique and can be recognized by its pitch and other properties. It also incorporates a transient-rich pulse to support rapid, accurate localization. When a called agent comes into near-field view, the reduced distance leads to a correspondingly higher sonic amplitude, for better observation.

Allobrain inspired us to further pursue multimodal, multiagent interaction. The largest, most complex of our multiagent systems is Time of Doubles (developed by Haru [Hyunkyung] Ji and Graham Wakefield). Each agent is an evolving artificial-life organism that explores the fluid simulation, seeking particles to consume to stay alive and reproduce (see Figure 4b).

Unlike Allobrain, Time of Doubles allows up to a thousand active agents at a time. Each agent incorporates a distinct program combining input sensors, kinetic actions, memory, and other operations. This program is generated at the agent's

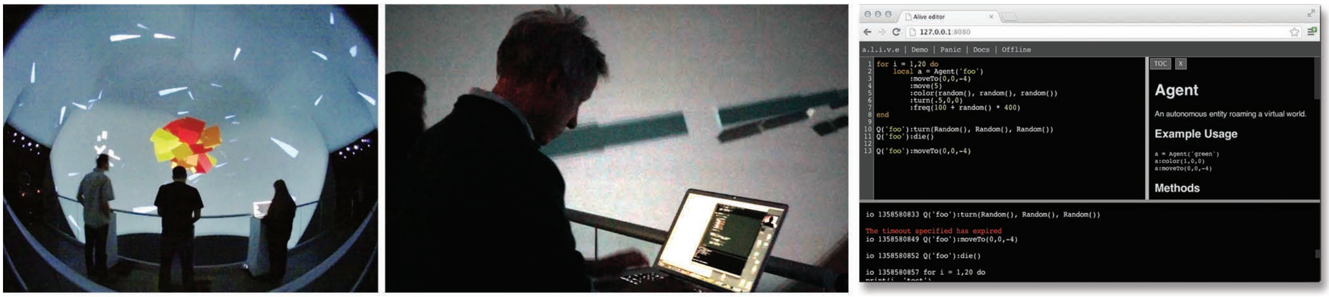


Figure 6. The prototype Alive system. (a) Multiuser live programming. (b) A close-up of a user. (c) The browser-based interface, showing syntax-highlighted code, console feedback, and contextual help. Users can define and modify the visual, sonic, and motile properties of the virtual world's agents by composing or reinvoking terse fragments of high-level code at runtime.

birth by genetic programming according to a mutated copy of the parent genome. The need to find food particles creates a selection pressure to evolve superior search strategies. Just-in-time compilation of the agent programs supports large populations without compromising complexity.

The strategy for sonifying agents drew inspiration from cricket chirps: short pulse trains of narrow-band frequency implemented by asynchronous granular synthesis. Users can easily localize the granular pulses, and the narrow bandwidth lets users identify many voices concurrently. Chirps are parameterized by the organism's genome. So, as populations grow and collapse, the soundscape develops from isolated pulses to dense clouds of sound, whose timbres vary with the evolving gene pool.

Interaction employs an array of commercially available depth cameras. Dense clouds of food particles mirror each participant's shape and movement. Three-dimensional optical-flow analysis of the clouds adds forces to the fluid simulation. The goal is to evoke an aesthetic experience through participation in a "natural" system. Participants can feed their doubles to growing populations of organisms and induce currents in the environment through unencumbered nuanced movements of their bodies.

Toward Ensemble-Style Interaction

Enabling groups of researchers to interact with immersive data in a shared environment presents a unique opportunity to divide the search responsibilities while sharing the results. The AlloSphere doesn't just privilege a single user with complete control over visualizations and sonifications. Instead, it provides concurrent users with configurable interfaces enabling them to

- adopt distinct responsibilities such as spatial navigation or agent control, and
- provide awareness of each other and raise attention as needed.

The multidisplay approach appears particularly well suited to foster such collaborative action. To continue the musical allusion of AlloSphere as an instrument, we characterize this as "ensemble-style interaction,"[12] underlining a creative aspect latent in the process of exploring models and data.

An ongoing challenge concerns how to support more open-ended, exploratory collaborative control and design. To this end, we're prototyping a promising form of ensemble-style interaction based on live coding, in which the users develop software as it runs (often as part of an audiovisual performance).[13] As with Graph Browser and Cloud Bridge, this prototype (named Alive) uses tablets or laptops for textual interaction. Specifically, Alive shares a virtual world's code with multiple users through an application-embedded Web server (see Figure 6).

Users can define and modify the visual, sonic, and motile properties of the world's agents by composing or reinvoking terse fragments of high-level code at runtime. Even a small amount of such coding can result in a complex behavioral environment. We believe such collaborative live programming could be productively applied to scientific data exploration. Users could design, dispatch, and supervise agents to search for specific features or to evaluate an in-the-moment insight. The agents could report the results by cross-modal display.

The use of spatialized autonomous agents connotes the possibility of hybrid spatial interaction that merges human and digital pattern-matching capabilities. Multiagent systems have a long history in AI and evolutionary design. However, systems for hybrid human-agent ensembles must take into account the perceptual and cognitive capabilities of the human as observer and supervisor.[14] The application of sense-limited multimodal display and natural user interfaces suggests an approach forming a seamless interconnection between the real and virtual worlds, to facilitate both intuitive and intellectual understanding.

Acknowledgments

This research was supported partly by US National Science Foundation grants IIS-0855279, IIS-1047678, and IIS-1058132; by the US Army Research Laboratory and the US Army Research Office under grant W911NF-09-1-0553 and cooperative agreement W911NF-09-2-0053; and by graduate fellowships from the Robert W. Deutsch Foundation. Special thanks to Dennis Adderton and the other researchers in the AlloSphere Research Group, without whom this work wouldn't be possible.

References

1. T. Höllerer, X. Amatriain, and J. Kuchera-Morin, "The AlloSphere: A Large-Scale Immersive Surround-View Instrument," *Proc. 2007 Workshop Emerging Display Technologies* (EDT 07), ACM, 2007, article 3.
2. X. Amatriain et al., "Experiencing Audio and Music in a Fully Immersive Environment," *Computer Music Modeling and Retrieval: Sense of Sounds*, LNCS 4969, Springer, 2008, pp. 380–400.
3. N.B. Sarter, "Multimodal Information Presentation in Support of Human-Automation Communication and Coordination," *Advances in Human Performance and Cognitive Engineering Research*, E. Salas, ed., JAI Press, 2002, pp. 13–36.
4. A. Simon, R. Smith, and R. Pawlicki, "Omnistereo for Panoramic Virtual Environment Display Systems," *Proc. IEEE Virtual Reality*, IEEE, 2004, pp. 67–73.
5. G. Kramer, "An Introduction to Auditory Display," *Auditory Display: Sonification, Audification, and Auditory Interfaces*, G. Kramer, ed., Addison-Wesley, 1994, pp. 1–77.
6. C. Roberts et al., "Dynamic Interactivity inside the AlloSphere," *Proc. 2010 Conf. New Interfaces for Musical Expression* (NIME 10), Univ. of Technology Sydney, 2010, pp. 57–62; www.mat.ucsb.edu/Publications/roberts_NIME2010.pdf.
7. R. Lala, "Quintilian: A Framework for Intuitive Interaction in Immersive Environments," master's thesis, Media Arts and Technology Graduate Program, Univ. of California, Santa Barbara, 2011; www.riteshlala.net/home/wp-content/uploads/2012/06/Quintilian-Ritesh-Lala.pdf.
8. C. Roberts, G. Wakefield, and N. Wright, "The Web Browser as Synthesizer and Interface," *Proc. 2013 Conf. New Interfaces for Musical Expression* (NIME 13), Korea Advanced Inst. for Science and Technology, 2013, pp. 313–318; http://nime2013.kaist.ac.kr/program/papers/day2/paper6/282/282_Paper.pdf.
9. D. Bisig, J.C. Schacher, and M. Neukom, "Flowspace–a Hybrid Ecosystem," *Proc. 2011 Conf. New Interfaces for Musical Expression* (NIME 11), Univ. of Oslo and Norwegian Academy of Music, 2011, pp. 260–263; www.nime2011.org/proceedings/papers/G13-Bisig.pdf.
10. R. Ahmad et al., "A Multi-agent Based Approach for Particle Swarm Optimization," *Proc. Int'l Conf. Integration of Knowledge Intensive Multi-agent Systems* (KIMAS 07), IEEE, 2007, pp. 267–271.
11. K.V. Nesbitt, "Metaphors for Information Display: Modeling the Multi-sensory Design Space," *Proc. 2001 Asia-Pacific Symp. Information Visualisation*, Australian Computer Soc., 2001, pp. 27–36.
12. J. Kuchera-Morin, "Performing in Quantum Space: A Creative Approach to N-Dimensional Computing," *Leonardo*, vol. 44, no. 5, 2011, pp. 462–463.
13. N. Collins et al., "Live Coding in Laptop Performance," *Organised Sound*, vol. 8, no. 3, 2003, pp. 321–330.
14. J.C. Rimland, M. McNeese, and D. Hall, "Conserving Analyst Attention Units: Use of Multi-agent Software and CEP Methods to Assist Information Analysis," *Next-Generation Analyst*, Proc. SPIE 8758, Int'l Soc. Optics and Photonics, 2013; http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=1692030.

Graham Wakefield *is a visiting professor in the Graduate School of Culture Technology at the Korea Advanced Institute for Science and Technology. Contact him at grrrwaaa@kaist.ac.kr.*

Tobias Höllerer *is a professor and codirector of the Four Eyes Laboratory in the Computer Science Department at the University of California, Santa Barbara. Contact him at holl@cs.ucsb.edu.*

JoAnn Kuchera-Morin *is a professor and the director of the AlloSphere Research Group in the Media Arts and Technology program at the University of California, Santa Barbara. Contact her at jkm@create.ucsb.edu.*

Charles Roberts *is a PhD candidate in the AlloSphere Research Group in the Media Arts and Technology program at the University of California, Santa Barbara. Contact him at charlie@charlie-roberts.com.*

Matthew Wright *is the research director of the Center for Research in Electronic Art Technology at the University of California, Santa Barbara. Contact him at matt@create.ucsb.edu.*

Contact department editors Bernd Froehlich at bernd.froehlich@uni-weimar.de and Mark Livingston at mark.livingston@nrl.navy.mil.