Composing Outdoor Augmented-Reality Sound Environments

Camille Goudeseune, Hank Kaczmarski

Integrated Systems Laboratory, Beckman Institute, University of Illinois at Urbana-Champaign
email: camilleg@isl.uiuc.edu, hankk@isl.uiuc.edu

Abstract

We have constructed a wearable outdoor computer system which presents sounds and images placed at particular locations by means of headphones and a head-mounted display. The apparent location of a sound source can be fixed relative to the earth, independent of the listener's position or orientation. For easy duplication, the system is built exclusively from off-the-shelf hardware and publicly available software.

1 Introduction

The field of augmented reality aims to combine virtual reality with "real" reality, to augment the domain of our everyday sensory input and motor output instead of demanding that we face a computer screen. Rapid improvements in battery technology and electronics miniaturization are making it possible to accurately measure a person's position and which way they are looking. Armed with a model of the environment they are walking around in, we can add objects to this environment which they can see and hear as behaving consistently. If they walk around one such object, they see and hear it as staying in one place.

Because we hear with less spatial precision than we see, and because the "field of view" of earphones is far larger than that of eyephones, the augmented reality illusion can be quite strong with sound. This is particularly the case when listener orientation is measured with a low-latency compass. Combining such a compass with headphones is startlingly effective because we naturally expect the audio field of headphones to be relative to head position, not to the outside world. So when a sound such as a birdsong remains locked in place when you turn your head, you instinctively think that the sound came from a real bird, not the headphones.

2 Infrastructure

The head-mounted display (HMD) is based on the i-O Display Systems i-glasses unit, which displays a small image to each eye and stereo sound via headphones. The sounds and images come from a generic Windows 98 laptop computer. Image and sound computation is driven primarily by the position and direction of the wearer, which are inferred from the values reported by a GPS receiver, flux-gate compass, tilt sensor, and gyroscope connected to the laptop with RS-232 cables.

2.1 Hardware

The Precision Navigation TCM2-50 compass module contains three orthogonal coils and a tilt sensor, which keep it accurate even when tilted 50 degrees from level. It reports pitch angle and roll angle as well as compass heading. These two angles are valuable particularly for aligning three-dimensional visual images with the real world. Visual registration between the real and virtual worlds is difficult because we see spatial relationships with such high precision.

Accurate estimation of position and head tilt. Like most handheld GPS receivers, the Garmin GPS 12 unit reports position data over its RS-232 link only once per second. To reduce position measurement latency we use a gyroscope for inertial dead reckoning between these reports.

When the wearer begins walking forwards, the tilt sensor reports a change of pitch angle because, like any accelerometer, it cannot distinguish between horizontal acceleration and tilt relative to Earth's gravitational field. In this scenario the translationally insensitive gyroscope does not also report tilt, so we reinterpret the reported pitch angle as a forwards acceleration. From this, successive integration yields velocity and then position. (Had the gyroscope also reported tilt, then we would conclude that head tilt actually occurred.)

The gyroscope has high drift, so this technique cannot entirely substitute for the GPS receiver if the latter's view of the sky is obscured for an extended period. The gyroscope is actually a Gyration GyroMouse Pro computer mouse, which normally operates in midair by pointing it in different directions.[1]

Without this combination of gyroscope and tilt-sensor data, the image would bob up and down whenever the wearer changed velocity causing horizontal acceleration to be misinterpreted as head tilt. The GyroMouse is mounted vertically on the HMD, in order to place its two rotationally

[1] Roth (2001) describes how GyroMice can be used by dancers to control Opcode Max.

sensitive axes in the horizontal plane where we need to distinguish tilt from acceleration (figure 1).

We connect the GyroMouse to another RS-232 port on the laptop, instead of to the more obvious PS/2 mouse port. This lets us capture the data directly, without having the operating system try to interpret the data as mouse motion.

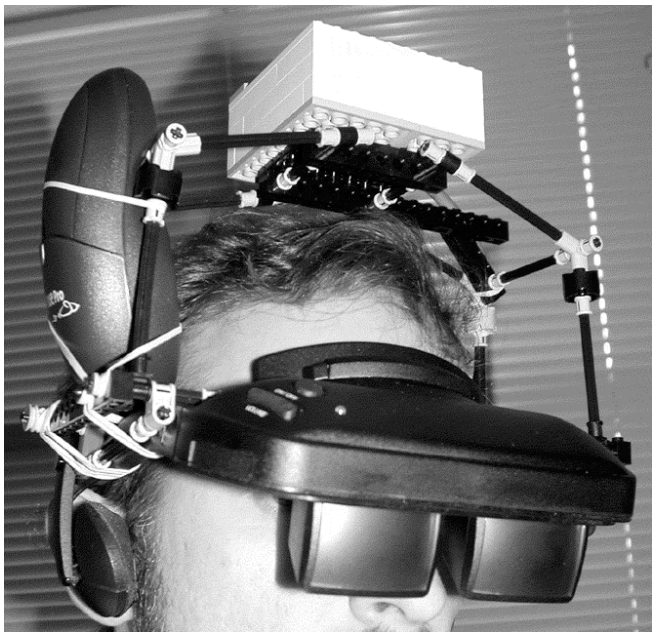
Mechanical Details. We chose to attach the compass and gyroscope to the HMD with Technic Lego because it is nonferrous, robust, lightweight, and easy to prototype with. (Glue and elastics can strengthen the Lego joints if needed.) Both the HMD and its superstructure fold up for easy transport.

To keep the HMD comfortably lightweight, all the other equipment is stowed in a backpack. (The Sony Vaio PCG-F520 laptop can be nondestructively modified to operate while folded up in the backpack. Unscrew the case, pry open the front edge, and in the locking mechanism flip over the plastic lever which triggers a microswitch.) The HMD weighs 460 g overall; the backpack, 4.6 kg.

The laptop contains a B&B Electronics model 232PCC2 PCMCIA card which provides extra RS-232 ports, since most laptops have only one such port and we need one each for the GPS receiver, compass, and gyroscope. An optional second PCMCIA card provides a wireless Ethernet connection to access real-time data from or report the listener's position to other computers.

Accuracy. The GPS positional accuracy we measure is 1.5 metres. Positional dead-reckoning latency is normally 200 msec but can be adjusted. Angular accuracy is under 1.5º, well within the spatial accuracy of human hearing; pitch and roll have a range of up to 50º from vertical. Visual contrast, gamut, and resolution are poor compared to nonportable display devices, but brightness is adequate for outdoor use. Audio quality is typical for laptop built-in sound hardware. This could be improved by using a self-powered USB audio interface such as the Roland UA-30, but in most outdoor environments the weak link in audio quality is background noise.

Power consumption. Battery life of the laptop is about two hours, somewhat less when using the wireless Ethernet card; for longer demonstrations we keep extra charged batteries on hand. The HMD and compass can both run from an unregulated 6-volt supply, so they are powered by a 1500 mAh NiMH battery pack for radio-controlled model cars (Radio Shack part no. 23-338; removing the innards from their 23-327 recharger produces a convenient mounting bracket for the battery). This battery pack lasts 5 hours, long enough to fully recharge a second one before the first runs out. Four AA batteries power the GPS receiver for 24 hours. They could actually be removed, and weight correspondingly reduced, by powering the GPS from the same source as the HMD.



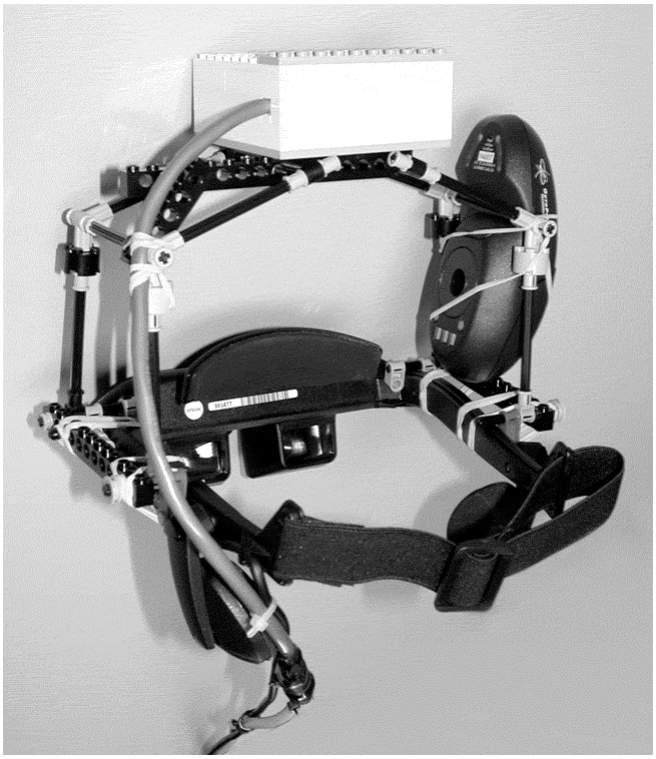


Figure 1. The head-mounted display. The white box on top contains the compass module. The dark ovoid on the side is the GyroMouse, which has a wireless connection to the backpack. The display units do not obstruct peripheral vision, so it is safe to walk (slowly) while wearing the HMD.

2.2 Software

The software which collects data from all these sensors, coordinates it, and generates images is based on the Syzygy open-source toolkit for implementing virtual environments on clusters of PC's (Schaeffer 2000). Its flexibility can be seen by the range of applications to date: psychology experiments (McConkie, Zheng, and Schaeffer 2001), a multi-user shared virtual environment incorporating an earlier version of the HMD (Goudeseune and Schaeffer 2000), and a high-performance fully enclosed virtual reality chamber. Syzygy, and the source code for this audio system, are at <http://www.isl.uiuc.edu/ClusteredVR/ClusteredVR.htm>.

Normally Syzygy runs on a network of computers (Irix, Windows, and Linux), but in this configuration it all runs on a single laptop. Since Syzygy is highly multi-threaded, it communicates efficiently with the RS-232 based sensors. This also means that it stays running if a part of the system is removed, either intentionally or accidentally (a loose connector, a low battery). The system automatically detects when the absent part starts running again: no restart is needed. This may seem over-engineered, and indeed once a hardware configuration has stabilized it is a luxury, but during development it is a necessity.

Sygyzy uses the "fmod" multiplatform low-level sound library (www.fmod.org), free for academic use. It provides a straightforward three-dimensional world model, uses very little CPU power, and takes advantage of several kinds of hardware acceleration in commercial sound cards.

3 Compositional Issues

A simple but rich class of sound environments which can be built with this system is the simulation of conventional sound sculptures or sound installations (with the advantage of nearly instant deployment and tear-down). The interaction between listeners and environment is deliberately restricted to moving around, though the software could in principle do much more. (Multiple HMD's can operate in the same physical space, of course.) Part of the success of these passive environments may be because most members of the public (adults, at least) prefer zero training before using the system. Their attention is adequately captured by discovering how walking around affects what they hear. The first environment we built for an earlier version of this system added sounds to a campus quadrangle. Various mechanical rumblings and clankings emanated from the buildings on the perimeter; a bird was placed in each of a dozen trees along a curved path crossing the quadrangle; at the central sundial/fountain sculpture, excerpts from Vivaldi's *The Four Seasons* emanated from the fountains marking the solstices and equinoxes. To this we later added spoken announcements of each bird, season, and building name, triggered when the listener approached the same.

Within this passive mode, we can play games with the listener's expectation that the sound environment entirely obey conventional acoustic laws. For example, as the listener approaches a sound source, instead of becoming gradually louder following the inverse square law, the sound may begin suddenly when the listener crosses a distance threshold. If such a boundary is used, hysteresis can gently lead listeners away from exploring the limitations of the system (by jumping back and forth across the boundary of a sound's audible region) and back into the sound world itself.

The sound environment can respond more actively to listener position. In the *Four Seasons* example, the fountains were only a few meters from each other. But since it was undesirable to simultaneously play several excerpts of tonal music, only the one nearest to the listener would be played. (Speech is a similar example.) The word interactive begins to apply honestly in a "mosquito" environment where buzzing sounds become louder and more insistent the longer listeners stand still, even pursuing them once walking begins, and ceasing suddenly and spectacularly when a bug-zapping location is approached.

Interactive sound installations which use video cameras as sensors such as the Very Nervous System (Rokeby 2000) are experienced immediately and even unsuspectingly by passersby. But we can hardly sneak a backpack and HMD onto a listener (though the hardware continues to shrink: our system is not far from running on a palmtop computer, and GPS-on-a-chip products are starting to appear). The advantage of GPS tracking and headphone presentation is primarily spatial. Sounds can emanate from any point at all, not only from speakers or other mounted objects. Sounds can be precisely superimposed on a familiar public space to transform it, and can even follow arbitrary paths through the space.

Of course the inherent flexibility of software can let us compose incomprehensible webs of relationships between sounds and listener, among sounds themselves, and among multiple listeners. If we want unprepared listeners to (after initial perplexity) enjoy discovery and increased understanding through feedback from actions they take, the lack of nonacoustic clues about the behavior of the sound environment argues for keeping things not too far removed from everyday experience.

Handheld input devices such as the tilt-sensitive Side-Winder Freestyle Pro USB gamepad can be used for direct interaction, but for purely audio environments this is often more distracting than helpful. This may be because the time scale of pushing buttons is so much faster than that of walking. Again, simple controls are satisfying to use: the gamepad's throttle controls overall volume; while pointing the gamepad at a sound source, holding one button mutes that sound while holding another "solos" it, *i.e.*, mutes all other sounds; in an environment with mobile sound sources, pressing another button can push all sounds a few metres farther away from you. This last example demonstrates listener control of structural parameters. Controlling these rather than directly acoustic parameters such as pitch or rhythm leads the listener away from the trap of neophyte

instrumental performance, an activity whose high feedback is immediately attractive but does not sustain interest and ultimately distracts from the environment as a whole.

4 Describing the Environment

Passive sound sculptures are defined in this system by simply listing the sounds in a text file, which is then read by the Syzygy program when it starts up. Preset spatial behaviors for looped sounds include things like tracing a circular path, oscillating back and forth on a line, proximity sensing with hysteresis, Brownian motion, and of course immobility. Triggered sounds, on the other hand, have finite duration so their spatial behaviors are somewhat different. These include a variant of proximity detection (useful for spoken announcements), a Poisson process with adjustable irregularity, and so on. All sound sources have a maximum radius beyond which they cannot be heard.

More active audio environments with advanced interaction are impractical to describe in terms of presets. C++ programming is then required for defining callback functions implementing custom spatial behavior, for defining interactions between sounds, and for creating and destroying sounds on the fly. The simpler environment description of a text file listing the sounds allows only for persistent, not temporary, sound objects.

The geographical coordinate system used is based on latitude and longitude. Handheld GPS receivers often report these values in degrees, minutes, and ten-thousandths of a minute (40º 06.9370').[2] Since this system is aimed at perambulatory rather than jet-powered listeners, limited range suggests using an origin for the coordinate system closer to home than where the Greenwich meridian meets the equator. So we let the environment designer specify a nearby origin, such as (40º 06.0000' N, 88º 13.0000' W) for Urbana, Illinois. All measurements are then made relative to this origin. For convenient manual data entry we use a unit of one ten-thousandth of a minute: a coordinate range from 0 to 1000 then covers about 150 meters, or about two minutes of uninterrupted walking. Positions can be determined directly from GPS measurements, or from maps or online satellite photographs.

5 Future Work

We plan to extend this work to an indoor/outdoor system with GPS-like transmitters called pseudolites or pseudo-satellites (Cobb and O'Connor 1998). Pseudolite-based GPS is typically accurate down to a centimeter. Such high resolution dramatically changes the kind of environments which can be composed: the world can respond to gestures which are much smaller and faster than taking a few steps. With multiple receivers in a jacket, even the position and orientation of individual limbs can be determined.

[2] One ten-thousandth of a minute of latitude is about 18.6 cm. One ten-thousandth of a minute of longitude is 18.6 cm at the equator, 13.1 cm at 45º latitude, 9.3 cm at 60º.

6 Acknowledgments

Ken Chen implemented an early prototype of a vehicle-based GPS-driven sound system at the Integrated Systems Laboratory. His work and our continuation of it were funded by Yamaha Motor Corporation's "Computer Companions" project.

References

Cobb, S., and M. O'Connor. 1998. "Pseudolites: Enhancing GPS with Ground-Based Transmitters." *GPS World* 9(3):55–60.

Goudeseune, C., and B. Schaeffer. 2000. *CARMEN: Collaborative Augmented-Reality Multimodal Environment*. <http://www.isl.uiuc.edu/Virtual%20Tour/TourPages/carmen.htm>.

McConkie, G. W., X. S. Zheng, and B Schaeffer. 2001. "Effects of Navigation Control Method on Spatial Updating in Virtual Environments." *ARL Federated Laboratory 5th Annual Symposium—ADID Consortium Proceedings*, pp. 59–64.

Rokeby, D. 2000. *Very Nervous System.* <http://www.interlog.com/~drokeby/vns.html>.

Roth, I. 2001. *How to use a Gyration GyroMouse with Max.* <http://ccrma-www.stanford.edu/~issac/max/gyromouse.html>.

Schaeffer, B. 2000. *A Software System for Inexpensive VR via Graphics Clusters.* <http://www.isl.uiuc.edu/ClusteredVR/paper/dgdpaper.pdf>.