



*this is a 360 / ambisonic video. that means that as long as you are viewing on desktop, you can drag the video around in 3 dimensions and hear the sound pan around you. wear headphones for best results.*

area~ (May, 2020)

**project paper: sonicscope issue two journal article**

## Summary

The area~ system, which stands loosely for 'augmented reality environmental audio' aims to afford users the ability to spectromorphologically (defined by Smalley to concern spatial, temporal and textural qualities of sound (1997)) manipulate sounds from their environment into a virtual audio environment. Through bone conduction headphones and head tracking, this sound field is heard in synchronicity with their actual environment. The system was created in order to explore and reveal the relationship between real and virtual environments.

Through an autobiographical design study, these experiences are discussed in relation to the research question: "How can we better understand relationships between virtual and real environments through gesture and virtually placed audio augmented reality objects?" This hypothesis study proposes that new aesthetic experiences can result from the system and are waiting to be tested through user studies. The adoption of the **North Star open-source AR head-mounted-display** could expand the possibilities of the area~ system by reducing the need to be tethered to a laptop and table for hand gesture input.

Overall, the results from ABD have shown that area~ is an effective tool for examining the combinatorial relationship between real and virtual environments. Despite the systems hardware setup needing to be worked on to allow for quicker start up and more accurate head tracking, it has provided me with novel aesthetic experience through:

- The blending of real and virtual auditory environments to create a third, augmented environment that was



greater in experiential nature than the sum of its parts (not simply a combinatorial layering)

- The ability to spectromorphologically manipulate sounds in real-time in this third environment with the body
- The potential for creating believable illusions of real-world sound sources from these manipulated and spatialised virtual sounds.

## Technologies

The three technologies used in area~, are gestural hand tracking, rotational head tracking, and ambisonics. The gestural hand tracker used in the system is a Leap Motion LM-010 Controller, a USB infrared camera device that provides location and orientation data output of individual finger joints (and therefore hands) when they are presented above the device. The Leap Motion Controller (LMC) has been adopted in a multitude of settings such as being mounted on VR headsets (Leap Motion 2016), and converting hand gestures to MIDI (Leap Motion 2017). More recently, UltraLeap are investigating the use of this same technology with gesture-based public information screens to help combat the “hygiene risks of touch screens” (2020).

Rotational head tracking is achieved via an inertial measurement unit (IMU). This small and inexpensive component provides orientational data output at 20 times a second. When affixed to the head via a headset or headphones, it is a relatively easy and cheap way of implementing head tracking into the system.

Ambisonics is an audio format that allows for full-spherical audio capture and playback (Gerzon 1973). There are four recorded channels (referred to as A-Format) that unlike regular mono, stereo, or surround sound formats, contain no information about the speakers it is delivering the signal to. Rather, these channels can be encoded in such a way as to describe a three-dimensional field of sound referred to as B-Format. B-Format can be decoded through “virtual microphones”, any number of which can be placed within this three-dimensional sound field to provide standard channel outputs.

For example, in area~, I have used a RØDE Soundfield NTSF-1 microphone array comprised of 4 microphones. The A-Format output is encoded to B-Format by an audio plugin. A software library I have configured, further decodes the B-Format to two responsive, binaural, virtual audio output channels. This all occurs in real-time; and as the user moves their head, the microphones inside the three-dimensional sound field rotate proportionally, providing realistic changes to what is heard.

## Experience

The recording or ‘sampling’ stage is initiated by making a left-hand grab above the LMC, the longer lasting the grab, the longer the portion of audio from the ambisonic palette is sampled. The three-dimensional coordinates of the hand above the LMC correlates with the location of audio recorded (this is achieved by mapping the hand coordinates to a virtual microphone inside the ambisonic palette), essentially allowing the user to record sounds around their person in three dimensions. Upon letting go of the grab gesture, the sample plays on repeat (using the karma~ Library (Constanzo 2015)) through the bone conduction headphones, thus setting up the session’s virtual audio environment.


The manipulation stage is automatically initiated after the ending of the previous grab gesture and uses translational (x, y, z) and rotational (roll, pitch) values from both hands when above the LMC. There are two audio effects being manipulated, with parameters from these effects mapped in different ways to the




translation and rotation of the user's hands. When the user decides to end manipulating the sample, they can do so by performing a grab with both hands. Once this happens, the band-pass filter and granular synthesis parameters are frozen for that sample.

The spatialise stage begins once the manipulation stage is ended by the user. The three-dimensional space above the LMC is mapped to the virtual audio environment, in which the user is currently listening to the sample that they have recorded. The user can use their right hand to move the sample around the virtual audio environment. For an example of the effect this has, moving the hand between the two extremes of the x-axis (left to right) results in hearing the sample move from ear to ear. The spatialise stage is ended by grabbing with the right hand.


Once the spatialise stage has ended, the user has the option to repeat the process 7 more times, allowing for the creation of a virtual audio environment comprised of up to 8 spatialised audio samples, or what I refer to as nodes. Below are some examples of these virtual audio environments.

 virtual audio environment 1 (inside)

 [SOUNDCLOUD](#)

 virtual audio environment 2 (birds)

 [SOUNDCLOUD](#)

 virtual audio environment 3 (meeting)

 [SOUNDCLOUD](#)

 virtual audio environment 4 (outside)

 [SOUNDCLOUD](#)