HearThere: Networked Sensory Prosthetics Through Auditory Augmented Reality

HearThere

Networked Sensory Prosthetics Through Auditory Augmented Reality

Spencer Russell
sfr@media.mit.edu

Gershon Dublon
gershon@media.mit.edu

Joseph A. Paradiso
joep@media.mit.edu

Responsive Environments Group
MIT Media Lab
75 Amherst St.
Cambridge, MA, USA

ABSTRACT

In this paper we present a vision for scalable indoor and outdoor auditory augmented reality (AAR), as well as HearThere, a wearable device and infrastructure demonstrating the feasibility of that vision. HearThere preserves the spatial alignment between virtual audio sources and the user's environment, using head tracking and bone conduction headphones to achieve seamless mixing of real and virtual sounds. To scale between indoor, urban, and natural environments, our system supports multi-scale location tracking, using fine-grained (20-cm) Ultra-WideBand (UWB) radio tracking when in range of our infrastructure anchors and mobile GPS otherwise. In our tests, users were able to navigate through an AAR scene and pinpoint audio source locations down to 1 m. We found that bone conduction is a viable technology for producing realistic spatial sound, and show that users' audio localization ability is considerably better in UWB coverage zones than with GPS alone. HearThere is a major step towards realizing our vision of *networked sensory prosthetics*, in which sensor networks serve as collective sensory extensions into the world around us. In our vision, AAR would be used to mix spatialized data sonification with distributed, livestreaming microphones. In this concept, HearThere promises a more expansive perceptual world, or umwelt, where sensor data becomes immediately attributable to extrinsic phenomena, externalized in the wearer's perception. We are motivated by two goals: first, to remedy a fractured state of attention caused by existing mobile and wearable technologies; and second, to bring the distant or often invisible processes underpinning a complex natural environment more directly into human consciousness.

CCS Concepts

•Computing methodologies → Mixed / augmented reality; •Hardware → Sound-based input / output; **•Information systems →** *Location based services;* **•Human-centered computing →** Mobile computing;

AH 2016, February 25 - 27, 2016, Geneva, Switzerland

ISBN 978-1-4503-3680-2/16/02…$15.00
DOI: http://dx.doi.org/10.1145/2875194.2875247

Keywords

sonification, auditory augmented reality, sensory augmentation, UWB, bone conduction

1. AN IMAGINED FUTURE

A woman stands at the foot of a path on an overcast summer day in late afternoon and looks into the woods. Beside her, an expanse of open marsh. Around her, the sounds of wildlife making the transition from day to night are beginning to swell. On her left, a frog croaks from the banks of a stream. Just behind her on the right, a turtle disturbs the surface of a still pond, creating a visible ripple. In front of her, birds shuffle about in the trees. She puts on a headset with bone conduction transducers in front of each ear and a small enclosure against the back of her head.

Through the headset, she is able to hear extraordinary sonic detail in her surroundings. She hears chicks in a nest in the woods, peeping softly. The turtle's meal becomes audible, a plop in the rippling water beside her. There are signs of activity below the surface—the otherworldly clicks and pops of frogs and fish swimming and feasting.

She fixates on the pond and a spatial richness begins to emerge in the sound from the water. What was before a single source becomes a complex mixture of sounds from hydrophones throughout the pond and its tributary stream. As the underwater sound comes to the foreground, the birds blur together to form a general background, still emplaced in space but no longer specific. Underwater sensors measure levels of dissolved oxygen, and a layer of unnatural droning fades into her awareness. The sounds closer to her are just slightly out of tune with the ones in the distance. She notes that the water closest to her is stagnant. She pauses for one more moment, taking it in, before returning her gaze to the woods. The sounds of the pond recede from the foreground as she continues down the path.

2. INTRODUCTION

Mobile devices provide immediate access to information at a distance—communications, world news, even sensor data. As mobile augmented reality (AR) applications have come into widespread use, the same devices are increasingly mediating our sensory experiences of the immediate environment (we might choose to check a weather app before going outside). This has led to a paradoxical reality where real-time data about the world has grown, but we find ourselves less and less present in it. Instead, we exist in a permanently fractured state of attention. Of course, sensors are all around us, capturing rich, 'sensory' data about our environments, but this data is typically presented through the same highly mediated and overloaded channels through which we communicate, create, and consume media.

Researchers across a variety of fields have explored spatial mappings of data through devices on the body, under the broad categories of AR and sensory augmentation, where the former refers to physical world mappings of media and the latter to haptic or auditory mappings of data derived from wearable sensors, using the language of prosthetics [2]. We envision a future in which data from distributed sensor networks would be incorporated into the sensorium, building on the spatial qualities of AR and the sensory qualities of prosthetics. Like glasses, *networked sensory prosthetics* exist between the body and world, working to alter a wearer's perception of their surroundings without becoming a site of the wearer's attention in themselves. While visual AR has been extensively explored in a variety of application domains from gaming to task support, we believe that auditory augmented reality (AAR) lends itself well to undirected, sensory experiences, where users can shift their attention fluidly from source to source in a 360° sound field. In this vision, sensor networks serve as collective sensory extensions into the world around us. AAR would be used to mix spatialized data sonification and sound from distributed, livestreaming microphones into real world sensory experience.

In this paper we present a vision for scalable indoor and outdoor auditory augmented reality (AAR), as well as HearThere, a wearable device and infrastructure demonstrating the feasibility of that vision. Rather than working towards and prototyping this vision with off-the-shelf equipment such as camera-based optical tracking systems or VR headsets, this work develops a set of technologies which, taken together, are suitable for building systems of this kind. HearThere preserves the spatial alignment between virtual audio sources and the user's environment, using head tracking and bone conduction headphones to achieve seamless mixing of real and virtual sounds. Taking advantage of low-cost, precision-ranging Ultra-WideBand (UWB) radio chips, our multi-scale head-tracking system uses UWB localization anchors for 20 cm resolution tracking, falling back to GPS when UWB is unavailable. The device also includes a 9-DOF inertial measurement unit (IMU) and embedded sensor fusion for orientation tracking. In our tests, users were able to navigate through an AAR scene and pinpoint audio source locations down to 1 m. We found that bone conduction is a viable technology for producing realistic spatial sound, and show that users' audio localization ability is considerably better in UWB coverage zones than with GPS alone. We believe this multi-scale tracking approach is necessary for expanding to large geographic areas.

Our work in this space is largely targeting *Tidmarsh*, an instrumented wetland restoration site where densely-sampled ecological measurements and sound are being continuously collected and streamed. This collaborative effort seeks to make *senseable* the invisible ecological processes that support a complex ecosystem. We seek to heighten human sensual experiences of the environment through these kinds of extrasensory extensions into the natural world.

As the opening speculative illustration shows, our vision extends beyond pure AAR to incorporate the wearer's selective attention, as well as strategies for dynamically foregrounding and backgrounding sources. For future work, we have designed experiments that assign attention-dependent tasks to subjects using HearThere and measure their physiological responses. Recognizing these responses, in turn, would allow us to build an attention-driven display. Many of these concepts are not yet possible with our current hardware, but we include the larger conceptual vision here because we believe it to be attainable with our approach, and offer it as a contribution to the field in itself.

3. BACKGROUND

3.1 Sensing

Our vision of a networked sensory prosthetic depends on two major research components: ubiquitous sensing and attention-sensitive AAR display. This paper is largely concerned with the latter, though we present our approach to the former here briefly for context. Over the past three years, we and our collaborators have been building a sensor network and wireless infrastructure on a wetland restoration site in southern Massachusetts, part of a larger research initiative called the Tidmarsh Living Observatory. This network consists of approximately 100 low-power wireless sensor nodes across 3 distinct areas, each capturing a variety of ecological signals including microclimate, light, soil conditions, wind, and water quality. We are also streaming audio from microphones and hydrophones. The data are stored and re-streamed from an off-site server for use by end-user applications via an HTTP and WebSocket-based API.

3.2 Indoor Localization

Indoor localization is a field with abundant applications, and is a very active research area. There are also a variety of commercial products available. **Hightower and Borriello** describe many of the foundational works in the field and include a well-developed taxonomy [14].

Optical tracking systems such as **OptiTrack**[1] from NaturalPoint are currently popular. While these systems support precision on the order of millimeters, they are expensive, difficult to scale, and have no way to distinguish individual markers. Recently Valve Corporation has introduced their **Lighthouse** system which scans a laser line through the tracked space, similar to the **iGPS** system from Nikon Metrology, described by Schmitt et al. [23].

SLAM (Simultaneous Location and Mapping) is a camera-based optical approach that places the camera on the object to be tracked. Though this approach is attractive because it does not require infrastructure to be installed, it requires heavy computation in the tag. Another infrastructure free approach, **Chung et al.**'s geomagnetic tracking system, builds a database of magnetic field distortions and then at runtime attempts to locate the tag by finding the most similar database entry [9]. This approach is known as fingerprinting and has also been widely explored with ambient WiFi signals, generally with a precision on the order of 1 or more meters.

[1] http://www.optitrack.com/

3.2.1 *Ultra-WideBand*

UWB describes radio frequency (RF) signals with an absolute bandwidth of greater than 500 MHz or relative bandwidth greater then 20 %[10]. A wide frequency spectrum corresponds to a time domain signal with very short pulses and sharp transitions. This property is what makes UWB particularly suitable for measuring time-of-flight of RF pulses, even in the presence of reflections off of walls, floors, and objects in the area. Note that reflected signals can still be a source of error in cases where the direct signal is blocked by an obstacle, known as non-line-of-site (NLOS) conditions. In these cases the receiver can mistake a reflected signal for the direct, which over-estimates the range.

Previous work has investigated combining GPS and UWB to cover both outdoor and indoor localization with promising results [12, 8]. We chose this approach for HearThere because it offered the best balance of performance, cost, and scalability. We use what is known as Symmetric Double-Sided Two-Way Ranging (SDS-TWR) to determine the distance between our *Tag* and several fixed *Anchors* that are at known locations. In this scheme a total of 2 round-trip exchanges occur between the tag and anchor, allowing us to compensate for clock drift between the two devices[1]. Without noise we could then solve for the location of the tag analytically using trilateration. Given that noisy signals are inevitable however, there is often no analytical solution to the trilateration problem, so we have implemented the particle filter described in the “Particle Filter Server” section.

While this approach works with for a single tag and small number of anchors, each ranging measurement takes four messages (three for the ranging and one for the anchor to report back the calculated range), and ranging to each anchor must be done sequentially, which adds error if the tag is in motion. Future work will implement a Time-Difference-of-Arrival (TDOA) approach which will only require a single outgoing message from the tag that will be received by all anchors within communication range.

3.3 Auditory Augmented Reality

Azuma[3] provides a simple and useful definition of Augmented Reality: it combines real and virtual, is interactive in real time, and is registered in 3-D. The third criterion is useful for separating *Auditory Augmented Reality* (AAR) from *Location-Based Sound* (LBS). The key difference is that in LBS the sound cannot be said to be registered to a particular location in 3D space. For example, **Audio Aura** [21] is an LBS system in an office environment, but not augmented reality audio because the sounds are simply triggered by the user’s location and played through headphones. Similarly, **ISAS** [5] presents spatialized audio content with a defined location in 3D space, but uses the user’s mobile to determine orientation rather than the user’s head.

Loco-Radio [15] uses a mobile phone mounted to the user’s head for orientation tracking and the aforementioned location tracker from Chung et al. for a location precision of about 1 m, updated at 4 Hz . **LISTEN** [30] includes an authoring system and focuses on context-awareness, providing content based on individualized profiles and inferences based on the user’s behavior. At SIGGRAPH 2000, AuSIM Inc. presented **InTheMix** [7], an installation with responsive musical content spatialized using HRTFs and room modeling. Their system used a number of commercial tracking systems and was limited to a 4 m radius circle; the user was tethered for audio and tracking purposes.

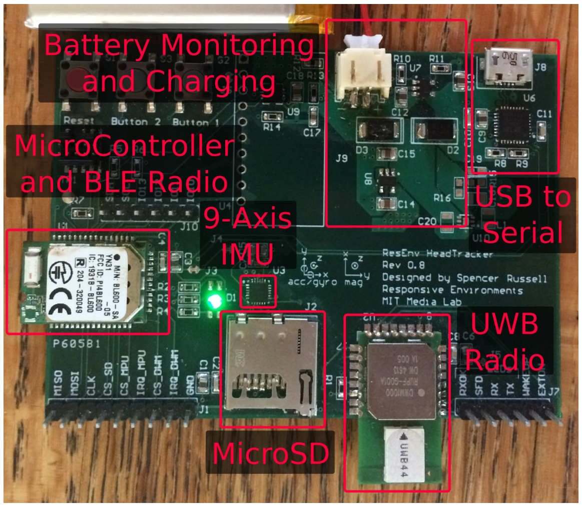

Figure 1: Overview of the HearThere hardware

3.4 Spatial Audio Delivery and Perception

AAR systems often use standard in-ear or over-ear headphones, which interferes with the user’s perception of the world around them. **Härmä et al.** present a system [13] that includes what they refer to as *hear-through* headphones integrate binaural microphone capsules into a pair of in-hear headphones. There have been several commercial products that have added orientation tracking to traditional headphones for virtual surround sound, including the **DSPeaker HeaDSPeaker**, **Smyth Research Realiser A8**, **Beyerdynamic DT 880 HT**, and the **Sony VPT**.

It has been shown that head motion plays an important role in our ability to localize sound [28, 25, 20], particularly in reducing front/back confusion errors. **Brimijoin** and **Akeroyd** [6] showed that as the test signal bandwidth goes from 500 Hz to 8 kHz, spectral cues become as important as head movement cues (in situations where they are contradictory). Even experiments that don’t allow head movement assume that the head orientation is *known*. Without this information the system is simply guessing. Head tracking is therefore a requirement for our application.

Excessive latency is detrimental to our ability to localize [22]. Azimuth error is significantly greater at 96ms latency than 29ms, and latency has a greater effect than update rate or HRTF measurement resolution.

Several studies have tried to measure these suitability of bone conduction headphones for spatialized content, though none have been particularly conclusive. **MacDonald et al.** [16] found that localization performance using the bone conduction headphones was almost identical to a pair of over-ear headphones. However, their measurements were coarse-grained (in 45° increments), proving suitability only for very basic localization. As part of the **SWAN**[29] project, Walker et al. evaluated navigation performance when following spatialized audio beacons using bone conduction headphones[27]. While performance was somewhat degraded from previous work with traditional headphones, the study at least confirms that unmodified HRTFs presented through bone conduction headphones can support basic spatialization.

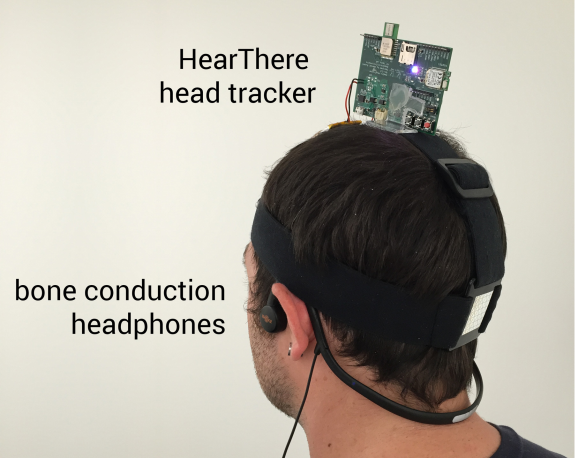

Figure 2: HearThere head tracker worn with bone conduction headphones

4. SYSTEM DESCRIPTION

4.1 Hardware

The HearThere HeadTracking hardware is the basis for our head tracking system, communicating with the wearer's mobile device over Bluetooth Low-Energy (BLE) and ranging to UWB anchors in the infrastructure. It is designed in a development board form-factor and optimized for development and testing. The main microcontroller is the nRF51822 from Nordic Semiconductor, which also handles BLE communication with the host. It communicates with the InvenSense MPU-9250 IMU and the DecaWave DWM1000 UWB module over the SPI bus. It includes several buttons and an RGB LED for user feedback. Users can switch between indoor (without magnetometer) and outdoor (with magnetometer) modes by pressing a button. The board is powered by a rechargeable LiPo battery and includes a battery monitoring chip. It has an on-board SD card slot for data logging. The next version of the board will be made significantly smaller.

4.2 Firmware

The firmware for the HearThere head tracker is written in C and runs on the nRF51822 chip, which is built around an ARM Cortex-M0. The main tasks of the firmware are:

- continually ranging to all available anchors
- reading from the IMU and running the Madgwick sensor fusion to compute an orientation estimate
- maintaining a BLE connection to a mobile device or PC (the host)
- notifying the BLE host of updated range and orientation information

HearThere uses a simple cooperative task scheduling design, in which each module has a tick function that is called from the main loop. Each module is responsible for maintaining their own state machine and in general the modules avoid busy-waiting so that other tasks can run.

Minimizing latency was a driving design factor, and one of the tightest latency deadlines came from managing the UWB ranging process. The ranging process can't wait more than 400 us. Because our processor does not have a hardware floating point unit, each iteration of the Madgwick algorithm takes 2.8 ms. We refactored the IMU Sensor Fusion algorithm into a state machine that breaks up the computation into separate pieces that can be run in successive function calls. We estimated that without partitioning the fusion calculations, we would need to slow down our ranging rate to under 3 Hz to make our deadlines. With partitioning we estimated we could run at 16.7 Hz, and in practice we were able to get 15 Hz. All tests were run while reading from the IMU and updating the sensor fusion algorithm at 200 Hz, and sending updated orientation over BLE at approximately 35 Hz to 40 Hz. In later experiments the Anchor range update rate was reduced to 7-10 Hz to ensure more reliable operation due to more timing headroom.

4.3 IMU Sensor Fusion

The HearThere head tracker relies on a MEMS inertial measurement unit (IMU) chip from InvenSense called the MPU-9250. It provides a 3-axis gyroscope (measures angular velocity), 3-axis accelerometer (measures a combination of gravity and translational acceleration), and 3-axis magnetometer (measures the local magnetic field vector).

In theory, with a starting orientation we could simply integrate the gyroscope signal to compute our orientation. In practice this method is hindered by gyroscope noise, which after integration becomes a random walk that causes our orientation estimate to gradually drift. The search for methods for correcting this drift by combining the available sensor data (sensor fusion) has been an active research area dating at least to the inertial guidance system development of the mid-twentieth century [17], and common approaches include complementary filters, extended Kalman filters, and unscented Kalman filters. HearThere uses the Madgwick algorithm [18] based on prior success [19] and the availability of efficient C-language source code that could be run on our microcontroller. One important note is that all rotations are represented as quaternions, to avoid gimbal lock and singularity instability.

Our iOS application also has a *ReZero* button that the user presses while looking directly north with a level head to determine set an initial heading in absence of a magnetometer reading. This operation typically happens at the beginning of an interaction, but can be repeated if the orientation estimate drifts.

4.4 iOS Software

The HearThere iOS application is built using the Unity3D game engine and written in C$^\sharp$. Though the main user-facing app is intended to focus on sound, we have implemented several features to display various system metrics, transmit raw and processed data over OpenSoundControl (OSC), and provide manual adjustment tools for testing purposes. The app manages the BLE connection and receives data, updating the orientation and location estimates. It synchronizes the in-game listener to the real-world user's head and places the virtual audio sources in the game world to create the binaural rendering. The app can display the user's position on a map, with tiles downloaded on demand from Google Maps. In map mode the app also renders virtual objects representing the audio sources; users can drag the objects on the map to manually reposition the sources. We have developed a framework that allows designers to place auditory objects in the scene either with absolute GPS coordinates, or with a relative position in meters. This allows for instance, a designer to create place a whole scene based on its real-world GPS location, and then place objects within that scene using their locations measured relative to the scene origin.

4.4.1 *Audio Engine*

While Unity provides a sophisticated authoring environment for placing sonic objects in our world, the built-in spatialization is very basic. It only models interaural level difference (ILD) and optionally a simple lowpass filter to approximate occlusion effects for sources behind the listener. With the recent resurgence of virtual reality, there are a number of more sophisticated spatial audio engines now available. We are using the 3DCeption plugin from Two Big Ears[2] which uses a generalized head-related transfer function (HRTF) that captures interaural level and time differences, as well as spectral cues. They also implement a simple shoebox model to generate physically-plausible first-order reflections from the environment.

4.5 Particle Filter Server

In principle it is possible to use ranging data from a number of different fixed locations and analytically solve for tag location, a process known as multilateration. In the presence of noise this problem becomes much more difficult, although many approaches are viable. We chose to solve the problem using a particle filter[24], which has been shown to perform well specifically in UWB-based localization systems [11]. Particle filters are particularly attractive because they provide a straightforward way to use measurable statistics (such as the variance of ranging data) to create a likelihood model that can generate a complex probability distribution.

We implemented our particle filter system in the Julia programming language[4] and added an HTTP interface. Clients can make an HTTP request to the root of the web server, which initializes a fresh set of particles and sends a response to the client with a link that they can use to submit sensor updates.

5. SYSTEM VALIDATION AND USER STUDY

5.1 Technical Evaluation

Our initial evaluation was focused on measuring and validating the tracking performance of the HearThere system. We collected data from the system while the user's head was also instrumented with optical motion capture markers tracked by a six-camera OptiTrack motion-capture system. Figure 3 shows an overhead view of the configuration with the cameras and anchors in a level plane near the ceiling (2.4 m from the floor), along with the path from the OptiTrack data projected onto the X-Z plane.

5.1.1 *Ranging Accuracy*

First we compare the raw range data measured over UWB to our expected range (computed from distances between OptiTrack's reported head location and the anchor locations). A range error histogram is shown in Figure 4, showing a mean error of 1.87 cm with standard deviation of 10.17 cm. We see that the HearThere UWB system tracks the expected ranges very well. Despite our calibration we have a mean error bias of 1.87 cm. This is most likely because of errors in calibration or inaccuracies in our anchor location measurements. The standard deviation of our range error is 10.17 cm, which is in line with our expectations of the DecaWave capabilities.

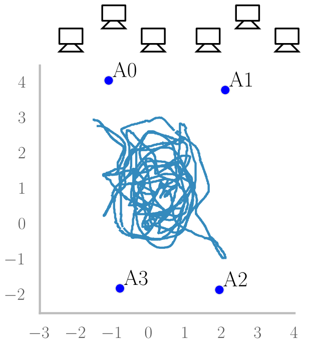

Figure 3: Experimental setup with six OptiTrack cameras and four UWB anchors, with the path walked during the experiment. Units are in meters.

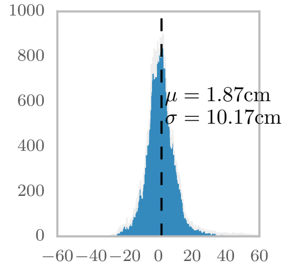

Figure 4: Overall ranging error histogram (in cm)

		Mean (cm)	Std. Dev. (cm)
Tracking	x	0.21	10.48
	y	37.21	18.08
	z	4.41	7.61
Overall	x	0.14	45.83
	y	30.25	28.51
	z	11.54	40.53

Table 1: Particle filter tracking error

5.1.2 *Localization Accuracy*

To evaluate HearThere's localization accuracy we used the measured range data and ran it through our particle filter implementation. Though these measurements were done offline, the data was processed sequentially using the same algorithm that our particle filter server uses, so the performance here should be representative of the algorithm running in real time. Runtime of the algorithm is substantially faster than real-time (on a 2011 MacBook Air), so compute performance is not an issue. Figure 5 shows the position measured by the OptiTrack system compared against 20 runs of the particle filter on our measured range data, to indicate the variance introduced by the nondeterministic algorithm.

From this data it is clear that while the filter is capable of tracking the location most of the time, it loses track during some portions of the test. Table 1 shows the error statistics for the tracking data. We analyzed both the overall performance, as well as the performance during the times when the filter was tracking well (approximately 20 s to 100 s).

5.1.3 *Discussion*

By comparing to the OptiTrack data we can see that the UWB hardware is successfully ranging between the tag hardware and the anchors, within expected errors. By feeding those range values into our particle filter-based tracking algorithm we show our ability to compute the 3D location of

[2] https://twobigears.com/

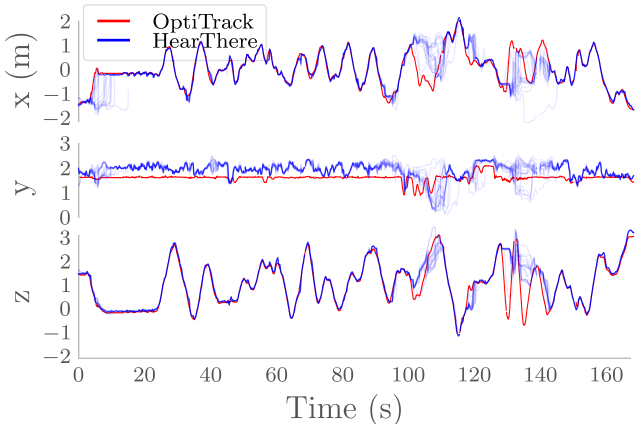

Figure 5: Location from OptiTrack compared to 20 runs of the particle filter fusion algorithm on the HearThere ranging data

the user's head. From Figure 5 we see that once the particle filter has locked on, error in the X and Z axes is near the ranging error, but the Y (vertical) error is greater. This is not surprising because the anchors are coplanar, not very far above the user's head (2.4 m from the floor). This means that small errors in range are magnified to larger errors in height. Though future work will focus on improving the tracking algorithm, it is encouraging that it recovers from losing track.

5.2 Outdoor User Study

We conducted a user study to demonstrate end-to-end system functionality and evaluate the auditory augmented reality experience in multi-scale tracking conditions. We also demonstrate the viability of HRTF-based spatial audio using bone-conduction headphones. The study took place in an outdoor plaza on a university campus, in an area approximately 60 m by 25 m. The test configuration is shown in Figure 7, showing the UWB Anchors, virtual audio sources, and tracking data from GPS and the UWB system. The UWB anchors were set to cover part of the test area, outside of which the system would be forced to rely on GPS for localization. We conducted the study with six volunteers (4 male, 2 female) between the ages of 23 and 36, all of whom were students or researchers with some level of technical or design expertise.

5.2.1 Procedure

We selected 4 audio samples (a female voice, bird sounds, chickens, and a solo saxophone) to use as virtual audio sources in our experiment. The user's task was to walk around the test area with the HearThere head tracker and attempt to locate the sources purely by listening. The test was conducted in two phases, A and B, each with the four sources in different locations. A random selection of half the participants used in-ear headphones (Etymotic ER-4) in phase A, followed by bone conduction headphones (Aftershokz Sportz 3) in phase B; the rest used bone conduction in A and in-ear in B. The sounds were not modified to account for differences in playback equipment (bone conduction vs. in-ear), but the users were able to adjust the volume freely. The audio sources were placed so that sources 0-4 were within range of the UWB anchors, and sources 5-7 required the use of the mobile GPS.

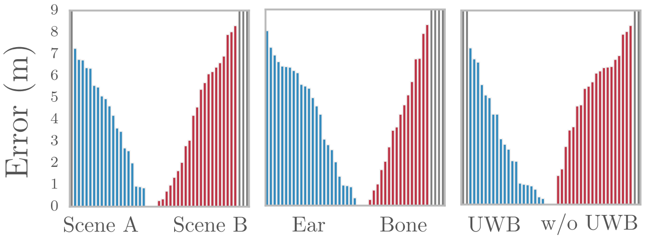

Figure 6: Localization error grouped by phase, headphone type, and whether the source was covered by UWB or not

During the tests the user placed a label on the ground where they thought the source was located, and after both phases we recorded the distance from the tape to the actual positions. The subjects knew the names of the sounds and in some cases some additional description was given, but they had not heard the sounds prior to the test. The results were marked *NA* either if the user was not able to decide on a location within the 10-minute time limit or if the error distance was greater than 8.5 m. After the task the user completed a short survey.

5.2.2 Subject Localization Error

Figure 6 shows the error data grouped by the phase (A vs. B), the headphone type (in-ear vs. bone conduction), and localization system (UWB vs. GPS). Average localization error was almost the same between phases A and B (4.2 m and 4.3 m). Error with the in-ear headphones was actually slightly higher than with the bone-conduction headphones(4.4 m and 4.0 m), though that result is somewhat biased towards the bone conduction because failures to localize are ignored in the average, and all four failures were with bone conduction. The most notable difference is between the sources localized with UWB and the ones solely relying on GPS (3.1 m and 5.3 m).

We qualitatively note a few things about these initial data. All four of the failures occurred with the bone conduction headphones. In three of those cases the stickers were not placed at all because the user was unable to hear the source. This is likely because the volume was much less with the bone conduction headphones, which was exacerbated by the overly-steep volume roll-off with distance. In one of the failures the user placed the marker tape but at a distance greater than 8.5 m, counting as a miss as well. Despite these challenges subjects were clearly still able to localize the sources that they could hear, and we expect performance would improve with sounds more tailored to bone conduction.

5.2.3 Discussion

The strongest result from the localization error is the difference between sources with and without UWB coverage. This supports the conclusion that users were able to use the higher-precision and lower-latency localization to get a more accurate idea of where the sources were. We expected that the task would be somewhat easier with the in-ear headphones than with bone conduction, as the bone conduction headphones have a more limited frequency response which can interfere with spectral cues, and also play at generally

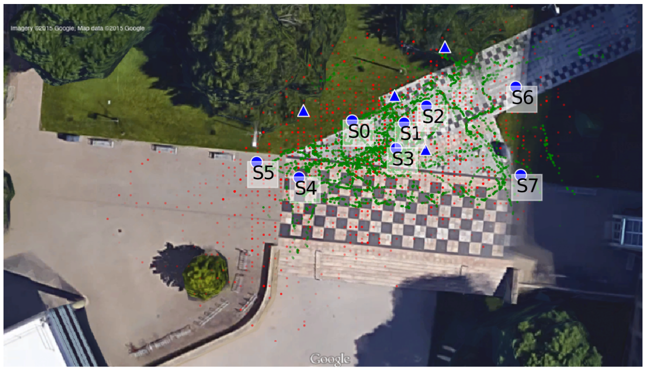

Figure 7: Tracking data accumulated for all users during the experiment. Triangles are UWB Anchors, Circles are audio sources. The green samples capture the output of the particle filter with the opacity representing the confidence. The red samples are GPS estimates

lower volume. Our survey results agreed with this hypothesis, with three reporting the task was easier with headphones, one with bone conduction, and two stating no preference. One user with previous spatial audio experience was able to localize sounds within approximately 1 m in the UWB zone but performed comparably to the other users in the GPS zone, indicating both that very high performance is possible and that skill plays an important role.

Five of six users reported confusing the real and virtual sounds at some point, particularly the birds. This is a sign that the spatialization and externalization are convincing, and supports our vision of HearThere as a device for sensory augmentation.

> When looking for the chickens, I couldn't help but look down as if I was searching a chicken, but for the voice sound I looked straight forward. some sounds caused an reaction *[sic]* as if someone appeared suddenly behind me.

User feedback offers several avenues for improvement in future work. Several subjects mentioned that the sounds would occasionally jump from one location to another, likely because of tracking discontinuities at the boundaries between UWB and GPS coverage. The jarring nature of these shifts could be mitigated with a slower filter on the location so that changes happen more smoothly. Subjects also mentioned that the volume differences between sources and low volume when using bone conduction made their task more difficult, indicating that more care should be taken to balance these levels in the future. Multiple users also noticed that the volume of the sounds dropped off too steeply with distance. The distance fading effects were exaggerated during this test in an effort to reduce distraction and help the users separate the sources, but this feedback indicates that a more natural rolloff would have been preferable. Users are attuned to the physical behavior of sound such as distance roll-off, so maintaining realistic distance effects is important. We also notice that the perceived realism of the sound is affected by that sound's plausibility in the real-world space, which opens interesting ground both for further design work and perceptual study.

With this hardware system and software framework in place we can begin to work on sound designs that cross more fluidly between scales, for instance covering a city with building-scale sounds for which GPS precision is sufficient, but including zones of local auditory objects that users can walk amongst, observe, and interact with. For example, a user in a UWB zone could perceive the air quality just across the street, where a truck is idling; a user gazing out a high-rise window might take an entire city block or even neighborhood into account.

6. CONCLUSIONS AND FUTURE WORK

In this paper, we articulated a vision of an attention-sensing wearable that would allow users to experience sensor data as sensory experience. This vision is motivated by two major goals: first, to remedy a fractured state of attention caused by existing mobile and wearable technologies; and second, to bring the distant or often invisible processes underpinning a complex natural environment into human consciousness. This approach promises a more expansive perceptual world, or umwelt [26], where sensor data becomes immediately attributable to extrinsic phenomena, externalized in the wearer's perception.

As a next step towards these goals, we developed a system that can represent data from sensors in realistic, scalable auditory augmented reality. Our system combines head tracking with spatial sound presented through bone conduction, enabling seamless mixing of real-world and virtual sound sources. To validate HearThere for the visionary use case, we conducted experiments that tested both the system's functionality and users' abilities to localize sources using bone conduction headphones. Subjects reported externalization as well as confusion between the virtual and real sound they were experiencing, indicating some degree of realism. HearThere can be used for both indoor and outdoor auditory display, supporting fine-grained tracking in areas of UWB coverage and GPS everywhere else.

Through our collaborative efforts to sonify ecological sensor data and a growing deployment of microphones, we have begun to integrate HearThere with our sensor network in the wild. After this field integration is complete we intend to test the end-to-end system with users. Finally, we have designed and are carrying out a set of experiments that use the HearThere hardware and a variety of additional sensors to capture physiological signals corresponding to subjects' top-down selective attention to spatial sound. If successful, these experiments will give HearThere the ability to sense and respond to its wearer's intent. While this goal remains in a speculative realm for now, we believe that in the near future, augmented humans will be able to freely extend their auditory perception into distributed microphones and sensors, listening across great distances, high in a tree, or into the deep.

7. REFERENCES

[1] Sources of Error in DW1000 Based Two-Way Ranging (TWR) Schemes. Application Note APS011, DecaWave, 2014.

[2] M. Auvray and E. Myin. Perception With Compensatory Devices: From Sensory Substitution to Sensorimotor Extension. *Cognitive Science*, 33(6):1036–1058, Aug. 2009.

[3] R. T. Azuma and others. A survey of augmented reality. *Presence*, 6(4):355–385, 1997.

[4] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah. Julia: A Fresh Approach to Numerical Computing. Nov. 2014. *arXiv: 1411.1607*.

[5] J. R. Blum, M. Bouchard, and J. R. Cooperstock. What's around me? Spatialized audio augmented reality for blind users with a smartphone. In *Mobile and Ubiquitous Systems: Computing, Networking, and Services*, pages 49–62. Springer, 2012.

[6] W. O. Brimijoin and M. A. Akeroyd. The role of head movements and signal spectrum in an auditory front/back illusion. *i-Perception*, 3(3):179–181, 2012.

[7] W. L. Chapin. InTheMix. Siggraph, 2000.

[8] D. S. Chiu and K. P. O'Keefe. Seamless outdoor-to-indoor pedestrian navigation using GPS and UWB. In *Proceedings of the 21st International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS 2008), The Institute of Navigation*, volume 1, pages 322–333, 2008.

[9] J. Chung, M. Donahoe, C. Schmandt, I. Kim, P. Razavai, and M. Wiseman. Indoor location sensing using geo-magnetism. In *Proceedings of the 9th international conference on Mobile systems, applications, and services*, pages 141–154. ACM, 2011.

[10] S. Gezici, Z. Tian, G. Giannakis, H. Kobayashi, A. Molisch, H. Poor, and Z. Sahinoglu. Localization via ultra-wideband radios: a look at positioning aspects for future sensor networks. *IEEE Signal Processing Magazine*, 22(4):70–84, July 2005.

[11] J. González, J. Blanco, C. Galindo, A. Ortiz-de Galisteo, J. A. Fernández Madrigal, F. A. Moreno, and J. L. Martínez. Mobile robot localization based on Ultra-Wide-Band ranging: A particle filter approach. *Robotics and Autonomous Systems*, 57(5):496–507, May 2009.

[12] J. Gonzalez, J. L. Blanco, C. Galindo, A. Ortiz-de Galisteo, J. A. Fernández Madrigal, F. A. Moreno, and J. L. Martinez. Combination of UWB and GPS for indoor-outdoor vehicle localization. In *Intelligent Signal Processing, 2007. WISP 2007. IEEE International Symposium on*, pages 1–6. IEEE, 2007.

[13] A. Härmä, J. Jakka, M. Tikander, M. Karjalainen, T. Lokki, J. Hiipakka, and G. Lorho. Augmented reality audio for mobile and wearable appliances. *Journal of the Audio Engineering Society*, 52(6):618–639, 2004.

[14] J. Hightower and G. Borriello. Location systems for ubiquitous computing. *Computer*, (8):57–66, 2001.

[15] W. Li. *Loco-Radio: designing high-density augmented reality audio browsers*. PhD thesis, Massachusetts Institute of Technology, 2013.

[16] J. A. MacDonald, P. P. Henry, and T. R. Letowski. Spatial audio through a bone conduction interface: Audición espacial a travės de una interfase de conducción ósea. *International Journal of Audiology*, 45(10):595–599, Jan. 2006.

[17] D. A. MacKenzie. *Inventing accuracy: A historical sociology of nuclear missile guidance*. MIT press, 1993.

[18] S. O. Madgwick, A. J. Harrison, and R. Vaidyanathan. Estimation of IMU and MARG orientation using a gradient descent algorithm. In *Rehabilitation Robotics (ICORR), 2011 IEEE International Conference on*, pages 1–7. IEEE, 2011.

[19] B. Mayton. WristQue: A Personal Sensor Wristband for Smart Infrastructure and Control. Master's thesis, Massachusetts Institute of Technology, 2012.

[20] P. Minnaar, S. K. Olesen, F. Christensen, and H. Moller. The importance of head movements for binaural room synthesis. In *Proceedings of the 7th International Conference on Auditory Display (ICAD2001), Espoo, Finland, July 29-August 1, 2001*, 2001.

[21] E. D. Mynatt, M. Back, R. Want, M. Baer, and J. B. Ellis. Designing audio aura. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 566–573. ACM Press/Addison-Wesley Publishing Co., 1998.

[22] J. Sandvad. Dynamic Aspects of Auditory Virtual Environments. In *Audio Engineering Society Convention 100*, May 1996.

[23] R. Schmitt, S. Nisch, A. Schönberg, F. Demeester, and S. Renders. Performance evaluation of iGPS for industrial applications. In *2010 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pages 1–8, Sept. 2010.

[24] S. Thrun, W. Burgard, and D. Fox. *Probabilistic Robotics*. The MIT Press, Cambridge, Mass, Aug. 2005.

[25] W. R. Thurlow and P. S. Runge. Effect of induced head movements on localization of direction of sounds. *The Journal of the Acoustical Society of America*, 42(2):480–488, 1967.

[26] J. Von Uexküll. A stroll through the worlds of animals and men: A picture book of invisible worlds. *Semiotica*, 1992.

[27] B. N. Walker and J. Lindsay. Navigation performance in a virtual environment with bonephones. In *Proceedings of the 2005 International Conference on Auditory Display (ICAD2005), Limerick, Ireland (July 6-10)*, 2005.

[28] H. Wallach. The role of head movements and vestibular and visual cues in sound localization. *Journal of Experimental Psychology*, 27(4):339, 1940.

[29] J. Wilson, B. N. Walker, J. Lindsay, C. Cambias, and F. Dellaert. Swan: System for wearable audio navigation. In *Wearable Computers, 2007 11th IEEE International Symposium on*, pages 91–98. IEEE, 2007.

[30] A. Zimmermann, A. Lorenz, and S. Birlinghoven. LISTEN: Contextualized presentation for audio-augmented environments. In *Proceedings of the 11th Workshop on Adaptivity and User modeling in Interactive Systems*, pages 351–357, 2003.