Sensor(y) Landscapes:
Technologies for New Perceptual Sensibilities

by

Gershon Dublon

B.S., Yale University (2008)
S.M., Massachusetts Institute of Technology (2011)

Submitted to the School of Architecture and Planning
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Media Arts and Sciences

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2018

Author . . . .
Program in Media Arts and Sciences
May 4, 2018

Certified by . . . .
Joseph A. Paradiso
Alexander W. Dreyfoos (1954) Professor of Media Arts and Sciences
Thesis Supervisor

Accepted by . . . .
Tod Machover
Academic Head, Program in Media Arts and Sciences

Sensor(y) Landscapes:
Technologies for New Perceptual Sensibilities

by

Gershon Dublon

Submitted to the School of Architecture and Planning
on May 4, 2018, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Media Arts and Sciences

Abstract

When we listen closely, there is a pervading sense that we could hear more if we could only focus a little more intently. Our own perceptual limits are a moving target that we cannot delineate and rarely reach. This dissertation introduces technologies that operate at that mysterious boundary. I envision *sensor(y) landscapes*, physical sites that meld distributed sensing and sensory perception to afford new perceptual sensibilities. Today's mainstream technologies are well designed for rapid consumption of information and linear, sequential action. A side effect of their effectiveness to task, however, is a loss of undirected, curiosity-driven exploration in the world. I propose alternative technologies that would extend perceptual presence, amplify attention, and leverage intuitions.

My focus is on turning rich sensor data into compelling sensory input, and as such, a substantial component of my work involved deploying sensor infrastructure in beautiful places. My projects center on a wetland restoration site, called *Tidmarsh*, where environmental data are densely and continuously collected and streamed. Using sound and vibration as the medium and nature as the setting, I undertook this work in two steps. The first constructs environments suffused with sensing and built for being present in. My projects in this space comprise sensor-driven virtual worlds, glass elevator sound installations, and vibrating forests that give oral histories. Building on lessons and infrastructure from the first approach, my culminating work uses non-occluding spatial audio to create situated perceptions of data. I developed a bone-conduction headphone device, called HearThere, that renders a live soundscape from distributed microphones and sensors, fully merged with the user's natural hearing. HearThere combines its wearer's inferred listening state with classification output from an AI engine to adjust the mix and spatial parameters of virtual audio sources. The device was developed based on findings from lab studies into spatial hearing and attention, and evaluated in a human subjects study with a panel of experts.

Through these projects, I found that deriving meaning in the medium is a matter of possessing or developing perceptual sensibilities, intuitions for how the permeated data can be teased out and contemplated. Carefully composed perceptual confusion—a blurring of place and distributed media—becomes an opportunity for the development of new *transpresence sensibilities.* How do users make sense of these new dimensions of perception, and how can technologies be designed to facilitate perceptual sense-making?

Thesis Supervisor: Joseph A. Paradiso
Title: Alexander W. Dreyfoos (1954) Professor of Media Arts and Sciences

This doctoral thesis has been examined by a Committee as follows:

Professor Joseph A. Paradiso . . . .
Thesis Supervisor
Alexander W. Dreyfoos (1954) Professor of Media Arts and Sciences
Massachusetts Institute of Technology

Professor Hiroshi Ishii . . . .
Member, Thesis Committee
Jerome B. Wiesner Professor of Media Arts and Sciences
Massachusetts Institute of Technology

Glorianna Davenport . . . .
Member, Thesis Committee
Visiting Scientist, MIT Media Lab
Founder, Living Observatory

Acknowledgments

I owe this work to the support, collaboration, friendship, and love of so many people throughout my time at MIT and before.

To my advisor, Joe Paradiso, who gave me his trust and the freedom to pursue crazy ideas; who encouraged me to instill the work with my values, and stood up for his too; who through countless conversations helped me turn my ideas into projects, and projects into research; who peppered our conversations with music history and political debates, high energy physics and science fiction; who saw the benefits of having creative hobbies and wacky side projects; who fomented restorative jam sessions when I needed them most; and who filled the Responsive Environments group with all the exceptionally smart and generous people who have become my extended family at the Media Lab, thank you.

To my thesis committee, Glorianna Davenport and Hiroshi Ishii, whose unique perspectives are imbued throughout the dissertation and will be with me always. Glorianna saw the potential for my work to engage with environmental restoration before I did, and worked tirelessly to spark that connection in my work and many others'. It was Glorianna's belief in the work that I turned to when it felt insurmountable, and that propelled me forward at the end with her beautiful accounts of her experiences using HearThere. In many ways, she now knows the project better than I do. Hiroshi was a member of my Master's thesis committee, and reprised his role for the dissertation. Hiroshi's exacting standards and the seriousness with which he takes my work constantly remind me that there is value in what I'm doing, because otherwise it would not be worth his time to so heartily critique it. His emphasis on a *crisp* articulation of my vision and conclusions pulled the work higher at its final stages, and his energy reflects academia at its best. Thank you Glorianna and Hiroshi for your wisdom, inspiration, and support.

To my Tidmarsh collaborators: Brian Mayton, Don Derek Haddad, Spencer Russell, Clement Duhart, and Evan Lynch. In pursuit of the vision we shared, Brian and I spent countless hours in blistering heat and frigid cold, sinking into boggy depths and climbing into trees with microphones in tow. At times we wondered how it was that two electrical engineers with no knowledge of ecology could end up at its cutting edge, literally hacking through thorny briars with repurposed wire cutters. I truly owe this work to Brian more than anyone. Don, Spencer, Clement, and Evan you were frequently out there with us, hacking away, and your theses are inextricably embedded in all the projects of my dissertation. Finally, we brought many folks along on our expeditions to paddle our boats, haul our cable, hammer our stakes, and dig our trenches. Thank you to everyone who helped bring Responsive Environments into the great boggy outdoors.

To my many lab mates and collaborators, with whom I spent countless sleepless nights trying to make a deadline, dreaming up a NIME project, throwing blowout parties, running through the dark halls of an immersive theater for a mid-show fix, or considering the acoustic properties of 3D printed bunny hair. In particular, I want to thank Brian Mayton, Spencer Russell, Clement Duhart, Evan Lynch, Don Derek Haddad, Edwina Portocarrero, Rebecca Kleinberger, Nick Joliat, Nan Zhao, Pol Pla, and Xin Liu for their direct contributions to the work in this thesis as collaborators and/or their support in finally getting it done. And thanks to my many mentors, teachers, advisors, and wise bystanders, among them: Peter Kindlmann, Mark Feldmeier, Joi Ito, Wendy Jacob, Kevin Slavin, Tod Machover, bunnie, and Pattie Maes.

Thanks to the Living Observatory and its many partners and contributors: Tidmarsh Farms, Glorianna Davenport, Evan Schulman and the Schulman family, Alex Hackman, MA Department of Ecological Restoration, Brian Spires, Jessica Norriss, Charis Durrance, Rob Vincent, Kate Ballantine, Christine Hatch, US EPA, USDA-NRCS, USFWS, NOAA, MA Environmental Trust (MET), MA Fish and Game, the Town of Plymouth, Bob Wilber, and Mass Audubon.

Thanks to my HearThere field study participants, who braved chilly mornings and mosquito-infested afternoons: Glorianna Davenport, Kate Ballantine, Jason Andras, Deborah Cramer, Chris Ickler, Bridget Hanson, Dan Gauger, Alex Hackman.

Thanks to the MIT Media Lab staff and the director's office. The facilities team, in particular Greg Tucker, Jessica Tsymbal, Kevin Davis, Cornelle King, Andrew Baird, Candido Monteiro, and Bill Lombardi have always been there to help make whatever weird thing we dreamed up pass inspection. Thanks to Tom Lutz, Seth Avecilla, and John DiFrancesco for putting up with our machine shop antics, keeping us safe, and helping us get the job done. Thanks to NECSYS for both enabling our free and open access and protecting us from the jaws of MIT's increasingly corporate network. Thanks to the video dream team and overall rays of sunshine Paula Aguilera and Jonathan Williams, and to Jimmy Day. Thanks to the Program in Media Arts in Sciences, for whom I was always a challenging case, but who squeezed me through in the end: Linda Peterson, Keira Horowitz, Monica Orta, Amanda Stoll, Kayla Williams, and Aaron Solle. Thanks to Amna, who not only keeps the Responsive Environments group running like a well-oiled machine, but also keeps us warm, safe, happy, fed, and properly observant of our own birthdays.

Thanks to the members of the MIT Media Lab, as well as the MIT Energy Initiative, Intel, and the Center for Terrestrial Sensing, all of which supported me through fellowships during my doctoral work. Thanks to the Living Observatory for its material support of the Tidmarsh sensing infrastrucure, which was pivotal to this work.

Finally, thanks to friends who not only put up with me through this long journey, but also understood and supported me. Thanks to Jessica Goldin, Sean Flynn, Alex Charrow, Tom Laakso, Spencer Russell, Kate Millington, Eleanor Roosevelt-Russell-Millington, Nan Zhao, Brian Mayton, Ben Bloomberg, Jifei Ou, Chelsea Barrabas, Edwina Portocarrero, Pol Pla i Conesa, Nan-Wei Gong, Chia-Wei Lin, Meow-Wei Gong-Lin, Asaf Azaria, Liz Carlin, Gabe Miller, Don Derek Haddad, Jie Qi, Nicole L'Huillier, Juan Necochea, Amanda Carlin, Adam Haar Horowitz, Xiao Xiao, Darthur Petron, David Cranor, Akito van Troyer, Christine Sun-Kim, Rebecca Kleinberger, Joi Ito, Nick Gillian, Jen Jacobs, Azza Gadir, Katia Vega, Laurel Pardue, Orcun Gogus, Pip Mothersill, Danielle Nadeau, Sandra Richter, Natan Linder, Sands Fish, Alexis Hope, Mar Gonzalez-Franco, Nathalie van Bockstaele, Clovis Theilhaber, and to my family: Dina Dublon, Giora Dublon, Amalle Dublon, Tina Zavitsanos, and Xin Liu; and to my extended family around the world, who ought to being seeing more of me now.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Preface: Listening-Looking

I hear across impossible boundaries—through closed windows and concrete walls, over considerable distances and astronomical durations. Not all the time, thankfully. That would be too much to handle. Only if I concentrate hard enough, and only if the conditions are right. Sometimes, if there's enough there to latch onto, I can grasp a vibration like a string, let it ring, pull it closer, jump the boundary. I can reel in slow settling buildings and melting ice, something rustling in the trees. I pick up alternating currents, private conversations, bird migrations, gravitational waves. I can't help but make a very particular face when I do it, eyes and mouth open soft, a kind of *listening-looking.* You know the one, so you'd probably know I'm up to something. That's no secret anyway. But I've never told anyone how good I am at it. They wouldn't believe me. They'd say everyone does it. Maybe they're right, maybe there are others. But it's my superpower, and I want you to feel it too.

I have to level with you though. For quite some time I've been interested in your listening-looking face. I've imagined all kinds of elaborate scenarios and invented justifications for research and experimentation that might elicit it. I've measured my work by it. Your listening-looking face captivates me because it embodies "a perception that can be both an absorption and an absence," as Jonathan Crary writes; your attention, "a sense of 'tension,' of being 'stretched,' and also of 'waiting'"; and together, ringing in all their contradictory resonances, "a suspension of perception":

> a looking or listening so rapt that it is an exemption from ordinary conditions,

> that it becomes a suspended temporality, a hovering out of time…It implies the possibility of a fixation, of holding something in wonder or contemplation, in which the attentive subject is both immobile and ungrounded. But at the same time, a suspension is also a cancellation or an interruption…a disturbance, even a negation of perception itself. [19]

So there they are—dual, interconnected motives: to give you the feeling of my sensory superpower; and to explore the beautiful contradictions I see in your face when you feel it.

1.2 Sensor(y) Landscapes

> Presence as access is as real as presence gets, and that's real enough.
>
> Alva Noë [94]

When we listen closely, there's a sense that we could always listen a little closer, go deeper, pull out more detail— if we could only focus our attention just so. We don't know our own perceptual limits because they are a moving target, and we rarely, if ever, reach them. In those rare moments of deliberate perception, we find ourselves touching a boundary we cannot delineate. This dissertation introduces technologies that meet us at that mysterious boundary. In pursuit of that feeling of rapt listening, its inner corporeality and its outer fringes, I developed a set of networked, sensed, activated, and vibrated landscapes, spaces, and devices to cultivate a relationship between technology and perceptual presence. In that relationship, I envisioned new sensory landscapes, places that meld distributed sensing and sensory perception. Once on those landscapes, I found new perceptual sensibilities, extensions of existing perceptual abilities and intuitions through information networks. I observed visitors describing their experiences as "sensory superpowers," a designation referencing both the permeability of perceptual limits and the joy of crossing them. In light of these experiences and observations, I propose a new category of *transpresence* technologies that would play a positive, restorative role in current and future "varieties of presence"[1], and discuss why we need new forms of technology to do so.

Notions of human "augmentation" by networked technologies are not new to followers of Weiser's fast materializing ubiquitous computing vision [131]. From wristbands to bus stops, connected devices provide us immediate access to data across distances, times, and scales. On my body alone, my watch and I feel my heartbeat together, my headphones are listening machines, and my glasses see what I see. As mobile augmented reality comes into widespread use, devices are culling and reconfiguring our senses, intermediaries between us and the world, and often between our measured bodies and minds. While these tools bring powerful new capabilities, their far-reaching application across work and all aspects of life leads them to prioritize efficient consumption of information and task completion by

---

[1] Alva Noë's umbrella designation of "varieties of presence," from his 2015 book by that name, provides a helpful framework for talking about new or changing modes of perception. In the book, he dramatically (and controversially) expands his sensorimotor contingency theory, upon which much contemporary sensory augmentation and substitution research has drawn, to various forms of thought [94].

Figure 1-1: Two aerial views of one place before and immediately after wetland restoration: at left, the gridded water control structures of an industrial cranberry farming operation enforce monoculture; at right, a sinewy stream channel strewn with fallen trees will play host to countless interdependent microhabitats.

design. In their effectiveness to task, today's tools of ubiquitous computing are supplanting undirected, curiosity-driven exploration in the world.

Paradoxically, as real-time data about the world have grown, we find ourselves increasingly deprived of the pleasures and sensibilities of physical presence. Instead, we exist in a permanently fractured state of attention, straining to allocate limited attentional bandwidth between our devices and our surroundings. Sensors are all around us, capturing rich, 'sensory' data in our environments, but these data are digested into knowledge and presented in the overloaded and highly mediated technological channels through which we communicate, create, and consume. In more ways than one, this is a thesis about restoration. Views to nature restore attention [121], and I seek for the same kind of feeling I get exploring a forest, where every moment presents something to discover not as a demand on attention, but as a quiet invitation to it. Environmental restoration offers a useful metaphor here: where farmers manage ecosystems for productivity with tiled, gridded water control structures, like the one in Figure 1-1, restoration practitioners use non-linear paths to create variation, and in turn biodiversity and interdependence. Technology has similarly hued to grids optimized for well-defined tasks, trading the richness of slow, deliberate perception for higher levels of productivity and efficiency.

I posit a mode of experience characterized by a blurring of digital media and the physical world that would extend perceptual presence, amplify attention, leverage intuitions, and heighten perceptual sensibilities. The notion of a digitally modulated undirected experience is premised on my development of technologies akin to augmented reality that, as a matter of perception, are not layered atop the universe, but are *of the universe.* My approach grows from calm technologies and sensory assistive devices with a focus on sensor networks and other distributed media. The selective but seamless blending of spaces and times also has roots in telepresence technologies, but where telepresence is principally about transport in support of communication or remote action, the melding together of place and distributed sensing has altogether different aims and outcomes. *Transpresence* perceptions are built from a blurring together of times and places that can only be teased apart again through pre-existing or newly acquired intuition. I envision a device that translates sensor data streams into extended spatial perceptions. Experientially akin to glasses, the *networked sensory prosthesis* exists between the body and world, working to alter one's perception of their surroundings without becoming a site of attention in itself.

I found that certain compositions of digitally-produced perceptual confusion (caused by the blurring of digital media and the physical world) would be resolved first by exploration and conscious reasoning and then by users' intuitions into qualitatively new perceptual experiences. To test this, I used sound and vibration as a medium, audiovibratory perception as a model, and, in much of the recent work, nature as a setting. I demonstrate new and altered temporal and spatial dimensions of perception, which I refer to as *perceptual sensibilities*, emerging in users of my experimental devices.

Based on the literature as well my own projects and studies, I map out a new design space for constructing non-linear perceptual experiences out of spatially-distributed media and sensor data. *How can a disparate set of remote measurements translate into egocentric proximate stimuli? Alternatively, how does an environment suffused with sensing become sensory?* Dimensions of this space include cognitive processes such as attention and externalization, as well as temporality, simultaneity, perspective, realism, sensation, and gestalt. In my studies, I found new perceptual sensibilities arising from a combination of confusion, rational intuition, and interest, raising new questions about which compositions of perceptual confusion would lead to that result, and how data might be curated by balancing individual selective attention and shared artificial intelligence. These and other questions are revisited throughout the thesis and discussed in Chapter 5.

- What intuitions does a participant draw on to make sense of their observations, particularly in situations of likely confusion between physical and virtual perceptions?

For example, if a networked sensory prosthetic warps physical space or blends the present moment with the recorded past, what kinds of insights can the participant glean? (space, time, linearity)

- What is ‘content,’ and can it be independent of the modality of its presentation? What characteristics should a perceptual ‘message’ share with the physical medium to leverage one’s intuitions? How can existing perceptual schema be leveraged or extended, or new schema learned? How realistic should these systems be? (modality, medium, intuition, learning)
- To what extent can, or should, interactions be cognitive or even subconscious? What effects should sensorimotor ‘gestures’, such as plugging one’s ears, have on a virtual sensory modality? (attention, interaction, feedback)

The next chapters take the reader through the projects that I, together with various research partners, undertook to explore these questions. These include, from early on, virtual worlds that explored the presence-making capacities of sensor networks, physical installations that suffused digital media into built and landscaped environments, and more playful sense-altering wearable devices.

My ultimate line of inquiry would not be possible without a *site*—a sensed environment from which to draw information and on which to conduct these probes and experiments. My experiments are built on sensor networks, which themselves require enormous effort to design, deploy, and maintain. Just the same as any pleasing perception, a *good* networked sensory experience is built on rich and compelling sense data in and of the world. As such, a substantial and foundational part of this thesis involves work deploying sensor networks and technological infrastructure in beautiful places, undertaken with and in support of close collaborators[2]. Someday, sites like these will be everywhere; for the purposes of my work, we had to build one from scratch.

My dissertation work culminates in a project on a wetland restoration site, called Tidmarsh, where I sought to realize the vision of this thesis on a larger scale. Over years, we introduced a dense sensor network to continuously collect and stream ecological measurements and sound,

---

[2] The sensor network centered aspect of this work is the subject of a forthcoming dissertation from my friend and collaborator Brian Mayton, with whom the vision of this dissertation was conceived and carried out. Mayton’s research focuses on the technological underpinnings of a flexible, long-term environmental sensor network as a medium for exploratory ecological science and creative expression.

and experimented with technologies that make perceptually senseable the interconnected instants and long-term processes that unfold there.

Building on this infrastructure and lessons from earlier work, I developed a non-occluding spatial auditory display called HearThere. HearThere renders a live soundscape from distributed microphones and sensor data sonification, fully merged with the user's natural hearing. HearThere combines its wearer's inferred attention and listening patterns with classification output from an AI engine to adjust the mix and spatial parameters of virtual audio sources. The device was developed based on findings from lab studies into spatial hearing and attention, and evaluated in a human subjects study with a panel of experts.

The next section introduces the Networked Sensory Landscape vision that grew out of our environmental sensing and interfacing work, and which serves as the backdrop for much of what I will present later in the dissertation. Both the vision and a complete set of associated projects are more extensively elaborated in [86].

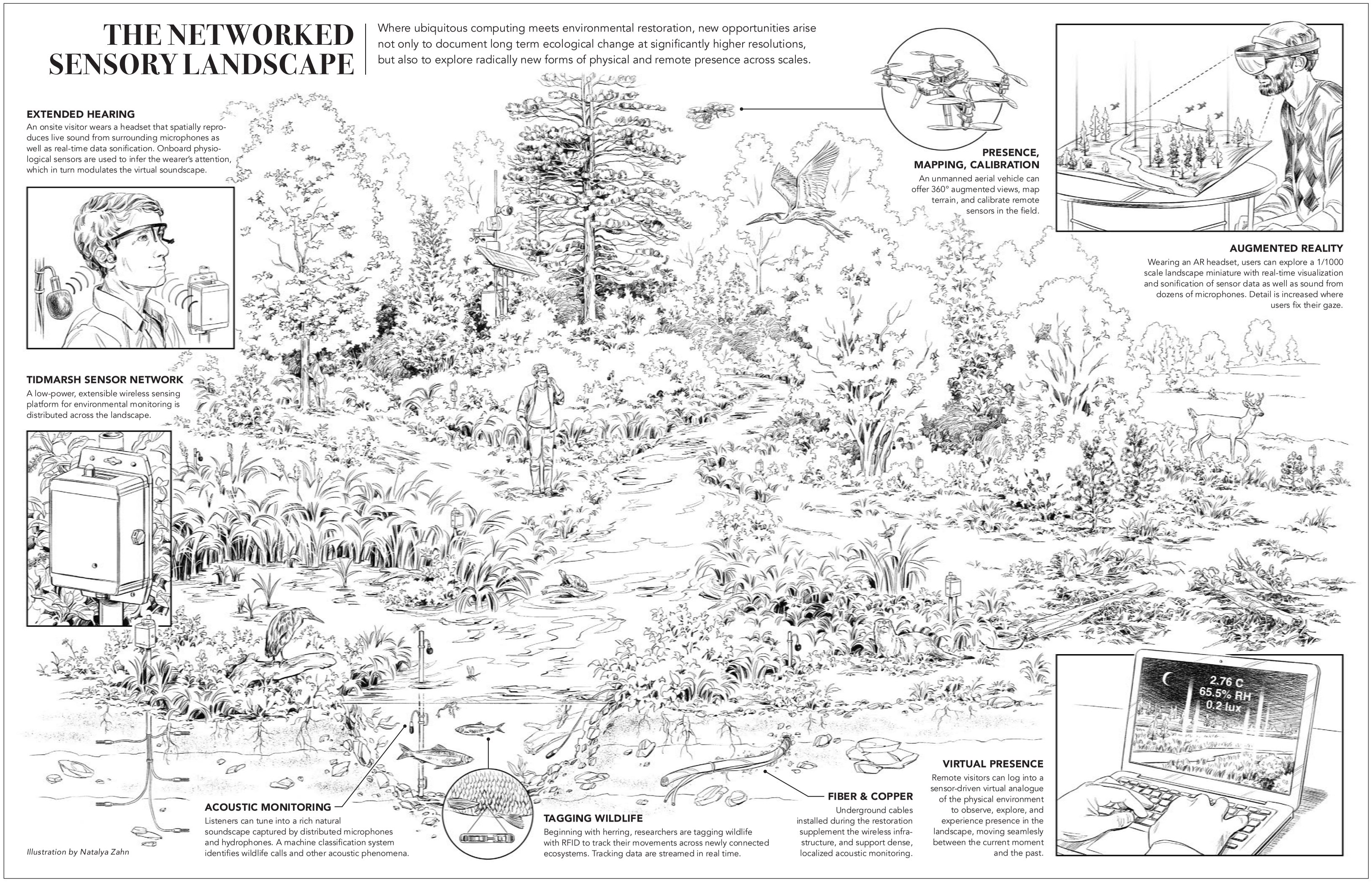
1.3 Site Vision: A Networked Sensory Landscape[3]

Landscape captures the complex exchange between the world we sense, the world we make, and the world we imagine. These worlds are often in tension, and perhaps no human endeavor captures this tension more than our pursuit of technology, the most significant driver of our impact on the environment. At the same time, technology provides our primary means of understanding the environment, preserving it, and expressing our relationship to it through art—from cave paintings to audio recordings of melting sea ice [52].

What is the role of ubiquitous sensing in the future of landscape? Since Szewczyk, et al. demonstrated the potential of wireless sensor networks as research tools in habitat monitoring [118], systems like theirs have been used with increasing frequency in primary ecological research, in conservation settings, and for agriculture [54, 79, 110, 129]. More than a decade on, in the era of mobile and ubiquitous computing, we are finding that environmental sensor networks embedded in the landscape can serve as a platform for a wide array of applications spanning research, outreach, and art. My dissertation work focuses on the intersection of presence and pervasive sensing as introduced in [28], and landscape is a natural site for this broad new field.

---

[3] This section contains modified excerpts from [86]. I only excerpted text I originally wrote, but substantial credit is due to my co-authors, who edited the text and wrote other parts of the article.



THE NETWORKED SENSORY LANDSCAPE

Where ubiquitous computing meets environmental restoration, new opportunities arise not only to document long term ecological change at significantly higher resolutions, but also to explore radically new forms of physical and remote presence across scales.

EXTENDED HEARING

An onsite visitor wears a headset that spatially reproduces live sound from surrounding microphones as well as real-time data sonification. Onboard physiological sensors are used to infer the wearer's attention, which in turn modulates the virtual soundscape.

TIDMARSH SENSOR NETWORK

A low-power, extensible wireless sensing platform for environmental monitoring is distributed across the landscape.

PRESENCE, MAPPING, CALIBRATION

An unmanned aerial vehicle can offer 360° augmented views, map terrain, and calibrate remote sensors in the field.

AUGMENTED REALITY

Wearing an AR headset, users can explore a 1/1000 scale landscape miniature with real-time visualization and sonification of sensor data as well as sound from dozens of microphones. Detail is increased where users fix their gaze.

ACOUSTIC MONITORING

Listeners can tune into a rich natural soundscape captured by distributed microphones and hydrophones. A machine classification system identifies wildlife calls and other acoustic phenomena.

TAGGING WILDLIFE

Beginning with herring, researchers are tagging wildlife with RFID to track their movements across newly connected ecosystems. Tracking data are streamed in real time.

FIBER & COPPER

Underground cables installed during the restoration supplement the wireless infrastructure, and support dense, localized acoustic monitoring.

VIRTUAL PRESENCE

Remote visitors can log into a sensor-driven virtual analogue of the physical environment to observe, explore, and experience presence in the landscape, moving seamlessly between the current moment and the past.

Illustration by Natalya Zahn

Figure 1-2: A vision for a landscape future at the intersection of ubiquitous computing and environmental conservation. The Networked Sensory Landscape presents opportunities to capture long-term ecological change at higher spatial and temporal resolutions than ever before, and to experiment with radically new forms of physical and remote presence across scales.

I call this field the *Networked Sensory Landscape*, comprising our own work developing an environmental sensor network, as well as its diverse applications and outcomes. What opportunities are created by weaving a continuously sampling, geographically dense web of sensors into the natural environment, from the ground up? And how can the data produced by this kind of network extend a renewed sense of presence to both onsite and remote visitors?

Currently the site of the largest freshwater wetland and riverine restoration in the state of Massachusetts, Tidmarsh serves as a shared field laboratory for both my work and a diverse array of research projects. Formerly an industrial-scale cranberry farm, Tidmarsh has recently undergone restoration [58] with the goals of re-establishing natural processes, such as the free flow of water, and ecosystem services, such as biodiversity support. At the time of this writing, the earth-moving portion of the restoration has been completed: water control structures have been removed and ditches filled; 3 miles of new sinuous stream channel were constructed and connected to a new pond, and riffles have been built to raise the water table. Formerly flat bog surfaces were sculpted with micro-topography, and many tons of fallen trees were distributed to create a wealth of new micro-habitats across the site. Nature has begun to take over, and the site is an officially designated publicly accessible wildlife sanctuary [84]. In addition, a non-profit organization, the Living Observatory (LO), has been formed to bring together the scientists, artists, and wetland restoration practitioners working on the land and engaging through it [9].

Over the past 5 years, as the restoration went from the design phase into active construction, we incorporated our ubiquitous sensing framework at each stage of the process, ultimately spanning sensor nodes, a generalized real-time sensor data API, and novel user experiences. Our continuously evolving sensor network was in place for 3 years documenting the pre-restoration environment, and is continuing to do so as nature takes its course. Its data, in conjunction with other, more targeted data collected by our research partners, are available to scientists, restoration engineers, and land managers. But beyond the environmental science and local environmental benefits of the project, the partnership behind the restoration is seeking to understand how future park visitors will interact with this new kind of landscape model. For this reason, our mandate has extended beyond sensing, both to constructing infrastructure and developing user interfaces that bring manifestations of the sensor data to the public. A significant portion of the underlying ecological change is invisible to the naked eye—some of it too slow for us to witness, some passing in the blink of an eye, and much of it occurring where we're not looking. Can we build technologies that allow us to witness, enjoy, and examine landscape-scale change in new ways?

It is with these broad questions in mind that we conceived of a Networked Sensory Landscape (NSL) on which multidisciplinary research would engage environmental science, technology, and the public. Figure 1-2 shows best my attempt to illustrate in one unifying representation both sides of the complex story I weave through this thesis—from sensor networks to sensory perception.

The system we deployed at Tidmarsh, like any end-to-end sensor network, has three layers (collection, back-end, and user interface) but in our case, each part is designed with the NSL concept at its core [86]. The sensors themselves and network infrastructure were built for both scientific inquiry and open-ended user experience, the back-end was built to support almost any kind of real-time, user-facing application, and the interfaces were built to encompass the myriad ways the public can engage with landscape. Details of the design and operation of these systems are given in Section 2.2 of this document.

1.4 Contributions

The contributions of this work are as follows:

- **Networked Sensory Landscape**: vision and case study development of a ubiquitous environmental sensing system at the Tidmarsh Wildlife Sanctuary, leading to many associated projects and theses, and supporting the studies in this dissertation. This collaborative work and associated projects were published recently in the journal *Presence* [86] and presented to the Wetlands research community in [25]. Data from the monitoring system continue to be collected, used in research, and publicly shared. The Networked Sensory Landscape vision was conceived with Brian Mayton, and the deployment was undertaken as a collaborator with Mayton and others.
    - Outdoor high-speed network infrastructure to support data collection and projects, embedded during environmental restoration
    - Environmental acoustic monitoring installation comprising more than 2 dozen synchronized live streaming and recorded audio channels distributed over a large area and running continuously for 1.5 years

  - Field deployment of an environmental sensor network, with more than 3 years of data collected across multiple sensing cells

- **Transpresence Environments**: a series of virtual environments and art installations exploring and developing transpresence (DoppelLab [30], Doppelmarsh [44, 86], Elevator Awareness, Moss Listening, ListenTree [29, 32, 98, 100])
  - In addition to their use as research and outreach platforms, the virtual environment projects include generalized software components used for transpresence wearables and other experimental HCI
  - Early installations brought the Tidmarsh soundscape into the built environment, allowing us to grasp the richness of the source material and experiment with different forms of spatial presentation and visitor interaction
  - ListenTree was installed in parks and arts venues worldwide, and experienced by tens of thousands of visitors. Lessons from its successes and failures led to an emphasis in the later work on the consideration of site-specific and pre-existing cultural and perceptual intuitions in interaction and content design

- **Transpresence Perception**: design and development of a networked sensory prosthetic platform, in the form of a modular, fully non-occluding spatial auditory display called HearThere [111], and user studies demonstrating its effects; studies in the lab investigating aspects of spatial hearing and auditory attention for application to wearable devices
  - Validation of the HearThere platform for long-term, real-world use
  - Demonstration of sensory integration, with new perceptual sensibilities arising within (typically) less than one hour of first use
  - User study outcomes with application to the physical and interaction design of future transpresence wearables
  - Findings pointing at the role of composed perceptual confusion and intuition in forming new perspectives and sensibilities on data and the environment
  - The construction of an experience of nature described by visitors as beautiful and transformational

- **Presence Future**: a well-defined path for future transpresence experimentation in the near term, as well as an attainable vision for wearables that would combine user attention with AI guidance for extended intelligence applications

1.5 Thesis Outline

Chapter 2 presents the background for the projects presented here: first, the related work across a number of fields; and second, my supporting role in the design and deployment of the Tidmarsh sensor network, which serves as the foundation for many of the key projects presented here. Chapter 3 introduces *transpresence environments*, place- and time-blurring sensate environments suffused with sensor data and built for open-ended exploration; the chapter goes on to present transpresence projects I've developed in both virtual and physical worlds. Chapter 4 presents the culminating projects of my thesis work, detailing the wearable audio-based approach I developed to extended perception, the field testing I undertook, and the exploratory lab studies I conducted along the way. Chapter 5 distills the results from earlier chapters into a set of guidelines and new research questions, introduces clear new lines of future inquiry brought about by the thesis, and provides an outlook for the future of networked sensor-driven extended presence and its growing relationship to extended intelligence.

Chapter 2

Background

This chapter is divided into two parts, in different ways comprising important background for the thesis. The first section reviews the supporting and related academic literature that I built on in developing the research and major projects of the dissertation. The second section, Deploying the Living Observatory, describes my collaboration with Brian Mayton and others to develop, deploy, and maintain the sensing and network infrastructure at the Tidmarsh Wildlife Sanctuary, upon which much of that research depends.

2.1 Related Work

To map the space of related work, I start with a broad, field-defining question: how can we make digital information viscerally perceptible and attentionally modulated? This question itself comprises several subquestions. Which data are suitable for perceptual mapping? How do different types or sources of data change the nature of the problem? What does it mean for data to be viscerally perceptible? What is attentional modulation, and how can it be modeled and applied? To what extent do different approaches generalize? Casting a wide net, what precedents can we build on to advance the field?

My work is largely focused on creating situated perceptions of data produced by distributed sensor networks, and my scope is therefore narrowed to physically-rooted sensor data—data in and of the world. Sensor data represent a fraction of all network-accessible digital information, a segment nonetheless widely recognized to be extremely important and fast-growing [46, 117]. Notwithstanding the significance of sensor networks across all aspects of connected life, it is important to note that I have chosen to exclude from my purview many other forms of

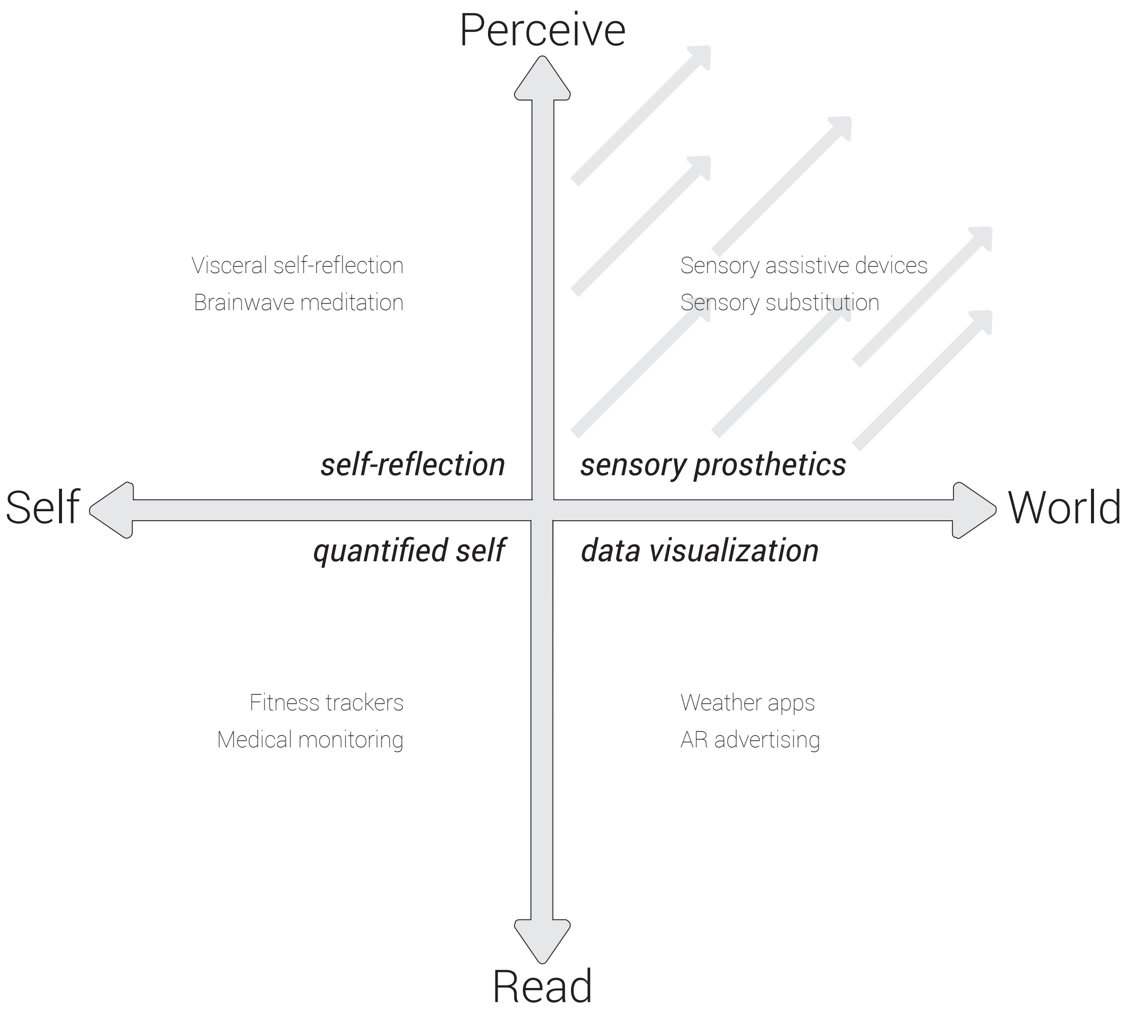


Figure 2-1: My work is situated in the space of technologically-mediated perceptual experience in and of the world.

data (emails, calendars, internet videos, social media, advertisements, etc). I am interested in the slow, visceral, corporeal engagement with the world that rapt listening brings, in contrast with digital interfaces that prioritize efficient consumption of information through layers of abstraction. Sensor data originate from the measurement of physical phenomena in space, a simple, physical rootedness I see as providing a more direct path to embodied interaction. However, my focus on sensors does not entirely rule out other forms of data, including person-to-person communication, from having embodied or embodying qualities. Brave, et al.'s InTouch, for example, offers the kind of physical directness of experience that I am after, in a desktop communication device [11]. When two users engage through InTouch, the physical distance between them is bridged not by disembodied abstraction but by the transmission of mutual touch: they *feel* one another.

On a practical level, rich sensor data vary over space and time, making them well suited to perceptual mapping as compared to spatially invariant data such as email. Numerous supporting examples can be found over decades of sensory substitution research, and in the successes of related sensory assistive technologies [3]. Finally, an underlying assumption in my work is that continuous spatiotemporal variance in data sources *and* sensorimotor engagement with these varying quantities is what distinguishes perceptual and non-perceptual awareness [95].

Figure 2.1 further clarifies the scope of my proposed work. In a 2d space of related work, where the first dimension spans *self* to *world* and the second *read* to *perceived*, my work occupies the first quadrant, representing technologies that construct perceptions of the surrounding physical world. The second quadrant captures work focused on perceptions of the self, such as Kleinberger's *Vocal Vibrations* [72] as well as biofeedback technologies, such as EEG-supported meditation [56]. The third quadrant covers technologies for quantifying and assessing the self, such as fitness trackers, and the fourth captures approaches to distilling and informatively presenting information about the world, such as climate data visualization, as in Hawkins' *Climate Spirals* [47], for example.

Given this scope, a diverse set of fields intersect to support my work: ubiquitous sensing, wearable computing and wearable assistive technology, virtual and augmented reality, ambient devices and environments, cognitive science and sensory neuroscience. In addition, my practice is strongly influenced by a long history of perception in philosophy and sensory experimentation in fine art. Selected references from these areas are briefly summarized in the sections that follow.

2.1.1 Sensor Networks & Ubiquitous Sensing

Wireless sensor networks (WSNs) emerged as a distinct research field in the late 1990s, building on advances in low-power microcomputing, the widespread proliferation of wireless networks, and the commoditization of sensor technologies [34]. While a large part of the field continues working to strengthen these pillars, a contemporary vision of ubiquitous computing [132] has begun to materialize with widespread WSN deployments, bringing with it the era of the Internet of Things (IoT) and ubiquitous sensing [103]. Proponents argue that as both sensors and distributed compute power get smaller, cheaper, and lower-power, they will weave seamlessly into the fabric of the physical world. We see this happening in new form factors for sensor networks [21] and body-integrated sensing for proxemic and personal interaction [136, 65]. Rapid prototyping of highly-integrated wearable technologies is now commonplace, and these developments were leveraged in building the systems of this thesis.

The convergence of WSNs and mobile technologies is also enabling a variety of applications planned, emergent, and expressive, from mobile air quality monitoring with distributed calibration [106] to Ryuichi Sakamoto's "Forest Symphony," a generative music installation based on worldwide forest sensing [114]. A number of early precedents exist for the Tidmarsh sensor network, from agricultural sensing and related UIs [13] to wildlife monitoring [119]. [110] reviews the space of sensor networks in ecological research, cataloging the various subdomains and their shared challenges. Most examples from the literature do not share our simultaneous focus on landscape-embedded infrastructure, long-term sensing, real-time data delivery, and a focus on user experience. Section 2.2 details our work in this space.

2.1.2 Sensor Network UI in Mixed Reality

General purpose UI tools for sensor networks have been developed by a number of researchers in HCI and embedded systems. In [83], Marquardt, et al. presented the "Visual Environment Explorer," a tool for assessing the states of multiple networked devices and exploring the network in various ways. In [87], Michel, et al. demonstrate an end-to-end environmental sensor network and visualization tool comprising environmental monitoring weather stations and various forms of graphical display, including 3D contour map overlays; their map-based, physically-linked visualization is an early precedent for the virtual environment projects presented later in the thesis.

A more direct precedent for that work can be found in [77], where Lifton, et al. proposed "cross-reality environments." They conceived of ubiquitous sensor and display nodes acting as portals between the physical world and pervasively shared virtual worlds, creating entirely new media from the sensor data and encouraging more open-ended exploration of sensor networks than traditional user interfaces would allow. Our work in this domain moved away from an initial emphasis on the social possibilities of shared cross-reality environments, and focused increasingly on evoking experiences of presence; by tying physical environments and sensors to immersive virtual counterparts, 3D visualization and spatial sonification are used to facilitate natural user interaction with sensor data [30, 63]. Platforms such as "DoppelLab" and its successors allowed us to explore sensor-driven 'responsive environments' tied to architecture or landscape but unencumbered by physical or temporal constraints. This characteristic open-endedness of the cross-reality world is shared by the wearable perceptual interfaces of this thesis.

Where cross-reality environments tunnel physical phenomena into virtual spaces through sensors, augmented reality (AR) can bring virtual representations of sensor data into the physical world [82]. All AR systems require position and head tracking to align the user's real and virtual point of view. There are too many examples of AR systems both commercially deployed and in the literature to name, with applications in geographical information systems (GIS), agriculture, and many other related areas. In one early example, King, et al. visually overlayed data such as harvest yield directly onto a vineyard in situ [69]. More recently, AR has been used to visualize predictions of future events in simulation on top of the affected objects [76, 61]. Generalized platforms for audiovisual AR are seeing wide commercial release, with notable entries from both major players like Microsoft and dedicated startups, like Magic Leap. These in turn are leading to commercial applications that will use real-time sensor data for task support in manufacturing and other domains. More broadly applicable sensory assistive systems using these platforms have yet to materialize but are sure to come, and this dissertation is positioned to offer guidance for the design of new perceptual, attentional, and cognitive dimensions they will bring to users.

2.1.3 Ambient, Wearable & Assistive Technologies

Stemming from a contemporaneous articulation of Weiser's vision of ubicomp [130], environmental and wearable interfaces promise to weave digital information into everyday *sensory* experience. Since Weiser's call in the late 1990s, artists and researchers have sought 'calm,' ambient technologies embedded into natural and built environments. A pioneering realization of this vision, Ishii, et al.'s *ambientROOM* was a platform for experimenting with pre-attentive, or peripheral interfaces through augmented architectural space [59]. Since then, technological advancements in low-power computing, networking, solid-state lighting, and ubiquitous sensing have brought that vision much closer to reality. Many recent examples, from proxemics and interactive digital signage [41, 10] to the ListenTree project presented later in the thesis [32] demonstrate resurgent interest.

Not limited to ambient environments, calm technologies can be located on the body as well. With its introduction of pervasive spatial auditory display, Sawhney and Schmandt's "Nomadic Radio," research well ahead of its time, paved the way for my work [116]. More recently, we have seen a revolution in always-on mobile and wearable devices, capable of collecting user input [22, 65], sensing their environments, and displaying dynamic information from the web [38]. These devices are getting smaller and coming closer to the body, weaving into clothing [101] and conforming over skin [125].

Meanwhile, interest in 'assistive' sensory prosthetics, such as hearing aids and vision replacements, has broadened considerably. Mobile assistive technologies precede even desktop HCI, as researchers going back to the 1960s have experimented with wearable sensors and actuators. Sensory substitution devices (SSDs) rely on neuroplasticity to map input from an external sensor to an existing sensory modality on the body [15], and recent efforts extend to implantation and even cross-sensory cortical stimulation [122]. Results from the literature suggest that a user's neuroplasticity can, under certain circumstances, transcend sensory substitution to enable pre-attentive cognition of 'extrasensory' stimuli [91, 74], though both the mechanism and the limitations of this approach continue to be debated [23].

From this perspective, some wearable technologies appear headed for convergence with SSDs and eventually implantables to become 'more' prosthetic, and efforts to engineer this convergence have indeed been intensifying. Evidence of this can be found in new academic communities with roots in human-computer interaction (HCI) advancing the 'augmented human' concept, though these ideas have been percolating in the wearable tech discourse for quite some time [108]. To date, assistive wearable and SSDs have largely been limited to sensors on the body and to relatively narrow conceptions of medicalized remedies (e.g. vision replacement with a front-facing wearable camera). Furthermore, advances in mobile sensing and networking have not been integrated into sensory prosthetics in the way that they have been in mainline wearable technologies. These capabilities include low-power wireless communications, fine-grained location and gaze tracking, and low-cost bio-sensing (EEG, ECG, GSR, heart rate). We are beginning to see crossovers, among them Empatica's electrodermal activity sensor [36], which received FDA approval for its seizure monitoring in early 2018. Similarly, technologies from sensory prosthetics have been slow to trickle down into wearables (bone conduction audio, electrotactile and vibrotactile display, implantation), though progress in these areas has been driven by recent demand for new input modalities for HCI. A number of other sensors (e.g. mobile eye-tracking and pupillary measures) remain challenging or expensive to measure in the field, but will drive new advances in both domains as commercial applications take hold and costs come down.

Wearables: Intelligent Agents vs Prosthetic Extensions

There is an evident split in the ongoing evolution of mainstream wearable and ambient technologies, particularly where AIs are concerned. On one hand, exemplified by the UI design guidelines of Google's Glass [40], devices become smaller, closer to the body, and more intelligent; their intelligence makes them better assistants (and by some accounts, better friends) to their users—more aware of users' goals and needs. Their third-party assistance translates into effective action in a goal-oriented timeline; AIs sense context to help users do what they're doing better and more efficiently, and they entertain on demand. Interaction occurs for the most part through direct engagement with the device/agent, which responds to touch and speech commands with natural language.

On the other hand, represented by the work in this dissertation, wearables inspired by sensory assistive technologies act non-linearly on their hosts, and not explicitly for task-support. They also rely on devices becoming smaller, closer to the body, and more intelligent, but modeled on a sense organ rather than an AI assistant, the sensory prosthesis does what it can to blend into its wearer's perceptual experience. Non-occluding presentation is a given. As such, it does not primarily *help* her achieve her immediate goals, and it may even distract from them with unrelated perceptual information. It may interrupt its wearer, in a bottom-up fashion, if some data in its stream meets a threshold of just noticeable difference; or its stimulus may be the focus of its wearer's attention as they explore its variations in the world. Stimuli are be mapped onto the space around the user, and the most powerful interface to the sensory wearable might be in its dynamic response to the user's smallest physical movement (self-movement being a key to the development of new sensorimotor contingencies [48, 91]). Following the conventions of a sense organ, the sensory wearable virtually spatializes stimuli to produce a sense of externalization, or distal attribution [2]. Ideally, the wearer's selective attention seamlessly interacts with the spatially coded data, dynamically bringing stimuli to the foreground or quieting them.

2.1.4 Sensation, Perception, Attention

The increasing convergence of consumer wearables and sensory assistive technologies has brought questions of perception to the forefront of HCI research. Examples of this include brain-computer interfaces (BCI) and related physiologically signaled interactive systems [120]. My own investigations have been strongly influenced by results from cognitive science assessing claims about perception through assistive devices such as SSDs. The literature offers guidelines for better understanding the conditions under which a generalized networked sensory prosthesis might work.

In recent years, a debate has centered on the locus of perception for users of SSDs. For example, given an SSD that maps image pixel intensities to intensities of vibration on the skin, under what conditions should the interface be considered visual or tactile? In [55], Hurley and Noë coin these states cortical *deference* and cortical *dominance*, respectively. In the preceding example, if the user's experience switches from tactile to visual (thereby constituting a distinct distal modality), the intrinsic modality (the channel) is considered deferent to the extrinsic one. From the literature, a number of practical design criteria may catalyze experiences of deference from a wearable device. First, the interface should not change rapidly; in a challenge to the traditional technology development cycle, changes and upgrades must be rare, and they should be gradual. Discomfort with the wearable or the sensation of its transducer appears to disrupt the SSD learning process [91]. A user should be able to interrupt the stimulus through self-motion (e.g. occlusion) in a consistent, repeatable way [2]. Finally, the user's interaction with the stimulus should be direct and low-latency [93]. While a new primary sensory modality might be difficult to prove, a sensory-focused wearable that meets these criteria ought to produce a secondary modality, something akin to reading for the visual sense, where the lines on a page are indistinguishable from the letters they represent. Techniques like spatialization and location, orientation, gaze, focus, and/or attention tracking would externalize and differentiate the stimuli.

Attention-sensing SSDs have not as yet appeared in the literature, though attentive user interfaces have shown promise in HCI research over the last decade, providing a basis for further investigation [126]. Attention is hard to observe, and even harder to measure, a problem well-described in Fritz, et al. [35]. It is "flickering and elusive," and has "highly variable selectivity, intensity, and duration." Particularly in the auditory study settings, it can be hard to control subjects' fleeting attention to different stimuli. In natural settings, this becomes even more of a challenge, as I will discuss in Chapter 4.

2.1.5 Philosophy & Art

In his 1637 *Dioptrique*, or *Treatise on Optics* [18], Descartes lays the groundwork for both a contemporary philosophy of the sensorium and a very modern understanding of sensory differentiation. A blind person using a walking stick might become so adept at its use, he argues, that "their stick is the organ of some sixth sense." In Descartes' conception, perception equates distinct physical phenomena with each other; that one might "see with their hands" is grounds to infer a relation between light and mechanical force. But how does a one's experience of an extrinsic device—a stick—develop from what Descartes describes as a "sensation [that] is somewhat confused and obscure" to one that could take the "place of sight"? Questions like these form the foundations of a long tradition of philosophy (more recently under the umbrella of sensory studies) and artistic experimentation in which some of my work takes part.

Particularly relevant is recent scholarship by Jonathan Crary, whose passage from [19] anchors the preface of this dissertation. In "Suspensions of Perception," Crary traces the recent history of industrialized human attention, both 'payable' and liable to incur 'deficit'. It is precisely this pervasive accounting that lays the groundwork both for my critique of contemporary computing and, ironically, for the attention sensing and inference I explore in Chapter 4 as a response. Alva Noë's pioneering work on perception and presence is equally critical background for my work. His sensorimotor contingency theory of perception, developed with Kevin O'Regan, provides a solid basis for pursuing a line of inquiry into the development of new perceptual modalities as well as new sensibilities on top of existing modalities [93]. They argue that we build and differentiate modalities by effecting repeatable dynamics in sensory stimuli through different kinds of self-motion. Noë's later work offers philosophical and phenomenological grounds for the same inquiry [94].

Beginning in the mid-late twentieth century, artists like Alfons Schilling (Dunkelkammerhut, Video-Head-Set), Kristof Wodiczko (Instruments), Rebecca Horn (Berlin Exercises, Feather Fingers), and later Kelly Dobson (Machine Therapy, Wearable Body Organs), Wendy Jacob (Autism Studio, architectural surfaces), Janet Cardiff (Her Long Black Hair, Forest for a thousand years) and others radically re-envisioned perception-action through their works and practices that incorporate the performativity, materiality, and corporeality of the act. Their works opened my eyes to the artistic resonances of my own research and creative experimental practice. As I gained clarity on the technological challenges I would face in the work, it also became clear that artists had been grappling with and creatively solving many of the same problems in their own ways, often by going back to basics. This insight was critical to the design of HearThere, a project ultimately about creating a completely new experience that is exactly like hearing has always been.

2.2 Living Observatory: Constructing the Sensor(y) Landscape

With the sensorimotor theory of perception as a basis, technologies for sensory augmentation have generally presumed the sensor data—the 'source material' of a digital perception—to be structured but otherwise generic. This premise has been fruitful in the beginning stages of sensory assistive technology development, where necessity has driven adoption. For the public at large, however, it becomes less clear what kinds of sense media would drive adoption of a *generic* sensory prosthetic, and to what ends, given the immense challenges of learning new contingencies as an adult. This has led me to two constraints in building my work: first, that I would leverage existing modalities (mostly audio-vibratory perception), and second, that I would draw my source material from the natural world. This section focuses on the development of the latter constraint.

I chose to work with environmental data for a number of reasons. Most importantly to me, the source itself is rich and compelling, a precious thing to behold. It not only feels good to be in nature, but it is also good *for* us [66]. Having a view to nature reduces stress, restores attention, and famously even improves test scores [121]. Beyond those direct benefits, we are dramatically changing the environment in ways and on scales we can barely comprehend, not least access perceptually; however, with the high spatiotemporal resolution of sensor networks we are able to capture environmental change over long periods of time. Finally, digital technologies as they exist today generally deprive us of the pleasures of being in nature, but I assert that designed to support perceptual attention, they would do the opposite. In a very different sense than in the previous section, the background for the work of this dissertation is in an installation of sensors and supporting infrastructure 50 miles down the coast from MIT. What follows is a summary of that deployment, undertaken over the four years prior to the culminating work of this dissertation. More detail about the Tidmarsh sensor network can be found in [86]. The design of the system is part of Brian Mayton's forthcoming dissertation (2019), and will be fully elaborated there.

2.2.1 Infrastructure

Figure 2-2 shows a map of most of the 600 acre Tidmarsh property, including the 225 acre active restoration area. Markings on the map show the network infrastructure we deployed and locations that we have instrumented with sensing. The sensing sites are connected by a high-bandwidth internet protocol (IP) network that allows real-time data streaming from the sensors, microphones and cameras, and provides standard WiFi internet connectivity to visitors. The IP backhaul consists of long-range TDMA Wi-Fi wireless links, single-mode and multi-mode fiber, and copper CAT6e for shorter runs. Most of the infrastructure was installed prior to and during the Tidmarsh restoration, allowing nature to return around it. A barn houses the head-end infrastructure, including the main router and on-site server; the wireless backhaul base station is mounted on its roof. Anywhere on the site with line of sight to the wireless backhaul base station can be IP connected using relatively inexpensive and widely available hardware.

Site 1 in Figure 2-2 contains the largest and longest-running sensor installation at Tidmarsh, hosting 60-70 sensor nodes, a network camera, and a stereo audio stream. This site is IP connected via the wireless backhaul and is powered by two 100 watt solar panels paired with a 2,400 watt-hour battery. The large power capacity is needed to withstand long periods of low sunlight during winter storms. Site 2 in Figure 2-2, known as the former impoundment, was once an artificial pond used in the cranberry farming operation that preceded the restoration. Restoration actions there began with opening dam spillways to let impounded water flow out. Several years later, the spillways and most of a berm were removed, connecting the river system in the former pond with the channel to the north. No excavation took place upstream of the dam site, and a naturalized river channel formed when the reservoir was drained.

Though there is no line of sight to the head end from the impoundment, buried optical fiber provides the IP connectivity there, and AC power is wired through as well. (Recent changes in the site infrastructure taking place after Figure 2-2 was drawn have resulted in a completely wired through IP connection to the impoundment site for the most reliable and

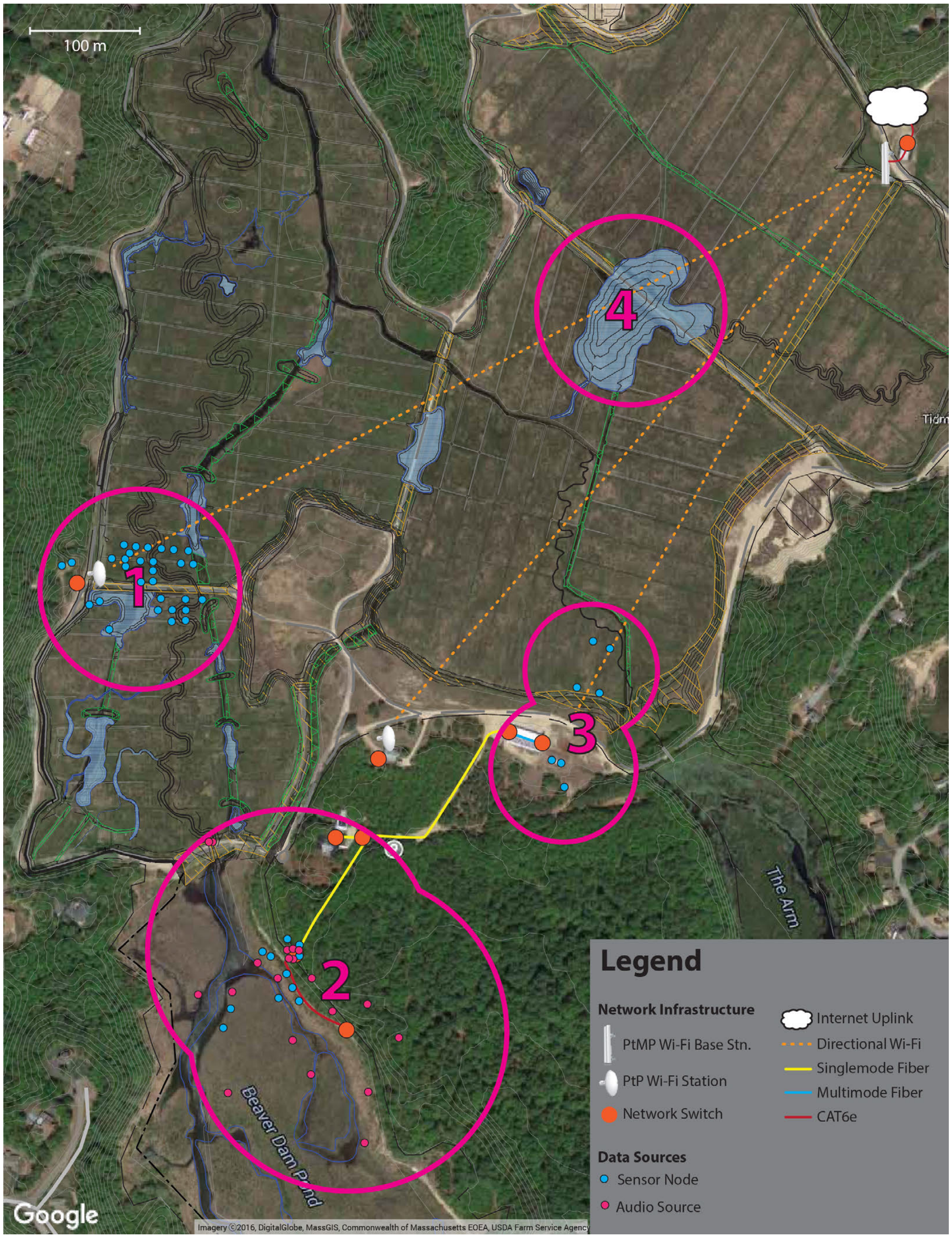


Figure 2-2: Tidmarsh network infrastructure and sensor sites. Figure by Brian Mayton, excerpted from [86].

highest bandwidth connectivity we can achieve.) Most of the work and experiments described in later chapters was centered on this site, which has sustained the highest density of acoustic monitoring largely due to its continuous supply of power. It also hosts two network cameras and another large set of sensor nodes.

Site 3 was not used in my studies, but hosts a small number of sensor nodes in nearby greenhouses and on a research test plot used by institutional collaborators (Mount Holyoke College), as well as a network camera. Finally, site 4 hosts a network camera and an audio stream at the edge of a highly ecologically active pond and stream system introduced in the restoration. In the spring of 2017 this site also hosted antennas that were installed by Living Observatory collaborators in the MIT Sea Grant program to monitor the movement of river herring through the site. Its systems are solar powered. Video and audio streams from site 4 were used in projects described later. Sites 3 and 4 both connect via the wireless backhaul.

2.2.2 Environmental Sensing

The Tidmarsh sensor node is the basic building block of the system. Designed by Brian Mayton in two major releases, the node has a number of advantages over commercial systems. It is tailored for the Tidmarsh setting, and was flexible to deploy, customize, and alter as the construction work progressed. While most environmental sensing deployments target a specific type of measurement over a fixed period of time, chosen to answer a specific question, the Tidmarsh node was designed to capture everything we could feasibly imagine measuring in a remote sensing application. In this way, it can support a wide spectrum of different applications, from traditional ecological research to experiments in presence and perception, at a cost (approximately \$80 per assembled node in the first generation and \$150 in the second, not including engineering) that allowed us to scale from prototype deployment to site-wide installation. The second generation node was designed to address shortcomings of the first, and made improvements to streamline expansion, extend operating life, and add additional sensing modalities.

Each node consists of a microcontroller, radio, power source, and sensors in a weatherproof enclosure. The platform is designed for deployment in large numbers, with internal sensors providing a common sensing baseline across the site. External probes such as soil moisture or water quality sensors can be connected to the node's expansion port, and this flexibility allows more targeted research studies to be quickly set up where needed. Baseline onboard sensors in the second-generation node include temperature and humidity, high dynamic range visible light as well as UVA, UVB and IR, atmospheric pressure, acceleration/vibration, passive IR (motion), and a microphone with integration, peak detection, and FFT capabilities. A small solar panel provides enough power for continuous operation without need for battery

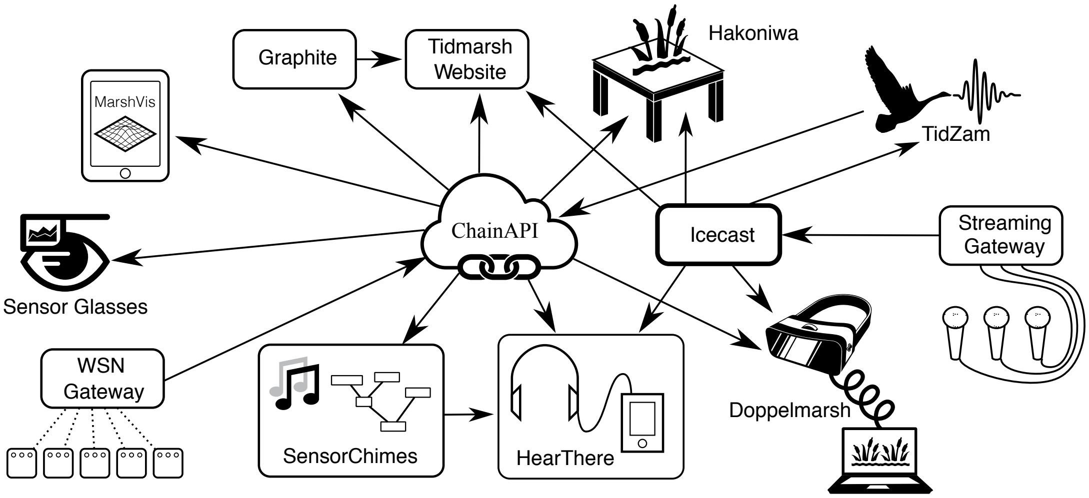


Figure 2-3: Data from sensors and other processes flows into ChainAPI, where it is distributed via HTTP and WebSockets to end-user applications. The Icecast server software distributes Ogg Opus, Ogg Vorbis, and MP3 audio streams. Figure excerpted from [86].

replacement for the life of the device. The node is wireless and supports basic multi-hop routing to extend the network.

50-100 first-generation Tidmarsh sensor nodes were continuously deployed during the studies described in this thesis, along with a very small number of prototype second-generation nodes. At the time of this writing, hundreds of second-generation nodes are being readied for deployment. A detailed account of the node design, evolution, and deployments will be included in Mayton's upcoming dissertation, expected in 2019.

2.2.3 Data

To handle the volume of data produced by the Tidmarsh sensors and flexibly support a diverse array of applications and use cases, we use a database front-end system called ChainAPI, described in [112]. Following the REST architecture, ChainAPI creates hyperlinks between data resources, allowing clients to discover both operations and data associated with a given resource. For example, a sensor node that has a real-time stream available will offer a link to the stream. This hyperlinked architecture enables applications to crawl through the resources at runtime, much like an online search engine. As sensors come online, they become discoverable through the establishment of new links.

Figure 2-3 shows a wide variety of clients that pull data from ChainAPI, many of which

Figure 2-4: Parts of the audio streaming installation. Top left: microphones awaiting installation. Bottom left: satellite input box. Center: microphones in trees and in the marsh. Right: main electrical box, showing mixer (top), computer (center), network switch and fiber (bottom left). Figure excerpted from [86].

are introduced in Chapters 3 and 4, as well as two sources of data (the Tidmarsh sensor network and audio classification output from Tidzam). The clients shown here are on a diverse set of platforms, including the Unity3D game engine on desktop and mobile, node.js web applications, Python scripts pulling the data into visualization tools, and browser-based JavaScript web applications. To support these applications we have developed client libraries in JavaScript, Python, and C#, each of which is used by multiple applications. In addition to requesting current or historical data over HTTP, clients can use ChainAPI to be notified of new or modified data using WebSockets. These notifications use the same JSON message format as the HTTP payloads, and are used by most of the real-time systems presented later to provide low-latency services (e.g. visualization, sonification, classification) to on-site (mobile) and remote users.

2.2.4 Audio Streaming & Acoustic Ecology

From the beginning of our work at Tidmarsh, preceding the sensor deployment, we sought to capture the sound there. When we walked the site, we heard the soundscape changing dramatically from moment to moment and place to place, and in surprising ways over time [42]. Microphones capture a rich trove of information about a site, including wildlife (measures of behavior, presence, and, in the aggregate, biodiversity), human-wildlife interactions, weather, as well as structural characteristics of the physical environment. Compared to other high-bandwidth media like video, sound is relatively inexpensive to capture and transmit, and is less impacted by occlusion (a challenging problem for cameras in dense marsh or forest). For evoking a sense of presence across a large area, sound in many ways surpasses video; where cameras are limited to single points-of-view from which a convincing 3D visual space is difficult to reconstruct, multiple audio streams can be spatially rendered for a sonically immersive user experience that feels like listening there.

Audio Infrastructure

Uniquely, because of our applications to presence, computing, and ecological science, our acoustic monitoring system has real-time/network as well as long-term storage requirements. With varying constraints and challenges posed by different areas on the landscape, we have taken a number of different approaches over the years. Where power was not available, we developed various WiFi-based systems centered around ARM-based platforms (Beaglebone, Raspberry Pi, etc) and either consumer audio interfaces or custom solutions, usually for two channels at a time. The wireless audio systems use solar panels (60 watt or more) and lead-acid batteries to sustain year-round 24/7 streaming. Where power is available, we have been able to bring in audio interfaces with large numbers of channels and build dense, wired installations.

Most commercial microphones are unsuited for outdoor use, much less long-term wetland application. As a result, we developed our own weatherproof omnidirectional microphone based around a low-noise electret capsule (Primo EM-172) paired with a circuit for buffering the signal and driving a differential cable.

The electronics are housed in a silicon-filled aluminum tube, with the electret capsule alone partially protruding from the rubber. The protruding element is fitted with a foam windscreen. For hydrophones, we experimented with many different designs, including oil-filled canisters, and ultimately found that a simple hack to the microphone design—filling the tube with enough silicone to just cover the capsule—provided both reasonable fidelity and effective water resistance. However, the proximity of Tidmarsh to high-voltage transmission lines has left us prone to significant amounts of 60 Hz hum in the water, which we reduced by filtering. The microphones and hydrophones connect to standard multichannel audio interfaces with balanced XLR connectors and 48 V phantom power. The microphones have been continuously deployed for more than two years without major issues. Voltage surges on buried AES50 lines have caused occasional, repairable breakdowns in the encoding and mixing hardware, even with protection in place. The weakest link in the system is the longer runs of microphone cabling, which, even when buried, are an attractive target for chewing by small rodents. Cable replacement and maintenance is frequent and can be taxing, though we have recently found that some kinds of cable insulation are less attractive to wildlife. The most promising solution has been to repurpose gel-flooded cabling commonly used for direct-burial networking; the insulation appears to be less delicious, and even when breached, the gel prevents water from wicking along the cable interior. Wireless audio is also a solution, but given the power requirements of continuous streaming, it has been more difficult to scale. Figure 2-4 shows various components of the audio installation, including microphones, audio interfaces, and network equipment.

Audio streams are published using Icecast, a ubiquitous streaming audio protocol. Large groups of microphones connected to one interface are published in a single multichannel stream, which maintains synchronization between the channels. The interface is also used to mix the audio down to a stereo stream for use by end-user applications that do not have multichannel decoding capabilities (e.g. web browsers).

Installations

The first medium-scale acoustic monitoring stations we developed at Tidmarsh were two 8-channel live-streaming systems using wired microphones. One was installed in an area of retired farmland that was not yet restored, and used a nearby barn to house the audio interface and network head end equipment. The other was installed at a transitional site known as the "Arm" (an area just above the legend in Figure 2-2), using a friendly neighbor's garage and home internet for shelter and network connection. The soundscapes captured by the two prototype systems were very different. At the time, the Arm was a dam-supported pond in the process of slowly drawing down. We installed microphones in trees at its edge as well as on a marshy island in the middle of the pond, where we were able to closely monitor a goose family and other wildlife. Used for various spatial sound art installations, the Arm site

Figure 2-5: Map showing the locations of the streaming microphones, hydrophones, and cameras in the Impoundment site at Tidmarsh.

was a powerful proof-of-concept for the rich experiential capacities of high-density acoustic monitoring at Tidmarsh (installations described in 3.2.2). In contrast, the soundscape of the retired farm site was characterized by road and wind noise, with little variation. It was only a short time after farming had ceased, and wildlife had not returned to the same extent. Had we not found the Arm site, we might not have continued the recording project; the Arm showed us the enormous potential for continuous monitoring in restoration settings. In retrospect, the recordings we made on the retired farm are a useful record of what *wasn't* there then. Post-restoration, the area is teeming with wildlife and new growth, and its soundscape reflects those dramatic changes.

We later built and deployed various smaller wireless monitoring stations at Site 1, but without synchronization and less densely distributed, they did not produce content as compelling as the Arm site had. This led us to aim for an even more ambitious and dense installation at the Impoundment site (Site 2). At its peak, our monitoring station at the

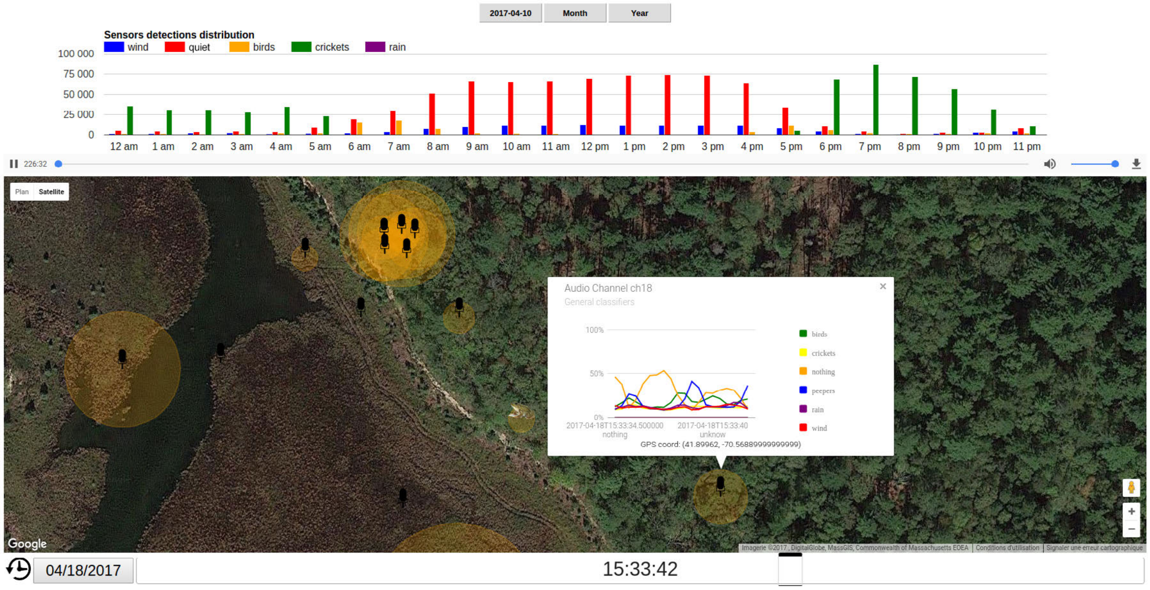


Figure 2-6: The Tidzam web interface visualizes density of clearly identified outdoor sonic events, such as wildlife and weather conditions, in space and over time. Figure by Clement Duhart, excerpted from [86].

Impoundment consisted of 22 wired microphones and 2 wired hydrophones, using more than 1 mile of cabling, distributed over a 900k square foot area of woods and marsh. Given the enormous efforts involved, the system was installed in stages over the course of approximately one year. At the time of this writing, 12 channels are operational, with the reduced numbers almost entirely attributable to cabling failures (caused by rodents and invasive species dredging operations at the Impoundment). We are planning to replace the cabling in the non-operational channels with more the more robust gel-flooded cables described in the previous section. Impoundment microphone, hyrdrophone, and camera locations are shown in Figure 2.2.4. Centered on a rack-mount, 32-channel combination mixer, DSP, and audio interface (Behringer X32R/S16), the capacity at the impoundment is likely the highest of real time environmental acoustic monitoring system. With the exception of the microphones on the network cameras, all channels are synchronized and encoded together.

Tidzam AI

Given the high density and long term nature of the acoustic monitoring project, the volumes of recorded data are extremely high, amounting at the time of this writing to approximately thirty years of recordings stacked end-to-end. As a result, automated analysis and classification of the audio are the only feasible ways to produce scientifically useful data and to intelligently index into the recordings for our end-user applications. To perform this task, we developed a system called *Tidzam* that analyzes large numbers of live streams, recognizes ambient acoustic scenes, and identifies the sources of transient sonic events from an array of wildlife (including dozens of bird species, frogs, and insects), vehicles, and visitors. Tidzam can also process temporary streams offered by users from their mobile devices. Resultant classifications are made available in real time for online use. This effort is headed by Dr. Clement Duhart, and is detailed in a forthcoming paper focused on it ([33], under review) as well as in [86]. I include a brief description of the system here because output from Tidzam is visualized in the VR and AR applications in Chapter 3.1.1, and because Tidzam was used by some of the wearable devices introduced later to tilt the audio mix towards wildlife sounds and suppress wind noise (Chapter 4.3.3). The output streams are accessible through ChainAPI (Section 2.2.3).

Tidzam uses convolutional neural networks to extract hierarchical feature representations describing the spectral content associated with different types of sound at Tidmarsh (e.g. rain, wind, frogs, birds, human voices, aircraft, etc.). Initial classifications are made to determine the type of sound, and for some classes (e.g. birds), the signal is forwarded to more specialized classifiers to identify the species. Particularly in the restoration context, certain types of sound (wildlife, human, etc) can appear and disappear both seasonally and over the long environmental recovery period. As a result, Tidzam is scalable to new classification tasks to accommodate newly resident and migrating wildlife; the system is able to detect when a new kind of sound appears and, with a human in the loop, build and label new classifiers as needed.

Chapter 3

Transpresence Environments

Since Marvin Minsky's 1980 call for the development of technologies that bring together remote sensing and action in [89], telepresence has for the most part retained defining characteristics of the preceding teleoperation technologies he cites: in particular, a focus on portal-based, sensorimotor *transport* from one place to another—presence for action. However, more recent developments in cross reality [30] and resynthesized reality [44] environments point to a calmer, more diffuse kind of mediated presence. Much like their physical world counterparts, these sensor-driven virtual worlds call for open-ended exploration, and getting there to effect an action is no longer the singular or even primary goal. Ambient interfaces, such as Ishii's ambientROOM [59] similarly aim for the background, open to and often inviting of attention but deliberately composed against demanding it.

At the intersection, then, of sensing and these data-suffused ambient environments is not telepresence but rather *transpresence*. In transpresence environments, multiple times and places might coexist, or even blur together. Sensemaking becomes a matter of possessing or developing perceptual sensibilities, intuitions for how different facets of the permeated data can be teased out again and contemplated. Transpresence environments are not portals, or windows *into* other places, but *are* other places, destinations for inhabiting, exploring, and probably staying a while. In them, generative spatial music pieces pipe in the sounds of the frogs and follow the phases of the moon; imaginary creatures feed on measurements, glass walls transparently pass sound, and living trees give oral histories. These situations certainly break visitors' expectations, but once the initial surprise is passed, embedded metaphors allow them to stretch or expand perceptual sensibilities.

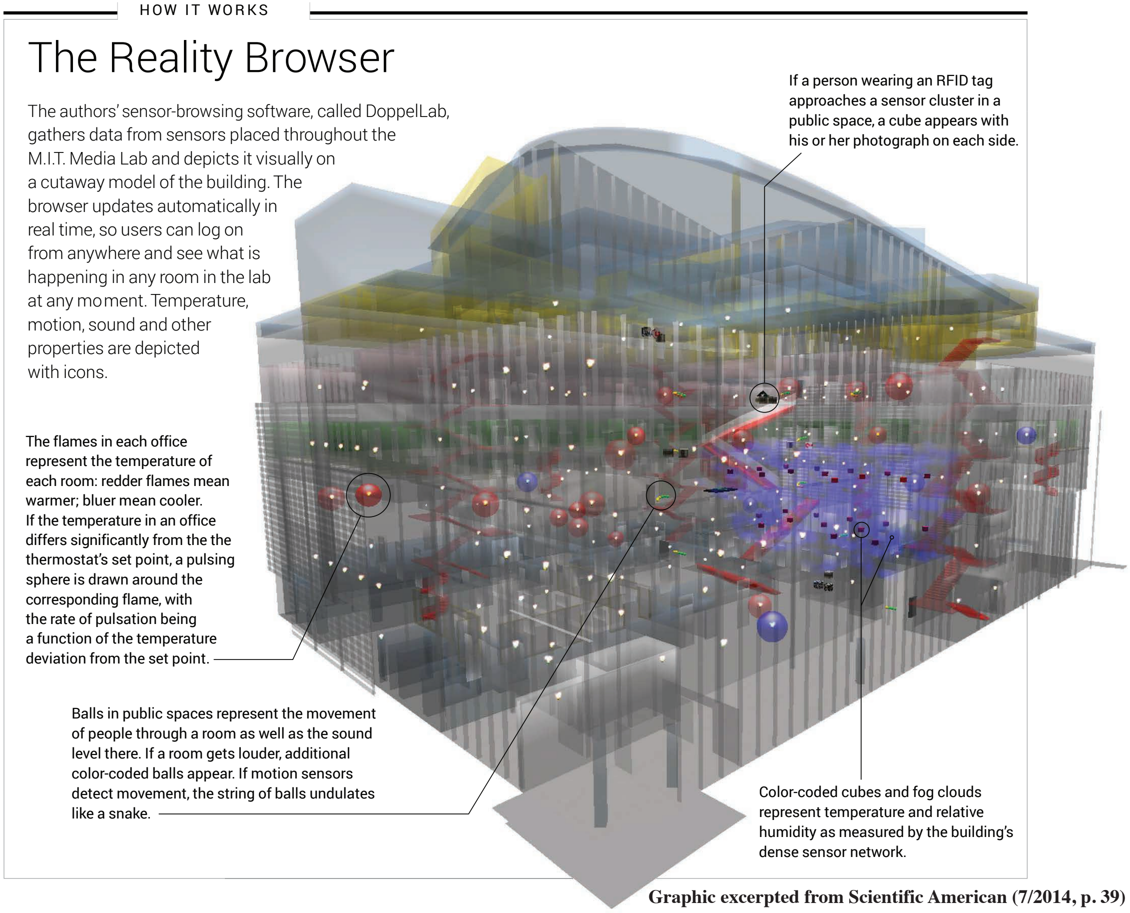
This chapter steps through transpresence environments I built in virtual and physical (built and natural) worlds. These projects are linked to each other and to the larger message of the dissertation through their progressively deeper embeddings of information into perceptible elements of the worlds they construct.

Major software components of the virtual world systems introduced in this chapter carry through to the wearable work in the next. Detailed descriptions of that software architecture are given in Section 3.1.2.

3.1 Transpresence in Virtual Worlds

This section presents projects exploring the sensory potentials of sensor networks in virtual worlds. My initial work in this area was developed as part of my master's thesis, which proposed various approaches to constructing sensory "contexts" in user interfaces to sensor data [24]. In one of those contexts, following a paradigm known as cross-reality [96], we sought to evoke feelings of presence in a virtual world modeled on a real physical environment. The virtual world is brought to life by real-time data from sensor networks distributed throughout the physical site. In this work, we used the virtual world to explore the design of immersive, animate audiovisual presentation for real-time presence and sensor data playback. What kinds of visual and auditory embedded representations could encode, in environmentally perceptible ways, absolute and relative magnitudes, rates of change, and higher-level inferences that relate data across time, space, and sensor modalities? From an engineering standpoint, how would sensor data be stored and made accessible to real-time and historical presence applications? Most importantly, what would it feel like?

The forerunner of the current work was DoppelLab, a PC-based cross-reality application based on real-time and recorded data produced by a building and its occupants [30]. In DoppelLab, representations of sensors are bound to architectural space, while users' movements are unconstrained by physical rules, and time is a dynamic parameter (to be traversed, stretched, and compressed). Virtual sensors can be represented in composite visualizations, and detail can be parameterized (representations are zoomable). DoppelLab's spatial live audio and sonification framework [63] constructed realistic sonic ambiences from a combination of live microphones and musical data sonification. In contrast with the current work, however, DoppelLab's dual conception as a remote presence platform *and* a smart building



HOW IT WORKS

The Reality Browser

The authors' sensor-browsing software, called DoppelLab, gathers data from sensors placed throughout the M.I.T. Media Lab and depicts it visually on a cutaway model of the building. The browser updates automatically in real time, so users can log on from anywhere and see what is happening in any room in the lab at any moment. Temperature, motion, sound and other properties are depicted with icons.

If a person wearing an RFID tag approaches a sensor cluster in a public space, a cube appears with his or her photograph on each side.

The flames in each office represent the temperature of each room: redder flames mean warmer; bluer mean cooler. If the temperature in an office differs significantly from the the thermostat's set point, a pulsing sphere is drawn around the corresponding flame, with the rate of pulsation being a function of the temperature deviation from the set point.

Balls in public spaces represent the movement of people through a room as well as the sound level there. If a room gets louder, additional color-coded balls appear. If motion sensors detect movement, the string of balls undulates like a snake.

Color-coded cubes and fog clouds represent temperature and relative humidity as measured by the building's dense sensor network.

Graphic excerpted from Scientific American (7/2014, p. 39)

Figure 3-1: DoppelLab renders 3D architecture together with the sensors that permeate it, opening up new dimensions for both building management and telepresence.

monitoring tool led us to prioritize efficient, high-contrast symbolic representations over more subtle embeddings of the data. This made the environment into more of an informative window than a destination in its own right.

Still, DoppelLab exemplifies the major leap in the evolution of our cross-reality work that came about in moving from the pervasively shared virtual world of Second Life into the Unity game engine, which immediately opened up new avenues for exploring cross-reality multimodally and through diverse forms of display. First, the world-building aspects of game development increased our focus on presence applications. In particular, our presence-focused approach to immersive audio in DoppelLab became the basis for the audio-related work that would follow. Finally, as the project grew to more platforms, we began applying of many of the software components underpinning the virtual environment to related projects in other

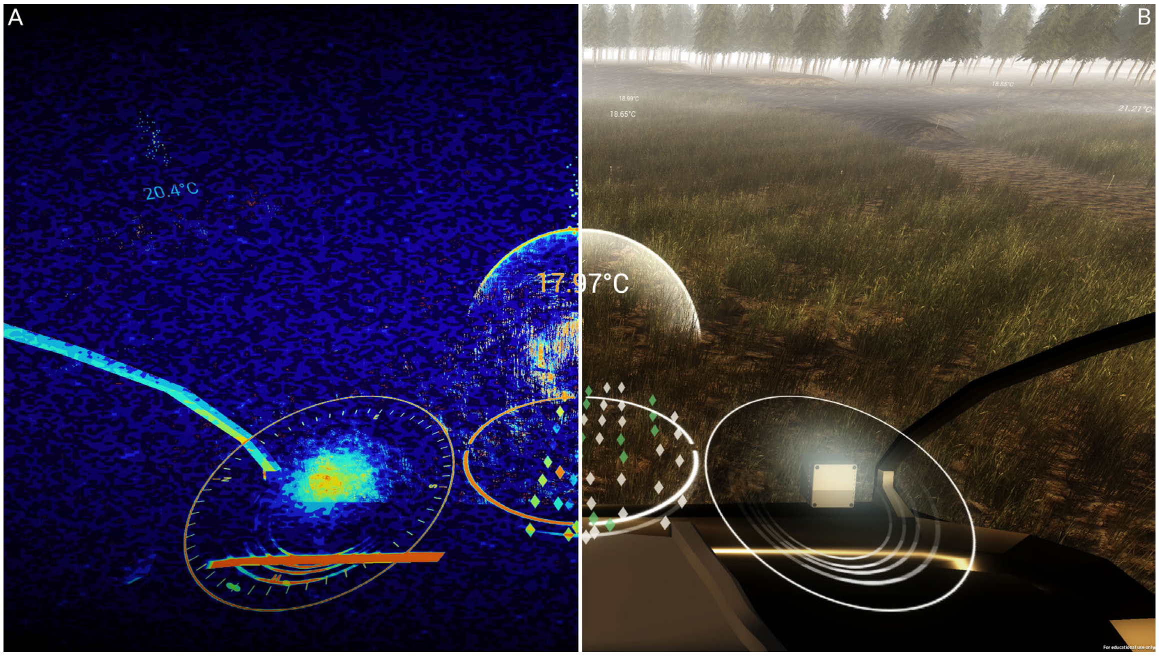


Figure 3-2: Different virtual 'lenses' highlight various aspects of the sensory world in Doppelmarsh: at left (A) with heatmaps on the terrain, and at right (B) with sensor-driven simulated microclimates. A 'mini-map' overlay shows the statuses and relative locations of sensor nodes around the user.

spheres, including in the augmented reality and in-situ sensory extension work introduced in Chapter 4.

3.1.1 Doppelmarsh

In the cross-reality work following DoppelLab, we sought to extend our approach to virtualizing the built environment to the outdoors. *Doppelmarsh*[1] is a virtual world based on the state of the physical environment at the Tidmarsh Wildlife Sanctuary. The virtual world re-synthesizes the physical environment in virtual reality [44], adding in interpretive elements that reflect measured processes on the physical site, and allowing users to explore the environment free of real-world constraints on physical space and time.

The Tidmarsh landscape is the basis for the virtual world onto which data are rendered in Doppelmarsh. The terrain models originated as topographic maps and United States Geological Survey (USGS) optical range scans (LIDAR) of Tidmarsh that were converted

---

[1] Doppelmarsh is a collaborative project, with contributions from me, Brian Mayton, Spencer Russell, Don Derek Haddad, and Evan Lynch. Recent development has been lead by Haddad, whose Master's thesis [43] introduced radical new visual representations to the virtual world.

and imported into the game engine as height maps. As the Tidmarsh restoration transformed the terrain, the models were updated to reflect the state of the landscape. The terrain contour itself continues to be updated to reflect changes from the restoration as new data are collected, including from UAV-based imaging and new LIDAR flyovers.

Visitors can inhabit Doppelmarsh through a variety of different perspectives and related technological interfaces. The primary perspective is a first-person view, shown in Figure 3-2, that allows users to see and hear the site at approximately the scale of a human visitor to the physical wetland. This view works well for virtual reality displays, where the user's head movements are tracked and used to position the virtual camera. The first person view provides a realistic sense of scale on the ground, but traversing large distances in this way is arbitrarily slow (though much faster and less taxing than crossing the rough wetland surface on foot). Gaming-inspired solutions include fast-moving vehicles and top-views or flying, which let visitors take in larger sections of the virtual landscape. More radical, not yet developed interface concepts have included zoom-based traversal, where navigation would occur across scales and allow users to move from the microbial to the macroscopic satellite view. In general, Doppelmarsh's front-end interface flexibility on top of a shared notion of terrain-locked data and media has facilitated its rapid adaptation to many different end uses.

Real-time sensor data visualizations are automatically placed on the terrain with reference to sensor metadata, in positions matching their physical world locations. In contrast to DoppelLab, where the sensor visualizations were strongly animated by the sensor values and dominated the view, representations in Doppelmarsh are more deeply embedded in the environment. Weather conditions are extracted from the sensor data, which are then synthesized into experiential dimensions of the virtual landscape. For instance, wind data from sensor nodes are used to control the virtual world wind speed and direction, which animate the movements of the grass and the trees in the virtual marsh. Measurements from humidity sensors manifest as virtual fog, and the intensity of virtual rain is determined by an on-site rain gauge. For slower transformations in the environment, such as shifting color palettes brought on by the changing seasons or the accumulation of snow on the ground, features are derived from images captured from on-site cameras. Haddad shows in preliminary studies that users are able to discern season, climate, and other qualities of the environment from those subtle mappings [43]. In recent developments, also detailed in [43] and shown in Figure 3-4, Haddad introduced strange creatures and translucent figures that roam the

Figure 3-3: Doppelmarsh virtual weather patterns: fog is controlled by nearby humidity sensors, and rain intensity by precipitation meters. Sensor data are combined with machine vision on the camera images to control rendering of snowpack and vegetation.

virtual landscape, their appearances and behaviors linked to the sensor measurements. As these types of embeddings multiply, Doppelmarsh comes closer to being a place to inhabit; the processes that unfold there are discernable on its surfaces but originate within in the ecological processes of the environment at Tidmarsh.

Sound in Doppelmarsh is similarly ambient and embedded. Live sound from approximately two dozen microphones at Tidmarsh is used to build a sense of a realistic, unpredictable sonic space. If a distant animal sounds a call at Tidmarsh, a Doppelmarsh visitor not only hears the sound in the virtual environment, but can move closer to hear it better. Several generative music-scapes, also highly spatial, are available to be mixed in with the live sound. As an example, one of these, inspired by Wendy Carlos' *Sonic Seasonings* [14], is based on a combination of plucked ukulele string and singing bowl samples. The samples' pitches are determined by local temperature and timbres by humidity. The musical scale changes from a pastoral daytime experience to a more mysterious nighttime musical setting. Other soundscape programs can be loaded by the user at runtime.


Images courtesy of Don Derek Haddad

Figure 3-4: In recent work by Don Derek Haddad, strange creatures and translucent figures roam the virtual landscape, outward appearances and subtle behaviors linked to the sensor measurements.

3.1.2 Modular Software Components

Doppelmarsh is built on a set of software modules that have been adapted for use in a number of other projects. Its development environment is the game engine Unity3D, which also hosts various native code plugins for different platforms and uses. These components are briefly described here, for reference in the next sections and chapters.

Static Resources

Doppelmarsh makes use of a large database of recorded data, which we refer to as static resources. These data can be statically linked with the applications or downloaded at runtime from ChainAPI (Section 2.2.3) and other file servers. They include recorded audio files, manually acquired landscape datasets such as ground penetrating radar and topographic maps, photographs, and more.

For example, our systems access pre-recorded audio files from a place-based audio documentary system called Roundware (http://roundware.org). Developed and curated by Halsey Burgund, Roundware allows us to download location-tagged, user-submitted recordings that can be spatially rendered through our applications in AR and VR.

Real-Time and Historical Data

We rely on ChainAPI to keep track of and in many cases directly access real-time and archived data streams; as such, the data handling system reflects the RESTful design of ChainAPI [112]. This approach lets us decouple the state of the sensor network from any UI

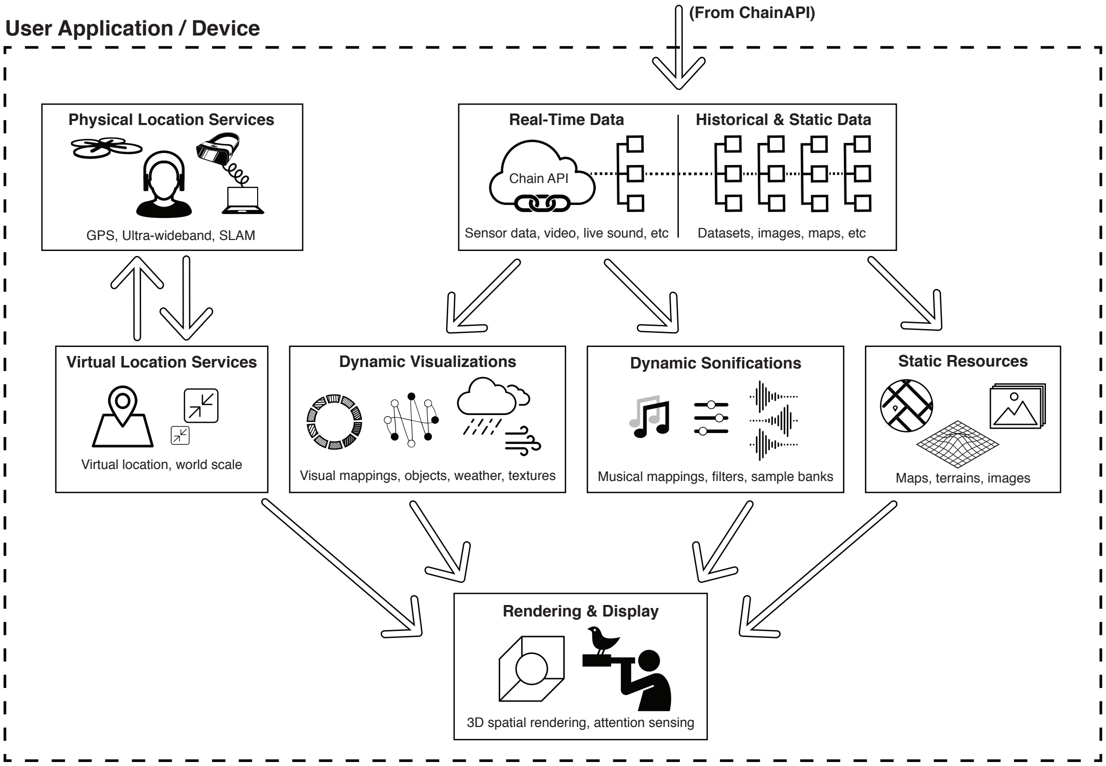


Figure 3-5: Software components of Doppelmarsh have been adapted to related AR and transpresence applications.

code; the application updates dynamically as new sensors come online or others go off.

On startup, the real-time data component, called ChainSync, downloads a summary digest of the sensors on a given top-level ChainAPI "site." This could be the closest site to the user in an onsite AR application, for example. The summary includes device IDs, geolocations, and a cache of recent data, as well as the unique device URLs. In this way, nodes and sensors can be instantiated in the virtual environment without any hard coding of link URLs, save for the first link to ChainAPI. After parsing the site summary, most of our applications subscribe to ChainAPI's real-time data streams, using a component called ChainSocket. Handlers within the application can subscribe to updates for complete sites, sensor nodes, or individual sensors.

As new data come in, they are forwarded appropriately, keyed on the unique ChainAPI URLs for each device. Some of our systems use ChainAPI to refer to data resources linked on other servers. For example, live audio streams are indexed in ChainAPI but served elsewhere. Some of these streams, such as the multichannel audio data, require special handling by the application and are passed to native code plugins before rendering.

Location and Navigation

Each platform and environment that runs Doppelmarsh provides different navigational affordances to the user. In our desktop version, the user moves with a mouse and keyboard, similar to traditional first-person video games. In VR, users can look around freely and move in a limited area, and can travel larger distances by pointing their controller in the direction they want to go and pressing a button to accelerate in that direction. In AR applications, data representations are overlaid onto the physical environment, allowing users full range of motion.

Because the data has a physical origin, we have developed a rich toolkit for managing geo-tagged data and mapping it into the virtual environment. Even in full-scale augmented reality environments, where the virtual and physical worlds are overlaid on top of one another in one-to-one mappings, global latitude, longitude, and elevation coordinates must be converted into local x, y, and z coordinates in meters. This transformation might require scaling up (to examine micro detail) or down into miniature (to take in activity across a large area). The module also enables virtual world control of geolocated devices on the physical site, such as a remotely operated UAV [105].

The location module requires the designer to place a single reference object with known geographic coordinates into the virtual scene. The library can then use that object as a reference to place objects in the rest of the scene, converting their geographic coordinates to virtual world coordinates. The module uses the Web Mercator projection so that the virtual world mapping can align with map tiles downloadable from many mapping services [7].

When designing a scene, it is possible to place objects directly using their virtual world coordinates. For example, in Doppelmarsh, sensor coordinates are retrieved from ChainAPI at runtime and placed using the location module, but other objects such as logs, trees, and terrain are statically placed in the scene.

Visualization

To enable a more spatial, embodied, and visceral interaction with the data, we created a toolkit for building modular 3D animated visualizations that can be mixed and matched in different applications. The architecture is oriented around a publish/subscribe model, where data modules (such as ChainSync) can publish to any number of subscribed consumers (such as the visualization modules). This allows us to build self-contained behaviors that can be easily added or removed.

Some visualizations affect the application as a whole, such as weather or virtual camera effects. Others are local and may vary throughout the site, such as a representation of temperature as measured by each of the sensor nodes. Both use the same publish/subscribe mechanism to subscribe to the data, but the code defining the behavior is instantiated differently. For global visualizations, a single instance of the visualization code can live at the top-level of the scene, and is placed when the scene is designed. For locally varying or per-node visualizations, we don't know ahead of time how many will be needed or where they will be. Game engines provide support for prototype objects that can be defined once and instantiated multiply, in our case at the location of each node. Adding a new representation involves designing a visual or sonic element and exposing one or more parameters to be mapped to a sensor value. For example, properties of a 3D object, such as color or size, could be mapped to humidity. The resulting prototype would then be automatically instantiated across the virtual terrain for each sensor. This framework allows visualizations to incorporate sensor-driven models that viewers can physically move through in virtual reality, such as a kinetic gas model driven by temperature [107].

Streaming Audio

The Doppelmarsh audio module passes decoded multichannel audio samples into the Unity spatial audio system, allowing individual channels to be presented as sources within the virtual environment. It significantly improves on the DoppelLab approach in both scalability and stability, and is built for a wide variety of platforms, including desktop, iOS, and HoloLens. Its primary purpose is to enable native decoding of Ogg Opus and Ogg Vorbis audio streams of high numbers of channels, a task for which no other software or library otherwise exists. Both Vorbis- and Opus-encoded streams are used for the Tidmarsh acoustic monitoring system. Once decoded, the streams are assignable to standard Unity AudioSources, making it easy to integrate them alongside other game engine assets. The module was originally written by Brian Mayton, with fixes and small changes from me, and was further adapted for use by the HearThere iOS application with features such as instant replay of the audio material.

Sonification and Generative Music

Much of the development of Doppelmarsh has focused on the sonic experience, which includes both informational auditory display and data-driven musical composition. While Unity offers a wide variety of visual effects and models, it lacks the facilities to compose spatial, richly layered generative music.

For his master's thesis work, Evan Lynch developed a framework to allow composers with limited or no knowledge about sensor data processing to compose musical pieces driven by ubiquitous sensing [80]. His framework, called *SensorChimes* is inspired by the wind chime, a prehistoric music-making wind 'sensor'. SensorChimes reimagines, generalizes, and augments the wind chime concept for transpresence environments.

Lynch's system was originally implemented as a library for the graphical programming environment Max/MSP, software commonly used by composers. SensorChimes streamlines the process of routing real-time and historical data from a sensor network into the composer's program. The library has also been ported to the PureData environment and made fully embeddable within the game engine, where it can be used to create rich and responsive real-time musical experiences that are coupled with the 3D visualizations. SensorChimes provides an interface for data from each device in the network as well as aggregate metrics over many devices, allowing for quick realization of innumerable musical mapping ideas. Building on Lynch's work, as well as improvements made by Spencer Russell, I adapted the PureData version of SensorChimes to run within Unity on mobile devices.

To date, four compositions have been written using SensorChimes and are available within the sonification software module. Each uses a different mapping strategy and explores a different part of the potential of sensor network-driven music as a new canvas for artists. Using the mobile version of the module, the pieces can be experienced directly on the landscape as fully geospatial generative music.

3.2 Transpresence in Built and Natural Environments

Sites for transpresence in the physical world happen when artists and designers suffuse built or natural environments with streams of sensory data. Occupants and passers-by can treat these responsive architectures as meditative spaces, prompts to sensory awareness, or background ambience, but in one way or another, visitors are invited to perceive the streams embedded within the physical space.

This is a huge area with too many precedents to cite and an extraordinary body of contemporary work, particularly from artists such as Janet Echelman, Natalie Jeremijenko, and David Bowen as well as composers such as Ryuichi Sakamoto. Imagining practical applications, computer interface researchers building on Ishii's ambient media [59] have considered the information carrying capacities and associated design factors of responsive spaces [45, 102]. The digitally and sensorially animate built environment is a relevant subject to address here, connecting interfaces, perception, art, and architecture.

In this section, I introduce three projects that constituted steps towards my development of transpresence: Data-Driven Elevator Music, Moss Listening, and ListenTree. The function of metaphor in the design of these systems is an important link to the larger dissertation and the subject of extended perception: the optical transparency of glass extending digitally to sound, for example, challenges perceptual expectations but maintains a perceptual sensibility. The spaces become natural vehicles for sensing outside of themselves.

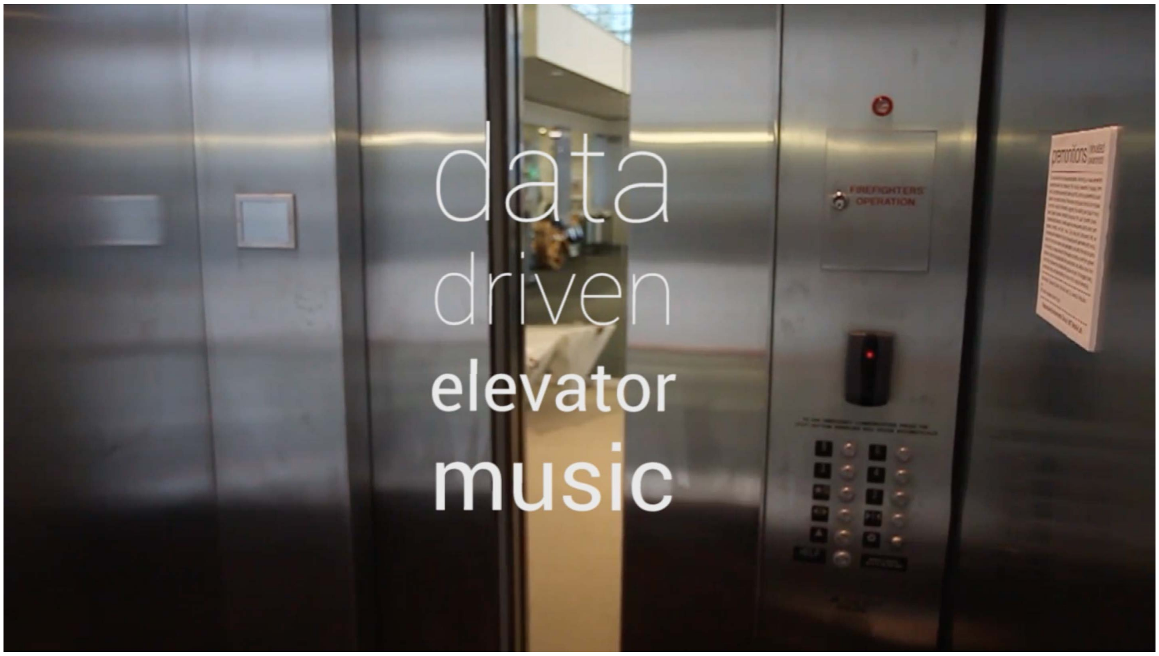


Figure 3-6: A still from a video of Data-Driven Elevator Music, available for viewing at [26].

3.2.1 Data-Driven Elevator Music

Fumihiko Maki, the architect of MIT Media Lab building E14, made extensive internal use of glass, designing for sight lines that would pass through large sections of the building and encourage connections between building occupants. Extending his idea, we considered that invisible sense media such as vibration and sound might also traverse the enclosing glass walls. In addition, networks of sensors inside and out capture many otherwise imperceivable dimensions of the environment, streams of which could also be embedded into the space.

A collaboration with Brian Mayton, *Data-Driven Elevator Music* brought live sound and sonification into the glass elevator at the MIT Media Lab, using the elevator's continuously measured altitude as its singular input [26]. The system was in place, with intermittent installed content, between 2011 and 2014. As a basic display platform, we introduced eight channels of sound to the elevator, one at each corner. A small box on the floor containing a computer, audio interface, and barometric pressure sensor (used as an altimeter in combination with a fixed reference) provided the means for spatial rendering of the sound. A subtle Doppler effect was applied overall to enhance a sense of motion in the sound.

In the first instance of the installation, shown in Figure 3-6, we mixed live sound from microphones throughout the building with musical sonification of occupant activity on each floor. The microphones and occupancy sensors were in place to support DoppelLab [63]. A generative musical piece was programmed by composer Ben Houge, with directional bursts of staccato bell-like sounds representing areas of activity. Subtle, privacy-preserving audio obfuscation at each microphone prevented any intelligible conversation from filtering through. The result, as the elevator moved through floors, was an ambient sense of where the action in the building was centered. In certain cases, an alignment between a sight line and the presented sound seemed to erase the glass walls. For example, as the elevator passed a floor with a ping pong table, the real-time sounds of the bouncing ball would enter the space with uncanny realism.

In the second instance of the installation, the elevator relayed live environmental sound from the microphones at Tidmarsh; vertical motion of the elevator translated into horizontal movement across the landscape. In this case, the transparency metaphor no longer held, and the motion effects were not physically intuitive. As a sound installation, the effect was beautiful and compelling for riders, but the space no longer evoked the same uncanny feeling of altered and extended presence.

3.2.2 Moss Listening

Not content with its presentation in the elevator installation but moved by the richness and enveloping spatiality of the audio from Tidmarsh, Brian Mayton and I sought again to bring the experience home with us. Presented at the MIT Media Lab's Other Festival in the Spring of 2013, *Moss Listening* was an 8-channel sound installation combining live microphone and hydrophone streams from the Arm site, our first dense acoustic monitoring station on the pre-restoration landscape.

For Moss Listening, shown in Figure 3-7, we wanted to create a contemplative multi-sensory space, with immersive spatial sound matched to the constructed setting. We collected and curated samples of activated organic matter from Tidmarsh, such as plants, moss, wood, mud, and peat. Inadvertently, we also collected quite a few insects and at least one toad found hiding in a bucket of mud. The approximately 8' by 12' indoor space was entirely shrouded with black curtains, and speakers were hidden in the corners. A hidden computer and audio interface were used to decode the live audio streams and create a spatial rendering convincing enough that, with eyes closed, one might believe she was truly there. A system of piping around the perimeter of the space was connected to a humidifier to feed in a steady

Figure 3-7: Moss Listening was a multisensory, 8-channel sound installation combining spatial live audio with curated samples of organic material from Tidmarsh in an enclosed, contemplative space.

stream of mist, and the organic materials were distributed so as to completely cover the ground. Two large found tree stumps were placed in the space for visitors to rest on.

Sitting on one of the tree stumps in the installation, visitors felt unmistakably in a building and in a marsh at the same time. Moss Listening was the culmination of our efforts to transmit essential sensorial elements of the Tidmarsh landscape to a remote place through physical installation, and as such it brought several significant insights to the work that followed. First, it demonstrated the presence-making potential of the spatialized live environmental sound in a multi-modal setting. Synchronized elements of the installation, achieved through the addition of scent, humidity, and dirt under visitors' feet, strongly enhanced the sense of presence and encouraged visitors to stay for longer periods of time. Second, the installation sparked my thinking about the power of selective attention to affect how and what we hear in rich, spatial sound. Because the source material was so active and spatially varied, two people sitting in the same confined space but attuned to different

Figure 3-8: Four images of ListenTree in Mexico City, for an installation titled *El Bosque de los Murmullos* (Murmuring Woods) at the National Center for the Arts.

sources of activity would come out of the space with very different experiences of what was heard. Watching visitors so outwardly focusing on sound in this way led me to consider how I might further extend individual attention.

Nice as it was, however, the problem with the installation was ultimately that the work itself was not novel. In our efforts to build an indoor contemplative space linked to Tidmarsh, we created a space closed off from everything else, immersive in the most isolating way. The preceding elevator work had succeeded by extending perception within its own sensory context, and failed when we tried to connect it, arbitrarily, to another. Moss Space, in turn, was "shadowed," as the anthropologist Stefan Helmreich puts it, by a problematic construction of the "listener as both apart from the world and immersed in it" [50].

3.2.3 ListenTree

Could a natural environment become a site for ambient transpresence? In the forest, perhaps more than anywhere, visible technology disturbs the scene, and mediating technologies deprive users of valuable, direct experiences of nature. Instead of removing visitors from their surroundings, could a technology bring visitors into closer contact? Considering these questions, and searching for unconventional ways of embedding information in the natural world, Edwina Portocarrero and I developed ListenTree, an audio-haptic device that would invisibly introduce sound and vibration into trees, enabling passers-by to hear sound through bone conduction when leaning against them.

Audition through bone conduction occurs when vibration is conducted through a listener's skull and into the inner ear, bypassing the eardrum. One of the earliest examples of bone conduction apparatuses is attributed to Beethoven, who is said to have compensated for his hearing loss by attaching one end of a metal rod to his piano while holding the other between his teeth. Exciter transducers have been used in art (as well as increasingly in consumer products) for several decades. Despite their widespread use, by seemingly magically producing sound through bone conduction and from inside objects, transducers continue to surprise users and audiences. In one early example, Laurie Anderson's installation *The Handphone Table* allowed participants, facing each other at a table, to hear sound when they placed their elbows on the table and their hands on their heads [1]. In his work *touched echo*, artist Marcus Kison used transducers attached to a bridge railing to reproduce through bone conduction the sound of a Dresden air raid that occurred at the same site [70]; the pose required to listen in that way, with arms shielding the head, makes visual reference to people sheltering in fear.

A visitor to our ListenTree installations would notice a faint sound appearing to emerge from a tree, and might feel a slight vibration under their feet as they approach it. By resting their head against the tree or its branches, they were able to both feel the vibration and hear sound emanating from within. To create this effect, a weatherproofed audio exciter transducer was attached to tree trunks or large roots several inches underground. A nearby solar-powered controller provided computation, audio decoding, and wireless network connectivity, and could drive multiple transducers independently using underground cables. Our approach let the technology disappear, producing an almost magical effect of sound seeming to emerge from inside the tree itself.

We presented ListenTree in four major installations[2], exploring a variety of audio source materials, including voices, live sound, music, and sonification. The installation contexts

---

[2] Portocarrero went on to do other ListenTree installations, described in her dissertation [99].

were very different, and elicited dramatically different responses from audiences. In each case, the technology worked flawlessly, but in my view, three of the installations could be considered failures, and one a resounding success. From ListenTree, I learned that the same technology applied in different contexts could serve as a mediating, distancing force or as a bridge to deep perceptual presence.

In installations at the MIT Museum (2014), at CHI Interactivity [98], and even in the fine art venue of the Montreal International Film Festival [31], where two composers were commissioned to write site-specific pieces for the trees, audiences' primary interests were always in the hidden operation of the technology. Was it a real tree? How was it done? What was the trickery? Once its operation was known, the experience was complete, and the knowledge became just one more piece of information to be shared. In each setting, we watched audiences queue up to press an ear to a tree, absorb several seconds of a longer sequence or musical work, exclaim in excitement, and move along.

Shown in Figure 3-8, the ListenTree installation in Mexico City was commissioned for the park of the National Center for the Arts, part of a festival celebrating the Mexican Day of the Dead holiday in 2014 [100]. In one area of the park, eight trees were activated with our devices, each telling a different story read by a different actor. The stories ranged in length from two minutes to ten. Over the course of one week, thousands of visitors visited the work, many of them young children. In stark contrast to the installations, audience engagement with the trees was deep and lasting, with significant numbers of visitors circulating amongst the trees for periods in excess of one hour, hugging each one tightly for minutes at a time. Visitors spoke of the living tree and the content we introduced as inextricably linked: "I felt like I was hugging a human being." Young children were seen speaking to trees outside the installation: "why don't you talk too?".

In Mexico, a deeply embedded mythology regards trees as connections to ancestry. There, the McLuhanian *message* of a sounding tree was at one time and inseparably to embrace our trees and embrace our ancestors. The content-carrying vibration served to erase the boundaries between what had been seen as distinct objects in earlier presentations of the work. Within the larger context of the thesis, this kind of blurring together of media and environment is the basis for the formation of new perceptual sensibilities. Here, the blurring comes about as a result of a site-specific cultural resonance. In the next chapter, physical and phenomenological links underpin the new perception.

Chapter 4

Transpresence Perception

4.1 A Vision of Extended Hearing

A woman stands at the foot of a path on an overcast fall day in early evening 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. 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 small rubber pads resting just forward of each ear.

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. She turns to face the pond, and 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 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.

In the woods, she takes her mobile phone from her pocket and observes her location on a map. She taps an icon of a sunrise, and in seconds the soundscape changes dramatically, now a blended mixture of the sounds of that morning's dawn and the dusk at present. What a difference only 12 hours can make! Returning to the phone, she taps another icon, representing 5 years back. The signs of morning wildlife disappear, replaced with a lonely whispering wind. At that time the site was an industrial farming operation, where widespread pesticide use eliminated the basis of the food web and its interdependent ecosystems.

4.2 To the Body

Where the last chapter introduced an environment-centered, top-down approach to cultivating extended perceptions through technology, this chapter switches to the bottom-up construction of first-person experience. Emphasizing that distinction, I call this chapter *Transpresence Perception*, and focus on work that aspires to the preface of the dissertation and the vision of the last section. My scope in this chapter is in the novel application of technologies to supporting and extending attention and sensory perception. More specifically, I focus on the carefully designed pairing of distributed sensing with wearable technologies to construct extraordinary perceptual perspectives[1].

I begin the chapter with a brief section describing three early wearable projects that allowed users to explore the world through sensorimotor action. In different ways, these projects explored the continuous, perceptual mapping of spatially variant parameters to haptic and/or auditory channels, and help contextualize the later work. Next, I introduce several AR projects that follow from the virtual world projects of the last chapter and lead into the major work of the thesis. Finally, in Section 4.3, I introduce HearThere, describing its design, development, and the results of the evaluations I conducted, leading into an extended discussion in the next chapter.

4.2.1 Early Work: Tongues, Fingers & Ears

Tongues, fingers, and ears are natural sites on the body for exploring augmented perception and extended sensorimotor experience, often serving as our first probes into the unknown.

---

[1] I originally titled this chapter "Sensory Superpowers," a label that several study subjects used to describe their experiences of my technology. I define a sensory superpower quite simply as a subjectively reported new and 'extraordinary' perceptual ability. By this definition, a sensory superpower would eventually become just as ordinary to its bearer as any part of them, extraordinary at times in the way that any sort of practiced perceptiveness might be. However, paired with the overused rhetoric of technological 'augmentation', the superpower label reifies a normative hierarchy of abilities I was not interested in endorsing here.

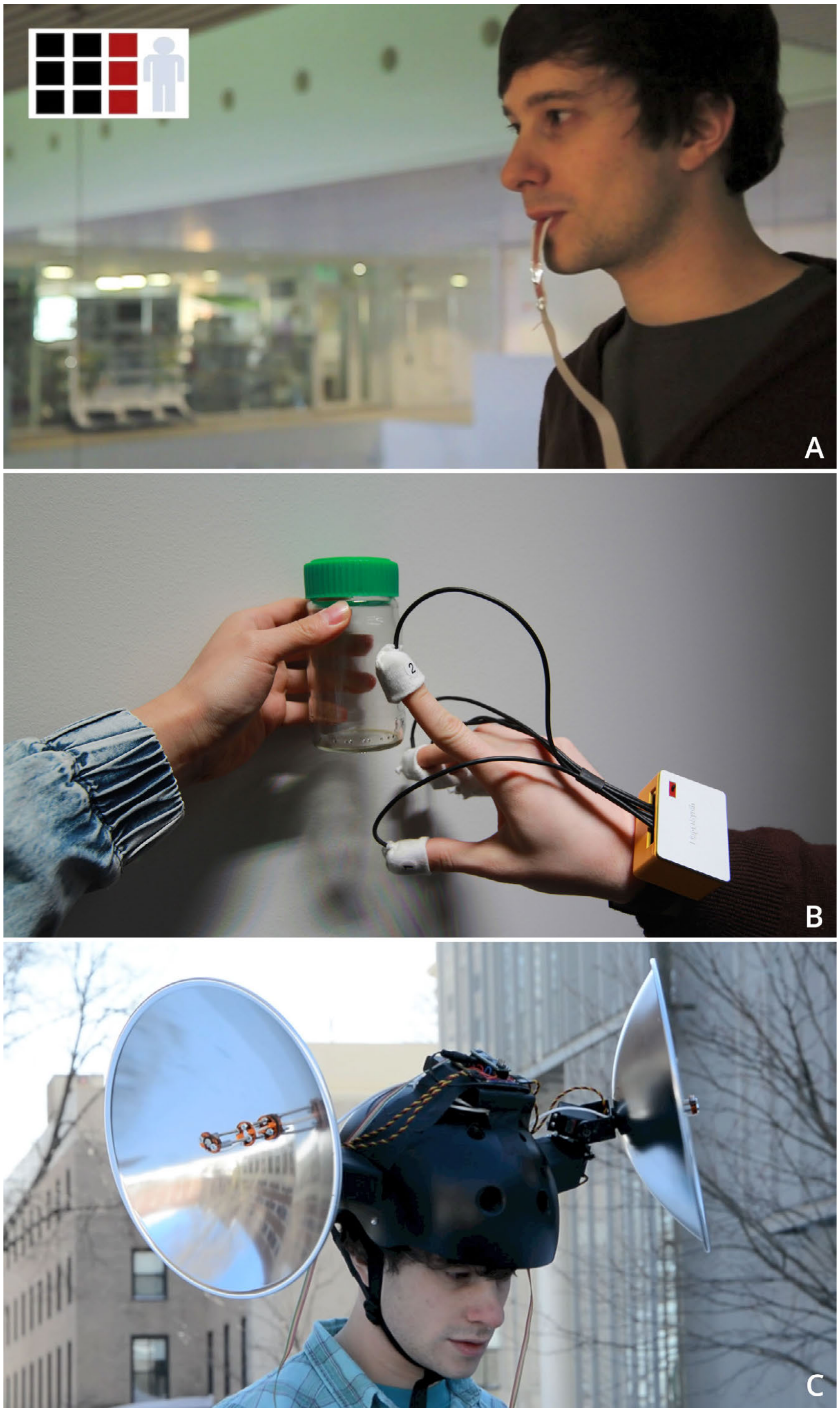


Figure 4-1: Playing with perception, from top, *sensation, action, and action in perception*: (A) Tongueduino, a low-cost, easy-to-fabricate sensory substitution device for experimenting with different sensor-to-tongue mappings [27]; (B) Fingersynth, a set of voice coil transducer rings and accompanying wrist-worn digital synthesizer for exploring object resonances and vibration conduction [29]; (C) PHOX Ears, an articulating pair of head-mounted parabolic reflector microphones, controlled by handheld joysticks and heard through stereo bone conduction, for dramatically heightening the wearer's hearing sensitivity [71].

They are extremely sensitive, receptive to the most minute vibrations, but just as often backgrounded or peripheral. They fill in the rich context—the detail that makes it a pleasure to sound out a word, run a hand along a grating, or listen closely to silence. My project-based experimentation on the body began with those sites, following a trajectory that parallels the evolution of my thinking on sensory augmentation: from a naïve notion of building new perception purely out of structured sensation, to the idea that perception emerges from action, and finally to an inquiry centered on the building and warping of physical space and time through the networked sensory prosthetic. This evolution is visible in Figure 4-1, which samples three, body-based projects from my early work.

My interest in sensory augmentation began when I read accounts of sensory externalization experienced by participants in Bach-y-Rita's tactile sensory substitution experiments in the late 1960s [5]. Presented with certain kinds of tactile stimuli on their backs, derived from a front-facing camera, subjects reported an eventual transition of the perception from the back to the space in front (I write more about this in Section 2.1). Even more than the directly practical benefits of the technology, my fascination with this gradual construction and elaboration of space—a cognitive transformation from inside to out—drew me into sensory research.

Bach-y-Rita later favored the tongue as the ideal site for both tactile-visual and tactile-vestibular substitution, and my first project in this space was a simplified tongue display unit (TDU), modeled on his early devices. With TDUs locked in patent-protected commercial development [6], building a device from scratch offered the path of least resistance. Made by printing, silver-plating, and laminating 25-pixel circuitry on polyethylene terephthalate (PET), and driven by a network-connected microcontroller, Tongueduino supported flexible mappings from wearable and distributed sensors [27]. I developed initial mappings from a pair of piezo whiskers and a magnetometer. In testing, users were able to repeatably discern shapes and pulse sequences presented through the display; however, I struggled to find a compelling reason to choose one sense/sensor over another. Acquiring a new sensory modality through sensory substitution requires a substantial commitment on the part of the user. Immediate needs felt by some users with disabilities are a driver of adoption (e.g. the already established visual TDU mapping for blind people and orientation mapping for vestibular assistance). But for more general-purpose sensory HCI, I see no broadly appealing sensibility worth the effort to learn a new mapping—no viable path to the commitment required for meaningful, SSD-based exploration in the world.

In a subsequent project, shown in Figure 4-1B, I considered ways of exploring the physical world sensorially through direct actuation. I built a musical instrument, called FingerSynth, in the form of a set of vibrating fingertip rings. Pitched in a scale chosen by the user, the rings would produce audible sound when the user's fingers came into contact with resonant objects in the environment [29]. In this way, FingerSynth used vibration to encourage a multimodal finger-object-ear practice, precipitating exploratory sensorimotor action in the world. In contrast to the tongue example, FingerSynth exemplifies the ways in which perceptual sensibilities, in this case a sensibility for object resonance, are all *out there*; technology can prompt us to discover them.

The third project in the sequence, shown in Figure 4-1C, explored the perceptual and social implications of a very radical transformation of hearing. Developed with Rebecca Kleinberger, PHOX Ears consisted of a pair of head-mounted, independently articulated parabolic microphones that allowed the user to sharply direct their hearing towards distant sound sources [71]. The microphones were mounted on 2-degree-of-freedom servo gimbals for precise positioning; joysticks in each hand independently controlled the gimbals. Signals from each microphone were presented to the corresponding ear through built-in bone conduction headphones, seamlessly mixing the amplified sound with the wearer's natural hearing. Users generally reported an uncanny but strongly compelling experience of the extreme sensitivity and radical warping of the sound the device brought them. With few exceptions, users quickly integrated the altered perception after several minutes of exploration. One user, a blind echolocation expert, expressed a desire to use the device while riding his bicycle, so as to increase his hearing range. Because users could not sense where their electronic ears were pointing, none found use of the electronic joystick method for control. Instead, users preferred either to leave the reflectors in one place or to refocus them by hand. I attribute the rapid perceptual integration we observed to two factors: first, learning required only sensory adaptation, and second, the learning process was conducive to physical world sensorimotor exploration.

My playful treatments of the tongue, the fingers, and the ears respectively demonstrate three facets of building and extending perception that carry forward to the later work: in the tongue display, structured sensation in its rawest form, absent sensorimotor contingency and therefore lacking sensibility; in the vibrating rings, the development of a new sensibility to object resonance through a sensorimotor action loop; and in the parabolic ears, an extension of an existing contingency through motor action, leading to integration.

4.2.2 Early Work: Sensor-Driven Augmented Reality

Augmented reality display of sensor data exists on a middle ground between sensory augmentation in its purest form and traditional data representation as disembodied abstraction. On the one hand, the location-based presentation of AR can bring sensor data back to the physical world they describe. On the other, most AR systems do not embed or blur representations into the physical world, and instead add layers on top of it[2]. Still, because AR *sets* information in the physical world, it raises perception-related issues such as externalization, as well as challenges of gaze and spatial selective attention. As a result, many of the design and technical/architectural aspects of AR sensor network display and interaction systems are shared by perceptual transpresence. This section describes two AR sensor data projects I developed that exemplify this middle ground.

Shown in Figure 4-2 (top), the Sensor Glasses app for Google Glass produces a location-aware visualization in the Google Glass head-mounted display. Sensor Glasses displays graphs of sensor data when the user looks at a sensor node, and allows the user to explore the available sensors on the node by swiping back and forth on the glasses' frame. Sensor Glasses crawls ChainAPI at runtime to build a structure of the available nodes, sensors, and data streams, and uses the device's onboard camera and QR codes printed on the nodes to sense which node is being examined. We designed concepts for an aggregate visualization feature, where GPS location and head orientation inferred from the onboard inertial sensors would be used to produce multi-sensor visualizations (e.g. heat maps) when a user gazed across a collection of nodes. However, we moved on from the project before fully engineering that feature.

Hakoniwa combined aspects of Doppelmarsh and concepts from the latter work on HearThere. With a name derived from the Japanese word for boxed garden, Hakoniwa renders a miniature version of the Tidmarsh terrain atop a physical table in the user's environment using Microsoft's HoloLens AR platform. Hakoniwa was co-developed with Spencer Russell. Representations of real-time data sources (e.g. sensor nodes and live

---

[2] A replacement of the senses with location-based data visualization of the corresponding sensor measurements would not be considered sensory. Taken to its extreme, AR can paper over the world so completely as to almost entirely hide it from view, as in Keiichi Matsuda's now infamous dystopian imagination [85].

Figure 4-2: Two examples of visual head mounted display and AR sensor data interfaces for exploring data from Tidmarsh. Top: Sensor Glasses for Google Glass displays data from the Tidmarsh sensor network when the wearer looks at QR code labels affixed to the sensor nodes. Bottom: *Hakoniwa* uses Microsoft HoloLens augmented reality display (A) to render a real-time sensor-driven miniature marsh landscape on a tabletop (B).

microphones) as well as static data sources (audio recordings) are shown as objects on the terrain. As shown in Figure 4-2A, the user is free to walk around the table and lean in to inspect the miniature Tidmarsh (Figure 4-2B). Sensor readings from the area are shown as floating text, and nearby audio sources (both live and pre-recorded) are played as spatialized sound, relative to the user's head position. Bird and insect calls detected by Tidzam (Section 2.2.4) are rendered on the scene as icons. As the user's attention (as indicated by the cursor in the center of the view) shifts to various parts of the scene, relevant information is displayed, culled from sensors within the user's field-of-view. The user's gaze not only determines the sensor data they see, but also amplifies the sound around that location, affecting the spatial roll-off curves for the nearby sources to allow the user to hear those sources more clearly. Multiple users can share a viewing experience, both in co-located scenarios (where the mini-landscape is rendered on the same table for all participants) and in remote scenarios (where two or more users can look at the same activity on the site from afar). Hakoniwa was one of our first radical departures of platform and scale from Doppelmarsh, forcing us to modularize many of the components that later went into the HearThere wearable platforms, and carefully consider mobile performance limitations. In addition, the simple, gaze-based culling and amplification of Hakoniwa is a precursor to the attention-based interaction in HearThere.

4.3 Extended Presence

The culminating project of this dissertation was envisioned early on as an interface that would leverage its user's intrinsic capacity for selective attention. Additional channels of information would become more voices in the cocktail party, and the device would compensate for the added cacophony by introducing mechanisms for enhancing and amplifying selectivity. As such, the interface would have to be spatially registered and allow for free exploration in the outside world. Auditory AR is well-suited to undirected, sensory experiences, as users can engage visually with the environment while shifting their attention fluidly from source to source in a 360° sound field. To allow truly free exploration, the device would not only have to be untethered, but also would need to be comfortable enough that a user could eventually forget they were wearing it [73].

HearThere, the hardware I built to realize this vision, is a bone-conducting head-mounted

Illustration by Nan Zhao

Figure 4-3: Early sketches of extended hearing at Tidmarsh.

auditory display device with touch gesture sensing, designed to interactively present spatial live and recorded sound along with AR musical data sonification. HearThere tracks the user's location and head pose to preserve the spatial alignment between virtual audio sources and the user's environment. Bone conduction transducers allow seamless mixing of real and virtual sounds. Users can hear through the device without obstructing hearing through their ears.

The associated user experience depends on a number of engineered components working together: the hardware and sensing device (HearThere) paired with a user-facing mobile application called Sensorium, a distributed sensing system (Tidmarsh Sensor Network), and several back-end database servers. This section focuses on the hardware and mobile application; the environmental sensing and server components are described in Chapter 2.2. HearThere and the accompanying Sensorium application have been through several significant versions and revisions leading to their current form.

Along the way to conceptualizing and designing HearThere, I undertook material explorations in the form of human listening studies in the lab. In these studies, I both replicated relevant psychoacoustic results from the literature and constructed an experimental context in which to examine firsthand how people listen to realistic 3D audio scenes like the one at Tidmarsh. These explorations led to concrete design choices in the development of HearThere, and informed the transpresence design guidelines introduced in Chapter 5. The next part describes the lab studies, along with relevant results and insights.

4.3.1 Exploring Spatial Hearing and Auditory Attention

Philosophers as well as researchers in cognitive science and neuroscience have widely theorized, modeled, and experimented with selective attention in both animals and humans, highlighting enormous challenges in the various disciplines within the field. My aims in investigating aspects of auditory attention were applied, not to advance the science within one of these disciplines but rather to design and field a very particular listening experience in which a participant's natural perception and attention would blend and project seamlessly into a digital realm. Inferring a user's selective attention in a spatial auditory display would enhance the experience by providing increasing detail in a particular representation as a user focuses their attention on it.

To that end, I designed a small lab study (n=12) informed by the literature that would allow me to observe listeners engaging with naturalistic audio scenes similar to the one at Tidmarsh. At the same time, I was interested in independently replicating existing experimental results as a way to learn more about canonical approaches to studying attention. I was surprised to find relatively little prior work on *spatial* auditory attention in humans, particularly of natural sound. This section details my study design and the findings I applied to the development of HearThere and perceptual transpresence. The study can be seen as a material exploration undertaken to better understand how to incorporate attention-like interaction into sensory wearables.

Psychoacoustic research has shown that a variety of physiological signal patterns are correlated with states of auditory attention. These include facial muscle tension patterns, pupillary measures, gaze, and heart rate variability. I engaged subjects in various 3D sound tasks, hoping to capture their task-evoked physiological responses using both cameras and off-the-shelf wearable sensor devices. I used observations made on individual subjects and comparisons between them to design the interactions for HearThere.

Study Design

The study consisted of two parts: first, a slightly modified implementation of a previously published listening test, called "Test of Attention in Listening" (TAIL), from [133]; and second, a set of tasks associated with playlist of recorded 3D audio scenes and synthesized sounds (sounds generated at runtime), interspersed with recorded instructions.

TAIL quantifies the modulation of auditory perception caused by attention, using reaction time as a measure and two main variables (frequency and location), where one or more variables are irrelevant to the task. The hypothesis, validated in [133], is that it will generally take longer to complete the task if a task-irrelevant dimension is varied. On an individual level, TAIL compares the subject's level of auditory focus on a simple discrimination task under distracted and non-distracted conditions. Variation in performance between subjects could indicate differences in the way they focus on sound: a subject who is strongly affected by the distractor, for example, might be more attuned to 'bottom up' attentional drivers or perhaps less able to follow a particular source in a natural setting.

The protocol for TAIL calls for subjects to listen to pairs of sine tones in rapid fire, varied randomly within the pairs in frequency and location (with additional very small random variations in duration and interval). There are equal chances that the frequencies will be the same or different and that the locations will be the same or different. If either are different, they are guaranteed to be sufficiently different so as to totally avoid confusion (the frequency difference between two tones being at least 2.1 equivalent rectangular bandwidths) [133]. In my implementation, one-hundred and twenty pairs were evenly distributed across three tasks conditions, each to be undertaken as quickly as possible: for the first forty pairs, subjects were asked to press a "yes" button if the *frequencies* were the same and a "no" button if different; for the next forty pairs, subjects were asked to press a "yes" button if the *locations* were the same and a "no" button if different; and for the final forty pairs, subjects were asked to press either button as soon as they heard the second tone. In contrast with the published study, which used headphones, the sine tones were presented through speakers at the front left and front right of the subject. This change was potentially significant, as my testing was conducted with spatial location as a primary variable.

The second part of the study sequence consisted of a playlist of recorded 3D audio scenes and synthesized sounds (sounds generated at runtime), interspersed with recorded instructions. The sequence is given in Table 4.1. In the first part of the experiment, subjects listened to the recorded scenes (e.g. a public square, a crowded theater, etc). Each scene was repeated three times. On the first two passes, the recorded instruction asks subjects to focus on listening, particularly on imagining the physical space represented by the recording. On the third pass, subjects were asked to draw their impression of the spatial arrangement of the scene on the paper provided, interpreting the sound as they saw fit. Five different

Scene	Description	Recording Type	Source
1	Prague public square from rooftop, ambulance	ambisonic recording	online/ambisonia
2	Tidmarsh April 11 morning, active with birds	spatial rendering of 6 microphone channels	self-produced
3	Audience sparse murmur	ambisonic recording	online/ambisonia
4	Tidmarsh nighttime mystery splash in water	spatial rendering of 6 microphone channels	self-produced
5	Desert bird call and response	ambisonic recording	online/ambisonia
6	Woman's voice (moving) and clapping distractor (moving)	spatial rendering of 2 recorded sources	self-produced

Table 4.1: Natural sound scenes in the auditory attention study sequence

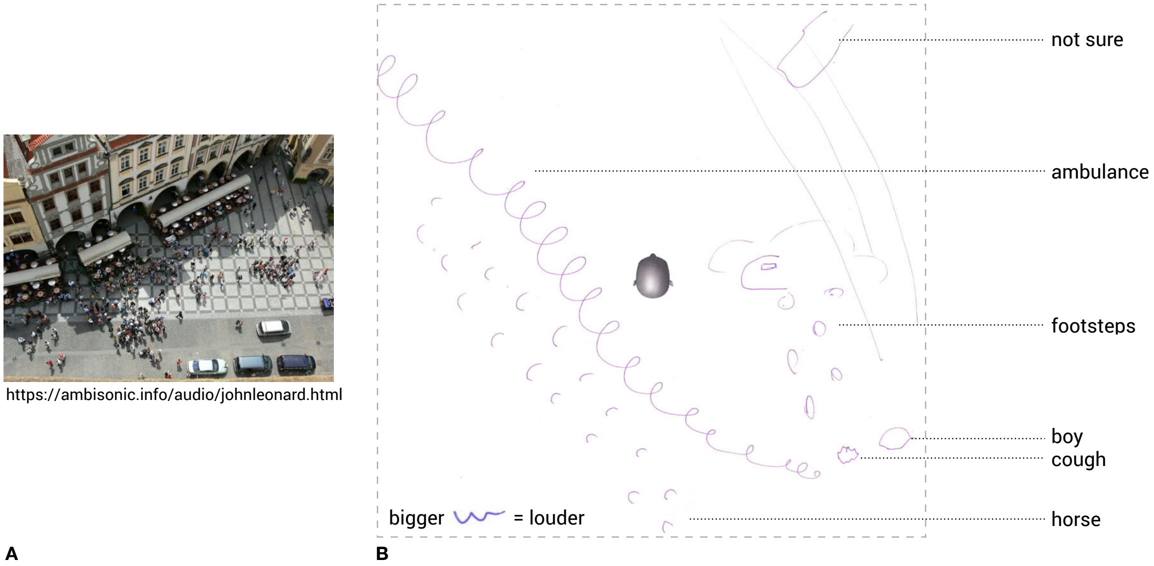
scenes were given this way, each approximately ninety seconds long. The sixth and final scene consisted of two moving voices (one target and one distractor) and no background ambience. Subjects were again instructed to listen first, and on the second play, to note down the path of one of the voices.

Study Protocol

After consent was given but before beginning the study sequence, subjects were fitted with a commercial heart-rate sensor (MedTronic Zephyr BioHarness 3), a head-mounted EEG device (Muse EEG headband), and an eye-facing camera (Pupil Labs Eye Tracker). They were then seated at a desk in an office room with facilities for running studies and synchronized data collection [134]. A camera capable of capturing pose, gaze, and facial expression was mounted in front of them. Six medium-sized loudspeakers ringed the walls of the room. A packet of prepared paper and a pen were provided, for use once the sequence begins. Once the subject was comfortably seated and ready to begin, the study administrator triggered the study sequence and synchronized data collection to begin on a timer. The study administrator then left the room and waited just outside the door until the subject had either completed the study or chosen to terminate the study early. At the conclusion of the sequence, the recorded instruction directed the subject to exit the room and notify the administrator.

Findings

The TAIL test I conducted confirmed the published results, where reaction time was slowed substantially by changes in the task-irrelevant dimension. This occurred despite the protocol



https://ambisonic.info/audio/johnleonard.html

Figure 4-4: On the third listen, subjects were asked to draw the spatial arrangement of the 3D scene. Some scenes were ambisonically recorded, others rendered spatially from mono files. At left, a view overlooking a public square in Prague from a rooftop, where a third party ambisonic recording was made; at right, one subject's drawing of the auditory scene.

changes I introduced. With respect to the headphones vs speakers (location) change, most subjects commented afterwards on the challenges of discerning location, particularly when the distractor was present: “frequency is easiest, location feels secondary and requires more thought”; “for location, it was hard for me to distinguish the source when the two sounds had different frequency.” Interestingly, several subjects commented on specific tonal characteristics that negatively influenced their performance on the location test: “‘I...interpret higher-pitched tones as being physically higher, too.” The strong effects of the distractor, particularly to force subjects to consciously process what otherwise felt unconscious, suggested that the perceptual interface of HearThere would have to avoid non-pertinent discontinuities. Distractions become flags for higher level cognitive processing to kick in where it needn't. As a result, transitions between audio states in HearThere, such as responses to Tidzam classifications and attentional dynamics, were later designed to occur slowly enough as to be almost imperceptible.

As discussed in Chapter 2.1.4, it is uncommon to measure auditory attention in natural settings due to the difficulty of managing the variables, including subjects' listening behaviors that are prone to sneaking auditory peaks, momentarily losing focus, etc. As my goal was observational and applied, I designed the study to free the subjects entirely from directed tasks during the first two listening periods, encouraging focused listening by reminding subjects of the upcoming spatial recall task on each repetition. I also designed and selected scenes I believed (and confirmed through piloting) would engage subjects' interest throughout. Post-study responses highlighted how engagement was indeed driven by realism in the spatial presentation and interest in the material: "The best image was the crickets and the huge splash. I was trying to figure out what the splash was—like a fish jumping or a person there trying to catch something." Figure 4-4 shows one of the ambisonically recorded scenes downloaded from the internet and used in the study ("Prague public square from rooftop"), along with the drawing of the scene produced by one subject. That scene in particular posed a localization challenge, as its rooftop recording caused unexpected spatial effects. The evocative drawing gestures in the figure are a sign that the scene not only created an image for that subject, but also a visual imagination of the sound absent any visual cues.

As expected, behaviors varied substantially in the uncontrolled natural listening portion, even within subjects, as interest and attention naturally waxed and waned. A common challenge reported by subjects was in the setup of the room, where the white walls did not match the richness of the sound, and subjects strongly desired to close their eyes (several subjects did close their eyes, despite being instructed to keep their eyes open for the sensing): "it was a little difficult to keep my eyes open, especially in the first, second, and fourth scenes, because I wanted to close them and imagine the space"; "I definitely think that my experience would be very different in a real world setting." Conversely, in the later study at Tidmarsh, I rarely observed subjects closing their eyes despite no instruction being given. Still, as subjects honed in on their individual spatial listening strategies, there were commonalities seen across most of the group. The most consistent observation was that during periods of focused listening, subjects remained still, eyes fixated within several smaller areas of the visual field. This pattern of fixation tended to relax on the second listening, as subjects looked around and sometimes lost interest. In the study, small area fixations closely followed general stillness of the head, a finding I later used in the attention-modeling interaction design of HearThere.

Figure 4-5 shows two subjects' normalized fixation positions in response to two different scenes, where fixations were defined as angular dispersion of the eye being less than 1°for a period of 0.5s or longer. The left panel shows the first listening, and the right panel the second. This data follows what the subjects reported when I asked them to describe their

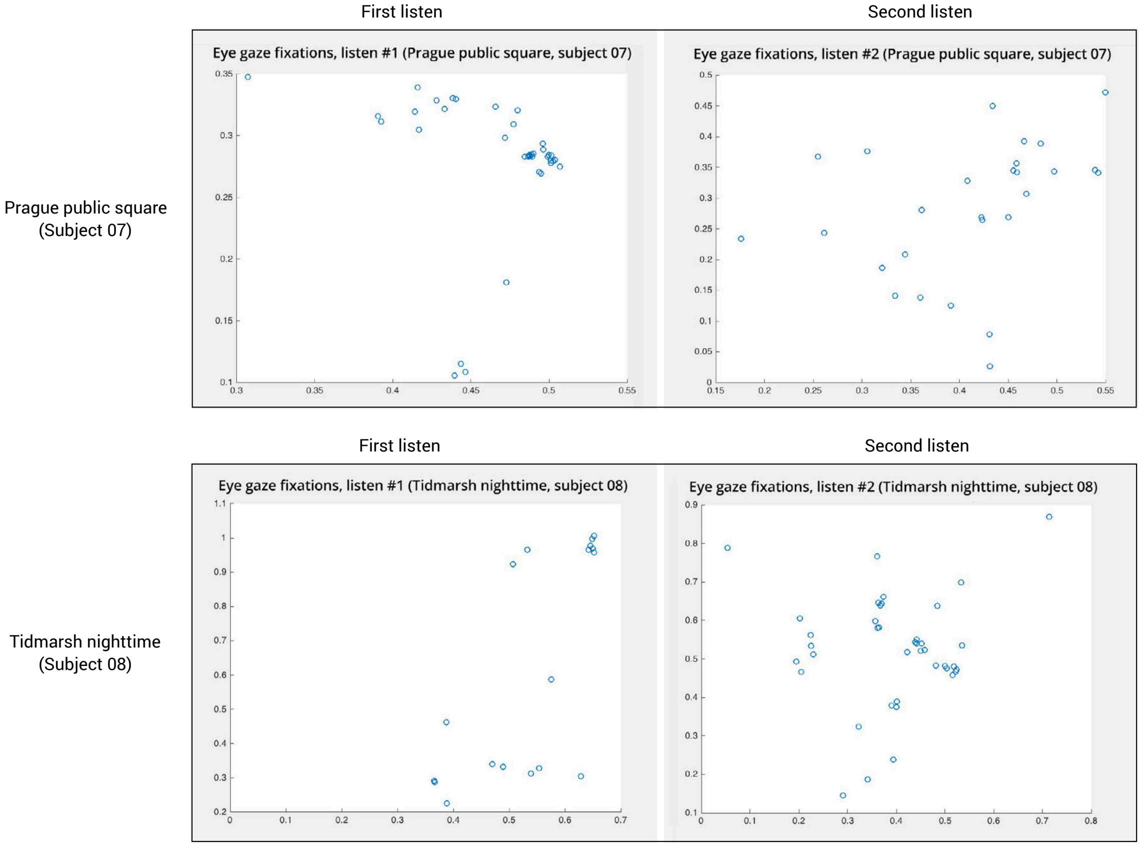


Figure 4-5: Normalized positions of eye gaze fixations (angular dispersion < 1°for > 0.5s) across two subjects listening twice to different 'natural' 3D scenes. On the first listening, intense concentration led subjects to fixate with significantly less variance than on the second. More relaxed, sometimes bored, and already anticipating the events of the scene, subjects began to look around.

strategies: “Occasionally I tried looking at specific speakers around me in order to gauge if a sound was emanating most strongly from it, but ultimately I found it most comfortable to just stare at the table, ground, or into the distance”; “I did move around but less so than I’d like to. When I looked around, it was mainly because to identify the location of a particular sound”; “I stayed mostly fixed, though I found myself looking toward particularly interesting or surprising noises.”

Ultimately, four findings carried forward from the lab study into the HearThere wearable. First, variations in auditory dimensions not relevant to a task cause conscious processing to kick in, particularly (according to subjects themselves) in disambiguating location, as well as slow reaction times. Second, subjects felt that a real-world setting would result in substantially different experiences and attentional behaviors than they exhibited in the conditions of the lab. Third, despite the sterile setting of the study, subjects felt that most of the 3D scenes were rich and realistic, highlighting in particular one of the rendered spatial scenes based on recordings from Tidmarsh; this was not only evident in subjects’ reports, but also in the evocative, gestural drawings they produced. Fourth, most subjects exhibited a pattern of focused listening behaviors that repeated across scenes, in which the first presentation of the scene was met with stillness and low variance in the position of eye fixations, and the second presentation was met with more looking around as well as a distribution of fixation positions with higher variance.

4.3.2 HearThere Hardware

The HearThere device is a bone conduction headphone with onboard sensing that can be used by paired applications to affect the real-time rendering of presented audio. Figure 4-6 shows the complete system paired with a custom-designed mobile application called *Sensorium*, detailed in Section 4.3.3, along with the associated Tidmarsh sensing and cloud infrastructure for producing the extended hearing experience.

Figure 4-7 shows the evolution of the hardware to its current form. The first version of the sensing component was designed by Spencer Russell as part of his 2015 master’s thesis work [113], which focused on auditory localization of virtual sound sources superimposed on indoor and outdoor environments. HearThere v.1 operates independently of the headphone system, and was used to validate the sensor technologies for indoor and outdoor auditory AR applications. Based on the results of that work, the current version, HearThere v.2,

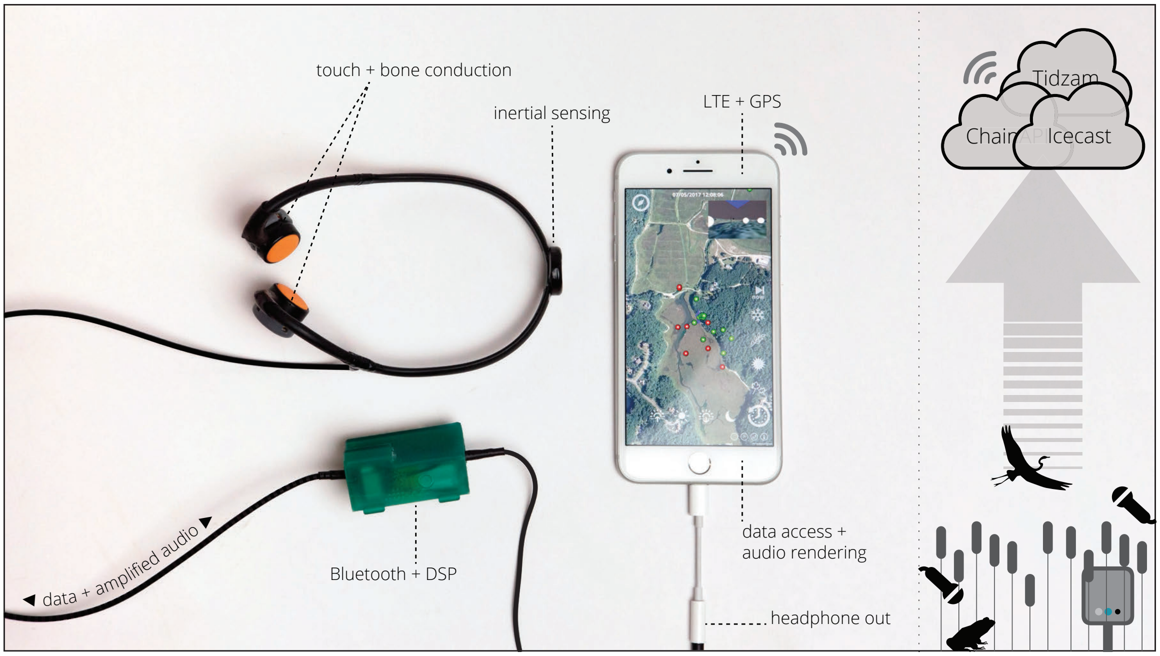


Figure 4-6: HearThere end-to-end system (final version), with wearable hardware, mobile software, distributed sensing, and back-end servers for processing and streaming data.

is designed around a comfortable headphone/transducer platform, while also allowing for additional custom and off-the-shelf wearable sensors to be flexibly added on.

The augmented reality functionality depends on two main sensing functions shared by all the HearThere systems: orientation (head) and location sensing (head/body). Head motion plays an important role in human sound localization ability, particularly where static cues can lead to challenging ambiguities, such as in front/back discrimination [128, 123, 88]. Brimijoin and Akeroyd show that for subjects facing contradictory information, head movement cues tend to dominate spectral cues for content below 8 kHz [12]. Finally, latency has been shown to be more important to localization than either sensor update rate or HRTF measurement resolution, and must be minimized [115].

HearThere v.1 Electronics

HearThere v.1 was designed by Spencer Russell as a development board for evaluating relevant sensor technologies and validating the auditory AR approach under different conditions. Parts of that work are summarized here. Information about its design and accompanying experiments can be found in much greater detail in Russell's master's thesis [113], as well as in a paper we jointly published [111].

The v.1 hardware integrates a 9-DOF inertial measurement unit (InvenSense MPU-9250 IMU), an ultra-wideband (UWB) radio module (DecaWave DWM1000), and a combination microcontroller and Bluetooth 4.0 (BLE) radio system-on-chip (Nordic Semiconductor nRF51822). The sensors communicate with the microcontroller over SPI. The device has three buttons for switching between operating modes, and a firmware-controlled RGB LED for feedback. It is powered by a rechargeable lithium polymer battery and includes a battery charging chip. An onboard SD card slot is available for data logging.

The UWB module allows HearThere v.1 to perform precise (20-cm) indoor localization when in range of fixed anchors, falling back to GPS on the host mobile device when out of range of the UWB anchors. This multi-scale localization scheme brings flexibility to the platform, opening up indoor/outdoor use cases. In an outdoor human subject test with 6 volunteers placing markers on their estimated locations of virtual audio sources, source localization ability fell from a 5.3m average error to 3.1m average error when subjects moved from GPS to UWB coverage. Switching from bone conduction (Aftershokz AS450) to in-ear monitors (Etymotic ER-4) only slightly reduced localization error, more significantly when counting sources that subjects failed to perceive at all. This effect can be attributed to the bone conduction headphones being quieter than the in-ear monitors, as well as the added environmental noise resulting from open ears in the bone conduction condition. Still, tests with HearThere v.1 confirmed that bone conduction with head tracking is a viable approach for creating auditory AR experiences. For all these tests, Bluetooth was used to connect the hardware to a purpose-built mobile application, in turn connected to a remote server. This model was later extended in pairing HearThere v.2 with the Sensorium application.

The HearThere v.1 development board does not have any specific head-attachment mechanism or headphone pairing, but rather functioned simply as an experimental head location and orientation tracker. A repurposed off-the-shelf head-mounted camera strap (GoPro Head Strap) and commercial headphones were used to conduct the user tests and experiments.

HearThere v.2 Electronics

HearThere v.1 served its purpose as a development platform for sensor and initial concept validation, but its size and form factor make it cumbersome for use in the field. The v.2

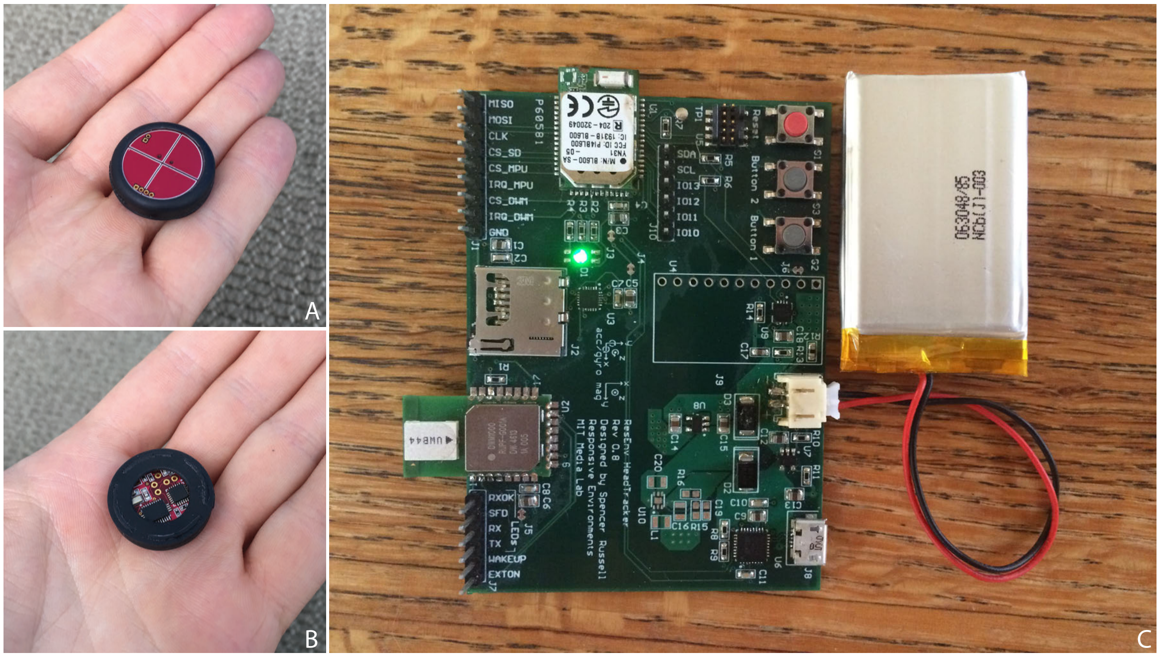


Figure 4-7: HearThere v.2 electronics (A, B) side-by-side with v.1 development board (C).

hardware is a ground-up redesign that takes into account lessons from the previous hardware and significantly shrinks the hardware footprint. Its shape is designed to closely integrate with a repurposed commercial bone conduction transducer that can be mounted on various platforms.

Head tracking performance was upgraded by switching the IMU to the Bosch BNO-055, a 9-DOF sensor with integrated orientation fusion and auto-calibration capabilities. The Bosch IMU can be programmed to provide quaternions directly, simplifying the firmware, increasing fusion performance and update rate, and all but eliminating orientation drift problems we experienced with the v.1 hardware. The UWB radio was removed, as the intended use of the v.2 device is outdoors with large area coverage, and only GPS would feasibly scale. The buttons and the SD card slot were removed and a touch sensing chip (Microchip CAP1188) was added for user interaction. The nRF51822 microcontroller and Bluetooth platform was maintained in the upgrade, though the module itself was swapped for the Raytac MDBT40, mounted on a breakout board (Redbearlab BLENano) with integrated power regulation.

The sensors communicate with the microcontroller over I2C, run over wires from both sides of the headphone assembly to a custom enclosure in line with the headphone wire. The same enclosure contains Aftershokz bone conduction audio signal conditioning circuitry and power management. A minor hardware revision added a separate power system for the sensing components, which had previously been causing intermittent noise in the audio. Over several design iterations, steps such as transmission line shielding were taken to improve signal integrity. Early versions also suffered from orientation drift due to magnetometer interference caused by proximity to the transducer magnets, corrected by moving the IMU to the back of the headphone assembly.

The touch sensor boards are housed in individual enclosures fitted to the backs of the transducers, with wiring run internally for maximal comfort. Significant care was taken to make the entire system as comfortable and simple to wear as possible, including in the selection of cabling, balancing signal integrity with physical compliance, size, and durability. To that end, a number of cable types were tested, including custom headphone prototyping Litz wire cables furnished by Bose. The latter provided maximal durability and comfort at the expense of signal integrity, due to lack of internal shielding over parallel runs of analog and digital signaling. The final design uses Mogami Ultraflexible W2880, a soft 6-conductor shielded cable intended for use in magnetic head leads for computer drives. After cable selection, significant effort was also put into strain relief design to maximize robustness in the field. These efforts were made in response to frequent failures of early versions due to unforeseen amounts of strain in regular usage.

Modular Flexibility: HearThere Pupil & Muse EEG

The combination bone conduction transducer and HearThere sensor apparatus were modeled for modular attachment to various head-mounted platforms, such as helmets and glasses. One of these platform alternatives is the Pupil Labs Eye Tracker, an open source glasses-mounted eye camera system [68]. Patterns of eye movement and variation in pupil diameter have long been associated with cognitive load and selective attention in the cognitive science literature [64, 16, 53]. As such, a glasses-mounted design could prove broadly useful for future auditory AR research, particularly where attentional dynamics are of interest. For my specific purposes, the eye tracking platform proved cumbersome in the field due to its obstruction of free movement (and reduced user comfort), susceptibility to sunlight, and significant compute resource requirements making it unsuitable to mobile, real-time use.

Another HearThere-compatible platform was the Muse EEG research variant, a consumer EEG device modified to strip away parts of the enclosure and use electrode stickers instead

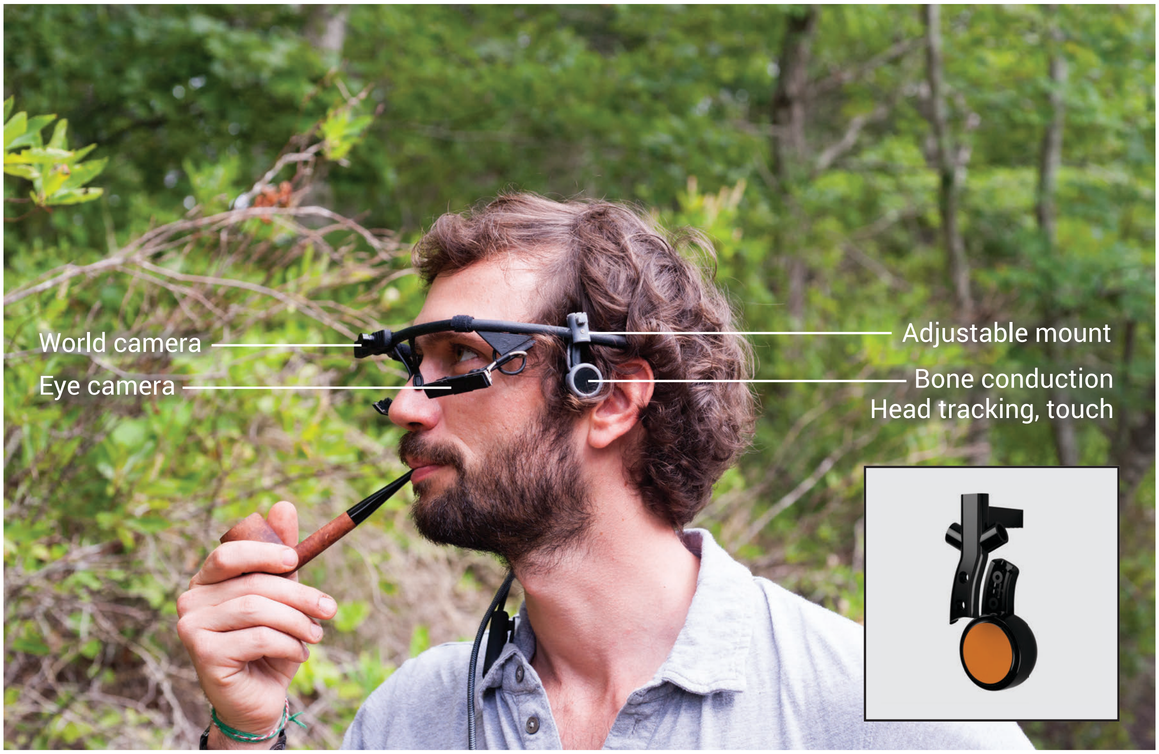


Figure 4-8: The HearThere system was designed for easy attachment to a variety of head-mounted platforms, including the Pupil Labs Eye Tracker, shown here. Inset: the bone conduction transducer and adjustable attachment mechanism.

of conductive rubber. To pair HearThere with the Muse, I used the HearThere headphone shown in Figure 4-6, which has a low-enough profile to allow both sensors to be worn comfortably. Sensorium, HearThere's mobile software application, supports simultaneous wireless connections to the HearThere hardware and the Muse EEG, as described in Section 4.3.3.

Figure 4-8 shows the assembled system. It uses the v.2 hardware but replaces the headphone platform with the glasses frame, using a custom-designed adjustable attachment mechanism. The system was validated and found to be field operational, tethered to a laptop in a backpack for the required computer vision processing and a network stream of eye tracking data feeding from the laptop to the Sensorium mobile app described in section 4.3.3. The software's responses to the eye measures are also described in that section.

Firmware

The core firmware is shared across the HearThere hardware family of devices, which were all built around the same microcontroller. The code is written in C and runs on the nRF51822

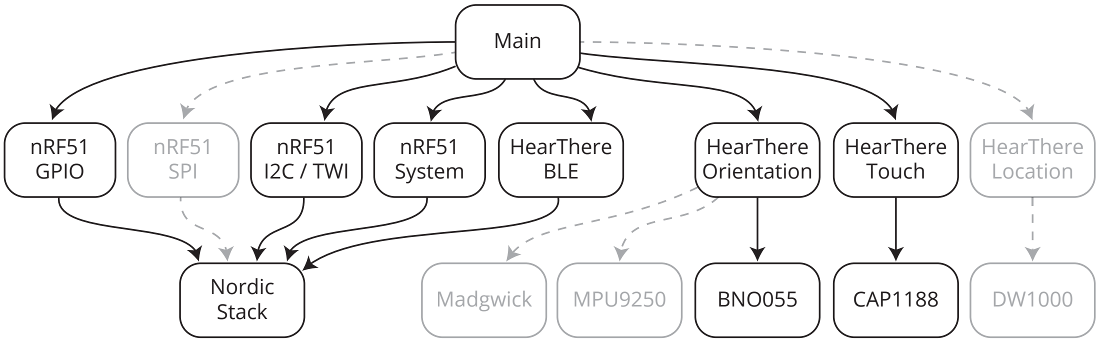


Figure 4-9: Firmware architecture inherits from HearThere v.1 [113]. Modules omitted from the v.2 revision are in grey.

chip, which is based on the ARM Cortex-M0 platform. In the v.1 hardware, the firmware continually ranges to all UWB available anchors, takes readings from the IMU, maintains a BLE connection to a host mobile device, and updates the device with range and orientation measurements. The Madgwick sensor fusion algorithm required for orientation runs on the microcontroller [81]. In the v.2 firmware, I2C (TWI) drivers and associated Bluetooth characteristics were added for both the Bosch IMU and the capacitive touch chip. Because the Bosch IMU performs sensor fusion internally, the Madgwick algorithm could be removed, streamlining the processing loop.

The firmware architecture is inherited from Russell's v.1 design, detailed in [113]. Briefly, as shown in Figure 4-9, the architecture is modular, where each module consists of a header (.h) and source (.c) file. Platform-specific hardware drivers such as the Nordic TWI interface are abstracted for easier portability of the modules. The orientation, touch, and location modules each initialize and read data from their respective peripherals, calling a callback when data is available. More information about the firmware can be found in [113].

Designing for Comfort

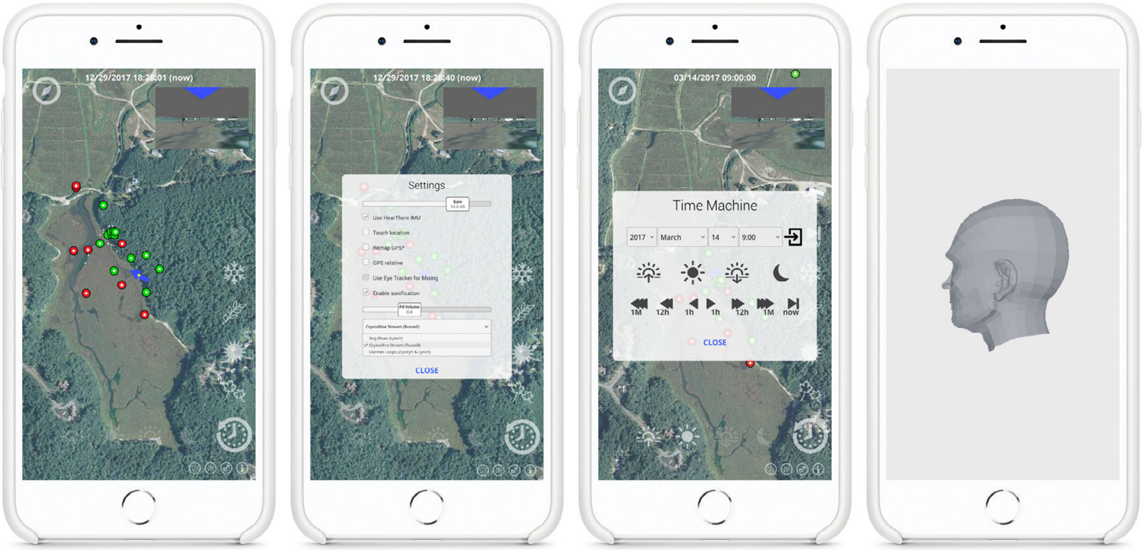
Issues of comfort and wearability are receiving increasing attention in the sensory substitution and augmentation literature [73]. In practice, researchers have observed a correlation between groups of users reporting inconvenience or discomfort of a device and a failure to integrate the substituted modality [92]. Particularly in cases where necessity is not the main driver for adoption, it has become clear that more comfortable devices offer shorter and less intrusive paths to sensorimotor integration.

As such, designing for comfort was a major concern in the development of the HearThere, which went through a number of iterations before settling on a hybrid custom and off-the-shelf solution. Several commercial headphone platforms were tested, along with a ground-up custom design. Bone conduction requires constant and extremely consistent contact with the head, adding a challenging constraint to the design. The hybrid solution leveraged the significant commercial investment in industrial design made to develop the off-the-shelf platform ultimately selected, the Aftershokz AS450. To build HearThere, the AS450 device was completely disassembled and rewired. Its internals were fit in custom enclosures along with the HearThere circuitry. As such, wearing the HearThere device feels exactly the same as wearing the unmodified commercial headphone, but with the added functionality of head tracking and dual touch gesture sensing.

Bone Conduction

Audition through bone conduction occurs when vibration is conducted through a listener's skull and into the inner ear, bypassing the eardrum. Tests with HearThere v.1 and an unmodified commercial headphone established the suitability of bone conduction with precise head tracking for spatial auditory display, with users reporting realism and successfully locating virtual sources in the physical environment, with only slightly reduced accuracy compared to in-ear monitors (IEM) [113, 111]. Perhaps due to the power of head tracking to compensate for spectral cue ambiguities [12], the substantially reduced accuracy previously reported in the literature was not observed in our case.

As reported in Section 4.3.2, the main challenge with bone conduction in a source location identification test rests with its lower perceived volume. In our tests, this resulted in occasional misses. Turning up the gain to match the loudness of the IEM results in a noticeable tactile vibration, which many users find uncomfortable or distracting. The most exciting discovery of using bone conduction for auditory AR relates to its perceived realism. This was initially observed in informal testing with both novices and experts, who when blind or blindfolded reported credible illusions that completely virtual sources were real, provided they were plausible. With fully non-occluding presentation, dry recordings of the virtual sources, and virtual room reverb matched to the user's physical space, it can be difficult to tell virtual from real without visual cues. This preliminary result was further confirmed in



Settings

Gain 12.0 dB

☑ Use HearThere IMU

☐ Touch location

☐ Remap GPS*

☐ GPS relative

☐ Use Eye Tracker for Mixing

☑ Enable sonification

Pd Volume 0.4

Crystalline Stream (Russell)

- Bog Blues (Lynch)
- ✓ Crystalline Stream (Russell)
- Clarinet Loops (Ziporyn & Lynch)

CLOSE

Figure 4-10: Selected control interfaces and display windows in the Sensorium application. From left: the default configuration, with the user position (GPS) and orientation (HearThere) indicated by the blue arrow, microphone audio sources indicated by green (active) and red (inactive) circles, and season preset buttons; the main settings panel, with user-selectable sonification and gain options; the fine-grained control 'time machine' interface; the data pass-through mode, showing the user's head orientation.

the HearThere field tests, discussed in Section 4.4.2 and in the next chapter.

4.3.3 HearThere Mobile Software

The HearThere hardware can be paired with one of several custom mobile applications that communicate over Bluetooth with the head-mounted electronics, handle network connectivity, and produce audio for the headset. The applications incorporate several modules from the Doppelmarsh software stack, in some cases containing improvements that have propagated back up to mainline Doppelmarsh development.

The first HearThere mobile application was used to validate the tracking system and conduct the early user experiments described in the previous section [111]. It had facilities for keeping statistics on hardware activity, forwarding data to a server, and spatially rendering static looping sounds. Its data-forwarding and system monitoring features have been incorporated into the current application, streamlining development by allowing the mobile device to pass through user interaction and wearable sensor readings to a desktop simulator.

The current software, called Sensorium, is visible on the mobile phone screen in Figure 4-6.

It plays a major role in delivering the extended hearing experience of HearThere. Developed for Apple's iOS operating system, Sensorium performs HearThere's audio rendering and user-sensing functions, allows the user to explore the recorded and live sound, and exposes a basic configuration interface.

Its minimal interface is designed to take as little user attention during the listening experience as possible, and rather spend the majority of its running time performing rendering tasks in the user's pocket. With the exception of the set-and-forget configuration options and external device pairing menus, direct user interaction is limited to the selection of historical time and date (when historical data is desired) or present time (real-time) audio. For those options, large preset buttons in the default view are there to encourage users to avoid any extended interaction. Figures 4-10 and 4-14 show several of the user interface panels. In regular operation, most of the user input is implicit, driven by onboard and on-body sensing.

The tasks of the Sensorium software are as follows:

- Managing the Bluetooth connection to HearThere for dual touch and head orientation
- Locating the user/listener with either GPS, GPS remapping, or manual input
- Synchronizing the virtual listener with the user's head position and orientation
- Retrieving and decoding live and recorded multi-channel audio streams
- Retrieving and queuing live and historical sensor data streams
- Generating sensor-driven music (sonification) from a user-selectable composition library
- Placing virtual audio sources at the correct world locations
- Binaural rendering of the virtual audio sources relative to the listener
- Mapping ear touch gestures to auditory effects such as 'zoom' and replay
- Pairing with external sensor devices such as Muse EEG and Pupil Labs eye tracker
- Feeding sensor data and inertial features to a channel-weighting attention model
- Using the Tidzam cloud AI data stream to promote wildlife sound over wind noise
- Providing a 'time machine' UI for switching between live and recorded audio and data

Figure 4-11: Two users experiment with the ‘touch zoom’ feature of HearThere and Sensorium.

- Displaying the user’s position/orientation and sensor locations on a map
- Offering configuration menus, storing settings and displaying internal data in the UI
- Transmitting internal data over OpenSoundControl (OSC)

The next sections discuss the design and software behind these tasks.

Features

The main function of Sensorium is the dynamic spatial auditory AR rendering of large numbers of audio channels, controlled by onboard and wireless (on-body) paired wearable sensors. The audio channels can be sourced from multiple multichannel web streams, generated at runtime, or played from static files. Because Sensorium uses ChainAPI to load sources at runtime, it could easily be adapted to other sites and configurations, and is flexible to the addition of new sensors and streaming audio channels.

Sensorium has several features that set it apart. It combines auditory AR presentation with in-situ distributed sound capture, creating not an AR ‘layer’ but an **extended hearing experience** unlike any that exist today. Its **time machine** allows a seamless flow between real-time (live) sound and an almost two-year database containing approximately 200,000 hours of continuously recorded audio; users can listen to an acoustically consistent rendering “what it sounded like” at any location within the sensed area and across the entire database to the present. It allows the balance between the audio streams to be adjusted smoothly on the basis of the user’s listening behavior, factoring in their stillness, eye movement, and brainwaves to **infer attentional state**, and allowing for new sensors to be added in the flexible channel weighting scheme. Paired with the HearThere hardware, it translates touch

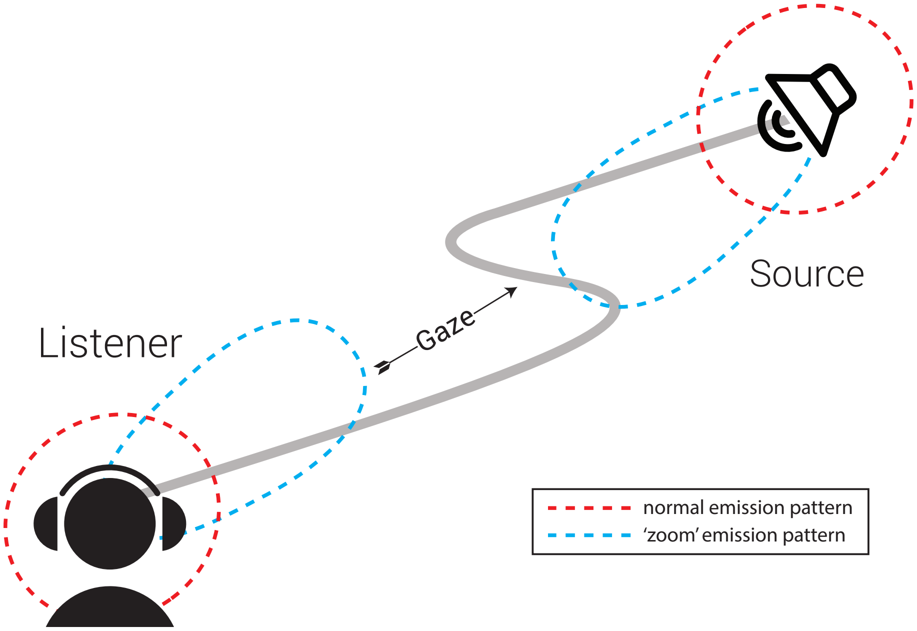


Figure 4-12: The auditory zoom feature changes the shape of both the sound emission and ear pickup patterns to create a smaller and more sparse but sharply focused sonic space in the direction the user is facing.

on the headphone into a unique **auditory zoom**, allowing users to hone in on individual audio sources using a quick-to-learn natural gesture. It hooks into Tidzam, a state-of-the-art, cloud-based, artificial intelligence audio classification system allowing wildlife identifications made alongside the listening human to additionally weight the audio channels, fielding a true **extended intelligence** system. It offers the first **geospatial and attention-modulated musical data sonification** by embedding the SensorChimes framework [80], including a set of generative compositions.

Auditory Zoom

The auditory zoom feature of HearThere with Sensorium allows a user to focus on individual sound sources by touching and holding the earpiece of the headphone. The intended use is both for locating the sound sources and dramatically isolating interesting events, using a conscious but intuitive gesture similar to cupping one's ear.

The touch has two effects on the sound. First, it directs all the energy of both the virtual sound sources and the ear's pickup pattern along the axis between the source and the listener's face, making the sonic space more sparse and sharply focused, as if it were a virtual sonic telescope. The physical microphones are omnidirectional, and the effect on the warped channel is similar to a gain increase, but reflections are reduced by the sharply elongated listener and source pickup and emission patterns. Second, it changes the master equalizer settings to sharply cut the lowest and highest frequencies and boost the mid frequencies on all the channels. The change is continuous, resulting in a pressure-like perceptual effect that feels similar to the effect of turning up noise cancellation. Combined, the gesture has the effect of cutting out wind noise and other background ambiences and highlighting activity in front of the listener when a microphone is present. When no microphone is present along the axis of the user's gaze, the feature makes the scene quiet, allowing the user to scan for source locations. Figure 4-12 depicts the effect of the auditory zoom feature graphically, showing the two pickup patterns. The morphing process between the two patterns is continuous and increases to a maximum while the earpiece is touched. Releasing the touch relaxes the scene into a more ambient spatial soundscape.

Modular Architecture

Figure 4-13 shows Sensorium's modular design, in which different configurations of on-body sensors and audio sources can be switched seamlessly into operation. For example, during operation, when the user pairs the EEG in the settings panel, the application will begin using the brainwave data to adjust channel weights in its attention module. Similarly, when the application begins receiving eye tracking data over the network, it begins ray tracing from the user's gaze to interactable objects in the scene, for precise targeting. If the sensors stop sending measurements, the system will seamlessly fall back to default operation.

This modular flexibility is achieved by keeping track of weights applied to each audio channel's rendering parameters. Each channel has an associated controller module, which keeps its state and receives updates, always interpolating smoothly between values to prevent noticeable discontinuities. Parameters include gains and equalizer settings, as well as spatial renderer settings such as directivity, sharpness, and roll-off curves. On each update, other modules within the application can pull a parameter value up or down by sending the channel controller a positive or negative weight. The channel manager keeps track of all the controllers, and forwards updated weights accordingly. Two high-level weights affecting multiple parameters are defined: loudness, controlling gains, EQs, and roll-offs; and spatial directivity, controlling the shape of the dispersed sound energy. This architecture can result in unexpected outcomes, such as where positive and negative weights equally cancel and no

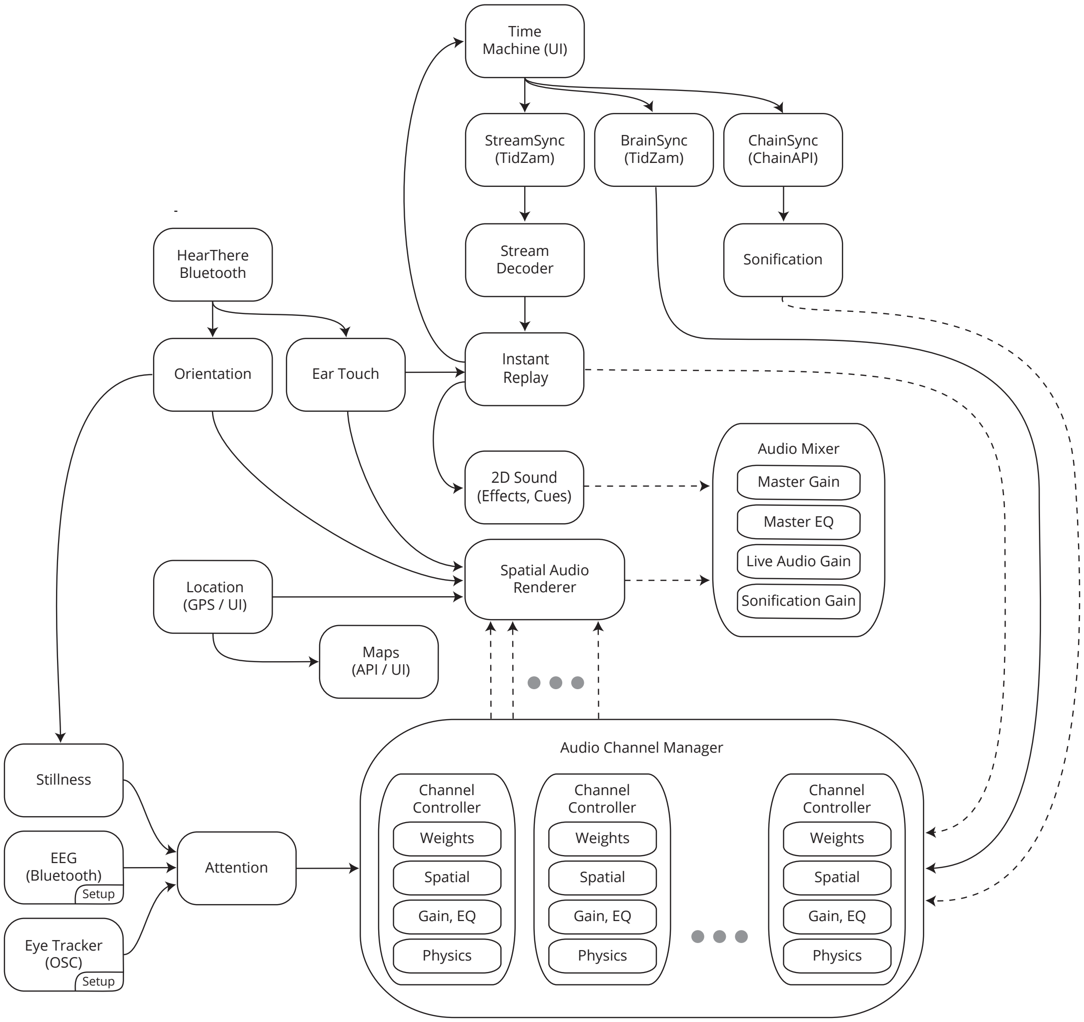
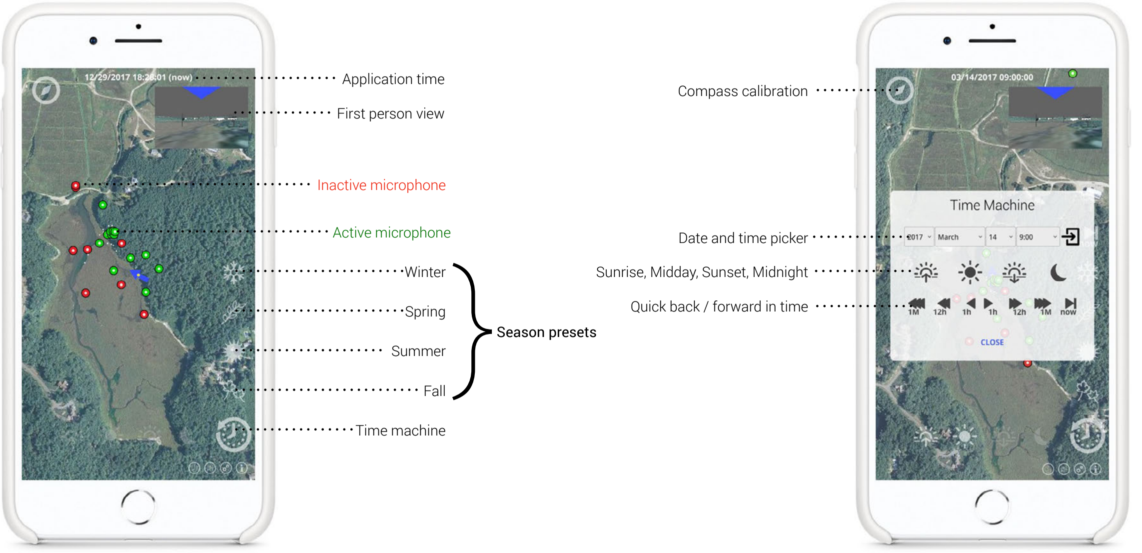


Figure 4-13: The Sensorium application's modular design allows for a variety of input and output configurations, with plug-and-play support for different wearable sensors (such as the EEG or eye tracker) and distributed audio sources. Internal audio connections are indicated with dotted lines.

effect is produced.

All internal sensor data are available throughout the architecture on a publish/subscribe model, allowing various modules to share the data to perform higher-level fusion (e.g. for generating channel weights) or for other wearable interaction purposes (e.g. head tracking and touch zoom).

The time machine module keeps track of application time, and makes requests to the sensor data and audio stream servers for chunks of data, which are buffered and dispensed



Default view

Time Machine

Figure 4-14: The Sensorium UI is designed to be as thin as possible, so as not to encourage its use during a nature walk, but common actions are made available through large preset buttons in the default view (at left). For fine-grained time selection, a series of drop-down menus are available (at right), but these options are cumbersome to use, and are intentionally kept out of view most of the time.

on the internal clock. The time machine also keeps track of the periods over which audio was recorded, so that users are prevented from accidentally seeking to gaps in the database; the module syncs its recordings database with that of the remote server.

Various modules are used to request and decode the data streams used by the application. Sensor data are delivered by websockets for real-time streams and downloaded in larger chunks for historical data access: ChainSync accesses the ChainAPI for data from the Tidmarsh Sensor Network, and BrainSync accesses the ChainAPI for Tidzam's AI analysis of the audio streams. The StreamSync module handles requests to Tidzam's audio file database server for historical sound retrieval.

Graphical User Interface

Sensorium's graphical UI, shown in Figure 4-14, is minimal. A default view renders local satellite imagery retrieved from the Mapbox API, and places interactable icons on the satellite image to represent user and sensor locations. The default view contains large icon buttons that activate time machine presets; touching the moon icon, for example, jumps the application forward to midnight. A smaller row of icons along the bottom of the screen access settings and monitoring panels, such as EEG pairing options and a mixer view. A window on the top right of the screen renders the first person (listener) view, useful for confirming the smooth operation of the head tracker. Finally, a hidden option renders a 3D head model and activates sensor data forwarding to a server using the OSC protocol, useful for prototyping and development. As previously stated, the intention behind the application's design is for the user to set up their session and stow their device in a pocket, interacting with the system through touch on the earphone as well as through natural exploring and listening behaviors.

Audio Decoding, Synthesis, and Rendering

Sensorium substantially extends the Doppelmarsh streaming audio module introduced in Chapter 3.1.2. The native code streaming audio module plugs into Unity to enable native decoding of high channel-count Ogg Opus and Ogg Vorbis audio streams. At the present time, no other software or library otherwise exists to do so. Currently, the system at Tidmarsh uses Opus-encoded streams for the multichannel audio and Vorbis streams for the one- and two-channel sources (e.g. camera-embedded microphones). Decoded audio is passed to standard Unity AudioSources, which are positioned using location metadata obtained from ChainAPI. In contrast with the Doppelmarsh system, which uses a now-defunct rendering plugin from 3dception, spatial rendering in Sensorium is performed by the Google VR audio components [39], which were stripped out of Google's comprehensive VR tools to make the application audio only[3]. I found the GVR libraries to meet or exceed the performance of competing platforms in terms of realism, features, and customizability.

Another custom plugin applies filters and other effects to the overall mix, after spatial rendering. This was used primarily to differentiate, by filtering, the audio zoom mode, cutting wind noise and bumping up mid-bands associated with wildlife sound. Chapter 3.1.2 introduces the sonification module used in Doppelmarsh, known as SensorChimes. I adapted the module for Sensorium, building it for the mobile platform, scaling it back to be performant with limited mobile resources, and allowing its musical sources to interact with the attentional systems used for other audio sources. A GUI panel allows users to

---

[3] Google's auditory and visual VR tools have since been separated and the audio-specific libraries have been rebranded as *Google Resonance Audio*.

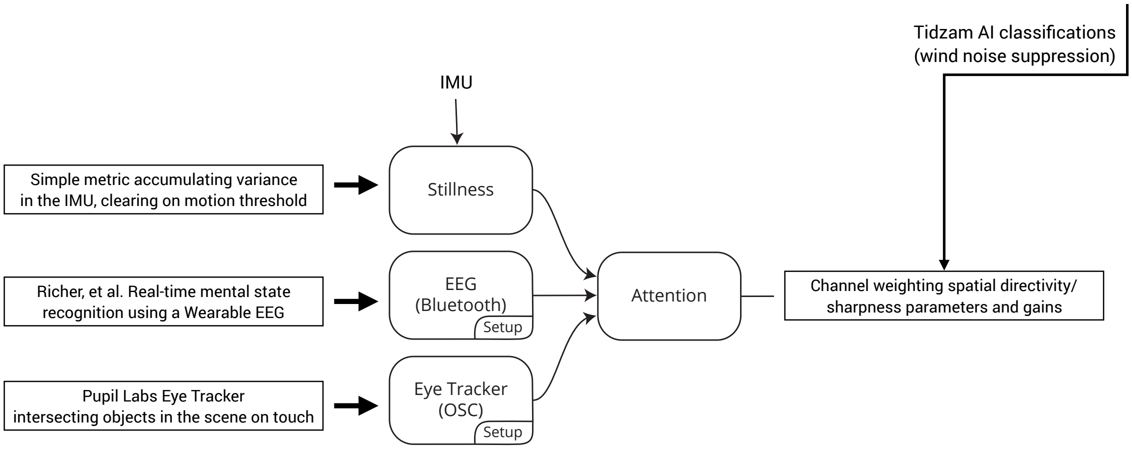


Figure 4-15: The Sensorium attention model flexibly combines input from multiple sensors and weights audio channels accordingly. The Tidzam AI can also weight the channels to promote wildlife sound or suppress wind noise.

choose which PureData (Pd) sonification patch to load, and to set other sonification-related parameters. In this way, Sensorium may represent the first platform to support playback of attentionally-modulated, spatially-rendered musical compositions.

Wearable Sensing and Interaction

A suite of internal and external wearable sensors are incorporated into the operation of Sensorium. The application supports a range of subtle perception-like and attention-like interactions, ranging from explicit input such as ear touch for auditory zoom to subtle rendering adjustments based on implicit input such as extended stillness. The modular architecture extends to sensor input, which is flexibly incorporated when available.

The most basic inputs used by the application are **location** and **head orientation**. GPS location measurements are taken at as high a resolution as the device will provide (1-m on current iOS devices). These are translated into virtual world coordinates registered with real-world positions of objects and sensors. Quaternions provided by the IMU are used to match the virtual head's orientation with the real-world orientation of the user's head.

Touch gesture interaction is available to Sensorium via two, four-pad capacitive sensors. The hardware and software support momentary touch, tap, swipe, and circular scroll gestures, though to simplify the interaction, only momentary touch was used, triggering the auditory zoom feature described in an earlier section.

Figure 4-15 shows the sensor inputs used by the attention module. The primary, always-on input to the module is a measure of **stillness** of the user's head. To compute this measure, the standard deviations of each axis of a circular buffer of head orientations are summed. If, on each update, that sum is below a threshold, the stillness metric is slowly increased. If the sum exceeds a threshold, the stillness metric is immediately cleared. If the sum of standard deviations is between its minimum threshold and maximum threshold, the metric is slowly brought to zero, interpolated downwards on each update step. The measure effectively tracks longer moments of stillness with a memory, but loses that memory at sufficiently large motions. **Attentional directionality** is inferred naïvely to be in front of the user (using the body and head to orient). This is a major shortcoming of the current approach. Significantly more lab studies, field experiments, and interaction design would be required to model and infer listening direction, if it is possible to do so.

The electroencephalography (**EEG**) measures are computed from sensor readings produced by a Muse EEG device [56]. Sensorium includes the Muse iOS native code plugin and derives its EEG measures online and in real time when a Muse device is paired to the application via Bluetooth. EEG measures are typically made after dividing the spectrum into five standard bands associated in the literature with different mental states. The Muse software internally defines those bands as follows: 1-4 Hz (Delta), 4-8 Hz (Theta), 7.5-13 Hz (Alpha), 13-30 Hz (Beta), 30-44 Hz (Gamma) [57]. Alpha waves are typically associated with calm and relaxed states. Beta and Gamma waves, on the other hand, are associated with focus and cognitive load. Sensorium uses a pair of entropy-based measures called "focus" and "relaxation." The measures were originally developed as inputs to responsive environments known as Mediated Atmospheres [135] and detailed in Robert Richer's master's thesis [109]. Both are based on relative spectral band powers furnished by the Muse EEG. Relaxation is based on Tsallis entropy in the Alpha band and focus on Renyi entropy in the Gamma band [109]. Richer's algorithm was ported to Unity, along with a C code plugin implementing the P-squared algorithm for efficiently calculating histograms of the EEG data [62].

I also tested similar (black box) measures included with the Muse SDK. In both cases, the focus measure was used to bolster the stillness measure. Increased focus increases the application's spotlight on the sources in front of the user as identified by the attentional directionality module). Increased relaxation was mapped inversely to reverb, drying out the audio to add clarity. Unfortunately, also in both cases, the EEG measure did not track focus as reliably as stillness. It is possible to drive the measure upwards by intentionally focusing (or even furrowing a brow), but moving around can cause similar effects and, just as often, cause the EEG to lose contact with the head. Because of these noise issues, I lowered the weight of the EEG measure to a point where it was difficult to identify any improvement to performance. After piloting the interaction, and given the added complexity, susceptibility to noise, and user discomfort associated with the EEG, I did not use it in the later studies.

The stillness and EEG focus measures were weighted and used to compute the application's **attention score**, which in turn increased the directivity and sharpness of the sources in front of the user, similar to but more subtly than the auditory zoom. Sources behind the user were ducked slightly. The overall effect was to add relative gain and clarity to the sound in front of the user when the user was still and focused, increasing slowly over time and cleared when the user moved again.

The **eye tracker** was originally included because of findings from the lab study (Section 4.3.1) showing a correlation between decreased variance in fixation positions and sustained auditory attention. However, a number of factors led to a much simpler use of the eye tracker in practice, and future work will investigate the possibility of using the eye fixation position variance in an attention model. Eye tracking data is provided by the Pupil Labs Eye Tracker [68] wired to a backpack-carried laptop computer running the Pupil Service application as well as a Python middleware server that culls and feeds fixation position data (or any other available eye measure requested by Sensorium) to the Sensorium app over the local network using OSC. The middleware server ensures that large volumes of data do not overwhelm the extremely limited compute and memory resources in the mobile application's update loop. In the current Sensorium app, when the user touches the earpiece to trigger auditory zoom while the Sensorium app is *also* receiving eye data, the rays projected with the eye data are used by the attentional directionality module to further pinpoint sources in the user's field of view. That way, a user can remain still and scan the surrounding environment for audio sources and microphones by eye. The limitation to acting only when the auditory zoom was triggered was implemented to limit noise and lost signals from the eye tracker in the field setting, caused by sunlight and user motion; users remained relatively still while engaging the auditory zoom, with an intention to focus. However, like the EEG, after piloting the Eye Tracking HearThere system (shown in Section 4.3.2), it was difficult to identify a net benefit, considering the discomfort and noise issues it brought in the field.

Both the EEG and Eye Tracker were not evaluated in the study for the reasons given, but the modular flexibility implemented to support flexibly adding and removing sensors from the system means that future work can evaluate and improve the sensors/interactions in more controlled scenarios.

Extracognitive Extensions

Also shown in Figure 4-15, Tidzam AI classifications are used in the interaction loop, providing an additional set of weights that, in the current version of Sensorium, are used to suppress channels picking up wind noise. Currently, audio stream buffering without embedded timecode results in unsynchronized Tidzam data streams and audio streams. As a result, the wind noise suppression module subscribes to the Tidzam data and integrates wind noise identifications over a rolling period of thirty seconds. High counts of wind in the period result in a low weight applied to the channel; low counts of wind have no effect. I implemented a similar scheme to promote general wildlife sounds, but did not use it in the later studies. Future work will experiment with more fine-grained, extracognitive weighting of the sources, supported by ongoing work to implement synchronization. This would require advance configuration to inform the system of the user's interests, which would then manifest in the field when Tidzam catches and promotes content in the live audio streams.

As previously described, a parameter weighting and interpolation scheme is used to allow multiple modules to simultaneously influence one channel on each update loop. Notably, the user is in this loop. For example, if the Tidzam AI promotes a particular channel, the likelihood is increased that the user's attention will be caught by the channel content and further promote the channel. The reverse, where user attention would be fed back into Tidzam, is not yet developed, but is part of the vision for future work described in Chapter 5.2.2.

4.4 Field Trials: HearThere at Tidmarsh

To learn more about the experience of extended auditory perception, I assembled a small group of domain expert users to try the end-to-end system in the field. Subjects were invited to participate in tailored activities as well as free exploration in an area of Tidmarsh specially instrumented with sensing and wireless infrastructure. The group was chosen carefully, and

Figure 4-16: Two of the four vantage points on the trail constructed for the HearThere field study. Both images captured during an actual study trial (stills extracted from video documentation by Glorianna Davenport).

each individual was asked to bring their experience to bear in learning to use the device and assessing their experience afterwards. As such, the evaluation described in this section can be seen both as a small-scale study and a panel of expert critics.

4.4.1 Method

Field Site

The field study was conducted at Tidmarsh Farms, later renamed the Mass Audubon Tidmarsh Wildlife Sanctuary. We chose the environmental sensing site known as the Impoundment, a 35-acre marshy area historically used to impound water for farming operations. Details

Figure 4-17: A 0.75 km looping trail through one of the sensor installations at Tidmarsh, designed for the HearThere field experiments. The trailhead is marked with a red circle, and resting vantage points (convenient locations for testing different aspects of the interface) are indicated with yellow squares.

about the Impoundment and its sensing and network infrastructure can be found in Chapter 2.2.

For the purposes of the study, we cleared a 0.75 km trail running along the former Impoundment bank and through an adjacent wooded area. Figure 4.4.1 shows the trail superimposed on a satellite image of the area, with the starting point marked with a red circle. The trail was designed to cover both forested (upland) and wetland areas. Half of its length straddled the boundary between the two ecosystems, aiding in the learning process by allowing subjects to observe the contrast between the different areas as they walked along the path. As the path formed a loop, subjects were free to move in either direction and repeat as they pleased, without concern of getting lost.

Several areas along the path offered convenient vantage points for observing large swaths of marsh or smaller framed views through cutaways in the trees. These were useful for stopping to take stock of the experience or compare real-time audio to the recorded database

Figure 4-18: Participants in the HearThere field study; the documentarian technologist, at right, was the sole participant in an ongoing (12+ week) repeated use pilot study.

(Time Machine feature). Figure 4.4 shows two of these points, which are also marked in 4.4.1.

Participants as Subject-Critics

The choice to invite a relatively small number of participants was made to balance breadth and depth. Adeptness at using the device takes time to develop, and the experience is likely to be extremely novel for any user, potentially skewing their critique. Experts in some aspect of the environment, hearing, or otherwise relevant area might have firmer ground to engage on than completely uninitiated novices. A second challenge was the remoteness of the setting, which required participants to travel long distances and in some cases spend the night.

The participants were selected to reflect a diversity of forms of engagement with the landscape and/or hearing (experienced birder, headphone inventor, ecologist, auditory perception expert, long-term repeated visitor, etc). I chose to vary participants' areas of expertise because I was interested in capturing the ways in which the system informs or alters users' ingrained behaviors either in the environment or with their auditory perception. For example, an expert listener, like a birder, might stand to gain by the added sensitivity HearThere provides, or might struggle with its interference in their habituated practice. For this reason, I consider the participants as subject-critics: experimental subjects with rare expertise, positioned to provide informed critique in an unstructured manner.

Given my focus on the different expertise and backgrounds of the subjects, in reporting and discussing the results in the next section I refer to them by the experience they brought to bear, using the characterizations in italics below. The panel assembled for the study was composed of the following people, shown in Figure 4.4.1:

- *Writer naturalist* Deborah Cramer writes books about ecological science, nature, and the environment, and is a visiting scholar at MIT's Environmental Solutions Initiative. She is a self-described tech novice and her participation was her first visit to the Tidmarsh site. Study period: 2017/09/27, 6:45AM - 8:15AM.

- *Restoration specialist* Alex Hackman is a project manager at the MA Division of Ecological Restoration, responsible for the Tidmarsh restoration project and others. He is a expert on the Tidmarsh site. Study period: 2017/09/28, 4:00PM-7:00PM.

- *Biogeochemist* Dr. Kate Ballantine is an Assistant Professor of Environmental Studies at Mount Holyoke College. She studies ecosystem processes and development on wetland restoration sites, including Tidmarsh, and is a scientific expert on the soil at Tidmarsh, where she has conducted research for years. She is Jason Andras's partner. Study period: 2017/09/30, 10:00AM-1:00PM.

- *Biologist* Dr. Jason Andras is an Assistant Professor of Biological Sciences at Mount Holyoke College. He studies ecological and evolutionary symbiotic interactions between animals and microbes. He is Kate Ballantine's partner. He had visited Tidmarsh a number of times before his participation. Study period: 2017/09/30, 10:00AM-1:00PM.

- *Headphone inventor* Dan Gauger is a Distinguished Engineer at Bose Corporation, and one of the inventors of its noise-cancelling headphone technology. His participation was his first visit to the Tidmarsh site. Study period: 2017/10/06, 5:00PM-7:30PM.

- *Psychoacoustician engineer* Chris Ickler is the Bose Fellow at Bose Corporation, and an expert in audio technologies and psychoacoustics. He is Bridget Hanson's partner. His participation was his first visit to the Tidmarsh site. Study period: 2017/10/11, 7:30AM-11:30AM.

- *Amateur birder* Bridget Hanson is involved in local environmental restoration and protection in Massachusetts. She is Chris Ickler's partner. Her participation was her first visit to the Tidmarsh site. Study period: 2017/10/11, 7:30AM-11:30AM.

- *Documentarian technologist* Glorianna Davenport is a documentary filmmaker, researcher, co-founder of the MIT Media Lab, founder of the Living Observatory at Tidmarsh, and a member of this dissertation's committee. She is the sole participant in a new, multi-session study, in which she took a HearThere prototype home and has used it repeatedly at Tidmarsh. As both a long-term resident of Tidmarsh and the principal driver of its restoration, she is an expert on the site. Davenport was also heavily involved with documenting other participants' experiences prior to and after her own. Study period: 2017/10/01, 7:30AM-11:00AM (with Evan Schulman) and regularly between 2017/12/21 - 2017/04/01 (ongoing).

Many others tried the device at Tidmarsh and elsewhere, some for extended periods. Some of their experiences and quotations appear in the later discussions, attributed to informal testers.

Procedure

A flexible study protocol was devised for the unique and challenging circumstances of the field at Tidmarsh, where issues like adverse weather and real-world technological problems are par for the course. Because of the travel required, some of the participants brought partners who typically wanted to sample the experience as well, adding an additional layer of logistical complexity. In those circumstances, each partner was taken through the steps below together with the main subject, given equal time with the device, and added to the study.

The subject was met at Tidmarsh at an appointed time and introduced to the broader goals of the environmental sensing project (section 2.2). Early mornings and late afternoons were preferred for the higher levels of wildlife activity, but not every subject was able to visit at those times. Next, the subject was brought to the trailhead, given a mobile device loaded with the control software (section 4.3.3), and fitted with the HearThere headphone. Volume levels were raised until the subject could register sound coming from the device. Basic instructions in the use of the control software, such as how to adjust volume and how to select database presets, were given. Finally, the IMU was calibrated with the subject heading due north.

The session consisted of two parts: a closely monitored training portion and a free walk of 20 minutes to 1 hour or longer, interrupted as needed. The training was designed to familiarize the subject with the operation of the device and control software, and just as importantly, to become comfortable with its perceptual effects. Before the free walk portion, subject and guide stopped at a vantage point to explore usage of the time traveling and zooming features (section 4.3.3), and to discuss the experience. Wherever possible, subjects were encouraged to discover the operation of subtle features (e.g. the auditory effect of head stillness, or the touch zooming) for themselves, sometimes prompted into a particular interaction and sometimes unprompted. Different subjects took different lengths of time to report a basic level of comfort operating independently with the device (see section 4.4.2 for reports of the learning process).

Finally, the subject was instructed to walk along the path ahead of the guide, with the guide following at some distance behind in case of problems. The subject was encouraged to explore freely and to use any feature of the control software. There was no strict time limit on the experience and subjects were free to end at any time. However, the majority of the sessions were time-limited to a maximum of two hours by audio disruptions associated with the device battery life; this hardware issue was resolved for the later trials. The free walk was followed by a wrap-up discussion. After the session, the subject was asked to fill out a post-experience questionnaire.

Unstructured Discussion and Data Collection

Throughout the accompanied portion of the experience, the participants were engaged in conversations about the experience they were having. The participants were regularly asked to describe what they were hearing, to put the functionalities of the device into their own words, and to voice their overall experience. These conversations were documented as thoroughly as possible in video and continuous audio recording.

In my initial designs, I believed that quantitative measures from wearable sensors might be useful in assessing the experience, particularly in conjunction with observation by the guide, qualitative self-reporting, and discussion. But the uncontrolled environment of the wetland made on-body sensing noisy and challenging to undertake, and the live experience itself was unpredictable. Separately, I found additional wearable components to interfere with the otherwise seamless experience afforded by the careful design of the HearThere headphone. Ultimately I determined that subjective descriptions of the experience elicited in discussion would be most useful in understanding the potential of the device and clarifying next steps, which would likely include quantitative assessment.

Repeated Use Pilot Study

The documentarian-technologist was given the device to take home and use regularly for a period of twelve weeks to date (and continuing through the time of this writing to what is expected to be several months). With her increasing habituation to the experience, her insights are expected to be different from those of the single session testers. She is personally documenting her experience in the form of journal entries and GPS logs, and has been in regular communication about it throughout.

The longer trial represents the first pilot of what would ultimately be a longitudinal study design (expected to be twelve weeks or more) into the longer term effects of the extended hearing experience offered by HearThere. The pilot study is aimed at validating the system under the much more challenging and more unpredictable conditions of unsupervised, repeated field usage, and uncovering the research questions that would arise under those unique circumstances. What would the experience feel like as the novelty wears off, and would there still be the same level of interest in it? Would subjects want to use the device at certain times but not others, or would it become the default experience of going for a walk? What would it feel like to go for an un-augmented walk? What kinds of insights would subjects glean from the regular use of the time machine features, and would they continue to show the same level of interest in exploring time overlay at the end as at the outset? What kinds of observations about nature would they make that they would not have made otherwise? How would they talk about their natural hearing abilities as compared to the their augmented experience?

4.4.2 Results

This section reports on the results from the HearThere expert panel study and field trial at Tidmarsh, assembled from a combination of observations of the participants, discussions during the sessions, and responses to the post-experience questionnaires. In-situ documentation was made using notes, audio recording (Zoom H6 portable recorder), and intermittent video recording (by Glorianna Davenport). Where there were hard documentation gaps, notated quotations were checked with the participants as part of the questionnaire process.

"I'm just overpowered thinking about why Mass Audubon would be interested in this. This is a tool for exploring a landscape. This space, what's interesting about this particular spot in the space, and the things that are happening there. Drawing your attention to different things, different aspects of the space. This is a great learning tool, for nature."

"I go to a superpower analogy. A tool is external to you—one manipulates it to achieve a goal. To explore the marsh from a distance I think of binoculars or a parabolic reflector mic. A superpower becomes part of you. The way you've integrated with head motion and the use of non-occluding audio playback makes it that. The more you can give people control of the experience with body motion, staying away from a screen as an interface, the more powerful this will become."

Headphone Inventor

"For me, the ability to go back in time, and hear sounds from a different season, was the best part. To stand there on a fall day, and hear spring or summer sounds from the same spot, was transporting. This is the feature I would probably use the most."

Restoration Specialist

"Makes you pause, and hear, really hear"

Documentarian Technologist

"The ambient, unzoomed experience was a pleasant enhancement of being outdoors in nature. I imagine walking through the natural landscape with a slightly more intense feeling of enjoyment of the overall experience, one that I might not directly attribute to sound."

I would say overall, my experience is that it intensifies the auditory experience of being out here. But I'm not finding it easy to get sensor fusion. I'm not finding it easy to key in on anything specific and locate it with eyes and ears, for example.

Psychoacoustician Engineer

Figure 4-19: Selected high-level comments about the HearThere experience, given during or shortly after participation.

The initial results stem from observations I made about each user during their augmented hearing experience combined with excerpts of transcribed remarks made during the experience and post-experience in situ discussion. Given the spread of backgrounds and orientations towards technology in particular, there was some variation in the feelings and language around the experience, but also a surprising amount of commonality. Particularly striking were the common models and metaphors that arose independently in different subjects' language as well as the common behaviors that emerged, both sanctioned (introduced during the learning orientation) and unsanctioned or emergent (invented by the subjects to resolve perceptual ambiguities, for example).

To parse the large volume of responses, I sorted the material into high-level categories that get at different aspects of the subjects' experiences. Representative samples are copied in the next sections, along with an elaboration of each category and subcategory. To begin, I separated out the highest level comments, which tended towards pronouncements about the experience overall, as well as the technology and its uses.

Figure 4.4.2 shows a small selection of those high-level comments. In different moments, subjects focused on very different aspects of their experience in assessing HearThere's impact:

from one subject, the Headphone Inventor, the device was described at one point as both a "learning tool for nature," and at another as a "superpower."

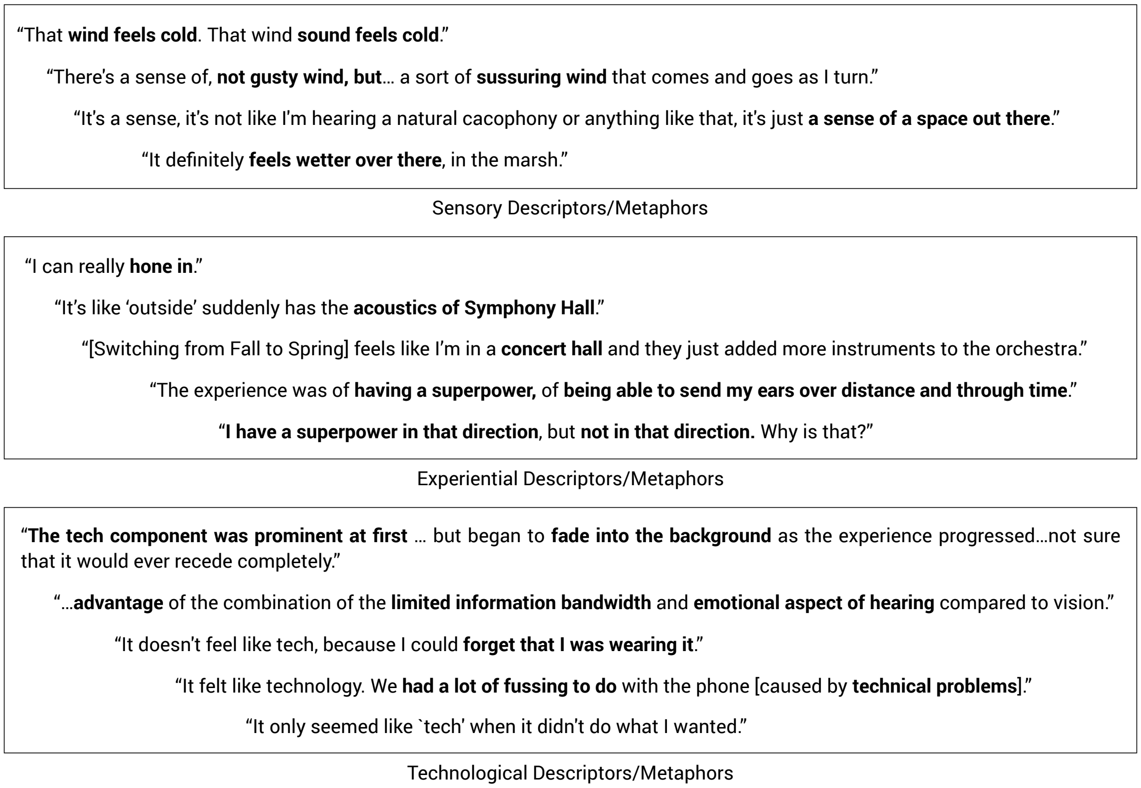
One of the critical remarks included here refers to the subject's challenge in achieving "sensor fusion" between their visual and auditory modalities. Standing in the field, the Psychoacoustician Engineer and I had an extended debate about this issue, focusing on what kinds of interactions would remain in the non-tech realm but allow for the deep focus into transient events he wished for. Two options were raised: first some kind of subtle visual AR indicator, which we both agreed would be difficult to produce using existing technologies without major interruption of the more flowing auditory experience. The second concept he proposed was for an instant replay function that would allow him to lock in a loop of the prior minute while searching for the multimodal he sought. This issue is elaborated further in the discussion in the next chapter.

Sensory Descriptors

Prompted to describe what they were experiencing, or often unprompted, participants frequently used sensory words and textural or even physical, gestural metaphors to describe the objects of their digital perception. Figure 4.4.2 excerpts samples of these, separated into three categories detailed below.

The notion of sensory descriptors has been explored in psychology and more recently by neuroscientists, who have established links between textural metaphors (e.g. 'a rough day') and their origins in perceptual experience (e.g. 'a rough surface') [75]. I contend that these descriptors are indications of a successful perceptual encoding extending an existing hearing ability and therefore picking up sensory- and hearing-related descriptive characteristics (in contrast with a more symbolic choice of language e.g. "it is windy over there"). A related, but different kind of sensory descriptor concerned one kind of perceptual confusion subjects commonly experienced at the beginning of the study, when the non-occluding display resulted in ambiguities about which source (ears vs. HearThere) was responsible the perceived soundscape: "[the sounds] blended so well that I was sometimes confused" (Psychoacoustician Engineer).

There were also frequent gestural indication of feeling, captured in a number of instances on video. To illustrate this visually, Figure 4.4.2 shows one of those moments using a series of video frames, in which the headphone inventor refers to an ebbing and flowing wind using



"That **wind feels cold**. That wind **sound feels cold**."

"There's a sense of, **not gusty wind, but**… a sort of **sussuring wind** that comes and goes as I turn."

"It's a sense, it's not like I'm hearing a natural cacophony or anything like that, it's just **a sense of a space out there**."

"It definitely **feels wetter over there**, in the marsh."

Sensory Descriptors/Metaphors

"I can really **hone in**."

"It's like 'outside' suddenly has the **acoustics of Symphony Hall**."

"[Switching from Fall to Spring] feels like I'm in a **concert hall** and they just added more instruments to the orchestra."

"The experience was of **having a superpower,** of **being able to send my ears over distance and through time**."

"**I have a superpower in that direction**, but **not in that direction.** Why is that?"

Experiential Descriptors/Metaphors

"**The tech component was prominent at first** … but began to **fade into the background** as the experience progressed…not sure that it would ever recede completely."

"…**advantage** of the combination of the **limited information bandwidth** and **emotional aspect of hearing** compared to vision."

"It doesn't feel like tech, because I could **forget that I was wearing it**."

"It felt like technology. We **had a lot of fussing to do** with the phone [caused by **technical problems**]."

"It only seemed like `tech' when it didn't do what I wanted."

Technological Descriptors/Metaphors

Figure 4-20: Selected sensory, experiential, and technology-related descriptors and metaphors used by subjects during the study (emphases mine).

an undulating hand gesture before pointing in the direction of the source.

Experience Descriptors and Contextual Metaphors

One level of abstraction above the sensory descriptors, I found commonality in the experiential descriptors subjects used. Some of these are excerpted in Figure 4.4.2. A number of common metaphors were brought up that resonate strongly with experiential qualities of other kinds of acoustic spaces, such as concert halls. The concert hall metaphor was particularly striking, as it arose independently with different subjects. It captures not only their surprise and appreciation of the high fidelity of the experience, but also their reimagining of the landscape as an acoustic stage that is consistent over time despite dramatic scenic changes. This is an effect I refer to in the upcoming discussion as *continuity*.

The experiential descriptors also related frequently to memory and emotion, which a number of participants referred to during and after the experience. Commenting after the experience, the Psychoacoustician Engineer wrote: "my memories of the marsh have more

Figure 4-21: The headphone inventor uses hand gestures to describes a feeling of undulating wind in the distance: "It's a wind that's coming and going, over there."

sound in them, a very spacious open, alive sound. A large space filled with sound."

Experience as Technology

Among the high-level responses from the participants, a common theme regarded surprise at the lack of technological dimensions of their felt experience, or appreciation for the extent of concealment of those dimensions. These responses, excerpted in Figure 4.4.2, came both unprompted and prompted by later questioning. Respondents whom expressed appreciation for the backgrounding of technology attributed this effect to the comfort of the device, its non-obstruction of their natural hearing, and the subtlety of the listening interactions. Those that expressed otherwise tended to have had experiences marred by technical failures causing interruptions or otherwise compromising the continuity of the study. Still others reported the "tech component" receding over time, as they became more comfortable operating the device.

The Psychoacoustician Engineer in particular made frequent reference to his surprise at HearThere's focus on perceptual experience over the underlying supporting technology, notable for his background in the area:

> "One of the subtle but important strengths of [HearThere] is that it's not obviously technology. It's more obviously experience. You're changing the experience that I have without having the technology itself be particularly noticeable or obtrusive... This is rare: you actually are using technology in a way that isn't obvious, the technical part isn't obvious. It's the nature part of it that's obvious. I think there's something important to try to preserve in that. I think you made the right choice not to let the technological, data-centric portion of it intrude into consciousness."

Others, such as the Author Naturalist, spoke of this effect—of HearThere's support of their focus on the environment above itself—in less technical, more experiential terms: "the walk did not feel like a tech experience in that many of my 'tech' experiences focus more on the technology... It's a great idea to use technology to bring someone closer to the outdoors rather than further away...expansive and illuminating." For subjects who experienced frequent technical problems and system failures, the effect was the opposite, as the Restoration Specialist noted: "it felt like technology because we had a lot of fussing to do with the phone [caused by technical problems]. If working, it is easy to imagine it would not feel like a tech-heavy experience." The idea captured here, that *technology feels like fussing*, is at the heart of this work, and shows how fragile perceptual transpresence experience can be.

Learning

With the exception of the longer term pilot subject (Documentarian Technologist), all participants went through a single session averaging two hours in length, resulting in a learning process that blurred together with the more independent part of the experience. Responses associated with the perceptual learning process in this period can be divided roughly into categories of sensory integration and perceptual intuition, the latter following the former. These categories represent a large body of comments, but a small number of excerpts

I had a very clear adjustment period. For the first hour, it was **hard to distinguish** background natural sounds from device augmented sounds. Then, **it all just clicked** and **I could tell the difference**, yet **use both sounds streams together.**

I didn't have enough time. I feel like, since this is **modifying a fundamental sense**, that it takes **mental time** to integrate the capability so as to master it and benefit from it.

It's a superpower to be **mastered**. But **it's not just the device**, it's the sensor-laden environment.

I did not know where the microphones were...When I heard nothing special in consequence, I did not know whether this was because there were no sounds there or no mics there. This was **frustrating**.

"Learning to coordinate movement and position on the phone map with the audio received from the headset took **10-20 minutes**, but after that felt much more intuitive."

"I **had enough time** to learn."

"Why is the sound moving around me?

Figure 4-22: Selected comments from subjects regarding perceptual confusion and the experience of learning to hear through HearThere (emphases mine).

are given in Figure 4.4.2. In the sensory integration category, participants made reference to sounds "moving around them" before ultimately registering that the spatial rendering was actually fixing the sounds in space (that it was self-motion causing the perceived rotation). The rapid spatial integration I observed was a welcome confirmation that head movement cues dominate spectral cues [12], overcoming the challenges of bone conduction for spatial sound presentation. Still, seeing that integration in action was notable, particularly for the Restoration Specialist, whose spatial understanding appeared to click in an instant, after nearly an hour of use. Others integrated the spatial presentation immediately, pointing out the locations of features of the environment. This occurred even when subjects had never been exposed to AR spatial presentation and did not readily understand it. The Author Naturalist, for example, insisted that the presented sounds did not have locations while simultaneously pointing out the directions from which the sounds originated.

A more interesting and less understood aspect of the learning process concerns the perceptual ambiguities between natural hearing and bone-conduction that occurred after spatial integration. Nearly all subjects expressed confusion about which source came from HearThere before slowly becoming more comfortable with the ambiguity and/or reporting to have perceptually resolved it. I later came to refer to this ambiguity as the "crow question," for the significant number of times that subjects specifically expressed that they "hear crows. Do you hear them?". The eventual resolution of the bone conduction ambiguity represents a new and striking result that calls for further study, and is discussed at length in the next chapter.

After spatial and perceptual integration, expressions of new perceptual sensibilities began to emerge, taking the form of subjects' knowledge-based approaches to disambiguation. These included observations about the timing of bird migration patterns, or an expectation of what the marsh would sound like compared to the forest. I noticed a feedback loop in which intuitions were applied to resolve perceptual ambiguities, and then later, once resolved, to generate new intuitions in the environment, such as to the response of wildlife from a distance when a subject approached.

Other kinds of confusion represent shortcomings of the current system, such as the warping of space caused by its representation of microphones as spatialized point sources rather than a reconstructed wavefront: "I did not know where the microphones were... When I heard nothing special in consequence, I did not know whether this was because there were no sounds there or no mics there. This was frustrating." After the experience, the headphone inventor articulated a very specific learning sequence he would want to take. His post-study written comments about pausing to let his mind integrate the experience are of particular interest here, and have bearing on the design of follow-up study:

> "My gut says it's a 3-step process: (1) Get exposed to it as I was that Friday but with little hands-on control; just get a feeling for what it can do. In hindsight I'd say I felt a bit overwhelmed toward the end of my visit; I needed to let my mind integrate. (2) In a 2nd visit I want hands-on the app, to explore the intersection of control and experience. Then I think I need some more integration time. (3) In a 3rd visit I think I could then use it, to really explore the Tidmarsh augmented by it. This may be my own learning process—I don't know how well it generalizes."

Behaviors & Interaction

Learned interaction and exploring behaviors with HearThere were partly taught and partly emergent. There were clear variations between subjects, as well as clear commonalities. The two subjects with engineering backgrounds (Psychoacoustician Engineer, Headphone Inventor) both made early and significant use of the more explicit technological features, such as auditory zooming, and were more systematic, at least initially, about probing the system and its auditory augmented reality rendering than other participants. The psychoacoustician

Figure 4-23: As she learned to use HearThere, the Amateur Birder repeatedly pulled the transducers away from her head in an effort to resolve the ambiguity between the bone conducting stimulus and her natural hearing. After her first determination that a source was indeed virtual, shown here, she asked: "Suppose I hear a little faint sound, and I wanted to hear a little more of it. What would I do?"

engineer in particular exhibited very slow, deliberate movements, rotating his head and body back and forth methodically and occasionally doubling back on the path.

All but the Author Naturalist, notably, made independent use of the touch zoom. Several of the participants seemed to grasp the interaction quickly, most of all the headphone inventor. They started to make what looked like automatic use of it after only several conscious attempts. I prompted exploring behaviors by asking subjects to describe what they were hearing, not unlike the drawing task from the lab study. Subjects tended to respond with rich descriptions that capture aspects of both the natural environment and the HearThere-specific experience, as the Biogeochemist illustrates:

> "As I move my head, I go from hearing the tree birds, and now I'm hearing more ducks and geese and froggy frogs [sic], and I put on my supersonic beam action [touches the headphone transducer], and now that's really all I'm getting, and now all I'm getting is the forest because I moved my head this way [towards the forest]. There must be [a microphone] right there... When I do this [touches

> the headphone transducer], there must be a sensor right over there, because I'm getting the sound from over there. I can really hone in."

In an attempt to resolve perceptual ambiguities, most of the participants developed an unsanctioned, emergent behavior of pulling the headphone transducers away from their heads to determine whether the sounds they were hearing were coming from the HearThere. Figure 4.4.2 shows the amateur birder discovering this strategy. Some subjects made use of the auditory zoom feature in a learning capacity as well, exemplified by comments from the Amateur Birder: "I think I would probe by zooming in on far-off microphones to see if there was sound too small for me to hear without amplification. It would be like scanning an area with binoculars instead of focussing on something I already had detected." The binoculars metaphor was anticipated, and common, as subjects settled on how and when to use the feature. The Restoration Specialist "felt like my hearing was zooming into something well outside of my normal range of hearing. It was akin to using binoculars and seeing something you normally wouldn't be able to." Similarly, the Biogeochemist commented that "the binocular effect seems fairly intuitive."

4.4.3 Summary Discussion

When the technology didn't fail for reasons of adverse weather or, more often, digital noise in the audio due to power management issues in early designs, the tests were largely successful. Failures never stopped a session entirely, but caused unwelcome interruptions that interfered with the learning process. These included the onset of digital noise as the battery voltage dropped, remedied by charging in the field, as well as orientation drift caused by magnetometer proximity to the transducers. The second half of the subject pool was given updated hardware that completely resolved these issues.

Even so, subjects across the board consistently reported their experiences in extraordinary terms, representative of new "sensory superpowers" and at the same time as fundamental, intuitive ways of exploring and learning on a landscape. Combined with the non-occluding, fully blended presentation of the hardware, the time machine feature of Sensorium introduced an entirely new, temporal dimension for new perceptual sensibilities to take root in; after as little as an hour of use, subjects began describing those new sensibilities about landscape over time. Similarly, HearThere dramatically extended subjects' perceptual space, prompting new sensibilities for acoustic interactions with distant wildlife. Finally, subjects universally reported experiences of extended presence—what some described as ‘immersion’.

A number of high-level takeaways can be distilled from my observations as well as from subjects’ comments during and after the field evaluation. The next chapter elaborates on the field study in an extended discussion connecting it to the larger message of the thesis. As such, only a brief summary discussion is included here.

First, though often neglected in research prototypes, both a high degree of physical comfort as well as high-quality data are a prerequisite for the extended perception to take root. When it worked, users in the field study were able to momentarily forget that there were wearing HearThere. This effect comes not only from the physical comfort of its design but also from fully non-occluding display. Low noise and high dynamic range data are always better in sensing, but in this context they allowed for the mixing of more sources without building an overwhelming ambience. The Networked Sensory Landscape of Tidmarsh made for an unusually dense and high-quality dataset, leading to interest and committed use from all the participants, as well as repeated use over months from the longer term pilot subject. The quality of the audio data in particular was also in evidence in the lab study, where subjects across the board reported that they most enjoyed paying attention to the Tidmarsh audio scenes.

Second, interactions based in sensorimotor contingencies, and experiences that maintain spatial and temporal continuity allow for rapid learning of exploratory gestures and heighten the sense that there is more out there to be attuned to. With both phenomenological and social bases, repeatable actions like cupping an ear or furrowing a brow quickly become second nature, as I observed with the immediate adoption of HearThere’s touch zoom on the part of all but one of the subjects. On the sensing side again, subjects recognized consistent acoustic properties of Tidmarsh that allowed them to directly perceive a single location across dramatic seasonal and annual changes.

Finally, designing for undirected exploration lets users cultivate sensibilities, own their discoveries, and develop individual paths to deepening their understanding. In the field study, confusion paired with interest resulted in the application of intuition, which led to a model, and in turn produced new perceptual sensibilities in time and space. Relinquishing control gave subjects agency in their exploration, which produced anticipation of discovery. With available interactions spanning explicit (phone-based), learned, and attention-based, subjects were able to strike their own balance.

Chapter 5

Towards a Networked Sensorium

In Chapters 3 and 4, I presented top-down and bottom-up approaches to constructing transpresence experiences out of distributed media, in particular sensor data. Inspired by calm technologies [130] and ambient interfaces [59], the top-down approach did so in both physical and virtual settings by embedding or suffusing data into sensory object properties and surfaces, as well as by subtly altering environmental mechanics; the process of separating out the data then relies on some combination of deliberate attention, contact with the object itself, and/or ambient perceptiveness of its embeddings. In the second approach, which built on principles of sensory assistive technology [91, 2, 74] and augmented reality [124], transpresence perceptions arose instead from the body outwards, blending networked sensor data with the physical world through a combination of lightweight non-occluding auditory display and realistic spatial rendering. In studies of the latter approach, I saw new perceptual sensibilities emerging from users' initial confusion at the blended experience. What began as conscious reasoning applied to assigning, at each instant, one set of stimuli to a physical source and another to the display technology, became second nature to users for reasons as yet undetermined (and theorized in the next sections). If the sensorium is understood as the sum of one's perception, then where percepts begin to traverse networks, I see the beginnings of a *networked sensorium*, in which digitally-produced transpresence adds new occasions for perception to existing sensory modalities[1].

---

[1] Uexküll's theory of the *umwelt*, or self-centered world, is commonly used to denote any organism's ensemble of sensory modalities which circumscribe its experience of the universe [127]. I prefer the term *sensorium* here because my focus is on presence, attention, and perceptual sensibilities in a Gibsonian sense, all of which can grow simply through practice. Although it may seem like only a framing, this is an important distinction, shifting the research aims from 'augmenting humans with new senses' to enabling humans to be perceptive of new things (and/or old things in new ways).

The next section distills the projects and studies from the last chapters into a set of guidelines and research questions for this new domain, drawing on the HearThere and transpresence work and opening new lines of inquiry. Section 5.2.1 details several future studies that would begin to investigate these questions, including one longitudinal study that is already in a pilot stage. Section 5.2.2 outlines my medium and long-term research visions for the sensor(y) landscape, which center on the dynamic it opens up between human perception and artificial intelligence. Finally, Section 5.3 concludes the dissertation.

5.1 Research Discussion: Occasions for Perception

Time spent in natural settings is well established to be beneficial to human psychological health. Kaplan's seminal treatment of its "restorative benefits" to attention offer a concrete framework for modeling experiences of nature [66]. This thesis introduces technologies designed to be experienced in nature while facilitating perceptual attention to sensor data. As such, Kaplan's framework is useful for analyzing the special case of HearThere at Tidmarsh, as well for grounding a broader discussion of transpresence experience. Kaplan identifies four major components of restorative environments, each of which, by extension, has bearing on the design of transpresence technologies: *being away*, *fascination*, *extent*, and *compatibility.*

I point the reader to [66] for an extended elaboration of these components. Briefly, *being away* describes a conceptual shift that could be brought about by "a change in the direction of one's gaze, or even an old environment viewed in a new way." *Fascination* can come from experiences of certain perceptual or cognitive processes (e.g. "predicting despite uncertainty") or arise directly from content (e.g. "wild animals and caves," or natural settings in general, which, as a result of so-called "soft fascination", provide "opportunit[ies] for reflection"). *Extent* describes a combination of richness and coherence that "constitutes a whole other world...of sufficient scope to engage the mind." Finally, *compatibility* is akin to a flow [20] between the environment and the explorer—a "responsive environment" with "prompt and useful feedback" not requiring of substantial conscious interaction. In the next part, I link these components to my design and engineering choices, and I return to them again later in the discussion.

In the HearThere field study, no specific tasks were given to participants besides an instruction to explore, and no specific material was prepared for subjects besides a set of

Design Constraint	Experiential Component
Physical comfort	Being away
Non-occlusion	Compatibility, extent, being away
Ease of interaction	Being away; compatibility
Content	Extent; fascination
Medium	Compatibility

Table 5.1: Matching design and engineering constraints of transpresence technologies to components of restorative experience from [66]

suggested timelines to explore. A significant proportion of the subjects' experiences were derived from uncontrolled real-time data streams. Those experiences, as a result, were all different, but, like any walk in the forest, characterized by uneventful stretches punctuated by exciting moments. Presented with a string of these unique, ephemeral, and often confusing *occasions for perception*, subjects slid into the role of explorers, in those moments the only people in the world with the ability to perceive through HearThere. As a result, they were driven to cultivate their sensibilities, own their discoveries, and develop individual paths to deepening their understanding of both the sensory landscape of Tidmarsh and the contours of their individual perceptions. But what are these occasions and how do we conjure them with technology? Through my projects, I identified a number of fundamental design and engineering constraints and criteria that, taken together, construct the transpresence experience.

5.1.1 Design Constraints

This subsection catalogs the design and engineering constraints that arose through my work and were met holistically in the HearThere project, producing the effects I observed in the Tidmarsh field study. Going a step further, Table 5.1 links each of the constraints to one or more of the high-level components of Kaplan's attention-restoration framework.

Physical Comfort

Physical comfort is always considered 'nice-to-have' but is uncommon in wearable research prototypes. I address comfort here in specific relation to the problem of acquiring new perceptual skill, which requires, on the one hand, conscious interest and driving curiosity on the part of the learner, and on the other an ability to lose sight of the task and be in the moment. To achieve the glasses-like experience of transparency described above, the user must be able to forget that they are wearing the device.

Difficult design tradeoffs here include balancing user-sensing features against the physical discomfort they tend to cause. In the case of HearThere, I found eye tracking to have real potential for improving the performance of the user attention model in the lab setting, but at too great a cost to mobility, sound quality, and comfort during extended use. The subtleties of faint natural sound were lost in the constant adjustment of the device. Similarly, while an EEG proved more comfortable than the eye tracker in my tests, its benefit over the much simpler stillness metric was negligible in real-world conditions. Ultimately, the comfort constraint forced me to find alternate, and generally simpler approaches to meeting my interaction design specifications.

We see the effects of comfort in perceptual interfaces in the literature, where failures to achieve sensory integration are correlated with complaints about discomfort and an inability to ignore (tune out) the stimulus [92]. I observed the same effect inversely throughout the HearThere subject pool, where headphone-averse and/or technology-averse participants reported their experience as being unlike either and were able to wear the device comfortably for hours at a time. This is not simply a matter of making the user experience more enjoyable; physical discomfort continually tethers the experience to the intervening technology, compromising Kaplan's dimension of *being away*. However, although I argue here for a proper treatment of physical comfort in the design of similar experimental systems, elaboration of general design principles is beyond the scope of this thesis; for that, I would encourage readers to apply a vast ergonomics and human factors literature to their specific designs, and to test many variants, as I did.

Transparency

True non-occluding display allows users to focus on the world without noticing the mediating layer of technology. This has long been a key goal for designers of AR systems. I emphasize *true* non-occlusion here because many consumer technologies are partially occluding. This requirement dramatically limits the available options for display, as there are currently few viable approaches to mobile, non-occluding audio besides bone conduction. Glasses-mounted, ear-directed speakers such as those on Microsoft's HoloLens offer a potentially viable approach, but in that particular case are not separable from the bulky and visually occluding visual platform. They can also be quite difficult to hear in loud environments.

A just-announced (but as yet unavailable) glasses-mounted, non-occluding, head-tracking audio platform from Bose is a promising candidate, and the product announcement cites a relationship to this work [17]. Other options on the market, such as collar-worn directional speakers (e.g. Bose's SoundWear product) suffer significant losses of fidelity and associated loss of effective spatialization outside of a centered (fixed) head orientation. Non-occluding audio is an active area of research, however, and lightweight alternatives to bone conduction will continue to emerge in the years following this thesis.

Visual augmented reality displays are further behind in this regard, with modality-specific challenges such as limited fields of view, as well as heavy compute, sensing, and power requirements. For auditory AR display, effective spatialization is produced through the use of head and location tracking, and an appropriately selected head-related transfer function (HRTF). For visual AR display, however, continuous high-resolution 3d mapping is required to merge information into the visual field. For the long distances involved in natural settings, there is a significantly greater tolerance for errors with spatial audio, and users quickly adapt to non-optimal HRTFs in situations of free movement and head tracking.

The transparency constraint underpins Kaplan's dimensions of *compatibility*, *extent*, and *being away*, as it reduces distraction, allows digital representations to merge into the physical world, and affords a transformation of perspective.

Interaction

Following naturally from the transparency constraint, designers must maintain conscious user interaction at an absolute minimum, even if doing so entails compromising useful system features. At the all-important beginning stages of the perceptual learning process, a low barrier to immediate engagement with the source material increases the surface area of the material relative to that of the mediating technology. With little in the user interface that would draw the user's attention away from their surroundings, learning effort is maximally focused on the content itself. Bootstrapping is essential: with both phenomenological and social bases, repeatable actions like cupping an ear or furrowing a brow quickly become second nature indicators of focused attention, for example.

There is no interface learning required to use HearThere; users simply put on the headset and go. From a user perspective, the device adds sound to their existing auditory perception without requiring input. In actuality, interaction in the active mode is a matter of moving and looking around, stopping occasionally, and, depending on the user's interest, touching the transducers to zoom into sound sources. When the context is appropriate and where possible, user sensors such as EEG or eye trackers operate in the background. To that end, I found gaze correlates to listening state in the lab setting (Section 4.3.1) and integrated eye tracking into one of the HearThere platforms (Section 4.3.3), but ultimately found the eye tracking both uncomfortable and not robust to outdoor use.

Still, in the field, I observed users reaching unconsciously for the touch zoom when something faint in the environment caught their interest; conversely, frustrations abounded when users failed to catch the source of interest in this way, and entered into a more conscious interaction as a result (a failure in the system due to limitations of the microphone network and naive spatial rendering described in Section 4.3.3). Other interactive features, such as the attentional effects of user stillness, were not perceptible unless explicitly pointed out. Once they were made aware of the effect, users increasingly stopped to listen as a conscious act, an interaction that could be seen either as an act of perception or an explicit instruction to the technology. Subjects in the field study consistently expressed surprise at the ease of use and negligible requirements compared to their prior impressions of technology. At all times, the content was a driver to learn new features, not the features themselves.

The most glaring exception to the unconscious interaction constraint was the Sensorium GUI, which allowed users to configure settings and select times in the database to explore. The compromise in HearThere's design was to keep all complex interactions to the touchscreen interface, and thus prevent them from slipping into the experience of exploration. A related interaction design tradeoff that came out in the field study concerned one subject's request for an "instant replay" feature that would allow him to trigger the soundscape to loop. I later implemented the feature as requested, and although it remains to be tested, I believe it would materially interfere with exploration by encouraging conscious interaction in general usage. Unfortunately, either option poses a problem, as the subject's desire for access to the transient sources reflects a failure of the system to be *compatible*: in Kaplan's terms, the environment did not provide the information needed to meet the subject's purposes. It remains to be seen whether additional practice might obviate the need for that feature, and otherwise highlights the complexity of balancing desirable features against fluid interaction in designing transpresence technologies.

When it works, however, unconscious interaction through a perceptual interface can lead to a state of *being away.* In the HearThere example, subjects' interests in an environmental signal led them into stillness, which in turn led to a greater focus on the source of the signal. This experience of a perspective shift was brought up repeatedly in the study feedback.

Content: More Than Meets the Eye

On the opposite end of the system design, the choice and quality of content are critically important properties of the transpresence experience, central to the success of HearThere and earlier projects. Most relevant to this dissertation were practical issues of noise floor and dynamic range on the one hand, and selection of data sources on the other. Of course, better sensing is always better, but particularly where large numbers of streams are added together to produce an effect, low noise allows for the mixing of greater numbers of sources without building an overwhelming background ambiance. Similarly, a high dynamic range enables the capture of faint environmental sounds, for example, without boosting gains (and increasing noise). The HearThere experience mixed dozens of audio sources without any noticeable hiss, and users were able to pick up distant sources easily. Significantly, no one reported the sound to be bothersome or distracting, despite its volume and the density of sources. I attribute this result largely to the low noise transduction.

Partly as a matter of personal taste, the selection of data sources is a much more complex problem to formalize. For new users of sensory assistive technologies, new perceptions [4] would generally arise through persistent practice rooted in necessity. Most sensory substitution devices are first and foremost tools for avoiding danger and operating independently in a poorly-designed, disability-unfriendly world. In HCI and sensor networks, where necessity is not the primary driver, interest in the source material will keep users in sensorimotor exploration longer. Interest begins with *richness* at the source. Richness in data is a function of a number of properties, among them non-linearity, a diversity of states, and a subjective sense of illimitable depth to the source of the data—"more than meets the eye." In Kaplan's model, richness is a part of the restorative components of *fascination* and *extent.*

Certain kinds of environmental data take on these qualities by offering limited but succinct views into complex ecological interrelationships: a tree's transpiration cycle, for example, alongside the time of day and soil moisture at its roots, offers a glimpse into a system with far more complex dynamics than a passing glance would imply. The *partial view of complexity* creates a new mysterious perceptual boundary for attention to linger on and cross.

Other forms of data have these qualities as well: by matching the audio content and cultural context, for example, the ListenTree installation in Mexico City (Section 3.2.3) gave audiences a sense of intrinsic depth that was not present in other ListenTree installations of the same technology and content. In that example, the Day of the Dead context provided a rich ground for audience interest to take root in and grow from, much more so than the content would alone.

Sometimes this sense of depth is simply window dressing. Numerous sensor-based art installations either intentionally or unintentionally operate with questionable or obviously nonsensical data using the premise of a connection with the environment to conjure richness. The resulting harm is that the richness would merely feed interest, but without structure in the data, perceptual sensibilities would not have a basis to form. This may be acceptable for environments we only visit on occasion, but certainly not for extended use. There is simply no learning to be done in noise.

Information in the Medium

Why use sound? Putting aside the current limitations of visual AR display, is there any sense to be made of the choice of presentation medium in perceptual interfaces? Continuing on the theme of compatibility, content is best served through a medium aligned with its physical, historical, and cultural resonances. Sound is continuous and omnidirectional. Sound sources add together without obstructing one another. Where vision can be occluded, sound is absorbed as vibration—an interior quality intrinsically linked to object materiality.

In [51], Stefan Helmreich writes about physicists' now iconic sonifications of gravitational wave detections. Why use sound to encode phenomena in the vacuum of space, a realm in which sound has no medium for propagation? Helmreich traces a historical precedent in sound as both a perceptual model and an icon of cosmological phenomena, from cosmic harmony to the Big Bang. He highlights its role in shaping the modern understanding of gravitational waves, which have taken on qualities of sound in both their popular conception and in the language of physicists; Helmreich quotes LIGO scientist Szabolcs Marka, writing in the New York Times: "everything else in astronomy is like the eye... Finally, astronomy grew ears. We never had ears before." This expression of a link between human-perceivable vibration and the utterly incomprehensible propagation of spacetime warping is repeated time and again by cosmologists, who as a result have developed remarkable auditory sensibilities for the relationship between the brief sonic chirps of LIGO and the cataclysmic collisions of black holes. The pre-existing connection between space and sound offers physicists and the public a phenomenological ledge to stand on, and continually reinforces itself.

Transpresence Design and Characterization

I developed an initial design and characterization space for transpresence interfaces based on the discussion above. Table 5.2 summarizes the space. Shaded regions in each dimension show where HearThere fits in. This is not an exhaustive listing of the design dimensions, but the beginnings of a more structured articulation of the available design choices to guide future work.

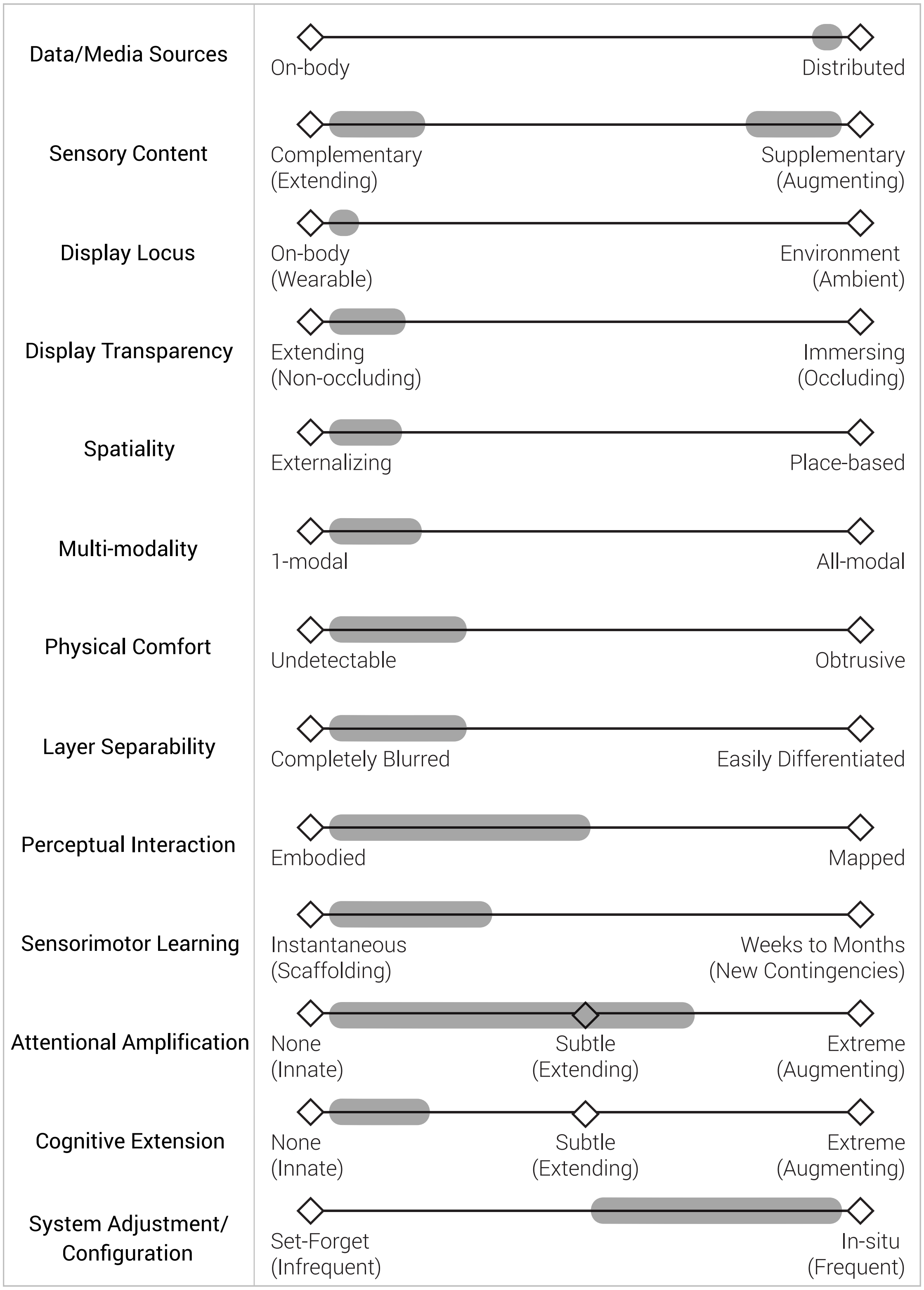
Data/Media Sources – Do the live or recorded data streams come from sensors/sources on the body or sensors/sources distributed in the user's surroundings? Traditional approaches to sensory augmentation and assistive technologies have assumed that the distal surface is contiguous with the body. Distributing the sources requires first-person perspectives to be generated for each user.

Sensory Content – Are the sources of data extensions of the user's existing sensory modalities (e.g. increasing sensitivity or range) or do they provide access to entirely new modalities (e.g. perceptual mappings of phenomena not otherwise senseable by humans)? HearThere mixes sound from microphones (extending) with spatial sonification of environmental sensor data (augmenting).

Display Locus – Is the display technology on the body (wearable) or embedded in the environment (ambient)? HearThere uses bone conduction on the body. An ambient counterexample from this thesis might be a ListenTree device at the edge of a pond that plays sound from under the water when a visitor leans against it.

Display Transparency – Does the see- or hear-through display technology in any way occlude a user's vision or hearing? Note that this dimension is related to but different from the mixed reality continuum. A selective hear-through technology that occludes the ears but allows audio from an external microphone through would fall on the right side of the spectrum, even though it might be experienced like augmented reality. A non-occluding device causes no change in the user's perception when worn but inactive.

Spatiality – Sources presented through an externalizing display are perceived by the user as emanating from the environment, having world positions and directions, and interacting



Data/Media Sources	On-body		Distributed
Sensory Content	Complementary (Extending)		Supplementary (Augmenting)
Display Locus	On-body (Wearable)		Environment (Ambient)
Display Transparency	Extending (Non-occluding)		Immersing (Occluding)
Spatiality	Externalizing		Place-based
Multi-modality	1-modal		All-modal
Physical Comfort	Undetectable		Obtrusive
Layer Separability	Completely Blurred		Easily Differentiated
Perceptual Interaction	Embodied		Mapped
Sensorimotor Learning	Instantaneous (Scaffolding)		Weeks to Months (New Contingencies)
Attentional Amplification	None (Innate)	Subtle (Extending)	Extreme (Augmenting)
Cognitive Extension	None (Innate)	Subtle (Extending)	Extreme (Augmenting)
System Adjustment/ Configuration	Set-Forget (Infrequent)		In-situ (Frequent)

Table 5.2: Design space of the transpresence approach to sensory extension/augmentation. Gray regions show where the HearThere prototype fits into the design space.

with the surrounding environment through reflection and other physical means. HearThere is a spatializing display, using an HRTF, physical reverb modeling, location and head-tracking to produce a sense of externalization. Place-based displays change in response to the user's location, but do not perform virtual spatialization.

Multi-modality – 1-modal displays are presented through a single sensory modality. HearThere is effectively 1-modal (auditory) but the tactility of its bone conduction presentation was described by some users as a partial secondary modality.

Physical Comfort – An undetectable interface is so comfortable as to be not noticeable to the wearer during use. Obtrusive interfaces are always noticeable to the wearer. Various HearThere platforms fell across this spectrum, depending on the user-sensing sensing platform used (eye tracker, EEG, etc).

Layer Separability – A completely blurred display presents sources that users cannot distinguish from their perceptual "reality," using some combination of non-occluding display transparency, externalizing spatiality, and plausibly complementary sensory content. This dimension can also be affected by multimodal presentation, such as with HearThere, where the subtle vibration of the bone conduction cued some users to the provenance of some sounds. With HearThere, characterization along this dimension was also seen to change over time, as users acclimated to the display and improved their ability to discriminate its sources. A purposefully easily differentiable display might even heighten this effect by introducing noticeable vibration in sync with the presentation.

Perceptual Interaction – Embodied interaction describes interaction that occurs as a result or byproduct of the user's sensorimotor actions, and not through indirect manipulation mapped to interface behavior. Examples include continuous head tracking driving spatial audio rendering, or user sensing used to infer attention. Between embodied and mapped interaction, intuitively mapped motor actions (e.g. cupping an ear or touch zoom in HearThere) can quickly become embodied second nature to users due to the interaction's inherent physicality.

Sensorimotor Learning – Instantaneous, or scaffolded sensorimotor learning builds on top of existing sensorimotor contingencies, such as spatial sound perception. Entirely new contingencies can take weeks to months to develop in adults, as is the case with many examples from the sensory substitution literature. For this reason, scaffolding is preferred wherever possible.

Attentional Amplification – Attentional amplification dynamically adjusts the intensity (levels, directivity, contrast, specularity, etc) of a source in response to the inferred selective attention of the user. This adjustment can be subtle, barely noticeable to the user, or as intense as a spotlight in the dark. HearThere's attentional system combines user sensing (eye tracking, EEG, motion) with explicit attentional gestures (touch zoom) to foreground different audio sources in the user's vicinity. Inferred selective attention results in subtle adjustment, while the touch gestures produce more obvious effects.

Cognitive Extension – Cognitive extension leverages AIs that observe the data/media sources alongside the user and adjust the intensities of associated perceptual stimuli to promote sources of interest (or reduce sources of distraction). This could be used to bring user attention to particular events (for learning or to promote rare events), or in the current version of HearThere, simply to suppress wind noise. Future contexts for cognitive extension include live musical performance, where AI-based source separation could be coupled with attentional amplification to enhance individual experience of performers of interest.

System Adjustment/Configuration – Transpresence systems designed for set-and-forget configuration do not allow any adjustment by users during operation in the field. HearThere's time navigation interface minimizes interruption of the experience through the use of prominent preset buttons on its default graphical user interface, and by limiting exposure to configurable features. Because of buffering delays, users are not incentivized to change the time setting often. The need for careful interaction design in this dimension will grow as system complexity increases (e.g. with deeper integration of the Tidzam AI).

5.1.2 Continuity

With our design constraints met, we can begin to investigate the higher level contours of transpresence experience. Results from the first field study point to a rich area. I focus this discussion on two related topics: *continuity* and, in the next section, *confusion*.

In a simple live media experience, continuity is my term to describe a quality of live action unfolding at a distance. As an illustrative example, consider the following hypothetical question: given the option of listening for one hour to a recording of a forest or to a comparable live stream of the forest, which would you choose? What if it were all a ruse, and both were just recordings? Would you be upset at the misdirection?

Continuity is about more than temporal liveness, however. In a complex system with extended spatial and temporal elements, like the one at Tidmarsh, another kind of continuity manifests in the knowledge that the sound one is perceiving has a real, physical origin in space and time, that it emanates (or emanated) from that place, and that the place has a continuous identity carrying it forward. One's ability to walk through the space of a sound from the past and relate it to the present place and present moment brings back the feeling of liveness; the recording is not being replayed but re-lived.

In the HearThere field study, I observed continuity expressed in multiple subjects' excited reactions to perceiving change over time and comments about the consistent acoustic properties of the sensory landscape (e.g. metaphors describing nature as a concert hall). One subject's references to her daily sense of "anticipation" with HearThere capture a desire to experience the sound perceptually through the device, as opposed to listening to it through simpler media (the always available web-based audio database for example). I attribute this sense of anticipation to excitement at continuity. Finally, I observed the effects of continuity in how interactions with other species became newly perceptible to some subjects, who changed not only their behaviors in the natural environment but also their social relationships with wildlife as a result.

In the system at Tidmarsh, continuity has some basic foundations: sensors remain in place, recording in a consistent manner over time, and are mapped to individual perception through a consistent spatial model. The continuous identity of place, however, raises deeper questions for this thesis and future work. *What are the limits of spatial and temporal extent* (to again borrow Kaplan's term) *within which a continuous identity of place is maintained?* How far back in time can we go from the present and still accept a continuity of time? How far out in space from the listener constitutes a coherent site? From Kaplan:

> An endless stream of stimuli both fascinating and different from the usual would not qualify as a restorative environment for two reasons. First, lacking extent, it does not qualify as an environment, but merely an unrelated collection of impressions. And second, a restorative environment must be of sufficient scope to engage the mind... Extent also functions at a more conceptual level. For example, settings that include historic artifacts can promote a sense of being connected to past eras and past environments and thus to a larger world. [66]

In conversation with the filmmaker Michel Gondry, Noam Chomsky poses this as a complex epistemological problem of what he calls *psychic continuity* [104]. Chomsky highlights, by example, the implicit assumptions that go into our conceptions of identity over time: a cutting of a willow tree, for example, could be planted and grow into a tree that we would consider to have a new identity, despite its apparent and genetic equivalence to the original. The Charles River could be split into tributaries, and flow with either clean water or toxic waste, and it would still be considered the Charles River. But, paraphrasing Chomsky's example, if the Charles were to be lined with concrete and filled with cargo ships it would become a canal. And if it were to be frozen solid and a yellow line painted down the middle, it would become a road [104]. Perhaps the most famous example of a continuity problem is the rebuilt Ship of Theseus [97], which, bearing no physical parts of the original, holds a stronger claim to the ship's identity than its physically original counterpart. Some things appear to maintain their identities through certain kinds of change, and others do not.

Environmental restoration raises specific questions about psychic continuity along the temporal dimension. The design of the Tidmarsh restoration was informed by research into what the site was like before industrialization dramatically changed it. It will, of course, never be the same again, but to what extent does restoration link those disjoint identities together? What do we make of the intervening period? And what of its glacial history, traces of which are evident all around? Kaplan suggests that certain artifacts can "promote" a sense of connection that establishes extent over time. Direct evidence of the glacial history was found in deep core samples taken prior to the restoration. Still, a great majority of the record has been lost or was never captured.

The environmental sensing on the site adds new complexity to the continuity problem because we now hold a continuous record linking the farm to the restoration wetland. In less than a century, the local effects of industrialized farming and the ensuing restoration process would have represented only a disjointed blip in the continuous identity of Tidmarsh, but they are now recorded in such a way that they can be directly experienced through transpresence. Sensing increases the granularity of the record by many orders of magnitude, and transpresence promises a perceptual time machine[2].

Along spatial dimensions, how does a stream of stimuli "qualify as an environment?" Again, this is an epistemological question complicated by culturally determined concepts like

---

[2] Should the sensor recordings persist for a century, will a transpresence experience of that media have any meaning? Given how preciously we now treat thin historical records such as hand-drawn maps and written accounts, and how powerful we find the deep core samples, maybe it will.

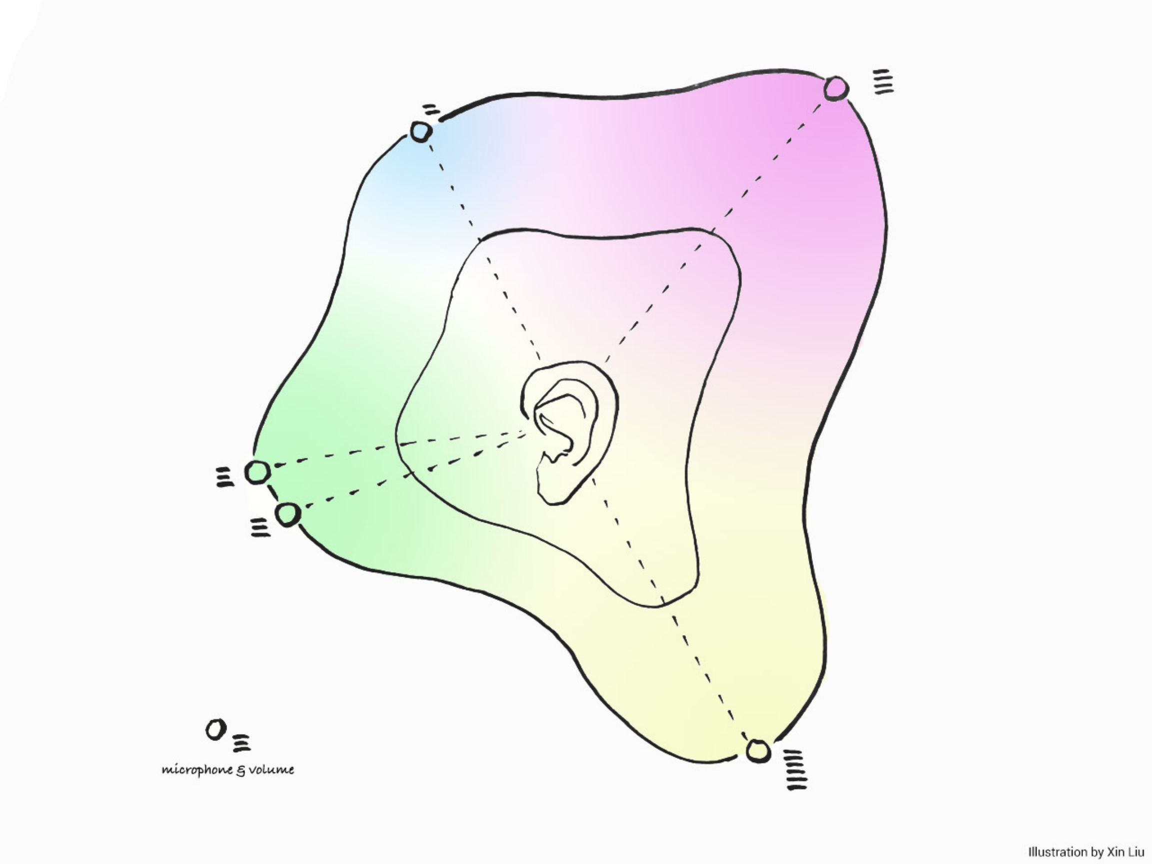


Illustration by Xin Liu

Figure 5-1: With HearThere, the shape of a listener's hearing pattern stems from the combination of several factors: the density of sensors, the distances between the listener and the sensors (roll-off curves), and the attentional state and heading of the listener. This shape is constantly morphing as the listener explores. Colors indicate different locations, and gradient the influence of each location.

property lines, ever-changing lines of sight, and an evolving understanding of perception itself. At the Grand Canyon, a vast expanse could constitute one place, whereas in a suburban subdivision, privacy bounds perception within a small plot of land or the walls of a house. Visual sight lines inform this question, as multimodal perception can create a sense of extent (e.g. visual-auditory synchrony). Sitting in MIT's Hayden Library, I can see across the Charles River to Storrow Drive; concentrating intently on the distant cars I can visually hear (or vEAR) them whooshing down the road [90]. That makes the highway, nearly 1 km away, feel like a part of the more immediate sensory landscape.

The shape of HearThere's auditory pickup pattern, illustrated in Figure 5.1.2, is not necessarily symmetric or isotropic, and the sensitivity is certainly not uniform. A greater

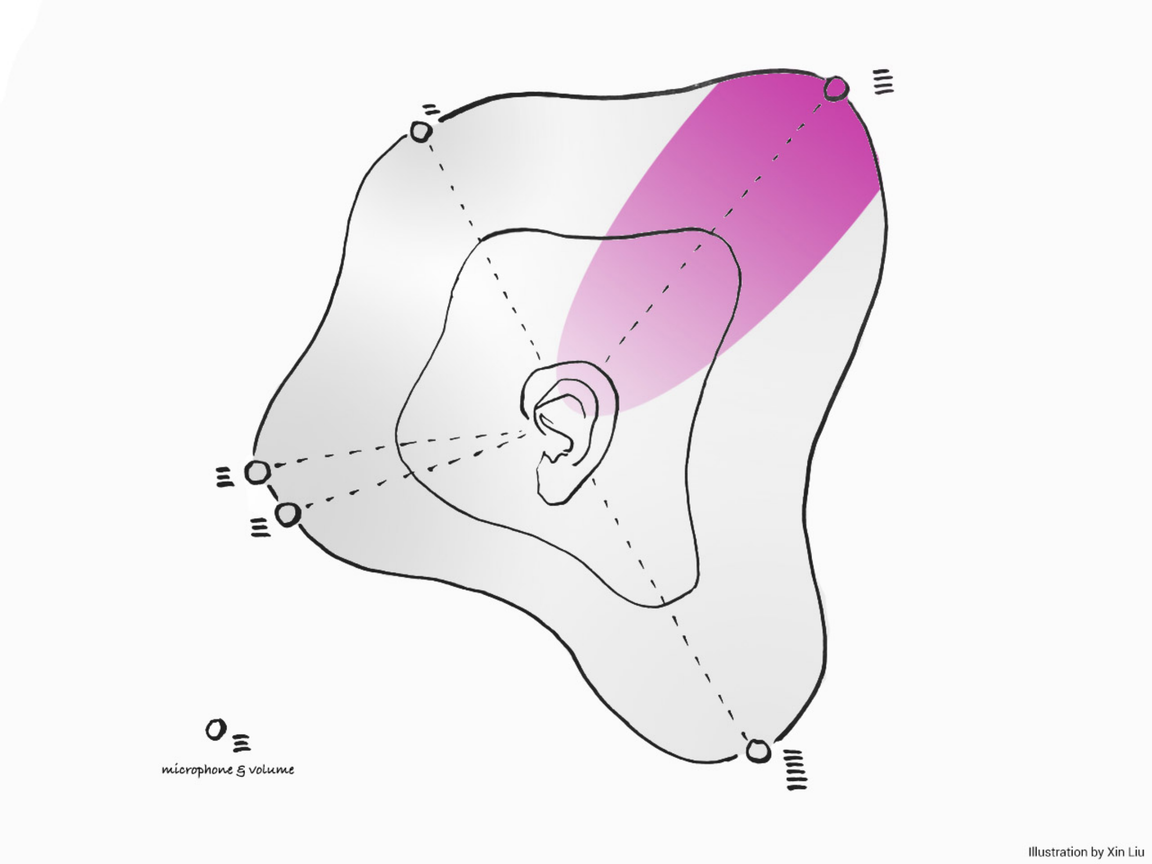


Illustration by Xin Liu

Figure 5-2: A natural gesture, such as remaining still, eyes fixated, for a long time, or cupping one's ear is used to dramatically reshape the hearing pattern into a narrow beam.

density of microphones in a given location sums to more pickup, and more interesting activity in a given place (as estimated by the Tidzam AI) can drive gains up there. The non-uniform pickup allows the extent to be larger while remaining coherent, as noisy or less important sources can be filtered out to allow more relevant sources into the mix. Further, focused attention narrows the listening beam substantially while extending its length, as shown in Figure 5.1.2. In the Storrow Drive example, HearThere could augment my 'visual hearing' with captured sound and still feel continuous, despite the great distances involved. This, too, is a matter of compatibility; it would only make sense to hear the road when my attention goes to it, and even then only faintly at first.

The epistemological nature of the continuity problem means that designing for continuity will always be a matter of the designer's own intuition combined with empirical study into what feels coherent, or "makes sense." The good news is that Chomskian psychic continuity is available even to young children, who intuitively understand that the frog is really just another form of the prince. But continuity will always be tricky at the boundaries, and as stated in the introduction, the boundaries are where we wanted to be all along. There, we start off wondering how it is that we hear something we cannot intuitively attribute to a source, and then incorporate a new model to resolve the ambiguity; some degree of confusion is a prerequisite for acquiring new perceptual sensibilities.

5.1.3 Confusion

New perceptions arise from confusion, if they arise. We see noise before we see structure, and only through self-produced sensorimotor exploration do perceptual pathways begin forming [48, 49]. In this way, perceptual confusion is an opportunity for the development of new sensibilities, produced by the repeated application of conscious reasoning and rational intuition, along with sustained attention, in attempts to distill structure or resolve an ambiguity. What differentiates the perceptual sensibility from its rational antecedents is that the source of new knowledge is sensory and comes about subjectively through individual exploration. After a new sensibility is acquired, upon encountering a new stimulus, the user takes no time to unambiguously attribute the proximate stimulus to its remote cause. This state of perceptual flow and its subconscious effects are what sensory augmentation researchers aspire to [67, 92] and why the broader HCI and Augmented Human communities are increasingly interested in this area.

HearThere extended and warped the spatial reach of perception through remote sensing, and added new temporal dimensions by blending timelines together. Those features initially prompted confusion on a number of levels, as users wondered: whether a sound came through their ears or the device; what the sound represents; where the source of sound is[3]; when the sound occurred. After the field study, I came to refer to the ear/device ambiguity as the “crow question” for the number of times it was expressed to the effect of “I hear crows. Do you hear them?”. Resolutions of any of these instances of confusion were difficult to observe directly, because I could not know what sounds subjects were internally labeling as

---

[3] With respect to confusion of location, the warping of space created by representing microphones as spatialized point sources rather than reconstructing the wavefront was a major challenge, breaking the continuity of the experience. With this naive rendering scheme, too high a density of microphones would cause unnatural effects like phasing, but too few would make the experience discontinuous. This is a shortcoming of the approach that would ideally be resolved by advances in signal processing but for now was only managed by controlling source density.

one source or another unless they voiced their response to a particular sonic event.

Remarkably, the frequency of crow questions went down to zero as users acclimated to the new perception. This suggests that something structural to the HearThere experience was providing cues subjects were able to use, subconsciously, to resolve their perceptual confusion. I hypothesize that subjects were able to *feel* enough of a difference between bone conducted audio and primary auditory perception that they began using tactility as a cue alongside higher-level contextual information. I saw anecdotal evidence of this in direct reports of the subtle vibration perception as well as in some subjects' comments that "all-of-a-sudden" they were able to tell the difference. Informal conversation with a person who lives with a permanently implanted bone conduction transducer also supported this theory. Given the mostly imperceptible tactility of the bone conduction presentation, I refer to this explanation as the "subtle multi-modality hypothesis," attributing differentiation to subjects' brains tagging bone-conducted auditory perceptions with faint somatosensory cues from the multimodal stimulus. I intend to test this hypothesis in future work.

Low-level sensory cues alone cannot account for the transformation of perceptual confusion into new sensibilities. Some inferences had a basis in pre-existing knowledge and rational intuition, such as expertise in ecology. Interestingly, expertise in hearing or headphone technology did not appear to advantage those subjects who possessed it, at least with respect to perceptual confusion and the learning process.

Temporal ambiguities occur when the transpresence technology blurs the present together with one or more instances of the recorded past. Many subjects in the field study made assignments based on expectations of seasonal soundscape characteristics, and several on prior knowledge of wildlife migration patterns. I model the experience of temporal confusion as a sequence of either logical or intuited assignments made for each perceived event (e.g. a bird call), at first consciously and later automatically. Figure 5.1.3 shows two overlapped timelines in which different parts of an ecological cycle (e.g. two seasons) are made to co-occur. Because of the continuity of the timeline, recognition of any part of a cycle would likely propagate intuition to new events going forward (particularly events related to the first assignment), an effect I observed in the field study. In the learning process, these knowledge-based inferences both reinforced and were reinforced by the low-level sensory cues discussed above.

In an ideal situation, users are able to quickly move out of an initial period of confusion

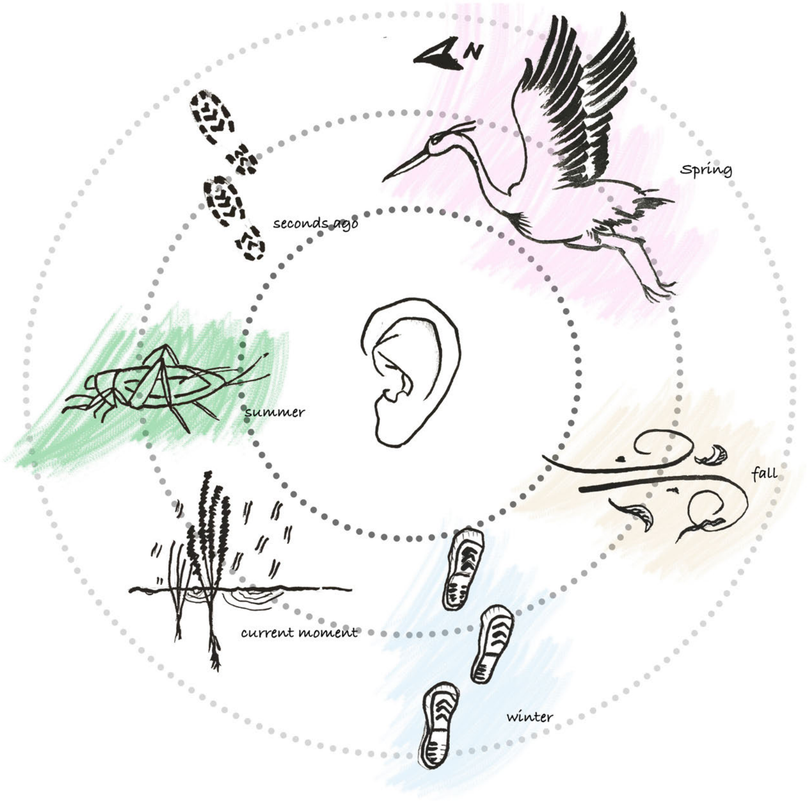


Illustration by Xin Liu

Figure 5-3: How would a listener make perceptual sense of two overlapping timelines? First, intuitions about what is likely to observed. Later, different parts of a single ecological cycle (e.g. season or animal migration pattern) may become apparent, leading to new intuitions about interactions between systems over time.

and into new and lasting perceptual sensibilities, but that may not always be the case. While the study strongly indicated increasing levels of skill being acquired over relatively short periods of time (often leading to self-reported breakthroughs), it is quite possible that some confusing stimuli were not resolved correctly. Mistakes could lead subjects to draw totally wrong conclusions, believing them to be true observations. There are clearly challenges to surmount, as well as ethical considerations in an era of fake news. Still, in the case of HearThere at Tidmarsh, the sensory cues did seem to lead to a consistent ability to make the correct assignments. If the subtle multimodal aspects of bone conduction did play a role in the learning process, there is an argument to be made for designing future AR wearables with similarly almost imperceptible cues for differentiating the digital layer from the physical world. At Tidmarsh, the odds were in favor of positively reinforced learning

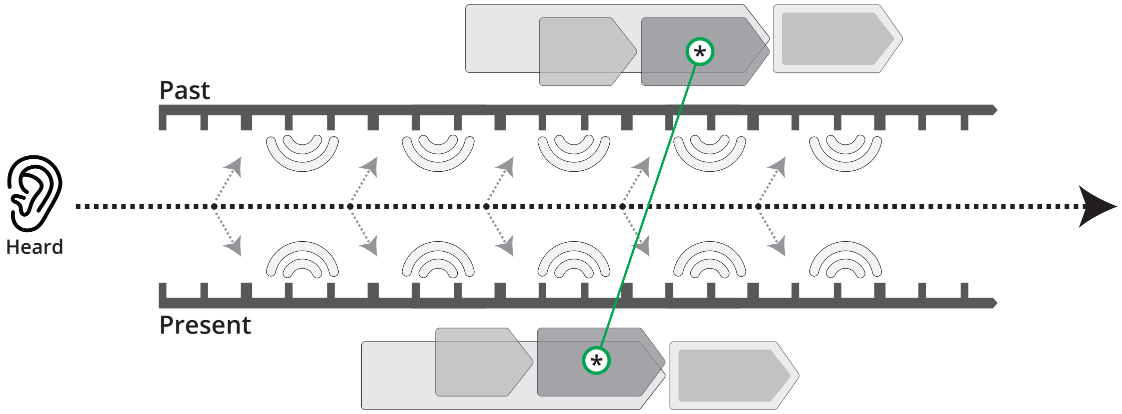


Figure 5-4: Two timelines, past and present, overlap, completely intertwined. A listener makes assignments at every time step they perceive, and recognizes *form in time* when two related events are teased apart.

because, as discussed in the last sections, the sources of data were physical phenomena. Occasional mistakes might be acceptable as long as they are significantly outnumbered by correct assignments and predicate valuable new perceptions.

In the work to date, only serendipitous alignments of the digital perception with intrinsic hearing created situations of confusion that led to positive resolution. These alignments mostly took the form of chance transient events (e.g. distant crows calling) or clearly identifiable seasonal patterns. But is there a way to increase the odds of these occurrences—to intentionally compose confusion for the purposes of encouraging certain kinds of encounters and creating particular new sensibilities? This is one of the big questions that will animate future work in transpresence. It points at a much greater role for AI in weighting different sources in the environment for consideration and attention, discussed in the next section.

5.2 Presence Future

This section outlines the future of transpresence research, beginning with near-term plans already underway to study the effects of extended use of HearThere. Finally, I discuss my vision for a future line of inquiry in which AI would meet perception halfway, producing real-world extended intelligence (EI) perceptions of the natural world. I introduce immediate applications of this future work to ecological science and public understanding of the environment and discuss its implications for the future of sensory and attentional user interfaces.

5.2.1 Longer Term Studies of Extended Presence

The immediate next step for transpresence research will be several new field studies of the use and effects of HearThere at Tidmarsh, including an evaluation of long-term independent use. Continuing challenges include the remoteness of the wetland site, barriers to public access, and robustness of the wearable device, sensing, and network infrastructure. With experience from the single-session field study and associated process and hardware improvements, I believe we can feasibly surmount these challenges. An extended study will provide the means to test new hypotheses and investigate some of the areas articulated in the last section. The planned studies have two main goals: first, to compare HearThere's bone-conducting presentation to recently released air-conducting non-occluding headphones (subtle multimodality hypothesis); and second, to observe the effects of long-term use, which preliminary data show would provide new insight for the design of transpresence and sensory augmentation technologies.

Comparing Bone Conduction with Non-Occluding Air Conduction

A recently-announced glasses-mounted headphone [17] provides a unique opportunity to test my hypothesis that subjects' rapidly improving skill at separating presented audio from natural hearing comes about partly as a result of the subtle tactile sensation of bone conduction. There are many ways to approach this question as a side-by-side comparison of the two platforms, but as my focus is applied, my interest leans towards field testing and observation at first. In one design under consideration, I would split a set of first-time users into two groups, one given the original HearThere and the other the head-tracking glasses platform paired with an appropriately ported mobile application. The experimental protocol for each group would match the one from the original HearThere study, detailed in Section 4.4. Having previously observed subjects' skill at separation improving over time with the bone-conducting HearThere, I would look for differences in subjects' descriptions of their perceptual confusion over the course of a session. A second experimental design under consideration is a cross-over study. This design would keep the group intact, but randomize the order in which subjects are exposed to each type of audio presentation, hypothesizing that a rise in confusion would follow the transition from bone conduction to air conduction but not vice versa. Of course, both experiments are premised on the air-conducting headphone being equivalently comfortable and non-occluding, and there are variables there beyond my control. Finally, to establish a difference between the two forms of presentation, a more controlled study in the lab is also called for. That study would introduce virtual, spatialized channels likely to be confused with sources presented through a speaker array in a room. As such, proper facilities for a controlled experiment would have to be arranged, tuned to match each headphone's frequency response. Specifics of that experimental design will have to await the release of the glasses-mounted headphone, as few details of the platform have been disclosed. The result, however, would have deep implications for the design of future sensory wearables that would aim to perfectly blend sources with the world, or alternatively to offer imperceptible cues for separation.

Long-Term Use

Responses from subjects in the initial field study strongly support further study. The case for new perceptual sensibilities arising from transpresence could be made more strongly through observations of extended use by independent explorer subjects.

The current design calls for a three-month cycle with five people independently using personal HearThere devices regularly, aiming to record at least three nature walk sessions per week. Quantitative data (GPS, head orientation, feature usage, physiological sensors, etc) would be stored on personal mobile devices for later upload, and diaries would be uploaded through a survey web interface. Both qualitative and quantitative data collection would be incorporated directly into the sessions, with recording automation afforded by the Sensorium mobile application.

As mentioned in the last chapter, I am already running a pilot with one subject, who has been regularly using the device for a period of approximately twelve weeks without major technical issues, and with exciting initial results indicated by her reporting. As before, I expect to see separation and localization abilities improving with repeated practice. Stemming from the pilot, new topic areas for observation include: rising skill at and comfort with use; new discoveries, such as trees falling in the forest and other mysteries uncovered; anticipation of use and broad impacts of the HearThere experience outside of the sessions; and appreciation of a more balanced relationship with wildlife, manifested first as increasing awareness of self sound and its disturbing effects on sensitive animals. In the pilot, we

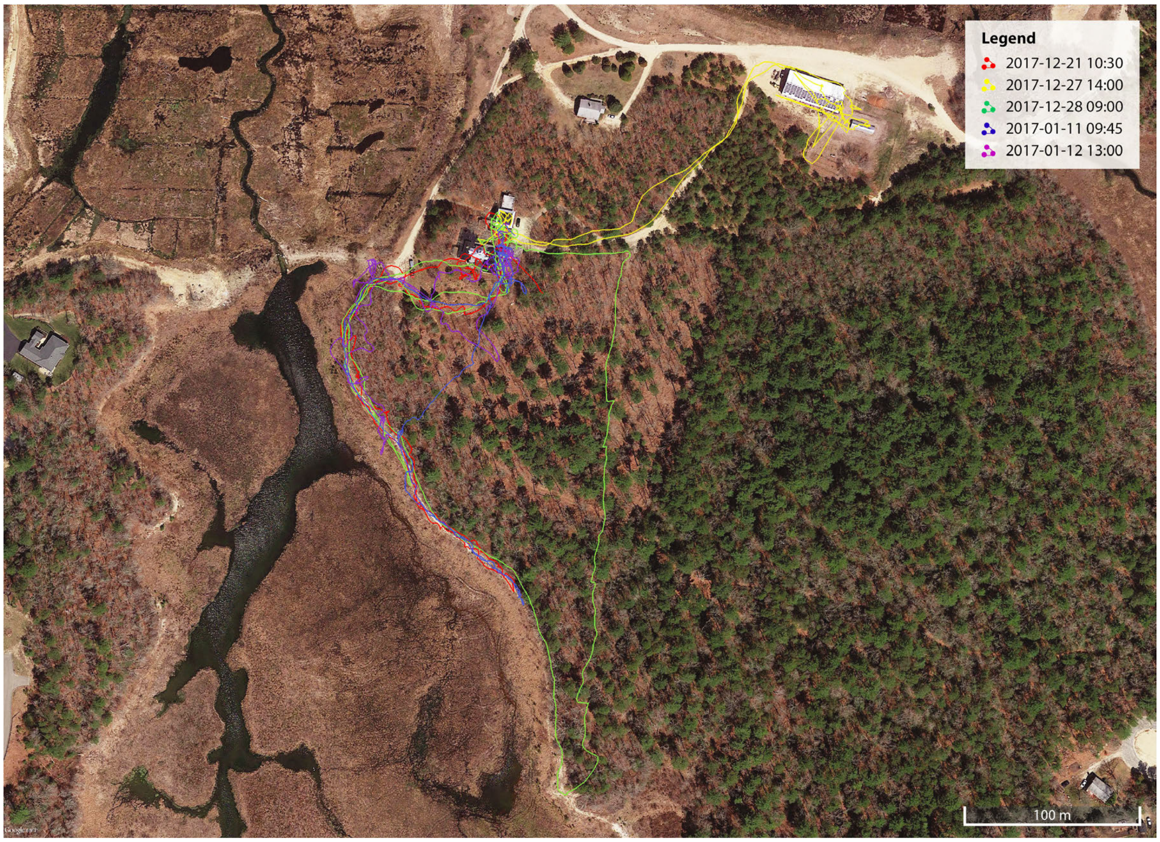


Figure 5-5: GPS tracks from a single subject: walks conducted in the first week of a pilot study running for twelve weeks.

have seen the hardware holding up well under real-world conditions, showing that small adjustments would make it essentially product-worthy. In light of this, finalizing a product design would also be a goal of the extended study.

There is hope to eventually facilitate use of HearThere by the visiting public at Tidmarsh, though a number of challenges remain to be surmounted first. Another iteration of industrial and software design, as well as software testing, are needed to ensure robustness and long-term stability of the platform in the hands of end users. At present, the microphone installation is limited to a relatively small area of the larger, publicly-accessible site. Finally, logistics will have to developed around the distribution and recovery of prototype hardware.

Another possible avenue to wider use would be to port the Sensorium application to Bose's AR product [17], which visitors could potentially buy themselves and bring to the site. Some features, such as touch, are not yet supported on that platform and would have to be disabled or adapted to screen-based input. Bluetooth audio latencies in that product may also prove to be a problem. In either case, maintaining HearThere devices and related software in the field would allow us to develop and experiment with a host of new features. This includes an expansion of the extended intelligence concept at the intersection of HearThere and the Tidzam AI, discussed in the next section.

5.2.2 Towards Extended Intelligence

By extending individual hearing into a distributed network of sensors and a database of recordings, HearThere opened up new spatial and temporal dimensions of perception not limited in the same way by the physiology of the ear and the physics of sound propagation. But there is another set of dimensions only barely explored to date in the work, which I see as the next frontier of this research. These are the *cognitive* dimensions of perception, at the interface with AI. In the same way that inferred direction of attention was used in HearThere to positively weight audio channels, high-level classifications of wildlife in real-time and recorded data streams could be used as inputs to the channel weighting scheme. Because transpresence technologies weave into existing perception, the result is a form of *extended intelligence* (EI) [60]. EI is a broad new field introduced by Joi Ito and the MIT Media Lab community, brought together by an understanding of intelligence (human and machine) as a fundamentally distributed phenomenon. One aspect of this reformulation of AI concerns the interface to distributed intelligence, which becomes less like that of a mysterious oracle, and more like a part of us.

The dynamic, channel-weighting design of HearThere is well-suited to this mold, allowing different agents, including but not limited to individual attention, to highlight different channels of interest while always maintaining balance and continuity. An external AI engine already acts in this way with HearThere: Tidzam was used in the initial study for wind noise suppression. Because we did not achieve tight synchronization between the AI's process and the live audio streams, Tidzam's classifications were integrated slowly over time, and as a result, only noise suppression was feasible. However, synchronization is required for most of the real-time applications of the AI, and work to achieve it is underway.

In the short term, a number of exciting opportunities beyond noise suppression are opened up by the convergence of transpresence and AI. With the technology already available, a user could pick a class of bird sound and have HearThere positively weight those microphones which most frequently observe that class. Automatic wildlife identification paired with user interest would guide users towards desired subjects or scenes without ever issuing explicit instructions. Of course, there are no guarantees that a user would find what they are looking for, but the odds of discovery could be increased by adjusting the weighting function towards the positive classification. With better audio synchronization, this could go from a slowly changing parameter to an instantaneous bump at the first transient classification. Those dynamics would have to be tuned to feel natural while catching as many instances of interest as possible. User preferences for certain classes would ideally be configured in advance, on a set-and-forget web interface with images and descriptions of each species. A user interested in learning about a particular species while improving their skill could start off by tilting the weighting function dramatically towards unnatural highlighting of the observation, perhaps even introducing artificial auditory feedback. Gradually, the system would get out of the way, allowing the user's identification skill to take over.

Also in the near term, extended intelligence could play a role in helping users understand the temporal dimensions of the environment. If a user were interested in what the previous summer sounded like, for example, the system could be directed to pick the most statistically representative day, or to contrast it with the most unusual day of the season, both chosen based on data produced by the AI. The system could even take ecological cycles or climate events into account, matching the present moment with its ecological precedents, such as a seasonal migration or a storm that precipitated current conditions, allowing users to discover interactions within or between ecological systems across time.

In the longer term, as both the user interfaces and the back-end connections between transpresence and AI mature, and as the AI itself improves, new applications will emerge. HearThere and Tidzam could become a practical tool for in-situ phenology, the study of periodic plant and animal life cycle events, facilitating new discoveries that would feed back into the AI/EI. HearThere could pull users towards the first sighting of a specimen or species of the season, or allow visitors to witness the last sighting before its extinction. Integrating an understanding of different species' perceptual systems would allow the system to facilitate encounters that do not disturb the wildlife under observation. Birdwatchers already take steps to protect the exact locations of sensitive species like owls, for example; HearThere could provide extraordinary opportunities to witness their activities at safe distances. Finally, in a true EI ecosystem, information could flow back up to the AI, which could learn from its human users' attentional dynamics where to set its computational sights.

5.3 Conclusions

This dissertation assembles a diverse body of work comprising environmental sensing, perception, interaction, and art, linked by two personal motives I gave at the outset:

> "...to give you the feeling of my sensory superpower; and to explore the beautiful contradictions I see in your face when you feel it."

Following this articulation, the prevailing, task-oriented technologies of ubiquitous computing were presented both as a foil to my aims and, reconfigured, as the tools to achieve them. I introduced an alternative composition of these technologies that, in rare moments of deliberate perceptual attention, would gently lead participants to a new perceptual plane, through what I termed the development of new perceptual sensibilities. I proposed the *networked sensory landscape* as the unique site for this work, and showed how one could be built. Finally, I placed this class of technologically-mediated perceptual experiences under a new umbrella called *transpresence*, highlighting that what they share is a melding together of distributed media and sensor mappings with existing perceptual abilities. The primary work comprising this investigation targeted spatial hearing and auditory attention, sensing the user and using non-occluding auditory display to produce situated perceptions of environmental sensor data.

Those who tested HearThere in the field experienced something not entirely unlike a hearing aid, a "turning up the world," but fluid in time and otherwise released from the physical constraints of sound propagation. Some described their experience as like having a superpower, in perceptual metaphors, in relation to their existing sense of hearing, or often just in terms of the world out there. My exploration of what I called *listening-looking* occurred in stages: early on, where it emerged unexpectedly in audiences of sound installations such as ListenTree and Moss Space; next, where I constructed lab-based contexts to elicit and closely observe it; later, in HearThere's attention model and interaction design; and finally, closing the loop, in users of HearThere at Tidmarsh. Still, I am only scratching the surface. Where future work might investigate applications to extended intelligence, improved sensing and modeling of attention, and the effects of subtle multimodal perception, my creative drive remains rooted in a desire to understand listening-looking in myself and others.

Visual and tactile modalities present wide open paths to extend this work. For the research in this dissertation, suitably lightweight and non-occluding visual AR displays were not ready for the wetland context, but they likely will be soon. The potential for multimodal sensory technologies to increase realism and produce powerful experiences of embodiment has been well-established by researchers in recent years [8, 37, 78].

What is it, to think deeply about or inside one's own perception as a way to momentarily increase its sensitivity? How and where can technology support this kind of thinking? Is that something technology should do everywhere, or only in special places, underscoring their fragile ephemeralities? I escaped the problems of privacy and human memory by focusing on environmental restoration, where few would dispute the benefits of being in nature while improving our appreciation of it through sensor technologies. Listening in on human spaces in the way we monitor Tidmarsh is a thing of dystopia. That said, given how much we enjoy watching and re-watching our grandparents on film, why not walk around inside of the recordings, in the places they were made, discovering new details each time? And what of the longer future? Will the sensory record bear any meaning to a site like Tidmarsh that is well below one hundred year projections of sea-level rise?

In daily life and on forest hikes, ubiquitous tools like GPS navigation direct us to follow the lowest-cost paths to our pre-programmed destinations in series of discrete steps. Any unexpected turn of events silently prompts a computation of the next best path. We have come to expect all technologies to follow this logic, getting us where we set off to go without concerning us with the contours of the way there. Transpresence, though made of the same technological ingredients, is not about getting to an endpoint. Instead, it is about expanding and enriching experience—in particular, developing and practicing new perceptual sensibilities—along the way. Vast new perceptual dimensions will be incorporated into everyday experience. But by design, the future of transpresence should feel like being there always did, that most everyday extraordinary thing where the boundaries of perception open rabbit holes to distant worlds when you seek for them, and close them down again when you continue on your way.

Bibliography

[1] Laurie Anderson. The handphone table. `https://www.moma.org/pdfs/docs/press_archives/5652/releases/MOMA_1978_0088_81.pdf`, August 1978.

[2] Malika Auvray, Sylvain Hanneton, Charles Lenay, and Kevin O'Regan. There is something out there: distal attribution in sensory substitution, twenty years later. *Journal of Integrative Neuroscience*, 4(04):505–521, 2005.

[3] Malika Auvray and Erik Myin. Perception With Compensatory Devices: From Sensory Substitution to Sensorimotor Extension. *Cognitive Science*, 33(6):1036–1058, August 2009.

[4] Malika Auvray and Erik Myin. Perception with compensatory devices: from sensory substitution to sensorimotor extension. *Cognitive Science*, 33(6):1036–1058, 2009.

[5] Paul Bach-y Rita, Carter C Collins, Frank A Saunders, Benjamin White, and Lawrence Scadden. Vision Substitution by Tactile Image Projection. *Nature*, 221(5184):963–964, March 1969.

[6] Paul Bach-y Rita and Kurt A. Kaczmarek. Tongue placed tactile output device, 1999.

[7] Sarah E Battersby, Michael P Finn, E Lynn Usery, and Kristina H Yamamoto. Implications of web mercator and its use in online mapping. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 49(2):85–101, 2014.

[8] Christopher C. Berger, Mar Gonzalez-Franco, Ana Tajadura-Jiménez, Dinei Florencio, and Zhengyou Zhang. Generic HRTFs may be good enough in virtual reality. Improving source localization through cross-modal plasticity. *Frontiers in Neuroscience*, 12(FEB), 2018.

[9] Jess Bidgood. The rewilding of a century-old cranberry bog. *The New York Times*, page A9, Jul 5, 2017.

[10] Michail Bletsas. The MIT Media Lab's Glass Infrastructure: An Interactive Information System. *IEEE Pervasive Computing*, 11(2):46–49, 2012.

[11] S Brave and A Dahley. inTouch: a medium for haptic interpersonal communication. In *ACM Conference on Human Factors in Computing Systems (CHI)*, 1997.

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

[13] Jenna Burrell, Tim Brooke, and Richard Beckwith. Vineyard Computing: Sensor Networks in Agricultural Production. *IEEE Pervasive Computing*, 3(1):38–45, Jan 2004.

[14] Wendy Carlos. Sonic Seasonings. Columbia Records, 1972.

[15] D R Chebat, F C Schneider, R Kupers, and M Ptito. Navigation with a sensory substitution device in congenitally blind individuals. *Neuroreport*, 2011.

[16] Barry H Cohen, Richard J Davidson, Joseph A Senulis, Clifford D Saron, and Douglas R Weisman. Muscle tension patterns during auditory attention. *Biological Psychology*, 33:133–156, jan 1992.

[17] Bose Corporation. Bose introduces audio augmented reality platform. `https://globalpressroom.bose.com/us-en/pressrelease/view/1905`, March 9, 2018. Accessed: 2018-03-10.

[18] J Cottingham, R Stoothoff, and D Murdoch. The Philosophical Writings of Descartes (I). Cambridge University Press, Cambridge, UK, 1993.

[19] Jonathan Crary. *Suspensions of Perception: Attention, Spectacle, and Modern Culture.* MIT Press, October 2001.

[20] Mihaly Csikszentmihalyi. *Flow and the Foundations of Positive Psychology.* Springer, Dordrecht, 2014.

[21] Artem Dementyev, Hsin-Liu (Cindy) Kao, and Joseph a. Paradiso. SensorTape: Modular and Programmable 3D-Aware Dense Sensor Network on a Tape. *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST)*, pages 649–658, 2015.

[22] Artem Dementyev and Joseph A. Paradiso. WristFlex: Low-Power Gesture Input with Wrist-Worn Pressure Sensors. In *Proceedings of the 27th annual ACM Symposium on User Interface Software and Technology (UIST)*, pages 161–166, 2014.

[23] Ophelia Deroy and Malika Auvray. Reading the world through the skin and ears: a new perspective on sensory substitution. *Frontiers in psychology*, 3, 2012.

[24] Gershon Dublon. *Beyond the lens : communicating context through sensing, video, and visualization.* PhD thesis, Massachusetts Institute of Technology, 2011.

[25] Gershon Dublon, Brian Mayton, Don Derek Haddad, Clement Duhart, Spencer Russell, Glorianna Davenport, and Joseph A. Paradiso. A Ubiquitous Sensing Model for Environmental Science , Interaction Research , and Public Engagement with Restored Wetlands. In *Society of Wetland Scientists 2017 Annual Meeting*, San Juan, Puerto Rico, 2017.

[26] Gershon Dublon, Brian Mayton, Ben Hogue, and Joseph A. Paradiso. Data-Driven Elevator Music (Video). `https://vimeo.com/152768612`, 2011.

[27] Gershon Dublon and Joseph A Paradiso. Tongueduino: hackable, high-bandwidth sensory augmentation. In *The ACM CHI Conference on Human Factors in Computing Systems*, Austin, TX, 2012.

[28] Gershon Dublon and Joseph A Paradiso. Extra Sensory Perception. *Scientific American*, 311(1):36–41, 2014.

[29] Gershon Dublon and Joseph A. Paradiso. FingerSynth: Wearable Transducers for Exploring the Environment and Playing Music Everywhere. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, 2014.

[30] Gershon Dublon, Laurel S Pardue, Brian Mayton, Noah Swartz, Nicholas Joliat, Patrick Hurst, and Joseph A Paradiso. DoppelLab: Tools for Exploring and Harnessing Multimodal Sensor Network Data. In *IEEE International Conference on Sensors*, Limerick, Ireland, 2011. MIT Media Lab.

[31] Gershon Dublon and Edwina Portocarrero. ListenTree. Art Installation, Rencontres Internationales de Documentaire de Montreal (Montreal International Documentary Festival). `http://ridm.ca/uploads/document/catalogue{_}planches{_}lowres.pdf`, Nov 2014.

[32] Gershon Dublon and E Portocarrerro. ListenTree: Audio-Haptic Display in the Natural Environment. In *International Community on Auditory Display*, 2014.

[33] Clement Duhart, Gershon Dublon, Brian Mayton, and Joseph A. Paradiso. Deep Learning Locally Trained Wildlife Sensing in Real Acoustic Wetland Environment. In *SUBMITTED, IN REVIEW*, 2018.

[34] D. Estrin, L. Girod, G. Pottie, and M. Srivastava. Instrumenting the world with wireless sensor networks. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 4, pages 2033–2036, 2001.

[35] Jonathan B Fritz, Mounya Elhilali, David Warland, and Shihab A Shamma. Auditory attention–focusing the searchlight on sound. *Current opinion in neurobiology*, 17(4):437–455, jan 2007.

[36] Maurizio Garbarino, Matteo Lai, Dan Bender, Rosalind W. Picard, and Simone Tognetti. Empatica E3: A wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition. *Proceedings of the 2014 4th International Conference on Wireless Mobile Communication and Healthcare (MOBIHEALTH)*, pages 39–42, 2015.

[37] Mar Gonzalez-Franco and Jaron Lanier. Model of illusions and virtual reality. *Frontiers in Psychology*, 8(June):1–8, 2017.

[38] Google. Google Glass Design Principles, 2015.

[39] Google. Google VR iOS API Reference. `https://developers.google.com/vr/ios/reference/`, September 2017.

[40] Google Inc. Google Glass Design Guidelines: User Interface, 2014.

[41] Saul Greenberg, Nicola Marquardt, Til Ballendat, Ro Diaz-Marino, and Miaose Wang. Proxemic interactions: The New Ubicomp? *interactions*, 18(1):9, January 2011.

[42] Daniel A. Gross. How a family is transforming its cranberry bog from environmental liability to climate change buffer. `https://www.pri.org/stories/2017-09-29/massachusettss-iconic-cranberry-bogs-leave-legacy-environmental-damage`, September 29, 2017. Accessed: 2018-03-10.

[43] Don Derek Haddad. Resynthesizing Reality: Driving Vivid Virtual Environments from Sensor Networks. *Master's Thesis, Massachusetts Institute of Technology, School of Architecture and Planning, Program in Media Arts and Sciences*, 2018.

[44] Don Derek Haddad, Gershon Dublon, Brian Mayton, Spencer Russell, Xiao Xiao, Ken Perlin, and Joseph A. Paradiso. Resynthesizing reality. In *ACM SIGGRAPH 2017 Talks*, pages 51:1—-51:2, Los Angeles, California, 2017. ACM.

[45] Lars Hallnäs and Johan Redström. Slow technology: Designing for reflection. *Personal and Ubiquitous Computing*, 5(3):201–212, 2001.

[46] Quentin Hardy. Looking Beyond the Internet of Things. *The New York Times*, January 2016. Accessed: 2016-06-01.

[47] Ed Hawkins. Climate spirals, 2016.

[48] Richard Held and Alan Hein. Movement-produced stimulation in the development of visually guided behavior. *Journal of Comparative and Physiological Psychology*, 56(5):872–876, October 1963.

[49] Richard Held, Yuri Ostrovsky, Beatrice DeGelder, Tapan Gandhi, Suma Ganesh, Umang Mathur, Pawan Sinha, Beatrice de Gelder, Tapan Gandhi, Suma Ganesh, Umang Mathur, and Pawan Sinha. The newly sighted fail to match seen with felt. *Nature neuroscience*, 14(5):551–553, may 2011.

[50] Stefan Helmreich. Listening Against Soundscapes. *Anthropology News*, 51(9):10–10, January 2011.

[51] Stefan Helmreich. GRAVITY'S REVERB: Listening to Space-Time, or Articulating the Sounds of Gravitational-Wave Detection. *Cultural Anthropology*, 31(4):464–492, 2015.

[52] Stefan Helmreich. Melt. theorizing the contemporary. *Cultural Anthropology website*, 2016.

[53] Linbi Hong, Jennifer M. Walz, and Paul Sajda. Your eyes give you away: Prestimulus changes in pupil diameter correlate with poststimulus task-related EEG dynamics. *PLoS ONE*, 9(3):e91321, mar 2014.

[54] Wen Hu, Nirupama Bulusu, Chun Tung Chou, Sanjay Jha, Andrew Taylor, and Van Nghia Tran. Design and evaluation of a hybrid sensor network for cane toad monitoring. *ACM Transactions on Sensor Networks*, 5(1):1–28, February 2009.

[55] Susan Hurley and Alva Noë. Neural Plasticity and Consciousness. *Biology and Philosophy*, 18(1):131–168, January 2003.

[56] InteraXon. Interaxon Muse EEG Technical Specification, Validation, and Research Use, 2015.

[57] InteraXon. Muse Research Tools - Available Data. `http://developer.choosemuse.com/research-tools/available-data`, 2017.

[58] Interfluve. Tidmarsh Restoration Project, 2015.

[59] Hiroshi Ishii, Craig Wisneski, Scott Brave, Andrew Dahley, Matt Gorbet, Brygg Ullmer, and Paul Yarin. ambientROOM: Integrating Ambient Media with Architectural Space. In *ACM Conference on Human Factors in Computing Systems (CHI)*, number April, pages 173–174, 1998.

[60] Joi Ito. Extended intelligence. `http://v2.pubpub.org/pub/extended-intelligence`, March 1, 2016. Accessed: 2018-03-20.

[61] Yuta Itoh, Jason Orlosky, Kiyoshi Kiyokawa, and Gudrun Klinker. Laplacian Vision: Augmenting Motion Prediction via Optical See-Through Head-Mounted Displays. In *Augmented Human International Conference*, Geneva, Switzerland, 2016.

[62] Raj Jain and Imrich Chlamtac. The P2 algorithm for dynamic calculation of quantiles and histograms without storing observations. *Communications of the ACM*, 28(10):1076–1085, 1985.

[63] N Joliat, B Mayton, and Joseph A Paradiso. Spatialized Anonymous Audio for Browsing Sensor Networks via Virtual Worlds. In *International Community on Auditory Display*, 2013.

[64] Daniel Kahneman, Bernard Tursky, David Shapiro, and Andrew Crider. Pupillary, Heart Rate, and Skin Resistance Changes During a Mental Task. *Journal of Experimental Psychology*, 79(1):164–167, 1969.

[65] Hsin-Liu (Cindy) Kao, Artem Dementyev, Joseph A. Paradiso, and Chris Schmandt. NailO: Fingernails as an Input Surface. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, pages 3015–3018, 2015.

[66] Stephen Kaplan. The restorative benefits of nature: Toward an integrative framework. *Journal of Environmental Psychology*, 15(3):169–182, 1995.

[67] Kai Kaspar, Sabine König, Jessika Schwandt, and Peter König. The experience of new sensorimotor contingencies by sensory augmentation. *Consciousness and Cognition*, 28(1):47–63, 2014.

[68] Moritz Kassner, William Patera, and Andreas Bulling. Pupil: An Open Source Platform for Pervasive Eye Tracking and Mobile Gaze-based Interaction. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, pages 1151–1160, New York, NY, USA, 2014. ACM.

[69] Gary R King, Wayne Piekarski, and Bruce H Thomas. ARVino: Outdoor Augmented Reality Visualisation of Viticulture GIS Data. In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 2005.

[70] Marcus Kison. touched echo. `http://www.markuskison.de/#touched_echo`, December 2007.

[71] R Kleinberger, G Dublon, J Paradiso, and T Machover. PHOX Ears: Parabolic, Head-mounted, Orientable, eXtrasensory Listening Device. In *New Interfaces for Musical Expression*, Baton Rouge, LA, 2015.

[72] Rébecca Kleinberger. *Singing about singing: using the voice as a tool for self-reflection.* Master of science, MIT, 2014.

[73] Árni Kristjánsson, Alin Moldoveanu, Ómar I. Jóhannesson, Oana Balan, Simone Spagnol, Vigdís Vala Valgeirsdóttir, and Rúnar Unnthorsson. Designing sensory-substitution devices: Principles, pitfalls and potential. *Restorative Neurology and Neuroscience*, 34(5):769–787, 2016.

[74] Ron Kupers, Daniel R Chebat, Kristoffer H Madsen, Olaf B Paulson, and Maurice Ptito. Neural correlates of virtual route recognition in congenital blindness. *Proceedings of the National Academy of Sciences*, 107(28):12716–12721, July 2010.

[75] Simon Lacey, Randall Stilla, and K. Sathian. Metaphorically feeling: Comprehending textural metaphors activates somatosensory cortex. *Brain and Language*, 120(3):416–421, 2012.

[76] Sang-won Leigh and Pattie Maes. AfterMath: Visualizing Consequences of Actions through Augmented Reality. *Extended Abstracts of the ACM CHI Conference on Human Factors in Computing Systems*, 2, 2015.

[77] Joshua Lifton, Mathew Laibowitz, Drew Harry, Nan-Wei Gong, Manas Mittal, and Joseph A Paradiso. Metaphor and Manifestation—Cross-reality with Ubiquitous Sensor/Actuator Networks. *IEEE Pervasive Computing*, 8(3), September 2009.

[78] Xin Liu. Inward to Outward. *Master's Thesis, Massachusetts Institute of Technology, School of Architecture and Planning, Program in Media Arts and Sciences*, 2018.

[79] Jaime Lloret, Miguel Garcia, Diana Bri, and Sandra Sendra. A wireless sensor network deployment for rural and forest fire detection and verification. *IEEE Sensors*, 9(11):8722–8747, 2009.

[80] Evan F. Lynch and Joseph A Paradiso. SensorChimes: Musical Mapping for Sensor Networks. In *Conference on New Interfaces for Musical Expression (NIME), to appear*, Australia, 2016.

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

[82] Gun Lee Mark Billinghurst, Adrian Clark. A Survey of Augmented Reality. *Foundations and Trends in Human-Computer Interaction*, 8(2-3):73–272, 2014.

[83] Nicolai Marquardt, Tom Gross, Sheelagh Carpendale, and Saul Greenberg. Revealing the invisible: visualizing the location and event flow of distributed physical devices. In *International conference on Tangible, Embedded, and Embodied Interaction*, 2010.

[84] Massachussetts Audubon Society. Discover Tidmarsh: Take a Walk on the Wildflower Side. *Mass Audubon Explore Newsletter*, March 2017.

[85] Keiichi Matsuda. Hyper-Reality (concept film). `http://hyper-reality.co`, May 2016.

[86] Brian Mayton, Gershon Dublon, Spencer Russell, Evan F Lynch, Don Derek Haddad, Vasant Ramasubramanian, Clement Duhart, Glorianna Davenport, and Joseph A Paradiso. The Networked Sensory Landscape: Capturing and Experiencing Ecological Change Across Scales. *Presence: Teleoperators and Virtual Environments*, 2018.

[87] Sebastian Michel, Ali Salehi, Liqian Luo, Nicholas Dawes, Karl Aberer, Guillermo Barrenetxea, Mathias Bavay, Aman Kansal, K. Ashwin Kumar, Suman Nath, Marc Parlange, Stewart Tansley, Catharine Van Ingen, Feng Zhao, and Yongluan Zhou. Environmental monitoring 2.0. *International Conference on Data Engineering*, pages 1507–1510, 2009.

[88] Pauli Minnaar, Søren Krarup Olesen, Flemming Christensen, and Henrik Møller. The Importance of Head Movements for Binaural Room Synthesis. In *International Conference on Auditory Display (ICAD)*, pages 21–25, Espoo, Finland, 2001.

[89] Marvin Minsky. Telepresence. *Omni Magazine*, pages 45–51, jun 1980.

[90] Heather Murphy. Why We ‘Hear’ Some Silent GIFs. *The New York Times*, page D2, Dec 8, 2017.

[91] S K Nagel, C Carl, T Kringe, R Märtin, and P König. Beyond sensory substitution—learning the sixth sense. *Journal of Neural Engineering*, 2:R13, 2005.

[92] Saskia K Nagel, Christine Carl, Tobias Kringe, Robert Märtin, and Peter König. Beyond sensory substitution–learning the sixth sense. *Journal of neural engineering*, 2(4):R13–26, 2005.

[93] A Noë and E Thompson. Vision and mind: Selected readings in the philosophy of perception, 2002.

[94] Alva Noë. *Varieties of Presence*. Harvard University Press, 2012.

[95] Kevin O’Regan and Alva Noe. A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24(05):939–1031, nov 2001.

[96] Joseph A Paradiso and James A Landay. Guest Editors’ Introduction: Cross-Reality Environments. *IEEE Pervasive Computing*, 8(3):14–15, 2009.

[97] Plutarch (ed. Arthur Hugh Clough). *Plutarch: Lives of the noble Grecians and Romans*. Project Gutenberg, AD 46-120, 1996 (Gutenberg).

[98] E. Portocarrero, G. Dublon, J. Paradiso, and M.V. Bove. ListenTree: Audio-haptic display in the natural environment. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*, volume 18, 2015.

[99] Edwina Portocarrero. Networked Playscapes: Redefining the Playground. *Dissertation, Massachusetts Institute of Technology, School of Architecture and Planning, Program in Media Arts and Sciences*, June 2018.

[100] Edwina Portocarrero and Gershon Dublon. El bosque de los murmullos (the whispering woods). Centro Nacional de Las Artes (National Center for the Arts), Mexico City, Nov 2014.

[101] Ivan Poupyrev, Nan-Wei Gong, Shiho Fukuhara, Mustafa E. Karagozler, Carsten Schwesig, and Karen Robinson. Project Jacquard: Manufacturing Digital Textiles at Scale. In *ACM CHI Conference on Human Factors in Computing Systems*, 2016.

[102] Zachary Pousman and John Stasko. A taxonomy of ambient information systems. In *Proceedings of the Conference on Advanced Visual Interfaces (AVI '06)*, page 67, 2006.

[103] Daniele Puccinelli and Martin Haenggi. Wireless sensor networks: Applications and challenges of ubiquitous sensing. *IEEE Circuits and Systems Magazine*, 5(3):19–29, 2005.

[104] Raffi Adlan, Georges Bermann, Julie Fong, and Michel Gondry (Producers), and Michel Gondry (director). Is the Man Who Is Tall Happy? *France: Partizan Films*, 2013.

[105] Vasant Ramasubramanian. Quadrasense: Immersive UAV-based Cross-Reality Environmental Sensor Networks. *Master's Thesis, Massachusetts Institute of Technology, School of Architecture and Planning, Program in Media Arts and Sciences*, 2015.

[106] David B. Ramsay and Joseph A. Paradiso. Automated characterization of consumer-grade sensor accuracy from supporting data in heterogeneous air quality monitoring networks. In *National Environmental Monitoring Conference*, Washington, DC, June 2017.

[107] Sara Remsen. *Visualizing Models Driven by Real-Time, Sensor-Based Data: A New Way to Experience the Inner Workings of Ecosystems.* Master of science, Massachusetts Institute of Technology, 2017.

[108] Bradley Rhodes. A brief history of wearable computing.

[109] Robert Richer. Exploring Interaction Concepts for a Context-aware Smart Office Prototype. *Master's Thesis, Friedrich-Alexander-Universitat Erlangen-Nurnberg*, 2017.

[110] Philip W. Rundel, Eric A. Graham, Michael F. Allen, Jason C. Fisher, and Thomas C. Harmon. Environmental sensor networks in ecological research. *New Phytologist*, 182(3):589–607, 2009.

[111] Spencer Russell, Gershon Dublon, and Joseph A. Paradiso. HearThere: Networked Sensory Prosthetics Through Auditory Augmented Reality. In *ACM Augmented Human International Conference.* ACM, 2016.

[112] Spencer Russell and Joseph A Paradiso. Hypermedia APIs for sensor data: a pragmatic approach to the web of things. *11th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, pages 30–39, December 2014.

[113] Spencer F. Russell. HearThere: Infrastructure for Ubiquitous Augmented-Reality Audio. *Thesis: S.M., Massachusetts Institute of Technology, Program in Media Arts and Sciences.*, 2015.

[114] Ryuichi Sakamoto. Forest symphony. `http://forestsymphony.ycam.jp/`, July 2013. Accessed: 2018-03-01.

[115] J. Sandvad. Dynamic Aspects of Auditory Virtual Environments. *100th AES Convention*, 4226:Convention Paper 4226, 1996.

[116] Nitin Sawhney and Chris Schmandt. Nomadic radio: speech and audio interaction for contextual messaging in nomadic environments. *ACM Transactions on Computer-Human Interaction*, 7(3):353–383, 2000.

[117] Frost & Sullivan. Global Wireless Sensor Networks Market (N92B-32). Technical report, Frost & Sullivan, 2016.

[118] Robert Szewczyk, Alan Mainwaring, Joseph Polastre, John Anderson, and David Culler. An analysis of a large scale habitat monitoring application. In *Proceedings of the 2nd international conference on Embedded networked sensor systems SenSys 04*, volume 2, pages 214–226, 2004.

[119] Robert; Szewczyk, Eric; Osterweil, Joseph; Polastre, Michael; Hamilton, Alan; Mainwaring, and Deb Estrin. Habitat Monitoring with Sensor Networks. *Communications of the ACM: Wireless sensor networks*, 47(6):34 – 40, 2004.

[120] Desney S Tan and Anton Nijholt. *Brain-Computer Interfaces: Applying Our Minds to Human-Computer Interaction*. Springer, January 2010.

[121] Carolyn M. Tennessen and Bernadine Cimprich. Views to nature: Effects on attention. *Journal of Environmental Psychology*, 15(1):77–85, 1995.

[122] Eric E Thomson, Rafael Carra, and Miguel A L Nicolelis. Perceiving invisible light through a somatosensory cortical prosthesis. *Nature Communications*, 4, February 2013.

[123] Willard R. Thurlow and Philip 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.

[124] D.W.F. van Krevelen and R. Poelman. A Survey of Augmented Reality Technologies, Applications and Limittions. *The International Journal of Virtual Reality*, 9(2):1–20, 2010.

[125] Katia Vega, Nan Jiang, Xin Liu, Viirj Kan, Nick Barry, Pattie Maes, Ali Yetisen, and Joe Paradiso. The Dermal Abyss : Interfacing with the Skin by Tattooing Biosensors. *Proceedings of the 2017 ACM International Symposium on Wearable Computers*, pages 138–145, 2017.

[126] Roel Vertegaal, Jeffrey S Shell, Daniel Chen, and Aadil Mamuji. Designing for augmented attention: Towards a framework for attentive user interfaces. *Computers in Human Behavior*, 22(4):771–789, July 2006.

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

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

[129] Carl J. Watras, Michael Morrow, Ken Morrison, Sean Scannell, Steve Yaziciaglu, Jordan S. Read, Yu Hen Hu, Paul C. Hanson, and Tim Kratz. Evaluation of wireless sensor networks (WSNs) for remote wetland monitoring: Design and initial results. *Environmental Monitoring and Assessment*, 186(2):919–934, 2014.

[130] M Weiser and J S Brown. The coming age of calm technology. *Beyond calculation*, 1997.

[131] Mark Weiser. The computer for the 21st century. *Mobile Computing and Communications Review*, 3(3):3–11, 1991.

[132] Mark Weiser. The computer for the 21 st century. *ACM Mobile Computing and Communications Review (SIGMOBILE)*, 3(September):3–11, 1999.

[133] Yu Xuan Zhang, Johanna G. Barry, David R. Moore, and Sygal Amitay. A New Test of Attention in Listening (TAIL) Predicts Auditory Performance. *PLoS ONE*, 7(12):1–12, 2012.

[134] Nan Zhao. *Mediated Atmospheres: Context-aware Adaptive Lighting and Multimodal Media Environments*. Doctor of philosophy, Massachusetts Institute of Technology, 2017.

[135] Nan Zhao. Mediated Atmospheres: Context-aware Adaptive Lighting and Multimodal Media Environments. *Dissertation, Massachusetts Institute of Technology, School of Architecture and Planning, Program in Media Arts and Sciences*, 2018.

[136] Nan Zhao, Gershon Dublon, Nicholas Gillian, Artem Dementyev, and Joseph A. Paradiso. EMI Spy: Harnessing electromagnetic interference for low-cost, rapid prototyping of proxemic interaction. In *IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, 2015.