Towards Perceptual Augmentation

by

Sam Chin

B.A. Wellesley College (2017)

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
in partial fulfillment of the requirements for the degree of

Master of Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2024

Author____
Sam Chin
Program in Media Arts and Sciences
July 19th, 2024

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

Accepted by____
Joseph A. Paradiso
Alexander W. Dreyfoos (1954) Professor of Media Arts and Sciences
Program in Media Arts and Sciences

Towards Perceptual Augmentation

by

Sam Chin

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
on July 19th, 2024, in partial fulfillment of the
requirements for the degree of
Master of Science

Abstract

This thesis explores the concept of perceptual augmentation, focusing on expanding human sensory capabilities beyond their biological limitations. It challenges traditional approaches to sensory enhancement by emphasizing the importance of perception over mere sensory input. Drawing inspiration from the diverse sensory abilities found in nature, the research aims to develop methods for meaningful augmentation of human perception that can impact daily life. The study adopts an ecological approach to perceptual augmentation, grounded in Gibsonian ecological psychology. Key principles include providing correct mental models of augmentation devices, leveraging environmental training and natural tasks, emphasizing multisensory interfaces with sensorimotor feedback, and creating affordances that mimic the natural world. This approach seeks to facilitate perceptual learning through natural interaction with the environment, rather than relying on extensive explicit training. The thesis presents early work in exploring and evaluating individual principles of this ecological framework for perceptual augmentation. While acknowledging the gap between the proposed theoretical approach and current research outcomes, the studies conducted focus on augmenting perception for specific tasks such as pitch interval perception, pilot situation awareness, and sleep staging. The research does not yet demonstrate a generalized, "all-purpose" augmented sense, but lays groundwork for future investigations, including a proposed experiment to mitigate age-related hearing loss using the developed principles.

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

Towards Perceptual Augmentation

by

Sam Chin

The following people served as readers for this thesis:

Thesis Reader____

Pattie Maes
Germeshausen Professor of Media Arts and Sciences
Program in Media Arts and Sciences

Thesis Reader____

Daniel Polley
Director of Eaton-Peabody Laboratories at Mass Eye and Ear
Harvard Medical School

Contents

List of Figures

List of Tables

Acknowledgements

I would first like to thank my readers: Joe Paradiso, Pattie Maes, and Dan Polley.

Joe, thank you for allowing me to join Responsive Environments. I'm so grateful for the opportunity to explore perceptual augmentation in such depth. You've pushed me to be more creative and bolder and I'm excited for what's to come.

Pattie, you've encouraged me toward practicality and kept me grounded. I've learned so much from your group meetings and am grateful you included me in your classes to present on sensory augmentation.

Dan, I am grateful for your emphasis on scientific rigor. Your incisive comments have really pushed me to improve my clarity and precision. I admire your drive to improve the lives of your patients in tangible ways.

Chapter 0 Ed Yong provided inspiration for many of the ideas in this thesis in his book *An Immense World*. I am grateful for this peek into the umwelt of so many other creatures.

Chapter 1 There were many sources that I used while preparing this chapter, but I would like to acknowledge two in particular that undergird my understanding of perception: SHBT 205 (Audition - Neural Mechanisms, Perception, and Cognition) and *Sensation and Perception ed. 11*. In addition to providing knowledge, *Sensation and Perception ed. 11* provides this thesis with many images that are (gratefully) used with permission.

I would also like to thank Fernanda, Ajani, and Ian, members of the McDermott lab, for many conversations about experiment design and perception overall.

Additionally, I would like to thank Manuel for our discussions on experimental design on evaluating perception. These conversations have inspired me and your questions have pushed me toward rigor and specificity.

Chapter 4 I am eternally grateful for my coauthors: Nikhil and Cathy. Cathy, I have loved exploring perception, haptics, art, and more with you. From the haptic actuators you gave me to our conversations about learning, I feel we have both explored in parallel and grown together. I'm inspired by your creativity and your cleverness and I am glad to have you as a confidant and collaborator.

Nikhil, your friendship, mentorship, and advice have shaped so many of my priors. Our conversations about Gibson, perception, and Molyneux's question have shaped much of my thinking on this thesis. Also, thank you for getting me to use R. I also want to thank Rose for consulting on pitch intervals and for providing experiential knowledge from teaching ear training.

Furthermore, I'd like to thank Abhi for teaching Extended Cognition and for providing me with suggestions for papers and ideas that I hadn't considered. Your feedback has been invaluable, and your jokes have provided much-needed levity.

Chapter 5 I'd like to thank my coauthor Nathan WW for working with me for over a year to take this from a discussion at tea to a real paper. I'm encouraged by your commitment to academic honesty and rigor as well as your pursuit of knowledge through self-experimentation. Getting to explore some of my ideas with an expert sleep staging has been illuminating.

Chapter 3 I want to thank my coauthors, Noam and Brittany, for sharing my excitement about interfaces for aviation (and for keeping the aero in AeroAstro). Noam, your insight about the human factors of aviation was invaluable and inspirational. I would also like to thank members of the Human Systems Lab, Andy, Chuck, and Katya, for their advice and for hosting an excellent journal club.

Additionally, I would like to thank Nathan P for constructing sensemux, the underlying software that enabled this project (and many of my early experiments). Our initial discussions about sensory augmentation on Slack jumpstared my academic pursuit of this topic. So many of the ideas in this thesis were developed in conversation and experimentation with you. It has been a joy to follow curiosity together. Thank you for all the exploration, building, and debugging

Additional Acknowledgments

Rob, Tobin, Kush, and Naana - thank you for being down for adventures and tomfoolery. You've made my time at lab much weirder and much brighter.

Manuel, Nikhil, Kimy, Jess, Jessie, and Ana, thank you for welcoming me into your space and treating me like an honorary member of Opera. I love how our discussions can range from auditory neuroscience to Geoguesser. This thesis truly would not have been possible without this generosity.

Patrick, Nathan, Perry, Brian, Mark, Lancelot, Cedric, Irmandy, and Xan - you are a group of people who are not only incredibly clever but also kind. When someone comes to our group looking for assistance, they are inundated with solutions. I feel lucky to be surrounded by such excellent humans.

I want to thank my mom and sister for their love, support, and constant supply of snacks.

Finally, I'd like to thank Udayan. Sometimes you're my biggest critic, but you also encourage me to go beyond what I think I can do. I think I am the best version of myself when I'm with you. I'm so grateful for the life we have with Sesame and Jupiter - thank you for helping make all of this possible.

Chapter 0

Preamble

Light, the visible reminder of invisible light

T.S. Eliot

0.1 Introduction

The light we see is a small part of the electromagnetic spectrum. Most forms of electromagnetic radiation, from radio waves to gamma rays, are invisible to humans. Visible light serves as a tangible reminder of all the invisible forms of energy that exist beyond our direct perception. This limitation of human perception extends beyond vision. All of our senses have biologically defined boundaries, outside of which an abundance of sensory information exists.

For instance, the black grouse notices a ripe ultraviolet berry the same way we notice a ripe red strawberry. A sea turtle navigates using the Earth's magnetic field like a map to return to the beach where she was born. The platypus hunts blind in dark, murky waters by detecting minute changes in electric fields generated by his prey. These examples from the animal kingdom are reminders that our biology shapes our experience of reality. Our senses do not provide an objective picture of the world around us. Instead, we sample a fraction of the available sensory information, constructing our reality from this limited input.

As we age, this sampling of reality becomes even more restricted. Vision begins to blur, and sounds become more difficult to hear. Such sensory losses have additional harmful effects on our mental and emotional well-being. For example, in this thesis, I discuss hearing loss, which is associated with depression [72] and dementia [50]. The interventions to mitigate hearing loss are limited. Hearing aids can improve hearing levels but do not address frequency losses. Cochlear implants, while effective, require surgical intervention. There is a lack of noninvasive approaches that address both level and frequency losses.

Although seemingly unrelated, assistive technologies that mitigate sensory loss and augmentation technologies to expand our sensory abilities are the same problem at the individual level. Both seek to transmit sensory information that is otherwise unavailable. In this thesis, I pragmatically focus on the perceptual augmentation of our existing senses. However, I also believe focusing on methods for our existing senses is, paradoxically, the most effective way to develop augmented senses. We first must have effective retraining methods for our existing senses before we can effectively work on developing augmented senses.

We accept the deterioration of our senses—the blurring of vision, the dulling of sounds—as an inevitable part of aging. But what if we could expand our sensory capabilities instead of accepting their decline? What if we could maintain our perception and enhance it, venturing into realms of awareness previously reserved for other species and sophisticated instruments? An abundance of sensory information exists beyond the boundaries of human biology. This thesis is the beginning of an exploration of methods and technologies for going beyond these boundaries.

0.2 Defining Perceptual Augmentation

Since my aim is to provide sensory signals to ultimately improve perception, I call my work perceptual augmentation. However, much of my work could also be called sensory augmentation or sensory substitution. In this thesis, I use the term perceptual augmentation because much of the discussion on sensory augmentation focuses on the process of encoding and transmitting the raw sensory data. Sensing is the process of getting raw sensory information. In contrast, perception is combining sensory information with context, memory, and expectations to form understanding. Without perception, sensory signals are noise.

Many approaches to sensory augmentation have primarily emphasized the sensory aspect, often neglecting the crucial role of perception. One of the most prominent examples is the "Mr. Potato Head" model of learning, which asserts that "it doesn't matter how you get information in there, the brain will figure out what to do with it" [36]. This model is incomplete and neglects the role of attention, memory, context, and other priors in transforming raw sensory data into a percept.

The most successful examples of perceptual augmentation are sensory substitution devices for accessibility. In contrast, sensory augmentation devices that attempt to give humans a sense beyond our physical abilities have more mixed results. Often, sensory augmentation work on these novel senses focuses on designing and engineering to convey physical phenomena with limited user studies. Demonstrating perceptual augmentation in an academic environment subject to a limited timeframe can be challenging, a factor that I have struggled with and attempted to address in this thesis.

Focusing on perception instead of only sensation is one way to improve outcomes when building assistive and augmentation technologies. As discussed earlier, both are trying to solve the same problem. Both seek to enable perception of something the person is physically unable to sense. In that case, what determines if something is an assistive or augmentation technology? The environment.

For example, consider me and a person with hearing loss. A person with hearing loss cannot hear specific frequencies, just as I cannot see ultraviolet (UV) light. However, hearing loss has significant adverse effects, while not seeing UV has no effect. This difference in outcomes is because what matters is not absolute sensory ability but sensory ability relative to the environment. In a deaf community, a person with hearing loss does not experience challenges related to their hearing loss because their environment does not assume hearing ability. Many in the deaf community prefer to call deafness a "difference" instead of a "loss" or "impairment" [105]. This further highlights how the user's or entity's environment is central to perception.

0.2.1 An Ecological Approach to Perceptual Augmentation

This focus on the environment in my thinking has led me to adopt principles from Gibsonian ecological psychology[47], which emphasizes the role of the environment in shaping perception and action. I use it as a lens through which I motivate my experiments. Suppose we believe Gibson that perception stems from active sensorimotor exploration. In that case, it then follows that the objective of perceptual augmentation should be to mimic the natural sensory acquisition experience, where learning occurs through interaction with the environment rather than through hours of explicit training. To achieve this goal, I consider several fundamental principles:

Principles of Ecological Perceptual Augmentation

1. Providing the correct mental model of the augmentation device in the environment
2. Leveraging environmental training and natural tasks to minimize explicit training sessions.
3. Emphasizing multisensory interfaces with sensorimotor feedback to create rich, perceptual experiences
4. Creating affordances that mimic the natural world, which are subtle without encoding or discretization.

There is a significant gap between the theoretical approach advocated here and what I have demonstrated in my research. This thesis presents early work exploring a framework for an ecological model of perceptual augmentation. The studies I present are focused on evaluating the validity of individual principles of the framework but do not evaluate it as a whole. Much of this framework was formed through challenges I encountered while conducting the studies shared in this thesis. The studies presented study perceptual augmentation in specific contexts, such as while recovering from unusual flight attitudes in aircraft" or performing sleep staging. However, this thesis aims to explore augmenting perception independent of context. In the final chapter, I propose a device and experiment aimed at a more general purpose - for mitigating age-related hearing loss in Chapter 3 that utilizes the framework I suggest in the thesis.

Chapter 1

Mechanisms of Human Sensing and Perception

1.1 Introduction

The human sensory system is a complex network of specialized organs and neural pathways that allow us to understand and interact with our environment. To ground the experiments to come, I provide a brief overview of the neurological mechanisms behind human sensing and perception as they relate to perceptual augmentation.

We begin by discussing multisensory integration and the brain's ability to combine information from various sensory modalities into a unified perceptual experience. I present concepts relevant to all perception and explain why multisensory interfaces are beneficial when augmenting perception.

Then, we delve into the physical systems and neural mechanisms underlying haptic perception, including the various types of mechanoreceptors in the skin and their roles in detecting different tactile stimuli. In this thesis, we then review other ways of measuring and evaluating haptic perception to inform the haptic devices.

Next, we move to auditory perception, discussing the biomechanical and neural mechanisms underlying hearing. This thesis examines audition from dual perspectives: as a complement to visual perception and as a modality enhanced through haptic feedback. Thus, we explore how these processes impact perceptual augmentation, highlighting their opportunities and constraints for multisensory enhancement.

We subsequently investigate neuroplasticity and perceptual learning. We consider Gibson's theory of ecological perception and discuss how many of the themes of context and action present in his theories appear in this thesis. We then investigate particular studies on neuroplasticity and observe how these can be applied to perceptual augmentation devices.

1.2 Perceptual Processes

Perception is a complex cognitive process that goes far beyond mere sensory input. This section explores three key aspects of perceptual processes: Unconscious Inference, Sensory Integration, and Perceptual Ambiguity. These interconnected phenomena demonstrate how our brains actively interpret and construct our experience of reality, often without our conscious awareness. Understanding these processes is crucial for grounding our work in designing perceptual interfaces and augmentation technologies.

1.2.1 Unconscious Inference

The brain integrates information from multiple sensory modalities (e.g., vision, audition, haptic) and prior knowledge to form a unified percept of the world through multisensory integration. For instance, we combine lip reading and sentence context in noisy environments to comprehend speech. Combining visual, auditory, and syntactic information into a more reliable perception is an example of unconscious inference.

Due to the inherent ambiguity of our senses, all our perceptions stem from these unconscious inferences. These inferences range from simple associations to complex processes like physics simulations (Battaglia et al. (2013) provide evidence that subconscious, intuitive physics simulations enable us to assess the stability of stacked objects or predict object movements [112, 9]).

1.2.2 Multisensory Integration

In addition to creating inferences from prior experiences, our subconscious dynamically selects which sensory streams to include in a percept. The brain combines sensory inputs to minimize variance and maximize the reliability of the integrated percept [37]. Having a multisensory stream of information allows each sense to function where it is most effective. For example, audition is more temporally precise, while vision is more spatially precise [119]. These spatial and temporal resolution differences may make a particular sensory modality preferable for a given signal. The brain can select the sensory modality with the most information for a given signal by presenting both an auditory and visual signal.

Studies have shown that multisensory integration is statistically optimal [119, 37] and that his optimal integration allows for more accurate perception than possible from any single sensory modality. This is also confirmed by experimental results, which have demonstrated that sound can enhance visual perception [116, 68].

Successful multisensory integration requires temporal, spatial, or semantic similarity of stimuli [68]. Temporal synchrony refers to the need for multiple sensory inputs to occur within a time window to be perceived as originating from the same event. Spatial congruence, however, requires that sensory stimuli originate from approximately the exact location in space to be integrated effectively. Semantic similarity refers to correspondence or congruence of content. For example, a video of a cat meowing paired with audio of a barking dog may not integrate [35] even if they were temporally matched. These principles ensure that the brain combines related sensory information while avoiding integrating unrelated stimuli from different events or sources. This topic of multisensory integration for perceptual interface design will be further explored in Chapter 2.

1.2.3 Perceptual Ambiguity and Sensory Illusions

Sometimes, our brains make the wrong unconscious inference based on incomplete information, resulting in a sensory illusion. These phenomena occur when our perception diverges from physical reality, often due to misinterpretation of contextual cues or insufficient sensory information. Studying the mechanisms of sensory illusions provides a unique window into the mechanisms of perception we can utilize when designing devices and experiments.

The McGurk effect, for example, demonstrates the intricate interplay between visual and auditory processing in speech perception. In this illusion, conflicting visual and auditory phoneme information leads to the perception of a third, distinct phoneme [81]. Interestingly, this effect diminishes when the phonemes are presented within the context of a sentence, suggesting the influence of higher-level linguistic processing. Another compelling example is the ventriloquist effect, where visual stimuli can alter the perceived location of auditory stimuli [20]. This illusion underscores the brain's tendency to integrate multisensory information, even when doing so leads to perceptual inaccuracies.

Ill-Posed Problems and Perceptual Ambiguity

A useful framework for understanding these perceptual phenomena is through the lens of ill-posed and well-posed problems from mathematics. An ill-posed problem has multiple equally valid solutions or interpretations. For instance, the equation $\mathrm{x} + \mathrm{y} = 10$ is ill-posed since multiple combinations of x and y could satisfy this condition. Similarly, in perceptually ambiguous situations, there could be multiple valid causes for the observed stimulus [14]. Well-posed problems, in contrast, have unique and stable solutions. In perception, a well-posed problem would be one where the sensory input unambiguously corresponds to a single interpretation of the physical world.

Many sensory illusions are ill-posed problems caused by the physical limitations of our sensory organs. For example, auditory front-back confusion is an illusion where listeners may perceive sounds originating from in front of them as coming from behind, or vice versa [48]. The root cause of this illusion lies in the symmetrical placement of our ears on either side of the head. This arrangement creates an ambiguity in sound localization, particularly for pure tones (consisting of a single frequency). Even under ideal listening conditions, the auditory system lacks sufficient information to definitively determine whether a pure tone is coming from the front or the back [17] and is thus an ill-posed problem.

Visual illusions, such as the Müller-Lyer illusion (Figure 1-1), can also be viewed as an ill-

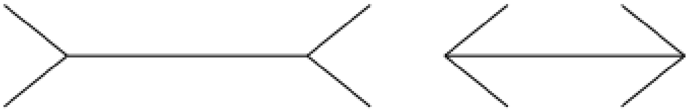

Figure 1-1: Müller-Lyer illusion. In this illusion, two lines of equal length appear different due to arrow-like figures pointing inward or outward at their ends. These illusions persist even when we know their nature, highlighting the automatic and unconscious nature of perceptual processes [48]. Image from Wikipedia under Creative Commons License.

posed problem arising from physical sensory limitations. The primary hypothesis for why the lines appear to be of different lengths is that the visual system incorrectly interprets 2D patterns based on ingrained 3D assumptions created by the arrows [48]. While this seems to be primarily a problem with the visual system's default processing assumptions, it is fundamentally a sensory limitation problem. Our visual system processes a 2D projection of a 3D world, requiring us to infer depth based on shadows and angles. This projection creates an ill-posed problem because there are multiple valid interpretations: a 2D image of lines and arrows or a 3D image depicting an object with depth. The image our eyes see is fundamentally a 2D projection of a 3D world, so it is difficult to resolve the cause of ambiguity. However, even in this case, the ill-posed problem framework is useful to analyze if it is even reasonable to augment a sense.

This thesis will employ the framework of well-posed and ill-posed problems as a metric for evaluating the effectiveness of perceptual augmentation devices. In perception, a well-posed problem is one where the sensory input unambiguously corresponds to a single interpretation of the physical world. Conversely, an ill-posed problem presents ambiguous sensory information that could result in multiple valid interpretations.

By applying this framework, we can assess how perceptual augmentation devices transform ill-posed perceptual problems into well-posed ones. Effective devices should reduce ambiguity in sensory input, minimizing the likelihood of misinterpretations and illusions. This approach provides a quantifiable means of measuring the success of these devices in enhancing human perception and reducing perceptual errors.

1.3 Haptic Perception

Haptic perception refers to the ability to acquire information through active tactile exploration. This perceptual modality is distinguished from passive tactile perception by its reliance on purposeful movement and interaction with surfaces and objects [46]. In haptic perception, the subject actively engages with the environment, utilizing cutaneous sensations and proprioceptive feedback to construct a comprehensive understanding of spatial and material properties [70]. This dynamic process involves the integration of sensory inputs from mechanoreceptors in the skin, muscles, and joints, enabling a more robust and detailed perception than static touch alone. While "haptic" can encompass various sensations, including touch, pressure, and weight, this section will concentrate on the mechanisms relevant to vibrotactile perception.

1.3.1 Haptic Mechanoreceptors

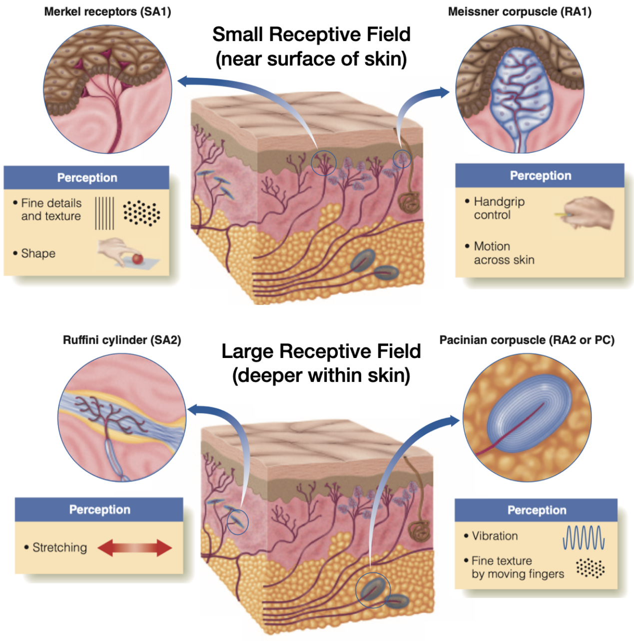

Figure 1-2: Four types of haptic mechanoreceptors. Merkel receptors and Meissner corpuscles are located close to the skin's surface and responsible for a small receptive field. In contrast, Ruffini cylinders and Pacinian corpuscules are located lower in the skin and are in a larger receptive field. Adapted from *Sensation and Perception ed. 11* [48]. Used with permission.

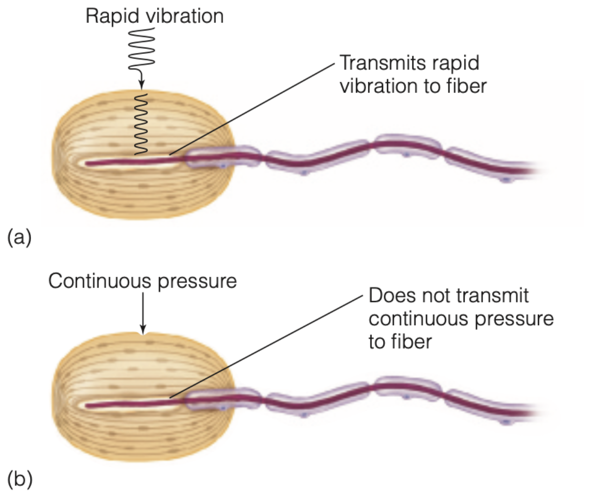

Figure 1-3: The layered structure of a Pacinian corpuscle acts as a mechanical high-pass filter. They efficiently transmit high-frequency pressure changes (vibrations) to the enclosed nerve fiber while attenuating sustained pressure. Adapted from *Sensation and Perception ed. 11.* [48] Used with permission.

Mechanoreceptors in the skin detect and transmit haptic information to the central nervous system. There are four types of mechanoreceptors: Merkel receptors, Ruffini cylinders, Meissner corpuscles, and Pacinian corpuscles.

Two mechanoreceptors are particularly relevant for vibrotactile stimuli: Meissner corpuscles and Pacinian corpuscles [60]. These two are considered fast-acting (FA) or rapid-acting (RA). Meissner corpuscles (RA I) respond to low-frequency vibrations (5-40 Hz), while Pacinian corpuscles (RA II) are sensitive to higher frequencies (40-400 Hz) [70]. The Pacinian corpuscle's selective responsiveness to rapid vibrations is attributed to its layered, onion-like structure. These layers, separated by fluid, act as a mechanical filter. They efficiently transmit high-frequency pressure changes (vibrations) to the enclosed nerve fiber while attenuating sustained pressure (Figure 1-3).

1.3.2 Haptic Processing in the Somatosensory Cortex

The primary somatosensory cortex (S1) and the secondary somatosensory cortex (S2) process tactile and haptic information. The somatosensory cortex is spatially organized: specific regions in the brain correspond to different parts of the body. This organization was mapped by Wilder Penfield (with collaborators Edwin Boldrey and Theodore Rasmussen), who stimulated various points on the cortex of awake patients undergoing epilepsy surgery. Their work revealed that the somatosensory cortex is not uniformly represented; instead, it exhibits a disproportionate allocation of cortical space to body parts with a high density of sensory receptors. This distorted representation is often depicted as a "sensory homunculus" (Figure 1-4).

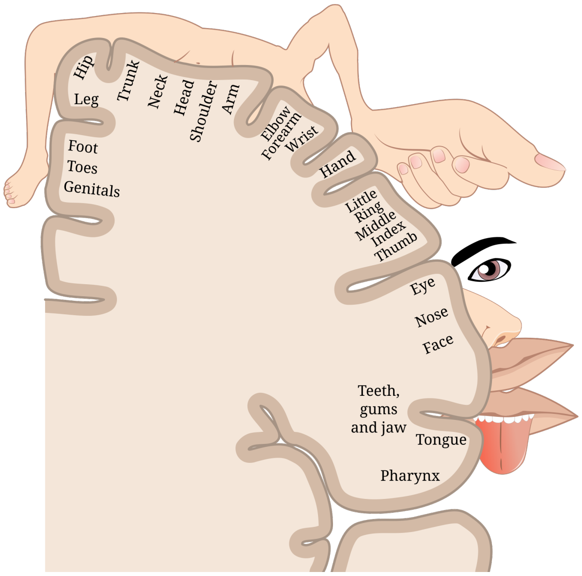

Figure 1-4: The sensory homunculus was developed based on work by Penfield et al., which mapped the primary somatosensory cortex (S1). Larger areas of the cortex represent parts of the body with higher sensory nerve density. [89] image from Wikipedia under Creative Commons License.

The processing of vibrotactile information in the brain involves both spatial and temporal components, as described by the duplex theory of texture perception [48]. This theory suggests that our perception of texture, including vibrotactile sensations, relies on two types of cues: **Spatial cues**: larger surface elements that can be felt through static touch or movement, which convey information about shape, size, and distribution of surface features. **Temporal cues**: vibrations generated when the skin moves across a textured surface.

1.3.3 Evaluating Haptic Perceptual Acuity

The capacity of mechanoreceptors determines haptic perceptual acuity and can vary depending on the body location and type of skin (glabrous, hairy). There are various psychophysical measures to measure different aspects of tactile and haptic perception, including: **Two-point discrimination**: The minimum distance at which two distinct points of stimulation can be perceived [48, 70], **Grating Acuity**: The narrowest spacing which orientation is accurately judged [48], **Frequency discrimination**: The ability to distinguish between different vibration frequencies [79, 16]. **Intensity discrimination**: The ability to detect differences in vibration strength [93].

As noted earlier in the chapter, these psychophysical measures do not directly match up with perception. In Chapter 2, we will further discuss methods of evaluating haptic perceptual acuity.

1.4 Auditory Perception

Audition, the sense of hearing, plays a crucial role in human perception and interaction with the environment and offers several unique advantages as a sensory modality. Perhaps most notably, it allows for sensing at a distance, enabling the detection and localization of objects and events beyond our field of view[17]. Moreover, auditory processing is remarkably fast, with neural responses to auditory stimuli occurring more rapidly than those to visual stimuli [68]. The human auditory system can detect a wide range of frequencies, typically 20 Hz to 20 kHz in healthy young adults [48].

1.4.1 The Mechanics of Audition

Audition is the perception of vibrations at particular frequencies. The mechanical process of hearing begins when the pinnae collects sound waves and channels them into the ear canal. These sound waves cause the eardrum (tympanic membrane) to vibrate. These vibrations are transmitted through the malleus, incus, and stapes, which act as mechanical amplifiers [48]. The stapes, in turn, transmit these vibrations to the fluid-filled cochlea in the inner ear, where hair cells are stimulated. (Figure 1-5) These hair cells play a crucial role in converting mechanical vibrations into electrical signals that the brain can interpret. The hair cells are arranged in a tonotopic manner along the cochlea's basilar membrane. High frequencies elicit peak vibrations near the cochlear base, while low frequencies resonate maximally at

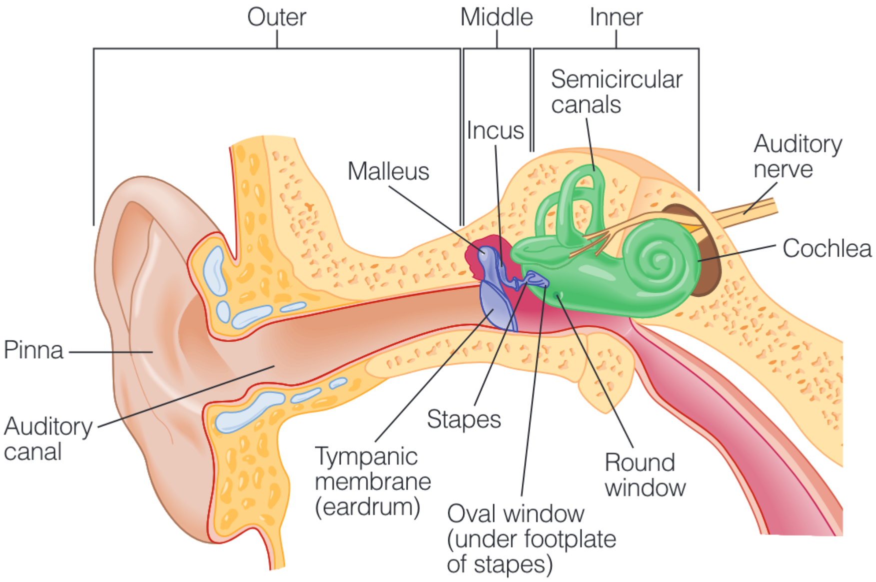

Figure 1-5: A diagram showing the ear structure going from the outer ear to the middle ear to the cochlea. From *Sensation and Perception ed. 11* [48]. Used with permission.

the apex. This spatial frequency discrimination activates hair cells differentially along the cochlea, effectively sorting auditory stimuli by frequency (Figure 1-7).

These delicate structures can be damaged or die due to exposure to loud noises and aging. Hair cell loss is a primary cause of sensorineural hearing loss, which is permanent as these cells do not regenerate in humans. Additionally, other forms of degradation, such as nerve damage or stereocilia damage, may also cause hearing loss. We will discuss these hair cells further in the section on sensorineural hearing loss and cochlear implants (CIs) below.

1.4.2 Processing in the Auditory Cortex

The auditory cortex, located in the temporal lobe, is organized in a tonotopic manner, mirroring the arrangement in the cochlea [98]. This tonotopic organization is preserved throughout the auditory pathway, from the cochlea to the primary auditory cortex, due to structural constraints of routing information through the auditory cortex [51]. Auditory cortex neurons select for many features of sound, such as frequency, amplitude (intensity), amplitude modulation (AM), frequency modulation (FM), and presentation rate. Neurons are primarily selective for frequency and are tonotopically clustered in the region of the

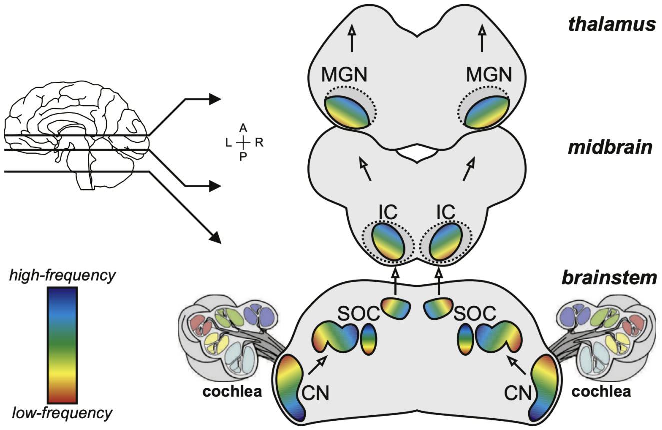

Figure 1-6: The central auditory system follows a hierarchical structure with tonotopic organization throughout its key components. This organization is present in the cochlear nucleus (CN), superior olivary complex (SOC), inferior colliculus (IC), and medial geniculate nucleus (MGN). Neuroimaging studies in humans consistently reveal at least two primary tonotopic gradients in the auditory cortex. These gradients follow a "high-to-low-to-high" frequency pattern. [98]. Used with permission via MIT Libraries.

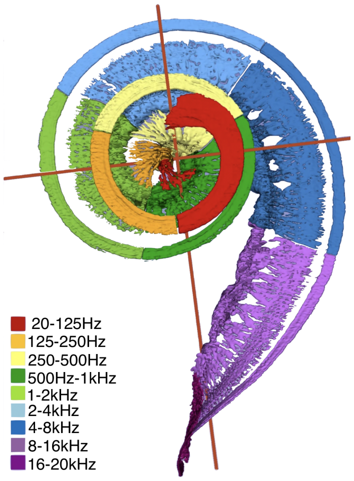

Figure 1-7: A color mapping of the human cochlea showing the tonotopic nature of the cochlea. Image adapted from Li et al. 2021 [73]. Used with permission via MIT Libraries.

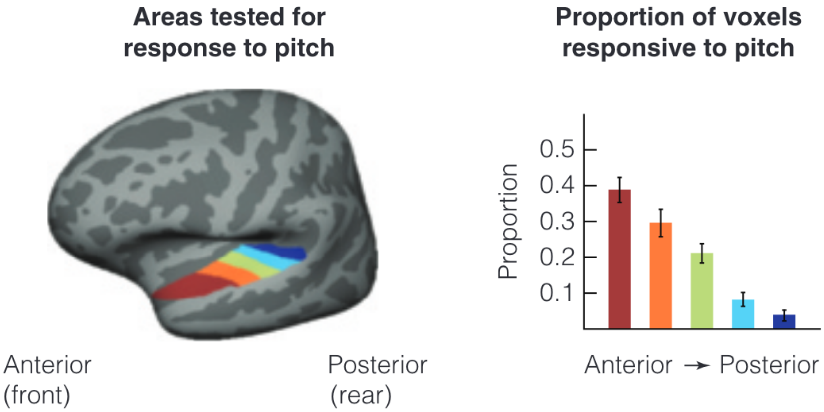

Figure 1-8: A diagram showing the relative proportion of voxels responsive to pitch in the auditory cortex measured by Norman-Haignere et al. (2013)[87]. Image from *Sensation and Perception ed. 11* [48]. Used with permission.

anterior auditory cortex [87]. Recent research has revealed that the auditory cortex is organized not just by frequency but also by more complex sound categories. Norman-Haignere et al. demonstrated that the human auditory cortex contains regions that respond selectively to specific sound categories such as speech and music [88]. This finding suggests a hierarchical organization of auditory processing, where some areas represent basic acoustic features while other areas represent more abstract sound categories.

1.4.3 Evaluating Auditory Perception

Two standard methods for assessing hearing ability are audiograms and psychophysical tests. Audiograms are graphical representations of an individual's hearing sensitivity across different frequencies [48, 17]. They provide a quantitative measure of hearing thresholds and are essential for diagnosing and monitoring hearing loss. Functional tests, such as speech audiometry, assess an individual's ability to understand speech in various listening conditions [66]. These tests provide valuable information about real-world hearing performance and can guide the selection and fitting of hearing aids or other assistive devices.

1.5 Neuroplasticity and Perceptual Learning

[115] Neural plasticity is the nervous system's ability to adapt its structure and function. This fundamental property underlies neural development, normal functioning, and the system's response to environmental changes, injury, or aging. The brain's capacity to reorganize itself by forming new neural connections encompasses a wide range of adaptive processes from acquiring new languages to relearning motor skills following neurological injuries such as strokes [48]. Neuroplasticity is more pronounced in children due to their developing brains; however, it persists throughout the lifespan, albeit to a lesser degree in adulthood. Plasticity is fundamental to sensory and perceptual augmentation, as it underpins the ability to learn and integrate novel sensory modalities.

1.5.1 Evidence of Maladaptive Neuroplasticity

Despite the growing body of knowledge on neuroplasticity, there remain significant gaps in our understanding. In this section, I review some examples of maladaptive neuroplasticity.

Such cases of maladaption present useful information for designing perceptual augmentation and a potential opportunity to use perceptual augmentation for sensory retraining.

Tinnitus as Maladaptive Neuroplasticity

Tinnitus is a persistent auditory perception without an external source that negatively impacts the quality of life for millions globally. It is often described as a ringing in the ear that will not go away. Tinnitus is often associated with presbycusis (age-related hearing loss) [101]. Recent research has shown these changes appear to be underpinned by maladaptive neural plasticity, resulting in increased spontaneous firing rates and synchrony among neurons in central auditory structures [101]. This results in the phantom percept.

Initial clinical research has demonstrated that bimodal (haptic, electric, etc.) [29] stimulation can reduce the symptoms of tinnitus, which provides an exciting application for perceptual augmentation devices.

Studying maladaptive systems, such as those observed in tinnitus and other perceptual disorders, provides valuable insights into the complex mechanisms of neuroplasticity and sensory processing. These investigations offer a unique window into the brain's adaptive capabilities and potential limitations. By examining instances where neuroplasticity leads to detrimental outcomes, we can better understand the underlying neural processes that govern adaptive and maladaptive changes.

Chapter 2

Sensory Substitution, Perceptual Augmentation

It will not help to try to imagine that one has very poor vision and perceives the surrounding world by a system of reflected high-frequency sound signals; [...] it tells me only what it would be like for me to behave as a bat behaves. But that is not the question.

Thomas Nagel, *"What is it like to be a bat?"*

2.1 Introduction

In this chapter, we will review prior art in sensory augmentation for accessibility and artistic/abstract experiences. Next, we will synthesize best practices for conveying information through auditory and haptic channels. Then, we will elaborate on the framework and the four principles: 1. Providing a correct mental model of the augmentation device in the environment. 2. leveraging environmental training and natural tasks to minimize explicit training sessions. 3. Emphasizing multisensory interfaces with sensorimotor feedback to create rich perceptual experiences. 4. Creating affordances that mimic the natural world are subtle without encoding or discretization.

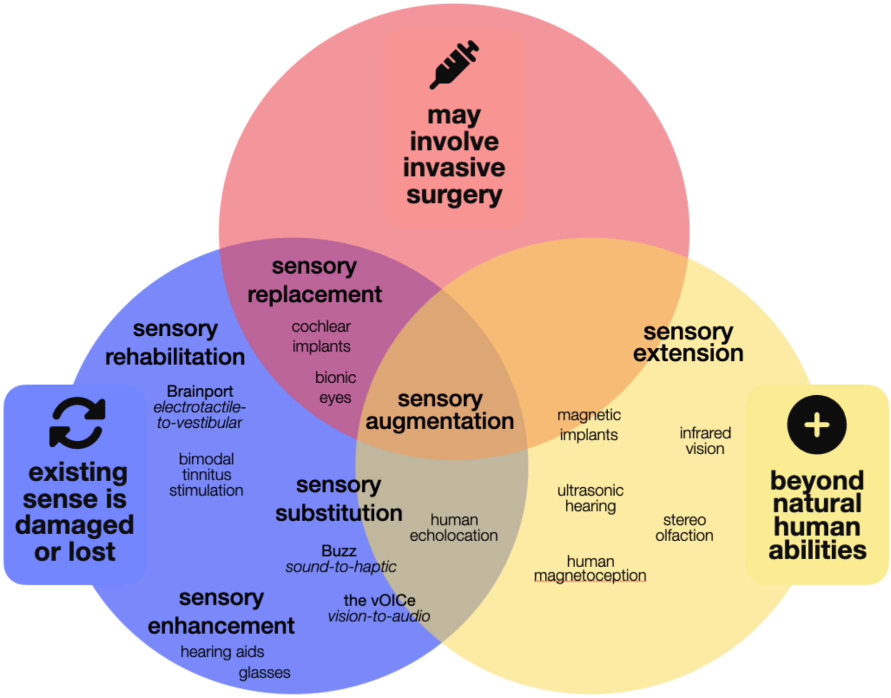

Figure 2-1: A Venn diagram depicting the different forms of sensory modification and how they relate to one another.

sectionDefining Perceptual Augmentation

2.1.1 Definitions

The field of sensory modification encompasses various terms, including sensory substitution, augmentation, extension, enhancement, and replacement. These systems universally comprise an input sensor and an output actuator, though their specific applications and contexts differ significantly. **Sensory substitution** systems typically operate in assistive technology contexts, where one sensory modality replaces another that has been lost. For instance, a device might transform visual input into auditory output for visually impaired users. Technically, any non-invasive device utilizing existing sensory modalities could be classified as sensory substitution, although this broad interpretation is not commonly employed in the literature.

Sensory enhancement devices, while similar in structure to substitution systems, are designed for partially impaired senses rather than completely lost ones. Hearing aids exemplify this category, amplifying and modifying auditory input to compensate for hearing deficits. **Sensory replacement** refers to invasive interventions that directly replace a lost sense, often interfacing directly with the nervous system. A cochlear implant is an example of a sensory replacement device. **Sensory augmentation** represents the broadest category, encompassing systems that detect signals beyond natural human sensory capabilities. Some researchers use the term **sensory extension** specifically to denote augmentation of existing senses beyond their typical range.

In my thesis, I use the phrase perceptual augmentation to mean something very specific: augmenting perception. As I mentioned in the Introduction, historically, the field of sensory augmentation has placed a disproportionate emphasis on the sensory components, often underestimating the pivotal role of perception. This narrow focus is particularly noticeable in work that focuses on a device. These approaches operate under the flawed assumption that merely introducing additional sensory data is sufficient to enhance perception. However, this perspective fails to account for the intricate feedback processes involved in learning and perceiving.

2.1.2 The Challenges of Sensory Augmentation

Since many of the first sensory augmentation devices were sensory substitution devices for replacing a "lost" sense, sensory substitution has proved multiple times that it is possible to remap "lost" senses to a new modality. Neosensory's Buzz sends sound through a vibrotactile bracelet such that deaf users feel like they can hear again. The Brainport sends sight through a tactile electrode array and enabled a blind man to climb Mt. Everest. Sensory substitution shows the brain's neuroplasticity and how it can adapt to new information. This idea of perception was built into the idea because most humans have a very clear idea about what hearing and vision are and thus can provide feedback to the person learning to use the sensory substitution device.

However, as we enter the realm of things people can't perceive naturally, with sensory augmentation, the importance of feedback becomes increasingly important. Perception is fundamentally a multisensory experience that connects sensory input with existing knowledge about the world. In some ways, multisensory interfaces have a closed feedback loop in which one sense can reinforce the other and bring multimodal attention to the stimulus.

2.1.3 Attention is Key for Perceptual Learning

Generally, the brain must pay attention to the stimulus to learn. Without attention, a sensory signal becomes noise because the brain cannot differentiate significant patterns from random stimuli. Generally, our ability to attend improves when we receive feedback and have a clear mental model about the stimulus [48]. Learning outcomes also improve when the tasks are sensorimotor [117, 45]. Our ability to interpret sensory information is not just a function of receiving stimuli but rather a complex process that integrates new inputs with existing sensory experiences and cognitive frameworks [44].

The attentional challenge of introducing new sensory modalities is perhaps best illustrated by the phenomenon of tetrachromacy. Typically, humans perceive colors using three types of cone cells in the retina: red, blue, and green. Individuals with red-green colorblindness have only two functional types of cones, limiting their color perception. In contrast, tetrachromats possess a genetic mutation that results in a fourth type of cone cell. Theoretically, this additional cone should allow tetrachromats to perceive a whole new dimension of colors, colors that are invisible to individuals with normal trichromatic vision [58].

From a neurobiological perspective, tetrachromacy represents the ultimate Brain-Computer Interface: a sensor in the eye directly connected to the brain. However, most individuals with tetrachromacy are functionally indistinguishable from trichromats in color discrimination tasks. They possess the "hardware” for enhanced color perception but fail to utilize this potential fully. Jordan and Mollon (2019) [61] reported that the only known functional tetrachromat is an artist, suggesting that mere possession of advanced sensory "hardware" is insufficient; active attention and appropriate context are crucial for developing enhanced sensory abilities.

This principle of attention also applies to more common perceptual learning in the auditory and visual domains. For example, piano tuners demonstrate increased perceptiveness to beat frequencies, accompanied by increased grey matter in associated brain regions[108]. Similarly, blind individuals with echolocation skills show improved sound localization abilities[114]. These examples all share a common thread: the individuals with above-average perceptual acuity had work that required them to attend to that percept for long periods of time. This leads to a critical point: mere exposure to stimuli is insufficient to develop new perceptual abilities. Instead, attention, active engagement, and targeted practice are key factors in expanding our perceptual horizons.

2.2 Sensory Substitution for Accessibility

This section reviews relevant prior art in the fields of sensory substitution and sensory augmentation. It is split into two parts: sensory substitution devices for accessibility and sensory augmentation devices for aesthetic experiences.

2.2.1 First Vision-to-Tactile Device (1969)

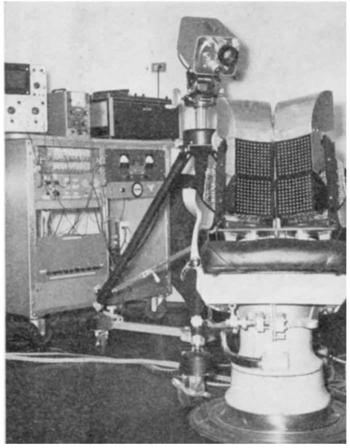

Figure 2-2: A photograph of the sensory substitution device that converts visual input from the camera into a tactile array. The tactile array was mounted to a dentist's chair and could only be used while seated [5] with permission via MIT Libraries.

Figure 2-3: An example of an image produced by Bach-y-Rita's tactile array depicting a face [5]. Used with permission via MIT Libraries.

Some of the first experimental evidence of human neuroplasticity was a visual-to-tactile sensory substitution device designed by Paul Bach-y-Rita in 1969 (Figure 2-2). This 40x40 tactile array system enabled blind individuals to differentiate between objects, recognize faces, and even discern the presence or absence of glasses (Figure 2-3)[5]. This marked the birth of sensory augmentation and demonstrated the brain's ability to adapt and interpret sensory information from unconventional sources.

2.3 Augmentation for Experiences or Abstract Understanding

Cherston and Paradiso (2017) used a similar approach of splitting complex information across sensory channels. Rotator was a web interface that allows users to switch data between the auditory and visual domains, however, this project was focused on analysis of high dimensional data instead of perceptual augmentation [23]. Similarly,[103] demonstrate a tool for interacting with and understanding high dimensional data, however, like Rotator, it is a tool for analysis rather than an interface for everyday use.

2.4 Prior Art in Perceptual Augmentation

In this section, I equate performance improvement with successful perceptual augmentation. There have been many nonvisual perceptual augmentation devices in the aeronautics and astronautics field since, frequently, pilots are overloaded. For instance, the Tactile Situation Awareness System (TSAS)[19] sends the orientation of a helicopter via a vibrotactile array. Ten helicopter pilots maintained a stable hover without visual cues using TSAS, which is otherwise impossible. Another example is Ueda Sakai et al. (2019)[111], who demonstrated that drivers can use a sonic representation of a car's position in a lane to navigate a visually occluded environment.

2.5 An Ecological Approach to Perceptual Augmentation

2.5.1 An Ecological Approach

J.J. Gibson (and later, E.J. Gibson) proposed an "ecological" approach to perception and perceptual learning. This approach focuses on the environment as the fundamental driver of perception and treats active (sensorimotor) exploration of the environment as the fundamental mechanism of perceptual learning [47, 45]. This thesis (somewhat incidentally) takes a Gibsonian approach and focuses on creating multisensory environments and sensorimotor-focused tasks.

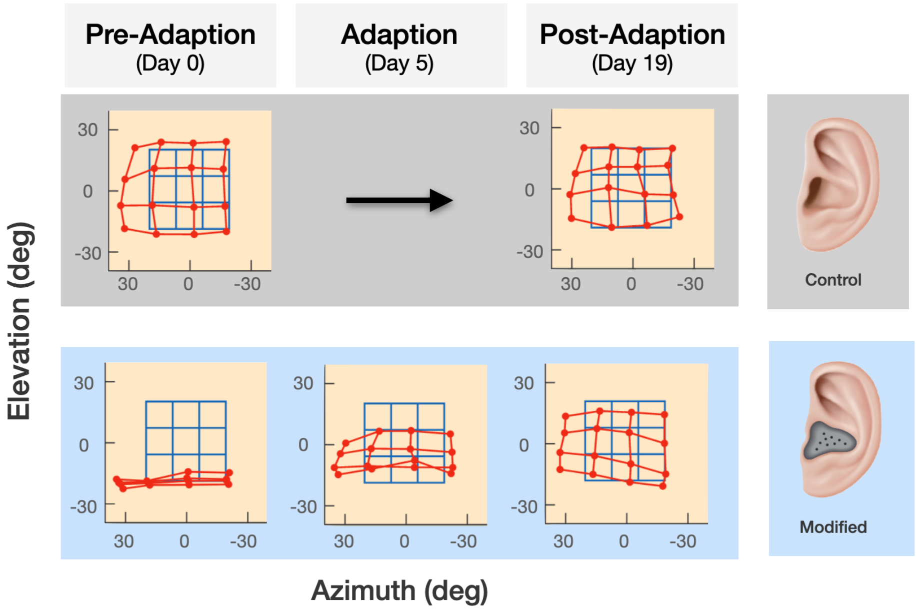

Figure 2-4: The grid in dark blue represents perfect localization with Elevation (degrees) on the y-axis and Azimuth (degrees) on the x-axis. On Day 0, participants localized elevation well without the modification (pre-adaption, control) but were unable to localize with the modified pinnae (pre-adaption, modified). In the middle of the adaption period, participants were somewhat able to localize elevation (adaption, modified). In the post-adaption period, we see that performance for both the control (post-adaption, control) and the modified (post-adaption, modified) has returned to a similar level as the pre-adaption control. Image adapted from *Sensation and Perception ed. 11* [48] and Hoffman et al. (1998) [55]. Used with permission.

2.5.2 Exploration Environments as Perceptual Training

Current explicit perceptual training methods are limited to a rehabilitative context. Such training is like sensory flashcards - the user receives a stimulus, provides an answer, and then gets feedback on if they were correct. This works but relies on memorization. This rote training may not capture attention as well as a more engaging, interactive approach. Video games can demonstrably improve perception skills, but commercial games require hours of play time to improve perception [10, 26]. Auditory training games have been developed for clinical use and proven effective at enhancing speech intelligibility in background noise [117, 113].

One relevant experiment on how a rich environment impacts outcomes of sensory learning is work by Hofman et al. on adult adaptation to altered pinnae shapes, which provides compelling evidence for ongoing plasticity in the mature auditory system [55]. In this experiment, Hofman et al. introduced an artificial pinnae that modified the head-related transfer function (HRTF) of participants and changed how they were able to localize elevation. Since pinnae only affect elevation localization, we expect the azimuthal localization to remain the same and elevation localization to adapt over time. Figure 2-4 depicts the adaption curves of one participant throughout the study.

This study yields several insights into auditory neuroplasticity and sensory adaptation. It demonstrates that within approximately one month, participants successfully adapted to altered spectral cues affecting sound localization. Remarkably, this adaptation occurred implicitly, without the aid of additional training or explicit feedback during the localization measurements. Furthermore, upon removal of the ear modifications, participants retained their ability to localize sounds accurately using both their original and newly learned spectral cues. This dual capacity indicates a flexible cognitive framework capable of maintaining multiple spatial maps concurrently rather than a simple overwriting of previous sensory associations.

This suggests a robust, autonomous mechanism for sensory re-calibration in the adult auditory system when in a rich, multisensory environment (like the physical world).

2.5.3 Appropriate Mental Models

Let us consider the viral illusion of *The Dress*[69]. Sensory illusions like this can give insight into how mental models and priors can impact perception.

Color is typically considered an invariant characteristic of an object, where most observers agree when asked to identify the color of something. However, in the case of *The Dress*, viewers perceived the dress's colors wildly differently depending on their mental model of the lighting conditions [69]. The dress is a photo of a blue and black dress taken during the day; however, if viewers thought the photo was taken at night, they saw a white and gold dress. The mental model of the environment caused unconscious inferences, which led to divergent perceptions of color.

Mental models play a crucial role in perceptual interfaces, influencing users' intuitive reactions and perceptions of stimuli. As exemplified by *The Dress* illusion, variations in mental models can lead to dramatic differences in perception. This concept's significance extends

Figure 2-5: Image of *The Dress* that went viral in 2015. Viewers perceived the dress's colors differently depending on their implicit assumptions about the lighting conditions. Image from Wikipedia under Creative Commons License.

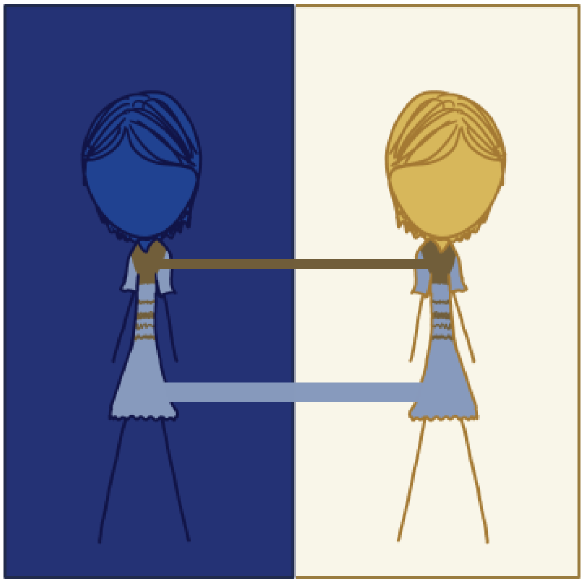

Figure 2-6: If viewers thought the photo was taken at night, they saw white and gold; if they believed the photo was taken during the day, they saw black and blue. Image from explainxkcd.com under Creative Commons License.

beyond sensory illusions and impacts the design and effectiveness of perceptual augmentation devices.

2.5.4 Multisensory Interfaces with Sensorimotor Feedback

Multisensory experiences have been found to be particularly effective in promoting neuroplasticity. The integration of multiple sensory modalities can enhance neural reorganization and learning. A meta-analysis by Li et al. (2022) [75] hypothesizes about three possible mechanisms by which a multisensory context may promote perceptual learning: 1. improved perceptual decision-making; 2. increased saliency and attentional focus; 3. additional or more reliable information.

More recent studies have further supported this notion, demonstrating the effectiveness of sensorimotor training over more passive approaches for developing perceptual understanding. Bramley et al. (2018) found that participants who interacted with a simulated physical environment with altered physics were more able to predict behaviors over passive participants [18]. Similarly, Whitton et al. found that the sensorimotor approach was more effective for training speech-in-noise recognition [117].

Instead of using a single sensory channel to encompass information, I propose a multisensory approach in which one sensory modality dominates, but other senses contribute to resolving the percept. This mimics the natural world in which physics creates multisensory phenomena.

2.5.5 Natural Affordances without Discretization

The way we treat auditory and haptic interfaces is currently discrete or codified. However, this misses out on a lot of dimensionality present in the raw signals. This section focuses new design paradigm that focuses on naturally presenting information without discretization to allow for natural perceptual boundaries to form.

I believe one reason this has not been explored further is that much of the research in HCI has focused on changing affordances to improve usability. Usability is generally defined as 5 Es: effective, efficient, engaging, error-tolerant, and easy-to-learn. This historical precedent creates an implicit philosophy that equates good design with being easy-to-learn. This section introduces the concept of alternative signals instead of those based on physical parameters such as amplitude or frequency. These signals may be more challenging to recognize initially but have the potential to create richer haptic and auditory interfaces.

Physically-based Signals and Perceptually-based Signals

Often, research on haptic and auditory will ground itself in psychophysical responses to basic physical parameters (for example, amplitude, frequency, and intensity). However, these physical parameters do not linearly map to perceptual space. Frequently, individuals will have an existing perceptual categorization that is different from the physical parameters.

Prior work by Anathabotla, Ramsay, and Paradiso (2019) found that individuals preferred to identify a sound's source and categorize it within their existing auditory framework of the world. This method of perceptual categorization is different from the typical psychophysically-motivated approach [3, 2]. Tan et al. (2020) caution that psychophysical measures do not necessarily provide a good metric for perceptual clarity in haptics: "Physical and perceptual dimensionalities may be the same, there also exist ample examples where the two are not equal" [107]. Tan et al. provide an example of this occurring in a haptic experiment in which two physical dimensions, number of cycles and amplitude, combined to form one percept of perceived intensity [107]. Recent computational research lends support to the notion that dynamically created perceptual categories are fundamental to human cognition. Tenenbaum et al. (2011) [109] demonstrated that computational models employing probabilistic inference over hierarchies of flexibly structured representations successfully replicated complex abstraction processes observed in humans. Their findings suggest that hierarchical categorization may play a crucial role in human thought and perception.

Chapter 3

Haptic Augmentation for Improved Situation Awareness

3.1 Introduction

This study marks the first of a series of three investigations exploring various multisensory interfaces designed to augment human perception. This investigation was a pilot study focused on a novel sensory augmentation device in the form of a hat, aimed at enhancing pilots' reaction times in a fixed-wing aircraft. The research was motivated by the potential of multisensory integration to improve human perception and thus, performance in complex operational environments. The development of this device and related pilot studies have yielded significant insights that have informed the subsequent studies of this thesis.

One major finding was the importance of environmental context in perceptual augmentation: The efficacy of the device varied significantly between day and night conditions, with nighttime operations demonstrating markedly improved stated utility and variation in performance. The variable utility across different environmental contexts proved to be a formative insight, steering subsequent research toward a more ecologically driven approach aligned with Gibsonian principles of perception.

Another key finding that emerged from this pilot study was the identification of perceptual ambiguity as a critical factor in the effectiveness of perceptual augmentation devices. The results suggested that sensory augmentation is most beneficial in situations where natural perceptual cues are unclear or insufficient. This insight led to the conclusion that augmentation should be selectively applied, focusing on scenarios where perceptual ambiguity is high, rather than providing constant additional sensory input. This "always on" nature renders some perceptual augmentation devices more annoying than useful.

Finally, the study highlighted the significant role of mental models in shaping both the initial adoption and ongoing use of perceptual augmentation devices. The pilots' existing mental models of what the haptic device represented were indicative of performance. (One described it as a bee that they wanted to move away from, while others imagined it as an indicator to move towards). While these directionality cues can be learned through repetition, providing users with a clear mental model of the device and what it represents before use can significantly improve outcomes.

This work was a joint collaboration between Noam Eisen, Sam Chin, and Brittany Bishop. All co-authors contributed equally and were involved with the conception, execution, and analysis of this work. Andy Liu, Katya Arquilla, and Joe Paradiso were involved with the revision and review of this work.

3.2 Background

Maintaining situation awareness is essential for safety in piloted aircraft; however, in crowded visual environments, it can be challenging to attend to and perceive the relevant information [86]. One approach, following the multiple resource model [118], is to expand information to other sensory resources. In modern-day cockpits, this primarily has been enacted with auditory cues. The auditory channel is well-suited for alarms because auditory alerts do not require active attention, unlike visual alerts. Some examples of auditory alarms include altitude alert tones and ground-proximity warning systems (GPWS), which provide aural instructions like “TERRAIN. PULL UP”. The prevalence and variety of auditory alerts combined with radio communications and crew coordination can render the auditory channel overloaded. One approach to addressing auditory or visual saturation is to expand into the haptic domain. Haptic displays have been found to be particularly useful when “visual or audio information is unavailable or deteriorated” or “the user’s sensory capacity is overloaded” [25]. There are few active haptic interfaces in most cockpits aside from the stick shaker, which vibrates to warn of an impending stall. Prior work on using haptics for situation awareness in aviation has proven reasonably successful. In 2000, haptic cues were demonstrated to improve spatial awareness while hovering a helicopter [95]. Shortly after, the US military began testing a haptic device called TSAS (Tactile Situation Awareness System), a vibrotactile belt that provides pilots an understanding of their surroundings haptically. The TSAS device was shown to assist helicopter pilots landing in visually-degraded states [59]. Helicopter pilots using an updated version of TSAS were able to "non-visually hover helicopters," and the technology improved situation awareness and decreased pilot cognitive load (Rupert et al., 2016). Additionally, we hypothesized that haptics’ effect would be more pronounced in low-visibility environments, where spatial disorientation and loss of situation awareness are more common. A secondary goal was to investigate the efficacy and comfort of head-mounted haptics in the context of aviation. Previous research has demonstrated the efficacy of head-worn haptics under visually-degraded conditions[12, 15]; however, there has not been extensive work on head-mounted haptics as a part of a multisensory display. Pilots routinely wear headwear such as helmets and headsets, so a head-mounted haptic system potentially has a lower barrier of adoption compared to a new type of body-mounted wearable device.

The purpose of our experiment was to answer the following research question: “Can additional haptic input improve pilot recovery from unusual flight attitudes?” The FAA’s Airplane Flying Handbook defines an unusual attitude as pitch beyond +25° or -10° or bank angles greater than 45°. We selected the task of recovery from unusual flight attitudes because the maneuvers used to recover are not part of normal flight operations and are likely to atrophy over time. We hypothesized that recovery time would be shorter with the addition of haptic feedback since a haptic alarm can directly indicate the action needed without requiring the time to rebuild situation awareness. A meta-analysis by Prewett et al. (2012) found that users benefited from haptic-visual inputs compared to visual-only.

3.3 Methods

We performed an exploratory study with human participants to investigate our research questions. The task was to correct from an unusual flight attitude to straight and level flight as quickly as possible. We tested seven participants meeting the criteria of at least five hours of flight experience to mitigate learning effects during task execution.

3.3.1 Setup & Hardware

Participants were tested using Digital Combat Simulator (DCS) simulating a Yak-52. Participants viewed the simulation on a 27" Dell monitor and used a Microsoft SideWinder Force Feedback 2 Joystick, Virpil APC ACE Flight Pedals, and Virpil VPC MongoosT-50CM3 Throttle for flight control (Figure 3-1).

Figure 3-1: Simulator Flight Controls

Haptic cues were provided through a prototype bone-conduction haptic device installed in a baseball cap (Figure 3-2). When activated, the haptic device simulated tapping on the wearer's head in any of four locations: anterior, posterior, left temporal, or right temporal. Participants were instructed that maneuvering "away" from the tapping would achieve straight and level flight (e.g., tapping on the forehead would indicate pulling the stick 'aft') (Brill et al. 2014). Tapping was activated when the pitch exceeded +25°/-10° or the roll exceeded 45°. When both roll and pitch thresholds were exceeded, the system prioritized zeroing roll before correcting pitch, to avoid overstressing the plane.

Figure 3-2: Prototype Bone Conduction Haptic Device

3.3.2 Procedure

Before the task, participants flew freely for six minutes to familiarize themselves with Yak-52 flight characteristics and the haptic cues. Two levels of the test condition (haptic-visual, visual-only) were tested at two environmental conditions (daytime and nighttime), yielding four total test conditions. For each test condition, three unusual flight attitudes were tested for a total of twelve trials per participant (Table 1). The order of conditions was counterbalanced among participants. Following the task, users completed a NASA TLX and qualitative preference survey.

3.4 Results

3.4.1 Time to Recovery

We hypothesized that users would recover more quickly in the haptic-visual condition compared to the visual-only condition. In order to calculate this, we subtracted the visual-only by the visual-haptic to calculate the difference in reaction time (μvisual-only - μvisual-haptic). We found no statistically significant difference in recovery times between the haptic-visual and haptic-only conditions using a Friedman Test (μ = - 0.36, σ = 1.44, p = 0.892). We observed inter-individual differences in response to the haptic feedback (Table 2). Anecdotal evidence from participants suggests lack of improvement in the haptic-visual condition

	Daytime	**Nighttime**
Haptic Feedback	• Pitch up, bank left/right • Pitch down, left/right • Inverted, nose low	• Pitch up, bank left/right • Pitch down, left/right • Inverted, nose low
No Haptic Feedback	• Pitch up, bank left/right • Pitch down, left/right • Inverted, nose low	• Pitch up, bank left/right • Pitch down, left/right • Inverted, nose low

Table 3.1: Performance conditions with and without haptic feedback during daytime and nighttime.

due to participants' perception of the indication direction (moving away from the haptic input) as unintuitive. One participant stated that he accidentally moved toward the tapping multiple times.

Participant	1	2	3	4	5	6	7
Mean Difference (seconds)	2.77	0.16	-3.1	1.42	-2.37	-0.81	-3.09

Table 3.2: Mean differences in performance times for each participant.

Table 2. Inter-individual Differences in Response Time One notable result was that the standard deviation of time to recovery was much larger for the night condition (σ-day $=$ 0.63, σ-night $=$ 2.24) (Figure 3-3). This may support our hypothesis that haptic inputs have a greater impact on recovery time in visually degraded conditions. During the day, pilots performed similarly. However, at night the magnitude of the effect of adding haptics was much larger.

3.4.2 NASA TLX and User Preference Survey

We calculate the results of the NASA TLX survey by subtracting visual-only from the visual-haptic to calculate the difference in rating. We found that on average, participants had higher workload when haptics were added haptics, but this result was not statistically

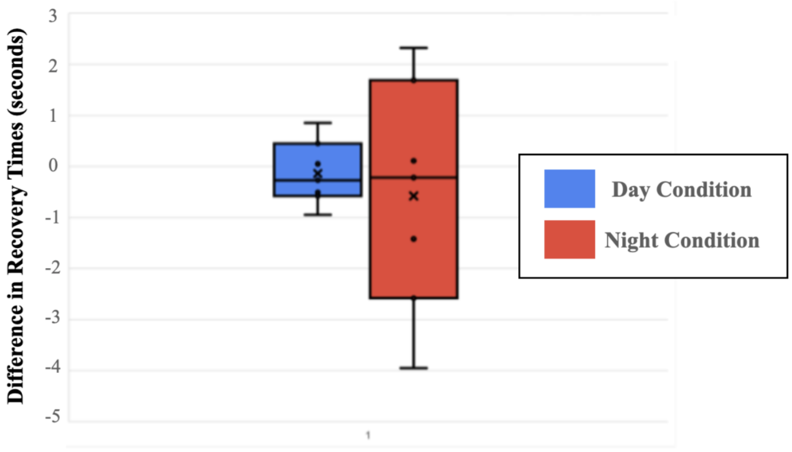

Figure 3-3: Large Standard Deviation of Time-to-Recovery

significant (μ= -5.75, σ = 5.52, p=0.171). This may be due to the unintuitive indication direction discussed previously. For our user preference survey (Table 3), we found that users had a slight preference for using the device (μ = 4.57 on a 7-point Likert-scale, σ = 1.72). Notably, multiple participants found the device annoying during the day but helpful at night when visual cues were ambiguous. Since we asked these questions at the end of the study, the responses represent a combination of both day and night conditions.

Statement	**Mean**	**Std. Dev.**
I would wear this device again.	4	1.29
The information I was receiving from the device was clear.	4.43	1.27
The sensations produced by this device were distracting.	3.71	1.38
My performance with the device active was better than my performance when it was inactive.	4.57	1.72

Table 3.3: Survey responses about the device.

3.5 Discussion

This study illustrates the importance of a user's mental representation and interpretation of the meaning of the haptic information. With sufficient training, muscle memory can be built, but providing users with an appropriate mental representation can significantly reduce the need for flight training by making stimuli intuitive. This mental representation also includes the selection of the proper corrective action and the ability to project the results of that action into the future vehicle state. During the exercise with unusual pitch and roll, the pilot who had trained on fighter jets commented “[the haptics] seemed to indicate the wrong direction” since he was taught to perform the maneuver as one action instead of correcting roll and then pitch. Further study of differences in the helpfulness of haptics in novice versus experienced users and design of the implications of primacy could facilitate a more holistic design plus training approach. One way we attempted to make the mental model more intuitive was to make the haptic input sharper by adjusting aspects of the waveform. One participant who performed well described imagining the haptic stimulus as a bee sting and wanting to move away from it as a result. Future studies should investigate how different forms of haptic feedback (e.g., square wave versus sinusoidal inputs) influence the speed and accuracy of comprehending the signal. One additional finding was the significance of context when providing haptic cues - users found haptics annoying during the day but helpful at night. Instead of having haptics that constantly transmit information, we recommend having haptics conditioned on sensory context (i.e. only providing haptic information when the visual field is ambiguous).

Chapter 4

Haptic Augmentation for Improved Pitch Interval Learning

4.1 Introduction

This marks the first actual foray into perceptual augmentation. This study combines ideas from perceptual learning and embodied cognition in an attempt to realize some of the ideas presented in Chapter 2.

Making sense of sensory information is the foundation of everyday activities. Individuals develop improved acuity for sensory information through experience or practice in specific sensory tasks, which for some results in permanent changes in their perceptual response [30]. For example, some paint makers develop acuity to color and can mix paint from fundamental colors by simply looking; wine connoisseurs can differentiate notes in the smell and tastes of different wine types; professional piano tuners have more pitch acuity to fix out-of-tuned pianos.

Different from learning math or physics, which is represented using symbols and abstractions, perceptual learning requires low-level changes and adaptation in how we respond to sensory stimuli. This is similar to learning a motor skill (e.g., learning how to bike or play sports), where completing bodily gestures is difficult to capture symbolically and thus hard to learn through written representations or observations alone [67, 38]. Both motor learning and perceptual learning often require strenuous practice to achieve mastery.

To that end, a few systems have been developed to aid perceptual learning. While motor skill learning requires specific knowledge, perceptual learning can happen implicitly [100]. Multisensory perceptual learning offers an advantage by integrating information from multiple sensory modalities. To illustrate with an example, consider learning a new skill like ceramics. For new ceramicists, learning to see when the clay is perfectly centered is primarily a visual task. However, learning to center the clay is much easier when you hear the rhythm of uneven spots and feel the clay pushing back against your hands. Auditory, haptic and visual all combine to form a percept about how centered the clay is.

In our work, we focused on the specific case of learning auditory perceptual skills. In particular, we looked at musical ear training, a common yet challenging task that is typically undertaken through only one modality. Ear training refers to gaining the ability to identify musical components (e.g., melodies and chords) by ear, an essential component of developing musical proficiency. For example, violinists need to tell the pitch based on hearing when needing to imitate recordings. Jazz musicians intuitively identify different chords and intervals to be able to improvise during a performance. However, identifying intervals and other musical elements by ear can be challenging for novices. Developing this skill requires extensive training, and the traditional rote approach of repetitive listening and identifying often requires long periods of strenuous, repetitive practice.

In this work, we explore a multisensory (audio and haptic) platform for training to identify musical intervals, or the differences in pitch between two tones (two audio frequencies). We develop a perceptual training platform that involves an interface with a haptic wearable placed on the back. This device provides vibrotactile feedback concurrently while the notes are played. To understand whether haptic feedback has a positive effect on learning to perceive musical intervals, we conducted a study with 18 participants. Participants had no prior musical ear training, and about half were assigned to an audio-only control condition. Initial results show that participants with haptic feedback could identify musical intervals more accurately while feeling less frustrated and more engaged.

Our main research questions were: (1) can haptic reinforcement improve the performance of interval **recognition**? and (2) what is the effect on auditory perceptual **learning**?

4.2 Background & Related Work

Our perception relies on the ability to interpret various stimuli through our senses. In particular, auditory perception plays an important role in our daily lives, allowing us to distinguish between sounds and understand their meaning. Such auditory skills can often be refined through focused training.

In this work, we focus on musical ear training as an example of learning an auditory perceptual skill. We first discuss the background of musical ear training, with a focus on musical interval recognition, and then we discuss learning methods and end with ones with haptic feedback.

4.2.1 Musical Ear Training & Pitch Interval Recognition

Ear training, or aural skills training, involves learning to identify musical elements like melodies, chords, rhythms, and their building blocks by how they sound. This centuries-old practice remains a core aspect of Western musical pedagogy, extending beyond early education to even the college level [63]. The ability to rapidly recognize musical structures is essential for many musicians, for example jazz musicians rely on this ability to interpret and respond to others in improvisational contexts.

Ear training often begins with learning to identify the distances between notes in a melody as this is considered "basic to good musicianship" [22], and the distance is known as a musical interval. We have selected pitch interval recognition for a number of reasons we discuss below.

Pitch interval identification is a well-studied perceptual task [102, 121, 83]. Many studies have been conducted on pitch interval perception and it has been established as a learnable but also challenging perceptual skill [76]. Trained musicians perceive musical intervals as discrete categories. In contrast, many novices perceive musical intervals as being perceptually ambiguous [21].

Unlike a perceptual task fabricated for a scientific study, pitch intervals are a real skill. As noted, ear training is an integral part of musical training and supports both fundamental pedagogy [90] and advanced activities like transcription [39]. Pitch interval identification is a skill that is difficult to pick up without training. Although most humans have the ability to recognize precise frequencies [80], pitch interval recognition is still a challenging task. This is useful because it ensures that participants are starting from a similar skill level. Pitch interval recognition is a class of tasks that are unpleasant to learn. Students often dislike learning musical intervals because the traditional rote approach to learning is repetitive and not engaging. Thus, techniques for making such learning more enjoyable or easier might indirectly improve learning outcomes.

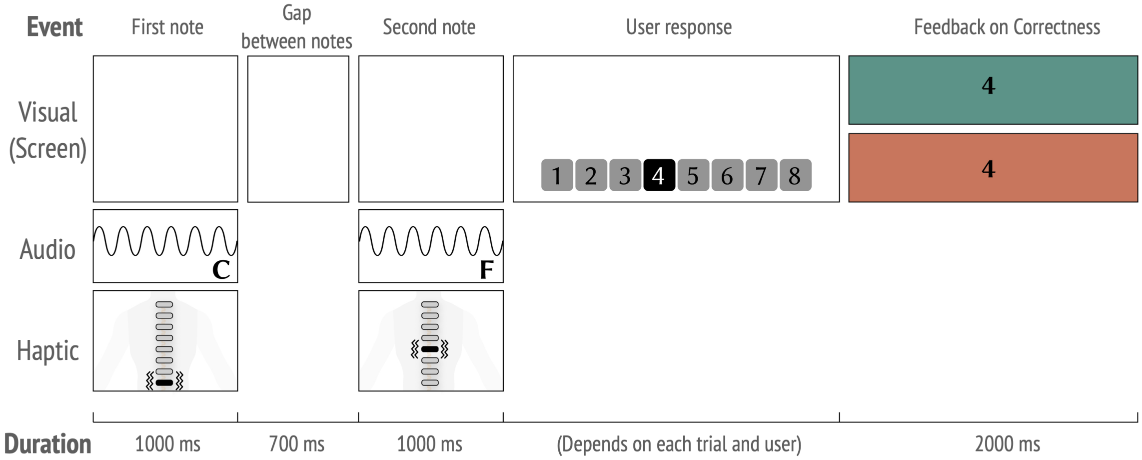

Figure 4-1: The user experience procedure of a single trial. The user hears (and feel) two musical tones in sequence with a small gap in between. Then they type a number representing their guessed interval and receive on-screen feedback about their guess.

4.3 Design of Multisensory Interval Training System: Purrfect Pitch

We set out to design a multisensory system that allows novice learners to identify musical intervals beyond rote training. Specifically, we have the following design goals: First, we want to make sure the system can improve the learning outcome in terms of the accuracy of the users' response. Second, usability. The system usage should be intuitive, and the additional haptic stimuli should not distract the learner. The wearable device should be comfortable to wear over time. Third, the system should be engaging, and novice learners should prefer to train with it over rote practice, as motivation plays an especially significant role in perceptual learning. Finally, we want to enable the community to build on top of our tools, and thus we have open-sourced our designs for future work to build upon.

We designed a multisensory platform consisting of a wearable device that provides haptic feedback and a digital interface that plays musical notes and displays information about the learner's performance. We chose the Western diatonic (major) scale and only evaluated ascending intervals, as it is a common starting point in Western ear training lessons. This gave us eight possible intervals. Additionally, we limited possible tones to mid-range frequencies between C2 (65.4 Hz) and B4 (493.88 Hz), and the training set combinatorially looped within this range.

An example of a user's experience is shown in Figure 4-1 and in the Video Figure. The training experience is as follows: Two tones of a randomly selected interval (between 1-8) are played sequentially for each trial. We chose a note length and vibration length of 500ms based on the average length of vibration in [99]. At the same time, the user feels vibrations sequentially in two places: first always from the bottom-most module near the lower back, then from another module certain distance (or "interval") apart. For example, if the trial was an interval of 8 (an octave), the module towards the lower back vibrates when the first note is played, and when the second note is played, the wearable system vibrates the eighth module from the lowest one (i.e., the top most module near the neck). After hearing and feeling the notes, the user responds by pressing a number key from 1 to 8. The interface then displays the correct interval number and a green or red screen for a correct or an incorrect response, respectively. The subsequent trial would again choose a random interval and use the second tone of the previous trial as the new first tone. Next, we describe the design considerations of the different components of our platform.

4.3.1 Hardware Design

Prior work shows the natural mapping "between the body position and egocentric orientation" is an intuitive way to guide an individual's attention [24]. We chose a vertical arrangement of the haptic modules to match the vertical spatial metaphor used in Western music, where higher pitches are perceived as "going up" or placed higher on the body. Following this spatial metaphor, we used the eight modules to map to the eight intervals (i.e., eight possible spatial differences) in an octave. We used the location of modules to encode interval information and kept the vibration pattern or intensity consistent. We considered several body study areas: back, forearm, and wrist. Ultimately, we chose the back because it has a large surface area, and thus the spatial difference between neighboring modules is more pronounced.

We built on the vibrotactile haptic platform (VHP) used for on-body haptic research [32]. We redesigned the flexible PCB layout to conform to the back along the spine. We modified the physical enclosure of the modules so that they can be stitched onto a piece of fabric (Figure 4-2). The Bluetooth module of the VHP wirelessly communicates with our learning interface running on a PC. The distance between two neighboring modules is 3cm, and the total length of the electronics is 75 cm. We followed the findings by Plaisier et al. [92] on individuals' perceptual distance of vibrotactile simulations around the spine. In their findings, the vertical layout resulted in the least overall variance of perceptual distance, and we chose the 3cm to fit all eight modules on the spine. Since our setup does not exactly replicate Plaisier et al.'s work, we also conducted a perceptual study to understand the spatial discrimination between the modules, detailed in Section 4.5.1.

4.3.2 Learning Interface

As discussed in section 4.2.1, perceptual learning cannot be performed passively – it requires active learning with attention to detail and repetition. We used a similar method of implicit active learning and sensory integration utilized in the work by Seim et al. [99].

The goal of the learning interface is two-fold: (1) to orchestrate the audio and haptic feedback stimuli in different training and test conditions, and (2) to provide the user feedback on the correctness of their response to enable active learning.

The audio and haptic stimuli are played concurrently to enable sensory integration while the screen displays nothing to limit visual interference effects [74]. The system plays the lower note first, followed by the higher note. After the user enters a number, they are given feedback about whether they gave the right or wrong answer, and they see the correct interval number (Figure 4-3). We choose to display a uniform color block of red or green instead of words (“correct” or “incorrect”) to allow users to advance quickly through the trials to avoid requiring extra cognitive processing of the feedback. Users can repeat a "question" by pressing the space key and hear the notes again, but they only have one chance to answer. Two seconds after the user guesses, the next trial automatically advances.

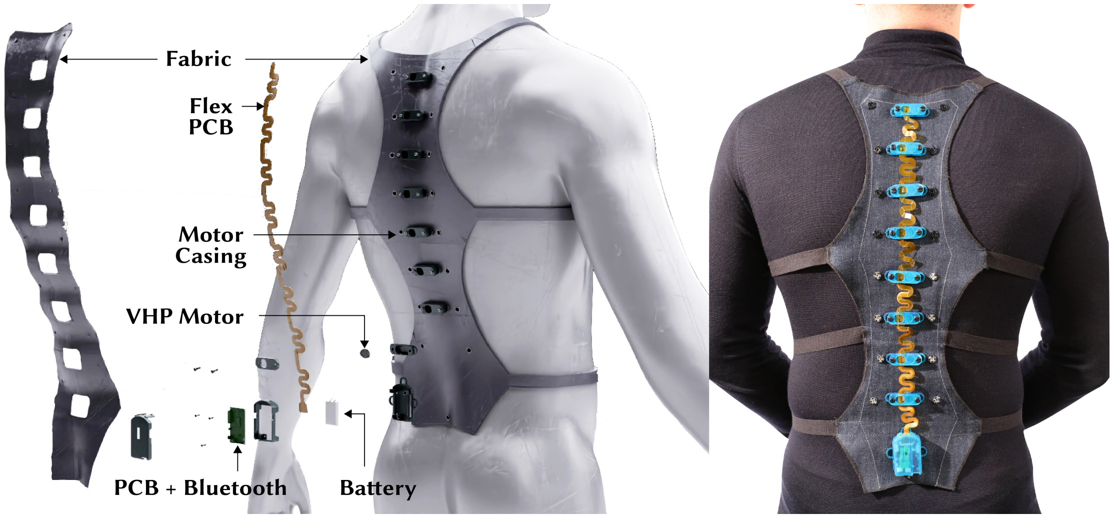

Figure 4-2: Overview of the haptic device. It is a vest like wearable with eight vibrotactile modules evenly distributed along the spine. The haptic device can be controlled by our web-based learning interface via Bluetooth.

Setting up the learning system is simple: the web-based interface can be easily deployed and connects to the haptic vest via Bluetooth. The interface offers different toggles for turning on/off the sensory feedback channels.

4.4 Methods

We conducted a between-subjects in-lab study to evaluate and understand our system as (1) a multi-modal feedback system and (2) a learning tool. Specifically, our research questions are as follows:

- RQ1: What is the **perceptibility** of the spatial distance between two stimuli if participants are only given haptic feedback or only given audio feedback?
- RQ2: Does haptic feedback improve the **accuracy** and **response time** of identifying musical intervals *during training*?
- RQ3: Does the introduction of haptic feedback affect people's cognitive load *during training*?
- RQ4: Finally, after training with our system, can novice learners **learn** musical intervals identification by ear only?

Figure 4-3: The web-based learning interface that plays the musical notes and connects to the haptic device. The color green represents a correct answer, with a number displayed representing the correct interval number.

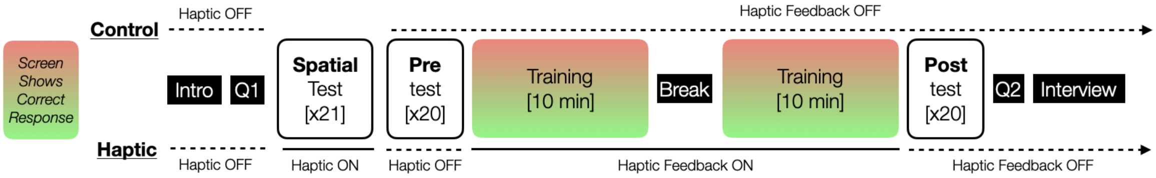

Figure 4-4: Experimental protocol. The top lines represent the control condition and the bottom lines represent the experimental condition. The dashed lines represent when haptic feedback is absent and a solid line represents the presence of haptic feedback. The numbers in brackets represent the number of trials.

4.4.1 Experimental Protocol

Conditions

Our experiment consists of the control (without haptics) condition and the experimental condition (with haptics). In the haptic condition, those participants wore the wearable on their back and felt haptic feedback while hearing the musical notes during the spatial perceptual experiment and the training sessions. We conducted a between-subjects experiment to avoid learning effects from one condition to another within a participant.

Participants

We recruited 18 participants (age 18-36; 5 identifying as female or non-binary) through email and Slack promotion. Our participants consisted of undergraduate and graduate students who reported having normal hearing, no to little prior training in identifying musical intervals, and none to minimal experience with playing musical instruments. Participants were recruited through our university mailing lists. The participants were randomly assigned to either the audio-only control (n = 10) or the audio-haptic condition (n = 8). The study lasted around 45 minutes, and the participants were compensated with $15 in the form of a gift card upon completion.

Procedure

Figure 4-4 shows the experimental procedure. All participants were asked to sit in front of a computer and given over-the-ear headphones (Sennheiser HD 560) to wear for the duration of the experiment. Only participants in the haptic condition were given our wearable device to wear on the back over tight clothing.

We start the experiment by first explaining the concept of musical intervals. The participants then filled out a demographics questionnaire (Q1). The participants in the haptic condition also went through an additional spatial perceptual experiment. The spatial perceptual experiment aims to measure how well a person can identify the felt spatial distance between two sequential vibrations on the back along the spine. There are eight possible pairs of vibrations, all starting with the bottom-most module closest to the tailbone. The user will respond to each pair of vibrations by typing a number that represents the felt spatial distance between the vibrations. They could use a non-zero, positive number, subject to the constraint that a larger number should correspond to a greater distance. The order of the eight pairs of vibrations is randomized. The experiment ends when all eight pairs of vibrations have been played.

For both conditions, participants used our learning interface in which they must rapidly identify pitch intervals. They were asked to go through as many trials as possible. The user response, number of correct responses, response time, and number of repeats were collected. To avoid fatigue, participants went through two ten-minute training sessions with a small break in between. Only participants in the haptic condition felt haptic feedback during the training sessions. We conducted a short test before and after the training sessions (a pre-test and a post-test). The tests aim to measure the participants' true auditory perceptual ability, meaning that they were done without haptics and without on-screen feedback on the learning interface about the correctness of their response.

Finally, at the end of the study, participants completed a brief questionnaire (Q2) containing NASA TLX (Task Load Index) questions and additional questions about such as levels of engagement and effectiveness. They also went through a semi-structured interview about the learning experience and shared any other comments with the experimenter.

4.5 Results & Analysis

Here we represent results and analysis from the quantitative experimental and survey results and qualitative responses to answer each of our research questions.

4.5.1 Perceptual Estimation of Spatial Distance

First, we want to understand the perceptibility of the spatial distances of each pair of haptic modules on the back. This part of the study is similar to the free magnitude estimation test [104, 92], but we are interested in the relative distance within a distance range determined by our hardware rather than the absolute scale.

Similar to the setup of a free magnitude estimation task, we asked participants to rate the felt distance on a scale of their choice. To account for the variations in individuals' range of distance values, we first normalize each participant's score by fitting their ratings to a 0-1 scale, where 0 is the minimum rated distance, and 1 is the maximum rated distance. Concretely, within a participant, for each score, we divide the difference between the score and the minimum value over the difference between the maximum and minimum values.

We first want to know if the perceptual distances of each pair of motors are distinct from each other, especially the neighboring pairs. Figure 4-5 shows the participants' distance estimation of the 8 pairs of vibrations. As expected, the perceived distance increases as the actual distance between the two haptic motors increases. Then we look at whether the

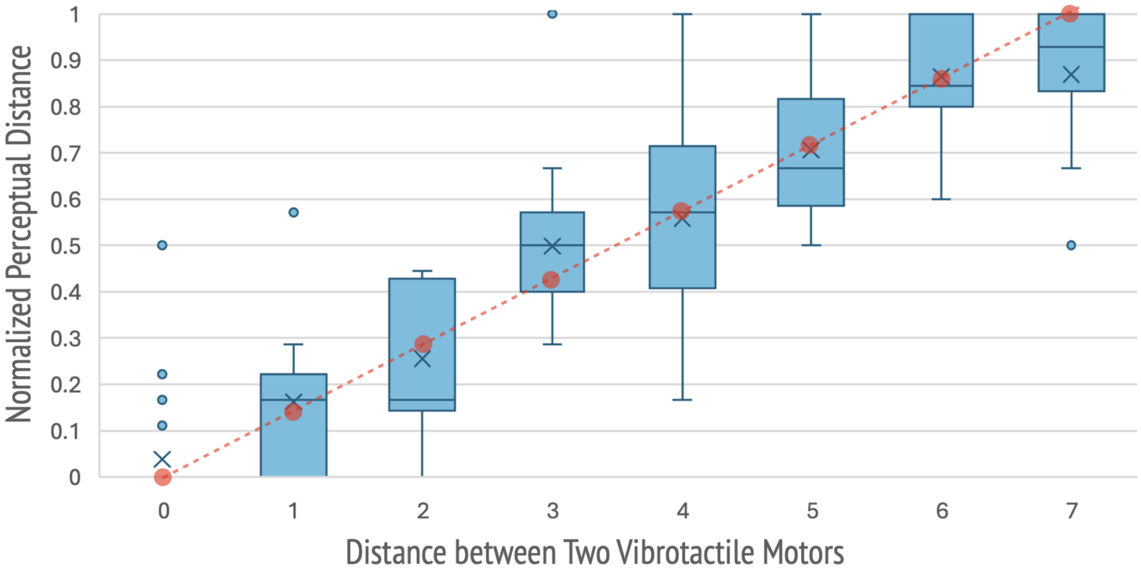

Figure 4-5: Box plots of the perceived distance between two vibrotactile motors on the back. The cross mark indicates the mean value. Red dots and the red dashed line indicate the ground truth.

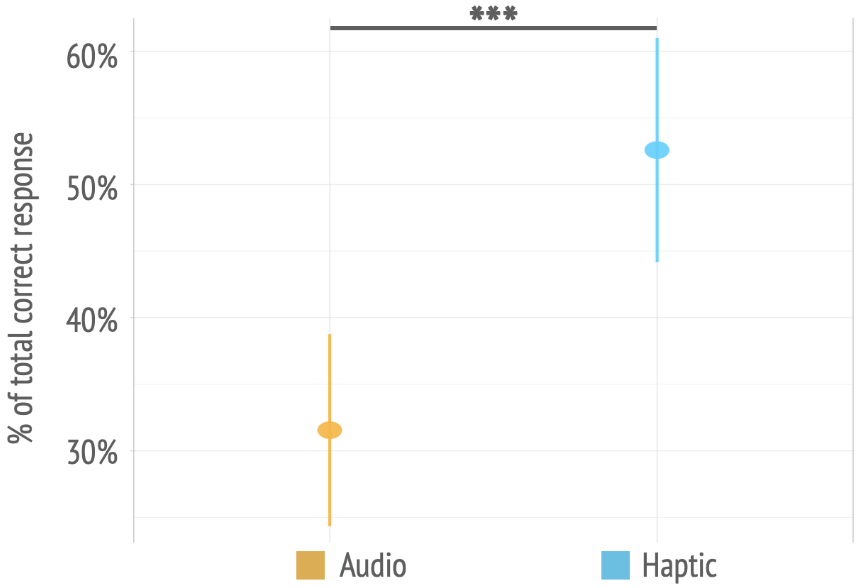

Figure 4-6: Total percent of correct response per condition during training based on mixed-effects logistic regression. The adjusted marginal prediction of a correct response is 34% (95% C.I.=[0.270, 0.410]) for the audio-only group and 54.3% (95% C.I.=[0.463, 0.624]) for the audio-haptic group. *** $p<0.001$

neighboring distances can be differentiated from each other. The differences between the neighboring distances are mostly clear, except for between distances 3 and 4 and between distances 6 and 7. We then look at the direction in which the estimated distances deviate from the ground truth. The red dots and the red dashed line indicate the ground truth distance responses. Specifically, distance 3 is estimated to be bigger and distance 7 is estimated to be smaller. The top motor's position for distance 7 is near the neck, farthest away from the bottom motor. It is also worth noting the significant variance of distance 4. This shows that distances where the top motor is around the middle of the spine are notably perceptually ambiguous.

4.5.2 Accuracy, Learning Rate and Response Time during Training Phase

Effect of haptic stimuli on accuracy during training

In Research Question 2, we want to understand if users would perform better when they receive both auditory and haptic feedback. To evaluate this, we performed a mixed-effect logistic regression (binomial family generalized linear mixed-effects model) in R using the *lme4* package [8]. Such models are a common analysis method in psychophysics for perceptual tasks [85]. We based our regression on the data collected from participants during the "Training" phase which resulted in 3052 observations total across groups. We modeled the likelihood of getting a question correct based on haptic input (a binary variable to indicate whether haptic input was present or not), the number of trials completed (to account for improvement over time), and the interaction between haptic input and the number of trials (to examine whether the rate of improvement/decline differed between the haptic and audio-only conditions). We accounted for individual differences with a random effect for participant ID.

We calculated the marginal adjusted predictions of a correct response for the audio-only group as 34% ($p<0.001$, 95% C.I.=[0.270, 0.410]) and audio-haptic group as 54.3% ($p<0.001$,

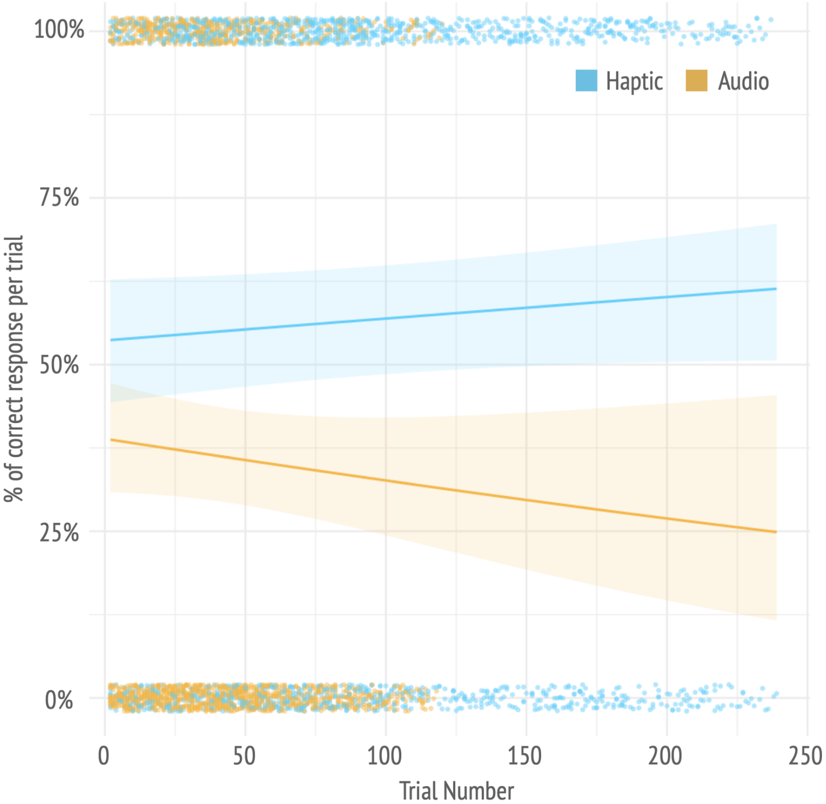

Figure 4-7: Percent correct response per trial over the number of trials during training. The dots represent the raw data which is binary (correct or incorrect). The lines and confidence intervals are average marginal predictions of the mixed-effect logistic regression.

95% C.I.=[0.463, 0.624]) using *marginaleffects* [4]. The difference in predictions can be seen in Figure 4.5.2. Adding haptic input to auditory input increases the likelihood of correctness by about 20% on average. We performed a hypothesis test on this contrast and found this accuracy difference was statistically significant (0.203, p<0.001, 95% C.I. = [0.0967, 0.31]).

Effect of haptic stimuli on learning rate Additionally, we want to compare the learning rate for both groups. We define the learning rate as how much the user is expected to improve for one additional trial (change in expected likelihood of correctness over the change in trial number). Figure shows a slightly positive learning rate (3.17e-4, p =0.172, 95% C.I. =[-0.000138, 0.000771]) for the haptic-audio participants and slightly negative (-5.92e-4, p=0.218, 95% C.I.=[-15.34e-4, 3.50e-3]) for the audio-only participants; however, both of these results are statistically insignificant. The average learning rate is slightly negative

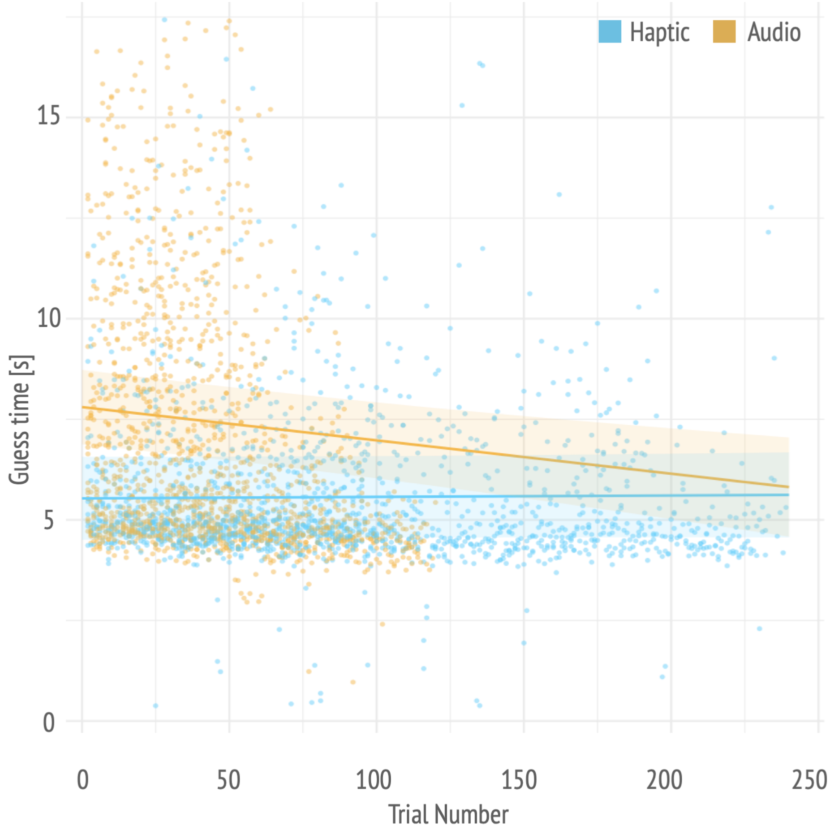

Figure 4-8: Guess time per trial over the number of trials during training. The dots represent the raw data. The lines and confidence intervals are average marginal predictions of the mixed-effect linear regression.

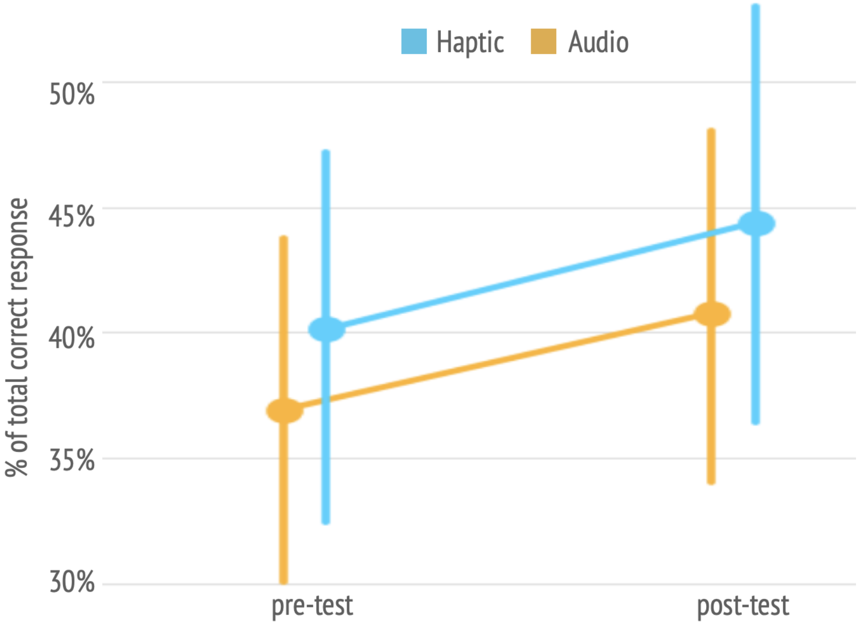

Figure 4-9: Accuracy of pre-test and post-test. Users who received audio-only training got on average 37% correct on the pre-test (95% C.I.=[0.297, 0.441]) and 41% on the post-test (95% C.I.=[0.335, 0.481]). Users who received haptic-audio training got on average 40% correct on the pre-test (95% C.I.=[0.324, 0.479]) and 44% on the post-test (95% C.I.=[0.364, 0.524]).

(-1.19e-4, p=0.647, 95% C.I.=[-6.31e-4, 3.92e-4] which is unexpected.

Effect of haptic stimuli on response time In addition to response accuracy, we also investigated the effect of introducing haptic feedback on users' response time (i.e., the time difference between the onset of the first note in the interval and when the participant pressed a number key). We dropped times below 1.2 s, which is the earliest a user could hear the second note because anything before that would be effectively a random guess or accidental entry. We also dropped outliers beyond 2σ (17.2s) – most of these long guess times involved the user asking a question and were not considered a valid trial. Figure 4-8 shows the mean response time over the course of the study for each condition. We performed another mixed-effect linear regression, modeling the response time based on haptic input, the number of trials, the interaction term, and a random effect for participant ID.

The audio-only group has a higher predicted average response time (6.918s, p<0.001, 95%

C.I.=[6.001, 7.836]) compared to the haptic-audio group (5.244s, $p<0.001$, 95% C.I.=[4.229, 6.259]). We performed a post-hoc contrast to compare these (audio-only vs. haptic-audio), and found a statistically significant difference of 1.674s ($p=0.0165$, 95% C.I. =[0.306, 3.043]). We also see that the slope for the audio-only group is negative (-0.00828, $p < 0.001$, 95% C.I.=[-0.0012.4, -0.00415]) which indicates that audio-only participants got faster over time. The haptic-audio group has a slope that is comparatively flat (0.356, $p = 0.72086$, 95% C.I. = [-0.0016, 0.00231]).

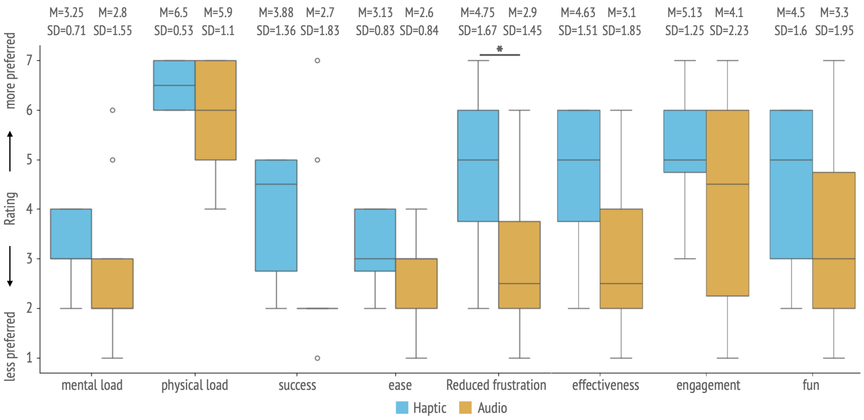

Figure 4-10: Box plots of the questionnaire results. M: mean, SD: standard deviation. *: $p<0.05$

4.5.3 Comparing accuracy of pre-test and post-test

We also measured learning improvement by comparing the scores from the pre-test and post-test. For both tests, each group received 20 audio-only questions without feedback on whether they responded correctly. Users who received audio-only training got on average 37% correct on the pre-test (μ=0.369, σ=0.036, 95% C.I.=[0.297, 0.441]) and 41% on the post-test (μ=0.408, σ=0.037, 95% C.I.=[0.335, 0.481]).Users who received haptic-audio training got on average 40% correct on the pre-test (μ=0.401, σ=0.039, 95% C.I.=[0.324, 0.479]) and 44% on the post-test (μ=0.444, σ=0.041, 95% C.I.=[0.364, 0.524]). In Figure 4-9, the change in percentage correct is about the same for both groups(Δ_{audio} = 0.0385, Δ_{haptic} = 0.0424), and the difference is not statistically significant (t=0.03, p=0.9747).

4.5.4 Cognitive load during training

We administered a partial NASA TLX questionnaire [52] and a separate set of questions. The full questionnaire and the original rating scale can be found in the Appendix. Figure 4-10 shows the questionnaire results. For readability, we converted the scores for this figure, where a higher score is more desirable (e.g., a higher score means the task is *less* mentally loaded and hence more preferred). We found that across the board, the haptic condition performed better.

The frustration score for the haptic condition is significantly lower (paired t-test, $p<0.05$). It is worth pointing out that learning a new skill presents challenges beyond improving one's performance, such as maintaining one's motivation [110]; in our case, learning musical intervals typically requires arduous rote ear-training. Figure 4-10 shows that participants reported feeling more engaged, and considered the learning experience with haptics more effective and fun. Anecdotally, participants in the haptic condition reported that they "relied more on instincts during haptic trials" (P8) and were "more confident when pressing the key (to give their response)" (P3).

4.6 Discussion

Overall, we found that participants in the auditory-haptic experimental condition performed better compared to the audio-only group during training. Our primary findings are: **(1) The haptic-auditory condition performed 20% more accurately compared to the audio-only control. (2) The haptic-auditory condition responded 1.674s faster than the audio-only control.**

We suspect that one reason haptic participants were able to perform much better during training is purely on an information-theoretic basis – they simply had more information to inform their guess. For example, Participant 2 reported that haptic feedback "allowed me to narrow down my interval guess within a range". Additionally, participants in the haptic condition commented on noticing how their sensory weights changed as their auditory perception improved: "[I] use the haptic distance first and connect it with audio" (P1) and

"Coming into this, [I was] relying exclusively on haptics. [I] started to notice patterns in sound." (P7). The spatial discrimination test (Figure 4-5) also shows that the design and placement of our haptic device on the back produced sufficiently distinct stimuli that benefited the judgment of audio intervals.

Both of these results support the idea that multisensory interfaces with haptics have the potential for augmenting perception in a short time (less than an hour). We believe we were able to see such rapid improvement because multisensory integration is a dynamic process which adjusts quickly and subconsciously. Multisensory integration is the mechanism through which the nervous system uses sensory information of varying reliability to create a coherent perception of the world. For example, speech perception is not purely an auditory task - depending on the noise in the environment humans may rely on vision more than sound in order to understand speech [20]. Often this process is considered dynamic Bayesian optimization problem in neuroscience [34] because the reliability of a given sense varies depending on the environment.

Even though the haptic group did not significantly outperform the audio (control group) in the pre-post training test results (Figure 4-9), we were able to find interesting insights when looking at the partcipants' response by interval.

First, the results show that removing the multisensory device did not *negatively* impact performance in our study. One concern with using technology to augment perception is overreliance - for instance, blind spot indicators are useful, but this overreliance may cause problems when a driver is in a different car. In our case, overreliance would look like users ignoring the audio information and only using the haptic device for the task. However, we did not see a lower in the accuracy count for the haptic group after training.

In addition, the haptic device positively influenced participants' judgment of specific intervals that performed worse than others during the pre-test. Figure 4-11 shows the pre-post test results broken down by intervals. We see that the haptic group shows lower and reduced variances in post-test guesses across intervals compared to pre-test, whereas the control group shows consistently large variances in post-test guesses and no significant reduction compared to pre-test. The post-test guesses of the haptic group also show less anchoring of pre-test guesses. However, these results are only empirical, and the implications of the observed trends require further analysis.

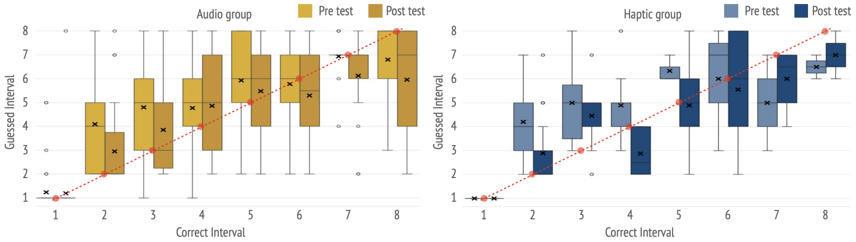

Figure 4-11: Box plots of the guessed interval for pre- and post- test broken down by interval. The left figure is for the audio control group and the right figure is for the haptic group. The cross mark indicates the mean value. Red dots and the red dashed line indicate the ground truth.

Small effect size. One limitation of this study was the short time period of training. Learning of all kinds inherently takes time, and we were limited by how long we could reasonably ask participants to listen to pairs of musical notes.

Confounding variables. Another challenge was the high variation in the prior acuity of musical notes of the participants, even though we recruited only people who did not have prior proper ear training. In post-hoc exploratory statistics, we found a correlation between experience in music and knowledge of musical intervals with the change in pre- and post-test results. We do not report on these results since they are not a primary contribution to the work. It is unclear if prior interest and experience in musical tasks reflect a difference in intrinsic motivation or existing knowledge. Future work on perceptual augmentation may wish to explore explicit counterbalancing based on exposure or intrinsic motivation.

Personalized learning. Additionally, the platform could tailor the learning experience to the skill level such as focusing on specific intervals. Some participants mentioned using existing song knowledge as a touchpoint for intervals. Future work could offer tailored haptic feedback to represent specific intervals based on prior knowledge like familiar songs or jingles [71].

Beyond learning musical intervals. We believe our wearable device can be used for applications beyond musical interval learning. Future work can investigate leveraging haptics on the back for other tasks involving the metaphor of moving in a vertical dimension, such as learning the tones of tonal languages like Mandarin.

4.6.1 Conclusion

Our findings revealed several limitations and areas for improvement in the experimental design. The game-based task, designed to simulate a sensorimotor experience akin to Guitar Hero or keyboard shortcuts, fell short of achieving the desired effect. The interface proved too cumbersome, and the task duration was insufficient to induce the anticipated sensorimotor learning. This outcome underscores the critical importance of interface design and task duration in studies of perceptual learning through embodied interaction. A recurring theme in our observations was the significance of participants' mental models and their understanding of the task's purpose. We found that when participants focused excessively on accuracy, they tended to engage in more cognitive processes rather than the intended perceptual ones. This finding aligns with previous research on the role of attention and cognitive load in perceptual learning. Despite these limitations, we maintain that pitch intervals present an excellent paradigm for evaluating perceptual learning. However, our experience suggests that a more longitudinal approach is necessary to capture the full spectrum of perceptual learning effects in this domain. This aligns with the work of Wright and Zhang [120], who demonstrated the benefits of extended training periods in auditory perceptual learning tasks.

Chapter 5

EEG Sonification for Enhancing Sleep Staging Performance

5.1 Introduction

During a night of sleep, humans transition through distinct physiological states, referred to as sleep stages. Sleep staging is a medical diagnostic procedure in which a technician classifies sleep into distinct sleep stages using a standardized set of rules applied to EEG, electromyogram, and electroculogram data. [13].

Our study aimed to examine whether adding an audio representation of the EEG to a sleep staging interface could improve sleep staging performance. Prior research has investigated developing algorithms to convert EEG into sound[49, 54, 84, 43], but has not yet addressed whether this technique can improve accuracy or reduce the workload of sleep staging.

This study integrates various concepts of multisensory learning and perception to investigate the potential for the intuitive acquisition of information for perceptual augmentation. It also marks a move towards greater scalability. While haptics offer numerous opportunities, implementing haptic solutions at scale presents significant challenges. In conducting the previous two studies, I found that sonification had many of the same benefits as haptification but provided a more scalable alternative.

The primary contribution of this is evaluating this more naturalistic approach to sonification that does very little additional processing compared to other sonification methods. More complicated musical or pattern-based encodings may increase the salience of irrelevant features.t. A simple transformation allows the brain's auditory system to naturally identify the relevant perceptual properties of the sound.

5.1.1 Manual Sleep Staging is Challenging

Sleep staging is a frequently performed medical diagnostic procedure. In the United States, it is performed more than a million times per year in order to test for conditions such as sleep apnea [106]. These stages are then used to compute clinical sleep indices such as sleep onset time, wake after sleep onset, and sleep fragmentation to diagnose specific sleep conditions such as apnea and periodic limb movement disorder [13].

Manual sleep staging is highly labor intensive and requires specialized training [97]. Furthermore, even highly trained sleep stagers show agreements only around 83% [97]. Thus, technology to improve sleep staging performance has the potential to improve both sleep staging accuracy and efficiency.

While machine learning algorithms have demonstrated good performance in sleep staging, [7] they are not used extensively in clinical practice due to concerns about their reliability and inability to provide justifications for their decisions. Regulatory agencies like the FDA have emphasized that medical AI algorithms should include a human-in-the-loop to verify decisions [41]; thus, manual sleep staging will likely continue to be necessary despite advances in AI sleep staging. It is, therefore, important to develop technologies that can facilitate faster and more accurate manual sleep staging.

5.1.2 EEG Sonification for Sleep Staging

A potential approach for improving human judgment in sleep staging is EEG sonification, or presenting an auditory representation of the EEG. We take a multi-sensory approach and present the user with both an auditory and visual representation since these can be integrated into the brain and improve the perception of a stimulus [116, 68].

There is significant prior art in sonifying bioelectric signals for analysis [23, 6], some of the earliest work on neurons used an auditory interface to monitor activations [57]. Previous research has shown that sonification can reduce the time required to classify brain activity as epileptic or normal, suggesting that sonification can reduce the workload of EEG interpretation [54, 49]. Research has also examined developing sleep-EEG sonification algorithms that balance aesthetic qualities with retaining important information in the EEG [84, 43]. However, to date, no research has examined if adding sonification to an existing visual display system for sleep staging can improve accuracy or reduce the workload of sleep staging.

5.1.3 How Could Sonification Improve Performance and Reduce Workload?

The brain combines information from sensory modalities (vision, audition, haptic, etc.) to form a unified percept of the world using multisensory integration. Multisensory integration maximizes reliability and minimizes the variance of the integrated percept [37]. Having a multisensory stream of information allows each sense to function where it is most effective. For example, audition is more temporally precise, while vision is more spatially precise [119]. These differences in the spatial and temporal resolution may make a certain sensory modality preferable for a given signal. By presenting both an auditory and visual signal, the brain can select the sensory modality with the most information for a given signal.

For example, in a loud room, we use lip reading to augment our auditory perception of speech [48]. By combining both visual and auditory information, the brain creates a more reliable percept than either modality alone. This is also confirmed by experimental results which have demonstrated that sound can enhance visual perception [116, 68]. According to the Multiple Resource Model, a multisensory presentation may also reduce workload by moving parts of a complex visual task to an alternate sensory modality [118].

One additional benefit is that this cross-modal association can be learned implicitly, without conscious attention. For example, through repeated exposure to the sight of objects falling and the accompanying sound, we develop an intuitive understanding of the physics of falling objects [9]. This means participants can continue to focus on the visual representation of the EEG and still implicitly learn and benefit from the auditory stimulus.

5.1.4 Study Design

We tested whether sonifying EEG could improve accuracy, speed, and cognitive load using an online sleep staging task where participants were asked to stage sleep either with or without sonification. Specifically, we aimed to test the following hypotheses:

> H1: Participants will perform more accurately when they view EEG and also hear a sonified representation than when they view the EEG without sonification.
>
> H2: Participants will stage faster when they view EEG and also hear a sonified representation than when they view the EEG without sonification.
>
> H3: Participants will report lower mental demand, effort, and frustration as measured by the NASA Task Load Index[53] when staging with sonification as compared to visual staging alone.
>
> H4: Benefits will be selective for sleep epochs which contain alpha waves or sleep spindles.

5.2 Methods

We preregistered our procedure prior to analysis at https://osf.io/2xz5e. Our analysis followed the procedures laid out in the preregistration with one exception. We chose not to analyze the number of transitions between epochs since fewer than 5 participants had transitions.

5.2.1 Participants and Screening Procedure

We recruited participants via advertisements distributed to sleep researchers on Twitter. Participants were eligible if they were at least 18 years of age, and reported some experience with AASM sleep staging. After observing a large number of probable bots attempting to complete the survey we also added an additional screening question requiring participants to correctly identify an obvious epoch of Wake to proceed; this question was added after the 386th participant. 536 participants completed the screening and proceeded to the practice test.

To exclude inattentive participants and bots, we excluded participants with implausibly short reaction times in the practice block. Participants were not allowed to proceed to the test blocks if their median time to stage an epoch during the practice block was less than 2 seconds. Any participants who failed this criterion were excluded from the data. The criterion was selected based on the reaction time distribution of 11 manually identified likely bots (μ of medians 0.59 seconds, $\sigma = 0.27$ seconds) vs 3 known valid sleep stagers (μ of medians 5.5 seconds, $\sigma = 2.1$ seconds).

There were 623 attempts to complete the practice block; of these, 46 met the reaction time criterion and proceeded to the test blocks. 40 of these participants completed all of the test phases and were included in the analysis. The included participants consisted of 24 males, 15 females, and 1 non-binary person with a mean age of 24 ± 3.11 years. Reported sleep staging experience is shown in Table 5.1.

333

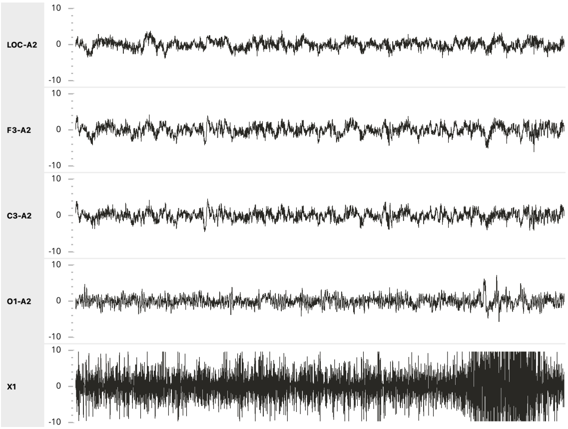

Figure 5-1: An example of the online sleep staging task interface used by participants. The O1-A2 channel was sonified using our stimulus generation procedure.

# **Polysomnograms Staged in Past Year**	**Count**
1-10	14
11-20	8
21-40	10
41-80	5
More than 80	3

Table 5.1: Count of participants by sleep staging experience level

5.2.2 Stimulus Generation Procedure

We used data from healthy participants in the ISRUC-SLEEP database [64]. We included only sleep epochs which were staged as the same sleep stage by both ISRUC-SLEEP stagers. For more details on the sleep dataset, see [64].

We created the visual and auditory representations of the data using a modified version of Visbrain Sleep[28], an open-source sleep staging tool. The visual representation included 5 channels: 3 EEG channels (F3, C3, O1), 1 electrooculogram channel (left outer canthus), and 1 chin EMG channel. All data were referenced to the right ear. All channels were scaled to ± 100 μV. An example of the visual display is shown in Figure 5-1.

To generate the audio data, we performed a minimal transformation where we converted 30 seconds of visually displayed data from the O1 channel into a 1.36-second sound clip. We used this approach rather than a more complex transformation to avoid destroying any information and to maximize participants' ability to learn. We chose a simple transformation over a musical or mapping-based approach because these risk increasing the salience of features irrelevant to sleep staging. A simple transformation allows the brain's auditory system to naturally identify the relevant properties of the sound. This approach enables the formation of perceptual boundaries based on the sound's intrinsic characteristics [2]. The original dataset was 30s epochs measured at 200Hz [65]. This was downsampled to 100Hz in Visbrain. In order to produce the auditory stimulus, we sped up the visual data by approximately 95% for a final audio length of 1.36 seconds. We also performed scaling and padding of the audio to reduce clipping and popping during playback. We observed that this method produced qualitatively distinct sounds for each sleep stage.

5.2.3 Experimental Procedure

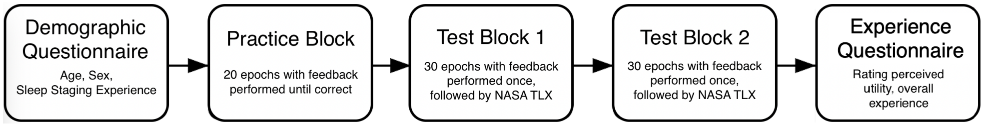

Figure 5-2: The experimental procedure.

On beginning the experiment, participants first completed a questionnaire on their age, sex, and sleep staging experience, which was quantified through two questions (self-described experience and the number of polysomnograms staged in the previous year).

Participants then proceeded to the online sleep staging task. The task began with an orientation to the display format (Figure 5-1), after which participants began the practice block of sleep staging. In this block, participants viewed and attempted to stage 20 30-second epochs of sleep. Participants were also able to toggle between the epoch to be staged, and the 30 seconds before and after this epoch as many times as desired before issuing a sleep stage. Sound corresponding to the O1 channel of the EEG was played immediately upon viewing the sleep epoch and any time the participant moved forward or backward in time.

After the participant issued a sleep stage, the sonification of the O1 channel was played again and the participant received feedback on the correct sleep stage. The participant could then proceed to the next epoch of sleep. Epochs were presented in random order until the participant issued the correct response; the epoch was then dropped from the rotation.

Once all epochs were correctly identified, the block ended.

Participants then proceeded to the first and second test blocks, which presented information identically to the practice block. Each included 30 epochs of sleep, and each epoch was shown only once, regardless of whether it was correctly staged. Participants were randomized and counterbalanced such that half received sonification on the first test block (but not the second) and half received sonification on the second test block (but not the first).

At the end of each test block, participants also performed the NASA Task Load Index [53] to assess cognitive load while performing the task. Following the completion of the last block, we asked participants for their overall impressions of the experiment.

5.2.4 Data analysis

For all statistical analyses, we compared performance in the no-sound block to performance in the sound block using a paired test. We used a Wilcoxon signed-rank test for all comparisons except for the comparison of task load index scales, where the differences between blocks met normality assumptions for using a paired t-test. For all analyses, we defined the correct sleep stage for each epoch as the stage given by the ISRUC stagers.

We measured sleep staging accuracy using Cohen's kappa to measure agreement between the participant's stages and the correct stages in each block. Cohen's kappa provides an agreement score from -1 (complete disagreement) to 1 (perfect agreement) corrected for the agreement expected from random chance [82]. Importantly, kappa compensates for the highly imbalanced classes found in sleep staging which can distort simpler measurements of percentage correct.

To determine whether participants' sleep staging experience mediated the effect of sound, we split participants into 5 categories based on the number of polysomnograms they reported staging in the prior year. We then performed a paired test comparing kappa for the sound vs no-sound block for each experience group.

When analyzing reaction times, we only included trials where the participant gave the correct sleep stage.We excluded trials with outlier reaction times (RTs) more than two sigma above or below the average reaction time consistent with standard practices. [11].

We hypothesized that sonification might selectively improve the ability to identify Wake and N2, which were associated with especially distinctive sounds. To determine if sonification improved participants' ability to identify specific sleep stages, we defined the true stage for each epoch as the stage given by the ISRUC stagers. We then measured the percent of "true stage X" that was identified as that stage by participants. For example, if participants recognized 5 of the 10 Wake epochs present in a block, their accuracy for Wake would be 50%. For this analysis, we quantified participant performance using percent correct rather than Cohen's kappa because kappa is undefined when the correct sleep stage is the same for all epochs.

We also computed how sonification affected the time required to identify stage Wake and stage N2 specifically (Figure 5-6). For these analyses, we only included trials where the RT was within 2 sigma of the mean and the sleep stage was correctly identified; 3 participants failed to correctly identify any instances of stage Wake in one or more blocks and were excluded from this analysis.

5.3 Results

5.3.1 Sonification improves staging accuracy for participants with the least experience

We found no significant difference between sleep staging performance in the sound and no-sound block [Wilcoxon signed-rank test, p=0.21], and therefore no support for H1. Similarly we did not find support for H2, as the time to complete sleep staging did not differ significantly between the sound and no-sound blocks for all epochs staged [Wilcoxon signed-rank test, p=0.65] or epochs staged correctly [Wilcoxon signed-rank test, p=0.83].

However, our exploratory analysis revealed that sound improved accuracy measured by kappa only for the group of participants who reported having staged 1-10 nights of sleep previously; for this subgroup kappa was significantly higher with sound [Wilcoxon signed-

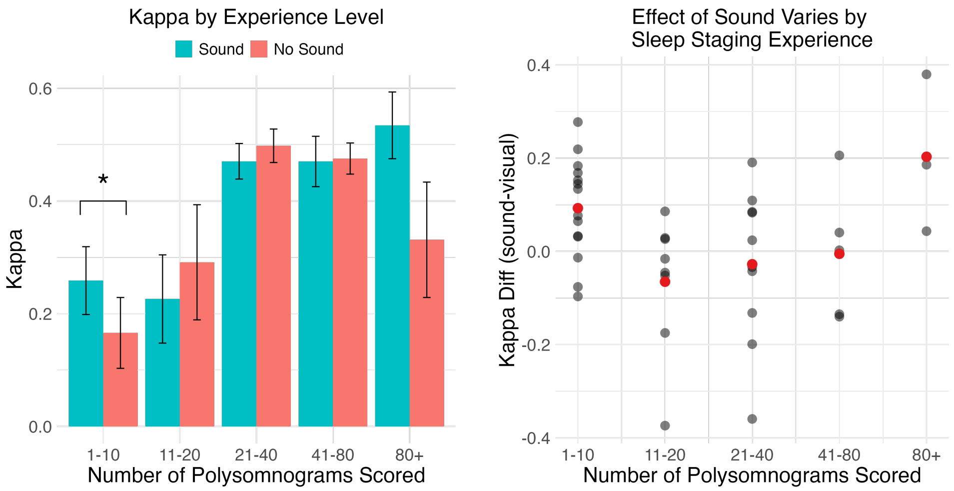

Figure 5-3: (Left) Sleep staging performance (Cohen's kappa of participants in the sound and no-sound blocks by experience level. (Right) Difference in sleep staging performance between the sound and no-sound blocks. The least experienced participants showed a significant improvement in kappa with sound (indicated by *). Higher kappa indicates better staging performance.

rank test, p=0.01] (Figure 5-3). No other group of participants showed a significant effect of sound.

5.3.2 Sonification did not significantly alter task load

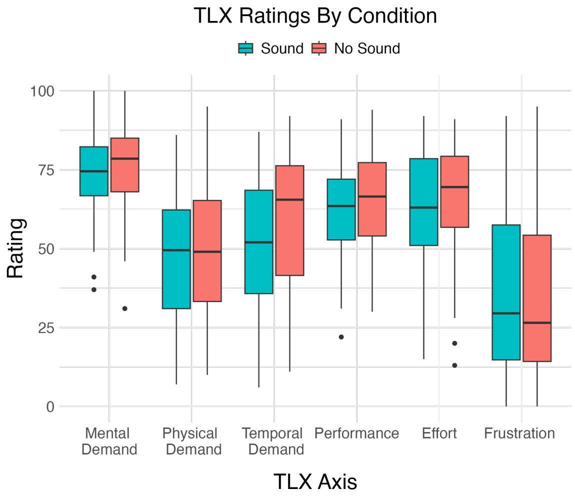

Figure 5-4: The effect of sonification on NASA TLX scores was not statistically significant. **Mental Demand** [t(39) = 1.052, p = 0.30], **Physical Demand** [t(39) = 0.524, p = 0.60], **Temporal Demand** [t(39) = 1.622, p = 0.11], **Performance** [t(39) = 0.826, p = 0.41], **Effort** [t(39) = 0.434,p = 0.67], **Frustration** [t(39) = -0.510, p = 0.61]

To test our prediction that sonification would reduce the demand of sleep staging [H3], we measured whether the NASA task load index of mental load, effort, and frustration differed between the sound block and the no-sound block. As the differences were normally distributed, we used a paired t-test. We did not observe any significant differences between the sound and no-sound conditions on any of the three axes (Figure 5-4). We also did not observe significant effects on the TLX for the subgroup of least experienced users (who reported scoring 1-10 nights of sleep).

5.3.3 Effects of sonification did not depend on sleep stage

We predicted [H4] that sonification might selectively improve the recognition of Wake and N2 as those sleep stages contain features that are particularly identifiable in sound. However, we found that neither accuracy in identifying Wake nor accuracy in identifying N2 was improved by sonification (Figure 5-5), [Wilcoxon signed-rank test |Wake: p = 0.57, N2: p = 0.28]. Therefore we did not find support for this hypothesis.

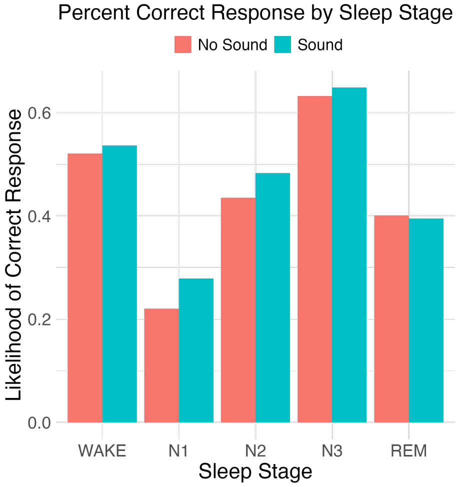

Figure 5-5: Effect of sonification on accuracy for each sleep stage. Accuracy was calculated by first finding the epochs identified as a sleep stage by the ISRUC sleep stagers and then measuring how many were correctly identified by the participant as the target sleep stage. We did not observe significant differences in the effect of sonification by sleep stage.

5.3.4 Sleep staging experience predicts test performance

We observed a significant correlation between the number of polysomnograms staged in the past year and the overall kappa across both test blocks [r(38)=0.49, p=0.001].

5.3.5 Participants found sonification useful

After completing the experiment, 30/40 participants reported some benefit from the sound in staging sleep. 26 participants reported that the sound was "Useful, a lot" for staging, while 4 participants reported it was "Useful, a little", 4 indicated that they were unsure, and 6 reported that it was not useful. 39/40 participants indicated that they would use a sleep staging platform that incorporated sonification.

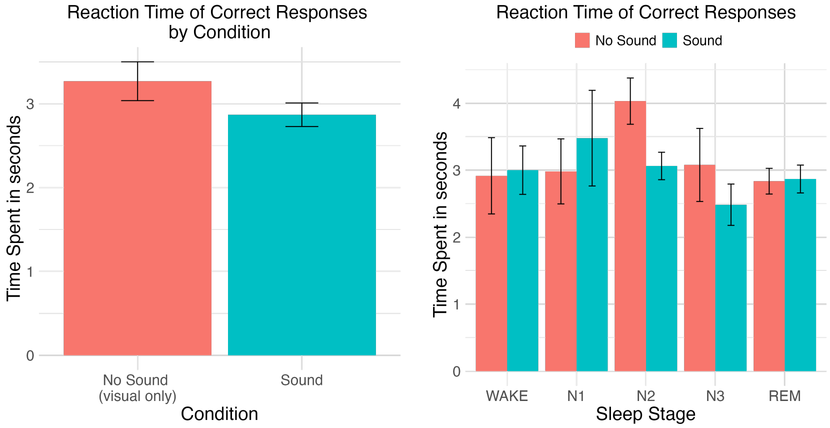

Figure 5-6: (Left) Time required to stage each epoch for correct responses. (Right) Time required to stage each epoch for correct responses by sleep stage. We did not observe significant effects of sonification for any sleep stage.

5.4 Discussion

In this study, we found that adding sonification to an EEG sleep staging task improved sleep staging accuracy only for the group of participants with the least experience. Contrary to our predictions, we did not find that sonification reduced the time to complete staging or the mental load of the staging task; though participants reported finding the sounds useful.

Why does sonification improve staging performance in novices, but not in experienced sleep stagers? One hypothesis is that experienced sleep stagers can extract all the information in the signal from its visual representation; thus, sound does not add any additional information. Another hypothesis is that experienced sleep stagers may experience a blocking effect where well-learned associations between visual features and sleep stages prevent effective learning of the relationship between sound features and sleep stages. The blocking effect is well-studied in associative learning ([62, 78]) where learning the association between a novel stimulus (such as sound) and outcome (sleep stage) is inhibited when the well-learned visual stimulus is also present.

While we did not find overall significant effects of sonification, participants did frequently report that sound was helpful for sleep staging, and we observed non-significant improvements in the time to complete sleep staging (Figure 5-6). It is possible that more experienced participants would also show benefits from sonification with a larger sample size or with a longer period of learning.

We observed that sonification enabled participants with 1-10 polysomnograms staged to perform at the same level as more experienced participants with 11-20 polysomnograms staged (Figure 5-3). Therefore, future studies can also test whether sonification can enable novice stagers to reach acceptable performance levels faster. Such a study may also show larger benefits as inexperienced sleep stagers may be better able to integrate audio and visual information without blocking effects. Future studies could also explore the effects of alternative sound transformations or whether an increased number of learning trials improves the effects of sonification.

As a compensated online study, there is a risk of participants engaging in deception to collect rewards; risks here include both automated and semi-automated bots and human participants who do not perform the task according to instructions [77]. We employed multiple measures to detect and exclude these response types, including reCAPTCHA, excluding participants who failed to identify an obvious epoch of Wake, and excluding participants who exhibited implausibly fast responses during the practice block. We also demonstrated that the reported level of sleep staging experience correlated with performance, suggesting that the bulk of our participants correctly reported their experience level and attempted to perform the sleep staging task. Nonetheless, it is possible that the recorded answers contain some non-genuine responses.

A unique feature of this study compared to prior work on the sonification of bioelectric signals is that we examined whether sonification combined with standard techniques could augment performance. In contrast, previous studies have examined the optimal techniques to create aesthetically pleasing sounds [84, 43] or whether sonification could replace visual interpretation [54, 49]. Our observation that sonification can increase sleep staging accuracy in inexperienced sleep stagers suggests that sonification may be particularly valuable for non-specialists with limited experience, such as internists and students in sleep research labs.

5.5 Future Work and Conclusion

This study explored the potential of sonification to perceptually augment EEG sleep staging performance. While our primary findings diverged from our preregistered hypotheses, they yielded valuable insights through exploratory analysis. Our results suggest several promising avenues for future research, and we plan to conduct a follow-up study, preregistering a design focused on teaching EEG sleep staging to novices instead of evaluating this perceptual augmentation for experienced sleep scorers. This approach is operationally easier and allows us to do the following: 1. Expand our participant pool to online platforms such as Prolific and Lab in the Wild 2. Address the observed lack of benefit for experienced users from additional sensory information 3. Study the results of perceptual augmentation through sonification on intrinsically motivated populations

Additionally, we aim to refine our statistical analysis methods. Although we presented our preregistered analysis as planned, the exploratory analysis revealed that linear regression treating each participant's question as a data point yielded improved results compared to the aggregative, kappa-based methods employed in this study.

In conclusion, while our initial hypotheses were not supported, this study has laid the groundwork for further investigation into the potential of sonification in EEG sleep staging for perceptual augmentation.

Chapter 6

Haptic Headphones: Designing Devices for Perceptual Augmentation

The previous chapters have explored domain-specific studies on perceptual augmentation measured by performance. The intent of these was to gain better implicit perceptual learning and how to make it more effective. The following chapters move away from these context-specific augmentation and examine a more generalized augmentation of auditory perception. This ongoing research aims to develop generalized methods for enhancing perceptual capabilities beyond task-specific contexts.

In the following section, I focus on hardware and a planned study on generalized perceptual augmentation.

This chapter presents the development of a hardware system called "haptic headphones." This hardware is designed to facilitate the exploration of auditory perceptual augmentation. By integrating haptic feedback with traditional audio output, I intend for this device to serve as a versatile tool for investigating the potential enhancement and expansion of auditory perception.

6.1 Improving Phoneme Perception for Age-Related Hearing Loss

I plan to conduct a speech-in-noise study on participants with age-related hearing loss. Age-related hearing loss (presbycusis) affects a significant portion of the elderly population, impacting their ability to communicate effectively and potentially leading to social isolation. While hearing aids and cochlear implants have significantly addressed this issue, challenges remain, particularly in complex auditory environments and distinguishing similar phonemes, as discussed earlier in Chapter 2. This study proposes a novel approach using customized haptic feedback to supplement visual and auditory input, potentially improving phoneme differentiation and overall speech perception.

This proposed study would investigate the potential of customized haptic feedback to improve phoneme differentiation in individuals with age-related hearing loss. By developing a generalized encoding technique that translates audio signals into haptic feedback, I aim to enhance the ability to distinguish between similar phonemes, particularly those that are challenging to differentiate based on the high-frequency content lost due to age-related hearing loss. This research builds upon previous work in haptic-assisted hearing [27, 40] but focuses specifically on non-semantic audio and integrating high-bandwidth, high-dimensionality information in the vibrotactile domain [94, 96].

At a high level, the procedure would be as follows:

1. Conduct pure-tone audiometry to establish hearing thresholds
2. Perform a phoneme discrimination task without haptic feedback. Present pairs of phonemes with audio/visual information (e.g., /p/ vs /k/, /m/ vs /n/, /s/ vs /z/) and ask participants to identify if they are the same or different.
3. Customize haptics based on an envelope-following method. This mapping would partially take an envelope following an approach from prior work by Fletcher et al. (2019)[40], which has demonstrated improved localization abilities. Additionally, this

method will focus on preserving spectral and temporal cues crucial for phoneme discrimination and localization [96].

4. Perform training for an hour with pre-encoded videos of television that have had the phonemes pre-encoding using an AI tool. Many speech recognition LLMs break the speech into phonemes. [1]. These videos would have various levels of noise and occlusion to encourage users to learn to perceive the phonemes actively.

6.2 Prior Work in Haptic Platforms

After completing the two studies utilizing existing haptics development platforms, it became evident that these platforms would not necessarily be effective for performing the types of experiments I am interested in because they lacked the flexibility and multisensory integration I require.

Haptic devices fundamentally require three key components: a waveform generator to create a signal, an amplifier to provide sufficient power, and a haptic actuator to produce the tactile sensation. This chapter examines two previous designs utilized in Chapters 4 and 3, serving as a foundation for discussing the improvements aimed for in the new design.

6.2.1 Overview of Hardware Used in Haptic Hat and Pitch Purrfect

In the Haptic Hat project, I used a board from Syntacts [91] in conjunction with a Focusrite Scarlett 18i20 and a voice coil actuator. The Syntacts system uses audio input from a computer to drive a multichannel sound card, which then routes the signal through a haptic driver to amplify haptic actuators. A similar setup, predating Syntacts, was used in Reed's Phonemic display [96], which utilized an off-the-shelf chip MAX98306 as an output driver.

Another system used was the Vibrotactile Haptics Platform (VHP), a modular haptics development platform [33]. VHP was implemented in the Pitch Purrfect project and comprises a Bluetooth module, microcontroller, haptic driver (AB amplifier), and current-sense haptic actuators. This design incorporated a "wave table," a repository of various pre-defined sounds, to generate haptic feedback. While fully wireless, VHP offered less flexibility in terms of input data compared to the audio-controlled system. The primary distinctions between these designs lie in their method of data transmission (Bluetooth versus direct line) and the type of amplifier employed.

6.2.2 Takeaways from These Two Haptics Projects

design should be flexible

VHP, while quite effective, did not offer the flexibility I wanted while developing. There are many unknowns, and I want to have low overhead when I am testing. Producing signals with a wavetable and the overhead of charging and transferring information over Bluetooth seem currently unnecessary. Additionally, I want my design to test a number of questions about what the most effective design is.

Smaller Form Factor and Improved Tactor Connections

The Syntacts setup [91] while functionally robust, presented practical issues. The audio interface was extensive and not portable. Additionally, the board connectors lacked durability and would often come out if transported. I want to improve the portability of this setup and have it be somewhat "wearable."

Experimenting with Number and Type of Actuators

At specific frequencies, the bone conduction actuators produced excessive audio content (as they were meant to!), which interfered with intended haptic sensations. I want my next design to use Linear Resonant Actuators (LRAs) for more focused vibrotactile feedback with a lower acoustic output. Additionally, I would like to improve the density of tactors. The previous four-tactor design limited the complexity of motion-based encoding. I will double the number of tactors in the device to enhance spatial resolution and enable more sophisticated haptic patterns.

6.3 Electronics Design of Haptic Headphones

The following section will review some of the technical considerations I had when building this system. In 6-2, I present a block diagram of the system. There are four major parts that I

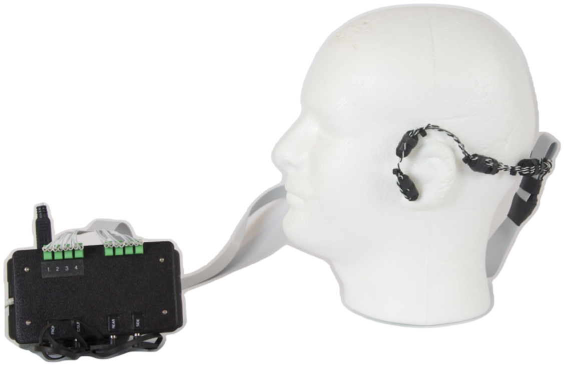

Figure 6-1: An image of the completed electromechanical haptic headphone system on a dummy

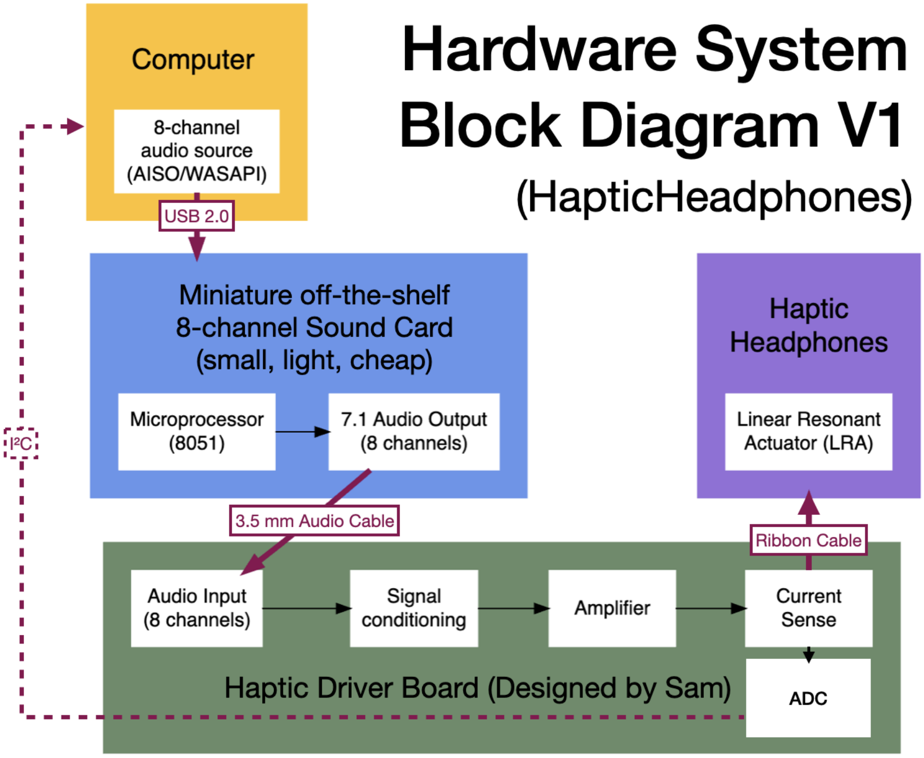

Figure 6-2: A block diagram of the Haptic Headphones hardware system

will discuss: The computer, which is used for sound generation (gold), a miniature 8-channel sound card (blue), the custom haptic driver board (green), and the haptic headphones (purple). The following sections will be organized following this diagram.

Work on this project is still ongoing, and this is a snapshot of the work. The current hardware design is a development board that I plan to use to make decisions about the final board.

6.4 Audio Sources for Driving Haptics

6.4.1 Waveform generation for audio

I have been using a digital modular synth program called VCV Rack for audio sound generation. Since this tool is used for musical performance, I can generate various waveforms dynamically. It also allows for precise tuning of multiple parameters such as level, channel, mixing, etc. Since it is based on electrical systems, this type of interface felt very natural to me.

Python is another reasonable way to generate waveforms. I used `sounddevice`, a Python package for Audio Input and Output. This can be used for mixing and addressing individual speakers in a speaker array [42].

6.4.2 Latency of Audio Driver APIs

Penzent et al. [91] characterized the latency of Windows audio drivers in Figure 6-3 and found that ASIO (Audio Stream Input/Output) and WASAPI (Windows Audio Session API) both are faster than the perceptual threshold of visual-haptic simultaneity for a variety of sound output devices / sound cards. The same work states that this latency is not an issue in the standard Mac system for audio input and output: CoreAudio.

6.5 Sound Cards

Sound cards are how a computer converts digital information into sound and may also include amplifiers to change the level of that sound. Most computers minimally have a 2-channel

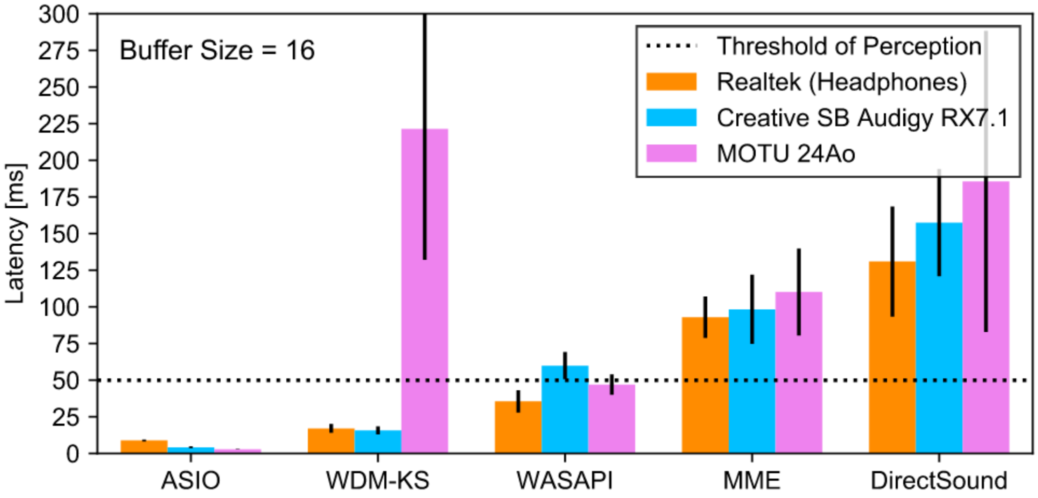

Figure 6-3: An image of the class D amplifier circuit design

sound card for a stereo output. An external sound card is needed if more than two channels are required. Most sound cards are made for audio engineering and music production, where some multiple microphones and speakers must be routed. Syntacts[91] lists some available options. However, these are still reasonably large, and the least expensive, Creative Soundblaster 7.1, is above $100.

I found a 7.1 (8-channel) audio driver on Amazon for $20 produced by a no-name company that has since disappeared. However, variations can be found when searching for "7.1 USB 3.5MM Sound Card". I have incorporated this small external sound card into the design and attached the two boxes in order to evaluate the chip handling. The USB to 8-channel conversion would be sufficient for this purpose.

I plan to design a new revision of the board in which the sound card functionality is inside the device. This would allow a user to plug the device into a USB port of any computer without sourcing a separate sound card. This seems like a significant barrier to any audio-controlled haptics device. I discuss this more in 6.8.1 of this section. Having such a small form factor also allows me to begin prototyping a device for age-related hearing loss, which I discuss more in 6.1.

6.6 Custom Haptic Driver Board

6.6.1 Audio Input and Signal Conditioning

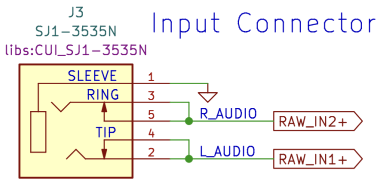

Figure 6-4: An image of a single 3.5mm audio jack that carries the signal from the sound card to the amplifier.

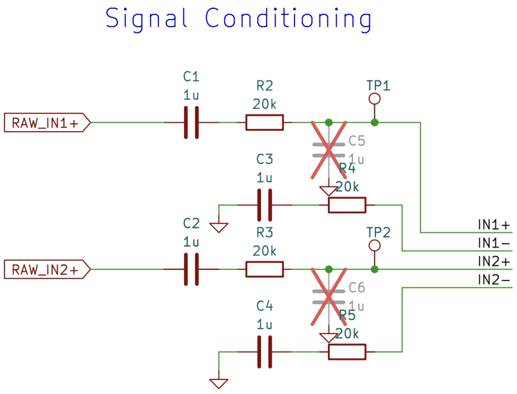

Figure 6-5: An image of the conditioning circuitry before the amplifier for two channels of the device.

The input stage consists 3.5mm Tip-Ring-Sleeve (TRS) audio jacks - each contains two channels of audio-controlled haptic input. This topology for the signal conditioning circuitry is based on the values from the Syntacts [91] design. This conditioning is important so that the input into the amplifier is in the correct range,

6.6.2 Amplifiers and Current Sensing

Amplifiers

Amplifiers provide the current to drive haptic actuators. Most of our actuators can draw up to 1W each, which is about 0.5 Amps. Our design has eight tactors, so our system could draw up to 4 Amps if all the haptics ran at full power. To provide this current, we need a dedicated circuit.

As seen in Table 6.1, the primary choice was around the amplifier type. Class AB amplifiers have been widely used in audio applications due to their linear operation and low distortion characteristics. In contrast, Class D amplifiers operate on a switching principle, rapidly

Project	**Type of Amplification**	**Part Number**
Syntacts [91]	Class AB Amplifier	TPA6211A1
Phoenemic Tactile Display [96] & Vibrotactile Haptics Platform [33]	Class D Amplifier	MAX98306

Table 6.1: Comparison of amplifiers evaluated in this thesis by project

turning the output transistors on and off at a high frequency. This switching may result in noise and deviations from the set voltage (voltage ripple) [56]. Pezent et al. (2021) [91] cited concerns with switching noise for measurements (motor encoders and EEG) as the reason they selected a Class AB amplifier.

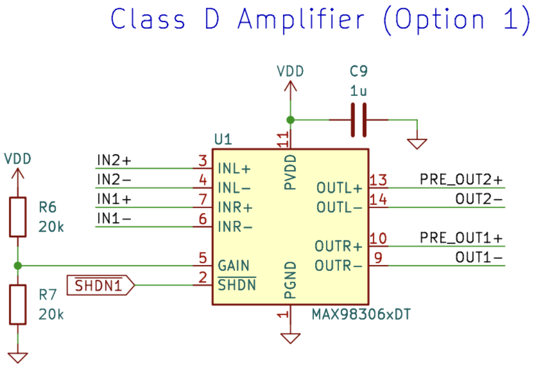

Figure 6-6: An image of the class D amplifier circuit design

For this iteration of the circuit board design, I elected to include both Class AB and Class D amplifier footprints. I wanted to evaluate potential distortions and efficiency/heat trade-offs myself, so this planned analysis is ongoing.

Current Sensing

Dementyev et al. (2020) [31] demonstrate that backEMF current sensing can be used to evaluate the fit of haptics. Proper connection to the skin is crucial for building a successful haptic device. This design follows the work of Dementyev et al. (2020) and is intended to connect to a separate Analog Digital Converter (ADC).

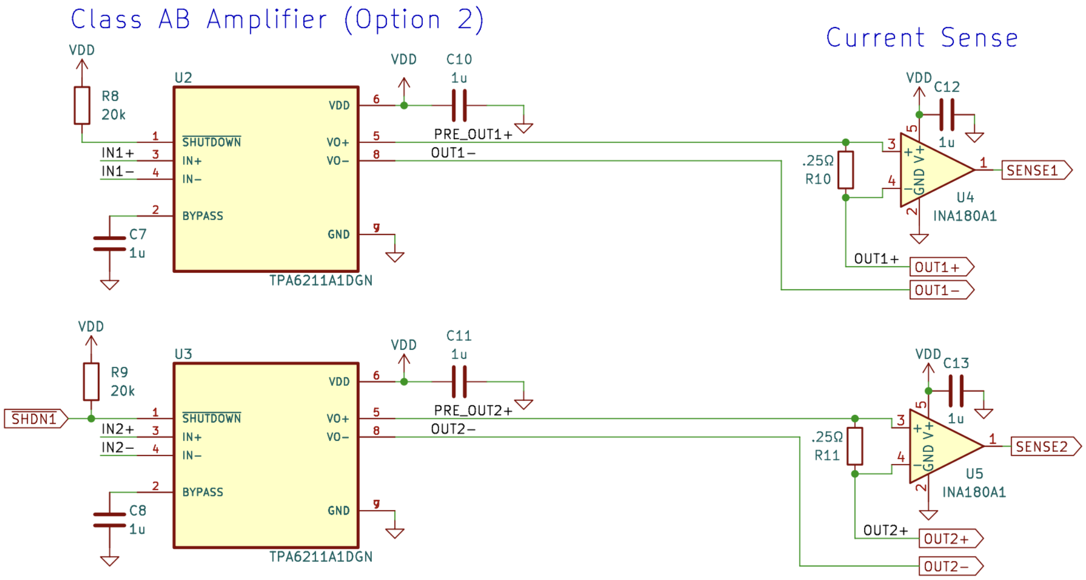

Figure 6-7: An image of the class AB amplifier circuit design with current sensing

6.7 Haptic Headphone Mechanical Design

6.7.1 Overall Mechanical Design

The mechanical design is inspired by bone conduction headphones, which enhance auditory perception without occluding natural hearing. This open-format design not only preserves the user's existing auditory capabilities but also accommodates the potential use of hearing aids. I have mounted them near the ears and around the head, following literature on embodiment and perception to leverage the benefits of sensory integration at the hearing location. The device aims to optimize sensory integration and provide a more intuitive and effective user experience by localizing tactile feedback to the hearing region.

The mechanical structure utilizes spring steel, currently bent to mimic the form factor of bone-conduction headphones. This ensures a secure and comfortable fit while maintaining an aesthetically pleasing appearance that resembles conventional headphones rather than assistive technology. The device aims to optimize sensory integration and provide a more intuitive and effective user experience by localizing tactile feedback to the hearing region.

6.7.2 Actuator Selection

Vibrotactile actuators come in several forms; however, the primary two used in haptics are LRAs and VCMs. **Linear Resonant Actuators (LRAs)** employ an electromagnet-driven mass on a spring, vibrating at specific frequencies. LRAs are both efficient in power consumption and react quickly. They generally have a small frequency range where they can best actuate. **Voice Coil actuators (VCMs)** are similar to LRAs but offer a broader frequency response, enabling richer tactile stimulation at the cost of higher power consumption. This broader frequency response may also produce unintended auditory noise when actuating the haptics. I have created an overview that summarizes the physical parameters of the vibrotactile actuators I have previously used (Table 6.2).

Name of Work	**Type of Actuator**	**Frequency Range (Hz)**	**Peak Frequency (Hz)**	V_{max} **(Volts)**	**Part Number**
Syntacts [91]	LRA	228-242	235	1.8V	VG0832022D
Phoenemic Tactile Display [96]	Voice Coil	50 - 2000	600	2.8V*	TEAX13C02-8/RH
Vibrotactile Haptics Platform [33]	LRA	150-300	170	2.5V	G1040003D
Haptic Hat	Voice Coil	300-19000	1600	2.4V	Adafruit Bone Conduction Transducer
Haptic Headphones	LRA	154 - 168	160	1.8V	VLV200634A
Haptic Headphones	LRA	150-300	170	2.5V	G1040003D

*Calculated value based on max power = 1W, impedance = 8Ω

Table 6.2: Comparison of actuators evaluated in this thesis by project

For the design of the Haptic Headphones, I chose to primarily work with LRAs due to their smaller form factor and the aforementioned audio artifacts produced by the voice coils. I still anticipate the haptics will produce sound, but hopefully, at a lower volume overall. In Figure 6-8, you can see both actuators in focus. The circular one is the G1040003D, and the rectangular one is the VG0832022D.

Similarly to the amplifiers, I have selected two different actuators to compare their performance. This characterization has not yet occurred.

Figure 6-8: A closeup featuring both actuators. The circular one is the G1040003D, and the rectangular one is the VG0832022D.

6.8 Future Plans

6.8.1 Hardware Improvements

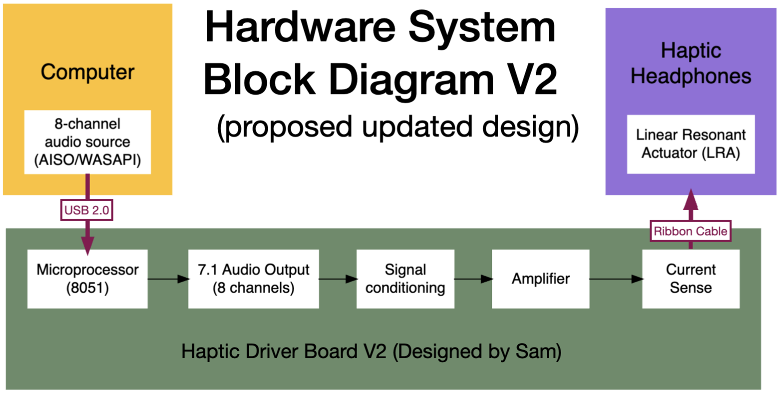

Figure 6-9: A block diagram of the proposed hardware which includes the sound card

As discussed earlier, I plan to redesign the current board using the USB to 7.1 audio chip, integrating sound card functionality directly into the device. The two candidate chips are the CM6212 and the CM6637. Both are chips that provide this desirable USB to 8-channel output. See Figure 6-9 for a more detailed diagram of the updated block diagram. See Figure /reffig:blockdiag for the current block diagram.

One other improvement I hope to make is regarding the mechanical design. Currently, this has been designed for my head, however, the tactors do not successfully touch other people as well. I intend to deconstruct existing bone conduction headphones to understand this mechanism better.

Chapter 7

Conclusion

This thesis explored the concept of perceptual augmentation through a series of studies investigating ways to enhance human sensory capabilities beyond their biological limitations. Each study provided unique insights into the potential and challenges of augmenting human perception:

The pilot study on haptic augmentation for improved situation awareness in aviation demonstrated the importance of environmental context in perceptual augmentation. While overall performance did not significantly improve with haptic feedback, the study revealed that such augmentation might be more beneficial in visually degraded conditions, such as nighttime flying. This highlighted the need for context-sensitive augmentation strategies.

The study on haptic augmentation for pitch interval learning showed promising results in improving musical perception. Participants using the haptic-audio interface performed 20% more accurately and responded 1.674 seconds faster than the audio-only control group during training. This study underscored the potential of multisensory interfaces in enhancing perceptual learning, particularly in domains requiring fine discrimination of sensory inputs.

The investigation into EEG sonification for enhancing sleep staging performance revealed that audio augmentation can selectively benefit novices in complex perceptual tasks. While experienced sleep stagers did not show significant improvement with sonification, those with the least experience demonstrated enhanced accuracy. This study highlighted the potential of perceptual augmentation in specialized fields and the importance of considering user expertise in design. These studies collectively emphasize the promise of perceptual augmentation while also revealing the complexities involved in effectively implementing such technologies. They underscore the need for careful consideration of factors such as environmental context, user expertise, and the specific perceptual task at hand. Future research should focus on developing more generalized methods for enhancing perceptual capabilities, exploring long-term learning effects, and addressing the challenges of individual differences in sensory integration and learning.

This thesis has presented the development and initial testing of a novel "haptic headphone" system designed to explore auditory perceptual augmentation through integrated haptic feedback for age-related hearing loss The hardware architecture, comprising waveform generation software, a compact 8-channel sound card, and a custom haptic driver board, represents an advancement in flexibility and portability over previous haptic platforms. The mechanical design, inspired by bone conduction technology, aims to preserve natural hearing while providing localized tactile feedback. While the system shows promise, further refinement and testing are necessary to fully evaluate its efficacy. Future work will focus on integrating sound card functionality directly into the device and improving the mechanical design for better fit across users. This platform may enable studies on generalized auditory perceptual augmentation, including a planned investigation into phoneme perception enhancement for individuals with age-related hearing loss.

Bibliography

[1] Hanan Aldarmaki et al. “Unsupervised Automatic Speech Recognition: A review”. en. In: *Speech Communication* 139 (Apr. 2022), pp. 76–91. ISSN: 01676393. DOI: `10.1016/j.specom.2022.02.005`. URL: `https://linkinghub.elsevier.com/retrieve/pii/S0167639322000292` (visited on 08/17/2024).

[2] Ishwarya Ananthabhotla. “Cognitive Audio: Enabling Auditory Interfaces with an Understanding of How We Hear”. en. Accepted: 2022-06-15T13:06:09Z. Thesis. Massachusetts Institute of Technology, Feb. 2022. URL: `https://dspace.mit.edu/handle/1721.1/143241` (visited on 09/04/2022).

[3] Ishwarya Ananthabhotla, David B. Ramsay, and Joseph A. Paradiso. “HCU400: an Annotated Dataset for Exploring Aural Phenomenology through Causal Uncertainty”. In: *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Brighton, United Kingdom: IEEE, May 2019, pp. 920–924. ISBN: 978-1-4799-8131-1. DOI: `10.1109/ICASSP.2019.8683147`. URL: `https://ieeexplore.ieee.org/document/8683147/` (visited on 03/22/2024).

[4] Vincent Arel-Bundock. *marginaleffects: Predictions, comparisons, slopes, marginal means, and hypothesis tests*. manual. 2024. URL: `https://marginaleffects.com/`.

[5] Paul Bach-Y-Rita et al. “Vision Substitution by Tactile Image Projection”. en. In: *Nature* 221.5184 (Mar. 1969). 541 citations (Crossref) [2024-02-23], pp. 963–964. ISSN: 0028-0836, 1476-4687. DOI: `10.1038/221963a0`. URL: `https://www.nature.com/articles/221963a0` (visited on 07/26/2022).

[6] Mark Ballora et al. “Heart Rate Sonification: A New Approach to Medical Diagnosis”. en. In: *Leonardo* 37.1 (Feb. 2004), pp. 41–46. ISSN: 0024-094X, 1530-9282. DOI: `10.1162/002409404772828094`. URL: `https://direct.mit.edu/leon/article/37/1/41-46/44575` (visited on 08/07/2024).

[7] Anuja Bandyopadhyay and Cathy Goldstein. “Clinical Applications of Artificial Intelligence in Sleep Medicine: A Sleep Clinician’s Perspective”. In: *Sleep and Breathing* 27.1 (Mar. 2023), pp. 39–55. ISSN: 1522-1709. DOI: `10.1007/s11325-022-02592-4`. (Visited on 06/16/2024).

[8] Douglas Bates et al. “Fitting Linear Mixed-Effects Models Using **lme4**”. en. In: *Journal of Statistical Software* 67.1 (2015). ISSN: 1548-7660. DOI: `10.18637/jss.v067.i01`. URL: `http://www.jstatsoft.org/v67/i01/` (visited on 04/03/2024).

[9] Peter W. Battaglia, Jessica B. Hamrick, and Joshua B. Tenenbaum. “Simulation as an engine of physical scene understanding”. en. In: *Proceedings of the National Academy of Sciences* 110.45 (Nov. 2013), pp. 18327–18332. ISSN: 0027-8424, 1091-6490. DOI: `10.1073/pnas.1306572110`. URL: `https://pnas.org/doi/full/10.1073/pnas.1306572110` (visited on 07/08/2024).

[10] Benoit Bediou et al. “Meta-analysis of action video game impact on perceptual, attentional, and cognitive skills.” en. In: *Psychological Bulletin* 144.1 (Jan. 2018), pp. 77–110. ISSN: 1939-1455, 0033-2909. DOI: `10.1037/bul0000130`. URL: `https://doi.apa.org/doi/10.1037/bul0000130` (visited on 08/08/2024).

[11] Alexander Berger and Markus Kiefer. “Comparison of Different Response Time Outlier Exclusion Methods: A Simulation Study”. In: *Frontiers in Psychology* 12 (June 2021). ISSN: 1664-1078. DOI: `10.3389/fpsyg.2021.675558`. (Visited on 07/17/2024).

[12] Matthias Berning et al. “ProximityHat: a head-worn system for subtle sensory augmentation with tactile stimulation”. en. In: *Proceedings of the 2015 ACM International Symposium on Wearable Computers - ISWC ’15*. Osaka, Japan: ACM Press, 2015, pp. 31–38. ISBN: 978-1-4503-3578-2. DOI: `10.1145/2802083.2802088`. URL: `http://dl.acm.org/citation.cfm?doid=2802083.2802088` (visited on 02/26/2024).

[13] Richard B Berry et al. "The AASM Manual for the Scoring of Sleep and Associated Events". In: *Rules, Terminology and Technical Specifications, Darien, Illinois, American Academy of Sleep Medicine* 176.2012 (2012), p. 7.

[14] M. Bertero, T.A. Poggio, and V. Torre. "Ill-posed problems in early vision". In: *Proceedings of the IEEE* 76.8 (Aug. 1988), pp. 869–889. ISSN: 00189219. DOI: `10.1109/5.5962`. URL: `http://ieeexplore.ieee.org/document/5962/` (visited on 08/06/2024).

[15] Craig Bertram et al. "Sensory Augmentation with Distal Touch: The Tactile Helmet Project". en. In: *Biomimetic and Biohybrid Systems*. Ed. by David Hutchison et al. Vol. 8064. Series Title: Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 24–35. ISBN: 978-3-642-39801-8. DOI: `10.1007/978-3-642-39802-5_3`. URL: `http://link.springer.com/10.1007/978-3-642-39802-5_3` (visited on 02/26/2024).

[16] Myriam Bikah, M. Susan Hallbeck, and John H. Flowers. "Supracutaneous vibrotactile perception threshold at various non-glabrous body loci". en. In: *Ergonomics* 51.6 (June 2008), pp. 920–934. ISSN: 0014-0139, 1366-5847. DOI: `10.1080/00140130701809341`. URL: `https://www.tandfonline.com/doi/full/10.1080/00140130701809341` (visited on 08/02/2024).

[17] Jens Blauert. *Spatial Hearing: The Psychophysics of Human Sound Localization*. en. The MIT Press, 1997. ISBN: 978-0-262-26868-4. DOI: `10.7551/mitpress/6391.001.0001`. URL: `https://direct.mit.edu/books/book/4885/Spatial-HearingThe-Psychophysics-of-Human-Sound` (visited on 09/23/2022).

[18] Neil R. Bramley et al. "Intuitive experimentation in the physical world". en. In: *Cognitive Psychology* 105 (Sept. 2018), pp. 9–38. ISSN: 00100285. DOI: `10.1016/j.cogpsych.2018.05.001`. URL: `https://linkinghub.elsevier.com/retrieve/pii/S001002851730347X` (visited on 08/05/2024).

[19] J. Christopher Brill, Ben D. Lawson, and Angus Rupert. "Tactile Situation Awareness System (TSAS) as a Compensatory Aid for Sensory Loss". en. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 58.1 (Sept. 2014). 1

citations (Crossref) [2024-02-23], pp. 1028–1032. ISSN: 2169-5067, 1071-1813. DOI: `10.1177/1541931214581215`. URL: `http://journals.sagepub.com/doi/10.1177/1541931214581215` (visited on 07/26/2022).

[20] Patrick Bruns. “The Ventriloquist Illusion as a Tool to Study Multisensory Processing: An Update”. In: *Frontiers in Integrative Neuroscience* 13 (Sept. 2019). 32 citations (Crossref) [2024-02-23], p. 51. ISSN: 1662-5145. DOI: `10.3389/fnint.2019.00051`. URL: `https://www.frontiersin.org/article/10.3389/fnint.2019.00051/full` (visited on 10/23/2022).

[21] Edward M Burns and W Dixon Ward. “Categorical perception—phenomenon or epiphenomenon: Evidence from experiments in the perception of melodic musical intervals”. In: *The Journal of the Acoustical Society of America* 63.2 (1978), pp. 456–468.

[22] Joe B Buttram. “The influence of selected factors on interval identification”. In: *Journal of Research in Music Education* 17.3 (1969), pp. 309–315.

[23] Juliana Cherston and Joseph A. Paradiso. “Rotator: Flexible Distribution of Data Across Sensory Channels”. en. In: *Proceedings of the 23rd International Conference on Auditory Display - ICAD 2017*. University Park Campus: The International Community for Auditory Display, June 2017, pp. 86–93. ISBN: 978-0-9670904-4-3. DOI: `10.21785/icad2017.009`. URL: `http://hdl.handle.net/1853/58351` (visited on 08/07/2024).

[24] Seungmoon Choi and Katherine J Kuchenbecker. “Vibrotactile display: Perception, technology, and applications”. In: *Proceedings of the IEEE* 101.9 (2012), pp. 2093–2104.

[25] Seungmoon Choi and Katherine J. Kuchenbecker. “Vibrotactile Display: Perception, Technology, and Applications”. In: *Proceedings of the IEEE* 101.9 (Sept. 2013). 298 citations (Crossref) [2024-02-23], pp. 2093–2104. ISSN: 0018-9219, 1558-2256. DOI: `10.1109/JPROC.2012.2221071`. URL: `http://ieeexplore.ieee.org/document/6353870/` (visited on 01/22/2023).

[26] Adrien Chopin, Benoit Bediou, and Daphne Bavelier. "Altering perception: the case of action video gaming". en. In: *Current Opinion in Psychology* 29 (Oct. 2019), pp. 168–173. ISSN: 2352250X. DOI: `10.1016/j.copsyc.2019.03.004`. URL: `https://linkinghub.elsevier.com/retrieve/pii/S2352250X18302331` (visited on 08/08/2024).

[27] K. Cieśla et al. "Effects of training and using an audio-tactile sensory substitution device on speech-in-noise understanding". en. In: *Scientific Reports* 12.1 (Dec. 2022). 13 citations (Crossref) [2024-02-23], p. 3206. ISSN: 2045-2322. DOI: `10.1038/s41598-022-06855-8`. URL: `https://www.nature.com/articles/s41598-022-06855-8` (visited on 09/03/2022).

[28] Etienne Combrisson et al. "Visbrain: A Multi-Purpose GPU-Accelerated Open-Source Suite for Multimodal Brain Data Visualization". In: *Frontiers in Neuroinformatics* 13 (Mar. 2019), p. 14. ISSN: 1662-5196. DOI: `10.3389/fninf.2019.00014`. URL: `https://www.frontiersin.org/article/10.3389/fninf.2019.00014/full` (visited on 07/01/2024).

[29] Brendan Conlon et al. "Different bimodal neuromodulation settings reduce tinnitus symptoms in a large randomized trial". en. In: *Scientific Reports* 12.1 (June 2022), p. 10845. ISSN: 2045-2322. DOI: `10.1038/s41598-022-13875-x`. URL: `https://www.nature.com/articles/s41598-022-13875-x` (visited on 08/07/2024).

[30] Kevin Connolly. "Perceptual Learning". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Summer 2017. Metaphysics Research Lab, Stanford University, 2017.

[31] Artem Dementyev, Alex Olwal, and Richard F. Lyon. "Haptics with Input: Back-EMF in Linear Resonant Actuators to Enable Touch, Pressure and Environmental Awareness". en. In: *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*. 12 citations (Crossref) [2024-02-23]. Virtual Event USA: ACM, Oct. 2020, pp. 420–429. ISBN: 978-1-4503-7514-6. DOI: `10.1145/3379337.3415823`. URL: `https://dl.acm.org/doi/10.1145/3379337.3415823` (visited on 09/03/2022).

[32] Artem Dementyev et al. "VHP: Vibrotactile Haptics Platform for On-body Applications". In: *The 34th Annual ACM Symposium on User Interface Software and Technology*. 2021, pp. 598–612.

[33] Artem Dementyev et al. "VHP: Vibrotactile Haptics Platform for On-body Applications". en. In: *The 34th Annual ACM Symposium on User Interface Software and Technology*. 2 citations (Crossref) [2024-02-23]. Virtual Event USA: ACM, Oct. 2021, pp. 598–612. ISBN: 978-1-4503-8635-7. DOI: `10.1145/3472749.3474772`. URL: `https://dl.acm.org/doi/10.1145/3472749.3474772` (visited on 01/22/2023).

[34] Sophie Deneve and Alexandre Pouget. "Bayesian multisensory integration and cross-modal spatial links". en. In: *Journal of Physiology-Paris* 98.1-3 (Jan. 2004), pp. 249–258. ISSN: 09284257. DOI: `10.1016/j.jphysparis.2004.03.011`. URL: `https://linkinghub.elsevier.com/retrieve/pii/S0928425704000841` (visited on 04/03/2024).

[35] Oliver Doehrmann and Marcus J. Naumer. "Semantics and the multisensory brain: How meaning modulates processes of audio-visual integration". en. In: *Brain Research* 1242 (Nov. 2008), pp. 136–150. ISSN: 00068993. DOI: `10.1016/j.brainres.2008.03.071`. URL: `https://linkinghub.elsevier.com/retrieve/pii/S0006899308008007` (visited on 08/06/2024).

[36] David M. Eagleman. 2022. URL: `https://www.youtube.com/watch?v=_K5qcm2kF6E`.

[37] Marc O. Ernst and Martin S. Banks. "Humans integrate visual and haptic information in a statistically optimal fashion". en. In: *Nature* 415.6870 (Jan. 2002), pp. 429–433. ISSN: 0028-0836, 1476-4687. DOI: `10.1038/415429a`. URL: `https://www.nature.com/articles/415429a` (visited on 06/16/2024).

[38] Cathy Mengying Fang et al. "An Accessible, Three-Axis Plotter for Enhancing Calligraphy Learning through Generated Motion". In: *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 2024, pp. 1–15.

[39] Connor Fletcher, Vedad Hulusic, and Panos Amelidis. "Virtual Reality Ear Training System: A study on Spatialised Audio in Interval Recognition". In: *2019 11th In-*

ternational Conference on Virtual Worlds and Games for Serious Applications (VS-Games). 2019, pp. 1–4. DOI: `10.1109/VS-Games.2019.8864592`.

[40] Connor Fletcher, Vedad Hulusic, and Panos Amelidis. "Virtual reality ear training system: A study on spatialised audio in interval recognition". In: *2019 11th international conference on virtual worlds and games for serious applications (VS-Games).* 5 citations (Crossref) [2024-02-23]. 2019, pp. 1–4. DOI: `10.1109/VS-Games.2019.8864592`.

[41] US Food, Drug Administration, et al. "Good Machine Learning Practice for Medical Device Development". In: *US Food and Drug Administration [Internet]* (2021).

[42] Andrew Francl and Josh H. McDermott. "Deep neural network models of sound localization reveal how perception is adapted to real-world environments". en. In: *Nature Human Behaviour* 6.1 (Jan. 2022). 22 citations (Crossref) [2024-02-23], pp. 111–133. ISSN: 2397-3374. DOI: `10.1038/s41562-021-01244-z`. URL: `https://www.nature.com/articles/s41562-021-01244-z` (visited on 08/24/2022).

[43] Pedro Franco and Alpo Värri. "Experiments of the Sonification of the Sleep Electroencephalogram". In: *Finnish Journal of eHealth and eWelfare* 7.2-3 (May 2015), pp. 65–74. ISSN: 1798-0798. (Visited on 06/16/2024).

[44] A Ghazanfar and C Schroeder. "Is neocortex essentially multisensory?" en. In: *Trends in Cognitive Sciences* 10.6 (June 2006), pp. 278–285. ISSN: 13646613. DOI: `10.1016/j.tics.2006.04.008`. URL: `https://linkinghub.elsevier.com/retrieve/pii/S1364661306001045` (visited on 08/01/2024).

[45] Eleanor Jack Gibson and Anne D. Pick. *An ecological approach to perceptual learning and development.* eng. OCLC: 646787762. Oxford: Oxford University Press, 2003. ISBN: 978-0-19-511825-4.

[46] James J. Gibson. "Observations on active touch." en. In: *Psychological Review* 69.6 (Nov. 1962), pp. 477–491. ISSN: 1939-1471, 0033-295X. DOI: `10.1037/h0046962`. URL: `https://doi.apa.org/doi/10.1037/h0046962` (visited on 08/04/2024).

[47] James J. Gibson. *The Ecological Approach to Visual Perception: Classic Edition.* en. 1st ed. Psychology Press, Nov. 2014. ISBN: 978-1-315-74021-8. DOI: `10.4324/9781315740218`. URL: `https://www.taylorfrancis.com/books/9781315740218` (visited on 08/04/2024).

[48] E. Bruce Goldstein and Laura Cacciamani. *Sensation and perception.* eng. Eleventh edition, student edition. Boston, MA: Cengage, 2022. ISBN: 978-0-357-44647-8.

[49] Sergi Gomez-Quintana et al. *A Method for AI-Assisted Human Interpretation of Biological Signals: Analysis of Neonatal EEG.* Feb. 2022. DOI: `10.21203/rs.3.rs-1232994/v1`. (Visited on 06/16/2024).

[50] Timothy D. Griffiths et al. "How Can Hearing Loss Cause Dementia?" en. In: *Neuron* 108.3 (Nov. 2020), pp. 401–412. ISSN: 08966273. DOI: `10.1016/j.neuron.2020.08.003`. URL: `https://linkinghub.elsevier.com/retrieve/pii/S0896627320306103` (visited on 09/06/2024).

[51] Troy A. Hackett. "Information flow in the auditory cortical network". en. In: *Hearing Research* 271.1-2 (Jan. 2011), pp. 133–146. ISSN: 03785955. DOI: `10.1016/j.heares.2010.01.011`. URL: `https://linkinghub.elsevier.com/retrieve/pii/S0378595510000249` (visited on 08/05/2024).

[52] Sandra G Hart and Lowell E Staveland. "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research". In: *Advances in psychology.* Vol. 52. Elsevier, 1988, pp. 139–183.

[53] Sandra G. Hart and Lowell E. Staveland. "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research". en. In: *Advances in Psychology.* Vol. 52. Elsevier, 1988, pp. 139–183. ISBN: 978-0-444-70388-0. DOI: `10.1016/S0166-4115(08)62386-9`. URL: `https://linkinghub.elsevier.com/retrieve/pii/S0166411508623869` (visited on 01/22/2023).

[54] Kyle Hobbs et al. "Rapid Bedside Evaluation of Seizures in the ICU by Listening to the Sound of Brainwaves: A Prospective Observational Clinical Trial of Ceribell's Brain Stethoscope Function". In: *Neurocritical Care* 29.2 (Oct. 2018), pp. 302–312. ISSN: 1556-0961. DOI: `10.1007/s12028-018-0543-7`. (Visited on 06/16/2024).

[55] Paul M. Hofman, Jos G.A. Van Riswick, and A. John Van Opstal. "Relearning sound localization with new ears". en. In: *Nature Neuroscience* 1.5 (Sept. 1998). 304 citations (Crossref) [2024-02-23], pp. 417–421. ISSN: 1097-6256, 1546-1726. DOI: `10.1038/1633`. URL: `https://www.nature.com/articles/nn0998_417` (visited on 08/02/2022).

[56] Paul Horowitz and Winfield Hill. *The art of electronics*. Third edition. New York, NY: Cambridge University Press, 2015. ISBN: 978-0-521-80926-9.

[57] David H. Hubel. "Tungsten Microelectrode for Recording from Single Units". en. In: *Science* 125.3247 (Mar. 1957), pp. 549–550. ISSN: 0036-8075, 1095-9203. DOI: `10.1126/science.125.3247.549`. URL: `https://www.science.org/doi/10.1126/science.125.3247.549` (visited on 06/26/2024).

[58] Kimberly A. Jameson, Susan M. Highnote, and Linda M. Wasserman. "Richer color experience in observers with multiple photopigment opsin genes". en. In: *Psychonomic Bulletin & Review* 8.2 (June 2001), pp. 244–261. ISSN: 1069-9384, 1531-5320. DOI: `10.3758/BF03196159`. URL: `http://link.springer.com/10.3758/BF03196159` (visited on 08/01/2024).

[59] S. Jennings et al. "Flight-Test of a Tactile Situational Awareness System in a Land-based Deck Landing Task". en. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 48.1 (Sept. 2004), pp. 142–146. ISSN: 2169-5067, 1071-1813. DOI: `10.1177/154193120404800131`. URL: `http://journals.sagepub.com/doi/10.1177/154193120404800131` (visited on 02/26/2024).

[60] K Johnson. "The roles and functions of cutaneous mechanoreceptors". In: *Current Opinion in Neurobiology* 11.4 (Aug. 2001), pp. 455–461. ISSN: 09594388. DOI: `10.1016/S0959-4388(00)00234-8`. URL: `https://linkinghub.elsevier.com/retrieve/pii/S0959438800002348` (visited on 08/02/2024).

[61] Gabriele Jordan and John Mollon. "Tetrachromacy: the mysterious case of extra-ordinary color vision". en. In: *Current Opinion in Behavioral Sciences* 30 (Dec. 2019). 7 citations (Crossref) [2024-02-23], pp. 130–134. ISSN: 23521546. DOI: `10.1016/j.cobeha.2019.08.002`. URL: `https://linkinghub.elsevier.com/retrieve/pii/S2352154619300270` (visited on 09/26/2022).

[62] L. J. Kamin. *Attention-like Processes in Classical Conditioning.* Tech. rep. TR-5. June 1967. (Visited on 07/08/2024).

[63] Gary Steven Karpinski. *Aural skills acquisition: The development of listening, reading, and performing skills in college-level musicians.* Oxford University Press, USA, 2000.

[64] Sirvan Khalighi et al. "ISRUC-Sleep: A Comprehensive Public Dataset for Sleep Researchers". In: *Computer Methods and Programs in Biomedicine* 124 (Nov. 2015). DOI: `10.1016/j.cmpb.2015.10.013`.

[65] Sirvan Khalighi et al. "ISRUC-Sleep: A comprehensive public dataset for sleep researchers". en. In: *Computer Methods and Programs in Biomedicine* 124 (Feb. 2016), pp. 180–192. ISSN: 01692607. DOI: `10.1016/j.cmpb.2015.10.013`. URL: `https://linkinghub.elsevier.com/retrieve/pii/S0169260715002734` (visited on 07/01/2024).

[66] Mead C. Killion et al. "Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners". en. In: *The Journal of the Acoustical Society of America* 116.4 (Oct. 2004), pp. 2395–2405. ISSN: 0001-4966, 1520-8524. DOI: `10.1121/1.1784440`. URL: `https://pubs.aip.org/jasa/article/116/4/2395/545472/Development-of-a-quick-speech-in-noise-test-for` (visited on 08/05/2024).

[67] David Kirsh. "Knowledge, explicit vs implicit". In: (2009).

[68] Armin Kohlrausch and Steven Van De Par. "Audio—Visual Interaction in the Context of Multi-Media Applications". en. In: *Communication Acoustics.* Ed. by Jens Blauert. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 109–138. ISBN: 978-3-540-22162-3. DOI: `10.1007/3-540-27437-5_5`. URL: `http://link.springer.com/10.1007/3-540-27437-5_5` (visited on 08/05/2024).

[69] Rosa Lafer-Sousa, Katherine L. Hermann, and Bevil R. Conway. "Striking individual differences in color perception uncovered by 'the dress' photograph". en. In: *Current Biology* 25.13 (June 2015), R545–R546. ISSN: 09609822. DOI: `10.1016/j.`

`cub.2015.04.053`. URL: `https://linkinghub.elsevier.com/retrieve/pii/S0960982215005357` (visited on 08/06/2024).

[70] S. J. Lederman and R. L. Klatzky. “Haptic perception: A tutorial”. en. In: *Attention, Perception & Psychophysics* 71.7 (Oct. 2009), pp. 1439–1459. ISSN: 1943-3921, 1943-393X. DOI: `10.3758/APP.71.7.1439`. URL: `http://link.springer.com/10.3758/APP.71.7.1439` (visited on 08/02/2024).

[71] Daniel J Levitin. “Absolute memory for musical pitch: Evidence from the production of learned melodies”. In: *Perception & Psychophysics* 56.4 (1994), pp. 414–423.

[72] Chuan-Ming Li et al. “Hearing Impairment Associated With Depression in US Adults, National Health and Nutrition Examination Survey 2005-2010”. en. In: *JAMA Otolaryngology–Head & Neck Surgery* 140.4 (Apr. 2014), p. 293. ISSN: 2168-6181. DOI: `10.1001/jamaoto.2014.42`. URL: `http://archotol.jamanetwork.com/article.aspx?doi=10.1001/jamaoto.2014.42` (visited on 09/06/2024).

[73] Hao Li et al. “Three-dimensional tonotopic mapping of the human cochlea based on synchrotron radiation phase-contrast imaging”. en. In: *Scientific Reports* 11.1 (Feb. 2021), p. 4437. ISSN: 2045-2322. DOI: `10.1038/s41598-021-83225-w`. URL: `https://www.nature.com/articles/s41598-021-83225-w` (visited on 08/05/2024).

[74] Jianhua Li and Sophia W Deng. “Facilitation and interference effects of the multisensory context on learning: a systematic review and meta-analysis”. In: *Psychological Research* (2022), pp. 1–19.

[75] Jianhua Li and Sophia W. Deng. “Facilitation and interference effects of the multisensory context on learning: a systematic review and meta-analysis”. en. In: *Psychological Research* (Sept. 2022). 2 citations (Crossref) [2024-02-23]. ISSN: 0340-0727, 1430-2772. DOI: `10.1007/s00426-022-01733-4`. URL: `https://link.springer.com/10.1007/s00426-022-01733-4` (visited on 01/19/2023).

[76] David F Little, Henry H Cheng, and Beverly A Wright. “Inducing musical-interval learning by combining task practice with periods of stimulus exposure alone”. In: *Attention, Perception, & Psychophysics* 81 (2019), pp. 344–357.

[77] Gemma Loebenberg et al. “Bot or Not? Detecting and Managing Participant Deception When Conducting Digital Research Remotely: Case Study of a Randomized Controlled Trial”. In: *Journal of Medical Internet Research* 25.1 (Sept. 2023), e46523. DOI: `10.2196/46523`. (Visited on 07/09/2024).

[78] David Luque et al. “The Blocking Effect in Associative Learning Involves Learned Biases in Rapid Attentional Capture”. In: *Quarterly Journal of Experimental Psychology* 71.2 (Feb. 2018), pp. 522–544. ISSN: 1747-0218. DOI: `10.1080/17470218.2016.1262435`. (Visited on 07/08/2024).

[79] D. A. Mahns et al. “Vibrotactile Frequency Discrimination in Human Hairy Skin”. en. In: *Journal of Neurophysiology* 95.3 (Mar. 2006), pp. 1442–1450. ISSN: 0022-3077, 1522-1598. DOI: `10.1152/jn.00483.2005`. URL: `https://www.physiology.org/doi/10.1152/jn.00483.2005` (visited on 08/02/2024).

[80] Josh H. McDermott et al. “Musical intervals and relative pitch: Frequency resolution, not interval resolution, is special”. In: *The Journal of the Acoustical Society of America* 128.4 (Oct. 2010), pp. 1943–1951. ISSN: 0001-4966. DOI: `10.1121/1.3478785`. eprint: `https://pubs.aip.org/asa/jasa/article-pdf/128/4/1943/13782920/1943\_1\_online.pdf`. URL: `https://doi.org/10.1121/1.3478785`.

[81] Harry Mcgurk and John Macdonald. “Hearing lips and seeing voices”. en. In: *Nature* 264.5588 (Dec. 1976), pp. 746–748. ISSN: 0028-0836, 1476-4687. DOI: `10.1038/264746a0`. URL: `https://www.nature.com/articles/264746a0` (visited on 06/16/2024).

[82] Mary L. McHugh. “Interrater Reliability: The Kappa Statistic”. In: *Biochemia Medica* 22.3 (Oct. 2012), pp. 276–282. ISSN: 1330-0962. (Visited on 06/25/2024).

[83] Ken’ichi Miyazaki. “Perception of musical intervals by absolute pitch possessors”. In: *Music Perception* 9.4 (1992), pp. 413–426.

[84] Foad Moradi et al. “A Novel Method for Sleep-Stage Classification Based on Sonification of Sleep Electroencephalogram Signals Using Wavelet Transform and Recurrent Neural Network”. In: *European Neurology* 83.5 (Oct. 2020), pp. 468–486. ISSN: 0014-3022. DOI: `10.1159/000511306`. (Visited on 06/16/2024).

[85] A. Moscatelli, M. Mezzetti, and F. Lacquaniti. “Modeling psychophysical data at the population-level: The generalized linear mixed model”. en. In: *Journal of Vision* 12.11 (Oct. 2012), pp. 26–26. ISSN: 1534-7362. DOI: `10.1167/12.11.26`. URL: `http://jov.arvojournals.org/Article.aspx?doi=10.1167/12.11.26` (visited on 04/03/2024).

[86] Mark I. Nikolic and Nadine B. Sarter. “Peripheral Visual Feedback: A Powerful Means of Supporting Effective Attention Allocation in Event-Driven, Data-Rich Environments”. en. In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 43.1 (Mar. 2001). 42 citations (Crossref) [2024-02-23], pp. 30–38. ISSN: 0018-7208, 1547-8181. DOI: `10.1518/001872001775992525`. URL: `http://journals.sagepub.com/doi/10.1518/001872001775992525` (visited on 07/26/2022).

[87] Sam Norman-Haignere, Nancy Kanwisher, and Josh H. McDermott. “Cortical Pitch Regions in Humans Respond Primarily to Resolved Harmonics and Are Located in Specific Tonotopic Regions of Anterior Auditory Cortex”. en. In: *The Journal of Neuroscience* 33.50 (Dec. 2013), pp. 19451–19469. ISSN: 0270-6474, 1529-2401. DOI: `10.1523/JNEUROSCI.2880-13.2013`. URL: `https://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.2880-13.2013` (visited on 08/05/2024).

[88] Sam Norman-Haignere, Nancy G. Kanwisher, and Josh H. McDermott. “Distinct Cortical Pathways for Music and Speech Revealed by Hypothesis-Free Voxel Decomposition”. en. In: *Neuron* 88.6 (Dec. 2015), pp. 1281–1296. ISSN: 08966273. DOI: `10.1016/j.neuron.2015.11.035`. URL: `https://linkinghub.elsevier.com/retrieve/pii/S0896627315010715` (visited on 08/05/2024).

[89] Wilder Penfield and Theodore Rasmussen. “The cerebral cortex of man; a clinical study of localization of function.” In: (1950). Publisher: Macmillan.

[90] Matevž Pesek et al. “Troubadour: A Gamified e-Learning Platform for Ear Training”. In: *IEEE Access* 8 (2020), pp. 97090–97102. DOI: `10.1109/ACCESS.2020.2994389`.

[91] Evan Pezent, Brandon Cambio, and Marcia K. O’Malley. “Syntacts: Open-Source Software and Hardware for Audio-Controlled Haptics”. In: *IEEE Transactions on Haptics* 14.1 (Jan. 2021), pp. 225–233. ISSN: 1939-1412, 2329-4051, 2334-0134. DOI:

10.1109/TOH.2020.3002696. URL: https://ieeexplore.ieee.org/document/9117187/ (visited on 08/16/2024).

[92] Myrthe A Plaisier, Lotte IN Sap, and Astrid ML Kappers. “Perception of vibrotactile distance on the back”. In: *Scientific Reports* 10.1 (2020), p. 17876.

[93] Ella Pomplun et al. “Vibrotactile Perception for Sensorimotor Augmentation: Perceptual Discrimination of Vibrotactile Stimuli Induced by Low-Cost Eccentric Rotating Mass Motors at Different Body Locations in Young, Middle-Aged, and Older Adults”. In: *Frontiers in Rehabilitation Sciences* 3 (July 2022), p. 895036. ISSN: 2673-6861. DOI: 10.3389/fresc.2022.895036. URL: https://www.frontiersin.org/articles/10.3389/fresc.2022.895036/full (visited on 08/02/2024).

[94] Ville Pulkki, Leo McCormack, and Raimundo Gonzalez. “Superhuman spatial hearing technology for ultrasonic frequencies”. en. In: *Scientific Reports* 11.1 (Dec. 2021). 8 citations (Crossref) [2024-02-23], p. 11608. ISSN: 2045-2322. DOI: 10.1038/s41598-021-90829-9. URL: http://www.nature.com/articles/s41598-021-90829-9 (visited on 07/27/2022).

[95] Anil K. Raj, Steven J. Kass, and James F. Perry. “Vibrotactile Displays for Improving Spatial Awareness”. en. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 44.1 (July 2000), pp. 181–184. ISSN: 2169-5067, 1071-1813. DOI: 10.1177/154193120004400148. URL: http://journals.sagepub.com/doi/10.1177/154193120004400148 (visited on 02/26/2024).

[96] Charlotte M. Reed et al. “A Phonemic-Based Tactile Display for Speech Communication”. In: *IEEE Transactions on Haptics* 12.1 (Jan. 2019). 48 citations (Crossref) [2024-02-23], pp. 2–17. ISSN: 1939-1412, 2329-4051, 2334-0134. DOI: 10.1109/TOH.2018.2861010. URL: https://ieeexplore.ieee.org/document/8423203/ (visited on 01/22/2023).

[97] Richard S. Rosenberg and Hout Steven Van. “The American Academy of Sleep Medicine Inter-scorer Reliability Program: Sleep Stage Scoring”. In: *Journal of Clinical Sleep Medicine* 09.01 (), pp. 81–87. DOI: 10.5664/jcsm.2350. (Visited on 06/25/2024).

[98] Melissa Saenz and Dave R.M. Langers. “Tonotopic mapping of human auditory cortex”. en. In: *Hearing Research* 307 (Jan. 2014), pp. 42–52. ISSN: 03785955. DOI: `10.1016/j.heares.2013.07.016`. URL: `https://linkinghub.elsevier.com/retrieve/pii/S0378595513001871` (visited on 08/05/2024).

[99] Caitlyn Seim, Tanya Estes, and Thad Starner. “Towards passive haptic learning of piano songs”. In: *2015 IEEE World Haptics Conference (WHC)*. IEEE. 2015, pp. 445–450.

[100] Lior Shmuelof and John W Krakauer. “Recent insights into perceptual and motor skill learning”. In: *Frontiers in human neuroscience* 8 (2014), p. 683.

[101] Susan E. Shore, Larry E. Roberts, and Berthold Langguth. “Maladaptive plasticity in tinnitus — triggers, mechanisms and treatment”. en. In: *Nature Reviews Neurology* 12.3 (Mar. 2016), pp. 150–160. ISSN: 1759-4758, 1759-4766. DOI: `10.1038/nrneurol.2016.12`. URL: `https://www.nature.com/articles/nrneurol.2016.12` (visited on 08/07/2024).

[102] Jane A Siegel and William Siegel. “Absolute identification of notes and intervals by musicians”. In: *Perception & Psychophysics* 21 (1977), pp. 143–152.

[103] Daniel Smilkov et al. *Embedding Projector: Interactive Visualization and Interpretation of Embeddings.* Version Number: 1. 2016. DOI: `10.48550/ARXIV.1611.05469`. URL: `https://arxiv.org/abs/1611.05469` (visited on 08/08/2024).

[104] Oliver Snyder et al. “Assessment of surface rendering with 1 DoF vibration”. In: *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization* 9.4 (2021), pp. 400–406.

[105] Amy Szarkowski and Patrick Brice. “Positive Psychology in Research with the Deaf Community: An Idea Whose Time Has Come”. en. In: *The Journal of Deaf Studies and Deaf Education* 23.2 (Apr. 2018), pp. 111–117. ISSN: 1081-4159, 1465-7325. DOI: `10.1093/deafed/enx058`. URL: `https://academic.oup.com/jdsde/article/23/2/111/4844093` (visited on 09/06/2024).

[106] Naoko Tachibana, Najib T. Ayas, and David P. White. "A Quantitative Assessment of Sleep Laboratory Activity in the United States". In: *Journal of Clinical Sleep Medicine* 01.01 (Jan. 2005), pp. 23–26. DOI: `10.5664/jcsm.26292`. (Visited on 07/16/2024).

[107] Hong Z. Tan et al. "Methodology for Maximizing Information Transmission of Haptic Devices: A Survey". In: *Proceedings of the IEEE* 108.6 (June 2020). 28 citations (Crossref) [2024-02-23], pp. 945–965. ISSN: 0018-9219, 1558-2256. DOI: `10.1109/JPROC.2020.2992561`. URL: `https://ieeexplore.ieee.org/document/9103350/` (visited on 10/06/2022).

[108] Sundeep Teki et al. "Navigating the Auditory Scene: An Expert Role for the Hippocampus". en. In: *The Journal of Neuroscience* 32.35 (Aug. 2012), pp. 12251–12257. ISSN: 0270-6474, 1529-2401. DOI: `10.1523/JNEUROSCI.0082-12.2012`. URL: `https://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.0082-12.2012` (visited on 08/01/2024).

[109] Joshua B. Tenenbaum et al. "How to Grow a Mind: Statistics, Structure, and Abstraction". en. In: *Science* 331.6022 (Mar. 2011), pp. 1279–1285. ISSN: 0036-8075, 1095-9203. DOI: `10.1126/science.1192788`. URL: `https://www.science.org/doi/10.1126/science.1192788` (visited on 08/05/2024).

[110] Dishita G Turakhia et al. "Adapt2Learn: A Toolkit for Configuring the Learning Algorithm for Adaptive Physical Tools for Motor-Skill Learning". In: *Designing Interactive Systems Conference 2021*. DIS '21. Virtual Event, USA: Association for Computing Machinery, 2021, pp. 1301–1312. ISBN: 9781450384766. DOI: `10.1145/3461778.3462128`. URL: `https://doi.org/10.1145/3461778.3462128`.

[111] Sayako Ueda, Hiroyuki Sakai, and Takatsune Kumada. "A Novel Approach to Sensorimotor Skill Acquisition Utilizing Sensory Substitution: A Driving Simulation Study". en. In: *Scientific Reports* 9.1 (Dec. 2019). 1 citations (Crossref) [2024-02-23], p. 17886. ISSN: 2045-2322. DOI: `10.1038/s41598-019-54324-6`. URL: `http://www.nature.com/articles/s41598-019-54324-6` (visited on 07/26/2022).

[112] Tomer D. Ullman et al. "Mind Games: Game Engines as an Architecture for Intuitive Physics". en. In: *Trends in Cognitive Sciences* 21.9 (Sept. 2017). 99 citations (Cross-

ref) [2024-02-23], pp. 649–665. ISSN: 13646613. DOI: `10.1016/j.tics.2017.05.012`. URL: `https://linkinghub.elsevier.com/retrieve/pii/S1364661317301134` (visited on 03/28/2023).

[113] Chiara Valzolgher et al. "Reaching to sounds in virtual reality: A multisensory-motor approach to promote adaptation to altered auditory cues". en. In: *Neuropsychologia* 149 (Dec. 2020), p. 107665. ISSN: 00283932. DOI: `10.1016/j.neuropsychologia.2020.107665`. URL: `https://linkinghub.elsevier.com/retrieve/pii/S0028393220303377` (visited on 08/08/2024).

[114] Tiziana Vercillo et al. "Enhanced auditory spatial localization in blind echolocators". en. In: *Neuropsychologia* 67 (Jan. 2015), pp. 35–40. ISSN: 00283932. DOI: `10.1016/j.neuropsychologia.2014.12.001`. URL: `https://linkinghub.elsevier.com/retrieve/pii/S0028393214004552` (visited on 08/01/2024).

[115] Rommy Von Bernhardi, Laura Eugenín-von Bernhardi, and Jaime Eugenín. "What Is Neural Plasticity?" In: *The Plastic Brain*. Ed. by Rommy Von Bernhardi, Jaime Eugenín, and Kenneth J Muller. Vol. 1015. Series Title: Advances in Experimental Medicine and Biology. Cham: Springer International Publishing, 2017, pp. 1–15. ISBN: 978-3-319-62815-8. DOI: `10.1007/978-3-319-62817-2_1`. URL: `http://link.springer.com/10.1007/978-3-319-62817-2_1` (visited on 09/03/2024).

[116] Jean Vroomen and Beatrice De Gelder. "Sound enhances visual perception: Cross-modal effects of auditory organization on vision." en. In: *Journal of Experimental Psychology: Human Perception and Performance* 26.5 (2000), pp. 1583–1590. ISSN: 1939-1277, 0096-1523. DOI: `10.1037/0096-1523.26.5.1583`. URL: `https://doi.apa.org/doi/10.1037/0096-1523.26.5.1583` (visited on 06/26/2024).

[117] Jonathon P. Whitton et al. "Audiomotor Perceptual Training Enhances Speech Intelligibility in Background Noise". en. In: *Current Biology* 27.21 (Nov. 2017). 49 citations (Crossref) [2024-02-23], 3237–3247.e6. ISSN: 09609822. DOI: `10.1016/j.cub.2017.09.014`. URL: `https://linkinghub.elsevier.com/retrieve/pii/S0960982217311788` (visited on 02/19/2024).

[118] Christopher D. Wickens. “Multiple Resources and Mental Workload”. en. In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 50.3 (June 2008), pp. 449–455. ISSN: 0018-7208, 1547-8181. DOI: `10.1518/001872008X288394`. URL: `http://journals.sagepub.com/doi/10.1518/001872008X288394` (visited on 02/26/2024).

[119] Ilana B. Witten and Eric I. Knudsen. “Why Seeing Is Believing: Merging Auditory and Visual Worlds”. en. In: *Neuron* 48.3 (Nov. 2005), pp. 489–496. ISSN: 08966273. DOI: `10.1016/j.neuron.2005.10.020`. URL: `https://linkinghub.elsevier.com/retrieve/pii/S0896627305008858` (visited on 06/26/2024).

[120] Beverly A Wright and Yuxuan Zhang. “A review of the generalization of auditory learning”. en. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 364.1515 (Feb. 2009), pp. 301–311. ISSN: 0962-8436, 1471-2970. DOI: `10.1098/rstb.2008.0262`. URL: `https://royalsocietypublishing.org/doi/10.1098/rstb.2008.0262` (visited on 08/08/2024).

[121] Robert J Zatorre and Andrea R Halpern. “Identification, discrimination, and selective adaptation of simultaneous musical intervals”. In: *Perception & Psychophysics* 26.5 (1979), pp. 384–395.