MusiKeys: Exploring Haptic-to-Auditory Sensory Substitution to Improve Mid-Air Text-Entry

Alexander Krasner and Joseph Gabbard

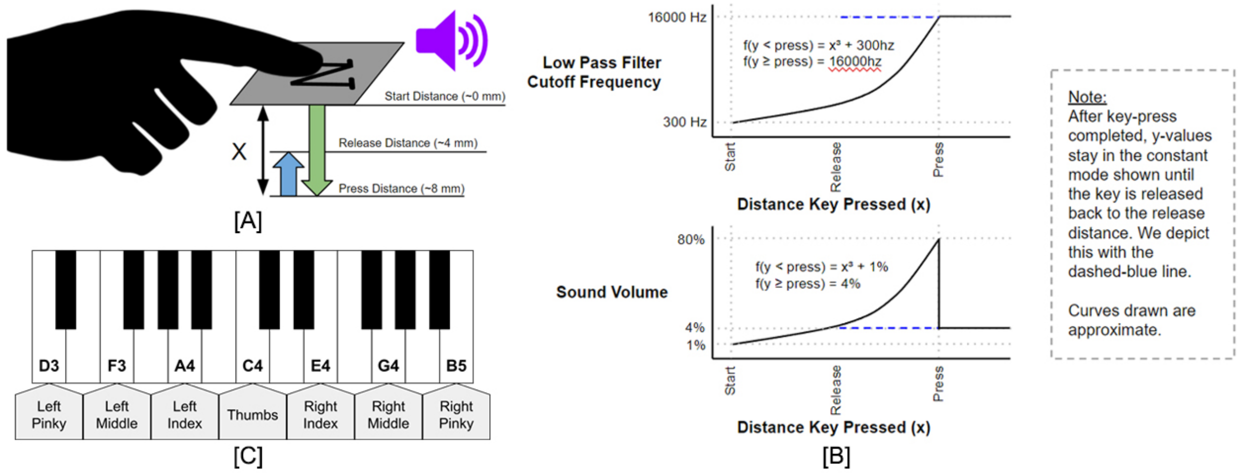

Fig. 1: Auditory augmentation design parameters. [A] Model of key-travel used by MusiKeys technique. [B] Modulation curves and functions applied to tones as keys are depressed further. [C] Mapping of fingers to notes for *Musical Sound* keyboard.

Abstract—Physical QWERTY keyboards are the current standard for performing precision text-entry with extended reality devices. Ideally, there would exist a comparable, self-contained solution that works anywhere, without requiring external keyboards. Unfortunately, when physical keyboards are recreated virtually, we currently lose critical haptic feedback information from the sense of touch, which impedes typing. In this paper, we introduce the MusiKeys Technique, which uses auditory feedback in virtual reality to communicate missing haptic feedback information typists normally receive when using a physical keyboard. To examine this concept, we conducted a user study with 24 participants which encompassed four mid-air virtual keyboards augmented with increasing amounts of feedback information, along with a fifth physical keyboard for reference. Results suggest that providing clicking feedback on key-press and key-release improves typing performance compared to not providing auditory feedback, which is consistent with the literature. We also found that audio can serve as a substitute for information contained in haptic feedback, in that users can accurately perceive the presented information. However, under our specific study conditions, this awareness of the feedback information did not yield significant differences in typing performance. Our results suggest this kind of feedback replacement can be perceived by users but needs more research to tune and improve the specific techniques.

Index Terms—Virtual reality, Extended reality, Spatial computing, Mid-air text-entry, Sensory substitution, Human-computer interaction (HCI)

1 Introduction

In modern life, precision text entry is an essential task that often goes unnoticed in its ubiquity. People rely on physical and simulated keyboards to compose long documents, fill out forms and spreadsheets, and write the code that drives our increasingly digital lifestyle. For many of these precision tasks, a typist needs to be able to specify exact characters, and thus often cannot rely on voice transcription or predictive keyboards. These tasks are currently most efficiently accomplished through use of a physical QWERTY keyboard.

In the coming decades, Augmented Reality (AR) and Virtual Reality (VR) technologies, often described together as Extended Reality (XR), are positioned to join laptops and smartphones as a primary form of personal, mobile computing [6] [21]. However, though XR head worn displays (HWDs) can provide audio and visual feedback, HWDs, as self contained systems, lack the ability to present haptic feedback to bare hands of a user. This creates an unfortunate obstacle for precision text-entry with this technology given that haptic feedback represents important perceptual feedback mechanisms when typing [10]. If we display a "mid-air keyboard", i.e. a purely virtual QWERTY keyboard floating in the air that reacts to touch in a realistic way, the lack of haptic feedback makes it very difficult to use effectively when compared to a physical keyboard [10, 22]. This aspect of XR HWDs has led to an absence of an established best practice for self-contained precision text entry in XR, and connecting an external physical QWERTY keyboard has become the current standard for performing these kinds of tasks.

When considering a future with mobile XR HWDs, it is critical that we develop a viable solution for self-contained precision text-entry that can be used anywhere. Given that a primary differentiating factor between physical and virtual keyboards is the presence or absence of haptic feedback, it is possible that providing information typically delivered via haptic feedback to typists via other means could augment the typing experience in XR. For example, we could substitute the

- *Alexander Krasner is with Virginia Tech. E-mail: akrasner19@gmail.com*
- *Joseph Gabbard is with Virginia Tech. E-mail: jgabbard@vt.edu*

Manuscript received 4 October 2023; revised 17 January 2024; accepted 24 January 2024. Date of publication 4 March 2024; date of current version 15 April 2024. This article has supplementary downloadable material available at https://doi.org/10.1109/TVCG.2024.3372065, provided by the authors. Digital Object Identifier no. 10.1109/TVCG.2024.3372065

missing haptic information with a type of feedback an XR HWD is capable of rendering, such as audio and/or visual. Since typing tasks are already visually demanding, auditory feedback is the ideal candidate. This approach led us to the following research questions:

- **RQ1:** Can information typically delivered through a haptic modality be presented via audio to improve typing on a mid-air QWERTY keyboard?
- **RQ2:** How does the amount of haptic information communicated through audio affect typing on a mid-air QWERTY keyboard?

Thus, in this research we explore the possibility of improving XR mid-air keyboards by targeting the information afforded to a typist via haptic feedback when using a physical keyboard, and encoding that information into auditory feedback presented while using a virtual mid-air keyboard in VR. We investigated this space via our solution "MusiKeys", a set of auditory feedback augmentations to a mid-air QWERTY keyboard. MusiKeys consists of four types of auditory feedback, *No Sound*, *Click Sound*, *Key-travel Sound*, and *Musical Sound*, each encoding a larger set of afforded information from physical keyboards into the auditory feedback. We evaluated this solution via a 24 participant within-subjects study where we assessed typist performance, typist workload, interface usability, and typist awareness of the feedback information.

2 Related Work

2.1 Mid-air Text Entry

Virtual text-entry is a well established problem in the field of XR interface research [4, 7, 28], with researchers continuing to experiment on new ideas for future standards of typing.

Mid-air text entry, a subset of this field, seeks to break away from physical keyboards and controllers, bringing the mobility of the text-entry system to the same level of mobility as the XR HWD itself and allowing for the consolidation of hardware. There have been a wide range of approaches researched in this area.

One approach is called unconstrained typing, where no virtual keyboard is displayed at all. Gil et al. [10] utilize a method where text is entered by capturing the hand motions of touch-typists who are mimicking typing motions in mid-air, then translating this data into key-presses via machine learning. This method was shown to be effective, with participants achieving 78.3% of their physical keyboard typing speeds. Singhal et al. [22] improved the capability further by applying machine learning correction to the finger classifiers based on what words are expected. These solutions are promising but create new restrictions that are not present with physical keyboards. For those performing precision text-entry, it is important that they know which key is being pressed for precise input, rather than having a predictive system introduce error. Yi et al. [29] use auditory and visual feedback for their similar study where they use Bayesian inference to detect key-presses in mid-air, though they did not isolate the feedback to find it effect it provides.

Another way researchers have approached the mid-air typing problem is to try to provide alternative forms of haptic feedback. Dudley et al. [8] found that using a flat surface for passive haptic feedback yielded better performance than typing in the air, though it is to be noted that this not a mid-air typing solution. They also found that using all 10 fingers when typing in the air performed worse than just the two index fingers, and suggests this is because it is harder to visually attend to all of the fingers during a typing task. This leaves open the possibility that we can attend to a greater number of fingers through audio, since typing does not normally require our auditory attention. Passive haptic feedback can also be accomplished through finger-to-finger contact, as shown by Fashimpaur et al. [9]. They mapped each finger to its corresponding set of characters in a QWERTY keyboard. Their study relies on a language model to predict the words being typed, which will not allow for precision text entry. Though we target precision inputs, we use the idea of mapping of fingers to sets of QWERTY characters as a method to compensate for tracking unreliability in our study. A similar study uses passive haptics of a users palm, but due to lack of other feedback, resulted in a very slow 4.3-6.2 Words-Per-Minute [26]. Another way to create haptic feedback is through vibro-tactile feedback to the hands via transducers, small devices that can provide isolated points of vibration. The goal is to give a similar kind of feedback as typing on a real keyboard. Though transducers are often placed on the hands [11], Sand et al. [18] placed an array of ultrasonic transducers on a VR headset to provide a standalone system with mid-air tactile feedback. The experiment was limited to only pressing six numbered virtual buttons instead of a full keyboard, and the study found that though there were not significant differences in performance, participants preferred the haptic feedback.

One particular study uses vibro-tactile haptics in a way that is very similar to how MusiKeys employs sound. Gupta et al. [11] evaluated text-entry on a mid-air keyboard in VR where participants had vibro-tactile feedback on different locations including fingertips, wrist with spatialized and unspatialized vibrations, and no vibrations. They provided visual and vibrational feedback of touching, pressing, and releasing the keys, including a continuous signal while touching. However, they only included auditory feedback for completed key-presses. They found that all versions were comparable, but that participants preferred the haptic feedback to no haptic feedback. They also found that the tactile feedback enabled participants to have better awareness of hand position relative to the keyboard, something we investigate in our study. It is notable that the visual component did not seem to contribute to this effect, likely because typist visual attention is on the text they are typing and not the keyboard. In contrast, audio, like vibration, does not require visual focus to be perceived. The auditory feedback in this study did not have this same depth of feedback signals as the other kinds of feedback, which leaves an open question: what will happen if the key-touch and key-release signals were also communicated through audio? This is exactly what we address in our work.

2.2 Auditory Feedback in XR

Auditory feedback is a staple element of digital user interface design, but it becomes even more important in XR, since there is no physical feedback. In a VR text-entry study using controllers to cast rays and select keys, Yildirim [30] found that providing auditory feedback when switching between keys increased accuracy, and that auditory feedback when clicking keys increased text-entry speed. It is important to note that the same study included a haptic vibration condition which performed equally well and in some cases better, but this does not discount the significant effect of auditory feedback. We can also relate their method of providing feedback when switching between keys to being a way of replacing the physical feedback one may get when running a finger across a keyboard. Similarly, Kim and Tan [13] examined applying auditory feedback to typing on a flat surface with vibro-tactile and auditory feedback. They found that, though the vibrations did communicate the feedback of pressing keys best, that auditory feedback also improved performance above the passive haptics of touching a flat surface. In a study by Zhang et al. [31], auditory feedback improved performance and awareness during virtual part assembly. It seems clearly established in the research that audio is an effective way to communicate system status, though it does generally have performance trade-offs when compared to haptic feedback. This pattern appears in our results.

Studies have also demonstrated that modifying the pitch of auditory feedback can be a valid approach to communicate information to a user. Batmaz et al. [2] found that providing differently pitched auditory feedback in a VR pointing experiment, using the pitch to encode speed and accuracy of the pointing selection, was effective in controlling how participants performed. This lends credence to the idea of pitch as a method of encoding information that falls on some kind of spectrum into the pitch of auditory feedback. While their study used it to show distance away from the desired performance threshold, it is possible it could instead map to distance to the left/right, for example. Batmaz and Stuerzlinger [1] also analyzed what ranges of pitch were most effective for use as feedback, finding that the lower pitches of a musical C8 note had participants respond more slowly, compared to a musical C4 note. They also conclude that it is important for auditory studies to report the pitches of sounds used. Based on this study, we follow their advice and report our pitches used in MusiKeys, as well as centering our selected musical notes around C4 to optimize effectiveness in the time sensitive task of typing.

2.3 Sensory Substitution

Throughout section 2 we have been able to relate the findings of some studies to sensory substitution, i.e. the idea of directly substituting sensory signals from one modality for another. This concept is at the core of the MusiKeys technique, as we try to replace the sense of touch with hearing. However, in contrast to our research which targets the XR typing experience, much of the work surrounding sensory substitution focuses on helping people with disabilities that bring sensory limitations, such as blindness or hearing impairments [15]. Belardinelli et al. [16] reviews various ways auditory feedback could be used to replace the information contained in visual feedback to help visually-impaired people explore spaces. Aside from methods using auditory feedback, research has tried to substitute haptic feedback as a substitute for similar cases [20, 25]. Thome et al. [25] used haptic feedback to assist users with high-frequency hearing loss. One study on music therapy also involves replacement of feedback types [14]. The study mentions visual and auditory feedback replacement for musical instruments and demonstrates successful association between the new cues and the instruments.

Meanwhile, there are certain studies that may implement forms of sensory substitution in XR, though none mention their inclusion of it by name. Shen et al. [19] used the color of a virtual interaction point to signify if a participant's finger had passed through the mid-air keyboard or was still above it. It is possible to view this as encoding tactile feedback into a visual form. Bermejo et al. [3] examined the effects of wrist-based haptics on virtual button presses for mid-air number entry. They included wrist vibrations to communicate touching as well as pressing the virtual buttons, very similarly to Gupta [11]. Given that wrist-based haptic representations are not directly linked to the sensations of fingers that should be experiencing touch feedback, there is room for exploration into presenting similar feedback via sound. In fact, most studies that substitute the sense of touch in one area seem to be replacing it with the sense of touch at a different area.

Despite much work in sensory substitution, there is a clear gap in the literature regarding audio as a direct substitute for touch feedback information in XR typing.

3 MusiKeys User Study

We conducted a study to evaluate the efficacy of the MusiKeys technique in augmenting mid-air typing. ***MusiKeys*** refers to our technique of using audio to target and replace the lost afforded feedback information when transitioning from using a physical keyboard to a mid-air keyboard. At its core, MusiKeys focuses on the specific haptic-information afforded by a physical keyboard and encodes that information into audio. For this study, we created 4 different mid-air VR keyboards with increasing levels of information encoding. The particular feedback signals and their respective keyboards are plotted in Table 1 and described in more detail in subsection 3.3.

3.1 Experimental Design

The study was designed as a within-subjects study with Keyboard Type as the independent variable with four conditions: *No Sound*, *Click Sound*, *Key-Travel Sound*, and *Musical Sound* (described below in section subsection 3.3). The order of presentation of these conditions was counterbalanced across 24 participants. There was an additional, fifth, typing condition of *Physical Baseline* keyboard, but it was treated as a separate baseline measurement of current real-world typing performance, and thus was excluded from the counterbalancing. Because of issues with hand-tracking reliability, we determined that direct comparison of virtual keyboards to a physical keyboard could not be made fair, so we opted to always present the physical baseline condition first. This exception was chosen to provide a stable reference point for prior typing skill of our participants. This study was approved by the authors' institutional review board.

3.2 Virtual Keyboard Design

The virtual keyboard for the MusiKeys technique consists of a set of virtual keys arranged in a QWERTY layout, and set at an angle of 17 degrees offset from horizontal towards the user. The Oculus SDK built-in hand-tracking V2 algorithm was used to track the position and movements of hands and fingers to allow for natural interaction with the keys. The Meta Quest 2 and its associated tracking system were selected as the development platform since they performed best among the available hand-tracking solutions at the time the application began development, including when tested alongside the Ultraleap hand-tracking sensor. Despite being the best available choice for tracking, there were a number of issues with the reliability of the hand-tracking when using this system. We found that ring-fingers were unusable for typing, since the tracking algorithm failed to detect any bending of them. Additionally, the pinky-fingers only tracked to a usable degree when fully outstretched from the hand. In contrast, middle-fingers tracked fairly reliably, while index-fingers and thumbs featured especially robust tracking. These compensation techniques were concluded to viable compromises because the focus of our study was on the viability of the auditory feedback substitution rather than pure mid-air typing performance.

Several constraints were placed on the keyboard to compensate for the above-mentioned issues with hand-tracking reliability. First, we eliminated the ability to press any keys used by the ring-fingers and pinky-fingers in the standard "home-row" typing technique. "Home-row" technique refers to a standard typing method taught in grade-school where typists orient their fingers above the keys in the center-row of letters, called the home-row, and certain columns of letters are designated to be typed by certain fingers. We greatly reduced the opacity for the disabled keys, as seen in Figure 2. This was done to make the usable keys visually distinct while still providing a sense of familiar shape to the keyboard so participants would know where to position their hands. We included an enlarged backspace key, intended to be pressed by the right pinky-finger, since pinky tracking worked when the finger was outstretched from the hand, and the location of the backspace key required that hand-shape to reach it. In testing we also discovered it was common to accidentally type letters with the thumb, since a user has no physical space-bar to rest the thumb on in mid-air. We fixed this by enlarging the space-bar and removing the ability of the thumbs to press anything except the space-bar.

The virtual keys themselves were designed to emulate various aspects of physical keyboards that are often missing from mid-air keyboards. Each key was represented as a small flat plane in 3D space, approximately the size of the keys on an Apple Macbook Pro. These planes were physics objects constrained to move in a linear path perpendicular to the keyboard. We refer to the movement of a key as "key-travel". The path and stages of key-travel can be seen in Figure 1, item A. When a key was pressed deep enough below the plane of the keyboard, it would trigger a key-press, typing the character. The depth to trigger a key-press is called the "press distance", which was calibrated from pilot test feedback. After a key-press, the key needed to be released a set amount before it could be typed again. This buffer area, called the "release distance", prevented accidental rapid, repeated presses from shaky hands or jittery tracking. By implementing this system, we model the states of a physical key on a mechanical keyboard: a finger can be touching a key but not yet pressing it down, a key travels a set distance down until it is pressed, and a key must be released a set distance before it can be pressed again.

The size of the virtual keyboard, the keys, and the key-press depth were determined through iterative pilot testing. Because the goal was to investigate the auditory augmentations rather than just mid-air typing performance, sizing parameters of the keyboard were initialized to match an apple magic keyboard, then adjusted during pilot-testing to maximize reported usability, given the unreliable hand-tracking. The final measurements determined that the keys were spaced out 2.3 cm from key-center to key-center, with each letter key being a square measuring 1.22 cm across. The enlarged spacebar was 8 cm across, and 4.3 cm tall, while the enlarged backspace was a square, 3 cm across. Key-press distance was set at 8mm, while key-release distance was

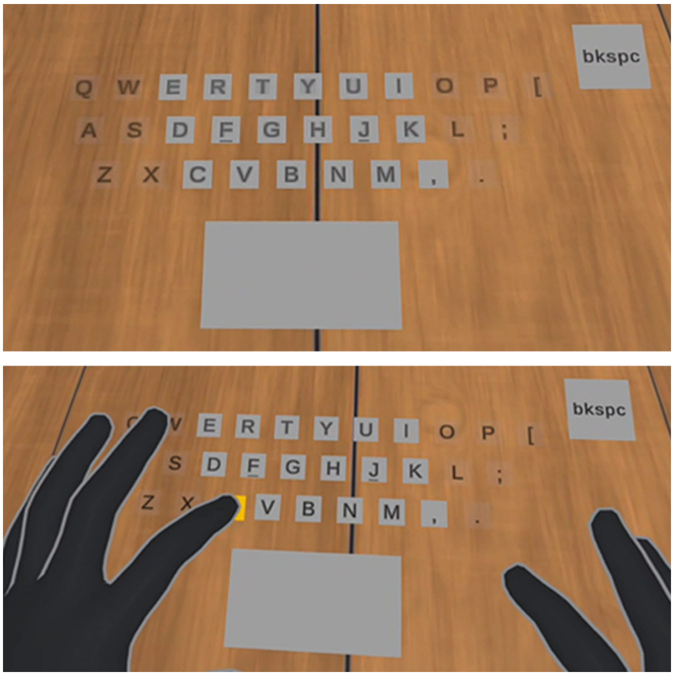

Fig. 2: Screen captures of the virtual keyboard.

half of that, at 4mm. In testing we found this was deep enough of a press to make the key-travel auditory augmentation (described in subsection 3.3) discernable, while minimizing depth to keep typing feeling as natural as possible.

3.3 Experimental Conditions

We examined four virtual mid-air keyboard conditions, as well as a single physical keyboard condition as a baseline, for reference. Each mid-air keyboard type was additive to the next, meaning that the amount of feedback information strictly increases as we consider the auditory augmentations of each keyboard. We assigned feedback information to each keyboard following Table 1. The implementation and audio samples used in this system can be found at https://github.com/akrasner19/MusiKeys.

***No Sound*:** As the name implies, this keyboard produced no auditory feedback of any kind, with the exception of the clicks when holding down backspace to indicate multiple deletions. This ability to hold backspace and hear repeated clicks for each subsequent deleted character was constant across all keyboards as a quality-of-life feature, so it could not be excluded from the *No Sound* condition. The *No Sound* condition was included in the study to give us a point of reference to differentiate effects of auditory feedback versus effects of visual interface design.

***Click Sound*:** This keyboard featured a medium-pitched click sound upon completing a key-press, and a higher, lighter click sound upon releasing the key past the release distance. In this keyboard and all of the other virtual keyboards, the spatial audio system, provided by the Oculus SDK, was used to emit each sound from the location of the respective key that made the sound. Though the effect was fairly minor in practice, we hoped it would provide an additional level of presence to sell the effect that each key was producing sound based on how you interacted with it. The *Click Sound* keyboard was included in the study as an approximation of the kind of auditory feedback present in state-of-the-art (SOTA) mid-air keyboards, such as the one found in Mixed Reality Toolkit SDK or the stock keyboards on the HoloLens 2 and Quest 2. It should finally be noted that the click sounds played by this keyboard were chosen to sound similar to those of SOTA keyboards, but due to certain requirements of the next two levels of auditory feedback, we employed a different set of click sounds on those two keyboards. This goes to say, the specific clicking sounds of the

Keyboard Feedback Information	Physical	No Sound	Click Sound	Key-Travel Sound	Musical Sound
Key-Press (Bottom-out)	h,a		a	a	a
Key-Release (Top-out)	h,a		a	a	a
Key-Touch	h			a	a
Key-Travel (Motion of Key)	h			a	a
Key-Hold	h			a	a
Finger Differentiation	h				a

h = Provides haptic feedback of the information
a = Provides auditory feedback of the information

Table 1: Comparison Table of Feedback Information Provided by Different Keyboards

Click Sound keyboard were exclusive to itself, but each subsequent keyboard detailed below still maintained the feedback paradigm of playing some type of click sound on press and release.

***Key-Travel Sound*:** This keyboard provides our first level of augmentation where we progress past kinds of auditory feedback currently found in SOTA mid-air keyboards. In addition to a set of tonal click sounds that play on press and release of a key, a constant, airy tone was added that plays while the key is being touched. The tone was produced by smoothly looping an audio sample created for this study, which consisted of an airy, droning, musical C4 note (using A440/pitch standard tuning). This sound is modulated by a low-pass audio filter and a volume envelope that increases the filter cutoff frequency and volume level of the tone as the key is pressed further down towards the press distance. When a key is pressed past the press distance, the tone volume cuts lower. The tone will continue to play at this low volume, indicating that the key is still held down, until the key is released past the release distance, returning the tone volume back to being a function of the distance the key is pressed down. The modulation curve for low-pass cutoff and volume is cubic in relation to distance pressed, as shown in Figure 1, item B. The purpose of adding this tone is to communicate that a finger is touching a key, the modulating filters provide the direction and status of the key-travel, and the volume-reduced tone after pressing the key informs the typist that they are still holding a key pressed down. These augmentations target the missing feedback information of a physical keyboard that are not provided by SOTA mid-air keyboards: awareness of touching, holding-down, and moving keys. Tones were used instead of white noise in *Key-Travel Sound* to avoid confounding comparisons with the next and final keyboard.

***Musical Sound*:** Once again, this included all of the parts of the prior keyboards: tonal click sounds on press and release, tones that play while touching a key, and the modulations of the volume and low-pass with key-travel and holding-down a key. However, *Musical Sound* takes all of the tonal elements above and maps the musical note of the tones played to the finger that is pressing each key. The mapping of each finger to each note is shown in Figure 1, item C. Notably, each finger is separated by a musical interval of a third to provide each finger with a distinct note while still sounding pleasant, and both thumbs share the same note since they both only ever press the same key, space-bar. This augmentation was designed to provide important missing feedback information of a physical keyboard that was not yet provided by *Key-Travel Sound*, that is, the knowledge of which fingers are currently interacting with the keys. In designing this experimental, novel augmentation, we do acknowledge that musical notes may not be intuitive or easy to differentiate for the average person, so along with collecting background data on musical experience, we hypothesized that even if participants cannot tell exactly which finger is touching a key at the moment, they may at least be able to localize which hand is touching a key.

***Physical Baseline*:** We created an analogous condition to the mid-air VR keyboard testbed in the real world so that we could examine the same task with a physical keyboard. The goal was to provide reference for grounding our measurements to the normal typing experience of our participants, not to be compared directly with the virtual keyboard conditions. A fair comparison was not possible due to the unreliability of hand-tracking used in our virtual keyboard implementation. For the physical baseline we used an Apple Magic Keyboard on an adjustable standing desk, and presented stimuli on a flat screen TV.

Finally, we positioned all virtual and baseline keyboards for use when standing, rather than sitting. This was because, in real world use cases of an XR virtual keyboard, if the user was seated, they would likely have a table or surface in front of them to use for passive haptics, a solution that has already been shown to work [8].

3.4 Apparatus

The study was conducted using both a VR testbed and a physical testbed. Our VR testbed consisted of an implementation of the virtual keyboard system, described in subsection 3.2, placed inside a virtual environment with a floor and the task screens detailed later in subsection 3.6. We built it in Unity as an application for the Meta Quest 2.

For testing the physical keyboard, the study room was setup to closely parallel to the VR testbed. A electronic variable-height desk held an Apple Magic Keyboard at a set distance from a 65-inch flat-screen TV. We placed the keyboard at the very edge of the standing desk, such that the participant could not rest their wrists while typing. In this way, we presented as similar a typing environment as possible to the VR testbed. The physical baseline task software was written in Python using the Kivy library, and was designed to have the same layout and features as the VR testbed task screens.

3.5 Participants

We recruited 24 participants from our university community for the study, 9 male and 15 female. Participants were screened to ensure safety as well as to guarantee all participants had perfect vision or corrected-to-perfect vision with contacts, since glasses would affect the Field-of-View (FOV) in the headset. It was important to maintain a comparable headset FOV between participants, since it impacted how much of the prompt and the keyboard could be kept in view at the same time. The participants were between the ages of 19 and 45, with the group having a median age of 22 and mean age of 24.46. Participants were given \$15 Amazon gift cards in compensation for their participation.

3.6 Task Design

We designed a precision text-entry typing task to evaluate the effect of the auditory augmentations. Participants read all words from a simulated document on the left and typed them into a simulated document on the right, with no interface indicators of correctness. Each task had a time limit of 1 minute and provided large samples to copy so that most participants would be unable to finish the task in the time provided. A submit button was provided in case of early completion. The VR testbed version of this is shown in Figure 3, and the physical keyboard task was identical but displayed on a flat-screen TV. Participants were instructed to balance their speed evenly with their accuracy, not prioritizing one over the other.

Because of our hand-tracking compensation method described in subsection 3.2, we could only present words that used a limited set of characters. A custom set of 1643 3-7 character words was manually created and randomly sampled to build all typing prompts. Commas were added randomly every 5-10 words, forming pseudo-sentences.

3.7 Measures

To measure performance of typing on virtual keyboards, we establish three metrics.

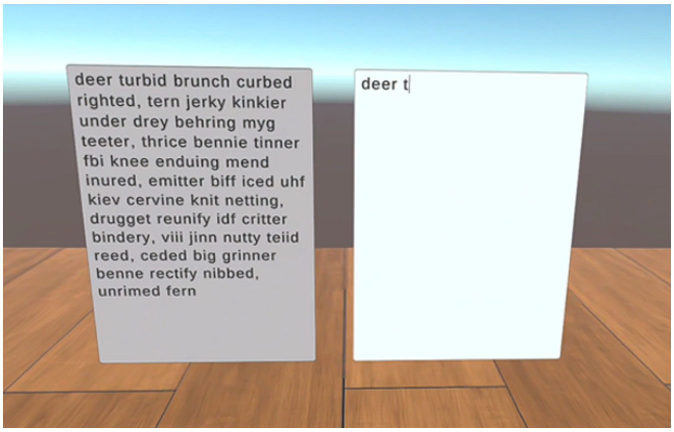

Fig. 3: Screen capture of VR task display.

First is Average Net Words-Per-Minute (Average Net WPM). This is calculated by taking the count of correct characters written, and dividing by the median word length of the data set, which for our study is 5 characters, then averaging measurements across the 5 tasks. To determine an approximate measure of how many correct characters were entered per minute, we utilized the work of Soukoreff and MacKenzie [23] to find the minimum string distance (MSD) between the closest correct result and the participant input. We treat the MSD as the number of incorrect characters entered per task, and subtract it from the total number of characters entered to find the net characters entered. We then take the resultant net characters entered, divide by the fraction of a minute spent typing, divide by 5 characters per word to obtain Net WPM, and finally average the Net WPM across the 5 tasks per trial to obtain Average Net WPM. This is summarized in the equation below, where n is the task number:

$$AvgNetWPM = \frac{1}{5}\sum_{n=1}^{5}\frac{CharactersEntered - MSD}{5}$$

Our second statistic for typing performance was Average Error Rate. This was found by using the method in [23] taking the MSD between the prompt and response, then dividing by whichever was higher, prompt length (A) or response length (B). We then average across the 5 tasks per trial to get Average Error Rate. We summarize this in the equation below where n is task number:

$$AvgErrorRate = \frac{1}{5}\sum_{n=1}^{5}\frac{MSD}{max(A,B)}$$

Note that our error rate does not account for corrected errors, it is an uncorrected error rate, nor does our WPM give an accurate idea of time lost from error correction. Thus, our final performance statistic is Average Keystrokes-Per-Character (Average KSPC). Using a method inspired by Soukoreff and MacKenzie [24], we calculate average keystrokes per character by summing the number of correct characters input by the participant with the number of backspace presses, then dividing the sum by the number of correct characters. This yields a statistic that tells us how many keystrokes it took to produce the number of correct characters reported in Net WPM. We finally averaged this number across all 5 tasks per trial to obtain Average KSPC. An equation to summarize this is below:

$$AvgKSPC = \frac{1}{5}\sum_{n=1}^{5}\frac{BackspaceCount + CorrectCharacterCount}{CorrectCharacterCount}$$

We used the unweighted NASA TLX [12] to measure workload for each keyboard condition.

One of the measures we used to measure perceived usability was the system usability scale (SUS) [5]. We also designed and utilized a study-specific MusiKeys user experience questionnaire (MUXQ) to answer specialized questions pertaining to our study. The MUXQ consisted of 20 questions ranging from how often a participant needed to look at the keys, to questions about how well they were able to understand and make use of the auditory feedback. Responses were given using a 7-point likert scale spanning from strongly disagree to strongly agree. In this paper, we only report results for the subset of the questions in the MUXQ that are relevant. The full contents of the MUXQ are published alongside this paper as supplemental materials.

The MUXQ was also a critical measure for evaluating how effectively each condition communicated its respective encoded feedback information. Without knowledge of this, it would be impossible to address if the information contained within the sounds had an effect or if the effect was caused by the sounds themselves, irrespective of the encoded information.

Finally, we collected keyboard preference rankings from the participants in an exit interview.

3.8 Procedure

The study was conducted in a series of 5 trials, one for each keyboard, starting with the physical keyboard, followed by the mid-air keyboards. Each trial followed the same structure.

At the start of each trial, participants were given a scripted explanation of how each keyboard worked. Next they were provided with unlimited free-practice time, typing prompts until they indicated their skill had reached a plateau for this keyboard. Following confirmation, participants completed 5, 1-minute long typing tasks, with short breaks between each. After completing the tasks, a NASA TLX survey was verbally administered. Participants then were directed back to the keyboard to spend time using the keyboard untimed while thinking about the experience of typing with it. When participants indicated they were ready, the trial concluded and they completed a series of post-trial surveys before beginning the next trial.

After completing all 5 trials, an exit interview was conducted.

4 RESULTS

The resulting data from the study was analyzed using ANOVAs on linear mixed effects models for a 2-way repeated measures design in both Minitab and R. Our primary fixed factor for investigation was Keyboard Type, but we discovered that, despite the incorporation of counterbalancing and practice periods into the study, a very strong order effect was present. To more accurately model the results, we incorporated Trial Number as a second fixed factor to isolate its effect on the model. Participants were treated as a random effect. Residuals were checked for normality using QQ plots and normality checks, and we used REML estimation with Kenward-Roger approximation for fixed effects, with significance at $p < 0.05$. For likert scale data, we applied an aligned rank transform (ART) prior to the analysis [27]. All post-hoc tests consisted of pairwise comparisons and contrast tests with Tukey HSD adjustment.

For all bar chart graphs, we display error bars of ± 1 SEM. Columns are labeled with post-hoc test groupings of significant difference. When physical baseline is shown, it is separated to signify that it was excluded from comparison testing. Physical baseline measurements are presented solely to provide a grounding reference value.

4.1 Typing Performance

We found that the inclusion of key-press and key-release feedback yields improved typing performance when compared to providing no auditory feedback. There was a significant main effect of Keyboard Type for both our Average Net WPM ($F_{3,60} = 6.55$, $p < .001$) and Average Error Rate ($F_{3,60} = 6.39$, $p < .001$) measures, with post-hoc tests showing that typists performed significantly worse using the *No Sound* keyboard compared to all three other mid-air, virtual keyboards. These three keyboard conditions all shared the aforementioned click sounds, though some offered additional auditory feedback. There was no significant change in typist performance across these keyboards that provided auditory feedback. The results are depicted in Figure 4.

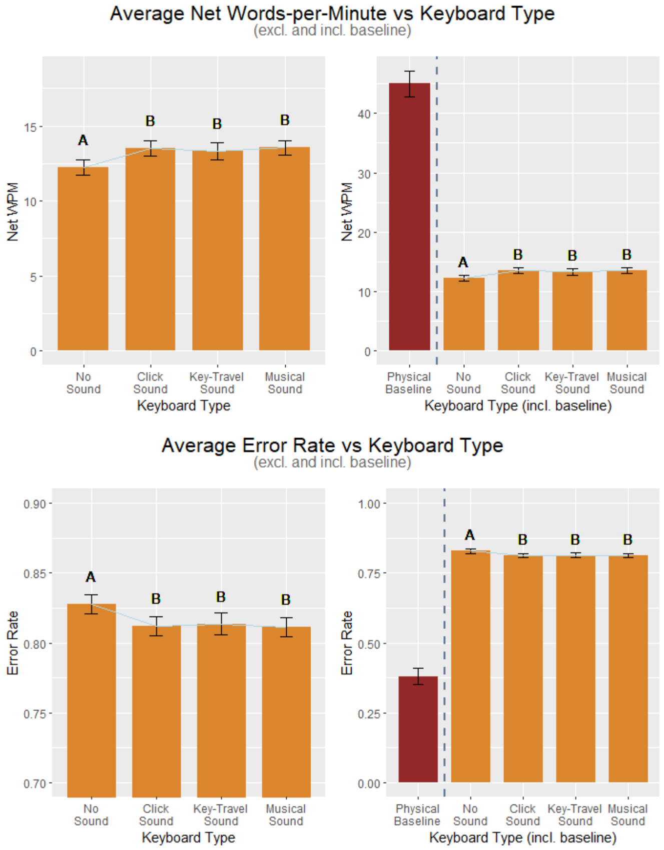

Fig. 4: Typing Performance for Average Net WPM and Error Rate

Average KSPC showed no significant difference across the four mid-air conditions, which indicates there is no significant difference in rate of backspace use between mid-air keyboards. Given that reductions in Average Net WPM and Error Rate were observed when typists used the *No Sound* keyboard, it is implied that typists correct their mistakes a constant amount for all conditions, leaving more errors when they struggled with accuracy, instead of correcting the errors. This is because uncorrected mistakes reduce both WPM and Error Rate measures.

4.2 Perceived Usability and User Preference

We analyzed perceived usability and found that none of our measures met our significance criteria: SUS Score ($F_{3,60} = 2.40$, $p < .077$) and MUXQ_20, a survey question on how natural typing felt ($F_{3,60} = 2.717$, $p < .053$). However, there are several trends visible in the data, pictured in Figure 5.

Our perceived usability results suggest that typists may find mid-air typing more usable when provided with auditory feedback of key-press and release. Additionally, the inclusion of varied pitches like in the *Musical Sound* condition may reduce usability to some degree (as suggested by Figure 5). However, without statistically significant differences, additional studies are needed to further examine this trend.

A similar trend exists in the user preference data, which we collected during the exit interview and where participants ranked the 4 mid-air virtual keyboards in order of preference (Figure 6). *No Sound* was not once chosen as the most preferred keyboard, and was ranked as least preferred keyboard most often. In line with the trend, *Musical Sound* was ranked most preferred the second fewest times, selected 7/24 times compared to *No Sound*'s 0/24 times, and *Musical Sound* ranked least preferred the second most times, 5/24 times compared to *No Sound*'s 13/24 times. Lastly, it should be noted that *Key-Travel Sound* was very rarely ranked last, by only 2 out of 24 participants.

4.3 Perceived Workload

Results from administering the NASA TLX find that typists perceive a reduction in overall workload when key-press and key-release feedback

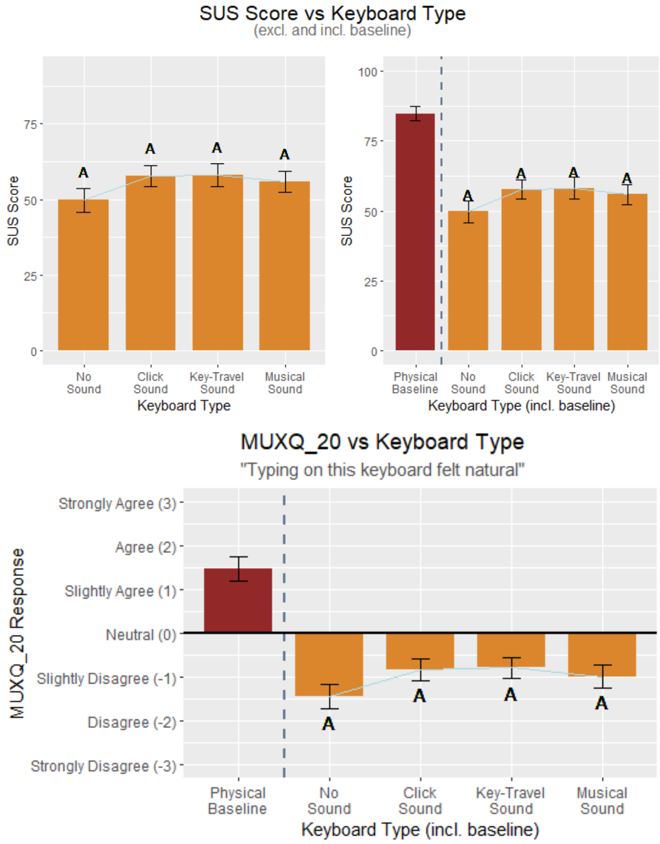

Fig. 5: Usability Results for SUS and MUXQ_20

are provided. There was a significant main effect of Keyboard Type on Average TLX score ($F_{3,60} = 4.84$, $p < .004$), TLX Performance ($F_{3,60} = 8.26$, $p < .001$), and TLX Frustration ($F_{3,60} = 3.97$, $p < .012$). All were scored significantly higher by participants (i.e. more workload) in *No Sound* condition compared to all others. TLX Effort appears to have a similar trend, with $p < .051$ for a significant difference between *Musical Sound* and *No Sound*.

Interestingly, typists also appear to perceive they perform better when key-touch, key-travel, and key-hold feedback are provided. In the TLX Performance scale, a lower score means participants believed they achieved closer to perfect performance. Both *Key-Travel Sound* and *Musical Sound* keyboards averaged lower in TLX Performance than the *Click Sound* keyboard, though significance was not proven by post-hoc testing. This is contrary to actual participant performance measures, where no such trend can be seen. We think that the added awareness of the typing system provided by the key-travel sounds, discussed in

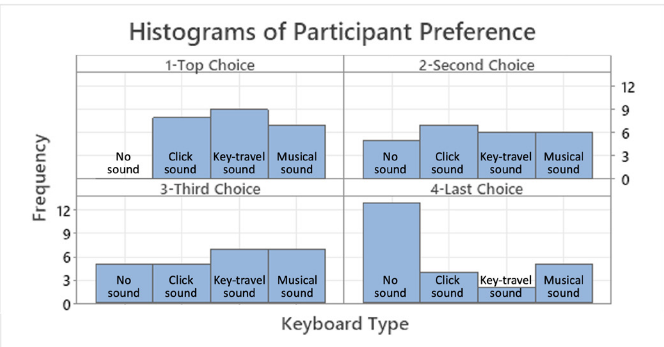

Fig. 6: Histograms of Participant Keyboard Preference by Ranking

MUXQ_6	I could tell when a key was being pressed-down (i.e. pressed deep enough to type a letter).
MUXQ_7	I could tell when a key was being touched (but not yet pressed-down).
MUXQ_8	I could tell which finger(s) were pressing-down key(s).
MUXQ_9	I could tell which finger(s) were touching key(s) (but not yet pressed-down).
MUXQ_10	I could tell which hand(s) had finger(s) that were pressing-down key(s).
MUXQ_11	I could tell which hand(s) had finger(s) that were touching key(s) (but not pressing down).
MUXQ_19	I felt like I had to look at the keys more than usual.

Table 2: Table of MUXQ questions 6-11 & 19, which probe participants' awareness of hands and fingers relative to keyboard.

the next section, may cause typists to intuit greater control over their responses, even though these perceived improvements do not affect real-world performance.

4.4 Awareness of Hands Relative to Keyboard

Finally we analyzed MUXQ questions that target participant awareness of hand and finger position with regards to state of the keyboard (MUXQ 6-11 & 19, see Table 2). In general, results show that the auditory feedback techniques of MusiKeys appear to effectively communicate the missing afforded information of the physical keyboard that is intended by each condition. While we found significant differences across MUXQ 6-11 & 19, for space constraints we present statistical results for MUXQ 6-9 & 19 only.

Participants responded that they are significantly more aware of pressing-down a key (MUXQ_6) when key-press and key-release feedback was provided , compared to no auditory feedback ($F_{3,60} = 35.31$, $p < .001$)(Figure 7, top left). Notably, *Key-Travel Sound* and *Musical Sound*keyboards were associated with significantly greater typist awareness of pressing-down a key than *No Sound* and *Musical Sound* was associated with greater awareness than *Click-Sound* and *No-Sound*. We assert that this increased awareness comes from key-hold sounds, since the tone continues to play while a key is held down. We see a similar finding in key-touch feedback (MUXQ_7), with typists reporting significantly greater awareness of touching the keys when using keyboards that provide key-touch feedback ($F_{3,60} = 35.31$, $p < .001$)(Figure 7, bottom left). Interestingly though, though the *Click Sound* keyboard does not provide key-touch feedback information, it still yields a significantly higher awareness of touching keys than with *No Sound*. We believe this is because having the sound play on key-press and key-release grounds the typists finger position to the auditory "landmark" at the press-distance, giving them some idea of finger location, whereas *No Sound* leaves them in the dark.

Responses also showed that providing different musical notes for the sounds made by each finger, i.e. finger differentiation feedback, does successfully communicate information of which fingers were pressing down keys. Participants were significantly more aware of which fingers were pressing-down keys (MUXQ_8) ($F_{3,60} = 9.591$, $p < .001$) (Figure 7, top right) and which fingers were touching keys (MUXQ_9) ($F_{3,60} = 35.42$, $p < .001$)(Figure 7, bottom right) when using the *Musical Sound* keyboard than with keyboards that did not provide finger differentiation.

Additionally, analysis revealed that participants reported the need to look at the keys significantly more for the *No Sound* condition than for *Click Sound*, *Key-Travel Sound*, and *Musical Sound* conditions (MUXQ_19) ($F_{3,60} = 18.51$, $p < .001$).

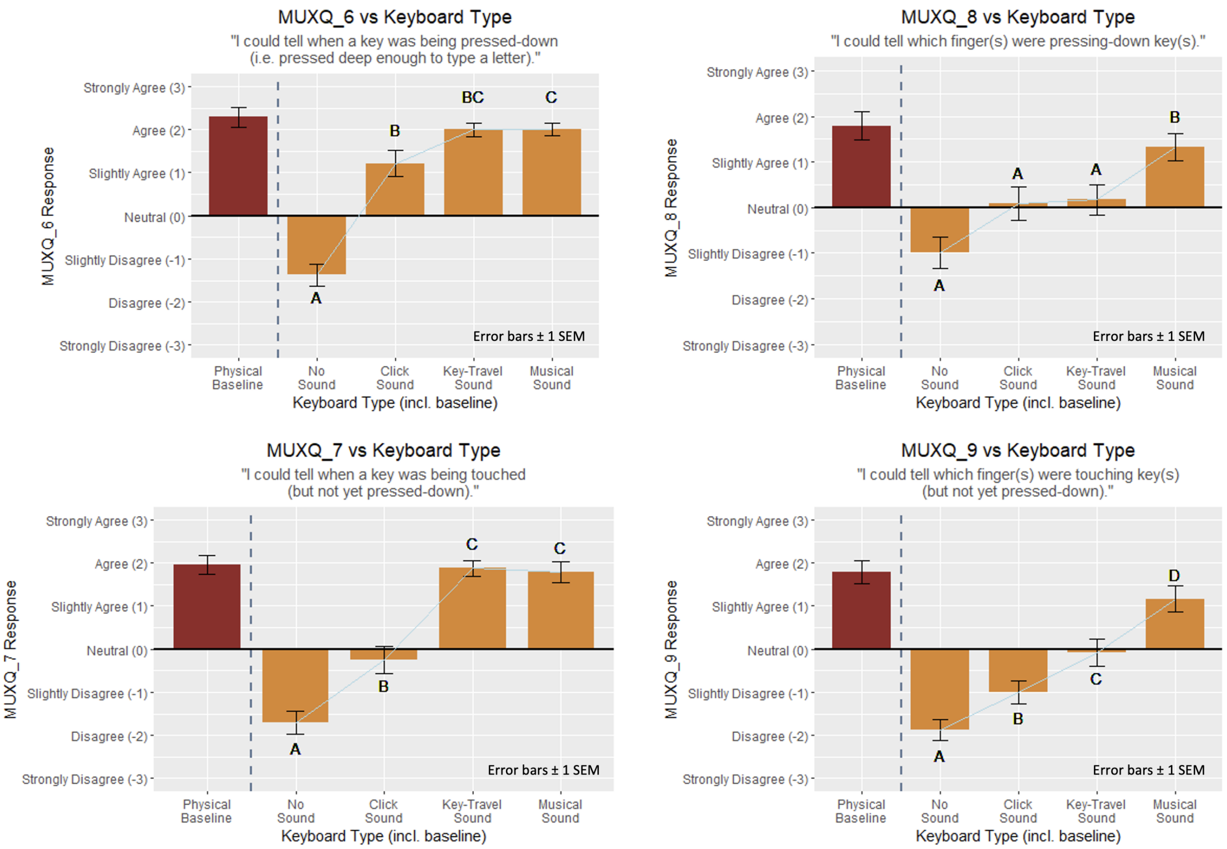

Fig. 7: Participants' Awareness of Hands Relative to Keyboard Across Several Measures

5 Discussion

We set out in this study to assess the effects of augmenting an XR mid-air QWERTY keyboard with audio to communicate the haptic feedback information typically afforded by a physical keyboard. In this section, we address our research questions and the implications of our results.

5.1 Capability of Sensory Substitution in Mid-air Typing

Our first research question centered on *whether or not it is possible* to provide haptic information typically afforded by a physical keyboard through audio modal, and if so, could that audio improve typing on an XR mid-air keyboard. Looking at the results in subsection 4.4 and the results in Figure 7, it seems clear that each auditory feedback augmentation was able to effectively communicate the haptic-feedback information encoded by it. For example, in results for MUXQ_6 (Figure 7, top left), we see that the keyboards that provided information of key-press yielded significantly greater awareness of key-press. Similarly, in MUXQ_7 the keyboards that present key-touch feedback were rated significantly higher for awareness of key-touch. This pattern continues through MUXQ_8 and MUXQ_9, with finger differentiation being significantly higher for *Musical Sound* keyboard, the only keyboard presenting finger differentiation feedback information. Key-touch is also reflected to be impacting the results in MUXQ_9.

However, though the feedback information appears to be understood by the participants, we do not see increases in our measures of typing interface effectiveness (ie. typing performance, workload, and usability) as further feedback information is provided. This raises a critical question: if mid-air typists understand the auditory feedback signals they are receiving, thereby giving them similar awareness of how they are interacting with the virtual keyboard as they have with a physical keyboard, why does effectiveness of the interface not grow closer to a physical keyboard for each physical feedback signal we emulate?

One possible explanation is that processing the substituted information to inform typing actions is possible, but is much slower, thus not becoming evident in time-limited tasks. We hypothesize that with extensive training time, beyond the scope of this study, it may be possible for participants to learn to reflexively use the feedback from the *Musical Sound* keyboard to improve their typing, similarly to how one learns to type for the first time and builds that muscle memory.

Another explanation of our findings could be that, at least in the case of typing and similar tasks, awareness of feedback signals is not utilizable in the same way as the physical feedback signal itself. Specifically, it may be that providing the information contained in haptic-feedback of a physical keyboard to a typist may not grant the typist similar ability to use a mid-air, non-physical keyboard. In this way of thinking, it may be that typists do not consciously process haptic information while typing on a physical keyboard, and instead may rely on muscle memory [17].

In conclusion, while our study did not show that including haptic-feedback information results in a correlated increase in typing effectiveness across the mid-air keyboards we examined, our results suggest that MusiKeys technique, i.e. communication of haptic feedback information through audio during mid-air typing , was effective at communicating this feedback information, and has promise.

5.2 Effect of Amount of Auditory Feedback

Our second research question examined how *the amount* of haptic information communicated through audio may affect typing on a mid-air QWERTY keyboard. The results for our measures of typing effectiveness, i.e. typing performance (as measured by Average Net WPM and error rate), usability, and workload, unanimously suggest an overarching finding that providing auditory feedback, at least to the level of the *Click Sound* condition, improves typing on mid-air keyboards. This replicates the findings of other XR and general user-interface studies [13, 30, 31]. These works found auditory feedback more effective than absence of feedback in improving task performance, including for typing tasks. However, the forms of auditory feedback studied in these works used audio that increased visibility of system status rather than audio that attempted to present equivalent information as the sense of touch.

We also find that adding additional information contained in *Key-Travel Sound* (i.e. key-touch, key-travel, key-hold) and *Musical Sound* (i.e. finger differentiation) does not significantly degrade effectiveness of a mid-air typing interface. Note that the absence of statistically significant differences between the three audio-producing virtual keyboards does not mean an absence of differences, simply that it can not be proven using the given data.

We can also see that *No Sound* is consistently significantly lower rated for hand and finger awareness relative to virtual keyboard position as compared to other virtual keyboards across MUXQ measures 6-9 & 19. It is notable that for MUXQ 7 & 9, *No Sound* is rated significantly lower than *Click Sound* even though *Click Sound* does not provide the additional information needed to convey key-touch and finger differentiation (MUXQ 7 & 9 respectively).Additionally, participants self-reported the need to look at the virtual keys significantly more often for the *No Sound* condition as compared to all other conditions where auditory feedback was provided (MUXQ_19).

We propose these findings may be caused by a feeling of disorientation in the absence of any auditory feedback. It seems likely that purely using proprioception to assess where fingers are tracked in space may not be effective. Without auditory feedback, typists are more likely to lose their orientation of hands and fingers to the keys (as seen in results for MUXQ_19). Conceptually, one would posit that with each "layer" of feedback information added, the typist is provided with more points of reference to derive awareness of their hands interacting with the keyboard. However, more research is needed since our particular study design did not unequivocally show this.

6 Limitations and Future Work

The native hand-tracking of the Meta Quest 2 was unreliable, leading us to implement a compensatory technique that limited usable fingers and keys. Not only does this limit the applicability of our work to systems that support 10-finger typing, but it may have been a contributing factor to how well participants understood the auditory feedback. Fewer fingers means fewer sounds playing simultaneously. Additionally, having to build our own word set due to limited available letters meant that we could not evaluate the system with any of the established standard prompts for text-entry tasks.

Another limitation lies in how strong our learning effect was. Despite in-depth explanations to bring participants up to speed and unlimited practice time, our learning effect was so strong that we needed to change our statistical design to address it. If the study was run in a multi-day time-trame, similar to the work of Gupta et al. [11], we may have had clearer results.

Finally, we need to address the possibility of priming caused by the aforementioned in-depth explanations of how each interface worked. We actively limited the confounding effect of our explanations by not only requesting before each survey that they give honest responses to the questions, but also giving them a free period before the surveys to focus on the experience of typing and form their own opinions. It is still possible, though, that telling participants how the auditory augmentations of each keyboard worked influenced their responses to a degree.

Looking towards future work, further implementations of the key-touch, key-travel, and key-hold feedback augmentations may lead to improvements not only for mid-air text-entry, but also for general interface design. For example, it may be valuable to apply them to other instances where a user may need to interact with an object outside of their field-of-view.

7 Conclusion

In this work, we set out to investigate if typing on XR mid-air QWERTY keyboards could be improved by using sound to provide typists with the same information that is normally afforded when typing on a physical QWERTY keyboard. We built a virtual reality testbed with a real-world counterpart and designed a study to explore the effects of the auditory augmentations on performance, usability, and workload, while additionally evaluating the ability of our technique to communicate the tactile-feedback information. We found that providing auditory feedback of key-press and key-release improves typing compared to not providing any auditory feedback, which replicates literature. We also show that it is possible to substitute the information contained in haptic-feedback using audio, to the extent that users are able to detect and understand these substitutions. However, our specific type of substitutions (along with other possible design considerations such as keyboard dimensions) did not significantly improve typing performance and usability across our four mid-air keyboards. While the supplied information was understood, we posit that making use of that information would potentially take time to develop the muscle memory reflexes that typists have built over years when using physical keyboards to type. Nonetheless, we recommend others consider incorporating auditory feedback of key-touch, key-travel, and key-hold into their mid-air keyboards, since it was associated with high levels of user preference. We hope that our work in this area helps to inform future XR auditory design research for mid-air interfaces, with the goal of one day living in a world where the only keyboard you need can be used anywhere, takes up no space, and is always with you.

References

[1] A. U. Batmaz and W. Stuerzlinger. Effects of different auditory feedback frequencies in virtual reality 3D pointing tasks. *Proceedings - 2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops, VRW 2021*, pp. 189–194, 3 2021. doi: 10.1109/VRW52623.2021.00042 2

[2] A. U. Batmaz, K. Yu, H. N. Liang, and W. Stuerzlinger. Improving Effective Throughput Performance using Auditory Feedback in Virtual Reality. *Proceedings - SUI 2022: ACM Conference on Spatial User Interaction*, 12 2022. doi: 10.1145/3565970.3567702 2

[3] C. Bermejo, L. H. Lee, P. Chojecki, D. Przewozny, and P. Hui. Exploring Button Designs for Mid-air Interaction in Virtual Reality: A Hexa-metric Evaluation of Key Representations and Multi-modal Cues. *Proceedings of the ACM on Human-Computer Interaction*, 5(EICS):26, 5 2021. doi: 10.1145/3457141 3

[4] D. A. Bowman, C. J. Rhoton, and M. S. Pinho. Text Input Techniques for Immersive Virtual Environments: An Empirical Comparison. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 46(26):2154–2158, 9 2002. doi: 10.1177/154193120204602611 2

[5] J. Brooke. SUS: A quick and dirty usability scale. *Usability Eval. Ind.*, 189, 1995. 5

[6] S. H.-W. Chuah. Why and Who Will Adopt Extended Reality Technology? Literature Review, Synthesis, and Future Research Agenda. *SSRN Electronic Journal*, 12 2018. doi: 10.2139/SSRN.3300469 1

[7] T. J. Dube and A. S. Arif. Text Entry in Virtual Reality: A Comprehensive Review of the Literature. In M. Kurosu, ed., *Human-Computer Interaction. Recognition and Interaction Technologies*, pp. 419–437. Springer International Publishing, Cham, 2019. 2

[8] J. Dudley, H. Benko, D. Wigdor, and P. O. Kristensson. Performance envelopes of virtual keyboard text input strategies in virtual reality. *Proceedings - 2019 IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2019*, pp. 289–300, 10 2019. doi: 10.1109/ISMAR.2019.00027 2, 5

[9] J. Fashimpaur, K. Kin, and M. Longest. PinchType: Text entry for virtual and augmented reality using comfortable thumb to fingertip pinches. *Conference on Human Factors in Computing Systems - Proceedings*, 4 2020. doi: 10.1145/3334480.3382888 2

[10] H. Gil, Y. Shin, H. Son, I. Hwang, I. Oakley, and J. R. Kim. Characterizing In-Air Eyes-Free Typing Movements in VR. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology, VRST*. Association for Computing Machinery, 11 2020. doi: 10.1145/3385956.3418963 1, 2

[11] A. Gupta, M. Samad, K. Kin, P. O. Kristensson, and H. Benko. Investigating Remote Tactile Feedback for Mid-Air Text-Entry in Virtual Reality. In *Proceedings - 2020 IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2020*, pp. 350–360. Institute of Electrical and Electronics Engineers Inc., 11 2020. doi: 10.1109/ISMAR50242.2020.00062 2, 3, 9

[12] S. G. Hart and L. E. Staveland. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in Psychology*, 52(C):139–183, 1 1988. doi: 10.1016/S0166-4115(08)62386-9 5

[13] J. R. Kim and H. Z. Tan. A study of touch typing performance with keyclick feedback. *IEEE Haptics Symposium, HAPTICS*, pp. 227–233, 2014. doi: 10.1109/HAPTICS.2014.6775459 2, 8

[14] L. H. Lee, K. Yung Lam, Y. P. Yau, T. Braud, and P. Hui. HIBEY: Hide the Keyboard in Augmented Reality. In *2019 IEEE International Conference on Pervasive Computing and Communications (PerCom*, pp. 1–10, 3 2019. doi: 10.1109/PERCOM.2019.8767420 3

[15] J. M. Loomis. Sensory replacement and sensory substitution: Overview and prospects for the future. *Converging technologies for improving human performance*, 213, 2002. 3

[16] M. Olivetti Belardinelli, S. Federici, F. Delogu, and M. Palmiero. Sonification of spatial information: Audio-tactile exploration strategies by normal and blind subjects. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5615 LNCS(PART 2):557–563, 2009. doi: 10.1007/978-3-642-02710-9
62 3

[17] V. Pujari. Original Research Muscle Memory and the Brain: How Physical Skills are Stored and Retrieved. *Journal of Advanced Medical and Dental Sciences Research |Vol*, 2019. 8

[18] A. Sand, I. Rakkolainen, P. Isokoski, J. Kangas, R. Raisamo, and K. Palovuori. Head-mounted display with mid-air tactile feedback. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology, VRST*, vol. 13-15-Nove, pp. 51–58. ACM, New York, NY, USA, 2015. doi: 10.1145/2821592.2821593 2

[19] J. Shen, J. Hu, J. J. Dudley, and P. O. Kristensson. Personalization of a Mid-Air Gesture Keyboard using Multi-Objective Bayesian Optimization. *Proceedings - 2022 IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2022*, pp. 702–710, 2022. doi: 10.1109/ISMAR55827.2022.00088 3

[20] P. B. Shull and D. D. Damian. Haptic wearables as sensory replacement, sensory augmentation and trainer - A review. *Journal of NeuroEngineering and Rehabilitation*, 12(1):1–13, 7 2015. doi: 10.1186/S12984-015-0055-Z/FIGURES/7 3

[21] A. Simeone, B. Weyers, S. Bialkova, and R. W. Lindeman. *Everyday Virtual and Augmented Reality*, chap. 1, pp. 1–20. Springer International Publishing, 2023. 1

[22] Y. Singhal, R. H. Noeske, A. Bhardwaj, and J. R. Kim. Improving Finger Stroke Recognition Rate for Eyes-Free Mid-Air Typing in VR. *Conference on Human Factors in Computing Systems - Proceedings*, 4 2022. doi: 10.1145/3491102.3502100 1, 2

[23] R. W. Soukoreff and I. S. MacKenzie. Measuring Errors in Text Entry Tasks: An Application of the Levenshtein String Distance Statistic. In *CHI '01 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '01, p. 319–320. Association for Computing Machinery, New York, NY, USA, 2001. doi: 10.1145/634067.634256 5

[24] R. W. Soukoreff and I. S. MacKenzie. Metrics for Text Entry Research: An Evaluation of MSD and KSPC, and a New Unified Error Metric. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, p. 113–120. Association for Computing Machinery, New York, NY, USA, 2003. doi: 10.1145/642611.642632 5

[25] R. Thome, C.-T. Leader, S. Offutt, M. E. Tyler, and M. S. PE. Tactile Auditory Sensory Substitution. *BME*, 200:300, 2006. 3

[26] C. Y. Wang, W. C. Chu, P. T. Chiu, M. C. Hsiu, Y. H. Chiang, and M. Y. Chen. Palm type: Using palms as keyboards for smart glasses. *MobileHCI 2015 - Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*, pp. 153–160, 8 2015. doi: 10.1145/2785830.2785886 2

[27] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins. The Aligned Rank Transform for nonparametric factorial analyses using only ANOVA procedures. *Conference on Human Factors in Computing Systems - Proceedings*, pp. 143–146, 2011. doi: 10.1145/1978942.1978963 6

[28] W. Xu, H. Liang, A. He, and Z. Wang. Pointing and Selection Methods for Text Entry in Augmented Reality Head Mounted Displays. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 279–288, 10 2019. doi: 10.1109/ISMAR.2019.00026 2

[29] X. Yi, C. Yu, M. Zhang, S. Gao, K. Sun, and Y. Shi. ATK: Enabling ten-finger freehand typing in air based on3D hand tracking data. In *UIST 2015 - Proceedings of the 28th Annual ACM Symposium on User Interface Software and Technology*, pp. 539–548. Association for Computing Machinery, Inc, 11 2015. doi: 10.1145/2807442.2807504 2

[30] C. Yildirim. Point and Select: Effects of Multimodal Feedback on Text Entry Performance in Virtual Reality. *International Journal of Human-Computer Interaction*, 2022. doi: 10.1080/10447318.2022.2107330 2, 8

[31] Y. Zhang, R. Sotudeh, and T. Fernando. The Use of Visual and Auditory Feedback for Assembly Task Performance in a Virtual Environment. *Proceedings of the 21st Spring Conference on Computer Graphics*, 2005. doi: 10.1145/1090122 2, 8