Conveying Language through Haptics: A Multi-sensory Approach

Nathan Dunkelberger[1], Jenny Sullivan[1], Joshua Bradley[1], Nickolas P Walling[1], Indu Manickam[1], Gautam Dasarathy[1], Ali Israr[2], Frances W. Y. Lau[2], Keith Klumb[2], Brian Knott[2], Freddy Abnousi[2], Richard Baraniuk[1], Marcia K O'Malley[1]

[1]Rice University, Houston, TX, USA
{nbd2, jls3, jmb25, npw1, indu.manickam, gautamd, richb, omalleym}@rice.edu

[2]Facebook Inc., 1 Hacker Way, Menlo Park, CA, USA
{aliisrar, flau, kklumb, brianknott, abnousi}@fb.com

ABSTRACT

In our daily lives, we rely heavily on our visual and auditory channels to receive information from others. In the case of impairment, or when large amounts of information are already transmitted visually or aurally, alternative methods of communication are needed. A haptic language offers the potential to provide information to a user when visual and auditory channels are unavailable. Previously created haptic languages include deconstructing acoustic signals into features and displaying them through a haptic device, and haptic adaptations of Braille or Morse code; however, these approaches are unintuitive, slow at presenting language, or require a large surface area. We propose using a multi-sensory haptic device called MISSIVE, which can be worn on the upper arm and is capable of producing brief cues, sufficient in quantity to encode the full English phoneme set. We evaluated our approach by teaching subjects a subset of 23 phonemes, and demonstrated an 86% accuracy in a 50 word identification task after 100 minutes of training.

ACM Classification Keywords

H.5.2. User Interfaces: Haptic I/O

Author Keywords

Haptics; Multi-sensory; Wearable; Speech.

INTRODUCTION

There are many everyday situations in which a haptic language would be beneficial. Whether the typical audio and visual channels are unavailable due to impairment, or they are occupied by other stimuli, the sense of touch is often left underutilized. Providing a haptic language is an elegant solution to this problem, enabling communication through a largely available organ — the human skin.

ISWC '18, October 8–12, 2018, Singapore, Singapore

ACM ISBN 978-1-4503-5967-2/18/10. . . $15.00

DOI: https://doi.org/10.1145/3267242.3267244

One commonly-used approach to convey language through the haptic channel is to de-construct acoustic signals into components that can be interpreted by humans through tactile sensations. Spectral displays are one example which converts acoustic signals into corresponding frequency bands, and amplitudes, and convey these features through a haptic display. Two examples are the tactile vocoder presented by Brooks et al. [2] in which a 16 channel solenoid array placed on the forearm was actuated with corresponding amplitudes from different frequency bands, and a vibrotactile vest presented by Novich and Eagleman [15]. These systems that encode acoustic signals are not very intuitive to learn and take a substantial amount of practice to be able to understand the underlying signal [17, 22].

A more intuitive approach that many researchers have utilized for haptic communication is using letter-based communication. There are significant benefits to using letters as the basis for the haptic code because spelling is independent of pronunciation. Several haptic systems have been designed to take advantage of letters as the basic building block, including virtual braille through touch screen and vibration on a phone requiring active touch [8], or through tangential skin strain [11]. Another example is through the use of tactual Morse code using a device to displace the fingertip to present either a dot or a dash [19]. However, since letters are the smallest building block of language, these methods take the longest amount of time to communicate, resulting in slow communication rates.

Words, on the other hand, are a relatively large building block of language, and therefore would be very fast to communicate. However, while presenting words may be fast, and may provide enough information for limited scenarios where small numbers of words are needed, there are more than 100,000 words in the English language, so creating a universal haptic language using words would be impractical.

An option that lies between letters and words is phonemes, which correspond to each distinct sound made in a language. Most English words either have the same or fewer phonemes than letters, which allows for a faster presentation rate. Because there are approximately 40 phonemes in the English language, the cue set size in increased slightly from letters. Encoding phonemes through haptics has been explored and validated using the haptic device, the TactuatorII [6], as well as through vibrotactile sleeves [9, 23]. With the TactuatorII, pairwise discrimination tests between similar phonemes were performed with high sensitivity indices. However, the TactuatorII requires three fingers on the device to function, which would prevent functionality during everyday activities. Similarly, the vibrotactile sleeve used by Jiao et al. [9] has shown promising results, but requires a large area of the lower arm to be used. Using phonemes rather than letters agrees with the hypothesis that in order to reach high information transfer rates, one should increase the information content of each cue, rather than continuing to speed up the presentation rates [20]. Therefore, we have chosen to to use phonemes for our implementation of a haptic language.

Regardless of the type of building-block used to convey language, we need to provide a distinct cue set of at least 26 cues. The main approach in current research to provide these large cue sets is through combinations of vibrotactors [5, 9, 13, 15, 16, 18, 22], due to the wide variety of possible variations of vibrotactile cues including amplitude, frequency, and waveforms. However, interference is a common problem in vibrotactile arrays when the vibrotactors are spaced too close together. It was found that on the back, vibrotactors must be spaced at least 2 cm [15] apart in order for users to be able to individually locate each of the vibrotactors. This often leads to large arrays worn across a large surface area of the body, such as the the chest and torso [15, 16]. While these devices can provide a large number of distinct cues, they typically cover either a large portion or an important part of the body, which can interfere with everyday activities.

A large cue set can be obtained without using large vibrotactile arrays through the use of temporal cues, or by using a portion of the body with higher sensory resolution. However, with the increased complexity of temporal cues [10], the duration of haptic cues extends to undesirable lengths, and can negatively impact presentation speed. Additionally, haptic language devices that rely on the high sensitivity of skin found on the fingertips [7, 8, 11, 19] or on the hands [13] occupy real estate that is necessary for many activities. While better perception can be obtained in these areas [14], these devices are not usable in an everyday environment due to their contact location.

In this paper, we introduce a haptic language through the use of the Multi-sensory Interface of Stretch, Squeeze, and Integrated Vibration Elements (MISSIVE) [4], a compact, wearable haptic device capable of producing multi-sensory cues. The MISSIVE is worn on the upper arm and leverages the ability to perceive different types of tactile sensations concurrently. This increases the size of the cue set without expanding the contact surface area or lengthening the cue presentation time.

We have validated the MISSIVE through a user study and have shown that subjects are able to accurately understand a set of phonemes and words in a limited training time. The remainder of the paper is organized as follows: in the Methods section, we present the hardware design and implementation of the MISSIVE; in the Experimental Evaluation section, we outline the phoneme mapping and the protocol used to evaluate the MISSIVE; and finally, in the Results and Conclusions sections, we analyze the results and outline key points and possible future work.

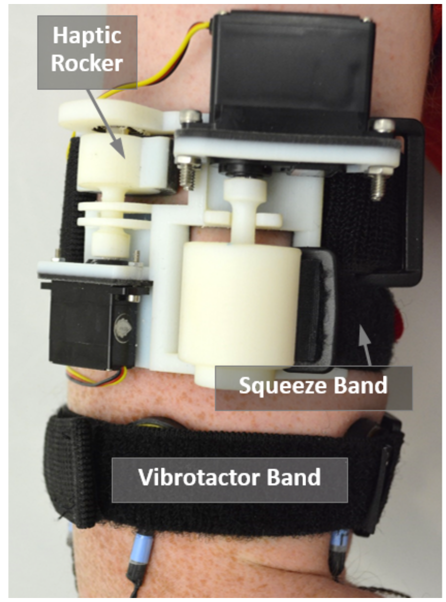

Figure 1. MISSIVE device, worn on the upper arm. The Proximal Band includes the Haptic Rocker and the Squeeze Band, which render lateral skin stretch and radial squeeze sensations, respectively. The Vibrotactor Band utilizes four vibrotactors to render a low, high, or double pulse on either the top, right, bottom, or left side of the arm.

METHODS

Hardware Design

MISSIVE is a compact device capable of delivering a variety of tactile cues to the upper arm of the user. It integrates three types of haptic actuators—a vibrotactor band, radial squeeze band, and haptic rocker—to produce concurrent sensations of vibration, radial squeeze, and lateral skin stretch, as shown in Figure. 1.

Vibrotactor Band

The Vibrotactor Band consists of four C2 Tactors (Engineering Acoustics Inc., USA), which are 1.2 inches in diameter and actuated by a voice coil mechanism. They are positioned on the top, right, bottom, and left sides of the arm and secured with an elastic strap.

Radial Squeeze Band

The Radial Squeeze Band consists of a strap that is connected to a servomotor on one end and wraps around the user's arm. When the servomotor is actuated, it tightens the band and squeezes the user's arm. It is mounted on the Proximal Band of the MISSIVE, approximately half an inch above the Vibrotactor Band. The design of the Squeeze Band is based on a similar device developed in the MAHI Lab, the Rice Squeeze Band [21]. The servomotor (HS-485HB, Hitec RCD USA, Inc.) has a maximum torque output of 588 mNm.

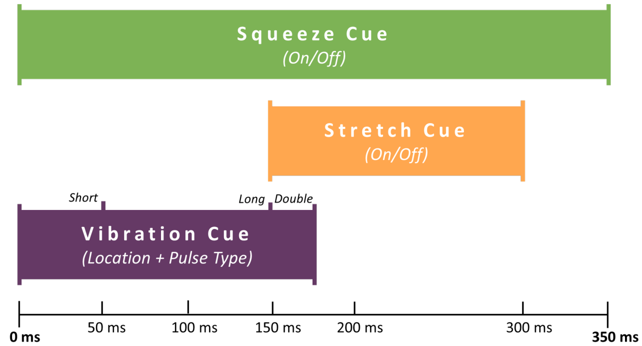

Figure 2. Timing diagram for a single MISSIVE cue. Both the squeeze and the stretch cue components can be either off or on, and the vibration always occurs. The stretch cue begins 150 ms after the squeeze and vibration cue components begin

Haptic Rocker

Lateral skin stretch is rendered by the Rice Haptic Rocker, which was designed by Clark and described in Battaglia et al. [1]. The device comprises a servomotor connected to a rubber-coated, semi-circular end-effector that is pressed against the user's arm. When the servomotor is actuated, it induces a mild skin-shear sensation by rotating the end-effector and stretching the skin. The Haptic Rocker is mounted on the proximal end of the Proximal Band, near the top side of the user's arm. The servomotor (HS-5070MH, Hitec RCD USA, Inc.) has a maximum torque of 375 mNm.

Haptic Cue Set

A set of 48 multi-sensory haptic cues was developed as part of the MISSIVE communication system. Each cue contains three components: a vibration component, a lateral skin stretch component, and a radial squeeze component. Through pilot testing, actuation patterns for each cue component were created that were both easily perceptible and of similar intensity.

The vibration cue components were rendered by activating a single tactor (top, right, bottom, or left) for a short pulse (50 ms), long pulse (150 ms), or double pulse (50 ms on, 75 ms off, 50 ms on), resulting in twelve unique cues. The vibrotactors were driven at a constant frequency of 265 Hz to maximize the amplitude of the vibration.

The radial squeeze and lateral skin stretch cue components were rendered as binary, on/off cues. The radial squeeze cue component was rendered by tightening the Squeeze Band to maximum torque and then releasing, for a total of 350 ms. Although it was possible to actuate the Squeeze Band faster, pilot testing showed a significant drop-off in perception accuracy. The lateral skin stretch cue component was rendered by rotating the haptic rocker 30° and then returning it back to its center position, resulting in a total cue duration of 150 ms. The initiation of the skin stretch component was delayed by 150 ms, which did not change the overall cue time, but made cues more discriminable in pilot testing. The relative timing of all three cue components is shown in Figure 2.

Mapping Haptic Cues to Phonemes

The mapping between cues and phonemes was designed with the objective of minimizing sentence-level comprehension errors by assigning haptic cue pairs more likely to be confused to phoneme pairs that are less likely to cause a misconception at the sentence level. The mapping was generated by optimizing the cost function defined in (2).

The cue-phoneme mapping problem may be cast as follows. Let N represent the set of phonemes and M represent the set of cues, with $|M| \geq |N|$; notice that in our situation $|N| = 40$ and $|M| = 48$. There are two input functions required to compute the expected cost of a given mapping. $D : M \times M \to R^+$ represents the probability of confusing one cue for another, and was determined empirically using previous studies [4]. $F : N \times N \to R^+$ represents the cost of confusing one phoneme for another.

F was estimated based on the principle that translation errors where the user determines that the received message is nonsense are preferable to errors where the user is unaware that an error has occurred and therefore misconceives the message. The cost of the phoneme pair (i, j) is therefore a function of the number of instances where mistaking phoneme i for phoneme j within a particular word results in a new valid word that is the same part of speech as the original word, with higher weight assigned to words that are more frequently used.[1] The part of speech was taken into consideration because confusion between words with different parts of speech most likely results in nonsense at the sentence level.

We formally compute F as follows. Let W represent the set of valid English words, and let W_i denote the subset of words within W that contain the phoneme i. Let $w_i(j)$ denote the new word that is created when all instances of the phomeme i within the word $w \in W_i$ are substituted by the phoneme j. The indicator function $I_D(w)$ evalutes to 1 if $w \in D$ and 0 otherwise. The binary function $P(w_1, w_2)$ evalutes to 1 if the part of speech of words w_1 and w_2 are equivalent, and 0 otherwise. Let $Q(w)$ represent the frequency that the word w appears in speech and literature. The cost between phonemes i and j is then defined as:

$$F(i, j) = \log(1 + \sum_{w \in W_i} Q(w) I(w_i(j) \in D) P(w, w_i(j))) \quad (1)$$

Using F as defined in (1) and D, we define the total expected cost of a mapping as:

$$C(\varphi) = \sum_{(i,j) \in N} F(i, j) D(\varphi(i), \varphi(j)) \quad (2)$$

In (2), the function $\varphi : N \to M$ denotes a particular mapping of phonemes to cues, where $\varphi(i)$ represents the cue mapped to phoneme i. Our objective therefore was to find the mapping $\varphi^* = \text{argmin}_\varphi C(\varphi)$. This problem is a variant of the Quadratic Assignment Problem (QAP), a long-standing combinatorial

[1] The following databases were used: COCA (https://corpus.byu.edu/coca/) for word frequency, WordNet (https://wordnet.princeton.edu/) for part of speech, and CMU-dict (http://www.speech.cs.cmu.edu/cgi-bin/cmudict) for word pronunciation.

optimization problem. The QAP has been shown to be NP hard, and therefore in practice approximate solutions are found using local search algorithms [12]. The approximated optimal mapping for this problem was determined using the genetic algorithm[2] [3].

After initial testing [4], the cost function was adjusted to make it easy for subjects to distinguish between consonants and vowels by constraining the assignment of consonant phonemes to cues with the squeeze enabled and vowel phonemes to cues with the squeeze disabled. This feature aids with memory recall, as it reduces the number of search items the subject needs to consider to either the set of vowels or the set of consonants based on whether the squeeze is included in the presented cue. Thus the updated mapping attempts to compromise between minimizing long-term confusion and increasing learnability (the time required for subjects to reach proficiency with the mapping) and ease of recall.

EXPERIMENTAL EVALUATION

A proof-of-concept experiment was performed using a subset of 23 phonemes and a list of 150 words that could be created with those phonemes. The subset of phonemes used in the study are shown in black in Figure 3. The goal was to see if novice users could learn to identify these 150 words in 100 minutes of training. The measure of accuracy was the result of a 50-word (subset of full 150 word list) untimed post-test following the last day of training.

Subjects

Ten Rice University undergraduate and graduate students, 4 male and 6 female, age range 19 to 30 years, provided informed consent according to the approved protocol and participated in the study. Subjects were recruited through the Mechanical Engineering Department and received a gift card for participating in the study.

Setup

All subjects wore the MISSIVE on their right arm, regardless of handedness. They were seated in front of a computer monitor and interacted with the MISSIVE using an interface built in Unity. Subjects listened to pink noise through headphones to block out sounds from servo motor and vibrotactor actuation.

Protocol

Results from pilot studies indicated that subjects learned more effectively when training was spread over multiple days and when they started learning a small subset of cues before progressing to larger ones. Therefore, we spread the 100 minutes of training across four daily sessions, and subjects focused on learning one subset of phonemes at a time. The phoneme subsets were determined by vibrotactor location: set A included cues that utilized the top vibrotactor, set B comprised the left vibrotactor cues, set C the bottom vibrotactor cues, and set D the right vibrotactor cues.

We developed a variety of exercises and assessments to guide subjects through the learning process and evaluate their progress along the way. A timeline of the protocol is shown in Table **??**. Descriptions of the exercises and assessments are provided below:

Cue familiarization: Since subjects had no prior experience with the MISSIVE device or multi-sensory cues, they were given 10 minutes on the first day to familiarize themselves with the haptic cues and the sensation of each actuator. The cues were displayed on the screen in a spatial representation of the MISSIVE layout, as shown in Figure. 3. Subjects could select a cue and click a button to feel the cue rendered on their arm. Although the corresponding phonemes were visible and audible, subjects were instructed to focus on the feeling of the cues and the individual components. Once the subjects had familiarized themselves with the various haptic sensations, they were given the option to test themselves with correct answer feedback to see how well they could identify the cues.

Learn Phoneme Set: Two exercises were developed to guide subjects through learning a new phoneme set. In the first, phonemes for a single set were displayed on the screen. Subjects could click on a phoneme to feel the haptic cue rendered on their arm and hear an audio clip of the phoneme sound. The audio clip served to clarify pronunciation and to encourage the association of the haptic cue with the phoneme's *sound*, as opposed to its written representation. When the subjects felt ready, they could advance to a self-test phase where a cue would be rendered on the MISSIVE, and they had to identify the corresponding phoneme. Correct answer feedback was provided. The second phoneme-learning exercise was designed to teach subjects how to combine phoneme sounds into words. The format was similar to the self-test phase, where subjects identified phonemes from the rendered haptic cues, except the sequence of phonemes would build a word. Subjects progressed through each phoneme at their own pace, rather than a prescribed time between phonemes, clicking a "next" button when they were ready to proceed. After all of the phonemes for a word had been presented, the user was asked to select the word that had just been sounded out from an alphabetical, multiple-choice list.

Cumulative Assessment: The cumulative assessment served as practice for word identification. The format was similar to the word-building exercise for learning new phoneme sets, in that haptic cues were presented one at a time, and the sequence of phonemes built a word. However, subjects did not respond to each phoneme individually; they only identified the word at the end. The words contained phonemes from all sets learned up until that point. Correct answer feedback was provided.

Post-Test: The Post-Test had the same format as the Cumulative Assessment, except that the subjects did not receive correct answer feedback. It was limited to 5 minutes instead of a fixed number of trials.

Pre-Test: On days 2-4, subjects began each session with a 5-minute, cumulative pre-test on all the phonemes they had learned so far. Haptic cues were rendered on the subject's arm, and they had to identify the corresponding phoneme. No correct answer feedback was provided.

[2]The implementation of the genetic algorithm was obtained from Yarpiz (www.yarpiz.com)

Day 1 *(23 mins training)*	**Day 2** *(37 mins training)*	**Day 3** *(30 mins training)*	**Day 4** *(10 mins training)*
• Cue Familiarization • Learn Phoneme Set A • Cumulative Assessment • Post-Test (A)	• Pre-Test & Review (A) • Learn Phoneme Set B • Learn Phoneme Set C • Phoneme Review • Cumulative Assessment • Post-Test (ABC)	• Pre-Test & Review (ABC) • Learn Phoneme Set D • Cumulative Assessment • Post-Test (ABCD)	• Pre-Test & Review (ABCD) • Final Test • Survey

Table 1. Training protocol: 100 minutes of training were spread over a 4-day period. Training was defined as any practice or learning exercise with correct answer feedback, as opposed to testing (pre-tests, post-tests, and final test), in which correct answer feedback was not provided.

Review: Following the pre-test, subjects were given time to review phonemes they had already learned. They could do this using the interface where they could click on phonemes to feel and hear them, as well as the interface where they could self-test. Accuracy scores for each phoneme (from pre-test results) were displayed so subjects could focus their practice accordingly.

Final Test: The Final Test consisted of a total of 50 pre-defined words (ranging from 1 to 6 phonemes, average 3.1 phonemes/word) which was a subset of the 150 words learned. The list is shown in Table 2. The test was untimed and was the same format as the Post-Test, meaning the users went at their own pace to render each of the phonemes in the word, and then were presented with a list of words to choose from. Subjects did not respond to each phoneme individually.

Survey: At the conclusion of the Final Test, subjects filled out a survey. The main questions were: *By the end of the protocol, how confident were you in your ability to correctly identify the phonemes?* and *By the end of the protocol, how confident were you in your ability to correctly identify the words?* Both items were scored on a scale of 1-5, where 1 was "not at all confident" and 5 was "very confident."

Data Analysis

Word accuracy scores were calculated for the Post-Test each day and for the Final Test. These scores were calculated as the number of correctly-identified words divided by the total number of trials. Note that since the Post-Tests are limited to 5 minutes, the distribution of words and number of trials vary across subjects.

The phoneme presentation rate was calculated by taking the average amount of time between when the phoneme was rendered and when the subject requested the next phoneme. The word response rate was calculated by taking the average time subjects took to record their response once they had the list of words to choose from.

RESULTS

The word accuracy results from the Post-Tests and the Final Test are shown in Figure. 4. The average accuracy in the Final Test was 86.6%. Seven of ten subjects scored higher than 85%, and five of ten subjects scored higher than 90% accuracy.

The average phoneme presentation rate was 3.5 seconds per phoneme, and the average word response rate was 7.7 seconds per word.

In response to the survey, on a scale of 1-5 (not at all confident - very confident), the average confidence scores were 3.2 for phoneme identification and 4.1 for word identification. These relative scores were consistent within-subjects as well: seven subjects were more confident in identifying words than phonemes, two were equally confident, and one was more confident in her phoneme responses than words.

DISCUSSION

The high accuracy scores achieved in the Final Test imply a successful preliminary validation of the MISSIVE system. With only 100 minutes of training, novice participants learned to interpret 23 multi-sensory haptic cues as phonemes and build those phonemes into a 150-word vocabulary.

Although word identification error rates were less than 15% in the Final Test and subjects were confident in their responses, user feedback provided insight into the learning experience and ways to further improve the MISSIVE system. Many subjects indicated that consonants, represented by cues without squeeze, were much easier to understand than vowels. One explanation for this trend is that since the vowel cues always had a squeeze component, they were on average more complex than the consonant cues. This additional complexity likely increased the cognitive load for novice users, making the vowels more difficult to interpret quickly and accurately. This shows room for further refinement of the haptic cue set to reduce confusion between cues.

Another possible factor is that the vowel phonemes themselves were more difficult to remember. Although most people are not accustomed to thinking about speech in terms of phonemes, the consonants were easy to learn because in most cases, there is a one-to-one correspondence between letters and the phonemes. On the other hand, each vowel (letter) is associated with multiple vowel sounds (phonemes), which not only have inconsistent spellings (for example, /UUH/ as in *put* and *foot*, but *boot* is /OO/ like *you*), but also can be quite similar to other vowel sounds (for example, /O/ as in *opt* versus /AW/ as in *awful*). This variability surely increased the difficulty of the learning process. However, subjects did not

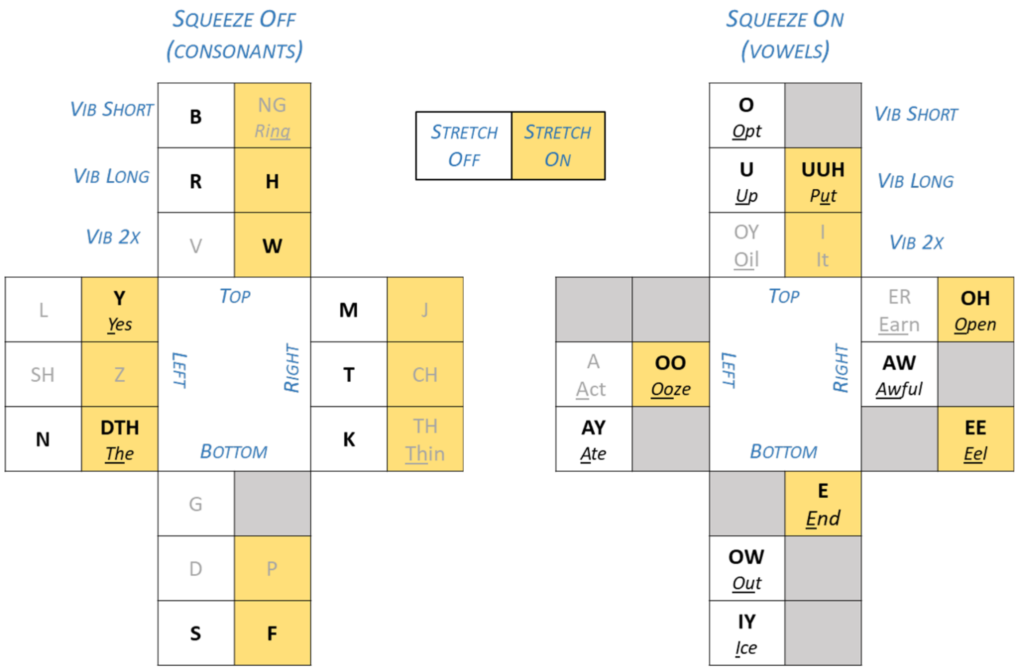

Figure 3. Spatial representation of the full cue-phoneme mapping for the MISSIVE system. Each square in the diagram represents a multi-sensory haptic cue based on its position and color and is labeled with the corresponding phoneme. The large cluster on the left represents cues with no squeeze (consonants), while the large cluster on the right represents cues with squeeze (vowels). Within each large cluster, there are 4 smaller (3x2) clusters labeled top, right, bottom and left, which correspond to the location of the actuated vibrotactor. Within each small cluster, the yellow and white squares indicate cues with and without stretch, respectively. Finally, each of the 3 rows in the small clusters corresponds to the vibrotactor pulse type: the top row is a low pulse, the middle row is a high pulse, and the bottom row is a double pulse. For example, the phoneme /AY/ is mapped to the haptic cue with the following components: squeeze on, stretch off, left vibrotactor, double pulse. The 23 phonemes used in the validation experiment are written in black text.

receive any kind of training with respect to phonemes, and doing this in the future may reduce confusion.

Despite these challenges, subjects indicated that they were more confident in their ability to identify words than individual phonemes, presumably because they were able to deduce the rendered word even if they did not understand all of the phonemes. This is partly due to familiarity with the language itself: even in spoken and written English, verbal context can provide similar cues that allow us to mentally fill in gaps when information is lost in transmission. However, the multiple choice setup of the system likely made it even easier for subjects to figure out the correct word, especially if they could confidently identify the consonant sounds. Ideally, this dependency should be mitigated, although it is difficult to do so in a way that does not significantly complicate data collection and analysis. None of the subjects indicated any discomfort with the device.

Among the low percentage of misidentified words, many of the incorrect responses were very similar to the actual rendered word. In these cases, it was obvious that the mistake was attributable to a similar phoneme sound (e.g. *who* /H/ /OO/ mistaken for *how* /H/ /OW/) or haptic cue (e.g. *coat* and *comb* differ only by the length of the vibrotactor pulse in the last phoneme). Along similar lines, subjects seemed to have particular trouble with words that did not have phonetic spellings. For example, when subjects incorrectly identified the word "who," they tended to pick words that began with an "h," because the phonemic rendering of "who" is /H/ /OO/. These errors are encouraging, as they suggest that subjects were able to understand the majority of the rendered communication. Furthermore, in any real-world application, these minor confusion errors would likely be clarified by context. We were unable to test whether the phoneme mapping — aimed at minimizing sentence-level confusion — was effective because sentences were not presented in this study. Future studies would need to be performed with sentence comprehension to test this.

While this study did not present a stream of phonemes at a fixed presentation rate, this would be a next step towards presenting full messages. A similar study encoding haptic cues to letters using vibrotactors on the back of a user's hand [13] achieved a presentation rate of 70 ms per letter with 5 hours of training. The current presentation rate of the MISSIVE is 350 ms per cue, but the device is located on a part of the body that would not be invasive for many daily activities, meaning that it does not have as fine of a perceptual resolution. Additionally,

Word	# Phonemes	Phonemes	Word	# Phonemes	Phonemes	Word	# Phonemes	Phonemes
a	1	\AY\	coat	3	\K\ \OH\ \T\	then	3	\DTH\ \E\ \N\
i	1	\IY\	comb	3	\K\ \OH\ \M\	try	3	\T\ \R\ \IY\
ace	2	\AY\ \S\	echo	3	\E\ \K\ \OH\	week	3	\W\ \EE\ \K\
are	2	\O\ \R\	face	3	\F\ \AY\ \S\	yes	3	\Y\ \E\ \S\
bye	2	\B\ \IY\	fake	3	\F\ \AY\ \K\	about	4	\U\ \B\ \OW\ \T\
he	2	\H\ \EE\	free	3	\F\ \R\ \EE\	area	4	\E\ \R\ \EE\ \U\
hi	2	\H\ \IY\	home	3	\H\ \OH\ \M\	baby	4	\B\ \AY\ \B\ \EE\
how	2	\H\ \OW\	mate	3	\M\ \AY\ \T\	from	4	\F\ \R\ \U\ \M\
me	2	\M\ \EE\	meet	3	\M\ \EE\ \T\	test	4	\T\ \E\ \S\ \T\
now	2	\N\ \OW\	name	3	\N\ \AY\ \M\	west	4	\W\ \E\ \S\ \T\
so	2	\S\ \OH\	night	3	\N\ \IY\ \T\	anyway	5	\E\ \N\ \EE\ \W\ \AY\
they	2	\DTH\ \AY\	phone	3	\F\ \OH\ \N\	anymore	6	\E\ \N\ \EE\ \M\ \AW\ \R\
us	2	\U\ \S\	right	3	\R\ \IY\ \T\	extra	6	\E\ \K\ \S\ \T\ \R\ \AH\
we	2	\W\ \EE\	room	3	\R\ \OO\ \M\	moment	6	\M\ \OH\ \M\ \U\ \N\ \T\
who	2	\H\ \OO\	same	3	\S\ \AY\ \M\	facebook	6	\F\ \AY\ \S\ \B\ \UUH\ \K\
why	2	\W\ \IY\	site	3	\S\ \IY\ \T\	tomorrow	6	\T\ \OO\ \M\ \O\ \R\ \OW\
but	3	\B\ \U\ \T\	sob	3	\S\ \O\ \B\			

Table 2. The word list for the Final Test. Each individual phoneme is indicated between two backslashes. The mapping of each phoneme to the haptic cue can be found in Figure 3.

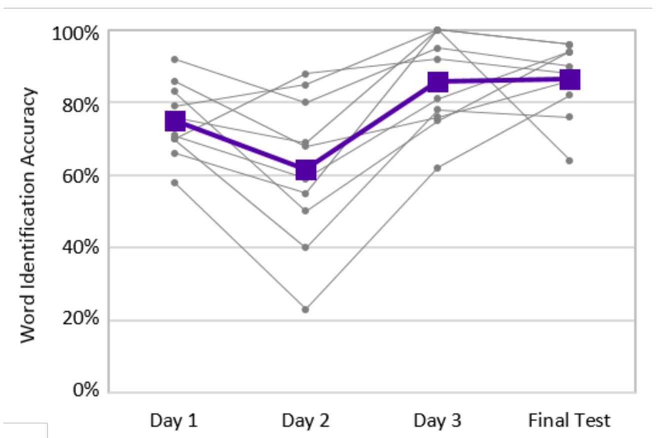

Figure 4. Word identification accuracy scores for individual subjects (gray) and on average (purple). Post-Tests on days 1-3 were a fixed duration (5 minutes), while the Final Test was a fixed number of trials (50 words).

given a similar amount of training time, users should be able to learn shorter cues and achieve a small gap in between cues.

CONCLUSIONS

A multi-sensory device, the MISSIVE, was designed and implemented to convey language through haptics. The system was validated through an experiment designed to train subjects to learn 150 words, comprising 23 phonemes, through the MISSIVE. Through the 100 minutes of training, subjects scored on average 86.6% on a final test of 50 words with an average phoneme presentation rate of 3.5 seconds per phoneme and and average word response rate of 7.7 seconds per word.

These results show promise in providing a haptic language using the MISSIVE, and there is room for further refinement of the haptic cue set in order to reduce confusion between haptic cues. There is also potential to decrease the actuation time of each cue by choosing more responsive actuators or adjusting transmission ratios, enabling words to be rendered faster.

Potential improvements could focus on extending the training time to help subjects move to a forced-pace environment with increasing speed, similar to how someone would see it in everyday use. Additionally, the phoneme set should be expanded to the full set to understand how long it would take subjects to learn.

REFERENCES

1. Edoardo Battaglia, Janelle P Clark, Matteo Bianchi, Manuel G Catalano, Antonio Bicchi, and Marcia K O'Malley. 2017. The Rice Haptic Rocker: skin stretch haptic feedback with the Pisa/IIT SoftHand. In *World Haptics Conf.* IEEE, 7–12.
2. P. L. Brooks and B. J. Frost. 1983. Evaluation of a tactile vocoder for word recognition. *The Journal of the Acoustical Society of America* 74, 1 (1983), 34–39.
3. Zvi Drezner. 2003. A new genetic algorithm for the quadratic assignment problem. *INFORMS Journal on Computing* 15, 3 (2003), 320 – 330.
4. Nathan Dunkelberger, Joshua Bradley, Jennifer L. Sullivan, Ali Israr, Frances Lau, Keith Klumb, Freddy

Abnousi, and Marcia K. O'Malley. 2018. Improving Perception Accuracy with Multi-sensory Haptic Cue Delivery. In *Haptics: Science, Technology, and Applications*. Springer International Publishing, Cham, 289–301.

5. Mario Enriquez, Karon MacLean, and Christian Chita. 2006. Haptic Phonemes: Basic Building Blocks of Haptic Communication. In *Proceedings of the 8th Int. Conf. on Multimodal Interfaces (ICMI '06)*. ACM, New York, NY, USA, 302–309.
6. Ali Israr, Peter H Meckl, Charlotte M Reed, and Hong Z Tan. 2009. Controller design and consonantal contrast coding using a multi-finger tactual display. *The Journal of the Acoustical Society of America* 125, 6 (2009), 3925–3935.
7. Ali Israr, Peter H Meckl, and Hong Z Tan. 2004. A two DOF controller for a multi-finger tactual display using a loop-shaping technique. In *Proceedings of the ASME Int. Mechanical Engineering Congress and Exposition (IMECE04)*. 1083–1089.
8. Chandrika Jayant, Christine Acuario, William Johnson, Janet Hollier, and Richard Ladner. 2010. V-braille: Haptic Braille Perception Using a Touch-screen and Vibration on Mobile Phones. In *Proceedings of the 12th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '10)*. 295–296.
9. Yang Jiao, Frederico M. Severgnini, Juan Sebastian Martinez, Jaehong Jung, Hong Z. Tan, Charlotte M. Reed, E. Courtenay Wilson, Frances Lau, Ali Israr, Robert Turcott, Keith Klumb, and Freddy Abnousi. 2018. A Comparative Study of Phoneme- and Word-Based Learning of English Words Presented to the Skin. In *Haptics: Science, Technology, and Applications*. Springer International Publishing, Cham, 623–635.
10. Edward T. Auer Jr., Lynne E. Bernstein, and David C. Coulter. 1998. Temporal and spatio-temporal vibrotactile displays for voice fundamental frequency: An initial evaluation of a new vibrotactile speech perception aid with normal-hearing and hearing-impaired individuals. *The Journal of the Acoustical Society of America* 104, 4 (1998), 2477–2489.
11. Vincent Lévesque, Jérôme Pasquero, Vincent Hayward, and Maryse Legault. 2005. Display of virtual braille dots by lateral skin deformation: feasibility study. *ACM Trans. on Applied Perception (TAP)* 2, 2 (2005), 132–149.
12. Eliane M. Loiola, Nair M. Maia de Abreu, Paulo O. Boaventura-Netto, Peter Hahn, and Tania Querido. 2007. A survey for the quadratic assignment problem. *European Journal of Operational Research* 176, 2 (2007), 657–690.
13. Granit Luzhnica, Eduardo Veas, and Viktoria Pammer. 2016. Skin Reading: Encoding Text in a 6-channel Haptic Display. In *Proceedings of the 2016 ACM International Symposium on Wearable Computers (ISWC '16)*. ACM, New York, NY, USA, 148–155. DOI: `http://dx.doi.org/10.1145/2971763.2971769`
14. Kimberly Myles and Mary S. Binseel. 2007. *The Tactile Modality: A Review of Tactile Sensitivity and Human Tactile Interfaces, U.S. Army Research Laboratory, ARL-TR-4115*.
15. Scott D. Novich and David M. Eagleman. 2015. Using space and time to encode vibrotactile information: toward an estimate of the skin's achievable throughput. *Experimental Brain Research* 233, 10 (2015), 2777–2788.
16. E. Piateski and L. Jones. 2005. Vibrotactile pattern recognition on the arm and torso. In *First Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems. World Haptics Conference*. 90–95.
17. Charlotte M Reed, Nathaniel I Durlach, and Louis D Braida. 1982. Research on tactile communication of speech: a review. *ASHA monographs* 20 (1982), 1.
18. Hong Tan, Robert Gray, J Jay Young, and Ryan Taylor. 2003. A haptic back display for attentional and directional cueing. *Haptics-e, The electronic journal of haptics research* (2003).
19. Hong Z. Tan, Nathaniel I. Durlach, William M. Rabinowitz, Charlotte M. Reed, and Jonathan R. Santos. 1997. Reception of Morse code through motional, vibrotactile, and auditory stimulation. *Perception & Psychophysics* 59, 7 (1997), 1004–1017.
20. Hong Z. Tan, Charlotte M. Reed, and Nathaniel I. Durlach. 2010. Optimum Information Transfer Rates for Communication through Haptic and Other Sensory Modalities. *IEEE Trans. on Haptics* 3, 2 (2010), 98–108.
21. Emma Treadway, Brent Gillespie, Darren Bolger, Amy Blank, Marcia K. O'Malley, and Alicia Davis. 2015. The role of auxiliary and referred haptic feedback in myoelectric control. In *World Haptics Conf.* IEEE, Northwestern University, 13–18.
22. Robert Turcott, Jennifer Chen, Pablo Castillo, Brian Knott, Wahyudinata Setiawan, Forrest Briggs, Keith Klumb, Freddy Abnousi, Prasad Chakka, Frances Lau, and Ali Israr. 2018. Efficient Evaluation of Coding Strategies for Transcutaneous Language Communication. In *Haptics: Science, Technology, and Applications*. Springer International Publishing, Cham, 600–611.
23. Siyan Zhao, Ali Israr, Frances Lau, and Freddy Abnousi. 2018. Coding Tactile Symbols for Phonemic Communication. In *Conference on Human Factors in Computing Systems*.