A Multi-sensory Approach to Present Phonemes as Language through a Wearable Haptic Device

Nathan Dunkelberger, *Student Member, IEEE,* Jennifer L. Sullivan, *Member, IEEE,* Joshua Bradley, Indu Manickam, Gautam Dasarathy, Richard Baraniuk, and Marcia K. O'Malley, *Fellow, IEEE*

Abstract—Communication is an important part of our daily interactions; however, communication can be hindered, either through visual or auditory impairment, or because usual communication channels are overloaded. When standard communication channels are not available, our sense of touch offers an alternative sensory modality for transmitting messages. Multi-sensory haptic cues that combine multiple types of haptic sensations have shown promise for applications, such as haptic communication, that require large discrete cue sets while maintaining a small, wearable form factor. This paper presents language transmission using a multi-sensory haptic device that occupies a small footprint on the upper arm. In our approach, phonemes are encoded as multi-sensory haptic cues consisting of vibration, radial squeeze, and lateral skin stretch components. Participants learned to identify haptically transmitted phonemes and words after training across a four day training period. A subset of our participants continued their training to extend word recognition free response. Participants were able to identify words after four days using multiple choice with an accuracy of 89% and after eight days using free response with an accuracy of 70%. These results show promise for the use of multi-sensory haptics for haptic communication, demonstrating high word recognition performance with a small, wearable device.

Index Terms—wearable haptics, language communication, multi-sensory haptics, tactile device, phoneme coding.

1 INTRODUCTION

Communication is an important part of our daily experiences as we interact with others. Our visual and auditory senses are the primary means by which we communicate; however, there are circumstances when these typical communication channels are unavailable, either through saturation or impairment. Communication saturation can occur when visual and auditory channels must be entirely focused on a demanding task, such as a surgeon operating on a patient, or a pilot attending to a complex array of visual displays in a cockpit. Visual or auditory impairment can also impede communication with the environment, requiring more information to be taken in through remaining available communication methods, or causing an individual to miss portions of information entirely. In cases where the standard auditory and visual communication channels are not available, haptics, or the sense of touch, can provide a means for communication through the skin, the largest organ in the human body.

Haptic communication has roots as far back as the 1890s [1] when the Tadoma method was developed for deaf-blind individuals to understanding spoken language. The technique involves users putting their hand on a speaking person's neck and face, to feel the movements made during speech and interpret those movements as words. The Tadoma method demonstrated the potential of the sense of touch to serve as an alternative communication channel to auditory and visual means [2].

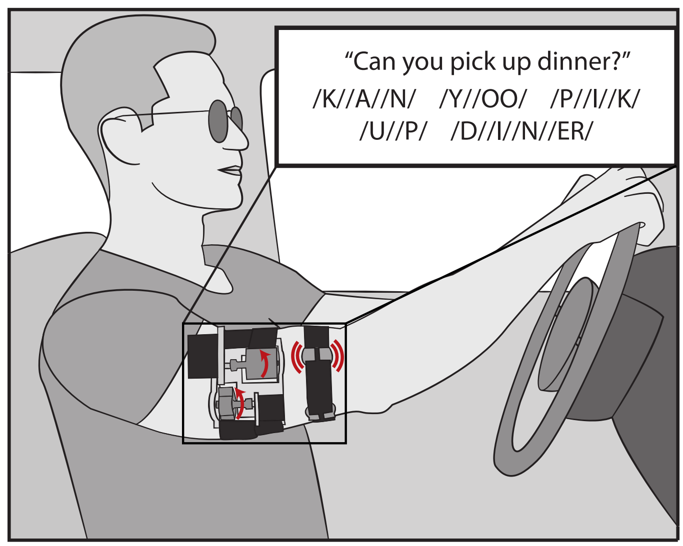

Fig. 1. The Multi-sensory Interface of Stretch, Squeeze, and Integrated Vibration Elements (MISSIVE) is a wearable communication device which can present phonemes in the English language through multi-sensory haptic cues, allowing communication to a user without sacrificing visual or auditory communication channels.

Researchers have developed a number of prototype communication systems [3], [4], [5] that are intended to present sounds or words using the sense of touch, without requiring the "listener" to place their hands on the speaker, while still maintaining an efficient and effective method of trans-

- *This work was supported by Facebook, Inc.*
- *N. Dunkelberger, J. L. Sullivan, J. Bradley, and M. K. O'Malley represent the Mechatronics and Haptic Interfaces Lab, Department of Mechanical Engineering, Rice University, Houston, TX, 77005. E-mails: {jls3, nbd2, jmb25, omalleym}@rice.edu*
- *I. Manickam and R. Baraniuk represent the Department of Electrical and Computer Engineering, Rice University, Houston, TX, 77005. E-mails: {indu.manickam, richb}@rice.edu*
- *G. Dasarathy represents the School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ, 85281. E-mail: gautamd@asu.edu*

Manuscript received DATE; revised DATE.

mitting those messages. As mobile devices and wearable technologies become commonplace, there has been heightened interest in developing *wearable* haptic devices to enable communication [6], [7], [8], [9]. In addition to developing the hardware itself, it is also necessary to consider the means by which words and messages will be encoded and transmitted.

We posit that a wearable haptic device for communication should have the following features.

1) **Convey an unconstrained vocabulary (e.g. not limited to a pre-defined list of words or phrases).**
To be able to communicate language in any capacity, we must break language down to its building blocks so that these components can be used to create complete phrases. This limits the options to letters (the most common building blocks of written communication), phonemes (discrete building blocks of aural communication), or continuous-time features (continuous building blocks of aural communication). Each of these options enables limitless communication via the English language.
2) **Transmit messages at a rate that is comparable to spoken language.**
The speed of communication is an important factor in making a device usable in an everyday setting. To achieve communication rates comparable to spoken language, it is important to both convey large *amounts* of information in each haptic cue, and to provide haptic cues *quickly*. To achieve fast communication rates, research indicates that the information per cue should be maximized [10]. Transmission speed of each haptic unit will also play a part in efficiency of communication. To provide high information transfer rates, we must design cues that balance the amount of information contained in each cue and the speed at which the cues can be transmitted.
3) **Fit comfortably and occupy a small surface area (so as to not limit activities of daily living).**
We must be able to present our information-rich haptic cues in a small form factor, to limit the intrusiveness of the device. Recently, multi-sensory haptic devices have been developed that simultaneously stimulate the different types of available mechanoreceptors in the skin. Multi-sensory haptic devices that stimulate multiple mechanoreceptors simultaneously have the potential to occupy smaller form factors than single modality (e.g. all vibration) haptic devices.
4) **Require less than eight hours of training to learn how to interpret the haptic cues as words.**
For a haptic communication device to become adopted, it must also be learnable in a reasonable amount of time. End-users will have varying levels of willingness to learn to use a new device, so the training time to achieve proficiency should aim to be as low as possible.

This paper is organized as follows. Section 2 provides related work in the field of haptic communication, addressing the state of the art in encoding language as haptic cues, and recent advances in the design of wearable haptic devices. Section 3 describes an experimental evaluation of MISSIVE (Multi-sensory Interface of Squeeze, Stretch, and Integrated Vibration Elements) and our method for encoding phonemes as multi-sensory haptic cues. Section 4 presents results from the user study, and Section 5 analyzes these results and discusses the implications.

2 Related Work

There are two fundamental areas of prior work that are relevant when developing a wearable haptic system that can convey words to a user. In this section, we summarize the state of the art for encoding language in ways that facilitate haptic transmission, followed by a review of wearable hardware systems that have been proposed for this task.

2.1 Language Encoding

There are two common methods for encoding language for use in a haptic communication device. One method uses *continuous-time features*, where features of an acoustic signal are represented using a haptic device. The other common method is to leverage the already-available *discrete language components* in written or spoken language, and map those to discrete haptic cues. Discrete language components include letters, the most basic building block of written language, and phonemes, the most basic building block of spoken language.

2.1.1 Continuous-Time Features

One of the earliest approaches to communicating speech through haptics used a class of devices called vocoders, which transform speech into its respective audio frequency components, then transmit these components through a haptic display mapped to each component. Because the continuous time information from the sound waves is mapped to a continuous haptic display, the approach is viable for transmitting an unconstrained vocabulary to the user. Haptic communication devices that use continuous-time features generally use continuous cues that represent some relationship to the desired acoustic feature set. A common way to use continuous-time features in a discrete display is to decompose the audio signal into frequency bands, and represent each of those bands through a different discrete cue on the haptic device. The relative magnitudes of the frequency bands are then displayed as continuously varying amplitudes of the discrete cues. Researchers using this approach have explored transmitting haptic cues to different locations on the skin, varying the distribution of frequency bands, and changing the acoustic features that are conveyed through the cues.

An early example of this approach was proposed by Gault et al. in 1928, who developed a device consisting of five vibrators that would be probed by each of the five fingers on a user's hand [11]. The system filtered the incoming audio signal into five frequency bands (0-250 Hz, 250-500 Hz, 500-1000 Hz, 1000-2000 Hz, and 2000+ Hz), amplified the signal, and displayed each of the bands to a different finger. More recently, a device was developed that provides vibration to the thumb, index, and middle fingers of a user's hand [3], [12]. This device, like the prior system, separated the incoming acoustic signal into five frequency bands, but the band regions differed from those used by Gault. Here, both the index and middle fingers were presented with the amplitude envelopes of two frequency bands, and the thumb received a filtered signal from the lowest frequency band. Experimental results indicated that users could accurately discriminate pairs of closely related phonemes, but this research did not evaluate users' ability to identify a comprehensive set of phonemes or words. Novich et al. developed a frequency-based haptic vest that used 27 Eccentric Rotating Mass (ERM) motors on the back. They presented different acoustic frequency bands to different motors [13], and taught users to identify words. The system required extensive training (300 training trials per day for an average of 11 days of training) and was only demonstrated in a small pilot study (N=2). The average word identification accuracy only reached approximately 60% when presented with four choices from an unseen set of words.

The continuous-time feature approach to encoding language can often take significant amounts of training to achieve satisfactory performance [14]. For example, while one study achieved 90% accuracy in word recognition, 70-80 hours of training was required to reach that level [15]. Further, the comprehension capabilities with training words generally does not translate well to unseen words [13], [14]. As a result, researchers have sought other methods for mapping language to haptics.

2.1.2 Discrete Language Components

Discrete components of language provide building blocks that, when assembled, create words and phrases that can be perceived by the recipient. In this approach to haptic language encoding, discrete haptic cues are created for each building block. Example building blocks can be letters (26 for the English language) or phonemes (35 to 44 for spoken English, depending on criteria [16]), which can be strung together to create an unconstrained vocabulary. In this section, a *cue* will refer to the haptic representation of a single language component (e.g. letter or phoneme).

Letters. There are several ways in which letters can be represented, including encoding letters directly to discrete haptic cues [7], [17], or using pre-defined sets of symbols that represent letters, such as Morse code [4] or Braille [5]. The concept of encoding letters with unique haptic patterns was explored as early as 1957 when Geldard used a system called vibratese, where letters and numbers were encoded to five vibrotactors on the chest [17], with cues varying in intensity and duration. Experiments were effective in demonstrating the potential to train users to perceive the letters, and participants were able to understand with "satisfactorily high" accuracy after 12 hours of training although the specific accuracy was not specified.

Another approach to directly encode letters used six vibration motors placed in a glove on the back of a user's hand, where letters were rendered by actuating specific motors in a specified temporal scheme [7]. Results showed 94% accuracy in letter identification, with training spread between five roughly one-hour sessions. Further testing was performed to show that these cues and associations could even be learned passively while participants were performing other tasks with low cognitive loads [18].

Haptic Morse code was explored using an electrodynamic minishaker that represented dots or dashes by the duration of a fingertip displacement [4]. This representation has the advantage that participants must only understand a small subset of haptic stimulations—dot and dash—and only remember how those cues correspond to letters or numbers. Morse code has been well-demonstrated as a reliable method of encoding messages; however, because the presentation is temporal in nature, the time required per cue is typically slower than when providing a letter or phoneme directly.

Phonemes. One of the more recent developments in haptic communication research has been in the use of phonemes, the phonetic building blocks of spoken language, as the building block for haptic communication. Typically, there are fewer phonemes per word than there are letters per word, meaning fewer haptic cues would be required to render a word. For example, the word *zoo* has three letters, but only two phonemes: /z/ and /u/, requiring three haptic cues to present each letter, but only two haptic cues to present each phoneme.

In one approach, researchers developed a 24-vibrotactor sleeve, placed around the user's forearm, that conveys 39 sensations mapped to 39 phonemes in the English language [9]. The system uses spatiotemporal vibration cues, providing sequential vibrations around the user's arm in a predefined order. Several participants learned to identify a subset of 10 phonemes and 51 words with 60 minutes of training with a 97.5% success rate, and one user was able to recognize the full set of 39 phonemes with 94% accuracy after a total of 80 minutes of training. The researchers did not explore the ability of participants to perceive strings of all 39 phonemes as words in this experiment.

Because so many phonemes must be learned when using this method of encoding, some exploratory studies have focused on training. For example, groups have experimented with small sets of phonemes (up to 13) combined into words and conveyed using a 6-vibrotactator sleeve [6], [19]. After training for 65 minutes, two-thirds of participants were able to select words from a list of 100 possibilities with over 90% accuracy using multiple choice, but testing was not expanded to further numbers of phonemes.

There are drawbacks to using phonemes as the building block when encoding words as haptic cues. The phonemic composition of words is determined by pronunciation and therefore is affected by accents and other regional pronunciation differences. For example, many people merge two vowel sounds in some contexts, such as the pin-pen merger [20], which is prevalent in the southern United States. Many people are also not experienced with consciously thinking about language in terms of phonemes, so they must learn how to think in terms of phonemes first, and then understand how the cues correspond to those phonemes. It would also be difficult to render stressing of specific syllables using haptics, requiring an additional cue component to represent the concept of stress if desired.

2.2 Wearable Devices

Recent research has studied how to make haptic language communication devices more wearable and therefore viable and practical for everyday use. Areas that have been considered include the forearm [6], [9], [19], [21], [22], upper arm [23], and the back [8]. The use of these areas of the body, while well-suited for wearable devices that will not interfere with everyday activities, introduces trade-offs in terms of perceptual performance, since the density of mechanoreceptors in the skin of the back and arm is not as great as in glaborous skin such as that on the fingertips or face.

The vibrotactile systems described in Section 2.1 often rely on arrays of vibrotactors to generate the discrete haptic cues that correspond to each frequency band, letter, or phoneme. For letters and phonemes in particular, large numbers of cues are necessary for encoding. Vibrotactile arrays contained many tactors covering significant portions of the body. Further, to realize large numbers of distinct haptic cues, researchers often vary frequency, amplitude, pattern, and duration to generate distinct cues that can be reliably perceived. For example, cue amplitude and frequency can vary temporally, and multiple tactors can be actuated in sequence to create cues with sequential cue components [6], [8], [9], [19], [21]. While this is an effective method for creating large cue sets, such an approach typically means that cue-length increases in duration, limiting the rate at which messages can be conveyed via vibrotactor arrays alone.

Multi-sensory haptic systems [24], [25], [26], [27], which render different types of haptic cues (vibration, squeeze, skin stretch, and more), offer the ability to generate large sets of haptic cues without covering the skin area necessary for single haptic systems, such as vibrotactor arrays. In addition, because different mechanoreceptors are targeted, haptic cue components can be presented simultaneously, reducing cue presentation time compared to vibration-only systems that utilize temporal presentations to provide enough information per cue [10]. In an experiment comparing cue identification accuracy with a multi-sensory device (MISSIVE) and a comparably-sized vibrotactile only device, performance was was superior in the multi-sensory device [23], motivating their use for haptic communication systems.

3 METHODS

An experiment was performed where participants learned to identify words conveyed through the MISSIVE. The phonemes in the English language were mapped to multi-sensory haptic cues of the MISSIVE, and a training protocol was developed to teach participants to recognize phonemes and words using the MISSIVE.

3.1 Hardware

The MISSIVE, shown in Fig. 2, consists of three actuation mechanisms: *radial squeeze*, *lateral skin stretch*, and *vibration* [23]. These three actuation types are housed in two bands, referred to as the distal band and the proximal band, which are positioned in the middle of the user's upper arm.

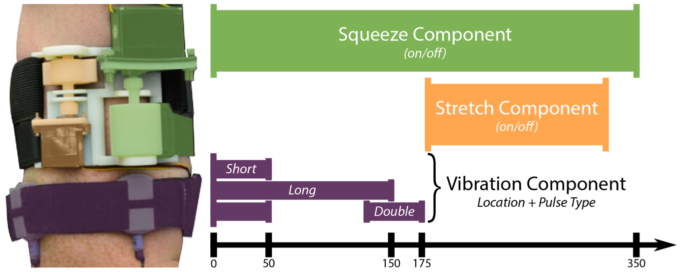

Fig. 2. The MISSIVE is shown on the left. Multi-sensory haptic cue timing for skin stretch, squeeze, and vibration components is shown on the right. The squeeze and stretch components can either be on or off, and the vibration component is specified by pulse type of short, long, or double, and location of top, right, bottom, or left.

3.1.1 Vibration Element

The distal band houses the vibration element for the MISSIVE, consisting of a Velcro band wrapped around four vibrotactors (C2 Tactors, Engineering Acoustics Inc.). These vibrotactors are enclosed in a 3D-printed case that allows them to slide along the length of the band to adjust for different user arm sizes. The vibrotactors are evenly spaced around the top, left, bottom, and right of the user's arm, with roughly 8.3 cm inter-actuator spacing for an averaged sized upper-arm [28].

3.1.2 Radial Squeeze Element

The proximal band houses both the radial squeeze and lateral skin stretch actuation elements on a 3D-printed base which is held around the arm with a velcro strap. The radial squeeze element is driven by a servomotor (HS-485HB, Hitec RCD USA, Inc.) mounted on the 3D-printed structure. This servomotor is connected to a shaft that contains another Velcro band that is wrapped around the user's arm. When the servomotor actuates, it decreases the radius of the band wrapped around the user's arm, causing a squeezing action, before returning to its starting position.

3.1.3 Lateral Skin Stretch Element

The skin stretch element is driven by a servomotor (HS-5070MH, Hitec RCD USA, Inc.) mounted on the 3D-printed structure and connected to a 3D-printed, semi-circular shaft with a rubber edge that makes contact with the skin. When the servomotor actuates, the skin in contact with the rubber portion of the shaft is stretched to one side, and then is returned to its original position. This element is based on the Haptic Rocker design by Battaglia et al. [29], [30].

3.2 Haptic Cue Set

The MISSIVE was used to provide discrete multi-sensory haptic cues corresponding to phonemes. The modes of actuation for each component were determined based on pilot tests performed by the researchers. A *component* refers to a particular actuation type: vibration, lateral skin stretch, or squeeze. A *multi-sensory haptic cue* refers to a combination of commanded actions to one or more components. The specific actuation strategies for each component are detailed below, and the timing of each component to create a single multi-sensory haptic cue is shown in Fig. 2.

3.2.1 Vibration

The vibration component was realized using one of four vibrotactors on the distal band on the user's arm (top, right, bottom, or left location, based on the orientation of a user's arm holding a computer mouse in their right hand). One of three different vibration patterns were used, either a short pulse (50 msec), a long pulse (150 msec), or a double pulse (one 50 msec pulse, a 75 msec pause, then another 50 msec pulse), as shown in Fig. 2. The pulses were all displayed at 265 Hz, the resonant frequency of the tactor, at maximum amplitude.

3.2.2 Lateral Skin Stretch

The lateral skin stretch component was a binary cue (actuation or absent). If the cue actuated, the rocker was commanded to rotate to the left for 75 ms, and then was commanded back to its original position, taking 150 ms and rotating approximately 30 degrees.

3.2.3 Radial Squeeze

The radial squeeze component was also a binary cue, either actuating or not. If the cue actuated, the motor was commanded to tighten the band for 175 ms, then was commanded back to its original position, taking 350 ms.

3.3 Mapping Multi-sensory Haptic Cues to Phonemes

An optimization algorithm was used to determine the mapping between multi-sensory haptic cues and phonemes. First, pilot testing data were used to identify common confusion patterns between multi-sensory haptic cues. Then, the optimization algorithm was used to map phonemes to cues such that if one of the common multi-sensory haptic cue identification errors occurred, the corresponding phoneme confusion would not produce a word that could logically replace it in context.

The mapping was generated by optimizing the cost function defined in (1). 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 $|N| = 39$ 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 *multi-sensory haptic cue* for another, and was determined empirically using pilot studies (as in [23]) to determine cue confusion over the full multi-sensory haptic cue set. $F : N \times N \to R^+$ represents the cost of confusing one *phoneme* for another.

The mapping was constrained to assign all consonants to cues with squeeze *off* and assigning all vowels to cues with squeeze *on*. This constraint is intended to aid with memory recall, as it reduces the number of search items that the participant needs to consider. For example, based on the presence or absence of the squeeze, cue, the user need only identify one of the set of vowel phonemes, or one of the set of consonant phonemes.

The term, F, was estimated based on the principle that translation errors where the user determines that the perceived word is nonsense are preferable to errors where the user is unaware that an error has occurred and therefore perceives the wrong word. 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. Higher weights were 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 would most likely result in nonsense at the sentence level. For example, if the word *cat* was understood as *cas*, due to the confusion between the multi-sensory haptic cues corresponding to the phonemes /t/ and /s/, the algorithm would assume the user could identify that there was an error. In contrast, if the word *cat* was confused with *cab*, the algorithm would penalize the confusion between /t/ and /b/ heavily because the replacement of one noun with another noun may confuse the user, even with context.

Using F and D, we define the total expected cost of a mapping as:

$$C(\varphi) = \sum_{(i,j)\in N} F(i,j) D(\varphi(i), \varphi(j)) \tag{1}$$

A visual representation of the cue-phoneme mapping used in the experiment is shown in Fig. 3.

In (1), 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 $\varphi^* = \operatorname{argmin}_{\varphi} C(\varphi)$. This problem is a variant of the Quadratic Assignment Problem (QAP), a long-standing combinatorial optimization problem. The QAP has been shown to be NP hard, and therefore in practice approximate solutions are found using local search algorithms [31]. The approximated optimal mapping for this problem was determined using the genetic algorithm[2] [32].

3.4 Experimental Protocol

After mapping all 39 phonemes optimally to a subset of the 48 possible multi-sensory haptic cues, we conducted an experiment to evaluate participants' ability to learn 23 phonemes and, subsequently, words that could be constructed with these phonemes. The phoneme set was selected by choosing phonemes used in a set of 50 words which would be useful in a messaging context. Using the 23 phonemes present in that word set, the word set was expanded to 150 words.

The 23 selected phonemes were divided into four sets (A, B, C, and D) based on location of active vibrotactor component in the corresponding multi-sensory haptic cue (top, left, bottom, and right). Participants were introduced to one or two sets of phonemes each day. All participants completed the same protocol for Days 1 through 4, while a subset of five participants continued to an extended version of the protocol, which included an extra four days, to see if their ability to recognize words could be transferred to a free response paradigm.

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 CMUdict (http://www.speech.cs.cmu.edu/cgi-bin/cmudict) for word pronunciation.

2. The source code for the genetic algorithm code was generated by Yarpiz (www.yarpiz.com)

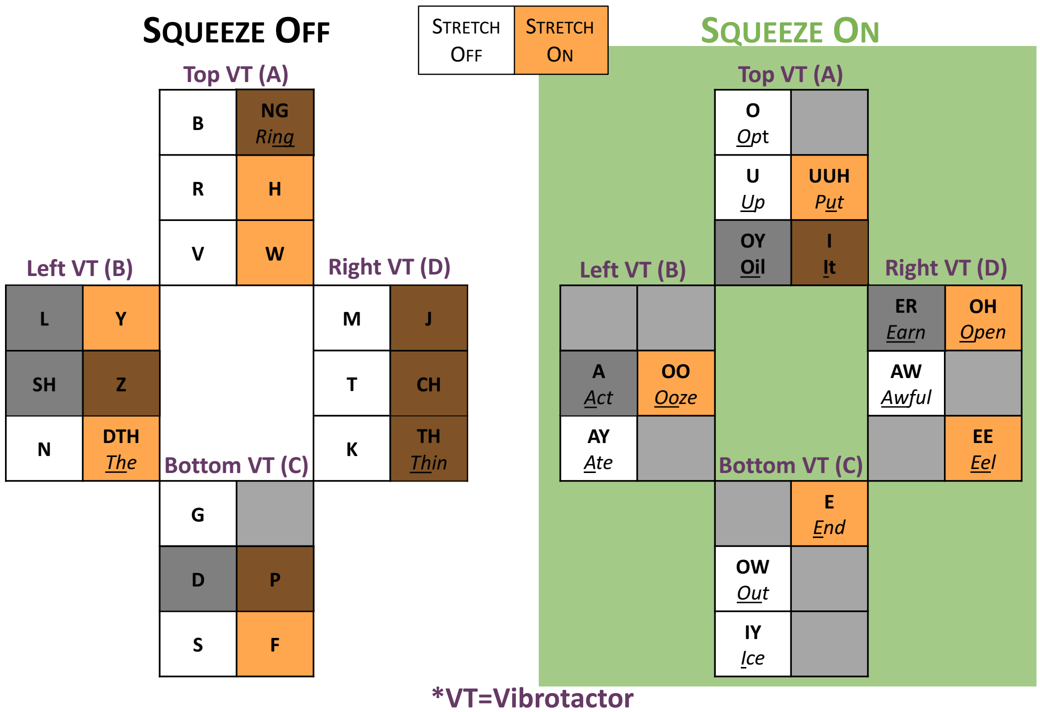

Fig. 3. Visual representation of the mapping showing the correspondence between the multi-sensory haptic cues and assigned phonemes. In this figure, the top, right, bottom, and left subsections of 6 cues each correspond to their respective vibrotactor location, with corresponding set labeled in parenthesis. Type of vibrotactor vibration is represented as being the top, middle, or bottom of these subsections representing short, long, and double pulses. Stretch is indicated as on by showing an orange filled square rather than the white squares. The squeeze component being activated is indicated as being in the right duplication of the diagram rather than the left. Box shading shows whether each phoneme was a part of the subset of 23 phonemes used for this experiment, with grayed out phonemes representing those not used in this experiment.

Training time only included exercises where participants received correct answer feedback. The term *test* was reserved for exercises in which participants did not receive correct answer feedback. Pink noise was played in headphones throughout the experiment so that participants could not hear the sound of the device actuating. Explanations of the different training exercises are detailed in Section 3.4.2. A detailed diagram of the protocol schedule, including the progression of haptic phoneme cue sets during training, is shown in Fig. 4.

3.4.1 Participants

Sixteen Rice University undergraduate and graduate students (7 female, average age 23.4, range 19-30) participated in the experiment. All participants gave informed consent and received a gift card for each day that they participated in the experiment. A subset of eleven participants completed only the first four days of the training protocol, referred to as *training part 1*. The other of five participants completed both training part 1, and a subsequent four days, referred to as *training part 2*, for a total of eight days of training.

3.4.2 Training and Testing Exercises

Cue Familiarization. All participants began the protocol with ten minutes to familiarize themselves with how the multi-sensory haptic cues felt. During this time, participants could click on a cue, shown in a spatial representation of the cues similar to Fig. 3, then click a button to render the cue on their arm. Participants could continue to explore the multi-sensory haptic cues until they had fully familiarized themselves with the different cue types and interactions. While participants were instructed to focus on the multi-sensory haptic cues, the phoneme text representation was shown on the screen and an audio clip corresponding to that phoneme was played. Once participants were comfortable with the cue presentations, they could proceed to a self-test exercise. This self-test entailed the rendering of random multi-sensory haptic cues followed by a prompt to identify the three cue components. Correct answer feedback was provided to participants.

Learn Set A/B/C/D. Participants learned each set of phonemes based on vibrotactor location as shown in Fig. 3. Because sets were based on vibrotactor location, participants experienced variation in all available types of cue components except for vibration location during each training set. Participants began learning each new set by viewing the interface with all sections grayed out except the one that they were currently learning. In this exercise, participants had five minutes to memorize the association between the phonemes and the cues. During this exercise, participants completed two different phases. In the first phase, when participants chose a phoneme, the corresponding cue was played on the MISSIVE and the sound corresponding to the phoneme was played in the participant's headphones. When the participant was confident that they had memorized the phonemes, they would move onto the second phase, where they could use any remaining time to test themselves by rendering a random phoneme from that set, and responding with what phoneme they thought it was. After responding, the correct cue played again on the user's arm, and the

EXERCISE	CUE SETS				TIME
Day 1 (23 mins training)					
Learn Haptic Cues	(All)				10 mins
Learn New Phonemes	**A**	B	C	D	5 mins
Intro to Words	**A**	B	C	D	3 mins
Cumulative Assessment	**A**	B	C	D	5 mins
Post-Test	**A**	B	C	D	(5 mins)
Day 2 (37 mins training)					
Pre-Test	**A**	B	C	D	(14 trials)
Review Day 1	**A**	B	C	D	5 mins
Learn New Phonemes	A	**B**	C	D	5 mins
Intro to Words	**A**	**B**	C	D	5 mins
Learn New Phonemes	A	B	**C**	D	5 mins
Intro to Words	**A**	**B**	**C**	D	5 mins
Cumulative Review	**A**	**B**	**C**	D	7 mins
Cumulative Assessment	**A**	**B**	**C**	D	5 mins
Post-Test	**A**	**B**	**C**	D	(5 mins)
Day 3 (30 mins training)					
Pre-Test	**A**	**B**	**C**	D	(34 trials)
Review Day 2	**A**	**B**	**C**	D	10 mins
Learn New Phonemes	A	B	C	**D**	5 mins
Intro to Words	**A**	**B**	**C**	**D**	5 mins
Cumulative Assessment	**A**	**B**	**C**	**D**	10 mins
Post-Test	**A**	**B**	**C**	**D**	(5 mins)
Day 4 (10 mins training)					
Pre-Test	**A**	**B**	**C**	**D**	(46 trials)
Review Day 3	**A**	**B**	**C**	**D**	10 mins
Final Test	**A**	**B**	**C**	**D**	(50 trials)
Days 5-8: Free Response (15 mins training)					
Pre-Test	**A**	**B**	**C**	**D**	(46 trials)
Review	**A**	**B**	**C**	**D**	5 mins
Cumulative Assessment*	**A**	**B**	**C**	**D**	10 mins
Final Test*	**A**	**B**	**C**	**D**	(75 trials)

Fig. 4. Protocol for the study. Total training times for each day are shown in parenthesis next to the day and this includes all time in the white squares, indicating there is correct answer feedback. The exercises with * indicate that the response to words presented are using free response rather than multiple choice.

correct phoneme was shown on the screen, along with the sound made by the phoneme in the user's headphones. Phonemes in the second phase were presented in such a way that subjects would receive one of each available phoneme before repeating any of the phonemes so that they practiced on a uniform distribution of phonemes.

Review. When Reviewing, participants had time to review the phonemes that they had learned so far and were shown what phonemes they missed in the Pre-Test for the day to help guide practice. This section was similar to the Learn Set A/B/C/D section in that participants could choose to play a desired cue on MISSIVE, or self-test themselves with random phonemes, except that all phonemes that participants had learned so far were available.

Introduction to Words Set A/B/C/D. In the Introduction to Words exercise, participants began to make the connection between phonemes and words. For a given word, each phoneme in the word was presented in an isolated manner, with the participant choosing a phoneme from a multiple choice list before proceeding to the next one. Once all of the phonemes for a word had been played, the participants were shown a multiple choice list of twelve words from which to choose their response. In this exercise, correct answer feedback was provided for both phonemes and words. Words could be comprised of phonemes from the current set or from prior sets, but every word presented had at least one phoneme from the current set to reinforce the learning of the new phonemes.

Cumulative Assessment. In the Cumulative Assessment section, participants were given phonemes sequentially to make up a word, as in the Introduction to Words section, but the participant did not identify each phoneme individually. After all of the phonemes of a given word had been played, the participant selected the word from a twelve-word multiple choice list, and they were provided with correct answer feedback.

Pre-Test. In the Pre-Test, participants were presented with a multi-sensory haptic cue and asked to select the corresponding phoneme from a multiple choice list of all of the phonemes they had learned so far. Two of every phoneme that the participant had learned so far were played and participants did not receive correct answer feedback.

Post-Test. The Post-Test measured word identification in the same way as the Cumulative Assessment section, but participants did not receive correct answer feedback.

Final Test. The Final Test section was the same in structure to the Post-Test section, except that in the Final Test, the participants responded to a predetermined subset of 50 words for consistency in data analysis. The number of words tested in this section was determined by pilot testing so that a comprehensive set of words could be tested, and the total time of the day would remain under an hour.

Free Response sections. The free response section was only presented in training part 2. Each free-response section was structured similarly to the corresponding section in Days 1 through 4, but rather than selecting from a set of multiple choice options to identify words, participants had to type in their answers and click a submit button. For these participants, the trial count for the Final Tests was increased from 50 to 75 during free response evaluations so that every word would be presented twice during Days 5 through 8.

3.5 Data Analysis

Phoneme recognition was assessed using data collected from the Pre-Tests that participants took each day. Word recognition was assessed based on the Post-Tests that participants took, and was further separated by whether the response type was multiple choice or free response.

3.5.1 *Phoneme Recognition*

Phoneme accuracy was calculated from the Day 4 Pre-Test, once all of the phonemes had been introduced, as well as on Day 8, the last day of training part 2.

A confusion matrix was created from the pretest on Day 4 to aid in identifying trends in the mistakes that participants were making. The confusion matrix also allows phoneme recognition accuracies to be broken down for each subgroup of consonants and vowels. In addition, the frequency that vowels were confused for consonants and that consonants were confused for vowels was calculated.

In some cases, participants proceeded to the next stage without completing all trials for the Pre-Test. When this occurred, the skipped phonemes were excluded from analysis of performance. This occurred in 36 out of 736 phoneme presentations on Day 4 assessments, and only one time out of 230 phoneme presentations on Day 8 assessments.

3.5.2 Word Recognition

Individual and average word recognition accuracies were calculated for the post-tests in training part 1. Individual and average word recognition accuracies were calculated for the Free Response Final Tests in training part 2. When comparing the free response accuracy from training part 1 with results from training part 2, only results from participants who participated in both parts were used to ensure consistency in analysis. For all Free Response Final Tests, the incorrect responses were checked and marked correct if they responded with a homonym of the correct answer.

Transitioning from multiple choice to free response assessments allowed further investigation as to how the participants were interpreting phonemes and words. To understand trends, the percentage of incorrect responses that were incorrect by only one phoneme were calculated. A response was considered to be incorrect by only one phoneme if it met one of two criteria: (1) The user response contained the same number of phonemes as the rendered word and differed by only one phoneme. For example, if the rendered word was *talk* but the user responded with *took*. (2) The presented and responded word differed in length by one phoneme, but all of the phonemes in the shorter word appear in the longer word in the correct order. For example, if the rendered word was *baby* but the user responded with *babe*.

4 Results

An experiment was performed in which sixteen participants underwent 100 minutes of training and were evaluated on their ability to recognize 23 phonemes presented as multi-sensory haptic cues via the multi-sensory device, the MISSIVE. Participants achieved an average phoneme identification accuracy of 61.4%, and an average word identification accuracy of 89.4% when presented with twelve multiple choice options. Five participants continued their training for an additional four days, and achieved an average phoneme identification accuracy of 85.2% and an average word identification accuracy of 70.9% when providing free responses instead of multiple choice.

4.1 Phoneme Identification Accuracy (Days 1-4)

Pre-Test phoneme identification accuracies for days 1 through 4 are presented in the left pannel of Fig. 6. Phoneme identification accuracy for participants on Day 4 (the last day of training part 1) was 61.4%. The subset of participants who completed training part 1 and training part 2 achieved an accuracy of 75.5%. Vowels and consonants were separately analyzed, showing that the accuracy for correctly identifying consonants was 82.3%, while the accuracy in correctly identifying vowels was 38.4%. The confusion matrix (Fig. 5) shows that participants identified presented vowels as consonants only 5.7% of the time, meaning that participants correctly identified presence of the squeeze cue 94.3% of the time. Similarly, participants identified presented consonants as vowels only 3.8% of the time, meaning that participants correctly identified the absence of the squeeze cue 96.2% of the time.

4.2 Phoneme Identification Accuracy (Days 5-8)

Phoneme identification accuracy for only those participants who continued to training part 2 rose by an average of 9.7% (from 75.5% to 85.2%). This group's accuracy in identifying vowels rose by 19.5%, from 54.1% to 73.6%, and their accuracy in identifying consonants remained high at 95.8%, a 0.8% increase from Day 4.

4.3 Word Identification Accuracy (Multiple Choice)

Post-test word identification accuracy for Days 1 through 4 is shown in Fig. 7. Throughout Days 1, 2, and 3, participants recorded 80.4%, 70.9%, and 90.3% accuracy, respectively in the 5-minute timed post-tests, corresponding to sets of 7, 17, and 23 phonemes, respectively, comprising the post-test word sets. On Day 4, participants correctly identified on average 89.4% of a 50-word set that comprised all 23 phonemes. The subset of participants who completed training part 1 and training part 2 achieved an average of 93.6%.

4.4 Word Identification Accuracy (Free Response)

Average word identification accuracy for only those participants who continued to training part 2 dropped from 93.6% on Day 4 (multiple choice) to an average of 66.8% on Day 5 (free response). Over the course of the training part 2, participants steadily increased their accuracy to reach 70.2% on Day 8. On the final day of testing, 67.3% of incorrect responses differed from the presented word by only one phoneme. Of these, 93.2% were attributable to an incorrect identification of a vowel phoneme, while the rest were the result of incorrect identification of a consonant phoneme.

5 Discussion

For a haptic communication device to be useful and realistic for everyday use, it should be able to convey unconstrained vocabulary, transmit messages at comparable rates to spoken language, not interfere with activities of daily living, and require less than eight hours of training. The MISSIVE can convey an unconstrained vocabulary because it uses phonemes as its building block. By using multi-sensory haptic cues, the device fits on a small portion of the upper-arm, which keeps it form impeding activities of daily living. In 100 minutes of training, users were able to achieve 61.4% accuracy in identifying 23 phonemes and 89.4% accuracy in identifying 50 words using multiple choice. Through a further 60 minutes of training, five participants were able to achieve 85.2% accuracy in identifying phonemes and 70.9% accuracy in identifying words using free response. With this only taking 160 minutes of training, users should be able to significantly improve within a reasonable total of eight hours of training. In this experiment, we did not test

User Responses \ Actual Phoneme	Vowels (38.4%)											Consonants (82.3%)											
	AW	AY	E	EE	IY	O	OH	OO	OW	U	UUH	B	DTH	F	H	K	M	N	R	S	T	W	Y
Vowels: AW	33	10	7	10	3	3	13	7	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0
AY	7	47	13	3	3	7	0	17	3	7	3	0	0	0	0	0	0	0	0	0	0	0	0
E	3	0	17	7	6	3	13	20	3	7	7	0	0	0	3	0	0	0	0	0	0	7	0
EE	13	0	17	47	16	0	3	7	0	7	0	0	0	3	0	0	0	0	0	0	0	3	0
IY	3	17	7	0	38	3	7	3	13	0	3	0	3	10	0	0	0	0	0	0	0	0	0
O	0	0	10	0	3	53	7	3	0	13	10	0	0	0	0	0	0	0	0	0	0	3	3
OH	7	3	10	17	6	3	23	0	13	3	0	0	0	0	0	0	0	0	0	0	0	0	0
OO	7	7	10	3	3	3	10	33	10	7	0	0	3	0	3	0	0	0	0	0	0	0	0
OW	10	3	0	3	13	3	7	0	48	3	7	0	0	0	0	0	0	0	0	0	0	0	0
U	17	0	0	0	0	0	3	0	6	40	23	0	0	0	0	0	0	0	0	0	0	0	0
UUH	0	0	3	3	3	10	10	0	0	7	43	0	0	0	3	0	0	0	0	0	0	0	0
Consonants: B	0	0	0	0	0	0	0	0	0	0	0	97	0	0	0	0	0	0	7	0	0	3	0
DTH	0	10	0	0	3	0	3	0	0	0	0	0	67	3	0	0	0	3	0	0	0	0	10
F	0	0	3	3	0	0	0	0	0	0	0	0	10	77	0	3	0	0	0	13	0	7	6
H	0	0	0	0	0	3	0	0	0	0	0	0	0	3	74	0	0	0	0	0	0	10	0
K	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	83	0	0	0	0	3	7	3
M	0	0	0	0	0	0	0	0	0	0	0	3	3	0	0	0	90	0	0	0	0	0	0
N	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	90	0	0	0	0	6
R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	93	0	0	0	0
S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	6	7	0	88	0	0	0
T	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	7	3	0	0	0	97	0	0
W	0	0	3	0	3	7	0	0	0	3	3	0	7	0	13	0	0	0	0	0	0	60	0
Y	0	3	0	3	0	0	0	10	0	0	0	0	7	0	0	0	0	0	0	0	0	0	71

Fig. 5. Confusion matrix showing percent accuracy in identifying phonemes on Day 4 for all with correct answers on the main diagonal. The actual phoneme presented is shown at the top, and the phoneme that the user chose is shown on the left. Thick outlines show the separation between the vowels and consonants.

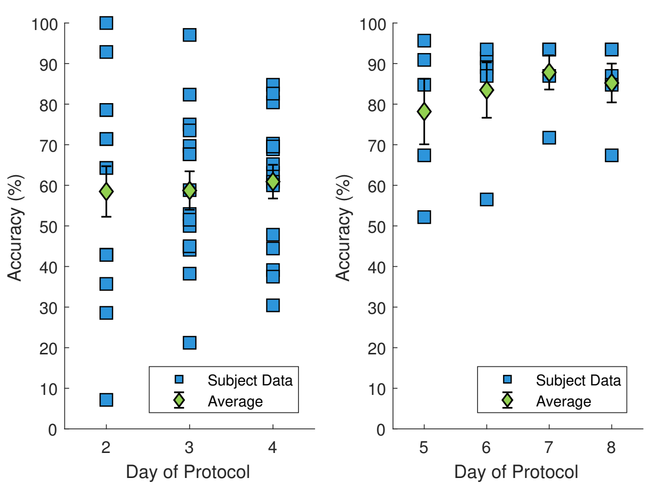

Fig. 6. Pre Test accuracies for all participants shown for each day during training part 1 (left) and training part 2 (right). Averages are shown in green, and error bars represent standard error.

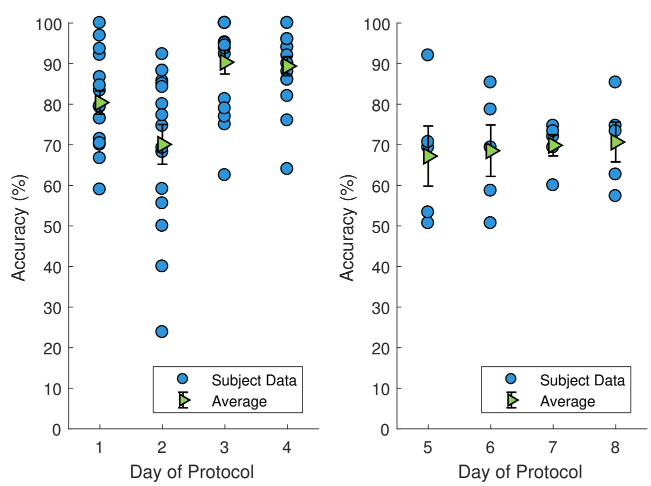

Fig. 7. Post Test accuracies for all participants shown for each day during training part 1 (left) and training part 2 (right). Averages are shown in green, and error bars represent the standard error. The right plot corresponds to the free response answering mechanism versus the multiple choice mechanism in the left plot.

transmission speeds, but this could be observed in future experiments.

The form factor of the MISSIVE is smaller compared to devices of similar capability which use the whole forearm [9], [19], the torso [13], or the valuable space of the hand [7] to display an array of vibrotactors. A recent study has shown subjects can identify 100 words made of 13 phonemes in 65 minutes of training, but subjects had the ability to replay haptic cues for 30 seconds [19]. A similar study showed that subjects were able to achieve above 90% accuracy in identifying 39 phonemes, with most subjects moving on to test their ability to identify 50 and 100 word lists using multiple choice. Of those who tested 100 words, accuracy ranged from 50% to 95% on the last day [33]. All subjects from our study learned to understand 150 words with the MISSIVE and scored comparable accuracies to previous studies in a similar amount of training time, while utilizing a device with a smaller form factor.

5.1 Vowel Identification Accuracy

Participants were noticeably better at identifying consonants (82.3% correct) than vowels (38.4% correct) after four days of training (see Section 4.1). Participants who continued to training part 2 improved their phoneme identification accuracy. Still, for 93.2% of words that were incorrectly identified, the error was attributable to incorrect identification of a vowel. It is clear that vowel identification remained challenging, despite further training. We suspect two factors may be contributing to the difficulty in accurate vowel identification, namely errors in multi-sensory haptic cue perception, and errors in phoneme interpretation.

First, vowel identification performance could be lower than consonant identification performance because vowels were mapped to multi-sensory haptic cues comprised of three components (vibration, stretch, and squeeze), whereas consonants were mapped to two-component cues (vibration and stretch only). Our prior work has identified that multi-sensory cues comprising vibration, stretch, and squeeze can lead to mis-identification of the multi-sensory haptic cue components [23], [34], showing how errors in vowel identification could be attributable to errors in multi-sensory haptic cue perception.

Second, the lower accuracy could be due to the increased difficulty of understanding the representation of vowels as phonemes. Since participants did not have any specific training for translating between sounds and phonemes, participants were not necessarily experienced with interpreting sounds and words in this manner. Vowel phonemes can in general correspond to many different word spellings, and similar letter combinations in different words can correspond to different vowel phoneme. For example, the sounds produced by the phoneme /AY/, as in **A**te, can correspond to the spellings "a", "ai", "ay", "a_e", "ei", "er", "et", "ey", or "ea", and the phoneme /U/, as in **U**p, can correspond to spellings with "a", "o", "oo", "ou", or "u". In contrast, most consonant phonemes have a direct one-to-one mapping to a letter of the alphabet, and generally have a small number of letter combinations to which they correspond. For example, the sound produced by the phoneme /b/ can only correspond to a spelling using the letters "b" or "bb", and the phoneme /p/ can only correspond to a spelling using the letters "p" or "pp". Because phonemes were presented one at a time in sequence to the participants, and because most people are used to thinking of words in terms of spelling rather than phonemes, participants tended to relate the phonemes that they were feeling with sequences of letters (as if spelling out words), instead of sequences of sounds, resulting in vowel identification errors.

5.2 Training Effects

With only 100 minutes of training, phoneme identification accuracies reached an average of only about 60%. For those participants who continued to training part 2, performance improved by 9.7%, from 75.5% to 85.2% during training part 2, and importantly, the accuracy in vowel identification improved from 54.1% to 73.6%. This increase shows that while the vowels were indeed more difficult to perceive compared to consonants, participants were able to improve their perception of vowels through additional training. Still, 73.6% accuracy in identifying vowels is lower than we would like to see for effective haptic transmission of words. Further training should likely show higher accuracies, and future research should determine the amount of training necessary to achieve consistent and high accuracy word identification using the MISSIVE.

In our training, we did not test with full sentences. It will be an important next step to combine strings of words into sentences to see what effects this has on comprehension. A sequence of words will need to be separated by a pause or some signal representing a space between words. While we did not test this, one study successfully used a gap between words of 270 ms when displaying two words in a sequence using letters (spaced as little as 100 ms apart) on the hand [7]. While this number would likely vary with phonemes on the MISSIVE, it provides some guidance as to how quickly people can receive and comprehend strings of haptic cues.

5.3 Word Identification Accuracy

Despite participants' relatively low accuracy in identifying phonemes (61.4% on Day 4), word identification accuracy was quite high, always greater than 70% during training days. For the 50 word un-timed final test on Day 4, participants were able to correctly identify 89.4% of the presented words. Still, in subjective responses on surveys, participants indicated that they continued to have trouble with the vowels, but were confident in their consonant responses. It is likely that participants relied on the multiple choice options to identify the word that matched the set of consonants that they understood from the cue presentations, making up for their somewhat poorer performance at identifying vowels. This subjective feedback supports the conclusion that participants will likely need longer than 100 minutes of training to become proficient in identifying arbitrary words with the MISSIVE, and that training should involve free responses instead of multiple choice responses that can lead to higher performance even without full command over the entire set of trained phonemes.

Word identification accuracy during post-tests decreased from Day 4 to Day 5 when participants switched from multiple choice to free response, from an average of 93.6% accuracy on Day 4 using multiple choice to 67.2% accuracy on Day 5 using free response. Without the multiple choices to rely on, participants had to correctly identify each phoneme in the presented words, and correctly combine them into the correct word. Participants improved each day to a final word identification accuracy of 70.7% on Day 8, but this was not notably different from Day 5. Interestingly, even though on word identification accuracy did not improve, participants improved their phoneme identification accuracy, specifically in vowel identification. It is clear that having multiple choice answers displayed to participants reduced most mistakes that participants made when compared to the free response portion of assessment.

Recall that 67.3% of mistakes in word identification were attributable to the mis-identification of only one phoneme. This implies that given some context or constraints on a given word being transmitted, users may be able to still correctly identify words they are unsure of.

5.4 Phonemes as a Building Block

In our training protocol, we did not include time to teach participants the concept of phonemes or how they are combined to make words. It seems that this led to participants having trouble correctly connecting phonemes to make words, and common error patterns were observed in free-response word identification assessments. Here, participants seemed to confuse the sound of the phonemes with the letters that were used to represent them. For example, on Day 8 of the protocol, when the word *phone* was presented, four of the five participants responded with a word that began with the letter "f". This is telling—even with initial training on the method of combining phonemes into words, participants still seemed to default to letter or spelling-based representations of words. The dialects used and the ways that participants interpret words can also have an impact on word identification success. All pronunciations were taken from the CMUdict database, and some words had pronunciations that differed compared to the participants' common usage. Therefore, some of the the 67.2% of word identification mistakes attributable to one incorrect phoneme identification could be due to misinterpretations of the phonemes, or how certain participants tend to pronounce words. For example, on the last day of training part 2, every participant responded with the word "were" to the phoneme representation /w/ /e/ /r/, instead of the word based on the the CMUdict database which was "where". This underlying confusion highlights a potential drawback of using phonemes as building blocks, instead of other options such as letters. While there exist some discrepancies in spellings between different regions and dialects, there are far fewer of these than pronunciation differences, which can differ wildly for people speaking the same language. Future work should investigate the trade-offs between the use of phonemes or letters as building blocks for haptic communication. Phonemes offer a higher density of information per cue, but spelling out words with letters may prove more intuitive for users. It is also worth noting that subjects learned a subset of 23 of 39 possible phonemes. Further testing would have to be performed to see the effects of learning all 39 phonemes.

6 Conclusions

Haptics offers a novel mechanism for communication, and new compact actuators and wearable sensors offer the potential for wearable, private communication devices. In this paper, we presented design criteria that should be considered when developing a device and system for haptic transmission of the English language. These criteria emphasize the need to convey unconstrained language at reasonable speeds, while maintaining a desirable form factor and learnability.

We mapped phonemes in the English language to 350 ms multi-sensory haptic cues comprised of vibration, skin stretch, and squeeze, and were displayed to users via the MISSIVE device. After four days and 100 minutes of training, participants were able to identify phonemes with 61.4% accuracy and words with 89.4% accuracy using multiple choice. Five participants continued through another 60 minutes of training and were able to identify phonemes with 85.2% accuracy and scored an average of 70.2% accuracy in identifying words using free response. Analysis of the phoneme responses showed that participants were less adept at identifying vowel phonemes compared to the consonant phonemes, and that participants likely had trouble combining several phonemes into words. Still, the demonstrated accuracy in identifying phonemes and words encourages further exploration into the use of multi-sensory haptic devices for transmission of words to users.

References

[1] C. M. Reed, W. M. Rabinowitz, N. I. Durlach, L. D. Braida, S. Conway-Fithian, and M. C. Schultz, "Research on the tadoma method of speech communication," *The Journal of the Acoustical Society of America*, vol. 77, no. 1, pp. 247–257, 1985.

[2] C. M. Reed, "The implications of the tadoma method of speechreading for spoken language processing," in *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP '96*, vol. 3, Oct 1996, pp. 1489–1492.

[3] A. Israr, P. Meckl, C. Reed, and H. Tan, "Controller design and consonantal contrast coding using a multi-finger tactual display," *The Journal of the Acoustical Society of America*, vol. 125, pp. 3925–35, 2009.

[4] H. Z. Tan, N. I. Durlach, W. M. Rabinowitz, C. M. Reed, and J. R. Santos, "Reception of Morse code through motional, vibrotactile, and auditory stimulation," *Perception & Psychophysics*, vol. 59, no. 7, pp. 1004–1017, 1997.

[5] V. Lévesque, J. Pasquero, V. Hayward, and M. Legault, "Display of virtual braille dots by lateral skin deformation: feasibility study," *ACM Trans. on Applied Perception*, vol. 2, no. 2, pp. 132–149, 2005.

[6] S. Zhao, A. Israr, F. Lau, and F. Abnousi, "Coding Tactile Symbols for Phonemic Communication," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, Montreal, QC, 2018, pp. 1–13.

[7] G. Luzhnica, E. Veas, and V. Pammer, "Skin Reading: Encoding Text in a 6-Channel Haptic Display," in *Proceedings of the International Symposium on Wearable Computers*, 2016, pp. 148–155.

[8] S. D. Novich and D. M. Eagleman, "Using space and time to encode vibrotactile information: toward an estimate of the skin's achievable throughput," *Experimental Brain Research*, vol. 233, no. 10, pp. 2777–2788, 2015.

[9] J. Jung, Y. Jiao, F. M. Severgnini, H. Z. Tan, C. M. Reed, A. Israr, F. Lau, and F. Abnousi, "Speech Communication Through the Skin: Design of Learning Protocols and Initial Findings," in *Design, User Experience, and Usability: Designing Interactions. DUXU 2018. Lecture Notes in Computer Science*, 2018, pp. 447–460.

[10] H. Z. Tan, C. M. Reed, and N. I. Durlach, "Optimum Information-Transfer Rates for Communication through Haptic and Other Sensor Modalities," *IEEE Transactions on Haptics*, vol. 3, no. 2, pp. 98–108, 2010.

[11] R. H. Gault and G. W. Crane, "Tactual patterns from certain vowel qualities instrumentally communicated from a speaker to a subject's fingers," *The Journal of General Psychology*, vol. 1, no. 2, pp. 353–359, 1928.

[12] A. Israr, P. H. Meckl, and H. Z. Tan, "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)*, 2004, pp. 1083–1089.

[13] S. D. Novich, "Sound-to-touch sensory substitution and beyond," Ph.D. dissertation, Rice University, 2015.

[14] C. M. Reed, N. I. Durlach, and L. D. Braida, "Research on tactile communication of speech: a review." *ASHA monographs*, no. 20, p. 1, 1982.

[15] S. Engelmann and R. Rosov, "Tactual hearing experiment with deaf and hearing subjects," *Exceptional Children*, vol. 41, no. 4, pp. 243–253, 1975.

[16] A. Bizzocchi, "How many phonemes does the english language have?" *International Journal on Studies in English Language and Literature (IJSELL)*, vol. 5, pp. 36–46, 10 2017.
[17] F. A. Geldard, "Adventures in tactile literacy," *American Psychologist*, vol. 12, no. 3, pp. 115–124, 1957.
[18] G. Luzhnica, E. Veas, and C. Seim, "Passive haptic learning for vibrotactile skin reading," 10 2018.
[19] J. Chen, R. Turcott, P. Castillo, W. Setiawan, F. Lau, and A. Israr, "Learning to feel words: A comparison of learning approaches to acquire haptic words," in *Proceedings of the 15th ACM Symposium on Applied Perception*, ser. SAP '18. New York, NY, USA: ACM, 2018, pp. 11:1–11:7.
[20] D. Bigham, "The pin˜ pen vowel merger in southern illinois english," *NWAV, University of Michigan, Ann Arbor, Michigan*, p. 33, 2004.
[21] C. M. Reed, H. Z. Tan, Z. D. Perez, E. C. Wilson, F. M. Severgnini, J. Jung, J. S. Martinez, Y. Jiao, A. Israr, F. Lau, K. Klumb, R. Turcott, and F. Abnousi, "A phonemic-based tactile display for speech communication," *IEEE Transactions on Haptics*, pp. 1–1, 2018.
[22] E. Y. Wong, A. Israr, and M. K. O'Malley, "Discrimination of consonant articulation location by tactile stimulation of the forearm," in *2010 IEEE Haptics Symp.*, 2010, pp. 47–54.
[23] N. Dunkelberger, J. Bradley, J. L. Sullivan, A. Israr, F. Lau, K. Klumb, F. Abnousi, and M. K. O'Malley, "Improving Perception Accuracy with Multi-sensory Haptic Cue Delivery," in *Haptics: Science, Technology, and Applications. EuroHaptics 2018. Lecture Notes in Computer Science*, vol. 10894 LNCS, 2018, pp. 289–301.
[24] M. A. Baumann, K. E. MacLean, T. W. Hazelton, and A. McKay, "Emulating human attention-getting practices with wearable haptics," in *IEEE Haptics Symposium*, 2010, pp. 149–156.
[25] S. Casini, M. Morvidoni, M. Bianchi, M. Catalano, G. Grioli, and A. Bicchi, "Design and realization of the CUFF - Clenching upper-limb force feedback wearable device for distributed mechano-tactile stimulation of normal and tangential skin forces," in *IEEE International Conference on Intelligent Robots and Systems*, vol. 2015-Decem, 2015, pp. 1186–1193.
[26] L. Meli, I. Hussain, M. Aurilio, M. Malvezzi, M. O'Malley, and D. Prattichizzo, "The hBracelet: A Wearable Haptic Device for the Distributed Mechanotactile Stimulation of the Upper Limb," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 1–1, 2018.
[27] M. Aggravi, F. Pause, P. R. Giordano, and C. Pacchierotti, "Design and Evaluation of a Wearable Haptic Device for Skin Stretch, Pressure, and Vibrotactile Stimuli," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 2166–2173, jul 2018.
[28] C. D. Fryar, Q. Gu, C. L. Ogden, and K. M. Flegal, "Anthropometric reference data for children and adults; united states, 2011-2014," *Vital and Health Statistics*, 2016.
[29] E. Battaglia, J. P. Clark, M. Bianchi, M. G. Catalano, A. Bicchi, and M. K. O'Malley, "The Rice Haptic Rocker: Skin stretch haptic feedback with the Pisa/IIT SoftHand," in *2017 IEEE World Haptics Conference, WHC 2017*, 2017, pp. 7–12.
[30] E. Battaglia, J. Clark, M. Bianchi, M. Catalano, A. Bicchi, and M. K. O'Malley, "Skin stretch haptic feedback to convey closure information in anthropomorphic, under-actuated upper limb soft prostheses," *IEEE Transactions on Haptics*, pp. 1–1, 2019.
[31] E. M. Loiola, N. M. M. de Abreu, P. O. Boaventura-Netto, P. Hahn, and T. Querido, "A survey for the quadratic assignment problem," *European Journal of Operational Research*, vol. 176, no. 2, pp. 657–690, 2007.
[32] Z. Drezner, "A new genetic algorithm for the quadratic assignment problem," *INFORMS Journal on Computing*, vol. 15, no. 3, pp. 320 – 330, 2003.
[33] Y. Jiao, F. M. Severgnini, J. S. Martinez, J. Jung, H. Z. Tan, C. M. Reed, E. C. Wilson, F. Lau, A. Israr, R. Turcott, K. Klumb, and F. Abnousi, "A comparative study of phoneme- and word-based learning of english words presented to the skin," in *Haptics: Science, Technology, and Applications*, D. Prattichizzo, H. Shinoda, H. Z. Tan, E. Ruffaldi, and A. Frisoli, Eds. Cham: Springer International Publishing, 2018, pp. 623–635.
[34] J. L. Sullivan, N. Dunkelberger, J. Bradley, J. Young, A. Israr, F. Lau, K. Klumb, F. Abnousi, and M. K. O'Malley, "Multi-sensory stimuli improve distinguishability of cutaneous haptic cues," *IEEE Transactions on Haptics*, pp. 1–1, 2019.

Nathan Dunkelberger holds a B.S. from Texas A&M University in College Station, TX, USA and a M.Sc. from Rice University, Houston, TX, USA, both in mechanical engineering. He is currently pursuing a Ph.D. in mechanical engineering from Rice University as a member of the Mechatronics and Haptic Interfaces Lab.

Jennifer L. Sullivan holds a B.S. in mechanical engineering from Rice University in Houston, TX, USA, and a M.A.Sc. in mechanical engineering from the University of British Columbia in Vancouver, BC, Canada. She is currently a Human Performance Engineer at NASA Johnson Space Center.

Joshua Bradley holds a B.S. in applied mathematics from the United States Air Force Academy in Colorado Springs, CO, USA, and a M.S. in mechanical engineering from Rice University in Houston, TX, USA.

Indu Manickam holds B.S.E.s in biomedical engineering and electrical and computer engineering from Duke University in Durham, NC, USA, and a M.Sc. in electrical and computer engineering from Rice University in Houston, TX, USA.

Gautam Dasarathy received the B.Tech degree in Electronics and Communication Engineering from Vellore Institute of Technology, Vellore, India, in 2008, and the M.S. and Ph.D degrees in electrical engineering from the University of Wisconsin, Madison, WI, USA in 2010 and 2014 respectively. He is currently an Assistant Professor for the school of Electrical Computer, and Energy Engineering with Arizona State University, Tempe, AZ, USA.

Richard G. Baraniuk is the Victor E. Cameron Professor of Electrical and Computer Engineering at Rice University and the Founding Director of OpenStax. His research interests lie in new theory, algorithms, and hardware for sensing, signal processing, and machine learning. He is a Fellow of the American Academy of Arts and Sciences, National Academy of Inventors, American Association for the Advancement of Science, and IEEE. He has received the DOD Vannevar Bush Faculty Fellow Award (National Security Science and Engineering Faculty Fellow), the IEEE Signal Processing Society Technical Achievement Award, and the IEEE James H. Mulligan, Jr. Education Medal, among others.

Marcia K. O'Malley (F '20) received the B.S. degree from Purdue University, West Lafayette, IN, USA, in 1996, and the M.S. and Ph.D. degrees from Vanderbilt University, Nashville, TN, USA, in 1999 and 2001, respectively, all in mechanical engineering. She is currently the Thomas Michael Panos Family Professor of mechanical engineering, of computer science, and of electrical and computer engineering with Rice University, Houston, TX, USA, and directs the Mechatronics and Haptic Interfaces Laboratory.