Identification of Words and Phrases Through a Phonemic-Based Haptic Display: Effects of Inter-Phoneme and Inter-Word Interval Durations

CHARLOTTE M. REED, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA, USA
HONG Z. TAN and YANG JIAO, Haptic Interface Research Lab, School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, USA
ZACHARY D. PEREZ and E. COURTENAY WILSON, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA, USA

13

Stand-alone devices for tactile speech reception serve a need as communication aids for persons with profound sensory impairments as well as in applications such as human-computer interfaces and remote communication when the normal auditory and visual channels are compromised or overloaded. The current research is concerned with perceptual evaluations of a phoneme-based tactile speech communication device in which a unique tactile code was assigned to each of the 24 consonants and 15 vowels of English. The tactile phonemic display was conveyed through an array of 24 tactors that stimulated the dorsal and ventral surfaces of the forearm. Experiments examined the recognition of individual words as a function of the inter-phoneme interval (Study 1) and two-word phrases as a function of the inter-word interval (Study 2). Following an average training period of 4.3 hrs on phoneme and word recognition tasks, mean scores for the recognition of individual words in Study 1 ranged from 87.7% correct to 74.3% correct as the inter-phoneme interval decreased from 300 to 0 ms. In Study 2, following an average of 2.5 hours of training on the two-word phrase task, both words in the phrase were identified with an accuracy of 75% correct using an inter-word interval of 1 sec and an inter-phoneme interval of 150 ms. Effective transmission rates achieved on this task were estimated to be on the order of 30 to 35 words/min.

CCS Concepts: • **Human-centered computing → Haptic devices**;

Additional Key Words and Phrases: Human haptics, speech communication, phoneme codes, tactile devices

ACM Reference format:
Charlotte M. Reed, Hong Z. Tan, Yang Jiao, Zachary D. Perez, and E. Courtenay Wilson. 2021. Identification of Words and Phrases Through a Phonemic-Based Haptic Display: Effects of Inter-Phoneme and Inter-Word Interval Durations. *ACM Trans. Appl. Percept.* 18, 3, Article 13 (July 2021), 22 pages.
https://doi.org/10.1145/3458725

This work was partially supported by a research grant funded by Facebook Inc., Menlo Park, CA, and by support from the National Science Foundation through NSF award numbers 1954842 and 1954886.
Authors' addresses: C. M. Reed, Room 36-751, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA; email: cmreed@mit.edu; H. Z. Tan, Haptic Interface Research Lab, School of Electrical and Computer Engineering, Purdue University, 465 Northwestern Avenue, West Lafayette, IN 47907, USA; email: hongtan@purdue.edu; Y. Jiao, The Future Lab, Tsinghua Universty, No. 160 Cheng Fu Road, Haidian District, Beijing, P. R. China 100086; email: jymars@live.cn; Z. D. Perez, Room 36-751, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA; email: zdperez@gmail.com; E. C. Wilson, Room 36-751, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA; email: ecwilson828@yahoo.com.

1544-3558/2021/07-ART13
https://doi.org/10.1145/3458725

1 INTRODUCTION

1.1 Overview of Tactile Speech Communication

The information capacity of the tactual sensory system has been demonstrated through studies of experienced deaf-blind users of several natural methods of tactual communication, in which the tactual stimulation of the deaf-blind "receiver" is achieved through direct physical contact with the "sender." These methods are referred to as natural in that they do not employ any devices for tactual stimulation. Among these are the Tadoma method of speechreading [1–3], as well as the tactual reception of fingerspelling and sign language [4–6]. Experienced users of the Tadoma method are able to understand speech by placing a hand over the face and neck of a talker to monitor a variety of cues that can be felt during the production of speech (such as airflow, lip and jaw movements, and laryngeal vibration). This information is sufficient to support the reception of key words in conversational sentences produced at slow-to-normal speaking rates at a level of roughly 80% correct. Other natural methods of communication that have evolved within the deaf-blind community also offer powerful demonstrations of the information-bearing capacity of the tactual sense, albeit for the reception of signals other than speech, including the tactual reception of fingerspelling and sign language [4–6]. In place of the visual observation of fingerspelled letters and signs as used by sighted deaf persons, these methods have been adapted for tactual reception by persons who are both deaf and blind: The hands of the deaf-blind person are placed over those of the fingerspeller or signer to feel the formational properties of these signals (such as handshape, movement, orientation, and location). Experienced deaf-blind users of each of these two methods were also tested on their ability to receive key words in conversational sentences. Scores for reception of tactual fingerspelling by experienced deaf-blind users were in the range of 85%–100% correct [4] and for tactual sign language, 60%–90% correct [5]. The information rates achieved with Tadoma and the tactual reception of sign language (12–14 bits/sec) are higher than those for tactual reception of fingerspelling (7.5 bits/sec), but are roughly one-half of those achieved for the auditory reception of speech and the visual reception of sign language [7].

In addition to these natural methods of tactual communication, there is a long history of research on the development of tactile devices for speech communication. Promising results have been reported for the reception of isolated words through spectral-based aids using either vibrotactile [8–10] or electrotactile [11, 12] stimulation. Even after extensive periods of training and use, however, tactile aids have functioned primarily as supplements to the visual information available on the face through lipreading for the reception of connected speech signals [13–20]. Many of these previous tactile devices have relied on the use of spectral-based displays of the acoustic speech spectrum, which suffers from drawbacks of inadequate resolution, temporal masking, and token variability. In addition, these displays have been relatively impoverished in their use of homogeneous stimulation of the cutaneous system compared to the richness of the "talking face" that is felt in Tadoma [21].

With recent advances in the areas of automatic speech recognition—ASR [22], haptic technologies [23], and learning theory [24], the opportunity now exists to revisit the challenge of developing a tactile device that will enable its users to receive speech through the tactual sense alone. Such devices have a broad range of applications, not only as communication aids for persons with profound visual and/or auditory impairments, but also in situations for persons with normal sensory abilities when hearing and/or sight are compromised or overloaded (such as for human-computer interfaces and remote communication).

Recent work has employed a phonemic-based strategy for a tactile display of speech, under the assumption that phonemic strings can be derived from ASR at the front end of the system [25–35]. This approach allows for the development of tactile displays that are no longer tied to displays of the acoustic speech spectrum, but instead allow for the design of displays that exploit the information-bearing properties of the tactual sense.

The current study is concerned with extending our previous research on the use of a phonemic-based haptic display for word recognition [25, 27, 29–31]. We present below a brief review of results obtained in previous studies concerned with the recognition of words through tactile-alone presentation for both spectral-based and phonemic-based tactile displays.

1.2 Relevant Background Studies

1.2.1 Spectral Displays. Studies of tactile-alone word recognition using spectral-based displays have employed either vibrotactile [8, 9, 10, 36] or electrotactile [11, 12] stimulation. Following training times on the order of 40 to 80 hours using a 16-channel vibrotactile vocoder applied to the forearm [9, 10], users of the display acquired vocabularies of 70 to 250 words. One normal-hearing participant in [9] achieved 76% correct identification of words from a 250-word vocabulary for tactile-alone presentation following 80 hrs of training. More recently, Novich (2015) [36] used a spectral-based approach to display speech signals through an array of 27 tactors arranged in a vest. Participants were trained on a word-recognition task in a 4-alternative response format (i.e., a chance level of performance of 25%) with a vocabulary of 50 words. Following 12 days of training with 300 trials per day, the mean word recognition rate was 35%–65% correct across participants. Using an electrotactile display applied to the fingers, Galvin et al. (1999) [12] demonstrated acquisition of a tactile-alone vocabulary that averaged 31 words following 12 hrs of training, and a vocabulary averaging 50 words with criterion performance of 80% correct was acquired following a mean of 18.6 hrs training. These results compare favorably to those reported for normal-hearing learners of the Tadoma method [37].

1.2.2 Alphabetic Displays. Another approach to tactile communication is through the use of alphabetic-based displays. Luzhnica and colleagues [38–40] coded the 26 letters of the English alphabet for display through a tactile glove consisting of six tactors. Following 5 hrs of training, the letter recognition rate was ≥90% correct on a set of 48 words transmitted at a mean duration of 0.6 sec/word [38]. Following modification of their alphabetic display to a 7-tactor layout on the back of the hand, Luzhnica and Veas (2019) [40] demonstrated improved letter recognition of 98% following five training sessions. Participants were also tested on a word-recognition task for which accuracy was reported in terms of a Levenshtein distance, averaging 0.97 in the final session. Retention of the isolated alphabetic codes after gaps of 10 or more days was high with scores averaging 94% correct. Furthermore, Luzhnica and Veas (2019) [39] demonstrated that participants could accurately identify a set of 10 tactile alphabetic codes in the presence of a competing primary task performed in the visual modality.

1.2.3 Phonemic-based Displays. Most relevant to the current study is previous work concerned with the development and testing of phonemic-based tactile codes. Zhao et al. (2018) [26] employed phonemic coding of nine tactile symbols (five consonants and four vowels) using a 2×3 array of tactors applied to the dorsal forearm. After training on a phoneme-identification task, participants then learned to recognize words from a 20-word vocabulary. Performance was 83% correct for the better of two phoneme mapping strategies. Turcott et al. (2018) [28] compared haptic word recognition for a spectral-based processing strategy (based on the display of dominant spectral peaks) to that obtained with a phoneme-based display. The spectral display was presented through a 32-tactor array applied to the upper arms and forearms of both sides of the body (with 8 tactors per site). The phonemic display was used to encode 10 phonemes through a16-tactor array applied to the dorsal and ventral surfaces of the left forearm. Following 50 minutes of training, participants were tested on two different lists of 10 words composed of the 10 phonemes. Across the two lists (where chance performance on each list was 10%), mean scores were 76.3% correct for phonemic encoding compared to 44.4% correct for the spectral-based display. In further work with the 32-tactor array used to display phonemic codes of 13 phonemes, Chen et al. (2018) [33] compared two different approaches to learning. One group of learners followed a fixed schedule of training, while the other group had control over their training activities. Following 65 minutes of training on identification of phonemes and words, participants were tested on word identification using a 12-alternative forced-choice response paradigm for words selected from a 100-word vocabulary. Mean performance was 86% for the group with fixed-schedule training and 72% correct for self-guided training, although significantly more participants from the former group were able to achieve performance >90% correct.

Dunkelberger et al. (2018) [34] developed tactile codes for 23 English phonemes using a display attached to the upper forearm that consisted of four tactors, a radial squeeze band, and a lateral skin stretch rocker. After

100 min of training on identification of the 23 individual phonemes and of a set of 150 words composed of these phonemes, participants were tested using a subset of 50 words selected from this vocabulary. Phonemes were presented at a self-paced rate (with an average phoneme presentation rate of 3.5 sec/phoneme) and participants selected responses from a closed set of the 50 stimulus words. Performance averaged 87% correct with a mean response time of 7.7 sec/word. Further results on this system were reported by Dunkelberger et al. (2021) [32] for a larger group of participants who received 100 mins of training on the identification of 23 isolated haptic phonemes and a set of 150 words formed from these phonemes. Mean scores were 61.4% correct for phoneme identification and 89.9% correct for identification of words in a 12-alternative forced-choice procedure. Fontana de Vargas et al. (2019) [35] designed tactile codes for 15 English consonants (based on salient features derived from acoustic waveforms) and 9 vowels and diphthongs (based on synthesized speech signals). The tactile signals were presented through two vibrotactile channels on the ventral surface of the forearm, in which level differences between the two vibrators were used to create illusions of location on the forearm. Following training on phoneme and word identification, participants achieved a score of 94% correct on a word-identification test using a 12-alternative forced-choice response from a 150-word stimulus set with an inter-phoneme interval of 1 sec. Further testing conducted with an open-set response paradigm yielded scores of 51% correct for words drawn from the 150-word training set and 39% correct for novel words.

1.3 Goals of Current Study

The current study is a continuation of work in which phonemes were encoded into haptic symbols for presentation through a device referred to as the **Tactile Phonemic Sleeve (TAPS)**, which consists of a 4×6 array of tactors applied to the dorsal and volar surfaces of the forearm. The 39 phonemes of English (24 consonants and 15 vowels) were encoded using multiple dimensions that included frequency of vibration, duration, and place of stimulation, selection of only a few possible values per dimension (for example, frequency of either 60 or 300 Hz), and the use of illusory movements to enhance the number of distinctive signals (see References [29, 41, 42]). Within two to four hours of training across participants, the 39 phonemes could be identified with an accuracy of 86% correct [29]. In a preliminary evaluation of the ability to recognize words composed of sequences of phonemes, users of the display were able to identify 10 phonemes and a set of 50 words derived from them with nearly 100% accuracy after one hour of training [25]. Word-identification tests were also conducted using the full set of 39 phonemes and a 100-word vocabulary [27]. Following 100 minutes of training, the best learners of the display were able to recognize words with an accuracy of 90% or greater (where chance performance was 1/100). Further study of word recognition through the TAPS device was reported by Tan et al. (2020) [31]. These experiments included 51 participants who advanced to testing with a 500-word vocabulary. The most proficient of these learners could recognize words with an accuracy of 65%–83% correct within 4.5 to 7 hrs of training, with an average acquisition rate of 1.3 words/min.

These results represent an advancement in the field of tactile speech communication devices, both in terms of the performance that was achieved on word recognition solely through tactual cues and in terms of the relatively short learning times. Performance must also be evaluated in terms of the communication rates that can be achieved with a particular tactile display. For the normal auditory reception of speech, two-way communication rates are on the order of 160–200 words/min [43]. These rates drop to roughly 60–80 words/min for the Tadoma method [3], which is comparable to the rates produced for slow conversational speech [44]. By comparison, communication rates obtained thus far with the phonemic-based haptic system [25, 27, 29] are estimated to be on the order of roughly 30 words/min.

2 STUDY 1: EFFECT OF INTER-PHONEME INTERVAL ON IDENTIFICATION OF WORDS

2.1 Overview

The goal of this experiment was to determine the effect of the **inter-phoneme interval (IPI)** on the ability to identify words composed of 2 or 3 haptic phonemic symbols. The participants were first trained to identify the

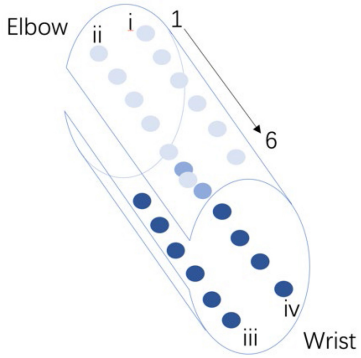

Fig. 1. Schematic illustration of layout of tactors on the forearm: four rows (i–iv) in the longitudinal direction (elbow to wrist) and six columns in the transversal direction (around the forearm).

individual haptic symbols used to code 39 English phonemes (24 consonants and 15 vowels). After training on the full set of phonemes was completed, participants then proceeded to learn to identify a set of 100 words constructed from the haptic codes. Four values of IPI were examined: 300, 150, 75, and 0 ms. The longest duration of 300 ms was chosen because it exceeds the major effects of forward and backward masking for tactile stimulation (e.g., References [45, 46]), 0 ms was selected as the minimum IPI, and two additional durations of 150 and 75 ms were selected between these two endpoints. Participants received practice on the word-identification task with each of the four values of IPI in decreasing order. Following this learning period, participants were tested on word identification with the IPIs presented in random order.

2.2 Methods

2.2.1 Apparatus. The tactile device (referred to as the **Tactile Phonemic Sleeve—TAPS**) consists of a 4-by-6 array of tactors (Tectonic elements, Model TEAX13C02-8/RH, Part #297-14) worn on the forearm, with four rows in the longitudinal direction (elbow to wrist) and six columns in the transversal direction (around the forearm). As shown in the schematic depicted in Figure 1 (taken from Figure 1 of Reed et al., 2019 [29]), two rows reside on the dorsal surface (light blue dots) and two on the volar surface (dark blue dots) of the forearm, respectively. The tactors are wide-bandwidth audio speakers with a constant impedance of ≈8Ω in the frequency region of 50 to 2 kHz, except for a peak in the vicinity of 600 Hz. Each of the 24 tactors is stimulated independently. Stimulus waveforms are generated in Matlab using a multi-channel *playrec* utility (online resource by Robert Humphrey, UK: http://www.playrec.co.uk) running on a desktop computer. A MOTU USB audio device (MOTU, model 24Ao) receives the 24-channel signal via the computer's USB port, performs synchronous D/A conversion of the signals, and sends the 24 waveforms via its 24 channels of analog output connectors to custom-built amplifier boards that among them carry 24 class D stereo amplifiers (Maxim, Model MAX98306) to drive each tactor independently.

2.2.2 Haptic Stimuli. A set of 39 tactile signals was created to correspond to each of the 24 consonants and 15 vowels of English [29]. The dimensions used to create the haptic codes included: frequency (60 and 300 Hz), duration (100 ms for short-duration and 400 ms for long-duration consonants; 240 ms for short-duration and 480 ms for long-duration vowels), place of stimulation (wrist, mid-forearm, and elbow; dorsal and ventral), waveform (e.g., modulated or unmodulated), and the use of different types of movement patterns for vowels (e.g., saltatory versus smooth apparent motion). Articulatory properties of speech sounds (such as voicing, manner, and place of articulation) were also used to guide the mapping of phonemes to tactile codes (e.g., modulated

versus unmodulated sinewaves were used to code voiced versus unvoiced phonemes, and sounds made at the front or back of the mouth were coded at the wrist or elbow, respectively). For example, the phoneme /p/ is coded as a smooth 300 Hz vibration at the dorsal wrist, and the phoneme /b/ is a 300 Hz vibration at the same location with amplitude modulation at 30 Hz to create a heavier feeling for voicing. The phoneme pair /k/ and /g/ are coded at the dorsal elbow with the same 300 Hz vibration without or with a 30 Hz modulation, respectively, to differentiate between the two. The four phonemes were recognized by their location and the presence of amplitude modulation or the lack thereof. In contrast, the haptic codes for vowels always moved on the arm. The vowel /i/ moved from dorsal wrist to dorsal elbow, and the shorter /I/ moved from dorsal elbow to the middle of the dorsal forearm. This way, the two vowels could be differentiated by their directions and their spatial lengths coded the long vs. short durations.

The tactile phonemic codes used in the current study were as those described in Tables 1 and 2 and Figures 2 and 3 of Reference [29], with the following exceptions: The duration of the six plosive phonemes (P (/p/), B (/b/), T, (/t/) D (/d/), K (k), and G (/g/) in Table 1 and Figure 2) was increased from 100 to 140 ms; and the durations of the 11 vowel and diphthong stimuli that were previously 480 ms were decreased to 400 ms. Videos visualizing the tactile stimuli for consonants can be found at https://youtu.be/Fr0-XucKGEY and for vowels at https://youtu.be/CYfqcdnvMyE.

2.2.3 Participants. Seven of the ten participants in the haptic phoneme identification study of Reed et al. (2019) [29] took part in the IPI study (P1, P2, P5, P7, P8, P9, and P10). The participants ranged in age from 19 to 32 years (mean of 23.0 yrs with standard deviation of 4.5 yrs) and included four females and three males. Five of the participants reported right-hand and two left-hand dominance. Of the seven participants, six had clinically normal hearing and one (P8) had a severe sensorineural hearing loss. None of the participants reported any history of problems with the sense of touch. Five were native speakers of English, one a native speaker of Romanian with English acquired at 10 yrs of age, and one a native speaker of Korean with English acquired at 5 yrs of age. All participants provided informed consent through a protocol approved by the **Internal Review Board (IRB)** of MIT and were paid for their time.

2.2.4 Procedure. The haptic device was fitted to the left forearm of each participant, over a Spandex sleeve that was used for hygienic purposes, such that each of the four rows of six tactors was spaced between the wrist and elbow with two rows on the volar and two on the dorsal surface. The tactors were roughly 30 mm in diameter and 9 mm in thickness, with a center-to-center spacing in the longitudinal direction of roughly 35 mm (for female arms) or 40 mm (male), and in the transverse direction of 50 mm (females) or 57 mm (males). The level at which the signals were presented was determined for each participant using a two-step process described in detail by Reference [29]. In short, the 300 Hz threshold was measured at a reference tactor in the middle of the array. The threshold and all subsequent measurements were specified in dB relative to the maximum output of the system. The perceived intensity of each of the other 23 tactors was then equalized to that of the reference tactor (presented at 35 dB above threshold) using a method of adjustment procedure. The tactile phonemic codes used in the experiments were presented at a level of approximately 25 dB sensation level for each participant. To mask any auditory sensations arising from the device, participants wore a pair of acoustic-noise-cancelling headphones (Bose QuietComfort 25) over which a pink masking noise was presented.

Testing was conducted in a sound-treated booth that contained the haptic device as well as a computer monitor, keyboard, and mouse that were connected to a desktop computer located outside the booth. Software developed in Matlab was used to generate the stimuli delivered to the tactile array, control the experiments, and record responses. A one-interval, forced-choice identification paradigm was employed in training and testing on the word-identification task. On each trial, one word in the set was selected for presentation at random with replacement from a set of 100 words. Each word was defined phonetically as a sequence of

haptic symbols that were presented through the haptic device. The participant was given unlimited time in which to respond by selecting one of the 100 words from the list, which was made available on a computer screen. Training runs were collected with the use of trial-by-trial correct-answer feedback, which was eliminated in testing runs. Data files were saved for each experimental run, where the stimulus and response were recorded for each trial along with the elapsed time between the offset of the stimulus and the onset of the response.

2.2.5 Experimental Conditions. Each of the participants had taken part in the haptic phoneme-identification study reported by Reed et al. (2019) [29], as well as having received previous training on the identification of words composed of sequences of the haptic phonemes. The duration of training for phoneme identification ranged from 1 to 2 hrs across these seven participants, with post-training identification scores in the range of 75% to 97% correct. The seven participants also received training and testing on the identification of words. In this initial phase of exposure to the word-identification task, the IPI was always set at 300 ms. For example, the word "but" (/b/-/ʌ/-/t/) was presented as a sequence of three haptic phonemes in order of B-UH-T (as defined in the first two tables of Reed et al., 2019 [29]), with a 300-ms interval between successive haptic codes. An alphabetized list of the 100-word set (with IPA transcriptions) used in the study is provided in Table 1. Of the 100 words, 31 are composed of 2 phonemes and the remaining 69 of three phonemes resulting in a mean of 2.69 phonemes/word. All 39 of the haptic symbols are represented in the word set.

In the training sessions, participants were given the option of responding by typing in the word from the list shown on the computer screen or by selecting a sequence of phonemes from a set of orthographic characters corresponding to the phonemic symbols that were also made available on the computer screen. Correct-answer feedback was made available during the training phase of the study. In the case of an incorrect response, the word corresponding to the correct answer was shown on the screen. Initial training runs included the use of smaller subsets of the words prior to runs with the full set of 100 words. Training (with IPI of 300 ms) was continued until participants were able to achieve scores of ≥70% correct on the 100-word task. Following training, each participant was tested without correct-answer feedback on three 50-trial runs of words selected randomly with replacement from the 100-word list. On these test runs, participants responded by typing a word selected from the full set of 100 words shown on the screen. Across participants, the mean of the three word-identification test scores ranged from 58.7% to 89.3% correct (mean of 72.7%±9.3), with total training times in the range of 2.2 to 6.2 hrs (mean of 4.1 hrs ±1.2). The range of training times is due to differences in the amount of time individual participants required to achieve criterion performance (≥80% correct on phonemes and ≥70% correct on words).

Immediately following this phase of training and testing, the participants were entered into the IPI study where they received training and testing on the word-identification task using four values of IPI: 300, 150, 75, and 0 ms. In all phases of the IPI study, participants chose their responses from the list of 100 words that was provided on the computer screen. For each IPI in decreasing order, participants received two 20-trial runs of practice with correct-answer feedback, followed by three 50-trial test runs without the use of correct-answer feedback.

Following the fixed-order practice runs in order of decreasing IPI, a final set of testing was conducted using a randomized order of IPIs. The test runs consisted of one 50-trial run with words randomly selected with replacement from the 100-word list at each IPI without the use of correct-answer feedback.

Data were collected in two-hour test sessions with breaks provided as needed. The number of sessions required to complete the training and testing on the IPI study ranged from four to six across participants and was dependent primarily on the length of time each subject required to complete a run as well as differences in the time needed for breaks.

Table 1. List of 100 Words with IPA Transcriptions Used in the Inter-Phoneme Interval (IPI) Study[1]

Word	IPA Ph 1	IPA Ph 2	Word	IPA Ph 1	IPA Ph 2	IPA Ph 3	Word	IPA Ph 1	IPA Ph 2	IPA Ph 3
ace	/eɪ/	/s/	azure	/ae/	/ʒ/	/ɝ/	mock	/m	/ɑ/	/k/
all	/ɔ/	/l/	bad	/b/	/ae/	/d/	mood	/m/	/u/	/d/
bow	/b/	/oʊ/	bath	/b/	/ae/	/Θ/	nut	/n/	/Λ/	/t/
chow	/ʧ/	/aʊ/	bike	/b/	/aɪ/	/k/	pawn	/p/	/ɔ/	/n/
cow	/k/	/aʊ/	but	/b/	/Λ/	/t/	pen	/p/	/Ɛ/	/n/
do	/d/	/u/	came	/k/	/eɪ/	/m/	pool	/p/	/u/	/l/
foe	/f/	/oʊ/	check	/ʧ/	/Ɛ/	/k/	raid	/r/	/eɪ/	/d/
gay	/g/	/eɪ/	chin	/ʧ/	/I/	/n/	rave	/r/	/eɪ/	/v/
guy	/g/	/aɪ/	choose	/ʧ/	/u/	/z/	read	/r/	/i/	/d/
how	/h/	/aʊ/	come	/k/	/Λ/	/m/	ring	/r/	/I/	/ŋ/
jay	/ʤ/	/eɪ/	cut	/k/	/Λ/	/t/	run	/r/	/Λ/	/n/
joy	/ʤ/	/ɔɪ/	dame	/d/	/eɪ/	/m/	sad	/s/	/ae/	/d/
knee	/n/	/i/	deem	/d/	/i/	/m/	same	/s/	/eɪ/	/m/
my	/m/	/aɪ/	den	/d/	/Ɛ/	/n/	seek	/s/	/i/	/k/
no	/n/	/oʊ/	dim	/d/	/I/	/m/	shame	/ʃ/	/eɪ/	/m/
now	/n/	/aʊ/	dirt	/d/	/ɝ/	/t/	shirt	/ʃ/	/ɝ/	/t/
oath	/oʊ/	/Θ/	dome	/d/	/oʊ/	/m/	shun	/ʃ/	/Λ/	/n/
on	/ɔ/	/n/	duck	/d/	/Λ/	/k/	sing	/s/	/I/	/ŋ/
ought	/ɔ/	/t/	fan	/f/	/ae/	/n/	some	/s/	/Λ/	/m/
row	/r/	/oʊ/	fool	/f/	/u/	/l/	tall	/t/	/ɔ/	/l/
say	/s/	/eɪ/	fowl	/f/	/aʊ/	/l/	then	/ð/	/Ɛ/	/n/
see	/s/	/i/	gun	/g/	/Λ/	/n/	thing	/Θ/	/I/	/ŋ/
she	/ʃ/	/i/	has	/h/	/ae/	/z/	thumb	/Θ/	/Λ/	/m/
shoe	/ʃ/	/u/	hen	/h/	/Ɛ/	/n/	ton	/t/	/Λ/	/n/
shy	/ʃ/	/aɪ/	home	/h/	/oʊ/	/m/	turn	/t/	/ɝ/	/n/
sigh	/s/	/aɪ/	keep	/k/	/i/	/p/	vase	/v/	/eɪ/	/s/
thee	/ð/	/i/	learn	/l/	/ɝ/	/n/	vowed	/v/	/aʊ/	/d/
tie	/t/	/aɪ/	limb	/l/	/I/	/m/	wake	/w/	/eɪ/	/k/
toy	/t/	/ɔɪ/	loud	/l/	/aʊ/	/d/	weed	/w/	/i/	/d/
way	/w/	/eɪ/	mace	/m/	/eɪ/	/s/	when	/w/	/Ɛ/	/n/
you	/j/	/u/	made	/m/	/eɪ/	/d/	wide	/w/	/aɪ/	/d/
			make	/m/	/eɪ/	/K/	wing	/w/	/I/	/ŋ/
			maze	/m/	/eɪ/	/z/	woke	/w/	/oʊ/	/k/
			mead	/m/	/i/	/d/	yawn	/j/	/ɔ/	/n/
			men	/m/	/Ɛ/	/n/				

[1]The stimulus list contains 31 two-phoneme words and 69 three-phoneme words. Thus, the probability of a correct response to a two-phoneme word assuming a two-phoneme word response is 0.0323; and for three-phoneme words assuming a three-phoneme word response this probability is 0.0145. Using the distribution of phonemes across the words, the probability of a correct response on the basis of chance alone was calculated for each phoneme position. For two-phoneme words, these probabilities were 0.0614 and 0.1113 for the first and second position, respectively, and for three-phoneme words .0645, 0.1117, and 0.1361 for the first, second, and third positions, respectively.

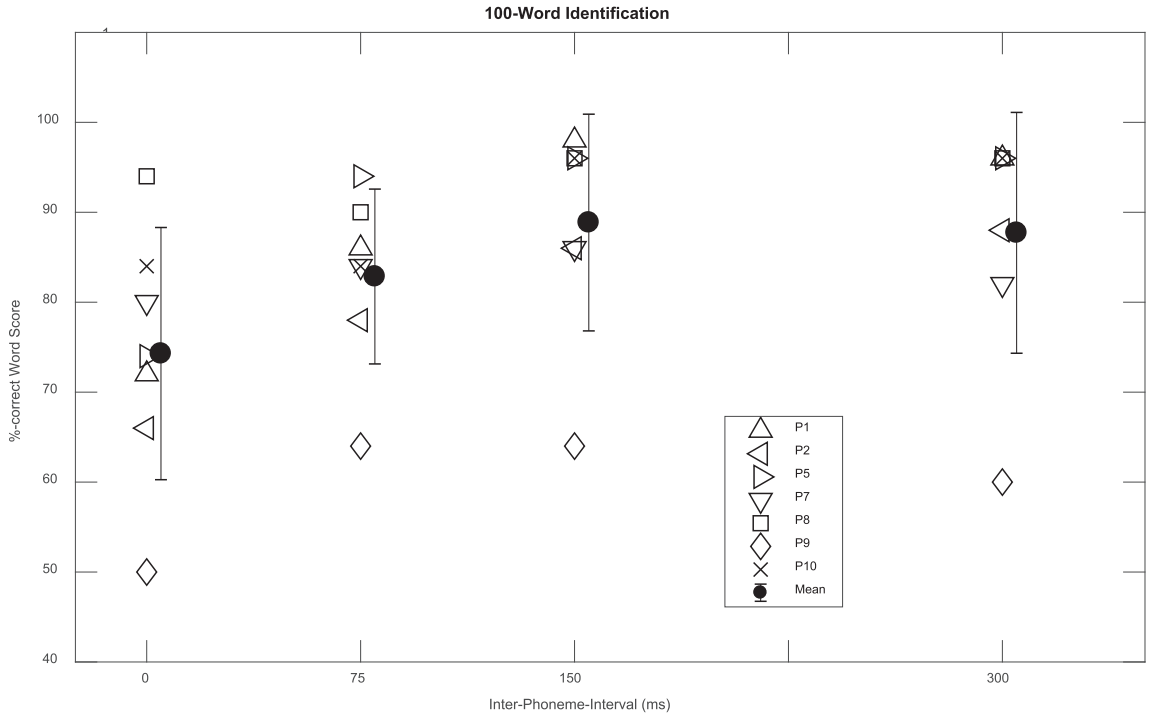

Fig. 2. Percent-correct word score as a function of Inter-Phoneme Interval for individual participants and also showing mean and standard deviation across participants. Data are for randomized-order presentation of IPI.

2.3 Results

The results of the final test runs conducted with randomized order of IPIs are shown in Figure 2. These are the scores of each participant following training and testing on the fixed-order IPI conditions. Thus, the participants were familiar with the word-identification task for each of the IPI conditions when the final tests were conducted. The percent-correct score on each 50-trial run for each of the seven participants is plotted as a function of IPI. Means and standard deviations across participants are also plotted at each IPI. Test scores across participants averaged 87.7, 88.9, 82.9, and 74.3% correct for IPI of 300, 150, 75, and 0 ms, respectively. A repeated-measures ANOVA was conducted to test for the main effect of IPI using the rationalized arcsine transformation of the percent-correct scores [47]. A significant effect of IPI was observed ($F(3, 18) = 11.1$, $p=0.0002$). Using a Tukey-Kramer post hoc comparison test at the .05 level of significance, the only significant differences that were observed were for a lower score at IPI of 0 ms than for scores at IPIs of 300 and 150 ms.

Response times (RTs), defined as the duration between the offset of the stimulus word and the onset of the participant's response, are shown in the box plots of Figure 3 as a function of IPI, aggregated across participants. Median response times increased from 5.4 sec for IPI of 300 ms to 7.1 sec for IPI of 0 ms. For individual participants (across all values of IPI), RTs ranged from 4.3 (P8) to 8.0 sec (P5).

Further analyses were conducted to examine patterns of errors made on the word-identification task using the 350 trials on the randomized order runs across participants at each IPI. The total number of errors for two-phoneme words (which make up 31% of the stimulus set) and three-phoneme words (making up 69% of the stimulus set) at each IPI is given in Table 2. The total percentage of errors on two-phoneme words was 16% and 84% on three-phoneme words. The results of chi-square analyses indicate that the number of errors made on

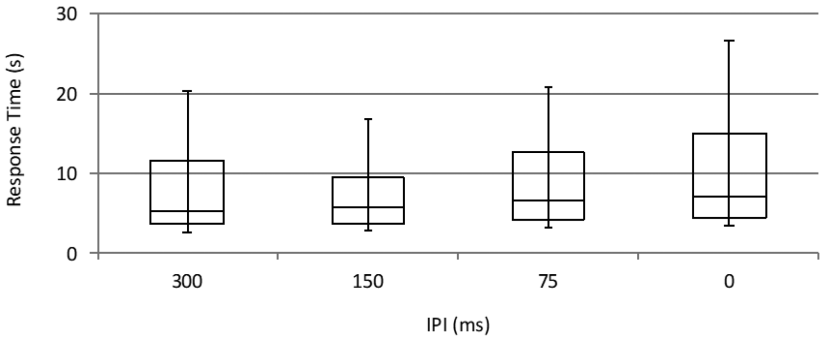

Fig. 3. Box plots of response times aggregated across the seven participants as a function of IPI.

Table 2. Total Number of Errors for 2-phoneme and 3-phoneme Words

IPI in ms:	0	75	150	300	Total
2-phoneme Words	19	5	10	4	38
3-phoneme Words	71	55	29	39	194
Total	90	60	39	43	232

Table 3. Number of Errors with 1 or 2 Missed Phonemes (for 2-phoneme words) or with 1, 2, or 3 Missed Phonemes for 3-phoneme Words

IPI in ms:	0	75	150	300	Total
2-phoneme Words:					
1 error	14	1	7	3	25
2 errors	5	4	3	1	13
Total	19	5	10	4	38
3-phoneme Words:					
1 error	16	13	10	8	47
2 errors	30	25	13	18	86
3 errors	25	17	6	13	61
Total	71	55	39	39	194

the two-phoneme words was lower than expected: χ^2 (3, 38) = 19.77, p = .0002, and the number of errors on the three-phoneme words was higher than expected: χ^2 (3, 194) = 8.93, p = .03.

The incorrect responses were also analyzed to determine the number of phonemes in error (Table 3). For two-phoneme words, the percentage of incorrect responses with only one phoneme in error (66% of total errors) was nearly twice as high as having two phonemes in error (34%), although a chi-square test showed that this effect was not significant (χ^2 (3, 38) = 5.41, p = .1439). Likewise for three-phoneme words, the percentage of phonemes in error (1, 2, or 3) was not significantly different, as shown by a chi-square test (χ^2 (6,194) = 3.12, p = .7930).

A further breakdown of the errors at each IPI was derived from an analysis of the word-identification errors, in which the phonetic transcription of the word that was presented on a given trial was compared with the phonetic transcription of the incorrect response (which was always one of the other 99 words from the closed-set list). For

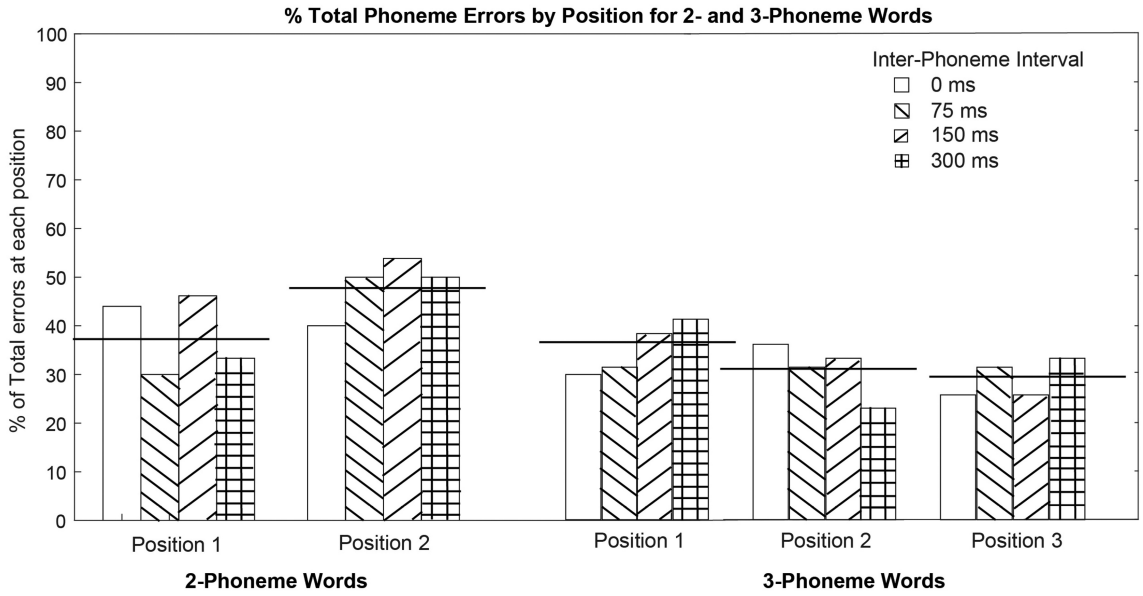

Fig. 4. Percentage of errors by position for 2-phoneme and 3-phoneme words, as a function of inter-phoneme interval. Horizontal bars show mean error rates for each category.

Table 4. Number of Errors by Position for 2-phoneme and 3-phoneme Words

IPI in msec:	0	75	150	300	Total
2-phoneme Words:					
Position 1	11	3	6	2	22
Position 2	10	5	7	3	25
Position 3	4	2	0	1	7
Total	25	10	13	6	54
3-phoneme Words:					
Position 1	47	36	21	35	139
Position 2	56	37	18	20	131
Position 3	48	41	15	28	132
Total	151	114	54	83	402

For 2-phoneme words, errors in Position 3 arise from use of a 3-phoneme response to a 2-phoneme stimulus.

each phoneme position (first or second for the two-phoneme words; first, second, or third for the three-phoneme words), comparisons were made between the phonetic transcriptions of the presented word and the incorrect response.

Figure 4 shows the percentage of the total errors for each word type that were made at each phoneme position in the word. For each word length, the total number of phonemes in error was tabulated and used to normalize the number of errors made at each position. On average across IPI for the two-phoneme words, the percentage of errors was 38% and 49% of the total for the first and second phoneme, respectively. The results of a chi-square analysis of results (see Table 4) indicate that error rate was not dependent on position: χ^2 (3, 47) = .6356,

p=.8882. The erroneous addition of a third phoneme to the response accounted for another 13% of the errors. For the three-phoneme words, an error was roughly equally likely on all three phoneme positions (percentage of errors ranging from 30%–35% across the three positions). The results of a chi-square analysis of data shown in Table 4 support the conclusion that error rate was not dependent on position: χ^2 (6, 402) = 6.1058, p = .4114. Misidentification of a three-phoneme word with a two-phoneme response also was observed, but this was an infrequent occurrence (roughly 5% on average across IPIs).

2.4 Discussion

Experienced users of the haptic display were able to identify words from a 100-item vocabulary (formed from 39 unique haptic phonemes) with a high degree of accuracy for IPIs in the range of 75–300 ms. Even when the IPI was reduced to 0 ms, performance remained in the vicinity of 75% correct. In fact, the highest-performing participant (P8 in Figure 2) maintained scores in excess of 90% correct at all four values of IPI. As IPI decreased from 300 to 0 ms, the mean duration of the 100 words in the set decreased from 1.4 to 0.89 sec. The analysis of errors made on the word-identification task indicates, not unexpectedly, a greater degree of difficulty with words consisting of three compared to two phonemes. In general, the three-phoneme words are longer in duration than the two-phoneme words, thus requiring a higher cognitive load and greater working memory for encoding compared to the two-phoneme words. For example, with IPI of 150 ms, the duration of the two-phoneme words ranged from 690 to 950 ms, compared to a range in duration of 820 to 1,500 ms for the three-phoneme words. No significant trends were observed in the data for a greater likelihood of errors as a function of phoneme position in the word or for the number of phonemes in error. Thus, it appears that participants were attending to the complete phonetic make-up of the word in selecting their responses.

The results indicate that IPI of 75 ms may provide a good solution to the challenge of correctly identifying a multi-phoneme word. For shorter values of IPI, temporal masking of successive signals may increase the difficulty of the task, due to effects of both forward and backward masking. Using the example of a two-phoneme word, the tactile representation of the first phoneme may persist through the inter-phoneme gap and combine with that of the succeeding phoneme, leading to reduced ability to identify these phonemes. In fact, studies of tactile temporal masking using multidimensional tactual patterns indicate strong effects of forward and backward masking in the region of a 0–75 ms gap between successive stimuli [48]. However, when the IPI is relatively long (e.g., >300 ms condition), a toll is placed on working memory to remember the tactile signal that was presented more than a second ago. Setting IPI to 75 ms could provide a tradeoff between tactile masking and working memory.

Because training and testing were conducted with the same 100-word vocabulary, the participants received multiple exposures to the individual words in this stimulus set. With such repeated exposure, the possibility exists that participants may have begun to recognize the haptic patterns individual words as a whole rather than through the sequential identification of individual phonemes. However, studies of Morse Code indicate that chunking ability requires many more hours of practice than that provided to the participants of the current study [49]. The ability to chunk individual tactile phonemes into words would represent a major step towards the practical use of TAPS as a communication. Given the limited amount of exposure of the participants to the TAPS system, it appears unlikely that the ability to chunk the haptic phonemes into larger units for deciphering words occurred in the current study. However, this step in the learning process is likely necessary for a practical system of communication. In addition, the current response times for word identification (on the order of 6 sec) are not conducive to practical use of the TAPS system, indicating the need for further training and exposure in its use.

The histogram shown in Figure 5 indicates that some words were more accurately identified than others. Among the 100 words in the full set, there were 23 on which no errors were made across all participants and values of IPI. However, there was a group of 18 words with error rates in excess of 30%: All but one of these words contained at least one of the 6 plosive phonemes, 12 contained one of the two nasal phonemes, and

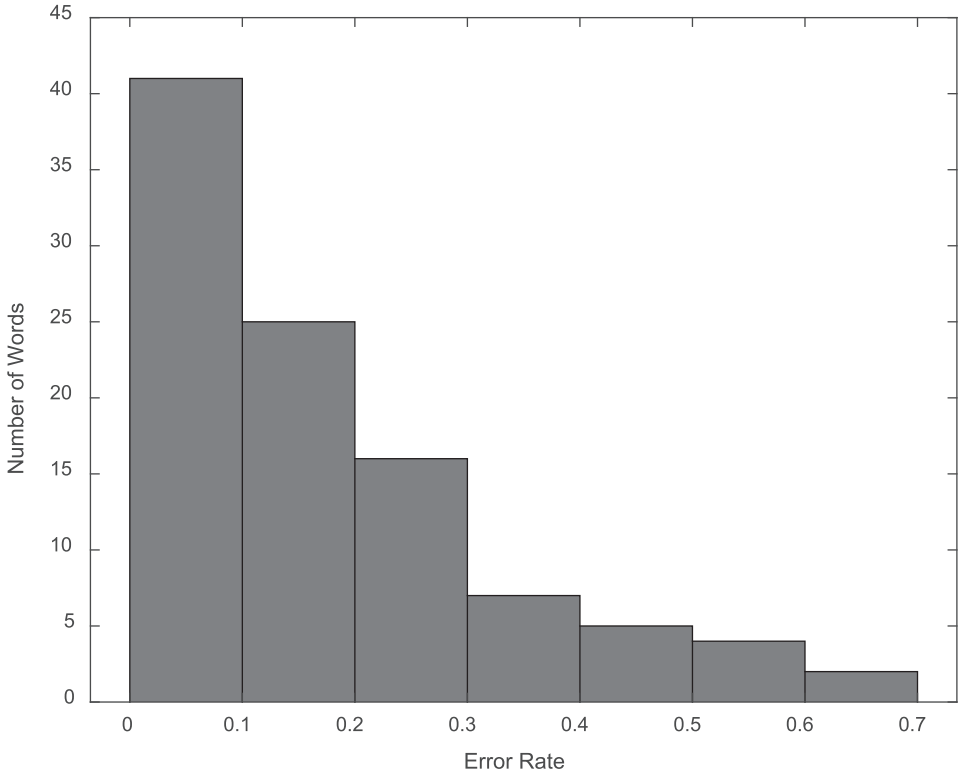

Fig. 5. Histogram showing number of words with error rates in bins with width of 0.1.

7 contained one of the 4 "short" vowels. It appears that the shorter phonemes (plosives and short vowels) as well as the nasals (characterized by the use of a 60 Hz carrier) contributed to difficulty on the word identification task, even though these haptic codes were well identified in isolation (Reed et al., 2019) [29].

3 STUDY 2: EFFECT OF INTER-WORD INTERVAL ON IDENTIFICATION OF TWO-WORD PHRASES

3.1 Overview

The goal of this study was to examine the ability to identify two-word sequences as a function of the inter-word interval. Each of the participants had been trained in previous experiments to identify the 39 tactile phonemic codes and to identify words composed of these codes. A set of 100 words was selected for use in creating two-word phrases that were semantically and syntactically valid. The participants received training on the identification of the individual words presented through the tactile display prior to being asked to identify two-word phrases. Five values of inter-word interval ranging from 300 ms to 2,000 ms were included in the testing. Participants were trained and tested on a task in which they were asked to identify both words of a two-word phrase. Training runs included correct-answer feedback, which was eliminated in the testing phase.

3.2 Methods

Data were collected at two different sites (MIT and Purdue University). The MIT subjects provided informed consent through a protocol approved by the IRB at MIT, and the Purdue University subjects provided informed

Table 5. Summary of Participants' Previous Experience in Studies with TAPS

Participant	PU 108	PU 205	PU 302	PU 305	MIT 046	MIT 047	MIT 052	MIT 055
Participant Number in Tan et al. (2020) [31]	(P18)	(P27)	(P35)	(P36)	(P40)	(P41)	(P48)	(P51)
Total Previous Exposure (hrs)	8.5	4.3	18.4	16.6	11.9	10.4	10.6	7.8
Score on 500 words	94%	90%	90%	80%	69%	78%	65%	77%

consent through a protocol approved by the IRB at Purdue. Participants were paid for their work on the study. At both sites, the apparatus and tactile stimulus codes were as described above for Study 1. The procedures for fitting the tactile device and adjusting presentation levels were also the same at both sites, as described above for Study 1, with the exception that testing was conducted in a quiet laboratory space at Purdue and within a sound-treated booth at MIT. Any further procedural differences between the two sites are noted below.

3.2.1 Participants. A total of eight participants were recruited for this study based on their having demonstrated proficiency in previous experiments concerned with the identification of phonemes and individual words presented through the tactile display. Half of the participants were tested at Purdue University and half at MIT. The participants (five female and three male) ranged in age from 18 to 23 yrs (mean of 21.1 yrs and standard deviation of 1.6 yrs), reported right-hand dominance and were native English speakers. The experience of these participants on previous studies with the TAPS system is summarized in Table 5. Prior to the current study, each of the participants had received training and testing for identification of words with a 500-word vocabulary. The duration of all previous exposure to the TAPS device along with final scores obtained with the 500-word vocabulary are provided in Table 5.

3.2.2 Experimental Conditions. A list of 100 words was selected for use in creating 218 two-word phrases. The word list, provided in alphabetical order in Table 6, consists of 1 one-phoneme word, 27 two-phoneme words, and 72 three-phoneme words. The words represent different parts of speech (nouns, pronouns, verbs, adjectives, adverbs) and were selected so various pairs of words could be combined to form meaningful phrases. Examples of 20 of the two-word phrases are provided in Table 7. Each of the phrases was selected to be semantically and syntactically valid and representative of word sequences that may occur naturally in spoken utterances. Words from the list could be used in either the first or second position of the phrase, and some words occurred in both positions as shown in the examples of "good job" and "no good" in Table 7.

The words were phonetically transcribed and presented as a sequence of tactile phoneme codes with an interphoneme interval of 150 ms. Although the results of Study 1 indicated no significant difference in performance with IPIs of 75 and 150 ms, the decision was made to use IPI of 150 ms in Study 2 considering both the word scores and the reaction times, which were lower at 150 ms and showed less variability across participants. Participants were first given practice with identification of single words from the 100-word vocabulary used to generate the phrases. The single-word identification task was conducted using the one-interval identification procedure described above for Study 1. After participants had demonstrated proficiency with the single-word identification task (i.e., scores greater than 80% correct), they then received training and testing of their ability to identify two-word phrases.

For the two-word phrase testing, the **inter-word interval (IWI)** included durations of 2,000, 1,000, 750, 500, and 300 ms. The specific values that were tested varied across participants, based on their performance on the task as well as their availability for the experiment. As shown in Table 8, all participants were tested with IWIs of 1,000 and 750 ms, and all except P108 (due to experimenter error) at IWI of 2,000 ms. Four participants proceeded to testing with IWI of 500 ms, and one of these (P205) advanced to an IWI of 300 ms after demonstrating good performance at 500 ms. The identification procedure described for Study 1 was also employed for the two-word phrase tests, with the following differences: On each trial, one of the two-word phrases was selected at random

Table 6. Alphabetical List of 100 Words Used for the Two-word Phrases

Words 1–25	Words 26–50	Words 51–75	Words 76–100
age	fish	men	soup
at	five	mood	tall
bad	fun	my	team
big	girl	no	ten
bird	good	noise	these
book	hat	nurse	they
both	have	off	thin
boy	hen	pan	this
cake	high	pay	time
catch	home	pipe	toe
cheese	hope	pitch	too
coat	I	rain	town
come	job	rich	toy
cool	lake	ring	us
cows	leg	rode	voice
day	less	run	way
deep	light	sad	we
dim	like	same	week
dog	loose	see	white
dude	loud	shade	wide
each	low	shoe	wife
face	mad	show	win
fat	man	shy	you
feet	me	so	young
fine	mean	some	zoo

Table 7. Examples of 20 of the Two-word Phrases Used in the Testing

Word 1	Word 2	Word 1	Word 2	Word 1	Word 2	Word 1	Word 2
I	see	cool	dude	light	rain	pay	off
bad	mood	each	day	like	this	see	me
big	bird	fat	hen	low	voice	tall	man
both	men	good	job	my	team	we	hope
come	home	have	fun	no	good	wide	lake

with replacement, and the participant's task was to type a two-word response with no limit on response time. The trial consisted of a sequence of phonemes corresponding to the particular two-word phrase selected for presentation (with IPI of 150 ms and the value of IWI for a specific test condition). Although participants had been exposed to the words that were used to make up the phrases, this 100-word vocabulary was not made available to the participants during the testing with phrases. Thus, it was possible for participants to use words not in the original vocabulary in their responses.

Table 8. Percent-correct Scores (%) and Training Times (mins) as a Function of Inter-word Interval (IWI) for Each of the 8 Individual Participants on the Two-word Identification Task

IWI (ms)	2,000	2,000	1,000	1,000	750	750	500	500	300	300
	Score %-Corr	Time Mins	Score %-Corr	Time Mins	Score %-Corr	Time Mins	Score %-Corr	Time Mins	Score %-Corr	Time Mins
PU108			86	34	90	58				
PU205	70	40	85	65	80	81	85	92	75	99
PU302	80	93	90	134	80	142	65	189		
PU305	80	57	85	94	85	104				
MIT046	30	63	58	105	34	126				
MIT047	68	79	76	128	64	169				
MIT052	74	63	52	121	70	149	74	161		
MIT055	60	37	68	63	60	88	76	102		

Training times are relative to the start of training on this task and are cumulative across decreasing values of IWI.

Participants received training and testing on the identification task with IWIs presented in decreasing order. For each IWI, training runs consisted of 20 or 25 trials in which trial-by-trial correct-answer feedback was provided in the form of the correct two-word phrase appearing above the response entered by the participant. Training continued with a given value of IWI until scores stabilized (i.e., improvement of less than 5 percentage points between two consecutive runs). At the MIT site, participants were then tested on two 25-trial runs at that IWI without the use of correct-answer feedback. At the Purdue site, the score reported at each IWI was based on the final run conducted with feedback (typically 20 trials). The total number of trials presented for training and testing varied across participants, depending on their performance and the number of conditions tested. With phrases drawn at random with replacement from the 218-phrase stimulus set, the maximum number of presentations of any given phrase over the course of training and testing for any participant was expected to be three to four.

3.3 Results

Results of the two-word phrase experiment are summarized in Table 8. For each participant, the percent-correct score for correct identification of both words in the phrase and the duration of the training time (relative to the start of the two-word training task and cumulative from that point) are provided for each IWI that was tested. Scores and training time varied across participants. The highest-performing participant (PU205) achieved a score of 75% correct with IWI of 300 ms after a total training time of 99 minutes, compared to MIT046 who scored 34% correct at IWI of 750 ms after 126 minutes of training.

All participants were tested at IWIs of 1,000 and 750 ms. At IWI of 1,000 ms, final scores ranged from 52%–90% (mean of 75.0%) with training times in the range of 34–134 min (mean of 93.0 min). At IWI of 750 ms, scores ranged from 34%–90% (mean of 70.4%) with training times in the range of 58–169 min (mean of 115.9 min). Four participants advanced to an IWI of 500 ms (mean score of 75% correct achieved with a mean training time of 136 min), and PU205 was able to identify 75% of the phrases correctly at an IWI of 300 ms within a cumulative training time of 99 min.

For the data of the seven subjects who were tested on IWI of 2,000, 1,000, and 750 ms, scores averaged 63.8%, 76.0%, and 67.6% correct, respectively. A one-way ANOVA conducted on the rationalized arcsine transformation of the percent-correct scores [31] indicated no significant effect of IWI ($F(2,18) = 1.22$, $p = .319$).

Across participants and IWIs, the results were examined to determine the percentage of responses made in each of the following categories: correct identification of both words (70%), first word correct/second word incorrect (15%), first word incorrect/second word correct (6%), and both words incorrect (9%). Thus, the most common error response was correct reception of the first word only.

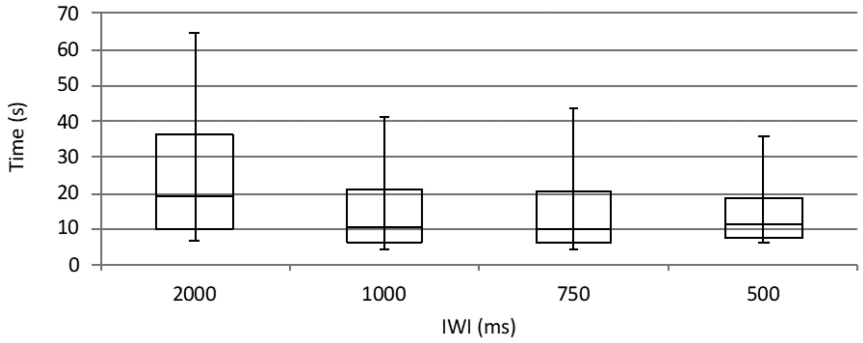

Fig. 6. Box plot of response times across participants at each of the four values of IWI (see Table 8 for a list of participants included at each value).

The **response times (RTs)**, defined as the duration between the offset of the second word in the phrase and the onset of the participant's response, are shown in the box plots of Figure 6. The RTs are shown as a function of IWI for data across all participants tested at a given IWI (see Table 8). Median RT was 19.1 sec for IWI of 2 sec and decreased to roughly 10 sec for RTs of 1, 0.75, and 0.5 sec. Median RTs varied widely across participants with a range of 4.8 (PU108) to 35.2 (PU302) sec.

Estimates of the rates of communication for the two-word phrases were computed in the following manner: First, the duration of each of the 100 words in the vocabulary was calculated using the durations of the phonemes in that particular word and a 150-ms inter-phoneme duration. For example, for the word "like," transcribed phonetically as /l/-/aI/-/k/, the duration of the three phonemes is 400 + 400 + 140 ms with two inter-phoneme intervals of 150 ms each, for a total duration of 1,240 ms. The mean duration across the 100 words was 1,150 ms. The communication rate was calculated for each of the five values of IWI used in the study in the following way: The mean duration was used to represent each of the words in the phrase (i.e., 1,150 × 2 = 2,300 ms) and a given value of IWI was added to this duration. For example, for an IWI of 500 ms, the total duration of the two-word phrase was 2,800 ms. Thus, the rate of presentation was 1.4 sec/word, which translates to 42.9 words/min. This represents the highest rate that can be achieved given the specified temporal parameters and assuming perfect recognition on the part of the user. These theoretically highest rates (shown by the asterisk symbols in Figure 7) ranged from 27.9 words/min (for IWI of 2,000 ms) to 46.2 words/min (for IWI of 300 ms).

Estimates of the effective rate at which phrases were transmitted (Effective Transmission Rate: ETR) in words/min took into account the intelligibility score for a given experimental condition for each participant by multiplying the **%-correct score (PC)** by the **presentation rate (PR)** in wpm:

$$\textbf{ETR}\ (\text{words/min}) = \textbf{PR}\ (\text{words/min}) \times \textbf{PC}$$

Thus, perfect reception leads to ETR = PR while a percent correct score of 50% results in ECR that is one-half of PR. This formulation of ETR = PR X PC is derived from previous work [50] on optimal **information-transfer (IT)** rates, showing a tradeoff between rate of presentation and information received to yield a constant peak IT rate (analogous to ECR above) that depends on degree of familiarity and training with the stimulus set.

In Figure 7, estimated transmission rates for each participant on the two-word phrase task are plotted as a function of IWI. The maximum rate for each IWI, shown by the asterisk (*) symbols in the plot, was multiplied by the percent-correct score for two-word phrase reception for each participant at that IWI (shown in Table 8). Maximum ETR was in the range of 21 to 26 words/min for two of the participants (MIT046 and MIT047), and from 30 to 35 words/min for the remaining six participants. Focusing on the data of the seven participants who were tested at IWI of 2,000, 1,000, and 750 ms, ETR averaged 17.8, 26.7, and 26.8 words/min, respectively. A one-way ANOVA of these data indicated a significant effect of IWI ($F(2, 18) = 5.67$, $p = 0123$), with a Tukey-Kramer

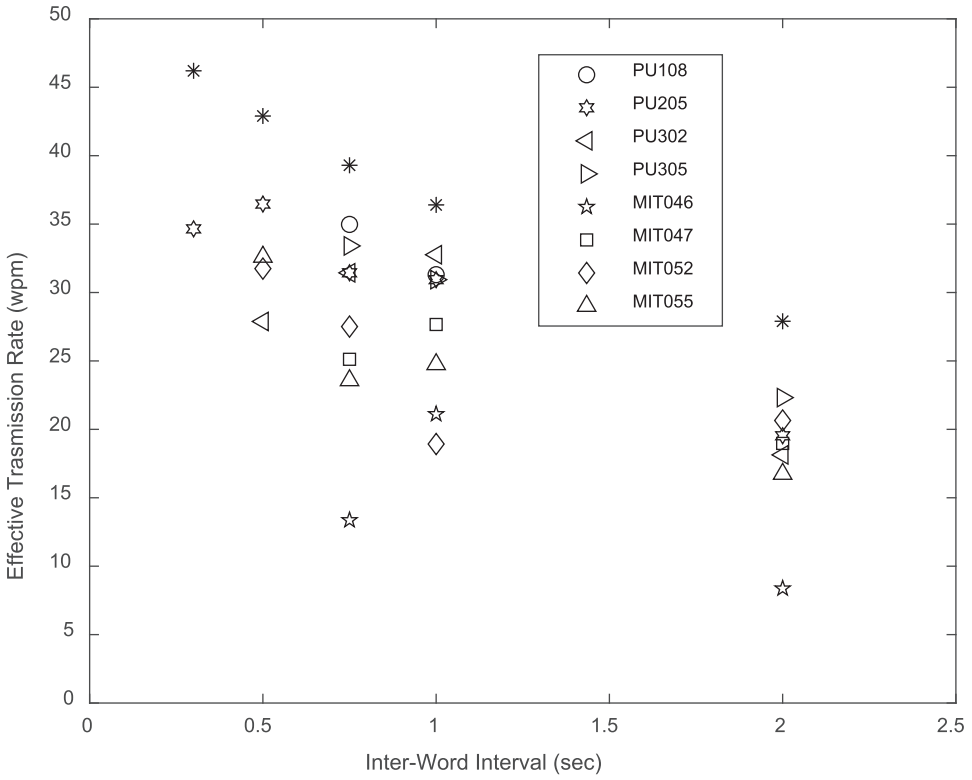

Fig. 7. Performance shown as effective transmission rate (ETR) in words/min (wpm) plotted as a function of the inter-word interval (IWI) on the two-word identification task for each of the eight individual participants. The maximum achievable rate for each of the five values of IWI used in the testing is shown by the asterisk (*) symbols. The percent-correct score achieved at each duration of IWI was multiplied by the maximum rate to estimate the effective transmission rate (ETR) in terms of the number of words per minute that were correctly identified.

multiple comparison at a significance level of .05 indicating that communication rates achieved with IWI of 2,000 msec were significantly lower than for the two lower values of IWI, which were not significantly different from each other.

3.4 Discussion

A range of performance was observed across participants in their ability to identify two-word phrases. Across all values of IWI after training, performance on individual participants ranged from a low of 30% correct (MIT046 for IWI of 2,000 msec) to 90% correct (PU302 at IWI of 1,000 msec and PU108 at IWI of 750 msec). The most frequent type of error involved correct identification of the first word in the phrase but with an incorrect response for the second word, occurring on 15% of the trials, compared to a 6% occurrence of errors in which the second word was correctly identified while the first word was incorrect. This pattern of errors suggests a possible strategy of sequential phoneme-by-phoneme processing across the time course of the two-word stimulus phrase, and that errors on the second word may have arisen due to limited attentional capacity for the second word while processing the first word. However, participants achieved correct identification of both words in the phrase on 70% of the trials, thus indicating the capability of processing both words accurately and holding them in memory before making a response.

Presentation rates were varied by changing the inter-word duration over a range of 2,000 to 300 ms. For presentation rates across the range of 750 to 2,000 msec, communication rates were significantly lower for IWI of 2,000 ms (18 words/min) than for IWIs of 1,000 and 750 ms (27 words/min). The results at IWI of 1,000 and 750 ms suggest a tradeoff between presentation rate and accuracy of reception, with a corresponding decrease in percent-correct scores with an increase in presentation rate. The effective transmission rates achieved here are roughly half of those reported for users of the Tadoma method, which are on the order of 60–80 words/min (which themselves are roughly half the rate at which speech is received through the normal auditory channel). Several limitations of the current study should be noted in making comparisons with Tadoma. Whereas sentence reception from a large vocabulary of English words was measured in Tadoma users, the current study was limited to two-word phrases drawn from a 100-vocabulary of words that were primarily one, two, or three phonemes in length. Another important consideration in the development of a practical system for tactile speech communication is the length of time required for users to respond to the messages. The response time for two-word phrases was on the order of 10 secs, somewhat higher than the 6 sec response time for single words with IPI of 150 msec (see Figures 3 and 6) and must also be taken into account in determining the rate at which two-way communication can be conducted. Further training and practice on TAPS are necessary to enable users to decrease their response times for use in a practical system.

4 GENERAL DISCUSSION

The results reported here on word recognition through TAPS may be compared to those obtained with other approaches for encoding and displaying words through the tactile sense, including spectral-based displays of speech, alphabetic-based displays of text, and phonemic-based codes.

We first consider previous studies of tactile-alone word recognition in which the tactile display is derived from spectrally based transformations of the acoustic speech signal (e.g., References [9, 10, 12]). Results from these studies indicate training times of 40 hrs to acquire a vocabulary of 70 words [9], 80 hrs for 250 words [10], and 18 hrs for 50 words [12]. Although these training times are longer than those observed for the phonemic-based TAPS system in Study 1 (less than 10 hrs to acquire 100 words), they compare favorably to those reported for normal-hearing learners of the Tadoma method [37]. For both spectral-based displays and Tadoma, however, studies employed live-voice speech utterances of the test stimuli. Thus, words are presented at normal speaking rates, and each stimulus is a fresh utterance, requiring participants to cope with the effects of both within- and across-talker token variability [51]. These factors likely contribute to the generally longer training times observed for spectral displays compared to the phonemic-based system studied here, where the codes are longer in duration than spoken phonemes and each phoneme is represented by a fixed tactile code.

Our results for tactile communication of words can also be compared to the approach of alphabetic-based tactile display. Luzhnica et al. [38] studied recognition of words presented through a six-tactor array for conveying 26 alphabetic codes at a mean duration of 0.6 sec/word. In further work with a seven-tactor array that led to improved perception of the tactile alphabetic codes, Luzhnica and Veas (2019) [40], conditions included codes with durations as short as 70 ms, 100 ms between letters, and an average of 3.5 letters/word. These settings correspond to transmission rates of roughly 0.5 sec/word, which are over twice as fast as those of the current study, which had a mean duration of 1.15 sec/word in Study 1. However, alphabetic coding is less efficient than phonemic coding in that, on average, more letters than phonemes are required to represent a given word. Because these authors did not report the scores for recognizing whole words themselves, it is not possible to compute effective transmission rates for comparison with the current results.

Most relevant to the current study is previous work concerned with the development and testing of phonemic-based tactile codes. In the studies reported here with the TAPS system, words were composed from a full set of 39 English haptic phonemes presented over a 24-channel tactile array. Phoneme durations ranged from 140 to 400 ms with IPI fixed by the experimenter. With an IPI of 150 ms, for example, the mean word duration in

Study 1 was 1.15 secs with an average phoneme presentation rate of 0.43 sec/phoneme. These stimulus parameters resulted in the identification of words from a 100-word practiced vocabulary of roughly 90% correct. Several other studies of phonemic-based tactile displays have achieved faster rates of phonemic presentation, but with substantially smaller sets of phonemes and words. For example, Zhao et al. (2018) [26] coded nine tactile phonemes at a rate of 0.28 sec/phoneme and reported a mean score of 83% correct on a 20-word vocabulary. Turcott et al. [28] coded 10 tactile phonemes at a rate of 0.22 sec/phoneme symbols, leading to a mean score of 76% correct on two different 10-word stimulus sets. Thus, it is not known whether these same rates of phonemic presentation would allow for comparable word-identification performance after an increase to the full set of 39 English phonemes. Dunkelberger et al. (2018) [34] developed tactile codes for 23 English phonemes and used a set of 150 words from which stimuli were drawn for training and testing. Unlike in the current study where the inter-phoneme interval was set by the experimenter, a self-paced rate of presentation was used between the phonemes in a given word, resulting in an average phoneme presentation rate of 3.5 sec/phoneme. Post-training word-identification scores averaged 87% correct, similar to the word scores with IPI in the range of 75 to 300 ms in Study 1 here. However, the rates at which word stimuli were presented here were roughly eight times faster than those of Dunkelberger et al. (2018) [34]. In the work of Fontana de Vargas et al. [35], 24 tactile phonemes were encoded with a phoneme presentation rate of 3.1 sec/phoneme, similar to that of Reference [34]. In post-training tests conducted with an open response format, mean score of 51% was achieved on words to which the participants had previously been exposed. Thus, these studies, along with the results of the current study, indicate promising results with the use of phonemic-based tactile displays for speech communication.

5 CONCLUDING REMARKS

The results reported here have demonstrated the feasibility of a phonemic-based haptic code for transmission of words in the English language. Within two to six hours of training, participants were able to receive individual words from a 100-word vocabulary made up of 39 haptic phonemes with an accuracy of 90% correct using an inter-phoneme interval of 150 ms. When words were combined into two-word phrases selected from a 100-word vocabulary including different parts of speech, participants were successful at identifying both words in the phrase with an accuracy of 75% correct using inter-word intervals on the order of 1 sec. The training time required to perform this task was relatively modest, ranging from roughly 1 to 3 hrs across participants. Estimates of the highest effective transmission rates achieved on the two-word phrase task were on the order of 30 to 35 words/min. Response times for two-word phrase identification, however, were roughly twice as long as for the identification of individual words and must also be taken into account in determining the rate at which two-way communication can be conducted.

Future research will be directed towards an increase in the effective communication rates that can be achieved with the phonemic-based haptic display, with the goal of at least doubling the rates demonstrated here. This will include work designed to increase the presentation rate at which the codes are delivered as well as the accuracy and speed at which they can be received. This research will address some of the limitations of the current study, in which we will expand the limited sets of words studied here to an unrestricted vocabulary of English words and increase the length of the two-word phrases to longer phrases and ultimately to full sentences. We will consider the use of a new set of phoneme codes that take into account the statistical properties of English [30] as well as the development of improved training methods to expedite learning. Once the phoneme codes have been mastered, it is theoretically possible to decode any word or phrase. An important question to be addressed in future studies is the degree to which training and testing on a particular vocabulary of words or phrases will transfer to new and larger sets of stimuli. Our long-term goal is to demonstrate that the skin is capable of receiving speech through an artificial tactual display at rates comparable to those achieved by experienced deaf-blind users of the natural method of Tadoma.

ACKNOWLEDGMENTS

We wish to acknowledge the contributions of Ali Israr, Frances Lau, Keith Klumb, Robert Turcott, and Freddy Abnousi of Facebook to this research through their valuable insights, advice, and support. The authors also wish to thank Patrick Zurek for his assistance with probability calculations.

REFERENCES

[1] S. J. Norton, M. C. Schultz, C. M. Reed, L. D. Braida, N. I. Durlach, W. M. Rabinowitz, and C. Chomsky. 1977. Analytic study of the Tadoma method: Background and preliminary results. *J. Speech Hear. Res.* 20, 3 (1977), 574–595.

[2] C. M. Reed, W. M. Rabinowitz, N. I. Durlach, L. D. Braida, S. Conway-Fithian, and M. C. Schultz. 1985. Research on the Tadoma method of speech communication. *J. Acoust. Soc. Amer.* 77, 1 (1985), 247–257.

[3] C. M. Reed. 1995. Tadoma: An overview of research. In *Profound Deafness and Speech Communication,* G. Plant and K.-E. Spens (Eds.). Whurr Publishers, London, 40–55.

[4] C. M. Reed, L. A. Delhorne, N. I. Durlach, and S. D. Fischer. 1990. A study of the tactual and visual reception of fingerspelling. *J. Speech Hear. Res.* 33 (1990), 786–797.

[5] C. M. Reed, L. A. Delhorne, N. I. Durlach, and S. D. Fischer. 1995. A study of the tactual reception of sign language. *J. Speech Hear. Res.* 38 (1995), 477–489.

[6] C. M. Reed, N. I. Durlach, and L. A. Delhorne. 1992. Natural methods of tactual communication. In *Tactile Aids for the Hearing Impaired,* I. R. Summers (Ed.). Whurr Publishers, London, 218–230.

[7] C. M. Reed and N. I. Durlach. 1998. Note on information transfer rates in human communication. *Pres.: Teleop. Virt. Environ.* 7, 5 (1998), 509–518.

[8] S. Engelmann and R. Rosov. 1975. Tactual hearing experiment with deaf and hearing subjects. *J. Except. Child.* 41, 4 (1975), 243–253.

[9] P. L. Brooks and B. J. Frost. 1983. Evaluation of a tactile vocoder for word recognition. *J. Acoust. Soc. Amer.* 74, 1 (1983), 34–39.

[10] P. L. Brooks, B. J. Frost, J. L. Mason, and K. Chung. 1985. Acquisition of a 250-word vocabulary through a tactile vocoder. *J. Acoust. Soc. Amer.* 77, 4 (1985), 1576–1579.

[11] M. P. Lynch, R. E. Eilers, D. K. Oller, and L. Lavoie. 1988. Speech preception by congenitally deaf subjects using an electrocutaneous vocoder. *J. Rehab. Res. Devel.* 25, 3 (1988), 41–50.

[12] K. L. Galvin, P. J. Blamey, M. Oerlemans, R. S. Cowan, and G. M. Clark. 1999. Acquisition of a tactile-alone vocabulary by normally hearing users of the Tickle Talker^{TM}. *J. Acoust. Soc. Amer.* 106, 2 (1999), 1084–1089.

[13] A. Boothroyd. 1989. Developing and evaluating a tactile speechreading aid. *Volta Rev.* 91 (1989), 101–112.

[14] L. Hanin, A. Boothroyd, and T. Hnath-Chisolm. 1988. Tactile presentation of voice fundamental frequency as an aid to the speechreading of sentences. *Ear Hear.* 9, 6 (1988), 335–341.

[15] C. M. Reed and L. A. Delhorne. 1995. Current results of a field study of adult users of tactile aids. *Semin. Hear.* 16, 4 (1995), 305–315.

[16] J. M. Weisenberger, S. P. Broadstone, and F. A. Saunders. 1989. Evaluation of two multichannel tactile aids for the hearing impaired. *J. Acoust. Soc. Amer.* 86 (1989), 1764–1775.

[17] H. Yuan, C. M. Reed, and N. I. Durlach. 2005. Tactual display of consonant voicing as a supplement to lipreading. *J. Acoust. Soc. Amer.* 118, 2 (2005), 1003–1015.

[18] J. Rönnberg, U. Andersson, B. Lyxell, and K. E. Spens. 1998. Vibrotactile speech tracking support: Cognitive prerequisites. *J. Deaf Stud. Deaf Educ.* 3, 2 (1998), 143–156.

[19] L. E. Bernstein, M. E. Demorest, D. C. Coulter, and M. P. O'Connell. 1991. Lipreading sentences with vibrotactile vocoders: Performance of normal-hearing and hearing-impaired subjects. *J. Acoust. Soc. Amer.* 90, 6 (1991), 2971–2984.

[20] P. L. Brooks, B. J. Frost, J. L. Mason, and D. M. Gibson. 1986. Continuing evaluation of the Queen's University tactile vocoder II: Identification of open set sentences and tracking narrative. *J. Rehab. Res. Devel.* 23, 1 (1986), 129–138.

[21] C. M. Reed, N. I. Durlach, L. A. Delhorne, W. M. Rabinowitz, and K. W. Grant. 1989. Research on tactual communication of speech: Ideas, issues, and findings. *Volta Rev.* 91 (1989), 65–78.

[22] M. Benzeghiba, R. De Mori, O. Deroo, S. Dupont, T. Erbes, D. Jouvet, L. Fissore, P. Laface, A. Mertins, C. Ris, and R. Rose. 2007. Automatic speech recognition and speech variability: A review. *Speech Commun.* 49, 10-11 (2007), 763–786.

[23] M. Sreelakshmi and T. D. Subash. 2017. Haptic technology: A comprehensive review on its applications and future prospects. *Mater. Today: Proc.* 4, 2 (2017), 4182–4187.

[24] T. Wade and L. L. Holt. 2005. Incidental categorization of spectrally complex non-invariant auditory stimuli in a computer game task. *J. Acoust. Soc. Amer.* 118, 4 (2005), 2618–2633.

[25] J. Jung, Y. Jiao, F. M. Severgnini, H. Z. Tan, C. M. Reed, A. Israr, F. Lau, and F. Abnousi. 2018. Speech communication through the skin: Design of learning protocols and initial findings. In *Proceedings of the International Conference of Design, User Experience, and Usability.* 447–460.

[26] S. Zhao, A. Israr, F. Lau, and F. Abnousi. 2018. Coding tactile symbols for phonemic communication. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems.* 1–13.

[27] 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. 2018. *A Comparative Study of Phoneme- and Word-based Learning of English Words Presented to the Skin*. Springer International Publishing AG.

[28] R. Turcott, J. Chen, P. Castillo, B. Knott, W. Setiawan, F. Briggs, K. Klumb, F. Abnousi, P. Chakka, F. Lau, and A. Israr. 2018. Efficient evaluation of coding strategies for transcutaneous language communication. In *Proceedings of EuroHaptics 2018* (Springer LNCS 10894). 600–611.

[29] 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. 2019. A phonemic-based tactile display for speech communication. *IEEE Trans. Haptics* 12, 1 (2019), 2–17.

[30] J. S. Martinez, H. Z. Tan, and C. M. Reed. 2021. Improving haptic codes for increased speech communication rates in a phonemic-based tactile display. *IEEE Trans. Haptics* 14, 1 (2021), 200–211.

[31] H. Z. Tan, C. M. Reed, Y. Jiao, Z. D. Perez, E. C. Wilson, J. Jung, J. S. Martinez, and F. Severgnini. 2020. Acquisition of 500 English words through a TActile phonemic sleeve (TAPS). *IEEE Trans. Haptics* 13, 4 (2020), 745–760.

[32] N. Dunkelberger, J. L. Sullivan, J. Bradley, I. Manickam, G. Dasarathy, R. G. Baraniuk, and M. K. O'Malley. 2021. A multi-sensory approach to present phonemes as language through a wearable haptic device. *IEEE Trans. Haptics*. 14, 1 (2021), 188–199.

[33] J. Chen, R. Turcott, P. Castillo, W. Setiawan, F. Lau, and A. Israr. 2018. Learning to feel words: A comparison of different learning approaches to acquire haptic words. In *Proceedings of the ACM Symposium on Applied Perception (SAP'18)*.

[34] N. Dunkelberger, J. Sullivan, J. Bradley, N. P. Walling, I. Manickam, G. Dasarathy, A. Israr, F. W. Y. Lau, K. Klumb, B. Knott, F. Abnousi, R. Baraniuk, and M. K. O'Malley. 2018. Conveying language through haptics: A multi-sensory approach. In *Proceedings of the ACM International Symposium on Wearable Computers*, 25–32.

[35] M. Fontana de Vargas, A. Weill-Duflos, and J. R. Cooperstock. 2019. Haptic speech communication using stimuli evocative of phoneme production. In *Proceedings of the IEEE World Haptics Conference (WHC'19)*, 610–615.

[36] S. Novich. 2015. *Sound-to-Touch Sensory Substitution and Beyond*. PhD Dissertation. Department of Electrical and Computer Engineering, Rice University, Houston, TX.

[37] C. M. Reed, M. J. Doherty, L. D. Braida, and N. I. Durlach. 1982. Analytic study of the Tadoma method: Further experiments with inexperienced observers. *J. Speech Hear. Res.* 25 (1982), 216–223.

[38] G. Luzhnica, E. Veas, and V. Pammer. 2016. Skin reading: Encoding text in a 6-channel haptic display. In *Proceedings of the International Symposium on Wearable Computers (ISWC'16)*.

[39] G. Luzhnica and E. Veas. 2019. Background perception and comprehension of symbols conveyed through vibrotactile wearable displays. In *Proceedings of the 24th International Conference on Intelligent User Interfaces (IUI'19)*, 57–64.

[40] G. Luzhnica and E. Veas. 2019. Optimising encoding for vibrotactile skin reading. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. Article 235 (2019), 1–14.

[41] H. Z. Tan, S. Choi, F. W. Y. Lau, and F. Abnousi. 2020. Methodology for maximizing information transmission of haptic devices: A survey. *Proc. IEEE* 108, 6 (2020), 945–965.

[42] L. A. Jones. 2011. Tactile communication systems: Optimizing the display of information. *Prog. Brain Res.* 192 (2011), 113–128.

[43] M. A. Picheny, N. I. Durlach, and L. D. Braida. 1986. Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *J. Speech, Lang. Hear. Res.* 29, 4 (1986), 434–446.

[44] J. C. Krause and L. D. Braida. 2002. Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility. *J. Acoust. Soc. Amer.* 112, 5 (2002), 2165–2172.

[45] G. A. Gescheider, S. J. Bolanowski, and R. T. Verrillo. 1989. Vibrotactile masking: Effects of stimulus onset asynchrony and stimulus frequency. *J. Acoust. Soc. Amer.* 85 (1989), 2059–2064.

[46] G. Gescheider and N. Migel. 1995. Some temporal parameters in vibrotactile forward masking. *J. Acoust. Soc. Amer.* 98, 6 (1995), 3195–3199.

[47] G. A. Studebaker. 1985. A "rationalized" arcsine transform. *J. Speech, Lang. Hear. Res.* 28, 3 (1985), 455–462.

[48] H. Z. Tan, C. M. Reed, L. A. Delhorne, N. I. Durlach, and N. Wan. 2003. Temporal masking of multidimensional tactual stimuli. *J. Acoust. Soc. Amer.* 114, 6 (2003), 3295–3308.

[49] W. L. Bryan and N. Harter. 1899. Studies on the telegraphic language: The acquisition of a hierarchy of habits. *Psychol. Rev.* 6, 4 (1899), 345–375.

[50] H. Z. Tan, C. M. Reed, and N. I. Durlach. 2010. Optimum information-transfer rates for communication through haptic and other sensory modalities. *IEEE Trans. Haptics* 3, 2 (2010), 98–108.

[51] R. M. Uchanski and L. D. Braida. 1998. Effects of token variability on our ability to distinguish between vowels. *Percept. Psychophys.* 60, 4 (1998), 533–543.

Received August 2019; revised March 2021; accepted March 2021