



Article

Tactile Speech Communication: Reception of Words and Two-Way Messages through a Phoneme-Based Display

Jaehong Jung ¹, Charlotte M. Reed ² , Juan S. Martinez ¹ and Hong Z. Tan ^{1,*}

¹ Haptic Interface Research Laboratory, Elmore Family School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907, USA; jung137@purdue.edu (J.J.); mart1304@purdue.edu (J.S.M.)

² Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA; cmreed@mit.edu

* Correspondence: hongtan@purdue.edu

Abstract: The long-term goal of this research is the development of a stand-alone tactile device for the communication of speech for persons with profound sensory deficits as well as for applications for persons with intact hearing and vision. Studies were conducted with a phoneme-based tactile display of speech consisting of a 4-by-6 array of tactors worn on the dorsal and ventral surfaces of the forearm. Unique tactile signals were assigned to the 39 English phonemes. Study I consisted of training and testing on the identification of 4-phoneme words. Performance on a trained set of 100 words averaged 87% across the three participants and generalized well to a novel set of words (77%). Study II consisted of two-way messaging between two users of TAPS (TACTile Phonemic Sleeve) for 13 h over 45 days. The participants conversed with each other by inputting text that was translated into tactile phonemes sent over the device. Messages were identified with an accuracy of 73% correct in conjunction with 82% of the words. Although rates of communication were slow (roughly 1 message per minute), the results obtained with this ecologically valid procedure represent progress toward the goal of a stand-alone tactile device for speech communication.

Keywords: human haptics; tactile speech communication; phoneme codes; human performance; tactile word identification; tactile two-way messaging



Citation: Jung, J.; Reed, C.M.; Martinez, J.S.; Tan, H.Z. Tactile Speech Communication: Reception of Words and Two-Way Messages through a Phoneme-Based Display. *Virtual Worlds* **2024**, *3*, 184–207. <https://doi.org/10.3390/virtualworlds3020010>

Academic Editors: Panagiotis Kourtesis and Domna Banakou

Received: 7 February 2024

Revised: 31 March 2024

Accepted: 26 April 2024

Published: 7 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The present study was motivated by recent research on phoneme-based tactile speech communication systems conducted at Facebook/Meta [1–3], Rice University [4,5], McGill University [6–8], and a collaborative effort between Purdue University and MIT [9–14]. Our approach assumes that the front end of the device contains a module for producing a string of phonemes extracted from either the acoustic speech signal (using automatic speech recognition) or written text (using a text-to-speech converter). Although this research is motivated in part by the development of tactile aids for persons with profound sensory impairments of hearing and/or sight, the research cited here has been addressed primarily towards applications in situations where the normal sensory channels are temporarily compromised or unavailable.

This type of phoneme-based approach differs from much of the previous research on tactile speech-communication aids that has relied heavily on a vocoder-based approach (see reviews by Kirman [15]; Reed et al. [16,17]; Kappers & Plaisier [18]). Vocoder systems employ a frequency-to-place transformation of the acoustic speech signal in which location of stimulation on the skin corresponds to a given acoustic frequency region. One major difference between phoneme-based and spectral-based approaches lies in the manner in which the inherent variability in speech tokens within and across talkers [19] is handled. In the spectral-based vocoder systems, the burden of interpreting such variations is placed on the user of the device in the tactile domain. By contrast, in the phonemic-based approach, these variations are handled by the recognition system, and a fixed set of tactile signals can

then by assigned to the phonemes. These tactile signals are selected with knowledge of the perceptual characteristics of the tactile sensory system and can be tailored to yield high performance within tens of hours of training.

1.1. Background

The current study is an extension of research conducted by the Purdue University/MIT team using a 24-channel tactile display referred to as TAPS (TActile Phonemic Sleeve). The device consists of four rows of six tactors spaced between the wrist and elbow, with two rows applied to the dorsal surface and two rows to the ventral surface of the forearm. Unique tactile signals were created to represent 24 consonants (using position-based stimuli) and 15 vowels (using patterns of movement across the array). In a study using 10 consonants and vowels, Jung et al. demonstrated that, with 60 min of training, participants were able to achieve near-perfect identification of the individual tactile phonemes as well as a set of 51 words constructed from these 10 phonemes using a 300 msec inter-phoneme interval [9]. In a study of the identification of the full set of 39 phonemes by Reed et al. [11], mean performance of 85% correct was achieved after 1.5 to 4 h of training across the 10 participants. Two different types of training were compared for a task requiring the identification of 100 words composed of the 39 phonemes with an inter-phoneme interval of 300 ms [10,12]. One group of twelve participants was trained on a bottom-up approach in which they learned to identify the 39 individual phonemes before proceeding to the word identification task. Another different group of twelve participants was trained using a “top-down” approach in which they learned to identify words without initial exposure to isolated phonemes. Both sets of participants received 100 min of training spread out over 10 days. The best learners of both methods achieved scores > 90% correct; however, the phoneme-based approach led to greater success across participants in a shorter period of time than the word-based approach. The word-identification task was expanded to a vocabulary of 500 words in a study by Tan et al. [12]. Using an inter-phoneme interval of 150 ms for words consisting primarily of two and three phonemes, word recognition scores on the 500-word vocabulary averaged 71.9% correct within 4.5 to 8.0 h of training across 20 participants. Finally, the ability of trained users of TAPS to identify two-word phrases was tested by Reed et al. [14] who used an inter-phoneme interval of 150 ms and varied the inter-word interval between 500 and 2000 ms. Optimal performance of roughly 75% correct was achieved with an inter-word interval of 1000 ms, leading to an estimate of the effective transmission rate in the range of 30–35 words/min.

Word-identification studies have been conducted with other phoneme-based tactile systems, including the MISSIVE device developed at Rice University by Dunkelberger et al. [4,5,20]. This device, mounted on the forearm, consists of four vibratory channels, a radial squeeze band, and a lateral stretch rocker, that were used to encode 23 English phonemes. Following 100 min of training, participants were able to identify a closed set of 50 words (selected from a trained 150-word vocabulary) at a level of 87% correct using self-paced delivery of phonemes. With a new group of participants who received 160 min of training, identification of a set of 50 words using free form response was 67% correct.

Another phoneme-based display, developed at McGill University and referred to as WhatsHap [6–8], consists of two tactors on the forearm used to encode 25 English phonemes based on salient acoustic features for 16 consonants and speech synthesis for nine vowels. Word-identification studies included probes of the ability to generalize performance from a set of trained words to novel words. Following 100 min of training, open-set identification of words from a 150-word vocabulary was 51% correct and fell to 39% correct for novel words [6]. In a later study, de Vargas et al. [7] reported isolated word scores of roughly 50% correct for trained words compared to 45% correct for novel words in tests where the inter-phoneme interval was under the participant’s control. These authors also reported word scores for words presented in phrases where the advancement of words was under the participant’s control. In this case, words were identified with an accuracy of roughly

65% correct in phrases with lengths of two to six words per phrase, independent of the ratio of trained to novel words that was used to construct the phrases. Table 1 provides a summary of the aforementioned studies.

Table 1. Performance comparison of recent phoneme-based speech communication systems.

References	Learning Time (Hour)	No. of Tactors	No. of Phonemes	No. of Words Tested	Word Scores
[9]	1	24	10	51	≈100%
[11]	1.5–4	24	39	—	85%*
[10,12]	1.7	24	39	100	>90% (best)
[12]	4.5–8	24	39	500	71.9%
[14]	1.5–9.3	24	39	phrases	75%
[4,5,20]	1.7–2.7	4(+2)	23	50	67–87%
[6–8]	1.7	2	25	150	51%

* Phoneme score.

Studies concerned with learning words through the tactile sense have also been conducted with spectral-based displays. Brooks and colleagues [21–23] conducted long-term training studies with a 16-channel tactile vocoder. Of the two participants, one achieved 80% correct performance on a 70-word vocabulary following 40 h of training. The other achieved 80% correct performance on a 150-word vocabulary after 55 h of training, and then advanced to the acquisition of a 250-word vocabulary after 80 h of training. A further study with this vocoder was conducted on a large vocabulary of 1000 words. After nearly 200 h of training, the word score through the vocoder alone was 8.8% correct. Galvin et al. [24] studied generalization of word learning using the Tickle Talker device, consisting of eight electrodes attached to four fingers which are used to convey information about the formant regions of speech. Mean performance on 20 new untrained words was 22.5% (chance level near 0), which was about half the score of 42.5% obtained on 20 trained words following a mean training time of 20 h. Generalization to unfamiliar talkers was also tested, resulting in a drop in performance of roughly 30 percentage points for the new talkers (40% correct for unfamiliar compared to 68% for familiar talkers).

1.2. Motivation for Current Study

The results obtained with the TAPS system to date have demonstrated that individual phonemes, words, and two-word phrases can be conveyed to users of the tactile device with high levels of accuracy of reception within tens of hours of training time. This research is limited, however, by (1) the content of the vocabulary that has been studied and (2) the restriction of connected-speech reception tasks to a predetermined set of two-word phrases. The current paper extends our previous research through two studies reported here. Study I is concerned with the reception of 4-phoneme words through TAPS and Study II consists of exploration of two-way communication of multi-word messages with TAPS.

One limitation of previous work with TAPS is that the vocabularies in the word identification tasks consisted primarily of short words containing two or three phonemes. Of the 500-word vocabulary tested by Tan et al. [12], for example, 71.8% were 3-phoneme words, 17.8% were 2-phoneme words, 9.8% were 4-phoneme words, and the remaining 0.06% consisted of two 1-phoneme words and one 5-phoneme word. The goal of “Study I: Acquisition of Four-Phoneme Words” was to examine the reception of a new vocabulary consisting entirely of 4-phoneme words. In this experiment, users of the device received training on the identification of a set of 100 4-phoneme words over ten experimental sessions. Participants were then tested on their ability to generalize this learning to a different set of 100 4-phoneme words. Performance was assessed through percent-correct identification scores, analysis of error responses, and response times.

The ultimate goal in the development of a tactile speech communication system is the ability to receive connected speech at rates that approximate slow-to-normal reception

of speech through the normal auditory channel. The only demonstration of such ability through the tactile sense alone is provided by the Tadoma method of speech communication used by persons who are profoundly deaf and blind [25–27]. In this method, the Tadoma user places one or both hands over the face and neck of the talker to monitor articulatory movements and actions associated with the production of speech. Highly experienced Tadoma users are able to understand 80–85% of the key words in sentences produced at slow-to-normal speaking rates with an estimate of communication rates of the order of 12 bits/s [28], roughly half that of speech reception through audition. The goal of the second study conducted here was to extend the research on TAPS from single words and a fixed set of two-word phrases to the reception of connected discourse. In “Study II: Two-Way Messaging via TAPS”, two experienced users of TAPS communicated with each other through two identical systems. These two participants took turns typing in messages (that were translated into tactile phonemes presented through TAPS) and receiving them. Performance was measured in terms of accuracy of reception of messages and words, response time, and rates of communication. Our ultimate goal is to achieve accuracy and communication rates through TAPS comparable to those of Tadoma.

2. General Methods

This section describes the methods that are common to both studies. It includes information on the participants, hardware and software setup, intensity calibration procedures, and phonemic-based coding of English words. Information that is unique to each of the two studies is included in the subsequent sections.

2.1. Participants

Three young adults (P1, P2, P3) participated in Study I, and two (P1 and P3) continued in Study II. The participants provided informed consent through a protocol approved by the IRB at Purdue University. All participants were right-handed without any sensorimotor impairments. Their ages ranged from 23 (P3) to 26 (P1, P2). Participant P1 is a native Korean speaker who started learning English at the age of 8 years old, and also speaks German as a second language. P2 is a native Spanish speaker who began to learn English at the age of 8 years old. As young adults, P1 and P2 (both males) now speak English fluently. Participant P3 (female) is a native English speaker who also speaks Chinese as a second language.

All three participants are experienced users of the TAPS system through participation in earlier studies that included Jung et al. [9], Jiao et al. [10], Tan et al. [12], and Reed et al. [14]. As a result, they were already familiar with the haptic symbols for the 39 English phonemes and had acquired 500 English words in the Tan et al. study [12]. Each of the three participants had received roughly 20 h of experience with TAPS across these previous studies, thus providing them with preparation for the more difficult tasks of the current study.

Prior to the present study, the three participants took part in the TAPS word-identification study of Tan et al. (2020) [12]. P1, P2 and P3 were P05, P23 and P18 in Tan et al. (2020) [12], respectively. Figure 7 in Tan et al. (2020) [12] shows the equivalent number of words learned by these three experienced participants. Specifically, P05 (P1 in the present study) acquired 448 words after 50 min of learning time, P23 (P2 in the present study) acquired 433 words after 237 min, and P18 (P3 in the present study) acquired 469 words after 88 min. They were among the five best-performing participants who learned the most number of words in Tan et al. (2020) [12]. The time interval between Tan et al. (2020) [12] and the present study was 48 weeks for P1 and P3, and 77 weeks for P2.

2.2. Tactile Phonemic Sleeve (TAPS) System

The TAPS system consists of a 4-by-6 tactor array worn on the left forearm, as shown in Figure 1. This device was developed for use in previous studies at Purdue University and MIT, e.g., refs. [9–14]. There are six tactors in the longitudinal direction (elbow to wrist)

and four tactors in the transversal direction (ring around the forearm). As seen in Figure 1, the 24 tactors are arranged in six groups of four, with three clusters on both the dorsal and volar sides of the forearm. For use of the TAPS device, a spandex sleeve was first placed on the participant's left forearm for hygienic purposes. The participant then placed the left forearm on the lower half of the tactor array (rows iii and iv shown in Figure 1) with the volar side facing down, wrapped the upper half of the tactor array (rows i and ii) on top of the dorsal forearm, and fastened the gauntlet with Velcro straps, typically with the assistance of the experimenter.

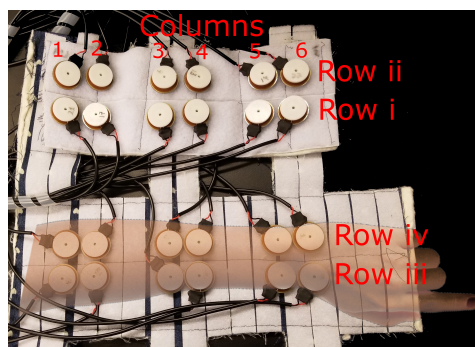


Figure 1. Layout of the tactor array of the TAPS system. As shown by the superimposed hand and forearm image, tactor rows iii and iv are placed on the volar (under) side of the forearm, and rows i and ii are on the dorsal (upper) side of the forearm. Reproduced from Figure 1 in Tan et al. (2020) [12].

The actuators were broadband voice-coil tactors (Tectonic Elements, Model TEAX13C02-8/RH, Part No. 297-214, sourced from Parts Express International, Inc.). A MOTU 24Ao audio device (MOTU, Cambridge, MA, USA) was used for delivering 24 channels of audio waveforms to the 24 tactors through custom-built stereo audio amplifiers. A Matlab program running on a desktop computer generated the 24-channel waveforms that were synchronously converted to 24 channels of analog signals by the MOTO device. With this setup, the 24 tactors could be driven independently with programmable waveforms.

During the experiments, the participant sat in front of a computer monitor. The forearm rested comfortably on the table with support at the wrist and elbow to avoid placing too much pressure on the tactors in rows iii and iv under the forearm.

2.3. Calibration of Perceived Intensity and Tactor Equalization

A two-step calibration procedure was used to control the perceived intensities of vibrotactile signals presented at different frequencies and locations on the forearm. First, detection thresholds at 60 Hz and 300 Hz were estimated for each participant with a reference tactor located in column 4, row ii in Figure 1. A three-interval, two-alternative, forced-choice, one-up two-down adaptive procedure was used for the measurement of thresholds for 70.7%-correct detection (see the review by Jones & Tan for a description of the psychophysical method [29]). This was followed by a tactor equalization procedure using the method of adjustment (see [29]). In this procedure, the signal on the reference tactor was a 300-Hz vibration with a level of -10 dB re maximum output. For each of the remaining 23 tactors, the participant adjusted the signal amplitude of the tactor until it felt equally strong as the signal at the reference tactor. This mapping was then used to adjust the level of each tactor so as to produce equal perceived strength across all tactors. A more detailed description of the calibration procedures is available in Reed et al. [11].

2.4. Haptic Symbols for Phonemes and Words

A phonemic-based approach was used to encode phonemes and words on the TAPS system after a survey of possible design choices for effective tactile speech communication systems (see a detailed description in Reed et al. [11] and the introduction in Tan et al. [12]). With this approach, 39 haptic symbols were designed, tested and revised for the 39 English phonemes that consist of 24 consonants and 15 vowels. Each symbol consists of vibrotactile

patterns using a prescribed subset of the 24 tactors in TAPS. The mapping of the phonemes and haptic symbols incorporates the articulatory features of the sounds of the English language, balanced by the need to maintain the distinctiveness of the 39 haptic symbols.

The stimulus properties included *amplitude* (in dB sensation level (SL), i.e., dB above individually measured detection thresholds), *frequency* (single or multiple sinusoidal components), *waveform* (sinusoids with or without modulation), *duration* (100 and 480 ms for short and long signals, respectively), *location* (place of stimulation along the TAPS array), *numerosity* (number of tactors turned on simultaneously or sequentially), and *movement* (smooth apparent motion or discrete saltatory motion varying in direction, spatial extent, and/or trajectory). Examples of the use of articulatory features to construct the phonemes include the use of location on the array to map place of articulation (e.g., front sounds are presented near the wrist and back sounds near the elbow), the use of unmodulated versus modulated waveforms to distinguish voiceless and voiced cognate pairs (i.e., vibrotactile modulation was used to encode vocal-fold vibration), and the use of short and long signal durations for distinguishing brief plosive bursts from longer fricative noises, respectively.

To further differentiate consonants and vowels, all haptic symbols for consonants occur at distinct locations on the forearm, and those for vowels involve several different types of illusory movements (e.g., a rumbling sensation of movement from the wrist to the elbow for the vowel sound /u/ in “moose”). Further details of the phoneme-to-symbol mapping strategies and the resultant haptic symbols can be found in Reed et al. [11]. The phoneme codes are as described in Tables 1 and 2 of Reed et al. [11] with the exception that the duration of the six plosive phonemes was increased from 100 to 140 ms and the duration of the 11 vowels and diphthongs that were previously 480 ms was reduced to 400 ms.

3. Study I: Acquisition of Four-Phoneme Words

Study I investigated the learning of 4-phoneme words with TAPS. This study was conducted to extend our work on word recognition beyond the 2- and 3-phoneme words that made up the bulk of the vocabularies in our previous studies [10–12]. Therefore, we used 4-phoneme words to probe the ability of participants to process “longer” words as would be required for a practical communication system.

3.1. Learning Materials

Two 100-word lists consisting entirely of 4-phoneme words were created as the new learning materials for Study I. The two-hundred 4-phoneme words were selected from the vocabulary used in the CUNY sentences [30]. The words were selected from these materials because they employ vocabulary representative of conversational speech across twelve different topic areas. These words were then randomly assigned to the two 100-word lists. Tables 2 and 3 show the two sets of 100 words in List #1 and List #2, respectively. The participants practiced with List #1 and were tested for word recognition afterwards. They were then tested with List #2, without any practice, for their ability to generalize learning of 4-phoneme words. In terms of vowel (V) and consonant (C) composition, each list consisted of 31%, 31%, 22%, and 6% of CVCC, CCVC, CVCV and VCVC words, respectively. The remaining 10% of the words in each list were composed of VCCV (4%), CVVC (3%), VCCC (1%), VCVV (1%), and CCVV (1%) words.

Each word was first transcribed into its corresponding phoneme sequence. To display an English word on TAPS, the haptic symbols corresponding to the phonemes making up the word were delivered to the tactor array in a sequential order, with an inter-phoneme interval inserted between successive phonemes. This interval was set to 150 ms throughout the study.

Prior to working with the two lists of 4-phoneme words, the three experienced participants reviewed and were tested on the recognition of the 39 haptic symbols for the 39 phonemes, and reviewed and were tested on a 500-word list to which they had been exposed in an earlier study by Tan et al. [12]. Of the 500 words in the review list, the

number of words (and the corresponding percentages) with 1, 2, 3, 4 and 5 phonemes were 2 (0.4%), 89 (17.8%), 359 (71.8%), 49 (9.8%) and 1 (0.2%), respectively. After the review of the 39 phonemes and the 500-word list over a period of three 10- to 20-min sessions across three days, the participants started the experiment with the 4-phoneme words.

Table 2. The 100 four-phoneme words in List #1.

List #1				
able	clean	first	mirage	steak
after	clothe	flower	moved	stock
apple	color	fourth	movie	store
aunts	cooked	fried	often	summer
baby	cost	from	only	sweet
band	cover	glove	oven	tailor
belt	crack	golf	paper	tank
berry	cream	green	park	teacher
black	cruise	guest	plan	tent
block	diet	hard	puppy	throat
bond	dollar	honey	quit	train
borrow	dress	huge	ranch	trick
break	drop	kids	rest	trip
bring	ducks	last	sand	turkey
brush	easier	later	silk	under
butter	Easter	lend	singer	water
card	eaten	lion	skirt	weeks
change	even	lunch	slice	wind
chimp	fever	major	slope	years
city	field	milk	snake	zipper

Table 3. The 100 four-phoneme words in List #2.

List #2				
about	cloth	fresh	neighbor	soccer
alter	coffee	fries	office	sold
asked	coins	fruit	onto	stall
avoid	cookie	garage	open	still
ballet	crash	gift	paint	stop
bank	cross	great	pairs	style
bark	cute	grill	parade	sunny
best	desk	guest	past	swim
better	dinner	heavy	pepper	tax
bleach	doing	horse	place	think
blood	drive	into	press	toast
boots	dryer	July	ready	today
breathe	eating	least	rent	track
bridge	enjoy	letter	risk	treat
broke	enough	lights	shirts	trim
busy	every	liver	shutter	twice
cats	fast	lost	skate	washer
child	fatty	money	sleeve	weather
chilly	find	month	slip	wrist
clear	fix	must	smoke	yard

3.2. Procedures

The main experiment was conducted over 12 days, with practice limited to 10 to 20 min per day to avoid fatigue. This was followed by daily testing to assess the participants' performance. Our past research has shown evidence of memory consolidation. According to the theory of memory consolidation [31], and following the practice of our previous studies, it is more efficient and effective to spread learning over a period of time with

concentrated practice time each day. For the first 10 days of the study, participants practiced with and were tested on words from List #1. The final two days were spent on generalization tests with words from List #2. The participants wore noise-reduction earphones to block any sounds from the factors.

The participants performed three types of tasks: (i) free play by presenting words through TAPS, (ii) a word identification task with correct-answer feedback, and (iii) word identification without feedback. The purpose of the first two tasks was to provide the participants with training on the words in List #1. During free play, the participant was able to select any word for presentation on TAPS. The word identification tasks were conducted using a one-interval identification paradigm where words from one of the lists were selected at random with replacement. The participant's task was to enter a response by typing a word on a keyboard. During the practice phase, trial-by-trial correct-answer feedback was provided. During the testing phase, the same paradigm was employed except that correct-answer feedback was not provided. The participants were instructed to respond as quickly and as accurately as possible. Response times (RT) were measured on each trial, defined as the duration between the end of a stimulus presentation and the start of the first key-down event of a typed response.

An overview of the tasks performed by the participants on each day is shown below, followed by a detailed description of each day's activities.

- Day 1: Practice with List #1 (20 min).
- Days 2 to 4: Practice with List #1 (10 min) and test (25 trials/block \times 2 blocks).
- Days 5 to 10: Practice with List 1 (10 min) and test (25 trials/block \times 3 blocks).
- Days 11 & 12: Test with List #2 (25 trials/block \times 4 blocks) without practice.

On Day 1, participants practiced with the 100 4-phoneme words in List #1 (see Table 2), spending 20 min on learning in the form of free play and/or practice testing. No word recognition test was conducted on Day 1.

From Day 2 to Day 4, the participants spent 10 min each day with the practice of List #1 and then performed a word recognition test without feedback in two blocks of 25 trials each. During the practice tests, the list of 100 words was shown on the computer screen and the participant was instructed to select one of the words as the response (that was entered via typing). During the tests with no feedback, however, the participants were not given access to the 100-word list and were free to type any word into a text box on the computer screen as their response.

Starting on Day 5, the participants again spent 10 min each day practicing with List #1. The number of word recognition tests without feedback increased to three blocks of 25 trials. The same procedure continued until Day 10.

On Days 11 and 12, the ability to generalize learning on List #1 to the new set of words in List #2 was examined. With no opportunity for practice, word identification testing without feedback was conducted with the 100 words in List #2 (see Table 3). Each participant was tested on four blocks of 25 trials each day.

At the conclusion of the 12 days of the experiment, each participant had completed a total of 600 trials with List #1 (24 25-trial runs) and 200 trials with List #2 (8 25-trial runs) of word recognition testing without correct-answer feedback.

3.3. Data Analysis

Starting on Day 2, the response time (RT) and responses from word recognition tests without feedback were logged on each trial. The responses entered by the participants were compared with the words presented, and homophones (i.e., words that are phonetically identical but are spelled differently and have different meanings) were accepted as correct answers. The percent-correct (PC) scores for words were then computed. Four dependent measures were calculated from the recorded data: (1) word PC score vs. Day; (2) number of phoneme errors per word; (3) phoneme errors vs. phoneme position in a word, and (4) RT vs. Day.

PC scores for words were plotted as a function of test day, to observe any learning trend. *T*-tests were performed on the arcsine-transformed PC scores between Days 10 and 11 and between Days 11 and 12 for each of the three participants. A Welch's unequal variance *t*-test was performed using the Welch–Satterthwaite approximation for obtaining degrees of freedom.

Further analyses of the incorrect responses were conducted to gain insight into the processing of the phoneme-based tactile words, and are limited to the data provided by the three participants studied here. To conduct these analyses, a phonetic transcription of the error response was compared with that of the stimulus word. Errors at the phoneme level were analyzed in two ways. First, the number of phonemes in error was calculated for the words in List #1 (combined over Days 2 to 10) and List #2 (combined over Days 11 and 12). Second, the number of phoneme errors was analyzed as a function of their position in each word. Again, the data from Day 2 to Day 10 for List #1 and those on Days 11 and 12 for List #2 were processed separately. To check if the observed error distributions were due to chance, χ^2 tests were performed for the analysis of both types of phonemic errors and the effect size was assessed using Cohen's *d*. These analyses were conducted on errors aggregated across the three participants to provide a general description and summary of the error patterns that occurred in the data.

The RTs recorded on the same day were processed to create box plots of RT as a function of test day, from Day 2 to Day 12. Data points more than 1.5 times the inter-quartile range above the third quartile or below the first quartile were regarded as outliers (where inter-quartile range is defined as the third quartile minus the first quartile). To examine the change in RT over the training period encompassing Day 2 to Day 10, a linear regression was derived from the RT functions for each participant and these slopes were compared to a slope of zero using a *t*-test. *t*-tests were also performed to examine changes in the size of the median RT between Day 10 and Day 11 and between Day 11 and Day 12 for individual participants. The effect size was assessed using Cohen's *d*.

3.4. Results

3.4.1. Word PC Scores

Figure 2 shows the word recognition PC scores on each day, separately for the three participants. The total amount of time spent in training and testing across the 12 days of the study was 2.8 h for P1, 4.7 h for P2, and 3.7 h for P3, respectively. Shown are the results for List #1 collected from Day 2 to Day 10 (open symbols) and those for List #2 collected on Days 11 and 12 (filled symbols). Different patterns of learning over the course of training on List #1 were observed across the three participants. P1, who began with a score of 62% correct on Day 1, showed improvement over the first five days of training and achieved asymptotic scores of roughly 80% correct over the final four days. P2 also began with an initial score of 60% correct but reached a score of 80% correct by Day 3, with further improvement to 90% correct by Day 10. P3 began with a score of 80% correct on Day 1, achieved a score of 90% correct on Day 3, and asymptoted at this score over the final four days of testing. Participant P1 achieved an average PC score of $74.7 \pm 10.4\%$ for List #1 and $80.5 \pm 11.2\%$ for List #2. Data for P2 show average PC scores of $78.3 \pm 13.2\%$ and $69.5 \pm 11.7\%$ for List #1 and List #2, respectively. P3 achieved the highest average word PC score of $87.7 \pm 8.3\%$ for List #1 and the lowest average PC score of $67.0 \pm 10.2\%$ for List #2.

As described previously, the participants selected their responses from a closed set of the 100 words in List #1 on the practice tests, but not during the testing without feedback, where responses were made from an open set. Therefore, the chance level from Day 2 to Day 10 on tests without feedback ranged between 1% (1 out of 100), assuming that the participants had memorized all 100 words to 0% if the participants did not remember any. The number of error responses to List #1 that were not in the word list (excluding homonyms) were: 43 out of 148 incorrect responses (29.1%) for P1, 76 out of 126 (60.3%) for P2, and 32 out of 74 incorrect responses (43.2%) for P3. Averaged across participants, incorrect responses consisting of words not contained in List #1 were less frequent than the

use of words within the list (43% versus 57%). Thus, participants relied somewhat heavily on their knowledge of the words in the list when making responses.

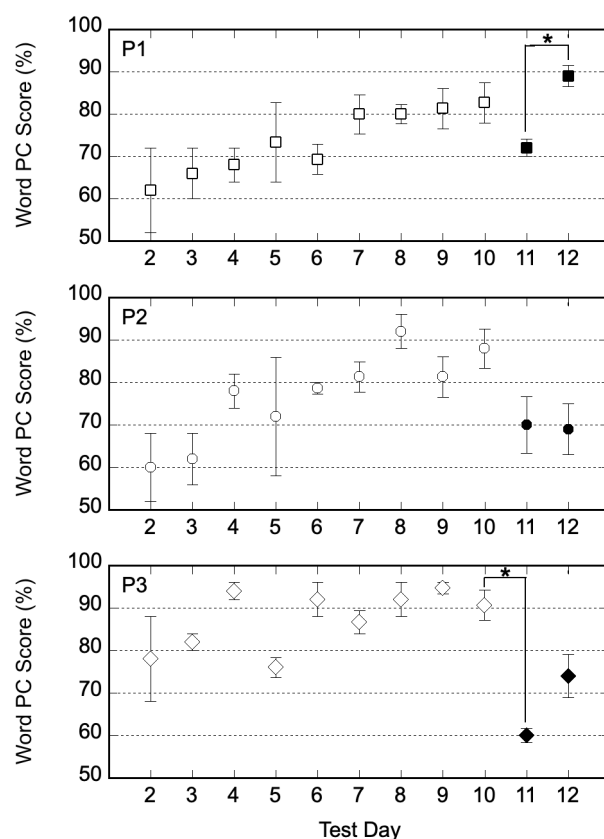


Figure 2. Percent-correct scores for the recognition of 4-phoneme words on each test day. Data for the three participants are shown in separate panels. Error bars indicate standard errors. Open symbols represent data for List #1 and filled symbols for List #2. An asterisk indicates statistical significance.

Because the words in List #2 were never shown to the participants, the chance level for Day 11 and Day 12 was 0%. It is clear from Figure 2 that the participants' word recognition percent-correct scores were well above the chance levels for both List #1 and List #2, demonstrating that the participants were able to decode phonemes sequentially in order to identify novel 4-phoneme words.

It was expected that a decrease in PC score from Day 10 to Day 11 might occur when the test materials changed from the learned List #1 to the new List #2. A significant decrease in PC scores from Day 10 to Day 11 was observed for participant P3 ($t(2) = 6.081, p = 0.018$) but not for either P1 ($t(3) = 1.683, p = 0.180$) or P2 ($t(5) = 2.120, p = 0.089$). The effect for P3 was large (Cohen's $d = 5.37$). It is further observed that participant P1 showed a significant increase in percent-correct score from Day 11 to Day 12 ($t(6) = 3.320, p = 0.018$) while P2 ($t(6) = -0.137, p = 0.896$) and P3 ($t(4) = 2.612, p = 0.068$) did not. The effect for P1 was large (Cohen's $d = 2.34$).

Further analyses of the errors made on the word-identification tests without feedback are provided below, where data have been pooled across the three participants.

3.4.2. Number of Phoneme Errors per Word

In the left column of Figure 3, the percentage of errors is plotted as a function of the number of phonemes in error (from 1 to 4), with the number of errors shown in parentheses. The results for List #1 (top) and List #2 (bottom) show different trends. For the words in List #1 there was a gradual increase in errors from 1 to 3 phonemes, followed by a sudden drop in 4-phoneme errors. For the words in List #2, there were more 1-phoneme and 2-phoneme errors than 3-phoneme errors, with few 4-phoneme errors. The results

of a chi-square test confirmed that the distribution of errors for both List #1 and List #2 was not uniform across the four error categories (List #1: $\chi^2(3, 348) = 64.989, p < 0.001$; List #2: $\chi^2(3, 163) = 65.859, p < 0.001$). The residuals for 1, 2, 3, and 4 phoneme errors, respectively, were $-1.90, -1.60, 5.55$, and -5.20 for List #1 and $4.58, 3.32, -2.15$, and -5.75 for List #2. For List #1, the residuals indicate that the effect arose from a greater number of 3-phoneme errors and fewer 4-phoneme errors. For List #2, the effects arose from fewer 3-phoneme and 4-phoneme errors in conjunction with greater than expected 1-phoneme and 2-phoneme errors.

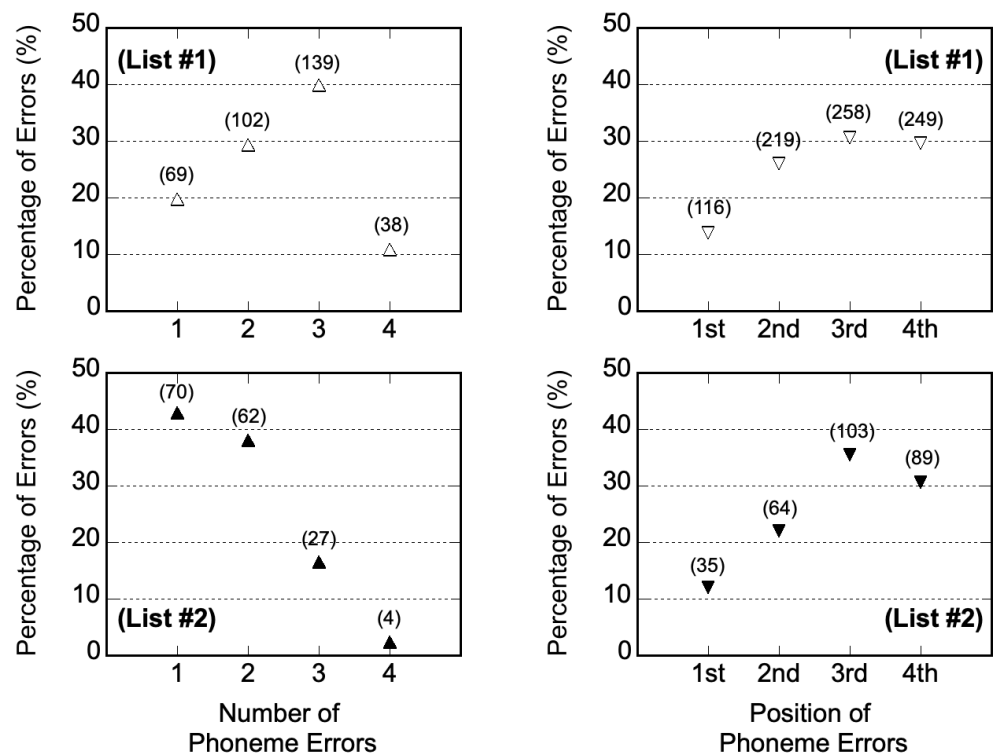


Figure 3. Percentage of errors by number of misidentified phonemes (left) and by phoneme positions (right) in a word. Results for List #1 and List #2 are shown on the top and bottom, respectively. Also shown are the corresponding error counts in parentheses.

3.4.3. Phoneme Errors vs. Position in a Word

The incorrect responses were then analyzed by looking at the positions of the phonemes that were incorrectly identified. Shown in the right column of Figure 3 are the proportion of phoneme errors at each of the four phoneme positions over the total number of phoneme errors (with the number of errors at each position shown in parentheses). The plots for List #1 on the top and List #2 at the bottom show similar patterns: phoneme errors, both in percentage and in number, increased from the first position to the third position, followed by a slight drop at the fourth position. A chi-square test confirmed that phoneme errors depended on the phoneme position within a word (List #1: $\chi^2(3, 842) = 60.527, p < 0.001$; List #2: $\chi^2(3, 291) = 36.849, p < 0.001$). The residuals for errors in the 1st, 2nd, 3rd, and 4th positions in the word, respectively, were $-6.51, 0.58, 3.27$, and 2.65 for List #1 and $-4.42, -1.02, 3.55$, and 1.90 for List #2. Thus, for both List #1 and List #2, the chi-square effects arose from fewer errors on the first phoneme and a greater number of errors on the third and fourth phoneme positions.

3.4.4. Reaction Time

RT as a function of test day is shown as box plots in Figure 4, separately for each participant. A trend was observed for a decrease in the median RT and variance from Day 2 to Day 10 for each of the participants. The median RT for P1 decreased from 2.55 s on

Day 2 to 0.45 s on Day 10, for P2 from 3.18 s to 1.06 s, and for P3 from 1.28 to 0.66 s. T-tests conducted on the slopes of the functions were significant for P1 ($t(7) = -6.72, p < 0.001$), P2 ($t(7) = -8.81, p < 0.001$), and P3 ($t(7) = -3.33, p = 0.013$). These results show that in general, RTs improved (became smaller) throughout the 9 test days with List #1 as the participants became more proficient at the task. These effects were large as evidenced by Cohen's $d = -5.08, -6.66$, and -7.52 for P1, P2, and P3, respectively. To determine if there was a change in median RT from Day 10 to Day 11 with the introduction of List #2, t -tests were conducted using a Welch two-sample t -test with unequal variance for P1 and P2 and equal variance for P3. These tests indicated significance for P2 ($t(324.04) = -2.89, p = 0.004$, showing a negligible effect with Cohen's $d = 0.12$) but not for P1 ($t(325.57) = -1.71, p = 0.088$) or P3 ($t(348) = -1.49, p = 0.138$). Finally, t -tests were conducted to determine if there was a change in median RT from Day 11 to Day 12 (using a Welch two-sample t -test with equal variance for P1 and unequal variance for P2 and P3). No significant difference was found between RT on Day 11 and Day 12 for any of the participants: P1 ($t(398) = -0.60, p = 0.725$), P2 ($t(390.54) = 0.46, p = 0.323$), and P3 ($t(392.4) = -0.77, p = 0.780$).

Comparing the PC scores in Figure 2 with the RT data in Figure 4, a general trend is observed for a decrease in response time with an increase in percent-correct score.

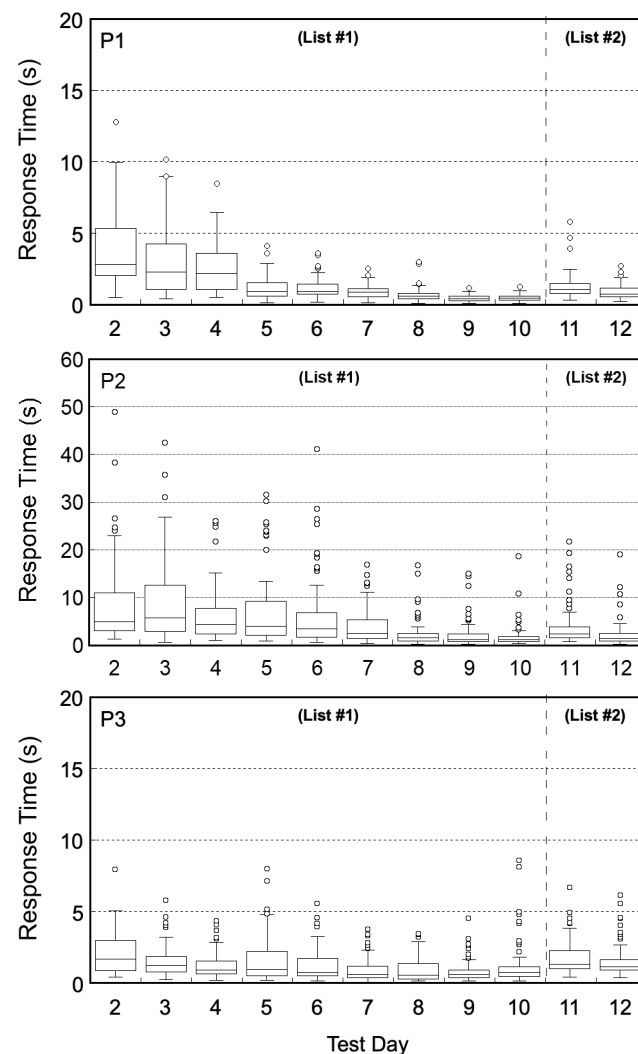


Figure 4. Box plots of daily RTs for participants P1 (top), P2 (middle), and P3 (bottom). Note the different ordinate scales on the three plots. Outliers are indicated by open circles.

3.5. Discussion

In this study with the TAPS device, the ability of three experienced participants to identify a set of 100 four-phoneme words improved from $66.67 \pm 9.87\%$ correct to $87.11 \pm 4.07\%$ within 110 min of training. Furthermore, this ability was generalized to a new set of untrained 100 four-phoneme words where the mean score across participants was $67.33 \pm 6.43\%$ correct on the first day of testing and $77.33 \pm 10.41\%$ correct on the second day of testing. Thus, participants were clearly performing well above the chance level of 1% (for the trained list) or near 0% (for the untrained list), indicating an ability to transfer learning from the practiced words to the novel words.

The participants showed a decreasing trend in response time with a concurrent increasing trend in word PC scores. By the end of the training, RTs were of the order of roughly 0.5 to 1 s across participants. These RTs compare quite favorably to those reported in previous studies with TAPS. For example, Reed et al. [14] reported RTs of roughly 6 s for identification of a closed set of 100 words with an IPI (inter-phoneme interval) of 150 ms as employed in the current study. The faster RTs reported here compared to those of [14] are likely due to the greater experience of the current group of observers with the TAPS device compared to those in our earlier study. Prior to their extended training in the current study, these observers had all received previous training with TAPS on the acquisition of a 500-word vocabulary [12].

For both the trained and novel words, participants appear to have processed the phonemes in order of presentation, given that participants were more successful at recognizing the first two phonemes in a word compared to the final two phonemes. The pattern observed for percentage of errors as a function of the number of phonemes in error, however, was markedly different across the two word lists. For the practiced words in List #1, participants were most likely to make an error in which three phonemes were incorrect, while for the novel words in List #2, they were most likely to have an error response with only one incorrect phoneme. This suggests that the participants used different strategies to identify practiced and novel words. For the practiced words, they appear to have used their knowledge of the test vocabulary to select one of the words in the list based primarily on recognition of the first phoneme, thus being most likely to have three phonemes in error. On the novel words, however, participants were much more likely to have only one or two phonemes in error than three or four phonemes. In this case, the test vocabulary was unknown and participants appear to have processed multiple phonemes before formulating a response.

The error patterns of the present study showed a different trend compared to those of Reed et al. (2021) [14] which focused primarily on the identification of two- and three-phoneme words. In that study, no evidence was found for the dependence of error rate on either number or position of phonemes in error. This difference in results might have been due to the change in the length of words. Compared to words with 2 or 3 phonemes, 4-phoneme words have a longer duration which required a greater amount of memory. The participants might have preferred to concentrate more on identifying the initial phoneme and deduce the words while receiving other phonemes rather than focusing equally on receiving all the four phonemes which demanded a higher cognitive load.

The word scores obtained in the present study with four-word phonemes are comparable to those obtained in previous word-acquisition studies with TAPS using vocabularies consisting primarily of two- and three-phoneme words. With 100-word vocabularies and training times of the order of 1.5 to 2 h, scores ranged from 80% correct [10] to 89% correct [14]. The results obtained here show good generalizability from trained words to a novel set of words. This transfer of learning has also been demonstrated in studies with other tactile speech devices. Using the WhatsHap device, de Vargas et al. [6] showed that performance was well above chance levels when new words were introduced into the vocabulary. In studies with the Tickle Talker, Galvin et al. [24] also demonstrated some carry-over in performance from trained to untrained vocabulary.

The present study supports the conclusion that users of the TAPS system can learn longer words within tens of hours and generalize this ability to a novel set of four-phoneme words. These conclusions are limited, however, to the performance of three participants for words with a length of four phonemes. Further study is necessary to determine the generalizability of these results to a larger set of users, and to determine whether longer words can be learned and acquired with high accuracy within a similar period of training. It is also possible that users might employ different strategies in the recognition of words containing five or more phonemes. For example, rather than decoding a long word phoneme-by-phoneme, practiced users may begin to employ chunking to recognize unique haptic patterns in longer signals. Although the current study represents progress in the use of TAPS, users would obviously need to be able to process words of any length for the reception of messages in a practical tactile communication system.

4. Study II: Two-Way Messaging via TAPS

Study II investigated the ability to conduct two-way communication using two identical TAPS systems worn by two experienced participants (P1 and P3). The results of Study II shed light on the feasibility of TAPS for tactile communication of spontaneous speech, which is the ultimate goal of any device-mediated tactile speech communication system.

4.1. Experimental Setup

Two identical stations were set up on two adjacent tables. Each station consisted of a personal computer, a monitor, and a complete TAPS system (see Figure 5). A TCP server-client protocol was implemented to enable communication between the two PCs. No earphones were used in Study II. The participants relied on verbal communication to start and end the experimental sessions, and needed to hear each other's voices. Taking the earphones on and off with their hands may inadvertently shift the tactors inside the TAPS systems.



Figure 5. Photo of two-way messaging via TAPS. Each participant wore a TAPS system on the left forearm to receive text messages. They recorded received messages and sent new messages on computer keyboards.

To enable automatic transcription of words into phoneme streams, a text-to-speech (TTS) front-end was added to the TAPS system. The TTS front-end was adapted from the FLITE system, an open source system intended for speech synthesis on small embedded machines [32]. In its typical usage, FLITE converts text input into a phonemic transcription that is fed into a speech synthesizer. In our case, the phonemic transcription was converted into the corresponding haptic symbols that served as the input to the TAPS system.

Another modification was the use of a haptic signal that primed the participant's attention prior to the transmission of each text message. The 1.2 s “knocking” signal designed by Shim & Tan (2020) [33], consisting of short pulses of superimposed 30 Hz and 300 Hz vibrations that feels distinctly different from the phonemic haptic symbols, was used for this purpose.

Two temporal parameters were set to control for the time between phonemes and the time between words. An inter-phoneme interval of 150 ms was used in Study II, as in Study I. The inter-word interval was initially set to 300 ms for the first ten days of the experiment, but was then increased to 500 ms for the remainder of the experiment.

4.2. Procedures

Between 18 March 2019 and 24 February 2020, participants P1 and P3 conducted the two-way messaging experiment on 45 days during the Spring and Fall semesters in 2019 and the beginning weeks of Spring semester of 2020. On the first day of testing, participants spent 47 min becoming familiar with the user interface for sending text messages and recording received messages. For the next 44 days of testing, the daily time ranged between 8 and 31 min (average: 16 ± 6 min). There was no specific time requirement per day. Instead, the participants met when both were available and worked together until they felt tired or lost concentration. An overview of the time spent per month over the course of the experiment is shown in Figure 6. There is a visible gap during the summer months of 2019, and a decrease in activities during November and December of 2019 due to holidays and winter break.

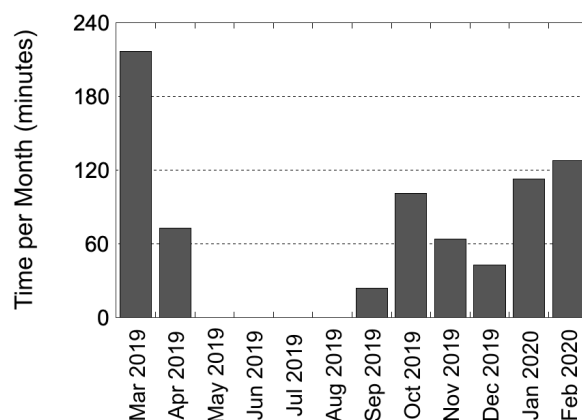


Figure 6. Testing time per calendar month.

On each testing day, the participants started by agreeing on a general topic for their conversation. The vocabulary was completely open: any English word could be used, and any number of words could be included. One participant, the “sender”, would start by typing a message into a text box in the user interface, and press a “Send” button when finished. The other participant, the “receiver”, would feel the knocking signal on TAPS, followed by the haptic symbols corresponding to the sequence of phonemes transcribed from the typed text message. A “Replay last” button was available if the receiver wanted to feel the message again before entering a response. There was no limit to how many times the replay button could be used. The receiver was required to type the received message into a text box so it could be recorded and analyzed later. After the response was entered, the receiver could click on a “Show” button to check if the message was received correctly. The receiver then becomes the sender and initiates the next message. As would happen with any natural conversation, the participants sometimes changed the topic in the midst of their chat and continued with the experiment. An example dialogue is shown in Table 4 along with the number of times each message was presented before a response was made.

Table 4. An example dialogue recorded from Study II, and the number of times a text message was presented prior to response.

Sender	Message	No. of Presentations
P1	how is your project	1
P3	should start testing	3
P1	how many participants	2
P3	I think five	2
P1	how long does it take	1
P3	I think three days	4
P1	starting this week	3
P3	yes how about you	1
P1	next week probably	2

4.3. Data Analysis

The details of the two-way conversation in terms of time stamps per message, number of repeats per message, total time spent on each day, the text messages sent, and the text messages received were recorded for the 45 experiment days. Figure 7 illustrates the four time stamps recorded for each text message: start time of the “knocking” signal, start time of a message being sent to TAPS, end time of the message transmission, and the first key-down event for recording the received message. The time interval between the knock-knock signal and the first haptic word was equivalent to the inter-phoneme interval (150 ms). As shown in Figure 7, response time was defined as the time between the end of a transmitted message and the first stroke of recording the message. If the participant replayed any messages on TAPS, that time was included in the response time.

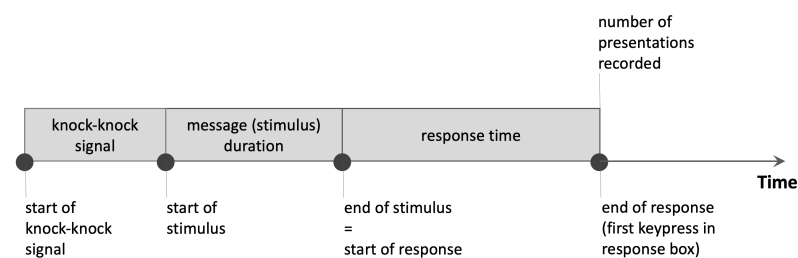


Figure 7. Timing diagram illustrating recorded time stamps (black dots).

By comparing the text messages sent and received, the percent-correct (PC) scores for messages or words, PC_{msg} and PC_{word} , were computed. Communication rates in terms of messages/min or words/min $MsgPM$ and WPM , were also calculated. Due to the small number of entire messages transmitted per day, the results for each of these measures were averaged over 5-day periods without overlap (resulting in nine 5-day periods across the 45 test days, e.g., Days 1 to 5, Days 6 to 10, etc.) to show trends more clearly. The summer break in testing occurred between Days 6–10 and Days 11–15, that is, Day 10 was the last day of testing in the spring semester of 2019 whereupon testing was resumed on Day 11 in the fall semester of 2019. The following paragraphs detail the calculations of the PC scores and communication rates.

4.3.1. PC for Messages (PC_{msg})

For each transmitted message, the receiver’s response was compared with the sender’s original message. A score of 1 was assigned if the response was identical to the sent message. Otherwise, a score of 0 was assigned. The analysis was performed on all messages regardless of who was the sender or receiver. The daily PC_{msg} was calculated by dividing the number of correctly received messages by the total number of messages transmitted on each day.

4.3.2. PC for Words (PC_{word})

It was recognized that successful communication could be accomplished even if not all the words were correctly received within a message. Therefore, PC scores for words were also calculated. For each transmitted message, the PC_{word} score was calculated by dividing the number of correctly received *words* by the total number of *words* transmitted. The daily PC_{word} score was then obtained by averaging the PC_{word} scores for the same day. We expected that the PC_{word} score would increase as time went on.

4.3.3. Communication Rates in Messages per Minute ($MsgPM$) and Words per Minute (WPM)

Similar to PC scores, communication rates were calculated based on messages or words. The $MsgPM$ score was computed by dividing the number of correctly received messages per day by the total time in minutes for the day. The total time per day was calculated by summing the intervals between the start and end of each experimental session, over all experimental sessions and both participants on a day. The communication rate in words per minute (WPM) was calculated as the total number of correctly received words on each day divided by the total time in minutes for the day.

To examine the change in PC_{msg} and PC_{word} scores over time, a linear regression was derived and its slope was compared to a slope of zero using a *t*-test. To conduct the *t*-test on PC scores, the rationalized arcsine transformation of PC_{msg} and PC_{word} scores was used [34,35].

4.4. Results

The participants each spent a total of 12.7 h (762 min) over the 45 days of this experiment, leading to a total of 732 messages with a mean number of 16.3 messages \pm 5.7 transmitted per day. The total number of messages for each of the non-overlapping 5-day periods used in the data summaries ranged from 58 (Days 16 to 20) to 121 (Days 6–10), with a mean of 81.3 \pm 22.5. The mean number of times a message was presented before a response was initiated was highly consistent throughout the study. Within the 5-day periods, the mean number of repetitions ranged between 2.2 (Days 1–5 and 6–10) and 3.0 (Days 21–25) with an overall mean of 2.6 \pm 0.3.

The results of PC_{msg} and PC_{word} scores averaged over 5-day periods for the two participants are shown in Figure 8. Over the 45 days, the average PC_{msg} was 69.0% \pm 18.9% for P1 and 77.8% \pm 21.1% for P3. The data show a great deal of fluctuation in scores over the duration of the testing. However, the overall results indicate that the two participants performed well above chance level, considering the fact that this was truly an “open vocabulary” task. The minimum five-day average score for P1 (60.0%) and P3 (64.1%) was significantly above a chance level of essentially 0%.

In Figure 8 (upper left panel), it can be seen that P1's PC_{msg} scores started and ended at similar levels (hovering around 70% correct), with a slight dip (to roughly 60%) across the range of Days 16 to 30. The slope of the linear regression compared to a slope of 0 is not significant ($t(43) = 0.34$, $p = 0.734$). In Figure 8 (lower left panel), the PC_{msg} scores for P3 show a slight improving trend from Days 1–15 with a dip around Days 16–25 and a smaller dip around Days 31–35. A *t*-test comparing the slope with 0 indicates that the linear trend is significant ($t(43) = 2.71$, $p = 0.010$), suggesting that P3's ability to communicate through TAPS as measured by the PC_{msg} scores improved over time. This effect was large as evidenced by Cohen's $d = 0.83$.

The PC_{word} scores show very similar patterns as the PC_{msg} scores. The average PC_{word} of P1 was 78.2% \pm 13.3% whereas that of P3 was 85.2% \pm 14.4%. Figure 8 shows the 5-day running averages for P1 (upper right panel) and P3 (lower right panel), respectively. Again, the slope of the regression line fit to P1's data was not significantly different from 0 ($t(43) = 0.51$, $p = 0.609$), whereas the data for P3 in Figure 8 show an increasing trend which reached significance ($t(43) = 2.51$, $p = 0.016$), with a medium effect as evidenced by Cohen's $d = 0.76$.

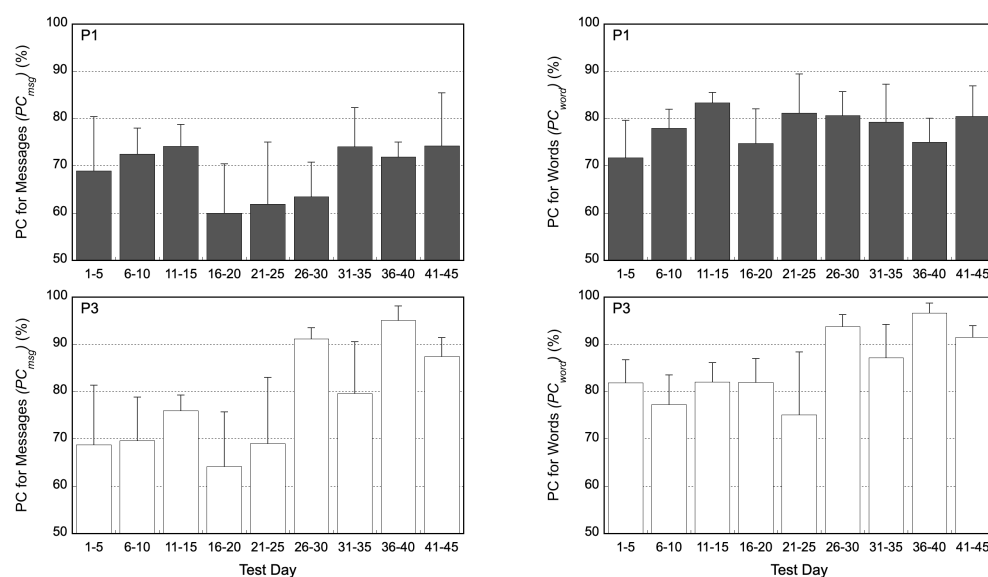


Figure 8. The two panels on the left show PC for messages (PC_{msg}) as a function of Test Day (binned into consecutive 5-day periods) for P1 (**upper panel**) and P3 (**lower panel**). The two panels on the right show PC for words (PC_{word}) as a function of Test Day for P1 (**upper panel**) and P3 (**lower panel**). The error bars represent standard errors.

The communication rates in messages per minute ($MsgPM$) were analyzed next. The average communication rates were calculated for the entire conversation on a day rather than for individual participants or individual messages. The 5-day averages are plotted in the top panel of Figure 9 to smooth the daily fluctuations in $MsgPM$. The minimum and maximum communication rates over the 5-day averages were roughly 0.5 (Days 1–5, 16–20, and 21–25) and 1.2 $MsgPM$ (Days 36–40), with an overall average of 0.784 ± 0.312 $MsgPM$, respectively. It can be seen that the communication rates in messages per minute have a pattern that is strikingly similar to those of the PC_{msg} and PC_{word} scores. An increasing trend in the data was significant as indicated by a t -test comparing the slope of the regression line with a slope of 0 ($t(43) = 3.98$, $p < 0.001$), an effect that was large as evidenced by Cohen's $d = 0.83$.

Curiously, the communication rates in words per minute (WPM) averaged over 5-day periods (middle panel of Figure 9) show a pattern that is inverse to that of the rates in messages per minute ($MsgPM$). The WPM graph started at a low value, reached a peak in the middle of the 45-day period, and ended at an intermediate level. The minimum, maximum and average communication rates over the 45 days were 0.9 (Days 1–5), 2.1 (Days 26–30), and 1.6 ± 0.4 WPM . The slope of the regression line was not significantly different from zero ($t(43) = 1.47$, $p = 0.148$).

To investigate the dissimilar trend between the communication rates in $MsgPM$ and WPM , the average words per message (WPM_{msg}) in a dialogue as a function of 5-day periods is plotted in the bottom panel of Figure 9. Across all days, the dialogs averaged 2.0 ± 0.7 words per message. The maximum number of words per message of a dialogue was 3.2 WPM_{msg} that occurred over Days 16–20, and the minimum was 1.2 WPM_{msg} which was observed over Days 36–40.

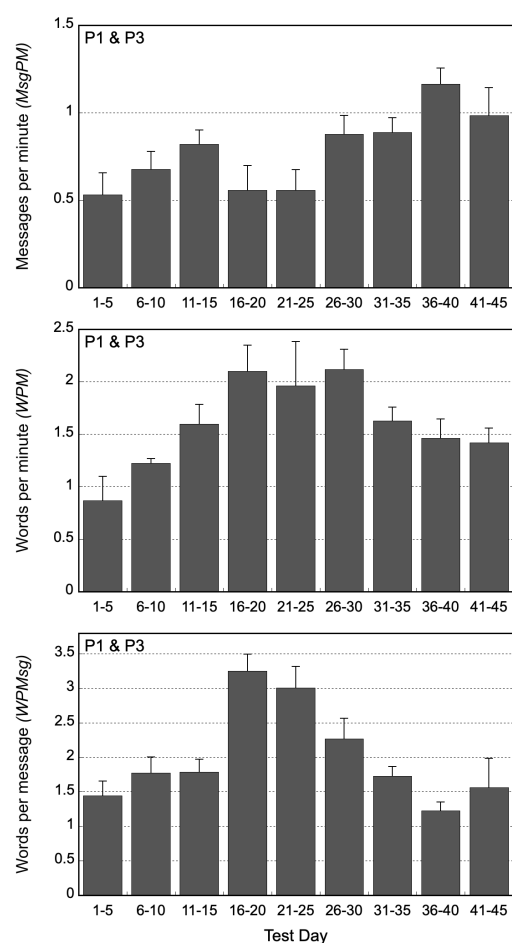


Figure 9. Communication rates combined over P1 and P3 plotted as a function of Test Day (binned into sets of five consecutive non-overlapping days). Upper panel shows rate in messages per minute (MsgPM), middle panel shows rate in words per minute (WPM), and lower panel shows rate in words per message (WPMMsg). The error bars represent standard errors.

4.5. Discussion

The two-way communication performance using the TAPS device is discussed and compared to previous studies in terms of results on (1) word reception, (2) message reception, (3) communication rates, and (4) directions for future research.

4.5.1. Word Reception

In the communication task reported here, the participants were able to repeat a message multiple times before making a response, resulting in an average of 2.6 repetitions prior to a response. Under these conditions, the overall reception of words on the messaging task averaged 81.7% correct across the two participants. The participants created messages spontaneously with no restriction on vocabulary, number of words in a message, or the number of phonemes in a word. Thus, the predictability of the words in the messages was fairly low, particularly at the beginning of a conversation. Of course, as the conversation progressed, contextual cues would lead to an increase in the predictability of future words. The performance of P1 and P3 may be compared to their individual results in the TAPS word-identification study of Tan et al. [12], with the caveat that performance in [12] was based on a single stimulus presentation. The average score of P1 and P3, who were among the top-performers in Tan et al.'s study, was $91.7 \pm 2.1\%$ for the task of identifying isolated words from a 500-word vocabulary. The performance of P1 and P3 in the previous study of Tan et al. [12] indicates high accuracy in the tactile reception of large vocabulary of isolated words. The present study has demonstrated their ability to receive impromptu text

messages on the skin of their forearms via the TAPS system, albeit allowing for multiple presentations before performance was scored.

4.5.2. Message Reception

Over the course of the study, 73.4% of the messages were received correctly, requiring correct reception of each word in the message. This performance is similar to that obtained in a previous study with TAPS, where participants were trained and tested on the identification of two-word phrases ([14]). Participants who had an average of 11 h of previous experience on phoneme and word identification with TAPS received additional training on the identification of a set of 218 two-word phrases. The phrases were semantically and syntactically valid and were composed of words from a previously acquired 500-word vocabulary. Mean performance was 75% correct for identification of the entire two-word phrase following 2.5 h of training, using an inter-phoneme interval of 150 ms and an inter-word interval of 1 s. Although the rate of reception is similar between this study and the present study, several differences must be noted. While stimuli in the previous study were restricted to 218 two-word phrases drawn from a previously trained vocabulary limited to 500 words, the messages transmitted in the current study were drawn from an unrestricted vocabulary and varied in length with a mean of 2 words per message and a range of 1.2 to 3.2 words per message. Again, participants had the option of repeating a message prior to making a response in the current study which was not an option in [14].

Reception of sentences for a tactile-alone condition has been reported by Galvin et al. [24] using the Tickle Talker device (described above in Sec. III-E). After receiving 12 to 33 h of training on word identification, four participants had acquired vocabularies of 34 to 77 words with an accuracy of 70–80% correct. Using each participant's acquired vocabulary, sentences were created using the trained words as key words and delivered to the device through live speech. Key word reception ranged from 6.5 to 22.8% correct across participants, scores that are substantially lower than the message or word reception rate reported here. However, these sentences were presented at slow-to-normal speaking rates which is substantially faster than the presentation rates achieved with the TAPS system.

A messaging study has also been reported using the WhatsHap device for the tactile display of phonemes [7,8], previously described above in Sec. III-E. Three experienced users of the device (with prior training on phoneme, word, and phrase identification) were each paired with a "speaker" to accomplish a problem-solving task. The "speaker" typed or spoke a message that was transformed into tactile phonemes presented through the device. The WhatsHap user first decoded the words in the tactile message (with the option of multiple replays of the tactile signal) and then typed a text message back to the "speaker" who received it as text. This exchange continued in an effort to complete the task. Across participants, 87.5% of the tasks were completed successfully, with a phonetic accuracy of 0.73 (computed as a Levenshtein edit distance). Over the course of the study, there was a decrease both in the length of the messages (from roughly 5.5 to 3 words/message) and in the number of repetitions needed before a response (from 3.4 to 1.9). A decrease in the number of words per message was also observed in the current study, peaking at 3.2 words/message on Days 16 to 20 and decreasing to roughly 1.6 words/message from Days 41 to 45. The number of repetitions per message, however, remained fairly constant over the course of the current study at 2 to 3 presentations per message.

Several major differences should be noted between the messaging studies conducted with WhatsHap [7,8] and in the current work with TAPS. First, the nature of the two tasks was quite different. The WhatsHap study required the two participants to complete a given task, whereas in the TAPS study, the two participants conducted a conversation on a selected topic. Second, there was a key difference between the two studies in how the messages were received. In the current study, two-way tactile messaging took place between the two participants with the TAPS devices. In the WhatsHap study, however, only one of the participants received a tactile message while the other participants received a

written text message. This may explain the difference in the trends observed for the number of presentations required per message. In the TAPS study, where the number of repetitions per message was constant between 2 and 3 throughout the study, there was a chance for errors on tactile message reception by both participants. In the WhatsHap study, however, only one of the participants was attempting to decode tactile messages while the other always had perfect reception via texting. This may have led to improved communication as the messaging progressed and contextual information accrued, and thus resulted in a decrease in the number of repetitions over the completion of a task. One similar finding across the two studies is a decrease in the length of a message over the course of the study, based on the participants' insight that performance was better with shorter messages.

4.5.3. Communication Rates

Although the participants were able to receive messages and individual words within a message with an accuracy sufficient for communication, the overall rates of communication were quite low. On average across the course of the study, it took roughly 2 min for the successful transmission of a message. This slow rate of transmission can be attributed to a number of factors in the manner in which the experiment was designed and conducted. One such factor is the increased duration of the tactile phonemic signals compared to speech, including the durations of the tactile phonemes themselves, as well as the timing between phonemes and words. Reed et al. [14] estimated maximum achievable communication rates of roughly 45 words/min with the phoneme durations, IPI (inter-phoneme interval), and IWI (inter-word interval) values employed in the current study. These rates are roughly four times longer than for normally produced speech. Another factor contributing to the message reception rate was the option to replay a message an unlimited number of times before making a response (on average a message was presented 2.6 times over the course of the experiment), thus adding to the time in sending a message. After accounting for the time required to enter a response, additional time was then needed for the "receiver" to switch roles and become the "sender". This included time to formulate and deliver a new message to their partner. This combination of temporal parameters contributes to the slow transmission rates observed here.

Within the context of the experiment itself, there were additional factors that contributed to variations in message transmission rates, including the length of the messages. The trough in *MsgPM* observed over the period of Days 16–20 and 21–25 in the upper panel of Figure 9 may be related to a temporary increase in the number of words per message over this same period of time as shown in the bottom panel of Figure 9. It appears that the longer messages led to a decrease in communication rate in *MsgPM*. It is quite possible that the participants became aware of increased difficulty with the task as they made their messages longer, and subsequently used a strategy of sending shorter messages that could be received more easily. The communication rate in *WPM*, however, showed an opposite effect in which the maximum rate in words/min was achieved for the longer messages. This somewhat puzzling result can be interpreted as the participants being able to identify correctly most of the words in a sentence while at the same time not receiving the entire message correctly. This phenomenon could help explain the dissimilar trend between communication rates of *WPM* and that of *MsgPM*. Scoring a high PC_{word} does not necessarily mean that the participants have a high PC_{msg} . In fact, a message is scored correct only when PC_{word} is 100% correct. The participants might have correctly identified short messages containing a small number of words but incorrectly identified complete sentences with a larger number of words. Hence, this could explain why the communication rate in *MsgPM* was lowest for the longest messages, but the rate in *WPM* was highest for the longest messages.

The results of the present study can be discussed relative to other studies that have examined rates for the communication of English words through tactile stimulation. One early study [36] used the "vibratese language", in which 45 symbols, such as English letters and numbers, were encoded on five vibrators. Following roughly 12 h of training in which

participants learned to identify the letters and numbers from their tactile codes, they were able to receive “vibratese” at a rate of 38 words/min (compared to a maximum transmission rate of 67 words/min). A similar transmission rate of roughly 30–35 words/min was obtained in a previous study of the identification of two-word phrases with the TAPS system [14]. The transmission rate of the present study cannot be directly compared to either the previous study with TAPS or to “vibratese”, however, due to the differences in the way that the time variable was calculated. Whereas the participants of other past studies were typically instructed to respond as quickly and accurately as they could upon receiving pre-constructed stimuli, the two participants in the present study recorded their responses to a transmitted message and then took the time to think about the next message before typing it into a message box on the screen. In other words, the “conversation” during the present study was more “natural” and the messages were not pre-determined. The time that the two participants took to record their responses and to think and type new messages penalized the calculation of communication rates in the present study.

4.5.4. Future Research

The current study has provided us with an existence proof of the ability of two-way communication through the sole use of tactile input using the phonemic-based TAPS device. Further work is necessary to address certain limitations of the current study, including the use of a larger number of participants, as well as the development of greater control over the content of the conversations between users. Future research will be concerned with the applications of the TAPS device for use by persons with profound sensory loss of hearing and/or sight. The utility of the TAPS device must also be evaluated in relation to other types of communication devices being developed for the community of persons with deafness and/or deaf-blindness (see summary in [37]). Examples of some recent research in this area include the presentation of Braille codes through the screen of a mobile device [38], the development of a device for presentation of Braille codes through the backs of the fingers [39], and an anthropomorphic arm-hand system for the tactile presentation of fingerspelling and sign language [40]. These different approaches reflect on the variety of methods of communication that are employed within the community of deaf-blind persons, requiring different types of technology to meet the needs of persons who rely on different types of communication, including speech, fingerspelling, and sign language.

5. Concluding Remarks

The experiments reported here have led to new results for the tactile reception of speech through the TAPS system, summarized below.

- Following training on the order of 3 to 5 h, participants could successfully identify a trained set of 4-phoneme words at a rate of 87% correct. This training generalized to a novel set of 4-phoneme words where performance dropped by only 10 percentage points.
- A two-way messaging task was carried out by two experienced users of TAPS. Over a 45-day period encompassing 13 h, 73% of the messages were identified accurately in conjunction with 82% of the words, based on an average of 2.6 presentations per message. Rates of communication using this ecologically valid setup, however, were slow, averaging of the order of 1 message per minute.
- The two-way messaging results with TAPS contribute to the literature on the use of tactile devices for speech communication in a realistic environment, adding to the work of de Vargas et al. [7] with WhatsHap. These studies represent an advancement over much of the previous research that has employed closed sets of word stimuli with one-way communication in a well-controlled laboratory setting. Our research has demonstrated the potential for real-life communication using spontaneous messages composed of an open vocabulary between two users of the TAPS tactile system.
- Future research will be focused on increasing the rates of communication achieved with TAPS. The time needed for communication can be reduced by a reduction in

temporal factors, including phoneme duration, inter-phoneme interval, and inter-word interval, as well as a further practice with the TAPS system.

Author Contributions: Conceptualization, C.M.R. and H.Z.T.; data curation, J.J.; funding acquisition, C.M.R. and H.Z.T.; investigation, J.J.; methodology, C.M.R., J.S.M. and H.Z.T.; project administration, H.Z.T.; software, J.J. and J.S.M.; supervision, C.M.R. and H.Z.T.; validation, J.S.M.; writing—original draft, J.J.; writing—review and editing, J.J., C.M.R., J.S.M. and H.Z.T. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by a research grant funded by Facebook Inc. and by Grants No. 1954842-IIS and 1954886-IIS from the National Science Foundation.

Institutional Review Board Statement: This study was conducted in accordance with the Declaration of Helsinki and approved by the Institute Review Board at Purdue University (Protocol No. 0611004681; Expiration date 7-Oct-2024).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author.

Acknowledgments: The authors thank Gina Clepper for feedback regarding the experimental setup.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Zhao, S.; Israr, A.; Lau, F.; Abnoui, F. Coding tactile symbols for phonemic communication. In Proceedings of the CHI Conference on Human Factors in Computing Systems, Montreal, QC, Canada, 21–26 April 2018; pp. 1–13.
2. Turcott, R.; Chen, J.; Castillo, P.; Knott, B.; Setiawan, W.; Briggs, F.; Klumb, K.; Abnoui, F.; Chakka, P.; Lau, F.; et al. Efficient evaluation of coding strategies for transcutaneous language communication. In Proceedings of the International Conference on Human Haptic Sensing and Touch Enabled Computer Applications, Pisa, Italy, 13–16 June 2018; pp. 600–611.
3. Chen, J.; Turcott, R.; Castillo, P.; Setiawan, W.; Lau, F.; Israr, A. Learning to feel words: A comparison of learning approaches to acquire haptic words. In Proceedings of the 15th ACM Symposium on Applied Perception, Vancouver, BC, Canada, 10–11 August 2018; pp. 1–7.
4. Dunkelberger, N.; Sullivan, J.; Bradley, J.; Walling, N.P.; Manickam, I.; Dasarathy, G.; Israr, A.; Lau, F.W.; Klumb, K.; Knott, B.; et al. Conveying language through haptics: A multi-sensory approach. In Proceedings of the 2018 ACM International Symposium on Wearable Computers, Singapore, 8–12 October 2018; pp. 25–32.
5. Dunkelberger, N.; Sullivan, J.L.; Bradley, J.; Manickam, I.; Dasarathy, G.; Baraniuk, R.; O'Malley, M.K. A multisensory approach to present phonemes as language through a wearable haptic device. *IEEE Trans. Haptics* **2020**, *14*, 188–199. [[CrossRef](#)] [[PubMed](#)]
6. de Vargas, M.F.; Weill-Duflos, A.; Cooperstock, J.R. Haptic speech communication using stimuli evocative of phoneme production. In Proceedings of the 2019 IEEE World Haptics Conference (WHC), Tokyo, Japan, 9–12 July 2019; pp. 610–615.
7. de Vargas, M.F.; Marino, D.; Weill, A.; Cooperstock, J.R.; et al. Speaking haptically: From phonemes to phrases with a mobile haptic communication system. *IEEE Trans. Haptics* **2021**, *14*, 479–490. [[CrossRef](#)] [[PubMed](#)]
8. Marino, D.; de Vargas, M.F.; Weill, A.; Cooperstock, J.R. Conversing Using WhatsApp: A Phoneme Based Vibrotactile Messaging Platform. In Proceedings of the IEEE World Haptics Conference (WHC), Montreal, QC, Canada, 6–9 July 2021; pp. 943–948.
9. Jung, J.; Jiao, Y.; Severgnini, F.M.; Tan, H.Z.; Reed, C.M.; Israr, A.; Lau, F.; Abnoui, F. Speech communication through the skin: Design of learning protocols and initial findings. In Proceedings of the International Conference of Design, User Experience, and Usability, Las Vegas, NV, USA, 15–20 July 2018; pp. 447–460.
10. Jiao, Y.; M Severgnini, F.; Martinez, J.S.; Jung, J.; Tan, H.Z.; Reed, C.M.; Wilson, E.C.; Lau, F.; Israr, A.; Turcott, R.; et al. A comparative study of phoneme-and word-based learning of English words presented to the skin. In Proceedings of the International Conference on Human Haptic Sensing and Touch Enabled Computer Applications, Pisa, Italy, 13–16 June 2018; pp. 623–635.
11. Reed, C.M.; Tan, H.Z.; Perez, Z.D.; Wilson, E.C.; Severgnini, F.M.; Jung, J.; Martinez, J.S.; Jiao, Y.; Israr, A.; Lau, F.; et al. A phonemic-based tactile display for speech communication. *IEEE Trans. Haptics* **2019**, *12*, 2–17. [[CrossRef](#)] [[PubMed](#)]
12. Tan, H.Z.; Reed, C.M.; Jiao, Y.; Perez, Z.D.; Wilson, E.C.; Jung, J.; Martinez, J.S.; Severgnini, F.M. Acquisition of 500 English words through a TActile phonemic sleeve (TAPS). *IEEE Trans. Haptics* **2020**, *13*, 745–760. [[CrossRef](#)] [[PubMed](#)]
13. Martinez, J.S.; Tan, H.Z.; Reed, C.M. Improving tactile codes for increased speech communication rates in a phonemic-based tactile display. *IEEE Trans. Haptics* **2021**, *14*, 200–211. [[CrossRef](#)] [[PubMed](#)]
14. Reed, C.M.; Tan, H.Z.; Jiao, Y.; Perez, Z.D.; Wilson, E.C. Identification of words and phrases through a phonemic-based haptic display: Effects of inter-phoneme and inter-word interval durations. *ACM Trans. Appl. Percept.* **2021**, *18*, 1–22. [[CrossRef](#)]
15. Kirman, J.H. Tactile communication of speech: A review and an analysis. *Psychol. Bull.* **1973**, *80*, 54–74. [[CrossRef](#)] [[PubMed](#)]
16. Reed, C.M.; Durlach, N.I.; Braida, L.D. Research on tactile communication of speech: A review. *ASHA Monogr.* **1982**, *80*, 1–23.

17. Reed, C.M.; Durlach, N.I.; Delhorne, L.A.; Rabinowitz, W.M. Research on tactual communication of speech: Ideas, issues, and findings. *Volta Rev.* **1989**, *91*, 65–78.
18. Kappers, A.M.; Plaisier, M.A. Hands-Free Devices for Displaying Speech and Language in the Tactile Modality—Methods and Approaches. *IEEE Trans. Haptics* **2021**, *14*, 465–478. [[CrossRef](#)]
19. Uchanski, R.M.; Braida, L.D. Effects of token variability on our ability to distinguish between vowels. *Percept. Psychophys.* **1998**, *60*, 533–543. [[CrossRef](#)] [[PubMed](#)]
20. Dunkelberger, N.; Bradley, J.; Sullivan, J.L.; Israr, A.; Lau, F.; Klumb, K.; Abnoui, F.; O'Malley, M.K. Improving perception accuracy with multi-sensory haptic cue delivery. In Proceedings of the International Conference on Human Haptic Sensing and Touch Enabled Computer Applications, Pisa, Italy, 13–16 June 2018; pp. 289–301.
21. Brooks, P.; Frost, B.J. Evaluation of a tactile vocoder for word recognition. *J. Acoust. Soc. Am.* **1983**, *74*, 34–39. [[CrossRef](#)]
22. Brooks, P.; Frost, B.; Mason, J.; Chung, K. Acquisition of a 250-word vocabulary through a tactile vocoder. *J. Acoust. Soc. Am.* **1985**, *77*, 1576–1579. [[CrossRef](#)] [[PubMed](#)]
23. Brooks, P.; Frost, B.; Mason, J.; Gibson, D. Continuing evaluation of the Queen's University tactile vocoder. I: Identification of open set words. *J. Rehabil. Res. Dev.* **1986**, *23*, 119–128.
24. Galvin, K.L.; Blamey, P.J.; Cowan, R.S.; Oerlemans, M.; Clark, G.M. Generalization of tactile perceptual skills to new context following tactile-alone word recognition training with the Tickle Talker™. *J. Acoust. Soc. Am.* **2000**, *108*, 2969–2979. [[CrossRef](#)] [[PubMed](#)]
25. Norton, S.J.; Schultz, M.C.; Reed, C.M.; Braida, L.D.; Durlach, N.I.; Rabinowitz, W.M.; Chomsky, C. Analytic study of the Tadoma method: Background and preliminary results. *J. Speech Hear. Res.* **1977**, *20*, 574–595. [[CrossRef](#)]
26. Reed, C.M.; Rabinowitz, W.M.; Durlach, N.I.; Braida, L.D.; Conway-Fithian, S.; Schultz, M.C. Research on the Tadoma method of speech communication. *J. Acoust. Soc. Am.* **1985**, *77*, 247–257. [[CrossRef](#)]
27. Reed, C.M. Tadoma: An Overview of Research. In *Profound Deafness and Speech Communication*; Whurr Publishers: London, UK, 1995; pp. 40–55.
28. Reed, C.M.; Durlach, N.I. Note on information transfer rates in human communication. *Presence* **1998**, *7*, 509–518. [[CrossRef](#)]
29. Jones, L.A.; Tan, H.Z. Application of psychophysical techniques to haptic research. *IEEE Trans. Haptics* **2013**, *6*, 268–284. [[CrossRef](#)]
30. Boothroyd, A.; Hanin, L.; Hnath, T. *A Sentence Test of Speech Perception: Reliability, Set-Equivalence, and Short-Term Learning*; RCI10; City University of New York, Speech and Hearing Sciences Research Center: New York, NY, USA, 1985.
31. Dudai, Y.; Karni, A.; Born, J. The consolidation and transformation of memory. *Neuron* **2015**, *88*, 20–32. [[CrossRef](#)]
32. Black, A.W.; Lenzo, K. Flite: A Small, Fast Speech Synthesis Engine. 2005. Available online: <http://cmufite.org/doc/flite.pdf> (accessed on 5 February 2024).
33. Shim, S.; Tan, H.Z. PalmScape: Calm and pleasant vibrotactile signals. In Proceedings of the Design, User Experience, and Usability, Interaction Design: 9th International Conference, DUXU 2020, Copenhagen, Denmark, 19–24 July 2020; Chapter 37, pp. 1–17.
34. Sherbecoe, R.L.; Studebaker, G.A. Supplementary formulas and tables for calculating and interconverting speech recognition scores in transformed arcsine units. *Int. J. Audiol.* **2004**, *43*, 442–448. [[CrossRef](#)] [[PubMed](#)]
35. Studebaker, G.A. A “rationalized” arcsine transform. *J. Speech, Lang. Hear. Res.* **1985**, *28*, 455–462. [[CrossRef](#)] [[PubMed](#)]
36. Geldard, F.A. Adventures in tactile literacy. *Am. Psychol.* **1957**, *12*, 115–124. [[CrossRef](#)]
37. Reed, C.M.; Tan, H.Z.; Jones, L.A. Haptic Communication of Language. *IEEE Trans. Haptics* **2023**, *6*, 134–153. [[CrossRef](#)]
38. Rantala, J.; Raisamo, R.; Lylykangas, J.; Surakka, V.; Raisamo, J.; Salminen, K.; Pakkanen, T.; Hippula, A. Methods for presenting braille characters on a mobile device with a touchscreen and tactile feedback. *IEEE Trans. Haptics* **2009**, *2*, 28–39. [[CrossRef](#)]
39. Amemiya, T.; Yamashita, J.; Hirota, K.; Hirose, M. Virtual leading blocks for the deaf-blind: A real-time way-finder by verbal-nonverbal hybrid interface and high-density RFID tag space. In Proceedings of the IEEE Virtual Reality, Chicago, IL, USA, 27–31 March 2004; pp. 165–287.
40. Chang, C.M.; Sanches, F.; Gao, G.; Johnson, S.; Liarokapis, M. An Adaptive, Affordable, Humanlike Arm Hand System for Deaf and DeafBlind Communication with the American Sign Language. In Proceedings of the 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Kyoto, Japan, 23–27 October 2022; pp. 871–878.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.