Improving Tactile Codes for Increased Speech Communication Rates in a Phonemic-Based Tactile Display

Juan S. Martinez, *Member, IEEE*, Hong Z. Tan, *Fellow, IEEE*, and Charlotte M. Reed

***Abstract*—Previous research has shown evidence of tactile speech acquisition of up to 500 English words presented as tactile phonemic patterns using a 4-by-6 tactor array worn on the forearm. This article describes modifications to some of the tactile codes encoding the 39 English phonemes, and ten additional codes as abbreviated patterns for the ten most frequent phoneme pairs in spoken English. The re-design aimed to reduce the duration of phonemes and phoneme pairs that occur most frequently, with the goal to increase tactile speech transmission rates. Code identification experiments were conducted with ten participants over three weeks using a video game. The average identification rate of the 49 modified codes (39 phonemes plus 10 phoneme pairs) was 83.3% with an average learning time of 6.2 hours. The average identification rate of the 49 codes in a retention test with 7 of the 10 participants after more than 90 days of no exposure to the tactile codes was 75.7%. An analysis using ideal transmission rates showed a 58% increase in transmission rate with the modified tactile codes as compared to the original codes, demonstrating that the improved codes can speed up tactile speech communication.**

***Index Terms*—Frequency-based tactile coding, phonemic-based tactile display, speech transmission rate, tactile codes for phonemes, tactile speech communication.**

I. Introduction

The sense of touch can be an alternative channel of communication when the auditory and/or visual modalities are absent or impaired. For example, deaf-and-blind users of the Tadoma method are able to perceive and recognize facial movements and other articulatory features from a talker's face during speech production by placing their hand on the face of the talker. Previous research on the Tadoma method used for speechreading demonstrates reception of connected speech at rates between 60 and 80 words per minute (wpm), and identification accuracy of segmental consonants and vowels, isolated monosyllabic words, and key words in conversational sentences at 55%, 40% and 80%, respectively [1], [2], indicating improved speech comprehension with contextual cues.

The existence of Tadoma and other natural tactile speech communication methods has motivated the development of various devices that employ specific codifications of speech or text information into tactile stimuli. These translations are based on different approaches; for example, mappings from spectral properties of acoustic speech signals to location of stimulation [3], [4], extraction of articulatory-based speech features from signals that are later encoded into tactile stimuli applied to the skin [5], and letter-based encoding schemes for text [6], [7]. The general consensus is that tactile aids can be useful as a supplement to lipreading or for the reception of environmental sounds by people with severe-to-profound hearing impairments, but cannot be used alone for speech communication [8]–[10].

Our most recent efforts in this area of research have employed an approach in which phonemic transcriptions of text or speech are encoded into a sequence of tactile stimuli composed of 39 tactile codes corresponding to the 39 English phonemes [11]. The tactile codes are presented through a 24-tactor array worn on the forearm, referred to as TAPS (TActile Phonemic Sleeve). This strategy proved to be highly effective in that the 39 codes were distinct (86% recognition accuracy), could be learned within a reasonable amount of time (1 to 4 hours), and eliminated the issue of token variations (the same sound pronounced differently within or across talkers) that increases the demands placed on learning and recognition.

Jiao *et al.* [12] conducted a study comparing two different training strategies for learning to recognize words using TAPS. In one strategy, participants were trained to identify the 39 phoneme codes before learning words, and in the second strategy, participants went directly to the word-learning task without explicit training on the individual phonemes. The results showed that the best participants were able to learn one English word per minute on average with a 100-word vocabulary using either strategy. A comparison of the learning curves across participants for both methods, however, indicated that learning individual phonemes prior to words was a more consistent path for improvement with practice. Furthermore, the third study from Tan *et al.* [13] using TAPS demonstrated that participants

Manuscript received March 1, 2020; revised June 12, 2020; accepted July 9, 2020. Date of publication July 14, 2020; date of current version March 19, 2021. This work was supported in part by a gift and a Faculty Research Award from Google LLC, a fellowship to the first author from Colfuturo, and Grant No. 1954842-IIS from the National Science Foundation.This article recommended for publication by Associate Editor Hiroyuki Kajimoto and Editor Domenico Prattichizzo upon evaluation of the reviewers' comments. *(Corresponding author: Juan S. Martinez)*

Juan S. Martinez and Hong Z. Tan are with the Haptic Interface Research Laboratory, School of Electrical and Computer Engineering, College of Engineering, Purdue University, West Lafayette, IN 47907 USA. (e-mail: mart1304@purdue.edu; hongtan@purdue.edu).

Charlotte M. Reed is with the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139 USA. (e-mail: cmreed@mit.edu).

Digital Object Identifier 10.1109/TOH.2020.3008869

can learn to recognize up to 500 English words at a maximum rate of 1 wpm with a training curriculum of up to 8 hours. The relatively large 500-word list was considered essentially an open set as the participants responded by typing words received without access to the full list of 500 words, and thus were unlikely to have memorized the complete vocabulary. These results demonstrated that phonemic-based encoding is theoretically capable of the transmission of any English word and can be learned in a reasonable amount of time.

The main goal of the present study was to increase stimulus presentation rates of the TAPS speech communication system. The stimulus presentation rate in the Reed *et al.* [11], Jiao *et al.* [12] and Tan *et al.* [13] studies was roughly 40 wpm, which is lower than the rates of 60 to 80 wpm demonstrated by Tadoma users. Faster rates may be achieved by a reduction in the duration of the phonemic codes and/or the inter-phoneme intervals. The present study tackled the first issue by assigning shorter tactile codes for commonly-used phonemes, and creating additional abbreviated codes for frequently co-occurring phoneme pairs. The modifications resulted in a revised set of tactile codes that can lead to increased communication rates by reducing the time it takes to deliver codes. Ten participants completed a three-week learning curriculum for the revised tactile codes. Their identification scores were collected and compared to those reported in our earlier work [11]. Seven of the ten participants also completed a retention test after periods of no exposure to the tactile codes for more than 90 days.

During the present study, a game-based training paradigm was developed for a more engaging learning experience. Past studies of game designs have proven the efficacy of games in learning a second language [14], [15]. Role playing games and immersive scenarios that support language learning in context have demonstrated, for example, that people with little knowledge of the Chinese language can learn to recognize Chinese words in visual and auditory tests after playing an immersive video game [14]. In this regard, training on phonemes and words may be achieved in a more engaging manner by using a role playing video game that follows the same game design ideas proven to be effective in second language acquisition. The immediate goal of the present study was to provide evidence that the new tactile codes are distinguishable, memorable and can be learned in a relatively short period of time. In support of the study objective, a video game based on game design principles for language learning was created to support the learning and testing of the new tactile codes. The evaluation of the efficacy of the game-based approach to training and learning (although worthy of eventual consideration) is beyond the scope of the present study.

Fig. 1. Experimental setup with the phonemic-based speech communication system TAPS (TActile Phonemic Sleeve).

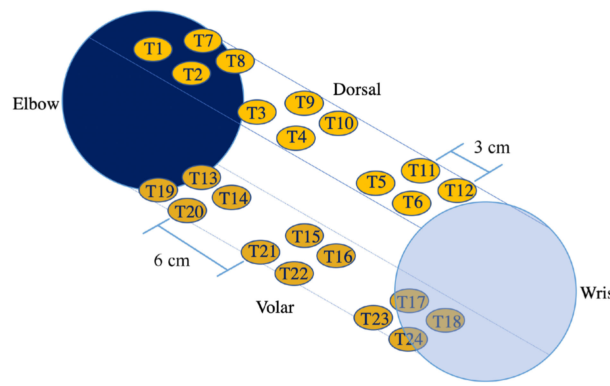


Fig. 2. Illustration of tactor layout on the forearm. The tactors are arranged in six groups of four tactors, with a distance of 3 cm between the center of tactors within a group, and 6 cm between the center of ipsilateral nearby tactors from different groups.

II. Methods

A. Participants

Five female and five male participants were recruited through an approved IRB protocol at Purdue University. The participants (P01 to P10) gave informed consent and were paid for their time. Their ages ranged from 18 to 26 years and included six native English speakers, two native Spanish speakers and two native Chinese speakers. The four non-native English speakers are fluent in English, having acquired it as a second language at the ages of 6, 7 or 9 years. The participants were naive in the sense that they had not used the TAPS system before. The participants had normal hearing and sight, and none reported any problems with their sense of touch.

B. Experimental Apparatus

The experimental apparatus was the TAPS speech communication system from [11]–[13]. It consists of a 4-by-6 tactor array worn on the left forearm (see Figure 1 to Figure 3). The individual tactors (Tectonic Elements, Model TEAX13C02-8/RH, 26.3 mm in diameter and 9.0 mm thick) form 2 rows of 6 tactors in the longitudinal direction on the dorsal side of the forearm, and 2 rows of 6 tactors on the volar side (see Figure 2). The device is worn as a gauntlet composed of 2 separate pieces attached with Velcro straps: one for the dorsal side (top of Figure 3) and the other for the volar side (bottom of Figure 3).

The array of 24 tactors is connected to the outputs of 24 class D audio amplifiers. The input of the amplifiers is connected to the 24 outputs of a MOTU 24Ao audio interface that

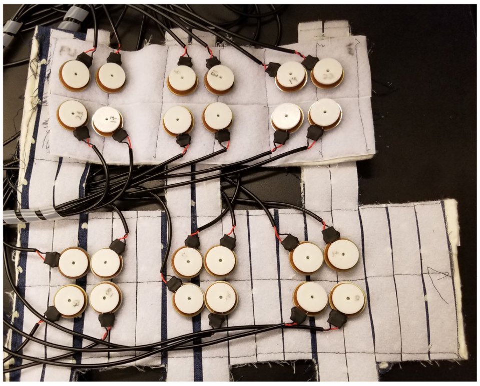


Fig. 3. The 4-by-6 tactor array on two gauntlet pieces.

TABLE I
MOST FREQUENTLY-USED PHONEMES AND MOST FREQUENTLY CO-OCCURRING PHONEME PAIRS

Most Frequent Phonemes		Most Frequent Phoneme Pairs	
Tactile Code	**Word Example**	**Tactile Code**	**Word Example**
UH	G**u**n	DH-UH	**The**
N	**N**o	UH-N	G**un**
T	**T**ea	T-UH	**Tou**ch
IH	Sh**i**p	IH-T	H**it**
S	**S**un	N-T	A**nt**
R	**R**ock	IH-NG	K**ing**
EE	K**ey**	N-D	Sa**nd**
L	**L**eaf	S-T	Gho**st**
D	**D**ish	Y-OO	**You**
EH	Br**ea**d	IH-N	P**in**
M	**M**ap		
DH	**Th**at		
K	**K**id		

is connected through a USB port to a Windows computer. More details are available in [11]–[13].

Our software consists of a Unity 3D application with a custom-built C++ library and a C# interface that communicates with the MOTU device using the PortAudio I/O audio library (http://www.portaudio.com/). The Unity application allowed the participants to train and test with different groups of tactile codes within the context of a video game. The participants navigated through the learning game with a keyboard and mouse as shown in Figure 1. They wore noise-reduction earmuffs that blocked auditory cues arising from TAPS throughout the experiment.

C. *Design of Tactile Stimuli*

Our earlier work took into account the following considerations in devising the tactile codes: (i) psychophysical properties of the tactile sensory system; (ii) principles for maximizing information transfer that include the use of multiple stimulus dimensions with only a few levels per dimension, and the use of movement patterns; (iii) articulatory properties of the speech stimuli to map phonemes to tactile codes, and (iv) heuristics to facilitate memorizing the mapping of phonemes to tactile codes. A detailed description of the 39 tactile codes developed to code the English phonemes (24 consonants and 15 vowels) is provided in [11] (see supplemental materials for [11]). The dimensions used to create the 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 volar), waveform (e.g., modulated or unmodulated), and the use of different types of movement patterns for vowels (e.g., saltatory versus smooth apparent motion; see [16] for a description of tactile movement illusions). 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). These tactile codes were recognized with a high degree of accuracy in isolation [11] and when used to form words [12], [13]. The communication rates at which words were transmitted with codes were estimated to be roughly 40 wpm.

The starting point of the present study was the set of 39 tactile codes developed in Reed *et al.* [11]. A subset of the codes was then modified, and another new set of codes was created. We aimed to achieve faster speech transmission and word recognition rates by incorporating the *statistics of spoken English*. Due to the distinctiveness of the tactile codes developed in [11], the modified codes adhered to most of the aforementioned characteristics, but relaxed the constraint on mapping related to articulatory properties.

Our design process can be summarized as follows. Firstly, the most frequent phonemes were coded as short (100 ms) single-frequency vibrations (either 60 or 300 Hz). This would reduce average durations of words and increase word presentation rates. We shortened 13 most frequently-used phonemes from the union of two lists, each containing 10 most frequently-used English phonemes according to two statistical studies of spoken English by Denes [17] and Mines *et al.* [18]. See the first two columns in Table I for the capital-letter symbols used for the phonemes and sample words containing the sounds. The tactile codes for the 13 phonemes were spatially distributed at distinct locations on the forearm, with constraints explained below. The remaining consonants were (re) distributed along the arm so that no more than 3 consonants were coded at the same location.

Secondly, the most frequent pairs of phonemes were coded in adjacent spatial locations to facilitate the creation of "chunks" – tactile codes that represent the most frequently co-occurring phoneme pairs. We selected the 10 phoneme pairs with the highest number of co-occurrences in Denes' study [17]. They appear in the third column of Table I with the corresponding word examples in the fourth column. As an example of how the spatial locations of phoneme pairs were adjusted, consider the pair S-T. The consonant S was originally coded with tactors at both the volar and dorsal elbow (T1, T7, T13 and T19 in Figure 2) in Reed *et al.* [11]. In the present study, S was moved to the volar elbow (T13, T14, T19 and T20) so that it was close to the consonant T, which was coded on the volar middle (T15, T16, T21, T22) in Reed *et al.*

TABLE II
THE 24 CONSONANT CODES USED IN THE PRESENT STUDY

Code	Word Example	Waveform			Location		Duration (ms)	Tactors Involved
		Intensity (dB SL)	Frequency (Hz)	Amplitude Modulation (Hz) or Shaping	Dorsal (D) vs. Volar (V)	Wrist (W), Mid-foream (M), or Elbow (E)		
P	**P**age	25	300		D	W	140	T5, T6, T11, T12
T	**T**ea	30	60		V	M	100	T15, T16, T21, T22
K	**K**ey	25	300		D	E	100	T1, T2, T7, T8
B	**B**ell	25	300	25	D	W	140	T5, T6, T11, T12
D	**D**ish	30	60		D	W	100	T5, T6, T11, T12
G	**G**old	25	300	25	V	M	140	T15, T16, T21, T22
CH	**Ch**ess	25	300	cos^2 window	D	W and E	400	T1, T6, T7, T12
J	**J**udge	25	300	9	D	W and E	400	T1, T6, T7, T12
F	**F**ish	25	300		D and V	W	400	T6, T12, T18, T24
V	Lo**v**e	25	300	9	D and V	W	400	T6, T12, T18, T24
TH	**Th**orn	25	300	20	D	M	140	T3, T4, T9, T10
DH	**Th**e	30	60		D	M	100	T3, T4, T9, T10
S	**S**un	30	60		V	E	100	T13, T14, T19, T20
Z	**Z**one	25	300	cos^2 window	V	E	400	T13, T14, T19, T20
SH	**Sh**ell	25	300	cos^2 window	V	W	400	T17, T18, T23, T24
ZH	A**z**ure	25	300	25	V	W	140	T17, T18, T23, T24
H	**H**at	25	60	cos^2 window	D and V	M	400	T4, T5, T10, T11, T16, T17, T22, T23
M	**M**ap	25	60	15	V	M	400	T15, T16, T21, T22
N	**N**oo**n**	30	60		D	E	100	T1, T2, T7, T8
NG	Wi**ng**	25	300	25	D	E	140	T1, T2, T7, T8
L	**L**eaf	25	300	15	D and V	M and W	100	T9, T10, T11, T12, T21, T22, T23, T24
R	**R**ock	25	300	30	Volar	E	100	T13, T14, T19, T20
W	**W**ine	25	60	9	D and V	M and W	400	T9, T10, T11, T12, T21, T22, T23, T24
Y	**Y**ellow	30	60		V	W	100	T17, T18, T23, T24

[11]. By modifying the location of S, we were later able to create a two-phoneme "chunk" code for S-T that occupied a relatively small region on the forearm, as explained below.

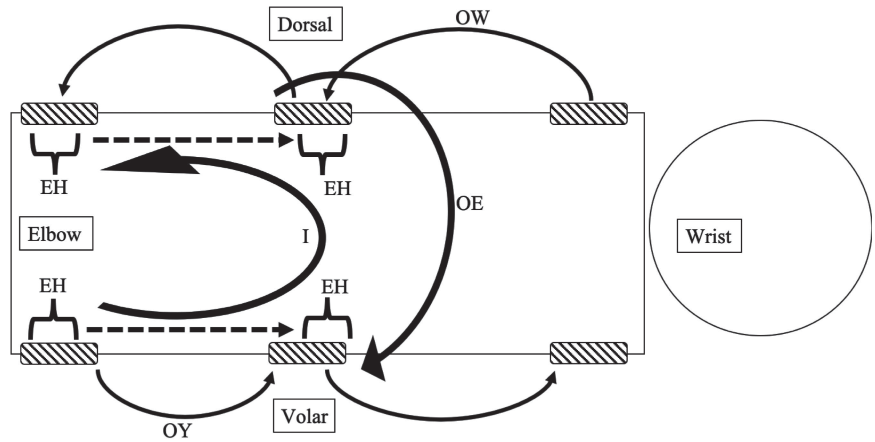
Thirdly, additional codes to represent the 10 two-phoneme "chunks" were created by concatenating shortened versions of the newly designed tactile codes for the two phonemes in a chunk code and presenting them sequentially. For example, for the S-T chunk, the duration of S was reduced from 100 to 50 msec and that of T from 100 to 50 msec. When presenting the chunk code S-T, the shortened version of S was presented first, followed by the shortened version of T immediately after. The shortened tactile codes preserved the *feel* of the tactile codes for the S and T phonemes when presented alone, and did not require the user to relearn the tactile codes. The combined duration of the S-T phoneme pair was shorter (100 msec) than the sum of their individual durations (200 msec), thereby leading to further increases in the presentation rates of English words and sentences. By having the shortened S and T occur at nearby locations on the skin, we were able to take advantage of the user's spatial attention being already focused on S, and the chance of the shortened T being missed was greatly reduced.

Finally, when a vowel was involved in the most frequent phoneme pairs, the movement sensations were re-designed so that the start (or end) of the vowel connected with the end (or start) of the other phoneme in a pair. For example, one of the most frequent phoneme pairs is DH-UH (as in "**the**"). In Reed *et al.* [11], DH was located at the dorsal middle (T3, T4, T9, T10) and UH was a "grabbing" sensation moving from near the wrist towards the middle of the forearm. In the present study, UH was modified to be a "twinkle"-like moving sensation on the dorsal side that started at T3 and T9 (to connect with DH) and finished off at T1, T2, T7 and T8 near the elbow. Another example is the IH-NG pair (as in "k**ing**"). In Reed *et al.* [11], IH was a short smooth motion from the elbow to the middle (T1-T7 $\rightarrow$ T2-T8 $\rightarrow$ T3-T9 $\rightarrow$ T4-T10, where "-" indicates simultaneous activation of the two tactors), and NG was delivered on the dorsal elbow with T1-T2-T7-T8. In the present study, IH remained a smooth motion but moved from the wrist to the middle (T6-T12 $\rightarrow$ T5-T11 $\rightarrow$ T4-T10 $\rightarrow$ T3-T9) to meet NG near the elbow without changing its movement direction. In general, care was taken so that the revised tactile codes resembled those developed earlier in Reed *et al.* [11] as much as possible, including the perceived intensity levels in dB SL (sensation level) and the frequency contents.

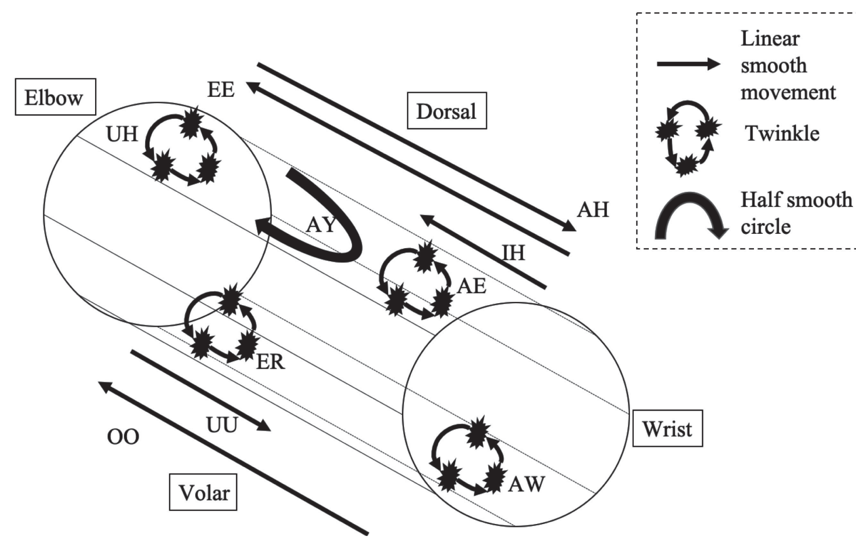
The resulting set of 24 consonant codes are described in Table II. Pictographic representations of the movement patterns associated with the newly-designed 15 vowels are described in Figure 4, in two plots for ease of viewing. A schematic diagram of the 10 chunk codes is shown in Figure 5. Readers interested in duplicating the new tactile codes can find a complete description online at https://juansmartinez.github.io/ImprovedTactileCodes/

D. Experimental Procedure

Tactile code identification experiments were performed with the ten participants to test the learnability and distinctiveness



(a) Illustration of the movement patterns associated with five of the fifteen vowels: OW and OY used sensory saltation (thin arrow arcs), I and OE used apparent motion (thick arrow arcs), and EH is a "grabbing" sensation that activates the tactors at the elbow prior to activating tactors at the middle of the forearm (dashed lines).



(b) Illustration of the movement patterns associated with the remaining ten vowels that all used apparent motion. The sensations were either smooth movements along a linear or half-circle trajectory (UU, OO, AH, IH, EE and AY) or "twinkle"-like (UH, AE, ER, AW).

Fig. 4. The 15 vowel codes used in this article.

of the new set of 49 codes (39 individual phonemes plus the 10 most common phoneme pairs). The experiments consisted of daily sessions that ranged from 30 minutes to 1 hour over a period of 11 to 17 days, depending on the pace of the individual participants. Throughout the sessions, participants learned and were tested on increasing sets of tactile codes. After a period of at least 3 months of no exposure to the codes, seven of the ten participants completed a retention test with all 49 codes. This final tests assessed the memorability of the newly designed codes. This section describes the detection threshold measurements that were conducted with each participant prior to the main experiment, the learning protocols, the testing procedures, and the retention tests. Each participant spent between 8 to 16.5 hours of total time in the main experiment and 1 hour in the retention test.

1) Threshold Measurements and Intensity Adjustments: At the beginning of the first session, detection thresholds were measured for each participant using a three-interval, two-alternative, forced-choice, one-up two-down adaptive procedure with trial-by-trial correct-answer feedback. Thresholds were measured at tactor T10 at the middle of the dorsal forearm (see Figure 2) at 300 Hz and 60 Hz. This was followed by adjustments of perceived intensities at the other tactors using the method of adjustment. For the adjustment procedure, tactor T10 was used as a reference. For each of the remaining 23 test tactors, the participant felt the sequence "reference-test-reference" and manually adjusted the intensity of the test tactor until it was perceived to be as strong as the reference. We used the same parameters for threshold estimation and intensity equalization as those described in Sec. IV.D-2 of Reed *et al.* [11]. The results were

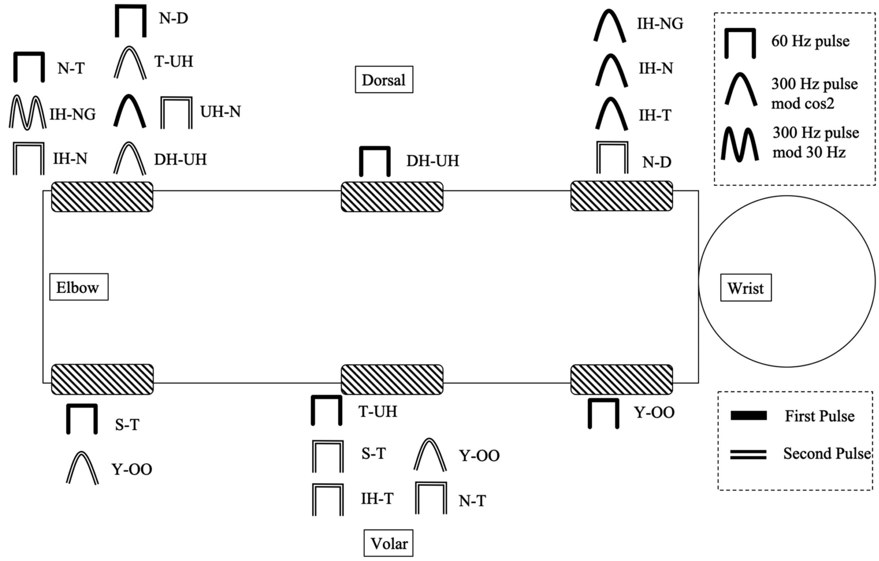


Fig. 5. Schematic of the 10 chunk codes used in the present study. Each chunk is composed of a first sinusoidal pulse (single black line) followed by a second sinusoidal pulse (double black line) at different locations of the forearm. Pulses can be of three types: 60 Hz, 300 Hz with a cos^2 envelope, and 300 Hz with an amplitude modulation at 30 Hz.

TABLE III
TACTILE CODES BY TRAINING ZONES

Training Zone	Phonemes Included
Z1	T, S, DH, N, NG, D
Z2	P, K, B, M, G, TH, Z
Z3	F, V, CH, J, SH, ZH
Z4	H, L, R, W, Y
Z5	UH, IH, OO, UU, EE, AH
Z6	AW, ER, AY, AE
Z7	I, OE, OW, OY, EH
Z8	DH-UH, UH-N, T-UH IH-T, N-T, IH-NG N-D, S-T, Y-OO, IH-N

TABLE IV
LEARNING LEVELS

Learning Level	Zones Included	Number of Codes
L1	Z1	6
L2	Z1 to Z2	13
L3	Z1 to Z3	19
L4	Z1 to Z4	24
L5	Z5	6
L6	Z5 to Z6	10
L7	Z5 to Z7	15
L8	Z8	10
L9	Z1 to Z8	49

used to ensure that the tactile signals were delivered at the same perceived intensity levels across participants (by setting signals levels relative to an individual's detection thresholds measured at T10) and across tactors (by equalizing intensity levels relative to that at T10) in subsequent learning and testing procedures.

2) Learning Protocols: Learning took place in a custom-designed role playing game called "Haptos." The participant was represented by a virtual avatar and tasked to learn the phoneme codes in the "Training School" of a fantasy village. The learning materials consisted of the 49 tactile codes that were grouped by *Training Zones* (see Table III). Learning progressed according to *Learning Levels* defined in Table IV. All participants learned increasing numbers of consonants first (L1 to L4), followed by vowels (L5 to L7) and the new "chunk" signals (L8). Level L9 included all 49 codes and was tested at the end of the curriculum.

During the daily sessions, the participants interacted with phonemes by entering a training zone and clicking on the tactile codes in any order they wished (see Figure 6a). Each phoneme was represented by a sphere and was accompanied by an object that represented how the phoneme was used in a word. For example, in Figure 6b, the phoneme "SH" was shown next to a shell. The phoneme could be played and felt on the left forearm by clicking on "PLAY PHONEME". It could be returned to its original position by clicking on "RETURN THE PHONEME". The participant could navigate to another training zone by typing a zone number in the top-right text field and clicking on "GO". This free exploration period was self-paced and no time limit was imposed. Participants were allowed to explore any zone regardless of their current Learning Level, but they were informed about the relevant zones for learning each day and generally confined their exploration to these zones. Within any zone, the participants could click on "TEST" to perform a mock test. A randomly-selected phoneme within the zone was played, and the participant clicked on the sphere corresponding to the recognized phoneme. Trial-by-trial correct-answer feedback was provided to reinforce learning during the mock tests.

3) Testing Procedures: When the participants were ready for a test, they were instructed to go to the "Lab" for testing, this time without any feedback (Figure 7a). The lab was a room where tactile code identification experiments were conducted. The participants were required to pass a test at the current level before progressing to a higher Learning Level (see Table IV). On their first day in the Lab, the participants were

(a) Overview of the Training School. The participant can interact with any zone by approaching the lighted area under the zone number. Zones are scattered around the 3D space of the scene.

(b) Zone Z3 includes phonemes J, F, ZH, V, SH and CH. The participant has clicked on the sphere "SH" to bring the phoneme and the image of a shell that includes the "SH" sound into foreground. Other phonemes in the zone may become occluded as a result.

Fig. 6. The Training School.

(a) Main screen in the Lab. The text at the top left corner of the interface instructs the participant to enter a level number in the text box (e.g., "1" for L1). The participant can then start a tutorial or a test by clicking on "MAKE A CRYSTAL."

(b) Screenshot of a test at L9. The participant enters responses by selecting a stimulus from the list. Consonants, vowels and chunks are represented by blue, orange and purple spheres, respectively. Chunks are represented by connected pairs of spheres labeled with the corresponding phoneme pairs.

Fig. 7. The Lab.

encouraged to run a tutorial to become familiar with the test procedures. To initiate a test, the participant specified the level to be tested and a code belonging to the level was randomly selected and presented on each trial. The participant responded to the stimulus by selecting the sphere with the recognized code from those shown on the screen (see Figure 7b). The main difference between the mock test in the Training School and the test in the Lab was that the latter provided no feedback. A passing score of $\geq$ 80% was required before the participant was allowed to move to a higher level. The test was repeated the following day if the participant failed to reach the passing score. This passing criterion was applied to levels L1 to L8 but not L9.

The total number of trials for each test was group-dependent. It was set by the software so that the probability of each code appearing at least once during the test was at least 70%. The number of trials ranged from 20 to 30 from L1 to L4, and remained 20 for L5 to L8. For L9, the participants completed one block of 60 trials on each of 3 consecutive days for a total of 180 trials per participant.

4) Retention Tests: After the completion of the main experiment, the participants were invited back to perform a retention test after a relatively long period (roughly 3 to 10 months) without exposure to the tactile codes or the TAPS device. The test consisted of one block of 60 trials of the identification test at L9. The participants were allowed to review all the training zones for as long as they needed before completing the 49-code identification test.

E. Data Analysis

Daily training times per training zone were recorded for each participant during the time spent in the Training School. For the code identification tests conducted in the Lab and the retention tests, response time (measured from the offset of the stimulus to the onset of the response) was logged for every trial.

For the 3 blocks of 60 trials at L9, the 180 trials from all participants were pooled into one stimulus-response confusion matrix with a total of 1800 trials (180 trials $\times$ 10 participants). This matrix was processed to obtain an overall percent correct (PC) score and a conservative lower-bound for information transfer (IT_{pc}). Due to the relatively small number of trials collected (1800) compared to the size of the confusion matrix (49$\times$49) and the high level of recognition rates, we used a lower-bound estimate of $IT_{pc} = (1 - 2e) \cdot \log_2 k$, where k=49 was the number of stimuli, and $e = 1 - PC$ was the error rate (see [19] for an explanation of the lower bound calculation).

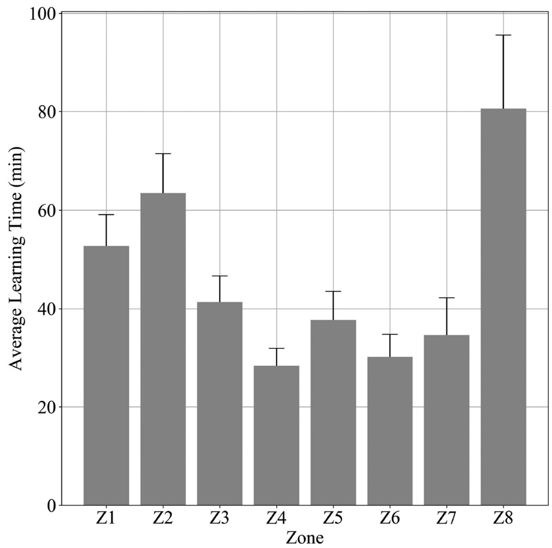


Fig. 8. Average learning time per training zone. Error bars denote +1 standard errors.

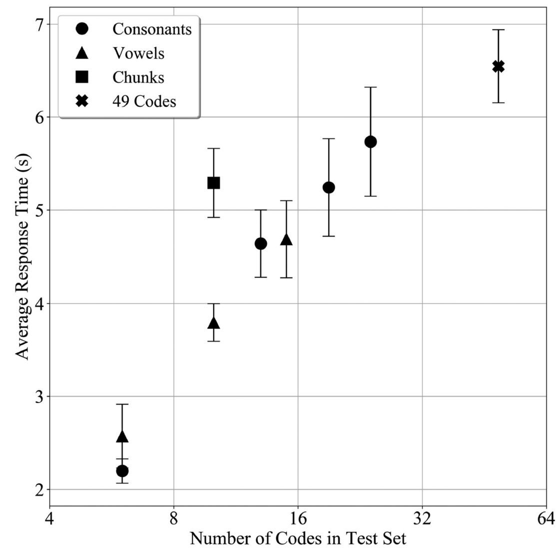


Fig. 9. Average response time for training levels L1 to L9. Error bars denote ± 1 standard errors.

The retention test results were analyzed by calculating the average retention PC score, time spent in the Training School to review tactile codes in all the training zones, and the observed decrease in PC compared to the average PC score at the end of the main experiment at level L9. Statistical analysis involving t-tests on PC scores were performed on arcsin-transformed percent-correct scores.

III. Results

A. Main Experiment

Figure 8 shows the average learning time the participants spent in each training zone. From the plot, it can be observed that the participants spent more time in Z1, Z2 and Z8 than in other zones. The participants spent more time in Z1 and Z2 for both the initial learning of the phonemes in these two zones and for later review of these phonemes when they reached L3, L4, and L9 (see Table IV for the zones included at each learning level). It appears that vowels were easier to learn as indicated by the relatively less time spent in zones Z5 to Z7. Finally, the participants spent more time in Z8 presumably due to several reasons. First, more time was needed to learn the new tactile "chunks" representing phoneme pairs. Second, the participants were aware that level L8 tested only the codes in zone Z8 so they spent the majority of the learning time reviewing the codes in Z8. Third, when the participants were tested at L9, they were observed to spend more time reviewing Z8 than any other zones. All this contributed to a relatively longer learning time spent in this zone.

Figure 9 shows the average response time across participants as a function of number of codes in a test set for different groups of codes tested (Consonants: L1 to L4; Vowels: L5 to L7; Chunks: L8; all 49 codes at the end of the main experiment: L9; see Table IV). The response times for 49 codes were averaged across 10 participants (1800 trials in total). Overall, there appeared to be a trend for the average response

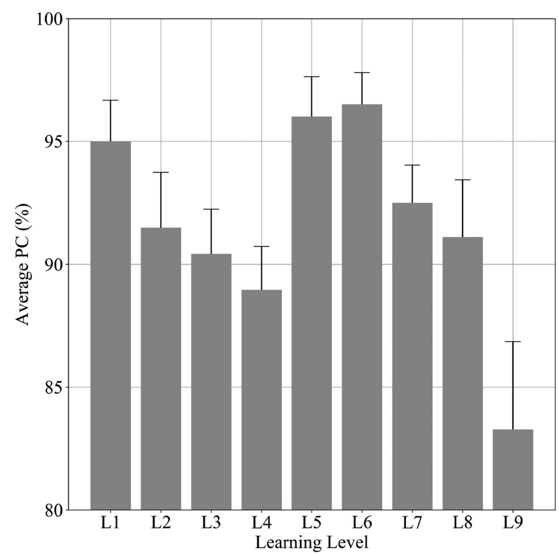


Fig. 10. Average PC scores at each learning level. Error bars denote +1 standard errors.

time to increase almost linearly with the log of the number of codes in the stimulus set, except for Chunks. A linear regression of the data points in Figure 9 showed a significant linear effect of the slope 1.2 $s/\log_2 n$ $(t(89) = 33.7, p < 0.001)$ where n is the number of codes. Each doubling of the number of codes led to a 1.2 s increase in the average response time.

In terms of identification performance, the participants generally passed the code identification tests with no more than two repeated tests throughout the whole curriculum, with the exception of participant P05, who repeated six tests by the end of the main experiment. The average PC scores at each Learning Level are shown in Figure 10. The figure shows scores from tests that passed the criterion of $\geq 80\%$. The results for L9 were pooled from the last 3 60-trial blocks of all participants. It can be observed that the participants performed well

TABLE V
RECOGNITION RESULTS AND STANDARD ERRORS FOR ALL 49 CODES FROM THE MAIN EXPERIMENT AND RETENTION TESTS

Test Case	Avg. PC (%)	Avg. RT (s)	Time (min)
Main exp. (N=10)	83.3±3.6	6.5±0.4	369.1±50.7*
Main exp. (N=7)	83.7±4.4	6.4±0.4	359.4±66.4*
Retention (N=7)	75.7±2.3	7.2±0.5	37.4±4.5**

*Total time spent in the Training School for learning
**Total time spent in the Training School for reviewing

at each level, with average scores ranging from 83.3% (in L9) to 96.5% (in L6).

Table V shows the recognition results for all 49 tactile codes at L9 in terms of the average PC score, average response time (RT) and the total time it took to achieve these results. The results for the ten participants in the main experiment are reported in Table V. It can be seen from the first row that the ten participants achieved a recognition accuracy of 83.3% with 49 tactile codes. The recognition accuracy for native and non-native English speakers was 79.9% (±5.5 std. err., $n = 6$) and 88.3% (±2.6 std.err., $n = 4$), respectively. A two-sample t-test showed no significant difference between the two groups ($t(8) = -1.00, p = 0.346$). The average response time was 6.5 s and the total learning time was 369.1 min, or roughly 6.2 hours. The IT_{PC} for the ten participants was 3.8 bits (derived from $\lfloor 2^{IT_{PC}} \rfloor$), indicating that the participants could identify 13 codes without any error.

B. Retention Test

Seven of the ten participants from the main experiment were able to return to the laboratory to perform the retention test. The periods of inactivity, defined as the number of days from the last session of the main experiment to the retention test, ranged between 90 and 300 days across the seven returning participants. The retention test session lasted one hour for every participant which included both review and test times.

The performance decrease in PC scores for each participant as a function of the respective inactivity period, shown in Figure 11, ranged from 6.7% to 23.3% with no clear correlation with the inactivity period. A linear regression showed a decrease of -0.03% per day with a poor fit, ($R^2 = 0.113$) and the trend was not significant ($t(5) = -0.80, p = 0.462$).

The second and third rows in Table V provide the performance comparisons of the seven participants at the end of the main experiment and retention tests. It can be seen that the performance of the seven participants in the main experiment (second row in Table V) was similar to that of the ten participants in the main experiment (first row). From the main experiment to the retention tests (N=7), the average PC scores dropped significantly from 83.7% to 75.7% by a paired t-test ($t(6) = 2.78, p = 0.032$). Nevertheless, the 8% drop was less than 10% of the original PC score of 83.7%, and the 75.7% retention performance remained significantly above the 2% (1 out of 49 alternatives) chance level ($t(6) = 34.56, p < 0.001$). The slight increase of RT from 6.4 s to 7.2 s was insignificant ($t(6) = -1.71, p = 0.138$). These results indicate that after many months without exposure to TAPS, the seven participants were able to retain their learning of the 49 tactile codes as indicated by the slight albeit significant drop in PC scores and similar response times.

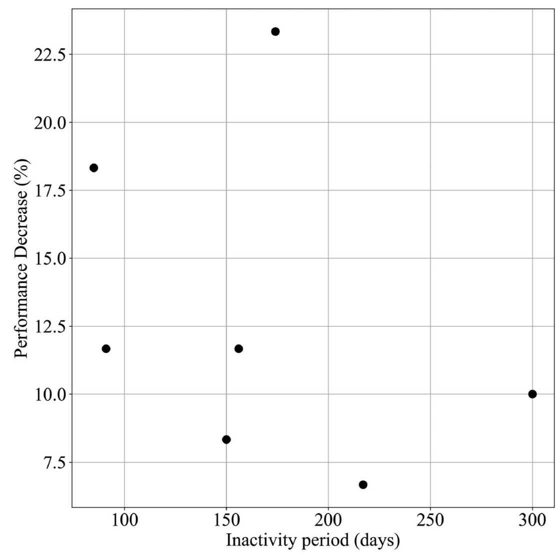


Fig. 11. Performance decrease in the retention test calculated as the PC score at L49 during the main experiment subtracted by the PC score from the retention test for each participant.

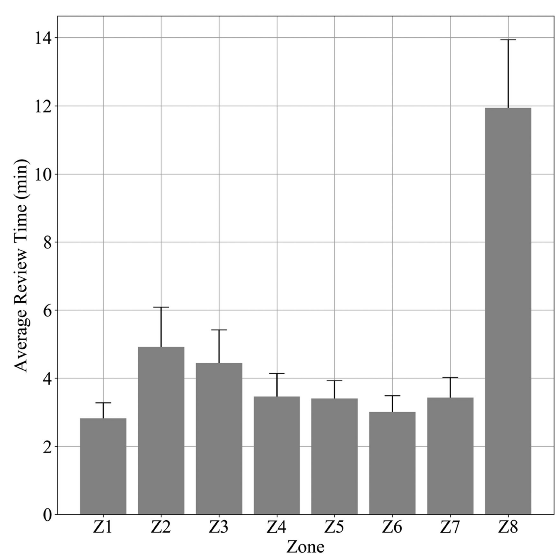


Fig. 12. Average review time per training zone prior to retention tests. Error bars denote +1 standard errors.

From the "Time" column in Table V, it can be seen that the seven participants spent 37.4 minutes on average reviewing the 49 codes before taking the retention tests. The distribution of the review time per training zone is shown in Figure 12. It can be observed that each participant needed less than 6 minutes on average to review the phonemes in each of the first seven zones. A longer time of 12 minutes was needed to review the 10 additional chunk codes from zone Z8, suggesting that the two-phoneme codes may have been less memorable than the single-phoneme codes.

TABLE VI
COMPARISON BETWEEN THE PRESENT STUDY AND AN EARLIER STUDY OF OURS ON TAPS

	Reed et al. [11]	**Present Study**
Number of Participants	10	10
Number of Codes	39	49
Total Number of Trials	1,560	1,800
Total Learning Time (hours)	1 to 4	6.2
PC Score (%)	85.8	83.3
Response Time (RT) (s)	4.2	6.5
RT Cost per Code ($s/\log_2(n)$)	0.07	1.2

IV. DISCUSSION

Our results validated the effectiveness of the new set of tactile codes that were designed based on the statistical properties of spoken English. The set included tactile codes for the 39 individual phonemes of English, revised to associate the shortest duration to the 13 most common phonemes, as well as 10 new tactile codes representing the 10 most frequently co-occurring two-phoneme pairs. The participants demonstrated the ability to identify the full set of 49 codes with above 80% accuracy following 6.2 hours of training. These results are promising for increasing communication rates through TAPS, our phonemic-based tactile display.

The results from the main experiment of the present study can be compared to those from our previous work on a similar set of 39 tactile codes encoding the 39 English phonemes by Reed *et al.* [11], as tabulated in Table VI. The PC scores in the two studies are not significantly different from each other ($t(9) = -0.29, p = 0.776$), suggesting that the modifications to the tactile codes and the addition of 10 two-phoneme codes did not impact the learning performance. However, the response time for the 10 participants in the present study with 49 codes was significantly longer than those reported by Reed *et al.* [11] ($t(9) = 5.97, p < 0.001$). This suggests that the augmented set of 49 tactile codes required an additional 2.3 s on average to process when the participants performed code identification. Furthermore, the response time results from both studies show an increasing trend as the number of codes in the stimulus set increases. While the results from the present study show an increase of 1.2 s in processing time whenever the number of codes in the test is doubled, the results from Reed *et al.* [11] show an increase of only 0.07 s per doubling of the number of codes. It thus appears that the modifications to the original 39 phonemic codes designed by Reed *et al.* [11] and the addition of 10 new tactile signals led to longer processing time. These comparisons reflect the challenge associated with learning a larger and more complex set of tactile codes. Differences in the grouping of codes in the learning curriculum and user interface for learning and testing may have also played a role and requires further investigation. Overall, the 6.2 hours spent by the participants in the present study demonstrate that the new set of codes can still be learned within a reasonable amount of time.

Results from the retention tests show that the participants were able to maintain code recognition accuracy at levels comparable to those achieved at the end of the main experiment after a period of 3 to 10 months of non-exposure to TAPS. It took the participants a little over half an hour on average to review the set of 49 tactile codes before taking the retention test. This attests that the set of new codes is highly memorable.

To quantify the potential increase in speech transmission rate with the new codes developed in the present study, we calculated the total time it would take to transmit the sentence "*The time it takes for you to learn and practice is in the order of hours*" (constructed based on the words found in https://www.talkenglish.com/vocabulary/top-2000-vocabulary.aspx, accessed on Dec. 18, 2019). This 16-word phrase contains some of the most frequently occurring words in spoken English. Without loss of generality, the total duration of the phrase was computed as the sum of the duration of all phonemes in the sentence with no inter-phoneme intervals inserted between adjacent phonemes or words. The sentence duration calculated with the 49 tactile codes developed in the present study was 9.5 s (101 wpm) where the 10 most frequently co-occurring phoneme pairs were coded with the 10 two-phoneme chunks. Using only the 39 single-phoneme codes from the present study or those from Reed *et al.* [11], the sentence duration was 11.4 s (84 wpm) and 15 s (64 wpm), respectively. Thus the speech transmission rate with the 49 codes in the present study represents a 58% increase over that with the original 39 codes developed by Reed *et al.* in [11].

Our previous work by Reed *et al.* [11] and the present study provide ample proof that it is feasible to transmit all 39 English phonemes, paving the way for transmitting any English word using our phonemic-based tactile display TAPS. In comparison, many previous studies have used a partial set of the English alphabet or phonemes and thus were unable to address the question of whether the acquisition of all alphabets or phonemes is achievable. For example, Zhao *et al.* encoded 9 phonemes with 6 tactors and reported a phoneme recognition rate of 88%[20]. Turcott *et al.* encoded 10 codes and demonstrated that the phonemic approach outperformed other acoustic encodings in word identification tests [21]. Other studies do not test identification of tokens in isolation but report the identification of tokens in words instead. For example, Dunkelberger *et al.* encoded 23 phonemes using a combination of four tactors, one radial squeeze band, and one lateral skin stretch rocker on the arm. They trained their participants on 150 words and reported a word recognition rate of 87% with 50 test words [22]. Finally, Luzhnica et al.'s "skin reading" tactile glove encoded the 26 letters of the alphabet using 6 tactors. They reported a letter recognition rate of $> 90\%$ with 48 words after 300 min of training [23]. In comparison, our recent results reported in Tan *et al.* [13] provide evidence of the acquisition of up to 500 English words using TAPS, with the best participants achieving a learning rate of one English word per minute.

The compilation of our previous studies and this work motivates the continued use of TAPS as a promising device to encode the English language for the use of any person, regardless of their sensory capabilities. Thus far, our research has focused on ascertaining that our phonemic-based encoding approach works well with normal-hearing individuals who have a good grasp of the English language, and to establish the maximum rate at which speech communication can be achieved via TAPS. Future work will involve participants from the hearing-impaired and deaf-blind communities as potential users of TAPS. For many of them, English may not be their first language, and they may instead be most proficient with a signed-language communication system such as the American Sign Language. Other methods of tactile communication are under development for people who are deaf-and-blind, including new approaches to Braille [24], alphabetic codes [23], tactile icons [25], [26], navigational cues [27] and supplementary exploration devices [28]. However, the goal of the current research is communication of English. Hence, before introducing such potential users to the TAPS system, we first need to ensure that we have established baseline performance levels with normal-hearing individuals with proficient levels of English including both native and non-native English speakers. We do not want to confound linguistic ability with the capabilities of the TAPS device in evaluating its performance. We are now ready to tackle the problem of devising a coding scheme for a phonemic-based haptic system and developing an effective training protocol with the TAPS system, with the goal for addressing the communication needs of people with sensory impairments.

In conclusion, the present study improved upon our previous phonemic-based coding scheme that encodes all 39 English phonemes by reducing phoneme transmission time using the statistics of spoken English. Our results show that the new set of 49 codes is highly distinct and memorable and can be learned in a reasonable amount of time. Our work promises to increase speech transmission rates significantly using the TAPS system. Future work will assess the ability to receive English words, phrases and sentences using the 49 tactile codes developed in the present study.

Acknowledgment

The authors would like to thank Jaehong Jung for providing feedback on the design of the tactile codes.

References

[1] C. M. Reed, W. M. Rabinowitz, N. I. Durlach, L. D. Braida, S. Conway-Fithian, and M. C. Schultz, "Research on the tadoma method of speech communication," *J. Acoust. Soc. Amer*, vol. 77, no. 1, pp. 247–257, 1985.
[2] C. M. Reed, "Tadoma: An overview of research," in *Proc. Profound Deafness Speech Commun.*, 1995, pp. 40–55.
[3] S. Engelmann and R. Rosov, "Tactual hearing experiment with deaf and hearing subjects," *Exceptional Children*, vol. 41, no. 4, pp. 243–253, 1975.
[4] D. Franklin, "Tactile aids, new help for the profoundly deaf," *Hearing J.*, vol. 37, no. 2, pp. 20–23, 1984.
[5] E. T. Auer Jr., L. E. Bernstein, and D. C. Coulter, "Temporal and spatio-temporal vibrotactile displays for voice fundamental frequency: An initial evaluation of a new vibrotactile speech perception aid with normal-hearing and hearing-impaired individuals," *J. Acoust. Soc. Amer*, vol. 104, no. 4, pp. 2477–2489, 1998.
[6] F. A. Geldard, "Adventures in tactile literacy," *Amer Psychologist*, vol. 12, no. 3, pp. 115–124, 1957.
[7] Y. Gaffary, F. Argelaguet, M. Marchal, A. Girard, F. Gosselin, M. Emily, and A. Lécuyer, "Toward haptic communication: Tactile alphabets based on fingertip skin stretch," *IEEE Trans. Haptics*, vol. 11, no. 4, pp. 636–645, Oct.–Dec. 2018.
[8] J. M. Weisenberger and M. E. Percy, "Use of the Tactaid II and Tactaid VII with children," *Volta Rev.*, vol. 96, no. 5, pp. 41–57, 1994.
[9] C. M. Reed and L. A. Delhorne, "Current results of a field study of adult users of tactile aids," *Seminars Hearing*, vol. 16, no. 04, pp. 305–315, 1995.
[10] C. M. Reed and L. A. Delhorne, "The reception of environmental sounds through wearable tactual aids," *Ear Hearing*, vol. 24, no. 6, pp. 528–538, 2003.
[11] C. M. Reed *et al.*, "A phonemic-based tactile display for speech communication," *IEEE Trans. Haptics*, vol. 12, no. 1, pp. 2–17, Jan.–Mar. 2019.
[12] Y. Jiao *et al.*, "A comparative study of phoneme- and word-based learning of English words presented to the skin," in *Haptics: Science, Technology, and Applications*, D. Prattichizzo, H. Shinoda, H. Z. Tan, E. Ruffaldi, and A. Frisoli, Eds., Berlin, Germany: Springer, 2018, pp. 623–635.
[13] H. Z. Tan *et al.*, "Acquisition of 500 English words through a TActile phonemic sleeve (TAPS)," *IEEE Trans. Haptics*, to be published, doi: 10.1109/TOH.2020.2973135.
[14] J. J. Shepherd, R. J. Doe, M. Arnold, Y. Zhu, and J. Tang, "A different approach to teaching Chinese through serious games," in *Proc. 6th Int. Conf. Found. Digital Games*, 2011, pp. 304–306.
[15] K. P. Jantke and T. Hume, "Effective learning through meaning construction in digital role playing games," in *Proc. IEEE Int. Conf. Consum. Electron.*, 2015, pp. 653–656.
[16] S. J. Lederman and L. A. Jones, "Tactile and haptic illusions," *IEEE Trans. Haptics*, vol. 4, no. 4, pp. 273–294, Oct.–Dec. 2011.
[17] P. B. Denes, "On the statistics of spoken English," *J. Acoust. Soc. Amer.*, vol. 35, no. 6, pp. 892–904, 1963.
[18] M. A. Mines, B. F. Hanson, and J. E. Shoup, "Frequency of occurrence of phonemes in conversational english," *Lang. Speech*, vol. 21, no. 3, pp. 221–241, 1978.
[19] H. Z. Tan, N. I. Durlach, C. M. Reed, and W. M. Rabinowitz, "Information transmission with a multifinger tactual display," *Percep-tion Psychophysics*, vol. 61, no. 6, pp. 993–1008, 1999.
[20] S. Zhao, A. Israr, F. Lau, and F. Abnousi, "Coding tactile symbols for phonemic communication," in *Proc. CHI Conf. Human Factors Comput. Syst.*, 2018, pp. 392:1–392:13.
[21] R. Turcott *et al.*, "Efficient evaluation of coding strategies for transcutaneous language communication," in *Proc. Haptics: Sci., Technol. Appl.*, 2018, pp. 600–611.
[22] N. Dunkelberger *et al.*, "Conveying language through haptics: A multi-sensory approach," in *Proc. ACM Int. Symp. Wearable Comput.*, 2018, pp. 25–32.
[23] G. Luzhnica, E. Veas, and V. Pammer, "Skin reading: Encoding text in a 6-channel haptic display," in *Proc. ACM Int. Symp. Wearable Comput.*, 2016, pp. 148–155.
[24] J. Rantala *et al.*, "Methods for presenting braille characters on a mobile device with a touchscreen and tactile feedback," *IEEE Trans. Haptics*, vol. 2, no. 1, pp. 28–39, Jan–Mar. 2009.
[25] A. Carrera, A. Alonso, R. De la Rosa, and E. J. Abril, "Sensing performance of a vibrotactile glove for deaf-blind people," *Appl. Sci.*, vol. 7, no. 4, p. 317, 2017.
[26] T. McDaniel, S. Krishna, D. Villanueva, and S. Panchanathan, "A haptic belt for vibrotactile communication," in *Proc. IEEE Int. Symp. Haptic Audio Visual Environ. Games*, 2010, pp. 1–2.
[27] T. Amemiya, J. Yamashita, K. Hirota, and M. Hirose, "Virtual leading blocks for the deaf-blind: A real-time way-finder by verbal-nonverbal hybrid interface and high-density RFID tag space," in *Proc. IEEE Virtual Reality*. 2004, pp. 165–287.
[28] C. Asakawa, H. Takagi, S. Ino, and T. Ifukube, "TAJODA: Proposed tactile and jog dial interface for the blind," *IEICE Trans. Inf. Syst.*, vol. 87, no. 6, pp. 1405–1414, 2004.

Juan S. Martinez (Member, IEEE) received the B.A.Sc. degree in electronic engineering and the B.A.Sc. degree in systems and computing engineering from Los Andes University, Bogotá, Colombia, in 2016 and 2017, respectively, and the M.S. degree in electrical and computer engineering from Purdue University, West Lafayette, IN, USA, in 2019. He is currently working toward the Ph.D. degree with Purdue University, West Lafayette, IN, USA, where he works as a Research Assistant with the Haptic Interface Research Lab, advised by Dr. H. Z. Tan.

Hong Z. Tan (Fellow, IEEE) received the bachelor's degree in biomedical engineering from Shanghai Jiao Tong University, Shanghai, China, in 1986, and the master's and Doctorate degrees in electrical engineering and computer science from the Massachusetts Institute of Technology Cambridge, MA, USA, in 1988 and 1996, respectively. She is currently a Professor of electrical and computer engineering, mechanical engineering (by courtesy) and psychological sciences (by courtesy) with Purdue University, West Lafayette, IN, USA. She has served three terms as an Associate Editor for the IEEE TRANSACTIONS ON HAPTICS (2007 to 2012; 2016 to 2019) and received a meritorious service award in 2012.

Charlotte M. Reed received the Bachelors of Science degree in education from Carlow College, Carlow, Ireland, in 1969 and the Ph.D. degree in bioacoustics from the University of Pittsburgh, Pittsburgh, PA, USA, in 1973. She is currently a Senior Research Scientist with the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA, USA.