A Phonemic-Based Tactile Display for Speech Communication

Charlotte M. Reed, Hong Z. Tan, *Fellow, IEEE*, Zachary D. Perez, E. Courtenay Wilson, Frederico M. Severgnini, Jaehong Jung, Juan S. Martinez, Yang Jiao, Ali Israr, Frances Lau, Keith Klumb, Robert Turcott, and Freddy Abnousi

***Abstract*—Despite a long history of research, the development of synthetic tactual aids to support the communication of speech has proven to be a difficult task. The current paper describes a new tactile speech device based on the presentation of phonemic-based tactile codes. The device consists of 24 tactors under independent control for stimulation at the forearm. Using properties that include frequency and waveform of stimulation, amplitude, spatial location, and movement characteristics, unique tactile codes were designed for 39 consonant and vowel phonemes of the English language. The strategy for mapping the phonemes to tactile symbols is described, and properties of the individual phonemic codes are provided. Results are reported for an exploratory study of the ability of 10 young adults to identify the tactile symbols. The participants were trained to identify sets of consonants and vowels, before being tested on the full set of 39 tactile codes. The results indicate a mean recognition rate of 86 percent correct within one to four hours of training across participants. Thus, these results support the viability of a phonemic-based approach for conveying speech information through the tactile sense.**

***Index Terms*—Human haptics, speech communication, phoneme codes, human performance, tactile devices, tactile display, rehabilitation.**

I. Introduction

THE sense of touch has evolved to provide humans with information about environmental stimuli through the reception of pressure, pain, and temperature changes as well as internal sensations providing kinesthetic information about positions and movements of the limbs [1]. The tactual sensory system has also been utilized as a channel of human communication, as in situations where the more typical channels of audition and sight are absent, compromised, or overburdened. For example, methods of tactual communication have arisen out of necessity within the community of deaf-blind individuals and their educators, as a means of conveying language in the absence of either visual or auditory input [2]. Over the years, a variety of methods of tactual communication have been employed as substitutes for hearing and/or vision. These include natural methods of tactual communication such as the Tadoma method of speechreading [3], as well as the tactual reception of fingerspelling [4] and sign language [5]. Generally, these three methods may be thought of as tactual adaptions of visual methods of communication used by sighted persons with profound auditory impairment.

Concurrent with the use of natural methods of tactual communication, there is also a long history of research on the development of artificial devices designed to convey acoustic information through the tactual sense (e.g., see older reviews [6], [7], [8] as well as research that continues to the present day [9], [10], [11]). These devices generally attempt to convey characteristics of the acoustic speech signal through tactual patterns generated on arrays of stimulators. From a signal-processing point of view, many devices have attempted to display spectral properties of speech to the skin. These displays rely on the principle of frequency-to-place transformation, where location of stimulation corresponds to a given frequency region of the signal [12]. Another approach to signal processing has been the extraction of speech features (such as voice fundamental frequency and vowel formants) from the acoustic signal prior to encoding on the skin [13]. For both classes of aids, devices have included variations in properties such as number of channels, geometry of the display, body site, transducer properties, and type of stimulation (e.g., vibrotactile versus electrotactile).

To date, however, no wearable tactile aid has yet been developed that is capable of allowing users to receive speech at levels comparable to those achieved by users of the Tadoma method of speechreading. Thus, the speech-reception results reported for Tadoma users may serve as a benchmark in the development of future tactile communication devices. Using only manual sensing of cues that are available on the face and neck of a talker during speech production (such as airflow, lip and jaw movements, vibration on the neck), proficient users of Tadoma are able to receive connected speech at a rate of

Manuscript received April 19, 2018; revised July 24, 2018; accepted July 25, 2018. Date of publication July 31, 2018; date of current version March 28, 2019. This paper was recommended for publication by Associate Editor M. L. Kappers upon evaluation of the reviewers' comments. This work was partially supported by a research grant funded by Facebook Inc. *(Corresponding: Charlotte M. Reed.)*

C. M. Reed, Z. D. Perez, and E. C. Wilson are with the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA (e-mail: cmreed@mit.edu; zperez@mit.edu; ecwilson@mit.edu).

H. Z. Tan, F. M. Severgnini, J. Jung, J. S. Martinez, and Y. Jiao 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: hongtan@purdue.edu; fmarcoli@purdue.edu; jung137@purdue.edu; mart1304@purdue.edu; jiao12@purdue.edu).

A. Israr, F. Lau, K. Klumb, R. Turcott, and F. Abnousi are with Facebook Inc., Menlo Park, CA 94025, USA (e-mail: aliisrar@fb.com; flau@fb.com; kklumb@fb.com; rturcott@fb.com; abnousi@fb.com).

Digital Object Identifier 10.1109/TOH.2018.2861010

approximately 60 to 80 words/min [14]. This rate, which is roughly one-third of that for normal auditory reception of speech, corresponds to an information transfer rate on the order of 12 bits/s [15]. Although the reception of speech segments through tactile devices is similar to that obtained with Tadoma [16], [17], these devices have shown minimal performance for reception of connected speech, and are often evaluated as aids to speechreading [18], [19].

Inspired by the success of Tadoma, artificial displays have been developed to mimic various properties of the Tadoma display. Research with an artificial talking face [20], [21] demonstrated that the discrimination and identification of speech segments by naïve observers with this display compared favorably to results obtained through Tadoma by experienced deaf-blind users of the method. A more stylized version of Tadoma properties was incorporated into another device referred to as the Tactuator [22], which consisted of three bars capable of stimulating the thumb, index, and middle fingers over a frequency range that included kinesthetic as well as cutaneous stimulation. Experiments conducted with this device demonstrated an information transfer rate of 12 bits/s for sets of multidimensional stimuli, similar to that estimated for speech reception by experienced deaf-blind users of Tadoma.

Despite these promising laboratory results, there is still a need for the development of wearable tactile devices as aids to communication. Such devices would have applications to a broad range of situations where input to the auditory and/or visual sense is absent or compromised, or when these sensory channels are overloaded in performing other tasks. The tactile sense can then serve as an additional communication channel for applications for persons with normal sensory abilities, such as human-computer interfaces and remote communication, in addition to being used as communication aids for persons with sensory disabilities of deafness and/or blindness. For the current study, a decision was made to use a phonemic-based approach to encoding speech. Other approaches are also worthy of consideration, including the use of alphabetic codes [23] and tactile icons [24], [25]. Advantages of the phonemic approach include the greater efficiency of phonemic versus textual codes [15] and its ability to encode any possible word or message in the language as opposed to the use of tactile icons which must be adapted to particular situations.

In the light of advancements in several technical areas, an opportunity exists to develop and evaluate a new generation of tactile devices which may be capable of significant improvements for speech reception. These advancements include developments in (1) technological areas such as signal processing and haptic displays [26]; (2) the field of automatic speech recognition (ASR) [27]; and (3) approaches to training and learning in the use of novel displays [28].

The current paper describes a new tactile speech communication device which is based on a multi-channel array applied to the forearm. The display, which is wearable but tethered to equipment that has yet to be miniaturized, employs a phonemic-based approach to the encoding of speech stimuli. This approach assumes that ASR can be employed at the front end of the device to recognize speech stimuli and to encode them as strings of phonemes. Thus, a set of distinct tactile symbols was created to represent the individual consonant and vowel phonemes of English. Work on the development and evaluation of the display is described in the following sections of the paper as follows. Section 2 is concerned with describing the design of the tactile device. Section 3 describes the manner in which phonemes were mapped to vibratory patterns on the tactile array. Section 4 reports results of experiments conducted to train naïve participants on the identification of the tactile symbols. Section 5 provides a discussion of the work on tactile coding of phonemes. Finally, Section 6 provides conclusions and directions for future research.

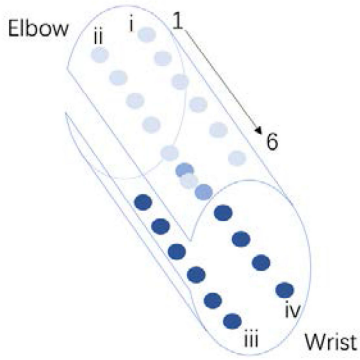


Fig. 1. Schematic illustration of the layout of the tactors in the experimental device.

II. Design of Tactile Device

The tactile device consists of a 4-by-6 tactor array worn on the forearm. The 24 tactors form four rows in the longitudinal direction (elbow to wrist) and six columns (rings) in the transversal direction (around the forearm). As shown in Fig. 1, two rows (i and ii) reside on the dorsal side of the forearm and the other two (iii and iv) on the volar side. The tactors (Tectonic Elements, Model TEAX13C02-8/RH, Part #297-214, sourced from Parts Express International, Inc.) were wide-bandwidth audio exciters with a constant impedance of $\approx 8\ \Omega$ in the frequency range of 50 to 2k Hz, except for a peak in the vicinity of 600 Hz. Each tactor measured about 30 mm in diameter and 9 mm in thickness. The current study used sinusoidal waveforms at 60 and 300 Hz, sometimes with an amplitude modulation at 8 or 30 Hz. We attached an accelerometer (Kistler 8794A500) to the tactors and ascertained that the tactors were able to respond to the driving waveforms at these frequencies.

A Matlab program generated 24 independently programmable waveforms and temporal onsets and offsets using the multichannel playrec utility (http://www.playrec.co.uk/index.html) running on a desktop computer. A MOTU USB audio device (MOTU, model 24Ao, Cambridge, MA, USA) received the 24-channel signal via the computer's USB port, performed synchronous digital-to-analog conversion of the signals, and sent the 24 audio waveforms through its 24 channels of analog output connectors to three custom-built amplifier boards. Each amplifier board catered 8 audio channels, and passed audio waveforms through four class D stereo amplifiers (Maxim,

Model MAX98306, sourced from Adafruit, New York, USA) to drive eight tactors independently. Each tactor was mounted with a temperature sensor that measured temperature of the tactor's chassis. An on-board protective circuit turned off the power supply and sounded an alarm if the temperature of any sensor rose above 50° C or the current drawn by any amplifier module (supporting two tactors) exceeded 600 mA. We verified with the same accelerometer (Kistler 8794A500) that the tactor responses followed the signal waveforms and did not saturate or clip at the maximum amplitudes allowed in the Matlab program.

The stimulus properties that were controlled by the software included amplitude (specified in dB sensation level, or dB above individually-measured detection thresholds), frequency (single or multiple sinusoidal components), waveform (such as temporal onset/offset characteristics and the use of amplitude modulation), duration, location, numerosity (single tactor activation or multiple tactors turned on simultaneously or sequentially), and movement (smooth apparent motion or discrete saltatory motion varying in direction, spatial extent, and trajectory).

III. Mapping of Phonemes to Tactile Symbols

The International Phonetic Alphabet (IPA) symbols of the 39 English phonemes that were coded for delivery through the tactile display are provided in Table I for consonants (24 phonemes) and Table II for vowels (15 phonemes). A unique vibrotactile pattern was mapped on the 4-by-6 array of tactors for each of the 39 phonemes. The mapping of the phonemes to tactile symbols was guided by three primary considerations, which included (1) the psychophysical properties of the tactile sensory system, (2) various articulatory properties of the phonemes, and (3) the need to generate a set of perceptually distinct tactile signals.

A major challenge in the development of tactile speech-communication devices lies in encoding the processed speech signals to match the perceptual properties of the skin. Compared to the sense of hearing, the tactile sensory system has a reduced frequency bandwidth (20-20,000 Hz for hearing compared to 0-1000 Hz for touch), reduced dynamic range (115 dB for hearing compared to 55 dB for touch), and reduced sensitivity for temporal, intensive, and frequency discrimination (see [29]). The tactile sense also lags behind the auditory sense in terms of its capacity for information transfer (IT) and IT rates [15]. For example, communication rates of up to 50 words/min are achieved by experienced operators of Morse code through the usual auditory route of transmission, compared to 25 words/min for vibrotactile reception of these patterns [30]. Taking these properties of the tactile sense into account, certain principles may be applied to create displays with high IT rate. One such principle is to include as many dimensions as possible in the display, while limiting the number of variables along each dimension [31], [32].

After an initial set of tactile codes was developed for the 39 phonemes, informal observations on the distinctiveness of the stimuli were made by members of the laboratory staff, and an iterative process was conducted to make adjustments to the codes to enhance their discriminability. Among these considerations were balancing the use of tactors across the transverse and longitudinal dimensions of the array. The sets of codes that were developed for use in the current study are described below for consonants (Section 3.1) and vowels (Section 3.2).

A. Consonant Codes

A description of the tactile codes generated for the 24 consonants (defined by IPA symbols and the phoneme codes adopted for the current study) is provided in Table I. Articulatory properties, including manner and place of articulation and voicing, were taken into consideration in the development of the tactile codes. The phonemes are organized in Table I by manner of articulation: six plosives (P, B, T, D, K, G), eight fricatives (F, V, TH, DH, S, Z, SH, ZH), two affricates (CH, J), three nasals (M, N, NG), and five semivowels (H, W, R, L, Y). In Table I, the consonant codes are described in terms of their waveforms, location on the tactile array, number of activated tactors, duration, and the tactors involved by their locations (see Fig. 1 for how we label tactor locations). A schematic description of the patterns generated on the tactile display for each of the consonant phonemes is provided in Fig. 2.

Two values of vibrational frequency (60 and 300 Hz) and two values of duration (100 and 400 ms) were used to code manner of articulation for the first four classes of sounds (from P to NG in Table I), all of which were coded with the use of four tactors. A duration cue was used to distinguish the plosives (100 ms) from the other classes of phonemes (400 ms). Frequency of vibration was used to distinguish the nasals (60 Hz) from the plosives, fricatives, and affricates.

Within each of these four classes of phonemes, place of articulation was generally mapped along the longitudinal direction of the display such that sounds made in the front of the mouth were presented near the wrist (e.g., P, B, F, V, M), those in middle of the mouth were presented in the middle of the forearm (e.g., T, D, TH, DH, N), and those made in the back of the mouth were presented near the elbow (e.g., K, G, NG). These rules were occasionally violated in order to create distinct signals (e.g., S and Z were presented at the elbow despite their alveolar place of articulation). The affricates (CH and J) were place-coded with stimulation at both the wrist and elbow to represent the change in place of articulation from alveolar to velar. Note that place was coded only at three positions on the longitudinal dimension of the array such that the two tactors at the wrist, the two at the middle of the forearm, and the two at the elbow, respectively, were always activated simultaneously. Likewise, along the transverse dimension of the array, position was coded using only dorsal versus volar location. This was done to ensure that the spacing between the tactors along both dimensions greatly exceeded the two-point limen reported for the forearm [33], [34]. (See Section 4.2 for further information regarding the spacing of tactors on the array).

To distinguish voiced from unvoiced cognates, amplitude modulation was applied to the sinusoidal frequency in generating the voiced consonants. In the case of the voiced plosives

TABLE I
DESCRIPTION OF TACTILE CODES DEVELOPED FOR 24 CONSONANT PHONEMES

IPA Symbol	Phoneme Code	Waveform		Location		# of Tactors	Duration (ms)	Tactors Involved
		Frequency (Hz)	Modulation (Hz) or Shaping	Transverse	Longitudinal	(Simultaneous)		
/p/	P	300		Dorsal	Wrist	4	100	i5, i6, ii5, ii6
/b/	B	300	30	Dorsal	Wrist	4	100	i5, i6, ii5, ii6
/t/	T	300		Volar	Mid-Forearm	4	100	iii3, iii4, iv3, iv4
/d/	D	300	30	Volar	Mid-Forearm	4	100	iii3, iii4, iv3, iv4
/k/	K	300		Dorsal	Elbow	4	100	i1, i2, ii1, ii2
/g/	G	300	30	Dorsal	Elbow	4	100	i1, i2, ii1, ii2
/f/	F	300		Dorsal+Volar	Wrist	4	400	i6, ii6, iii6, iv6
/v/	V	300	8	Dorsal+Volar	Wrist	4	400	i6, ii6, iii6, iv6
/ɵ/	TH	300		Dorsal	Mid-Forearm	4	400	i3, i4, ii3, ii4
/ð/	DH	300	8	Dorsal	Mid-Forearm	4	400	i3, i4, ii3, ii4
/s/	S	300		Dorsal+Volar	Elbow	4	400	i1, ii1, iii1, iv1
/z/	Z	300	8	Dorsal+Volar	Elbow	4	400	i1, ii1, iii1, iv1
/ʃ/	SH	300		Volar	Wrist	4	400	iii5, iii6, iv5, iv6
/□/	ZH	300	8	Volar	Elbow	4	400	iii1, iii2, iv1, iv2
/ʧ/	CH	300	cos^2	Dorsal	Wrist+Elbow	4	400	i1, i6, ii1, ii6
/ʤ/	J	300	8	Dorsal	Wrist+Elbow	4	400	i1, i6, ii1, ii6
/m/	M	60	8	Dorsal	Wrist	4	400	i5, i6, ii5 ,ii6
/n/	N	60	8	Volar	Mid-Forearm	4	400	iii3, iii4, iv3, iv4
/ŋ/	NG	60	8	Dorsal	Elbow	4	400	i1, i2, ii1, ii2
/h/	H	60	cos^2	Dorsal+Volar	Mid-Forearm	8	400	i4, i5, ii4, ii5, iii4, iii5, iv4, iv5
/w/	W	60	8	Dorsal+Volar	Wrist-Mid	8	400	i3, i4, i5, i6, iii3, iii4, iii5, iii6
/r/	R	300	30	Volar	Elbow	4	400	iii1, iii2, iv1, iv2
/l/	L	300	30	Volar	Wrist	4	400	iii5, iii6, iv5, iv6
/j/	Y	60		Dorsal+Volar	Wrist-Mid	8	400	i3, i4, i5, i6, iii3, iii4, iii5, iii6

The IPA notation and the orthographic representation for each consonant are provided in columns 1 and 2, respectively. A description of the waveform is provided in columns 3 and 4, location on the tactile array in columns 5 and 6, number of tactors employed in the code in column 7, and stimulus duration in column 8. Finally, the tactors used for each code are provided in the last column. The tactors are defined using the conventions described in the schematic illustration of Fig. 1. The default shaping was a 10-ms Hanning window on/off ramp, except for the cos^2 windows as noted in column 4.

(B, D, G), the 300 Hz waveform was modulated sinuosoidally between full and half amplitude at a rate of 30 Hz, in contrast with their unvoiced counterparts which contained no modulation (P, T, K). This same principle was applied to the cognate pairs of fricatives as described in Table I, where the 300 Hz tone was modulated sinusoidally between full and half amplitude at a rate

TABLE II
DESCRIPTION OF THE TACTILE CODES DEVELOPED FOR 15 VOWELS

IPA Symbol	Phoneme Code	Waveform or Shaping		Location on Array	Movement	# of Tactors	Duration (ms)	Subjective Impression
		Frequency (Hz)	Modulation (Hz)			(In succession)		
/i/	EE	300		Top dorsal row	Longitudinal: Wrist-to-Elbow	6	480	Smooth movement
/ɑ/	AH	60		2 dorsal rows	Longitudinal: Elbow-to-Wrist	6 (on each of 2 rows)	480	Wide movement
/u/	OO	300	30	2 volar rows	Longitudinal: Wrist-to-Elbow	6 (on each of 2 rows)	480	Rumbling motion
/ae/	AE	300		2 dorsal rows; 3 columns near elbow	Use of pulses to create circular motion near elbow	12 (6 tactors activated twice)	480	"Twinkle" Sensation
/ɔ/	AW	300		2 volar rows; 3 columns near wrist	Use of pulses to create circular motion near wrist	12 (6 tactors activated twice)	480	"Twinkle" Sensation
/ɝ/, /ɚ/	ER	300		2 volar rows; 3 columns near elbow	Use of pulses to create circular motion near elbow	12 (6 tactors activated twice)	480	"Twinkle" Sensation
/ʌ/	UH	300		4 rows on 2 columns in front and middle of array	From front to middle of array on 2 columns	2 on each of 4 rows	240	Grabbing sensation near wrist and middle of arm
/I/	IH	300		Top dorsal row	From elbow to middle of forearm across 4 columns	4	240	Quick smooth movement
/ɛ/	EH	300		4 rows on 2 columns in back and middle of array	From elbow to middle of forearm on 2 columns	2 on each of 4 rows	240	Grabbing sensation near elbow and middle of arm
/U/	UU	300	30	2 volar rows	From elbow to middle of forearm across 4 columns	4 on each of 2 rows	240	Quick movement
/eI/	AY	300	30	2 dorsal rows near wrist	From mid-forearm to wrist and back	8	480	Forms tents shape; rumbling sensation
/aI/	I	300	30	2 volar and 2 dorsal rows near elbow	From elbow to mid-forearm and back	2 rows of four columns in each direction	480	Sweeping motion; rumbling sensation
/aU/	OW	300		Top dorsal row	From wrist to elbow; Cutaneous rabbit	3 taps on each of 3 tactors	480	Tapping sensation
/OU/	OE	300	Cos^2 window	2 middle columns; 4 rows	Moves across tactors to create a ring	5 tactors on each of 2 columns (1 tactor activated both at beginning and end of sequence)	480	Smooth circular ring
/ɔI/	OY	300		1 volar row	From elbow to wrist; Cutaneous rabbit	3 taps on each of 3 tactors	480	Tapping sensation

The IPA notation and the orthographic representation for each vowel are provided in columns 1 and 2, respectively. A description of the waveform is provided in columns 3 and 4, location on the tactile array (as described in Fig. 1) in column 5, type of movement in column 6, number of tactors employed in the code in column 7, stimulus duration in column 8, and a description of the subjective impression in column 9. The default shaping was a 10-ms Hanning window on/off ramp, except for the cos^2 windows as noted in column 4.

of 8 Hz for the voiced cognates and left unmodulated for their voiceless counterparts. For the affricate cognate pair, the 300 Hz tone was unmodulated for CH and modulated between full and one-fifth amplitude at a rate of 8 Hz for J.

The final five phoneme codes in Table I represent the semivowels, three of which were coded with the use of 8 tactors at frequency of 60 Hz (H, W, Y) and two with the use of 4 tactors at 300 Hz (R, L). As described in Table I, the codes for these

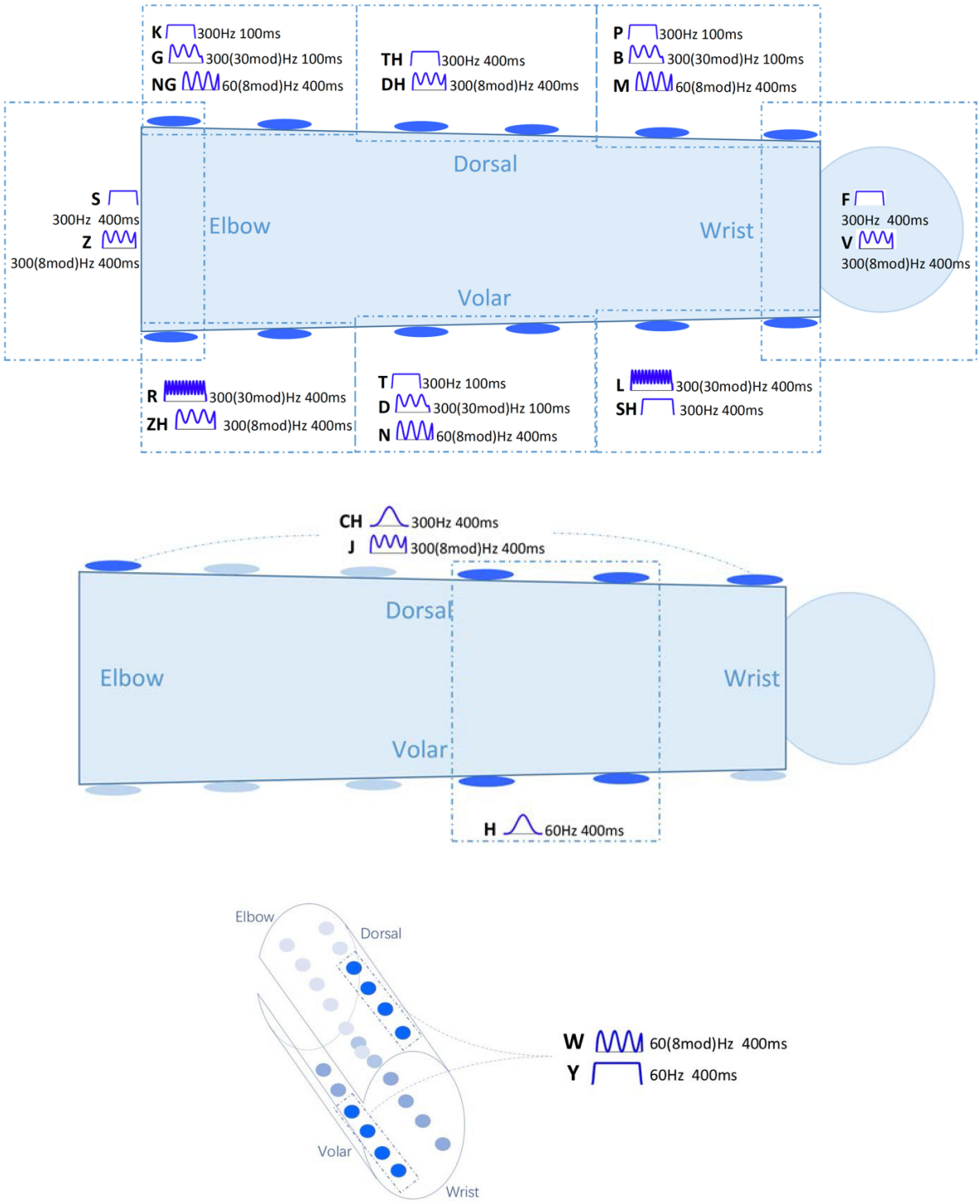


Fig. 2. Schematic description of patterns generated on the tactile display for each of the 24 consonant phonemes. The description for each phoneme includes properties of the stimulus waveform, duration, and location on the array in the dorsal-volar and wrist-to-elbow dimensions. To eliminate crowding on the visual displays and for ease in interpreting the codes, the phoneme descriptions are divided among three different layouts.

phonemes employed different locations along the longitudinal and transverse directions of the array. Among all classes of consonant sounds, the transverse direction of the array (described in Table I and Fig. 2 as dorsal or volar) was used primarily to provide physical separation and distinctions among the codes, rather than being used to represent any particular attribute of these sounds. A video visualizing the tactile stimuli for constants can be found at https://youtu.be/Fr0-XucKGEY.

B. Vowel Codes

A description of the tactile codes generated for the 15 vowels (defined by IPA symbols and the phoneme codes adopted

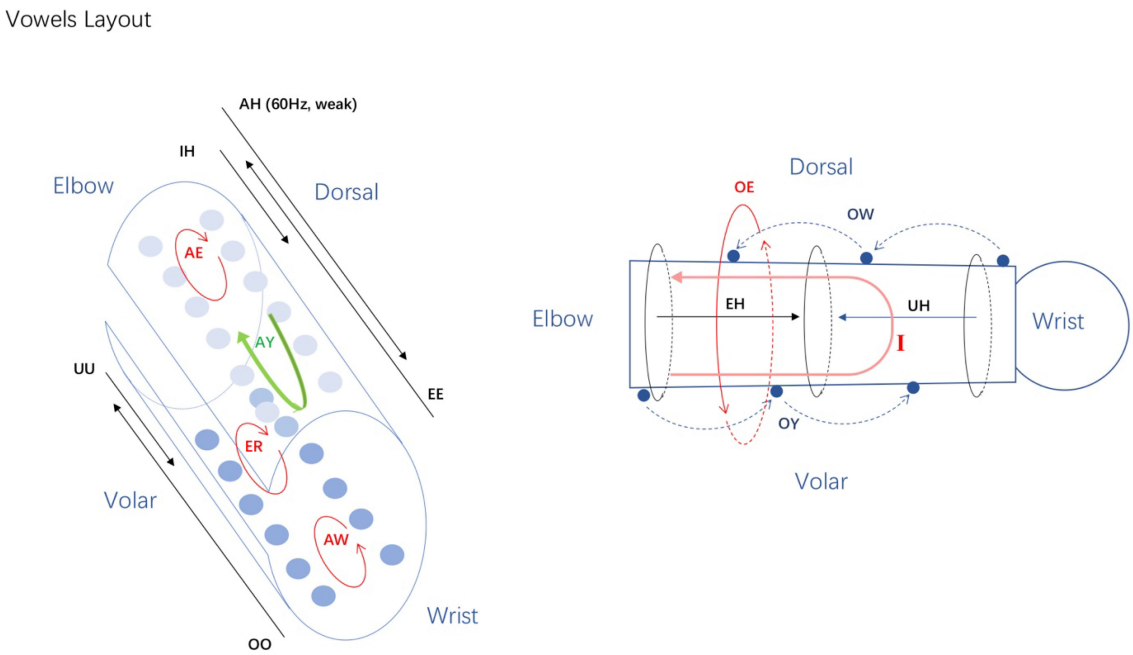


Fig. 3. Schematic description of the movement patterns generated on the tactile display for each of the 15 vowels and diphthongs. For each phoneme, the description includes the location of the signal on the array in the dorsal-volar and wrist-to-elbow dimensions as well as properties related to the direction and extent of movement on the array.

for the current study) is provided in Table II. A key aspect of the vowel codes was to employ different patterns of movement across the tactors for each vowel, in order to exploit the high information-bearing capacity of movement cues for the sense of touch [15], [35]. Although there was some use of articulatory and acoustic properties of vowels in coding decisions (e.g., the use of duration to distinguish tense from lax vowels and the presentation of high-front vowels on the top dorsal row of the array), the vowel codes relied less on these features and more heavily on principles relevant to generating a set of perceptually distinct stimuli. These included creating stimuli that invoked different directions, extent, and trajectory of movement, as well as invoking smooth apparent motion versus discrete saltatory motion.

The 15 vowels were classified into three major groups of six "long" vowels (EE, AH, OO, AE, AW, ER), four "short" vowels (UH, IH, EH, UU), and five diphthongs (AY, I, OW, OE, OY). In Table II, the codes for the vowels are described in terms of their waveforms, location on the tactile array, duration, and movement patterns. In addition, the final column provides a description of the overall impression created by the vibratory pattern. A schematic depiction of the movement patterns associated with each of the 15 vowels is provided in Fig. 3.

Among the 15 vowels and diphthongs, 300 Hz unmodulated waveforms were employed for 9 of the codes. The remaining waveforms included a 60 Hz vibration for AH; the use of sinusoidal modulation of a 300 Hz tone between full and half amplitude at a rate of 30 Hz for OO, UU, AY, and I; and the use of a $\cos^2$ window on the 300 Hz vibration for OE. The duration of the 6 long vowels and 5 dipthongs was 480 ms, while that of the 4 short vowels was 240 ms.

Generally, sensations of movement across the array were created through the use of pulsatile stimuli delivered in a defined temporal order to a specified set of tactors. To create smooth apparent motions, the selection of pulse durations and temporal overlap between successive tactors were guided by the studies of Israr and Poupyrev [36], [37]. The codes for the long vowel EE and the short vowel IH generated this type of smooth apparent motion on the top dorsal row. These two codes differed in the longer duration, larger extent of movement, and direction of movement for EE (wrist to elbow) compared to IH (elbow to middle of forearm). Signals were also constructed to convey saltatory motion, using parameters described in studies of the "cutaneous rabbit" [38], [39]. Saltation was invoked in the codes for the diphthongs OW and OY through the use of 3 tapping pulses at each of three successively stimulated tactors. For OW, the saltation was created on the top dorsal row in the direction from wrist to elbow, and for OY, the movement was on a volar row in the direction of elbow to wrist.

Other types of patterns were created to invoke sensations of circular motion. Pulses moving across 6 tactors twice in succession were used to create codes for AE (at a location on the dorsal rows near the elbow), AW (on the volar rows near the wrist), and ER (on the volar rows near the elbow). The codes for these three vowels led to a "twinkling" type of sensation. For the vowel OE, a smooth circular ring was created to imitate the shape of an "O" through the stimulation of successive pulses on 5 tactors simultaneously across two rows (with the use of one tactor on each row at both the beginning and end of the sequence).

The codes that were generated with the 60 Hz tone and with amplitude-modulated 300 Hz tones led to a heavy or rumbling type of movement. This type of signal construction was used

for the long vowel OO and the short vowel UU, both of which were located on the two volar rows. These two vowels contrasted in their durations, the larger extent of motion for OO compared to UU, and in the direction of longitudinal motion (elbow to wrist for OO and wrist to middle of forearm for UU). For the diphthong AY, tactors were activated on the dorsal rows near the wrist to form a tent shape with a rumbling type of movement. For the diphthong I, a sweeping back-and-forth motion was created with a rumbling sensation on tactors between the elbow and middle of the forearm. Two short vowels were generated to create the sensation of a grabbing type of movement. For UH, pulses moved from the wrist to the middle of the array along two columns, while for EH the movement went from the elbow to the middle of forearm. Finally, the code for the vowel AH involved successive stimulation of 6 tactors along each of the two dorsal rows in the direction of elbow to wrist, leading to the sensation of a wide movement pattern.

Further details regarding the locations on the tactile array of the patterns generated for each of the vowels are provided in Table II and Fig. 3. Generally, decisions on placement of the vowels were made to make full use of both the longitudinal and transversal dimensions of the array. A video visualizing the tactile stimuli for vowels can be found at https://youtu.be/CYfqcdnvMyE. Detailed timing diagrams for the vowels are provided in the Supplemental Materials, which can be found on the IEEE Xplore Digital Library at https://ieeexplore.ieee.org/document/8423203/.

IV. Experiment on Identification of Tactile Phonemic Codes

An absolute identification study was conducted to evaluate the effectiveness of the tactile codes for use in a speech communication device. Participants were trained and tested on their ability to identify the 39 tactile symbols.

A. Participants

The participants were 10 young adults (7 female, 3 male) who were recruited from universities in the Boston area. The participants provided informed-consent through a protocol approved by the IRB at MIT and were paid for their participation in the study. The participants (P1 through P10) ranged in age from 19 to 32 years with a mean of 22.1 and S. D. of 3.9 years. Eight participants reported right-hand and two reported left-hand dominance. The device was always applied to the left arm. None of the participants reported having any history of problems with their sense of touch. Clinical audiograms indicated hearing within normal limits (defined as 15 dB hearing level or better at octave frequencies within the range of 500 to 4000 Hz) for 9 of the participants and a severe hearing loss (mean hearing levels of 75 dB) for one participant. Seven of the participants described themselves as native English speakers, one as bilingual in English and Spanish, one as a native speaker of Romanian with English acquired at age 10 years, and one as a native speaker of Korean with English acquired at age 5 years.

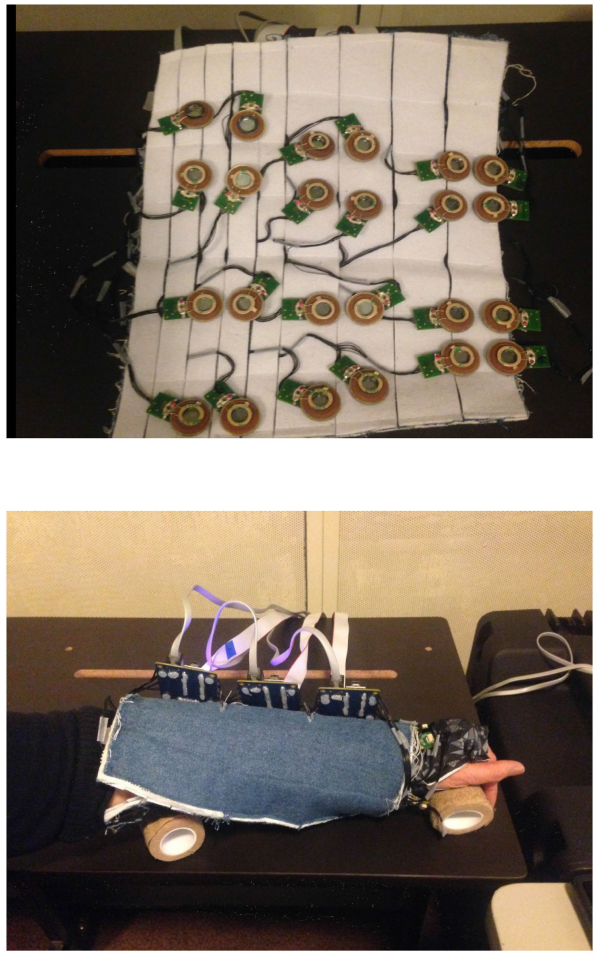


Fig. 4. Top photograph (a) shows the layout of the 4×6 array of tactors with Velcro attachment to a denim gauntlet. Bottom photograph (b) demonstrates the device as placed on an experimenter's forearm, with the fabric gauntlet wrapped snugly around the arm.

B. Assembly of Wearable Array

The tactile device described in Section 2 above was used to deliver the phonemic codes to the participants. To enable application of the device to the forearm of the participants, the 4x6 array of tactors (depicted schematically in Fig. 1) was attached via Velcro to a 13.5 $\times$ 10.5 inch piece of denim fabric (see photograph in Fig. 4a). The placement of the tactors on the fabric gauntlet was determined for each individual participant in the following way. A Spandex sleeve was first placed on the participant's left forearm for hygienic purposes. The forearm was then placed on the gauntlet with the volar side facing down and the forearm lying in the transverse plane with the elbow-to-wrist direction the same as the back-to-abdomen direction. Then the six tactors in each row were adjusted so that two were placed near the wrist, two in the middle of the forearm, and two near the elbow. Two rows were evenly spaced on the dorsal surface and two on the volar surface. Stimulus patterns were presented to the array to ensure that the tactors in each row were perceived as lying on a straight line on the forearm. The final spatial layout, which varied across participants due to differences in the length and

circumference of the forearm, was traced onto a sheet of paper, and was used for subsequent fittings of the device for that particular participant. The fabric gauntlet containing the tactors was then wrapped snugly around the forearm to ensure good contact between the tactors and the skin (see photograph in Fig. 4b). The forearm rested on two supports at the elbow and wrist so that the tactors on the volar side did not touch the table. The center-to-center spacing of the tactors as fit on a typical female forearm was approximately 35 mm in the longitudinal direction and 50 mm in the transverse direction. For a typical male forearm, these distances increased to roughly 40 mm and 57 mm, respectively. Thus, the distance between adjacent tactors in both dimensions of the array exceeded the estimate of 30 mm as the two-point limen on the forearm [33], [34].

C. General Test Protocols

Testing was conducted in a sound-treated booth that contained the components of the tactile device. A monitor, keyboard, and mouse were connected to a desktop computer located outside the booth. This computer ran the experiments with custom-made Matlab programs. At the start of each test session, the tactile device was fit on the participant's forearm, as described in Section 4.2 above, followed by the delivery of sample stimuli to ensure that the tactors were making proper contact with the skin. For all measurements, the participant wore a pair of acoustic-noise cancelling headphones (Bose QuietComfort 25) over which a pink masking noise was presented at a level of roughly 63 dB SPL. This was done to mask any auditory signals arising from the tactile device.

Participants were tested in 2-hour sessions with breaks as needed between experimental tasks. The number of sessions required for completing the phoneme training and testing tasks varied across participants. Two participants required 4 sessions (P1 and P6), two required 3 sessions (P9 and P10), and the remaining six participants completed the tasks within 2 sessions.

D. Level Settings

Two steps were taken to control the perceived intensity of vibrotactile signals at different frequencies and different locations on the forearm: threshold measurements were obtained at one tactor followed by perceived intensity adjustments at the other tactors. These settings were established during the first test session for each participant and then used in subsequent sessions.

1) Threshold Measurements: Individual detection thresholds were measured at 60 and 300 Hz for one tactor on the dorsal side of the forearm near the center of the array (i.e., the tactor in row ii, column 4; see Fig. 1). Thresholds were measured using a three-interval, two-alternative, one-up two-down adaptive forced-choice procedure with trial-by-trial correct-answer feedback. The level of the vibration was adjusted adaptively using the one-up, two-down rule to estimate the stimulus level required for 70.7 percent correct detection [40]. A step size of 5 dB was employed for the first four reversals, and changed to 2 dB for the next 12 reversals. A 400 ms signal (including a 10 ms Hanning window on/off ramp for smoothing onsets and offsets, avoiding energy spread in the frequency domain, and ensuring that the signal begins and ends at 0) was presented with equal *a priori* probability in one of the three intervals, and no signal was presented during the remaining two intervals. The participant's task was to identify the interval containing the signal. Each interval was cued visually on a computer monitor during its 400 ms presentation period with a 500 ms interstimulus interval. Signal levels were specified in dB relative to the maximum output of the system. The threshold measurements began with a signal level set at –20 dB re maximum output. The threshold level was estimated as the mean across the final 12 reversals.

Across the 10 participants, thresholds at 300 Hz ranged from –57.0 to –35.0 dB re maximum output with a mean of –45.0 dB and standard deviation of 6.9 dB. At 60 Hz, thresholds ranged from –41.8 to –26.2 dB re maximum output, with a mean of –32.8 dB and a standard deviation of 6.3 dB. The greater sensitivity at 300 Hz compared to 60 Hz is consistent with previous measurements in the literature [41] and possibly included the differences in tactor responses at these two frequencies.

2) Intensity Adjustments: For each participant, the perceived intensity of the 24 tactors was equalized using a method of adjustment procedure. The reference signal was a 300 Hz sinewave delivered at a level of –10 dB re maximum output to the tactor that was used in the detection threshold measurements. The level of each of the 23 remaining tactors was then adjusted so that its strength matched that of the reference tactor. For each adjustment, the reference tactor and the selected test tactor were delivered in a repeating sequence of three signals consisting of Reference-Test-Reference. Following each sequence, the participant was asked to judge whether the strength of the test signal was lower or higher than the reference signal, and its level was then adjusted accordingly in 2-dB steps. The sequence was repeated until the participant was satisfied that the reference and test signals were at equal perceived strength. The signals in the sequence were presented at a duration of 400 ms with a 300 ms inter-stimulus interval. The experimenter controlled the selection of the test tactor and the level adjustments based on the judgments of the participant. This procedure yielded a level-adjustment table consisting of a level in dB relative to maximum output for each tactor, chosen to produce equal perceived strength across all tactors.

On average across participants, the reference signal was presented at 35 dB sensation level (SL) relative to the 300 Hz threshold. The map derived for the 300 Hz signal was applied to other signal levels using the relative differences between levels of the test tactors and the reference tactor. This map was also used for a 60 Hz signal, taking into account the threshold measurement at 60 Hz and using the same relative differences between levels of the test tactors and the reference tactor as was obtained at 300 Hz. The use of the same intensity adjustments at both signal frequencies is based on the shape of the subjective magnitude contours reported by Verrillo,

TABLE III
EXAMPLE OF EQUALIZATION RESULTS ACROSS THE SET OF 24 TACTORS FOR P9

Tactor Row\Column	Column 1	Column 2	Column 3	Column 4	Column 5	Column 6
Row i	–9	–11	–9	–9	–13	–9
Row ii	–9	–10	–10	–10	–13	–9
Row iii	–13	–13	–16	–13	–13	–10
Row iv	–10	–13	–13	–12	–13	–13

The reference was a 300-Hz signal at a level of –10 dB relative to maximum output, presented at the tactor in row ii, column 4 of the array. The levels of the other tactors (shown in dB relative to maximum output) represent the matches made by the participant for equal strength with the reference tactor. See Fig. 1 for tactor row and column layout.

Fraioli, and Smith [42]. The growth of perceived magnitude for frequencies in the range of 20 to 400 Hz is roughly linear as sensation level is increased from threshold to 55 dB SL.

Representative equalization results across the set of tactors are shown in dB re maximum output for P9 in Table III. This participant's adjustments indicate that signals on the volar surface required less amplitude than the reference signal and the tactors on the dorsal surface (suggesting greater sensitivity on the volar surface for this participant). These adjustments were used to control the intensity of the tactile signals used in the phoneme identification study described below. The level of the tactile stimuli was set for each participant at 25 dB SL relative to the threshold measured on the reference tactor (row ii, column 4) at 300 Hz.

E. Tactile Phoneme Identification Study

The participants were provided training and testing on the identification of the 39 tactile symbols created for the consonants and vowels, as described in Tables I and II and Figs. 2 and 3.

1) Phoneme Sets: The tactile phonemes were introduced to the participants in the order described in Table IV, with consonants preceding vowels. The consonants and vowels were introduced gradually, with each new set building on the previous set. The training sets generally consisted of stimuli from within a given class of phonemes that were constructed along similar principles as described in Sections 3.1 and 3.2 above. This approach was employed to help participants learn to distinguish minimal contrasts that were used to generate the phonemic stimuli. For example, one set of consonants contained the six plosives, all of which had the same duration and frequency of vibration, but differed according to location on the array and amplitude modulation (see Table I).

For consonants, the initial set C1 consisted of the 6 plosives; set C2 consisted of C1 plus six fricatives; set C3 consisted of C2 plus six additional stimuli (2 affricates, 2 fricatives, and two semivowels); and set C4 consisted of C3 plus the remaining 3 semivowels and 3 nasals. After training was completed on the 24 consonants in set C4, three sets of vowels were introduced. Set V1 contained 6 long vowels, set V2 consisted of V1 plus 4 short vowels, and set V3 consisted of V2 plus 5 diphthongs. After training was completed on the full set of 15 vowels, sets V3 and C4 were combined to form the full set of 39 stimuli (CV39).

2) Training Procedure: For each of the sets defined in Table IV, participants engaged in two types of training activities. The first was an unstructured mode of training, referred to as free-play, in which presentation of the stimuli was under the participant's control. This was followed by the use of an identification paradigm which employed trial-by-trial correct-answer feedback along with the option for the participant to replay stimuli arising from error trials. Both procedures were implemented in Matlab.

In the unstructured mode of free-play training, participants were seated in front of a monitor that contained icons labeled with the orthographic representations of the phonemes within a given set. They were able to control the presentation of the tactile signal associated with any given member of the set by selecting an icon and using a computer mouse to click on "Play." Based on previous research suggesting that visual displays may benefit observers in the learning of a tactile task [43], participants were also given the option of clicking on "Show" to activate a visual representation of the duration, frequency, modulation, and motion of the tactile signal associated with the selected phoneme. This representation was displayed on a 4x6 visual plot corresponding to the tactor array. During free-play, participants were free to select icons for tactile or visual presentation and could use as much time as they wished on this activity. A log was kept of the participant's activity, including a record of the stimuli selected for

TABLE IV
STIMULUS SETS EMPLOYED IN TRAINING AND TESTING FOR PHONEME IDENTIFICATION

Set	Number of Items	Stimuli in Set
C1	6	P B T D K G
C2	12	C1 plus F V TH DH S Z
C3	18	C2 plus CH J SH ZH H W
C4	24	C3 plus M N NG R L Y
V1	6	EE AH OO AE AW ER
V2	10	V1 plus UH IH EH UU
V3	15	V2 plus AY I OW OE OY
CV39	39	C4 plus V3

Consonant sets are labeled as "C" followed by a number; vowel sets are labeled as "V" followed by a number.

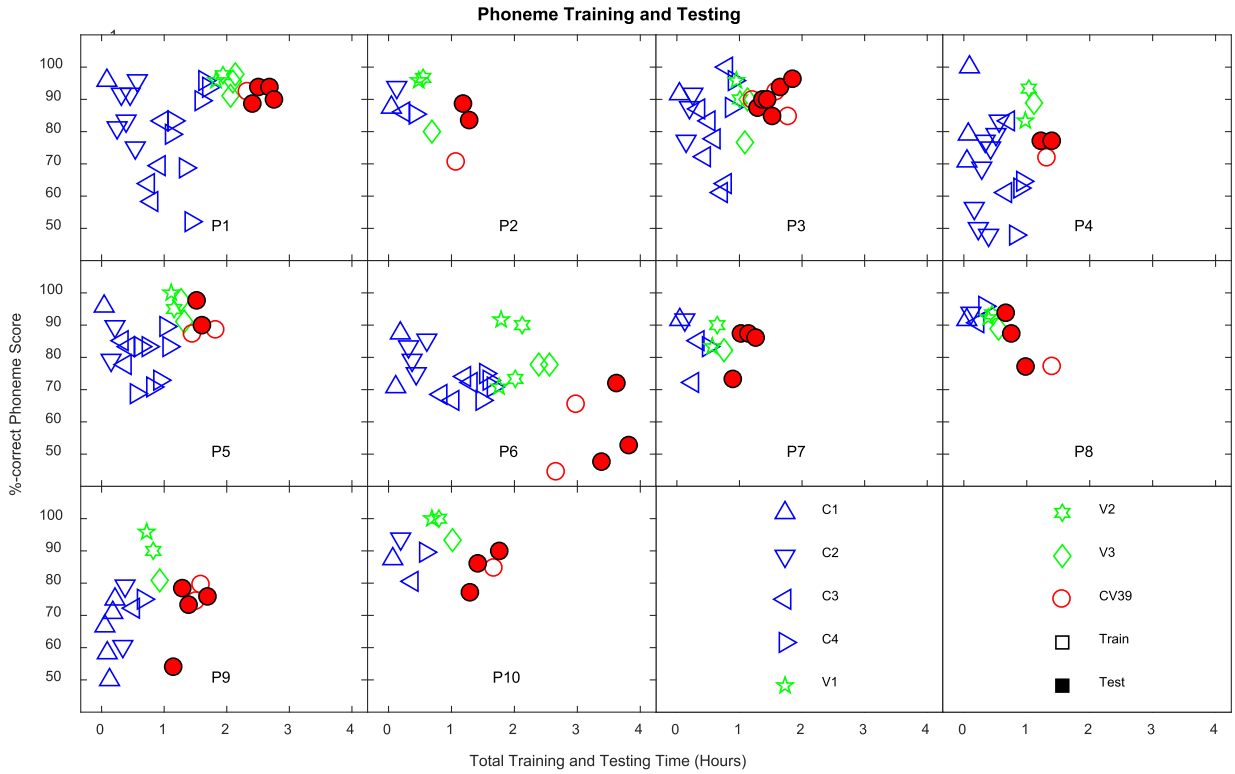


Fig. 5. Results of phoneme training and testing. Each panel shows results for one of the 10 participants. Percent-correct phoneme scores are plotted as a function of total training and testing time (in hours). Different symbols represent different phoneme sets. Unfilled data points show scores obtained on training runs with correct-answer feedback, and filled data points show test scores obtained without feedback. The phoneme sets are defined in Table 4.

presentation, the modality of presentation, and a time stamp for each selection.

After the participant finished using the free-play mode on any given set, training was continued through the use of a one-interval, forced-choice identification paradigm with trial-by-trial correct-answer feedback. On each trial, one of the stimuli was selected randomly for presentation and the participant's task was to select one of the alternatives from the set, which were displayed on the screen. On error trials, the stimulus and the incorrect response were illuminated in different colors on the screen. The participant was given the option of comparing the signals associated with the incorrect response and the stimulus, using either "Play" for tactile presentation or "Show" for visual display of these signals. Participants were given unlimited time for replaying the stimuli. For sets C1, V1, and CV39 (C4+V3), stimuli were selected randomly with replacement. For sets C2, C3, C4, V2, and V3, half the trials were devoted to new phonemes that had been added to the set and half to phonemes that had been introduced in a previous set. The number of trials presented in the training runs increased with the number of stimuli in the set. Training runs were conducted until a criterion level of performance was obtained (in the range of 80-90 percent correct) before proceeding to training on the next phoneme set in the sequence shown in Table IV.

3) Testing Procedure: Testing began immediately after training was completed. Testing was conducted using the identification paradigm described above, except that the use of any type of feedback was eliminated. These tests were conducted on all participants for CV39, where at least two 78-trial runs were collected. On each run, each of the phonemes was presented twice in a randomly selected order. Stimulus-response confusion matrices were used to calculate percent-correct scores and were analyzed to examine patterns of confusion among the tactile symbols. An analysis was also conducted of the response times that were recorded on each trial of a test run, measured as the duration between the offset of the stimulus and the onset of the participant's response.

4) Training and Test Results: A summary of performance on tactile phoneme identification is shown in Fig. 5, where each panel contains results for one of the ten participants. Percent-correct phoneme recognition scores on the various stimulus sets are plotted as a function of the cumulative duration of training (open symbols) and testing (filled symbols). [The time spent on training within the free-play mode was added into the cumulative duration at the time periods when it occurred.] Note that once criterion performance was achieved with a given stimulus set, the next set was introduced, generally resulting in a decrease in performance until criterion was achieved again. Thus, these are not traditional learning curves with monotonically increasing levels of performance. The total duration of time required to meet the criteria for training, as well as final scores on the full set of CV39 stimuli, varied across participants. The length of time required to complete the training ranged from roughly 50 minutes (P8) to 230 minutes (P6). Of the 10 participants, seven achieved a test

score on the full set of CV39 stimuli within the range of 85 to 97 percent correct, while three of the participants (P4, P6, and P9) were less successful in mastering the task with maximal test scores in the range of 71 to 76 percent correct.

Trends were also examined in the participants' use of the unstructured free-play mode as part of the training protocol. Within the free-play program itself, participants' activities were analyzed in terms of the time spent accessing each phoneme. Across phonemes and participants, a mean duration of 48.3 s per phoneme (standard deviation of 13.2 s) was calculated. Across participants, the use of the Play option for presentation of stimuli through the tactile device far exceeded that of the Show option for a corresponding visual display. On average, across stimulus sets and participants, the Play option (mean across phoneme groups and participants of 243.0 s) was used roughly 16.5 times more often than the Show option (mean of 14.7 s).

A 39-by-39 stimulus-response confusion matrix (with rows corresponding to each phoneme presented) was constructed from the two maximally scoring test runs on the CV39 set from each of the 10 participants (a total of 1,560 trials arising from 10 participants $\times$ 4 trials per phoneme $\times$ 39 phonemes). Each entry $n_{ij}(\mathrm{i,j} = 1, \ldots, 39)$ represents the number of times that phoneme i is called phoneme j. The diagonal entries $n_{ii}(\mathrm{i} = 1, \ldots, 39)$ are the correct responses. The off-diagonal entries $n_{ij}(\mathrm{i} \neq \mathrm{j})$ are the error trials. The row sum $n_i = \sum_{j=1}^{39} n_{ij}$ represents the total number of times that the i-th phoneme was presented. The error percentages are calculated as $e_{ij} = 100 \times \sum_{j=1}^{39} n_{ij}/n_i$ for entries where $i \neq j$. For the confusion matrix constructed here (provided in the Supplemental Materials, available online), the overall correct-response rate was 85.77 percent (thus an overall error rate of 14.23 percent).

A visualization procedure was used to depict patterns of confusion among the phonemes in the matrix. In this procedure, the user specifies the range of errors to be visualized. Any e_{ij} that falls into the range is then shown as the phoneme pair i-j with a line connecting two circles labeled with the phoneme stimulus and phoneme response, respectively. It follows that there might be multiple lines connecting one phoneme to multiple other phonemes. The locations of the circles representing the phonemes and their relative distances can be adjusted by the user to change the layout of visualization, and they do not carry any additional information. The layout shown in Fig. 6 uses a minimum error percentage of 7.5 percent (i.e., at least 3 errors in any off-diagonal cell) and a maximum error of 25 percent which corresponded to the maximal error rate of any off-diagonal entry observed in the data. These results indicate that error patterns existed within the 15 vowels and within the 24 consonants, but not across these two major categories.

The confusions observed within the vowels are depicted on the top of Fig. 6. Seven of the 15 vowel stimuli (ER AH AW IH I OE OO) were highly identifiable with no off-diagonal errors greater than 7.5 percent. Two pairs of vowels formed clusters with confusion rates of roughly 10 percent for UH–EH and 12 percent for AY-AE. The remaining notable confusions were grouped into a four-stimulus cluster containing EE, OW, OY, and UU. The stimulus OW was confused with EE

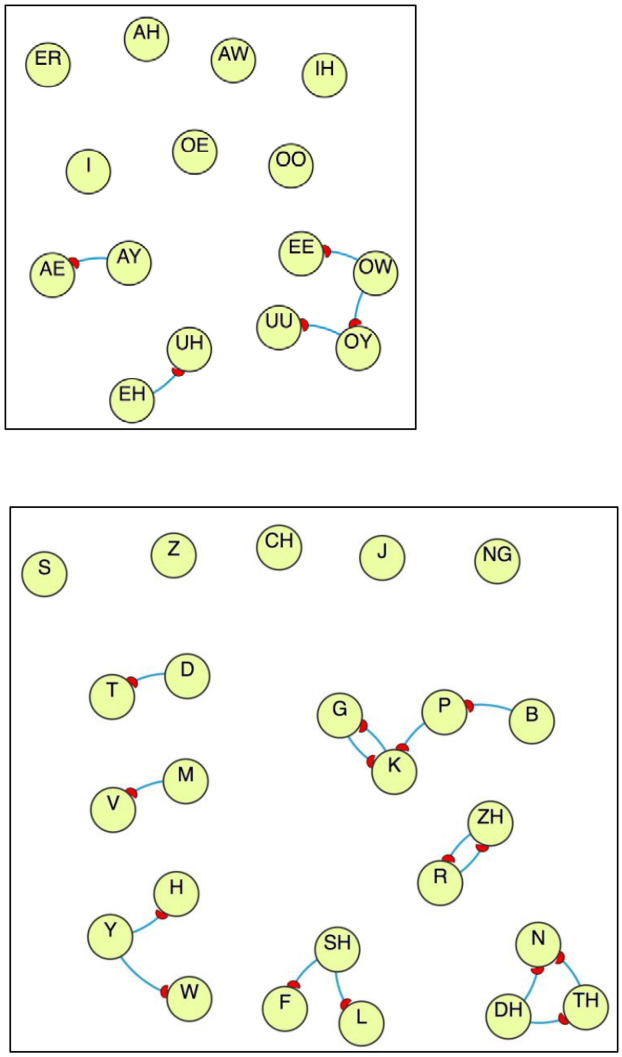


Fig. 6. Results of visualization procedure used to depict confusion patterns observed in the stimulus-response matrix for 39 phonemes, with a minimum error percentage of 7.5 percent. Vowel confusions are shown on the top, and consonants on the bottom. Each phoneme is represented by one of the circles. Pairs of phonemes with confusion rates greater than 0.075 are connected by lines; the semi-circle ending indicates the direction in which the error was made. For example, EH was misidentified as UH whereas errors in the opposite direction were not observed. For the cases of G-K and R-ZH, it can be seen that errors occurred in both directions.

(error rate of 10 percent) and OY (7.5 percent), and OY in turn was confused with UU (10 percent). Nearly all of these confused pairs consist of tactile symbols with the same duration and direction of movement but differing in other properties including location along the longitudinal dimension (e.g., AY-AE) and the particular type of movement that was employed (e.g., EE-OW and OY-UU both contrast the saltatory sensation with a different type of movement). For the confused pairs UH-EH and OW-OY, on the other hand, the members of each pair evoke the same type of movement but differ in its direction and location on the array.

Confusions observed among the consonant stimuli are shown on the bottom of Fig. 6. Among the consonants, five stimuli were identified with no off-diagonal error rates greater than 7.5 percent (S Z CH J NG). Confusion patterns on the

remaining 19 stimuli were arranged into seven clusters. Two of these clusters described confusions solely among plosive stimuli. These were confusions of D-T (error rate of 16 percent) and a 4-item cluster containing P, B, K, and G. These confusions included B-P (at an error rate of 17.5 percent), P-K (7.5 percent), and K-G (15 percent). Among this set of confusions, all stimuli were the same duration (100 ms) and included the three pairs of voicing contrasts (P-B, T-D, and K-G). Within each of these pairs, the two confused stimuli occupy the same location on the array and differ only in the contrast of an unmodulated with a modulated 300 Hz sinewave. The P-K confusion represents an error of location (wrist versus elbow). Two additional two-item clusters were observed, both at an error rate of 7.5 percent: V-M (which differ in location as well as frequency of vibration) and ZH-R (which differ in modulation frequency). The three final clusters each contained three items. These were confusions at a rate of 7.5 percent between TH-DH (unmodulated versus modulated tone), DH-N, and TH-N (dorsal/volar and frequency confusions). Another group highlighted errors of SH with F at a confusion rate of 12.5 percent (dorsal/volar confusion) and SH with L at a rate of 7.5 percent (modulated versus unmodulated tone). The final group contained confusions of Y-W (at a rate of 25 percent) and Y-H (12.5 percent). All three phonemes used a 60 Hz sinewave, but had differences in modulation and/or location (see Fig. 2).

Response times were also examined as an indication of the processing demands placed on the participants. In Fig. 7, mean response times across participants are plotted as a function of the number of stimuli in the set (on a base 2 logarithmic scale). For consonants, mean response times increased from 2.2 to 3.8 s as the number of items in the set increased from 6 to 24. For vowels, the response time increased from 2.0 to 3.3 s as the set size increased from 6 to 15. When all 39 phonemes were included in the set, mean response time increased to 4.2 s. The slope of the function, for number of items in the set regardless of phoneme group, is roughly 0.07 s per doubling of the items in the set.

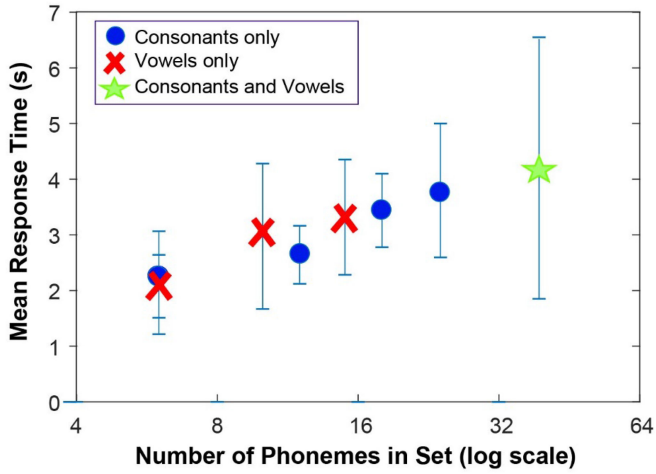


Fig. 7. Plot of mean response times as a function of number of items in the stimulus sets (base 2 logarithmic scale), as defined in Table 4. Response times, measured as the duration between signal offset and initiation of participant's response on each trial of the identification paradigm, are averaged across the 10 participants. Error bars represent $\pm$ 1 standard deviation around the mean.

V. Discussion

After modest amounts of training (from one to four hours), the tactile codes generated to convey the 39 English phonemes through a 4 $\times$ 6 array of tactors could be identified by naïve, young adult participants at high rates of accuracy. The data plotted in Fig. 5 indicate that generally longer amounts of training were required for the consonants compared to the vowels, with likely reasons for this being that consonants were trained on first and contained a larger number of items in the full set. However, other factors may also be related to the greater ease with which vowels were acquired. Comparing the performance on the first consonant set (C1, containing 6 items) with the first vowel set (V1, also containing 6 items), it can be seen that scores on C1 generally began at much lower levels than on V1. Although the participants' previous experience with the consonant codes may be related to their greater facility with vowel acquisition, it is possible that characteristics of the construction of the codes for consonants and vowels may also play a role. The different movement patterns employed in the vowel codes may have been more easily learned than the static contrasts of spatial location and waveform employed in the plosive sounds of the C1 set. With sufficient training, however, it appears that both vowel and consonant stimuli could be acquired by the participants.

The error patterns observed in the stimulus-response confusion analysis indicate that the vowels were perceived separately from the consonants, as no appreciable confusions were observed between these two classes of codes. Among the vowel stimuli, the direction of movement was rarely confused. Instead, vowel errors arose from confusions with the location on the array at which stimuli were presented, as well as with confusions between different types of evoked movement (e.g., apparent versus saltatory movement). For consonants, errors were primarily concentrated on confusions of voiced and unvoiced pairs of stimuli which were coded by the use of a modulated versus unmodulated sinewave, as well as on errors related to location on the array in both the dorsal-volar and elbow-to-wrist dimensions.

In the approach to training taken here, participants were given the opportunity to practice with stimuli presentations under their own control, including the option for presentation in an alternative modality (vision) as well as through the tactile device. In terms of their usage of these options, the participants were much more likely to initiate tactile rather than visual stimulus presentations as they were learning the stimuli. This observation suggests that the visual display employed here did not provide participants with useful information to promote their learning of the signals. It is not clear whether this is due to characteristics of the particular visual display employed here, or whether alternative modality training, in general, does not transfer readily to the tactile task. Exploration of alternative visual displays as well as training options in another modality such as hearing is warranted.

Another approach to phonemic coding of tactile symbols has been reported recently by Zhao *et al.* [44] using a 2×3 array of tactors applied to the dorsal forearm. Five consonants were coded using single-point activation on the array at a duration of 180 ms, and four vowels were coded using sequential two-point activation at a duration of 770 ms. Phoneme labels were assigned to these tactile codes either with a random association or using a place-of-articulation mapping similar to that employed in the current study. Following training with feedback on identification of the 9 phonemes in an AXB paradigm (where X was the phoneme to be identified), participants performed similarly (roughly 80 percent correct), regardless of the mapping strategy. The articulatory mapping approach, however, proved to be advantageous for recognition of words constructed from sequences of tactile phonemes.

The levels of performance obtained on the phonemic-based tactile codes described in the current study are promising for use in further research concerned with the identification of tactile words and phrases. The phoneme-recognition rate achieved here of 86 percent correct compares well to that reported previously for laboratory-trained users of acoustic-based tactile displays of speech [45], [46]. Weisenberger and Percy [45] studied phonemic reception through a seven-channel tactile aid that was applied to the volar forearm and provided a spectral display of the acoustic speech signal. The ability of laboratory-trained normal-hearing users to identify items in six different sets of 8 consonants or vowels produced by a live talker ranged from 16 to 32 percent correct across sets. Performance on a set of 24 consonants dropped to 12 percent correct. Weisenberger *et al.* [46] conducted studies of phonemic identification using a 6×5 array of tactors attached to the forearm (similar in size and applied to the same body site as that used here). The tactile device provided information about properties of speech that were derived from the acoustic waveform. In a group of normal-hearing participants who were highly experienced in laboratory use of tactile speech displays (and had roughly 5 hrs of experience with this particular device), performance on sets of 9 vowels or 10 consonants was roughly 40 percent correct, and fell to 23 percent correct for a set of 19 consonants. It is important to point out a major difference between these studies and the one reported here. In the current work, each phoneme is represented by one tactile code. When the raw acoustic speech signal of live talkers is used to extract information for the tactile display, however, the users of the display must cope with the token-to-token variability that arises in the representation of each phoneme, thus increasing the difficulty of the task.

Evidence that the phonemic recognition rate achieved here is sufficient to support the recognition of tactile words and phrases is provided by results obtained with experienced users of the Tadoma method of speechreading [3]. Even though the segmental reception ability of Tadoma users for consonants and vowels in nonsense syllables is roughly 55 percent correct, they are nonetheless able to understand conversational sentences spoken at slow-to-normal rates with 80 percent correct reception of key words. These Tadoma results indicate that partial information at the phonemic level can be combined with knowledge of supra-segmental properties of speech as well as semantic and linguistic cues to support the recognition of spoken language. Thus, the tactile phonemic results obtained here offer support for the successful use of these tactile phonemic codes in the reception of tactile words and phrases. In fact, preliminary studies indicate that participants exhibit memory capacity sufficient for using the tactile phonemic codes to interpret words [47], [48].

Although the results reported here support the feasibility of a phonemic-based tactile aid, there are still a number of challenges that must be addressed in the realization of a practical device for speech communication. In addition to the need for accurate real-time ASR at the front end of the system, there is also the need to cope with the complex listening environments associated with real-world situations. This includes the need to distinguish among multiple speech sources arising from different directions as well as the ability to separate the target speech from background interference. Whereas the attentional systems of persons with normal hearing allow them to cope with such complex auditory situations, signal-processing algorithms must be developed that will allow the user of a tactile aid to focus on the intended source and filter out unwanted interference. Further research is required to address these issues in the development of a wearable tactile speech-communication system for use in real-world situations.

VI. Conclusions

A set of tactile symbols corresponding to 39 English phonemes was developed for use in a tactile speech communication device. This approach assumes that a string of phonemes corresponding to an utterance can be produced at the front end of the device by an automatic speech recognizer. The tactile codes were developed for presentation through a 4×6 array of independently activated tactors applied to the forearm. Preliminary studies with a group of 10 naïve participants indicated that the tactile codes could be identified at high rates of proficiency within several hours of training. These results support the feasibility of a phonemic-based approach to the development of tactile speech communication devices. Future research will address the reception of words and sentences composed of strings of tactile phonemes.

References

[1] S. J. Lederman and R. L. Klatzky, "Haptic perception: A tutorial," *Attention Perception Psyschophysics*, vol. 71, no. 7, pp. 1439–1459, 2009.

[2] R. T. Enerstvedt, *Legacy of the Past: Those Who Are Gone But Have Not Left.* Dronninglund, Denmark: Forlaget Nord-Press, 1996, pp. 235–330.

[3] 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.

[4] C. M. Reed, L. A. Delhorne, N. I. Durlach, and S. D. Fischer, "A study of the tactual and visual reception of fingerspelling," *J. Speech Hearing Res.*, vol. 33, no. 4, pp. 786–797, 1990.

[5] C. M. Reed, L. A. Delhorne, N. I. Durlach, and S. D. Fischer, "A study of the tactual reception of sign language," *J. Speech Hearing Res.*, vol. 38, no. 2, pp. 477–489, 1995.

[6] J. H. Kirman, "Tactile communication of speech: A review and analysis," *Psychol. Bull.*, vol. 80, no. 1, pp. 54–75, 1973.

[7] C. M. Reed, N. I. Durlach, and L. D. Braida, "Research on tactile communication of speech: A review," *Amer. Speech-Language-Hearing Assoc. Monographs*, no. 20, 1982.
[8] C. M. Reed, N. I. Durlach, L. A. Delhorne, W. M. Rabinowitz, and K. W. Grant, "Research on tactual communication of speech: Ideas, issues, and findings," *Volta Rev.*, vol. 91, no. 5, pp. 65–78, 1989.
[9] H. Yuan, C. M. Reed, and N. I. Durlach, "Tactual display of consonant voicing as a supplement to lipreading," *J. Acoust. Soc. Amer.*, vol. 118, no. 2, pp. 1003–1015, 2005.
[10] A. Israr, C. M. Reed, and H. Z. Tan, "Discrimination of vowels with a multi-finger tactual display," in *Proc. Symp. Haptic Interfaces Virtual Environ. Teleoperator Syst.*, 2008, pp. 17–24.
[11] S. D. Novich and D. M. Eagleman, "Using space and time to encode vibrotactile information: Toward an estimate of the skin's achievable throughput," *Exp. Brain Res.*, vol. 233, no. 10, pp. 2777–2788, 2015.
[12] S. Engelmann and R. Rosov, "Tactual hearing experiment with deaf and hearing subjects," *J. Exceptional Children*, vol. 41, pp. 243–253, 1975.
[13] E. T. Auer Jr. and L. E. Bernstein, "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.
[14] C. M. Reed, "Tadoma: An overview of research," in *Profound Deafness and Speech Communication*, G. Plant and K.-E. Spens, Eds. London, U.K.: Whurr Publishers, 1995, pp. 40–55.
[15] C. M. Reed and N. I. Durlach, "Note on information transfer rates in human communication," *Presence*, vol. 7, no. 5, pp. 509–518, 1998.
[16] M. C. Clements, L. D. Braida, and N. I. Durlach, "Tactile communication of speech: Comparison of two computer-based displays," *J. Rehab. Res. Develop.*, vol. 25, no. 4, pp. 25–44, 1988.
[17] J. M. Weisenberger, S. M. Broadstone, and F. A. Saunders, "Evaluation of two multichannel tactile aids for the hearing impaired," *J. Acoust. Soc. Amer.*, vol. 86, no. 5, pp. 1764–1775, 1989.
[18] L. Hanin, A. Boothroyd, and T. Hnath-Chisolm, "Tactile presentation of voice fundamental frequency as an aid to the speechreading of sentences," *Ear Hearing*, vol. 9, no. 6, pp. 335–341, 1988.
[19] C. M. Reed and L. A. Delhorne, "Current results of a field study of adult users of tactile aids," *Seminars Hearing*, vol. 16, pp. 305–315, 1995.
[20] D. F. Leotta, W. M. Rabinowitz, C. M. Reed, and N. I. Durlach, "Preliminary results of speech-reception tests obtained with the synthetic Tadoma system," *J. Rehab. Res.*, vol. 25, no. 4, pp. 45–52, 1988.
[21] D. R. Henderson, "Tactile speech reception: Development and evaluation of an improved synthetic Tadoma system," M.S. thesis, Dept. Elect. Eng. Comput. Sci., Massachusetts Institute of Technology, Cambridge, MA, USA, 1989.
[22] H. Z. Tan, N. I. Durlach, C. M. Reed, and W. M. Rabinowitz, "Information transmission with a multifinger tactual display," *Perception Psychophysics*, vol. 61, no. 6, pp. 993–1008, 1999.
[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] 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, 2017, Art. no. 317; doi:10.3390/app7040317.
[25] T. McDaniel, S. Krishna, D. Villanueva, and S. Panchanathan, "A haptic belt for vibrotactile communication," *Proc. IEEE Int. Symp. Haptic Audio Visual Environ. Games*, 2010, pp. 1–2.
[26] M. Sreelakshmi and T. D. Subash, "Haptic technology: A comprehensive review on its applications and future prospects," *Mater. Today: Proc.*, vol. 4, pt. B, no. 2, pp. 4182–4187, 2017.
[27] M. Benzeghiba *et al.*, "Automatic speech recognition and speech variability: A review," *Speech Commun.*, vol. 49, no. 10-11, pp. 763–786, 2007.
[28] T. Wade and L. L. Holt, "Incidental categorization of spectrally complex non-invariant auditory stimuli in a computer game task," *J. Acoust. Soc. Amer.*, vol. 188, no. 4, pp. 2618–2633, 2005.
[29] R. T. Verrillo and G. A. Gescheider, "Perception via the sense of touch," in *Tactile Aids for the Hearing Impaired*, I. R. Summers, Ed. London, U.K.: Whurr Publishers, 1992, pp. 1–36.
[30] H. Z. Tan, N. I. Durlach, W. M. Rabinowitz, C. M. Reed, and J. R. Santos, "Reception of Morse code through motional, vibrotactile, and auditory stimulation," *Perception Psychophysics*, vol. 59, no. 7, pp. 1004–1017, 1997.
[31] G. A. Miller, "The magical number seven, plus or minus two: Some limits on our capacity for processing Information," *Psychological Rev.*, vol. 63, pp. 81–97, 1956.
[32] I. Pollack and L. Ficks, "Information of elementary multidimensional auditory displays," *J. Acoust. Soc. Amer.*, vol. 26, no. 2, pp. 155–158, 1954.
[33] S. Weinstein, "Intensive and extensive aspects of tactile sensitivity as a function of body part, sex, and laterality," in *The Skin Senses*, D. R. Kenshalo, Ed. Springfield, IL, USA: Charles C. Thomas, 1968, pp. 195–222.
[34] J. Tong, O. Mao, and D. Goldreich, "Two-point orientation discrimination versus the traditional two-point test for tactile spatial acuity assessment," *Frontiers Human Neuroscience*, vol. 7, 2013, Art. no. 579.
[35] H. Z. Tan, S. Choi, F. W.Y. Lau, and F. Abnousi, "Maximizing information transmission of man-made haptic systems," *Proc. IEEE*, to be published.
[36] A. Israr and I. Poupyrev, "Tactile brush: Drawing on skin with a tactile grid display," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2011, pp. 2019–2028.
[37] A. Israr and I. Poupyrev, "Control space of apparent haptic motion," in *Proc. IEEE World Haptics Conf.*, 2011, pp. 457–462.
[38] F. A. Geldard and C. E. Sherrick, "The cutaneous 'rabbit': A perceptual illusion," *Sci.*, vol. 178, no. 4057, pp. 178–179, 1972.
[39] H. Z. Tan, R. Gray, J. J. Young, and R. Traylor, "A haptic back display for attentional and directional cueing," *Haptics-e*, vol. 3, no. 1, Jun. 11, 2003. [Online]. Available: http://www.haptics-e.org
[40] H. Levitt, "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Amer.*, vol. 49, no. 2B, pp. 467–477, 1971.
[41] S. J. Bolanowski, Jr., G. A. Gescheider, R. T. Verrillo, and C. M. Checkosky, "Four channels mediate the mechanical aspects of touch," *J. Acoust. Soc. Amer.*, vol. 84, no. 5, pp. 1680–1694, 1988.
[42] R. T. Verrillo, A. J. Fraioli, and R. L. Smith, "Sensation magnitude of vibrotactile stimuli," *Perception Psychophysics*, vol. 6, no. 6, pp. 366–372, 1969.
[43] A. D. Hall and S. E. Newman, "Braille learning: Relative importance of seven variables," *Appl. Cogn. Psychol.*, vol. 1, pp. 133–141, 1987.
[44] S. Zhao, A. Israr, F. Lau, and F. Abnousi, "Coding tactile symbols for phonemic communication," in *Proc. Conf. Human Factors Comput. Syst.*, Apr. 21-26, 2018, Art. no. 392.
[45] J. M. Weisenberger and M. E. Percy, "The transmission of phoneme-level information by multichannel speech perception aids," *Ear Hearing*, vol. 16, no. 4, pp. 392–406, 1995.
[46] J. M. Weisenberger, J. C. Craig, and G. D. Abbott, "Evaluation of a principal-components tactile aid for the hearing-impaired," *J. Acoust. Soc. Amer.*, vol. 90, no. 4, pp. 1944–1957, 1991.
[47] Y. Jiao *et al.*, "A comparative study of phoneme- and word-based learning of English words presented to the skin," in *Proc. EuroHaptics*, 2018, pp. 623–635.
[48] J. Jung *et al.*, "Speech communcation through the skin: Design of learning protocols and intial findings," in *Proc. Int. Conf. Des. User Exp. Usability*, Jul. 15-20, 2018, pp. 447–460.

Charlotte M. Reed received the B.S. 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.

Hong Z. Tan received the bachelor's degree in biomedical engineering from Shanghai Jiao Tong University, Shanghai, China, in 1986, and the master's degree in electrical engineering and the doctorate degree in 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 engineerg (by courtesy), and psychological sciences (by courtesy) with Purdue University, West Lafayette, IN, USA. She was an Associate Editor of the IEEE TRANSACTIONS ON HAPTICS from 2007 to 2012 and since 2016. She received a Meritorious Service Award in 2012.

Zach Perez received the bachelor's degree in chemistry from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 2005 and the Juris Doctor degree from the Georgetown University Law Center, Washington, DC, USA, in 2012. He is currently a Senior Research Support Associate with the Research Laboratory of Electronics, Massachusetts Institute of Technology.

E. Courtenay Wilson received the B.S. degree in computer science and engineering from the University of Connecticut, Storrs, CT, USA, in 1994, the M.S. degree in computer engineering from the University of Nevada, Reno, NV, USA, in 2001, and the Ph.D. degree in speech and hearing bioscience and technology from the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, in 2010. She is currently a Researcher with MIT and a Lecturer with the Math Department, Northeastern University, Shenyang, China.

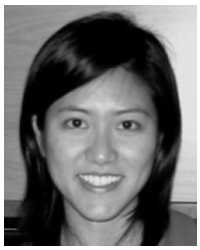
Frederico M. Severgnini received the B.A.Sc. degree in electronic and telecommunication engineering from PUC Minas University, Belo Horizonte, Brazil, in 2015 and the M.S. degree in electrical and computer engineering from Purdue University, West Lafayette, IN, USA, in 2018. He was a Graduate Research Assistant with the Haptic Interface Research Lab advised by Dr. H. Z. Tan.

Jaehong Jung is working toward the B.S. degree in the School of Mechanical Engineering, Purdue University, West Lafayette, IN, USA. He is an Undergraduate Research Assistant with the Haptic Interface Research Lab, Purdue University, under the supervision of Dr. H. Z. Tan.

Juan S. Martinez received the B.A.Sc. degrees in electronic engineering and systems and computer engineering from Los Andes University, Bogota, Colombia, in 2016 and 2017, respectively. He is currently working toward the M.S. degree at Purdue University, West Lafayette, IN, USA, where he is a Research Assistant in the Haptic Interface Research Lab, advised by Dr. H. Z. Tan.

Yang Jiao received the B.E. degree in electronic information science and technology from the Beijing University of Posts and Telecommunications, Beijing, China, in 2009, the M.S. degree in wireless communications from the University of Southampton, Southampton, U.K., in 2010, and the Ph.D. degree in design from Tsinghua University, Beijing, China, in 2017. He is currently a Postdoctoral Research Associate at Purdue University, West Lafayette, IN, USA.

Ali Israr received the B.S. degree in mechanical engineering from the University of Engineering & Technology, Lahore, Pakistan, and the M.S. and Ph.D. degrees in mechanical engineering from Purdue University, West Lafayette, IN, USA.



Frances Lau received the B.A.Sc. degree in electrical engineering from the University of Toronto, Toronto, ON, Canada, in 2005, and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, USA, in 2007 and 2013, respectively. She is currently working on research and development of haptic communication systems at Facebook, Menlo Park, CA, USA.

Keith Klumb received the bachelor's degree in electrical engineering technology from Purdue Unviersity, West Lafayette, IN, USA, in 2001. He is currently a Research Program Manager with Facebook, Menlo Park, CA, USA.

Robert Turcott received the Ph.D. degree in electrical engineering with a focus on signal processing from Columbia University, New York, NY, USA, and the M.D. degree from Stanford University, Stanford, CA, USA. When he is not practicing cardiology, he serves as a consultant in the medical device and technology industries.

Freddy Abnousi received the M.D. degree from the Stanford University School of Medicine, Stanford, CA, USA, the M.B.A. degree from Oxford University, Oxford, U.K., and the M.Sc. degree in health policy, planning, and financing from the London School of Economics, London, U.K. He is an interventional cardiologist specializing in coronary and structural interventions. He currently leads haptic communication efforts at Facebook's Building 8.