Discrimination of Consonant Articulation Location by Tactile Stimulation of the Forearm

Elaine Y. Wong, Ali Israr, Marcia K. O'Malley

Mechatronics and Haptic Interfaces Research Laboratory, Rice University

ABSTRACT

In this paper, we evaluate the ability of four able-bodied participants to discriminate the articulation location for spoken consonants, using tactile cues presented on the dorsal side of their forearm. Additionally, we determine the processing capability of the dorsal forearm's skin with a tactile sleeve worn by ten participants using two psychophysical studies. Our first study shows that 2-3 tactors arranged along the length of the forearm can be reliably identified by human users, when only the location of vibration is varied. Our second study indicates that the physical placement of localized vibrations map linearly to the perceived physical arrangement. Based on these findings, the subsequent speech experiment uses six tactors placed equidistant from each other and maps location of the constriction inside the mouth as directional cues on the forearm. Results of the speech study show that participants are able to indicate which of two randomly presented tactile cues (derived from consonant-vowel-consonant (CVC) non-sense syllables) has the preceding consonant closer to the lips. The discrimination performance is better (i) with fricatives than plosives, (ii) when the consonants are produced further apart inside the mouth, and (iii) when both place and manner of articulation feature is varied. The study also shows that discrimination performance with cues applied to the forearm is inferior to that with the fingerpads utilized in a previous study.

KEYWORDS: Perception and Psychology, sensory substitution, speech communication, tactile perception.

INDEX TERMS: H.1.2 [Models and Principles]: User/Machine Systems—Human Factors; H.5.2 [Information Interfaces and Representation]: User Interfaces—Haptic I/O.

1 INTRODUCTION

It has been a long term goal to use touch as a substitute for speech reception for individuals with hearing impairments. The effort of developing artificial displays for speech communication using the skin started in the early 1900s when a bone vibrator (a long tube placed in front of the mouth and held by a receiver) was used to transmit voicing information to the receivers [1]. Later, more sophisticated schemes were used to extract speech features and present them through a single or several vibrators arranged as a matrix-array or along a line (i.e., as vocoders) [2, 3]. These displays were successful in transmitting coarse spectral features, such as formant distributions in vowels and amplitude variations; however, fine spectral features that were critical for consonant distinctions were not transmitted by these displays.

The ability to use touch for speech communication is demonstrated by a natural (non-device based) method, the Tadoma method, in which deaf-blind individuals process facial variations during speech production through the sensitive fingerpads of their hands [4]. These individuals converse in a daily setting with successful information transmission rates of about 12 bits/sec, roughly the half the rate achieved by hearing individuals [5]. However, most commercially available tactile aids transmit no greater than 5 bits/sec [5]. One reason for the success of the Tadoma method is that users have access to rich tactual cues (vibrations and motions) deduced directly from the speech production mechanism and processed by sensitive fingers of the receiver hand. In contrast, users of tactile aids are exposed to homogenous vibrations presented to relatively less innervated skin such as that of the forearm, abdomen, thigh, or in some cases the fingerpad.

Recently, Yuan et al. [6] used a three channel tactual stimulator, the Tactuator, to present low frequency (<350 Hz) and high frequency (>3000 Hz) energy envelopes as low- and high-frequency vibrations on the thumb and index finger, respectively. The time of onset of the two envelopes was a viable cue for discrimination of voicing features in consonants. This coding scheme was also effective in providing a substantial benefit to lip-reading in closed-set consonant identification tasks, nonetheless, place and manner of articulation features were not sensed by the participants using tactual-only cues.[1] Using the same stimulator, Israr et al. [7] developed a speech-to-touch coding scheme that extracted features from three spectral bands corresponding to the F0 (fundamental frequency of speech signal), F1 (first formant) and F2 (second formant) frequency regions and presented them as vibrational and motional cues to the fingerpads of the thumb, middle finger and index finger, respectively. The formants are resonances in the speech spectrum that occur due to the shape of the vocal tract, the location of constriction (such as the tongue touching the ceiling of the vocal tract) as well as the place where

---

6100 Main Street MS-321, Houston, Texas 77005. The second author is now affiliated with Disney Research, Pittsburgh. 4615 Forbes Ave, Suite 420, Pittsburgh, PA 15213. E-Mail: eyw1-, omalleym@rice.edu, israr@disneyresearch.com.

[1] Voicing is a phonetic feature related to the presence of vocal cord vibration during the production of consonants. About half of the consonants are voiced and other half are unvoiced (for example, /b/ is a voiced consonant and /p/ is an unvoiced consonant). The place and manner of articulations are related to the location where the consonant is produced and the manner it is uttered, respectively. The consonant /b/ and /p/ are produced by joining the lips, thus bilabials defines the place of articulation feature, and a burst of air is released after a small occlusion pause, defines plosives as the manner of articulation. The only distinction between /b/ and /p/ is the voicing. Almost all consonants in the English and other languages can be differentiated by the three consonantal features, i.e. voicing, place and manner articulation, as described by the International Phonetic Association.

IEEE Haptics Symposium 2010
25 - 26 March, Waltham, Massachusetts, USA
978-1-4244-6822-5/10/$26.00 ©2010 IEEE

stress is concentrated in vocal tract during the production of a speech sound [8]. The formants are numbered by their corresponding resonance number (i.e. the first resonance in the spectrum is called the first formant). The first formant typically has a range of 300 Hz to 800 Hz for male speakers and slightly higher for female speakers. Similarly, the second formant ranges from about 900 Hz to 2400 Hz for males and slightly higher for females. The coding scheme showed promising performance in vowel identification and pair-wise consonant discrimination tests when only tactual cues were presented to the trained users [7, 9]. Unlike previous efforts, all three consonantal features (voicing, place, and manner of articulations) were successfully transmitted, mainly due to the multidimensional nature of the Tactuator device that stimulated low-frequency motional and high-frequency vibrational waveforms in a continuum,

The uniqueness of the study by Israr et al. [7] was that it provided a representation of the articulation location feature that was not transmitted with prior tactile aids (see a review of prior efforts in [7]). The coding scheme monitored variations in the first two formant energies and presented these transitions as motional cues to the two fingerpads. So, if the formant energy increased in frequency, the corresponding finger channel extended, and conversely it flexed for decreasing formant frequency. Analysis of the second formant transitions showed that the index finger extended more for consonants that were produced well inside the mouth than consonants that were produced close to the lips.

This second formant transition coding is utilized in the present study, where instead of the sensitive finger motion, the transitions are mapped to the skin of the dorsal forearm using a custom designed tactile sleeve worn by the users. The vibrators (called tactors) are mounted inside the sleeve and temporally turned on/off while keeping the fixed frequency and amplitude of stimulation. The number of tactors and spacing of tactors are determined by conducting two psychophysical experiments. The first experiment is conducted to determine the information transmission (IT) and localization performance of the forearm skin. This was done to determine the number of tactors used in the subsequent speech study. The second experiment determines the relationship between the physical spacing of the tactors on the forearm and the mental representation of location of vibrations perceived by the human participants. Thus using the relationship, the tactors can be mounted on the forearm having optimal localization performance. The results of these two psychophysical experiments are then used to design an effective tactile coding system for the subsequent speech experiment. The goal of the present effort is to device simple tactile translations of constrictions inside the mouth (that is not directly seem by the deaf using lip-reading) as cues on the forearm that can be supplemented with lip-reading for speech communication by impaired individuals.

The organization of the paper is as follows: The design and description of the tactile stimulator is presented in Section 2. Section 3 presents the methods and results of the two psychophysical experiments. The speech experiment is discussed in Section 4 and the paper concludes with some remarks in Section 5.

2 TACTILE STIMULATOR

We used a custom designed tactile sleeve as the tactile stimulator in this study. The tactile sleeve consisted of a leather arm wrap and a control box shown in Fig. 1. The arm wrap was worn along the length of the arm and fastened with Velcro straps. The inside of the arm wrap had Velcro attachments used to mount tactors along a line or in an array configuration. A thin replaceable layer of padding was placed between the skin and tactors in order to reduce itchiness and discomfort for the user. Two sleeves were made to accommodate different forearm lengths. The smaller sleeve could accommodate forearms up to 23 cm long, while the longer sleeve could be used for individuals with forearms up to 30 cm long. The length of the forearm (between the wrist and the elbow) was measured for each participant before donning the sleeve and the appropriate sized sleeve was used for all testing. In this study, the tactors were always arranged along a line and touching the posterior (dorsal) side of the left forearm. The tactors were placed equidistant from each other covering the entire length of the participant's forearm. The number of tactors used varied across experiment.

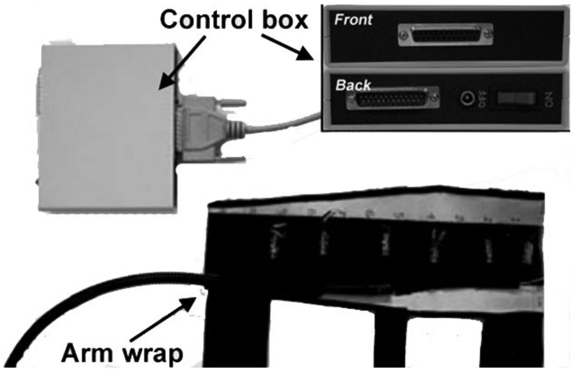

Figure 1. The tactile Sleeve. Shown is the control box and six tactors mounted on the arm wrap

The control box housed circuitry to power and control up to 12 tactors. The tactors were linear motors each with an unbalanced mass attached to the shaft (model 256090, Jameco Electronics, Belmont, CA) and housed in acrylic blocks. Each block along with the motor weighed approximately 3.5 grams, provided a contact area of 125 mm^2, and vibrated at constant frequency of about 180-200 Hz with perceivable amplitude. The circuitry allowed tactors to operate in a logical on/off fashion. The 25-pin D-connector at the front of the control box connected it to the tactors and the back end connector connected it to the 24 pin USB based digital input-output module (model NI USB-6501, National Instruments Corp., Austin, TX). Details of the circuitry and operation of the control box are described in [10].

3 PSYCHOPHYSICAL EVALUATION

We conducted two experiments to determine how spatial information presented on the skin of the forearm could be used by human users wearing the tactile stimulator. In the first experiment, we used the absolute identification paradigm to determine number of vibratory locations reliably identified by the participants on the dorsal forearm. In the second experiment, we determined the mental mapping of physical locations of the forearm by asking participants to rate the location of vibration along the length of the dorsal forearm between numbers 0-100, where 0 corresponded to the location at elbow and 100 corresponded to that at the wrist. The experimental procedures for the two experiments are explained below.

3.1 Participants

Eight participants (six males and two females) between ages 19 and 27 years old (average age 21 years) took part in the psychophysical evaluation experiments. All participants were healthy students at Rice University and had no known sensory/tactile impairments. They signed consent forms (approved through Rice University's IRB) at the beginning of the experiments and were compensated with extra credit in a required undergraduate course.

3.2 Absolute Identification Paradigm

3.2.1 Procedures

In this experiment, we asked participants to identify the location of the vibration they felt on their forearm for a brief period of time. Participants sat comfortably in front of the computer screen in the upright posture and wore the tactile sleeve on the left forearm such that the mounted tactors stimulated the skin of the dorsal side of the forearm, as shown in Fig. 2. A one-interval five-alternative forced-choice (1I-5AFC) absolute identification paradigm was used to determine information transfer capabilities of the forearm's skin. Five tactors were placed equidistant in a line along the length of the forearm using the tactile sleeve. In each trial, one randomly selected tactor was turned "on" for 50-milliseconds. Participants were instructed to respond by pressing a button corresponded to the location of the vibration with a right-handed mouse. After their response was recorded a new trial began. No correct answer feedback was provided at the end of a trial.

Each participant completed two runs of 75 trials. In a run, each tactor was turned "on" an equal number of times. The duration of a run was no more than 6 minutes for any participant, and they were asked to rest between runs. Both runs were completed in a single 20 minute test session. Before the start of the experiment, participants completed a few training trials to familiarize themselves with the set-up, apparatus and vibratory stimulus. During the experiment, participants donned noise cancellation head-phones that played pink noise in order to mask environment and mechanical noise. Visual instructions on the computer screen and auditory tones were provided to indicate the start of a trial during the experiment, and a pictorial illustration of the location of five tactors on the forearm was also displayed on the screen (see monitor screen in Fig. 2).

Figure 2. A participant sitting in the upright posture during the experiment and a graphical instruction of tactors on the forearm.

3.2.2 Data Analysis

Results of the identification experiments were expressed in terms of information transfer (IT) [11]. A 5×5 stimulus-response confusion matrix was formed and the maximum likelihood estimate of IT was calculated by using:

$$IT_{est} = \sum_{j=1}^{k}\sum_{i=1}^{k} \frac{n_{ij}}{n_i n_j} \log_2\left(\frac{n_{ij} n}{n_i n_j}\right) \tag{1}$$

where k = 5 was the number of stimulus alternatives, n was the total number of trials, n_{ij} was the number of times the joint event (S_i, R_j) occurred, and $n_i = \sum_{j=1}^{k} n_{ij}$ and $n_j = \sum_{i=1}^{k} n_{ij}$ were the sum of trials for each row and column, respectively. Localization performance was also determined by computing the percentage-correct (PC) scores of identifying each tactor location correctly.

3.2.3 Results

Individual and combined performance in terms of information transfer (IT) and percentage correct (PC) scores obtained from the forced-choice absolute identification experiment are shown in Table 1. The IT scores ranged from 1.32 to 1.84 bits with a combined IT of 1.49 bits. This corresponded to participants' ability to identify 2-3 categories (or tactors) correctly. Based on the IT scores, six out of eight participants would identify three tactors every time while the other two participants would identify two tactors arranged on the forearm.

Table 1. Performance scores of eight participants in 1I-5AFC forced-choice absolute identification paradigm

Participants	IT_{est} (bits)	Percent Correct	Categories
S1	1.75 bits	88%	3.4
S2	1.78 bits	87%	3.4
S3	1.58 bits	77%	3.0
S4	1.32 bits	75%	2.5
S5	1.78 bits	87%	3.4
S6	1.57 bits	79%	3.0
S7	1.29 bits	67%	2. 5
S8	1.84 bits	88%	3.6
Overall	**1.49 bits**	**81%**	**2.80**

The localization performance of the combined data (among participants) is shown in Fig. 3. Each data point shows the percent correct identification of tactors located along the length of the forearm and the error bars show the standard error of the mean. An ANOVA analysis showed that the localization performance was significantly different [$F(4,35)=3.32$, $p<0.05$] at different locations. Participants were able to correctly identify each tactor with at least 70% accuracy, with the extreme and center tactors resulting in better performance than the middle two tactors.

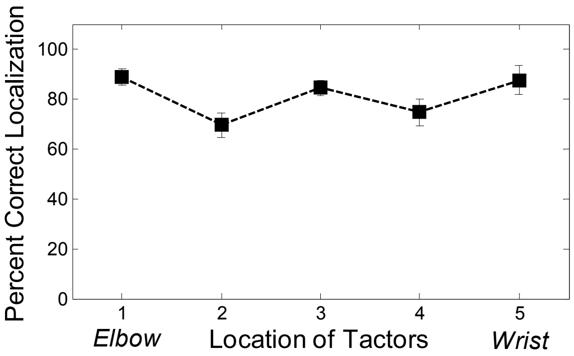

Figure 3. Mean vibrotactile localization scores for eight participants. The error bars show standard error of the mean

3.3 Numeric Location Estimation

3.3.1 Procedure

In this experiment, participants were instructed to rate the location of one out of ten randomly selected vibrating tactors as a fraction of the total length of the forearm. The sitting posture and experimental set up were the same as that used in the identification experiment. The procedure used in the present study followed those described in the literature for magnitude estimation [12]. Ten equally spaced tactors were mounted on the tactile sleeve that covered the entire length of the dorsal left forearm. In each trial, participants first felt vibrations through the extreme two tactors (one closest to the elbow and the other closest to the wrist) for 200-milliseconds. After a pause of 700-milliseconds, a randomly selected vibrator (test stimulus) was turned ‘on’ for 100-milliseconds. The reason for turning the extreme tactors ‘on’ at the beginning of the trials was that participants could calibrate the length of the forearm before making a judgment for the location of the test stimulus. They were instructed to rate the location of the test stimulus in the range 0 to 100. The number 0 corresponded to the zero distance from the elbow and the number 100 corresponded to the full length of the forearm. The rated number could be an integer, a fraction or a decimal numeric value. Participants were asked not to rate the location of the vibration based on their ratings in the previous trials. They were given an option to either enter the number with their right hand using the number pad of a computer keyboard or verbally say it loud for the experimenter to record it for them. Each participant completed 80 trials in a single session that lasted no more than 10 minutes.

Training trials were provided before the experiment in order for the participants to familiarize themselves with the vibrations and to understand the experiment procedure. In the training, participants could feel a vibration very close to the elbow, midpoint between the elbow and the wrist, and, very close to the wrist. They were told that these vibrations corresponded to a very small number (close to 0), a number in the middle of the 0 to 100 range and a very large number (close to 100), respectively.

3.3.2 Data Analysis

The numerical subjective ratings were normalized by dividing the rating of each test stimulus by the average rating of the corresponding session and then multiplying by the overall average of all ratings in the experiment [13]. There were a total of 10 test stimuli, one at location 0 (closest to the elbow) and the remaining 9 at successive $1/9^{th}$ fractions of the forearm length apart. The normalized ratings were regressed against the normalized length of the forearm (normalized as the fraction of the forearm length) using a straight-line function in order to determine the mapping of physical locations of the tactors and mental (internal) representation of the stimulating locations. The mean and standard errors of the normalized subjective ratings were also computed at each of the 10 fraction points and plotted against the normalized length of the forearm.

3.3.3 Results

Figure 4 shows the relation between the normalized length (fractional length) of the forearm (in abscissa) and the normalized ratings judged by the participants (in ordinate). Each data point shows the average rating at a specific fractional length and the error bars indicate the standard error of the mean. The regressed straight line function fits very well with the average data points (correlation coefficient, R = 0.93, degree-of-freedom = 638) indicating that the mental mapping of the physical locations on the forearm is linearly projected. However, the slope of 0.89 ($p<0.001$) indicated that the perceived range is slightly condensed.

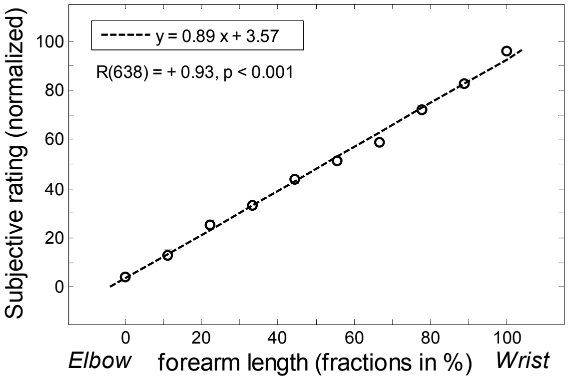

Figure 4. Mean subjective rating (normalized) as a function of the normalized forearm length

3.4 Discussion

In this section, two psychophysical experiments were conducted to understand users’ ability to process tactile information presented spatially along the length of the dorsal forearm. The vibrations presented on the forearm were fixed, i.e. their location, amplitude and frequency remained the same, and the only changing variable was the temporal onset of the spatially distributed cues. This type of tactile feedback scheme is simple to control and economical to produce. However, understanding how the spatial information is processed by the human somatosensory system is necessary for this approach to be worthy of further study.

The processing capabilities of the human tactile sensory system have been studied for various parts of the body, such as fingertips [14, 15], hand [16], tongue [17], head [18], thigh [19], velar forearm [20], abdomen [21] and back [19], but that for dorsal forearm is not available. The present study covers the gap by presenting results of the identification and subjective location estimation experiments.

The results of the present experiments showed that 2-3 tactors placed on the dorsal forearm could be reliably identified by human users, when only location of the vibration was varied. The IT was similar to that with the finger when only one joint was moved [22]. The observation that the processing ability of fingers is superior to other regions of the body is due to the fact that multiple joints of the fingers are involved in finger motion and dense population of receptors embedded in the skin of fingerpads. The localization performance, on average, was also mediocre with five tactors – participants could not locate any vibration greater than 90% of the time. The poor ability of participants was likely due to scarce and less densely populated mechanoreceptors in the skin of the forearm. The identification and localization capabilities were, however, similar to those on the velar forearm [20] and on some other parts of the human body. Participants in the previous studies were able to localize extreme vibrators better than the vibrators located in the middle. Similarly, in the present study the localization performance was better with the extreme tactors. Additionally, the performance to localize the center tactor was comparable to that of the extreme tactors.

Another objective of these psychophysical studies was to determine the optimal spacing of the tactors arranged along the forearm. The optimal spacing should based on the equally spaced perceived distance between the tactors, whether the perceived spacing is a logarithmic function of the physical distance (as in visual perception and ratio scales in tactile intensity perception) or some other function representing the mental-physical mapping more accurately. The present experiment showed that the skin of the forearm maps linearly on its mental representations in a way that could be well represented by a straight line function. However, less than unity slope indicates that participants were underestimating the true physical range. Hence, for providing tactile feedback through the forearm, tactors could be linearly spaced to elicit an optimized response from the users.

4 SPEECH STUDY

With knowledge of the information transfer capabilities of the skin of the forearm and the perceived relationship between the physical placement of tactors and their perceived location, we conducted an experimental evaluation focused on speech transmission via the tactile sleeve. We conducted a pair-wise discrimination experiment to analyze a human's ability to determine tactile cues derived from the place of articulation in consonants using the skin of the forearm. The details of the experiment are as follows.

4.1 Speech Material, Speech Acoustics and Speech Processing

The speech material comprised of consonant-vowel-consonant (C_1VC_2) non-sense syllables spoken by two female speakers of American descent. The syllables were converted into digital segments and processed offline (for digitizing methods, refer to [7]). Six tokens (3 tokens per speaker) of three plosives (/b/, /d/, /g/), four fricatives (/v/, /tx/ as in those, /z/, /zh/ as in vision) and one affricate (/j/) at the initial consonant location were joined by a medial vowel /ah/ and the final consonant was randomly selected from a set of 21 consonants. The total duration of segments was always less than 2 seconds, including the frames before and after the lip motion to produce syllables.

The speech processing scheme tracked the frequency of the energy peak in the 1150-4000 Hz band of the acoustic signal and presented it as localized vibration along the length of the forearm. This frequency band corresponded to the second and third formant in the speech spectrum. Speech analyses have shown that the transition of formants between consonants and vowels is useful for the distinction of place of articulation feature in stops, fricatives and affricates, as indicated in [23], [24] and [25].

The frequency of energy peak was extracted by passing the band-limited signal through ten contiguous band-pass filters in parallel, and the temporal envelope of each band was then obtained. The envelopes were compared and the center frequency of the band with the largest envelope value was noted at each sample instant. (Refer to [7] cf. Sec. IIIC for details on the peak extraction.) This center frequency was corresponded to the energy peak in the band and linearly mapped to six tactors arranged along the length of the forearm. If the energy peak was around 1150 Hz then the tactor closest to the wrist vibrated, while peaks around 4000 Hz stimulated the tactor closest to the elbow. Thus, the quasi-static location of the vibration on the forearm was intended to present discriminatory cues for the place of articulation in consonants. During the silence of speakers before and after the utterance of a segment, the third vibrator from the wrist was turned "on". This was done to imitate forward and backward masking during the continuous exposure of speech. The extracted data was down-sampled to 100 Hz.

Analysis of the vibratory location data extracted from the speech segments and processing scheme showed that for consonants produced closest to the lips (such as bilabials /b/ and labiodentals /v/) the vibratory signal started stimulating at the middle of the forearm (due to the silence before the utterance) and then quickly moved towards the wrist due to the low second formant of bilabials/labiodentals and the vowel /ah/. When the consonant was produced further inside the mouth (such as dental /tx/ or alveolar /d/) then after the middle vibration due to silence of the speaker, the vibration moved towards the elbow before it dropped towards the wrist. For consonant produced furthest inside the mouth (such as post-alveolar /zh/), the vibration started at the middle, moved towards the elbow and stayed for an instant before it dropped towards the wrist. Thus, the further the consonant was produced inside the mouth, the longer the vibration was cued to skin near the elbow before dropping towards the wrist.

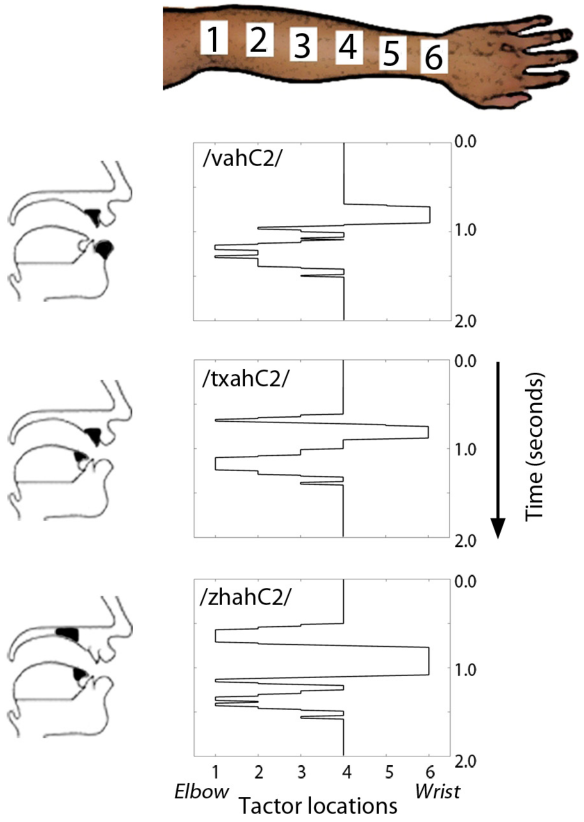

Figure 5. Location of /v/, /tx/ and /zh/ production inside the mouth and associated tactile cues on the forearm.

Figure 5 shows sample temporal sequences of vibrations presented on the forearm and the corresponding location of constriction inside the mouth during the production of consonants /v/ (libiodental, close to the lips), /tx/ (dental, slightly moved away from the lips) and /zh/ (post-alveolar, close to the throat). The figure shows that for /v/ the vibrations start at the middle location on the forearm and fall rapidly towards the wrist due to the medial vowel /ah/. The vibrations for /tx/ start from the middle and slightly rise towards the elbow before falling towards the wrist. Finally, the vibrations start at the middle location on the forearm, rise towards and stay at the elbow before falling towards the wrist for /zh/.

4.2 Experimental Methods

4.2.1 Participants

Two males and two females (18-33 years old, average age 22 years) participated in the study. All were normal healthy students of Rice University with no hearing/visual disabilities. One male (P1) and one female (P2) worked on the development of the speech study; however, only P1 had prior experience with tactile devices and participated in speech studies prior to this one. All participants signed consent forms approved by Institutional Review Board at Rice University.

4.2.2 Procedure

The sitting posture and experimental setup were the same as those used in the psychophysical evaluation experiments above. Participants wore the tactile sleeve, with six tactors arranged equidistantly, on the left forearm. They sat comfortably in front of the computer screen and wore noise cancellation headphones that played pink noise throughout the experiment.

The ability to discriminate consonants that differ in place of articulation features was tested for 12 pairs of initial consonants. The pairs were chosen from voiced plosives, fricatives, and affricate and are shown in the first row of Table 2. The discrimination tests were conducted using a two-interval two-alternative forced-choice paradigm (2I-2AFC) [26]. For each trial, the participant was presented with two stimuli associated with a randomly selected pair of consonants. The order of the two consonants in a pair was randomized with equal *a-priori* probability in each trial. The participant was instructed to press a button corresponding to the order of the consonants presented. The duration of each stimulus interval was 2-seconds with an inter-stimulus-interval of 450-milliseconds. A 150-millisecond auditory tone and a visual phrase indicating “stimulus 1” or “stimulus 2” was presented 200-milliseconds before the start of each stimulus to mark the beginning of each stimulus interval. Participants P3 and P4 received correct answer feedback in the text form “`Your answer was CORRECT`” at the end of each trial, while P1 and P2 did not.

All consonant pairs were intermixed together and randomly presented 60 times in 12 runs. Each participant completed four 60-trial runs in a 45-minute test session for three consecutive days, thus resulting in a total of 720 (60 repetitions of 12 consonant pairs) test trials per participant. They were encouraged to take breaks between the runs. Before each testing run, participants experienced training trials to get familiarized with the test trial and cues associated with different consonants. The training trials were similar to the test trial except that correct answer feedback were presented at the end of the trial and the response was not recorded for future analysis. The participants were also given the option of performing slow training trials. These were similar to the training trials except that the stimuli were extended in time and presented at twice the normal duration. These trials were available for the participants to better understand the tactile cues. Participants terminated training when they felt comfortable enough to begin the experiment.

4.2.3 Data Analysis

For each consonant pair, a 2×2 stimulus-response confusion matrix was obtained, from which the percentage correct (PC) scores, the sensitivity index *d'*, and the response bias *β* were calculated using the following expressions of signal detection theory [26]

$$d' = Z(H) - Z(F) \tag{2}$$

and

$$\beta = -\frac{z(H) + z(F)}{2} \tag{3}$$

where H is the hit rate and F is the false alarm rate. Z(.) is the inverse function of the normal Gaussian distribution and is determined by transforming the cumulative probability under the normalized Gaussian density curve to standard deviation units. An unbiased response is indicated by β=0. The criterion for discrimination is d'=1. The sensitivity index was set to 4.65 (corresponding to a hit rate of 0.99 and a false-alarm rate of 0.01) when the performance was perfect.

4.3 Results

Table 2 shows the performance scores of discriminating all twelve consonant pairs by the four participants as well as the overall performance scores. The discrimination scores varied among participants. The performance of P1 (who participated in previous similar speech studies and received no correct answer feedback) was in general better than that of the other three participants. Participants who received correct answer feedback (P3 and P4) performed similar to the participant who obtained no correct answer feedback and had no prior experience (P2).

On average, the sensitivity indices for the eight out of twelve pairs were greater than 1 (criterion for discrimination). The exception was for discriminating /b/ with /g/, /d/ with /g/, /v/ with /g/ and /z/ with /zh/. The pairs /b-g/ and /z-zh/ were discriminated by participants P1 and P2 showing d'>1. The pair /v-g/ was only discriminated by P1 and the final pair /d-g/ was confused by the participants as shown by the non-positive sensitive index. The average bias for discriminating each pair was low, indicating that participants did not prefer one response over another during the experiment.

Table 2. Performance summary in the consonant discrimination experiment. Reported values are the sensitivity indices *d'*.

Participants		**/b-d/**	**/b-g/**	**/b-zh/**	**/b-j/**	**/d-g/**	**/v-tx/**	**/v-zh/**	**/v-g/**	**/v-j/**	**/tx-z/**	**/tx-j/**	**/z-zh/**
P1*		2.08	1.01	4.16	4.65	-1.47	2.80	4.16	1.35	3.83	1.47	0.84	1.57
P2*		0.62	1.05	3.34	1.81	0.52	0.77	2.22	0.26	0.96	1.05	0.62	1.07
P3		1.27	0.93	3.61	2.08	-0.17	0.36	2.22	-0.26	2.95	1.54	1.70	0.46
P4		2.00	0.60	3.00	3.00	-1.05	0.96	3.83	0.68	3.67	1.25	2.23	0.17
Overall	***d'***	**1.41**	***0.87***	**3.38**	**2.60**	***-0.5***	**1.06**	**2.83**	***0.48***	**2.39**	**1.29**	**1.22**	***0.79***
	β	**0.11**	**-0.12**	**0.04**	**0.16**	**0.07**	**-0.15**	**0.09**	**0.08**	**0.09**	**0.17**	**0.06**	**0.23**
	PC	**76%**	**67%**	**95%**	**90%**	**40%**	**70%**	**92%**	**60%**	**88%**	**74%**	**73%**	**65%**

* Participants did not receive correct answer feedback

4.4 Discussion

We conducted an experiment to examine the appropriateness of using the skin of the forearm as a communication medium for speech communication. A speech-to-touch coding scheme was developed in [7] that extracted spectral features from three different regions of speech bands and presented them at three fingertips of the left hand. A subset of the coding scheme was used in this study that monitored frequency variations in the second formant frequency region and presented it as location of vibration along the length of the forearm. A similar scheme has been utilized with other tactile aids, such as Queen's University tactile vocoder and Tactaid 7, which utilized a broad frequency spectrum range of speech acoustics [27, 28] and presented vibratory patterns on the forearm. The difference between the previous and present schemes is the features extracted from the band of acoustic information. The purpose of the present coding scheme was to translate the location of constriction inside the mouth (place of articulation) during production of consonants as simple tactile translations. The reason for focusing on the place of articulation was that this cue was not directly seen by the deaf using lip-reading, and the proposed tactile coding could be supplemented with lip-reading for speech communication.

This paper extends the analysis of the efficacy of the coding scheme developed in [7]. The scheme was able to discriminate vowels (which are usually longer and more stable than consonants) with an ideal vibrational observer (computational model) as well as by using it on a human participant [9]. The consonants are, however, shorter in time and have varying spectral features (for example a stop in plosives) and are difficult to detect even with a trained speech recognition system. In [7], the discrimination ability of selected consonants by two users using tactile and motional cues were promising at the hand, however, the pair of consonants presented remained the same throughout the run. In the present study, the pairs of consonants were pooled together and presented together in the test runs, thus the users did not focus on a specific difference, generalize the discrimination and making the discrimination harder. In addition, the present paper utilized discrete locations of simple vibrations on the forearm rather than the continuous motion of the fingers, as was done in [7].

This paper proposes a simple scheme to present place of articulation feature in consonants by extending vibration cues towards the elbow depending on the location of constriction inside the mouth for consonant production. When the consonants were produced at the lip, such as for /v/ and /b/, the vibrations stayed at the midpoint of arm and did not extend closer to the elbow. The vibrations closer to the elbow occurred most for consonants produced closer to the velar, such as /zh/. All CVC nonsense syllables used in the present study had an /ah/ middle vowel that had lower second formant. Thus the vibrations eventually moved towards the wrist, (higher formant corresponded to vibrations closer to the elbow and vice versa) and the direction of cueing (whether it went directly towards the wrist or extended towards the elbow before moved towards the wrist) was an indication about the place of articulation of initial consonants in the CVC syllables.

The performance of contrasting consonants at the initial location of CVC nonsense syllables showed that it was in general better than that with prior tactile aids, and slightly inferior to that reported in [7]. The discrimination performance of place of articulation among selected consonants was close to chance level, i.e. $d'=0$, in almost all prior studies using tactile cues only, (see a review in [7]) however, with the hand the performance was much higher in [7], i.e. d' ranged from 1.46 to 3.80. The present study showed that for some consonants the discrimination was good, i.e. $d'>2$ but the performance varied among participants. The discrimination performance was better when the contrasting consonants varied in more than one feature (i.e., contrasting bilabial plosives (such as /b/) with post-alveolar fricative (/zh/) or affricate (/j/)). This was consistent with previous findings that consonants were better discriminated when two features were changed than when only one was changed. (see [27] and [29] cf. Table 1.3 at page 19-23 for comparison of several past studies.) Moreover, when the consonants were produced further apart inside the mouth, they were easier to discriminate (for example, compare /v/ and /tx/ with /v/ and /zh/). The exception was observed with plosives, where the discrimination was better with /b/ and /d/ than with /b/ and /g/. The reason for not discriminating plosives was highlighted by the fact that consonants have fine high frequency contents (such as both second and third formants) that were required for discrimination of consonants [30]. The current coding scheme combined second and third frequency bands together and extracted a single feature from the broad high frequency (1150-4000 Hz) band. It is, however, not clear if the performance differences in the present study and in [7] were due to *1)* presentation of continuous temporal variation of sensitive finger motion as opposed to the discrete vibration points along less innervated forearm, or *2)* training of participants. In any case, the current coding scheme requires further investigation and adjustment based on the loci of tactile stimulation.

5 CONCLUDING REMARKS

In this paper, the ability to process broadband information, such as that extracted from speech signals, is analyzed on the dorsal side of the human left forearm. The psychophysical evaluation was carried out using two methods, one to determine localization and information transmission capability, and other to determine the perceived mapping of vibrating locations on the skin of the forearm. The skin of the forearm is generally inferior to that of the finger in its ability to localize vibrations as well as in terms of processing vibrational information presented. Users were able to correctly identify about three tactors on the forearm at the 80% correct level. The information transfer also corresponded to the same number of categories users were able to identify (IT=1.49 bits corresponding to 2-3 vibration locations). We found that the physical locations of vibrations were mapped linearly in the processing mechanism of the brain. This indicated that when placing tactors (or vibration contacts) on the forearm, no special placement techniques are required and they can be mounted based on physical displacements that the designer wishes to convey.

Using psychophysical parameters determined in the two preliminary psychophysical studies, six equally spaced tactors were mounted on the dorsal skin of the left forearm that were used to map variations in energy in speech acoustic as location of vibrations. The extracted acoustical features were corresponded to the second formant and quasi-static formant transitions between consonant-vowel pair, critical for distinction of consonants that varied in place of articulation. The formant transitions were mapped to the six discrete vibration points on the forearm. Thus, if the formants decreased from initial-consonant to medial-vowel then this variation would be presented as movement of vibrations towards the wrist. Analysis of pair-wise consonant discrimination test showed that the speech-to-touch method was adequate for distinction of place of articulation in consonants, however, the performance varied among participants. In general, the present coding scheme worked well with fricatives and affricated, and needs some tweaks for plosives. This paper shows a simple technique to present place of articulation features that was not transmitted through tactile aids in prior studies.

ACKNOWLEDGEMENTS

Elaine Wong was supported by the Brown Undergraduate Research Internship Program (BURIP) at Rice University.

REFERENCES

[1] R. H. Gault. Progress in Experiments of Tactual Interpretation of Oral Speech. *Journal of Abnormal and Social Psychology*, 19:155-159, 1924.

[2] H. Z. Tan and A. Pentland. Tactual Displays for Sensory Substitution and Wearable Computers. In W. Barfield & T. Caudell (Eds.), *Fundamentals of Wearable Computers and Augmented Reality*, Mahwah, NJ: Lawrence Erlbaum Associates, pages 579-598, 2001.

[3] L. E. Bernstein, M. E. Demorest, D. C. Coulter, and M. P. O'Connell. Lipreading Sentences with Vibrotactile Vocoders: Performance of Normal-hearing and Hearing-impaired Subjects. *The Journal of the Acoustical Society of America*, 90(6): 2971-2984, 1991.

[4] C. M. Reed, W. M. Rabinowitz, N. I. Durlach, L. D. Braida, S. Conway-Fithian, and M. C. Schultz. Research on the Tadoma Method of Speech Communication. *The Journal of the Acoustical Society of America*, 77(1):247-257, 1985.

[5] C. M. Reed and N. I. Durlach. Note on Information Transfer Rates in Human Communication. *Presence: Teleoperators and Virtual Environments*, 7(5):509-518, 1998.

[6] H. Yuan, C. M. Reed, and N. I. Durlach. Tactual Display of Consonant Voicing as a Supplement to Lipreading. The Journal of the Acoustical society of America, 118(2):1003-1015, 2005.

[7] A. Israr, P. H. Meckl, C. M. Reed, and H. Z. Tan. Controller Design and Consonantal Contrast Coding using a Multi-finger Tactual Display. *The Journal of the Acoustical Society of America*, 125(6): 3925-3935, June 2009.

[8] R. Jakobson, C. G. M. Fant, and M. Halle. *Preliminaries to Speech Analysis*. Cambridge: MIT Press, 1963.

[9] A. Israr, C. M. Reed, and H. Z. Tan. Discrimination of Vowels with a Multi-finger Tactual Display. *Proceedings of 16th Symposium on Haptic Interfaces for Virtual Environments and Teleoperator Systems,* pages 17–24, March 2008.

[10] A. Israr, D. P. Eng, S. S. Vedam, and M. K. O'Malley. A Low Cost Vibrotactile Array to Manage Respiratory Motion. *Proceedings of Third Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, pages 619-620, March 2009.

[11] W. R. Garner. *Uncertainty and Structure as Psychological Concepts*. New York, USA: John Wiley & Sons, Inc, 1962.

[12] J. J. Zwislocki and D. A. Goodman. Absolute Scaling of Sensory Magnitudes: A Validation. *Perception and Psychophysics*, 28(1):28-38, 1980.

[13] R. L. Klatzky, S. J. Lederman, C. Hamilton, M. Grindley, and R. H. Swendsen. Feeling Textures through a Probe: Effects of Probe and Surface Geometry and Exploratory Factors. *Perception and Psychophysics*, 65(4):613-631, 2003.

[14] H. Z. Tan, N. I. Durlach, C. M. Reed, and W. M. Rabinowitz. Information Transmission with a Multifinger Tactual Display. *Perception and Psychophysics,* 61(6): 993-1008, 1999.

[15] A. Israr, H. Z. Tan, and C. M. Reed. Frequency and Amplitude Discrimination along the Kinesthetic-cutaneous Continuum in the Presence of Masking Stimuli. *The Journal of the Acoustical Society of America*, 120(5):2789-2800, 2006.

[16] S. J. Bolanowski, G. A. Gescheider, R. T. Verrillo, and C. M. Checkosky. Four Channels Mediate the Mechanical Aspects of Touch. *The Journal of the Acoustical society of America*, 84(5): 1680-1694, 1988.

[17] H. Tang and D. J. Beebe. Tactile Sensitivity of the Tongue on Photolithographically Fabricated Patterns. *Proceedings of the First Joint BMES/EMBS Conference*, page 633, 1999.

[18] K. Myles and J. T. Kalb. Vibrotactile Sensitivity of the Head. Army Research Laboratory Report ARL-TR-4696, 2009.

[19] L. A. Scadden. Tactile Pattern Recognition and Body Loci. *Perception*, 2(3):333-336, 1973.

[20] R. W. Cholewiak and A. A. Collins. Vibrotactile Localization on the Arm: Effects of Place, Space, and Age. Perception and Psychophysics, 65(7):1058-1077, 2003.

[21] R. W. Cholewiak, J. C. Brill, and A. Schwab. Vibrotactile Localization on the Abdomen: Effects of Place and Space. *Perception and Psychophysics*, 66(6): 970-987, 2004.

[22] H. Z. Tan, M. A. Srinivasan, C. M. Reed, and N. I. Durlach. Discrimination and Identification of Finger Joint-angle Positions using Active Motion. *ACM Transactions on Applied Perception*, 4(2): Article 10, 14 pages, 2007.

[23] A. M. A. Ali, J. Van der Spiegel, and P. Mueller, Acoustic-Phonetic Features for the Automatic Classification of Stop Consonants. *IEEE Transactions on Speech and Audio Processing*, 9(8):833-841, 2001.

[24] A. M. A. Ali, J. Van der Spiegel, and P. Mueller. Acoustic-Phonetic Features for the Automatic Classification of Fricatives. *The Journal of the Acoustical Society of America*, 109(5):2217-2235, 2001.

[25] A. Jongman, R. Wayland, and S. Wong. Acoustic Characteristics of English Fricatives. *The Journal of the Acoustical Society of America*, 108(3):1252-1263, 2000.

[26] N. A. Macmillan and C. D. Creelman. *Detection Theory: A User's Guide*. Second ed., New York: Lawrence Erlbaum Associates. 2004.

[27] J. M. Weisenberger and M. E. Percy. The Transmission of Phoneme-level Information by Multichannel Tactile Speech Perception Aids. *Ear and Hearing*, 16(4):392-406, 1995.

[28] P. L. Brooks and B. J. Frost. Evaluation of a Tactile Vocoder for Word Recognition. *The Journal of the Acoustical Society of America*, 74 (1):34-39, 1983.

[29] A. Israr. *Tactual Transmission of Phonetic Features*. Ph.d Thesis, Purdue University, West Lafayette, USA, 2007.

[30] J. P. Olive, A. Greenwood, and J. Coleman. *Acoustics of American English Speech: A Dynamic Approach*. New York: Spinger-Verlag, 1993.