Tactile Communication of Speech

RLE Group
Sensory Communication Group

Sponsor
National Institutes of Health/National Institute on Deafness and Other Communication Disorders
Grant 2 R01 DC00126

Project Staff
Lorraine A. Delhorne, Nathaniel I. Durlach, Joseph Frisbie, Eleanora Luongo, Theodore Moallem, Charlotte M. Reed, Hanfeng Yuan

Goals and Significance

The long-term goal of this research is to develop tactual aids for persons who are profoundly deaf or deaf-blind to serve as a substitute for hearing in the reception of speech and environmental sounds. This research can contribute to improved speech reception and production, language competence, and environmental-sound recognition in such individuals. This research is also relevant to the development of improved tactual and haptic displays for a broad class of applications (including virtual-environment and teleoperator systems in addition to sensory aids).

Research is being conducted in two major areas. Work in *Area 1* (Basic Studies of Human Touch) is designed to increase our knowledge concerning the transmission of information through the sense of touch. This research includes theoretical and experimental studies concerned with dynamic information transfer as well as experimental work designed to increase our understanding of the psychophysical properties of the sense of touch. Work in *Area 2* (Tactual Displays of Speech and Environmental Sounds) is concerned with the application of tactual displays to sensory aids for persons who are profoundly deaf or deaf-blind. This research includes studies related to the processing and display of speech and environmental sounds through the tactual sense as well as studies concerned with evaluations of performance achieved through these displays.

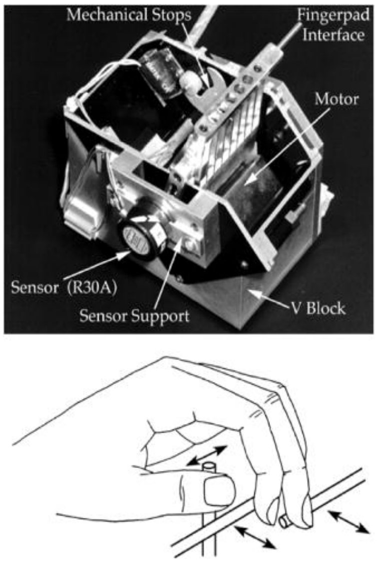

FIG 1. Diagram of Tactuator.

Current Studies

Development of Tactual Stimulating Devices

Work has been conducted on the implementation of an upgraded version of the Tactuator device (Tan, 1996; Tan and Rabinowitz, 1996) used in our experimental work at MIT as well on the development of a new tactual stimulating device for use in experiments at our sub-contractual site of Purdue University.

Hardware and Software Development for Upgrade of Tactuator System

The tactual stimulating device developed for our research is a three-finger display capable of presenting a broad range of tactual movement to the human fingers. It consists of three mutually perpendicular rods that interface with the thumb, index finger and middle finger in a manner that allows for a natural hand configuration (see bottom panel of Fig. 1). A photograph of the motor assembly (with labeled components) associated with one of the rods is provided in the upper panel of Fig. 1. The device is well-suited for use in psychophysical tactual experiments based on its capability for delivering frequencies along a continuum from dc to 300 Hz over a wide range of motions, its linearity, and its insensitivity to loading of a finger on the rods. Modifications were made to this system to improve its performance capabilities for use in experiments requiring multi-modal signal control and real-time processing of speech signals. The hardware components of the current system (including motor assemblies, fingerpad interface, angular-position sensor, power amplifier, and supporting structures -- see Fig. 1) remained unchanged from those of the original system. The changes to the controller components of the device include a new host computer with a more powerful platform for multimedia operations, a new DSP card which supports real-time processing of speech signals, and an electronic analog PID controller. The performance of the current system is similar to that of the original Tactuator in terms of its frequency response, linearity, low harmonic distortion, and little crosstalk (see Brughera, 2002 and Yuan, 2003 for a complete description). This system has been utilized in the new studies undertaken at MIT over the past four years.

Development of Tactual Stimulating System for use at Sub-Contractual Site (Purdue University)

A multi-finger tactual stimulator device, Tactuator II, has been developed for use in experimental work at our sub-contractual site (Israr et al., 2004, 2006a). An innovation in the design of this device is the preservation of the relative amplitude of spectral components in terms of their perceived intensity by human observers. This specification was achieved through the use of a two degree-of-freedom controller consisting of a feedback controller and a pre-filter, and its digital implementation. The device employs a hardware configuration identical to that of the original Tactuator together with a new DSP system that implements the new controller. Measurements with the new system confirm that the steady-state frequency response closely follows the design specification. Measurements of human detection threshold curves indicate a close correspondence with those modeled on the basis of threshold measurements reported by Bolanowski et al. (1988) for frequencies above 30 Hz, and higher thresholds by roughly 10 dB at frequencies below 30 Hz (consistent with measurements made on the MIT Tactuator systems by Tan and Rabinowitz, 1996 and Yuan, 2003). The pre-filter has subsequently been re-shaped to compensate for this deviation. The current system can be driven with a broadband signal (up to 300 Hz) while preserving the relative intensity of different spectral components in terms of the relative sensation levels delivered by the Tactuator II.

Basic Studies of Human Touch.

Work was conducted to increase our basic understanding of information transmission through the tactual sensory system, including studies of dynamic information-transfer capabilities through the tactual sense and studies of perceptual interactions between signals presented through different components of the tactual sense, as well as studies of masking, temporal order resolution, and effects of roving parameters on amplitude and frequency discrimination.

Measurements of Information-Transfer (IT) rate for Multidimensional Tactual Signals

Research in this area is directed towards an improved understanding of the properties that contribute to optimizing information-transfer (IT) rate. Such knowledge is important not only in connection with the design of improved aids for persons with sensory impairments, but also in connection with the design of improved displays for normally-sensed users of data visualization systems and synthetic environments. In previous research (Tan et al., 1999; Tan et al., 2003), we relied on a set of hypotheses to estimate tactual information-transfer (IT) rates from experiments where the subject’s task was to identify one signal in a specified location within a sequence of two or three consecutive stimuli (using backward, forward, and sandwiched masking paradigms) -- see timeline depiction in Fig. 2a.

In the current experiment (Reed et al., 2003), IT rate was measured for the identification of sequential streams of multidimensional tactual stimuli. Signals were composed using spectral components in the low-frequency (2 or 4 Hz), mid-frequency (30 Hz), and high-frequency (300 Hz) regions of the tactual sensory system. Seven waveforms, including three single-frequency stimuli, three two-frequency stimuli, and one three-frequency stimulus, were presented at four different finger configurations (thumb alone, index finger alone, middle finger alone, or all three digits stimulated simultaneously). The 28 resulting signals were presented through our tactual stimulating device (see Fig. 1). Three subjects were trained to identify signals (at each of two durations - 250 or 125 msec) presented in sequences of two (Fig. 2b) or three (Fig. 2c) items using interstimulus intervals (ISIs) in the range 0-640 msec. For purposes of comparison with previous experimental results (Tan et al., 1999; Tan et al., 2003), subjects were also tested in an AXB paradigm where the task is to identify only the middle signal X in a series of three stimuli. IT rate was estimated on the basis of three parameters: the amount of information in the stimulus set (IS= $\log_2$ (K), where K is the number of stimulus alternatives), the identification error rate (e), and stimulus-onset asynchrony (SOA, defined as the sum of signal duration and inter-stimulus interval). Specifically, IT rate=[(IS)x(1-2e)]/SOA (see derivation of this formula in Tan et al., 1999; Reed and Durlach, 1998). Across subjects, the maximum values of IT rate ranged from roughly 6-18 bits/sec for the AXB paradigm, 4-12 bits/sec for two-stimulus sequences, and 4-9 bits/sec for three-stimulus sequences. The IT rates observed here on the sequential identification tasks (4-12 bits/sec) are within the range of IT rates estimated for communication under several different natural methods of tactual communication employed by deaf-blind individuals and thus should be sufficient for supporting communication (e.g., through tactual displays designed to provide supplemental cues to lipreading). We expect that subjects will be able to maintain these levels of performance for longer sequences presented at stimulus rates which correspond to the maximal values of IT rate achieved in the current study.

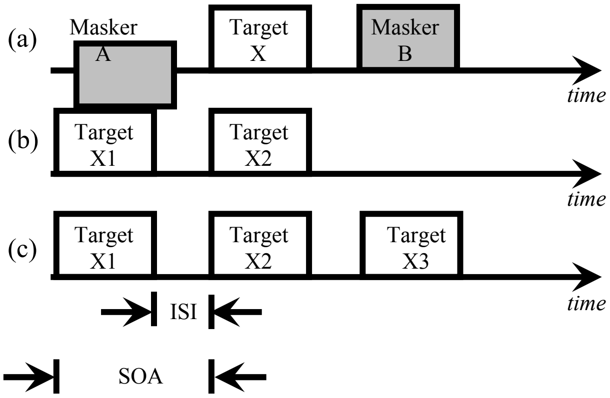

FIG 2. Timeline depiction of three different experimental paradigms.

Masking and Temporal Integration of Tactual Spectral Components

Work was completed on data analysis of experiments concerned with temporal masking properties of multidimensional tactual stimuli delivered to the left index finger (Tan et al., 2003). Seven stimuli composed of one, two, or three spectral components at durations of 125 and 250 msec were employed under three different masking paradigms (forward, backward, and sandwiched masking) in which target identification was studied as a function of ISI in the range of 0 to 640 msec. Performance was similar for backward and forward masking and exceeded that obtained for sandwiched masking with equivalent signal duration and ISI. Analyses of error trials revealed that subjects showed a tendency to respond, more often than chance, with the masker (supporting the notion of masker-response competition), with a composite of the masker and target (providing evidence for temporal integration of masker and target), or with a combination of the target and a component of the masker (indicating greater spatial/temporal summation of mid- and high-frequency components relative to low-frequency components). These basic trends are similar to those obtained in previous studies of tactual recognition masking with brief cutaneous spatial patterns (e.g., Evans and Craig, 1986, 1992; Craig and Evans, 1987, 1995; Evans, 1987; Craig, 1996; Mahar and Mackenzie, 1993).

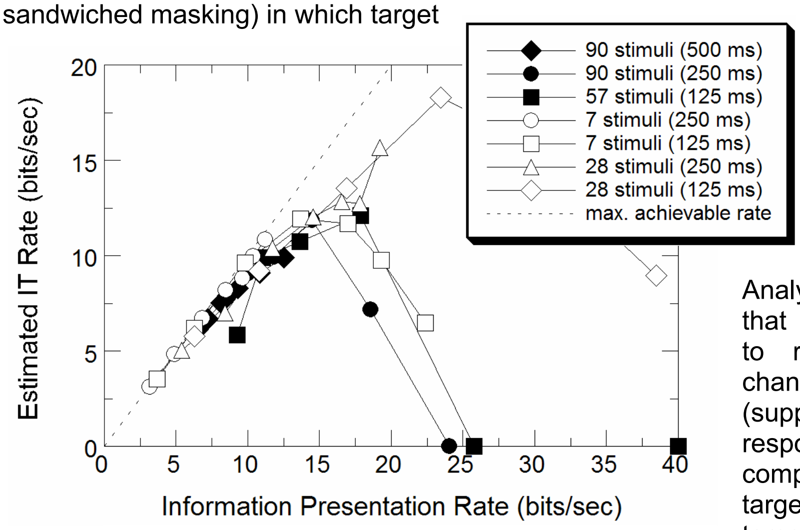

FIG 3. IT rates for 7 different stimulus sets.

The results were also analyzed in terms of IT rate and compared with our previous results (Tan et al., 1999). A summary of results presented in Fig. 3 indicates that (with the exception of an anomalous data point at an IT rate of 18 bit/sec) the peak IT rates clustered around roughly 12 bits/sec independent of stimulus uncertainty (IS) in the range of 2.8 to 6.5 bits and independent of stimulus duration over the range of 125 to 500 msec. For a fixed value of IS, this result is associated with a decrease in optimal stimulus-onset asynchrony (or conversely, an increase in the optimal delivery rate in items/sec) as signal duration decreases. This result contradicts the generally accepted notion of an optimal delivery rate of roughly 2-3 items/sec independent of IS or duration (see Garner, 1962, p. 91; Klemmer and Muller, 1953). These issues are considered in a paper concerned with factors governing the optimization of IT rate (Reed et al., 2006).

Temporal Order Resolution for Tactual Signals

The temporal resolution of the tactual sense was assessed in a series of experiments examining discrimination of temporal onset order (Yuan et al., 2004b, 2005a; Parachuru, 2003) and temporal offset order (Yuan and Reed, 2005; Yuan et al., 2004c). These studies are motivated by, and highly related to, our work on tactual presentation of a temporal-based cue to consonant voicing as a supplement to lipreading.

Temporal Onset-Order Resolution using Redundant Coding of Location and Vibratory Frequency

Tactual temporal-onset order thresholds were measured for two sinusoidal vibrations of different frequencies delivered to two separate locations (left thumb and index finger) of our upgraded tactual stimulating device. The frequency delivered to the thumb was fixed at 50 Hz (T50) and that to the index finger at 250 Hz (I250). The amplitude and duration of each of the two sinusoidal vibrations were roved independently from trial to trial in a one-interval, two-alternative, forced-choice procedure (1I-2AFC), depicted in Fig. 4.

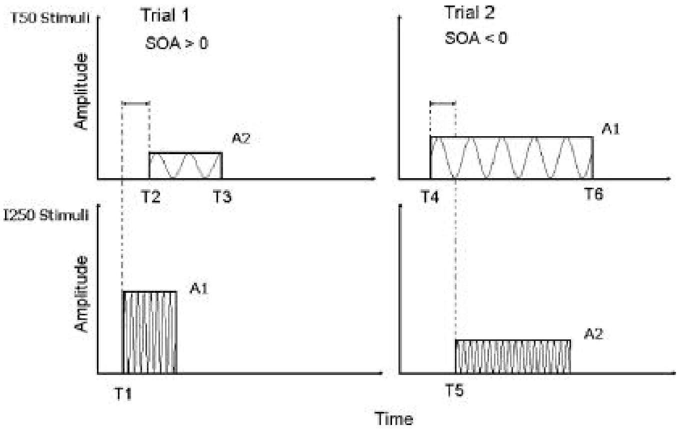

FIG. 4. Timeline for trials in 1I2AFC procedure.

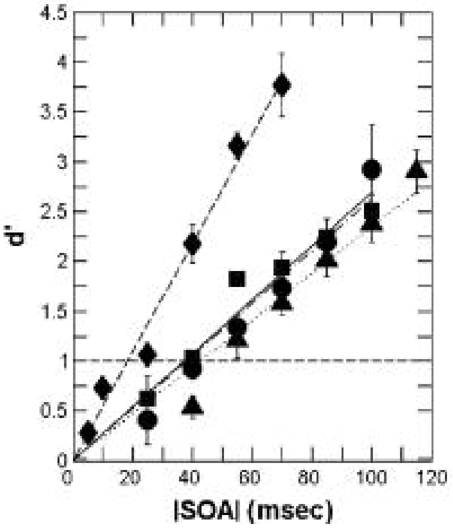

FIG. 5. Performance d' for each of 4 subjects.

Sensitivity d', measured as a function of stimulus-onset asynchrony (SOA), indicated that the temporal onset-order thresholds (defined as SOA required for d'=1.0) averaged 34 msec across subjects (results from each of the four subjects are plotted in Fig. 5.) These results are similar in magnitude to those obtained in previous studies of tactual temporal order employing brief or transient stimuli (Hirsh and Sherrick, 1961; Sherrick, 1970; Craig and Baihua, 1990; Shore et al., 2002). The data were further classified into subsets according to both the amplitude and duration of the two stimuli in each trial of the roving-discrimination paradigm. The results indicated that the amplitude differences of the two stimuli in each trial had a substantial effect on onset-order discrimination, while duration differences generally had little effect. The effects of amplitude differences can be explained qualitatively in terms of amplitude latency relationships and stimulus interactions such as temporal masking. The results of this experiment are relevant to our speech-based measurements of a temporal asynchrony cue to voicing, as well as our perceptual studies of consonant discrimination and identification using a tactual display of a temporal-onset based cue to voicing.

Effects of Frequency and Site of Stimulation on Temporal Onset-Order Discrimination

This research (Yuan et al., 2006; 2004b) extends our study of temporal order resolution of the tactual sensory system through measurements of temporal-onset order discrimination for tonal signals addressing (a) the effects of frequency separation of the two stimuli whose onset orders are to be discriminated and (b) the effects of redundant coding of frequency and site of stimulation on performance. Sinusoidal vibrations were presented either at two separate digits (thumb and index finger of the left hand) or at a single site of stimulation (left index finger) using our tactual stimulation device (see Fig. 1). Measurements were obtained using the 1I-2AFC procedure in which each interval consisted of the random-order presentation of two different stimuli with roving values of amplitude and duration. Thresholds were estimated from psychometric functions of sensitivity d' versus stimulus-onset asynchrony (SOA). On average, temporal onset-order thresholds were larger for one-finger conditions (mean of 74.8 msec) than for two-finger conditions (mean of 48.5 msec) and decreased as frequency separation increased, particularly for single-site presentation. Redundant coding of frequency and site of stimulation resulted in higher resolution by a factor of 1.5 compared to frequency alone and by a factor of 1.2 compared to site alone. The observed advantage for redundant coding supports previous results obtained by Taylor (1978) and Craig and Busey (2003). In terms of the design of tactual displays of speech, these results suggest that site of stimulation may be used to provide a temporal onset-order cue while frequency is used to encode an additional speech parameter.

Temporal Offset-Order Resolution

Tactual resolution for temporal offset order (Yuan and Reed, 2005; Yuan et al., 2004c) was examined in experiments employing a 2I-2AFC procedure with correct answer feedback with two signals: 250 Hz at the index finger (I250) and 50 Hz at the thumb (T50). The amplitude and duration of the two signals were selected at random on each trial from a range of 35 to 40 dB Sensation Level (SL) and 300 to 800 msec, respectively. Performance was measured as a function of stimulus-offset asynchrony (SOFA), defined as $\text{SOFA}=|\text{Offset Time}_{I250} - \text{Offset Time}_{T50}|$. On one interval of each trial (selected at random) SOFA=0 and in the other interval SOFA was greater than 0. The subject's task was to select the interval in which SOFA≠0. For each of four subjects, performance was measured at two values of SOFA (tested in separate runs) selected to yield performance in the range of 60-85% correct. Data were summarized through psychometric plots of d' versus SOFA and threshold was defined as SOFA in msec required for d'=1.0. Thresholds ranged from 97.6 to 231.2 msec across subjects and averaged 142.4 msec. Temporal resolution for offset order is substantially poorer than for onset order for tactual stimulation (by a factor of roughly 4), and may reflect effects of neural adaptation and persistence (Gescheider et al., 1992). The results of this experiment are related to the discrimination of final consonant voicing.

Effects of Roving Background Parameters on Amplitude and Frequency Discrimination

Using the new controller system (Israr et al., 2004) developed at our sub-contractual site at Purdue University, data have been collected on four subjects for amplitude and frequency discrimination of a given reference stimulus (Israr et al., 2006b). These experiments provide insight into interactions between the kinesthetic and cutaneous components of the tactual sensory system and are relevant to the design of tactual display schemes. Twelve reference signals were selected with 6 values of frequency (2 and 4 Hz in the low-frequency region; 15 and 30 Hz in the mid-frequency region; 80 and 200 Hz in the high-frequency region) and 2 values of amplitude (20 and 35 dB SL). Frequency and amplitude discrimination were tested for each of the 12 reference signals under four types of background conditions: no-roving background (C1), one roving background stimulus selected from each of the two frequency regions of which the reference signal was not a member (C2 and C3), and two roving background signals from the two non-member frequency regions (C4). For example, when the reference signal was 2 Hz, the masker in C2 was selected from the mid-frequency region, the masker in C3 from the high-frequency region, and the maskers in C4 from the mid- and high-frequency regions, respectively. A three-interval forced choice (3IFC) paradigm with a one-up three-down adaptive procedure was used to estimate thresholds for both frequency and amplitude discrimination. For both measures, performance decreased in the presence of roving-background maskers. For frequency discrimination, the average Weber fraction ($\Delta F/F$) increased significantly from a value of 0.22 at C1 to values of 0.4, 0.5, and 0.7 at conditions C3, C2, and C4, respectively. Likewise, for amplitude discrimination, ΔA increased significantly from 2.3 dB for C1 to values of 3.4, 3.6, and 4.1 dB at C3, C2, and C4, respectively. In general, there appeared to be greater independence of low- and high-frequency channels, suggesting that they can be employed simultaneously in the development of tactual displays for encoding speech information.

Tactual Displays of Speech and Environmental Sounds

This research is concerned with the development of tactual aids for the deaf for use in the communication of speech and environmental sounds. In the area of tactual displays of speech, studies are focused on the development and perceptual evaluation of displays of consonant voicing. In the area of tactual displays of environmental sounds, our research has been concerned with the design and administration of a survey concerned with the interest of a broad class of deaf and hard-of-hearing individuals in the use of tactual devices for awareness and recognition of environmental sounds.

Improved Tactual Displays of Consonant Voicing

This research is concerned with the development of tactual displays to supplement the information available through lipreading (which is an important means of communication for many persons with profound hearing impairment or deafness). Because voicing carries a high informational load in speech and is not well-transmitted through lipreading, our efforts have been focused on the development of tactual displays of voicing cues to supplement the information available on the lips of the talker. Our research includes (i) the development of signal-processing schemes to extract information about voicing from the acoustic speech signal, (ii) methods of displaying this information through a multi-finger tactual display, and (iii) perceptual evaluations of voicing reception through the tactual display alone and in combination with lipreading.

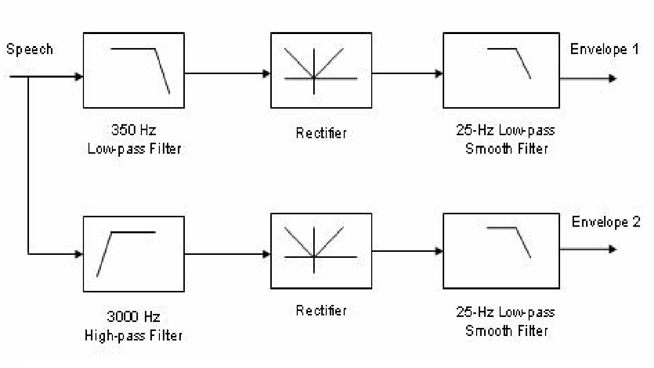

FIG. 6. Block diagram for envelope extraction.

Acoustic Cues to Initial and Final Consonant Voicing

Based on the underlying processes associated with the production of the voicing contrast in speech, we have derived a novel acoustic cue for voicing (Yuan et al., 2003a, b; 2004a,c; 2005b,d; Yuan and Reed, 2005). The signal-processing scheme for extraction of voicing information employs amplitude-envelope signals derived from two filtered bands of speech: a lowpass-filtered band at 350 Hz and a highpass-filtered band at 3000 Hz—see Fig. 6). Acoustic measurements of temporal properties of these envelope signals (that provide a reliable and robust cue to voicing) were derived from audiovisual recordings of C_1VC_2 syllables (V = /i a u/) spoken by two female talkers.

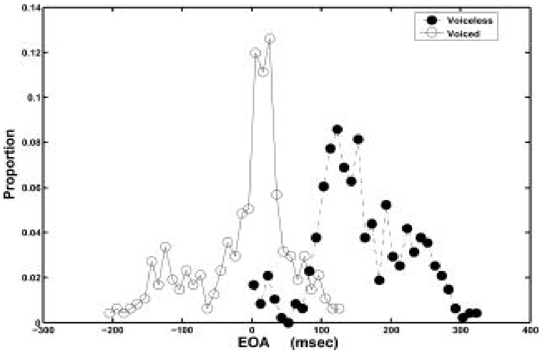

FIG. 7. Probability distribution of EOA for voiced vs. voiceless consonants.

Envelope-Onset-Asynchrony (EOA) Cue for Initial Consonants

Acoustic measurements were made on a set of 1024 C_1VC_2 syllables representing 16 different values of C_1: /p b t d k g f v th tx s z sh zh ch j/ (Yuan et al., 2004a). EOA was defined as the difference in time between the onset of the high-frequency (HF) envelope and the onset of the low-frequency (LF) envelope: $EOA=OnsetTime_{LF} - OnsetTime_{HF}$. The EOA measurements of the 64 tokens representing each value of C_1 were used to derive a probability-distribution function (pdf), from which a cumulative distribution function (cdf) was computed. A Gaussian fit was then made to the cdf of each C_1. For individual voiceless consonants, the means of the best-fitting Gaussian distributions ranged from 54.6 to 233.5 msec (with s.d.s in the range of 24.9 to 107.5 msec). For individual voiced consonants, the means ranged from -122.4 to 68.6 msec (with s.d.s in the range of 2.6 to 69.8 msec). Distribution functions and Gaussian fits were also derived for the two general classes of voiced versus voiceless consonants by pooling the measurements for the 8 voiced consonants /b d g v tx z zh j/ and the 8 voiceless consonants /p t k f th s sh ch/. The distribution functions of EOA values for these two general classes of sounds are shown in Fig. 7. The results of the best Gaussian fits to the cdfs for these two categories indicated a mean EOA of 142.5 msec for the voiceless category (s.d. of 77.2 msec) and a mean EOA of -12.4 msec (s.d. of 66.5 msec) for the voiced category.

Using the EOA as a perceptual-distance measurement, we calculated the performance of an ideal observer in making the voiced-voiceless distinction using the sensitivity measure d'. The calculations make use of the means and s.d.s of the best Gaussian fits and are computed for a 2I-2AFC procedure. Across the 8 pairs of voiced-voiceless contrasts, d' values ranged from 3.5 (for the pair /f-v/) to 13.0 (for the pair /s-z/). These results indicate excellent sensitivity: in terms of percent-correct scores the results fall in the range 96% to 100% correct. The EOA measurement thus provides a reliable and robust cue to initial consonant voicing across a variety of vowel contexts for two different speakers and is independent of manner and place of consonant production. The performance of the EOA measure compares favorably with that reported for voicing-detection algorithms incorporated into schemes for Automatic Speech Recognition (ASR) (e.g., see Nyogit and Ramesh, 2003; Thomson and Chenglavarayan, 2002).

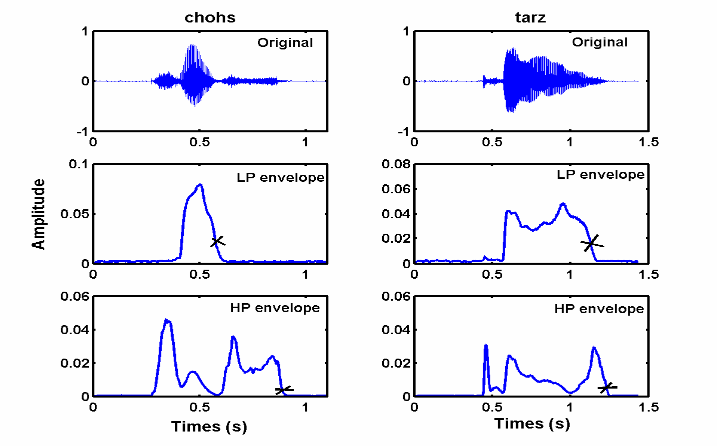

FIG. 8. Illustration of EOFA for two syllables contrasting final /s/ (EOFA=.32sec) and /z/ (EOFA=.12 sec).

Envelope-Offset-Asynchrony (EOFA) Cue to Final Consonant Voicing

Work parallel to that described above for initial consonants was conducted to identify a temporal cue to voicing in final consonants. In this case, EOFA was defined as the difference in time between the offset of the high-frequency (HF) envelope and the offset of the low-frequency (LF) envelope: EOFA=OffsetTime$_{LF}$ – OffsetTime$_{HF}$. Measurements of EOFA were made on at least 63 tokens (and as many as 94 tokens) of each of 16 values of C_2 in C_1VC_2 syllables produced by two female talkers with 16 vowels. An illustration of the EOFA measurement for two syllables contrasting final /s/ and /z/ is provided in Fig. 8. The same methodologies described above for EOA were employed in the current work. That is, for each C_2, a Gaussian distribution was fit to the measurements of EOFA, and the resulting distributions for pairs of voiced-voiceless contrasts were used to calculate the performance of an ideal observer. For voiceless consonants, means of the Gaussian fits ranged from 219.2 to 422.0 msec (s.d. range of 37.4 to 156.7 msec) and for voiced consonants the means ranged from -44.1 to 185.5 msec (s.d. range of 67.6 to 90.6 msec). For the two general categories, the best-fitting Gaussian distributions yielded means of 348.3 msec (s.d. of 104.4 msec) for voiceless consonants and 65.5 msec (s.d. of 98.1 msec) for voiced consonants. Using the Gaussian fits to the EOFA measurements, the performance of the ideal observer (calculated for a 2I-2AFC procedure) ranged from d' of 3.0 (for /p-b/) to 6.3 (for /s-z/). For the two general categories, d' was 4.0. These results indicate excellent separability of the two classes of final consonants on the basis of the EOFA measurement.

Development of Tactual Display of Temporal Cue to Consonant Voicing

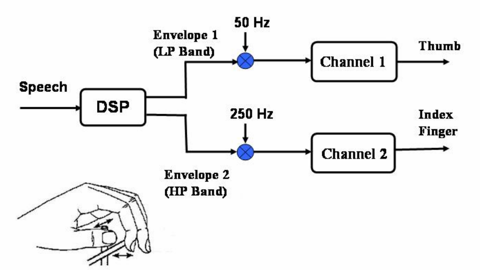

FIG. 9. Schematic diagram of envelope-based tactual speech display.

A tactual display was designed for the presentation of envelope-asynchrony cues (EOA or EOFA) through the upgraded version of the Tactuator system. The amplitude envelope of each of the two filtered bands of speech (350 Hz lowpass and 3000 Hz highpass) was extracted in real time through the operations of filtering, rectification, and smoothing. Thresholds were established for the levels of each of the two smoothed amplitude envelopes to eliminate signals arising primarily from random noise fluctuations in the passband but yet sufficient for passing signals driven by the speech waveform. The specific scheme for the presentation of the envelope-asynchrony cues involved the delivery of the two modulated envelopes to two different fingers and employed a different modulation frequency at each finger (selections made to minimize cross-channel masking; e.g., see Verrillo et al., 1983; Tan, 1996). Specifically, the low-frequency envelope was used to modulate a 50-Hz sinewave at the left thumb and the high-frequency envelope was used to modulate a 250-Hz sinewave at the left index finger (see schematic in Fig. 9). The level of these speech-derived signals was generally in the range of 20 to 50 dB SL.

Perceptual Evaluations with Speech Signals

The efficacy of the EOA cue was evaluated through perceptual studies that included pairwise discrimination of eight pairs of voicing contrasts in the initial consonant of C_1VC_2 syllables, identification of a set of 16 initial consonants, and open-set recognition of sentences. In addition, the EOFA cue was evaluated through pairwise discrimination of eight pairs of voicing contrasts in the final consonant of C_1VC_2 syllables. These evaluations were generally concerned with examining benefits to lipreading through the addition of the tactual cue to voicing. Three conditions were studied in the discrimination and identification tests: lipreading alone (L), tactual signals alone (T), and the combined condition (L+T). For the sentence-reception task, only conditions L and L+T were examined (the T condition was not included because no significant information was expected for delivery of the tactual cue alone without extended training). These studies employed normal-hearing subjects who wore earplugs and headphones through which masking noise was presented to eliminate the possibility of any auditory cues arising from the vibration of the Tactuator.

Pairwise Discrimination of Voicing in Initial Consonants

The ability to discriminate eight pairs of initial voicing contrasts in C_1VC_2 syllables was examined using a 2I-2AFC procedure. Following training sessions which employed trial-by-trial correct-answer feedback, testing was conducted using a fresh set of speech tokens where each C_1 was represented by 24 tokens (2 talkers X 3 vowels X 4 utterances) selected at random on each trial. The sensitivity index d' averaged 0.09 for L (indicating chance-level performance) and was 2.4 for conditions of both T and L+T. Some inter-pair variability was observed on the T and L+T conditions, where scores ranged from d' of roughly 3.3 for the pair /◁–↗/ to roughly 1.6 for the pair /t◁-d↗/. Overall, performance on the T and L+T conditions averaged roughly 90% correct (compared to chance performance of 50% on L) and represents an improvement over that obtained with previous tactual devices (e.g., Reed et al., 1992; Waldstein and Boothroyd, 1995a, b). A rough correspondence was observed between subjects' resolution in the temporal-onset-order discrimination task and their performance on the speech-discrimination task. A comparison of the performance of human observers (HO) to that of an ideal observer with infinite resolution of temporal-onset-order resolution (IO) and limited temporal resolution of 34 msec (IO-Lim) is shown in Table C-1 below for each of the eight consonant pairs. The mean performance of the ideal observer is reduced by a factor of roughly 1.6 when a limitation is imposed on temporal resolution corresponding to the average human performance on the temporal onset-order discrimination task (Yuan et al., 2005a). The observed average human performance of d'=2.7 is lower than IO-Lim by a factor of roughly 1.6 indicating that in addition to limitations imposed by temporal resolution, other imperfections in human processing (e.g., related to memory and attention) lead to degradations in human performance.

TABLE 1. Comparison of performance of human observers (HO) to that of an ideal observer (IO) for eight pairs of consonants.

		Consonant Pair								
		/≺-⇕/	/↔-○/	/ -•/	/↔⊲-○↗/	/≻-⌟/	/ -ƛ/	/↩- /	/⊲-↗/	**Mean**
d′	IO	5.1	8.1	6.8	3.9	3.5	4.0	13.0	10.6	6.9
	IO-Lim	3.1	4.1	3.5	2.5	3.1	3.8	8.2	7.2	4.4
	HO	2.5	3.0	2.9	1.7	2.1	3.1	2.9	3.3	2.7

Identification of Initial Consonants

The tactual presentation of the EOA cue was highly effective for pair-wise discrimination of initial voicing contrasts. The purpose of this experiment was to examine the contribution of the tactual voicing cue to the task of consonant identification. The ability to identify the initial consonant of C_1VC_2 syllables was tested using a 1-interval, 16-alternative forced-choice procedure. Following training sessions which employed trial-by-trial correct-answer feedback, testing was conducted without feedback and with a fresh set of speech tokens (which included 24 tokens of each C_1). Overall performance averaged 34%, 12%, and 49%-correct for L, T, and L+T, respectively, demonstrating that the addition of the tactual cue improved lipreading ability by 15 percentage points. Performance was also examined in terms of percentage of unconditional IT for features of voicing, manner, and place (see Fig. 10). For the feature voicing, %-feature IT was ordered as L<L+T<T, representing an improvement of 30 percentage points for T over L alone and 10 percentage points for T over L+T (indicating that subjects may have had difficulty attending to the tactual cue in the presence of lipreading). For the features of manner and place, information was transmitted primarily through lipreading alone (with 0 %-IT through T alone) and was nearly identical for conditions of L and L+T. Predictions of overall performance under the combined condition L+T, computed using various models of bimodal integration (see Yuan, 2003; Yuan et al., 2005b), indicated that observed performance was inferior to that of model predictions. The non-optimal integration of the tactual cue with lipreading may arise in part from a lack of experience attending to the tactual modality.

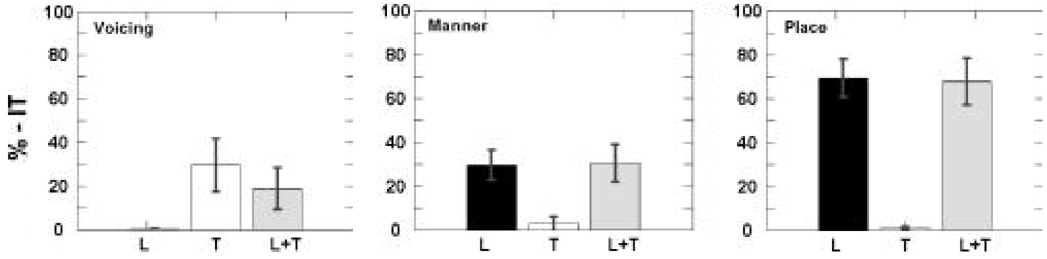

FIG. 10. Reception of the speech features of Voicing, Manner, and Place.

Pairwise Discrimination of Voicing in Final Consonants

The ability to discriminate eight pairs of voicing contrasts in the final position of C_1VC_2 syllables was examined using the same basic procedures described above for initial consonants (except that the syllables included 16, rather than 3, vowels). The sensitivity index d' averaged 1.2 for L, 2.9 for T, and 3.1 for L+T. Unlike the case of initial consonants, performance through L alone was greater than chance, presumably on the basis of vowel-duration cues to final voicing (Raphael, 1971). Performance through T was higher than through L, however, and provided a substantial supplement to lipreading. Compared to the predictions of an ideal observer operating on the EOFA measurements, the human performance (under T alone) was lower by a factor of roughly 1.4 (reflecting limitations at the peripheral or central processing levels for the human observers).

Connected-Speech Reception

Sentence testing was conducted to determine whether the benefits observed for the tactual cue at the segmental level would have immediate carry-over to the task of connected speech reception. The ability to recognize words in CUNY sentences (Boothroyd et al., 1985) was examined using recorded materials presented through conditions of L and L+T. Subjects received only a minimal amount of training which consisted of multiple repetitions (under the subject's control) of 36 sentences per condition and correct-answer feedback following each response. The testing itself employed only one presentation of each sentence (27 lists of 12 sentences each under L and L+T, respectively) and did not employ correct-answer feedback. The results indicated no improvement in performance over time and no difference in performance between the two conditions. Mean word-recognition scores in sentences ranged from 12 to 67% correct across subjects for L (averaging 32.7%) and 11 to 70% for L+T (averaging 32.0%). Segmental performance (which occupied a narrow range from 32-35% correct across subjects) does not appear to be a good predictor of sentence performance (as has been also observed by Bernstein et al., 2000). The lack of benefit for the tactual cue to lipreading in these tests may be due to a variety of factors, including the limited amount of training with the tactual cue, difficulty integrating tactual and visual cues, the potentially greater role of temporal masking of the tactual signals in continuous speech signals, and an increased variability of the temporal cue arising from additional complexity of continuous speech.

Tactual Displays of Environmental Sounds

Progress in this area includes the publication of two papers concerned with the reception of environmental sounds through tactual aids (Reed and Delhorne, 2003) and cochlear implants (Reed and Delhorne, 2005). The main focus of the research in this area is determining the interest of a broad class of deaf and hard-of-hearing (HOH) individuals (including those with oral and manual backgrounds) in the awareness and recognition of environmental sounds. This goal was accomplished through the development and administration of a survey to determine potential interest in tactual aids as a source of information about non-speech environmental sounds. The work includes development of a preliminary survey which was administered to 20 deaf and HOH adults (Delhorne et al., 2003). Based on analysis of the preliminary survey, a final questionnaire was developed for administration to several hundred deaf and HOH adults.

The revised survey (see Appendix B) consists of 52 questions in seven categories: personal information and linguistic history, hearing-aid use, cochlear-implant use, tactual-aid use, alerting-system use, interest in a variety of types of environmental sounds (other than speech), and suggestions for future devices. Respondents to the survey were recruited through advertising on websites and in the newsletters of a variety of organizations for the deaf and hard-of-hearing. The survey was made available for completion through a website which could be accessed by potential respondents or through pen and paper. The 258 respondents ranged in age from 16 to 82 years and were divided into two groups based on their cultural preference with the hearing community (Hearing Group, 131 respondents) or with the Deaf community (Deaf Group, 127 respondents). The two groups were similar in terms of level of education and employment. The individuals in the Deaf Group were more likely to have learned American Sign Language (ASL) as their first language, to make use of interpreters and transliterators throughout their school years, and to rely more on technical support than the respondents in the Hearing group. An equal number of respondents from both groups used hearing aids. A greater number from the Hearing Group used cochlear implants and a greater number from the Deaf Group used some type of alerting systems to receive information about environmental sounds (not including speech). The majority of people from both groups were "moderately" satisfied with the current technology for receiving information about environmental sounds, but felt a need for improvements in the technology. Both groups were most interested in receiving information about warning sounds and least interested in sounds from machines or vehicles. In terms of future devices, a greater number of respondents from the Deaf Group preferred one large multipurpose device whereas a greater number of people form the Hearing Group preferred several small devices that were more specific in the number of sounds that they could detect. Both groups of respondents indicated a high level of preference in the use of tactual displays (as opposed, for example, to flashing-light displays) for receiving information about environmental sounds.

Publications

Bratakos, M.S., Reed, C.M., Delhorne, L.A., and Denesvich, G. (2001). "A Single-Band Envelope Cue as a Supplement to Speechreading of Segmentals: A Comparison of Auditory versus Tactual Presentation. Ear and Hearing, 22(3), 225-235.

Israr, A., Meckl, P.H., and Tan, H.Z. (2004). "A Two DOF Controller for a Multi-Finger Tactual Display using a Loop-Sharing Technique," Proceedings of IMECE2004, 2004 ASME International Mechanical Engineering Congress and RD&D Expo, November 13-19, 2004, Anaheim, CA, pp. 1083-1089.

Israr, A., Meckl, P.H., and Tan, H.Z. (2006a). "A 2-DOF Controller to Preserve Tactual Sensitivity in a Multi-Channel Tactual Display," submitted to IEEE/ASME Transactions on Mechatronics.

Israr, A., Tan, H.Z., and Reed, C.M. (2006b). "Frequency and Amplitude Discrimination along the Kinesthetic-Cutaneous Continuum with Fixed and Roving Background," Journal of the Acoustical Society of America, undergoing revision.

Israr, A., Tan, H. Z., and Reed, C. M. (2006c). "Tactual frequency and amplitude discrimination with fixed and roving background," to appear in Proceedings of the International Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, Mar. 2006.

Parachuru, L. (2003). "Temporal-Order Discrimination through the Tactual Sense: Effects of Variations in Signal Amplitude and Duration," Undergraduate Final Paper, MIT, May 2003.

Reed, C.M., and Braida, L.D. (2003). "Frequency compression." In R.D. Kent (Ed.), MIT Encyclopedia of Communication Disorders. Cambridge, MA: MIT Press, pp. 471-475.

Reed, C.M., and Delhorne, L.A. (2003). "The Reception of Environmental Sounds through Wearable Tactual Aids," Ear and Hearing, 24, 528-538.

Reed, C.M., and Delhorne, L.A. (2005). "The Reception of Environmental Sounds through Cochlear Implants," Ear and Hearing, 26, 48-61.

Reed, C.M., and Delhorne, L.A. (2006). "A Study of the Combined Use of a Hearing Aid and a Tactual Aid in an Adult with Profound Hearing Impairment," submitted to The Volta Review.

Reed, C.M., Durlach, N.I., and Tan, H.Z. (2006). "Optimum Information-Transfer Rates in Human Communication," submitted to PRESENCE (undergoing revision).

Tan, H.Z., Reed, C.M., Delhorne, L.A., Durlach, N.I., and Wan, N. (2003). "Temporal Masking of Multidimensional Tactual Stimuli," Journal of the Acoustical Society of America, 114, 3295-3308.

Tan, H.Z., Srinivasan, M.A., Reed, C.M., and Durlach, N.I. (2006). "Discrimination and identification of finger joint-angle positions using active motion," ACM Transactions on Applied Perception, undergoing revision.

Yuan, H. (2003). Tactual Display of Consonant Voicing to Supplement Lipreading. Ph.D. Thesis, Department of Electrical Engineering and Computer Science, MIT, Cambridge, MA.

Yuan, H., Reed, C.M., and Durlach, N.I. (2004a). “Envelope-Onset Asynchrony as a Cue to Voicing in Initial English Consonants,” Journal of the Acoustical Society of America, 116, 3156-3167.

Yuan, H. Reed, C.M., and Durlach, N.I. (2005a). “Temporal Onset-Order Discrimination through the Tactual Sense,” Journal of the Acoustical Society of America, 117, 3139-3148.

Yuan, H. Reed, C.M., and Durlach, N.I. (2005b). “Tactual Display of Consonant Voicing to Supplement Lipreading,” Journal of the Acoustical Society of America, 118, 1003-1015.

Yuan, H., Reed, C.M.., and Durlach, N.I. (2006). “Temporal Onset-Order Discrimination through the Tactual Sense: Effects of Frequency and Site of Stimulation,” Journal of the Acoustical Society of America, in press for August.

Meeting Papers, Presented

Delhorne, L.A., Reed, C.M., McKee, B., and Fischer, S.D. (2003). “Preliminary Survey on Reception of Environmental Sounds: Opinions of Deaf and Hard-of-Hearing Persons,” Seventh International Sensory Aids Conference (ISAC-03), Portland, ME, May 8.

Reed, C.M., Delhorne, L.A., Brughera, A., Durlach, N.I., Tan, H.Z., and Wong, A. (2003). “Information-Transfer rates for Sequences of Multidimensional Tactual Signals,” Seventh International Sensory Aids Conference (ISAC-03), Portland, ME, May 8.

Reed, C.M., Durlach, N.I., and Tan, H.Z. (2006). “Optimizing Information-Transfer Rates for Human Communication,” Journal of the Acoustical Society of America, 119, 3237 (A).

Yuan, H., Reed, C.M., and Durlach, N.I. (2003). “Tactual Displays of Consonant Voicing to Supplement Speechreading,” Seventh International Sensory Aids Conference (ISAC-03), Portland, ME, May 8.
Yuan, H., Reed, C.M., and Durlach, N.I. (2003). “Tactual Displays of Consonant Voicing to Supplement Speechreading,” Journal of the Acoustical Society of America, 113, 2291 (A).

Yuan, H., Reed, C.M., and Durlach, N.I. (2004b). “Tactual Temporal-Onset Order Discrimination: Effects of Frequency and Site of Stimulation,” Journal of the Acoustical Society of America, 115, 2386 (A).

Yuan, H., Reed, C.M., and Durlach, N.I. (2004c). “Envelope Offset Asynchrony as a Cue to Voicing in Final English Consonants,” Journal of the Acoustical Society of America, 116, 2628 (A).

Yuan, H., and Reed, C.M. (2005). “A Tactual Display of Envelope Offset-Asynchrony as a Cue to Voicing of Consonants in the Final Position,” Eighth International Sensory Aids Conference (ISAC-05), Portland, ME, May 12.

References

Bernstein, L. E., Demorest, M. E., and Tucker, P. E. (2000). "Speech perception without hearing," Perception and Psychophysics 62, 233-252.

Bolanowski S. J., Gescheider, G. A., Verrillo, R. T., and Checkosky, C. M. (1988). "Four channels mediate the mechanical aspects of touch," Journal of the Acoustical Society of America 84, 1680-1694.

Boothroyd, A., Hnath-Chislm, T., and Hanin, L. (1985). “A sentence test of speech perception: Reliability, set-equivalence, and short-term learning,” City University of New York, Rep. No. RCI10.

Brughera, A. (2002). Upgrade of the Tactuator Computer, DSP, and Control System. Sensory Communication Group, Research Laboratory of Electronics, MIT, Cambridge, MA.

Craig, J. C. (1996). "Interference in identifying tactile patterns: Response competition and temporal integration," Somatosensory and Motor Research 13(3–4), 199–213.

Craig, J. C. and Baihua, X. (1990). "Temporal order and tactile patterns," Perception and Psychophysics 47, 22-34.

Craig, J.C. and Busey, T.A. (2003). "The effect of motion on tactile and visual temporal order judgments," Perception and Psychophysics 65, 81-94.

Craig, J. C., and Evans, P. M. (1987). "Vibrotactile masking and the persistence of tactual features," Perception and Psychophysics 42(4), 309-317.

Craig, J. C., & Evans, P. M. (1995). "Tactile selective attention and temporal masking," Perception and Psychophysics 57(4), 511-518.

Delhorne, L.A., Reed, C.M., McKee, B., and Fischer, S.D. (2003). "Preliminary Survey on Reception of Environmental Sounds: Opinions of Deaf and Hard-of-Hearing Persons," Seventh International Sensory Aids Conference (ISAC-03), Portland, ME, May 8.

Evans, P. M. (1987). "Vibrotactile Masking: Temporal Integration, Persistence, and Strengths of Representations," Perception and Psychophysics 42(6), 515-525.

Evans, P. M. and Craig, J. C. (1986). "Temporal integration and vibrotactile backward masking," Journal of Experimental Psychology: Human Perception and Performance 12(2), 160–168.

Evans, P. M. and Craig, J. C. (1992). "Response competition: A major source of interference in a tactile identification task," Perception and Psychophysics 51(2), 199–206.

Garner, W.R. (1962). Uncertainty and Structure as Psychological Concepts. Wiley, NY.

Gescheider, G.A., Valetutti, A.A., Padula, M.C., and Verrillo, R.T. (1992). "Vibrotactile forward masking as a function of age," Journal of the Acoustical Society of America 91, 1690-1696.

Hirsh, I. J. and Sherrick, C. E. (1961). ‘‘Perceived order in different sensory modalities,’’ Journal of Experimental Psychology 62, 423-432.

Israr, A., Meckl, P.H., and Tan, H.Z. (2004). “A Two DOF Controller for a Multi-Finger Tactual Display using a Loop-Sharing Technique,” Proceedings of IMECE2004, 2004 ASME International Mechanical Engineering Congress and RD&D Expo, November 13-19, 2004, Anaheim, CA, pp. 1083-1089.

Israr, A., Meckl, P.H., and Tan, H.Z. (2006a). “A 2-DOF Controller to Preserve Tactual Sensitivity in a Multi-Channel Tactual Display,” submitted to IEEE/ASME Transactions on Mechatronics.

Israr, A., Tan, H.Z., and Reed, C.M. (2006b). “Frequency and Amplitude Discrimination along the Kinesthetic-Cutaneous Continuum with Fixed and Roving Background,” Journal of the Acoustical Society of America, undergoing revision.

Klemmer, E.T. and Muller, P.F. (1953). The rate of handling information: Key pressing responses to light patterns. HFORL Memo Report No. 34, Human Factors Operations Research

Laboratories, Air Research and Development Command, Bolling Air Force Base, Washington, D.C., March 1953.

Mahar, D. P. and Mackenzie, B. D. (1993). "Masking, information integration, and tactile pattern perception: A comparison of the isolation and integration hypotheses," Perception 22, 483-496

Niyogi, P., and Ramesh, P. (2003). “The Voicing Veature for Stop Consonants: Recognition Experiments with Continuously Spoken Alphabets,” Speech Communication, 41, 349-367.

Parachuru, L. (2003). “Temporal-Order Discrimination through the Tactual Sense: Effects of Variations in Signal Amplitude and Duration,” Undergraduate Final Paper, MIT, May 2003.

Raphael, L. J. (1971). "Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English," Journal of the Acoustical Society of America 51, 1296-1303.

Reed, C.M., and Delhorne, L.A. (2003). “The Reception of Environmental Sounds through Wearable Tactual Aids,” Ear and Hearing, 24, 528-538.

Reed, C.M., and Delhorne, L.A. (2005). “The Reception of Environmental Sounds through Cochlear Implants,” Ear and Hearing, 26, 48-61.

Reed, C. M., Delhorne, L. A., and Durlach, N. I. (1992). “Results obtained with Tactaid II and Tactaid VII,” in A. Risberg, S. Felicetti, G. Plant, and K. –E. Spens (Eds.), *Proceedings of the Second International Conference on Tactile Aids, Hearing Aids and Cochlear Implants (pp. 149-155).* Stockholm, Sweden: Royal Institute of Technology.

Reed, C.M. and Durlach, N.I. (1998). "Note on information transfer rates in human communication," PRESENCE 7, 509-518.

Reed, C.M., Delhorne, L.A., Brughera, A., Durlach, N.I., Tan, H.Z., and Wong, A. (2003). "Information-transfer rates for sequences of multidimensional tactual signals," Seventh International Sensory Aids Conference (ISAC-03), Portland, ME, May 8.

Reed, C.M., Durlach, N.I., and Tan, H.Z. (2006). "Optimum information-transfer rates in human communication," submitted to PRESENCE.

Sherrick, C. E. (1970). ‘‘Temporal ordering of events in haptic space,’’ IEEE Trans. Man-Machine Syst. MMS-11, 25-28.

Shore, D. I., Spry, E., and Spence, C. (2002). ‘‘Confusing the mind by crossing the hands,’’ Brain research. Cognitive Brain Research 14, 153-16.

Tan, H. Z. (1996). Information transmission with a multi-finger tactual display, Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Tan, H. Z., and Rabinowitz, W. M. (1996). ‘‘A new multi-finger tactual display,’’ Proceedings of the Dynamic Systems and Control Division, DSC-Vol. 58, pp. 515-522.

Tan, H.Z., Durlach, N.I., Reed, C.M., and Rabinowitz, W.M. (1999). “Information Transmission with a Multifinger Tactual Display,” Perception and Psychophysics, 61, 993-1008.

Tan, H.Z., Reed, C.M., Delhorne, L.A., Durlach, N.I., and Wan, N. (2003). “Temporal Masking of Multidimensional Tactual Stimuli,” Journal of the Acoustical Society of America, 114, 3295-3308.

Taylor, B. (1978). Dimensional Redundancy in the processing of vibrotactile temporal order. Ph. D dissertation, Princeton University, Princeton, NJ.

Thomson, D.L., and Chengalvarayan, R. (2002). “Use of Voicing Features in HMM-Based Speech Recognition,” Speech Communication, 37, 197-211.

Verillo, R.T., Gescheider, B.A., Calman, B.G., and Van Doren, C.L. (1983). “Vibrotactile masking: Effects of one and two-site Stimulation,” Perception and Psychophysics, 33, 379-387.

Waldstein, R. S., and Boothroyd, A. (1995a). “Comparison of two multichannel tactile devices as supplements to speechreading in a postlingually deafened adult,” Ear Hear. **16**, 198-208.

Waldstein, R. S., and Boothroyd, A. (1995b). “Speechreading supplemented by single-channel and multichannel tactile displays of voice fundamental-frequency,” J. Speech Hear. Res. **38**, 690-705.

Yuan, H. (2003). Tactual Display of Consonant Voicing to Supplement Lipreading. Ph.D. Thesis, Department of Electrical Engineering and Computer Science, MIT, Cambridge, MA.

Yuan, H., Reed, C.M., and Durlach, N.I. (2004a). “Envelope-Onset Asynchrony as a Cue to Voicing in Initial English Consonants,” Journal of the Acoustical Society of America, 116, 3156-3167.

Yuan, H., Reed, C.M., and Durlach, N.I. (2004b). “Tactual Temporal-Onset Order Discrimination: Effects of Frequency and Site of Stimulation,” Journal of the Acoustical Society of America, 115, 2386 (A).

Yuan, H., Reed, C.M., and Durlach, N.I. (2004c). “Envelope Offset Asynchrony as a Cue to Voicing in Final English Consonants,” Journal of the Acoustical Society of America, 116, 2628 (A).

Yuan, H., Reed, C.M., and Durlach, N.I. (2005a). “Temporal Onset-Order Discrimination through the Tactual Sense,” Journal of the Acoustical Society of America, 117, 3139-3148,

Yuan, H., and Reed, C.M. (2005b). “A Tactual Display of Envelope Offset-Asynchrony as a Cue to Voicing of Consonants in the Final Position,” Eighth International Sensory Aids Conference (ISAC-05), Portland, ME, May 12, 2005.

Yuan, H., Reed, C.M., and Durlach, N.I. (2006). “Tactual Display of Consonant Voicing to Supplement Lipreading” Journal of the Acoustical Society of America, in press for August.