Controller design and consonantal contrast coding using a multi-finger tactual display[a)]

Ali Israr[b)]
Haptic Interface Research Laboratory, Purdue University, 465 Northwestern Avenue, West Lafayette, Indiana 47907-2035

Peter H. Meckl
Ruth and Joel Spira Laboratory for Electromechanical Systems, 585 Purdue Mall, West Lafayette, Indiana 47907-2088

Charlotte M. Reed
Research Laboratory of Electronics, Massachusetts Institute of Technology, Room 36-751, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139

Hong Z. Tan
Haptic Interface Research Laboratory, Purdue University, 465 Northwestern Avenue, West Lafayette, Indiana 47907-2035

(Received 8 August 2008; revised 24 March 2009; accepted 29 March 2009)

This paper presents the design and evaluation of a new controller for a multi-finger tactual display in speech communication. A two-degree-of-freedom controller consisting of a feedback controller and a prefilter and its application in a consonant contrasting experiment are presented. The feedback controller provides stable, fast, and robust response of the fingerpad interface and the prefilter shapes the frequency-response of the closed-loop system to match with the human detection-threshold function. The controller is subsequently used in a speech communication system that extracts spectral features from recorded speech signals and presents them as vibrational-motional waveforms to three digits on a receiver's left hand. Performance from a consonantal contrast test suggests that participants are able to identify tactual cues necessary for discriminating consonants in the initial position of consonant-vowel-consonant (CVC) segments. The average sensitivity indices for contrasting voicing, place, and manner features are 3.5, 2.7, and 3.4, respectively. The results show that the consonantal features can be successfully transmitted by utilizing a broad range of the kinesthetic-cutaneous sensory system. The present study also demonstrates the validity of designing controllers that take into account not only the electromechanical properties of the hardware, but the sensory characteristics of the human user.
 [DOI: 10.1121/1.3124771]

PACS number(s): 43.66.Wv, 43.66.Ts, 43.66.Gf, 43.60.Ek [ADP] Pages: 3925–3935

I. INTRODUCTION

The motivation for this research is to utilize touch as a sensory substitute for hearing in speech communication for individuals with severe hearing impairments. That such a goal is attainable is demonstrated by users of the Tadoma method who receive speech by placing a hand on the face of a speaker to monitor facial movements and airflow variations associated with speech production. Previous research has documented the speech-reception performance of highly experienced deaf-blind users of the Tadoma method at the segmental, word, and sentence levels (Reed *et al.*, 1985). An analysis of information-transfer (IT) rates for a variety of methods of human communication (Reed and Durlach, 1998) suggests that the communication rates achieved through Tadoma are roughly half of those achieved through normal auditory reception of spoken English. By comparison, the estimated communication rates for speech transmission through artificial tactile aids are substantially below those of the Tadoma method (Reed and Durlach, 1998). The limited success demonstrated thus far with artificial tactual communication systems may be due to a variety of factors, including (1) the homogeneous nature of displays that utilize single or multiple actuators to deliver only high-frequency cutaneous stimulation, and (2) the use of body sites with relatively sparse nerve innervation, such as forearm, abdomen, or neck (Plant, 1989; Waldstein and Boothroyd, 1995; Weisenberger *et al.*, 1989; Galvin *et al.*, 1999; Summers *et al.*, 2005). In contrast, Tadoma users have access to a rich set of stimulus attributes, including kinesthetic movements of the face and jaw, cutaneous vibrations at the neck, airflow at the lips, and muscle tensions in the face, jaw, and neck, which are received through the hands.

To more fully exploit the capabilities of the tactual sensory system that are engaged in the use of the Tadoma

[a)]Part of this work concerning the controller design was presented at the 2004 ASME International Mechanical Engineering Congress and Exposition, Anaheim, CA, Nov. 13-19, 2004.
[b)]Author to whom correspondence should be addressed. Electronic mail: israr@rice.edu

method, an artificial device, the Tactuator, was developed to deliver kinesthetic (motions) as well as cutaneous (vibrations) stimuli through the densely innervated fingertips of the left hand (Tan and Rabinowitz, 1996). Previous research has examined IT rates for multidimensional stimuli delivered through the Tactuator device (Tan *et al.*, 1999, 2003). For example, in Tan *et al.*, 2003, IT rates of up to 21.9 bits/s were achieved using multidimensional synthetic waveforms presented at a single contact site. These rates, which are among the highest reported to date for a touch-based display, are at the lower end of the range of IT rates obtained for auditory reception of speech (Reed and Durlach, 1998).

The present research was concerned with the utilization of the broad kinesthetic-to-cutaneous stimulation range (nearly 0–300 Hz) of the TactuatorII[1] for the display of speech. In particular, this research was designed to extend the work of Yuan (2003) in which speech was encoded for display through the Tactuator device. Yuan (2003) examined the ability to discriminate the voicing cue in consonants using a two-channel speech-coding scheme in which the amplitude envelope of a low-frequency band of speech was used to modulate a 50-Hz waveform delivered to the thumb, and the amplitude envelope of a high-frequency band of speech was used to modulate a 250-Hz waveform at the index finger. Noise-masked normal-hearing participants achieved high levels of performance on the pairwise discrimination of consonants contrasting the feature of voicing through the tactual display alone. This coding scheme was also effective in providing a substantial benefit to lipreading in closed-set consonant identification tasks.

Encouraged by the results of Yuan (2003), the present study investigated consonant discriminability for the features of place and manner of articulation in addition to voicing. A speech-to-touch coding scheme was developed to extract envelope information from three major spectral regions of the speech signal and present them as kinesthetic motional and cutaneous vibrational cues. The three spectral bands included a low-frequency region (intended to convey information about fundamental frequency), a mid-frequency region (intended to convey information about the first formant of speech), and a high-frequency region (intended to convey second-formant information). These bands were somewhat consistent with the assessment of bands of modulated noise required for speech recognition by Shannon *et al.* (1995), as well as those used in previous studies on tactile aids (Weisenberger and Percy, 1995; Clements *et al.*, 1988; Summers, 1992). Amplitude-envelope information from each of these spectral regions was encoded tactually through the use of mid- and high-frequency vibrations at one of the three contactor sites of the TactuatorII (thumb, middle finger, and index finger, respectively). The absolute amplitude of the vibrations at each finger provided information about the energy in the corresponding frequency band. The relative amplitudes of the two vibrations (modulated at 30 and 200 Hz) at each finger channel provided information about energy spread in the corresponding frequency band. In addition to the tactile waveforms, the coding scheme monitored energy peaks within each band and presented this information as low-frequency motional cues—extending the finger for high-frequency contents and flexing the finger for low-frequency contents in the corresponding finger band. These more pronounced representations of formant and formant transition cues were employed in an effort to improve the transmission of cues related to place of articulation, which have been poorly transmitted through previous tactile aids (Clements *et al.*, 1988; Weisenberger *et al.*, 1989; Waldstein and Boothroyd, 1995; Plant, 1989; Summers *et al.*, 2005; Weisenberger and Percy, 1995; Galvin *et al.*, 1999). Acoustical analyses of plosive and fricative consonants have shown that place of articulation is well correlated with the frequency values of the first two formants (spectral peaks in the speech spectrum due to the shape of the mouth), F1 and F2, and their transitions (Ali *et al.*, 2001a, 2001b; Jongman *et al.*, 2000). Therefore, motions indicating changes in F1 and F2 were used to encode information concerning place of articulation. Although the location of the energy peaks was presented as high-frequency vibrations, redundant presentations of the same information as quasi-static positions of the fingers were intended to reduce inter-channel effects that may arise, such as those due to masking. It is well known that masking reduces internal representations of proximal tactual stimuli (Craig and Evans, 1987; Evans, 1987; Tan *et al.*, 2003) and redundant presentation of speech information can lead to improved perceptual performance (Yuan, 2003; Summers *et al.*, 1994).

One challenge associated with the use of broadband signals with an electromechanical system such as the TactuatorII is that the system frequency response is not uniform across its operating range. Therefore, the input signals are distorted spectrally before they are presented to a human user. To solve this problem, a closed-loop two-degree-of-freedom (2DOF) controller was developed to reshape the overall system response. Specifically, the controller compensated for both the frequency response of the TactuatorII and the frequency-dependent human detection-thresholds (HDTs) for tactual stimulation so that when a broadband input signal is applied to the TactuatorII, the relative amplitude of spectral components in the input signal is preserved in terms of perceived intensity when the signal reaches the user's fingers. The 2DOF controller consists of a feedback controller and a prefilter. The feedback controller (referred to as the low-frequency kinesthetic or motion controller) counters the effects of low-frequency disturbances due to a user's finger loading the device, increases the closed-loop bandwidth, and reduces the high-frequency in-line noise. The prefilter (referred to as the broadband cutaneous or vibration controller) shapes the overall system frequency response so that two equal-amplitude spectral components at the reference input would be perceived as equally intense by the human user.

The remainder of this paper describes the controller design and implementation of the TactuatorII system (Sec. II) and the speech-to-touch coding scheme (Sec. III). An experimental study on the pairwise discrimination of consonants with two human observers is reported (Sec. IV) before the paper concludes with a general discussion (Sec. V).

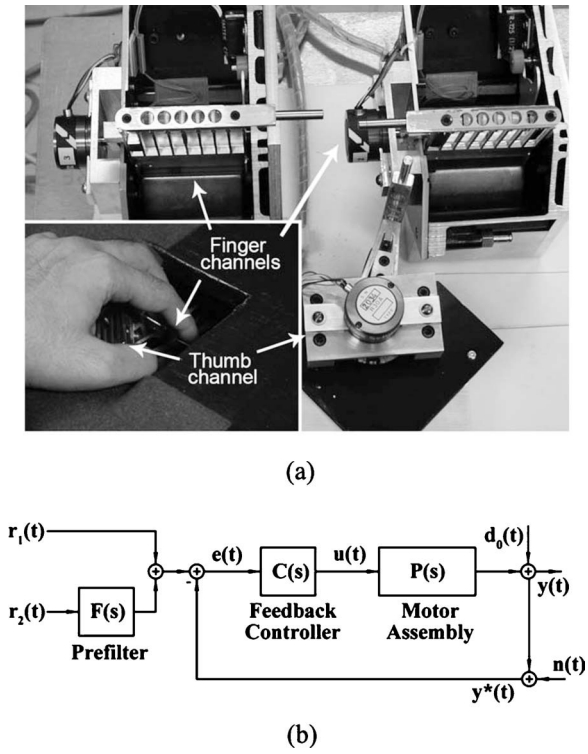

FIG. 1. (a) Three channels of the TactuatorII system and the three hand contact points rested lightly on the "fingerpad interface" rods (inset). (b) The block diagram representation of the 2DOF controller.

II. CONTROLLER DESIGN

A. Apparatus

The TactuatorII consists of three independently-controlled channels interfaced with the fingerpads of the thumb, the index finger, and the middle finger, respectively [Fig. 1(a)]. The range of motion for each digit is about 25 mm. Each channel has a *continuous* frequency response from dc to 300 Hz, delivering stimuli from the kinesthetic range (i.e., low-frequency gross motion) to cutaneous range (i.e., high-frequency vibration) as well as in the mid-frequency range. Across the frequency range of dc to 300 Hz, an amplitude of 0 dB sensation level (SL) (decibels above HDT) to at least 47 dB SL can be achieved at each frequency, thereby matching the dynamic range of tactual perception (Verrillo and Gescheider, 1992). Details of the TactuatorII can be found in Israr *et al.*, 2006 (cf. Sec. II A, on pp. 2790–2791) and Tan and Rabinowitz, 1996.

The frequency response of the motor assembly was obtained by measuring the input-output voltage ratio over the frequency range dc to 300 Hz. It was modeled by a second-order transfer function $P(s)=2875/(s^2+94s+290)$ (see also Tan and Rabinowitz, 1996).

B. Controller design

The main design objective was to shape the frequency response of the TactuatorII so that when driven with a broadband signal (up to 300 Hz), the relative intensities of different spectral components were preserved in terms of the relative sensation levels (SLs) delivered by the TactuatorII. In addition, the controller should be able to reduce the effects of low-frequency finger load disturbance, and achieve fast and stable motion tracking. Because of the similarities among all three channels, controller design for one channel assembly of the TactuatorII is discussed in this paper.

Our approach was to have two main components in the controller: one for the low-frequency kinesthetic movements, and the other for the broadband high-frequency cutaneous region, as explained in Fig. 1(b). The high-frequency broadband reference position signal, $r_2(t)$, was first passed through a prefilter, $F(s)$, and then added to the low-frequency motional reference position, $r_1(t)$. The combined signal was then compared to the measured position signal, $y^*(t)$, to form an error signal, $e(t)$, as the input to the feedback controller, $C(s)$. The output of the feedback controller or the command signal, $u(t)$, was used to drive the motor assembly, $P(s)$, to achieve a position trajectory of $y(t)$ at the point where the fingerpad rests. The effects of finger loading and sensor noise are represented by $d_0(t)$ and $n(t)$, respectively.

Major steps in the design of the feedback controller and the prefilter are outlined below. More details can be found in Israr, 2007 (cf. Chap. 2).

1. Feedback controller for kinesthetic stimulus region

The feedback controller or the motional (kinesthetic) controller $C(s)$ was designed using a lead-lag frequency loop-shaping technique that shaped the frequency response of the open-loop transfer function, $L(s)=C(s)P(s)$, to lie within the constraints determined by the required closed-loop response, $T(s)=C(s)P(s)/[1+C(s)P(s)]$ (Maciejowski, 1989). It consists of an integrator for maintaining the 0 dB closed-loop gain, a pair of zeros for increasing the stability margin, and a high-frequency pole for suppressing inline noise and for the proper structure (causality) of the controller $C(s)$. The final design of the feedback (kinesthetic) controller is given by

$$C(s) = 12.264\frac{s^2+111s+530}{s(s+260)}.$$

Figure 2 shows the magnitude [panel (a)] and the phase [panel (b)] of the frequency response for the open-loop system (dashed-dotted curve) and the closed-loop system (solid curve). The stability gain and phase margins achieved with the controller $C(s)$ are also shown in Fig. 2. A quantitative analysis of the system showed that the feedback controller was able to reject or reduce unwanted noise. The 60-Hz inline noise was imperceptible by human users due to rapidly falling slope of the closed-loop magnitude frequency response at 60 Hz. The finger load was rejected by keeping the closed-loop response close to the 0 dB line at low frequencies, and by selecting an appropriate bandwidth of about 30 Hz. In the loaded conditions (where the fingerpad was lightly placed on the fingerpad interface), the average deviations of the closed-loop response were 0.34 dB at 1 Hz, 1.43 dB at 8 Hz, 0.64 dB at 40, 0.3 dB at 100, and 0.65 dB at 260 Hz from the unloaded conditions (where the fingerpad interface was displaced with no finger load), measured at four intensity levels.

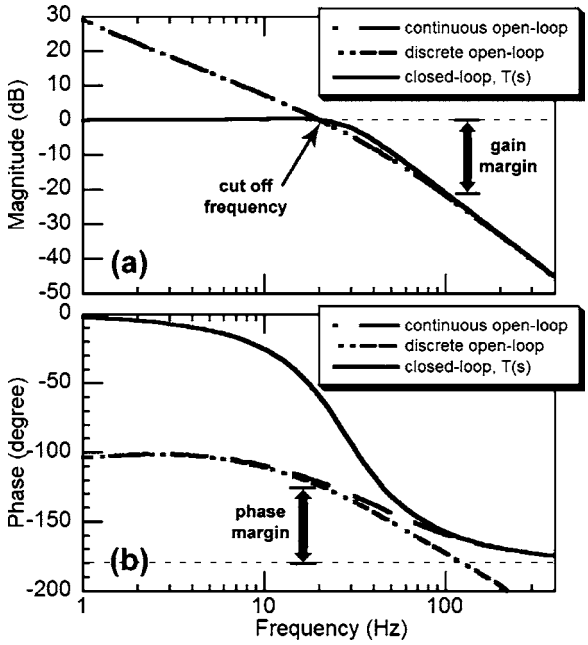

FIG. 2. Comparison of the frequency response of open-loop and closed-loop systems. (a) shows the magnitude response of the open-loop and closed-loop systems. A solid curve shows the input-output transfer function model of the closed-loop TactuatorII assembly. The response of the continuous and discrete open-loop responses overlaps. (b) shows phase response of the open-loop and closed-loop systems. Also shown in the figure are gain and phase margins, which are important criteria for system stability.

2. Prefilter controller for cutaneous stimulus region

For the design of the broadband controller component, i.e., the prefilter $F(s)$, we first considered the typical HDT curve as a function of sinusoidal stimulus frequencies[2] (Bolanowski *et al.*, 1988) (shown as solid curve in Fig. 3). The inverse of this detection-threshold curve was regarded as the sensitivity curve, or equivalently, the "frequency response" of the human user. The *perceived intensity* of a signal, in dB SL, is roughly determined by the distance between the physical intensity of the signal and the detection-threshold at the corresponding frequency (Verrillo and Gescheider, 1992). The effect of the human sensitivity curve on system performance is illustrated in Fig. 4. When a broadband cutaneous

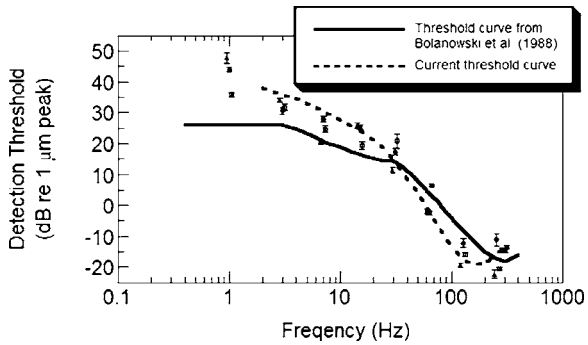

FIG. 3. A typical HDT curve as a function of frequency adapted from Bolanowski *et al.* (1988) (solid curve) and a HDT curve obtained in the present study (dashed curve). Also shown here are data points from three participants (S1—○, S2—□, and S3—△) and the standard errors of their threshold levels. The dashed curve is a first order approximation of the detection-threshold levels for three participants along the frequency continuum.

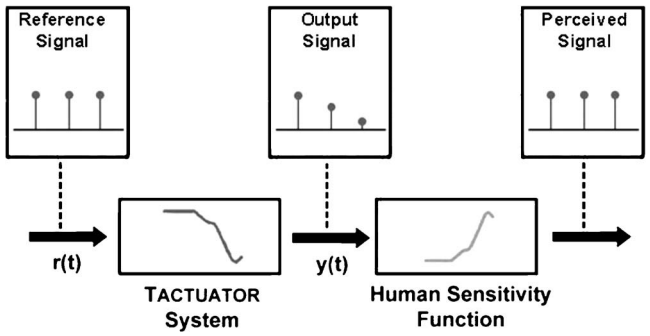

FIG. 4. Graphical illustration of the objectives for the high-frequency cutaneous controller. When the frequency-response function of the mechanical system matches with that of the HDT, the frequency function cancels the effects of variable human sensitivity function and preserves spectral components of the reference input signal.

controller is used to compensate for the human sensitivity function, equal intensities in the input signal (shown as equal intensities in the reference signal, Fig. 4) spectrum will result in equally strong sensations when received by a human user. Therefore, the steady-state response of the overall closed-loop system, $H(s)=F(s)T(s)$, should follow the target frequency function of the HDT curve in the frequency range dc to 300 Hz, i.e., $H(s)=\mathrm{HDT}(s)$.

It was anticipated that the HDT function with the fingerpad interface of the TactuatorII system would differ from that reported in Bolanowski *et al.*, 1988 based on the known

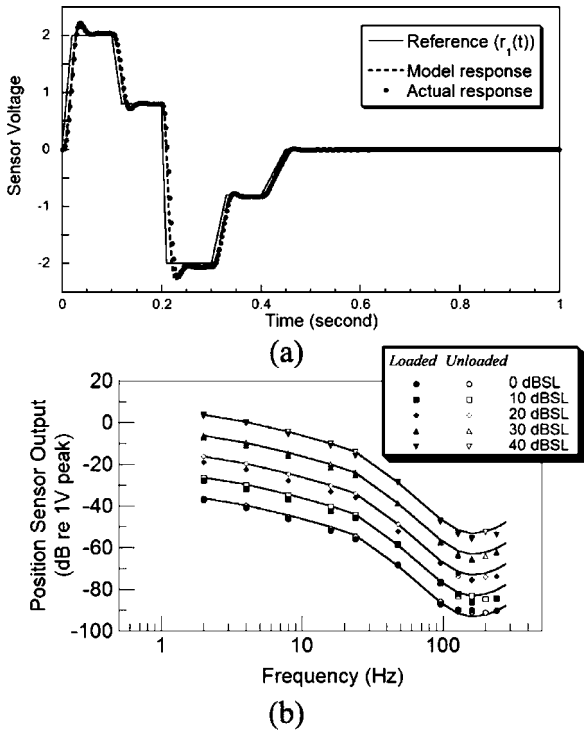

FIG. 5. (a) Response of TactuatorII for ramp signal applied at the reference input, $r_1(t)$, shown as solid line. The response of the model (dashed line) and actual mechanical assembly (dots) showed fast response time and low overshoot. (b) A comparison of the measured sensor outputs (individual data points) and the predicted output levels (solid lines) at 0, 10, 20, 30, and 40 dB SLs without the influence of human finger loading (unloaded condition shown as unfilled symbols) and with the influence of human finger loading (loaded condition as filled symbols).

variation in tactual thresholds with experimental conditions such as contact site, direction of vibrations, use of an annulus surround to restrict penetration of vibrations, etc. (Verrillo and Gescheider, 1992; Brisben *et al.*, 1999). Thus, the detection-thresholds for three highly trained participants were estimated in a psychophysical experiment. Detection-thresholds for 1-s stimulus at nine test frequencies (1, 3, 7, 15, 31, 63, 127, 255, and 300 Hz) were determined with a three-interval forced-choice paradigm combined with a one-up three-down adaptive procedure (Leek, 2001). Thresholds obtained this way corresponded to the 79.4-percentile point on the psychometric function. The results are shown in Fig. 3. Compared with the HDT curve determined by Bolanowski *et al.* (1988), which were measured at the thenar eminence, the newly measured thresholds measured on the index fingerpad followed the same general trend; however, our absolute-threshold measurements were somewhat higher than those of Bolanowski *et al.* (1988) at the lower frequencies and lower than theirs at the higher frequencies. These results were consistent with those found in other studies (Gescheider *et al.*, 1978; Van Doren, 1990; Goble *et al.*, 1996), and those taken earlier with the Tactuator using a Proportional-Integral-Derivative (PID) controller (Tan and Rabinowitz, 1996; Yuan, 2003).

The TactuatorII-specific HDTs were subsequently incorporated into the parameters of the prefilter controller, $F(s)$. A new HDT curve based on the measured data (dashed line in Fig. 3) was obtained and used as the required frequency function $H(s)=F(s)T(s)$. The Laplace transform of the resulting prefilter is

$$F(s) = 0.51\frac{s^4 + 1797s^3 + 1.822 \times 10^6 s^2 + 9.779 \times 10^8 s + 1.955 \times 10^{11}}{s^4 + 1134s^3 + 4.313 \times 10^6 s^2 + 1.337 \times 10^9 s + 1.995 \times 10^{10}}.$$

C. Controller response analysis

The 2DOF controller was implemented on a SBC6711 standalone DSP card (Innovative Integration, Simi Valley, CA) with a 16-bit Analog-to-Digital Converter (ADC) and a 16-bit Digital-to-Analog Converter (DAC) at a sampling rate of 4 kHz. The 2DOF controller design was analyzed by measuring closed-loop reference tracking and overall closed-loop frequency response in unloaded and loaded conditions. In order to readily compare the sensor feedback signal in volts with the threshold levels, the controller input reference was scaled by a factor of 0.003 97. This factor accommodated the magnitude level of the flat portion of the HDT function at lower frequencies (26 dB with regard to 1 μm peak or -34 dB with regard to 1 mm peak in Fig. 3, or equivalently 0.019 95) and the sensor gain of 0.198 98 V/mm.

1. Motion tracking

Figure 5(a) shows the response of the TactuatorII system for ramp trajectories applied at the reference input $r_1(t)$ without the influence of human finger load. Shown are the responses of the model (dashed line) and the actual hardware assembly (dots). The slopes of the reference trajectories were 0.1, -0.06, -0.28, 0.04, and 0.016 V/ms, respectively. The output (position) response of the hardware assembly showed that the low-frequency kinesthetic controller maintained stability of the device, and tracked the reference input with low response time (about 10 ms) and with a small response overshoot.

2. Frequency response

Sinusoidal reference input signals of 2-s duration at various frequencies, $r_2(t)$, were applied to the TactuatorII system, and the position-sensor readings were recorded. The results for unloaded (without finger load) and loaded (with finger placed lightly on the fingerpad interface) conditions are shown in Fig. 5(b). The bottom solid curve corresponds to the HDT curve measured with the TactuatorII (dashed line in Fig. 3), i.e., the 0 dB SL curve. The other four solid curves are at 10, 20, 30, and 40 dB SLs, respectively. The open symbols show the measured outputs at the five SLs with no finger load and the filled symbols show loaded results with a finger resting lightly on the fingerpad interface. There was generally a close match between the measured data points (filled and unfilled symbols) and the expected output levels (solid curves). Deviations at a few frequencies (especially at the highest frequencies) were likely due to signal noise and non-linear finger loading effects at such a low signal level. Therefore, the 2DOF controllers were successful at compensating for the frequency response of the motor assembly and the HDT curve, and the feedback controller was effective at rejecting the low-frequency disturbances caused by the finger load.

The engineering performance measurements presented above indicate that the 2DOF controller met the original design objectives in accurate and fast motion tracking, disturbance rejection, and broadband response shaping. Most importantly, we demonstrated the achievement of the main design objective of preserving the relative intensities of input signal spectral components in terms of dB SLs.

III. SPEECH-TO-TOUCH CODING

Speech features were extracted off-line in MATLAB (The MathWorks, Inc., Natick, MA) from the digitized speech segments and were converted into tactual signals presented through all three channels of the TactuatorII. Before the processing, the speech signal was passed through a pre-emphasis filter that amplified the energy above 1000 Hz at a

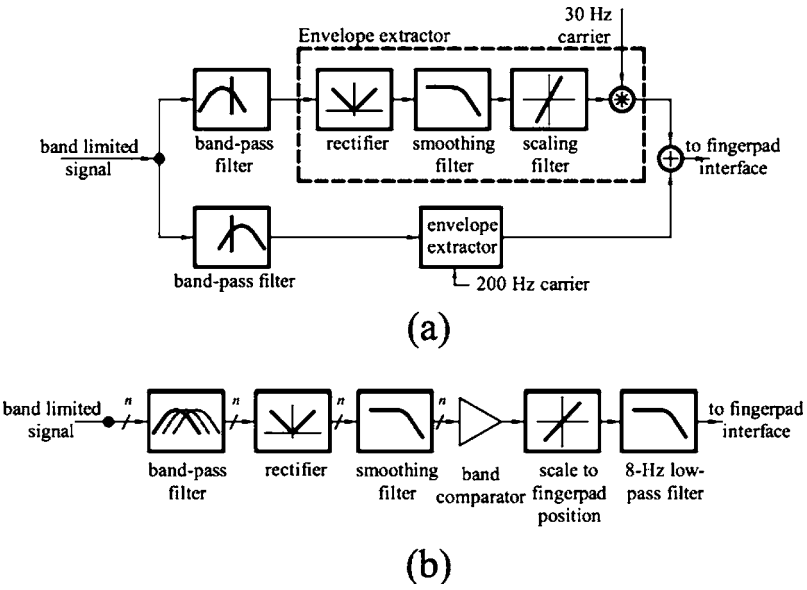

FIG. 6. (a) Block diagram illustration of tactile coding scheme used in the formant bands. (b) Block diagram illustration of motional coding scheme. Spectral features are extracted from three bands of speech signal and presented as motional waveforms.

typical rate of 6 dB/octave in order to compensate for the falling speech spectrum above 1000 Hz. Three major signal processing schemes were used for the extraction of spectral features: (1) low-pass filtering, (2) band-pass filtering, and (3) envelope extraction scheme of Grant *et al.*, (1985). In this scheme, the band-limited signal is rectified and passed through a low-pass filter to extract its temporal envelope, which is then scaled and output with a carrier frequency, as shown in Fig. 6(a). The coding scheme incorporates both high-frequency tactile vibrations and low-frequency motional waveforms.

A. Speech material

The speech materials consisted of C_1VC_2 nonsense syllables spoken by two female speakers of American descent. Each speaker produced eight tokens of each of 16 English consonants (the plosives, fricatives, and affricates: /p, t, k, b, d, g, f, θ, s, ʃ, v, ð, z, ʒ, tʃ, dʒ/) at the initial consonant (C_1) location with medial V = /ɑ/. The final consonant (C_2) was randomly selected from a set of 21 consonants (/p, t, k, b, d, g, f, θ, s, ʃ, v, ð, z, ʒ, tʃ, dʒ, m, n, ŋ, l, r/). The syllables were converted into digital segments and stored as a .mov (QuickTime Movie) file on the hard drive of a desktop computer (see details of conversion in Yuan, 2003). The .mov files were then converted into .wav (waveform audio) files by using CONVERTMOVIE 3.1 (MOVAVI, Novosibirsk, Russia) and with audio format set at a sampling rate of 11 025 Hz and 16-bit mono. The duration of the segments varied from 1.268 to 2.002 s with a mean duration of 1.653 s.

B. Tactile coding scheme

The coding scheme extracted envelopes from three distinct frequency bands (F0-, F1-, and F2-bands) of the speech spectrum and presented them as vibrations (mid- and high-frequency waveforms) through the three channels of the TactuatorII. Table I lists the numerical values for the frequency bands and corresponding finger channels. Spectral energy from the fundamental frequency (F0) region was presented at the thumb channel by passing the low-pass filtered speech signal directly through the 2DOF controller described in Sec. II. Information from the first-formant band (F1) was presented through the middle finger channel and the second-formant band (F2) information through the index finger channel, using processing units described in Fig. 6(a). The formant band-limited signal was processed through two band-pass filters, and amplitude envelopes of these two bands were extracted and modulated with carrier frequencies of 30 and 200 Hz. The 30-Hz waveforms modulated the envelope of the lower-frequency band and the 200-Hz waveforms modulated the higher-frequency band. The two vibratory signals were added and passed through the fingerpad interface. Since the digitized speech segments were normalized to one, the vibrations were scaled to a maximum intensity of 40 dB SL.

Figure 7 illustrates the vibration cues associated with two CVC segments spoken by two female speakers. The top two panels show the 30- and 200-Hz vibrations for segment /bɑC_2/ spoken by speaker 1 and the bottom two panels show the same by speaker 2. Note the similar waveforms at the two fingerpads associated with the same medial vowel /ɑ/. The vowel /ɑ/ has a high first formant and a low second formant. This is indicated by stronger 200-Hz vibrations than the 30-Hz vibrations at the middle fingerpad (see the two left panels in Fig. 7) and significantly stronger 30-Hz vibrations at the index fingerpad (see the two right panels). Cues associated with similar initial consonants are difficult to judge

TABLE I. Speech bands and the corresponding vibrations through the three channels.

TactuatorII channel	Speech bands (Hz)	Envelope bands (Hz)	Carrier frequency (Hz)
Middle finger	F1-band (300–1200)	300–650	30
		650–1200	200
Index finger	F2-band (1150–4000)	1150–1750	30
		1750–4000	200
Thumb	F0-band (80–270)	Low-pass filtered at 270 Hz	

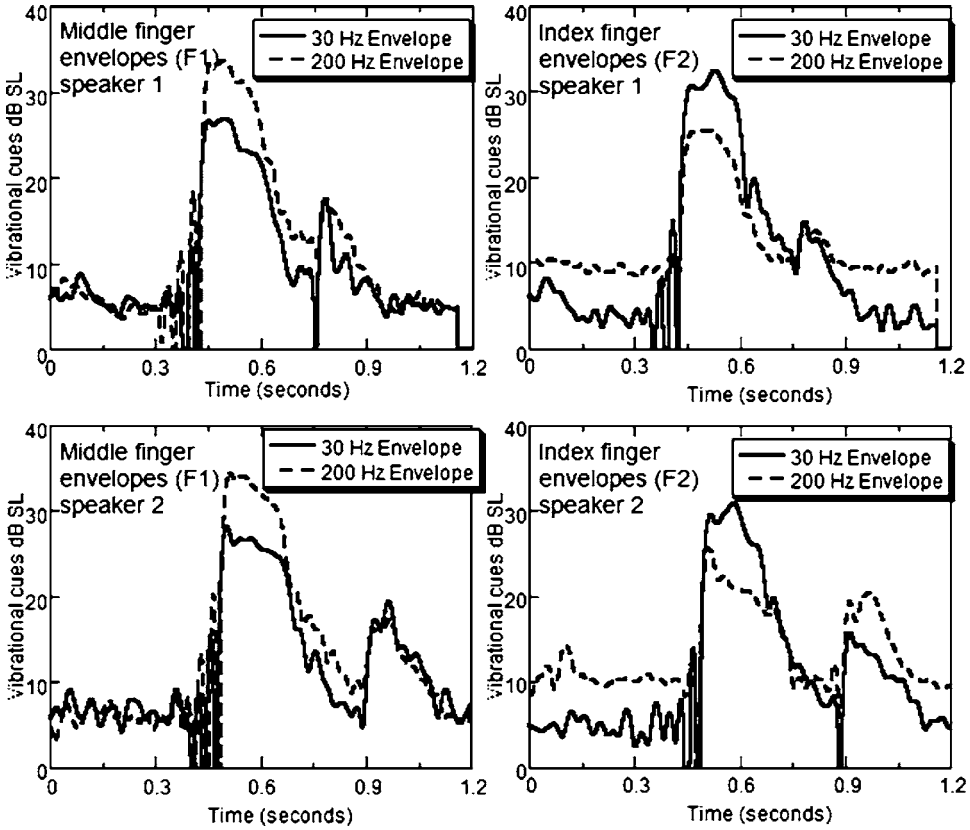

FIG. 7. Illustration of vibration waveforms extracted by using the speech-to-touch coding scheme. The figure shows vibration cues presented on the middle fingerpad (left panel) and on the index fingerpad (right panel). The cues are associated with multiple segments of /baC$_2$/ spoken by two female speakers (Sp1 or Sp2).

because they are not resolved for a small duration of time (either visually or through the tactual sensory system).

C. Motional coding scheme

The coding scheme extracted frequency variations in the F0-, F1-, and F2-bands using processing blocks shown in Fig. 6(b) and presented them through three channels of the TactuatorII. These motion cues indicated variations of spectral energy such as formant transition cues in the consonant-vowel segments and the quasi-static positions of the fingerpad interface redundantly indicated the frequency locations of energy peaks in the frequency band of each channel. As illustrated in Fig. 6(b), the formant band-limited signal was passed through contiguous band-pass filters in parallel and the temporal envelope of each band was obtained. The envelopes were compared and the center frequency of the band with the largest envelope value was noted at each sample instant. The center frequency was linearly mapped to the absolute reference position of the fingerpad interface that ranged ± 12.5 mm from the neutral zero position and was low-pass filtered with a gain crossover at 8 Hz. Thus, the finger extended for high-frequency contents and flexed for the low-frequency contents in the finger band. As with the tactile coding scheme, the features from the F0-, F1-, and F2-bands were presented to the thumb, middle finger, and index finger channels, respectively. The center frequencies and bands of each band-pass filters are shown in Table II. The frequency ranges covered by the middle finger and thumb channels were divided into eight bands, while the larger range encompassed by the index finger channel was divided into ten bands. Illustration of the motion cues associated with two segments of six initial consonants in CVC format spoken by the two speakers is shown in Fig. 8.

IV. PRELIMINARY STUDY OF CONSONANT DISCRIMINATION

A perception study was conducted on the pairwise discriminability of consonants that were processed for display through the three finger-interfaces of the TactuatorII system.

A. Methods

The ability to discriminate consonantal features was tested for 20 pairs of initial consonants that contrasted in voicing, place, and manner features. Each pair contrasted one

TABLE II. Frequency bands for motional cues.

	Frequency band (Hz)		
Filter index	Middle finger	Index finger	Thumb
1	300–400	1150–1300	80–100
2	400–500	1300–1500	100–120
3	500–600	1500–1700	120–140
4	600–700	1700–1900	140–160
5	700–800	1900–2100	170–200
6	800–900	2100–2300	200–220
7	900–1000	2300–2500	220–240
8	1000–1200	2500–3000	240–260
9	N/A	3000–4000	N/A
10	N/A	4000 – 5000	N/A

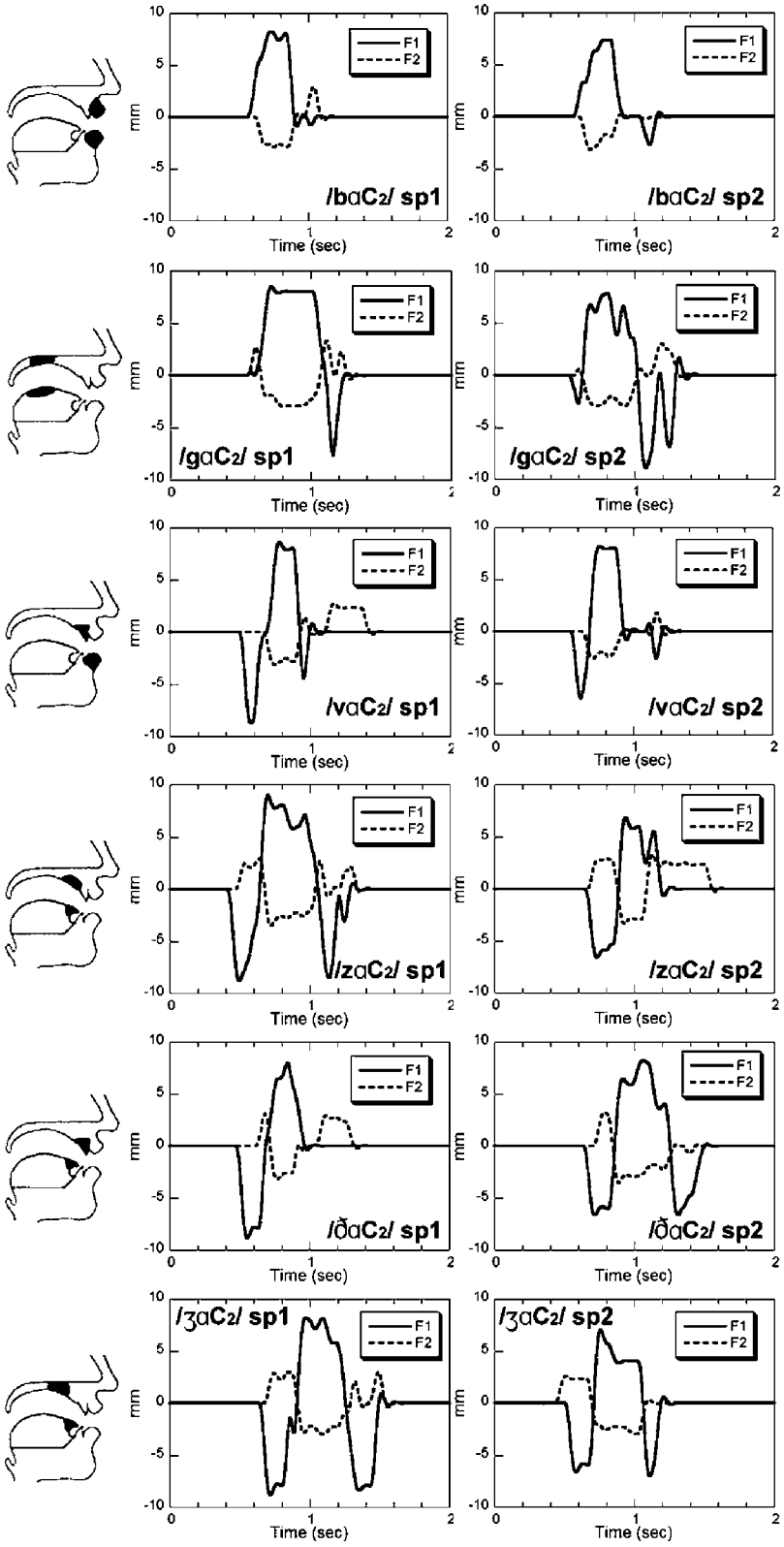

FIG. 8. Illustration of motion waveforms extracted by using the speech-to-touch coding scheme. Each row shows the waveforms associated with two segments of the same initial consonant spoken by two female speakers (sp1 or sp2). The waveforms correspond to the formant location and formant transitions in the first-formant band (solid line, motion waveforms at the middle finger) and in the second-formant band (dashed line, motion waveforms at the index finger). Also shown are the locations of constriction during the production of the initial consonant.

of the three features (and had the same value for each of the other two features). The pairs used in the present study along with their contrasting features are shown in Table III. Out of the 20 pairs, 5 pairs contrasted in voicing, 8 pairs contrasted in place, and 7 pairs contrasted in manner. Two male participants (ages 30 and 22 years old) took part in the experiments. S1, who is one of the authors, was highly experienced with the TactuatorII system, but S2 had not used the device prior to the present study. Both S1 and S2 were research staff members with previous experience in other types of haptic experiments.

The tests were conducted using a two-interval two-alternative forced-choice paradigm (Macmillan and Creelman, 2004). On each trial, the participant was presented with two tactual stimuli associated with a specific pair of consonants. The order of the two consonants was randomized with

TABLE III. Contrasting consonant pairs, associated articulatory and phonetic features, and average evaluation scores in C3.

Pairs	Articulatory features	Contrasting distinction	d'
/p-b/	Bilabial plosives	Voicing	4.65
/k-g/	Velar plosives	Voicing	2.90
/f-v/	Labiodental fricatives	Voicing	2.36
/s-z/	Alveolar fricatives	Voicing	3.78
/tʃ-dʒ	Affricates	Voicing	3.81
/p-t/	Unvoiced plosives-bilabial/alveolar	Place	2.38
/t-k/	Unvoiced plosives-alveolar/velar	Place	1.66
/b-d/	Voiced plosives-bilabial/alveolar	Place	3.80
/d-g/	Voiced plosives-alveolar/velar	Place	2.13
/f-s/	Unvoiced fricatives labiodental/alveolar	Place	1.46
/v-z/	Voiced fricatives labiodental/alveolar	Place	3.80
/θ-ʃ/	Unvoiced fricatives dental/post-alveolar	Place	2.66
/ð-ʒ/	Voiced fricatives dental/post-alveolar	Place	3.60
/p-f/	Unvoiced bilabial plosives/labiodental fricative	Manner	3.12
/b-ð/	Voiced bilabial plosives/dental fricative	Manner	3.47
/t-s/	Unvoiced alveolar plosive/fricative	Manner	3.60
/d-ʒ/	Voiced alveolar plosive/post-alveolar fricative	Manner	4.65
/d-dʒ/	Voiced alveolar plosive/affricate	Manner	2.37
/s-tʃ/	Unvoiced alveolar fricative/affricate	Manner	3.00
/ʃ-tʃ/	Unvoiced post-alveolar fricative/affricate	Manner	3.38

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 s with an inter-stimulus-interval of 500 ms. A 150-ms auditory tone and a visual phrase indicating "stimulus 1" or "stimulus 2" were presented 250 ms before the start of each stimulus to mark the beginning of each stimulus interval.

Data were collected for each consonant pair under three different experimental conditions tested in a single session: A 20-trial initial run without any feedback (C1), up to four 20-trial runs with trial-by-trial correct-answer feedback (C2), and a 50-trial final run without feedback (C3). Condition C2 was terminated if a percent-correct score above 90% was obtained in a single run or when the participant had completed all four runs. Conditions C1 and C3 could be viewed as the initial and final assessments of the participants' performance, while C2 provided training as needed (although one could argue that S1 was already "trained" prior to C1). Half of the 256 total speech tokens were used in conditions C1 and C3 and the other half were used in condition C2. Thus, the two sounds associated with each discrimination test were represented by eight tokens apiece (four from each of the two speakers). Each consonant within a pair was presented once or twice to the participant before C1 to familiarize the participant with its tactual cues. The order in which consonant pairs were tested was randomized for each participant. Each participant was tested for no more than two 40–45 min sessions on a single day, and frequent rests were encouraged.

For each experimental run, a 2×2 stimulus-response confusion matrix was obtained, from which the percentage-correct (PC) score, the sensitivity index d', and the response bias β were calculated using signal-detection theory (Macmillan and Creelman, 2004). 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.

During the experiment, the TactuatorII was placed to the left of the participant's torso. It was covered by a padded wooden box that served as an armrest for the participant's left forearm. The top of the box had an opening that allowed the participant to place the thumb, the index finger, and the middle finger on the "fingerpad interface" rods (see inset in Fig. 1). Earmuff (Twin-Cup type, H10A, NRR 29, Peltor, Sweden) and pink noise (presented at roughly 80 dB SPL) were used to eliminate possible auditory cues.

B. Results

In general, performance indices increased as the participants gained more experience with the stimuli. Overall, the average sensitivity index of all pairs increased from $d'=2.66$ (PC=83%) in C1 to $d'=3.13$ (PC=91%) in C3. A pairwise two-sided t-test showed that sensitivity scores in C3 were significantly higher than in C1 [$t(39)=2.16$, $p<0.05$]. The sensitivity indices averaged over the two participants for each contrasting consonant in condition C3 are shown in Table III. For consonants contrasting in the voicing, place, and manner of articulation features, the performance levels of the two participants were similar in C3: $d'=3.5$ for S1 and $d'=3.5$ for S2 in voicing, $d'=2.8$ for S1 and $d'=2.6$ for S2 in place, and $d'=3.2$ and $d'=3.6$ for S1 and S2, respectively, in manner distinction. The response bias across the 20 consonant pairs ranged from $\beta=-0.74$ to $\beta=0.63$ and averaged $\beta=0.008$, indicating that the participants generally demonstrated little or no bias in their use of the two response choices. Both participants performed perfectly in discriminating the two consonant pairs /p,b/ and /d,ʒ/. For the remaining pairs, the participants' relative performance levels were mixed as one participant performed better than the other with some pairs but not others. In all cases, d' was greater than 1.0, a typical criterion for discrimination threshold, indicating that the coding scheme succeeded in providing the cues needed for the discrimination of the consonant pairs.

V. DISCUSSION

The coding scheme used in the present study was an extension of the scheme presented in Yuan, 2003, where the envelope of the low-frequency speech band (<350 Hz) was modulated with a 50-Hz vibration at the thumb and the envelope of the high-frequency speech band (>3000 Hz) was modulated with a 250-Hz vibration at the index finger. This scheme was successful in pairwise discrimination of initial consonants that contrasted in voicing only. On average, discriminability of roughly 90% and d' of 2.4 were obtained for eight voiced-unvoiced pairs in four participants. Our coding scheme presented the low-frequency speech band (fundamental frequency band) directly at the thumb and the envelopes of the high-frequency speech band (second-formant band) at the index fingerpad, consistent with the scheme presented in Yuan,2003. The results of the two studies show similar performance: An average discriminability of 94% and d' of 3.5 were obtained in the present study when contrasting

five voiced-unvoiced consonant pairs, indicating that the coding scheme used in Yuan, 2003 was a subset of the scheme used in the present study. The performance level obtained in the present study also appears to compare favorably with the results reported by earlier studies of tactual displays, where discrimination scores were generally less than 75% (Plant, 1989; Clements *et al.*, 1988; Galvin *et al.*, 1999; Waldstein and Boothroyd, 1995; Weisenberger *et al.*, 1989; Summers *et al.*, 2005).

In addition to incorporating the amplitude information from low- and high-frequency speech bands, as in Yuan, 2003, our coding scheme displays energy information from the mid-frequency speech band in the form of temporal envelopes as well as low-frequency motion cues from the three speech bands to the corresponding fingerpads. To the best of our knowledge, this is the first time that low-frequency motion cues have been used to encode speech spectral information. These cues provide both frequency *location* and frequency *transition* information of formants to the receiver's fingerpads. The transition of formants is useful for the distinction of the place of articulation feature in consonants as indicated in Ali *et al.*, 2001a,2001b and Jongman *et al.*,2000. Although some studies have presented contradictory results arguing that formant transitions are not useful for distinguishing place of articulation in consonants [e.g., see a review by Jongman *et al.* (2000)], motion waveforms extracted from the speech-to-touch coding scheme in the present study (see Fig. 8) indicate distinction in transitions as the place of constriction during the production of consonants varies from lips to velum. The cues associated with transition of the second formant can be observed in Fig. 8 (dashed lines). The index finger flexes at the onset of the initial bilabial plosive /b/ and stays flexed for the medial vowel /ɑ/ (first row). The index finger extends at the onset of the initial velar plosive /g/ and flexes at the onset of the medial vowel /ɑ/ (second row). Similarly, for fricatives, the index finger stays at the neutral zero position at the onset of the initial labiodental consonant /v/ (third row) and slightly extends at the onset of the initial dental fricative /ð/ (fifth row). The index finger extends for a longer duration at the onset of the initial alveolar and the post-alveolar fricatives /z/ and /ʒ/ before it flexes at the onset of the medial vowel /ɑ/ (fourth and sixth rows). Thus, as the place of articulation of consonant moves from near lips (bilabials) to near velum, the index finger extends more for the latter initial consonants (corresponding to an increase in F2 associated with an effective shortening of the vocal tract for velar as opposed to labial constrictions). This may explain the better performance level we have achieved in the present study due to the utilization of place of articulation cues.

Results of the pairwise consonantal discrimination experiments in the present study showed that both participants were able to discriminate all eight consonant pairs that differed in the place of articulation feature with an average discriminability of 88% and a d' of 2.7. The results of previous studies with tactual displays indicate poor transmission of place cues. For example, Clements *et al.* (1988) used a 12-by-12 pin tactual matrix display to present acoustic features as vibrations along the two dimensions of the display similar to that in the spectrogram used for speech analysis. The pairwise discrimination performance of the manner of articulation and voicing features was satisfactory (71% for voicing and 80% for manner) but discriminability of place of articulation distinction was poorer, i.e., 66%. Even with the multi-channel spectral display of the Queen's vocoder studied by Weisenberger *et al.* (1989), place of articulation was not discriminated as well as other features (65% for place compared to 75% for manner and 70% for voicing). In other studies, discriminability of place of articulation was at chance level (Waldstein and Boothroyd, 1995; Plant, 1989; Summers *et al.*, 2005; Weisenberger and Percy, 1995; Galvin *et al.*, 1999). Therefore, it appears that the present coding scheme was able to transmit the place of articulation feature more successfully than has been demonstrated previously.

Consonants contrasting manner of articulation have been shown to be well discriminated with the tactile displays of previous studies, i.e., 80% in Clements *et al.*, 1988, 75% in Weisenberger *et al.*, 1989, ≤90% in Weisenberger and Percy, 1995, 70% in Plant, 1989, and <85% in Summers *et al.*, 2005. In the present study, the discriminability of manner of articulation was always greater than 90% except for the /s/-/tʃ/ contrast (88%) by S1 and the /d/-/dʒ/ contrast (84%) by S2. The manner of articulation distinction is associated with coarse spectral variations in speech such as abrupt or smooth temporal variations (e.g., plosives vs fricatives) or the combination of both (as in affricates). The manner discrimination results obtained in the present study are comparable to the best performance obtained with previous tactile speech displays.

A major distinction between the present and previous studies is that the previous displays utilized either the tactile or the kinesthetic sensory system, but not both, to transmit acoustic and phonetic features associated with consonantal and vocalic contrasts (Bliss, 1962; Tan *et al.*, 1997). The two sensory systems are perceptually independent (Bolanowski *et al.*, 1988; Israr *et al.*, 2006) and can be utilized simultaneously to improve the transmission of features associated with speech signals. Tan *et al.* (1999, 2003) formed a set of synthetic waveforms from the two sensory systems and demonstrated that relatively high rates of information could be transmitted through the tactual sense. In the present study, we utilized the entire kinesthetic-cutaneous sensory continuum in an effort to broaden the dynamic range of tactual perception, similar to the Tadoma method, and to improve speech transfer through the human somatosensory system. Our results demonstrate that with the new controller and the coding scheme reported here that engage both the kinesthetic and cutaneous aspects of the somatosensory system, normal-hearing participants were able to discriminate consonantal features at a level that is similar to or higher than those reported by previous studies with other types of tactual speech-information displays.

ACKNOWLEDGMENTS

This research was supported by Research Grant No. R01-DC00126 from the National Institute on Deafness and

Other Communication Disorders, National Institutes of Health.

[1]The first Tactuator was developed at MIT (Tan and Rabinowitz, 1996). A second device, the TactuatorII, was subsequently developed at Purdue University with essentially the same hardware.

[2]The unit "dB with regard to 1 μm peak" is commonly used with HDTs. It is computed as $20 \log_{10}(A/1.0)$ where A denotes the amplitude (in μm) of the sinusoidal signal that can be detected by a human participant at a specific frequency.

Ali, A. M. A., Van der Spiegel, J., and Mueller, P. (**2001a**). "Acoustic-phonetic features for the automatic classification of fricatives," J. Acoust. Soc. Am. **109**, 2217–2235.

Ali, A. M. A., Van der Spiegel, J., and Mueller, P. (**2001b**). "Acoustic-phonetic features for the automatic classification of stop consonants," IEEE Trans. Speech Audio Process. **9**, 833–841.

Bliss, J. C. (**1962**). "Kinesthetic-tactile communications," IRE Trans. Inf. Theory **8**, 92–99.

Bolanowski, S. J., Gescheider, G. A., Verrillo, R. T., and Checkosky, C. M. (**1988**). "Four channels mediate the mechanical aspects of touch," J. Acoust. Soc. Am. **84**, 1680–1694.

Brisben, A. J., Hsiao, S. S., and Johnson, K. O. (**1999**). "Detection of vibration transmitted through an object grasped in the hand," J. Neurophysiol. **81**, 1548–1558.

Clements, M. A., Braida, L. D., and Durlach, N. I. (**1988**). "Tactile communication of speech: Comparison of two computer-based displays," J. Rehabil. Res. Dev. **25**, 25–44.

Craig, J. C., and Evans, P. M. (**1987**). "Vibrotactile masking and the persistence of tactual features," Percept. Psychophys. **42**, 309–371.

Evans, P. M. (**1987**). "Vibrotactile masking: Temporal integration, persistence and strengths of representations," Percept. Psychophys. **42**, 515–525.

Galvin, K. L., Mavrias, G., Moore, A., Cowan, R. S. C., Blamey, P. J., and Clark, G. M. (**1999**). "A comparison of Tactaid II+ and Tactaid 7 use by adults with a profound hearing impairment," Ear Hear. **20**, 471–482.

Gescheider, G. A., Capraro, A. J., Frisina, R. D., Hamer, R. D., and Verrillo, R. T. (**1978**). "The effects of a surround on vibrotactile thresholds," Sens Processes **2**, 99–115.

Goble, A. K., Collins, A. A., and Cholewiak, R. W. (**1996**). "Vibrotactile threshold in young and old observers: The effects of spatial summation and the presence of a rigid surround," J. Acoust. Soc. Am. **99**, 2256–2269.

Grant, K. W., Ardell, L. H., Kuhl, P. K., and Sparks, D. W. (**1985**). "The contribution of fundamental frequency, amplitude envelope, and voicing duration cues to speechreading in normal-hearing subjects," J. Acoust. Soc. Am. **77**, 671–677.

Israr, A. (**2007**). "Tactual transmission of phonetic features," Ph.D. thesis, Purdue University, West Lafayette, IN.

Israr, A., Tan, H. Z., and Reed, C. M. (**2006**). "Frequency and amplitude discrimination along the kinesthetic-cutaneous continuum in the presence of masking stimuli," J. Acoust. Soc. Am. **120**, 2789–2800.

Jongman, A., Wayland, R., and Wong, S. (**2000**). "Acoustic characteristics of English fricatives," J. Acoust. Soc. Am. **108**, 1252–1263.

Leek, M. R. (**2001**). "Adaptive procedures in psychophysical research," Percept. Psychophys. **63**, 1279–1292.

Maciejowski, J. M. (**1989**). *Multivariable Feedback Design* (Addison-Wesley, Reading, MA).

Macmillan, N. A., and Creelman, C. D. (**2004**). *Detection Theory: A User's Guide* (Lawrence Erlbaum Associates, New York).

Plant, G. (**1989**). "A comparison of five commercially available tactile aids," Aust. J. Audiol. **11**, 11–19.

Reed, C. M., and Durlach, N. I. (**1998**). "Note on information transfer rates in human communication," Presence—Teleoperators & Virtual Environments **7**, 509–518.

Reed, C. M., Rabinowitz, W. M., Durlach, N. I., Braida, L. D., Conway-Fithian, S., and Schultz, M. C. (**1985**). "Research on the Tadoma method of speech communication," J. Acoust. Soc. Am. **77**, 247–257.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.

Summers, I. R. (**1992**). *Tactile Aids for the Hearing Impaired* (Whurr, London).

Summers, I. R., Dixon, P. R., Cooper, P. G., Gratton, D. A., Brown, B. H., and Stevens, J. C. (**1994**). "Vibrotactile and electrotactile perception of time-varying pulse trains," J. Acoust. Soc. Am. **95**, 1548–1558.

Summers, I. R., Whybrow, J. J., Gratton, D. A., Milnes, P., Brown, B. H., and Stevens, J. C. (**2005**). "Tactile information transfer: A comparison of two stimulation sites," J. Acoust. Soc. Am. **118**, 2527–2534.

Tan, H. Z., Durlach, N. I., Rabinowitz, W. M., Reed, C. M., and Santos, J. R. (**1997**). "Reception of Morse code through motional, vibrotactile and auditory stimulation," Percept. Psychophys. **59**, 1004–1017.

Tan, H. Z., Durlach, N. I., Reed, C. M., and Rabinowitz, W. M. (**1999**). "Information transmission with a multifinger tactual display," Percept. Psychophys. **61**, 993–1008.

Tan, H. Z., and Rabinowitz, W. M. (**1996**). "A new multi-finger tactual display," *Proceedings of the International Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, edited by K. Danai (American Society of Mechanical Engineers, New York), Vol. **58**, pp. 515–522.

Tan, H. Z., Reed, C. M., Delhorne, L. A., Durlach, N. I., and Wan, N. (**2003**). "Temporal masking of multidimensional tactile stimuli," J. Acoust. Soc. Am. **114**, 3295–3308.

Van Doren, C. L. (**1990**). "The effects of a surround on vibrotactile thresholds: Evidence for spatial and temporal independence in the non-Pacinian I (NPI) channel," J. Acoust. Soc. Am. **87**, 2655–2661.

Verrillo, R. T., and Gescheider, G. A. (**1992**). "Perception via the sense of touch," in *Tactile Aids for the Hearing Impaired*, edited by I. R. Summers (Whurr, London), pp. 1–36.

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

Weisenberger, J. M., Broadstone, S. M., and Saunders, F. A. (**1989**). "Evaluation of two multichannel tactile aids for the hearing impaired," J. Acoust. Soc. Am. **86**, 1764–1775.

Weisenberger, J. M., and Percy, M. E. (**1995**). "The transmission of phoneme-level information by multichannel tactile speech perception aids," Ear Hear. **16**, 392–406.

Yuan, H. (**2003**). "Tactual display of consonant voicing to supplement lipreading," Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA.