Temporal and spatio-temporal vibrotactile displays for voice fundamental frequency: An initial evaluation of a new vibrotactile speech perception aid with normal-hearing and hearing-impaired individuals

Edward T. Auer, Jr. and Lynne E. Bernstein
Spoken Language Processes Laboratory, House Ear Institute, 2100 West Third Street, Los Angeles, California 90057

David C. Coulter
Coulter Associates, Fairfax, Virginia

(Received 6 March 1997; revised 27 April 1998; accepted 30 June 1998)

Four experiments were performed to evaluate a new wearable vibrotactile speech perception aid that extracts fundamental frequency ($F0$) and displays the extracted $F0$ as a single-channel temporal or an eight-channel spatio-temporal stimulus. Specifically, we investigated the perception of intonation (i.e., question versus statement) and emphatic stress (i.e., stress on the first, second, or third word) under Visual-Alone (VA), Visual-Tactile (VT), and Tactile-Alone (TA) conditions and compared performance using the temporal and spatio-temporal vibrotactile display. Subjects were adults with normal hearing in experiments I–III and adults with severe to profound hearing impairments in experiment IV. Both versions of the vibrotactile speech perception aid successfully conveyed intonation. Vibrotactile stress information was successfully conveyed, but vibrotactile stress information did not enhance performance in VT conditions beyond performance in VA conditions. In experiment III, which involved only intonation identification, a reliable advantage for the spatio-temporal display was obtained. Differences between subject groups were obtained for intonation identification, with more accurate VT performance by those with normal hearing. Possible effects of long-term hearing status are discussed. [S0001-4966(98)03010-0]

PACS numbers: 43.71.Ma [WS]

INTRODUCTION

The substitution of the sense of touch for the sense of hearing has been explored intermittently across almost the entire twentieth century (Reed *et al.*, 1993), with the primary goal to develop portable devices to enhance speech communication by hearing-impaired individuals. Fundamental issues continue to be: (1) what speech information to encode for presentation to the skin (e.g., spectral patterns or extracted acoustic-phonetic cues); and (2) how to present the information (e.g., as mechanical or electrical stimulation; as rate displays, or spatial displays). To date, several different tactile speech perception aids have been developed and tested, encoding a variety of sources of speech information in a variety of presentation schemes designed to enhance speechreading (for reviews see Kishon-Rabin *et al.*, 1996; Reed *et al.*, 1993; Summers, 1992). Reported enhancements to speechreading have generally been in the range of 4–20 percentage points for words correct in open-set sentence identification tasks.

One type of speech perception aid, which has motivated a fair amount of research, delivers extracted voice fundamental frequency ($F0$) (Bernstein *et al.*, 1989, 1998; Boothroyd and Hnath, 1986; Eberhardt *et al.*, 1990; Grant, 1987; Hanin *et al.*, 1988; Hnath-Chisolm and Kishon-Rabin, 1988; Kishon-Rabin *et al.*, 1996; Plant and Risberg, 1983; Rothenberg and Molitor, 1979; Summers, 1992; Waldstein and Boothroyd, 1995a, b). Voice $F0$ has been studied for several reasons:(1) $F0$ characteristics contribute to speech perception at several different linguistic levels; (2) voice $F0$ is generated at the glottis, which is invisible to the speechreader; and (3) the presentation of simple *acoustic* stimuli composed of pulses generated as a function of $F0$ greatly enhances speechreading. Normal-hearing adults have been shown to improve 40–50 percentage points over speechreading alone when they speechread with an acoustic $F0$ supplement (Boothroyd *et al.*, 1988; Breeuwer and Plomp, 1984, 1985, 1986; Grant, 1987; Rosen *et al.*, 1981).

A brief inventory of the linguistic levels at which $F0$ has been shown to be a factor readily illustrates the first point above. The onset time of vocal fold vibration relative to supralaryngeal articulator release (i.e., voice onset time) is a primary contributor to perception of prevocalic consonantal voicing distinctions (Lisker and Abramson, 1967). $F0$ direction and height at the onset of prevocalic consonants also vary systematically with the consonantal voicing distinctions (Haggard *et al.*, 1970; House and Fairbanks, 1953). $F0$ is an acoustical correlate of lexical stress (e.g., *CONvert* versus *conVERT*) and sentential stress (Lehiste, 1970; Bolinger, 1958; Fry, 1955, 1958) also referred to as accent (Sluijter and van Heuven, 1996). Lexical stress patterns may be involved in the perception of word boundaries (Cutler and Butterfield, 1992). $F0$ also systematically signals syntactic structure (Bolinger, 1978; Lehiste, 1970; Pike, 1945) (e.g.,

the contrast between a sentence spoken as a statement or a question) and can signal phrase boundaries (Streeter, 1978).

The present study reports on an initial evaluation of a new design for a wearable vibrotactile speech perception aid that employs a digital processor for the automatic extraction of $F0$. The new aid was evaluated using closed-set identification of emphatic sentential stress (on the first, second, or third word of the stimulus) and intonation (question versus statement). The main questions were: (1) whether one of the two different vibrotactile displays of $F0$ (one a single-channel temporal display and one an eight-channel spatio-temporal display) was more successful for conveying $F0$; and (2) whether auditory speech perception experience was a factor in the successful perception of $F0$ information.

I. DISPLAYS

A main issue in designing an $F0$ vibrotactile speech perception aid is how to tailor the linguistically relevant characteristics of the $F0$ signal to vibrotactile perceptual characteristics. $F0$ varies over the range of approximately 70–500 Hz across children, women, and men, with individuals' voices exhibiting ranges of 1–1.5 octaves (Hess, 1983). Based on their research on $F0$, Hnath-Chisolm and Boothroyd (1992) suggested that 1/8-octave or better resolution is needed to effectively encode $F0$ tactually, but semitone (1/12-octave) resolution has also been suggested in the literature (Hermes and van Gestel, 1991).

Vibrotactile psychophysical experiments by Rothenberg *et al.* (1977) showed that the vibrotactile frequency difference limen results in at least ten discriminable steps between approximately 10 and 90 Hz with some additional steps, if the range is extended to 300 Hz. Beyond 300 Hz, frequency discrimination falls off rapidly. Therefore, a direct one-to-one transformation from acoustic pitch periods to vibrotactile rate is unlikely to be optimal. For example, the entire range of a child's voice (approximately 157–444 Hz; Hess, 1983) would be above the range at which frequency is best resolved by the skin. The $F0$ range of adult males was even found to be too large and poorly resolved with a direct acoustic-to-vibrotactile transformation (Bernstein *et al.*, 1989).

A. Single-channel temporal displays

Rothenberg and Molitor (1979) proposed to solve the problem of transforming acoustic $F0$ to vibrotactile $F0$ by using frequency range compression and frequency shifting. Detected voice $F0$ was encoded as vibrotactile pulses delivered by a single vibrator. The most favorable results were obtained when frequency was shifted so as to center around 50 Hz, with the scale factor at 1:1 or 1:0.75 (for a male talker whose untransformed 80-Hz range was centered at 125 Hz). Their experimental task was forced-choice identification of the stressed word and intonation type of a sentence, *Ron will win*, presented as a series of vibrotactile pulses. The sentence was spoken with emphasis on one of the three words and with the intonation pattern of either a question (rising) or a statement (falling). The results confirmed the prediction that perception is most accurate when vibrotactile stimuli make use of the frequencies below approximately 100 Hz. Analysis of Fig. 5 in the Rothenberg and Molitor (1979) study suggests that stress judgments were 51% correct ($n=8$ normal-hearing individuals, chance at 33%), and intonation judgments were approximately 90% correct ($n=8$, chance at 50%).

Using a similar paradigm to Rothenberg and Molitor (1979), Bernstein *et al.* (1989) tested several alternate $F0$-to-vibrotactile transformations. However, Bernstein *et al.* employed a much larger and more varied stimulus sentence set leading to higher uncertainty. In addition, because use of an $F0$ speech perception aid would necessarily involve speechreading under practical (not laboratory) conditions, they compared stress and intonation identification under tactile-alone (TA), visual-tactile (VT), and visual-alone (VA) conditions. When $F0$ was shifted into the range below approximately 100 Hz, stress and intonation patterns were perceived most accurately. Differences in display effectiveness were only observed under TA conditions. Under the TA condition, the most effective transformation produced somewhat more accurate identification of stress (approximately 60% correct for the best subject and a group mean of 52%, $n=5$ normal-hearing individuals) and somewhat less accurate identification of intonation (approximately 80% correct for the best subject and a group mean of 70%, $n=5$) than obtained by Rothenberg and Molitor (1979).

B. Multichannel spatio-temporal displays

An alternate method for solving the problem of matching the acoustic $F0$ characteristics to the skin's perceptual capabilities is to engage spatial perception, following the suggestion of Kirman (1973) who argued that temporal pattern perception is not a forte of the skin. Spatial resolution for touch is extremely good. Phillips and Johnson (1985) reported spatial resolution of the finger pad to be 0.7 mm for a moving stimulus. Blind readers of Braille can achieve rates of 80–200 words per minute (Foulke, 1991), indicating potential for high information transfer rates. Spatial $F0$ displays might deliver greater information than temporal displays and also result in more effective processing under cross-modal conditions. However, vibrotactile perception is also susceptible to distortions due to interactions between time and space (Geldard, 1985; Geldard and Sherrick, 1972). These distortions could enhance or degrade the perception of speech information (Kirman, 1973).

Boothroyd and colleagues implemented a spatio-temporal vibrotactile $F0$ display scheme (Boothroyd and Hnath, 1986; Boothroyd *et al.*, 1988; Hnath-Chisolm and Medwetsky, 1988; Hanin *et al.*, 1988; Waldstein and Boothroyd, 1995a) (for an aid employing a purely spatial electrotactile display scheme, see Grant *et al.*, 1986). $F0$ was extracted using an analog circuit and was encoded via vibration rate and vibrator location in a single-dimensional vibrator array (Yeung *et al.*, 1988). Under this scheme, frequency resolution is theoretically limited only by vibrotactile spatial resolution and the number of vibrotactile stimulators (Hnath-Chisolm and Kishon-Rabin, 1988). The spatial component of the Boothroyd display used a change of 0.16 octaves in input frequency to cause a change in selection of the vibrator for

activation.[1] The temporal component of the display was the output of a pulse on the spatially appropriate stimulator for every other detected pitch pulse.

Hnath-Chisolm and Kishon-Rabin (1988) compared this spatio-temporal display with a single-channel temporal-only display of $F0$, which also employed the every-other-pulse scheme. Their experimental tasks included identification of stress and intonation in separate conditions, and results were interpreted as showing some advantage for the spatio-temporal display (VT=80% correct and TA=78% correct, $n=6$ normal-hearing individuals).[2] However, Hnath-Chisolm and Kishon-Rabin's temporal display of intonation (VT=70% correct and TA=64% correct, $n=6$) failed to replicate TA intonation identification accuracy obtained by Rothenberg and Molitor (1979), and Bernstein *et al.* (1989) with several different temporal displays, calling into question the conclusion that there is a general advantage for a spatio-temporal display.

II. CURRENT STUDY

The present study used the same task employed in Bernstein *et al.* (1989) to investigate whether an eight-channel spatio-temporal or single-channel temporal display was more effective at conveying intonation and sentential stress information. The temporal display consisted of the presentation of every other detected pitch period, and the spatio-temporal display consisted of presentation of the same pulses but distributed across an eight-channel single-dimensional array. Experiment I investigated whether vibrotactile $F0$ enhanced stress and intonation identification beyond VA identification, and whether there were differential effects of display. Experiment II assessed the perception of stress and intonation information and display in the TA condition. Experiment III further investigated display effects for intonation identification under VT and TA conditions using a within-subjects design to compare displays. In addition to group comparisons, individual performance differences were examined as a function of display.

The goal of sensory substitution is to develop vibrotactile speech perception aids for hearing-impaired individuals. However, the relationship between the specialized experience of hearing-impaired individuals and vibrotactile speech perception has received little attention in the literature (e.g., Bernstein *et al.*, in press; Rothenberg and Molitor, 1979). Therefore, experiment IV examined effects of auditory experience on vibrotactile $F0$ perception by hearing-impaired adults.

III. EXPERIMENT I. VISUAL-ALONE VERSUS VISUAL-TACTILE JUDGMENTS OF STRESS AND INTONATION

In experiment I, under VA and VT conditions, subjects identified both the position of an emphatically stressed word in a sentence and the sentence intonation pattern.

A. Subjects

The 12 subjects (all female), were age 20–29 years old, with normal hearing, 20/30 or better normal or corrected vision, and English as their native language. Six were randomly assigned to the single-channel temporal display and six to the eight-channel spatio-temporal display. Subjects were paid for their time in the experiment.

B. Stimuli

1. Sentence stimuli

Four sentences were used to generate stimuli for this and subsequent experiments.

(1) We owe you a yoyo.
(2) We will weigh you.
(3) Pat cooked Pete's breakfast.
(4) Chuck caught two cats.

Phonetic content was controlled to examine the influence of continuously voiced versus interrupted vibrotactile stimulus patterns: the first two sentences (referred to as ''continuous'') have voiced continuants (resulting in almost uninterrupted voicing), and the second two (referred to as ''discontinuous'') have voiceless stop and fricative consonants (except for /b/ in ''breakfast'') (resulting in brief periods of silence). It was predicted that vibrotactile stress identification would be easier for sentences with voiceless consonants, because the silences would facilitate identifying syllable location and duration (Bernstein *et al.*, 1989). At the same time, inasmuch as those gaps interrupt intonation contours, they might reduce vibrotactile intonation identification accuracy (Bernstein *et al.*, 1989). Twenty-four stimulus sentences were generated from orthogonal combination of intonation pattern (statement or question) and position of the word receiving emphatic stress (first, second, or third). For example, the base sentence ''We owe you a yoyo.'' generated the following six sentences in which the capitalized word received emphatic stress:

WE owe you a yoyo.
WE owe you a yoyo?
we OWE you a yoyo.
we OWE you a yoyo?
we owe YOU a yoyo.
we owe YOU a yoyo?

Two tokens of each sentence in its six different instantiations were spoken by a male and a female talker (96 tokens total).

Stimuli were recorded on videotape and then stored on video laserdisc (Bernstein and Eberhardt, 1986). The 96 stimuli had been previously evaluated for production accuracy via auditory testing (98% correct stress and intonation identification) (Bernstein *et al.*, 1989). A linear-phase electroglottograph (EGG) (Synchrovoice) signal, which was passed through a high-pass filter (Glottal Enterprise) was recorded on one audio channel of the video recording. The dominant frequency in the EGG signal is the voice $F0$ (Fourcin, 1981). The EGG signals were used as input to the vibrotactile aid following gating to eliminate spurious vibrotactile output due to shifts in the noise floor as the videodisc player moved from pause to play. Video stimuli were presented on a color monitor.

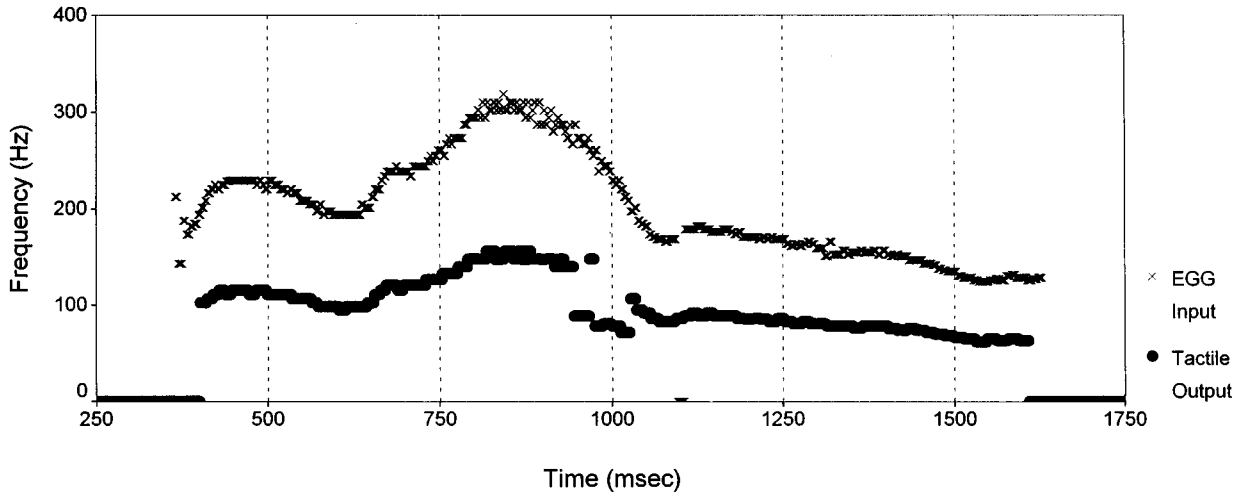

FIG. 1. Frequency in Hz is displayed as a function of time for the input EGG signal (indicated by the ×'s) and the output tactile signal (indicated by the filled circles). The statement ''We WILL weigh you.'' was spoken by the female talker. Tactile-alone stress (chance=33%) judgment accuracy for this token was 30% correct with the temporal display and 45% correct with the spatio-temporal display. Tactile-alone intonation (chance=50%) judgment accuracy was 85% correct with the temporal display and 100% correct with the spatio-temporal display.

2. Vibrotactile stimulus generation

Both vibrotactile displays in these experiments employed the same real-time $F0$ extraction algorithm. The algorithm was based on a time domain model of the decay of the acoustic speech waveform following glottal excitation. The algorithm was implemented on a Texas Instruments TMS320 digital signal processing chip. Using a set of conditions prespecified in software, the algorithm compared the decline in amplitude over a sequence of digital samples with the decay rate of the model. Criteria known to the model were adaptively updated. Initial conditions were based on likely first formant values, which determine the largest upward excursion in the speech waveform. When the amplitude of the sampled waveform exceeded criteria for decay, or the duration of the putative pitch period was within the limits of the model, a pitch period was reported by the processor. Over time, in the absence of input, the model returned to its initial values.

The performance of the current $F0$ extraction algorithm has not yet been subjected to direct comparison with other possible real-time extraction $F0$ extraction algorithms (for a comparison of three alternate algorithms, see Bosman and Smoorenburg, 1997). However, we have examined several randomly selected examples of the pitch tracking by the current algorithm (see Figs. 1–4). Frequency in Hz is displayed as a function of time for the handpicked pitch of the input EGG signal and the output vibrotactile signal. Figures 1 and 3 show better frequency tracking for continuously voiced sentences, as would be expected given the difficulty of extracting pitch from speech with many stop consonants. The figures also show that errors were made in tracking $F0$.

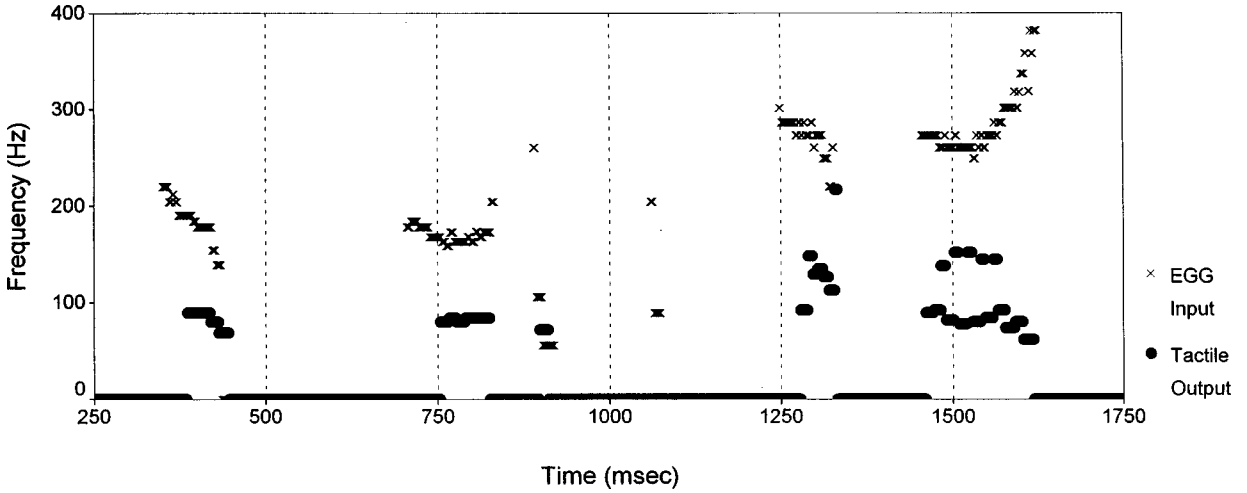

FIG. 2. Frequency in Hz is displayed as a function of time for the input EGG signal (indicated by the ×'s) and the output tactile signal (indicated by the filled circles). The question ''Chuck CAUGHT two cats?'' was spoken by the female talker. Tactile-alone stress (chance=33%) judgment accuracy for this token was 85% correct with the temporal display and 60% correct with the spatio-temporal display. Tactile-alone intonation (chance=50%) judgment accuracy was 100% correct with the temporal display and 90% correct with the spatio-temporal display.

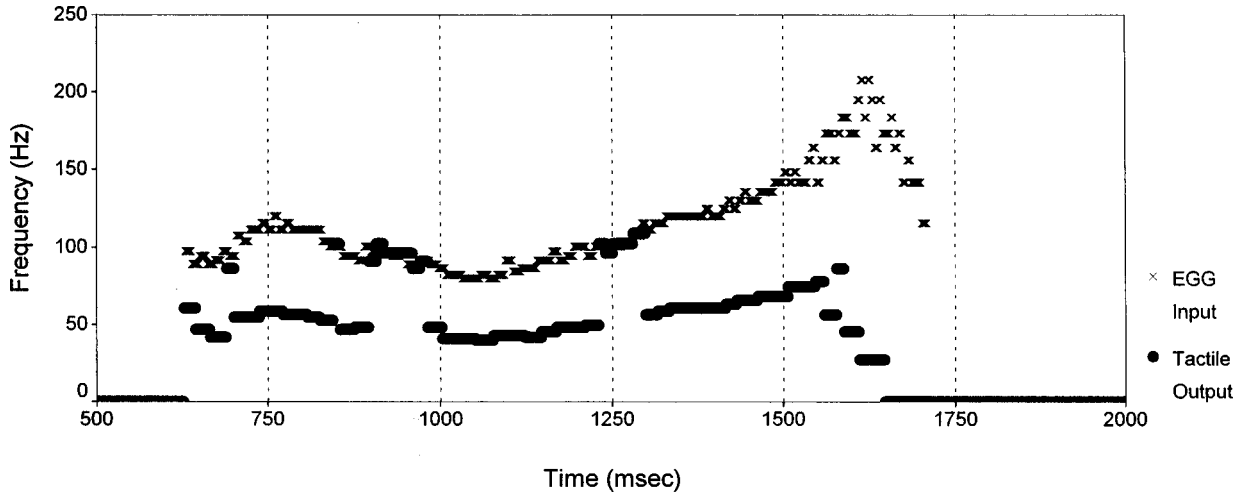

FIG. 3. Frequency in Hz is displayed as a function of time for the input EGG signal (indicated by the ×'s) and the output tactile signal (indicated by the filled circles). The question ''We will WEIGH you?'' was spoken by the male talker. Tactile-alone stress (chance=33%) judgment accuracy for this token was 40% correct with the temporal display and 70% correct with the spatio-temporal display. Tactile-alone intonation (chance=50%) judgment accuracy was 100% correct with the temporal display and 100% correct with the spatio-temporal display.

However, examination of the human performance data taken from experiment II (see captions of Figs. 1–4) suggests that with tactile information alone, stress and intonation judgments can be accurate even when tracking errors occurred. For example, despite substantial errors in the tracking of the sentence final frequency contour for the token displayed in Fig. 2, intonation judgments were highly accurate.

The processor in the vibrotactile aid was programmed to send a signal to a custom electronic circuit that initiated a square pulse for every second detected pitch pulse. This scheme was required by the spatio-temporal display in order to determine the appropriate output channel. Discarding the initial pulse introduced a variable delay in the output. Analysis of 8 of the 96 tokens suggested that onset delays ranged between 0 and 50 ms with an average onset delay (~30 ms). However, this delay was within the 40-ms limit that McGrath and Summerfield (1985) suggested as sufficient to *not* dramatically affect audio-visual speech understanding.

The scheme used to assign output channels for the eight-channel aid is shown in Fig. 5. Output channel number is plotted as a function of input rate. The lines with the open squares represent the channel selection scheme for the male talker. Channel selection was programmed on the basis of prior analysis of the frequency ranges of the two talkers (Bernstein *et al.*, 1989) and was designed to center the eight channels over the range of frequencies containing the center 90% of detected pitch periods for each talker. Frequency ranges were allocated to channels equally except for the highest and lowest channels, which were expanded to accommodate extreme pitch excursions. Spatial resolution was approximately 0.083 octaves per vibrotactile stimulator location for the male talker and 0.125 octaves for the female talker. Vibrotactile stimulation was delivered via small solenoids with contact points of 2 mm in diameter arrayed with 4-mm separation between them. The single-channel display presented every pulse on a single solenoid. Amplitude of the

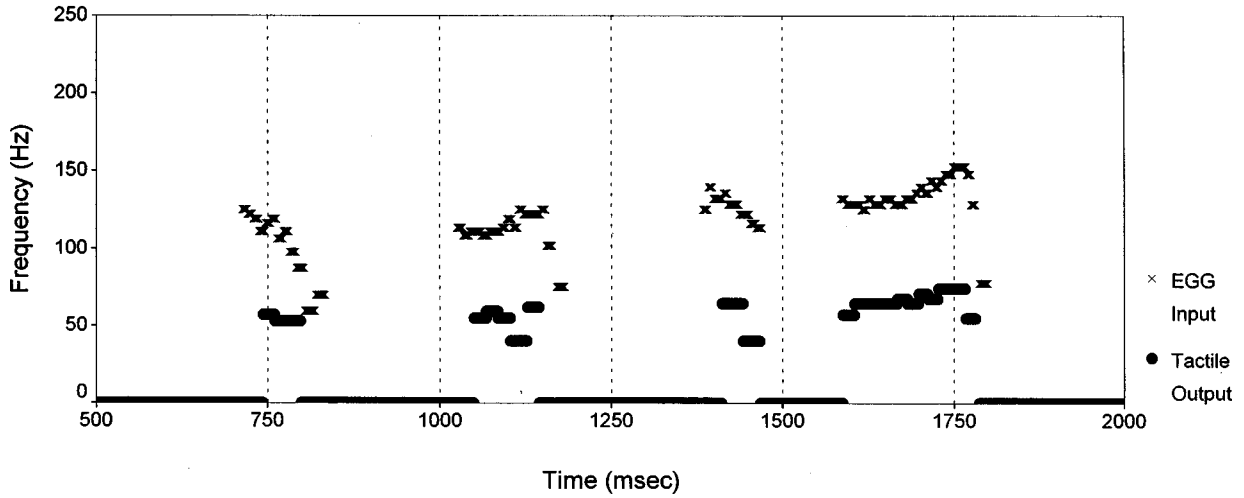

FIG. 4. Frequency in Hz is displayed as a function of time for the input EGG signal (indicated by the ×'s) and the output tactile signal (indicated by the filled circles). The question ''Chuck CAUGHT two cats?'' was spoken by the male talker. Tactile-alone stress (chance=33%) judgment accuracy for this token was 55% correct with the temporal display and 30% correct with the spatio-temporal display. Tactile-alone intonation (chance=50%) judgment accuracy was 75% correct with the temporal display and 90% correct with the spatio-temporal display.

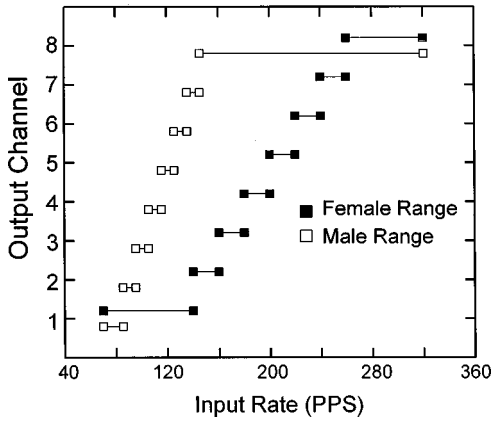

FIG. 5. Channel selection for spatio-temporal display of voice *F*0. The lines delimited by the open squares indicate the input ranges for the male display. The lines delimited by the filled squares indicate the ranges for the female display.

output signal was not modulated by the amplitude input signal. All stimuli were presented to the hypothenar of the left hand.[3]

C. Design and procedure

Stimulus presentation and response collection were computer controlled. The subject's task was to make a six-alternative forced-choice identification of the emphatically stressed word (first, second, or third word) and the intonation pattern (question or statement) of each stimulus sentence. Responses were collected using a six-button response box with ''question'' responses on the right side of the box and ''statement'' responses on the left. Buttons on each side were numbered to correspond to the position in the sentence of the stressed word.

LEDs on the response box signaled the beginning of a trial. The first frame of a video stimulus was then presented and paused briefly, with the talker in a relaxed position. The video stimulus was then played in real-time. The final frame of the stimulus was then paused with the talker again in a relaxed position. Following each response, feedback was provided by lighting an LED over the button associated with the correct response.

Subjects were tested in both VA and VT conditions, with the order of conditions counterbalanced. Sentences for a single talker were presented in blocks of 48 (2 tokens×4 base sentences×2 types of intonation×3 positions of stress). Subjects received 20 blocks in the visual condition and 30 blocks in the VT condition. The talker was alternated every five blocks, and the order of talker presentation was counterbalanced across subjects. During VT trials, subjects wore earplugs and received shaped noise through headphones to mask possible sounds caused by the stimulator. The stimulus blocks lasted approximately 60 mins, and subjects were tested in seven sessions. The subjects received at most 3 h of vibrotactile experience.

Subjects were informed of both the structure and content of the stimulus sentences and practice was administered prior to data collection. Practice preceding the VA condition consisted of audiovisual trials with 48 sentences (two talkers ×four base sentences×two intonation×three stress) and visual trials with 48 sentences, each sentence followed by the subject's response and feedback. Practice prior to the VT condition consisted of audiovisual presentation of the 48 sentences, the same sentences presented with video and EGG signals auditorily, and then presented with VT stimulation. Feedback was given following each trial.

TABLE I. Experiment I: Individual subject data in percent correct for *stress* judgments under visual-alone and visual-tactile conditions (Chance =33%). Display=1 indicates the subject used the temporal display; Display=8 indicates that the subject used the spatio-temporal display. Cont.=Continuous Sentences; Discont.=discontinuous sentences.

Subject (Display)	Visual-alone			Visual-tactile		
	Cont.	Discont.	Combined	Cont.	Discont.	Combined
1(1)	89	78	83	85	79	82
2(1)	82	74	78	74	81	78
3(1)	93	84	88	92	84	88
4(1)	85	75	80	83	85	84
5(1)	90	80	85	93	93	93
6(1)	97	93	95	88	87	88
7(8)	82	73	78	75	81	78
8(8)	77	71	74	78	69	73
9(8)	84	80	82	84	80	82
10(8)	93	84	89	86	80	83
11(8)	86	79	83	78	80	79
12(8)	82	83	82	91	88	90
Mean	87	79	83	84	82	83

D. Analyses

Separate $2\times2\times2\times2\times2$ (condition×talker×sentence type×order×display) repeated measures analyses of variance were performed on the mean percentage correct responses separately for stress and intonation.[4] Condition, talker, and sentence type were within-subjects variables, and order and display were between-subjects variables. The condition factor corresponded to whether the trials were performed VA or VT. The sentence-type factor compared the two sentences with voiced consonants to the two sentences with unvoiced consonants. Order corresponded to whether the subjects were tested in the VA condition first or second. Display corresponded to the temporal or spatio-temporal vibrotactile stimuli. Only the last five blocks of data for each talker, for each condition, were analyzed. Those blocks were considered representative of the subjects' most experienced performance within a given condition.

E. Results and discussion

VA stress and intonation identifications were reliably above chance (chance=33.3% for stress and 50% for intonation) according to the binomial test, but not at ceiling. (See Tables I and II for individual subject data.)

Stress identification did not differ as a function of condition [$F(1,8)=0.00$, $p>0.96$] (see Table I). The overall mean percent correct stress was 79% for the female talker and 87% for the male talker (chance=33%) [$F(1,8)=24.94$, $p<0.01$]. This result was consistent with previous results suggesting that the female talker was more difficult to speechread than the male (Demorest and Bernstein, 1987;

TABLE II. Experiment I: Individual subject data in percent correct for *intonation* judgments under visual-alone and visual-tactile conditions (Chance=50%). Display=1 indicates the subject used the temporal display; Display=8 indicates that the subject used the spatio-temporal display. Cont.=Continuous Sentences; Discont.=discontinuous sentences.

Subject (Display)	Visual-alone			Visual-tactile		
	Cont.	Discont.	Combined	Cont.	Discont.	Combined
1(1)	63	53	58	87	74	80
2(1)	63	55	59	70	62	66
3(1)	67	65	66	86	89	88
4(1)	76	66	71	93	88	91
5(1)	58	63	66	91	81	86
6(1)	79	68	74	81	75	78
7(8)	65	53	59	92	88	90
8(8)	61	60	61	85	70	78
9(8)	68	62	65	92	86	89
10(8)	68	65	67	84	78	81
11(8)	63	68	65	83	86	85
12(8)	62	54	58	73	66	70
Mean	67	61	64	85	79	82

Eberhardt *et al.*, 1990). Stress was more accurately identified in continuous sentences (85%) than in discontinuous sentences (81%) [$F(1,8)=20.96$, $p<0.01$]. This result was counter to the prediction that stress would be more perceptible when sentences had unvoiced consonants to mark syllable boundaries. However, because the visual stimulus appeared to provide most of the information about stress, it seemed that the mostly labial place of articulation for consonants in the continuous sentences provided the syllable boundary information. This result provides support for the hypothesis that visible changes in the rhythm of alternating mouth opening during vowels and closure or approximation during consonants provides cues to the location of emphatically stressed words in sentences (Bernstein *et al.*, 1989). Both talker and sentence-type interacted with condition. Examination of the means for both of these interactions suggested that the sentence-type (see Table I) and talker differences (VA: male=88% correct, female=78% correct; VT: male=86% correct, female=80% correct) were reduced in magnitude in the visual-tactile condition relative to the visual-alone condition. These interactions are not readily interpretable in terms of the influence of the vibrotactile speech perception aid.

Identification accuracy for *intonation* differed across conditions (82% VT and 64% VA) [$F(1,8)=58.39$, $p<0.01$] (see Table II). However, the spatio-temporal and temporal display were not different [$F(1,8)=0.20$, $p=0.66$], and display did not interact with condition (VA: temporal=66%, spatio-temporal=62%; VT: temporal=82%, spatio-temporal=82% correct). A significant main effect of sentence-type and a sentence-type by talker interaction were also obtained. These effects did not interact with condition and therefore were not interpretable in relation to the tactile information. This pattern of results provides evidence of a benefit from the tactile information for making intonation judgments. However, as in the VT condition of study I reported in Bernstein *et al.* (1989), no evidence was observed for any display differences.

IV. EXPERIMENT II. TACTILE-ALONE JUDGMENTS OF STRESS AND INTONATION

Experiment II was conducted to assess how accurately stress and intonation could be perceived by vibrotaction alone and to determine whether a difference between displays would emerge in the absence of visual stimulation.

A. Stimuli and subjects

The vibrotactile stimuli and the stimulus delivery system were the same as in experiment I. Seven subjects from experiment I were tested in experiment II. An eighth subject, a native-English speaking female, was recruited. This subject performed the experiment I protocol before proceeding to experiment II. Subjects ranged from 20 to 27 years of age. The eight subjects were divided into two groups according to their previously assigned display and performed experiment II with that display.

B. Design and procedure

A block of 48 VT trials was presented to refamiliarize the subjects with the stimuli. They then judged stress and intonation for 25 blocks of TA trials for each talker. Five blocks per talker were presented in each session. The talker was alternated every five blocks, and the order of talker presentation was counterbalanced across subjects. After day 1, subjects were informed of a monetary bonus that could be earned based on day 5 accuracy for stress identification. To earn the bonus, the subjects were required to maintain intonation accuracy while improving stress accuracy.

C. Analyses

Each subject's last five blocks of stress and intonation identification for each talker were analyzed. Separate $2\times2\times2$ (talker×display×sentence-type) repeated measures analyses of variance were performed on the mean percentages of correct responses for intonation and stress. Talker and sentence-type were within-subjects factors, and display was a between-subjects factor.

D. Results and discussion

Stress information was available from the vibrotactile stimuli at levels reliably above chance according to the binomial test (56% for the temporal display and 48% for the spatio-temporal display) (chance=33.3%) (see Table III). TA stress identification accuracy did not differ significantly as a function of displays [$F(1,6)=3.11$, $p=0.13$]. Stress was more accurately identified in discontinuous sentences (59%) than in continuous sentences (45%) [$F(1,6)=82.14$, $p<0.01$], as predicted. In addition, sentence-type interacted with talker [$F(1,6)=26.24$, $p<0.01$], such that performance was better with the female talker only for the continuously voiced sentences (Continuous Male=40% correct, Continuous Female=50% correct, and Discontinuous Male=58% correct, Discontinuous Female=61% correct). Analysis of the results showed that the bonus did not influence performance levels.

TABLE III. Experiment II: Individual subject data in percent correct for *stress* judgments under the tactile-alone condition (Chance=33%). Display=1 indicates the subject used the temporal display; Display=8 indicates that the subject used the spatio-temporal display. * indicates subject did not participate in experiment I.

Subject (Display)	Experiment I subject number	Tactile-alone		
		Continuous	Discontinuous	Combined
1(1)	3	54	69	61
2(1)	*	47	60	53
3(1)	5	40	58	49
4(1)	6	50	71	61
	Mean	48	64	56
5(8)	7	40	54	47
6(8)	8	38	42	40
7(8)	11	48	65	56
8(8)	12	43	56	50
	Mean	42	54	48

Accuracy was comparable to that for stress identification obtained earlier by Rothenberg and Molitor (1979), and by Bernstein *et al.* (1989), who used the same stimuli and a different set of vibrotactile displays based on hand-picked pitch periods. Scores were also roughly comparable to those obtained by Hnath-Chisolm and Kishon-Rabin (1988) (56% correct temporal and 51% correct spatio-temporal).

Although no statistically significant differences between displays were observed for stress judgments, the temporal display tended to convey stress more accurately. This trend was also present in the TA performance reported by Hnath-Chisolm and Kishon-Rabin (1988). However, given the evidence, obtained in experiment I, that emphatic stress tends to be highly visible, this trend was not explored further in subsequent experiments.

Intonation information was available from the vibrotactile stimuli at levels reliably above chance according to the binomial test (73% for the temporal display and 83% for the spatio-temporal display) (chance=50%) (see Table IV). The difference between performance with the temporal and spatio-temporal displays in the current study was not statistically significant [$F(1,6)=2.58$, $p=0.16$]. However, the absence of statistical significance may be a result of the subject sample size.

TABLE IV. Experiment II: Individual subject data in percent correct for *intonation* judgments under the tactile-alone condition (Chance=50%). Display=1 indicates the subject used the temporal display; Display=8 indicates the subject used the spatio-temporal display.* indicates subject did not participate in experiment I.

Subject (Display)	Experiment I subject number	Tactile-alone		
		Continuous	Discontinuous	Combined
1(1)	3	85	92	89
2(1)	*	61	68	65
3(1)	5	68	59	63
4(1)	6	69	79	74
	Mean	71	74	73
5(8)	7	93	84	88
6(8)	8	92	76	84
7(8)	11	85	85	85
8(8)	12	76	74	75
	Mean	86	80	83

Intonation identification accuracy with the temporal display was worse than that reported by Rothenberg and Molitor (1979) (90% correct) and the same as Bernstein *et al.* (1989) (70% correct). However, average performance was better than reported by Hnath-Chisolm and Kishon-Rabin (1988) (64%) for their temporal display. Intonation identification accuracy with the spatio-temporal display (83% correct) was greater than spatio-temporal performance reported by Hnath-Chisolm and Kishon-Rabin (1988) (78% correct).

Examination of the subject data suggested that the accuracy of intonation judgments by several individuals matched or exceeded the best subject's performance (81% correct) in Bernstein *et al.*'s experiment II in which pitch was hand-picked. Thus even with errors introduced by real-time $F0$ extraction, levels of accuracy were achieved with both displays that were comparable to results obtained when the tactile signal was generated off-line for the same sentence tokens (Bernstein *et al.*, 1989).

Individual differences were observed in the ability to use the tactile information (see Table IV). For example, examination of the intonation judgment performance shows that subject 1 performed exceptionally well with the temporal-display, and subject 8 performed somewhat less accurately with the spatio-temporal display. Unfortunately, the between subjects design employed limits our ability to interpret the relationship of the individual differences and display type. Experiment III examined the relationship of individual differences in performance accuracy and display type.

V. EXPERIMENT III. TEMPORAL VERSUS SPATIO-TEMPORAL DISPLAYS FOR INTONATION JUDGMENT

Given the theoretically motivated prediction of an advantage with the spatio-temporal display and the possibility of individual differences in display effectiveness, an additional experiment was conducted to attempt more sensitively to observe differential effects of display. Experiment II provided an indication, although not statistically significant, that the two displays were differentially effective for both stress and intonation judgments. However, the results of experiment I provided evidence that *only* intonation judgments were influenced by the addition of vibrotactile information. Experiment III examined the advantage for the spatio-temporal display for intonation judgments.

Arguably, the complexity of the task and the stimulus set in experiments I and II could result in performance errors that reduced sensitivity. Therefore in experiment III, one talker was presented, the subjects judged intonation only, and testing was under the two conditions employing vibrotactile stimuli, VT and TA. To enhance the power of the design, subjects received both displays in counterbalanced order and repeated measures analysis was performed.

A. Subjects and stimuli

The 21 subjects were age 18–45 years old, with normal hearing, 20/30 or better normal or corrected vision, and English as their native language. Complete data were obtained for 16 of the 21,[5] who were divided into groups of eight and received both displays in counterbalanced order. Subjects received approximately 3 h of experience with each display and had not participated in experiments I or II.

In experiments I and II, two talkers were used to increase the generalizability of the findings (Demorest *et al.*, 1996). However, the use of multiple talkers is known to increase task difficulty (Mullennix *et al.*, 1989). In the current experiment, we chose to use a single-talker to reduce task difficulty. The sentence stimuli were the 48 tokens spoken by the female talker.

B. Design, procedure, and analyses

Stimuli were presented under TA and VT conditions. Subjects made two-alternative forced-choice intonation identifications by pressing one of two buttons (one labeled ''question,'' and one ''statement''). Practice trials were presented on the first day of testing with each of the two displays. For each display, 15 blocks of trials were presented, 5 on each of 3 days. Only the data from the last five blocks of trials for each display were analyzed. For each subject, order of the 48 sentences was independently randomized within each block and condition, and the 2 conditions (VT and TA) were randomly ordered across the 5 blocks presented on a given day with the constraint that no more than 3 blocks were presented in a given condition.

A repeated measures analysis of variance was performed on percent correct intonation identification. The within-subjects factors were display (temporal versus spatio-temporal), condition (TA versus VT), sentence-type (continuous versus discontinuous), and block. The between-subjects factor was display order (temporal first versus spatio-temporal first).

C. Results and discussion

Across VT and TA conditions, intonation identification was more accurate with the spatio-temporal (91% correct) than with the temporal display (75% correct) [$F(1,14)=18.53$, $p<0.01$] (see Table V). Also, VT identification differed significantly from TA identification [$F(1,14)=9.29$, $p<0.01$]. However, a statistically significant display×condition interaction was also obtained [$F(1,14)=10.25$, $p<0.01$]: Spatio-temporal display resulted in the same performance levels across conditions (91% correct VT and TA), whereas performance varied across conditions with the *temporal* display (78% correct VT versus 72% correct TA). These results demonstrated an advantage for the spatio-temporal display under the TA and VT conditions. The advantage observed under VT conditions is important in that it suggests a difference may exist under practical (not just laboratory) conditions.

The increased accuracy with the spatio-temporal display observed in this experiment has two possible sources. First, the two displays differed in the extent to which they had been optimized for this female talker. The spatio-temporal display was optimized specifically for the presentation of the female talker (see Fig. 5). The temporal display was designed to deliver both male and female talkers and was known to be less than optimal for this female talker. Specifically, the female talker's mean $F0$ was 200 Hz, with her mean for the every-other-pulse scheme being 100 Hz. The best frequency resolution for the skin is below 100 Hz, thus the temporal display of the female talker's $F0$ is not optimally suited to the temporal resolution capabilities of the skin. Second, the advantage could be due to the use of the spatial dimension.

Although the current data alone do not allow us to decide between these two possible interpretations, examination of previous performance with tactile transformation of these stimuli can provide some insight. In their Fig. 2, Bernstein *et al.* (1989) present data on temporal displays optimized for the presentation of the same stimuli spoken by the female talker as were used in the current experiment. Average TA performance levels with the best of these temporal displays was comparable (74% correct) to the current TA performance with a temporal display (72% correct). Thus the increase in performance with the spatio-temporal display was likely due to the use of the spatial dimension and not the optimization of the display for a specific talker.

Examination of the individual subject data (see Table V) suggests that performance level varied more for the temporal display than for the spatio-temporal display. Within-subject performance levels (in percent correct) were not significantly correlated across displays for either condition (TA: $r=0.377$ and VT: $r=0.398$). The pattern of results suggests an advantage for spatio-temporal display in the stability of performance across individuals. That is, most individuals could successfully use the spatio-temporal display, whereas only some of those individuals could successfully use the temporal display.

TABLE V. Experiment III: Individual subject data in percent correct for *intonation* judgments under the visual-tactile and tactile-alone condition (Chance=50%).

	Temporal		Spatio-temporal	
Subject	Tactile-alone	Visual-tactile	Tactile-alone	Visual-tactile
1	66	65	95	94
2	55	68	89	90
3	85	91	98	95
4	93	88	83	81
5	88	90	85	92
6	65	70	93	98
7	45	53	89	78
8	87	94	93	98
9	68	75	92	94
10	83	86	95	95
11	69	78	95	95
12	45	57	72	63
13	44	65	90	97
14	84	87	95	98
15	95	97	97	96
16	83	90	88	86
Mean	72	78	90	91

TABLE VI. Display-type, age of onset in years, three-frequency (0.5 kHz, 1 kHz, 2 kHz) pure-tone averages (dB HL) and speechreading ability of sentences in percent words correct for subjects in experiment IV. NMH=no measurable hearing.

Subject (onset)	Display	Age of onset of hearing loss	Left ear	Right ear	Speechreading of sentences
1(Post)	Temporal	4	NMH	NMH	67
2(Post)	Temporal	4	NMH	NMH	37
3(Post)	Spatio-Temporal	3	110	105	61
4(Post)	Spatio-Temporal	15	103	95	28
5(Pre)	Temporal	Birth	83	92	67
6(Pre)	Temporal	Birth	87	100	72
7(Pre)	Spatio-Temporal	Birth	115	110	62
8(Pre)	Spatio-Temporal	Birth	100	105	71

VI. EXPERIMENT IV. VISUAL-ALONE VERSUS VISUAL-TACTILE JUDGMENTS OF STRESS AND INTONATION BY HEARING-IMPAIRED INDIVIDUALS

The goal of sensory substitution is to develop vibrotactile speech perception aids for hearing-impaired individuals. Having observed several hearing-impaired subjects who were less successful identifying vibrotactile stress and intonation than hearing subjects, Rothenberg and Molitor (1979) suggested that vibrotactile pitch perception could be based on experience with auditory pitch. Alternatively, the ability to judge stress and intonation could be based on experience with spoken language. That is, individuals with reduced experience perceiving spoken language, might have more difficulty with identification of stress and intonation regardless of the form of stimulation. Experiment IV, which was a modified version of experiment I, employed pre- and postlingually hearing-impaired adult subjects. In addition to testing stress and intonation judgments, a test was also given to estimate the speechreading ability of our subjects with early ages-of-onset of hearing impairment.

A. Subjects

The subjects were eight hearing-impaired adults, age 19–30 years, four with pre- and four with post-lingual hearing impairments. They were all from the Gallaudet University community. Table VI shows the three-frequency pure-tone averages (dB HL) for each of the subjects, their age at onset of hearing impairment, the display they received and their accuracy for speechreading words in sentences. All subjects reported English as their native language. They were paid for their participation.

B. Stimuli, design, and procedure

The stimuli and task in experiment IV were the same as in experiment I. The sequence of conditions was held constant across subjects. VT stimuli were used to help explain and give initial practice. It was hypothesized that the vibrotactile stimuli would aid in conveying the concept of stress and intonation, in the event that the subjects were unfamiliar with these linguistic characteristics.

Subjects were given written instructions explaining the task followed by a spoken explanation with manual sign accompaniment. The experimenter presented the subjects with examples of the different types of intonation patterns and stress patterns using live voice (and no manual signs), which was processed and transformed into vibrotactile stimulation in real-time. It was verified that the subjects had a basic understanding of the task. Subjects were then presented with practice consisting of 48 VT trials and 48 VA trials. Each trial was followed by feedback. Then subjects identified stress and intonation in 4 blocks of 48 sentences in the VA condition followed by 30 blocks of VT trials, followed by 20 blocks of VA trials. Half of the blocks were produced by the female talker and half by the male talker, with order of talker counterbalanced across subjects. The total number of trials in each condition was equivalent to that in experiment I. At the end of testing, each subject was presented with a set of 100 prerecorded sentences spoken by a different male talker to assess speechreading ability. Sentences were presented one at a time under computer control and after each sentence subjects typed what they thought the talker said.

C. Analyses

Separate $2\times2\times2\times2\times2$ (condition×talker×sentence-type×pre-post×display) repeated measures analyses of variance were performed on the mean percentages of correct responses for intonation and stress, as in experiment I, except that pre–post corresponded to whether the subjects had pre- or post-lingual hearing impairments. Again, only the last five blocks of data within the VA and VT conditions were analyzed.

D. Results and discussion

VA *stress* identification was reliably above chance for both talkers (80% correct for the female talker and 84% correct for the male talker) according to the binomial test (chance=33.3%) (see Table VII). VA *intonation* identification was reliably above chance for the male talker (66% correct) but not for the female talker (56% correct), according to the binomial test (chance=50%) (see Table VIII). The accuracy of speechreading sentences (see Table VI) is consistent with our previous studies assessing speechreading ability (Bernstein *et al.*, 1998; Auer, 1997). Specifically, individuals with early onset hearing impairments (subjects 1, 3, 5, 6, 7, 8) frequently far outperform individuals with late-onset hearing impairments (e.g., subject 4) or normal hearing.

TABLE VII. Experiment IV: Individual subject data in percent correct for *stress* judgments under visual-alone and visual-tactile conditions (Chance=33%). Display=1 indicates the subject used the temporal display; Display=8 indicates that the subject used the spatio-temporal display. Post indicates the subject had a post-lingual hearing impairment. Pre indicates the subject had a pre-lingual hearing impairment.

Subject (Display) (Onset)	Visual-alone			Visual-tactile		
	Cont.	Discont.	Combined	Cont.	Discont.	Combined
1(1)(Post)	87	82	85	75	73	74
2(1)(Post)	88	75	82	77	75	76
3(8)(Post)	95	88	92	98	88	93
4(8)(Post)	71	62	66	63	60	62
5(1)(Pre)	91	87	89	89	82	85
6(1)(Pre)	70	63	66	74	65	69
7(8)(Pre)	93	84	89	91	83	87
8(8)(Pre)	89	83	86	85	75	80
Mean	86	78	82	81	75	78

The results of experiment IV were essentially consistent with those of experiment I. The analysis of variance for *stress* identification revealed statistically significant effects of talker [$F(1,4)=15.18$, $p<0.02$] and sentence type [$F(1,4)=109.44$, $p<0.01$]. As in experiment I, stress identification was easier with the male talker and easier for continuous sentences (see Table VII). Furthermore, as in experiment I, no significant effects or interactions related to the variables of interest were revealed.

Intonation identification accuracy was enhanced by the addition of the vibrotactile $F0$ information (61% correct VA and 71% correct VT) [$F(1,4)=9.68$, $p<0.04$]. Consistent with the results of experiment I, intonation identification was easier for continuous sentences [$F(1,4)=46.07$, $p<0.01$]. However, this effect did not interact with condition and was not interpretable in relation to the tactile information. No statistically significant effects of display or pre- versus post-lingual hearing impairment were observed. Thus benefit for intonation judgment accuracy obtained with the addition of the tactile information does not appear related to the age-of-onset of the hearing impairment or the type of vibrotactile display used.

The similarity of the results for individuals with pre- and post-lingual onset of hearing impairment does not support the suggestion that vibrotactile pitch perception may be based on auditory pitch perception experience (Rothenberg *et al.*, 1977). However, experiential differences between these two populations were relevant to VA intonation judgment accuracy. The post-lingual group was better able to make intonation judgments from visual speech information [$t(6)=4.197$, $p<0.01$; Pre-lingual=56% correct, Post-lingual=66% correct]. Interestingly, this difference in VA intonation judgment accuracy was not related to speechreading ability or the magnitude of enhancement observed with the addition of tactile information.

VII. GENERAL DISCUSSION

This study was conducted within the context of a project to develop a new, wearable vibrotactile speech perception aid. The main questions with practical implications were: (1) whether one of the two different vibrotactile displays of $F0$ was more successful for conveying $F0$; and (2) whether auditory speech perception experience was a factor in the successful perception of $F0$ information.

The current studies demonstrated an advantage in intonation judgment accuracy for a spatio-temporal display over a temporal display. The advantage was observed under both VT and TA presentation conditions, but was reduced in magnitude in the VT condition. Furthermore, the advantage was

TABLE VIII. Experiment IV: Individual subject data in percent correct for *intonation* judgments under visual-alone and visual-tactile conditions (Chance=50%). Display=1 indicates the subject used the temporal display; Display=8 indicates the subject used the spatio-temporal display. Post indicates the subject had a post-lingual hearing impairment. Pre indicates the subject had a pre-lingual hearing impairment.

Subject (Display) (Onset)	Visual-alone			Visual-tactile		
	Cont.	Discont.	Combined	Cont.	Discont.	Combined
1(1)(Post)	72	63	67	70	69	70
2(1)(Post)	64	59	62	80	77	79
3(8)(Post)	73	66	69	90	77	84
4(8)(Post)	70	61	65	81	71	76
5(1)(Pre)	66	53	59	75	68	72
6(1)(Pre)	52	50	51	57	53	55
7(8)(Pre)	60	57	58	83	76	79
8(8)(Pre)	63	49	56	57	56	56
Mean	65	57	61	74	68	71

observed in the context of a temporal display that also provided a significant enhancement over the visually available information. Thus the current results appear to provide stronger evidence in favor of a spatio-temporal display than was reported in Hnath-Chisolm and Kishon-Rabin (1988). Interestingly, evidence was observed that performance was more stable with the spatio-temporal display across individual subjects. Thus if the goal of the speech perception aid is to accurately convey sentential intonation contours, then the current study supports the use of a spatio-temporal display. However, this conclusion may not hold for other linguistic levels at which $F0$ has been shown to be a factor.

A coarse level comparison of previous perceptual and linguistic experience was possible by comparing the performance of the normal-hearing subjects in experiment I and hearing-impaired subjects in experiment IV. The pattern of results for *stress* identification did not appear to differ as a function of hearing group nor the pre–post lingual onset distinction. It is interesting to note that because hearing-impaired individuals rely on visual information for speech communication, more accurate stress identification might have been predicted on their part.

VA *intonation* identification accuracy was comparable across subject groups (64% correct for normal-hearing and 61% correct for hearing-impaired subjects). VT *intonation* identification accuracy was higher for the hearing subjects (82% for normal-hearing subjects and 71% for hearing-impaired subjects).[6] This pattern of results is consistent with the suggestion that vibrotactile pitch perception may be based on auditory pitch perception experience (Rothenberg *et al.*, 1997). However, this conclusion was not supported by the comparison of pre- versus post-lingually hearing-impaired individuals (see discussion of experiment IV). Thus although it is clear that hearing-impaired subjects received less benefit than the normal-hearing subjects, determination of the nature of the experiential factors responsible for this difference between hearing groups awaits further study (e.g., Bernstein *et al.*, 1998).

VIII. CONCLUSION

The results of the present study clearly support the use of a spatio-temporal display, if the goal is to accurately convey sentential intonation contours. Furthermore, enhancements in performance over VA performance are related to previous perceptual/linguistic experience, with normal-hearing individuals obtaining the largest enhancements. However, the issues raised in Sec. VII point to the need for additional studies investigating both of these conclusions. Direct comparison of alternate display schemes, holding signal processing constant, is relatively rare (cf., Eberhardt *et al.*, 1990; Hnath-Chisolm and Kishon-Rabin, 1988; Rothenberg and Molitor, 1979). When such studies have been conducted, they have provided insights into vibrotactile speech perception and have shown that the vibrotactile system can be sensitive to how the vibrotactile stimulus is constructed. These demonstrations encourage us to believe that systematic research and engineering methods can lead to speech perception aids that optimally engage the vibrotactile system.

ACKNOWLEDGMENTS

E. T. Auer, Jr. and L. E. Bernstein collected the majority of the data for these studies while employed at Gallaudet University. The authors thank Drs. Marilyn E. Demorest and Robin S. Waldstein, Paula E. Tucker, and two anonymous reviewers for their comments. They thank Dr. Marilyn E. Demorest for her advice on the design and analysis of experiment III. They also acknowledge the very capable assistance of Paula E. Tucker, Jennifer Johnson, and Deborah Yakel. This study was supported by a grant from NIH (DC00695). Correspondence concerning this article should be addressed to Edward T. Auer, Jr., Spoken Language Processes Laboratory, House Ear Institute, Fifth Floor, 2100 West Third Street, Los Angeles, CA 90057. Electronic mail for Edward T. Auer, Jr. may be sent via Internet to eauer@hei.org.

[1]In Hnath-Chisolm and Kishon-Rabin (1988) two different figures are given for the frequency change that corresponds to a change in vibration location, 0.14 and 0.16 octaves.

[2]In Hnath-Chisolm and Kishon-Rabin (1988) percent correct was reported after correction for guessing. The values reported here are not corrected for guessing in order to facilitate comparison of results between studies.

[3]Spatial resolution on the hypothenar is less than that on the fingerpad (Cholewiak and Collins, 1991).

[4]Using the methods described in Kirk (1968) (pp. 63–67), it was determined for experiments I and II that a data transformation was not required.

[5]Of the five who did not provide complete data, two performed a portion of the training without any masking noise, one was provided with an erroneous display, and two had scheduling difficulties.

[6]The order of the VA and VT conditions was counterbalanced for normal-hearing subjects whereas hearing-impaired subjects always received the VA condition second. Therefore, a simple practice effect could explain the difference between the two groups. However, comparison of the hearing-impaired subjects' performance with only those normal-hearing subjects who received the VA condition second resulted in the same difference between the populations.

Auer, Jr., E. T. (**1997**). ''The scope of individual differences in cognitive models of spoken language understanding,'' J. Acoust. Soc. Am. **102**, 3115A.

Bernstein, L. E., Demorest, M. E., and Tucker, P. E. (**1998**). ''Speech perception without hearing,'' Manuscript submitted for publication.

Bernstein, L. E., and Eberhardt, S. P. (**1986**). Johns Hopkins lipreading corpus I–II: Disc 1 [videodisc] (The Johns Hopkins University, Baltimore, MD).

Bernstein, L. E., Eberhardt, S. P., and Demorest, M. E. (**1989**). ''Single-channel vibrotactile supplements to visual perception of intonation and stress,'' J. Acoust. Soc. Am. **85**, 397–405.

Bernstein, L. E., Tucker, P. E., and Auer, Jr., E. T. (**1998**). ''Potential perceptual bases for successful use of a vibrotactile speech perception aid,'' Scand. J. Psych. **39**, 181–186.

Bolinger, D. L. (**1958**). ''A theory of pitch accent in English,'' Word **14**, 109–149.

Bolinger, D. L. (**1978**). ''Intonation across languages,'' in *Universals of Human Language: Volume 2. Phonology*, edited by J. P. Greenberg, C. A. Ferguson, and E. A. Moravcsik (Stanford U.P., Stanford, CA).

Boothroyd, A., and Hnath, T. (**1986**). ''Tactile supplements to speechreading,'' J. Rehab. Res. Develop. **23**, 139–146.

Boothroyd, A., Hnath-Chisolm, T., Hanin, L., and Kishon-Rabin, L. (**1988**).

''Voice fundamental frequency as an auditory supplement to the speechreading of sentences,'' Ear Hear. **9**, 306–312.
Bosman, A. J., and Smoorenburg, G. F. (**1997**). ''Evaluation of three pitch tracking algorithms at several signal-to-noise ratios,'' Acustica Acta Acust. **83**, 567–571.
Breeuwer, M., and Plomp, R. (**1984**). ''Speechreading supplemented with frequency-selective sound-pressure information,'' J. Acoust. Soc. Am. **76**, 686–691.
Breeuwer, M., and Plomp, R. (**1985**). ''Speechreading supplemented with formant-frequency information from voiced speech,'' J. Acoust. Soc. Am. **77**, 314–317.
Breeuwer, M., and Plomp, R. (**1986**). ''Speechreading supplemented with auditorily presented speech parameters,'' J. Acoust. Soc. Am. **79**, 481–499.
Cholewiak, R. W., and Collins, A. A. (**1991**). ''Sensory and physiological bases of touch,'' in *The Psychology of Touch*, edited by M. A. Heller and W. Schiff (Erlbaum, Hillsdale, NJ).
Cutler, A., and Butterfield, S. (**1992**). ''Rhythmic cues to speech segmentation: Evidence from juncture misperception,'' J. Mem. Lang. **31**, 218–236.
Demorest, M. E., and Bernstein, L. E. (**1987**). ''Reliability of individual differences in lipreading,'' J. Acoust. Soc. Am. **82**, S24.
Demorest, M. E., Bernstein, L. E., and De Haven, G. P. (**1996**). ''Generalizability of speechreading performance on nonsense syllables, words, and sentences: Subjects with normal hearing,'' J. Speech Hear. Res. **39**, 697–713.
Eberhardt, S. P., Bernstein, L. E., Demorest, M. E., and Goldstein, M. H. (**1990**). ''Lipreading sentences with single-channel vibrotactile transformations of voice fundamental frequency,'' J. Acoust. Soc. Am. **88**, 1274–1285.
Foulke, E. (**1991**). ''Braille,'' in *The Psychology of Touch*, edited by M. A. Heller and W. Schiff (Erlbaum, Hillsdale, NJ).
Fourcin, A. J. (**1981**). ''Laryngographic assessment of phonatory function,'' Am. Speech Hear. Assoc. Rep. **11**, 116–127.
Fry, D. B. (**1955**). ''Duration and intensity as acoustic correlates of linguistic stress,'' J. Acoust. Soc. Am. **35**, 765–769.
Fry, D. B. (**1958**). ''Experiments on the perception of stress,'' Lang. Speech **1**, 126–152.
Geldard, F. (**1985**). ''The mutability of time and space on the skin,'' J. Acoust. Soc. Am. **77**, 233–237.
Geldard, F., and Sherrick, C. (**1972**). ''The cutaneous rabbit: A perceptual illusion,'' Science **178**, 178–179.
Grant, K. W. (**1987**). ''Encoding voice pitch for hearing-impaired listeners,'' J. Acoust. Soc. Am. **82**, 423–432.
Grant, K. W., Ardell, L. H., Kuhl, P. K., and Sparks, D. W. (**1986**). ''The transmission of prosodic information via an electrotactile speechreading aid,'' Ear Hear. **7**, 328–335.
Haggard, M., Ambler, S., and Callow, M. (**1970**). ''Pitch as a voicing cue,'' J. Acoust. Soc. Am. **47**, 613–617.
Hanin, L., Boothroyd, A., and Hnath-Chisolm, T. (**1988**). ''Tactile presentation of voice fundamental frequency as an aid to speechreading of sentences,'' Ear Hear. **9**, 329–334.
Hermes, D. J., and van Gestel, J. C. (**1991**). ''The frequency scale of speech intonation,'' J. Acoust. Soc. Am. **90**, 97–102.
Hess, W. (**1983**). *Pitch Determination of Speech Signals* (Springer-Verlag, New York).
Hnath-Chisolm, T., and Boothroyd, A. (**1992**). ''Speechreading enhancement by voice fundamental frequency: The effects of contour distortions,'' J. Speech Hear. Res. **35**, 1160–1168.
Hnath-Chisolm, T., and Kishon-Rabin, L. (**1988**). ''Tactile presentation of voice fundamental frequency as an aid to the perception of speech pattern contrasts,'' Ear Hear. **9**, 329–334.
Hnath-Chisolm, T., and Medwetsky, L. (**1988**). ''Perception of frequency contours via temporal and spatial tactile transforms,'' Ear Hear. **9**, 322–328.
House, A. S., and Fairbanks, G. (**1953**). ''The influence of consonant environment upon the secondary acoustical characteristics of vowels,'' J. Acoust. Soc. Am. **25**, 105–113.
Kirk, R. E. (**1968**). *Experimental Design Procedures for the Behavioral Sciences* (Wadsworth, Belmont, CA).
Kirman, J. H. (**1973**). ''Tactile communication of speech: A review and analysis,'' Psychol. Bull. **80**, 54–74.
Kishon-Rabin, L., Boothroyd, A., and Hanin, L. (**1996**). ''Speechreading enhancement: A comparison of spatial-tactile display of voice fundamental frequency ($F0$) with auditory $F0$,'' J. Acoust. Soc. Am. **100**, 593–602.
Lehiste, I. (**1970**). *Suprasegmentals* (MIT Press, Cambridge, MA).
Lisker, L., and Abramson, A. S. (**1967**). ''The voicing dimension: Some experiments in comparative phonetics,'' in *Proceedings of the Sixth International Congress of Phonetic Sciences*, Prague (Academia, Prague), pp. 563–567.
McGrath, M., and Summerfield, Q. (**1985**). ''Intermodal timing relations and audiovisual speech recognition by normal-hearing adults,'' J. Acoust. Soc. Am. **77**, 678–685.
Mullennix, J. W., Pisoni, D. B., and Martin, C. S. (**1989**). ''Some effects of talker variability on spoken word recognition,'' J. Acoust. Soc. Am. **85**, 365–378.
Phillips, J. R., and Johnson, K. O. (**1985**). ''Neural mechanisms of scanned and stationary touch,'' J. Acoust. Soc. Am. **77**, 220–224.
Pike, K. (**1945**). *The Intonation of American English* (University of Michigan Press, Ann Arbor, MI).
Plant, G., and Risberg, A. (**1983**). ''The transmission of fundamental frequency variations via a single channel vibrotactile aid,'' Stockholm, Sweden, Speech Transmission Laboratory, Quarterly Report, **2–3**, 61–84.
Reed, C. M., Delhorne, L. A., and Durlach, N. I. (**1993**). ''Historical overview in tactile aid research,'' in *Proceedings of the Second International Conference on Tactile Aids, Hearing Aids, and Cochlear Implants*, edited by A. Risberg, S. Felicetti, and K.-E. Spens (Akademitryck AB, Edsbruk, Sweden).
Rosen, S. M., Fourcin, A. J., and Moore, B. C. J. (**1981**). ''Voice pitch as an aid to lipreading,'' Nature (London) **291**, 150–152.
Rothenberg, M., and Molitor, R. D. (**1979**). ''Encoding voice fundamental frequency as vibrotactile frequency,'' J. Acoust. Soc. Am. **66**, 1029–1038.
Rothenberg, M., Verrillo, R. T., Zahorian, S. A., Brachman, M. L., and Bolanowski, S. J. (**1977**). ''Vibrotactile frequency for encoding a speech parameter,'' J. Acoust. Soc. Am. **62**, 1003–1012.
Sluijter, A. M. C., and van Heuven, V. J. (**1996**). ''Spectral balance as an acoustic correlate of linguistic stress,'' J. Acoust. Soc. Am. **100**, 2471–2485.
Streeter, L. A. (**1978**). ''Acoustic determinants of phrase boundary perception,'' J. Acoust. Soc. Am. **64**, 1582–1592.
Summers, I. R. (**1992**). ''Signal processing strategies for single channel aids,'' in *Tactile Aids for the Hearing Impaired*, edited by I. R. Summers (Whurr, London).
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 displays of voice fundamental frequency,'' J. Speech Hear. Res. **38**, 690–705.
Yeung, E., Boothroyd, A., and Redmond, C. (**1988**). ''A wearable multichannel tactile display of voice fundamental frequency,'' Ear Hear. **9**, 342–350.