The transmission of fundamental frequency variations via a single channel vibrotactile aid

Plant, G. and Risberg, A.

journal: STL-QPSR
volume: 24
number: 2-3
year: 1983
pages: 061-084

http://www.speech.kth.se/qpsr
**III. SPEECH AND HEARING DEFECTS AND AIDS**

**A. THE TRANSMISSION OF FUNDAMENTAL FREQUENCY VARIATIONS VIA A SINGLE CHANNEL VIBROTACTILE AID**

Geoff Plant* and Arne Risberg

**Abstract**

A series of experiments were carried out to evaluate a single channel vibrotactile aid which extracts a speaker's fundamental frequency and presents it in the range where the skin is maximally sensitive to frequency change. The aid's ability to convey emphatic stress in sentences and word syllable number and type was tested using both English and Swedish materials. The results indicate that the aid transmits such information at a high level although it appears that extensive training would be needed with some tasks before optimal performance is attained. The information provided by this aid should assist deaf persons in both speech perception and production. The significance of the results obtained and the direction of future research efforts are discussed.

**Introduction**

Early research into the use of tactile aids with the deaf was aimed at using "touch as a substitute for hearing in the interpretation and control of speech" (Gault, 1926; p. 121). This aim, however, has not as yet been realized and the research emphasis has shifted to the development of tactile aids which provide supplementary information to assist lipreading and provide some feedback as to speech production. A number of factors have been cited to account for the lack of success in the development of a tactile substitute for hearing. Liberman et al (1968) hypothesized that speech perception requires a special decoder which is biologically tied to the auditory system. Display schemes which present a visual or tactual transform of the speech signal do not have access to this decoder and, as a result, can provide only limited speech perceptual skills. This view has been challenged by Cole et al (1980) who examined the ability of a skilled spectrogram reader with more than 2,000 hours experience. They found that he was able to correctly identify about 85% of all phonetic segments from the visual information available in spectrograms, and concluded that "speech can be interpreted in a modality other than audition" (Cole et al, 1980; p. 45). Additional counter evidence to Liberman et al's (1968) hypothesis is found in experimental studies of deaf-blind persons using Tadoma (Norton et al, 1977; Reed et al, 1981), a manual method of lipreading "in which the listener receives speech information by placing his hand on the talker's face and monitoring the articulation process" (Durlach et al, 1977; p.

* from National Acoustics Laboratory, Millers Point, Australia. Guest researcher at KTH from February - July, 1983.
59). This method, which has been used by a number of deaf-blind persons to acquire expressive and receptive speech and language skills, offers compelling evidence that the tactile system can be used for speech perception and indicates that it may be possible to develop an aid which provides full speech comprehension via the tactile system. Kirman (1974) in a review of tactile research concluded that "neither the results of past work on tactile displays, nor contemporary theories of speech perception, nor psychophysical studies of tactile perception have provided reasonable grounds for believing that the skin lacks the capacity to comprehend a suitable display of speech" (Kirman, 1974; p. 71). Rather, he believes that the major factor contributing to the relative lack of success with tactile aids was that the coding techniques selected were not appropriate for the tactual sense.

This view has been given partial support by later studies such as those of Sparks et al (1978; 1979), Scilley (1980) and Green et al (1983) which indicate that much speech information can be interpreted via the tactile sense alone. Other variables which have most probably influenced results obtained with tactile systems include the amount and type of training provided and the hearing status of the experimental subjects. For example, many studies have used normal hearing subjects artificially deafened for the duration of the experiment. The training times provided are necessarily limited and the motivation of the subjects is probably not as great as would be found with deaf persons. As a result work may have been discontinued on many potentially useful aids because the time allocated for training was not sufficient to allow for optimal performance. Research efforts aimed at the development of a tactile aid have also been motivated by the need to provide deaf persons with a wearable aid which can be used at all times. Most aids developed as tactile substitutes for audition are extremely bulky and many researchers feel that it is essential to develop an interim tactile aid which will provide at least some assistance for deaf persons until more complex systems have been developed and miniaturized for everyday use.

An important consideration in the development of any tactile aid is the selection of the speech parameters to be displayed. One feature which appears to be of great importance in both speech production and perception is fundamental frequency (F0) variations. Studies by Risberg (1974), Risberg and Lubker (1978) and Risberg and Agelfors (1978) have all shown that great improvements in lipreading ability resulted when the visual signal was supplemented by auditorily presented F0 information.
Ardell (1980) investigated the contribution of F0 information as a supplement to lipreading using the tracking technique developed by De Filippo and Scott (1978). She found that by lipreading alone normal hearing persons were able to track speech at only 38% of the rate achieved by hearing subjects when both auditory and visual information were fully available. When lipreading was supplemented by auditorily presented time, intensity and F0 changes, the score obtained rose to 78% of the normal rate. Grant (1980) replicated this study using a tactile aid consisting of 10 electrodes placed 2 cm apart on the forearm to encode F0 variations. He found that, although it took more time to learn how to use the tactile information, the final tracking rate obtained was about the same as that reported by Ardell (1980). He concluded that the "tactile aid appeared to be as useful to speechreaders as were the auditory cues" (Grant, 1980; p. 91).

Attempts to present F0 variations via a single vibrator have been less successful due, presumably, to the poor frequency resolution of the tactile sense. Rothenberg and Molitor (1979) attempted to overcome this limitation by recoding F0 information into the frequency region where the skin appears to be maximally sensitive to frequency change. An earlier study by Rothenberg et al (1977) had found that there are perhaps 10 discriminable steps in vibrotactile frequency from 10 to 100 Hz and the processing scheme adopted by Rothenberg and Molitor (1979) aimed at presenting F0 variations within this range. The task presented to their subjects was both the identification of the stressed word in the sentence "Bev loves Bob" plus a decision as to whether the sentence presented was a statement or a question. The scores obtained when the range was expanded by a factor of 2 or shifted down to a centre frequency of 50 Hz, were around 50% correct rising to 65% with training. This result indicates that the use of a single vibrator to present F0 variations does appear to be a feasible approach and, as Rothenberg and Molitor (1979) conclude, better scores may be attained if longer training times are provided. Such an aid offers the possibility of serving as a very useful lipreading aid in the short term and may also become part of a more complex aid using a spatial array to encode spectral features.

An aid which presents F0 variations should also be a useful speech training device for use with both congenitally and adventitiously deaf subjects. Studies of the speech of the congenitally deaf (see Osberger and McGarr, 1982 for a review of published research) have shown that the
average range of pitch variations is "more reduced than those of speakers with normal hearing" (Osberger and McGarr, 1982; p. 249). Similarly, the speech of persons with an acquired profound hearing loss is characterised by a restricted range of F0 variations and is often described as being "monotonous" and lacking in appropriate intonational movements (Silverman and Calvert, 1978; Plant, 1983). Aids which present F0 information visually (Mártony, 1968; Abberton et al, 1978) or tactually (Stratton, 1974) have been shown to result in improved intonational control by deaf speakers.

The research results summarized above indicate that a vibrotactile aid presenting F0 variations may be a useful means of providing profoundly deaf persons with information which will assist both speech production and perception. This current paper reports preliminary results obtained with a single-channel vibrotactile aid designed to transmit F0 information.

**Equipment**

A block diagram of the experimental equipment is presented in Fig. 1. The speaker's F0 was extracted using a conventional pitch extractor consisting of a low pass filter set at 150 Hz followed by an infinite clipper. The positive going zero-crossings from the clipper triggered a monostable multivibrator, and the pulses from this were integrated in a low pass filter set at 32 Hz (slope = 18 dB/octave). The signal from the low pass filter was then a DC-voltage proportional to the speaker's F0. This voltage controlled the frequency of a signal-generator (Wave-tec Model 114). By means of the controls on the signal generator, the duration of the pulses from the monostable multivibrator, the setting of the potentiometer R3 and finally the germanium diode D it was possible to achieve the required relationship between the speaker's pitch range and the output frequency. Fig. 2 shows the adjustment used in Experiment 4 with speaker AR.

The output signal from the signal generator was controlled by the speech cutoff frequency in the frequency range 68-1500 Hz. The high pass cutoff frequency was selected to filter out some low frequency disturbances on the tape with the low pass filter frequency selected to avoid fricative sounds activating the vibrator. The signal from the band pass filter was rectified and then smoothed in a low pass filter with a cutoff frequency of 32 Hz (slope = 18 dB/octave). The modulation was linear, but by the use of the rectifier a linear or logarithmic setting
Fig. 1. Block diagram of the experimental equipment used in the study.
Fig. 2. Relationship between the input and output frequencies for the equipment as used in Experiment 4 for speaker AR.

Fig. 3. Intensity input/output characteristics of the experimental equipment for the linear and logarithmic settings.
could be selected. The resulting input-output characteristics for the 2 settings are shown in Fig. 3. The signal from the modulator was fed to a power amplifier which allowed the subject to vary the intensity of the vibrator. The frequency response of the amplifier was adjusted so that a sweep in the input frequency of the total device between 100 and 200Hz did not result in any apparent change of "loudness" of the vibration.

The output transducer used was that developed for the Sentiphone by Traumüller (1980). This vibrator consists of a small DC-motor with the stator fixed. When the rotor is fed with an AC-voltage it makes a backward and forward movement. The diameter of the vibrator is 32 mm and it was designed to be held in the hand.

**Experiment 1**

**Materials**

The aim of this experiment was to assess the ability of untrained observers to detect the stressed word in 3- and 4-word English sentences. Five test lists were drawn up presenting the following sentences: List 1, "We will win"; List 2, "They run well"; List 3, "Bev loves Bob"; List 4, "Their room was warm"; List 5, "We love warm bread". Each of the lists consisted of a number of repetitions of the test sentence with randomized placement of stress on one of the words. For the 3-word sentences each word was stressed 6 times in a random order for a total of 18 test items. The 4-word sentences involved random stress being placed on each of the words on 5 occasions to form lists consisting of 20 test items. Practice items were also provided to familiarize the subjects with the test alternatives. These presented repetitions of the test sentences with each of the words stressed in turn. The test lists were recorded by one of the authors (GP), a native speaker of Australian English. The speaker used a normal rate of articulation and attempted to produce the stress at a level only slightly stronger than that found in everyday speech. The recordings were made using a 2-channel Revox A77 tape recorder with inputs from an air microphone and an accelerometer attached to the speaker’s throat. The recordings were analyzed using the program developed by Ternström (1983) which plots F0 contours as a function of time. Fig. 4 presents the F0 contour associated with the 3 stress patterns for the sentence "Bev loves Bob". It can be seen that emphatic stress is marked by an F0 peak on the emphasized word. This is consistent with reported data for
Fig. 4. F0 contours for two presentations of the sentence "Bev loves Bob" spoken by GP with emphasis placed on "Bev" (a), "loves" (b), and "Bob" (c).
The analysis revealed that the sentences were produced with a mean F0 of
125 Hz ranging from 90-180 Hz.

Procedure
The settings on the vibrotactile aid were adjusted so that the
speaker's F0 range (90-180 Hz) was presented in the frequency range 30-
100 Hz with the intensity variations in the speech signal also availa-
ble. The materials were presented to 10 normal-hearing non-native
speakers of English. Although the vibrator used in this study has very
little radiated sound, the subjects wore headphones with masking noise
set at a level at which they were confident that they could not hear the
vibrator. The subjects were given scoring sheets on which were presented
the test sentences, and were asked to ring the emphasized word after
each presentation. Prior to the presentation of the test the subjects
received 2 repetitions of each of the stress patterns in order to familiarize them with the test stimuli. The subjects were divided into 2
groups of 5. Group 1 were presented with lists 1, 2 and 4 while Group 2
received lists 1, 3 and 5. To ensure that the individual subjects
understood and were able to complete the task successfully List 1 was
presented auditorily at the subjects' preferred listening level. The
remaining lists were presented via the hand-held vibrator set to the
subject's preferred level.

Results and Discussion
The results obtained by the 2 groups are presented in Table I. All
subjects scored 100% for the auditory condition with the mean scores for
the tactually presented lists being 81%, 80%, 86% and 64% respectively.
These scores are all well above that which could be reasonably attribu-
ted to chance, and indicate that the aid presents relatively unambig-
uous cues to emphatic stress in English sentences. The only discrepency
in the test results is the large differences in the scores obtained by
the 2 groups for the 4 word sentences. Examination of the F0 plots for
the 2 sentences used (Fig. 5) revealed that the stress patterns for the
sentence used in List 5 are less marked than for the sentence used in
List 4. This is especially true when stress is placed on either the
third or forth word of the sentences. A second factor which may have
contributed to the differences between the 2 groups was the performance
of Subject 4 in Group 2 who scored at a chance level for the 2 lists
presented tactually.
The scores obtained in this experiment compare favorably with other reported studies. Plant (1982) presented 4 experienced users of vibrotactile aids with 4 word sentences and asked them to identify the stressed word. The mean score obtained for this task was 60% but an analysis of the subjects' responses indicated that although stress on either the first or second word was correctly identified on 85% of presentations the score obtained when stress was placed on either the third or fourth words was only 35%. An examination of the F0 contours presented in Figs. 4 and 5 gives some insights into the factors leading to this result. It can be seen that initial stress in sentences is marked by a
Fig. 5. F0 contours for the sentences:
A. "Their room was warm".
B. "We love warm bread".
Emphatic stress is placed in turn on the first (a), second (b), third (c) and fourth (d) words.
far more prominent FO excursion than that which accompanies final stress. Given the very poor frequency discrimination ability of the skin above 100 Hz (Rothenberg et al, 1977) it is not surprising that only relatively large changes can be detected. This is especially so when it is considered that the speaker in Plant’s (1982) study was a female with a mean FO of 180 Hz. The results obtained in this experiment are also much better than those obtained by Rothenberg and Molitor (1979) despite the use of a similar processing scheme in both studies. A number of factors probably account for this difference. The most obvious is that the vibrator was hand-held in the present study whereas it was positioned on the forearm in that of Rothenberg and Molitor (1979). Rothenberg et al (1977) concluded that there were probably 7 discriminable steps on the forearm in the frequency range 10-90 Hz compared to 10 differentiable steps on the fingertip. Rothenberg and Molitor (1979) also believed that the poor temporal resolution of the skin led to extreme difficulty in making temporal judgements, and although the authors of the current paper are not aware of any studies comparing the temporal capacities of the fingertip and the forearm it is possible that the fingertip is, again, more sensitive to such changes. This is obviously an area which warrants further investigation. Other factors which may have led to the differences in scores between the 2 studies are the speakers used and the possibility that the processing used in the present study was optimal for the speaker. The score obtained in this study also compares favorably with that reported by Martin et al (1983) for a subject using a multi-channel cochlear implant. The subject in Martin et al’s (1983) study was able to identify the stressed word in a 4 word sentence on 55% of presentations.

The results obtained in the present study indicate that a single channel vibrotactile aid, with appropriate processing of the speech signal, provides useful information as to emphatic stress in English sentences. This finding motivated further experiments aimed at evaluating the aid’s capacity to transmit other prosodic information.

**Experiment 2**

**Materials**

This study was aimed at evaluating the aid’s ability to transmit word syllable number and type. The test used consists of 4 lists of 40 items contrasting English monosyllables, trochees, spondees and trisyllables. Twenty of the contrasts presented were designed to be relative-
ly easy in that the syllable boundaries were marked by either stop consonants or unvoiced fricatives. The remaining 20 test items represent more difficult contrasts with their syllable boundaries marked by voiced, continuant consonants. Studies by Mártony (1974) and Zeiser and Erber (1977) have shown that this latter situation can lead to great difficulty in the actual identification of syllabic structure. The 4 test lists were recorded by GP using the equipment described for the previous experiment and with the speaker attempting to produce the words with normal syllabic stress.

Procedure

The settings used for the vibrotactile aid were those described for Experiment 1. Ten non-native speakers of English were divided into 2 experimental groups with Group 1 receiving Lists 1 and 2 whilst Group 2 were administered Lists 3 and 4. Masking noise was presented through headphones to ensure that the subjects received no auditory information from the vibrator. A test sheet was prepared which presented the 4 possible responses for each stimulus item, and the subjects were instructed to underline the word they thought had been spoken.

Results and Discussion

The results obtained in this experiment are presented in Table II. The scores for the 2 groups (46.75% and 51.25% respectively) indicate that this was an extremely difficult task for all the subjects. It should be noted, however, that the scores are still well above chance level performance. Confusion matrices which show the subjects' response patterns are given in Fig. 6. These reveal that the subjects' errors occur in a relatively systematic fashion. For example, monosyllables are very rarely confused with spondees or trisyllables but often with trochees. Almost all of the subjects commented that the task was extremely difficult and believed that the major contributing factors were the use of English words as stimuli and the inter-stimulus interval time of 5 seconds. The subjects felt that this allowed them insufficient time to scan the alternatives especially when it is considered they were unfamiliar with the stress patterns associated with many of the words. For this reason it was decided to administer similar materials in Swedish to determine whether the use of English words had led to unrealistically low scores in the present study.
Table 1 presents the data from Experiment 2. The table shows the mean number of errors made by three groups of participants: Group A, Group B, and Group C. The data are presented in three conditions: STIMULUS, THUMBS, and THUMBS plus pooled data. The mean number of errors for each condition is shown for Group 1 and Group 2.

**Group A**

<table>
<thead>
<tr>
<th>STIMULUS</th>
<th>THUMBS</th>
<th>THUMBS plus pooled</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>49.8</td>
<td></td>
</tr>
</tbody>
</table>

**Group B**

<table>
<thead>
<tr>
<th>STIMULUS</th>
<th>THUMBS</th>
<th>THUMBS plus pooled</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>51.25</td>
<td></td>
</tr>
</tbody>
</table>

**Group C**

<table>
<thead>
<tr>
<th>STIMULUS</th>
<th>THUMBS</th>
<th>THUMBS plus pooled</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>46.75</td>
<td></td>
</tr>
</tbody>
</table>
Table II. Scores (% correct) obtained by Group 1 (Lists 1 & 2) and 2 (Lists 3 & 4) for Experiment 2.

<table>
<thead>
<tr>
<th>Subject No.</th>
<th>Group 1</th>
<th>Group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>50</td>
<td>52.5</td>
</tr>
<tr>
<td>2</td>
<td>47.5</td>
<td>42.5</td>
</tr>
<tr>
<td>3</td>
<td>50</td>
<td>35</td>
</tr>
<tr>
<td>4</td>
<td>65</td>
<td>55</td>
</tr>
<tr>
<td>5</td>
<td>40</td>
<td>30</td>
</tr>
<tr>
<td>Mean</td>
<td>50.5</td>
<td>43.0</td>
</tr>
<tr>
<td>S.D.</td>
<td>9.08</td>
<td>10.81</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>60</td>
<td>37.5</td>
</tr>
<tr>
<td>2</td>
<td>50</td>
<td>45</td>
</tr>
<tr>
<td>3</td>
<td>55</td>
<td>62.5</td>
</tr>
<tr>
<td>4</td>
<td>40</td>
<td>60</td>
</tr>
<tr>
<td>5</td>
<td>45</td>
<td>57.5</td>
</tr>
<tr>
<td>Mean</td>
<td>50.0</td>
<td>52.5</td>
</tr>
<tr>
<td>S.D.</td>
<td>7.90</td>
<td>10.75</td>
</tr>
</tbody>
</table>

Experiment 3

Materials

The materials used in this experiment were taken from a Swedish rhyme test described by Mártony (1974). Five test lists were used presenting the following contrasts:

1. One- and two-syllable words with an intervocalic stop consonant.
   For example, bad/bada.

2. One- and two-syllable words with the syllable boundary marked by a voiced continuant.
   For example, mål/måla.

3. One-, two- and three-syllable words with stop consonants in the intervocalic position.
   For example, bad/bada/bada/de.

4. One-, two- and three-syllable words with voiced continuants in the intervocalic position.
   For example, mål/måla/måläre.

5. Monosyllabic words with long or short vowels.
   For example, sil/sill.

A tape recording of the test lists with a male Swedish speaker saying each stimulus word twice was used to present this material. Practice materials designed to familiarize the subjects with the test
stimuli were presented prior to the administration of each list.

Procedure

Analysis of the test tape revealed that the speaker's mean F0 and range were approximately the same as for the speaker in Experiments 1 and 2. Consequently the vibrotactile aid's settings were those used in the previous experiments. The subjects in the study were 10 native speakers of Swedish. Masking noise was again presented through headphones to ensure that the subject's responses represented tactual rather than auditory stimulation. The vibrator was hand-held and the subjects were given answer sheets showing the alternative responses, and were asked to circle the word presented.

Results and Discussion

The results for this experiment are given in Table III and indicate that the subjects do score considerably better in tasks involving word syllable identification when they are more familiar with the stimulus material. The scores contrasting different syllable types with intervocalic stops are extremely high with many subjects scoring 100% and the lowest score obtained being 75%. The scores for the lists which contrasted words with intervocalic voiced continuants, although slightly poorer, still represent a high degree of proficiency with the task. The scores for long and short vowel identification varied considerably from subject to subject. Four subjects scored 80% while the remaining 6 scored 70% or less. It is probable that the scores obtained by 5 of the subjects represent chance level performance. The subjects reported feeling more comfortable with this material than they had with the English material presented in Experiment 2. It would appear that the F0 and intensity contour provides very useful information which enables subjects to accurately identify word syllable number. This indicates that the present aid may be a worthwhile means of presenting information to assist the deaf in both speech production and perception.
Table III. Scores (% correct) obtained by subjects for Experiment 3.

<table>
<thead>
<tr>
<th>Subject No.</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>83.3</td>
<td>91.7</td>
<td>100</td>
<td>94.5</td>
<td>80</td>
</tr>
<tr>
<td>2</td>
<td>100.0</td>
<td>87.5</td>
<td>100</td>
<td>77.8</td>
<td>80</td>
</tr>
<tr>
<td>3</td>
<td>91.7</td>
<td>91.7</td>
<td>100</td>
<td>83.3</td>
<td>80</td>
</tr>
<tr>
<td>4</td>
<td>75.0</td>
<td>87.5</td>
<td>88.9</td>
<td>66.6</td>
<td>80</td>
</tr>
<tr>
<td>5</td>
<td>91.7</td>
<td>83.3</td>
<td>100</td>
<td>83.3</td>
<td>65</td>
</tr>
<tr>
<td>6</td>
<td>95.8</td>
<td>91.7</td>
<td>100</td>
<td>83.3</td>
<td>65</td>
</tr>
<tr>
<td>7</td>
<td>100.0</td>
<td>95.8</td>
<td>100</td>
<td>100.0</td>
<td>55</td>
</tr>
<tr>
<td>8</td>
<td>100.0</td>
<td>100.0</td>
<td>94.5</td>
<td>66.6</td>
<td>55</td>
</tr>
<tr>
<td>9</td>
<td>100.0</td>
<td>100.0</td>
<td>100</td>
<td>94.5</td>
<td>70</td>
</tr>
<tr>
<td>10</td>
<td>100.0</td>
<td>100.0</td>
<td>100</td>
<td>88.9</td>
<td>65</td>
</tr>
</tbody>
</table>

Mean          | 93.75 | 92.92 | 98.34 | 83.88 | 69.5 |
S.D.          | 8.61  | 5.91  | 3.74  | 11.28 | 10.12 |

Experiment 4

Materials

The final experiment in this study involved the perception of emphatic stress in Swedish sentences. The test materials comprised the following sentences consisting of 3 two-syllable words:

Olle jagar hunden. (Olle hunts the dog.)

Anna lurar Gunnar. (Anna plays a trick on Gunnar.)

Lena lagar nallen. (Lena fixes the teddy-bear.)

Eleven lists, each with 18 test items, were recorded, presenting the 3 sentences with emphatic stress placed on either the first, second or third words. The speaker was a Swedish male (AR) who attempted to produce the sentences at a normal rate with very few intensity or durational cues. As a result it is felt that the subjects would have to rely upon frequency information if they were to correctly identify the stressed word. Analysis of the test materials confirmed that the primary cue to emphatic stress was frequency change, with relatively weak intensity and durational cues. The speaker’s mean F0 was found to be 113 Hz with a range from 80-150 Hz. The F0 contours also revealed that in most cases emphatic stress was marked by a rise in F0 on the second syllable of the stressed word. The rises in F0 accompanying emphatic stress also appeared to be less marked than those found for the English
materials in Experiment 1. It was anticipated that this would make the task difficult even for native speakers of Swedish. The 11 lists were prepared for use on a programmed learning machine to determine whether subject performance would improve with training. Control tones were recorded onto the second audio track of the test recording to stop the tape after each presentation and also to inform the subject of the correctness of her/his response.

Procedure

Four subjects were used in this experiment. Two of the subjects (1 and 2) were native speakers of Swedish while the 2 other subjects (3 and 4) had English and Portuguese respectively as their first language. The subjects were instructed that they would feel a sentence of the type "Lena lagar nallen" and their task was to indicate whether the first, second or third word had been stressed. This was to be done by pushing 1 of 3 buttons on the teaching machine. The subjects received immediate feedback as to the correctness of their response and a running total of correct responses was automatically displayed. Each subject was presented with 5 lists a day (90 sentences) for a period of 5 days, and their scores were recorded over the training period. In an attempt to maximize the F0 information available the vibrotactile aid was set so that the F0 range 90-150 Hz was presented in the frequency range 30-100 Hz. Intensity cues were also provided. The subjects wore headphones providing sufficient masking noise to ensure that they were receiving no auditory information from the vibrator.

Results and Discussion

The mean scores obtained for the 5 training sessions are presented in Table IV. These indicate that the subjects' performance remained relatively stable over the duration of the training period. The mean scores for the 4 subjects across the 5 training sessions were 58.16%, 59.06%, 53.33% and 52.16% respectively. It is interesting to note that the 2 non-native speakers of Swedish had lower scores than the native speakers. The subjects' performance overall is relatively poor, especially when compared to the results obtained in Experiment 1. The complexity of the test material was probably an important factor, and it was decided to provide intensive training for 1 subject using a variety of processing schemes. One of the authors of this paper (GP) acted as a subject for 10 days and the results obtained are presented in Fig. 7.
Fig. 7. Results for subject CQ over 40 training sessions using the material presented in Experiment 4.

Session No. 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40

Percent Correct

-75 -50 -25 0 25 50 75 100

Guessing
Each datum point in the figure represents 5 of the test lists presented in a random order.

<table>
<thead>
<tr>
<th>Subject No.</th>
<th>Session Number</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>55.5</td>
<td>56.6</td>
<td>58.8</td>
<td>57.7</td>
<td>62.2</td>
<td>57.15</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>60.0</td>
<td>57.7</td>
<td>55.5</td>
<td>54.4</td>
<td>67.7</td>
<td>59.06</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>62.2</td>
<td>45.5</td>
<td>53.3</td>
<td>55.5</td>
<td>50.0</td>
<td>53.3</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>47.7</td>
<td>43.3</td>
<td>58.8</td>
<td>52.2</td>
<td>58.8</td>
<td>52.16</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>56.35</td>
<td>50.77</td>
<td>56.6</td>
<td>54.95</td>
<td>59.67</td>
<td></td>
</tr>
<tr>
<td>S.D.</td>
<td></td>
<td>6.40</td>
<td>7.42</td>
<td>2.69</td>
<td>2.28</td>
<td>7.41</td>
<td></td>
</tr>
</tbody>
</table>

Table IV. Performance (% correct) for the 5 subjects in Experiment 4.

The first 10 sessions were conducted with the FO range 90-150 presented in the frequency range 30-100 Hz with intensity cues available. Sessions 11-20 involved the use of a different frequency range as all subjects involved in the training experiment had commented that the use of processing which presented 90Hz as 30 Hz was uncomfortable and made the speech feel unnatural. In an attempt to overcome this, and also to maximize the available frequency range the FO range 90-150 Hz was encoded in the frequency range 40-140 Hz with intensity cues also available. The final 20 sessions (21-40) were conducted using the same frequency setting as that used in sessions 11-20 but with reduced intensity cues. This was achieved by setting the rectifier in the "log" position. This final setting represented an attempt to eliminate any spurious frequency cues which may arise with changes in intensity. Kirman (1974), for example, has noted that "perceived vibratory pitch is very considerably influenced by the amplitude as well as the frequency of the vibratory stimulus" (Kirman, 1974; p. 55). It can be seen that this subject's performance improves steadily throughout the training period rising from a starting point around 45% to a final performance over 80%. This score compares well with those obtained in Experiment 1 but the amount of training necessary to reach this level was considerable. It is felt that the scores represent true improvements in per-
formance rather than merely increased familiarity with the test lists. Attempts were made to guard against this factor by randomizing the order of presentation of the test lists from training session to training session. The final 5 sessions were conducted without feedback as to the correctness of the response in a further attempt to ensure that the subject did not receive any cues as to list order.

**Conclusion**

The results obtained in all of the experiments reported in this study indicate that a single channel aid which presents FO and intensity variations provides much useful information which could assist hearing impaired persons in both speech production and perception. The use of an expanded and lowered frequency range enabled inexperienced subjects to identify sentence stress and word syllable number and type with a high degree of proficiency. It would also appear that performance would improve with extended training periods. It should be remembered that the subjects used in this study have had extremely limited experience with vibrotactile stimulation and yet their performance level was extremely high. It would appear that Rothenberg and Molitor's (1979) contention that the use of the frequency range below 100Hz would lead to a signal which provides some frequency perception, is justified. It is also tempting to speculate that subject performance would improve if the speaker used a rate of articulation that is slower than normal. Some caution needs to be exercised on this point, however, for the use of an unnaturally slow rate of articulation may worsen an already existing problem. One of the major faults in the speech of both the congenitally and adventitiously deaf is their slow and laboured rate of articulation (Osberger and McGarr, 1982; Plant, 1983). For this reason the use of an overly slow rate of articulation may lead to improvements in speech perception, but be accompanied by an unacceptable decline in speech quality. It is probably better to investigate whether the temporal resolution of the skin can be improved with training before deciding to introduce a speaking strategy which leads to a decline in speech production.

Further research also needs to be done to determine whether the signal provided by the aid is a useful supplement to assist in lipreading and speech production. Further studies involving the use of Connected Discourse Tracking (De Filippo and Scott, 1978) are planned to
evaluate the effectiveness of the aid as a lipreading supplement. The subjects in this study will be a group of adventitiously deaf adults experienced in the use of vibrotactile aids. The aid's use as a means of providing feedback for speech production will also be evaluated, again using adventitiously deafened adults as the experimental subjects. The areas to be examined in this latter study will include the use of the aid as a means of monitoring F0 range, which has been shown to be restricted in the speech of the deafened (Silverman and Calvert, 1978). Consideration also needs to be given to the provision of other information which will serve to improve the signal available. This may include the use of a spatial array presenting spectral information, or may be as simple as a single vibrator which signals the presence of high frequency information in the speech signal. Finally there exists a very real need to develop a wearable version of this aid so that deaf persons can receive assistance at all times, rather than having a unit which is confined to use in an experimental setting. The use of radio-frequency transmission offers deaf persons the opportunity to receive a speech signal which is not degraded by distance or adverse signal-to-noise ratios. If such an aid is developed, the long-term effects of using a processing system aimed at overcoming some of the limitations of the tactile sense can be evaluated.

References


