SIMULTANEOUS MEASUREMENTS OF VOWELS PRODUCED BY A HEARING-IMPAIRED SPEAKER

NANCY S. MCGARR AND CAROLE E. GELFER
Haskins Laboratories
and
Graduate School, The City University of New York

Perceptual judgments, acoustic measurements and electromyographic (EMG) records were obtained for one deaf speaker producing the vowels [i, ɪ, æ, ə, u, ʊ] in an [hVD] frame. Overall listener judgments were consistent with spectral measurements. In general, front vowels were perceived as more similar to targets than back vowels, and high vowels were perceived correctly more often than low vowels. Experienced and inexperienced listeners were found to differ significantly in their categorization of the point vowels [i, æ, ə, and u] but not for [ɪ and ʊ]. The vowel space, as determined by the formant frequency measures, was reduced with respect to normal values particularly in the region appropriate to high back vowels. However, EMG records of genioglossus and orbicularis oris do not entirely account for the perceptual and acoustic data. In particular, genioglossus activity is relatively undifferentiated across all vowels when compared to data from normals. The results of this study generally support the widespread notion of reduced vowel space secondary to a reduced range of tongue movement in this deaf speaker. The physiological records were also characterized by a significant degree of variability from token to token. In this regard, these data are different from acoustic and physiological patterns that have been previously reported for vowels produced by deaf speakers.

INTRODUCTION

There is a considerable literature that describes the typical vowel errors produced by hearing-impaired speakers. These studies usually rely on perceptual assessments wherein experienced or inexperienced listeners transcribe the productions and the error patterns are noted (e.g., Hudgins and Numbers, 1942; Smith, 1975). In these studies, hearing-impaired speakers have been found to produce back vowels correctly more often than front vowels (Boone, 1966; Geffner, 1980; Mangan, 1961; Nober, 1967; Smith, 1975) and low vowels correctly more often than those with mid or high tongue positions (Geffner, 1980; Nober, 1967; Smith 1975). On the other hand, Stein's (1980) cineradiographic study of five deaf speakers shows “fronting” of back vowels. Similarly, Crouter (1963) reported greater variation in tongue shape for [i] than for [u] and [ə] as measured by cinefluorography.

* We are grateful to Mrs. Margaret Hilbok for enthusiastically serving as a subject for this study, to Rosemary Szcesniak for assisting in the data analysis and to Judith Rubin, Katherine S. Harris, and Peter Alfonso for providing helpful comments. This work was supported by NINCDS Grants NS-13617 and NS-13870, and NIH Biomedical Research Support Grant RR-05596 to Haskins Laboratories.
Hearing-impaired speakers also fail to distinguish between what has traditionally been referred to as the "tense-lax" distinction between vowel pairs such as [i–ɪ]. Often the substitution is for the "tense" member of the pair (Mangan, 1961; Monsen, 1974; Smith 1975), although other less closely related vowel substitutions have also been reported (Hudgins and Numbers, 1942; Markides, 1970).

The acoustic characteristics of vowels produced by deaf speakers have also been examined using techniques such as spectrographic analysis (Angelocci, Kopp and Holbrook, 1964; Monsen, 1976; Bush 1981) and linear predictive coding (LPC) (Osberger, Levitt and Slosberg, 1979). Formant frequency measures show a reduced phonological space with formant values tending toward the neutral vowel [ʌ]. Monsen (1976) noted that the second formant of vowels produced by hearing-impaired children remained around 1800 Hz rather than varying as different vowels were articulated. Perceptual judgments and acoustic analyses have, thus, led some researchers (e.g., Angelocci et al., 1964; Horwich, 1977) to propose that hearing-impaired speakers use a limited amount of tongue movement and consequently do not achieve vowel differentiation. Some studies (Bush, 1981; Martony, 1968) suggest that deaf speakers who produce vowel distinctions do so by exaggerated variations in $F_0$, particularly for high vowels such as [i] and [u]. Existing physiological studies of deaf speech production — electromyography (Huntington, Harris and Sholes, 1968; McGarr and Harris, 1983; Rothman, 1977) and cineradiography (Crouter, 1963; Stein 1980; Zimmermann and Retallack, 1981) — are few and provide minimal information regarding vowel production.

Each type of investigation — descriptive, acoustic and physiological — contributes partial insight into a deaf speaker's vowel production. However, only a few studies (cf. Huntington et al., 1968; Rothman, 1977) incorporated simultaneous acoustic and articulatory measures of production with listener judgments or phonetic transcriptions. The paucity of such studies is undoubtedly related to the considerable effort and specialized technology required to obtain such measures from deaf speakers. However, the information potentially gained from such simultaneous measures could greatly enhance our knowledge of speech organization in the deaf population.

This study was undertaken as a preliminary investigation of the hypothesis that deaf speakers fail to vary tongue position in their attempt to achieve vowel differentiation. EMG activity was recorded from the posterior genioglossus muscle and superior and inferior orbicularis oris of one deaf speaker. Listener judgments were obtained and acoustic analyses were performed in order to reconcile these measures with physiological records.

**METHODS AND PROCEDURES**

The pre-lingually deaf speaker (pure tone average for 0.5, 1 and 2 KHz = 105 dB ISO) was a woman who attended an oral school for the deaf and also received remedial speech classes as an adult. Speech samples obtained from the subject were analyzed in several ways. First, a listener highly experienced with the speech of the deaf rated her spontaneous speech samples for overall intelligibility. Following the format described by
Subtelny (1975), this subject was classified as difficult to understand, producing only occasional intelligible words or phrases. Second, judgments of vowel identity were obtained from five listeners experienced with the deaf and 18 listeners who had no previous experience with deaf speech. Listeners were asked to identify the vowel they heard from a closed set of vowels and diphthongs. The data, confusion matrices were derived. Third, narrow phonetic transcriptions were made by a phonetician. The listener judgments and phonetic transcriptions will be described further below.

Simultaneous acoustic and electromyographic recordings were made of the speaker's production of 10 randomized repetitions each of the vowels [i, i, æ, a, u, u] in an [hVd] frame. Because of technical problems, only five repetitions of [u] could be analyzed perceptually and acoustically; the EMG signals for this vowel could not be analyzed. Conventional hooked-wire electrodes were inserted into the posterior fibers of the genioglossus muscle, which elevates and buncles the main body of the tongue (Raphael and Bell-Berti, 1975; Raphael, Bell-Berti, Collier and Baer, 1979). The electrode preparation and insertion techniques for this muscle have been reported in detail elsewhere (Hirose, 1971). Patterns of peak genioglossus activity for vowels produced by a hearing speaker are shown in Figure 1 for purposes of comparison with our data.
Vowels of a Hearing-Impaired Speaker

(Alfonso and Baer, 1982). This figure shows that greater muscle activity occurs for the front vowels [i] and [i'], and to a lesser extent, [u]; the genioglossus shows relatively little activity for [a]. Thus, genioglossus appears to be active for high vowels in general and for front vowels in particular.

Measures were also made of lip-rounding activity using surface electrodes to record from the superior and inferior orbicularis oris muscles (Allen, Lubker and Harrison, 1972). It was assumed that only [u] and [u'] would show significant orbicularis oris activity.

The acoustic and electromyographic (EMG) data obtained from the deaf speaker were analyzed in the following ways. First, the experienced listeners' judgments were used to sort the production tokens into three categories: perceptually correct productions (at least four of the five listeners agreed with the intent of the talker), perceptually incorrect productions (four or more listeners disagreed with the intent of the talker), and perceptually equivocal (two or three listeners heard the vowel as intended; the remaining heard it as incorrect). Second, spectral analyses and vowel duration measurements were performed on an interactive computer system at Haskins Laboratories. Third, the EMG signals were rectified, integrated and further analyzed as previously described (Kewley-Port, 1973).

Results

Listener judgments

Table 1 shows the confusion matrices obtained from the listeners' scores. Fifty judgments were obtained from the experienced listeners (5 listeners x 10 repetitions) for each vowel; 180 judgments were obtained from the inexperienced listeners (18 listeners x 10 repetitions). Percentages are reported for each listening group for each vowel. In general, the pattern of correct responses is similar for the two groups of listeners. Overall, listeners perceived the front vowels [i], [i'], and [æ] as correct more often than the back vowels [a, u] or [u']. Confusions for the high front vowels [i] and [i'] were most often restricted to this "tense-lax" pair, although this was not the case for other vowel pairs. Substitution errors occurred across the vowel space for other target vowels. Of significance is the considerable number of [i] or [i'] substitutions for [u] or [u'] targets. Percentages of correct judgments for experienced and inexperienced listeners across all vowel types, (taken from Table 1) and their averages, are summarized in Table 2. Table 3 shows the ranking of the most common combined listener responses (again taken from Table 1) for each vowel. It is interesting that vowels tended to be judged as more fronted than their targets. A two-by-two Chi-square analysis was performed on the most common listener response versus all other choices in order to ascertain if the two groups of listeners differed in their categorizations. There was a significant difference between experienced and inexperienced listeners for the vowels [i] ($\chi^2 = 16.4, p < 0.01$), [æ] ($\chi^2 = 17.3, p < 0.01$), [u] ($\chi^2 = 18.3, p < 0.01$), [u'] ($\chi^2 = 4.5, p < 0.05$) but not for the vowels [i and u']. That is, both groups of listeners tended to cluster their responses for
TABLE 1

Confusion matrices of listeners' judgments for vowels produced by the deaf speaker. Scores are reported as percentages.

**Experienced Listeners**

<table>
<thead>
<tr>
<th>Target vowel</th>
<th>i</th>
<th>e</th>
<th>æ</th>
<th>ø</th>
<th>o</th>
<th>u</th>
<th>u</th>
<th>e</th>
<th>æ</th>
<th>Other vowels</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>74</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>i</td>
<td>48</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>æ</td>
<td>2</td>
<td>4</td>
<td>32</td>
<td>42</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>14</td>
</tr>
<tr>
<td>ø</td>
<td>6</td>
<td>4</td>
<td>26</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>u</td>
<td>32</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td>16</td>
<td>32</td>
<td></td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>u</td>
<td>46</td>
<td>24</td>
<td>2</td>
<td></td>
<td></td>
<td>8</td>
<td>14</td>
<td></td>
<td></td>
<td>6</td>
</tr>
</tbody>
</table>

**Inexperienced Listeners**

<table>
<thead>
<tr>
<th>Target vowel</th>
<th>i</th>
<th>e</th>
<th>æ</th>
<th>ø</th>
<th>o</th>
<th>u</th>
<th>u</th>
<th>e</th>
<th>æ</th>
<th>Other vowels</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>42</td>
<td>21</td>
<td>1</td>
<td>1</td>
<td>.5</td>
<td>1</td>
<td>4</td>
<td>13</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>i</td>
<td>27</td>
<td>34</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>8</td>
<td>4</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>æ</td>
<td>5</td>
<td>11</td>
<td>11</td>
<td>15</td>
<td>5</td>
<td>.5</td>
<td>7</td>
<td>5</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>ø</td>
<td>8</td>
<td>13</td>
<td>7</td>
<td>14</td>
<td>6</td>
<td>2</td>
<td>9</td>
<td>4</td>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td>u</td>
<td>22</td>
<td>17</td>
<td>2</td>
<td>1</td>
<td></td>
<td>16</td>
<td>25</td>
<td>5</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>u</td>
<td>30</td>
<td>24</td>
<td>4</td>
<td>2</td>
<td>.5</td>
<td>1</td>
<td>8</td>
<td>9</td>
<td>7</td>
<td>2</td>
</tr>
</tbody>
</table>

the lax vowels, while the experienced listeners' responses also clustered for the point vowels. Inexperienced listeners, on the other hand, were more scattered in their responses for the point vowels.

**Acoustic measures**

Figure 2 shows the values for $F_1$ and $F_2$ for all tokens of all vowels. These measurements were taken at the center, and relatively steady-state, portion of the vowel. Formant values for $F_1$ grossly differentiate between high and low vowels, while the range of $F_2$
variation is restricted. These latter values imply limited backward movement of the tongue. In an attempt to produce the back vowel [ʊ], this speaker succeeds only in approaching mid range. Thus, the values for the low vowels [æ] and [ɑ] cluster, and the tendency for listeners to perceive [ɑ] as [æ] is not surprising. The $F_2$ values for [u] are grouped with [ɪ] and [i], and thus, an acoustic basis for the listeners' perceptual judgments becomes somewhat more apparent. Some formant values for [ʊ] are similarly found to have a high $F_2$, although two tokens show a more appropriate formant range.

Because these acoustic data are not totally adequate in explaining listener identification accuracy, particularly in discriminating [ɪ] from [i], it seemed reasonable to assume that some other acoustic cue must be available to the listeners. Figure 3 shows $F_2$ plotted against duration for all vowels. It can be seen that the vowels [ɪ] and [i] are differentiated on the basis of duration, with values for [ɪ] considerably longer than those for
Formant values of \( F_1 \) and \( F_2 \) for 5 vowels produced by the deaf speaker. Values in squares are the average formant values for (non-deaf) women reported by Peterson and Barney (1952). Values for [v] are from Fischer-Jørgensen (1960).

\( [i] \) Differentiation of vowels such as [i] and [ɪ] on the basis of durational cues has been noted previously for deaf speakers (Angelocci et al., 1964; Levitt, Osberger and Stromberg, 1979; Monsen, 1974). There is no clear differentiation of other vowels based on durational cues. Overall durations of vowels produced by this deaf speaker were considerably longer than those reported for normals, which is frequently observed for hearing-impaired speakers (Calvert, 1961; Osberger and Levitt, 1979).

**EMG analysis**

Figure 4 shows the patterns of peak posterior genioglossus activity for the five vowels analyzed for this deaf speaker. EMG activity for [v] could not be analyzed due to
technical problems. Perceptually correct productions, perceptually incorrect productions, and perceptually equivocal productions are plotted. The data show an obvious lack of differentiated peak genioglossus activity across nearly all vowels regardless of perceptual category. However, one would expect more genioglossus activity for [i] and [ɪ], somewhat less for [ʊ], and still less for [æ] and [ɑ] (see Fig. 1). This pattern was not observed even for this speaker's correct productions. Furthermore, peak genioglossus activity was not greater for the vowel [ɪ] than for the vowel [i], as might be observed in the productions of hearing speakers (Alfonso and Baer, 1982; Raphael and Bell-Berti, 1975). Furthermore, values of peak genioglossus activity for all incorrect categories of [u] were greater than values obtained for any perceptually correct high front vowel.

Because of this unexpected pattern of genioglossus activity for [u] as well as the number of listeners who judged the production as [ɪ] (cf. Table 1), narrow phonetic transcriptions were obtained. Eight of the 10 tokens intended as [ʊ] were transcribed by a trained phonetician as [ʏ], a high front rounded vowel not typical in American English.
Fig. 4. Peak genioglossus activity in microvolts (μV) for vowels produced by the deaf speaker. At the top of each column, the number of experienced listeners whose judgments fell into each category (perceptually correct, equivocal or incorrect) is noted. At the bottom of each column are noted the vowel judgments assigned by the listener to the corresponding token. See text for more detailed discussion.

Figure 5 shows a comparison of genioglossus activity for selected tokens intended and transcribed as [i] with those intended as the vowel [u] but transcribed as [y] (and perceived as [i] by our listeners, cf. Figure 4). Both sets of tokens are distinguished by variability in the onset and offset of genioglossus activity. In some instances (e.g., token 1 for correct [i] productions), onset of genioglossus occurs quite early, while for other tokens (e.g., token 3), the onset is considerably later. It is noteworthy that, despite token-to-token variability for both correct and incorrect productions, the overall pattern of activity for the two categories is nearly identical. That is, no single distinguishable peak of muscle activity is identifiable with production of a high front vowel.

Figure 6 shows genioglossus and orbicularis oris activity for three utterances: [i] correct, [u] equivocal, and [y/u] substitutions (i.e., [u] incorrect). There is no token that four of the five experienced listeners judged correctly as [u]. For both the equivocal productions of [u] and those transcribed as [y], there is the expected orbicularis oris activity associated with lip rounding. However, while it is difficult to state with certainty
Fig. 5. Genioglossus activity in microvolts (µV) for selected tokens of vowels produced by the speaker. At the left, tokens transcribed by the phonetician as a correct production of [i], at the right, tokens intended as [u] but transcribed as [y]. Data plots show the ensemble average for 7 tokens of [i], and 8 tokens of [u] for the genioglossus muscle. Four individual tokens are shown below. The vertical line, the line-up point at 0 msec for these measures, is the onset of voicing for the vowel. The offset of voicing occurs 500 msec after the line-up point.

what differentiates the last two categories, in the equivocal case, orbicularis oris activity is maintained as long as that for genioglossus, while for [y], orbicularis oris activity ceases earlier and genioglossus activity begins sooner. Thus, it is possible that the temporal relationship between orbicularis oris and genioglossus represents at least one of the underlying bases for the acoustic cues that lead to different listener impressions.

DISCUSSION

The acoustic results of the present study are in general agreement with those of previous studies in demonstrating a reduced vowel space. However, the reduction appears to occur mostly in the front–back dimension, with the high vowels [i, ɪ, u] and some tokens of [u] clustered around a high $F_2$ (range = 1975–2300 Hz) and the low vowels [æ, a] clustered around the mid-range $F_2$ (range = 1600–2075 Hz). In general, the
Fig. 6. Selected individual tokens of the EMG potentials from the genioglossus and orbicularis oris muscles as produced by the deaf speaker. The line-up point is as in Figure 5. Offset of voicing occurs 500 msec after the line-up point. Tokens shown are [i] judged as correct (top), [u] as equivocal (mid), [u] as [y] (bottom). See text for further discussion.

judgments of both experienced and inexperienced listeners are consistent with the acoustic measures, although the experienced listeners, on average, made more correct judgments than the inexperienced listeners (cf. Table 2). The higher scores achieved by the experienced listeners may be attributed to this group's ability to disambiguate [i] from [ɪ], and [æ] from other front vowels (cf. Table 1). The data also show that this speaker tends to produce front vowels more often than back vowels, whether correct or incorrect, and to produce high vowels correctly more often than low vowels. These data thus differ from previous descriptive studies of vowels produced by deaf speakers (Boone, 1966; Geffner, 1980; Mangan, 1961; Nober, 1967; Smith 1975) that report better production of back or low vowels, although the data concur with results obtained in cineradiographic studies (Crouter, 1963; Stein, 1980).

It is apparent that there is an acoustic basis for the listeners' judgments. Formant values of [i] and [ɪ] fall roughly in the appropriate range so that the relatively high number of correct judgments for these vowels can be explained. Similarly, formant values for this speaker's intended productions of [u] account for the high percentage of [i]
and [i] listener judgments and the [y] judgments of the phonetician. This speaker had considerable success in differentiating high and low vowels, although the formant values for the low vowels are inappropriate with respect to normal productions. Thus, the acoustic basis for the very low percentage of [a] judgments is readily explained. In fact, overall there is a fairly straightforward relationship between the acoustic measures and the listener judgments.

We are limited in our inferences regarding the physiological basis of the acoustic data in that only one tongue muscle (posterior fibers of the genioglossus) was studied. Therefore, the implied failure to produce back tongue movements from the acoustics cannot be confirmed physiologically. However, we can address ourselves to the relative appropriateness of the degree of genioglossus activity for all vowels. As noted in Figure 3, genioglossus activity for this deaf speaker is, on average, relatively undifferentiated across all vowels studied. This is in striking contrast to the results for a normal speaker (cf. Fig. 1). Furthermore, even for tokens that were perceived as correct, onset and offset of genioglossus activity was highly variable from token to token. It is not surprising, then, that there are so many equivocal and incorrect productions. Furthermore, the pattern of EMG activity also does not readily distinguish between [i] and [i], so that the corresponding listener judgments seem to be based primarily on duration cues. However, the relatively uniform level of genioglossus activity for [i, 1, a] does explain the general tendency for $F_2$ values to occur in regions expected for high front vowels.

Therefore, based on acoustic and physiological measures, we conclude that this deaf speaker fails to vary tongue position, particularly in the front–back dimension, in order to achieve vowel differentiation. Although the vowel space is reduced overall, there is considerable differentiation in the high–low plane, as is evident from the ranked listener responses in Table 3. Productions of [a] differed from [i] primarily on the basis of lip-rounding, as noted in the electromyographic records of the orbicularis oris. Such a production strategy is not surprising when we consider that, owing to difficulty in perceiving acoustic cues, deaf speakers rely heavily on visual information for deriving cues to place of articulation. Examples of these would include lip-rounding, as noted above, and jaw-lowering for production of low vowels. When acoustic cues can be perceived with limited residual hearing, e.g., vowel duration, the speaker employs these, as noted for the “tense–lax” pair [i–1].

While this study is only preliminary, it provides some insight into the physiological differences between deaf and hearing speakers in vowel production. We are intrigued by the token-to-token variability noted in the onset and offset of genioglossus records. We intend to examine this issue further by examining other tongue muscles that are known to be important in vowel production, particularly the extrinsic muscles, hyoglossus and styloglossus. In addition, we will investigate the hypothesis that deaf speakers, such as our subject, who do not vary tongue position achieve vowel differentiation by exaggerated variation in larynx height and fundamental frequency.
REFERENCES


