INTRODUCTION

While many children who are born severely or profoundly deaf, or become deaf in infancy achieve intelligible speech, the vast majority do not. Speech intelligibility is fairly well correlated with residual hearing (Boothroyd, 1970; Smith, 1972) at least until 90dB, and overall intelligibility is well correlated with the percent of segmental errors, and to a lesser extent with suprasegmental deviancy (Levitt, Smith, & Stromberg, 1974). While many educators of the deaf would claim that the characteristic unintelligibility of deaf speakers is a consequence of faulty teaching practices (Haycock, 1933; Ling, 1976), independent investigations have been remarkably consistent in showing similar patterns of segmental and suprasegmental errors in the speech of deaf talkers trained in a wide variety of programs (Hudgins & Numbers, 1942; Smith, 1972; Levitt, Stark, McGarr, Carp, Stromberg, Gaffney, Barry, Velez, Osberger, Leiter, & Freeman, Note 1; Johnson, 1975). Furthermore, experienced teachers of the deaf can discriminate between deaf and non-deaf speakers from disyllables produced by both groups (Calvert, 1961), and experienced listeners of the deaf are better than naive listeners in decoding deaf utterances (McGarr, 1978). If we accept the point of view that there is a generic "deaf speech" pattern, not dependent at least on the fine-grained details of the training procedure, we may ask what are its characteristics? Why do the deaf sound as they do? Why are they unintelligible?

One hypothesis, primarily concerned with consonant articulation, is that deaf speakers place their articulators fairly accurately—especially for those places of articulation that are highly visible—but fail to coordinate the movements of several articulators normally (Huntington, Harris, & Sholes, 1968; Levitt et al., 1974). Thus, we may suggest that the errors in deaf speech are the consequences of incorrect motor planning in time.
A second hypothesis, primarily concerned with vowel articulation, is that deaf speakers move their articulators through a relatively restricted range, thereby "neutralizing" vowels (Angelocci, Kopp, & Holbrook, 1964; Monsen, 1974). However, this hypothesis fails to account for the great variability in the speech production of deaf talkers, a point we will discuss in further detail later.

A third hypothesis is that the inability of deaf speakers to control the suprasegmental characteristics of their speech makes both segmental and suprasegmental characteristics more difficult for listeners to decode (Harris & McGarr, 1980). Suprasegmental aspects of speech may be so abnormal as to mislead the listener. Deaf speakers may not preserve phonological contrasts or may produce them in a way that makes information about the intended contrast unavailable to the listener, and perhaps block information about other contrasts. That fundamental frequency (McGarr & Osberger, 1978) and overall duration levels (e.g., Osberger & Levitt, 1979) are often deviant in deaf speakers is well known. These deviations alone might interfere with a listener's ability to decode a speech signal, even if other suprasegmental contrasts were preserved in either a normal or an abnormal way.

On an entirely different level, poor control of the speech source function may simply provide inadequate support for the acoustic realization of upper articulator movement. Deaf speakers characteristically take in less air in speech respiration (Forner & Hixon, 1977; Whitehead, in press) and may, in addition, convert air into acoustic energy inefficiently due to poor control of the larynx.

This paper presents a preliminary attempt to assess these hypotheses by examining a number of productions of some simple utterances by a single deaf talker using listeners to judge production accuracy utterance-by-utterance. While it is obvious that more subjects must be studied in order to reach firm conclusions, we believe that the general technique of examining interarticulator programming in depth with combined perceptual, acoustic, and physiological techniques is a promising avenue for investigation.

**METHODS AND PROCEDURES**

The prelingually deaf speaker in this study is a woman in her mid-forties who graduated from an oral school for the deaf, and has received remedial speech classes as an adult. Her pure tone average is 105dB ISO. Informal ratings of spontaneous speech samples suggest that her productions would be characterized as fairly typical of her group. For purposes of comparison, productions of a hearing speaker who has frequently served as an experimental subject were also examined.

Each subject produced approximately 20 repetitions of each of six utterance types. These utterances were nonsense words of the type /ə pip ə p/, /ə pip/, and /ɒp ə pip/ with stress on either the /i/ or the /ə/. For this paper, data will be presented primarily for the first and third utterance types. Paint-on surface electrodes were used to record from the orbicularis oris muscle (Allen, Lubker, & Harrison, 1972); conventional hooked-wire electrodes were inserted into the genioglossus muscle. The electrode prepara-
tion and insertion techniques for the genioglossus muscle electrodes have been reported in detail elsewhere (Hirose, 1971). Conventional acoustic recordings were made at the same time as the electromyography.

The acoustic and electromyographic (EMG) data obtained from the two speakers were analyzed in several ways. First, for the deaf speaker, the acoustic recordings of six utterance types were randomized and presented to listeners inexperienced in hearing deaf speech. The listeners were required to select one of the six utterance types presented on an answer sheet, for each item they heard. Confusion matrices were obtained. The hearing subject's productions were not checked perceptually, but informal listening suggested that perceptual errors would not be made by listeners to her speech. Second, acoustic measurements were made on an interactive computer system at the Haskins Laboratories and with conventional sound spectrography. Third, the EMG signals were rectified, integrated, and then further analyzed, as we will describe below.

RESULTS

Listener Judgments

First, examining the results of the listening test, we found that the deaf speaker was judged as being fairly intelligible (at least as measured by a closed response listening task). Table 1 shows the confusion matrix obtained from the listeners' scores. An item was considered to be correct if 9 out of 10 listeners identified it as the originally intended utterance. The average percent correct for all utterance types was 75%. Overall, there were more errors of stress than of the type (i.e., a vowel identity error). In fact, only for the utterance /əpə'pip/ was there a significant number of vowel errors. In this case, the listeners perceived the utterance as /ə'pip/ 32% of the time.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. ə'pi pap</td>
<td>88</td>
<td>08</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. əpi'pap</td>
<td>25</td>
<td>75</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. ə'pi pip</td>
<td>83</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. əpi'pip</td>
<td>07</td>
<td>91</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. əpa'pip</td>
<td>67</td>
<td>29</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. apa'pip</td>
<td>32</td>
<td>16</td>
<td>51</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1
Confusion Matrix of Listeners' Judgments for the Deaf Speakers
Using these listener judgments, all tokens (repetitions) of an item were divided into two categories: "perceived correct" utterances and "stress error" utterances. Only for the intended utterance /æ pəˈpip/ was there an additional category (that of a vowel error).

**Acoustic Measurements**

The acoustic cues used to convey contrastive stress in normal speech production have been extensively studied (Fry, 1958, 1964; Harris, 1978). In general, speakers convey changes in contrastive stress to listeners by differences in acoustic cues such as vowel duration, fundamental frequency, amplitude, and formant frequency. For the deaf speaker, two questions are of interest. First, what acoustic cues does a deaf speaker use to convey contrastive stress to the listener and how do these cues compare to those used by the normal speaker? Second, can productions perceived as being incorrect in the speech of the deaf be explained as differing systematically from those utterances perceived as being correct?

If stress may be conveyed at least in part by differences in vowel duration, we might expect that for "perceived correct" utterances in the speech of the deaf, the stressed vowel would be longer than the unstressed vowel. Conversely, "stress error" utterances may be due, in part, to an inappropriate vowel duration ratio.

The measurements of vowel duration show that the deaf speaker was like the hearing speaker in some ways, but not in others. Figure 1 shows the measurements of vowel duration for the hearing speaker (FBB) and the deaf speaker's "perceived correct" utterances (MH↑) and "stress error" utterances (MH↓). Dark bars represent stressed vowels; open bars represent unstressed vowels. As expected, overall duration of the vowels produced by the deaf speaker was considerably longer than that of the hearing speaker.

For the hearing speaker, there is always a shift towards longer relative duration for a vowel when it is stressed than when it is not, although this pattern is apparently complicated by differences in intrinsic vowel duration in that productions of /艾滋/ are in general longer than productions of /艾滋/ in the same phonetic environment. An acoustic analysis of a second hearing speaker shows less effect of intrinsic vowel duration. However, the deaf speaker did not show consistent differences in intrinsic vowel duration between /艾滋/ and /艾滋/ within the same phonetic context.

On average, the deaf speaker appears to be conveying contrastive stress by varying vowel duration in the sense that intended stressed vowels were always longer than unstressed vowels in the same utterance, and across utterances. For example, in the utterance /'ai-ə/, when perceived as intended (↑), the average duration of /艾滋/ was 334 msec; in the contrastive pair /'ai-ə/, when /艾滋/ was not stressed, its duration was 267 msec. The same pattern—stressed vowels longer than unstressed—holds for all vowels perceived as correct. However, we find nearly the same pattern for "stress error" utterances. That is, when an unstressed /艾滋/ was perceived in the first contrast /'ai-ə/, the duration of the /艾滋/ was 380 msec, and when a stressed /艾滋/ was perceived in the contrastive pair /'ai-ə/, the /艾滋/ was 285 msec. Thus, the same pattern of vowel durations was found in both "perceived correct" and "stress error" utterances.
Figure 1. Mean duration of vowels for the hearing speaker and the deaf speaker.
In Figure 2, the data show the mean vowel durations and their standard deviations. The durations of the hearing speaker's utterances show very little variability, as reflected in the small standard deviations. In contrast, the deaf speaker was exceedingly variable. Standard deviations were fairly large for the deaf speaker and vowel durations for correct and incorrect utterances often fell within the same range.

The data in Figures 1 and 2 suggest that the deaf speaker is not conveying stress contrasts primarily by differences in vowel duration and also that perceived stress errors are not due simply to a consistently used incorrect pattern of duration. Instead, it would seem that the deaf speaker learned the stress rules of relative vowel duration but is unable to use them to produce an acoustically constant output.

Figure 3 shows measurements of fundamental frequency ($F_0$) obtained from extracting individual pitch periods from the middle portion of each vowel and calculating the frequency from the period. In making these measurements, we noted frequent abnormalities of the waveform. For the hearing speaker, $F_0$ is higher for stressed than for unstressed vowels, as expected. For the deaf speaker, $F_0$ is higher for the intended stressed vowel in three of the four utterance types, but for /ə-ɪ/, $F_0$ is slightly lower for the intended stressed vowel in both "perceived correct" and "stress error" utterances. Again, as with duration, patterns are the same for "perceived correct" and "stress error" utterances.

In Figure 4, the data show mean $F_0$ and its standard deviation. For the hearing speaker, the standard deviations are small, again reflecting little variability. Obviously, the standard deviations for the deaf speaker are large, indicating that the utterances were quite variable. Again, these data suggest that perceived errors are not due simply to a consistently used incorrect pattern of $F_0$.

Figure 5 shows measurements of the amplitudes of the vowels relative to a standard, the first production of an unstressed /ə/ in the utterance /ə ˈpipə/. For the hearing speaker, not surprisingly, stressed /ə/ had greater amplitude than stressed /i/ and the amplitude of a given vowel increased with stress. For the deaf speaker, the stressed vowel always had a higher amplitude than the unstressed vowel. But again, it is clear that this deaf speaker is not conveying contrastive stress to the listener by differences in relative amplitude since "correct" and "incorrect" productions show the same pattern.

Another way in which stress change may be conveyed acoustically is by differences in vowel color. Fry (1964) has shown that listeners are more likely to perceive a syllable as unstressed if the formant values are less extreme, or more like the neutral schwa. Physiological explanations for the effect have been proposed by Lindblom (1963) and by Harris (1978). Without going into the details, it should be noted that the Harris study included measurements of productions of the same disyllables by the same speaker, FBB. We therefore measured the values for the deaf speaker, as presented in Table 2. The results show neither a consistent pattern overall, nor a systematic difference between "correct" and "incorrect" utterances. However, it should be noted that measurements were extremely difficult to make either because of
Figure 2. Mean and standard deviations of vowel duration for the hearing and deaf speaker.
Figure 3. Mean fundamental frequency ($F_0$) for the hearing and deaf speaker.
Figure 4. Mean and standard deviations of $F_0$ for hearing and deaf speakers.
the mismatch between spectrograph filter and fundamental frequency (cf. Huggins, 1980), or because of source function abnormalities.

This deaf speaker appears, at least on average, to have learned some rules for conveying stress increase: vowel duration longer, \( F_0 \) higher, and amplitude higher. Furthermore, it is not likely that these were specifically included in this deaf speaker's training program since theoretical discussions of suprasegmental production at this level are relatively recent in the literature on training deaf speakers. More likely, this speaker has extracted this information from her low frequency residual hearing and then generalized it to abstract rules. However, the variability in her production suggests an inability to coordinate the production mechanism so as to achieve these stress contrasts in a consistent acoustic manner. Furthermore, although she communicates the information that should allow listeners to judge stress, they evidently cannot use it.

**EMG Results**

The electromyographic (EMG) results were examined to see if they revealed any systematic differences between normal and deaf interarticulator programming, or between correctly and incorrectly perceived utterances. In these utterances, orbicularis oris (OR) activity is associated with pursing and closing the lips as for the /p/. For the vowel /i/, the genioglossus (GG) bunches the tongue and brings it forward in the mouth (Raphael & Bell-Berti, 1975; Raphael, Bell-Berti, Collier, & Baer, 1979).

Figure 6 shows data for the hearing speaker producing the utterance type /ə/ papip/. At the top of each column (genioglossus at the left, orbicularis oris at the right) is the ensemble average of the EMG waveforms. This was obtained by rectifying and integrating the EMG potentials for each repetition and aligning them with respect to an acoustic event. The signals were digitized and the ensemble average calculated by averaging each sample for each repetition of an utterance type (Kewley-Port, 1973). A sample of four of the 20 repetitions is seen in the columns below the average. For this utterance type, the line-up point for averaging the EMG and acoustic events, indicated by the vertical line at 0 msec, is the release burst of the second /p/.

The data for orbicularis oris show three well-defined peaks of activity corresponding to the lip gestures for the three /p/ closures in /ə pəˌpip/. The line-up point falls between peaks 2 and 3. The duration of the interval between peaks 1 and 2 is greater than that between peaks 2 and 3, reflecting the longer duration of the /ə/. One notable feature of these data is the striking similarity of the EMG patterns for all tokens. For the genioglossus, there is a peak of activity for the /i/ and no activity for the /ə/ as expected, since the genioglossus is active in raising and bunching the tongue. Indeed, peak genioglossus activity (for the vowel) occurs approximately at the time of the acoustic line-up event—the /p/ burst-release. This is not surprising since EMG activity precedes the articulatory event to which it is related by about 50-100 msec.

Figure 7 shows data for the utterance /ə pəˌpip/ again for the hearing speaker. The interval between the second and third peaks of orbicularis oris
Figure 6. /'papip/ as produced by a hearing speaker. Data plots at the top show the EMG averaged for about 20 tokens for the genioglossus and orbicularis oris muscles. Four individual tokens are shown below. The vertical line indicates the acoustic release of the /p/ closure.
Figure 7. /əpə'pip/ as produced by a hearing speaker. Data presented as in Figure 6.
activity is greater than that between the first and second peaks since the vowel in the final syllable is longer. Also, the duration of genioglossus activity is longer in this utterance type, since /i/ is stressed. Note, however, that peak activity for the genioglossus still occurs at the release of the second /p/, between peaks 2 and 3. Once again, the pattern of activity for all these tokens looks remarkably similar.

Figure 8 shows parallel data for several of the deaf subject's productions of /ə'pə'pip/. Each of these tokens was a "perceived correct" utterance. Examining the EMG activity for orbicularis oris we see that, as for the hearing subject, there are three well-defined peaks of activity and the interval between the second and third peaks is greater than that between the first and second peaks. However, the duration of each peak is prolonged. The /p/ release falls between the second and third peaks as for the hearing speaker.

Turning to the genioglossus EMG, peak activity is less well defined and occurs later than for the hearing speaker; it follows /p/ release. Further, there is considerable variability from token to token in the duration of genioglossus activity. In some instances, this activity starts fairly early (token 3) and at other times, later (token 4).

Figure 9 shows the data for the deaf speaker's production of /ə pə'pip/. Here again, the overall duration of EMG activity is prolonged for both muscles, but the pattern more closely resembles that of the hearing speaker for orbicularis oris than for genioglossus. The variability and "lateness" of the genioglossus are again observed. These data show that the deaf speaker was somewhat like the hearing speaker with respect to "the visible aspects of articulation," but quite variable with respect to the timing of lingual control. This variability appears to be particularly manifested in what we would describe as abnormal interarticulator coordination. To illustrate this notion further, the data for selected tokens of orbicularis oris and genioglossus were plotted.

For purposes of comparison, Figure 10 shows the averaged EMG activity for these muscles for the hearing speaker. Onset of the genioglossus activity is closely coordinated with the second peak of orbicularis oris activity. Shifting of stress from the first vowel (Fig. 10a) to the second vowel (Fig. 10b) does not disrupt this temporal relationship. Indeed, this closely timed interarticulator relationship has been shown for several other hearing speakers (Tuller & Harris, 1980).

Figure 11a shows one of the tokens, perceived as correct, that most closely resembles those of the hearing speaker. Peak genioglossus activity occurs between the second and third peaks of orbicularis oris activity, but the peak is late relative to the acoustic event. Timing between the articulators differs from the hearing speaker in that genioglossus activity begins after the second orbicularis oris peak occurs, and continues well into the third burst of orbicularis oris activity.

Figure 11b shows a token perceived as a stress error. Genioglossus activity begins quite late relative to orbicularis oris activity, and in fact, it peaks simultaneously with the third orbicularis oris peak. This pattern was never seen for the hearing speaker.
Figure 8. /ə'pæpip/ as produced by a deaf speaker. Data presented as in Figure 6.
Figure 9. /əpaˈpip/ as produced by a deaf speaker. Data presented as in Figure 6.
Figure 10. Ensemble average of the EMG potentials for genioglossus and orbicularis oris for the utterance type /eəpəpɪp/ produced by the hearing speaker. The vertical line indicates the acoustic release of the /p/ closure.
Figure 11. A single selected token of the EMG potential from the genioglossus and orbicularis oris muscles as produced by the deaf speaker. The vertical line indicates the acoustic release of the /p/ closure. In Figure 11a, peak genioglossus activity occurs between the second and third orbicularis oris peaks, but is late relative to the acoustic event. This pattern was most like normal. In Figure 11b and 11c, the single tokens show that genioglossus activity was either too late or too early respectively. (N.B. Single tokens filtered with settings used for the average in Figure 8.)
Figure 11c shows another token perceived as a stress error. Genioglossus activity begins too soon in this case, although a peak occurs between the second and third peaks of orbicularis oris activity. However, the genioglossus activity continues beyond the final burst of orbicularis oris activity.

Figures 12a and 12b show respective examples of: (1) a perceived vowel error, and (2) an instance in which there was inappropriate genioglossus activity for the /ɑ/, but listeners perceived the vowel as correct. These two final examples were quite unusual with respect to the normal. It should be emphasized that while there was substantial token-to-token variation in the deaf speaker, the types of physiological patterns do not differ systematically from "correct" to "incorrect" tokens.

**DISCUSSION**

While this study obviously does not allow definitive answers to questions about other deaf speakers, it does suggest some further directions for research. First, these results give ample evidence of the instability of deaf production. The speaker does not produce a "wrong" pattern in a stereotyped way; rather, production is variable in all acoustic and physiological measurements we examined. If the results for this speaker are replicated in further work, we cannot assume the deaf speaker simply operates in a reduced or deviant phonological space, whether the distortion of phonology is produced by explicit teaching or some other aspect of the speaker's experience. While the instability has been noted in transcription studies (e.g., Oller & Eilers, in press), it is better documented by studies that go beyond traditional techniques (Fisher, King, Parker, & Wright, in press).

At a segmental level, there is an apparent failure of consistent interarticulator programming. Overall, a tight temporal coupling of activity in articulatory muscles is lacking. For the normal hearing speaker producing a stop consonant-vowel syllable, activity of the tongue muscles for the vowel is well underway when acoustic release for the stop takes place—this may not be so in this deaf speaker. However, the more important difference between deaf and normal subjects is that the relationship between lip and tongue activity varies from token-to-token in the deaf speaker. It is interesting that the variability of the relationship arises from the lingual rather than the labial component—that is, it is the invisible rather than the visible aspect of articulation that varies.

The second hypothesis about deaf speech, described above, is that the tongue is relatively immobile in this group, as inferred from acoustic measures of formant positions, and this contributes to the unintelligibility of the speech (Monsen, 1976). This hypotheses is, in some sense, an extension of the common observation that deaf vowels are neutralized. When we examine our deaf speaker's data, we note that she is capable of contracting an appropriate muscle for /i/, and leaving it relatively inactive for /ɑ/. Thus, the tongue cannot be in the same position for the two vowels. Of course, the present EMG technique cannot be used to ascertain absolute tongue position. The absolute level of EMG activity is not interpretable, since, in addition to the relative strength of muscle contraction, the amplitude of recorded EMG activity reflects the distance of the active electrode from the firing muscle. 328
Figure 12. Figure 12a shows an example of a perceived vowel error, with genioglossus activity occurring between the first and second orbicularis oris peaks. This token was perceived as /əpipip/. Figure 12b shows an example of an utterance perceived as correct although genioglossus activity clearly occurs between the first and second orbicularis oris peaks as seen above. Data after Figure 11.
fibers. With respect to the vowel neutralization, we note that her formant values for /i/ and /a/ are more similar to each other than those for the "average" female speaker of Peterson and Barney (1952).

A third hypothesis about deaf speech is that source function control is a substantial source of unintelligibility. The present speaker apparently knew the rules for conveying stress by varying F0, duration, and intensity, even though she showed the characteristic overall durational lengthening of deaf speech. What is puzzling is that listeners were not able to extract this information from the signal, as shown by the similarity of "correct" and "incorrect" tokens in acoustic measures. We examined the possibility that "incorrect" tokens were those in which conflicting cues were presented, but no such readily apparent pattern emerged. It is possible that the contours of intensity and F0 were abnormal although the syllable center values were in appropriate ratio.

A question we could not answer within the framework of the present study is what contribution source function irregularities may contribute to segmental unintelligibility. The present experiment suggests an articulatory variable, interarticulator timing, which deserves greater attention. However, it would also be interesting to know how much a deviant and inadequate source in and of itself prevents the listener from interpreting the segmental cues that are received, however inadequate they may be. We intend to pursue this question further, by examining simple nonsense syllables within a wider range of phonetic structures, attempting to use various instrumental techniques to manipulate the source function.

REFERENCE NOTE


REFERENCES


Fry, D. B. Experiments in the perception of stress. Language and Speech, 1958, 1, 126-152.


FOOTNOTE

For convenience in the ensuing discussion, we will call the speech characteristic of the group "deaf speech" and for the purposes of the paper, speakers of "deaf speech" will be called deaf. By making this identification, we wish to acknowledge the fact that persons who are severely to profoundly hearing impaired do not necessarily produce this characteristic speech.