The effects of a long-term hearing loss on speech production

Plant, G.

journal: STL-QPSR
volume: 24
number: 1
year: 1983
pages: 018-035

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

A. THE EFFECTS OF A LONG-TERM HEARING LOSS ON SPEECH PRODUCTION

Geoff Plant

Abstract

Analysis of the speech of an 18-year-old female speaker deafened at age 9 revealed a number of deviations from normal speech production. These included restricted variation in fundamental frequency (FO) contour, absence of FO declination, the use of non-frequency cues to signal emphatic stress in sentences, a slow rate of articulation due to vowel elongation and a restricted range of vowel formant values. The implications of these findings are discussed with special reference to the use of audition as a means of monitoring long-term speech production. The need for more extensive studies of the effects of an acquired hearing loss is especially emphasized. Such studies should include a large number of speakers with varying ages of onset of deafness.

Introduction

Although the effects of a congenital profound hearing loss on speech production have been extensively studied (see Osberger and McGarr, 1982, for an excellent review) comparatively little research has been conducted into the effects of an acquired profound hearing loss on speech quality and intelligibility. Zimmermann and Rettaliata (1981) conducted a cinefluorographic study of the speech of a 34 year old male who had acquired a profound hearing loss during his teenage years. The subject's speech differed from that of normal hearing persons in "movement transition time, voice termination time and tongue dorsum-jaw coordination during the VC gesture" (Zimmermann and Rettaliata, 1981; p. 177). Monsen (1981) reported on two adult males deafened at ages 9 and 11. He found that one of the subjects produced fairly normal second formant movements while the other showed "a sharp reduction in second formant movement, such as is so typically seen in the speech of the profoundly deaf" (Monsen, 1981; p. 25). Plant (1983) in a study of a 14-year-old male deafened at age 11 found a deterioration in both speech intelligibility and quality. Deviations from normal speech production included a slower rate of utterance, reduced pitch movement, distortion of vowel length cues, substitution of /tr/ and /dr/ for the affricatives /tʃ/ and /dʒ/ respectively and an apparent decline in the range of second formant values. Recordings of the subject made 2 months and 30 months after the onset of deafness were rated by a panel of normal hearing listeners. The recording made 2 months after onset was rated as normal by the listeners who expressed surprise that the speaker was, in fact, deaf. The recording made 30 months after onset was, however, rated as having a number of deviations from normal speech
production with the listeners judging pitch and intonation as being most affected. It should be noted that the deterioration in speech resulting from an acquired hearing loss occurs slowly over time. This gradual decline in speech production is cited by Zimmermann and Rettaliata (1981) as support for the view that speakers do not rely on continuous auditory feedback of their speech but rather use audition to recalibrate and retune the production system when movements exceed the normal range of variability. The process of deterioration is gradual as "it takes many instances of exceeding the normal range of variability and making errors without knowledge of errors for these patterns to degenerate" (Zimmermann and Rettaliata, 1981; p. 177).

The age of onset seems to be a critical variable in determining the long term effects of an acquired profound hearing loss on speech production. Madison and Fucci (1971) cited in Edmondson (1977) state "that 7 to 9 years may be a critical time related to phonological development and related skills. During this age-time span we expect both articulation development and speech sound discrimination ability reach a point of maturational stability" (quoted in Edmondson, 1977, p. 2). Edmondson (1977) believes that when the acquisition period is over, at around puberty, "that speech production and speech perception become less interactive as processes. At this stage, and hereafter, articulation can proceed largely independently of acoustic monitoring; the monitoring becomes kinesthetic" (Edmondson, 1977; p. 2). Edmondson speculates that as a result the speech quality degradation of persons deafened before and after the onset of puberty will differ both in severity and type. Binnie et al (1982) studied the "phonetic disintegration" of a 5-year-old boy deafened by meningitis. The child's speech intelligibility declined markedly over the observation period of 9 months. The final intelligibility score obtained by normal hearers listening to the child's speech was around 30% which is comparable to the speech intelligibility of congenitally deaf speakers (Osberger and McGarr, 1982). Analysis of the subject's speech revealed a rise in fundamental frequency (FO) with a reduction in its range, a slow tempo with extended vowel duration plus prolongation of consonant articulation and perceptively excess nasalization.

The present study is part of a long-term investigation aimed at acquiring an understanding of the effects of an acquired profound hearing loss on speech production with a view to developing effective rehabilitative measures. Such an investigation necessarily involves the
study of a large number of subjects varying in the age of onset of
deafness and the time elapsed since onset. The long term effects of an
acquired hearing loss should also provide interesting insights into the
nature of auditory feedback mechanisms and may, as Ellmonson (1977) has
suggested, provide information on the nature of articulatory maturation.

Method

SUBJECT

The subject of the present study was an 18-year-old female deafened
by meningitis at age 8. The subject is a native speaker of Swedish who
uses the Stockholm dialect. Pure tone audiometry revealed responses to
sound only for the low frequencies at high intensities, see Fig. 1. It
is probable that these responses result from tactile rather than audi-\niory sensations. The subject reported receiving no assistance from a
hearing aid and relies upon lipreading for speech perception. Following
the onset of deafness the subject was educated at a school for partially
hearing children and a secondary school for hearing children. She
successfully completed her secondary education and intends to commence
university studies in the near future. At the time of this study she
had recently been fitted with a vibrotactile aid (Spens, 1983) and was
receiving instruction in its use as a lipreading aid.

MATERIALS

In order to obtain a sample of connected discourse a recording was
made of the subject reading a short story of 81 words. A recording was
also made of the subject reading the sentence "Uno belânade gården i
Boden" (Uno mortgaged the farm in Boden). The subject was first asked
to read the sentence in a neutral manner and then to produce 4 different
versions in which the words "Uno", "belânade", "gården" and "Boden" were
each stressed in turn. The question form "Belânade Uno gården i Boden?"
was also recorded. The test sentence was chosen as it was previously
Finally a recording was made of the subject reading a list of 37 simple
Swedish words which was designed to provide a representative sampling of
Swedish consonants and vowels (Öster and Mörtony, 1983). The recordings
were made in 2 sessions. A Revox tape recorder was used to record the
short story and the versions of the test sentence. As these recordings
were to be used for fundamental frequency analysis an accelerometer mi-
crophone was attached to the subject’s neck below the larynx. Askenfelt
Fig. 1. The subject's air conduction thresholds.
et al (1980) have reported that the resultant signal is "related to vocal fold vibrations and the sound pressure in the trachea" and "is reasonably independent of the articulation because of the high glottal impedance" (Askenfelt et al, 1980; p. 259). In order to obtain normative data four recordings were also made of 10 normal hearing Swedish subjects (6 male and 4 female) reading the same materials. At the second session the short story and the word list were recorded using a TEAC A-3340S 4-channel tape recorder with inputs from an air microphone plus accelerometers attached to the subject's throat and nose. For comparative purposes a recording was later made of a 17-year-old normally hearing, female speaker reading the word list.

ANALYSIS

The recordings made in session 1 and those of the control subjects were analyzed using a program developed by Ternström (1983). This program plots F0 contours as a function of time, calculates mean F0 and standard deviation and produces an F0 histogram. Individual spectrograms were made for each of the words in the word list using a Voice Print Spectrograph while the outputs from the 3 separate channels were displayed using an ink-jet Mingograph (Svenska Siemens Model 34i) recorder. Spectrograms were also made of the word list recording of the 17-year-old normal hearing speaker.

Results

FUNDAMENTAL FREQUENCY ANALYSIS

Normal hearing subjects

The results obtained for the reading of the short story by the 10 normal hearing subjects are given in Table 1. The results for the male and female subjects are plotted separately and apart from the obvious difference in mean F0 female subjects appear to use a wider F0 range. Bennett and Weinberg (1979) in a study of the sexual characteristics of pre-adolescent voices, have reported that listeners associate monotonicity in intonation patterns with male speakers. Fig. 2 shows the F0 contours of one of the normal hearing female speakers in the production of the 6 versions of the test sentence. This pattern is typical of all of the normal hearing speakers and shows a substantial rise in F0 associated with the stressed word in the sentence. These prominence peaks are in some cases more than 100 Hz above the rest of the sentence's F0 contour. This result is consistent with the findings of Carlson et al
Table I. Results obtained with a group of normally-hearing subjects reading a short passage.

<table>
<thead>
<tr>
<th>Male subjects</th>
<th>Age</th>
<th>Mean F0 (Hertz)</th>
<th>Standard Deviation (Hertz)</th>
<th>Time (Seconds)</th>
<th>Pauses &gt; .20 sec as a % of total reading time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>26</td>
<td>139</td>
<td>21.5</td>
<td>28.6</td>
<td>29.02</td>
</tr>
<tr>
<td>2.</td>
<td>28</td>
<td>131</td>
<td>14.75</td>
<td>30.6</td>
<td>27.48</td>
</tr>
<tr>
<td>3.</td>
<td>27</td>
<td>115</td>
<td>11.19</td>
<td>23.6</td>
<td>18.57</td>
</tr>
<tr>
<td>4.</td>
<td>22</td>
<td>107</td>
<td>9.51</td>
<td>24.2</td>
<td>22.79</td>
</tr>
<tr>
<td>5.</td>
<td>35</td>
<td>125</td>
<td>20.7</td>
<td>25.3</td>
<td>24.17</td>
</tr>
<tr>
<td>6.</td>
<td>30</td>
<td>147</td>
<td>18.5</td>
<td>20.5</td>
<td>14.93</td>
</tr>
<tr>
<td><strong>MEAN</strong></td>
<td>28</td>
<td>127</td>
<td>16.1</td>
<td>25.5</td>
<td>22.8</td>
</tr>
</tbody>
</table>

Female subjects

<table>
<thead>
<tr>
<th>Female subjects</th>
<th>Age</th>
<th>Mean F0 (Hertz)</th>
<th>Standard Deviation (Hertz)</th>
<th>Time (Seconds)</th>
<th>Pauses &gt; .20 sec as a % of total reading time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>17</td>
<td>222</td>
<td>32.2</td>
<td>25.9</td>
<td>32.09</td>
</tr>
<tr>
<td>2.</td>
<td>38</td>
<td>177</td>
<td>35.0</td>
<td>28.4</td>
<td>26.06</td>
</tr>
<tr>
<td>3.</td>
<td>37</td>
<td>215</td>
<td>38.3</td>
<td>24.3</td>
<td>23.69</td>
</tr>
<tr>
<td>4.</td>
<td>40</td>
<td>162</td>
<td>34.5</td>
<td>24.7</td>
<td>23.50</td>
</tr>
<tr>
<td><strong>MEAN</strong></td>
<td>33</td>
<td>194</td>
<td>35.0</td>
<td>25.8</td>
<td>26.47</td>
</tr>
</tbody>
</table>

Table II. Results obtained for the hearing-impaired subject's reading of the short passage

<table>
<thead>
<tr>
<th>Mean F0 (Hertz)</th>
<th>Standard Deviation (Hertz)</th>
<th>Time (Seconds)</th>
<th>Pauses &gt; .20 sec as a % of total reading time</th>
</tr>
</thead>
<tbody>
<tr>
<td>219</td>
<td>11.2</td>
<td>40.65</td>
<td>31.66</td>
</tr>
</tbody>
</table>
Fig. 2. F0 contours for the sentence "Uno belånade gården i Boden" produced by a normally-hearing female speaker.
who found that F0 change and duration were primary cues in the perception of stressed words in sentences. Similar results have been reported by Löfqvist (1975) who concluded that F0 seems to be the primary cue to emphatic stress in Swedish. The question form of the sentence was marked by a rising F0 contour at the beginning of the sentence. The final measure taken was of F0 declination during the production of the test sentence. This was calculated by measuring the F0 frequency at the beginning and the end of each of the 6 versions of the test sentence produced by the normal hearing subjects. The mean fall across the sentences for the male subjects was 18.9 Hz (Ranges 6–27 Hz) with that for the female subjects being 19.75 Hz (Ranges 11–30 Hz).

Hearing impaired subject

The deafened subject’s results for the reading of the short story are given in Table 2. Although the mean F0 is within normal limits the subject appears to use a much more restricted F0 range. This can be seen in the standard deviation of 11.2 Hz which represents a range of only 5% around the mean F0. Another obvious deviation from the results obtained with normal hearing subjects is the time taken to read the passage. The subject’s total time of 40.65 seconds is 15 seconds greater than the means obtained with both the male and female normal hearing speakers. The factors contributing to this relative slowness of articulation are considered in detail later in this paper. Fig. 3 shows the F0 contours for the 6 productions of the test sentence by the deafened subject. It can be seen that there is very little F0 movement during the sentences, and emphatic stress is signalled by pausing after the intended stressed word. Similarly the question form of the sentence is not marked by an initial rise in the F0 contour. The rate of F0 change across the 6 productions of the sentence was calculated and found to differ greatly from that of the hearing subjects. Examination of Fig. 3 reveals that on most occasions the F0 contour rose rather than fell towards the end of the sentence, with the mean final frequency being 5.5 Hz above the starting frequency.

SPECTROGRAPHIC ANALYSIS

The spectrograms obtained from the subject’s reading of the word list were used to calculate the first and second formant values for the vowels. These are presented in Fig. 4 which also gives mean values obtained with 10 normal hearing female speakers of Stockholm Swedish.
Fig. 3. F0 contours for the sentence "Uno belånade gården i Boden" as produced by the experimental subject. See Fig. 2 for letter key. The area of intended stress is marked.
Fig. 4. F1/F2 plots for vowels produced by the experimental subject (x) compared with mean values for 10 normally-hearing, female Swedish speakers (o). Normative values from Stålhammar (1983).
(Stålhammar, 1983). The subject’s first formant values appear to be within normal limits, but the range of second formant values is restricted relative to that of the normal hearing speakers. Comparative measures were also made of the word and vowel duration for the tokens produced by the hearing impaired speaker and the hearing control. These revealed a number of striking differences between the 2 speakers. The mean duration of monosyllabic words produced by the hearing impaired subject and the control subject were 906.5 ms and 629 ms, respectively, while the relative durations of disyllabic words were 950.7 ms and 636.8 ms. Much of this difference is due to vowel elongation on the part of the hearing impaired speaker. Mean vowel durations for the subject and the control respectively were 240 ms and 157.5 ms for initial vowels, 343.75 ms and 193.25 ms for medial vowels and 311.5 ms and 160.75 ms for final vowels. The overall vowel duration for the control subject was 57.33% that of the experimental subject. Other measures taken were the mean duration of stop closures which were 222.8 ms for the experimental subject and 183.9 ms for the control and mean duration of nasal consonants which were 220 ms and 152.5 ms, respectively.

Spectrograms were also made of the experimental subject and the control subject’s reading of the initial phrase (“En pojke kom en dag”) of the short story. These are shown in Figs. 5 and 6 and give a good illustration of the durational differences between the 2 speakers. The time taken for the control subject to produce this phrase was approximately 1.1 seconds while the experimental subject took around 1.9 seconds.

**Nasalization**

Analysis of the Mingograph tracing of the output from the nasal accelerometer showed that in most instances the subject appears to have fairly normal velopharyngeal control. Contextual nasalization of vowels adjoining nasal consonants was apparent but only in a few instances was there evidence of nasalization of vowels in a non-nasal context. There were, however, many instances of substantial nasal emission during the closure for voiced stops.

**Discussion**

The results obtained in this study reveal a number of deviations from normal speech production which presumably result from the acquisition of a profound hearing loss. The hearing impaired subject’s F0 contour is monotonous and lacks the systematic rises and falls which are
Fig. 5. Spectrogram of the phrase "En pojke kom en dag" as produced by the experimental subject.
Fig. 6. Spectrogram of the phrase "In poyke kon en dag" as produced by a 17-year-old normally hearing female speaker.
characteristic of normal speech production. Ladsfoged (1967) has hypo-
thesized that audition is used to monitor intonation, a view disputed by
Mallard et al (1978). They found that auditory masking of normal hear-
ing subjects had little effect on F0 control in conversational speech
and concluded that "intonation patterns are largely open loop in nature
being programmed into the speech production system at a high level and
not subject to control linked to peripheral monitoring" (Mallard et al,
1978; p. 210-211). Studies such as this, however, involve only short-
term deprivation of audition and bear little similarity to the situation
of deafened subjects. It is interesting to note that in Mallard et al’s
study, auditory masking resulted in a higher than normal F0 and greater
F0 variability. In the present study the subject’s F0 appears to be
within normal limits (Stoicheff, 1981) but the range of variability is
restricted relative to that of normal hearing subjects. It would appear
consequently, that the effects of a long term hearing loss are quite
different from those resulting from short term deprivation. The results
obtained in this study tend to support the view of Zimmermann and Ratta-
lata (1981) who see the auditory system being used for calibration and
tuning of speech movements. Thus, F0 control may at first be relatively
unaffected by the loss of auditory monitoring, but in the long term it
will result in a monotonous pitch contour with little frequency variabi-

Another interesting feature of F0 control by the subject is the
lack of declination (downdrift) in her production of the test sentences.
Declination is found in most languages, and is normally used by speakers
to signal clause and sentence boundaries (Ohala, 1978). Ohala (1978)
believes that this gradual fall in pitch across utterances is due to
"active laryngeally-caused changes in vocal cord tension and that it is
not an automatic effect ..... but purposeful" (Ohala, 1978; p. 100). The
findings in this study tend to support this view of declination as a
purposeful change rather than one resulting from purely physiological
mechanisms. It would also seem to follow that this feature of intona-
tion requires auditory monitoring for its long term maintenance.

The subject’s means of signalling word prominence in a sentence also
differs markedly from that of normal hearing subjects. Contrastive
stress was signalled by the subject using means other than frequency
change. This is not to say that the subject is incapable of signalling
contrastive stress but rather that the mechanisms used are not the same
as those primarily used by normal hearing speakers. Gandour and Wein-
berg (1982), for example, have found in a study of contrastive stress produced by alaryngeal speakers, that "fundamental frequency based cues need not always be primary" (Gandour and Weinberg, 1982; p. 356). It appears that the means used by the subject in the present study—duration, pausing and perhaps intensity change—are those which allow some feedback from movement receptors.

The subject’s slow rate of articulation is observable in her reading of the short story, her reading of the test sentences and her productions of the test words. In all these cases the time taken by the subject was one-third to one-half longer that that of the normal hearing controls. This appears to result primarily from vowel elongation with the subjects mean duration for long vowels being 458 ms compared to the hearing control’s 234 ms while the mean durations of the short vowels were 229 ms and 152 ms, respectively. A slow overall speaking rate is a feature of the speech of the congenitally hearing impaired. Osberger and McGarr (1982) in a summary of previous research findings attribute this reduced speaking rate "to the excessive prolongation of speech segments and the insertion of pauses" (p. 241). Osberger and Levitt (1979) found that the primary factor contributing to syllable elongation was vowel prolongation. Binnie et al. (1982) commented on the similarities between the speech of their deafened subject and that of the congenitally deaf. They concluded that it is tempting "to speculate that many of the phenomena manifest in this child’s speech represent his efforts to maximize interoceptive feedback and to slow the tempo of movement in an effort to maintain articulatory control" (Binnie et al, 1982; p. 189). Silverman and Calvert (1978) have suggested that deafened subject’s must learn to make "effective use of kinesthetic cues" (Silverman and Calvert, 1978; p. 394) to compensate for the loss of auditory feedback, but these appear to be a very inferior substitute. Vowel elongation in the speech of the hearing impaired may result from an attempt to maximize available tactile and kinesthetic feedback on tongue shape and posture. Zimmermann and Rettaliata (1981) believe that positioning and co-ordination of the tongue dorsum is biologically dependent upon auditory information. This view is supported by the relatively restricted range of second formant values in the experimental subject’s speech, indicating that tongue fronting and backing have been adversely affected by her acquired deafness. This is consistent with the results reported in an earlier study of the speech of a deafened subject (Plant, 1983).

Although the subject’s speech deviates in many ways from normal, her
overall speech intelligibility appears to be relatively high. Three normal hearing subjects listened to the subject’s recording of the word list and were asked to write down each of the words as they were spoken. The intelligibility scores obtained ranged from 90-100% correct. This is much higher than scores normally obtained with congenitally deaf speakers which are typically around 20% (Osberger and McGarr, 1982). Some caution should be used, however, in the interpretation of the subject’s speech intelligibility. The results obtained were in quiet and it may be that the subject’s speech would be more adversely affected by competing noise than that of normal hearing speakers. A study of the effects of noise on the intelligibility of the speech of deafened speakers is planned in the near future. In real life situations the effects of noise on speech intelligibility may be even more marked. Deprived of auditory monitoring, deafened subjects do not compensate in their speech for changes in environmental noise levels.

The results of this study again highlight the disruptive effect the loss of auditory monitoring can have on speech production. Lane and Tranel (1971) in a study of the Lombard Sign concluded that after the speech acquisition stage is over, "listening and speaking are not interlocked but are quite separate temporally and that we are no more obliged to listen to our own speech than to that of others" (Lane and Tranel, 1971; p. 701). They also stated that touch and bodily movement receptors are "quite sufficient to provide whatever feedback is needed to maintain fluent articulation after experimental or traumatic deafness" (Lane and Tranel, 1971; p. 694). This hypothesis has not been adequately tested as yet for the majority of studies of the speech of deafened persons have been of subjects deafened prior to the onset of puberty (Monsen, 1981; Binnie et al, 1982; Plant, 1983). Only the study of Zimmermann and Rettaliata provides some counter evidence to this claim. As Edmondson (1977) has suggested the study of subjects with differing ages of onset of deafness may serve to provide interesting insights into the processes underlying articulatory maturation. Information on the effects of age of onset is also essential if adequate rehabilitative programmes and aids are to be developed.

Acknowledgements

This study was conducted while the author was a visiting researcher at the Department of Speech Communication and Music Acoustics. Special thanks are due to the Swedish Institute for the provision of a visiting
fellowship. I would also like to gratefully acknowledge the invaluable assistance of Arne Risberg, Angelica Risberg, Sten Ternström, Anne-Marie Öster, Inger Karlsson, Ulf Stålhammar and Tony de Serpa-Leitao. Without their help this study would not have been possible. Special thanks also to Katerina for acting as the subject of this paper.

References


