ACOUSTIC EVIDENCE THAT POSTLINGUALLY ACQUIRED DEAFNESS AFFECTS SPEECH PRODUCTION

R.Cowie, E.Douglas-Cowie and J.Rahilly

Queen's University, Belfast, Northern Ireland

ABSTRACT

31 controls and 23 speakers with severe to total acquired hearing losses were recorded reading a set passage and describing a day in their life. Samples were digitised using 1/3 octave filters. Their output was passed to programs which characterised the signal statistically. Control and deafened speakers showed a range of significant differences. Deafened speakers' average spectra showed an overall upward shift, plus over-concentration of energy and more particularly change around 1-2 kHz. Their F0 showed an upward shift and increased spread, plus sex-dependent changes in tune shape, and their reading showed fewer stretches where F0 inflected more than once. Their speech amplitude showed higher variance than controls'. Rises in amplitude were too protracted, and falls too large.

1. INTRODUCTION

This paper is concerned with the speech of deafened people, i.e. people with postlingually acquired hearing losses.

It is controversial whether acquired deafness leads to speech deterioration. Goehl and Kaufman [2] argued with some justification that studies which claimed to show speech deterioration were inconclusive for various reasons, including subjectivity of measures, small sample size, and lack of adequate controls. We report research which meets those points. It uses totally objective measures, and it compares speech from a substantial sample of deafened people with speech from a similar number of controls.

The approach was prompted by looking at spectrograms of deafened speech. It is often visually obvious that these are abnormal, but hard to pinpoint the problem in terms of local phonetic features. This led us to develop methods which focus on gross statistical attributes of the speech signal. That approach allows us to demonstrate undeniable differences between deafened speakers and controls. Other aspects of our work follow up and provide more detailed, linguistically oriented descriptions.

2. METHOD

The sample consisted of 54 subjects, 23 controls and 31 deafened.The deafened subjects almost all had losses over 80dB in the worse ear, so that the picture was not confused by the less severely affected speech of speakers with milder losses. Subjects were tape-recorded reading a short passage, and describing a day in their lives. This gave a range of styles from formal reading to a more spontaneous style.

Analysis used an ARIEL spectrum analyser housed in an IBM PC. It contains 31 filters with centre frequencies running from 20Hz to 16kHz in 1/3 octave steps, and a 32nd filter for the amplitude of the signal. A signal capture program sampled the output of these filters at 40ms intervals, and stored the results in files. Gain control was adjusted so as to use the full output range of the filters. Amplitude measures are relative to the peak amplitude in a passage (which was set to 100). Hence the analysis cannot address problems with absolute volume. But though these certainly do occur, they were not salient in our speech sample.

The analysis program takes files from the first as its input. The analysis can be thought of as involving three phases. The first extends the description of the signal. The second obtains graphs which summarise some aspect of a signal. The third extracts a range of statistical parameters which are associated with each graph.

The first phase provided four descriptions of the signal. These were the basic spectrum obtained by the filter bank, the trace of amplitude provided by the 32nd filter, and a trace of fundamental frequency. The filters are not an ideal basis for extracting fundamental frequency, but we developed a reasonably robust algorithm. Its output was always checked, and we rejected passages where we were not confident of its output. The fourth description we call a sharpened spectrum. It measures the salience of each point in the spectrum relative to the points immediately above and below it. The value at each point is the value of the corresponding point in the basic spectrum minus a proportion of the values just above and below it.

Most of the graphs generated in the second phase are histograms. Amplitude and F0 contours were also used to generate scattergrams, mainly by plotting each point against its predecessor. This kind of treatment has interesting properties, but it led to few significant results here and so it will not be reported.

In the largest block of histograms each column is associated with one of the frequency channels in the spectrum analyser. The simplest of these show the average level at each frequency in the basic spectrum and the sharpened spectrum, and the peak level at each freqency. More complex descriptions deal with change in the spectrum.

One set of histograms deals with sample-to-sample change. For each channel we obtain a measure which is the average (root mean square) of the differences between each value in the channel and its predecessor. This is done for both the basic and the sharpened spectrum, giving two more derived histograms.

A parallel set of histograms is derived from a measure which we call peak-to-peak change. Roughly speaking, it deals with change between successive syllable centres, whereas the sample-to-sample measure is dominated by change within syllables. The peak-to-peak measure only uses samples where overall amplitude is at a maximum. At each maximum, the value associated with each channel is compared with the value associated with the same channel at the last maximum. The differences between them are used to construct a family of histograms analogous to the histograms for sample-to-sample change.

From these descriptions another set follow. They involve the ratios of different measures in corresponding channels. For instance it is sensible to consider change/average energy: high rates of change in a channel with low average energy mean something different from similar rates in a channel where the signal is generally strong.

Histograms of a different kind were used to summarise the amplitude and F0 traces. Both were again considered on two levels, one based on point by point description and the other based on the identification of higher order structure in the trace.

For amplitude, the point by point treatment generated two histograms. In one, each column showed the number of observations at a particular amplitude. In the other, it showed the number of observations which differed from their predecessor by a particular amount (using signed, not absolute differences). Higher order structure was found by picking maxima and minima in the contour, and looking at the properties of segments which ran from a maximum to the next minimum or vice versa. Histograms were formed specifying the distributions of amplitudes at all inflections, at maxima, and at minima; the distribution of rises in amplitude between points of inflection and the distribution of falls in amplitude between all points of inflection; the distribution of the durations of rises in amplitude between points of inflection; and the distribution of the durations of falls in amplitude between all points of inflection.

For F0, the point by point treatment generated one histogram, showing the number of observations at a particular amplitude. Higher order structure involved two types of limit. The contour was divided into continuous stretches, bounded by intervals where F0 was absent. Maxima and minima were then marked on each stretch. Stretches were then assigned to one of six types: rises, rise/falls, levels, fall-rises, falls, and compound stretches. The last type contains stretches with more than one inflection. One histogram showed the distribution of these types. A second showed the distribution of stretch durations. A third set out the distribution of pitch changes in segments (i.e. the interval from the highest point in each segment to the lowest).

In the third phase statistical parameters were derived from each histogram. To summarise the central tendency and spread of each histogram we calculated its mean, variance, and quartile points. Histograms whose x axis was frequency were also described in another way, by summing the values associated with four frequency bands. These were chosen to span the usual range of F0, F1, F2, and frication respectively, using values cited by Baken [1] to set boundaries (which were slightly different for males and females).

3. RESULTS

Inferential statistics were applied to the measures provided by the third phase to establish where deafened and control speakers differed systematically. Unless otherwise stated all effects reported here emerged as significant effects or interations from analyses of variance with two between variables, sex and hearing level (control or deafened); and one within variable, passage.

3.1 Spectral Abnormalities. Overall, the mean of the spectrum is shifted upwards by about 1/3 octave in the deafened speakers. The deafened also show too much overall change in the spectrum. This is true on any measure of change. More specifically, the deafened show an abnormal concentration of change in the centre of the spectrum. This is shown by the significantly lower variances associated with most of the distributions of change across the spectrum.

The measures which use formant related frequency bands provide more detail.

The F2 band is anomalous on almost any measure. Among the deafened speakers the average energy there is too high, change there is too great on any criterion, and energy is too sharply peaked at any given instant. The effect is particularly marked among females in the reading passage.

In the F1 region, the problem is more restricted. The deafened show excessive rates of change. High change in this region is also consistently associated with the reading passage and with males.

There is a related problem in the F0 region, but once again it is more restricted. With one measure of change, the peak-to peak measure, the deafened show significantly raised change relative to the absolute energy in the region. The measure is also affected by style. In the controls, change is higher relative to energy in free speech than it is in reading. That effect is much less marked in the deaf.

At the other end of the spectrum, the fricative region shows no effect of hearing on any simple measure we used. However the deafened show a high ratio of average to peak energy in the region - that is to say the energy in that region is spread too evenly across time. That is the opposite of the kind of effect that occurred in the F0 band, at the other end of the spectrum.

3.2 The F0 contour. This topic is complicated by problems in extraction. Initially we believed that F0 was showing no large scale abnormalities, but a different picture has emerged from reanalysis using measures which are insensitive to the shortcomings of our F0 extraction.

The median was taken as the most robust index of each subject's central pitch. The table below summarises average values of the medians. Both sex and hearing have significant effects.

Table 1: Averages of subjects' median pitch.

	hearing	deafened
females	185.6Hz	199.7Hz
males	119.7Hz	134.8Hz

As a robust measure of pitch range we took the distance between the lowest observation and the point below which 75% of the observations lay. There is a relatively consistent pattern of increased range among the deafened, and this is mirrored in an analysis of variance which shows a marginal effect of hearing ($0.1>p> 0.05$).

The other abnormalities in F0 involved high order structure. The controls show a marked increase in compound features in the reading passages - that is, there are more stretches where F0 continues unbroken through more than one inflection. This pattern is greatly reduced in the deafened. The natural inference is that they fail to make a style shift towards rather elaborate phrases in reading.

A separate effect emerges from grouping simpler features into those which end with a fall and those which end with a rise. (Levels are ignored). A significant interaction is found between hearing, sex, and feature type. Deafened males use features which end in a rise much less than any other group do, and features which end in a fall much more. It is tempting to link this to the concept of declination as a universal of intonation. However deafened females show too many of both categories.

3.3 Amplitude. The average variance of amplitude was too high in the deaf, particularly in the reading passage. Table 2 shows how variance differs between the two groups.

Table 2: variance of amplitude as a function of hearing and passage.

	read	spontaneous
controls	67	71
deaf	86	79

High variance means that the deafened spent too little of their time at amplitudes which were near their average. Statistics concerned with maxima and minima augment the picture of how this happened.

One way of spending too much time far from the average is to oscillate between extremes. If the deafened did that, then mean amplitude at maxima should be too great and the mean amplitude at minima too low. In fact there was no significant effect of hearing on mean amplitude at either maxima minima. Conversely both variance of amplitude at maxima and variance of amplitude at minima were too great in the deafened. Again, we would expect the opposite if the deafened were simply oscillating between loud and silent.

More detail comes from the properties of the segments between maxima and minima. Overall, the variance of change per segment was too high among the deafened. However that measure combines rises and falls, and they behaved rather differently. There were no significant abnormalities in the behaviour of amplitude change per rise, but both the mean duration of rises and the variance of rise duration were too high among the deaf. This is to say that rises tended to be big enough, but drawn out too long. Conversely, both the mean amplitude change per fall and its variance tended to be too high among the deafened, whereas the duration of falls showed similar means and variances for both groups. This suggests that the deafened tended to make drops in amplitude which were too big, though they lasted about the right time.

Combining these observations, only one obvious explanation for the general high variance of amplitude remains. It is that deafened speakers protract relatively extreme events (vowels at one extreme, pauses at the other) for too long. Subjectively this seems true, but it needs direct confirmation.

3.4 Style shift. We have mentioned some effects which relate to style shift already. Choosing the right register is an important part of speech, and a speaker who cannot do so has a non-trivial problem. There are consistent indications that deafened people have that kind of difficulty, but we will only mention a few.

Among the controls, the variance of amplitude was lower in the reading passage. The deaf reversed that trend, showing slightly more variance in the reading. The controls showed a lower mean change per rise in the reading passage: that effect was minimal in the deafened. We also found style effects in correlations measuring the relationship between change in one segment and change in the next. Among the controls, these correlations were stronger in the reading passage, than in free speech - i.e. volume became less like a sequence of rises followed by similar sized falls. In the deafened, we found the opposite pattern.

4. DISCUSSION

There is promise in the technique of using a battery of statistical descriptors to characterise speech as a distribution of energy, and we are applying it to other domains. In this domain, it makes clear the existence of quite gross abnormalities in deafened speech. It also establishes that deafened speech shows strong common trends: it does not just drift unpredictably and idiosyncratically. This is not universally expected.

The trends which we have reported provide a focus for closer study. Since the concentration of energy around 1-2kHz emerges as a strong trend, looking at possible explanations should be a high priority, as should looking at explanations of the high variance of amplitude. The existence of problems with style shift has clear methodological implications, and since it presumably involves central control, raises theoretical issues. We are following through such questions in more detailed studies.

5. REFERENCES

[1] BAKEN, R. (1987), *"Clinical measurement of speech and voice"*, London: Taylor & Francis.

[2] GOEHL, H. & KAUFMAN, D. (1984), "Do the effects of adventitious deafness include disordered speech?", *Journal of Speech and Hearing Disorders*, *49*, 58-64.