COMPUTER BASED SPEECH TRAINING METHODS FOR THE HEARING IMPAIRED

M.I. Dawson, S. Sridharan

Queensland University of Technology

ABSTRACT - The general characteristics of the speech of hearing impaired individuals are reviewed to gain an understanding of the requirements of an appropriate computer based visual speech training aid. Current speech training aids are reviewed with regard to this and a direction for future research in the field is discussed.

INTRODUCTION

Hearing impaired individuals require special educational services as a result of their loss of auditory feedback in the speech generation process. This education is currently carried out using a mixture of oral/aural methods of communication and some form of manual complement, usually signed English, finger spelling, and lip reading. Research into electronic speech training aids for hearing impaired people has been carried out since the beginning of the modern electronic era (1920's) and since then over 100 different aids have been developed. It is difficult to judge the usefulness of many of these aids because of the general lack of clinical evaluation. Also, most of the aids are one off experimental models and never come into widespread use (Bernstein *et al,* 1988).

This paper will firstly discuss the generic characteristics of the speech of the hearing impaired. It will then go on to discuss the particular speech aids needed to address their disability and review some of the current computer based speech training aids.

CHARACTERISTICS OF THE SPEECH OF HEARING IMPAIRED INDIVIDUALS

The loss of auditory feedback produces a distinct class of deficiencies in the speech of an individual. The various speech deficiencies of hearing impaired individuals may be grouped into five categories: Timing, Pitch, Velar control, Articulation, and Voice quality (Nickerson, 1975).

Timing

Improper timing control is regarded to be one of the major causes of unintelligibility in deaf speech. A major contributing factor to poor timing is inappropriate breath control. The speech rate of deaf persons is much slower and contains more pauses of longer duration often in inappropriate places than in normal hearing speech. The deaf tend to use more breath during speaking than when they are not speaking. They are thus more likely to interrupt the speech flow more often to take a breath. This problem may be linked to their inability to completely close the glottis during voiced sounds.

The training procedures deaf persons use tend to place emphasis on the articulation of individual speech sounds in isolation. This results in connected speech that has minimal differences in the duration of stressed and unstressed syllables. Final position stressed syllables should be three to five times longer than the preceding stressed syllable in normal speech - deaf people have a much smaller ratio.

Pitch

The pitch characteristics of the speech of deaf people exhibit many different extremes. Pitch variation during speaking may be monotonous or excessive. The mean pitch level is often inappropriate and tends to be higher than normal or even falsetto. Terminal pitch rises and falls are often inappropriate or insufficient.

These problems arise from a difficulty in teaching and understanding the concept of pitch and its variation. Deaf people quite often try to alter pitch by increasing volume. They need to *feel* some form of feedback from the speech generation process and as a result employ excessive laryngeal variations leading to excessive pitch change and high average pitch during speaking. This is not surprising, since many hearing pop singers employ the same methods during singing to alter pitch and intonation, resulting eventually in a raspy voice quality and ultimately in corrective surgery to remove nodes from the glottis.

Velar Control

The velum is used to control the air flow between the oral and nasal cavities. The type of hearing loss of an individual tends to dictate the amount of nasality in their speech. People with conductive hearing loss tend to be hyponasal because they can still detect nasalisation due to bone conduction. People with sensorineural hearing loss do not have this form of negative feedback from bone conduction and thus tend to be hypernasal.

Poor velar control leads to confusion of the nasals /m/, /n/, and /ng/ with the nasal stops /b/, /d/, and /g/ respectively. Words with abutting nasal and stop consonants require accurate velar control also. Control of the velum is also difficult to demonstrate because it is not a visible gesture and its movement results in minimal kinesthetic feedback.

Articulation

The articulatory deficiencies of the speech of deaf persons appear to show a strong correlation with the visibility of articulatory movements. Consonant pronunciation from best to worst are: Bilabial, labiodental, glottal, linguadental, linguaalveolar, linguapalatal, and linguavelar. Articulatory movements from best to worst are: glides, stops, nasals, and fricatives (Nober, 1967). Deaf people find great difficulty with the pronunciation of fricatives and africatives (e.g. /s/ and /z/). They quite often fail to distinguish between voiced and voiceless consonants and plosives. Omission of arresting and releasing consonants is a problem also, for example the confusion of the words "sheep" and "cheap", and "share" and "chair".

Voice Quality

The perceived quality of deaf person's speech has been described as tense, flat, breathy, harsh, and throaty, and can be contributed to the relative intensity of F0 and it's harmonics as well as transition gestures from one articulation position to another (Calvert, 1962). Speech volume level is also a problem: too soft, too loud, or varying erratically. This is particularly a problem with sensorineural loss individuals where there is no bone conduction feedback. People with conductive hearing loss tend to speak too softly.

Subjective voice quality is closely linked with intelligibility. The different voice characteristics of a deaf person tend to present too much of a distraction to listeners not familiar with deaf speech. However, an experienced listener can quite often overcome this distraction and become familiar enough with the particular acoustic cues in a deaf persons speech to enable correlation with the acoustic cues in normal speech.

All the particular characteristics of the speech of hearing impaired people can be attributed to the fact that they are trying to operate with a largely open loop control system with very little feedback. What is needed is to provide an alternate feedback loop for the speech centre using one or more of the deaf person's remaining senses. These senses would obviously be either visual or tactual since the olfactory and taste senses have somewhat doubtable bandwidths. This paper is concerned with primarily visual aids and the next section will develop the specifications of the necessary aids given the particular characteristics of the speech of deaf people.

DEVELOPMENT OF VISUAL SPEECH AIDS

We will first look at the particular requirements of a visual speech aid with regard to the five basic speech deficiencies.

Timing

A speech timing display would need to address two fundamental requirements. The first is the timing of articulatory movements within phonemes, i.e. voiced/unvoiced boundaries, onset of nasalisation, fricative/vowel transitions, etc. The second is sentence timing, i.e. speed and rhythm, length of pauses, duration of stressed and unstressed syllables, etc.

Bate *et al*, (1982) produced a useful frication and timing aid display shown in Figure 1. below. In this a band develops across a computer screen in real time. During voiced speech the band is grey, during frication it is chequered, and during silence it is black. There is no discrimination between different voiced or fricative sounds and the training method uses a double-trace, target and attempt mode (student-teacher).

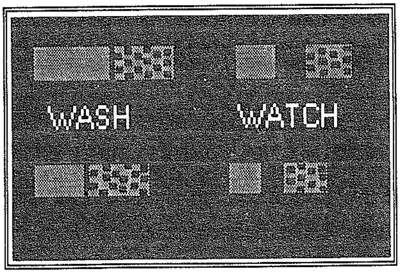

Figure 1. Frication and Timing Aid Display (Bate *et al*, 1982).

This display has been found to be highly useful in the training of the timing of stops (note the period of silence in the word "watch") and in the effect of surrounding phonemes on vowel duration, which are otherwise difficult to indicate and teach.

The format of this display could be expanded quite easily to incorporate a few more acoustic features. Fricative identification is possible in real time (MacKinnon & Lee, 1976) and could be represented as different colours of chequering on the bar. Pitch information could be used to modulate the width of the bar and intensity information could be displayed as a background to the bar or even have a separate bar of it's own. The variations are endless, however, the complexity of the display would be limited by the intended student. Young children and possibly some adults would have difficulty correlating too many sources of feedback. In this case, a modular display would be necessary so that one or two features at a time could be concentrated on, later changing or adding displayed features to work on a particular speech defect as needed.

Pitch

There have been many pitch period indicator displays developed over the years since this has been recognised as a ripe area for improvement of the speech of the deaf. Typical displays resemble games where some object has to be coaxed into a hoop or window by appropriate variations in the pitch of the students speech. Other displays utilise a maze or a tunnel to simulate the appropriate pitch contour which the student has to navigate. An important feature to include in a display of this type would be an accurate indicator of voice intensity since pitch period variations are often confused with intensity variations by the hearing impaired.

Current research at QUT has developed a real time cepstral pitch tracker using a PC based Ariel DSP 16 TMS320C25 processing card. The cepstrum is calculated using base two logarithms and has proved to be quite an accurate indicator of pitch period and voiced/unvoiced transitions.

Articulation and Voice Quality

The development of a visual articulatory movement display is extremely important because current training methods have great difficulty in conveying the many subtle articulator movements inherent in normal speech. Early research in this area employed the use of speech spectrograms as a basic aid (Stewart, 1976). The results were impressive; the formant detail contained in the display provided an accurate model of the speech generation process. However, the conceptual difficulties in using the aid were high. It is difficult to convey the concept of a spectrum to lay-people and children, and the spectrogram is quite often a complex series of cryptic entwined lines.

Crichton and Fallside (1974) used an LPC derived vocal tract display as a speech training aid with reasonable success. A real time cross-sectional representation of the speakers vocal tract from glottis to lips was displayed on a computer screen. The model provided good feedback for vowel production, however, the all pole LPC analysis could not model nasal sounds and had difficulty representing other sounds.

A real time vocal tract model has been developed at QUT using the above mentioned Ariel card. Research is currently looking into the effect of anti-alias filtering, lip impedance, and pre-emphasis on the accuracy of the model.

A real time colour formant vector display was developed by Akira Watanabe in 1985 (Watanabe *et al*, 1985). In this system the lowest three formants were extracted from the speech and were converted to the three primary colour vectors. The resultant colour was plotted as a bar graph on a display, with the length of the bar corresponding to the instantaneous pitch of the speech signal. The display proved to be very useful in the training of connected vowels and has the capability of representing voiced consonants with further development.

Perhaps the most significant development in speech training aids in recent times is the Johns Hopkins Speech Training Aid (Mahshie, 1988). This consists of two related workstations based around IBM PC compatible machines. These are the STS, or Speech Training Station, and the SPS, or Speech Practice Station. Internal signal processing hardware consists of a SKY320 - TMS32010 processor card with a Data Translation DT2821 aquisition card and some signal conditioning hardware to support the external electroglottograph (EGG) and pneumotachograph (PTG) sensors. The SPS is a simpler version of the STS and is meant to be used as a remedial practice station to be situated at the student's home.

The suite of software accompanying the stations consists of mainly speech training drills presented in a game format designed for children between the ages of three and eleven. The speech skills dealt with are sustained vocalisation, production of repeated syllables, and control of voice intensity and fundamental frequency. The software also employs various forms of progressive delayed feedback in the displays in an attempt to wean the students from the aid as their progress improves.

Clinical trials of the training stations to date were concerned primarily with the ergonomics of the aids and were largely subjective. It was found that the aids were easy to use and were accepted well by the children. The primary benefits found were the increased participation of the children in speech training skills, and the ability for unsupervised training.

SUMMARY AND CONCLUSIONS

The loss of auditory feedback produces distinct deficiencies in the speech of an individual that may be categorised under the five areas of timing control, pitch control, velar control, articulation, and subjective voice quality. Computer based speech training aids have been developed for these deficiencies in the past, but have largely been isolated experimental models that have not been adopted to widespread use. The benefits of these aids have not been researched thoroughly in the past and as a result their appraisal has been largely subjective. What is needed to improve the specifications and future development of these aids is a program of thorough formal clinical evaluation. At QUT, we plan to develop a suite of integrated speech aids using signal processing of the speech signal entirely without recourse to expensive sensors such as electroglottographs. This will place the computer based speech training aid within the financial reach of more people.

REFERENCES

Bate, E. M. *et al* (1982) *A speech training aid for the deaf with display of voicing, frication and silence,* Proceedings of the IEEE ICASSP82, pp. 743-746.

Bernstein, L. E., *et al* (1988) *Speech Training Aids for Hearing-Impaired Individuals: I. Overview and Aims* J. Rehab. R&D v25n4 pp. 53-62.

Calvert, D. R. (1964) *An approach to the study of deaf speech,* Proceedings of the International Congress on Education of the Deaf, pp. 242-245.

Crichton, M. A. & Fallside, F. (1974) *Linear prediction model of speech production with applications to deaf speech training,* Proceedings of the IEE Control & Science, v121, pp.865-873.

MacKinnon, D. A. & Lee, H. C. (1976) *Realtime Recognition of Unvoiced Fricatives in Continuous Speech to Aid the Deaf,* Proceedings of the IEEE ICASSP76, pp. 586-589.

Mahshie, J. J. *et al* (1988) *Speech training aids for hearing-impaired individuals: III. Preliminary Observations in the clinic and children's homes,* J. Rehab. R&D v25n4 pp. 69-82.

Nober, E. H. (1967) *Articulation of the Deaf,* Exceptional Children, v33, pp. 611-621.

Stewart, L. C. *et al* (1976) *A Real Time Sound Spectrograph with Implications for Speech Training for the Deaf,* Proceedings of the IEEE ICASSP76, pp. 590-593.

Watanabe, A. *et al* (1895) *Color Display System for Connected Speech to be Used for the Hearing Impaired,* IEEE Transactions on Acoustics, Speech, and Signal Processing, v33n1, pp. 164-173.