Proc. Int. Conf. Recent Advances in Biomedical Engineering, Osmania University, Hyderabad, 6-8 January, 1994

A SPEECH TRAINING AID FOR THE DEAF

N. D. Khambete[1] and P. C. Pandey[2]

1. Sree Chitra Tirunal Institute For Medical Sciences & Technology, Trivandrum.

2. Indian Institute of Technology, Bombay.

ABSTRACT

Prelingual, profoundly deaf children have difficulty in achieving intelligible speech, obvious reason being absence of auditory feedback of their own speech. This paper describes a speech training aid based on visual feedback of speech in the form of vocal tract shape display. A PC with an add-on DSP board, having on-board memory shareable between the dsp chip and the PC, is used as hardware setup. Vocal tract shape is derived from the speech signal using LPC technique. Real-time performance is achieved by first generating the vocal tract image by the dsp chip and then transferring it to the display memory of the PC. Vocal tract shapes obtained for vowels and vowel-consonant-vowel sequences are found closely matching with the actual ones. The system can be further modified by displaying more realistic vocal tract shape and designing appropriate training strategy.

INTRODUCTION

The absence of auditory feedback of one's own speech results in poor development of spoken language. Speech training in such cases can be achieved by providing feedback of speech via an alternate sense modality like vision or touch. Lipreading is simplest and widely adopted natural method of learning speech from visual clues. But the detail positions and movements of major articulators like tongue are not visible and the information about other features of speech is not available. It can be assisted by devices like Upton glasses [1] where miniature lamps mounted on eyeglasses are flashed in synchronism with speech to indicate presence of the lacking features. Spectrographic displays have been used as speech training aids as they are capable of providing more information including pitch variations, formants and their transitions, etc. However, as rightly pointed out by Liberman et al [2], this information is still in uncoded form and hence very difficult to use for correcting errors in speech of the deaf. In their

All correspondence may be addressed to Mr. Niranjan D. Khambete, Scientist, Sree Chitra Tirunal Institute For Medical Sciences & Technology, Satelmond Palace, Poojapura, Trivandrum - 695012.

opinion, since most of the encoding takes place at the conversion of muscle contraction to the vocal tract shape, providing information about articulatory muscle contraction may be much more effective for speech training of the deaf. Recent advances in signal processing techniques has made it possible to derive the vocal tract shape from the speech signal. Early training aids based on this principle displayed a simple line graph indicating variation of vocal tract area from glottis to lips [3]. The recent ones have more realistic display and it is possible to identify different articulators clearly [4, 5]. The most important advantage of these aids is that the display has a direct relation to the actual physical effort involved while producing a particular sound. However, no single system seems to be completely successful as a speech training aid. In an attempt to find out the probable reasons for this, we realize that it is first necessary to establish the reliability of the technique used for deriving vocal tract shape by extensively testing the system for various sounds. The deaf people may not remember the effort they have put in while producing a particular sound even for a short while. Therefore a real-time display may be more useful in achieving better control over the articulatory muscles. Finally it is necessary to investigate whether the information in this form is truly helpful for deaf people in achieving intelligible speech. However, this can be only known after clinical trials. Our aim in this study is to develop a system that has a real-time display as well as slow-motion display of vocal tract shape. An appropriate choice of hardware and software is made to achieve real-time performance. The system has been extensively tested for different sounds and vocal tract shapes are found closely matching to the actual ones. Clinical testing will finally prove the usefulness of the system as a speech training aid.

THEORY

The vocal tract from glottis to lips can be modeled as loss less acoustic tubes of varying cross-section connected to each other as shown in Fig. 1. Transfer function of vocal tract can be considered equivalent to that of an all-pole digital filter at least in case of vowels. Variation in cross-section of vocal tract with respect to length can be derived using Linear Predictive Coding (LPC). The set of reflection coefficients, required for this computation, can be obtained directly from normalized auto-correlation coefficients using Leroux-Gueguen algorithm [6]. This algorithm is suitable for fractional fixed point arithmetic, thus reducing the processing time.

HARDWARE SETUP

The basic factor governing the choice of hardware setup is achieving real-time performance. It is already demonstrated earlier [7] that dsp chip TMS32010 (Texas Instruments) is fast enough to complete the processing in real-time. It is more important to select a configuration that will reduce the time required for transferring processed data to the PC for display. In this context, add-on DSP board PCLDSP-25 (Dynalog Microsystems, Bombay) has an advantage over other systems like stand alone DSP trainer kits because of possibility of fast data transfer via the on-board memory (128 k max.), shareable between the dsp chip and the PC. Thus we can save time required for handshaking involved in other modes of data transfer like serial port or bidirectional parallel port. In addition, the add-on card uses next version of signal processor TMS320C25. As the ADC (conversion time 35 μs) and the programmable timer, for sampling the speech signal, are available on the add-on DSP board, ~~and~~ only the analog circuit (microphone amplifier and antialiasing filter) is to be assembled externally making the system compact. The complete system is in Fig. 2.

SOFTWARE SETUP

The software development work is aimed at working out a scheme that will help in achieving real-time display of vocal tract. The screen indicates an outline of human face as shown in Fig. 3. To achieve the real-time display, the image of the vocal tract (enclosed within rectangular window) has to be updated within the time interval of one speech frame (30 ms), that is before the vocal tract shape data of the following frame is ready for display. The size of this window is governed by two factors; too big the window, more time is required for updating since larger memory block is to be manipulated; too small the window, it becomes difficult to identify individual articulators. A program running on PC is not fast enough to update the image within the selected optimum size window. Therefore, considering much higher speed of dsp chip TMS320C25, it can be expected that a program running on it will be able to update the image within the required time. Thus, the program running on the dsp chip, acquires and processes the speech signal and generates an image of vocal tract on the on-board memory only within 5.2 ms. A program, simultaneously running on PC, that transfers this image directly to the display memory, requires only 20 ms, thus assuring the real-time performance. This program also stores the vocal tract shape data for slow-motion and frame-by-frame display.

TEST RESULTS

Vocal tract shapes obtained for front vowels /e/ and /i/ indicate reducing area towards the lips while those for back vowels /a/ and /u/ have reducing area away from the lips as shown in Fig. 4. Vocal tracts obtained for vowel-consonant-vowel sequences are observed in frame-by-frame mode to locate the transition frames. The location of the minimum area in these frames is found close to the place of constriction for the particular consonant.

CONCLUSION

The configuration of the hardware setup and the scheme developed for display are found appropriate for real-time performance of the system. Vocal tract shapes obtained for vowels are found matching to actual ones, thus proving the reliability of the processing technique used. Presently the vocal tract shape is indicated as sections of cascaded tubes. More realistic display can be obtained by suitably interpolating between the discrete steps. The LPC technique fails to extract information about place of articulation for stop consonants as the speech signal is absent during this short time interval. It may be possible to derive it from the position of minimum area in the transition frames and display it in review mode.

REFERENCES

1. Upton H. (1968). Wearable eye glasses speechreading aid, *Am. Ann. Deaf.*, 113, p 222.
2. Liberman A., Cooper F., Shankweiler D., & Studdert M. (1968). Why are speech spectrograms hard to read?, *Am. Ann. Deaf.*, 113, p 127.
3. Crichton R., Fallside F.,(1974), Linear prediction model of speech production with applications to deaf speech training, *Proc. IEE*, 121, p 865.
4. Pardo J. (1982). Vocal tract shape analysis for children, *Proc. ICASSP*, p 763.
5. Shigenaga M., & Kubo H. (1986). Speech training systems for handicapped children using vocal tract lateral shape, *Proc. ICASSP*, p 637.
6. Leroux J., and Gueguen C., (1977). A fixed point computation of PARCOR coefficients, *IEEE Trans. ASSP*, 25, p 257.
7. Gupte M. (1990). *A Speech Processor and Display for Speech Training of Hearing Impaired*, M. Tech. Dissertation, Dept. Elect. Engg., Indian Institute of Technology, Bombay.

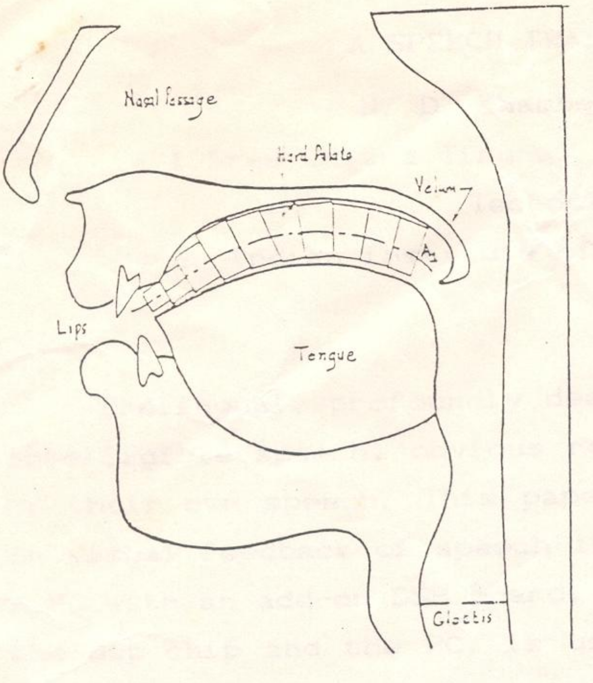

Fig. 1. Acoustic model of vocal tract.

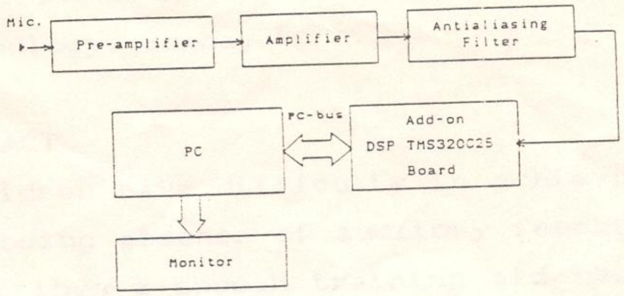

Fig. 2. Hardware setup for speech training aid.

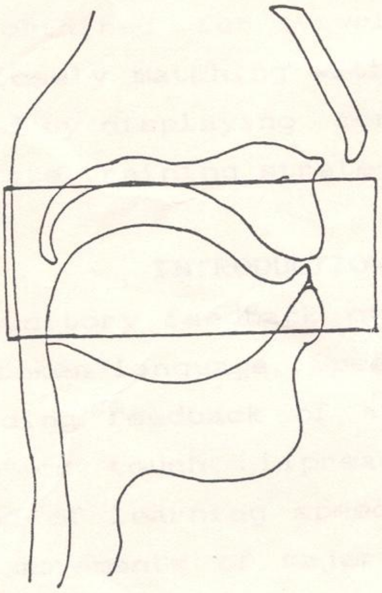

Fig. 3. Selection of appropriate window size for real-time image updating.

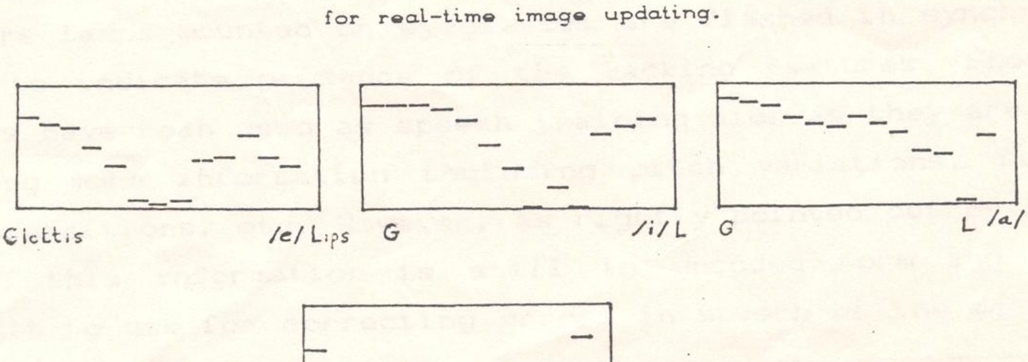

Fig. 4. Vocal tract shapes for front and back vowels.