Veterans Administration

Journal of Rehabilitation Research
and Development Vol. 25 No. 4
Pages 63–68

Speech training aids for hearing-impaired individuals: II. Configuration of the Johns Hopkins aids

JAMES B. FERGUSON, MS, LYNNE E. BERNSTEIN, PhD, and MOISE H. GOLDSTEIN, Jr, ScD
Speech Processing Laboratory, Department of Electrical and Computer Engineering, The Johns Hopkins University, Baltimore, MD 21218

Abstract—Two interrelated personal computer (PC) based speech training aids have been developed: one for use in a school or clinic, the Speech Training Station (STS); and the other for the deaf child's home, the Speech Practice Station (SPS). The STS monitors speech production by microphone, electroglottograph, and pneumotachograph. The SPS system uses only the microphone input. Both systems utilize commercially available board-level hardware and a custom analog preprocessor board for the analysis of the acoustic and/or physiologic inputs. The school system has been used by speech therapists for diagnosis, training by game playing, and specification of exercises for the SPS. The home system provides directed speech practice between therapy sessions.

Key words: *acoustic/physiological signals, computer-based speech training aids, electroglottograph, fundamental frequency, hearing impaired, microphone, pneumotachograph, Speech Practice/Training Stations.*

INTRODUCTION

In an effort to facilitate the speech training of prelingually profoundly deaf children, our laboratory has developed 2 interrelated personal computer (PC) based aids: 1 for the specialist in a school or clinic, the Speech Training Station (STS); and the other, the Speech Practice Station (SPS), for the child's use at home. This paper describes the design of the 2 training aids. There have been previous attempts to improve speech intelligibility through computer technology: Bernstein, Goldstein, and Mahshie (1) review earlier work as well as several other current computer-based aids. Early systems resulted in useful schemes for automatic processing of acoustic, physiologic, and aerodynamic measures. However, they were strictly laboratory systems and not easily transferred to the school or home environment. The aids described here were designed for use outside the laboratory.

DESCRIPTION

Hardware configuration

Figure 1 is a functional block diagram of the STS, showing its major components as configured for the IBM PC/AT system bus. The SPS was designed to be simpler, using only the microphone input.[1] Both systems have incorporated commercially available hardware and a customized analog preprocessor board for the analysis of acoustic and physiologic

Address all correspondence and requests for reprints to James B. Ferguson, MS, Applied Physics Laboratory, Johns Hopkins University, Johns Hopkins Road, Laurel, MD 20707-6099.

Lynne E. Bernstein is currently at the Center for Auditory and Speech Sciences, Gallaudet University, 800 Florida Ave. NE, Washington, DC 20002.

[1]When the STS and SPS were originally designed, the PC/XT was chosen for the SPS, following considerations of cost. Since that time, changes in the PC market have been such that IBM PC/AT's and their look-alikes have become priced so that both systems are now intended for the AT type machines.

Journal of Rehabilitation Research and Development Vol. 25 No. 4 Fall 1988

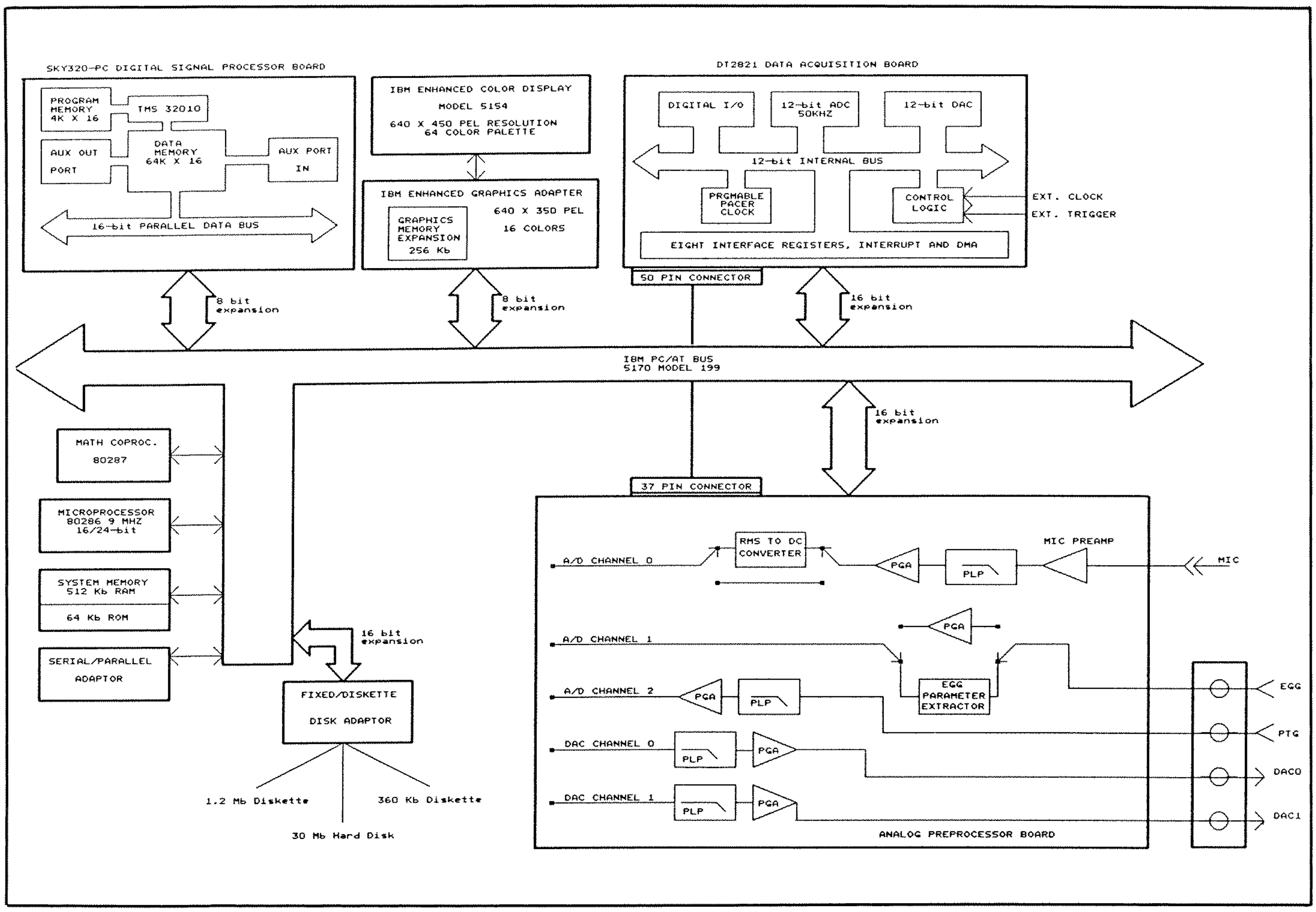

Figure 1.
Functional block diagram of the Speech Training System.

speech signals. The STS system is described below, with comments about differences between the SPS and STS.

Work on the STS made use of an IBM PC/AT running under the PC-DOS operating system. The IBM PC/AT is a single-user, 80286 microprocessor-based, 16-bit computer system that can support multiple expansion boards for specialized applications such as data acquisition, digital signal processing, and display graphics. This machine was chosen as the basis for the STS because of its superior performance, virtually unlimited memory size, and overall compatibility with less powerful systems such as the IBM PC/XT.

The STS uses speech and physiologic signals from a microphone, an electroglottograph (EGG) and a pneumotachograph (PTG). An analog preprocessor board was designed for the system bus to perform various processes associated with the 3 available input channels. The analog preprocessor contains the following unique components: programmable gain/attenuator (PGA) amplifiers, programmable low-pass (PLP) filters, acoustic RMS-to-DC converters, and an EGG parameter extractor circuit. The programmable components on the board can accept a wide range of input signal levels and perform multiple functions on 1 or more input signals. In addition, channels now used by the EGG and PTG can accept other physiologic inputs (e.g., an accelerometer) that may be added to the aid in the future.

The acoustic channel

A functional block diagram of the acoustic channel is shown in **Figure 2**. Acoustic signals are received via an Audio-Technica AT857AM unidirectional, condenser microphone. This microphone was chosen for its excellent off-axis sound rejection at all frequencies, phantom power capability, and small diameter gooseneck design. The microphone preamplifier was designed to maximize the signal-to-noise ratio, maintain a wide dynamic range, and preserve the full available frequency response. The low impedance rating (600 ohms) of this microphone allowed for a transformer-input microphone preamplifier for optimal performance with minimal noise distortion. The acoustic PLP filter serves as an anti-aliasing filter with stopband attenuation down by greater than 51 dB at one-third octave above the selected cut-off. The acoustic PGA stage provides a means for defining an appropriate dynamic range for individual talkers. The signal output from the PGA stage can be directed under program control to either the RMS converter stage or directly to the Data Translation Corporation DT2821 12-bit A/D converter channel for digitization, thus bypassing the RMS converter.

The RMS-to-DC converter was incorporated into the design of the analog preprocessor board for the reduction of incoming acoustic data. The RMS signal is used for real-time speech training of phonatory control of acoustic intensity, duration, and rhythm. The output from the RMS converter goes to an analog switch that isolates the direct acoustic signal from the RMS signal. The output from the RMS converter is typically sampled at 100 Hz because of the low frequency nature of the signal.

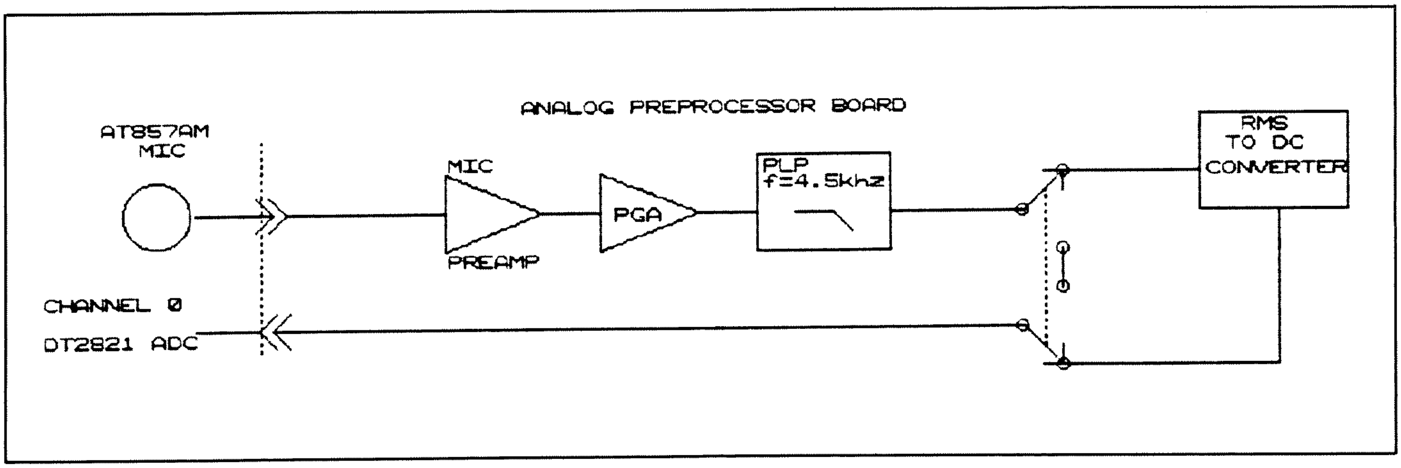

Figure 2.
Functional block diagram of the Acoustic channel.

The electroglottograph channel

A functional block diagram of the EGG channel is shown in **Figure 3**. The EGG is a non-invasive physiologic measurement instrument that detects changes in the amount of contact between the vocal folds (2). Hence, the EGG is useful for obtaining information about voice fundamental frequency (F_o) and the laryngeal dynamics necessary for phonation. The SynchroVoice Research Electroglottograph is currently being used. A spectral analysis of the EGG signal indicates strong frequency components below 50 Hz. This low frequency energy is primarily a result of slowly varying larynx position with respect to the fixed EGG electrodes (i.e., from swallowing or neck movements). In order to effectively reduce the undesired frequency components below 50 Hz, without introducing phase distortion of the desired EGG signal above 50 Hz, the EGG signal is fed to a Glottal Enterprises Linear Phase High Pass Filter (LPHP2). This high-pass filter obtains its high-pass characteristic by subtracting the output of a low-pass forward path filter from the output of an all-pass delayed network (7).

The filtered EGG signal can follow 1 of 2 analysis paths. The first path is used strictly for digitizing the EGG signal. It includes a PGA amplifier stage and a direct path to Channel One of the A/D converter. The second path is used for voice parameter extraction and currently includes 2 modes of operation: one is called the F_0 mode, and the other, the vocal quality mode. The choice of paths is made via a programmable analog switch.

The pneumotachograph channel

The use of a pneumotachograph (PTG) for airflow feedback has been investigated (6). The average airflow signal from the PTG contains laryngeal articulatory information that has been used most extensively for training correct consonant voicing. The Glottal Enterprises Ms-100 Model A2 PTG is a non-invasive instrument which has been successfully integrated into the STS for work with the voiceless segments, such as /h, p, t, k, s, f/.

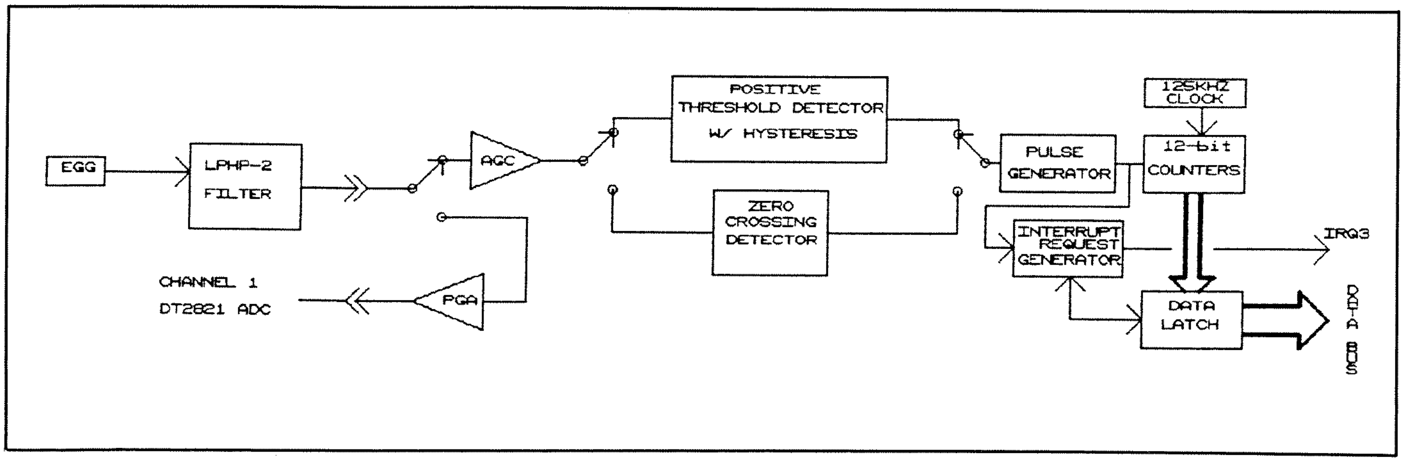

Figure 3.
Functional block diagram of the EGG channel.

The use of the EGG for parameter extraction has been well documented (2,3). The EGG parameter extraction circuit can be programmed for monitoring either fundamental frequency (F_0), or the degree of abduction or adduction (duty cycle) of the vocal folds during voiced speech. Depending on the mode of operation, the filtered EGG signal can pass to 1 of 2 different comparator configurations. The basis of operation of the comparator stage for F_0 detection is, that as the EGG signal rises in voltage (caused by adduction of the vocal folds), the comparator will trigger at a preset voltage (which must be set to a level greater than the ambient noise to avoid false triggers, but less than the minimum peak voltage level to avoid missed triggers). In the second mode of operation, the vocal fold duty cycle is measured from the zero crossings of the filtered EGG waveform.

DATA ACQUISITION

The STS utilizes the Data Translation DT2821 board, designed specifically for the IBM PC/AT computer to perform analog-to-digital conversions (ADC), digital-to-analog conversions (DAC), or digital input and output (DIO) transfers. The DT2821 board can be configured to perform the analog I/O functions in programmed I/O mode or direct memory access (DMA), both with or without interrupt capability. The DT2821 ADC subsystem features 8 differential channels, each with 12-bit resolution, 50 kHz aggregate throughput and a unique RAM channel-gain list which allows any of the channels at any gain to be sampled in any sequence at the full throughput rate.[2]

Digital signal processing hardware

The requirement for complex real-time digital signal processing (DSP) dictated incorporation of a processor in parallel with that of the host computer. The SKY320-PC was chosen for this purpose. It is a fixed-point, digital signal processor board for the IBM PC/AT or XT. The basis for the SKY320-PC's computational power is the Texas Instruments TMS32010 signal processing chip, with an efficient (200 nanosecond cycle-time) instruction set specific for digital signal processing. The TMS320 digital signal processor and the associated SKY320 board are well-suited for the iterative operations necessary to implement real-time speech processing algorithms. The host interface provides direct access to the TMS32010 internal data memory and program memory. The current system configuration of the SKY320-PC board, DT2821, and host PC-BUS are shown in **Figure 1**.

Development work has been completed on 2 signal processing applications utilizing the SKY320-PC system for speech training on the home system. A pitch detection algorithm has been implemented on the SKY320 board for real-time extraction of pitch from the acoustic signal via an external microphone. This algorithm is currently being evaluated within a speech training game for the control of voice pitch. A speaker-dependent isolated word/vowel recognizer is also being developed for the SKY320 board within the speech training system.

The isolated word recognition algorithm relies on real-time parameterization of the acoustic signal based on linear predictive coding (LPC). The segmented and parameterized speech signal is used to train the computer (i.e., to create single-syllable word templates). The actual discrete word recognition is implemented through a dynamic time-warping algorithm, which utilizes a likelihood ratio of LPC residuals as the method of comparing reference templates to the current word productions (4).

System software

Software development has been a principal activity in the development of the STS and SPS. Initial design goals called for a flexible system that is easy for therapists and young children to use. Technical requirements include program portability, real-time applications, and powerful device-independent graphics. Most of the user programs have been written in the C programming language, with support from assembly language when required. The graphics subroutines come from the Media Cybernetics' HALO libraries. A description of the training games that have been developed is in a companion paper in this issue (5).

SUMMARY

Two computer-based speech training aids have been designed, one for use in the home and the other for use in the school or clinic. Design made use of commercial hardware and software whenever possible. A design goal was to achieve a powerful system that can accommodate a range of input sources (such as microphone, EGG, and PTG) and perform complex signal processing: that goal was achieved. Software has been developed for the aids. However, the full potential of the hardware design can only be achieved with the development of much additional software, which is a realistic and ongoing aim.[3] The philosophy of the development effort has been, and continues to be, that software design must proceed in tandem with basic and clinical research on speech production and speech training of the deaf.

ACKNOWLEDGMENTS

The work reported here was supported by NIH/NINCDS (NS-4-2372).

REFERENCES

1. **Bernstein LE, Goldstein Jr MH, Mahshie JJ:** Speech training aids for hearing-impaired individuals: I. Overview and aims. *J Rehabil Res Dev* 25(4), 1988 (this issue).

[2]Original design of the SPS utilized the Data Translation DT2801-A board, which is also a high-speed (27 kHz) data acquisition board specifically made for the IBM PC and PC/XT computers.

[3]Current work on the STS and SPS is being conducted in collaboration with Mr. David C. Coulter with the support of a Small Business Innovation Research Grant from NIH/NINCDS.

2. **Fourcin AJ:** Laryngographic assessment of phonatory function. *ASHA Reports* 11:116-127, 1981.
3. **Hasegawa A, Mahshie JJ, Herbert E, Brandt F, Metzler M, Pickett J:** Real-time extraction of vocal quality parameters from electroglottograph signal. Paper presented at the 107th Meeting of the Acoustical Society of America, 1984.
4. **Itakura F:** Minimum prediction residual principle applied to speech recognition. *IEEE Trans Acoust Speech Sig Proc* ASSP-23:67-72, 1975.
5. **Mahshie JJ, Alquist-Vari D, Waddy-Smith B, Bernstein LE:** Speech training aids for hearing-impaired individuals: III. Preliminary observations in the clinic and childrens' homes. *J Rehabil Res Dev* 25(4), 1988 (this issue).
6. **Mahshie JJ, Herbert E, Hasegawa A:** Use of airflow feedback to modify deaf speakers' consonant voicing errors. Paper presented at American Speech-Hearing-Language Association Convention, 1985.
7. **Ng TK, Rothenberg M:** A matched delay approach to subtractive linear phase high-pass filtering. *IEEE Trans Circ Sys* CAS-29(8):584-587, 1982.