A COMPUTER-CONTROLLED SPEECH SYNTHESIZER

Sotaro Sekimoto

Summary

A speech synthesizer of a more or less conventional terminal analog type for on-line computer control has been constructed as one of the standard research tools for this laboratory. It takes nine control variables for the synthesis parameters, and their time functions as well as the buzz source waveforms are generated by software. The synthesizer is controlled by several simple IOT (Input/Output Transfer) instructions in the same way as other standard input/output devices are. By using this speech synthesizer, some test tapes for A-B and A-B-X listening tests have been produced.

Speech Synthesis System

The configuration of our speech synthesis system is shown in Fig. 1. The PDP-9 central processor unit which controls the synthesizer has 16,384 18-bit words of core memory. The system is equipped with a Control Ball, a Switch Box and a computer-controlled audio tape-recorder. The Control Ball is used for a graphic input of two-dimensional curves for the speech synthesis parameters as time functions, in conjunction with an oscilloscope display monitor. A new computer program for this graphic input system is now under development. The Switch Box has 18 switches, each of which is connected to a bit on the accumulator respectively and is used for synthesis parameter assignment. The audio tape recorder is used to record and reproduce the synthesized speech signal with specified pauses.

Speech Synthesizer

As shown in Fig. 2, the speech synthesizer has a main branch consisting of a series of formant circuits, and a parallel branch for fricative sounds (frication circuit). The main branch (vowel circuit) can be excited with either buzz or noise source.

There are nine control parameters in total. The vowel circuit consists of four pole circuits (P1-P4), which represent the four lowest formants, and a higher pole correction circuit. Of the frequencies and bandwidths of the

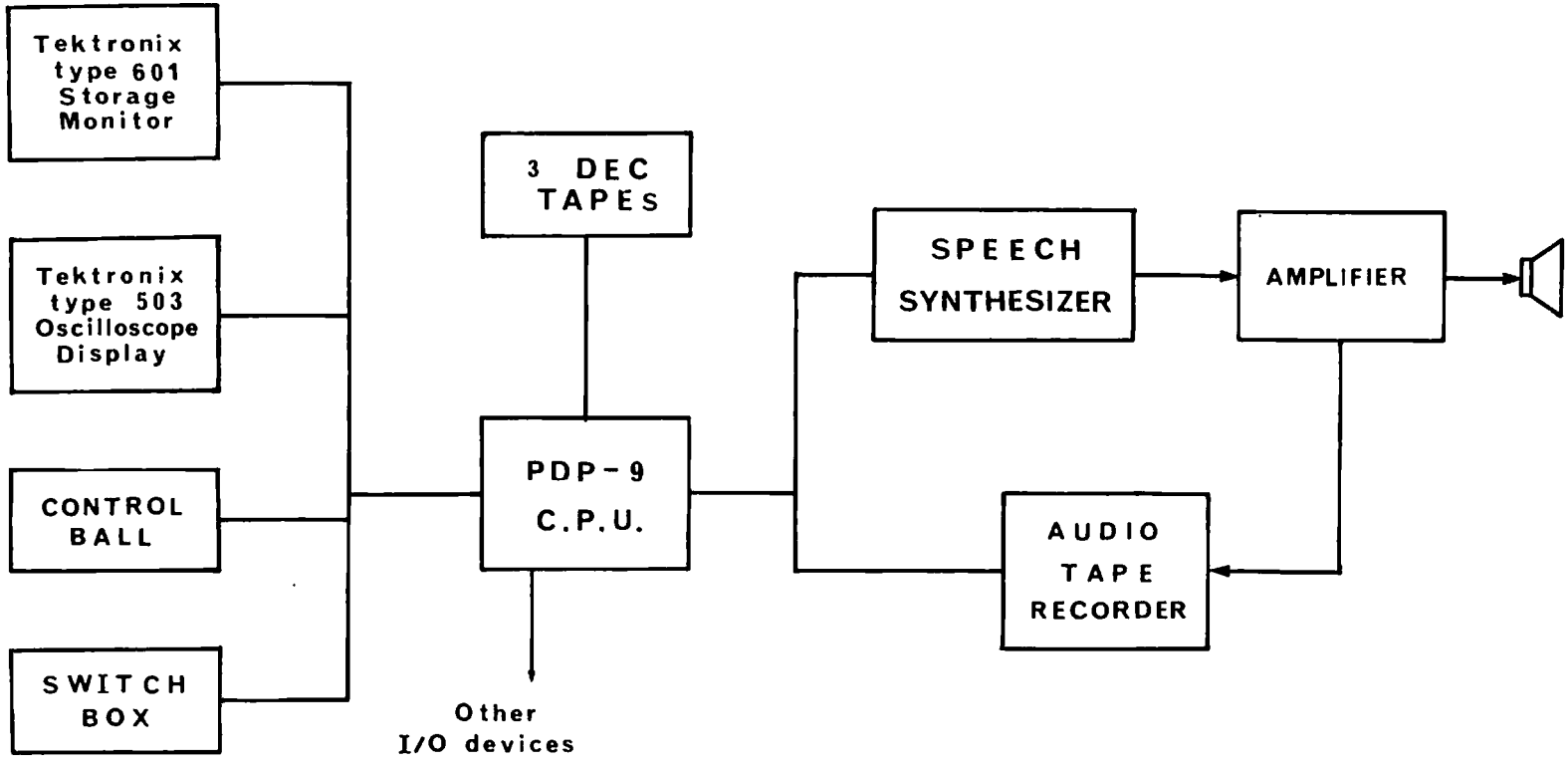

Fig. 1. The configuration of the speech synthesis system.

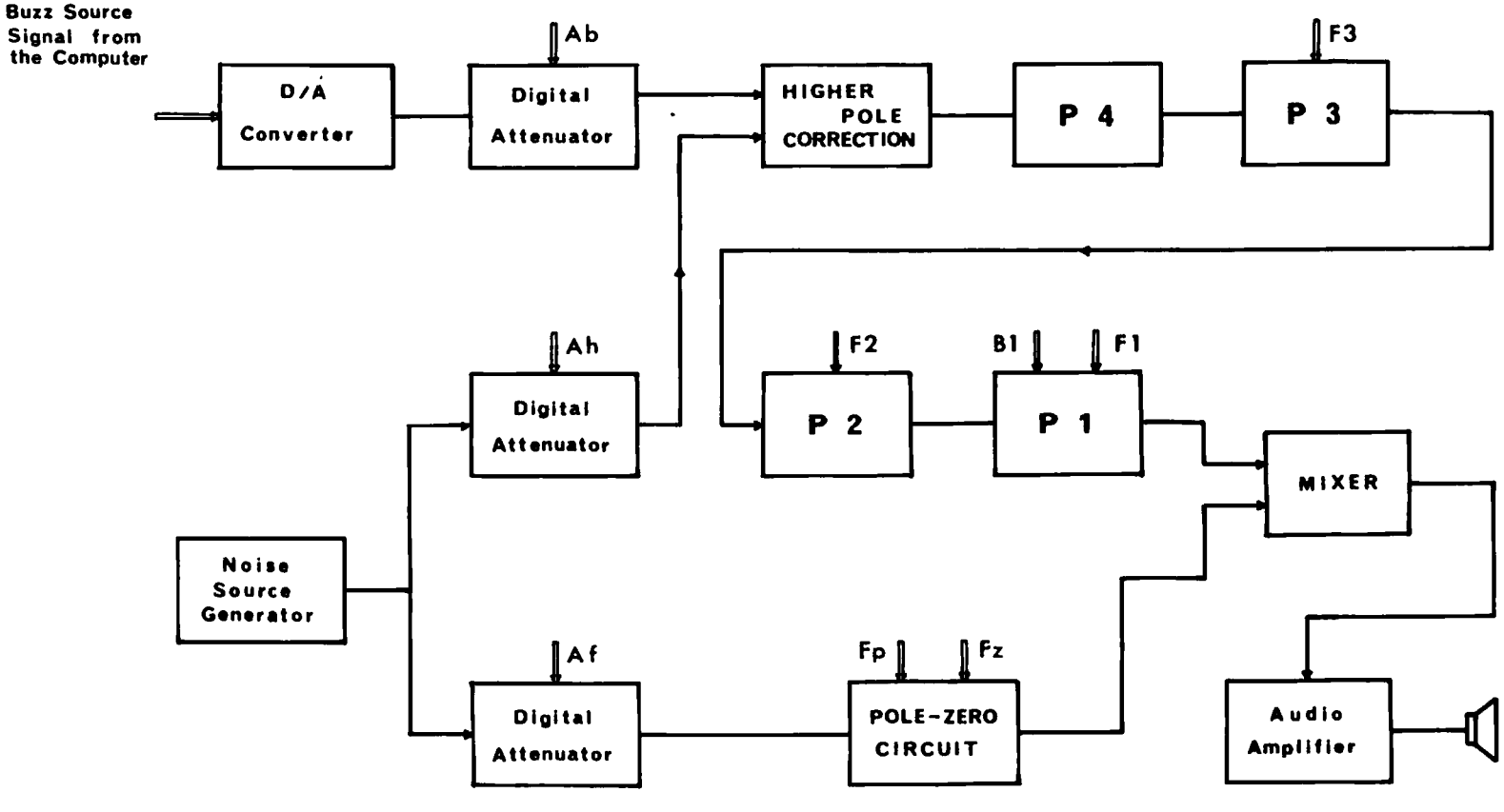

Fig. 2. Block diagram of the speech synthesizer.

four formants, the frequencies of the first, second and third formants (F1, F2, and F3) and the bandwidth of the first formant (B1) are controlled. The last parameter is useful for controlling nasality.

The pole circuit, shown in Fig. 3, is an analog simulator of a second-order differential equation which employs three operational amplifiers. The resonance frequency and the bandwidth (in the case of the first formant) are independently varied by digital control signals. The frequency change is almost linear with respect to the control signal, but in order to obtain satisfactory accuracy, lists for control values are prepared in core memory for the ranges 170 - 1330 Hz for F1, 620 - 3170 Hz for F2, and 1290 - 3850 Hz for F3, in an 18 Hz step for F1, and in a 40 Hz step for F2 and F3. A pole-zero pair circuit for the fricative filter is another analog simulator circuit. The pole frequency and the zero frequency are independently controlled, but no bandwidth control is provided. All the control signals are given in 6-bit quantized values.

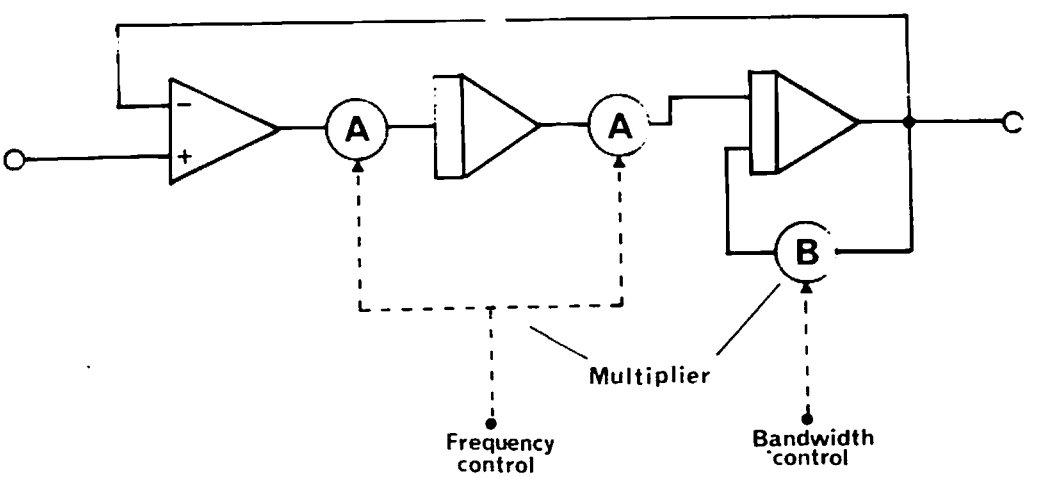

Fig. 3. Block diagram of the pole circuit.

The source waves of the buzz excitation for voiced sounds are generated in the computer and supplied through a D/A converter. In place of the usual radiation circuit, a differentiated waveform is used for this voice source signal. The noise source is included in the speech synthesizer. The noise is generated by a Zener diode which is operated at its characteristic knee point. The amplitudes of the buzz and noise sources are controlled by digital attenuators. The attenuation ratio is varied in a 1-dB step over a 64-dB range according to the 6-bit control signal from the computer.

Control

Each of the nine control parameters has a 6-bit control buffer. The three 6-bit buffers are treated as a single 18-bit register and the content of the 18-bit accumulator is transferred to this register by a single IOT instruction. Thus, three IOT commands are provided for the nine control parameters. In addition to this, the buzz source has a 9-bit control buffer and one more IOT command is used for this parameter. The bit assignment in the accumulator to each control buffer is shown in Fig. 4.

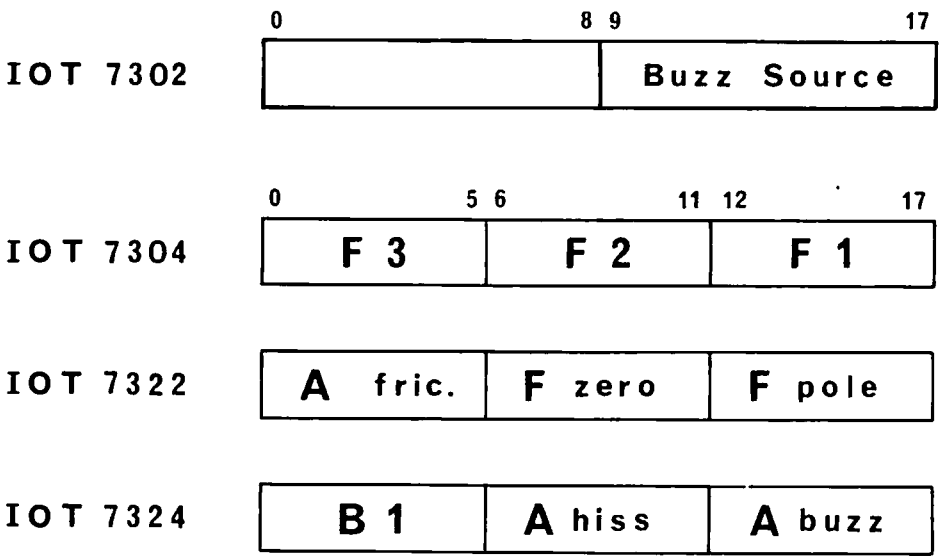

Fig. 4.

A real-time clock is provided in the speech synthesizer to regulate the control timing of the synthesizer, time pulses being produced every 16 microseconds.

Listening Test Program

By the on-line operation of the speech synthesizer, production and recording (for the ordinary magnetic tape recorder) of test stimuli can be automatically performed for producing A-B or A-B-X (or other triplet-type) listening test material. The sequence of the stimuli is controlled by a computer program.

The parameter values for synthesizing each sound stimulus are typed in via a Teletype. Up to sixteen different kinds of stimuli are handled by the present program. First, a linear list of all the possible combinations of the stimulus sounds (in doublets or triplets for an A-B or A-B-X test, respectively) is automatically constructed. Then, the order of the test tokens is randomized by generating a series of random numbers, and for each random number n generated, the test item at the n-th position in the list is interchanged with the one at the initial position. This process is repeated for a sufficiently high number of items. The test stimuli are generated through the speech synthesizer and recorded in the randomized order. After every twentieth stimulus, a peep of about 400 Hz is inserted. The list of the randomized stimuli is printed out on the Teletype.

A triangular source wave is available in the current program for the buzz excitation, its wave form being specified by the opening time, closing time and fundamental period. The amplitude envelope of the sound stimulus has a linear initial rise and a linear final fall with prescribed slopes. The time intervals between the sound stimuli and between the test tokens are given as prescribed. Certain types of fluctuation can be given to the amplitude, pitch, and/or waveform of the source wave if desired.