Speech Synthesis Project*

David Meltzer

*Sponsored in part by the National Science Foundation through Grant GN-534.1 from the Office of Science Information Service to the Computer and Information Science Research Center, The Ohio State University.

Speech Synthesis Project

David Meltzer

The purpose of this project has been to attach a terminal analog speech synthesizer of the Glace-Holmes type (JAWORD) to the PDP-10 computer to allow on-line speech synthesis in an interactive environment.

The JAWORD synthesizer is an advanced terminal analog type with 7 resonators and two forms of excitation (see Figure 1). The frequency and amplitude of each resonator are controlled by an externally supplied voltage of from 0 to +3 volts. Of the 14 parameters thus needed, only 10 are used at any one time; which of these 14 are used is determined by an 11th parameter which also controls which form of excitation is applied to the resonators (noise or pulse excitation). The important characteristics of this synthesizer from the viewpoint of interfacing it to a computer are:

1. Eleven analog voltages must be supplied simultaneously to the synthesizer.

2. A new set of voltages must be available every 10 milliseconds, i.e., the control voltages change at a very slow rate.

Previous experience had shown that the normal configuration of JAWORD did not allow sufficient control of the time of occurrence of the fundamental excitation pulses. The synthesizer was therefore modified so that the pulse could be supplied from an external source, in this case the computer.

The configuration used to control the synthesizer is shown in Figure 2. The digital control information generated by the control program is fed to the synthesizer controller via the PDP-10 Input-Output Bus. This controller, designed and built at Ohio State University, converts this information into the required analog voltages and supplies them to the synthesizer. The controller (see Figure 3a) uses an analog multiplexor feeding a series of capacitive hold circuits to generate the 9 continuously variable outputs from the output of one Digital-to-Analog Converter (DAC). The control word (Figure 3b) includes the information to set the DAC as well as the address of the hold circuit which will receive the value. Although considerably slower than a one DAC pre-channel system, its response is adequate for this application. Additional bits are provided in the control word for control functions, including a bit for the fundamental excitation pulse (1=pulse, 0=no pulse).

The logic of the controller is implemented in Digital Equipment Corporation B-series discrete component logic, the same family as used in the PDP-10 central processing unit (CPU). This family was chosen for reasons of ease of interfacing to the CPU. The DAC is a high precision 10-bit unit also made by Digital Equipment Corporation. The analog hold and amplifier circuits as well as miscellaneous level conversion circuits were designed and built at Ohio State University. The hardware is completely operational on-line with the CPU.

The next major element in the system is the control program for the synthesizer. The program is the basic interface between the user and the synthesizer. This control program operates as a privileged user job within the time-sharing environment so that other jobs may continue to use the system during speech synthesis jobs. There is a definite random degradation in output speech quality due to the presence of other activity in the system during synthesis, but this is not a problem for trial synthesis runs. A future revision of the control program is planned which will optionally turn off the time-sharing for the duration of synthesis output and then return to time-sharing mode. The actual output is of short enough duration so that this will not cause appreciable response time degradation.

The control program as now in operation performs several functions necessary for effective interaction with the experimenter. The primary level controls the operation of the I/o bus and appropriately sequences the outputting of control words to the synthesizer controller. These control words are generated by the next level of program, the conversion routine which translates from a standardized code (Carlson, 1969) to the appropriate channel addresses and DAC level, adds offset and calibration data and packs the words into a table in CPU core memory. This table becomes the input to the first program.

The basic interface to the experimenter is one of the Teletype consoles attached to the CPU. It is anticipated that this function can be taken over by the CRT display at some future date but the lack of adequate systems programming support and hardware character generator feature on the display makes this a difficult task.

Synthesis of a speech sample using the current program involves the following steps.

1. Typing in the sample in coded form and in-core editing it to the experimenter's satisfaction. The program provides for selective display and alteration of portions of the parameter table.

2. Outputting the speech sample and then making needed corrections.

3. Dumping the table in coded form on paper-tape for later use. The paper tape so prepared may be used as input in place of step 1.

The speech output may also be recorded on a built-in tape recorder. Future program versions will allow dumping the output on DEC tape so that many words or phrases may be kept on one tape. At present, however, there are not enough DEC tape drums available to allow each user of the system to have one tape and still let the synthesis job have two, one for program storage and one for speech sample storage.

The program as described above with paper-tape storage is fully operational. Evolutionary changes are being made to improve the output quality as experience with the synthesizer grows. Synthesis parameter tables distributed through SOUGHS (Society of Users, Glace-Holmes Synthesizer) are being used to gain experience with the system.

Reference

Carlson, W. A. 1969, "On the Establishment of a Standard Format for Exchange of Data for the Synthesizer Among the Users," (private communication, distributed through SOUGHS).

Figure 1. JAWORD SPEECH SYNTHESIZER (SIMPLIFIED)

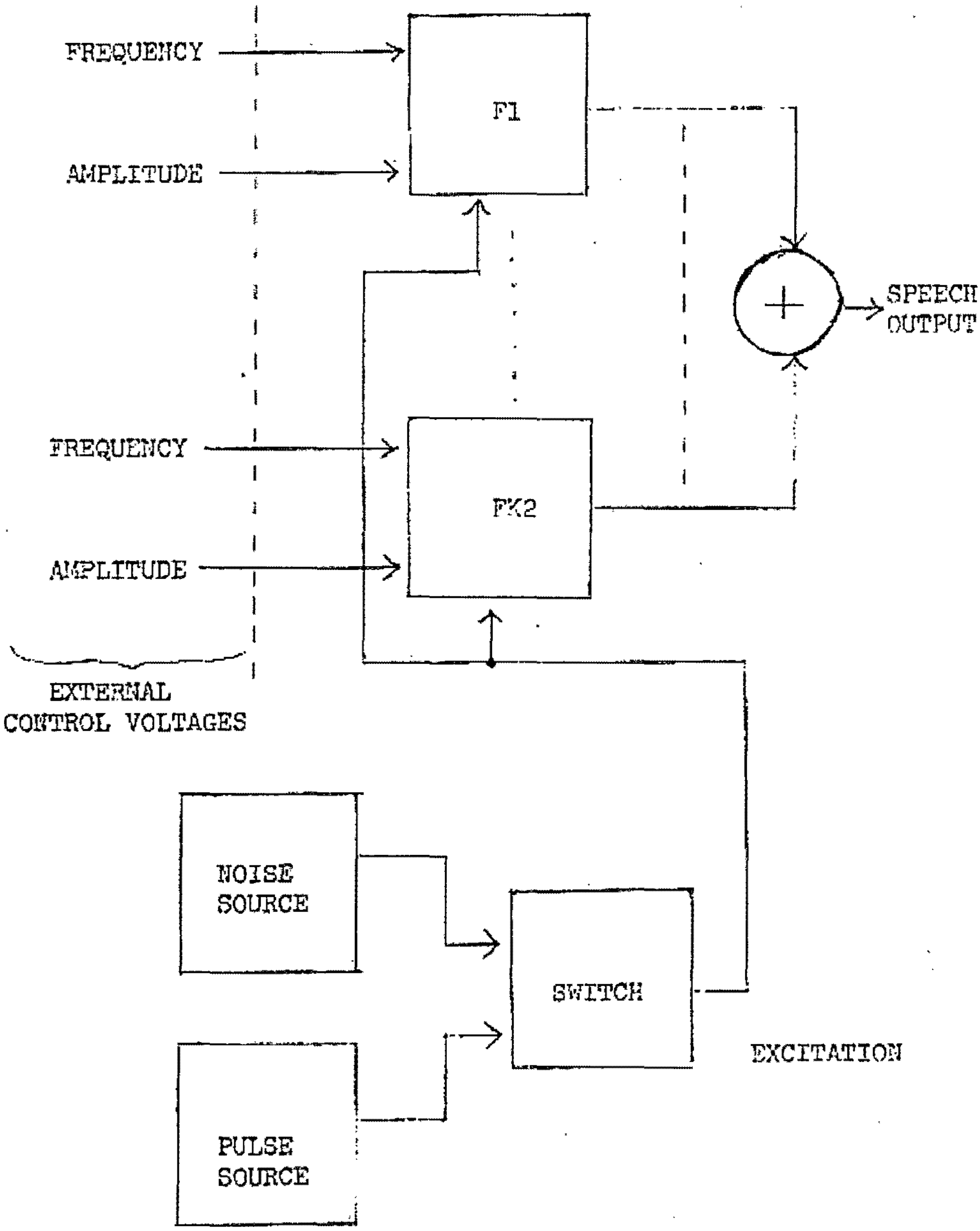


Figure 2. SYSTEM CONFIGURATION

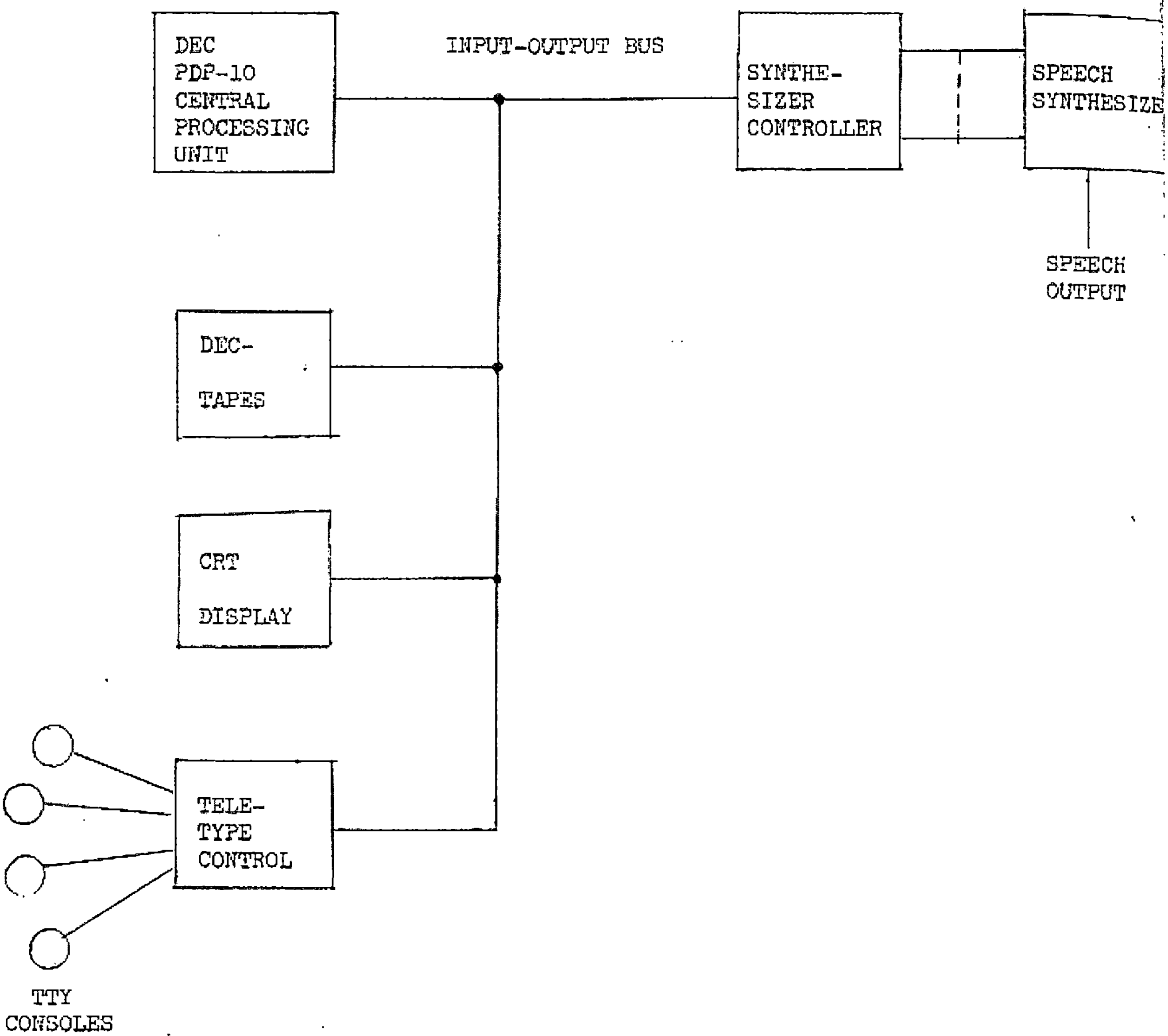
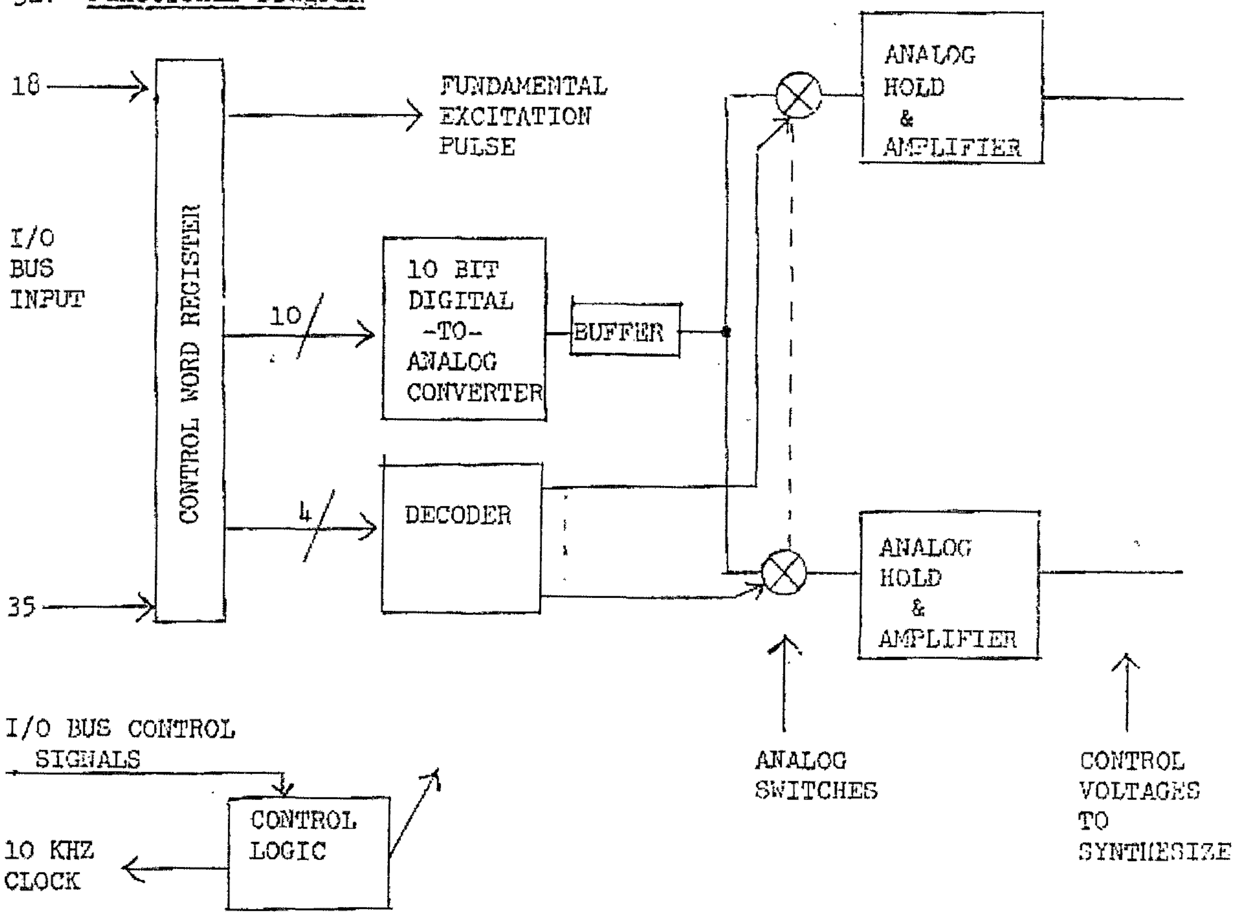


Figure 3. SYNTHESIZER CONTROLLER

3a. Functional Diagram



3b. Control Word Format