XX. SPEECH COMMUNICATION*

Prof. K. N. Stevens	Dr. T. T. Sandel	W. L. Henke
Prof. M. Halle	Jane B. Arnold	A. P. Paul
Prof. J. B. Dennis	J. M. Heinz	J. R. Sussex
Dr. A. S. House		E. C. Whitman

RESEARCH OBJECTIVES

The objectives of our work are to further our understanding of: (a) the process whereby human listeners decode an acoustic speech signal into a sequence of discrete linguistic symbols such as phonemes; and (b) the process whereby human talkers encode a sequence of discrete linguistic symbols into an acoustic signal.

Current research activities related to these objectives include experiments on the generation of speech by electrical analog speech synthesizers, development of means for controlling analog speech synthesizers by a digital computer, measurements of movements of the speech-generating structures during speech production, studies of methods of speech analysis, accumulation of data on the acoustic characteristics of utterances corresponding to phonemes in various linguistic contexts, and studies of the perception of speechlike sounds.

K. N. Stevens, A. S. House, M. Halle

A. A COMPUTER PROGRAM FOR CONTROLLING THE DYNAMIC VOCAL-TRACT ANALOG (DAVO)

A program for the TX-0 computer has been prepared to allow flexible operation of the dynamic analog speech synthesizer constructed by Rosen.[1] The development of this control program has been the subject of a Master of Science thesis.[2] The objective of this work is to replace the control system originally incorporated into DAVO by a more versatile one involving the TX-0 computer and interconnection equipment consisting mainly of digital-to-analog converters.[3]

Control of the vocal-tract analog is accomplished through 24 analog voltages that specify cross-section area as a function of distance along the tract, and 3 voltages that control amplitude of buzz and noise excitation, and coupling between the oral tract and an analog of the nasal cavity. Voicing frequency is controlled through the timing of glottal pulses supplied to the vocal-tract model. The original control system for DAVO employed trapezoidal waveform generators as sources for the control voltages, and was limited to essentially monosyllabic utterances without resort to tape-splicing.

In using the control program to operate DAVO, a number of control signals are specified as sequences of piecewise quadratic segments. The control voltages for the vocal-tract area function are represented as a linear combination of up to three area functions selected from a library of vocal-tract configurations. The coefficients of the linear combination are provided by control signals. One control signal is used within

*This research was supported in part by the U.S. Air Force (Electronic Systems Division) under Contract AF 19(604)-6102; in part by the National Science Foundation (Grant G-16526); and in part by the National Institutes of Health (Grant MH-04737-02).

the program to generate glottal pulses at a corresponding rate, and three others become control voltages for excitation and nasal coupling. The input to the control program is a time-ordered list containing the parameters of each quadratic control-signal segment, the assignment of configurations from the library, and other functions. The input list is processed by the control program, and gives a series of digital samples representing the motion of the control voltages with time. The samples are stored, then read out later to perform the synthesis in real time. The size of the magnetic core memory in the TX-0 computer limits the length of an utterance to approximately 1 sec; however, this will be greatly extended through the use of the digital tape unit that is now available. Initial experiments with the program show that the processing time is approximately 50 times real time when the samples are computed at 1-msec intervals. When the interval is increased to 6 msec, the processing-to-real-time ratio is improved, and is 25 times. New instructions added to the TX-0 computer and some improvements in the program might possibly reduce the ratio to 10.

The internal structure of the program has been designed so that changes and additions to its function can be easily accomplished. It should be pointed out that the program will not become really useful for general speech-synthesis work until a compiler is available to simplify the preparation of input lists and make on-line changes in the utterance conveniently possible. Such a compiler is now being developed.

The control program also employs stored correction curves to correct for irregularities in the characteristics of some components of the analog. At present, a method of automatically calibrating the vocal-tract analog is being investigated. This involves obtaining the correction curves through the observation of available variables in the system, and without having to dismantle or isolate any of its parts.

J. R. Sussex, J. B. Dennis

References

1. G. Rosen, Dynamic Analog Speech Synthesizer, Technical Report 353, Research Laboratory of Electronics, M. I. T., February 10, 1960.

2. J. R. Sussex, Computer Control of a Dynamic Analog Speech Synthesizer, S. M. Thesis, Department of Electrical Engineering, M. I. T., September 1962.

3. J. B. Dennis, Speech synthesis, Quarterly Progress Report No. 67, Research Laboratory of Electronics, M. I. T., October 15, 1962, pp. 157-162.

B. A TRANSISTORIZED ARTICULATORY SPEECH SYNTHESIZER

Work is progressing on the design of an improved dynamic analog of the vocal organs to replace the vacuum tube analog (DAVO) constructed by Rosen.[1]

Like Rosen's device, the new analog will represent the acoustical system of the vocal tract as a lumped-element transmission line in the electrical domain. The theory

of such analogs is treated by Kasowski,[2] Fant,[3] and Stevens, Kasowski, and Fant.[4] Essentially, the vocal tract is considered a cascade of cylindrical segments of common length but variable areas. In the electrical system, in which voltage is analogous to pressure, and current to acoustic volume velocity, each segment is represented by a series inductance and a shunt capacitance, and a cascade of such sections forms an artificial line (Fig. XX-1). It can be shown that

$$A = \frac{\rho c}{k}\left(\frac{C}{L}\right)^{1/2} \qquad \text{and} \qquad \ell = c(LC)^{1/2}, \tag{1}$$

where A is the area of the section, ℓ its length, ρ the density of air, c the sound speed, and k an arbitrary constant of the transmission line, depending upon the choice of units and the impedance level.

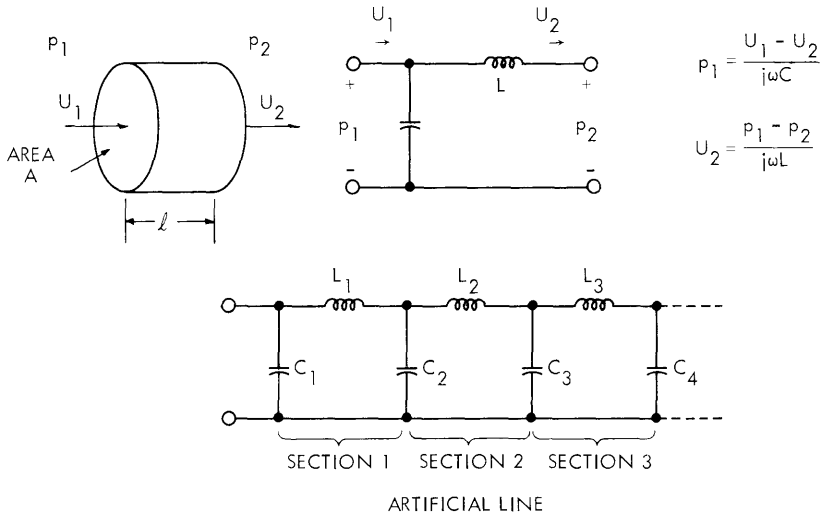
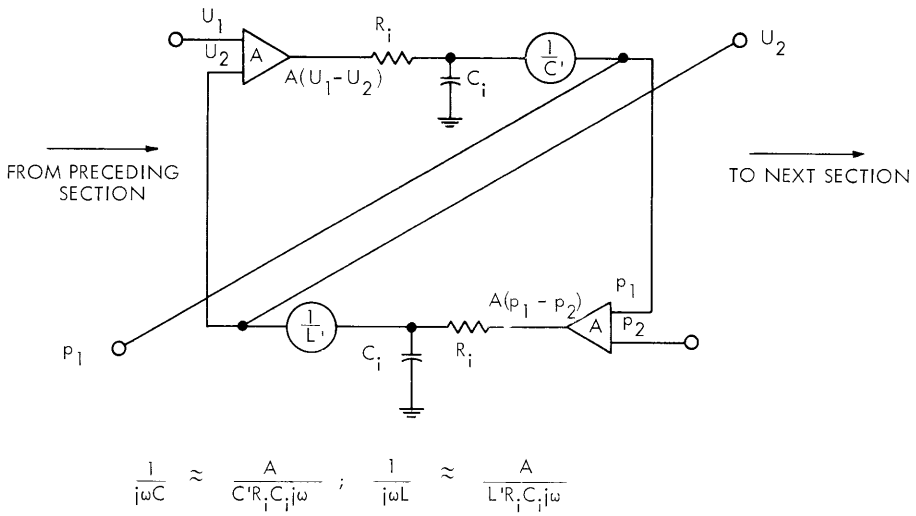


Fig. XX-1. Electrical analog of a short lossless tube.

In the static analog of Stevens, Kasowski, and Fant,[4] each inductance and capacitance was set manually and could not be changed during the course of an utterance. In Rosen's device, each L and C is electrically controlled by an elaborate timing system, and it thus can be used to synthesize speech sounds and transitions that depend upon a gross motion of the articulatory mechanism for their performance.

In the existing DAVO, the tract configuration represented by the analog is controlled by a set of analog voltages supplied to each LC section by a potentiometer matrix that must be set by the operator. In each section, the product of L and C is constrained to be constant and independent of the area represented, thus keeping the section length invariant (by Eq. 1). Recent attempts to control DAVO from the TX-0 computer have



$$\frac{1}{j\omega C} \approx \frac{A}{C'R_iC_i j\omega} \; ; \quad \frac{1}{j\omega L} \approx \frac{A}{L'R_iC_i j\omega}$$

Fig. XX-2. Proposed analog section.

demonstrated that the present device is poorly suited to digital control. The instability of the vacuum tube circuitry employed and the analog nature of the control system have caused great difficulty.

The new analog will be completely transistorized and designed for operation by a digital computer. Hence the control signals will be in digital form. The LC constraint can then be included in the computer program or discarded to allow small changes in the tract length during phonation. This eliminates from the system a complex piece of hardware, which in Rosen's DAVO serves as a frequent source of trouble in calibration and maintenance.

But the greatest difference between the existing DAVO and that which is proposed here is the very manner in which the inductance and capacitance are obtained. For the former, Rosen uses a saturable reactor; for the latter, the input impedance of a variable-gain Miller amplifier – in both cases a physical element. In the proposed design (Fig. XX-2), each section is constructed as a self-contained analog computer that receives as inputs two signal voltages representing the volume velocity at the input of the section and the pressure at the input of the next section. It performs on these the mathematical operations associated with a series inductance and shunt capacitance and delivers two output voltages, again representing pressure and volume velocity, to be used as inputs by the adjoining sections in which the process is repeated. As shown in Fig. XX-2, both integrations required are performed by simple RC circuits following differential amplifiers. Thus,

$$\frac{1}{L} \approx \frac{1}{R_iC_i} \cdot \frac{1}{L'} \cdot A \qquad \frac{1}{C} \approx \frac{1}{R_iC_i} \cdot \frac{1}{C'} \cdot A. \tag{2}$$

Here, 1/L' and 1/C' represent two step attenuators, controlled by digital signals and capable of attenuating from 1 db to 64 db in steps of 1 db. The degree of attenuation effectively varies the value of L and C and thus sets the desired area for the section. This area will be variable over a range 15-0.01 sq cm, and a complete closure of the tract will be available for the first time.

It is also possible to implement this approach with RC differentiating circuits in the loop. This approach has the advantage of requiring a much smaller gain A for the differential amplifiers, but leads to stability problems. Experiments have been conducted with both circuits, but still no decision has emerged as to which will prevail.

The design and successful testing of the digital attenuators, differential amplifiers, and simplified models of a single section have been completed. At present, a detailed study is being carried through to determine how the effects of acoustical losses should be accounted for in the system. It appears at this stage that the nonideal integration (or differentiation) of the RC circuits introduces damping that closely approximates the acoustical situation; that is, that the exact equations of the analog of Fig. XX-2 are practically identical with those of an RLC circuit with a Q of approximately 20.

Considerable work must be expended in several areas to complete the project. First, an analytical study must be made of the section design and a cascade of such sections to determine the nature and extent of errors arising in this approximation of the vocal tract. Also, the details of combining the hardware components into a compact and stable system have not been settled, and no consideration whatever has been given to the representation of the glottal source and radiation impedance at the lips.

E. C. Whitman

References

1. G. Rosen, Dynamic Analog Speech Synthesizer, Technical Report 353, Research Laboratory of Electronics, M. I. T., February 10, 1960.

2. S. E. Kasowski, A Speech Sound Synthesizer, S. M. Thesis, Department of Electrical Engineering, M. I. T., January 1952.

3. C. G. M. Fant, Acoustic Theory of Speech Production (Mouton and Company, 's-Gravenhage, 1960).

4. K. N. Stevens, S. Kasowski, and C. G. M. Fant, An electric analog of the vocal tract, J. Acoust. Soc. Am. 22, 734-742 (1953).

C. COMPUTER CONTROL OF A TERMINAL ANALOG SPEECH SYNTHESIZER

A system has been developed for the control of a terminal analog (or resonance) type of speech synthesizer by the TX-0 digital computer, which is similar to the application of a computer to the control of an articulatory analog speech synthesizer recently described by Dennis.[1] The system includes the necessary equipment for the coupling

of the computer to the synthesizer and a program for the computer which effects the desired operation of the synthesis system.

The synthesizer consists of a cascaded series of electrically variable resonators. An electrical signal that is an analog of the glottal excitation to the human vocal tract is generated and then passed through the synthesizer. Each of the variable resonators impresses one formant on the signal. By varying the formant frequencies, as well as the amplitude and fundamental frequency of the source generator, sounds that closely approximate the voiced sounds of human speech can be produced.

The purpose of the new control system is to provide a means by which the parameters of the synthesizer may be independently and simultaneously varied as a function of time in a manner specified by input data in a numerical format. A computer program uses these input data to produce the necessary temporal sequence of digital outputs that, after being converted to analog form by a digital-to-analog converter, control the

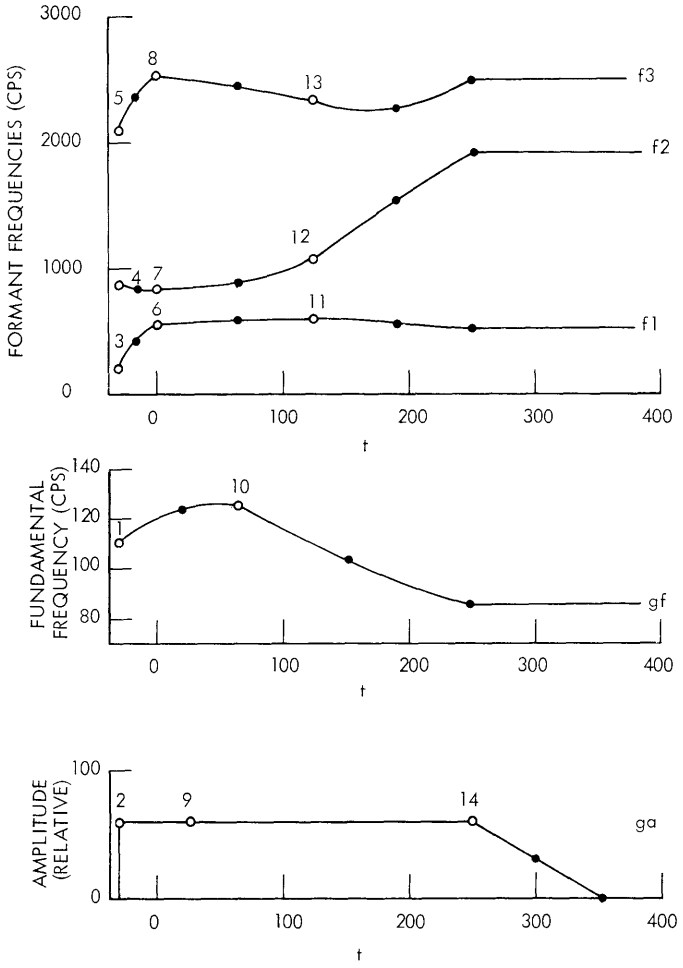


Fig. XX-3. Sample parameter specification.

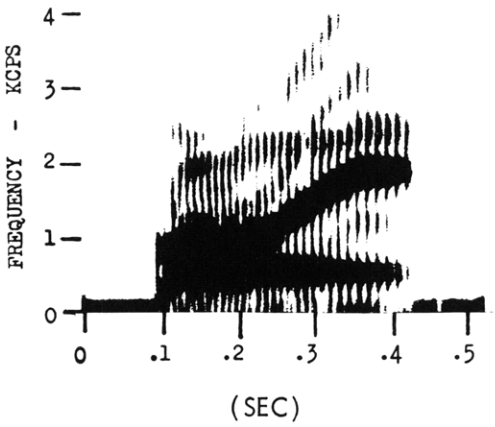


Fig. XX-4. Sound spectrogram of output produced by the synthesizer for the input data shown in Fig. XX-3.

synthesizer parameters.

Temporal variations for each of the parameters are described by a sequence of confluent segments of second-degree curves. The input data that specify each of these time segments include the value of the parameter at the initial, central, and final values of time of that segment. From these data the program calculates the parameter value for all intervening time on the basis of the quadratic function that the data uniquely determine. This scheme allows a high degree of flexibility, inasmuch as the segment density can be varied according to the complexity of the desired temporal variation. The specification of each time segment also includes data that identify the particular parameter being described, the duration of the time segment, and data that determine the time relationship between the segments of the various parameters.

The input data are first sketched graphically and then transcribed directly into the numerical format used by the computer. An example of this process is shown by Fig. XX-3. The small circles denote the beginning of a new time segment, and the numbers associated with each circle identify the time sequence in which they occur. This particular example produces a speech sample approximating the word boy. A sound spectrogram of the output of the synthesizer for this input is shown in Fig. XX-4.

Measurements of the acoustic output of the system have indicated satisfactory performance, and several speech samples have been synthesized.

Further details concerning the physical realization of the synthesis system and its use can be found in the author's thesis.[2]

W. L. Henke

References

1. J. B. Dennis, Speech synthesis, Quarterly Progress Report No. 67, Research Laboratory of Electronics, M. I. T., October 15, 1962, pp. 157-162.

2. W. L. Henke, Computer Control of a Terminal Analog Speech Synthesizer, S. M. Thesis, Department of Electrical Engineering, M. I. T., August 1962.