Dept. for Speech, Music and Hearing

Quarterly Progress and Status Report

The OVE III speech synthesizer

Liljencrants, J.

journal: STL-QPSR
volume: 8
number: 2-3
year: 1967
pages: 076-081

IV. SPEECH SYNTHESIS

A. THE OVE III SPEECH SYNTHESIZER*

J. Liljencrants

The OVE III speech synthesizer has been developed at the Speech Transmission Laboratory in 1966 and 1967. The basic terminal analog has a block diagram, Fig. IV-A-2, similar to that of the earlier OVE II system [(1)]. The constructional details are however entirely different, especially since the synthesis parameters are digitally controlled from a small computer. The computer programs handle all communication between the operator and the synthesizer as well as other input/output devices.

General hardware layout

The overall equipment configuration used for synthesis is shown in Fig. IV-A-1. The computer is a CDC 1700 with an 8K, 16 bit memory with 1.1 μsec cycle time. Attached to it is a disk storage and a general I/O interface. This interface handles digital transmission between the computer and the laboratory built hardware. It also gives analog output to the monitor oscilloscope. The control table contains push-buttons with indicator lamps and a number of rotary knobs coupled to digital encoders. Natural and synthetic speech samples may be compared using a filter bank spectrum analyzer with digitized output [(2)].

Terminal analog synthesizer

The outlines of the synhtesizer are given in Fig. IV-A-2. Some of its key elements may be described closer:

Control amplifier/attenuators

The control principle is similar to that of the MIT synthesis system [(3)], i.e. the analog circuits are directly controlled in a step-wise manner by digital signals. The complete absence of analog control signals is a major promotor of long-time stability. As a consequence,

* Paper to be presented at the 1967 Conference on Speech Communication and Processing, Cambridge, Mass., Nov. 6-8, 1967.

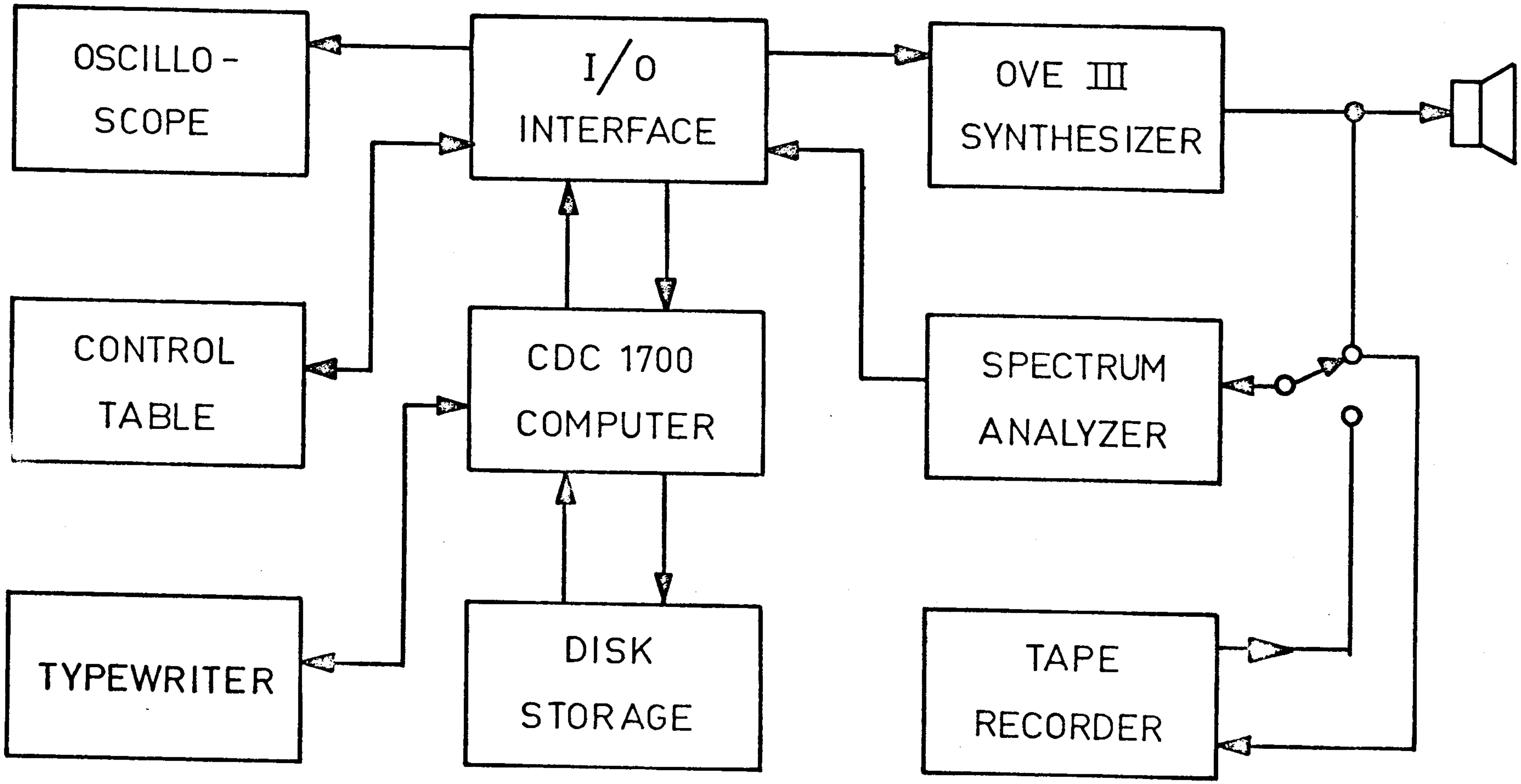

Fig. IV-A-1. Block diagram of complete synthesis setup.

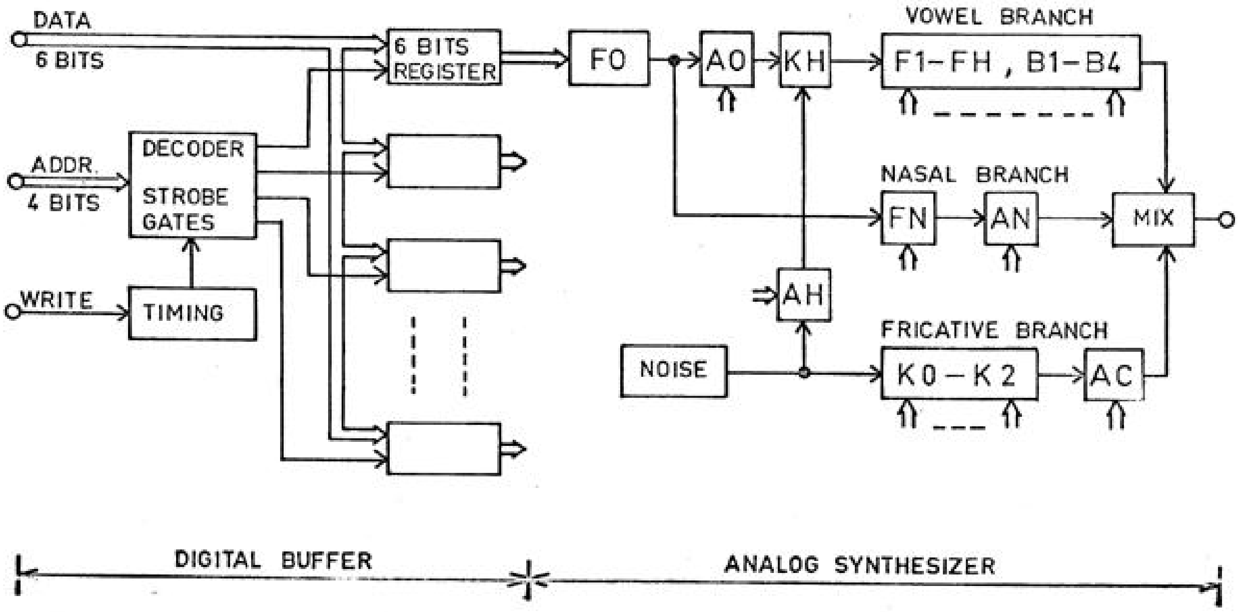

Fig. IV-A-2. Block diagram of the synthesizer with its digital control signal buffer.

however, any "smoothing" of the control signals has to be done on the digital side. This will necessitate a higher bit-rate for the control than with a conventional system using analog smoothing.

It is a well-known problem to maintain a reasonable signal to noise ratio in a series synthesizer. This is much relieved by keeping all losses and gains to numerically small dB values throughout the system. In the case of the control elements this is done by alternate cascading of loss pads and feedback controlled amplifiers as in Fig. IV-A-3a. The controlled gain range can thus be made roughly symmetrical around 0 dB. The control elements have a 32 dB range with a .5 dB increment (6 bits). Identical units are used for control of formant frequencies and signal levels. As an exception the pitch frequency controller has only a 16 dB range.

In cases of level control the amplifiers are cascaded with gates that cut the signal off entirely when the code for minimum gain is given.

Formant (pole) circuits

The block diagram of Fig. IV-A-3b is used. Again in order to minimize gains, an LC combination is used. A secondary advantage from the low gains is now that it is easy to keep the amplifier phase shifts within tolerable limits. This is important since phase errors will affect the resonance bandwidth and may eventually cause instability.

The bandwidths can be controlled by external shunt resistors as indicated in the figure. In OVE these resistors cause rather drastic bandwidth increments. This is used for the synthesis of nasalized vowels and certain consonants. The smaller variations in bandwidth with formant frequency called for in vowel synthesis are cared for the conventional way with series and shunt resistors to the tuning capacitor. The shunt resistor will then give a constant contribution while the series resistor causes bandwidth rising with frequency. The latter also serves as a final corrector for amplifier phase errors.

Antiformant (zero) circuit

A conjugate zero in the fricative branch is implemented as in Fig. IV-A-3d. No external bandwidth control is exercised in this circuit.

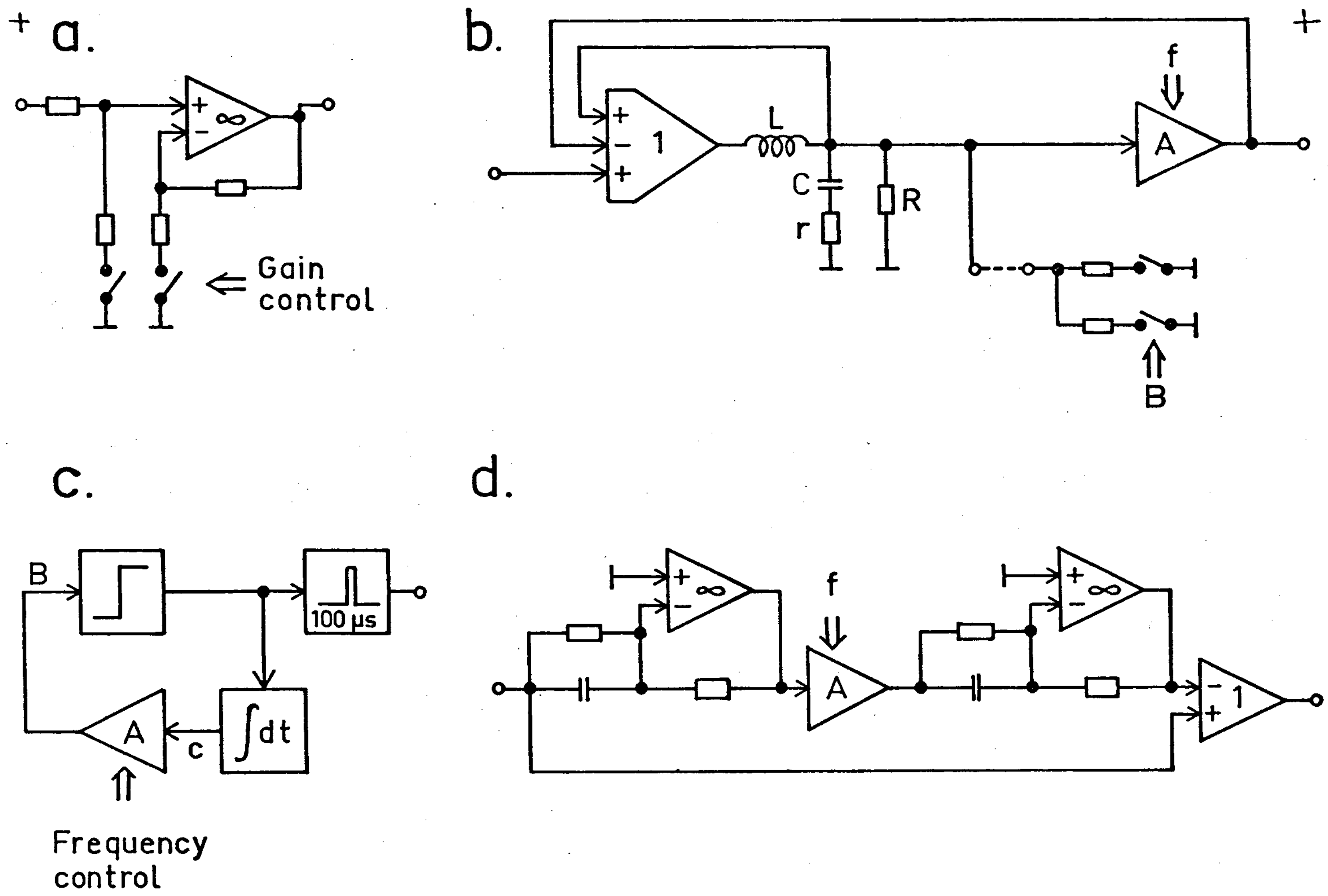

Fig. IV-A-3. a. One stage of a gain control module. Each module contains three cascaded stages.
b. Formant (pole) circuit with frequency and bandwidth controls.
c. Pitch pulse generator.
d. Antiformant (zero) circuit.

Pitch pulse generator

The oscillator shown in Fig. IV-A-3c contains a large hysteresis trigger. Its square wave output is integrated. The resultant triangular wave is fed to the control amplifier and returned to the trigger. As soon as one of the trigger levels is reached the slope will be reversed. At point B the peak to peak voltage then is constant and equal to the trigger hysteresis. Thus at point C it is proportional to the inverse of the control gain. Here instead the waveform slope is constant. It is now easily understood that the frequency must be proportional to the gain A.

The square wave starts a 100 μsec one-shot whose output is used for vowel and nasal excitation after appropriate analog pulse shaping. Part of this shaping is done in the "KH" network which corrects the spectrum for the lacking influence of F5 and higher formants not simulated in the vowel branch (see Ref. [4]).

Digital buffer

The digital control signals are buffered in 14 flip-flop registers with 6 bits each. These are in turn loaded from the computer one at a time using a de-multiplexer. The computer output to load one register will then have to be 10 bits in parallel. Four of these are interpreted as an address or parameter name while the other six give the parameter value. With this arrangement it is necessary to transmit data only when a parameter has to be changed.

Control parameter

In Table IV-A-I the synthesizer control parameters are listed together with specifications.

Control program

The present first version of the synthesis control program is primarily intended for initial synthesis strategy evaluation. The data input is thus not yet automatized but taken from the typewriter. To ease the manual operations required a special language has been developed.

All control parameters for a specific time instant are packed into 8 machine words in the computer. This sample also contains information used by the program. Especially there is a pseudo-parameter

Table IV-A-I. OVE III control parameters.

Name	Address code	Data bits no.	Range	Incre-ment	Remarks
F0	1	0-5	50-300 Hz	3 %	Pitch fundamental
F1	2	0-5	200-1200 Hz	3 %	
F2	3	0-5	500-3100 Hz	3 %	Vowel formants
F3	4	0-5	1000-6200 Hz	3 %	
A0	5	0-5	32 dB	5 dB	Vowel level
AC	6	3-5	28 dB	4 dB	Fricative level
AH	6	0-2	28 dB	4 dB	Aspirative level
AN	7	4-5	24 dB	8 dB	Nasal level
FN	7	0-3	200-1100 Hz	12 %	Nasal formant
FH	8	2-5	2500-6200 Hz	12 %	F4 and part of KH
K0	9	0-5	1000-6200 Hz	3 %	Fricative antiformant
K1	10	0-5	1000-6200 Hz	3 %	Fricative formants
K2	11	0-5	2500-9800 Hz	3 %	
B1	12	1-2		100 Hz	
B2	12	4-5		100 Hz	Vowel formant
B3	12	0		200 Hz	bandwidths
B4	12	3		200 Hz	
01	13	0-5			For optional adden-
02	14	0-5			da to the circuits

"DR" giving the time duration between the present sample and the previous one. Later, in controlling the synthesizer, the program interpolates linearly between adjacent samples for each time increment. This way a comparatively small number of samples has to be specified, typically two or three per phoneme in connected synthetic speech.

To build up a sample string the operator can choose among a set of one-character control statements preceded by pertinent sample numbers. Among the operations possible are the insertion or removal of single or groups of samples in the string, listing and alteration of parameters within a sample, and a number of copying operations. Standard sample groups can be loaded by typing pseudo-phonetic characters.

Since all essential I/O equipment is operated using the computer interrupt system the building up of the sample string can be monitored continuously from the oscilloscope. It is also possible to make the computer repetitively output control to the synthesizer in time-sharing with these functions. This gives an immediate auditory feedback to the operator.

From the manual control table the operator can select parameters to be displayed and move the plot along the time axis with a rotary knob. With another knob he can select points in time to get the corresponding output to the synthesizer. Simultaneously the output spectrum is analyzed and displayed on the screen. The plots may be recorded for future reference, see Fig. IV-A-4.

The oscilloscope plot can also be changed to show sections and comparisons for synthetic and natural sounds, Fig. IV-A-5. The natural reference spectra are then pre-recorded on the disk storage using the spectrum analyzer. The time coordinate is again taken from the control table knobs.

The programs occupy approximately half the computer memory while the rest is available for storage of samples. This memory space is sufficient for approximately 30 seconds of speech corresponding to a storage bit rate of about 2,000 bits/sec. Considering the scheme used this is a high figure, due partly to inefficient packing oriented to simplify the programming. Also, to preserve unlimited

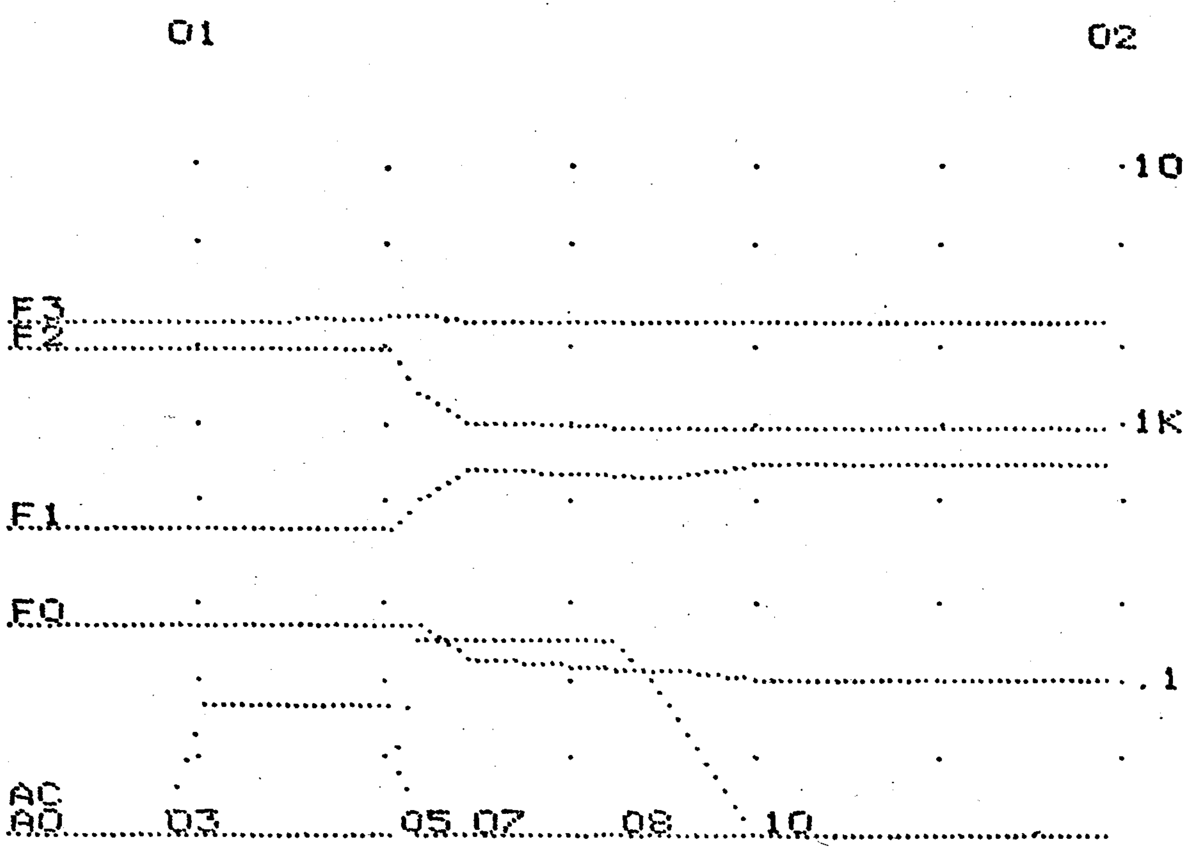

Fig. IV-A-4. Computer plot of some control parameters for the synthesis of /sa:/. Characters indicate: left, parameter names; top, time (seconds); right, frequency (kHz); and bottom, sample number.

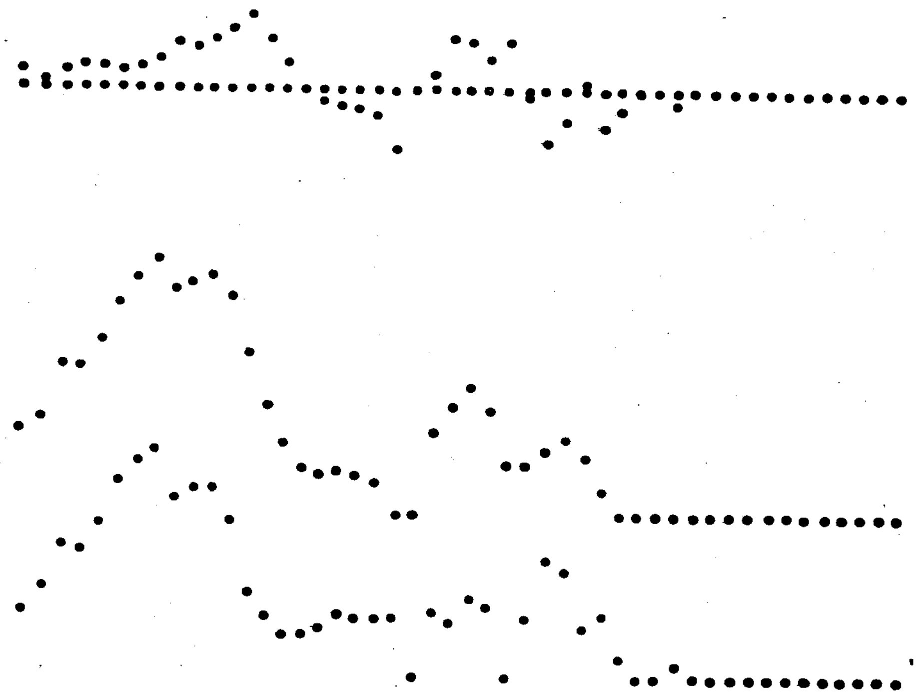

Fig. IV-A-5. Computer plot of spectra for the /a:/ of Fig. IV-A-4. From bottom to top: original sound, synthetic sound, and spectrum difference.

flexibility at this initial stage of development there is a big redundancy in the data stored. Some future work will be devoted to minimizing the storage volume needed, important in computer vocal response applications.

References:

(1) Fant, G. and Mártony, J.: "Instrumentation for Parametric Synthesis (OVE II)", STL-QPSR No. 2/1962, pp. 18-24.

(2) Liljencrants, J.: "A Filter Bank Speech Spectrum Analyzer", paper A27 in Proc. Vth ICA, Liège, Sept. 1965.

(3) Tomlinson, R.S.: "SPASS - an Improved Terminal Analog Speech Synthesizer", MIT, RLE, QPR No. 80, Jan. 1966.

(4) Fant, G.: "Acoustic Analysis and Synthesis of Speech with Applications to Swedish", Ericsson Technics, 15, No. 1 (1959), pp. 3-108.