

# THE SYNTHESIS OF SPEECH USIING A DIGITAL COMPUTER 

by

Robert Peel Futrelle

# SUBMITTED IN PARTIAL FULFILLMENT OF THE <br> REQUIREMENTS FOR THE DEGREE OF BACHELOR OF SCIENCE 

at the<br>MASSACHUSETTS INSTITUTE OF TECHNOLOGY<br>June, 1959

```
Signature of Author . . ... . . . . ...................
                                    Denartment of Bhwsica May &5, 1959
Certified by. . . : . . . .
                                    Thesis Supervisor
Accepted by .

\begin{abstract}
A program has been written for the IBM-704 digital computer to synthesize speech. The input requires detailed specification of the speech spectrum as a function of time. The output is audible samples of 2.4 seconds duration with a signal-to-noise ratio up to 36 db and a frequency response up to 7 kc . Included is a complete SAP listing and flow--charts of the program. The synthesis of each sample takes from 6 to 10 seconds. The IBM-704 "Direct-Data" attachment and a digital-to-analog converter is required.
\end{abstract}

\section*{PABLE OF CONTENTS}
Introduction
pg.
Acoustics of the Vocal Tract ..... 1
Electrical Analogy ..... 2
Solution of the Circuit Equations ..... 3
The Production of Vowels and Consonants ..... 4
Previous Methods of Synthesis ..... 5DIGITAL COMPUTER METHODS
Introduction ..... 6
Digital to Analog Conversion ..... 7
The Instructions ..... 8
Summation ..... 10
Preparation of Instructions ..... 11
The P1ece-Wise-Linear Program ..... 11
Pre-emphasis ..... 13
Program Speed, Testing, and Future Plans ..... 14
Bibliography ..... 16
APPENDICES
A -- Sub-word Arithmetic ..... 19
B -- The 704 Program in Detail ..... 22
C -- Flow-Charts ..... 27
D -- SAP Program (Iisting) ..... 30
E -- The Format of the Instructions ..... 50
F -- The Instructions Used for Testing ..... 51

\section*{Introduction}

The purpose of this work is the development of a device to synthesize speech which will be able to manipulate the acoustic "building blocks" of speech in a very flexible manner. Today a synthesizer is primarily useful for studying speech by testing various hypotheses about the interactions of various speech sounds and the nature of our perception of speech. Eventually it is hoped that speech may be added to our methods of communicating with machines. Before this can be done we need to have a fairly detailed knowledge of the acoustics of speech.

Acoustics of the Vocal Tract
The vocal tract begins at the vocal cords in the larynx and then divides, one section terminating at the mouth approximately 17 cm . away and the other section terminating at the nose a slightly greater distance away as shown in Fig. 1.

For a large class of speech sounds we need only consider the oral cavity, as the nasal cavity is virtually isolated from the system. Thus in essence the vocal tract can be considered to


Fig. 1 be a tube of fixed length and variable cross-sectional area. The area varies from 0 to \(15 \mathrm{~cm} .^{2}\).

\section*{Electrical Analogy}

If we define an acoustical impedance as the ratio of the applied pressure to the volume velocity ( \(\left.\mathrm{cm} .^{3} / \mathrm{sec}.\right)\), we find that the vocal tract is analogous to an electrical transmission line with distributed inductance \(L=p / S\) and capacitance \(C=S / c^{2} p\) per unit length. Here \(p\) is the density of air, \(c\) is the velocity of sound and \(S\) is the cross-sectional area of the tract at the corresponding point.

A transmission line may be simulated by lumped elements, inductors and capacitors, provided the wavelength \(\lambda\) is long compared to the section being represented. For speech the highest frequency of normal interest is of the order of 6 kc . with a corresponding wavelength of \(\lambda=c / f\) which is \(3 \times 10^{2} / 6 \times 10^{4}=0.05 \mathrm{~m}\). or 5 cm . Thus the vocal tract can be adequately represented by approximately 15 sections. In this analogue the pressure in the system corresponds to the voltage and the current to the volume velocity. The source (for vowels) is represented by a high impedance (current) generator. The output is taken as the pressure (voltage) at the lips which are loaded by a dissipative impedance representing the loss of energy to the surrounding air. See Fig. 2 below.


Fig. 2
-2-

\section*{Solution of the Circuit Equations}

The elements in the circuit obey the simple equations:
\(i=C\)
\(C \frac{d e}{d t}, \quad e=L \frac{d i}{d t}, \quad e=i R\).

By ordinary circuit analysis we obtain a differential equation of the form:
\(b_{n} \frac{d^{n} e}{d t^{n}}+b_{n-1} \frac{d^{n-1}}{d t_{n-1}}+\ldots+b_{0} e=i\)
We are primarily interested in the homogeneous or force-free ( \(i=0\) ) solution. The standard method for solving the above linear; homogeneous, differential equation with constant coefficients is to make the substitution \(e=e^{s t}\). Doing this we obtain the algebraic equation:
\[
b_{n} s^{n}+b_{n-1} s^{n-1}+\ldots+b_{0}=0
\]

Denoting the roots of this equation by \(S_{n} w e\) have the homogeneous solution:

where the values of the constants, \(B_{n}\), are determined by the initial conditions.

The actual system has distributed losses due to viscosity, cavity wall absorption and due to the fact that the process is not adiabatic as the calculations suppose. In the case of losses the roots are in general complex numbers, \(\delta_{i}=\sigma_{i}+j \omega_{i}\), and the solution for the force-free condition can be written in the form:
\(e=B_{n} \epsilon^{\sigma_{n} t} \sin \left(\omega_{n} t+a_{n}\right)+B_{n-1} \epsilon^{\sigma_{n-1} t} \sin \left(w_{n-1} t+a_{n-1}\right)+\ldots\)

\section*{The Production of Vowels and Consonants}

Vowels: The current source can be thought of as delivering impulses spaced at intervals of \(l / f_{0}\) where \(f_{0}\) is the fundamental pitch or voicing frequency. The impulse determines the initial conditions and for the remainder of the interval the circuit decays by the force-free solution above. See Fig. 3 below.


Fig. 3
In general the acoustic network acts as a filter on whatever source spectrum is present, as shown in Fig. 4 below.


Fig. 4
Consonants: The production of a fricative consonant such as (s) proceeds in this manner: A constriction, in this case caused by the tongue pressing against the front of the roof of the mouth, causes a high velocity, turbulent flow of air. This is a source of noise which is filtered by the cavities ahead of it, around the teeth and to some extent by those back of it, giving a spectrum of the type shown in Fig. 5 on the next page.


Fig. 5
The equivalent circuit for the production of fricative consonants seems to involve a noise voltage source at the place of the constriction as shown in Fig. 6 below.


\section*{Fig. 6}

\section*{Previous Methods of Synthesis}

For the synthesis of speech a number of methods have been used. Probably the two most flexible ones are those at \(\mathrm{MIT}^{16 \%}\) and the Haskins Laboratory. 4 The first uses a dynamic analog vocal tract. It is an electrical analogy in which the "cross-section" of the tract is continuously adjustable by purely electrical means (saturable reactors and reactance tube circuits.) The other at the Haskins Laboratory in New York is a device allowing playback of hand drawn speech patterns. The patterns are drawn to any desired complexity and the sound is reproduced by the playback device.

A number of other devices have been built \(9,10,11,13,14,15\), but they all suffer in being difficult to "program" for complex or rapidly changing speech sounds. *Superscripts refer to the bibliography on pgs. 16 - 18 .

\section*{DIGITAL COMPUTER METHODS}

Introduction
It is possible with certain high speed digital computers both to compute the waveforms and to have the waveforms presented at a loudspeaker in their normal audible form, ie., as speech sounds. The computer can also manipulate at high speed the instructions specifying the nature of the waveforms.

The differential equation describing the vocal tract could be converted to difference form and solved containuously to yield the desired output. However, this involves rather high order differences and is thus time -consuming. It is intriguing because it can be made quite exact and still retains the analogue form fairly well. A faster approach, and the one utilized, is to compute a number of samples of solutions and have the program simply extract the desired sections of the solutions from its table and present them at the output. The stored solutions would contain damped sine waves of many different frequencies and filtered noise of various bandwidths and frequencies. This produces a quazi-static solution quite similar to the WKB solution of a wave equation. This approach is justified by the fact that the vocal tract parameters change much more slowly than the characteristic decay rates of the cavity responses.

A program has been written by myself to perform these operations on the MIT-IBM 704 computer. The program has been designed primarily for high speed and flexible operation. An outline of the program is shown in Fig. 7 below.


\section*{Digital to Analog Conversion}

The conversion of the synthesized samples in their digital form to an analog voltage which can be recorded or reproduced in audible form is simply accomplished by a converter such as the one shown in reference 26, pg. 530. Within the machine samples are computed to six binary places of accuracy. This gives a maximum of 36 db signal-to-quantization level ratio, a rough measure of the maximum signal-to-noise ratio. Frequency response is limited by the maximum obtainable read-out rate of 14,000 points per second. The highest component that can be transmitted is \(7,000 \mathrm{cps}\); all higher are reflected into the 0 to \(7,000 \mathrm{cps}\). region and appear as noise.

\section*{The Instructions}

When summing a number of sine and noise samples, the instructions must specify the nature of the samples (frequency, amplitude, phase, bandwidth, etc.), the point at which the sample is to begin, and finally the duration of the sample. The four instructions used to specify this information are: "Time", "Duration," "Formant," and "Noise." Time A marker is kept within the machine which denotes the present point of time. This point is a certain location in the 704 magnetic core memory. When any sample is specified it will bevin at this point. When the Time instruction is given (abbreviated by \(T\) ) the marker is moved ahead by the amount specified by the instruction. Thus a series of \(T\) instructions interspersed with other instructions might give the output shown in Fig. 8 below.

\(T_{1}\) given \(T_{2}\) given \(\quad T_{3}\) given
Fig. 8
Duration It is sometimes desired to limit the duration of a sample. Often a noise burst is desired which is much shorter than the particular sample in storage. By giving a duration instruction (abbreviated by D) we can accomplish this limiting. The amplitude of a noise sample is set and
cannot be changed while it is being summed. Therefore, if a burst of increasing amplitude is desired, a series of short bursts of length limited by \(D\) can be specified. Each of these bursts would have a greater amplitude than the preceding one. The value of the duration is always that of the last given; it need not be stated frequently. The effect of the \(D\) instruction is shown in Fig. 9 below.




Formant
\(D_{2}{ }^{\uparrow}\) given The Formant instruction specifies a single damped sine wave. This instruction (abbreviated F) specifies the phase(as plus or minus), the frequency, and the amplitude. There are normally about 50 frequencies available between 100 and \(5,000 \mathrm{cps}\). The amplitude is adjustable in 3 db steps over a range of 36 db .

Noise The Noise instruction (abbreviated N) specifies the amplitude in 6 db steps and the frequency of the noise. The exact character of the noise samples is yet to be decided.

Summation
It is normally necessary to sum a number of different damped sines or noise. This is done very simply by letting them overlap or coincide in time, when specifying them in the instructions. Thus, if a number of different F instructions are given without an intervening \(T\) instruction the sum of the various F's will be formed, similarly for noise. There are no restrictions on the ordering of the instructions -- they are completely independent in that sense. To produce a section of a vowel we might give a sequence of the form FFFHFFFTFFF...... which would produce a sum as shown in Fig. 10 below.



 \(\stackrel{\uparrow}{\text { feet }}\)
 \(\uparrow\) Fig. 10

\section*{Preparation of Instructions}

The instructions are punched on IBM cards to be read into the machine. Once a large number of cards is punched they need not be re-punched, they need only be re-ordered to form new sounds. If a number of identical cards are required they can simply be reproduced from a "master" (cf. the last 15 instructions in Appendix F.)

The program considers the instructions in order, one-by-one. It is designed so that the parameters such as the frequencies and the bandwidths of the samples can easily be altered. The noise samples are actually computed by the synthesis program itself using random numbers, and the \(T, D\), and \(F\) instructions. The noise can thus be made to have almost any spectral characteristics desired. The program is also designed to be readily modified to receive new instructions that might be wanted to store the samples on magnetic tape, display the samples on an oscillograph screen or present them at the audio output for recording.

\section*{The Piece-wise-linear Program}

The primary disadvantage of the program outlined above is that we must write too many instructions for each sample we wish synthesized. Often there is a great deal of redundancy in these many detailed instructions because the sounds are changing slowly or in some very regular way, egg., the fundamental pitch or some of the
formats may be changing approximately linearally with time during some interval. For cases such as these we use a coding scheme to eliminate a great deal of this redundancy. Even simple schemes can reduce the number of instructions necessary for a 2 second sample from many hundred to easily less than one hundred. The method of doing this is to write a second program which accepts the more concise instructions and translates them into a series of many more of the \(T, D, F\), and instructions. This operation is outlined in Fig. 11 below.


Fig. 11
At this point it is useful to discuss a convenient representation of the speech we want to synthesize.

A common way to represent the dynamic acoustical properties of speech is by "visible speech" patterns \({ }^{21}\). These are produced by certain devices used for the analysis of speech. The patterns are constructed so that the vertical axis represents frequency, the horizontal time, and the density of the shading represents intensity. Vowels then appear as horizontal bars called formats and fricatives appear as grey regions of no very definite structure in the high frequency region. This is shown in Fig. 12 at the top of the next page.


\section*{Time}

\section*{Fig. 12}

To code for the piece-wise-linear program we first approximate the movements of the formant by straight-line segments. At each break-point we state the properties of the formant somewhat as an \(F\) instruction does. We do the same for the noise and for the fundamental frequency. This is shown in Fig. 13 below.

Freq


Piece. Wise-Lincar Coding of the Above Sample

Time
Fig. 13
These patterns are then coded and presented to the piece-wise-linear program which interprets them and composes the proper detailed instructions.

The present writer has not worked on the above program in detail but another student is studying the problem. Pre-emphasis

By programming unnaturally large amplitudes for the higher frequencies we can, in an approximate way, pre-emphasize them. Filtering the output will bring
them back to their"natural" level but now with less noise in the higher frequencies. This is done, of course, at the expense of adding noise and reducing the dynamic range in the lower frequency region. It has been found that the higher frequencies are quite important in the production and recognition of speech even though their normal amplitudes are lower than the low frequency region. The action of the noise and filtering is shown in Fig. 14 below.


As Syathesized

After Filtering
to "Natural Level"

Program Speed, Testing, and Future Plans
The maximum length for a single continuous sample
that can be read out of the machine is 2.4 seconds at 14 kc .
In synthesizing a sample involving the first three formants the program can run as fast as twice real time, i.e., it would take the machine 1.2 seconds to synthesize a sample which would last 2.4 seconds. The slowest rate a which synthesis proceeds is half of real time. The major delay in time is preparing the sample for the output device which takes half of real time. Thus it would normally take between 6 and 10 seconds to compute the sample and ready it for read-cut.

The program has been tested and debugged to a large extent. A cursory examination of the scope display of the synthesis using the instructions in Appendix \(F\) indicated that the program was functioning properly though no films have been received as yet. Further testing using scope display is in order.

The next steps in the development of this program might be the inclusion of tape routines for storing the bank samples, the piece-wise-linear program, and finally the setting up and use of the digital-to-analog equipment. One of the runs of the program should be devoted to computing the noise samples.

A reasonable arrangement for the actual operation of the program would be to have all of the programs and bank samples on tape. Then the piece-wise-linear instructions would be read in from the on-line card reader. Synthesis would then proceed, the samples finally being recorded on an audio tape recorder for listening tests.
* * *

ACKNOWLEDGMENT
The author would like to thank Professor Kenneth N. Stevens for his helpful advice relating to all problems of speech acoustics. Thanks are due also to Professor Dean N. Arden and Mr. Ercolino Ferretti of the RLE for stimulating discussions on various computational techniques.

\section*{BIBLIOGRAPHY}
(The Journal of the Acoustical Society of America is
denoted by JASA.)
1. Don Lewis, "Vocal Resonance," JASA, 8, (1936), 91-99.
2. Ralph K. Potter and Gordon E. Peterson, "The Representation of Vowels and their Movements, " JASA, 20, (1948), 528-535.
3. R. M. Fane, "The Information Theory Point of View in Speech Communication," JASA, 22, (1950), 691-696.
4. Franklin S. Cooper, "Spectrum Analysis," JASA, 22, (1950), 761-762.
5. W. H. Huggins, "System-Function Analysis of Speech Sounds," JASA, 22, (1950), 765-767.
6. Ralph K. Potter and J. C. Steinberg, "Toward the Specification of Speech," JASA, 22, (1950), 807-820.
7. J. C. R. Licklider and Irwin Pollack, "Effect of Differentiation, Integration, and Infinite Peak Clipping upon the Intelligibility of Speech," JASA, 20, (1948), 42-51.
8. Franklin S. Cooper, Pierre C. Delattre, Alvin M. Liberman, John M. Borst, and Louis J. Gerstam, "Some Experiments on Synthetic Speech Sounds," JASA, 24, (1952), 597-606.
9. Cyril M. Harris, "A Speech Synthesizer," JASA, 25, (1953), 970-975.
10. R. I. Miller, "Auditory Tests with Synthetic Vowels," JASA, 25, (1953), 114-121.
11. Kenneth N. Stevens, S. Karowski and C. Gunnar M. Fant, "Electric Analog of the Vocal Tract," JASA, 25, (1953) 734-742.
12. Kenneth N. Stevens and Arthur S. House, "Development of a duantitative Description of Vowel Articulation," JASA, 27, (1955), 484-493.
13. Harry F. Olson and Herbert Belar, "Electronic Music Synthesizer," JASA, 27, (1955), 595-612.
14. E. S. Weibel, "Vowel Synthesis by Means of Resonant Circuits," JASA, 27, (1955), 858-865.
15. James L. Flannagan, "Note on the Design of 'Terminal--Analog' Speech Synthesizers," JASA, 29, (1957), 306-310.
16. George Rosen, "Dynamic Analog Speech Synthesizer," JASA, 30, (1958), 201-209.
17. E. E. David, Jr., "Artificial Auditory Recognition in Telephony," The IBM Journal of Research and Development, 2, (1958), 294-309.
18. R-ili. S. Heffner, Genergl Phonetics, (Madison, Jisconsin, 1952)
19. Colin Cherry, On Human Communication, (New York, 1957)
20. G. E. Shannon and W. Weaver, The Mathematical Theory of Communication, (Urbana, Illinois, 1949)
21. R. K. Potter, G. A. Kopp, and H. C. Green, Visible Speech, (New York, 1947)
22. P. M. Morse, Vibration and Sound, (New York, 1948)
23. E. A. Guillemin, Communication Networks, vol. II, (New York, 1935)
24. M. F. Gardner and J. L. Barnes, Transients in Linear Systems, (New York, 1942)
25. C. G. M. Pant, Acoustic Theory of Speech Production, (The Hague, 1959)
26. Notes on Analog to Digital Conversion Techniques, ed. A. K. Susskind, (Cambridge, Mass., 1957)
27. Coding for the MIT-IBM 704 Computer, ed. F. Helwig, (Cambridge, Mass., 1958)
28. \(\frac{704 \text { Electronic Data-Processing Machine - Manual of }}{\text { Operation, IBM Corporation, (New York, } 1954 \text { to date) }}\)

\section*{Appendix A}

\section*{Sub-word Arithmetic}

When little accuracy is desired it is useful to divide a 36 binary digit 704 word into a number of smeller words called sub-words. Each sub-word carries a sign with it and one position must be left open for a carry so that one sub-word will not interfere with another. The sign and carry bit are indicated by \(S\) and \(C\) respectively. The negative is handled by using the two's complement. These numbers are convenient because addition, subtraction and multiplication can be performed on all of the subwords in a 704 word without ever "unpacking" them. In this program five sub-words are stored in a 704 word as follows:


A single n-place signed sub-word is of the following form:


We have assumed the binary point to be to the right of the first place but the argument below can simply be generalized to this case (normally the sub-word is surrounded by others.)

The \(n+l\) st place designate the sign, \(S\). A zero in \(S\) indicates a positive number, a one indicates a negative number.

The \(n+2 n d\) place is a carry bit for absorbing any changes of sign -- it is erased after each computation. If a number of these sub-words in a 704 word are added to another similar 704 word, each of the \(n\)-place numbers will be added algebraically to the corresponding number (sub-word) in the other 704 word.

The negative of a number is obtained by subtracting it from \(2^{n+1}\) (complementing it.) (and then adding 1 )

First we restrict all numbers of interest so that \(0 \leq a, b<2^{n}\) and \(|a|+|b|<2^{n}\)

We will show that the negative of a number always has a one in S. Trivially a number< \(2^{n}\) has a zero in \(S\) so that complementing it always produces a one in \(S\).

It is interesting to note that the number \(2^{n}\) has the properties \(\left(-2^{n}\right)=2^{n}\) and \(2^{n}+2^{n}=0\). Laws of Addition
\((+a)+(+b)=(a+b)\) by the normal addition process. For \((+a)+(-b)=a-b+2^{(x)}\) we have \(a>b\) or \(2^{n}>a-b>0\) so that \(2^{n+1}+2^{n}>(+a)+(-b)>2^{n+1}\) giving a zero in \(S\), a positive number as we had assumed.

If \(a=b,(+a)+(-b)=2^{n+1}\), a zero.
If \(a<b\) then \(0>a-b>-2^{n}\) or \(2^{n+1}>(+a)+(-b)>2^{n+1}-2^{n}=2^{n}\), giving a one in \(S\).

For \((-a)+(-b)\) we have \(2^{n+2}>(-a)+(-b)>2^{n+2}-2^{n}\), giving a minus sign in \(S\).

All of the sums considered here are \(<2^{n+2}\) meaning that no C ever overflows into and interferes with the sub-word on the left.

Subtraction of the sub-words "in bulk" is similar to addition but the \(C\) bits of the minuend must be loaded with I's to prevent borrowing from the sub-word on the left. Multiplication

We can multiply a number by a factor of \(2^{-m}\), if it is positive, by shifting it \(m\) places to the right and inserting zeros on the left. If we consider the digits shifted past position 1 to be lost we have, originally,
\[
c=k_{n-1} 2^{n-1}+k_{n-2} 2^{n-2}+\ldots+k_{0} 2^{0}
\]

After shifting \(m\) places we have decreased each exponent by \(m\), giving,
\[
c^{\circ}=2^{-m} c \quad \text { to within } 2^{-\theta}=1 \text {. }
\]

If \(c\) is the negative of a number, \(c=2^{n+1}-a\), we insert I's on the left after shifting.
 and we add to this 0 ....1 \(0 . .00\). The number added is \(\left.c^{n+1}-1\right)-\left(2^{n-m+1}-1\right)=2^{n+1}-2^{n-m+1}\) so that \(c^{\prime} \cong 2^{-m}\left(2^{n+1}-a\right)+2^{n+1}-2^{n-m+1} \cong 2^{n+1}-2^{-m} a\) which is the negative of the number \(2^{-m} a\) as desired.

If a digital-to-analog converter is used which only handles absolute values we add \(\mathbf{2 n}^{\boldsymbol{n}}\) to each of the sub-words. This is a 704 word with l's in the \(S\) bits and 0 's elsewhere. We can also accomplish the same operation by simply changing the sign which can be seen to be equivalent.

Appendix B The 704 Program in Detail
The 704 program as tested in May 1959 is made of two basic sections. First is the synthesis program itself, and second the auxiliary routines -- one to initially compute the damped sines and another to display synthesized waveforms on the cathode ray tube unit attached to the computer.

\section*{Synthesis Program}

The present point of time may occur at a sub-word within a 704 word so that addition, when it occurs, will involve shifting the stored samples to the right or left so that their beginning will correspond with the present point of time. On the other hand the present point of time may be at the beginning of a 704 word which means no shifting is required. Since the second operation is much simpler than the first, an optimum program could be written that would make the simple routine run much faster. Also, multiplication may or may not be required end there is quite a difference of speed when multiplication is omitted.. To take advantage of these possibilities to gain speed, routines have been designed to handle each case separately. The phase (plus or minus) is also handed separately at times. There are three working areas in the machine: the bank or sample block containing the damped sines and noise, a short work space called the multiply block, and finally the assembly block in which the entire synthesized sample is formed.

The various possible sequences of operations are shown in Fig. 15 below.


The instructions are tested by the program one at a time and completely executed before the next instruction is studied. Whenever the synthesis (sometimes called "assembly") program is entered all of the markers are initialized so that it begins at the origin of the assembly block. After an instruction has been executed a return is made to obtain the next instruction, or an error count is stepped by one if an illegal instruction was given (amplitude or frequency out of range), or the assembly may terminate. Termination occurs if an all zero instruction is found (this is a \(T\) zero instruction, or if the present point of time becomes located near the end of the assembly block.

The \(T\) instruction causes the setting of three markers: the present point of time measured in sub-words from the beginning of the assembly block, the number of 704 registers from the beginning of the assembly block, and the position of the present point of time in sub-words as measured from the beginning of the 704 word which contains it.

The \(D\) instruction sets the duration as measured in sub-words. If the duration is found to be greater than the length of the longest sample (a noise sample), then it is reduced to this value.

The \(F\) instruction first computes the phase of the sample as plus or minus and sets a marker denoting this. Since the multiplication routine can only handle shifts of multiples of 6 db in amplitude (powers of 2), and 3 db steps are desired, each sine is stored at two amplitudes separated by three db . Thus both the amplitude and the frequency enter into the calculation of the location in storage. Both the number of places to shift, for multiplication, and the sample location are then computed. Finally, advantage is taken of the fact that the sine waves damp out in shorter times for lower amplitudes. If this damping time is shorter than the duration time already specified, the routines handle the sample for the minimum time necessary and do not compute useless strings of zero's. The \(F\) routine then transfers control to the proper assembly routine for the actual synthesis.

The \(N\) instruction is quite similar to the \(F\) except that the phase is always plus and because of storage requirements the amplitude is in 6 db steps.

In this program the end of assembly is followed by the (TV) routine for scope display. In a production version the end of the assembly would cause the sample to be stored or to be read out. Then a new set of instructions would be obtained and the next sample synthesized. Auxiliary Routines

To compute the damped sine waves, one cycle of a 55 cps . tone and one exponential of somewhat longer duration are computed. The nth harmonic is then formed by indexing through the block of the sine wave, picking every nth point. The computations for the indexing are done modulo the block length (which is conveniently \(256=2^{8}\) ), so that the point picked is always constrained to lie within the one cycle block. This is a very fast way to produce multiples of a waveform in time. These harmonics are multiplied by the decaying exponential, scaled to two different values and the resulting samples stored. All harmonics of 55 cps . up to a pre-set limit are computed and then every even harmonic up to a second limit. The limits in the present Version of the program are 550 and 5500 cps, respectively. This routine is finished at a point called "End Computation." At this point a routine can be inserted to store the samples on tape so that they need not be computed for each run.

The scope routine first titles the run and then takes blocks of 100 words ( 500 sample points) and displays them a number of times, numbering each block sequentially from 1 . The right edge, left edge and the x-axis are also displayed. The role of this routine is quite flexible, as it is only needed when there is some question of whether or not synthesis is proceeding correctly -- it is quite useful in debugging.
Appendix C
FLOW - CHARTS
General Synthesis Program ..... 28
Auxiliary Routines. . . . . . . . . page ..... 29
On the flow-charts the symbols in parentheses referto points of entry and data words in the SAP program.The actual locations of the most important of these inthe SAP program are indexed on page 30.




\begin{tabular}{l} 
SYMBOLS IN PARENTHESES \\
REFER TO LOCATIONS IN \\
THE SAP PROGRAM \\
\hline
\end{tabular}








Program Indez (continued)
Symbol Instruction Number
PMP ..... 349
PSKO ..... 207
S ..... 107
SNMPY ..... 381
T ..... 133
TD ..... 193
Auxiliary Routines
A ..... 80
DSPIC ..... 427
EC ..... 79
HARML ..... 27
HOUTI ..... 54
NEXTT ..... 458
SETV ..... 413
SMWL ..... 434
SMWIR ..... 430A
START ..... 1
THNH ..... 75
TV ..... 405
TVE ..... 468







> RFSFT PPK
PRESENT TIME ROUTINE
\(T=0, ~ E N D ~ A S S E M P L Y ~\)
RFSFT PPK
FORM PPR
FORM PPSK
 FND ASSEMBLY
DURATION ROUTINF
TFST FOR D MAXIMUM
SET DK
FORMANT ROUTINF
SET PHASE
FREQ. HIGH
STORE AMPLITUDF



\(T\) 1 1
\(\vdots\) \(\begin{array}{ll}: & : \\ \vdots & \end{array}\)

C
35














\(A\) SUR 9
A SUR 7
A SUR 5



\section*{Appendix E}

\section*{The Format of the Instructions}

Octal numbers are based on the numbers 0 through 7 as our decimal system is based on 0 through 9. Octal numbers are especially useful in computer work as there is a one-to-one correspondence between an octal digit and a group of three binary digits. The instructions will be composed of six octal digits so that two will occupy each 704 word. The positions are numbered from the left.

T will have a 0 in position l. Positions 2-6 contain the value of the shift in units of one sub-word, from 0 to 77,7778 ( 0 to 32,76710 .) A 0 will terminate assembly.

D will have a 1 in position 1. Positions 2-6 contain the duration in units of one sub-word.

F will have a 2 in position 1. Position 2 is even for a starting phase of \(0^{\circ}\) (plus) and odd for a starting phase of \(180^{\circ}\) (minus.) Positions 3 and 4 contain the number of the sample and thus determines the frequency. Positions 5 and 6 contain the amplitude.

N will have a 3 in position 1. Positions 2 and 3 specify the nature of the noise desired. The samples would be numbered and their characteristics listed for the programmer's convenience. Position 4 specifies the amplitude. Positions 5 and 6 are disregarded.

THE INSTRUCTIONS USED TO TEST
THE SYNTHESIS PROGRAM
```

