393

National Aeronautics and Space Administration

HOUSTON, TEXAS

N70-35836

(ACCESSION NUMBER) 78

(PAGES)

TMX-65003

(NASA CR OR TMX OR AD NUMBER)

(THRU) 1

(CODE) 07

(CATEGORY)

FACILITY FORM 602

Manned Spacecraft Center

393

MSC INTERNAL NOTE NO. 66-EE-10

DEVELOPMENT OF A NARROW BANDWIDTH SPEECH PROCESSOR

Prepared by C. H. Stewart

Approved:

Approved:

Approved: R S Sawyer

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

MANNED SPACECRAFT CENTER

HOUSTON, TEXAS

December 1966

TABLE OF CONTENTS

INTRODUCTION

A narrow bandwidth speech processor has been developed and a breadboard model exists as part of the Audio and Voice Communications Laboratory. This breadboard demonstrates that human speech can be transmitted in a 160 Hz communications channel. Also, the type circuits used provides means for maintaining minimum size and weight requirements imposed by spacecraft applications. This report will present a discussion of this development and some possibilities the system has for spacecraft applications.

This report is divided into four sections. Section one will present the basic concepts of the processor. Section two will discuss some aspects of how the processor was developed. Section three will present a simplified analysis of how the processor operates. Finally, section four will discuss applications of the processor with particular reference to its development status. An appendix is provided which contains indirectly related information.

I. PROCESSOR CONCEPT

The concept of the narrow bandwidth speech processor lies within the discovery of a technique known as the single equivalent formant (SEF). This technique is a method by which parameters are extracted from speech that describes an equivalence of the formant structure of speech. However, before discussing this further, it is necessary that the basic structure of human speech be understood. The following few paragraphs are an attempt to provide this understanding.

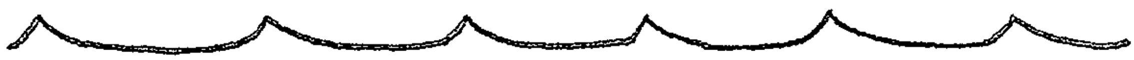
Speech may be defined as humanly generated utterances, logically formed to produce meaningful sounds, as related to written text. The source of these utterances is, in part, the product of air from the lungs and vocal cord control. Air is generated from the lungs which is controlled by the vocal cord in the form of periodic bursts. The bursts of air set up ringing effects in a number of cavities along the vocal track. The vocal track is the hollow volume of the human head and neck, terminated on one end by the vocal cord, and on the other end by the lips, teech, mouth, tongue, and nasal cavity, or passage.

There are times during speech production when air from the lungs is unaffected by the vocal cord. At such time, constrictive parts of the vocal track (primarily lips and teeth) play an important role in speech generation. They, along with the cavities of the vocal tract, are instrumental in providing the logic that forms the utterances into meaningful sounds. How this is done can be shown with the following definitions, explanations and diagrams.

The structure of speech is known to be made up of a number of basic parameters which associate themselves with the operations performed on the air from the lungs by the vocal track, for example, the periodic bursts of air caused by the vocal cord. This operation provides a parameter in speech structure known as pitch and is defined as the rate at which the bursts of air by the vocal cord are permitted to occur.

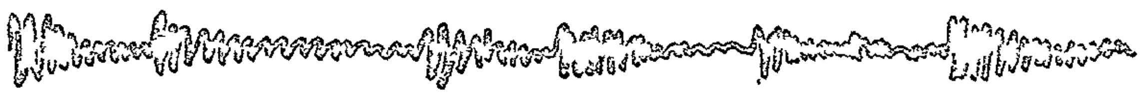
Likewise, the bursts of air cause resonances in the cavities of the vocal track. The resonances are called formants and are associated in a composite manner with a term known as voiced sounds.

During the time when the vocal cord does not operate on the air from the lungs, the constrictions of the vocal track provide another parameter in the form of white noise and is termed unvoiced sound.

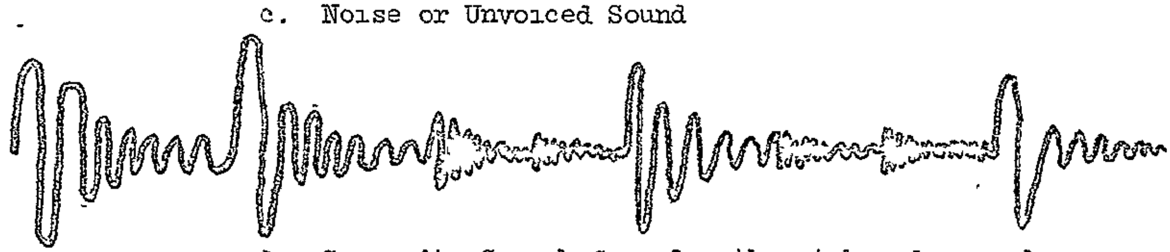


a. Speech pitch (air bursts)

b. Composite formant structure (voiced sound with pitch)



c. Noise or Unvoiced Sound



d. Composite Speech Signal with pitch and voiced or unvoiced sound.

FIGURE 1.

If a visual perception of these parameters is required, one might think of their appearance as shown in Figure 1. a, b, c and d.

In a frequency spectrum the composite voice signal would appear as shown in Figure 2

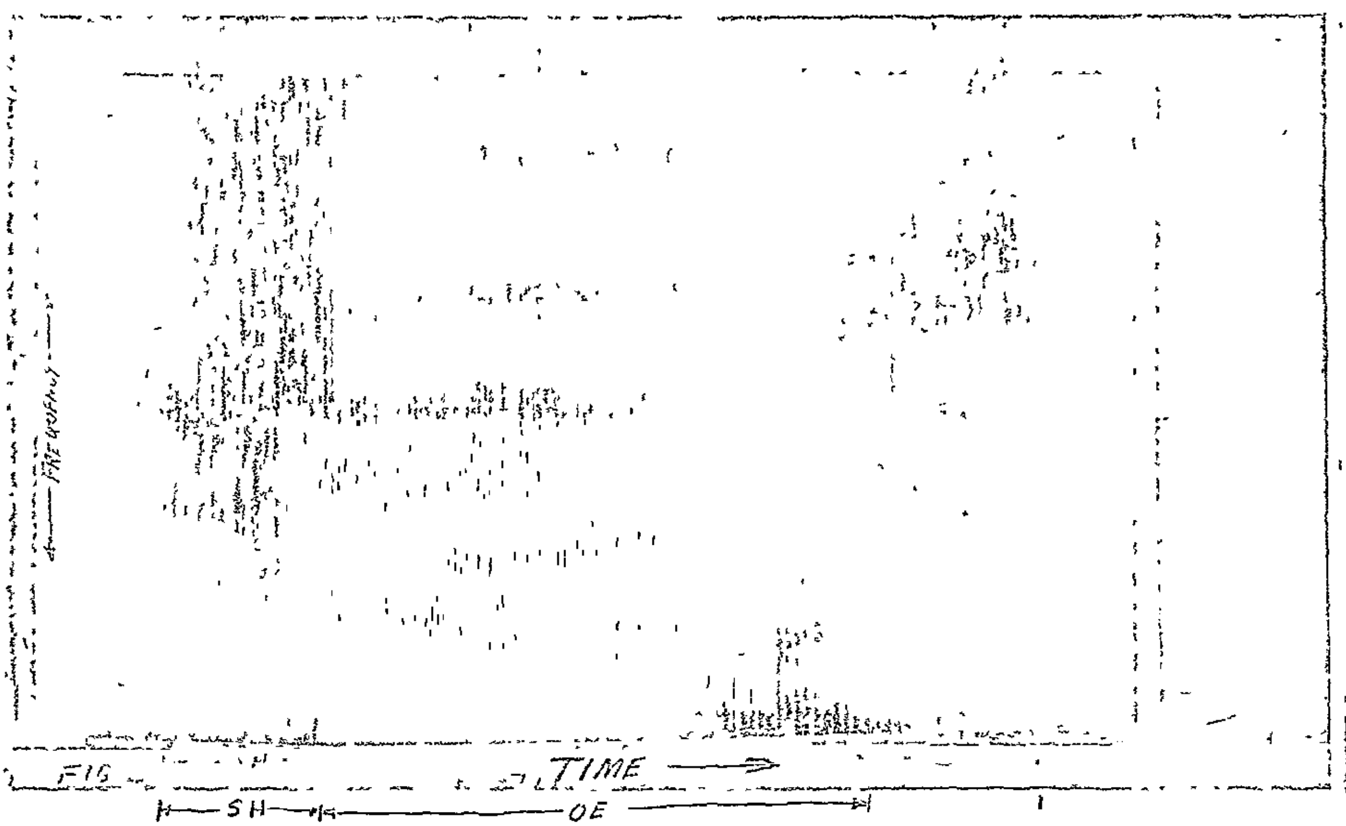


Fig 2

To this point, it may be said that the structure of speech is made up of three basic parameters: (1) Pitch, (2) Formants (voiced sounds) and (3) Noise (unvoiced sounds).

(1) Pitch is the rate at which the vocal cord permits voiced sounds to occur.

(2) Formants are resonances that occur in the cavities of the vocal track.

(3) Noise or unvoiced sounds are those sounds produced by the constrictive parts of the vocal tract.

All of these parameters may be related to the phonetic structure of a language thus providing a correlation between written text and humanly generated utterances.

The phonetic structure of sound, as applied to the English language, is broken up into two general classes, vowels and consonants. In relation to speech utterances, these classes associate themselves in a manner such that vowels correspond to voiced sounds and consonants correspond to unvoiced sounds. However, not all consonants are unvoiced. There are some consonants that correspond to voiced sounds as well as unvoiced ones. These sounds are called combination sounds and are treated as such.

The voiced speech sound, as found from basic speech studies, is known to consist of many formants. However, if the frequency content of speech is limited to below 3.5 khz, only three formants are known to exist.

These three formants, along with pitch and unvoiced sound, are known to be sufficient for providing a fairly accurate replica of speech. It is from these formants that the SEF technique was discovered.

From an extensive study [1] of vowel sound perception and the three speech formants, it was learned that an equivalence existed for the three formants. The idea stemmed from the fact that when one hears, he does not distinguish between individual formants. Instead, the ear seeks to hear an equivalence. The concept is the same as that of two colors in hue. As two different colors approach each other at the point of incident, the eye sees neither of the original colors. Instead, a third color unlike the other two is seen. With this in mind, the graph of Figure 3 provides an approximation of this occurrence with reference to vowel sounds perception and speech formants.

The concept of the single equivalent formant was formed from these ideas As shown in Figure 3, the three formants are described as functions of vowels and frequency. The heavy black curve represents the single equivalent formant (SEF). As indicated for back vowels, the first formant dominates for central vowels an average of the first and second formants dominates and finally, for the front vowels, the second formant dominates. Thus, the SEF appears to shift, depending on which vowel or group of vowels occurs in the spoken text.

Although indicated on the graph of Figure 3, the heavy black curve does not exist physically in speech. Hence, the concept lies in being able to recognize the dominate formant from real time speech. By being able to do this, and realizing that the change from one dominate form to another is never greater than at a 25 Hz rate, a low bandwidth processor can be constructed. Theoretically, this could be done by describing the SEF in frequency and amplitude, along with pitch and unvoiced sounds. Each of the four parameters can be made to vary at a 20 Hz rate

A development was undertaken to implement this theory and concept, A discussion of that effort follows. As will be seen, direct implementation of this concept was not achieved. However, the existing hardware is a direct result of the effort to do so.

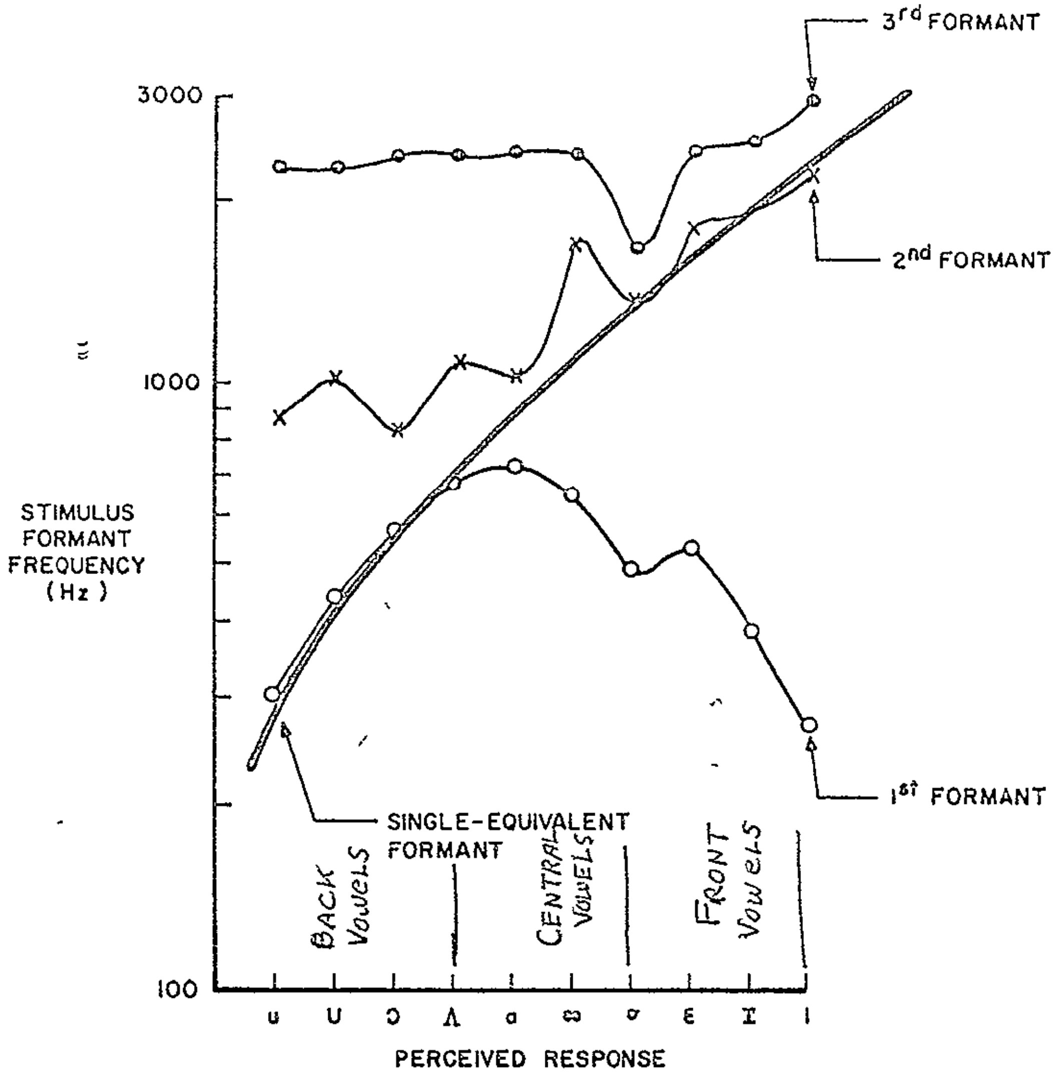


Figure 3. Correlation of the Single-Equivalent Formant and Human-Vowel Formant Locations

II. SOME ASPECTS AS TO HOW THE PROCESSOR WAS DEVELOPED

The development put forth to implement the SEF concept was the result of funds expended by MSC in a program outlined by contract NAS 9-4523. The purpose of the program was to develop a laboratory narrow bandwidth speech processor for spacecraft applications. The characteristics of the processor were to be:

(1) 160 Hz in communication channel bandwidth

(2) Able to meet as a design goal 80% phonetically balanced (PB) word intelligibility for a signal-to-noise ratio of 20 db

(3) Capable of meeting as a design goal an analyzer volume of 35 in.^{3}

(4) Capable of being put into microcircuit form

(5) Able to have a power consumption of no more than 5 watts

The SEF technique appeared to offer great promise of meeting these requirements.

The program was initially focused on utilizing the SEF concept directly. An analyzer, built from earlier work, existed which provided the extracted SEF parameters (SEF period, SEF amplitude, pitch and voiced/unvoiced decision). Thus, work was concentrated on developing a synthesizer that would provide an intelligible speech replica from the given parameters.

Three attempts were made to design a suitable synthesizer. The first involved use of a pitch oscillator to cohere a square wave single equivalent formant oscillator. The second, was construction of a three formant synthesizer. The third, was construction of a three formant synthesizer with digital formant shapers. Neither of these was able to provide a suitable speech replica However, it was discovered later that more accurate definition of the SEF was necessary from the analyzer. This required designing a new analyzer.

The new analyzer was defined as a SEF tracking filter analyzer It appeared to work well. However, suitable definition of all vowels was not possible. Either the A to i, or u, V, region could be defined, but not both simultaneously.

A two formant analyzer of the same type was used to overcome this problem. The u, V, , vowel region was defined by the first formant and the A to i region was defined by the second formant. The entire vowel range could now be defined.

Because the analyzer was now of the two formant type, the synthesizer also was designed to be of the two formant type. Thus, the final system is a two formant speech processor.

The final system does not exhibit direct implementation of the SEF technique, however, the concept remains the same. Also, the final system meets all of the intended requirments.

A simplified analysis of how this processor operates is provided in the next section.

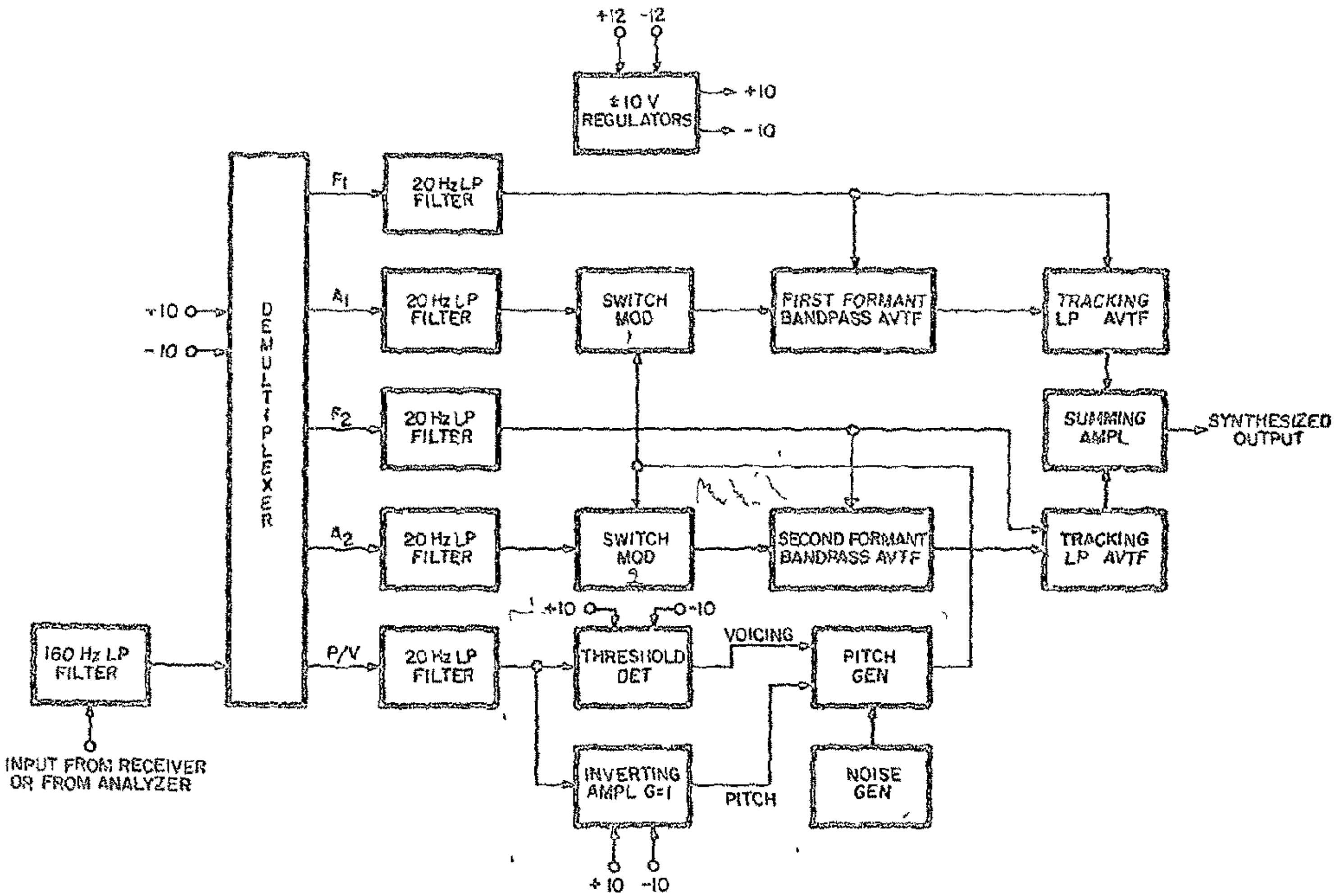


FIG 5 Synthesizer Block Diagram

III SIMPLIFIED ANALYSIS OF HOW THE PROCESSOR OPERATES

The narrow bandwidth speech processor developed is a two formant processor and operates functionally in accordance with the diagrams of Figures 4 and 5. Figure 4 is a block diagram of the processor analyzer. Figure 5 is a block diagram of the processor synthesizer.

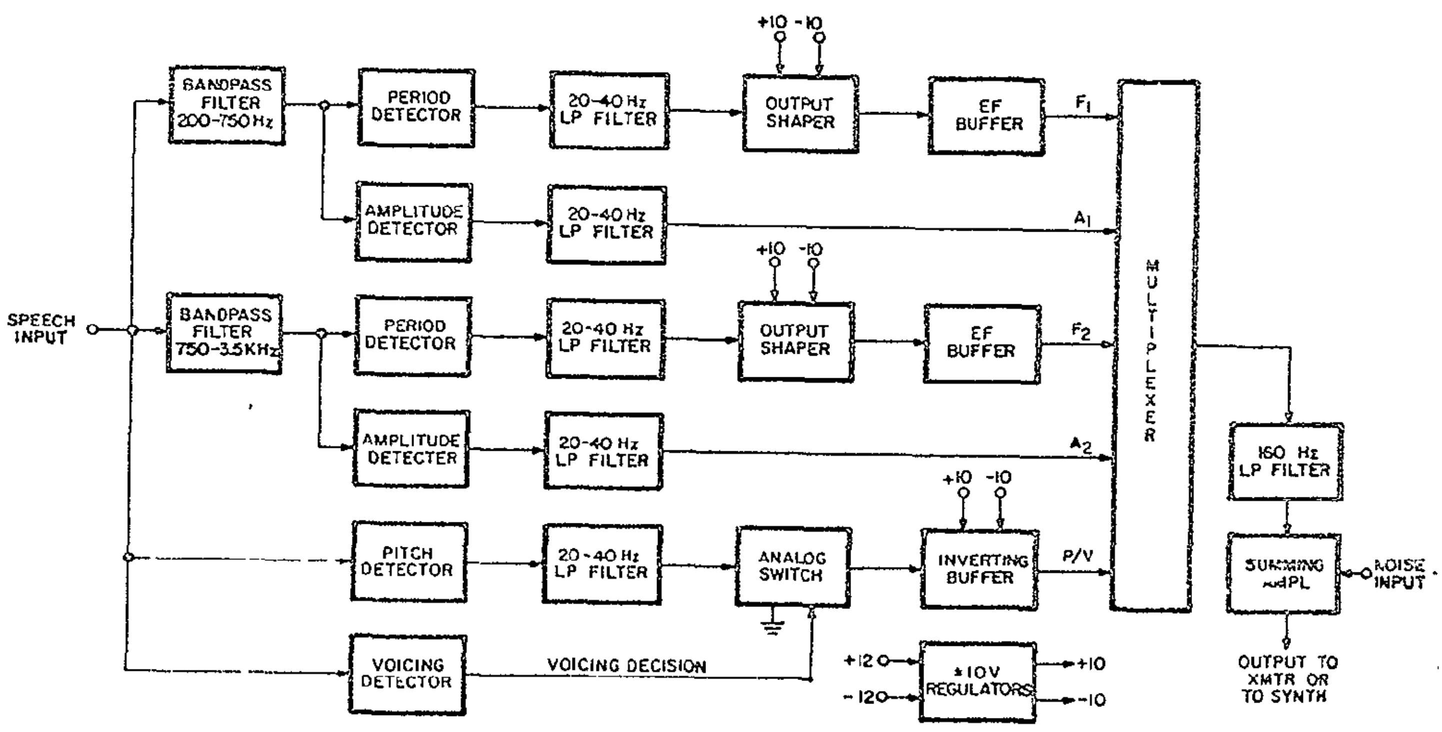


Fig 4 Two - Formant Analyzer

The analyzer extracts directly from the speech signal the parameters that provide the low bandwidth. The synthesizer accepts these low band limited parameters and from them constructs a two formant replica of the speech input. The following detailed analysis, using waveforms, will provide some idea as to how this is done and what the final speech replica looks like

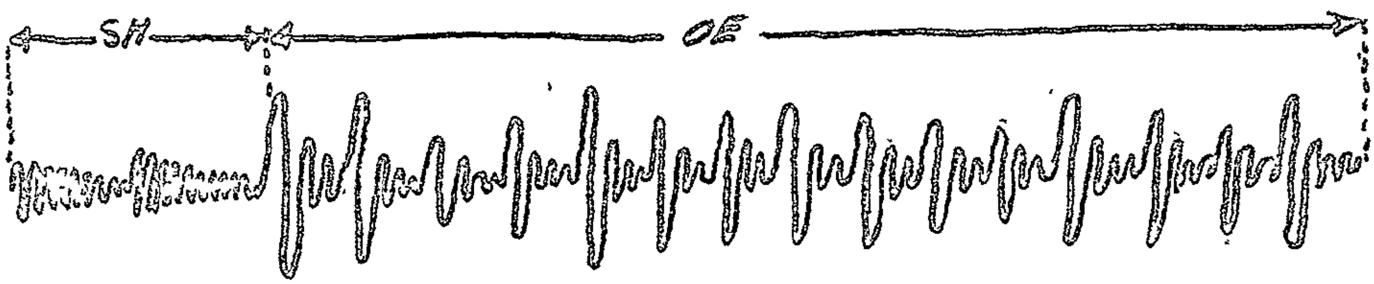


Figure 6 Typical Processor Speech Input

A signal similar to that of Figure 6 is chosen as the input for the analysis. However, like that of Figure 6, this signal represents the word "shoe." The word shoe was chosen because it contains all of the basic structures of speech pertinent to the operation of the processor. Also, the time span of the word when spoken normally is within the limitations of this analysis. The word is also a good representation of a typical speech input.

As shown in Figure 6, the word shoe contains two basic speech sounds which are unvoiced and voiced. The unvoiced appears as random noise and the voiced as a periodic wavetrain. These two properties are fundamental in the processor operation. The periodic wavetrain contains the formant structure and the noise like wave contains factors pertinent to the intelligibility of the spoken text. Thus, both properties must be processed in such a manner that they exist in the speech replica.

The task of performing this processing may be understood from the explanations and series of diagram to follow. The series of diagrams is an illustration of the processing of the word "shoe." The diagrams show how the word is broken down into five individual parameters, multiplexed, made to vary at a 160 Hz rate, demultiplexed and logically formed to provide a replica of the word at the processor synthesizer output.

The five parameter breakdown of the word is (1) first formant period (F_1), (2) first formant amplitude A_1, (3) second formant period F_2, (4) second formant amplitude A_2, and (5) pitch/unvoiced decision P/V. In the actual processor extraction of these parameters is done in a simultaneous manner. However, the diagrams here follow a sequence showing, in the analyzer, F_1 extraction, A_1 extraction, F_2 extraction, A_2 extraction and P/V extraction. In the synthesizer, P/V reconstruction is presented first, followed by combining P/V with F_1 and A_1 to form the first formant, and commanding P/V with F_2 and A_2 to form the second formant. Finally, the two formants are summed to form the output speech replica.

A. ANALYZER ANALYSIS

The analyzer block diagram of Figure 4 is repeated in Figure 7 to provide reference for the analyzer analysis. As shown in this diagram, the speech is fed into the processor at the point indicated as the speech input. The speech then is passed through a 200 Hz to 750 Hz bandpass filter, a 750 Hz to 3.5 kHz bandpass filter, a pitch detector and a voicing detector.

The route through the 200 to 750 Hz bandpass filter represents isolation of the first formant and is the first series of diagram to be discussed. In Figure 8, a block diagram shows first formant isolation and F_1, A_2 extraction. From Figure 9, it is seen that the speech input is passed through a 200 Hz to 750 Hz bandpass filter. The output of this filter, as shown, is the variation of the first formant in both period and amplitude. This output, as illustrated in Figure 10, is first fed to the first formant period detector. Here, period averaging of each damped wave is performed and the output is the variation in the first formant period F_1. However, as shown in Figure 11, this output is passed through a 20 Hz low pass filter. This filter is designed such that its output voltage is a function of the relative change in formant period. Thus, resulting in the final form for F_1. The theory that makes this approach feasible is that of recognizing that the rate of change of formant frequency in human speech is less than 20 Hz. This can be seen by analyzing the sonagraphs provided in the appendix of this report.

The amplitude parameter associated with the first formant, as shown in Figure 12, is obtained by passing the bandpass filtered speech into an amplitude detector. The amplitude detector is a fullwave envelope type detector Thus, the output represents the peak-to-peak envelope of the input. This output, as indicated in Figure 13, is passed through a 20 Hz low pass filter. The filter output is the relative change of the formant peak amplitude. A theory similar to that discussed for the formant period is found to be true of the amplitude peaks and is applied in this process.

The second formant parameter extraction process is exactly the same as that for the first formant and is illustrated in Figure 14 through Figure 19. However, because the second formant region occurs between 750 Hz and 3.5 Hz, the period appears shorter than the first formant. The relative change in period and peak amplitudes remains in the same order as those of formant one. As a result of this operation, the third and fourth parameters are defined.

The final extraction process, as illustrated in Figure 20 through Figure 24, is that of pitch/unvoiced decision. As seen in Figure 20, the speech input is fed simultaneously to a pitch detector and a voicing detector.

The pitch detector analyzes the periodic part of the speech signal and determines in the time domain the beginning of each damped wave. As a result, a series of pulses is produced as the pitch detector output

Each pulse corresponds to the beginning of a pitch period The period is then the time between successive pulses. After production, these pulses are passed through a 20 Hz low pass filter. The output voltage of the filter then produces a signal proportional to the period.

The voicing detector, like the pitch detector, analyzes the speech signal. However, it is designed to provide a bi-level output signal for all zero crossings of the speech signal above 2200 Hz. Thus, the occurrence of unvoiced sounds is determined. This, is possible because the highest significant voiced sound zero crossing occurs at or below 2200 Hz.

The output of the voicing detector, along with the pitch detector output, is used to drive an analog gate. Here, the two are combined to form the pitch/unvoiced decision parameter.

Figure 21 shows the pitch detector input and output. This output is proportional to the input occurrence of the pitch pulses.

The output of the voicing detector, as shown in Figure 23, is a bi-level signal which says that when the level is +v, the speech input is unvoiced, and when the level is -v, the speech input is voiced. Figure 24 shows the addition of the unvoiced decision and pitch. These two parameters can be combined as one because in speech they never occur simultaneously. The speech is either voiced or unvoiced, but not both.

The combined parameterized speech output of the processor analyzer is shown in Figure 25. These parameters are individually fed to the multiplexer, shown in Figure 26. The multiplexer operates on these parameters and combines them in a process similar to that described in Figure 27 through Figure 30. The diagram of Figure 30 represents the composite analyzer output. However, this output is passed through a 160 Hz low pass filter and the final compressed speech output of the analyzer is shown in Figure 31.

Figure 32 relates this compressed speech output to the speech input. As can be seen, the two signals are different.

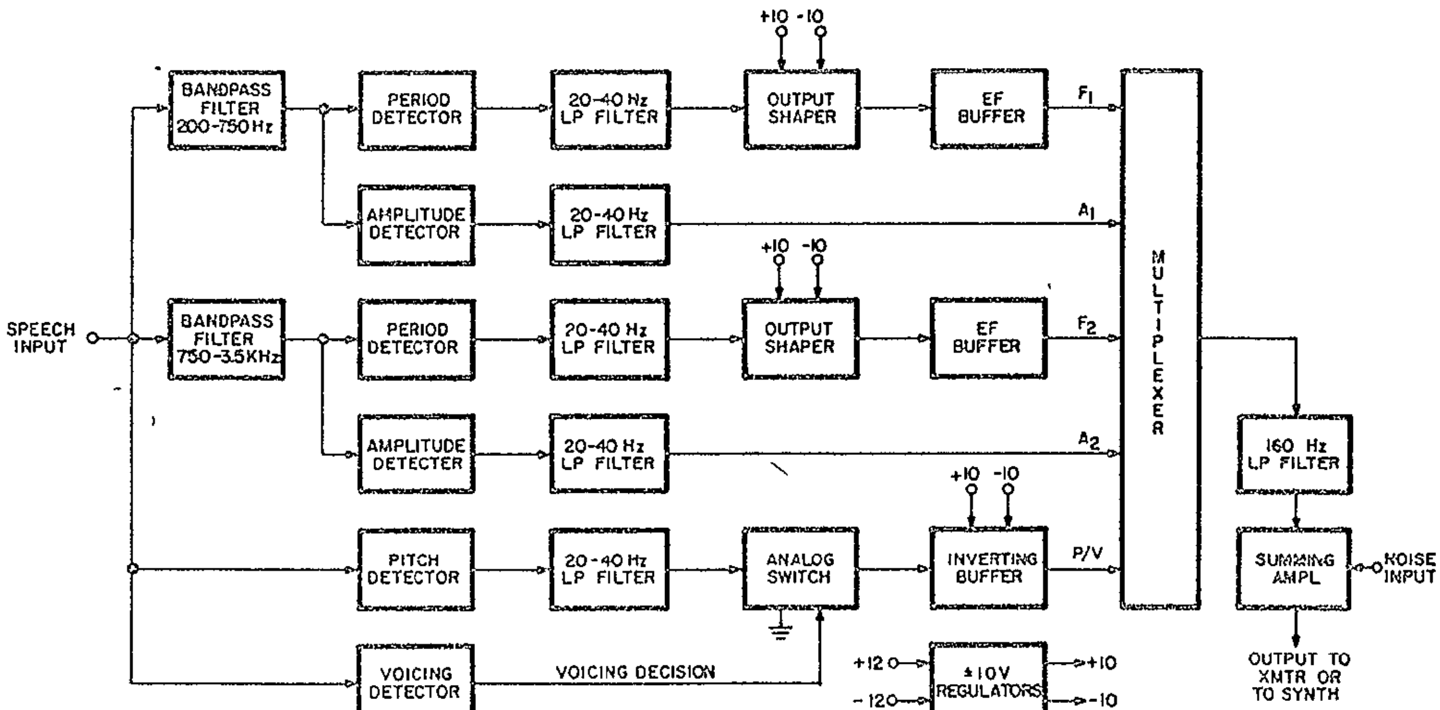


Fig. 7 Two-Formant Analyzer

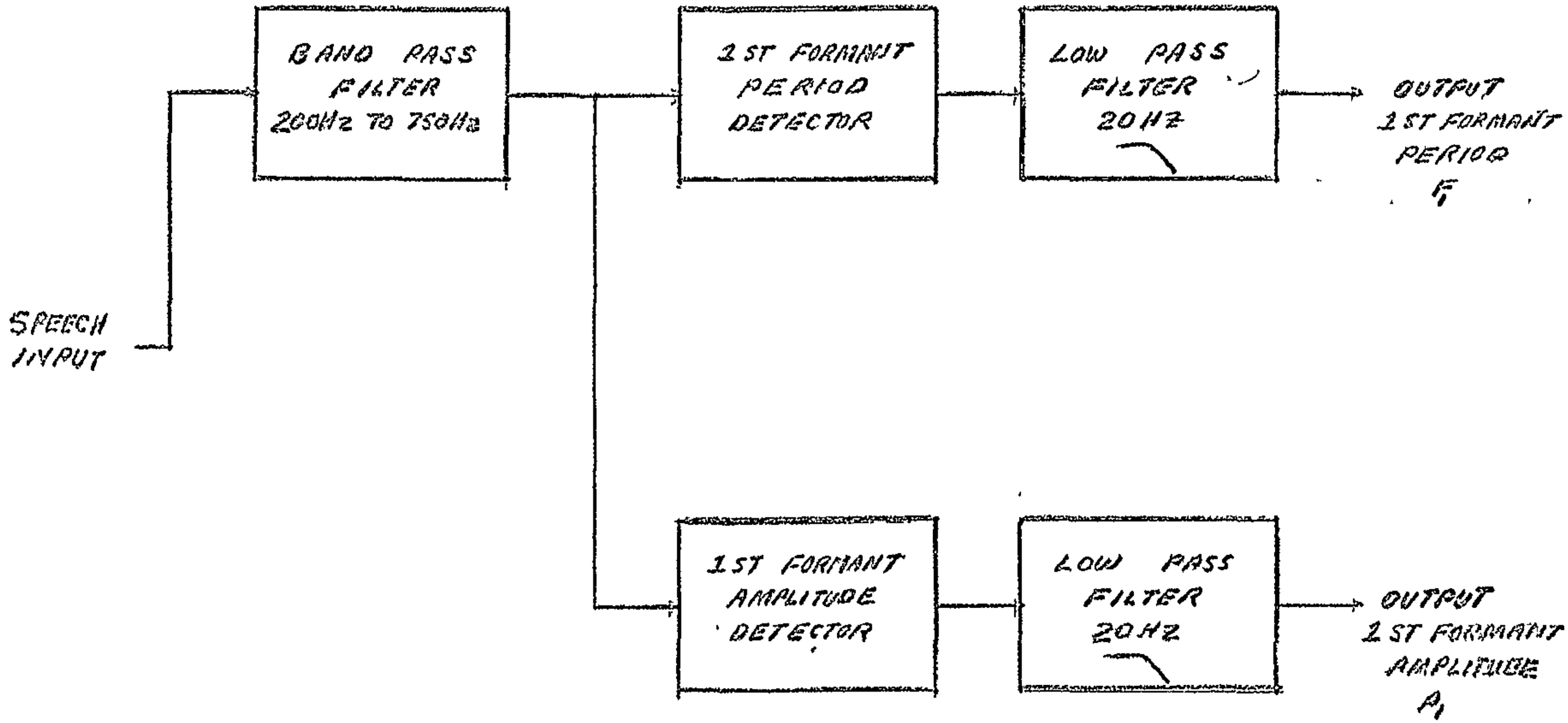


FIG. 8 BLOCK DIAGRAM OF 1 ST FORMANT PERIOD AND AMPLITUDE DETECTORS CHANNELS 1 & 2

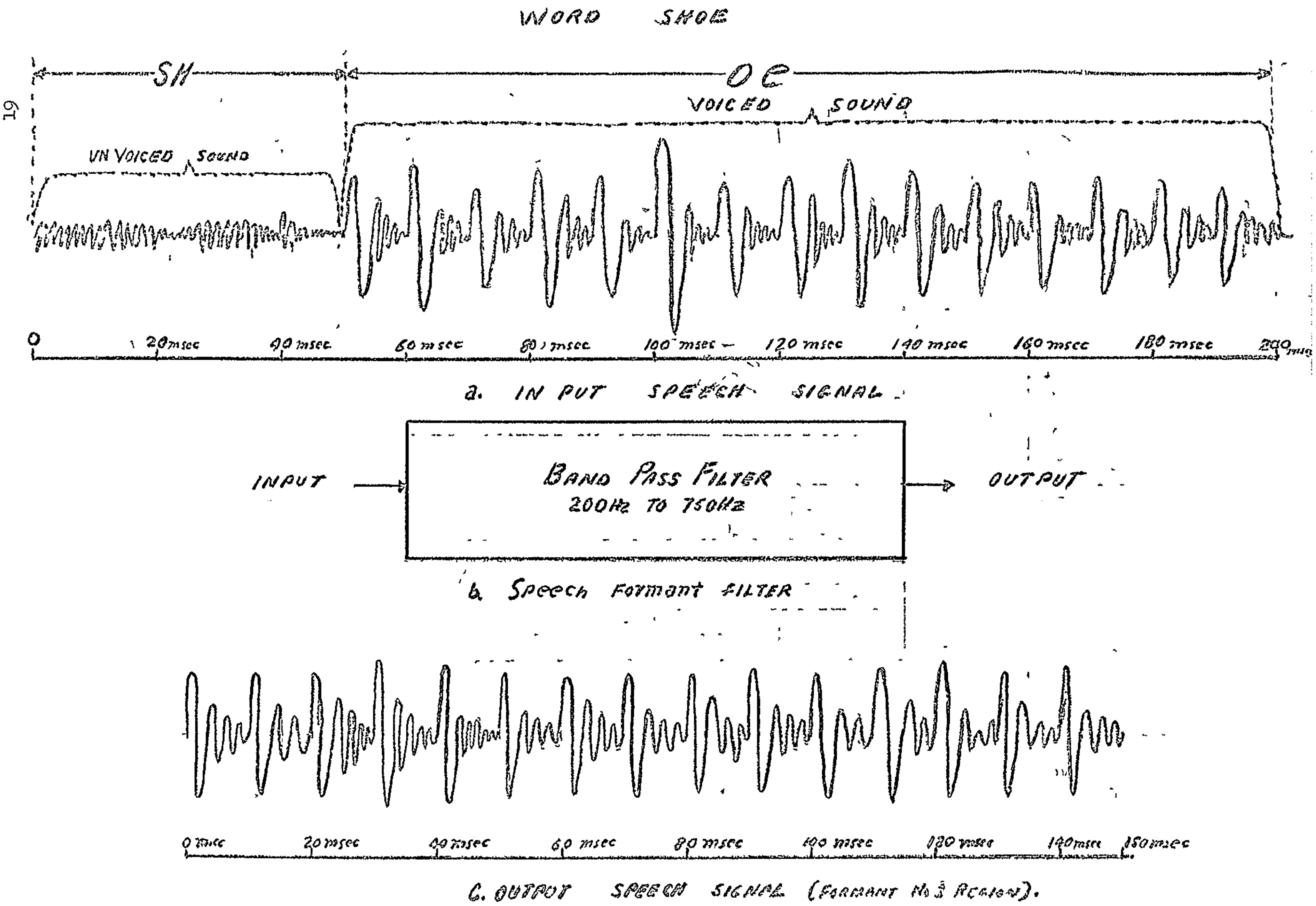

WORD SHOE
SH
OE
VOICED SOUND
UN VOICED SOUND
0
20 msec
40 msec
60 msec
80 msec
100 msec
120 msec
140 msec
160 msec
180 msec
200 msec
a. IN PUT SPEECH SIGNAL
INPUT
BAND PASS FILTER
200Hz TO 750Hz
OUTPUT
b. Speech Formant FILTER
0 msec
20 msec
40 msec
60 msec
80 msec
100 msec
120 msec
140 msec
150 msec
C. OUTPUT SPEECH SIGNAL (FORMANT No 3 REGION).

FIG. 9

20

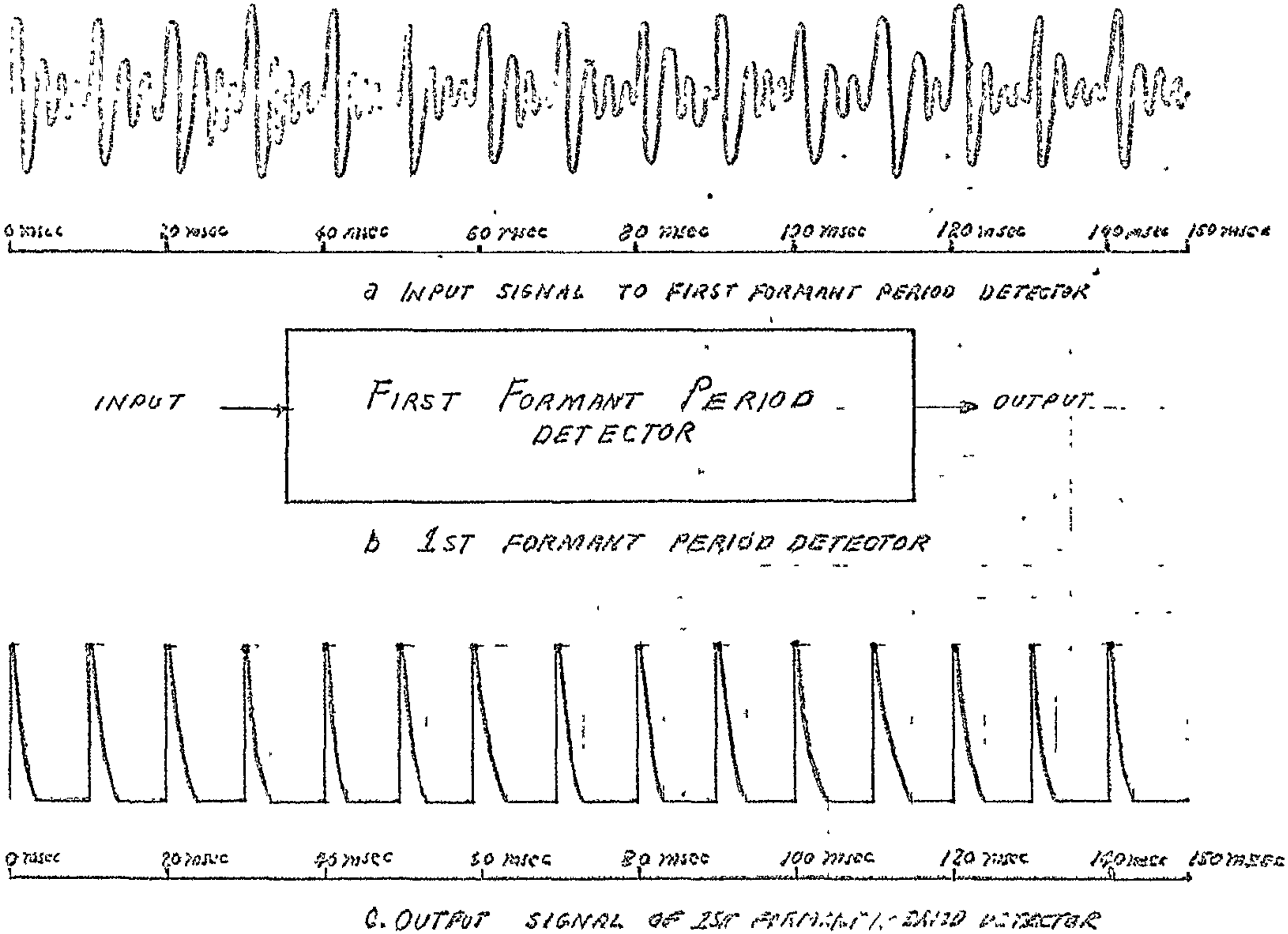
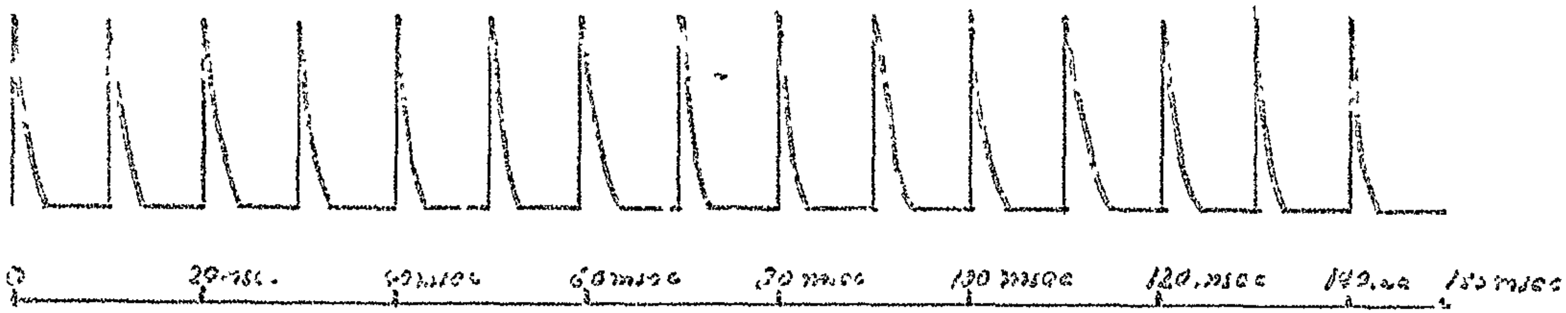
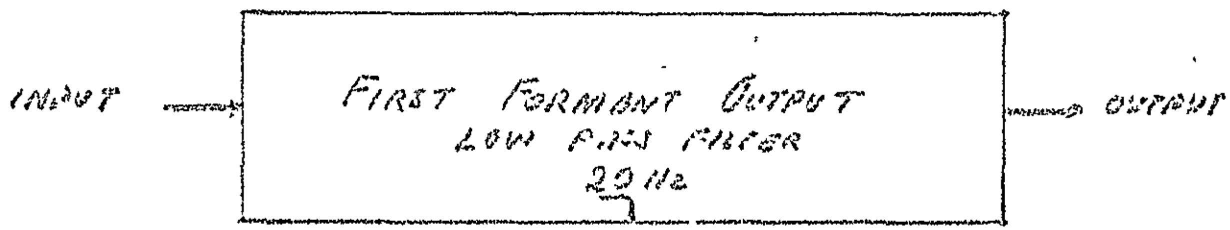

0 msec
20 msec
40 msec
60 msec
80 msec
100 msec
120 msec
140 msec
150 msec
a INPUT SIGNAL TO FIRST FORMANT PERIOD DETECTOR
INPUT
FIRST FORMANT PERIOD DETECTOR
OUTPUT
b 1ST FORMANT PERIOD DETECTOR
0 msec
20 msec
40 msec
60 msec
80 msec
100 msec
120 msec
140 msec
150 msec
c. OUTPUT SIGNAL OF 1ST FORMANT PERIOD DETECTOR

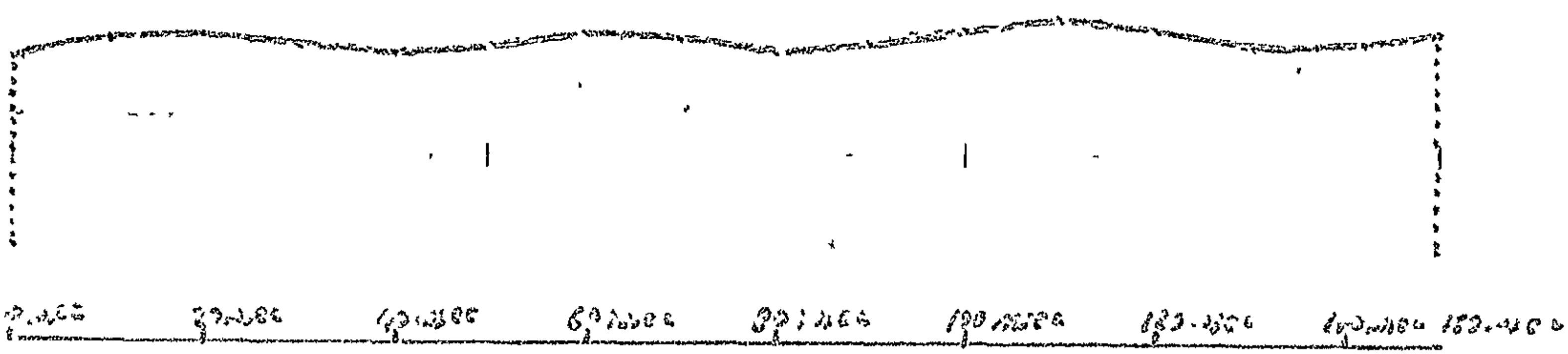
FIG 10



a. INPUT TO 1ST FORMANT OUTPUT LOW PASS FILTER



b 1ST FORMANT OUTPUT LOW PASS FILTER 20Hz



c. OUTPUT SIGNAL OF 1ST FORMANT OUTPUT LOW PASS FILTER 20Hz

FIG 11

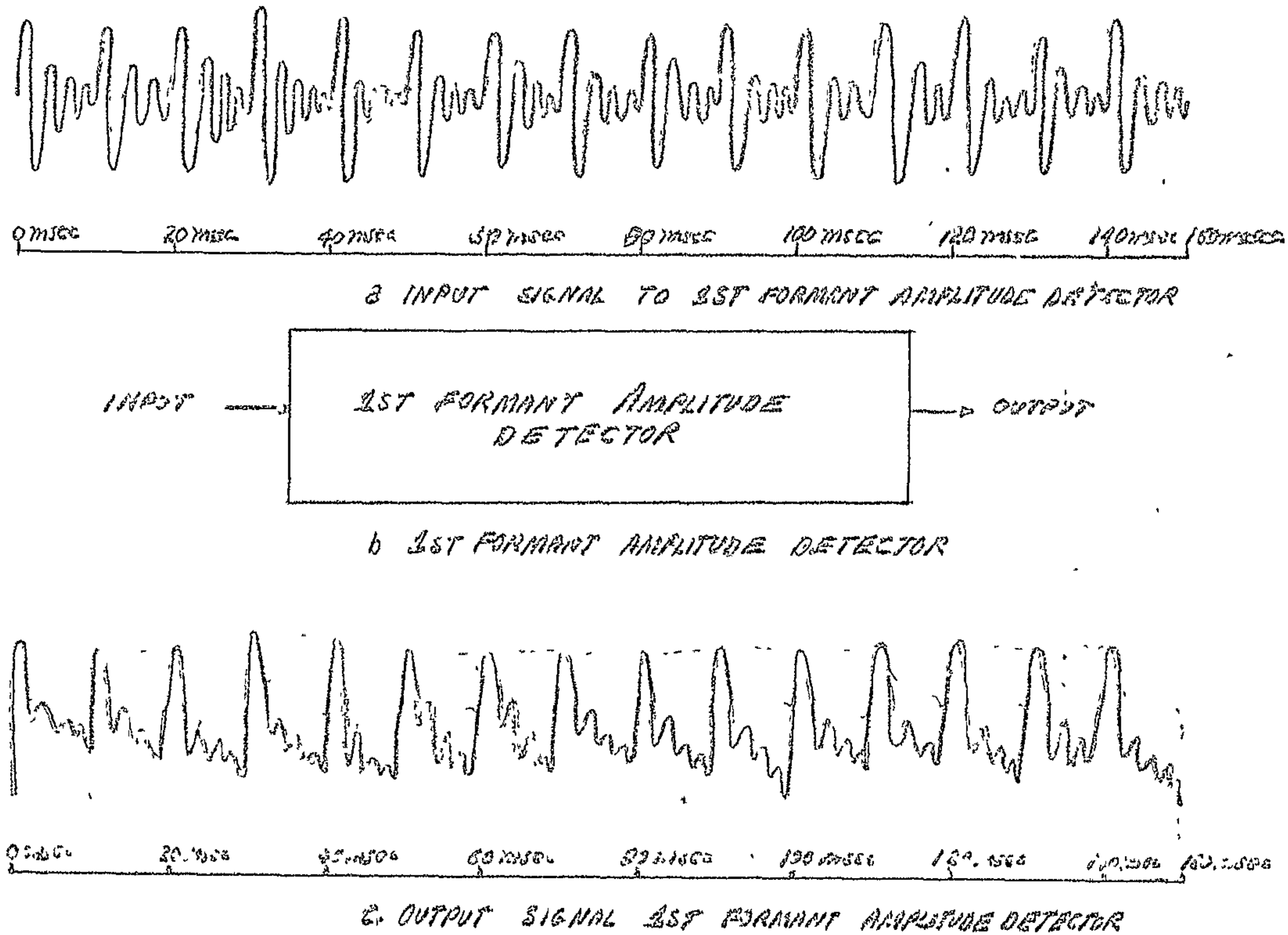

0 msec
20 msec
40 msec
60 msec
80 msec
100 msec
120 msec
140 msec
160 msec
a INPUT SIGNAL TO 1ST FORMANT AMPLITUDE DETECTOR
INPUT
1ST FORMANT AMPLITUDE DETECTOR
OUTPUT
b 1ST FORMANT AMPLITUDE DETECTOR
c OUTPUT SIGNAL 1ST FORMANT AMPLITUDE DETECTOR

FIG 12

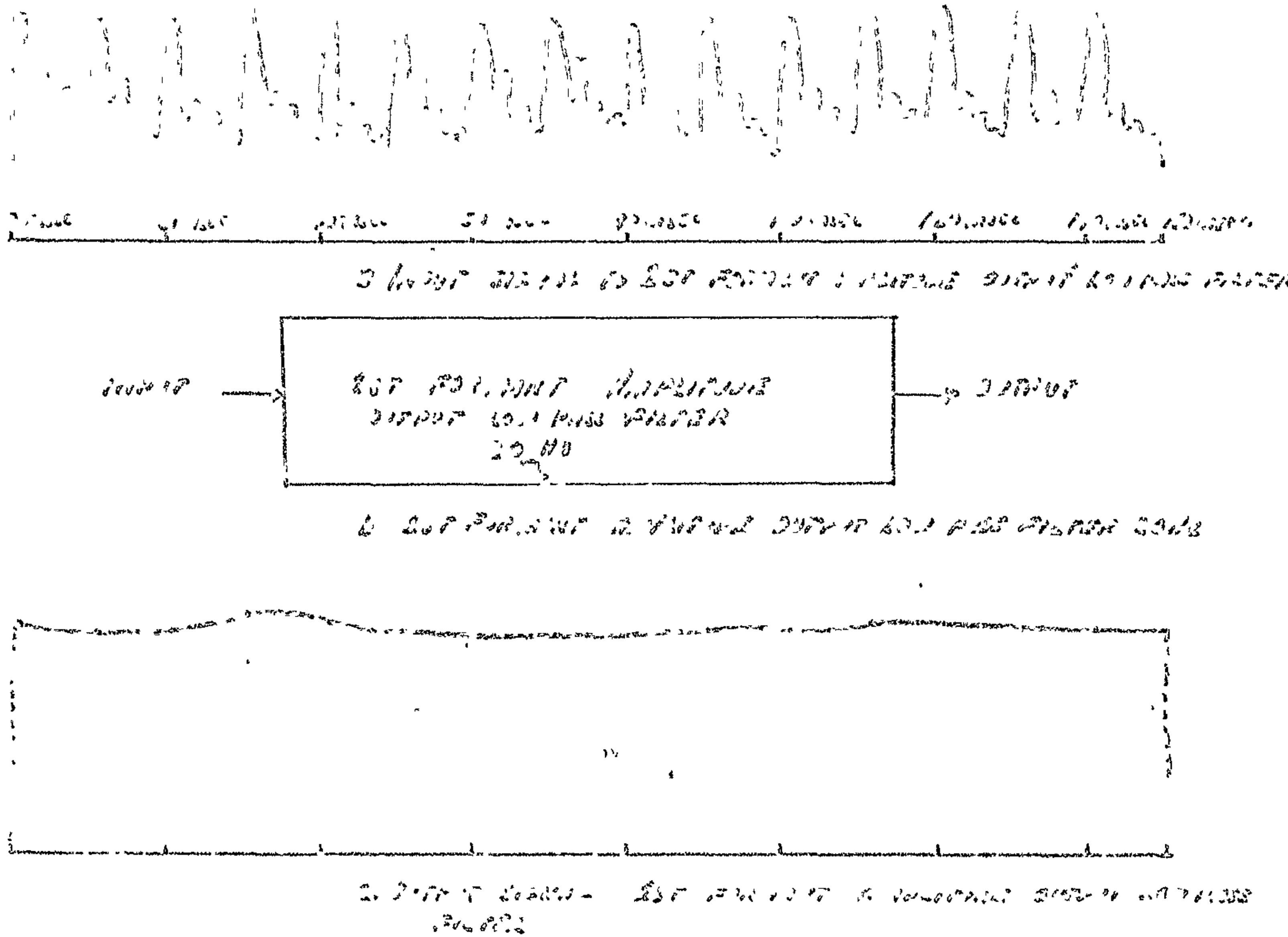


FIG. 12

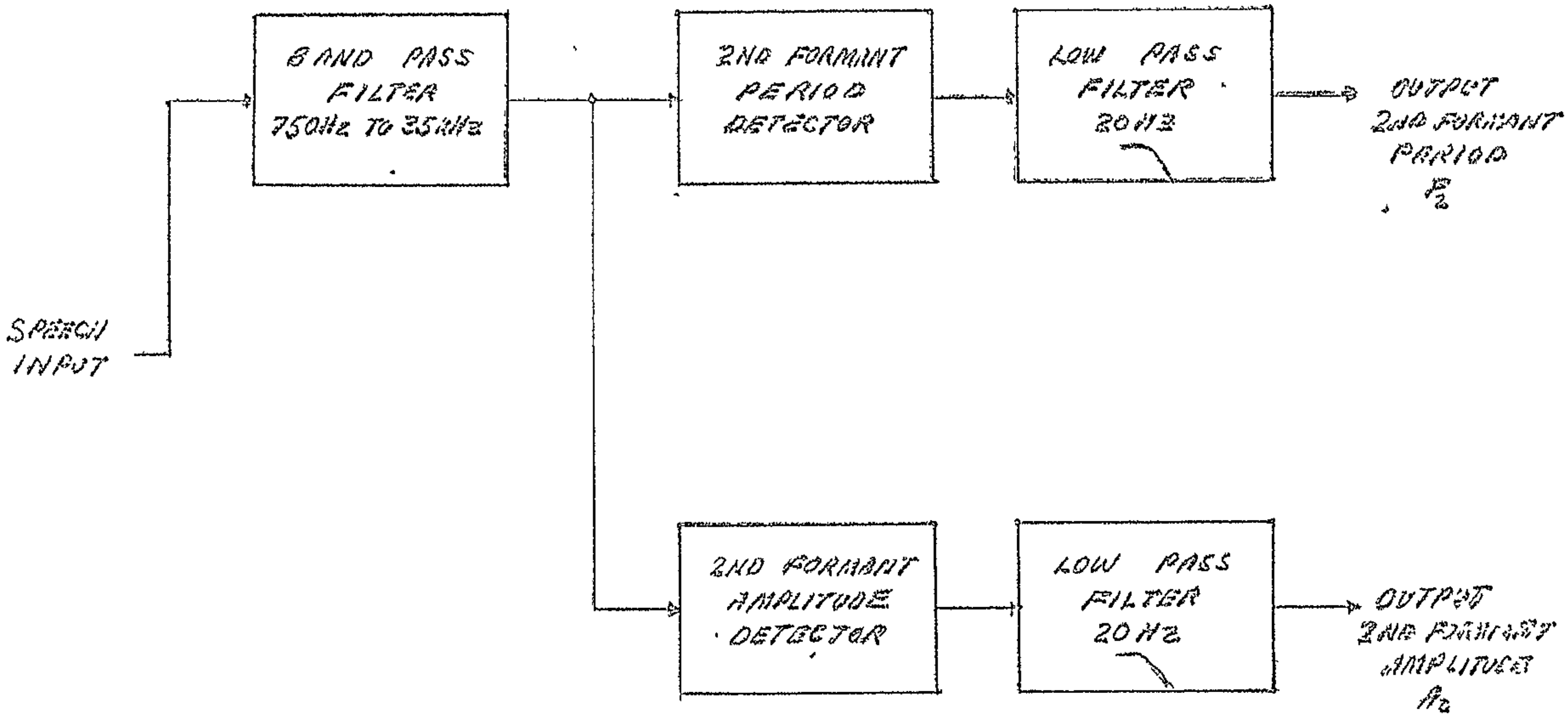


FIG 13 BLOCK DIAGRAM OF 2ND FORMANT PERIOD AND AMPLITUDE DETECTORS CHANNELS 3 & 4

FIG 14

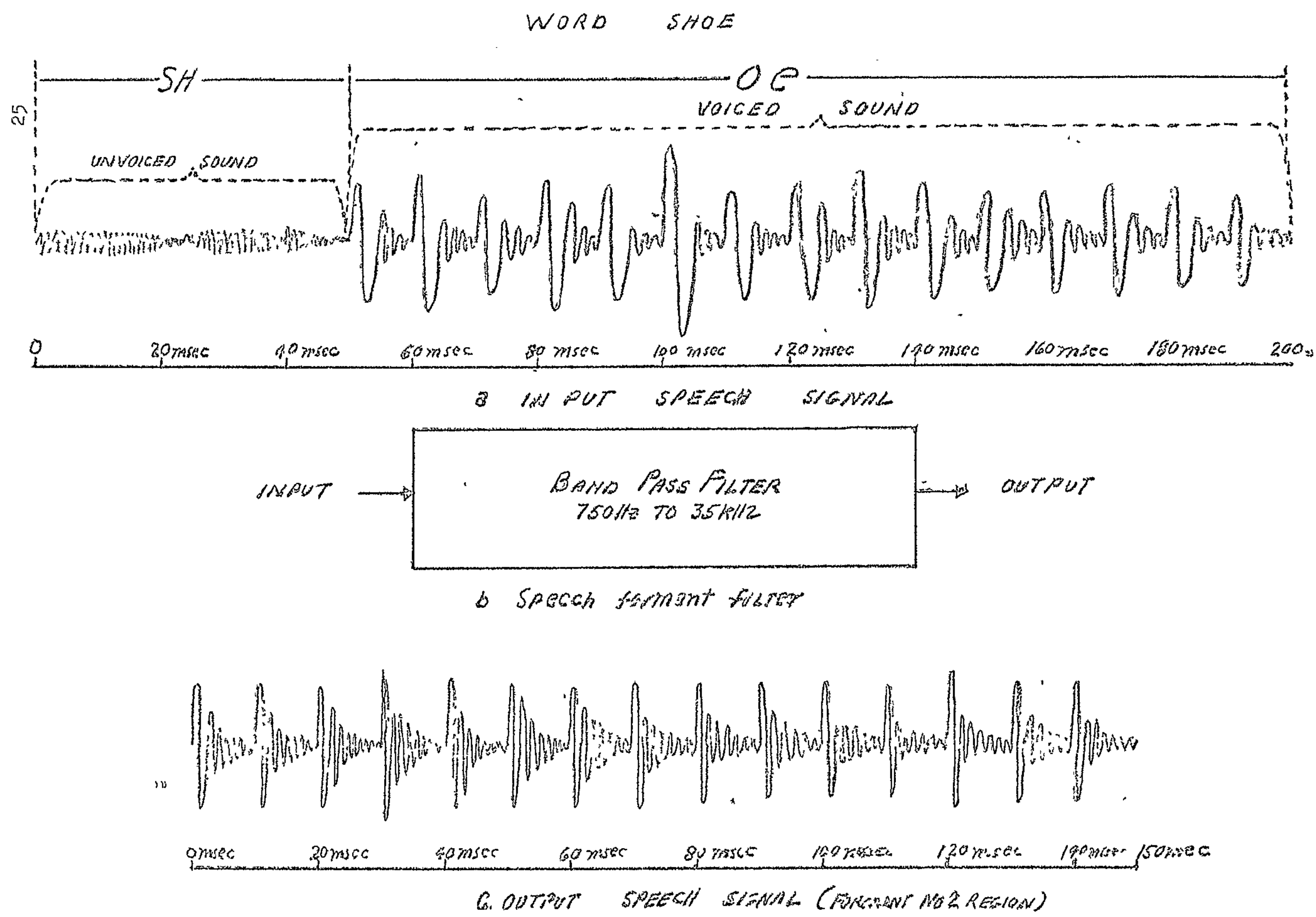
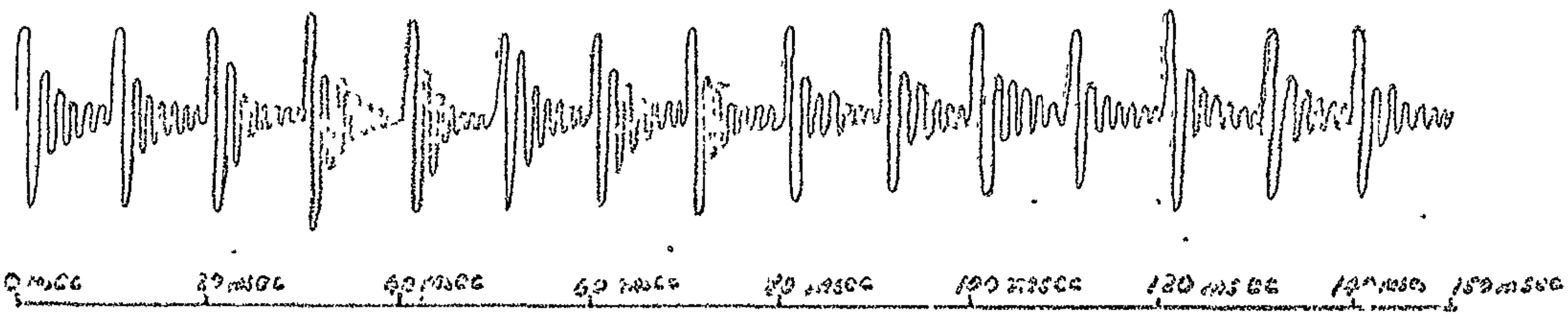
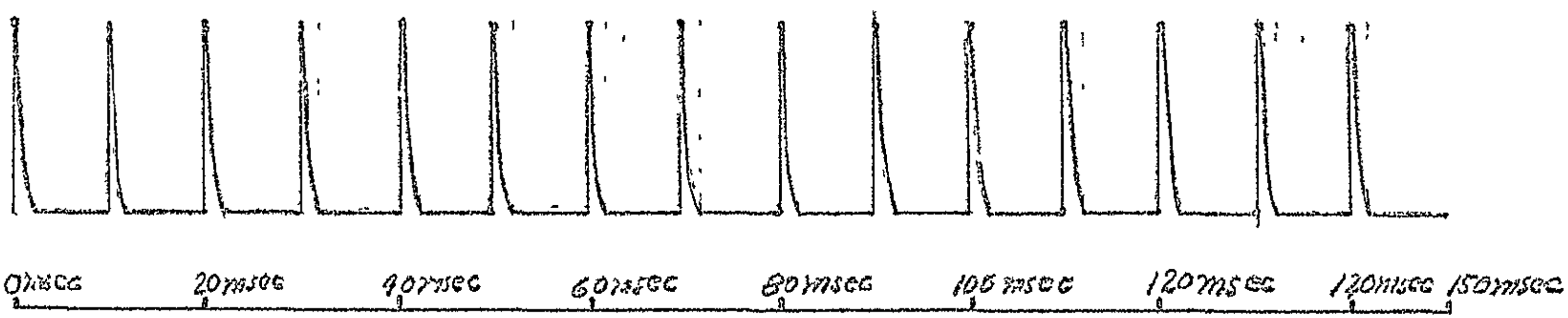

WORD SHOE
SH
OE
VOICED SOUND
UNVOICED SOUND
0
20 msec
40 msec
60 msec
80 msec
100 msec
120 msec
140 msec
160 msec
180 msec
200
a INPUT SPEECH SIGNAL
INPUT
BAND PASS FILTER
750Hz TO 3.5KHZ
OUTPUT
b SPEECH FORMANT FILTER
0 msec
20 msec
40 msec
60 msec
80 msec
100 msec
120 msec
140 msec
150 msec
C. OUTPUT SPEECH SIGNAL (FORMANT NO 2 REGION)

FIG. 15



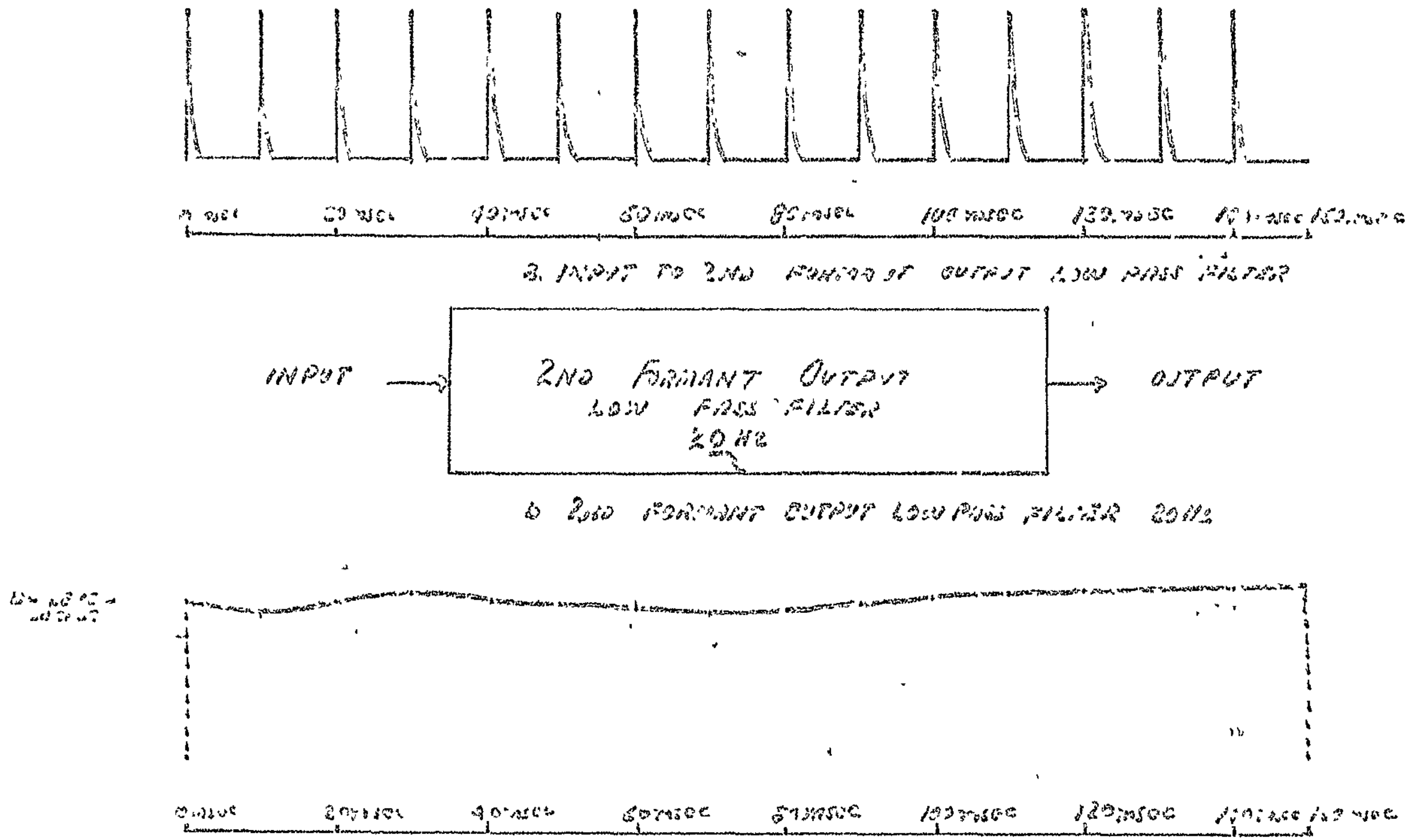
a. INPUT SIGNAL TO SECOND FORMANT PERIOD DETECTOR

b. 2nd FORMANT PERIOD DETECTOR



c. OUTPUT SIGNAL 2ND FORMANT PERIOD DETECTOR

FIG 16



a. INPUT TO 2ND FORMANT OUTPUT LOW PASS FILTER
INPUT
2ND FORMANT OUTPUT
LOW PASS FILTER
20 HZ
OUTPUT
b 2ND FORMANT OUTPUT LOW PASS FILTER 20 HZ

FIG. 17

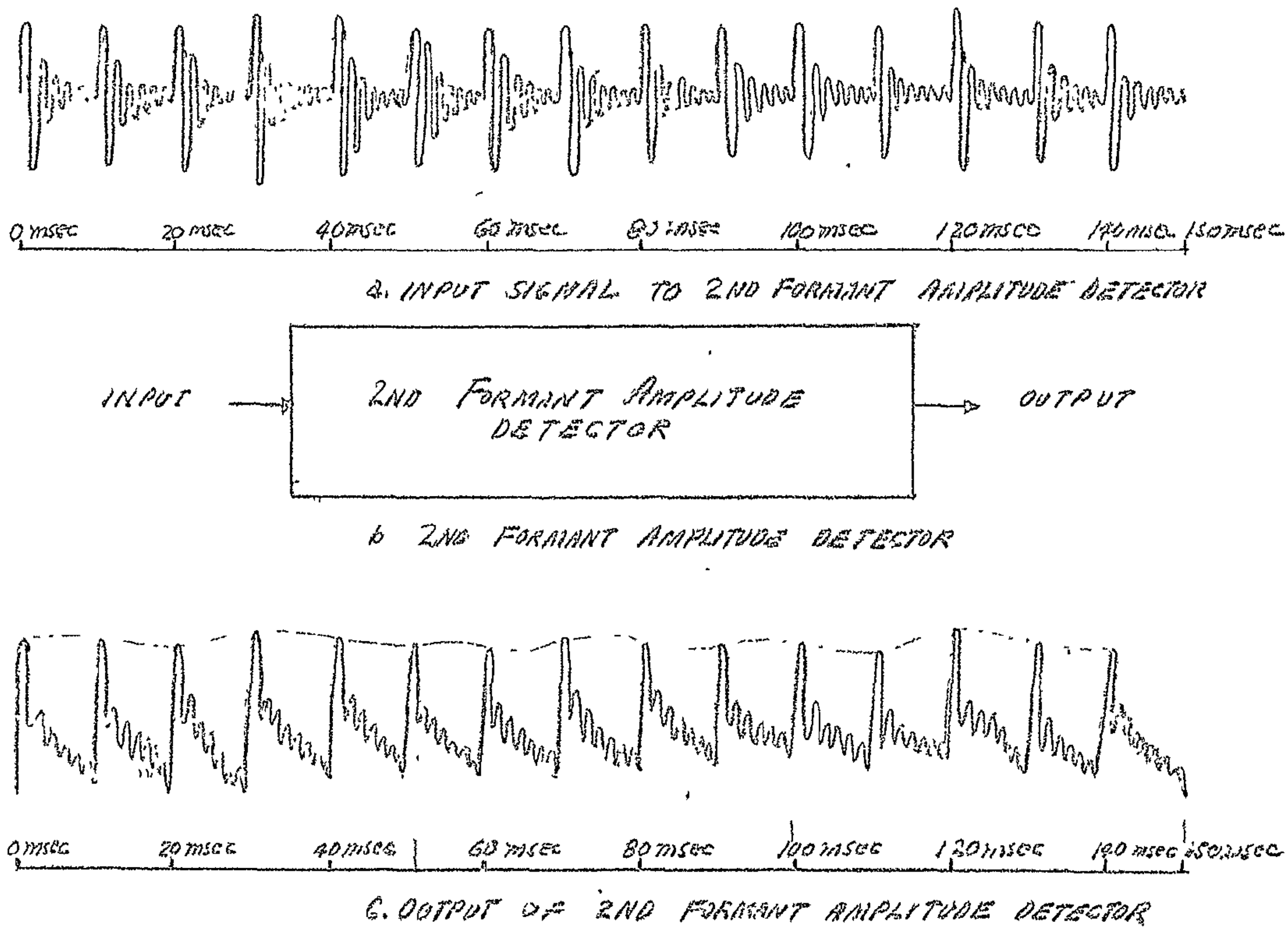
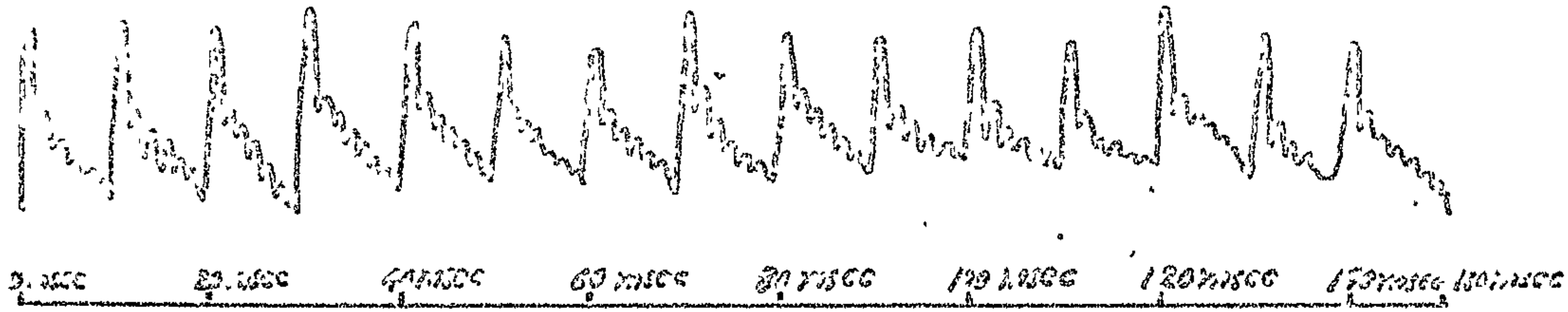
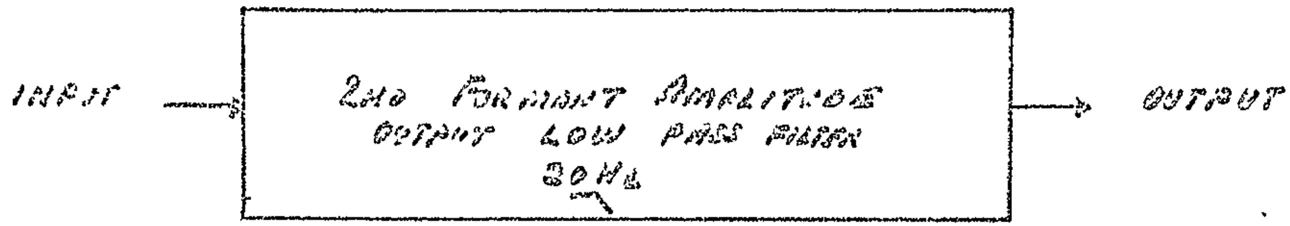

0 msec
20 msec
40 msec
60 msec
80 msec
100 msec
120 msec
140 msec
150 msec
a. INPUT SIGNAL TO 2ND FORMANT AMPLITUDE DETECTOR
INPUT
2ND FORMANT AMPLITUDE DETECTOR
OUTPUT
b 2ND FORMANT AMPLITUDE DETECTOR
0 msec
20 msec
40 msec
60 msec
80 msec
100 msec
120 msec
140 msec
150 msec
c. OUTPUT OF 2ND FORMANT AMPLITUDE DETECTOR

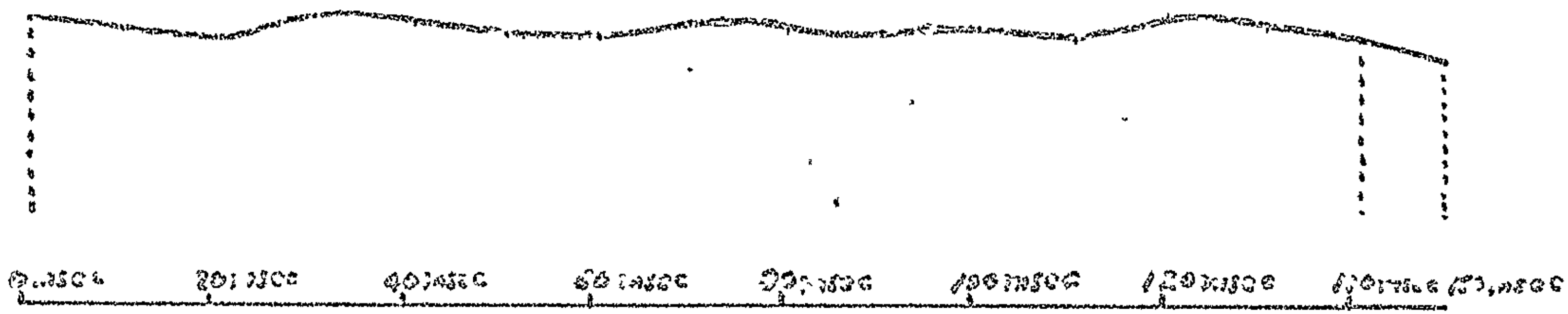
FIG 18



a. INPUT TO 2ND FORMANT AMPLITUDE LOW PASS FILTER



b. 2ND FORMANT AMPLITUDE OUTPUT LOW PASS FILTER 30HZ



c. OUTPUT 2ND FORMANT AMPLITUDE OUTPUT LOW PASS FILTER

FIG 19

30

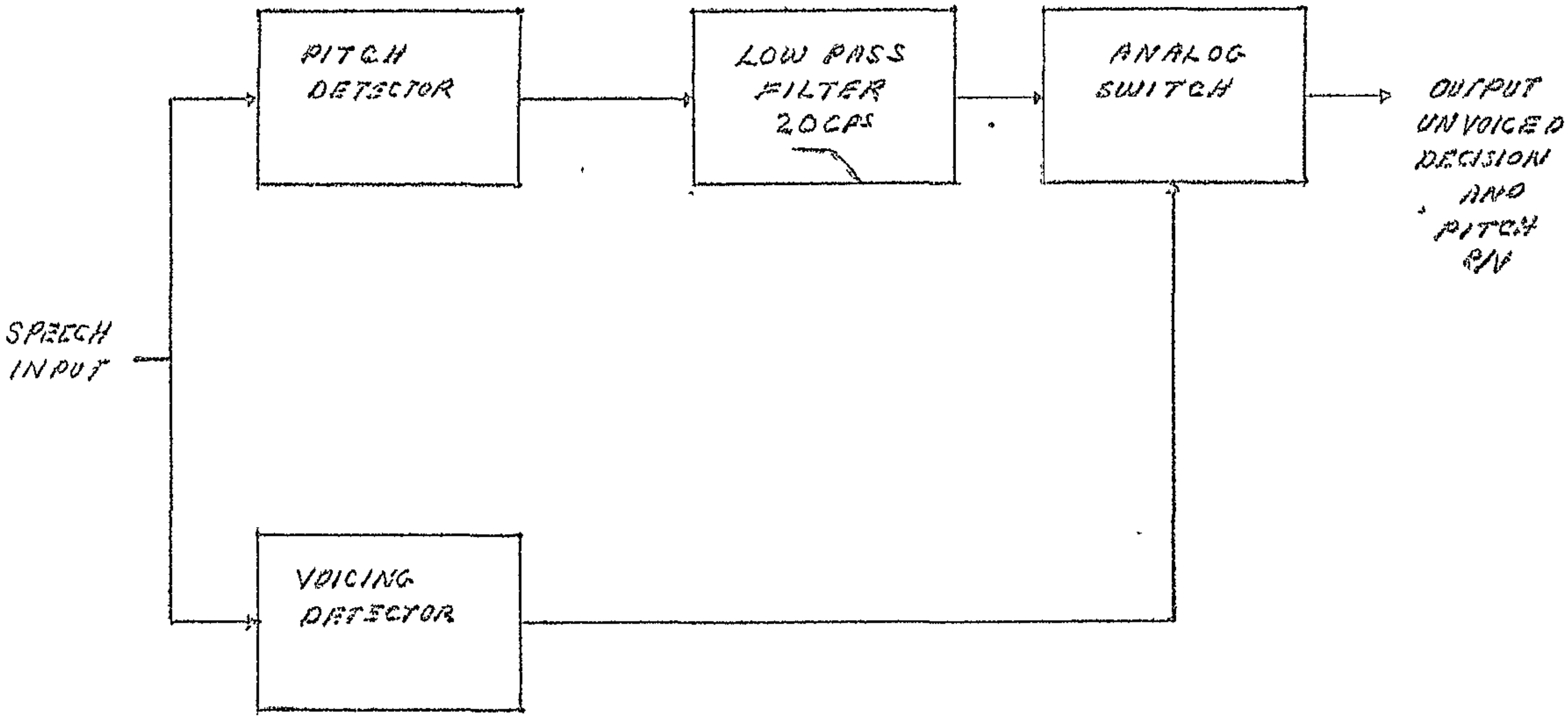


FIG BLOCK DIAGRAM OF PITCH AND VOICING DETECTORS CHANNEL 5

FIG 20

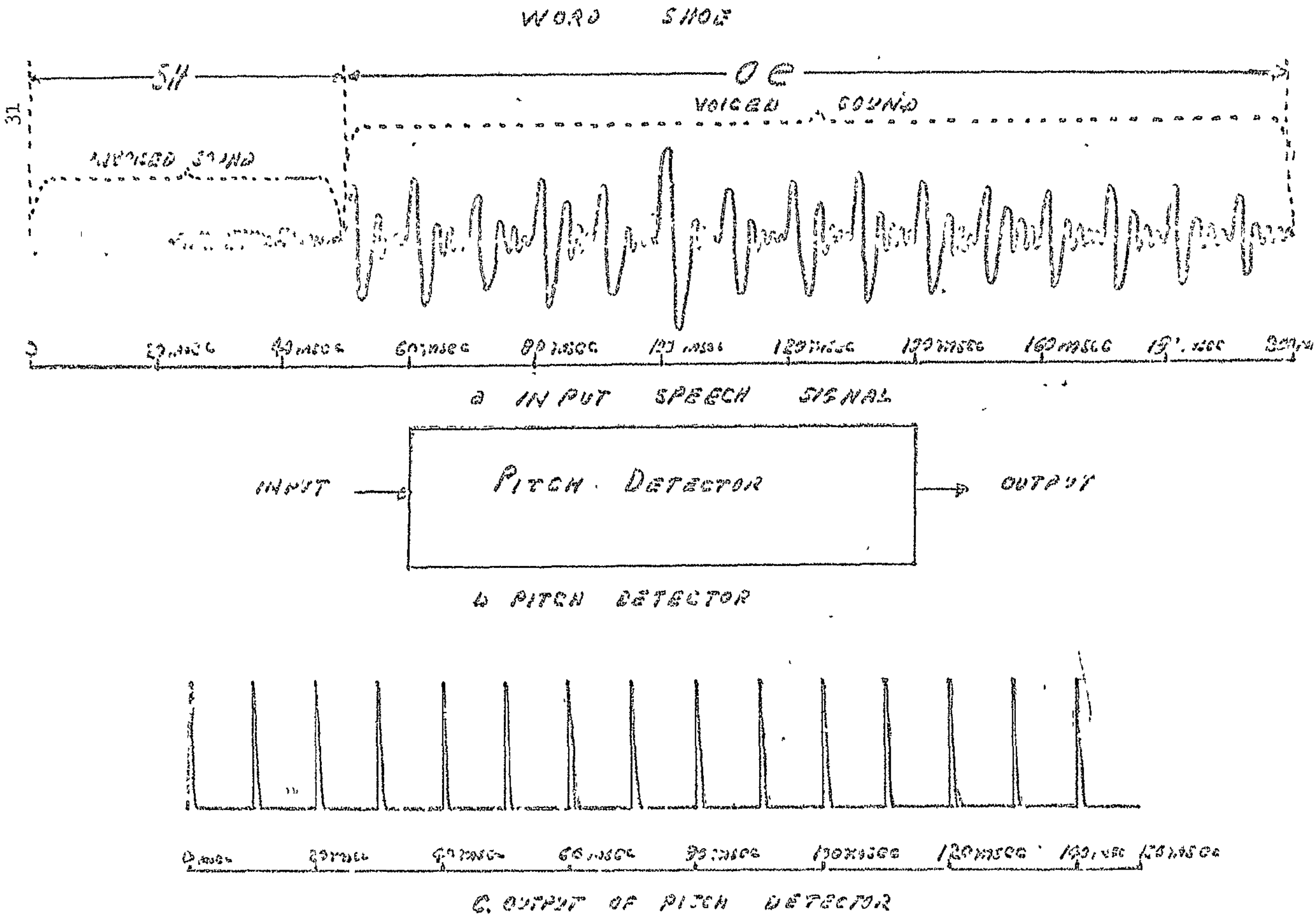

WORD SHOE
SH
OE
VOICED SOUND
UNVOICED SOUND
a INPUT SPEECH SIGNAL
INPUT
PITCH DETECTOR
OUTPUT
b PITCH DETECTOR
C. OUTPUT OF PITCH DETECTOR

FIG 21

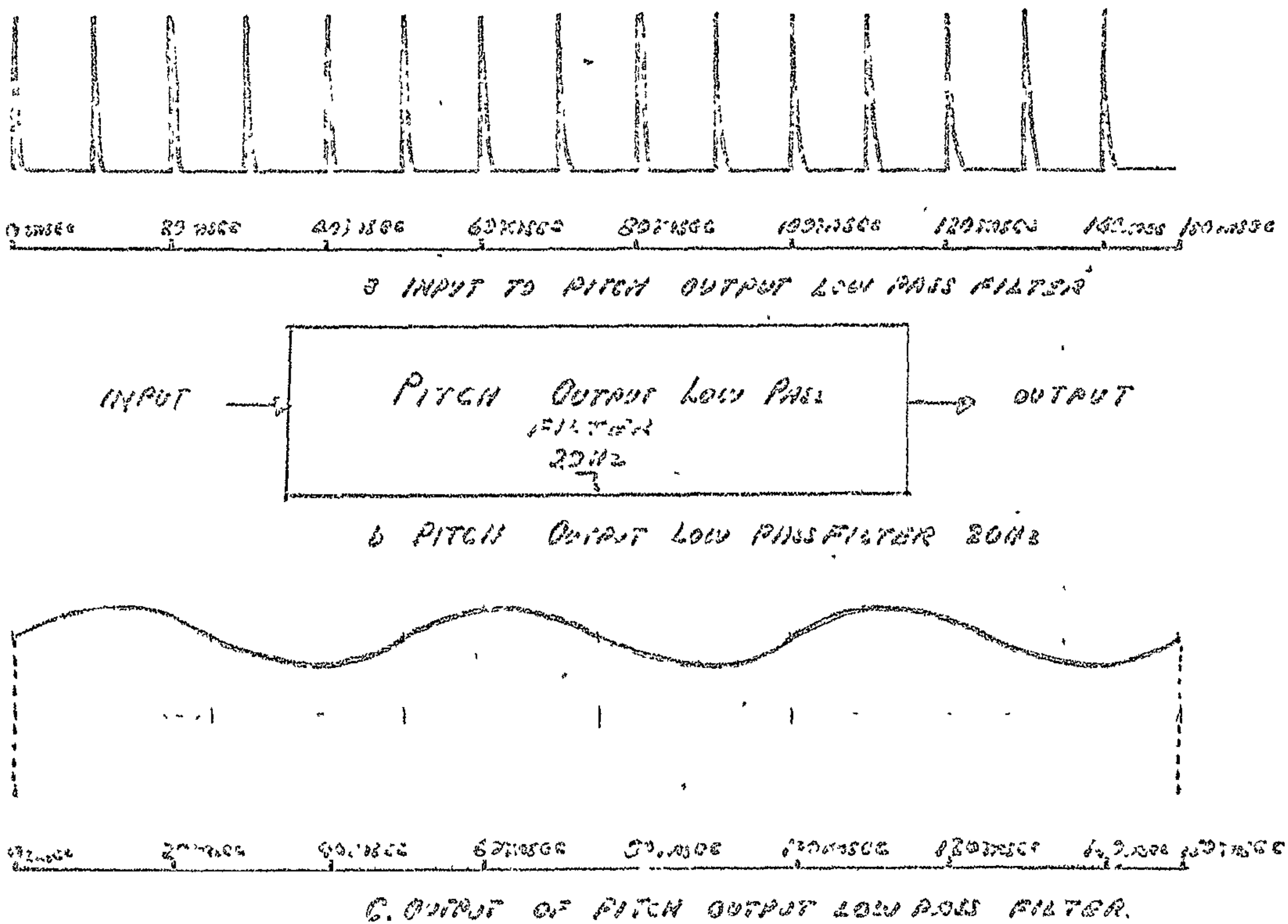

0 mSEC
20 mSEC
40 mSEC
60 mSEC
80 mSEC
100 mSEC
120 mSEC
140 mSEC
160 mSEC
a INPUT TO PITCH OUTPUT LOW PASS FILTER
INPUT
PITCH OUTPUT LOW PASS FILTER 20Hz
OUTPUT
b PITCH OUTPUT LOW PASS FILTER 20HZ
c. OUTPUT OF PITCH OUTPUT LOW PASS FILTER.

FIG 22

33

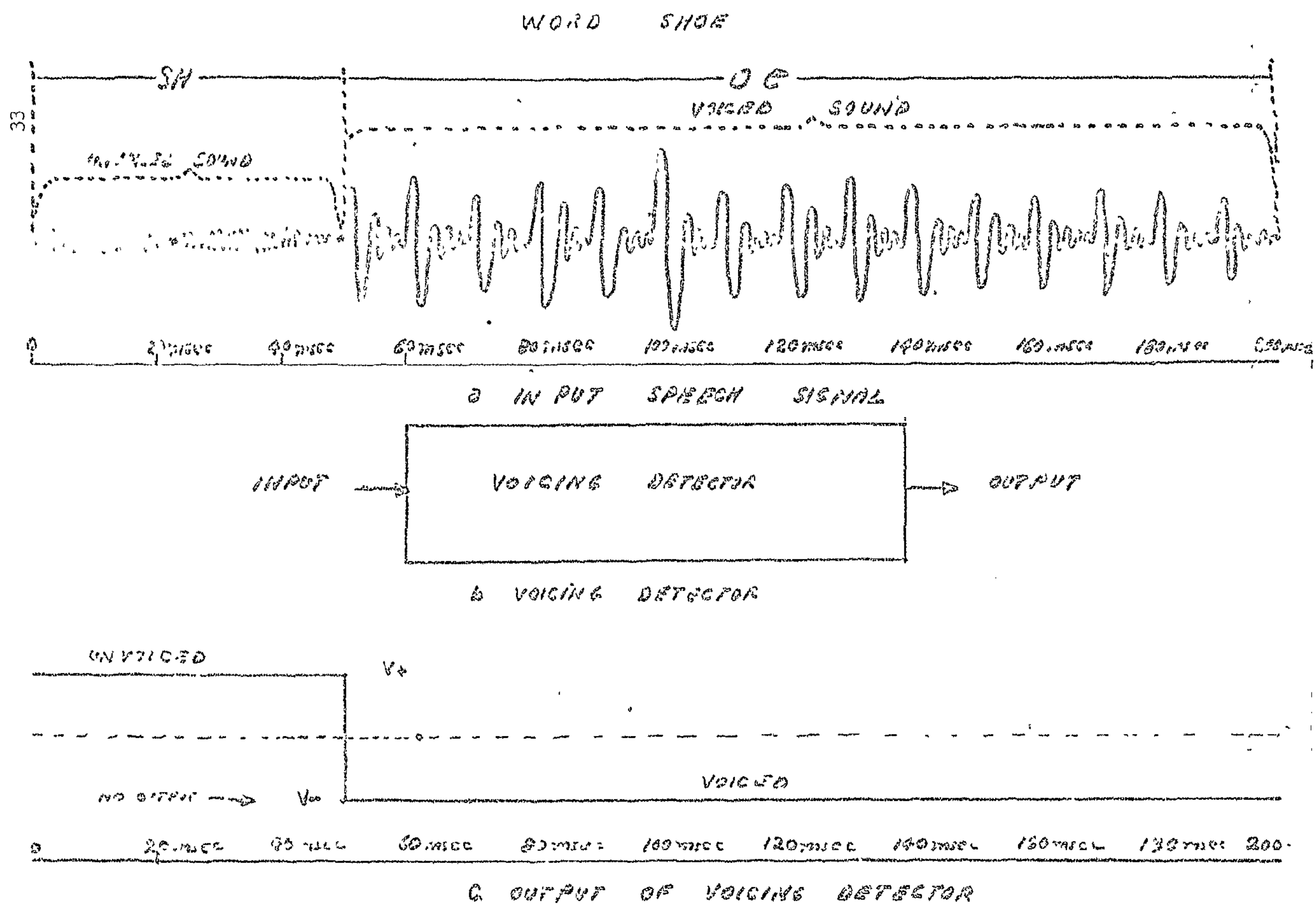
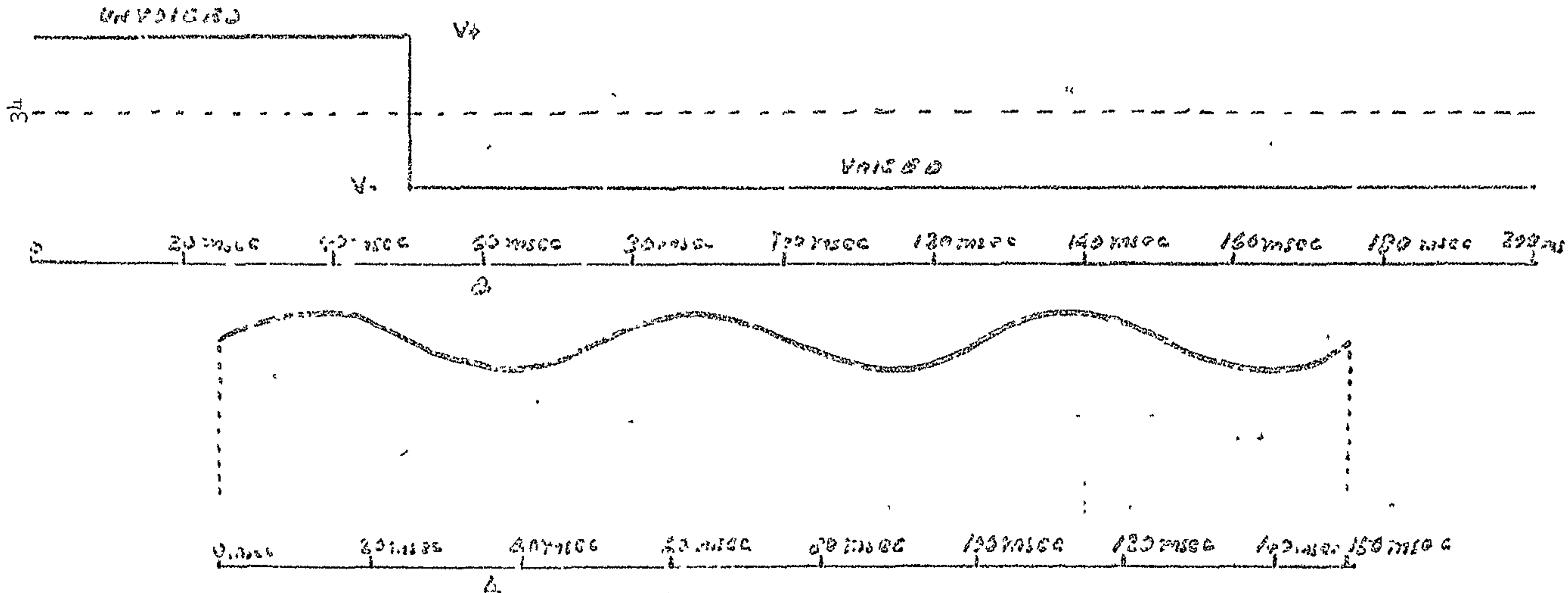
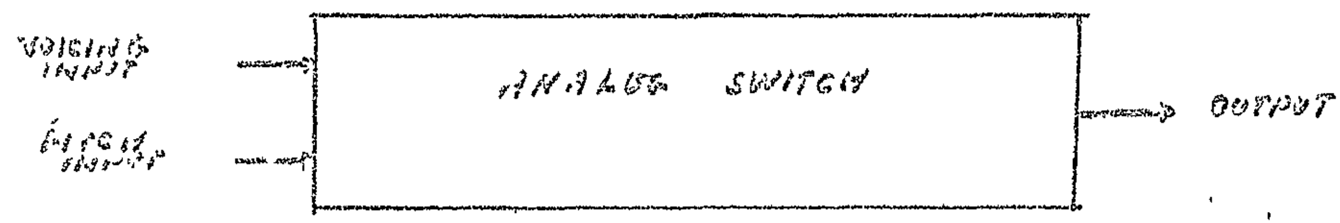

WORD SHOE
SH
OE
VOICED SOUND
0
20 msec
40 msec
60 msec
80 msec
100 msec
120 msec
140 msec
160 msec
180 msec
200 msec
a INPUT SPEECH SIGNAL
INPUT
VOICING DETECTOR
OUTPUT
b VOICING DETECTOR
UNVOICED
V+
VOICED
NO OUTPUT
V-
c OUTPUT OF VOICING DETECTOR

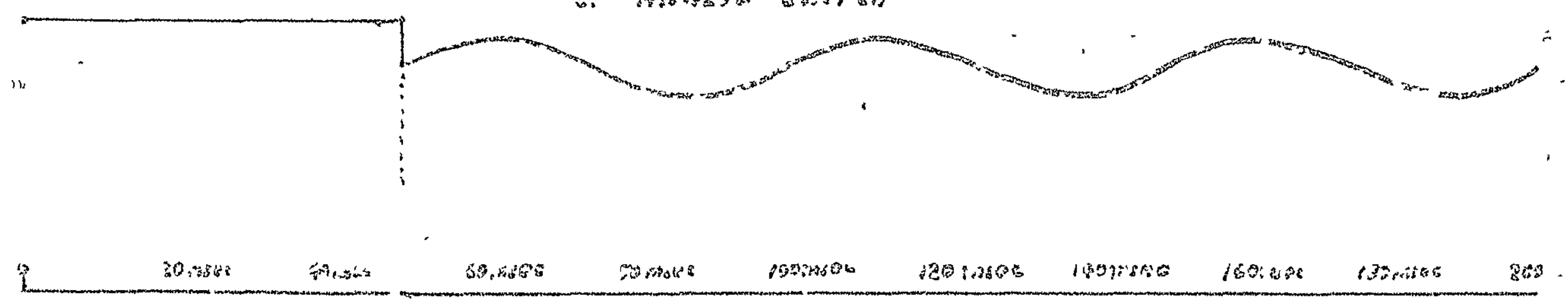
Fig 22

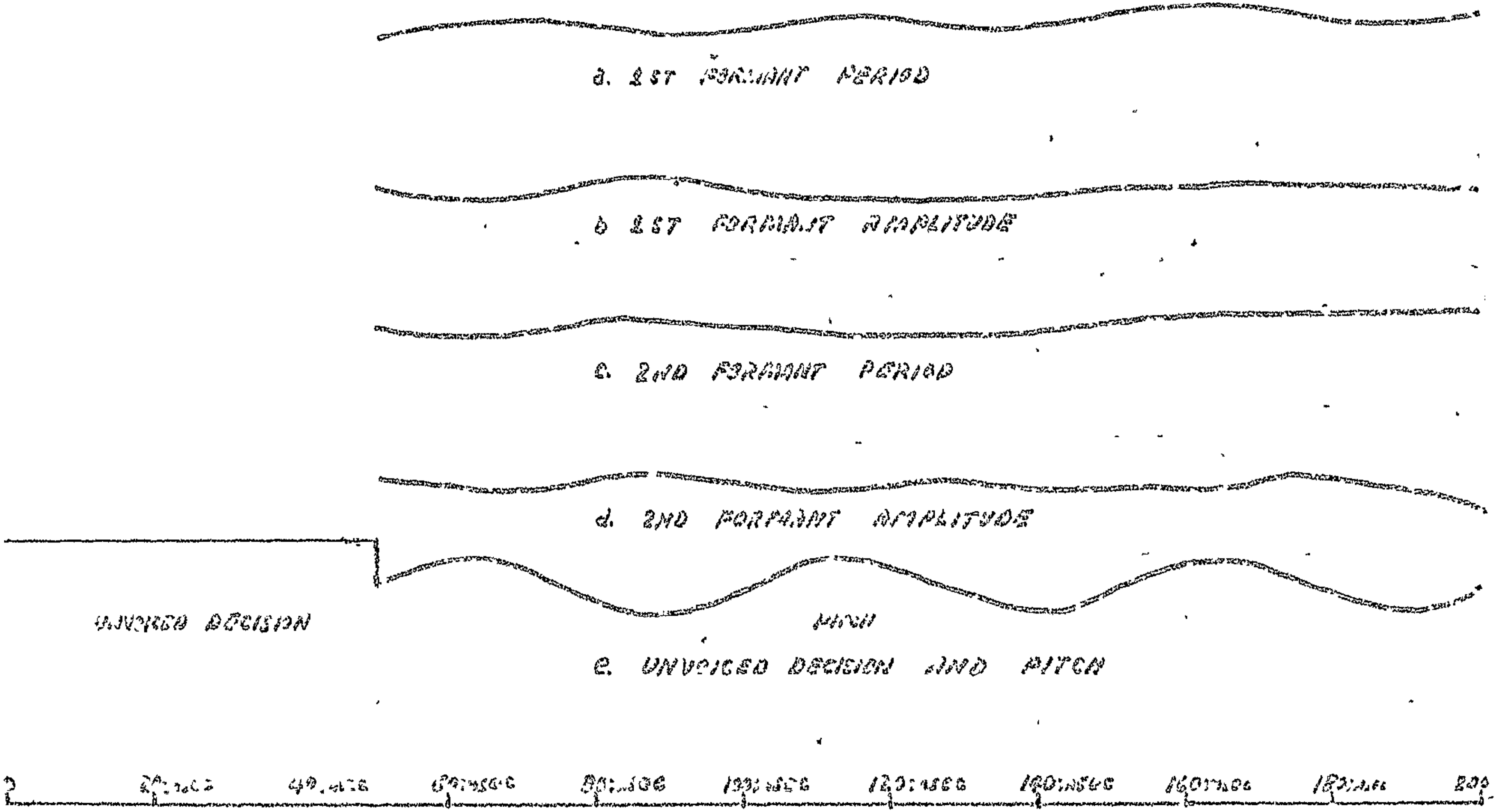


a. & b. INPUTS TO ANALOG SWITCH



c. ANALOG SWITCH





SPEECH PROCESSOR ANALYZER OUTPUT BEFORE MULTIPLEXING
FOR THE WORD "[illegible]"

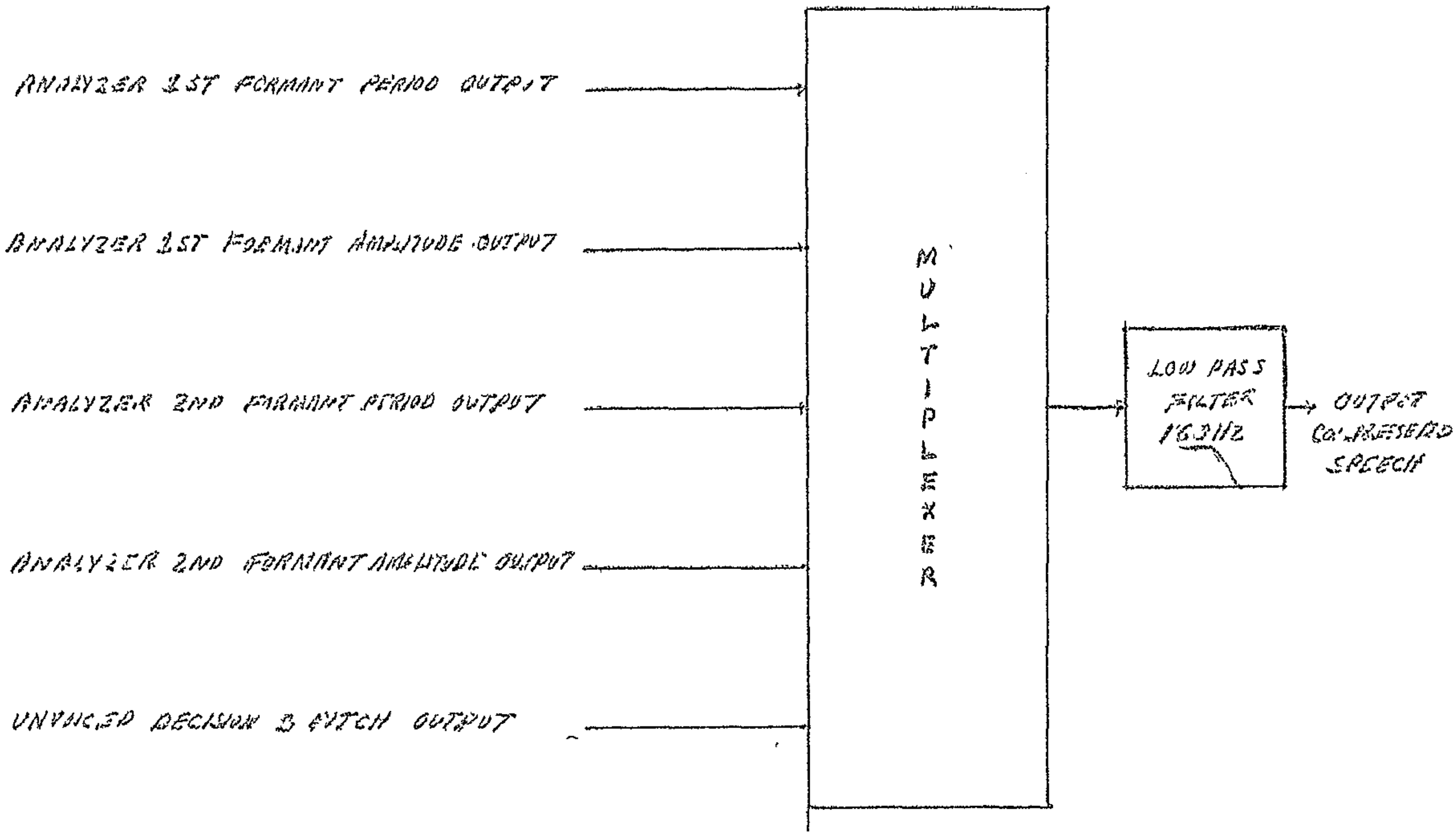


FIG 26 BLOCK DIAGRAM OF ANALYZER MULTIPLEXER

37

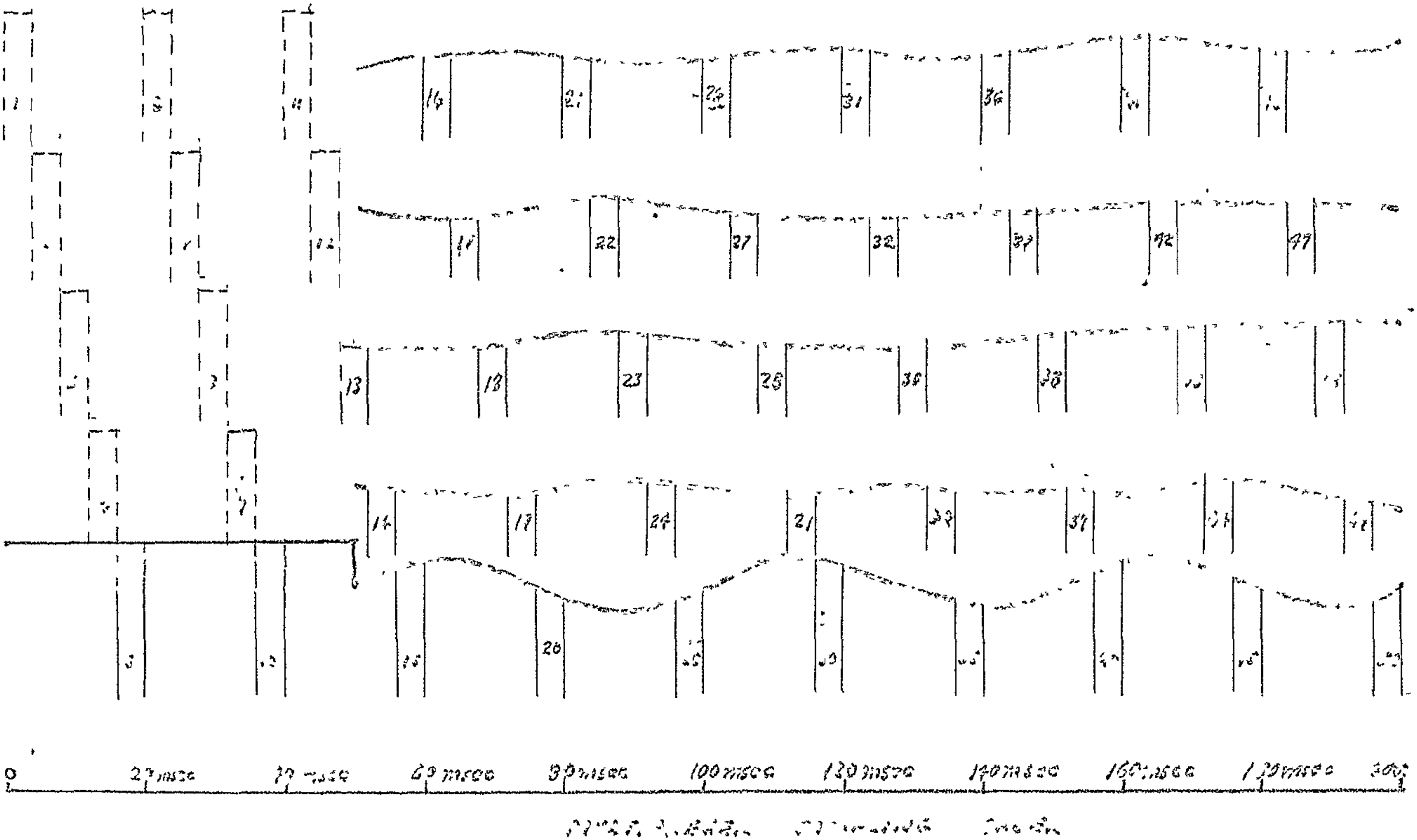
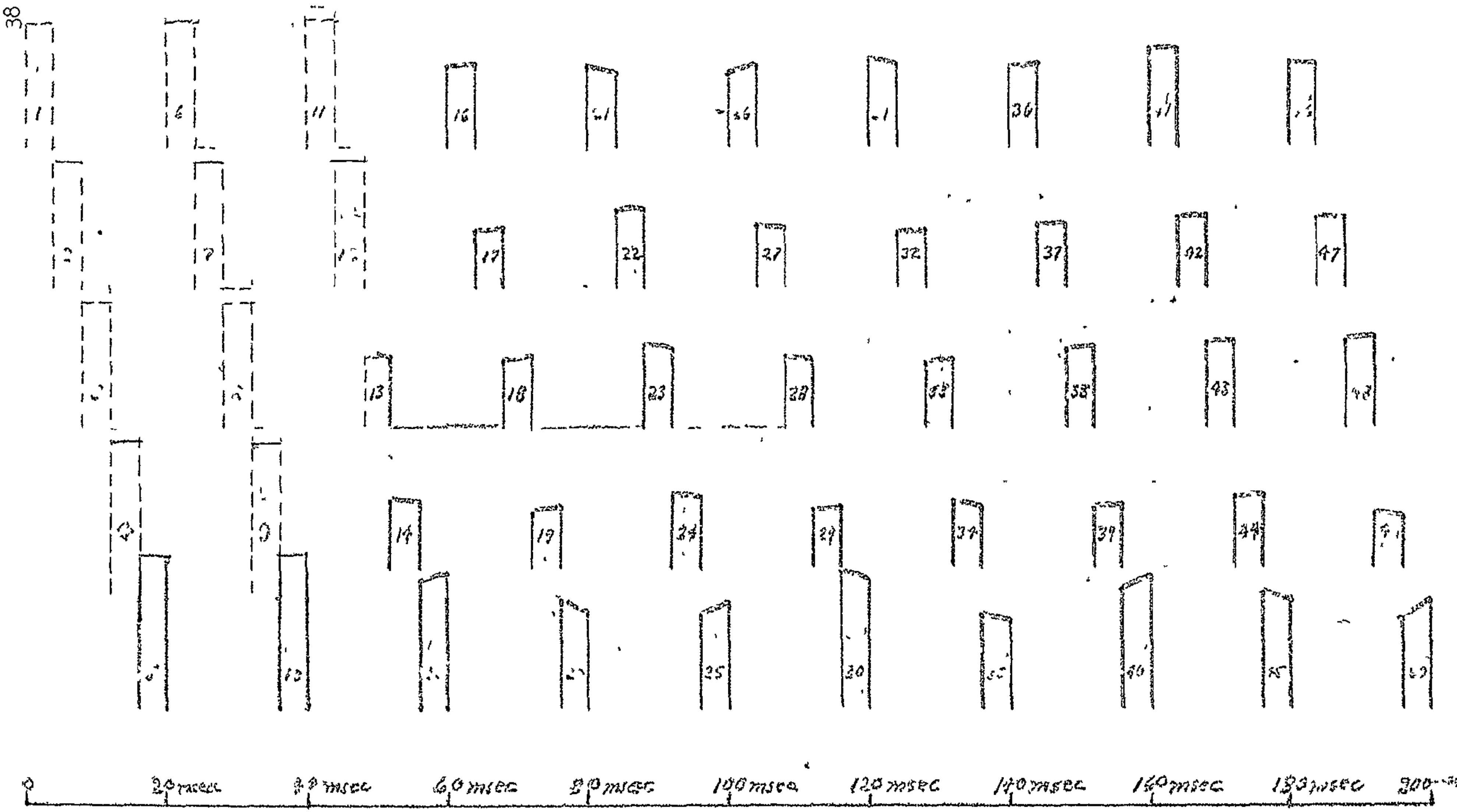
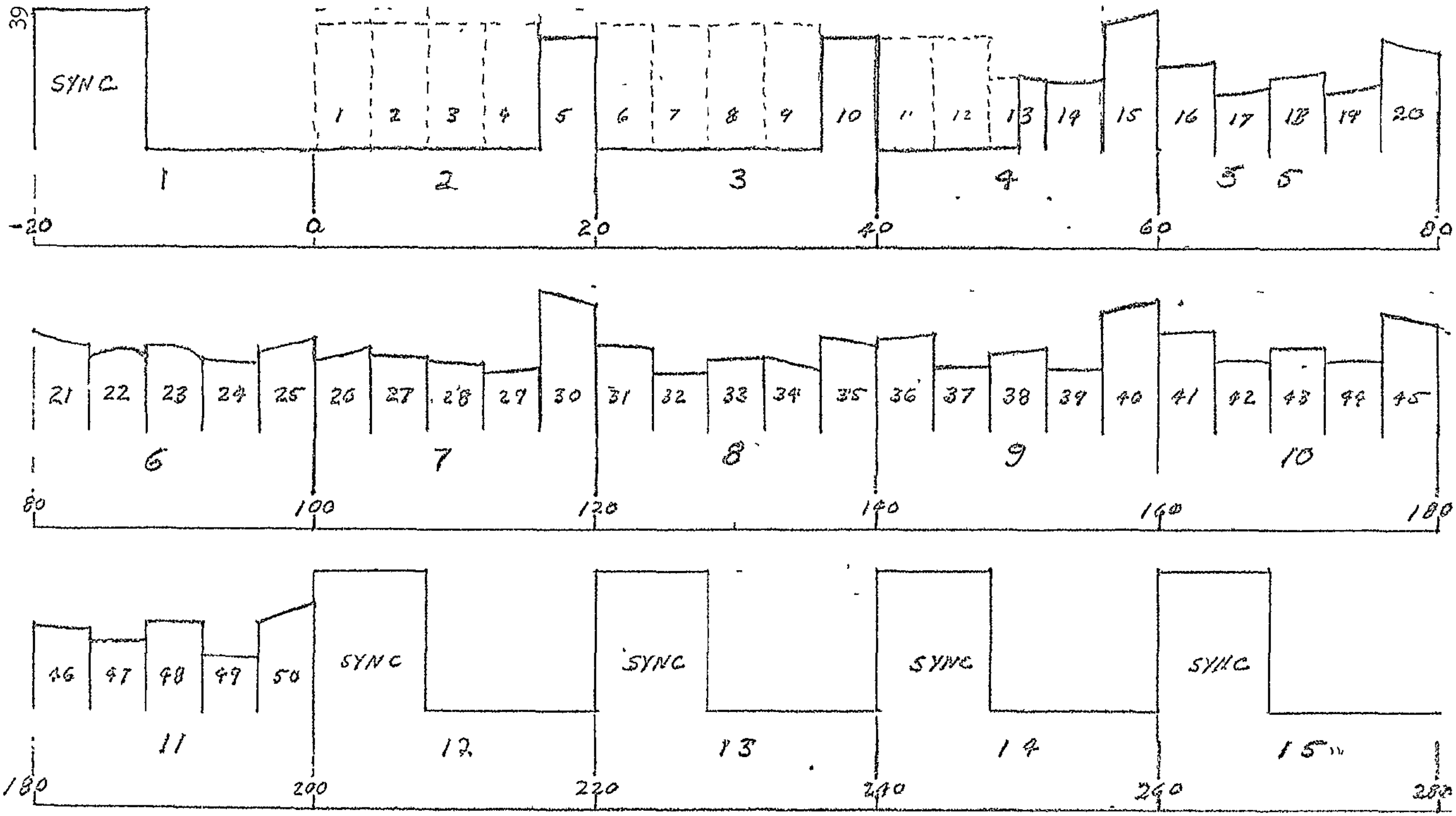


Fig 27



SAMPLED ANALYZER OUTPUT FOR WORD SHOE

FIG 29



MULTIPLEXER OUTPUT BEFORE FILTERING

Fig 29

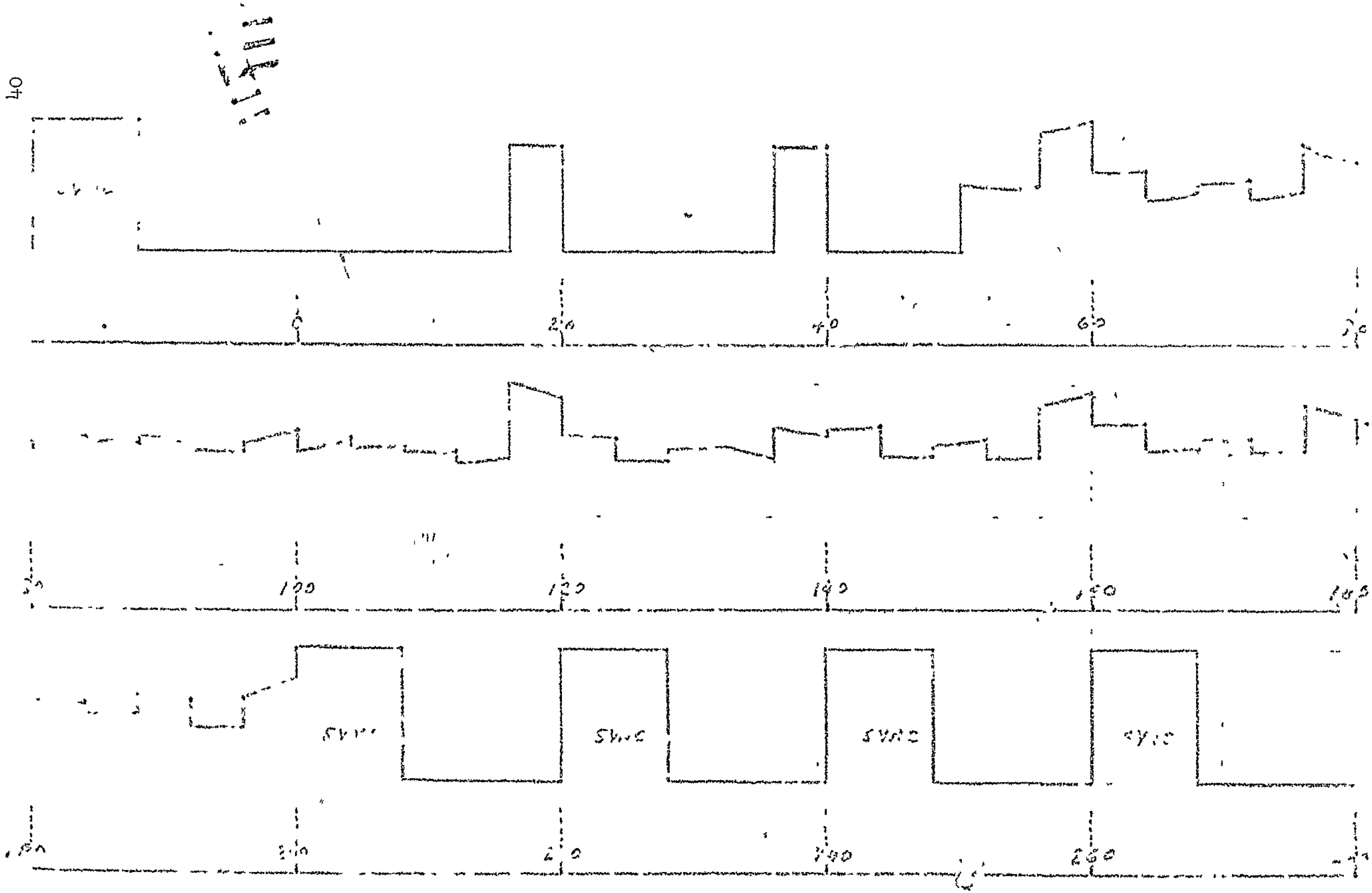
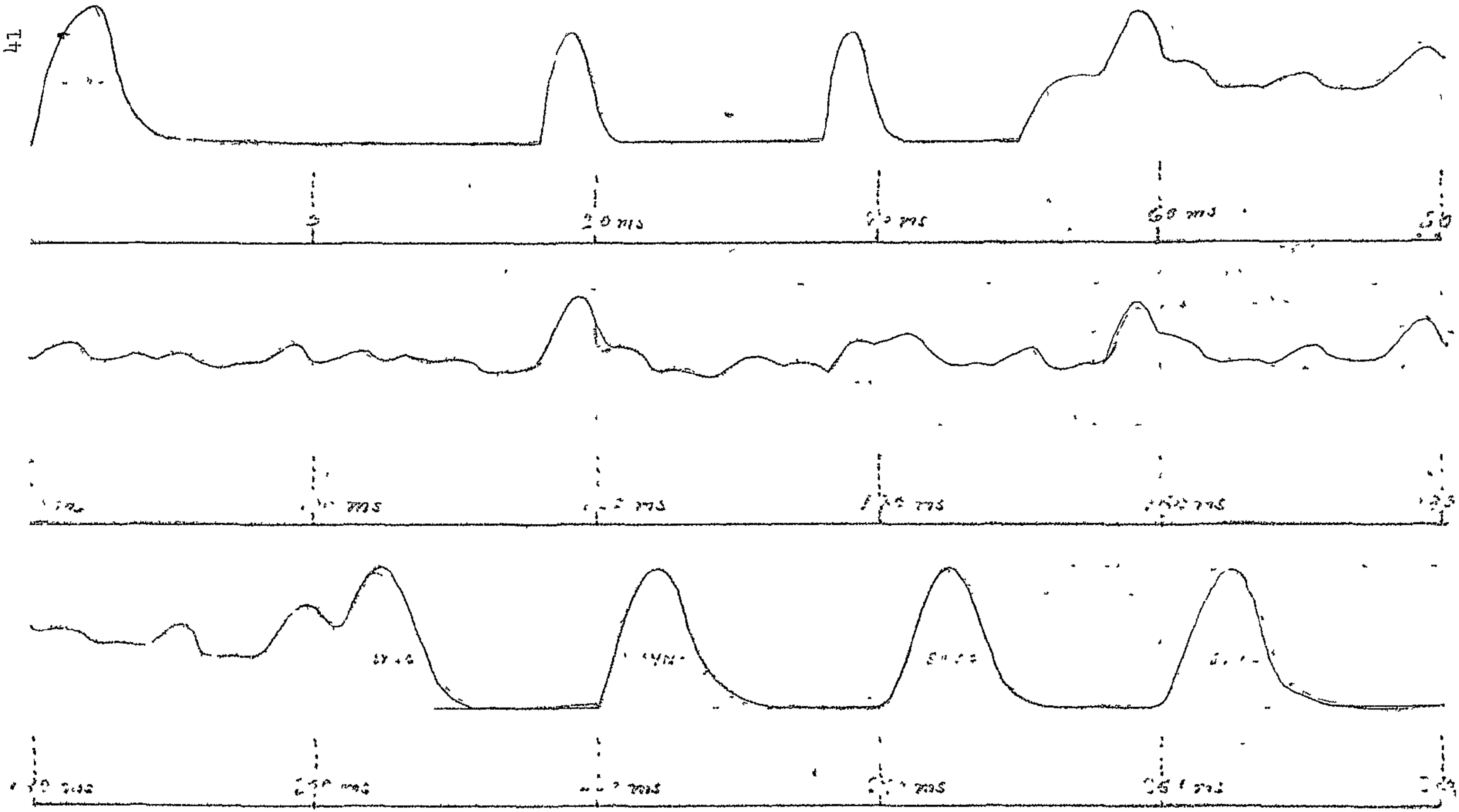
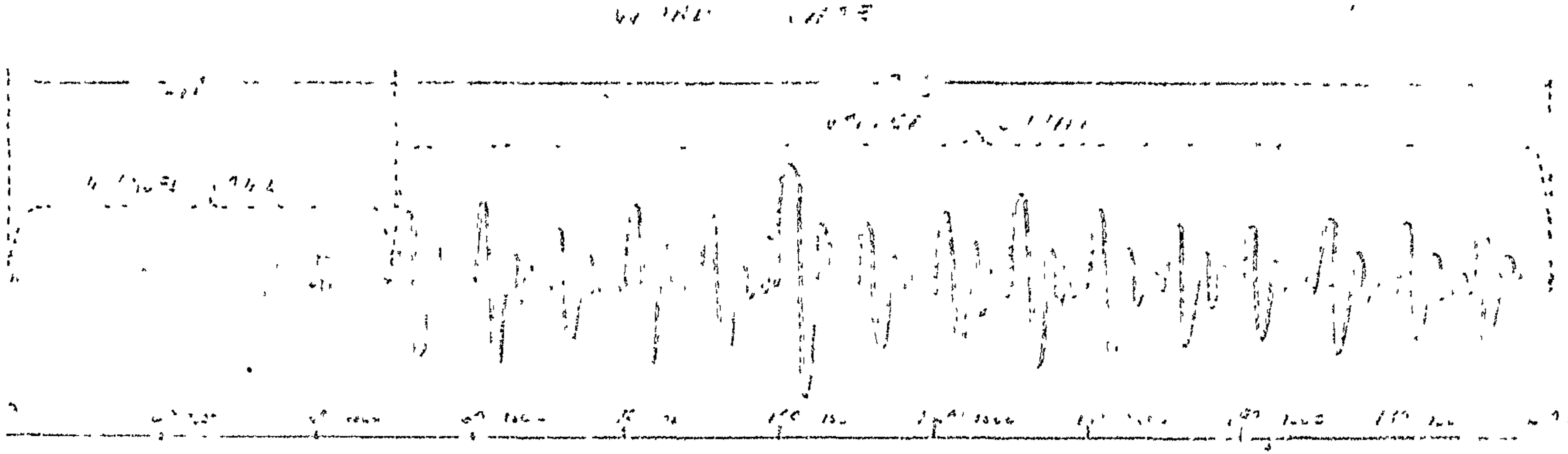


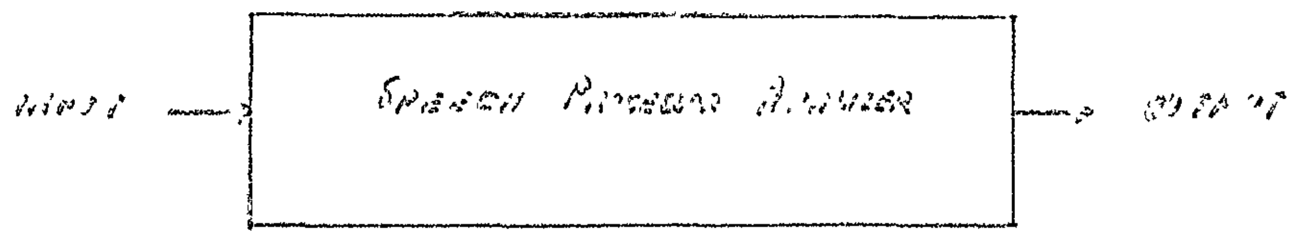
FIG 30

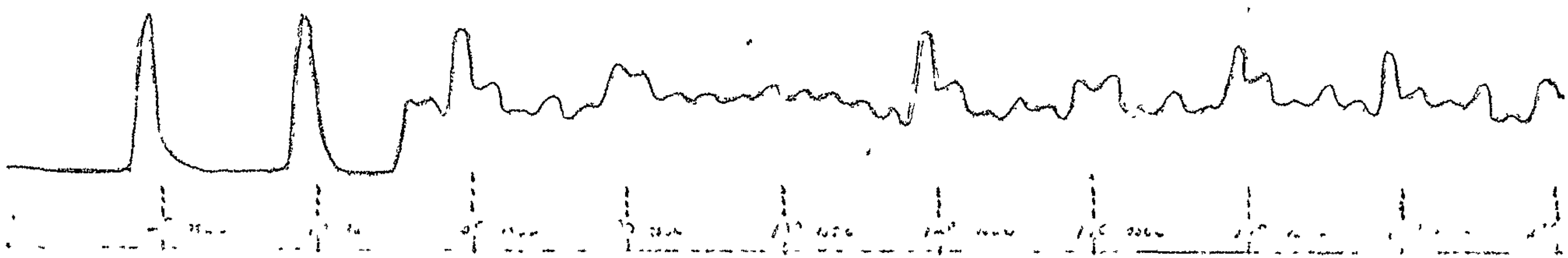


MULTIPLEXED SIGNAL FILTERED - 160 Hz LOW PASS

FIG 31







B. SYNTHESIZER ANALYSIS

The synthesizer or reconstruction processor is somewhat the reverse of the analyzer operation The compressed signal is fed first to the demultiplexer via a 160 Hz low pass filter. The demultiplexed output is passed through a bank of 20 Hz low pass filters, thus, recovering the five speech parameters. A block diagram of the complete synthesis process is shown in Figure 33. The demultiplexing operation is shown in Figure 34 through Figure 38. Figure 34 shows a block diagram of the demultiplexer. Figure 35 shows the multiplexed analyzer output as the input to the demultiplexer. The demultiplexer output is shown in Figure 36 This output is fed through the bank of 20 Hz low pass filters shown in Figure 37. The outputs of these filters, shown in Figure 38, are the recovered speech parameters. By comparing Figure 38 to Figure 25, a degree of degradation to the speech parameters due to the multiplexing process can be realized However, it must be kept in mind that the idea is somewhat over simplified.

The speech parameters of Figure 38 are the ones used to reconstruct a replica of the original speech input. As noted in Figure 38, there are five individual parameters (1) F_1, (2) A_1, (3) F_2, (4) A_2 and (5) P/V. explanation and series of diagram Figures 29 through 49 to follow, will show how a speech replica is formed from these five parameters.

Considering parameter five first, and as illustrated in Figure 39, the Pitch/Unvoiced (P/V) decision is fed to a threshold detector and an inverting amplifier of unity gain The output of the threshold detector, as shown in Figure 40, is the unvoiced decision and the output of the inverting amplifier also shown in Figure 40, is the pitch. It is obvious why this is true.

Both of these outputs are fed to a pitch generator and noise source The output of the pitch generator is shown in Figure 41. As illustrated, this output contains random noise and periodic pulses. The random noise represents unvoiced sound and the periodic pulses represent speech pitch. Both of these signals in their series occurrence are fed to two switch modulators along with first formant amplitude and second formant amplitude, respectively. As a result of this operation, first and second formant reconstruction begins.

Formant reconstruction is illustrated in the diagrams of Figures 42 through 47. The first formant is shown in Figure42 through 44, and the second formant in Figure 45 through 47. Figure 42 represents a block diagram of the first formant reconstruction process. The first formant amplitude parameter, the unvoiced decision and pitch are fed to the inputs of switch modulator number one, as shown in Figure 43. The output of switch modulator one, also shown in Figure 43, is a composite of unvoiced sound and amplitude modulated pitch pulses. This output is then used to drive the first formant bandpass active voltage tuned filter (AVTF) and tracking low pass AVTF along with the first formant period parameter.

The output of these two AVTF's, as shown in Figure 44, is unvoiced sound plus the first formant The theory of reconstructing this formant lies in the fact that the modulated pulses from the switch modulator one cause the bandpass AVTF to ring and produce periodic damped sinusoids At the same time, the ringing frequency is made to vary as a function of the first formant period parameter F_1. Thus, as a result of these operations, the first formant is reconstructed

The second formant is reconstructed in the same manner as the first and is illustrated in Figure 45 through Figure 47. However, after being reconstructed, it is combined with the first formant in a summing process.

This summing process is performed in a summing amplifier as shown in Figure 48. The summed output, shown in Figure 49, is the replica of the speech input. As illustrated, the signal contains noise representing unvoiced sound and a periodic wavetrain representing voiced sound also corresponding to the SH and OE in the word "shoe."

Figure 50 gives a summary illustration of the synthesis process. As shown, the input is the multiplexed output of the analyzer and the processor synthesizer output is a replica of the analyzer speech input.

Finally, Figure 51 is a general summary of the complete processor operation. The word "shoe" is provided as the analyzer input. It is then processed, made to vary at a 160 Hz rate and used as the synthesizer input. The synthesizer expands the compressed word and presents a replica of the input word shoe as its output. Thus, as a result of this, the processing operation is complete.

This concludes the picturial analysis of the narrow bandwidth speech processor and it is hoped that some understanding of the system operation has been achieved. The following section is a discussion of applications of processor.

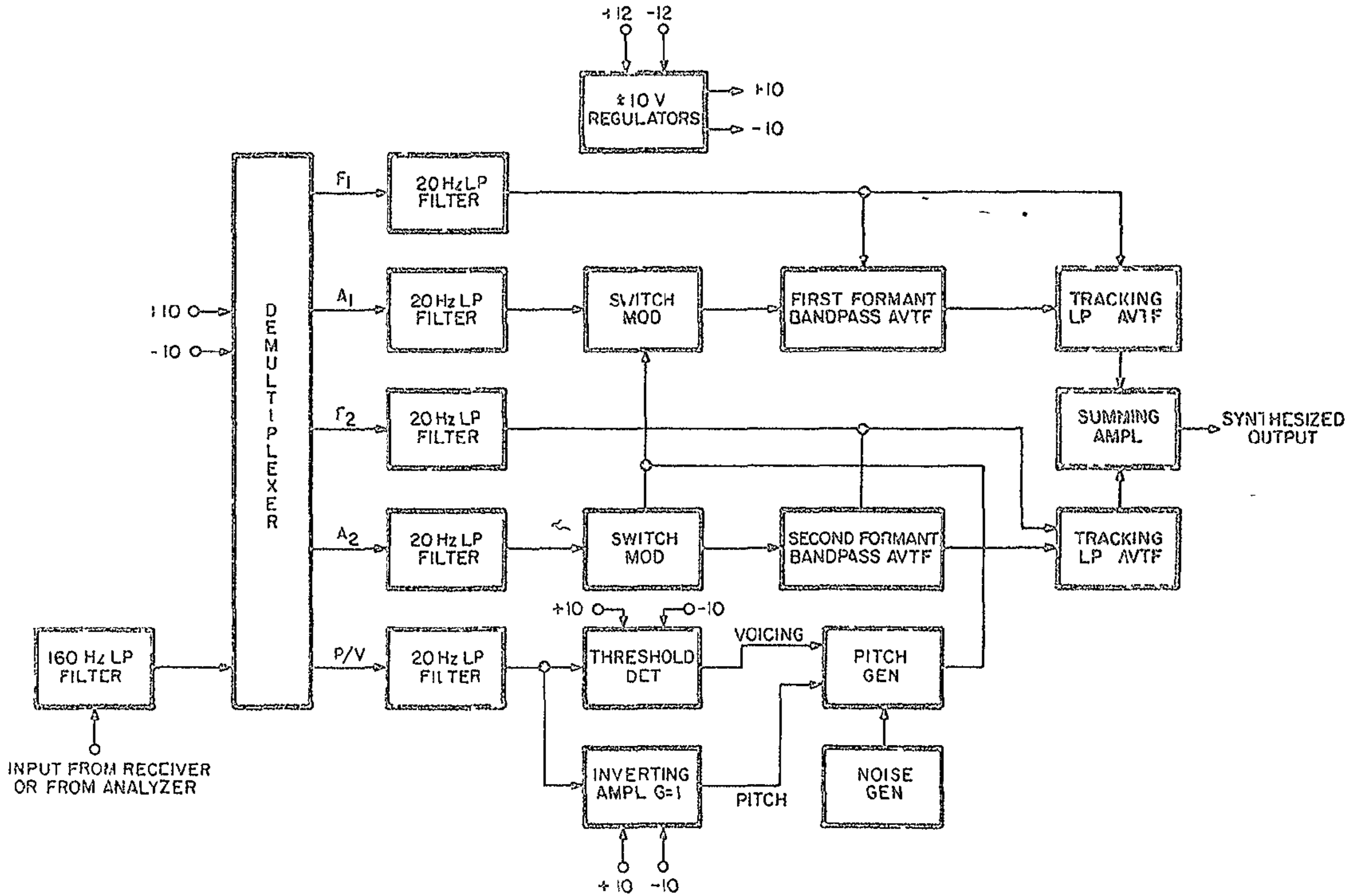


Fig 33 Synthesizer Block Diagram

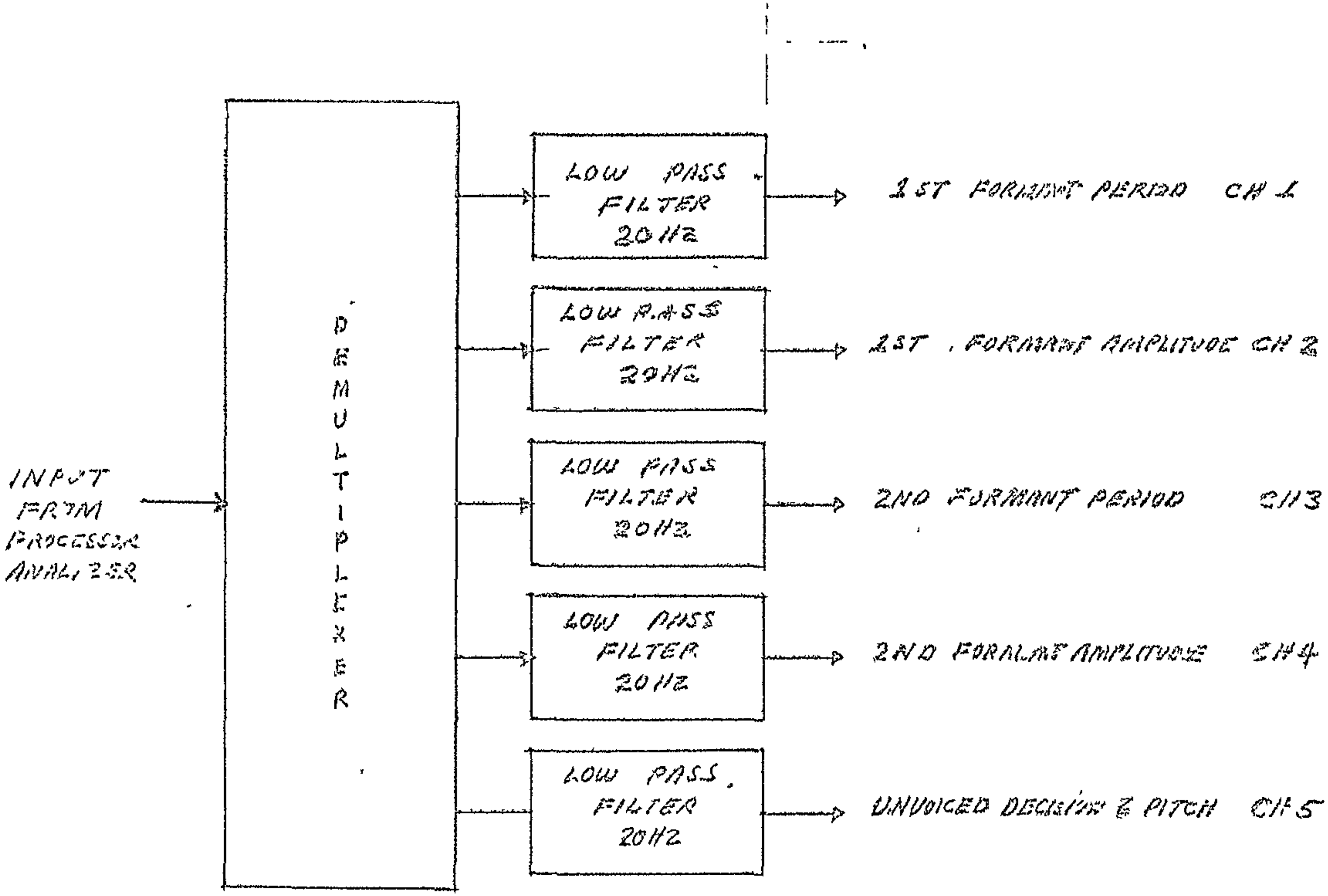
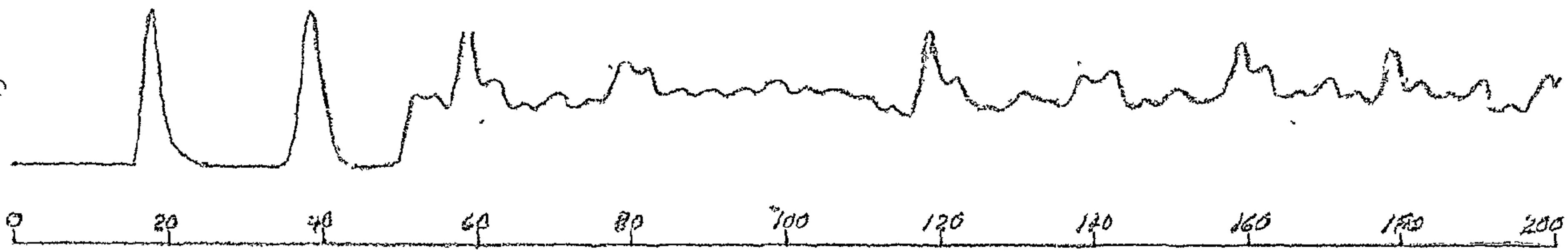
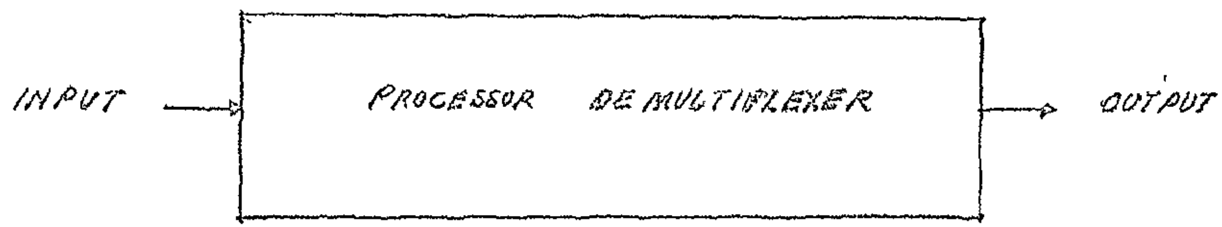


Fig 34 Block diagram of synthesizer
DE MULTIPLEXER

F

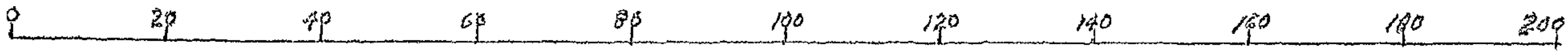


a. INPUT TO PROCESSOR DEMULTIPLEXER



b. PROCESSOR DEMULTIPLEXER

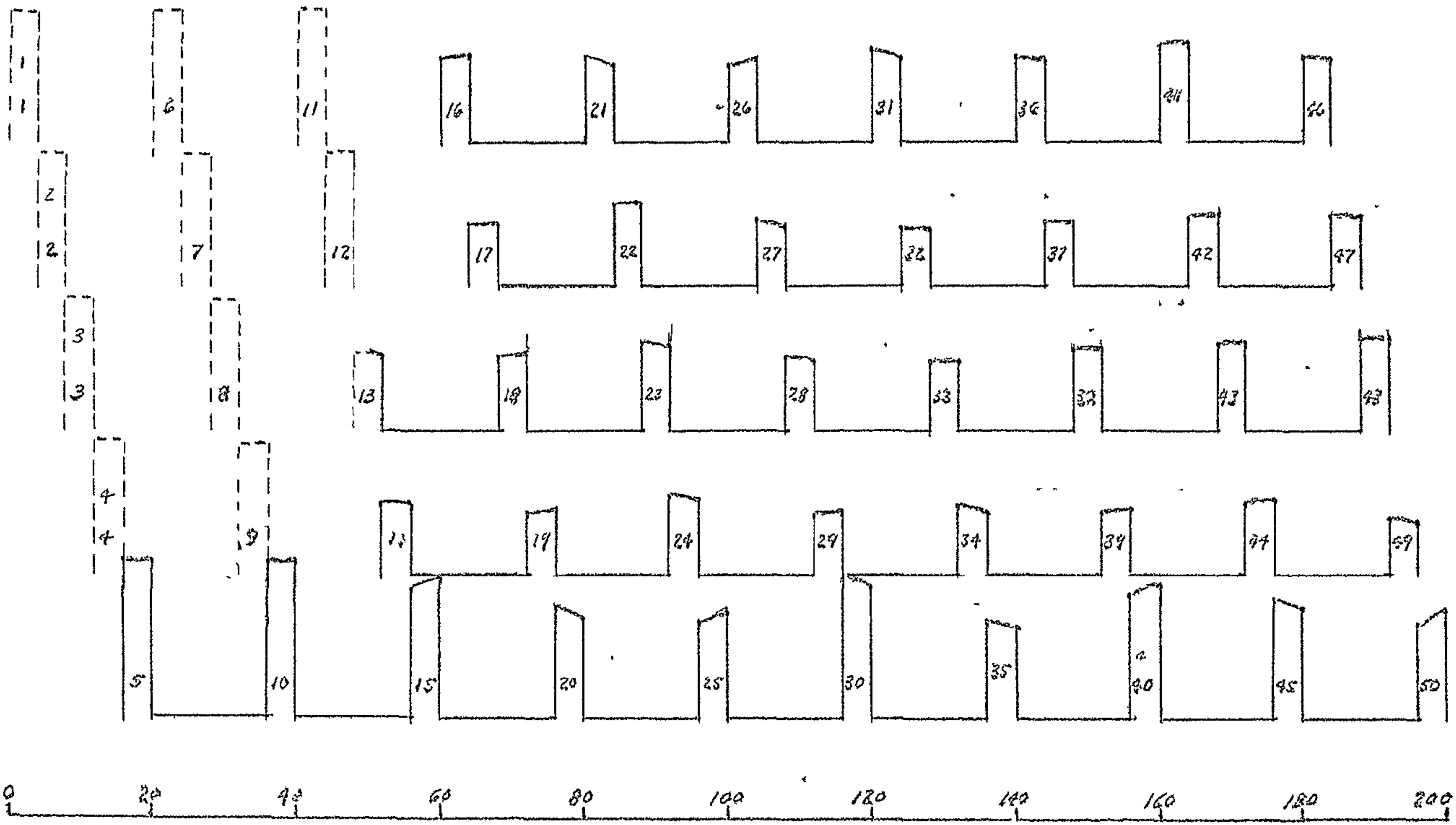
OUTPUT SHOWN IN FOLLOWING DIAGRAM



c. DEMULTIPLEXER OUTPUT

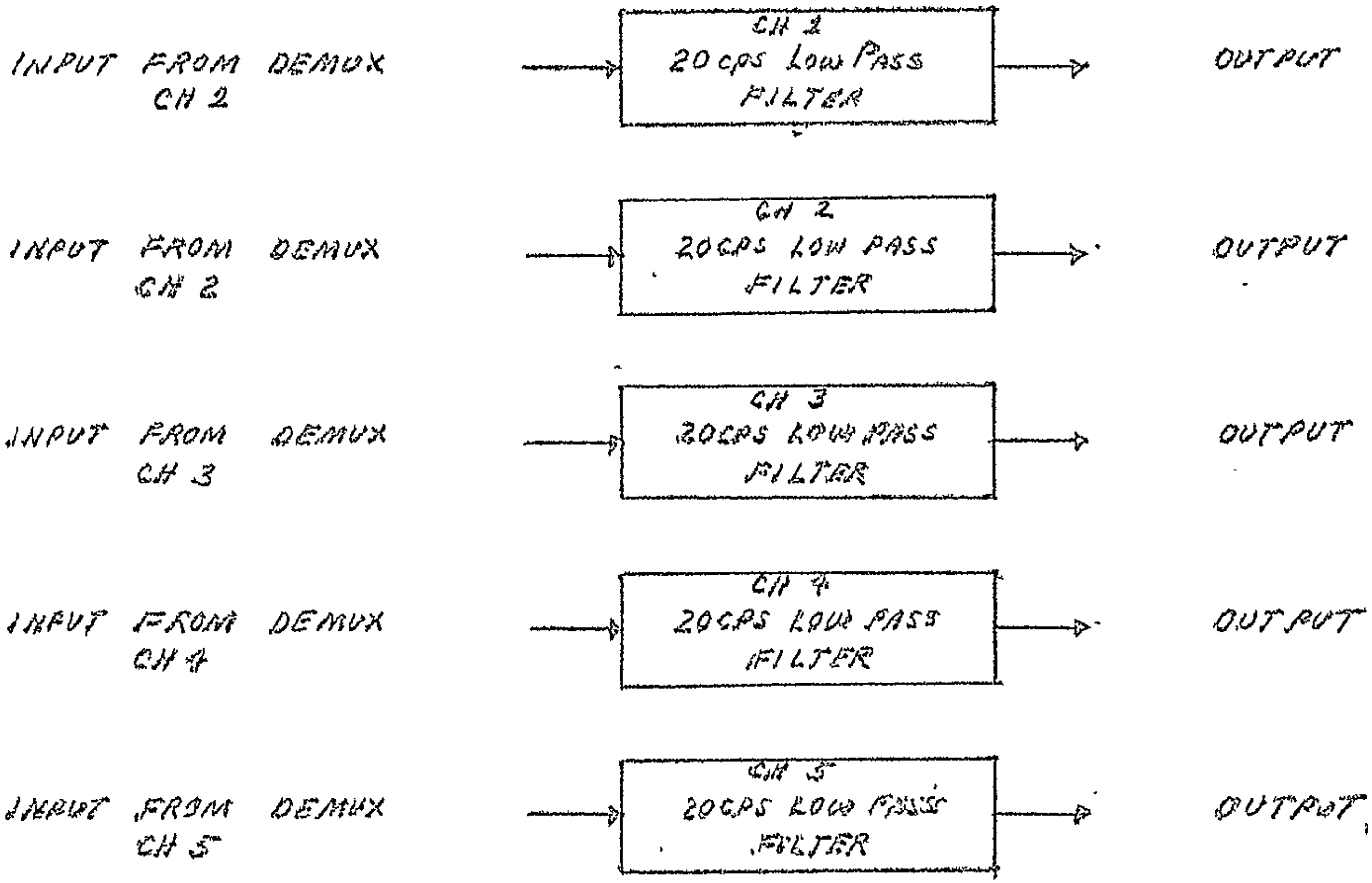
FIG. 35

50



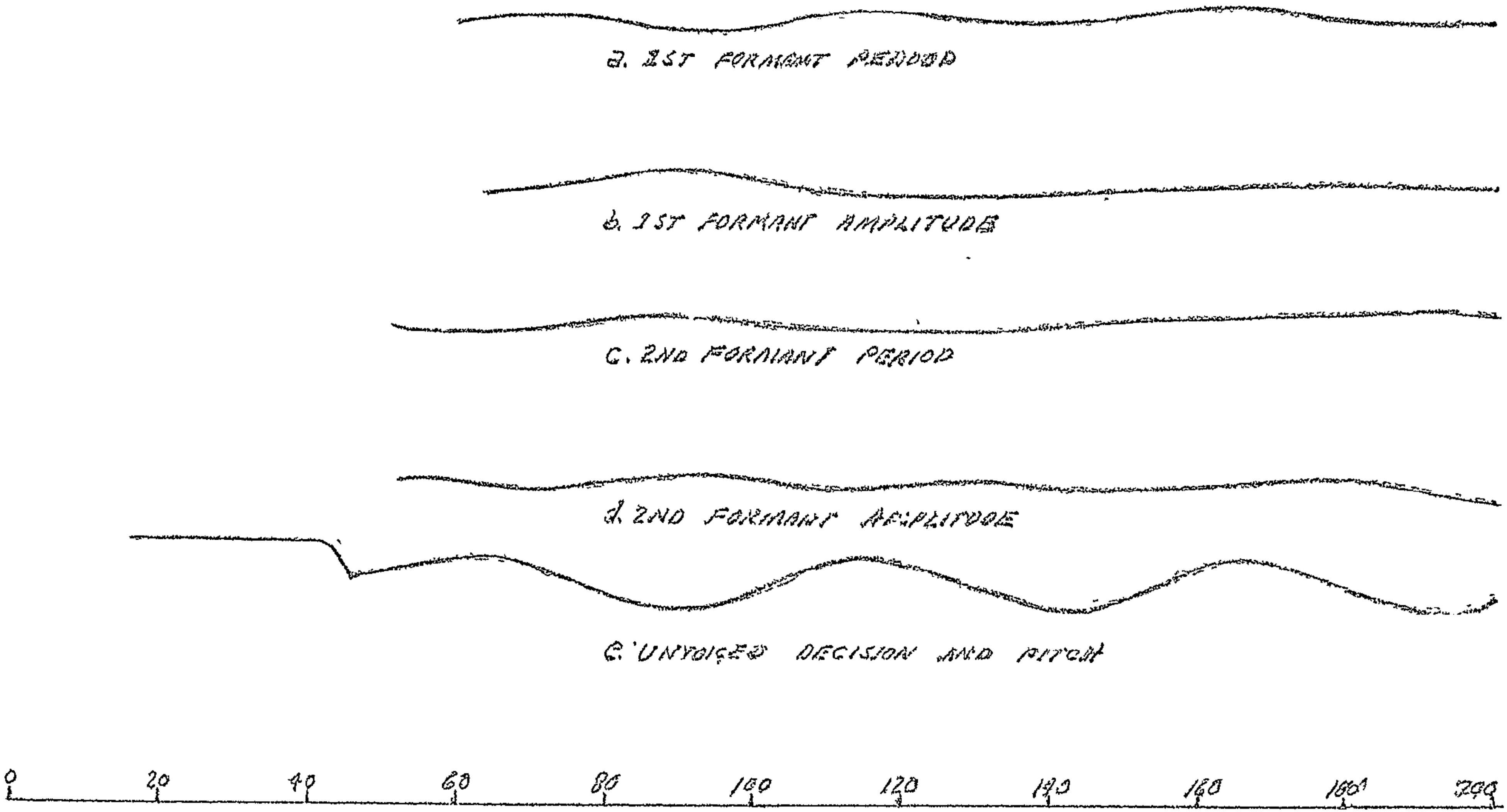
DE MULTIPLEXER OUTPUT CHANNELS 1 THROUGH 5

Fig. 36



DEMULTIPLEXER OUTPUT LOW PASS FILTERS
CHANNEL 2 THROUGH 5

Fig 37



OUTPUT OF DEMULTIPLEXER OUTPUT LOW PASS FILTERS CHANNELS 1 THROUGH 5

Fig 38

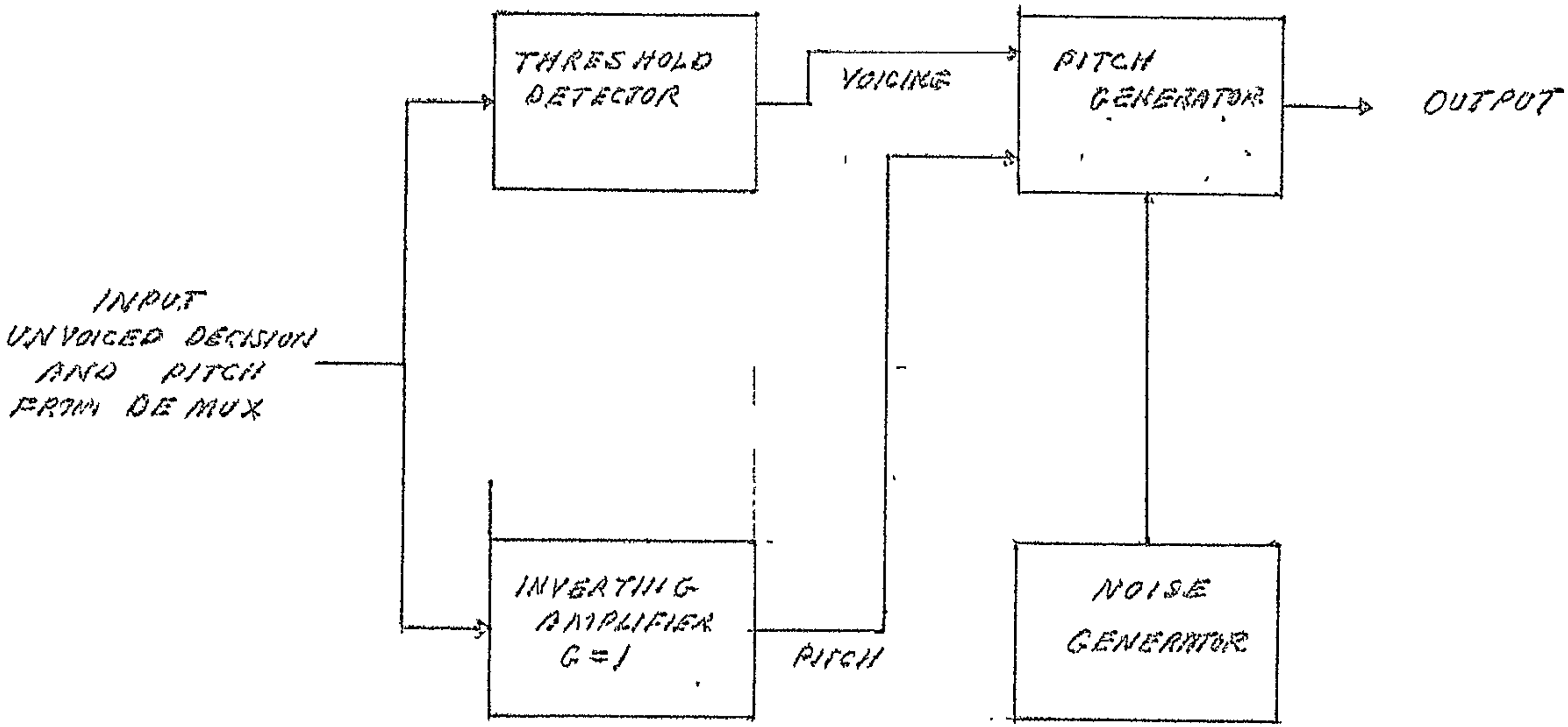
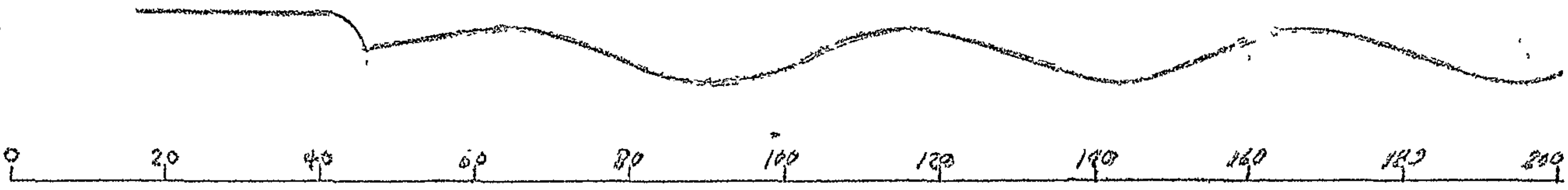
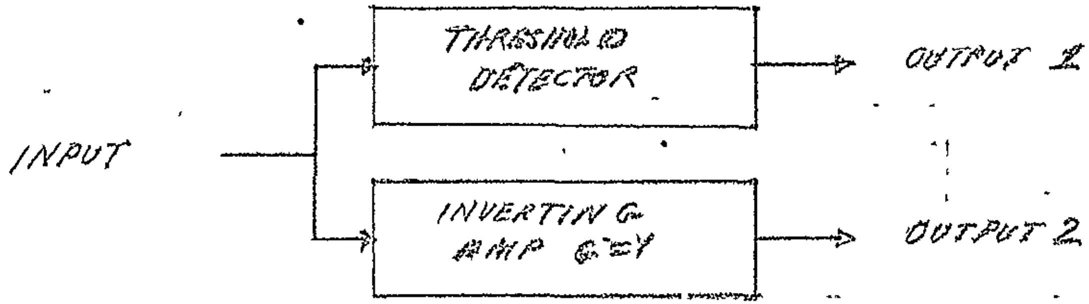


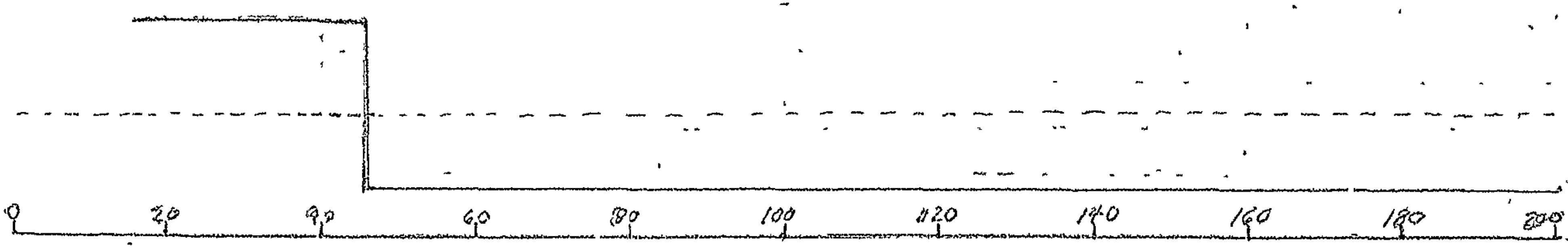
FIG 39 BLOCK DIAGRAM OF SYNTHESIZER UNVOICED SOUND AND PITCH GENERATOR



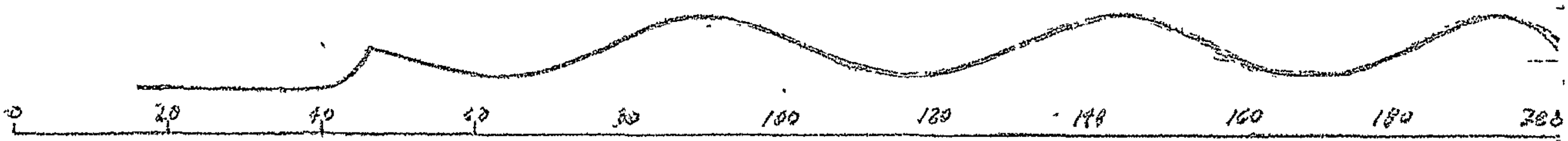
a UNVOICED DECISION AND PITCH



b SEPARATION OF UNVOICED DECISION AND PITCH



c UNVOICED DECISION (OUTPUT 1)



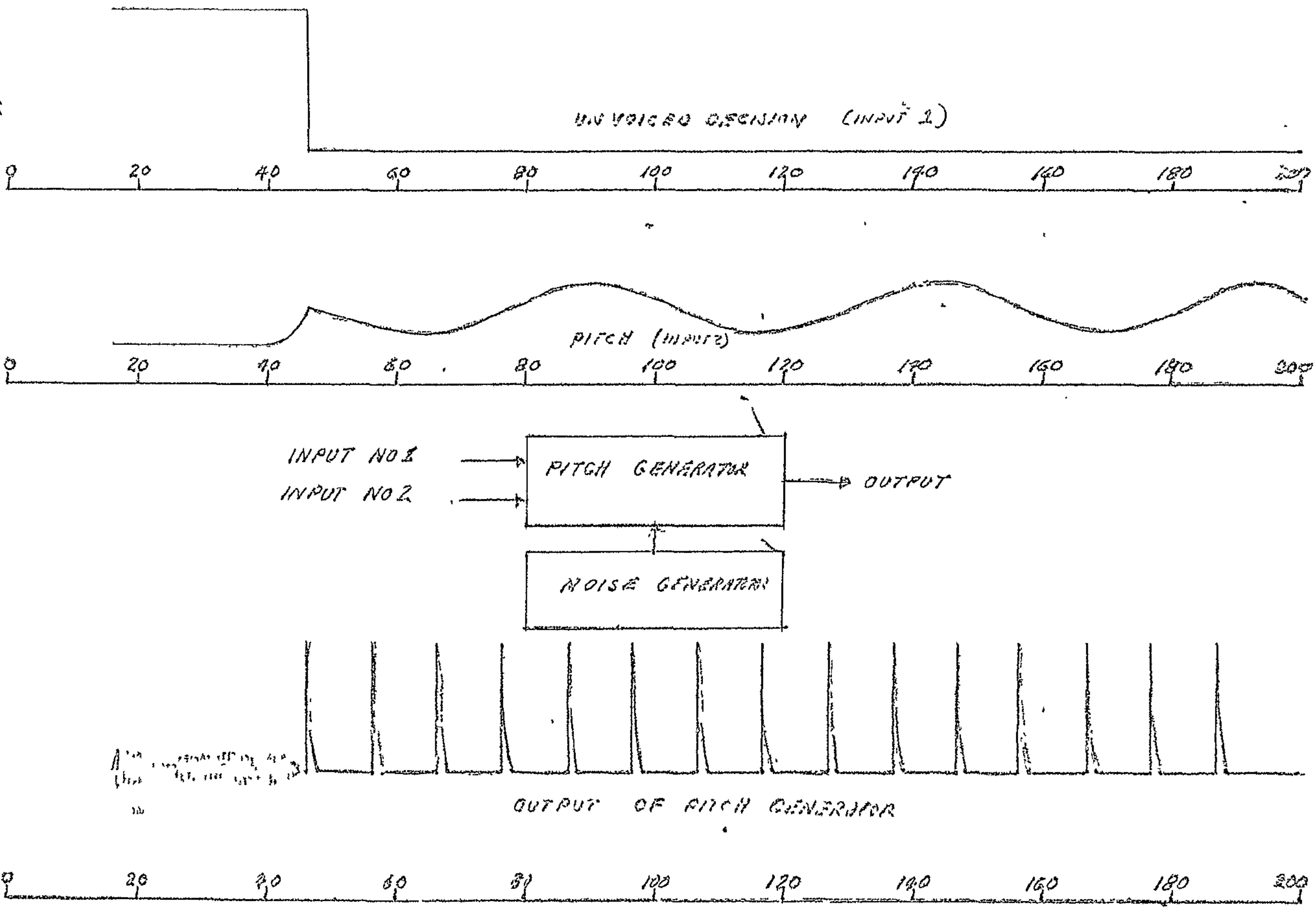

UNVOICED DECISION (INPUT 1)
PITCH (INPUT 2)
0 20 40 60 80 100 120 140 160 180 200
INPUT NO 1
INPUT NO 2
PITCH GENERATOR
OUTPUT
NOISE GENERATOR
OUTPUT OF PITCH GENERATOR

FIG 41

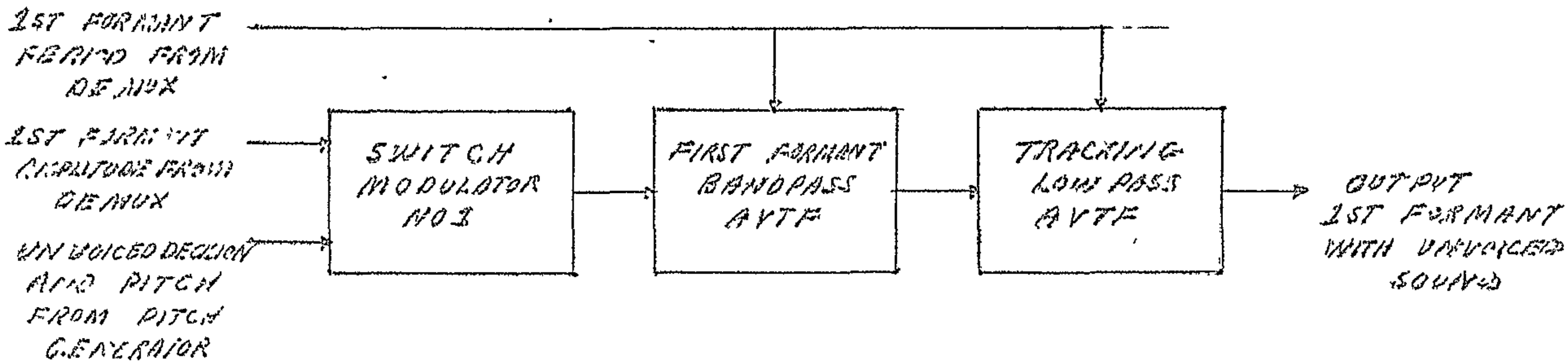


FIG 42 BLOCK DIAGRAM OF SYNTHESIZER 1ST FORMANT RECONSTRUCTION WITH UN VOICED SOUND

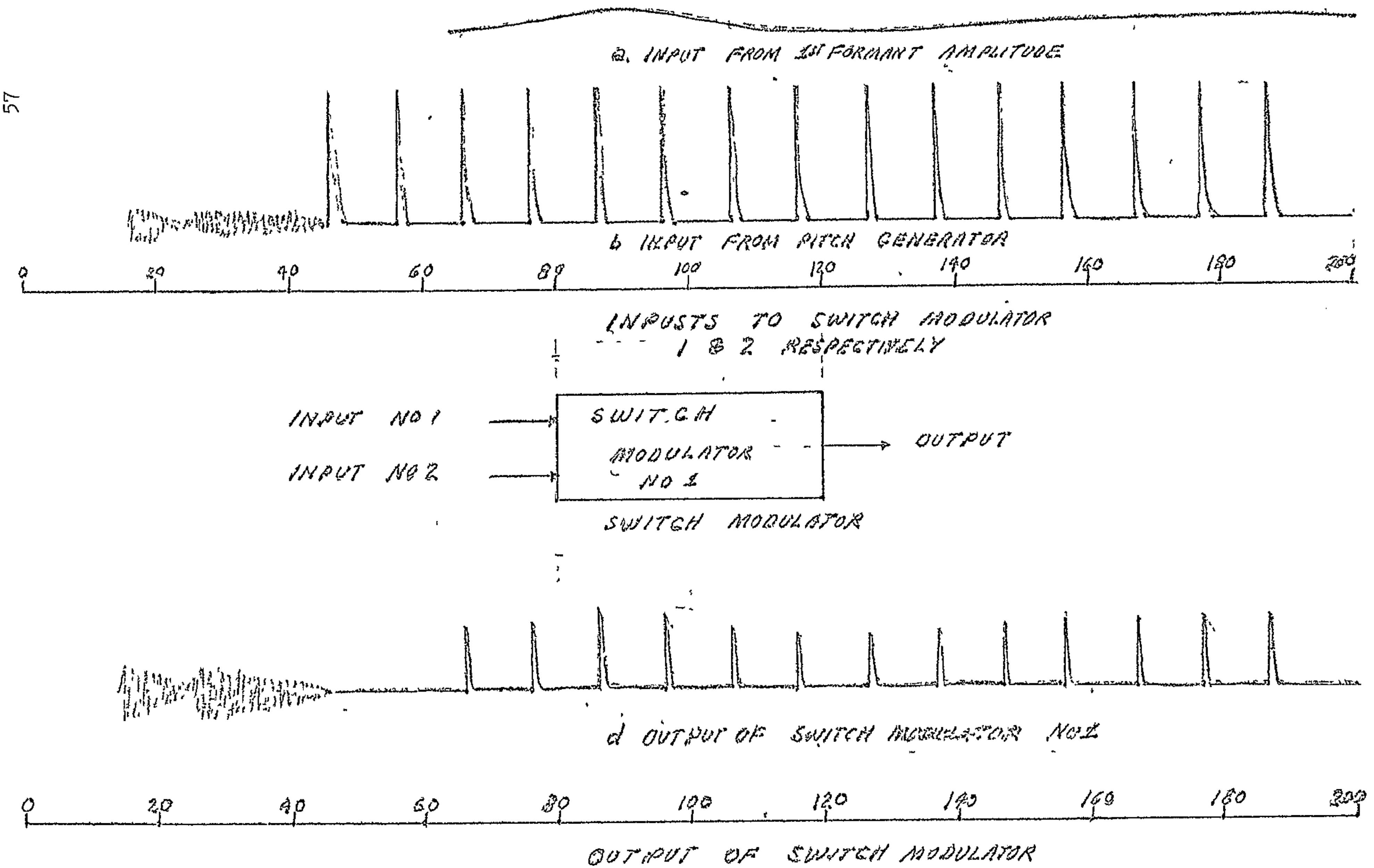

a. INPUT FROM 1st FORMANT AMPLITUDE
b INPUT FROM PITCH GENERATOR
0 20 40 60 80 100 120 140 160 180 200
INPUSTS TO SWITCH MODULATOR
1 & 2 RESPECTIVELY
INPUT NO 1
INPUT NO 2
SWITCH MODULATOR NO 1
OUTPUT
SWITCH MODULATOR
d OUTPUT OF SWITCH MODULATOR NO 1
0 20 40 60 80 100 120 140 160 180 200
OUTPUT OF SWITCH MODULATOR

FIG 43

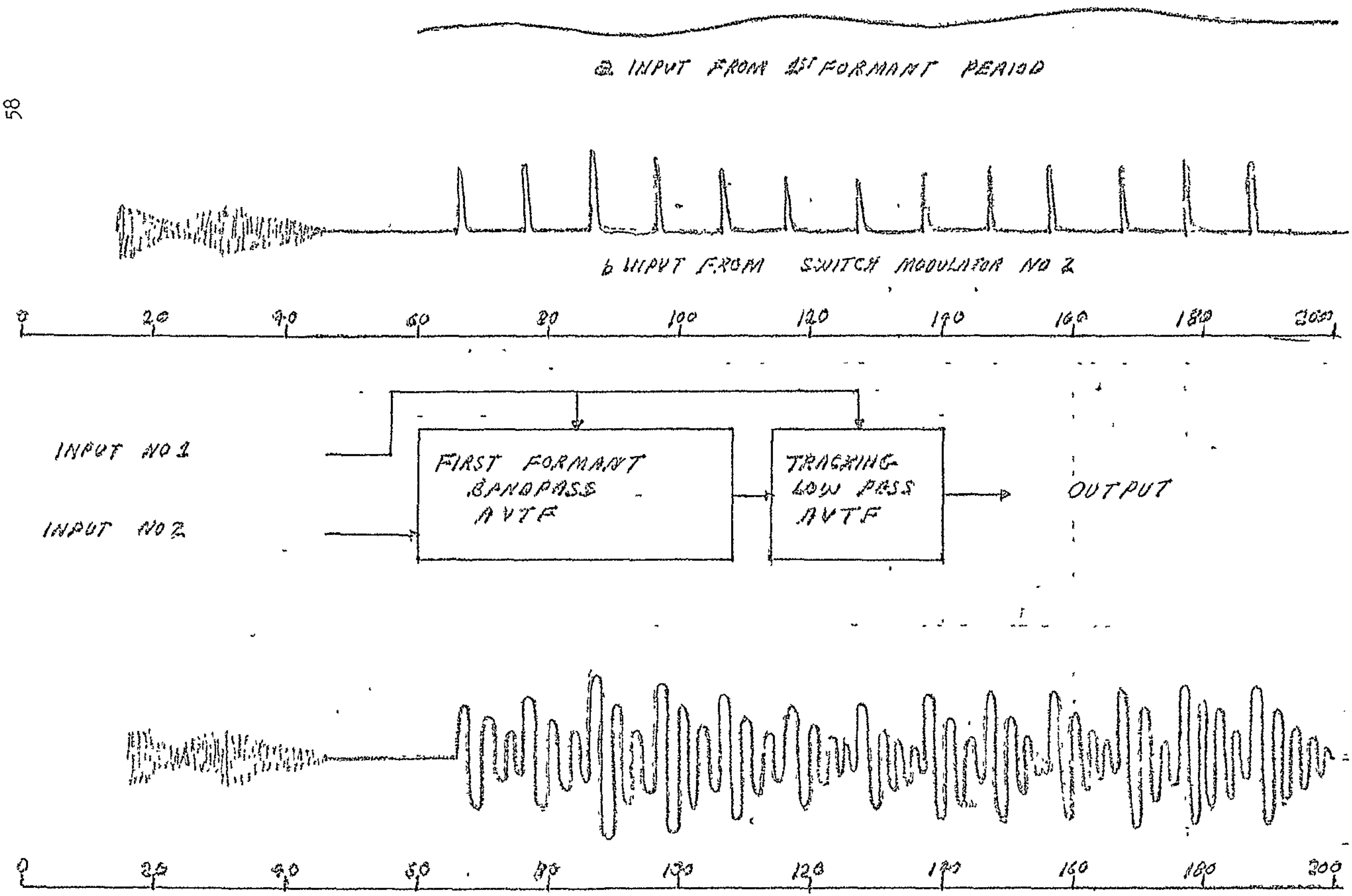

a INPUT FROM 1st FORMANT PERIOD
b INPUT FROM SWITCH MODULATOR NO 2
0 20 40 60 80 100 120 140 160 180 200
INPUT NO 1
INPUT NO 2
FIRST FORMANT BANDPASS AVTF
TRACKING LOW PASS AVTF
OUTPUT
0 20 40 60 80 100 120 140 160 180 200

FIG 44

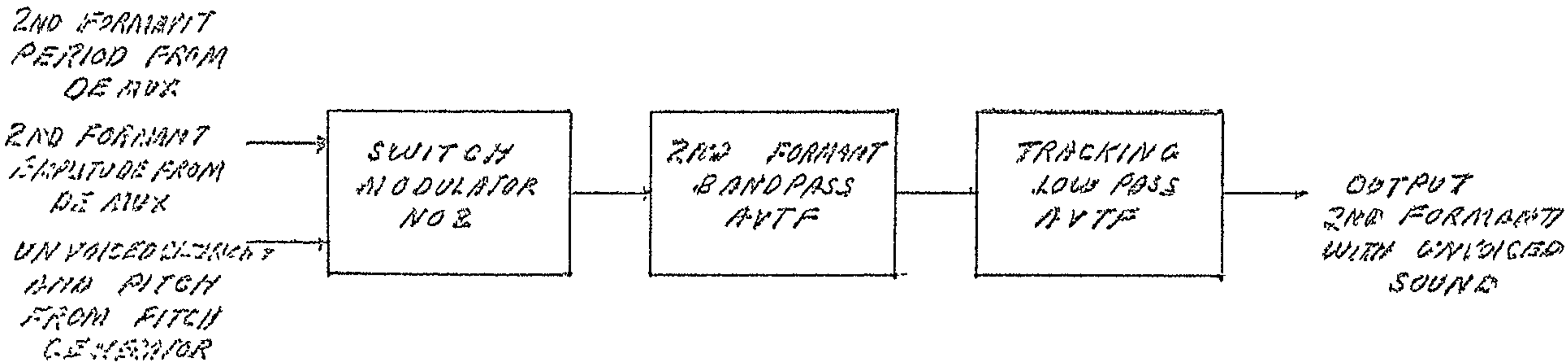


FIG 45 BLOCK DIAGRAM OF SYNTHESIZER
2ND FORMANT RECONSTRUCTION
WITH UNVOICED SOUND

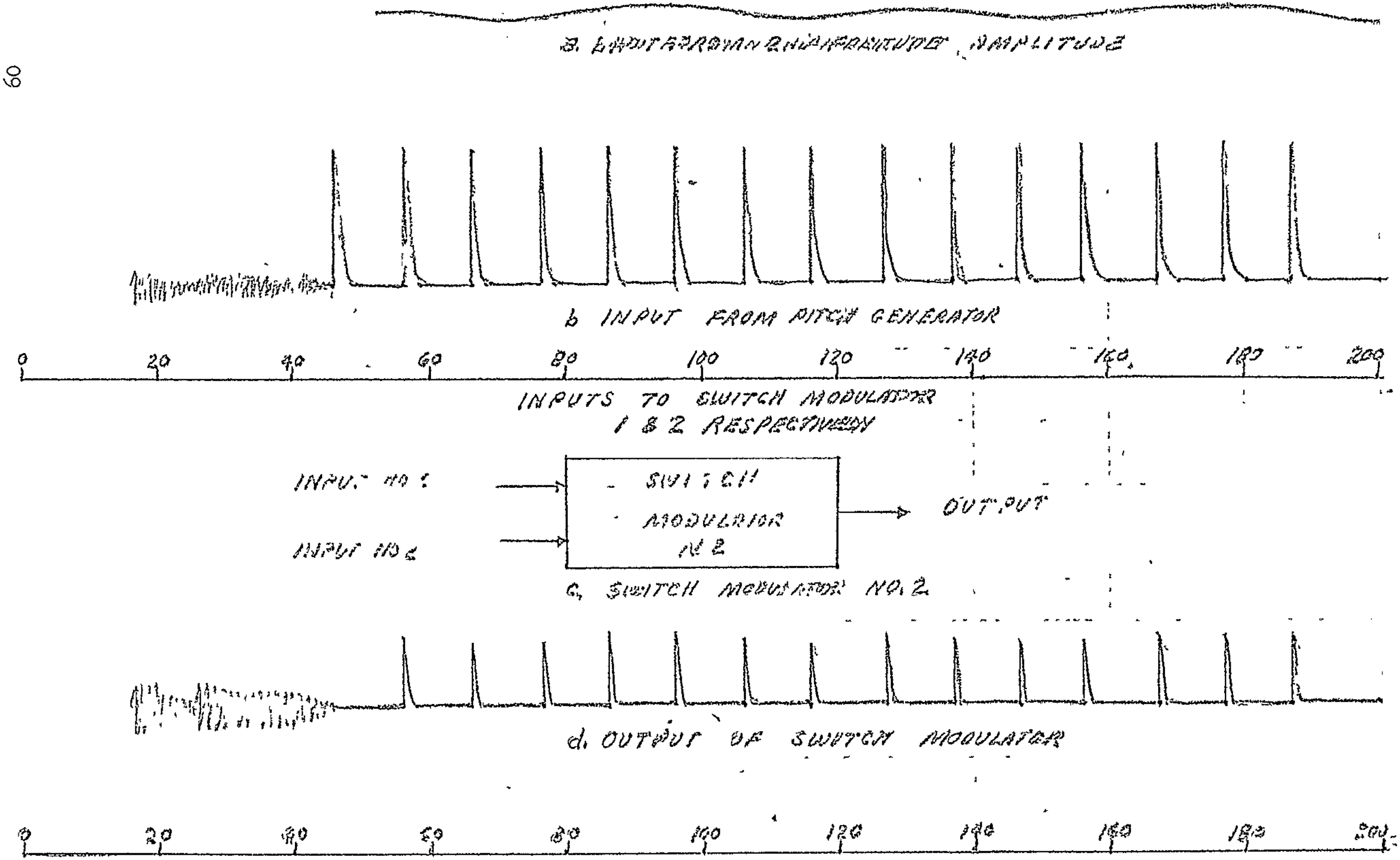

AMPLITUDE
b INPUT FROM PITCH GENERATOR
0 20 40 60 80 100 120 140 160 180 200
INPUTS TO SWITCH MODULATOR
1 & 2 RESPECTIVELY
INPUT NO 1
INPUT NO 2
SWITCH
MODULATOR
NO 2
OUTPUT
c. SWITCH MODULATOR NO. 2
d. OUTPUT OF SWITCH MODULATOR
0 20 40 60 80 100 120 140 160 180 200

FIG. 46

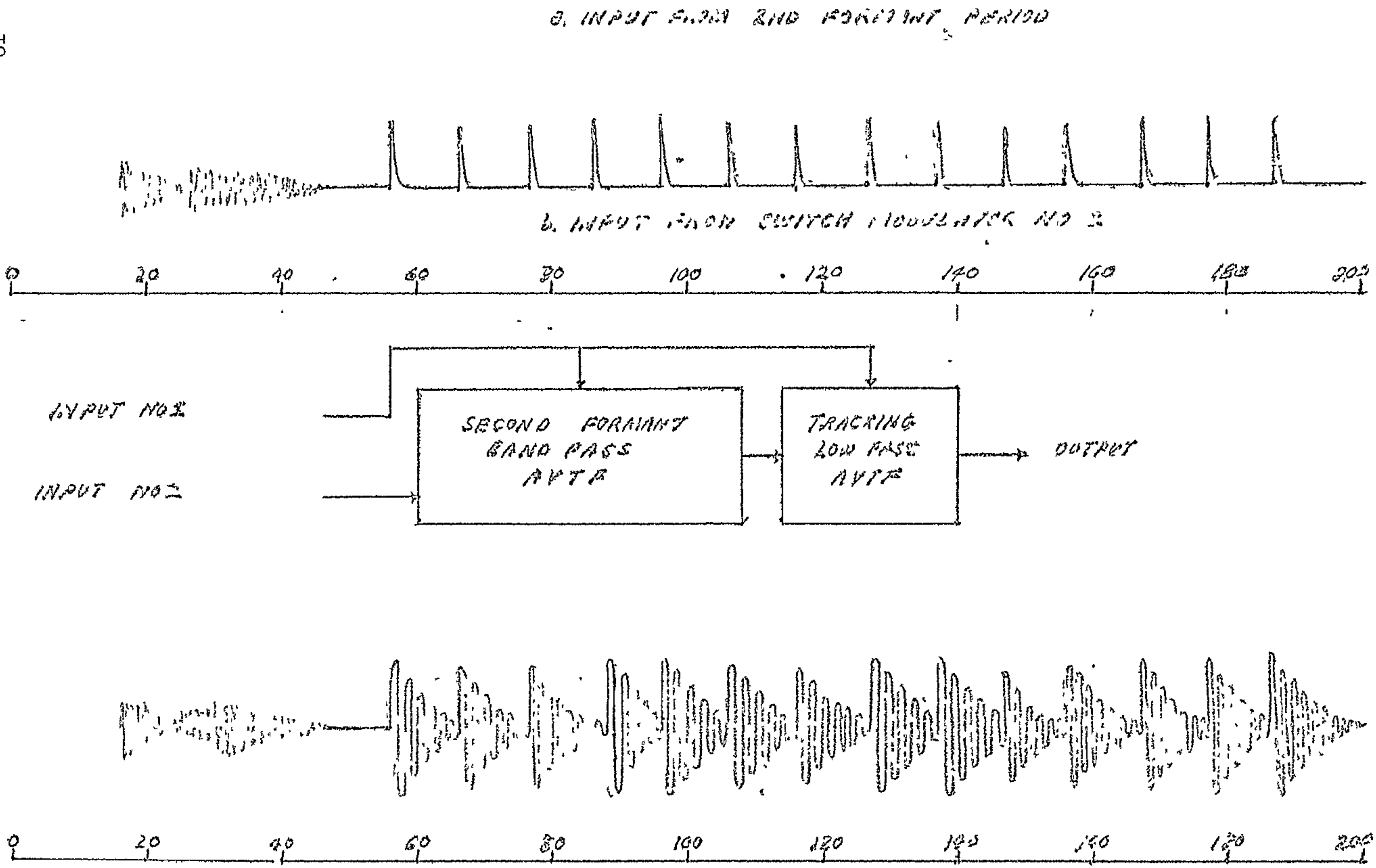

INPUT NO 1
INPUT NO 2
SECOND FORMANT
BAND PASS
AVTF
TRACKING
LOW PASS
AVTF
OUTPUT
0
20
40
60
80
100
120
140
160
180
200

FIG 47

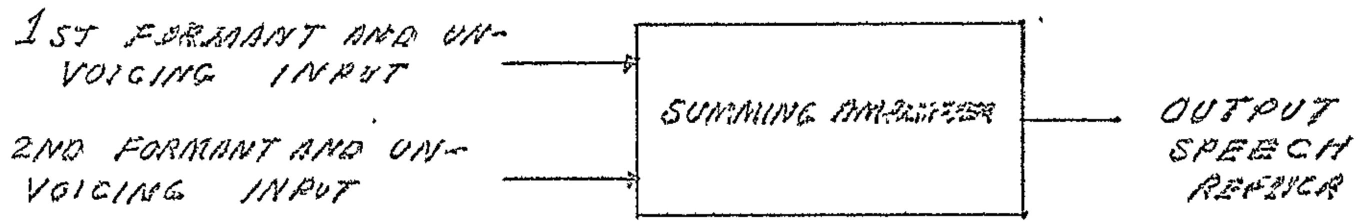
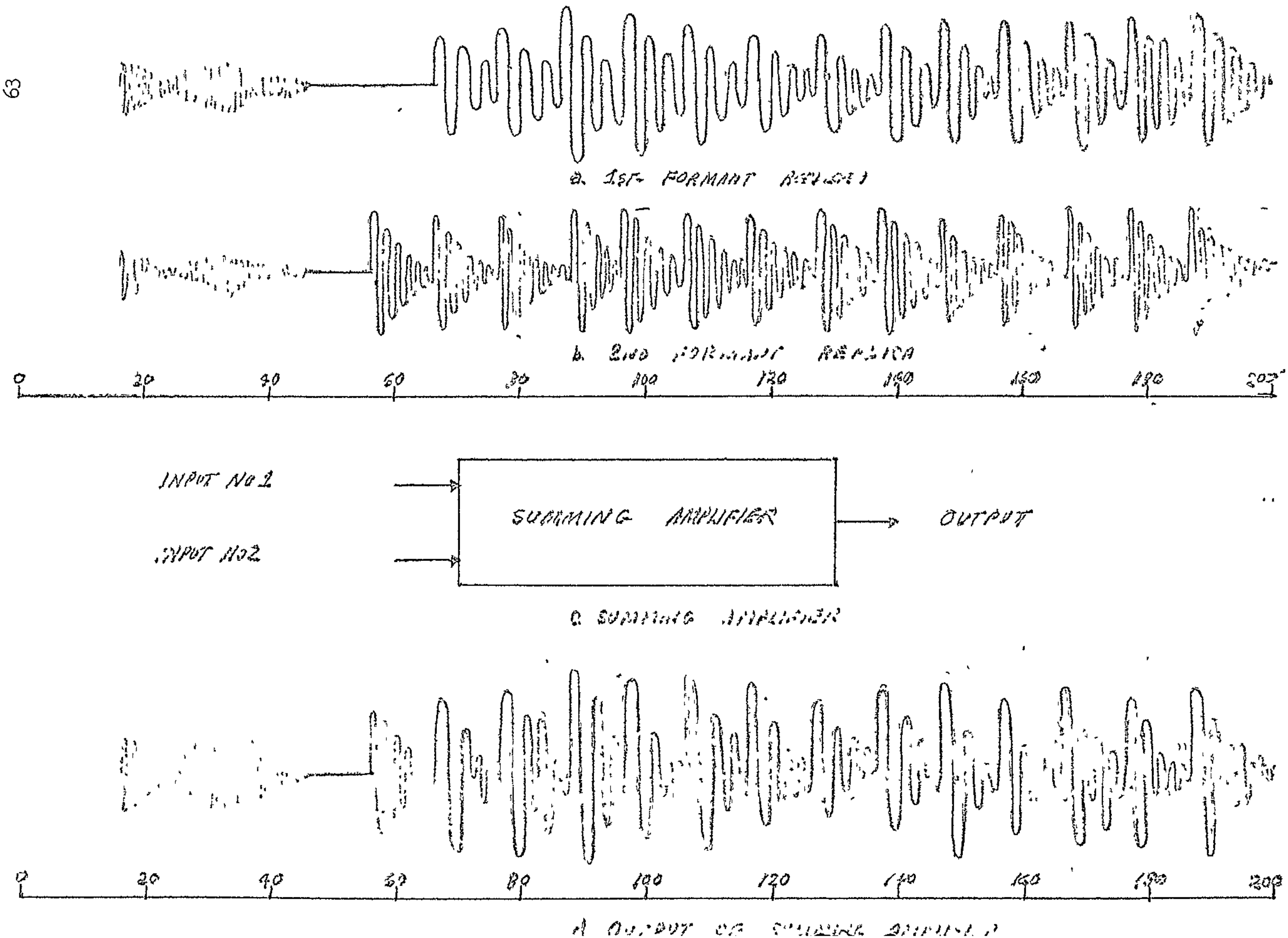


FIG BLOCK DIAGRAM OF SYNTHESIZER
SUMMING AMPLIFIER

FIG 48


a. 1ST FORMANT
b. 2ND FORMANT REPLICA
0 20 40 60 80 100 120 140 160 180 200
INPUT No 1
INPUT No 2
SUMMING AMPLIFIER
OUTPUT
c. SUMMING AMPLIFIER
0 20 40 60 80 100 120 140 160 180 200
d. OUTPUT OF SUMMING AMPLIFIER

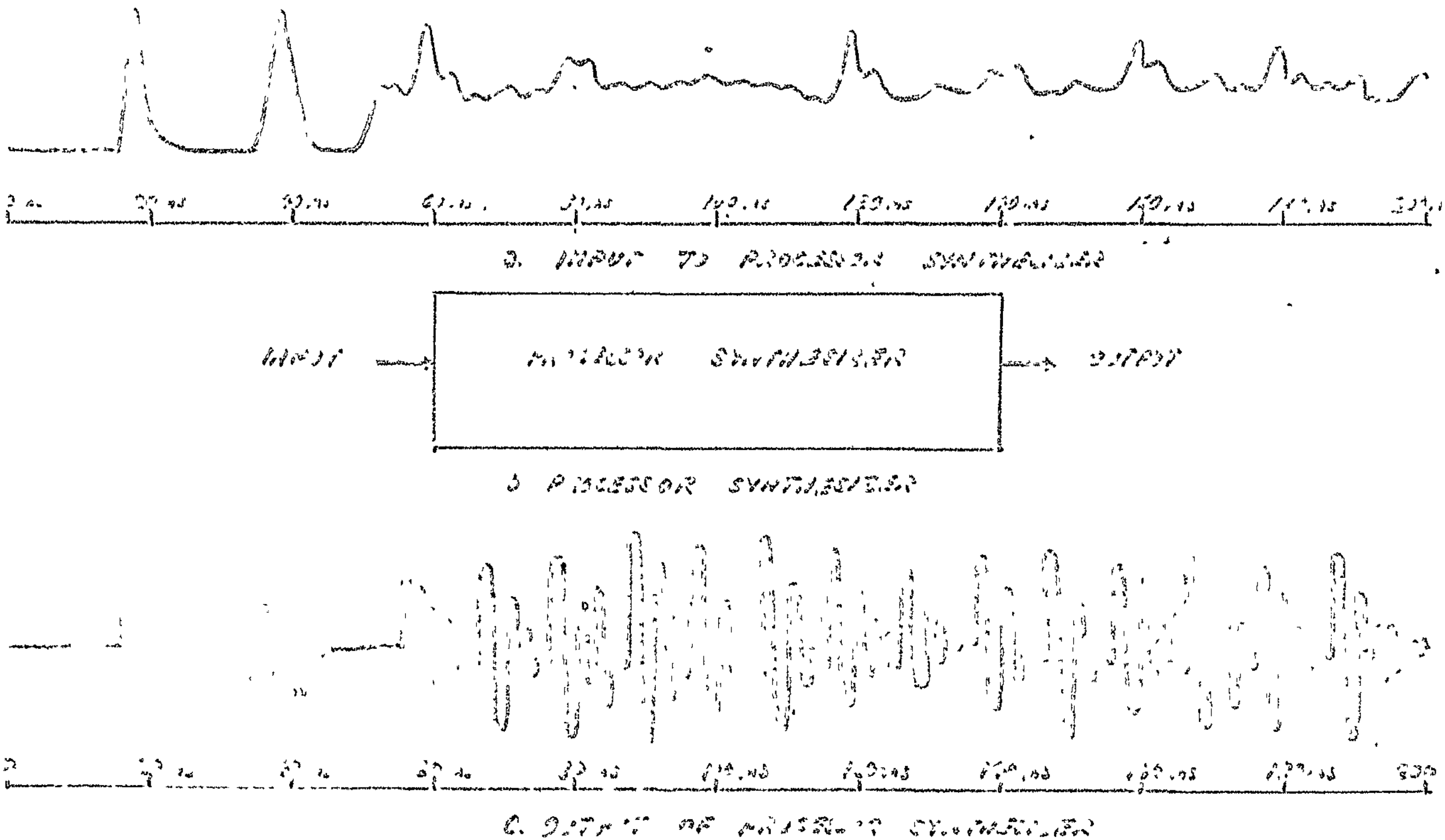
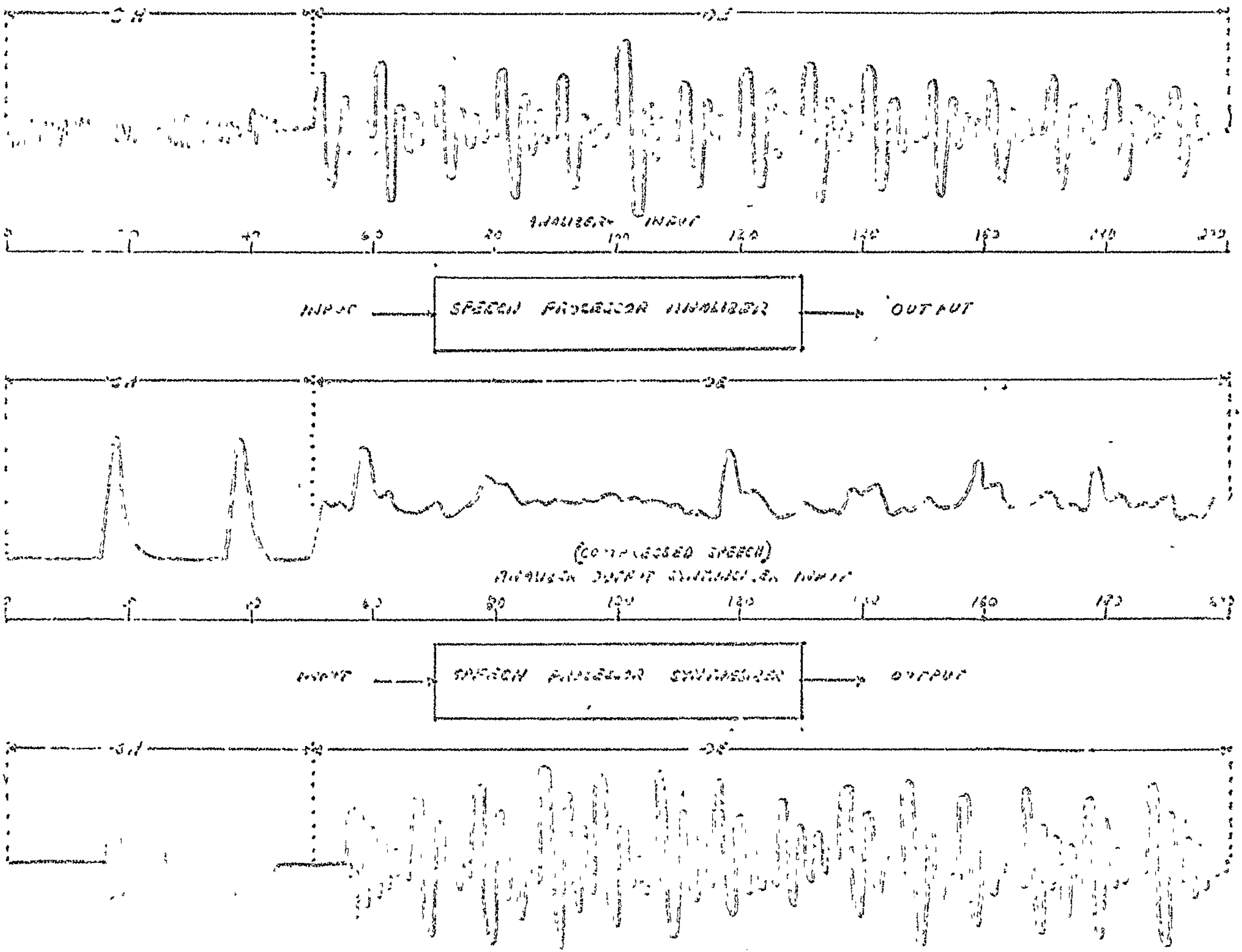


FIG 50


INPUT
SPEECH
OUTPUT
OUTPUT

IV APPLICATIONS

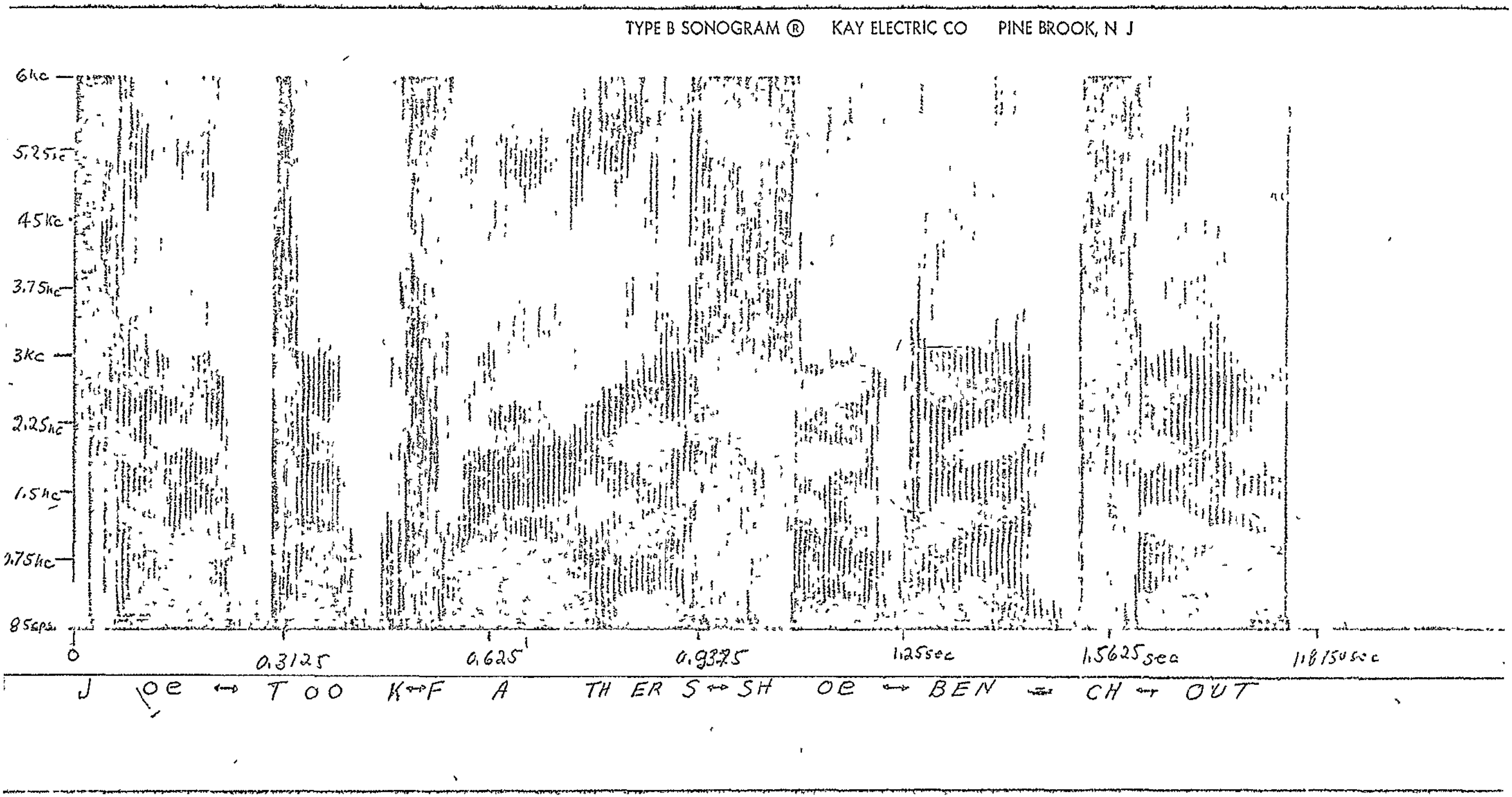
Studies in speech processing were initiated to find a means of lessening power requirements for voice communications from spacecraft to earth. In the course of performing these studies, the general consensus was that if the bandwidth could be reduced, then transmission power being a function of bandwidth, could also be reduced. Under this consensus, the idea for the narrow bandwidth processor was conceived. Thus, it was hoped that such a processor could some day be used as spececraft flight hardware. To date, and in a general sense, this seems possible for the developed processor, in two areas (1) narrow bandwidth, and (2) size-weight requirements. However, there remain factors that still must be carefully considered before some definite application commitment can be made. In particular, the processor itself is presently limited to laboratory tests and evaluations. Reasons for this lie within the initial development intent, which was to demonstrate that human speech can be transmitted within a 160 Hz bandwidth with an acceptable degree of intelligibility. No attempt was made to optimize intelligibility or to include voice quality. Also, no attempt was made to eliminate a listener learning requirement that became obvious as a result of the development effort. Thus, the output voice replica of the processor is presently an unhumanlike sound that requires an adjustment period for optimum appreciation.

It should be noted that a significant achievement has been made with the development of this processor. It is a first and does exhibit advancement to the state-of-the-art. Before, no speech processor existed that reduced the speech bandwidth to 160 Hz and provided acceptable intelligibility. However, for this system to be useful for the intended application, improvements are necessary

The areas of particular interest are optimum intelligibility, voice quality and elimination of the dependency of listener learning. It is very possible that from extensive studies of the existing breadboard improvements in these areas can be defined along with establishing techniques for implementation.

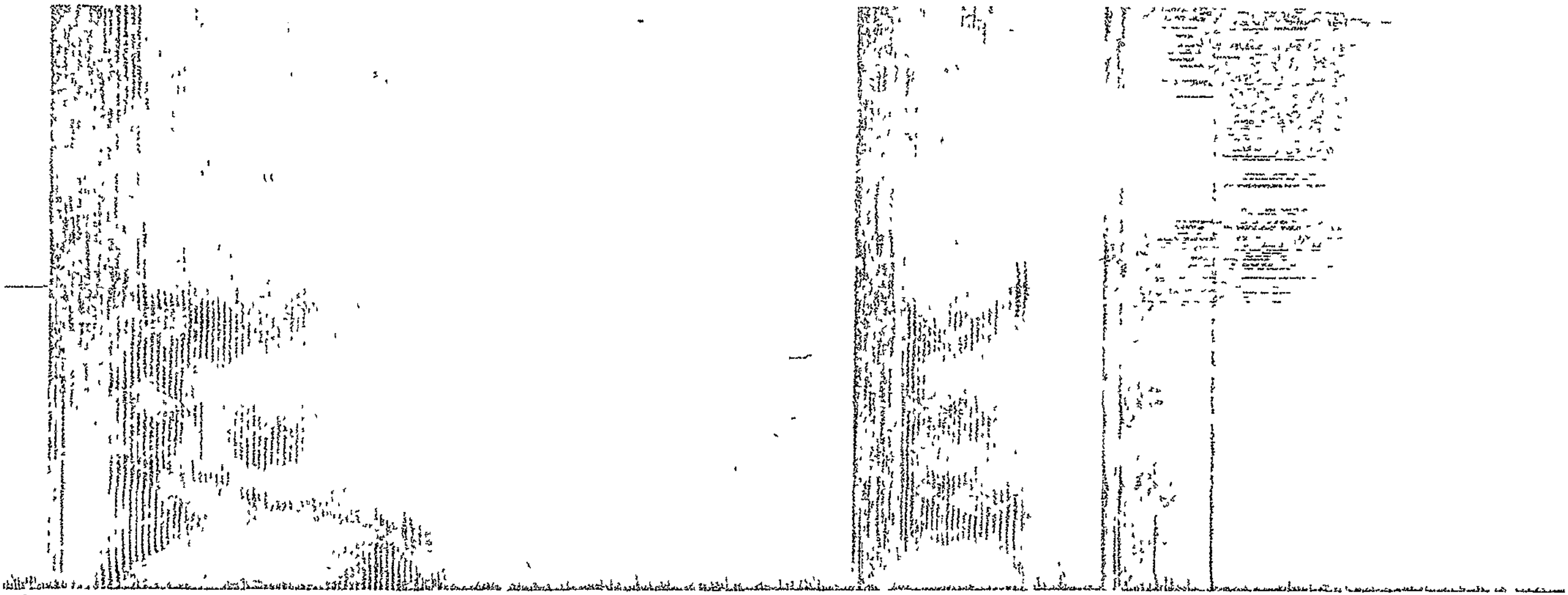
In conclusion, it is felt that the concept used to develop the narrow bandwidth speech processor, is quite an advancement to the state-of-the-art and can be developed to possibly achieve the intended application. However, continued development should make this a fact.

APPENDIX



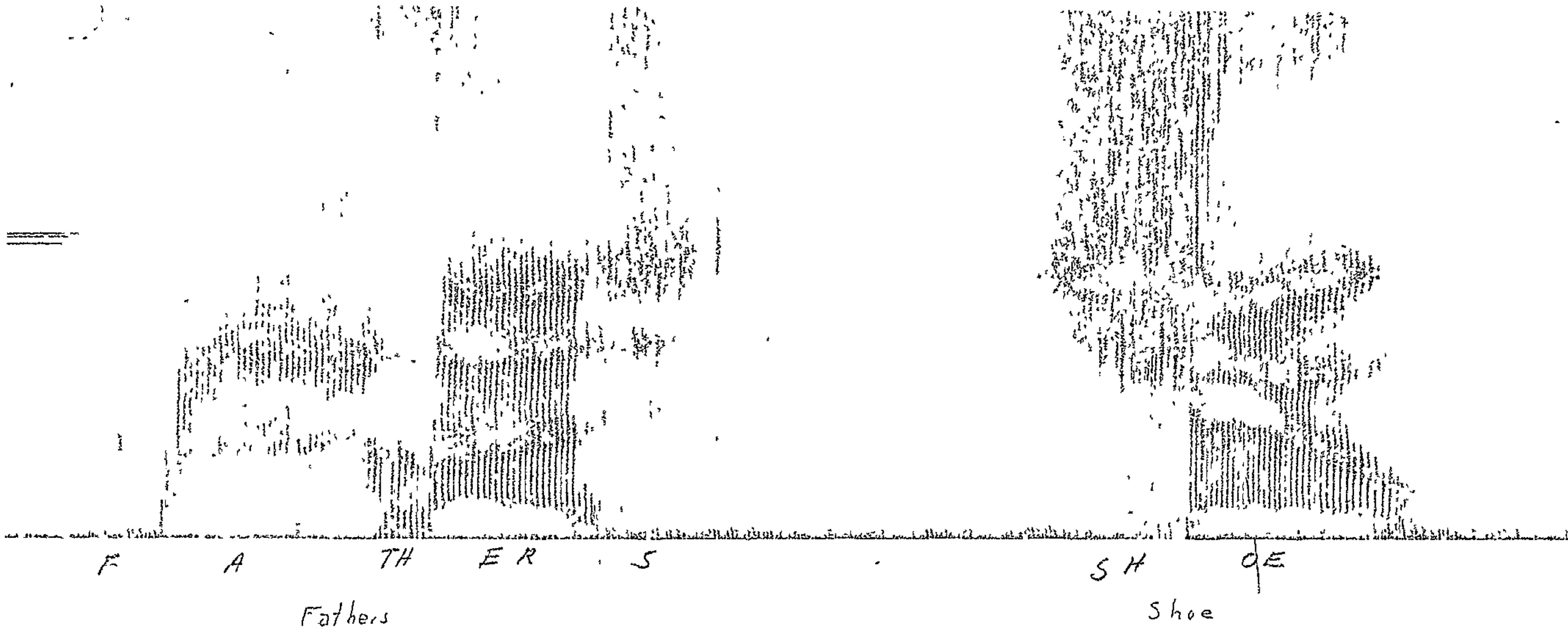
TYPE B SONOGRAM ® KAY ELECTRIC CO PINE BROOK, N J
6kc
5.25kc
4.5kc
3.75kc
3kc
2.25kc
1.5kc
.75kc
85cps
0
0.3125
0.625
0.9375
1.25sec
1.5625sec
1.8150sec
J oe ↔ T OO K↔F A TH ER S↔SH OE ↔ BEN ↔ CH ↔ OUT

TYPE B SONOGRAM ® KAY ELECTRIC CO PINE BROOK, N J

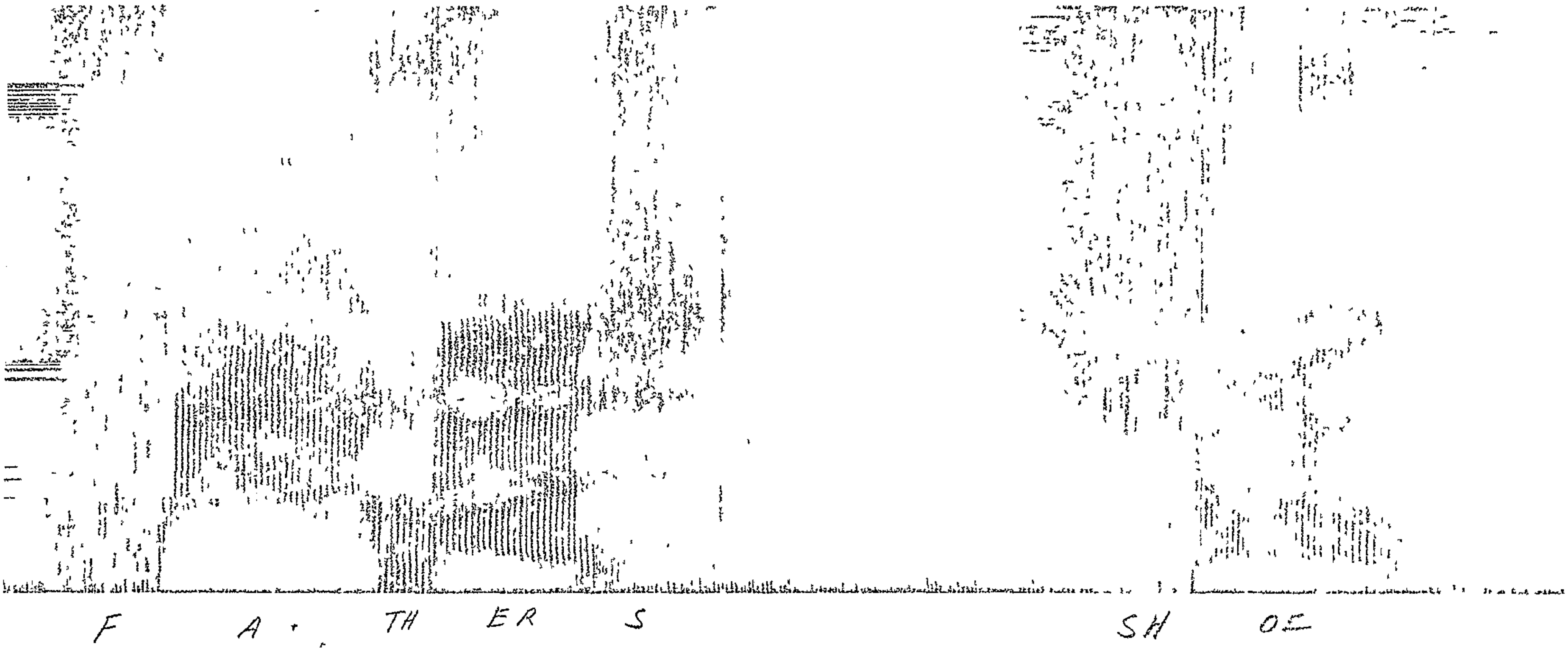


J O e T O O K

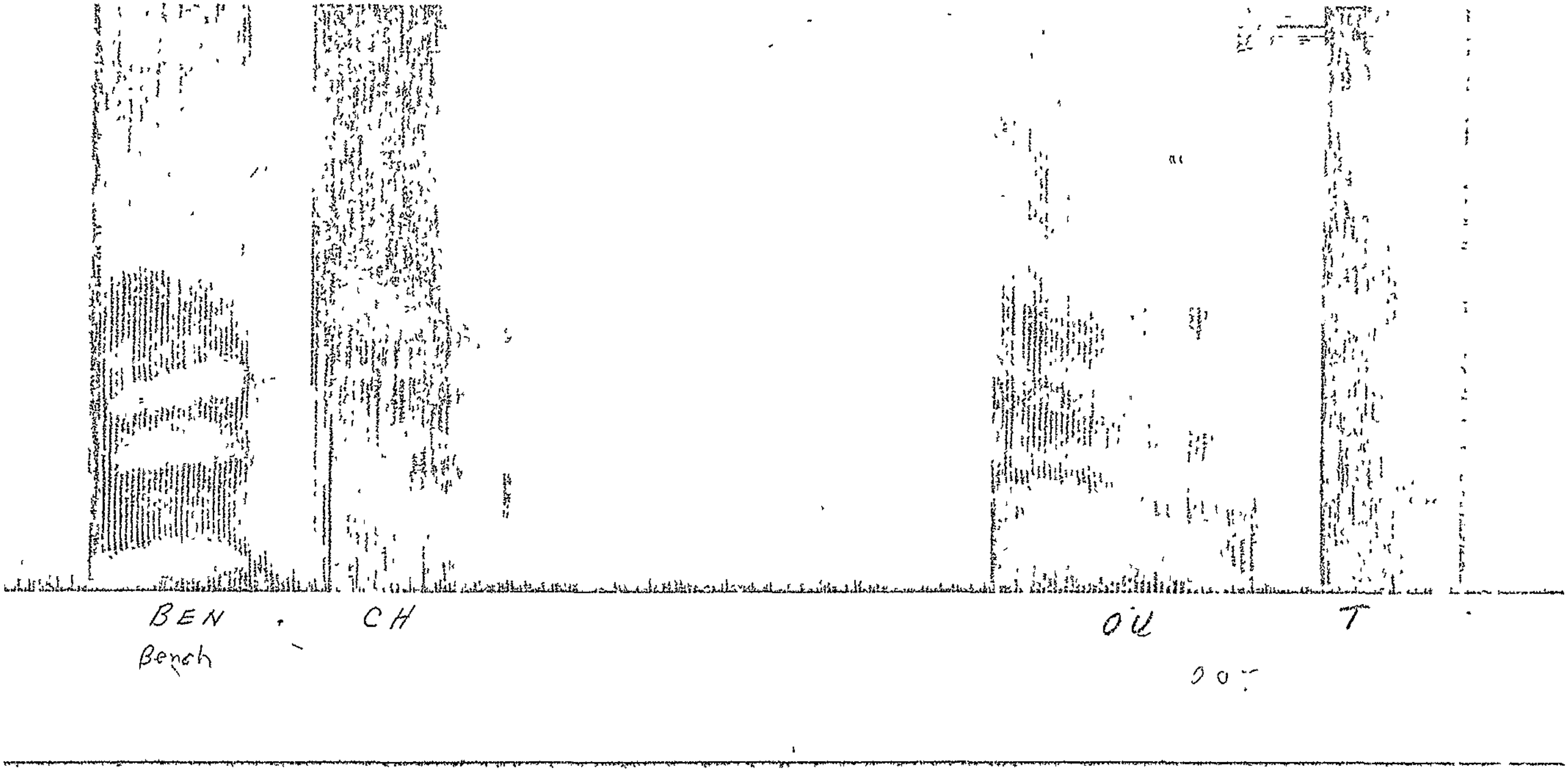
Joe Took


F A TH ER S
SH OE
Fathers
Shoe

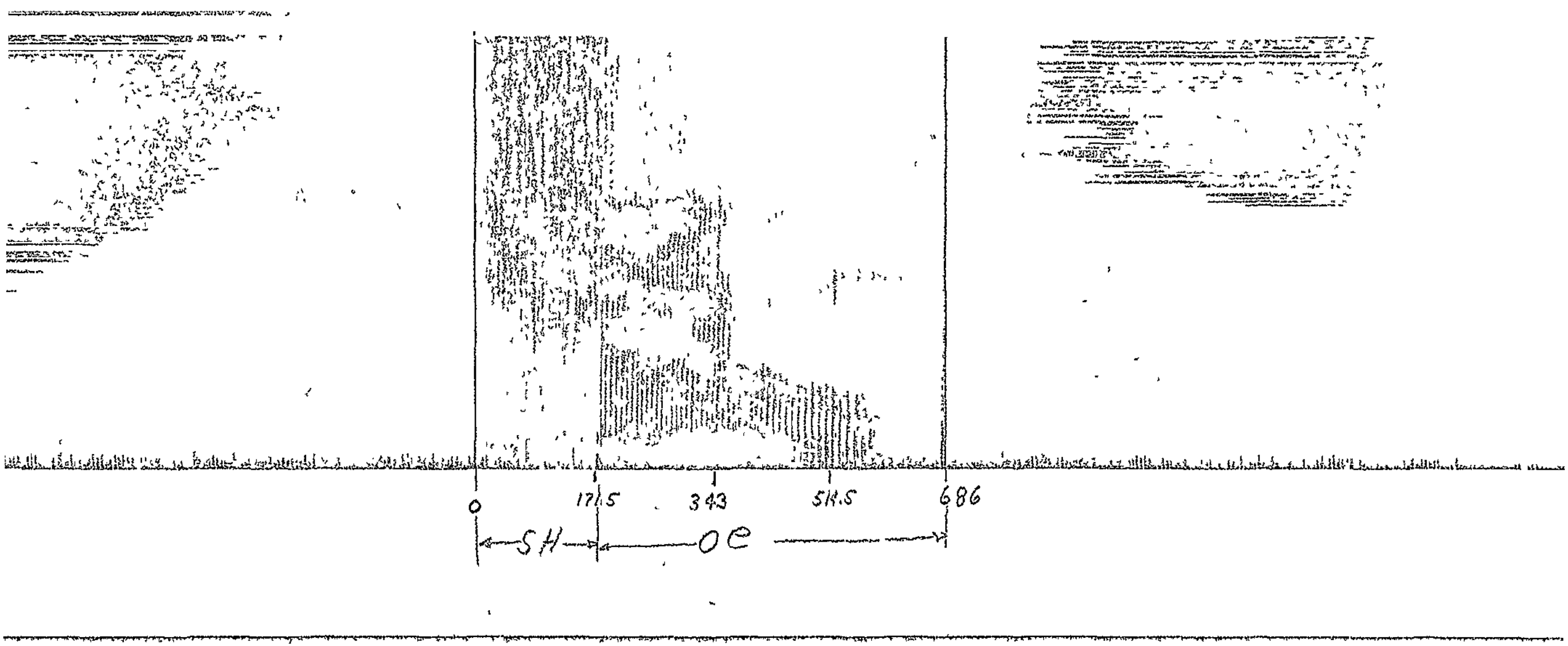
TYPE B SONOGRAM ® KAY ELECTRIC CO PINE BROOK, N J



TYPE B SONOGRAM ® KAY ELECTRIC CO PINE BROOK, N J



TYPE B SONOGRAM ® KAY ELECTRIC CO PINE BROOK, N J



LIST OF REFERENCES

1. Research and Development of an Optimized Laboratory Speech-Compression System for Spacecraft Application. Final Report Contract NAS 9-4523, Manned Spacecraft Center - NASA, Houston.

2. Speech Analysis Synthesis and Perception, J. L. Flanagan, Academic Press, Inc., 1965.