N71-23208

NASA TECHNICAL MEMORANDUM

NASA TM X-58058
APRIL 1971

REAL-TIME STATISTICAL TIME SERIES ANALYZER FOR SPEECH SEGMENTATION

A Thesis Presented to
the Faculty of the
Department of Electrical Engineering
of the University of Houston
In Partial Fulfillment
of the Requirements for the Degree
Master of Science

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
MANNED SPACECRAFT CENTER
HOUSTON, TEXAS

NASA TM X-58058

REAL-TIME STATISTICAL TIME SERIES ANALYZER FOR SPEECH SEGMENTATION

Carrington H. Stewart
Manned Spacecraft Center
Houston, Texas

ACKNOWLEDGEMENTS

The author wishes to express his gratitude to the chairman of his thesis committee, Dr. H. S. Hayre, without whose advice and encouragement this work would have been impossible. Also acknowledged is Professor W. P. Schneider and Dr. R. L. Motard, who served on the thesis committee and made useful suggestions concerning this work.

ABSTRACT

This thesis presents results of a research effort designed to advance the development of an acoustic speech segmentation procedure reported by an earlier researcher. The procedure is known as "moment analysis of the reciprocal zero crossing distances of speech." The development of this procedure is advanced through the design and construction of a real-time statistical time-series analyzer to eliminate the need for computer analysis.

This work discusses the general needs to which the advancement of the development of this segmentation procedure can be useful. Then it is shown that an electronic real-time statistical time-series analyzer is an effective method to advance the development of the segmentation procedure. In addition, a discussion of the design concepts and design feasibility is presented. Finally, the paper shows that the design concept is feasible and that the analyzer design is physically realizable, thus illustrating that the need for computer analysis to study and utilize the most promising aspect of the segmentation procedure is essentially eliminated.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE PAGE

LIST OF TABLES

TABLE PAGE

CHAPTER I

INTRODUCTION

During the past thirty years a very high degree of interest in the field of speech processing and bandwidth compression has been shown by many United States government agencies and private companies (Fant, 1960; Flanagan, 1965). Moreover, in many cases the agencies and private firms have shared their interest through joint research and development programs. However, the concern of the agencies has been by and large in the area of achieving low power, narrow bandwidth, and secure voice communications. On the other hand, the private firms have directed their attention to the development of voice recognition devices, except where work has been applied to government contracts. At any rate, both of these areas of development complement each other. Furthermore, the ultimate aim of the total effort by the agencies and private firms is to produce better hardware to improve existing conditions. For example, with the nearing of the end to the Apollo program, the National Aeronautics and Space Administration is looking toward spacecraft that can house from fifty to one hundred men. With this large population of people, multiple voice communication to and from earth as well as to remote manned satellites will be required. So the need for low power, narrow bandwidth voice systems will become a demand.

Supplementary to this growing need for new type voice systems by government agencies, private firms are seeking to meet the anticipated requirement which is to provide computers that can be programed vocally as well as in the present software manner. Not only is this method of vocal address advantageous to the users of computers, it has many other uses, automatic control systems with vocal address is an example. The need for practical voice processing equipment is a realistic one which can only be met through the results of applied research. This statement exemplifies the purpose on which the present work has been established.

Further exemplification of the purpose is seen through a brief recognition of a fundamental study in the area of speech segmentation performed by an earlier researcher (Sitton, 1969). The results of this work revealed the discovery of a new speech segmentation concept with a great promise toward advancing the present state-of-the-art in voice processing equipment. However, the illustration of the concept is only a computer simulation in the form of an algorithm. In order to make this concept more useful, additional research is needed. This need was also pointed out by Sitton. The work herein is the result of further research directed toward the implementation of the concept into a domain more conducive to practical hardware.

This implementation is done through an effort to design and develop a real-time statistical time-series analyzer to segmentize speech in real time.

CHAPTER II

THEORETICAL BACKGROUND

A comprehensive search of the speech research literature shows that the development of a statistical analyzer for speech research has been by and large a means of achieving other ends. This postulation is illustrated by the history of research in speech processing, bandwidth compression, and recognition, as well as in the recent developments in computer technology (Gold, 1969; Kock, 1962). In the past many researchers have developed all kinds of techniques to analyze human speech; one in particular is found in the work performed in 1952 by Davenport. This work dealt with an experimental study of speech wave probability distributions, and it required the use of two statistical analyzers (although not called statistical analyzers at that time). One was used to define amplitude distributions while the other was used to define zero-crossing distributions. Neither analyzer was emphasized as a possible universal test instrument for speech research but they were more or less developed as a means of achieving the desired ends.

Since the effort by Davenport, studies in speech recognition have established a definite need for statistical analysis (Gold, 1969; Sitton, 1969; Velichkin, 1963). However, the trend is now toward utilizing computers for analysis rather than the construction of test instruments (Gold, 1969; Reddy, 1967; Leytes, 1966; Sitton, 1969; Weiss, 1963). The reasons for the deemphasis on developing test instruments are not that test instruments are not needed, but rather that a computer is accessible and can expidite the major emphasis of the research. In other words, the primary objective of speech research generally is to learn about the nature of the signal itself by any means available. So a concentration on the development of test instruments is somewhat devious to the main objective of the speech research as well as to the talents of the people carrying out the research.

At any rate, a great deal of work has gone into statistical analysis for speech research be it by computer analysis or by analysis performed by an instrument designed for that particular purpose. Nevertheless more work is needed in the development of tailored instrumentation. Tailored instruments are important because they offer a bridging between theoretical abstraction and practical utilization of the theoretical concepts. This is particularly true of statistical analysis. Moreover, the design of the real-time statistical time-series analyzer, mentioned in the previous chapter, offers such bridging.

However the analyzer itself has an interesting theoretical background which is taken from Sitton's emphasis on the use of moment analysis of the reciprocal time distances between successive zero crossings of the speech signal as a segmentation procedure.

This procedure has been chosen because it seems to be quite promising for further research among a number of other methods tried. The fundamental philosophy for using the reciprocal zero crossing distance moment analysis is that these moments were found to give a reliable measure of changes in stationarity indicating phoneme transitions in speech. This moment analysis may be described mathematically by Eqs. (2-1) and (2-2).

$$\mu_{\bar{q}}(d_j^{-1}) = \frac{1}{P_j} \sum_{i=1}^{P_j} (d_{j,i}^{-1} - \overline{d_j^{-1}})^q \qquad (2\text{-}1)$$

$$\overline{d_j^{-1}} = \mu_{j,i}(d_j^{-1}) = \frac{1}{P_j} \sum d_{j,i}^{-1} \qquad (2\text{-}2)$$

where d = distance between zero crossings

d^{-1} = reciprocal of d and $\overline{d_j^{-1}}$ = average or first moment

$\mu_q(d_j^{-1})$ = all higher moments above $(q = 1)$ as a function of d_j^{-1}

$P_j = 1, 2, 3, \ldots\ldots N$

Prior to implementing Eqs. (2-1) and (2-2) into the algorithm as being mathematically description of the segmentation procedure, a process of linear detrending was performed on the speech signal. Mathematically linear detrending is a least squares line problem which is illustrated by Fig. 2-1 for a data sample. The reason for applying the detrending process was to establish a meaningful zero reference line which in effect limited the analysis to dealing with only the high frequency part of the input speech signal. This detrending process, along with the aforementioned theoretical criterion, establishes the theoretical framework of the present research effort.

With the theoretical basis of the problem known, the carry on question is what approaches should be taken to solve the problem. The research work being reported herein is an attempt to show that the method of solution can be realized through an electronic simulation of the theoretical relationships and is illustrated as follows.

Given a speech signal $S(t)$, one wishes now to define through electronic analog circuit means the first four central statistical moments. The first requirement by the theory is to perform the process of linear detrending and measuring the distances between zero crossings. This step can be achieved by passing the signal $S(t)$ through an analog differentiator with the appropriate time constant, as shown in Fig. 2-2.

The next theoretical requirement is to obtain the reciprocal of the distances between the zero crossings. This can be done by operating on the signal with an analog divider

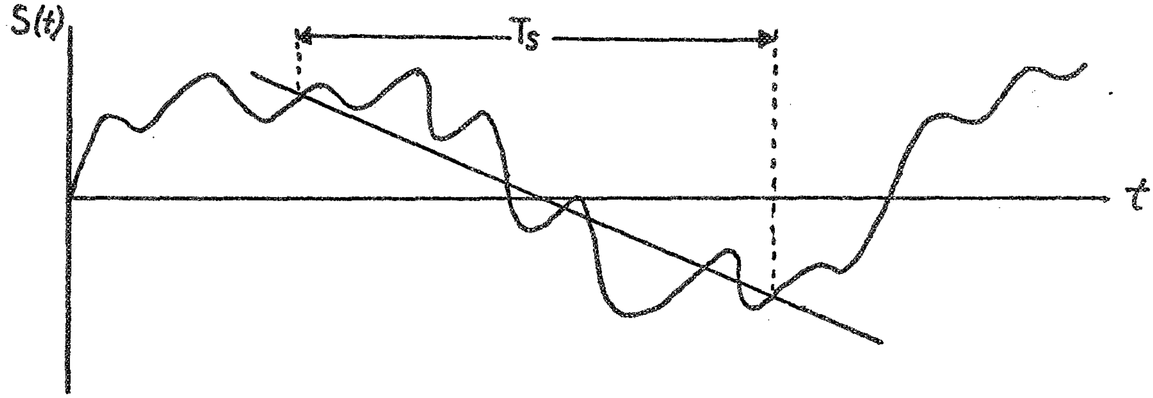


Fig. 2-1. Linear Detrending Establishing
Zero Reference

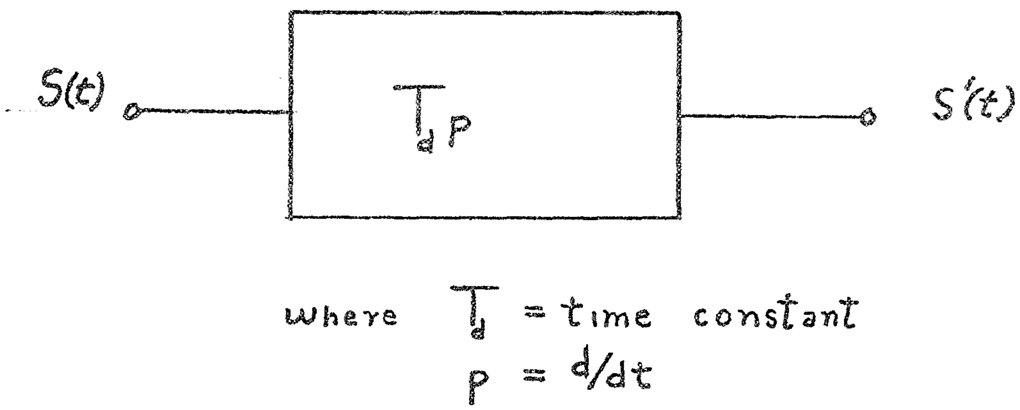


Fig. 2-2. Analog Differentiator

(see Fig. 2-3) after zero crossing detection, and frequency to amplitude conversion.

The next step is to find the average value or first central moment of the reciprocal $[\overline{S(T_i^{-1})} = \bar{S}]$ zero crossing distances. This is done by using an analog averaging network or integrator with the appropriate integration time constant (see Fig. 2-4).

The next process is to develop a moment generating function. This is done by using a high quality operational amplifier as a difference network (see Fig. 2-5) and taking the difference between $S(T_i^{-1})$ and $\bar{S}$.

With this difference quantity $[S(T_i^{-1}) - \bar{S}]$ defined, any central moment above 1 can be found by using the proper arrangement of analog multipliers; this is shown in Fig. 2-6 for moments 2nd, 3rd, and 4th order. The output of the multipliers should be a signal representative of the moments that are described by Eq. (2-1).

In summary, the approach to solving the problem is to use a set of properly ordered analog networks to achieve the signal moment for analysis as illustrated by the block diagram of Fig. 2-7. The proper ordered set of electronic analog networks represent the real-time statistical-time series analyzer. This name is chosen to describe the electronic device because the final products of the assembly of electrical networks are statistical quantities and they are derived

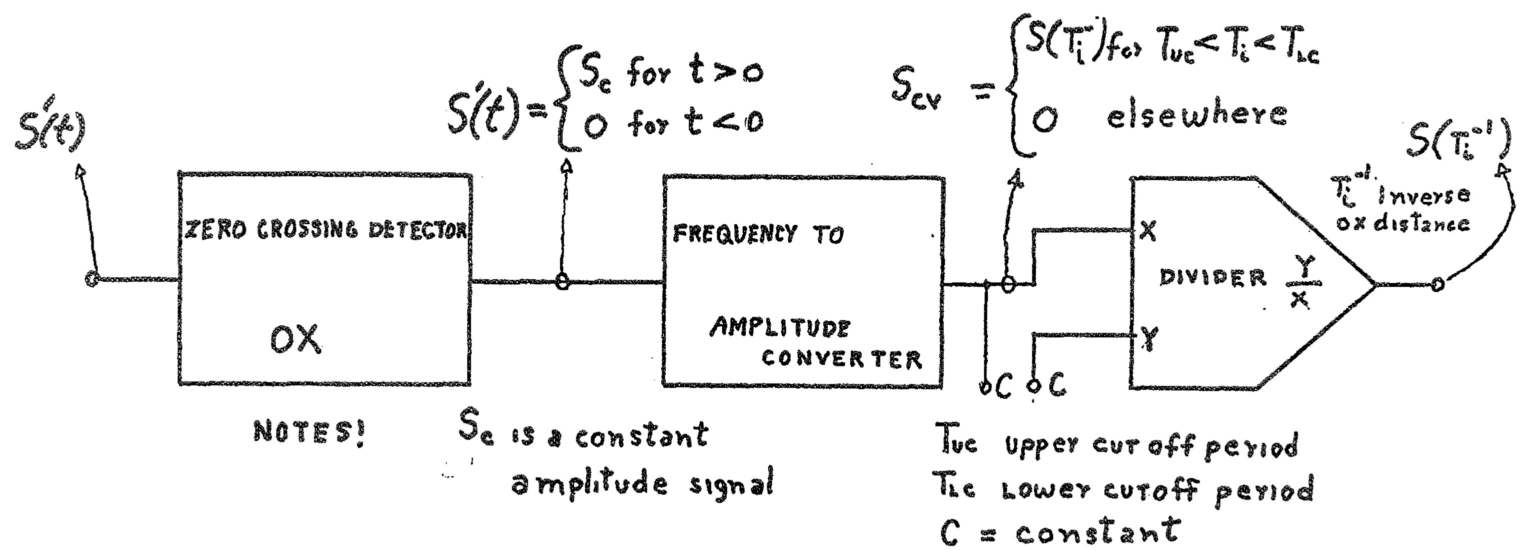


Fig. 2-3. Zero Crossing Detector and Frequency to Amplitude Converter

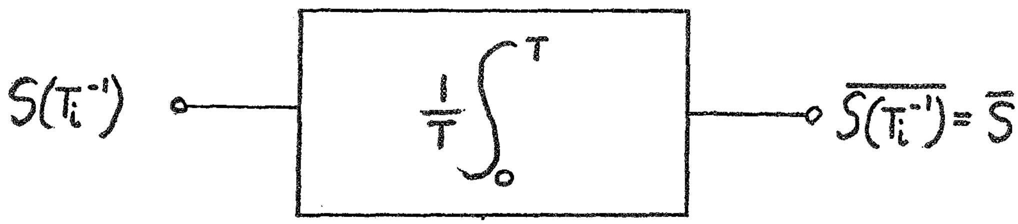


Fig. 2-4. Signal Integrator

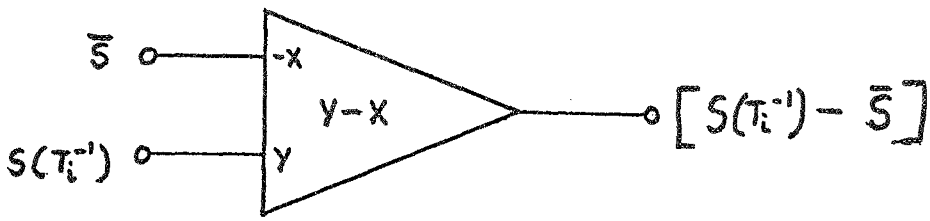


Fig. 2-5. Difference Network

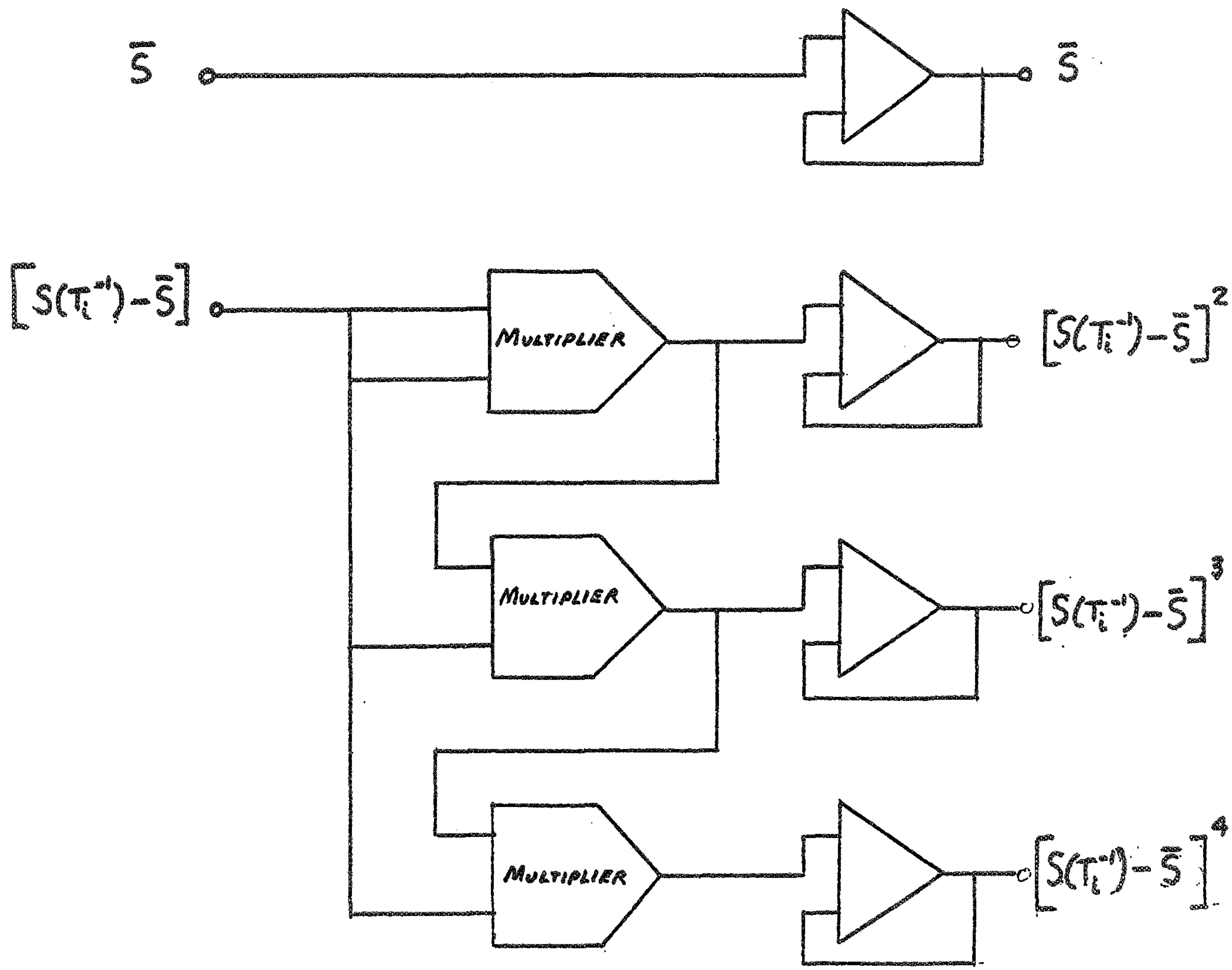
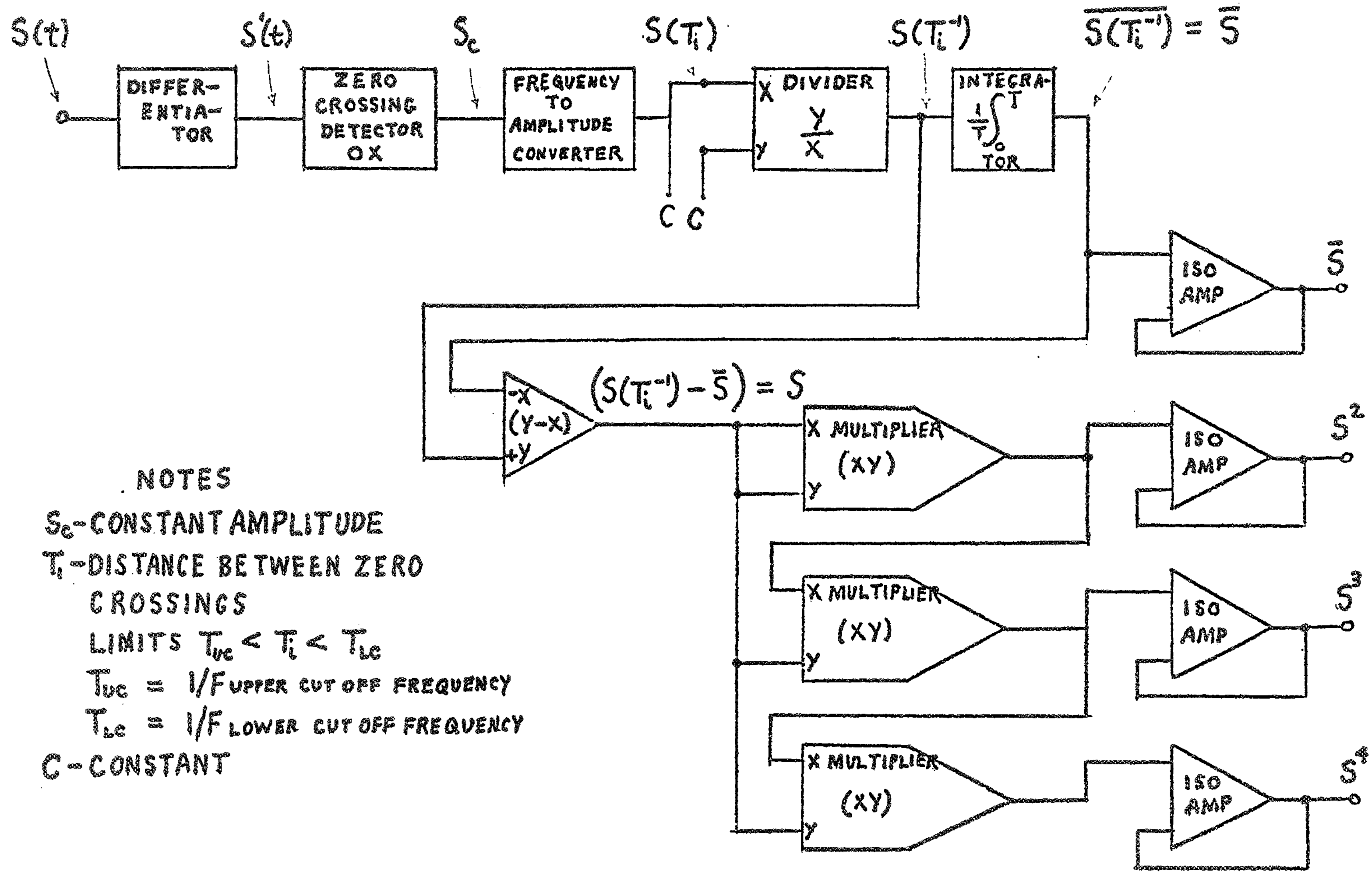


Fig. 2-6. Arrangement of Analog Multipliers



NOTES

S_c - CONSTANT AMPLITUDE

T_i - DISTANCE BETWEEN ZERO CROSSINGS

LIMITS $T_{UC} < T_i < T_{LC}$

$T_{UC} = 1/F_{\text{UPPER CUT OFF FREQUENCY}}$

$T_{LC} = 1/F_{\text{LOWER CUT OFF FREQUENCY}}$

C - CONSTANT

FIG. 2-7 COMPLETE ANALYZER

from the real-time domain representation of the input signal.

CHAPTER III

SYSTEM DESIGN

The problem expressed in Chapter II is translated into an electronic analog system simulating the theoretical relationships of Eqs. (2-1) and (2-2).

The system design has been carried out in four phases. The first phase is called signal conditioning; the second, signal conversion and inversion; the third, signal integration and differing; and the fourth, signal moment generation. In order to provide a clear understanding of the overall design concept, this chapter gives a more detailed discussion of the four design phases.

Phase 1: Signal Conditioning

The signal of concern in this investigation is human speech. So the problem in this system design phase is to provide the proper signal conditioning. As stated in Chapter II, the operation to be performed on the speech signal is a moment analysis of the reciprocal zero crossing distances. Therefore a meaningful zero reference must be established and its zero crossings extracted in addition to the speech signal being of sufficient signal amplitude. Thus the signal conditioning network design consists of an input amplifier, a differentiator, and a zero crossing detector. The input amplifier provides the signal voltage range for establishing a high signal to noise ratio; the differentiator establishes the meaningful zero reference whereas the zero crossing detector extracts the signal zero crossings. Figure 3-1 shows such a signal conditioning network.

Phase 2: Signal Conversion and Inversion

After the signal zero crossings are detected, the next phase measures the reciprocal zero crossing distances of the signal. The essential task for this phase is to express the reciprocal zero crossing distances as a function of amplitude and time. This was accomplished by developing a network, illustrated in Fig. 3-2, called a frequency to amplitude converter.

This network functions as follows. The appearance of the first zero crossing starts a high speed digital clock. During the time interval between the first zero crossing and the appearance of a second zero crossing the clock drives a digital counter with an eight bit parallel output. The appearance of the second zero crossing stops the clock and the digital value of the counter is inverted and dumped into an eight bit digital-to-analog converter via an eight bit storage register. Then the output voltage of the digital-to-analog converter is directly proportional to the reciprocal zero crossing distance. More circuit details of the design

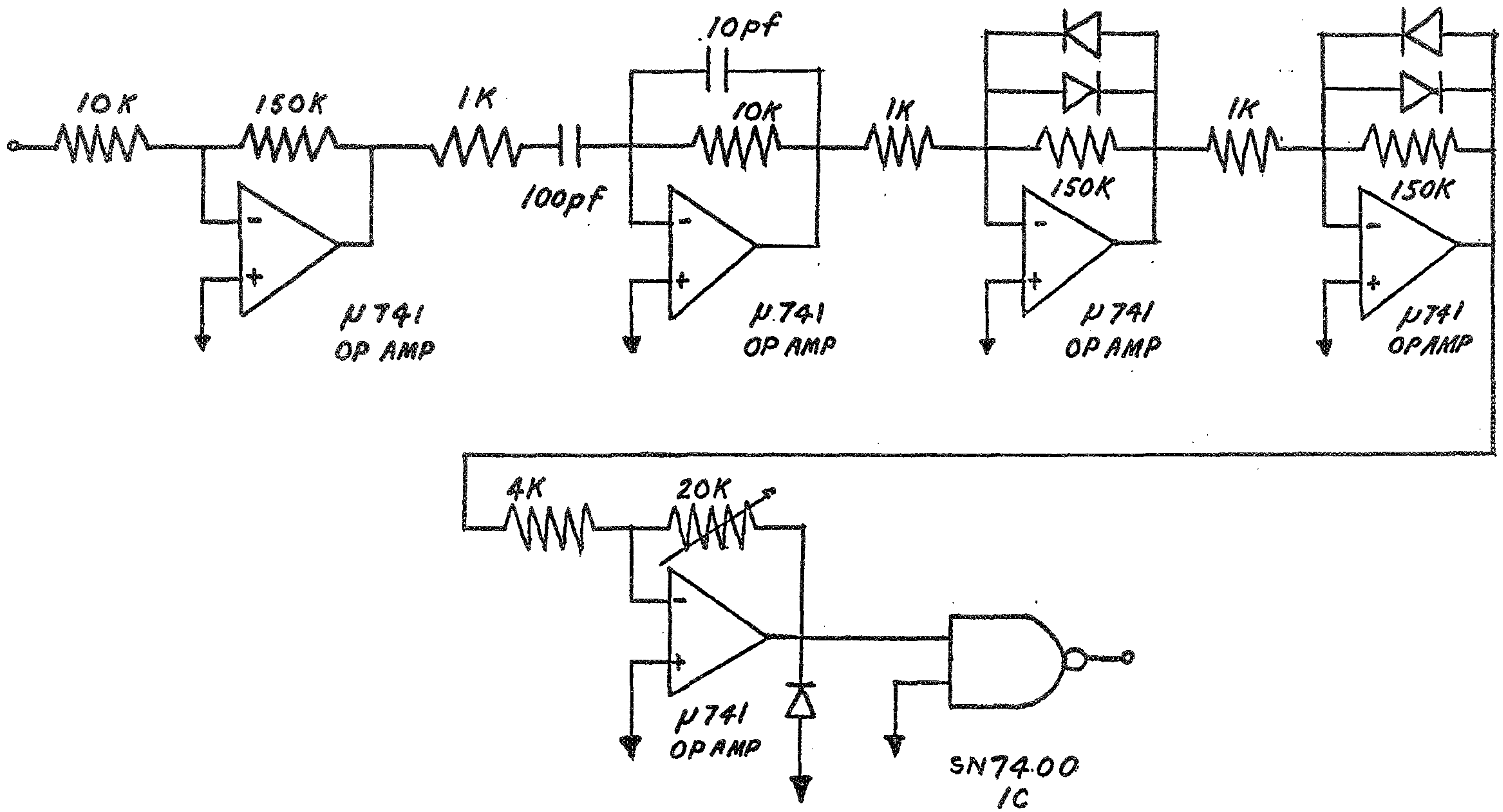


FIG 3-1 ZERO CROSSING DETECTOR AND PRE-AMPLIFIER

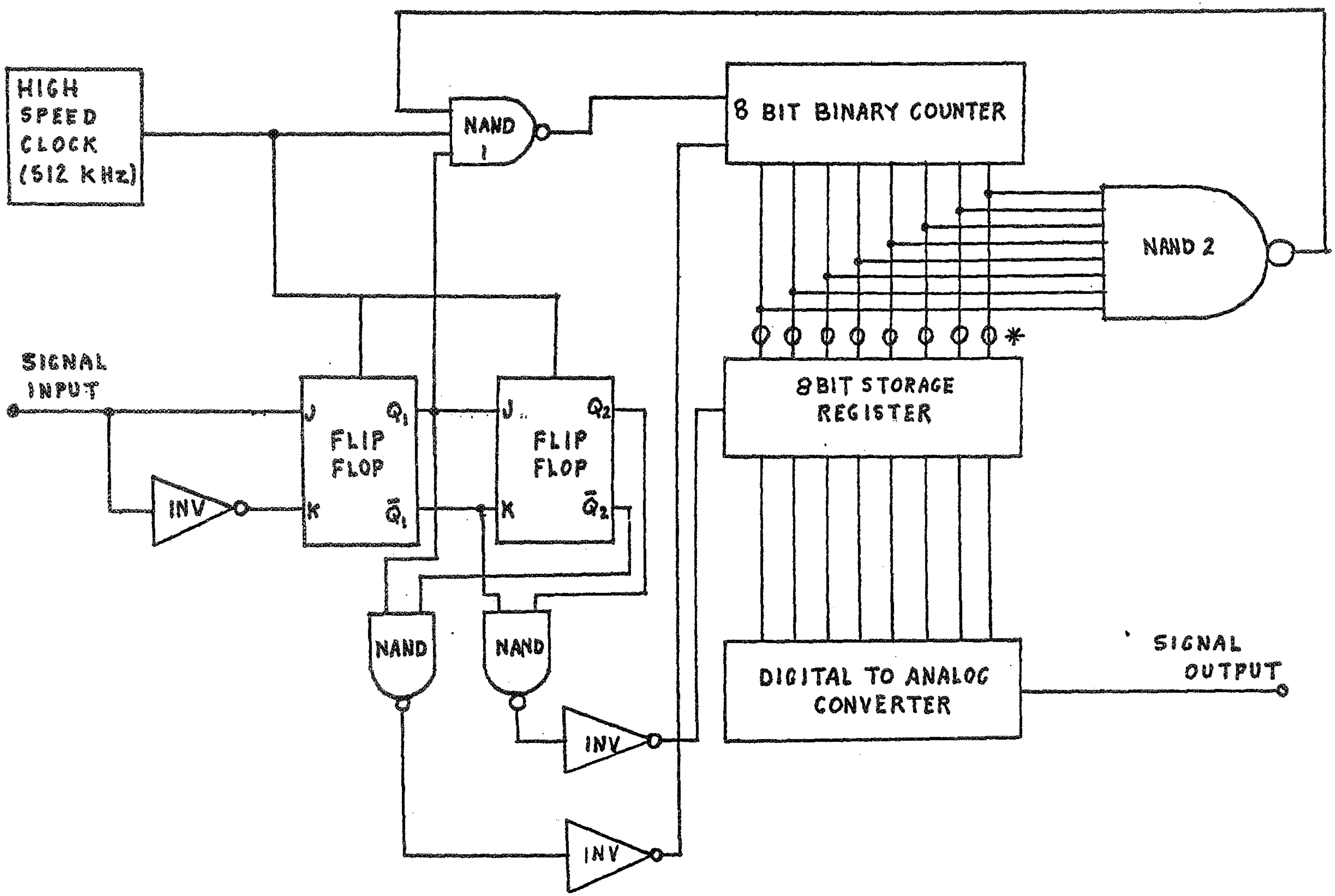


FIGURE 3-2 FREQUENCY TO AMPLITUDE CONVERTER

of this signal conversion and inversion network are provided in the Appendix. Figure 3-2 shows the typical design of this network.

Phase 3: Signal Integration and Difference

The networks defined in Phases 1 and 2, transformed the speechsignal into a signal with amplitudes directly proportional to the reciprocal of the zero crossing distances in order to perform the moment analysis. The signal integration and difference network described in this phase begins the moment analysis and basically consists of an integrate-and-hold circuit and a difference amplifier. The circuit and amplifier operate on alternate 3.75 millisecond time periods. In other words, the integrator and difference amplifier are controlled by 7.5 millisecond square wave drive to four analog gates. The integrator is turned on and integrates the input signal for 3.75 milliseconds. Also, during this time period no input is applied to the difference amplifier. At the end of 3.75 milliseconds the output of the integrator along with the input to the integrator are connected to the two inputs of the difference amplifier. The output of the amplifier then gives the difference between the integrator input and its output for a period of about 3.50 milliseconds. During the remaining 0.25 milliseconds, the integrator is reset to integrate the next alternate 3.75 millisecond period and repeat the entire process over again.

The time period of 3.75 milliseconds was chosen because it is a submultiple of 15 milliseconds which has been found to be a period over which speech is assumed to be stationary (Sitton, 1969).

Figure 3-3 shows a circuit diagram of the complete integration and difference network. As a supplement to this circuit diagram, Fig. 3-4 shows a timing diagram which illustrates the proper time relationships of the integration and hold process as well as the difference network. It also shows the analog gate control logic signals.

Phase 4: Signal Moment Generation

The signal output of the difference network may be thought of as a statistical moment generating function. A moment generating network was required in order to provide the moments of the signal. This network was designed by properly employing a set of analog multipliers. Figure 3-5 shows this network for the generation of moments 2, 3, and 4.

These four phases, designed and developed a system simulating Eqs. (2-1) and (2-2). The final outputs of the system are time-series-patterns which are later correlated with those achieved through the use of Sitton's computer algorithm. Spectral patterns are made and presented as part of this work; the speech material used to obtain these patterns is the same word set used by Sitton. Some comparison

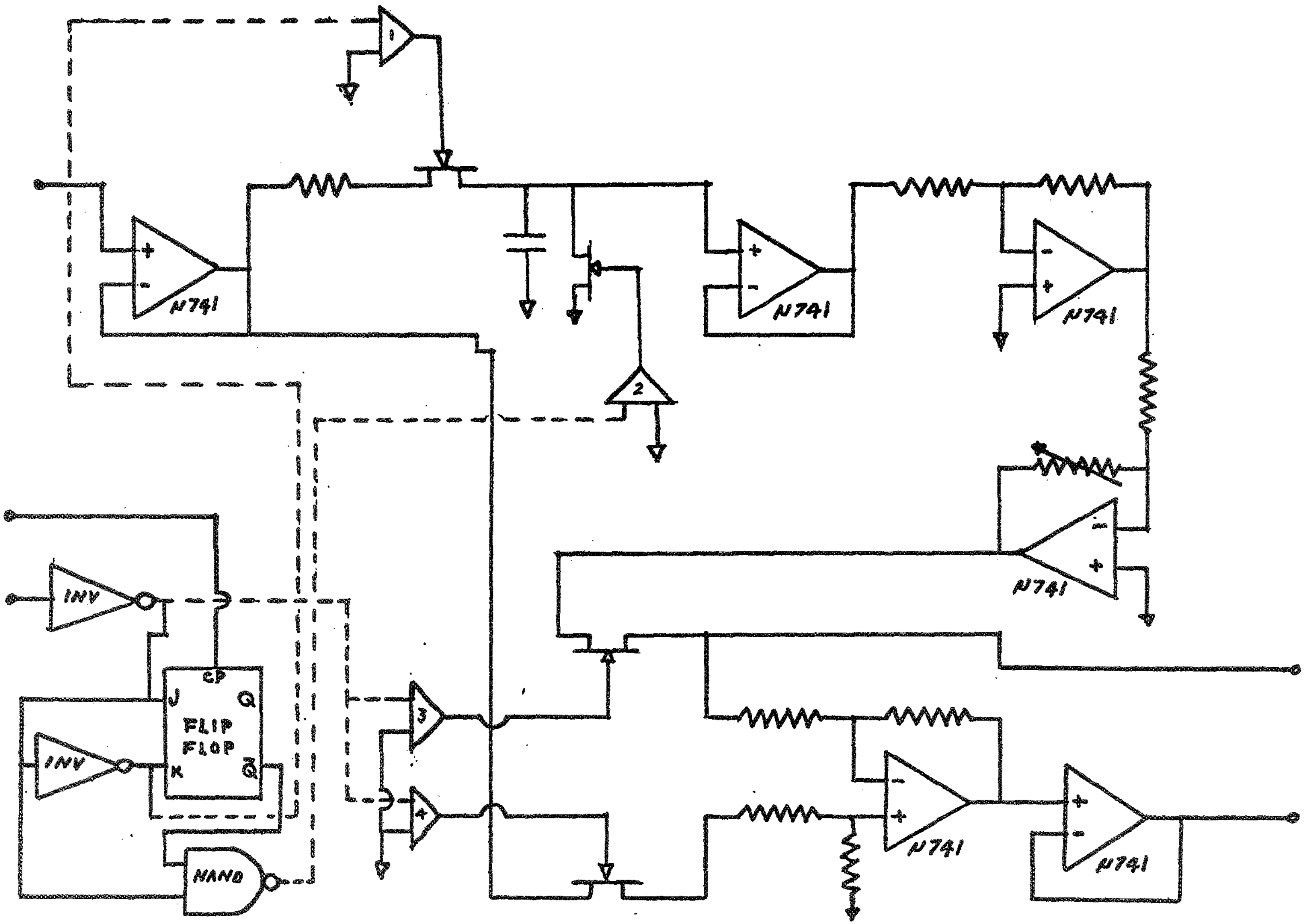


Fig. 3-3. Integration, Hold and Difference Network

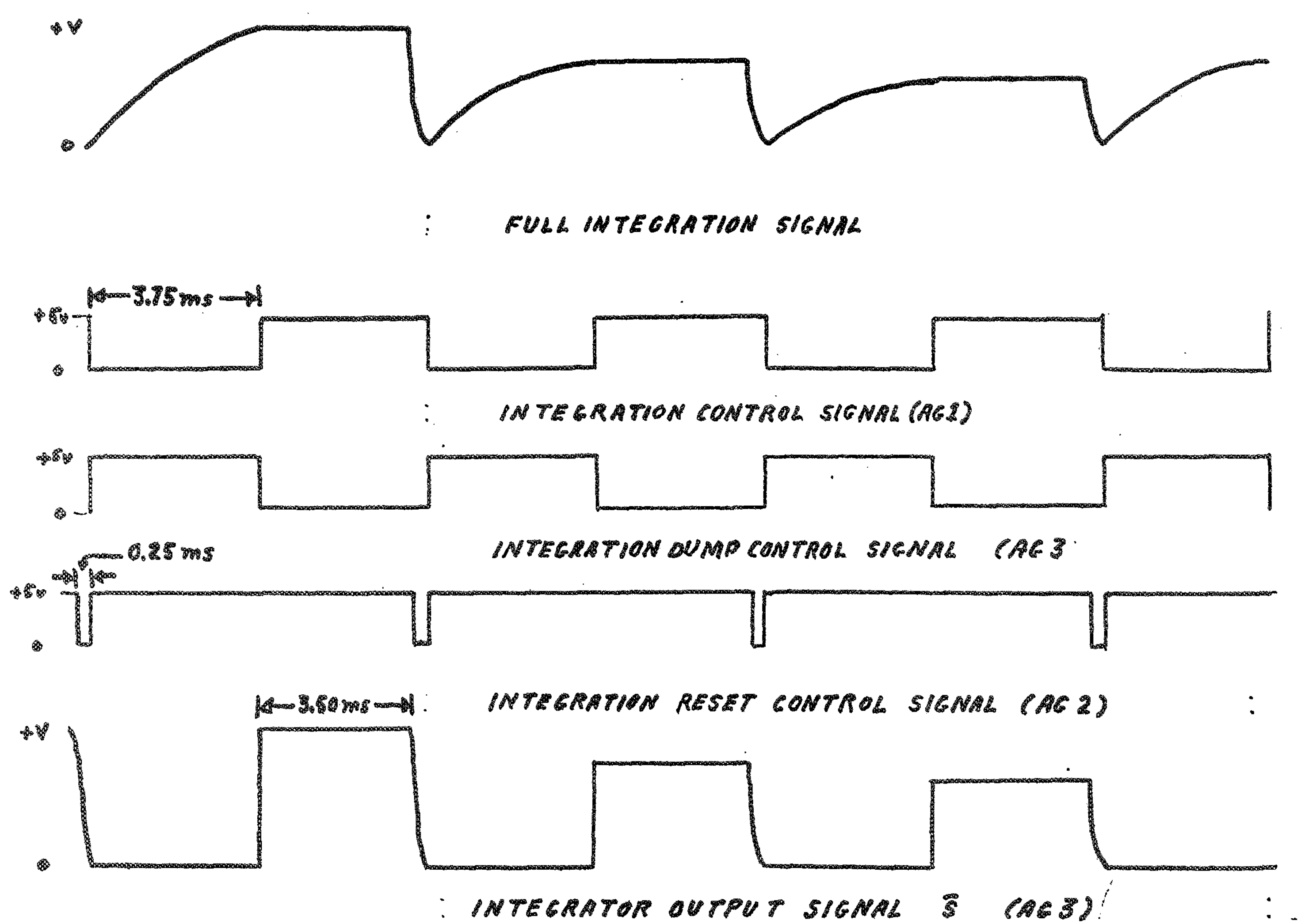


Fig. 3-4. Timing Diagram for Integrate, Hold, and Difference Network

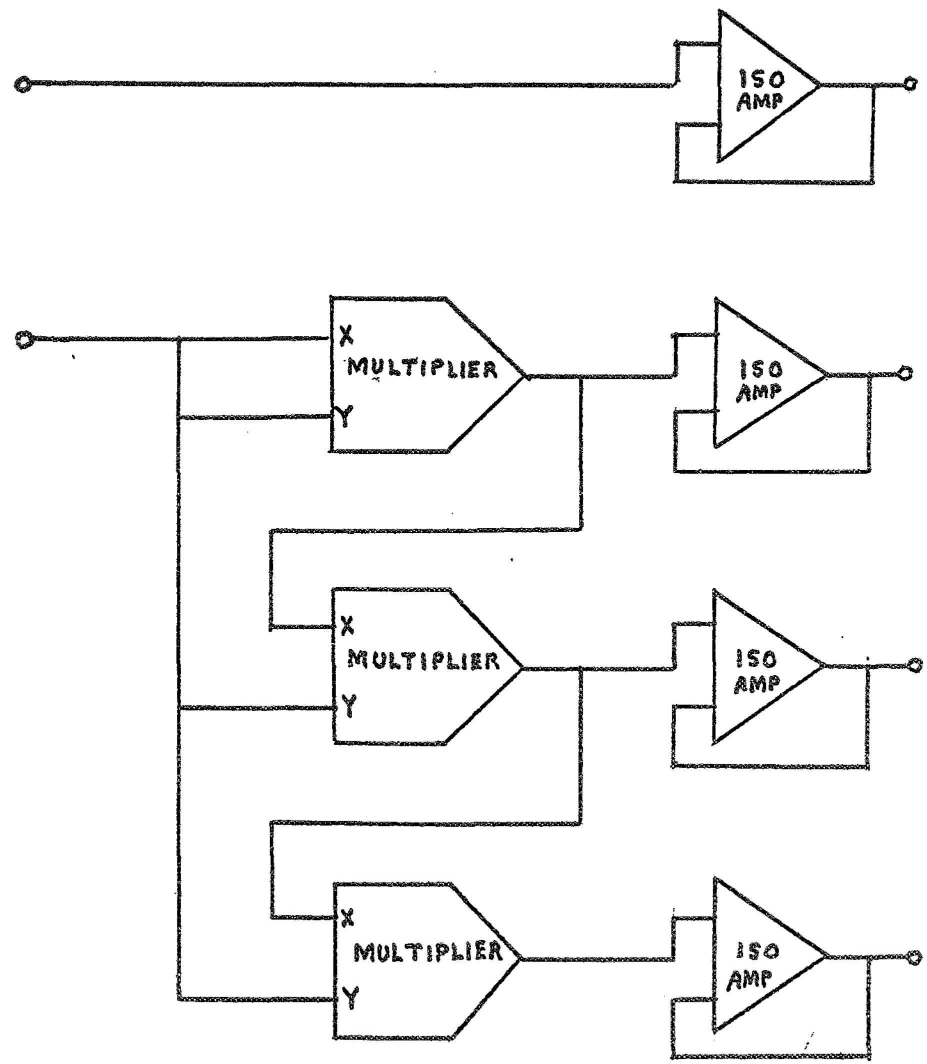


FIGURE 3-5 MOMENT GENERATING NETWORK

between the two spectral patterns was made by the present researcher but the most critical comparison is left for further research. More is said about this point and the total achievement of the present work in the next two chapters.

Figure 3-6 shows an overall diagram of the final system design and the system is called a real-time statistical time-series analyzer. This name is given to the system because the analysis is statistical in nature and it is done in real time as well as in the signal time domain.

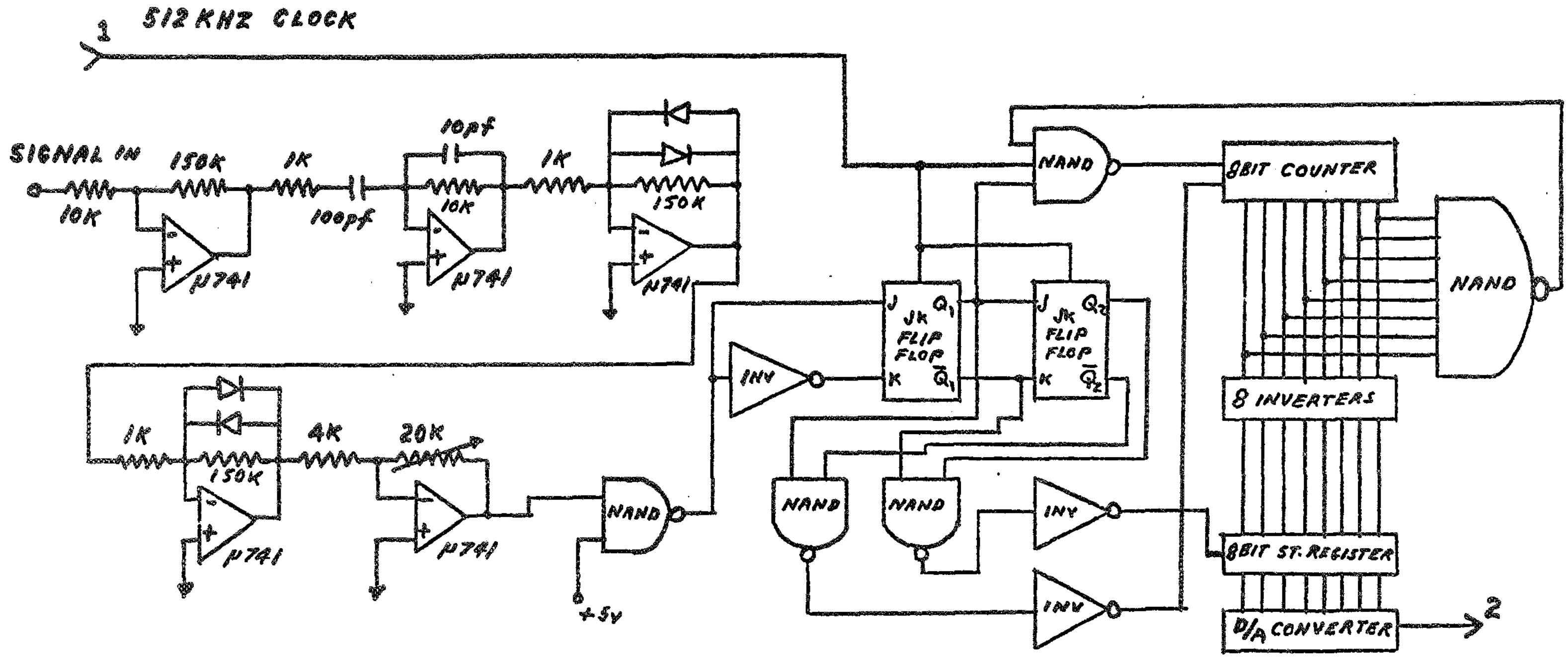
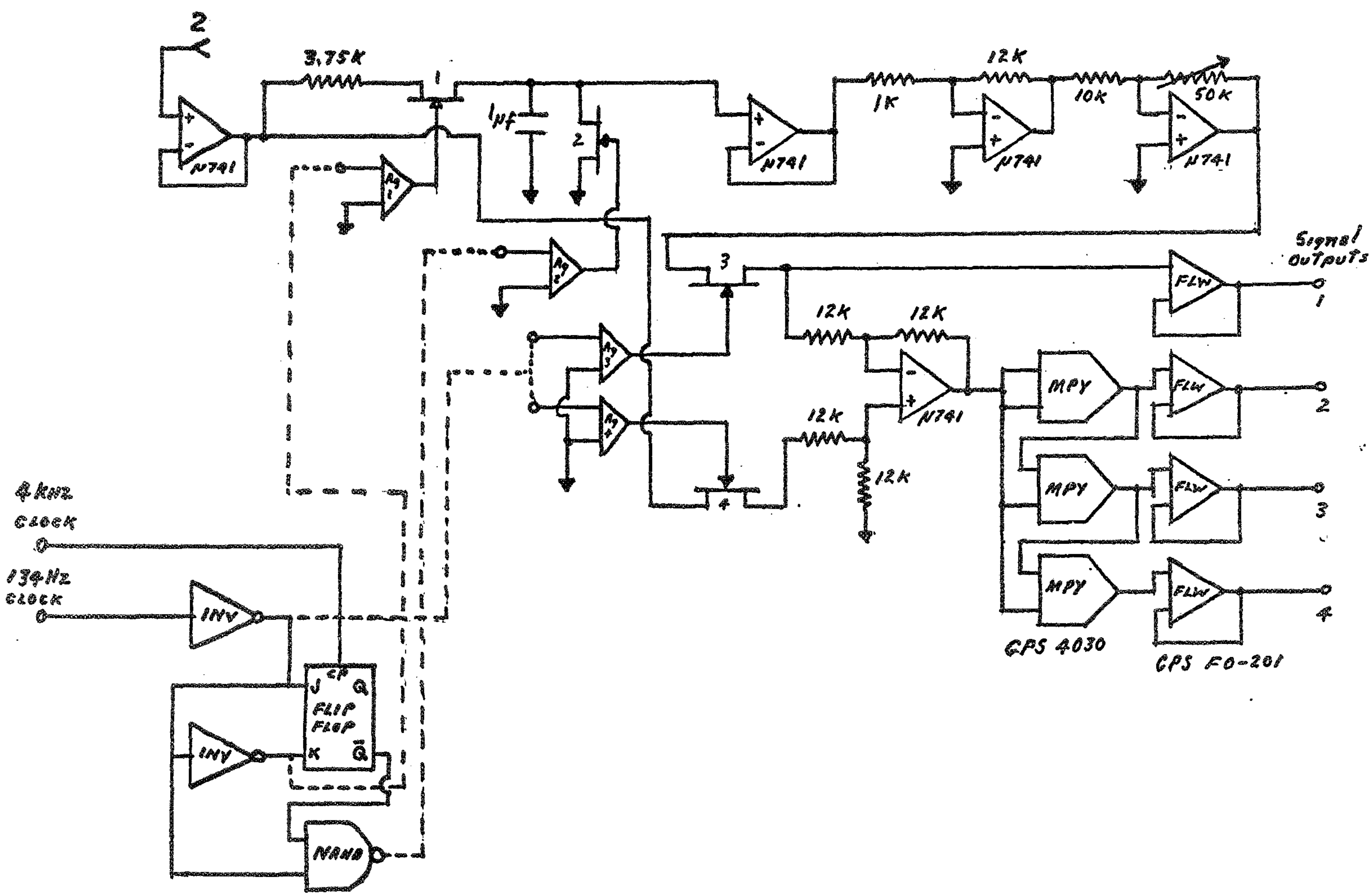


Fig. 3-6. Circuit Diagram of Complete Analyzer



2
3.75K
1
1μf
2
μ741
1K
12K
10K
50K
Ag 1
Ag 2
Ag 3
Ag 4
3
4
12K
12K
12K
12K
Signal outputs
FLW
MPY
1
2
3
4
CPS 4030
CPS FO-201
4KHz clock
134Hz clock
INV
J
CP
Q
FLIP FLOP
K
NAND

Fig. 3-6 (Cont'd)

CHAPTER IV

PROTOTYPE AND FEASIBILITY STUDY

The experimental prototype refered to in this chapter is the real-time statistical time-series analyzer illustrated by the diagram of Fig. 4-5. Up to the present point, however, much has been said about the analyzer design approach and method of implementation. On the contrary, this chapter attempts to present data which defines and proves the physical realizability of the analyzer. This data is presented through a short discussion of the final analyzer design and the presentation of the results of a feasibility study conducted on the prototype.

Figure 3-6 shows the complete and final circuit design of the analyzer and this diagram is repeated in Fig. 4-1. Moreover, this final design is a more detailed representation of the conceptual block diagram defined by Fig. 2-7 and repeated in Fig. 4-2. However, there is a basic difference in that the final design did not require the use of an analog divider to achieve the reciprocal zero crossing distances as indicated by y/x in Fig. 4-2. Instead, the reciprocal zero crossing distances were obtained in the frequency-to-amplitude conversion process by inverting the 8 bit counter output shown in Fig. 3-2, prior to digital-to-analog conversion. This inversion makes the highest input signal frequency

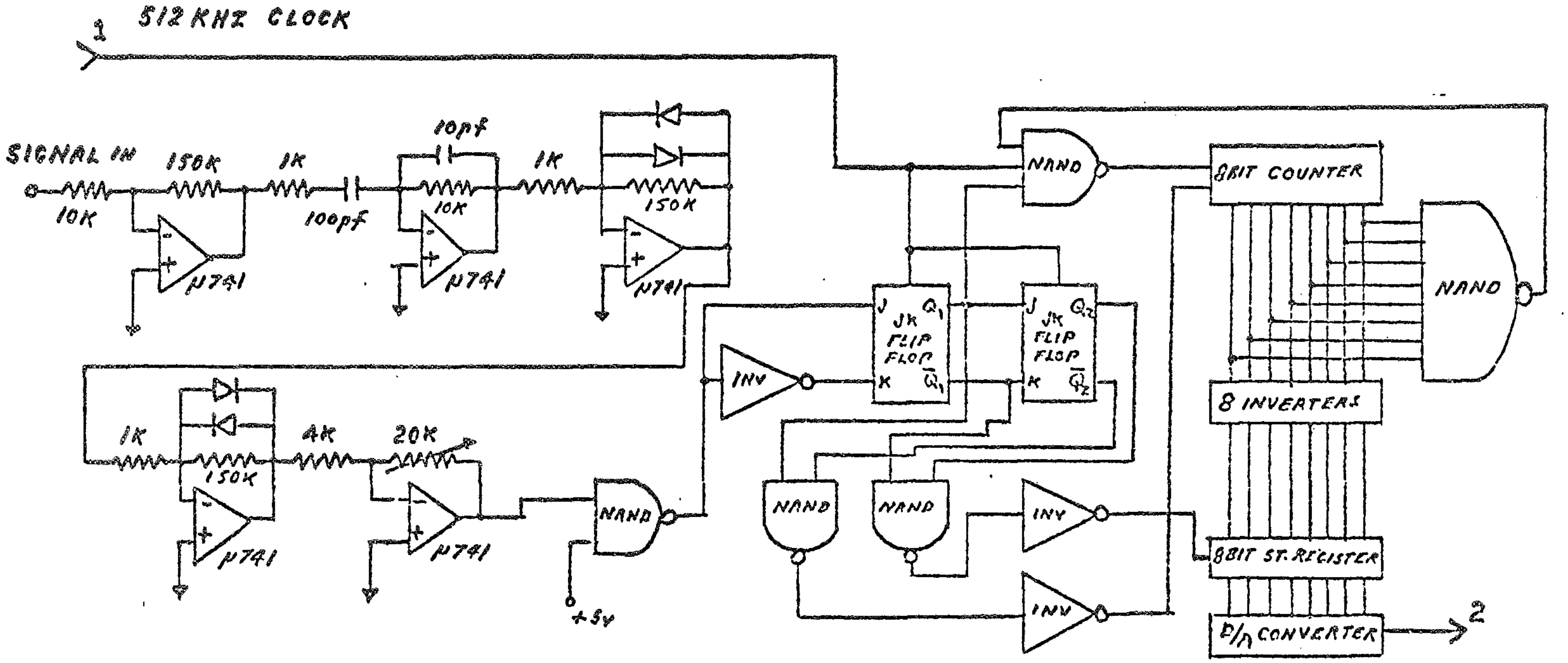


Fig. 4-1. Circuit Diagram of Complete Analyzer

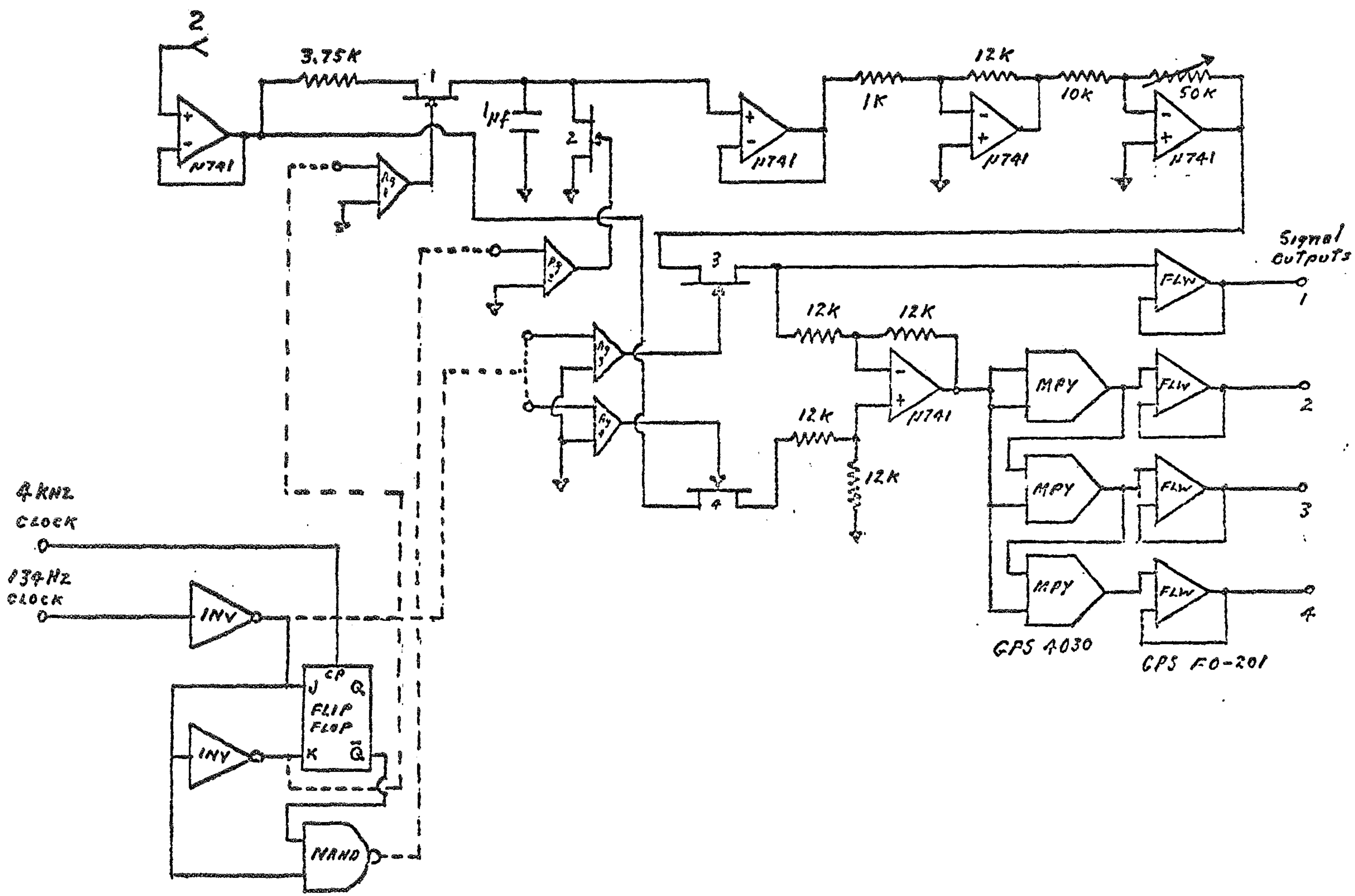
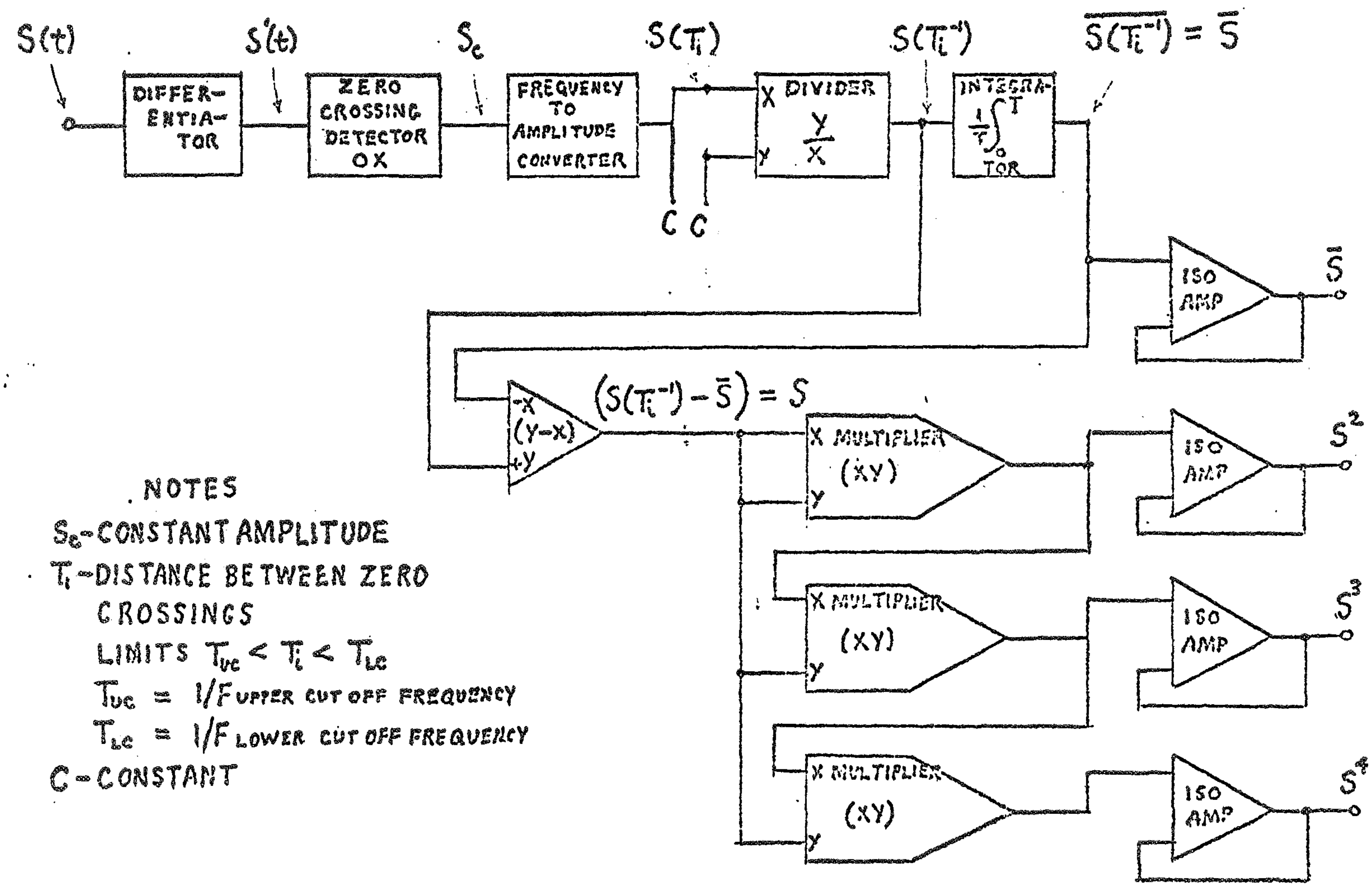

2
3.75K
1
1μf
2
μ741
1K
12K
10K
50K
3
4
12K
12K
12K
12K
Signal Outputs
1
2
3
4
FLW
MPY
GPS 4030
GPS FO-201
4 KHz CLOCK
134 Hz CLOCK
INV
J
CP
Q
FLIP FLOP
K
NAND

Fig. 4-1 (Cont'd)



NOTES

S_c - CONSTANT AMPLITUDE

T_i - DISTANCE BETWEEN ZERO CROSSINGS

LIMITS $T_{UC} < T_i < T_{LC}$

$T_{UC} = 1/F_{\text{UPPER CUT OFF FREQUENCY}}$

$T_{LC} = 1/F_{\text{LOWER CUT OFF FREQUENCY}}$

C - CONSTANT

FIG 4-2 COMPLETE ANALYZER

correspond to the highest amplitude of the digital-to-analog converter output voltage. This relationship is the same affect as desired through the use of the analog divider.

Other points of interest about the final design are the time constants associated with the presentation of Figs. 3-2 and 3-4. These figures illustrate, among other things, the use of appropriate differentiation and integration time constants. In the final design, these constants were defined. The differentiator constant, for example, was found to be (T_d = 1ns) one nanosecond; and the integration constant was found to be (T_i = 3.75 ms) three-point-seven-five miliseconds. One nanosecond was chosen for T_d because it provides an optimum trade-off between establishing a true zero reference for the input signal and maintaining system noise immunity. Likewise, three-point-seven-five miliseconds was chosen for T_i because it represents a fair trade-off between obtaining a workable average of the input signal and remaining within the limits of signal stationarity in order to present a useful signal analysis in real-time. More is said about this point in the following chapter.

The feasibility study conducted using the prototype model was designed to uncover the correlation between the statistical moments obtained by the earlier researcher's algorithm and the present researcher's electronic model. A sufficient degree of correlation between the data obtained through the use of the algorithm and data obtained from the electronic model establishes the physical realizability of the analyzer.

The study consisted of an analysis of key speech samples taken from the word space outlined in Table I. Out of this word space, the speech samples were the words "which" and "sunless" spoken by a speaker with a general American accent. Although both words were used as speech samples, only the word "sunless" was used to establish the desired degree of correlation between the algorithm data and the data taken from the electronic model. "Sunless" was used because the algorithm data consisted only of the word "sunless" as its speech sample.

Nevertheless, the data from the electronic model was obtained by requiring the speaker to record on magnetic tape the complete word space. Then the tape recording was used as the signal source for the operational test. The test setup is shown in Fig. 4-3. In this figure the recorded speech is shown to be provided, via a tape recorder, to the analyzer. Then the analyzer is connected to an oscilloscope. The oscilloscope is triggered by the input speech signal while the analyzed speech signal is displayed as the scope trace. Also, when the scope trace appears, a polaroid picture is taken of the trace and used as the data collection and storage technique. This technique of polaroid pictures was used because it proved to be the best technique out of

TABLE I

WORD SPACE USED IN STUDY (Sitton, 1969)

Orthographic	Phonemic
which*	|wItʃ|
into	|In'tu|
only	|on'li|
some	|sAm|
did	|dId|
many	|me'ni|
sunless*	|sAn'les|
Monday	|man'di|
zero	|zi'ro|
himself	|him'self|
speechless*	|spitʃ'les|

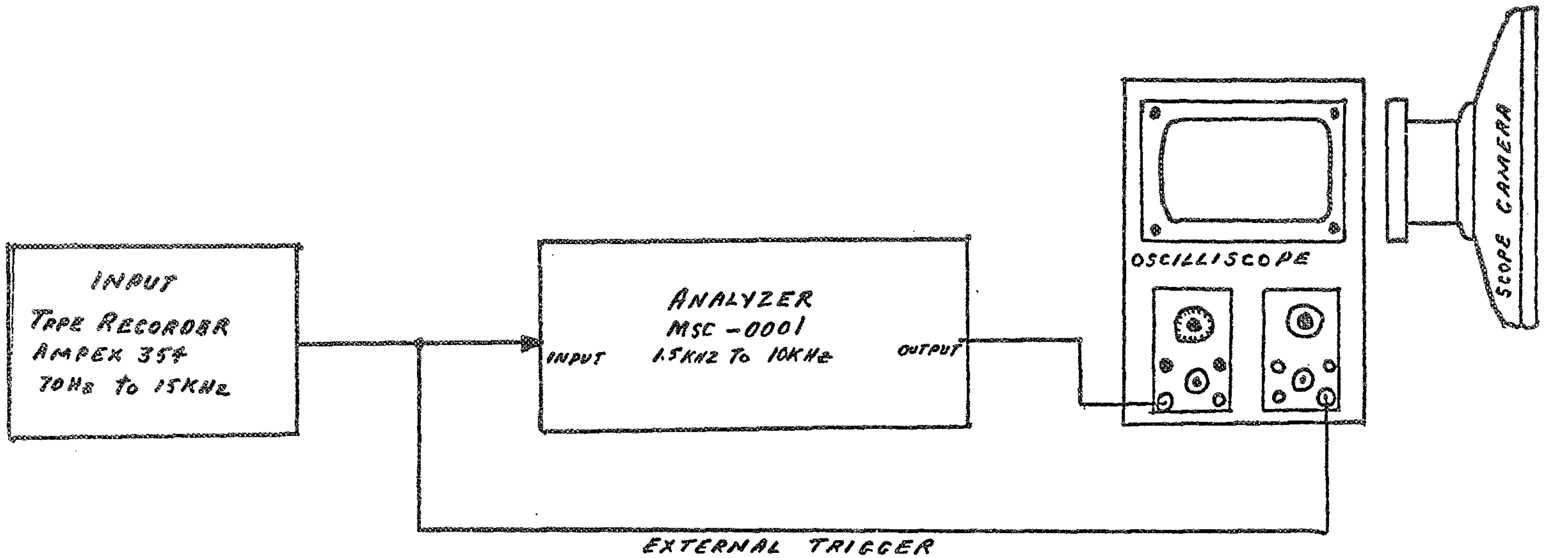


Fig. 4-3. Test Setup Used in Feasibility Study

a number of other methods tried, such as xy plotter, oscillograph, visicorder, etc.

Five different pictures were made of the analyzed speech samples and are distinguished by the following equations:

$$f(t) = s \qquad \text{Picture 1} \tag{4-1}$$

$$f(t) = \bar{s} \qquad \text{Picture 2} \tag{4-2}$$

$$f(t) = (s - \bar{s})^2 \qquad \text{Picture 3} \tag{4-3}$$

$$f(t) = (s - \bar{s})^3 \qquad \text{Picture 4} \tag{4-4}$$

$$f(t) = (s - \bar{s})^4 \qquad \text{Picture 5} \tag{4-5}$$

These five pictures represented by the five equations are: The signal (Eq. 4-1), the signal first, second, third, and fourth central statistical moments (Eqs. 4-2 through 4-5, respectively). Where the signal "s" is the reciprocal of the distance in time between the zero axis crossings of the input speech samples. The speech samples used, as stated earlier, are the spoken words "which" and "sunless." Other words in the word space were observed through the analyzer but no data was collected and stored; as a result, no data on these words appear in this report except for the word "speechless." At any rate, a set of five pictures is presented for each speech sample.

The primary procedure for proving the physical realizability of the analyzer was to establish some degree of correlation between the data obtained using the algorithm and the data obtained using the electronic model, namely, the real-time statistical time-series analyzer. Figures 4-4 and 4-5 represent the data obtained from the algorithm which are the first four central statistical moments of the signal. On the other hand, Figs. 4-6 through 4-11 present the same data obtained using the electronic model with the addition of the signal itself and the words "which" and "speechless." A comparison study was made of these data and the findings are as follows: A close observation of Sitton's algorithm data (Figs. 4-4 and 4-5), of the spoken word "sunless" shows that in the third moment, the transition from and to the phoneme/s/ at the beginning and end of the word are significantly emphasized. Likewise, in other moments, the first, second and fourth, an emphasis of the transition from and to /s/ is also present. However, these transition illustrations are not as pronounced as that of the third moment.

A similar observation of the data obtained on the spoken word "sunless" using the electronic model shows that an emphasis, represented by the third moment, of the transition to the final phoneme/s/ is pronounced and correlates with the like moment presented by the algorithm data. Traces of the same appear in the representation of the other three

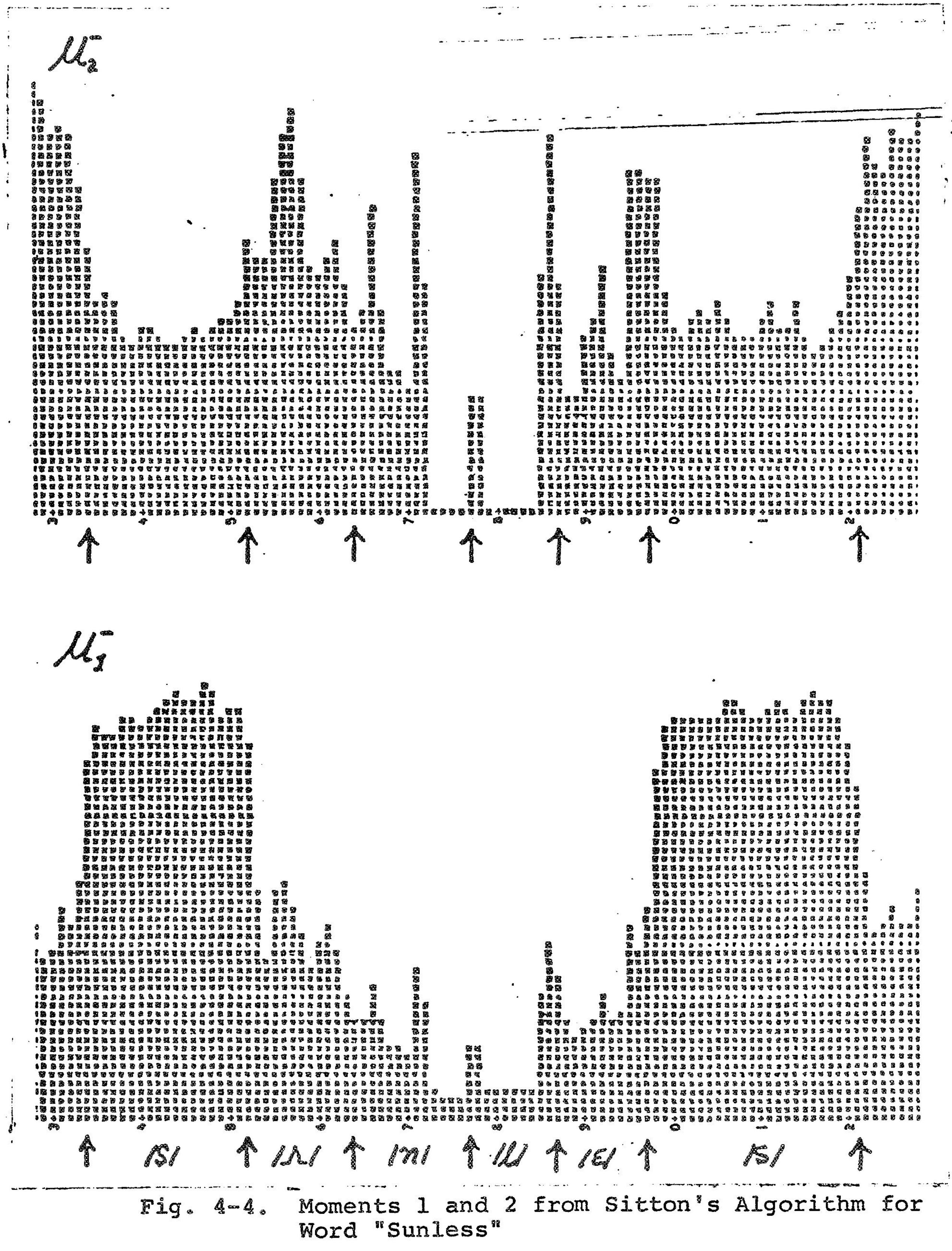


Fig. 4-4. Moments 1 and 2 from Sitton's Algorithm for Word "Sunless"

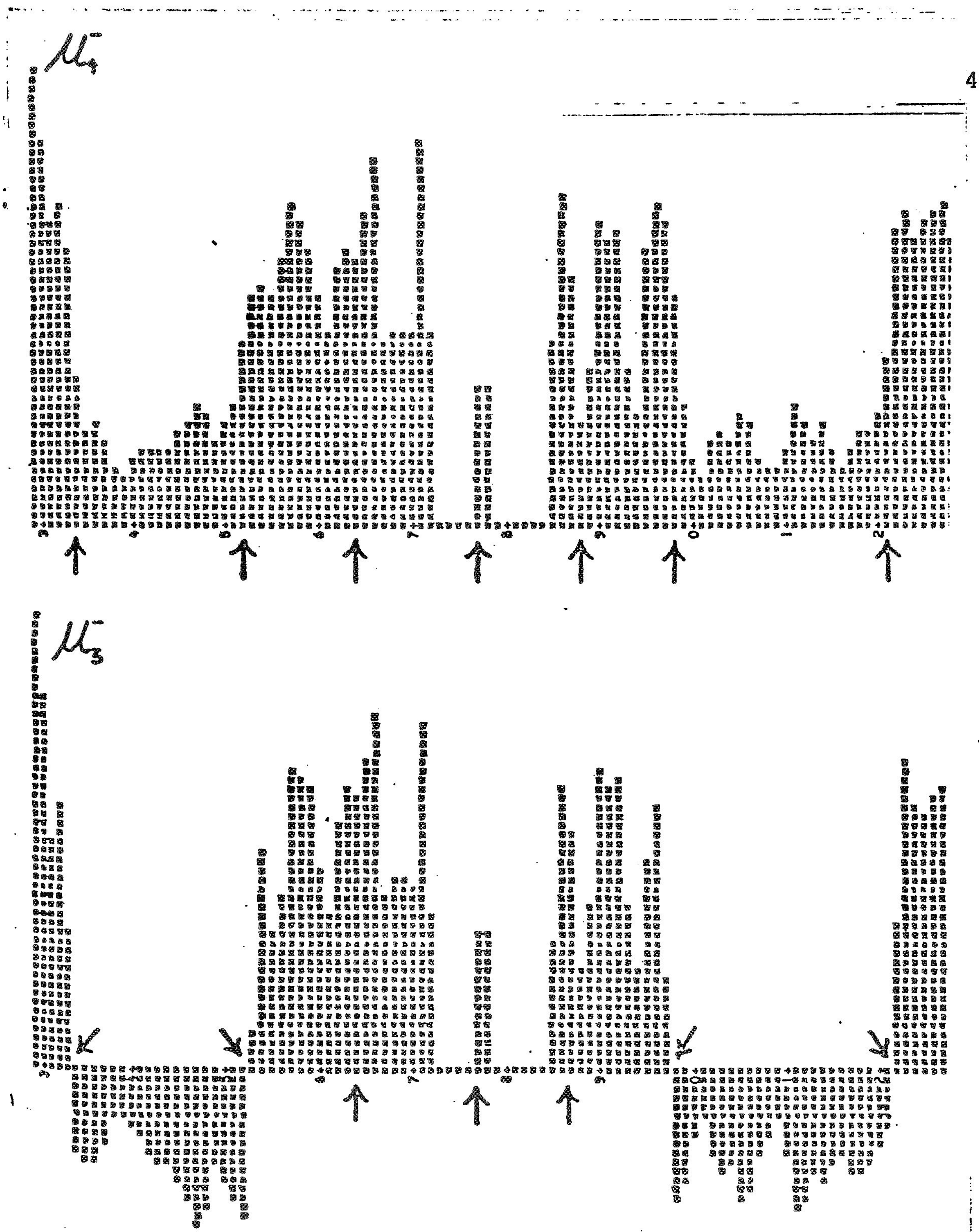
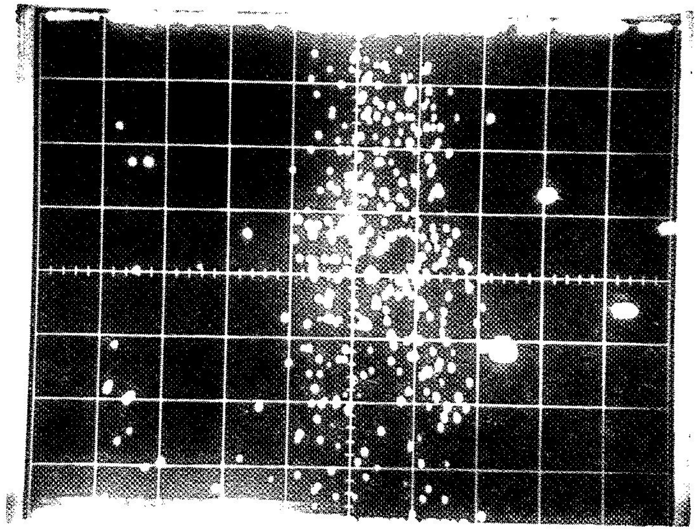


Fig. 4-5. Moments 3 and 4 from Sitton's Algorithm for Word "Sunless"

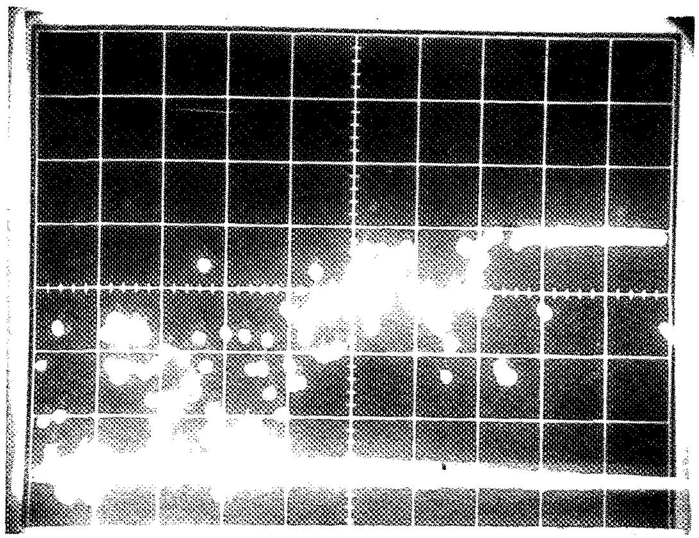


f(t) = S for the word "Sunless"

Scope settings 1 volt/division by 0.1 sec/division

Single Sweep

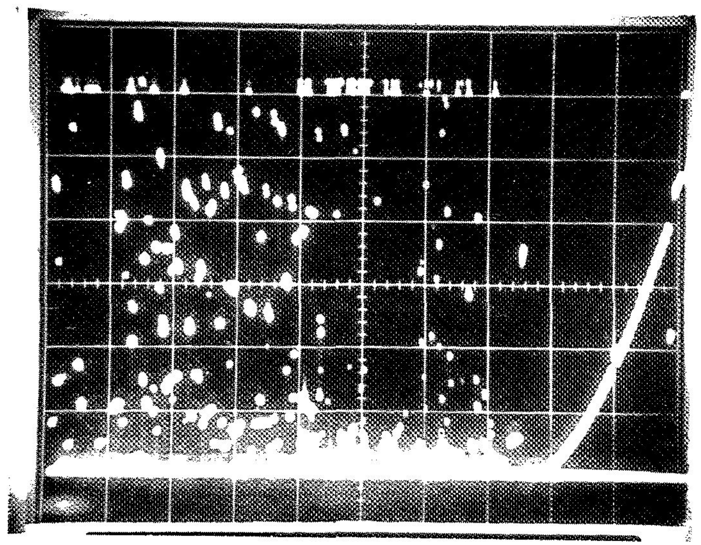
Fig. 4-6. Zero Crossing Distance Distribution "S" of Word "Sunless"



$f(t) = \overline{S}$ for the word "Sunless"

Scope Settings 1 volt/division by 0.1 sec/division

Single Sweep

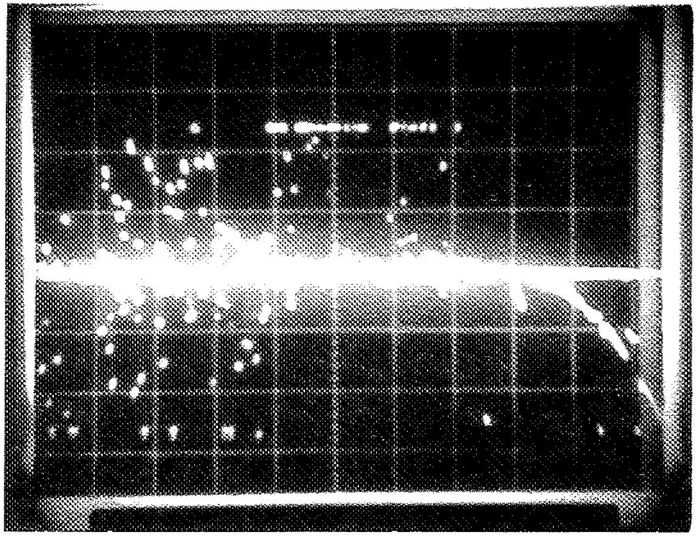


$f(t) = (S - \overline{S})^2$ for the word "Sunless"

Scope Settings 2 volts/division by 0.1 sec/division

Single Sweep

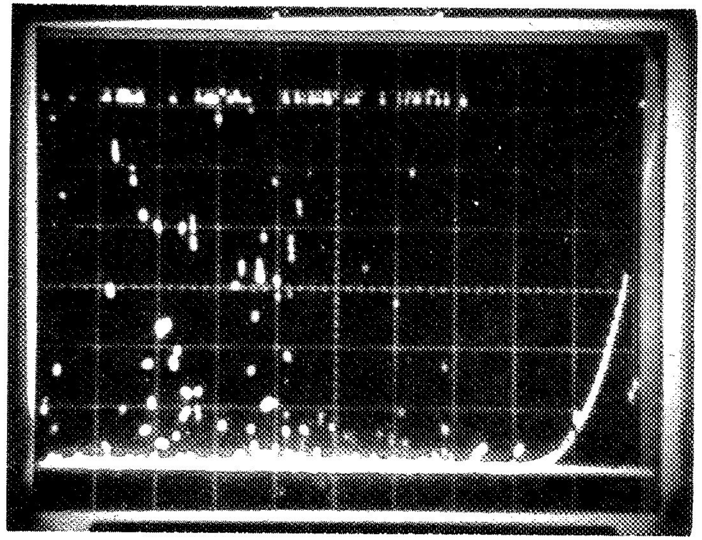
Fig. 4-7. 1st, 2nd, 3rd and 4th Order Moments of "S" for the Word "Sunless"



$f(t) = (S - \overline{S})^3$ for the word "Sunless"

Scope Settings 5 volts/div by 0.1 sec/div

Single Sweep

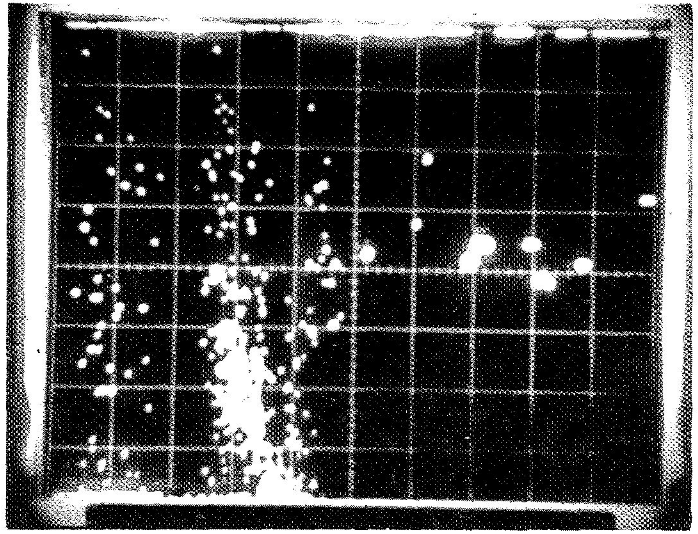


$f(t) = (S - \overline{S})^4$ for the word "Sunless"

Scope Settings 2 volts/div by 0.1 sec/div

Single Sweep

Fig. 4-7 (Cont'd)

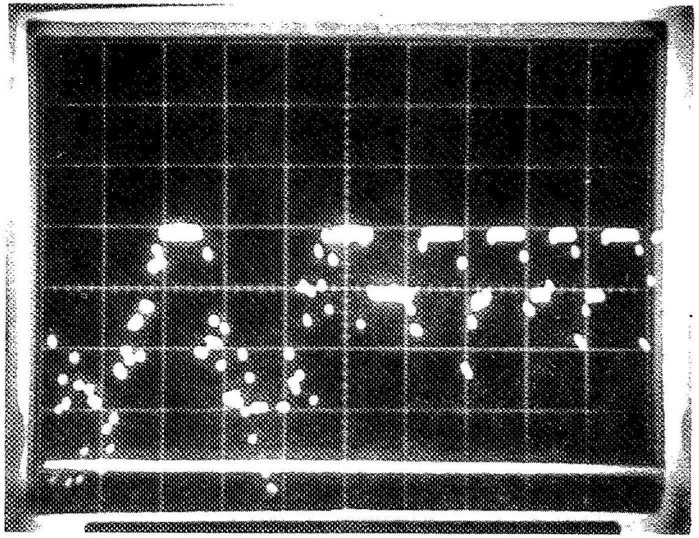


$f(t) = S$ for the word "Which"

Scope Settings 1 volt/div by 0.1 sec/div

Single Sweep

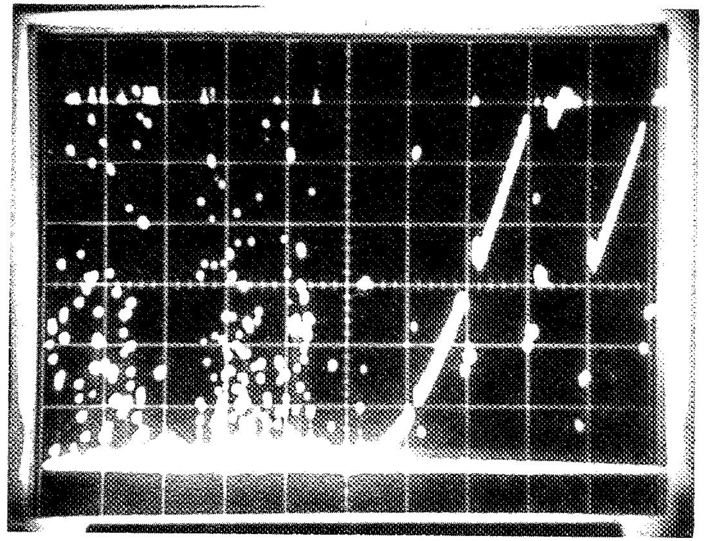
Fig. 4-8. Zero Crossing Distance Distribution "S" of Word "Which"



$f(t) = \overline{S}$ for the word "Which"

Scope Settings 1 volt/div by 0.1 sec/div

Single Sweep

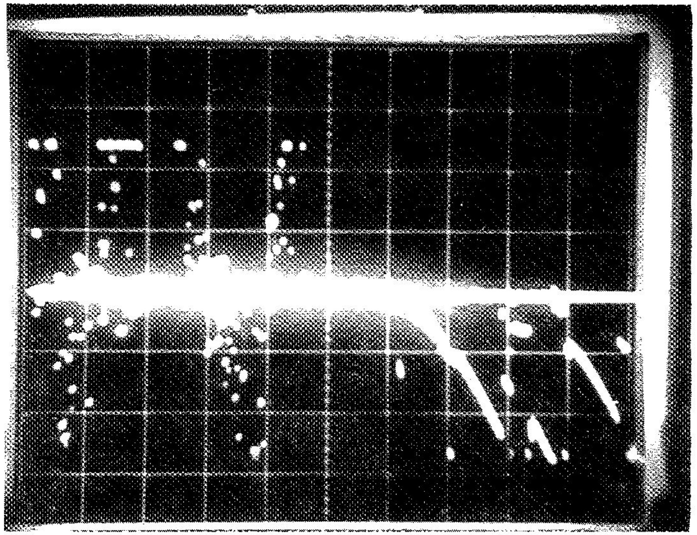


$f(t) = (S - \overline{S})^2$ for the word "Which"

Scope Settings 2 volts/div by 0.1 sec/div

Single Sweep

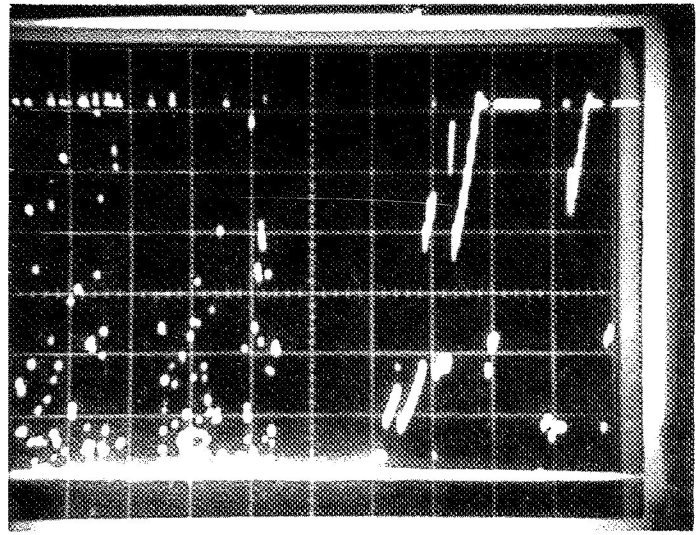
Fig. 4-9. 1st, 2nd, 3rd and 4th Order Moments of "S" for the Word "Which"



$f(t) = (S - \overline{S})^3$ for the word "Which"

Scope Settings 5 volt/div by 0.1 sec/div

Single Sweep

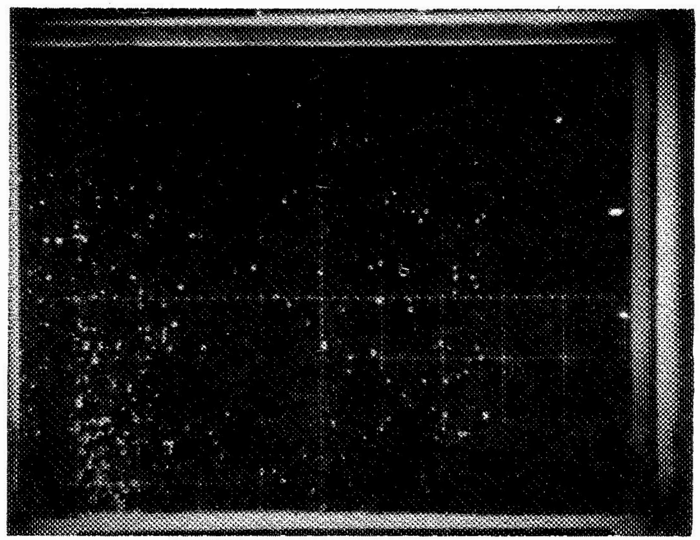


$f(t) = (S - \overline{S})^4$ for the word "Which"

Scope Settings 2 volts/div by 0.1 sec/div

Single Sweep

Fig. 4-9 (Cont'd)

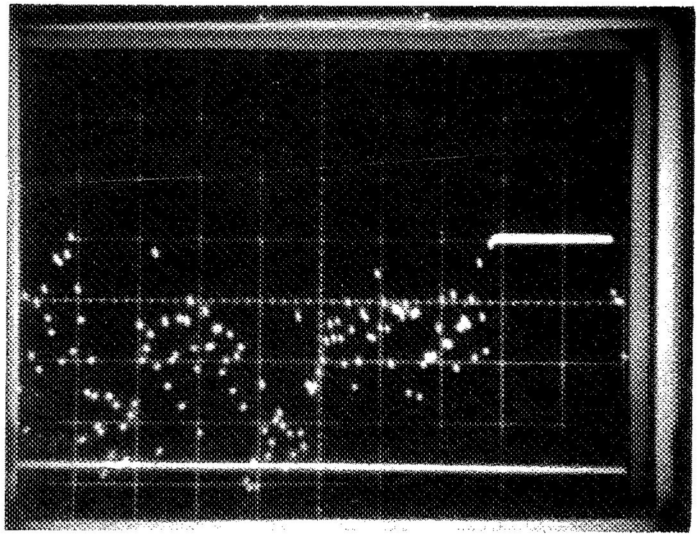


$f(t) = S$ for the word "Speechless"

Scope Settings 1 volt/div by 0.1 sec/div

Single Sweep

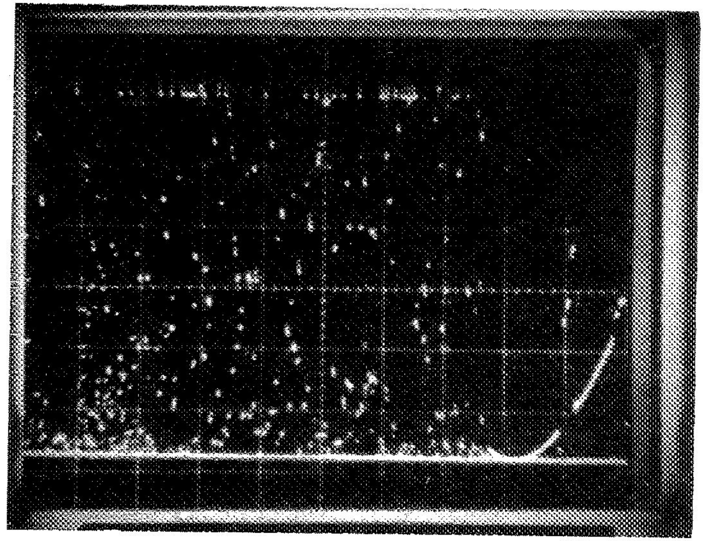
Fig. 4-10. Zero Crossing Distance Distribution "S" of Word "Speechless"



$f(t) = \overline{S}$ for the word "Speechless"

Scope Settings 1 volt/div by 0.1 sec/div

Single Sweep

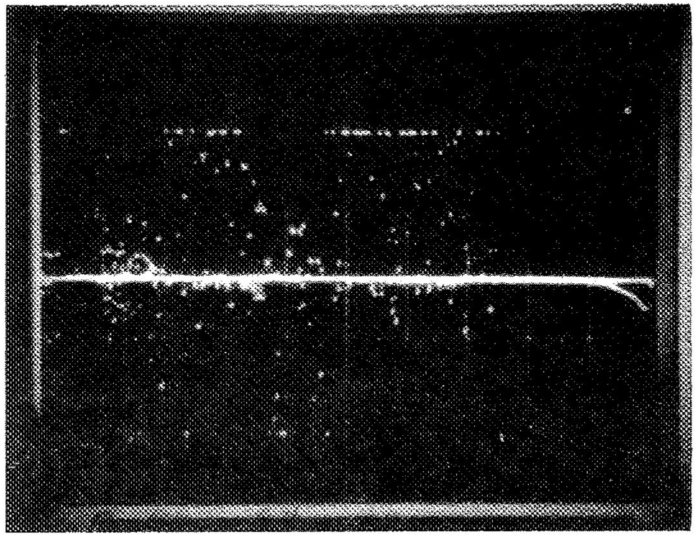


$f(t) = (S - \overline{S})^2$ for the word "Speechless"

Scope Settings 2 volts/div by 0.1 sec/div

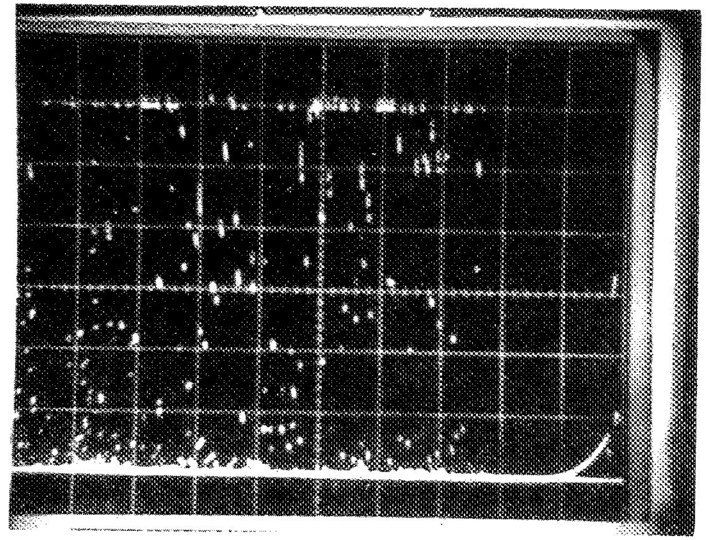
Single Sweep

Fig. 4-11. 1st, 2nd, 3rd, and 4th Order Moments of "S" for the Word "Speechless"



$f(t) = (S - \overline{S})^3$ for the word "Speechless"

Scope Settings 5 volts/div by 0.1 sec/div



$f(t) = (S - \overline{S})^4$ for the word "Speechless"

Scope Settings 2 volts/div by 0.1 sec/div

Fig. 4-11 (Cont'd)

moments generated from the electronic model. However, these occurrences are not as pronounced as the algorithm data; as a result, a solid comparison is not possible. There are a number of reasons why this comparison is not solid. These are discussed later in this report.

It is obvious also that the transition to the phoneme/s/ illustrated by the data from the electronic model is not an exact replica of the algorithm data. There are reasons why this is true and these reasons likewise are discussed later in this report. Nevertheless, the correlation illustrated by the third moment of the electronic model data and the third moment of the algorithm data shows that the electronic model can provide similar results to the algorithm thus indicating that the design of the electronic model is feasible.

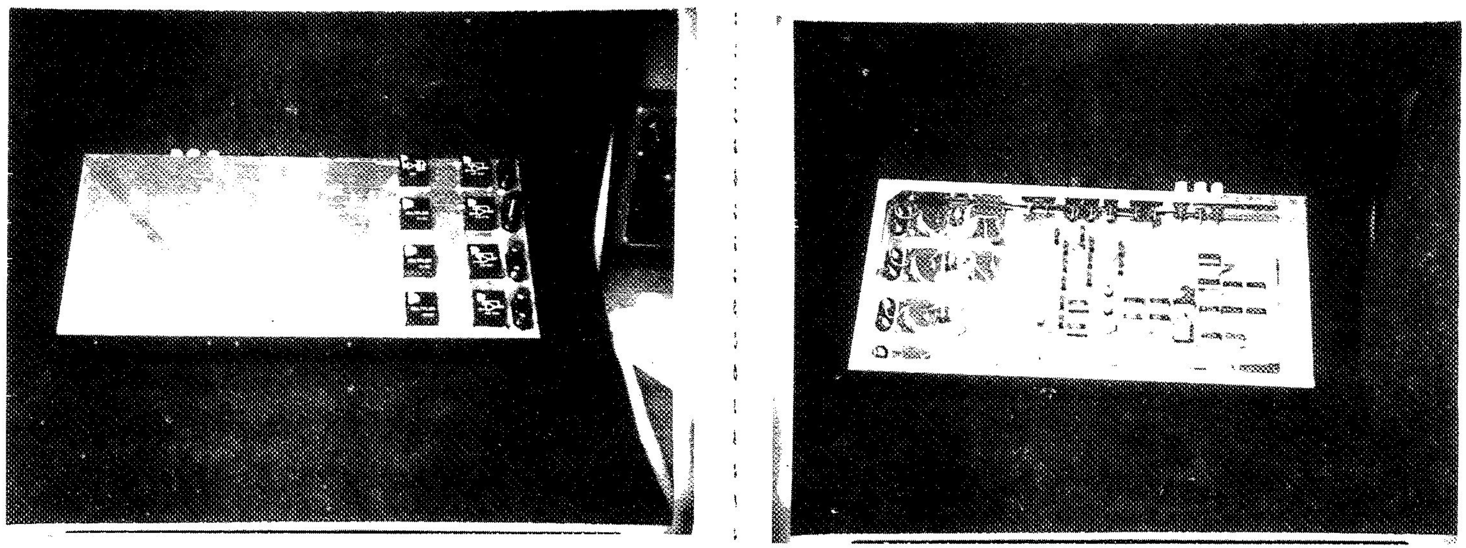
To further substantiate the feasibility of the electronic model, another comparison study was made; this study equated the spoken words "sunless" and "which" with the word "speechless." From an observation of the third moments of the words "sunless" and "which" phoneme transitions are indicated going to the /s/ of "sunless" and to the /ch/ of "which." Moreover, both of these phonemes /s/ and /ch/ appear in the word "speechless." So an observation of the third moment of the word "speechless: emphasizes the transition to the phoneme /ch/ and from the phoneme /ch/ to /s/.

These transitions are not quite as apparent on one observation as one would like them to be, but continued review of the data over a series of observations points out more clearly the transition. This comparison study of the words "sunless" and "which" with the word "speechless" along with the comparison of the data on the word "sunless" obtained from the electronic model with that obtained for the algorithm data establishes the feasibility of design and that the model is physically realizable. As stated, however, there are reasons why the complete set of data taken on the word "sunless" did not correlate in an exact sense to the data obtained by Sitton's algorithm. These are all covered in the next chapter.

CHAPTER V

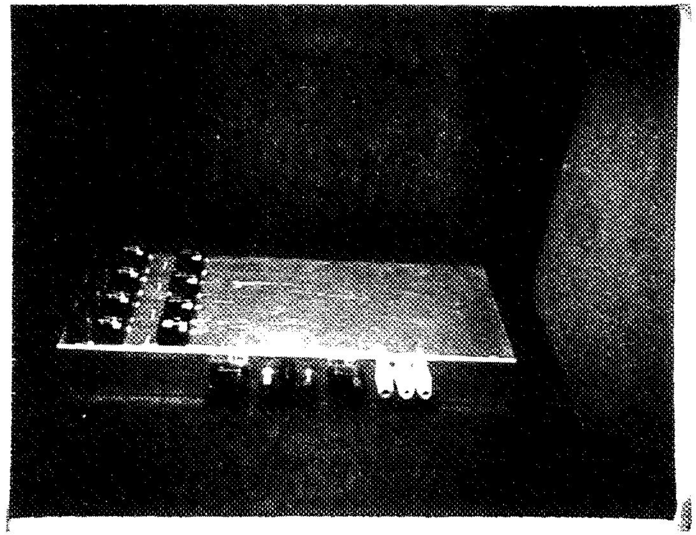
EXPERIMENTAL RESULTS

In general the real-time statistical time-series analyzer design reported herein represents a new tool for use in conducting speech research and implementing the benefits of real-time statistical time-series analysis into speech processing equipment. Prior to the research effort at Rice University, only two other works are known to have dealt with speech analysis beyond the first central statistical moment. As a result, the study efforts are relatively new. To this author's knowledge, no known research other than the present has made any attempt to design and construct a statistical time series analyzer which could be used to study the statistical moments of speech beyond the first moments and yet be far less complicated than a computer. So the physical implementation of the analyzer designed herein is considered to be somewhat of a first. One important application for this analyzer in speech research is where a great deal of statistical data is needed about the speech signal and computer usage is at a premium. Also, the device may be easily used as a laboratory instrument which can easily be stored because of its small size. Figures 5-1(a), (b) and (c) show three photographs of the device.



(a) Front and Top View of Analyzer (b) Bottom View of Analyzer

Fig. 5-1. Photograph of Prototype Analyzer



(c) Rear and Top View of Analyzer

Fig. 5-1 (Cont'd)

Like any prototype design model, there are always problems that need to be overcome. This analyzer design is no exception. Furthermore, it is this fact that accounts for the lack of exactness in the data used to correlate with the algorithm data. In other words, the comparison of these data in Chapter IV showed that the data from the electronic model is not an exact duplicate of the algorithm data, but sufficient likeness does exist in order to establish that the design concept of the model is feasible and physically realizable.

There are, however, very tangible reasons why the two sets of data do not agree in an exact sense. The first is that the two methods of data display are not the same because they are not the same only and indication of likeness can be established.

The other reason the two sets of data do not agree in an exact sense is found in the physical implementation of the design itself. In a general sense, the reason is found in a combination of problems in three areas: (1) The process of establishing a true zero reference and zero crossing detection, (2) Frequency to amplitude conversion, and, finally, (3) Development of a moment generating function.

The problem with establishing the true zero reference and zero crossing detection was that of maintaining noise immunity and proper zero crossing detection. In the physical design the zero reference was established using a differentiator with a time constant of one nanosecond. This differentiator acted as a high pass filter thus emphasizing high frequency noise as well as signal, which ultimately affected the zero crossing detection in the form of stability errors. To minimize the effect of the noise produced by the input filter, the zero crossing detector was not allowed to operate exactly at zero. Instead, the input signal was amplified many times with a high gain amplifier, then clipped and amplified again. This high gain amplification allows the detection of the zero crossing to occur at a level, offset from zero with minimum error. The rule used to determine the minimum error was the measurement of the time interval between the point of the offset level and the actual zero crossing point. The ratio of this time interval to the minimum distance in time between successive zero crossings determined the error. The error was designed to be less than 1% for a minimum time interval of 100 microseconds. In other words, the time interval between the offset level and the actual zero crossing was maintained at one microsecond. The technique offered a high degree of noise immunity but instability in the axis crossing itself.

This negative effect of the zero crossing detection was not apparent in the detector itself but it caused problems in the frequency to amplitude converter. Proper operation of the frequency to amplitude converter depends heavily on the accuracy of the zero crossing detection. This is recognized in that the occurrence of successive zero crossings were used to derive the control function for the frequency to amplitude converter. For error free conversion, the converter requires a high degree of stability from the zero crossing detector. Because of the limitation on noise immunity, this stability could not be maximized; so, apparent conversion error occured from time to time. These errors were represented in the form of premature and post transfer of the measurement of the zero crossing distance from an 8 bit digital counter to an 8 bit digital-to-analog converter. This kind of error is not a function of the converter but the stability and accuracy of the zero crossing detector.

The problem that exists with the frequency to amplitude converter is one of using the proper digital-to-analog (D/A) converter. At the time the design was physically implemented, the only suitable D/A converter available was a bipolar one. This means that, besides the fact that the output amplitude ranges from a negative minimum to a positive maximum, the minimum time interval between successive zero crossing corresponded to an amplitude nearly equal to the positive maximum voltage of the D/A converter. On the other hand, the maximum time interval between successive zero crossings corresponded to a D/A voltage level nearly equal to the negative minimum. With the D/A voltage ranging from a $-V_{min}$ to a $+ V_{max}$, the weight of the maximum time interval between successive zero crossings was essentially equal to that of the minimum time interval. The desire, however, was to weight the minimum time interval more heavily than the maximum and allowing the intervals in between to be weighted proportionally. So actually a unipolar D/A converter would be more proper to use.

To compensate for the equal weighting of the minimum and maximum time interval, a diode was used to cancel out the D/A voltage levels from near $V = 0$ to $V = -V_{min}$. This compensation offered more closely the desired weighting of the time intervals. Also the signal analysis, as a result of the compensation, was limited to time intervals of 0.5 milliseconds and less. Because the compensation used limited the D/A voltage only to near $V = 0$, errors appeared in the data from time intervals producing D/A output voltages between $V = 0$, $V = -V$ near zero. These errors were not extremely significant but were found to be significant enough to affect the exactness of the analyzer data compared to the algorithm data. This problem can easily be overcome with the use of the proper unipolar D/A converter.

Finally, the problem associated with the development of the moment generating function is not really known to have affected the exactness of the data comparison. However, when coupled with the other two, the affect could be significant. At any rate the problem involves the physical derivation of the function

$$f(t) = (s - \bar{s}) \tag{5-1}$$

where s = the signal

$\bar{s}$ = the average of the signal

Basically, the components "s" and "$\bar{s}$" occur at different time intervals. In other words, the signal "s" is averaged over a time period of 3.75 milliseconds, and then held for another 3.5 milliseconds more while being compared with "s" via a difference network. A more suitable method might be to store "s" during the 3.75 millisecond period when "$\bar{s}$" is being derived. Then, during the 3.5 millisecond period, "$\bar{s}$" could be compared in the difference network with the stored value of "s" rather than the alternate value of "s." Hoever, using the alternate value of "s" is a valid technique because "s" the signal has been assumed to be stationary over a period not to exceed 15 milliseconds.

As illustrated, these three problems can be solved and, if solved, should without any doubt improve greatly the exactness of the analyzer data. No other problems were apparent during both the design and study of the analyzer but to conclude this chapter, a brief discussion of the perspective of the kind of statistical analysis the physical analyzer performs would be in order.

As a final note to the results which have been presented, it should be pointed out that the text of Chapter II implies that the analyzer performs a continuous statistical analysis. However, a more accurate description of the analysis is to say that it is piece-wise continuous. In other words, each statistical moment presented is made up of a series of 3.75 millisecond samples. Mathematically these moments may be described by the following five equations:

$$S = \sum_{i = Tsp}^{i = nTsp} S_i \tag{5-2}$$

where $n = \frac{Tw}{2Tsp}$

Tw = word or speech duration time

Tsp = sample period

S_i = value of s during period Tsp

$$\overline{S} = \sum_{i = Tsp}^{i = nTsp} \overline{S}_i \tag{5-3}$$

$$(S - \overline{S})^2 = \sum_{i = Tsp}^{i = nTsp} (S_i - \overline{S}_i)^2 \tag{5-4}$$

$$(S - \overline{S})^3 = \sum_{i = Tsp}^{i = nTsp} (S_i - \overline{S}_i)^3 \tag{5-5}$$

$$(S - \overline{S})^4 = \sum_{i = Tsp}^{i = nTsp} (S_i - \overline{S}_i)^4 \qquad (5\text{-}6)$$

Another point that should be made about the analyzer is that two different type test signals were used to insure that the system was operating properly prior to any test. These signals were a sine wave and a triangular wave. The sine wave offered system checkout from the analyzer input through the output of the analyzer difference network. The triangular wave, on the other hand, offered system checkout from the input of the analyzer analog multiplier arrangement through the system outputs. The main idea of using these two signals is that the sine wave has well known characteristics and from these characteristics the system average and difference networks can be checked. Moreover, the triangular wave have well known characteristics for checking the system analog multiplier arrangement. By using these two signals to insure the proper operation of the average and difference network along with the analog multiplier arrangement assured the proper operation of the entire system. The three functions are the most critical to the proper operation of the whole system.

CHAPTER VI

CONCLUSIONS

The research effort herein is considered to be a success. The effort achieved its primary goal which was to advance the development of a new speech segmentation concept such that the concept can be more practically utilized. At the start of this research effort, the segmentation concept was contained within a computer algorithm; as a result, practical application of the concept required the use of the associated computer or its equivalent. Now, this concept may be studied and utilized without the need for a computer.

Because the need for a computer has been eliminated, speech researchers, both pure and applied, may explore much further the promising aspect that the concept has toward solving speech recognition problems and speech processing equipment development. The segmentation concept is known to have great potential in solving speech recognition problems because today all speech researchers, by and large, agree that the problem of speech recognition is a runner-up to speech segmentation. In addition, the increased accessibility to the segmentation concept opens the doorway to many new ideas for speech processing equipment with applications in voice communication systems. Moreover, the fact that the design model itself uses simple design techniques and small circuit components, alone, makes the concept more conducive for use in equipment development. Finally, the real factor that makes this research effort a success is that the method chosen to advance the development of the segmentation concept has been proven to be both feasible and physically realizable.

BIBLIOGRAPHY

American Standards Association. "American Standard-Graphic Symbols for Logic Diagrams," ASA, Vol. No. Y32.14, UDC 621.3.003.6, 1962.

Blackman, R. B. and J. W. Tukey. *The Measurement of Power Spectra*. New York: Dover Publications, Inc., 1958.

Bricker, Peter D. "Technique for Objective Measurement of Speech Levels," *J. Acous. Soc. Am.*, Vol. 38, September, 1965, pp. 361-62.

Capon, Jack. "On the Properties of an Active Time-Variable Network: The Coherent Memory Filter," *Proc. of the Symp. on Active Networks and Feedback Systems*, Polytechnic Institute of Brooklyn, New York, 19-21 April 1960, pp. 561-581.

Cooper, F. S. "Spectrum Analysis," *J. Acous. Soc. Am.*, Vol. 22, 1950, pp. 761-62.

Davenport, Wilbur B., Jr. "An Experimental Study of Speech-Wave Probability Distributions," *J. Acous. Soc. Am.*, Vol. 24, July, 1952, pp. 390-99.

Dunkel'skiy, N. I. "Principles of Segmentation of the Speech Stream," *U. S. Dept. of Commerce Joint Publications Research Service*, Acc. No. N66-12957, November, 1965.

Dunn, H. K. and H. L. Barney. "Artificial Speech in Phonetics and Communications," *J. Speech and Hearing Research*, Vol. 1, 1958, pp. 23-39.

Dunn, H. K. and S. P. White. "Statistical Measurements on Conversational Speech," *J. Acous. Soc. Am.*, Vol. 11, 1940, pp. 278-88.

Elphick, Michael, E. L. Renschler, Thomas Cate and Howard Handler. "Designer's Guide: Analog Multipliers," *EEE*, Vol. 14, May, 1969, pp. 51-77.

Essignmann, Martin, Sze-Hou-Chang and George E. Phil. "Representations of Speech Sounds and Some of Their Statistical Properties," *Pro. of IRE*, Vol. 39, February, 1951, pp. 147-53.

Fano, R. M. "The Information Theory Point of View in Speech Communication," J. Acous. Soc. Am., Vol. 22, November, 1950, pp. 691-97.

Fano, R. M. "Short-Time Autocorrelation Functions and Power Spectra," J. Acous. Soc. Am., Vol. 22, September, 1950, pp. 546-50.

Fant, G. and K. N. Stevens. "Systems for Speech Compression," Fortschr, Hochfrequvenztenchn, Vol. 5, 1960, pp. 229-62.

Flanagan, J. S. Speech Analysis Synthesis and Perception. New York: Academic Press, 1965.

Gill, J. S. "A Versatile Method for Short-Term Spectrum Analysis in 'Real-Time'," Nature, No. 4759, 14 January 1961.

Gold, Bernard. Digital Processing of Signals. New York: McGraw-Hill, 1969.

Golden, Roger M. "Digital Computer Simulation of a Samples-Data Voice Excited Vocider," J. Acous. Soc. Am., Vol. 35, September, 1963, pp. 1358-66.

Griffiths, John David. "Speech Communications from an Information Theory Viewpoint," Syracuse University, Doctoral Thesis, January, 1965.

Kintner, Paul M. Electronic Digital Techniques. New York: McGraw-Hill, 1968.

Kock, Winston E. "Speech Communication Systems," Proc. of IRE, Vol. 50, May, 1962, pp. 769-76.

Leybman, Uy. A. and V. N. Sobolev. "Analog-to-Digital Converter for Feeding a Speech Signal into a Computer," Telecommunications and Radio Engr., Vol. 17, August, 1963, pp. 42-50.

Leytes, R. D. "Mathematical Modeling of Speech Signal Transmission Systems," Telecommunications and Radio Engr., Vol. 17, August, 1963, pp. 32-42.

Philbrick/Nexus Research. Applications Manual for Operational Amplifiers for Modeling, Measuring, Manipulating and Much Else. Philbrick/Nexus Research, a Teledyne Company, Dedham, Mass., August, 1969.

Myasnikova, Ye N. _Objective Recognition of Speech Sounds_. U. S. Dept. of Commerce Joint Publications Research Service, Acc. No. N68-1401, January, 1968.

Reddy, D. R. "Computer Recognition of Connected Speech," _J. Acous. Soc. Am._, Vol. 42, 1967, pp. 329-47.

Schroeder, M. R. and B. S. Atal. "Generalized Short-Time Power Spectra and Autocorrelation Functions," _J. Acous. Soc. Am._, Vol. 34, November, 1962, pp. 1679-83.

Sitton, G. A. _Acoustic Segmentation of Speech_. Rice Univ. Rept. No. ORO-2572-18, AEC Contract AT-(40-1)-2572, Thesis, June, 1969.

Tuffs, D. "Some Results in Automatic Speech Segmentation Using Wide-Band Filtering and Subtraction," _J. Acous. Soc. Am._, Vol. 38, September, 1965, pp. 362-65.

Velichkin, A. I. "Correlation Function and Spectral Density of Quantized Speech," _Soviet Physic-Acous_, Vol. 9, September, 1963, pp. 10-14.

Weiss, Mark R. and Cyril M. Harris. "Computer Technique for High-Speed Extraction of Speech Parameters," _J. Acous. Soc. Am._, Vol. 35, February, 1963, pp. 207-14.

APPENDIX A

ANALYZER THEORY OF OPERATION AND SPECIAL NOTES

In general, the theory of operation of the real-time statistical time-series analyzer is quite simple and straight-forward. First of all, the input signal is converted to a signal which possesses only zero crossing information. Then this zero crossing information is converted into proportional amplitude information by measuring the time interval between successive zero crossings. These amplitudes which represent the time interval between successive zero crossings are averaged over a period of 3.75 milliseconds. Then, during the following 3.50 milliseconds, the averaged value is simultaneously held and subtracted from the amplitude values which occurs at the average circuit input during the same 3.50 millisecond period.

During the remaining .25 milliseconds, recognizing a total of 3.75 milliseconds per period, the averager is reset and is prepared to accept another 3.75 milliseconds of amplitude values. Meanwhile, the difference between the averaged amplitude quantities and the actual amplitude quantities occuring at the averagecircuit input, are squared, cubed, and raised to the fourth power simultaneously.

Finally, the averaged value of the amplitude along with the square, the cube, and the fourth power of the amplitude differences represent a 3.50 millisecond sample displaying the first four central statistical moments of the distance between successive zero crossings of the input signal.

The process as described above assumes that the input signal is human speech. However, the analyzer may be used to perform the same kind of analysis on any input signal. This use, if desired, may require some minor changes to the analyzer depending on the nature of the signal to be analyzed. Such changes would probably occur in the analysis sample time and the frequency to amplitude converter. In other words, the analysis sample time of 3.50 milliseconds may not meet the stationarity requirements of another signal. However, all of the required changes are a direct function of the nature of the signal to be analyzed.

APPENDIX B

ANALYZER PARTS LIST

1. Input Amplifier, Differentiator and Zero Crossing Detector

Operational Amplifiers	- Seven μ741 Fairchild
Digital Integrated Circuit	- One SN7400N TI
Diodes	- Five 1N457 Fairchild
Resistors (Fixed)	- Three 1K $\pm$10%
	- One 4K $\pm$10%
	- Two 10K $\pm$10%
	- Three 150K $\pm$10%
Resistors (Variable)	- One 20K $\pm$10%
Capacitors	- One 100 pf and one 10 pf

2. Frequency to Amplitude Converter

Operational Amplifiers	- None
Digital Integrated Circuits TTL	- Three SN 7400N TI
	- One SN 7410A TI
One Diode	- One SN 7430N TI
	- One SN 7473N TI
	- Two SN 7475N TI
	- Two SN 7493N TI

Special Digital to Analog Converter - One Beckman 845-B10

3. Integration, Hold, and Difference Network

Operational Amplifier - Six μ741

Digital Integrated Circuits TTL	- One SN 7400N
	- One SN 7473N
Analog Gates	- Four CAG13-6952 Sicloconix
Diodes	- None
Resistors (Fixed)	- One 1K
	- One 3.75K
	- One 10K
	- Four 12K
	- One 120K
Resistors (Variable)	- One 50K
Capacitors	- One 1μf

4. Moment Generating Circuit

Operational Amplifiers	- Four GPS-F0201
	- One GPS-801
Digital Integrate Circuits	- None
Diodes	- None
Resistors	- None
Capacitors	- None
Multipliers	- Three GPS-4030

5. Special Circuit Timing

Operational Amplifiers	- None
Digital Integrated Circuits	- One SN 7400N TI
	- One MC 851P Fairchild
	- Four SN 7493N

Diodes	- None
Resistors (Fixed)	- One 10K
Resistors (Variable)	- One 20K
Capacitor	- One 390 pf
Crystal	- One 1024 KHz