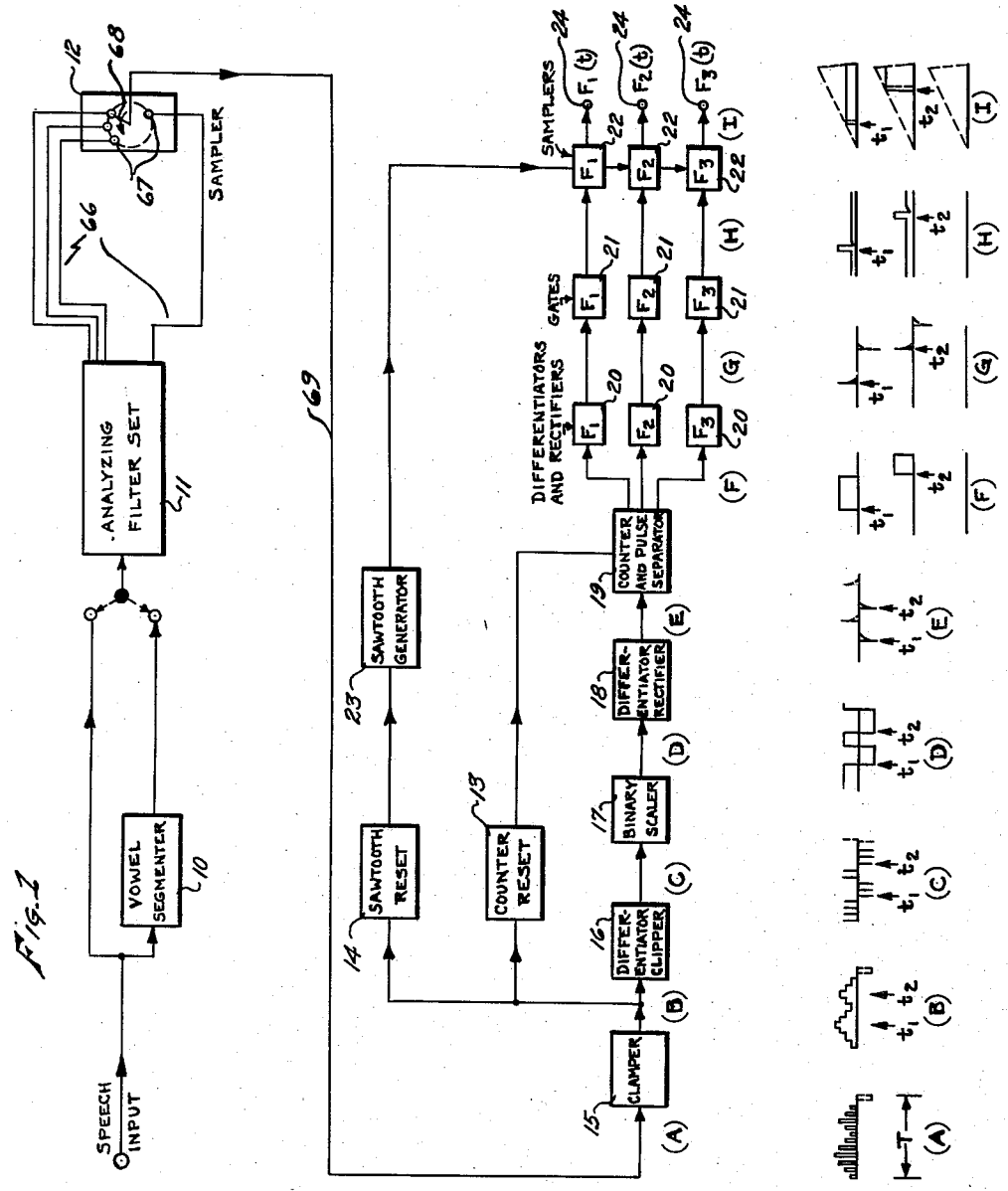
June 27, 1961 J. L. FLANAGAN 2,990,453

AUTOMATIC SPECTRUM ANALYZER

Filed Dec. 6, 1955 8 Sheets-Sheet 1



INVENTOR.
JAMES L. FLANAGAN

BY Wade Koontz and Martin J. Finnegan
ATTORNEYS

June 27, 1961 J. L. FLANAGAN 2,990,453

AUTOMATIC SPECTRUM ANALYZER

Filed Dec. 6, 1955 8 Sheets-Sheet 2

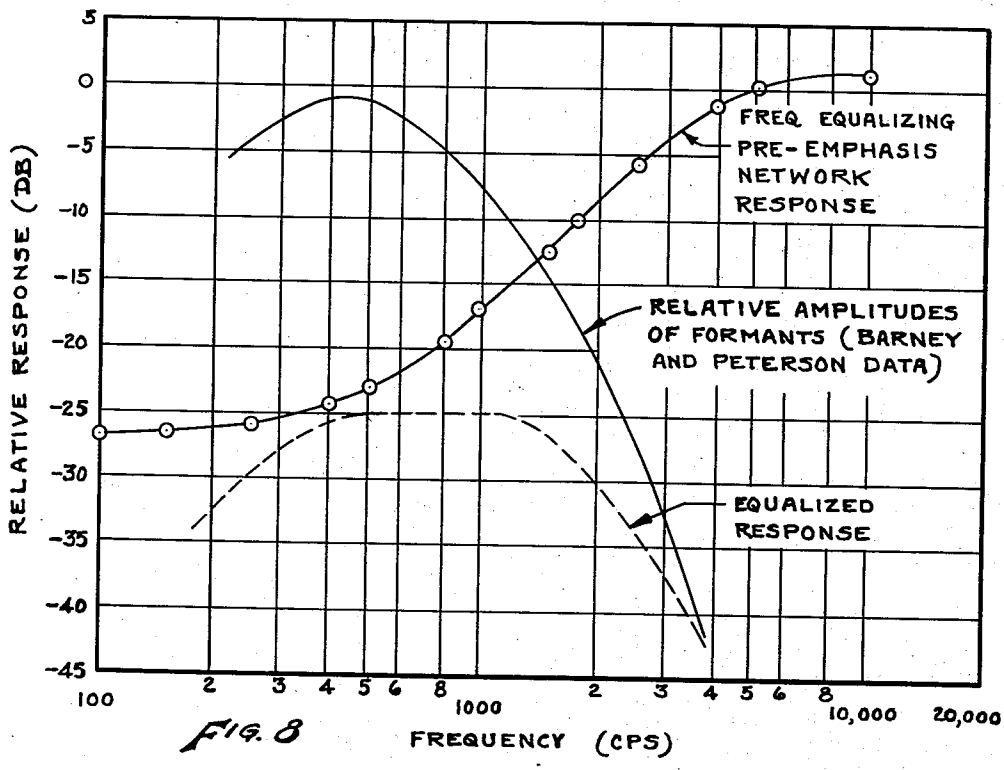


FIG. 8

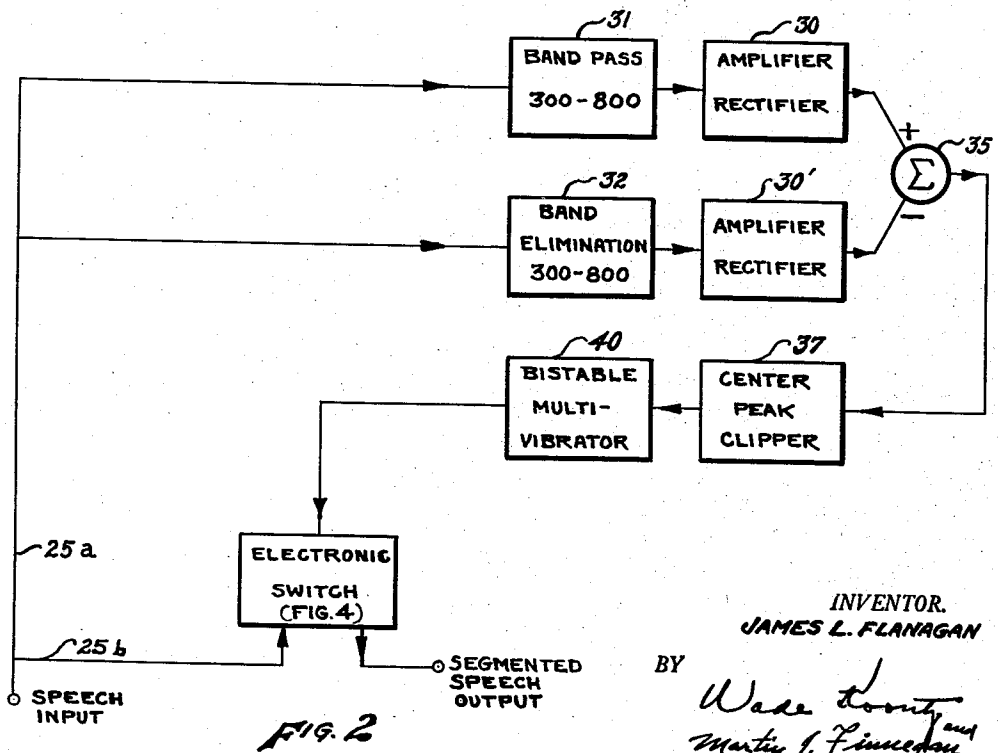


FIG. 2

INVENTOR.
JAMES L. FLANAGAN

BY Wade Doorty and Martin J. Finnegan
ATTORNEYS

June 27, 1961 J. L. FLANAGAN 2,990,453

AUTOMATIC SPECTRUM ANALYZER

Filed Dec. 6, 1955 8 Sheets-Sheet 3

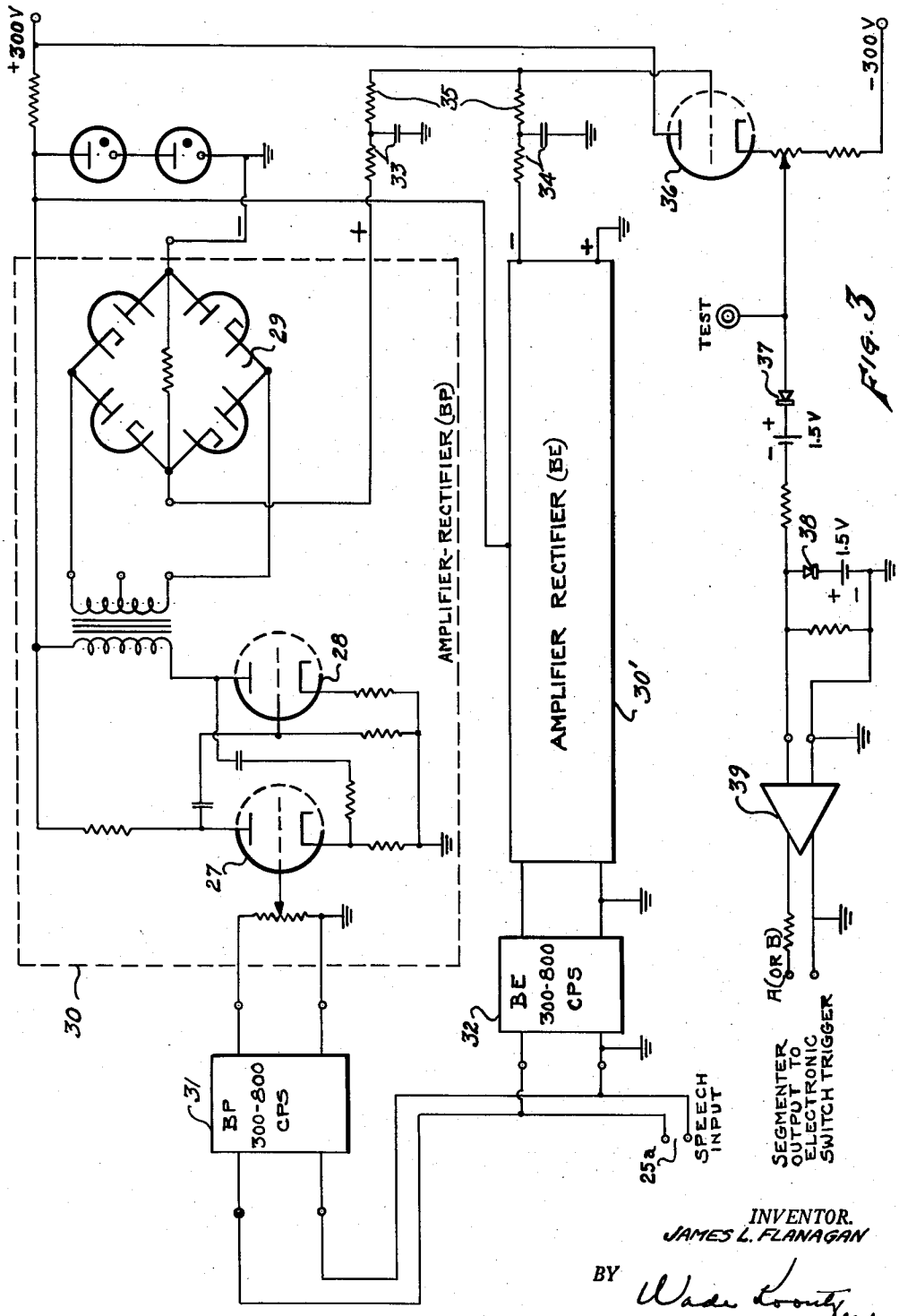


FIG. 3

INVENTOR.
JAMES L. FLANAGAN

BY Wade Kountz and Martin J. Finnegan
ATTORNEYS

June 27, 1961 J. L. FLANAGAN 2,990,453

AUTOMATIC SPECTRUM ANALYZER

Filed Dec. 6, 1955 8 Sheets-Sheet 4

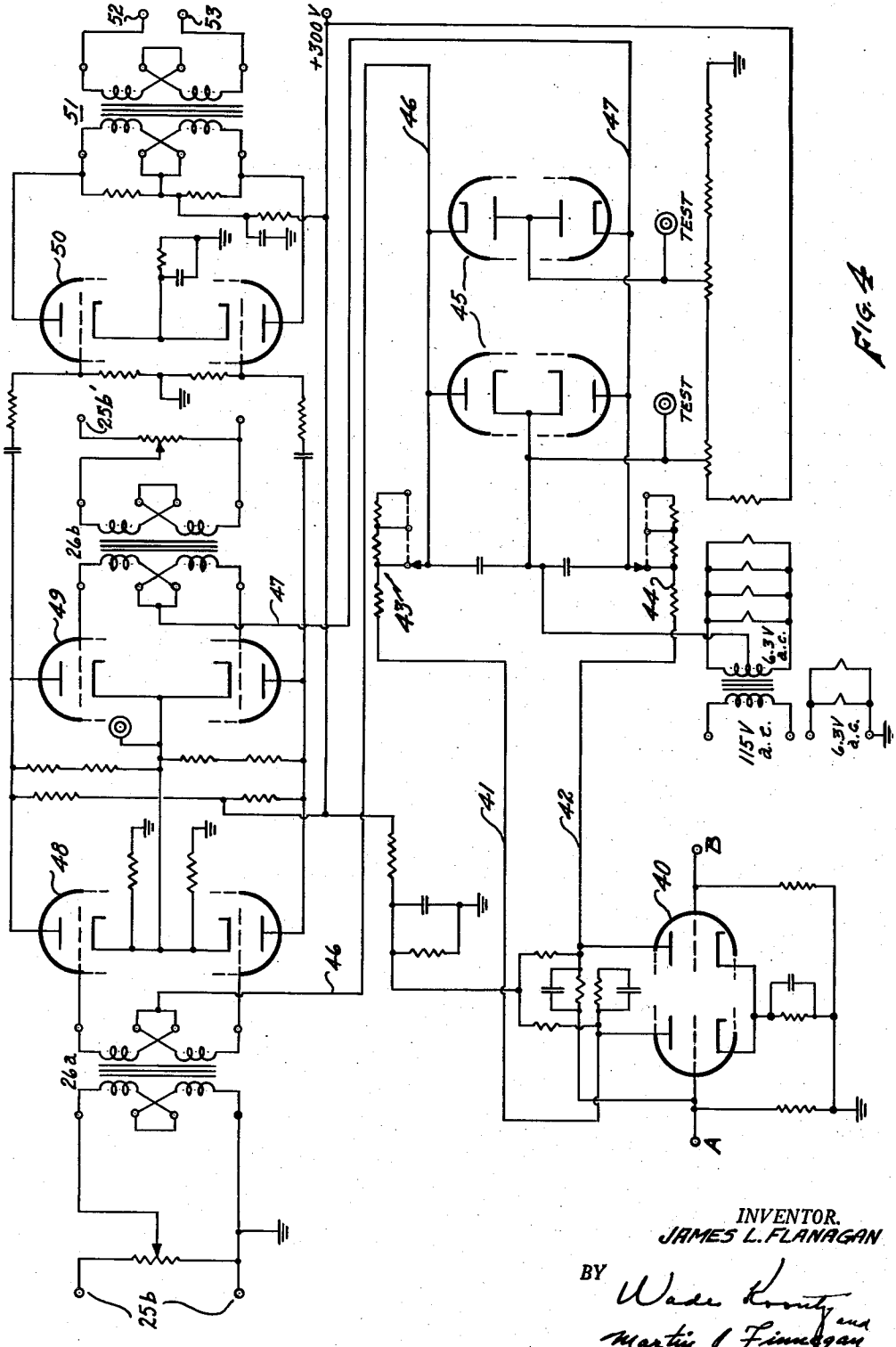


FIG. 4

INVENTOR.
JAMES L. FLANAGAN

BY Wade Kenty and Martin J. Finnegan
ATTORNEYS

June 27, 1961 J. L. FLANAGAN 2,990,453

AUTOMATIC SPECTRUM ANALYZER

Filed Dec. 6, 1955 8 Sheets-Sheet 5

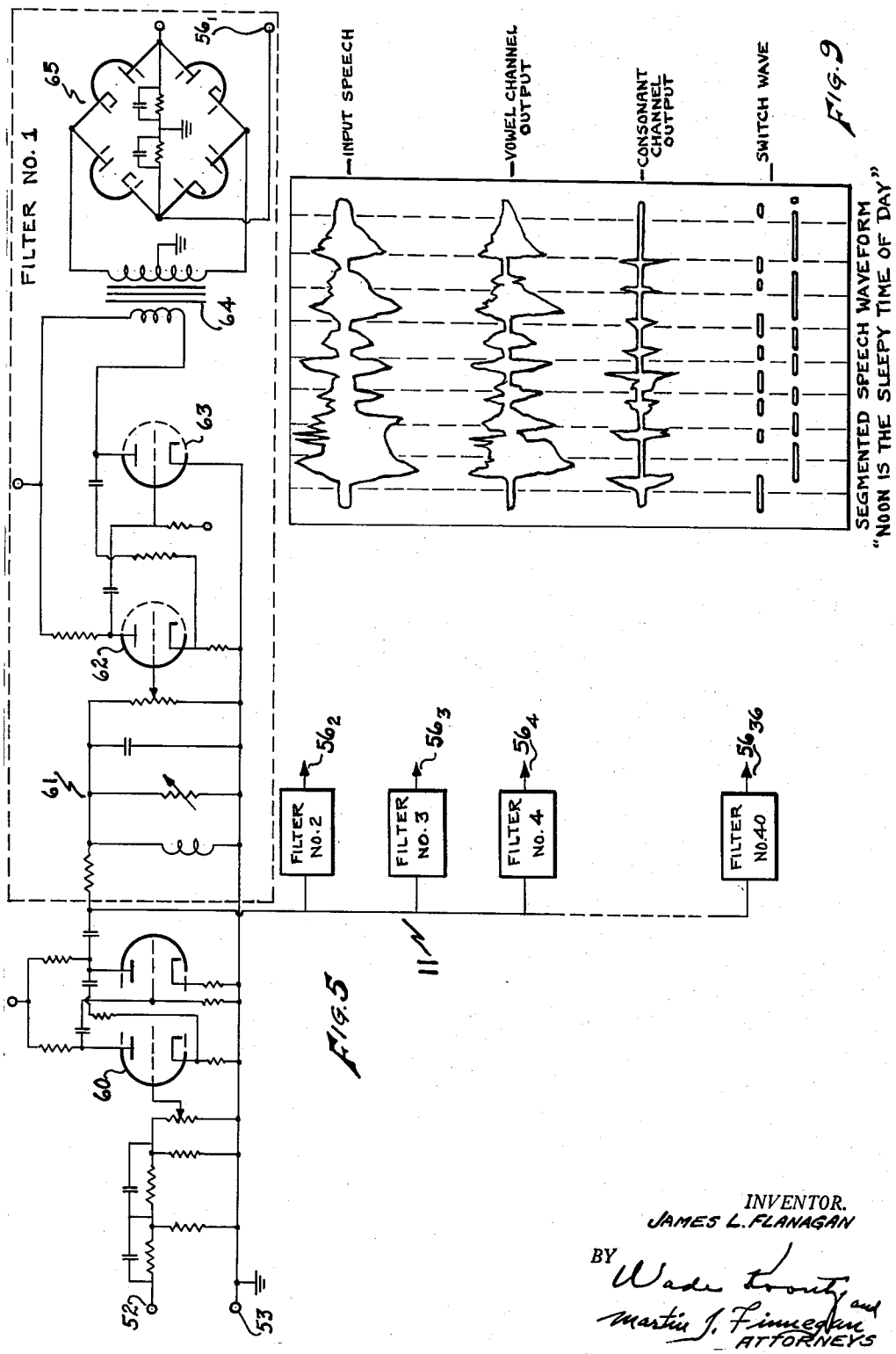


FIG. 5

FIG. 9

INVENTOR.
JAMES L. FLANAGAN

BY Wade Lowry and Martin J. Finnegan
ATTORNEYS

June 27, 1961 J. L. FLANAGAN 2,990,453

AUTOMATIC SPECTRUM ANALYZER

Filed Dec. 6, 1955 8 Sheets-Sheet 6

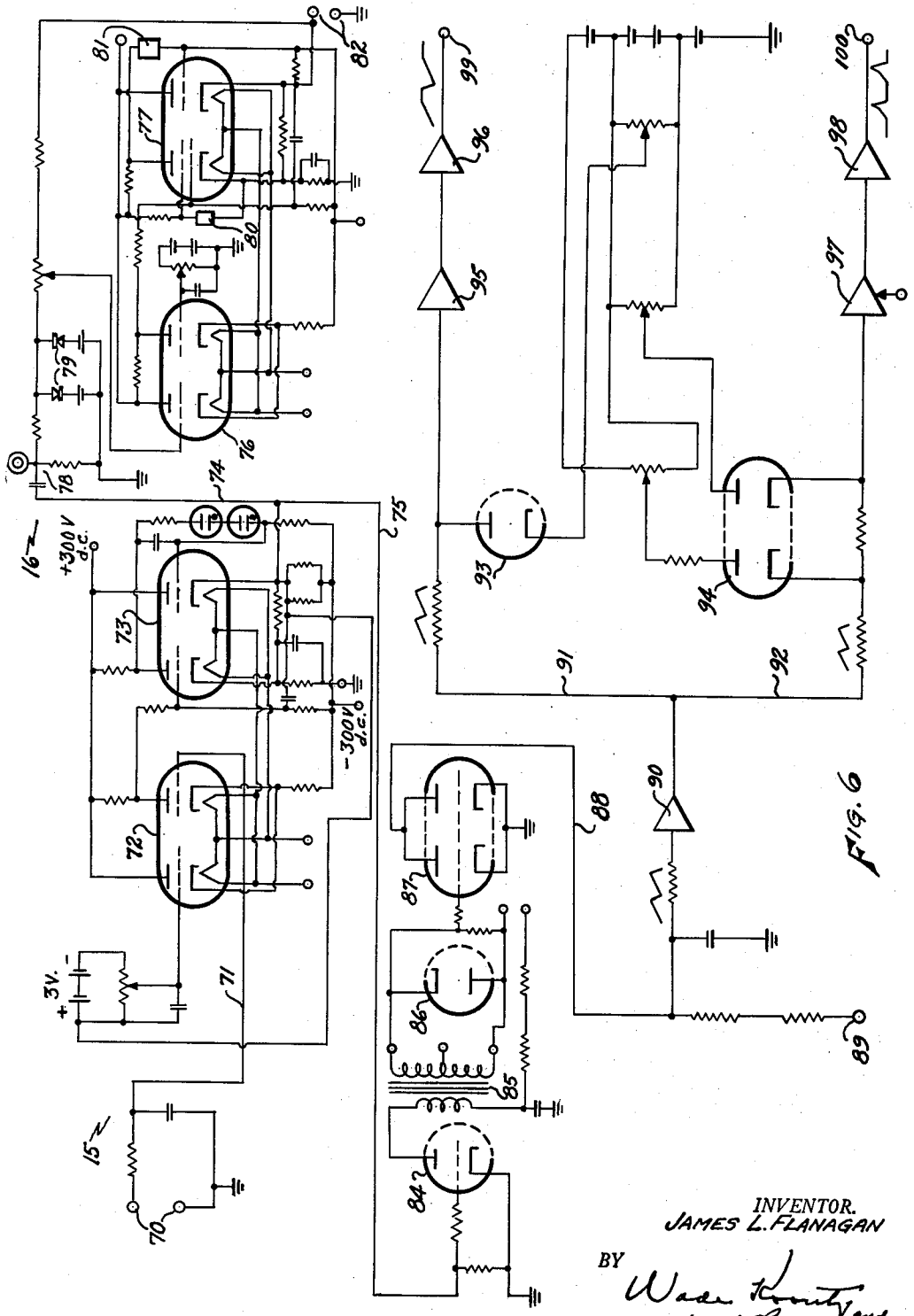


FIG. 6

INVENTOR.
JAMES L. FLANAGAN

BY Wade Hampton and Martin J. Finnegan
ATTORNEYS

June 27, 1961 J. L. FLANAGAN 2,990,453

AUTOMATIC SPECTRUM ANALYZER

Filed Dec. 6, 1955 8 Sheets-Sheet 7

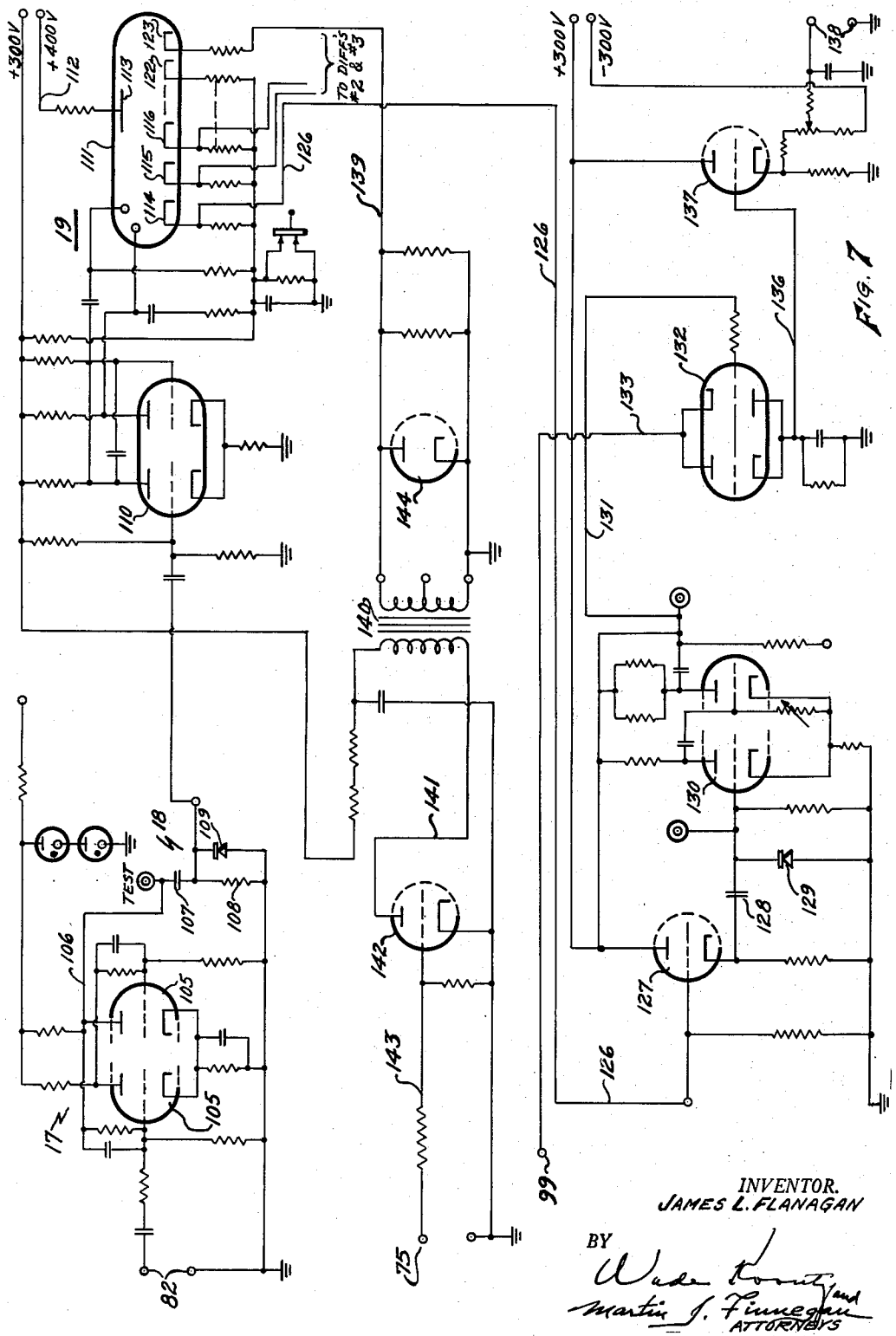


FIG. 7

INVENTOR.
JAMES L. FLANAGAN

BY
Wade Kroutz and
Martin J. Finnegan
ATTORNEYS

June 27, 1961 J. L. FLANAGAN 2,990,453

AUTOMATIC SPECTRUM ANALYZER

Filed Dec. 6, 1955 8 Sheets-Sheet 8

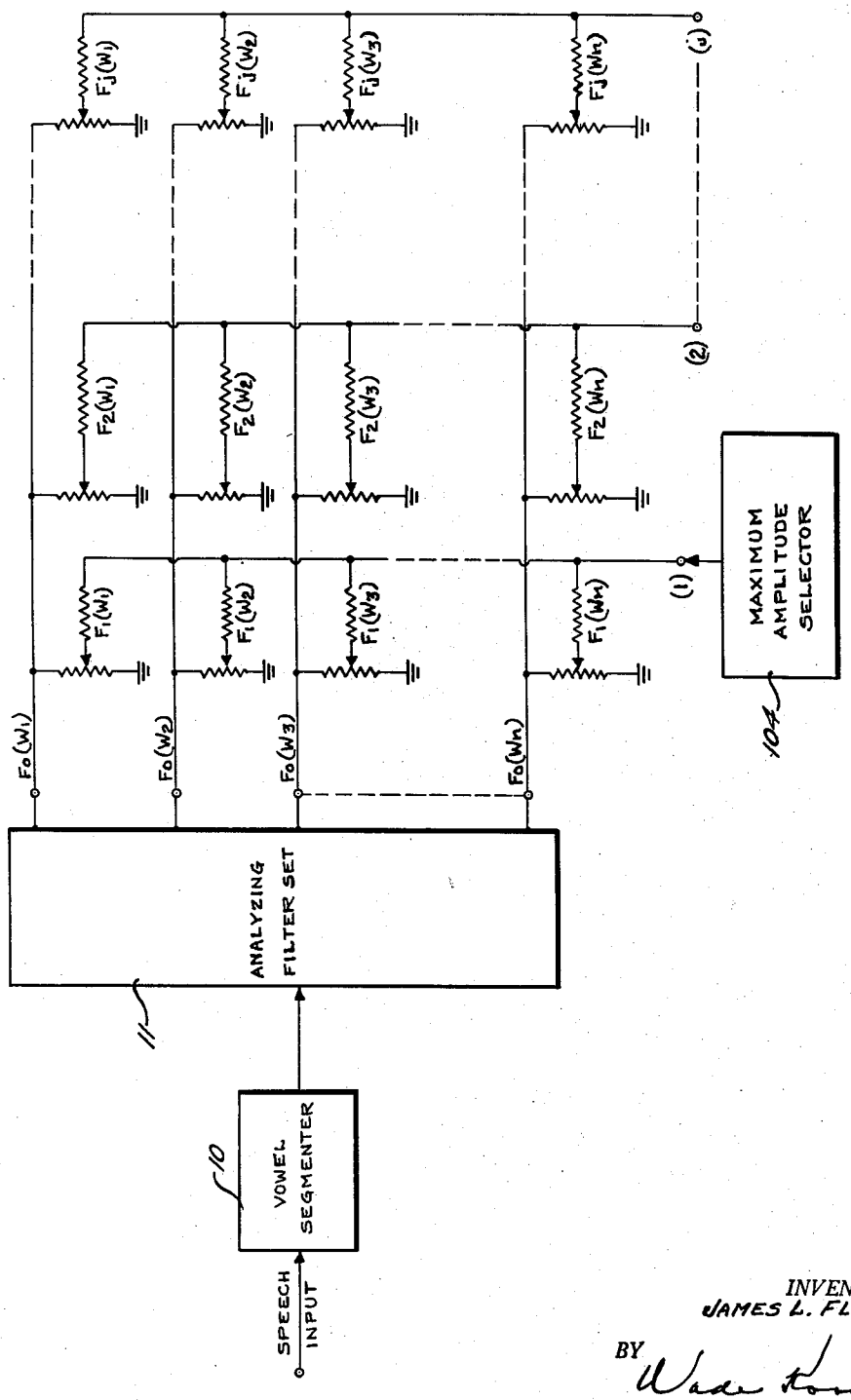


FIG. 10

INVENTOR.
JAMES L. FLANAGAN

BY Wade Kountz and Martin J. Finnegan
ATTORNEYS

United States Patent Office 2,990,453
Patented June 27, 1961

2,990,453

AUTOMATIC SPECTRUM ANALYZER

James L. Flanagan, Cambridge, Mass., assignor to the United States of America as represented by the Secretary of the Air Force

Filed Dec. 6, 1955, Ser. No. 551,478
10 Claims. (Cl. 179—15.55)
(Granted under Title 35, U.S. Code (1952), sec. 266)

The invention described herein may be manufactured and used by or for the United States Government for governmental purposes without payment to me of any royalty thereon.

This invention relates to the transmission of speech over a communication channel highly restricted in transmission bandwidth and capacity. The invention is pertinent to a speech bandwidth-compression system of the analysis-synthesis type, in which the speech information is coded in terms of signals representing the major vocal resonances (or formants) and the nature of the excitation of the vocal tract, both as functions of time during the production of speech. The major vocal resonances, or formants, are manifested as maxima in the frequency spectrum of the speech radiated by a talker. It is basic to the operation of the speech compression system that the frequencies of these spectral maxima (i.e. the formant frequencies) be determined by automatic analysis as the speech is uttered, and that signals representing these frequencies be transmitted to the speech synthesizer in order that the speech may be reproduced with negligible time delay.

The conventional telephone channel, which is a "waveform" transmission system, requires a transmission bandwidth of the order of 3000 cps. and a signal-to-noise ratio of the order of 30 db. Speech analysis synthesis systems for bandwidth compression heretofore known have required a total transmission bandwidth of the order of 300 cps. and a signal-to-noise ratio of about 30 db. The present invention makes possible a speech bandwidth compression system requiring a total transmission bandwidth of the order of 50 cps. and a signal-to-noise ratio of approximately 30 db.

The invention described herein, when incorporated as a bandwidth compression system of the character indicated, operates to accept continuous speech at its input and to deliver at its output electrical signals, varying slowly with time, whose amplitudes represent the frequencies of the first three major vocal resonances, or the first three formant frequencies of the input speech. The accomplishment of this task in the manner herein described involves, first, obtaining a gross Fourier transformation of the input speech, and by means of a set of contiguous band-pass filters, to provide a short-time amplitude spectrum of the speech. Appropriate analyzing and sampling circuitry then examines this short-time spectrum to determine, and properly indicate, the frequencies of the first three maxima.

The circuitry herein illustrated (reproduced, in substance, from the applicant's Doctoral Thesis entitled "A Speech Analyzer for a Forman-Coding Compression System," submitted to Massachusetts Institute of Technology in May 1955, and published in abridged form in the 1956 Journal of the Acoustical Society of America, is representative of a successfully operated embodiment of the invention. Other characteristics and advantages of the invention will appear upon examination of the following description of the invention as illustrated in the accompanying drawings wherein:

FIG. 1 is a block diagram indicating major components of the speech analysis apparatus adapted to receive electrical signals corresponding to the oral intelligence to be analyzed and to convert such received signals into electrical output signals of a character suitable to be handled in an analog type of speech synthesizer, in accordance with the principles of the present invention;

FIG. 2 is a block diagram of the sub-components entering into the circuitry constituting the vowel segmenter component diagrammatically shown in FIG. 1;

FIG. 3 is a circuit diagram showing vowel segmenter equipment indicated in block form in FIG. 2;

FIG. 4 is a circuit diagram of the multivibrator and switching branch of the vowel segmenter equipment indicated in block form in FIG. 2;

FIG. 5 is a circuit diagram of the analyzing filter component shown in block form in FIG. 1;

FIGS. 6 and 7 show the circuitry of some of the sequentially arranged components constituting the block diagram of FIG. 1;

FIG. 8 is a graph showing the frequency response characteristics of the driving amplifier for the analyzing filter set;

FIG. 9 is a pictorial representation of speech waveforms as they appear at the indicated points in the vowel segmenter;

FIG. 10 is a circuit diagram indicating an alternate method for speech formant extraction.

The basic principles of the spectrum-sampling formant extractor are illustrated in FIG. 1. The task of the system is to accept continuous speech at its input and to yield output voltages whose magnitudes, as functions of time, represent the frequencies of the first three speech formants. Continuous speech is fed directly, or through the vowel segmenting apparatus, into the analyzing filter set. The vowel segmenter (designated by reference numeral **10** in FIG. 1) is designed to allow only the vowel sounds to enter the analyzing system **11**. It is for these sounds that the formant structure is best defined. The filter set is composed of 36 contiguous band-pass filters, each with an associated amplifier, rectifier and smoothing network. The filter set provides, therefore, a short-time spectrum of the input speech. The outputs of the filter set **11** are scanned at 60 times per second by a constant speed motor-operated sampling switch **12**, to produce a time function of voltage samples representing the short-time spectrum of the input speech.

The time function obtained during one scan (in time period "T") of the sampling switch is illustrated in FIG. 1 by waveform (A) which is a waveform for a spectrum having only two maxima (or formants). The single negative sample shown in dash lines as part of said sample (A) is utilized in counter-resetting units **13** and **14** for resetting and synchronizing the equipment at the end of each scan. The output samples from the switch are sent through a clamper circuit **15** to produce the stairstep waveform (B). The waveforms (B) is differentiated and then peak-clipped (in circuitry **16**) to render the apparatus reasonably insensitive to the level of the input speech, yielding waveform (C). The differentiation produces a train of pulses which changes polarity at the times when the spectral maxima are passed: namely, t_1 and t_2. These pulses are allowed to trigger a binary scaler circuit **17** which changes state each time the derivative pulses change sign; the scaler yields waveform (D). The output of the scaler is differentiated and rectified (at unit **18**) to produce individual pulses indicating the time t_1 and t_2 within the sampling interval "T." These pulses are shown in (E). The next function is to separate these pulses marking the maxima into individual channels, and to convert the times t_1 and t_2 into voltages. Separation is accomplished by allowing the pulses (E) to trigger a scale-of-ten counter (shown at **19**) which is reset by a pulse from unit **13** at the end of each sampling period; (i.e. 60 times per sec.). The counter **19** utilizes a multi-cathode glow-

transfer tube shown in FIG. 7, and more fully described hereinafter. Trigger pulses cause conduction to move successively from one cathode to the next. The waveforms (F) are obtained respectively at the first three cathodes of the glow-transfer tube when the counter **19** is triggered by (E). The individual outputs from the counter are differentiated and rectified (in units **20**) to obtain (G), which represents an effective separation of the pulses in (E). The separated pulses, still indicating t_1 and t_2 independently generate (in units **21**) short gate pulses of accurately controlled duration, shown at (H). (The duration of the gate pulses is set to be 500 microseconds.) The gates **21** are delivered to units **22**, where they sample a calibrating (or sweep) waveform (shown at (I) and derived from generator **23**) as a linearly increasing (sawtooth) voltage. This calibrating waveform is generated in synchronism with the sampling switch. The amplitude values of the calibrating wave, read at times t_1 and t_2, are stored in the sampler circuits until the next sampling intervals. The voltages stored by the samplers are illustrated by the waveforms (I). These stored voltages are made available as outputs **24**. The sampler outputs may be smoothed by a low-pass network having a cutoff frequency chosen in accordance with the expected bandwidth of the formant signal. This cutoff frequency is of the order of 10 cps. The outputs **24** of the device, therefore, are smoothed voltages whose magnitudes, as functions of time, represent the formant frequencies.

A block diagram illustrating the principles of operation of the vowel segmenter **10**, and the inter-connection of its components, is shown in FIG. 2. The circuit diagram for the vowel segmenter is shown in FIGS. 3 and 4. The upper rectangle **30** of FIG. 3 shows one channel of the amplifier-rectifier components **27**, **28** and **29** of the vowel segmenter, the other rectangle **30'** containing identical components yielding an output of opposite polarity. Opposite polarity outputs are taken, therefore, from the band-pass and the band-pass elimination channels **31** and **32**, respectively. These outputs are smoothed with RC networks **33** and **34**, and differenced in a resistive summing network **35**. The difference voltage is applied to cathode follower triode **36** and then passes through center and peak clippers **37** and **38**. The center and peak clippers are arranged to have an output voltage-vs.-input voltage characteristic that is essentially a step function, +1.5 volts in magnitude, occurring at an input voltage of +1.5 volts. The clipped difference voltage is amplified in amplifier **39**, and then passed to input terminal A or B of FIG. 4, to switch the bistable multivibrator **40** of the electronic switch circuit, shown in FIG. 4. Terminals A and B of said multivibrator are also indicated in FIG. 3.

The output of the bistable multivibrator circuit is delivered to lines **41** and **42** (FIG. 4) and is passed through a symmetrical peak clipper **43**, **44** with variable build-up and decay time control. The clipped multivibrator output is then delivered to lines **46** and **47**, and allowed to gate a set of four balanced amplifier stages, constituted by twin triodes **48** and **49**. The four balanced amplifier stages, having two push-pull inputs and a single balanced output, act as a single pole, double throw switch. The speech to be segmented passes through this switch and is amplified in twin triode **50**, whose plate circuit supplies the output to terminals **52** and **53** of transformer **51**, which terminals may also serve as the input terminals (see FIG. 5) of the analyzer filter set. Connection of a speech input to terminals **25**(*b*) yields the segmented vowel sounds at the output terminals **52** and **53**, while connection to the input terminals **25***b'* yields the segmented consonants at the output terminals **52** and **53**.

From the vowel segmenter the signal energy passes to the analyzing filter set, consisting of 36 contiguous band-pass filters, all branching from the above-described single set of input terminals, as indicated at **52** and **53** in FIG. 5, and having individual output terminals corresponding to the output terminals shown at 56_1—56_2—56_3—56_4 and 56_{36} respectively. Each channel includes components corresponding to those indicated at **61**, **62**, **63**, **64** and **65** as the components of channel No. 1. These components include a single-tuned circuit **61**, a feedback amplifier **62**, **63**, a transformer **64** and a full-wave rectifier **65**. Either a positive or negative polarity output is available. For the present arrangement of the equipment the positive output from the filter set is used.

The positive D.C. outputs of the filter channels are carried over leads **66** (FIG. 1) to the circumferentially spaced points **67** of the sampler **12**, for successive transfer of signals to the rotating brush **68** whose pivoting end serves as a point of connection for the terminal of the output lead **69** which delivers the successively commutated signals to clamper **15**, as shown in FIG. 6, includes input terminals **70**, a lead **71** to the righthand grid of the first of two twin triode amplifiers **72**, **73**, and a cathode follower output circuit which divides into two branches **74** and **75**, the former supplying the signal to a differentiator-clipper network **78**—**79**, thence through amplifiers **76** and **77**, in sequence, to the cathode-follower output terminals **82**. Branch **75** supplies excitation voltage to the grid of triode **84** in the primary circuit of transformer **85**. The secondary circuit of transformers **85** includes the rectifier **86** and the grid circuit of dual amplifier **87**, the plate circuit **88** of said dual amplifier **87** being supplied from the positive source of high voltage indicated at **89**. The output of amplifier **87** is delivered by way of said conductor **88** to amplifier **90** and the amplified output is divided between branches **91** and **92**. The circuitry **84** through **90** generates the sawtooth calibrating waveform to be used in the sampler circuits **22** (FIG. 1). The sawtooth calibrating waves for the F_1 and for the F_2 and F_3 samplers are delivered to output terminals **99** and **100** by way of amplifier stages **95** and **96** on the one hand and **97** and **98** on the other. The sawtooth pulses are reshaped in the two channels leading to terminals **99** and **100** by the diodes **93** and **94**. The sweep voltages supplied the F_1 sampler and the F_2 and F_3 samplers are thereby constrained to the frequency ranges normally occupied by the speech formants. Typical waveforms for the F_1 and F_2—F_3 samplers as shown at terminals **99** and **100**, respectively.

The output of the differentiator-clipper **16** as delivered to the output terminals **82**, heretofore described, is supplied to the binary scaler **17** shown in FIG. 7 as consisting of the twin triode **105**, and an output conductor **106** leading to a differentiator-rectifier network **18** (as designated in the block diagram of FIG. 1), and shown in FIG. 7 as including capacitor **107**, resistor **108**, and crystal diode **109**. After such differentiation and rectification the successive signals are delivered to twin triode **110** which coacts with multi-cathode glow discharge tube **111** to perform the pulse counting and pulse separating operations heretofore referred to. At the end of each counting cycle the counting circuit is reset for the commencement of a new cycle by delivery of a resetting pulse to the final cathode element **123** of the glow discharge tube **111**, each resetting pulse being delivered over line **139** constituting the output circuit of transformer **140** whose alternating current output is rectified by action of diode **144**. The primary circuit of transformer **140** receives high voltage energization from the indicated 300 v. source, the action being controlled by triode **142** whose plate element is connected to the transformer primary by lead **141**. Triode **142** obtains its input from conductor **75** (FIG. 6) and is carried beyond cutoff by the negative sample pulse shown at A in FIG. 1.

The outputs from the counter circuitry **19** are delivered by way of conductor **126**, as shown in FIG. 7 to cathode follower triode **127** from which the cathode follower output which is differentiated and rectified by the combined action of capacitor **128** and crystal diode **129**, after which the differentiated and rectified output is allowed to trig-

ger a one-shot multivibrator, shown as the twin triode **130** with its associated timing network. This results in the delivery of gate pulses over conductor **131** to sampler tube **132** whose plate circuit includes a lead **133** which receives the sawtooth pulse train generated in the element **23** and delivered thereto by way of either of terminals **99** or **100**, previously described in the discussion of FIG. 6. The voltage "read" and stored by sampler tube **132** is delivered to triode cathode follower **137** and the output is supplied to the output terminals **138** indicated in FIG. 7. These terminals **138** correspond to the terminals **24** of FIG. 1, the circuitry being duplicated for each formant sampler F_1—F_2—F_3.

FIG. 10 illustrates alternate circuitry for achieving the purposes of the invention. In this FIG. 10 the elements **10** and **11** correspond to those similarly designated in FIG. 1. The filter set **11** yields essentially the values of the short-time spectrum, $F_0(w)$, at discrete frequency points, w_1, w_2, w_n. The spectrum $F_0(w)$ is effectively multiplied by the values of the spectra $F_1(w)$, $F_2(w)$, $F_j(w)$, consisting of standard spectral patterns stored in the circuitry. The circuitry operates on the principle of cross-correlating the short-time spectrum of the input speech with this group of standard spectral patterns. The system periodically indicates the spectral pattern which affords the best correlation. As speech is fed into the system, the indicated network performs a computation of the correlation between each successively stored pattern and the short-time spectrum of input speech. This computation is performed for zero delay between the correlated functions. The stored pattern affording the maximum correlation during any particular selecting intervals is indicated by the maximum amplitude selector circuitry indicated in block form at **104** in FIG. 10, which circuitry may involve the multicathode glow-transfer tube **111**, and associated appropriate circuitry, or any equivalent maximum amplitude selector apparatus may be substituted. Selection of a particular stored pattern is equivalent to specifying a standard set of speech formants. Therefore, it is possible to utilize the system of FIG. 10 as a formant tracker to attain the same ultimate result as is attained by the circuitry of FIGS. 1 to 7 inclusive.

The spectrum $F_0(w)$ is effectively multiplied by the values of the stored spectra at the corresponding discrete frequency points by appropriately setting the arms on each column of the potentiometers associated with said stored spectral patterns. The outputs from the arms of each vertical column of potentiometers as indicated in FIG. 10, are subjected to a summing operation and the maximum amplitude selector **104** periodically indicates the column (that is, the standard pattern) which affords the best correlation. A signal representing the column thus selected may then be supplied to the speech synthesizer equipment, said signal effectively specifying the formant pattern of the speech entering the analyzing filter set, and said signal hence being equivalent to the outputs **24** (FIG. 1) of the system illustrated in FIGS. 1 to 7 inclusive.

What is claimed is:

1. A speech analyzing apparatus for the development of electrical energy impulses suitable for application to speech synthesizing equipment, said apparatus comprising signal receiving means for accepting signals of a frequency corresponding to the frequency of sound waves generated by selected formant components of continuous speech, means for cyclically scanning said signals to produce a waveform for each scanning cycle, which waveform will include as many amplitude peaks as there are selected formant components in the speech received during said scanning cycle, means for converting said amplitude peaks to pulses whose polarity reverses on each occurrence of signal components of maximum voltage amplitude within a predetermined spectral range constituting the identifying frequency envelope for said selected formant-derived signals, and means for counting the polarity reversals.

2. A speech analyzing apparatus for the development of electrical energy impulses suitable for application to speech synthesizing equipment of the terminal-analog type, or equivalent computer mechanism, said apparatus comprising signal receiving means for accepting signals of a frequency corresponding to the frequency of sound waves generated by selected formant components of continuous speech, means for cyclically scanning said signals to produce a waveform for each scanning cycle, which waveform will include as many amplitude peaks as there are selected formant components in the speech received during said scanning cycle, means for converting said amplitude peaks to pulses whose polarity reverses on each occurrence of signal components of maximum voltage amplitude within a predetermined spectral range constituting the identifying frequency envelope for said selected formant-derived signals, means for counting the polarity reversals, and means for obtaining an output voltage whose magnitude is proportional to the counting total achieved by said counting means in a given counting interval.

3. Speech analyzing apparatus as defined in claim 1, including means associated with said counting means for separating said polarity reversals into separate circuits representative of the respective formant components, in accordance with the relative time displacements between said polarity reversals.

4. Speech analyzing apparatus as defined in claim 1 wherein said signal receiving means includes vowel segmenting circuitry responsive to the frequency pattern generated by the speech under analysis for interrupting the flow of energy during intervals when there are no vowel sounds in the speech under analysis.

5. Speech analyzing apparatus as defined in claim 1 including means for dividing the received signal energy into a plurality of channels, each having a frequency range constituting a selected fraction of the preselected total range, and means for electrically connecting each of said fractional signal channels to said signal converting means in predetermined succession.

6. Speech analyzing apparatus as defined in claim 1 wherein said pulse generating means includes binary scaler means receiving the entire energy content of said signal receiving means, and means for conditioning the output of said binary scaler means for application of said scaled energy to said counting means.

7. Speech analyzing apparatus as defined in claim 1 including means for obtaining a plurality of voltage outputs varying each from the other in accordance with the relative time displacements between said pulses.

8. Speech analyzing apparatus as defined in claim 1 including means for processing the total energy content of said signal receiving means in a series of successive stages having a time sequence corresponding to the progressively varying frequency stages constituting the total range of said envelope.

9. Speech analyzing apparatus as defined in claim 2 including means for recycling said pulse-counting means at regularly occurring time intervals.

10. Speech analyzing apparatus as defined in claim 2 including means for progressively sampling the signal content of said signal receiving means, and means for restoring said counting means to the zero position at the termination of each cycle of operation of said sampling means.

References Cited in the file of this patent

UNITED STATES PATENTS

2,098,956	Dudley	Nov. 16, 1937
2,458,227	Vermeulen et al.	Jan. 4, 1949
2,629,017	Dahlbom et al.	Feb. 17, 1953
2,646,465	Davis et al.	July 21, 1953
2,672,512	Mathes	Mar. 16, 1954
2,681,385	Oliver	June 15, 1954
2,766,325	Di Toro	Oct. 9, 1956