Aug. 8, 1967 S. J. CAMPANELLA ETAL 3,335,225

FORMANT PERIOD TRACKER

Filed Feb. 20, 1964 2 Sheets-Sheet 1

FIG. 1

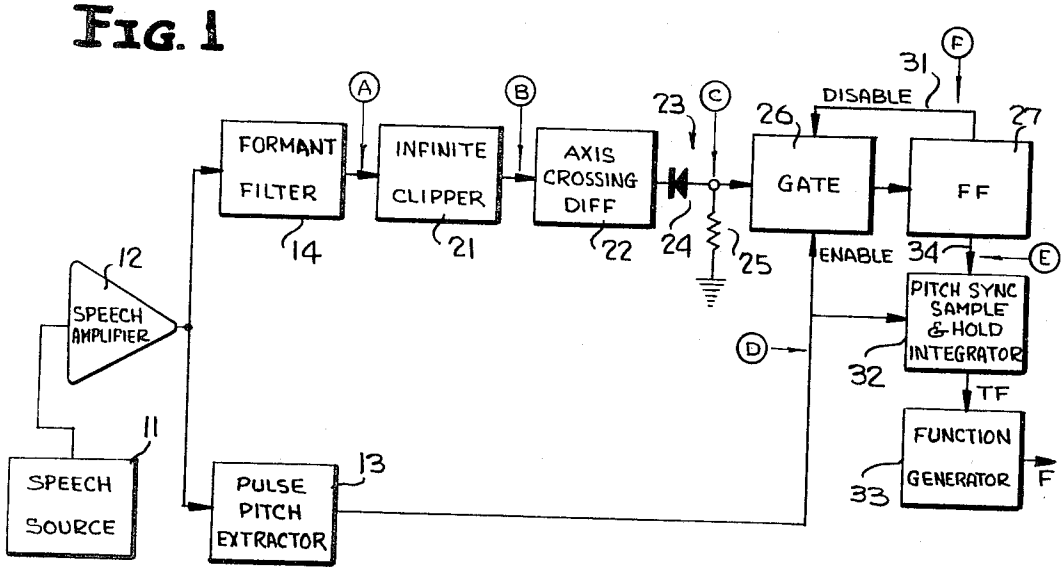


FIG. 2

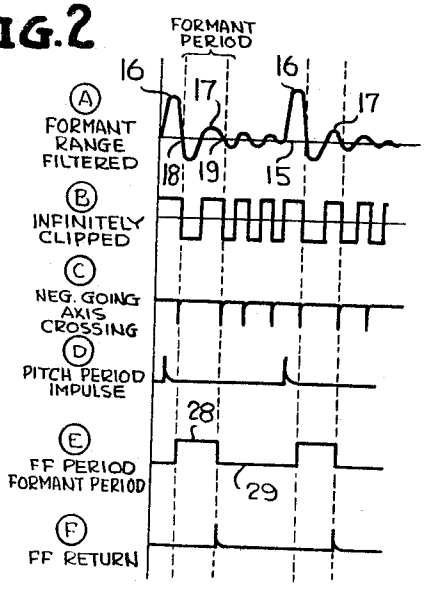
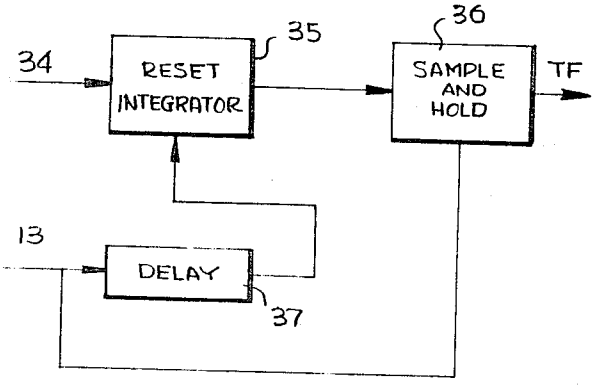


FIG. 3



INVENTORS
S. JOSEPH CAMPANELLA
& DAVID C. COULTER

BY Hurwitz & Rose

ATTORNEYS

Aug. 8, 1967 S. J. CAMPANELLA ETAL 3,335,225

FORMANT PERIOD TRACKER

Filed Feb. 20, 1964 2 Sheets-Sheet 2

FIG. 4

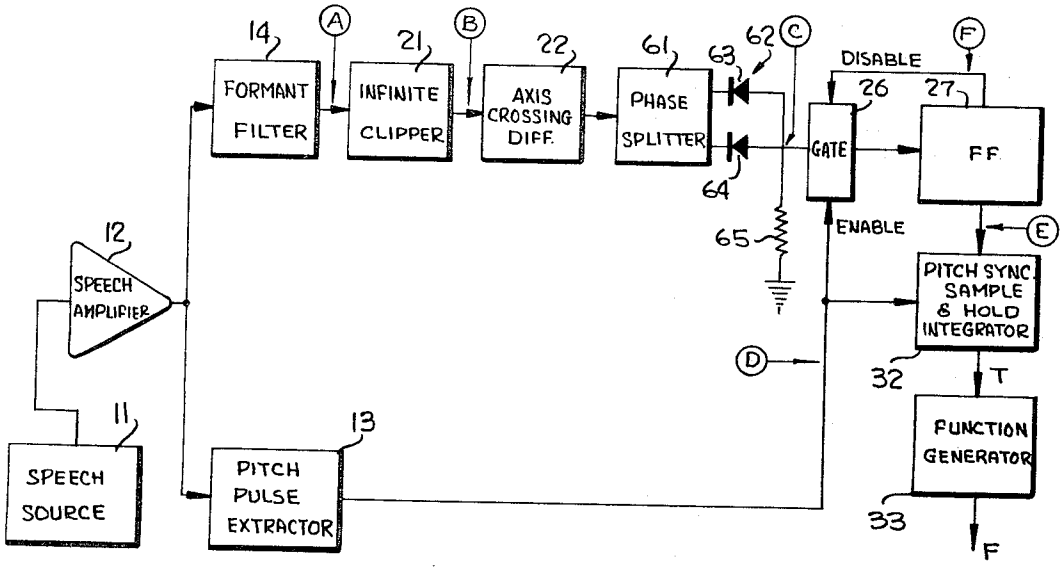
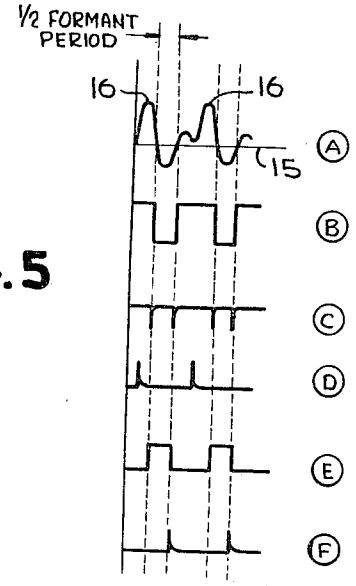


FIG. 5



INVENTORS
S. JOSEPH CAMPANELLA
& DAVID C. COULTER

BY Hurwitz & Ross

ATTORNEYS

United States Patent Office 3,335,225 Patented Aug. 8, 1967

3,335,225

FORMANT PERIOD TRACKER

Samuel Joseph Campanella and David C. Coulter, Springfield, Va., assignors to Melpar, Inc., Falls Church, Va., a corporation of Delaware

Filed Feb. 20, 1964, Ser. No. 346,185
13 Claims. (Cl. 179—1)

ABSTRACT OF THE DISCLOSURE

Formant frequency measurements are provided, for formant tracking in a speech compression system, by passing the frequency range of the formant of interest in the incoming speech wave to produce a periodic waveform representative of that formant, and detecting the length of the time interval over which an integral number of consecutive half cycles of the periodic waveform occur, only within the time segment between successive pitch excitation discontinuities of the speech wave.

The present invention relates generally to speech analyzing systems. More particularly, the invention relates to a device for deriving information regarding the frequency of energy in one of the speech formants as measured by the period of the damped sinusoid following each larynx excitation.

In speech bandwith compression and recognition systems, an important but frequently overlooked parameter is the frequency of each formant that arises in response to each larynx excitation. As is well known, each time the larynx is excited it produces a set of exponentially damped sinusoidal waves. This set of waves, which occurs for voiced utterences, includes frequency components that generally lie in three ranges or formants; these ranges for the average male being 200 to 1,000 c.p.s., 800 to 2,300 c.p.s., and 2,300 to 3,800 c.p.s. Each time the larynx is re-excited the previous set of sinusoidal waves is usually completely damped because the Q of the previously existing resonant cavity drops virtually to zero in response to opening of the glottis. Thus, there is virtualy no phase interference between waves deriving from adjacent larynx excitations and the damped sinusoids are easily identified by filters that segment the frequency ranges occupied by the formants.

We have found that formant frequencies can be ascertained by measuring the period of the damped sinusoid following each larynx excitation in the formant of interest. This period is inversely proportional to the formant frequency. The period can be measured as a function of the time it takes a predetermined number of half cycles of the damped sinusoid to be completed. The length of each half cycle is measured in response to the time duration between adjacent crossings of the sinusoid as it crosses a zero reference. As long as the number of predetermined half cycles is such that the damped sinusoid is still being generated while the period measurement is being conducted, this approach is satisfactory. If, however, the larynx is excited while the time duration of the half cycles is being measured, the information derived would be meaningless because there is no phase relation between the waves of adjacent larynx excitations. Thus, for low pitched speakers, the number of half cycles can frequently be two or more. But for the first formant of high pitched female speakers, who frequently excite their larynx before the completion of a full formant frequency period, after the main larynx or pitch excitation, the number must be limited to one.

Because all measurements according to the present invention are made in response to measeurements on continuous waves, i.e. the damped sinusoid, the problem of pitch harmonic interference is avoided. Pitch harmonic interference arises in prior art devices because the damped sinusoid will not generally be crosing the reference axis at the same time a new larynx excitation occurs. Hence, the waves of adjacent damped sinusoids are not usually continuous and the average number of axis crossings does not provide an accurate measure of each half cycle duration. A similar problem arises in attempts to measure formant frequency location from spectrum analysis data since the closest that a spectrum sample can fall to a formant location is a frequency equal to an integral multiple of the fundamental pitch frequency.

It is accordingly an object of the present invention to provide a new and improved system for deriving formant period information.

Another object of the invention is to provide a system for measuring formant period in response to only the continuous portions of the damped sinusoid deriving from a larynx excitation.

An additional object is to provide a formant period measuring device that provides accurate information for low and high pitched speakers and is insensitive to phase relations of adjacent damped sinusoidal segments deriving from successive larynx excitations.

It is a further object of the invention to provide a device for measuring formant period in repsonse to the time duration between a predetermined number of axis crossings of the damped sinusoid following a larynx excitation.

The above and still further objects, features and advantages of the present invention will become apparent upon consideration of the following detailed description of several specific embodiments thereof, especially when taken in conjunction with the accompanying drawings, wherein:

FIGURE 1 is a block diagram of a preferred embodiment of the present invention wherein formant period is measured in response to the completion of a full cycle of the damped sinusoid following larynx excitation;

FIGURES 2A–2F are wave forms to aid in describing the operation of FIGURE 1;

FIGURE 3 is a block diagram of a portion of the circuitry of FIGURE 1;

FIGURE 4 is a block diagram of a further embodiment of the invention wherein formant period is measured in response to the completion of a half cycle of the damped sinusoid following larynx excitation; and

FIGURES 5A–5F are wave forms to aid describing the operation of FIGURE 4.

Reference is now made to FIGURE 1 of the drawings wherein speech source **11** feeds amplifier **12** having slow response AGC which normalizes, to a certain extent, the signal amplitude applied in parallel to pitch pulse extractor **13** and formant filter **14**. Filter **14** is a band pass filter having a center frequency and width commensurate with the formant being analyzed; for the first formant it is a filter having a pass band between 200 and 1000 c.p.s. The waveform deriving from filter **14** is a series of damped sinusoids centered about axis **15**, as indicated in FIGURE 2A. In response to each larynx excitation deriving from speech source **11**, there is produced a large amplitude positive wave **16**. The time separation between peak values of adjacent ones of waves **16** is generally termed the pitch period. Following each positive wave **16** there is produced an exponentially damped sinusoid **17** having a repetition rate equal to the formant frequency of the wave passing through filter **14**. Thus, the time separation between axis crossings of adjacent negative going segments of wave **17**, i.e., between points **18** and **19**, is inversely proportional to the formant frequency of interest. According to the present invention the time between these axis crossings or a predetermined number of other

crossings, is measured to provide an indication of formant frequency.

The waveform of FIGURE 2A deriving from formant filter **14** is applied to infinite clipper **21** which generates the rectangular wave of FIGURE 2B. In response to any segment of the wave of FIGURE 2A being above axis **15**, clipper **21** derives a positive voltage of constant amplitude. Clipper **21** produces a negative voltage of constant amplitude as soon as the wave of FIGURE 2A goes below axis **15**. Thereby, the wave derived by clipper **21** consists of a series of constant amplitude positive and negative voltages which are in phase with the variations of the damped sinusoid illustrated in FIGURE 2A.

The output of clipper **21**, FIGURE 2B, is applied to RC differentiator **22** that derives positive and negative going pulses in response to the positive and negative going edges, respectively, of the infinitely clipped wave. These pulses are applied to half wave rectifier **23** comprising diode **24** and load resistor **25**. Diode **24** is poled such that only the negative pulses are passed to resistor **25**. In consequence, the waveform across resistor **25** comprises a series of negative going pulses, FIGURE 2C, each of the pulses being derived simultaneously with a negative going crossing of the damped sinusoid across axis **15**, FIGURE 2A. The output of rectifier **23** is selectively coupled through bi-stable gate **26** to the input of bistable flip flop **27**.

For the initial positive going segment **16** of the damped sinusoid, gate **26** is closed, in a manner seen infra, to prevent coupling between the output of rectifier **23** and the input of flip flop **27**. When the peak value of wave segment **16** is reached, pitch pulse extractor **13** derives the positive pulse indicated in FIGURE 2D. Extractor **13** may take any conventional form, such as described by Gruenz, Jr. et al. in an article published in the "Journal of Acoustic Society of America," September 1949, p. 487.

The positive pulses deriving from extractor **13** are applied to one input of gate **26**. Thereby, gate **26** opens and stays open to enable the axis crossing pulses, FIGURE 2C, to be applied to flip flop **27**. In response to the first pulse immediately following opening of gate **26**, indicative of axis crossing **18**, flip flop **27** is switched so it derives the positively going rectangular wave **28**, FIGURE 2E. Flip flop **27** remains in this state until the next pulse is generated by rectifier **23**. In response to the next pulse, derived in response to axis crossing **19**, flip flop **27** is restored to its initial state so its output voltage goes negatively and voltage level **29** is reached.

In response to flip flop **27** being restored to its initial state, it derives on lead **31** the pulse indicated in FIGURE 2F. This pulse is coupled from flip flop **27** to the disable input of gate **26**, and causes the gate to close. Thereby, the negative going pulses in the waveform of FIGURE 2C occurring between axis crossing **19** and the next large wave segment **16** have no effect on flip flop **27**. It is thus seen that the time interval during which flip flop **27** is activated to generate wave segment **28** is exactly equal to the period between axis crossings **18** and **19**, hence inversely proportional to the formant frequency passing through filter **14**.

To convert the wave of FIGURE 2E into a voltage directly proportional to the formant frequency between each larynx excitation that causes wave segments **16**, sample and hold integrator **32** and reciprocal function generator **33** are provided. The output of flip flop **27** on lead **34**, as depicted in FIGURE 2E, is applied to pitch sync, sample and hold integrator **32**. As indicated in FIGURE 3, circuit **32** includes a resettable integrator **35** having a time constant selected such that rectangular wave **28** is converted into a substantially linear sawtooth for all frequencies in the formant range of interest. The output of reset integrator **35** is coupled to sample and hold circuit **36** that is retriggered in response to each pitch period impulse deriving from extractor **13**. The pitch period impulses are also applied as reset pulses to integrator **35** via delay element **37**. The length of time introduced by delay **37** is such that the output of integrator **35** is at a level indicative of the time duration of the preceding wave **28** and is then reset to a zero level prior to the occurrence of the leading edge of the next wave **28**. Thus, the output of circuit **36** comprises a series of varying amplitude steps, the level of each being proportional to the preceding formant period. The voltage deriving from circuit **32** is coupled to function generator **33** that derives a signal equal in value to the reciprocal of its input amplitude, hence directly proportional to the formant frequency.

While the circuit of FIGURE 1 is suitable for many speakers in the first and second formants, it does not provide an accurate measurement for some people who speak at a very high pitch. The inaccuracy is caused in such cases because the larynx may be excited before occurrence of the second negative going axis crossing **19**, FIGURE 2A. To cure this situation, the system of FIGURE 4, wherein the period between adjacent axis crossings is used to measure formant frequency, was developed.

The system of FIGURE 4 is identical to the one of FIGURE 1 except that phase splitter **61** and full wave rectifier **62** have been substituted for half wave rectifier **23**. The opposite polarity outputs of phase splitter **61** are applied to the cathodes of diodes **63** and **64** in rectifier **62** so that the voltage across load resistor **65** appears as a series of negative pulses, one pulse being derived in response to each axis crossing of the wave deriving from filter **14**.

To provide a better understanding of the operation of FIGURE 4, reference is now made to the waveforms of FIGURES 5A–5F. It is assumed that the speech analysis is being performed for a high pitched speaker so that a larynx excitation, indicated by large positive wave segment **16**, occurs before a pair of negative going axis crossings.

In response to each crossing of axis **15** about which the wave of FIGURE 5A oscillates there is derived a negative going pulse across resistor **65**, as indicated by the waveform of FIGURE 5C. Thus, rectifier **62** always generates a pulse at the beginning and end of each half cycle of the damped sinusoid except at the beginning of large amplitude segment **16**. The first and second of these pulses for each larynx excitation are coupled to flip flop **27** since gate **26** is then open in response to the pitch period impulse, FIGURE 5D, deriving from circuit **13**. The second pulse coupled to flip flop **27** causes the flip flop to change state to prevent gate **26** from passing additional pulses until the next pitch period impulse is generated. Thus, flip flop **27** stays in the switched state whereby wave **28** is generated for a time interval equal to one half the formant period.

The rectangular wave deriving from flip flop **27** is coupled to circuits **32** and **33**. The wave is converted into a series of analog voltages, each of which represents the formant frequency characteristic of the particular larynx excitation. Function generator **33** is calibrated with appropriate valued resistors so that the voltage deriving from it is directly proportional to formant frequency, rather than twice the formant frequency.

While we have described and illustrated several specific embodiments of our invention, it will be clear that variations of the details of construction which are specifically illustrated and described may be resorted to without departing from the true spirit and scope of the invention as defined in the appended claims.

We claim:

1. In a system for measuring the formant frequency of speech resulting from larynx excitations followed by damped sinusoidal waves, the combination comprising means responsive to a speech formant for deriving therefrom an indication of the length of time required for a predetermined number of half cycles of the damped sinusoidal wave following the larynx excitation, and means for enabling said deriving means in response to each

larynx excitation and for disabling said indicating means before the next larynx excitation.

2. In a system for measuring formant frequency of speech resulting from larynx excitations followed by damped sinusoidal waves, the combination comprising means responsive to speech formants for deriving therefrom an indication of the length of time required for a predetermined number of half cycles of the damped sinusoidal wave following the larynx excitation, and means for enabling said deriving means in response to each larynx excitation and for disabling said indicating means in response to completion of said predetermined number of half cycles.

3. The system of claim 2 wherein said number equals 1.

4. The system of claim 2 wherein said number equals 2.

5. In a system for measuring the formant frequency of speech energy, the combination comprising a formant filter responsive to said speech energy, said filter deriving a damped sinusoid in response to each larynx excitation causing said speech energy, said sinusoid oscillating about a reference axis, means for deriving an indication of the length of time required for a predetermined number of axis crossings of said sinusoid following each larynx excitation, and means for enabling said deriving means in response to each larynx excitation and for disabling said indicating means before the next larynx excitation.

6. In a system for measuring the formant frequency of speech energy, the combination comprising a formant filter responsive to said speech energy, said filter deriving a damped sinusoid in response to each larynx excitation causing said speech energy, said sinusoid oscillating about a reference axis, means for deriving an indication of the length of time required for a predetermined number of axis crossings of said sinusoid following each larynx excitation, and means for enabling said deriving means in response to each larynx excitation and for disabling said indicating means in response to completion of said predetermined number of said predetermined number of half cycles.

7. The system of claim 6 wherein said predetermined number equals 1.

8. The system of claim 6 wherein said predetermined number equals 2.

9. In a system for measuring the formant frequency of speech energy, the combination comprising a formant filter responsive to said speech energy, said filter deriving a damped sinusoid in response to each larynx excitation causing said speech energy, said sinusoid oscillating about a reference axis, means for deriving an impulse in response to each axis crossing of said sinusoid, means for measuring the time duration between adjacent ones of said impulses, means for deriving an indication in response to each larynx excitation, means for coupling said indication to said measuring means to enable said measuring means, and means for disabling said measuring means in response to the second impulse occurring after said indication.

10. In a system for measuring the formant frequency of speech energy, the combination comprising a formant filter responsive to said speech energy, said filter deriving a damped sinusoid in response to each larynx excitation causing said speech energy, said sinusoid oscillating about a reference axis, means for deriving an impulse in response to alternate axis crossings of said sinusoid, means for measuring the time duration between adjacent ones of said impulses, means for deriving an indication in response to each larynx excitation, means for coupling said indication to said measuring means to enable said measuring means, and means for disabling said measuring means in response to the second impulse occurring after said indication.

11. In a system for measuring the formant frequency of speech energy, the combination comprising a formant filter responsive to said speech energy, said filter deriving a damped sinusoid in response to each larynx excitation causing said speech energy, said sinusoid oscillating about a reference axis, means for deriving an impulse in response to each cycle of said sinusoid attaining a predetermined phase position, means for measuring the time duration between adjacent ones of said impulses, means for deriving an indication in response to each larynx excitation, means for coupling said indication to said measuring means to enable said measuring means, and means for disabling said measuring means in response to the second impulse occurring after said indication.

12. The system of claim 11 wherein said measuring means includes means for deriving a signal of constant value between adjacent larynx excitations, said signal being proportional to the time between said adjacent ones of said impulses occurring during the previous pair of larynx excitations.

13. In a speech compression system,

means for measuring the frequency of speech formants,

said measuring means comprising filter means responsive to incoming speech signal for passing the frequency range of the formant of interest to derive therefrom a periodic waveform representative of said formant,

means responsive to said waveform for detecting the length of the time interval encompassed by an integral number of consecutive half cycles of said waveform,

means further responsive to said incoming speech signal for activating said detecting means only between successive pitch excitation discontinuities of said speech signal, and

means responsive to the detected time interval length for deriving therefrom an indication of formant frequency.

References Cited

UNITED STATES PATENTS

3,020,344 2/1962 Prestigiacomo ---------- 179—1

KATHLEEN H. CLAFFY, *Primary Examiner.*

R. MURRAY, *Assistant Examiner.*