SYSTEM FOR DETERMINING CONSONANT FORMANT LOCI

Fig 1

INVENTOR
DAVID C. COULTER

ATTORNEYS
SYSTEM FOR DETERMINING CONSONANT FORMANT LOCI

David C. Coulter, Springfield, Va., assignor to Melpar, Inc., Falls Church, Va., a corporation of Delaware
Filed Apr. 9, 1962, Ser. No. 186,080
14 Claims. (Cl. 179—1)

The present invention relates generally to speech analyzing apparatus and more particularly to a speech analyzer for voiced consonants wherein the frequency loci of voice consonants are determined by extrapolating the slope of the consonant sound to its initial enunciation.

It is known that voiced consonant phonemes have fixed, predetermined frequencies at which they originate to form sounds in combination with vowels. For example, the consonants, B, D, and G appear to originate at frequencies of 720, 1900, and 3000 cycles per second, respectively. If when these consonants are coupled at the end of vowel sounds, they terminate at a frequency equal to the same frequency as when the consonants are initiated. The point at which a frequency appears to emanate from or be directed toward is known as the frequency locus of the consonant.

As a consonant phoneme is enunciated, the frequency changes at a substantially linear rate from the frequency locus to the vowel frequency which remains constant as the vowel is sounded. As the vowel phoneme is terminated and the terminal consonant is sounded, the frequency changes again at a substantially constant rate toward the frequency locus of the particular consonant. Thus, while pronouncing the word bed, B, D, the consonant "b" appears to be initiated at a frequency of 700 cycles per second and increases at a uniform rate to a frequency of 2300 cycles. When the sound "e" is pronounced the frequency is maintained substantially constant at 2300 cycles per second until the beginning of the letter "d". At this point, the frequency decreases at a linear rate to 1800 cycles.

In the present invention, the frequency locus of the consonant is obtained by calculating the rate of change of the first consonant in the sound, computing the duration which the consonant is voiced, and extrapolating from the constant frequency of the vowel. Extrapolation for the initial consonant is performed by subtracting the product of the slope and the time which it takes the consonant to be spoken from the vowel frequency. To ascertain the terminal consonant, the product of the time of enunciation and the slope of the consonant is added to the vowel frequency.

While the consonant loci are constant, the frequency at which the consonant is first heard is not usually constant because of the silent interval between the initiation of the sound and the time at which it is first heard. A similar silent interval occurs at the end of a sound so that the terminal frequencies of the same consonant are variable. The silent intervals in both the initial and terminal consonants of a sound are substantially one-half the time which it takes the frequency to change from the consonant locus to the vowel frequency. By setting the voiced portion of the consonant begins and when the consonant terminates, and multiplying this factor by 2, the actual time including both voiced and silent periods for the consonant utterance may easily be computed.

It is an object of the present invention to provide a system for deriving the initial loci of consonant sounds in speech.

It is another object of the present invention to provide a system for automatically ascertaining terminal loci of voiced consonants.

It is still another object of the present invention to provide a system for ascertaining both initial and terminal loci of voiced consonants when coupled with vowel sounds.

It is another object of the present invention to provide a system for ascertaining consonant loci by extrapolating the slope of the frequency variation of the consonant between the time at which the consonant is first uttered and the time at which the vowel sound is initiated.

It is still a further object of the present invention to provide a system for determining frequency loci of voiceless consonants regardless of the slope of the consonant sound or phoneme as it approaches or leaves the vowel with which it is coupled.

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

FIGURE 1 is a block diagram of one form of the present invention; and
FIGURES 2B-2F are diagrams illustrating wave forms for a typical sound applied to the system of FIGURE 1.

Reference is now made to FIGURE 1 of the drawings which discloses a microphone 11 responsive to an audio speech signal. The output of the microphone 11 is applied in parallel to a formant analyzer 12 and a voice/unvoiced-silence detector 13. The format analyzer 12 derives an output signal commensurate with the frequency of the signal derived from microphone 11 in the second format, i.e. between 700 and 3000 cycles per second. The output of formant analyzer 12 is a low frequency signal, 0 to 25 cycles per second, commensurate with the centroid frequency of the speech in the second format.

The voice/unvoiced-silence detector 13 derives an output signal of a binary nature, either 0 or 1. When a voiced signal, i.e. one having a particular pitch frequency is derived from microphone 11, voice/unvoiced-silence detector 13 derives a binary 0 output. When, however, an unvoiced or silent input thereto is derived from microphone 11, a binary 1 output is generated by detector 13. The purpose of the binary 0 and 1 outputs of detector 13 is described infra. The analyzer 12 and detector 13 are described in the co-pending application of Campanella et al., titled Speech Compression System, filed July 31, 1958, bearing Serial No. 752,253, now Patent No. 3,078,345 and assigned to the same assignee as the present invention.

The output of formant analyzer 12, having an amplitude representing the frequency of the speech applied to microphone 11, is applied to differentiator 14 which derives an output signal commensurate with the rate of change of the frequency components in the speech input to microphone 11. The output of differentiator 14 is applied to a further differentiator 15 and to an integrator 16. Differentiator 15 derives a signal proportional to the second derivative of the output of differentiator analyzer 12. Differentiators 14 and 15 may comprise conventional R.C. differentiating circuits having isolating amplifiers connected in their output circuits to prevent loading.

The output signal of differentiator 15, consisting essentially of pulse spikes, is applied to a pulse generator 17 which generates constant amplitude and width pulses of short duration. A pulse is derived from generator 17 upon the generation of each spike from differentiator 15. The output of pulse generator 17 is applied in parallel to each stage of a four-stage shift register 18.

The first stage of shift register 18 is loaded with a pulse by Schmitt trigger 19. Schmitt trigger 19 is responsive to the output of absolute value circuit 21, which is fed by the output signal of integrator 16. As successive
3,268,661 pulses are derived from generator 17, the binary ones loaded into the first stage of shift register 18 are translated from the first shift register stage to the succeeding stages.

The output of each stage of the shift register 18 is coupled in parallel to the reset circuit of integrator 16 so that as each shift register stage is switched from an off to an on condition the integrator 16 is reset to its initial value of zero. Integrator 16 is returned to its initial state upon the completion of each phoneme in the speech signal applied to microphone 11 and is adapted to derive a separate value commensurate with the average value of the preceding phoneme over the phoneme duration. The output of the integrator 16 is applied through the absolute value circuit 21, which produces an output proportional to the absolute value of the integrator output, i.e. a signal of varying polarity but of varying magnitude, dependent upon the integrator output.

When the output of the integrator 16 achieves a predetermined level, Schmitt trigger 19 is activated to generate a short duration pulse which loads the first stage of shift register 18. As described supra, subsequent generation of pulses from generator 17 causes the loaded signal in the first stage of shift register 18 to be shifted from one shift register stage to the next. When the last stage of the shift register is reached, the pulse is dropped and does not circulate.

The output of pulse generator 17 is applied in parallel to time to voltage generator 22, one bit delay 23, two bit delay 24, and to the various stages of shift register 18.

Time to voltage generator 22 comprises a saw tooth generator and a sampling capacitor which stores the maximum voltage achieved by the saw tooth during each of its cycles. The saw tooth flyback and initial increase is initiated upon the application of a pulse thereto from generator 17. The step of desynchronization is derived from generator 22 commensurate in amplitude with the time between the preceding successive pulses from generator 17.

One bit delay 23 samples the amplitude of the output of differentiator 14 and stores it for a period between successive outputs of generator 17. Delay 23 includes an analog storage device which may comprise a stepping motor having a plurality of capacitors connected to its armature. The voltage derived from differentiator 14 is stored in one of the capacitors and the armature is caused to step in response to each pulse from generator 17. The output circuit of the delay element 23 is connected at a point one step from the leading to the point of the storage capacitor. Two bit delay 24 is responsive to the output of formant analyzer 12 and comprises essentially the same structure as one bit delay 23. The two bit delay 24, however, is arranged so that its output is derived at a point two steps from the point at which the capacitor or storage element was loaded.

The output signal from time to voltage generator 22 is applied to an extrapolator circuit 25 which comprises an analog multiplying circuit having a multiplication constant of substantially two. The outputs of delay element 23 and extrapolator 25 are applied to multiplier 26 which derives an output signal that is applied to the negative input of summing amplifier 27. Summing amplifier 27 is responsive at its positive input terminal to the output of formant analyzer 12 and derives a signal commensurate with the present formant frequency minus twice the derivative of the consonant slope times the time of the consonant utterance, i.e.

\[ F(t) - 2\Delta t F'(t) \vert_{t=t_1} \]

where \( F(t) \) is the present value of frequency, \( \Delta t \) is the time of consonant utterance, and \( t_1 \) is a time during the consonant utterance. This signal is indicative of the initial consonant loci frequency. The output signal of multiplier 26 is also applied to a positive input terminal of summing amplifier 28, which has a further positive input terminal responsive to the output of two bit delay 24. The output signal magnitude of amplifier 28 is proportional to the sum of the formant frequency output plus the time slope product and represents the terminal consonant loci.

Adder 27 linearly combined the difference between the frequency and the time-slope product because positive slopes at the initiation of a sound result in lower loci frequencies while negative slopes result in higher loci frequencies than the vowel frequency. In an opposite sense, amplifier 28 linearly combines the sum of the formant frequency and the time-slope product because a positive slope at the end of a sound indicates that the consonant frequency is greater than the vowel frequency while a negative slope indicates the consonant loci is less than the vowel frequency.

To insure that the signals from adding amplifiers 27 and 28 are derived only for the initial and terminal loci, respectively, gates 29 and 31 are provided. Gate 29, responsive to the output signal of only the second stage of shift register 18, derives an output signal proportional in magnitude to the output of summing amplifier 27 when the second stage of shift register 18 is in the "on" state. Gate 31, responsive to the second and fourth stages of shift register 18, provides an output voltage proportional to the output of summing amplifier 28 only when both the second and fourth stages of shift register 18 are in the "on" state. Gate 29 is not activated when the second and fourth stages of shift register are "on" because of the inhibit terminal connected to the fourth stage.

Reference is now made to FIGURE 2 of the drawings, which illustrates various wave forms derived in the circuit of FIGURE 1, for providing a better understanding of the manner in which the circuit functions. For purposes of description, it will be assumed that the signal applied to microphone 11 corresponds with the sound "bead," phonetically sounded as b e d. The frequency locus of "b" is approximately 700 cycles and the duration of the phoneme is approximately 0.10 of a second, as depicted in FIGURE 2A. The first half or 0.05 second of the phoneme are silent so that the first sound heard is approximately 1400 cycles. The frequency increases linearly to approximately 2300 cycles at which time the vowel "a" is formed. The vowel continues for approximately 0.2 of a second until T=0.3 second. At this time, enunciation of the phoneme "d" begins and the frequency begins to fall linearly towards 1800 cycles. However, the last half of the phoneme "d" is silent so that the lowest frequency uttered in enunciating "d" is approximately 2050 cycles.

To compute the initial and terminal loci of the consonant utterances, the linear or straight line slopes of the consonants "b" and "d" must be extrapolated to the initiation and termination thereof from the substantially discontinuous points which the linear frequency variations intercept the vowel frequency.

The output of formant analyzer 12, illustrated in the time v. frequency characteristic of FIGURE 2A is applied to differentiator 14 which derives an output signal, indicated in FIGURE 2B of the drawings. Since the output of formant analyzer 12 is zero in the interval 0 to T=0.05 second, the differentiator output is likewise zero. A substantially constant amplitude signal 32 is derived from the differentiator in response to the constant time rate of change of the consonant "b" as it varies from an utterance of 1500 to 2300 c.p.s. The amplitude of wave form 32 derived from differentiator 14 measures both in polarity and magnitude with the slope. When the output of formant analyzer 12 is constant, zero output 33 is derived from differentiator 14. As the consonant "d" is uttered, the output of differentiator 14 decreases from the zero portion 33 to a finite constant negative amplitude 34, proportional to the slope of the
When the consonant utterance is terminated, the output of differentiator 14 is returned to zero.

The output signal of differentiator 14, as indicated in FIGURE 2B is applied to cascaded differentiator 15 which derives output spikes as indicated in FIGURE 2C. For each sudden variation in the output of differentiator 14 as depicted in FIGURE 2B a spike is derived by differentiator 15. These spikes from differentiator 15 are applied to pulse generator 17 which derives the constant amplitude and width pulses, indicated in FIGURE 2D. The pulses in FIGURE 2D occur at varying times, dependent upon the time of occurrence of the voltage pulses from differentiator 15.

The time to voltage generator 22, derives an output voltage proportional to the time of occurrence between the succeeding pair of pulses from generator 17. Thus, the amplitude of wave form 35, FIGURE 2E which is generated in response to the time between pulses 36 and 37, FIGURE 2D, is 3/4 as great as the amplitude of wave form 38 which is generated in response to the time interval between pulses 37 and 39, FIGURE 2D.

To insure multiplication of the proper time and slope components, the one bit delay 23 is inserted to derive the wave form of FIGURE 2F. The wave form portion 41, FIGURE 2F, is equal in amplitude to the wave form 32, FIGURE 2B but extends for a time period between the occurrence of pulses 37 and 39, FIGURE 2D. Similarly, the portions 42 and 43, FIGURE 2F correspond in amplitude with the portions 33 and 34, FIGURE 2B, but correspond in duration with the time between pulses 39 and 44 and 44 and 45, FIGURE 2D.

The wave form of FIGURE 2F is multiplied by the extrapolation factor in multiplier 25 and further multiplied with the output of one bit delay element 23. Accordingly the wave form of FIGURE 2G is derived in response to the product of the wave forms of FIGURES 2E and 2F.

To derive the initial voltage, the wave forms of FIGURES 2A and 2G are subtracted and the initial locus of the consonant "d" is found to be proportional to the amplitude at portion 44 of the wave form depicted in FIGURE 2I. The wave form portion 44 is fed through gate 29 at the appropriate time to provide an indication of the initial consonant locus value.

To derive the terminal locus, the output wave form of analyzer 12 is delayed two bits in element 24 so that the wave form indicated in FIGURE 2H is derived therefrom. Actually, the output of delay 24 comprises a series of samples of the wave form derived from analyzer 12. This is suitable for the purposes of the present invention since we are interested only in vowel phoneme of a substantially constant value for fairly long durations of time.

To calculate the terminal loci, the wave forms of FIGURES 2H and 2F are added to produce the wave form indicated in FIGURE 2I. This wave form has a positive portion 45 commensurate in amplitude with the frequency locus of the consonant "d."

Control of gates 29 and 31 is accomplished through the action of shift register 18. When the output of differentiator 14, FIGURE 2B, is of at least a predetermined amplitude for a sufficient time period, the output of integrator 16 achieves a predetermined voltage. This voltage is rectified in absolute value circuit 21 and applied to Schmidt trigger 19. The Schmidt trigger generates a short duration pulse which loads shift register 18.

When the word "feed" loading occurs in the first phoneme of the word prior to the occurrence of pulse 37 from generator 17. The Schmidt trigger output pulse causes the first stage of shift register 18 to be loaded to an "on" state. When pulse 37 is derived from generator 17, the "on" state of the first stage of shift register 18 is transferred to the second stage and the first stage is returned to its "off" state. The shift register remains in this condition throughout the pronunciation of the phoneme "e" because the output of differentiator 14 is zero. When, however, the pulse 37 was generated, integrator 16 was reset to its initial value so that Schmidt trigger 19 could be triggered again.

When the second stage of shift register 18 is in the "on" condition, a gating pulse is applied to open gate 29 to permit the output of summing amplifier 27 to be supplied through it. When pulse 39 is derived from generator 17, the shift register is activated so that its third stage is in the "on" condition while the other three stages are in the "off" state.

During the period of time between the occurrence of pulses 39 and 41, a finite output voltage, indicated by portion 34 of the wave form depicted in FIGURE 2B, is applied to the integrator 16. The integrator is charged to a sufficient value by the wave form portion 34 to reactive Schmidt trigger 19 and load the first stage of shift register 18 to its "on" state. Accordingly, upon the occurrence of pulse 44, the first and third stages of shift register 18 are loaded or are in the "on" condition. Pulse 44 then causes the stages of shift register 18 to be shifted so that the second and fourth stages are in the "on" condition.

The output signal from the second and fourth stages are coupled to initial and terminal loci gates 29 and 31. Gate 31 passes the output signal of summing amplifier 28 only when the second and fourth stages of shift register 18 are in the "on" condition so that the output of integrator 16 is not effective. The output of integrator 16 is connected to the output of shift register 18. This binary one is derived from delay element 13. The binary one is applied in parallel to time-to-voltage generator 22, integrator 16 and shift register 18. The binary one serves to react each of these elements to their initial state. In the time-to-voltage generator 22 and integrator 16 the time-to-voltage generator 22 and integrator 16 is accomplished by short circuiting the capacitors in the wave form generating circuits while shift register 18 is responsive to the binary one so that each of its stages is reset to the "off" condition.

It is to be understood that digital as well as analog apparatus may be employed for the computation functions indicated. Also other systems may be utilized to gate the initial and terminal loci information from summing amplifiers 27 and 28. In a complete speech recognition system, a system of consonant recognition of initial and terminal consonants coupled about a word may be provided for gating the initial and terminal loci signals to the appropriate read-out apparatus.

While I have described and illustrated one specific embodiment of my 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.

I claim:
1. In a system for determining the frequency loci of voiced consonants in a speech signal comprising means responsive to said speech signal for deriving a first signal representing in magnitude the frequency of the speech signal, means for transmitting said signal to said speech signal for deriving a second signal representing the rate of change of the first signal, means responsive to said first signal for deriving third signals representing phoneme time duration, and means for multiplying said second and third signals together.
2. In a system for determining the frequency loci of voiced consonants in a speech signal comprising means responsive to said speech signal for deriving a first signal...
representing in magnitude the frequency of the speech signal, differentiating means responsive to said first signal for deriving a second signal representing the rate of change of the first signal, means responsive to said first signal for deriving third signals representing phoneme time durations, means for multiplying said second and third signals together for deriving a fourth signal, and means for algebraically adding said first and fourth signals.

3. A computer for ascertaining the value at time \( t_1 \) of a function \( F(t) \) represented by an input signal, comprising means responsive to said input signal for computing

\[
\frac{dF(t)}{dt}
\]

at a time when the frequency versus time characteristic of \( F(t) \) is substantially a straight line of non-zero slope, means responsive to said input signal for computing \( t_2 - t_0 \), \( t_2 \) being a time other than \( t_0 \), wherein \( F(t) \) has said substantially straight line characteristic between times \( t_1 \) and \( t_2 \), means responsive to both of said computing means for calculating

\[
(t_2 - t_1) \frac{dF(t)}{dt}
\]

and means for linearly combining \( F(t_2) \) and

\[
(t_2 - t_1) \frac{dF(t)}{dt}
\]

4. A computer for ascertaining at a time \( t_1 \) the value of a function \( F(t) \) represented by a signal, said function having substantially straight line variations between substantially discontinuous points at times \( t_1, t_2, \ldots, t_n \) comprising first means responsive to said signal for computing

\[
\frac{dF(t)}{dt}
\]

in the interval between \( t_1 \) and \( t_2 \), second means responsive to said signal for determining \( dT \), the time between \( t_1 \) and \( t_2 \), third means responsive to said first and second means for computing

\[
\Delta T \frac{dF(t)}{dt}
\]

and combining means responsive to said third means and said signal for algebraically adding \( F(t_2) \) and

\[
\Delta T \frac{dF(t)}{dt}
\]

5. The system of claim 4 wherein said combining means derives an output signal proportional to

\[
F(t_2) + \Delta T \frac{dF(t)}{dt}
\]

6. The system of claim 4 wherein said combining means derives an output signal proportional to

\[
F(t_2) - \Delta T \frac{dF(t)}{dt}
\]

7. A system for ascertaining the frequency loci of voiced consonants in a speech signal comprising means responsive to said signal for deriving another signal varying in magnitude in accordance with frequency components of said speech signal, said another signal being substantially constant when representing a vowel and having a substantially constant rate of change when representing a voiced consonant, computing means responsive to said another signal for deriving a further signal representing the product of said rate of change and the total time of occurrence of the voiced consonant, and means responsive to said another and further signals for combining said said rate of change and said product.

8. The system of claim 7 wherein said combining means adds said constant amplitude and said product to derive a signal representing the terminal frequency locus of the voiced consonant.

9. The system of claim 8 including means for gating the terminal frequency locus signal in response to the occurrence of a terminal consonant in said speech signal.

10. The system of claim 7 including means for gating the initial frequency locus signal in response to the occurrence of an initial consonant in said speech signal.

11. The system of claim 7 further including means for deriving a control signal in response to unvoiced and silent conditions in said speech signal, and means responsive to said control signal for resetting said computing means and said combining means to a predetermined condition.

12. The system of claim 7 wherein said combining means subtracts said constant amplitude and said product from one another to derive a signal representing the initial frequency locus of the voiced consonant.

13. A system for computing the frequency loci of voiced consonants in a speech signal, voiced consonants including silent intervals removed in time from the vowel with which the consonant is coupled and voiced intervals contiguous in time with said vowel, comprising means responsive to said signal for deriving another signal varying in magnitude in accordance with frequency components of said speech signal, said another signal being substantially constant when representing a vowel and having a substantially constant rate of change when representing a voiced consonant, computing means responsive to said another signal for deriving a further signal having a magnitude varying in accordance with the time rate of change of said another signal, means responsive to said another signal for deriving an indication of the time duration of said voiced intervals, means for modifying said indication in accordance with the ratio of said voiced and silent intervals, means for multiplying said modified indication and said further signal to derive a product signal, and means for linearly combining said product signal with said another signal when it is substantially constant and representing a vowel.

14. In a system for determining the frequency loci of voiced consonant phonemes in a speech signal comprising means responsive to said input signal for deriving a signal, \( F(t) \), representing in magnitude the frequency of the speech signal, means responsive to \( F(t) \) for computing

\[
\frac{dF(t)}{dt}
\]

at a time when the frequency versus time characteristic of \( F(t) \) is substantially a straight line of non-zero slope, said last named means deriving a signal represented by

\[
\frac{dF(t)}{dt}
\]

means responsive to \( F(t) \) for computing the phoneme duration as \( t_2 - t_1 \), wherein \( t_1 \) and \( t_2 \) are times at the beginning and end of each voiced consonant and \( F(t) \) has said substantially straight line characteristic between \( t_1 \) and \( t_2 \), means responsive to both said computing means for calculating

\[
(t_2 - t_1) \frac{dF(t)}{dt}.
\]

and means for linearly combining \( F(t_2) \) and

\[
(t_2 - t_1) \frac{dF(t)}{dt}.
\]

(References on following page)
### References Cited by the Examiner

<table>
<thead>
<tr>
<th>Patent Number</th>
<th>Date</th>
<th>Inventor</th>
<th>Primary Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>2,243,526</td>
<td>5/1941</td>
<td>Dudley</td>
<td>179-1</td>
</tr>
<tr>
<td>2,243,527</td>
<td>5/1941</td>
<td>Dudley</td>
<td>179-1</td>
</tr>
<tr>
<td>2,652,194</td>
<td>9/1953</td>
<td>Hirsch</td>
<td>253-193</td>
</tr>
<tr>
<td>2,666,576</td>
<td>1/1954</td>
<td>Hirsch</td>
<td>235-193</td>
</tr>
<tr>
<td>2,960,268</td>
<td>11/1940</td>
<td>Chope et al.</td>
<td>235-192</td>
</tr>
</tbody>
</table>

### Other References


KATHLEEN H. CLAFFY, Primary Examiner.

WILLIAM C. COOPER, ROBERT H. ROSE, Examiners.

A. J. SANTORELLI, R. MURRAY, Assistant Examiners.