

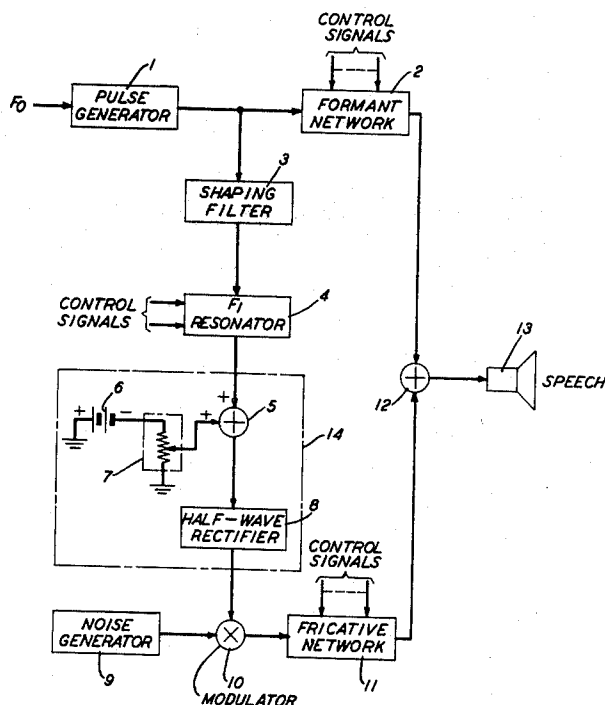
[72] Inventor **Lawrence R. Rabiner**
Rockaway, N.J.
 [21] Appl. No. **668,560**
 [22] Filed **Sept. 18, 1967**
 [45] Patented **Dec. 22, 1970**
 [73] Assignee **Bell Telephone Laboratories, Incorporated**
Murray Hill, N.J.
a corporation of New York

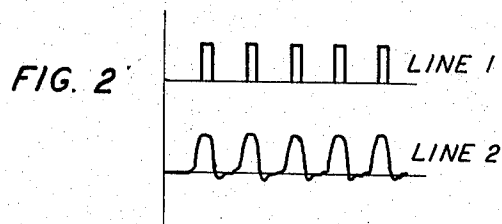
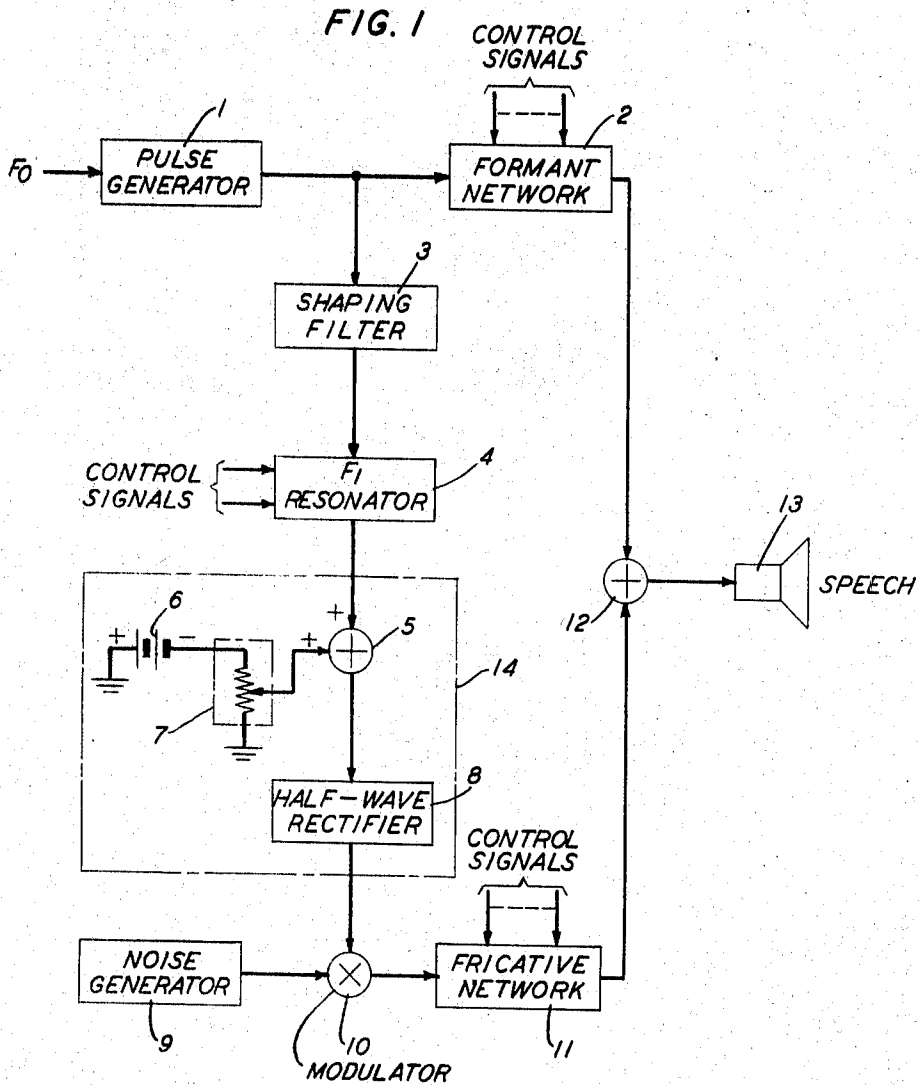
[56] **References Cited**
UNITED STATES PATENTS
 3,395,249 7/1968 Clapper..... 179/1
 3,190,963 6/1965 David et al..... 179/1
Primary Examiner—Kathleen H. Claffy
Assistant Examiner—Douglas W. Olms
Attorneys—R. J. Guenther and William L. Keefauver

[54] **VOICED FRICATIVE SYNTHESIZER**
5 Claims, 2 Drawing Figs.

[52] U.S. Cl..... 179/1
 [51] Int. Cl..... G101 1/10
 [50] Field of Search..... 179/1AS,
 15.55

ABSTRACT: A speech synthesizer is disclosed in which voiced fricatives are synthesized from a voiced component and an unvoiced component, the unvoiced component being generated only during those times the voiced component exceeds a selected amplitude.





INVENTOR
 BY **L.R. RABINER**
G.E. Hirsch Jr.
 ATTORNEY

VOICED FRICATIVE SYNTHESIZER

BACKGROUND OF THE INVENTION

This invention relates to the synthesis of speech and, in particular, to the synthesis of voiced fricatives.

Speech synthesizers are devices which produce speech from coded signals representing the fundamental components of speech. Best known are vocoder synthesizers which produce replicas of input speech from narrowband control signals derived from the input speech. But recently, speech synthesizers for so-called "voice response systems," which provide artificially generated spoken responses to coded requests for information, have become of increasing interest.

Most current synthesizers produce speech of low quality. Partly, this is because a high quality synthesizer must reproduce accurately the complicated process by which a person speaks. This process is not completely understood. As a result, synthesized speech is often highly unnatural with rough transitions between sounds, unrealistic sound durations, and disconcerting background noise.

A particularly difficult sound to synthesize has been the "voiced fricative." Voiced fricatives are produced by passing quasi-periodic puffs of air, released from the lungs by vibrating vocal cords, through constrictions in the vocal tract. For a volume flow of air above a certain minimum flow, the air flow through the constriction becomes turbulent. Thus, an analysis of voiced fricatives shows that they contain both turbulent noise energy—the so-called unvoiced component—and quasi-periodic energy—the so-called voiced component. An accurate synthesis of voiced fricatives must reproduce both of these components.

SUMMARY OF THE INVENTION

This invention improves the quality of synthesized voiced fricatives by accurately simulating the processes in the human vocal tract which produce both the unvoiced and voiced components of voiced fricatives. As a result, the synthesized voiced fricatives sound highly natural.

According to this invention, quasi-periodic pulses from a pulse at a desired fundamental frequency are sent along two parallel paths. In one path, these pulses are passed through a formant network to produce a signal representing the voiced component of a voiced fricative. In the other path, these pulses are first shaped to resemble, as a function of time, the air flow through the vocal cords. A threshold circuit and a half-wave rectifier pass only those portions of the shaped pulses which exceed a selected amplitude. The resulting rectified pulse portions then modulate the output signal from a noise generator. Finally, a spectral shaping network produces, from the modulated noise signal, a signal representing the unvoiced component of the voiced fricative. The voiced fricative is synthesized by summing the output signals from the first and second paths and passing the resulting sum signal through a loudspeaker.

The threshold circuit and the half-wave rectifier simulate the fact that turbulence is introduced by the constriction only when the volume flow of air exceeds a certain minimum value. Otherwise, no turbulence is produced. Hence the synthesizer of this invention duplicates the actual mechanism by which the vocal tract produces voiced fricatives.

This invention may be more fully understood from the following detailed description taken together with the drawings.

BRIEF DESCRIPTION OF THE DRAWING

FIG. 1 shows in schematic from the network used in this invention to produce voiced fricatives; and

FIG. 2 shows the shape of certain pulses at selected points in the network of FIG. 1.

DETAILED DESCRIPTION

Linguists consider speech to be composed of a variety of basic sounds called phonemes. Phonemes themselves are further subdivided according to their manner and place of production within the vocal tract. In general, as explained by J. L. Flanagan in his book "Speech Analysis, Synthesis and Perception," Academic Press Inc. 1965, Chapter 1, English speech (American dialect) is composed of vowels, normally produced exclusively by vocal cord excitation of the vocal tract, and consonants, both voiced and unvoiced, produced by various combinations of quasi-periodic and incoherent excitation of the vocal tract. In particular, the voiced fricatives are consonants produced by combinations of vocal cord and noise excitation of the vocal tract.

As explained by Flanagan, sounds are produced by expelling air from the lungs along the trachea through the vocal cords into the vocal tract and out of the mouth. The vocal tract is an irregular tube whose cross-sectional area is controlled by movement of the lips, jaw, and velum—so-called soft palate—which controls access to the nasal cavity. When the velum is open, sounds radiate from both the mouth and the nostrils.

To produce a voiced fricative, quasi-periodic puffs of air, released from the lungs by the vocal cords, are passed through a constriction created at some point in the vocal tract, by, for example, the tongue. For a volume flow of air above a critical minimum value, the constriction introduces turbulence, with the result that the voiced fricative contains both quasi-periodic and incoherent energy components. An accurate synthesis of a voice fricative, therefore, must take into consideration both the vocal cord quasi-periodic excitation and the incoherent noise noiselike energy superimposed on this excitation by the vocal tract constrictions. A system to do this is shown in FIG. 1.

Pulse generator 1 produces pulses, shown on line 1, FIG. 2, at a selected frequency F_0 , which, in general, can vary with time. The pulses from generator 1 are sent along two paths.

First, they are passed through formant network 2. Network 2 modifies the frequency components of these pulses by the vocal tract transfer function so as to produce a signal representing the voiced component of a voiced fricative with the correct amplitude spectrum. A voiced spectrum synthesizer suitable for use as formant network 2 in the practice of the invention is described in E. E. David, Jr. et al. U.S. Pat. No. 3,190,963, granted June 22, 1965, particularly with reference to FIG. 5. Generation of control signals for energizing the network is described in the patent, particularly with regard to FIG. 1.

Second, these pulses are passed through shaping filter 3, resonator 4, threshold circuit 14, modulator 10 and network 11 to produce a signal representing the unvoiced component of a voiced fricative.

Filter 3 produces output pulses closely resembling in shape output pulses from the vocal cords. The shape of these pulses is described by J. L. Flanagan in an article entitled "Some Properties of the Glottal Sound Source" published in Volume 1, Journal of Speech and Hearing Research, pages 99 to 116 (1958). These shaped output pulses, shown on line 2 of FIG. 2, are then passed through resonator 4 which represents the acoustical characteristics of the vocal tract "back cavity"—the volume between the vocal cords and the point of constriction in the vocal tract. This resonator essentially reproduces the first formant of the speech spectrum. A typical resonator is described, for example, on page 184 of the above-cited book by Flanagan. FIG. 3A of the David et al. U.S. Pat. No. 3,190,963, shows a similar resonator and the associated text discusses the form of suitable control signals and the manner of generating them. Resonators to take into account higher frequency formants of the speech spectrum are unnecessary because the 12 db. per octave attenuation of the vocal tract frequency characteristics makes the amplitudes of the higher frequency formants quite small relative to the am-

plitude of the first formant. In addition, the amplitudes of the high frequency components of the pulses from the vocal cords are smaller than the amplitudes of the low frequency components of these pulses.

Now, in any flow constriction, turbulence is not produced until the volume velocity of the flowing fluid exceeds some threshold value. This is true in the vocal tract. Thus, threshold circuit 14, composed of summing network 5, together with battery 6, attenuator 7, and half-wave rectifier 8, simulates the threshold at which the flow through a constriction changes from laminar to turbulent. Half-wave rectifier 8 passes only the positive components of the output signal from summing network 5. But summing network 5, in turn, produces a positive output signal only when the output signal from resonator 4 exceeds the absolute value of the bias voltage from battery 6 as adjusted by attenuator 7. Thus, in effect, rectifier 8 passes only those portions of the pulses from resonator 4 which exceed a selected amplitude. Because the volume flow at which the flow through the constriction becomes turbulent is a function of the diameter of the constriction, attenuator 7 can be adjusted as a function of time, if desired, to represent changes in constriction size with time.

The output signal from rectifier 8 multiplies the output noise signal from noise generator 9 in modulator 10. Modulator 10, as a result, produces an output signal only when half-wave rectifier 8 produces an output signal. Thus, the output signal from modulator 10 represents quasi-periodic bursts of noise energy. Fricative network 11 shapes this modulated noise energy by superimposing the formant structure of the remainder of the vocal tract downstream from the constriction on this energy. A suitable unvoiced spectrum synthesizer for generating unvoiced component signals is described in the aforementioned David et al. patent. Summing network 12 combines the modulated noise from fricative network 11 with the shaped quasi-periodic energy from formant network 2 to produce an electrical signal representing the synthesized voiced fricative. Loudspeaker 13 converts this electrical signal into an acoustic sound.

Of course, the vocal tract is, in general, continuously changing in shape. Thus, both the center frequency and the bandwidth of resonator 4, which produces the shape of the first formant of the vocal tract, are controlled as a function of time by control signals from sources not shown. Such control signals might, for example, be derived in a well known manner from a table of stored formant data similar to the table shown on page 208 of the above-cited book by Flanagan. In addition, the center frequencies and bandwidths of the formants produced by formant network 2 and fricative network 10 are also controlled in the same manner. Network 2, for example, which produces the shape of a selected number of the formants in the vocal tract, is controlled so that the center frequencies and bandwidths of the first three formants can be varied in response to control signals. Network 10 is similarly controlled.

Voiced fricatives synthesized by the network shown in FIG. 1 contain a quasi-periodic, or voiced component, and, if the quasi-periodic component exceeds a selected magnitude, an aperiodic or unvoiced component. The resulting synthesized voiced fricatives are superior in quality to the voiced fricatives synthesized by prior art systems. Yet the voiced fricative synthesizer itself is quite simple.

Other embodiments incorporating the principles of this invention will be apparent from this disclosure to those skilled in the art of speech synthesis.

I claim:

1. Apparatus for synthesizing voiced fricative sounds which comprises:

means for producing first and second signals representing the voiced and the unvoiced components, respectively, of a selected voiced fricative; and

means, supplied with said first signal and, whenever said first signal exceeds a selected amplitude threshold, with said second signal, for combining said signals to produce a composite signal representing said selected voiced fricative.

2. Apparatus for the synthesis of voiced fricatives which comprises:

means for producing pulse signals at selected intervals;

means for processing said pulses to produce a first signal representative of the voiced component of a selected voiced fricative;

means for generating a second signal containing noise energy during the time said pulse signals exceed a selected amplitude and containing no energy at all other times, said second signal representing the unvoiced component of said voiced fricative; and

means for combining said first and second signals to produce said voiced fricative.

3. Apparatus as in claim 2 in which said means for processing comprises a network for superimposing on said pulse signals the spectral shape of the vocal tract formants.

4. Apparatus as in claim 2 in which said means for generating comprises:

a shaping network for converting said pulse signals into a shaped signal resembling, as a function of time, the air flow through vibrating vocal cords;

a resonator for superimposing on said shaped signal the formant structure of the back cavity of the vocal tract;

a threshold circuit for passing only those portions of said shaped signal from said resonator which exceed a selected amplitude;

a noise source;

means for modulating the noise from said source with the output signals from said threshold circuit to produce bursts of noise during the time said shaped signal from said resonator exceeds said selected amplitude; and

a fricative network for superimposing on said bursts of noise the formant structure of a selected portion of the vocal tract.

5. Apparatus as in claim 4 in which said threshold circuit comprises:

a battery, the positive terminal of which is connected to ground;

a variable attenuator connected between the negative terminal of said battery and ground, said attenuator possessing a movable output tap; and,

a summing network, one input terminal of which is connected to said movable output tap from said attenuator and the other input terminal of which is connected to receive the output signals from said resonator, said summing network producing a positive output signal only when the output signal from said resonator exceeds the absolute value of the bias voltage on the output tap of said attenuator; and

a half-wave rectifier for passing the positive output signals from said summing network, said positive output signals comprising the output signals from said threshold circuit.