

[54] SPEECH-SYNTHEISER

[75] Inventor: **Claude Dechaux**, Paris, France

[73] Assignee: **Thomson-CSF**, Paris, France

[22] Filed: **Mar. 3, 1972**

[21] Appl. No.: **231,558**

[30] Foreign Application Priority Data

Mar. 26, 1971 France 71.10824

[52] U.S. Cl. **179/1 SA**

[51] Int. Cl. **G101 1/00**

[58] Field of Search **179/1 SA, 1 SB, 15.55 R**

[56] References Cited

UNITED STATES PATENTS

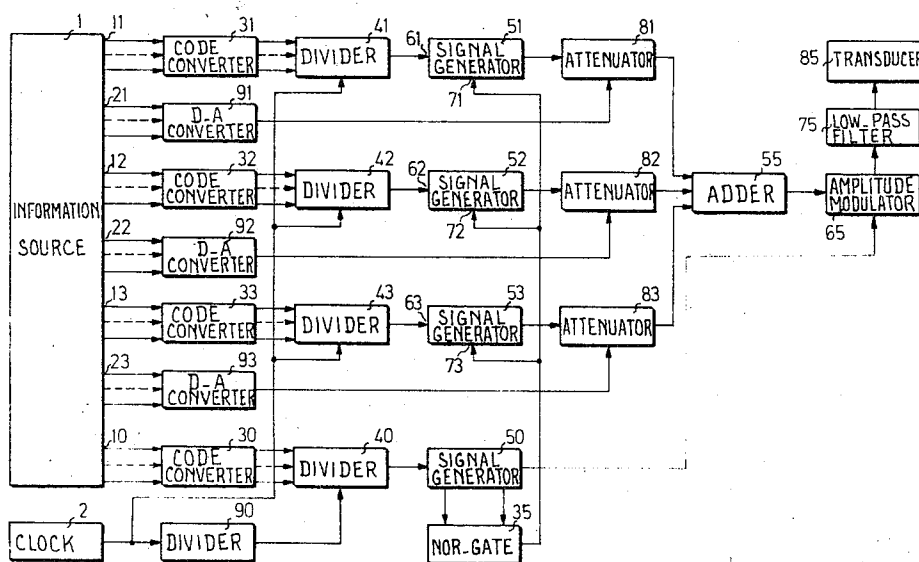
3,268,660	8/1966	Flanagan.....	179/1 SA
3,394,228	7/1968	Flanagan.....	179/1 SA
3,491,205	1/1970	Focht.....	179/15.55 R
3,499,991	3/1970	Cassel.....	179/1 SA
3,532,821	10/1970	Nakata.....	179/1 SA

Primary Examiner—Kathleen H. Claffy
Assistant Examiner—Jon Bradford Leaheey
Attorney, Agent, or Firm—Cushman, Darby & Cushman

[57] ABSTRACT

In a synthesiser which, for each sampling period, reconstitutes a language element by means of three sinusoidal components obtained with the help of variable-frequency generators and variable-attenuators, those components are simultaneously subject to predetermined rephasing operations carried out at an auxiliary frequency identified with the pitch frequency (or frequency of vibration of the voice) at the time of emission of vowels or voiced consonants. This auxiliary frequency is delivered by a further variable-frequency generator. In addition, the signal representing the sum of these components is amplitude-modulated by a modulating signal at the auxiliary frequency.

2 Claims, 2 Drawing Figures



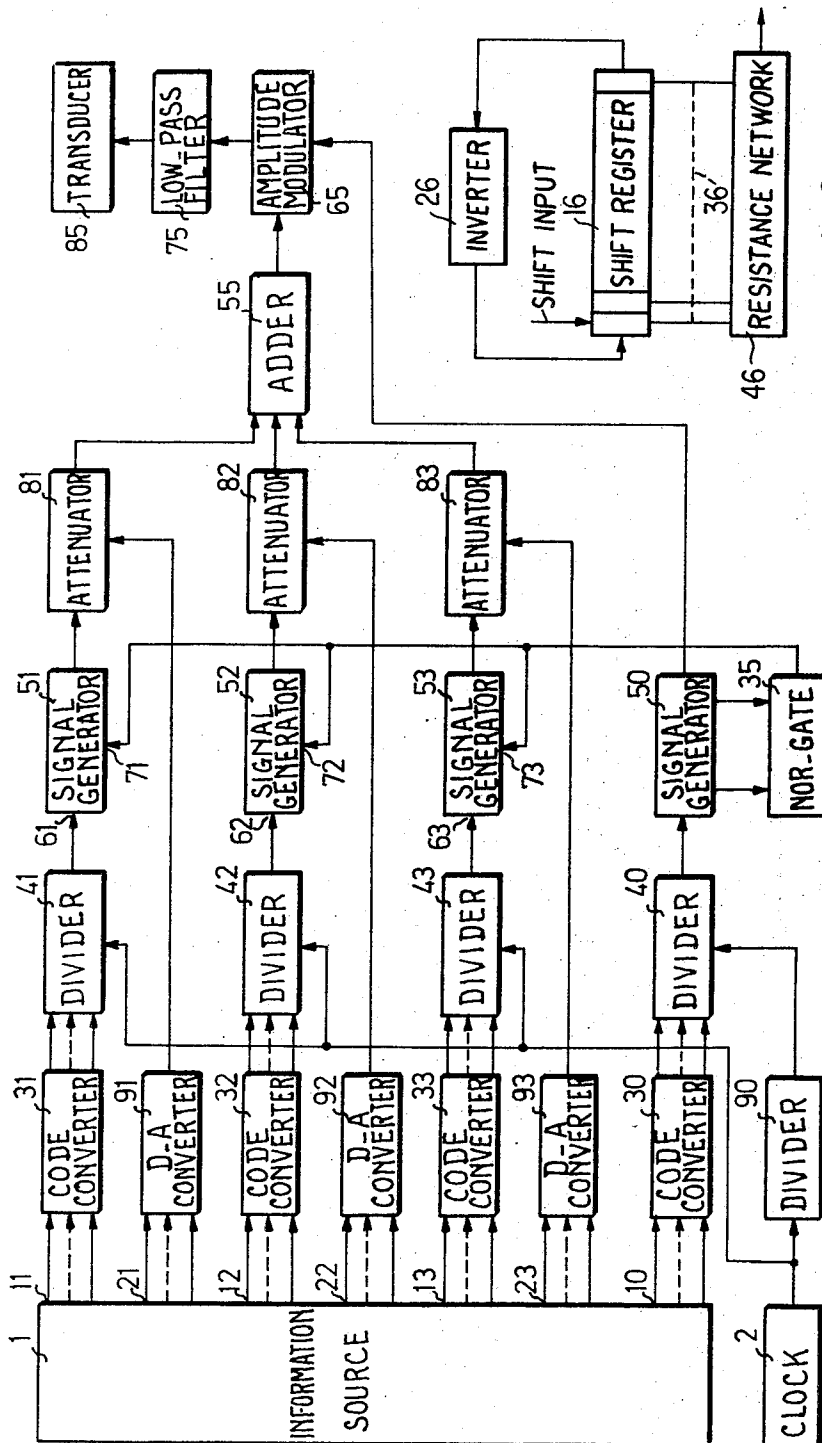


FIG. 1

FIG. 2

SPEECH-SYNTHESISER

The present invention relates to an improvement in speech-synthesisers supplied with a recurrence periodicity of T , corresponding to a sampling periodicity T , with digital information translating an information that will be referred to hereinafter as the main information, the latter making it possible to reconstitute in an approximate manner a language element by adding up, in respect of each sampling period, a certain number p , equal to or less than a fixed number n , of sinusoidal signals, whose frequency and amplitude form the main information, and which will be referred to here as "main components."

The aforesaid kind of main information or data, was proposed in French Pat. No. 69.15712, publication No. 2,044,290. Briefly, a main component is a portion, in the course of time, of a formant, a formant being defined as a succession, (in time) of spectral components whose frequencies are identical or vary little, and corresponding to an absolute or relative maximum of energy in the speech spectrum. The formant portions which are transmitted with each speech sampling operation are determined in accordance with criteria set out in the aforementioned Patent (the precise way in which the main information is selected being, however, outside of the scope of the present invention which only concerns the use of it in a synthesiser).

Because analysis of the sound is effected by means of a bank of filters, the frequencies of the main components are only transmitted to within a certain degree of accuracy.

The object of the present invention is the utilisation in synthesisers of the above-mentioned kind, of an auxiliary information relating to the person speaking and making it possible, to a certain extent, to identify him.

This auxiliary information is the "pitch information." It is constituted during the emission of vowels and voiced consonants by a frequency which is the frequency of vibration of the vocal chords of the person speaking, and this will be referred to abbreviatedly as the "pitch frequency;" it generally ranges between 80 and 350 c/s. The spectral components relating to the vowels and voiced consonants are harmonics of this pitch frequency, unlike the case with the unvoiced consonants.

A variety of devices have been described for the measurement of the pitch frequency at emission. A corresponding bibliography is listed in the work by J. L. Flanagan, entitled "Speech Analysis, Synthesis and Perception," published by Springer-Verlag, Berlin-Heidelberg-New York, 1965.

The simplest of these devices is of the peak detector type:

The speech signal is applied to a bandpass filter or, even safer, to two bandpass filters, normally producing a signal which is amplitude-modulated at the pitch frequency at the time of emission of vowels or voiced consonants; each of these filters is followed by an amplitude detector itself followed by a peak detector, the frequency of the peaks corresponding to the pitch frequency when the latter appears in the signal; in the contrary case, the measuring device produces an output frequency which fluctuates very widely.

In many digital speech transmission systems, the frequency produced by the measuring device and referred

to here as the auxiliary frequency, is permanently transmitted through an auxiliary channel, irrespective of whether it is in fact the pitch frequency or not, but there is also additionally transmitted a special signal whose production requires complex equipment and which indicates if the auxiliary frequency is the pitch frequency, in which case the speech synthesis is carried out by means of a generator producing harmonics of this frequency, which is obviously only reconstituted to within the limits dictated by the quantizing operations; in the contrary case, the auxiliary frequency is not utilised and synthesis is effected with the help of a noise generator.

In the speech synthesiser described in the aforementioned French Patent, where synthesis is carried out in accordance with entirely different principles by means of a small number of "main components," the pitch information was not used.

The present invention makes it possible to exploit the pitch information by carrying out predetermined, simultaneous rephasing on all the main components, these predetermined rephasing operations being effected at the pitch frequency; they enable the sum signal of the main components to be rendered periodic at this frequency.

According to invention, there is provided a speech synthesiser designed to carry out speech synthesis on the basis of a periodic information comprising a main information relating to a language element, and constituted by the frequencies and amplitudes of p sinusoidal components hereinafter referred to as main components (p being a variable number which is equal at the most at a fixed number n greater than 1), and an auxiliary information constituted by a frequency, hereinafter referred to as the auxiliary frequency, which, during the emission of vowels and voiced consonants, is the frequency of vibration of the vocal chords of the speaker, referred to as the pitch frequency, said synthesiser comprising: n variable-frequency generators and n amplitude control devices respectively associated with said n generators, said n generators and said amplitude-control devices being controlled by said main information in order to reconstitute said main components; an adder, having an output, for delivering the sum of said reconstituted main components; a device hereinafter referred to as rephasing device, for rephasing, i.e., restoring to a predetermined phase, each of said reconstituted main components, said rephasing operations taking place simultaneously on said components and being controlled by said auxiliary information for taking place, and this at said auxiliary frequency, at least when said auxiliary frequency is the pitch frequency.

The invention will be better understood and other of its features rendered apparent, from a consideration of the ensuing description and the drawing relating thereto, FIG. 1 being the diagram of a preferred embodiment of the synthesiser in accordance with the invention and FIG. 2 illustrating an element of the diagram of FIG. 1.

In FIG. 1, the information source 1 is the source designed to supply the synthesiser.

This source may, for example, be the final stage of a receiver device which produces the information required by the synthesiser, in the form of parallel binary signals. Each of these signals is maintained at the out-

puts of the source 1 for a period T equal to the analysis period.

It has been assumed here that the maximum number n of main components is three.

The source 1 thus has seven multiple outputs (each multiple output having a plurality of terminals each of which corresponds to a binary digit), namely the outputs 11, 12 and 13 respectively delivering the values of the frequencies F_1 , F_2 and F_3 of the main components, the outputs 21, 22 and 23 delivering the values of the corresponding amplitudes A_1 , A_2 and A_3 , and, finally, the output 10 delivering the value of the auxiliary frequency f . As will be seen there is no need to deliver to the apparatus a special signal for indicating whether or not the auxiliary frequency being received is in fact the pitch frequency.

The outputs 11, 12 and 13 are respectively connected to three code converter circuits 31, 32, 33 which respectively produce in relation to the digital data representing F_1 , F_2 , F_3 (these data are generally constituted by the identification number of an analysis filter), the numbers

$$N_1 = F_0/2qF_1, N_2 = F_0/2qF_2, N_3 = F_0/2qF_3$$

where F_0 is a frequency very much higher than the acoustic frequency band (300 to 3,500 c/s for example) analysed at emission, and where q is a fixed whole number in the order of 10 for example.

Three variable dividers 41, 42, 43 of the counter type, receive the pulses from a clock 2 at frequency F_0 . These dividers are respectively provided with multiple control inputs respectively connected to the outputs of the code converters 31, 32 and 33, and their respective outputs are connected to the frequency control inputs 61, 62 and 63 of three signal generators 51, 52 and 53 which they respectively supply with pulses of frequency $2qF_1$, $2qF_2$ and $2qF_3$.

Each of the three generators (FIG. 2) is of a known type essentially comprising a shift register 16 with q stages (only the first two and the last one of which have been symbolically shown by means of partition lines), the output of this register being connected to the input through a level inverter 26 which converts a "1" digit to a "0" digit, and vice-versa. The q stages are respectively provided with auxiliary outputs 36 connected to a circuit 46 comprising a resistance network supplied from a source. The register is initially loaded in a suitable manner, for example with q "0" digits. The register load varies with each shift pulse and the position of the "1" digits in the register determines which resistors contribute to the formation of the output signal from the generator. This output signal is thus constituted by a step voltage. The resistance network is designed so that for $2q$ shift pulses produced at a fixed recurrence periodicity, the envelope of the signal constitutes one cycle of a sinusoidal waveform (disregarding a direct component).

For a periodic shift pulse train, the generator produces a step signal comprising a certain number of cycles, which simply has to be fed into a low-pass filter in order to convert it to a sinusoidal signal, the direct component normally being thereafter eliminated through an amplification process for example.

Generators of this kind have been described in an article by Anthony C. Davies, entitled "Digital Generation of Low-Frequency Sine Waves" published in IEEE Transactions of Instrumentation and Measurement,

volume IM 18, No. 2, June 1969, pages 97 to 105. These generators, and more generally speaking all generators of periodic signals, which utilise a resistance network in which the resistors are switched via a shift-register, are referred to in the claims as "shift-register and resistance network generators."

This being so, in the drawing the inputs 61, 62 and 63 of the generators 51, 52 and 53 are the inputs of the shift device of the registers involved.

It is easy to see that under these circumstances, the three generators will respectively produce output voltages whose envelopes (apart from a direct component) are portions of sinusoidal waveforms having respective frequencies F_1 , F_2 and F_3 .

It is likewise easy to appreciate that the load of the register at a given instant, determines the phase of the sinusoidal signal at said same instant.

The three generators furthermore comprise inputs 71, 72 and 73 making it possible to reset to zero all the stages of the respective registers, this register condition corresponding to the phase $+270^\circ$ in the sinusoidal signal.

The three generators respectively supply three variable attenuators 81, 82 and 83 the control inputs of which are respectively connected to the outputs 21, 22 and 23 of the source 1 through three digital-analogue converters 91, 92, 93.

It will be observed here that if, for a given sampling period, p is less than n or even zero (periods of silence), this information is readily translated by a zero amplitude information in the channels corresponding to non-existent main components, and that accordingly the manner in which the corresponding signal generators behave during the period in question does not matter in the least, any output signals which they produce being blocked by the variable attenuators. To clarify the language, in the claims, it will be considered that p is always equal to n , but that part or all of the main components may have, for certain sampling periods, a zero amplitude.

The output signals from the attenuators 81, 82 and 83 are added in an adder 55. The signals appearing at the output of the adder 55 have lost their direct components because of the presence of a capacitor arranged at some suitable point. The output of the adder 55 supplies the carrier input of an amplitude-modulator 65 whose output is coupled, through a low pass filter 75, to the input of an electro-acoustic transducer 85 which represents the output element of the synthesiser.

The auxiliary frequency, in this preferred embodiment of the synthesiser, is utilised not only for the aforesaid operations of rephasing, but also to carry out amplitude modulation (in the modulator 65) of the output signal from the adder 55, this indeed in such a manner that the instantaneous value of the modulating signal of frequency f passes through a minimum at the time of the rephasing operations. The overall effect thus obtained is highly satisfactory.

In the embodiment described, each main component is rephased to the same fixed value, corresponding to a zero in all the stages of each register.

Finally, in this example, there is not modification in the circuits if the auxiliary frequency does not correspond with the pitch frequency, and this obviates the need for the special signal referred to hereinbefore. The erratic frequencies which successively constitute the auxiliary frequency on emission of unvoiced conso-

nants, lead to operations of rephasing and to an erratically varying amplitude-modulation from one synthesis period to another, and it has been observed that the auditory effect obtained in this way for the synthesis of unvoiced consonants, is superior in quality to that obtained when the functions of rephasing and amplitude-modulation are not used for the synthesis of these consonants. This simplifies the circuits as well.

This being the case, in the figure the output 10 of the source 1, in order to produce a signal of frequency f , supplies a circuit of the same kind as that which, as far as the frequency is concerned, is employed for the step signals of frequencies F_1 , F_2 and F_3 .

This circuit contains a code converter circuit 30 followed by a variable divider 40 and a signal generator 50.

However, since the frequency f is less than the frequencies of most of the main components, the variable divider 40 is supplied with pulses by a fixed divider 90 itself supplied from the clock 2.

The output signal of frequency f used by the generator 50 is applied to the modulating input of the modulator 65. In this example, the modulating signal is the step signal produced by the generator 50, as yet not riddled of its direct component, so that it passes through a minimum value of zero when the register 50 contains nothing but zeros; moreover, the modulation depth is adjusted so that the modulated signal becomes zero at the same time as the modulating signal. The output signal from the modulator 65 is then supplied to the low-pass filter 75 which smooths it, removing the discontinuities which are due to the steps in both the modulated and the modulating signals.

The zero transits in the output signal from the generator 50 are detected by means of a decoding circuit which is reduced here to a simple NOR-gate 35 with two inputs respectively connected to the two end stages of the shift-register of the generator 50. It can readily be confirmed that this register can only simultaneously carry a zero in each of these two stages, when all its stages carry the zero condition. The output signal from the gate 35 is applied to the inputs 71, 72 and 73 of the generators 51, 52 and 53 and thus effects the rephasing operations at the desired instants.

Another feature which improves the reconstitution of the vowels and voice consonants consists in using as the modulating signal a periodic signal of periodicity $\theta = 1/f$, a cycle of which is formed by two sinusoidal half-cycles (apart from the direct component which give them a minimum value of zero) of respective durations $\theta/4$ for the rise from zero to the maximum and $3\theta/4$ for the descent from the maximum to zero.

Because of this asymmetry, it is thus no longer possible to utilise a generator whose shift register has q stages, in order to obtain a signal of frequency f using shift pulses whose recurrence frequency is $2qf$.

However, it is possible for example to utilise a similar generator in which the shift register has $2q$ stages and initially contains a "1" digit and $(2q-1)$ "0" digits, the loop between the input and the output of the register containing no inverter.

Then, a network of $2q$ resistors is used each of which is selectively connected to the source of the generator for a corresponding position of the "1" digit, this making it possible to give any desired shape to a cycle of the step signal.

A given phase of the envelope of the output signal is indicated by the position, in a given stage, of the single "1" digit circulating through the register, and the output of this stage can directly control the operations of predetermined rephasing.

Another solution consists in retaining the generator 50 of FIG. 1 and supplying it with shift pulses whose frequency is three times higher in the case of the rise portions of the signal (register changes from "0" throughout to "1" throughout), than it is for the descent portions (return to the "0" throughout condition).

Of course, the invention is not limited to the embodiments described and shown which were given solely by way of example.

What is claimed, is:

1. A speech synthesiser comprising means for receiving for each one of successive sampling periods, first digital signals representative of the amplitude of n formants in a speech wave, n being an integer greater than 1, second digital signals representative of the frequencies of said n formants, and an auxiliary digital signal representative of an auxiliary frequency which, for the time intervals corresponding in said speech wave to the emission of vowels and voiced consonants, is the pitch frequency, said synthesiser comprising: n variable-frequency signal generators having respective outputs, said n signal generators being signal generators of the shift register and resistance network type, each one of said n signal generators comprising a shift register having stages and a shift pulse input, said shift register of each one of said n signal generators having a further input for resetting its stages to predetermined states; n amplitude-control devices respectively coupled to said outputs of said n signal generators, and having respective amplitude control inputs and respective outputs; means for deriving from said first signals n amplitude control signals and respectively applying them to said amplitude control inputs of said n amplitude control devices; a frequency control circuit for deriving from each one of said second digital signals a series of pulses having a recurrence frequency proportional to the frequency of which this digital signal is representative and applying the n series of pulses respectively to the shift pulse inputs of said n signal generators; a further control circuit for deriving from said auxiliary signal, at least when said auxiliary signal is representative of said pitch frequency, phase control pulses at said pitch frequency and simultaneously applying them to said further inputs of said n signal generators for resetting each one of said signal generators to a fixed phase; and an adder having n inputs respectively coupled to said outputs of said n amplitude control devices and an output.

2. A speech synthesiser as claimed in claim 1, wherein said further control circuit comprises a further signal generator, having an output, for deriving from said auxiliary signal a modulation signal at said auxiliary frequency, said further signal generator being of the shift register and resistance network type and comprising a shift register having stages and a shift pulse input, means for deriving from said auxiliary signal a further series of pulses having a recurrence frequency proportional to said auxiliary frequency and applying said further series to said shift pulse input of said further signal generator, and decoding means coupled to said stages of said further signal generator for generating said

7

phase control pulses at the instants when the value of said modulation signal passes through a minimum; said speech synthesiser further including an amplitude modulator having a carrier input coupled to said output of

8

said adder and a modulation input coupled to said output of said further signal generator for receiving said modulation signal.

* * * * *

5

10

15

20

25

30

35

40

45

50

55

60

65