

(19)



Europäisches Patentamt
European Patent Office
Office européen des brevets



(11) Publication number:

0 482 699 A2

(12)

EUROPEAN PATENT APPLICATION(21) Application number: **91202675.4**(51) Int. Cl.⁵: **G10L 9/14**(22) Date of filing: **16.10.91**(30) Priority: **23.10.90 NL 9002308**(43) Date of publication of application:
29.04.92 Bulletin 92/18(84) Designated Contracting States:
AT BE CH DE DK ES FR GB GR IT LI LU NL SE(71) Applicant: **Koninklijke PTT Nederland N.V.**
P.O. Box 95321
NL-2509 CH The Hague(NL)(72) Inventor: **Beerends, John Gerard**
20 Kikkerstraat
NL-2515 NB The Hague(NL)
Inventor: **Muller, Frank**
24 Meerkoetlaan
NL-2623 NJ Delft(NL)
Inventor: **van Ravesteijn, Robertus Lambertus**
Adrianus
46 Hoekweg
NL-2275 TB Voorburg(NL)(54) **Method for coding and decoding a sampled analog signal having a repetitive nature and a device for coding and decoding by said method.**

(57) Frequency components are calculated from the STP-filtered speech signal. The amplitudes of these are combined in a manner such that the resultant values are associated with frequencies which are situated equidistantly on a linear Bark scale. Said components are quantised, possibly after scaling. In the decoder, the components are again distributed over the frequency spectrum.

In the coder, the fundamental regularity D is determined with an LTP technique, after which it is transmitted. In the decoder the phases of the reconstructed signal at the spacing D in the past are determined. These phases are combined with the amplitudes already present in the frequency spectrum, after which transformation back to the time domain takes place. Inverse STP filtering is then carried out.

EP 0 482 699 A2

The invention relates to a method for coding a sampled analog signal having a repetitive nature, in which the sampled signal is split into consecutive segments each containing a predetermined number of samples; in which a short-term prediction analysis is performed on said segments and in which the coefficients determined in said analysis are transmitted and are also fed to a short-term prediction filter, in which a long-term prediction analysis is performed on the residual signal available at the output of said filter and the information determined in said analysis is also transmitted, and in which the information present in the residual signal is coded and transmitted.

The invention also relates to a method for decoding a signal coded in the manner described above, in which the long-term prediction analysis information received and the other information received from the residual signal are combined and the combined signal, together with the short-term prediction analysis coefficients received, is fed to an inverse short-term prediction filter at whose output a series of samples is delivered which forms a reconstruction of the sampled analog signal.

The invention also relates to a device for coding and decoding by the method described above.

It is known that analog signals having a strongly consistent nature such as, for example, speech signals can be efficiently coded after sampling by consecutively performing a number of different transformations on consecutive segments of the signal which each have a particular time duration. One of the known transformations for this purpose is linear predictive coding (LPC), for an explanation of which reference can be made to the book entitled "Digital Processing of Speech Signals" by L.R. Rabiner and R.W. Schafer; Prentice Hall, New Jersey; chapter 8. As stated, LPC is always used for signal segments having a particular time duration, in the case of speech signals, for example, 20 ms, and is considered as short-term coding. It is also known to make use not only of a short-term prediction but also a long-term prediction (LTP) in which a very efficient coding is obtained by a combination of these two techniques. The principle of LTP is described in Frequenz, (Frequency), volume 42, no. 2-3, 1988; pages 85-93; P. Vary et al.: "Sprachcodec für das Europäische Funkfernsprechnet" ("Speech coder/decoder for that European Radiotelephone Network"), while an improved version of the LTP principle is described in the Dutch Patent Application 9001985.

The object of the invention is to provide a method for very efficiently transmitting, i.e. with a small number of bits/sec, the information relevant to the human ear in the residual signal remaining after applying the STP principle without the quality, experienced by the listener, of the speech reconstructed by the decoder at the receiving side being impaired.

For this purpose, the method for coding according to the invention is characterised in that the residual signal is transformed to the frequency domain, in that the amplitudes of at least a number of the frequency components obtained in transforming to the frequency domain are combined in a manner such that the frequencies associated with the combined amplitudes are situated equidistantly on a linear Bark scale, and in that a signal is transmitted which is representative of said combined amplitudes.

The method for decoding according to the invention is characterised in that the original amplitudes in the frequency domain are reconstructed from the combined amplitude values received, in that the information transmitted as a result of the long-term prediction analysis is used to calculate the phase values associated with said amplitudes, and in that the calculated phase values, together with the associated amplitudes, are transformed to the time domain.

According to the present invention, the residual signal is coded perceptively, which means that only that information is transmitted which is relevant for differences in the decoded received signal which can be detected by the human ear.

In the first place, use is made for this purpose of the known fact that the human ear is not sensitive to absolute phase values, but only to phase relationships, so that it is not necessary in principle to transmit the phase information from the residual signal to be coded, provided only that it is possible to reconstruct the original phase relationships at the receiving end.

In addition, the present invention makes use of the insight known for some time that human hearing functions in fact as a chain consisting of a number of filters having adjacent frequency bands but having different bandwidths, the so-called critical bands or Barks, the bandwidth of such critical bands being much smaller for low frequencies than for high frequencies. A frequency scale formed in accordance with this insight is referred to as a linear Bark scale. For a further explanation of the principle of the Bark scale, reference is made to B. Scharf and S. Buus, "Stimulus, Physiology, Thresholds" in L. Kaufman, K.R. Boff and J. P. Thomas, editors, Handbook of Perception and Human Performance, chapter 14, pages 1-43, Wiley, New York, 1986.

It is also pointed out that the principle of first transforming a residual signal to be transmitted in speech coding to the frequency domain and then transmitting the information available after this transformation has already been put forward earlier. For this purpose reference can be made, for example, to the paper entitled

"Fourier Transform Vector Quantisation for Speech Coding" by P. Chang et al. in IEEE Transactions on Communications, Vol. COM 35, No. 10, pages 1059-1068. According to this publication, however, after the transformation use is made of vector quantisation and there is no mention of transmitting purely amplitude information.

5 The invention will be explained in greater detail below on the basis of an exemplary embodiment with reference to the drawing, wherein:

Figure 1a shows a block diagram of an exemplary embodiment of a coding unit for the device according to the invention.

10 Figure 1b shows a block diagram of an exemplary embodiment of a decoding unit for the device according to the invention.

An analog signal delivered by a microphone 1 is limited in bandwidth by a low-pass filter 2 and converted in an analog/digital converter 3 into a series of amplitude and time-discrete samples which are representative of the analog signal. The output signal of the converter 3 is fed to the input of a short-term analysis unit 4 and to the input of a short-term prediction filter 5. These two units cater for the
15 abovementioned short-term prediction (STP) on segments of, for example, 160 samples and the analysis unit 4 provides an output signal in the form of short-term prediction filter coefficients which are quantised, coded and transmitted to the decoder unit shown in Figure 1b. The structure and the function of the filter 5 and the unit 4 are well known to those skilled in the art in the field of speech coding and are of no further importance for the essence of the present invention, so that a further explanation can be omitted.

20 The STP-filtered signal is fed to a long-term prediction (LTP) analysis unit 6. In this analysis unit, an LTP analysis is applied twice per segment of 160 samples in a manner such as that described, for example, in Dutch Patent Application 9001985. In such an LTP analysis, for a signal segment to be coded, a search is always made, in accordance with a particular search strategy, for a segment which is as similar as possible in a signal period preceding said segment having a particular duration and a signal is transmitted
25 in coded form which is representative of the number of samples D situated between the starting instant of the segment found and the starting instant of the segment to be coded.

The output signal of the STP filter unit 5 is referred to as the residual signal and, according to the invention, said residual signal is transmitted in coded form in a manner such that only the information which, seen perceptively, is relevant is transmitted. For this purpose, the segments of 160 samples in said residual
30 signal are divided into 8 subsegments of 30 samples in the circuit 7. This is done by first dividing the segment supplied into eight subsegments of 20 samples and then completing these at the leading edge with the ten last samples of the previous subsegment. This implies that the last ten samples of every segment have to be stored in order to also be able to complete the first subsegment of the subsequent segment. Then every subsegment of 30 samples is multiplied in a circuit 8 by a window function such as,
35 for example, a cosine function. The window function is so chosen that, for every sample in the overlapping parts of the subsegments, the sum of the squares of the two multiplication factors is unity. The reason that this has to be the case for the squares is that the multiplication by the window function takes place both in the coding unit and in the decoding unit shown in Figure 1b. A Discrete Fourier Transform (DFT) is performed on the windowed subsegment in a circuit 9, 16 different frequency components being obtained
40 for every subsegment. Of these 16 frequency components, numbered 0 to 15 inclusive, the amplitudes A of the components 1 to 13 inclusive are calculated in a circuit 10. The components 0, 14 and 15 can be ignored because they are situated outside the frequency band of 300 - 3,400 Hz chosen for speech communication. If a greater or a smaller frequency band is relevant, the number of amplitude components taken into consideration can be adjusted accordingly. Starting from the said 13 components, four so-called
45 Bark amplitude components are calculated in a circuit 11. These are amplitudes associated with frequencies which are situated equidistantly on a linear Bark scale. The Bark amplitude components B₁ to B₄ inclusive can, for example, be calculated as follows from the DFT amplitudes A₁ to A₁₃ inclusive:

50

55

$$B_1 = \sqrt{A_1^2 + A_2^2}$$

5

$$B_2 = \sqrt{A_3^2 + A_4^2 + A_5^2}$$

10

$$B_3 = \sqrt{A_6^2 + A_7^2 + A_8^2 + A_9^2}$$

15

$$B_4 = \sqrt{A_{10}^2 + A_{11}^2 + A_{12}^2 + A_{13}^2}$$

20 If desired, a gain factor G is calculated as a scaling value in circuit 12 from the four Bark amplitude components in accordance with:

$$G = \sqrt{B_1^2 + B_2^2 + B_3^2 + B_4^2}$$

25

30 The application of the scaling value G has the advantage that the scaled amplitudes can be coded more efficiently. The value of G is quantised in a circuit 13 and then transmitted to the decoding unit. If the scale factor G has been calculated, every Bark component is divided by the quantised gain factor $\hat{G}$ in a circuit 14. The result of this division is quantised in a circuit 15, coded and then also transmitted to the decoding unit.

If no use is made of a scaling value, the circuits 12, 13 and 14 can be omitted and the four calculated values for the Bark amplitude components can be transmitted directly after quantisation in circuit 15.

35 After decoding in a circuit 16 in the decoder unit, the four scaled Bark amplitude components are multiplied in a multiplier 18 by the gain factor, $\hat{G}$, decoded in a circuit 17, as a result of which the reconstructed Bark amplitude components $\hat{B}_1$ to $\hat{B}_4$ inclusive are obtained. This is of course not applicable if no scaling factor is used in the coding unit. In a circuit 19, the amplitudes in the frequency domain $\hat{A}_1$ to $\hat{A}_{13}$ inclusive (equidistant on the Hz scale) are calculated by means of the following formulae

40

$$\hat{A}_1 = \hat{A}_2 = \frac{\hat{B}_1}{\sqrt{2}}$$

45

$$\hat{A}_3 = \hat{A}_4 = \hat{A}_5 = \frac{\hat{B}_2}{\sqrt{3}}$$

50

$$\hat{A}_6 = \hat{A}_7 = \hat{A}_8 = \hat{A}_9 = \frac{\hat{B}_3}{2}$$

55

$$\hat{A}_{10} = \hat{A}_{11} = \hat{A}_{12} = \hat{A}_{13} = \frac{\hat{B}_4}{2}$$

In order to be able to transform the 13 frequency components considered in the coder back to the time domain by means of an inverse DFT (IDFT) in the IDFT circuit, the amplitudes and the phases are required.

The phases are determined in the following manner with the aid of the $\overline{LTP}$ information decoded in a circuit 23 and consisting of the sample spacing D.

5 The 120 most recent samples of the reconstructed STP residue such as are present at the output of the circuit 22 to be discussed in greater detail below are stored in each case. In a circuit 24, the subsegment is determined which is situated at a spacing of D samples in the past with respect to the present subsegment and this subsegment is multiplied in a circuit 25 by the same window function as was used in the circuit 8 in the coder unit. A DFT is then applied to said subsegment in a circuit 26, after which the phases of the 13
10 components considered can be calculated in a circuit 27. With the aid of the phases determined in this way and the amplitudes already calculated, an IDFT is performed in the circuit 20, the amplitudes of $\hat{A}_0$, $\hat{A}_{14}$, $\hat{A}_{15}$ and $\hat{A}_{16}$ being set equal to zero.

At the output of the circuit 20 a reconstruction of the subsegment, 30 samples long, is now available, but this has also been modified by the window function performed in the coder unit. The reconstructed
15 subsegment is therefore multiplied again by the window function in a circuit 21. In the case of the first ten samples of the subsegment now multiplied twice by the window function, the last ten samples, stored for this purpose, of the previous subsegment multiplied twice by the window function are added in a circuit 22. As a result of this, the sum of the multiplication factors in the resultant ten samples is equal to unity.

The last ten samples in this subsegment are stored. The first twenty samples form a portion of the
20 reconstruction of a segment of the STP residue. After eight subsegments have been reconstructed and combined, a completely reconstructed segment of the STP residue is obtained, and this is situated ten samples in the past with respect to the segment on which the STP analysis has been performed in the coding unit.

An inverse STP filtering is performed on this segment in a filter circuit 28 in a manner known per se
25 with the aid of the STP coefficients received, the filter coefficients from the previous segment being used for the first ten samples.

The output signal of the filter 28 is converted in a digital/analog converter 29 into an analog signal which is fed via a low-pass filter 30 to a loudspeaker 31 which gives a high-fidelity reproduction of the speech
30 signal supplied to the microphone 1, it having been possible to transmit said speech signal in coded form with a low number of bits due to the measures according to the invention.

If desired, a circuit 23' can be included between the circuits 23 and 24 to first subject the value of D received by the decoder additionally to a number of operations in order to obtain an optimum value of D for the reconstruction of the speech signal. These may be three consecutive operations.

1) If the series of values of D received exhibit a trend, the present D received, if it falls outside said trend
35 by a certain margin, is replaced by a value which is in keeping with said trend. Algorithms for determining a trend in a series of consecutive values and for determining a replacement value for a signal which falls outside said trend are well known per se to those skilled in the art.

2) Three intermediate values (I_1 , I_2 and I_3) are calculated between two consecutive values of D (D_1 and D_2), possibly adjusted with the aid of such an algorithm, by means of interpolation. This is done, for
40 example, in the following manner:

$$I_1 = 0.75 * D_1 + 0.25 * D_2$$

$$I_2 = 0.5 * D_1 + 0.5 * D_2$$

$$I_3 = 0.25 * D_1 + 0.75 * D_2$$

45 The interpolation is carried out because the spacing D is determined in the coding unit twice per segment. Without interpolation, decoding of four consecutive subsegments would be carried out with the same value of D. If no fundamental regularity is present in the signal in the coding unit, a regularity would consequently wrongly be provided in the decoder during four subsegments. This problem is
50 overcome by the interpolation.

If fundamental regularity is in fact present in the speech signal, the repetition spacing in the signal will in general vary slowly. Due to the interpolation, the variation in the value of D now also has a smooth nature in the decoder.

3) After equalising the values of D by, if necessary, calculating a replacement value and after
55 interpolation, the calculated spacing D corresponds as well as possible with the actual repetition spacing present in the signal. If, however, said spacing D is less than 30, D is multiplied by an integer which is

chosen in a manner such that the result is as a minimum equal to 30. This is necessary because all the samples of a subsegment at a spacing of less than 30 with respect to the present segment have not yet been reconstructed, so that they can therefore not be used to calculate the phases.

The reason that spaces D of less than 30 are nevertheless transmitted is that, if the fundamental regularity in the signal encompasses a number of samples less than 30, this prevents the decoded spacing D assuming values which are mutually unequal multiples of the actual repetition spacing. As a result of this, the equalisation algorithm would have less opportunity of detecting a trend.

Claims

1. Method for coding a sampled analog signal having a repetitive nature, in which the sampled signal is split into consecutive segments each containing a predetermined number of samples; in which a short-term prediction analysis is performed on said segments and in which the coefficients determined in said analysis are transmitted and are also fed to a short-term prediction filter, in which a long-term prediction analysis performed on the residual signal available at the output of said filter and the information determined in said analysis is also transmitted, and in which the information present in the residual signal is coded and transmitted, characterised in that the residual signal is transformed to the frequency domain, in that the amplitudes of at least a number of the frequency components obtained in the transformed to the frequency domain are combined in a manner such that the frequencies associated with the combined amplitudes are situated equidistantly on a linear Bark scale, and in that a signal is transmitted which is representative of said combined amplitudes.
2. Method for decoding a signal coded by the method according to Claim 1, in which the long-term prediction analysis information received and the other information received from the residual signal are combined and the combined signal, together with the short-term prediction analysis coefficients received, is fed to an inverse short-term prediction filter at whose output a series of samples is delivered which is representative of the sampled analog signal, characterised in that original amplitudes in the frequency domain are reconstructed from the combined amplitude values received, in that the information transmitted as a result of the long-term prediction analysis is used to calculate the phase values associated with said amplitudes, and in that the calculated phase values, together with the associated amplitudes are transformed to the time domain.
3. Method according to Claim 1, characterised in that the amplitudes of thirteen frequency components A_1 to A_{13} inclusive obtained in the transformed to the frequency domain are transformed into amplitudes B_1 to B_4 inclusive of four frequency components situated equidistantly on a Bark scale in accordance with:

$$B_1 = \sqrt{A_1^2 + A_2^2}$$

$$B_2 = \sqrt{A_3^2 + A_4^2 + A_5^2}$$

$$B_3 = \sqrt{A_6^2 + A_7^2 + A_8^2 + A_9^2}$$

$$B_4 = \sqrt{A_{10}^2 + A_{11}^2 + A_{12}^2 + A_{13}^2};$$

and in that these values for B are transmitted after quantisation.

4. Method according to Claim 3, characterised in that a scaling factor G is calculated for the four frequency components B_1 to B_4 inclusive situated equidistantly on a Bark scale in accordance with:

$$G = \sqrt{B_1^2 + B_2^2 + B_3^2 + B_4^2};$$

5

in that this value for G is quantised, and in that the values of B_1 to B_4 inclusive are divided by the quantised scaling factor before they are quantised.

5. Method according to Claim 2 and 3 or 4, characterised in that combined amplitude values $\hat{B}_1$ to $\hat{B}_4$ are constructed from the information received, in that amplitude values $\hat{A}_1$ to $\hat{A}_{13}$ inclusive are obtained therefrom in accordance with:

10

15

$$\hat{A}_1 = \hat{A}_2 = \frac{\hat{B}_1}{\sqrt{2}}$$

20

$$\hat{A}_3 = \hat{A}_4 = \hat{A}_5 = \frac{\hat{B}_2}{\sqrt{3}}$$

25

$$\hat{A}_6 = \hat{A}_7 = \hat{A}_8 = \hat{A}_9 = \frac{\hat{B}_3}{2}$$

30

$$\hat{A}_{10} = \hat{A}_{11} = \hat{A}_{12} = \hat{A}_{13} = \frac{\hat{B}_4}{2}$$

and in that the information transmitted as a result of the long-term prediction analysis is representative of the number of samples D which is situated between the starting instant of a group of samples found with the aid of the long-term prediction analysis and transmitted earlier and the starting instant of a group of samples to be decoded.

35

6. Method according to Claim 5, characterised in that the group of samples transmitted earlier which is situated at a spacing D with respect to a group of samples to be decoded is transformed to the frequency domain, in that the phase value is determined of at least a number of the frequency components calculated with said transformation, in that said phase values are combined with the amplitude values A_1 to A_{13} inclusive, and in that these combinations are transformed to the time domain.

40

7. Method according to Claim 5 or 6, characterised in that the variation in the received values of D is equalised according to a predetermined algorithm by, if necessary, calculating a replacement value for a received value of D and in that three intermediate values are calculated for D between two consecutive values of D by means of interpolation.

45

8. Method according to Claim 7, characterised in that three intermediate values I_1 , I_2 and I_3 are calculated from the known values D_1 and D_2 in accordance with:

50

$$I_1 = 0.75 * D_1 + 0.25 * D_2$$

$$I_2 = 0.5 * D_1 + 0.5 * D_2$$

$$I_3 = 0.25 * D_1 + 0.75 * D_2$$

55

9. Device for performing the method according to at least one of Claims 1 to 8 inclusive.

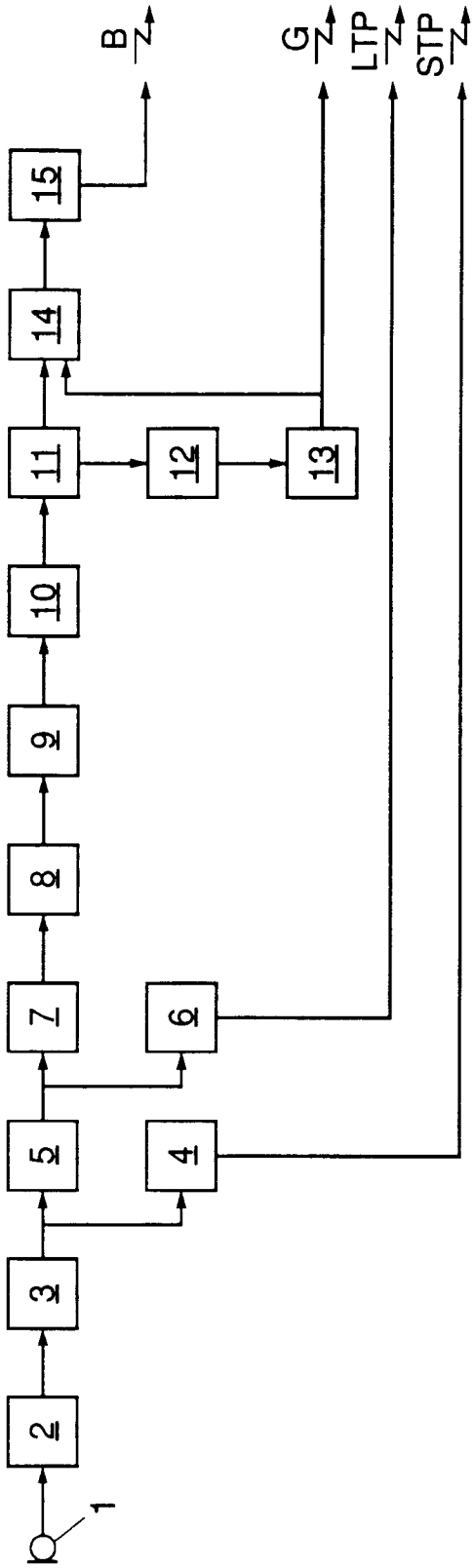


FIG. 1a

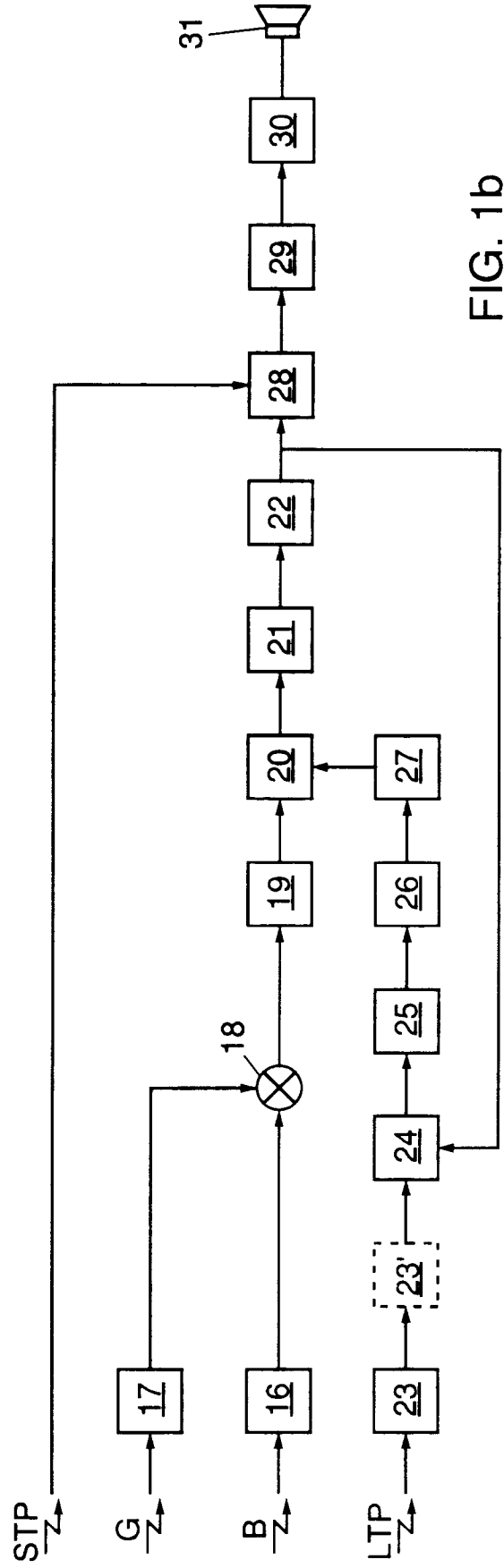


FIG. 1b