(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) Publication number: **0 361 432 A2**

(12) EUROPEAN PATENT APPLICATION

(21) Application number: **89117837.8**

(51) Int. Cl.5: **G10L 9/14**

(22) Date of filing: **27.09.89**

(30) Priority: **28.09.88 IT 6786888**

(43) Date of publication of application:
04.04.90 Bulletin 90/14

(84) Designated Contracting States:
AT BE CH DE ES FR GB GR LI NL SE

(71) Applicant: **SIP SOCIETA ITALIANA PER L'ESERCIZIO DELLE TELECOMUNICAZIONI P.A.**
Via San Dalmazzo, 15
I-10122 Torino(IT)

Applicant: **ITALTEL SOCIETA ITALIANA TELECOMUNICAZIONI s.p.a.**
P.le Zavattari, 12
I-20149 Milano(IT)

(72) Inventor: **Omologo, Maurizio**
Via Isonzo, n. 6
Sarmeola di Rubano Padova(IT)
Inventor: **Sereno, Daniele**
Via Isernia, 7/A
Torino(IT)

(74) Representative: **Riederer Freiherr von Paar zu Schönau, Anton et al**
Van der Werth, Lederer & Riederer Freyung
615 Postfach 2664
D-8300 Landshut(DE)

(54) Method of and device for speech signal coding and decoding by means of a multipulse excitation.

(57) A coding-decoding method using a mulipulse analysis-by-synthesis excitation technique comprises, in the decoding phase, cascaded long-term and short-term synthesis filterings. The lag and gain of the long-term synthesis and the excitation pulses are determined during the coding phase within the analysis-by-synthesis procedure in two subsequent steps, in the first of which the lag and the gain are determined, while in the second the positions and the amplitudes of the excitation pulses are determined. This allows either attainment of a higher quality for a given bit rate or maintenance of a given quality at reduced bit rate. The invention concerns also the device performing the method.

EP 0 361 432 A2

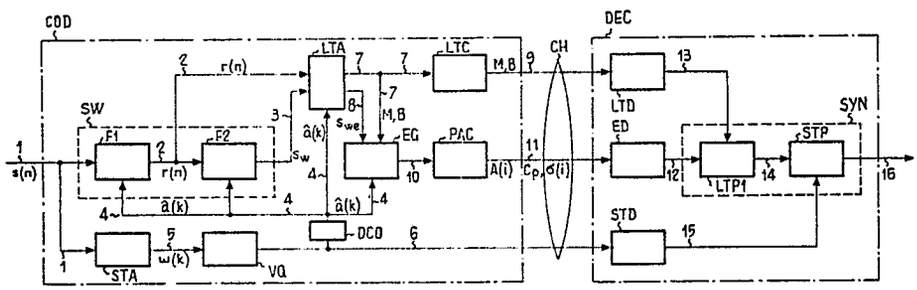

FIG. 1

Method of and Device for Speech Signal Coding and Decoding by Means of a Multipulse Excitation

The present invention concerns medium-low bit-rate speech signal coding systems, and more particularly it relates to a coding-decoding method and device using a multipulse analysis-by-synthesis excitation technique.

Multipulse linear prediction coding is one of the most promising techniques for obtaining high quality synthetic speech at bit rates below 16 kbit/s. This technique has been originally proposed by B. S. Atal and J. R. Remde in the paper entitled "A new method of LPC excitation for producing natural-sounding speech at low bit rates", International Conference on Acoustic, Speech, Signal Processing (ICASSP), pages 614-617, Paris, 1982. According to this technique, the excitation signal for the synthesis filter consists of a train of pulses whose amplitudes and time positions are determined so as to minimize a perceptually-meaningful distorsion measurement; such a measurement is obtained by comparing the samples at the synthesis filter output with the original speech samples and simultaneous weighting the difference by a function which takes into account how the human perception evaluates the distorsion introduced (analysis-by-synthesis procedure)

Different coding-decoding systems using this excitation technique have been suggested. Among those systems, the ones where the synthesizer comprises the cascade of a long-term and a short-term synthesis filter are of particular interest: in fact they provide signals whose quality gradually decreases as the bit rate decreases and do not present a dramatic performance deterioration below a threshold rate.

Examples of said systems are described e.g. in the papers "High quality multipulse speech coder with pitch prediction: presented by K. Ozawa and T. Araseki at the conference ICASSP 86, Tokyo, 7-11 April 1986, and published at pages 33.3.1 - 33.3.4 of the conference proceedings, and "Experimental evaluation of different approaches to the multipulse coder", presented by P. Kroon and E. F. Deprettere at the conference ICASSP 84, San Diego, 19-21 March 1984, and published at pages 10.4.1-10.4.4 of the conference proceedings.

In those systems all parameters relevant to long-term synthesis filter and to excitation are optimized within the analysis-by-synthesis procedure . This procedure gives highly-complex optimum algorithms. If the optimum procedure is not followed, there is a performance reduction for a given transmission rate, or a transmission rate increase is required to maintain a certain performance level.

The invention provides on the contrary a method and a device allowing quality to be increased leaving bit rate unchanged or a given quality to be maintained even at lower bit rate. This can be achieved by using a combined optimization technique, of sequential type, of the parameters of the long-term synthesis filter and of the excitation within the analysis-by-synthesis procedure; the sequential procedure is sub-optimum with respect to the original optimum one, but it is easier to be implemented.

The method according to the invention comprises a coding phase including the following operations:

- speech signal conversion into frames of digital samples;
- short-term analysis of the speech signal, to determine a group of linear prediction coefficients relevant to a current frame and a representation of said coefficients as line spectrum pairs;
- coding of said representation of the linear prediction coefficients;
- spectral shaping of the speech signal, by weighting the digital samples in a frame by a first and a second weighting functions, the weighting according to the first weighting function generating a residual signal which is then weighted by the second function to generate the spectrally-shaped speech signal;
- long-term analysis of the speech signal, by using said residual signal and said spectrally-shaped signal, to determine the lag separating a current sample from a preceding sample used to process said current sample, and the gain by which said preceding sample is weighted for the processing;
- determination of the locations and amplitudes of the excitation pulses, by exploiting the results of the short-term and long-term analysis;
- coding of the values of said lag and gain of the long-term analysis and of said positions and amplitudes of the excitation pulses, the coded values forming, together with said coded representation of the linear prediction coefficients and with coded r.m.s. values of said excitation pulses, the coded speech signal;

and a decoding phase, wherein the excitation is reconstructed starting from the coded values of the amplitudes, the locations and the r.m.s. values of the pulses, and, by using the reconstructed excitation, a synthesized filtering followed by a short-term synthesis filtering, using long-term analysis parameters and respectively the linear prediction coefficients reconstructed starting from said coded line spectrum pairs or line pair differences, which method is characterized in that the long-term analysis and excitation pulse generation are performed in successive steps, in the first of which the long-term analysis gain and lag are determined by minimizing a mean squared error between the spectrally-shaped speech signal and a further

signal obtained by weighting by said second weighting function the signal resulting from a long-term synthesis filtering in which the input signal used for the synthesis is a null signal, while in the second step the amplitudes and positions of the excitation pulses are actually determined, and in that said coding of the representation of the linear prediction coefficients consists of a vector quantization of the line spectrum pairs or of the differences between adjacent line pairs, according to a split-codebook quantization technique.

The invention provides also a device for implementing the method, comprising, for speech signal coding:

- means for converting the speech signal into digital sample frames;
- means for the short-term analysis of the speech signal, which means receive a group of samples from said converting means, compute a set of linear prediction coefficients relevant to a current frame, and emit a representation of said linear prediction coefficients as line spectrum pairs;
- means for coding said representation of the linear prediction coefficients;
- means for obtaining quantized linear prediction coefficients from said coded representation;
- a circuit for the spectral shaping of the speech signal, connected to the converting means and to the means obraining the quantized coefficients, and comprising a pair of cascaded weighting digital filters, which weight the digital samples according to a first and a second weighting function, respectively, said first filter supplying a residual signal;
- means for the long-term analysis of the speech signal, connected to the output of said first filter and of the spectral shaping circuit, to determine the lag which separates a current sample from a preceding sample used to synthesize said current sample and the gain by which said preceding sample is weighted for the synthesis;
- an excitation generator for determining the positions and the amplitudes of the excitation pulses, which generator is connected to said means for obtaining the quantized coefficients and to said long-term analysis means;
- means for coding the values of said long-term analysis lag and gain and excitation pulse positions and amplitudes, the coded values forming, jointly with the coded respresentation of said coefficients and a coded r.m.s. value of said excitation pulses, the coded speech signal;

and also comprising, for speech signal decoding (synthesis):

- means for reconstructing the excitation, the lag and gain of long-term analysis and the linear prediction coefficients starting from the coded signal;
- a synthesizer, comprising the cascade of a first long-term synthesis filter, which receives the reconstructed excitation pulses, gain and lag and filters them according to a first transfer function dependent on said gain and lag, and of a short-term synthesis filter, having a second transfer function reciprocal of said first spectral weighting function,

characterized in that the long-term analysis means are arranged to determine said lag and said gain in two successive steps, preceding a step in which the amplitudes and positions of the excitation pulses are determined by said excitation generator, and comprise:

- a second long-term synthesis filter, which is fed with a null signal and in which, for the lag computation, there is used a predetermined set of values of the number of samples separating a current sample being synthesized from a previous sample used for the synthesis, and, for the gain computation, a predetermined set of values of the gain itself is used;
- a multiplexer which receives at a first input a sample of residual signal and at a second input a sample of the output signal of the second long-term synthesis filter and supplies the samples present at either input depending on whether or not said number of samples is lower than the frame length;
- a third weighting filter, having the same transfer function as said second digital filter in the spectral shaping means, which third filter is connected to the output of said second long-term synthesis filter and is enabled only during determination of long-term analysis gain;
- a first adder, which receives at a first input the spectrally-shaped signal and at a second input the output signal of said third weighting filter and supplies the difference between the signals present at its first and second input;
- a first processing unit, which receives in said first operation step the signal outgoing from said multiplexer and determines the optimum value of said number of samples, and in a second operation step receives the output signal of said first adder and determines, by using the lag computed in the first step, the value of the gain which minimizes the mean squared error, within a gain validity period, between the input signals of the first adder;

and in that the excitation pulse generating means comprise:

- a third long-term synthesis filter, which has the same transfer function as the first one and is supplied with the excitation pulses which are being generated;

- a fourth weighting filter, connected to the output of the third synthesis filter and having the same transfer function as said second and third weighting filters;
- a second adder, which receives at a first input the output signal of said first adder and at a second input the output signal of the fourth weighting filter, and supplies at the output the difference between the signals present at its first and second inputs;
- a second processing unit which is connected to the output of said adder and determines the amplitudes and positions of said pulses by minimizing the mean squared error, within a pulse validity period, between the input signals of the fourth adder.

The invention will be better understood from the following description of an exemplary embodiment thereof, with reference to the annexed drawings, in which:

Fig. 1 a block diagram of the coder-decoder according to the invention;

Fig. 2 is a flow chart of the operation concerning the determination of long-term analysis gain;

Fig. 3 is a block diagram of the circuits for long-term analysis and excitation pulse generation.

With reference to Fig. 1, a generic speech signal coding-deconding system can be schematized by a coder COD, a transmission channel CH and a decoder DEC.

In case of a system based on a multipulse excitation technique and exploiting speech signal long-term and short-term correlations, coder COD receives digital samples s(n) of the original speech signal, organized into frames comprising each a predetermined number of samples, and sends onto channel CH, for each sample frame, the coding of a suitable representation $\omega(k)$ of a group of linear prediction coefficients a(k) obtained by a short-term analysis of the speech signal, the coded amplitudes and positions A(i), Cp of the pulses forming the excitation signal, the coded r.m.s. values $\sigma(i)$ of the excitation pulses, and the codings of two parameters (gain B and lag M) determined by the long-term analysis. Decoder DEC reconstructs the excitation and generates a synthesized speech signal on the basis of the reconstructed excitation, the linear prediction coefficients reconstructed starting from the transmitted representation thereof, and long-term analysis parameters.

By way of example, whenever necessary, reference will be made to a 15 ms frame duration, which corresponds to 120 samples if a 8 kHz sampling frequency is assumed.

For the coding in COD, the digital sample frames, present on connection 1, are supplied to a spectral shaping circuit SW and to a short-term analysis circuit STA.

Spectral shaping circuit SW performs a frequency-shaping of the speech signal in order to render the differences between the original and the reconstructed speech signals less perceptible in correspondence with the formants of the original speech signal. Such a circuit consists of a pair of cascaded digital filters F1, F2, whose transfer functions, in z transform, are given in a non-limiting example respectively by relations

$$A(z) = 1 - \sum_{k=1}^{p} \hat{a}(k) z^{-k}$$

and

$$1/A(z,\gamma) = 1/\left[1 - \sum_{k=1}^{p} \gamma^{k} \cdot \hat{a}(k) z^{-k}\right]$$

where z represents a sampling interval delay; $\hat{a}(k)$ is a quantized linear prediction coefficient vector ($1 \leq k \leq p$, where p is the filter order) reconstructed from the coded representation of the linear prediction coefficients short-term analysis result; γ is an experimentally determined constant correcting factor, determining the bandwidth increase around the formants. Spectral weighting circuit SW as a whole has a transfer function $W(z) = A(z)/A(z,\gamma)$. A signal r(n), hereinafter referred to as "residual signal", is obtained on output connection 2 of F1, and spectrally weighted speech signal $s_w(n)$ is obtained on output connection 3 of F2: both signals are used in long-term analysis.

Short-term analysis circuit STA is to determine linear prediction coefficients a(k), which depend on short-term correlations deriving from a non-flat spectral envelope of speech signal. Circuit STA calculates coefficients a(k) according to the classical autocorrelation method, as described in "Digital Signal Process-

ing of Speech Signals" by L.R. Rabiner and R.W. Schafer (Prentice-Hall, Englewood Cliffs, N.J., USA, 1978), page 401, and uses to this aim a set of digital samples $s_h(n)$ which can comprise, besides the samples of the current frame, a certain number of samples of both the preceding and the following frames.

More particularly, with reference to the exemplary frame length, the set of samples $s_h(n)$ can comprise 200 samples, overlapping the frame which is being processed. Block STA also comprises circuits for transforming the coefficients into a group of parameters $\omega(k)$ in the frequency domain, known as "line spectrum pairs", which are presented on output 5 of STA. As known, line spectrum pairs denote the resonant frequencies at which the acoustic tube, the vocal tract can be assimilated to, exhibits a line spectrum structure under extreme boundary conditions corresponding to complete opening and closure at the glottis.

The conversion of linear prediction coefficients into line spectrum pairs is described e.g. by N. Sugamura and F.Itakura in the paper "Speech analysis and synthesis method developed at ECL in NTT - From LPC to LSP", Speech Communication, Vol.5, No.2, June 1986, pages 199-215.

Line spectrum pairs $\omega(k)$ or the differences $\Delta\omega$ between adjacent line pairs are then vectorially quantized in a vector quantization circuit VQ exploiting techniques of the type described in published European Patent application EP-A-186763 (CSELT), applied to a set of codebooks. In other wordsl, leaving unchanged the number of bits by which each vector of ω (or $\Delta\omega$) is desiderably coded, that vector, instead of being coded by a single word with that number of bits, is quantized by a group of words of smaller size chosen out of suitable sub-codebooks. The modality of quantization of the above patent application are applied to obtain each of said words. The presence of vector quantizer VQ is one of the characteristics of the present invention and allows a reduction in the number of bits necessary to code the results of the short-term analysis, while maintaining the same quality of the coded signal, from about 36-34 bits (scalar quantization) to 24 (vector quantization). By way of example, differences $\Delta\omega$, organized into three vectors of 3, 3 and 4 components respectively, may be quantized with 24 bits organized into three groups of 256 words, each group corresponding to one of said vectors. The indices of the vectors are sent by VQ on a connection 6 which belongs to channel CH.

A circuit DCO obtains from said indices quantized linear prediction coefficients $\hat{a}(k)$ which are supplied, through connection 4, to filters F1, F2 of circuit SW, to an excitation generator EG and to a long-term analysis circuit LTA.

Long-term analysis circuit LTA supplies information dependent on the fine spectral structure of the signal, which information is used to make the synthesized signal more natural-sounding. For the analysis concerning a sample frame, the samples relevant to M preceding sampling instants, weighted by a weighting factor (gain) B, are used. LTA is just to determine both M and B. Lag M, in case of a voiced sound, corresponds to the pitch period. In the example considered, the lag can range from 20 to 83 samples and it is updated every frame. The gain is on the contrary updated every half frame. Values M and B are emitted on a connection 7 and are supplied to excitation generator EG which also receives, through a connection 8, a signal $s_{we}(n)$, obtained from $s_w(n)$ in a manner which will be described hereinafter. Values M and B are also sent to a coder LTC, which transfers the coded signals onto a connection 9 belonging to channel CH.

The structure and the operation of a device such as LTC are known in the art.

Long-term analysis circuit LTA performs a closed-loop analysis as a part of the procedure for determining the pulse positions, with modalities allowing a good coder performance to be maintained even if a sub-optimum procedure is used, as will be better described hereinafter.

Excitation generator EG is to supply the sequence of Ns pulses (e.g. 6), distributed within a time period Ls (more particularly corresponding to half a frame), forming the excitation signal; such a signal is computed so as to minimize a mean squared error, frequency-weighted as mentioned, between the original signal and the reconstructed one.

The operations carried out by blocks LTA and EG will be described in more details hereinafter, making also reference to Fig. 3.

Excitation generator EG supplies, through a connection 10, the pulses it has generated to a circuit PAC coding the amplitudes and the positions of such pulses, which circuits calculate and code also the r.m.s. values of said pulses. The coded values $\sigma(i)$, A(i) ($1 \leq i \leq Ns$) and Cp are emitted on a connection 11, also belonging to channel CH.

The structure of circuit PAC is known to the skilled in the art.

In decoder DEC, an excitation decoder ED reconstructs the excitation starting from the coded values σ-(i), A(i), Cp. Through a connection 12, reconstructed excitation pulses $\hat{e}$ are supplied by ED to a long-term synthesis filter LTPI which, together with a short-term synthesis filter STP, forms synthesizer SYN. The long-term synthesis filter is a filter whose transfer function, in z transform, is $1/P(z) = 1/(1-B \cdot z^{-M})$, where M,

B, have the meanings stated above and are supplied to LTP1, through a connection 13, by a circuit LTD decoding the long-term analysis parameters.

Reconstructed residual signal $\hat{r}$ is present at the output of LTP1 and is sent via a connection 14 to short-term synthesis filter STP. This is a filter whose transfer function in z transform is 1/A(z), where A(z) is the function already examined for filter F1 of spectral weighting circuit SW. Coefficients $\hat{a}(k)$ for filter STP are supplied through a connection 15 from a circuit STC, which reconstructs them by decoding the information relevant to line spectrum pairs.

Filter STP emits on connection 16 the reconstructed or synthesized speech signal $\hat{s}$.

To simplify the drawing we have not represented the devices for converting speech signal into sample frames, the buffer for the samples to be processed and the time base for timing the various operations. On the other hand said devices are wholly conventional.

Considering again long-term analysis and excitation generation, the optimum solution would be determining, for each pair of possible values m, b of the lag and gain used to determine the optimum values M, B to be exploited in the synthesis, the combination of excitation pulses, gain and lag minimizing the mean squared error between the original signal and the reconstructed signal. However, the optimum solution is too complex and hence, according to the invention, the determination of M and B is separated from that of the excitation pulses. There are hence two successive operation phases.

In the first phase (determination of M, B) values M, B of m and b are to be found which minimize mean squared error

$$E(m, b) = \sum_{n=1}^{Ls} swe^2(n) = \sum_{n=1}^{Ls} [s_w(n) - s_{w0}(n)]^2 \qquad (1)$$

between frequency-shaped speech signal $s_w(n)$ and a signal $s_{w0}(n)$ obtained by weighting, in the same way as the residual signal, a signal r_0 obtained as a response from a long-term synthesis filter (similar to the one of the synthesizer) when at the filter input a zero has been forced (long-term synthesis filter memory). In the second phase the positions and amplitudes of the excitation pulses are actually determined, so as to minimize, in a perceptually meaningful way, a squared error

$$dw = \sum_{n=1}^{Ls} [s_{we}(n) - \hat{s}_{we}(n)]^2 \qquad (2)$$

where $s_{we}(n)$ has the meaning above and $\hat{s}_{we}(n)$ is the signal obtained by filtering excitation pulses e(i) according to a function H(z) = 1/[P(z)A(z)].

For the first phase an analytical approach could be followed, by taking into account that determining the minimum of E(m,b) corresponds to determining the maximum of a function

$$R(m) = x^2(m)/y(m) \qquad (3)$$

where

$$x(m) = \sum_{n=1}^{L} [r(n)r_0(n-m)]$$

$$y(m) = \sum_{n=1}^{L} r_0^2(n-m) \qquad (4)$$

L being the frame length.

This can be easily deduced by deriving the relation which gives the error and equalling the derivative to 0. However, for a generic value of n and m, signal $r_0(n-m)$ can be unavailable, unless the lag exceeds the frame duration.

According to the invention two sub-optimum solutions allowing elimination of the constraint between the lag and the duration of the frame are proposed for computing B and M.

According to the first sub-optimum solution a predetermined value b is allotted to the gain and the error is minimized for each value m of lag: once found optimum lag M, the successive step is that of determining the optimum gain B.

A second and simpler solution is that of computing M by using a signal $\tilde{r}_0$ which consists of the signal r_0, when the lag is greater than the frame length (or, more generally, when a sample of the current frame is processed by using a sample of the preceding frame), while in the opposite case it is equal to residual signal r(n), and minimising the error

$$dr = \sum_{n=1}^{L} [r(n) - \tilde{r}_0(n-m)]^2 \qquad (5)$$

Under said conditions the previous constraint for the lag is eliminated, since signals r_0 are always available, and hence M can be determined as the number m of samples which maximizes the function

$$R'(m) = X^2(m)/Y(m) \qquad (6)$$

where

$$X(m) = \sum_{n=1}^{L} [r(n)\tilde{r}_0(n-m)]$$

$$Y(m) = \sum_{n=1}^{L} \tilde{r}_0^2(n-m) \qquad (7)$$

Once M has been determined, gain B can be determined either in exhaustive manner or by the following procedure, which reduces the necessary amount of computations. First, value s'_{w0} of s_{w0} when b = 1 is determined, according to relation

$$s'_{w0}(n) = r'_0(n) + \sum_{k=1}^{p} \gamma^k \hat{a}(k) s'_{w0}(n-k) \qquad (8)$$

where $r'0(n)$ is the value of r_0 for b = 1, and mean squared error E(M,1) is calculated. For each b ≠ 1, sw0 is calculated starting from $s'w_0$, according to relations:

$s_{w0}(n) = bs'_{w0}(n)$ for $n \leq M$

$s_{w0}(n) = b[s_{w0}(n-M) + s'_{w0}(n) - s'_{w0}(n-M)]$ for $n > M$ (9)

and the corresponding error E(M, b) is determined. Lastly, value B of b is chosen which renders E(M, b) minimum. Once found M, B, the positions of the individual pulses e(i) of the excitation signal and then the amplitudes of same are determined so as to minimize dw, e.g. by the modalities described in the paper "Efficient computation and encoding of the multipulse excitation for LPC" by M. Berouti, H. Garten, P. Kabal and P. Mermelstein, presented at the already mentioned conference ICASSP 84 and published at pages 10.1.1-10.1.4 of the conference proceedings.

As said, B is computed every half frame, and hence also the excitation pulses will be computed every half frame.

Fig. 3 shows a block diagram of the devices of LTP and EG in case signal $\tilde{r}_0$ is used to determine M and B.

In circuit LTA a synthesis filter LTP2, having a transfer function similar to that of LTP1 (Fig. 1), is fed with a null signal. When M is being determined, filter LTP2 successively uses the different values m and, for each of them, an optimum value b(ott) which is implicitly obtained in the above-mentioned derivative operation. When B is being determined, LTP2 uses value M of the lag determined in the preceding step and different values b. Values m and b are supplied to LTP2 by a processing unit CMB, carrying out the computation and comparisons mentioned above. Signal $\tilde{r}_0$ is present on output 20 of LTP2.

Output 20 is connected to a first input of a multiplexer MX1 receiving at a second input the residual signal r(n) present on connection 3, and letting through signal r_0 or signal r depending on the relative value of m and n. Hence signal r_0 is present on output connection 21 of MX1, and that signal is delayed by a time equal to m samples in a delay element DL1 before being sent to CMB. The latter receives also signal r(n) and, for each frame and for all values m, calculates function $R'(m)$ and determines the value M of m which maximizes such function. The value is stored into a register RM and made available on wires 7a of connection 7.

Output 20 of LTP2 is also connected to a weighting filter F3, which is enabled only while B is being computed and has the same transfer function $1/A(z,\gamma)$ as filter F2 in SW (Fig. 1). Filter F3 weights signal r_0 (or r'_0, when the gain used in LTP2 is 1) giving at output 22 signal s_{w0} (s'_{w0}). The latter is supplied at an input of an adder SM1 where it is subtracted from signal s_w coming from spectral weighting filter SW (Fig. 1) via connection 3. SM1 supplies on output 8 signal s_{we}. By using the procedure above (relations (8) and (9)), device CMB determines, every half frame, value B of $\underline{b}$ which minimizes E and stores it into register RB which keeps it available, for the whole half frame, on a group of wires 7b of connection 7.

Values B, M computed by CMB are supplied to LTC (Fig.1) and to a long-term synthesis filter LTP3 which is part of the excitation generator EG and is followed by a weighting filter F4. Filters LTP3, F4 have transfer functions similar to those of LTP1 and F2, respectively; LTP3 is fed, during the analysis-by-synthesis procedure, with the excitation pulses e(i) supplied via connection 10 by a processing unit CE which sequentially determines the positions and the amplitudes of the various pulses. F4 emits on output 24 signal $\hat{s}_{we}$ which is supplied to a first input of an adder SM2 receiving at a second input signal s_{we} outgoing from SM1. The difference between the two signals is then supplied via connection 25 to CE, which determines pulses e(i) by minimizing mean squared error dw.

It is clear that what described has been given only by way of non limiting example and that variations and modifications are possible without going out of the scope of the invention as defined in the following claims.

Claims

1 - A method of speech signal coding and decoding, using a multipulse analysis-by-synthesis excitation technique, which method comprises a coding phase including the following operations:

- speech signal conversion into frames of digital samples [s(n)];
- short-term analysis of the speech signal, to determine a group of linear prediction coefficients [a(k)] relevant to a current frame and a representation thereof as line spectrum pairs;
- coding of said representation of the linear prediction coefficients;
- spectral shaping of the speech signal, by weighting the digital samples [s(n)] in a frame by a first and a second weighting functions [A(z); $1/A(z,\gamma)$], the weighting by the first weighting function generating a residual signal [r(n)], which is then weighted by the second function to generate a spectrally-weighted speech signal [$s_w(n)$];
- long-term analysis of the speech signal, by using said residual signal [r(n)] and said spectrally weighted signal [$s_w(n)$], to determine the lag (M) separating a current sample from a preceding sample [r(n)-M)] used to process said current sample, and the gain (B) by which said preceding sample is weighted for the processing;
- determination of the positions and amplitudes of the excitation pulses, by exploiting the results of short-term and long-term analysis, said determination being performed in closed loop as a part of the procedure by which excitation pulse positions are determined;
- coding of the values of said lag and gain of long-term analysis and of said amplitudes and position of the excitation pulses, the coded values forming, jointly with the coded representation of the linear prediction coefficients and with coded r.m.s. values of said excitation pulses, the coded speech signal;

and also comprising a decoding phase, where the excitation is reconstructed starting from the coded values

of the amplitudes, the positions and the r.m.s. values of the pulses and where, by using the reconstructed excitation ($\hat{e}$, a synthesized speech signal [$\hat{s}(n)$] is generated, by means of a long-term synthesis filtering followed by a short-term synthesis filtering, which filterings exploit the long-term analysis parameters and respectively the quantized linear prediction coefficients, characterized in that said long-term analysis and excitation pulse generation are performed in successive steps, in the first of which long-term analysis gain (B) and lag (M) are determined by minimizing a mean squared error between the spectrally-shaped speech signal [$s_w(n)$] and a further signal [$s_{w0}(n)$] obtained by weighting by said second weighting function the signal resulting from a long-term synthesis filtering, which is similar to that performed during decoding and in which the signal used for the synthesis is a null signal, while in the second step the amplitudes and positions of the excitation pulses [e(i)] are actually determined by minimizing the mean squared error between a signal [$s_{we}(n)$] representing the difference between the spectrally-shaped speech signal [$s_w(n)$] and said further signal [$s_{w0}(n)$], and a third weighted signal [$\hat{s}_{we}(n)$], obtained by submitting the excitation pulses to a long-term synthesis filtering and to a weighting by said second weighting function, and in that the coding of said representation of the linear prediction coefficients consists in a vector quantization of the line spectrum pairs or of the adjacent line pair differences according to a split-codebook quantization technique.

2. A method as claimed in claim 1, characterized in that the lag (M) and the gain (B) are determined in two successive steps, in the first of which an optimum value of the lag is determined by minimizing said error for a predetermined gain value, while in the second the optimum gain value is determined, by using said optimum lag value.

3. A method as claimed in claim 1, characterized in that the lag (M) and the gain (B) are determined in two successive steps, in the first of which the mean squared error is minimized between the residual signal [r(n)] and a signal [$\tilde{r}_0(n)$] which is the signal [$r_0(n)$] resulting from said long-term synthesis filtering with null input, if the synthesis relevant to a sample of the current frame is performed on the basis of a sample of a preceding frame, and is said residual signal [r(n)] if the synthesis relevant to a sample of the current frame is performed on the basis of a preceding sample of the same frame, while in the second step the gain (B) is calculated with the following sequence of operations: a value [$s'_{w0}(n)$] of said further signal is determined for a unitary gain value; a first error value E(M,1) is hence determined, and the operations for determining the value of the signal weighted with said second weighting function and of the error are repeated for each value possible for the gain, the value adopted being the one which minimizes the error.

4. A method as claimed in claim 3, characterized in that the lag (M) is computed every frame, and the gain (B) every half frame.

5. Device for speech signal coding and decoding by multipulse analysis-by-synthesis excitation techniques, for implementing the method as claimed in any of claims 1, 3 or 4, comprising, for speech signal coding:

- means for converting the speech signal into frames of digital samples [s(n)];
- means (STA) for the short-term analysis of the speech signal, which means receive a group of samples from said converting means, compute a set of linear prediction coefficients [a(k)] relevant to a current frame and emit a representation of said linear prediction coefficients [a(k)] as line spectrum pairs;
- means (VQ) for coding said representation of the linear prediction coefficients;
- means (DCO) for obtaining quantized linear prediction coefficients [$\hat{a}(k)$] from said coded representation;
- a circuit (SW) for the spectral shaping of the speech signal, connected to the converting means and to the means (DCO) obtaining the quantized linear prediction coefficients and comprising a pair of cascaded weighting digital filters (F1, F2), weighting the digital samples [s(n)] according to a first and a second weighting function [A(z); $1/A(z,\gamma)$], respecively, said first filter (F1) supplying a residual signal;
- means (LTA) for the long-term analysis of the speech signal, connected to the outputs of said first filter (F1) and of the spectral shaping circuit (SW) to determine the lag (M) which separates a current sample from a preceding sample [r(n-M)], used to process said current sample, and the gain (B) by which said preceding sample is weighted for the processing;
- an excitation generator (EG) for determining the positions and the amplitudes of the excitation pulses, connected to said short-tem and long-term analysis means (STA, LTA) and to said spectral shaping circuit (SW);
- means (LTC, PAC) for coding the values of said long-term analysis lag and gain and excitation pulse positions and amplitudes, the coded values forming, jointly with the coded representation of the linear prediction coefficients and with r.m.s. values of said excitation pulses, the coded speech signal;

and also comprising, for speech signal decoding (synthesis):

- means (ED, LTD, STD) for reconstructing the excitation, the long-term analysis lag (M) and gain (B) and the linear prediction coefficients [a(k)] starting from the coded signal; and

- a synthesizer, comprising the cascade of a first long-term synthesis filter (LTP1), which receives the reconstructed excitation pulses, gain and lag and filters them according to a first transfer function [1/P(z)] dependent on said gain and lag, and a short-term synthesis filter (STP) having a second transfer function [1/A(z)] which is the reciprocal of said first spectral weighting function [A(z)],

characterized in that the long-term analysis means (LTA) are apt to determine said lag (M) and gain (B) in two successive steps, preceding a step in which the amplitudes and positions of the excitation pulses are determined by said excitation generator (EG), and comprise:

- a second long-term synthesis filter (LTP2), which is fed with a null signal and in which, for the computation of the lag (M), there is used a predetermined set of values of the number of samples separating a current sample being synthesized from a previous sample used for the synthesis, and, for the computation of the gain (B), a predetermined set of possible values of the gain itself is used;
- a multipleser (MX1) receiving at a first input a residual signal sample and at a second input a sample of the output signal of the second long-term synthesis filter (LTP2) and supplies the samples present at either input depending on whether or not said number of samples is lower than a frame length;
- a third weighting filter (F3), which has the same transfer function as said second digital filter (F2) of the spectral shaping circuit (SW), is connected to the output of said second long-term synthesis filter (LTP2) and is enabled only during the determination of the long-term analysis gain (B);
- a first adder (SM1), which receives at a first input the spectrally-shaped signal (s_w) and at a second input the output signal of said third weighting filter (F3) and supplies the difference between the signals present at its first and second input;
- a first processing unit (CMB), which receives in said first operation step the signal outgoing from said multiplexer (MX1) and determines the optimum value of said number of samples, and in a second operation step receives the output signal of said first adder (SM1) and determines, by using the lag computed in the first step, the value of the gain which minimizes the mean squared error, within a validity period of the excitation pulses, between the input signals of the first added (SM1);

and in that the excitation generator (EG) for generating the excitation pulses [e(i)] comprise:

- a third long-term synthesis filter (LTP3), which has the same transfer function as the first one and is fed with the excitation pulses generated;
- a fourth weighting filter (F4), connected to the output of the third synthesis filter (LTP3) and having the same transfer function as said second and third weighting filters (F2, F3);
- a second adder (SM2), which receives at a first input the output signal of said first adder (SM1) and at a second input the output signal of the fourth weighting filter (F4), and supplies the difference between the signals present at its first and second input;
- a second processing unit (CE) which is connected to the output of said second adder and determines the amplitudes and positions of said pulses by minimizing the mean squared error, within a pulse validity period, between the input signals of the fourth adder (SM2).

6. A device as claimed in claim 5 characterized in that the means (VQ) coding said representation of the linear prediction coefficient consist of a vector quantizer (VQ) for split-codebook vector quantization of the line spectrum pairs or of the differences between adjacent line spectrum pairs.

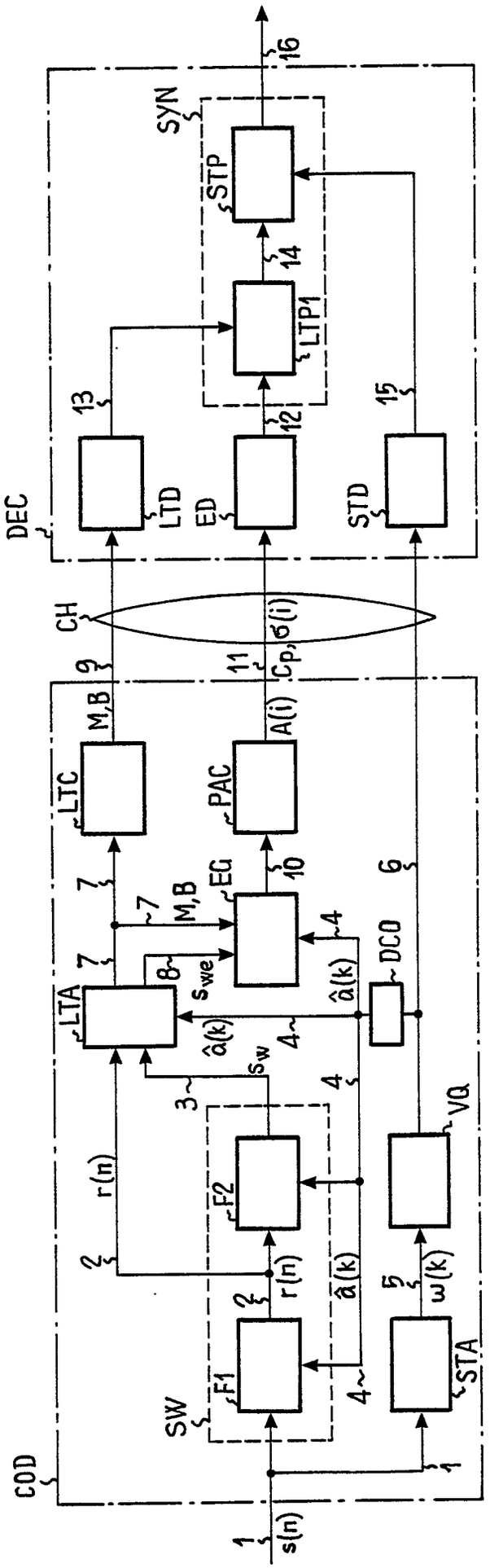
COD
DEC
SW
F1
F2
STA
VQ
LTA
LTC
PAC
EG
DCO
LTD
ED
STD
SYN
LTP1
STP
CH
s(n)
r(n)
ŝ(k)
ŝw
swe
M,B
A(i)
cp,σ(i)
ω(k)

FIG. 1

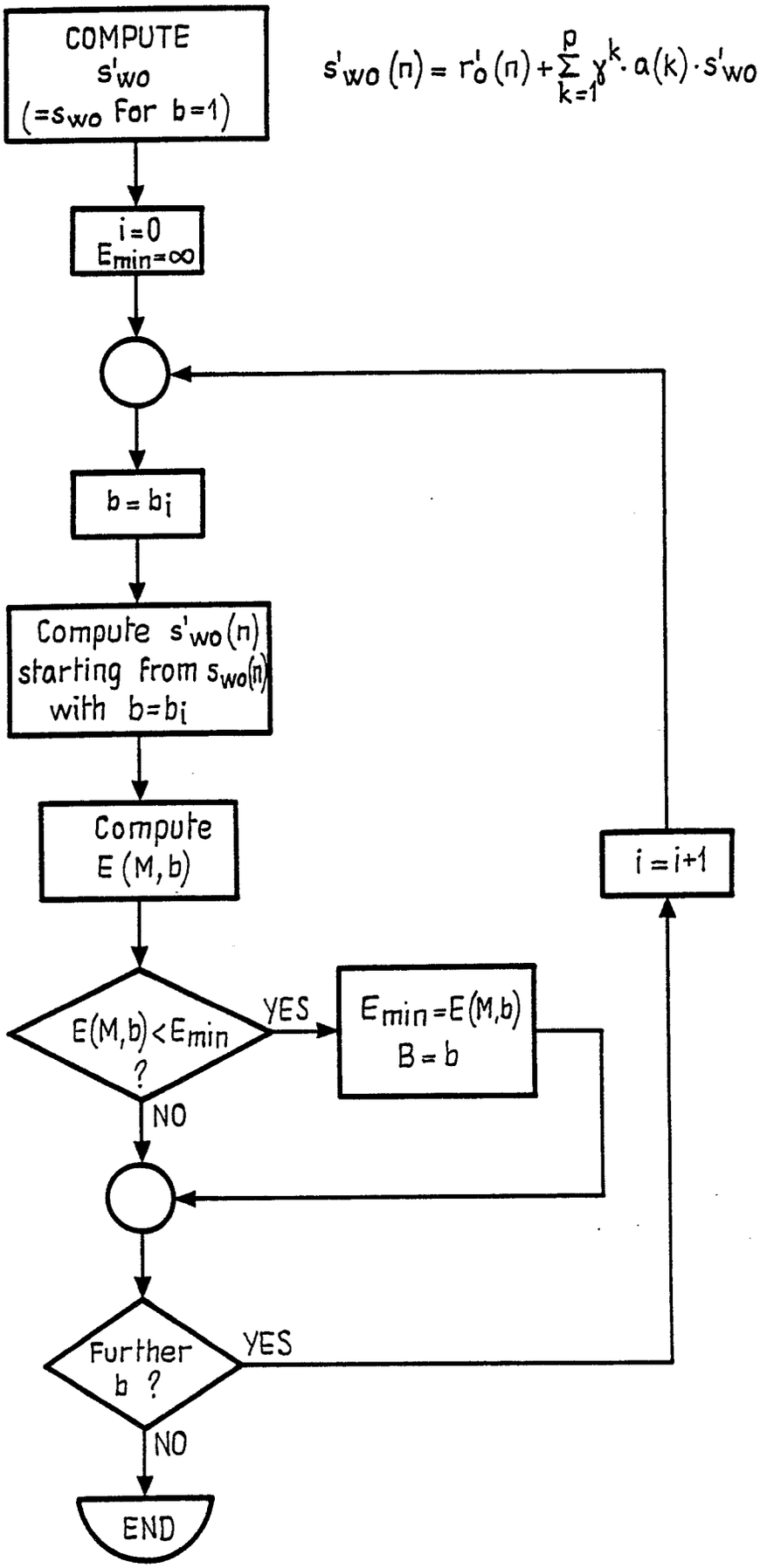
COMPUTE
s'wo
(=swo For b=1)
$s'_{wo}(n) = r'_o(n) + \sum_{k=1}^{p} \gamma^k \cdot a(k) \cdot s'_{wo}(n-k)$
i=0
Emin=∞
b = bi
Compute s'wo(n)
starting from swo(n)
with b=bi
Compute
E (M, b)
E(M,b)<Emin
?
YES
Emin=E(M,b)
B = b
NO
i = i+1
Further
b ?
YES
NO
END

FIG. 2

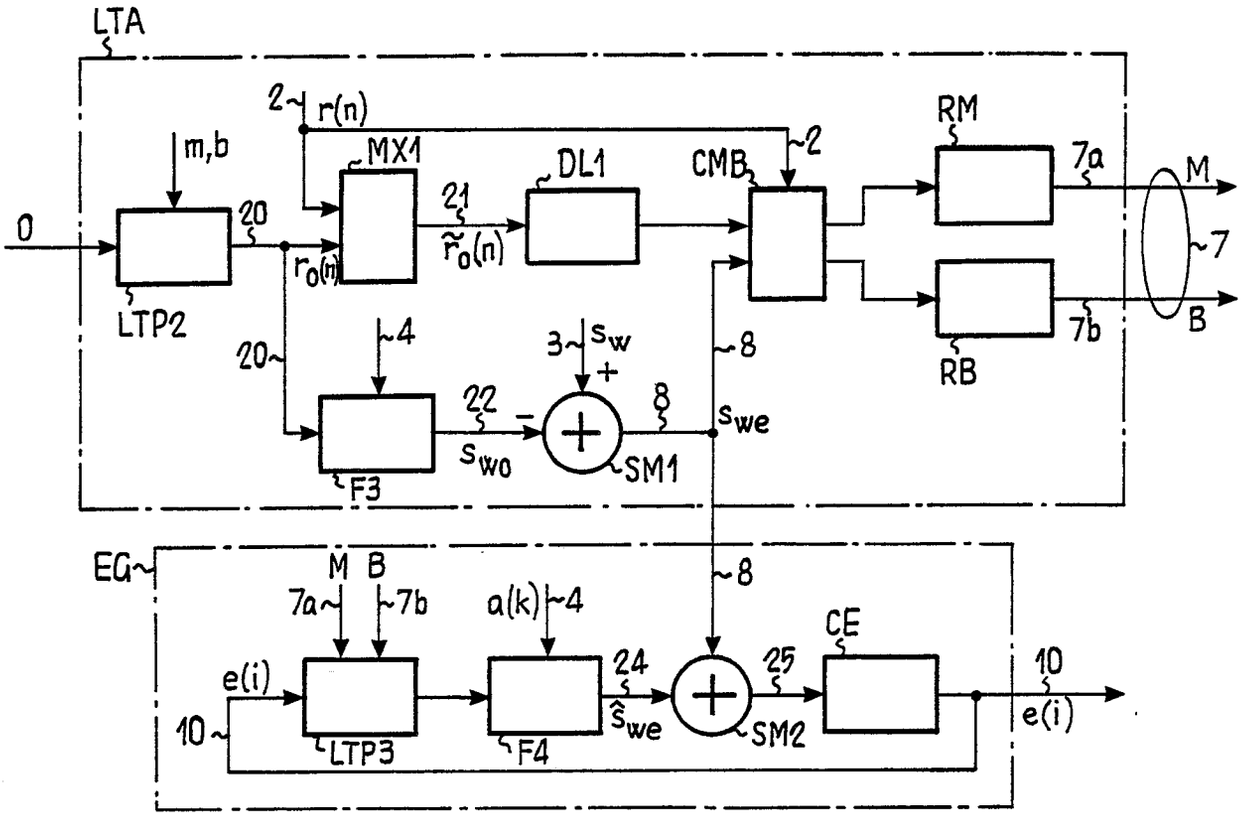
LTA
2
r(n)
m,b
MX1
DL1
CMB
2
RM
7a
M
0
20
21
$\tilde{r}_0(n)$
$r_0(n)$
LTP2
7
20
4
3
s_w
+
8
RB
7b
B
22
8
s_{we}
s_{wo}
SM1
F3
EG
M B
8
7a
7b
a(k)
4
CE
e(i)
24
25
10
e(i)
10
$\hat{s}_{we}$
SM2
LTP3
F4

FIG.3