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

(11) EP 2 633 521 B1

(12) **EUROPEAN PATENT SPECIFICATION**

(45) Date of publication and mention of the grant of the patent:
01.08.2018 Bulletin 2018/31

(21) Application number: **11835383.8**

(22) Date of filing: **24.10.2011**

(51) Int Cl.:
G10L 19/08 *(2013.01)* ***G10L 19/20*** *(2013.01)*
G10L 19/02 *(2013.01)*

(86) International application number:
PCT/CA2011/001182

(87) International publication number:
WO 2012/055016 (03.05.2012 Gazette 2012/18)

(54) **CODING GENERIC AUDIO SIGNALS AT LOW BITRATES AND LOW DELAY**

KODIERUNG GENERISCHER AUDIOSIGNALE BEI NIEDRIGEN BITRATEN UND GERINGER VERZÖGERUNG

CODAGE DE SIGNAUX AUDIO GÉNÉRIQUES À FAIBLE DÉBIT BINAIRE ET À FAIBLE RETARD

(84) Designated Contracting States:
AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR

(30) Priority: **25.10.2010 US 406379 P**

(43) Date of publication of application:
04.09.2013 Bulletin 2013/36

(60) Divisional application:
17175692.7 / 3 239 979

(73) Proprietor: **VoiceAge Corporation**
Quebec H3R 2H6 (CA)

(72) Inventors:
- **VAILLANCOURT, Tommy**
  Sherbrooke, Québec J1N 2K1 (CA)
- **JELINEK, Milan**
  Sherbrooke, Québec J1L 2W8 (CA)

(74) Representative: **Ipside**
6, Impasse Michel Labrousse
31100 Toulouse (FR)

(56) References cited:
EP-A1- 2 146 344 US-A1- 2007 225 971

- **LEFEBVRE R ET AL: "High quality coding of wideband audio signals using transform coded excitation (TCX)", PROCEEDINGS OF ICASSP '94. IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING; 19-22 APRIL 1994; ADELAIDE, SA, AUSTRALIA, IEEE SERVICE CENTER, PISCATAWAY, NJ, vol. i, 19 April 1994 (1994-04-19), pages I/193-I/196, XP010133560, DOI: 10.1109/ICASSP.1994.389322 ISBN: 978-0-7803-1775-8**
- **SCHNITZLER J ET AL: "Wideband speech coding using forward/backward adaptive prediction with mixed time/frequency domain excitation", SPEECH CODING PROCEEDINGS, 1999 IEEE WORKSHOP ON PORVOO, FINLAND 20-23 JUNE 1999, PISCATAWAY, NJ, USA,IEEE, US, 20 June 1999 (1999-06-20), pages 4-6, XP010345568, DOI: 10.1109/SCFT.1999.781465 ISBN: 978-0-7803-5651-1**
- **JUIN-HWEY CHEN ET AL: "Transform predictive coding of wideband speech signals", 1996 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING CONFERENCE PROCEEDINGS, vol. 1, 1 January 1996 (1996-01-01), pages 275-278, XP055161975, DOI: 10.1109/ICASSP.1996.540411 ISBN: 978-0-78-033192-1**

- YELDENER ET AL.: 'A Mixed Sinusoidally Excited Linear Prediction Coder at 4 kb/s and Below' PROCEEDINGS OF THE 1998 INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING vol. 2, 1998, pages 589 - 592, XP010279254
- YELDENER ET AL.: 'A High Quality Speech Coding Algorithm Suitable for Future INMARSAT Systems' PROCEEDINGS OF THE 7TH EUROPEAN SIGNAL PROCESSING CONFERENCE (EUSIPCO-94) September 1994, pages 407 - 410, XP008169171
- YELDENER ET AL.: 'Multiband Linear Predictive Speech Coding at Very Low Bit Rates' IEE PROCEEDINGS - VISION, IMAGE AND SIGNAL PROCESSING vol. 141, no. 5, October 1994, pages 289 - 296, XP006002237
- GRIFFIN ET AL.: 'Multiband Excitation Vocoder' IEEE TRANSACTIONS ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING vol. 36, no. 8, August 1988, pages 1223 - 1235, XP011516260

Description

FIELD

[0001] The present disclosure relates to mixed time-domain / frequency-domain coding devices and methods for coding an input sound signal, and to corresponding encoder and decoder using these mixed time-domain / frequency-domain coding devices and methods.

BACKGROUND

[0002] A state-of-the-art conversational codec can represent with a very good quality a clean speech signal with a bit rate of around 8 kbps and approach transparency at a bit rate of 16 kbps. However, at bitrates below 16 kbps, low processing delay conversational codecs, most often coding the input speech signal in time-domain, are not suitable for generic audio signals, like music and reverberant speech. To overcome this drawback, switched codecs have been introduced, basically using the time-domain approach for coding speech-dominated input signals and a frequency-domain approach for coding generic audio signals. However, such switched solutions typically require longer processing delay, needed both for speech-music classification and for transform to the frequency domain.

[0003] "High quality coding of wideband audio signals using transform coded excitation (TCX)", Lefebvre et al., Proceedings of ICASSP ’94, proposes an approach using a combination of time-domain and frequency-domain techniques. The proposed approach utilizes a synthesis model similar to that of linear prediction coders such as CELP. However, at the encoder, the high complexity analysis-by-synthesis is by-passed by direct quantizing of the so-called target signal (the perceptually weighted signal with removed filter ringing and pitch correlations) in the frequency domain. The innovative excitation is derived at the decoder by inverse filtering the quantized target signal.

[0004] To overcome the above drawback, a more unified time-domain and frequency-domain model is proposed.

SUMMARY

[0005] The present disclosure relates to a mixed time-domain / frequency-domain coding device for coding an input sound signal, comprising: a calculator of a time-domain excitation contribution in response to the input sound signal; a calculator of a cut-off frequency for the time-domain excitation contribution in response to the input sound signal; a filter responsive to the cut-off frequency for adjusting a frequency extent of the time-domain excitation contribution; a calculator of a frequency-domain excitation contribution in response to the input sound signal; and an adder of the filtered time-domain excitation contribution and the frequency-domain excitation contribution in the frequency domain to form a mixed time-domain / frequency-domain excitation constituting a coded version of the input sound signal.

[0006] The present disclosure also relates to an encoder using a time-domain and frequency-domain model, comprising: a classifier of an input sound signal as speech or non-speech; a time-domain only coder; the above described mixed time-domain / frequency-domain coding device; and a selector of one of the time-domain only coder and the mixed time-domain / frequency-domain coding device for coding the input sound signal depending on the classification of the input sound signal.

[0007] The present disclosure further relates to a decoder for decoding a sound signal coded using the mixed time-domain / frequency-domain coding device as described above, comprising: a converter of the mixed time-domain / frequency-domain excitation in time-domain; and a synthesis filter for synthesizing the sound signal in response to the mixed time-domain / frequency-domain excitation converted in time-domain.

[0008] The present disclosure is also concerned with a mixed time-domain / frequency-domain coding method for coding an input sound signal, comprising: calculating a time-domain excitation contribution in response to the input sound signal; calculating a cut-off frequency for the time-domain excitation contribution in response to the input sound signal; in response to the cut-off frequency, adjusting a frequency extent of the time-domain excitation contribution; calculating a frequency-domain excitation contribution in response to the input sound signal; and adding the adjusted time-domain excitation contribution and the frequency-domain excitation contribution in the frequency domain to form a mixed time-domain / frequency-domain excitation constituting a coded version of the input sound signal.

[0009] In the present disclosure, there is further described a method of encoding using a time-domain and frequency-domain model, comprising: classifying an input sound signal as speech or non-speech; providing a time-domain only coding method; providing the above described mixed time-domain / frequency-domain coding method, and selecting one of the time-domain only coding method and the mixed time-domain / frequency-domain coding method for coding the input sound signal depending on the classification of the input sound signal.

[0010] In the present disclosure, there is still further described a method of decoding a sound signal coded using the mixed time-domain / frequency-domain coding method as described above, comprising: converting the mixed time-domain / frequency-domain excitation in time-domain; and synthesizing the sound signal through a synthesis filter in response to the mixed time-domain / frequency-domain excitation converted in time-domain.

[0011] The foregoing and other features will become more apparent upon reading of the following non restrictive description of an illustrative embodiment of the proposed time-domain and frequency-domain model, given by way of example only with reference to the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0012] In the appended drawings:

Figure 1 is a schematic block diagram illustrating an overview of an enhanced CELP (Code-Excited Linear Prediction) encoder, for example an ACELP (Algebraic Code-Excited Linear Prediction) encoder;

Figure 2 is a schematic block diagram of a more detailed structure of the enhanced CELP encoder of Figure 1;

Figure 3 is a schematic block diagram of an overview of a calculator of cut-off frequency;

Figure 4 is a schematic block diagram of a more detailed structure of the calculator of cut-off frequency of Figure 3;

Figure 5 is a schematic block diagram of an overview of a frequency quantizer; and

Figure 6 is a schematic block diagram of a more detailed structure of the frequency quantizer of Figure 5.

DETAILED DESCRIPTION

[0013] The proposed more unified time-domain and frequency-domain model is able to improve the synthesis quality for generic audio signals such as, for example, music and/or reverberant speech, without increasing the processing delay and the bitrate. This model operates for example in a Linear Prediction (LP) residual domain where the available bits are dynamically allocated among an adaptive codebook, one or more fixed codebooks (for example an algebraic codebook, a Gaussian codebook, etc.), and a frequency-domain coding mode, depending upon the characteristics of the input signal.

[0014] To achieve a low processing delay low bit rate conversational codec that improves the synthesis quality of generic audio signals like music and/or reverberant speech, a frequency-domain coding mode may be integrated as close as possible to the CELP (Code-Excited Linear Prediction) time-domain coding mode. For that purpose, the frequency-domain coding mode uses, for example, a frequency transform performed in the LP residual domain. This allows switching nearly without artifact from one frame, for example a 20 ms frame, to another. Also, the integration of the two (2) coding modes is sufficiently close to allow dynamic reallocation of the bit budget to another coding mode if it is determined that the current coding mode is not efficient enough.

[0015] One feature of the proposed more unified time-domain and frequency-domain model is the variable time support of the time-domain component, which varies from quarter frame to a complete frame on a frame by frame basis, and will be called sub-frame. As an illustrative example, a frame represents 20 ms of input signal. This corresponds to 320 samples if the inner sampling frequency of the codec is 16 kHz or to 256 samples per frame if the inner sampling frequency of the codec is 12.8 kHz. Then a quarter of a frame (the sub-frame) represents 64 or 80 samples depending on the inner sampling frequency of the codec. In the following illustrative embodiment the inner sampling frequency of the codec is 12.8 kHz giving a frame length of 256 samples. The variable time support makes it possible to capture major temporal events with a minimum bitrate to create a basic time-domain excitation contribution. At very low bit rate, the time support is usually the entire frame. In that case, the time-domain contribution to the excitation signal is composed only of the adaptive codebook, and the corresponding pitch information with the corresponding gain are transmitted once per frame. When more bitrate is available, it is possible to capture more temporal events by shortening the time support (and increasing the bitrate allocated to the time-domain coding mode). Eventually, when the time support is sufficiently short (down to quarter a frame), and the available bitrate is sufficiently high, the time-domain contribution may include the adaptive codebook contribution, a fixed-codebook contribution, or both, with the corresponding gains. The parameters describing the codebook indices and the gains are then transmitted for each sub-frame.

[0016] At low bit rate, conversational codecs are not capable of coding properly higher frequencies. This causes an important degradation of the synthesis quality when the input signal includes music and/or reverberant speech. To solve this issue, a feature is added to compute the efficiency of the time-domain excitation contribution. In some cases, whatever the input bitrate and the time frame support are, the time-domain excitation contribution is not valuable. In those cases, all the bits are reallocated to the next step of frequency-domain coding. But most of the time, the time-domain excitation contribution is valuable up only to a certain frequency (the cut-off frequency). In these cases, the time-domain excitation contribution is filtered out above the cut-off frequency. The filtering operation permits to keep valuable information coded with the time-domain excitation contribution and remove the non-valuable information above the cut-off frequency. In an illustrative embodiment, the filtering is performed in the frequency domain by setting the frequency bins above a certain frequency to zero.

[0017] The variable time support in combination with the variable cut-off frequency makes the bit allocation inside the integrated time-domain and frequency-domain model very dynamic. The bitrate after the quantization of the LP filter can be allocated entirely to the time domain or entirely to the frequency domain or somewhere in between. The bitrate allocation between the time and frequency domains is conducted as a function of the number of sub-frames used for the time-domain contribution, of the available bit budget, and of the cut-off frequency computed.

[0018] To create a total excitation which will match more efficiently the input residual, the frequency-domain coding mode is applied. A feature in the present disclosure is that the frequency-domain coding is performed on a vector which contains the difference between a frequency representation (frequency transform) of the input LP residual and a frequency representation (frequency transform) of the filtered time-domain excitation contribution up to the cut-off frequency, and which contains the frequency representation (frequency transform) of the input LP residual itself above that cut-off frequency. A smooth spectrum transition is inserted between both segments just above the cut-off frequency. In other words, the high-frequency part of the frequency representation of the time-domain excitation contribution is first zeroed out. A transition region between the unchanged part of the spectrum and the zeroed part of the spectrum is inserted just above the cut-off frequency to ensure a smooth transition between both parts of the spectrum. This modified spectrum of the time-domain excitation contribution is then subtracted from the frequency representation of the input LP residual. The resulting spectrum thus corresponds to the difference of both spectra below the cut-off frequency, and to the frequency representation of the LP residual above it, with some transition region. The cut-off frequency, as mentioned hereinabove, can vary from one frame to another.

[0019] Whatever the frequency quantization method (frequency-domain coding mode) chosen, there is always a possibility of pre-echo especially with long windows. In this technique, the used windows are square windows, so that the extra window length compared to the coded signal is zero (0), i.e. no overlap-add is used. While this corresponds to the best window to reduce any potential pre-echo, some pre-echo may still be audible on temporal attacks. Many techniques exist to solve such pre-echo problem but the present disclosure proposes a simple feature for cancelling this pre-echo problem. This feature is based on a memory-less time-domain coding mode which is derived from the "Transition Mode" of ITU-T Recommendation G.718; Reference [ITU-T Recommendation G.718 "Frame error robust narrow-band and wideband embedded variable bit-rate coding of speech and audio from 8-32 kbit/s", June 2008, section 6.8.1.4 and section 6.8.4.2]. The idea behind this feature is to take advantage of the fact that the proposed more unified time-domain and frequency-domain model is integrated to the LP residual domain, which allows for switching without artifact almost at any time. When a signal is considered as generic audio (music and/or reverberant speech) and when a temporal attack is detected in a frame, then this frame only is encoded with this special memory-less time-domain coding mode. This mode will take care of the temporal attack thus avoiding the pre-echo that could be introduced with the frequency-domain coding of that frame.

ILLUSTRATIVE EMBODIMENT

[0020] In the proposed more unified time-domain and frequency-domain model, the above mentioned adaptive codebook, one or more fixed codebooks (for example an algebraic codebook, a Gaussian codebook, etc.), i.e. the so called time-domain codebooks, and the frequency-domain quantization (frequency-domain coding mode can be seen as a codebook library, and the bits can be distributed among all the available codebooks, or a subset thereof. This means for example that if the input sound signal is a clean speech, all the bits will be allocated to the time-domain coding mode, basically reducing the coding to the legacy CELP scheme. On the other hand, for some music segments, all the bits allocated to encode the input LP residual are sometimes best spent in the frequency domain, for example in a transform-domain.

[0021] As indicated in the foregoing description, the temporal support for the time-domain and frequency-domain coding modes does not need to be the same. While the bits spent on the different time-domain quantization methods (adaptive and algebraic codebook searches) are usually distributed on a sub-frame basis (typically a quarter of a frame, or 5 ms of time support), the bits allocated to the frequency-domain coding mode are distributed on a frame basis (typically 20 ms of time support) to improve frequency resolution.

[0022] The bit budget allocated to the time-domain CELP coding mode can be also dynamically controlled depending on the input sound signal. In some cases, the bit budget allocated to the time-domain CELP coding mode can be zero, effectively meaning that the entire bit budget is attributed to the frequency-domain coding mode. The choice of working in the LP residual domain both for the time-domain and the frequency-domain approaches has two (2) main benefits. First, this is compatible with the CELP coding mode, proved efficient in speech signals coding. Consequently, no artifact is introduced due to the switching between the two types of coding modes. Second, lower dynamics of the LP residual with respect to the original input sound signal, and its relative flatness, make easier the use of a square window for the frequency transforms thus permitting use of a non-overlapping window.

[0023] In a non limitative example where the inner sampling frequency of the codec is 12.8 kHz (meaning 256 samples per frame), similarly as in the ITU-T recommendation G.718, the length of the sub-frames used in the time-domain CELP coding mode can vary from a typical ¼ of the frame length (5 ms) to a half frame (10 ms) or a complete frame length (20 ms). The sub-frame length decision is based on the available bitrate and on an analysis of the input sound signal, particularly the spectral dynamics of this input sound signal. The sub-frame length decision can be performed in a closed loop manner. To save on complexity, it is also possible to base the sub-frame length decision in an open loop manner. The sub-frame length can be changed from frame to frame.

[0024] Once the length of the sub-frames is chosen in a particular frame, a standard closed-loop pitch analysis is performed and the first contribution to the excitation signal is selected from the adaptive codebook. Then, depending on the available bit budget and the characteristics of the input sound signal (for example in the case of an input speech signal), a second contribution from one or several fixed codebooks can be added before the transform-domain coding. The resulting excitation will be called the time-domain excitation contribution. On the other hand, at very low bit rates and in case of generic audio, it is often better to skip the fixed codebook stage and use all the remaining bits for the transform-domain coding mode. The transform domain coding mode can be for example a frequency-domain coding mode. As described above, the sub-frame length can be one fourth of the frame, one half of the frame, or one frame long. The fixed-codebook contribution is used only if the sub-frame length is equal to one fourth of the frame length. In case the sub-frame length is decided to be half a frame or the entire frame long, then only the adaptive-codebook contribution is used to represent the time-domain excitation, and all remaining bits are allocated to the frequency-domain coding mode.

[0025] Once the computation of the time-domain excitation contribution is completed, its efficiency needs to be assessed and quantized. If the gain of the coding in time-domain is very low, it is more efficient to remove the time-domain excitation contribution altogether and to use all the bits for the frequency-domain coding mode instead. On the other hand, for example in the case of a clean input speech, the frequency-domain coding mode is not needed and all the bits are allocated to the time-domain coding mode. But often the coding in time-domain is efficient only up to a certain frequency. This frequency will be called the cut-off frequency of the time-domain excitation contribution. Determination of such cut-off frequency ensures that the entire time-domain coding is helping to get a better final synthesis rather than working against the frequency-domain coding.

[0026] The cut-off frequency is estimated in the frequency-domain. To compute the cut-off frequency, the spectrums of both the LP residual and the time-domain coded contribution are first split into a predefined number of frequency bands. The number of frequency bands and the number of frequency bins covered by each frequency band can vary from one implementation to another. For each of the frequency bands, a normalized correlation is computed between the frequency representation of the time-domain excitation contribution and the frequency representation of the LP residual, and the correlation is smoothed between adjacent frequency bands. The per-band correlations are lower limited to 0.5 and normalized between 0 and 1. The average correlation is then computed as the average of the correlations for all the frequency bands. For the purpose of a first estimation of the cut-off frequency, the average correlation is then scaled between 0 and half the sampling rate (half the sampling rate corresponding to the normalized correlation value of 1). The first estimation of the cut-off frequency is then found as the upper bound of the frequency band being closest to that value. In an example of implementation, sixteen (16) frequency bands at 12.8 kHz are defined for the correlation computation.

[0027] Taking advantage of the psychoacoustic property of the human ear, the reliability of the estimation of the cut-off frequency is improved by comparing the estimated position of the 8^{th} harmonic frequency of the pitch to the cut-off frequency estimated by the correlation computation. If this position is higher than the cut-off frequency estimated by the correlation computation, the cut-off frequency is modified to correspond to the position of the 8^{th} harmonic frequency of the pitch. The final value of the cut-off frequency is then quantized and transmitted. In an example of implementation, 3 or 4 bits are used for such quantization, giving 8 or 16 possible cut-off frequencies depending on the bit rate.

[0028] Once the cut-off frequency is known, frequency quantization of the frequency-domain excitation contribution is performed. First the difference between the frequency representation (frequency transform) of the input LP residual and the frequency representation (frequency transform) of the time-domain excitation contribution is determined. Then a new vector is created, consisting of this difference up to the cut-off frequency, and a smooth transition to the frequency representation of the input LP residual for the remaining spectrum. A frequency quantization is then applied to the whole new vector. In an example of implementation, the quantization consists in coding the sign and the position of dominant (most energetic) spectral pulses. The number of the pulses to be quantized per frequency band is related to the bitrate available for the frequency-domain coding mode. If there are not enough bits available to cover all the frequency bands, the remaining bands are filled with noise only.

[0029] Frequency quantization of a frequency band using the quantization method described in the previous paragraph does not guarantee that all frequency bins within this band are quantized. This is especially true at low bitrates where the number of pulses quantized per frequency band is relatively low. To prevent the apparition of audible artifacts due to these non-quantized bins, some noise is added to fill these gaps. As at low bit rates the quantized pulses should dominate the spectrum rather than the inserted noise, the noise spectrum amplitude corresponds only to a fraction of the amplitude of the pulses. The amplitude of the added noise in the spectrum is higher when the bit budget available is low (allowing more noise) and lower when the bit budget available is high.

[0030] In the frequency-domain coding mode, gains are computed for each frequency band to match the energy of the non-quantized signal to the quantized signal. The gains are vector quantized and applied per band to the quantized signal. When the encoder changes its bit allocation from the time-domain only coding mode to the mixed time-domain / frequency-domain coding mode, the per band excitation spectrum energy of the time-domain only coding mode does not match the per band excitation spectrum energy of the mixed time-domain / frequency domain coding mode. This energy mismatch can create some switching artifacts especially at low bit rate. To reduce any audible degradation created by this bit reallocation, a long-term gain can be computed for each band and can be applied to correct the energy of each frequency band for a few frames after the switching from the time-domain coding mode to the mixed time-domain / frequency-domain coding mode.

[0031] After the completion of the frequency-domain coding mode, the total excitation is found by adding the frequency-domain excitation contribution to the frequency representation (frequency transform) of the time-domain excitation contribution and then the sum of the excitation contributions is transformed back to time-domain to form a total excitation. Finally, the synthesized signal is computed by filtering the total excitation through a LP synthesis filter. In one embodiment, while the CELP coding memories are updated on a sub-frame basis using only the time-domain excitation contribution, the total excitation is used to update those memories at frame boundaries. In another possible implementation, the CELP coding memories are updated on a sub-frame basis and also at the frame boundaries using only the time-domain excitation contribution. This results in an embedded structure where the frequency-domain quantized signal constitutes an upper quantization layer independent of the core CELP layer. In this particular case, the fixed codebook is always used in order to update the adaptive codebook content. However, the frequency-domain coding mode can apply to the whole frame. This embedded approach works for bit rates around 12 kbps and higher.

1) Sound type classification

[0032] Figure 1 is a schematic block diagram illustrating an overview of an enhanced CELP encoder 100, for example an ACELP encoder. Of course, other types of enhanced CELP encoders can be implemented using the same concept. Figure 2 is a schematic block diagram of a more detailed structure of the enhanced CELP encoder 100.

[0033] The CELP encoder 100 comprises a pre-processor 102 (Figure 1) for analyzing parameters of the input sound signal 101 (Figures 1 and 2). Referring to Figure 2, the pre-processor 102 comprises an LP analyzer 201 of the input sound signal 101, a spectral analyzer 202, an open loop pitch analyzer 203, and a signal classifier 204. The analyzers 201 and 202 perform the LP and spectral analyses usually carried out in CELP coding, as described for example in ITU-T recommendation G.718, sections 6.4 and 6.1.4, and, therefore, will not be further described in the present disclosure.

[0034] The pre-processor 102 conducts a first level of analysis to classify the input sound signal 101 between speech and non-speech (generic audio (music or reverberant speech)), for example in a manner similar to that described in reference [T.Vaillancourt et al., "Inter-tone noise reduction in a low bit rate CELP decoder," Proc. IEEE ICASSP, Taipei, Taiwan, Apr. 2009, pp. 4113-16], of which the full content is incorporated herein by reference, or with any other reliable speech/non-speech discrimination methods.

[0035] After this first level of analysis, the pre-processor 102 performs a second level of analysis of input signal parameters to allow the use of time-domain CELP coding (no frequency-domain coding) on some sound signals with strong non-speech characteristics, but that are still better encoded with a time-domain approach. When an important variation of energy occurs, this second level of analysis allows the CELP encoder 100 to switch into a memory-less time-domain coding mode, generally called Transition Mode in reference [Eksler, V., and Jelínek, M. (2008), "Transition mode coding for source controlled CELP codecs", IEEE Proceedings of International Conference on Acoustics, Speech and Signal Processing, March-April, pp. 4001-40043], of which the full content is incorporated herein by reference.

[0036] During this second level of analysis, the signal classifier 204 calculates and uses a variation σ_C of a smoothed version C_{st} of the open-loop pitch correlation from the open-loop pitch analyzer 203, a current total frame energy E_{tot} and a difference between the current total frame energy and the previous total frame energy E_{diff}. First the variation of the smoothed open loop pitch correlation is computed as:

$$\sigma_C = \sqrt{\sum_{i=0}^{i=-10}\left(\frac{\left(C_{st}(i)-\overline{C_{st}}\right)^2}{10}\right)}$$

where:

C_{st} is the smoothed open-loop pitch correlation defined as:

$$\underline{C_{st} = 0.9 \cdot C_{ol} + 0.1 \cdot C_{st}};$$

C_{ol} is the open-loop pitch correlation calculated by the analyzer 203 using a method known to those of ordinary skill in the art of CELP coding, for example, as described in ITU-T recommendation G.718, Section 6.6;

$\overline{C_{st}}$ is the average over the last 10 frames of the smoothed open-loop pitch correlation C_{st};

σ_C is the variation of the smoothed open loop pitch correlation.

[0037] When, during the first level of analysis, the signal classifier 204 classifies a frame as non-speech, the following verifications are performed by the signal classifier 204 to determine, in the second level of analysis, if it is really safe to use a mixed time-domain / frequency-domain coding mode. Sometimes, it is however better to encode the current frame with the time-domain coding mode only, using one of the time-domain approaches estimated by the pre-processing function of the time-domain coding mode. In particular, it might be better to use the memory-less time-domain coding mode to reduce at a minimum any possible pre-echo that can be introduced with a mixed time-domain/frequency-domain coding mode.

[0038] As a first verification whether the mixed time-domain / frequency-domain coding should be used, the signal classifier 204 calculates a difference between the current total frame energy and the previous frame total energy. When the difference E_{diff} between the current total frame energy E_{tot} and the previous frame total energy is higher than 6 dB, this corresponds to a so-called "temporal attack" in the input sound signal. In such a situation, the speech/non-speech decision and the coding mode selected are overwritten and a memory-less time-domain coding mode is forced. More specifically, the enhanced CELP encoder 100 comprises a time-only/time-frequency coding selector 103 (Figure 1) itself comprising a speech/generic audio selector 205 (Figure 2), a temporal attack detector 208 (Figure 2), and a selector 206 of memory-less time-domain coding mode. In other words, in response to a determination of non-speech signal (generic audio) by the selector 205 and detection of a temporal attack in the input sound signal by the detector 208, the selector 206 forces a closed-loop CELP coder 207 (Figure 2) to use the memory-less time-domain coding mode. The closed-loop CELP coder 207 forms part of the time-domain-only coder 104 of Figure 1.

[0039] As a second verification, when the difference E_{diff} between the current total frame energy E_{tot} E_{tot} and the previous frame total energy is below or equal to 6 dB, but:

- the smoothed open loop pitch correlation C_{st} is higher than 0.96; or
- the smoothed open loop pitch correlation C_{st} is higher than 0.85 and the difference E_{diff} between the current total frame energy E_{tot} and the previous frame total energy is below 0.3 dB ; or
- the variation of the smoothed open loop pitch correlation σ_C is below 0.1 and the difference E_{diff} between the current total frame energy $\underline{E_{tot}}$ and the last previous frame total energy is below 0.6 dB; or
- the current total frame energy E_{tot} is below 20 dB;

and this is at least the second consecutive frame ($cnt \geq 2$) where the decision of the first level of the analysis is going to be changed, then the speech/generic audio selector 205 determines that the current frame will be coded using a time-domain only mode using the closed-loop generic CELP coder 207 (Figure 2).

[0040] Otherwise, the time/time-frequency coding selector 103 selects a mixed time-domain/frequency-domain coding mode that is performed by a mixed time-domain/frequency-domain coding device disclosed in the following description.

[0041] This can be summarized, for example when the non-speech sound signal is music, with the following pseudo code:

$$
\begin{aligned}
& if\left(generic\ audio\right)\\
& \qquad if\left(E_{diff} > 6dB\right)\\
& \qquad\qquad \text{coding mode} = \text{Time domain memory less}\\
& \qquad\qquad \text{cnt} = 1\\
& \qquad else\ if\left(C_{st} > 0.96 \mid \left(C_{st} > 0.85\ \&\ E_{diff} < 0.3dB\right) \mid \left(\sigma_C < 0.1\ \&\ E_{diff} < 0.6dB\right) \mid E_{tot} < 20dB\right)\\
& \qquad\qquad cnt++\\
& \qquad\qquad if\left(cnt >= 2\right)\\
& \qquad\qquad\qquad \text{coding mode} = \text{Time domain}\\
& \qquad else\\
& \qquad\qquad \text{coding mode} = \text{mix time/frequency domain}\\
& \qquad\qquad cnt = 0
\end{aligned}
$$

[0042] Where E_{tot} is a current frame energy expressed as:

$$E_{tot} = 10\log\left(\frac{\sum_{i=0}^{i=N} x(i)^2}{N}\right)$$

(where $x(i)$ represents the samples of the input sound signal in the frame) and E_{diff} is the difference between the current total frame energy E_{tot} and the last previous frame total energy.

2) Decision on sub-frame length

[0043] In typical CELP, input sound signal samples are processed in frames of 10-30 ms and these frames are divided into several sub-frames for adaptive codebook and fixed codebook analysis. For example, a frame of 20 ms (256 samples when the inner sampling frequency is 12.8 kHz) can be used and divided into 4 sub-frames of 5 ms. A variable sub-frame length is a feature used to obtain complete integration of the time-domain and frequency-domain into one coding mode. The sub-frame length can vary from a typical ¼ of the frame length to a half frame or a complete frame length. Of course the use of another number of sub-frames (sub-frame length) can be implemented.

[0044] The decision as to the length of the sub-frames (the number of sub-frames), or the time support, is determined by a calculator of the number of sub-frames 210 based on the available bitrate and on the input signal analysis in the pre-processor 102, in particular the high frequency spectral dynamic of the input sound signal 101 from an analyzer 209 and the open-loop pitch analysis including the smoothed open loop pitch correlation from analyzer 203. The analyzer 209 is responsive to the information from the spectral analyzer 202 to determine the high frequency spectral dynamic of the input signal 101. The spectral dynamic is computed from a feature described in the ITU-T recommendation G.718, section 6.7.2.2, as the input spectrum without its noise floor giving a representation of the input spectrum dynamic. When the average spectral dynamic of the input sound signal 101 in the frequency band between 4.4 kHz and 6.4 kHz as determined by the analyzer 209 is below 9.6 dB and the last frame was considered as having a high spectral dynamic, the input signal 101 is no longer considered as having high spectral dynamic content in higher frequencies. In that case, more bits can be allocated to the frequencies below, for example, 4 kHz, by adding more sub-frames to the time-domain coding mode or by forcing more pulses in the lower frequency part of the frequency-domain contribution.

[0045] On the other hand, if the increase of the average dynamic of the higher frequency content of the input signal 101 against the average spectral dynamic of the last frame that was not considered as having a high spectral dynamic as determined by the analyser 209 is greater than, for example, 4.5 dB, the sound input signal 101 is considered as having high spectral dynamic content above, for example, 4 kHz. In that case, depending on the available bit rate, some additional bits are used for coding the high frequencies of the input sound signal 101 to allow one or more frequency pulses encoding.

[0046] The sub-frame length as determined by the calculator 210 (Figure 2) is also dependent on the bit budget available. At very low bit rate, e.g. bit rates below 9 kbps, only one sub-frame is available for the time-domain coding otherwise the number of available bits will be insufficient for the frequency-domain coding. For medium bit rates, e.g. bit rates between 9 kbps and 16 kbps, one sub-frame is used for the case where the high frequencies contain high dynamic spectral content and two sub-frames if not. For medium-high bit rates, e.g. bit rates around 16 kbps and higher, the four (4) sub-frames case becomes also available if the smoothed open loop pitch correlation C_{st}, as defined in paragraph [0037] of sound type classification section, is higher than 0.8.

[0047] While the case with one or two sub-frames limits the time-domain coding to an adaptive codebook contribution only (with coded pitch lag and pitch gain), i.e. no fixed codebook is used in that case, the four (4) sub-frames allow for adaptive and fixed codebook contributions if the available bit budget is sufficient. The four (4) sub-frame case is allowed starting from around 16 kbps up. Because of bit budget limitations, the time-domain excitation consists only of the adaptive codebook contribution at lower bitrates. Simple fixed codebook contribution can be added for higher bit rates, for example starting at 24 kbps. For all cases the time-domain coding efficiency will be evaluated afterward to decide up to which frequency such time-domain coding is valuable.

3) Closed loop pitch analysis

[0048] When a mixed time-domain / frequency-domain coding mode is used, a closed loop pitch analysis followed, if needed, by a fixed algebraic codebook search are performed. For that purpose, the CELP encoder 100 (Figure 1) comprises a calculator of time-domain excitation contribution 105 (Figures 1 and 2). This calculator further comprises an analyzer 211 (Figure 2) responsive to the open-loop pitch analysis conducted in the open-loop pitch analyzer 203 and the sub-frame length (or the number of sub-frames in a frame) determination in calculator 210 to perform a closed-loop pitch analysis. The closed-loop pitch analysis is well known to those of ordinary skill in the art and an example of implementation is described for example in reference [ITU-T G.718 recommendation; Section 6.8.4.1.4.1], the full content thereof being incorporated herein by reference. The closed-loop pitch analysis results in computing the pitch parameters, also known as adaptive codebook parameters, which mainly consist of a pitch lag (adaptive codebook index T) and pitch gain (or adaptive codebook gain b). The adaptive codebook contribution is usually the past excitation at delay T or an interpolated version thereof. The adaptive codebook index T is encoded and transmitted to a distant decoder. The pitch gain b is also quantized and transmitted to the distant decoder.

[0049] When the closed loop pitch analysis has been completed, the CELP encoder 100 comprises a fixed codebook 212 searched to find the best fixed codebook parameters usually comprising a fixed codebook index and a fixed codebook gain. The fixed codebook index and gain form the fixed codebook contribution. The fixed codebook index is encoded and transmitted to the distant decoder. The fixed codebook gain is also quantized and transmitted to the distant decoder. The fixed algebraic codebook and searching thereof is believed to be well known to those of ordinary skill in the art of CELP coding and, therefore, will not be further described in the present disclosure.

[0050] The adaptive codebook index and gain and the fixed codebook index and gain form a time-domain CELP excitation contribution.

4) Frequency transform of signal of interest

[0051] During the frequency-domain coding of the mixed time-domain / frequency-domain coding mode, two signals need to be represented in a transform-domain, for example in frequency domain. In one embodiment, the time-to-frequency transform can be achieved using a 256 points type II (or type IV) DCT (Discrete Cosine Transform) giving a resolution of 25 Hz with an inner sampling frequency of 12.8 kHz but any other transform could be used. In the case another transform is used, the frequency resolution (defined above), the number of frequency bands and the number of frequency bins per bands (defined further below) might need to be revised accordingly. In this respect, the CELP encoder 100 comprises a calculator 107 (Figure 1) of a frequency-domain excitation contribution in response to the input LP residual $r_{es}(n)$ resulting from the LP analysis of the input sound signal by the analyzer 201. As illustrated in Figure 2, the calculator 107 may calculate a DCT 213, for example a type II DCT of the input LP residual $r_{es}(n)$. The CELP encoder 100 also comprises a calculator 106 (Figure 1) of a frequency transform of the time-domain excitation contribution. As illustrated in Figure 2, the calculator 106 may calculate a DCT 214, for example a type II DCT of the time-domain excitation contribution. The frequency transform of the input LP residual f_{res} and the time-domain CELP excitation contribution f_{exc} can be calculated using the following expressions:

$$f_{res}(k) = \begin{cases} \sqrt{\frac{1}{N}} \cdot \sum_{n=0}^{N-1} r_{es}(n) \cdot \cos\left(\frac{\pi}{N}\left(n + \frac{1}{2}\right)k\right), k = 0 \\ \sqrt{\frac{2}{N}} \cdot \sum_{n=0}^{N-1} r_{es}(n) \cdot \cos\left(\frac{\pi}{N}\left(n + \frac{1}{2}\right)k\right), & 1 \le k < N-1 \end{cases}$$

and:

$$f_{exc}(k) = \begin{cases} \sqrt{\frac{1}{N}} \cdot \sum_{n=0}^{N-1} \mathbf{e}_{td}(n) \cdot \cos\left(\frac{\pi}{N}\left(n - \frac{1}{2}\right)k\right), k = 0 \\ \sqrt{\frac{2}{N}} \cdot \sum_{n=0}^{N-1} \mathbf{e}_{td}(n) \cdot \cos\left(\frac{\pi}{N}\left(n + \frac{1}{2}\right)k\right), & 1 \le k < N-1 \end{cases}$$

where $r_{es}(n)$ is the input LP residual, $\mathbf{e}_{td}(n)$ is the time-domain excitation contribution, and N is the frame length. In a possible implementation, the frame length is 256 samples for a corresponding inner sampling frequency of 12.8 kHz. The time-domain excitation contribution is given by the following relation:

$$e_{td}(n) = bv(n) + gc(n)$$

where $v(n)$ is the adaptive codebook contribution, b is the adaptive codebook gain, $c(n)$ is the fixed codebook contribution, and g is the fixed codebook gain. It should be noted that the time-domain excitation contribution may consist only of the adaptive codebook contribution as described in the foregoing description.

5) Cut-off frequency of time-domain contribution

[0052] With generic audio samples, the time-domain excitation contribution (the combination of adaptive and/or fixed algebraic codebooks) does not always contribute much to the coding improvement compared to the frequency-domain coding. Often, it does improve coding of the lower part of the spectrum while the coding improvement in the higher part of the spectrum is minimal. The CELP encoder 100 comprises a finder of a cut-off frequency and filter 108 (Figure 1) that is the frequency where coding improvement afforded by the time-domain excitation contribution becomes too low to be valuable. The finder and filter 108 comprises a calculator of cut-off frequency 215 and the filter 216 of Figure 2. The cut-off frequency of the time-domain excitation contribution is first estimated by the calculator 215 (Figure 2) using a computer 303 (Figures 3 and 4) of normalized cross-correlation for each frequency band between the frequency-transformed input LP residual from calculator 107 and the frequency-transformed time-domain excitation contribution from calculator 106, respectively designated f_{res} and f_{exc} which are defined in the foregoing section 4. The last frequency L_f included in each of, for example, the sixteen (16) frequency bands are defined in Hz as:

$$L_f = \begin{Bmatrix} 175, 375, 775, 1175, 1575, 1975, 2375, 2775, \\ 3175, 3575, 3975, 4375, 4775, 5175, 5575, 6375 \end{Bmatrix}$$

[0053] For this illustrative example, the number of frequency bins per band B_b, the cumulative frequency bins per band C_{Bb}, and the normalized cross-correlation per frequency band $C_c(t)$ are defined as follows, for a 20 ms frame at 12.8 kHz sampling frequency:

$$B_b = \begin{Bmatrix} 8, 8, 16, 16, 16, 16, 16, 16, \\ 16, 16, 16, 16, 16, 16, 16, 32 \end{Bmatrix}$$

$$C_{Bb} = \begin{Bmatrix} 0,8,16,32,48,64,80,96, \\ 112,128,144,160,176,192,208,224 \end{Bmatrix}$$

$$C_C(i) = \frac{\sum_{j=C_{Bb}(i)}^{j=C_{Bb}(i)+B_b(i)} f_{exc}(j) \cdot f_{res}(j)}{\sqrt{\left(S'_{f_{exc}}(i) \cdot S'_{f_{res}}(i)\right)}}$$

Where

$$S'_{f_{exc}}(i) = \sum_{j=C_{Bb}(i)}^{j=C_{Bb}(i)+B_b(i)} f_{exc}(j)^2$$

and

$$S'_{f_{res}}(i) = \sum_{j=C_{Bb}(i)}^{j=C_{Bb}(i)+B_b(i)} f_{res}(j)^2$$

where B_b is the number of frequency bins per band B_b, C_{Bb} is the cumulative frequency bins per bands, $C_{Bb}C_c(i)C_c(i)$ is the normalized cross-correlation per frequency band, $S'_{f_{exc}}$ is the excitation energy for a band and similarly $S'_{f_{res}}$ is the residual energy per band.

[0054] The calculator of cut-off frequency 215 comprises a smoother 304 (Figures 3 and 4) of cross-correlation through the frequency bands performing some operations to smooth the cross-correlation vector between the different frequency bands. More specifically, the smoother 304 of cross-correlation through the bands computes a new cross-correlation vector C_{c_2} using the following relation:

$$C_{c_2}(i) = \begin{Bmatrix} 2 \cdot \left(\min\left(0.5,\ \alpha \cdot C_c(0) + \delta C_c(1)\right) - 0.5\right) & \text{for } i = 0 \\ 2 \cdot \left(\min\left(0.5,\ \alpha \cdot C_c(i) + \beta C_c(i+1) + \beta C_c(i-1)\right) - 0.5\right) & \text{for } 1 \le i < N_b \end{Bmatrix}$$

where

$$\alpha = 0.95; \qquad \delta = (1-\alpha); \qquad N_b = 13; \qquad \beta = \delta/2$$

[0055] The calculator of cut-off frequency 215 further comprises a calculator 305 (Figures 3 and 4) of an average of the new cross-correlation vector C_{c_2} over the first N_b bands (N_b =13 representing 5575 Hz).

[0056] The calculator 215 of cut-off frequency also comprises a cut-off frequency module 306 (Figure 3) including a limiter 406 (Figure 4) of the cross-correlation, a normaliser 407 of the cross-correlation and a finder 408 of the frequency band where the cross-correlation is the lowest. More specifically, the limiter 406 limits the average of the cross-correlation vector to a minimum value of 0.5 and the normaliser 408 normalises the limited average of the cross-correlation vector between 0 and 1. The finder 408 obtains a first estimate of the cut-off frequency by finding the last frequency of a frequency band L_f which minimizes the difference between the said last frequency of a frequency band L_f and the normalized average $\overline{C}_{c2}$ of the cross-correlation vector C_{c2} multiplied by the width F/2 of the spectrum of the input sound signal:

$$i_{min} = \min_{0 \le i < N_b}\left(L_f(i) - \overline{C_{c_2}} \cdot \left(\frac{\mathrm{F}_s}{2}\right)\right) \quad and \quad f_{tc_1} = L_f(i_{min})$$

where

$$\mathrm{F}_s = 12800\ Hz \text{ and } \overline{C_{c_2}} = \frac{\sum_{i=0}^{i=N_b-1}\left(C_{c_2}(i)\right)}{N_b}$$

[0057] f_{tc1} is the first estimate of the cut-off frequency.

[0058] At low bit rate, where the normalized average $\overline{C}_{c2}$ is never really high, or to artificially increase the value of f_{tc1} to give a little more weight to the time domain contribution, it is possible to upscale the value of $\overline{C}_{c2}$ with a fix scaling factor, for example, at bit rate below 8 kbps, f_{tc1} is multiplied by 2 all the time in the example implementation.

[0059] The precision of the cut-off frequency may be increased by adding a following component to the computation. For that purpose, the calculator 215 of cut-off frequency comprises an extrapolator 410 (Figure 4) of the 8th harmonic computed from the minimum or lowest pitch lag value of the time-domain excitation contribution of all sub-frames, using the following relation:

$$h_{8^{th}} = \frac{8 \cdot F_s}{\min_{0 \le i < N_{sub}}\left(T(i)\right)}$$

where F_s = 12800 *Hz*, N_{sub} is the number of sub-frames and *T(i)* is the adaptive codebook index or pitch lag for sub-frame *i*.

[0060] The calculator 215 of cut-off frequency also comprises a finder 409 (Figure 4) of the frequency band in which the 8th harmonic h_{8th} is located. More specifically, for all $i<N_b$, the finder 409 searches for the highest frequency band for which the following inequality is still verified:

$$\left(h_{8^{th}} \ge L_f(i)\right) \quad h_{8th} \ge L_f(i)$$

The index of that band will be called i_{8th} and it indicates the band where the 8th harmonic is likely located.

[0061] The calculator 215 of cut-off frequency finally comprises a selector 411 (Figure 4) of the final cut-off frequency f_{tc}. More specifically, the selector 411 retains the higher frequency between the first estimate f_{tc1} of the cut-off frequency from finder 408 and the last frequency of the frequency band in which the 8th harmonic is located ($L_f(i_{8th})$), using the following relation:

$$f_{tc} = \max\left(L_f\left(i_{8th}\right), f_{tc1}\right)$$

[0062] As illustrated in Figures 3 and 4,

- the calculator 215 of cut-off frequency further comprises a decider 307 (Figure 3) on the number of frequency bins to be zeroed, itself including an analyser 415 (Figure 4) of parameters, and a selector 416 (Figure 4) of frequency bins to be zeroed; and
- the filter 216 (Figure 2), operating in frequency domain, comprises a zeroer 308 (Figure 3) of the frequency bins decided to be zeroed. The zeroer can zero out all the frequency bins (zeroer 417 in Figure 4), or (filter 418 in Figure 4) just some of the higher-frequency bins situated above the cut-off frequency f_{tc} supplemented with a smooth transition region. The transition region is situated above the cut-off frequency f_{tc} and below the zeroed bins, and it allows for a smooth spectral transition between the unchanged spectrum below f_{tc} and the zeroed bins in higher

frequencies.

[0063] For the illustrative example, when the cut-off frequency f_{tc} from the selector 411 is below or equal to 775 Hz, the analyzer 415 considers that the cost of the time-domain excitation contribution is too high. The selector 416 selects all frequency bins of the frequency representation of the time-domain excitation contribution to be zeroed and the zeroer 417 forces to zero all the frequency bins and also force the cut-off frequency f_{tc} to zero. All bits allocated to the time-domain excitation contribution are then reallocated to the frequency-domain coding mode. Otherwise, the analyzer 415 forces the selector 416 to choose the high frequency bins above the cut-off frequency f_{tc} for being zeroed by the zeroer 418.

[0064] Finally, the calculator 215 of cut-off frequency comprises a quantizer 309 (Figures 3 and 4) of the cut-off frequency f_{tc} into a quantized version f_{tcQ} of this cut-off frequency. If three (3) bits are associated to the cut-off frequency parameter, a possible set of output values can be defined (in Hz) as follows:

$$f_{tcQ} = \{0, 1175, 1575, 1975, 2375, 2775, 3175, 3575\}$$

[0065] Many mechanisms could be used to stabilize the choice of the final cut-off frequency f_{tc} to prevent the quantized version f_{tcQ} to switch between 0 and 1175 in inappropriate signal segment. To achieve this, the analyzer 415 in this example implementation is responsive to the long-term average pitch gain G_{lt} 412 from the closed loop pitch analyzer 211 (Figure 2), the open-loop correlation C_{ol} 413 from the open-loop pitch analyzer 203 and the smoothed open-loop correlation C_{st}. To prevent switching to a complete frequency coding, when the following conditions are met, the analyzer 415 does not allow the frequency-only coding, i.e. f_{tcQ} cannot be set to 0:

$$f_{tc} > 2375Hz$$

or

$$f_{tc} > 1175\,Hz \text{ and } C_{ol} > 0.7 \text{ and } G_{lt} \geq 0.6$$

or

$$f_{tc} \geq 1175\,Hz \text{ and } C_{st} > 0.8 \text{ and } G_{lt} \geq 0.4$$

or

$$f_{tcQ}(t-1) != 0 \text{ and } C_{ol} > 0.5 \text{ and } C_{st} > 0.5 \text{ and } G_{lt} \geq 0.6$$

where C_{ot} is the open-loop pitch correlation 413 and C_{st} corresponds to the smoothed version of the open-loop pitch correlation 414 defined as $C_{st} = 0.9 \cdot C_{ol} + 0.1 \cdot C_{st}$. Further, G_{lt} (item 412 of Figure 4) corresponds to the long term average of the pitch gain obtained by the closed loop-pitch analyzer 211 within the time-domain excitation contribution. The long term average of the pitch gain 412 is defined as $G_{lt} = 0.9 \cdot \overline{G}_{\overline{\mu}} - 0.1 \cdot G_{lt}$ and $\overline{G}_{\overline{\mu}}$ is the average pitch gain over the current frame. To further reduce the rate of switching between frequency-only coding and mixed time-domain/frequency-domain coding, a hangover can be added.

6) Frequency domain encoding

Creating a difference vector

[0066] Once the cut-off frequency of the time-domain excitation contribution is defined, the frequency-domain coding is performed. The CELP encoder 100 comprises a subtractor or calculator 109 (Figures 1, 2, 5 and 6) to form a first portion of a difference vector f_d with the difference between the frequency transform f_{res} 502 (Figures 5 and 6) (or other frequency representation) of the input LP residual from DCT 213 (Figure 2) and the frequency transform f_{exc} 501 (Figure 5 and 6) (or other frequency representation) of the time-domain excitation contribution from DCT 214 (Figure 2) from zero up to the cut-off frequency f_{tc} of the time-domain excitation contribution. A downscale factor 603 (Figure 6) is applied to the frequency transform f_{exc} 501 for the next transition region of f_{trans}=2 kHz (80 frequency bins in this example implementation) before its subtraction of the respective spectral portion of the frequency transform f_{res}. The result of the subtraction constitutes the second portion of the difference vector f_d representing the frequency range from the cut-off frequency f_{tc} up to $f_{tc}+f_{trans}$. The frequency transform f_{res} 502 of the input LP residual is used for the remaining third portion of the vector f_d. The downscaled part of the vector f_d resulting from application of the downscale factor 603 can be performed with any type of fade out function, it can be shortened to only few frequency bins, but it could also be omitted when the available bit budget is judged sufficient to prevent energy oscillation artifacts when the cut-off frequency f_{tc} is changing. For example, with a 25 Hz resolution, corresponding to 1 frequency bin f_{bin} = 25 Hz in 256 points DCT at 12.8 kHz, the difference vector can be built as:

$$f_d(k) = f_{res}(k) - f_{exc}(k)$$

where $0 \le k \le f_{tc} / f_{bin}$

$$f_d(k) = f_{res}(k) - f_{exc}(k) \cdot \left(1 - \sin\left(\frac{\pi}{2} \cdot \frac{f_{bin}}{f_{trans}} \cdot (k - \frac{f_{tc}}{f_{bin}})\right)\right)$$

where $f_{tc} / f_{bin} < k \le (f_{tc} + f_{trans}) / f_{bin}$

$f_d(k) = f_{res}(k)$*, otherwise*

where $f_{res}f_{res}$, f_{exc} and f_{ts} have been defined in previous sections 4 and 5.

Searching for frequency pulses

[0067] The CELP encoder 100 comprises a frequency quantizer 110 (Figures 1 and 2) of the difference vector f_d. The difference vector f_d can be quantized using several methods. In all cases, frequency pulses have to be searched for and quantized. In one possible simple method, the frequency-domain coding comprises a search of the most energetic pulses of the difference vector f_d across the spectrum. The method to search the pulses can be as simple as splitting the spectrum into frequency bands and allowing a certain number of pulses per frequency bands. The number of pulses per frequency bands depends on the bit budget available and on the position of the frequency band inside the spectrum. Typically, more pulses are allocated to the low frequencies.

Quantized difference vector

[0068] Depending on the bitrate available, the quantization of the frequency pulses can be performed using different techniques. In one embodiment, at bitrate below 12 kbps, a simple search and quantization scheme can be used to code the position and sign of the pulses. This scheme is described herein below.

[0069] For example for frequencies lower than 3175 Hz, this simple search and quantization scheme uses an approach based on factorial pulse coding (FPC) which is described in the literature, for example in the reference [Mittal, U., Ashley, J.P., and Cruz-Zeno, E.M. (2007), "Low Complexity Factorial Pulse Coding of MDCT Coefficients using Approximation of Combinatorial Functions", IEEE Proceedings on Acoustic, Speech and Signals Processing, Vol. 1, April, pp. 289-292], the full content thereof being incorporated herein by reference.

[0070] More specifically, a selector 504 (Figures 5 and 6) determines that all the spectrum is not quantized using FPC. As illustrated in Figure 5, FPC encoding and pulse position and sign coding is performed in a coder 506. As illustrated in Figure 6, the coder 506 comprises a searcher 609 of frequency pulses. The search is conducted through all the frequency bands for the frequencies lower than 3175 Hz. An FPC coder 610 then processes the frequency pulses. The coder 506 also comprises a finder 611 of the most energetic pulses for frequencies equal to and larger than 3175 Hz, and a quantizer 612 of the position and sign of the found, most energetic pulses. If more than one (1) pulse is allowed within a frequency band then the amplitude of the pulse previously found is divided by 2 and the search is again conducted over the entire frequency band. Each time a pulse is found, its position and sign are stored for quantization and the bit packing stage. The following pseudo code illustrates this simple search and quantization scheme:

$$
\begin{aligned}
&\mathit{for}\ k = 0 : N_{BD}\\
&\quad \mathit{for}\ \ i = 0 : N_p\\
&\qquad p_{max} = 0\\
&\qquad \mathit{for}\ j = C_{Bb}(k) : C_{Bb}(k) + B_b(k)\\
&\qquad\quad \mathit{if}\ \ f_d(j)^2 > p_{max}\\
&\qquad\qquad p_{max} = f_d(j)^2\\
&\qquad\qquad f_d(j) = \frac{f_d(j)}{2}\\
&\qquad\qquad p_p(i) = j\\
&\qquad\qquad p_s(i) = sign\left(f_d(j)\right)\\
&\qquad\quad \mathit{end}\\
&\qquad \mathit{end}\\
&\quad \mathit{end}\\
&\mathit{end}
\end{aligned}
$$

Where N_{BD} is the number of frequency bands (N_{BD} = 16 in the illustrative example), N_p is the number of pulses to be coded in a frequency band k, B_b is the number of frequency bins per frequency band B_b, C_{Bb} is the cumulative frequency bins per band as defined previously in section 5, $p_p^{p_p}$ represents the vector containing the pulse position found, $p_s p_s$ represents the vector containing the sign of the pulse found and $p_{\max}$ □ p_{max} represents the energy of the pulse found.

[0071] At bitrate above 12 kbps, the selector 504 determines that all the spectrum is to be quantized using FPC. As illustrated in Figure 5, FPC encoding is performed in a coder 505. As illustrated in Figure 6, the coder 505 comprises a searcher 607 of frequency pulses. The search is conducted through the entire frequency bands. A FPC processor 610 then FPC codes the found frequency pulses.

[0072] Then, the quantized difference vector f_{dQ} is obtained by adding the number of pulses *nb_pulses* with the pulse sign p_s to each of the position p_p found. For each band the quantized difference vector f_{dQ} can be written with the following pseudo code:

$$\mathit{for}\ j = 0, \ldots, j < nb\_pulses$$

$$f_{dQ}\left(p_p(j)\right) += p_s(j)$$

Noise filling

[0073] All frequency bands are quantized with more or less precision; the quantization method described in the previous section does not guarantee that all frequency bins within the frequency bands are quantized. This is especially the case at low bitrates where the number of pulses quantized per frequency band is relatively low. To prevent the apparition of audible artifacts due to these unquantized bins, a noise filler 507 (Figure 5) adds some noise to fill these gaps. This noise addition is performed over all the spectrum at bitrate below 12 kbps for example, but can be applied only above the cut-off frequency f_{tc} of the time-domain excitation contribution for higher bitrates. For simplicity, the noise intensity varies only with the bitrate available. At high bit rates the noise level is low but the noise level is higher at low bit rates.

[0074] The noise filler 504 comprises an adder 613 (Figure 6) which adds noise to the quantized difference vector f_{dQ} after the intensity or energy level of such added noise has been determined in an estimator 614 and prior to the per band gain has been determined in a computer 615. In the illustrative embodiment, the noise level is directly related to the encoded bitrate. For example at 6.60 kbps the noise level N'_L is 0.4 times the amplitude of the spectral pulses coded in a specific band and as it goes progressively down to a value of 0.2 times the amplitude of the spectral pulses coded in a band at 24 kbps. The noise is added only to section(s) of the spectrum where a certain number of consecutives frequency bins has a very low energy, for example when the number of consecutives very low energy bins N_z is half the number of bins included in the frequency band. For a specific band i, the noise is injected as:

$$for\ j = C_{Bb}(i),\ ...,\ j < C_{Bb}(i) + B_b(i)$$

$$if \sum_{k=j}^{j+N_z} f_{dQ}(k)^2 < 0.5$$

$$for\ k = j,\ ...,\ k < j + N_z$$

$$f_{dQ}(k) = f_{dQ}(k) + N'_L(i) \cdot r_{and}(\ )$$

$$j += N_z$$

Where $N_z = \frac{B_b(i)}{2}$ where, for a band *i*, C_{Bb} is the cumulative number of bins per bands, B_b is the number of bins in a specific band *i*, N'_L is the noise level, and r_{and} is a random number generator which is limited between -1 to 1.

7) Per band gain quantization

[0075] The frequency quantizer 110 comprises a per band gain calculator/quantizer 508 (Figure 5) including a calculator 615 (Figure 6) of per band gain and a quantizer 616 (Figure 6) of the calculated per band gain. Once the quantized difference vector f_{dQ}, including the noise fill if needed, is found, the calculator 615 computes the gain per band for each frequency band. The per band gain for a specific band $G_b(i)$ is defined as the ratio between the energy of the unquantized difference vector f_d signal to the energy of the quantized difference vector f_{dQ} in the log domain as:

$$G_b(i) = \log_{10}\left(\frac{S'_{f_d}(i)}{S'_{f_{dQ}}(i)}\right)$$

Where $S'_{f_d}(i) = \sum_{j=C_{Bb}(i)}^{j=C_{Bb}(i)+B_b(i)} f_d(j)^2$ *and* $S'_{f_{dQ}}(i) = \sum_{j=C_{Bb}(i)}^{j=C_{Bb}(i)+B_b(i)} f_{dQ}(j)^2$ where C_{Bb} and B_b are defined hereinabove in section 5.

[0076] In the embodiment of Figures 5 and 6, the per band gain quantizer 616 vector quantizes the per band frequency gains. Prior to the vector quantization, at low bit rate, the last gain (corresponding to the last frequency band) is quantized separately, and all the remaining fifteen (15) gains are divided by the quantized last gain. Then, the normalized fifteen (15) remaining gains are vector quantized. At higher rate, the mean of the per band gains is quantized first and then removed from all per band gains of the, for example, sixteen (16) frequency bands prior the vector quantization of those per band gains. The vector quantization being used can be a standard minimization in the log domain of the distance between the vector containing the gains per band and the entries of a specific codebook.

[0077] In the frequency-domain coding mode, gains are computed in the calculator 615 for each frequency band to match the energy of the unquantized vector f_d to the quantized vector f_{dQ}. The gains are vector quantized in quantizer 616 and applied per band to the quantized vector f_{dQ} through a multiplier 509 (Figures 5 and 6).

[0078] Alternatively, it is also possible to use the FPC coding scheme at rate below 12 kbps for the whole spectrum by selecting only some of the frequency bands to be quantized. Before performing the selection of the frequency bands, the energy E_d of the frequency bands of the unquantized difference vector f_d, are quantized. The energy is computed as :

$$E_d(i) = \log_{10}(S_d(i))$$

$$S_d(i) = \sum_{j=C_{Bb}(i)}^{j=C_{Bb}(i)+B_b(i)} f_d(j)^2$$

where where C_{Bb} and B_b are defined hereinabove in section 5.

[0079] To perform the quantization of the frequency-band energy E'_d, first the average energy over the first 12 bands out of the sixteen bands used is quantized and subtracted from all the sixteen (16) band energies. Then all the frequency bands are vectors quantized per group of 3 or 4 bands. The vector quantization being used can be a standard minimization in the log domain of the distance between the vector containing the gains per band and the entries of a specific codebook. If not enough bits are available, it is possible to only quantize the first 12 bands and to extrapolate the last 4 bands using the average of the previous 3 bands or by any other methods.

[0080] Once the energy of frequency bands of the unquantized difference vector are quantized, it becomes possible to sort the energy in decreasing order in such a way that it would be replicable on the decoder side. During the sorting, all the energy bands below 2 kHz are always kept and then only the most energetic bands will be passed to the FPC for coding pulse amplitudes and signs. With this approach the FPC scheme codes a smaller vector but covering a wider frequency range. In others words, it takes less bits to cover important energy events over the entire spectrum.

[0081] After the pulse quantization process, a noise fill similar to what has been described earlier is needed. Then, a gain adjustment factor G_a is computed per frequency band to match the energy E_{dQ} of the quantized difference vector f_{dQ} to the quantized energy E'_d of the unquantized difference vector f_d. Then this per band gain adjustment factor is applied to the quantized difference vector f_{dQ}.

$$G_a(i) = 10^{E'_d(i) - E_{dQ}(i)}$$

where

$$E_{dQ}(i) = \log_{10}\left(\sum_{j=C_{Bb}(i)}^{j=C_{Bb}(i)+B_b(i)} f_{dQ}(j)^2 \right)$$

and E'_d is the quantized energy per band of the unquantized difference vector f_d as defined earlier

[0082] After the completion of the frequency-domain coding stage, the total time-domain / frequency domain excitation is found by summing through an adder 111 (Figures 1, 2, 5 and 6) the frequency quantized difference vector f_{dQ} to the filtered frequency-transformed time-domain excitation contribution f_{excF}. When the enhanced CELP encoder 100 changes its bit allocation from a time-domain only coding mode to a mixed time-domain / frequency-domain coding mode, the excitation spectrum energy per frequency band of the time-domain only coding mode does not match the excitation spectrum energy per frequency band of the mixed time-domain / frequency domain coding mode. This energy mismatch can create switching artifacts that are more audible at low bit rate. To reduce any audible degradation created by this bit reallocation, a long-term gain can be computed for each band and can be applied to the summed excitation to correct the energy of each frequency band for a few frames after the reallocation. Then, the sum of the frequency quantized difference vector f_{dQ} and the frequency-transformed and filtered time-domain excitation contribution f_{excF} is then transformed back to time-domain in a converter 112 (Figures 1, 5 and 6) comprising for example an IDCT (Inverse DCT) 220.

[0083] Finally, the synthesized signal is computed by filtering the total excitation signal from the IDCT 220 through a LP synthesis filter 113 (Figures 1 and 2).

[0084] The sum of the frequency quantized difference vector f_{dQ} and the frequency-transformed and filtered time-domain excitation contribution f_{excF} forms the mixed time-domain / frequency-domain excitation transmitted to a distant decoder (not shown). The distant decoder will also comprise the converter 112 to transform the mixed time-domain / frequency-domain excitation back to time-domain using for example the IDCT (Inverse DCT) 220. Finally, the synthesized signal is computed in the decoder by filtering the total excitation signal from the IDCT 220, i.e. the mixed time-domain / frequency-domain excitation through the LP synthesis filter 113 (Figures 1 and 2).

[0085] In one embodiment, while the CELP coding memories are updated on a sub-frame basis using only the time-domain excitation contribution, the total excitation is used to update those memories at frame boundaries. In another possible implementation, the CELP coding memories are updated on a sub-frame basis and also at the frame boundaries using only the time-domain excitation contribution. This results in an embedded structure where the frequency-domain quantized signal constitutes an upper quantization layer independent of the core CELP layer. This presents advantages in certain applications. In this particular case, the fixed codebook is always used to maintain good perceptual quality, and the number of sub-frames is always four (4) for the same reason. However, the frequency-domain analysis can apply to the whole frame. This embedded approach works for bit rates around 12 kbps and higher.

[0086] The foregoing disclosure relates to non-restrictive, illustrative embodiments, and these embodiments can be modified at will, within the scope of the appended claims.

Claims

1. A mixed time-domain / frequency-domain coding device for coding an input sound signal (101), **characterized in that** it comprises:

    a calculator (105) of a time-domain excitation contribution in response to the input sound signal (101);
    a calculator (215) of a cut-off frequency for the time-domain excitation contribution in response to the input sound signal (101);
    a filter (216) responsive to the cut-off frequency for adjusting a frequency extent of the time-domain excitation contribution;
    a calculator (107) of a frequency-domain excitation contribution in response to the input sound signal (101); and
    an adder (111) of the filtered time-domain excitation contribution and the frequency-domain excitation contribution in the frequency domain to form a mixed time-domain / frequency-domain excitation constituting a coded version of the input sound signal (101).

2. A mixed time-domain / frequency-domain coding device according to claim 1, **characterized in that** the time-domain excitation contribution includes (a) only an adaptive codebook contribution, or (b) the adaptive codebook contribution and a fixed codebook contribution.

3. A mixed time-domain / frequency-domain coding device according to claim 1 or 2, **characterized in that** it comprises a calculator (210) of a number of sub-frames to be used in a current frame, the calculator (210) of the number of sub-frames in the current frame is responsive to at least one of an available bit budget and a high frequency spectral dynamic of the input sound signal (101), and the calculator (105) of time-domain excitation contribution uses in the current frame the number of sub-frames determined by the sub-frame number calculator (210) for said current frame.

4. A mixed time-domain / frequency-domain coding device according to any one of claims 1 to 3, **characterized in that** the calculator (107) of frequency-domain excitation contribution performs a frequency transform (213) of a LP residual obtained from an LP analysis (201) of the input sound signal (101) to produce a frequency representation of the LP residual.

5. A mixed time-domain / frequency-domain coding device according to claim 4, **characterized in that** the calculator (215) of cut-off frequency comprises a computer (303) of cross-correlation, for each of a plurality of frequency bands, between the frequency representation of the LP residual and a frequency representation of the time-domain excitation contribution, and the coding device comprises a finder (408) of an estimate of the cut-off frequency in response to the cross-correlation.

6. A mixed time-domain / frequency-domain coding device according to claim 4 or 5, **characterized in that** it comprises a smoother (304) of the cross-correlation through the frequency bands to produce a cross-correlation vector, a calculator (305) of an average of the cross-correlation vector over the frequency bands, and a normalizer (407) of the average of the cross-correlation vector, and the finder (408) of the estimate of the cut-off frequency determines a first estimate of the cut-off frequency by finding a last frequency of one of the frequency bands which minimizes a difference between said last frequency and the normalized average of the cross-correlation vector multiplied by a spectrum width value.

7. A mixed time-domain / frequency-domain coding device according to claim 6, **characterized in that** the calculator (215) of cut-off frequency comprises a finder (409) of one of the frequency bands in which a harmonic computed from the time-domain excitation contribution is located, and a selector (411) of the cut-off frequency as the higher frequency between said first estimate of the cut off-frequency and a last frequency of the frequency band in which said harmonic is located.

8. A mixed time-domain / frequency-domain coding device according to any one of claims 1 to 7, **characterized in that** the filter (216) comprises a zeroer (418) of frequency bins which forces the frequency bins of a plurality of frequency bands above the cut-off frequency to zero.

9. A mixed time-domain / frequency-domain coding device according to any one of claims 1 to 8, **characterized in that** the filter (216) comprises a zeroer (417) of frequency bins which forces all the frequency bins of a plurality of frequency bands to zero when the cut-off frequency is lower than a given value.

10. A mixed time-domain / frequency-domain coding device according to any one of claims 1 to 9, **characterized in that** the calculator (107) of frequency-domain excitation contribution comprises a calculator (109) of a difference between a frequency representation an LP residual of the input sound signal (101) and a filtered frequency representation of the time-domain excitation contribution.

11. A mixed time-domain / frequency-domain coding device according to claim 4, **characterized in that** the calculator (107) of frequency-domain excitation contribution comprises a calculator (109) of a difference between the frequency representation of the LP residual and a frequency representation of the time-domain excitation contribution up to the cut-off frequency to form a first portion of a difference vector, a downscale factor (603) is applied to the frequency representation of the time-domain excitation contribution in a determined frequency range following the cut-off frequency to form a second portion of the difference vector, and the difference vector is formed by the frequency representation (604) of the LP residual for a third remaining portion above the determined frequency range.

12. A mixed time-domain / frequency-domain coding device according to claim 11, **characterized in that** it comprises a quantizer (110) of the difference vector, and the adder (111) adds, in the frequency domain, the quantized difference vector and a frequency-transformed version of the filtered, time-domain excitation contribution to form the mixed time-domain / frequency-domain excitation.

13. A mixed, time-domain / frequency-domain coding device according to any one of claims 1 to 12, **characterized in that** it comprises means for dynamically allocating a bit budget between the time-domain excitation contribution and the frequency-domain excitation contribution.

14. An encoder (100) using a time-domain and frequency-domain model, **characterized in that** it comprises:

    a classifier (204) of an input sound signal (101) as speech or non-speech;
    a time-domain only coder (104);
    the mixed time-domain / frequency-domain coding device of any one of claims 1 to 13; and
    a selector (103) of one of the time-domain only coder and the mixed time-domain / frequency-domain coding device for coding the input sound signal (101) depending on the classification of the input sound signal.

15. An encoder as defined in claim 14, **characterized in that** it comprises a selector (206) of a memory-less time-domain coding mode which, when the classifier (204) classifies the input sound signal (101) as non-speech and detects a temporal attack in the input sound signal (101), forces the memory-less time-domain coding mode for coding the input sound signal (101) in the time-domain only coder (207).

16. A decoder for decoding a sound signal coded using the mixed time-domain / frequency-domain coding device of any one of claims 1 to 13, **characterized in that** it comprises:

    a converter of the mixed time-domain / frequency-domain excitation of any of claims 1 to 13 in time-domain; and
    a synthesis filter for synthesizing the sound signal in response to the mixed time-domain / frequency-domain excitation converted in time-domain.

17. A mixed time-domain / frequency-domain coding method for coding an input sound signal (101), **characterized in that** it comprises:

    calculating (105) a time-domain excitation contribution in response to the input sound signal (101);
    calculating (215) a cut-off frequency for the time-domain excitation contribution in response to the input sound signal (101);
    in response to the cut-off frequency, adjusting (216) a frequency extent of the time-domain excitation contribution;
    calculating (107) a frequency-domain excitation contribution in response to the input sound signal (101); and
    adding (111) the adjusted time-domain excitation contribution and the frequency-domain excitation contribution in the frequency domain to form a mixed time-domain / frequency-domain excitation constituting a coded version of the input sound signal (101).

18. A mixed time-domain / frequency-domain coding method according to claim 17, **characterized in that** the time-domain excitation contribution includes (a) only an adaptive codebook contribution, or (b) the adaptive codebook contribution and a fixed codebook contribution.

19. A mixed time-domain / frequency-domain coding method according to claim 17 or 18, **characterized in that** it comprises calculating (210) a number of sub-frames to be used in a current frame in response to at least one of an available bit budget and a high frequency spectral dynamic of the input sound signal (101), and calculating (105) the time-domain excitation contribution comprises using in the current frame the number of sub-frames determined for said current frame.

20. A mixed time-domain / frequency-domain coding method according to any one of claims 17 to 19, **characterized in that** calculating (107) the frequency-domain excitation contribution comprises performing a frequency transform (213) of a LP residual obtained from an LP analysis of the input sound signal (101) to produce a frequency representation of the LP residual.

21. A mixed time-domain / frequency-domain coding method according to claim 20, **characterized in that** calculating (215) the cut-off frequency comprises computing (303) a cross-correlation, for each of a plurality of frequency bands, between the frequency representation of the LP residual and a frequency representation of the time-domain excitation contribution, and the coding method comprises finding (408) an estimate of the cut-off frequency in response to the cross-correlation.

22. A mixed time-domain / frequency-domain coding method according to claim 21, **characterized in that** it comprises smoothing (304) the cross-correlation through the frequency bands to produce a cross-correlation vector, calculating (305) an average of the cross-correlation vector over the frequency bands, and normalizing (407) the average of the cross-correlation vector, and finding (408) the estimate of the cut-off frequency comprises determining a first estimate of the cut-off frequency by finding a last frequency of one of the frequency bands which minimizes a difference between said last frequency and the normalized average of the cross-correlation vector multiplied by a spectrum width value.

23. A mixed time-domain / frequency-domain coding method according to claim 22, **characterized in that** calculating (215) the cut-off frequency comprises finding (409) one of the frequency bands in which a harmonic computed from the time-domain excitation contribution is located, and selecting (411) the cut-off frequency as the higher frequency between said first estimate of the cut off-frequency and a last frequency of the frequency band in which said harmonic is located.

24. A mixed time-domain / frequency-domain coding method according to any one of claims 17 to 23, **characterized in that** adjusting (216) the frequency extent of the time-domain excitation contribution comprises zeroing (418) frequency bins to force the frequency bins of a plurality of frequency bands above the cut-off frequency to zero.

25. A mixed time-domain / frequency-domain coding method according to any one of claims 17 to 24, **characterized in that** adjusting (216) the frequency extent of the time-domain excitation contribution comprises zeroing (417) frequency bins to force all the frequency bins of a plurality of frequency bands to zero when the cut-off frequency is lower than a given value.

26. A mixed time-domain / frequency-domain coding method according to any one of claims 17 to 25, **characterized in that** calculating (107) the frequency-domain excitation contribution comprises calculating (109) a difference between a frequency representation an LP residual of the input sound signal (101) and a filtered frequency representation of the time-domain excitation contribution.

27. A mixed time-domain / frequency-domain coding method according to any one of claims 17 to 25, **characterized in that** calculating (107) the frequency-domain excitation contribution comprises calculating (109) a difference between the frequency representation of the LP residual and a frequency representation of the time-domain excitation contribution up to the cut-off frequency to form a first portion of a difference vector, a downscale factor (603) is applied to the frequency representation of the time-domain excitation contribution in a determined frequency range following the cut-off frequency to form a second portion of the difference vector, and the difference vector is formed with the frequency representation (604) of the LP residual for a third remaining portion above the determined frequency range.

28. A mixed time-domain / frequency-domain coding method according to claim 27, **characterized in that** it comprises quantizing (110) the difference vector, and adding (111) the adjusted time-domain excitation contribution and the frequency-domain excitation contribution to form the mixed time-domain / frequency-domain excitation comprises adding, in the frequency domain, the quantized difference vector and a frequency-transformed version of the adjusted, time-domain excitation contribution.

29. A mixed, time-domain / frequency-domain coding method according to any one of claims 17 to 28, **characterized in that** it comprises dynamically allocating a bit budget between the time-domain excitation contribution and the frequency-domain excitation contribution.

30. A method (100) of encoding using a time-domain and frequency-domain model, **characterized in that** it comprises:

classifying (204) an input sound signal as speech or non-speech;
providing a time-domain only coding method (104);
providing the mixed time-domain / frequency-domain coding method of any one of claims 17 to 29; and
selecting (103) one of the time-domain only coding method and the mixed time-domain / frequency-domain coding method for coding the input sound signal (101) depending on the classification of the input sound signal (101).

31. A method of encoding as defined in claim 30, **characterized in that** it comprises selecting (206) a memory-less time-domain coding mode which, when the input sound signal (101) is classified (204) as non-speech and a temporal attack in the input sound signal (101) is detected (208), forces the memory-less time-domain coding mode for coding the input sound signal (101) using the time-domain only coding method (207).

32. A method of decoding a sound signal coded using the mixed time-domain / frequency-domain coding method of any one of claims 17 to 31, **characterized in that** it comprises:

converting the mixed time-domain / frequency-domain excitation of any of claims 17 to 31 in time-domain; and
synthesizing the sound signal through a synthesis filter in response to the mixed time-domain / frequency-domain excitation converted in time-domain.

Patentansprüche

1. Gemischte Zeitdomänen-/Frequenzdomänen-Codiervorrichtung zum Codieren eines Eingabeschallsignals (101), **dadurch gekennzeichnet, dass** sie umfasst:

einen Rechner (105) eines Zeitdomänen-Anregungsbeitrag als Reaktion auf das Eingabeschallsignal (101);
einen Rechner (215) einer Grenzfrequenz für den Zeitdomänen-Anregungsbeitrag als Reaktion auf das Eingabeschallsignal (101);
einen Filter (216), der auf die Grenzfrequenz zum Einstellen einer Frequenzerweiterung des Zeitdomänen-Anregungsbeitrags reagiert;
einen Rechner (107) eines Frequenzdomänen-Anregungsbeitrags als Reaktion auf das Eingabeschallsignal (101); und
einen Addierer (111) des gefilterten Zeitdomänen-Anregungsbeitrags und des Frequenzdomänen-Anregungsbeitrags an der Frequenzdomäne, um eine gemischte Zeitdomänen-/Frequenzdomänen-Anregung zu bilden, die eine Codeversion des Eingabeschallsignals (101) bildet.

2. Gemischte Zeitdomänen-/Frequenzdomänen-Codiervorrichtung nach Anspruch 1, **dadurch gekennzeichnet, dass** der Zeitdomänen-Anregungsbeitrag (a) nur einen adaptiven Codebuchbeitrag oder (b) den adaptiven Codebuchbeitrag und einen festen Codebuchbeitrag aufweist.

3. Gemischte Zeitdomänen-/Frequenzdomänen-Codiervorrichtung nach Anspruch 1 oder 2, **dadurch gekennzeichnet, dass** sie einen Rechner (210) aus einer Anzahl von Subframes umfasst, die in einem aktuellen Frame zu verwenden sind, der Rechner (210) der Anzahl von Subframes in dem aktuellen Frame auf mindestens ein verfügbares Bitbudget und eine Hochfrequenz-Spektraldynamik des Eingabeschallsignals (101) reagiert und der Rechner (105) den Zeitdomänen-Anregungsbeitrag in dem aktuellen Frame die Anzahl der Subframes verwendet, die durch den Subframe-Anzahlrechner (210) für den aktuellen Frame bestimmt wird.

4. Gemischte Zeitdomänen-/Frequenzdomänen-Codiervorrichtung nach einem der Ansprüche 1 bis 3, **dadurch gekennzeichnet, dass** der Rechner (107) des Frequenzdomänen-Anregungsbeitrags eine Frequenztransformation (213) eines LP-Rests durchführt, der von einer LP-Analyse (201) des Eingabeschallsignals (101) zum Erzeugen einer Frequenzdarstellung des LP-Restes erhalten wird.

5. Gemischte Zeitdomänen-/Frequenzdomänen-Codiervorrichtung nach Anspruch 4, **dadurch gekennzeichnet, dass** der Rechner (215) der Grenzfrequenz einen Kreuzkorrelationscomputer (303) für jede einer Vielzahl von Frequenzbändern zwischen der Frequenzdarstellung des LP-Restes und einer Frequenzdarstellung des Zeitdomänen-Anregungsbeitrags umfasst, und die Codiervorrichtung einen Finder (408) einer Schätzung der Grenzfrequenz als Reaktion auf die Kreuzkorrelation umfasst.

6. Gemischte Zeitdomänen-/Frequenzdomänen-Codiervorrichtung nach Anspruch 4 oder 5, **dadurch gekennzeichnet, dass** sie einen Glätter (304) der Kreuzkorrelation durch die Frequenzbänder zum Erzeugen eines Kreuzkorrelationsvektors, einen Rechner (305) eines Durchschnitts des Kreuzkorrelationsvektors über die Frequenzbänder und einen Normalisierer (407) des Mittelwerts des Kreuzkorrelationsvektors umfasst, und der Finder (408) der Schätzung der Grenzfrequenz eine erste Schätzung der Grenzfrequenz durch Finden einer letzten Frequenz eines der Frequenzbänder bestimmt, die eine Differenz zwischen der letzten Frequenz und dem normalisierten Durchschnitt des Kreuzkorrelationsvektors multipliziert mit einem Spektrumsbreitenwert minimiert.

7. Gemischte Zeitdomänen-/Frequenzdomänen-Codiervorrichtung nach Anspruch 6, **dadurch gekennzeichnet, dass** der Rechner (215) der Grenzfrequenz einen Finder (409) eines der Frequenzbänder, in dem eine Oberschwingung, die aus der Zeitdomänen-Anregungsbeitrag berechnet wird, angeordnet ist, und einen Selektor (411) der Grenzfrequenz als die höhere Frequenz zwischen der ersten Schätzung der Grenzfrequenz und einer letzten Frequenz des Frequenzbandes umfasst, in dem die Oberschwingung angeordnet ist.

8. Gemischte Zeitdomänen-/Frequenzdomänen-Codiervorrichtung nach einem der Ansprüche 1 bis 7, **dadurch gekennzeichnet, dass** der Filter (216) einen Nullsteller (418) von Frequenz-Bins umfasst, der die Frequenz-Bins einer Vielzahl von Frequenzbändern über der Grenzfrequenz auf Null drückt.

9. Gemischte Zeitdomänen-/Frequenzdomänen-Codiervorrichtung nach einem der Ansprüche 1 bis 8, **dadurch gekennzeichnet, dass** der Filter (216) einen Nullsteller (417) von Frequenz-Bins umfasst, der alle Frequenz-Bins einer Vielzahl von Frequenzbändern auf Null drückt, wenn die Grenzfrequenz kleiner als ein vorgegebener Wert ist.

10. Gemischte Zeitdomänen-/Frequenzdomänen-Codiervorrichtung nach einem der Ansprüche 1 bis 9, **dadurch gekennzeichnet, dass** der Rechner (107) des Frequenzdomänen-Anregungsbeitrag einen Rechner (109) einer Differenz zwischen einer Frequenzdarstellung eines LP-Restes des Eingabeschallsignals (101) und einer gefilterten Frequenzdarstellung der Zeitdomänen-Anregungsbeitrag umfasst.

11. Gemischte Zeitdomänen-/Frequenzdomänen-Codiervorrichtung nach Anspruch 4, **dadurch gekennzeichnet, dass** der Rechner (107) des Frequenzdomänen-Anregungsbeitrags einen Rechner (109) einer Differenz zwischen der Frequenzdarstellung des LP-Restes und einer Frequenzdarstellung des Zeitdomänen-Anregungsbeitrags bis zur Grenzfrequenz zum Bilden eines ersten Abschnitts eines Differenzvektors umfasst, ein Herunterskalierungsfaktor (603) an die Frequenzdarstellung des Zeitdomänen-Anregungsbeitrags in einem bestimmten Frequenzbereich angelegt wird, der auf die Grenzfrequenz folgt, um einen zweiten Abschnitt des Differenzvektors zu bilden, und der Differenzvektor durch die Frequenzdarstellung (604) des LP-Rests für einen dritten restlichen Teil oberhalb des bestimmten Frequenzbereichs gebildet wird.

12. Gemischte Zeitdomänen-/Frequenzdomänen-Codiervorrichtung nach Anspruch 11, **dadurch gekennzeichnet, dass** sie einen Quantisierer (110) des Differenzvektors umfasst und der Addierer (111) in der Frequenzdomäne den quantisierten Differenzvektor und eine frequenztransformierte Version des gefilterten Zeitdomänen-Anregungsbeitrags addiert, um die gemischte Zeitdomänen-/Frequenzdomänen-Anregung zu bilden.

13. Gemischte Zeitdomänen-/Frequenzdomänen-Codiervorrichtung nach einem der Ansprüche 1 bis 12, **dadurch gekennzeichnet, dass** sie Einrichtungen zum dynamischen Zuweisen eines Bitbudgets zwischen dem Zeitdomänen-Anregungsbeitrag und dem Frequenzdomänen-Anregungsbeitrag umfasst.

14. Codierer (100), der ein Zeitdomänen- und Frequenzdomänenmodell verwendet, **dadurch gekennzeichnet, dass** er umfasst:

ein Klassifizierer (204) eines Eingabeschallsignals (101) als Sprache oder Nichtsprache;
einen Nur-Zeitdomänen-Codierer (104);
die gemischte Zeitdomänen-/Frequenzdomänen-Codiervorrichtung nach einem der Ansprüche 1 bis 13; und
einen Selektor (103) eines von Nur-Zeitdomänen-Codierer und der gemischten Zeitdomänen-/Frequenzdomänen-Codiervorrichtung zum Codieren des Eingabeschallsignals (101) abhängig von der Klassifizierung des Eingabeschallsignals.

15. Codierer nach Anspruch 14, **dadurch gekennzeichnet, dass** er einen Selektor (206) eines speicherlosen Zeitdomänen-Codiermodus umfasst, der, wenn der Klassifizierer (204) das Eingabeschallsignal (101) als Nicht-Sprache klassifiziert und eine zeitliche Attacke in dem Eingabeschallsignal (101) detektiert, den speicherlosen Zeitdomänen-Codiermodus zum Codieren des Eingabeschallsignals (101) in dem Nur-Zeitdomänen-Codierer (207) erzwingt.

16. Decodierer zum Decodieren eines Schallsignals, das unter Verwendung der gemischten Zeitdomänen-/Frequenzdomänen-Codiervorrichtung nach einem der Ansprüche 1 bis 13 codiert ist, **dadurch gekennzeichnet, dass** er umfasst:

einen Wandler der gemischten Zeitdomänen-/Frequenzdomänen-Anregung nach einem der Ansprüche 1 bis 13 in der Zeitdomäne; und
einen Synthesefilter zum Synthetisieren des Schallsignals als Reaktion auf die gemischte Zeitdomänen-/Frequenzdomänen-Anregung.

17. Gemischtes Zeitdomänen-/Frequenzdomänen-Codierverfahren zum Codieren eines Eingabeschallsignals (101), **dadurch gekennzeichnet, dass** es umfasst:

Berechnen (105) eines Zeitdomänen-Anregungsbeitrags als Reaktion auf das Eingabeschallsignal (101);
Berechnen (215) einer Grenzfrequenz für den Zeitdomänen-Anregungsbeitrag als Reaktion auf das Eingabeschallsignal (101);
als Reaktion auf die Grenzfrequenz, Einstellen (216) einer Frequenzerweiterung des Zeitdomänen-Anregungsbeitrags;
Berechnen (107) eines Frequenzdomänen-Anregungsbeitrags als Reaktion auf das Eingabeschallsignal (101); und
Addieren (111) des eingestellten Zeitdomänen-Anregungsbeitrags und des Frequenzdomänen-Anregungsbeitrags an der Frequenzdomäne, um eine gemischte Zeitdomänen-/Frequenzdomänen-Anregung zu bilden, die eine Codeversion des Eingabeschallsignals (101) bildet.

18. Gemischtes Zeitdomänen-/Frequenzdomänen-Codierverfahren nach Anspruch 17, **dadurch gekennzeichnet, dass** der Zeitdomänen-Anregungsbeitrag (a) nur einen adaptiven Codebuchbeitrag oder (b) den adaptiven Codebuchbeitrag und einen festen Codebuchbeitrag aufweist.

19. Gemischte Zeitdomänen-/Frequenzdomänen-Codierverfahren nach Anspruch 17 oder 18, **dadurch gekennzeichnet, dass** es das Berechnen (210) einer Anzahl von Subframes umfasst, die in einem aktuellen Frame zu verwenden sind, als Reaktion auf mindestens ein verfügbares Bitbudget und eine Hochfrequenz-Spektraldynamik des Eingabeschallsignals (101) und das Berechnen (105) des Zeitdomänen-Anregungsbeitrags das Verwenden, in dem aktuellen Frame, der Anzahl von Subframes umfasst, die für den aktuellen Frame bestimmt wird.

20. Gemischtes Zeitdomänen-/Frequenzdomänen-Codierverfahren nach einem der Ansprüche 17 bis 19, **dadurch gekennzeichnet, dass** das Berechnen (107) des Frequenzdomänen-Anregungsbeitrags das Durchführen einer Frequenztransformation (213) eines LP-Rests umfasst, der von einer LP-Analyse des Eingabeschallsignals (101) zum Erzeugen einer Frequenzdarstellung des LP-Rests erhalten wird.

21. Gemischtes Zeitdomänen-/Frequenzdomänen-Codierverfahren nach Anspruch 20, **dadurch gekennzeichnet, dass** das Berechnen (215) der Grenzfrequenz das Berechnen (303) eines Kreuzkorrelationscomputers für jede einer Vielzahl von Frequenzbändern zwischen der Frequenzdarstellung des LP-Restes und einer Frequenzdarstellung des Zeitdomänen-Anregungsbeitrags umfasst, und das Codierverfahren das Finden (408) einer Schätzung der Grenzfrequenz als Reaktion auf die Kreuzkorrelation umfasst.

22. Gemischtes Zeitdomänen-/Frequenzdomänen-Codierverfahren nach Anspruch 21, **dadurch gekennzeichnet, dass** es das Glätten (304) der Kreuzkorrelation durch die Frequenzbänder zum Erzeugen eines Kreuzkorrelationsvektors, das Berechnen (305) eines Durchschnitts des Kreuzkorrelationsvektors über die Frequenzbänder und das Normalisieren (407) des Durchschnitts des Kreuzkorrelationsvektors umfasst, und das Finden (408) der Schätzung der Grenzfrequenz das Bestimmen einer ersten Schätzung der Grenzfrequenz durch Finden einer letzten Frequenz eines der Frequenzbänder umfasst, die eine Differenz zwischen der letzten Frequenz und dem normalisierten Durchschnitt des Kreuzkorrelationsvektors multipliziert mit einem Spektrumsbreitenwert minimiert.

23. Gemischtes Zeitdomänen-/Frequenzdomänen-Codierverfahren nach Anspruch 22, **dadurch gekennzeichnet, dass** das Berechnen (215) der Grenzfrequenz das Finden (409) eines der Frequenzbänder, in dem eine Oberschwingung, die aus dem Zeitdomänen-Anregungsbeitrag berechnet wird, angeordnet ist, umfasst, sowie das Auswählen (411) der Grenzfrequenz als die höhere Frequenz zwischen der ersten Schätzung der Grenzfrequenz und einer letzten Frequenz des Frequenzbandes, in dem die Oberschwingung angeordnet ist.

24. Gemischtes Zeitdomänen-/Frequenzdomänen-Codierverfahren nach einem der Ansprüche 17 bis 23, **dadurch gekennzeichnet, dass** das Einstellen (216) der Frequenzerweiterung des Zeitdomänen-Anregungsbeitrags das Nullstellen (418) von Frequenz-Bins umfasst, um die Frequenz-Bins einer Vielzahl von Frequenzbändern oberhalb der Grenzfrequenz auf Null zu drücken.

25. Gemischtes Zeitdomänen-/Frequenzdomänen-Codierverfahren nach einem der Ansprüche 17 bis 24, **dadurch gekennzeichnet, dass** das Einstellen (216) der Frequenzerweiterung des Zeitdomänen-Anregungsbeitrags das Nullstellen (417) von Frequenz-Bins umfasst, um alle Frequenz-Bins einer Vielzahl von Frequenzbändern auf Null zu drücken, wenn die Frequenz kleiner als ein vorgegebener Wert ist.

26. Gemischtes Zeitdomänen-/Frequenzdomänen-Codierverfahren nach einem der Ansprüche 17 bis 25, **dadurch gekennzeichnet, dass** das Berechnen (107) des Frequenzdomänen-Anregungsbeitrags das Berechnen (109) einer Differenz zwischen einer Frequenzdarstellung eines LP-Restes des Eingabeschallsignals (101) und einer gefilterten Frequenzdarstellung des Zeitdomänen-Anregungsbeitrags umfasst.

27. Gemischtes Zeitdomänen-/Frequenzdomänen-Codierverfahren nach einem der Ansprüche 17 bis 25, **dadurch gekennzeichnet, dass** das Berechnen (107) des Frequenzdomänen-Anregungsbeitrags das Berechnen (109) einer Differenz zwischen der Frequenzdarstellung des LP-Restes und einer Frequenzdarstellung des Zeitdomänen-Anregungsbeitrags bis zur Grenzfrequenz zum Bilden eines ersten Abschnitts eines Differenzvektors umfasst, ein Herunterskalierungsfaktor (603) an die Frequenzdarstellung des Zeitdomänen-Anregungsbeitrags in einem bestimmten Frequenzbereich angelegt wird, der auf die Grenzfrequenz folgt, um einen zweiten Abschnitt des Differenzvektors zu bilden, und der Differenzvektor mit der Frequenzdarstellung (604) des LP-Rests für einen dritten restlichen Teil oberhalb des bestimmten Frequenzbereichs gebildet wird.

28. Gemischtes Zeitdomänen-/Frequenzdomänen-Codierverfahren nach Anspruch 27, **dadurch gekennzeichnet, dass** es das Quantisieren (110) des Differenzvektors umfasst und das Addieren (111) des eingestellten Zeitdomänen-Anregungsbeitrags und des Frequenzdomänen-Anregungsbeitrags, um die gemischte Zeitdomänen-/Frequenzdomänenanregung zu bilden, das Addieren des quantisierten Differenzvektors und einer frequenztransformierten Version des eingestellten Zeitdomänen-Anregungsbeitrags in der Frequenzdomäne umfasst.

29. Gemischtes Zeitdomänen-/Frequenzdomänen-Codierverfahren nach einem der Ansprüche 17 bis 28, **dadurch gekennzeichnet, dass** es das dynamische Zuweisen eines Bitbudgets zwischen dem Zeitdomänen-Anregungsbeitrag und dem Frequenzdomänen-Anregungsbeitrag umfasst.

30. Codierverfahren (100) mithilfe des Zeitdomänen- und Frequenzdomänenmodells, **dadurch gekennzeichnet, dass** es umfasst:

Klassifizieren (204) eines Eingabeschallsignals als Sprache oder Nichtsprache;
Bereitstellen eines Nur-Zeitdomänen-Codierverfahrens (104);
Bereitstellen des gemischten Zeitdomänen-/Frequenzdomänen-Codierverfahrens nach einem der Ansprüche 17 bis 29; und
Auswählen (103) eines von Nur-Zeitdomänen-Codierverfahren und gemischtem Zeitdomänen-/Frequenzdomänen-Codierverfahren zum Codieren des Eingabeschallsignals (101) abhängig von der Klassifizierung des Eingabeschallsignals (101).

31. Codierverfahren nach Anspruch 30, **dadurch gekennzeichnet, dass** es das Auswählen (206) eines speicherlosen

Zeitdomänen-Codiermodus umfasst, der, wenn das Eingabeschallsignal (101) als Nicht-Sprache klassifiziert (204) wird und eine zeitliche Attacke in dem Eingabeschallsignal (101) detektiert (208) wird, den speicherlosen Zeitdomänen-Codiermodus zum Codieren des Eingabeschallsignals (101) unter Verwendung des Nur-Zeitdomänen-Codierverfahrens (207) erzwingt.

32. Decodierverfahren zum Decodieren eines Schallsignals, das unter Verwendung des gemischten Zeitdomänen-/Frequenzdomänen-Codierverfahrens nach einem der Ansprüche 17 bis 31 codiert ist, **dadurch gekennzeichnet, dass** es umfasst:

Umwandeln der gemischten Zeitdomänen-/Frequenzdomänen-Anregung nach einem der Ansprüche 17 bis 31 in der Zeitdomäne; und
Synthetisieren des Schallsignals durch einen Synthesefilter als Reaktion auf die gemischte Zeitdomänen-/Frequenzdomänen-Anregung, die in der Zeitdomäne umgewandelt wurde.

Revendications

1. Dispositif de codage mixte de domaine temporel/domaine fréquentiel pour coder un signal audio d'entrée (101), **caractérisé en ce qu'**il comprend :

un calculateur (105) d'une contribution d'excitation de domaine temporel en réponse au signal audio d'entrée (101) ;
un calculateur (215) d'une fréquence de coupure pour la contribution d'excitation de domaine temporel en réponse au signal audio d'entrée (101) ;
un filtre (216) sensible à la fréquence de coupure pour régler une ampleur de fréquence de la contribution d'excitation de domaine temporel ;
un calculateur (107) d'une contribution d'excitation de domaine fréquentiel en réponse au signal audio d'entrée (101) ; et
un additionneur (111) de la contribution d'excitation de domaine temporel filtrée et de la contribution d'excitation de domaine fréquentiel dans le domaine fréquentiel pour former une excitation mixte de domaine temporel/domaine fréquentiel constituant une version codée du signal audio d'entrée (101).

2. Dispositif de codage mixte de domaine temporel/domaine fréquentiel selon la revendication 1, **caractérisé en ce que** la contribution d'excitation de domaine temporel comprend (a) seulement une contribution de livre de codes adaptatif, ou (b) la contribution de livre de codes adaptatif et une contribution de livre de codes fixe.

3. Dispositif de codage mixte de domaine temporel/domaine fréquentiel selon la revendication 1 ou 2, **caractérisé en ce qu'**il comprend un calculateur (210) d'un nombre de sous-trames à utiliser dans une trame actuelle, le calculateur (210) du nombre de sous-trames dans la trame actuelle est sensible à au moins un parmi un budget de bits disponible et une dynamique spectrale haute fréquence du signal audio d'entrée (101), et le calculateur (105) de contribution d'excitation de domaine temporel utilise dans la trame actuelle le nombre de sous-trames déterminé par le calculateur du nombre de sous-trames (210) pour ladite trame actuelle.

4. Dispositif de codage mixte de domaine temporel/domaine fréquentiel selon l'une quelconque des revendications 1 à 3, **caractérisé en ce que** le calculateur (107) de contribution d'excitation de domaine fréquentiel effectue une transformée fréquentielle (213) d'un résidu LP obtenu à partir d'une analyse LP (201) du signal audio d'entrée (101) pour produire une représentation fréquentielle du résidu LP.

5. Dispositif de codage mixte de domaine temporel/domaine fréquentiel selon la revendication 4, **caractérisé en ce que** le calculateur (215) de fréquence de coupure comprend un calculateur (303) de corrélation croisée, pour chacune d'une pluralité de bandes de fréquences, entre la représentation fréquentielle du résidu LP et une représentation fréquentielle de la contribution d'excitation de domaine temporel, et le dispositif de codage comprend un dispositif de détermination (408) d'une estimation de la fréquence de coupure en réponse à la corrélation croisée.

6. Dispositif de codage mixte de domaine temporel/domaine fréquentiel selon la revendication 4 ou 5, **caractérisé en ce qu'**il comprend un dispositif de lissage (304) de la corrélation croisée à travers les bandes de fréquences pour produire un vecteur de corrélation croisée, un calculateur (305) d'une moyenne du vecteur de corrélation croisée sur les bandes de fréquences, et un dispositif de normalisation (407) de la moyenne du vecteur de corrélation croisée, et le dispositif de détermination (408) de l'estimation de la fréquence de coupure détermine une première estimation de la fréquence de coupure en trouvant une dernière fréquence d'une des bandes de fréquences qui minimise une différence entre ladite dernière fréquence et la moyenne normalisée du vecteur de corrélation croisée multiplié par une valeur de largeur de spectre.

7. Dispositif de codage mixte de domaine temporel/domaine fréquentiel selon la revendication 6, **caractérisé en ce que** le calculateur (215) de fréquence de coupure comprend un chercheur (409) d'une des bandes de fréquences dans laquelle se situe une harmonique calculée à partir de la contribution d'excitation de domaine temporel, et un sélecteur (411) de la fréquence de coupure comme la plus haute fréquence entre ladite première estimation de la fréquence de coupure et une dernière fréquence de la bande de fréquence dans laquelle se situe ladite harmonique.

8. Dispositif de codage mixte de domaine temporel/domaine fréquentiel selon l'une quelconque des revendications 1 à 7, **caractérisé en ce que** le filtre (216) comprend un dispositif de mise à zéro (418) de cases de fréquences qui met à zéro les cases de fréquences d'une pluralité de bandes de fréquences au-dessus de la fréquence de coupure.

9. Dispositif de codage mixte de domaine temporel/domaine fréquentiel selon l'une quelconque des revendications 1 à 8, **caractérisé en ce que** le filtre (216) comprend un dispositif de mise à zéro (417) de cases de fréquences qui met à zéro toutes les cases de fréquences d'une pluralité de bandes de fréquences quand la fréquence de coupure est inférieure à une valeur donnée.

10. Dispositif de codage mixte de domaine temporel/domaine fréquentiel selon l'une quelconque des revendications 1 à 9, **caractérisé en ce que** le calculateur (107) de contribution d'excitation de domaine fréquentiel comprend un calculateur (109) d'une différence entre une représentation fréquentielle du résidu LP du signal audio d'entrée (101) et une représentation fréquentielle filtrée de la contribution d'excitation de domaine temporel.

11. Dispositif de codage mixte de domaine temporel/domaine fréquentiel selon la revendication 4, **caractérisé en ce que** le calculateur (107) de contribution d'excitation de domaine fréquentiel comprend un calculateur (109) d'une différence entre la représentation fréquentielle du résidu LP et une représentation fréquentielle de la contribution d'excitation de domaine temporel jusqu'à la fréquence de coupure pour former une première portion d'un vecteur de différence, un facteur de réduction d'échelle (603) est appliqué à la représentation fréquentielle de la contribution d'excitation de domaine temporel dans une plage de fréquences déterminée à la suite de la fréquence de coupure pour former une deuxième portion du vecteur de différence, et le vecteur de différence est formé par la représentation fréquentielle (604) du résidu LP pour une troisième portion restante au-dessus de la plage de fréquences déterminée.

12. Dispositif de codage mixte de domaine temporel/domaine fréquentiel selon la revendication 11, **caractérisé en ce qu'**il comprend un quantificateur (110) du vecteur de différence, et l'additionneur (111) additionne, dans le domaine fréquentiel, le vecteur de différence quantifié et une version transformée en fréquence de la contribution d'excitation de domaine temporel filtrée pour former l'excitation mixte de domaine temporel/domaine fréquentiel.

13. Dispositif de codage mixte de domaine temporel/domaine fréquentiel selon l'une quelconque des revendications 1 à 12, **caractérisé en ce qu'**il comprend des moyens pour allouer de manière dynamique un budget de bits entre la contribution d'excitation de domaine temporel et la contribution d'excitation de domaine fréquentiel.

14. Codeur (100) utilisant un modèle de domaine temporel et de domaine fréquentiel, **caractérisé en ce qu'**il comprend :

   un classificateur (204) d'un signal audio d'entrée (101) comme vocal ou non vocal ;
   un codeur uniquement de domaine temporel (104) ;
   le dispositif de codage mixte de domaine temporel/domaine fréquentiel selon l'une quelconque des revendications 1 à 13 ; et
   un sélecteur (103) d'un du codeur uniquement de domaine temporel et du dispositif de codage mixte de domaine temporel/domaine fréquentiel pour coder le signal audio d'entrée (101) en fonction de la classification du signal audio d'entrée.

15. Codeur selon la revendication 14, **caractérisé en ce qu'**il comprend un sélecteur (206) d'un mode de codage de domaine temporel sans mémoire qui, quand le classificateur (204) classe le signal audio d'entrée (101) comme non vocal et détecte une attaque temporelle dans le signal audio d'entrée (101), force le mode de codage de domaine temporel sans mémoire pour coder le signal audio d'entrée (101) dans le codeur uniquement de domaine temporel (207).

16. Décodeur pour décoder un signal audio codé en utilisant le dispositif de codage mixte de domaine temporel/domaine fréquentiel selon l'une quelconque des revendications 1 à 13, **caractérisé en ce qu'**il comprend :

un convertisseur de l'excitation mixte de domaine temporel/domaine fréquentiel selon l'une quelconque des revendications 1 à 13 dans le domaine temporel ; et
un filtre de synthèse pour synthétiser le signal audio en réponse à l'excitation mixte de domaine temporel/domaine fréquentiel convertie dans le domaine temporel.

17. Procédé de codage mixte de domaine temporel/domaine fréquentiel pour coder un signal audio d'entrée (101), **caractérisé en ce qu'**il comprend :

le calcul (105) d'une contribution d'excitation de domaine temporel en réponse au signal audio d'entrée (101) ;
le calcul (215) d'une fréquence de coupure pour la contribution d'excitation de domaine temporel en réponse au signal audio d'entrée (101) ;
en réponse à la fréquence de coupure, le réglage (216) d'une ampleur de fréquence de la contribution d'excitation de domaine temporel ;
le calcul (107) d'une contribution d'excitation de domaine fréquentiel en réponse au signal audio d'entrée (101) ; et
l'addition (111) de la contribution d'excitation de domaine temporel réglée et de la contribution d'excitation de domaine fréquentiel dans le domaine fréquentiel pour former une excitation mixte de domaine temporel/domaine fréquentiel constituant une version codée du signal audio d'entrée (101).

18. Procédé de codage mixte de domaine temporel/domaine fréquentiel selon la revendication 17, **caractérisé en ce que** la contribution d'excitation de domaine temporel comprend (a) seulement une contribution de livre de codes adaptatif, ou (b) la contribution de livre de codes adaptatif et une contribution de livre de codes fixe.

19. Procédé de codage mixte de domaine temporel/domaine fréquentiel selon la revendication 17 ou 18, **caractérisé en ce qu'**il comprend le calcul (210) d'un nombre de sous-trames à utiliser dans une trame actuelle en réponse à au moins un parmi un budget de bits disponible et une dynamique spectrale haute fréquence du signal audio d'entrée (101), et le calculateur (105) de contribution d'excitation de domaine temporel comprend l'utilisation dans la trame actuelle du nombre de sous-trames déterminé pour la trame actuelle.

20. Procédé de codage mixte de domaine temporel/domaine fréquentiel selon l'une quelconque des revendications 17 à 19, **caractérisé en ce que** le calcul (107) de la contribution d'excitation de domaine fréquentiel comprend l'exécution d'une transformée fréquentielle (213) d'un résidu LP obtenu à partir d'une analyse LP du signal audio d'entrée (101) pour produire une représentation fréquentielle du résidu LP.

21. Procédé de codage mixte de domaine temporel/domaine fréquentiel selon la revendication 20, **caractérisé en ce que** le calcul (215) de la fréquence de coupure comprend le calcul (303) d'une corrélation croisée, pour chacune d'une pluralité de bandes de fréquences, entre la représentation fréquentielle du résidu LP et une représentation fréquentielle de la contribution d'excitation de domaine temporel, et le procédé de codage comprend la détermination (408) d'une estimation de la fréquence de coupure en réponse à la corrélation croisée.

22. Procédé de codage mixte de domaine temporel/domaine fréquentiel selon la revendication 21, **caractérisé en ce qu'**il comprend le lissage (304) de la corrélation croisée à travers les bandes de fréquences pour produire un vecteur de corrélation croisée, le calcul (305) d'une moyenne du vecteur de corrélation croisée sur les bandes de fréquences, et la normalisation (407) de la moyenne du vecteur de corrélation croisée, et la détermination (408) de l'estimation de la fréquence de coupure comprend la détermination d'une première estimation de la fréquence de coupure en trouvant une dernière fréquence d'une des bandes de fréquences qui minimise une différence entre ladite dernière fréquence et la moyenne normalisée du vecteur de corrélation croisée multiplié par une valeur de largeur de spectre.

23. Procédé de codage mixte de domaine temporel/domaine fréquentiel selon la revendication 22, **caractérisé en ce que** le calcul (215) de la fréquence de coupure comprend la recherche (409) d'une des bandes de fréquences dans laquelle se situe une harmonique calculée à partir de la contribution d'excitation de domaine temporel, et la sélection (411) de la fréquence de coupure comme la plus haute fréquence entre ladite première estimation de la fréquence de coupure et la dernière fréquence de la bande de fréquence dans laquelle se situe ladite harmonique.

24. Procédé de codage mixte de domaine temporel/domaine fréquentiel selon l'une quelconque des revendications 17 à 23, **caractérisé en ce que** le réglage (216) de l'ampleur de fréquence de la contribution d'excitation de domaine temporel comprend la mise à zéro (418) de cases de fréquences qui met à zéro les cases de fréquences d'une pluralité de bandes de fréquences au-dessus de la fréquence de coupure.

25. Procédé de codage mixte de domaine temporel/domaine fréquentiel selon l'une quelconque des revendications 17 à 24, **caractérisé en ce que** le réglage (216) de l'ampleur de fréquence de la contribution d'excitation de domaine temporel comprend la mise à zéro (417) de cases de fréquences qui met à zéro toutes les cases de fréquences d'une pluralité de bandes de fréquences quand la fréquence de coupure est inférieure à une valeur donnée.

26. Procédé de codage mixte de domaine temporel/domaine fréquentiel selon l'une quelconque des revendications 17 à 25, **caractérisé en ce que** le calcul (107) de la contribution d'excitation de domaine fréquentiel comprend le calcul (109) d'une différence entre une représentation fréquentielle d'un résidu LP du signal audio d'entrée (101) et une représentation fréquentielle filtrée de la contribution d'excitation de domaine temporel.

27. Procédé de codage mixte de domaine temporel/domaine fréquentiel selon l'une quelconque des revendications 17 à 25, **caractérisé en ce que** le calcul (107) de la contribution d'excitation de domaine fréquentiel comprend le calcul (109) d'une différence entre la représentation fréquentielle du résidu LP et une représentation fréquentielle de la contribution d'excitation de domaine temporel jusqu'à la fréquence de coupure pour former une première portion d'un vecteur de différence, un facteur de réduction d'échelle (603) est appliqué à la représentation fréquentielle de la contribution d'excitation de domaine temporel dans une plage de fréquences déterminée à la suite de la fréquence de coupure pour former une deuxième portion du vecteur de différence, et le vecteur de différence est formé par la représentation fréquentielle (604) du résidu LP pour une troisième portion restante au-dessus de la plage de fréquences déterminée.

28. Procédé de codage mixte de domaine temporel/domaine fréquentiel selon la revendication 27, **caractérisé en ce qu'**il comprend la quantification (110) du vecteur de différence, et l'addition (111) de la contribution d'excitation de domaine temporel réglée et de la contribution d'excitation de domaine fréquentiel pour former l'excitation mixte de domaine temporel/domaine fréquentiel, dans le domaine fréquentiel, du vecteur de différence quantifié et d'une version transformée en fréquence de la contribution d'excitation de domaine temporel réglée.

29. Procédé de codage mixte de domaine temporel/domaine fréquentiel selon l'une quelconque des revendications 17 à 28, **caractérisé en ce qu'**il comprend l'allocation dynamique d'un budget de bits entre la contribution d'excitation de domaine temporel et la contribution d'excitation de domaine fréquentiel.

30. Procédé (100) de codage utilisant un modèle de domaine temporel et de domaine fréquentiel, **caractérisé en ce qu'**il comprend :

la classification (204) d'un signal audio d'entrée comme vocal ou non vocal ;
la fourniture d'un procédé de codeur uniquement de domaine temporel (104) ;
la fourniture du procédé de codage mixte de domaine temporel/domaine fréquentiel selon l'une quelconque des revendications 17 à 29 ; et
la sélection (103) d'un du codeur uniquement de domaine temporel et du dispositif de codage mixte de domaine temporel/domaine fréquentiel pour coder le signal audio d'entrée (101) en fonction de la classification du signal audio d'entrée (101).

31. Procédé de codage selon la revendication 30, **caractérisé en ce qu'**il comprend la sélection (206) d'un mode de codage de domaine temporel sans mémoire qui, quand le signal audio d'entrée (101) est classé (204) comme non vocal et une attaque temporelle dans le signal audio d'entrée (101) est détectée (208), force le mode de codage de domaine temporel sans mémoire pour coder le signal audio d'entrée (101) en utilisant le procédé de codage uniquement de domaine temporel (207).

32. Procédé de décodage d'un signal audio codé en utilisant le procédé de codage mixte de domaine temporel et de domaine fréquentiel selon l'une quelconque des revendications 17 à 31, **caractérisé en ce qu'**il comprend :

la conversion de l'excitation mixte de domaine temporel/domaine fréquentiel selon l'une quelconque des revendications 17 à 31 dans le domaine temporel ; et
la synthèse du signal audio par le biais d'un filtre de synthèse en réponse à l'excitation mixte de domaine temporel/domaine fréquentiel convertie dans le domaine temporel.

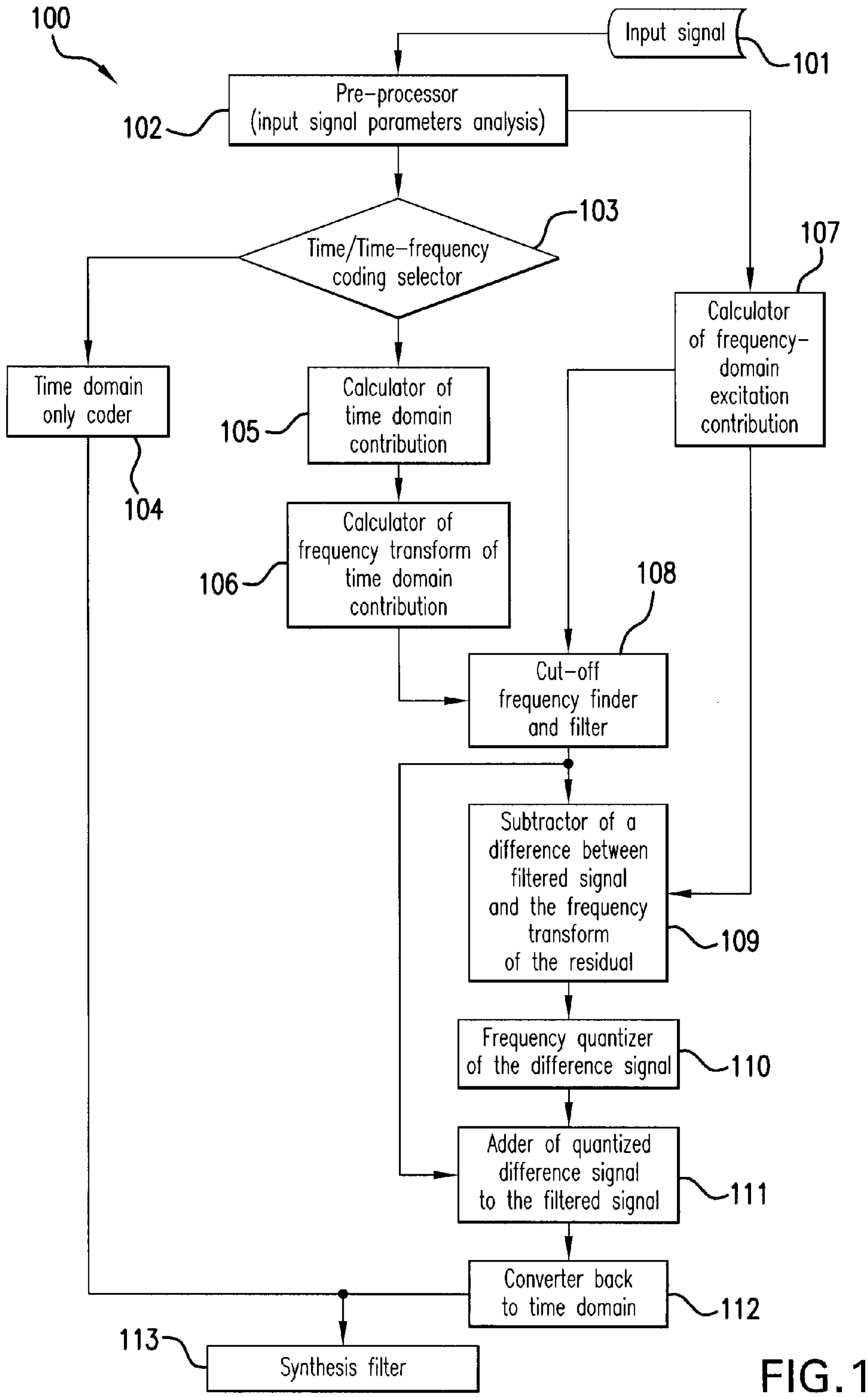
100
Input signal
101
Pre–processor
(input signal parameters analysis)
102
Time/Time–frequency
coding selector
103
107
Time domain
only coder
104
Calculator of
time domain
contribution
105
Calculator
of frequency–
domain
excitation
contribution
Calculator of
frequency transform of
time domain
contribution
106
108
Cut–off
frequency finder
and filter
Subtractor of a
difference between
filtered signal
and the frequency
transform
of the residual
109
Frequency quantizer
of the difference signal
110
Adder of quantized
difference signal
to the filtered signal
111
Converter back
to time domain
112
113
Synthesis filter

FIG.1

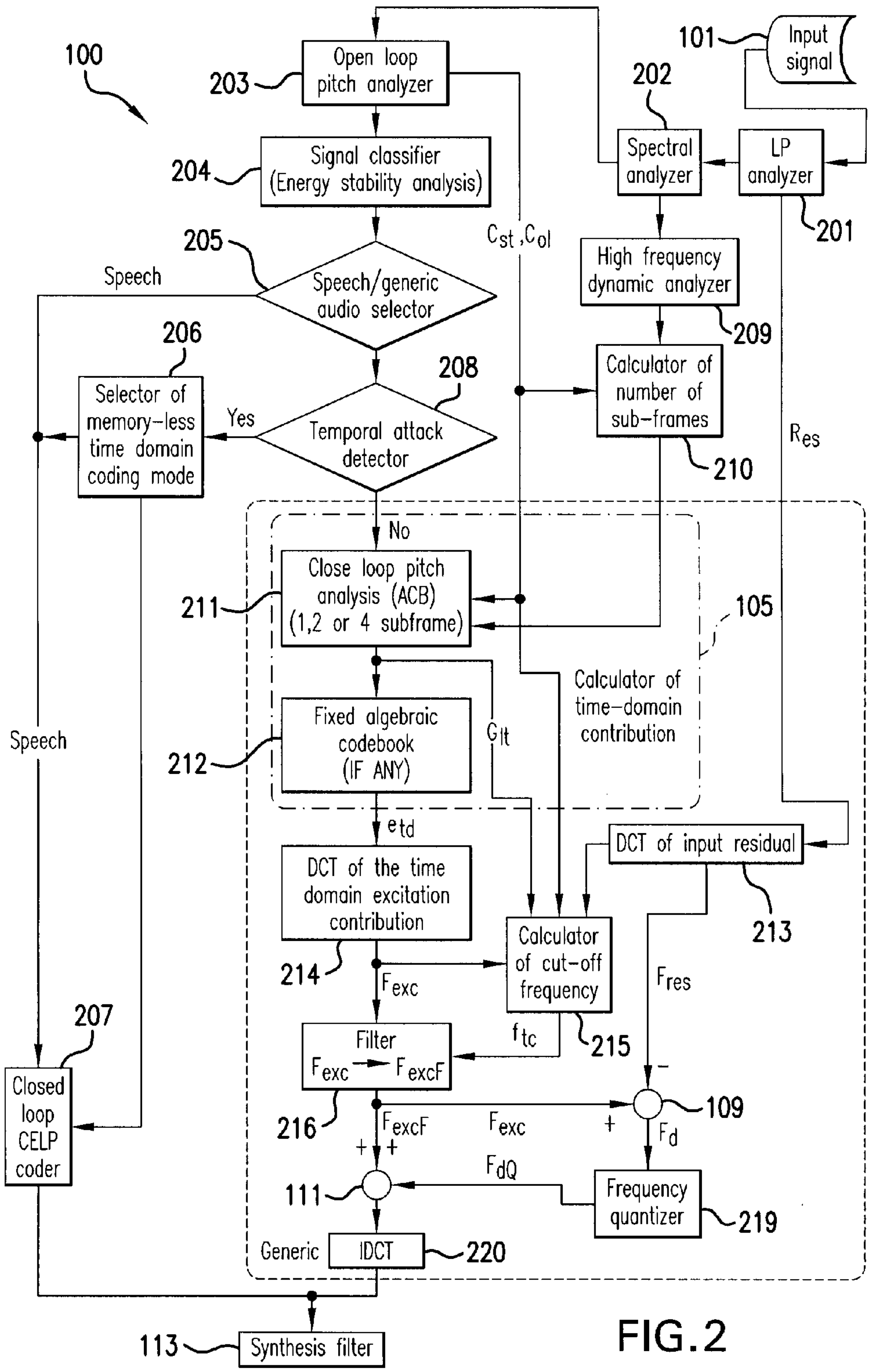
100
101
Input signal
201
LP analyzer
202
Spectral analyzer
203
Open loop pitch analyzer
204
Signal classifier (Energy stability analysis)
205
Speech/generic audio selector
Speech
206
Selector of memory-less time domain coding mode
Yes
208
Temporal attack detector
No
C_{st}, C_{ol}
High frequency dynamic analyzer
209
Calculator of number of sub-frames
210
R_{es}
105
211
Close loop pitch analysis (ACB) (1,2 or 4 subframe)
Calculator of time-domain contribution
212
Fixed algebraic codebook (IF ANY)
G_{lt}
e_{td}
DCT of input residual
213
DCT of the time domain excitation contribution
214
Calculator of cut-off frequency
215
F_{exc}
F_{res}
Speech
207
Closed loop CELP coder
Filter $F_{exc} \rightarrow F_{excF}$
f_{tc}
216
F_{excF}
F_{exc}
109
F_d
111
F_{dQ}
Frequency quantizer
219
Generic
IDCT
220
113
Synthesis filter

FIG.2

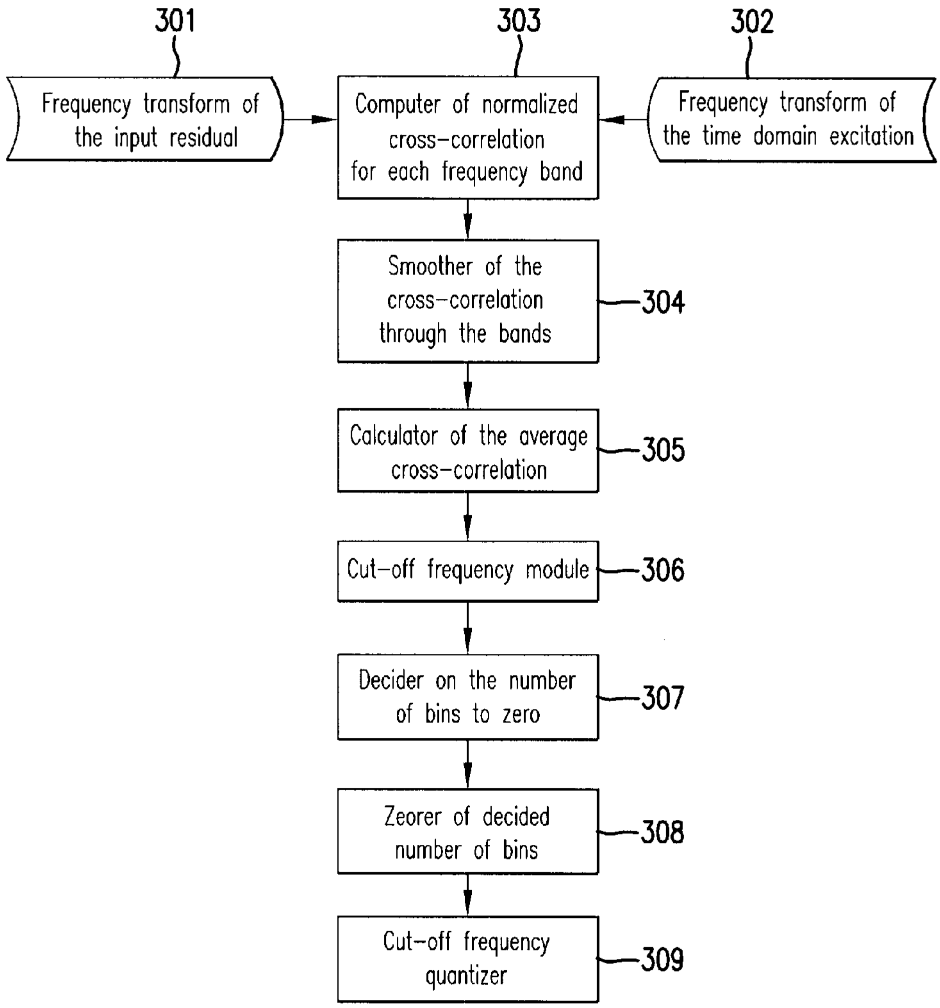
301
303
302
Frequency transform of the input residual
Computer of normalized cross-correlation for each frequency band
Frequency transform of the time domain excitation
Smoother of the cross-correlation through the bands
304
Calculator of the average cross-correlation
305
Cut-off frequency module
306
Decider on the number of bins to zero
307
Zeorer of decided number of bins
308
Cut-off frequency quantizer
309

FIG.3

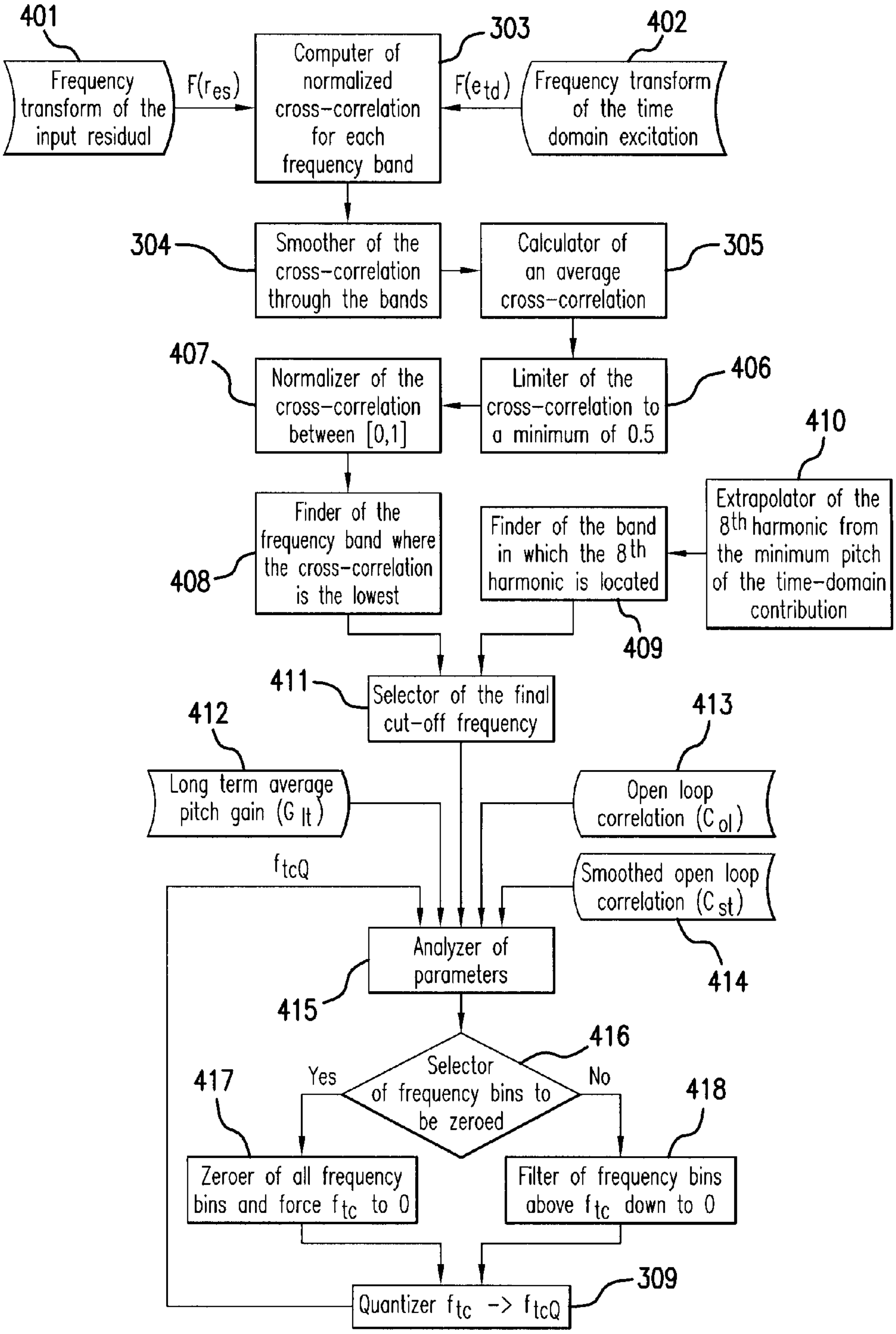
401
Frequency transform of the input residual
F(res)
Computer of normalized cross-correlation for each frequency band
303
F(etd)
402
Frequency transform of the time domain excitation
304
Smoother of the cross-correlation through the bands
Calculator of an average cross-correlation
305
407
Normalizer of the cross-correlation between [0,1]
Limiter of the cross-correlation to a minimum of 0.5
406
410
Finder of the frequency band where the cross-correlation is the lowest
408
Finder of the band in which the 8th harmonic is located
Extrapolator of the 8th harmonic from the minimum pitch of the time-domain contribution
409
411
Selector of the final cut-off frequency
412
Long term average pitch gain (G lt)
413
Open loop correlation (C ol)
ftcQ
Smoothed open loop correlation (C st)
Analyzer of parameters
415
414
416
Selector of frequency bins to be zeroed
Yes
No
417
418
Zeroer of all frequency bins and force ftc to 0
Filter of frequency bins above ftc down to 0
Quantizer ftc -> ftcQ
309

FIG.4

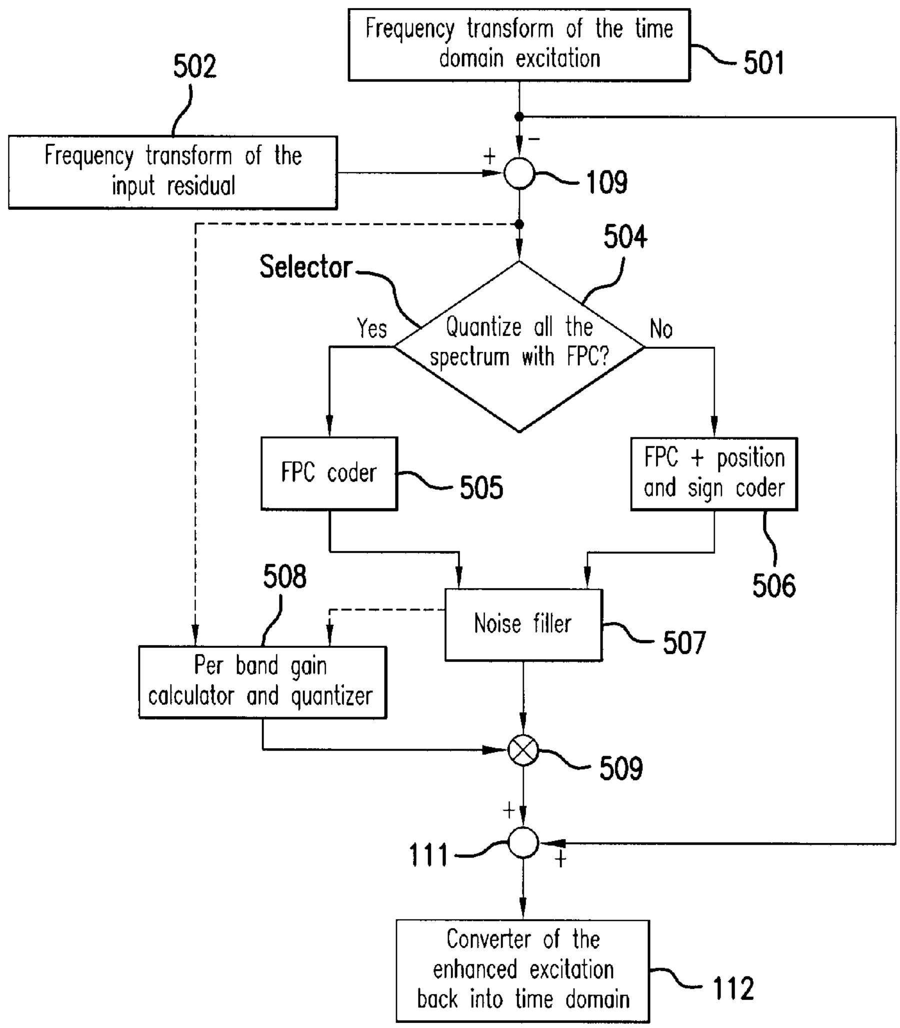
Frequency transform of the time domain excitation
501
502
Frequency transform of the input residual
+
−
109
504
Selector
Yes
Quantize all the spectrum with FPC?
No
FPC coder
505
FPC + position and sign coder
506
508
Noise filler
507
Per band gain calculator and quantizer
509
+
111
+
Converter of the enhanced excitation back into time domain
112

FIG.5

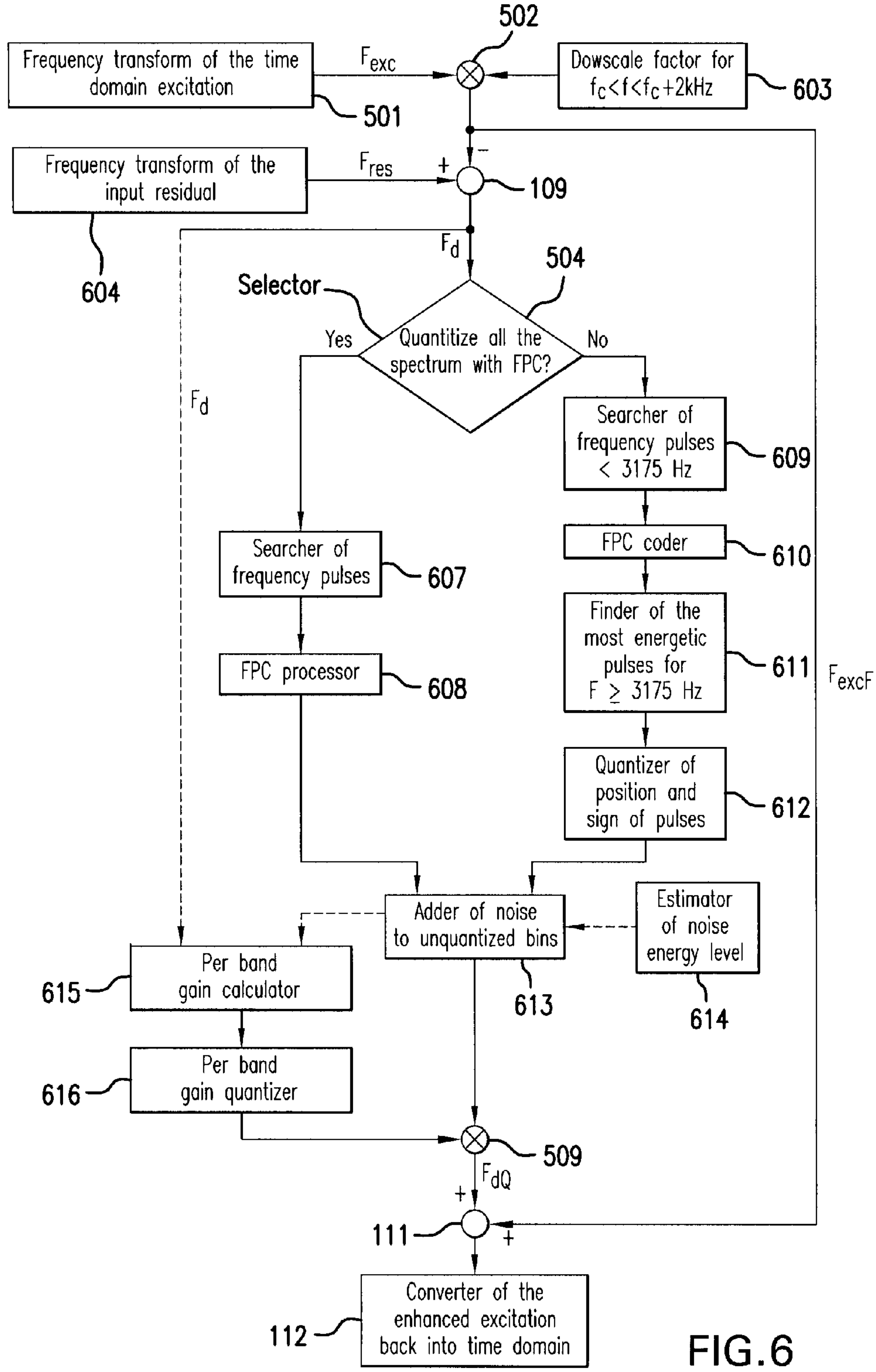
Frequency transform of the time domain excitation
501
F_{exc}
502
Dowscale factor for $f_c<f<f_c+2kHz$
603
Frequency transform of the input residual
604
F_{res}
+
−
109
F_d
504
Selector
Quantitize all the spectrum with FPC?
Yes
No
F_d
Searcher of frequency pulses < 3175 Hz
609
FPC coder
610
Finder of the most energetic pulses for F ≥ 3175 Hz
611
F_{excF}
Quantizer of position and sign of pulses
612
Searcher of frequency pulses
607
FPC processor
608
Adder of noise to unquantized bins
613
Estimator of noise energy level
614
615
Per band gain calculator
616
Per band gain quantizer
509
F_{dQ}
+
111
+
112
Converter of the enhanced excitation back into time domain

FIG.6

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Non-patent literature cited in the description

- **LEFEBVRE et al.** High quality coding of wideband audio signals using transform coded excitation (TCX). *Proceedings of ICASSP '94* **[0003]**
- Frame error robust narrow-band and wideband embedded variable bit-rate coding of speech and audio from 8-32 kbit/s. *ITU-T Recommendation G.718,* June 2008 **[0019]**
- **T.VAILLANCOURT et al.** Inter-tone noise reduction in a low bit rate CELP decoder. *Proc. IEEE ICASSP,* April 2009, 4113-16 **[0034]**
- **EKSLER, V. ; JELÍNEK, M.** Transition mode coding for source controlled CELP codecs. *IEEE Proceedings of International Conference on Acoustics, Speech and Signal Processing,* March 2008, 4001-40043 **[0035]**
- **MITTAL, U. ; ASHLEY, J.P. ; CRUZ-ZENO, E.M.** Low Complexity Factorial Pulse Coding of MDCT Coefficients using Approximation of Combinatorial Functions. *IEEE Proceedings on Acoustic, Speech and Signals Processing,* April 2007, vol. 1, 289-292 **[0069]**