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

(11) EP 2 863 390 B1

(12) **EUROPEAN PATENT SPECIFICATION**

(45) Date of publication and mention of the grant of the patent:
31.01.2018 Bulletin 2018/05

(51) Int Cl.:
G10L 19/26 *(2013.01)* *G10L 25/18* *(2013.01)*

(21) Application number: **15151693.7**

(22) Date of filing: **05.03.2009**

(54) **System and method for enhancing a decoded tonal sound signal**

System und Verfahren zur Verbesserung eines dekodierten tonalen Schallsignals

Système et procédé d'amélioration d'un signal de son tonal décodé

(84) Designated Contracting States:
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 SE SI SK TR

(30) Priority: **05.03.2008 US 64430**

(43) Date of publication of application:
22.04.2015 Bulletin 2015/17

(62) Document number(s) of the earlier application(s) in accordance with Art. 76 EPC:
09717868.5 / 2 252 996

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

(72) Inventors:
- **Vaillancourt, Tommy**
  Sherbrooke
  Québec J1N 2K1 (CA)
- **Jelinek, Milan**
  Sherbrooke
  Québec J1L 2W8 (CA)
- **Malenovsky, Vladimir**
  Sherbrooke
  Québec J1K 1L7 (CA)
- **Salami, Redwan**
  Saint Laurent
  Québec H4R 2Y3 (CA)

(74) Representative: **Ipside**
7-9 Allées Haussmann
33300 Bordeaux Cedex (FR)

(56) References cited:
US-A- 6 138 093

- **RAPPORTEUR Q9/16: "Updated draft new of new ITU-T Recommendation G.VBR-EV", ITU-T SG16 MEETING; 22-4-2008 - 2-5-2008; GENEVA,, no. T05-SG16-080422-TD-WP3-0338, 24 April 2008 (2008-04-24), XP030100513,**
- **WANG F M ET AL: "Frequency domain adaptive postfiltering for enhancement of noisy speech", SPEECH COMMUNICATION, ELSEVIER SCIENCE PUBLISHERS, AMSTERDAM, NL, vol. 12, no. 1, 1 March 1993 (1993-03-01), pages 41-56, XP026658543, ISSN: 0167-6393, DOI: 10.1016/0167-6393(93)90017-F [retrieved on 1993-03-01]**
- **Juin-Hwey Chen ET AL: "Adaptive postfiltering for quality enhancement of coded speech", IEEE Transactions on Speech and Audio Processing, 1 January 1995 (1995-01-01), pages 59-71, XP055104008, DOI: 10.1109/89.365380 Retrieved from the Internet: URL:http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=365380**

Note: Within nine months of the publication of the mention of the grant of the European patent in the European Patent Bulletin, any person may give notice to the European Patent Office of opposition to that patent, in accordance with the Implementing Regulations. Notice of opposition shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European Patent Convention).

EP 2 863 390 B1

Description

FIELD OF THE INVENTION

[0001] The present invention relates to a system and method for enhancing a decoded tonal sound signal, for example an audio signal such as a music signal coded using a speech-specific codec. For that purpose, the system and method reduce a level of quantization noise in regions of the spectrum exhibiting low energy.

BACKGROUND OF THE INVENTION

[0002] The demand for efficient digital speech and audio coding techniques with a good trade-off between subjective quality and bit rate is increasing in various application areas such as teleconferencing, multimedia, and wireless communications.

[0003] A speech coder converts a speech signal into a digital bit stream which is transmitted over a communication channel or stored in a storage medium. The speech signal is digitized, that is, sampled and quantized with usually 16-bits per sample. The speech coder has the role of representing the digital samples with a smaller number of bits while maintaining a good subjective speech quality. The speech decoder or synthesizer operates on the transmitted or stored bit stream and converts it back to a sound signal.

[0004] *Cade-Excited Linear Prediction* (CELP) coding is one of the best prior art techniques for achieving a good compromise between subjective quality and bit rate. The CELP coding technique is a basis of several speech coding standards both in wireless and wireline applications. In CELP coding, the sampled speech signal is processed in successive blocks of *L* samples usually called *frames,* where *L* is a predetermined number of samples corresponding typically to 10-30 ms. A linear prediction (LP) filter is computed and transmitted every frame. The computation of the LP filter typically uses a *lookahead,* for example a 5-15 ms speech segment from the subsequent frame. The *L*-sample frame is divided into smaller blocks called *subframes.* Usually the number of subframes is three (3) or four (4) resulting in 4-10 ms subframes. In each subframe, an excitation signal is usually obtained from two components, a past excitation and an innovative, fixed-codebook excitation. The component formed from the past excitation is often referred to as the adaptive-codebook or pitch-codebook excitation. The parameters characterizing the excitation signal are coded and transmitted to the decoder, where the excitation signal is reconstructed and used as the input of the LP filter.

[0005] In some applications, such as music-on-hold, low bit rate speech-specific codecs are used to operate on music signals. This usually results in bad music quality due to the use of a speech production model in a low bit rate speech-specific codec.

[0006] In some music signals, the spectrum exhibits a tonal structure wherein several tones are present (corresponding to spectral peaks) and are not harmonically related. These music signals are difficult to encode with a low bit rate speech-specific codec using an all-pole synthesis filter and a pitch filter. The pitch filter is capable of modeling voice segments in which the spectrum exhibits a harmonic structure comprising a fundamental frequency and harmonics of this fundamental frequency. However, such a pitch filter fails to properly model tones which are not harmonically related. Furthermore, the all-pole synthesis filter fails to model the spectral valleys between the tones. Thus, when a low bit rate speech-specific codec using a speech production model such as CELP is used, music signals exhibit an audible quantization noise in the low-energy regions of the spectrum (inter-tone regions or spectral valleys). An approach for reducing such inter-tone quantization noise is for example disclosed in RAPPORTEUR Q9/16: "Updated draft new of new ITU-T Recommendation G.VBR-EV", ITU-T SG16 MEETING, 22-4-2008 - 2-5-2008, GENEVA, no. T05-SG16-080422-TD-WP3-0338, 24 April 2008.

SUMMARY OF THE INVENTION

[0007] An objective of the present invention is to enhance a tonal sound signal decoded by a decoder of a speech-specific codec in response to a received coded bit stream, for example an audio signal such as a music signal, by reducing quantization noise in low-energy regions of the spectrum (inter-tone regions or spectral valleys).

[0008] More specifically, according to the present invention, there is provided a system for enhancing a decoded tonal sound signal according to claim 2.

[0009] The present invention also relates to a method for enhancing a decoded tonal sound signal according to claim 1.

[0010] The foregoing and other objects, advantages and features of the present invention will become more apparent upon reading of the following non restrictive description of illustrative embodiments thereof, given by way of example only with reference to the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0011] In the appended drawings:

Figure 1 is a schematic block diagram showing an overview of a system and method for enhancing a decoded tonal sound signal;

Figure 2 is a graph illustrating windowing in spectral analysis;

Figure 3 is a schematic block diagram showing an overview of a system and method for enhancing a decoded tonal sound signal;

Figure 4 is a schematic block diagram illustrating tone gain correction;

Figure 5 is a schematic block diagram of an example of signal type classifier; and

Figure 6 is a schematic block diagram of a decoder of a low bit rate speech-specific codec using a speech production model comprising a LP synthesis filter modeling the vocal tract shape (spectral envelope) and a pith filter modeling the vocal chords (harmonic fine structure).

DETAILED DESCRIPTION

[0012] In the following detailed description, an inter-tone noise reduction technique is performed within a low bit rate speech-specific codec to reduce a level of inter-tone quantization noise for example in musical content. The inter-tone noise reduction technique can be deployed with either narrowband sound signals sampled at 8000 samples/s or wideband sound signals sampled at 16000 samples/s or at any other sampling frequency. The inter-tone noise reduction technique is applied to a decoded tonal sound signal to reduce the quantization noise in the spectral valleys (low energy regions between tones). In some music signals, the spectrum exhibits a tonal structure wherein several tones are present (corresponding to spectral peaks) and are not harmonically related. These music signals are difficult to encode with a low bit rate speech-specific codec which uses an all-pole LP synthesis filter and a pitch filter. The pitch filter can model voiced speech segments having a spectrum that exhibits a harmonic structure with a fundamental frequency and harmonics of that fundamental frequency. However, the pitch filter fails to properly model tones which are not harmonically related. Further, the all-pole LP synthesis filter fails to model the spectral valleys between the tones. Thus, using a low bit rate speech-specific codec with a speech production model such as CELP, the modeled signals will exhibit an audible quantization noise in the low-energy regions of the spectrum (inter-tone regions or spectral valleys). The inter-tone noise reduction technique is therefore concerned with reducing the quantization noise in low-energy spectral regions to enhance a decoded tonal sound signal, more specifically to enhance quality of the decoded tonal sound signal.

[0013] In one embodiment, the low bit rate speech-specific codec is based on a CELP speech production model operating on either narrowband or wideband signals (8 or 16 kHz sampling frequency). Any other sampling frequency could also be used.

[0014] An example 600 of the decoder of a low bit rate speech-specific codec using a CELP speech production model will be briefly described with reference to Figure 6. In response to a fixed codebook index extracted from the received coded bit stream, a fixed codebook 601 produces a fixed-codebook vector 602 multiplied by a fixed-codebook gain *g* to produce an innovative, fixed-codebook excitation 603. In a similar manner, an adaptive codebook 604 is responsive to a pitch delay extracted from the received coded bit stream to produce an adaptive-codebook vector 607; the adaptive codebook 604 is also supplied (see 605) with the excitation signal 610 through a feedback loop comprising a pitch filter 606. The adaptive-codebook vector 607 is multiplied by a gain G to produce an adaptive-codebook excitation 608. The innovative, fixed-codebook excitation 603 and the adaptive-codebook excitation 608 are summed through an adder 609 to form the excitation signal 610 supplied to an LP synthesis filter 611; the LP synthesis filter 611 is controlled by LP filter parameters extracted from the received coded bit stream. The LP synthesis filter 611 produces a synthesis sound signal 612, or decoded tonal sound signal that can be upsampled/downsampled in module 613 before being enhanced using the system 100 and method for enhancing a decoded tonal sound signal.

[0015] For example, a codec based on the AMR-WB ([1] - 3GPP TS 26.190, "Adaptive Multi-Rate - Wideband (AMR-WB) speech codec; Transcoding functions") structure can be used. The AMR-WB speech codec uses an internal sampling frequency of 12.8 kHz, and the signal can be re-sampled to either 8 or 16 kHz before performing reduction of the inter-tone quantization noise or, alternatively, noise reduction or audio enhancement can be performed at 12.8 kHz.

[0016] Figure 1 is a schematic block diagram showing an overview of a system and method 100 for enhancing a decoded tonal sound signal.

[0017] Referring to Figure 1, a coded bit stream 101 (coded sound signal) is received and processed through a decoder 102 (for example the decoder 600 of Figure 6) of a low bit rate speech-specific codec to produce a decoded sound signal 103. As indicated in the foregoing description, the decoder 102 can be, for example, a speech-specific decoder using a CELP speech production model such as an AMR-WB decoder.

[0018] The decoded sound signal 103 at the output of the sound signal decoder 102 is converted (re-sampled) to a sampling frequency of 8 kHz. However, it should be kept in mind that the inter-tone noise reduction technique disclosed herein can be equally applied to decoded tonal sound signals at other sampling frequencies such as 12.8 kHz or 16 kHz.

[0019] Preprocessing can be applied or not to the decoded sound signal 103. When preprocessing is applied, the decoded sound signal 103 is, for example, pre-emphasized through a preprocessor 104 before spectral analysis in the spectral analyser 105 is performed.

[0020] To pre-emphasize the decoded sound signal 103, the preprocessor 104 comprises a first order high-pass filter (not shown). The first order high-pass filter emphasizes higher frequencies of the decoded sound signal 103 and may have, for that purpose, the following transfer function:

$$H_{\text{pre-emph}}(z) = 1 - 0.68z^{-1} \qquad (1)$$

where z represents the Z-transform variable.

[0021] Pre-emphasis of the higher frequencies of the decoded sound signal 103 has the property of flattening the spectrum of the decoded sound signal 103, which is useful for inter-tone noise reduction.

[0022] Following the pre-emphasis of the higher frequencies of the decoded sound signal 103 in the preprocessor 104:

- Spectral analysis of the pre-emphasized decoded sound signal 10b is performed in the spectral analyser 105. This spectral analysis uses Discrete Fourier Transform (DFT) and will be described in more detail in the following description.

- The inter-tone noise reduction technique is applied in response to the spectral parameters 107 from the spectral analyser 107 and is implemented in a reducer 108 of quantization noise in the low-energy spectral regions of the decoded tonal sound signal. The operation of the reducer 108 of quantization noise will be described in more detail in the following description.

- An inverse analyser and overlap-add operator 110 (a) applies an inverse DFT (Discrete Fourier Transform) to the inter-tone noise reduced spectral parameters 109 to convert those parameters 109 back to the time domain, and (b) uses an overlap-add operation to reconstruct the enhanced decoded tonal sound signal 111. The operation of the inverse analyser and overlap-add operator 110 will be described in more detail in the following description.

- A postprocessor 112 post-processes the reconstructed enhanced decoded tonal sound signal 111 from the inverse analyser and overlap-add operator 110, This post-processing is the inverse of the preprocessing stage (preprocessor 104) and, therefore, may consist of de-emphasis of the higher frequencies of the enhanced decoded tonal sound signal. Such de-emphasis will be described in more detail in the following description.

- Finally, a sound playback system 114 may be provided to convert the post-processed enhanced decoded tonal sound signal 113 from the postprocessor 112 into an audible sound.

[0023] For example, the speech-specific codec in which the inter-tone noise reduction technique is implemented operates on 20 ms frames containing 160 samples at a sampling frequency of 8 kHz. Also according to this example, the sound signal decoder 102 uses a 10 ms lookahead from the future frame for best frame erasure concealment performance. This lookahead is also used in the inter-tone noise reduction technique for a better frequency resolution. The inter-tone noise reduction technique implemented in the reduced 108 of quantization noise follows the same framing structure as in the decoder 102. However, some shift can be introduced between the decoder framing structure and the inter-tone noise reduction framing structure to maximize the use of the lookahead. In the following description, the indices attributed to samples will reflect the inter-tone noise reduction framing structure.

Spectral analysis

[0024] Referring to Figure 3, DFT (Discrete Fourier Transform) is used in the spectral analyser 105 to perform a spectral analysis and spectrum energy estimation of the pre-emphasized decoded tonal sound signal 106. In the spectral analyser 105, spectral analysis is performed in each frame using 30 ms analysis windows with 33% overlap. More

specifically, the spectral analysis in the analyser 105 (Figure 3) is conducted once per frame using a 256-point Fast Fourier Transform (DFT) with the 33.3 percent overlap windowing as illustrated in Figure 2. The analysis windows are placed so as to exploit the entire lookahead. The beginning of the first analysis window is shifted 80 samples after the beginning of the current frame of the sound signal decoder 102.

[0025] The analysis windows are used to weight the pre-emphasized, decoded tonal sound signal 106 for frequency analysis. The analysis windows are flat in the middle with sine function on the edges (Figure 2) which is well suited for overlap-add operations. More specifically, the analysis window can be described as follow:

$$w_{FFT}(n) = \begin{cases} \sin\left(\dfrac{\pi n}{2L_{window}/3}\right), & n = 0,\ldots,L_{window}/3-1 \\ 1, & n = L_{window}/3,\ldots,2L_{window}/3-1 \\ \sin\left(\dfrac{\pi(n-L_{window}/3)}{2L_{window}/3}\right), & n = 2L_{window}/3,..,L_{window}-1 \end{cases}$$

where L_{Window}= 240 samples is the size of the analysis window. Since a 256-point FTT (L_{FFT} = 256) is used, the windowed signal is padded with 16 zero samples.

[0026] An alternative analysis window could be used in the case of a wideband signal with only a small lookahead available. This analysis window could have the following shape:

$$w_{FFT_{WB}}(n) = \begin{cases} \sin\left(\dfrac{\pi n}{2.L_{window_{WB}}/9}\right), & n = 0,\ldots,L_{window_{WB}}/9-1 \\ 1, & n = L_{window_{WB}}/9,\ldots,8.L_{window_{WB}}/9-1 \\ \sin\left(\dfrac{\pi(n-L_{window_{WB}}/9)}{2.L_{window_{WB}}/9}\right), & n = 8.L_{window_{WB}}/9,\ldots,L_{window_{WB}}-1 \end{cases}$$

where $L_{windowWB}$ = 360 is the size of the wideband analysis window, In that case, a 512-point FFT is used. Therefore, the windowed signal is padded with 152 zero samples. Other radix FFT can potentially be used to reduce as much as possible the zero padding and reduce the complexity.

[0027] Let $s'(n)$ denote the decoded tonal sound signal with index 0 corresponding to the first sample in the inter-tone noise reduction frame (As indicated hereinabove, in this embodiment, this corresponds to 80 samples following the beginning of the sound signal decoder frame). The windowed decoded tonal sound signal for the spectral analysis can be obtained using the following relation:

$$x_w^{(1)}(n) = \begin{cases} w_{FFT}(n)s'(n), & n = 0,\ldots,L_{window}-1 \\ 0, & n = L_{window},\ldots,L_{FFT}-1 \end{cases} \qquad (2)$$

where $s'(0)$ is the first sample in the current inter-tone noise reduction frame.

[0028] FFT is performed on the windowed, decoded tonal sound signal to obtain one set of spectral parameters per frame:

$$X^{(1)}(k) = \sum_{n=0}^{N-1} x_w^{(1)}(n)e^{-j2\pi\frac{kn}{N}}, \quad k = 0,\ldots,L_{FFT}-1 \qquad (3)$$

where $N = L_{FFT}$.

[0029] The output of the FFT gives real and imaginary parts of the spectrum denoted by $X_R(k)$, k=0 to $\frac{L_{FFT}}{2}$, and $X_I(k)$, k=1 to $\left(\frac{L_{FFT}}{2}-1\right)$. Note that $X_R(0)$ corresponds to the spectrum at 0 Hz (DC) and $X_R(\frac{L_{FFT}}{2})$ corresponds to the spectrum at $\frac{F_S}{2}$ Hz, where F_S corresponds to the sampling frequency. The spectrum at these two (2) points is only real valued and usually ignored in the subsequent analysis.

[0030] After the FFT analysis, the resulting spectrum is divided into critical frequency bands using the intervals having the following upper limits; (17 critical bands in the frequency range 0-4000 Hz and 21 critical frequency bands in the frequency range 0-8000 Hz) (See [2]: J. D. Johnston, "Transform coding of audio signal using perceptual noise criteria," IEEE J. Select. Areas Commun., vol. 6, pp. 314-323, Feb. 1988).

[0031] In the case of narrowband coding, the critical frequency bands = {100.0, 200.0, 300.0, 400.0, 510.0, 630.0, 770.0, 920.0, 1080.0, 1270.0, 1480.0, 1720.0, 2000.0, 2320.0, 2700.0, 3150.0, 3700.0, 3950.0} Hz.

[0032] In the case of wideband coding, the critical frequency bands = {100.0, 200.0, 300.0, 400.0, 510.0, 630.0, 770.0, 920.0, 1080.0, 1270.0, 1480.0, 1720.0, 2000.0, 2320.0, 2700.0, 3150.0, 3700.0, 4400.0, 5300.0, 6700.0, 8000.0} Hz.

[0033] The 256-point or 512-point FFT results in a frequency resolution of 31.25 Hz (4000/128=8000/256). After ignoring the DC component of the spectrum, the number of frequency bins per critical frequency band in the case of narrowband coding is M_{CB}= {3, 3, 3, 3, 3, 4, 5, 4, 5, 6, 7, 7, 9, 10, 12, 14, 17, 12}, respectively, when the resolution is approximated to 32Hz. In the case of wideband coding M_{CB}= {3, 3, 3, 3, 3, 4, 5, 4, 5, 6, 7, 7, 9, 10, 12, 14, 17, 22, 28, 44, 41}.

[0034] The average spectral energy per critical frequency band is computed as follows:

$$E_{CB}(i) = \frac{1}{(L_{FFT}/2)^2 M_{CB}(i)} \sum_{k=0}^{M_{CB}(i)-1} \left(X_R^2(k+j_i) + X_I^2(k+j_i)\right), \quad i = 0,...,17, \qquad (4)$$

where $X_R(k)$ and $X_I(k)$ are, respectively, the real and imaginary parts of the k^{th} frequency bin and j_i is the index of the first bin in the i^{th} critical band given by j_i = {1, 4, 7, 10, 13, 16, 20, 25, 29, 34, 40, 47, 54, 63, 73, 85, 99, 116} in the case of narrowband coding and j_i = {1, 4, 7, 10, 13, 16, 20, 25, 29, 34, 40, 47, 54, 63, 73, 85, 99, 116, 13 8, 166, 210} in the case of wideband coding.

[0035] The spectral analyser 105 of Figure 3 also computes the energy of the spectrum per frequency bin, $E_{BIN}(k)$, for the first 17 critical bands (115 bins excluding the DC component) using the following relation:

$$E_{BIN}(k) = X_R^2(k) + X_I^2(k), \qquad k = 0,...,114 \qquad (5)$$

[0036] Finally, the spectral analyser 105 computes a total frame spectral energy as an average of the spectral energies of the first 17 critical frequency bands calculated by the spectral analyser 105 in a frame using, the following relation:

$$E_{fr}^t = 10\log\left(\sum_{i=0}^{i=16} \bar{E}_{CB}(i)\right), \quad \text{dB} \qquad (6)$$

[0037] The spectral parameters 107 from the spectral analyser 105 of Figure 3, more specifically the above calculated average spectral energy per critical band, spectral energy per frequency bin, and total frame spectral energy are used in the reducer 108 to reduce quantization noise and perform gain correction.

[0038] It should be noted that, for a wideband decoded tonal sound signal sampled at 16000 samples/s, up to 21 critical frequency bands could be used but computation of the total frame energy E_{fr}^t at time t will still be performed on the first 17 critical bands.

Signal type classifier:

[0039] The inter-tone noise reduction technique conducted by the system and method 100 enhances a decoded tonal sound signal, such as a music signal, coded by means of a speech-specific codec. Usually, non-tonal sounds such as speech are well coded by a speech-specific codec and do not need this type of frequency based enhancement.

[0040] The system and method 100 for enhancing a decoded tonal sound signal further comprises, as illustrated in Figure 3, a signal type classifier 301 designed to further maximize the efficiency of the reducer 108 of quantization noise by identifying which sound is well suited for inter-tone noise reduction, like music, and which sound is not, like speech.

[0041] The signal type classifier 301 comprises the feature of not only separating the decoded sound signal into sound signal categories, but also to give instruction to the reducer 108 of quantization noise to reduce at a minimum any possible degradation of speech.

[0042] A schematic block diagram of the signal type classifier 301 is illustrated in Figure 5. In the presented embodiment, the signal type classifier 301 has been kept as simple as possible. The principal input to the signal type classifier 301 is the total frame spectral energy E_i as formulated in Equation (6).

[0043] First, the signal type classifier 301 comprises a finder 501 that determines a mean of the past forty (40) total frame spectral energy (E_i) variations calculated using the following relation:

$$\overline{E}_{diff} = \frac{\left(\sum_{t=-40}^{t=-1} \Delta_E^t\right)}{40}, \quad where \quad \Delta_E^t = E_{fr}^t - E_{fr}^{(t-1)} \tag{7}$$

[0044] Then, the finder 501 determines a statistical deviation of the energy variation history σ_E over the last fifteen (15) frames using the following relation:

$$\sigma_E = 0.7745967 \cdot \sqrt{\sum_{t=-15}^{t=-1} \frac{\left(\Delta_E^t - \overline{E}_{diff}\right)^2}{15}} \tag{8}$$

[0045] The signal type classifier 301 comprises a memory 502 updated with the mean and deviation of the variation of the total frame spectral energy E_i as calculated in Equations (7) and (8).

[0046] The resulting deviation σ_E is compared to four (4) floating thresholds in comparators 503-506 to determine the efficiency of the reducer 108 of quantization noise on the current decoded sound signal. In the example of Figure 5, the output 302 (Figure 3) of the signal type classifier 301 is split into five (5) sound signal categories, named sound signal categories 0 to 4, each sound signal category having its own inter-tone noise reduction tuning.

[0047] The five (5) sound signal categories 0-4 can be determined as indicated in the following Table:

Category	Enhanced band (narrowband)	Enhanced band (wideband)	Allowed reduction
	Hz	Hz	dB
0	NA	NA	0
1	[2000, 4000]	[2000, 8000]	6
2	[1270, 4000]	[1270, 8000]	9
3	[700, 4000]	[700, 8000]	12
4	[400, 4000]	[400, 8000]	12

[0048] The sound signal category 0 is a non-tonal sound signal category, like speech, which is not modified by the inter-tone noise reduction technique. This category of decoded sound signal has a large statistical deviation of the spectral energy variation history. When detection of categories 1-4 by the comparators 503-506 is negative, a controller 511 instructs the reducer 108 of quantization noise not to reduce inter-tone quantization noise (Reduction = 0 dB).

[0049] The tree in between sound signal categories includes sound signals with different types of statistical deviation of spectral energy variation history.

[0050] Sound signal category 1 (biggest variation after "speech type" decoded sound signal) is detected by the comparator 506 when the statistical deviation of spectral energy variation history is lower than a Threshold 1. A controller

510 is responsive to such a detection by the comparator 506 to instruct, when the last detected sound signal category was ≥ 0, the reducer 108 of quantization noise to enhance the decoded tonal sound signal within the frequency band 2000 to $\frac{F_S}{2}$ Hz by reducing the inter-tone quantization noise by a maximum allowed amplitude of 6 dB.

[0051] Sound signal category 2 is detected by the comparator 505 when the statistical deviation of spectral energy variation history is lower than a Threshold 2. A controller 509 is responsive to such a detection by the comparator 505 to instruct, when the last detected sound signal category was ≥ 1, the reducer 108 of quantization noise to enhance the decoded tonal sound signal within the frequency band 1270 to $\frac{F_S}{2}$ Hz by reducing the inter-tone quantization noise by a maximum allowed amplitude of 9 dB.

[0052] Sound signal category 3 is detected by the comparator 504 when the statistical deviation of spectral energy variation history is lower than a Threshold 3. A controller 508 is responsive to such a detection by the comparator 504 to instruct, when the last detected sound signal category was ≥ 2, the reducer 108 of quantization noise to enhance the decoded tonal sound signal within the frequency band 700 to $\frac{F_S}{2}$ Hz by reducing the inter-tone quantization noise by a maximum allowed amplitude of 12 dB.

[0053] Sound signal category 4 is detected by the comparator 503 when the statistical deviation of spectral energy variation history is lower than a Threshold 4. A controller 507 is responsive to such a detection by the comparator 503 to instruct, when the last detected signal type category was ≥ 3, the reducer 108 of quantization noise to enhance the decoded tonal sound signal within the frequency band 400 to $\frac{F_S}{2}$ Hz by reducing the inter-tone quantization noise by a maximum allowed amplitude of 12 dB.

[0054] In the embodiment of Figure 5, the signal type classifier 301 uses floating thresholds 1-4 to split the decoded sound signal into the different categories 0-4. These floating thresholds 1-4 are particularly useful to prevent wrong signal type classification. Typically, decoded tonal sound signal like music gets much lower statistical deviation of its spectral energy variation than non-tonal sound signal like speech. But music could contain higher statistical deviation and speech could contain lower statistical deviation. It is unlikely that speech or music content changes from one to another on a frame basis. The floating thresholds acts like reinforcement to prevent any misclassification that could result in a suboptimal performance of the reducer 108 of quantization noise.

[0055] Counters of a series of frames of sound signal category 0 and of a series of frames of sound signal category 3 or 4 are used to respectively decrease or increase thresholds.

[0056] For example, if a counter 512 counts a series of more than 30 frames of sound signal category 3 or 4, the floating thresholds 1-4 will be increased by a threshold controller 514 for the purpose of allowing more frames to be considered as sound signal category 4. Each time the count of the counter 512 is incremented, the counter 513 is reset to zero.

[0057] The inverse is also true with sound signal category 0. For example, if a counter 513 counts a series of more than 30 frames of sound signal category 0, the threshold controller 514 decreases the floating thresholds 1-4 for the purpose of allowing more frames to be considered as sound signal category 0. The floating thresholds 1-4 are limited to absolute maximum and minimum values to ensure that the signal type classifier 301 is not locked to a fixed category.

[0058] The increase and decrease of the thresholds 1-4 can be illustrated by the following relations:

$$IF\left(Nbr_cat4_frame > 30\right)$$

$$Thres(i) = Thres(i) + TH_UP\Big|_{i=1}^{4}$$

$$Thres(i) = Thres(i) + TH_UP\Big|_{i=1}^{4}$$
$$ELSE\ IF\ \left(Nbr_cat0_frame > 30\right)$$

$$Thres(i) = Thres(i) - TH_DWN\Big|_{i=1}^{4}$$

$$Thres(i) = MIN(Thres(i), MAX_TH)\Big|_{i=1}^{4}$$

$$Thres(i) = MAX(Thres(i), MIN_TH)\Big|_{i=1}^{4}$$

[0059] In the case of frame erasure, all the thresholds 1-4 are reset to theirs minimum values and the output of the signal type classifier 301 is considered as non-tonal (sound signal category 0) for three (3) frames including the lost frame.

[0060] If information from a Voice Activity Detector (VAD) (not shown) is available and is indicating no voice activity (presence of silence), the decision of the signal type classifier 301 is forced to sound signal category 0.

[0061] According to an alternative of the signal type classifier 301, the frequency band of allowed enhancement and/or the level of maximum inter-tone noise reduction could be completely dynamic (without hard step).

[0062] In the case of a small lookahead, it could be necessary to introduce a minimum gain reduction smoothing in the first critical bands to further reduce any potential distortion introduced with the inter-tone noise reduction. This smoothing could be performed using the following relation:

$$\mathrm{RedGain}_i = 1.0\Big|_{i=[0,FEhBand]} ;$$

$$\mathrm{RedGain}_i = \mathrm{RedGain}_{i-1} - \frac{(1.0 - Allow_red)}{(10 - FEhBand)}\Big|_{i=]FEhBand,10]} ;$$

$$\mathrm{RedGain}_i = Allow_red\Big|_{i=]10,max_band]}$$

where $RedGain_i$ is a maximum gain reduction per band, *FEhBand* is the first band where the inter-tone noise reduction is allowed (vary typically between 400Hz and 2kHz or critical frequency bands 3 and 12), *Allow_red* is the level of noise reduction allowed per sound signal category presented in the previous table and *max_band* is the maximum band for the inter tone noise reduction (17 for Narrowband (NB) and 20 for Wideband (WB)).

Inter-tone noise reduction:

[0063] Inter-tone noise reduction is applied (see reducer 108 of quantization noise (Figure 3)) and the enhanced decoded sound signal is reconstructed using an overlap and add operation (see overlap add operator 303 (Figure 3)). The reduction of inter-tone quantization noise is performed by scaling the spectrum in each critical frequency band with a scaling gain limited between g_{min} and 1 and derived from the signal-to-noise ratio (SNR) in that critical frequency band. A feature of the inter-tone noise reduction technique is that for frequencies lower than a certain frequency, for example related to signal voicing, the processing is performed on a frequency bin basis and not on critical frequency band basis. Thus, a scaling gain is applied on every frequency bin derived from the SNR in that bin (the SNR is computed using the bin energy divided by the noise energy of the critical band including that bin). This feature has the effect of preserving the energy at frequencies near harmonics or tones preventing distortion while strongly reducing the quantization noise between the harmonics. In the case of narrow band signals, per bin analysis can be used for the whole spectrum. Per bin analysis can alternatively be used in all critical frequency bands except the last one.

[0064] Referring to Figure 3, inter-tone quantization noise reduction is performed in the reducer 108 of quantization noise. According to a first possible implementation, per bin processing can be performed over all the 115 frequency bins in narrowband coding (250 frequency bins in wideband coding) in a noise attenuator 304.

[0065] In an alternative implementation, noise attenuator 304 perform per bin processing to apply a scaling gain to each frequency bin in the first voiced *K* bands and then noise attenuator 305 performs per band processing to scale the spectrum in each of the remaining critical frequency bands with a scaling gain. If *K*=0 then the noise attenuator 305 performs per band processing in all the critical frequency bands.

[0066] The minimum scaling gain g_{min} is derived from the maximum allowed inter-tone noise reduction in dB, NR_{max}. As described in the foregoing description (see the table above), the signal type classifier 301 makes the maximum allowed noise reduction NR_{max} varying between 6 and 12 dB. Thus minimum scaling gain is given by the relation:

$$g_{min} = 10^{-NR_{max}/20} \qquad (9)$$

[0067] In the case of a narrowband tonal frame, the scaling gain can be computed in relation to the SNR per frequency bin then per bin noise reduction is performed. Per bin processing is applied only to the first 17 critical bands corresponding to a maximum frequency of 3700 Hz. The maximum number of frequency bins in which per bin processing can be used is 115 (the number of bins in the first 17 bands at 4 kHz).

[0068] In the case of a wideband tonal frame, per bin processing is applied to all the 21 critical frequency bands corresponding to a maximum frequency of 8000 Hz. The maximum number of frequency bins for which per bin processing can be used is 250 (the number of bins in the first 21 bands at 8kHz).

[0069] In the inter-tone noise reduction technique, noise reduction starts at the fourth critical frequency band (no reduction performed before 400 Hz). To reduce any negative impact of the inter-tone quantization noise reduction technique, the signal type classifier 301 could push the starting critical frequency band up to the 12th. This means that the first critical frequency band on which inter-tone noise reduction is performed is somewhere between 400 Hz and 2 kHz and could vary on a frame basis.

[0070] The scaling gain for a certain critical frequency band, or for a certain frequency bin, can be computed as a function of the SNR in that frequency band or bin using the following relation:

$$(g_s)^2 = k_s\, SNR + c_s \text{, bounded by } g_{\min} \le g_s \le 1 \qquad (10)$$

[0071] The values of k_s and c_s are determined such that $g_s = g_{\min}$ for SNR =1 dB, and $g_s = 1$ for SNR = 45 dB. That is, for SNRs at 1 dB and lower, the scaling gain is limited to g_s and for SNRs at 45 dB and higher, no inter-tone noise reduction is performed in the given critical frequency band (g_s =1). Thus, given these two end points, the values of k_s and c_s in Equation (10) can be calculated using the following relations:

$$k_s = (1 - g_{\min}^2)/44 \quad \text{and} \quad c_s = (45 g_{\min}^2 - 1)/44. \qquad (11)$$

[0072] The variable SNR of Equation (10) is either the SNR per critical frequency band, $SNR_{CB}(i)$, or the SNR per frequency bin, $SNR_{BIN}(k)$, depending on the type of per bin or per band processing.

[0073] The SNR per critical frequency band is computed as follows:

$$SNR_{CB}(i) = \frac{0.3E_{CB}^{(1)}(i) + 0.7E_{CB}^{(2)}(i)}{N_{CB}(i)} \qquad i = 0,\ldots,17 \qquad (12)$$

where $E_{CB}^{(1)}(i)$ and $E_{CB}^{(2)}(i)$ denote the energy per critical frequency band for the past and current frame spectral analyses, respectively (as computed in Equation (4)), and $N_{CB}(i)$ denote the noise energy estimate per critical frequency band.

[0074] The SNR per frequency bin in a certain critical frequency band i is computed using the following relation:

$$SNR_{BIN}(k) = \frac{0.3E_{BIN}^{(1)}(k) + 0.7E_{BIN}^{(2)}(k)}{N_{CB}(i)}, \qquad k = j_i,\ldots, j_i + M_{CB}(i) - 1 \qquad (13)$$

where $E_{BIN}^{(1)}(k)$ and $E_{BIN}^{(2)}(k)$ denote the energy per frequency bin for the past (1) and the current (2) frame spectral analysis, respectively (as computed in Equation (5)), $N_{CB}(i)$ denote the noise energy estimate per critical frequency band, j_i is the index of the first frequency bin in the i^{th} critical frequency band and $M_{CB}(i)$ is the number of frequency bins in critical frequency band i as defined herein above.

[0075] According to another, alternative implementation, the scaling gain could be computed in relation to the SNR per critical frequency band or per frequency bin for the first voiced bands. If $K_{VOIC} > 0$ then per bin processing can be performed in the first K_{VOIC} bands. Per band processing can then be used for the rest of the bands. In the case where $K_{VOIC} = 0$ per band processing can be used over the whole spectrum.

[0076] In the case of per band processing for a critical frequency band with index i, after determining the scaling gain using Equation (10) and the SNR as defined in Equation (12) or (13), the actual scaling is performed using a smoothed scaling gain updated in every spectral analysis by means of the following relation:

$$g_{CB,LP}(i) = \alpha_{gs} g_{CB,LP}(i) + (1 - \alpha_{gs}) g_s \qquad (14)$$

[0077] According to a feature, the smoothing factor α_{gs} used for smoothing the scaling gain g_s and can be made adaptive and inversely related to the scaling gain g_s itself. For example, the smoothing factor can be given by α_{gs}=1-g_s. Therefore, the smoothing is stronger for smaller gains g_s. This approach prevents distortion in high SNR segments preceded by low SNR frames, as it is the case for voiced onsets. In the proposed approach, the smoothing procedure is able to quickly adapt and use lower scaling gains upon occurrence of, for example, a voiced onset.

[0078] Scaling in a critical frequency band is performed as follows:

$$\begin{aligned} X'_R(k + j_i) &= g_{CB,LP}(i)\, X_R(k + j_i), \quad \text{and} \\ X'_I(k + j_i) &= g_{CB,LP}(i)\, X_I(k + j_i), \quad k = 0,\ldots, M_{CB}(i) - 1, \end{aligned} \qquad (15)$$

where j_i is the index of the first frequency bin in the critical frequency band i and $M_{CB}(i)$ is the number of frequency bins in that critical frequency band.

[0079] In the case of per bin processing in a critical frequency band with index i, after determining the scaling gain using Equation (10) and the SNR as defined in Equation (12) or (13), the actual scaling is performed using a smoothed scaling gain updated in every spectral analysis as follows:

$$g_{BIN,LP}(k) = \alpha_{gs} g_{BIN,LP}(k) + (1 - \alpha_{gs}) g_s \qquad (16)$$

where the smoothing factor α_{gs} = 1-g_s is similar to Equation (14).

[0080] Temporal smoothing of the scaling gains prevents audible energy oscillations, while controlling the smoothing using α_{gs} prevents distortion in high SNR speech segments preceded by low SNR frames, as it is the case for voiced onsets for example.

[0081] Scaling in a critical frequency band i is then performed as follows:

$$\begin{aligned} X'_R(k + j_i) &= g_{BIN,LP}(k + j_i)\, X_R(k + j_i), \quad \text{and} \\ X'_I(k + j_i) &= g_{BIN,LP}(k + j_i)\, X_I(k + j_i), \quad k = 0,\ldots, M_{CB}(i) - 1, \end{aligned} \qquad (17)$$

where j_i is the index of the first frequency bin in the critical frequency band i and $M_{CB}(i)$ is the number of frequency bins in that critical frequency band.

[0082] The smoothed scaling gains $g_{BIN,LP}(k)$ and $g_{CB,LP}(i)$ are initially set to 1.0. Each time a non-tonal sound frame is processed (music_flag = 0), the value of the smoothed scaling gains are reset to 1.0 to reduce a possible reduction of these smoothed scaling gains in the next frame.

[0083] In every spectral analysis performed by the spectral analyser 105, the smoothed scaling gains $g_{CB,LP}(i)$ are updated for all critical frequency bands (even for voiced critical frequency bands processed through per bin processing - in this case $g_{CB,LP}(i)$ is updated with an average of $g_{BIN,LP}(k)$ belonging to the critical frequency band i). Similarly, the smoothed scaling gains $g_{BIN,LP}(k)$ are updated for all frequency bins in the first 17 critical frequency bands, that is up to frequency bin 115 in the case of narrowband coding (the first 21 critical frequency bands, that is up to frequency bin 250 in the case of wideband coding). For critical frequency bands processed with per band processing, the scaling gains are updated by setting them equal to $g_{CB,LP}(i)$ in the first 17 (narrowband coding) or 21 (wideband coding) critical frequency bands.

[0084] In the case of a low-energy decoded tonal sound signal, inter-tone noise reduction is not performed. A low-energy sound signal is detected by finding the maximum noise energy in all the critical frequency bands, max($N_{CB}(i)$), i = 0,...,17, (17 in the case of narrowband coding and 21 in the case of wideband coding) and if this value is lower than or equal to a certain value, for example 15 dB, then no inter-tone noise reduction is performed.

[0085] In the case of processing of narrowband signals, the inter-tone noise reduction is performed on the first 17 critical frequency bands (up to 3680 Hz). For the remaining 11 frequency bins between 3680 Hz and 4000 Hz, the spectrum is scaled using the last scaling gain g_s of the frequency bin corresponding to 3680 Hz.

Spectral gain correction

[0086] The Parseval theorem shows that the energy in the time domain is equal to the energy in the frequency domain. Reduction of the energy of the inter-tone noise results in an overall reduction of energy in the frequency and time domains.

An additional feature is that the reducer 108 of quantization noise comprises a per band gain corrector 306 to rescale the energy per critical frequency band in such a manner that the energy in each critical frequency band at the end of the resealing will be close to the energy before the inter-tone noise reduction.

[0087] To achieve such rescaling, it is not necessary to rescale all the frequency bins but to rescale only the most energetic bins. The per band gain corrector 306 comprises an analyser 401 (Figure 4) which identifies the most energetic bins prior to inter-tone noise reduction as the bins scaled by a scaling gain between]0.8, 1.0] in the inter-tone noise reduction phase. According to an alternative, the analyser 401 may also determine the per bin energy prior to inter-tone noise reduction using, for example, Equation (5) in order to identify the most energetic bins.

[0088] The energy removed from inter-tone noise will be moved to the most energetic events (corresponding to the most energetic bins) of the critical frequency band. In this manner, the final music sample will sound clearer than just doing a simple inter-tone noise reduction because the dynamic between energetic events and the noise floor will further increase.

[0089] The spectral energy of a critical frequency band after the inter-tone noise reduction is computed in the same manner as the spectral energy before the inter-tone noise reduction:

$$E_{CB}(i) = \frac{1}{(L_{FFT}/2)^2 M_{CB}(i)} \sum_{k=0}^{M_{CB}(i)-1} \left(X_R^2(k+j_i) + X_I^2(k+j_i) \right), \quad i = 0,...,16 \tag{18}$$

[0090] In this respect, the per band gain corrector 306 comprises an analyser 402 to determine the per band spectral energy prior to inter-tone noise reduction using Equation (18), and an analyser 403 to determine the per band spectral energy after the inter-tone noise reduction using Equation (18).

[0091] The per band gain corrector 306 further comprises a calculator 404 to determine a corrective gain as the ratio of the spectral energy of a critical frequency band before inter-tone noise reduction and the spectral energy of this critical frequency band after inter-tone noise reduction has been applied.

$$G_{corr}(i) = \sqrt{\left(E_{CB}(i) / E_{CB}'(i) \right)}, \qquad i = 0,...,16 \tag{19}$$

where E_{CB} is the critical band spectral energy before inter-tone noise reduction and E_{CB}' is the critical frequency band spectral energy after inter-tone noise reduction, The total number of critical frequency bands covers the entire spectrum from 17 bands in Narrowband coding to 21 bands in Wideband coding.

[0092] The rescaling along the critical frequency band *i* can be performed as follows:

$$\begin{aligned} &IF\left(g_{BIN,LP}(k+j_i) > 0.8 \;\&\; i > 4\right) \\ &\quad X_R''(k+j_i) = G_{corr}(k+j_i)\, X_R'(k+j_i), \text{ and} \\ &\quad X_I''(k+j_i) = G_{corr}(k+j_i)\, X_I'(k+j_i), \qquad k = 0,...,M_{CB}(i)-1, \end{aligned} \tag{20}$$

ELSE

$$X_R''(k+j_i) = X_R'(k+j_i),$$

and

$$X_I''(k+j_i) = X_I'(k+j_i), \qquad k = 0,...,M_{CB}(i)-1$$

where j_i is the index of the first frequency bin in the critical frequency band *i* and $M_{CB}(i)$ is the number of frequency bins in that critical frequency band. No gain correction is applied under 600 Hz because it is assumed that spectral energy at very low frequency has been accurately coded by the low bit rate speech-specific codec and any increase of inter-

harmonic tone will be audible.

Spectral gain boost

[0093] It is possible to further increase the clearness of a musical sample by increasing furthermore the gain G_{corr} in critical frequency bands where not many energetic events occur. A calculator 405 of the per band gain corrector 306 determines the ratio of energetic events (ratio of the number of energetic bins on total number of frequency bins) per critical frequency band as follow:

$$REv_{CB} = \frac{NumBin_{\max}}{NumBin_{total}} \quad k = 0,...,M_{CB}(i-1)$$

$$NumBin_{\max} = \sum \left(g_{BIN,LP} > 0.8 \right)$$

$$NumBin_{total} = Total\ bin\ in\ a\ critical\ band$$

[0094] The calculator 405 then computes an additional correction factor to the corrective gain using the following formula:

$$IF\left(NumBin_{\max} > 0\right)$$

$$C_F = -0.2778 \cdot REv_{CB} + 1.2778$$

[0095] In a per band gain corrector 406, this new correction factor C_F multiplies the corrective gain G_{corr} by a value situated between [1.0, 1.2778]. When this correction factor C_F is taken into consideration, the rescaling along the critical frequency band *i* becomes:

$$IF\left(g_{BIN,LP}(k+j_i) > 0.8\ \&\ i > 4\right)$$

$$X_R''(k+j_i) = G_{corr} \cdot \mathrm{C_F} \cdot (k+j_i)\, X_R'(k+j_i),$$

and

$$X_I''(k+j_i) = G_{corr} \cdot \mathrm{C_F} \cdot (k+j_i)\, X_I'(k+j_i), \qquad k = 0,...,M_{CB}(i)-1$$

ELSE

$$X_R''(k+j_i) = X_R'(k+j_i),$$

and

$$X_I''(k+j_i) = X_I'(k+j_i), \qquad k = 0,...,M_{CB}(i)-1$$

[0096] In the particular case of Wideband coding, the rescaling is performed only in the frequency bins previously scaled by a scaling gain between] 0.96, 1.0] in the inter-tone noise reduction phase. Usually, higher the bit rate is closer will be the energy of the spectrum to the desired energy level. For that reason the second part of the gain correction, the gain correction factor C_F, might not be always used. Finally, at very high bit rate, it could be benefical to perform gain resealing only in the frequency bins which were previously not modified (having a scaling gain of 1.0).

Reconstruction of enhanced, denoised sound signal

[0097] After determining the scaled spectral components 308, $X'_R(k)$ of $X_R''(k)$ and $X'_I(k)$ or $X_I''(k)$, a calculator 307 of the inverse analyser and overlap add operator 110 computes the inverse FFT. The calculated inverse FFT is applied to the scaled spectral components 308 to obtain a windowed enhanced decoded sound signal in the time domain given by the following relation:

$$x_{w,d}(n) = \frac{1}{N}\sum_{k=0}^{N-1} X(k) e^{j2\pi\frac{kn}{N}}, \qquad n = 0,\ldots,L_{FFT}-1 \tag{21}$$

[0098] The signal is then reconstructed in operator 303 using an overlap add operation for the overlapping portions of the analysis. Since a sine window is used on the original decoded tonal sound signal 103 prior to spectral analysis in the spectral analyser 105, the same windowing is applied to the windowed enhanced decoded tonal sound signal 309 at the output of the inverse FFT calculator prior to the overlap add operation. Thus, the doubled windowed enhanced decoded tonal sound signal is given by the relation:

$$x_{ww,d}^{(1)}(n) = w_{FFT}(n) x_{w,d}^{(1)}(n), \qquad n = 0,\ldots,L_{FFT}-1 \tag{22}$$

[0099] For the first third of the Narrowband analysis window, the overlap add operation for constructing the enhanced sound signal is performed using the relation:

$$s(n) = x_{ww,d}^{(0)}\left(n + 2\cdot {}^{L_{window}}\!/_{3}\right) + x_{ww,d}^{(1)}(n), \qquad n = 0,\ldots,L_{window}/3-1 \tag{23}$$

and for the first ninth of the Wideband analysis window, the overlap-add operation for constructing the enhanced decoded tonal sound signal is performed as follows:

$$s(n) = x_{ww,d}^{(0)}\left(n + 2\cdot {}^{L_{window_{WB}}}\!/_{9}\right) + x_{ww,d}^{(1)}(n), \qquad n = 0,\ldots,L_{window_{WB}}/9-1$$

where $x_{ww,d}^{(0)}(n)$ is the double windowed enhanced decoded tonal sound signal from the analysis of the previous frame.

[0100] Using an overlap add operation, since there is a 80 sample shift (40 in the case of Wideband coding) between the sound signal decoder frame and inter-tone noise reduction frame, the enhanced decoded tonal sound signal can be reconstructed up to 80 samples from the lookahead in addition to the present inter-tone noise reduction frame.

[0101] After the overlap add operation to reconstruct the enhanced decoded tonal sound signal, deemphasis is performed in the postprocessor 112 on the enhanced decoded sound signal using the inverse of the above described preemphasis filter. The postprocessor 112 therefore comprises a deemphasis filter which, in this embodiment, is given by the relation:

$$H_{\text{de-emph}}(z) = 1/(1-0.68z^{-1}) \tag{24}$$

Inter-tone noise energy update

[0102] Inter-tone noise energy estimates per critical frequency band for inter-tone noise reduction can be calculated for each frame in an inter-tone noise energy estimator (not shown), using for example the following formula:

$$N_{CB}^{0}(i)=\frac{\left(0.6\cdot E_{CB}^{0}(i)+0.2\cdot E_{CB}^{1}(i)+0.2\cdot N_{CB}^{1}(i)\right)}{16.0},\qquad i=0,...,16 \qquad (25)$$

where N_{CB}^{0} and E_{CB}^{0} represent the current noise and spectral energies for the specified critical frequency band *(i)* and N_{CB}^{1} and E_{CB}^{1} represent the noise and the spectral energies for the past frame of the same critical frequency band.

[0103] This method of calculating inter-tone noise energy estimates per critical frequency band is simple and could introduce some distortions in the enhanced decoded tonal sound signal. However, in low bit rate Narrowband coding, these distortions are largely compensated by the improvement in the clarity of the synthesis sound signals.

[0104] In wideband coding, when the inter-tone noise is present but less annoying, the method to update the inter-tone noise energy have to be more sophisticated to prevent the introduction of annoying distortion. Different technique could be use with more or less computational complexity.

Inter-tone noise energy update using weighted average per band energy:

[0105] In accordance with this technique, the second maximum and the minimum energy values of each critical frequency band are used to compute an energy threshold per critical frequency band as follow:

$$thr_ener_{CB}(i)=1.85\cdot\left(\frac{\max_2\left(E_{CB}^{0}(i)\right)+\min\left(E_{CB}^{0}(i)\right)}{2}\right),\qquad i=0,...,20$$

where max_2 represents the frequency bin having the second maximum energy value and *min* the frequency bin having the minimum energy value in the critical frequency band of concern.

[0106] The energy threshold (thr_ener_{CB}) is used to compute a first inter-tone noise level estimation per critical band (tmp_ener_{CB}) which corresponds to the mean of the energies (E_{BIN}) of all the frequency bins below the preceding energy threshold inside the critical frequency band, using the following relation:

$$\begin{aligned}
&mcnt=0\\
&tmp_ener_{CB}(i)=0\\
&for\ \left(k=0:M_{CB}(i)\right)\\
&\quad if\left(E_{BIN}(k)<thr_ener_{CB}\right)\\
&\quad\quad tmp_ener_{CB}(i)=tmp_ener_{CB}(i)+E_{BIN}(k)\\
&\quad\quad mcnt=mcnt+1\\
&\quad endif\\
&endfor\\
&tmp_ener_{CB}(i)=\frac{tmp_ener_{CB}(i)}{mcnt}
\end{aligned}$$

where *ment* is the number of frequency bins of which the energies (E_{BIN}) are included in the summation and $mcnt \leq M_{CB}(i)$. Furthermore; the number *mcnt* of frequency bins of which the energy (E_{BIN}) is below the energy threshold is compared to the number of frequency bins (M_{CB}) inside a critical frequency band to evaluate the ratio of frequency bins below the energy threshold. This ratio $accepted_ratio_{CB}$ is used to weight the first, previously found inter-tone noise level estimation (tmp_ener_{CB}).

$$accepted_ratio_{CB}(i) = \frac{mcnt}{M_{CB}(i)}, \qquad i = 0,...,20$$

[0107] A weighting factor β_{CB} of the inter-tone noise level estimation is different among the bit rate used and the *accepted_ratio*$_{CB}$. A high *accepted_ratio*$_{CB}$ for a critical frequency band means that it will be difficult to differentiate the noise energy from the signal energy. In that case it is desirable to not reduce too much the noise level or that critical frequency band to not risk any alteration of the signal energy. But a low *accepted_ratio*$_{CB}$ indicates a large difference between the noise and signal energy levels then the estimated noise level could be higher in that critical frequency band without adding distortion. The factor β_{CB} is modified as follow:

$$\begin{aligned}
&IF\left((accepted_ratio(i) < 0.6 \mid accepted_ratio(i-1) < 0.5) \;\&\; i > 9\right)\\
&\qquad \beta_{CB}(i) = 1\\
&ELSE\,IF\left(accepted_ratio(i) < 0.75 \;\&\; i > 15\right)\\
&\qquad \beta_{CB}(i) = 2\\
&ELSE\,IF\left(\begin{pmatrix} accepted_ratio(i) > 0.85 \quad \& \\ accepted_ratio(i-1) > 0.85 \& \\ accepted_ratio(i-2) > 0.85 \end{pmatrix} \& \; bitrate > 16000\right), \qquad i = 0,...,20\\
&\qquad \beta_{CB}(i) = 30\\
&ELSE\,IF\left(bitrate > 16000\right)\\
&\qquad \beta_{CB}(i) = 20\\
&ELSE\\
&\qquad \beta_{CB}(i) = 16
\end{aligned}$$

[0108] Finally the inter-tone noise estimation per critical frequency band can be smoothed differently if the inter-tone noise is increasing or decreasing.
Noise decreasing:

$$N^0_{CB}(i) = (1-\alpha)\left(\frac{tmp_ener_{CB}(i)}{\beta_{CB}(i)}\right) + \alpha \cdot N^1(i)$$

Noise increasing: *i* = 0,...,20

$$N^0_{CB}(i) = (1-\alpha_2)\left(\frac{tmp_ener_{CB}(i)}{\beta_{CB}(i)}\right) + \alpha_2 \cdot N^1(i)$$

Where

$$\alpha = 0.1$$

$$\alpha_2 = \begin{cases} 0.98 \;\; for \;\; bitrate > 16000\,bps \\ 0.95 \;\; otherwise \end{cases}$$

where N^0_{CB} represents the current noise energy for the specified critical frequency band *(i)* and N^1_{CB} represents the noise energy of the past frame of the same critical frequency band.

[0109] Although the present invention has been described in the foregoing description by way of non restrictive illustrative embodiments thereof, many other modifications and variations may be possible within the scope of the appended claims.

REFERENCES

[0110]

[1] 3GPP TS 26.190, "Adaptive Multi-Rate - Wideband (AMR-WB) speech codec; Transcoding functions".
[2] J. D. Johnston, "Transform coding of audio signal using perceptual noise criteria," IEEE J. Select. Arenas Commun., vol. 6, pp. 314-323, Feb. 1988.

Claims

1. A method (100) for enhancing a decoded tonal sound signal, comprising:

spectrally analysing (105) the decoded tonal sound signal to produce spectral parameters (107) representative of the decoded tonal sound signal, wherein spectrally analysing (105) the decoded tonal sound signal comprises dividing a spectrum resulting from the spectral analysis into a set of critical frequency bands each comprising a number of frequency bins;
reducing (108) a quantization noise in low-energy spectral regions of the decoded tonal sound signal in response to the spectral parameters (107) from the spectral analysis, wherein reducing (108) the quantization noise comprises scaling (108, 304, 305, 306) the spectrum of the decoded tonal sound signal per critical frequency band, per frequency bin or per both critical frequency band and frequency bin;

performing signal type classification comprising:

determining (501) (a) a mean $\overline{E}_{diff}$ of variations of a total frame spectral energy over past 40 frames of the decoded sound signal using the relation

$$\overline{E}_{diff} = \frac{\left(\sum_{i=t-40}^{t-1} \Delta_E^i\right)}{40}, \quad where \ \Delta_E^t = E_{fr}^t - E_{fr}^{(t-1)}$$

where E^t_{fr} is the total frame spectral energy for a current frame t and $E^{(t-1)}_{fr}$ is the total frame spectral energy for a previous frame (t-1), and (b) a statistical deviation σ_E of the energy variation over last 15 frames of the decoded sound signal using the relation

$$\sigma_E = 0.7745967 \cdot \sqrt{\sum_{i=t-15}^{t-1} \frac{\left(\Delta_E^i - \overline{E}_{diff}\right)^2}{15}}.$$

storing the mean $\overline{E}_{diff}$ and the statistical deviation σ_E in a memory (50);
comparing (503-506), by a first to fourth comparator, the statistical deviation σ_E to four floating thresholds including threshold 1, threshold 2, threshold 3 and threshold 4, to classify the decoded sound signal into sound signal category 0, sound signal category 1, sound signal category 2, sound signal category 3, and sound signal category 4;
counting (512), by a first counter, frames of sound signal category 3 or 4 and increasing (514) the floating thresholds 1 to 4 by a value *TH_UP* when a series of more than 30 frames of sound signal category 3 or 4 is counted by the first counted; and
counting (513), by a second counter, frames of sound signal category 0, and decreasing (514) the floating thresholds 1 to 4 by a value *TH_DOWN* when a series of more than 30 frames of sound signal category 0 is counted by the second counted, wherein thresholds 1 to 4 are limited to absolute maximum and minimum values and wherein each time the count of the first counter is incremented, the second counter is reset to zero;

characterized in that the signal type classification comprises:

- controlling (510), by a first controller, the quantization noise reduction (108) to enhance the decoded tonal sound signal within a frequency band 2000 to $F_s/2$ Hz by reducing inter-tone quantization noise by a maximum allowed amplitude of 6 dB, when (a) sound signal category 1 is detected by the first comparator (506) showing a statistical deviation σ_E lower than threshold 1 and (b) the last detected sound signal category was ≥ 0, wherein F_s is a sampling frequency of the decoded sound signal;
- controlling (509), by a second controller, the quantization noise reduction (108) to enhance the decoded tonal sound signal within a frequency band 1270 to $F_s/2$ Hz by reducing the inter-tone quantization noise by a maximum allowed amplitude of 9 dB, when (a) sound signal category 2 is detected by the second comparator (505) showing a statistical deviation σ_E lower than threshold 2 and (b) the last detected sound signal category was ≥ 1;
- controlling (508), by a third controller, the quantization noise reduction (108) to enhance the decoded tonal sound signal within a frequency band 700 to $F_s/2$ Hz by reducing the inter-tone quantization noise by a maximum allowed amplitude of 12 dB, when (a) sound signal category 3 is detected by the third comparator (504) showing a statistical deviation σ_E lower than threshold 3 and (b) the last detected sound signal category was ≥ 2;
- controlling (507), by a fourth controller, the quantization noise reduction (108) to enhance the decoded tonal sound signal within a frequency band 400 to $F_s/2$ Hz by reducing the inter-tone quantization noise by a maximum allowed amplitude of 12 dB, when (a) sound signal category 4 is detected by the fourth comparator (503) showing a statistical deviation σ_E lower than threshold 4 and (b) the last detected sound signal category was ≥ 3; and
- controlling (511), by a fifth controller, the quantization noise reduction (108) not to reduce inter-tone quantization noise when sound signal category 0 is detected, when detection of sound signal categories 1 to 4 by the first to fourth comparator is negative.

2. A system (100) for enhancing a decoded tonal sound signal, comprising:

a spectral analyser (105) of the decoded tonal sound signal adapted to produce spectral parameters (107) representative of the decoded tonal sound signal, wherein the spectral analyser (105) is adapted to divide a spectrum resulting from spectral analysis into a set of critical frequency bands, and wherein each critical frequency band comprises a number of frequency bins;

a reducer (108) of quantization noise in low-energy spectral regions of the decoded tonal sound signal using the spectral parameters (107) from the spectral analyser (105), wherein the reducer (108) of quantization noise comprises a noise attenuator (108, 304, 305, 306) that is adapted to scale the spectrum of the decoded tonal sound signal per critical frequency band, per frequency bin or per both critical frequency band and frequency bin; and

a signal type classifier (301) comprising:

- a finder (501) for determining (a) a mean $\overline{E}_{diff}$ of variations of a total frame spectral energy over past 40 frames of the decoded sound signal using the relation

$$\overline{E}_{diff} = \frac{\left(\sum_{t=-40}^{t=-1} \Delta_E^t\right)}{40}, \text{ where } \Delta_E^t = E_{fr}^t - E_{fr}^{(t-1)}$$

where E^t_{fr} is the total frame spectral energy for a current frame t and $E^{(t-1)}_{fr}$ is the total frame spectral energy for a previous frame (t-1), and (b) a statistical deviation σ_E of the energy variation over last 15 frames of the decoded sound

$$\sigma_E = 0.7745967 \cdot \sqrt{\sum_{t=-15}^{t=-1} \frac{\left(\Delta_E^t - \overline{E}_{diff}\right)^2}{15}}$$

- a memory (502) adapted to be updated with the mean $\overline{E}_{diff}$ and the statistical deviation σ_E;
- first, second, third and fourth comparators (503-506) for comparing the statistical deviation σ_E to four floating thresholds including threshold 1, threshold 2, threshold 3 and threshold 4, to classify the decoded sound signal into sound signal category 0, sound signal category 1, sound signal category 2, sound signal category 3, and sound signal category 4;

- a first counter (512) of frames of sound signal category 3 or 4 and a threshold controller (514) adapted to increase the floating thresholds 1 to 4 by a value *TH_UP* when a series of more than 30 frames of sound signal category 3 or 4 is counted by the first counter, and
- a second counter (513) of frames of sound signal category 0, the threshold controller (514) being adapted to decrease the floating thresholds 1 to 4 by a value *TH_DOWN* when a series of more than 30 frames of sound signal category 0 is counted by the second counter, wherein thresholds 1 to 4 are limited to absolute maximum and minimum values and wherein each time the count of the first counter is incremented, the second counter is reset to zero;

characterized in that the signal type classifier comprises:

- a first controller (510) for instructing the reducer of quantization noise (108) to enhance the decoded tonal sound signal within a frequency band 2000 to $F_s/2$ Hz by reducing inter-tone quantization noise by a maximum allowed amplitude of 6 dB, when (a) the first comparator (506) detects sound signal category 1 by detecting a statistical deviation σ_E lower than threshold 1 and (b) the last detected sound signal category was ≥ 0, wherein F_s is a sampling frequency of the decoded sound signal;
- a second controller (509) for instructing the reducer of quantization noise (108) to enhance the decoded tonal sound signal within a frequency band 1270 to $F_s/2$ Hz by reducing the inter-tone quantization noise by a maximum allowed amplitude of 9 dB, when (a) the second comparator (505) detects sound signal category 2 by detecting a statistical deviation σ_E lower than threshold 2 and (b) the last detected sound signal category was ≥ 1;
- a third controller (508) for instructing the reducer of quantization noise (108) to enhance the decoded tonal sound signal within a frequency band 700 to $F_s/2$ Hz by reducing the inter-tone quantization noise by a maximum allowed amplitude of 12 dB, when (a) the third comparator (504) detects sound signal category 3 by detecting a statistical deviation σ_E lower than threshold 3 and (b) the last detected sound signal category was ≥ 2;
- a fourth controller (507) for instructing the reducer of quantization noise (108) to enhance the decoded tonal sound signal within a frequency band 400 to $F_s/2$ Hz by reducing the inter-tone quantization noise by a maximum allowed amplitude of 12 dB, when (a) the fourth comparator (503) detects sound signal category 4 by detecting a statistical deviation σ_E lower than threshold 4 and (b) the last detected sound signal category was ≥ 3; and
- a fifth controller (511) for instructing the reducer of quantization noise (108) not to reduce inter-tone quantization noise when sound signal category 0 is detected, when detection of sound signal categories 1 to 4 by the first to fourth comparators is negative.

Patentansprüche

1. Verfahren (100) zum Verbessern eines decodierten Klangsignals, umfassend:

spektrales Analysieren (105) des decodierten Klangsignals zum Erzeugen von spektralen Parametern (107), die repräsentativ für das decodierte Klangsignal sind, wobei das spektrale Analysieren (105) des decodierten Klangsignals Aufteilen eines Spektrums, das aus der Spektralanalyse resultiert, in einen Satz von kritischen Frequenzbändern umfasst, die jeweils eine Anzahl von Frequenzabschnitten umfassen;
Reduzieren (108) eines Quantifizierungsrauschens in niederenergetischen Spektralbereichen des decodierten Klangsignals als Reaktion auf die spektralen Parameter (107) aus der Spektralanalyse, wobei das Reduzieren (108) des Quantifizierungsrauschens Skalieren (108, 304, 305, 306) des Spektrums des decodierten Klangsignals pro kritischem Frequenzband, pro Frequenzabschnitt oder sowohl pro kritischem Frequenzband als auch Frequenzabschnitt umfasst;

Ausführen der Signaltypklassifikation, umfassend:

Bestimmen (501) (a) eines Mittelwertes $\overline{E}_{diff}$ von Variationen einer spektralen Gesamtrahmenenergie über die vorherigen 40 Rahmen des decodierten Klangsignals unter Verwendung der Gleichung

$$\overline{E}_{diff} = \frac{\sum_{t=-40}^{t=-1} \Delta_E^t}{40}, \quad \text{wobei} \quad \Delta_E^t = E_{fr}^t - E_{fr}^{(t-1)}$$

wobei E^t_{fr} die spektrale Gesamtrahmenenergie für einen aktuellen Rahmen t ist, und $E^{(t-1)}_{fr}$ die spektrale Gesamtrahmenenergie für einen vorherigen Rahmen (t-1) ist, und (b) einer statistischen Abweichung σ_E der Energievariation über die letzten 15 Rahmen des decodierten Klangsignals unter Verwendung der Beziehung

$$\sigma_E = 0.7745967 \cdot \sqrt{\sum_{t=-15}^{t=-1} \frac{\left(\Delta_{fr}^{\ t} - \overline{E}_{diff}\right)^2}{15}}$$

Speichern des Mittelwertes $\overline{E}_{diff}$ und der statistischen Abweichung σ_E in einem Speicher (50);

Vergleichen (503-506), durch einen ersten bis vierten Komparator, der statistischen Abweichung σ_E mit vier flexiblen Schwellenwerten, die Schwellenwert 1, Schwellenwert 2, Schwellenwert 3 und Schwellenwert 4 umfassen, um das decodierte Klangsignal in Klangsignalkategorie 0, Klangsignalkategorie 1, Klangsignalkategorie 2, Klangsignalkategorie 3 und Klangsignalkategorie 4 zu klassifizieren;

Zählen (512), durch einen ersten Zähler, von Rahmen der Klangsignalkategorie 3 oder 4 und Erhöhen (514) der flexiblen Schwellenwerte 1 bis 4 um einen Wert *TH_UP*, wenn eine Reihe von mehr als 30 Rahmen der Klangsignalkategorie 3 oder 4 vom ersten Zähler gezählt wird; und

Zählen (513), durch einen zweiten Zähler, von Rahmen der Klangsignalkategorie 0, und Verringern (514) der flexiblen Schwellenwerte 1 bis 4 um einen Wert *TH_DOWN*, wenn eine Reihe von mehr als 30 Rahmen der Klangsignalkategorie 0 vom zweiten Zähler gezählt wird, wobei die Schwellenwerte 1 bis 4 auf absolute Maximal- und Minimalwerte beschränkt sind, und wobei jedes Mal, wenn die Zählung des ersten Zählers erhöht wird, der zweite Zähler auf null zurückgesetzt wird;

dadurch gekennzeichnet, dass die Signaltypklassifikation umfasst:

- Steuern (510), durch einen ersten Controller, der Reduzierung des Quantifizierungsrauschens (108), um das decodierte Klangsignal innerhalb eines Frequenzbandes von 2000 bis $F_s/2$ Hz durch Reduzieren des Quantifizierungsrauschens zwischen den Tönen um eine maximal zulässige Amplitude von 6 dB zu verstärken, wenn (a) die Klangsignalkategorie 1 durch den ersten Komparator (506) festgestellt wird, die eine statistische Abweichung σ_E zeigt, die kleiner als der Schwellenwert 1 ist, und (b) die letzte festgestellte Klangsignalkategorie ≥0 war, wobei F_s eine Abtastfrequenz des decodierten Klangsignals ist;
- Steuern (509), durch einen zweiten Controller, der Reduzierung des Quantifizierungsrauschens (108), um das decodierte Klangsignal innerhalb eines Frequenzbandes von 1270 bis $F_s/2$ Hz durch Reduzieren des Quantifizierungsrauschens zwischen den Tönen um eine maximal zulässige Amplitude von 9 dB zu verstärken, wenn (a) die Klangsignalkategorie 2 durch den zweiten Komparator (505) festgestellt wird, die eine statistische Abweichung σ_E zeigt, die kleiner als Schwellenwert 2 ist, und (b) die letzte festgestellte Klangsignalkategorie ≥1 war;
- Steuern (508), durch einen dritten Controller, der Reduzierung des Quantifizierungsrauschens (108), um das decodierte Klangsignal innerhalb eines Frequenzbandes von 700 bis $F_s/2$ Hz durch Reduzieren des Quantifizierungsrauschens zwischen den Tönen um eine maximal zulässige Amplitude von 12 dB zu verstärken, wenn (a) die Klangsignalkategorie 3 durch den dritten Komparator (504) festgestellt wird, die eine statistische Abweichung σ_E zeigt, die kleiner als Schwellenwert 3 ist, und (b) die letzte festgestellte Klangsignalkategorie ≥2 war;
- Steuern (507), durch einen vierten Controller, der Reduzierung des Quantifizierungsrauschens (108), um das decodierte Klangsignal innerhalb eines Frequenzbandes von 400 bis $F_s/2$ Hz durch Reduzieren des Quantifizierungsrauschens zwischen den Tönen um eine maximal zulässige Amplitude von 12 dB zu verstärken, wenn (a) die Klangsignalkategorie 4 durch den vierten Komparator (503) festgestellt wird, die eine statistische Abweichung σ_E zeigt, die kleiner als Schwellenwert 4 ist, und (b) die letzte festgestellte Klangsignalkategorie ≥3 war; und
- Steuern (511), durch einen fünften Controller, der Reduzierung des Quantifizierungsrauschens (108), um das Quantifizierungsrauschen zwischen den Tönen nicht zu reduzieren, wenn die Klangsignalkategorie 0 festgestellt wird, wenn die Feststellung von Klangsignalkategorien 1 bis 4 durch den ersten bis vierten Komparator negativ ist.

2. System (100) zum Verstärken eines decodierten Klangsignals, umfassend:

einen Spektralanalysator (105) des decodierten Klangsignals, der dafür ausgelegt ist, spektrale Parameter (107) zu erzeugen, die repräsentativ für das decodierte Klangsignal sind, wobei der Spektralanalysator (105) dafür ausgelegt ist, ein Spektrum, das aus der Spektralanalyse resultiert, in einen Satz von kritischen Frequenzbändern aufzuteilen, und wobei jedes kritische Frequenzband eine Anzahl von Frequenzabschnitten umfasst;
einen Abschwächer (108) des Quantifizierungsrauschens in niederenergetischen Spektralbereichen des decodierten Klangsignals unter Verwendung der spektralen Parameter (107) aus dem Spektralanalysator (105), wobei der Abschwächer (108) des Quantifizierungsrauschens einen Rauschdämpfer (108, 304, 305, 306) umfasst, der dafür ausgelegt ist, das Spektrum des decodierten Klangsignals pro kritischem Frequenzband, pro kritischem Frequenzabschnitt oder pro sowohl kritischem Frequenzband als auch Frequenzabschnitt zu skalieren; und
einen Signaltypklassifikator (301), umfassend:

- einen Sucher (501) zum Bestimmen (a) eines Mittelwertes $\overline{E}_{diff}$ von Variationen einer spektralen Gesamtrahmenenergie über die vorherigen 40 Rahmen des decodierten Klangsignals unter Verwendung der Beziehung

$$\overline{E}_{diff} = \frac{\sum_{t=-40}^{t=-1} \Delta_E^t}{40}, \quad \text{wobei} \quad \Delta_E^t = E_{fr}^t - E_{fr}^{(t-1)}$$

wobei E^t_{fr} die spektrale Gesamtrahmenenergie für einen aktuellen Rahmen t ist, und $E^{(t-1)}_{fr}$ die spektrale Gesamtrahmenenergie für einen vorherigen Rahmen (t-1) ist, und (b) einer statistischen Abweichung σ_E der Energievariation über die letzten 15 Rahmen des decodierten Klangsignals unter Verwendung der Beziehung

$$\sigma_E = 0.7745967 \cdot \sqrt{\sum_{t=-15}^{t=-1} \frac{\left(\Delta_{fr}^t - \overline{E}_{diff}\right)^2}{15}}$$

- einen Speicher (502), der dafür ausgelegt ist, mit dem Mittelwert $\overline{E}_{diff}$ und der statistischen Abweichung σ_E aktualisiert zu werden;
- erste, zweite, dritte und vierte Komparatoren (503 - 506) zum Vergleichen der statistischen Abweichung σ_E mit vier flexiblen Schwellenwerten, die Schwellenwert 1, Schwellenwert 2, Schwellenwert 3 und Schwellenwert 4 umfassen, um das decodierte Klangsignal in Klangsignalkategorie 0, Klangsignalkategorie 1, Klangsignalkategorie 2, Klangsignalkategorie 3 und Klangsignalkategorie 4 zu klassifizieren;
- einen ersten Zähler (512) von Rahmen der Klangsignalkategorie 3 oder 4 und einen Schwellenwertcontroller (514), der dafür ausgelegt ist, die flexiblen Schwellenwerte 1 bis 4 um einen Wert *TH_UP* zu erhöhen, wenn eine Reihe von mehr als 30 Rahmen der Klangsignalkategorie 3 oder 4 vom ersten Zähler gezählt wird; und
- einen zweiten Zähler (513) von Rahmen der Klangsignalkategorie 0, wobei der Schwellenwertcontroller (514) dafür ausgelegt ist, die flexiblen Schwellenwerte 1 bis 4 um einen Wert *TH_DOWN* zu verringern, wenn eine Reihe von mehr als 30 Rahmen der Klangsignalkategorie 0 vom zweiten Zähler gezählt wird,

wobei die Schwellenwerte 1 bis 4 auf absolute Maximal- und Minimalwerte beschränkt sind und wobei jedes Mal, wenn die Zählung des ersten Zählers erhöht wird, der zweite Zähler auf null zurückgesetzt wird;

dadurch gekennzeichnet, dass der Signaltypklassifikator umfasst:

- einen ersten Controller (510) zum Instruieren des Abschwächers des Quantifizierungsrauschens (108), das decodierte Klangsignal innerhalb eines Frequenzbandes von 2000 bis $F_s/2$ Hz durch Reduzieren des Quantifizierungsrauschens zwischen den Tönen um eine maximal zulässige Amplitude von 6 dB zu verstärken, wenn (a) der erste Komparator (506) die Klangsignalkategorie 1 durch Feststellen einer statistischen Abweichung σ_E feststellt, die niedriger als Schwellenwert 1 ist, und (b) die letzte festgestellte Klangsignalkategorie ≥ 0 war, wobei F_s eine Abtastfrequenz des decodierten Klangsignals ist;
- einen zweiten Controller (509) zum Instruieren des Abschwächers des Quantifizierungsrauschens (108), das decodierte Klangsignal innerhalb eines Frequenzbandes von 1270 bis $F_s/2$ Hz durch Reduzieren des Quanti-

fizierungsrauschens zwischen den Tönen um eine maximal zulässige Amplitude von 9 dB zu verstärken, wenn (a) der zweite Komparator (505) die Klangsignalkategorie 2 durch Feststellen einer statistischen Abweichung σ_E feststellt, die niedriger als Schwellenwert 2 ist, und (b) die letzte festgestellte Klangsignalkategorie ≥1 war;
- einen dritten Controller (508) zum Instruieren des Abschwächers des Quantifizierungsrauschens (108), das decodierte Klangsignal innerhalb eines Frequenzbandes von 700 bis $F_s/2$ Hz durch Reduzieren des Quantifizierungsrauschens zwischen den Tönen um eine maximal zulässige Amplitude von 12 dB zu verstärken, wenn (a) der dritte Komparator (504) die Klangsignalkategorie 3 durch Feststellen einer statistischen Abweichung σ_E feststellt, die niedriger als Schwellenwert 3 ist, und (b) die letzte festgestellte Klangsignalkategorie ≥2 war;
- einen vierten Controller (507) zum Instruieren des Abschwächers des Quantifizierungsrauschens (108), das decodierte Klangsignal innerhalb eines Frequenzbandes von 400 bis $F_s/2$ Hz durch Reduzieren des Quantifizierungsrauschens zwischen den Tönen um eine maximal zulässige Amplitude von 12 dB zu verstärken, wenn (a) der vierte Komparator (503) die Klangsignalkategorie 4 durch Feststellen einer statistischen Abweichung σ_E feststellt, die niedriger als Schwellenwert 4 ist, und (b) die letzte festgestellte Klangsignalkategorie ≥3 war; und
- einen fünften Controller (511) zum Instruieren des Abschwächers des Quantifizierungsrauschens (108), das Quantifizierungsrauschen zwischen den Tönen nicht zu reduzieren, wenn die Klangsignalkategorie 0 festgestellt wird, wenn die Feststellung von Klangsignalkategorien 1 bis 4 durch den ersten bis vierten Komparator negativ ist.

Revendications

1. Procédé (100) d'accentuation d'un signal de son tonal décodé, comportant les étapes consistant à :

analyser spectralement (105) le signal de son tonal décodé pour produire des paramètres spectraux (107) représentatifs du signal de son tonal décodé, l'analyse spectrale (105) du signal de son tonal décodé comportant la division d'un spectre résultant de l'analyse spectrale en un ensemble de bandes de fréquences critiques comportant chacune une multiplicité de canaux fréquentiels ;
réduire (108) une distorsion de quantification dans des régions spectrales à faible énergie du signal de son tonal décodé en réponse aux paramètres spectraux (107) issus de l'analyse spectrale, la réduction (108) de la distorsion de quantification comportant la mise à l'échelle (108, 304, 305, 306) du spectre du signal de son tonal décodé par bande de fréquence critique, par canal fréquentiel ou à la fois par bande de fréquence critique et par canal fréquentiel ;

effectuer une classification du type de signal comportant les étapes consistant à :

déterminer (501) (a) une moyenne $\overline{E}_{diff}$ de variations d'une énergie spectrale totale de trame sur 40 dernières trames du signal de son décodé à l'aide de la relation

$$\overline{E}_{diff} = \frac{\sum_{t=-40}^{t=-1} \Delta_E^t}{40}, \quad \text{où} \quad \Delta_E^t = E_{fr}^t - E_{fr}^{(t-1)}$$

où E_{fr}^t est l'énergie spectrale totale de trame pour une trame actuelle t et $E_{fr}^{(t-1)}$ est l'énergie spectrale totale de trame pour une trame précédente (t-1), et (b) un écart statistique σ_E de la variation d'énergie sur 15 dernières trames du signal de son décodé à l'aide de la relation

$$\sigma_E = 0.7745967 \cdot \sqrt{\sum_{t=-15}^{t=-1} \frac{\left(\Delta_{fr}^t - \overline{E}_{diff}\right)^2}{15}}$$

conserver la moyenne $\overline{E}_{diff}$ et l'écart statistique σ_E dans une mémoire (50) ;
faire comparer (503-506), par un premier à quatrième comparateur, l'écart statistique σ_E à quatre seuils flottants comprenant le seuil 1, le seuil 2, le seuil 3 et le seuil 4, pour classifier le signal de son décodé en une catégorie 0 de signaux de son, une catégorie 1 de signaux de son, une catégorie 2 de signaux de son, une catégorie 3 de signaux de son, et une catégorie 4 de signaux de son ;
faire compter (512), par un premier compteur, des trames de catégorie 3 ou 4 de signaux de son et augmenter

(514) les seuils flottants 1 à 4 d'une valeur TH_UP lorsqu'une série de plus de 30 trames de catégorie 3 ou 4 de signaux de son est comptée par le premier compteur ; et
faire compter (513), par un deuxième compteur, des trames de catégorie 0 de signaux de son, et diminuer (514) les seuils flottants 1 à 4 d'une valeur TH_DOWN lorsqu'une série de plus de 30 trames de catégorie 0 de signaux de son est comptée par le deuxième compteur, les seuils 1 à 4 étant limités à des valeurs maximales et minimales absolues et le deuxième compteur étant réinitialisé à zéro chaque fois que le comptage du premier compteur est incrémenté ;

caractérisé en ce que la classification du type de signal comporte les étapes consistant à :

- faire commander (510), par une première unité de commande, la réduction (108) de distorsion de quantification pour accentuer le signal de son tonal décodé à l'intérieur d'une bande de fréquence de 2000 à $F_s/2$ Hz en réduisant une distorsion de quantification entre tonalités d'une amplitude maximale admise de 6 dB, lorsque (a) une catégorie 1 de signaux de son est détectée par le premier comparateur (506) indiquant un écart statistique σ_E inférieur au seuil 1 et (b) la dernière catégorie de signaux de son détectée était ≥0, F_S étant une fréquence d'échantillonnage du signal de son décodé ;
- faire commander (509), par une deuxième unité de commande, la réduction (108) de distorsion de quantification pour accentuer le signal de son tonal décodé à l'intérieur d'une bande de fréquence de 1270 à $F_s/2$ Hz en réduisant la distorsion de quantification entre tonalités d'une amplitude maximale admise de 9 dB, lorsque (a) une catégorie 2 de signaux de son est détectée par le deuxième comparateur (505) indiquant un écart statistique σ_E inférieur au seuil 2 et (b) la dernière catégorie de signaux de son détectée était ≥ 1 ;
- faire commander (508), par une troisième unité de commande, la réduction (108) de distorsion de quantification pour accentuer le signal de son tonal décodé à l'intérieur d'une bande de fréquence de 700 à $F_s/2$ Hz en réduisant la distorsion de quantification entre tonalités d'une amplitude maximale admise de 12 dB, lorsque (a) une catégorie 3 de signaux de son est détectée par le troisième comparateur (504) indiquant un écart statistique σ_E inférieur au seuil 3 et (b) la dernière catégorie de signaux de son détectée était ≥ 2 ;
- faire commander (507), par une quatrième unité de commande, la réduction (108) de distorsion de quantification pour accentuer le signal de son tonal décodé à l'intérieur d'une bande de fréquence de 400 à $F_s/2$ Hz en réduisant la distorsion de quantification entre tonalités d'une amplitude maximale admise de 12 dB, lorsque (a) une catégorie 4 de signaux de son est détectée par le quatrième comparateur (503) indiquant un écart statistique σ_E inférieur au seuil 4 et (b) la dernière catégorie de signaux de son détectée était ≥ 3 ; et
- faire commander (511), par une cinquième unité de commande, la réduction (108) de distorsion de quantification pour ne pas réduire la distorsion de quantification entre tonalités lorsqu'une catégorie 0 de signaux de son est détectée, lorsque la détection des catégories 1 à 4 de signaux de son par les premier à quatrième comparateurs est négative.

2. Système (100) d'accentuation d'un signal de son tonal décodé, comportant :

un analyseur spectral (105) du signal de son tonal décodé prévu pour produire des paramètres spectraux (107) représentatifs du signal de son tonal décodé, l'analyseur spectral (105) étant prévu pour diviser un spectre résultant d'une analyse spectrale en un ensemble de bandes de fréquences critiques, et chaque bande de fréquence critique comportant une multiplicité de canaux fréquentiels ;
un réducteur (108) de distorsion de quantification dans des régions spectrales à faible énergie du signal de son tonal décodé utilisant les paramètres spectraux (107) issus de l'analyseur spectral (105), le réducteur (108) de distorsion de quantification comportant un atténuateur (108, 304, 305, 306) de bruit qui est prévu pour mettre à l'échelle le spectre du signal de son tonal décodé par bande de fréquence critique, par canal fréquentiel ou à la fois par bande de fréquence critique et par canal fréquentiel ; et
un classificateur (301) de type de signal comportant :

- un moyen (501) de détermination servant à déterminer (a) une moyenne $\overline{E}_{diff}$ de variations d'une énergie spectrale totale de trame sur 40 dernières trames du signal de son décodé à l'aide de la relation

$$\overline{E}_{diff} = \frac{\sum_{t=-40}^{t=-1} \Delta_E^t}{40}, \quad \text{où} \quad \Delta_E^t = E_{fr}^t - E_{fr}^{(t-1)}$$

où E_{fr}^{t} est l'énergie spectrale totale de trame pour une trame actuelle t et $E_{fr}^{(t-1)}$ est l'énergie spectrale totale de trame pour une trame précédente (t-1), et (b) un écart statistique σ_E de la variation d'énergie sur 15 dernières trames du signal de son décodé à l'aide de la relation

$$\sigma_E = 0.7745967 \cdot \sqrt{\sum_{t=-15}^{t=-1} \frac{\left(\Delta_{fr}^{t} - \overline{E}_{diff}\right)^2}{15}}$$

- une mémoire (502) prévue pour être mise à jour avec la moyenne $\overline{E}_{\text{diff}}$ et l'écart statistique σ_E ;
- des premier, deuxième, troisième et quatrième comparateurs (503-506) servant à comparer l'écart statistique σ_E à quatre seuils flottants comprenant le seuil 1, le seuil 2, le seuil 3 et le seuil 4, pour classifier le signal de son décodé en une catégorie 0 de signaux de son, une catégorie 1 de signaux de son, une catégorie 2 de signaux de son, une catégorie 3 de signaux de son, et une catégorie 4 de signaux de son ;
- un premier compteur (512) de trames de catégorie 3 ou 4 de signaux de son et une unité (514) de commande de seuils prévue pour augmenter les seuils flottants 1 à 4 d'une valeur TH_UP lorsqu'une série de plus de 30 trames de catégorie 3 ou 4 de signaux de son est comptée par le premier compteur, et
- un deuxième compteur (513) de trames de catégorie 0 de signaux de son, l'unité (514) de commande de seuils étant prévue pour diminuer les seuils flottants 1 à 4 d'une valeur TH_DOWN lorsqu'une série de plus de 30 trames de catégorie 0 de signaux de son est comptée par le deuxième compteur, les seuils 1 à 4 étant limités à des valeurs maximales et minimales absolues et le deuxième compteur étant réinitialisé à zéro chaque fois que le comptage du premier compteur est incrémenté ;

caractérisé en ce que le classificateur de type de signal comporte :

- une première unité (510) de commande servant à donner comme consigne au réducteur (108) de distorsion de quantification d'accentuer le signal de son tonal décodé à l'intérieur d'une bande de fréquence de 2000 à $F_s/2$ Hz en réduisant une distorsion de quantification entre tonalités d'une amplitude maximale admise de 6 dB, lorsque (a) le premier comparateur (506) détecte une catégorie 1 de signaux de son en détectant un écart statistique σ_E inférieur au seuil 1 et (b) la dernière catégorie de signaux de son détectée était ≥ 0, F_S étant une fréquence d'échantillonnage du signal de son décodé ;
- une deuxième unité (509) de commande servant à donner comme consigne au réducteur (108) de distorsion de quantification d'accentuer le signal de son tonal décodé à l'intérieur d'une bande de fréquence de 1270 à $F_s/2$ Hz en réduisant la distorsion de quantification entre tonalités d'une amplitude maximale admise de 9 dB, lorsque (a) le deuxième comparateur (505) détecte une catégorie 2 de signaux de son en détectant un écart statistique σ_E inférieur au seuil 2 et (b) la dernière catégorie de signaux de son détectée était ≥ 1 ;
- une troisième unité (508) de commande servant à donner comme consigne au réducteur (108) de distorsion de quantification d'accentuer le signal de son tonal décodé à l'intérieur d'une bande de fréquence de 700 à Fs/2 Hz en réduisant la distorsion de quantification entre tonalités d'une amplitude maximale admise de 12 dB, lorsque (a) le troisième comparateur (504) détecte une catégorie 3 de signaux de son en détectant un écart statistique σ_E inférieur au seuil 3 et (b) la dernière catégorie de signaux de son détectée était ≥ 2 ;
- une quatrième unité (507) de commande servant à donner comme consigne au réducteur (108) de distorsion de quantification d'accentuer le signal de son tonal décodé à l'intérieur d'une bande de fréquence de 400 à $F_s/2$ Hz en réduisant la distorsion de quantification entre tonalités d'une amplitude maximale admise de 12 dB, lorsque (a) le quatrième comparateur (503) détecte une catégorie 4 de signaux de son en détectant un écart statistique σ_E inférieur au seuil 4 et (b) la dernière catégorie de signaux de son détectée était ≥ 3 ; et
- une cinquième unité (511) de commande servant à donner comme consigne au réducteur (108) de distorsion de quantification de ne pas réduire la distorsion de quantification entre tonalités lorsqu'une catégorie 0 de signaux de son est détectée, lorsque la détection des catégories 1 à 4 de signaux de son par les premier à quatrième comparateurs est négative.

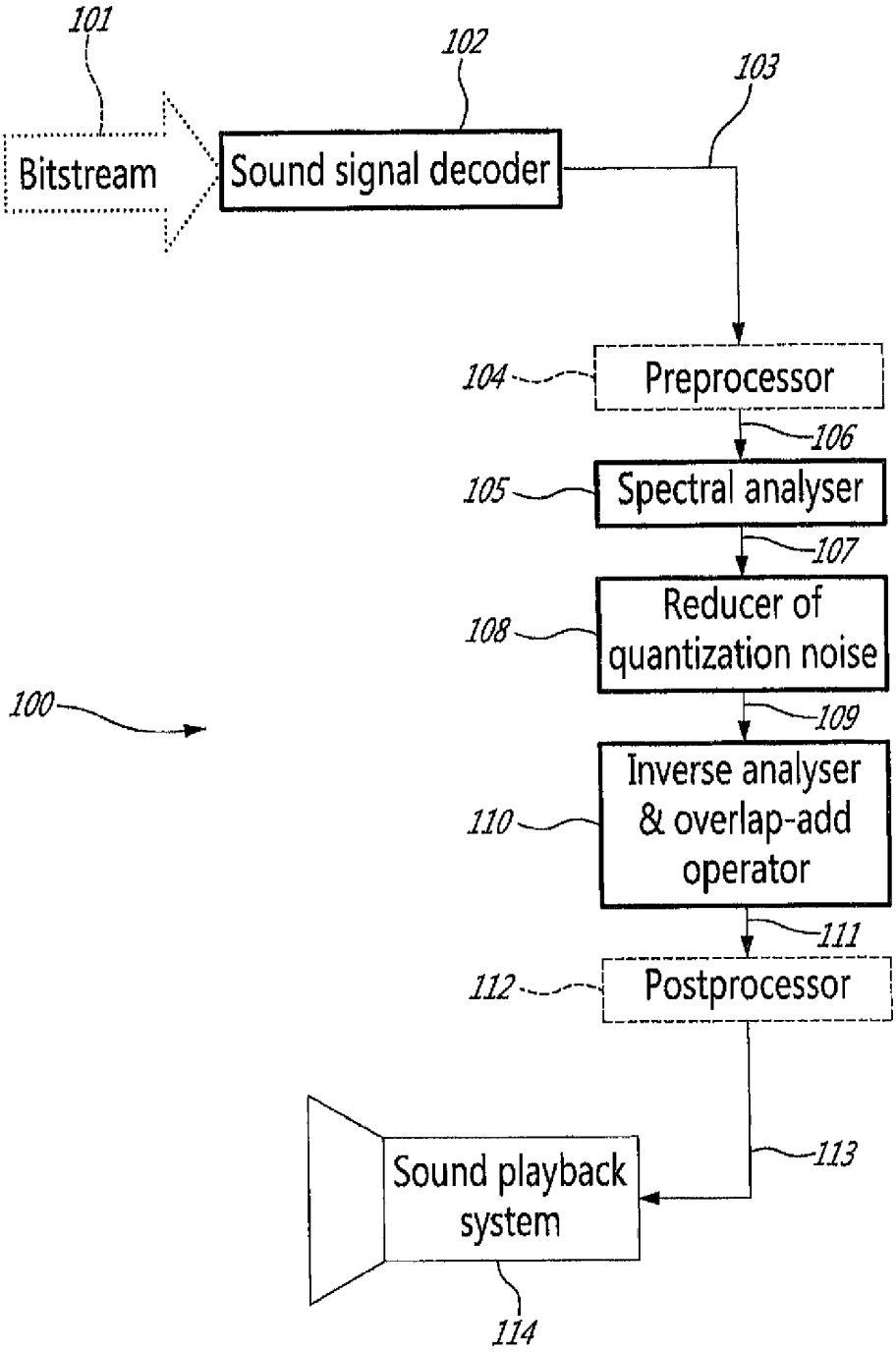
101
102
103
Bitstream
Sound signal decoder
104
Preprocessor
106
105
Spectral analyser
107
108
Reducer of quantization noise
100
109
Inverse analyser & overlap-add operator
110
111
112
Postprocessor
113
Sound playback system
114

Fig. 1

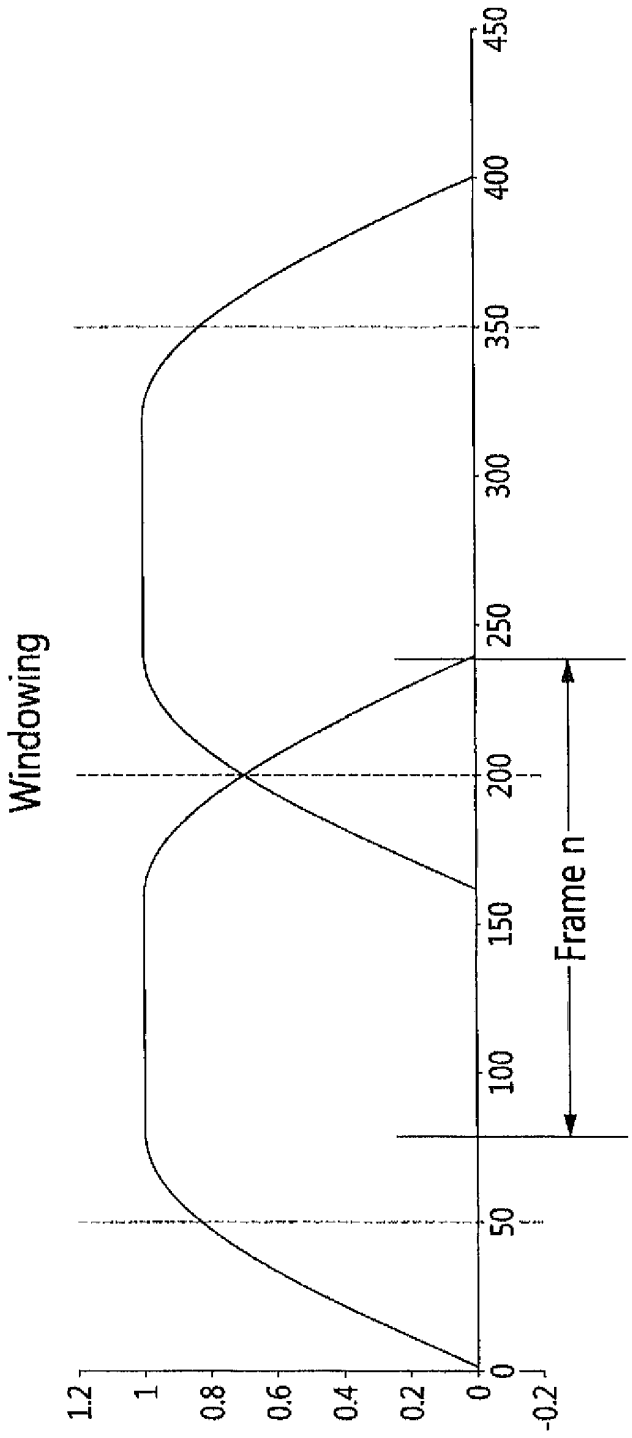
Windowing
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
0
50
100
150
200
250
300
350
400
450
Frame n

Fig. 7

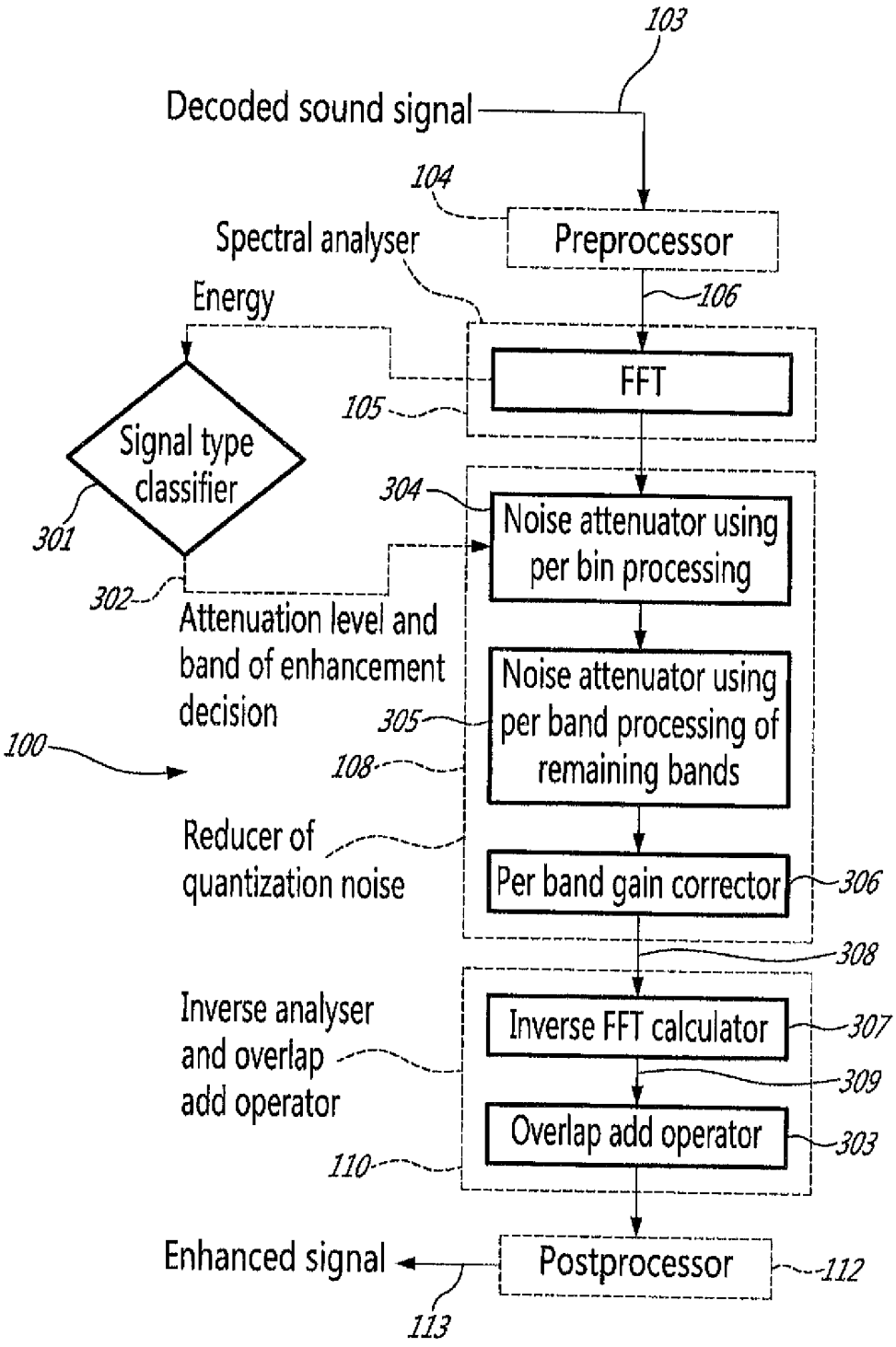
103
Decoded sound signal
104
Preprocessor
Spectral analyser
Energy
106
FFT
105
Signal type classifier
301
304
Noise attenuator using per bin processing
302
Attenuation level and band of enhancement decision
305
Noise attenuator using per band processing of remaining bands
100
108
Reducer of quantization noise
Per band gain corrector
306
308
Inverse analyser and overlap add operator
Inverse FFT calculator
307
309
Overlap add operator
303
110
Enhanced signal
Postprocessor
112
113

Fig. 3

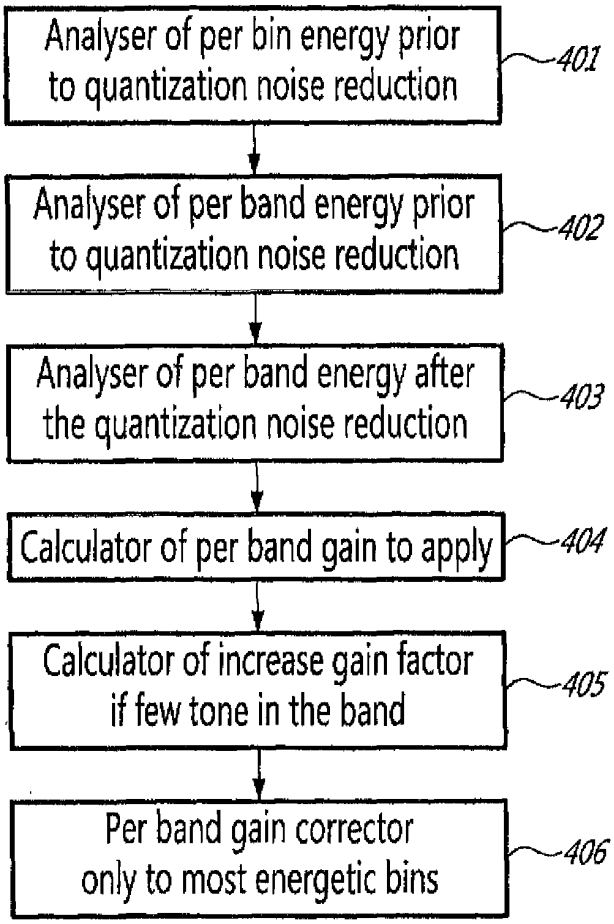
Analyser of per bin energy prior to quantization noise reduction
401
Analyser of per band energy prior to quantization noise reduction
402
Analyser of per band energy after the quantization noise reduction
403
Calculator of per band gain to apply
404
Calculator of increase gain factor if few tone in the band
405
Per band gain corrector only to most energetic bins
406

Fig. 4

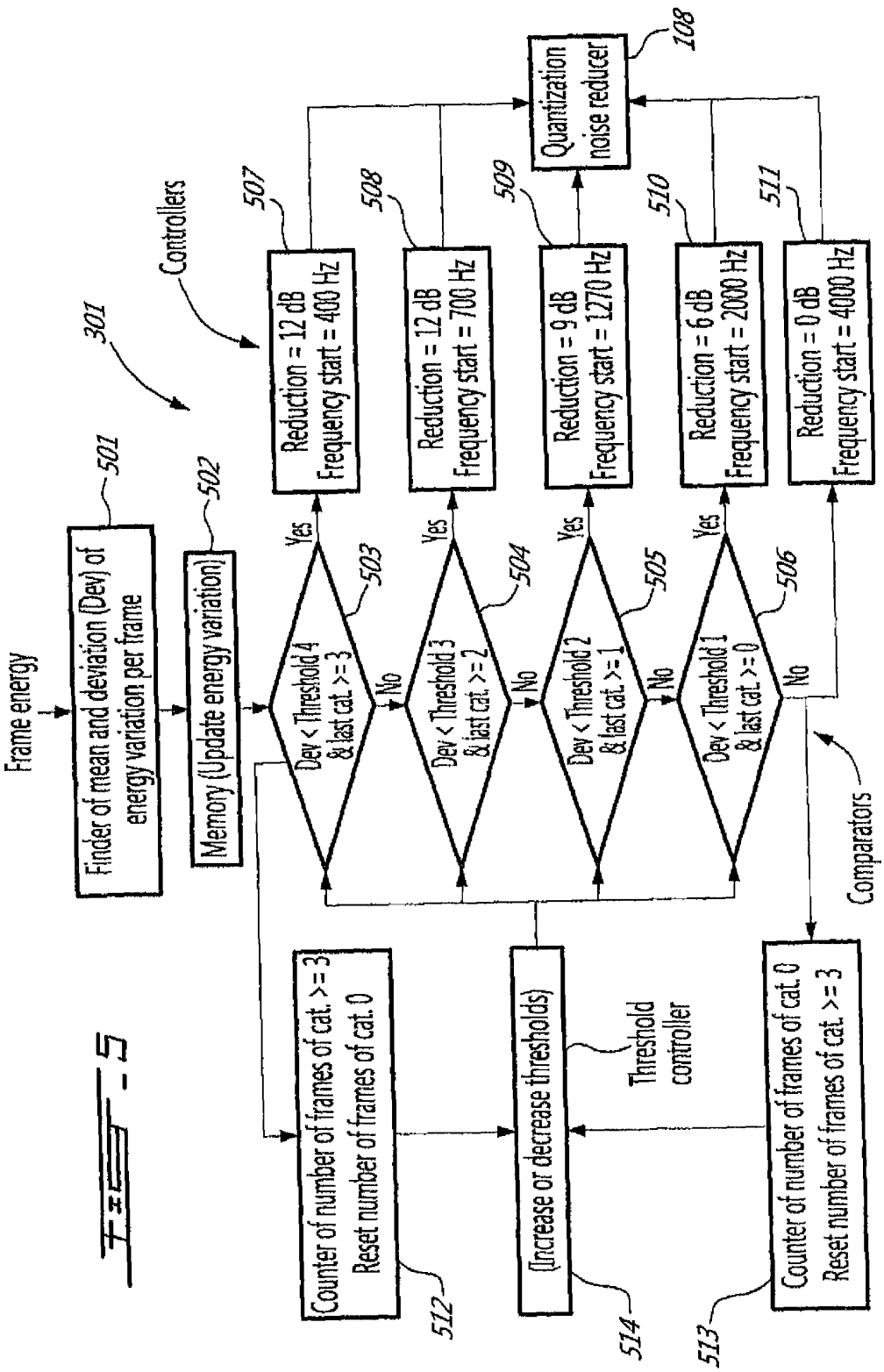
Frame energy
Finder of mean and deviation (Dev) of energy variation per frame
501
Memory (Update energy variation)
502
301
Controllers
Dev < Threshold 4 & last cat. >= 3
503
Dev < Threshold 3 & last cat. >= 2
504
Dev < Threshold 2 & last cat. >= 1
505
Dev < Threshold 1 & last cat. >= 0
506
Yes
No
Reduction = 12 dB Frequency start = 400 Hz
507
Reduction = 12 dB Frequency start = 700 Hz
508
Reduction = 9 dB Frequency start = 1270 Hz
509
Reduction = 6 dB Frequency start = 2000 Hz
510
Reduction = 0 dB Frequency start = 4000 Hz
511
Quantization noise reducer
108
Comparators
Counter of number of frames of cat. >= 3 Reset number of frames of cat. 0
512
(Increase or decrease thresholds)
514
Threshold controller
Counter of number of frames of cat. 0 Reset number of frames of cat. >= 3
513

Fig - 5

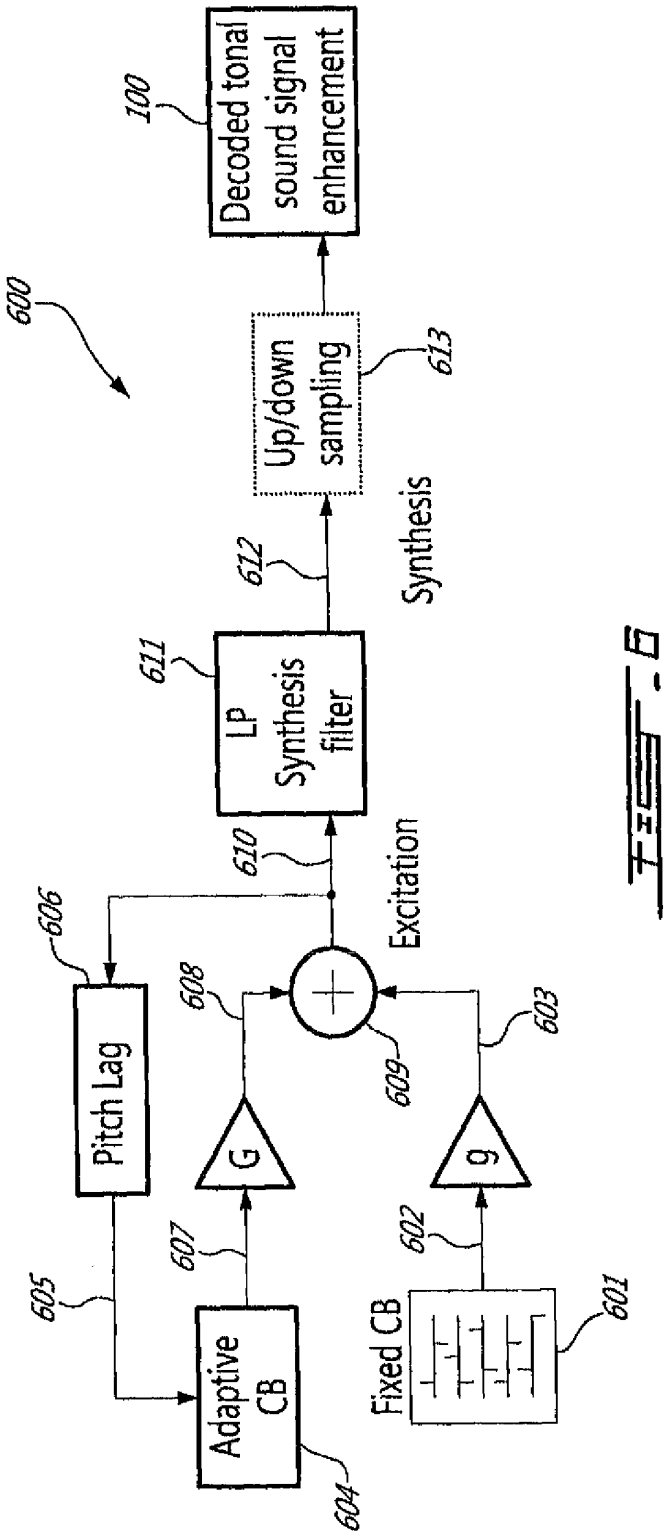
600
100
Decoded tonal sound signal enhancement
613
Up/down sampling
612
Synthesis
611
LP Synthesis filter
610
Excitation
606
Pitch Lag
608
G
609
603
g
605
607
602
Adaptive CB
Fixed CB
604
601

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

- **J. D. JOHNSTON.** Transform coding of audio signal using perceptual noise criteria. *IEEE J. Select. Areas Commun.,* February 1988, vol. 6, 314-323 **[0030]**
- **J. D. JOHNSTON.** Transform coding of audio signal using perceptual noise criteria. *IEEE J. Select. Arenas Commun,* February 1988, vol. 6, 314-323 **[0110]**