HU000033069T2

(19) **HU** (11) Lajstromszám: **E 033 069** (13) **T2**

**MAGYARORSZÁG**
**Szellemi Tulajdon Nemzeti Hivatala**

# EURÓPAI SZABADALOM
## SZÖVEGÉNEK FORDÍTÁSA

(21) Magyar ügyszám: **E 12 790692**
(22) A bejelentés napja: **2012. 10. 30.**

(51) Int. Cl.: ***G10L 19/028*** *(2006.01)*
***G10L 19/038*** *(2006.01)*

(96) Az európai bejelentés bejelentési száma:
**EP 20120790692**
(97) Az európai bejelentés közzétételi adatai:
**EP 2831874 A1** **2013. 10. 03.**
(97) Az európai szabadalom megadásának meghirdetési adatai:
**EP 2831874 B1** **2017. 05. 03.**

(86) A nemzetközi (PCT) bejelentési szám:
**PCT/SE 12/051177**
(87) A nemzetközi közzétételi szám:
**WO 13147666**

(30) Elsőbbségi adatok:
**201261617216 P** **2012. 03. 29.** **US**

(72) Feltaláló(k):
**GRANCHAROV, Volodya, S-171 67 Solna (SE)**
**TOFTGÅRD, Tomas, S-754 35 Uppsala (SE)**
**NÄSLUND, Sebastian, S-16933 Solna (SE)**
**POBLOTH, Harald, S-187 54 Täby (SE)**

(73) Jogosult(ak):
**Telefonaktiebolaget LM Ericsson (publ), 164 83 Stockholm (SE)**

(74) Képviselő:
**SBGK Szabadalmi Ügyvivői Iroda, Budapest**

(54) **Harmonikus hangjelek átalakítási kódolása/dekódolása**

Az európai szabadalom ellen, megadásának az Európai Szabadalmi Közlönyben való meghirdetésétől számított kilenc hónapon belül, felszólalást lehet benyújtani az Európai Szabadalmi Hivatalnál. (Európai Szabadalmi Egyezmény 99. cikk(1))

A fordítást a szabadalmas az 1995. évi XXXIII. törvény 84/H. §-a szerint nyújtotta be. A fordítás tartalmi helyességét a Szellemi Tulajdon Nemzeti Hivatala nem vizsgálta.

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

(11) EP 2 831 874 B1

(12) **EUROPEAN PATENT SPECIFICATION**

(45) Date of publication and mention of the grant of the patent:
**03.05.2017 Bulletin 2017/18**

(21) Application number: **12790692.3**

(22) Date of filing: **30.10.2012**

(51) Int Cl.:
***G10L 19/028*** $^{(2013.01)}$ ***G10L 19/038*** $^{(2013.01)}$

(86) International application number:
**PCT/SE2012/051177**

(87) International publication number:
**WO 2013/147666 (03.10.2013 Gazette 2013/40)**

# (54) TRANSFORM ENCODING/DECODING OF HARMONIC AUDIO SIGNALS
TRANSFORMATIONSCODIERUNG/-DECODIERUNG VON HARMONISCHEN AUDIOSIGNALEN
CODAGE/DÉCODAGE DE TRANSFORMÉE DE SIGNAUX AUDIO HARMONIQUES

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

(30) Priority: **29.03.2012 US 201261617216 P**

(43) Date of publication of application:
**04.02.2015 Bulletin 2015/06**

(73) Proprietor: **Telefonaktiebolaget LM Ericsson (publ)**
**164 83 Stockholm (SE)**

(72) Inventors:
- **GRANCHAROV, Volodya**
  **S-171 67 Solna (SE)**
- **TOFTGÅRD, Tomas**
  **S-754 35 Uppsala (SE)**
- **NÄSLUND, Sebastian**
  **S-16933 Solna (SE)**
- **POBLOTH, Harald**
  **S-187 54 Täby (SE)**

(74) Representative: **Ericsson Patent Development**
**Torshamnsgatan 21-23**
**164 80 Stockholm (SE)**

(56) References cited:
**US-A1- 2007 238 415** **US-A1- 2012 029 923**
**US-A1- 2012 046 955**

- **BARTKOWIAK MACIEJ ET AL: "Harmonic Sinusoidal + Noise Modeling of Audio Based on Multiple F0 Estimation", AES CONVENTION 125; OCTOBER 2008, AES, 60 EAST 42ND STREET, ROOM 2520 NEW YORK 10165-2520, USA, 1 October 2008 (2008-10-01), XP040508748,**

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).

Printed by Jouve, 75001 PARIS (FR)

## Description

TECHNICAL FIELD

**[0001]** The proposed technology relates to transform encoding/decoding of audio signals, especially harmonic audio signals.

BACKGROUND

**[0002]** Transform encoding is the main technology used to compress and transmit audio signals. The concept of transform encoding is to first convert a signal to the frequency domain, and then to quantize and transmit the transform coefficients. The decoder uses the received transform coefficients to reconstruct the signal waveform by applying the inverse frequency transform, see Fig. 1. In Fig. 1 an audio signal $X(n)$ is forwarded to a frequency transformer 10. The resulting frequency transform $Y(k)$ is forwarded to a transform encoder 12, and the encoded transform is transmitted to the decoder, where it is decoded by a transform decoder 14. The decoded transform $\hat{Y}(k)$ is forwarded to an inverse frequency transformer 16 that transforms it into a decoded audio signal $\hat{X}(n)$. The motivation behind this scheme is that frequency domain coefficients can be more efficiently quantized for the following reasons:

1) Transform coefficients ($Y(k)$ in Fig. 1) are more uncorrelated than input signal samples ($X(n)$ in Fig. 1).

2) The frequency transform provides energy compaction (more coefficients $Y(k)$ are close to zero and can be neglected), and

3) The subjective motivation behind the transform is that the human auditory system operates on a transformed domain, and it is easier to select perceptually important signal components on that domain.

**[0003]** In a typical transform codec the signal waveform is transformed on a block by block basis (with 50% overlap), using the Modified Discrete Cosine Transform (MDCT). In an MDCT type transform codec a block signal waveform $X(n)$ is transformed into an MDCT vector $Y(k)$. The length of the waveform blocks corresponds to 20-40 ms audio segments. If the length is denoted by $2L$, the MDCT transform can be defined as:

$$Y(k)=\sqrt{\frac{2}{L}}\sum_{n=0}^{2L-1}\sin\left[\left(n+\frac{1}{2}\right)\frac{\pi}{L}\right]\cos\left[\left(n+\frac{1}{2}+\frac{L}{2}\right)\left(k+\frac{1}{2}\right)\frac{\pi}{L}\right]X(n) \qquad (1)$$

for $k = 0,\ldots, L-1$. Then the MDCT vector $Y(k)$ is split into multiple bands (sub-vectors), and the energy (or gain) $G(j)$ in each band is calculated as:

$$G(j)=\sqrt{\frac{1}{N_j}\sum_{k=m_j}^{m_j+N_j-1}Y^2(k)} \qquad (2)$$

where $m_j$ is the first coefficient in band $j$ and $N_j$ refers to the number of MDCT coefficients in the corresponding bands (a typical range contains 8-32 coefficients). As an example of a uniform band structure, let $N_j = 8$ for all $j$, then $G(0)$ would be the energy of the first 8 coefficients, $G(1)$ would be the energy of the next 8 coefficients, etc.

**[0004]** These energy values or gains give an approximation of the spectrum envelope, which is quantized, and the quantization indices are transmitted to the decoder. Residual sub-vectors or shapes are obtained by scaling the MDCT sub-vectors with the corresponding envelope gains, e.g. the residual in each band is scaled to have unit Root Mean Square (RMS) energy. Then the residual sub-vectors or shapes are quantized with different number of bits based on the corresponding envelope gains. Finally, at the decoder, the MDCT vector is reconstructed by scaling up the residual sub-vectors or shapes with the corresponding envelope gains, and an inverse MDCT is used to reconstruct the time-domain audio frame.

**[0005]** The conventional transform encoding concept does not work well with very harmonic audio signals, e.g. single instruments. An example of such a harmonic spectrum is illustrated in Fig. 2 (for comparison a typical audio spectrum without excessive harmonics is shown Fig. 3). The reason is that the normalization with the spectrum envelope does not result in a sufficiently "flat" residual vector, and the residual encoding scheme cannot produce an audio signal of acceptable quality. This mismatch between the signal and the encoding model can be resolved only at very high bitrates, but in most cases this solution is not suitable.

**[0006]** US 2012/0029923 discloses a scheme for coding a set of transform coefficients that represent an audio frequency range of a signal uses a harmonic model to parameterize a relationship between the locations of regions of significant energy in the frequency domain.

## SUMMARY

**[0007]** An object of the proposed technology is a transform encoding/decoding scheme that is more suited for harmonic audio signals.

**[0008]** The proposed technology involves a method of encoding Modified Discrete Cosine Transform coefficients of a harmonic audio signal. The method includes the steps of:

locating spectral peaks having magnitudes exceeding a predetermined threshold, wherein the spectral peaks are located by comparing coefficients to said threshold to form a vector of peak candidates, and extracting elements from the peak candidates vector in decreasing order, wherein said threshold is calculated as

$$\theta = \left( \frac{\overline{E}_p}{\overline{E}_{nf}} \right)^{\gamma} \overline{E}_{nf},$$

where $\overline{E}_p$ is an average peak energy, $\overline{E}_{nf}$ is an average noise-floor energy and $\gamma$ has a fixed predetermined value, and wherein a peak energy is calculated as $E_p(k)=\beta E_p(k)+(1-\beta)|Y(k)|$ and a noise-floor energy is calculated as $E_{nf}(k)=\alpha E_{nf}(k)+(1-\alpha)|Y(k)|$, wherein contribution of high-energy coefficients is emphasized in calculation of the peak energy and contribution of low-energy coefficients is emphasized in calculation of the noise-floor energy;
encoding peak regions including and surrounding the located peaks, wherein the spectral peaks are quantized together with neighboring MDCT bins;
encoding, using a number of reserved bits, a first low-frequency, LF, set of coefficients outside the peak regions and below a crossover frequency that depends on the number of bits used to encode the peak regions, wherein encoding comprises encoding one or more further low-frequency set of coefficients outside the peak regions if there are non-reserved bits available after encoding the peak regions;
encoding, using a number of reserved bits, a noise-floor gain of at least one high-frequency set of not yet encoded coefficients outside the peak regions.

**[0009]** The proposed technology also involves an encoder for encoding Modified Discrete Cosine Transform coefficients of a harmonic audio signal. The encoder includes:

a peak locator configured to locate spectral peaks having magnitudes exceeding a predetermined threshold, wherein the spectral peaks are located by comparing coefficients to said threshold to form a vector of peak candidates, and extracting elements from the peak candidates vector in decreasing order, wherein said threshold is calculated as

$$\theta = \left( \frac{\overline{E}_p}{\overline{E}_{nf}} \right)^{\gamma} \overline{E}_{nf},$$

where $\overline{E}_p$ is an average peak energy, $\overline{E}_{nf}$ is an average noise-floor energy and $\gamma$ has a fixed predetermined value, and wherein a peak energy is calculated as $E_p(k)=\beta E_p(k)+(1-\beta)|Y(k)|$ and a noise-floor energy is calculated as $E_{nf}(k)=\alpha E_{nf}(k)+(1-\alpha)|Y(k)|$, wherein contribution of high-energy coefficients is emphasized in calculation of the peak energy and contribution of low-energy coefficients is emphasized in calculation of the noise-floor energy;
a peak region encoder configured to encode peak regions including and surrounding the located peaks, wherein the spectral peaks are quantized together with neighboring MDCT bins;
a low-frequency set encoder configured to encode, using a number of reserved bits, a first low-frequency set of coefficients outside the peak regions and below a crossover frequency that depends on the number of bits

used to encode the peak regions, and to encode one or more further low-frequency set of coefficients outside the peak regions if there are non-reserved bits available after encoding the peak regions; and
a noise-floor gain encoder configured to encode, using a number of reserved bits, a noise-floor gain of at least one high-frequency set of not yet encoded coefficients outside the peak regions.

**[0010]** The proposed technology also involves a user equipment (UE) including such an encoder.

**[0011]** The proposed technology also involves a method of reconstructing Modified Discrete Cosine Transform coefficients of an encoded frequency transformed harmonic audio signal. The method includes the steps of:

decoding spectral peak regions of the encoded frequency transformed harmonic audio signal, wherein said spectral peak regions includes and surround spectral peaks having magnitudes exceeding a predetermined threshold, wherein said threshold is calculated as

$$\theta = \left( \frac{\overline{E}_p}{\overline{E}_{nf}} \right)^{\gamma} \overline{E}_{nf} ,$$

where $\overline{E}_p$ is an average peak energy, $\overline{E}_{nf}$ is an average noise-floor energy and $\gamma$ has a fixed predetermined value, and wherein a peak energy is calculated as $E_p(k)=\beta E_p(k)+(1-\beta)|Y(k)|$ and a noise-floor energy is calculated as $E_{nf}(k)=\alpha E_{nf}(k)+(1-\alpha)|Y(k)|$, wherein contribution of high-energy coefficients is emphasized in calculation of the peak energy and contribution of low-energy coefficients is emphasized in calculation of the noise-floor energy;
decoding at least one low-frequency set of coefficients;
distributing coefficients of each low-frequency set outside the peak regions;
decoding a noise-floor gain of at least one high-frequency set of coefficients outside of the peak regions;
filling each high-frequency set with noise having the corresponding noise-floor gain.

**[0012]** The proposed technology also involves a decoder for reconstructing Modified Discrete Cosine Transform coefficients of an encoded frequency transformed harmonic audio signal. The decoder includes:

a peak region decoder configured to decode spectral peak regions of the encoded frequency transformed harmonic audio signal, wherein said spectral peak regions includes and surround spectral peaks having magnitudes exceeding a predetermined threshold, wherein said threshold is calculated as

$$\theta = \left( \frac{\overline{E}_p}{\overline{E}_{nf}} \right)^{\gamma} \overline{E}_{nf} ,$$

where $\overline{E}_p$ is an average peak energy, $\overline{E}_{nf}$ is an average noise-floor energy and $\gamma$ has a fixed predetermined value, and wherein a peak energy is calculated as $E_p(k)=\beta E_p(k)+(1-\beta)|Y(k)|$ and a noise-floor energy is calculated as $E_{nf}(k)=\alpha E_{nf}(k)+(1-\alpha)|Y(k)|$, wherein contribution of high-energy coefficients is emphasized in calculation of the peak energy and contribution of low-energy coefficients is emphasized in calculation of the noise-floor energy;
a low-frequency set decoder configured to decode at least one low-frequency set of coefficients;
a coefficient distributor configured to distribute coefficients of each low-frequency set outside the peak regions;
a noise-floor gain decoder configured to decode a noise-floor gain of at least one high-frequency set of coefficients outside of the peak regions;
a noise filler configured to fill each high-frequency set with noise having the corresponding noise-floor gain.

**[0013]** The proposed technology also involves a user equipment (UE) including such a decoder.

**[0014]** The proposed harmonic audio coding encoding/decoding scheme provides better perceptual quality than the conventional coding schemes for a large class of harmonic audio signals.

## BRIEF DESCRIPTION OF THE DRAWINGS

**[0015]** The present technology, together with further objects and advantages thereof, may best be understood by making reference to the following description taken together with the accompanying drawings, in which:

Fig. 1 illustrates the frequency transform coding concept;
Fig. 2 illustrates a typical spectrum of a harmonic audio signal;
Fig. 3 illustrates a typical spectrum of a non-harmonic audio signal;
Fig. 4 illustrates a peak region;
Fig. 5 is a flow chart illustrating the proposed encoding method;
Fig. 6A-D illustrates an example embodiment of the proposed encoding method;
Fig. 7 is a block diagram of an example embodiment of the proposed encoder;
Fig. 8 is a flow chart illustrating the proposed decoding method;
Fig. 9A-C illustrates an example embodiment of the proposed decoding method;
Fig. 10 is a block diagram of an example embodiment of the proposed decoder;
Fig. 11 is a block diagram of an example embodiment of the proposed encoder;
Fig. 12 is a block diagram of an example embodiment of the proposed decoder;
Fig. 13 is a block diagram of an example embodiment of a UE including the proposed encoder;
Fig. 14 is a block diagram of an example embodiment of a UE including the proposed decoder;
Fig. 15 is a flow chart of an example embodiment of a part of the proposed encoding method;
Fig. 16 is block diagram of an example embodiment of a peak region encoder in the proposed encoder;
Fig. 17 is a flow chart of an example embodiment of a part of the proposed decoding method;
Fig. 16 is block diagram of an example embodiment of a peak region decoder in the proposed decoder.

## DETAILED DESCRIPTION

**[0016]** Fig. 2 illustrates a typical spectrum of a harmonic audio signal, and Fig. 3 illustrates a typical spectrum of a non-harmonic audio signal. The spectrum of the harmonic signal is formed by strong spectral peaks separated by much weaker frequency bands, while the spectrum of the non-harmonic audio signal is much smoother.

**[0017]** The proposed technology provides an alternative audio encoding model that handles harmonic audio signals better. The main concept is that the frequency transform vector, for example an MDCT vector, is not split into envelope and residual part, but instead spectral peaks are directly extracted and quantized, together with neighboring MDCT bins. At high frequencies, low energy coefficients outside the peaks neighborhoods are not coded, but noise-filled at the decoder. Here the signal model used in the conventional encoding, {spectrum envelope + residual} is replaced with a new model {spectral peaks + noise-floor}. At low frequencies, coefficients outside the peak neighborhoods are still coded, since they have an important perceptual role.

### Encoder

**[0018]** Major steps on the encoder side are:

- Locate and code spectral peak regions
- Code low-frequency (LF) spectral coefficients. The size of coded region depends on the number of bits remaining after peak region coding.
- Code noise-floor gains for spectral coefficients outside the peak regions

**[0019]** First the noise-floor is estimated, then the spectral peaks are extracted by a peak picking algorithm (the corresponding algorithms are described in more detail in APPENDIX I-II). Each peak and its surrounding 4 neighbors are normalized to unit energy at the peak position, see Fig. 4. In other words, the entire region is scaled such that the peak has amplitude one. The peak position, gain (represents peak amplitude, magnitude) and sign are quantized. A Vector Quantizer (VQ) is applied to the MDCT bins surrounding the peak and searches for the index $I_{shape}$ of the codebook vector that provides the best match. The peak position, gain and sign, as well as the surrounding shape vectors are quantized and the quantization indices $\{I_{position}\ I_{gain}\ I_{sign}\ I_{shape}\}$ are transmitted to the decoder. In addition to these indices the decoder is also informed of the total number of peaks.

**[0020]** In the above example each peak region includes 4 neighbors that symmetrically surround the peak. However it is also feasible to have both fewer and more neighbors surrounding the peak in either symmetrical or asymmetrical fashion.

**[0021]** After the peak regions have been quantized, all available remaining bits (except reserved bits for noise-floor coding, see below) are used to quantize the low frequency MDCT coefficients. This is done by grouping the remaining unquantized MDCT coefficients into, for example, 24-dimensional bands starting from the first bin. Thus, these bands will cover the lowest frequencies up to a certain crossover frequency. Coefficients that have already been quantized in the peak coding are not included, so the bands are not necessarily made up from 24 consecutive coefficients. For this reason the bands will also be referred to as "sets" below.

**[0022]** The total number of LF bands or sets depends on the number of available bits, but there are always enough bits reserved to create at least one set. When more bits are available the first set gets more bits assigned until a threshold for the maximum number of bits per set is reached. If there are more bits available another set is created and bits are assigned to this set until the threshold is reached. This procedure is repeated until all available bits have been spent. This means that the crossover frequency at which this process is stopped will be frame dependent, since the number of peaks will vary from frame to frame. The crossover frequency will be determined by the number of bits that are available for LF encoding once the peak regions have been encoded.

**[0023]** Quantization of the LF sets can be done with any suitable vector quantization scheme, but typically some type of gain-shape encoding is used. For example, factorial pulse coding may be used for the shape vector, and scalar quantizer may be used for the gain.

**[0024]** A certain number of bits are always reserved for encoding a noise-floor gain of at least one high-frequency band of coefficients outside the peak regions, and above the upper frequency of the LF bands. Preferably two gains are used for this purpose. These gains may be obtained from the noise-floor algorithm described in APPENDIX I. If factorial pulse coding is used for the encoding the low-frequency bands some LF coefficients may not be encoded. These coefficients can instead be included in the high-frequency band encoding. As in the case of the LF bands, the HF bands are not necessarily made up from consecutive coefficients. For this reason the bands will also be referred to as "sets" below.

**[0025]** If applicable, the spectrum envelope for a bandwidth extension (BWE) region is also encoded and transmitted. The number of bands (and the transition frequency where the BWE starts) is bitrate dependent, e.g. 5.6 kHz at 24 kbps and 6.4 kHz at 32 kbps.

**[0026]** Fig. 5 is a flow chart illustrating the proposed encoding method from a general perspective. Step S1 locates spectral peaks having magnitudes exceeding a predetermined frequency dependent threshold. Step S2 encodes peak regions including and surrounding the located peaks. Step S3 encodes at least one low-frequency set of coefficients outside the peak regions and below a crossover frequency that depends on the number of bits used to encode the peak regions. Step S4 encodes a noise-floor gain of at least one high-frequency set of not yet encoded (still uncoded or remaining) coefficients outside the peak regions.

**[0027]** Fig. 6A-D illustrates an example embodiment of the proposed encoding method. Fig. 6A illustrates the MDCT transform of the signal frame to be encoded. In the figure there are fewer coefficients than in an actual signal. However, it should be kept in mind that purpose of the figure is only to illustrate the encoding process. Fig. 6B illustrates 4 identified peak regions ready for gain-shape encoding. The method described in APPENDIX II can be used to find them. Next the LF coefficients outside the peak regions are collected in Fig. 6C. These are concatenated into blocks that are gain-shape encoded. The remaining coefficients of the original signal in Fig. 6A are the high-frequency coefficients illustrated in Fig. 6D. They are divided into 2 sets and encoded (as concatenated blocks) by a noise-floor gain for each set. This noise-floor gain can be obtained from the energy of each set or by estimates obtained from the noise-floor estimation algorithm described in APPENDIX I.

**[0028]** Fig. 7 is a block diagram of an example embodiment of a proposed encoder 20. A peak locator 22 is configured to locate spectral peaks having magnitudes exceeding a predetermined frequency dependent threshold. A peak region encoder 24 is configured to encode peak regions including and surrounding the extracted peaks. A low-frequency set encoder 26 is configured to encode at least one low-frequency set of coefficients outside the peak regions and below a crossover frequency that depends on the number of bits used to encode the peak regions. A noise-floor gain encoder 28 is configured to encode a noise-floor gain of at least one high-frequency set of not yet encoded coefficients outside the peak regions. In this embodiment the encoders 24, 26, 28 use the detected peak position to decide which coefficients to include in the respective encoding.

## Decoder

**[0029]** Major steps on the decoder are:

- Reconstruct spectral peak regions
- Reconstruct LF spectral coefficients

- Noise-fill non-coded regions with noise, scaled with the received noise-floor gains.

**[0030]** The audio decoder extracts, from the bit-stream, the number of peak regions and the quantization indices $\{I_{position}\ I_{gain}\ I_{sign}\ I_{shape}\}$ in order to reconstruct the coded peak regions. These quantization indices contain information about the spectral peak position, gain and sign of the peak, as well as the index for the codebook vector that provides the best match for the peak neighborhood.

**[0031]** The MDCT low-frequency coefficients outside the peak regions are reconstructed from the encoded LF coefficients.

**[0032]** The MDCT high-frequency coefficients outside the peak regions are noise-filled at the decoder. The noise-floor level is received by the decoder, preferably in the form of two coded noise-floor gains (one for the lower and one for the upper half or part of the vector).

**[0033]** If applicable, the audio decoder performs a BWE from a pre-defined transition frequency with the received envelope gains for HF MDCT coefficients.

**[0034]** Fig. 8 is a flow chart illustrating the proposed decoding method from a general perspective. Step S11 decodes spectral peak regions of the encoded frequency transformed harmonic audio signal. Step S12 decodes at least one low-frequency set of coefficients. Step S13 distributes coefficients of each low-frequency set outside the peak regions. Step S14 decodes a noise-floor gain of at least one high-frequency set of coefficients outside the peak regions. Step S15 fills each high-frequency set with noise having the corresponding noise-floor gain.

**[0035]** In an example embodiment the decoding of a low-frequency set is based on a gain-shape decoding scheme.

**[0036]** In an example embodiment the gain-shape decoding scheme is based on scalar gain decoding and factorial pulse shape decoding.

**[0037]** An example embodiment includes the step of decoding a noise-floor gain for each of two high-frequency sets.

**[0038]** Fig. 9A-C illustrates an example embodiment of the proposed decoding method. The reconstruction of the frequency transform starts by gain-shape decoding the spectral peak regions and their positions, as illustrated in Fig. 9A. In Fig. 9B the LF set(s) are gain-shape decoded and the decoded transform coefficient are distributed in blocks outside the peak regions. In Fig. 9C the noise-floor gains are decoded and the remaining transform coefficients are filled with noise having corresponding noise-floor gains. In this way the transform of Fig. 6A has been approximately reconstructed. A comparison of Fig. 9C with Fig. 6A and 6D shows that the noise filled regions have different individual coefficients but the same energy, as expected.

**[0039]** Fig. 10 is a block diagram of an example embodiment of a proposed decoder 40. A peak region decoder 42 is configured to decode spectral peak regions of the encoded frequency transformed harmonic audio signal. A low-frequency set decoder 44 is configured to decode at least one low-frequency set of coefficients. A coefficient distributor 46 configured to distribute coefficients of each low-frequency set outside the peak regions. A noise-floor gain decoder 48 is configured to decode a noise-floor of at least one high-frequency set of coefficients outside the peak regions. A noise filler 50 is configured to fill each high-frequency set with noise having the corresponding noise-floor gain. In this embodiment the peak positions are forwarded to the coefficient distributor 46 and the noise filler 50 to avoid overwriting of the peak regions.

**[0040]** The steps, functions, procedures and/or blocks described herein may be implemented in hardware using any conventional technology, such as discrete circuit or integrated circuit technology, including both general-purpose electronic circuitry and application-specific circuitry.

**[0041]** Alternatively, at least some of the steps, functions, procedures and/or blocks described herein may be implemented in software for execution by suitable processing equipment. This equipment may include, for example, one or several micro processors, one or several Digital Signal Processors (DSP), one or several Application Specific Integrated Circuits (ASIC), video accelerated hardware or one or several suitable programmable logic devices, such as Field Programmable Gate Arrays (FPGA). Combinations of such processing elements are also feasible.

**[0042]** It should also be understood that it may be possible to reuse the general processing capabilities already present in the encoder/decoder. This may, for example, be done by reprogramming of the existing software or by adding new software components.

**[0043]** Fig. 11 is a block diagram of an example embodiment of the proposed encoder 20. This embodiment is based on a processor 110, for example a micro processor, which executes software 120 for locating peaks, software 130 for encoding peak regions, software 140 for encoding at least one low-frequency set, and software 150 for encoding at least one noise-floor gain. The software is stored in memory 160. The processor 110 communicates with the memory over a system bus. The incoming frequency transform is received by an input/output (I/O) controller 170 controlling an I/O bus, to which the processor 110 and the memory 160 are connected. The encoded frequency transform obtained from the software 150 is outputted from the memory 160 by the I/O controller 170 over the I/O bus.

**[0044]** Fig. 12 is a block diagram of an example embodiment of the proposed decoder 40. This embodiment is based on a processor 210, for example a micro processor, which executes software 220 for decoding peak regions, software 230 for decoding at least one low-frequency set, software 240 for distributing LF coefficients, software 250 for decoding at least one noise-floor gain, and software 260 for noise filling. The software is stored in memory 270. The processor

210 communicates with the memory over a system bus. The incoming encoded frequency transform is received by an input/output (I/O) controller 280 controlling an I/O bus, to which the processor 210 and the memory 280 are connected. The reconstructed frequency transform obtained from the software 260 is outputted from the memory 270 by the I/O controller 280 over the I/O bus.

**[0045]** The technology described above is intended to be used in an audio encoder/decoder, which can be used in a mobile device (e.g. mobile phone, laptop) or a stationary device, such as a personal computer. Here the term User Equipment (UE) will be used as a generic name for such devices.

**[0046]** Fig. 13 is a block diagram of an example embodiment of a UE including the proposed encoder. An audio signal from a microphone 70 is forwarded to an A/D converter 72, the output of which is forwarded to an audio encoder 74. The audio encoder 74 includes a frequency transformer 76 transforming the digital audio samples into the frequency domain. A harmonic signal detector 78 determines whether the transform represents harmonic or non-harmonic audio. If it represents non-harmonic audio, it is encoded in a conventional encoding mode (not shown). If it represents harmonic audio, it is forwarded to a frequency transform encoder 20 in accordance with the proposed technology. The encoded signal is forwarded to a radio unit 80 for transmission to a receiver.

**[0047]** The decision of the harmonic signal detector 78 is based on the noise-floor energy $\overline{E}_{nf}$ and peak energy $\overline{E}_p$ in APPENDIX I and II. The logic is as follows: IF $\overline{E}p / \overline{E}_{nf}$ is above a threshold AND the number of detected peaks is in a predefined range THEN the signal is classified as harmonic. Otherwise the signal is classified as non-harmonic. The classification and thus the encoding mode is explicitly signaled to the decoder.

**[0048]** Fig. 14 is a block diagram of an example embodiment of a UE including the proposed decoder. A radio signal received by a radio unit 82 is converted to baseband, channel decoded and forwarded to an audio decoder 84. The audio decoder includes a decoding mode selector 86, which forwards the signal a frequency transform decoder 40 in accordance with the proposed technology if it has been classified as harmonic. If it has been classified as non-harmonic audio, it is decoded in a conventional decoder (not shown). The frequency transform decoder 40 reconstructs the frequency transform as described above. The reconstructed frequency transform is converted to the time domain in an inverse frequency transformer 88. The resulting audio samples are forwarded to a D/A conversion and amplification unit 90, which forwards the final audio signal to a loudspeaker 92.

**[0049]** Fig. 15 is a flow chart of an example embodiment of a part of the proposed encoding method. In this embodiment the peak region encoding step S2 in Fig. 5 has been divided into sub-steps S2-A to S2-E. Step S2-A encodes spectrum position and sign of a peak. Step S2-B quantizes peak gain. Step S2-C encodes the quantized peak gain. Step S2-D scales predetermined frequency bins surrounding the peak by the inverse of the quantized peak gain. Step S2-E shape encodes the scaled frequency bins.

**[0050]** Fig. 16 is block diagram of an example embodiment of a peak region encoder in the proposed encoder. In this embodiment the peak region encoder 24 includes elements 24-A to 24-D. Position and sign encoder 24-A is configured to encode spectrum position and sign of a peak. Peak gain encoder 24-B is configured to quantize peak gain and to encode the quantized peak gain. Scaling unit 24-C is configured to scale predetermined frequency bins surrounding the peak by the inverse of the quantized peak gain. Shape encoder 24-D is configured to shape encode the scaled frequency bins.

**[0051]** Fig. 17 is a flow chart of an example embodiment of a part of the proposed decoding method. In this embodiment the peak region decoding step S11 in Fig. 8 has been divided into sub-steps S11-A to S11-D. Step S11-A decodes spectrum position and sign of a peak. Step S11-B decodes peak gain. Step S11-C decodes a shape of predetermined frequency bins surrounding the peak. Step S11-D scales the decoded shape by the decoded peak gain.

**[0052]** Fig. 18 is block diagram of an example embodiment of a peak region decoder in the proposed decoder. In this embodiment the peak region decoder 42 includes elements 42-A to 42-D. A position and sign decoder 42-A is configured to decode spectrum position and sign of a peak. A peak gain decoder 42-B is configured to decode peak gain. A shape decoder 42-C is configured to decode a shape of predetermined frequency bins surrounding the peak. A scaling unit 42-D is configured to scale the decoded shape by the decoded peak gain.

**[0053]** Specific implementation details for a 24 kbps mode are given below.

- The codec operates on 20 ms frames, which at a bit rate of 24 kbps gives 480 bits per-frame.
- The processed audio signal is sampled at 32 kHz, and has an audio bandwidth of 16 kHz.
- The transition frequency is set to 5.6 kHz (all frequency components above 5.6 kHz are bandwidth-extended).
- Reserved bits for signaling and bandwidth extension of frequencies above the transition frequency: ~30-40.
- Bits for coding two noise-floor gains: 10.
- The number of coded spectral peak regions is 7-17. The number of bits used per peak region is ~20-22, which gives a total number of ~140-340 for coding all peaks positions, gains, signs, and shapes.
- Bits for coding low frequency bands: ~100-300
- Coded low frequency bands: 1-4 (each band contains 8 MDCT bins). Since each MDCT bin corresponds to 25 Hz,

coded low-frequency region corresponds to 200-800 Hz

- The gains used for bandwidth extension and the peak gains are Huffman coded so the number of bits used by these might vary between frames even for a constant number of peaks.
- The peak position and sign coding makes use of an optimization which makes it more efficient as the number of peaks increase. For 7 peaks, position and sign requires about 6.9 bits per peak and for 17 peaks the number is about 5.7 bits per peak.
- This variability in how many bits are used in different stages of the coding is no problem since the low frequency band coding comes last and just uses up whatever bits remain. However the system is designed so that enough bits always remain to encode one low frequency band.

**[0054]** The table below presents results from a listening test performed in accordance with the procedure described in ITU-R BS. 1534-1 MUSHRA (Multiple Stimuli with Hidden Reference and Anchor). The scale in a MUSHRA test is 0 to 100, where low values correspond to low perceived quality, and high values correspond to high quality. Both codecs operated at 24 kbps. Test results are averaged over 24 music items and votes from 8 listeners.

| System Under Test | MUSHRA Score |
|---|---|
| Low-pass anchor signal (bandwidth 7 kHz) | 48.89 |
| Conventional coding scheme | 49.94 |
| Proposed harmonic coding scheme | 55.87 |
| Reference signal (bandwidth 16 kHz) | 100.00 |

**[0055]** It will be understood by those skilled in the art that various modifications and changes may be made to the proposed technology without departure from the scope thereof, which is defined by the appended claims.

APPENDIX I

**[0056]** The noise-floor estimation algorithm operates on the absolute values of transform coefficients $|Y(k)|$. Instantaneous noise-floor energies $E_{nf}(k)$ are estimated according to the recursion:

$$E_{nf}(k) = \alpha E_{nf}(k) + (1-\alpha)|Y(k)| \quad (3)$$

where

$$\alpha = \begin{cases} 0.9578 & if \quad |Y(k)| > E_{nf}(k-1) \\ 0.6472 & if \quad |Y(k)| \le E_{nf}(k-1) \end{cases} \quad (4)$$

**[0057]** The particular form of the weighting factor $\alpha$ minimizes the effect of high-energy transform coefficients and emphasizes the contribution of low-energy coefficients. Finally the noise-floor level $\overline{E}_{nf}$ is estimated by simply averaging the instantaneous energies $E_{nf}(k)$.

APPENDIX II

**[0058]** The peak-picking algorithm requires knowledge of noise-floor level and average level of spectral peaks. The peak energy estimation algorithm is similar to the noise-floor estimation algorithm, but instead of low-energy, it tracks high-spectral energies:

$$E_p(k) = \beta E_p(k) + (1-\beta)|Y(k)| \quad (5)$$

where

$$\beta = \begin{cases} 0.4223 & if \quad |Y(k)| > E_p(k-1) \\ 0.8029 & if \quad |Y(k)| \le E_p(k-1) \end{cases} \tag{6}$$

**[0059]** In this case the weighting factor $\beta$ minimizes the effect of low-energy transform coefficients and emphasizes the contribution of high-energy coefficients. The overall peak energy $\overline{E}_p$ is estimated by simply averaging the instantaneous energies.

**[0060]** When the peak and noise-floor levels are calculated, a threshold level $\theta$ is formed as:

$$\theta = \left( \frac{\overline{E}_p}{\overline{E}_{nf}} \right)^{\gamma} \overline{E}_{nf} \tag{7}$$

with $\gamma = 0.88579$. Transform coefficients are compared to the threshold, and the ones with amplitude above it, form a vector of peak candidates. Since the natural sources do not typically produce peaks that are very close, e.g., 80 Hz, the vector with peak candidates is further refined. Vector elements are extracted in decreasing order, and the neighborhood of each element is set to zero. In this way only the largest element in certain spectral region remain, and the set of these elements form the spectral peaks for the current frame.

ABBREVIATIONS

**[0061]**

| | |
|---|---|
| ASIC | Application Specific Integrated Circuit |
| BWE | BandWidth Extension |
| DSP | Digital Signal Processors |
| FPGA | Field Programmable Gate Arrays |
| HF | High-Frequency |
| LF | Low-Frequency |
| MDCT | Modified Discrete Cosine Transform |
| RMS | Root Mean Square |
| VQ | Vector Quantizer |

## Claims

1. A method of encoding Modified Discrete Cosine Transform, MDCT, coefficients ($Y(k)$) of a harmonic audio signal, said method including the steps of:

   locating (S1) spectral peaks having magnitudes exceeding a predetermined threshold, wherein the spectral peaks are located by comparing coefficients to said threshold to form a vector of peak candidates, and extracting elements from the peak candidates vector in decreasing order, wherein said threshold is calculated as

   $$\theta = \left( \frac{\overline{E}_p}{\overline{E}_{nf}} \right)^{\gamma} \overline{E}_{nf},$$

   where $\overline{E}_p$ is an average peak energy, $\overline{E}_{nf}$ is an average noise-floor energy and $\gamma$ has a fixed predetermined value, and wherein a peak energy is calculated as $E_p(k)=\beta E_p(k)+(1-\beta)|Y(k)|$ and a noise-floor energy is calculated as $E_{nf}(k)=\alpha E_{nf}(k)+(1-\alpha)|Y(k)|$, wherein contribution of high-energy coefficients is emphasized in calculation of the peak energy and contribution of low-energy coefficients is emphasized in calculation of the noise-floor energy;
   encoding (S2) peak regions including and surrounding the located peaks, wherein the spectral peaks are quantized together with neighboring MDCT bins;

encoding (S3), using a number of reserved bits, a first low-frequency, LF, set of coefficients outside the peak regions and below a crossover frequency that depends on the number of bits used to encode the peak regions, wherein encoding (S3) comprises encoding one or more further low-frequency set of coefficients outside the peak regions if there are non-reserved bits available after encoding the peak regions;
encoding (S4), using a number of reserved bits, a noise-floor gain of at least one high-frequency set of not yet encoded coefficients outside the peak regions.

2. The encoding method of claim 1, where a weighting factor $\alpha$ is defined as

$$\alpha = \begin{cases} 0.9578 & if \quad |Y(k)| > E_{nf}(k-1) \\ 0.6472 & if \quad |Y(k)| \le E_{nf}(k-1) \end{cases},$$

and a weighting factor $\beta$ is defined as

$$\beta = \begin{cases} 0.4223 & if \quad |Y(k)| > E_{p}(k-1) \\ 0.8029 & if \quad |Y(k)| \le E_{p}(k-1) \end{cases}.$$

3. The encoding method of claim 1 or 2, wherein the step (S2) of encoding peak regions comprises:

encoding (S2-A) spectrum position and sign of a peak;
quantizing (S2-B) peak gain;
encoding (S2-C) the quantized peak gain;
scaling (S2-D) predetermined frequency bins surrounding the peak by the inverse of the quantized peak gain;
shape encoding (S2-E) the scaled frequency bins.

4. The encoding method of any of claims 1 to 3, wherein the peak region comprises the peak and four MDCT bins surrounding said peak.

5. The encoding method of any of the preceding claims, wherein the step (S3) of encoding low-frequency set of coefficients comprises grouping remaining un-quantized MDCT coefficients into 24-dimensional bands.

6. The encoding method of any of the preceding claims, wherein encoding of a low-frequency set is based on a gain-shape encoding scheme, said gain-shape encoding scheme being based on scalar gain quantization and factorial pulse shape encoding.

7. The encoding method of any of the preceding claims, including the step of encoding a noise-floor gain for each of two high-frequency sets.

8. An encoder for encoding Modified Discrete Cosine Transform, MDCT, coefficients ($Y(k)$) of a harmonic audio signal, said encoder including:

a peak locator (22) configured to locate spectral peaks having magnitudes exceeding a predetermined threshold, wherein the spectral peaks are located by comparing coefficients to said threshold to form a vector of peak candidates, and extracting elements from the peak candidates vector in decreasing order, wherein said threshold is calculated as

$$\theta = \left( \frac{\overline{E}_p}{\overline{E}_{nf}} \right)^{\gamma} \overline{E}_{nf} ,$$

where $\overline{E}p$ is an average peak energy, $\overline{E}_{nf}$ is an average noise-floor energy and $\gamma$ has a fixed predetermined

value, and wherein a peak energy is calculated as $E_p(k)=\beta E_p(k)+(1-\beta)|Y(k)|$ and a noise-floor energy is calculated as $E_{nf}(k)=\alpha E_{nf}(k)+(1-\alpha)|Y(k)|$, wherein contribution of high-energy coefficients is emphasized in calculation of the peak energy and contribution of low-energy coefficients is emphasized in calculation of the noise-floor energy;
a peak region encoder (24) configured to encode peak regions including and surrounding the located peaks, wherein the spectral peaks are quantized together with neighboring MDCT bins;
a low-frequency set encoder (26) configured to encode, using a number of reserved bits, a first low-frequency set of coefficients outside the peak regions and below a crossover frequency that depends on the number of bits used to encode the peak regions, and to encode one or more further low-frequency set of coefficients outside the peak regions if there are non-reserved bits available after encoding the peak regions; and
a noise-floor gain encoder (28) configured to encode, using a number of reserved bits, a noise-floor gain of at least one high-frequency set of not yet encoded coefficients outside the peak regions.

**9.** The encoder of claim 8, wherein the peak region encoder (24) includes:

a position and sign encoder (24-A) configured to encode spectrum position ($I_{position}$) and sign ($I_{sign}$) of a peak;
a peak gain encoder (24-B) configured to quantize peak gain and to encode ($I_{gain}$) the quantized peak gain;
a scaling unit (24-C) configured to scale predetermined frequency bins surrounding the peak by the inverse of the quantized peak gain;
a shape encoder (24-D) configured to shape encode the scaled frequency bins.

**10.** A user equipment (UE) including an encoder (20) in accordance with claim 8 or 9.

## Patentansprüche

**1.** Verfahren zur Codierung von MDCT (modifizierte diskrete Cosinus-Transformation)-Koeffizienten ($Y(k)$) eines harmonischen Audiosignals, wobei das Verfahren die folgenden Schritte umfasst:

Lokalisieren (S1) von Spektralspitzen mit Größen, die eine vorbestimmte Schwelle überschreiten, wobei die Spektralspitzen durch Vergleichen von Koeffizienten mit der Schwelle, um einen Vektor von Spitzenkandidaten zu bilden, und Extrahieren von Elementen aus dem Spitzenkandidatenvektor in absteigender Reihenfolge lokalisiert werden, wobei die Schwelle berechnet wird als:

$$\theta = \left(\frac{\bar{E}_p}{\bar{E}_{nf}}\right)^{\gamma} \bar{E}_{nf},$$

wobei $\overline{E}_p$ eine mittlere Spitzenenergie ist, $\overline{E}_{nf}$ eine mittlere Grundrauschenergie ist, und $\gamma$ einen festen vorbestimmten Wert hat, und wobei eine Spitzenenergie als $E_p(k) = \beta E_p(k) + (1 - \beta)|Y(k)|$ berechnet wird, und eine Grundrauschenergie als $E_{nf}(k) = \alpha E_{nf}(k) + (1-\alpha)|Y(k)|$ berechnet wird, wobei bei der Berechnung der Spitzenenergie ein Beitrag von Koeffizienten hoher Energie betont wird, und bei der Berechnung der Grundrauschenergie ein Beitrag von Koeffizienten niedriger Energie betont wird;
Codieren (S2) von Spitzenregionen, welche die lokalisierten Spitzen umfassen und umgeben, wobei die Spektralspitzen zusammen mit benachbarten MDCT-Bins quantisiert werden;
Codieren (S3) unter Verwendung einer Anzahl von reservierten Bits eines ersten Niederfrequenz, LF,-Satzes von Koeffizienten außerhalb der Spitzenregionen und unter einer Übergangfrequenz, die von der Anzahl von Bits abhängt, die zum Codieren der Spitzenregionen verwendet werden, wobei das Codieren (S3) ein Codieren eines oder mehrerer weiterer Niederfrequenzsätze von Koeffizienten außerhalb der Spitzenregionen umfasst, wenn nach dem Codieren der Spitzenregionen nichtreservierte Bits vorhanden sind;
Codieren (S4) unter Verwendung einer Anzahl von reservierten Bits einer Grundrauschverstärkung mindestens eines Hochfrequenzsatzes von noch nicht codierten Koeffizienten außerhalb der Spitzenregionen.

**2.** Codierverfahren nach Anspruch 1, wobei ein Gewichtungsfaktor $\alpha$ definiert wird als

$$\alpha = \begin{cases} 0{,}9578 & si \quad |Y(k)| > E_{nf}(k-1) \\ 0{,}6472 & si \quad |Y(k)| \le E_{nf}(k-1)' \end{cases}$$

und ein Gewichtungsfaktor β definiert wird als

$$\beta = \begin{cases} 0{,}4223 & si \quad |Y(k)| > E_p(k-1) \\ 0{,}8029 & si \quad |Y(k)| \le E_p(k-1) \end{cases}.$$

3. Codierverfahren nach Anspruch 1 oder 2, wobei der Schritt (S2) des Codierens von Spitzenregionen umfasst:

   Codieren (S2-A) von Spektralposition und Vorzeichen einer Spitze;
   Quantisieren (S2-B) von Spitzenverstärkung;
   Codieren (S2-C) der quantisierten Spitzenverstärkung;
   Skalieren (S2-D) von vorbestimmten Frequenz-Bins, welche die Spitze umgeben, um die Inverse der quantisierten Spitzenverstärkung;
   Formcodieren (S2-E) der skalierten Frequenz-Bins.

4. Codierverfahren nach einem der Ansprüche 1 bis 3, wobei die Spitzenregion die Spitze und vier die Spitze umgebende MDCT-Bins umfasst.

5. Codierverfahren nach einem der vorhergehenden Ansprüche, wobei der Schritt (S3) des Codierens eines Niederfrequenzsatzes von Koeffizienten ein Gruppieren von restlichen nichtquantisierten MDCT-Koeffizienten in 24-dimensionale Bänder umfasst.

6. Codierverfahren nach einem der vorhergehenden Ansprüche, wobei das Codieren eines Niederfrequenzsatzes auf einem Verstärkungs-/Formcodierschema basiert, wobei das Verstärkungs-/Formcodierschema auf skalarer Verstärkungsquantisierung und faktorieller Impulsformcodierung basiert.

7. Codierverfahren nach einem der vorhergehenden Ansprüche, umfassend den Schritt des Codierens von Grundrauschverstärkung für jeden von zwei Hochfrequenzsätzen.

8. Encoder zum Codieren von MDCT (modifizierte diskrete Cosinus-Transformation)-Koeffizienten ($Y(k)$) eines harmonischen Audiosignals, wobei der Encoder umfasst:

   eine Spitzenpositionsgeber (22), der so konfiguriert ist, dass er Spektralspitzen mit Größen lokalisiert, die eine vorbestimmte Schwelle überschreiten, wobei die Spektralspitzen durch Vergleichen von Koeffizienten mit der Schwelle, um einen Vektor von Spitzenkandidaten zu bilden, und Extrahieren von Elementen aus dem Spitzenkandidatenvektor in absteigender Reihenfolge lokalisiert werden, wobei die Schwelle berechnet wird als:

$$\theta = \left(\frac{\bar{E}_p}{\bar{E}_{nf}}\right)^{\gamma} \bar{E}_{nf},$$

   wobei $\overline{E}_p$ eine mittlere Spitzenenergie ist, $\overline{E}_{nf}$ eine mittlere Grundrauschenergie ist, und γ einen festen vorbestimmten Wert hat, und wobei eine Spitzenenergie als $E_p(k) = \beta E_p(k) + (1 - \beta)|Y(k)|$ berechnet wird, und eine Grundrauschenergie als $E_{nf}(k) = \alpha E_{nf}(k) + (1-\alpha)|Y(k)|$ berechnet wird, wobei bei der Berechnung der Spitzenenergie ein Beitrag von Koeffizienten hoher Energie betont wird, und bei der Berechnung der Grundrauschenergie ein Beitrag von Koeffizienten niedriger Energie betont wird;
   einen Spitzenregion-Encoder (24), der so konfiguriert ist, dass er Spitzenregionen codiert, welche die lokalisierten Spitzen umfassen und umgeben, wobei die Spektralspitzen zusammen mit benachbarten MDCT-Bins quantisiert werden;
   einen Niederfrequenzsatz-Encoder (26), der so konfiguriert ist, dass er unter Verwendung einer Anzahl von reservierten Bits einen ersten Niederfrequenz, LF,-Satz von Koeffizienten außerhalb der Spitzenregionen und unter einer Übergangfrequenz codiert, die von der Anzahl von Bits abhängt, die zum Codieren der Spitzenregionen verwendet werden, und einen oder mehrere weitere Niederfrequenzsätze von Koeffizienten außerhalb der Spitzenregionen codiert, wenn nach dem Codieren der Spitzenregionen nichtreservierte Bits vorhanden sind; und
   Grundrauschverstärkungs-Encoder (28), der so konfiguriert ist, dass er unter Verwendung einer Anzahl von reservierten Bits eine Grundrauschverstärkung mindestens eines Hochfrequenzsatzes von noch nicht

codierten Koeffizienten außerhalb der Spitzenregionen codiert.

**9.** Encoder nach Anspruch 8, wobei der Spitzenregion-Encoder (24) umfasst:

einen Positions- und Vorzeichen-Encoder (24-A), der so konfiguriert ist, dass er Spektralposition ($I_{position}$) und Vorzeichen ($I_{sign}$) einer Spitze codiert;
einen Spitzenverstärkungs-Encoder (24-B), der so konfiguriert ist, dass er Spitzenverstärkung quantisiert und die quantisierte Spitzenverstärkung codiert ($I_{gain}$) ;
eine Skaliereinheit (24-C), die so konfiguriert ist, dass sie vorbestimmte Frequenz-Bins, welche die Spitze umgeben, um die Inverse der quantisierten Spitzenverstärkung skaliert;
einen Form-Encoder (24-D), der so konfiguriert ist, dass er die skalierten Frequenz-Bins formcodiert.

**10.** Benutzereinrichtung (UE), umfassend einen Encoder (20) nach Anspruch 8 oder 9.

## Revendications

**1.** Procédé de codage de coefficients de transformée cosinusoïdale discrète modifiée, MDCT, (Y(k)) d'un signal audio harmonique, ledit procédé comprenant les étapes de :

la localisation (S1) de crêtes spectrales ayant des grandeurs dépassant un seuil prédéterminé, dans lequel les crêtes spectrales sont localisées par la comparaison de coefficients au dit seuil pour former un vecteur de candidats de crête, et l'extraction d'éléments du vecteur de candidats de crête par ordre décroissant, dans lequel ledit seuil est calculé par :

$$\theta = \left(\frac{\bar{E}_p}{\bar{E}_{nf}}\right)^{\gamma} \bar{E}_{nf} ,$$

où $\overline{E}p$ est une énergie de crête moyenne, $\overline{E}_{nf}$ est une énergie de plancher de bruit moyenne et $\gamma$ a une valeur prédéterminée fixe, et dans lequel une énergie de crête est calculée par $E_p(k)=\beta E_p(k) + (1-\beta)|Y(k)|$ et une énergie de plancher de bruit est calculée par $E_{nf}(k)=\alpha E_{nf}(k) + (1-\alpha)|Y(k)|$, dans lequel la contribution de coefficients de haute énergie est accentuée dans le calcul de l'énergie de crête et une contribution de coefficients de basse énergie est accentuée dans le calcul de l'énergie de plancher de bruit ;
le codage (S2) de régions de crête incluant et entourant les crêtes localisées, dans lequel les crêtes spectrales sont quantifiées avec des bins MDCT voisines ;
le codage (S3), en utilisant un nombre de bits réservés, d'un premier ensemble de basses fréquences, LF, de coefficients à l'extérieur des régions de crête et au-dessous d'une fréquence de croisement en fonction du nombre de bits utilisés pour coder les régions de crête, dans lequel le codage (S3) comprend le codage d'un ou plusieurs autres ensembles de basses fréquences de coefficients à l'extérieur des régions de crête si des bits non réservés sont disponibles après le codage des régions de crête ;
le codage (S4), en utilisant un nombre de bits réservés, d'un gain de plancher de bruit d'au moins un ensemble de hautes fréquences de coefficients pas encore codés à l'extérieur des régions de crête.

**2.** Procédé de codage selon la revendication 1, dans lequel un facteur de pondération $\alpha$ est défini en tant que :

$$\alpha = \begin{cases} 0{,}9578 & si \quad |Y(k)| > E_{nf}(k-1) \\ 0{,}6472 & si \quad |Y(k)| \leq E_{nf}(k-1) \end{cases}'$$

et un facteur de pondération $\beta$ est défini en tant que :

$$\beta = \begin{cases} 0{,}4223 & si \quad |Y(k)| > E_p(k-1) \\ 0{,}8029 & si \quad |Y(k)| \leq E_p(k-1) \end{cases}.$$

**3.** Procédé de codage selon la revendication 1 ou 2, dans lequel l'étape (S2) du codage de régions de crête comprend :

le codage (S2-A) d'une position de spectre et d'un signe d'une crête ;
la quantification (S2-B) d'un gain de crête ;
le codage (S2-C) du gain de crête quantifié ;
la mise à l'échelle (S2-D) des bins de fréquences prédéterminées entourant la crête par l'inverse du gain de crête quantifié ;
le codage de forme (S2-E) des bins de fréquences mises à l'échelle.

**4.** Procédé de codage selon l'une quelconque des revendications 1 à 3, dans lequel la région de crête comprend la crête et quatre bins MDCT entourant ladite crête.

**5.** Procédé de codage selon l'une quelconque des revendications précédentes, dans lequel l'étape (S3) du codage d'un ensemble de basses fréquences de coefficients comprend le regroupement de coefficients MDCT non quantifiés restants dans 24 bandes dimensionnelles.

**6.** Procédé de codage selon l'une quelconque des revendications précédentes, dans lequel le codage d'un ensemble de basses fréquences est basé sur un schéma de codage de forme de gain, ledit schéma de codage de forme de gain étant basé sur une quantification de gain scalaire et un codage de forme d'impulsion factorielle.

**7.** Procédé de codage selon l'une quelconque des revendications précédentes, comprenant l'étape du codage d'un gain de plancher de bruit pour chacun de deux ensembles de hautes fréquences.

**8.** Codeur de codage de coefficients de transformée cosinusoïdale discrète modifiée, MDCT, (Y(k)) d'un signal audio harmonique, ledit codeur comprenant :

un localisateur de crête (22) configuré pour effectuer la localisation de crêtes spectrales ayant des grandeurs dépassant un seuil prédéterminé, dans lequel les crêtes spectrales sont localisées par la comparaison de coefficients au dit seuil pour former un vecteur de candidats de crête, et l'extraction d'éléments du vecteur de candidats de crête par ordre décroissant, dans lequel ledit seuil est calculé par :

$$\theta = \left(\frac{\bar{E}_p}{\bar{E}_{nf}}\right)^{\gamma} \bar{E}_{nf},$$

où $\overline{Ep}$ est une énergie de crête moyenne, $\overline{E}_{nf}$ est une énergie de plancher de bruit moyenne et $\gamma$ a une valeur prédéterminée fixe, et dans lequel une énergie de crête est calculée par $E_p(k)=\beta E_p(k) + (1-\beta)|Y(k)|$ et une énergie de plancher de bruit est calculée par $E_{nf}(k)=\alpha E_{nf}(k) + (1-\alpha)|Y(k)|$, dans lequel la contribution de coefficients de haute énergie est accentuée dans le calcul de l'énergie de crête et une contribution de coefficients de basse énergie est accentuée dans le calcul de l'énergie de plancher de bruit ;
un codeur de régions de crête (24) configuré pour effectuer le codage de régions de crête incluant et entourant les crêtes localisées, dans lequel les crêtes spectrales sont quantifiées avec des bins MDCT voisines ;
un codeur d'ensemble de basses fréquences (26) configuré pour effectuer le codage, en utilisant un nombre de bits réservés, d'un premier ensemble de basses fréquences de coefficients à l'extérieur des régions de crête et au-dessous d'une fréquence de croisement en fonction du nombre de bits utilisés pour coder les régions de crête, et le codage d'un ou plusieurs autres ensembles de basses fréquences de coefficients à l'extérieur des régions de crête si des bits non réservés sont disponibles après le codage des régions de crête ; et
un codeur de gain de plancher de bruit (28) configuré pour effectuer le codage, en utilisant un nombre de bits réservés, d'un gain de plancher de bruit d'au moins un ensemble de hautes fréquences de coefficients pas encore codés à l'extérieur des régions de crête.

**9.** Codeur selon la revendication 8, dans lequel le codeur de régions de crête (24) comprend :

un codeur de position et de signe (24-A) configuré pour effectuer le codage d'une position de spectre ($I_{position}$) et d'un signe ($I_{sign}$) d'une crête ;
un codeur de gain de crête (24-B) configuré pour effectuer la quantification d'un gain de crête et le codage du gain de crête quantifié ($I_{gain}$) ;

une unité de mise à l’échelle (24-C) configurée pour effectuer la mise à l’échelle des bins de fréquences prédéterminées entourant la crête par l’inverse du gain de crête quantifié ;
un codeur de forme (24-D) configuré pour effectuer le codage de forme des bins de fréquences mises à l’échelle.

**10.** Equipement d’utilisateur (UE) comprenant un codeur (20) selon la revendication 8 ou 9.

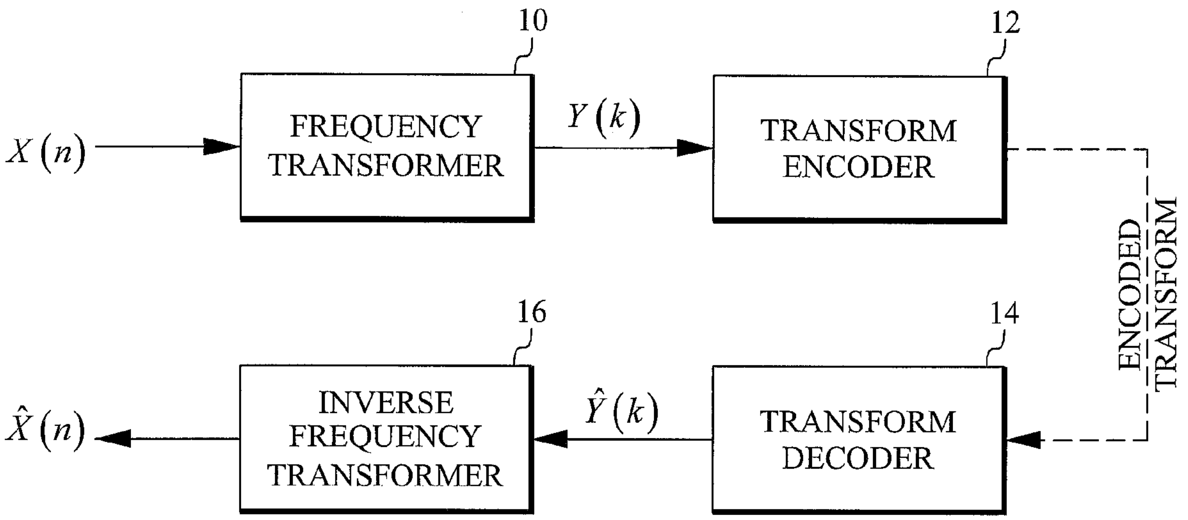
10
12
$X(n)$
FREQUENCY TRANSFORMER
$Y(k)$
TRANSFORM ENCODER
ENCODED TRANSFORM
16
14
$\hat{X}(n)$
INVERSE FREQUENCY TRANSFORMER
$\hat{Y}(k)$
TRANSFORM DECODER

FIG. 1

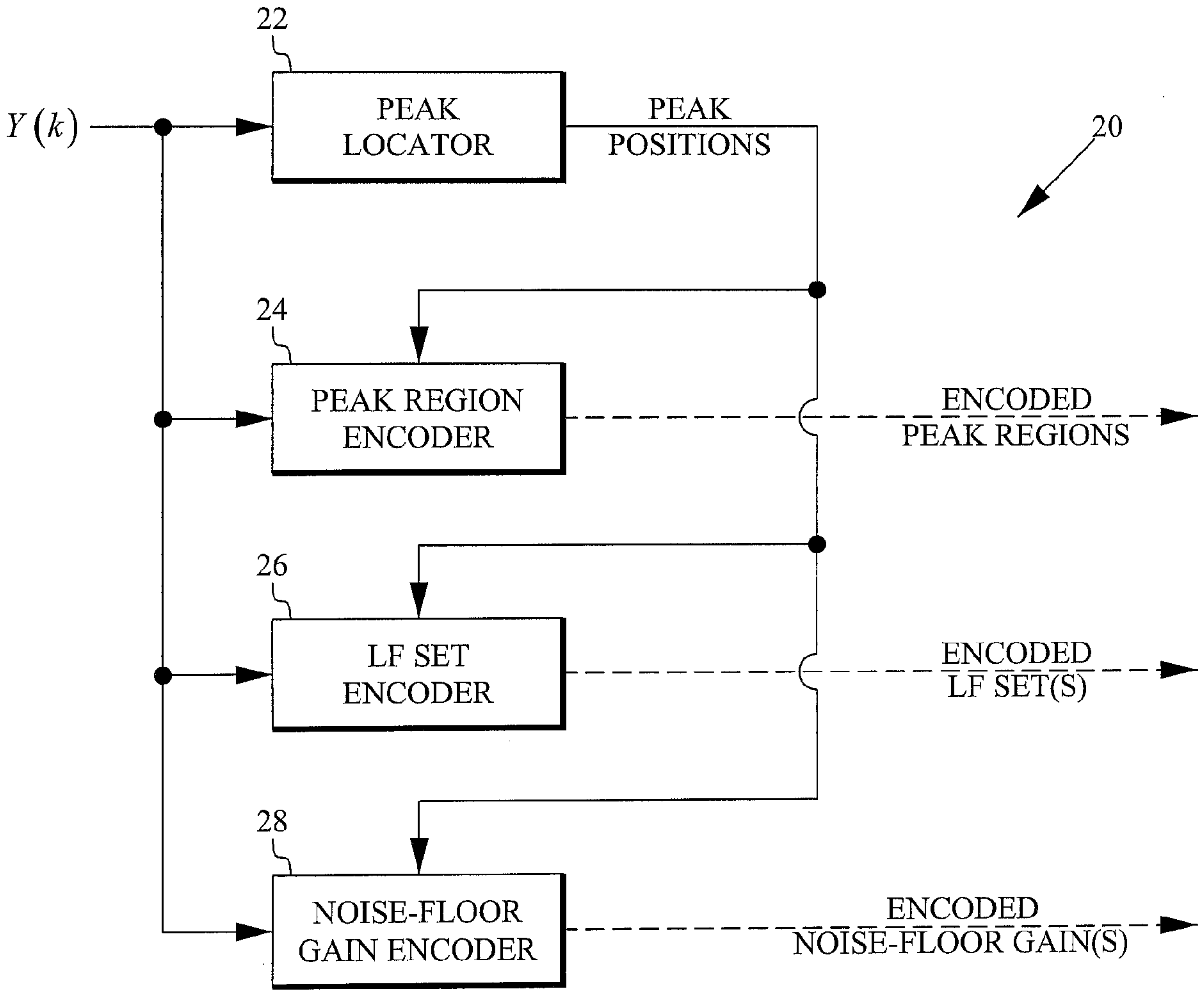
22
$Y(k)$
PEAK LOCATOR
PEAK POSITIONS
20
24
PEAK REGION ENCODER
ENCODED PEAK REGIONS
26
LF SET ENCODER
ENCODED LF SET(S)
28
NOISE-FLOOR GAIN ENCODER
ENCODED NOISE-FLOOR GAIN(S)

FIG. 7

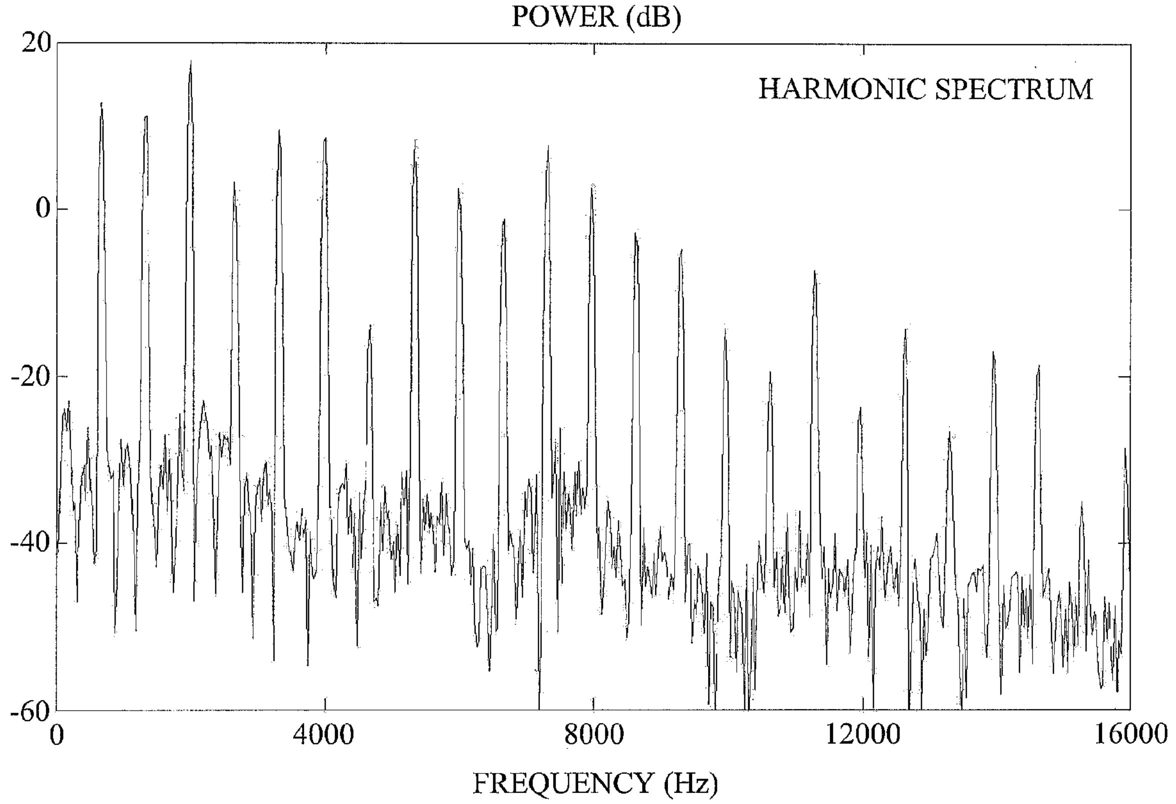
POWER (dB)
HARMONIC SPECTRUM
20
0
-20
-40
-60
0
4000
8000
12000
16000
FREQUENCY (Hz)

Fig. 2

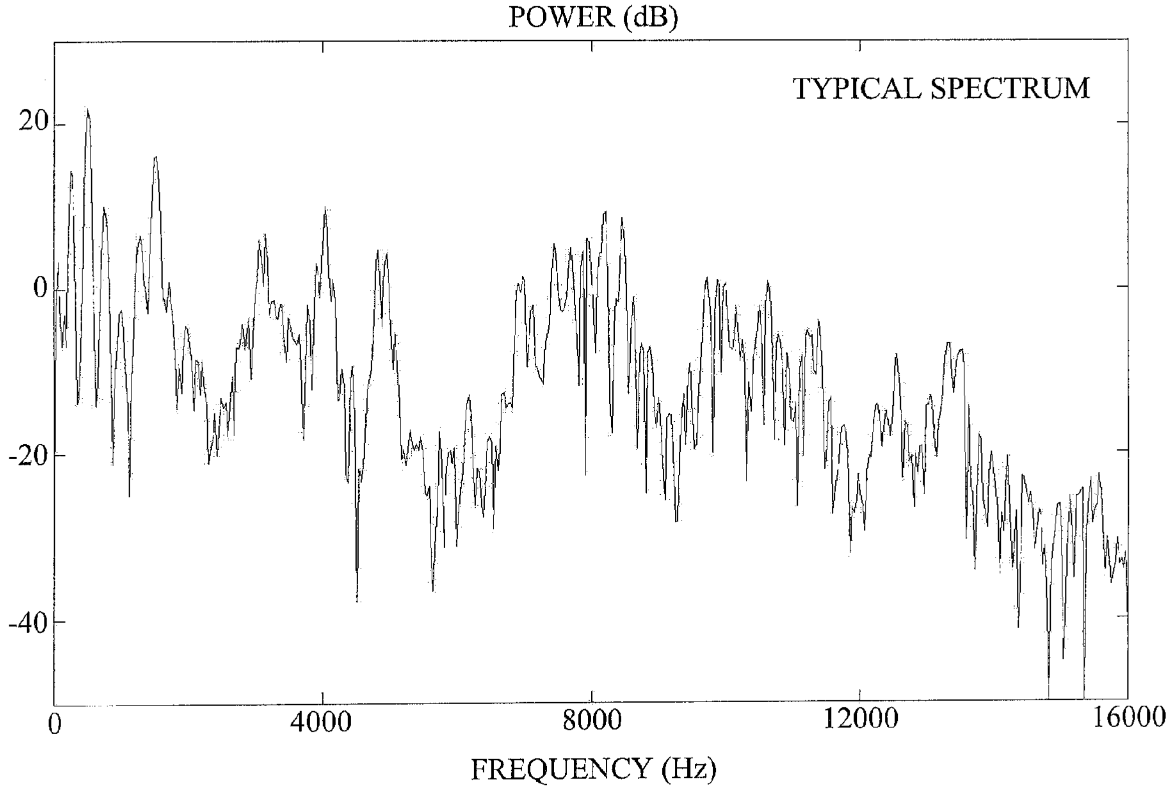
POWER (dB)
TYPICAL SPECTRUM
20
0
-20
-40
0
4000
8000
12000
16000
FREQUENCY (Hz)

Fig. 3

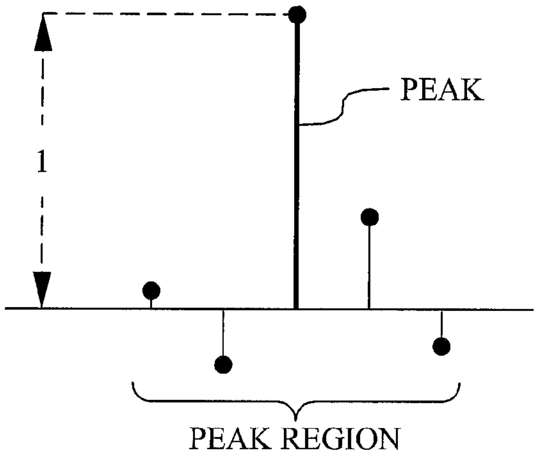
PEAK
1
PEAK REGION

FIG. 4

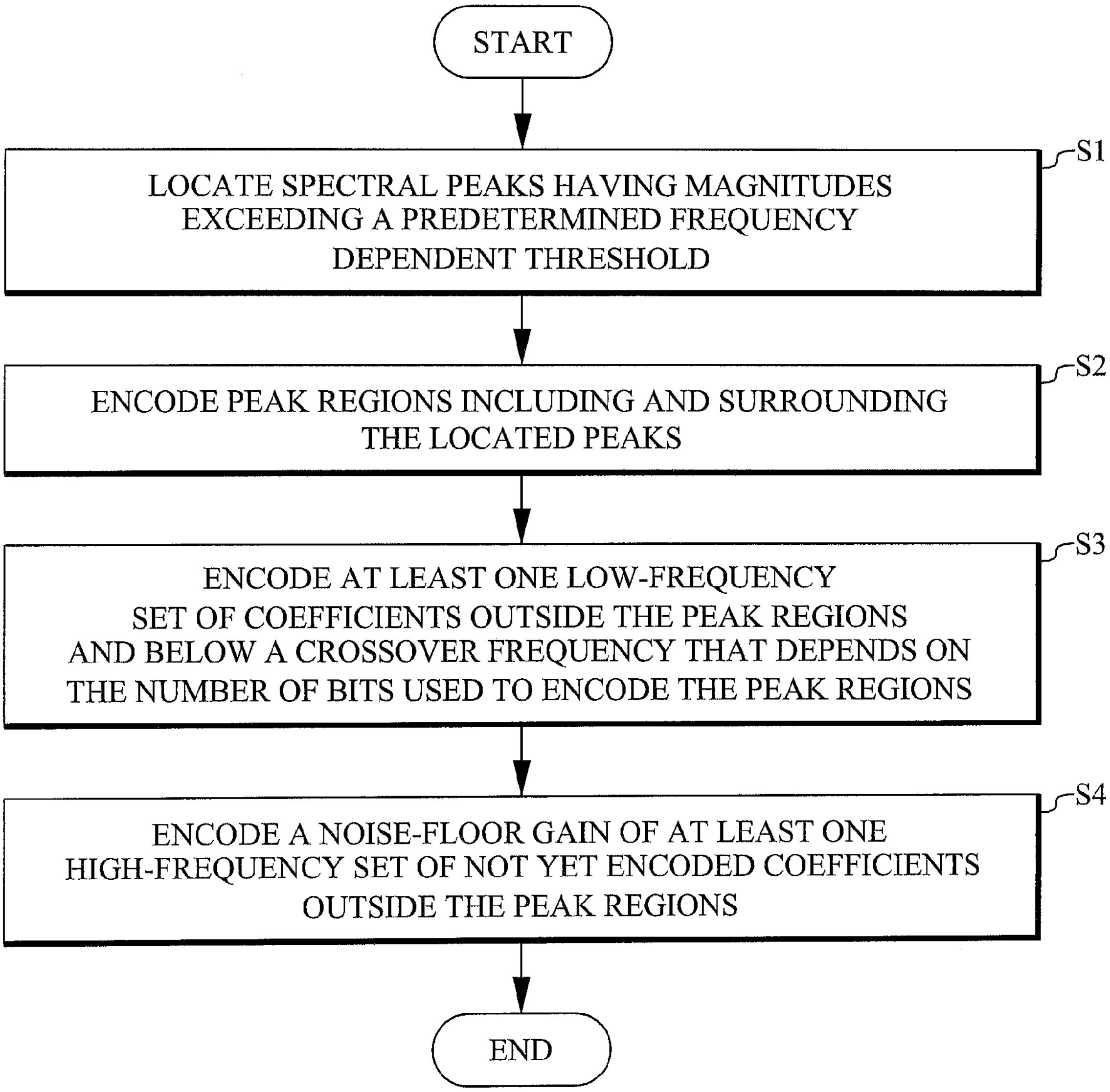
START
LOCATE SPECTRAL PEAKS HAVING MAGNITUDES EXCEEDING A PREDETERMINED FREQUENCY DEPENDENT THRESHOLD
S1
ENCODE PEAK REGIONS INCLUDING AND SURROUNDING THE LOCATED PEAKS
S2
ENCODE AT LEAST ONE LOW-FREQUENCY SET OF COEFFICIENTS OUTSIDE THE PEAK REGIONS AND BELOW A CROSSOVER FREQUENCY THAT DEPENDS ON THE NUMBER OF BITS USED TO ENCODE THE PEAK REGIONS
S3
ENCODE A NOISE-FLOOR GAIN OF AT LEAST ONE HIGH-FREQUENCY SET OF NOT YET ENCODED COEFFICIENTS OUTSIDE THE PEAK REGIONS
S4
END

FIG. 5

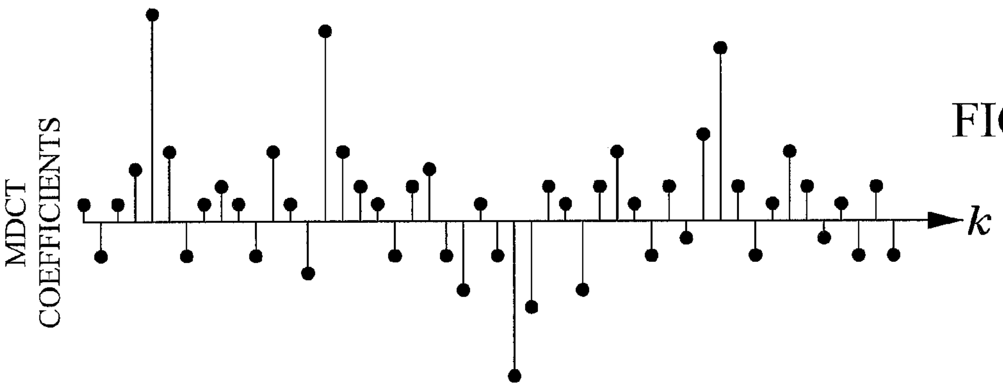

FIG. 6A

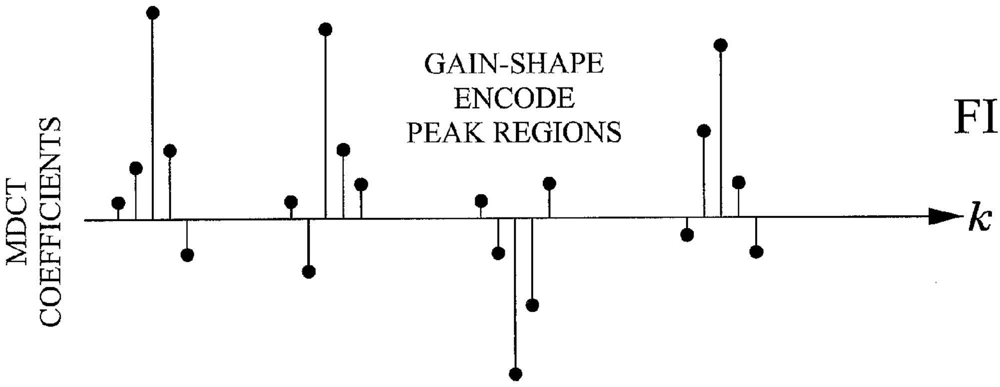

FIG. 6B

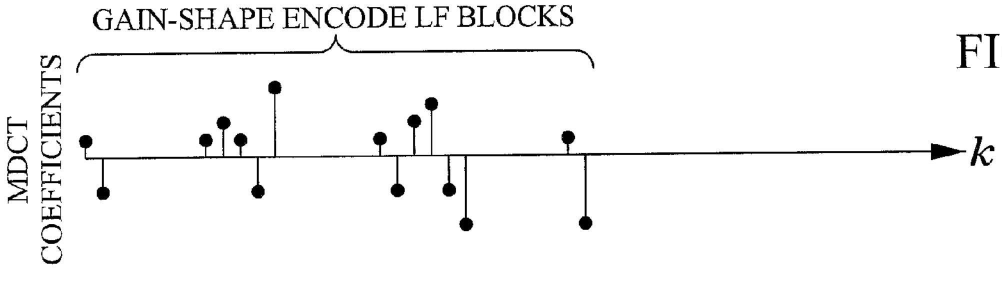

FIG. 6C

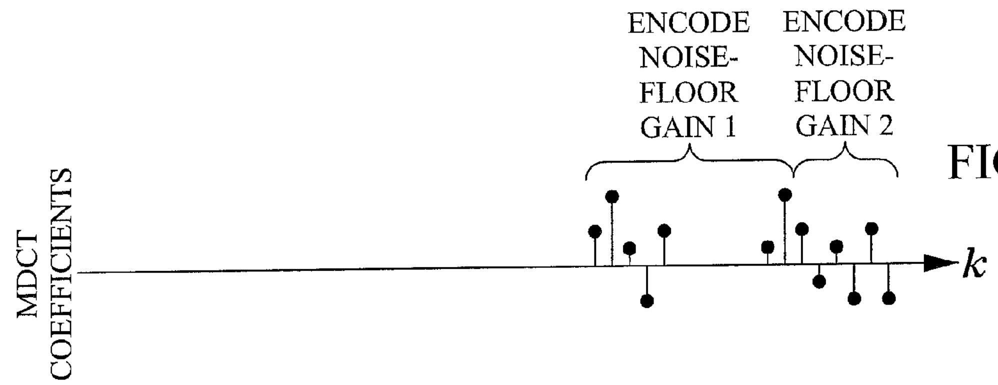

FIG. 6D

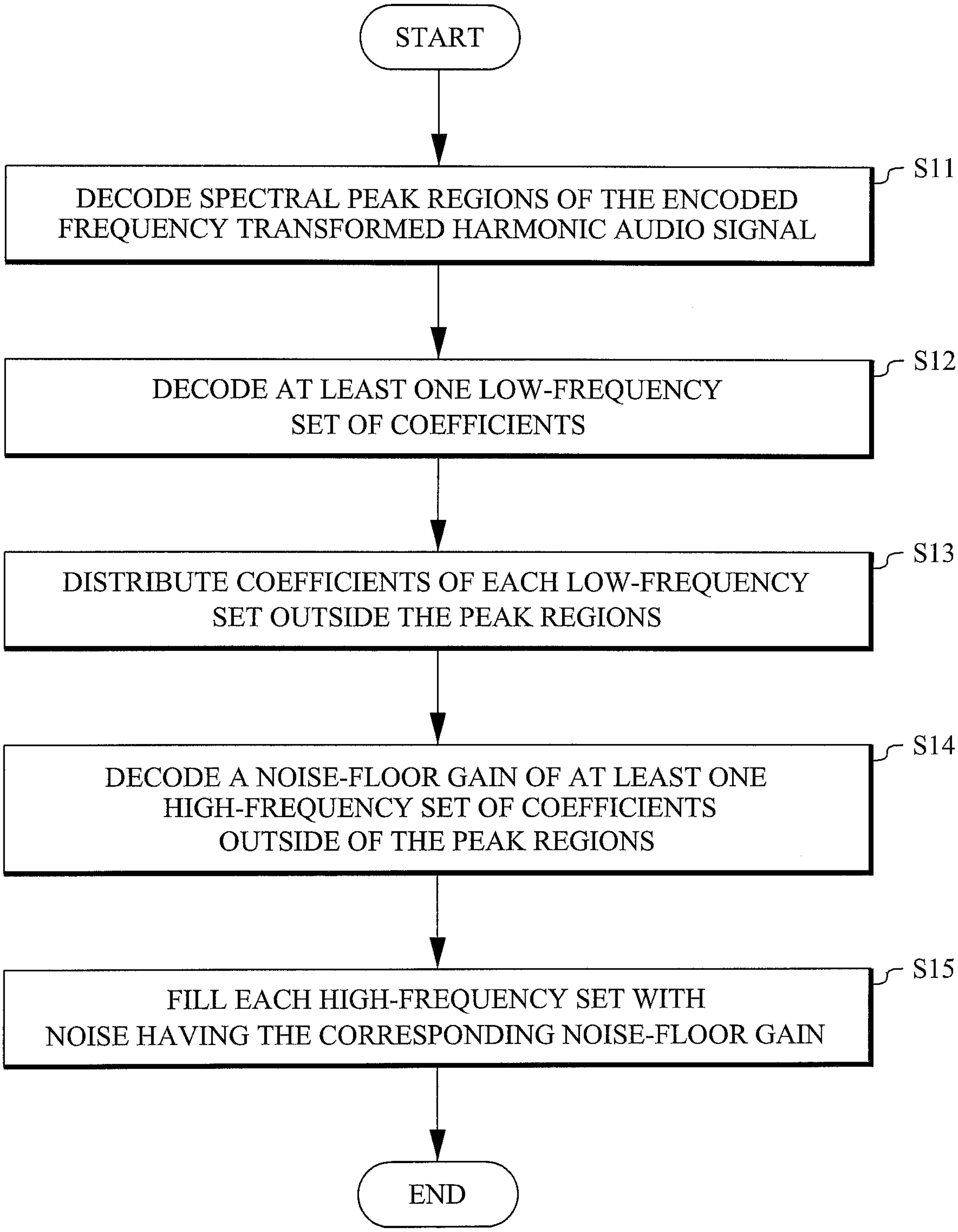
START
DECODE SPECTRAL PEAK REGIONS OF THE ENCODED FREQUENCY TRANSFORMED HARMONIC AUDIO SIGNAL
S11
DECODE AT LEAST ONE LOW-FREQUENCY SET OF COEFFICIENTS
S12
DISTRIBUTE COEFFICIENTS OF EACH LOW-FREQUENCY SET OUTSIDE THE PEAK REGIONS
S13
DECODE A NOISE-FLOOR GAIN OF AT LEAST ONE HIGH-FREQUENCY SET OF COEFFICIENTS OUTSIDE OF THE PEAK REGIONS
S14
FILL EACH HIGH-FREQUENCY SET WITH NOISE HAVING THE CORRESPONDING NOISE-FLOOR GAIN
S15
END

FIG. 8

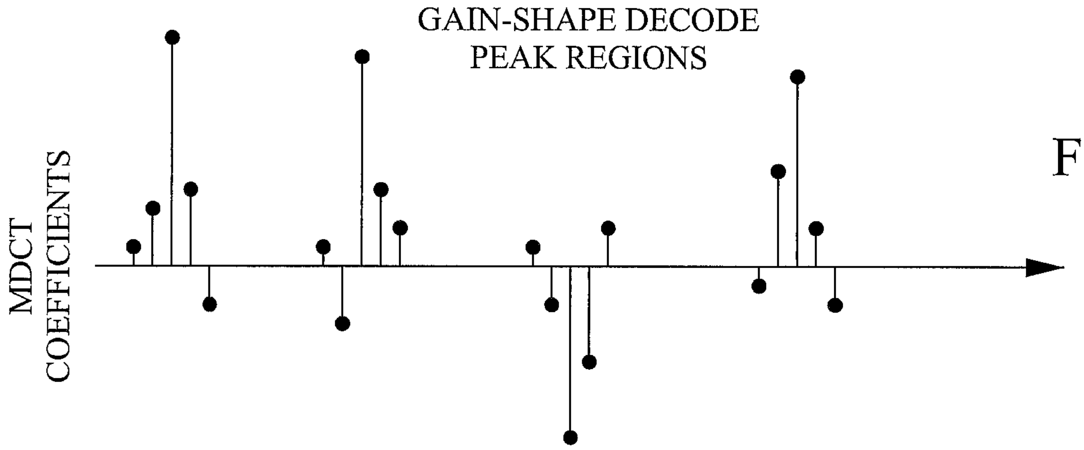

FIG. 9A

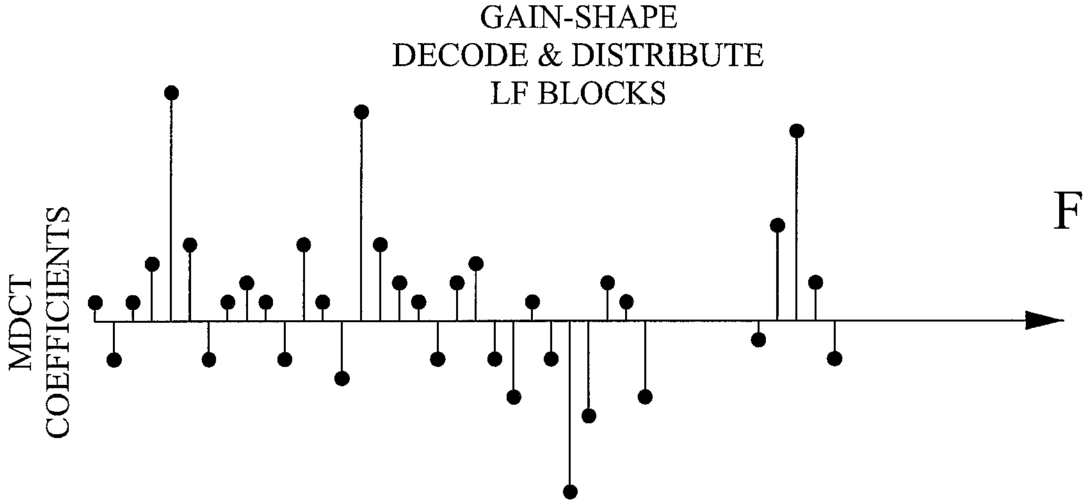

FIG. 9B

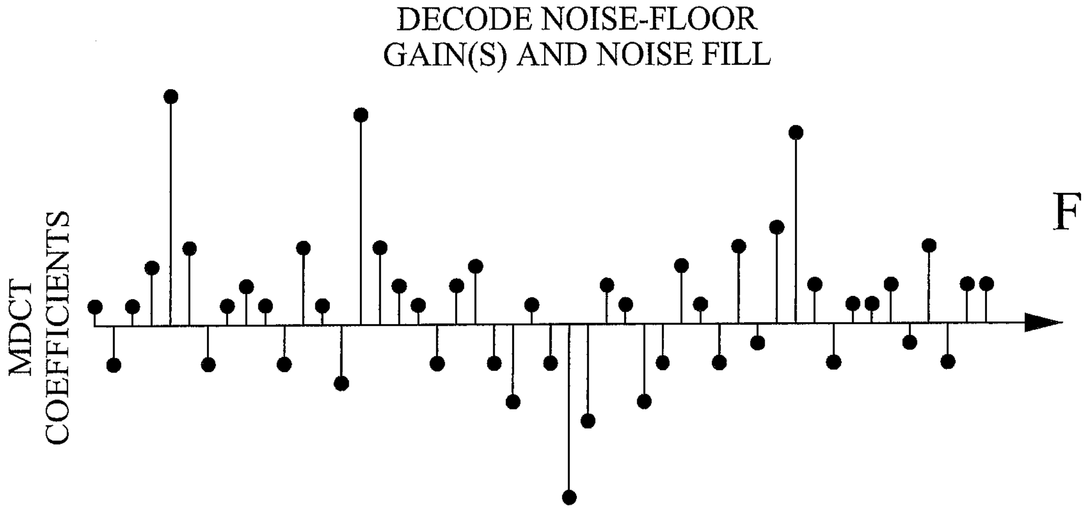

FIG. 9C

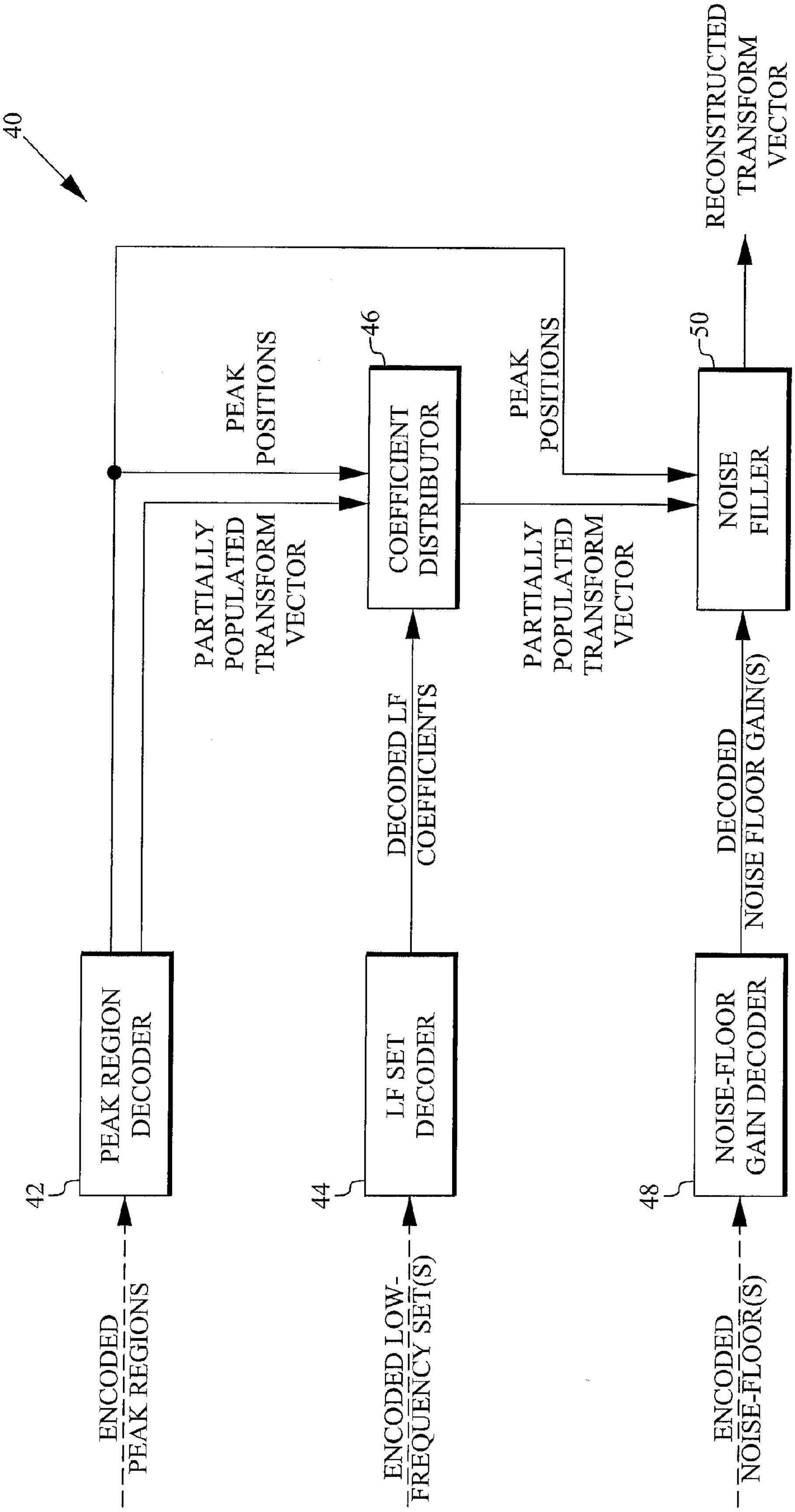
40
ENCODED PEAK REGIONS
42
PEAK REGION DECODER
PEAK POSITIONS
PARTIALLY POPULATED TRANSFORM VECTOR
46
COEFFICIENT DISTRIBUTOR
ENCODED LOW-FREQUENCY SET(S)
44
LF SET DECODER
DECODED LF COEFFICIENTS
PEAK POSITIONS
PARTIALLY POPULATED TRANSFORM VECTOR
50
NOISE FILLER
RECONSTRUCTED TRANSFORM VECTOR
ENCODED NOISE-FLOOR(S)
48
NOISE-FLOOR GAIN DECODER
DECODED NOISE FLOOR GAIN(S)

FIG. 10

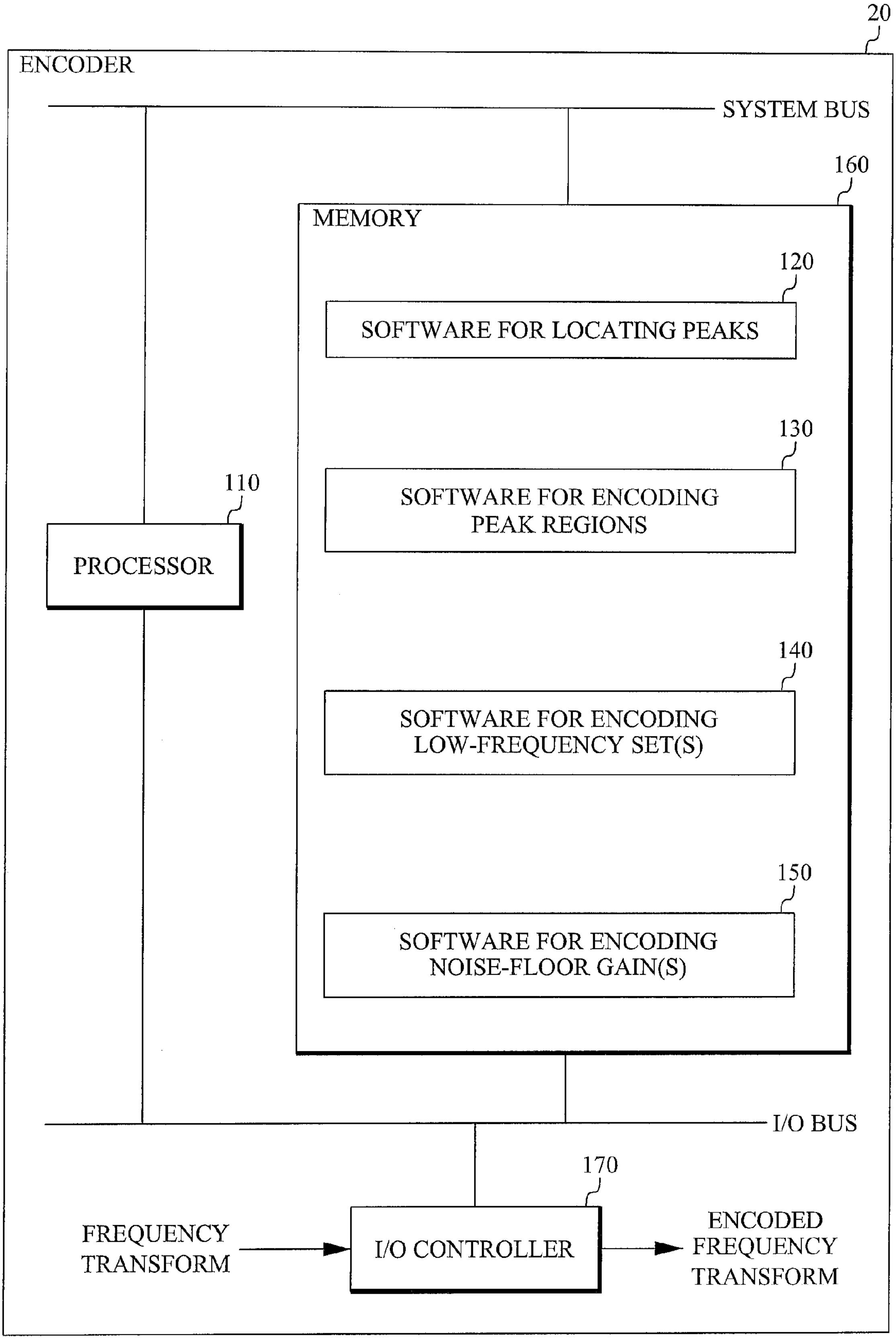
20
ENCODER
SYSTEM BUS
160
MEMORY
120
SOFTWARE FOR LOCATING PEAKS
130
SOFTWARE FOR ENCODING PEAK REGIONS
110
PROCESSOR
140
SOFTWARE FOR ENCODING LOW-FREQUENCY SET(S)
150
SOFTWARE FOR ENCODING NOISE-FLOOR GAIN(S)
I/O BUS
170
FREQUENCY TRANSFORM
I/O CONTROLLER
ENCODED FREQUENCY TRANSFORM

FIG. 11

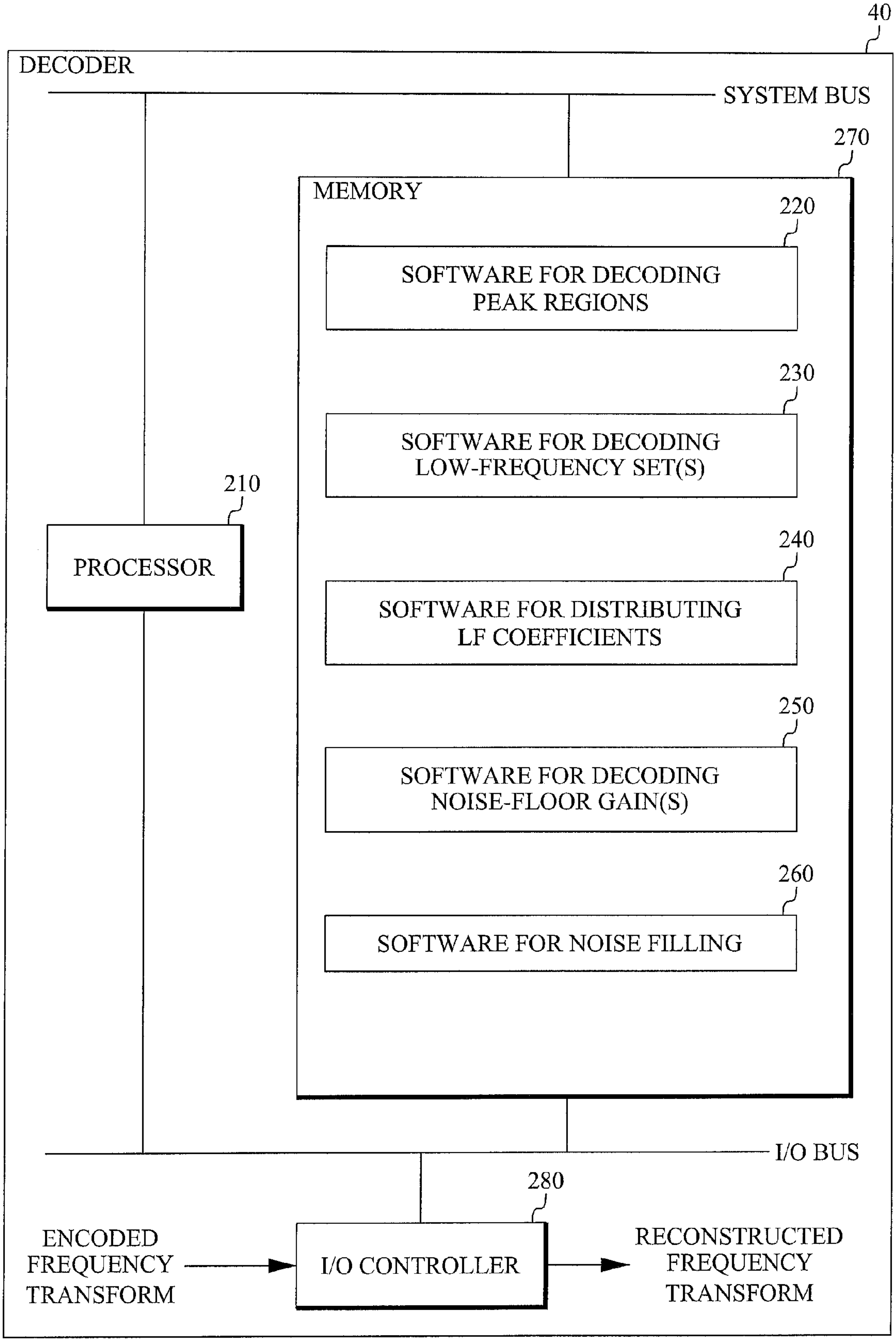
40
DECODER
SYSTEM BUS
270
MEMORY
220
SOFTWARE FOR DECODING PEAK REGIONS
230
SOFTWARE FOR DECODING LOW-FREQUENCY SET(S)
210
PROCESSOR
240
SOFTWARE FOR DISTRIBUTING LF COEFFICIENTS
250
SOFTWARE FOR DECODING NOISE-FLOOR GAIN(S)
260
SOFTWARE FOR NOISE FILLING
I/O BUS
280
ENCODED FREQUENCY TRANSFORM
I/O CONTROLLER
RECONSTRUCTED FREQUENCY TRANSFORM

FIG. 12

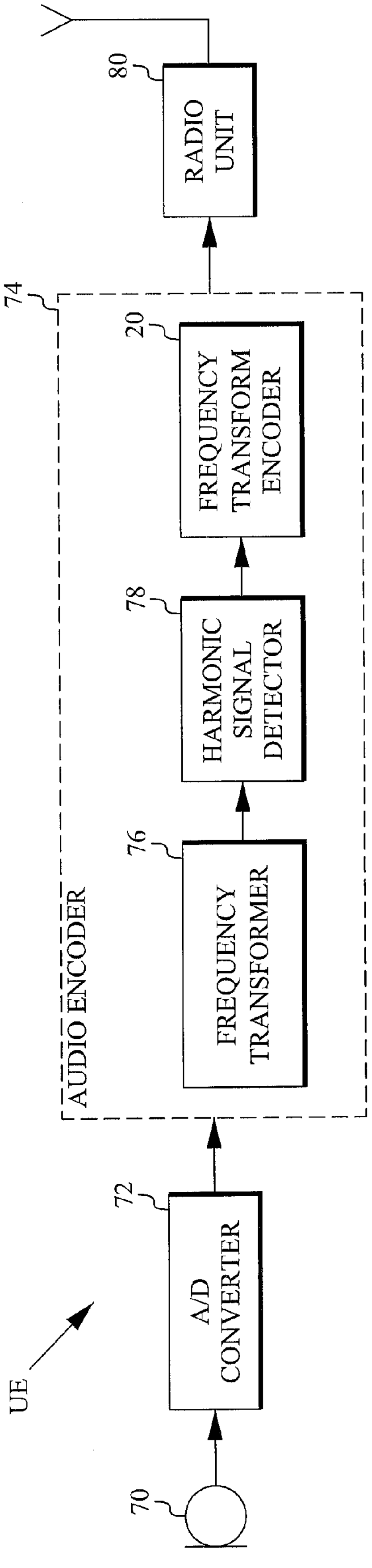
UE
70
72
A/D CONVERTER
74
AUDIO ENCODER
76
FREQUENCY TRANSFORMER
78
HARMONIC SIGNAL DETECTOR
20
FREQUENCY TRANSFORM ENCODER
80
RADIO UNIT

FIG. 13

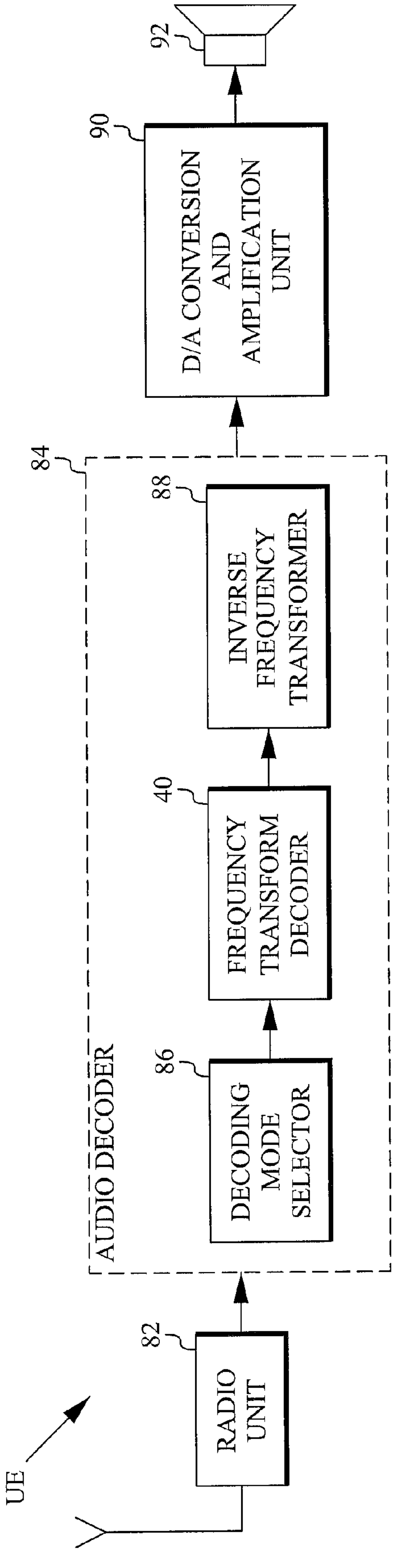
UE
82
RADIO UNIT
84
AUDIO DECODER
86
DECODING MODE SELECTOR
40
FREQUENCY TRANSFORM DECODER
88
INVERSE FREQUENCY TRANSFORMER
90
D/A CONVERSION AND AMPLIFICATION UNIT
92

FIG. 14

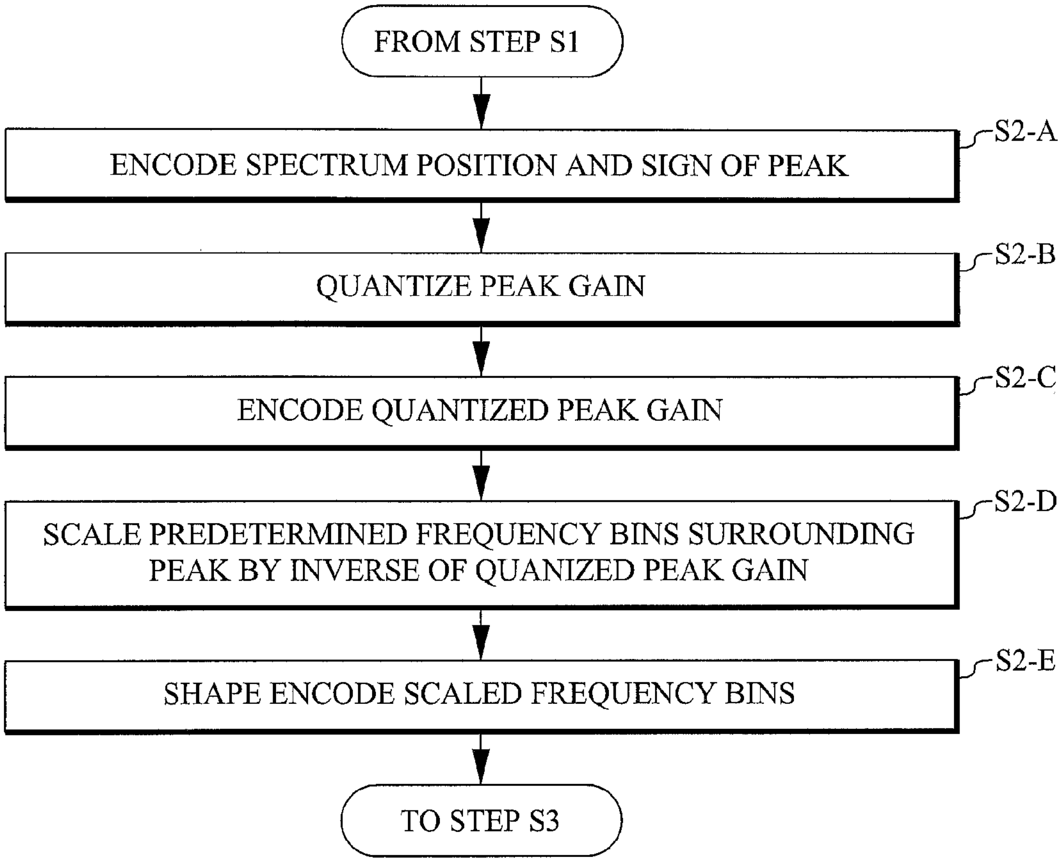
FROM STEP S1
ENCODE SPECTRUM POSITION AND SIGN OF PEAK
S2-A
QUANTIZE PEAK GAIN
S2-B
ENCODE QUANTIZED PEAK GAIN
S2-C
SCALE PREDETERMINED FREQUENCY BINS SURROUNDING PEAK BY INVERSE OF QUANIZED PEAK GAIN
S2-D
SHAPE ENCODE SCALED FREQUENCY BINS
S2-E
TO STEP S3

FIG. 15

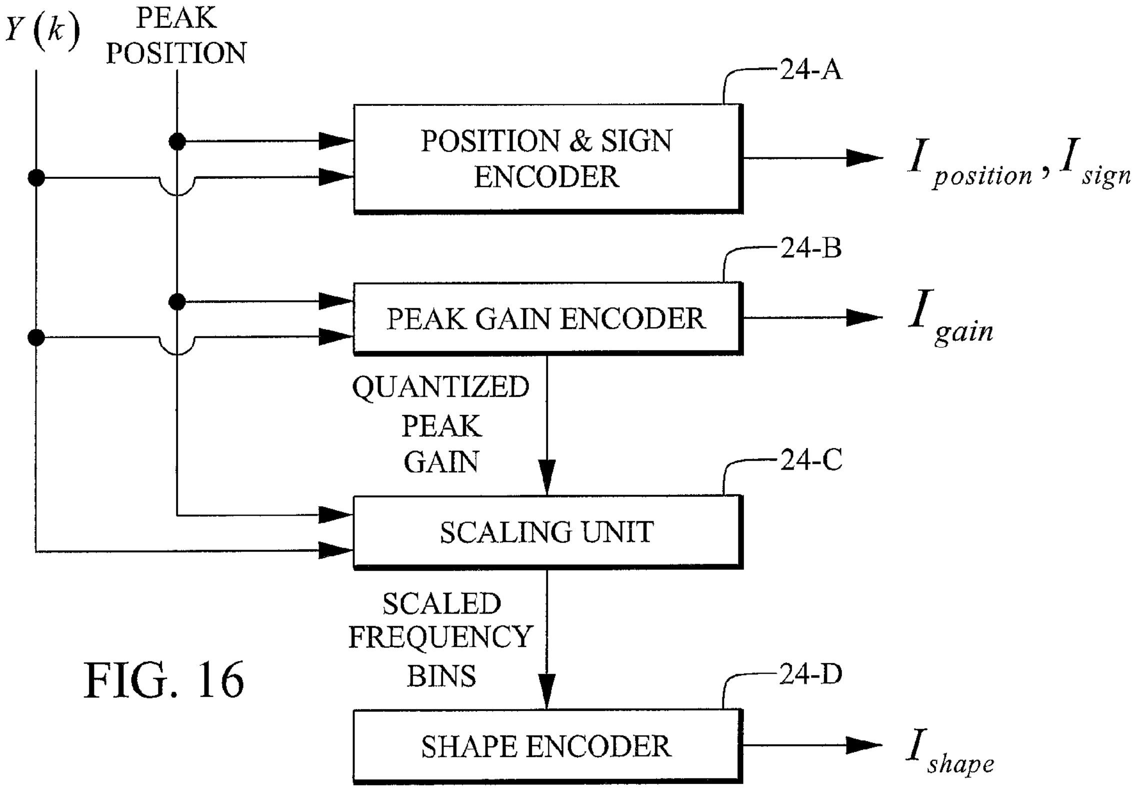
Y(k)
PEAK POSITION
POSITION & SIGN ENCODER
24-A
I position, I sign
PEAK GAIN ENCODER
24-B
I gain
QUANTIZED PEAK GAIN
SCALING UNIT
24-C
SCALED FREQUENCY BINS
SHAPE ENCODER
24-D
I shape

FIG. 16

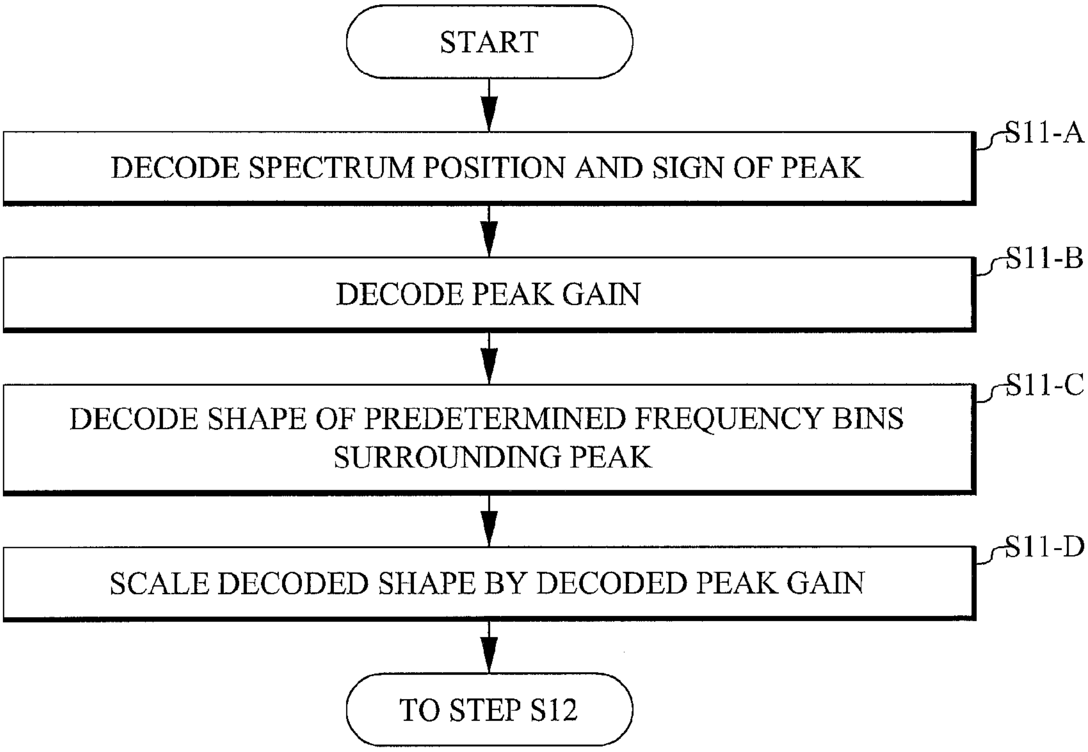
START
DECODE SPECTRUM POSITION AND SIGN OF PEAK
S11-A
DECODE PEAK GAIN
S11-B
DECODE SHAPE OF PREDETERMINED FREQUENCY BINS SURROUNDING PEAK
S11-C
SCALE DECODED SHAPE BY DECODED PEAK GAIN
S11-D
TO STEP S12

FIG. 17

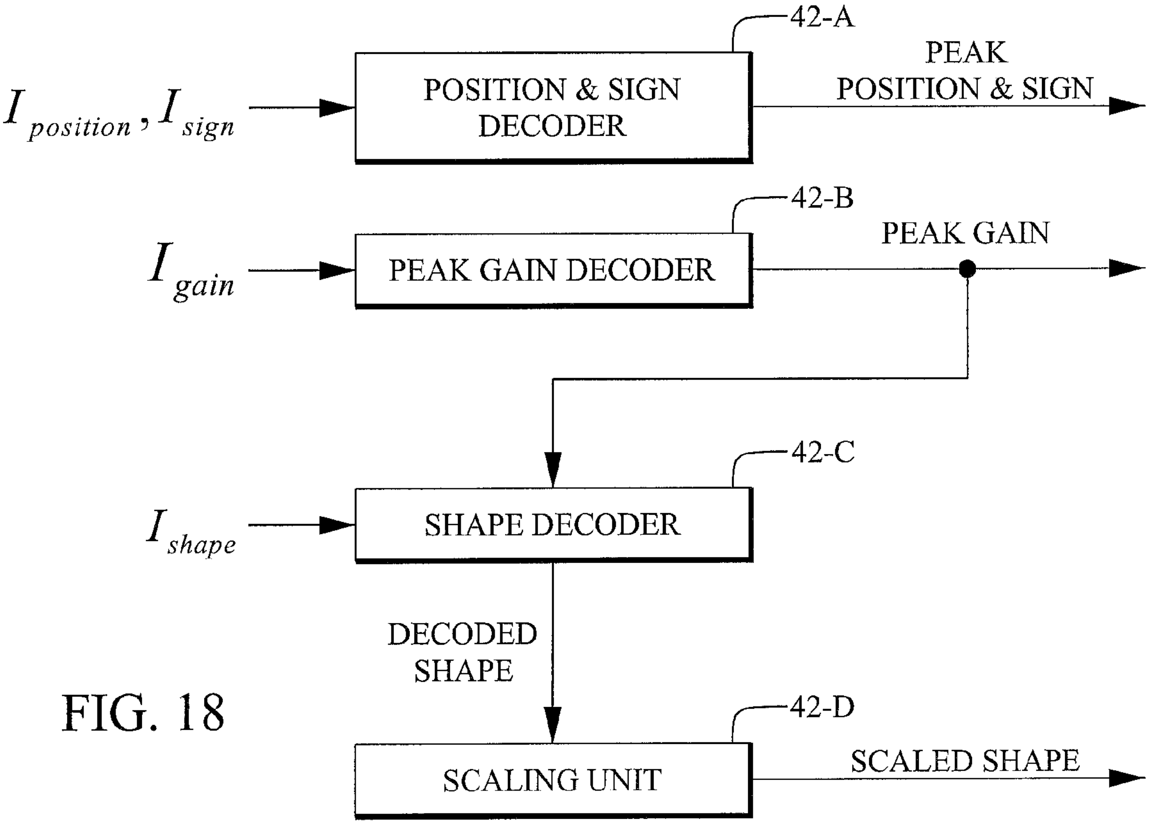
42-A
$I_{position}, I_{sign}$
POSITION & SIGN DECODER
PEAK POSITION & SIGN
42-B
$I_{gain}$
PEAK GAIN DECODER
PEAK GAIN
42-C
$I_{shape}$
SHAPE DECODER
DECODED SHAPE
42-D
SCALING UNIT
SCALED SHAPE

FIG. 18

## REFERENCES CITED IN THE DESCRIPTION

### Patent documents cited in the description

- US 20120029923 A **[0006]**

**716467/KOT**

# HARMONIKUS HANGJELEK ÁTALAKÍTÁSI KÓDOLÁSA/DEKÓDOLÁSA

## Szabadalmi igénypontok

**1.** Eljárás harmonikus hangjel MDCT módosított diszkrét koszinusz transzformációs együtthatóinak ($Y(k)$) kódolására, az eljárás a következő lépéseket tartalmazza:

> előre meghatározott küszöbértéket túllépő mennyiségekkel rendelkező spektrális csúcsok helymeghatározása (S1), ahol a spektrális csúcsok helymeghatározása az együtthatóknak a küszöbhöz való hasonlításával történik, hogy csúcsjelölt vektort képezzünk, és kivonjuk az elemeket a csúcsjelöltek vektorból, ahol a küszöb kiszámítása:

$$\theta = \left(\frac{\bar{E}_p}{\bar{E}_{nf}}\right)^{\gamma} \bar{E}_{nf} \,,$$

ahol $\bar{E}_p$ az átlagos csúcsenergia, $\bar{E}_{nf}$ az átlagos zajszintű energia, γ pedig rögzített előre meghatározott értékkel rendelkezik, és ahol a csúcsenergia kiszámítása: $E_p(k)=\beta E_p(k)+(1-\beta)|Y(k)|$ a zajszintű energia kiszámítása pedig: $E_{nf}(k)=\alpha E_{nf}(k)+(1-\alpha)|Y(k)|$, ahol a nagy energiás együtthatók hozzájárulása ki van emelve a csúcsenergia kiszámításában, a kis energiás együtthatók hozzájárulása pedig ki van emelve a zajszintű energia kiszámításában;
csúcsterületek kódolása (S2), beleértve és körbevéve a helymeghatározott csúcsokat, ahol a spektrális csúcsok együtt kvantálódnak a szomszédos MDCT tárolókkal;
a csúcsterületeken kívüli és olyan átkapcsolási gyakoriság alatti együtthatók LF kisfrekvenciás halmazának kódolása (S3) bizonyos számú fenntartott bitek segítségével, amely a csúcsterületek kódolásához használt bitek számától függ, ahol a kódolás (S3) tartalmazza a csúcsterületeken kívüli együtthatók egy vagy több kisfrekvenciás halmazának kódolását, ha nem fenntartott bitek állnak rendelkezésre a csúcsterületek kódolása után;
a csúcsterületeken kívüli még nem kódolt együtthatók legalább egy nagyfrekvenciás halmaza zajszintű nyereségének kódolása (S4) bizonyos számú fenntartott bit segítségével.

**2.** Az 1. igénypont szerinti eljárás, ahol az α súlyozási tényező meghatározása:

$$\alpha = \begin{cases} 0{,}9578 & si \quad |Y(k)| > E_{nf}(k-1) \\ 0{,}6472 & si \quad |Y(k)| \leq E_{nf}(k-1) \end{cases}'$$

a β súlyozási tényező meghatározása pedig

$$\beta = \begin{cases} 0{,}4223 & si \quad |Y(k)| > E_p(k-1) \\ 0{,}8029 & si \quad |Y(k)| \leq E_p(k-1) \end{cases}.$$

**3.** Az 1. vagy 2. igénypont szerinti kódolási eljárás, ahol a csúcsterületek kódolásának lépése (S2) a következőket tartalmazza:

egy csúcs spektrumhelyének és előjelének kódolása (S2-A);
csúcsnyereség kvantálása (S2-B);
a kvantált csúcsnyereség kódolása (S2-C);
a csúcsot körülvevő előre meghatározott frekvenciatárolók skálázása (S2-D) a kvantált csúcsnyereség fordítottjával;
a skálázott frekvenciatárolók alakkódolása (S2-E).

**4.** Az 1 – 3. igénypontok egyike szerinti kódolási eljárás, ahol a csúcsterület tartalmazza a csúcsot és a csúcsot körülvevő négy MDCT tárolót.

**5.** Az előző igénypontok egyike szerinti kódolási eljárás, ahol az együtthatók kisfrekvenciás halmazának kódolási lépése (S3) tartalmazza a fennmaradó nem kvantált MDCT együtthatók 24-dimenziós sávokba való csoportosítását.

**6.** Az előző igénypontok egyike szerinti kódolási eljárás, ahol egy kisfrekvenciás halmaz kódolása nyereségalakú kódolási sémán alapul, a nyereségalapú kódolási séma skaláris nyereségkvantáláson és faktoros impulzusalakú kódoláson alapul.

**7** Az előző igénypontok egyike szerinti kódolási eljárás, amely tartalmazza egy zajszintű nyereség kódolási lépését mindkét nagyfrekvenciás halmazra.

**8.** Kódoló harmonikus hangjel MDCT módosított diszkrét koszinusz transzformációs együtthatóinak ($Y(k)$) kódolására, a kódoló a következőket tartalmazza:

> csúcshely-meghatározó (22), amely előre meghatározott küszöbértéket túllépő mennyiségekkel rendelkező spektrális csúcsok helymeghatározására van kialakítva, ahol a spektrális csúcsok helymeghatározása az együtthatóknak a küszöbhöz való hasonlításával történik, hogy csúcsjelölt vektort képezzünk, és kivonjuk az elemeket a csúcsjelöltek vektorból, ahol a küszöb kiszámítása:

$$\theta = \left(\frac{\bar{E}_p}{\bar{E}_{nf}}\right)^{\gamma} \bar{\bar{E}}_{nf},$$

ahol $\bar{E}_p$ az átlagos csúcsenergia, $\bar{E}_{nf}$ az átlagos zajszintű energia, γ pedig rögzített előre meghatározott értékkel rendelkezik, és ahol a csúcsenergia kiszámítása: $E_p(k)=\beta E_p(k)+(1-\beta)|Y(k)|$ a zajszintű energia kiszámítása pedig: $E_{nf}(k)=\alpha E_{nf}(k)+(1-\alpha)|Y(k)|$, ahol a nagy energiás együtthatók hozzájárulása ki van emelve a csúcsenergia kiszámításában, a kis energiás együtthatók hozzájárulása pedig ki van emelve a zajszintű energia kiszámításában;

csúcsterület-kódoló (24), amely csúcsterületek kódolására van kialakítva, beleértve és körülvéve a helymeghatározott csúcsokat,

ahol a spektrális csúcsok együtt kvantálódnak a szomszédos MDCT tárolókkal;

kisfrekvenciás halmazkódoló (26), amely arra van kialakítva, hogy fenntartott bitek segítségével kódolja a csúcsterületeken kívüli és olyan átkapcsolási gyakoriság alatti területek együtthatóinak első kisfrekvenciás halmazát, amely a csúcsterületek kódolásához használt bitek számától függ, valamint kódolja a csúcsterületeken kívüli együtthatók egy vagy több további kisfrekvenciás halmazát, ha vannak a csúcsterületek kódolása után rendelkezésre álló nem fenntartott bitek; és

zajszintű nyereségkódoló (28), amely arra van kialakítva, hogy bizonyos számú fenntartott bit segítségével kódolja a csúcsterületeken kívüli, még nem kódolt együtthatók legalább egy nagyfrekvenciás halmazának zajszintű nyereségét.

**9.** A 8. igénypont szerinti kódoló, ahol a csúcsterület kódoló (24) a következőket tartalmazza:

hely- és előjelkódoló (24-A), amely arra van kialakítva, hogy kódolja egy csúcs spektrumhelyét ($I_{position}$) és előjelét ($I_{sign}$).

csúcsnyereség-kódoló (24-B), amely arra van kialakítva, hogy kvantálja a csúcsnyereséget és kódolja ($I_{gain}$) a kvantált csúcsnyereséget.

skálázó egység (24-C), amely arra van kialakítva, hogy a kvantált csúcsnyereség fordítottjával skálázza a csúcsot körülvevő előre meghatározott frekvenciatárolókat;

alakkódoló (24-D), amely a skálázott frekvenciatárolók alakkódolására van kialakítva.

**10.** Felhasználói berendezés (UE), beleértve egy kódolót (20) a 8. vagy 9. igénypont szerint.