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

(11) **EP 1 852 851 A1**

(12) **EUROPEAN PATENT APPLICATION**
published in accordance with Art. 158(3) EPC

(43) Date of publication:
07.11.2007 Bulletin 2007/45

(21) Application number: **05738242.6**

(22) Date of filing: **01.04.2005**

(51) Int Cl.:
G10L 19/04 *(2006.01)*

(86) International application number:
PCT/CN2005/000441

(87) International publication number:
WO 2005/096274 (13.10.2005 Gazette 2005/41)

(84) Designated Contracting States:
AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IS IT LI LT LU MC NL PL PT RO SE SI SK TR

(30) Priority: **01.04.2004 CN 200410030945**

(71) Applicants:
- **BEIJING MEDIA WORKS CO., LTD**
  Haidian District
  Beijing
  100044 (CN)
- **Beijing E-World Technology Co. Ltd.**
  Beijing 100044 (CN)
- **Coding Technologies AB**
  113 52 Stockholm (SE)

(72) Inventors:
- **PAN, Xingde**
  Xicheng District,
  Beijing 100044 (CN)
- **MARTIN, Dietz**
  90429 Nürnberg (DE)
- **EHRET, Andreas**
  90429 Nürnberg (DE)
- **HÖRICH, Holger**
  90429 Nürnberg (DE)
- **ZHU, Xiaoming**
  Xicheng District,
  Beijing 100044 (CN)
- **SCHUG, Michael**
  90429 Nürnberg (DE)
- **REN, Weimin**
  Xicheng District,
  Beijing 100044 (CN)
- **WANG, Lei**
  Xicheng District,
  Beijing 100044 (CN)
- **DENG, Hao**
  Xicheng District,
  Beijing 100044 (CN)
- **HENN, Fredrik**
  S-113 52 Stockholm (SE)

(74) Representative: **Zinkler, Franz et al**
Schoppe, Zimmermann, Stöckeler & Zinkler
Postfach 246
82043 Pullach bei München (DE)

(54) AN ENHANCED AUDIO ENCODING/DECODING DEVICE AND METHOD

(57) An enhanced audio encoding device, which is consisted of a signal type analyzing module, a psychoacoustical analyzing module, a time-frequency mapping module, a quantization and entropy encoding module, a frequency-domain linear prediction and vector quantization module, and a bit stream multiplexing module, in which the singal type analyzing module is configured to anaylze the signal type of the input audio signal and output to the psychoacoustical analyzing module, the time-frequency mapping module and the bit stream multiplexing module, the frequency-domain linear prediction and vector quantization module is configured to perform a linear prediction of the frequency-domain coefficients and a multi-level vector quantization, and output the residual sequences to the quantization and entropy encoding module while outputting a lateral information to the bit stream multiplexing module. The device is adapted to perform a compressive encoding of the audio signal with high fidelity, which is sampled at multi sampling rates and has multi audio channel configurations, the device can support the audio signal whose sampling rate is between 8kHz and 192kHz and all possible audio channel configurations, and the range of the objective code rate of the audio encode/decode is very wide.

EP 1 852 851 A1

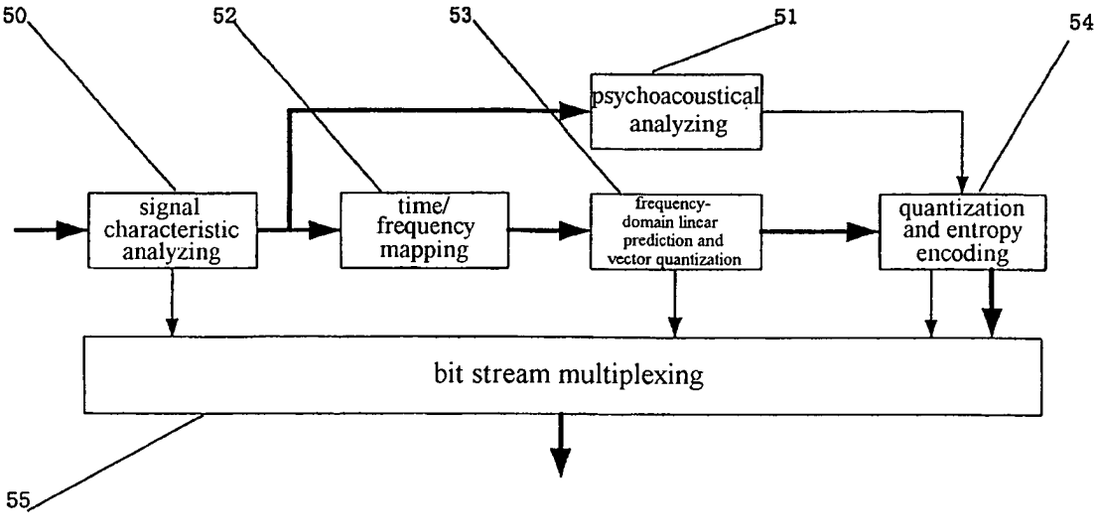
50
52
53
51
54
psychoacoustical analyzing
signal characteristic analyzing
time/ frequency mapping
frequency-domain linear prediction and vector quantization
quantization and entropy encoding
bit stream multiplexing
55

Fig. 5

Description

Technical Field

[0001] The invention relates to audio encoding and decoding, and in particular, to an enhanced audio encoding/decoding device and method based on a sensor model.

Background Art

[0002] In order to obtain Hi-Fi digital audio signals, the digital audio signals need to be audio encoded or audio compressed for storage and transmission. The object of encoding the audio signals is to realize transparent representation thereof by using as less number of bits as possible, for example, the originally input audio signals are almost the same as the output audio signals after being encoded.

[0003] In early 1980s, CD came into existence, which reflects many advantages of representing the audio signals by digits, such as high fidelity, large dynamic range and great robustness. However, all these advantages are achieved at the cost of a very high data rate. For example, the sampling rate requested by the digitization of the stereo signal of CD quality is 44.1kHz, and each sampling value has to be uniformly quantized by 15 bits, thus the non-compressed data rate reaches 1. 41Mb/s which brings great inconvenience to the transmission and storage of data, and the transmission and storage of data are limited by the bandwidth and cost especially in the situation of multimedia application and wireless transmission application. In order to maintain high-quality audio signals, the data rate in new network and wireless multimedia digital audio system must be reduced without damaging the quality of the audio. With respect to said problem, various audio compression techniques have been put forward that can both obtain high compression ratio and generate hi-fi audio signals, among which the typical ones are the MPEG-1/-2/-4 technique of ISO/IEC, AC-2/AC-3 technique of Dolby, ATRAC/MiniDisc/SDDS technique of Sony, and PAC/EPAC/MPAC technique of Lucent Technologies, etc. The MPEG-2 AAC technique and the AC-3 technique of Dolby are described specifically below.

[0004] MPEG-1 and MPEG-2 BC techniques are high sound quality encoding technique mainly used for mono and stereo audio signals. With the increasing demand in the multi-channel audio encoding that achieves high encoding quality at a relatively low code rate, since the MPEG-2 BC encoding technique gives emphasis to backward compatibility with the MPEG-1 technique, it is impossible to realize high sound quality encoding of five sound channels at a code rate lower than 540kbps. With respect to this shortage, the MPEG-2 AAC technique was put forward, which can realize a high quality encoding of the five channel signals at a rate of 320kbps.

[0005] Fig. 1 is a block diagram of the MPEG-2 AAC encoder. Said encoder comprises a gain controller 101, a filter bank 102, a time-domain noise shaping module 103, an intensity/coupling module 104, a psychoacoustical model, a second order backward adaptive predictor 105, a sum-difference stereo module 106, a bit allocation and quantization encoding module 107, and a bit stream multiplexing module 108, wherein the bit allocation and quantization encoding module 107 further comprises a compression ratio/distortion processing controller, a scale factor module, a non-uniform quantizer, and an entropy encoding module.

[0006] The filter bank 102 uses a modified discrete cosine transformation (MDCT), whose resolution is signal-adaptive, that is, an MDCT transformation of 2048 dots is used for the steady state signal, while an MDCT transformation of 256 dots is used for the transient state signal, thus for a signal sampled at 48kHz, the maximum frequency resolution is 23Hz and the maximum time resolution is 2. 6ms. Meanwhile, sine window and Kaiser-Bessel window can be used in the filter bank 102, and the sine window is used when the harmonic wave interval of the input signal is less than 140Hz, while the Kaiser-Bessel window is used when the strong component interval in the input signal is greater than 220Hz.

[0007] Audio signals enter the filter bank 102 through the gain controller 101, and are filtered according to the different signals, then the time-domain noise shaping module 103 processes the frequency spectrum coefficients output by the filter bank 102. The time-domain noise shaping technique performs linear prediction analysis on the frequency spectrum coefficients in the frequency domain, then controls the shape of the quantized noise in the time domain according to said analysis to thereby control the pre-echo.

[0008] The intensity/coupling module 104 is used for stereo encoding of the signal intensity. With respect to signals of high frequency channel (greater than 2kHz), the sense of direction of audition is related to the change in the relevant signal intensity (signal envelope), but is irrelevant to the waveform of the signal, that is, a constant envelope signal has no influence on the sense of direction of audition. Therefore, this characteristic and the relevant information among multiple sound channels can be utilized to combine several sound channels into one common sound channel to be encoded, thereby forming the intensity/coupling technique.

[0009] The second order backward adaptive predictor 105 is used for removing the redundancy of the steady state signal and improving the encoding efficiency. The sum-difference stereo (M/S) module 106 operates on sound channel pairs. The sound channel pair refers to the two sound channels of the left-right sound channels or the left-righ't surround sound channels in, for example, dual sound channel signal or multiple sound channel signal. The M/S module 106

achieves the effect of reducing code rate and improving encoding efficiency by means of the correlation between the two sound channels in the sound channel pair. The bit allocation and quantization encoding module 107 is realized by a nested loop, wherein the non-uniform quantizer performs lossy encoding, while the entropy encoding module performs lossless encoding, thus removing redundancy and reducing correlation. The nested loop comprises inner layer loop and outer layer loop, wherein the inner layer loop adjusts the step size of the non-uniform quantizer until the provided bits are used up, and the outer layer loop estimates the encoding quality of signal by using the ratio between the quantized noise and the masking threshold. Finally, the encoded signals are formed into an encoded audio stream through the bit stream multiplexing module 108 to be output.

[0010] Under scalable sampling rate, four frequency bands of equal bandwidth are generated in the multi-phase filter bank of four frequency channels (PQF) while inputting signals. Each frequency band generating 256 frequency spectrum coefficients using MDCT, resulting in altogether 1024 frequency spectrum coefficients. The gain controller 101 is used in each frequency band. The high frequency PQF frequency band can be neglected in the decoder to obtain signals of low sampling rate.

[0011] Fig. 2 is a schematic block diagram of the corresponding MPEG-2 AAC decoder. Said decoder comprises a bit stream demultiplexing module 201, a lossless decoding module 202, an inverse quantizer 203, a scale factor module 204, a sum-difference stereo (M/S) module 205, a prediction module 206, an intensity/coupling module 207, a time-domain noise shaping module 208, a filter bank 209 and a gain control module 210. The encoded audio stream is demultiplexed by the bit stream demultiplexing module 201 to obtain the corresponding data stream and control stream. Said signals are then decoded by the lossless decoding module 202 to obtain an integer representation of the scale factors and the quantized values of the signal spectrum. The inverse quantizer 203 is a non-uniform quantizer bank realized by a companding function, which is used for transforming the integer quantized values into a reconstruction spectrum. The scale factor module in the encoder differentiates the current scale factor from the previous scale factor and performs a Huffman encoding on the difference, so the scale factor module 204 in the decoder can obtain the corresponding difference through Huffman encoding, from which the real scale factor can be recovered. The M/S module 205 converts the sum-difference sound channel into a left-right sound channel under the control of the side information. Since the second order backward adaptive predictor 105 is used in the encoder to remove the redundancy of the steady state signal and improve the encoding efficiency, a prediction module 206 is used in the decoder for performing prediction decoding. The intensity/coupling module 207 performs intensity/coupling decoding under the control of the side information; then outputs to the time domain noise shaping module 208 to perform time domain noise shaping decoding, and in the end, the filter bank 209 performs integration filtering by an inverse modified discrete cosine transformer (IMDCT) technique.

[0012] In the case of scalable sampling rate, the PQF frequency band of high frequency can be neglected through the gain control module 210 so as to obtain signals of low sampling rate.

[0013] The MPEG-2 AAC encoding/decoding technique is suitable for audio signals of medium and high code rate, but it has a poor encoding quality for low code rate or very low code rate audio signals; meanwhile, said encoding/decoding technique involves a lot of encoding/decoding modules, so it is highly complex in implementation and is not easy for real-time implementation.

[0014] Fig. 3 is a schematic drawing of the structure of the encoder using the Dolby AC-3 technique, which comprises a transient state signal detection module 301, a modified discrete cosine transformation filter MDCT 302, a frequency spectrum envelope/index encoding module 303, a mantissa encoding module 304, a forward-backward adaptive sensing model 305, a parameter bit allocation module 306, and a bit stream multiplexing module 307.

[0015] The audio signal is determined through the transient state signal detection module 301 to be either a steady state signal or a transient state signal. Meanwhile, the time-domain data is mapped to the frequency-domain data through the signal adaptive MDCT filter bank 302, wherein a long window of 512 dots is applied to the steady state signal, and a pair of short windows are applied to the transient state signal.

[0016] The frequency spectrum envelope/index encoding module 303 encodes the index portion of the signal according to the requirements of the code rate and frequency resolution in three modes, i.e. D15 encoding mode, D25 encoding mode and D45 encoding mode. The AC-3 technique uses differential encoding for the spectrum envelope in frequency, because an increment of ± 2 is needed at most, each increment representing a level change of 6dB. An absolute value encoding is used for the first DC item, and differential encoding is used for the rest of the indexes. In D15 frequency spectrum envelope index encoding, each index requires about 2.33 bits, and three differential groups are encoded in a word length of 7 bits. The D15 encoding mode sacrifices the time resolution to provide refined frequency resolution. Since only relative steady signals require refined frequency resolution, and the frequency spectrums of such signals are kept relatively constant on many blocks, with respect to the steady state signals, D15 is transmitted occasionally, usually the frequency spectrum envelope of every 6 sound blocks (one data frame) is transmitted at one time. When the signal frequency spectrum is not steady, the frequency spectrum estimate needs to be frequently updated. The estimate is encoded with lower frequency resolution generally using D25 and D45 encoding modes. The D25 encoding mode provides the appropriate frequency resolution and time resolution, and differential encoding is performed in every other

frequency coefficient, thus each index needs about 1.15 bits. If the frequency spectrum is steady on two to three blocks but changes abruptly, the D25 encoding mode can be used. The D45 encoding mode performs differential encoding in every three frequency coefficients, thus each index needs about 0.58 bit. The D45 encoding mode provides very high time resolution but low frequency resolution, so it is generally used for encoding of transient state signals.

[0017] The forward-backward adaptive sensing model 305 is used for estimating the masking threshold of each frame of signals, wherein the forward adaptive portion is only applied to the encoder to estimate a group of optimal sensing model parameters through iterative loop under the restriction of the code rate, then said parameters are transferred to the backward adaptive portion to estimate the masking threshold of each frame. The backward adaptive portion is applied both to the encoder and the decoder.

[0018] The parameter bit allocation module 306 analyzes the frequency spectrum envelope of the audio signals according to the masking rule to determine the number of bits allocated to each mantissa. Said module 306 performs an overall bit allocation for all the sound channels by using a bit reservoir. When encoding in the mantissa encoding module 304, bits are taken recurrently from the bit reservior to be allocated to all the sound channels. The quantization of the mantissa is adjusted according to the number of bits that can be obtained. In order to realize compression encoding, the AC-3 encoder also uses the high frequency coupling technique, in which the high frequency portion of the coupled signal is divided into 18 sub-frequency channels according to the critical bandwidth of the human ear, then some of the sound channels are selected to be coupled starting from a certain sub-band. Finally, AC-3 audio stream is formed through the bit stream multiplexing module 307 to be output.

[0019] Fig. 4 is a schematic drawing of the flow of decoding using Dolby AC-3. First, the bit stream that is encoded byAC-3 encoder is input, and data frame synchronization and error code detection are performed on the bit stream. If a data error code is detected, error code covering or muting processing is performed, then the bit stream is de-packaged to obtain the primary information and the side information, and then index decoding is performed thereon. When performing index decoding, two pieces of side information are needed, one is the number of packaged indexes, the other is the index strategy that is adopted, such as D15, D25 or D45 mode. The decoded indexes and the bit allocation side information again perform the bit allocation to indicate the number of bits used by each packaged mantissa, thereby obtaining a group of bit allocation pointers, each corresponding to an encoded mantissa. The bit allocation pointers point out the quantizer for the mantissa and the number of bits occupied by each mantissa in the code stream. The single encoded mantissa value is de-quantized to be transformed into a de-quantized value, and the mantissa that occupies zero bit is recovered to zero or is replaced by a random jitter value under the control of the jitter mark. Then the decoupling operation is carried out, which recovers the high frequency portion of the coupled sound channel, including the index and the mantissa, from the common coupling sound channel and the coupling factor. When using the 2/0 mode to encode at the encoding terminal, a matrix processing is used for a certain sub-band, then at the decoding terminal, the sum and difference sound channel value of said sub-band should be converted into the left-right sound channel value through matrix recovery. The code stream includes the dynamic range control value of each audio block. A dynamic range compression is performed on said value to change the amplitude of the coefficient, including index and mantissa. The frequency-domain coefficients are inversely transformed into time-domain samples, then the time-domain samples are processed by adding window, and adjacent blocks are superposed to reconstruct the PCM audio signal. When the number of sound channels decoded and output is less than the number of sound channels in the encoded bit stream, a down-mixing processing should be performed on the audio signal to finally output the PCM stream.

[0020] The Dolby AC-3 encoding technique is mainly for high bit rate signals of multi-channel surround sound, but when the encoding bit rate of 5.1 sound channel is lower than 384kbps, the encoding effect is bad; besides, the encoding efficiency of stereo of mono and dual sound channels is also low.

[0021] In summary; the existing encoding and decoding techniques cannot ensure the encoding and decoding quality of audio signals of very low code rate, low code rate and high code rate and of signals of mono and dual channels, and the implementation thereof is complex.

Summary of Invention

[0022] The technical problem to be solved by this invention is to provide an enhanced audio encoding/decodirig device and method so as to overcome the low encoding efficiency and poor encoding quality with respect to the low code rate audio signals in the prior art.

[0023] The enhanced audio encoding device of the invention comprises a signal type analyzing module, a psychoacoustical analyzing module, a time-frequency mapping module, a frequency-domain linear prediction and vector quantization module, a quantization and entropy encoding module, and a bit-stream multiplexing module. The signal type analyzing module is configured to analyze the signal type of the input audio signal and output the audio signal to the psychoacoustical analyzing module and the time-frequency mapping module, and to output the result of signal type analysis to the bit-stream multiplexing module at the same time; the psychoacoustical analyzingmodule is configured to calculate a masking threshold and a signal-to-masking ratio of the input audio signal, and output them to said quan-

tization and entropy encoding module; the time-frequency mapping module is configured to convert the time-domain audio signal into frequency-domain coefficients; the frequency-domain linear prediction and vector quantization module is configured to perform linear prediction and multi-stage vector quantization on the frequency-domain coefficients and output the residual sequence to the quantization and entropy encoding module and output the side information to the bit-stream multiplexing module; the quantization and entropy encoding module is configured to perform quantization and entropy encoding on the residual sequence under the control of the signal-to-masking ratio output from the psycho-acoustical analyzing module and to output it to the bit-stream multiplexing module; and the bit-stream multiplexing module is configured to multiplex the received data to form audio encoded code stream.

[0024] The enhanced audio decoding device of the invention comprises a bit-stream demultiplexing module, an entropy decoding module, an inverse quantizer bank, an inverse frequency-domain linear prediction and vector quantization module, and a frequency-time mapping module. The bit-stream demultiplexing module is configured to demultiplex the compressed audio data stream and output the corresponding data signals and control signals to the entropy decoding module and the inverse frequency-domain linear prediction and vector quantization module; the entropy decoding module is configured to decode said signals, recover the quantized values of the spectrum so as to output to the inverse quantizer bank; the inverse quantizer bank is configured to reconstruct the inverse quantization spectrum and output it to the inverse frequency-domain linear prediction and vector quantization module; the inverse frequency-domain linear prediction and vector quantization module is configured to perform inverse quantization and inverse linear prediction filtering on the inverse quantization spectrum to obtain the spectrum-before-prediction, and to output it to the frequency-time mapping module; and the frequency-time mapping module is configured to perform frequency-time mapping on the spectrum coefficients to obtain the time-domain audio signals of low frequency band.

[0025] The invention is applicable to the Hi-Fi compression encoding of audio signals with the configuration of multiple sampling rates and sound channels, and it supports audio signals with the sampling rate of 8kHz to 192kHz, meanwhile, it supports all possible sound channel configurations and supports audio encoding/decoding of wide range of target code rate.

Description of Drawings

[0026]

Fig. 1 is a block diagram of the MPEG-2 AAC encoder;
Fig. 2 is a block diagram of the MPEG-2 AAC decoder;
Fig. 3 is a schematic drawing of the structure of the encoder using the Dolby AC-3 technique;
Fig. 4 is a schematic drawing of the decoding flow using the Dolby AC-3 technique;
Fig. 5 is a schematic drawing of the structure of the audio encoding device according to the present invention;
Fig. 6 is a schematic drawing of the structure of the audio decoding device according to the present invention;
Fig. 7 is a schematic drawing of the structure of embodiment one of the encoding device according to the present invention;
Fig. 8 is a schematic drawing of the filtering structure using wavelet transformation of Harr wavelet basis;
Fig. 9 is a schematic drawing of the time-frequency division obtained by using wavelet transformation of Harr wavelet basis;
Fig. 10 is a schematic drawing of the structure of embodiment one of the decoding device according to the present invention;
Fig. 11 is a schematic drawing of the structure of embodiment two of the encoding device according to the present invention;
Fig. 12 is a schematic drawing of the structure of embodiment two of the decoding device according to the present invention;
Fig. 13 is a schematic drawing of the structure of embodiment three of the encoding device according to the present invention;
Fig. 14 is a schematic drawing of the structure of embodiment three of the decoding device according to the present invention;
Fig. 15 is a schematic drawing of the structure of embodiment four of the encoding device according to the present invention;
Fig. 16 is a schematic drawing of the structure of embodiment four of the decoding device according to the present invention;
Fig. 17 is a schematic drawing of the structure of embodiment five of the encoding device according to the present invention;
Fig. 18 is a schematic drawing of the structure of embodiment five of the decoding device according to the present invention;

Fig. 19 is a schematic drawing of the structure of embodiment six of the encoding device according to the present invention;

Fig. 20 is a schematic drawing of the structure of embodiment six of the decoding device according to the present invention;

Fig. 21 is a schematic drawing of the structure of embodiment seven of the encoding device according to the present invention;

Fig. 22 is a schematic drawing of the structure of embodiment seven of the decoding device according to the present invention.

Preferred Embodiments

[0027] Figs. 1-4 are the schematic drawings of the structures of the encoders of the prior art, which have been introduced in the background art, so they will not be elaborated herein.

[0028] It has to be noted that to facilitate a convenient and clear description of the present invention, the following preferred embodiments of the encoding device and decoding device are described in a corresponding manner, but it is not necessary that the encoding device and the decoding device must be of one-to-one correspondence.

[0029] As shown in Fig. 5, the audio encoding device of the present invention comprises a signal type analyzing module 50, a psychoacoustical analyzing module 51, a time-frequency mapping module 52, a frequency-domain linear prediction and vector quantization module 53, a quantization and entropy encoding module 54, and a bit-stream multiplexing module 55. The signal type analyzing module 50 is configured to analyze the signal type of the input audio signal; the psychoacoustical analyzing module 51 is configured to calculate a masking threshold and a signal-to-masking ratio of the audio signal; the time-frequency mapping module 52 is configured to convert the time-domain audio signal into frequency-domain coefficients; the frequency-domain linear prediction and vector quantization module 53 is configured to perform linear prediction and multi-stage vector quantization on the frequency-domain coefficients and to output the residual sequence to the quantization and entropy encoding module 54, and to output the side information to the bit-stream multiplexing module 55 at the same time; the quantization and entropy encoding module 54 is configured to perform quantization and entropy encoding of the residual coefficients under the control of the signal-to-masking ratio output from the psychoacoustical analyzing module 51 and to output them to the bit-stream multiplexing module 55; and the bit-stream multiplexing module 55 is configured to multiplex the received data to form audio encoding code stream.

[0030] After the digital audio signal is input to the signal pre-processing module 50, the signal type is analyzed and then the signal is input to the psychoacoustical analyzing module 51 and the time-frequency mapping module 52. On the one hand, the masking threshold and the signal-to-masking ratio of said frame of audio signal are calculated in the psychoacoustical analyzing module 51, and the signal-to-masking ratio is transmitted as a control signal to the quantization and entropy encoding module 54, and on the other hand, the time-domain audio signal is converted into frequency-domain coefficients through the time-frequency mapping module 52. Said frequency-domain coefficients are transmitted to the frequency-domain linear prediction and vector quantization module 53. If the gain threshold of the frequency-domain coefficients meets the given condition, linear prediction filtering is performed on the frequency-domain coefficients, and the resulted prediction coefficients are transformed into line spectrum frequency (LSF) coefficients. Then the optimal distortion measurement criterion is used to search and calculate the the code word indexes for the respective levels of code books, and the code word index is used as side information to be transferred to the bit-stream multiplexing module 55, while the residual sequence obtained through prediction analysis is output to the quantization and entropy encoding module 54. Under the control of the signal-to-masking ratio output from the psychoacoustical analyzing module 51, said residual sequence/frequency-domain coefficients are quantized and entropy encoded in the quantization and entropy encoding module 54. The encoded data and the side information are input to the bit-stream multiplexing module 55 to be multiplexed to form a code stream of enhanced audio encoding.

[0031] The modules composing said audio encoding device will be described below in detail.

[0032] In the present invention, the signal type analyzing module 50 is configured to analyze the signal type of the input audio signal. The signal type analyzing module 50 determines if the signal is a slowly varying signal or a fast varying signal by analyzing the forward and backward masking effects based on the adaptive threshold and waveform prediction. If the signal is of a fast varying type, the relevant parameter information of the abrupt component is calculated, such as the location where the abrupt signal occurs and the intensity of the abrupt signal, etc.

[0033] The psychoacoustical analyzing module 51 is mainly configured to calculate the masking threshold, the sensing entropy and the signal-to-masking ratio of the input audio signal. The number of bits needed for the transparent encoding of the current signal frame can be dynamically analyzed based on the sensing entropy calculated by the psychoacoustical analyzing module 51, thereby adjusting the bit allocation among frames. The psychoacoustical analyzing module 51 outputs the signal-to-masking ratio of each sub-band to the quantization and entropy encoding module 54 to control it.

[0034] The time-frequency mapping module 52 is configured to convert the audio signal from a time-domain signal into a frequency-domain coefficient, and it is formed of a filter bank which can be specifically discrete Fourier transfor-

mation (DFT) filter bank, discrete cosine transformation (DCT) filter bank, modified discrete cosine transformation (MDCT) filter bank, cosine modulation filter bank, and wavelet transformation filter bank, etc.

[0035] The frequency-domain coefficients obtained from the time-frequency mapping are transmitted to the frequency-domain linear prediction and vector quantization module 53 to undergo linear prediction and vector quantization. The frequency-domain linear prediction and vector quantization module 53 consists of a linear prediction analyzer, a linear prediction filter, a transformer, and a vector quantizer. Frequency-domain coefficients are input to the linear prediction analyzer for prediction analysis to obtain the prediction gain and prediction coefficients. If the prediction gain meets a certain condition, the frequeny-domain coefficients are input to the linear prediction filter to be filtered, thereby obtaining the predicted residual sequence of the rfequency-domain coefficients. The residual sequence is directly output to the quantization and entropy encoding module 54, while the prediction coefficients are transformed into line spectrum pair frequency LSF coefficients through the transformer, and then are sent to the vector quantizer for a multi-stage vector quantization, and the quantized relevant side information is transmitted to the bit-stream multiplexing module 55.

[0036] Performing a frequency-domain linear prediction processing on the audio signals can effectively suppress the pre-echo and obtain greater encoding gain. Given that the real signal is x(t), and the square Hilbert envelope e(t) thereof is $e(t) = e(t) = F^{-1}\{\int C(\xi)\cdot C^*(\xi - f)d\xi\}$, wherein C(f) is the one-side spectrum corresponding to the positive frequency component of signal x(t), that is, the Hilbert envelope of the signal is relevant to the autocorrelation function of said signal spectrum. The relationship between the power spectrum density function of the signal and the autocorrelation function of the time-domain waveform thereof is $PSD(f) = F\{\int x(\tau)\cdot x^*(\tau - t)d\tau\}$, so the square Hilbert envelope of the signal at the time-domain and the power spectrum density function of the signal at the frequency-domain are corresponding to each other. It can be seen that with respect to some of the band-pass signals in each predetermined range, if the Hilbert envelope thereof is constant, the autocorrelation of the adjacent spectrum values is also constant, which implies that the sequence of the spectrum coefficients is a steady state sequence with respect to the frequency, thus the prediction encoding technique can be used to process the spectrum values and a group of common prediction coefficients can be used to effectively represent said signal.

[0037] The quantization and entropy encoding module 54 further comprises a non-linear quantizer bank and an encoder, wherein the quantizer can be either a scalar quantizer or a vector quantizer. The vector quantizer can be further divided into the two categories of memoryless vector quantizer and memory vector quantizer. As for the memoryless vector quantizer, each input vector is separately quantized independent of the previous vectors; while the memory vector quantizer quantizes a vector taking into account the previous vectors, i.e. using the correlation among the vectors. Main memoryless vector quantizers include full searching vector quantizer, tree searching vector quantizer, multi-stage vector quantizer, gain/waveform vector quantizer and separating mean vector quantizer; and the main memory vector quantizers include prediction vector quantizer and finite state vector quantizer.

[0038] If the scalar quantizer is used, the non-linear quantizer bank further comprises M sub-band quantizers. In each sub-band quantizer, the scale factor is mainly used to perform the quantization, specifically, all the frequency-domain coefficients of the M scale factor sub-band are non-linearly compressed, then the frequency-domain coefficients of said sub-band are quantized by using the scale factors to obtain the quantization spectrum represented by an integer to be output to the encoder, The first scale factor in each frame of signal is used as the common scale factor to be output to the bit-stream multiplexing module 55, and the rest of the scale factors are output to the encoder after differential processing with respect to their respective preceding scale factors.

[0039] The scale factors in said step are constantly varying values, which are adjusted according to the bit allocation strategy. The present invention provides an overall sensing bit allocation strategy with the minimum distortion, and details are as follows:

First, each sub-band quantizer is initialized to select an appropriate scale factor, so that the quantization values of the spectrum coefficients of all the sub-bands are zero. The quantization noise of each sub-band at this time equals to the energy value thereof, and the noise masking ratio NMR of each sub-band equals to its signal-to-masking ratio SMR, the number of bit consumed by the quantization is zero, and the number of remaining bits B_1 equals to the number of target bits B.

Second, the sub-band with the largest noise-to-masking ratio NMR is searched. If the largest noise-to-masking ratio NMR is not more than 1, the scale factor remains unchanged and the allocation result is output, thus ending the bit allocation; otherwise, the scale factor of the corresponding sub-band quantizer is reduced by one unit, then the number of bits $\Delta B_i(Q_i)$ that needs to be added for said sub-band is calculated. If the number of remaining bits of said sub-band $B_i \geq \Delta B_i(Q_i)$, the modification of said scale factor is confirmed and the number of remaining bits B_i is subtracted by $\Delta B_i(Q_i)$ to recalculate the noise-to-masking ratio NMR of said sub-band, then continue searching for the sub-band with the largest noise-to-masking ratio NMR and repeat the subsequent steps. If the number of remaining bits $B_i < \Delta B_i(Q_i)$, said modification is canceled and the previous scale factor and number of remaining bits are retained. Finally, the allocation result is output and the bit allocation is ended.

[0040] If the vector quantizer is used, the frequency-domain coefficients form a plurality of M-dimensional vectors to be input to the non-linear quantizer bank. Each M-dimensional vector is spectrum smoothed according to the smoothing factor, i.e. reducing the dynamic range of the spectrum. Then the vector quantizer finds the code word from the code book that has the shortest distance from the vector to be quantized according to the subjective perception distance measure criterion, and transfers the corresponding code word index to the encoder. The smoothing factor is adjusted based on the bit allocation strategy of vector quantization, while the bit allocation strategy of vector quantization is controlled according to the priority of sensing among different sub-bands.

[0041] After said quantization processing, the entropy encoding technique is used to further remove the statistical redundancy of the quantized coefficients and the side information. Entropy encoding is a source encoding technique, whose basic idea is allocating shorter code words to symbols that have greater probability of appearance, and allocating longer code words to symbols that have less probability of appearance, thus the average code word length is the shortest. According to Shannon noiseless encoding theorem, if the transmitted N symbols of the source messages are independent from each other, appropriate variable length encoding is used, and the average length $\overline{n}$ of the code word satisfies $\left[\frac{H(x)}{\log_2(D)}\right] \leq \overline{n} < \left[\frac{H(x)}{\log_2(D)} + \frac{1}{N}\right]$, wherein H(x) represents the entropy of the signal source, x represents the symbol variable. Since entropy H(x) is the shortest limit of the average code word length, and said formula shows that the average length of the code word is very close to its lower limit entropy *H(x),* said variable length encoding technique is also called "entropy encoding". Entropy encoding mainly includes Huffman encoding, arithmetic encoding or run length encoding, etc. The entropy encoding in the present invention can be any of said encoding methods.

[0042] Entropy encoding is performed on the quantization spectrum quantized and output by the scalar quantizer and the differentially processed scale factors in the encoder to obtain the code book sequence numbers, the encoded values of the scale factors, and the lossless encoding quantization spectrum, then the code book sequence numbers are entropy encoded to obtain the encoded values of the code book sequence numbers, then the encoded values of the scale factors, the encoded values of the code book sequence numbers, and the lossless encoding quantization spectrum are output to the bit-stream multiplexing module 55.

[0043] The code word indexes quantized by the vector quantizer are one-dimensional or multi-dimensional entropy encoded in the encoder to obtain the encoded values of the code word indexes, then the encoded values of the code word indexes are output to the bit-stream multiplexing module 55.

[0044] The bit-stream multiplexing module 55 receives the side information output from the frequency-domain linear prediction and vector quantization module 53 and the code stream including the common scale factor, encoded values of the scale factors, encoded values of the code book sequence numbers and the quantization spectrum of lossless encoding or the encoded values of the code word indexes output from the quantization and entropy encoding module 54, and then multiplexes them to obtain the compressed audio data stream.

[0045] The encoding method based on said encoder as described above includes analyzing the signal type of the input audio signal; calculating the signal-to-masking ratio of the signal whose signal type has been analyzed; performing a time-frequency mapping on the signal whose signal type has been analyzed to obtain the frequency-domain coefficients of the audio signal; performing a standard linear prediction analysis on the frequency-domain coefficients to obtain the prediction gain and prediction coefficients; determining if the prediction gain exceeds the predetermined threshold, if it does, a frequency-domain linear prediction error filtering is performed on the frequency-domain coefficients based on the prediction coefficients to obtain the linear prediction residual sequence of the frequency-domain coefficients; transforming the prediction coefficients into line spectrum pair frequency coefficients, and performing a multi-stage vector quantization on said line spectrum pair frequency coefficients to obtain the side information; quantizing and entropy encoding the residual sequence; if the prediction gain does not exceed the predetermined threshold, quantizing and entropy encoding the frequency-domain coeffcients; and multiplexing the side information and the encoded audio signal to obtain the compressed audio code stream.

[0046] The signal type analyzing step determines if the signal is of a fast varying type or of a slowly varying type by performing forward and backward masking effect analysis based on the adaptive threshold and waveform prediction, and the specific steps thereof are: decomposing the input audio data into frames; decomposing the input frames into a plurality of sub-frames and searching for the local extremal vertexes of the absolute values of the PCM data on each sub-frame; selecting the sub-frame peak value from the local extremal vertexes of the respective sub-frames; for a certain sub-frame peak value, predicting the typical sample value of a plurality of (typically four) sub-frames that are forward delayed with respect to said sub-frame by means of apluralityof (typically three) sub-frame peak values before said sub-frame; calculating the difference and ratio between said sub-frame peak value and the predicted typical sample value; if the predicted difference and ratio are both larger than the predetermined threshold, determining that said sub-frame has jump signal and confirming that said sub-frame has the local extremal vertex with the capability of backward masking pre-echo, if there is a sub-frame between the front end of said sub-frame and the position that is 2. 5ms in front of the masking vertex, whose peak value is small enough, determining that said frame of signal is a fast varying type

signal; if the predicted difference and ratio are not larger than the predetermined threshold, repeating the above steps until it is determined that said frame of signal is fast varying type signal or until reaching the last sub-frame; and if it is still not determined whether said frame of signal is a fast varying type signal when the last sub-frame has been reached, said frame of signal is a slowly varying type signal.

[0047] There are many methods for performing a time-frequency transformation of the time-domain audio signals, such as discrete Fourier transformation (DFT), discrete cosine transformation (DCT), modified discrete cosine transformation (MDCT), cosine modulation filter bank, wavelet transformation, etc. The modified discrete cosine transformation MDCT and cosine modulation filtering are taken as examples to illustrate the process of time-frequency mapping.

[0048] With respect to using modified discrete cosine transformation MDCT to perform the time-frequency transformation, the time-domain signals of M samples from the previous frame and the time domain signals of M samples of the present frame are selected first, then a window adding operation is performed on the time-domain signal of altogether 2M samples of these two frames, and then, MDCT transformation is performed on the window added signal to obtain M frequency-domain coefficients.

[0049] The pulse response of MDCT analysis filter is:

$$h_k(n) = w(n)\sqrt{\frac{2}{M}}\cos\left[\frac{(2n+M+1)(2k+1)\pi}{4M}\right]$$

then MDCT transformation is:

$$X(k) = \sum_{n=0}^{2M-1} x(n)h_k(n) \quad 0 \le k \le M-1,$$

wherein *w(n)* is a window function, *x(n)* is the input time-domain signal of MDCT transformation, and *X(k)* is the output frequency-domain signal of MDCT transformation.

[0050] To meet the requirement for complete signal reconstruction, the window function *w(n)* of MDCT transformation must satisfy the following two conditions:

$$w(2M-1-n) = w(n) \text{ and } w^2(n)+w^2(n+M) = 1.$$

[0051] In practice, Sine window can be used as the window function. Of course, double orthogonal transformation can also be used, and said limitation to the window function is modified by specific analysis filter and synthesis filter.

[0052] With respect to using cosine modulation filtering to perform the time-frequency transformation, the time-domain signals of M samples from the previous frame and the time domain signals of M samples of the present frame are selected first, then a window adding operation is performed on the time-domain signal of altogether 2M samples of these two frames, and then, cosine modulation filtering is performed on the window added signal to obtain M frequency-domain coefficients.

[0053] The impulse responses of conventional cosine modulation filtering technique are:

$$h_k(n) = 2p_a(n)\cos\left(\frac{\pi}{M}(k+0.5)(n-\frac{D}{2})+\theta_k\right)$$
$$n = 0,1,\cdots,N_h-1$$

$$f_k(n) = 2p_a(n)\cos\left(\frac{\pi}{M}(k+0.5)(n-\frac{D}{2})-\theta_k\right), \quad n=0,1,\cdots,N_f-1$$

wherein, $0 \le k < M-1$, $0 \le n < 2KM-1$, K is an integer greater than 0, and

$$\theta_k = (-1)^k\frac{\pi}{4}$$

[0054] Suppose that the impulse response length of the analysis window (analysis prototype filter) $P_a(n)$ of M sub-bands cosine modulation filter bank is N_a, and the impulse response length of integration window (integration prototype filter) $P_s(n)$ is N_s. When the analysis window equals to the integration window, i.e. $P_a(n) = P_s(n)$ and $Na = N_s$, the cosine modulation filter bank represented by the above two formulae is an orthogonal filter bank, and matrixes *H* and *F* ($[H]_{n,k} = h_k(n)$, $[F]_{n,k} = f_k(n)$) are orthogonal transformation matrixes. In order to obtain linear phase filter bank, it is further specified that the symmetrical windows satisfy $P_a(2KM-1-n) = P_a(n)$. In order to ensure the complete reconstruction of the orthogonal and double orthogonal systems, the window function further needs to satisfy a certain condition, details can be found in the document "Multirate Systems and Filter Banks", P. P. Vaidynathan, Prentice Hall, Englewood Cliffs, NJ, 1993.

[0055] The calculation of the masking threshold and signal-to-mas king ratio of the pre-processed audio signal includes the following steps:

Step 1 : mapping the signal from time-domain to frequency-domain. Fast Fourier transformation and Hanning window techniques can be used to transform the time-domain data into frequency-domain coefficient X[*k*], X[*k*] is represented by amplitude r[*k*] and phaseφ [*k*] as $X[k] = r[k]e^{j\varphi[k]}$, then the energy e[b] of each sub-band is the sum of all the spectrum lines within said sub-band, i.e. $e[b] = \sum_{k=k_l}^{k=k_h} r^2[k]$, wherein k_l and k_h are respectively the upper and lower boundaries of the sub-band b.

Step 2: determining the tone and non-tone components in the signal. The tonality of signal is estimated by performing inter-frame prediction on each spectrum line. The Euclidean distances of the prediction value and real value of each spectrum line are mapped into unpredictable measure, and spectrum component of high predictability is considered as having strong tonality, while spectrum component of low predictability is considered as quasi-noise.

The amplitude r_{pred} and phaseφ_{pred} can be represented by the following equations:

$$r_{pred}[k] = r_{t-1}[k] + (r_{t-1}[k] - r_{t-2}[k])$$

$$\phi_{pred}[k] = \phi_{t-1}[k] + (\phi_{t-1}[k] - \phi_{t-2}[k]),$$

Wherein, t represents the coefficient of the present frame, t-1 represents the coefficient, of the previous frame, and t-2 represents the coefficient of the previous two frames. The unpredictable measure c[k] is calculated by the equation of

$$c[k]=\frac{dist(X[k],X_{pred}[k])}{r[k]+|r_{pred}[k]|}$$

wherein the Euclidean distance $dist\ (X[k], X_{pred[k]})$ is calculated by the equation of

$$dist(X[k],X_{pred}[k])=|X[k]-X_{pred}[k]|$$
$$=\left((r[k]\cos(\phi[k])-r_{pred}[k]\cos(\phi_{pred}[k]))^2+(r[k]\sin(\phi[k])-r_{pred}[k]\sin(\phi_{pred}[k]))^2\right)^{1/2}$$

Thus the unpredictability c[b] of each sub-band is the weighted sum of the energy of all the spectrum lines within said sub-band

[0056] with respect to its unpredictability, i.e. $c[b]=\sum_{k=k_l}^{k=k_h}c[k]r^2[k]$. A convolution operation is respectively performed for the sub-band energy e[b] and the unpredictability c[b] with respect to the spread function to obtain the sub-band energy spread $e_s[b]$ arid the sub-band unpredictability spread $c_s[b]$, the spread function of masking i with respect to sub-band b being represented by s[i, b]. In order to eliminate the influence on the energy transformation caused by the spread function, the sub-banduripredictability spread $c_s[b]$ has to be normalized, and the result of normalization $\tilde{c}_s[b]$ is represented by $\tilde{c}_s[b]=\frac{c_s[b]}{e_s[b]}$. Similarly, in order to eliminate the influence on the sub-band energy caused by the spread function, the normalized energy spread $\tilde{e}_s[b]$ is defined to be $\tilde{e}_s[b]=\frac{e_s[b]}{n[b]}$, wherein the normalization factor n[b] is $n[b]=\sum_{i=1}^{b_{max}}s[i,b]$, b_{max} is the number of sub-bands allocated to said frame of signal.

[0057] The tonality t[b] of the sub-band can be calculated according to the normalized unpredictability spread $\tilde{c}_s[b]$, i.e. $t[b]= -0.299-0.43\log_e(\tilde{c}_s[b])$, and $0\le t[b]\le 1$. When t [b] =1, said sub-band signal is pure tone, and when t [b] =0, said sub-band signal is white noise.

[0058] Step 3: calculating the signal-to-noise ratio (SNR) needed for each sub-band. The value of the noise-masking-tone (NMT) of all the sub-bands is set to be 5dB, and the value of the tone-masking-noise (TMN) is set to be 18dB; if the noise is to be made imperceptible, the signal-to-noise ratio SNR[b] of each sub-band should be that SNR[b] =18t[b] +6(1-t[b]).

[0059] Step 4: calculating the masking threshold of each sub-band and the sensing entropy of the signal. The noise energy threshold n[b] of each sub-band is calculated to be $n[b] = \tilde{e}_s[b]\ 10^{-SNR[b]/10}$ based on the normalized signal energy of each sub-band and the needed signal-to-noise ratio SNR obtained in the above steps.

[0060] In order to avoid the influence of the pre-echo, the noise energy threshold n[b] of the present frame is compared to the noise energy threshold $n_{prev}[b]$ of the previous frame, and the masking threshold of the signal is obtained to be $n[b] =\min (n[b], 2n_{prev}[b])$, thereby ensuring that there will not be any deviation in the masking threshold owing to the generation of high-energy impact at the near end of the analysis window.

[0061] Further, while taking into account the influence of the still masking threshold qsthr [b], the final masking threshold of the signal is selected to be the larger one of the still masking threshold and said calculated masking threshold, i.e. n [b] = max (n[b], qsthr[b]). Then the sensing entropy is calculated by the equation of $pe=-\sum_{b=0}^{b_{max}}(cbwidth_b\times\log_{10}(n[b]/(e[b]+1)))$, wherein $cbwidth_b$ represents the number of spectrum lines included in each sub-band.

[0062] Step 5: calculating the signal-to-masking ratio (SMR) of each sub-band signal. The signal-to-masking ratio SMR[b] of each sub-band is $SMR[b] = 10\log_{10}\left(\frac{e[b]}{n[b]}\right)$.

[0063] After obtaining the frequency-domain coefficients, a liner prediction and vector quantization is performed on the frequency-domain coefficients. First, a standard linear prediction analysis is performed on the frequency-domain coefficients, including calculating the autocorrelation matrix, obtaining the prediction gain and the prediction coefficients by recursively executing the Levinson-Durbin algorithm. Then it is determined whether the calculated prediction gain exceeds a predetermined threshold, and if it does, a frequency-domain linear prediction error filtering is performed on the frequency-domain coefficients based on the prediction coefficients, otherwise, the frequency-domain coefficients are not processed and the next step is executed to quantize and entropy encoding the frequency-domain coefficients.

[0064] Linear prediction includes forward prediction and backward prediction. Forward prediction refer to predicting the current value by using the values before a certain moment, while the backward prediction refers to predicting the current value by using the values after a certain moment. The forward prediction will be used as an example to explain the linear prediction error filtering. The linear predicton filtering function is $A(z) = 1 - \sum_{i=1}^{p} a_i z^{-i}$, wherein a_i denotes the prediction coefficient, p is the prediction order. The frequency-domain coefficient $X(k)$ that has undergone the time-frequency transformation is filtered to obtain the prediction error $E(k)$, which is also called the residual sequence, and there is the relationship of $E(k) = X(k) \cdot A(z) = X(k) - \sum_{i=1}^{p} a_i X(k-i)$.

[0065] Thus, after frequency-domain linear prediction filtering, the frequency-domain coefficients $X(k)$ output after the time-frequency transformation can be represented by the residual sequence $E(k)$ and a group of prediction coefficients a_i. Then said group of prediction coefficients a_i are transformed into the linear spectrum frequency LSF coefficients, and multi-stage vector quantization is performed thereon. The vector quantization uses the optimal distortion measurement criterion (e.g. nearest neighboring criterion) to search and calculate the code word indexes of the respective stages of code book, thereby determining the code words corresponding to the prediction coefficients and outputting the code word indexes as the side information. Meanwhile, the residual sequence $E(k)$ is quantized and entropy encoded. It can be seen from the encoding principle of linear prediction analysis that the dynamic range of the residual sequence of the spectrum coefficients is smaller than that of the original spectrum coefficients, so less number of bits are allocated thereto during quantization, or under the condition of same number of bits, improved encoding gain can be obtained.

[0066] After obtaining the signal-to-masking ratio of the sub-band signal, the frequency-domain coefficients or the residual sequence is quantized and entropy encoded based on said signal-to-masking ratio, wherein the quantization can be scalar quantization or vector quantization.

[0067] The scalar quantization comprises the steps of non-linearly compressing the frequency-domain coefficients in all the scale factor bands; using the scale factor of each sub-band to quantize the frequency-domain coefficients of said sub-band to obtain the quantization spectrum represented by an integer; selecting the first scale factor in each frame of signal as the common scale factor; and differentiating the rest of the scale factors from their respective previous scale factors.

[0068] The vector quantization comprises the steps of forming a plurality of multi-dimensional vector signals with the frequency-domain coefficients; performing spectrum smoothing for each M-dimensional vector according to the smoothing factor; searching for the code word from the code book that has the shortest distance from the vector to be quantized according to the subjective perception distance measure criterion to obtain the code word indexes.

[0069] The entropy encoding step comprises entropy encoding the quantization spectrum and the differentiated scale factors to obtain the sequence numbers of the code book, the encoded values of the scale factors and the quantization spectrum of lossless encoding; and entropy encoding the sequence numbers of the code book to obtain the encoded values thereof.

[0070] Or, a one-dimensional or multi-dimensional entropy encoding is performed on the code word indexes to obtain the encoded values of the code word indexes.

[0071] The entropy encoding method described above can be any one of the existing Huffman encoding, arithmetic encoding or run length encoding method.

[0072] After quantization and entropy encoding, the encoded audio signal is obtained, which is multiplexed together with the common scale factor, side information and the result of signal type analysis to obtain the compressed audio code stream.

[0073] Fig. 6 is a schematic drawing of the structure of the audio decoding device according to the present invention.

The audio decoding device comprises a bit-stream demultiplexing module 801, an entropy decoding module 802, an inverse quantizer bank 803, an inverse frequency-domain linear prediction and vector quantization module 804, and a frequency-time mapping module 805. The compressed audio code stream is demultiplexed by the bit-stream demultiplexing module 801 to obtain the corresponding data signal and control signal which are output to the entropy decoding module 802 and the inverse frequency-domain linear prediction and vector quantization module 804; the data signal and control signal are decoded in the entropy decoding module 802 to recover the quantized values of the spectrum. Said quantized values are reconstructed in the inverse quantizer bank 803 to obtain the inversely quantized spectrum. The inversely quantized spectrum is then output to the inverse frequency-domain linear prediction and vector quantization module 804 for inverse quantization and inverse linear prediction filtering to obtain the spectrum-before-prediction, which is output to the frequency-time mapping module 805, then the time-domain audio signal of low frequency band is obtained after the frequency-time mapping.

[0074] The bit-stream demultiplexing module 801 decomposes the compressed audio code stream to obtain the corresponding data signal and control signal and to provide the corresponding decoding information for other modules. The compressed audio data stream is demultiplexed, and the signals output to the entropy decoding module 802 include the common scale factor, the scale factor encoded values, the encoded values of the code book sequence numbers, and the quantized spectrum of the lossless encoding, or the encoded values of the code word indexes; the control information of inverse frequency-domain linear prediction and vector quantization is output to the inverse frequency-domain linear prediction and vector quantization module 804.

[0075] If, in the encoding device, the quantization and entropy encoding module 54 use the scalar quantizer, then in the decoding device, what the entropy decoding module 802 receives are the common scale factor, the scale factor encoded values, the encoded values of the code book sequence numbers, and the quantized spectrum of the lossless encoding as output from the bit-stream demultiplexing module 801, then code book sequence number decoding, spectrum coefficient decoding and scale factor decoding are performed thereon to reconstruct the quantization spectrum and to output the integer representation of the scale factors and the quantized values of the spectrum to the inverse quantizer bank 803. The decoding method used by the entropy decoding module 802 corresponds to the encoding method used by entropy encoding in the encoding device, which is, for example, Huffman decoding, arithmetic decoding or run length decoding, etc.

[0076] Upon receipt of the quantized values of the spectrum and the integer representation of the scale factor, the inverse quantizer bank 803 inversely quantizes the quantized values of the spectrum into reconstructed spectrum without scaling (inverse quantization spectrum), and outputs the inverse quantization spectrum to the inverse frequency-domain linear prediction and vector quantization module 804. The inverse quantizer bank 803 can be either a uniform quantizer bank or a non-uniform quantizer bank realized by a companding function.

[0077] In the encoding device, the quantizer bank uses the scalar quantizer, so in the decoding device, the inverse quantizer bank 803 also uses the scalar inverse quantizer. In the scalar inverse quantizer, the quantized values of the spectrum are non-linearly expanded first, then all the spectrum coefficients (inverse quantization spectrum) in the corresponding scale factor band are obtained by using each scale factor.

[0078] If the quantization and entropy encoding module 54 uses the vector quantizer, then in the decoding device, the entropy decoding module 802 receives the encoded values of the code word indexes output from the bit-stream demultiplexing module 801. The encoded values of the code word indexes are decoded by the entropy decoding method corresponding to the entropy encoding method used in encoding, thereby obtaining the corresponding code word indexes.

[0079] The code word indexes are output to the inverse quantizer bank 803. By looking up the code book, the quantized values (inverse quantization spectrum) are obtained and output to the frequency-time mapping module 805. The inverse quantizer bank 803 uses the inverse vector quantizer.

[0080] In the encoder, the technique of frequency-domain linear prediction vector quantization is used to suppress the pre-echo and to obtain greater encoding gain. Therefore, in the decoder, the control information of inverse frequency-domain linear prediction vector quantization output from the inverse quantization spectrum and bit-stream demultiplexing module 801 is input to the inverse frequency-domain linear prediction and vector quantization module 804 to recover the spectrum-before- linear-prediction.

[0081] The inverse frequency-domain linear prediction and vector quantization module 804 comprises an inverse vector quantizer, an inverse transformer, and an inverse linear prediction filter, wherein the inverse vector quantizer is used for inversely quantizing the code word indexes to obtain the line spectrum pair frequency LSF coefficients; the inverse transformer is used for inversely transforming the line spectrum frequency LSF coefficients into prediction coefficients, and the inverse linear prediction filter is used for inversely filtering the inverse quantization spectrum based on the prediction coefficients to obtain the spectrum-before-prediction and output it to the frequency-time mapping module 805.

[0082] Time-domain audio signals of low frequency channel can be obtained by a mapping processing of the inverse quantization spectrum or the spectrum-before-prediction by the frequency-time mapping module 805. The frequency-time mapping module 805 can be a filter bank of inverse discrete cosine transformation (IDCT), a filter bank of inverse

discrete Fourier transformation (IDFT), a filter bank of inverse modified discrete cosine transformation (IMDCT), a filter bank of inverse wavelet transformation, and a cosine modulation filter bank, etc.

[0083] The decoding method based on the above-mentioned decoder comprises: demultiplexing the compressed audio code stream to obtain the data information and control information; entropy decoding said information to obtain the quantized value of the spectrum; inversely quantizing the quantized values of the spectrum to obtain the inverse quantization spectrum; determining if the control information contains information concerning that the inverse quantization spectrum needs to undergo the inverse frequency-domain linear prediction vector quantization, if it does, performing the inverse vector quantization to obtain the prediction coefficients, and performing an inverse linear prediction filtering on the inverse quantization spectrum according to the prediction coefficients to obtain the spectrum-before-prediction; frequency-time mapping the spectrum-before-prediction to obtain the time-domain audio signals of low frequency band; if the control information does not contain information concerning that the inverse quantization spectrum needs to undergo the inverse frequency-domain linear prediction vector quantization, frequency-time mapping the inverse quantization spectrum to obtain the time-domain audio signals of low frequency band.

[0084] If the demultiplexed information include the encoded values of the code book sequence numbers, the common scale factor, the encoded values of the scale factors, and the quantization spectrum of the lossless encoding, then it means that the spectrum coefficients in the encoding device are quantized by the scalar quantization technique. Accordingly, the entropy decoding steps include: decoding the encoded values of the code book sequence numbers to obtain the code book sequence numbers of all the scale factor bands; decoding the quantization coefficients of all the scale factor bands according to the code book corresponding to the code book sequence number; and decoding the scale factors of all the scale factor bands to reconstruct the quantization spectrum. The entropy decoding method used in said process corresponds to the entropy encoding method used in the encoding method, which is, for example, run length decoding method, Huffman decoding method, or arithmetic decoding method, etc.

[0085] The entropy decoding process is described below by using as examples the decoding of the code book sequence numbers by the run length decoding method, the decoding of the quantization coefficients by the Huffman, decoding method, and the decoding of the scale factors by the Huffman decoding method.

[0086] First, the code book sequence numbers of all the scale factor bands are obtained through the run length decoding method. The decoded code book sequence numbers are integers within a certain range. Suppose that said range is [0, 11], then only the code book sequence numbers within said valid range, i.e. between 0-11, are corresponding to the Huffman code book of the spectrum coefficients. As for the all-zero sub-band, a certain code book sequence can be selected to correspond to it, typically, the 0 sequence number can be selected.

[0087] When the code book number of the respective scale factor band is obtained through decoding, the Huffman code book of spectrum coefficients corresponding to said code book number is used to decode the quantization coefficients of all the scale factor bands. If the code book number of a scale factor band is within the valid range, for example between 1-11 in this embodiment, then said code book number corresponds to a spectrumcoefficient code book, and said code book is used to decode the quantization spectrum to obtain the code word indexes of the quantization coefficients of the scale factor bands, subsequently, the code word indexes are de-packaged to obtain the quantization coefficients. If the code book number of the scale factor band is not between 1 and 11, then said code book number is not corresponding to any spectrum coefficient code book, and the quantization coefficients of said scale factor band does not need to be decoded, but they are all directly set to be zero.

[0088] The scale factor is used to reconstruct the spectrum value on the basis of the inverse quantization spectrum coefficients. If the code book number of the scale factor band is within the valid range, each code book number corresponds to a scale factor. When decoding said scale factors, the code stream occupied by the first scale factor is read first, then the rest of the scale factors are Huffman decoded to obtain the differences between each of the scale factors and their respective previous scale factors, and said differences are added to the values of the previous scale factors to obtain the respective scale factors. If the quantization coefficients of the present sub-band are all zero, then the scale factors of said sub-band do not have to be decoded.

[0089] After said entropy decoding, the quantized values of the spectrum and the integer representation of the scale factors are obtained, then the quantized values of the spectrum are inversely quantized to obtain the inverse quantization spectrum. The inverse quantization processing includes non-linearly expanding the quantized values of the spectrum, and obtaining all the spectrum coefficients (inverse quantization spectrum) in the corresponding scale factor band according to each scale factor.

[0090] If the demultiplexed information contain the encoded values of the code word indexes, it means that the encoding device uses the vector quantization technique to quantize the spectrum coefficients, then the entropy decoding steps include: decoding the encoded values of the code word indexesby means of the entropy decoding method corresponding to the entropy encoding method used in the encoding device so as to obtain the code word indexes, then inversely quantizing the code word index to obtain the inverse quantization spectrum.

[0091] An inverse frequency-domain linear prediction vector quantization is performed on the inverse quantization spectrum. First, it is determined if said frame of signal has undergone the frequency-domain linear prediction vector

quantization according to the control information, if it has, the code word indexes resulted from the vector quantization of the prediction coefficients are obtained from the control information; then the quantized line spectrum frequency LSF coefficients are obtained according to the code word indexes, on the basis of which the prediction coefficients are calculated; subsequently, a linear prediction synthesizing is performed on the inverse quantization spectrum to obtain the spectrum-before-prediction.

[0092] The transfer function of the linear prediction error filtering is $A(z) = 1 - \sum_{i=1}^{p} a_i z^{-i}$, wherein a_i denotes the prediction coefficient, p is the prediction order. The residual sequence $E(k)$ and the spectrum $X(k)$ before prediction has a relationship of

$$X(k) = E(k) \cdot \frac{1}{A(z)} = E(k) + \sum_{i=1}^{p} a_i X(k-i).$$

[0093] Thus the residual sequence $E(k)$ and the calculated prediction coefficient a_i are synthesized by frequency-domain linear prediction to obtain the spectrum X(k) before prediction which is then frequency-time mapped.

[0094] If the control information indicates that said signal frame has not undergone the frequency-domain linear prediction vector quantization, the inverse frequency-domain linear prediction vector quantization will not be performed, and the inverse quantization spectrum is directly frequency-time mapped.

[0095] The method of performing a frequency-time mapping on the inverse quantization spectrum corresponds to the time-frequency mapping method in the encoding method, which can be inverse discrete cosine transformation (IDCT), inverse discrete Fourier transformation (IDFT), inverse modified discrete cosine transformation (IMDCT), and inverse wavelet transformation, etc.

[0096] The frequency-time mapping process is illustrated below by taking inverse modified discrete cosine transformation IMDCT as an example. The frequency-time mapping process includes three steps: IMDCT transformation, time-domain window adding processing and time-domain superposing operation.

[0097] First, IMDCT transformation is perform on the spectrum-before-prediction or the inverse quantization spectrum to obtain the transformed time-domain signal $x_{i,n}$. The expression of IMDCT transformation is

$$x_{i,n} = \frac{2}{N} \sum_{k=0}^{\frac{N}{2}-1} spec[i][k] \cos\left(\frac{2\pi}{N}(n+n_0)(k+\frac{1}{2})\right)$$

wherein, n is the sequence number of the sample, and $0 \le n < N$, N represents the number of time-domain samples which is 2048, $n_0 = (N/2+1)/2$; i represents the frame sequence number; k represents the spectrum sequence number.

[0098] Second, window adding is performed on the time-domain signal obtained from IMDCT transformation at the time domain. In order to satisfy the requirement for complete reconstruction, the window function w(n) must meet the two conditions of $w(2M-1-n) = w(n)$ and $w^2(n) + w^2(n+M) = 1$.

[0099] Typical window functions include, among others, Sine window and Kaiser-Bessel window. The present invention uses a fixed window function, which is $w(N+k) = \cos(pi/2*((k+0.5)/N - 0.94*\sin(2*pi/N*(k+0.5))/(2*pi)))$, wherein pi is the circular constant, $k=1...N-1$; $w(k)$ represents the kth coefficient of the window function and $w(k) = w(2*N-1-k)$; N represents the number of samples of the encoded frame, and N=1024. In addition, said restriction to the window function can be modified by using double orthogonal transformation with a specific analysis filter and synthesizing filter.

[0100] Finally, said window added time-domain signal is superposed to obtain the time-domain audio signal. Specifically, the first N/2 samples of the signals obtained by the window adding are superposed with the last N/2 samples of the previous frame of signal to obtain N/2 output time-domain audio samples, i.e., $timeSam_{i,n} = preSam_{i,n} + preSam_{i-1,n+N/2}$, wherein i denotes the frame sequence number, n denotes the sample sequence number, $0 \le n \le \frac{N}{2}$, and N is 2048.

[0101] After processing of the compressed audio data stream through the above-described steps, time-domain audio

signals of low frequency band are obtained.

[0102] Fig. 7 is a schematic drawing of the structure of embodiment one of the encoding device of the present invention. On the basis of Fig. 5, this embodiment has a multi-resolution analyzing module 56 added between the output of the frequency-domain linear prediction and vector quantization module 53 and the input of the quantization and entropy encoding module 54.

[0103] With respect to signals of a fast varying type, in order to effectively overcome the pre-echo produced during the encoding and to improve the encoding quality, the encoding device of the present invention increases the time resolution of the encoded fast varying signals by means of a multi-resolution analyzing module 56. The residual sequence or frequency-domain coefficients output from the frequency-domain linear prediction and vector quantization module 53 are input to the multi-resolution analyzing module 56. If the signal is of a fast varying type, a frequency-domain wavelet transformation or frequency-domain modified discrete cosine transformation (MDCT) is performed to obtain the multi-resolution representation for the residual sequence/frequency-domain coefficients to be output to the quantization and entropy encoding module 54. If the signal is of a slowly varying type, the residual sequence/frequency-domain coefficients are directly output to the quantization and entropy encoding module 54 without being processed.

[0104] The multi-resolution analyzing module 56 performs a time-and-frequency-domain reorganization of the input frequency-domain data to improve the time resolution of the frequency-domain data at the cost of reducing the precision of the frequency, thereby to automatically adapt to the time-frequency characteristics of the signals of fast varying type, accordingly, the effect of suppressing the pre-echo is achieved without adjusting the form of the filter bank in the time-frequency mapping module 52 at any time. The multi-resolution analyzing module 56 comprises a frequency-domain coefficient transformation module and a reorganization module, wherein the frequency-domain coefficient transformation module is used for transforming the frequency-domain coefficients into a time-frequency plane coefficients; and the reorganization module is used for reorganizing the time-frequency plane coefficient according to a certain rule. The frequency-domain coefficient transformation module can use the filter bank of frequency-domain wavelet transformation, or the filter bank of frequency-domain MDCT transformation, etc.

[0105] The operation process of multi-resolution analysis module 56 is described below by taking frequency-domain wavelet transformation and frequency-domain MDCT transformation as examples.

1) Frequency-domain wavelet transformation

[0106] Suppose that the time series is x(i), i=0, 1, ...2M-1, and the frequency-domain coefficients obtained through time-frequency mapping is X(k), k=0, 1, ... M-1. The frequency-domain wavelet or the wavelet basis of wavelet package transformation may either be fixed or adaptive.

[0107] The multi-resolution analysis on the frequency-domain coefficients is illustrated below by taking the simplest wavelet transformation of Harr wavelet basis as an example.

[0108] The scale coefficient of Harr wavelet basis is $[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}]$, and

[0109] the wavelet coefficient is $[\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}]$. Fig. 8 shows the schematic drawing of the filtering structure that performs wavelet transformation by using Harr wavelet basis, wherein H_0 represents low-pass filtering (the filtering coefficient is $[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}]$), H_1 represents high-pass filtering (the filtering coefficient is $[\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}]$), "↓ 2" represents a duple down sampling operation. No wavelet transformation is performed for the medium and low frequency portions $X_1(k)$, k=1, ..., k_1 in the frequency-domain coefficients, and Harr wavelet transformation is performed for the high frequency portions in the MDCT coefficients to obtain coefficients $X_2(k)$, $X_3(k)$, $X_4(k)$, $X_5(k)$, $X_6(k)$, $X_7(k)$, of different time-frequency intervals, and the division of the corresponding time-frequency plane is as shown in Fig. 9. By selecting different wavelet bases, different wavelet transformation structures can be used for processing so as to obtain other similar time-frequency plane divisions. Therefore, the time-frequency plane division during signal analysis can be discretionarily adjusted as desired so as to meet different requirements of the analysis of the time and frequency resolution.

[0110] The above-mentioned time-frequency plane coefficients are reorganized in the reorganization module according to a certain rule, for example, the time-frequency plane coefficients can be organized in the frequency direction first, and the coefficients in each frequency band are organized in the time direction, then the organized coefficients are arranged in the order of sub-window and scale factor band.

2) Frequency-domain MDCT transformation

[0111] Suppose that the frequency-domain data input to the filter bank of the frequency-domain DMCT transformation is X(k), k= 0, 1, ..., N-1. M dot MDCT transformation is performed on said N dot frequency-domain data sequentially, so that the frequency precision of the time frequency domain data is reduced, while the time precision is increased. MDCT transformations of different lengths are used in different frequency-domain ranges, thereby to obtain different time-frequency plane divisions, i.e. different time and frequency precision. The reorganization module reorganizes the time-frequency data output from the filter bank of the frequency-domain MDCT transformation. One way of reorganization is to organize the time-frequency plane coefficients in the frequency direction first, and the coefficients in each frequency band are organized in the time direction at the same time, then the organized coefficients are arranged in the order to sub-window and scale factor band.

[0112] With respect to the encoding method based on the encoding device as shown in Fig. 7, the basic flow thereof is the same as that of the encoding method based on the encoding device as shown in Fig. 5, and the difference therebetween is that the former further includes the following steps: before quantizing and entropy encoding the residual sequence/frequency-domain coefficients, if the signal is a fast varying signal, performing a multi-resolution analysis on the residual sequence/frequency-domain coefficients; if the signal is not a fast varying signal, directly quantizing and entropy encoding the residual sequence/frequency-domain coefficients.

[0113] The multi-resolution analysis can use frequency-domain wavelet transformation method or frequency-domain MDCT transformation method. The frequency-domain wavelet analysis method includes: wavelet transforming the frequency-domain coefficients to obtain the time-frequency plane coefficients; reorganizing said time-frequency plane coefficients according to a certain rule. The MDCT transformation includes: MDCT transforming the frequency-domain coefficients to obtain the time-frequency plane coefficients; reorganizing said time-frequency plane coefficients according to a certain rule. The reorganization method includes: organizing the time-frequency plane coefficients in the frequency direction, and organizing the coefficients in each frequency band in the time direction, then arranging the organized coefficients in the order of sub-window and scale factor band.

[0114] Fig. 10 is a schematic drawing of embodiment one of the decoding device of the present invention. Said decoding device has a multi-resolution integration module 806 added on the basis of the decoding device as shown in Fig. 6. Said multi-resolution integration module 806 is between the output of the inverse quantizer bank 803 and the input of the inverse frequency-domain linear prediction and vector quantization module 804 for multi-resolution integrating the inverse quantization spectrum.

[0115] In the encoder, the technique of multi-resolution filtering is used for the fast varying type signals to increase the time resolution of the encoded fast varying type signals. Accordingly, in the decoder, the multi-resolution integration module 806 is used to recover the frequency-domain coefficients of the fast varying signals before multi-resolution analysis. The multi-resolution integration module 806 comprises a coefficient reorganization module and a coefficient transformation module, wherein the coefficient transformation module may use a filter bank of frequency-domain inverse wavelet transformation or a filter bank of frequency-domain IMDCT transformation.

[0116] With respect to the decoding method of the decoding device as shown in Fig. 10, the basic flow thereof is the same as that of the decoding,method of the decoding device as shown in Fig. 6, and the difference is that the former further includes the steps of after obtaining the inverse quantization spectrum, performing a multi-resolution integration thereon, and then determining if it is necessary to perform an inverse frequency-domain linear prediction vector quantization on the multi-resolution integrated inverse quantization spectrum.

[0117] The method of multi-resolution integration is described below by taking the frequency-domain IMDCT transformation as an example. The method specifically includes: reorganizing the coefficients of the inverse quantization spectrum, performing a plurality of IMDCT transformation on each coefficient to obtain the inverse quantization spectrum before the multi-resolution analysis. This process is described in detail by using 128 IMDCT transformation (8 inputs and 16 outputs). Firstly, the coefficients of the inverse quantization spectrum are arranged in the order of sub-window and scale factor band; then they are reorganized in the order of frequency, thus the 128 coefficients of each sub-window are organized together in the order of frequency. Subsequently, the coefficients that are arranged in the order of sub-window are organized in frequency direction with 8 in each group, and the 8 coefficients in each group are arranged in time sequence, thus there are altogether 128 groups of coefficients in the frequency direction. An IMDCT transformation of 16 dots is performed on each group of coefficients, and the 16 coefficients output after the IMDCT transformation of each group are added in an overlapping manner to obtain 8 frequency-domain data. 128 times of such operation are performed from the low frequency direction to the high frequency direction to obtain 1024 frequency-domain coefficients.

[0118] Fig. 11 is the schematic drawing of the second embodiment of the encoding device of the present invention. On the basis of Fig. 5, said embodiment has a sum-difference stereo (M/S) encoding module 57 added between the output of the frequency-domain linear prediction and vector quantization module 53 and the input of the quantization and entropy encoding module 54. The psychoacoustical analyzing module 51 outputs the masking threshold of the sum-difference sound channel to the quantization and entropy encoding module 54. With respect to multi-channel signals,

the psychoacoustical analyzing module 51 calculates not only the masking threshold of the single sound channel of the audio signals, but also the masking threshold of the sum-difference sound channels. The sum-difference stereo encoding module 57 can also be located between the quantizer bank and the encoder in the quantization and entropy encoding module 54.

[0119] The sum-difference stereo module 57 makes use of the correlation between the two sound channels in the sound channel pair to equate the freuqency-domain coefficients/residual sequence of the left-right sound channels to the freuqency-domain coefficients/residual sequence of the sum-difference sound channels, thereby reducing the code rate and improving the encoding efficiency. Hence, it is only suitable for multi-channel signals of the same signal type. While as for mono signals or multi-channel signals of different signal types, the sum-difference stereo encoding is not performed.

[0120] The encoding method of the encoding device as shown in Fig. 11 is substantially the same as the encoding method of the encoding device as shown in Fig. 5, and the difference is that the former further includes the steps of determining whether the audio signals are multi-channel signals before quantizing and entropy encoding the residual sequence/frequency-domain coefficients, if they are multi-channel signals, determining whether the types of the signals of the left-right sound channels are the same, if the signal types are the same, determining whether the scale factor bands corresponding to the two sound channels meet the conditions of sum-difference stereo encoding, if they meet the conditions, performing a sum-difference stereo encoding on the residual sequence/frequency-domain coefficients to obtain the residual sequence/frequency-domain coefficients of the sum-difference sound channels; if they do not meet the conditions, the sum-difference stereo encoding is not performed. If the signals are mono signals or multi-channel signals of different types, the frequency-domain coefficients are not processed.

[0121] The sum-difference stereo encoding can be applied not only before the quantization, but also after the quantization and before the entropy encoding, that is, after quantizing the residual sequence/frequency-domain coefficients, it is determined if the audio signals are multi-channel signals; if they are, it is determined if the signals of the left-right sound channels are of the same type, and if the signal types are the same, it is determined if the scale factor bands meet the encoding condition. If they meet the condition, performing a sum-difference stereo encoding on the quantization spectrum to obtain the quantization spectrum of the sum-difference sound channels. If they do not meet the conditions, the sum-difference stereo encoding is not performed. If the signals are mono signals or multi-channel signals of different types, the frequency-domain coefficients are not processed.

[0122] There are many methods for determining whether a sum-difference stereo encoding can be performed on the scale factor band, and the one used in the present invention is K-L transformation. The specific process of determination is as follows:

[0123] Suppose that the spectrum coefficient of the scale factor band of the left sound channel is $l(k)$, and the corresponding spectrum coefficient of the scale factor band of the right sound channel is r(k), the correlation matrix thereof is $C = \begin{pmatrix} C_{ll} & C_{lr} \\ C_{lr} & C_{rr} \end{pmatrix}$, wherein

$$C_{ll} = \frac{1}{N}\sum_{k=0}^{N-1} l(k) * l(k); \quad C_{lr} = \frac{1}{N}\sum_{k=0}^{N-1} l(k) * r(k);$$

$C_{rr} = \frac{1}{N}\sum_{k=0}^{N-1} r(k) * r(k)$ is the number of spectrum lines of the scale factor band. The K-L transformation is performed on the correlation matrix C to obtain $RCR^T = \Lambda = \begin{pmatrix} \lambda_{ll} & 0 \\ 0 & \lambda_{os} \end{pmatrix}$, wherein,

$$R = \begin{pmatrix} \cos a & -\sin a \\ \sin a & \cos a \end{pmatrix} \quad a \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right].$$

The rotation angle a satisfies the equation of $\tan(2a) = \frac{2C_{lr}}{C_{ll} - C_{rr}}$. When a = $\pm\pi/4$, it is the sum-difference stereo encoding mode. Therefore, when the absolute value of the rotation angle a deviates from $\pi/4$ by a small amount, e.g. $3\pi/16 < |a| < 5n/16$, a sum-difference stereo encoding can be performed on the corresponding scale factor band.

[0124] If the sum-difference stereo encoding is applied before the quantization, the residual sequence/frequency-domain coefficients of the left-right sound channels at the scale factor band are linearly transformed and are replaced with the residual sequence/frequency-domain coefficients of the sum-difference sound channels : $\begin{bmatrix} M \\ S \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} L \\ R \end{bmatrix}$, wherein M denotes the residual sequence/frequency-domain coefficients of the sum sound channel, S denotes the residual sequence/frequency-domain coefficients of the difference channel, L denotes the residual sequence/frequency-domain coefficients of the left sound channel, and R denotes the residual sequence/frequency-domain coefficients of the right sound channel.

[0125] If the sum-difference stereo encoding is applied after the quantization, the quantized residual sequence/frequency-domain coefficients of the left-right sound channels at the scale factor band are linearly transformed and are replaced with the residual sequence/frequency-domain coefficients of the sum-difference sound channels: $\begin{bmatrix} \hat{M} \\ \hat{S} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} \hat{L} \\ \hat{R} \end{bmatrix}$, wherein $\hat{M}$ denotes the quantized residual sequence/frequency-domain coefficients of the sum sound channel, $\hat{S}$ denotes the quantized residual sequence/frequency-domain coefficients of the difference channel, $\hat{L}$ denotes the quantized residual sequence/frequency-domain coefficients of the left sound channel, and $\hat{R}$ denotes the quantized residual sequence/frequency-domain coefficients of the right sound channel.

[0126] Putting the sum-difference stereo encoding after the quantization can effectively eliminate the correlation between the left-right sound channels, meanwhile, a lossless encoding can be realized since the encoding is after the quantization.

[0127] Fig. 12 is a schematic drawing of embodiment two of the decoding device of the present invention. On the basis of the decoding device of Fig. 6, said decoding device has a sum-difference stereo decoding module 807 added between the output of the inverse quantizer bank 803 and the input of the inverse frequency-domain linear prediction and vector quantization module 804 to receive the result of signal type analysis and the sum-difference stereo control signal output from the bit-stream demultiplexing module 801, and to transform the inverse quantization spectrum of the sum-difference sound channels into the inverse quantization spectrum of the left-right sound channels according to said control information.

[0128] In the sum-difference control signal, there is a flag bit for indicating if the present sound channel pair needs a sum-difference stereo decoding. If they need, then there is also a flag bit on each scale factor to indicate if the corresponding scale factor band needs to be sum-difference stereo decoded, and the sum-difference stereo decoding module 66 determines, on the basis of the flag bit of the scale factor band, if it is necessary to perform sum-difference stereo decoding on the inverse quantization spectrum in some of the scale factor bands. If the sum-difference stereo encoding is performed in the encoding device, then the sum-difference stereo decoding must be performed on the inverse quantization spectrum in the decoding device.

[0129] The sum-difference stereo decoding module 807 can also be located between the output of the entropy decoding module 802 and the input of the inverse quantizer bank 803 to receive the sum-difference stereo control signal and the result of signal type analysis output from the bit-stream demultiplexing module 601.

[0130] The decoding method of the decoding device as shown in Fig. 12 is substantially the same as the decoding method of the decoding device as shown in Fig. 6, and the difference is that the former further includes the followng steps: after obtaining the inverse quantization spectrum, if the result of signal type analysis shows that the signal types are the same, it is determined whether it is necessary to perform a sum-difference stereo decoding on the inverse quantization spectrum according to the sum-difference stereo control signal. If it is necessary, it is determined, on the basis of the flag bit on each scale factor band, if said scale factor band needs a sum-difference stereo decoding. If it needs, the inverse quantization spectrum of the sum-difference sound channels in said scale factor band is transformed into inverse quantization spectrum of the left-right sound channels before the subsequent processing; if the signal types are not the same or it is unnecessary to perform the sum-difference stereo decoding, the inverse quantization spectrum is not processed and the subsequent processing is directly performed.

[0131] The sum-difference stereo decoding can also be performed after the entropy decoding and before the inverse quantization, that is, after obtaining the quantized values of the spectrum, if the result of signal type analysis shows that

the signal types are the same, it is determined whether it is necessary to perform a sum-difference stereo decoding on the quantized value sof the spectrum according to the sum-difference stereo control signal. If it is necessary, it is determined, on the basis of the flag bit on each scale factor band, if said scale factor band needs a sum-difference stereo decoding, if it needs, the quantized values of the spectrum of the sum-difference sound channels in said scale factor band are transformed into the quantized values of the spectrum of the left-right sound channels before the subsequent processing; if the signal types are not the same or it is unnecessary to perform the sum-difference stereo decoding, the quantized values of the spectrum are not processed and the subsequent processing is directly performed.

[0132] If the sum-difference stereo decoding is after the entropy decoding and before the inverse quantization, then the frequency-domain coefficients of the left-right sound channels in the scale factor band are obtained from the frequency-domain coefficients of the sum-difference sound channels through the equation of $\begin{bmatrix} \hat{l} \\ \hat{r} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \hat{m} \\ \hat{s} \end{bmatrix}$, wherein m denotes the quantized frequency-domain coefficients of the sum sound channel, $\hat{s}$ denotes the quantized frequency-domain coefficients of the difference channel, $\hat{l}$ denotes the quantized frequency-domain coefficients of the left sound channel, and $\hat{l}$ denotes the quantized frequency-domain coefficients of the right sound channel.

[0133] If the sum-difference stereo decoding is after the inverse quantization, then the inversely quantized frequency-domain coefficients of the left-right sound channels in the sub-band are obtained from the frequency-domain coefficients of the sum-difference sound channels through the computation of the matrix of $\begin{bmatrix} l \\ r \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} m \\ s \end{bmatrix}$, wherein, m denotes the frequency-domain coefficients of the sum sound channel, s denotes the frequency-domain coefficients of the difference channel, l denotes the frequency-domain coefficients of the left sound channel, and r denotes the frequency-domain coefficients of the right sound channel.

[0134] Fig. 13 is a schematic drawing of the structure of the third embodiment of the encoding device of the present invention. On the basis of the encoding device as shown in Fig. 5, said embodiment has a frequency band spreading module 58 and a re-sampling module 59 added. The frequency band spreading module 58 is used for analyzing the originally input audio signal on the entire frequency band to extract the spectrum envelope of the high frequency portion and the parameters representing the correlation between the low and high frequency spectrum, and to output them as the frequency band spreading information to the bit-stream multiplexing module 55; and the re-sampling module 59 is used for re-sampling the originally input audio signal to change the sampling rate thereof.

[0135] The re-sampling includes up-sampling and down-sampling. The re-sampling is described below using down-sampling as an example. In this embodiment, the re-sampling module 59 comprises a low-pass filter and a down-sampler, wherein the low-pass filter is used for limiting the frequency band of the audio signals and eliminating the aliasing that might be caused by down-sampling. The input audio signal is down-sampled after being low-pass filtered. Suppose that the input audio signal is s(n), and said signal is output as v(n) after being filtered by the low-pass filter having a pulse response of h(n), then $v(n) = \sum_{k=-\infty}^{\infty} h(k)s(n-k)$; the sequence of an M times of down-sampling on v(n) is x(n), then $x(m) = v(Mm) = \sum_{k=-\infty}^{\infty} h(k)s(Mm-k)$. Thus the sampling rate of the re-sampled audio signal x(n) is reduced by M times as compared to the sampling rate of the originally input audio signal s(n).

[0136] The basic principle of frequency band spreading is that with respect to most audio signals, there is a strong correlation between the characteristic of the high frequency portion thereof and the characteristic of the low frequency portion thereof, so the high frequency portions of the audio signals can be effectively reconstructed through the low frequency portions, thus the high frequency portions of the audio signals may not be transmitted. In order to ensure a correct reconstruction of the high frequency portions, only few frequency band spreading information are transmitted in the compressed audio code stream.

[0137] The frequency band spreading module 58 comprises a parameter extracting module and a spectrum envelope extracting module. Signals are input to the parameter extracting module which extracts the parameters representing the spectrum characteristics of the input signals at different time-frequency regions, then in the spectrum envelope extracting module, the spectrum envelope of the high frequency portion of the signal is estimated at a certain time-frequency resolution. In order to ensure that the time-frequency resolution is most suitable for the characteristics of the present input signals, the time-frequency resolution of the spectrum envelope can be selected freely. The parameters of the spectrum characteristics of the input signals and the spectrum envelope of the high frequency portion are used as the

output of frequency band spreading to be sent to the bit-stream multiplexing module 55 for multiplexing.

[0138] The encoding method based on the encoding device as shown in Fig. 13 is substantially the same as the encoding method based on the encoding device as shown in Fig. 5, and the difference is that the former further includes the following steps: re-sampling the audio signal before analyzing the type thereof; analyzing the input audio signal on the entire frequency band to extract the high frequency spectrum envelope and the parameters of the signal spectrum characteristics thereof as the control signal of the frequency-band spreading, which are multiplexed together with the audio encoded signal and the side information to obtain the compressed audio code stream. Wherein the re-sampling includes the two steps of limiting the frequency band of the audio signal and performing a multiple down-sampling on the audio signal whose frequency band is limited.

[0139] Fig. 14 is a schematic drawing of the structure of embodiment three of the decoding device of the present invention. On the basis of the decoding device as shown in Fig. 6, said decoding device has a frequency band spreading module 808 added, which receives the frequency band spreading control information output from the bit stream demultiplexing module 801 and the time-domain audio signal of low frequency channel output from the frequency-time mapping module 805, and which reconstructs the high frequency signal portion through spectrum shift and high frequency adjustment to output the wide band audio signal.

[0140] The decoding method based on the decoding device as shown in Fig. 14 is substantially the same as the decoding method based on the decoding device as shown in Fig. 6, and the difference lies in that the former further includes the step of reconstructing the high frequency portion of the time-domain audio signal according to the frequency band spreading control information and the time-domain audio signal; thereby to obtain the wide band audio signal.

[0141] Fig. 15 is a schematic drawing of the structure of the fourth embodiment of the encoding device of the present invention, which has a frequency band spreading module 58 and a re-sampling module 59 added on the basis of the encoding device as shown in Fig. 7. In this embodiment, the connection between said frequency band spreading module 58 and re-sampling module 59 and other modules, and the function and operation principle of these two modules are the same as those shown in Fig. 13, so they will not be elaborated herein.

[0142] The encoding method based on the encoding device as shown in Fig. 15 is substantially the same as the encoding method based on the encoding device as shown in Fig. 7, and the difference is that the former further includes the following steps: re-sampling the audio signal before analyzing the type thereof; analyzing the input audio signal on the entire frequency band to extract the high frequency spectrum envelope and the parameters of the spectrum characteristics thereof; and multiplexing them together with the audio encoded signal and the side information to obtain the compressed audio code stream.

[0143] Fig. 16 is a schematic drawing of embodiment four of the decoding device of the present invention. On the basis of the decoding device as shown in Fig. 10, said decoding device has a frequency band spreading module 808 added. In this embodiment, the connection between said frequency band spreading module 808 and other modules, and the function and operation principle thereof are the same as those shown in Fig. 14, so they will not be elaborated herein:

[0144] The decoding method based on the decoding device as shown in Fig. 16 is substantially the same as the decoding method based on the decoding device as shown in Fig. 10, and the difference is that said decoding method further includes the step of reconstructing the high frequency portion of the audio signal according to the frequency band spreading control information and the time-domain audio signal, thereby to obtain audio signal of wide frequency band.

[0145] Fig. 17 is a schematic drawing of the structure of the fifth embodiment of the encoding device of the present invention. On the basis of the encoding device as shown in Fig. 7, said embodiment has a sum-difference stereo encodingmodule 57 added between the output of the multi-resolution analyzing module 56 and the input of the quantization and entropy encoding module 54, or between the quantizer bank and the encoder in the quantization and entropy encoding module 54. In this embodiment, the function and operation principle of the sum-difference stereo encoding module 57 are the same as those shown in Fig. 11, so they will not be elaborated herein.

[0146] The encoding method of the encoding device as shown in Fig. 17 is substantially the same as the encoding method of the encoding device as shown in Fig. 7, and the difference is that the former further includes the steps of determining whether the audio signals are multi-channel signals after multi-resolution analysis of the residual sequence/frequency-domain coefficients. If they are multi-channel signals, determining whether the types of the signals of the left-right sound channels are the same, and if the signal types are the same, determining whether the scale factor bands meet the encoding conditions. If they meet the conditions, performing a sum-difference stereo encoding on residual sequence/frequency-domain coefficients to obtain the residual sequence/frequency-domain coefficients of the sum-difference sound channels; if they do not meet the conditions, the sum-difference stereo encoding is not performed. If the signals are mono signals or multi-channel signals of different types, the frequency-domain coefficients are not processed. The specific flow thereof has been described above, so it will not be elaborated again.

[0147] Fig. 18 is a schematic drawing of the structure of embodiment five of the decoding device of the present invention. On the basis of the decoding device as shown in Fig. 10, said decoding device has a sum-difference stereo decoding module 807 added between the output of the inverse quantizer bank 803 and the input of the multi-resolution

integration module 806, or between the output of the entropy decoding module 802 and the input of the inverse quantizer bank 803. In this embodiment, the function and operation principle of the sum-difference stereo decoding module 807 are the same as those shown in Fig. 12, so they will not be elaborated herein.

[0148] The decoding method of the decoding device as shown in Fig. 18 is substantially the same as the decoding method of the decoding device as shown in Fig. 10, and the difference is that the former further includes the following steps: after obtaining the inverse quantization spectrum, if the result of signal type analysis shows that the signal types are the same, it is determined whether it is necessary to perform a sum-difference stereo decoding on the inverse quantization spectrum according to the sum-difference stereo control signal. If it is necessary, it is determined, on the basis of the flag bit on each scale factor band, if said scale factor band needs a sum-difference stereo decoding, and if it needs, the inverse quantization spectrum of the sum-difference sound channels in said scale factor band is transformed into inverse quantization spectrum of the left-right sound channels before the subsequent processing; if the signal types are not the same or it is unnecessary to perform the sum-difference stereo decoding, the inverse quantization spectrum is not processed and the subsequent processing is directly performed. The specific flow thereof has been described above, so it will not be elaborated again.

[0149] Fig. 19 is the schematic drawing of the sixth embodiment of the encoding device of the present invention. On the basis of Fig. 17, this embodiment has a frequency band spreading module 58 and a re-sampling module 59 added. In this embodiment, the connection between said frequency band spreading module 58 and re-sampling module 59 and other modules, and the functions and operation principles of these two modules are the same as those in Fig. 13, so they will not be elaborated herein.

[0150] The encoding method based on the encoding device as shown in Fig. 19 is substantially the same as the encoding method based on the encoding device as shown in Fig. 17, and the difference is that the former further includes the following steps: re-sampling the audio signal before analyzing the type thereof; analyzing the input audio signal on the entire frequency band to extract the high frequency spectrum envelope and the parameters of the spectrum characteristics thereof; and multiplexing them together with the audio encoded signal and the side information to obtain the compressed audio code stream.

[0151] Fig. 20 is a schematic drawing of embodiment six of the decoding device of the present invention.On the basis of the decoding device as shown in Fig. 18, said decoding device has a frequency band spreading module 808 added. In this embodiment, the connection between said frequency band spreading module 808 and other modules, and the function and principle thereof are the same as those shown in Fig. 14, so they will not be elaborated herein.

[0152] The decoding method based on the decoding device as shown in Fig. 20 is substantially the same as the decoding method based on the decoding device as shown in Fig. 18, and the difference is that said decoding method further includes the step of reconstructing the high frequency portion of the audio signal according to the frequency band spreading control information and the time-domain audio signal, thereby to obtain audio signals of wide frequency band.

[0153] Fig. 21 is a schematic drawing of the seventh embodiment of the encoding device of the present invention. On the basis of Fig. 11, said embodiment has a frequency band spreading module 58 and a re-sampling module 59 added. In this embodiment, the connection between said frequency band spreading module 58 and re-sampling module 59 and other modules, and the functions and operation principles of said two modules are the same as those in Fig. 14, so they will not be elaborated herein.

[0154] The encoding method of the encoding device as shown in Fig. 21 is substantially the same as the encoding method of the encoding device as shown in Fig. 11, and the difference is that said encoding method further includes the steps of re-sampling the audio signal before analyzing the type thereof; analyzing the input audio signal on the entire frequency band to extract the high frequency spectrum envelope and the parameters of the spectrum characteristics thereof; and multiplexing them together with the audio encoded signal and the side information to obtain the compressed audio code stream.

[0155] Fig. 22 is a schematic drawing of embodiment seven of the decoding device of the present invention. On the basis of the decoding device as shown in Fig. 12, said decoding device has a frequency band spreadingmodule 808 added. In this embodiment, the connection between said frequency band spreading module 808 and other modules, and the function and principle thereof are the same as those shown in Fig. 14, so they will not be elaborated herein.

[0156] The decoding method based on the decoding device as shown in Fig. 22 is substantially the same as the decoding method based on the decoding device as shown in Fig. 12, and the difference is that said decoding method further includes the step of reconstructing the high frequency portion of the audio signal according to the frequency band spreading control information and the time-domain audio signal, thereby to obtain audio signals of wide frequency band.

[0157] The seven embodiments of the encoding device as described above may also include a gain control module which receives the audio signals output from the signal type analyzing module 59, controls the dynamic range of the fast varying type signals, and eliminates the pre-echo in audio processing. The output thereof is connected to the time-frequency mapping module 52 and the psychoacoustical analyzing module 51, meanwhile, the amount of gain adjustment is output to the bit-stream multiplexing module 55.

[0158] According to the signal type of the audio signal, the gain control module controls only the fast varying type

signal, while the slowly varying signal is directly output without being processed. As for the fast varying type signal, the gain control module adjusts the time-domain energy envelope of the signal to increase the gain value of the signal before the fast varying point, so that the amplitudes of the time-domain signal before and after the fast varying point are close to each other; then the time-domain signal whose time-domain energy envelope is adjusted is output to the time-frequency mapping module 52 Meanwhile, the amount of gain adjustment is output to the bit-stream multiplexing module 55.

[0159] The encoding method based on said encoding device is substantially the same as the encoding method based on the above described encoding device, and the difference lies in that the former further includes the step of performing a gain control on the signal whose signal type has been analyzed.

[0160] The seven embodiments of the decoding device as described above may also include an inverse gain control module which is located after the output of the frequency-time mapping module 805 to receive the result of signal type analysis and the information of the amount of gain adjustment output from the bit-stream demultiplexing module 801, thereby adjusting the gain of the time-domain signal and controlling the pre-echo. After receiving the reconstructed time-domain signal output from the frequency-time mapping module 805, the inverse gain control module controls the fast varying signals but leaves the slowly varying signals unprocessed. As for the signal of a fast varying type, the inverse gain control module adjusts the energy envelope of the reconstructed time-domain signal according to the information of the amount of gain adjustment, reduces the amplitude value of the signal before the fast varying point, and adjusts the energy envelope back to the original state of low in the front and high in the back. Thus the amplitude value of the quantified noise before the fast varying point will be reduced along with the amplitude value of the signal, thereby controlling the pre-echo.

[0161] The decoding method based on said decoding device is substantially the same as the decoding method based on the above described decoding device, and the difference lies in that the former further includes the step of performing an inverse gain control on the reconstructed time-domain signals.

[0162] Finally, it has to be noted that the above-mentioned embodiments illustrate rather than limit the technical solutions of the invention. While the invention has been described in conjunction with preferred embodiments, those skilled in the art shall understand that that modifications or equivalent substitutions can be made to the technical solutions of the present invention without deviating from the spirit and scope of the technical solutions of the present invention. Accordingly, it is intended to embrace all such modifications or equivalent substitutions as fall within the scope of the appended claims

Claims

1. An enhanced audio encoding device, comprising a psychoacoustical analyzing module, a time-frequency mapping module, a quantization and entropy encoding module, and a bit-stream multiplexing module, **characterized in that** said device further comprises a signal type analyzing module and a frequency-domain linear prediction and vector quantization module; wherein
   the signal type analyzing module is configured to analyze the signal type of the input audio signal and output the audio signal to the psychoacoustical analyzing module and the time-frequency mapping module, and to output the result of signal type analysis to the bit-stream multiplexing module at the same time;
   the psychoacoustical analyzing module is configured to calculate a masking threshold and a signal-to-masking ratio of the audio signal whose signal type has been analyzed, and output them to said quantization and entropy encoding module; the time-frequency mapping module is configured to convert the time-domain audio signal into frequency-domain coefficients;
   the frequency-domain linear prediction and vector quantization module is configured to perform a linear prediction on the frequency-domain coefficients, convert the produced prediction coefficients into the line spectrum pair frequency coefficients, and to perform a multi-stage vector quantization on the line spectrum pair frequency coefficients, and then to output the prediction residual sequence of the frequency-domain coefficients to the quantization and entropy encoding module, and output the side information to the bit-stream multiplexing module;
   the quantization and entropy encoding module is configured to perform quantization and entropy encoding on the residual sequence/frequency-domain coefficients under the control of the signal-to-masking ratio output from the psychoacoustical analyzing module, and to output them to the bit-stream multiplexing module; and
   the bit-stream multiplexing module is configured to multiplex the received data to form audio encoding code stream.

2. The enhanced audio encoding device according to claim 1, **characterized in that** the frequency-domain linear prediction and vector quantization module consists of a linear prediction analyzer, a linear prediction filter, a transformer, and a vector quantizer;
   the linear prediction analyzer is used for predictive analyzing the frquency-domain coefficients to obtain the prediction gain and prediction coefficients, and for outputting the frequency-domain coefficients that meet a certain condition

to the linear prediction filter; while the prediction coefficients that do not meet the condition are directly output to said quantization and entropy encoding module;
the linear prediction filter is used for filtering the frequency-domain coefficients to obtain the residual sequence of the frequency-domain coefficients, and for outputting the residual sequence to the quantization and entropy encoding module and outputting the prediction coefficients to the transformer;
the transformer is used for transforming the prediction coefficients into line spectrum pair frequency coefficients;
the vector quantizer is used for performing multi-stage vector quantization on the line spectrum pair frequency coefficients, and the quantized signals are transmitted to the bit-stream multiplexing module.

3. The enhanced audio encoding device according to claim 1, further comprising a sum-difference stereo encoding module located between the output of the frequency-domain linear prediction and vector quantization module or the multi-resolution analyzing module and the input of the quantization and entropy encoding module, or between the quantizer bank and the encoder in the quantization and entropy encoding module, which is used for transforming the frequency-domain coefficents/ residual sequence of the left-right sound channels into the frequency-domain coefficents/residual sequence of the sum-difference sound channel.

4. The enhanced audio encoding device according to any one of claims 1-3, further comprising a re-sampling module and a frequency band spreading module; wherein
the re-sampling module is used for re-sampling the input audio signal to change the sampling rate thereof; said re-sampling module comprises a low-pass filter and a down-sampler; wherein the low-pass filter is used for limiting the frequency band of the audio signal, and the down-sampler is used for down-sampling the signal to reduce the sampling rate of the signal;
the frequency-band spreading module is used for analyzing the original input audio signal on the entire frequency band to extract the spectrum envelope of the high frequency portion and the parameters representing the correlation between the low and high frequency portion and to output them to the bit-stream multiplexing module; said frequency-band spreading module comprises a parameter extracting module and a spectrum envelope extracting module; said parameter extracting module is used for extracting the parameters representing the spectrum characteristics of the input signal at different time-frequency regions, and said spectrum envelope extracting module is used for estimating the spectrum envelope of the high frequency portion of the signal at a certain time-frequency resolution, and then outputting the parameters of the spectrum characteristics of the input signal and the spectrum envelope of the high frequency portion to the bit-stream multiplexing module.

5. An enhanced audio encoding method, comprising the following steps:

step 1: analyzing the type of the input audio signal and using the result of analysis of the signal type as a part of signal multiplexing;
step 2: calculating the signal-to-masking ratio of the signal whose signal type has been analyzed;
step 3: time-frequency mapping the type analyzed signal to obtain the frequency-domain coefficients of the audio signal; step 4: performing a standard linear prediction analysis on the frequency-domain coefficients to obtain the prediction gain and the prediction coefficients; determining if the prediction gain exceeds the predetermined threshold, and if it does, a frequency-domain linear prediction error filtering is performed on the frequency-domain coefficients based on the prediction coefficients to obtain the residual sequence; transforming the prediction coefficients into line spectrum pair frequency coefficients, and performing a multi-stage vector quantization on said line spectrum pair frequency coefficients to obtain the side information; if the prediction gain does not exceed the predetermined threshold, proceeding to step 5 without processing the frequency-domain coefficients;
step 5: quantizing and entropy encoding the residual sequence/frequency-domain coefficients;
step 6: multiplexing the side information and the encoded audio signal to obtain the compressed audio code stream.

6. The enhanced audio encoding method according to claim 5, **characterized in that** the quantization in step 5 is scalar quantization which comprises non-linearly companding the frequency-domain coefficients in all the scale factor bands; using the scale factor of each sub-band to quantize the frequency-domain coefficients of said sub-band to obtain the quantization spectrum represented by an integer; selecting the first scale factor in each frame of signal as the common scale factor; and differentiating the rest of the scale factors from their respective previous scale factor;
the entropy encoding comprises entropy encoding the quantization spectrum and the differentiated scale factors to obtain the sequence numbers of the code book, the encoded values of the scale factors and the quantization

spectrum of lossless encoding; and entropy encoding the sequence numbers of the code book to obtain the encoded values thereof.

7. The enhanced audio encoding method according claim 5 or 6, **characterized in that** said step 5 further comprises quantizing the residual sequence/frequency-domain coefficients; determining if the audio signals are multi-channel signals, and if they are, determining if the signals of the left-right sound channels are of the same type; if the signal types are the same, determining if the scale factor bands corresponding to the two sound channels meet the conditions of sum-difference stereo encoding, and if they meet the conditions, performing a sum-difference stereo encoding on the residual sequence/frequency-domain coefficients in said scale factor bands of the two sound channels to obtain the residual sequence/frequency-domain coefficients of the sum-diffrence sound channels; if they do not meet the conditions, not performing the sum-difference stereo encoding on the residual sequence/frequency-domain coefficientsinsaidscale factor bands; if the signals are mono signals or multi-channel signals of different types, not processing the residual sequence/frequency-domain coefficients; and entropy encoding the residual sequence/frequency-domain coefficients; wherein the method for determining whether the scale factor band meets the condition for encoding is K-L transformation, specifically, the correlation matrix of the spectrum coefficients of the scale factor bands of the left-right sound channels are calculated; a K-L transformation is performed on the correlation matrix; if the absolute value of the rotation angle a deviates from $\pi/4$ by a small amount, e.g. $3n/16<|a|<5\pi/16$, a sum-difference stereo encoding can be performed on the corresponding scale factor bands; said sum-difference stereo encoding is $\begin{bmatrix}\hat{M}\\ \hat{S}\end{bmatrix}=\begin{bmatrix}1 & 0\\ 1 & -1\end{bmatrix}\begin{bmatrix}\hat{L}\\ \hat{R}\end{bmatrix}$, wherein $\hat{M}$ denotes the quantized frequency-domain coefficients of the sum sound channel, $\hat{S}$ denotes the quantized frequency-domain coefficients of the difference channel, $\hat{L}$ denotes the quantized frequency-domain coefficients of the left sound channel, and $\hat{R}$ denotes the quantized frequency-domain coefficients of the right sound channel.

8. The enhanced audio encoding method according to any one of claims 5-7, **characterized in that** before said step 1, there is a step of re-sampling the input audio signal, which specifically includes: limiting the frequency band of the audio signal and performing a multiple down-sampling on the audio signal whose frequency band is limited; and after said step 6, there is a step of analyzing the original input audio signal before the re-sampling on the entire frequency band to extract the high frequency spectrum envelope and the parameters of the signal spectrum characteristics thereof; and multiplexing them together with the audio encoded signal and the side information to obtain the compressed audio code stream.

9. An enhanced audio decoding device, comprising a bit-stream demultiplexing module, an entropy decoding module, an inverse quantizerbank, a frequency-time mapping module,**characterized in that** said device further comprises an inverse frequency-domain linear prediction and vector quantization module;
the bit-stream demultiplexing module is configured to demultiplex the compressed audio data stream and output the corresponding data signal and control signal to the entropy decoding module and the inverse frequency-domain linear prediction and vector quantization module;
the entropy decodingmodule is configured to decode said signals, recover the quantized values of the spectrum so as to output to the inverse quantizer bank;
the inverse quantizer bank is configured to reconstruct the inverse quantization spectrum and output it to the inverse frequency-domain linear prediction and vector quantization module;
the inverse frequency-domain linear prediction and vector quantization module is configured to perform inverse linear prediction filtering on the inverse quantization spectrum to obtain the spectrum-before-prediction and output it to the frequency-time mapping module; and
the frequency-time mapping module is configured to perform a frequency-time mapping on the spectrum coefficients to output the time-domain audio signal.

10. The enhanced audio decoding device according to claim 9, **characterized in that** the inverse frequency-domain linear prediction and vector quantization module comprises an inverse vector quantizer, an inverse transformer, and an inverse linear prediction filter ; wherein the inverse vector quantizer is used for inversely quantizing the code word indexes to obtain the line spectrum pair frequency coefficients, the inverse transformer is used for inversely transforming the line spectrum pair frequency coefficients into prediction coefficients, and the inverse linear prediction filter is used for inversely filtering the inverse quantization spectrum based on the prediction coefficients to obtain the spectrum-before-prediction.

11. The enhanced audio decoding device according to claim 9 or 10, further comprising a sum-difference stereo decoding module located between the output of the inverse quantizer bank and the input of the multi-resolution integration module or the inverse frequency-domain linear prediction and vector quantization module, or between the output of the entropy decoding module and the input of the inverse quantizer bank, to receive the result of signal type analysis and the sum-difference stereo control signal output from the bit-stream demultiplexing module, and to transform the inverse quantization spectrum of the sum-difference sound channels into the inverse quantization spectrum of the left-right sound channels based on said control information.

12. An enhanced audio decodingmethod, comprising the following steps:

step 1: demultiplexing the compressed audio data stream to obtain the data information and the control information;
step 2: entropy decoding' said information to obtain the quantized values of the spectrum;
step 3: inversely quantizing the quantized values of the spectrum to obtain the inverse quantization spectrum;
step 4: determining if the control information contains information concerning that the inverse quantization spectrum has to undergo the inverse frequency-domain linear prediction vector quantization, and if it.does, performing the inverse vector quantization to obtain the prediction coefficients, and performing a linear prediction synthesizing on the inverse quantization spectrum according to the prediction coefficients to obtain the spectrum-before-prediction; if it does not, proceeding to step 5 without processing the inverse quantization spectrum;
step 5: performing a frequency-time mapping on the spectrum-before-prediction/inverse quantization spectrum to obtain the time-domain audio signal of low frequency band.

13. The enhanced audio decoding method according to claim 12, **characterized in that** said inverse vector quantization step further comprises obtaining from the control information the code word indexes resulted from the vector quantization of the prediction coefficients; and obtaining the quantized line spectrum pair frequency coefficients according to the code word indexes and calculating the prediction coefficients therefrom.

14. The enhanced audio decoding method according to claim 12, **characterized in that** said step 5 further comprises performing inverse modified discrete cosine transformation on the inverse quantization spectrum to obtain the transformed time-domain signals; performing a window adding processing on the transformed time-domain signals in the time domain; superposing said window added time-domain signals to obtain the time-domain audio signals; wherein the function window in said window adding processing is: $w(N+k)$ =cos ($pi/2$*((k+0.5)/N-0.94*sin $(2*pi/N*(k+0.5))/(2*pi)$)), wherein k=0...N-1; $w(k)$ represents the kth coefficient of the window function and $w(k) = w(2*N-1-k)$; N represents the number of samples of the encoded frame.

15. The enhanced audio decoding method according to any one of claims 12-14, **characterized in that** between said steps 2 and 3, there are also the steps that if the result of signal type analysis shows that the signal types are the same, it is determined whether it is necessary to perform a sum-difference stereo decoding on the quantized value of the spectrum according to the sum-difference stereo control signal, and if it is necessary, it is determined, on the basis of the flag bit on each scale factor band, if said scale factor band needs a sum-difference stereo decoding; if it needs, the quantized values of the spectrum of the sum-difference sound channels in said scale factor band are transformed into the quantized values of the spectrum of the left-right sound channels, and proceed to step 3; if the signal types are not the same or it is unnecessary to perform the sum-difference stereo decoding, the quantized values of the spectrum are not processed and proceed to step 3;

wherein the sum-difference stereo decoding is $\begin{bmatrix} \hat{l} \\ \hat{r} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \hat{m} \\ \hat{s} \end{bmatrix}$, wherein $\hat{m}$ denotes the quantized value of the spectrum of the quantized sum sound channel, $\hat{s}$ denotes the quantized value of the spectrum of the quantized difference channel, $\hat{l}$ denotes the quantized value of the spectrum of the quantized left sound channel, and $\hat{r}$ denotes the quantized value of the spectrum of the quantized right sound channel.

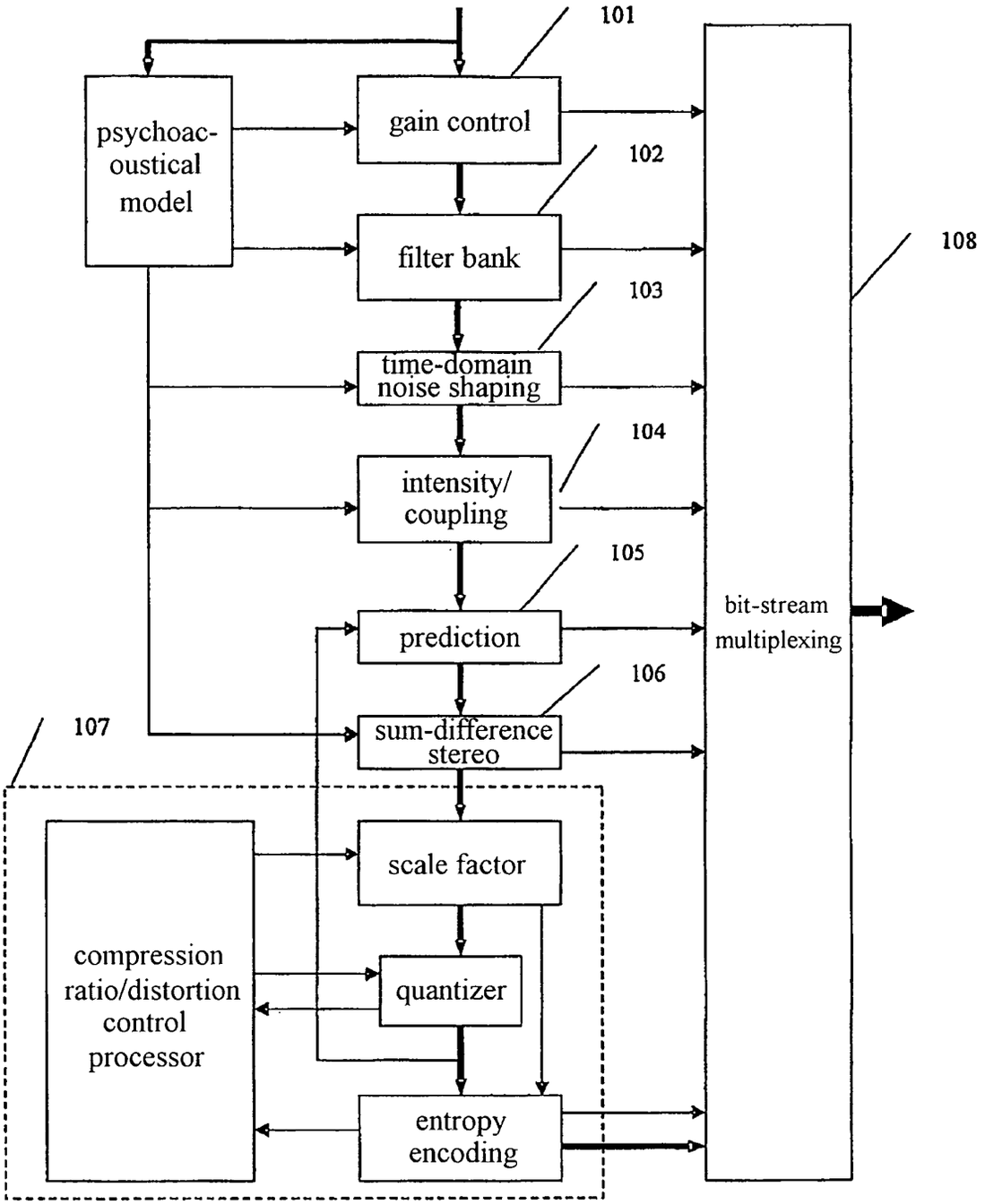
101
psychoac-
oustical
model
gain control
102
filter bank
103
time-domain
noise shaping
104
intensity/
coupling
105
prediction
106
sum-difference
stereo
107
scale factor
compression
ratio/distortion
control
processor
quantizer
entropy
encoding
bit-stream
multiplexing
108

Fig. 1

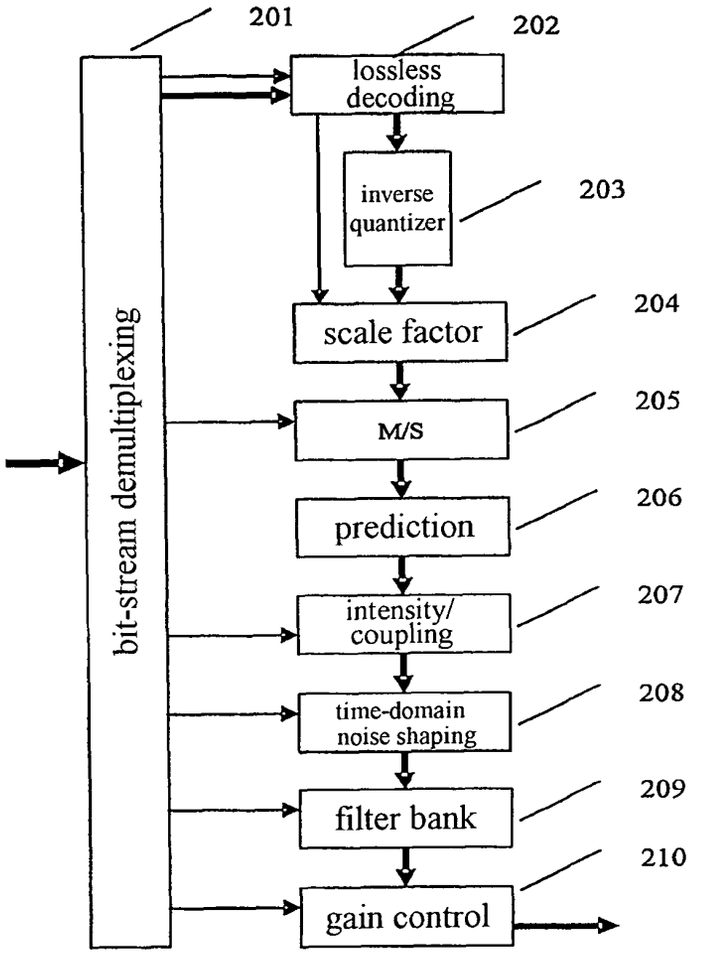
201
202
lossless decoding
inverse quantizer
203
bit-stream demultiplexing
scale factor
204
M/S
205
prediction
206
intensity/ coupling
207
time-domain noise shaping
208
filter bank
209
gain control
210

Fig. 2

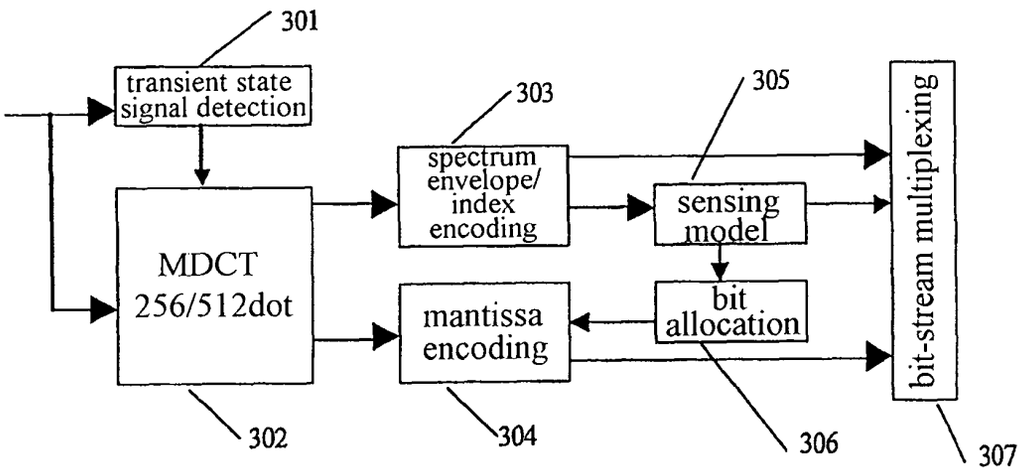
301
transient state signal detection
303
305
spectrum envelope/ index encoding
sensing model
bit-stream multiplexing
MDCT 256/512dot
mantissa encoding
bit allocation
302
304
306
307

Fig. 3

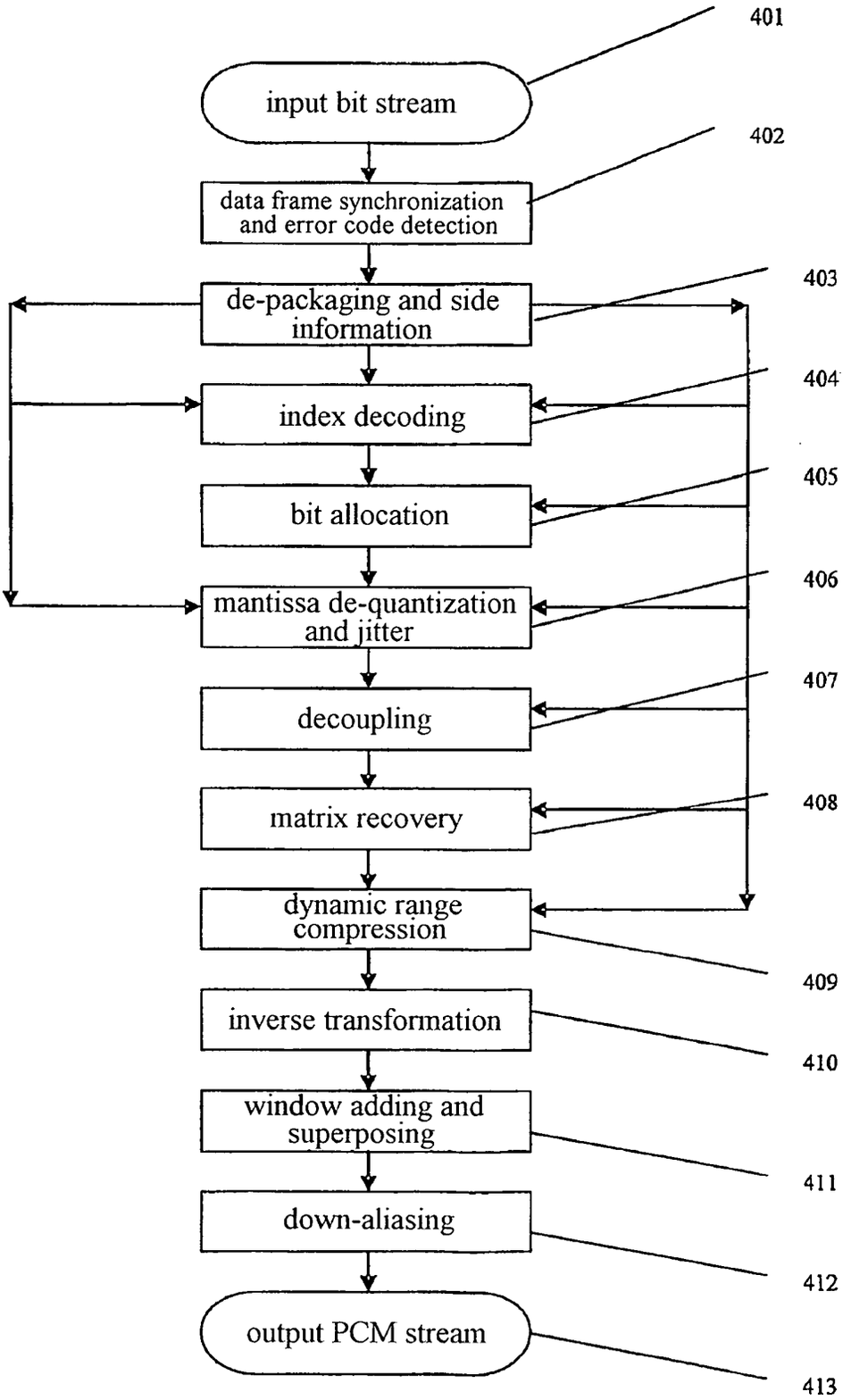
401
input bit stream
402
data frame synchronization and error code detection
403
de-packaging and side information
404
index decoding
405
bit allocation
406
mantissa de-quantization and jitter
407
decoupling
408
matrix recovery
dynamic range compression
409
inverse transformation
410
window adding and superposing
411
down-aliasing
412
output PCM stream
413

Fig. 4

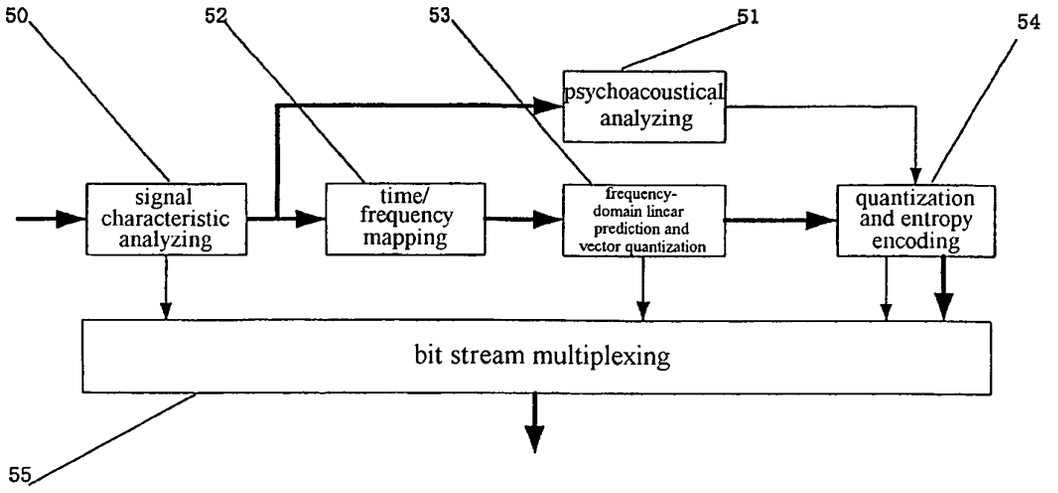
50
52
53
51
54
psychoacoustical analyzing
signal characteristic analyzing
time/ frequency mapping
frequency-domain linear prediction and vector quantization
quantization and entropy encoding
bit stream multiplexing
55

Fig. 5

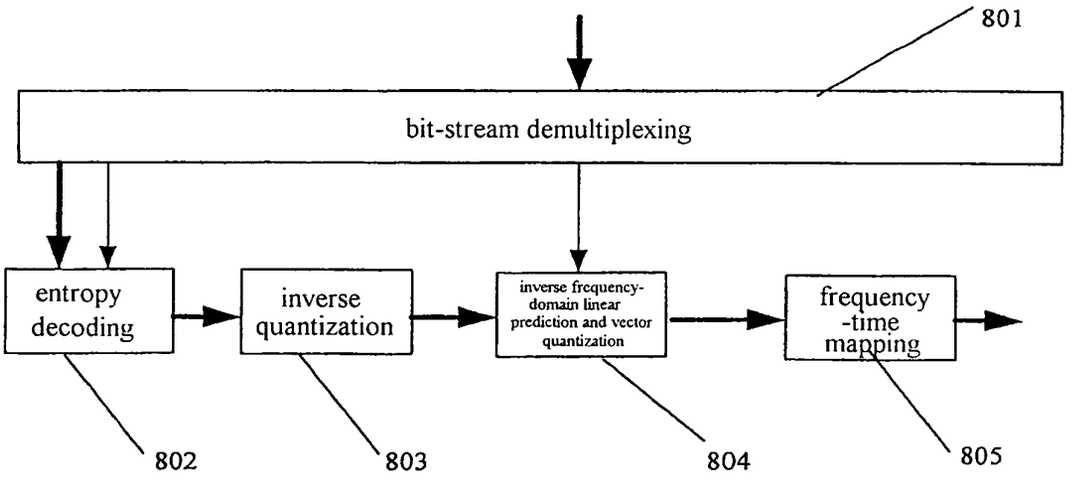
801
bit-stream demultiplexing
entropy decoding
inverse quantization
inverse frequency-domain linear prediction and vector quantization
frequency -time mapping
802
803
804
805

Fig. 6

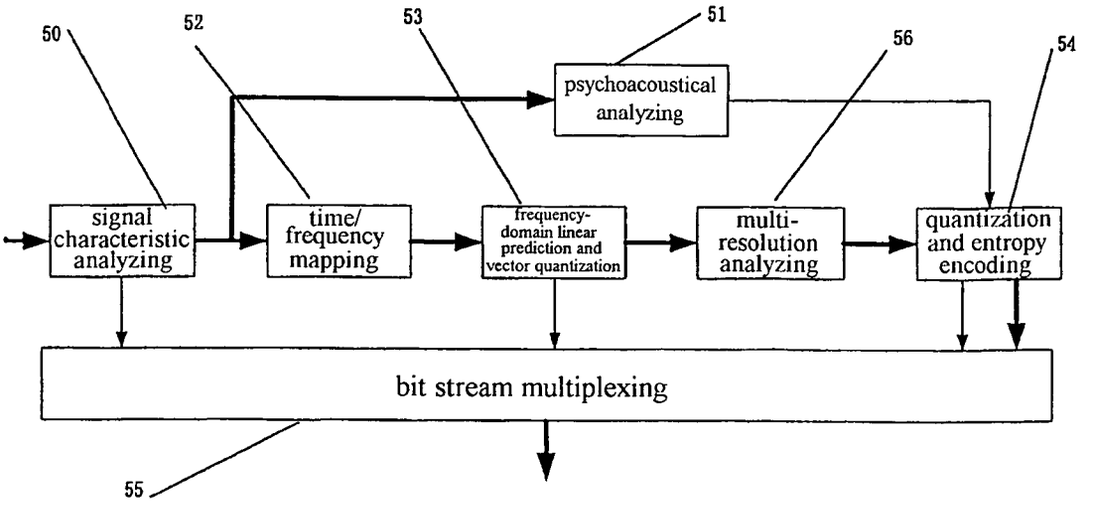

Fig. 7

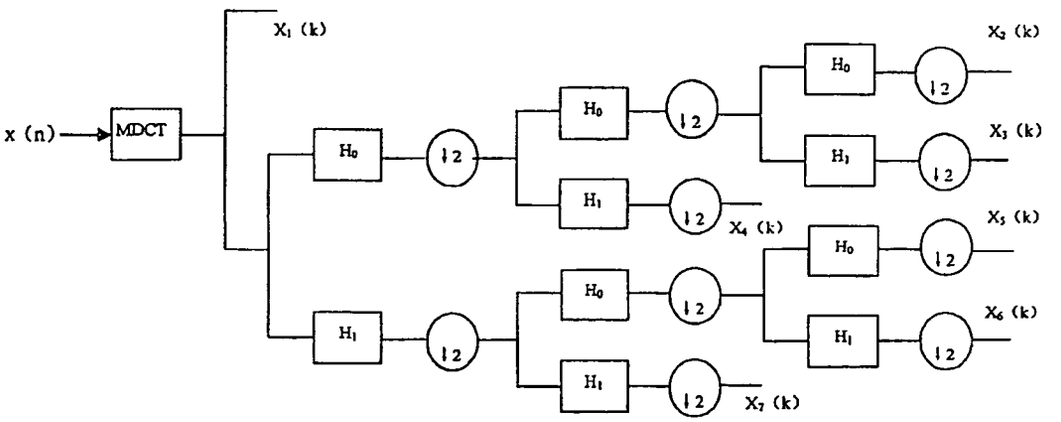

Fig. 8

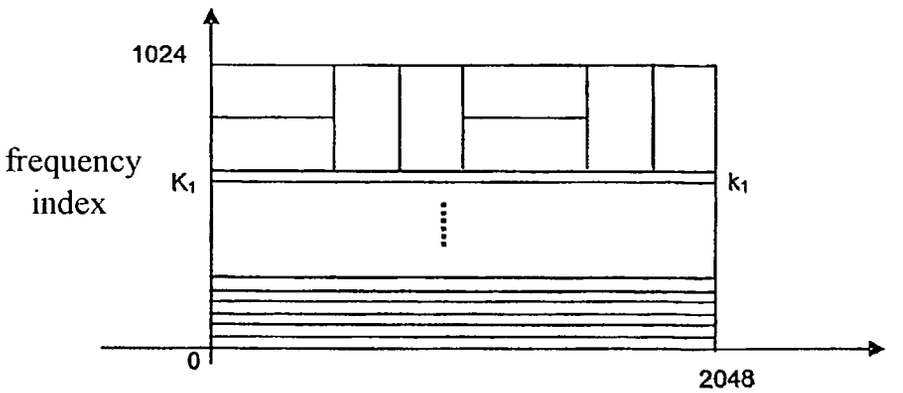

Fig. 9

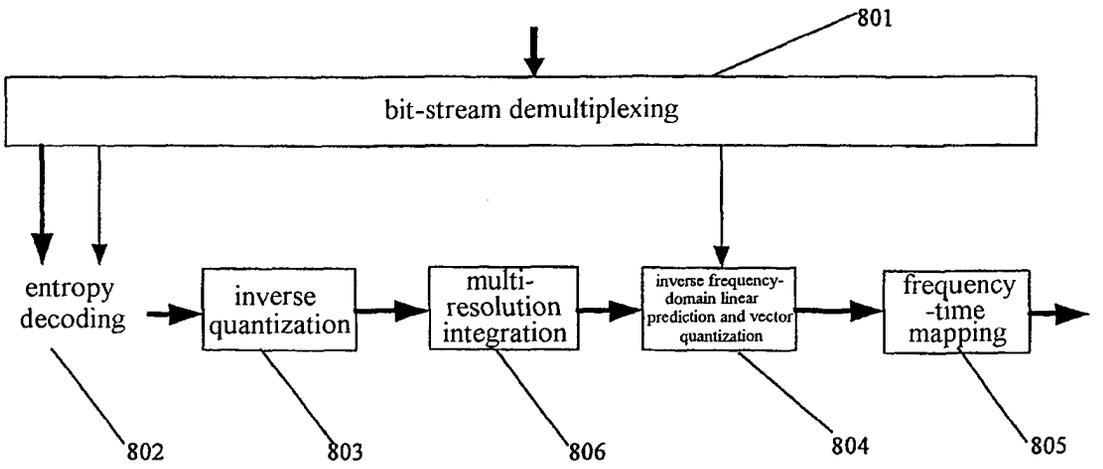
801
bit-stream demultiplexing
entropy decoding
inverse quantization
multi-resolution integration
inverse frequency-domain linear prediction and vector quantization
frequency -time mapping
802
803
806
804
805

Fig. 10

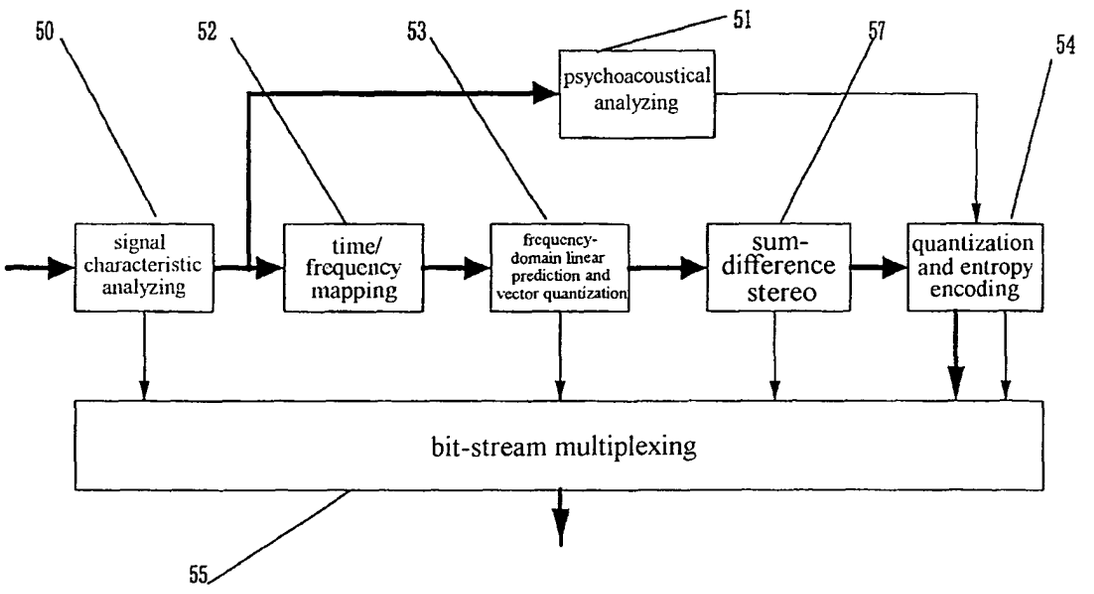
50
52
53
51
57
54
psychoacoustical analyzing
signal characteristic analyzing
time/ frequency mapping
frequency-domain linear prediction and vector quantization
sum-difference stereo
quantization and entropy encoding
bit-stream multiplexing
55

Fig. 11

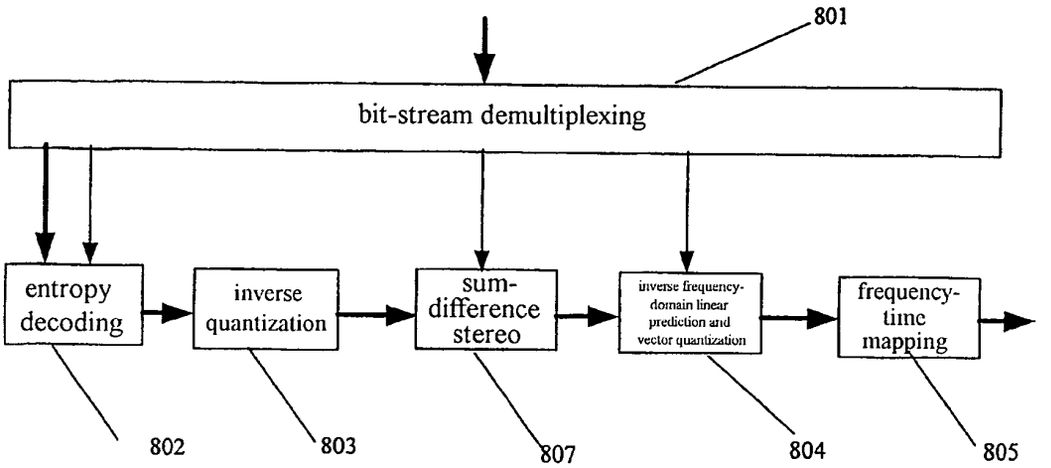
801
bit-stream demultiplexing
entropy decoding
inverse quantization
sum-difference stereo
inverse frequency-domain linear prediction and vector quantization
frequency-time mapping
802
803
807
804
805

Fig. 12

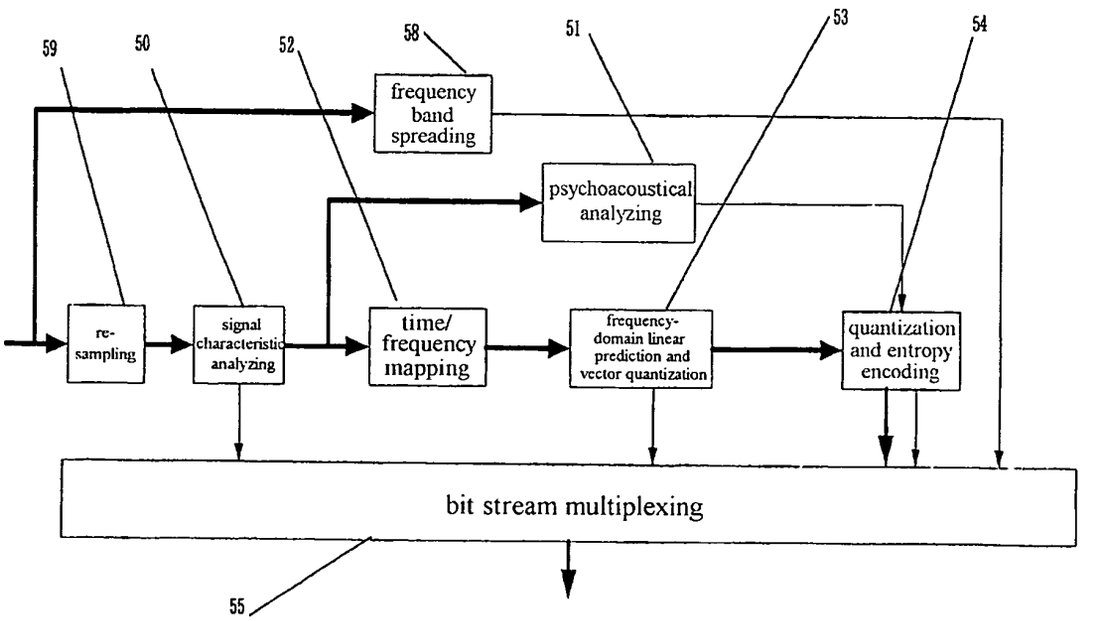
59
50
52
58
51
53
54
frequency band spreading
psychoacoustical analyzing
re-sampling
signal characteristic analyzing
time/ frequency mapping
frequency-domain linear prediction and vector quantization
quantization and entropy encoding
bit stream multiplexing
55

Fig. 13

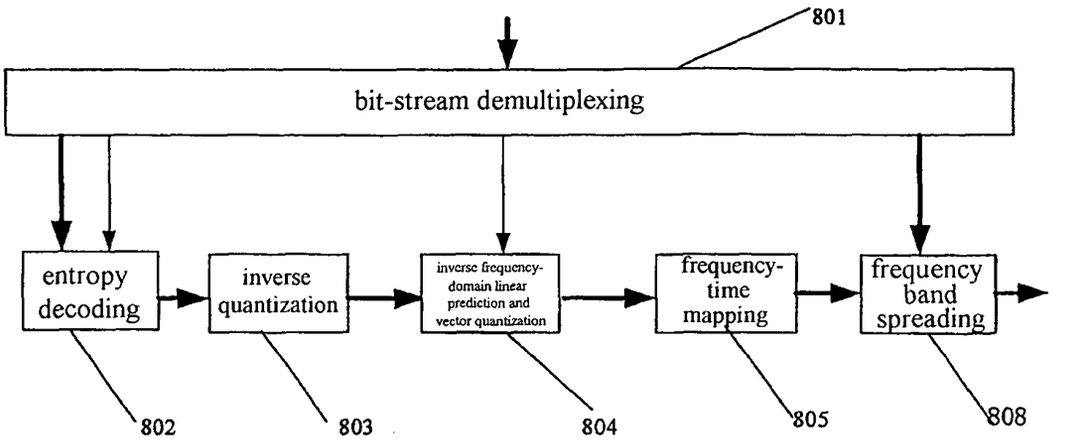
801
bit-stream demultiplexing
entropy decoding
inverse quantization
inverse frequency-domain linear prediction and vector quantization
frequency-time mapping
frequency band spreading
802
803
804
805
808

Fig. 14

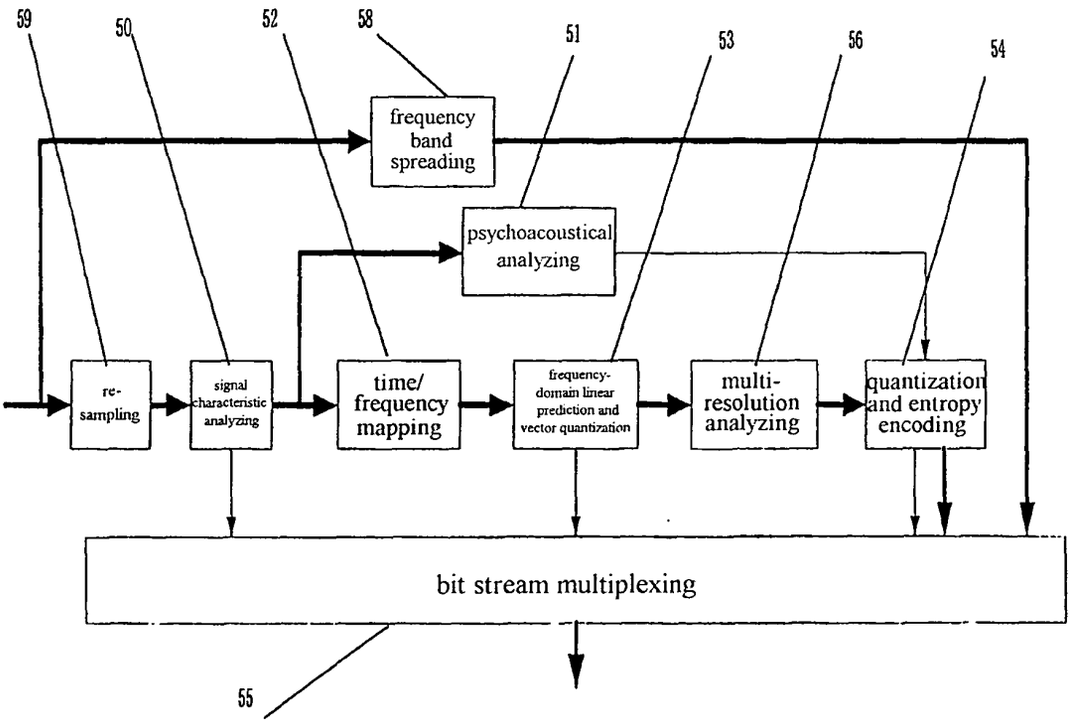
59
50
52
58
51
53
56
54
frequency band spreading
psychoacoustical analyzing
re-sampling
signal characteristic analyzing
time/ frequency mapping
frequency-domain linear prediction and vector quantization
multi-resolution analyzing
quantization and entropy encoding
bit stream multiplexing
55

Fig. 15

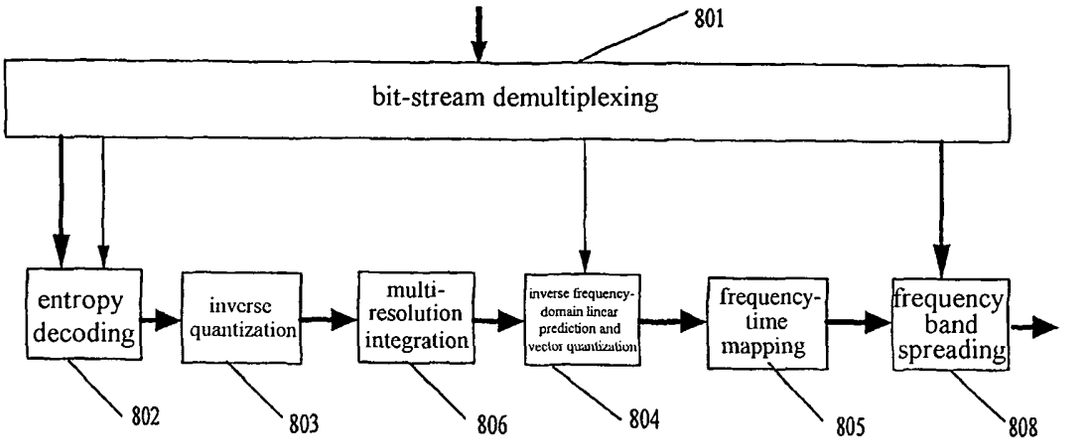
801
bit-stream demultiplexing
entropy decoding
inverse quantization
multi-resolution integration
inverse frequency-domain linear prediction and vector quantization
frequency-time mapping
frequency band spreading
802
803
806
804
805
808

Fig. 16

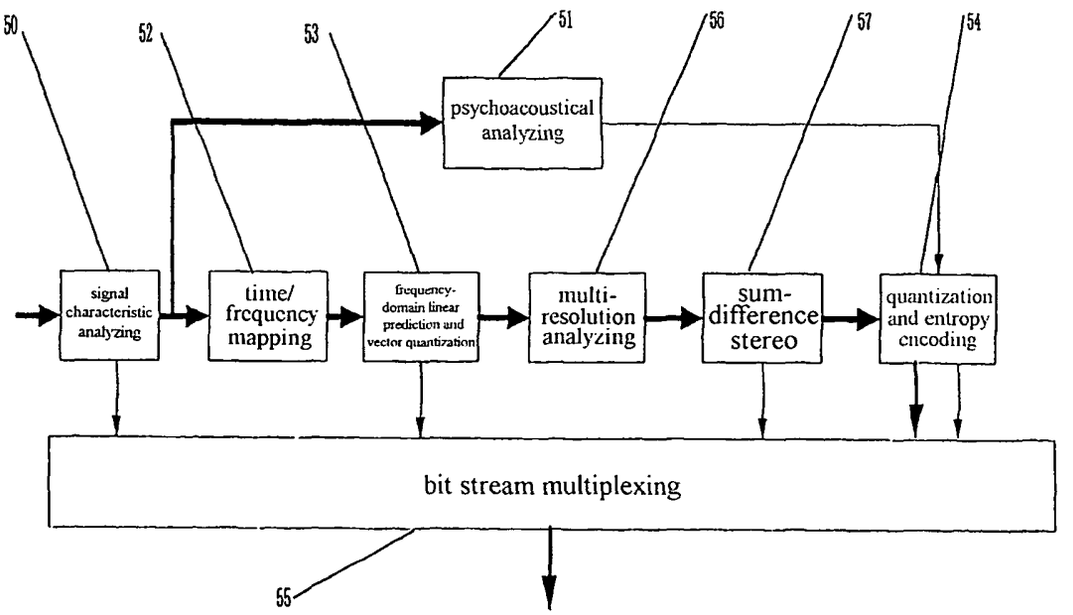
50
52
53
51
56
57
54
psychoacoustical analyzing
signal characteristic analyzing
time/ frequency mapping
frequency-domain linear prediction and vector quantization
multi-resolution analyzing
sum-difference stereo
quantization and entropy encoding
bit stream multiplexing
55

Fig. 17

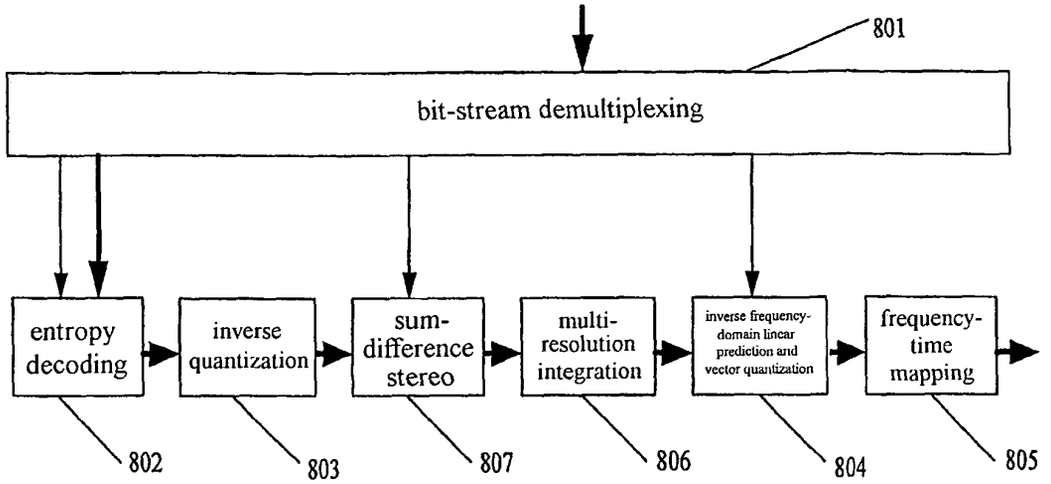
801
bit-stream demultiplexing
entropy decoding
inverse quantization
sum-difference stereo
multi-resolution integration
inverse frequency-domain linear prediction and vector quantization
frequency-time mapping
802
803
807
806
804
805

Fig. 18

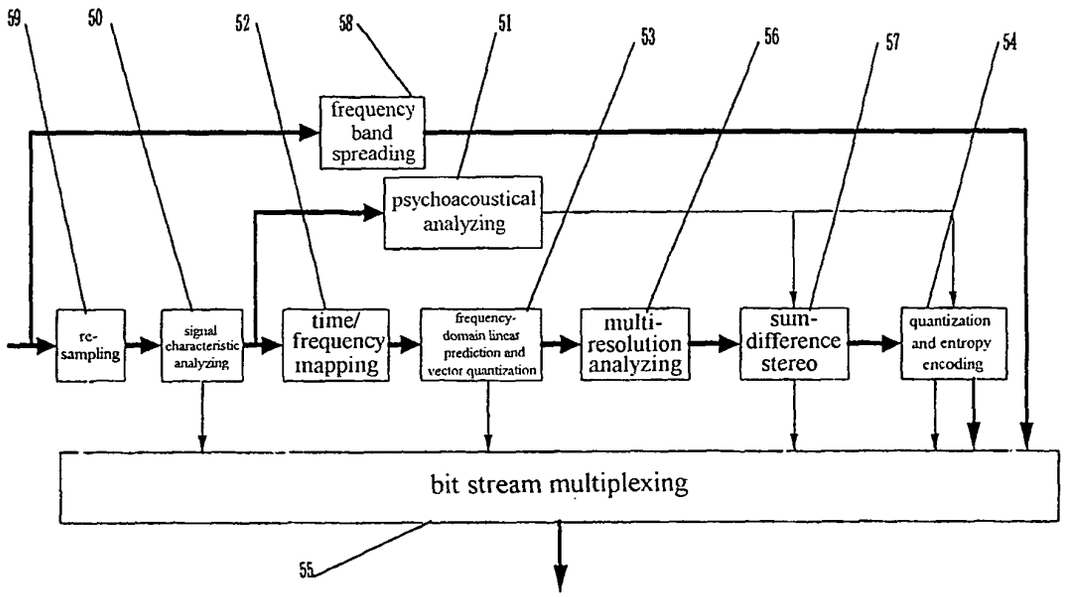
59
50
52
58
51
53
56
57
54
frequency band spreading
psychoacoustical analyzing
re-sampling
signal characteristic analyzing
time/frequency mapping
frequency-domain linear prediction and vector quantization
multi-resolution analyzing
sum-difference stereo
quantization and entropy encoding
bit stream multiplexing
55

Fig. 19

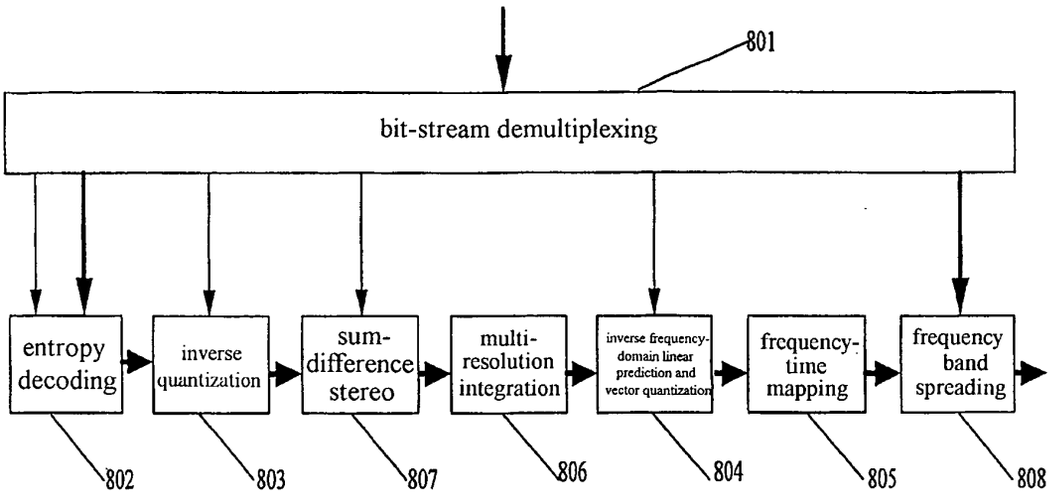
801
bit-stream demultiplexing
entropy decoding
inverse quantization
sum-difference stereo
multi-resolution integration
inverse frequency-domain linear prediction and vector quantization
frequency-time mapping
frequency band spreading
802
803
807
806
804
805
808

Fig. 20

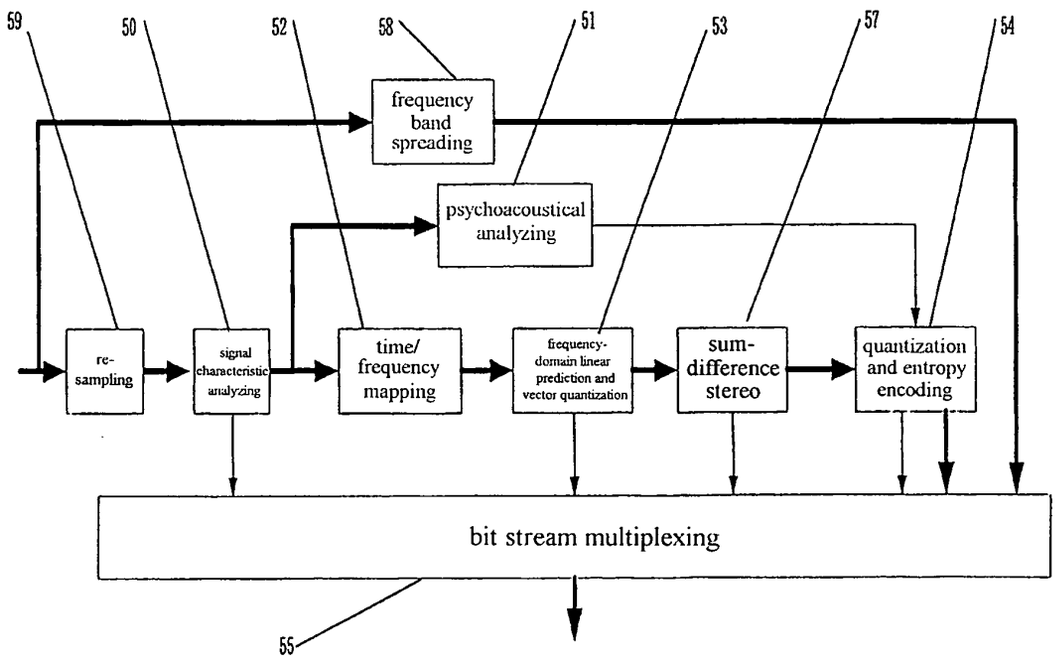
59
50
52
58
51
53
57
54
frequency band spreading
psychoacoustical analyzing
re-sampling
signal characteristic analyzing
time/ frequency mapping
frequency-domain linear prediction and vector quantization
sum-difference stereo
quantization and entropy encoding
bit stream multiplexing
55

Fig. 21

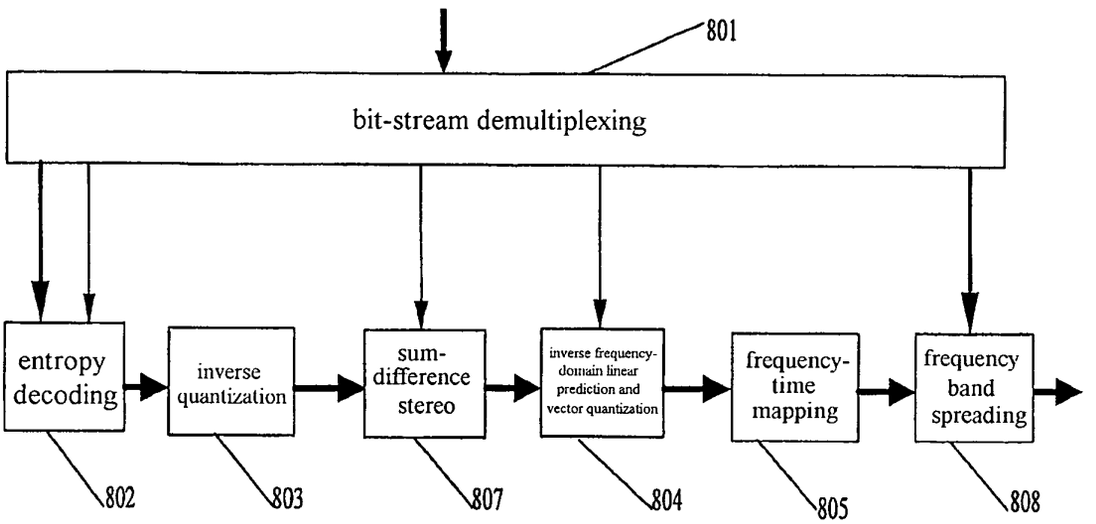
801
bit-stream demultiplexing
entropy decoding
inverse quantization
sum-difference stereo
inverse frequency-domain linear prediction and vector quantization
frequency-time mapping
frequency band spreading
802
803
807
804
805
808

Fig. 22

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2005/000441

A. CLASSIFICATION OF SUBJECT MATTER

IPC[7]: G10L19/04

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC[7]: G10L19/00 G10L19/04 G10L19/06 G10L19/08

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Chinese Patent Document （1985 ～ ）

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT CNKI： 音频 语音 声音 编码 解码 量化 熵 时域 频域 掩蔽 阀值 线性预测

WPI EPODOC PAJ : AUDIO SPEECH SOUND CODE+ DECODE+ ENCODE+ QUANTIZE+ ENTROPY TIME FREQUENCY DOMAIN MASK+ THRESHOLD LINEAR PREDICTION

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US, A, 5613035 (Daewoo Electronics Co. Ltd.,) 18. Mar.1997(18.03.1997) The whole document	1－15
A	CN, A, 1461112 （BEIJING E-WORLD TECHNOLOGY CO.,LTD） 10.Dec.2003（10.12.2003） The whole document	1－15
A	CN, A, 1388517 （BEIJING E-WORLD TECHNOLOGY CO.,LTD） 01.Jan.2003(01.01.2003) The whole document	1－15
A	US, A, 5537510 (Daewoo Electronics Co. Ltd.,) 16.July.1996 （16.07.1996） The whole document	1－15

☐ Further documents are listed in the continuation of Box C. ☒ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim (S) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&"document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
10.Jun.2005（10.06.2005）	30 · JUN 2005 (30 · 06 · 2005)
Name and mailing address of the ISA/CN The State Intellectual Property Office, the P.R.China 6 Xitucheng Rd., Jimen Bridge, Haidian District, Beijing, China 100088 Facsimile No. 86-10-62019451	Authorized officer YANG, San Telephone No. 86-10-62085725

Form PCT/ISA /210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2005/000441

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
US, A, 5613035	18.Mar.1997	KR, B, 9612475	20. Sep.1996
		CN, A, 1119376	27.Mar.1996
		JP, A, 8051366	20.Feb.1996
		EP, A, 0663740	19.Jul.1995
CN, A, 1461112	10.Dec.2003	NONE	
CN, A, 1388517	01.Jan.2003	NONE	
US, A, 5537510	16.July.1996	DE, T, 69422051	13.Jul.2000
		DE, D, 69422051	13.Jan.2000
		EP, A, 0720316	13.Jul.1996

Form PCT/ISA /210 (patent family annex) (April 2005)