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

(11) **EP 4 550 318 A1**

(12) **EUROPEAN PATENT APPLICATION**
published in accordance with Art. 153(4) EPC

(43) Date of publication:
07.05.2025 Bulletin 2025/19

(21) Application number: **22948636.0**

(22) Date of filing: **30.06.2022**

(51) International Patent Classification (IPC):
G10L 19/008 *(2013.01)* ***H04S 3/00*** *(2006.01)*
G10L 19/16 *(2013.01)* *G10L 19/24* *(2013.01)*

(52) Cooperative Patent Classification (CPC):
G10L 19/008; G10L 19/167; G10L 19/24

(86) International application number:
PCT/CN2022/103170

(87) International publication number:
WO 2024/000534 (04.01.2024 Gazette 2024/01)

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

(71) Applicant: **Beijing Xiaomi Mobile Software Co., Ltd.**
Beijing 100085 (CN)

(72) Inventor: **GAO, Shuo**
Beijing 100085 (CN)

(74) Representative: **Dehns Germany Partnerschaft mbB**
Theresienstraße 6-8
80333 München (DE)

(54) **AUDIO SIGNAL ENCODING METHOD AND APPARATUS, AND ELECTRONIC DEVICE AND STORAGE MEDIUM**

(57) Disclosed in the embodiments of the present disclosure are an audio signal encoding method and apparatus, and an electronic device and a storage medium. The method comprises: acquiring a scenario-based audio signal; determining the number of channels and an encoding rate of the audio signal; and performing encoding processing on the audio signal according to the number of channels and the encoding rate, so as to generate an encoded code stream. Therefore, by means of performing encoding processing on an audio signal according to the number of channels and an encoding rate, the number of bits that can be used can be fully utilized during an encoding process, the waste of the number of bits is avoided, and an audio service matching the encoding rate is provided for a remote user.

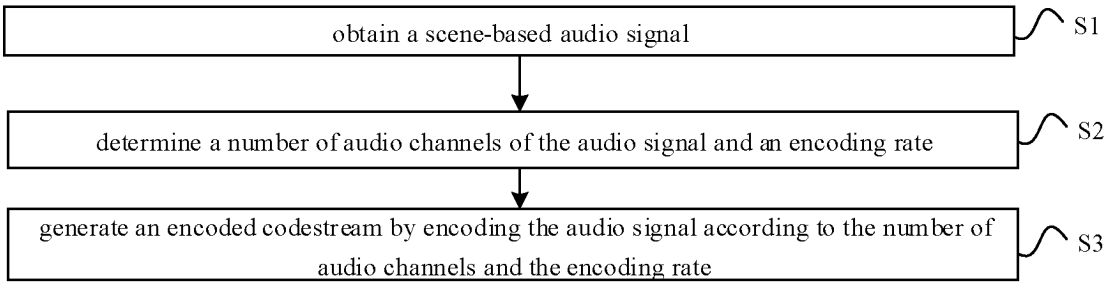

FIG. 1

EP 4 550 318 A1

Description

TECHNICAL FIELD

[0001] The disclosure relates to a field of communication technologies, in particular to an audio signal encoding method, an audio signal encoding apparatus, an electronic device and a storage medium.

BACKGROUND

[0002] In the related art, after acquiring an audio signal, the audio signal is subjected to a uniform encoding process. In the uniform encoding process, without considering the different encoding rates, a number of bits available for each audio channel is different, which causes the number of bits available for each audio channel to exceed or be less than a number of bits necessary for encoding, resulting in the waste of bits or the inability to provide an audio service that matches the encoding rate for remote users, which is an urgent problem to be solved.

SUMMARY

[0003] Embodiments of the disclosure provide an audio signal encoding method, an audio signal encoding apparatus, an electronic device and a storage medium, to encode the audio signal according to a number of audio channels and an encoding rate, which may make full use of bits available during the encoding process, avoid a waste of bits, and provide audio services that match the encoding rate for remote users.

[0004] According to a first aspect of embodiments of the disclosure, an audio signal encoding method is provided. The method includes: obtaining a scene-based audio signal; determining a number of audio channels of the audio signal and an encoding rate; and generating an encoded codestream by encoding the audio signal according to the number of audio channels and the encoding rate.

[0005] In this technical solution, a scene-based audio signal is obtained, and a number of audio channels of the audio signal and an encoding rate are determined, and then an encoded codestream is generated by encoding the audio signal according to the number of audio channels and the encoding rate. The audio signal is thus encoded according to the number of audio channels and the encoding rate, and during the encoding process, the bits available may be fully utilized, avoiding a waste of bits, and providing audio services that match the encoding rate for remote users.

[0006] In some embodiments, generating the encoded codestream by encoding the audio signal according to the number of audio channels and the encoding rate, includes: performing a down-mixed processing on the audio signal according to the number of audio channels and the encoding rate to generate a down-mixed parameter and a down-mixed audio channel signal; encoding the down-mixed audio channel signal to generate an encoding parameter; and generating the encoded codestream by performing codestream multiplexing on the down-mixed parameter and the encoding parameter.

[0007] In some embodiments, performing the down-mixed processing on the audio signal according to the number of audio channels and the encoding rate to generate the down-mixed parameter and the down-mixed audio channel signal, includes: determining a target control parameter for the audio signal according to the number of audio channels and the encoding rate; determining a down-mixed processing algorithm according to the target control parameter; and performing the down-mixed processing on the audio signal according to the down-mixed processing algorithm to generate the down-mixed parameter and the down-mixed audio channel signal.

[0008] In some embodiments, determining the target control parameter for the audio signal according to the number of audio channels and the encoding rate, includes: calculating an initial average rate of each channel according to the number of audio channels and the encoding rate; determining a target average rate according to the initial average rate and a preset average rate threshold; and determining the target control parameter for the audio signal according to the initial average rate and the target average rate.

[0009] In some embodiments, before encoding the audio signal, the method further includes: performing a pre-emphasis preprocessing and/or a high-pass filtering preprocessing on the audio signal.

[0010] According to a second aspect of embodiments of the disclosure, an audio signal encoding apparatus is provided. The apparatus includes: a signal obtaining unit, configured to obtain a scene-based audio signal; an information determining unit, configured to determine a number of audio channels of the audio signal and an encoding rate; and an encoding processing unit, configured to generate an encoded codestream by encoding the audio signal according to the number of audio channels and the encoding rate.

[0011] In some embodiments, the encoding processing unit includes: a down-mixed processing module, configured to perform a down-mixed processing on the audio signal according to the number of audio channels and the encoding rate to generate a down-mixed parameter and a down-mixed audio channel signal; a parameter generating module, configured to encode the down-mixed audio channel signal to generate an encoding parameter; and a codestream generating module,

configured to generate the encoded codestream by performing codestream multiplexing on the down-mixed parameter and the encoding parameter.

[0012] In some embodiments, the down-mixed processing module includes: a parameter determining sub-module, configured to determine a target control parameter for the audio signal according to the number of audio channels and the encoding rate; an algorithm determining sub-module, configured to determine a down-mixed processing algorithm according to the target control parameter; and a down-mixed processing sub-module, configured to perform the down-mixed processing on the audio signal according to the down-mixed processing algorithm to generate the down-mixed parameter and the down-mixed audio channel signal.

[0013] In some embodiments, the parameter determining sub-module is further configured to: calculate an initial average rate of each channel according to the number of audio channels and the encoding rate; determine a target average rate according to the initial average rate and a preset average rate threshold; and determine the target control parameter for the audio signal according to the initial average rate and the target average rate.

[0014] In some embodiments, the apparatus further includes: a preprocessing unit, configured to perform a pre-emphasis preprocessing and/or a high-pass filtering preprocessing on the audio signal.

[0015] According to a third aspect of embodiments of the disclosure, an electronic device is provided. The electronic device includes at least one processor, and a memory communicatively connected to the at least one processor. The memory stores instructions executable by the at least one processor, and the instructions are executed by the at least one processor to cause the at least one processor to execute the method described in the first aspect.

[0016] According to a fourth aspect of embodiments of the disclosure, a non-transitory computer-readable storage medium having computer instructions stored thereon is provided. The computer instructions are configured to cause a computer to execute the method described in the first aspect.

[0017] According to a fifth aspect of embodiments of the disclosure, a computer program product including computer instructions is provided. When the computer instructions are executed by a processor, the method described in the first aspect is implemented.

[0018] It is understood that both the foregoing general description and following detailed description are exemplary and explanatory only and are not for limiting the disclosure.

BRIEF DESCRIPTION OF THE DRAWINGS

[0019] In order to clearly illustrate technical solutions of embodiments of the disclosure or background technologies, a description of drawings used in the embodiments or the background technologies is given below.

FIG. 1 is a flowchart of an audio signal encoding method according to an embodiment of the disclosure.

FIG. 2 is a schematic diagram of a coordinate of an audio signal in a Firs-Order Ambisonics (FOA) format according to an embodiment of the disclosure.

FIG. 3 is a flowchart of another audio signal encoding method according to an embodiment of the disclosure.

FIG. 4 is a flowchart of an audio signal encoding method in the related art according to an embodiment of the disclosure.

FIG. 5 is a flowchart of yet another audio signal encoding method according to an embodiment of the disclosure.

FIG. 6 is a flowchart of substeps of step S30 in the audio signal encoding method according to an embodiment of the disclosure.

FIG. 7 is a flowchart of substeps of step S301 in the audio signal encoding method according to an embodiment of the disclosure.

FIG. 8 is a structural diagram of an audio signal encoding apparatus according to an embodiment of the disclosure.

FIG. 9 is a structural diagram of an encoding processing unit in an audio signal encoding apparatus according to an embodiment of the disclosure.

FIG. 10 is a structural diagram of a down-mixed processing module in an audio signal encoding apparatus according to an embodiment of the disclosure.

FIG. 11 is a structural diagram of another audio signal encoding apparatus according to an embodiment of the disclosure.

FIG. 12 is a structural diagram of an electronic device according to an embodiment of the disclosure.

DETAILED DESCRIPTION

[0020] In order to make those skilled in the field better understand the technical solutions of this disclosure, the technical solutions in the embodiment of this disclosure will be described clearly and completely in combination with the attached drawings.

[0021] Unless the context indicates otherwise, throughout the specification and the claims, the term "comprise" is

interpreted as open, inclusive, i.e., "includes, but is not limited to". In the description of the specification, "some embodiments" is intended to indicate that certain features, structures, materials or characteristics related to the embodiments or examples are included in at least one embodiment or example of the disclosure. The schematic representation of the above term does not necessarily refer to the same embodiment or example. Furthermore, the features, structures, materials, or characteristics described above may be included in any one or more embodiments or examples in any suitable manner.

[0022] It should be noted that the terms "first" and "second" in the specification and the claims of this disclosure and the drawings are used to distinguish similar objects, and are not necessarily used to describe a specific order or sequence. The terms "first" and "second" are only used for descriptive purposes, and cannot be understood as indicating or implying relative importance or implicitly indicating the number of indicated technical features. Therefore, the features defined with the terms "first" and "second" may explicitly or implicitly include one or more of these features. It should be understood that data so used may be interchanged under appropriate circumstances, so that the embodiments of the disclosure described herein may be implemented in other orders than those illustrated or described herein. The implementations described in the following exemplary embodiments do not represent all implementations consistent with the disclosure. Rather, they are merely examples of devices and methods consistent with some aspects of the disclosure as detailed in the appended claims.

[0023] The term "at least one" in the disclosure may also be described as one or more, and the term "multiple" may be two, three, four, or more, which is not limited in the disclosure. In the embodiments of the disclosure, for a type of technical features, "first", "second", and "third", and "A", "B", "C" and "D" are used to distinguish different technical features of the type, the technical features described using the "first", "second", and "third", and "A", "B", "C" and "D" do not indicate any order of precedence or magnitude.

[0024] The correspondences shown in the tables in this disclosure may be configured or may be predefined. The values of information in the tables are merely examples and may be configured to other values, which are not limited by the disclosure. In configuring the correspondence between the information and the parameter, it is not necessarily required that all the correspondences illustrated in the tables must be configured. For example, the correspondences illustrated in certain rows in the tables in this disclosure may not be configured. For another example, the above tables may be adjusted appropriately, such as splitting, combining, and the like. The names of the parameters shown in the titles of the above tables may be other names that may be understood by the communication device, and the values or representations of the parameters may be other values or representations that may be understood by the communication device. Each of the above tables may also be implemented with other data structures, such as, arrays, queues, containers, stacks, linear tables, pointers, chained lists, trees, graphs, structures, classes, heaps, and Hash tables.

[0025] Those skilled in the art may realize that the units and algorithmic steps of the various examples described in combination with the embodiments disclosed herein are capable of being implemented in the form of electronic hardware, or a combination of computer software and electronic hardware. Whether these functions are performed in the form of hardware or software depends on the specific application and design constraints of the technical solution. Those skilled in the art may use different methods to implement the described functions for each particular application, but such implementations should not be considered as beyond the scope of the disclosure.

[0026] The first generation (1G) mobile communication technology is the first generation wireless cellular technology, which belongs to an analog mobile communication network. When 1G is upgraded to 2G, a mobile phone switches from an analog communication to a digital communication, and a global system for mobile communication (GSM) network standard is adopted. A voice encoder adopts an adaptive multi rate (AMR) narrow band speech codec, an enhanced full rate (EFR), a full rate (FR), and a half rate (HR), and a communication provides a single-channel narrowband voice service. A 3G mobile communication system was proposed by the International Telecommunication Union (ITU) for international mobile communications in 2000, which can adopt Time Division-Synchronous Code Division Multiple Access (TD-SCDMA), Code Division Multiple Access 2000 (CDMA2000), or Wideband Code Division Multiple Access (WCDMA), and the voice encoder of which adopts an adaptive multi-rate wideband (AMR-WB) to provide a single-channel broadband voice service. A 4G is improved base on the 3G technology. Both data and voice are transmitted in an all-IP manner, for providing a real-time high definition (HD)+Voice service of voice audio. An enhanced voice service (EVS) codec adopted by the 4G can balance a high-quality compression of voice and audio.

[0027] The voice and audio communication service provided above have expanded from a narrowband signal to an ultra-wideband service or even a full-band service, but they are all signal-audio channel services. With the increasing demand for high-quality audio, compared with the signal-audio channel audio, stereo audio has a sense of orientation and distribution for each sound source, and can improve a clarity.

[0028] With an increase of transmission bandwidth, an upgrade of a terminal device signal collection device, an improvement of performance of a signal processor, and an upgrade of a terminal playback device, three signal formats, namely audio channel-based multi-channel audio signals, object-based audio signals, and scene-based audio signals, can provide three-dimensional audio services. An immersive voice and audio service (IVAS) codec that is being standardized by the 3rd Generation Partnership Project (3GPP) SA4 can support encoding and decoding requirements

of the above three signal formats. Terminal devices that can support 3D audio services include a mobile phone, a computer, a Pad, a conference system device, an augmented reality/virtual reality (AR/VR) device, a vehicle, etc.

[0029] A Firs-Order Ambisonics/High-Order Ambisonics (FOA/HOA) signal is a main scene-based audio signal. The FOA/HOA signal represents audio information collected at a certain position in an audio scene and is an immersive audio format whose audio quality gradually gets better with the increase of order. Different Ambisonics orders represent different numbers of audio signal components. That is, for an N-order Ambisonics signal, the number of Ambisonics coefficients is (N+1)*(N+1).

Table 1: the relationship between the Ambionics signal order and the Ambionics coefficient

Ambisonics order	Ambisonics coefficient/number of audio channels
0	1
1	4
2	9
3	16
4	25
5	36
6	49

[0030] As shown in Table 1, the number of audio channels of Ambisonics increases rapidly with the increase of order. Correspondingly, an amount of encoded data also increases rapidly, as well as an encoding complexity. Meanwhile, due to the limitation of encoding rate, an encoding performance is greatly reduced. In order to reduce the encoding complexity, it is necessary to perform a down-mixed processing on an input initial audio channel. After the down-mixed process, the number of audio channels decreases, and the encoding complexity is reduced, so as to achieve a balance between the encoding complexity and the encoding performance.

[0031] In response to the problem of the waste of bits or the inability to provide the audio services that match the encoding rate for a remote user in the related art, the embodiment of the disclosure provides an audio signal encoding method and an audio signal encoding apparatus to solve the problems existing in the related art at least to some extent, so as to make full use of the available bits, provide the audio services that match the encoding rate for the remote user, and improve the user experience.

[0032] As illustrated in FIG. 1, FIG. 1 is a flowchart of an audio signal encoding method according to an embodiment of the disclosure.

[0033] As illustrated in FIG. 1, the method includes but is not limited to the following steps.

[0034] At step S1, a scene-based audio signal is obtained.

[0035] It is understood that when a local user establishes a voice communication with any remote user, the local user can establish a voice communication with a terminal device of the any remote user through a terminal device of the local user. The terminal device of the local user may obtain sound information of an environment where the local user is located in real time and obtain the scene-based audio signal.

[0036] The sound information of the environment where the local user is located includes sound information made by the local user and sound information of surrounding things. The sound information of surrounding things may be, for example, sound information of vehicle driving, sound information of birds, sound information of wind, and sound information of other users around the local user, and so on.

[0037] It should be noted that the terminal device is an entity on a user side for receiving or transmitting signals. For example, the terminal device may be a mobile phone, a computer, a Pad, a watch, an interphone, a conference system device, an augmented reality/virtual reality (AR/VR) device, a vehicle, etc. The terminal device may also be referred to as a user equipment (UE), a mobile station (MS), a mobile terminal (MT), and the like. The terminal device may be a vehicle with communication functions, a smart vehicle, a mobile phone, a wearable device, a Pad, a computer with wireless transceiver functions, a VR terminal device, an AR terminal device, a wireless terminal device in industrial control, a wireless terminal device in self-driving, a wireless terminal device in remote medical surgery, a wireless terminal device in smart grid, a wireless terminal device in transportation safety, a wireless terminal device in smart city, a wireless terminal device in smart home, etc. The specific technology and specific device form adopted by the terminal device are not limited in embodiments of the disclosure.

[0038] In the embodiment of the disclosure, when acquiring the scene-based audio signal, the terminal device of the local user can acquire the sound information of the environment where the local user is located via a recording apparatus, such as a microphone, arranged in the terminal device or cooperating with the terminal device, and then generate the

scene-based audio signal to obtain the scene-based audio signal.

[0039] In the embodiment of the disclosure, the scene-based audio signal may be an audio signal in a FOA format or an audio signal in a HOA format.

[0040] At step S2, a number of audio channels of the audio signal and an encoding rate are determined.

[0041] In the embodiment of the disclosure, after obtaining the scene-based audio signal, the number of audio channels of the audio signal and the encoding rate are determined.

[0042] For example, as illustrated in FIG. 2, in a case where the scene-based audio signal is an audio signal in the FOA format, it is determined that the number of audio channels of the audio signal is 4, which may be represented by W, X, Y and Z, in which W represents a component containing all sounds in all directions in a sound field superimposed with the same gain and phase, X represents a component in a front-back direction in the sound field, Y represents a component in a left-right direction in the sound field, and Z represents a component in an up-down direction in the sound field. It is further determined that the selected encoding rate is 96kbps.

[0043] At step S3, an encoded codestream is generated by encoding the audio signal according to the number of audio channels and the encoding rate.

[0044] In the embodiment of the disclosure, the scene-based audio signal is obtained, the number of audio channels of the audio signal and the encoding rate are determined, and the encoded codestream is generated by encoding the audio signal according to the number of audio channels and the encoding rate.

[0045] When encoding the audio signal according to the number of audio channels and the encoding rate, the encoding rate of each audio channel may be determined according to the number of audio channels and the encoding rate. For example, an average encoding rate of each audio channel, the maximum encoding rate of each audio channel, or the encoding rate of each audio channel may be determined. The average encoding rate of each audio channel may be determined by dividing the encoding rate by the number of audio channels, the maximum encoding rate of each audio channel is equal to the encoding rate, and the encoding rate of each audio channel is the encoding rate.

[0046] In a base of determining the encoding rate of each audio channel, the number of bits available for each audio channel may be considered at the different encoding rates according to the encoding rate of each audio channel, so that the bits available is able to be fully utilized during the encoding process, to avoid the waste of bits and provide the audio services matching the encoding rate for the remote user. The generated encoded codestream is able to provide clear, stable and understandable audio services when the encoding rate is low, and is able to provide high-definition, stable and immersive audio services when the encoding rate is high. In this way, it can provide the remote user with the audio services matching the encoding rate, thus improving the user experience.

[0047] In some embodiments, before encoding the audio signal, the method further includes: performing a pre-emphasis preprocessing and/or a high-pass filtering preprocessing on the audio signal.

[0048] In the embodiment of the disclosure, in a case where the scene-based audio signal is obtained and the number of audio channels of the audio signal and the encoding rate are determined, the pre-emphasis preprocessing may be performed on the audio signa, which may enhance a high-frequency portion of the audio information and increase a high-frequency resolution of the audio information.

[0049] In the embodiment of the disclosure, in a case where the scene-based audio signal is obtained and the number of audio channels of the audio signal and the encoding rate are determined, the high-pass filtering preprocessing may be performed on the audio signal, to filter signal components in the audio signal lower than a certain frequency threshold. A starting frequency in the high-pass filtering processing may be set as required, for example, the starting frequency may be set as 20Hz.

[0050] After performing the high-pass filtering preprocessing on the audio signal, an audio signal component of the required encoding frequency band may be obtained. When the audio signal is encoded, an influence of an ultra-low frequency signal on encoding processing effects may be avoided.

[0051] By implementing the embodiments of the disclosure, the scene-based audio signal is obtained, the number of audio channels of the audio signal and the encoding rate are determined, and the encoded codestream is generated by encoding the audio signal according to the number of audio channels and the encoding rate. In this way, the audio signal is encoded according to the number of audio channels and the encoding rate, and the bits available are able to be fully utilized during the encoding process, so that the waste of bits may be avoided, and the audio services that match the encoding rate may be provided for the remote user.

[0052] FIG. 3 is a flowchart of an audio signal encoding method according to an embodiment of the disclosure.

[0053] As illustrated in FIG. 3, the method includes but is not limited to the following steps.

[0054] At step S10, a scene-based audio signal is obtained.

[0055] At step S20, a number of audio channels of the audio signal and an encoding rate are determined.

[0056] In the embodiment of the disclosure, the related descriptions of steps S10 and S20 can be referred to the related descriptions in the above embodiments, and the same contents will not be repeated here.

[0057] At step S30, a down-mixed processing is performed on the audio signal according to the number of audio channels and the encoding rate, to generate a down-mixed parameter and a down-mixed audio channel signal.

[0058] At step S40, the down-mixed audio channel signal is encoded to generate an encoding parameter.

[0059] At step S50, the encoded codestream is generated by performing codestream multiplexing on the down-mixed parameter and the encoding parameter.

[0060] In the embodiment of the disclosure, the scene-based audio signal is obtained, the number of audio channels of the audio signal and the encoding rate are determined, and the encoded codestream is generated by encoding the audio signal according to the number of audio channels and the encoding rate. Encoding the audio signal according to the number of audio channels and the encoding rate may include performing the down-mixed processing on the audio signal according to the number of audio channels and the encoding rate to generate the down-mixed parameter and the down-mixed audio channel signal. The down-mixed audio channel signal is then encoded to generate the encoding parameter. The encoded codestream is generated by codestream multiplexing according to the down-mixed parameter and the encoding parameter.

[0061] As illustrated in FIG. 4, in the related art, after acquiring an audio signal (an audio signal in the FOA format or an audio signal in the HOA format), the audio signal is subjected to a uniform down-mixed process, and the number of audio channels after down-mixed is less than the initial number of audio channels. All the remaining channels are encoded by a core encoder, and down-mixed parameters generated by the down-mixed processing and output parameters of the core encoder are performed by codestream multiplexing to output the encoded codestream.

[0062] The uniform down-mixed processing of the audio signal does not consider that the number of bits available for each audio channel is different under different encoding rates, resulting in the number of audio channels after the down-mixed processing does not match the number of audio channels that the core encoder is able to encode. Therefore, when the number of audio channels after the down-mixed processing is much less than the number of input audio channels, better audio services cannot be provided to remote users at a high encoding rate (because the number of bits available for each audio channel exceeds the number of bits necessary for encoding, which may lead to the waste of bits). When the number of audio channels after the down-mixed processing is slightly different from the number of input audio channels, the remote users cannot be provided with audio services that match the encoding rate at a low encoding rate (because the number of bits available for each audio channel is much less than the number of bits necessary for encoding, which may lead to a poor encoding quality of each audio channel).

[0063] However, as illustrated in FIG. 5, in the embodiment of the disclosure, a scene-based audio signal (an audio signal in the FOA format or an audio signal in the HOA format) is input to an encoder end, the encoder end may determine the number of audio channels of the audio signal and the encoding rate and input the encoding rate, the number of audio channels and the audio signal to a pattern analysis module, or the encoder end may perform a high-pass filtering preprocessing on the audio signal and then input the preprocessed audio signal into the pattern analysis module.

[0064] The pattern analysis module may output a control parameter according to the selected encoding rate and the number of audio channels, and use the control parameter to guide a down-mixed processing module to select a corresponding down-mixed processing algorithm. The down-mixed processing module outputs a down-mixed parameter and a down-mixed audio channel signal after processing the audio signal. An encoding parameter is output after encoding the down-mixed audio channel signal by the core encoder. The encoding parameter and the down-mixed parameter are input to a codestream multiplexer to output an encoded codestream.

[0065] In the embodiment of the disclosure, when the input scene-based audio signal is the audio signal in the FOA format/the audio signal in the HOA format, a matching down-mixed processing algorithm is adaptively selected according to the number of audio channels of the input audio signal and the number of bits available, so that the number of audio channels after the down-mixed processing matches the number of audio channels that may be encoded by the core encoder at this encoding rate, and a full (optimal) utilization of bits available may be achieved. That is, at a low rate, it may ensure the provision of clear, stable and understandable audio services, and at a high rate, it may ensure the provision of high-definition, stable immersive audio services, which may improve the user experience.

[0066] In the embodiment of the disclosure, after the encoder outputs the encoded codestream, the encoded codestream may be sent to a decoder end for decoding, so that the remote terminals may obtain sound information transmitted by the local terminal.

[0067] As illustrated in FIG. 6, in some embodiments, step S30 of performing the down-mixed processing on the audio signal according to the number of audio channels and the encoding rate to generate the down-mixed parameter and the down-mixed audio channel signal, includes the following steps.

[0068] At step S301, a target control parameter for the audio signal is determined according to the number of audio channels and the encoding rate.

[0069] In the embodiment of the disclosure, when performing the down-mixed processing on the audio signal according to the number of audio channels and the encoding rate, the target control parameter for the audio signal may be determined according to the number of audio channels and the encoding rate.

[0070] When determining the target control parameter for the audio signal according to the number of audio channels and the encoding rate, the encoding rate of each audio channel may be determined according to the number of audio channels and the encoding rate. For example, an average encoding rate of each audio channel, the maximum encoding

rate of each audio channel, or the encoding rate of each audio channel may be determined. The average encoding rate of each audio channel is determined by dividing the encoding rate by the number of audio channels, the maximum encoding rate of each audio channel is equal to the encoding rate, and the encoding rate of each channel is the encoding rate.

[0071] In the embodiment of the disclosure, on the basis of determining the encoding rate of each audio channel according to the number of audio channels and the encoding rate, the target control parameter for the audio signal is determined according to the encoding rate of each audio channel.

[0072] Certainly, when determining the target control parameter for the audio signal according to the number of audio channels and the encoding rate, by pre-setting corresponding relationships between the number of audio channels and the encoding rate and the control parameter, in a case of determining the number of audio channels of the audio signal and the encoding rate, the target control parameter for the audio signal may be determined.

[0073] Alternatively, a target number of audio channels may be determined according to the number of audio channels and the encoding rate, and then the target control parameter for the audio signal may be determined according to the target number of audio channels.

[0074] The target number of audio channels is determined according to the number of audio channels and the encoding rate. For example, N thresholds of the average encoding rate are preset, where N is a positive integer, and N+1 threshold ranges are determined by the N thresholds. Different threshold ranges are set to correspond to different numbers of audio channels after the down-mixed process. On the basis, an initial average encoding rate is calculated according to the number of audio channels and the encoding rate, and the target number of audio channels may be determined according to the threshold range to which the initial average rate belongs, and then the target control parameter for the audio signal is determined according to the target number of audio channels.

[0075] It is understood that in a case where the encoding rate and the number of audio channels after the down-mixed processing are known, an average rate that is able to be allocated to each audio channel after the down-mixed processing may be obtained, and the target control parameter for the audio signal may be determined according to the target number of audio channels and/or the average rate that is able to be allocated to each audio channel after the down-mixed processing.

[0076] When determining the target control parameter of the audio signal according to the target number of audio channels and/or the average rate that is able to be allocated to each audio channel after the down-mixed processing, corresponding relationships between the target number of audio channels and/or the average rate that is able to be allocated to each audio channel after the down-mixed processing and the control parameter may be preset, and the target control parameter for the audio signal may be determined according to the target number of audio channels and/or the average rate that is able to be allocated to each audio channel after the down-mixed processing.

[0077] At step S302, a down-mixed processing algorithm is determined according to the target control parameter.

[0078] In the embodiment of the disclosure, in a case where the target control parameter for the audio signal is determined according to the number of audio channels and the encoding rate, the down-mixed processing algorithm may be determined according to the target control parameter. Determining the down-mixed processing algorithm may be determining the down-mixed processing algorithm corresponding to each audio channel, and the determined down-mixed processing algorithms for different channels may be the same or different.

[0079] At step S303, the down-mixed processing is performed on the audio signal according to the down-mixed processing algorithm, to generate the down-mixed parameter and the down-mixed audio channel signal.

[0080] In the embodiment of the disclosure, in a case where the down-mixed processing algorithm corresponding to each audio channel is determined, the audio signal may be performed by the down-mixed processing according to the down-mixed processing algorithm to generate the down-mixed parameter and the down-mixed audio channel signal.

[0081] As illustrated in FIG. 7, in some embodiments, step S301 of determining the target control parameter for the audio signal according to the number of audio channels and the encoding rate, includes the following steps.

[0082] At step S3011, an initial average rate of each audio channel is calculated according to the number of audio channels and the encoding rate.

[0083] At step S3012, a target average rate is determined according to the initial average rate and a preset average rate threshold.

[0084] At step S3013, the target control parameter for the audio signal is determined according to the initial average rate and the target average rate.

[0085] According to the number of audio channels and the encoding rate, the initial average rate of each audio channel may be calculated by dividing the encoding rate by the number of audio channels. For example, if the number of audio channels is 4 and the encoding rate is 96kbps, the initial average rate of each audio channel is calculated to be 24kbps according to the number of audio channels and the encoding rate.

[0086] In the embodiment of the disclosure, in a case where the initial average rate of each audio channel is calculated, the target average rate may be determined according to the initial average rate and the preset average rate threshold.

[0087] The preset average rate threshold may be set according to the scene-based audio signal. For example, a first average rate threshold Thres1 is set to 13.2kbps, and a second average rate threshold Thres2 is set to 32kbps. According

to the above two average rate thresholds, ranges corresponding to the average rate is divided into three average rate ranges, as follows,

an average rate range 1: less than or equal to 13.2kbps;
an average rate range 2: greater than 13.2 kbps and less than 32 kbps; and
an average rate range 3: greater than or equal to 32 kbps.

[0088] In the embodiment of the disclosure, the target average rate is determined according to the initial average rate and the preset average rate threshold. If the average rate threshold range is determined according to the average rate threshold, the corresponding number of output audio channels is set for each average rate threshold range, so that the corresponding target number of output audio channels may be determined according to the average rate threshold range to which the initial average rate belongs.

[0089] On the basis, in a case where the target number of output audio channels is determined, the target average rate may be calculated according to the target number of output audio channels and the encoding rate.

[0090] For example, the number of output audio channels corresponding to the average rate range 1 is 2, the number of output audio channels corresponding to the average rate range 2 is 3, and the number of output audio channels corresponding to the average rate range 3 is 4. If the initial average rate is 24kbps and belongs to the average rate range 2, it is determined that the target number of output audio channels is 3, and the target average rate may be calculated to be 96kbps/3=32kbps. It may be seen that the target average rate in the average rate range 2 is increasing compared to the initial average rate, so that the appropriate target control parameter may be determined when determining the target control parameter for the audio signal in subsequent processes, and the down-mixed processing algorithm may be determined according to the target control parameter. Therefore, the number of output audio channels after the down-mixed processing matches the number of audio channels that may be encoded by the core encoder at this encoding rate, and the optimal use of available bits may be achieved. That is, at a low rate, it may ensure the provision of clear, stable and understandable audio services, and at a high rate, it may ensure the provision of high-definition, stable immersive audio services, which may improve the user experience.

[0091] In the embodiment of the disclosure, for three average rate ranges, three different types of down-mixed processing algorithms may be selected for scene-based audio signals. After the selected down-mixed processing, an average rate available for each audio channel in the average rate range 1 and the average rate range 2 are increasing after the down-mixed processing. The average rate range 3 chooses not to perform the down-mixed processing because the encoding rate is rich enough, that is, an input signal is directly used as an output signal of the down-mixed processing, which means that the average rate available for each audio channel after the down-mixed processing remains unchanged.

[0092] For example, Table 2 shows some kinds of scene-based audio signals, initial average rates (average rates that may be allocated to each audio channel initially), preset average rate thresholds, as well as corresponding numbers of output audio channels (numbers of audio channels after the down-mixed processing) and determined target average rates (average rates that may be allocated to each audio channel after the down-mixed processing).

[0093] As can be seen from Table 2 below, the average rate that may be allocated to each audio channel after the down-mixed processing is greater than or equal to an average number of bits available for each audio channel, which may make full use of the available bits, avoid the waste of bits, and provide audio services that match the encoding rate for the remote users.

Table 2

Scene-based audio signal	Number of audio channels	Coding rate (kbps)	Initial average rate (kbps) that may be allocated to each audio channel	Number of audio channels after the down-mixed processing	Average rate (kbps) that may be allocated to each audio channel after down-mixed processing
FOA	4	less than or equal to 52.8	less than or equal to 13.2	2	less than or equal to 26.4
		greater than 52.8 and less than 128	greater than 13.2 and less than 32	3	greater than 52.8/3 and less than 128/3
		greater than or equal to 128	greater than or equal to 32	4	greater than or equal to 32
HOA2	9	less than or equal to 118.8	less than or equal to 13.2	5	less than or equal to 118.8/5
		greater than 118.8 and less than 288	greater than 13.2 and less than 32	7	greater than 118.8/7 and less than 288/7
		greater than or equal to 288	greater than or equal to 32	9	greater than or equal to 32
HOA3	16	less than or equal to 211.2	less than or equal to 13.2	8	less than or equal to 26.4
		greater than 211.2 and less than 512	greater than 13.2 and less than 32	12	greater than 211.2/12 and less than 512/12
		greater than or equal to 512	greater than or equal to 32	16	greater than or equal to 32
HOA4	25	less than or equal to 330	less than or equal to 13.2	14	less than or equal to 330/14
		greater than 330 and less than 800	greater than 13.2 and less than 32	20	greater than 330/16 and less than 800/20
		greater than or equal to 800	greater than or equal to 32	25	greater than or equal to 32

[0094] It is understood that each element in Table 2 exists independently, and these elements are listed in the same table by way of example, but it does not mean that all the elements in the table must exist simultaneously as shown in the table. The value of each element is independent of any other element value in Table 2. Therefore, it is understood by those skilled in the art that the value of each element in Table 2 is an independent embodiment.

[0095] In the embodiment of the disclosure, a target average rate is determined according to an initial average rate and a preset average rate threshold. In addition to the above-mentioned exemplary method, an average rate threshold closest to the initial average rate may be determined as the target average rate, or the initial average rate may be directly determined as the target average rate, or an average rate threshold, among average rate thresholds greater than the initial average rate, closest to the initial average rate may be determined as the target average rate, which is not specifically limited in the embodiment of the disclosure.

[0096] In the embodiment of the disclosure, after the target average rate is determined, when determining a target control parameter for an audio signal according to the initial average rate and the target average rate, corresponding relationships between the initial average rate and the target average rate and the control parameter may be preset. For example, corresponding relationships between the initial average rate and the target average rate and the control parameter are set, or corresponding relationships between the control parameter and a difference between the initial average rate and the target average rate are set, or corresponding relationships between the control parameter and an absolute value of a difference between the initial average rate and the target average rate are set, or corresponding relationships between the control parameter and a sum of the initial average rate and the target average rate, etc., which is not specifically limited in the embodiment of the disclosure.

[0097] A down-mixed processing algorithm is to design a down-mixed conversion matrix according to a target number of output audio channels and a number of audio channels for acquiring scene-based audio signals. For example, if the number of audio channels is N and the target number of output audio channels is M, the conversion matrix is M*N, and both N and M are positive integers, and M is less than or equal to N.

[0098] The conversion matrix M*N satisfies the following equation:

$$[M*1]=[M*N]*[N*1]$$

where [M*1] represents a matrix of M times 1, [M*N] represents a matrix of M times N, and [N* 1] represents a matrix of N times 1.

[0099] For convenience of understanding, the embodiment of the disclosure provides an exemplary embodiment.

[0100] In an exemplary embodiment, the scene-based audio signal obtained is an audio signal in a FOA format, the number of audio channels is 4, namely, W, X, Y, Z, and the selected encoding rate is 96kbps. After the down-mixed processing, the target number of output audio channels is 3, where W represents a component containing all sounds in all directions in a sound field superimposed with the same gain and phase, X represents a component in a front-back direction in the sound field, Y represents a component in a left-right direction in the sound field, and Z represents a component in an up-down direction in the sound field. The schematic diagram of the coordinate is shown in FIG. 2.

[0101] If the target number of audio channels is 3 after the down-mixed processing, the component Z in the up-down direction is omitted, and only three channel components, W, X and Y are reserved. This strategy is made in consideration of two aspects. Firstly, when reconstructing the sound field, the listener at a playback end is sensitive to the component in the front-back direction and the component in the left-right direction, but less sensitive to the component in the up-down direction. Secondly, there are fewer sound sources for the component in the up-down direction in the sound field of a general audio scene. After the down-mixed processing, the number of audio channels is 3, and the average encoding rate that may be allocated to each audio channel is 96kbps/3=32kbps. An encoding core may encode and reconstruct high-quality audio signals at this average encoding rate, and provides high-definition, stable and immersive audio services to remote users.

[0102] FIG. 8 is a structural diagram of an audio signal encoding apparatus provided by an embodiment of the disclosure.

[0103] As illustrated in FIG. 8, the audio signal encoding apparatus 1 includes: a signal obtaining unit 11, an information determining unit 12, and an encoding processing unit 13.

[0104] The signal obtaining unit 11 is configured to obtain a scene-based audio signal.

[0105] The information determining unit 12 is configured to determine a number of audio channels of the audio signal and an encoding rate.

[0106] The encoding processing unit 13 is configured to generate an encoded codestream by encoding the audio signal according to the number of audio channels and the encoding rate.

[0107] By implementing the embodiment of the disclosure, the signal obtaining unit 11 obtains a scene-based audio signal, the information determining unit 12 determines a number of audio channels of the audio signal and an encoding

rate, and the encoding processing unit 13 generates a coded codestream by encoding the audio signal according to the number of audio channels and the encoding rate. In this way, the audio signal is encoded according to the number of audio channels and the encoding rate, and the bits available may be fully utilized during the encoding process, so that the waste of bits may be avoided, and audio services that match the encoding rate may be provided for remote users.

[0108] As illustrated in FIG. 9, in some embodiments, the encoding processing unit 13 includes: a down-mixed processing module 131, a parameter generating module 132, and a codestream generating module 133.

[0109] The down-mixed processing module 131 is configured to perform a down-mixed processing on the audio signal according to the number of audio channels and the encoding rate to generate a down-mixed parameter and a down-mixed audio channel signal.

[0110] The parameter generating module 132 is configured to encode the down-mixed audio channel signal to generate an encoding parameter.

[0111] The codestream generating module 133 is configured to generate the encoded codestream by performing codestream multiplexing on the down-mixed parameter and the encoding parameter.

[0112] As illustrated in FIG. 10, in some embodiments, the down-mixed processing module 131 includes: a parameter determining sub-module 1311, an algorithm determining sub-module 1312, and a down-mixed processing sub-module 1313.

[0113] The parameter determining sub-module 1311 is configured to determine a target control parameter for the audio signal according to the number of audio channels and the encoding rate.

[0114] The algorithm determining sub-module 1312 is configured to determine a down-mixed processing algorithm according to the target control parameter.

[0115] The down-mixed processing sub-module 1313 is configured to perform the down-mixed processing on the audio signal according to the down-mixed processing algorithm to generate the down-mixed parameter and the down-mixed audio channel signal.

[0116] In some embodiments, the parameter determining sub-module 1311 is further configured to:

calculate an initial average rate of each channel according to the number of audio channels and the encoding rate;
determine a target average rate according to the initial average rate and a preset average rate threshold; and
determine the target control parameter for the audio signal according to the initial average rate and the target average rate.

[0117] As illustrated in FIG. 11, in some embodiments, the audio signal encoding apparatus 1 further includes: a preprocessing unit 14.

[0118] The preprocessing unit 14 is configured to perform a pre-emphasis preprocessing and/or a high-pass filtering preprocessing on the audio signal.

[0119] With regard to the apparatus in the above embodiments, the specific way in which each module performs operations has been described in detail in the embodiments of the method, and will not be described in detail here.

[0120] The audio signal encoding apparatus according to the embodiment of the disclosure may execute the audio signal encoding methods as described in some of the above embodiments, and its beneficial effects are the same as those of the above audio signal encoding methods, which are not repeated here.

[0121] FIG. 12 is a structural diagram of an electronic device 100 for performing an audio signal encoding method illustrated by an exemplary embodiment.

[0122] For example, the electronic device 100 may be a mobile phone, a computer, a digital broadcasting terminal, a message transceiver device, a game console, a tablet device, a medical device, a fitness device or a personal digital assistant.

[0123] As illustrated in FIG. 12, the electronic device 100 may include one or more of the following components: a processing component 101, a memory 102, a power component 103, a multimedia component 104, an audio component 105, an input/output (I/O) interface 106, a sensor component 107, and a communication component 108.

[0124] The processing component 101 typically controls overall operations of the electronic device 100, such as the operations associated with display, telephone calls, data communications, camera operations, and recording operations. The processing component 101 may include one or more processors 1011 to perform all or part of the steps in the above described methods. Moreover, the processing component 101 may include one or more modules which facilitate the interaction between the processing component 101 and other components. For example, the processing component 101 may include a multimedia module to facilitate the interaction between the multimedia component 104 and the processing component 101.

[0125] The memory 102 is configured to store various types of data to support the operation of the electronic device 100. Examples of such data include instructions for any applications or methods operated on the electronic device 100, contact data, phonebook data, messages, pictures, video, etc. The memory 102 may be implemented using any type of volatile or non-volatile memory devices, or a combination thereof, such as a Static Random-Access Memory (SRAM), an Electrically-

Erasable Programmable Read Only Memory (EEPROM), an Erasable Programmable Read Only Memory (EPROM), a Programmable Read Only Memory (PROM), a Read Only Memory (ROM), a magnetic memory, a flash memory, a magnetic or optical disk.

[0126] The power component 103 provides power to various components of the electronic device 100. The power component 103 may include a power management system, one or more power sources, and any other components associated with the generation, management, and distribution of power in the electronic device 100.

[0127] The multimedia component 104 includes a touch screen providing an output interface between the electronic device 100 and the user. In some embodiments, the screen may include a Liquid Crystal Display (LCD) and a Touch Panel (TP). The touch panel includes one or more touch sensors to sense touches, swipes, and gestures on the touch panel. The touch sensor may not only sense a boundary of a touch or swipe action, but also sense a period of wakeup time and a pressure associated with the touch or swipe action. In some embodiments, the multimedia component 104 includes a front-facing camera and/or a rear-facing camera. When the electronic device 100 is in an operating mode, such as a shooting mode or a video mode, the front-facing camera and/or the rear-facing camera can receive external multimedia data. Each front-facing camera and rear-facing camera may be a fixed optical lens system or has focal length and optical zoom capability.

[0128] The audio component 105 is configured to output and/or input audio signals. For example, the audio component 105 includes a microphone (MIC) configured to receive an external audio signal when the electronic device 100 is in an operation mode, such as a call mode, a recording mode, and a voice recognition mode. The received audio signal may be further stored in the memory 102 or transmitted via the communication component 108. In some embodiments, the audio component 105 further includes a speaker to output audio signals.

[0129] The I/O interface 106 provides an interface between the processing component 101 and peripheral interface modules, such as a keyboard, a click wheel, buttons, and the like. The buttons may include, but are not limited to, a home button, a volume button, a starting button, and a locking button.

[0130] The sensor component 107 includes one or more sensors to provide status assessments of various aspects of the electronic device 100. For instance, the sensor component 107 may detect an open/closed status of the electronic device 100, relative positioning of components, e.g., the display and the keypad, of the electronic device 100, a change in position of the electronic device 100 or a component of the electronic device 100, a presence or absence of user contact with the electronic device 100, an orientation or an acceleration/deceleration of the electronic device 100, and a change in temperature of the electronic device 100. The sensor component 107 may include a proximity sensor configured to detect the presence of nearby objects without any physical contact. The sensor component 107 may also include a light sensor, such as a Complementary Metal Oxide Semiconductor (CMOS) or Charge-Coupled Device (CCD) image sensor, for use in imaging applications. In some embodiments, the sensor component 107 may also include an accelerometer sensor, a gyroscope sensor, a magnetic sensor, a pressure sensor, or a temperature sensor.

[0131] The communication component 108 is configured to facilitate communication, wired or wirelessly, between the electronic device 100 and other devices. The electronic device 100 can access a wireless network based on a communication standard, such as Wi-Fi, 2G or 3G, or a combination thereof. In an exemplary embodiment, the communication component 108 receives a broadcast signal or broadcast associated information from an external broadcast management system via a broadcast channel. In an exemplary embodiment, the communication component 108 further includes a Near Field Communication (NFC) module to facilitate short-range communication. For example, the NFC module may be implemented based on a Radio Frequency Identification (RFID) technology, an Infrared Data Association (IrDA) technology, an Ultra-Wide Band (UWB) technology, a Blue Tooth (BT) technology, and other technologies.

[0132] In some exemplary embodiments, the electronic device 100 may be implemented with one or more Application Specific Integrated Circuits (ASICs), Digital Signal Processors (DSPs), Digital Signal Processing Devices (DSPDs), Programmable Logic Devices (PLDs), Field Programmable Gate Arrays (FPGAs), controllers, micro-controllers, micro-processors or other electronic components, for performing the above described methods. It should be noted that the implementation process and technical principle of the electronic device in this embodiment can be referred to the above explanation of the audio signal encoding method in the embodiment of the disclosure, and will not be repeated here.

[0133] The electronic device 100 provided by the embodiment of the disclosure may execute the audio signal encoding method as described in some of the above embodiments, and its beneficial effects are the same as those of the above audio signal encoding method, which will not be repeated here.

[0134] In order to realize the above embodiments, the disclosure also provides a storage medium.

[0135] When the instructions stored in the storage medium are executed by the processor of the electronic device, the electronic device is caused to perform the audio signal encoding method described above. For example, the storage medium may be a ROM, a Random Access Memory (RAM), Compact Disc-ROM (CD-ROM), a magnetic tape, a floppy disk, an optical data storage device, etc.

[0136] In order to realize the above embodiments, the disclosure also provides a computer program product. When the computer program is executed by the processor of the electronic device, the electronic device is caused to perform the

audio signal encoding method as described above.

[0137] Other embodiments of the disclosure will be apparent to those skilled in the art from consideration of the specification and practice of the disclosure disclosed here. This application is intended to cover any variations, uses, or adaptations of the disclosure following the general principles thereof and including such departures from the disclosure as come within known or customary practice in the art. It is intended that the specification and examples are considered as illustrative only, with a true scope and spirit of the disclosure being indicated by the following claims.

[0138] It is clearly understood by those skilled in the art that for the convenience and conciseness of description, the specific working processes of the systems, devices and units described above can be referred to the corresponding processes in the aforementioned method embodiments, and will not be repeated here.

[0139] The above descriptions are specific implementations of the disclosure, but the protection scope of the disclosure is not limited thereto. Any technician familiar with the technical field can easily think of changes or substitutions within the technical scope disclosed in the disclosure, which should be included in the protection scope of the disclosure. Therefore, the protection scope of the disclosure should be based on the protection scope of the attached claims.

Claims

1. An audio signal encoding method, comprising:

obtaining a scene-based audio signal;
determining a number of audio channels of the audio signal and an encoding rate; and
generating an encoded codestream by encoding the audio signal according to the number of audio channels and the encoding rate.

2. The method of claim 1, wherein generating the encoded codestream by encoding the audio signal according to the number of audio channels and the encoding rate, comprises:

performing a down-mixed processing on the audio signal according to the number of audio channels and the encoding rate to generate a down-mixed parameter and a down-mixed audio channel signal;
encoding the down-mixed audio channel signal to generate an encoding parameter; and
generating the encoded codestream by performing codestream multiplexing on the down-mixed parameter and the encoding parameter.

3. The method of claim 2, wherein performing the down-mixed processing on the audio signal according to the number of audio channels and the encoding rate to generate the down-mixed parameter and the down-mixed audio channel signal, comprises:

determining a target control parameter for the audio signal according to the number of audio channels and the encoding rate;
determining a down-mixed processing algorithm according to the target control parameter; and
performing the down-mixed processing on the audio signal according to the down-mixed processing algorithm to generate the down-mixed parameter and the down-mixed audio channel signal.

4. The method of claim 3, wherein determining the target control parameter for the audio signal according to the number of audio channels and the encoding rate, comprises:

calculating an initial average rate of each audio channel according to the number of audio channels and the encoding rate;
determining a target average rate according to the initial average rate and a preset average rate threshold; and
determining the target control parameter for the audio signal according to the initial average rate and the target average rate.

5. The method of any one of claims 1-4, wherein before encoding the audio signal, the method further comprises:
performing a pre-emphasis preprocessing and/or a high-pass filtering preprocessing on the audio signal.

6. An audio signal encoding apparatus, comprising:

a signal obtaining unit, configured to obtain a scene-based audio signal;

an information determining unit, configured to determine a number of audio channels of the audio signal and an encoding rate; and
an encoding processing unit, configured to generate an encoded codestream by encoding the audio signal according to the number of audio channels and the encoding rate.

7. The apparatus of claim 6, wherein the encoding processing unit comprises:

a down-mixed processing module, configured to perform a down-mixed processing on the audio signal according to the number of audio channels and the encoding rate to generate a down-mixed parameter and a down-mixed audio channel signal;
a parameter generating module, configured to encode the down-mixed audio channel signal to generate an encoding parameter; and
a codestream generating module, configured to generate the encoded codestream by performing codestream multiplexing on the down-mixed parameter and the encoding parameter.

8. The apparatus of claim 7, wherein the down-mixed processing module comprises:

a parameter determining sub-module, configured to determine a target control parameter for the audio signal according to the number of audio channels and the encoding rate;
an algorithm determining sub-module, configured to determine a down-mixed processing algorithm according to the target control parameter; and
a down-mixed processing sub-module, configured to perform the down-mixed processing on the audio signal according to the down-mixed processing algorithm to generate the down-mixed parameter and the down-mixed audio channel signal.

9. The apparatus of claim 8, wherein the parameter determining sub-module is further configured to:

calculate an initial average rate of each audio channel according to the number of audio channels and the encoding rate;
determine a target average rate according to the initial average rate and a preset average rate threshold; and
determine the target control parameter for the audio signal according to the initial average rate and the target average rate.

10. The apparatus of any one of claims 6-9, further comprising:
a preprocessing unit, configured to perform a pre-emphasis preprocessing and/or a high-pass filtering preprocessing on the audio signal.

11. An electronic device, comprising:

at least one processor; and
a memory communicatively connected to the at least one processor;
wherein the memory stores instructions executable by the at least one processor, and the instructions are executed by the at least one processor to cause the at least one processor to perform the method of any one of claims 1-5.

12. A non-transitory computer-readable storage medium having computer instructions stored thereon, wherein the computer instructions are used to cause a computer to execute the method of any one of claims 1-5.

13. A computer program product comprising computer instructions, wherein when the computer instructions are executed by a processor, the method of any one of claims 1-5 is implemented.

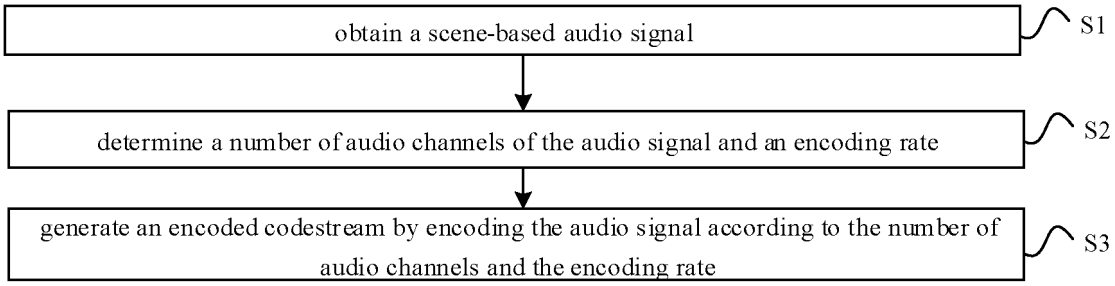

FIG. 1

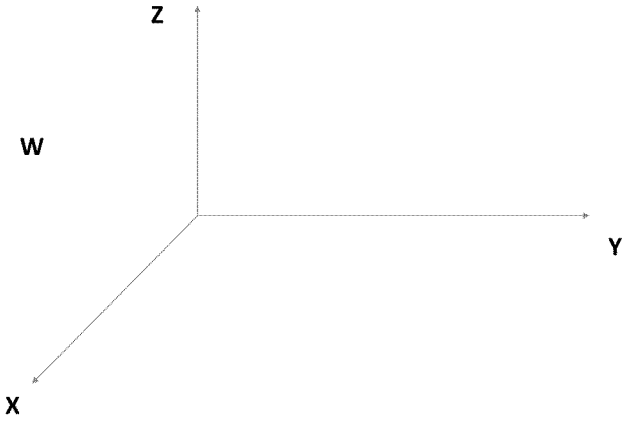

FIG. 2

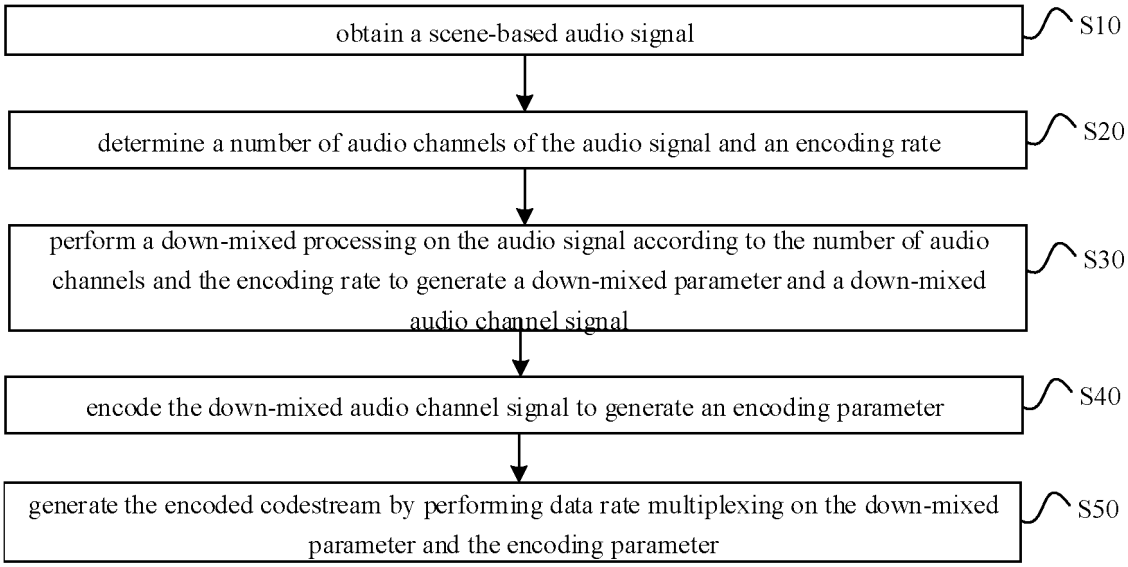

FIG. 3

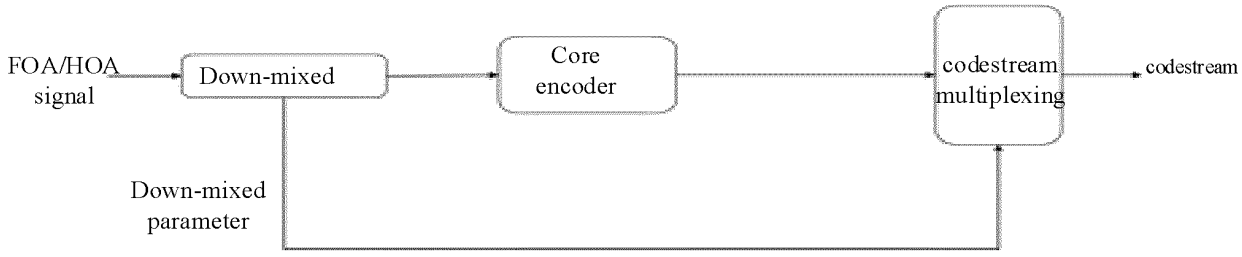

FIG. 4

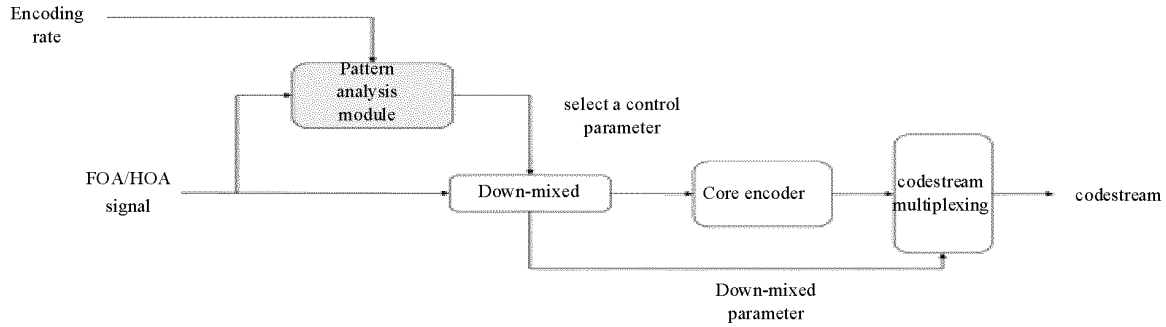

FIG. 5

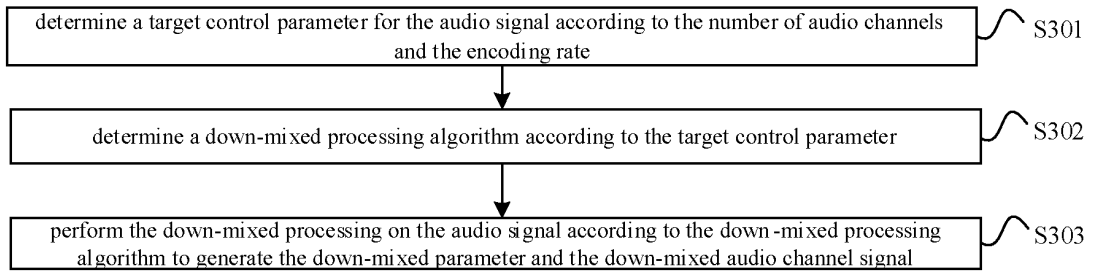

FIG. 6

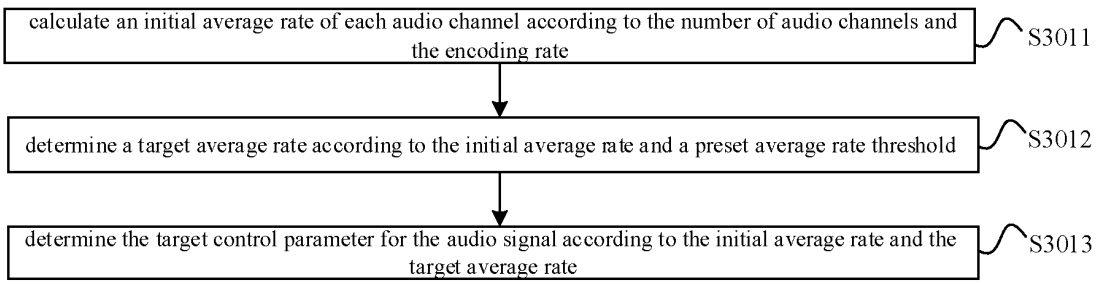

FIG. 7

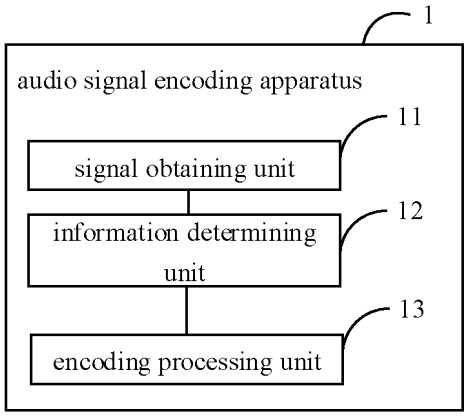

FIG. 8

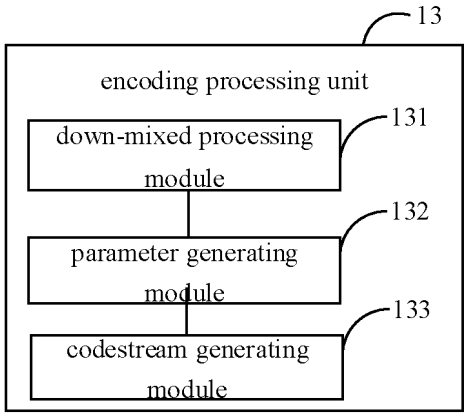

FIG. 9

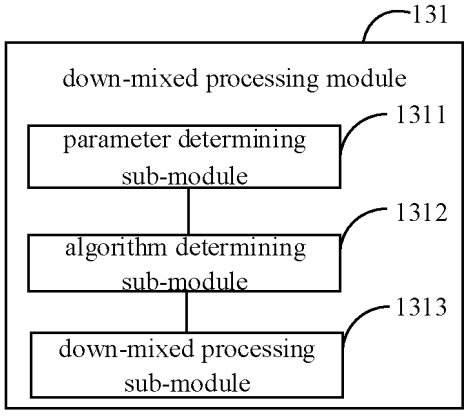

FIG. 10

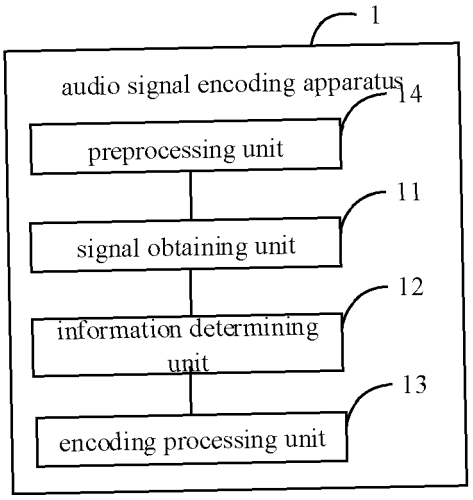
1
audio signal encoding apparatus
14
preprocessing unit
11
signal obtaining unit
12
information determining unit
13
encoding processing unit

FIG. 11

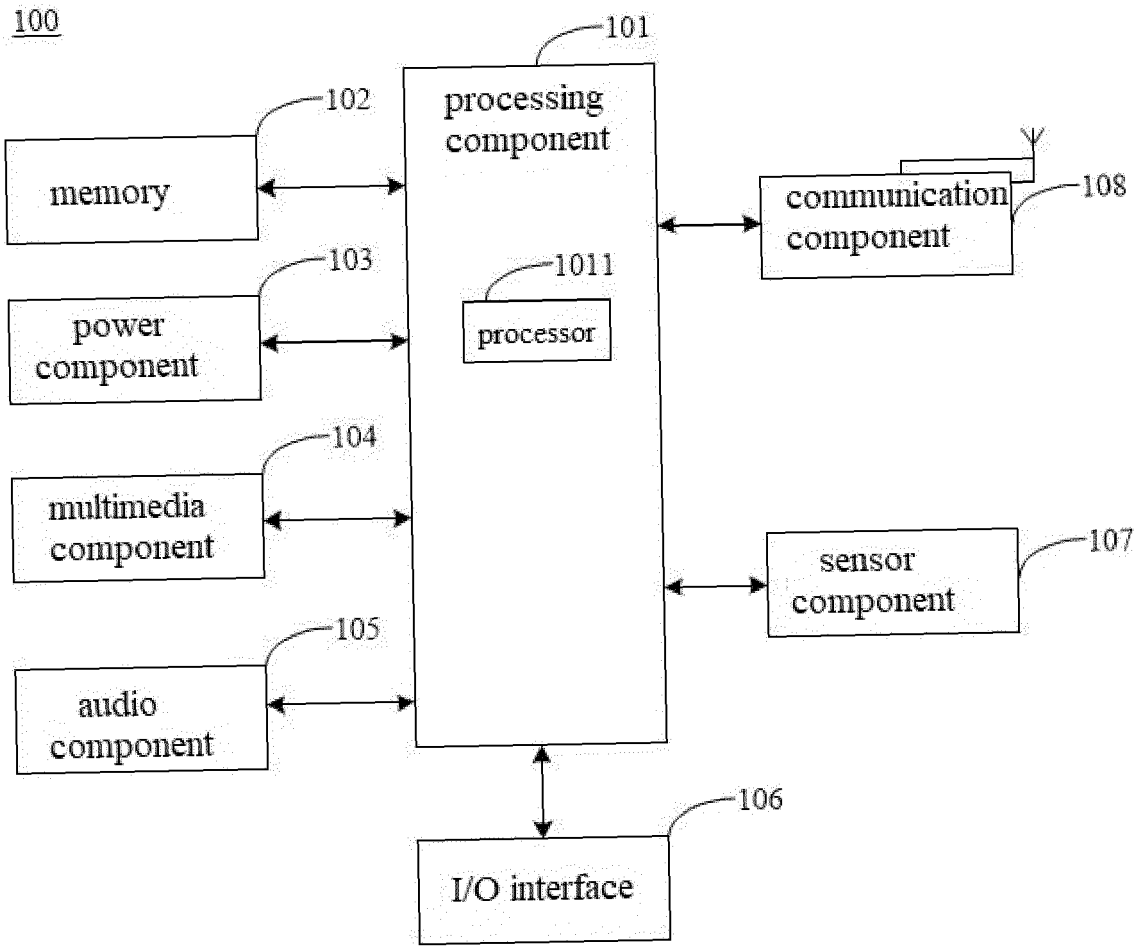
100
101
processing component
102
memory
103
power component
1011
processor
104
multimedia component
105
audio component
communication component
108
sensor component
107
106
I/O interface

FIG. 12

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2022/103170

A. CLASSIFICATION OF SUBJECT MATTER

H04S 3/00(2006.01)i;G10L 19/20 (2013.01)i

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:H04S G10L

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

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

CNABS; CNTXT; CNKI; VEN; WOTXT; USTXT; EPTXT: 音频, 信号, 声道, 速率, 码流, 编码, 下混, 滤波, voice, signal, channel, audio, rate, encode, flow, downmix, filter

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 110335615 A (BEIJING BYTEDANCE NETWORK TECHNOLOGY CO., LTD.) 15 October 2019 (2019-10-15) description, paragraphs [0047]-[0154]	1-13
X	CN 114582357 A (HUAWEI TECHNOLOGIES CO., LTD. et al.) 03 June 2022 (2022-06-03) description, paragraphs [0214]-[0597]	1-13
A	CN 109243488 A (TENCENT MUSIC ENTERTAINMENT TECHNOLOGY (SHENZHEN) CO., LTD.) 18 January 2019 (2019-01-18) entire document	1-13
A	US 2011002393 A1 (FUJITSU LTD.) 06 January 2011 (2011-01-06) entire document	1-13

☐ 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
"D" document cited by the applicant in the international application
"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
03 March 2023	**23 March 2023**
Name and mailing address of the ISA/CN **China National Intellectual Property Administration (ISA/CN)** **China No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088**	Authorized officer
Facsimile No. **(86-10)62019451**	Telephone No.

Form PCT/ISA/210 (second sheet) (July 2022)

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2022/103170

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN 110335615 A	15 October 2019	None	
CN 114582357 A	03 June 2022	None	
CN 109243488 A	18 January 2019	None	
US 2011002393 A1	06 January 2011	US 8818539 B2	26 August 2014

Form PCT/ISA/210 (patent family annex) (July 2022)