



(11)

EP 4 220 637 A1

(12)

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

(43) Date of publication:
02.08.2023 Bulletin 2023/31

(51) International Patent Classification (IPC):
G10L 21/0208 ^(2013.01) **H04R 3/00** ^(2006.01)

(21) Application number: **21870910.3**

(52) Cooperative Patent Classification (CPC):
G10L 21/0208; G10L 21/0216; H04R 3/00; H04S 7/00

(22) Date of filing: **29.06.2021**

(86) International application number:
PCT/CN2021/103110

(87) International publication number:
WO 2022/062531 (31.03.2022 Gazette 2022/13)

(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: **GUANGDONG OPPO MOBILE TELECOMMUNICATIONS CORP., LTD.**
Dongguan, Guangdong 523860 (CN)

(72) Inventor: **WANG, Wendong**
Dongguan, Guangdong 523860 (CN)

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

(30) Priority: **25.09.2020 CN 202011027264**

(54) **MULTI-CHANNEL AUDIO SIGNAL ACQUISITION METHOD AND APPARATUS, AND SYSTEM**

(57) A multi-channel audio signal acquisition method, comprising: acquiring a main audio signal acquired by a main device when photographing a target photographed object, and performing first multi-channel rendering to acquire an environmental multi-channel audio signal (201); acquiring an audio signal acquired by an additional device on the target photographed object, and determining a first additional audio signal (202); performing environmental sound suppression processing on the first additional audio signal and the main audio signal to acquire a target audio signal (203); performing second multi-channel rendering on the target audio signal to acquire a target multi-channel audio signal (204); and mixing the environmental multi-channel audio signal and the target multi-channel audio signal to acquire a mixed multi-channel audio signal (205). Also disclosed are a corresponding apparatus, a system, a terminal device and a computer-readable storage medium.

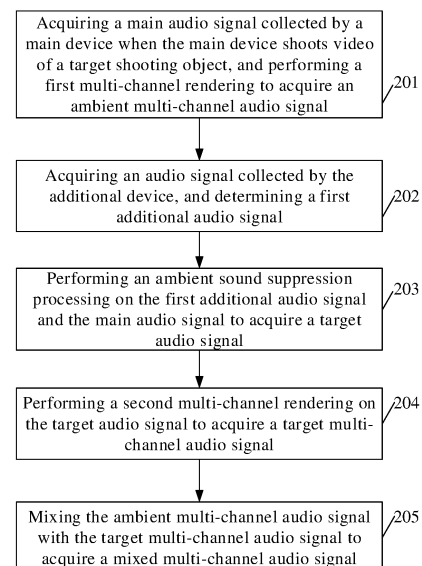


FIG. 2A

EP 4 220 637 A1

Description

TECHNICAL FIELD

[0001] The present disclosure relates to the field of audio technologies, in particular to a multi-channel audio signal acquisition method, a multi-channel audio signal acquisition device and a multi-channel audio signal acquisition system.

BACKGROUND

[0002] With the development of technology, people put forward higher requirements for performance of shooting and audio recording of mobile devices. At present, with a popularity of a true wireless stereo (TWS) Bluetooth headset, a distributed audio capture solution has been provided. This solution uses a microphone on the TWS Bluetooth headset to capture a high-quality close-up audio signal far away from a user, and mixes the spatial audio signals collected by the microphone array in a main device and performs a binaural rendering to simulate a point shaped auditory target in a spatial sound field, which creates a more real immersive experience. However, this solution only mixes the distributed audio signals, and does not suppress an ambient sound. When the user uses a mobile device to shoot video in an environment with multiple sound sources or in a noisy environment, a sound that a user is really interested in is mixed with various irrelevant sound sources, or even submerged in background noise. Therefore, solutions in related art may be affected by the ambient sound, such that the recording effect of audio signal is poor.

SUMMARY

[0003] A multi-channel audio signal acquisition method, a multi-channel audio signal acquisition device and a multi-channel audio signal acquisition system are provided in embodiments of the present disclosure, which can use a relationship between distributed audio signals to suppress an ambient sound and improve a recording effect of an audio signal.

[0004] In order to solve the above technical problems, the embodiments of the present disclosure are implemented as follows.

[0005] In a first aspect, a multi-channel audio signal acquisition method is provided in the embodiments of the present disclosure and includes following operations.

[0006] The method includes: acquiring a main audio signal collected by a main device when the main device shoots video, and performing a multi-channel rendering to acquire an ambient multi-channel audio signal.

[0007] The method includes: acquiring an audio signal collected by an additional device, and determining a first additional audio signal, a distance between the additional device and the target shooting object is less than the first threshold.

[0008] The method includes: performing an ambient sound suppression processing on the first additional audio signal and the main audio signal to acquire a target audio signal.

5 [0009] The method includes: performing a multi-channel rendering on the target audio signal to acquire a target multi-channel audio signal.

[0010] The method includes: mixing the ambient multi-channel audio signal and the target multi-channel audio signal to acquire a mixed multi-channel audio signal.

10 [0011] In a second aspect, a multi-channel audio signal acquisition device is provided and includes following components.

[0012] The multi-channel audio signal acquisition device includes an acquisition module configured to acquire a main audio signal collected by a main device when the main device shoots video of a target shooting object, and perform a first multi-channel rendering to acquire an ambient multi-channel audio signal, acquire an audio signal collected by an additional device, and determine a first additional audio signal, a distance between the additional device and the target shooting object is less than the first threshold.

20 [0013] The multi-channel audio signal acquisition device includes a processing module configured to perform an ambient sound suppression processing on the first additional audio signal and the main audio signal to acquire a target audio signal.

[0014] The processing module is configured to perform a multi-channel rendering on the target audio signal to acquire a target multi-channel audio signal.

30 [0015] The processing module is configured to mix the ambient multi-channel audio signal and the target multi-channel audio signal to acquire a mixed multi-channel audio signal.

35 [0016] In a third aspect, a terminal device is provided and includes a processor, a memory storing a computer program capable of running on the processor. The computer program is executed by the processor to perform the multi-channel audio signal acquisition method in the first aspect.

[0017] In a fourth aspect, a terminal device is provided and includes the multi-channel audio signal acquisition device in the second aspect and a main device.

40 [0018] The main device is configured to collect the main audio signal when the main device shoots video, and send the main audio signal to the multi-channel audio signal acquisition device.

[0019] In a fifth aspect, a multi-channel audio signal acquisition system is provided and includes the multi-channel audio signal acquisition device in the second aspect, a main device and an additional device, and the main device and the additional device establish a communication connection with the multi-channel audio signal respectively.

55 [0020] The main device is configured to collect a main audio signal when the main device shoots video, and send the main audio signal to the multi-channel audio

signal acquisition device.

[0021] The additional device is configured to collect a second additional audio signal, and send the second additional audio signal to the multi-channel audio signal acquisition device.

[0022] A distance between the additional device and the target shooting object is less than the first threshold.

[0023] In a six aspect, a computer-readable storage medium storing a computer program is provided, the computer program is executed by a processor to perform the multi-channel audio signal acquisition method in the first aspect.

[0024] In the embodiments of the present disclosure, the multi-channel audio signal acquisition method may include: acquiring a main audio signal collected by a main device when the main device shoots video, and performing a multi-channel rendering to acquire an ambient multi-channel audio signal; acquiring an audio signal collected by the additional device, and determining a first additional audio signal, a distance between the additional device and the target shooting object being less than the first threshold; performing an ambient sound suppression processing on the first additional audio signal and the main audio signal to acquire a target audio signal; performing a multi-channel rendering on the target audio signal to acquire a target multi-channel audio signal; mixing the ambient multi-channel audio signal and the target multi-channel audio signal to acquire a mixed multi-channel audio signal. In this way, distributed audio signals may be acquired from the main device and additional device, and the relationship between distributed audio signals may be used to perform the ambient sound suppression processing according to the first additional audio signal collected by the additional device and the main audio signal collected by the main device, so as to suppress the ambient sound in a recording process and acquire the target multi-channel audio signal. Then the ambient multi-channel audio signal (which is acquired by performing multi-channel rendering on the main audio signal) is mixed with the target multi-channel audio signal. Not only the distributed audio signals are mixed, and the point shaped auditory target in the space sound field is simulated, but also the ambient sound is suppressed, thereby improving the recording effect of the audio signal.

BRIEF DESCRIPTION OF THE DRAWINGS

[0025] In order to make the technical solution described in embodiments of the present disclosure more clearly, the drawings used for the description of the embodiments will be simply described. Apparently, the drawings in the following description are only some embodiments of the present disclosure. Other drawings may be acquired according to the drawings.

FIG. 1 is a schematic diagram of a multi-channel audio signal acquisition system according to some embodiments of the present disclosure.

FIG. 2A is a first flowchart of a multi-channel audio signal acquisition method according to some embodiments of the present disclosure.

FIG. 2B is a schematic diagram of an interface of a terminal device according to some embodiments of the present disclosure.

FIG. 3 is a second flowchart of a multi-channel audio signal acquisition method according to some embodiments of the present disclosure.

FIG. 4 is schematic diagram of a multi-channel audio signal acquisition device according to some embodiments of the present disclosure.

FIG. 5 is a structural schematic diagram of a terminal device according to some embodiments of the present disclosure.

FIG. 6 is a schematic diagram of a hardware structure of a terminal device according to some embodiments of the present disclosure.

DETAILED DESCRIPTION

[0026] The technical solutions in the embodiments of the present disclosure are clearly and completely described in conjunction with the drawings in the embodiments of the present disclosure. It is obvious that the described embodiments are only some embodiments of the present disclosure, and not all embodiments. All other embodiments acquired by those skilled in the art based on the embodiments in the present disclosure without the creative work are all within the scope of the present disclosure.

[0027] In the embodiments of the present disclosure, terms such as "exemplary" or "for example" are used as examples, exemplification or descriptions. Any embodiment or design solution described as "exemplary" or "for example" in the embodiments of the present disclosure should not be interpreted as more preferred or advantageous than other embodiments or designs. Specifically, the terms such as "exemplary" or "for example" are used to present relevant concepts in a specific manner. In addition, in the description of the embodiments of the present disclosure, unless otherwise specified, terms "multiple" or "a plurality of" mean two or more.

[0028] The term "and/or" in the embodiments of the present disclosure is just an association relationship that describes association objects, and it indicates three kinds of relationships. For example, A and/or B can indicate that there are three cases including: A alone, A and B together, and B alone.

[0029] The embodiments of the present disclosure provide a multi-channel audio signal acquisition method, a device and a system, which may be applied to video shooting scenes, especially applied to situations with multiple sound sources or noisy environments. The distributed audio signals are mixed, the point shaped auditory target in the space sound field is simulated, and the ambient sound is suppressed, thereby improving the recording effect of the audio signal.

[0030] As shown in FIG. 1, FIG. 1 is a schematic diagram of a multi-channel audio signal acquisition system according to some embodiments of the present disclosure. The system may include a main device, an additional device, and an audio processing device (such as a multi-channel audio acquisition device in embodiments of the present disclosure). The additional device in FIG. 1 may be a true wireless stereo (TWS) Bluetooth headset configured to collect audio streams (that is, an additional audio signal according to some embodiments of the present disclosure). The main device may be configured to collect video streams and audio streams (that is, a main audio signal according to some embodiments of the present disclosure). The audio processing device may include the following modules such as a target tracking module, a scene-sound-source classification module, a delay compensation module, an adaptive filtering module, a spatial filtering module, a binaural rendering module, and a mixer module, etc. Specific functions of each module are described in combination with the multi-channel audio signal acquisition method described in following embodiments, which is not repeated here.

[0031] It should be noted that the main device and the audio processing device in the embodiments of the present disclosure may be two independent devices. In some embodiments, the main device and the audio processing device may also be integrated in a device. For example, the integrated device may be a terminal device that integrates functions of the main device and the audio processing device.

[0032] In the embodiments of the present disclosure, a connection manner between the additional device and the terminal device, or between the additional device and the audio processing device may be a wireless communication such as a Bluetooth connection, or a wireless fidelity (WiFi) connection. In the embodiments of the present disclosure, the connection manner is not specifically limited.

[0033] The terminal device in the embodiments of the present disclosure may include a mobile phone, a tablet, a laptop, an ultra-mobile personal computer (UMPC), a handheld computer, a netbook, a personal digital assistant (PDA), a wearable device (such as a watch, a wrist, a glass, a helmet, or a headband, etc.), etc. The embodiments of the present disclosure do not make special limits on a specific form of the terminal device.

[0034] In the embodiments of the present disclosure, the additional device may be a terminal device independent of the main device and the audio processing device, and the mobile terminal device may be a portable terminal device such as a Bluetooth headset, a wearable device (such as a watch, a wrist, a glass, a helmet, and a headband, etc.), etc.

[0035] In a video shooting scene, the main device may shoot video, acquire the main audio signal, and send the main audio signal to the audio processing device. Since the additional device is close to a target shooting object in the video shooting scene (for example, a distance be-

tween the additional device and the target shooting object is less than a first threshold), the additional device may acquire the additional audio device, and then send it to the audio processing device.

[0036] In some embodiments, the target shooting object may be a person or a musical instrument in the video shooting scene.

[0037] In some embodiments, generally, a plurality of shooting objects may be occurred in the video shooting scene, and the target shooting object may be one of the plurality of shooting objects.

[0038] As shown in FIG. 2A, FIG. 2A is a flowchart of a multi-channel audio signal acquisition method according to some embodiments of the present disclosure. For example, the method may be performed by the audio processing device (i.e., the multi-channel audio acquisition device) as shown in FIG. 1, or performed by the terminal device that integrates functions of the audio processing device and the main device as shown in FIG. 1. In this case, the main device may be a functional module or functional entity that collects audio and video in the terminal device. In following embodiments, the terminal device performing the method is taken an example.

[0039] The method is described in detail below, as shown in FIG. 2A. The method may include following operations.

[0040] Operation 201 includes: acquiring a main audio signal collected by a main device when the main device shoots video of a target shooting object, and performing a first multi-channel rendering to acquire an ambient multi-channel audio signal.

[0041] A distance between the target shooting object and the additional device may be less than the first threshold.

[0042] In some embodiments, the user may arrange an additional device arranged on the target shooting object to be tracked, start a video shooting function of the terminal device, and select the target shooting object in a video content by clicking the video content displayed in a display screen. A radio module in the main device of the terminal device and a radio module in the additional device may start recording and collecting audio signals.

[0043] In some embodiments, the radio module in the main device may be a microphone array and the microphone array may be configured to collect the main audio signal. The radio module in the additional device may be a microphone.

[0044] As shown in FIG. 2B, FIG. 2B is a schematic diagram of an interface of the terminal device, and the display screen of the terminal device may display the video content. The user may click a character 21 displayed in the interface to determine the character 21 as the target shooting object. The character 21 may carry a Bluetooth headset (i.e., the additional device) to collect audio signal near the character 21, and the Bluetooth headset may send the audio signal to the terminal device.

[0045] In the embodiment of the present disclosure, the multi-channel may be dual channels, four channels,

5.1 channels or more channels.

[0046] When the audio signal acquired in the embodiments of the present disclosure is a dual channel audio signal, a binaural rendering may be performed on the main audio signal through a head related transfer function (HRTF) to acquire an ambient binaural audio signal.

[0047] For example, the binaural rendering may be performed on the main audio signal through the binaural renderer in FIG. 1 to acquire the environment binaural audio signal.

[0048] Operation 202 includes: acquiring an audio signal collected by an additional device, and determining a first additional audio signal.

[0049] In some embodiments, methods of acquiring an audio signal acquired by the additional device, and determining a first additional audio signal may include two implementation operations.

[0050] A first implementation operation includes: acquiring a second additional audio signal collected by the additional device arranged on the target shooting object, and determining the second additional audio signal as the first additional audio signal.

[0051] A second implementation operation includes: acquiring the second additional audio signal collected by the additional device arranged on the target shooting object, aligning the second additional audio signal with the main audio signal in a time domain to acquire the first additional audio signal.

[0052] Since there may be a distance between the main device and the additional device, there may be a delay between a time acquiring the main audio signal and a time acquiring the second additional audio signal. According to the delay, the main audio signal and the second additional audio signal may be aligned in a time domain to acquire the first additional audio signal.

[0053] Generally, in an audio signal acquisition system such as the multi-channel audio signal acquisition system shown in FIG. 1, there is also a system delay (for example, a delay caused by a Bluetooth transmission, and a delay caused by decoding module decoding), which may be measured. In some embodiments of the present disclosure, an actual delay may be acquired by combining an estimated acoustic wave propagation delay (i.e., the delay between the main audio signal and the second additional audio signal) with the system delay, and the main audio signal and the second additional audio signal may be aligned in the time domain according to the actual delay to acquire the first additional audio signal.

[0054] A delay compensator in FIG. 1 may be configured to align the additional audio signal with the main audio signal in the time domain according to the delay between the main audio signal and the second additional audio signal to acquire the first additional audio signal.

[0055] Operation 203 includes: performing an ambient sound suppression processing on the first additional audio signal and the main audio signal to acquire a target audio signal.

[0056] In the embodiments of the present disclosure,

for a situation that the target shooting object is within a shooting field of view (FOV) of the main device and a situation that the target shooting object is outside the shooting FOV of the main device, operations of performing the ambient sound suppression processing on the first additional audio signal and the main audio signal to acquire the target audio signal are different.

(1) For the situation that the target shooting object is within the shooting FOV of the main device

[0057] According to the shooting FOV of the main device, the spatial filtering is performed on the main audio signal in an area outside the shooting FOV of the main device to acquire reverse focusing audio signal. The reverse focusing audio signal is taken as a reference signal, and an adaptive filtering is performed on the first additional audio signal to acquire the target audio signal.

[0058] In this way, firstly, the spatial filtering is performed on the main audio signal in the area outside the shooting FOV of the main device to acquire the reverse focusing audio signal, which suppresses a sound signal at a location of the target shooting object included in the main audio signal to acquire a purer ambient audio signal. Then the reverse focusing audio signal is taken as a reference signal, and the adaptive filtering is performed on the first additional audio signal, the ambient sound in the additional audio signal may be further suppressed.

(2) For the situation that the target shooting object is outside the shooting FOV of the main device

[0059] According to the shooting FOV of the main device, the spatial filtering is performed on the main audio signal within the shooting FOV to acquire a focusing audio signal. The first additional audio signal is taken as the reference signal, and an adaptive filtering is performed on the focusing audio signal to acquire the target audio signal.

[0060] In this way, firstly, the spatial filtering is performed on the main audio signal in the area within the shooting FOV to acquire the focusing audio signal, which suppresses part of the ambient sound in the main audio signal. Then, the first additional audio signal is taken as the reference signal and the adaptive filtering is performed on the focusing audio signal, which may further suppress the ambient sound outside a focusing area that cannot be completely suppressed in the focusing audio signal, in particular a sound at a location of the target shooting object included in the ambient sound.

[0061] A spatial filter in FIG. 1 may be configured to perform the spatial filtering on the main audio signal to acquire a directionally enhanced audio signal. When the target shooting object is within the shooting FOV of the main device, since a high-quality close-up audio signal has been acquired through the first additional audio signal, a main purpose of the spatial filtering is to acquire a purer ambient audio signal. A target area of the spatial

filtering is an area outside the shooting FOV, and an acquired signal is called reverse focusing audio signal. When the target shooting object is outside the shooting FOV of the main device, the close-up audio signal in the area within the shooting FOV needs to be acquired through the spatial filtering, so the target area of spatial filtering is an area within the shooting FOV, and an acquired signal is the focusing audio signal.

[0062] The spatial filtering method may be based on a beamforming method such as a minimum variance distortionless response (MVDR) method, or a beamforming method of a general sidelobe canceller (GSC).

[0063] FIG. 1 includes two sets of adaptive filters. The two sets of adaptive filters are applied to the target audio signal acquired in the above two cases respectively. Specifically, only one set of adaptive filter may be enabled according to a change of the target shooting object in the shooting FOV. When the target shooting object is within the shooting FOV of the main device, the adaptive filter applied to the first additional audio signal is enabled, and the reverse focusing audio signal is taken as the reference signal and input to further suppress the ambient sound from the first additional audio signal, and make a sound near the target shooting object more prominent. When the target shooting object is outside the shooting FOV of the main device, the adaptive filter applied to the focusing audio signal is enabled, and the first additional audio signal is taken as the reference signal and input to further suppress the sound outside the shooting FOV from the focusing audio signal, especially a sound at the location of the target shooting object.

[0064] The adaptive filtering method may be a least mean square (LMS) method.

[0065] Operation 204 includes: performing a second multi-channel rendering on the target audio signal to acquire a target multi-channel audio signal.

[0066] For example, three sets of binaural renderers in FIG. 1 are applied to the main audio signal, the target audio signal performed on the adaptive filtering in above case (1), and the target audio signal performed on the adaptive filtering in above case (2) respectively to acquire three sets of binaural signals, i.e., an ambient binaural signal, an additional binaural signal, and a focusing binaural signal.

[0067] Since above cases (1) and (2) do not exist at a same time, the binaural renderer applied to the target audio signal of above case (1) and the binaural renderer applied to the target audio signal of above case (2) may not be enabled at the same time, and the two binaural renderers may be selected to be enabled according to the change of the target shooting object in the shooting FOV of the main device. The binaural renderer applied to the main audio signal is always enabled.

[0068] Further, when the target shooting object is within the shooting FOV of the main device, the binaural renderer applied to the target audio signal in above case (1) is enabled. When the target shooting object is outside the shooting FOV of the main device, the binaural ren-

derer applied to the target audio signal in above case (2) is enabled.

[0069] In some embodiments, the binaural renderer may include a deccorelator and a convolver inside, and needs an HRTF corresponding to a target location to simulate a perception of an auditory target in desired direction distance.

[0070] In some embodiments, the scene-sound-source classification module may be used to determine a rendering rule according to a determined current scene and the sound source type of the target shooting object, the determined rendering rule may be applied to the deccorelator to acquire different rendering styles, and an azimuth and a distance between the additional device and the main device may be used to control to generate the HRTF. A HRTF corresponding to a particular location may be acquired by interpolating on a set of previously stored HRTF, or by using a method based on a deep neural network (DNN).

[0071] Operation 205 includes: mixing the ambient multi-channel audio signal with the target multi-channel audio signal to acquire a mixed multi-channel audio signal.

[0072] In the embodiments of the present disclosure, mixing the ambient multi-channel audio signal and the target multi-channel audio signal means the ambient multi-channel audio signal adding up the target multi-channel audio signal according to a gain. Specifically, the ambient multi-channel audio signal adding up the target multi-channel audio signal according to a gain may indicate that signal sampling points in the ambient multi-channel audio signal add up, and then add up signal sampling points in the target multi-channel audio signal.

[0073] The gain may be a preset fixed value or a variable gain.

[0074] In some embodiments, the variable gain may be determined according to the shooting FOV.

[0075] A mixer in FIG. 1 is configured to mix two of the three sets of binaural signals mentioned above. When the target shooting object is within the shooting FOV of the main device, the ambient binaural signal and the additional binaural signal are mixed. When the target shooting object is outside the shooting FOV of the main device, the ambient binaural signal and the focusing binaural signal are mixed.

[0076] In the embodiments of the present disclosure, the method may include: acquiring the main audio signal acquired by the main device when the main device shoots video of the target shooting object, and performing the first multi-channel rendering to acquire an ambient multi-channel audio signal; acquiring the audio signal acquired by the additional device arranged on the target shooting object, the distance between the audio signal acquired by the additional device and the target shooting object being less than the first threshold, and determining a first additional audio signal; performing ambient sound suppression processing on the first additional audio signal and the main audio signal to acquire the target audio

signal; performing the second multi-channel rendering on the target audio signal to acquire the target multi-channel audio signal; and mixing the ambient multi-channel audio signal and the target multi-channel audio signal to acquire the mixed multi-channel audio signal. In this way, the distributed audio signals may be acquired from the main device and additional device, and the relationship between distributed audio signal may be used to perform the ambient sound suppression processing according to the first additional audio signal collected by the additional device and the main audio signal collected by the main device, so as to suppress the ambient sound in the recording process and acquire the target multi-channel audio signal. Then the ambient multi-channel audio signal (which is acquired by performing multi-channel rendering on the main audio signal) is mixed with the target multi-channel audio signal, not only the distributed audio signals are mixed, and the point shaped auditory target in the space sound field is simulated, but also the ambient sound is suppressed, thereby improving the recording effect of the audio signal.

[0077] As shown in FIG. 3, the embodiments of the present disclosure also provide a multi-channel audio signal acquisition method, which includes following operations.

[0078] Operation 301 includes: acquiring a main audio signal collected by a microphone array in a main device.

[0079] Operation 302 includes: acquiring a second additional audio signal collected by an additional device.

[0080] After the user selects a target shooting object on the main device and starts shooting video, a terminal device may perform the operations 301 and 302 described above. The terminal device may continuously track a movement of the target shooting object in the shooting FOV in response to the change of the shooting FOV the main device.

[0081] In some embodiments, the method may include: acquiring video data (including the main audio signal) shot by the main device and the second additional audio signal collected by the additional device.

[0082] Further, the method may include: determining a type of current scene and a type of the target shooting object according to above video data and/or the second additional audio signal, matching a rendering rule through the type of the current scene and the type of the target shooting object, performing a multi-channel rendering on a subsequent audio signal according to the determined rendering rule.

[0083] In some embodiments, the method may include: performing the second multi-channel rendering on the target audio signal according to the determined rendering rule to acquire a target multi-channel audio signal, and performing a first multi-channel rendering on the main audio signal according to the determined rendering rule to acquire an ambient multi-channel audio signal.

[0084] In some embodiments, the operation of performing a multi-channel rendering on the target audio signal according to the determined rendering rule to ac-

quire a target multi-channel audio signal may include following operations.

[0085] The operations include: acquiring video data shot by the main device and the second additional audio signal collected by the additional device.

[0086] The operations include: determining a type of a current scene and a type of the target shooting object.

[0087] The operations include: performing the multi-channel rendering on the target audio signal through the first rendering rule matching the type of the current scene and the type of the target shooting object to acquire the target multi-channel audio signal.

[0088] In some embodiments, the operation of performing a multi-channel rendering on the main audio signal according to the determined rendering rule to acquire an ambient multi-channel audio signal may include following operations.

[0089] The operations include: acquiring the main audio signal collected by the main device when the main device shoots video of the target shooting object.

[0090] The operations include: determining a type of a current scene.

[0091] The operations include: performing the first multi-channel rendering on the main audio signal through the second rendering rule matching the type of the current scene to acquire the ambient multi-channel audio signal.

[0092] In FIG. 1, the scene-sound-source classification module may include two paths, video stream information is applied to one of the two paths using and audio stream information is applied to another path. The two paths may include a scene analyzer and a voice/instrument classifier. The scene analyzer may analyze a current space where the user is according to the video or audio, the current space includes a small room, a medium room, a large room, a concert hall, a stadium, or outdoor, etc. The voice/instrument classifier may analyze a current sound source near the target shooting object according to the video or audio, the current sound source includes a male voice, a female, a child, an accordion, a guitar, a bass, a piano, a keyboard and a percussion instrument.

[0093] In some embodiments, both the scene analyzer and the voice/instrument classifier may be used based DNN methods. The video is input by each frame of images, and the audio is input by a Mel spectrum or a Mel-frequency cepstrum coefficient (MFCC) of sound.

[0094] In some embodiments, a rendering rule to be used in a following binaural rendering module may also be determined by combining a result of spatial scene analysis and the voice/instrument classifier with user preferences.

[0095] Operation 303 may include: generating a first multi-channel transfer function according to a type of the microphone array in the main device, performing the multi-channel rendering on the main audio signal according to the first multi-channel transfer function to acquire the ambient multi-channel audio signal.

[0096] It should be noted that when the multi-channel in the embodiments of the present disclosure is a dual

channel, the first multi-channel transfer function may be an HRTF function.

[0097] In the embodiments of the present disclosure, a set of preset HRTF function and binaural rendering method may be set in the binaural renderer in FIG. 1. The preset HRTF function is determined according to the type of the microphone array in the main device, and the binaural rendering is performed on main audio signal the by the HRTF function to acquire the ambient binaural audio signal.

[0098] Operation of 304 includes: judging whether the target shooting object is within the shooting FOV of the main device.

[0099] When it is detected that the target shooting object is within the shooting FOV of the main device, following operations 305 to 312, and 320 to 323 are performed. When it is detected that the target shooting object is outside the shooting FOV of the main device, following operations 313 to 319, and 320 to 323 are performed.

[0100] A target tracking module in FIG. 1 may include a visual target tracker and an audio target tracker configured to determine a position of the target shooting object, and estimate an azimuth and a distance between the target shooting object and the main device by using visual data and/or an audio signal. When the target shooting object is within the shooting FOV of the main device, the visual data and the audio signal may be used to determine the position of the target shooting object. At this time, the visual target tracker and the audio target tracker are enabled at the same time. When the target shooting object is outside the shooting FOV of the main device, the audio signal may be used to determine the position of the target shooting object. At this time, only the audio target tracker may be enabled.

[0101] In some embodiments, when the target shooting object is within the shooting FOV of the main device, one of the visual data and the audio signal may also be used to determine the position of the target shooting object.

[0102] Operation 305 includes: determining a first azimuth between the target shooting object and the main device according to video information and shooting parameters acquired by the main device, acquiring a first active duration of the second additional audio signal and a first distance, and determining a second active duration of the main audio signal according to the first active duration and the first distance.

[0103] The first distance is a target distance between a last determined target shooting object and the main device.

[0104] Operation 306 includes: performing a direction-of-arrival (DOA) estimation by using the main audio signal in the second active duration to acquire a second azimuth between the target shooting object and the main device, performing a smoothing processing on the first azimuth and the second azimuth to acquire a target azimuth.

[0105] Operation 307 includes: determining a second

distance between the target shooting object and the main device according to the video information acquired by the main device, and calculating a second delay according to the second distance and the sound speed.

[0106] Operation 308 includes: performing a beamforming processing on the main audio signal toward the target azimuth to acquire a beamforming signal, and determining a first delay between the beamforming signal and the second additional audio signal.

[0107] In FIG. 1, a sound source direction measurement and a beamformer may be used to perform the beamforming processing on the main audio signal toward the target azimuth to acquire the beamforming signal, and a delay estimator may be configured to further determine the first delay between the beamforming signal and the second additional audio signal.

[0108] Operation 309 includes: performing the smoothing processing on the second delay and the first delay to acquire a target delay, and calculating the target distance according to the target delay and the sound speed.

[0109] When the target shooting object is within the shooting FOV of the main device, the video data acquired at this time includes the target shooting object. At this time, the first azimuth may be acquired according to the position of the target shooting object in a video frame shot by the video frame combined with prior information such as camera parameters (such as a focal length) and zoom scale (different shooting fields correspond to different zoom scales). The azimuth and distance between the target shooting object and the main device may be determined by the audio signal to acquire the second azimuth. The target azimuth is acquired by performing the smoothing processing on the first azimuth and the second azimuth.

[0110] Further, through comparing a range of the target shooting object shot in the video frame with a typical range of the target shooting object recorded in advance, and combining with the prior information such as the camera parameters (such as a focal length) and the zoom scale (different shooting fields correspond to different zoom scales), a rough distance estimation may be performed to acquire the above second distance. According to the second distance, the sound speed and a predicted system delay, the second delay may be acquired. The delay (i.e., the first delay) between the second additional audio signal and the main audio signal may be calculated. The target delay may be acquired by performing the smoothing processing on the first delay and the second delay.

[0111] In the embodiments of the present disclosure, the smoothing processing may include calculating an average value. When the target azimuth is acquired by performing the smoothing processing on the first azimuth and the second azimuth, an average value of the first azimuth and the second azimuth may be calculated as the target azimuth. The target delay may be acquired by performing the smoothing processing on the first delay

and the second delay, and an average value of the first delay and the second delay may be taken as the target delay.

[0112] When the target shooting object is within the shooting FOV of the main device, the visual target tracker in FIG. 1 may be configured to detect the target azimuth and the target distance between the target shooting object and the main device through the shot video. An advantage of the visual target tracker is that its tracking results are more accurate than the audio target tracker in noisy environments or when there are many sound sources.

[0113] Further, the visual target tracker and the audio target tracker are configured to simultaneously detect the target azimuth and the target distance between the target shooting object and the main device, thereby further improving an accuracy.

[0114] Operation 310 includes: aligning, according to the target delay, the second additional audio signal with the main audio signal in the time domain to acquire the first additional audio signal.

[0115] Operation 311 includes: performing, according to the shooting FOV of the main device, the spatial filtering on the main audio signal in the area outside the shooting FOV to acquire the reverse focusing audio signal.

[0116] Operation 312 includes: taking the reverse focusing audio signal as the reference signal, performing the adaptive filtering on the first additional audio signal to acquire the target audio signal.

[0117] Operation 313 includes: acquiring the first active duration of the second additional audio signal and the first distance, and determining the second active duration of the main audio signal according to the first active duration and the first distance.

[0118] The first distance is the target distance between the last determined target shooting object and the main device.

[0119] In the embodiments of the present disclosure, an active duration of the audio signal is a duration when there is an effective audio signal in the audio signal. In some embodiments, a first active duration of the second additional audio signal may be a duration when there is an effective audio signal in the second additional audio signal.

[0120] In some embodiments, the effective audio signal may be voice or instrument voice. Exemplarily, the effective audio signal may be a sound of the target shooting object.

[0121] In the embodiments of the present disclosure, the delay between the second additional audio signal and the main audio signal may be determined according to the first distance and the sound speed, and then the audio signal of the second active duration corresponding to the second additional audio signal in the main audio signal may be determined according to the delay and the first active duration.

[0122] Operation 314 includes: performing the DOA estimation by using the main audio signal in the second

active duration to acquire the target azimuth between the target shooting object and the main device.

[0123] Operation 315 includes: performing the beamforming processing on the main audio signal toward the target azimuth to acquire the beamforming signal, and determining the first delay between the beamforming signal and the second additional audio signal.

[0124] Operation 316 includes: calculating the target distance between the target shooting object and the main device according to the first delay and the sound speed.

[0125] When the target shooting object is outside the shooting FOV of the main device, the video data acquired at this time does not include the target shooting object. At this time, the audio signal may be used to determine the position of the target shooting object.

[0126] In FIG. 1, the audio target tracker may estimate the target azimuth and target distance between the target shooting object and the main device by using the main audio signal and the additional audio signal, operations of estimating the target azimuth and target distance between the target shooting may specifically include a sound source direction measurement, a beamforming, and a delay estimation.

[0127] Specifically, the target azimuth may be acquired by performing the DOA estimation on the main audio signal. In order to avoid the impact of noisy environment or multiple sound sources on DOA estimation, the second additional audio may be analyzed before performing DOA estimation, and a duration corresponding to an active part of effective audio signal (which may be an audio signal with the sound of the target shooting object) of the second additional audio may be acquired, that is, the first active duration may be acquired. The delay (i.e., the first delay) between the second additional audio signal and the main audio signal may be acquired according to a last estimated target distance, and the first active duration is corresponded to the second active duration in the main audio signal. Then a segment of the main audio signal at the second active duration is cut out and performed conduct DOA estimation to acquire an azimuth between the target shooting object and the main device, and the azimuth is taken as the above target azimuth.

[0128] In some embodiments, when the DOA estimation is performed, a generalized cross correlation (GCC) method of phase transform (PHAT) may be used to perform a time-difference-of-arrival (TDOA) estimation, and then the DOA may be acquired by combining type information of the microphone array. After the DOA estimation is acquired, the multi-channel main audio signal acquires the beamforming signal through a fixed direction beamformer, and a directional enhancement is performed toward the direction of the above target azimuth to improve an accuracy of a next delay estimation. The beamforming method may be a delay-sum or a minimum variance distortion response (MVDR). The above first delay estimation is also performed between the main audio beamforming signal and the second additional audio signal by using the TDOA method. Similarly, the TDOA estimation

is also performed only during the active duration of the second additional audio signal. According to the first delay, the sound speed and the predicted system delay, the distance between the target shooting object and the main device may be acquired, that is, the target distance may be acquired.

[0129] Operation 317 includes: aligning, according to the first delay, the second additional audio signal with the main audio signal in the time domain to acquire the first additional audio signal.

[0130] When the target shooting object is outside the shooting FOV of the main device, the first delay is taken as the target delay between the main audio signal and the second additional audio signal, and according to the first delay, the second additional audio signal is aligned with the main audio signal in the time domain to acquire the first additional audio signal.

[0131] The delay compensator in FIG. 1 may align, according to the first delay to acquire the first additional audio signal, the second additional audio signal with the main audio signal in the time domain.

[0132] Operation 318 includes: performing, according to the shooting FOV of the main device, the spatial filtering on the main audio signal within the shooting FOV to acquire the focusing audio signal.

[0133] Operation 319 includes: taking the first additional audio signal as the reference signal, performing the adaptive filtering on the focusing audio signal to acquire the target audio signal.

[0134] When the target shooting object is within the shooting FOV of the main device, since the high-quality close-up audio signal has been acquired through the additional audio signal, a main purpose of spatial filtering is to acquire a purer ambient audio signal, so a target area of spatial filtering is outside the shooting FOV, and an acquired signal is hereinafter referred to as the reverse focusing audio signal. When the target shooting object is outside the shooting FOV, a close-up audio signal within the shooting FOV needs to be acquired through the spatial filtering, so the target area of spatial filtering is the shooting FOV, and an acquired signal is hereinafter referred to as the focusing audio signal.

[0135] Further, when spatial filtering is performed, the shooting FOV of the main device is combined, a change of the shooting FOV of the main device may be followed, such that a local audio signal is directionally enhanced.

[0136] In FIG. 1, two sets of adaptive filters are applied to the focusing audio signal and the additional audio signal respectively. Only one set of adaptive filter is enabled according to the change of the target shooting object in the shooting FOV. When the target shooting object is within the shooting FOV, the adaptive filter applied to the additional audio signal is enabled, and the reverse focusing audio signal is taken as the reference signal and input to further suppress the ambient sound from the additional audio signal, such that a sound near the target shooting object is more prominent. When the target shooting object is outside the shooting FOV, the adaptive

filter applied to the focusing audio signal is enabled, and the additional audio signal is taken as the reference signal and input to further suppress the sound outside the shooting FOV from the focusing audio signal. The adaptive filtering method may be the LMS, etc.

[0137] Operation 320 includes: generating a second multi-channel transfer function according to the target distance and the target azimuth.

[0138] Operation 321 includes: performing the multi-channel rendering on the target audio signal according to the second multi-channel transfer function to acquire the target multi-channel audio signal.

[0139] Operation 322 includes: determining a first gain of the ambient multi-channel audio signal and a second gain of the target multi-channel audio signal according to shooting parameters of the main device.

[0140] Operation 323 includes: mixing the ambient multi-channel audio signal with the target multi-channel audio signal according to the first gain and the second gain to acquire the mixed multi-channel audio signal.

[0141] In FIG. 1, a mixed gain controller may determine a mixed gain according to the user's shooting FOV, that is, the mixed gain is a proportion of two groups of signals in the mixed signal. For example, when a zoom level of the camera is increased, that is, when the FOV of the camera is reduced, a gain of the ambient binaural audio signal is reduced, a gain of the additional binaural audio signal (that is, the determined target multi-channel audio signal when the target shooting object is within the FOV) or the focusing binaural audio signal (that is, the determined target multi-channel audio signal when the target shooting object is outside the FOV) is increased. In this way, when the shooting FOV of the video is focused on a particular area, the audio is also focused on the particular area.

[0142] In the embodiments of the present disclosure, the range of the shooting FOV is determined according to the shooting parameters of the main device (such as the zoom level of the camera), and the first gain of the ambient multi-channel audio signal and the second gain of the target multi-channel audio signal are determined accordingly, such that when the shooting FOV of the video is focused to the particular area, the audio is also be focused to the particular area, thereby creating an effect of "immersive, sound follows image".

[0143] The multi-channel audio signal acquisition method provided by the embodiments of the present disclosure is a distributed recording and audio focusing method that may create a more realistic sense of presence. This method may simultaneously use the microphone array in the main device and the microphone in the additional device (TWS Bluetooth headset) of the terminal device for a distributed audio acquisition and fusion. The microphone array of the terminal device collects the spatial audio (that is, the main audio signal in the embodiments of the present disclosure) at the location of the main device, and the TWS Bluetooth headset may be arranged on the target shooting object to be

tracked, move along with the movement of the target shooting object to collect the high-quality close-up audio signal (that is, the first additional audio signal in the embodiments of the present disclosure) in the distance, , perform a corresponding adaptive filtering on the two groups of collected signals by combining with a FOV change in the video shooting process to achieve the ambient sound suppression, perform the spatial filtering on the spatial audio signal in the specified area to achieve the directional enhancement, track and locate the interested target shooting object in combination with the two positioning methods of vision and sound, perform a HRTF binaural rendering and an up mixing or a down mixing on the three groups signals respectively including the spatial audio, the high-quality close-up audio and the directional enhancement audio, acquire three sets of binaural signals including the ambient binaural signals, the additional binaural signals and the focusing binaural signal are acquired, determine a mixing proportion of the three sets of binaural signals according to the range of the FOV, and mix the three sets of binaural signals.

[0144] This technical solution may have following technical effects.

[0145] When the finally output binaural audio signal is played in a stereo headset, a spatial sound field and a point shaped auditory target at the specified position may simultaneously simulated.

[0146] A good directional enhancement effect may be acquired by using the distributed audio signal, and interference sound and ambient sound may be suppressed obviously when the distributed audio signal is focused.

[0147] The sounds that the user is interested in are easily focused and tracked by following the changes of the FOV, thereby creating an immersive experience of "immersive, sound follows image".

[0148] As shown in FIG. 4, the embodiments of the present disclosure provide a multi-channel audio signal acquisition device 400, which may include following modules.

[0149] The multi-channel audio signal acquisition device 400 includes an acquisition module 401 configured to acquire a main audio signal collected by a main device when the main device shoots video of a target shooting object, and perform a first multi-channel rendering to acquire an ambient multi-channel audio signal, acquire an audio signal collected by an additional device, and determine a first additional audio signal. A distance between the additional device and the target shooting object is less than a first threshold.

[0150] The multi-channel audio signal acquisition device 400 includes a processing module 402 configured to perform an ambient sound suppression processing on the first additional audio signal and the main audio signal to acquire a target audio signal.

[0151] The processing module 402 is configured to perform a second multi-channel rendering on the target audio signal to acquire a target multi-channel audio signal.

[0152] The processing module 402 is configured to mix the ambient multi-channel audio signal and the target multi-channel audio signal to acquire a mixed multi-channel audio signal.

5 **[0153]** In some embodiments, the processing module 402 is configured to determine a first gain of the ambient multi-channel audio signal and a second gain of the target multi-channel audio signal according to shooting parameters of the main device.

10 **[0154]** The processing module 402 is configured to mix the ambient multi-channel audio signal with the target multi-channel audio signal according to the first gain and the second gain to acquire the mixed multi-channel audio signal.

15 **[0155]** In some embodiments, the acquisition module 401 is configured to acquire the main audio signal collected by a microphone array in the main device.

[0156] The acquisition module 401 is configured to generate a first multi-channel transfer function according to a type of the microphone array in the main device.

20 **[0157]** The acquisition module 401 is configured to perform a multi-channel rendering on the main audio signal according to the first multi-channel transfer function to acquire the ambient multi-channel audio signal.

25 **[0158]** In some embodiments, the acquisition module 401 is configured to acquire a second additional audio signal collected by the additional device arranged on the target shooting object, and determine the second additional audio signal as the first additional audio signal.

30 **[0159]** In some embodiments, the acquisition module 401 is configured to acquire the second additional audio signal collected by the additional device arranged on the target shooting object, and align the second additional audio signal with the main audio signal in a time domain to acquire the first additional audio signal.

35 **[0160]** In some embodiments, the processing module 402 is configured to acquire a target azimuth between the target shooting object and the main device.

40 **[0161]** The processing module 402 is configured to perform a beamforming processing on the main audio signal toward the target azimuth to acquire a beamforming signal.

45 **[0162]** The processing module 402 is configured to determine a target delay between the main audio signal and the second additional audio signal.

[0163] The processing module 402 is configured to align, according to the first delay, the second additional audio signal with the main audio signal in a time domain to acquire the first additional audio signal.

50 **[0164]** In some embodiments, the processing module 402 is configured to acquire a target distance and the target azimuth between the target shooting object and the main device.

55 **[0165]** The processing module 402 is configured to generate a second multi-channel transfer function according to the target distance and the target azimuth.

[0166] The processing module 402 is configured to perform the multi-channel rendering on the target audio

signal according to the second multi-channel transfer function to acquire the target multi-channel audio signal.

[0167] In some embodiments, the acquisition module 401 is configured to acquire a first active duration of the second additional audio signal and a first distance when it is detected that the target shooting object is outside the shooting field of view of the main device. The first distance is the target distance between a last determined target shooting object and the main device.

[0168] The acquisition module 401 is configured to determine a second active duration of the main audio signal according to the first active duration and the first distance. The acquisition module 401 is specifically configured to perform a direction-of-arrival (DOA) estimation by using the main audio signal in the second active duration to acquire a target azimuth between the target shooting object and the main device.

[0169] In some embodiments, the acquisition module 401 is configured to perform the beamforming processing on the main audio signal toward the target azimuth to acquire the beamforming signal when the target shooting object is detected to be outside the shooting field of view of the main device.

[0170] The acquisition module 401 is configured to determine the first delay between the beamforming signal and the second additional audio signal.

[0171] The acquisition module 401 is configured to calculate the target distance between the target shooting object and the main device according to the first delay and the sound speed.

[0172] In some embodiments, the processing module 402 is configured to perform the spatial filtering on the main audio signal within the shooting field of view according to the shooting field of view of the main device to acquire a focusing audio signal when the target shooting object is detected to be outside the shooting field of view of the main device.

[0173] The processing module 402 is configured to take the first additional audio signal as the reference signal, perform an adaptive filtering on the focusing audio signal to acquire the target audio signal.

[0174] In some embodiments, the acquisition module 401 is configured to determine a first azimuth between the target shooting object and the main device according to video information and shooting parameters acquired by the main device when it is detected that the target shooting object is within the shooting field of view of the main device.

[0175] The acquisition module 401 is configured to acquire a first active duration of the second additional audio signal and a first distance. The first distance is a target distance between a last determined target shooting object and the main device.

[0176] The acquisition module 401 is configured to determine a second active duration of the main audio signal according to the first active duration and first distance.

[0177] The acquisition module 401 is configured to perform the DOA estimation by using the main audio signal

in the second active duration to acquire a second azimuth between the target shooting object and the main device.

[0178] The acquisition module 401 is configured to perform a smoothing processing on the first azimuth and the second azimuth to acquire the target azimuth.

[0179] In some embodiments, the acquisition module 401 is configured to determine a second distance between the target shooting object and the main device according to the video information acquired by the main device when it is detected that the target shooting object is within the shooting field of view of the main device.

[0180] The acquisition module 401 is configured to calculate a second delay according to the second distance and the sound speed.

[0181] The acquisition module 401 is configured to perform a beamforming processing on the main audio signal toward the target azimuth to acquire a beamforming signal.

[0182] The acquisition module 401 is configured to determine a first delay between the beamforming signal and the second additional audio signal.

[0183] The acquisition module 401 is configured to perform a smoothing processing on the second delay and the first delay to acquire a target delay.

[0184] The acquisition module 401 is specifically configured to calculate a target distance according to the target delay and the sound speed.

[0185] In some embodiments, the processing module 402 is configured to perform, according to the shooting field of view of the main device, the spatial filtering on the main audio signal in the area outside the shooting field of view to acquire the reverse focusing audio signal when the target shooting object is detected to be within the shooting field of view of the main device.

[0186] the processing module 402 is configured to take the reverse focusing audio signal as the reference signal, perform the adaptive filtering on the first additional audio signal to acquire the target audio signal.

[0187] In some embodiments, the processing module 402 is configured to acquire the video data shot by the main device and the second additional audio signal collected by the additional device.

[0188] The processing module 402 is configured to determine a type of the current scene and a type of target shooting object.

[0189] The processing module 402 is configured to perform the multi-channel rendering on the target audio signal through a first rendering rule matching the type of the current scene and the type of the target shooting object to acquire the target multi-channel audio signal.

[0190] In some embodiments, the processing module 402 is configured to acquire the main audio signal collected by the main device when the main device shoots video of the target shooting object.

[0191] The processing module 402 is configured to determine a type of a current scene.

[0192] The processing module 402 is configured to perform the first multi-channel rendering on the main au-

dio signal through the second rendering rule matching the type of the current scene to acquire the ambient multi-channel audio signal.

[0193] The embodiments of the present disclosure provide a terminal device including a processor, a memory, and a computer program stored on the memory and capable of running on the processor. The computer program is executed by the processor to perform the multi-channel audio signal acquisition method provided by the embodiment of the above method.

[0194] As shown in FIG. 5, the embodiments of the present disclosure also provide a terminal device including a multi-channel audio signal acquisition device 400 and a main device 500.

[0195] The main device is configured to collect the main audio signal when the main device shoots video, and send the main audio signal to the multi-channel audio signal acquisition device.

[0196] As shown in FIG. 6, the embodiments of the present disclosure also provide a terminal device including but not limited to a radio frequency (RF) circuit 601, a memory 602, an input unit 603, a display unit 604, a sensor 605, an audio circuit 606, a WiFi module 607, a processor 608, a Bluetooth module 609, a camera 610 and other components. The RF circuit 601 includes a receiver 6011 and a transmitter 6012. Those skilled may understand that the terminal device shown in FIG. 6 does not limit to the terminal device, the terminal device may include more or fewer components than the terminal device shown in FIG. 6, combine some components, or include different component arrangements.

[0197] The RF circuit 601 may be configured to receive and send information or receive and send signal during a call. Specifically, a downlink information of a base station is received and sent to the processor 608 for processing. In addition, designed uplink data is sent to the base station. Generally, the RF circuit 601 includes but is not limited to an antenna, at least one amplifier, a transceiver, a coupler, a low noise amplifier (LNA), and a duplexer, etc. In addition, the RF circuit 601 may also communicate with the network and other devices through the wireless communication. The above wireless communication may use any communication standard or protocol including but not limited to the global system of mobile communication (GSM), the general packet radio service (GPRS), the code division multiple access (CDMA), the wideband code division multiple access (WCDMA), the long term evolution (LTE), the E-mail, and the short messaging service (SMS), etc.

[0198] The memory 602 may be configured to store software programs and modules, and the processor 608 may execute various functional applications and data processing of the terminal device by running the software programs and modules stored in the memory 602. The memory 602 may mainly include a program storage area and a data storage area, the program storage area may store an operating system, an application program required for at least one function (such as a sound playing

function or an image playing function, etc.), etc. The data storage area may store data (such as audio signal or phone book, etc.) that has been created and used during using the terminal device. In addition, the memory 602 may include a high-speed random access memory, and may also include non-volatile memory such as at least one disk storage component, a flash memory component, or other volatile solid-state storage components.

[0199] The input unit 603 may be configured to receive input digital or character information and generate key signal input related to user settings and function control of the terminal device. Specifically, the input unit 603 may include a touch panel 6031 and other input devices 6032. The touch panel 6031 known as the touch screen may collect the user's touch operations (such as the user's operation on or near the touch panel 6031 with any suitable object or accessory such as fingers, and stylus, etc.) on or near it, and drive a corresponding connection device according to a preset program. In some embodiments, the touch panel 6031 may include two parts: a touch detection device and a touch controller. The touch detection device detects a user's touch position and a signal brought by the touch operation, and transmits the signal to the touch controller. The touch controller receives touch information from the touch detection device, converts the touch information into contact coordinates, and then sends the contact coordinates to the processor 608, and may receive commands from the processor 608 and execute the commands. In addition, the touch panel 6031 may be realized by a resistance, a capacitance, an infrared ray, and a surface acoustic wave, etc. In addition to the touch panel 6031, the input unit 603 may also include the other input devices 6032. Specifically, the other input devices 6032 may include but are not limited to one or more of a physical keyboard, a function key (such as a volume control key or a switch key, etc.), a trackball, a mouse, and a joystick, etc.

[0200] The display unit 604 may be configured to display information input by the user, information provided to the user and various menus of the terminal device. The display unit 604 may include a display panel 6041. In some embodiments, the display panel 6041 may be configured in a form of a liquid crystal display (LCD), or an organic light emitting diode (OLED), etc. Further, the touch panel 6031 may cover the display panel 6041. When the touch panel 6031 detects the touch operation on or near it, the touch operation is transmitted to the processor 608 to determine a touch event, and then the processor 608 provides a corresponding visual output on the display panel 6041 according to the touch event. In FIG. 6, although the touch panel 6031 and the display panel 6041 are two independent components to realize the input and output functions of the terminal device, in some embodiments, the touch panel 6031 may be integrated with the display panel 6041 to perform the input and output functions of the terminal device.

[0201] The terminal device may also include at least one sensor 605 such as a light sensor, a motion sensor,

and other sensors. Specifically, the light sensor may include an ambient light sensor and a proximity sensor. The ambient light sensor may adjust a brightness of the display panel 6041 according to a brightness of the ambient light, and the proximity sensor may exit the display panel 6041 and/or backlight when the terminal device moves to the ear. As a kind of motion sensor, an accelerometer sensor may detect value of acceleration in all directions (generally three-axis), and may detect value and direction of gravity when the accelerometer sensor is stationary, which may be configured to identify applications of pose of the terminal device (such as horizontal and vertical screen switching, related games, magnetometer pose calibration), functions related to vibration recognition (such as a pedometer, knocking), etc. A gyroscope, a barometer, a hygrometer, a thermometer, an infrared sensor and other sensors that may also be arranged in the terminal device are not described here. In the embodiments of the present disclosure, the terminal device may include an acceleration sensor, a depth sensor, and a distance sensor, etc.

[0202] The audio circuit 606, a loudspeaker 6061 and a microphone 6062 may provide audio interfaces between the user and the terminal device. The audio circuit 606 may transmit a converted electrical signal of the received audio signal to the loudspeaker 6061, and then the loudspeaker 6061 convert the electrical signal into a sound signal for output. On the other hand, the microphone 6062 converts the collected sound signal into an electrical signal, which is received by the audio circuit 606 and converted into an audio signal, and then the audio signal is output to the processor 608 for processing, then the audio signal is sent to another terminal device through the RF circuit 601, or the audio signal is output to the memory 602 for further processing. The microphone 6062 may be a microphone array.

[0203] The WiFi is a short-range wireless transmission technology. The terminal device may help the user send and receive e-mails, browse web pages and access streaming media through the WiFi module 607. The WiFi provides the user with wireless broadband Internet access. Although FIG. 6 shows the WiFi module 607, it may be understood that the WiFi module 607 has no need to be included in the terminal device, and may be omitted as needed without changing the essence of the present disclosure.

[0204] The processor 608 is a control center of the terminal device, which connect various parts of the entire terminal device through various interfaces and circuits, and performs various functions and processes data of the terminal device by running or executing software programs and/or modules stored in the memory 602, and calling data stored in the memory 602, so as to monitor the terminal device. In some embodiments, the processor 608 may include one or more processing units. In some embodiments, the processor 608 may integrate an application processor and a modem processor, the application processor mainly processes an operating system,

a user interface, and an application program, etc., and the modem processor mainly processes wireless communication. It should be understood that the above modem processor may not be integrated into the processor 608.

[0205] The terminal device may also include a Bluetooth module 609 configured for short distance wireless communication and may be divided into a Bluetooth data module and a Bluetooth voice module according to functions. The Bluetooth module is a basic circuit set of chips integrated with Bluetooth function, which is configured for wireless network communication. The Bluetooth module may be roughly divided into three types: a data transmission module, a Bluetooth audio module, and a Bluetooth module combining audio and data, etc.

[0206] Although not shown, the terminal device may also include other functional modules, which will not be repeated here.

[0207] In the embodiments of the present disclosure, the microphone 6062 may be configured to collect the main audio signal, and the terminal device may connect to the additional device through the WiFi module 607 or the Bluetooth module 609, and receive the second additional audio signal collected by the additional device.

[0208] The processor 608 is configured to acquire the main audio signal, perform the multi-channel rendering, acquire the ambient multi-channel audio signal, acquire the audio signal collected by the additional device, determine the first additional audio signal, perform the ambient sound suppression processing through the first additional audio signal and the main audio signal to acquire a target audio signal, perform the multi-channel rendering on the target audio signal to acquire the target multi-channel audio signal, and mix the ambient multi-channel audio signal with the target multi-channel audio signal to acquire the mixed multi-channel audio signal. The distance between the additional device and the target shooting object is less than the first threshold value.

[0209] In some embodiments, the processor 608 may also be configured to perform other processes implemented by the terminal device in the above method embodiments, which is not be repeated here.

[0210] The embodiments of the present disclosure also provide a multi-channel audio signal acquisition system including a multi-channel audio signal acquisition device, a main device, and an additional device. The main device and the additional device establish communication connections with the multi-channel audio signal respectively.

[0211] The main device is configured to collect the main audio signal when main device shoots video of the target shooting object, and send the main audio signal to the multi-channel audio signal acquisition device.

[0212] The additional device is configured to collect the second additional audio signal and send the second additional audio signal to the multi-channel audio signal acquisition device.

[0213] For example, the multi-channel audio signal ac-

quisition system may be as shown in FIG. 1, the audio processing device in FIG. 1 may be the multi-channel audio signal acquisition device.

[0214] The embodiments of the present disclosure also provide a computer-readable storage medium including a computer program, and the multi-channel audio signal acquisition method in the above method embodiments are performed when the computer program is executed by a processor.

[0215] In order to enable those skilled in the art to better understand the solutions of the present disclosure, the technical solutions in the embodiments of the present disclosure is described below in combination with the drawings in the embodiments of the present disclosure. Obviously, the described embodiments are only some embodiments of the present disclosure, not all embodiments. Other embodiments based on the embodiments of the present disclosure all belong to the protection scope of the present disclosure.

[0216] Those skilled in the art may clearly understand that for the convenience and conciseness of description, specific working processes of the system, the device and the units described above may refer to corresponding processes in the above method embodiments, which is not repeated here.

[0217] In embodiments of the present disclosure, it should be understood that the system, the device and the method may be realized in other ways. For example, the device embodiments described above is only exemplary. For example, a division of the units is only a division according to logical function, and there may be another division mode when it is actually implemented. For example, multiple units or components may be combined or integrated into another system, or some features may be ignored or not performed. On the other hand, mutual coupling, direct coupling or communication connection shown or discussed above may be indirect coupling or communication connection through some interfaces, indirect coupling or communication connection of devices or units may be electrical, mechanical or other forms.

[0218] The units spaced apart may or may not be physically spaced apart, and the displayed unit may or may not be a physical unit, that is, the displayed unit may be located in one place or distributed to multiple network units. Some or all of the units may be selected according to the practical needs to achieve the purpose of the embodiments.

[0219] In addition, each functional unit in the embodiments of the present disclosure may be integrated in a processing unit, or each unit may physically exist independently, or two or more units may be integrated in a unit. The above integrated units may be realized in a form of hardware or a software functional unit.

[0220] When the integrated unit is realized in the form of the software functional unit and sold or used as an independent product, the integrated unit may be stored in a computer-readable storage medium. Based on this understanding, the technical solution of the present dis-

closure may be embodied in the form of software product in essence, or the part that contributes to the related art may be embodied in the form of software product, or the whole or part of the technical solution may be embodied in the form of software product. The computer software product is stored in a storage medium including a number of instructions to enable a computer device (which may be a personal computer, a server, or a network device, etc.) to perform all or some the operations of the method described in various embodiments of the present invention. The aforementioned storage medium may include a USB flash disk, a mobile hard disk, a read-only memory (ROM), a random access memory (RAM), a magnetic disc or an optical disc and other medium that may store program codes.

[0221] As mentioned above, the above embodiments are only used to illustrate the technical solutions of the present disclosure, not to limit it. Although the present disclosure has been described in detail with reference to the aforementioned embodiments, those skilled in the art should understand that the technical solutions recorded in the aforementioned embodiments may be modified, or some of the technical features may be equally substituted. These modifications or substitutions do not make the essence of the corresponding technical solutions separate from the spirit and scope of the technical solutions of the embodiments of the present disclosure.

Claims

1. A multi-channel audio signal acquisition method, **characterized by** comprising:

acquiring a main audio signal collected by a main device when the main device shoots video of a target shooting object, and performing a first multi-channel rendering to acquire an ambient multi-channel audio signal;
acquiring an audio signal collected by an additional device, and determining a first additional audio signal, wherein a distance between the additional device and the target shooting object is less than a first threshold;
performing an ambient sound suppression processing on the first additional audio signal and the main audio signal to acquire a target audio signal;
performing a second multi-channel rendering on the target audio signal to acquire a target multi-channel audio signal; and
mixing the ambient multi-channel audio signal and the target multi-channel audio signal to acquire a mixed multi-channel audio signal.

2. The method as claimed in claim 1, wherein the mixing the ambient multi-channel audio signal and the target multi-channel audio signal to acquire a mixed multi-

channel audio signal, comprises:

determining a first gain of the ambient multi-channel audio signal and a second gain of the target multi-channel audio signal according to shooting parameters of the main device; and mixing the ambient multi-channel audio signal with the target multi-channel audio signal according to the first gain and the second gain to acquire the mixed multi-channel audio signal.

3. The method as claimed in claim 1, wherein the acquiring a main audio signal collected by a main device when the main device shoots video of a target shooting object, and performing a first multi-channel rendering to acquire an ambient multi-channel audio signal, comprises:

acquiring the main audio signal collected by a microphone array in the main device; generating a first multi-channel transfer function according to a type of the microphone array in the main device, and performing a first multi-channel rendering on the main audio signal according to the first multi-channel transfer function to acquire the ambient multi-channel audio signal.

4. The method as claimed in claim 1, wherein the acquiring an audio signal collected by an additional device, and determining a first additional audio signal, comprises:

acquiring a second additional audio signal collected by the additional device, and determining the second additional audio signal as the first additional audio signal; or acquiring a second additional audio signal collected by the additional device, and aligning the second additional audio signal with the main audio signal in a time domain to acquire the first additional audio signal.

5. The method as claimed in claim 4, wherein the aligning the second additional audio signal with the main audio signal in a time domain to acquire the first additional audio signal, comprises:

acquiring a target azimuth between the target shooting object and the main device; determining a target delay between the main audio signal and the second additional audio signal; and aligning, according to the target delay, the second additional audio signal with the main audio signal in the time domain to acquire the first additional audio signal.

6. The method as claimed in claim 1, wherein the performing a second multi-channel rendering on the target audio signal to acquire a target multi-channel audio signal, comprises:

acquiring a target distance and a target azimuth between the target shooting object and the main device; generating a second multi-channel transfer function according to the target distance and the target azimuth; and performing the second multi-channel rendering on the target audio signal according to the second multi-channel transfer function to acquire the target multi-channel audio signal.

7. The method as claimed in claim 6, wherein when the target shooting object is detected to be within a shooting field of view of the main device, the acquiring a target azimuth between the target shooting object and the main device, comprises:

determining a first azimuth between the target shooting object and the main device according to video information and shooting parameters acquired by the main device; acquiring a first active duration of the second additional audio signal and a first distance, determining a second active duration of the main audio signal according to the first active duration and first distance; wherein the first distance is a target distance between a last determined target shooting object and the main device; and performing a direction-of-arrival estimation by using the main audio signal in the second active duration to acquire a second azimuth between the target shooting object and the main device, performing a smoothing processing on the first azimuth and the second azimuth to acquire the target azimuth.

8. The method as claimed in claim 7, wherein the acquiring a target distance between the target shooting object and the main device, comprises:

determining a second distance between the target shooting object and the main device according to video information acquired by the main device, calculating a second delay according to the second distance and a sound speed; performing a beamforming processing on the main audio signal toward the target azimuth to acquire a beamforming signal, determining a first delay between the beamforming signal and the second additional audio signal; and performing a smoothing processing on the second delay and the first delay to acquire a target delay, calculating the target distance according

to the target delay and the sound speed.

9. The method as claimed in any one of claims 1 to 8, wherein when the target shooting object is detected to be within a shooting field of view of the main device, the performing an ambient sound suppression processing on the first additional audio signal and the main audio signal to acquire a target audio signal, comprises:

performing a spatial filtering in a region outside the shooting field of view of the main device according to the shooting field of view of the main device to acquire a reverse focusing audio signal; and
taking the reverse focusing audio signal as a reference signal, performing an adaptive filtering on the first additional audio signal to acquire the target audio signal.

10. The method as claimed in claim 6, wherein when the target shooting object is detected to be outside a shooting field of view of the main device, the acquiring a target azimuth between the target shooting object and the main device, comprises:

acquiring a first active duration of the second additional audio signal and a first distance, wherein the first distance is a target distance between a last determined target shooting object and the main device;
determining a second active duration of the main audio signal according to the first active duration and first distance; and
performing a direction-of-arrival estimation by using the main audio signal in the second active duration to acquire the target azimuth between the target shooting object and the main device.

11. The method as claimed in claim 6, wherein when the target shooting object is detected to be outside a shooting field of view of the main device, the acquiring a target distance between the target shooting object and the main device, comprises:

performing a beamforming processing on the main audio signal toward the target azimuth to acquire a beamforming signal; determining a first delay between the beamforming signal and the second additional audio signal; and
calculating the target distance between the target shooting object and the main device according to the first delay and a sound speed.

12. The method as claimed in any one of claims 1 to 6, 10 and 11, wherein when the target shooting object is detected to be outside a shooting field of view of the main device, the performing an ambient sound

suppression processing on the first additional audio signal and the main audio signal to acquire a target audio signal, comprises:

performing a spatial filtering in a region within the shooting field of view of the main device according to the shooting field of view of the main device to acquire a focusing audio signal; and
taking the first additional audio signal as a reference signal, performing an adaptive filtering on the first additional audio signal to acquire the target audio signal.

13. The method as claimed in claim 1, wherein performing a second multi-channel rendering on the target audio signal to acquire a target multi-channel audio signal, comprises:

acquiring video data shot by the main device and a second additional audio signal collected by the additional device;
determining a type of a current scene and a type of the target shooting object; and
performing the second multi-channel rendering on the target audio signal through a first rendering rule matching the type of the current scene and the type of the target shooting object to acquire the target multi-channel audio signal.

14. The method as claimed in claim 1, wherein the acquiring a main audio signal collected by a main device when the main device shoots video of a target shooting object, and performing a first multi-channel rendering on the main audio signal to acquire an ambient multi-channel audio signal, comprises:

acquiring the main audio signal collected by the main device when the main device shoots video of the target shooting object;
determining a type of a current scene; and
performing the first multi-channel rendering on the main audio signal through a second rendering rule matching the type of the current scene to acquire the ambient multi-channel audio signal.

15. A multi-channel audio signal acquisition device, **characterized by** comprising:

an acquisition module, configured to acquire a main audio signal collected by a main device when the main device shoots video of a target shooting object, perform a first multi-channel rendering to acquire an ambient multi-channel audio signal; acquire an audio signal collected by an additional device, and determine a first additional audio signal, wherein a distance between the additional device and the target shoot-

ing object is less than a first threshold; and
 a processing module, configured to perform an
 ambient sound suppression processing on the
 first additional audio signal and the main audio
 signal to acquire a target audio signal; perform
 a second multi-channel rendering on the target
 audio signal to acquire a target multi-channel
 audio signal; and mix the ambient multi-channel
 audio signal and the target multi-channel audio
 signal to acquire a mixed multi-channel audio
 signal.

16. A terminal device, **characterized by** comprising:

a processor;
 a memory, storing a computer program capable
 of running on the processor;
 wherein the processor is configured to:

acquire a main audio signal collected by a
 main device when the main device shoots
 video of a target shooting object, and per-
 forming a first multi-channel rende on the
 main audio signal acquire an ambient multi-
 channel audio signal;
 acquire an audio signal collected by an ad-
 ditional device, and determine a first addi-
 tional audio signal, wherein a distance be-
 tween the additional device and the target
 shooting object is less than a first threshold;
 perform an ambient sound suppression
 processing on the first additional audio sig-
 nal and the main audio signal to acquire a
 target audio signal;
 perform a second multi-channel rendering
 on the target audio signal to acquire a target
 multi-channel audio signal; and
 mix the ambient multi-channel audio signal
 and the target multi-channel audio signal to
 acquire a mixed multi-channel audio signal.

17. The terminal device as claimed in claim 16, wherein
 the processor is configured to:

determine a first gain of the ambient multi-chan-
 nel audio signal and a second gain of the target
 multi-channel audio signal according to shooting
 parameters of the main device; and
 mix the ambient multi-channel audio signal with
 the target multi-channel audio signal according
 to the first gain and the second gain to acquire
 the mixed multi-channel audio signal.

18. The terminal device as claimed in claim 16, wherein
 the processor is configured to:

acquire the main audio signal collected by a mi-
 crophone array on the main device;

generate a first multi-channel transfer function
 according to a type of the microphone array on
 the main device; and
 perform the first multi-channel rendering on the
 main audio signal according to the first multi-
 channel transfer function to acquire the ambient
 multi-channel audio signal.

19. The terminal device as claimed in claim 16, wherein
 the processor is configured to:

acquire a second additional audio signal collect-
 ed by the additional device, and determine the
 second additional audio signal as the first addi-
 tional audio signal; or
 acquire a second additional audio signal collect-
 ed by the additional device, align the second ad-
 ditional audio signal with the main audio signal
 in a time domain to acquire the first additional
 audio signal.

20. The terminal device as claimed in claim 19, wherein
 the processor is configured to:

acquire a target azimuth between the target
 shooting object and the main device;
 determine a target delay between the main au-
 dio signal and the second additional audio sig-
 nal; and
 align, according to the target delay, the second
 additional audio signal with the main audio sig-
 nal in the time domain to acquire the first addi-
 tional audio signal.

21. The terminal device as claimed in claim 16, wherein
 the processor is configured to:

acquire a target distance and a target azimuth
 between the target shooting object and the main
 device;
 generate a second multi-channel transfer func-
 tion according to the target distance and the tar-
 get azimuth; and
 perform the second multi-channel rendering on
 the target audio signal according to the second
 multi-channel transfer function to acquire the
 target multi-channel audio signal.

22. The terminal device as claimed in claim 21, wherein
 the processor is configured to:

acquire a first active duration of the second ad-
 ditional audio signal and a first distance, wherein
 the first distance is a target distance between a
 last determined target shooting object and the
 main device;
 determine a second active duration of the main
 audio signal according to the first active duration

- and first distance;
perform a direction-of-arrival estimation by using the main audio signal in the second active duration to acquire a second azimuth between the target shooting object and the main device;
and
perform a smoothing processing on the first azimuth and the second azimuth to acquire the target azimuth.
23. The terminal device as claimed in claim 22, wherein the processor is configured to:
- determine a second distance between the target shooting object and the main device according to video information acquired by the main device;
calculate a second delay according to the second distance and a sound speed;
perform a beamforming processing on the main audio signal toward the target azimuth to acquire a beamforming signal;
determine a first delay between the beamforming signal and the second additional audio signal;
perform a smoothing processing on the second delay and the first delay to acquire a target delay;
and
calculate the target distance according to the target delay and the sound speed.
24. The terminal device as claimed in any one of claims 16 to 23, wherein the processor is configured to:
- perform a spatial filtering in a region outside the shooting field of view of the main device according to the shooting field of view of the main device to acquire a reverse focusing audio signal;
and
take the reverse focusing audio signal as a reference signal, perform an adaptive filtering on the first additional audio signal to acquire the target audio signal.
25. The terminal device as claimed in claim 22, wherein the processor is configured to:
- acquire a first active duration of the second additional audio signal and a first distance, wherein the first distance is a target distance between a last determined target shooting object and the main device;
determine a second active duration of the main audio signal according to the first active duration and first distance; and
perform a direction-of-arrival estimation by using the main audio signal in the second active duration to acquire the target azimuth between
- the target shooting object and the main device.
26. The terminal device as claimed in claim 22, wherein the processor is configured to:
- perform a beamforming processing on the main audio signal toward the target azimuth to acquire a beamforming signal;
determine a first delay between the beamforming signal and the second additional audio signal; and
calculate the target distance between the target shooting object and the main device according to the first delay and a sound speed.
27. The terminal device as claimed in any one of claims 16 to 22, 25 and 26, wherein the processor is configured to:
- perform a spatial filtering in a region within the shooting field of view of the main device according to the shooting field of view of the main device to acquire a focusing audio signal; and
take the first additional audio signal as a reference signal, perform an adaptive filtering on the first additional audio signal to acquire the target audio signal.
28. The terminal device as claimed in claim 16, wherein the processor is configured to:
- acquire video data shot by the main device and a second additional audio signal collected by the additional device; and
determine a type of a current scene and a type of the target shooting object; and
perform the second multi-channel rendering on the target audio signal through a first rendering rule matching the type of the current scene and the type of the target shooting object to acquire the target multi-channel audio signal.
29. The terminal device as claimed in claim 16, wherein the processor is configured to:
- acquire the main audio signal collected by the main device when the main device shoots video of the target shooting object;
determine a type of a current scene; and
perform the first multi-channel rendering on the main audio signal through a second rendering rule matching the type of the current scene to acquire the ambient multi-channel audio signal.
30. A terminal device, **characterized by** comprising the multi-channel audio signal acquisition device as claimed in claim 15 and a main device;
wherein the main device is configured to collect the

main audio signal when the main device shoots video of a target shooting object, and send the main audio signal to the multi-channel audio signal acquisition device.

5

31. A multi-channel audio signal acquisition system, **characterized by** comprising the multi-channel audio signal acquisition device as claimed in claim 15, a main device and an additional device, the main device and the additional device establishing a communication connection with the multi-channel audio signal respectively; wherein

10

the main device is configured to collect a main audio signal when the main device shoots video of a target shooting object, and send the main audio signal to the multi-channel audio signal acquisition device;

15

the additional device is configured to collect a second additional audio signal, and send the second additional audio signal to the multi-channel audio signal acquisition device;

20

wherein a distance between the additional device and the target shooting object is less than the first threshold.

25

32. A computer-readable storage medium, **characterized by** storing a computer program, the computer program being executed by a processor to perform the multi-channel audio signal acquisition method as claimed in any one of claims 1 to 14.

30

35

40

45

50

55

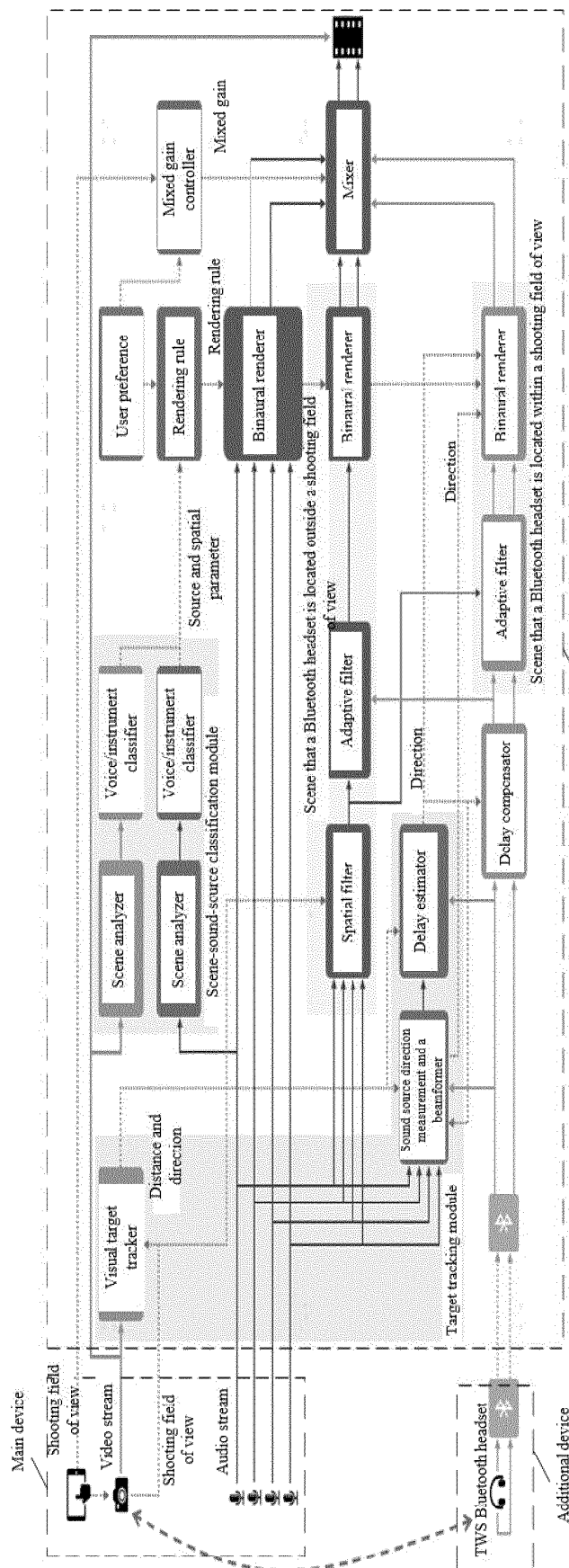


FIG. 1
device

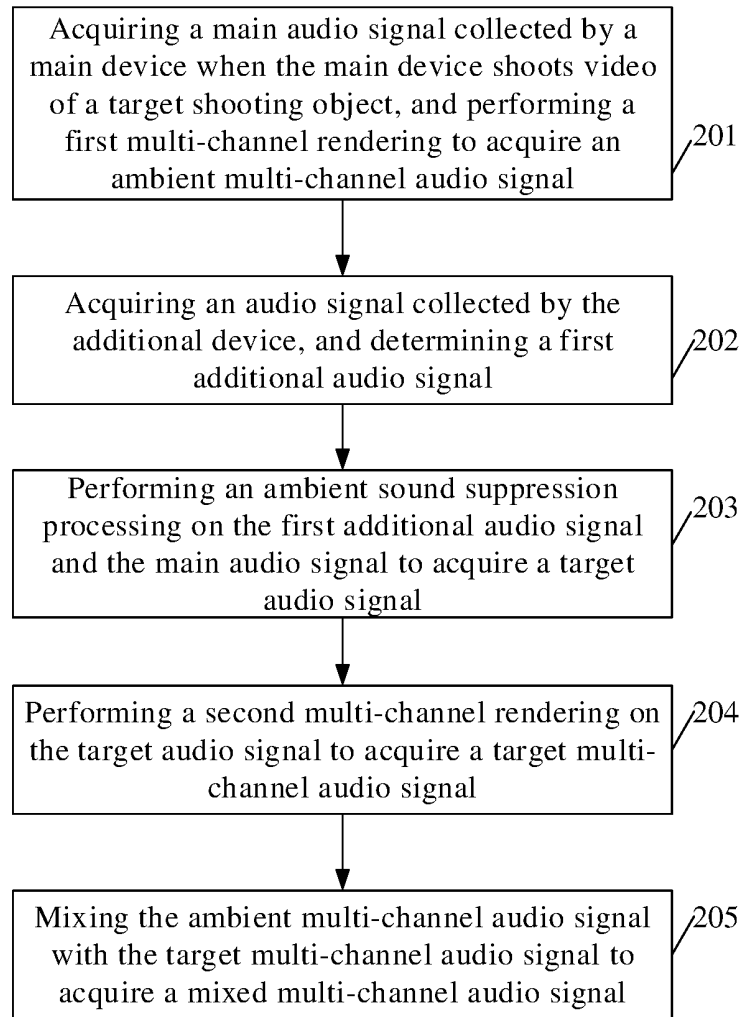


FIG. 2A

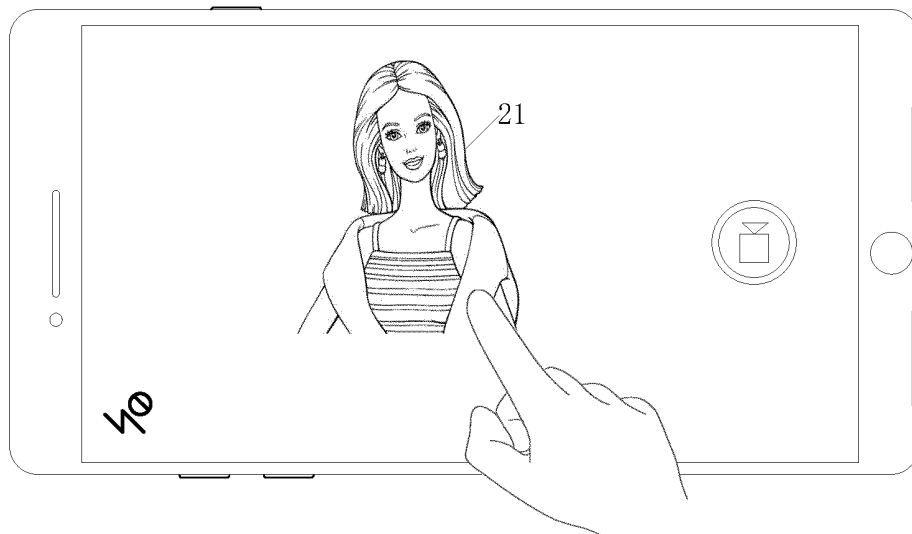


FIG. 2B

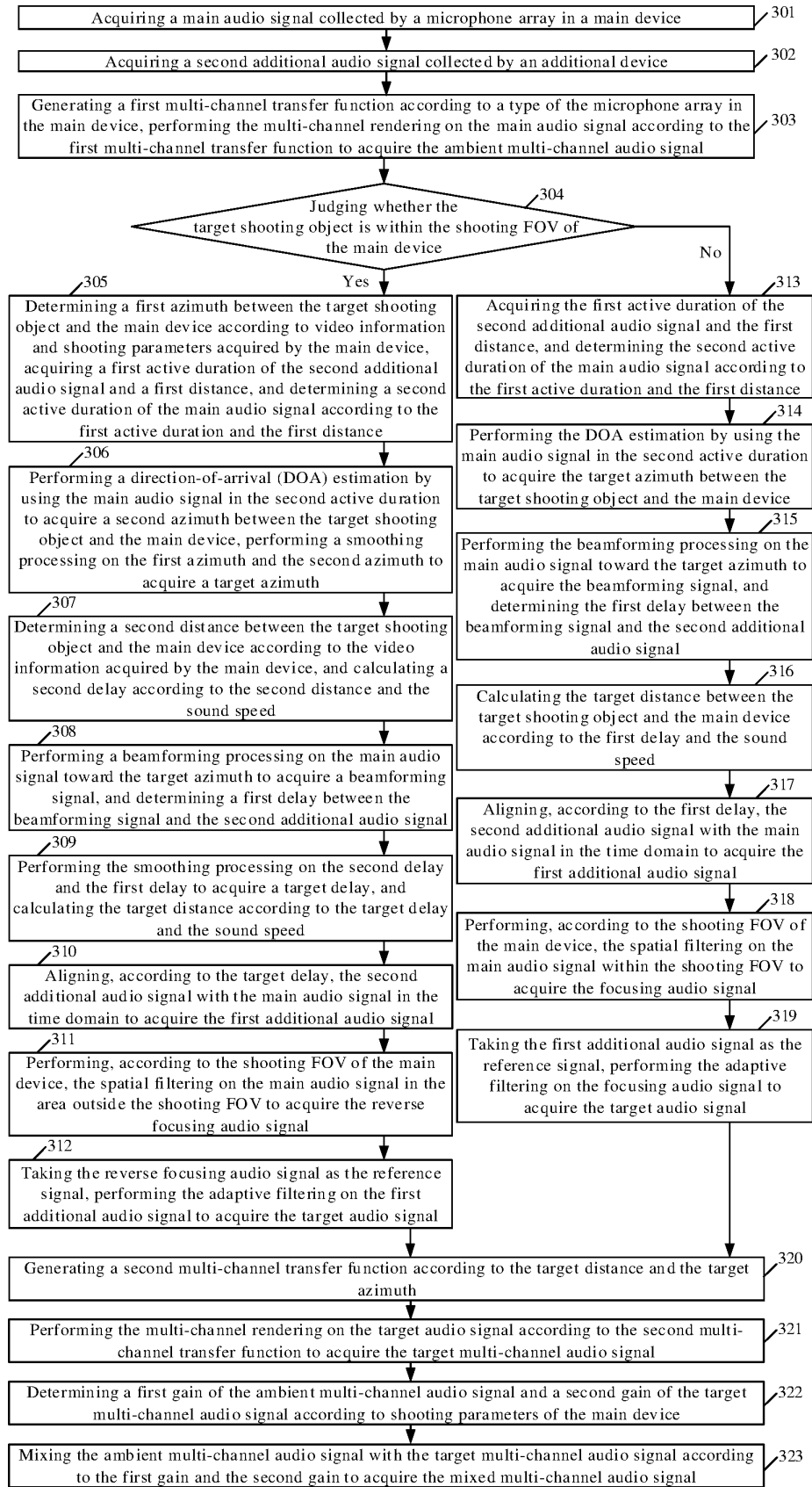


FIG. 3

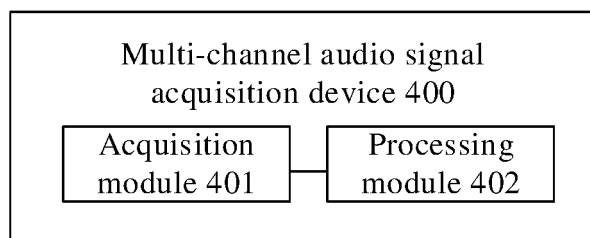


FIG. 4

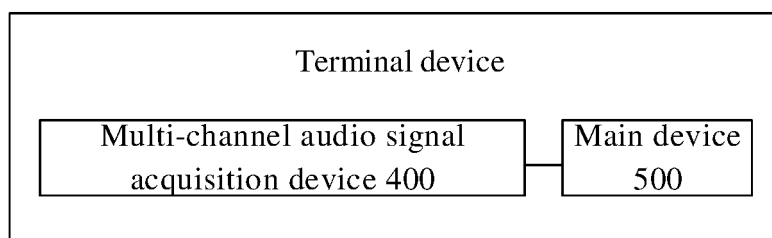


FIG. 5

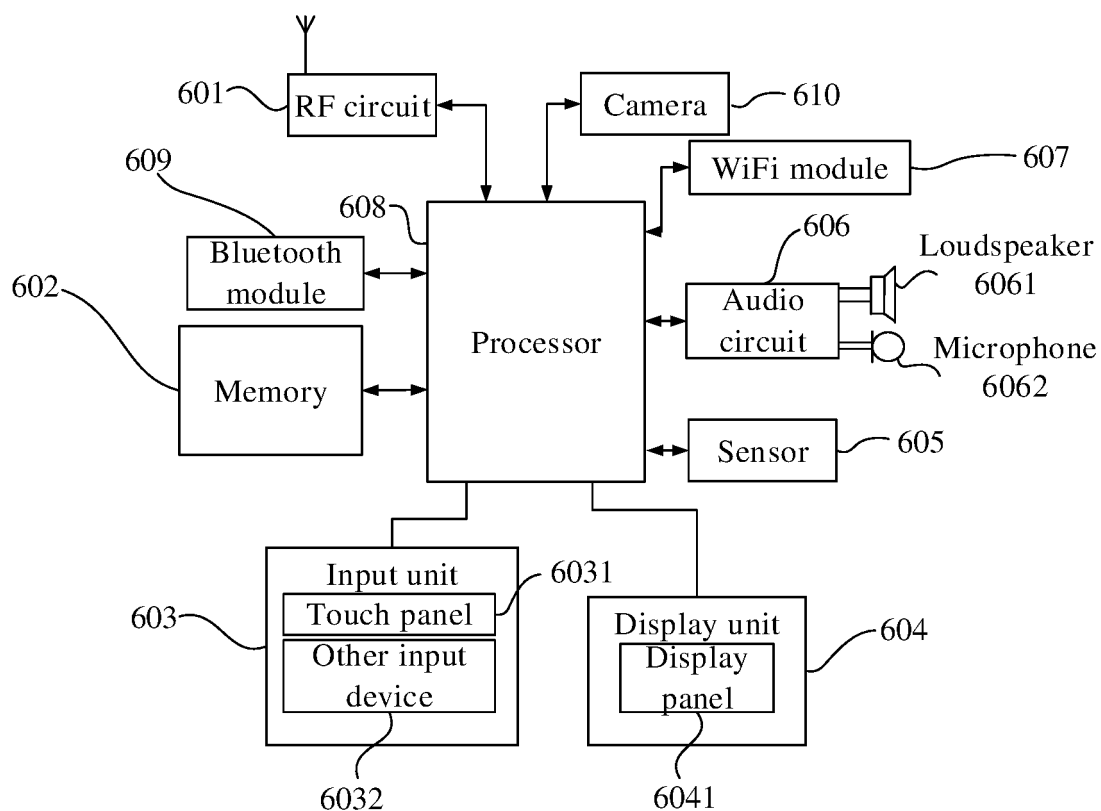


FIG. 6

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2021/103110

A. CLASSIFICATION OF SUBJECT MATTER

G10L 21/0208(2013.01)i; H04R 3/00(2006.01)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)

G10L21/-;H04R3/00

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)

EPTXT; USTXT; VEN; WOTXT; CNABS; CNTXT; CNKI; IEEE; OPPO, 王文东, 视频, 录像, 摄影, 摄像, 拍摄, 拍照, 录音, 音频, 声音, 主音频, 第一, 第二, 附加音频, 渲染, 混音, 噪声, 噪音, 环境声, 背景声, 场景类别, 场景分析, 空间音频, 聚焦音频, 分布式, 麦克风阵列, 耳机, 穿戴, 附加, 从属, 近场麦克风, 参考信号, video+, recording?, audio, sound, auxiliary, additional, earphone?, second+, other, another, render+, mix+, distribut+, mic+

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 108370471 A (NOKIA TECHNOLOGY CO., LTD.) 03 August 2018 (2018-08-03) description paragraphs [0033], [0083], [0088-0110], [0114], [0157], [0216-0224], [0249-0254], figures 1, 3a, 3b, 9	1, 3-6, 15-16, 18-21, 30-32
Y	CN 108370471 A (NOKIA TECHNOLOGY CO., LTD.) 03 August 2018 (2018-08-03) description [0033], [0083], [0088-0110], [0114], [0157], [0216-0224], [0249-0254], figures 1, 9	2, 13-14, 17, 28-29
Y	CN 108389586 A (NINGBO SANGDENA ELECTRONIC TECHNOLOGY CO., LTD.) 10 August 2018 (2018-08-10) description, paragraph [0090]	2, 17
Y	US 2019222950 A1 (APPLE INC.) 18 July 2019 (2019-07-18) description paragraphs [0005-0006]	13-14, 28-29
A	CN 111050269 A (HUAWEI TECHNOLOGIES CO., LTD.) 21 April 2020 (2020-04-21) entire document	1-32
A	CN 110970057 A (HUAWEI TECHNOLOGIES CO., LTD.) 07 April 2020 (2020-04-07) entire document	1-32

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

* Special categories of cited documents:	"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
"A" document defining the general state of the art which is not considered to be of particular relevance	"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
"E" earlier application or patent but published on or after the international filing date	"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
"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)	"&" document member of the same patent family
"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	

Date of the actual completion of the international search

03 August 2021

Date of mailing of the international search report

10 September 2021

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing
100088
China

Authorized officer

Facsimile No. (86-10)62019451

Telephone No.

Form PCT/ISA/210 (second sheet) (January 2015)

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2021/103110

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	EP 3683794 A1 (NOKIA TECHNOLOGIES OY.) 22 July 2020 (2020-07-22) entire document	1-32
A	US 2017359467 A1 (NORRIS, Glen A. et al.) 14 December 2017 (2017-12-14) entire document	1-32
A	CN 110089131 A (NOKIA TECHNOLOGY CO., LTD.) 02 August 2019 (2019-08-02) entire document	1-32
A	CN 102969003 A (DONGGUAN YULONG COMMUNICATION TECHNOLOGY CO., LTD. et al.) 13 March 2013 (2013-03-13) entire document	1-32
A	CN 104599674 A (XI'AN QIANYI ENTERPRISE MANAGEMENT CONSULTATION CO., LTD.) 06 May 2015 (2015-05-06) entire document	1-32
A	CN 108352155 A (HEWLETT-PACKARD DEVELOPMENT COMPANY, L.P.) 31 July 2018 (2018-07-31) entire document	1-32

Form PCT/ISA/210 (second sheet) (January 2015)

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2021/103110

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN 108370471 A	03 August 2018	EP 3363212 A4	03 July 2019
		GB 2543276 A	19 April 2017
		GB 201518025 D0	25 November 2015
		US 2018310114 A1	25 October 2018
		EP 3363212 A1	22 August 2018
		WO 2017064368 A1	20 April 2017
		US 10397722 B2	27 August 2019
CN 108389586 A	10 August 2018	None	
US 2019222950 A1	18 July 2019	US 10178490 B1	08 January 2019
		US 2019007780 A1	03 January 2019
		US 10848889 B2	24 November 2020
CN 111050269 A	21 April 2020	WO 2020078237 A1	23 April 2020
CN 110970057 A	07 April 2020	WO 2020062900 A1	02 April 2020
EP 3683794 A1	22 July 2020	WO 2020148109 A1	23 July 2020
US 2017359467 A1	14 December 2017	US 2019387102 A1	19 December 2019
		US 2021092232 A1	25 March 2021
		US 10225410 B2	05 March 2019
		US 2019174002 A1	06 June 2019
		US 10863032 B2	08 December 2020
		US 2018227426 A1	09 August 2018
		US 9998606 B2	12 June 2018
		US 10432796 B2	01 October 2019
CN 110089131 A	02 August 2019	US 2019349677 A1	14 November 2019
		EP 3542549 A4	08 July 2020
		US 10785565 B2	22 September 2020
		WO 2018091777 A1	24 May 2018
		EP 3542549 A1	25 September 2019
		GB 2556058 A	23 May 2018
		CN 110089131 B	13 July 2021
CN 102969003 A	13 March 2013	None	
CN 104599674 A	06 May 2015	None	
CN 108352155 A	31 July 2018	EP 3342187 A4	08 May 2019
		WO 2017058192 A1	06 April 2017
		US 10616681 B2	07 April 2020
		EP 3342187 A1	04 July 2018
		US 2018220231 A1	02 August 2018

Form PCT/ISA/210 (patent family annex) (January 2015)