encoding component 110, and a channel encoding component 132. The collection component 131 is connected to the encoding component 110, and the encoding component 110 is connected to the channel encoding component 132.

[0082] Optionally, the mobile terminal 140 may include an audio playing component 141, a decoding and rendering component 120, and a channel decoding component 142. The audio playing component 141 is connected to the decoding and rendering component 120, and the decoding and rendering component 120 is connected to the channel decoding component 142.

[0083] After collecting an audio signal through the collection component 131, the mobile terminal 130 encodes the audio signal through the encoding component 110, to obtain an audio signal encoded bitstream; and then encodes the audio signal encoded bitstream through the channel encoding component 132, to obtain a transmission signal.

[0084] The mobile terminal 130 sends the transmission signal to the mobile terminal 140 through the wireless or wired network.

[0085] After receiving the transmission signal, the mobile terminal 140 decodes the transmission signal through the channel decoding component 142, to obtain the audio signal encoded bitstream; decodes the audio signal encoded bitstream through the decoding and rendering component 120, to obtain a to-be-processed audio signal, and renders the to-be-processed audio signal through the decoding and rendering component 120, to obtain a rendered audio signal; and plays the rendered audio signal through the audio playing component 141. It may be understood that the mobile terminal 130 may alternatively include the components included in the mobile terminal 140, and the mobile terminal 140 may alternatively include the components included in the mobile terminal 130.

[0086] In addition, the mobile terminal 140 may further include an audio playing component, a decoding component, a rendering component, and a channel decoding component. The channel decoding component is connected to the decoding component, the decoding component is connected to the rendering component, and the rendering component is connected to the audio playing component. In this case, after receiving the transmission signal, the mobile terminal 140 decodes the transmission signal through the channel decoding component, to obtain the audio signal encoded bitstream; decodes the audio signal encoded bitstream through the decoding component, to obtain a to-be-processed audio signal; renders the to-be-processed audio signal through the rendering component, to obtain a rendered audio signal; and plays the rendered audio

signal through the audio playing component.

[0087] FIG. 3 is a structural block diagram of an audio signal receiving apparatus according to an embodiment of this application. Referring to FIG. 3, an audio signal receiving apparatus 20 in this embodiment of this application may include at least one processor 21, a memory 22, at least one communications bus 23, a receiver 24, and a transmitter 25. The communications bus 203 is used for connection and communication between the processor 21, the memory 22, the receiver 24, and the transmitter 25. The processor 21 may include a signal decoding component 211, a decoding component 212, and a rendering component 213.

[0088] Specifically, the memory 22 may be any one or any combination of the following storage media: a solid-state drive (Solid State Drives, SSD), a mechanical hard disk, a magnetic disk, a magnetic disk array, or the like, and can provide an instruction and data for the processor 21.

[0089] The memory 22 is configured to store the following data: correspondences between a plurality of preset positions and a plurality of HRTFs: (1) a plurality of positions relative to a left ear position, and HRTFs that are centered at the left ear position and that correspond to the positions relative to the left ear position; (2) a plurality of positions relative to a right ear position, and HRTFs that are centered at the right ear position and that correspond to the positions relative to the right ear position; (3) a plurality of positions relative to a head center, and HRTFs that are centered at the head center and that correspond to the positions relative to the head center.

[0090] Optionally, the memory 22 is further configured to store the following elements: an operating system and an application program module.

[0091] The operating system may include various system programs, and is configured to implement various basic services and process a hardware-based task. The application program module may include various application programs, and is configured to implement various application services.

[0092] The processor 21 may be a central processing unit (CPU), a general-purpose processor, a digital signal processor (DSP), an application-specific integrated circuit (ASIC), a field programmable gate array (FPGA) or another programmable logic device, a transistor logic device, a hardware component, or any combination thereof. The processor may implement or execute various example logical blocks, modules, and circuits described with reference to content disclosed in this application. The processor may alternatively be a combination of processors implementing

example, 0.1 m.

[0131] For example, a process of obtaining the HRTF 1 that is centered at the head center and that corresponds to the position c (100°, 50°, 1 m) is as follows: The first sound source 1 is placed at a position at which an azimuth relative to the head center is 100°, an elevation relative to the head center is 50°, and a distance from the head center is 1 m; and a corresponding HRTF that is used to transmit, to the head center 62, an audio signal sent by the first sound source 1 is measured, so as to obtain the HRTF 1 centered at the head center. The measurement method is an existing method, and details are not described herein.

[0132] For another example, a process of obtaining the HRTF 2 that is centered at the head center and that corresponds to the position d (100°, 45°, 1 m) is as follows: The first sound source 2 is placed at a position at which an azimuth relative to the head center is 100°, an elevation relative to the head center is 45°, and a distance from the head center is 1 m; and a corresponding HRTF that is used to transmit, to the head center 62, an audio signal sent by the first sound source 2 is measured, so as to obtain the HRTF 2 centered at the head center.

[0133] For another example, a process of obtaining the HRTF 3 that is centered at the head center and that corresponds to a position e (95°, 45°, 1 m) is as follows: A first sound source 3 is placed at a position at which an azimuth relative to the head center is 95°, an elevation relative to the head center is 45°, and a distance from the head center is 1 m; and a corresponding HRTF that is used to transmit, to the head center 62, an audio signal sent by the first sound source 3 is measured, so as to obtain the HRTF 3 centered at the head center.

[0134] For another example, a process of obtaining the HRTF 4 that is centered at the head center and that corresponds to a position f (95°, 50°, 1 m) is as follows: A first sound source 4 is placed at a position at which an azimuth relative to the head center is 95°, an elevation relative to the head center is 50°, and a distance from the head center is 1 m; and a corresponding HRTF that is used to transmit, to the head center 62, an audio signal sent by the first sound source 4 is measured, so as to obtain the HRTF 4 centered at the head center.

[0135] For another example, a process of obtaining the HRTF 5 that is centered at the head center and that corresponds to a position g (100°, 50°, 1.1 m) is as follows:A first sound source 5 is placed at a position at which an azimuth relative to the head center is 100°,an elevation relative to the head center is 50°, and a distance from the head center is 1.1 m; and a corresponding HRTF that is used to transmit,to the head center 62,an audio signal sent by the first sound source 5 is

relative to the left ear position 82 is 50°, and a distance from the left ear position 82 is 1 m; and a corresponding HRTF that is used to transmit, to the left ear position, an audio signal sent by the second sound source 1 is measured, so as to obtain the HRTF 1 centered at the left ear position.

[0167] For another example, a process of obtaining the HRTF 2 that is centered at the left ear position and that corresponds to the position d (100°, 45°, 1 m) is as follows: The second sound source 2 is placed at a position at which an azimuth relative to the left ear position 82 is 100°, an elevation relative to the left ear position 82 is 45°, and a distance from the left ear position 82 is 1 m; and a corresponding HRTF that is used to transmit, to the left ear position, an audio signal sent by the second sound source 2 is measured, so as to obtain the HRTF 2 centered at the left ear position.

[0168] For another example, a process of obtaining an HRTF 3 that is centered at the left ear position and that corresponds to a position e (95°, 50°, 1 m) is as follows: A second sound source 3 is placed at a position at which an azimuth relative to the left ear position 82 is 95°, an elevation relative to the left ear position 82 is 50°, and a distance from the left ear position 82 is 1 m; and a corresponding HRTF that is used to transmit, to the left ear position, an audio signal sent by the second sound source 3 is measured, so as to obtain the HRTF 3 centered at the left ear position.

[0169] For another example, a process of obtaining an HRTF 4 that is centered at the left ear position and that corresponds to a position f (95°, 45°, 1 m) is as follows: A second sound source 4 is placed at a position at which an azimuth relative to the left ear position 82 is 95°, an elevation relative to the left ear position 82 is 40°, and a distance from the left ear position 82 is 1 m; and a corresponding HRTF that is used to transmit, to the left ear position, an audio signal sent by the second sound source 4 is measured, so as to obtain the HRTF 4 centered at the left ear position.

[0170] For another example, a process of obtaining an HRTF 5 that is centered at the left ear position and that corresponds to a position g (100°, 50°, 1.2 m) is as follows: A second sound source 5 is placed at a position at which an azimuth relative to the left ear position 82 is 100°, an elevation relative to the left ear position 82 is 50°, and a distance from the left ear position 82 is 1.2 m; and a corresponding HRTF that is used to transmit, to the left ear position, an audio signal sent by the second sound source 5 is measured, so as to obtain the HRTF 5 centered at the left ear position.

[0171] For another example, a process of obtaining an HRTF 6 that is centered at the left ear position and that corresponds to a position h (95°, 50°, 1.1 m) is as follows: A second sound source

6 is placed at a position at which an azimuth relative to the left ear position 82 is 95°, an elevation relative to the left ear position 82 is 50°, and a distance from the left ear position 82 is 1.1 m; and a corresponding HRTF that is used to transmit, to the left ear position, an audio signal sent by the second sound source 6 is measured, so as to obtain the HRTF 6 centered at the left ear position.

[0172] It may be understood that, an azimuth ranges from –180° to 180°, and an elevation ranges from –90° to 90°. In this case, if the first angle is 5°, the second angle is 5°, the first distance is 0.1 m, and a total distance is 2 m, 72 x 36 x 21 HRTFs centered at the left ear position may be obtained.

[0173] According to the foregoing method, correspondences between a plurality of positions and a plurality of HRTFs centered at the left ear position may be obtained through measurement. It may be understood that, during measurement of the HRTFs centered at the left ear position, the plurality of positions at which the second sound sources are placed may be referred to as preset positions. Therefore, according to the foregoing method, the correspondences between the plurality of preset positions and the plurality of HRTFs centered at the left ear position may be obtained through measurement. The correspondences may be referred to as first correspondences, and the first correspondences may be stored in the memory 22 shown in FIG. 3.

[0174] Then, the determining, based on the M first positions and first correspondences, that M HRTFs corresponding to the M first positions are the M first HRTFs, where the first correspondences are prestored correspondences between a plurality of preset positions and a plurality of HRTFs centered at the left ear position includes:

determining M first preset positions associated with the M first positions, where the M first preset positions are preset positions in the first correspondences; and

determining, based on the first correspondences, that M HRTFs that are centered at the left ear position and that correspond to the M first preset positions are the M first HRTFs. The M HRTFs centered at the left ear position are actually M HRTFs that are centered at the left ear position 82 and that are used to transmit, to the left ear position 82, audio signals sent by sound sources at the M first preset positions.

[0175] The first preset position associated with the first position may be the first position; or

an elevation included in the first preset position is a target elevation that is closest to a third elevation included in the first position, an azimuth included in the first preset position is a target azimuth that is closest to a third azimuth included in the first position, and a distance

value is 6°, the fourth position is (58°, 73°, 0.5 m).

[0187] The three-dimensional coordinate system in this embodiment is the three-dimensional coordinate system corresponding to the foregoing audio signal receive end.

[0188] In step S303, before step S303, correspondences between a plurality of preset positions and a plurality of HRTFs centered at the head center need to be obtained in advance. For a method for obtaining the correspondences between a plurality of preset positions and a plurality of HRTFs centered at the head center, refer to the descriptions in the embodiment shown in FIG. 4. Details are not described again in this embodiment.

[0189] The determining, based on the M fourth positions and second correspondences, that M HRTFs corresponding to the M fourth positions are the M first HRTFs, where the second correspondences are prestored correspondences between a plurality of preset positions and a plurality of HRTFs centered at the head center includes:

determining, based on the M fourth positions, M second preset positions associated with the M fourth positions, where the M second preset positions are preset positions in the prestored second correspondences; and

determining, based on the second correspondences, that HRTFs that are centered at the head center and that correspond to the M second preset positions are the M first HRTFs.

[0190] Specifically, the second preset position associated with the fourth position may be the fourth position; or

an elevation included in the second preset position is a target elevation that is closest to an elevation included in the fourth position, an azimuth included in the second preset position is a target azimuth that is closest to an azimuth included in the fourth position, and a distance included in the second preset position is a target distance that is closest to a distance included in the fourth position. The target azimuth is an azimuth included in a corresponding preset position during measurement of the HRTF centered at the head center, namely, an azimuth of a placed first sound source relative to the head center during measurement of the HRTF centered at the head center. The target elevation is an elevation in a corresponding preset position during measurement of the HRTF centered at the head center, namely, an elevation of the placed first sound source relative to the head center during measurement of the HRTF centered at the head center. The target distance is a distance in a corresponding preset position during measurement of the HRTF centered at the head center, namely, a distance between the placed first sound source and the head center

[0196] Step S402: Determine M seventh positions based on the M third positions, where the M third positions are in a one-to-one correspondence with the M seventh positions, one seventh position and a corresponding third position include a same elevation and a same distance, a difference between an azimuth included in the one seventh position and a first preset value is a first azimuth included in the corresponding third position, where the correspondences are prestored correspondences between a plurality of preset positions and a plurality of HRTFs centered at the head center.

[0197] Step S403: Determine, based on the M seventh positions and second correspondences, that M HRTFs corresponding to the M seventh positions are the M first HRTFs, where the second correspondences are prestored correspondences between a plurality of preset positions and a plurality of HRTFs centered at the head center.

[0198] Specifically, for step S401 in this embodiment, refer to step S301 in the embodiment shown in FIG. 8. Details are not described herein again.

[0199] In step S402, a three-dimensional coordinate system in this embodiment is the three-dimensional coordinate system corresponding to the foregoing audio signal receive end.

[0200] Specifically, for each third position, a second elevation included in the third position is used as an elevation included in a corresponding seventh position, a second distance included in the third position is used as a distance included in the corresponding seventh position, and a second azimuth included in the third position plus the first preset value is an azimuth included in the corresponding seventh position. For example, if the third position is (52°, 73°, 0.5 m), and the first preset value is 5°, the seventh position is (57°, 73°, 0.5 m).

[0201] The first preset value is a preset value without consideration of a size of the head of a listener. In the foregoing embodiment, the first value is the difference between the first included angle and the second included angle, and this considers a size of the head of a current listener. Optionally, the first preset value is the same as the first preset angle in the embodiment shown in FIG. 4.

[0202] In step S403, before step S403, correspondences between a plurality of preset positions and a plurality of HRTFs centered at the head center need to be obtained in advance. For a method for obtaining the correspondences between a plurality of preset positions and a plurality of HRTFs centered at the head center, refer to the descriptions in the embodiment shown in FIG. 4. Details are not described again in this embodiment.

third sound source 3 is measured, so as to obtain the HRTF 3 centered at the right ear position.

[0219] For another example, a process of obtaining an HRTF 4 that is centered at the right ear position and that corresponds to a position f (95°, 45°, 1 m) is as follows: A third sound source 4 is placed at a position at which an azimuth relative to the right ear position is 95°, an elevation relative to the right ear position is 45°, and a distance from the right ear position is 1 m; and a corresponding HRTF that is used to transmit, to the right ear position, an audio signal sent by the third sound source 4 is measured, so as to obtain the HRTF 4 centered at the right ear position.

[0220] For another example, a process of obtaining an HRTF 5 that is centered at the right ear position and that corresponds to a position g (100°, 50°, 1.2 m) is as follows: A third sound source 5 is placed at a position at which an azimuth relative to the right ear position is 100°, an elevation relative to the right ear position is 50°, and a distance from the right ear position is 1.2 m; and a corresponding HRTF that is used to transmit, to the right ear position, an audio signal sent by the third sound source 5 is measured, so as to obtain the HRTF 5 centered at the right ear position.

[0221] For another example, a process of obtaining an HRTF 6 that is centered at the right ear position and that corresponds to a position h (95°, 50°, 1.1 m) is as follows: A third sound source 6 is placed at a position at which an azimuth relative to the right ear position is 95°, an elevation relative to the right ear position is 50°, and a distance from the right ear position is 1.1 m; and a corresponding HRTF that is used to transmit, to the right ear position, an audio signal sent by the third sound source 6 is measured, so as to obtain the HRTF 6 centered at the right ear position.

[0222] It may be understood that, an azimuth ranges from –180° to 180°, and an elevation ranges from –90° to 90°. In this case, if the first preset angle is 5°, the second preset angle is 5°, the first preset distance is 0.1 m, and a total distance is 2 m, 72 x 36 x 21 HRTFs centered at the right ear position may be obtained.

[0223] According to the foregoing method, correspondences between a plurality of positions and a plurality of HRTFs centered at the right ear position may be obtained through measurement. It may be understood that, during measurement of the HRTFs centered at the right ear position, the plurality of positions at which the third sound sources are placed may be referred to as preset positions. Therefore, according to the foregoing method, the correspondences between the plurality of preset positions and the plurality of HRTFs centered at the right ear position may be obtained through measurement. The correspondences are referred to as third correspondences, and the third correspondences may be stored in the memory 22 shown in FIG. 3.

descriptions about the first preset position associated with the first position. If the fourth elevation included in the second position is between two target elevations, for a method for determining the elevation included in the fourth preset position, refer to the descriptions about the first preset position associated with the first position. Details are not described herein again.

[0227] For example, if in the second position, obtained through measurement in step S501, of an n^{th} second virtual speaker relative to the right ear position, the fourth azimuth is 88°, the fourth elevation is 46°, and the fourth distance is 1.02 m, the correspondences between the plurality of preset positions and the plurality of HRTFs centered at the right ear position include an HRTF that is centered at the right ear position and that corresponds to a position (90°, 45°, 1 m), an HRTF that is centered at the right ear position and that corresponds to a position (85°, 45°, 1 m), an HRTF that is centered at the right ear position and that corresponds to a position (90°, 50°, 1 m), an HRTF that is centered at the right ear position and that corresponds to a position (85°, 50°, 1 m), an HRTF that is centered at the right ear position and that corresponds to a position (90°, 45°, 1.1 m), an HRTF that is centered at the right ear position and that corresponds to a position (85°, 45°, 1.1 m), an HRTF that is centered at the right ear position and that corresponds to a position (90°, 50°, 1.1 m), and an HRTF that is centered at the right ear position and that corresponds to a position (85°, 50°, 1.1 m). 88° is between 85° and 90°, but is closer to 90°, 46° is between 45° and 50°, but is closer to 45°, and 1.02 m is between 1 m and 1.1 m, but is closer to 1 m. Therefore, it is determined that the position (90°, 45°, 1 m) is a fourth preset position n associated with the second position of the n^{th} second virtual speaker relative to the right ear position.

[0228] After the N fourth preset positions associated with the N second positions are determined, it is determined that the N HRTFs that are centered at the right ear position and that correspond to the N fourth preset positions are the N second HRTFs. For example, in the foregoing examples, based on the third correspondences, the HRTF that is centered at the right ear position and that corresponds to the position (90°, 45°, 1 m) is an HRTF that is centered at the right ear position and that corresponds to the second position of the n^{th} second virtual speaker relative to the right ear position. In other words, based on the third correspondences, the HRTF that is centered at the right ear position and that corresponds to the fourth preset position n (90°, 45°, 1 m) is an n^{th} second HRTF, or a second HRTF corresponding to the n^{th} second virtual speaker.

[0229] In this embodiment, the N second HRTFs are N HRTFs that are centered at the right ear position and that are obtained through actual measurement. The obtained N second HRTFs can

the second virtual speaker and the current head center.

[0236] In step S602, for each fifth position, a second elevation included in the fifth position is used as an elevation included in a corresponding sixth position, a second distance included in the fifth position is used as a distance included in the corresponding sixth position, and a second azimuth included in the fifth position minus the second value is an azimuth included in corresponding M sixth positions. For example, if the fifth position is (52°, 73°, 0.5 m), and the second value is 6°, the sixth position is (46°, 73°, 0.5 m).

[0237] The three-dimensional coordinate system in this embodiment is the three-dimensional coordinate system corresponding to the foregoing audio signal receive end.

[0238] In step S603, before step S603, correspondences between a plurality of preset positions and a plurality of HRTFs centered at the head center need to be obtained in advance. For a method for obtaining the correspondences between a plurality of preset positions and a plurality of HRTFs centered at the head center, refer to the descriptions in the embodiment shown in FIG. 4. Details are not described again in this embodiment.

[0239] The determining, based on the N sixth positions and second correspondences, that N HRTFs corresponding to the N sixth positions are the N second HRTFs, where the second correspondences are prestored correspondences between a plurality of preset positions and a plurality of HRTFs centered at the head center includes:

determining N fifth preset positions based on the N sixth positions, where the N fifth preset positions are preset positions in the second correspondences; and

determining, based on the second correspondences, that N HRTFs that are centered at the head center and that correspond to the N fifth preset positions are the N second HRTFs.

[0240] For the fifth preset position associated with the sixth position, refer to the explanation of the second preset position associated with the fourth position. Details are not described herein again.

[0241] After the N fifth preset positions associated with the N sixth positions are determined, it is determined that the N HRTFs that are centered at the head center and that correspond to the N fifth preset positions are the N second HRTFs. For example, if a fifth preset position associated with a sixth position is (40°, 60°, 0.5 m), based on the second correspondences, an HRTF that is centered at the head center and that corresponds to the position (40°, 60°, 0.5 m) is an HRTF that is centered at the head center and that corresponds to the sixth position. In other words, based on

the second correspondences, the HRTF that is centered at the head center and that corresponds to the position (40°, 60°, 0.5 m) is one second HRTF in the N second HRTFs.

[0242] In this embodiment, the N second HRTFs are converted from HRTFs centered at the head center, and efficiency of obtaining the second HRTFs is comparatively high.

[0243] Next, a third process of obtaining N second HRTFs in step S102 in the embodiment shown in FIG. 4 is described. FIG. 13 is a flowchart 7 of an audio processing method according to an embodiment of this application. Referring to FIG. 13, the method in this embodiment includes the following steps.

[0244] Step S701: Obtain N fifth positions of N second virtual speakers relative to a current head center, where the fifth position includes a second azimuth and a second elevation of the second virtual speaker relative to the current head center, and includes a second distance between the current head center and the second virtual speaker.

[0245] Step S702: Determine N eighth positions based on the N fifth positions, where the N fifth positions are in a one-to-one correspondence with the N eighth positions, one eighth position and a corresponding fifth position include a same elevation and a same distance, and a sum of an azimuth included in the one eighth position and a first preset value is a second azimuth included in the corresponding fifth position.

[0246] Step S703: Determine, based on the N eighth positions and second correspondences, that N HRTFs corresponding to the N eighth positions are the N second HRTFs, where the second correspondences are prestored correspondences between a plurality of preset positions and a plurality of HRTFs centered at the head center.

[0247] Specifically, for step S701 in this embodiment, refer to step S601 in the embodiment in FIG. 12. Details are not described herein again.

[0248] In step S702, a three-dimensional coordinate system in this embodiment is the three-dimensional coordinate system corresponding to the foregoing audio signal receive end.

[0249] Specifically, for each fifth position, a second elevation included in the fifth position is used as an elevation included in a corresponding eighth position, a second distance included in the fifth position is used as a distance included in the corresponding eighth position, and a second azimuth included in the fifth position minus the first preset value is an azimuth included in the corresponding eighth position. For example, if the fifth position is (52°, 73°, 0.5 m), and the first preset value is 5°, the eighth position is (47°, 73°, 0.5 m).

[0302] The units described as separate parts may or may not be physically separate, and parts displayed as units may or may not be physical units, may be located in one position, or may be distributed on a plurality of network units. Some or all of the units may be selected based on an actual requirement to achieve the objectives of the solutions of the embodiments.

[0303] In addition, function units in the embodiments of this application may be integrated into one processing unit, or each of the units may exist alone physically, or two or more units are integrated into one unit. The integrated unit may be implemented in a form of hardware, or may be implemented in a form of hardware combined with a software functional unit.

[0304] The foregoing descriptions are merely specific implementations of the present disclosure, but are not intended to limit the protection scope of the present disclosure. Any variation or replacement readily figured out by a person skilled in the art within the technical scope disclosed in the present disclosure shall fall within the protection scope of the present disclosure. Therefore, the protection scope of the present disclosure shall be subject to the protection scope of the claims.

CLAIMS

What is claimed is:

1. An audio processing method, comprising:

processing, by M first virtual speakers, a to-be-processed audio signal to obtain M first audio signals;

processing, by N second virtual speakers, the to-be-processed audio signal to obtain N second audio signals, wherein the M first virtual speakers are in a one-to-one correspondence with the M first audio signals, the N second virtual speakers are in a one-to-one correspondence with the N second audio signals, and M and N are positive integers;

obtaining M first head-related transfer functions (HRTFs) and N second HRTFs, wherein all the M first HRTFs are centered at a left ear position, all the N second HRTFs are centered at a right ear position, the M first HRTFs are in a one-to-one correspondence with the M first virtual speakers, and the N second HRTFs are in a one-to-one correspondence with the N second virtual speakers;

obtaining a first target audio signal based on the M first audio signals and the M first HRTFs; and

obtaining a second target audio signal based on the N second audio signals and the N second HRTFs.

2. The audio processing method according to claim 1,

wherein obtaining the first target audio signal based on the M first audio signals and the M first HRTFs comprises:

convolving each of the M first audio signals with a corresponding first HRTF, to obtain M first convolved audio signals; and

obtaining the first target audio signal based on the M first convolved audio signals, or

wherein obtaining the second target audio signal based on the N second audio signals and the N second HRTFs comprises:

convolving each of the N second audio signals with a corresponding second HRTF, to obtain N second convolved audio signals; and

obtaining the second target audio signal based on the N second convolved audio signals.

3. The audio processing method according to claim 1,

wherein correspondences between a plurality of preset positions and a plurality of HRTFs are prestored, and

wherein:

obtaining the M first HRTFs or the N second HRTFs comprises:

obtaining M first positions of the M first virtual speakers relative to the current left ear position; and

determining, based on the M first positions and the correspondences, that M HRTFs corresponding to the M first positions are the M first HRTFs; or

obtaining the N second HRTFs comprises:

obtaining N second positions of the N second virtual speakers relative to the current right ear position; and

determining, based on the N second positions and the correspondences, that N HRTFs corresponding to the N second positions are the N second HRTFs.

4. The audio processing method according to claim 1,

wherein correspondences between a plurality of preset positions and a plurality of HRTFs are prestored, and

wherein:

obtaining the M first HRTFs or the N second HRTFs comprises:

obtaining M third positions of the M first virtual speakers relative to a current head center, wherein the third position comprises a first azimuth and a first elevation of the first virtual speaker relative to the current head center, and comprises a first distance between the current head center and the first virtual speaker;

determining M fourth positions based on the M third positions, wherein the M third positions are in a one-to-one correspondence with the M fourth positions, one fourth position and a corresponding third position comprise a same elevation and a same distance, and a difference between an azimuth comprised in the one

fourth position and a first value is a first azimuth comprised in the corresponding third position, and the first value is a difference between a first included angle and a second included angle, the first included angle is an included angle between a first straight line and a first plane, the second included angle is an included angle between a second straight line and the first plane, the first straight line is a straight line that passes through the current left ear and a coordinate origin of a three-dimensional coordinate system, the second straight line is a straight line that passes through the current head center and the coordinate origin, and the first plane is a plane constituted by an X axis and a Z axis of the three-dimensional coordinate system; and

determining, based on the M fourth positions and the correspondences, that M HRTFs corresponding to the M fourth positions are the M first HRTFs; or

obtaining the N second HRTFs comprises:

obtaining N fifth positions of the N second virtual speakers relative to the current head center, wherein the fifth position comprises a second azimuth and a second elevation of the second virtual speaker relative to the current head center, and comprises a second distance between the current head center and the second virtual speaker;

determining N sixth positions based on the N fifth positions, wherein the N fifth positions are in a one-to-one correspondence with the N sixth positions, one sixth position and a corresponding fifth position comprise a same elevation and a same distance, and a sum of an azimuth comprised in the one sixth position and a second value is a second azimuth comprised in the corresponding fifth position, and the second value is a difference between a third included angle and a second included angle, the second included angle is an included angle between a second straight line and a first plane, the third included angle is an included angle between a third straight line and the first plane, the second straight line is the straight line that passes through the current head center and the coordinate origin, the third straight line is a straight line that passes through the current right ear and the coordinate origin, and the first plane is the plane constituted by the X axis and the Z axis of the three-dimensional coordinate system; and

determining, based on the N sixth positions and the correspondences, that N HRTFs corresponding to the N sixth positions are the N second HRTFs.

5. The audio processing method according to claim 1,

wherein correspondences between a plurality of preset positions and a plurality of HRTFs are prestored, and

wherein:

obtaining the M first HRTFs or the N second HRTFs comprises:

obtaining M third positions of the M first virtual speakers relative to a current head center, wherein the third position comprises a first azimuth and a first elevation of the first virtual speaker relative to the current head center, and comprises a first distance between the current head center and the first virtual speaker;

determining M seventh positions based on the M third positions, wherein the M third positions are in a one-to-one correspondence with the M seventh positions, one seventh position and a corresponding third position comprise a same elevation and a same distance, and a difference between an azimuth comprised in the one seventh position and a first preset value is a first azimuth comprised in the corresponding third position; and

determining, based on the M seventh positions and the correspondences, that M HRTFs corresponding to the M seventh positions are the M first HRTFs; or

obtaining the N second HRTFs comprises:

obtaining N fifth positions of the N second virtual speakers relative to the current head center, wherein the fifth position comprises a second azimuth and a second elevation of the second virtual speaker relative to the current head center, and comprises a second distance between the current head center and the second virtual speaker;

determining N eighth positions based on the N fifth positions, wherein the N fifth positions are in a one-to-one correspondence with the N eighth positions, one eighth position and a corresponding fifth position comprise a same elevation and a same distance, and a sum of an azimuth comprised in the one eighth position

and the first preset value is a second azimuth comprised in the corresponding fifth position; and

determining, based on the N eighth positions and the correspondences, that N HRTFs corresponding to the N eighth positions are the N second HRTFs.

6. The audio processing method according to claim 1,

wherein before obtaining the M first audio signals by processing the to-be-processed audio signal by the M first virtual speakers, the audio processing method further comprises:

obtaining a target virtual speaker group, wherein the target virtual speaker group comprises M target virtual speakers, and the M target virtual speakers are in a one-to-one correspondence with the M first virtual speakers; and

determining M tenth positions of the M first virtual speakers relative to a coordinate origin of a three-dimensional coordinate system based on M ninth positions of the M target virtual speakers relative to the coordinate origin, wherein the M ninth positions are in a one-to-one correspondence with the M tenth positions, one tenth position and a corresponding ninth position comprise a same elevation and a same distance, and a difference between an azimuth comprised in the one tenth position and a second preset value is an azimuth comprised in the corresponding ninth position, and

wherein obtaining the M first audio signals by processing the to-be-processed audio signal by the M first virtual speakers comprises processing the to-be-processed audio signal based on the M tenth positions to obtain the M first audio signals.

7. The audio processing method according to claim 1,

wherein M = N,

wherein before obtaining the N second audio signals by processing the to-be-processed audio signal by the N second virtual speakers, the audio processing method further comprises:

obtaining a target virtual speaker group, wherein the target virtual speaker group comprises M target virtual speakers, and the M target virtual speakers are in a one-to-one correspondence with the N second virtual speakers; and

determining N eleventh positions of the N second virtual speakers relative to a coordinate origin of a three-dimensional coordinate system based on M ninth positions of

the M target virtual speakers relative to the coordinate origin, wherein the M ninth positions are in a one-to-one correspondence with the N eleventh positions, one eleventh position and a corresponding ninth position comprise a same elevation and a same distance, and a sum of an azimuth comprised in the one eleventh position and a second preset value is an azimuth comprised in the corresponding ninth position, and

wherein obtaining the N second audio signals by processing the to-be-processed audio signal by the N second virtual speakers comprises processing the to-be-processed audio signal based on the N eleventh positions to obtain the N second audio signals.

8. The audio processing method according to claim 1, wherein the M first virtual speakers are speakers in a first speaker group and the N second virtual speakers are speakers in a second speaker group, wherein M = N, and wherein either:

the first speaker group and the second speaker group are two independent speaker groups; or

the first speaker group and the second speaker group are a same speaker group.

9. An audio processing apparatus, comprising:

at least one processor; and

a memory coupled to the at least one processor and storing computer executable instructions for execution by the at least one processor, wherein the computer executable instructions instruct the at least one processor to:

process, by M first virtual speakers, a to-be-processed audio signal to obtain M first audio signals;–

process, by N second virtual speakers, the to-be-processed audio signal to obtain N second audio signals, wherein the M first virtual speakers are in a one-to-one correspondence with the M first audio signals, the N second virtual speakers are in a one-to-one correspondence with the N second audio signals, and M and N are positive integers;

obtain M first head-related transfer functions (HRTFs) and N second HRTFs, wherein all the M first HRTFs are centered at a left ear position, all the N second HRTFs are centered at a right ear position, the M first HRTFs are in a one-to-one correspondence

with the M first virtual speakers, and the N second HRTFs are in a one-to-one correspondence with the N second virtual speakers;

obtain a first target audio signal based on the M first audio signals and the M first HRTFs; and

obtain a second target audio signal based on the N second audio signals and the N second HRTFs.

10. The audio processing apparatus according to claim 9, wherein the computer executable instructions further instruct the at least one processor to:

convolve each of the M first audio signals with a corresponding first HRTF to obtain M first convolved audio signals, and obtain the first target audio signal based on the M first convolved audio signals; or

convolve each of the N second audio signals with a corresponding second HRTF, to obtain N second convolved audio signals, and obtain the second target audio signal based on the N second convolved audio signals.

11. The audio processing apparatus according to claim 9, wherein the computer executable instructions further instruct the at least one processor to:

obtain M first positions of the M first virtual speakers relative to the current left ear position; and determine, based on the M first positions and correspondences, that M HRTFs corresponding to the M first positions are the M first HRTFs, wherein the correspondences are prestored correspondences between a plurality of preset positions and a plurality of HRTFs; or

obtain N second positions of the N second virtual speakers relative to the current right ear position; and determine, based on the N second positions and correspondences, that N HRTFs corresponding to the N second positions are the N second HRTFs, wherein the correspondences are prestored correspondences between a plurality of preset positions and a plurality of HRTFs.

12. The audio processing apparatus according to claim 9, wherein the computer executable instructions further instruct the at least one processor to:

obtain M third positions of the M first virtual speakers relative to a current head center, wherein the third position comprises a first azimuth and a first elevation of the first virtual speaker

relative to the current head center, and comprises a first distance between the current head center and the first virtual speaker; determine M fourth positions based on the M third positions, wherein the M third positions are in a one-to-one correspondence with the M fourth positions, one fourth position and a corresponding third position comprise a same elevation and a same distance, and a difference between an azimuth comprised in the one fourth position and a first value is a first azimuth comprised in the corresponding third position, and the first value is a difference between a first included angle and a second included angle, the first included angle is an included angle between a first straight line and a first plane, the second included angle is an included angle between a second straight line and the first plane, the first straight line is a straight line that passes through the current left ear and a coordinate origin of a three-dimensional coordinate system, the second straight line is a straight line that passes through the current head center and the coordinate origin, and the first plane is a plane constituted by an X axis and a Z axis of the three-dimensional coordinate system; and determine, based on the M fourth positions and correspondences, that M HRTFs corresponding to the M fourth positions are the M first HRTFs, wherein the correspondences are prestored correspondences between a plurality of preset positions and a plurality of HRTFs; or

obtain N fifth positions of the N second virtual speakers relative to the current head center, wherein the fifth position comprises a second azimuth and a second elevation of the second virtual speaker relative to the current head center, and comprises a second distance between the current head center and the second virtual speaker; determine N sixth positions based on the N fifth positions, wherein the N fifth positions are in a one-to-one correspondence with the N sixth positions, one sixth position and a corresponding fifth position comprise a same elevation and a same distance, and a sum of an azimuth comprised in the one sixth position and a second value is a second azimuth comprised in the corresponding fifth position, and the second value is a difference between a third included angle and a second included angle, the second included angle is an included angle between a second straight line and a first plane, the third included angle is an included angle between a third straight line and the first plane, the second straight line is the straight line that passes through the current head center and the coordinate origin, the third straight line is a straight line that passes through the current right ear and the coordinate origin, and the first plane is the plane constituted by the X axis and the Z axis of the three-dimensional coordinate system; and determine, based on the N sixth positions and correspondences, that N HRTFs

corresponding to the N sixth positions are the N second HRTFs, wherein the correspondences are prestored correspondences between a plurality of preset positions and a plurality of HRTFs.

13. The audio processing apparatus according to claim 9, wherein the computer executable instructions further instruct the at least one processor to:

obtain M third positions of the M first virtual speakers relative to a current head center, wherein the third position comprises a first azimuth and a first elevation of the first virtual speaker relative to the current head center, and comprises a first distance between the current head center and the first virtual speaker; determine M seventh positions based on the M third positions, wherein the M third positions are in a one-to-one correspondence with the M seventh positions, one seventh position and a corresponding third position comprise a same elevation and a same distance, and a difference between an azimuth comprised in the one seventh position and a first preset value is a first azimuth comprised in the corresponding third position; and determine, based on the M seventh positions and correspondences, that M HRTFs corresponding to the M seventh positions are the M first HRTFs, wherein the correspondences are prestored correspondences between a plurality of preset positions and a plurality of HRTFs; or

obtain N fifth positions of the N second virtual speakers relative to the current head center, wherein the fifth position comprises a second azimuth and a second elevation of the second virtual speaker relative to the current head center, and comprises a second distance between the current head center and the second virtual speaker; determine N eighth positions based on the N fifth positions, wherein the N fifth positions are in a one-to-one correspondence with the N eighth positions, one eighth position and a corresponding fifth position comprise a same elevation and a same distance, and a sum of an azimuth comprised in the one eighth position and the first preset value is a second azimuth comprised in the corresponding fifth position; and determine, based on the N eighth positions and correspondences, that N HRTFs corresponding to the N eighth positions are the N second HRTFs, wherein the correspondences are prestored correspondences between a plurality of preset positions and a plurality of HRTFs.

14. The audio processing apparatus according to claim 9, wherein the computer executable instructions further instruct the at least one processor to:

obtain a target virtual speaker group, wherein the target virtual speaker group comprises M target virtual speakers, and the M target virtual speakers are in a one-to-one correspondence with the M first virtual speakers;

determine M tenth positions of the M first virtual speakers relative to a coordinate origin of a three-dimensional coordinate system based on M ninth positions of the M target virtual speakers relative to the coordinate origin, wherein the M ninth positions are in a one-to-one correspondence with the M tenth positions, one tenth position and a corresponding ninth position comprise a same elevation and a same distance, and a difference between an azimuth comprised in the one tenth position and a second preset value is an azimuth comprised in the corresponding ninth position; and

process the to-be-processed audio signal based on the M tenth positions to obtain the M first audio signals.

15. The audio processing apparatus according to claim 9, wherein M = N, and the computer executable instructions further instruct the at least one processor to:

obtain a target virtual speaker group, wherein the target virtual speaker group comprises M target virtual speakers, and the M target virtual speakers are in a one-to-one correspondence with the N second virtual speakers;

determine N eleventh positions of the N second virtual speakers relative to a coordinate origin of a three-dimensional coordinate system based on the M ninth positions of the M target virtual speakers relative to the coordinate origin, wherein the M ninth positions are in a one-to-one correspondence with the N eleventh positions, one eleventh position and a corresponding ninth position comprise a same elevation and a same distance, and a sum of an azimuth comprised in the one eleventh position and a second preset value is an azimuth comprised in the corresponding ninth position; and

process the to-be-processed audio signal based on the N eleventh positions to obtain the N second audio signals.

16. The audio processing apparatus according to claim 9, wherein the M first virtual speakers are speakers in a first speaker group, the N second virtual speakers are speakers in a second speaker group, wherein M = N, and wherein either:

the first speaker group and the second speaker group are two independent speaker groups; or

the first speaker group and the second speaker group are a same speaker group.

17. A non-transitory computer-readable storage medium storing computer instructions that, when executed by one or more processors, cause the one or more processors to:

process, by M first virtual speakers, a to-be-processed audio signal to obtain M first audio signals;

process, by N second virtual speakers, the to-be-processed audio signal to obtain N second audio signals, wherein the M first virtual speakers are in a one-to-one correspondence with the M first audio signals, the N second virtual speakers are in a one-to-one correspondence with the N second audio signals, and M and N are positive integers;

obtain M first head-related transfer functions (HRTFs) and N second HRTFs, wherein all the M first HRTFs are centered at a left ear position, all the N second HRTFs are centered at a right ear position, the M first HRTFs are in a one-to-one correspondence with the M first virtual speakers, and the N second HRTFs are in a one-to-one correspondence with the N second virtual speakers;

obtain a first target audio signal based on the M first audio signals and the M first HRTFs; and

obtain a second target audio signal based on the N second audio signals and the N second HRTFs.

18. The non-transitory computer-readable storage medium according to claim 17, wherein the computer instructions further cause the one or more processors to:

convolve each of the M first audio signals with a corresponding first HRTF to obtain M first convolved audio signals, and obtain the first target audio signal based on the M first convolved audio signals; or

convolve each of the N second audio signals with a corresponding second HRTF to obtain N second convolved audio signals, and obtain the second target audio signal based on the N second convolved audio signals.

19. The non-transitory computer-readable storage medium according to claim 17, wherein the computer instructions further cause the one or more processors to:

obtain M first positions of the M first virtual speakers relative to the current left ear position; and determine, based on the M first positions and correspondences between a plurality of preset positions and a plurality of HRTFs, that M HRTFs corresponding to the M first positions are the M first HRTFs; or

obtain N second positions of the N second virtual speakers relative to the current right ear position; and determine, based on the N second positions and the correspondences, that N HRTFs corresponding to the N second positions are the N second HRTFs.

20. The non-transitory computer-readable storage medium according to claim 17, wherein the computer instructions further cause the one or more processors to:

obtain M third positions of the M first virtual speakers relative to a current head center, wherein the third position comprises a first azimuth and a first elevation of the first virtual speaker relative to the current head center, and comprises a first distance between the current head center and the first virtual speaker; determine M fourth positions based on the M third positions, wherein the M third positions are in a one-to-one correspondence with the M fourth positions, one fourth position and a corresponding third position comprise a same elevation and a same distance, and a difference between an azimuth comprised in the one fourth position and a first value is a first azimuth comprised in the corresponding third position, and the first value is a difference between a first included angle and a second included angle, the first included angle is an included angle between a first straight line and a first plane, the second included angle is an included angle between a second straight line and the first plane, the first straight line is a straight line that passes through the current left ear and a coordinate origin of a three-dimensional coordinate system, the second straight line is a straight line that passes through the current head center and the coordinate origin, and the first plane is a plane constituted by an X axis and a Z axis of the three-dimensional coordinate system; and determine, based on the M fourth positions and correspondences between a plurality of preset positions and a plurality of HRTFs, that M HRTFs corresponding to the M fourth positions are the M first HRTFs; or

obtain N fifth positions of the N second virtual speakers relative to the current head center, wherein the fifth position comprises a second azimuth and a second elevation of the second virtual

speaker relative to the current head center, and comprises a second distance between the current head center and the second virtual speaker; determine N sixth positions based on the N fifth positions, wherein the N fifth positions are in a one-to-one correspondence with the N sixth positions, one sixth position and a corresponding fifth position comprise a same elevation and a same distance, and a sum of an azimuth comprised in the one sixth position and a second value is a second azimuth comprised in the corresponding fifth position, and the second value is a difference between a third included angle and a second included angle, the second included angle is an included angle between a second straight line and a first plane, the third included angle is an included angle between a third straight line and the first plane, the second straight line is the straight line that passes through the current head center and the coordinate origin, the third straight line is a straight line that passes through the current right ear and the coordinate origin, and the first plane is the plane constituted by the X axis and the Z axis of the three-dimensional coordinate system; and determine, based on the N sixth positions and the correspondences, that N HRTFs corresponding to the N sixth positions are the N second HRTFs.

21. The non-transitory computer-readable storage medium according to claim 17, wherein the computer instructions further cause the one or more processors to:

obtain M third positions of the M first virtual speakers relative to a current head center, wherein the third position comprises a first azimuth and a first elevation of the first virtual speaker relative to the current head center, and comprises a first distance between the current head center and the first virtual speaker; determine M seventh positions based on the M third positions, wherein the M third positions are in a one-to-one correspondence with the M seventh positions, one seventh position and a corresponding third position comprise a same elevation and a same distance, and a difference between an azimuth comprised in the one seventh position and a first preset value is a first azimuth comprised in the corresponding third position; and determine, based on the M seventh positions and correspondences between a plurality of preset positions and a plurality of HRTFs, that M HRTFs corresponding to the M seventh positions are the M first HRTFs; or

obtain N fifth positions of the N second virtual speakers relative to the current head center, wherein the fifth position comprises a second azimuth and a second elevation of the second virtual speaker relative to the current head center, and comprises a second distance between the current head center and the second virtual speaker; determine N eighth positions based on the N fifth

positions, wherein the N fifth positions are in a one-to-one correspondence with the N eighth positions, one eighth position and a corresponding fifth position comprise a same elevation and a same distance, and a sum of an azimuth comprised in the one eighth position and the first preset value is a second azimuth comprised in the corresponding fifth position; and determine, based on the N eighth positions and the correspondences, that N HRTFs corresponding to the N eighth positions are the N second HRTFs.