



(11)

EP 4 542 542 A1

(12)

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

(43) Date of publication:
23.04.2025 Bulletin 2025/17

(51) International Patent Classification (IPC):
G10L 15/22 ^(2006.01)

(21) Application number: **22946317.9**

(52) Cooperative Patent Classification (CPC):
G10L 15/22

(22) Date of filing: **17.06.2022**

(86) International application number:
PCT/CN2022/099595

(87) International publication number:
WO 2023/240636 (21.12.2023 Gazette 2023/51)

(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

• **Beijing Xiaomi Pinecone Electronics Co., Ltd.**
Beijing 100085 (CN)

(72) Inventor: **ZHOU, Lingsong**
Beijing 100085 (CN)

(74) Representative: **Stöckeler, Ferdinand et al**
Schoppe, Zimmermann, Stöckeler
Zinkler, Schenk & Partner mbB
Patentanwälte
Radlkoferstrasse 2
81373 München (DE)

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

(54) **SOUND BOX PLAYING CONTROL METHOD, SOUND BOX PLAYING CONTROL APPARATUS, AND STORAGE MEDIUM**

(57) The present disclosure relates to a sound box playing control method, a sound box playing control apparatus, and a storage medium. The sound box playing control method comprises: in response to an audio playing function being enabled, acquiring audio data from a serving end, and determining a target subspace from among subspaces to which a plurality of sub-sound boxes belong, wherein the target subspace is a subspace where a user is currently located; and sending the audio

data to a sub-sound box in the target sub-space, and controlling the sub-sound box in the target sub-space to perform audio playing on the basis of the audio data. By means of the present disclosure, the waste of energy is reduced; moreover, channel resources between a main control sound box and a sub-sound box, which does not play audio data, are released, thereby providing a more stable auditory experience for users.

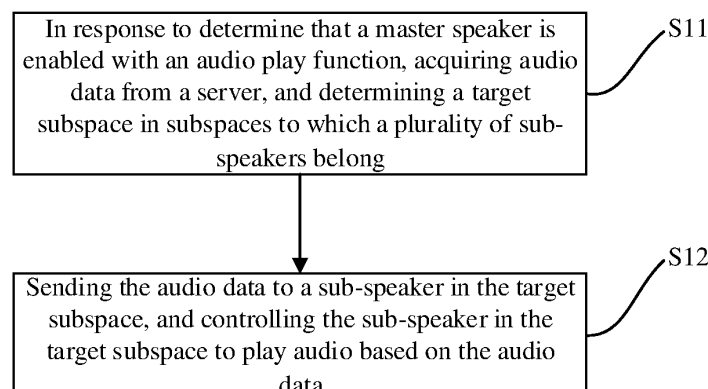


Fig. 1

EP 4 542 542 A1

Description**FIELD**

- 5 **[0001]** The present disclosure relates to the field of speakers, in particular to a speaker play control method, a speaker play control device and a storage medium.

BACKGROUND

- 10 **[0002]** With the development of smart home technology and people's pursuit of high-quality home life, the whole house speaker brings shocking auditory enjoyment to people's home life. A plurality of smart speakers are connected to each other through WiFi and play music synchronously, so that users can have an immersive listening experience while walking. The terminal speaker gets audio data from the server and transmits it to the speakers in all rooms through the wireless network. Through a certain synchronization mechanism, the same audio frame can be played at the same time.
- 15 **[0003]** However, the users usually only play music in one or part of the rooms, and playing music in the rooms without the users will lead to the loss of electric energy on the one hand. On the other hand, since the speakers sending the data will occupy the network channel, the more network resources are occupied, the more likely it will lead to instability of the system. For example, the problem such as unsynchronized play will occur.

20 **SUMMARY**

[0004] In order to overcome the problems existing in the related art, the present disclosure provides a speaker play control method, a speaker play control device and a storage medium.

- 25 **[0005]** According to embodiments of a first aspect of the present disclosure, there is provided a speaker play control method, which is applied to a master speaker, and the master speaker is configured to control a plurality of sub-speakers, and the plurality of sub-speakers belong to different subspaces of a same space. The method includes: in response to enabling an audio play function, acquiring audio data from a server, and determining a target subspace in the subspaces to which the plurality of sub-speakers belong, wherein the target subspace is a subspace where a user is currently located; and sending the audio data to a sub-speaker in the target subspace, and controlling the sub-speaker in the target subspace to play audio based on the audio data.

- 30 **[0006]** In an embodiment, the determining the target subspace in the subspaces to which the plurality of sub-speakers belong includes: determining a user detection result, wherein the user detection result is determined by a set sub-speaker in the subspace based on detection of a human body activity of the user, and the user detection result comprises a presence or an absence of the user; and determining a subspace to which a target sub-speaker whose detection result is the presence of the user belongs as the target subspace.

- 35 **[0007]** In an embodiment, the plurality of sub-speakers correspond to spatial identifiers, and the spatial identifier is configured to identify the subspace to which the sub-speaker belongs; the sending the audio data to the sub-speaker in the target subspace and controlling the sub-speaker in the target subspace to play audio based on the audio data includes: determining a target spatial identifier corresponding to the target subspace; and sending the audio data to the sub-speaker in the target subspace identified by the target spatial identifier, and controlling the sub-speaker in the target subspace based on the target spatial identifier to play audio based on the audio data.

- 40 **[0008]** In an embodiment, the target spatial identifier comprises a first target spatial identifier and a second target spatial identifier, a first target subspace identified by the first target spatial identifier is a subspace where the user is before the user moves, and a second target subspace identified by the second target spatial identifier is a subspace where the user is after the user moves; the controlling the sub-speaker in the target subspace based on the target spatial identifier to play audio based on the audio data includes: sending a first control play instruction to the sub-speaker in the first target subspace, and controlling the sub-speaker in the first target subspace based on the first control play instruction to play audio based on the first control play instruction and the audio data, wherein the first control play instruction is configured to control a playing sound of the sub-speaker to change from large to small; and sending a second control play instruction to the sub-speaker in the second target subspace, and controlling the sub-speaker in the second target subspace based on the second control play instruction to play audio based on the second control play instruction and the audio data, wherein the second control play instruction is configured to control the playing sound of the sub-speaker to change from small to large.

- 45 **[0009]** In an embodiment, the method further includes: determining a non-target subspace, wherein the non-target subspace is a subspace without the user to which each set sub-speaker belongs; and stopping sending the audio data to the non-target subspace.

- 50 **[0010]** According to embodiments of a second aspect of the present disclosure, there is provided a speaker play control method, which is applied to a sub-speaker, and a subspace to which the sub-speaker belongs is a target subspace, and the target subspace is a subspace where a user is currently located. The method includes: acquiring audio data sent by a

master speaker; and playing audio based on the audio data.

[0011] In an embodiment, the sub-speaker is a set sub-speaker, and the set sub-speaker is configured to detect whether there is the user in a subspace to which a target sub-speaker belongs, and the method further includes: detecting a human body activity, and determining a user detection result based on a human body activity detection result, wherein the user detection result comprises a presence or an absence of the user; and sending the user detection result to the master speaker.

[0012] In an embodiment, the sub-speaker corresponds to a spatial identifier, and the spatial identifier is configured to identify the subspace to which the sub-speaker belongs; and the acquiring the audio data sent by the master speaker includes: acquiring the audio data sent by the master speaker based on the spatial identifier.

[0013] In an embodiment, the playing audio based on the audio data includes: in response to that the spatial identifier is a first target spatial identifier, acquiring a first control play instruction sent by the master speaker, wherein a first target subspace identified by the first target spatial identifier is a subspace where the user is before the user moves, and the first control play instruction is configured to control a playing sound of the sub-speaker to change from large to small; and based on the first control play instruction, playing audio based on the audio data in a manner that the playing sound changes from large to small.

[0014] In an embodiment, the playing audio based on the audio data includes: in response to that the spatial identifier is a second target spatial identifier, acquiring a second control play instruction sent by the master speaker, wherein a second target subspace identified by the second target spatial identifier is a subspace where the user is after the user moves, and the second control play instruction is configured to control a playing sound of the sub-speaker to change from small to large; and based on the second control play instruction, playing audio based on the audio data in a manner that the playing sound changes from small to large.

[0015] In an embodiment, the playing audio based on the audio data includes: calibrating a system clock between the sub-speaker and the master speaker; and based on the calibrated system clock and the audio data, playing audio synchronously with the master speaker.

[0016] According to embodiments of a third aspect of the present disclosure, there is provided a speaker play control device, which includes: a determining unit configured to acquire audio data from a server in response to enabling an audio play function, and determining a target subspace in subspaces to which a plurality of sub-speakers belong, wherein the target subspace is a subspace where a user is currently located; and a playing unit configured to send the audio data to a sub-speaker in the target subspace and control the sub-speaker in the target subspace to play audio based on the audio data.

[0017] In an embodiment, the determining unit determines the target subspace in the subspaces to which the plurality of sub-speakers belong in a following manner: determining a user detection result, wherein the user detection result is determined by a set sub-speaker in the subspace based on detection of a human body activity of the user, and the user detection result comprises a presence or an absence of the user; and determining a subspace to which a target sub-speaker whose detection result is the presence of the user belongs as the target subspace.

[0018] In an embodiment, the plurality of sub-speakers correspond to spatial identifiers, and the spatial identifier is configured to identify the subspace to which the sub-speaker belongs; the playing unit sends the audio data to the sub-speaker in the target subspace and controls the sub-speaker in the target subspace to play audio based on the audio data in a following manner: determining a target spatial identifier corresponding to the target subspace; and sending the audio data to the sub-speaker in the target subspace identified by the target spatial identifier, and controlling the sub-speaker in the target subspace based on the target spatial identifier to play audio based on the audio data.

[0019] In an embodiment, the target spatial identifier comprises a first target spatial identifier and a second target spatial identifier, wherein a first target subspace identified by the first target spatial identifier is a subspace where the user is before the user moves, and a second target subspace identified by the second target spatial identifier is a subspace where the user is after the user moves; the playing unit controls the sub-speaker in the target subspace based on the target spatial identifier to play audio based on the audio data in a following manner: sending a first control play instruction to the sub-speaker in the first target subspace, and controlling the sub-speaker in the first target subspace based on the first control play instruction to play audio based on the first control play instruction and the audio data, wherein the first control play instruction is configured to control a playing sound of the sub-speaker to change from large to small; and sending a second control play instruction to the sub-speaker in the second target subspace, and controlling the sub-speaker in the second target subspace based on the second control play instruction to play audio based on the second control play instruction and the audio data, wherein the second control play instruction is configured to control the playing sound of the sub-speaker to change from small to large.

[0020] In an embodiment, the playing unit is further configured to: determine a non-target subspace, wherein the non-target subspace is a subspace without the user to which each set sub-speaker belongs; and stop sending the audio data to the non-target subspace.

[0021] According to embodiments of a fourth aspect of that present disclosure, there is provided a speaker play control device, which includes: an acquiring unit configured to acquire audio data sent by a master speaker; and a playing unit

configured to play audio based on the audio data.

[0022] In an embodiment, the sub-speaker is a set sub-speaker, and the set sub-speaker is configured to detect whether there is the user in a subspace to which a target sub-speaker belongs, and the playing unit is further configured to: detect a human body activity, and determining a user detection result based on a human body activity detection result, wherein the user detection result comprises a presence or an absence of the user; and send the user detection result to the master speaker.

[0023] In an embodiment, the sub-speaker corresponds to a spatial identifier, and the spatial identifier is configured to identify the subspace to which the sub-speaker belongs; the acquiring unit acquires the audio data sent by the master speaker in a following way: acquiring the audio data sent by the master speaker based on the spatial identifier.

[0024] In an embodiment, the playing unit plays audio based on the audio data in a following manner: in response to that the spatial identifier is a first target spatial identifier, acquiring a first control play instruction sent by the master speaker, wherein a first target subspace identified by the first target spatial identifier is a subspace where the user is before the user moves, and the first control play instruction is configured to control a playing sound of the sub-speaker to change from large to small; and based on the first control play instruction, playing audio based on the audio data in a manner that the playing sound changes from large to small.

[0025] In an embodiment, the playing unit plays audio based on the audio data in a following manner: in response to that the spatial identifier is a second target spatial identifier, acquiring a second control play instruction sent by the master speaker, wherein a second target subspace identified by the second target spatial identifier is a subspace where the user is after the user moves, and the second control play instruction is configured to control a playing sound of the sub-speaker to change from small to large; and based on the second control play instruction, playing audio based on the audio data in a manner that the playing sound changes from small to large.

[0026] In an embodiment, the playing unit plays audio based on the audio data in a following manner: calibrating a system clock between the sub-speaker and the master speaker; and based on the calibrated system clock and the audio data, playing audio synchronously with the master speaker.

[0027] According to embodiments of a fifth aspect of that present disclosure, there is provided a speaker play control device, which includes: a processor; and a memory for storing instructions executable by the processor. The processor is configured to perform the method described in the first aspect or any one of the embodiments of the first aspect.

[0028] According to embodiments of a sixth aspect of the present disclosure, there is provided a speaker play control device, which includes: a processor; and a memory for storing instructions executable the processor. The processor is configured to perform the method described in the second aspect or any one of the embodiments of the second aspect.

[0029] According to embodiments of a seventh aspect of the present disclosure, there is provided a computer-readable storage medium in which instructions are stored. When the instructions in the storage medium are executed by a processor of a network device, the network device is enabled to perform the method described in the first aspect or any one of the embodiments of the first aspect.

[0030] According to embodiments of an eighth aspect of the present disclosure, there is provided a computer-readable storage medium in which instructions are stored. When the instructions in the storage medium are executed by a processor of a terminal, the terminal is enabled to perform the method described in the second aspect or any one of the embodiments of the second aspect.

[0031] The technical solution provided by the embodiments of the present disclosure can include the following beneficial effects. When it is determined that the speaker is enabled with the audio play function, the audio data is acquired from the server through the master speaker, and the space where the user is currently located is determined. Further, the master speaker sends the audio data acquired from the server to the sub-speaker in the space where the user is currently located, and the sub-speaker plays audio based on the audio data. Based on this, the speaker in the space selectively plays the audio data according to the space where the user is currently located, thus reducing the waste of energy. That is, in the current space, the speaker which does not detect that the user is in this space is in a non-working state. At the same time, the channel resources between the master speaker and the sub-speakers that do not play the audio data are released, thus providing the user with a more stable listening experience.

[0032] It is to be understood that both the foregoing general description and the following detailed description are illustrative and explanatory only, and are not restrictive of the present disclosure.

BRIEF DESCRIPTION OF THE DRAWINGS

[0033] The accompanying drawings, which are incorporated in and constitute a part of the specification, illustrate embodiments consistent with the present disclosure and together with the description, serve to explain the principles of the present disclosure.

Fig. 1 is a flow chart of a speaker play control method according to an illustrative embodiment.

Fig. 2 is a flow chart of determining a target subspace according to an illustrative embodiment.

Fig. 3 is a flow chart of controlling a sub-speaker in a target subspace to play audio based on audio data according to an illustrative embodiment.

Fig. 4 is a flow chart of controlling a sub-speaker in a target subspace to play audio based on audio data according to an illustrative embodiment.

Fig. 5 is a flow chart of a speaker play control method according to an illustrative embodiment.

Fig. 6 is a flow chart of detecting whether a user exists in a subspace to which a target sub-speaker belongs according to an illustrative embodiment.

Fig. 7 is a flow chart of acquiring audio data sent by a master speaker according to an illustrative embodiment.

Fig. 8 is a flow chart of playing audio based on audio data according to an illustrative embodiment.

Fig. 9 shows a schematic diagram of calibrating a system clock between a master speaker and a sub-speaker.

Fig. 10 shows a schematic diagram of speaker play control.

Fig. 11 is a block diagram of a speaker play control device according to an illustrative embodiment.

Fig. 12 is a block diagram of a speaker play control device according to an illustrative embodiment.

Fig. 13 is a block diagram of a device for speaker play control according to an illustrative embodiment.

DETAILED DESCRIPTION

[0034] Reference will now be made in detail to illustrative embodiments, examples of which are illustrated in the accompanying drawings. When the following description refers to the drawings, unless otherwise indicated, the same numbers in different drawings indicate the same or similar elements. The implementations described in the following illustrative embodiments do not represent all the implementations consistent with the present disclosure.

[0035] In the accompanying drawings, the same or similar reference numerals indicate the same or similar elements or elements with the same or similar functions throughout the specification. The described embodiments are part of the embodiments of the present disclosure, but not all the embodiments. The embodiments described below with reference to the accompanying drawings are illustrative and are intended to explain the present disclosure, and should not be construed as limiting the present disclosure. Based on the embodiments in the present disclosure, all other embodiments acquired by those ordinary skilled in the art without creative work belong to the scope of protection of the present disclosure. Hereinafter, the embodiments of the present disclosure will be described in detail with reference to the accompanying drawings.

[0036] A speaker play control method provided by embodiments of the present disclosure may be applied to a plurality of smart devices connected by wireless communication technology, so as to realize an application scene of sharing music between users. In particular, the plurality of smart devices are in a set space. The plurality of smart devices belong to different subspaces respectively, and the plurality of smart devices include a master smart device and a sub-smart device. The smart device may be a smart speaker or other smart devices.

[0037] With the popularity of the smart speaker and the increasing demand of the user listening to music, major companies are committed to improving the music play function of their speakers. One of the breakthrough progresses is that the speakers are interconnected through a wireless network to realize the whole house play function. A plurality of smart speakers are connected to each other through WiFi and play music synchronously, so that the user can have an immersive listening experience while walking. The speakers get the music data from the server and send it to the speakers in all rooms through the wireless network. Through a certain synchronization mechanism, the same audio frame can be played at the same time, so that all the smart speakers in the family can play music synchronously.

[0038] In the related art, the information interaction among the plurality of smart speakers is realized through WiFi interconnection. The master speaker gets the audio data from the server, frames the audio data by an equal time length, marks corresponding time information for each audio frame, and packages it and sends it to other sub-speakers in all rooms through WiFi. The other sub-speakers receive the audio frame packets sent by the master speaker through WiFi and analyze the audio data and the time information. At the same time of interaction of the audio frame packet, the other sub-speakers should have a system clock interaction with the master speaker, so as to achieve the alignment of the master speaker and the other sub-speakers in terms of the system clock, thus ensuring that the data processing of the master speaker and the other sub-speakers is on the same clock reference. Based on the clock synchronization of the master speaker and the other sub-speakers, the audio data is processed respectively. The master speaker and the other sub-speakers are set to play the same audio frame after a fixed time, so that the plurality of speakers can play the audio synchronously, thus achieving the listening experience of the whole house play. However, in the family, the user usually only plays music in one or part of the rooms, and playing music in the room without the user will waste electric energy on the one hand. On the other hand, sending the data will occupy the network channel. The more network resources are occupied, the more likely it will lead to instability of the system. For example, the plurality of speakers cannot play the audio data synchronously.

[0039] In view of this, the present disclosure provides a speaker play control method. When it is determined that the speaker is enabled with the audio play function, the audio data is acquired from the server through the master speaker, and

a target subspace is determined in subspaces to which a plurality of sub-speakers belong. The target subspace is a subspace where the user is currently located, and the subspace is one of a plurality of spaces into which a complete space is divided. Further, the master speaker sends the audio data acquired from the server to the sub-speaker in the space where the user is currently located, and the sub-speaker plays audio based on the audio data. Based on this, the speaker in the space selectively plays the audio data according to the space where the user is currently located, thus reducing the waste of energy, that is, for a current space, when no user is detected in this space, the speaker in this space is in a non-working state. At the same time, the channel resources between the master speaker and the sub-speakers that do not play the audio data are released, which reduces the network load as much as possible, thus improving the stability of the system and providing the user with a more stable listening experience. Therefore, compared with the way of controlling the speaker to play in the related art, the speaker play control method provided by the present disclosure is more flexible and intelligent.

[0040] Fig. 1 is a flow chart of a speaker play control method according to an illustrative embodiment. As shown in Fig. 1, the method is applied to a master speaker and includes the following steps.

[0041] In the following disclosed embodiments, a speaker communicating with the server and other speakers is called a master speaker, and a speaker communicating with the master speaker is called a sub-speaker. The master speaker is used to control a plurality of sub-speakers, and different sub-speakers are included in different subspaces of the same space. One or more different sub-speakers may be included in the same subspace.

[0042] In step S11, when it is determined that the master speaker is enabled with an audio play function, audio data is acquired from a server, and a target subspace is determined in subspaces to which a plurality of sub-speakers belong.

[0043] The target subspace is a subspace where the user is currently located. The subspace may be understood as any room, such as a living room, a kitchen, a bedroom and so on.

[0044] In embodiments of the present disclosure, when it is determined that the user has enabled the audio play functions of all the speakers in the rooms, the master speaker acquires the audio data from the server through a wireless transmission technology, and determines the room where the user is currently located, by receiving a signal sent by the sub-speaker in the subspace.

[0045] In step S12, the audio data is sent to the sub-speaker in the target subspace, and the sub-speaker in the target subspace is controlled to play audio based on the audio data.

[0046] In embodiments of the present disclosure, in case that the master speaker determines which space has the user, it sends the audio data acquired from the server to the sub-speaker in the space where the user exists, and controls the sub-speaker in the space to play the decoded audio data.

[0047] In the present disclosure, the master speaker determines a non-target subspace through an instruction sent by the sub-speaker in the subspace, and the non-target subspace is a subspace to which a set sub-speaker whose detection result is the absence of the user belongs. In case that the master speaker determines the non-target subspace, it stops sending the audio data to the sub-speaker in the non-target subspace.

[0048] In embodiments of the present disclosure, in case that the master speaker determines that there is no user in the subspace, the master speaker stops sending the audio data to the sub-speaker in the space and closes a channel between the master speaker and the sub-speaker in the space. The channel resources between the master speaker and the sub-speaker that does not play the audio data are released, which improves the stability of the system and provides the user with a more stable listening experience.

[0049] In the present disclosure, after it is determined that all the speakers are enabled with the audio play function, the audio data is acquired from the server, and the target subspace is determined in the subspaces to which the plurality of sub-speakers belong. The audio data is sent to the sub-speaker in the target subspace, and the sub-speaker in the target subspace is controlled to play audio based on the audio data. Through the present disclosure, the speaker in the space selectively plays the audio data according to the space where the user is currently located, thereby reducing the waste of energy. It can be understood that only when there is the user in the space, the master speaker sends the audio data to the sub-speaker in the space.

[0050] Based on the above embodiments, it can be seen that it is a key step that the master speaker determines the target subspace. Therefore, in the following disclosed embodiments, how the master speaker determines the target subspace will be specifically explained.

[0051] Fig. 2 is a flow chart of determining a target subspace according to an illustrative embodiment. As shown in fig. 2, determining the target subspace in the subspaces to which the plurality of sub-speakers belong includes the following steps.

[0052] In step S21, a user detection result is determined.

[0053] The user detection result is determined by a set sub-speaker in the subspace based on the detection of the human body activity of the user, and the user detection result includes the presence or absence of the user.

[0054] In step S22, the subspace to which a target sub-speaker whose detection result is the presence of the user belongs is determined as the target subspace.

[0055] In embodiments of the present disclosure, if the sub-speaker set in the subspace detects the human body activity

of the user in the space, the master speaker receives an instruction confirming the presence of the user from the sub-speaker, and determines the subspace with the user to which the target sub-speaker belongs as the target subspace. If the sub-speaker set in the subspace does not detect the human body activity of the user in the space, the master speaker receives an instruction for confirming the absence of the user from the sub-speaker.

[0056] In the present disclosure, the user detection result is determined, and the subspace to which the target sub-speaker whose detection result is the presence of the user belongs is determined as the target subspace. Through the present disclosure, based on the detection result instruction transmitted from the sub-speaker to the master speaker, the master speaker sends a corresponding instruction to control the sub-speaker, thus realizing the interaction between the sub-speaker and the master speaker.

[0057] Fig. 3 is a flow chart of controlling a sub-speaker in a target subspace to play audio based on audio data according to an illustrative embodiment. As shown in Fig. 3, sending the audio data to the sub-speaker in the target subspace and controlling the sub-speaker in the target subspace to play audio based on the audio data includes the following steps.

[0058] In step S31, a target spatial identifier corresponding to the target subspace is determined.

[0059] The plurality of sub-speakers correspond to spatial identifiers, and the spatial identifier is used to identify the subspace to which the sub-speaker belongs.

[0060] In embodiments of the present disclosure, the room environment is divided into a plurality of subspaces, such as a kitchen, a living room, a bedroom and other subspaces. When configuring the network for the first time, the user can select the space where the speaker is located. For example, the user can select the space where the speaker is currently located through a resizable list on a display screen of the speaker, so that each speaker knows the space environment where it is currently located. For example, speaker A and speaker B are currently placed in the living room, and speaker C and speaker D are currently placed in the master bedroom and so on. Secondly, the sub-speakers in the target subspace perceive each other through ultrasonic communication to determine the subspace to which each sub-speaker belongs. The sub-speaker sends its own spatial identifier to the master speaker, and the master speaker stores IP addresses or other identification IDs of all the sub-speakers in the subspace in a list of devices, so that a target spatial identifier corresponding to the target subspace can be determined.

[0061] In step S32, the audio data is sent to the sub-speaker in the target subspace identified by the target spatial identifier, and the sub-speaker in the target subspace are controlled based on the target spatial identifier to play audio based on the audio data.

[0062] In an example, the user can select one of the sub-speakers in the target subspace in advance, and this sub-speaker detects whether there is the user activity in the space, and sends the detection result to the master speaker. It is also possible to set the speaker with the best performance to detect whether there is the user activity in the space through the mutual perception of the sub-speakers in the target subspace, and send the detection result to the set speaker of the master speaker. Assuming that the user is currently active in the master bedroom and the sub-speaker set in the master bedroom detects the human body activity, the speaker sends a confirmation instruction to the master speaker, and the master speaker sends the audio data to the sub-speaker in the master bedroom and controls the sub-speaker in the master bedroom to play audio.

[0063] In the present disclosure, the target spatial identifier corresponding to the target subspace is determined. The audio data is sent to the sub-speaker in the target subspace identified by the target spatial identifier, and the sub-speaker in the target subspace is controlled based on the target spatial identifier to play audio based on the audio data. Through the present disclosure, the master speaker can accurately control the sub-speaker in the space where the user exists to play audio.

[0064] Fig. 4 is a flow chart of controlling a sub-speaker in a target subspace to play audio based on audio data according to an illustrative embodiment. As shown in Fig. 4, controlling the sub-speaker in the target subspace based on the target spatial identifier to play audio based on the audio data includes the following steps.

[0065] In step S41, a first control play instruction is sent to a sub-speaker in a first target subspace, and the sub-speaker in the first target subspace is controlled based on the first control play instruction to play audio based on the first control play instruction and the audio data.

[0066] The first control play instruction is used to control a playing sound of the sub-speaker to change from large to small.

[0067] In embodiments of the present disclosure, when the user leaves the living room (i.e., the first target subspace), the sub-speaker in the living room cannot detect the user's human body activity at this time, and the sub-speaker sends the instruction for confirming the absence of the user to the master speaker. After receiving this instruction, the master speaker sends a volume reduction control instruction (i.e. the first control play instruction) to the sub-speaker in the living room, and controls the playing sound of the sub-speaker to change from large to small until there is no sound.

[0068] In step S42, a second control play instruction is sent to the sub-speaker in a second target subspace, and the sub-speaker in the second target subspace is controlled to play audio based on the second control play instruction and the audio data.

[0069] The second control play instruction is used to control the playing sound of the sub-speaker to change from small

to large.

[0070] In embodiments of the present disclosure, when the user walks into the bedroom (i.e., the second target subspace), the sub-speaker in the bedroom detects the human body activity of the user at this time, and the sub-speaker sends the instruction for confirming the presence of the user to the master speaker. After receiving this instruction, the master speaker sends a volume increase control instruction (i.e. the second control play instruction) to the sub-speaker in the bedroom to control the playing sound of the sub-speaker to change from small to large.

[0071] In the present disclosure, the first control play instruction is sent to the sub-speaker in the first target subspace, and the sub-speaker in the first target subspace is controlled based on the first control play instruction to play audio based on the first control play instruction and the audio data. The second control play instruction is sent to the sub-speaker in the second target subspace, and the sub-speaker in the second target subspace is controlled based on the second control play instruction to play audio based on the second control play instruction and the audio data. Through the present disclosure, the master speaker can also control the change of volume of the audio played by the sub-speaker according to the detection result of the sub-speaker in the subspace, so that the sub-speaker can achieve a gradual effect in volume and bring a good listening experience to the user.

[0072] The embodiments of the present disclosure provide the speaker play control method, which is applied to the master speaker, and the master speaker communicates with the server and receives the audio data from the server. The received audio data is decoded to obtain the decoded audio data. Then, the decoded audio data is divided according to a preset duration to obtain audio frame data. According to the received instructions of the sub-speakers in the respective subspaces, further operations are carried out. If the instruction that the sub-speaker informs the presence of the user is received, the audio data and the play control instruction are continued to be sent to the sub-speaker in the space, and the clock data for calibration also needs to be transmitted while the audio data is transmitted. If the instruction that the sub-speaker informs the absence of the user is received, the audio data is stopped from being sent to the sub-speaker in the space. According to the embodiments of the present disclosure, the speaker in the space selectively plays the audio data according to the space where the user is currently located, thereby reducing the waste of energy. At the same time, the channel resources between the master speaker and the sub-speakers that do not play the audio data are released, which improves the stability of the system.

[0073] Fig. 5 is a flow chart of a speaker play control method according to an illustrative embodiment, which is applied to a sub-speaker as shown in Fig. 5 and includes the following steps.

[0074] In step S51, the sub-speaker in the target subspace acquires the audio data sent by the master speaker.

[0075] In step S52, audio is played based on the audio data.

[0076] In the present disclosure, the audio data sent by the master speaker is acquired, and the audio is played based on the audio data. According to the present disclosure, the sub-speaker is controlled by the master speaker, and the interaction between the sub-speaker and the master speaker is realized.

[0077] Fig. 6 is a flow chart of detecting whether a user exists in a subspace to which a target sub-speaker belongs according to an illustrative embodiment. As shown in Fig. 6, the sub-speaker is a set sub-speaker, and the set sub-speaker is used to detect whether the user exists in the subspace to which the target sub-speaker belongs, and the method includes the following steps.

[0078] In step S61, the human body activity is detected, and the user detection result is determined based on a detection result of the human body activity.

[0079] The user detection results include the presence or absence of the user.

[0080] In embodiments of the present disclosure, one sub-speaker is selected randomly or selected according to the performance of the sub-speaker in each subspace to transmit and receive an ultrasonic wave. When the audio play functions of all the speakers are enabled, the selected sub-speaker in each subspace emits the ultrasonic wave to detect the human body activity of the user. In embodiments of the present disclosure, a differential channel impulse response (dCIR) is used to detect the human body activity of the user. The essence of this detection method is through calculating the dCIR in the environment in real-time. When the user is not active in subspace A at present, the dCIR value of the sub-speaker in subspace A receiving the ultrasonic wave back approaches zero. At this time, the sub-speaker set in the target subspace will send the instruction that the user does not exist in the space to the master speaker. After receiving the instruction, the master speaker sends the corresponding instruction operation to the sub-speaker. When the user is currently in subspace B, the dCIR value of the sub-speaker in subspace B receiving the ultrasonic wave back will reflect the amplitude change on the dCIR. At this time, if the dCIR amplitude of the sub-speaker set in the target subspace changes, the speaker will send the instruction that the user exists in the space to the master speaker. After receiving the instruction, the master speaker sends the corresponding instruction operation to the sub-speaker. By detecting the overall amplitude state of the dCIR, it can be detected whether there is the user in the current subspace.

[0081] In embodiments of the present disclosure, the application principle of the dCIR is based on the following formula:

$$\underbrace{\begin{pmatrix} s_1 & s_2 & \cdots & s_L \\ s_2 & s_3 & \cdots & s_{L+1} \\ \vdots & \vdots & \ddots & \vdots \\ s_P & s_{P+1} & \cdots & s_{P+L-1} \end{pmatrix}}_{\text{TrainingMatrix}} \underbrace{\begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_L \end{pmatrix}}_{\text{CIR}} = \underbrace{\begin{pmatrix} r_{L+1} \\ r_{L+2} \\ \vdots \\ r_{L+P} \end{pmatrix}}_{R[n]}$$

[0082] In the above formula, S is an ultrasonic signal emitted by a loudspeaker, r is the ultrasonic signal collected by a microphone, and H is a CIR vector. The calculation formula of h is: $h = (S^T S)^{-1} S^T R$. The dCIR is described as: $dCIR_m = \hat{h} = h_m - h_{m-1}$, where m is a current frame. The amplitude statistic of the dCIR is:

$$AMP_{dCIR} = \sum_{i=1}^L \hat{h}_i$$

. When the AMP_{dCIR} is greater than a set threshold, it is determined that there is the user in the current subspace, and the set threshold may be 3 after being verified by experiments.

[0083] In step S62, the user detection result is sent to the master speaker.

[0084] In embodiments of the present disclosure, when the sub-speaker in the subspace determines that the user is currently in the same room as the sub-speaker, the instruction is sent to the master speaker, informing the master speaker that it can continue to send the audio frame data to the sub-speaker in the subspace. In this case, when the master speaker receives the instruction sent from the sub-speaker, it can be determined that the user detection result determined by the set sub-speaker in this subspace is that there is the user. The master speaker determines the subspace to which the target speaker whose detection result is the presence of the user belongs as the target subspace.

[0085] In embodiments of the present disclosure, the sub-speaker in the subspace may also detect that there is no user in the current subspace.

[0086] In the present disclosure, non-target subspace information is sent to the master speaker, and a non-target subspace is a subspace without the user to which each set sub-speaker belongs. The master speaker is controlled to stop sending the audio data to the non-target subspace.

[0087] In embodiments of the present disclosure, when the dCIR value of the ultrasonic wave received back by the sub-speaker in the subspace approaches zero, it is determined that there is no user in the current subspace. When it is determined that there is no user in this subspace, the sub-speaker will send the instruction to the master speaker to inform that there is no user in the space to which it belongs.

[0088] In embodiments of the present disclosure, when the master speaker receives the non-target subspace information sent from the sub-speaker, it is determined that there is no user in the space where the sub-speaker is located. In order to avoid waste of resources, the master speaker stops sending the audio data to the non-target subspace.

[0089] In the present disclosure, the human body activity is detected, and the user detection result is determined based on the human body activity detection result. The user detection result is sent to the master speaker. Through the present disclosure, the audio data is stopped in time from being sent to the sub-speaker in the subspace where the user does not exist, and the channel resources are released dynamically, thus ensuring the stability of the system and providing a better listening experience.

[0090] Fig. 7 is a flow chart of acquiring audio data sent by a master speaker according to an illustrative embodiment. As shown in Fig. 7, the acquiring audio data sent by the master speaker includes the following steps.

[0091] In step S71, the audio data sent by the master speaker based on the spatial identifier is acquired.

[0092] In step S72, the audio is played based on different spatial identifiers.

[0093] In the present disclosure, in response to that the spatial identifier is a first target spatial identifier, the first control play instruction sent by the master speaker is acquired, and the first target subspace identified by the first target spatial identifier is the subspace where the user is before the user moves, and the first control play instruction is used to control the playing sound of the sub-speaker to change from large to small. Based on the first control play instruction, the audio is played based on the audio data in a manner that the playing sound changes from large to small.

[0094] In embodiments of the present disclosure, when the user leaves the living room (i.e., the first target subspace), the sub-speaker in the living room cannot detect the user's human body activity at this time, and the sub-speaker sends the instruction for confirming the absence of the user to the master speaker. After receiving the instruction, the master speaker sends the volume reduction control instruction (i.e. the first control play instruction) to the sub-speaker in the living room,

and controls the playing sound of the sub-speaker to change from large to small until there is no sound.

[0095] In the present disclosure, in response to that the spatial identifier is a second target spatial identifier, the second control play instruction sent by the master speaker is acquired, and the second target subspace identified by the second target spatial identifier is the subspace where the user is after the user moves, and the second control play instruction is used to control the playing sound of the sub-speaker to change from small to large. Based on the second control play instruction, the audio is played based on the audio data in a manner that the playing sound changes from small to large.

[0096] In embodiments of the present disclosure, when the user walks into the bedroom (i.e., the second target subspace), the sub-speaker in the bedroom detects the user's human body activity at this time, and the sub-speaker sends the instruction for confirming the presence of the user to the master speaker. After receiving this instruction, the master speaker sends the volume increase control instruction (i.e. the second control play instruction) to the sub-speaker in the bedroom to control the playing sound of the sub-speaker to change from small to large.

[0097] In the present disclosure, the audio data sent by the master speaker based on the spatial identifier is acquired. The audio is played based on different spatial identifiers. Through the present disclosure, the master speaker can also control the change of volume of the sub-speaker playing audio according to the detection result of the sub-speaker in the subspace, so that the sub-speaker can achieve a gradual effect in volume and bring a good listening experience to the user.

[0098] Fig. 8 is a flow chart of playing audio based on audio data according to an illustrative embodiment. As shown in Fig. 8, playing audio based on the audio data includes the following steps.

[0099] In step S81, a system clock between the sub-speaker and the master speaker is calibrated.

[0100] In an embodiment of the present disclosure, Fig. 9 shows a schematic diagram of calibrating the system clock between the master speaker and the sub-speaker. As shown in Fig. 9, while the audio data is transmitted between the master speaker and the sub-speaker, the master speaker and the sub-speaker also need to have an interaction of clock information. The sub-speaker needs to calculate the system clock difference between itself and the master speaker, and adjust itself to the system clock consistent with the master speaker. According to its own clock, the sub-speaker sends time information to the master speaker through WiFi at time TB0, the master speaker receives the time information of the sub-speaker at time TA0 according to its own clock, then the master speaker sends time information to the sub-speaker through WiFi at time TA1, and the sub-speaker receives the time information of the master speaker at time TB1. The sub-speaker can calculate the system clock difference between itself and the master speaker by using the four time information of TB0,

TA0, TA1 and TB1 as follows: $T_{A0} - T_{B0} = \Delta + \tau_0$ and $T_{B1} - T_{A1} = -\Delta + \tau_1$, where Δ represents a system clock error between the sub-speaker and the master speaker, and τ_0 and τ_1 are WiFi transmission delays between the master speaker and the sub-speaker. Then, the system clock error can be calculated as:

$$\Delta = \frac{(T_{A0} - T_{B0}) - (T_{B1} - T_{A1}) + (\tau_1 - \tau_0)}{2}$$

The sub-speaker compensates its own system clock with the clock difference Δ , and then it is calibrated to the system clock consistent with the master speaker. On the basis that all the speakers are aligned with the master speaker in clock, the audio frames at time T0 are played at time T0+1s, so that the synchronous play of all the speakers can be achieved.

[0101] In step S82, based on the calibrated system clock and the audio data, the audio is played synchronously with the master speaker.

[0102] In the present disclosure, the system clock between the sub-speaker and the master speaker is calibrated. Based on the calibrated system clock and the audio data, the audio is played synchronously with the master speaker. Through the present disclosure, based on the system clock criterion between the sub-speaker and the master speaker, the synchronous audio play of the respective speakers is realized, and a good listening experience is brought to the user.

[0103] Fig. 10 shows a schematic diagram of a speaker play control. As shown in Fig. 10, first, the space where each speaker is located is divided, and the division of the subspace can be realized by user configuration and automatic perception. The user configuration means that each time the user configures the network for the speaker, the user chooses the subspace in which the speaker is located. For example, the living room is chosen for speakers A and B, and the master bedroom is chosen for speakers C and D, so that speakers A and B mutually know that they are in a same subspace, and speakers C and D mutually know that they are in a same subspace. The automatic perception means that the user does not need to choose, and the speakers perceive each other through ultrasonic communication. For example, only speaker B can receive the ultrasonic information sent by speaker A, and speaker C and speaker D cannot receive it due to wall blocking, so that speaker B knows that it is in the same subspace with speaker A. Each speaker perceives in sequence, and then the subspaces where all the speakers are located can be finally divided. Speaker A is used as the master speaker, the information of the spaces where all the devices are located is finally summarized into speaker A, and the IP addresses or other identification IDs of the devices in the respective subspaces are stored in the list. In this case, the relationship between the speakers is established. In subspace 1, speaker M detects the user by ultrasonic technology. If speaker M

detects that the user is in subspace 1, it sends an instruction to the master speaker to inform that there is the user in subspace 1. When the master speaker receives the instruction from speaker M, it continues to send the audio data to speaker M and speaker N in subspace 1. If speaker M does not detect that the user is in subspace 1, it sends an instruction to the master speaker to inform that there is no user in subspace 1. When the master speaker receives the instruction from speaker M, it stops sending the audio data to speaker M and speaker N in subspace 1, and closes the channels between the master speaker and speakers M and N. Through the present disclosure, the play function of the speaker in the room without the user is stopped. Thus, on the one hand, the unnecessary waste of electric energy is saved, and on the other hand, it saves the occupation of network resources, reduces the network load and helps to improve the stability of the system.

[0104] The embodiments of the present disclosure provide the speaker play control method, which is applied to the sub-speakers in the respective divided subspaces, and the sub-speakers communicate with the master speaker. When the sub-speaker detects that there is the user in the space through ultrasonic technology, it sends the instruction that there is the user to the master speaker. When the sub-speaker detects that there is no user in the space through ultrasonic technology, it sends the instruction that there is no user to the master speaker. The sub-speaker receives the audio data and the play control instruction from the master speaker, decodes the data and plays the decoded data, under the condition of informing the master speaker that there is the user in the space.

[0105] Based on the same concept, the embodiments of the present disclosure also provide a speaker play control device.

[0106] It can be understood that, in order to realize the above functions, the speaker play control device provided by the embodiments of the present disclosure includes corresponding hardware structures and/or software modules for executing various functions. In combination with the units and algorithm steps of various examples disclosed in the embodiments of the present disclosure, the embodiments of the present disclosure can be realized in the form of hardware or a combination of hardware and computer software. Whether a function is executed by hardware or in a manner of computer software driving hardware depends on the specific application and design constraints of the technical solution. Those skilled in the art can use different methods to realize the described functions for each specific application, but this realization should not be considered beyond the scope of the technical solution of the embodiment of the present disclosure.

[0107] Fig. 11 is a block diagram of a speaker play control device according to an illustrative embodiment. Referring to Fig. 11, the device 100 can be provided as the master speaker according to the above embodiments, and include a determining unit 101 and a playing unit 102.

[0108] The determining unit 101 is configured to acquire audio data from a server in response to enabling an audio play function, and determine a target subspace in subspaces to which a plurality of sub-speakers belong. The target subspace is a subspace where a user is currently located. The playing unit 102 is configured to send the audio data to the sub-speaker in the target subspace and control the sub-speaker in the target subspace to play audio based on the audio data.

[0109] In an embodiment, the determining unit 101 determines the target subspace in the subspaces to which the plurality of sub-speakers belong in the following ways: determining a user detection result, where the user detection result is determined by a set sub-speakers in the subspace based on detection of a human body activity of the user, and the user detection result includes a presence or an absence of the user; and determining the subspace to which the target sub-speaker whose detection result is the presence of the user belongs as the target subspace.

[0110] In an embodiment, the plurality of sub-speakers correspond to spatial identifiers, and the spatial identifier is used for identifying the subspace to which the sub-speaker belongs. The playing unit 102 sends the audio data to the sub-speaker in the target subspace and control the sub-speaker in the target subspace to play audio based on the audio data in the following ways: determining a target spatial identifier corresponding to the target subspace; sending the audio data to the sub-speaker in the target subspace identified by the target spatial identifier, and controlling the sub-speaker in the target subspace based on the target spatial identifier to play audio based on the audio data.

[0111] In an embodiment, the target spatial identifier includes a first target spatial identifier and a second target spatial identifier, a first target subspace identified by the first target spatial identifier is a subspace where the user is before the user moves, and a second target subspace identified by the second target spatial identifier is a subspace where the user is after the user moves. The playing unit 102 controls the sub-speaker in the target subspace based on the target spatial identifier to play audio based on the audio data in the following manners: sending a first control play instruction to the sub-speaker in the first target subspace, and controlling the sub-speaker in the first target subspace based on the first control play instruction to play audio based on the first control play instruction and the audio data, where the first control play instruction is used to control a playing sound of the sub-speaker to change from large to small; sending a second control play instruction to the sub-speaker in the second target subspace, and controlling the sub-speaker in the second target subspace based on the second control play instruction to play audio based on the second control play instruction and the audio data, where the second control play instruction is used to control the playing sound of the sub-speaker to change from small to large.

[0112] In an embodiment, the playing unit 102 is also used to: determine a non-target subspace, which is a subspace

without the user to which each set sub-speaker belongs; and stop sending the audio data to the non-target subspace.

[0113] Fig. 12 is a block diagram of a speaker play control device according to an illustrative embodiment. Referring to Fig. 12, the device 200 can be provided as the sub-speaker according to the above embodiments, and include an acquiring unit 201 and a playing unit 202.

[0114] The acquiring unit 201 is configured to acquire audio data sent by a master speaker. The playing unit 202 is used to play audio based on the audio data.

[0115] In an embodiment, the sub-speaker is a set sub-speaker, and the set sub-speaker is used to detect whether there is a user in the subspace to which a target sub-speaker belongs. The playing unit 202 is also used to: detect a human body activity and determine a user detection result based on a human body activity detection result, where the user detection result includes a presence or an absence of a user; and send the user detection result to the master speaker.

[0116] In an embodiment, the sub-speaker corresponds to a spatial identifier, and the spatial identifier is used to identify the subspace to which the sub-speaker belongs. The acquiring unit 201 acquires the audio data sent by the master speaker in a following way: acquiring the audio data sent by the master speaker based on the spatial identifier.

[0117] In an embodiment, the playing unit 202 plays audio based on the audio data in the following ways: in response to that the spatial identifier is a first target spatial identifier, acquiring a first control play instruction sent by the master speaker, where a first target subspace identified by the first target spatial identifier is a subspace where the user is before the user moves, and the first control play instruction is used to control a playing sound of the sub-speaker to change from large to small; based on the first control play instruction, playing audio based on the audio data in a manner that the playing sound changes from large to small.

[0118] In an embodiment, the playing unit 202 plays audio based on the audio data in the following ways: in response to that the spatial identifier is a second target spatial identifier, acquiring a second control play instruction sent by the master speaker, where the second target subspace identified by the second target spatial identifier is a subspace where the user is after the user moves, and the second control instruction is used to control the playing sound of the sub-speaker to change from small to large; and based on the second control play instruction, playing audio based on the audio data in a manner that the playing sound changes from small to large.

[0119] In an embodiment, the playing unit 202 plays audio based on the audio data in the following ways: calibrating a system clock between the sub-speaker and the master speaker; and based on the calibrated system clock and the audio data, playing audio synchronously with the master speaker.

[0120] With regard to the devices in the above embodiments, the specific way in which each module performs operations has been described in detail in the embodiments of the methods, and will not be described in detail here.

[0121] Fig. 13 is a block diagram of a device for speaker play control according to an illustrative embodiment. For example, a device 300 may be a mobile phone, a computer, a digital broadcasting terminal, a messaging device, a game console, a tablet device, a medical device, a fitness device, a personal digital assistant and the like.

[0122] Referring to Fig. 13, the device 300 may include one or more of the following components: a processing component 302, a memory 304, a power component 306, a multimedia component 308, an audio component 310, an input/output (I/O) interface 312, a sensor component 314, and a communication component 316.

[0123] The processing component 302 generally controls the overall operation of the device 300, such as operations associated with display, telephone call, data communication, camera operation and recording operation. The processing component 302 may include one or more processors 320 to execute instructions to complete all or part of the steps of the method described above. In addition, the processing component 302 can include one or more modules to facilitate the interaction between the processing component 302 and other components. For example, the processing component 302 can include a multimedia module to facilitate interaction between the multimedia component 308 and the processing component 302.

[0124] The memory 304 is configured to store various types of data to support operations in the device 300. Examples of the data include instructions for any application or method operating on the device 300, contact data, phone book data, messages, pictures, videos, and the like. The memory 304 can be realized by any type of volatile or nonvolatile memory device or their combination, such as static random access memory (SRAM), electrically erasable programmable read-only memory (EEPROM), erasable programmable read-only memory (EPROM), programmable read-only memory (PROM), read-only memory (ROM), magnetic memory, flash memory, magnetic disk or optical disk.

[0125] The power component 306 provides power for various components of the device 300. The power component 306 may include a power management system, one or more power sources, and other components associated with generating, managing and distributing power for the device 300.

[0126] The multimedia component 308 includes a screen that provides an output interface between the device 300 and the user. In some embodiments, the screen may include a liquid crystal display (LCD) and a touch panel (TP). If the screen includes a touch panel, the screen may be implemented as a touch screen to receive an input signal from the user. The touch panel includes one or more touch sensors to sense touch, sliding and gestures on the touch panel. The touch sensor may not only sense the boundary of a touch or sliding action, but also detect the duration and pressure related to the touch or sliding operation. In some embodiments, the multimedia component 308 includes a front camera and/or a rear camera.

When the device 300 is in an operation mode, such as a shooting mode or a video mode, the front camera and/or the rear camera can receive external multimedia data. Each front camera and rear camera can be a fixed optical lens system or have a focal length and an optical zoom capability.

[0127] The audio component 310 is configured to output and/or input audio signals. For example, the audio component 310 includes a microphone (MIC) configured to receive external audio signals when the device 300 is in an operation mode, such as a call mode, a recording mode and a voice recognition mode. The received audio signal may be further stored in the memory 304 or transmitted via the communication component 316. In some embodiments, the audio component 310 further includes a speaker for outputting audio signals.

[0128] The I/O interface 312 provides an interface between the processing component 302 and peripheral interface modules, which can be keyboards, click wheels, buttons, etc. These buttons may include, but are not limited to, a home button, a volume button, a start button and a lock button.

[0129] The sensor assembly 314 includes one or more sensors for providing various aspects of the state evaluation for the device 300. For example, the sensor component 314 can detect the on/off state of the device 300, the relative positioning of components, such as the display and keypad of the device 300, the position change of the device 300 or a component of the device 300, the presence or absence of user contact with the device 300, the orientation or acceleration/deceleration of the device 300 and the temperature change of the device 300. The sensor assembly 314 may include a proximity sensor configured to detect the presence of a nearby object without any physical contact. The sensor assembly 314 may also include an optical sensor, such as a CMOS or CCD image sensor, for use in imaging applications. In some embodiments, the sensor assembly 314 may further include an acceleration sensor, a gyro sensor, a magnetic sensor, a pressure sensor or a temperature sensor.

[0130] The communication component 316 is configured to facilitate wired or wireless communication between the device 300 and other devices. The device 300 can access a wireless network based on communication standards, such as WiFi, 2G or 3G, or a combination thereof. In an illustrative embodiment, the communication component 316 receives a broadcast signal or broadcast related information from an external broadcast management system via a broadcast channel. In an illustrative embodiment, the communication component 316 further includes a near field communication (NFC) module to facilitate short-range communication. For example, the NFC module can be implemented based on radio frequency identification (RFID) technology, infrared data association (IrDA) technology, ultra-wideband (UWB) technology, Bluetooth (BT) technology and other technologies.

[0131] In an illustrative embodiment, the device 300 may be implemented by one or more application specific integrated circuits (ASIC), digital signal processors (DSP), digital signal processing devices (DSPD), programmable logic devices (PLD), field programmable gate arrays (FPGA), controllers, microcontrollers, microprocessors or other electronic components, for performing the above methods.

[0132] In an illustrative embodiment, there is also provided a non-transitory computer-readable storage medium including instructions, such as the memory 304 including instructions, and the instructions can be executed by the processor 320 of the device 300 to complete the above method. For example, the non-transitory computer-readable storage medium can be ROM, random access memory (RAM), CD-ROM, magnetic tape, floppy disk, optical data storage device, etc.

[0133] It can be understood that "a plurality of" in the present disclosure refers to two or more, and other quantifiers are similar. The term "and/or", which describes the relationship of related objects, means that there may be three kinds of relationships. For example, A and/or B can mean that A exists alone, A and B exist together, and B exists alone. The character "/" generally indicates that the former and latter objects have an OR relationship. The singular forms "a", "said" and "the" are also intended to include the plural forms, unless the context clearly indicates otherwise.

[0134] It is further understood that the terms "first" and "second" are used to describe various information, but the information should not be limited to these terms. These terms are only used to distinguish the same type of information from each other and do not indicate a specific order or importance. In fact, the expressions "first" and "second" can be used interchangeably. For example, without departing from the scope of the present disclosure, the first information may also be called the second information, and similarly, the second information may also be called the first information.

[0135] It can be further understood that unless otherwise specified, "connection" includes direct connection between two without other components, and indirect connection between two with other components.

[0136] It can be further understood that although the operations are described in a specific order in the drawings in the embodiments of the present disclosure, it should not be understood as requiring that these operations be performed in the specific order or serial order shown, or that all the operations shown should be performed to obtain the desired results. In certain circumstances, multitasking and parallel processing may be beneficial.

[0137] Other embodiments of the present disclosure will easily occur to those skilled in the art after considering the specification and practicing the invention disclosed herein. This application is intended to cover any variations, uses or adaptations of the present disclosure, which follow the general principles of the present disclosure and include the common sense or common technical means in the related art that are not disclosed in the present disclosure. The specification and embodiments are to be regarded as illustrative only, with the true scope and spirit of the present

disclosure being indicated by the following claims.

[0138] It should be understood that the present disclosure is not limited to the precise structure described above and shown in the drawings, and various modifications and changes can be made without departing from the scope of the present disclosure. The scope of the present disclosure is limited only by the scope of the appended claims.

Claims

1. A speaker play control method, applied to a master speaker, the master speaker being configured to control a plurality of sub-speakers, the plurality of sub-speakers belonging to different subspaces of a same space, and the method comprising:

in response to enabling an audio play function, acquiring audio data from a server, and determining a target subspace in the subspaces to which the plurality of sub-speakers belong, wherein the target subspace is a subspace where a user is currently located; and
 sending the audio data to a sub-speaker in the target subspace, and controlling the sub-speaker in the target subspace to play audio based on the audio data.

2. The method according to claim 1, wherein the determining the target subspace in the subspaces to which the plurality of sub-speakers belong comprises:

determining a user detection result, wherein the user detection result is determined by a set sub-speaker in the subspace based on detection of a human body activity of the user, and the user detection result comprises a presence or an absence of the user; and
 determining a subspace to which a target sub-speaker whose detection result is the presence of the user belongs as the target subspace.

3. The method according to claim 1 or 2, wherein the plurality of sub-speakers correspond to spatial identifiers, and the spatial identifier is configured to identify the subspace to which the sub-speaker belongs;
 the sending the audio data to the sub-speaker in the target subspace and controlling the sub-speaker in the target subspace to play audio based on the audio data comprises:

determining a target spatial identifier corresponding to the target subspace; and
 sending the audio data to the sub-speaker in the target subspace identified by the target spatial identifier, and
 controlling the sub-speaker in the target subspace based on the target spatial identifier to play audio based on the audio data.

4. The method according to claim 3, wherein the target spatial identifier comprises a first target spatial identifier and a second target spatial identifier, a first target subspace identified by the first target spatial identifier is a subspace where the user is before the user moves, and a second target subspace identified by the second target spatial identifier is a subspace where the user is after the user moves;
 the controlling the sub-speaker in the target subspace based on the target spatial identifier to play audio based on the audio data comprises:

sending a first control play instruction to the sub-speaker in the first target subspace, and controlling the sub-speaker in the first target subspace based on the first control play instruction to play audio based on the first control play instruction and the audio data, wherein the first control play instruction is configured to control a playing sound of the sub-speaker to change from large to small; and
 sending a second control play instruction to the sub-speaker in the second target subspace, and controlling the sub-speaker in the second target subspace based on the second control play instruction to play audio based on the second control play instruction and the audio data, wherein the second control play instruction is configured to control the playing sound of the sub-speaker to change from small to large.

5. The method according to claim 1, further comprising:

determining a non-target subspace, wherein the non-target subspace is a subspace without the user to which each set sub-speaker belongs; and
 stopping sending the audio data to the non-target subspace.

6. A speaker play control method, applied to a sub-speaker, a subspace to which the sub-speaker belongs being a target subspace, the target subspace being a subspace where a user is currently located, and the method comprising:

acquiring audio data sent by a master speaker; and
playing audio based on the audio data.

7. The method according to claim 6, wherein the sub-speaker is a set sub-speaker, and the set sub-speaker is configured to detect whether there is the user in a subspace to which a target sub-speaker belongs, and the method further comprises:

detecting a human body activity, and determining a user detection result based on a human body activity detection result, wherein the user detection result comprises a presence or an absence of the user; and
sending the user detection result to the master speaker.

8. The method according to claim 7, wherein the sub-speaker corresponds to a spatial identifier, and the spatial identifier is configured to identify the subspace to which the sub-speaker belongs; and
the acquiring the audio data sent by the master speaker comprises:
acquiring the audio data sent by the master speaker based on the spatial identifier.

9. The method according to claim 8, wherein the playing audio based on the audio data comprises:

in response to that the spatial identifier is a first target spatial identifier, acquiring a first control play instruction sent by the master speaker, wherein a first target subspace identified by the first target spatial identifier is a subspace where the user is before the user moves, and the first control play instruction is configured to control a playing sound of the sub-speaker to change from large to small; and
based on the first control play instruction, playing audio based on the audio data in a manner that the playing sound changes from large to small.

10. The method according to claim 8, wherein the playing audio based on the audio data comprises:

in response to that the spatial identifier is a second target spatial identifier, acquiring a second control play instruction sent by the master speaker, wherein a second target subspace identified by the second target spatial identifier is a subspace where the user is after the user moves, and the second control play instruction is configured to control a playing sound of the sub-speaker to change from small to large; and
based on the second control play instruction, playing audio based on the audio data in a manner that the playing sound changes from small to large.

11. The method according to claim 6, wherein the playing audio based on the audio data comprises:

calibrating a system clock between the sub-speaker and the master speaker; and
based on the calibrated system clock and the audio data, playing audio synchronously with the master speaker.

12. A speaker play control device, comprising:

a determining unit configured to acquire audio data from a server in response to enabling an audio play function, and determining a target subspace in subspaces to which a plurality of sub-speakers belong, wherein the target subspace is a subspace where a user is currently located; and
a playing unit configured to send the audio data to a sub-speaker in the target subspace and control the sub-speaker in the target subspace to play audio based on the audio data.

13. The device according to claim 12, wherein the determining unit determines the target subspace in the subspaces to which the plurality of sub-speakers belong in a following manner:

determining a user detection result, wherein the user detection result is determined by a set sub-speaker in the subspace based on detection of a human body activity of the user, and the user detection result comprises a presence or an absence of the user; and
determining a subspace to which a target sub-speaker whose detection result is the presence of the user belongs as the target subspace.

14. The device according to claim 12 or 13, wherein the plurality of sub-speakers correspond to spatial identifiers, and the spatial identifier is configured to identify the subspace to which the sub-speaker belongs;
the playing unit sends the audio data to the sub-speaker in the target subspace and controls the sub-speaker in the target subspace to play audio based on the audio data in a following manner:

determining a target spatial identifier corresponding to the target subspace; and
sending the audio data to the sub-speaker in the target subspace identified by the target spatial identifier, and
controlling the sub-speaker in the target subspace based on the target spatial identifier to play audio based on the audio data.

15. The device according to claim 14, wherein the target spatial identifier comprises a first target spatial identifier and a second target spatial identifier, wherein a first target subspace identified by the first target spatial identifier is a subspace where the user is before the user moves, and a second target subspace identified by the second target spatial identifier is a subspace where the user is after the user moves;

the playing unit controls the sub-speaker in the target subspace based on the target spatial identifier to play audio based on the audio data in a following manner;
sending a first control play instruction to the sub-speaker in the first target subspace, and controlling the sub-speaker in the first target subspace based on the first control play instruction to play audio based on the first control play instruction and the audio data, wherein the first control play instruction is configured to control a playing sound of the sub-speaker to change from large to small; and
sending a second control play instruction to the sub-speaker in the second target subspace, and controlling the sub-speaker in the second target subspace based on the second control play instruction to play audio based on the second control play instruction and the audio data, wherein the second control play instruction is configured to control the playing sound of the sub-speaker to change from small to large.

16. The device according to claim 12, wherein the playing unit is further configured to:

determine a non-target subspace, wherein the non-target subspace is a subspace without the user to which each set sub-speaker belongs; and
stop sending the audio data to the non-target subspace.

17. A speaker play control device, comprising:

an acquiring unit configured to acquire audio data sent by a master speaker; and
a playing unit configured to play audio based on the audio data.

18. The device according to claim 17, wherein the sub-speaker is a set sub-speaker, and the set sub-speaker is configured to detect whether there is the user in a subspace to which a target sub-speaker belongs, and the playing unit is further configured to:

detect a human body activity, and determining a user detection result based on a human body activity detection result, wherein the user detection result comprises a presence or an absence of the user; and
send the user detection result to the master speaker.

19. The device according to claim 18, wherein the sub-speaker corresponds to a spatial identifier, and the spatial identifier is configured to identify the subspace to which the sub-speaker belongs;
the acquiring unit acquires the audio data sent by the master speaker in a following way:
acquiring the audio data sent by the master speaker based on the spatial identifier.

20. The device according to claim 19, wherein the playing unit plays audio based on the audio data in a following manner:

in response to that the spatial identifier is a first target spatial identifier, acquiring a first control play instruction sent by the master speaker, wherein a first target subspace identified by the first target spatial identifier is a subspace where the user is before the user moves, and the first control play instruction is configured to control a playing sound of the sub-speaker to change from large to small; and
based on the first control play instruction, playing audio based on the audio data in a manner that the playing sound changes from large to small.

21. The device according to claim 19, wherein the playing unit plays audio based on the audio data in a following manner:

in response to that the spatial identifier is a second target spatial identifier, acquiring a second control play instruction sent by the master speaker, wherein a second target subspace identified by the second target spatial identifier is a subspace where the user is after the user moves, and the second control play instruction is configured to control a playing sound of the sub-speaker to change from small to large; and based on the second control play instruction, playing audio based on the audio data in a manner that the playing sound changes from small to large.

22. The device according to claim 17, wherein the playing unit plays audio based on the audio data in a following manner:

calibrating a system clock between the sub-speaker and the master speaker; and based on the calibrated system clock and the audio data, playing audio synchronously with the master speaker.

23. A speaker play control device, comprising:

a processor; and
a memory for storing instructions executable by the processor,
wherein the processor is configured to perform a method according to any one of claims 1 to 5 or a method according to any one of claims 6 to 11.

24. A computer-readable storage medium, wherein instructions are stored in the storage medium, and when the instructions in the storage medium are executed by a processor, the processor is enabled to perform a method according to any one of claims 1 to 5, or a method according to any one of claims 6 to 11.

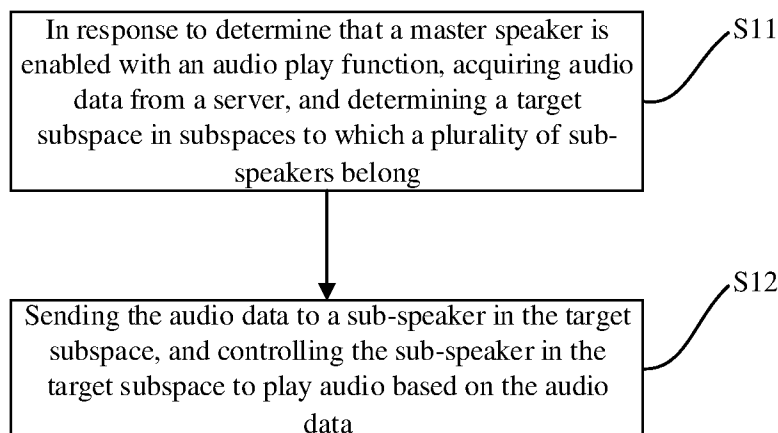


Fig. 1

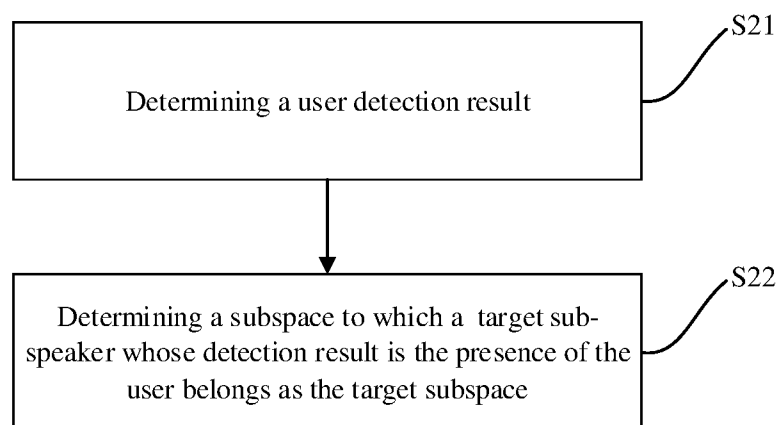


Fig. 2

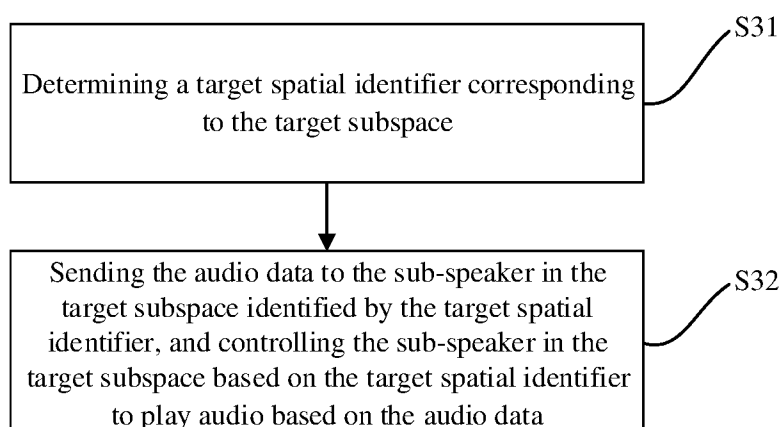


Fig. 3

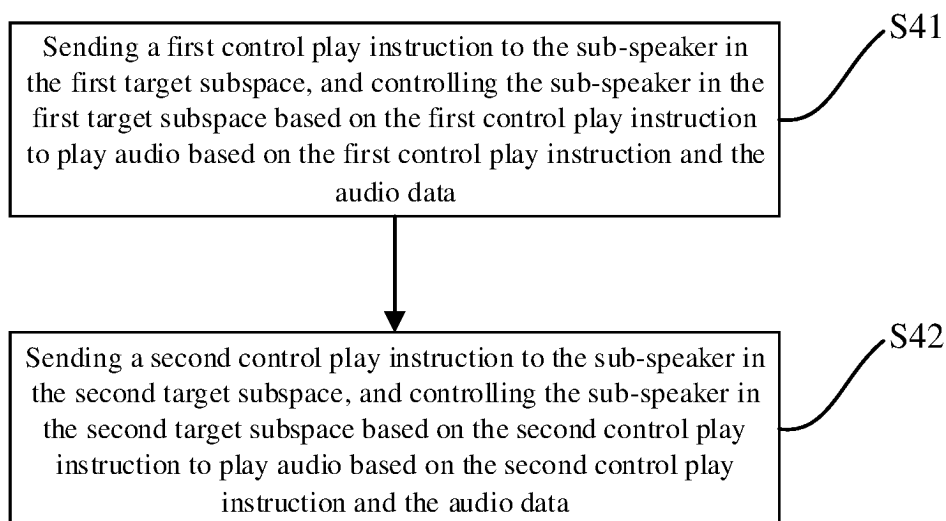


Fig. 4

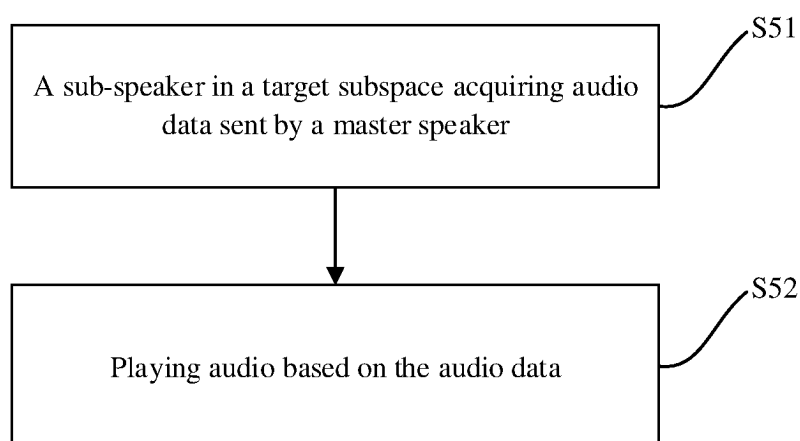


Fig. 5

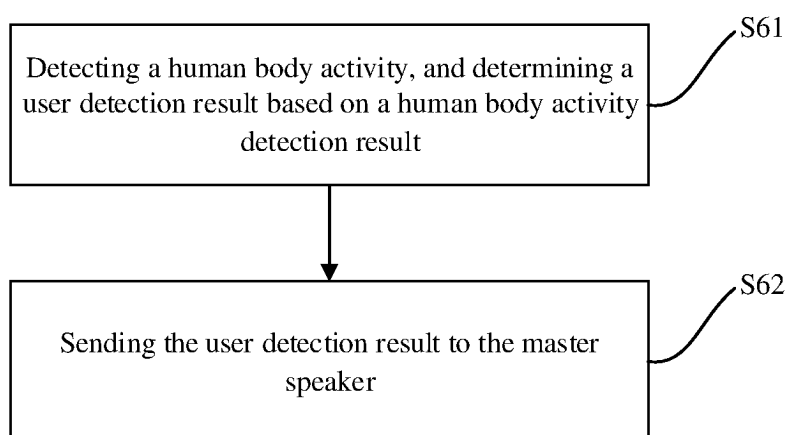


Fig. 6

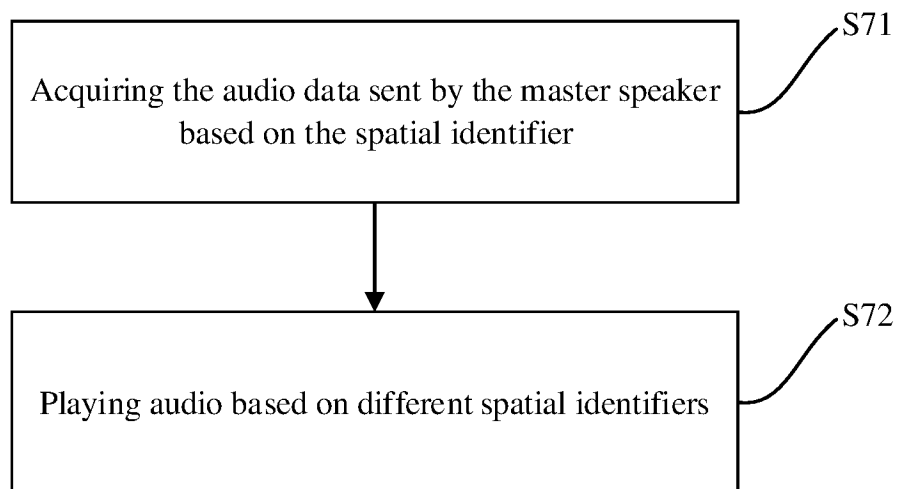


Fig. 7

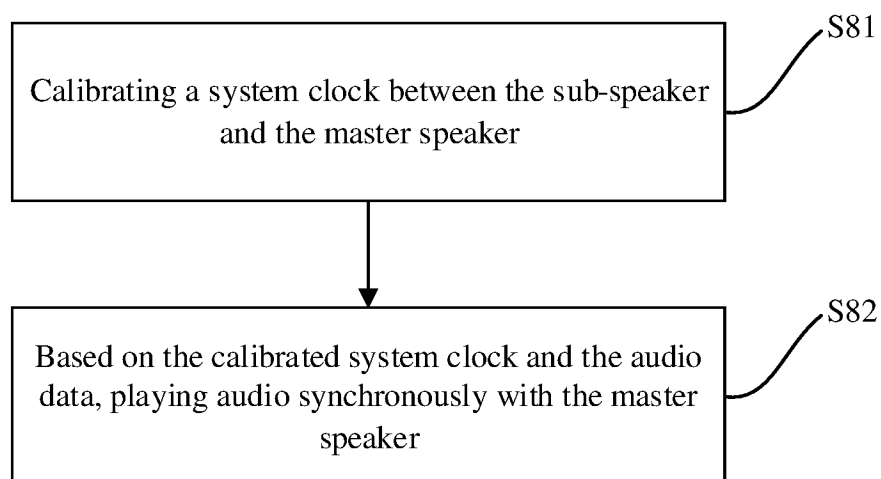


Fig. 8

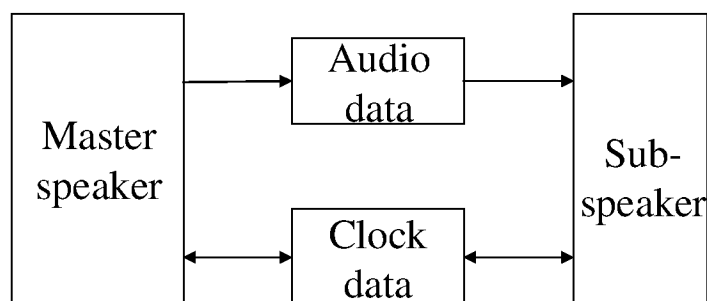


Fig. 9

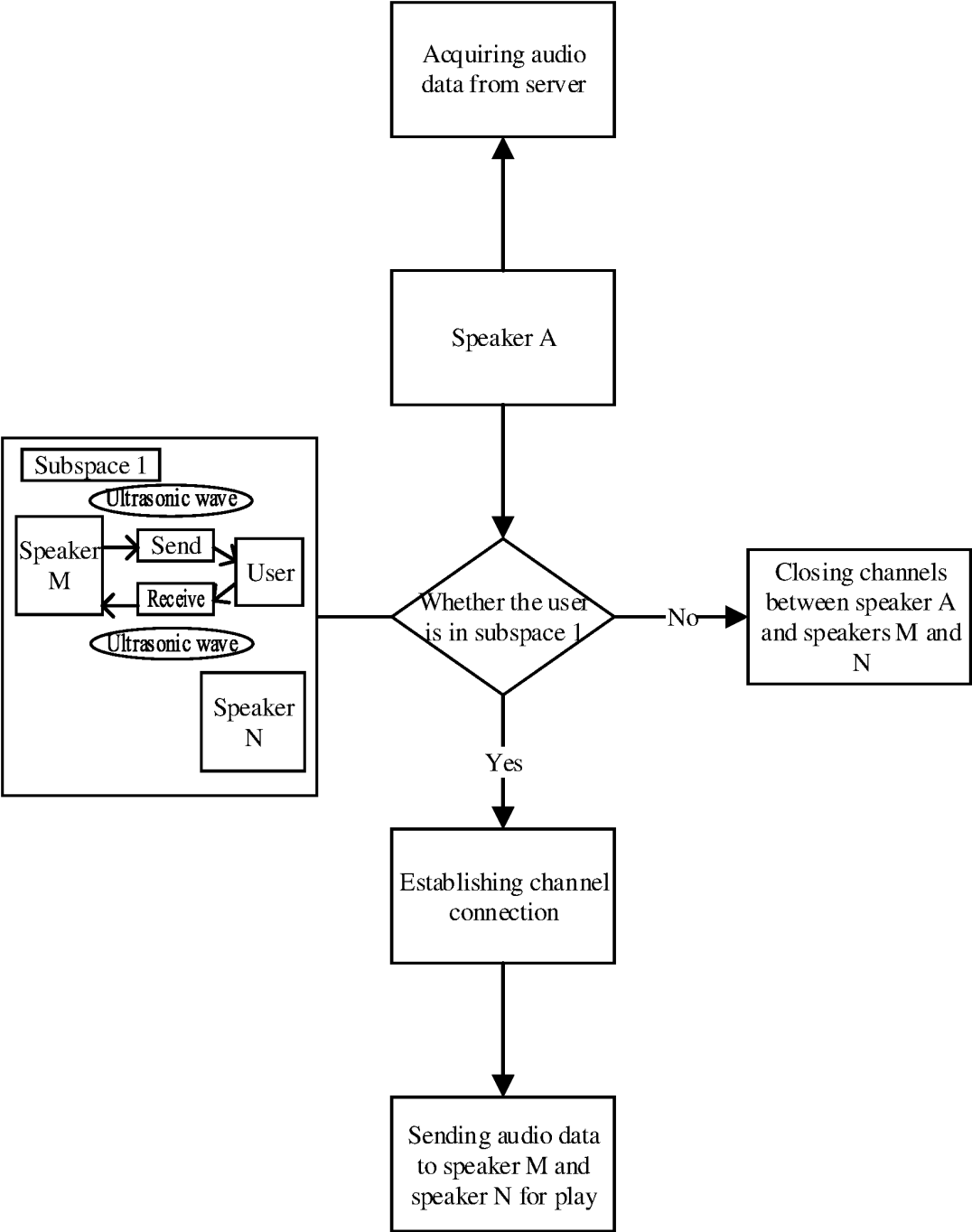


Fig. 10

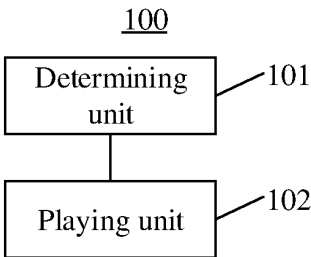


Fig. 11

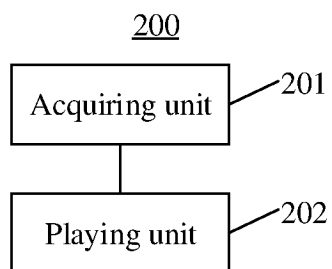


Fig. 12

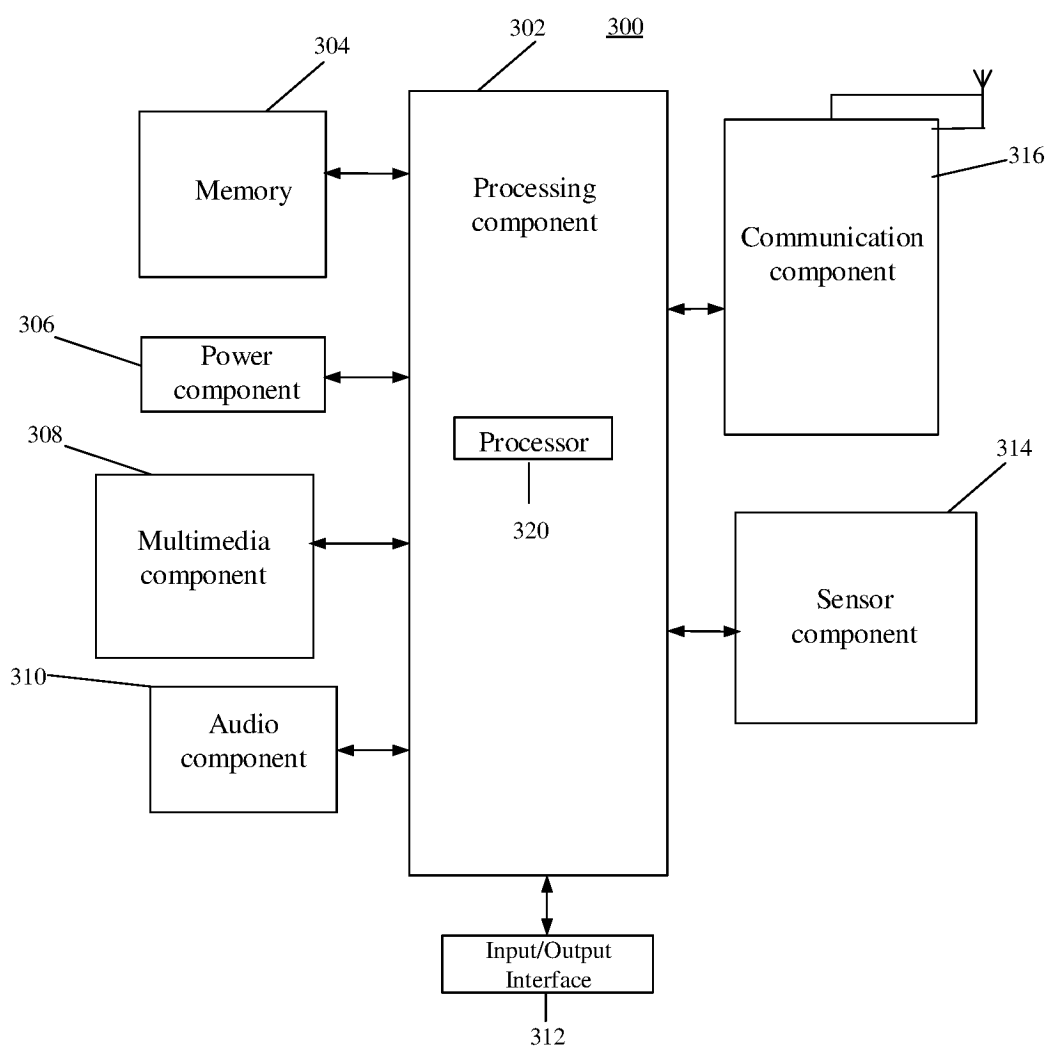


Fig. 13

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2022/099595

A. CLASSIFICATION OF SUBJECT MATTER

G10L 15/22(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)

G10L15 H04R3

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

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

CNABS, CNTXT, CNKI: 音箱, 主控, 中控, 子音箱, 从音箱, 播放, 用户, 对象, 空间, 房间, 标识; VEN, USTXT, EPTXT, WOTXT, JPTXT: loudspeaker box, speaker, master control, central control, slave loudspeaker box, play, user, object, space, room, identifier

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 113506568 A (HISENSE GROUP CO., LTD.) 15 October 2021 (2021-10-15) description, paragraphs 41-112, and figures 1-11	1-24
X	CN 112035086 A (HAIER YOUJIA INTELLIGENT TECHNOLOGY (BEIJING) CO., LTD.) 04 December 2020 (2020-12-04) description, paragraphs 36-119, and figures 1-10	1-24
A	CN 114299951 A (LENOVO (BEIJING) LTD.) 08 April 2022 (2022-04-08) entire document	1-24
A	CN 114446290 A (QINGDAO HAIER INTELLIGENT TECH RESEARCH AND DEVELOPMENT CO., LTD. et al.) 06 May 2022 (2022-05-06) entire document	1-24
A	CN 111757218 A (SHENZHEN GRANDSUN ELECTRONIC CO., LTD.) 09 October 2020 (2020-10-09) entire document	1-24
A	CN 112449285 A (GUANGZHOU DBL ELECTRONIC CO., LTD.) 05 March 2021 (2021-03-05) entire document	1-24

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

* Special categories of cited documents:

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

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

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

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

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

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

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

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

“&” document member of the same patent family

Date of the actual completion of the international search

18 November 2022

Date of mailing of the international search report

07 December 2022

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/CN2022/099595

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 201328824 A1 (ELLIPTIC LABORATORIES AS) 21 October 2021 (2021-10-21) entire document	1-24

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2022/099595

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN 113506568 A	15 October 2021	None	
CN 112035086 A	04 December 2020	None	
CN 114446290 A	06 May 2022	None	
CN 111757218 A	09 October 2020	None	
CN 112449285 A	05 March 2021	None	
US 2021328824 A1	21 October 2021	WO 2020046137 A1	05 March 2020
		EP 3844607 A1	07 July 2021
		NO 201801210 A	02 March 2020

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