

INTERACTION METHOD AND DEVICE

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims priority to Chinese Patent Application No. 201810806493.6, entitled “INTERACTION METHOD AND DEVICE” and filed on July 20, 2018, which is incorporated herein by reference in its entirety.

TECHNICAL FIELD

One or more embodiments of this specification relate to the field of electronic technologies, and in particular, to an interaction method and apparatus.

BACKGROUND

With the continuous development of intelligent technologies, electronic devices become more and more intelligent, which can interact with users to some extent and assist the users in completing related events. For example, an electronic device can complete a process of interaction with a user in the form of displaying related content on a screen, playing related content through voice, or the like.

SUMMARY

In view of this, one or more embodiments of this specification provide an interaction method and apparatus.

To achieve the foregoing objective, the one or more embodiments of this specification provide the following technical solutions.

According to a first aspect of the one or more embodiments of this specification, an interaction method is provided, including:

- detecting one or more users located in a sensing area;

- providing interaction content to the one or more users in the sensing area; and

- when a target interaction object of the interaction content is a portion of the one or more users in the sensing area, displaying information about the target interaction object to the one or more users in the sensing area.

According to a second aspect of the one or more embodiments of this specification, an interaction apparatus is provided, including:

a detection unit, configured to detect one or more users located in a sensing area;

a providing unit, configured to provide interaction content to the one or more users in the sensing area; and

a display unit, configured to: when a target interaction object of the interaction content is a portion of the one or more users in the sensing area, display information about the target interaction object to the one or more users in the sensing area.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a schematic architectural diagram of an interaction system, according to an exemplary embodiment;

FIG. 2 is a flowchart of an interaction method, according to an exemplary embodiment;

FIG. 3 is a schematic diagram of an interaction scenario, according to an exemplary embodiment;

FIG. 4 is a schematic diagram of performing interaction with an internal employee, according to an exemplary embodiment;

FIG. 5 is a schematic diagram of guiding on user positions through interaction content, according to an exemplary embodiment;

FIG. 6 is a schematic diagram of actively initiating interaction with a user by an interaction device, according to an exemplary embodiment;

FIG. 7 is another schematic diagram of guiding on user positions through interaction content, according to an exemplary embodiment;

FIG. 8 is a schematic diagram of an interaction scenario under a normal circumstance, according to an exemplary embodiment;

FIG. 9 is a schematic diagram of adjusting interaction content according to an associated event, according to an exemplary embodiment;

FIG. 10 is a schematic diagram of designating a speaker by an interaction device, according to an exemplary embodiment;

FIG. 11 is another schematic diagram of designating a speaker by an interaction device, according to an exemplary embodiment;

FIG. 12 is still another schematic diagram of designating a speaker by an interaction device, according to an exemplary embodiment;

FIG. 13 is a schematic diagram of designating a speaking order of outsiders, according to an exemplary embodiment;

FIG. 14 is a schematic diagram of marking an interaction object, according to an exemplary embodiment;

FIG. 15 is a schematic diagram of marking a target interaction object, according to an exemplary embodiment;

FIG. 16 is a schematic diagram of determining a source user of a user voice, according to an exemplary embodiment;

FIG. 17 is a schematic diagram of determining a source direction of an audio message, according to an exemplary embodiment;

FIG. 18 is a schematic diagram of marking a source user of a user voice, according to an exemplary embodiment;

FIG. 19 is a schematic structural diagram of a device, according to an exemplary embodiment; and

FIG. 20 is a block diagram of an interaction apparatus, according to an exemplary embodiment.

DETAILED DESCRIPTION

Exemplary embodiments are described in detail herein, and the embodiments are illustratively shown in the accompanying drawings. When the following descriptions are made with reference to the accompanying drawings, unless indicated otherwise, the same numbers in different accompanying drawings represent the same or similar elements. The implementations described in the following exemplary embodiments are not all the implementations consistent with one or more embodiments of this specification, but are instead only examples that are described in detail in the appended claims and are consistent with some aspects of the apparatuses and methods in one or more embodiments of this specification.

It should be noted that in other embodiments, the steps of corresponding methods are not necessarily performed according to sequences shown and described in this specification. In some other embodiments, the methods may include more or fewer steps than those described in this specification. In addition, a single step described in this specification may be divided into a plurality of steps for description in other embodiments, and a plurality of steps described in this specification may be combined into a single step for description in other embodiments.

In an embodiment, an interaction solution in this specification may be applied to an interaction device. The interaction device may be an electronic device dedicated to implement an interaction function. Alternatively, the interaction device may be a multifunctional electronic

device having an interaction function. For example, the interaction device may include a personal computer (PC), a tablet device, a notebook, a wearable device (such as smart glasses), and the like, which is not limited in one or more embodiments of this specification. During an execution process, the interaction device may execute an interaction system, to implement the interaction solution. An application program of the interaction system may be pre-installed on the interaction device, and then may be started and executed on the interaction device. Certainly, when a technology such as HTML5 is used, the interaction system may be obtained and executed without installing the application program on the interaction device.

In an embodiment, FIG. 1 is a schematic architectural diagram of an interaction system, according to an exemplary embodiment. As shown in FIG. 1, the interaction system may include a server 11, a network 12, and an interaction device 13. During an execution process, the server 11 may execute a program on the server of the interaction system, to implement related functions such as processing; and the interaction device 13 may execute a program of a client of the interaction system, to implement related functions such as information presentation and human-computer interaction, thereby implementing the interaction system through cooperation between the server 11 and the interaction device 13.

The server 11 may be a physical server including an independent host, or the server 11 may be a virtual server carried by a host cluster. The interaction device 13 may be an electronic device dedicated to implement an interaction function. Alternatively, the interaction device 13 may be a multifunctional electronic device having an interaction function. For example, the interaction device may include a PC, a tablet device, a notebook, a wearable device (such as smart glasses), and the like, which is not limited in one or more embodiments of this specification. The network 12 through which the interaction device 13 and the server 11 interact with each other may include a plurality of types of wired or wireless networks. In an embodiment, the network 12 may include a public switched telephone network (PSTN) and the Internet. It should be noted that an application program of a client of the interaction system may be pre-installed on the interaction device, and then may be started and executed on the interaction device. Certainly, when an online “client” in a technology such as HTML5 is used, the client may be obtained and executed without installing a corresponding application program on the interaction device.

In an embodiment, the interaction system may be implemented based on a mobile group office platform. The mobile group office platform may implement a communication function, and may be used as an integrated function platform of many other functions, for example, processing of an internal event of a group such as an examination and approval event (for

example, examination and approval events such as asking for time off, applying for office supplies, and financial affairs), an attendance checking event, a task event, or a log event, or processing of an external event of the group such as ordering food or purchasing items, which is not limited in one or more embodiments of this specification. Similarly, the mobile group office platform may implement the interaction system.

For example, the mobile group office platform may be carried on an instant messaging application in related technologies, for example, an enterprise instant messaging (EIM) application, such as Skype For Business®, Microsoft Teams®, Yammer®, Workplace®, Slack®, WeCom®, FXiaoKe®, Fetion enterprise®, or EasyChat enterprise®. Certainly, the instant messaging function is merely one of the communication functions supported by the mobile group office platform, and the mobile group office platform may further implement many other functions described above, which are not elaborated herein. The “group” in this specification may include organizations such as an enterprise, a school, a troop, a hospital, and an institution, and is not limited in this application.

In an embodiment, the interaction system is not limited to the mobile group office platform or a similar scenario, but may be implemented based on any other types of application, for example, a common instant messaging application, which is not limited in this specification.

FIG. 2 is a flowchart of an interaction method, according to an exemplary embodiment. As shown in FIG. 2, the method may be applied to an interaction device, and the method may include the following steps:

Step 202: detecting one or more users located in a sensing area.

In an embodiment, the interaction device has a specific sensing distance, and a coverage area of the sensing distance forms the sensing area, such as an area in a shape of a sector with a radius of 3 meters (m) (or in any other shapes). Whether there are one or more users located in the sensing area may be determined by performing detection in the sensing area.

In an embodiment, the interaction device may detect the one or more users in the sensing area in any manners. For example, the interaction device may determine whether there are one or more users in the sensing area through face detection.

Step 204: providing interaction content to the one or more users in the sensing area.

In an embodiment, the interaction device may provide the interaction content in any one or a combination of a plurality of manners, which is not limited in this specification. For example, the interaction device may include a display screen and display the interaction content on the display screen, to provide the interaction content to the one or more users in the sensing area. In another example, the interaction device may include a speaker, and play the interaction

content through voice by using the speaker, to provide the interaction content to the one or more users in the sensing area. In still another example, the interaction device may include a plurality of indicator lights, and control on/off states, colors, or flickering manners of the indicator lights, to provide the interaction content to the one or more users in the sensing area.

In an embodiment, the sensing area may include a near-field sensing area and a far-field sensing area. The near-field sensing area is closer to the interaction device than the far-field sensing area. In other words, the “near-field” and “far-field” herein are in a relative relationship. For example, a range of 0 m to 1.5 m is designated as the near-field sensing area, and a range of 1.5 m to 3 m is designated as the far-field sensing area.

In an embodiment, the interaction device may provide interaction content to a user in the near-field sensing area; and the interaction device may send guidance information to a user in the far-field sensing area, to guide the user to enter the near-field sensing area from the far-field sensing area. Then the user becomes a user in the near-field sensing area, and the interaction content is provided to the user. When the user is located in the far-field sensing area, the user may desire to interact with the interaction device by some chance. However, the far-field sensing area cannot provide a relatively good interaction effect due to a relatively far distance. Therefore, the guidance information may be sent to the user. In this way, on one hand, whether the user actually desires to implement interaction may be determined, and on the other hand, the user may obtain a better interaction effect after entering the near-field sensing area. The interaction device may send the guidance information in any one or a combination of a plurality of manners, which is not limited in this specification. For example, the interaction device may display the guidance information on the display screen. In another example, the interaction device may play the guidance information through voice by using the speaker. In still another example, the interaction device may turn on an indicator light or make the indicator light flicker, to guide the user to enter the near-field sensing area.

In an embodiment, during an attendance checking operation, the interaction device may identify the users in the sensing area, to automatically implement the attendance checking operation on a user who is located in the sensing area and has not been checked for attendance within an attendance checking time period no matter whether the user is located in the near-field sensing area or the far-field sensing area. When a stay duration of the user in the far-field sensing area reaches a first preset duration, or a stay duration of the user in the near-field sensing area reaches a second preset duration, the interaction device may determine that the user may desire to implement an interaction operation and then provide interaction content to the user. Generally, the user approaches the interaction device only when the user desires to

interact with the interaction device. Thus, the second preset duration may be appropriately less than the first preset duration, to reduce a waiting time of the user.

In an embodiment, the interaction device may actively provide interaction content to a user in the sensing area, which is similar to a “greeting” behavior between users. For example, the interaction content may include “what can I do for you,” etc., thereby guiding the user in the sensing area, to assist the user in completing a related event.

Further, the interaction device may determine whether the users in the sensing area meet a preset condition, to provide interaction content only to a user meeting the preset condition. For example, the preset condition may include at least one of the following: a stay duration in the far-field sensing area reaches the first preset duration, a stay duration in the near-field sensing area reaches the second preset duration, a user is looking at the interaction device, a user faces the interaction device or an angle between the face of the user and the interaction device is less than a preset angle, or the like, which is not limited in this specification.

In an embodiment, the interaction device may obtain an associated event of a user in the sensing area, so that when the interaction content is related to the associated event, the interaction device adjusts the interaction content provided to the user in the sensing area according to status information of the associated event. When there is no associated event related to the interaction content, default interaction content may be provided. For example, when the interaction content is related to attendance checking, for a user in the sensing area who is detected in a working period under a normal circumstance, the interaction content may be “are you sure to leave early?” When the associated event of the user in the sensing area includes a sick leave examination and approval event, and it is within a time period asked for leave involved in the sick leave examination and approval event, even in the working period under a normal circumstance, the interaction content may be adjusted to “are you sure to clock out?” In another example, when the interaction content is related to a visit of an outsider, if the user in the sensing area is an outsider and it is obtained that a corresponding associated event is a visit reservation event, the interaction content may be “would you like me to call the person you will visit?” When there is no corresponding visit reservation event, the interaction content may be “please tell me who you will visit.”

In an embodiment, the interaction device may determine identity types of the users in the sensing area, and then adjust interaction content provided to the users in the sensing area according to the identity types. For example, the identity types may include: whether the users in the sensing area are insiders belonging to a group or outsiders of the group, a department to which the users in the sensing area belong, and the like, thereby providing interaction content

meeting the identity types to the users in the sensing area.

Step 206: When a target interaction object of the interaction content is a portion of the one or more users in the sensing area, display information about the target interaction object to the one or more users in the sensing area.

In an embodiment, when there are a plurality of users in the sensing area, the target interaction object may be only a portion of the users (the number of the portion of the users may be one or more). By displaying the information about the target interaction object to the users in the sensing area, it can be ensured that each user in the sensing area can clearly learn whether himself/herself is the target interaction object of the interaction content.

In an embodiment, image information of the each user in the sensing area may be obtained, and a corresponding profile image may be generated for the each user. Then, before interaction is implemented, profile images of all the users may be shown. When interaction with the target interaction object is performed, only a profile image of the target interaction object may be shown and profile images of other users are blocked, or the profile image corresponding to the target interaction object is displayed distinctively from the profile images corresponding to the other users. For example, the distinctive display may include displaying the profile image of the target interaction object in a central area and displaying the profile images of the other users in edge areas. In another example, the distinctive display may include displaying the profile image of the target interaction object in a larger size and displaying the profile images of the other users in a normal or smaller size. In still another example, the distinctive display may include displaying the profile image of the target interaction object normally (colorfully), and graying the profile images of the other users to display. This is not limited in this specification.

In an embodiment, when the target interaction object of the interaction content is all the users (the number of the users may be one or more) in the sensing area, there is no need to display the information about the target interaction object to the users in the sensing area, so that the users more focus on the interaction content. Certainly, even the information about the target interaction object is displayed in this scenario, implementation of the technical solution of this specification is not affected.

In an embodiment, the users in the sensing area may change continuously. When the target interaction object of the interaction content changes from a portion of the users in the sensing area to all the users (for example, a user who is not the target interaction object leaves the sensing area), displaying the information about the target interaction object to the users in the sensing area may be suspended, to complete a smooth scenario transition from “the target interaction object of the interaction content is a portion of the users in the sensing area” to “the

target interaction object of the interaction content is all the users in the sensing area.”

In an embodiment, the interaction device may determine identity information of a user acting as the target interaction object and use that information as the information about the target interaction object; and then display the identity information to the users in the sensing area. For example, the interaction device may display the identity information to the users in the sensing area while providing the interaction content; or the interaction device may provide interaction content and display the identity information at different moments, which is not limited in this specification.

In an embodiment, the interaction device may identify the users in the sensing area. For example, the interaction device may identify the users in a physiological feature recognition manner such as face recognition, fingerprint recognition, iris recognition, gait recognition, or voiceprint recognition, or in any other manners, which is not limited in this specification. When an identity of a first user acting as the target interaction object is successfully identified, identity information may include an appellation (for example, a name, a nick name, a title, or other types) of the first user. For example, when the appellation of the first user is “Xiao Bai,” the interaction device may display “Xiao Bai, what can I do for you” to the users in the sensing area, where “Xiao Bai” is the information about the target interaction object, and “what can I do for you” is the interaction content. When an identity of a second user acting as the target interaction object is not successfully identified, identity information may include visual feature description information about the second user. For example, the visual feature description information may include at least one of the following: an estimated gender, an estimated height, an estimated age, a skin color, clothing, accessories, a distance to the interaction device, an orientation toward the interaction device, and the like. For example, the interaction device may display “this man in a black coat, what can I do for you?” to the users in the sensing area, where for “this man in a black coat,” an estimated gender and clothing are used as the information about the target interaction object, and “what can I do for you” is the interaction content.

In an embodiment, the interaction device may display user reference information corresponding to the users in the sensing area. The interaction device may determine user reference information corresponding to a user acting as the target interaction object and use that information as the information about the target interaction object; then, the interaction device may conspicuously display the determined user reference information to the users in the sensing area. For example, the interaction device may perform image acquisition on the users in the sensing area, and display acquired user images as the user reference information. Accordingly, the interaction device may enhance a visual effect of a user image corresponding

to the target interaction object (for example, circle the corresponding user image or display an arrow mark around the corresponding user image), or weaken a visual effect of a user image corresponding to a non-target interaction object (for example, block the content of the corresponding user image), to facilitate the users in the sensing area to learn about the target interaction object.

In an embodiment, the interaction device may obtain an event assistance request sent by a user in the sensing area, and assist in completing a corresponding event in response to the event assistance request. For example, the user in the sensing area may send an interaction voice “call the employee Xiao Hei” to the interaction device, and the interaction device may determine that the event assistance request of the user is a call request for the “employee Xiao Hei” and call the “employee Xiao Hei.” Certainly, in addition to the voice form, a user in the sensing area may send the event assistance request in other manners such as making preset body actions in space, which is not limited in this specification.

In an embodiment, the interaction device may receive response information returned by the user in the sensing area in response to the interaction content. The response information includes the event assistance request. For example, when the interaction content is “Xiao Bai, what can I do for you,” the user in the sensing area may answer “call the employee Xiao Hei,” and the interaction device may determine that the event assistance request is a call request for the “employee Xiao Hei” and call the “employee Xiao Hei.” The manner in which the interaction device provides the interaction content is not necessarily associated with the manner in which the user in the sensing area returns the response information. The two manners may be the same or different. This is not limited in this specification.

In an embodiment, when there are a plurality of users in the sensing area, the interaction device may select the users as assisted objects according to a preset sequence. Then the interaction device sequentially prompts the selected users to send corresponding event assistance requests. In this way, the users in the sensing area can sequentially send event assistance requests, avoiding a situation where the interaction device cannot accurately obtain the event assistance requests corresponding to the users due to the chaos caused when the plurality of users send the event assistance requests simultaneously, thereby helping to improve assistance efficiency and a success rate of the interaction device for the users.

In an embodiment, the interaction device may perform semantic recognition on an acquired user voice, to obtain the event assistance request; and the interaction device may perform sound feature recognition on the user voice, to determine a source user of the user voice. In this way, even a plurality of users in the sensing area speak simultaneously, the interaction device may

recognize speech content respectively corresponding to the users, and respond to corresponding event assistance requests, to improve assistance efficiency for the users. The interaction device may identify the users in the sensing area in advance, to obtain identity information of the users. When there are a plurality of users in the sensing area, the interaction device may perform sound feature recognition on an acquired user voice according to recognized sound features corresponding to the users, to determine a source user of the user voice. Compared with the practice of performing sound feature recognition on the user voice by using a full range of sound features, time consumed for the sound feature recognition may be greatly reduced.

In an embodiment, when the user in the sensing area is an insider of a group, the interaction device may assist in completing a corresponding group management event in response to the event assistance request. When the user in the sensing area is an outsider of the group, the interaction device may send an alert message to an associated insider of the group, assist the outsider of the group in establishing communication with the associated insider of the group, or guide the outsider of the group to a processing location of a visit event, in response to the event assistance request. When the user in the sensing area is an administrator, the interaction device may assist in completing a corresponding device management event in response to the event assistance request.

In an embodiment, the interaction device may receive a user voice sent by a user in the sensing area, and respond to the user voice. For example, the user in the sensing area may actively send the user voice to the interaction device, and the user voice is used for sending an event assistance request to the interaction device, giving a daily greeting to the interaction device, sending a control instruction to the interaction device, or the like, which is not limited in this specification. In another example, the user in the sensing area may send a corresponding user voice to the interaction device in response to an interaction operation implemented by the interaction device. For example, when the interaction operation implemented by the interaction device is asking whether the user in the sensing area needs help, the user voice sent by the user may tell the interaction device what kind of help is needed. This is not limited in this specification.

In an embodiment, the interaction device may perform semantic recognition on the user voice. Because the same pronunciation may correspond to a plurality of words or phrases, and distortion or noise interference may exist when the interaction device picks up the user voice, the interaction device may obtain a plurality of semantic recognition results after performing the semantic recognition on the user voice. The interaction device may score the semantic recognition results according to a predefined semantic recognition algorithm, to obtain

corresponding confidence levels. When a confidence level reaches a preset value, it indicates that the reliability of a corresponding semantic recognition result is high enough. Further, if there are a plurality of semantic recognition results with a degree of matching with the user voice reaching the preset value, the interaction device may display a plurality of corresponding semantic recognition result options to the user in the sensing area for the user to select and accurately express a true intention of the user, and then respond to the user voice according to a semantic recognition result corresponding to a selected semantic recognition result option. In a selection process, the user may speak out a semantic recognition result corresponding to a semantic recognition result option the user desires to select, or speak out a position (for example, “the first one” or “the leftmost one”) of a semantic recognition result option the user desires to select, which is not limited in this specification.

In an embodiment, the interaction device may determine a source direction of the user voice, and respond to a user located in the source direction of the user voice. In one situation, after determining the source direction of the user voice, the interaction device directly determines by default that only the user sending the user voice exists in the source direction, so that the interaction device may directly respond to the source direction of the user voice, for example, playing an interaction voice. In another situation, the interaction device may determine one or more users existing in the source direction of the user voice. If there are a plurality of users, the interaction device may further determine a source user of the user voice, to respond to the source user.

In an embodiment, a microphone array is built in the interaction device, and the user voice may be received by using the microphone array. The microphone array includes a first microphone disposed on the left relatively and a second microphone disposed on the right relatively. The source direction of the user voice is determined according to a difference between the time at which the first microphone and the second microphone receive the user voice. For example, when a user in the sensing area is located on a side close to the left, the first microphone may receive the user voice earlier than the second microphone, and when a user in the sensing area is located on a side close to the right, the second microphone may receive the user voice earlier than the first microphone. For a specific solution of how to determine a source direction of a user voice according to a receiving time difference, reference may be made to related technical solutions in existing technologies, and details are not described herein.

In an embodiment, when there are a plurality of users located in the source direction of the user voice, the interaction device may determine the source user of the user voice according to

a facial action (for example, the facial action is obtained through image acquisition by using a camera built in the interaction device) of each of the plurality of users, and respond to the source user. The facial action of the user may include actions of one or more parts such as cheek, mouth, and chin, which is not limited in this specification. Using mouth actions as an example, when there are a plurality of users in the source direction of the user voice, but only one user performs a mouth open-close action, the user may be determined as the source user of the user voice. In another example, although the plurality of users all perform a mouth open-close action, but only one user's action matches the user voice in the number of times and in amplitudes of opening and closing, the user may be determined as the source user of the user voice.

In an embodiment, when the interaction device is mounted on a wall, a user generally can pass only before the interaction device and send a user voice. When the interaction device is mounted in other manners, a user may appear before or behind the interaction device, so that an audio message acquired by the interaction device may come from a user located before or behind the interaction device. If there is a user in the sensing area of the interaction device, and another user speaks exactly when passing behind the interaction device, the interaction device may mistakenly consider that the user in the sensing area sends a user voice. Therefore, after receiving an audio message, the interaction device may determine whether the audio message is a user voice sent by a user in the sensing area based on a source direction of the audio message and whether there is a user in the sensing area.

For example, a microphone array is built in the interaction device, and the microphone array includes a third microphone disposed relatively close to the sensing area and a fourth microphone disposed relatively far from the sensing area. When an audio message is received by using the microphone array, a source direction of the audio message may be determined according to how much the third microphone and the fourth microphone receive a high frequency part in the audio message. When the source direction is relatively close to the sensing area, because the high frequency part in the audio message is affected by absorption of the housing of the interaction device, a high frequency part of the audio message received by the fourth microphone is less than a high frequency part of the audio message received by the third microphone. When the source direction is relatively far from the sensing area, because the high frequency part in the audio message is also affected by absorption of the housing of the interaction device, a high frequency part of the audio message received by the third microphone is less than a high frequency part of the audio message received by the fourth microphone. Therefore, the source direction of the audio message may be accurately determined according

to how much the third microphone and the fourth microphone receive the high frequency part in the audio message.

When the source direction is the side relatively close to the sensing area and there is a user in the sensing area, the interaction device may determine that the audio message is a user voice sent by the user in the sensing area. Otherwise, when the source direction is the side relatively far from the sensing area, or when the source direction is the side relatively close to the sensing area but there is no user in the sensing area, the interaction device may determine that the audio message is not a user voice sent by a user in the sensing area.

With respect to the first microphone, the second microphone, the third microphone, and the fourth microphone, it should be noted that, to distinguish microphones in a left-right direction, different microphones are distinguished as the first microphone and the second microphone, and to distinguish microphones in a front-rear direction, different microphones are distinguished as the third microphone and the fourth microphone. However, a number of microphones actually included in the microphone array is not limited in this specification. For example, the microphone array may include one or more first microphones and one or more second microphones. In another example, the microphone array may include one or more third microphones and one or more fourth microphones. In addition, when the microphone array needs to distinguish source directions in the left-right direction and the front-rear direction at the same time, the microphone array does not necessarily need to include four microphones. In other words, the first microphone, the second microphone, the third microphone, and the fourth microphone are merely roles microphones play when implementing related functions. In practice, the microphone array may include fewer microphones. For example, the microphone array may include three microphones, where a microphone 1 and a microphone 2 are located at the same row in the front-rear direction but spaced apart in the left-right direction, and a microphone 3 is located before or behind the microphone 1 and the microphone 2, so that the microphones 1 to 3 form a position relationship similar to a triangle. The microphone 1 and the microphone 2 may be used as a first microphone and a second microphone to distinguish a source direction of a user voice in the left-right direction, the microphone 1 and the microphone 2 may be used as a third microphone, and the microphone 3 may be used as a fourth microphone (the microphone 3 is located behind the microphone 1 and the microphone 2). Alternatively, the microphone 3 may be used as a third microphone, and the microphone 1 and the microphone 2 may be used as a fourth microphone (the microphone 3 is located before the microphone 1 and the microphone 2) to distinguish a source direction of an audio message in the front-rear direction.

In an embodiment, the interaction device may detect a number of the users in the sensing area, for example, performing acquisition by using a camera and then performing face detection and counting, which is not limited in this specification. When there are a plurality of users in the sensing area, the interaction device may respectively display profile images corresponding to the users, to represent the plurality of users. When a user enters or leaves the sensing area or the number of the users changes, the profile images displayed by the interaction device may change accordingly. When the interaction device receives an audio message and determines that the audio message is from a user in the sensing area, the interaction device may display a profile image of the source user of the user voice distinctively from profile images of other users, so that the user may determine that the interaction device has successfully received the user voice and distinguished the source user by viewing the change of the profile images without worrying that the interaction device fails to receive the user voice or mistakenly recognizes the source user. The profile image of the source user of the user voice may be displayed distinctively from the profile images of the other users in any manners, which is not limited in this specification. For example, the profile image of the source user of the user voice may be displayed in a central area and the profile images of the other users may be displayed in edge areas. In another example, the profile image of the source user of the user voice may be displayed in a larger size and the profile images of the other users may be displayed in a normal or smaller size. In still another example, the profile image of the source user of the user voice may be displayed normally (colorfully) and the profile images of the other users may be grayed and then displayed.

For ease of understanding, technical solutions of one or more embodiments of this specification are described by using the enterprise instant messaging application “WeCom” as an example. FIG. 3 is a schematic diagram of an interaction scenario, according to an exemplary embodiment. As shown in FIG. 3, it is assumed that an interaction device 3 is disposed in an office of an enterprise AA, and a WeCom client is executed on the interaction device 3, so that the interaction device 3 may implement the interaction solution of this specification based on the WeCom client.

In an embodiment, a camera 31 is mounted on the interaction device 3, and the camera 31 may form a corresponding shooting area 32 as a sensing area corresponding to the interaction device 3. Accordingly, the interaction device 3 may determine a user entering the shooting area 32, for example, a user 4 entering the shooting area 32 in FIG. 3, according to an image obtained by shooting the shooting area 32 by using the camera 31. Certainly, in addition to using the camera 31, the interaction device 3 may determine a user entering the sensing area through

sound detection, infrared detection, or other methods, which is not limited in this specification.

FIG. 4 is a schematic diagram of performing interaction with an internal employee, according to an exemplary embodiment. As shown in FIG. 4, the interaction device 3 may be equipped with a screen 33, and the screen 33 may be configured to display a user image 41 corresponding to the user 4 and acquired by the camera 31. The interaction device 3 may identify the user 4, for example, performing face recognition based on a face image acquired by the camera 31, which is not limited in this specification. Assuming that the interaction device 3 recognizes that the user 4 is an internal employee “Xiao Bai,” corresponding identity information 42 may be shown on the screen 33. For example, the identity information 42 may be an appellation “Xiao Bai” of the user 4.

During an attendance checking time period, after the interaction device 3 recognizes that the user 4 is the internal employee “Xiao Bai,” attendance checking data of the internal employee “Xiao Bai” may be queried, and an attendance checking operation for the internal employee “Xiao Bai” may be automatically implemented if attendance has not been checked. To make the user 4 learn that the attendance checking operation has been completed, the interaction device 3 may provide corresponding interaction content to the user 4. For example, the interaction content may include a label 43 shown on the screen 33, and information included in the label 43 is “clock in,” which indicates that the type of the attendance checking operation is “clock in.” The interaction content may alternatively be provided to the user 4 in other forms. For example, when the interaction device 3 includes a speaker 34, voice information such as “Xiao Bai, you have clocked in successfully” may be played by using the speaker 34. Similarly, the interaction device 3 may implement automatic attendance checking operations of clocking in for other internal employees of the enterprise AA, and the interaction device 3 may also implement automatic attendance checking operations of clocking out for the internal employees of the enterprise AA. Details are not described herein.

In an embodiment, the sensing area of the interaction device 3 may be divided into a plurality of sub-areas according to distances to the interaction device 3. For example, in FIG. 3, the shooting area 32 is divided into a far-field shooting area 321 (having a distance to the interaction device 3 of 1.5 m to 3.0 m) and a near-field shooting area 322 (having a distance to the interaction device 3 of 0 m to 1.5 m). If it is within the attendance checking time period currently, the interaction device 3 may implement the automatic attendance checking operation for the user 4 no matter whether the user is located in the far-field shooting area 321 or the near-field shooting area 322. If it is within the attendance checking time period but the user 4 has been checked for attendance, or if it is within another time period, when the user 4 is located

in the far-field shooting area 321, the interaction device 3 may determine by default that the user 4 merely temporarily passes by and does not have an interaction intention, so that the interaction device may not initiate interaction with the user 4 actively (that is, not provide interaction content to the user 4). However, if a stay duration of the user 4 in the far-field shooting area 321 reaches a first preset duration (for example, 3 seconds (s)), the interaction device 3 may determine that the user 4 has an interaction intention, so that the interaction device may provide interaction content to the user 4. Similarly, if it is within the attendance checking time period but the user 4 has been checked for attendance, or if it is within another time period, when the user 4 is located in the near-field shooting area 322, the interaction device 3 may determine by default that the user 4 merely temporarily passes by and does not have an interaction intention, so that the interaction device may not initiate interaction with the user 4 actively (that is, not provide interaction content to the user 4). However, if a stay duration of the user 4 in the near-field shooting area 322 reaches a second preset duration, the interaction device 3 may determine that the user 4 has an interaction intention, so that the interaction device may provide interaction content to the user 4. Because the near-field shooting area 322 is closer to the interaction device 3, the behavior that the user 4 actively enters the near-field shooting area 322 may include an interaction intention, so that the second preset duration may be appropriately less than the first preset duration. For example, the first preset duration is 3s and the second preset duration is 1s. In a relatively special situation, the second preset duration may be 0, which means that the interaction device 3 determines by default that the user 4 has an interaction intention once entering the near-field shooting area 322, so that interaction content may be provided to the user 4 immediately.

When the user 4 is located in the far-field shooting area 321, to ensure effective communication between the interaction device 3 and the user 4 and to improve interaction efficiency, the interaction device 3 may guide the user 4 to move to the near-field shooting area 322 from the far-field shooting area 321 through interaction content. For example, FIG. 5 is a schematic diagram of guiding on user positions through interaction content, according to an exemplary embodiment. As shown in FIG. 5, the interaction device 3 may show interaction content 511 in a text form in an interaction display area 51 on the screen 33. For example, the interaction content 511 is “please get closer to be within 1.5 meters,” to guide the user 4 to move to the near-field shooting area 322 from the far-field shooting area 321. In addition to the interaction content 511 in a text form, the interaction device 3 may play interaction content in a voice form, for example, “Xiao Bai, you are a little far from me” (“Xiao Bai” is identity information, and “you are a little far from me” is the interaction content) by using the speaker

34, to guide the user 4 to move to the near-field shooting area 322 from the far-field shooting area 321. In addition, the interaction device 3 may further control an indicator light 35 to flicker regularly, to attract attention of the user 4, which is equivalent to conveying interaction content to the user 4, to guide the user 4 to move to the near-field shooting area 322 from the far-field shooting area 321. Certainly, the interaction device 3 may convey interaction content in any one of the manners such as the text form, the voice form, and the light form, which is not limited in this specification.

When the user 4 enters the near-field shooting area 322 (actively or under the guidance), the interaction device 3 may guide the user 4 to speak out an interaction purpose of the user. For example, FIG. 6 is a schematic diagram of actively initiating interaction with a user by an interaction device, according to an exemplary embodiment. As shown in FIG. 6, the interaction device 3 may play interaction content in a voice form by using the speaker 34, for example, “Xiao Bai, what can I do for you?” (“Xiao Bai” is identity information, and “what can I do you for” is the interaction content). In addition, the interaction device 3 may display interaction content 512 in a text form in the interaction display area 51. For example, the interaction content 512 may be “try to say something like this” and “call Zhang San,” to guide the user 4 to express an interaction purpose to the interaction device 3 in a voice form.

It should be noted that, the interaction device 3 does not necessarily need to guide the user 4 from the far-field shooting area 321 to the near-field shooting area 322. For example, the interaction device 3 may alternatively directly guide the user 4 in the far-field shooting area 321 to speak out the interaction purpose of the user. For example, the interaction device 3 may further perform detection on noise from a surrounding environment. When a noise level is greater than a preset value, the user 4 is first guided from the far-field shooting area 321 to the near-field shooting area 322, and then the user 4 is guided to speak out the interaction purpose of the user. When the noise level is less than the preset value, the user 4 in the far-field shooting area 321 is directly guided to speak the interaction purpose of the user.

In an embodiment, when the user 4 is an internal employee of the enterprise AA, the interaction device 3 may obtain the appellation “Xiao Bai” of the user 4, so that the interaction device 3 may guide the user 4 from the far-field shooting area 321 to the near-field shooting area 322 by using “Xiao Bai, you are a little far from me” shown in FIG. 5, where the appellation “Xiao Bai” is used as identity information, and “you are a little far from me” is interaction content. For an outsider of the enterprise AA, the interaction device 3 may fail to obtain a corresponding appellation. Therefore, identity information used in an interaction process may be different from that in the embodiment shown in FIG. 5. For example, FIG. 7 is

another schematic diagram of guiding on user positions through interaction content, according to an exemplary embodiment. As shown in FIG. 7, it is assumed that the interaction device 3 acquires a user image 71 of a user by using the camera 31, but the user is an outsider of the enterprise AA, and the interaction device 3 fails to obtain an appellation of the user. Thus, when guiding the user to move from the far-field shooting area 321 to the near-field shooting area 322, the interaction device 3 may display interaction content 513 in a text form in the interaction display area 51. For example, the interaction content 513 is “please get closer to be within 1.5 meters.” The interaction device 3 may alternatively play interaction content in a voice form, for example, “hello, you are a little far from me” (identity information of the user is omitted) by using the speaker 34. The interaction device 3 may alternatively control the indicator light 35 to flicker regularly, to guide the user to move to the near-field shooting area 322 from the far-field shooting area 321.

In an embodiment, the interaction device 3 may learn an associated event of a user in the sensing area by accessing a server of WeCom, and may change to-be-provided interaction content based on the associated event. For example, FIG. 8 is a schematic diagram of an interaction scenario under a normal circumstance, according to an exemplary embodiment. As shown in FIG. 8, assuming that the interaction device 3 detects the user 4 located in the shooting area 32 in a working time period and recognizes that the user 4 is the internal employee “Xiao Bai” of the enterprise AA, if the interaction device 3 determines that an interaction purpose of the user 4 is to clock out but it is still within the working time period, when the interaction device 3 does not find an associated event related to clocking out of the user 4, interaction content 514 shown in the interaction display area 51 may be “are you sure to leave early?” FIG. 9 is a schematic diagram of adjusting interaction content according to an associated event, according to an exemplary embodiment. As shown in FIG. 9, assuming that the interaction device 3 detects the user 4 located in the shooting area 32 in a working time period and recognizes that the user 4 is the internal employee “Xiao Bai” of the enterprise AA, if the interaction device 3 finds that the user 4 has a sick leave examination and approval event that has been submitted and it is within a sick leave time approved in the sick leave examination and approval event, interaction content 515 shown in the interaction display area 51 may be “are you sure to clock out?”

In an embodiment, there may be a plurality of users in the shooting area 32 of the interaction device 3, and the interaction device 3 may implement communication with these users by using proper interaction content. FIG. 10 is a schematic diagram of designating a speaker by an interaction device, according to an exemplary embodiment. As shown in FIG.

10, assuming that there are a plurality of users in the sensing area 32 respectively corresponding to user images 81 to 82 shown on the screen 33, the interaction device 3 may recognize a user corresponding to the user image 81 as “Xiao Bai” and a user corresponding to the user image 82 as “Xiao Hei,” and display appellations of the users as identity information around corresponding user images. For example, identity information 91 of a corresponding user shown above the user image 81 is the appellation “Xiao Bai,” and identity information 92 of a corresponding user shown above the user image 82 is the appellation “Xiao Hei.” Because of a limited interaction capability of the interaction device 3, and to enable the interaction device 3 to clearly learn interaction purposes of users, the interaction device 3 may interact with only a portion of the users at a time. The interaction device 3 may select a target interaction object (the portion of the users) in a specific manner, for example, in ascending order of distances between the users and the interaction device 3, ascending order of angles between faces of the users and a shooting direction of the camera 31, or descending order of heights of the users, which is not limited in this specification. Assuming that the interaction device 3 intends to interact with the user “Xiao Bai” corresponding to the user image 81, to avoid misunderstanding of other users in the sensing area, the interaction device 3 needs to express that a target interaction object corresponding to interaction content is the user “Xiao Bai” to the users in the sensing area when providing the interaction content. For example, when the interaction device 3 plays interaction content “what can I do for you” by using the speaker 34, the identity information of the user “Xiao Bai” may be additionally played, so that actual played content may be “Xiao Bai, what can I do for you?” for the other users to be clear that the target interaction object of the interaction content “what can I do for you?” is “Xiao Bai.”

FIG. 11 is another schematic diagram of designating a speaker by an interaction device, according to an exemplary embodiment. As shown in FIG. 11, if a plurality of users such as a user “Xiao Bai” and a user “Xiao Hei” speak out their interaction purposes at the same time, the interaction device 3 may not accurately learn the interaction purpose of each user due to chaos of sound, or the interaction device 3 may not respond to the interaction purposes of the plurality of users at the same time, or there are other reasons, so that the interaction device 3 may provide interaction content to guide these users to express their interaction purposes sequentially. For example, the interaction device 3 may show interaction content 516 in the interaction display area 51. The interaction content 516 may include “please speak one by one.” Further, when the interaction device 3 determines, according to the foregoing embodiment, a speaking order of the users is that the user “Xiao Bai” speaks first and the user “Xiao Hei” speaks later, the interaction device 3 may additionally play the identity information of the user

“Xiao Bai” while playing interaction content “I cannot hear clearly. xx, please speak first” by using the speaker 34, so that actual played content may be “I cannot hear clearly. Xiao Bai, please speak first,” for other users to learn that a target interaction object of the interaction content is the user “Xiao Bai.”

In addition to the voice form in the embodiments shown in FIG. 10 and FIG. 11, a speaking order of a plurality of users may be designated in a plurality of other manners. For example, FIG. 12 is still another schematic diagram of designating a speaker by an interaction device, according to an exemplary embodiment. As shown in FIG. 12, when the interaction device 3 determines a speaking order of the users is that the user “Xiao Bai” speaks first and the user “Xiao Hei” speak later, the interaction device 3 may mark the user image 81 corresponding to the user “Xiao Bai”, for example, adding a marking box 810 to a face area. In this way, even interaction content is “what can I do for you,” “please speak,” or the like, the users may learn that a target interaction object of the interaction content is the user “Xiao Bai.” Certainly, in the embodiment shown in FIG. 12, when the interaction device 13 shows an interaction text 517 in the interaction display area 51, the interaction text 517 further includes the identity information of the user “Xiao Bai” in addition to interaction content “xx, please speak,” so that full content of the interaction text 517 is “Xiao Bai, please speak,” which may also indicate to the users that the current target interaction object is the user “Xiao Bai.”

FIG. 13 is a schematic diagram of a speaking order of outsiders, according to an exemplary embodiment. As shown in FIG. 13, it is assumed that users corresponding to user images of 81 and 82 are both outsiders of the enterprise AA, the interaction device 3 cannot obtain appellations of these users, but identity information of the users may be expressed in other manners to specify a target interaction object corresponding to interaction information. For example, when the interaction device 3 determines that the target interaction object is the user corresponding to the user image 81, if the user image 81 corresponds to a female user and the user image 82 corresponds to a male user, the identity information of the users may be expressed through genders such as “this lady” or “this gentleman.” Therefore, when voice content played by the interaction device 3 by using the speaker 34 is “I cannot hear clearly. Let’s have this lady speak first please,” both users in the shooting area 32 may determine that the interaction content is “I cannot hear clearly. xx, please speak first,” and determine that the target interaction object is the user corresponding to the user image 81 based on the identity information of “this lady.”

When interaction content is provided to a user in the shooting area 32, in some situations, the user does not need to respond to the interaction content, such as the interaction content “you

have clocked in successfully” in the embodiment shown in FIG. 4. In some other situations, the user may respond to the interaction content, and the response may include an event assistance request initiated by the user, for the interaction device 3 to assist the user in completing a corresponding event. For example, for the interaction content “are you sure to clock out?” in the embodiment shown in FIG. 9, when a response returned by the user “Xiao Bai” is “yes,” the interaction device 3 may determine, based on semantic analysis, that the user “Xiao Bai” initiates an event assistance request for an “attendance checking event of clocking out,” so that the interaction device 3 may assist in completing the “attendance checking event of clocking out.” Similarly, for example, in a multi-user scenario, after the interaction device 3 sends a voice “I cannot hear clearly. Let’s have this lady speak first please” in FIG. 13, if a response returned by the female user is “call Baibai,” the interaction device 3 may determine, based on semantic analysis, that the user “Xiao Bai” initiates an event assistance request for a “call event” and a called object is a user “Baibai,” so that the interaction device 3 may call the user “Baibai,” to assist in completing the “call event.”

Certainly, in addition to the situations of responding to the interaction content, the user in the shooting area 32 may directly initiate an event assistance request to the interaction device 3, and the interaction device 3 may assist in completing in a corresponding event. This is similar to the situations of responding to the interaction content, and details are not repeated herein.

In the embodiments shown in FIG. 10 to FIG. 13, the interaction device 3 may ensure that the plurality of users in the shooting area 32 speak sequentially, so that the interaction device 3 may determine the event assistance requests initiated by the users respectively, to assist in completing corresponding events respectively. In an embodiment, the interaction device 3 may receive a plurality of user voices sent by a plurality of users at the same time, and accurately separate the plurality of user voices based on sound features, and may further determine a mapping relationship between each user voice and a user in the shooting area 32 through sound feature recognition (for example, voiceprint recognition), so that the interaction device 3 can simultaneously learn the event assistance requests of the plurality of users and assist in completing the corresponding events, greatly improving assistance efficiency for the plurality of users.

In one situation, the interaction device 3 may directly compare an acquired user voice with a sound feature database, which, for example, may include voiceprint features of all internal employees in the enterprise AA, so as to determine an internal employee corresponding to the acquired user voice based on a comparison result. In addition, the interaction device 3 may identify the user in the shooting area 32 in other manners such as face recognition, and compare

an identification result with the comparison result obtained based on the sound feature database, to prevent the internal employee of the enterprise AA from being pretended. For example, it is determined through sound features that user voices are from a user A and a user B in the enterprise AA respectively, but it is determined based on face recognition that users in the shooting area are the user A and an outsider that cannot be recognized. Then the outsider may pretend the user B and use a recorded voice of the user B. In this situation, the interaction device 3 may refuse to complete a corresponding assistance event, and send a warning to the user B.

In another situation, the interaction device 3 may first identify the users in the shooting area 32 in a manner such as face recognition, to recognize that the users in the shooting area 32 are the user A and the user B in the enterprise AA. Then, when the interaction device 3 acquires two user voices, the interaction device may only compare the user voices with voiceprint features of the user A and the user B, to determine which user voice is from the user A and which user voice is from the user B without comparing the user voices with other voiceprint features in the sound feature database, thereby greatly improving the comparison efficiency.

In addition, the users in the shooting area 32 may include an administrator, and the interaction device 3 may assist, in response to an event assistance request of the administrator, in completing a corresponding device management event, for example, adjusting welcome speech content on the screen 33, adjusting a volume of the speaker 34, or adjusting ranges of the far-field shooting area 321 and the near-field shooting area 322.

FIG. 14 is a schematic diagram of marking an interaction object, according to an exemplary embodiment. After shooting the shooting area 32 by using the camera 31, the interaction device 3 may mark a detected user located in the shooting area 32, for the user to clearly determine whether the user is detected by the interaction device 3 and whether the user can interact with the interaction device 3. As shown in FIG. 14, when detecting that there is a user in the shooting area 32, the interaction device 3 may generate a corresponding profile image 1401 for the user according to a shot image, and display the profile image 1401 on the screen 33. When another user is also detected by the interaction device 3 to be located in the shooting area 32, a profile image 1402 corresponding to the user may also be displayed on the screen 33. Similarly, when still other users enter the shooting area 32, the interaction device 3 may also display corresponding profile images on the screen 33, and details are not repeated herein.

When the user corresponding to the profile image 1402 leaves the shooting area 32, the interaction device 3 may delete the profile image 1402 from the screen. Situations of other

users are similar to that of the user, and details are not repeated herein.

Therefore, when seeing that the screen 33 includes the profile images 1401 and 1402, the corresponding users may be sure that themselves have been detected by the interaction device 3 and determined by the interaction device 3 as interaction objects, and that they can interact with the interaction device 3. When another user desires to interact with the interaction device 3, but does not see a profile image corresponding to himself/herself on the screen 33, it indicates that the user may have not entered the shooting area 32, or the user enters the shooting area 32 but is not detected by the interaction device 3 successfully, and the user may take measures like entering or re-entering the shooting area 32 until a profile image of the user is displayed on the screen 33.

FIG. 15 is a schematic diagram of marking a target interaction object, according to an exemplary embodiment. It is assumed that the interaction device 3 recognizes a user “Xiao Bai” and a user “Xiao Hei” located in the shooting area 32 respectively, and the interaction device 3 determines the user “Xiao Bai” as a target interaction object. As shown in FIG. 15, the interaction device 3 may display a profile image 1401 corresponding to the user “Xiao Bai” in a central area (relatively far from edges of the screen 33) of the screen 33 in a normal proportion, and display a profile image 1402 corresponding to the user “Xiao Hei” in an edge area of the screen 33 in a smaller proportion. Then, when the interaction device 3 sends an interaction voice of “what can I do for you” by using the speaker 34, according to display proportions and display positions of the profile images 1401 and 1402, it may be determined that the target interaction object corresponding to the interaction voice is the user “Xiao Bai” corresponding to the profile image 1401 rather than the user “Xiao Hei” corresponding to the profile image 1402.

Certainly, in addition to configuring a display proportion and a display position of a profile image, other aspects of adjustments may be performed on a display attribute of the profile image, to distinguish the profile image corresponding to the target interaction object from the profile images of the other interaction objects, helping to determine the corresponding target interaction object according to the profile images. This is not limited in this specification.

In addition to the practice of initiating interaction with a user in the shooting area 32 by the interaction device 3, the user in the shooting area 32 may initiate interaction with the interaction device 3, for example, send a user voice to the interaction device 3, for the interaction device 3 to respond to the user voice to meet the requirement of a source user of the user voice. The user voice may be used for responding to an interaction voice sent by the interaction device 3, or may be actively sent by the user in the shooting area 32 to the interaction

device 3, which is not limited in this specification.

In an embodiment, there may be a plurality of users in the shooting area 32 at the same time, so that when receiving a user voice sent by a user in the shooting area 32, the interaction device 3 needs to determine a source user of the user voice, that is, to detect which of the plurality of users in the shooting area 32 sends the user voice.

For example, FIG. 16 is a schematic diagram of determining a source user of a user voice, according to an exemplary embodiment. As shown in FIG. 16, a microphone array may be built in the interaction device 3, the microphone array may include a microphone 36 and a microphone 37, and the microphone 36 is disposed relatively on the left and the microphone 37 is disposed relatively on the right. Therefore, when a user in the shooting area 32 sends a user voice such as “I need to reserve a conference room for 15 persons,” if the microphone 36 receives the user voice earlier than the microphone 37, it indicates that a source user of the user voice is relatively close to the microphone 36 and relatively far from the microphone 37. Therefore, it may be determined that the source user is located relatively close to the left in the shooting area 32. For example, it may be determined that the source user is the user “Xiao Bai” in combination with the image acquired in FIG. 10.

Similarly, if the microphone 37 receives the user voice earlier than the microphone 36, it indicates that a source user of the user voice is relatively close to the microphone 37 and relatively far from the microphone 36. Therefore, it may be determined that the source user is located relatively close to the right in the shooting area 32. For example, it may be determined that the source user is the user “Xiao Hei” in combination with the image acquired in FIG. 10. Alternatively, if the microphone 36 and microphone 37 receive the user voice at the same time or nearly at the same time, it indicates that a source user of the user voice is located at a middle point between the microphone 36 and the microphone 37, i.e., located right before the interaction device 3. Therefore, it may be determined that the source user is located at a middle position of the shooting area 32.

In an embodiment, the user may be located behind the interaction device 3 based on an installation position or installation manner of the interaction device 3, rather than located before an orientation of the screen 33 and the camera 31. Accordingly, the user is obviously not located in the shooting area 32, but the user may be exactly located around the interaction device 3. In this situation, when the interaction device 3 receives an audio message such as “I need to reserve a conference room for 15 persons,” the audio message is not necessarily a user voice sent by a user in the shooting area 32, but an interference voice sent by the user located behind the interaction device. Therefore, to prevent the interference voice from being mistakenly

determined as a user voice, a source direction of the audio message needs to be determined. An audio message from the front may be the user voice sent by the user in the shooting area 32, and an audio message from the rear is an interference voice.

For example, FIG. 17 is a schematic diagram of determining a source direction of an audio message, according to an exemplary embodiment. As shown in FIG. 17, a microphone array may be built in the interaction device 3, and the microphone array may include a microphone 36, a microphone 37, and a microphone 38. In a left-right direction (that is, a horizontal direction in FIG. 17), the microphone 36 is disposed on the left side, and the microphone 37 is disposed on the right side. In a front-rear direction (that is, a top-bottom direction in FIG. 17), the microphone 36 and the microphone 37 are disposed in a front part of the interaction device 3 that are relatively close to a video area 32, and the microphone 38 is disposed in a rear part of the interaction device 3 that is relatively far from the video area 32. The microphones 36 to 38 are all located in the interaction device 3. Therefore, for an audio message sent by a user located around the interaction device 3, if the user is located before the interaction device 3, the audio message is transmitted through the interaction device 3 from the front of the interaction device 3, a high frequency part of the audio message is absorbed by a housing part of the interaction device 3, and when the microphones 36 to 38 receive the high frequency part of the audio message, an intensity of a high-frequency signal received by the microphone 38 located in the rear part of the interaction device 3 is caused to be less than an intensity of a high-frequency signal received by each of the microphones 36 and 37 due to the absorption of the housing of the interaction device 3. If the user is located behind the interaction device 3, the audio message is transmitted through the interaction device 3 from the rear of the interaction device 3, the high frequency part of the audio message is absorbed by the housing part of the interaction device 3, and when the microphones 36 to 38 receive the high frequency part of the audio message, an intensity of a high-frequency signal received by each of the microphones 36 and 37 located in the front of the interaction device 3 is caused to be less than an intensity of a high-frequency signal received by the microphone 38 due to the absorption of the housing of the interaction device 3.

Therefore, it may be determined whether the source direction of the audio message is the front or the rear of the interaction device 3 according to how much the microphones 36 to 38 receive the high frequency part of the audio message. When it is determined that the source direction of the audio message is the rear of the interaction device 3, the source user of the audio message is definitely not a user in the shooting area 32. In other words, the audio message is an interference voice. When it is determined that the source direction of the audio message

is the front of the interaction device 3, the source user of the audio message may be a user in the shooting area 32. Certainly, to improve accuracy and reduce a probability of mistaken determination, further determination may be performed in combination with other conditions:

In an embodiment, image acquisition may be performed by using the camera 31 on the interaction device 3, and if there is a user in the shooting area 32, it may be determined that the audio message is from the user.

In an embodiment, image acquisition may be performed by using the camera 31 on the interaction device 3, and if there are a plurality of users in the shooting area 32, a user with a facial action matching the audio message is determined as a source user of the audio message with reference to facial actions of the users, for example, whether a mouth open-close action is performed in a processing of receiving the audio message and whether the moment at which the action occurs is consistent with the moment when a signal change of the audio message occurs.

In an embodiment, image acquisition may be performed by using the camera 31 on the interaction device 3, and if there are a plurality of users in the shooting area 32, a user corresponding to a corresponding direction may be determined as a source user of the audio message with reference to whether a source direction of the audio message determined by the microphones 36 and 37 is the left, the right, or the middle. If there are still a plurality of users in the same direction, a user with a facial action matching the audio message may be screened out with reference to facial actions of the users, and determined as a source user of the audio message.

FIG. 18 is a schematic diagram of marking a source user of a user voice, according to an exemplary embodiment. As shown in FIG. 18, assuming that the user “Xiao Bai” is determined as the source user of the user voice, the interaction device 3 may display the profile image 1401 in an original colorful mode, and gray the profile image 1402 and the like corresponding to the other users to display, for the users in the shooting area 32 to quickly determine whether the interaction device 3 accurately recognizes the source user of the user voice, to ensure that no error occurs in a subsequent interaction process.

In an embodiment, when the interaction device 3 recognizes the user voice, unfavorable factors such as an accent of the source user, excessive noise of an external environment, or distortion occurring during a sound picking up process all affect accuracy of semantic recognition of the interaction device 3. Therefore, in the recognition process, the interaction device 3 may score each candidate semantic recognition result. The interaction device 3 may directly discard a candidate semantic recognition result whose confidence level is relatively

low (for example, lower than a preset score). If the number of candidate semantic recognition results whose confidence levels are relatively high (for example, higher than a preset score) is one, the candidate semantic recognition result may be directly used as a semantic recognition result. If the number of candidate semantic recognition results whose confidence levels are relatively high is more than one, the interaction device 3 may display, to the source user, candidate semantic recognition result options corresponding to the candidate semantic recognition results whose confidence levels are relatively high, for example, an option 1801 “1. I need to reserve a conference room for 15 persons” and an option 1802 “2. I need to reserve a conference room for 45 persons” in FIG. 18, for the source user “Xiao Bai” to select and confirm.

For example, the user “Xiao Bai” may send a confirmation voice including content such as “the first,” “the former,” or “the one for 15 persons,” to tell the interaction device 3 that the option 1801 is selected. The interaction device 3 then may determine that a semantic recognition result corresponding to the user voice is “I need to reserve a conference room for 15 persons” and further respond to the user voice, for example, assist the user “Xiao Bai” in completing reservation for a relevant conference room.

FIG. 19 is a schematic structural diagram of a device, according to an exemplary embodiment. Referring to FIG. 19, at a hardware level, the device includes a processor 1902, an internal bus 1904, a network interface 1906, an internal memory 1908, and a non-volatile memory 1910, and certainly may further include hardware required for other services. The processor 1902 reads a corresponding computer program from the non-volatile memory 1910 into the internal memory 1908 and then executes the computer program, to form an interaction apparatus at a logic level. Certainly, in addition to a software implementation, one or more embodiments of this specification do not exclude other implementations, for example, a logical device or a software-hardware combination. In other words, execution bodies of the following processing procedures are not limited to logical units and may alternatively be hardware or logical devices.

Referring to FIG. 20, in a software implementation, the interaction apparatus may include:
a detection unit 2001, configured to detect one or more users located in a sensing area;
a providing unit 2002, configured to provide interaction content to the one or more users in the sensing area; and

a first display unit 2003, configured to: when a target interaction object of the interaction content is a portion of the one or more users in the sensing area, display information about the target interaction object to the one or more users in the sensing area.

Optionally, the first display unit 2003 is configured to:
determine identity information of a user acting as the target interaction object as the information about the target interaction object; and

display the identity information to the one or more users in the sensing area.

Optionally, the interaction apparatus further includes:

an identification unit 2004, configured to identify the one or more users located in the sensing area, where

when an identity of a first user acting as the target interaction object is successfully identified, identity information of the first user includes an appellation of the first user; and when an identity of a second user acting as the target interaction object is not successfully identified, identity information of the second user includes visual feature description information about the second user.

Optionally, the interaction apparatus further includes:

a second display unit 2005, configured to display user reference information corresponding to the one or more users in the sensing area, where

the first display unit 2003 is configured to: determine user reference information corresponding to a user acting as the target interaction object as the information about the target interaction object; and conspicuously display the determined user reference information to the one or more users in the sensing area.

Optionally, the second display unit 2005 is configured to:

perform image acquisition on the one or more users in the sensing area, and display one or more acquired user images as the user reference information.

Optionally, the interaction apparatus further includes:

a management unit 2006, configured to: when the target interaction object of the interaction content changes from the portion of the one or more users in the sensing area to all the one or more users, suspend displaying the information about the target interaction object to the one or more users in the sensing area.

Optionally, the interaction apparatus further includes:

a request obtaining unit 2007, configured to obtain an event assistance request sent by a user in the sensing area; and

an assisting unit 2008, configured to assist in completing a corresponding event in response to the event assistance request.

Optionally, the request obtaining unit 2007 is configured to:

receive response information returned by the user in the sensing area in response to the

interaction content, where the response information includes the event assistance request.

Optionally, the request obtaining unit 2007 is configured to:

when there are a plurality of users in the sensing area, select users as assisted objects according to a preset sequence; and

prompt the selected users sequentially to send corresponding event assistance requests.

Optionally, the request obtaining unit 2007 is configured to:

perform semantic recognition on an acquired user voice, to obtain the event assistance request; and

perform sound feature recognition on the user voice, to determine a source user of the user voice.

Optionally, the assisting unit 2008 is configured to:

when the user in the sensing area is an insider of a group, assist in completing a corresponding group management event in response to the event assistance request;

when the user in the sensing area is an outsider of the group, send an alert message to an associated insider of the group, assist the outsider of the group in establishing communication with the associated insider of the group, or guide the outsider of the group to a processing location of a visit event, in response to the event assistance request; and

when the user in the sensing area is an administrator, assist in completing a corresponding device management event in response to the event assistance request.

Optionally, the sensing area includes a near-field sensing area and a far-field sensing area; and the providing unit 2002 is configured to:

provide interaction content to a user in the near-field sensing area; and

send guidance information to a user in the far-field sensing area, to guide the user to enter the near-field sensing area from the far-field sensing area.

Optionally, the interaction apparatus further includes:

an event obtaining unit 2009, configured to obtain an associated event of the one or more users in the sensing area, where

the providing unit 2002 is configured to: when the interaction content is related to the associated event, adjust the interaction content provided to the one or more users in the sensing area according to status information of the associated event.

Optionally, the interaction apparatus further includes:

a determining unit 2010, configured to determine an identity type of the one or more users in the sensing area, where

the providing unit 2002 is configured to: adjust the interaction content provided to the one

or more users in the sensing area according to the identity type.

Optionally, the interaction apparatus further includes:

a voice receiving unit 2011, configured to receive a user voice sent by a user in the sensing area; and

a responding unit 2012, configured to respond to the user voice.

Optionally, the responding unit 2012 is configured to:

perform semantic recognition on the user voice;

when confidence levels of a plurality of semantic recognition results reach a preset value, display a plurality of corresponding semantic recognition result options to the user in the sensing area; and

respond to the user voice according to a semantic recognition result corresponding to a selected semantic recognition result option.

Optionally, the responding unit 2012 is configured to:

determine a source direction of the user voice; and

respond to a user located in the source direction of the user voice.

Optionally, the responding unit 2012 determines the source direction of the user voice in the following manner:

receiving the user voice by using a microphone array, where the microphone array includes a first microphone disposed on the left relatively and a second microphone disposed on the right relatively; and

determining the source direction of the user voice according to a difference between time at which the first microphone and the second microphone receive the user voice.

Optionally, the responding unit 2012 responds to the user located in the source direction of the user voice in the following manner:

when there are a plurality of users located in the source direction of the user voice, determining a source user of the user voice according to a facial action of each of the plurality of users; and

responding to the source user.

Optionally, the interaction apparatus further includes:

an audio receiving unit 2013, configured to receive an audio message by using a microphone array, where the microphone array includes a third microphone disposed relatively close to the sensing area and a fourth microphone disposed relatively far from the sensing area;

a direction determining unit 2014, configured to determine a source direction of the audio message according to how much the third microphone and the fourth microphone receive a

high frequency part in the audio message; and

a source determining unit 2015, configured to: when the source direction is a side relatively close to the sensing area and there is a user in the sensing area, determine that the audio message is a user voice sent by the user in the sensing area.

Optionally, the interaction apparatus further includes:

a profile image display unit 2016, configured to: when there are a plurality of users in the sensing area, display profile images corresponding to the users respectively; and

a distinctive display unit 2017, configured to display a profile image of a source user of the user voice distinctively from profile images of other users.

The system, the apparatus, the module, or the unit described in the foregoing embodiments can be implemented by a computer chip or an entity or implemented by a product having a specific function. A typical implementation device is a computer. The form of the computer may be a PC, a laptop computer, a cellular phone, a camera phone, a smartphone, a personal digital assistant, a media player, a navigation device, an email transceiver device, a game console, a tablet computer, a wearable device, or any combination thereof.

In a typical configuration, the computer includes one or more processors (such as CPUs), an input/output interface, a network interface, and an internal memory.

The internal memory may include forms such as a volatile memory, a random access memory (RAM) and/or a non-volatile memory in computer-readable media, for example, a read-only memory (ROM) or a flash RAM. The internal memory is an example of the computer-readable medium.

The computer-readable medium includes a persistent medium and a non-persistent medium, a removable medium and a non-removable medium, which may implement storage of information by using any method or technology. The information may be a computer-readable instruction, a data structure, a program module, or other data. Examples of a storage medium of a computer include, but are not limited to, a phase-change memory (PRAM), a static random-access memory (SRAM), a dynamic random access memory (DRAM), a RAM of another type, a ROM, an electrically erasable programmable read-only memory (EEPROM), a flash memory or another memory technology, a compact disc read-only memory (CD-ROM), a digital versatile disc (DVD) or another optical storage device, a cassette tape, a disk storage, a quantum memory, a graphene-based storage medium or another magnetic storage device, or any other non-transmission media, which may be configured to store information that can be accessed by a computing device. According to the definition in this specification, the computer-readable medium does not include transitory computer-readable media (transitory media) such

as a modulated data signal and a carrier.

It should also be noted that the terms “include,” “comprise” and any other variants mean to cover the non-exclusive inclusion. Thereby, the process, method, article, or device which includes a series of elements not only includes those elements, but also includes other elements which are not clearly listed, or further includes the inherent elements of the process, method, article, or device. Without further limitation, the element defined by a phrase “include one...” does not exclude other same elements in the process, method, article or device which include the element.

Some embodiments of this specification are described above. Other embodiments fall within the scope of the appended claims. In some situations, the actions or steps recorded in the claims may be performed in sequences different from those in the embodiments and an expected result may still be achieved. In addition, the processes depicted in the accompanying drawings do not necessarily require specific sequences or consecutive sequences to achieve an expected result. In some implementations, multitasking and parallel processing may be feasible or beneficial.

The terms used in one or more embodiments of this specification are merely used to describe the embodiments but are not intended to limit one or more embodiments of this specification. The “a,” “the,” and “said” in a singular form used in one or more embodiments of this specification and the appended claims are also intended to include a plural form, unless other meanings are clearly indicated in the context. It should also be understood that, the term “and/or” used herein indicates and includes any or all possible combinations of one or more associated listed items.

It should be understood that although the terms “first,” “second,” “third,” and the like may be used to describe various information in one or more embodiments of this specification, but such information should not be limited to these terms. These terms are merely used to distinguish between information of the same type. For example, without departing from the scope of one or more embodiments of this specification, the first information may also be referred to as the second information. Similarly, the second information may also be referred to as the first information. Depending on the context, the term “if” as used herein may be interpreted as “when ...” or “upon ...” or “in response to determining.”

The foregoing descriptions are merely exemplary embodiments of one or more embodiments of this specification, but are not intended to limit the one or more embodiments of this specification. Any modification, equivalent replacement, or improvement made within the spirit and principle of one or more embodiments of this specification shall fall within the

protection scope of the one or more embodiments of this specification.

CLAIMS

What is claimed is:

1. An interaction method, comprising:
detecting one or more users located in a sensing area;
providing interaction content to the one or more users in the sensing area; and
when a target interaction object of the interaction content is a portion of the one or more users in the sensing area, displaying information about the target interaction object to the one or more users in the sensing area.
2. The method according to claim 1, wherein the displaying information about the target interaction object to the one or more users in the sensing area comprises:
determining identity information of a user acting as the target interaction object as the information about the target interaction object; and
displaying the identity information to the one or more users in the sensing area.
3. The method according to claim 2, further comprising:
identifying the one or more users located in the sensing area, wherein
when an identity of a first user acting as the target interaction object is successfully identified, identity information of the first user comprises an appellation of the first user; and
when an identity of a second user acting as the target interaction object is not successfully identified, identity information of the second user comprises visual feature description information about the second user.
4. The method according to claim 1, further comprising:
displaying user reference information corresponding to the one or more users in the sensing area, wherein
the displaying information about the target interaction object to the one or more users in the sensing area comprises: determining user reference information corresponding to a user acting as the target interaction object as the information about the target interaction object; and conspicuously displaying the determined user reference information to the one or more users in the sensing area.
5. The method according to claim 4, wherein the displaying user reference information corresponding to the one or more users in the sensing area comprises:
performing image acquisition on the one or more users in the sensing area, and displaying one or more acquired user images as the user reference information.
6. The method according to claim 1, further comprising:

when the target interaction object of the interaction content changes from the portion of the one or more users in the sensing area to all the one or more users, suspending displaying the information about the target interaction object to the one or more users in the sensing area.

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

obtaining an event assistance request sent by a user in the sensing area; and

assisting in completing a corresponding event in response to the event assistance request.

8. The method according to claim 7, wherein the obtaining an event assistance request sent by a user in the sensing area comprises:

receiving response information returned by the user in the sensing area in response to the interaction content, wherein the response information comprises the event assistance request.

9. The method according to claim 7, wherein the obtaining an event assistance request sent by a user in the sensing area comprises:

when there are a plurality of users in the sensing area, selecting users as assisted objects according to a preset sequence; and

prompting the selected users sequentially to send corresponding event assistance requests.

10. The method according to claim 7, wherein the obtaining an event assistance request sent by a user in the sensing area comprises:

performing semantic recognition on an acquired user voice, to obtain the event assistance request; and

performing sound feature recognition on the user voice, to determine a source user of the user voice.

11. The method according to claim 7, wherein the assisting in completing a corresponding event in response to the event assistance request comprises:

when the user in the sensing area is an insider of a group, assisting in completing a corresponding group management event in response to the event assistance request;

when the user in the sensing area is an outsider of the group, sending an alert message to an associated insider of the group, assisting the outsider of the group in establishing communication with the associated insider of the group, or guiding the outsider of the group to a processing location of a visit event, in response to the event assistance request; and

when the user in the sensing area is an administrator, assisting in completing a corresponding device management event in response to the event assistance request.

12. The method according to claim 1, wherein the sensing area comprises a near-field sensing area and a far-field sensing area; and the providing interaction content to the one or more users in the sensing area comprises:

providing interaction content to a user in the near-field sensing area; and
sending guidance information to a user in the far-field sensing area, to guide the user to enter the near-field sensing area from the far-field sensing area.

13. The method according to claim 1, further comprising:
obtaining an associated event of the one or more users in the sensing area, wherein
the providing interaction content to the one or more users in the sensing area comprises:
when the interaction content is related to the associated event, adjusting the interaction content provided to the one or more users in the sensing area according to status information of the associated event.

14. The method according to claim 1, further comprising:
determining an identity type of the one or more users in the sensing area, wherein
the providing interaction content to the one or more users in the sensing area comprises:
adjusting the interaction content provided to the one or more users in the sensing area according to the identity type.

15. The method according to claim 1, further comprising:
receiving a user voice sent by a user in the sensing area; and
responding to the user voice.

16. The method according to claim 15, wherein the responding to the user voice comprises:
performing semantic recognition on the user voice;
when confidence levels of a plurality of semantic recognition results reach a preset value,
displaying a plurality of corresponding semantic recognition result options to the user in the sensing area; and

responding to the user voice according to a semantic recognition result corresponding to a selected semantic recognition result option.

17. The method according to claim 15, wherein the responding to the user voice comprises:
determining a source direction of the user voice; and
responding to a user located in the source direction of the user voice.

18. The method according to claim 17, wherein the determining a source direction of the user voice comprises:

receiving the user voice by using a microphone array, wherein the microphone array comprises a first microphone disposed on the left relatively and a second microphone disposed on the right relatively; and

determining the source direction of the user voice according to a difference between time at which the first microphone and the second microphone receive the user voice.

19. The method according to claim 17, wherein the responding to a user located in the source direction of the user voice comprises:

when there are a plurality of users located in the source direction of the user voice, determining a source user of the user voice according to a facial action of each of the plurality of users; and

responding to the source user.

20. The method according to claim 15, further comprising:

receiving an audio message by using a microphone array, wherein the microphone array comprises a third microphone disposed relatively close to the sensing area and a fourth microphone disposed relatively far from the sensing area;

determining a source direction of the audio message according to how much the third microphone and the fourth microphone receive a high frequency part in the audio message; and

when the source direction is a side relatively close to the sensing area and there is a user in the sensing area, determining that the audio message is a user voice sent by the user in the sensing area.

21. The method according to claim 15, further comprising:

when there are a plurality of users in the sensing area, displaying profile images corresponding to the users respectively; and

displaying a profile image of a source user of the user voice distinctively from profile images of other users.

22. An interaction apparatus, comprising:

a detection unit, configured to detect one or more users located in a sensing area;

a providing unit, configured to provide interaction content to the one or more users in the sensing area; and

a first display unit, configured to: when a target interaction object of the interaction content is a portion of the one or more users in the sensing area, display information about the target interaction object to the one or more users in the sensing area.

23. The apparatus according to claim 22, wherein the first display unit is configured to:

determine identity information of a user acting as the target interaction object as the information about the target interaction object; and

display the identity information to the one or more users in the sensing area.

24. The apparatus according to claim 23, further comprising:

an identification unit, configured to identify the one or more users located in the sensing area, wherein

when an identity of a first user acting as the target interaction object is successfully identified, identity information of the first user comprises an appellation of the first user; and when an identity of a second user acting as the target interaction object is not successfully identified, identity information of the second user comprises visual feature description information about the second user.

25. The apparatus according to claim 22, further comprising:

a second display unit, configured to display user reference information corresponding to the one or more users in the sensing area, wherein

the first display unit is configured to: determine user reference information corresponding to a user acting as the target interaction object as the information about the target interaction object; and conspicuously display the determined user reference information to the one or more users in the sensing area.

26. The apparatus according to claim 25, wherein the second display unit is configured to:

perform image acquisition on the one or more users in the sensing area, and display one or more acquired user images as the user reference information.

27. The apparatus according to claim 22, further comprising:

a management unit, configured to: when the target interaction object of the interaction content changes from the portion of the one or more users in the sensing area to all the one or more users, suspend displaying the information about the target interaction object to the one or more users in the sensing area.

28. The apparatus according to claim 22, further comprising:

a request obtaining unit, configured to obtain an event assistance request sent by a user in the sensing area; and

an assisting unit, configured to assist in completing a corresponding event in response to the event assistance request.

29. The apparatus according to claim 28, wherein the request obtaining unit is configured to:

receive response information returned by the user in the sensing area in response to the interaction content, wherein the response information comprises the event assistance request.

30. The apparatus according to claim 28, wherein the request obtaining unit is configured to:

when there are a plurality of users in the sensing area, select users as assisted objects according to a preset sequence; and

prompt the selected users sequentially to send corresponding event assistance requests.

31. The apparatus according to claim 28, wherein the request obtaining unit is configured to:

perform semantic recognition on an acquired user voice, to obtain the event assistance request; and

perform sound feature recognition on the user voice, to determine a source user of the user voice.

32. The apparatus according to claim 28, wherein the assisting unit is configured to:

when the user in the sensing area is an insider of a group, assist in completing a corresponding group management event in response to the event assistance request;

when the user in the sensing area is an outsider of the group, send an alert message to an associated insider of the group, assist the outsider of the group in establishing communication with the associated insider of the group, or guide the outsider of the group to a processing location of a visit event, in response to the event assistance request; and

when the user in the sensing area is an administrator, assist in completing a corresponding device management event in response to the event assistance request.

33. The apparatus according to claim 22, wherein the sensing area comprises a near-field sensing area and a far-field sensing area; and the providing unit is configured to:

provide interaction content to a user in the near-field sensing area; and

send guidance information to a user in the far-field sensing area, to guide the user to enter the near-field sensing area from the far-field sensing area.

34. The apparatus according to claim 22, further comprising:

an event obtaining unit, configured to obtain an associated event of the one or more users in the sensing area, wherein

the providing unit is configured to: when the interaction content is related to the associated event, adjust the interaction content provided to the one or more users in the sensing area according to status information of the associated event.

35. The apparatus according to claim 22, further comprising:

a determining unit, configured to determine an identity type of the one or more users in the sensing area, wherein

the providing unit is configured to: adjust the interaction content provided to the one or more users in the sensing area according to the identity type.

36. The apparatus according to claim 22, further comprising:

a voice receiving unit, configured to receive a user voice sent by a user in the sensing area; and

a responding unit, configured to respond to the user voice.

37. The apparatus according to claim 36, wherein the responding unit is configured to:
perform semantic recognition on the user voice;

when confidence levels of a plurality of semantic recognition results reach a preset value,
display a plurality of corresponding semantic recognition result options to the user in the
sensing area; and

respond to the user voice according to a semantic recognition result corresponding to a
selected semantic recognition result option.

38. The apparatus according to claim 36, wherein the responding unit is configured to:
determine a source direction of the user voice; and
respond to a user located in the source direction of the user voice.

39. The apparatus according to claim 38, wherein the responding unit determines the
source direction of the user voice in the following manner:

receiving the user voice by using a microphone array, wherein the microphone array
comprises a first microphone disposed on the left relatively and a second microphone disposed
on the right relatively; and

determining the source direction of the user voice according to a difference between time
at which the first microphone and the second microphone receive the user voice.

40. The apparatus according to claim 38, wherein the responding unit responds to the user
located in the source direction of the user voice in the following manner:

when there are a plurality of users located in the source direction of the user voice,
determining a source user of the user voice according to a facial action of each of the plurality
of users; and

responding to the source user.

41. The apparatus according to claim 36, further comprising:

an audio receiving unit, configured to receive an audio message by using a microphone
array, wherein the microphone array comprises a third microphone disposed relatively close to
the sensing area and a fourth microphone disposed relatively far from the sensing area;

a direction determining unit, configured to determine a source direction of the audio
message according to how much the third microphone and the fourth microphone receive a
high frequency part in the audio message; and

a source determining unit, configured to: when the source direction is a side relatively
close to the sensing area and there is a user in the sensing area, determine that the audio message
is a user voice sent by the user in the sensing area.

42. The apparatus according to claim 36, further comprising:

a profile image display unit, configured to: when there are a plurality of users in the sensing area, display profile images corresponding to the users respectively; and

a distinctive display unit, configured to display a profile image of a source user of the user voice distinctively from profile images of other users.