



- (51) **International Patent Classification:**
G01S 5/02 (2010.01) *G01S 5/20* (2006.01)
- (21) **International Application Number:**
PCT/CN2018/105500
- (22) **International Filing Date:**
13 September 2018 (13.09.2018)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (71) **Applicant:** **ALIBABA GROUP HOLDING LIMITED**
[—/CN]; Fourth Floor, One Capital Place, P.O. Box 847,
George Town, Grand Cayman (KY).
- (72) **Inventors:** **FENG, Jinwei**; 500 108th Av NE, #800, Belle-
vue, Washington 98004 (US). **LI, Xinguo**; B3F, Radiance

JinHui Tower, Building No.6, The 4th District Wangjing-
DongYuan, Chaoyang District, Beijing 100102 (CN).

- (74) **Agent:** **BEIJING TSINGYUANHUI INTELLECTUAL
PROPERTY LAW FIRM**; Room 362, 3rd Floor, Suzhou
Street No.1, Haidian District, Beijing 100080 (CN).

- (81) **Designated States** (*unless otherwise indicated, for every
kind of national protection available*): AE, AG, AL, AM,
AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ,
CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO,
DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN,
HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP,
KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME,
MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ,
OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA,
SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN,
TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(54) **Title:** METHODS AND DEVICES FOR PROCESSING AUDIO INPUT USING UNIDIRECTIONAL AUDIO INPUT DEVICES

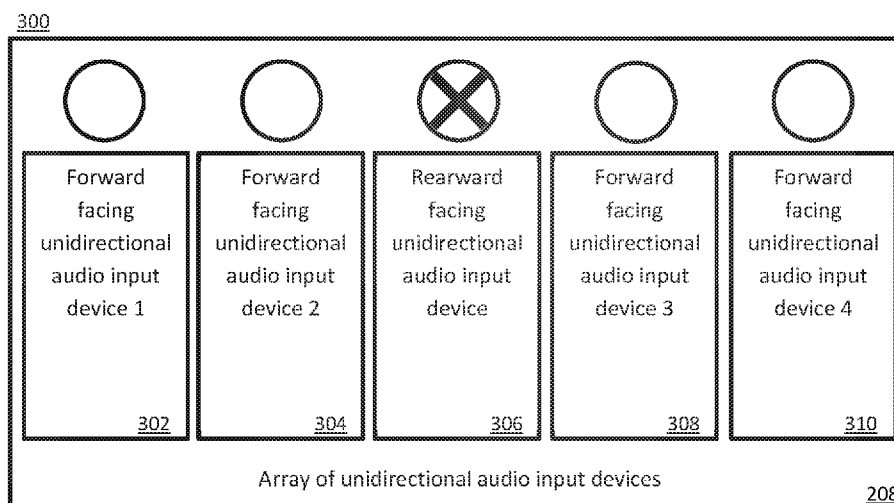


FIG. 3

(57) **Abstract:** A method for processing audio input includes receiving audio inputs from a plurality of unidirectional audio input devices, including at least two unidirectional audio input devices oriented to face forward and at least one unidirectional audio input device oriented to face rearward; analyzing the received audio inputs from the plurality of unidirectional audio input devices to obtain audio information corresponding to the received audio inputs; comparing the audio information corresponding to the audio input from at least one of the at least two forward facing unidirectional audio input devices and the audio information corresponding to the audio input from the rearward facing unidirectional audio input device; obtaining, from the comparison, a subset of sections of the audio information corresponding to the received audio inputs from the sound source; determining a position of the sound source based on the audio information corresponding to the audio inputs from the plurality of unidirectional audio input devices and the subset of sections of the audio information; and filtering the audio information corresponding to the audio inputs from the plurality of unidirectional audio input devices based on the subset of sections of the audio information and the position of the sound source..

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

METHODS AND DEVICES FOR PROCESSING AUDIO INPUT USING UNIDIRECTIONAL AUDIO INPUT DEVICES

TECHNICAL FIELD

[1] The present disclosure relates to the technical field of communications and, more particularly, to methods and devices for processing audio input using unidirectional audio input devices.

BACKGROUND

[2] With the development of smart terminals with audio input devices, a user may receive various services in environments with high levels of background noise, such as ordering a beverage, reserving a table, and the like. To use such services, a user of a terminal associated with the service often needs to speak their request, and the smart terminal needs to receive and accurately interpret the request.

[3] Conventionally, the reception and interpretation of the service request in environments with high levels of background noise relies on filtering out the background noise by employing a linear array of audio input devices, (e.g., microphones) to accurately receive and interpret the request. Filtering methods, however, are unable to determine if the desired sound source is in the front or in the back of the smart terminals, and are therefore unable to determine if the sound source is the user of the smart terminal or if the sound source is a nonuser speaking behind the terminal. This inability to determine the relative position of the sound source to the smart terminal results in inefficiencies in reception and inaccurate interpretation.

SUMMARY

[4] The present disclosure provides a method for processing audio input. The method includes receiving audio inputs from a plurality of unidirectional audio input devices, including

at least two unidirectional audio input devices oriented to face forward and at least one unidirectional audio input device oriented to face rearward; analyzing the received audio inputs from the plurality of unidirectional audio input devices to obtain audio information corresponding to the received audio inputs; comparing the audio information corresponding to the audio input from at least one of the two forward facing unidirectional audio input devices and the audio information corresponding to the audio input from the rearward facing unidirectional audio input device; obtaining, from the comparison, a subset of sections of the audio information corresponding to the received audio inputs from the sound source; determining a position of the sound source based on the audio information corresponding to the audio inputs from the plurality of unidirectional audio input devices and the subset of sections of the audio information; and filtering the audio information corresponding to the audio inputs from the plurality of unidirectional audio input devices based on the subset of sections of the audio information and the position of the sound source.

[5] Consistent with some embodiments, the present disclosure provides another method for processing audio input. The method includes receiving audio inputs from a plurality of unidirectional audio input devices, including at least two unidirectional audio input devices oriented to face forward and at least one unidirectional audio input device oriented to face rearward; analyzing the received audio inputs from the plurality of unidirectional audio input devices to obtain audio information corresponding to the received audio inputs; comparing the audio information corresponding to the audio input from at least one of the two forward facing unidirectional audio input devices and the audio information corresponding to the audio input from the rearward facing unidirectional audio input device; and determining if a sound source,

corresponding to the received audio inputs, confronts the at least two forward facing unidirectional audio input devices, based on the comparison.

[6] Consistent with some embodiments, the present disclosure also provides a device for processing audio input. The device includes an array of a plurality of unidirectional audio input devices, at least two of the unidirectional audio input devices oriented to face forward and at least one of the unidirectional audio input devices oriented to face rearward.

[7] Consistent with some embodiments, the present disclosure provides another device for processing audio input. The device includes a memory configured to store a set of instructions; and a processor configured to execute the set of instructions to cause the device to: receive audio inputs from a plurality of unidirectional audio input devices, including at least two unidirectional audio input devices oriented to face forward and at least one unidirectional audio input device oriented to face rearward; analyze the received audio inputs from the plurality of unidirectional audio input devices to obtain audio information corresponding to the received audio inputs; compare the audio information corresponding to the audio input from at least one of the two forward facing unidirectional audio input devices and the audio information corresponding to the audio input from the rearward facing unidirectional audio input device; obtain, from the comparison, a subset of sections of the audio information corresponding to the received audio inputs from the sound source; determine a position of the sound source based on the audio information corresponding to the audio inputs from the plurality of unidirectional audio input devices and the subset of sections of the audio information; and filter the audio information corresponding to the audio inputs from the plurality of unidirectional audio input devices based on the subset of sections of the audio information and the position of the sound source.

[8] Consistent with some embodiments, the present disclosure provides another device for processing audio input. The device includes a memory configured to store a set of instructions; and a processor configured to execute the instructions to cause the device to: receive audio inputs from a plurality of unidirectional audio input devices, including at least two unidirectional audio input devices oriented to face forward and at least one unidirectional audio input device oriented to face rearward; analyze the received audio inputs from the plurality of unidirectional audio input devices to obtain audio information corresponding to the received audio inputs; compare the audio information corresponding to the audio input from at least one of the two forward facing unidirectional audio input devices and the audio information corresponding to the audio input from the rearward facing unidirectional audio input device; and determine if a sound source, corresponding to the received audio inputs, confronts the at least two forward facing unidirectional audio input devices, based on the comparison.

[9] Consistent with some embodiments, the present disclosure further provides a non-transitory computer-readable medium that stores a set of instructions executable by at least one processor of a device for processing audio input to cause the device to perform a method for processing audio input. The method includes receiving audio inputs from a plurality of unidirectional audio input devices, including at least two unidirectional audio input devices oriented to face forward and at least one unidirectional audio input device oriented to face rearward; analyzing the received audio inputs from the plurality of unidirectional audio input devices to obtain audio information corresponding to the received audio inputs; comparing the audio information corresponding to the audio input from at least one of the two forward facing unidirectional audio input devices and the audio information corresponding to the audio input from the rearward facing unidirectional audio input device; obtaining, from the comparison, a

subset of sections of the audio information corresponding to the received audio inputs from the sound source; determining a position of the sound source based on the audio information corresponding to the audio inputs from the plurality of unidirectional audio input devices and the subset of sections of the audio information; and filtering the audio information corresponding to the audio inputs from the plurality of unidirectional audio input devices based on the subset of sections of the audio information and the position of the sound source.

[10] Consistent with some embodiments, the present disclosure provides another non-transitory computer-readable medium that stores a set of instructions executable by at least one processor of a device for processing audio input to cause the device to perform a method for processing audio input. The method includes receiving audio inputs from a plurality of unidirectional audio input devices, including at least two unidirectional audio input devices oriented to face forward and at least one unidirectional audio input device oriented to face rearward; analyzing the received audio inputs from the plurality of unidirectional audio input devices to obtain audio information corresponding to the received audio inputs; comparing the audio information corresponding to the audio input from at least one of the two forward facing unidirectional audio input devices and the audio information corresponding to the audio input from the rearward facing unidirectional audio input device; and determining if a sound source, corresponding to the received audio inputs, confronts the at least two forward facing unidirectional audio input devices, based on the comparison.

[11] Additional features and advantages of the disclosed embodiments will be set forth in part in the following description, and in part will be apparent from the description, or may be learned by practice of the embodiments. The features and advantages of the disclosed

embodiments may be realized and attained by the elements and combinations set forth in the claims.

[12] It is to be understood that both the foregoing general description and the following detailed description are exemplary and explanatory only and are not restrictive of the disclosed embodiments, as claimed.

BRIEF DESCRIPTION OF THE DRAWINGS

[13] The accompanying drawings, which are incorporated in and constitute a part of this specification, illustrate embodiments consistent with the invention and, together with the description, explain the principles of the invention.

[14] **FIG. 1** illustrates an exemplary system environment for implementing methods and devices for processing audio input using unidirectional audio input devices, consistent with some embodiments of this disclosure.

[15] **FIG. 2** is a block diagram of an exemplary service terminal, consistent with some embodiments of this disclosure.

[16] **FIG. 3** is a block diagram of an exemplary array of unidirectional audio input devices, consistent with some embodiments of this disclosure.

[17] **FIGS. 4A-4D** are diagrams of exemplary audio pickup patterns for unidirectional audio input devices, consistent with some embodiments of this disclosure.

[18] **FIG. 5** is a block diagram of an exemplary device for processing audio input from unidirectional audio input devices, consistent with some embodiments of this disclosure.

[19] **FIG. 6** is a diagram of an exemplary spatial layout of an array of unidirectional audio input devices, consistent with some embodiments of this disclosure.

[20] **FIG. 7** is a flowchart of an exemplary method for processing audio input from unidirectional audio input devices, consistent with some embodiments of this disclosure.

DESCRIPTION OF THE EMBODIMENTS

[21] Reference will now be made in detail to exemplary embodiments, examples of which are illustrated in the accompanying drawings. The following description refers to the accompanying drawings in which the same numbers in different drawings represent the same or similar elements unless otherwise represented. The implementations set forth in the following description of exemplary embodiments do not represent all implementations consistent with the invention. Instead, they are merely examples of apparatuses and methods consistent with aspects related to the invention as recited in the appended claims.

[22] In conventional systems in environments with high levels of background noise, to isolate and interpret information spoken by a user of a smart terminal, two methods may be used: either omnidirectional audio input devices are used to receive the spoken information, or an array of unidirectional audio input devices are used with all of the unidirectional audio input devices facing the same direction toward the user. However, both of these methods have drawbacks and shortcomings.

[23] When omnidirectional audio input devices are used to receive spoken information from a user of a smart terminal, such as a service terminal for providing a service to a customer, there may be significant pickup of background noise, and complex algorithms may be needed to filter the spoken information from the background noise. Further, MEMS omnidirectional audio input devices have a relatively low signal-to-noise ratio and signal-to-interference rejection as compared to unidirectional audio input devices. As a result, the reception and filtering of the spoken information from such omnidirectional audio input devices is inefficient when it is only

desired to isolate and interpret spoken information from a single direction. This leads to longer processing times and less predictable results.

[24] More importantly, when omnidirectional audio input devices forming a linear array mounted on the top of smart terminal are used to receive spoken information from a user of a smart terminal, there is no way to determine if the speaker is in front of the array or behind the array. For example, if there are two smart terminals placed back to back or if there is another speaker behind the smart terminal, the omnidirectional audio input devices will receive spoken information from both in front of and behind the array. There is no way for the smart terminal to determine if the speaker is in front of the array. As a result, the reception and filtering of the desired spoken information will include undesired spoken information from behind the array. This also leads to less predictable results.

[25] Alternatively, a linear array of unidirectional audio input devices with all of the unidirectional audio input devices facing the same direction toward the user may be mounted at the top of a smart display to receive spoken information from a user of a smart terminal, such as a service terminal for providing a service to a customer. When these arrays are used, there may still be pickup of some background noise from behind the array. Additionally, there may be significant pickup of background noise from sides of the array. As a result, the reception and filtering of the spoken information from the array of unidirectional devices is unreliable, especially in environments in which high levels of background noise from the sides of the array are expected. Results of interpreting a user's spoken information are therefore less predictable, which consequently affects user experience.

[26] Additionally, when arrays of unidirectional audio input devices with all of the unidirectional audio input devices facing the same direction toward the user are used to receive

spoken information from a user of a smart terminal, there is no way to determine if the speaker is in front of the array or behind the array. For example, if there are two smart terminals placed back to back or if there is another speaker behind the smart terminal, the arrays of unidirectional audio input devices with all of the unidirectional audio input devices facing the same direction toward the user will receive spoken information from both in front of and behind the array. There is no way for the smart terminal to determine if the speaker is in front of the array. As a result, the reception and filtering of the desired spoken information will include undesired spoken information from behind the array. This also leads to less predictable results.

[27] Embodiments of the present disclosure are directed to methods and devices for processing audio input using unidirectional audio input devices. For example and without limitation, embodiments of the present disclosure may include a service terminal in a service facility, such as a coffee shop, which may have high levels of background noise, e.g., operating machinery and people talking. The service terminal may be automated and controlled by one or more processors to receive spoken information from a user, e.g., a customer who orally makes a request for a service, e.g., an order for a beverage. The spoken information from the user is received by an array of unidirectional audio input devices, the outputs of which are sent to the one or more processors operating on software that causes the one or more processors to perform filtering and interpretation to determine the service requested by the user.

[28] According to embodiments of the present disclosure, the array of unidirectional audio input devices includes at least one unidirectional audio input device oriented to face away from the user, i.e., opposite the direction of the other unidirectional audio input devices, which face toward the user, to receive spoken information from a sound source, such as a customer. The at least one unidirectional audio input device oriented to face opposite the direction of the

other unidirectional audio input devices is hereinafter referred to as the unidirectional audio input device oriented to face rearward or as the at least one rearward facing unidirectional audio input device. The remaining unidirectional audio input devices are hereinafter referred to as the unidirectional audio input devices oriented to face forward or as the forward facing unidirectional audio input devices.

[29] Each of the unidirectional audio input devices in the array sends audio inputs, corresponding to the received spoken information to one or more processors. The one or more processors receive and analyze the audio inputs from each of the unidirectional audio input devices to obtain audio information corresponding to the received audio inputs.

[30] The one or more processors then compare the audio information corresponding to the audio inputs from at least one of the forward facing unidirectional audio input devices and the audio information corresponding to the audio input from the rearward facing unidirectional audio input device to determine if a sound source, corresponding to the received audio inputs, confronts the forward facing unidirectional audio input devices. A sound source confronts the forward facing unidirectional audio input devices if the sound source is in a position in front of the array, defined as a position in front of the forward facing unidirectional audio input devices and the service terminal. A position behind the array is defined as a position in front of the rearward facing unidirectional audio input device.

[31] The one or more processors then obtain from the comparison of the audio information a subset of sections of the audio information corresponding to the received audio inputs from the desired sound source.

[32] The one or more processors then determine a position of the desired sound source based on the audio information corresponding to the audio inputs from the unidirectional audio

input devices, the comparison of the audio information, and the subset of sections of the audio information.

[33] In some embodiments, the determination of the position of the sound source is based only on the subset of sections of the audio information corresponding to the received audio inputs from the desired sound source. As a result, the sound source localization is more precise since audio information corresponding to the positions in the rear, is suppressed.

[34] In some embodiments, the position from which the audio input is being received may be used to determine, for example, whether a user is speaking to the service terminal to initiate a service request, e.g., place an order, or whether a nonuser behind or to the side of the array is having a conversation.

[35] The one or more processors then filter the audio information corresponding to the audio inputs from the unidirectional audio input devices based on the audio information corresponding to the audio inputs from the unidirectional audio input devices, the comparison of the audio information, the subset of sections of the audio information, and the position of the sound source.

[36] In some embodiments, the filtering of the audio information is based only on the subset of sections of the audio information corresponding to the received audio inputs from the sound source. As a result of the more precise sound source localization, these beamforming algorithms are able to further optimize the audio information and interpret information from the sound source with very minimal, if any, background noise included.

[37] **FIG. 1** illustrates an exemplary system environment 100 for implementing methods and devices for processing audio input using unidirectional audio input devices, consistent with some embodiments of this disclosure. As shown in **FIG. 1**, the system

environment 100 may be, for example, a busy coffee shop with background noise including people conversing. As shown in **FIG. 1**, devices and methods for processing audio input using unidirectional audio input devices may receive desired spoken information from a sound source, e.g., a spoken request by a user ordering a beverage, e.g., coffee, while filtering out undesired background noise, e.g., a conversation between nonusers. As shown in **FIG. 1**, the exemplary system environment 100 for implementing methods and devices for processing audio input using unidirectional audio input devices includes a service terminal 102.

[38] In some embodiments, the system environment 100 may include users not directly in front of the service terminal 102, or nonusers in positions all around the service terminal. For example, although in an exemplary system environment the user would be directly in front of and facing the service terminal 102, the system is designed to be able to isolate and interpret spoken information from the user in a multitude of positions relative to the service terminal 102. Likewise, while an exemplary system environment would include sources of background noise, e.g., nonusers, conversing at the side of or behind the service terminal 102, the devices and methods disclosed herein are still able to effectively filter the desired spoken information from the user regardless of the positions of nonusers.

[39] **FIG. 2** is a block diagram of an exemplary service terminal 200 that may serve as service terminal 102, consistent with some embodiments of this disclosure. As shown in **FIG. 2**, the exemplary service terminal 200 includes one or more processors 202, a display screen 204, a memory 206 that may further include a nonvolatile storage, and an array of unidirectional audio input devices 208.

[40] The one or more processors 202 may be configured to perform various functions and data processing under control of instructions in operating programs and modules stored in

the memory 206. For example, the one or more processors 202 may be configured to execute instructions to perform all or part of steps of methods disclosed herein. In some embodiments, instructions for processing audio input using unidirectional audio input devices may be read from the nonvolatile storage in the memory 206, such as instructions that are executable by the one or more processors 202 to perform the methods disclosed herein. Each of the display screen 204, the memory 206, and the array of unidirectional audio input devices 208 is operatively coupled to the one or more processors 202. In some embodiments, each unidirectional audio input device in the array of unidirectional audio input devices 208 is separately coupled to the one or more processors 202. Similar structures may be implemented in the service terminal 200 to perform methods for processing audio input using unidirectional audio input devices described above. The service terminal 200 may also include other components not shown in **FIG. 2**.

[41] The one or more processors 202 are provided for executing instructions and performing data processing. In some embodiments, the one or more processors 202 include a digital signal processor configured to receive and process digital signals from the array of unidirectional audio input devices 208. Additionally, the one or more processors 202 also include a processor capable of communicating through a bus and running a real time operating system (RTOS).

[42] The display screen 204 is provided for displaying service options to a user, e.g. a customer, and the results of the customer placing a service request, e.g., an order. In some embodiments, the display screen 204 may be a touch screen used to both display information and receive user input. The display screen 204 may also be a standard computer monitor capable only of displaying information.

[43] The memory 206 is provided for storing instructions for the one or more processors 202 and other necessary data. In some embodiments, the memory 206 may include nonvolatile storage and random access memory (RAM). The instructions may be stored in the nonvolatile storage, read into the RAM, and then executed by the one or more processors 202. The memory 206 may further include a signal processing program and instructions for a method for processing audio input using unidirectional audio input devices.

[44] The array of unidirectional audio input devices 208 is provided for receiving spoken information (e.g., an order request) from a sound source (e.g., a customer) and converting it into digital signals. In some embodiments, the array of unidirectional audio input devices 208 includes one unidirectional audio input device oriented to face rearward, away from a customer placing an order request at the service terminal 200, and four unidirectional audio input devices oriented to face forward toward a customer placing an order request at the service terminal 200.

[45] **FIG. 3** is a block diagram of an exemplary array of unidirectional audio input devices 300, corresponding to the array of unidirectional audio input devices 208, consistent with some embodiments of this disclosure. As shown in **FIG. 3**, the exemplary array of unidirectional audio input devices 300 includes four unidirectional audio input devices oriented to face forward 302, 304, 308, 310, and one unidirectional audio input device oriented to face rearward 306 (denoted by the “X”). The array of unidirectional audio input devices 300 may be configured to include one or more rearward facing audio input devices 306 and two or more forward facing audio input devices 302, 304, 308, 310. For example, the array of unidirectional audio input devices 300 may be configured to include only one rearward facing audio input device 306 and four forward facing audio input devices 302, 304, 308, 310. In some embodiments, the forward facing unidirectional audio input devices 302, 304, 308, and 310 are placed at an equal distance

apart from one another. Additionally, the rearward facing unidirectional audio input device 306 is placed next to one of the forward facing unidirectional audio input devices, i.e., the device 308, closer than the spacing between the forward facing unidirectional audio input devices 302, 304, 308, 310. In some embodiments, the rearward facing unidirectional audio input device 306 is abutting one of the forward facing unidirectional audio input devices, i.e., the device 308. Similar structures may be implemented in the array of unidirectional audio input devices 300. The array of unidirectional audio input devices 300 may also include other components not shown in **FIG. 3**.

[46] **FIG. 3** is intended solely as a block diagram to illustrate components of the array of unidirectional audio input devices and one spatial layout of the array of unidirectional audio input devices. The array, however, may be configured to include fewer or greater than five unidirectional audio input devices. For example, the array may be configured to include the rearward facing unidirectional audio input device in positions other than the center of the array. The array may also be configured to include more than one rearward facing unidirectional audio input devices. Additionally, the array may be configured to include different spatial arrangements for the unidirectional audio input devices such as an even spacing between forward facing unidirectional audio input devices and the rearward facing unidirectional audio input device placed next to one of the forward facing unidirectional audio input devices, i.e., device 308.

[47] In some embodiments, the array of unidirectional audio input devices 300 may include audio input devices oriented to face forward and rearward 302, 304, 306, 308, and 310, each exhibiting a unidirectional audio pickup pattern. These patterns may include one or more of the following: a cardioid audio pickup pattern, a subcardioid audio pickup pattern, a

supercardioid audio pickup pattern, a hypercardioid audio pickup pattern, or combinations of these patterns. The array of unidirectional audio input devices 300 may also include other unidirectional audio pickup patterns not explicitly disclosed herein.

[48] **FIGS. 4A-4D** are diagrams of exemplary audio pickup patterns for unidirectional audio input devices, corresponding to audio pickup patterns exhibited by the unidirectional audio input devices 302, 304, 306, 308, and 310, consistent with some embodiments of this disclosure. As shown in **FIGS. 4A-4D**, the exemplary audio pickup pattern exhibited by each of the unidirectional audio input devices may correspond to a polar plot of: a cardioid polar audio pickup pattern shown in **FIG. 4A**, a supercardioid audio pickup pattern shown in **FIG. 4B**, a hypercardioid audio pickup pattern shown in **FIG. 4C**, or a subcardioid audio pickup pattern shown in **FIG. 4D**, or a combination of one or more of these patterns. These unidirectional audio pickup patterns as shown in **FIGS. 4A-4D** as well as the combinations of one or more of these patterns are referred to as cardioid-type audio pickup patterns. In each of these polar plots of exemplary audio pickup patterns, the unidirectional audio input device is situated at the center of the circle depicting the audio pickup pattern. The top of each polar plot represents a position in front of the unidirectional audio input device whereas the bottom of the polar plot represents a position behind the unidirectional audio input device. Additionally, the concentric circles around the center represent decibel levels that increase as the concentric circles become larger, i.e., with distance from the center.

[49] **FIG. 5** is a block diagram of an exemplary device 500 for processing audio input using unidirectional audio input devices, consistent with some embodiments of this disclosure. As shown in **FIG. 5**, the exemplary device 500 for processing audio input using unidirectional audio input devices 500 includes an array of unidirectional audio input devices 502

corresponding to the array of unidirectional audio input devices 208, a receiving unit 504, an analyzing unit 506, a comparing unit 508, a frontal talk determining unit 510, an obtaining unit 512, a position determining unit 514, and a filtering unit 516. The receiving unit 504, analyzing unit 506, comparing unit 508, frontal talk determining unit 510, obtaining unit 512, position determining unit 514, and filtering unit 516 may be configured to perform various functions and data processing. For example, the receiving unit 504, analyzing unit 506, comparing unit 506, frontal talk determining unit 510, obtaining unit 512, position determining unit 514, and filtering unit 516 may be configured to perform all of or a part of the steps in the methods described herein.

[50] The array of unidirectional audio input devices 502 is operatively coupled to the receiving unit 504. The receiving unit 504 is operatively coupled to the analyzing unit 506. The analyzing unit 506 is operatively coupled to the comparing unit 508, the position determining unit 514, and the filtering unit 516. The comparing unit 508 is operatively coupled to the frontal talk determining unit 510 and the obtaining unit 512. The obtaining unit 512 is operatively coupled to the position determining unit 514 and the filtering unit 516. The position determining unit 514 is operatively coupled to the filtering unit 516.

[51] In some embodiments, as shown in **FIG. 5**, the spoken information from the user is received by the array of unidirectional audio input devices 502. The array of unidirectional audio input devices 502 then sends the audio inputs to the receiving unit 504. For example, the array of unidirectional audio input devices 502 may receive spoken information and convert it into a digital signal to send to the receiving unit. Consistent with the exemplary array of unidirectional audio input devices 300, array 502 includes five unidirectional audio input devices of which one faces rearward (denoted by the “X”).

[52] In some embodiments, as shown in **FIG. 5**, the receiving unit 504 receives the audio inputs from the array of unidirectional audio input devices 502 and then sends the audio inputs to the analyzing unit 506. For example, the receiving unit 504 may have a different connection for each device of the array of unidirectional audio input devices 502. The receiving unit 504 may receive a digital signal corresponding to spoken information from each of the array of unidirectional audio input devices 502 and send the digital signal to the analyzing unit.

[53] In some embodiments, as shown in **FIG. 5**, the analyzing unit 506 then analyzes the audio inputs and obtains audio information that it sends to the comparing unit 508. For example, the analyzing unit 506 may analyze the received audio inputs to obtain waveforms for each frequency band of the audio inputs, power for each frequency band of the audio inputs, and times corresponding to the reception of the audio inputs, for each of the unidirectional audio input devices.

[54] In some embodiments, as shown in **FIG. 5**, the comparing unit 508 compares the relevant audio information and sends the result of its comparison to the frontal talk determining unit 510. The relevant audio information corresponds to the two arrows comparing from the analyzing unit representing the audio information from the rearward facing unidirectional audio input device and the neighboring forward facing audio input device. Additionally, the comparing unit 508 may also send the result of its comparison to the obtaining unit 512. For example, for one or more selected frequency bands, the comparing unit 508 may compare the power for each frequency band between the forward facing unidirectional audio input device and the rearward facing unidirectional audio input device.

[55] In some embodiments, as shown in **FIG. 5**, the frontal talk determining unit 510 determines whether a sound source corresponding to the received audio inputs confronts the

array of unidirectional audio input devices. For example, based on the comparison result from the comparing unit 510, the frontal talk determining unit may make a determination that a sound source confronts the array of unidirectional audio input devices 502 if there are more frequency bands with higher power in the audio input from the forward facing unidirectional audio input device than from the rearward facing unidirectional audio input device.

[56] In some embodiments, as shown in **FIG. 5**, the obtaining unit 512 obtains a subset of sections of audio information based on the comparison result from the comparing unit 510. For example, the obtaining unit 512 may obtain the frequency bands in which the power corresponding to the audio input received by the forward facing unidirectional audio input device is greater than the power corresponding to the audio input received by the rearward facing unidirectional audio input device based on the comparison results from the comparing unit 508. These obtained frequency bands are referred to herein as frontal dominant frequency bands.

[57] In some embodiments, as shown in **FIG. 5**, the position determining unit 514 determines a position of a sound source based on the audio information corresponding to the audio inputs from the array of unidirectional audio input devices 502 and the subset of sections of the audio information obtained by the obtaining unit 512. For example, the one or more processors may only use power and waveforms from the frontal dominant frequency bands to localize the sound source with triangulation and/or time difference algorithms.

[58] In some embodiments, as shown in **FIG. 5**, the filtering unit 516 filters the audio information corresponding to the audio inputs from the array of unidirectional audio input devices 502 based on the audio information corresponding to the audio inputs from the plurality of unidirectional audio input devices, the subset of sections of the audio information, and the position of the sound source. For example, the one or more processors may only use power and

waveforms from the frontal dominant frequency bands to employ beamforming algorithms to filter the audio information. These beamforming algorithms isolate the sections of the spoken information corresponding to the desired sound source. The filtered audio inputs corresponding to the spoken information may be subjected to speech recognition to enable processing of the customer's order. As a result of the beamforming algorithms, the filtered audio input has a high signal-to noise ratio and a speech recognition algorithm applied to the filtered audio output can provide a more accurate result.

[59] **FIG. 6** is a diagram of an exemplary spatial layout of an array 600 of unidirectional audio input devices, consistent with some embodiments of this disclosure. As shown in **FIG. 6**, the array of unidirectional audio input devices 600, corresponding to the array of unidirectional audio input devices 208 and 300, includes the forward facing unidirectional audio input devices 302, 304, 308, and 310 placed at an equal distance apart from one another. Additionally, the rearward facing unidirectional audio input device 306 is placed next to one of the forward facing unidirectional audio input devices, i.e., the device 308. As a result, the unidirectional polar audio pickup pattern of these unidirectional audio input devices is nearly mirrored across an imaginary axis running through the unidirectional audio input devices 302, 304, 306, 308, and 310. This mirroring allows the rearward facing unidirectional audio input device 306 to be utilized to reduce unwanted background noise from behind and the sides of the forward facing unidirectional audio input device 308.

[60] For example, as shown in **FIG. 6**, a sound source may be positioned behind the array (above the array in **FIG. 6**). The unidirectional audio input devices 302, 304, 306, 308, and 310 receive the spoken information from the sound source and send audio input corresponding to the spoken information to the one or more processors 202. The one or more processors 202 then

analyze the audio inputs and determine power corresponding to the audio input from the rearward facing audio input device 306 and the power corresponding to the audio input from the neighboring forward facing audio input device 308. These determined powers are compared by the one or more processors 202 for each frequency band. In the example shown in **FIG. 6**, the one or more processors 202 determine that the sound source does not confront the array of unidirectional audio input devices because there are fewer frequency bands with higher power corresponding to the audio input from the forward facing audio input device 308 than frequency bands with higher power corresponding to the audio input from the rearward facing audio input device 306.

[61] In some embodiments, the forward facing unidirectional audio input devices 302, 304, 308, and 310 may be placed at unequal distances apart from one another.

[62] In some embodiments, the unidirectional audio input devices 302, 304, 306, 308, and 310 may exhibit audio pickup patterns other than the cardioid audio pickup pattern shown in **FIG. 6**. For example, the unidirectional audio input devices 302, 304, 306, 308, and 310 may exhibit a subcardioid audio pickup pattern, a supercardioid audio pickup pattern, a hypercardioid audio pickup pattern, or a combination of one or more of these patterns.

[63] **FIG. 7** is a flowchart of an exemplary method 700 for processing audio input using unidirectional audio input devices, consistent with some embodiments of this disclosure. The exemplary method 700 may be performed by a processor (e.g., one or more processors 202) of a service terminal, such as a smart phone, a tablet, a Personal Computer (PC), or the like.

[64] In step 702, the processor may act as receiving unit 504 to receive audio inputs. For example, the array of unidirectional audio input devices 208 receives spoken information such as a spoken request by a user. The processor receives audio inputs corresponding to the

spoken information from each of the unidirectional audio input devices 302, 304, 306, 308, and 310 of the array of unidirectional audio input devices 300.

[65] In some embodiments, the reception by all unidirectional audio input devices may be caused by user input into the service terminal 200. For example, a user may select an option on the display screen 204 of the service terminal 200, and the one or more processors 202 may begin reception for the unidirectional audio input devices.

[66] In step 704, the processor may act as analyzing unit 506 to analyze the audio inputs and obtain audio information corresponding to the audio inputs from the array of unidirectional audio input devices 208. For example, the processor may sample the audio inputs to obtain waveforms for each frequency band within a predetermined frequency range, e.g., the human vocal range. The processor may then perform frequency analysis, such as performing a Fourier Transform, on the waveforms to determine power for each frequency band. The processor may also record the times corresponding to the reception of the audio inputs, for the unidirectional audio input devices. In step 704, the processor may act a receiving unit 506.

[67] In some embodiments, the processor may then select the unidirectional audio input device oriented to face rearward 306 and the nearest neighboring unidirectional audio input device oriented to face forward 308.

[68] In some embodiments, the analysis of the audio input from the unidirectional audio input devices 302, 304, 306, 308, and 310 may be performed using a subset of the audio information. For example, the processor may sample the audio inputs to obtain waveforms only for frequency bands associated with the human vocal range. The processor may then perform frequency analysis, such as performing a Fourier Transform, on the waveforms to determine power only for the frequency bands associated with the human vocal range.

[69] In some embodiments, the reception and analysis for all unidirectional audio input devices may be continually executed until the highest power received from the unidirectional audio input devices meets a certain threshold. For example, the processor may be continuously receiving and analyzing the audio inputs for all unidirectional audio input devices until the highest power determined by the frequency analysis, such as performing a Fourier Transform, meets a certain threshold such as a decibel level consistent with a user speaking near the unidirectional audio input devices.

[70] In some embodiments, the reception and analysis may be caused by sensory input into the service terminal 200. For example, a motion detector may detect presence of a user within a predetermined distance and send information corresponding to a user standing in front of the service terminal 200 to the processor.

[71] In step 706, the processor detects frontal talk. The processor may act as comparing unit 508 to compare the audio information corresponding to the audio input from at least one of the forward facing unidirectional audio input devices 308 with the audio information corresponding to the audio input from the rearward facing unidirectional audio input device 306. The processor may act as a determining unit 510 to determine if the sound source confronts the forward facing unidirectional audio input devices based on this comparison. For example, the processor may compare the number of frequency bands with higher power in the neighboring forward facing unidirectional audio input device 308 to the number of frequency bands with higher power in the rearward facing unidirectional audio input device 306. If there are more frequency bands with higher power in the forward facing unidirectional audio input device 308 than in the rearward facing unidirectional audio input device 306, then the processor makes a determination that frontal talk is detected.

[72] In some embodiments, if the processor determines that frontal talk is not detected, the algorithm ends and does not perform steps 708, 710, and 712. The processor will then start the algorithm again and begin receiving audio inputs as explained for step 702.

[73] In step 708, the processor may act as obtaining unit 512 to obtain frontal dominant frequency bands. The processor compares the audio information corresponding to the audio input from at least one of the forward facing unidirectional audio input devices 308 with the audio information corresponding to the audio input from the rearward facing unidirectional audio input device 306. The processor then obtains a subset of sections of the audio information corresponding to the received audio inputs from the desired sound source. For example, the processor may compare the power corresponding to the audio inputs from the forward facing unidirectional audio input devices 308 and the rearward facing unidirectional audio input device 306 in each frequency band. The processor may then obtain the frontal dominant frequency bands as the frequency bands in which the power from the forward facing unidirectional audio input device 308 is greater than the power from the rearward facing unidirectional audio input device 306.

[74] In some embodiments, the processor obtains the frontal dominant frequency bands as the frequency bands in which the power from one of the forward facing unidirectional audio input device 308 is a predetermined amount greater than the power from the rearward facing unidirectional audio input device 306.

[75] In some embodiments, steps 706 and 708 may be performed using the same comparison of power to both detect frontal talk and obtain the frontal dominant frequency bands.

[76] In step 710, the processor may act as position determining unit 514 to localize the sound source using the audio information corresponding to the audio inputs from the array of

unidirectional audio input devices 208 and the subset of sections of the audio information. For example, once the frontal dominant frequency bands are obtained in step 708, the processor may analyze audio information for all unidirectional audio input devices 302, 304, 306, 308, and 310 but only from the frontal dominant frequency bands to localize the sound source via time-difference and/or triangulation algorithms.

[77] In some embodiments, the processor may use other methods to localize the sound source such as particle velocity or intensity vectors, steered response power (SRP) methods, or steered response power phase transform (SRP-PHAT). These methods would use the same audio information previously described, including waveforms for each frequency band, power for each frequency band, and times corresponding to the reception of audio inputs. The sound source localization may also include other methods not explicitly disclosed here.

[78] In step 712, the processor may act as filtering unit 516 to employ a beam forming algorithm to filter remaining background noise. The processor filters the audio information corresponding to the audio inputs from the array of unidirectional audio input devices 208 based on the subset of sections of the audio information and the position of the desired sound source. For example, once the sound source is localized in step 708, the processor may employ a beamforming algorithm, such as Generalized Extreme Value Distribution (GEVD) or Minimum Variance Distortionless Response (MDVR) using the sound source localization information determined in step 710 as well as the frontal dominant frequency bands obtained in step 708. Additionally, the beamforming algorithm may use the audio information corresponding to the audio inputs from the array of unidirectional audio input devices 208. The beam forming algorithm may also include other methods not explicitly disclosed here. The beamforming algorithm isolates the sections of the spoken information corresponding to the desired sound

source. The filtered audio inputs corresponding to the spoken information may be subjected to speech recognition to enable processing of the customer's order. As a result of the beamforming algorithm, the filtered audio input has a high signal-to noise ratio and the speech recognition algorithm applied to the filtered audio output can provide a more accurate result.

[79] In some embodiments, the processor may employ the beamforming algorithm separately on individual frequency bands such as each of the frontal dominant frequency bands found in step 706.

[80] In some embodiments, a non-transitory computer-readable storage medium including instructions is also provided, and the instructions may be executed by a device (such as a terminal, a personal computer, or the like), for performing the above-described methods. Common forms of non-transitory media include, for example, a floppy disk, a flexible disk, hard disk, solid state drive, magnetic tape, or any other magnetic data storage medium, a CD-ROM, any other optical data storage medium, any physical medium with patterns of holes, a RAM, a PROM, and EPROM, a FLASH-EPROM or any other flash memory, NVRAM, a cache, a register, any other memory chip or cartridge, and networked versions of the same. The device may include one or more processors (CPUs), an input/output interface, a network interface, and/or a memory.

[81] It should be noted that, the relational terms herein such as "first" and "second" are used only to differentiate an entity or operation from another entity or operation, and do not require or imply any actual relationship or sequence between these entities or operations. Moreover, the words "comprising," "having," "containing," and "including," and other similar forms are intended to be equivalent in meaning and be open ended in that an item or items

following any one of these words is not meant to be an exhaustive listing of such item or items, or meant to be limited to only the listed item or items.

[82] One of ordinary skill in the art will understand that the above described embodiments can be implemented by hardware, or software (program codes), or a combination of hardware and software. If implemented by software, it may be stored in the above-described computer-readable media. The software, when executed by the processor can perform the disclosed methods. The computing units and other functional units described in this disclosure can be implemented by hardware, or software, or a combination of hardware and software. One of ordinary skill in the art will also understand that multiple ones of the above described modules/units may be combined as one module/unit, and each of the above described modules/units may be further divided into a plurality of sub-modules/sub-units.

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

[84] It will be appreciated that the present invention is not limited to the exact construction that has been described above and illustrated in the accompanying drawings, and that various modifications and changes can be made without departing from the scope thereof. It is intended that the scope of the invention should only be limited by the appended claims.

WHAT IS CLAIMED IS:

1. A method for processing audio input, comprising:
 - receiving audio inputs from a plurality of unidirectional audio input devices, including at least two unidirectional audio input devices oriented to face forward and at least one unidirectional audio input device oriented to face rearward;
 - analyzing the received audio inputs from the plurality of unidirectional audio input devices to obtain audio information corresponding to the received audio inputs;
 - comparing the audio information corresponding to the audio input from at least one of the two forward facing unidirectional audio input devices and the audio information corresponding to the audio input from the rearward facing unidirectional audio input device;
 - obtaining, from the comparison, a subset of sections of the audio information corresponding to the received audio inputs from the sound source;
 - determining a position of the sound source based on the audio information corresponding to the audio inputs from the plurality of unidirectional audio input devices and the subset of sections of the audio information; and
 - filtering the audio information corresponding to the audio inputs from the plurality of unidirectional audio input devices based on the subset of sections of the audio information and the position of the sound source.

2. The method according to claim 1, wherein the audio information comprises:
waveforms for each frequency band of a received signal obtained from the audio inputs from the plurality of unidirectional audio input devices;
power for each frequency band of a received signal obtained from the waveforms for each frequency band of a received signal; and
times corresponding to the reception of audio inputs from the at least three unidirectional audio input devices; and
wherein the subset of sections of the audio information comprises:
frequency bands in which the power corresponding to the audio input from the at least one of the at least two forward facing unidirectional audio input devices in the frequency band is greater than the power corresponding to the audio input from the rearward facing unidirectional audio input devices in the frequency band.

3. The method according to any one of claims 1 and 2, wherein the plurality of unidirectional audio input devices exhibit a cardioid-type audio pickup pattern.

4. The method according to claim 3, wherein the cardioid-type audio pickup pattern includes at least one:

- a cardioid audio pickup pattern,
- a subcardioid audio pickup pattern,
- a supercardioid audio pickup pattern, or
- a hypercardioid audio pickup pattern.

5. A method for processing audio input, comprising:

receiving audio inputs from a plurality of unidirectional audio input devices, including at least two unidirectional audio input devices oriented to face forward and at least one unidirectional audio input device oriented to face rearward;

analyzing the received audio inputs from the plurality of unidirectional audio input devices to obtain audio information corresponding to the received audio inputs;

comparing the audio information corresponding to the audio input from at least one of the two forward facing unidirectional audio input devices and the audio information corresponding to the audio input from the rearward facing unidirectional audio input device; and

determining if a sound source, corresponding to the received audio inputs, confronts the at least two forward facing unidirectional audio input devices, based on the comparison.

6. The method according to claim 5, wherein the audio information comprises:

waveforms for each frequency band of a received signal obtained from the audio inputs from the plurality of unidirectional audio input devices;

power for each frequency band of a received signal obtained from the waveforms for each frequency band of a received signal; and

times corresponding to the reception of audio inputs from the at least three unidirectional audio input devices.

7. The method according to any one of claims 5 and 6, wherein the plurality of unidirectional audio input devices exhibit a cardioid-type audio pickup pattern.

8. The method according to claim 7, wherein the cardioid-type audio pickup pattern includes at least one:

- a cardioid audio pickup pattern,
- a subcardioid audio pickup pattern,
- a supercardioid audio pickup pattern, or
- a hypercardioid audio pickup pattern.

9. A device for processing audio input, comprising:
an array of a plurality of unidirectional audio input devices, at least two of the unidirectional audio input devices oriented to face forward and at least one of the unidirectional audio input devices oriented to face rearward.

10. The device according to claim 9, wherein the plurality of unidirectional audio input devices exhibit a cardioid-type audio pickup pattern.

11. The device according to claim 10, wherein the cardioid-type audio pickup pattern includes at least one:

- a cardioid audio pickup pattern,
- a subcardioid audio pickup pattern,
- a supercardioid audio pickup pattern, or
- a hypercardioid audio pickup pattern.

12. A device for processing audio input, comprising:

a memory configured to store a set of instructions; and

a processor configured to execute the set of instructions to cause the device to:

receive audio inputs from a plurality of unidirectional audio input devices,
including at least two unidirectional audio input devices oriented to face forward and at
least one unidirectional audio input device oriented to face rearward;

analyze the received audio inputs from the plurality of unidirectional audio input
devices to obtain audio information corresponding to the received audio inputs;

compare the audio information corresponding to the audio input from at least one
of the two forward facing unidirectional audio input devices and the audio information
corresponding to the audio input from the rearward facing unidirectional audio input
device;

obtain, from the comparison, a subset of sections of the audio information
corresponding to the received audio inputs from the sound source;

determine a position of the sound source based on the audio information
corresponding to the audio inputs from the plurality of unidirectional audio input devices
and the subset of sections of the audio information; and

filter the audio information corresponding to the audio inputs from the plurality of
unidirectional audio input devices based on the subset of sections of the audio information
and the position of the sound source.

13. The device according to claim 12, wherein the audio information comprises:

waveforms for each frequency band of a received signal obtained from the audio inputs from the plurality of unidirectional audio input devices;

power for each frequency band of a received signal obtained from the waveforms for each frequency band of a received signal; and

times corresponding to the reception of audio inputs from the at least three unidirectional audio input devices; and

wherein the subset of sections of the audio information comprises:

frequency bands in which the power corresponding to the audio input from the at least one of the at least two forward facing unidirectional audio input devices in the frequency band is greater than the power corresponding to the audio input from the rearward facing unidirectional audio input devices in the frequency band.

14. The device according to any one of claims 12 and 13, wherein the plurality of unidirectional audio input devices exhibit a cardioid-type audio pickup pattern.

15. The device according to claim 14, wherein the cardioid-type audio pickup pattern includes at least one:

- a cardioid audio pickup pattern,
- a subcardioid audio pickup pattern,
- a supercardioid audio pickup pattern, or
- a hypercardioid audio pickup pattern.

16. A device for processing audio input, comprising:

a memory configured to store a set of instructions; and

a processor configured to execute the instructions to cause the device to:

receive audio inputs from a plurality of unidirectional audio input devices,
including at least two unidirectional audio input devices oriented to face forward and at
least one unidirectional audio input device oriented to face rearward;

analyze the received audio inputs from the plurality of unidirectional audio input
devices to obtain audio information corresponding to the received audio inputs;

compare the audio information corresponding to the audio input from at least one
of the two forward facing unidirectional audio input devices and the audio information
corresponding to the audio input from the rearward facing unidirectional audio input
device; and

determine if a sound source, corresponding to the received audio inputs, confronts
the at least two forward facing unidirectional audio input devices, based on the comparison.

17. The device according to claim 16, wherein the audio information comprises:

waveforms for each frequency band of a received signal obtained from the audio inputs
from the plurality of unidirectional audio input devices;

power for each frequency band of a received signal obtained from the waveforms for
each frequency band of a received signal; and

times corresponding to the reception of audio inputs from the at least three unidirectional
audio input devices.

18. The device according to any one of claims 16 and 17, wherein the plurality of unidirectional audio input devices exhibit a cardioid-type audio pickup pattern.

19. The device according to claim 18, wherein the cardioid-type audio pickup pattern includes at least one:

- a cardioid audio pickup pattern,
- a subcardioid audio pickup pattern,
- a supercardioid audio pickup pattern, or
- a hypercardioid audio pickup pattern.

20. A non-transitory computer-readable medium that stores a set of computer executable instructions that are executable by at least one processor of a device for processing audio input to cause the device to perform a method for processing audio input, the method comprising:

receiving audio inputs from a plurality of unidirectional audio input devices, including at least two unidirectional audio input devices oriented to face forward and at least one unidirectional audio input device oriented to face rearward;

analyzing the received audio inputs from the plurality of unidirectional audio input devices to obtain audio information corresponding to the received audio inputs;

comparing the audio information corresponding to the audio input from at least one of the two forward facing unidirectional audio input devices and the audio information corresponding to the audio input from the rearward facing unidirectional audio input device;

obtaining, from the comparison, a subset of sections of the audio information corresponding to the received audio inputs from the sound source;

determining a position of the sound source based on the audio information corresponding to the audio inputs from the plurality of unidirectional audio input devices and the subset of sections of the audio information; and

filtering the audio information corresponding to the audio inputs from the plurality of unidirectional audio input devices based on the subset of sections of the audio information and the position of the sound source.

21. The computer readable medium according to claim 20, wherein the audio information comprises:

waveforms for each frequency band of a received signal obtained from the audio inputs from the plurality of unidirectional audio input devices;

power for each frequency band of a received signal obtained from the waveforms for each frequency band of a received signal; and

times corresponding to the reception of audio inputs from the at least three unidirectional audio input devices; and

wherein the subset of sections of the audio information comprises:

frequency bands in which the power corresponding to the audio input from the at least one of the at least two forward facing unidirectional audio input devices in the frequency band is greater than the power corresponding to the audio input from the rearward facing unidirectional audio input devices in the frequency band.

22. A non-transitory computer-readable medium that stores a set of computer executable instructions that are executable by at least one processor of a device for processing audio input to cause the device to perform a method for processing audio input, the method comprising:

receiving audio inputs from a plurality of unidirectional audio input devices, including at least two unidirectional audio input devices oriented to face forward and at least one unidirectional audio input device oriented to face rearward;

analyzing the received audio inputs from the plurality of unidirectional audio input devices to obtain audio information corresponding to the received audio inputs;

comparing the audio information corresponding to the audio input from at least one of the two forward facing unidirectional audio input devices and the audio information corresponding to the audio input from the rearward facing unidirectional audio input device; and

determining if a sound source, corresponding to the received audio inputs, confronts the at least two forward facing unidirectional audio input devices, based on the comparison.

23. The computer readable medium according to claim 22, wherein the audio information comprises:

waveforms for each frequency band of a received signal obtained from the audio inputs from the plurality of unidirectional audio input devices;

power for each frequency band of a received signal obtained from the waveforms for each frequency band of a received signal; and

times corresponding to the reception of audio inputs from the at least three unidirectional audio input devices.

100

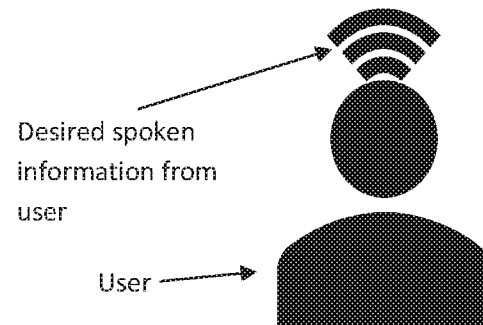
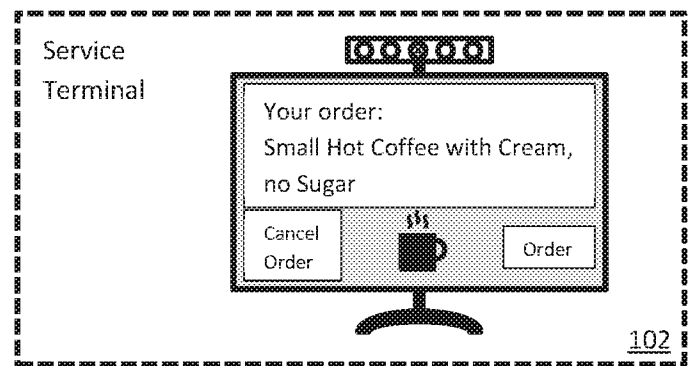
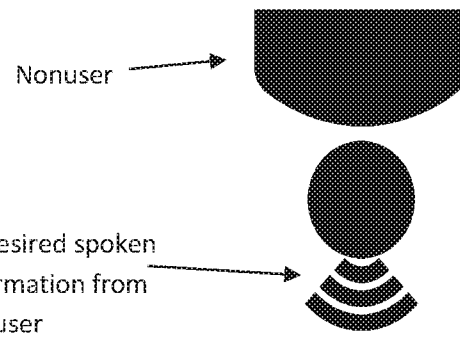
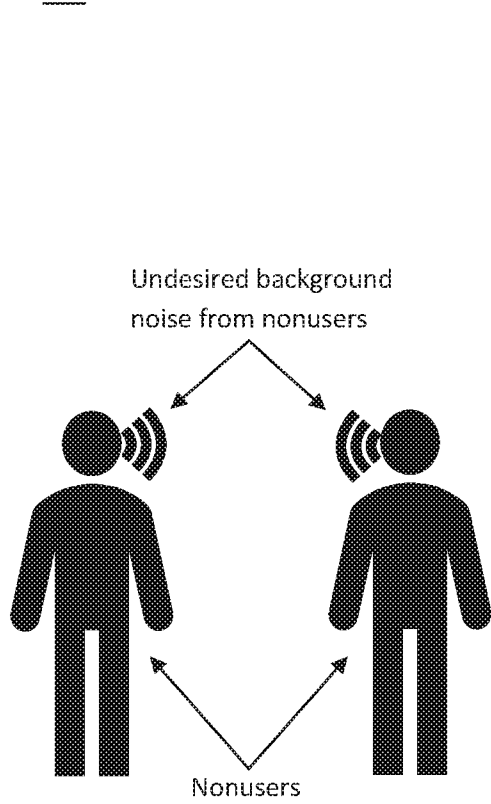


FIG. 1

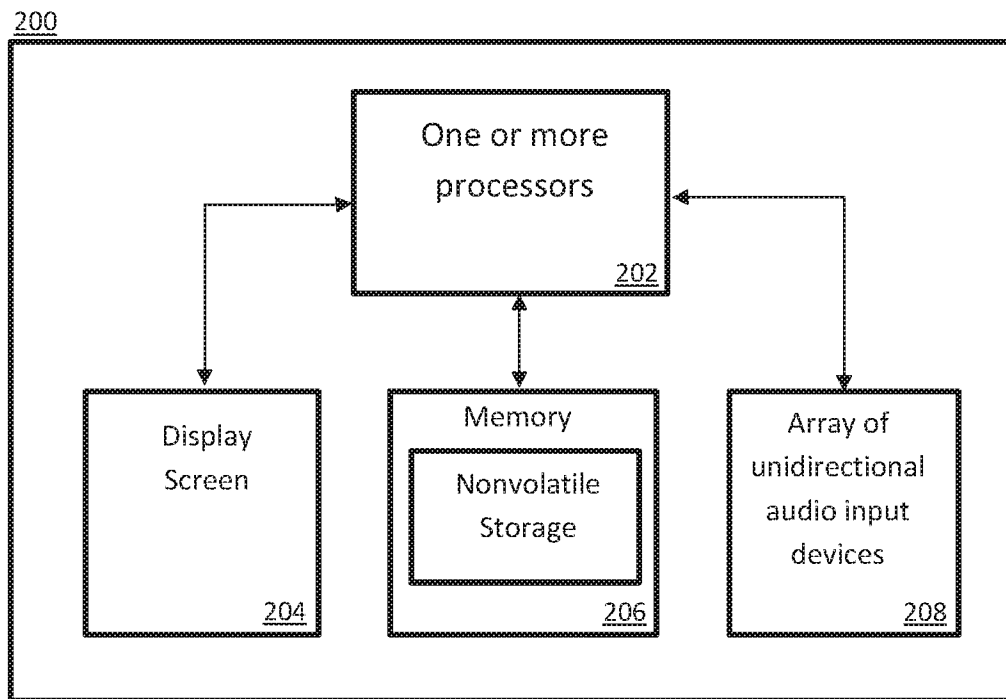


FIG. 2

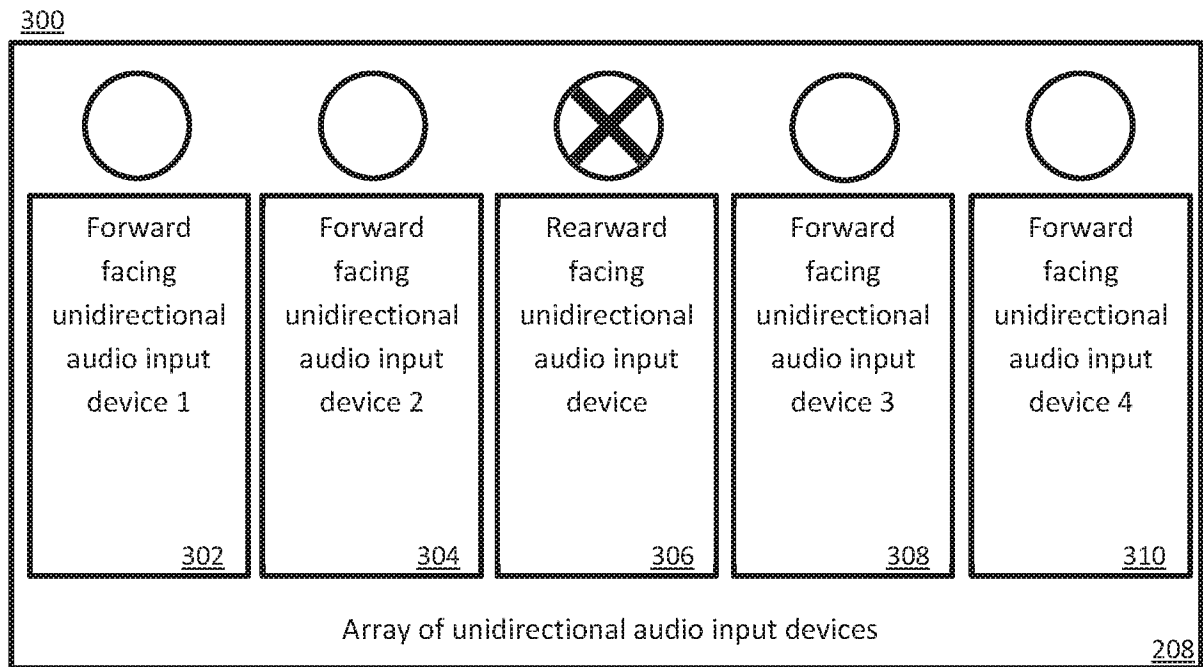


FIG. 3

FIG. 4A

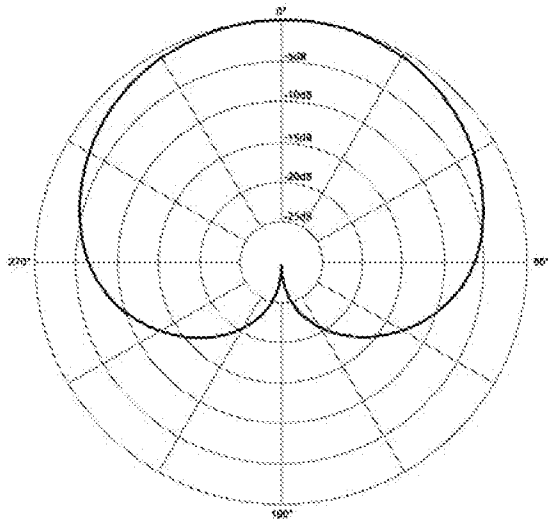


FIG. 4B

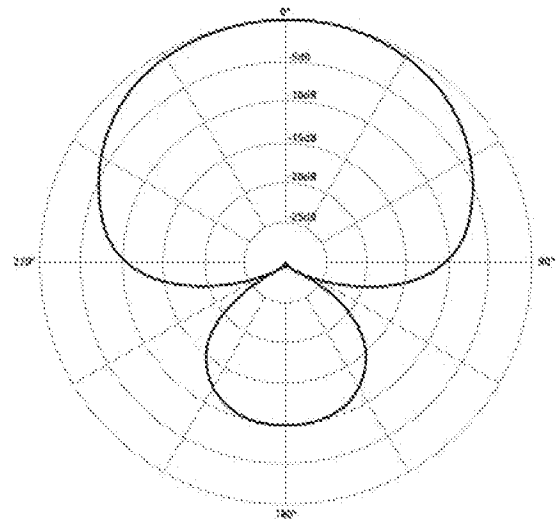


FIG. 4C

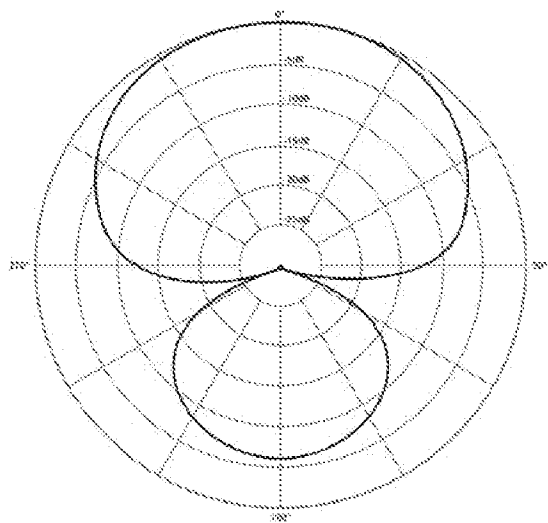
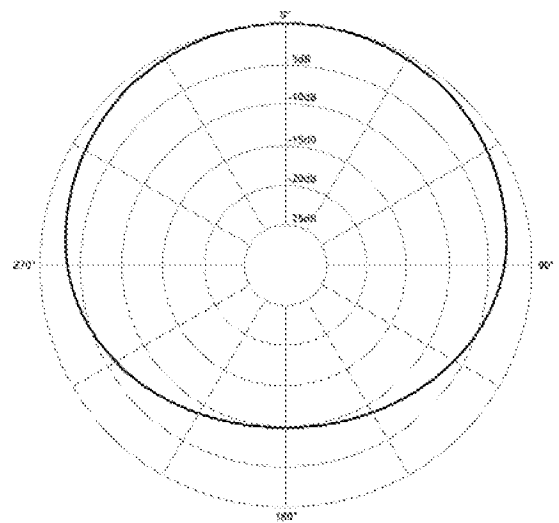


FIG. 4D



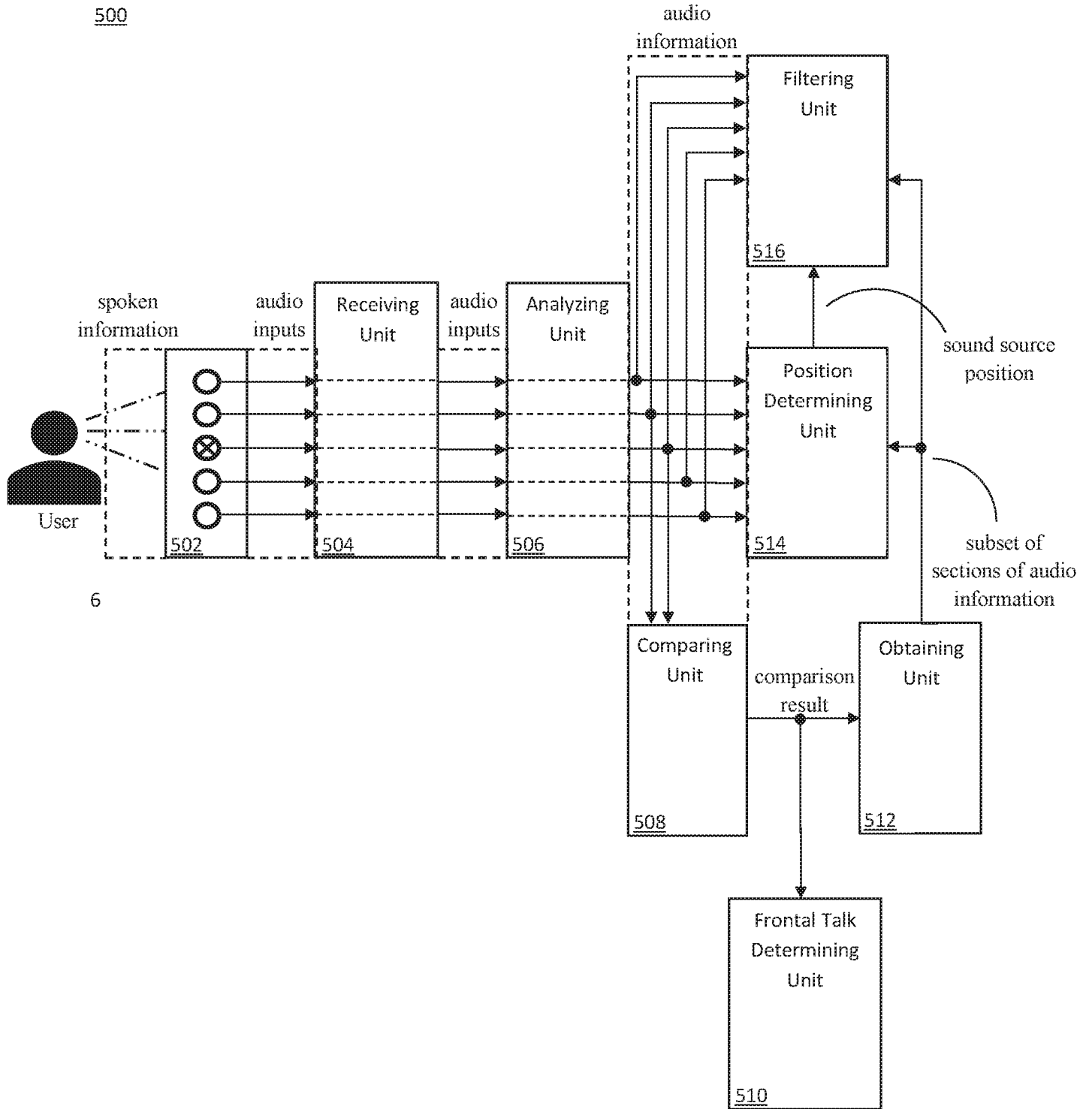


FIG. 5

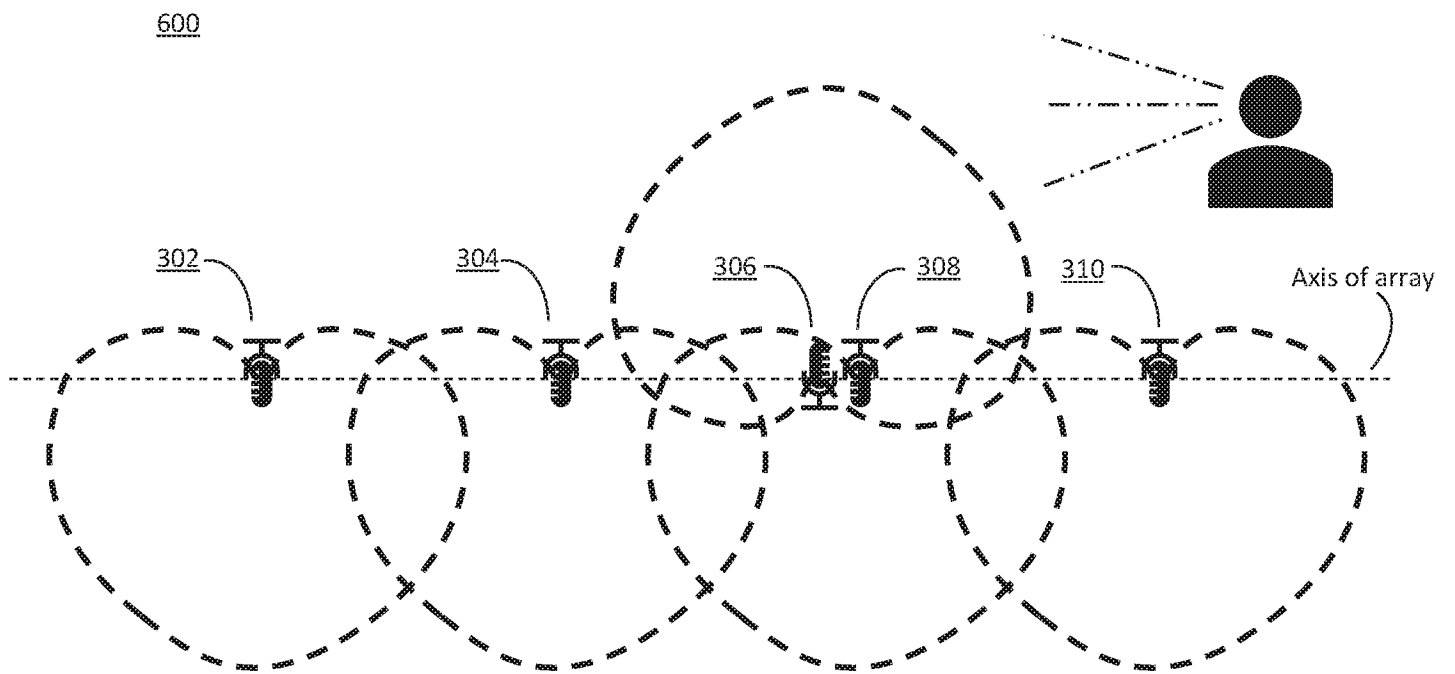


FIG. 6

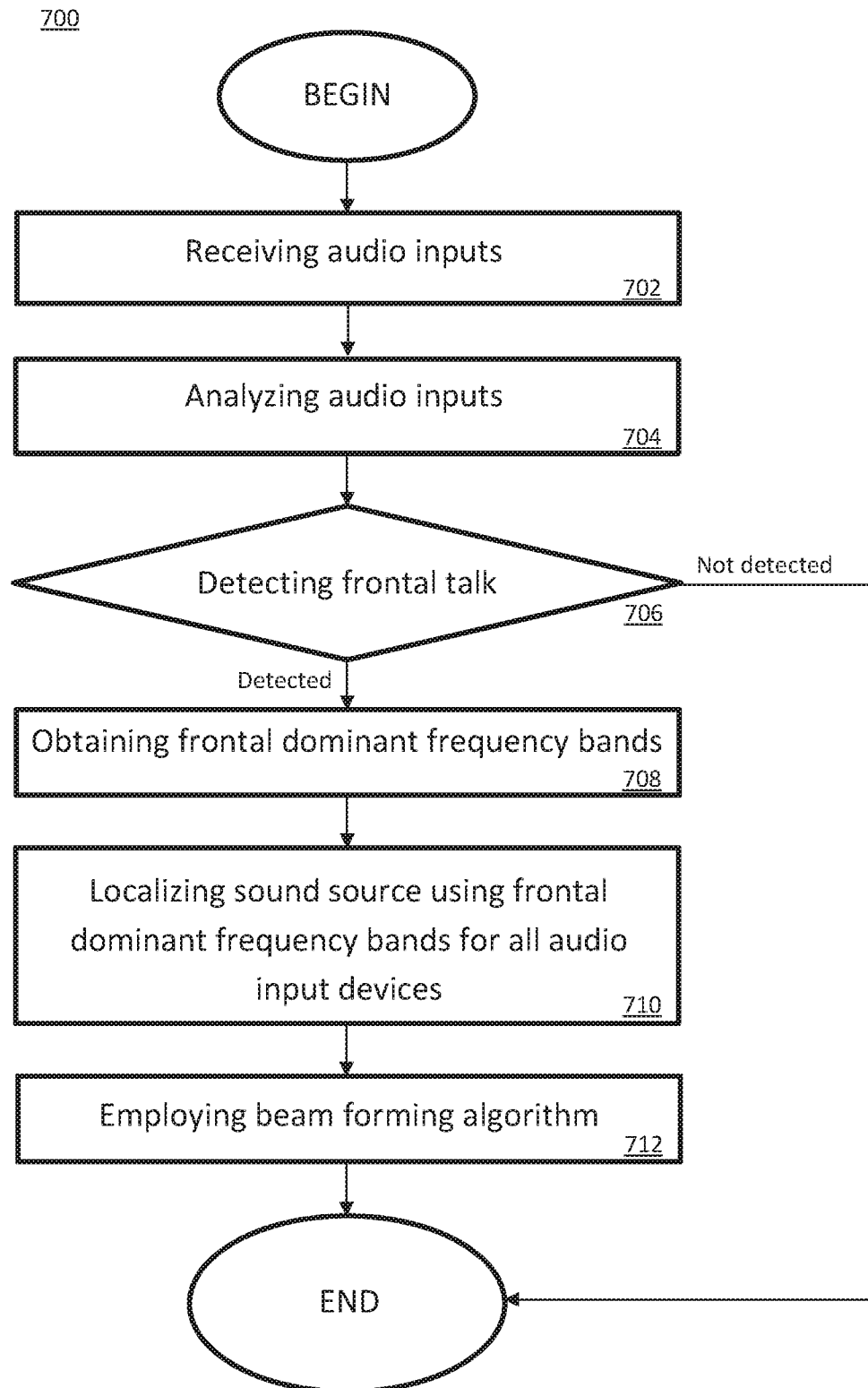


FIG. 7

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2018/105500

A. CLASSIFICATION OF SUBJECT MATTER

G01S 5/02(2010.01)i; G01S 5/20(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G01S G06F H04R

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

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

CNKI,CNPAT,WPI,EPODOC,IEEE: audio, input, sound, source, receive, unidirectional, face, forward, rearward, back, in front of, compare, filter, position, location

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 108254721 A (GOERTEK TECHNOLOGY CO., LTD.) 06 July 2018 (2018-07-06) description, paragraphs [0030]-[0197] and figures 1-9	1-23
A	CN 108073381 A (TENCENT TECHNOLOGY SHENZHEN COMPANY LIMITED) 25 May 2018 (2018-05-25) the whole document	1-23
A	CN 103856877 A (LENOVO BEIJING CO., LTD.) 11 June 2014 (2014-06-11) the whole document	1-23
A	WO 2013176890 A2 (QUALCOMM INCORPORATED) 28 November 2013 (2013-11-28) the whole document	1-23

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

* Special categories of cited documents:

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

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

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

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

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

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

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

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

“&” document member of the same patent family

Date of the actual completion of the international search

17 May 2019

Date of mailing of the international search report

31 May 2019

Name and mailing address of the ISA/CN

National Intellectual Property Administration, PRC
6, Xitucheng Rd., Jimen Bridge, Haidian District, Beijing
100088
China

Authorized officer

CHEN, Wenjing

Facsimile No. (86-10)62019451

Telephone No. 86-(10)-53961739

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2018/105500

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	108254721	A	06 July 2018	None			
CN	108073381	A	25 May 2018	None			
CN	103856877	A	11 June 2014	None			
WO	2013176890	A2	28 November 2013	CN	104321812	A	28 January 2015
				JP	2015523594	A	13 August 2015
				WO	2013176959	A1	28 November 2013
				KR	20150021052	A	27 February 2015
				EP	2856464	A2	08 April 2015
				US	2016005408	A1	07 January 2016
				US	2013315402	A1	28 November 2013
				US	2013317830	A1	28 November 2013