



(51) International Patent Classification:
H04R 3/00 (2006.01)

(21) International Application Number:
PCT/CN2014/075663

(22) International Filing Date:
18 April 2014 (18.04.2014)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
201310202043.3 27 May 2013 (27.05.2013) CN

(71) Applicant: **TENCENT TECHNOLOGY (SHENZHEN) COMPANY LIMITED** [CN/CN]; Room 403, East Block 2, SEG Park, Zhenxing Road, Futian District, Shenzhen, Guangdong 518044 (CN).

(72) Inventors: **LIU, Hong**; Room 403, East Block 2, SEG Park, Zhenxing Road, Futian District, Shenzhen, Guangdong 518044 (CN). **PENG, Chao**; Room 403, East Block 2, SEG Park, Zhenxing Road, Futian District, Shenzhen, Guangdong 518044 (CN). **PENG, Yuanjiang**; Room 403, East Block 2, SEG Park, Zhenxing Road, Futian District, Shenzhen, Guangdong 518044 (CN). **LONG, Xingping**; Room 403, East Block 2, SEG Park, Zhenxing Road, Futian District, Shenzhen, Guangdong 518044 (CN).

(74) Agent: **ADVANCE CHINA I.P. LAW OFFICE**; 39/F, Tower A, G.T. Land Plaza, No. 85 Huacheng Avenue,

Zhujiang New Town, Tianhe District, Guangzhou, Guangdong 510623 (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, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, 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.

(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, 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))

(54) Title: METHOD, SYSTEM AND COMPUTER STORAGE MEDIUM FOR DETECTING AN AUDIO INPUT INTERFACE

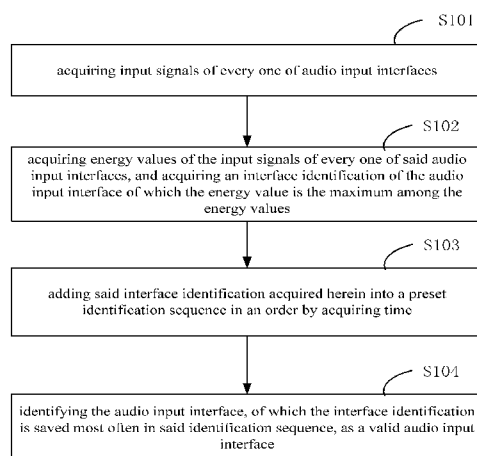


FIG. 1

(57) Abstract: Provided is a method for detecting an audio input interface, comprising: acquiring input signals of every one of audio input interfaces; acquiring energy values of said input signals of every one of said audio input interfaces, and acquiring an interface identification of the audio input interface of which the energy value is the maximum among the energy values; adding said interface identification acquired herein into an identification sequence in an order by acquiring time; and identifying the audio input interface, of which the interface identification is saved most often in said identification sequence, as a valid audio input interface. Also provided is a system for detecting an audio input interface. The present disclosure is capable of effectively identifying the audio input interface connected to the microphone that the user is using, so that the user does not need to switch manually.



Description

METHOD, SYSTEM AND COMPUTER STORAGE MEDIUM FOR DETECTING AN AUDIO INPUT INTERFACE

[1] **【FIELD OF THE INVENTION】**

[2] This application claims priority to Chinese Application No. 201310202043.3, entitled 'Method and System for Detecting an Audio Input Interface', filed on May 27, 2013, which is hereby incorporated by reference in its entirety.

[3] The present disclosure relates to the field of audio processing , and more particularly to a method and a device for detecting an audio input interface.

[4] **【BACKGROUND OF THE INVENTION】**

[5] Along with the popularity of voice software, it is becoming more accepted by most computer users, and has been an indispensable part in people's daily life gradually. The existing computer device often provides an option for choosing an audio input interface, which needs the user to manually switch to choose among different audio input interfaces, however, the switching method herein requires the user to manually try to choose each audio input interface one by one until a voice signal is heard, which is very inconvenient. Moreover, the user often makes misconnection due to not knowing the correct audio input interface, as a result, the correct voice input can not be acquired.

[6] **【SUMMARY OF THE INVENTION】**

[7] In view of the defects existing in conventional method and device mentioned above that, the microphone, to which a user is inputting voice signals, can not be identified automatically, but the user needs to manually switch to choose among audio input interfaces one by one, which is very inconvenient, an object of the present disclosure is to provide a method for detecting an audio input interface, by means of detecting the input data of every one of audio input interfaces, the audio input interface connected to the microphone, into which a voice signal is being input, can be effectively identified, which does not need the user to switch manually and is very convenient.

[8] In one aspect, the present disclosure is realized by the following technical scheme:

[9] A method for detecting an audio input interface comprises:

[10] acquiring input signals of every one of audio input interfaces;

[11] acquiring energy values of said input signals of every one of said audio input interfaces, and acquiring an interface identification of the audio input interface of which the energy value is the maximum among the energy values;

[12] adding said interface identification acquired herein into an identification sequence in an order by acquiring time; and

- [13] identifying the audio input interface, of which the interface identification is saved most often in said identification sequence, as a valid audio input interface.
- [14] In another aspect, the present disclosure is to provide a system for detecting an audio input interface, comprising:
- [15] an input detecting module, configured to acquire input signals of every one of audio input interfaces;
- [16] an energy detecting module, configured to acquire energy values of the input signals of every one of said audio input interfaces, and acquire an interface identification of the audio input interface of which the energy value is the maximum among the energy values;
- [17] an interface identification acquiring module, configured to add said interface identification acquired herein into an identification sequence in an order by acquiring time; and
- [18] an identifying module, configured to identify the audio input interface, of which the interface identification is saved most often in said identification sequence, as a valid audio input interface.
- [19] According to the method and system for detecting an audio input interface of the present disclosure, input signals of every one of audio input interfaces are acquired at a preset time interval, energy values of said input signals of every one of said audio input interfaces are detected, and an interface identification of the audio input interface of which the energy value is the maximum among the energy values is acquired. Each of the interface identifications acquired is added to a preset identification sequence in an order by acquiring time. If the same interface identification is saved more often in said identification sequence, it means that, there are more such situations that the energy value of the input signal of the audio input interface corresponding to said interface identification is the maximum, and the audio input interface corresponding to said interface identification will be identified as a valid audio input interface. The present disclosure is capable of effectively identifying the audio input interface connected to the microphone that the user is using, and the user does not need to switch manually, which is very convenient, moreover, the influence of the noise on the result of identifying is reduced, thereby the result of identifying is more accurate .
- [20] **【BRIEF DESCRIPTION OF THE DRAWINGS】**
- [21] FIG. 1 is a schematic flow diagram illustrating the method for detecting an audio input interface according to the first embodiment of the present disclosure ;
- [22] FIG. 2 is a schematic flow diagram illustrating the method for detecting an audio input interface according to the second embodiment of the present disclosure;
- [23] FIG. 3 is a schematic flow diagram illustrating the method for detecting an audio input interface according to the third embodiment of the present disclosure ;

[24] FIG. 4 is a schematic flow diagram illustrating the method for detecting an audio input interface according to the fourth embodiment of the present disclosure;

[25] FIG. 5 is a structure diagram illustrating the system for detecting an audio input interface according to the first embodiment of the present disclosure;

[26] FIG. 6 is a structure diagram illustrating the system for detecting an audio input interface according to the third embodiment of the present disclosure;

[27] FIG. 7 is a structure diagram illustrating the system for detecting an audio input interface according to the fourth embodiment of the present disclosure.

[28] **【DETAILED DESCRIPTION OF THE EMBODIMENTS】**

[29] In order to make the purpose, technical solutions and advantages of the present disclosure to be understood more clearly, the present disclosure will be described in further details with the accompanying drawings and the following embodiments. It should be understood that the specific embodiments described herein are merely examples to illustrate the disclosure, not to limit the present disclosure.

[30] As shown in Fig. 1, which is a schematic flow diagram illustrating the method for detecting an audio input interface according to the first embodiment of the present disclosure, the method for detecting an audio input interface of said embodiment comprises following steps:

[31] S101: acquiring input signals of every one of audio input interfaces ;

[32] S102: acquiring energy values of the input signals of every one of said audio input interfaces, and acquiring an interface identification of the audio input interface of which the energy value is the maximum among the energy values;

[33] S103: adding said interface identification acquired herein into a preset identification sequence in an order by acquiring time ;

[34] S104: identifying the audio input interface, of which the interface identification is saved most often in said identification sequence, as a valid audio input interface.

[35] According to the method for detecting an audio input interface of the present disclosure, input signals of every one of audio input interfaces are acquired at a preset time interval, energy values of said input signals of every one of said audio input interfaces are detected, and an interface identification of the audio input interface of which the energy value is the maximum among the energy values is acquired. Each of the interface identifications acquired is added to a preset identification sequence in an order by acquiring time. If the same interface identification is saved more often in said identification sequence, it means that, there are more such situations that the energy value of the input signal of the audio input interface corresponding to said interface identification is the maximum, and the audio input interface corresponding to said interface identification will be identified as a valid audio input interface. The method of the present disclosure is capable of effectively identifying the audio input interface

connected to the microphone that the user is using, and the user does not need to switch manually, which is very convenient, moreover, the influence of the noise on the result of identifying is reduced, thereby the result of identifying is more accurate.

- [36] Wherein, with respect to step S101, the input signals of every one of audio input interfaces can be acquired through monitoring for and collecting data from each of said audio input interfaces. The input signals of every one of said audio input interfaces comprise input signals from a microphone hardware-connected to any one of said audio input interfaces, noise signals, and so on. In one embodiment, all of the audio input interfaces can be enumerated by means of calling a function of Dsound API (Direct Sound Capture Enumerate()).
- [37] After all of the audio input interfaces of the device having been detected, the input data of every one of the audio input interfaces is acquired by means of collecting the input signals of every one of said audio input interfaces. Preferably, the parameter of each of the audio input interfaces is preset to unify the audio collecting format for each of the audio input interfaces, such as using an audio collecting format with mono-channel, and 44.1 KHz of sampling rate. By unifying the audio collecting format of each of the audio input interfaces, a large amount of computational load can be reduced in the post processing of the input signals, and the speed of identifying the microphone can be increased.
- [38] In one embodiment, input signals of every one of audio input interfaces are acquired simultaneously at a preset time interval, which comprises following sub-steps:
- [39] S1011: simultaneously collecting input signals of every one of said audio input interfaces, and encapsulating said input signals of every one of said audio input interfaces collected at the same time into a frame of detection data.
- [40] S1012: de-interleaving each of frames of said detection data so as to acquire the input signals of every one of said audio input interfaces. Moreover, each of frames of said detection data can be saved; one of said frames of detection data is extracted every preset frames and further de-interleaved so as to acquire the input signals of every one of said audio input interfaces contained in said frame of detection data.
- [41] In one embodiment, when encapsulating the input signals of every one of said audio input interfaces, each of the input signals is collected in an unit of 20 milliseconds, and then signals of every one of the input signals are put in a buffer with a $M \times 20$ (M represents the number of the devices enumerated) milliseconds length, the corresponding data is acquired from said buffer and encapsulated.
- [42] In this way, the input signals (such as N channels of input signals, N is a nature number) of every one of audio input interfaces are encapsulated in respective frames of detection data. By simply extracting one frame of said detection data every certain frames, the purpose of sampling said input signals at a preset time interval can be

realized. By simply de-interleaving the detection data extracted, the input signals of every one of audio input interfaces can be re-acquired, which is very convenient.

[43] Furthermore, after the input signals of every one of said audio input interfaces have been acquired, each of the input signals is pre-processed so as to ensure the accuracy of the result of the detection. The pre-processing comprises high-pass filtering, filtering certain frequency interferences, noise suppression, and so on, so as to reduce the influence of the noise on the detection of the input signals.

[44] With respect to S102, the energy values of the input signals of every one of said audio input interfaces acquired herein are detected and acquired, each energy value represents the intensity of a respective input signal. The energy value of an input signal is the maximum among the energy values, means that the intensity of the input signal is the strongest, and the audio input interface corresponding to said input signal is most likely connected to a microphone being used by the user. The interface identification of said audio input interface is acquired to identify the audio input interface of the input signal whose intensity is the strongest in the present detection.

[45] With respect to S103, said interface identification acquired herein is added into an identification sequence in an order by acquiring time. Said identification sequence may be created in a buffer or in other types of memories, so that it is easy to be accessed.

[46] Preferably, said identification sequence follows the rule of first-in-and-first-out in an order by acquiring time, and the quantity of the interface identifications saved each time is less than or equal to a preset quantity. That is, if the quantity of interface identifications saved in said identification sequence reaches said preset quantity, whenever a new interface identification is added, an interface identification of which the acquiring time (in another word, the save time) is the earliest will be discarded. The quantity of the interface identifications saved in said identification sequence is maintained in said preset quantity, such as 25, and the most recently acquired 25 of interface identifications will always be saved. The interface identifications saved in the identification sequence may be the same, that is, said identification sequence may contain multiple of the same interface identification.

[47] S104: identifying the audio input interface, of which the interface identification is saved most often in said identification sequence, as a valid audio input interface.

[48] The more often the same interface identifications is saved in said identification sequence, means that the more often the audio frequency with the maximum value is input into the audio input interface corresponding to said interface identification, and the audio input interface corresponding to said input signal is most likely connected to a microphone being used by the user, and the correct chance will be high to identify said audio input interface corresponding to said interface identification as a valid audio input interface. And with the increasing of the quantity of the interface identifications

saved in said identification sequence, the accuracy of identifying will be further increased. Moreover, after the valid audio input interface has been identified, said audio input interface may be automatically matched to an audio software in the background for processing. Alternatively, the input signal of said audio input interface may be subject to processing such as filtering and so on, before it is output into said audio software in the background for processing. After said audio input interface has been identified, a user interface may be further displayed, on which the audio input interface identified currently or the microphone or other audio input devices connected to said audio input interface will be shown. The start and the end of the method for detecting an audio input interface of the present disclosure may be triggered through many different modes, for example, it may be set to start the detection when it is detected that the microphone is inserted, or it may be set to start the detection when a start instruction is received; also for example, it may be set to end the detection when the right input interface is detected, or it may be set to end the detection when the microphone is removed, or it may be set to end the detection when an end instruction is received, and so on.

[49] As shown in Fig. 2, which is a schematic flow diagram illustrating the method for detecting an audio input interface according to the second embodiment of the present disclosure, the step S102 is specified as following sub-steps:

[50] S1022 : acquiring energy values of the input signals of every one of said audio input interfaces ;

[51] S1024: judging whether said maximum energy value among the energy values of every one of said input signals is no less than (greater than or equal to) a preset energy value;

[52] if it is, then going to S1026; if it is not, then judging the input signals of every one of the audio input interfaces as all invalid, and re-acquiring input signals of every one of said input interfaces;

[53] S1026: acquiring the interface identification of the audio input interface of which the energy value is the maximum among the energy values.

[54] The method for detecting an audio input interface in this embodiment is mainly different from the first embodiment as shown in Fig. 1 in that, prior to said acquiring an interface identification of the audio input interface of which the energy value is the maximum among the energy values (S1026), the method further comprises step S1024.

[55] According to the method for detecting an audio input interface in this embodiment, the interface identification of the audio input interface of which the energy value is the maximum among the energy values is acquired only when it is judged that said maximum energy value among the energy values of every one of said input signals is greater than or equal to a preset energy value, otherwise, the input signals of every one

of said audio input interfaces are judged as all invalid. As a result, if the audio input with the maximum energy value is created by a noise, said audio input will be judged as an invalid signal as long as the intensity of noise is less than said preset energy value, and the result of the identification will not be affected, the influence of the noise on the result of the identification is effectively reduced.

[56] As shown in Fig. 3, which is a schematic flow diagram illustrating the method for detecting an audio input interface according to the third embodiment of the present disclosure, the method for detecting an audio input interface in this embodiment is mainly different from the first embodiment as shown in Fig. 1 in that, after said acquiring input signals of every one of audio input interfaces, the method further comprises the following step S105:

[57] S105: executing VAD detection on the input signals of every one of said audio input interfaces;

[58] if all results of said VAD detection on every one of said audio input interfaces are zero (0), then judging said input signals of every one of said audio input interfaces as all invalid, and re-acquiring input signals of every one of said input interfaces;

[59] if at least one of the results of said VAD detection of every one of said audio input interfaces is one (1), then judging the input signals of every one of audio input interfaces as valid, and going to step S102 to acquire energy values of the input signals of every one of said audio input interfaces, and acquire an interface identification of the audio input interface of which the energy value is the maximum among the energy values.

[60] Said VAD detection is so called voice activity detection, which can effectively detect the activity of the input signals, identify the input signal which may be the audio input, and increase the speed of identifying the activities audio input interface. If all results of the VAD detection are 0, that means every one of said audio input interfaces are currently in a muted state; if at least one of results of said audio input interfaces is 1, then at least one of said audio input interfaces has an audio input, and the input signals of every one of audio input interfaces collected herein can be judged as valid, and then the next step (acquiring energy values of the input signals of every one of said audio input interfaces, and acquiring an interface identification of the audio input interface of which the energy value is the maximum among the energy values) is executed.

[61] As shown in Fig. 4, which is a schematic flow diagram illustrating the method for detecting an audio input interface according to the fourth embodiment of the present disclosure, the method for detecting an audio input interface in this embodiment is mainly different from the first embodiment as shown in Fig. 1 in that, after said acquiring input signals of every one of audio input interfaces, the method further comprises the following step S106:

- [62] acquiring signal-noise ratios of the input signals of every one of said audio input interfaces ;
- [63] if each of said signal-noise ratios of said input signals is less than a preset value of signal-noise ratio, then judging the input signals of every one of said audio input interfaces as invalid, and re-acquiring input signals of every one of said audio input interfaces;
- [64] if at least one of said signal-noise ratios of each of said input signals is no less than said preset value of signal-noise ratio, then judging every one of said input signals as valid, and going to step S102 to acquire energy values of the input signals of every one of said audio input interfaces, and acquire an interface identification of the audio input interface of which the energy value is the maximum among the energy values.
- [65] The so called signal-noise ratio is a ratio of a normal sound signal to a noise signal (power) with no-sound signal, and it is often represented in the unit of dB. In this embodiment, the signal-noise ratios of every one of audio input signals are detected, and the input signals of the audio input interface are judged as valid only when at least one of the signal-noise ratios of said audio input interfaces is no less than the preset value of signal-noise ratio, otherwise, the input signals of the audio input interface are judged as invalid. As a result, the influence of the noise on identifying an active audio input interface is reduced, and the accuracy of identifying is improved.
- [66] In a preferred embodiment, any two or any three of steps S1021, S105 and S106 can be selected and executed in combination simultaneously, thereby further improving the accuracy and efficiency of identifying.
- [67] In another preferred embodiment of the method, after said identifying the audio input interface, of which the interface identification is saved most often in said identification sequence, as a valid audio input interface, further executing a step of post-processing to the audio input interface identified by adjusting parameters related to the device so as to make the microphone connected to said device in the best working state.
- [68] Said post-processing comprises: executing a self-adaptive microphone volume adjustment to said valid audio input interface, that is, software/hardware AGC (Automatic Gain Control) processing, self-adaptively adjusting the volume of the microphone, so as to make the microphone in the best volume state; executing a signal-noise ratio detection, starting a noise suppression according to the result of the detection, and so on.
- [69] According to the method of the present disclosure, the user does not need to participate in setting the device, but the device will automatically select the microphone, to which the user is inputting voice signals, and the problem of no sound will never happen, and the corresponding result of the configuration will be displayed in a user interface. When there are multiple microphones with different acoustic characteristics

connected to one device, the user simply needs to speak to the microphone he wants, then the automatically switching of the audio input interfaces will be achieved, without the need of manual setting. If the microphone is broken and no sound is collected, the device can also automatically switch off the corresponding audio input interface, without the need of manual setting. With the method of the present disclosure, the user who doesn't know how to set the microphone is relieved from the difficulty, which is a convenience for the user.

- [70] As shown in Fig. 5, which is a structure diagram illustrating the system for detecting an audio input interface according to the first embodiment of the present disclosure, said system for detecting an audio input interface of present disclosure comprises:
- [71] an input detecting module 11, configured to acquire input signals of every one of audio input interfaces;
- [72] an energy detecting module 12, configured to acquire signals of every one of said audio input interfaces, and acquire an interface identification of the audio input interface of which the energy value is the maximum among the energy values;
- [73] an interface identification acquiring module 13, configured to add said interface identification acquired herein into an identification sequence in an order by acquiring time;
- [74] an identifying module 14, configured to the audio input interface, of which the interface identification is saved most often in said identification sequence, as a valid audio input interface.
- [75] According to the system for detecting an audio input interface of the present disclosure, input signals of every one of audio input interfaces are acquired at a preset time interval, energy values of said input signals of every one of said audio input interfaces are detected, and an interface identification of the audio input interface of which the energy value is the maximum among the energy values is acquired. Each of the interface identifications acquired is added to a preset identification sequence in an order by acquiring time. If the same interface identification is saved more often in said identification sequence, it means that, there are more such situations that the energy value of the input signal of the audio input interface corresponding to said interface identification is the maximum, and the audio input interface corresponding to said interface identification will be identified as a valid audio input interface. The system of the present disclosure is capable of effectively identifying the audio input interface connected to the microphone that the user is using, and the user does not need to switch manually, which is very convenient, moreover, the influence of the noise on the result of identifying is reduced, thereby the result of identifying is more accurate.
- [76] Wherein, with respect to said input detecting module 11, the input signals of every one of audio input interfaces can be acquired through monitoring for and collecting data from each of said audio input interfaces. The input signals of every one of said

audio input interfaces comprises input signals from a microphone hardware-connected to any one of said audio input interfaces, noise signals, and so on. In one embodiment, all of the audio input interfaces can be enumerated by means of calling a function of Dsound API (Direct Sound Capture Enumerate()).

- [77] After all of the audio input interfaces of the device having been detected, the input data of every one of the audio input interfaces is acquired by means of collecting the input signals of every one of said audio input interfaces. Preferably, the parameter of each of the audio input interfaces is preset to unify the audio collecting format for each of the audio input interfaces, such as using an audio collecting format with mono-channel, and 44.1 KHz of sampling rate. By unifying the audio collecting format of each of the audio input interfaces, a large amount of computational load can be reduced in the post processing of the input signals, and the speed of identifying the microphone can be increased.
- [78] In one embodiment, said input detecting module 11 comprises following sub-modules:
- [79] a collecting unit, configured to simultaneously collecting input signals of every one of said audio input interfaces ;
- [80] an encapsulating unit, configured to encapsulate said input signals of every one of said audio input interfaces being collected at the same time into a frame of detection data;
- [81] an extracting unit, configured to de-interleave each of frames of said detection data so as to acquire the input signals of every one of said audio input interfaces contained in said frame of detection data.
- [82] Moreover, said input detecting module 11 further comprises a saving unit, configured to save each of frames of said detection data; said extracting unit is further configured to extract one of said frames of detection data every preset frames, and then de-interleave said frame extracted.
- [83] For example, when the encapsulating unit encapsulates the input signals of every one of said audio input interfaces, each of the input signals is collected in an unit of 20 milliseconds, and then signals of every one of the input signals are put in a buffer with a $M*20$ (M represents the number of the devices enumerated) milliseconds length, the corresponding data is acquired from said buffer and encapsulated.
- [84] In this way, the input signals (such as N channels of input signals, N is a nature number) of every one of audio input interfaces are encapsulated in respective frames of detection data. By simply extracting one frame of said detection data every certain frames, the purpose of sampling said input signals at a preset time interval can be realized. By simply de-interleaving the detection data extracted, the input signals of every one of audio input interfaces can be re-acquired, which is very convenient.

- [85] Furthermore, after the input detecting module 11 has acquired the input signals of every one of said audio input interfaces, each of the input signals is pre-processed so as to ensure the accuracy of the result of the detection. The pre-processing comprises high-pass filtering, filtering certain frequency interferences, noise suppression, and so on, so as to reduce the influence of the noise on the detection of the input signals.
- [86] With respect to said energy detecting module 12, the energy values of the input signals of every one of said audio input interfaces acquired herein are detected and acquired, each energy value represents the intensity of a respective input signal. The energy value of an input signal is the maximum among the energy values, means that the intensity of the input signal is the strongest, and the audio input interface corresponding to said input signal is most likely connected to a microphone being used by the user. The interface identification of said audio input interface is acquired to identify the audio input interface of the input signal whose intensity is the strongest in the present detection.
- [87] With respect to said interface identification acquiring module 13, said interface identification acquired herein is added into an identification sequence in an order by acquiring time. Said identification sequence may be created in a buffer or in other types of memories, so that it is easy to be accessed.
- [88] Preferably, said identification sequence follows the rule of first-in-and-first-out in an order by acquiring time, and the quantity of the interface identifications saved each time is less than or equal to a preset quantity. That is, if the quantity of interface identifications saved in said identification sequence reaches said preset quantity, whenever a new interface identification is added, an interface identification of which the acquiring time (in another word, the save time) is the earliest will be discarded. The quantity of the interface identifications saved in said identification sequence is maintained in said preset quantity, such as 25, and the most recently acquired 25 of interface identifications will always be saved. The interface identifications saved in the identification sequence may be the same, that is, said identification sequence may contain multiple of the same interface identification.
- [89] With respect to said identifying module 14, it is configured to identify the audio input interface, of which the interface identification is saved most often in said identification sequence, as a valid audio input interface.
- [90] The more often the same interface identifications is saved in said identification sequence, means that the more often the audio frequency with the maximum value is input into the audio input interface corresponding to said interface identification, and the audio input interface corresponding to said input signal is most likely connected to a microphone being used by the user, and the correct chance will be high to identify said audio input interface corresponding to said interface identification as a valid audio

input interface. And with the increasing of the quantity of the interface identifications saved in said identification sequence, the accuracy of identifying will be further increased.

- [91] Moreover, after the valid audio input interface has been identified, said audio input interface may be automatically matched to an audio software in the background for processing. Alternatively, the input signal of said audio input interface may be subject to processing such as filtering and so on, before it is output into said audio software in the background for processing. After said audio input interface has been identified, a user interface may be further displayed, on which the audio input interface identified currently or the microphone or other audio input devices connected to said audio input interface will be shown.
- [92] The start and the end of the system for detecting an audio input interface of the present disclosure may be triggered through many different modes, for example, it may be set to start the detection when it is detected that the microphone is inserted, or it may be set to start the detection when a start instruction is received; also for example, it may be set to end the detection when the right input interface is detected, or it may be set to end the detection when the microphone is removed, or it may be set to end the detection when an end instruction is received, and so on.
- [93] The present disclosure further provides a second embodiment of the system for detecting an audio input interface, which is mainly different from the first embodiment as shown in Fig. 5 in that, prior to acquiring an interface identification of said audio input interface of which the energy value is the maximum among the energy values, said interface identification acquiring module 13 further judges whether said maximum energy value among the energy values of every one of said input signals is no less than (greater than or equal to) a preset energy value; if it is, then acquires the interface identification of the audio input interface of which the energy value is the maximum among the energy values; if it is not, then judges the input signals of every one of the audio input interfaces as all invalid, and re-acquires input signals of every one of said input interfaces;
- [94] According to the system for detecting an audio input interface in this embodiment, said interface identification acquiring module 13 acquires the interface identification of the audio input interface of which the energy value is the maximum among the energy values only when it is judged that said maximum energy value among the energy values of every one of said input signals is greater than or equal to a preset energy value, otherwise, the input signals of every one of said audio input interfaces are judged as all invalid. As a result, if the audio input with the maximum energy value is created by a noise, said audio input will be judged as an invalid signal as long as the intensity of noise is less than said preset energy value, and the result of the identi-

fication will not be affected, the influence of the noise on the result of the identification is effectively reduced.

[95] As shown in Fig. 6, which is a structure diagram illustrating the system for detecting an audio input interface according to the third embodiment of the present disclosure, the system for detecting an audio input interface in this embodiment is mainly different from the first embodiment as shown in Fig. 5 in that, the system further comprises a VAD detecting module 15;

[96] Said VAD detecting module 15 is configured to execute VAD detection on the input signals of every one of said audio input interfaces;

[97] if all results of said VAD detection on every one of said audio input interfaces are zero (0), then judging said input signals of every one of said audio input interfaces as all invalid, and re-acquiring input signals of every one of said input interfaces; if at least one of the results of said VAD detection of every one of said audio input interfaces is one (1), then judging the input signals of every one of audio input interfaces as valid, and said energy detecting module 12 acquires energy values of the input signals of every one of said audio input interfaces, and acquire an interface identification of the audio input interface of which the energy value is the maximum among the energy values.

[98] Said VAD detection is so called voice activity detection, which can effectively detect the activity of the input signals, identify the input signal which may be the audio input, and increase the speed of identifying the activities audio input interface. If all results of the VAD detection are 0, that means every one of said audio input interfaces are currently in a muted state; if at least one of results of said audio input interfaces is 1, then at least one of said audio input interfaces has an audio input, and the input signals of every one of audio input interfaces collected herein can be judged as valid, and then the next step (acquiring energy values of the input signals of every one of said audio input interfaces, and acquiring an interface identification of the audio input interface of which the energy value is the maximum among the energy values) is executed.

[99] As shown in Fig. 7, which is a structure diagram illustrating the system for detecting an audio input interface according to the fourth embodiment of the present disclosure, the system is mainly different from the first embodiment as shown in Fig. 5 in that, the system further comprises a signal-noise ratio detecting module 16, which is configured to acquire signal-noise ratios of the input signals of every one of said audio input interfaces; if each of said signal-noise ratios of said input signals is less than a preset value of signal-noise ratio, then judging the input signals of every one of said audio input interfaces as invalid, and re-acquiring input signals of every one of said audio input interfaces; if at least one of said signal-noise ratios of each of said input signals is no less than said preset value of signal-noise ratio, then judging every one of

said input signals as valid, said energy detecting module 12 acquires energy values of the input signals of every one of said audio input interfaces, and acquire an interface identification of the audio input interface of which the energy value is the maximum among the energy values.

[100] The so called signal-noise ratio is a ratio of a normal sound signal to a noise signal (power) with no-sound signal, and it is often represented in the unit of dB. In this embodiment, the signal-noise ratios of every one of audio input signals are detected, and the input signals of the audio input interface are judged as valid only when at least one of the signal-noise ratios of said audio input interfaces is no less than the preset value of signal-noise ratio, otherwise, the input signals of the audio input interface are judged as invalid. As a result, the influence of the noise on identifying an active audio input interface is reduced, and the accuracy of identifying is improved.

[101] In a preferred embodiment, any two or any three of said interface identification acquiring module 13, said VAD detecting module 15 and said signal-noise ratio detecting module 16 can be selected and adopted in combination, thereby further improving the accuracy and efficiency of identifying.

[102] In another preferred embodiment of the system, after said identifying module 14 identifies the audio input interface, of which the interface identification is saved most often in said identification sequence, as a valid audio input interface, post-processing is further executed to the audio input interface identified by adjusting parameters related to the device so as to make the microphone connected to said device in the best working state.

[103] Said post-processing comprises: executing a self-adaptive microphone volume adjustment to said valid audio input interface, that is, software/hardware AGC (Automatic Gain Control) processing, self-adaptively adjusting the volume of the microphone, so as to make the microphone in the best volume state; executing a signal-noise ratio detection, starting a noise suppression according to the result of the detection, and so on.

[104] According to the system of the present disclosure, the user does not need to participate in setting the device, but the device will automatically select the microphone, to which the user is inputting voice signals, and the problem of no sound will never happen, and the corresponding result of the configuration will be displayed in a user interface. When there are multiple microphones with different acoustic characteristics connected to one device, the user simply needs to speak to the microphone he wants, then the automatically switching of the audio input interfaces will be achieved, without the need of manual setting. If the microphone is broken and no sound is collected, the device can also automatically switch off the corresponding audio input interface, without the need of the user's manual setting. With the system of the present

disclosure, the user who doesn't know how to set the microphone is relieved from the difficulty, which is a convenience for the user.

[105] The person skilled in the art can understand that, all of or part of the processes and the corresponding system implementing the embodiments mentioned above, may be achieved by means of relevant hardware commanded by computer programs, the computer programs may be saved in the computer readable storage medium, and they may comprise the processes of embodiments of the respective methods and systems mentioned above when the programs are executed. Wherein, the storage medium may be a disk or CD or read-only memory or random access memory, and so on.

[106] The foregoing examples are preferred embodiments of the present disclosure only and not intended to limit the present disclosure. It should be understood that, to the person skilled in the art, various modifications and improvements can be made without departing from the spirit and principle of the present disclosure, which should all be included within the scope of the present disclosure. Therefore, the protection scope of the present disclosure shall be defined by the appended claims.

Claims

- [Claim 1] A method for detecting an audio input interface, comprising :
acquiring input signals of every one of audio input interfaces;
acquiring energy values of said input signals of every one of said audio input interfaces, and acquiring an interface identification of the audio input interface of which the energy value is the maximum among the energy values;
adding said interface identification acquired herein into an identification sequence in an order by acquiring time ; and
identifying the audio input interface, of which the interface identification is saved most often in said identification sequence, as a valid audio input interface.
- [Claim 2] The method for detecting an audio input interface according to claim 1, further comprising:
when the quantity of interface identifications saved in said identification sequence reaches a preset quantity, whenever a new interface identification is added, one saved interface identification of which the acquiring time is the earliest will be discarded.
- [Claim 3] The method for detecting an audio input interface according to claim 1 or 2, wherein, prior to said acquiring an interface identification of the audio input interface of which the energy value is the maximum among the energy values, the method further comprises the following sub-step: judging whether said maximum energy value among the energy values of every one of said input signals is no less than a preset energy value; when said maximum energy value among the energy values of every one of said input signals is no less than the preset energy value, then acquiring the interface identification of the audio input interface of which the energy value is the maximum among the energy values; when said maximum energy value among the energy values of every one of said input signals is less than the preset energy value, then judging the input signals of every one of the audio input interfaces as all invalid, and re-acquiring input signals of every one of said input interfaces.
- [Claim 4] The method for detecting an audio input interface according to claim 1 or 2, wherein, after said acquiring input signals of every one of audio input interfaces, the method further comprises the following step: executing VAD detection on the input signals of every one of said

audio input interfaces;

when all results of said VAD detection on every one of said audio input interfaces are zero, then judging said input signals of every one of said audio input interfaces as all invalid, and re-acquiring input signals of every one of said input interfaces;

when at least one of the results of said VAD detection of every one of said audio input interfaces is one, then judging the input signals of every one of audio input interfaces as valid, acquiring energy values of the input signals of every one of said audio input interfaces, and acquiring an interface identification of the audio input interface of which the energy value is the maximum among the energy values.

[Claim 5]

The method for detecting an audio input interface according to claim 1 or 2, wherein, after said acquiring input signals of every one of audio input interfaces, the method further comprises the following step: acquiring signal-noise ratios of the input signals of every one of said audio input interfaces;

when each of said signal-noise ratios of said input signals is less than a preset value of signal-noise ratio, then judging the input signals of every one of said audio input interfaces as invalid, and re-acquiring input signals of every one of said audio input interfaces;

when at least one of said signal-noise ratios of each of said input signals is no less than said preset value of signal-noise ratio, then judging every one of said input signals as valid, acquiring energy values of the input signals of every one of said audio input interfaces, and acquiring an interface identification of the audio input interface of which the energy value is the maximum among the energy values.

[Claim 6]

The method for detecting an audio input interface according to claim 1 or 2, wherein, said acquiring input signals of every one of audio input interfaces comprises following sub-steps:

simultaneously collecting input signals of every one of said audio input interfaces, and encapsulating said input signals of every one of said audio input interfaces collected at the same time into a frame of detection data;

de-interleaving each of frames of said detection data so as to acquire the input signals of every one of said audio input interfaces.

[Claim 7]

The method for detecting an audio input interface according to claim 6, wherein, said de-interleaving each of frames of said detection data comprises the following sub-steps:

saving each of frames of said detection data ;

extracting one of said frames of detection data every preset frames, and then de-interleaving said frame extracted herein.

[Claim 8]

A system for detecting an audio input interface , comprising:

an input detecting module, configured to acquire input signals of every one of audio input interfaces;

an energy detecting module, configured to acquire energy values of the input signals of every one of said audio input interfaces, and acquire an interface identification of the audio input interface of which the energy value is the maximum among the energy values;

an interface identification acquiring module, configured to add said interface identification acquired herein into an identification sequence in an order by acquiring time; and

an identifying module, configured to identify the audio input interface, of which the interface identification is saved most often in said identification sequence, as a valid audio input interface.

[Claim 9]

The system for detecting an audio input interface according to claim 8 , wherein, the interface identification is configured to discard one saved interface identification of which the acquiring time is the earliest whenever a new interface identification is added, if the quantity of interface identifications saved in said identification sequence reaches a preset quantity.

[Claim 10]

The system for detecting an audio input interface according to claim 8 or 9, wherein, said interface identification acquiring module is configured to judge whether said maximum energy value among the energy values of every one of said input signals is no less than a preset energy value;

when said maximum energy value among the energy values of every one of said input signals is no less than a preset energy value, then the system acquires the interface identification of the audio input interface of which the energy value is the maximum among the energy values; when said maximum energy value among the energy values of every one of said input signals is less than a preset energy value, then the system judges the input signals of every one of the audio input interfaces as all invalid, and re-acquires input signals of every one of said input interfaces.

[Claim 11]

The system for detecting an audio input interface according to claim 8 or 9, furthering comprising:

a VAD detecting module, configured to execute VAD detection on the input signals of every one of said audio input interfaces;
when all results of said VAD detection on every one of said audio input interfaces are zero, then the system judges said input signals of every one of said audio input interfaces as all invalid, and re-acquires input signals of every one of said input interfaces;
when at least one of the results of said VAD detection of every one of said audio input interfaces is one, then the system judges the input signals of every one of audio input interfaces as valid, the energy detecting module acquires energy values of the input signals of every one of said audio input interfaces, and acquires an interface identification of the audio input interface of which the energy value is the maximum among the energy values.

[Claim 12]

The system for detecting an audio input interface according to claim 8 or 9, furthering comprising :

a signal-noise ratio detecting module, configured to acquire signal-noise ratios of the input signals of every one of said audio input interfaces;
when each of said signal-noise ratios of said input signals is less than a preset value of signal-noise ratio, then the system judges the input signals of every one of said audio input interfaces as invalid, and re-acquires input signals of every one of said audio input interfaces;
when at least one of said signal-noise ratios of each of said input signals is no less than said preset value of signal-noise ratio, then the system judges every one of said input signals as valid, the energy detecting module acquires energy values of the input signals of every one of said audio input interfaces, and acquires an interface identification of the audio input interface of which the energy value is the maximum among the energy values.

[Claim 13]

The system for detecting an audio input interface according to claim 8 or 9, wherein, said input detecting module comprises following sub-modules:

a collecting unit, configured to simultaneously collecting input signals of every one of said audio input interfaces ;
an encapsulating unit, configured to encapsulate said input signals of every one of said audio input interfaces being collected at the same time into a frame of detection data;
an extracting unit, configured to de-interleave each of frames of said

detection data so as to acquire the input signals of every one of said audio input interfaces contained in said frame of detection data.

[Claim 14]

The system for detecting an audio input interface according to claim 13, wherein, said input detecting module further comprises a saving unit, which is configured to save each of frames of said detection data; and wherein said extracting unit is further configured to extract one of said frames of detection data every preset frames, and then de-interleave said frame extracted.

[Claim 15]

One or more non-transitory computer readable storage medium, including computer executable instructions, aid executable instructions are configured to execute a method for detecting an audio input interface, wherein, said method comprising:
acquiring input signals of every one of audio input interfaces;
acquiring energy values of said input signals of every one of said audio input interfaces, and acquiring an interface identification of the audio input interface of which the energy value is the maximum among the energy values;
adding said interface identification acquired herein into an identification sequence in an order by acquiring time ; and
identifying the audio input interface, of which the interface identification is saved most often in said identification sequence, as a valid audio input interface.

[Claim 16]

The one or more non-transitory computer readable storage medium according to claim 15, wherein, said method further comprising:
when the quantity of interface identifications saved in said identification sequence reaches a preset quantity, whenever a new interface identification is added, one saved interface identification of which the acquiring time is the earliest will be discarded.

[Claim 17]

The one or more non-transitory computer readable storage medium according to claim 16, wherein, prior to said acquiring an interface identification of the audio input interface of which the energy value is the maximum among the energy values, the method further comprises the following sub-step:
judging whether said maximum energy value among the energy values of every one of said input signals is no less than a preset energy value;
when said maximum energy value among the energy values of every one of said input signals is no less than the preset energy value, then acquiring the interface identification of the audio input interface of

which the energy value is the maximum among the energy values;
when said maximum energy value among the energy values of every one of said input signals is less than the preset energy value, then judging the input signals of every one of the audio input interfaces as all invalid, and re-acquiring input signals of every one of said input interfaces.

[Claim 18]

The one or more non-transitory computer readable storage medium according to claim 16, wherein, after said acquiring input signals of every one of audio input interfaces, the method further comprises the following step:

executing VAD detection on the input signals of every one of said audio input interfaces;

when all results of said VAD detection on every one of said audio input interfaces are zero, then judging said input signals of every one of said audio input interfaces as all invalid, and re-acquiring input signals of every one of said input interfaces;

when at least one of the results of said VAD detection of every one of said audio input interfaces is one, then judging the input signals of every one of audio input interfaces as valid, acquiring energy values of the input signals of every one of said audio input interfaces, and acquiring an interface identification of the audio input interface of which the energy value is the maximum among the energy values.

[Claim 19]

The one or more non-transitory computer readable storage medium according to claim 16, wherein, after said acquiring input signals of every one of audio input interfaces, the method further comprises the following step:

acquiring signal-noise ratios of the input signals of every one of said audio input interfaces;

when each of said signal-noise ratios of said input signals is less than a preset value of signal-noise ratio, then judging the input signals of every one of said audio input interfaces as invalid, and re-acquiring input signals of every one of said audio input interfaces;

when at least one of said signal-noise ratios of each of said input signals is no less than said preset value of signal-noise ratio, then judging every one of said input signals as valid, acquiring energy values of the input signals of every one of said audio input interfaces, and acquiring an interface identification of the audio input interface of which the energy value is the maximum among the energy values.

[Claim 20]

The one or more non-transitory computer readable storage medium according to claim 16, wherein, said acquiring input signals of every one of audio input interfaces comprises following sub-steps:
simultaneously collecting input signals of every one of said audio input interfaces, and encapsulating said input signals of every one of said audio input interfaces collected at the same time into a frame of detection data;
de-interleaving each of frames of said detection data so as to acquire the input signals of every one of said audio input interfaces.

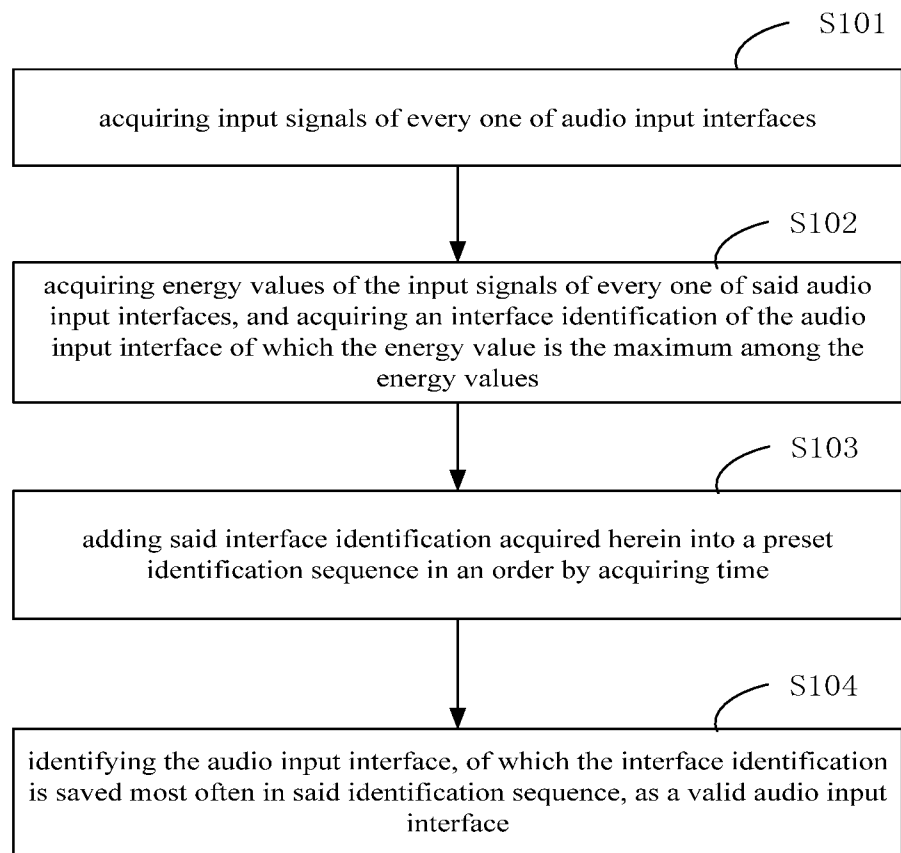


FIG. 1

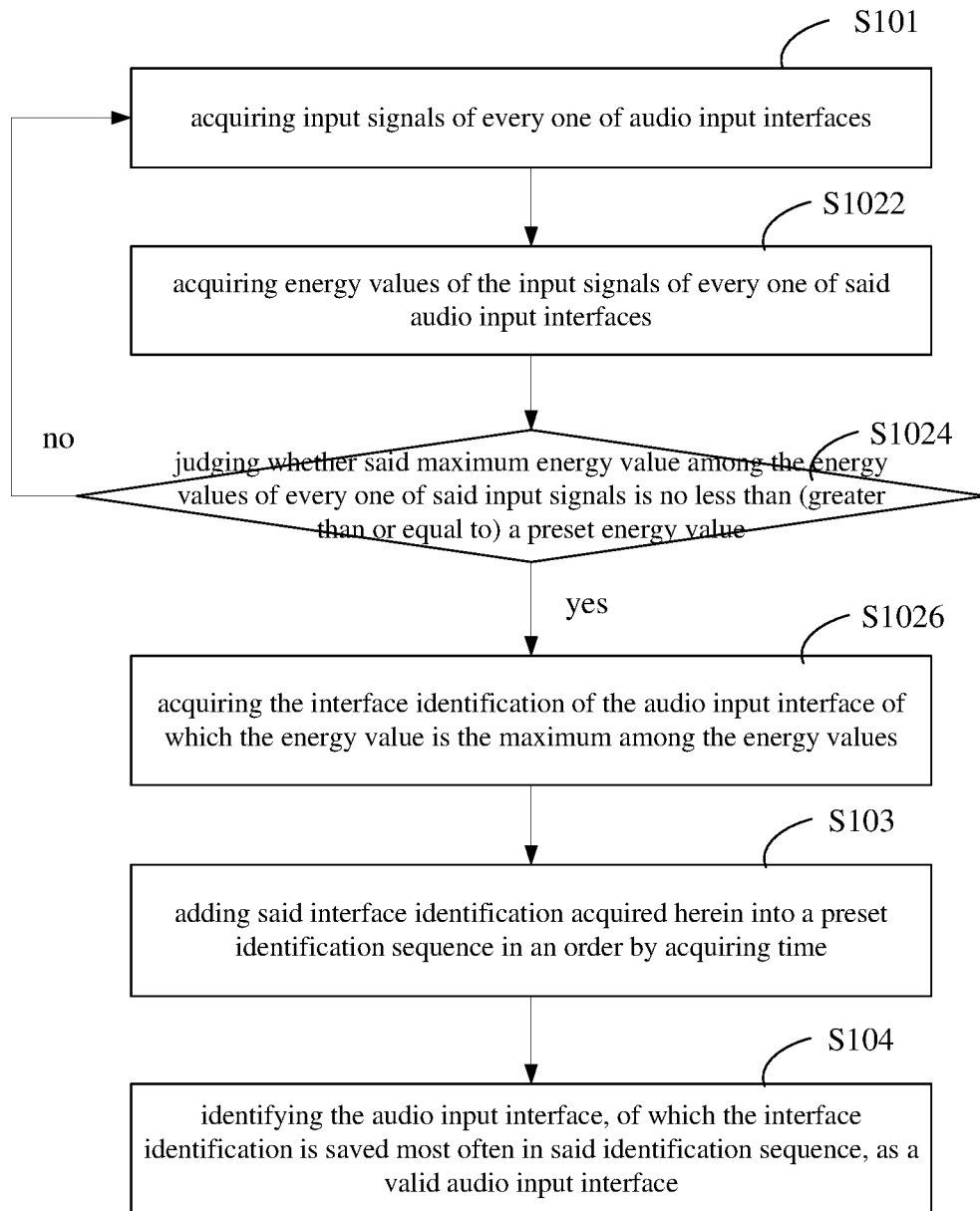


FIG. 2

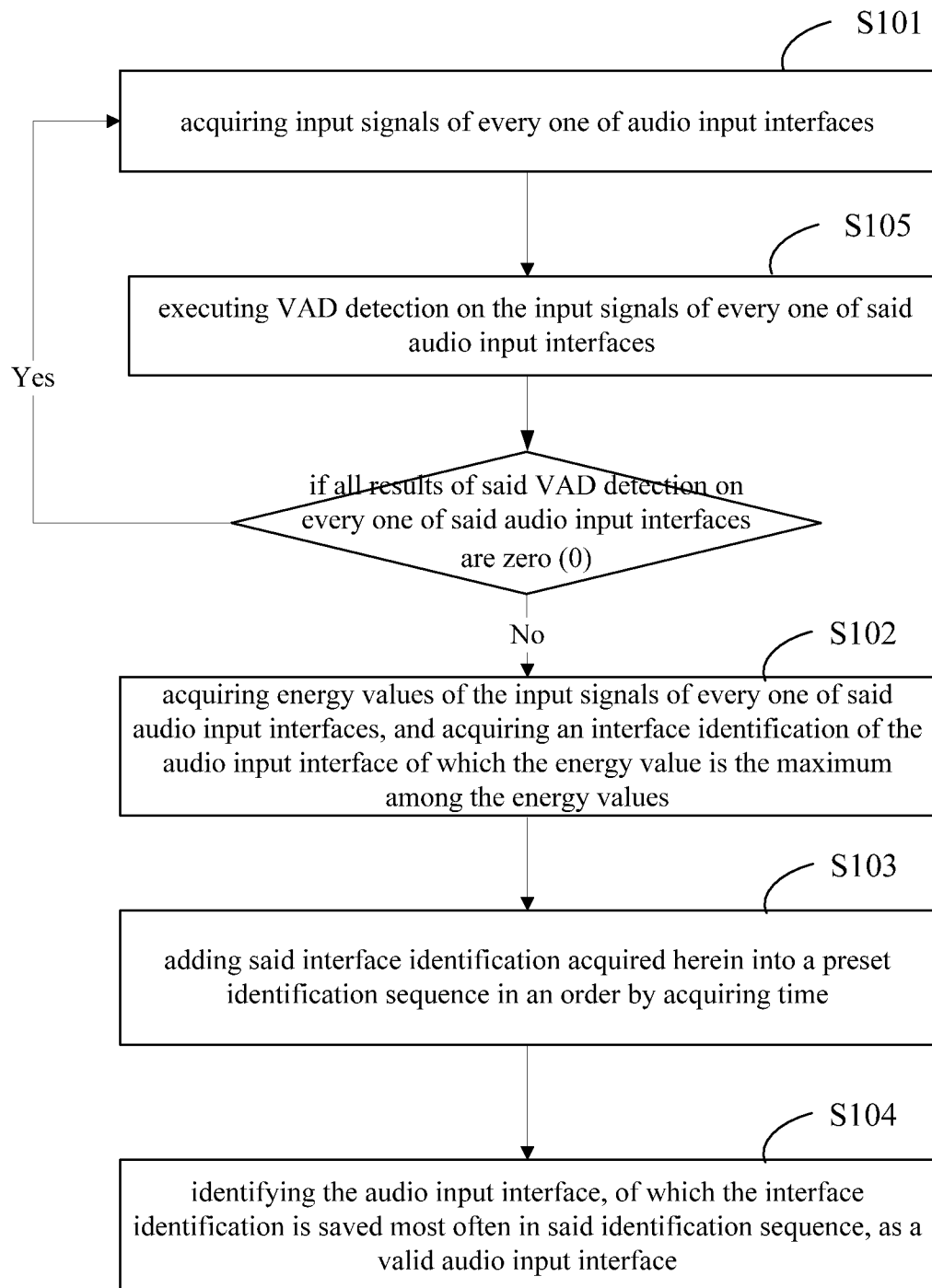


FIG. 3

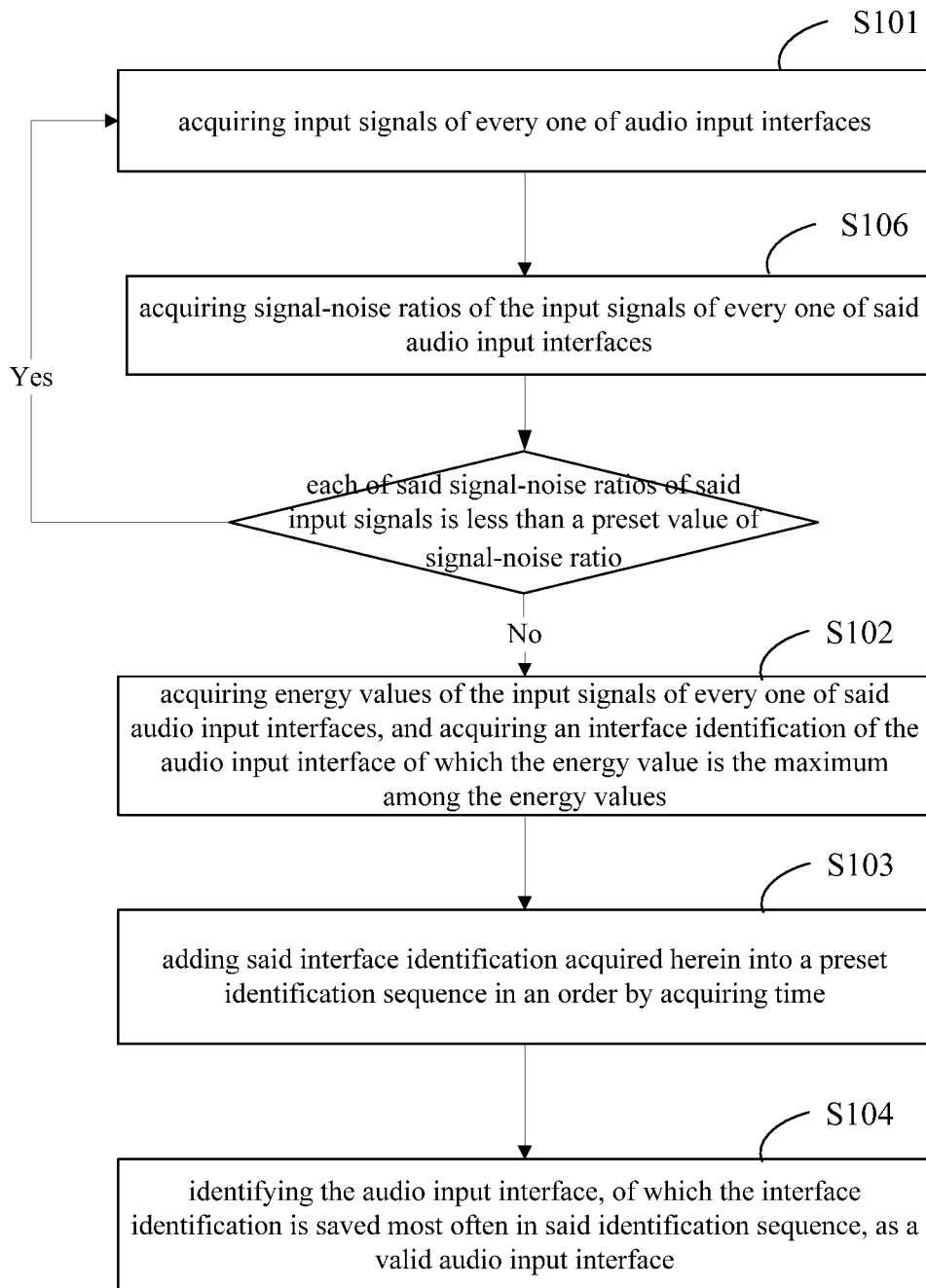


FIG. 4

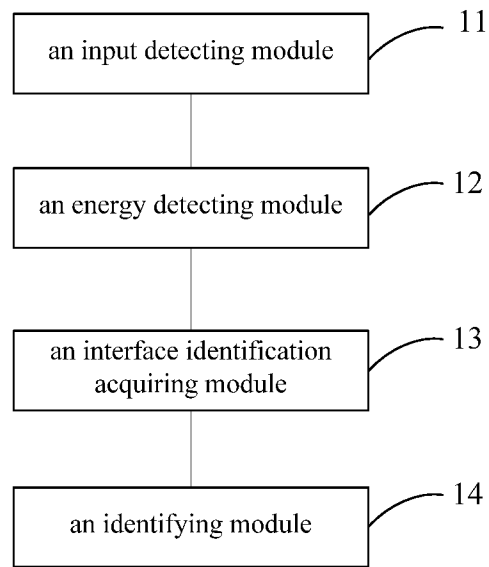


FIG. 5

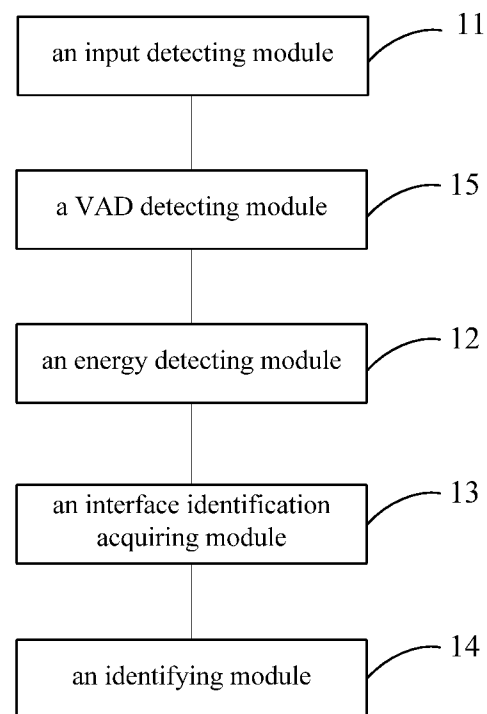


FIG. 6

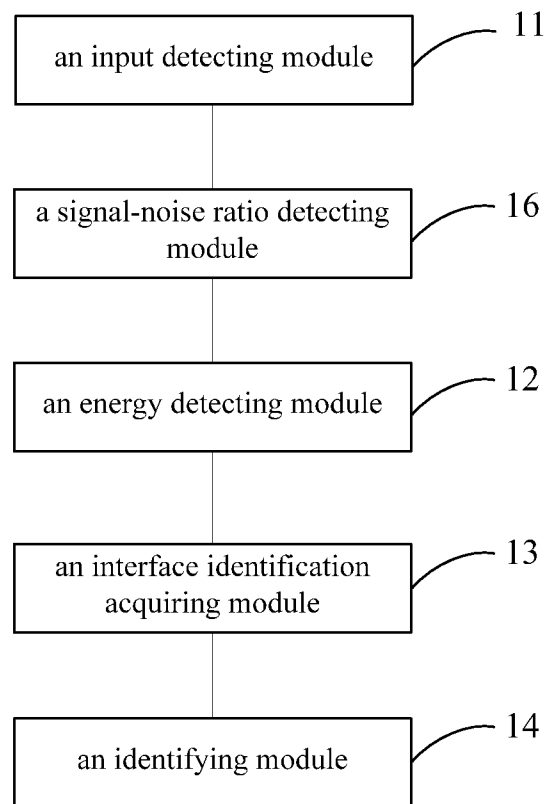


FIG. 7

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2014/075663

A. CLASSIFICATION OF SUBJECT MATTER

H04R 3/00(2006.01)i

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

H04R, H04M

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)

CPRSABS;CNTXT;CNKI;VEN:mic+, audio, sound, voice, input, interface?, channel?, port?, multiple, several, many, series, every, each, select+, choos+, switch+, detect+, identif+, energy, strength, intensity, amplitude, decibel, db, volume, number, amount, frequency, often, frequently

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 103327433A (TENCENT TECHNOLOGY SHENZHEN CO LTD) 25 September 2013 (2013-09-25) claims 1-14	1-14
X	US 2008040117A1 (YU SHUIAN) 14 February 2008 (2008-02-14) description paragraphs [0029]-[0034], [0061]-[0067]	1-5, 8-12, 15-19
A	US 2008040117A1 (YU SHUIAN) 14 February 2008 (2008-02-14) description paragraphs [0029]-[0034], [0061]-[0067]	6, 7, 13, 14, 20
X	CN 1805489A (HUAWEI TECHNOLOGIES CO LTD) 19 July 2006 (2006-07-19) claims 1 and 2	1-5, 8-12, 15-19
A	CN 1805489A (HUAWEI TECHNOLOGIES CO LTD) 19 July 2006 (2006-07-19) claims 1 and 2	6, 7, 13, 14, 20
A	CN 102056053A (ZTE CORP) 11 May 2011 (2011-05-11) the whole document	1-20

☐ 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	“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
“E”	earlier application or patent but published on or after the international filing date	“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
“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)	“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
“O”	document referring to an oral disclosure, use, exhibition or other means	“&”	document member of the same patent family
“P”	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search

11 June 2014

Date of mailing of the international search report

24 June 2014

Name and mailing address of the ISA/

STATE INTELLECTUAL PROPERTY OFFICE OF THE
P.R.CHINA(ISA/CN)
6,Xitucheng Rd., Jimen Bridge, Haidian District, Beijing
100088 China

Authorized officer

TIAN,Shan

Facsimile No. (86-10)62019451

Telephone No. (86-10)62089397

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/CN2014/075663

Patent document cited in search report		Publication date (day/month/year)	Patent family member(s)		Publication date (day/month/year)
CN	103327433A	25 September 2013	None		
US	2008040117A1	14 February 2008	CN	100466671C	04 March 2009
			US	8335686B2	18 December 2012
			CN	1697472A	16 November 2005
			WO	2005112413A1	24 November 2005
CN	1805489A	19 July 2006	WO	2006081728A1	10 August 2006
			CN	100484175C	29 April 2009
CN	102056053A	11 May 2011	WO	2012079459A1	21 June 2012