(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
18 August 2016 (18.08.2016)

WIPO | PCT

(10) International Publication Number
WO 2016/130520 A1

(51) International Patent Classification:
G10L 15/22 (2006.01) *G10L 15/28* (2006.01)

(21) International Application Number:
PCT/US2016/017110

(22) International Filing Date:
9 February 2016 (09.02.2016)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/115,898 13 February 2015 (13.02.2015) US

(71) Applicant: **KNOWLES ELECTRONICS, LLC** [US/US]; 1151 Maplewood Drive, Itasca, Illinois 60143 (US).

(72) Inventor: **RAGHUVIR, Ramanujapuram**; 3600 York Road, Oak Brook, Illinois 60523 (US).

(74) Agents: **SMITH, Troy D.** et al.; Foley & Lardner LLP, 3000 K Street, NW, 6th Floor, Washington, DC 20007-5109 (US).

(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, 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, 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))*

(54) Title: AUDIO BUFFER CATCH-UP APPARATUS AND METHOD WITH TWO MICROPHONES

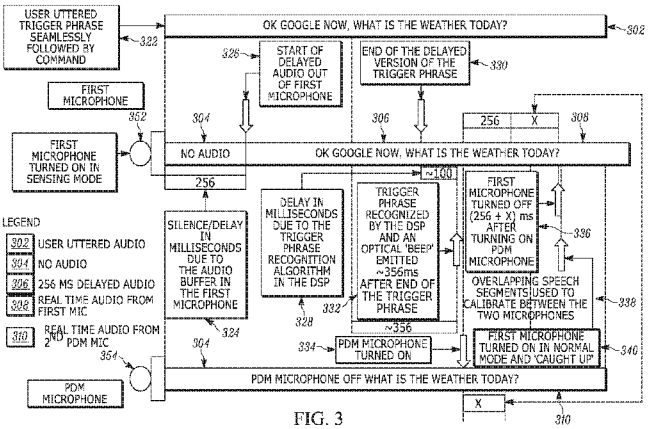

FIG. 3

(57) Abstract: A first microphone is operated in a low power sensing mode, and a buffer at the first microphone is used to temporarily store at least some of the phrase. Subsequently the first microphone is deactivated, then the first microphone is re-activated to operate in normal operating mode where the buffer is no longer used to store the phrase. The first microphone forms first data that does not include the entire phrase. A second microphone is maintained in a deactivated mode until the trigger portion is detected in the first data, and when the trigger portion is detected, the second microphone is caused to operate in normal operating mode where no buffer is used. The second microphone forms second data that does not include the entire phrase. A first electronic representation of the phrase as received at the first microphone and a second electronic representation of the phrase as received at the second microphone are formed from selected portions of the first data and the second data.

WO 2016/130520 A1

AUDIO BUFFER CATCH-UP APPARATUS AND METHOD WITH TWO MICROPHONES

CROSS-REFERENCE TO RELATED APPLICATION

[0001] This patent claims benefit under 35 U.S.C. §119(e) to United States Provisional Application No. 62115898 entitled "Audio Buffer Catch-up Apparatus and Method with Two Microphones" filed February 13, 2015, the content of which is incorporated herein by reference in its entirety.

TECHNICAL FIELD

[0002] This application relates to microphones and, more specifically, to approaches for operating these microphones.

BACKGROUND

[0003] Microphones are used to obtain a voice signal from a speaker. Once obtained, the signal can be processed in a number of different ways. A wide variety of functions can be provided by today's microphones and they can interface with and utilize a variety of different algorithms.

[0004] Voice triggering, for example, as used in mobile systems is an increasingly popular feature that customers wish to use. For example, a user may wish to speak commands into a mobile device and have the device react in response to the commands. In these cases, a voice activity detector may first detect whether there is voice in an audio signal captured by a microphone, and then, subsequently, analysis is performed on the signal to predict what the spoken word was in the received audio signal. Various voice activity detection (VAD) approaches have been developed and deployed in various types of devices such as cellular phones and personal computers.

[0005] Microphones that are always on are often equipped with internal oscillators and operate at very low power. Low power microphones are used in various applications and sometimes two or more microphones are used when the device is brought out of the low power mode. Although the low power aspect allows some of the microphones to be on all the time in a low power listening mode, the microphones may also use buffers to aid in voice activity detection, which introduce processing delays. The processing delays may cause problems at the far end of the system where the signals frequently need to be processed as quickly as possible.

[0006] The problems of previous approaches have resulted in some user dissatisfaction with these previous approaches.

BRIEF DESCRIPTION OF THE DRAWINGS

[0007] For a more complete understanding of the disclosure, reference should be made to the following detailed description and accompanying drawings wherein:

[0008] FIG. 1 comprises a system using two microphones that implements an audio buffer catch-up approach;

[0009] FIG. 2 comprises a flowchart showing an audio buffer catch-up approach;

[0010] FIG. 3 comprises a time line and flowchart showing one example of an audio buffer catch-up approach;

[0011] FIG. 4 comprises the time line showing an example of an audio buffer catch-up approach.

[0012] Skilled artisans will appreciate that elements in the figures are illustrated for simplicity and clarity. It will further be appreciated that certain actions and/or steps may be described or depicted in a particular order of occurrence while those skilled in the art will understand that such specificity with respect to sequence is not actually required. It will also be understood that the terms and expressions used herein have the ordinary meaning as is accorded to such terms and expressions with respect to their corresponding

respective areas of inquiry and study except where specific meanings have otherwise been set forth herein.

DETAILED DESCRIPTION

[0013] The present approaches utilize two (or potentially more) microphones to obtain speech phrases from an utterance of a speaker. The effects of delays caused by buffers in one of the microphones are significantly reduced or eliminated. The approaches described herein are easy to implement and eliminate problems and limitations associated with prior approaches.

[0014] Referring now to FIG. 1, one example of a system 100 for audio buffer catch-up is described. The system 100 includes a first microphone 102 (a microphone with a built in voice activity detector, manufactured by Knowles, Inc., in one example), a second microphone 104 (a pulse density modulation (PDM) microphone, in one example), and a processor 106. The processor 106 may include a codec 108 and an application processor 110.

[0015] The first microphone 102 and second microphone 104 may be micro electro mechanical system (MEMS) microphones. In one example, these microphones are assemblies including a sensing element (diaphragm and back plate) and an application specific integrated circuit (which includes a buffer in the case of microphone 102 and potentially performs other processing functions). Sound energy is received by the microphones, moves the diaphragms and produces an electrical signal (which may or may not be buffered).

[0016] The processing device 106 may include a codec 108 and an application processor 110. The codec 108 in this example may supply the clock signals to the microphones 102 and 104, and may perform other signal processing functions. The application processor 110 may also perform processing related to the device in which the microphones 102 and 104 are deployed. For example, if the microphones 102 and 104 are deployed in a cellular phone, the application processor 110 may perform processing associated with the cellular phone. Although both a codec 108 and an application

processor 110 are shown here, it will be appreciated that these devices can be merged together into a single processing device.

[0017] A clock signal 112 is applied to the microphones. Application of the clock signal, when applying the power signal to the microphone, causes the first microphone 102 to operate in a normal operating mode where incoming data to the ASIC is not buffered, but passed through to the output of the microphone 102. Non-application of the clock signal after power has been applied to the microphone, causes the first microphone 102 to operate in a low power operating mode. In this mode, incoming data to the ASIC is buffered and not directly passed through to the output of the microphone 102, thereby introducing a buffering delay, in one example of 256 milliseconds. The clock signal may be applied at one frequency when the microphone is in low power mode after acoustic activity has been detected and may be applied at the same or different frequency in the normal operating mode.

[0018] In one example of the operation of the system of FIG. 1, the first microphone 102 is on and uses its buffer (i.e., incoming audio data must pass through the buffer). The second microphone 104 is off. Incomplete segments of a phrase (e.g., OK GOOGLE NOW, WHAT IS THE WEATHER TODAY?) are received by at least one, and in some embodiments both, microphones 102 and 104. When the processing device 106 detects speech in data provided by the first microphone, the second microphone 104 is activated and begins to provide unbuffered (real time) data to the processing device. The second microphone may be activated by applying the clock signal 112 to the second microphone 104.

[0019] In one embodiment, the processing device determines whether the phrase includes a trigger phrase (e.g., OK GOOGLE NOW) within a phrase segment received at the first microphone 102. By "trigger phrase", it is meant any phrase that signifies that a command is immediately present after the trigger phrase. The second microphone 104 is turned on by the processor 106 as a result of the trigger phrase having been detected at the first microphone; the second microphone 104 after activation captures voice data in real time.

[0020] In one embodiment, the first microphone 102 is turned off by the processor 106 after a time period of (delay + x) where delay is the buffering delay of the first microphone 102 and x is the period of common speech information that has been received at each of the microphones 102 and 104. In one example, x can be determined by the algorithm required to calibrate the two microphones. This calibration may include determination of and compensation for the acoustic delay and gain difference between microphones 102 and 104. In embodiments where the first microphone is turned off, the first microphone is quickly turned back on after being turned off (e.g., within approximately 20 milliseconds) and placed in a normal mode of operation by receiving a clock signal 112 from the processor 106.

[0021] As will be more fully apparent from the discussion below, at least one and in some cases both microphones do not detect the entire uttered phase (e.g., OK GOOGLE NOW, WHAT IS THE WEATHER TODAY?) and thus the one or more microphones do not provide data derived from or corresponding to the entire phrase to the processor 106 for further processing. At the processor 106, the entire phrase (e.g., OK GOOGLE NOW, WHAT IS THE WEATHER TODAY?) is stitched together for each microphone 102 and 104 based upon information received from both microphones. An output 114 from processor 106 includes assembled phrases 116 and 118 (e.g., each being OK GOOGLE NOW, WHAT IS THE WEATHER TODAY?) with the first phrase 116 being associated with the first microphone 102 and the second phrase 118 being associated with the second microphone 104. It will be appreciated that this various processing described above can occur at either the codec 108 or the application processor 110, or at other processing devices (not shown in FIG. 1).

[0022] Referring now to FIG. 2, one example of an approach for audio buffer catch-up is described. This processing occurs at the processor 106 (and in one specific example at the codec 108 although this functionality can be moved to or shared with the application processor 110).

[0023] At step 202, the first microphone is on and uses the buffer. The second microphone is off.

[0024] At step 204, the trigger phrase is detected from data received from the first microphone. At step 206, the second microphone is turned on as a result of the trigger phrase having been detected. At step 208, the second microphone captures voice data in real time.

[0025] At step 210, the first microphone is turned off after a time period of (delay + x) where the delay is the buffering delay of the first microphone (i.e., how long data takes to move through its buffer) and x is the period of common speech between the two microphones. In one example, x can be determined by the algorithm required to calibrate the two microphones.

[0026] At step 212, the first microphone is quickly turned on after being turned off (e.g., the microphone is activated approximately 20 milliseconds after being deactivated) and placed in a normal mode of operation (i.e., a non-buffering mode of operation as explained elsewhere herein). At step 214, data derived from segments of the phrase received by the plural microphones are stitched together using suitable algorithms to form electronic representations of the entire phrase, one associated with each microphone. One example of assembling data from different microphones to form two separate electronic representations of complete phrases (one for first microphone and the other for the second microphone) is described below with respect to FIGs. 3 and 4.

[0027] Referring now to FIG. 3, one example of an approach for audio catch-up using two microphones is described. It will be appreciated that this is one example of an approach using two microphones and that other examples are possible.

[0028] At step 322, a user utters a trigger phrase (e.g., OK GOOGLE NOW) seamlessly followed by a command (WHAT IS THE WEATHER TODAY?). In this example, the trigger phrase and command are collectively labeled as 302. A first microphone 352 and a second microphone 354 (PDM microphone) detect parts of the complete phrase (e.g., the trigger and command in the example above). The first microphone 352 is in a low power sensing mode and all signals in this mode are buffered before being output, consequently introducing a buffer delay 324 (e.g., 256 milliseconds). During this delay time, a time period 304 exists at the output of the first microphone 352

where no audio is being supplied. At time 326, the start of the audio output of the first microphone 352 occurs. As mentioned, the buffer delay is approximately 256 milliseconds and the 256 millisecond delayed output occurs at the output of the first microphone 352 during period 306.

[0029] Another delay 328 (in this case an approximately 100 millisecond delay) may be introduced by the trigger phrase recognition algorithm in the processor (e.g., codec, applications processor, or digital signal processor to mention a few examples). The trigger phrase recognition algorithm compares the received audio to a predefined trigger word or phrase, to determine whether the trigger word or phrase has been uttered. The delay 328 may occur after time 330, which is the end of the delayed version of the trigger phrase. At time 332, after an approximately 256 plus 100 millisecond delay, a beep (or other signal) is emitted or presented to the user signifying that the trigger phrase has been detected. In some examples, no beep may be used and in other examples the "beep" may be inaudible to humans. This signal may be a marker used in later processing and may be removed before stitching together the various speech segments.

[0030] At time 334, the second microphone 354 is turned on. Prior to the second microphone 354 being turned on, a time period 304 exists at its output where no audio is being produced.

[0031] At time 336, the first microphone 352 is turned off a predetermined time after turning on the second microphone 354. The predetermined time may be 256 milliseconds (the buffer delay of the first microphone 352) plus x, where x is a time period 338 of overlapping speech segments that is used to determine the acoustic delay between microphones 352 and 354. As shown here in this example, x relates to the phrase "E WEA" because "E WEA" is the audio that has been received at both of the microphones 352 and 354. In these regards, the processor can first determine the common audio information received, and then use that to calibrate the microphone signals. This common time period (in this case, the length in time of the phrase "E WEA") is the value of x.

[0032] At step 340, the first microphone 352 is quickly or immediately turned on after a very small (e.g., after an approximately 20 millisecond delay) to operate in a normal processing mode. By normal processing or operating mode, it is meant that the microphone 352 does not buffer the incoming signal, but passes data through without buffering the data. In one example, normal operating mode may be entered by applying a clock signal when applying the power signal, while in low power mode no clock signal is applied, till acoustic activity is detected.

[0033] Referring now to FIG. 4, further explanation of assembling the phrases detected at two microphones 352 and 354 is provided. FIG. 4 illustrates assembling the information together to form the complete phrases 430 and 432 for each of the microphones 352 and 354 respectively. When each of the phrases is assembled, each phrase "OK GOOGLE NOW, WHAT IS THE WEATHER TODAY?" will be complete, but the internal characteristics (e.g., amplitude, acoustic delay) within the phrase may be different for each microphone. These phrases may be used in further processing (e.g., transporting to the 'cloud' for speech recognition in a cellular phone, tablet, or personal computer).

[0034] As shown, user uttered audio 402 is in this example OK GOOGLE NOW, WHAT IS THE WEATHER TODAY? There is a large missing segment of missing audio (at period 410) at the second microphone 354. However, the first microphone 352 has obtained this audio at time period 404. Thus, audio (for time period 404) can be used to obtain audio (for time period 410) for the second microphone, after appropriate calibration.

[0035] Time period 406 is missing from the first microphone 352. But, the second microphone 354 has obtained this audio and so this can be included in the audio for the first microphone 322.

[0036] The first microphone 352 has obtained audio for time period 408 in real time. Consequently, the complete audio phrase (430) ("OK GOOGLE NOW, WHAT IS THE WEATHER TODAY?") has been assembled for the first microphone 352 since time periods 404, 406, and 408 have been filled in.

[0037] The second microphone 354 obtains real time audio for time period 412. Consequently, the audio phrase 432 ("OK GOOGLE NOW, WHAT IS THE WEATHER TODAY?") has been assembled for the second microphone 354 since time periods 410 and 412 have been filled in.

[0038] In this way, the audio phrases 430 (from the first microphone 352) and 432 (for the second microphone 354) are assembled or stitched together. This processing may occur at a codec, application processor, or digital system processor to mention a few examples. The phrases 430 and 432 may be further processed by other processing devices as needed.

[0039] Preferred embodiments of this disclosure are described herein, including the best mode known to the inventors. It should be understood that the illustrated embodiments are exemplary only, and should not be taken as limiting the scope of the appended claims.

WHAT IS CLAIMED IS:

1. A method of forming multiple electronic representations of an uttered phrase, the method comprising:

receiving buffered data at a microphone interface of a processor, the buffered data derived from a first segment of the uttered phrase by a first microphone;

receiving first unbuffered data at the microphone interface of the processor, the first unbuffered data derived from a second segment of the uttered phrase by the first microphone, the first segment different than the second segment;

receiving second unbuffered data at the microphone interface of the processor, the second unbuffered data derived from less than the entire uttered phrase by a second microphone;

forming a first electronic representation of the entire uttered phrase based on buffered data and based on first unbuffered data; and

forming a second electronic representation of the entire uttered phrase based on buffered data and based on second unbuffered data.

2. The method of Claim 1, the buffered data and the first unbuffered data derived from less than the entire uttered phrase by the first microphone, forming the first electronic representation of the entire uttered phrase based on the second unbuffered data.

3. The method of Claim 1,

providing a second microphone enablement signal at the microphone interface after receiving at least some of the buffered data at the microphone interface,

receiving the second unbuffered data at the microphone interface after providing the second microphone enablement signal.

4. The method of Claim 3,

detecting speech in the buffered data,

providing the second microphone enablement signal after detecting speech in the buffered data.

5. The method of Claim 1,

receiving the second unbuffered data at the microphone interface after receiving buffered data,

providing a first microphone clock signal to the microphone interface after receiving buffered data,

receiving the first unbuffered data at the microphone interface after providing the first microphone clock signal.

6. The method of Claim 5,

providing a second microphone enablement signal at the microphone interface after receiving buffered data at the microphone interface,

receiving the second unbuffered data at the microphone interface after providing the second microphone enablement signal.

7. The method of Claim 6,

detecting speech in the buffered data received at the microphone interface,

providing the second microphone enablement signal after detecting speech.

8. The method of Claim 5, calibrating the first microphone and second microphone based on common information received from the first microphone and the second microphone at the microphone interface.

9. The method of Claim 1, providing a first microphone deactivation signal and subsequently providing a first microphone reactivation signal at the microphone interface while receiving the second unbuffered data at the microphone interface, receiving the first unbuffered data at the microphone interface after providing the first microphone reactivation signal.

10. The method of Claim 9, providing the first microphone deactivation signal after a delay period occurring when the second unbuffered data is received at the microphone interface.

11. A method of forming multiple electronic representations of a phrase that is uttered by a user and received at different microphones, the method comprising:

receiving buffered first data from a first microphone that is being operated in a low power sensing mode, the buffered first data having been buffered at the first microphone;

determining whether speech is present in the buffered first data received from the first microphone;

when speech is determined to be present in the buffered first data, activating a second microphone;

subsequently receiving unbuffered second data from the second microphone in real time, the unbuffered second data not having been buffered at the second microphone;

deactivating the first microphone after a delay period while continuing to receive the unbuffered second data in real time from the second microphone, the deactivating causing the first microphone to fail to sense a part of the phrase, the unbuffered second data including the part of the phrase not sensed by the first microphone;

re-activating the first microphone, the re-activating causing the first microphone to operate in a normal operating mode, wherein the first microphone produces unbuffered first data when in normal operating mode;

forming a first electronic representation of the entire phrase, the first electronic representation being assembled from the buffered first data, the unbuffered second data, and the unbuffered first data;

forming a second electronic representation of the entire phrase, the second electronic representation being assembled from the buffered first data and the unbuffered second data.

12. The method of claim 11,

the phrase includes a trigger and a command,

the first data is different than the second data,

determining whether speech is present includes determining whether a trigger is present in the buffered first data received from the first microphone,

activating the second microphone when the trigger is determined to be present in the buffered first data.

13. The method of claim 11, wherein the first microphone and the second microphone are micro electro mechanical system (MEMS) microphones.

14. The method of claim 11, wherein the first microphone includes a voice activity detector.

15. The method of claim 11, wherein the second microphone is a pulse density modulation (PDM) microphone.

16. A system of forming multiple electronic representations of a phrase uttered by a user and received at different microphones, the system comprising:

a first microphone that includes a buffer, wherein the buffer at the first microphone temporarily stores data corresponding to at least some of the phrase;

a second microphone;

a control apparatus coupled to the first microphone and the second microphone, the control apparatus having an output, the first microphone forming first data including buffered data corresponding to at least a portion of the phrase, the control apparatus further configured to maintain the second microphone in a deactivated mode until speech is detected in the first data, and when speech is detected, cause the second microphone to provide unbuffered second data , the second data corresponding to less than the entire phrase, the control apparatus configured to form a first electronic representation of the entire phrase and a second electronic representation of the entire phrase , the first electronic representation and second electronic representation being assembled from

selected portions of the first data and the second data, the first electronic representation and the second electronic representation be transmitted at the output.

17. The system of claim 16,

the first data is different than the second data,

the first microphone includes a voice activity detector and is operated in a low power sensing mode when the buffer at the first microphone temporarily stores data,

the control apparatus configured to maintain the second microphone in a deactivated mode until the control apparatus detects speech in the buffered data, and when speech is detected, cause the second microphone to provide unbuffered second data to the control apparatus.

18. The system of Claim 17,

after activating the second microphone, the control apparatus configured to deactivate the first microphone and then re-activate the first microphone to operate in a mode where data is not buffered,

the first data including buffered data and unbuffered data from the first microphone, the first data corresponding to less than the entire phrase,

the control apparatus configured to form the first electronic representation of the entire phrase from the first data and at least a portion of the second data, and the control apparatus configured to form the second electronic representation of the entire phrase from the buffered data and the second data.

19. The system of claim 17, the first microphone and the second microphone are micro electro mechanical system (MEMS) microphones.

20. The system of claim 17, wherein the second microphone is a pulse density modulation (PDM) microphone.

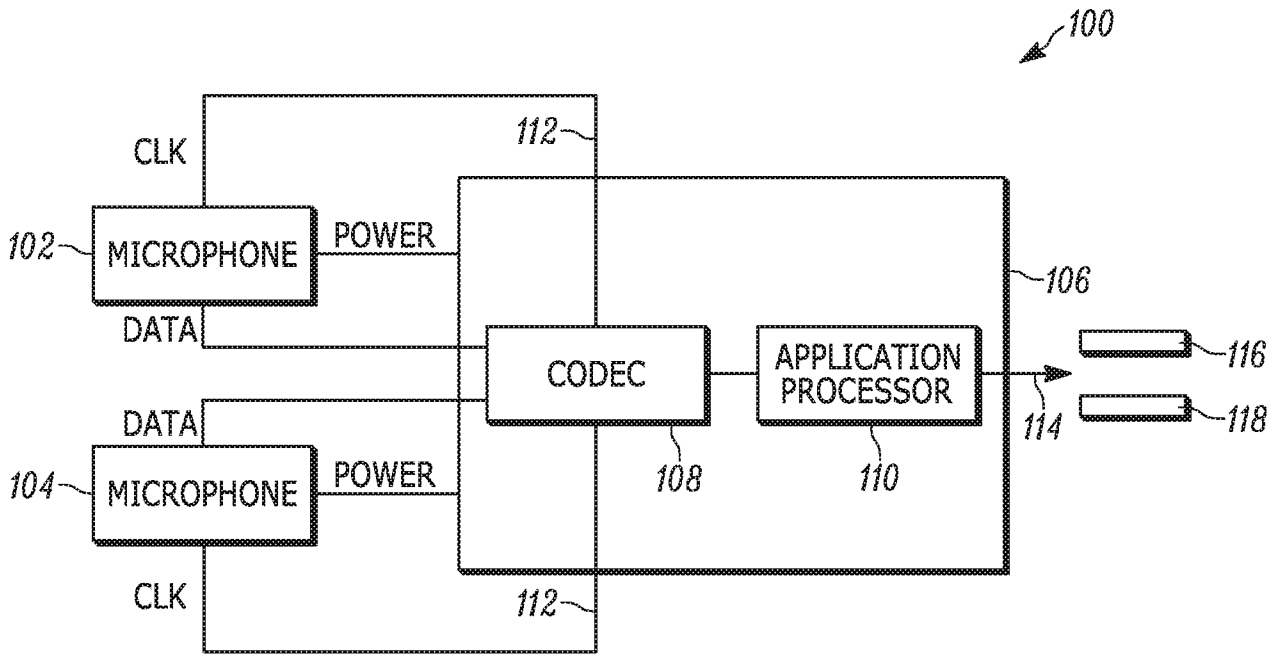
100
112
CLK
102
MICROPHONE
POWER
106
DATA
CODEC
APPLICATION PROCESSOR
116
DATA
114
118
104
MICROPHONE
POWER
108
110
CLK
112

FIG. 1

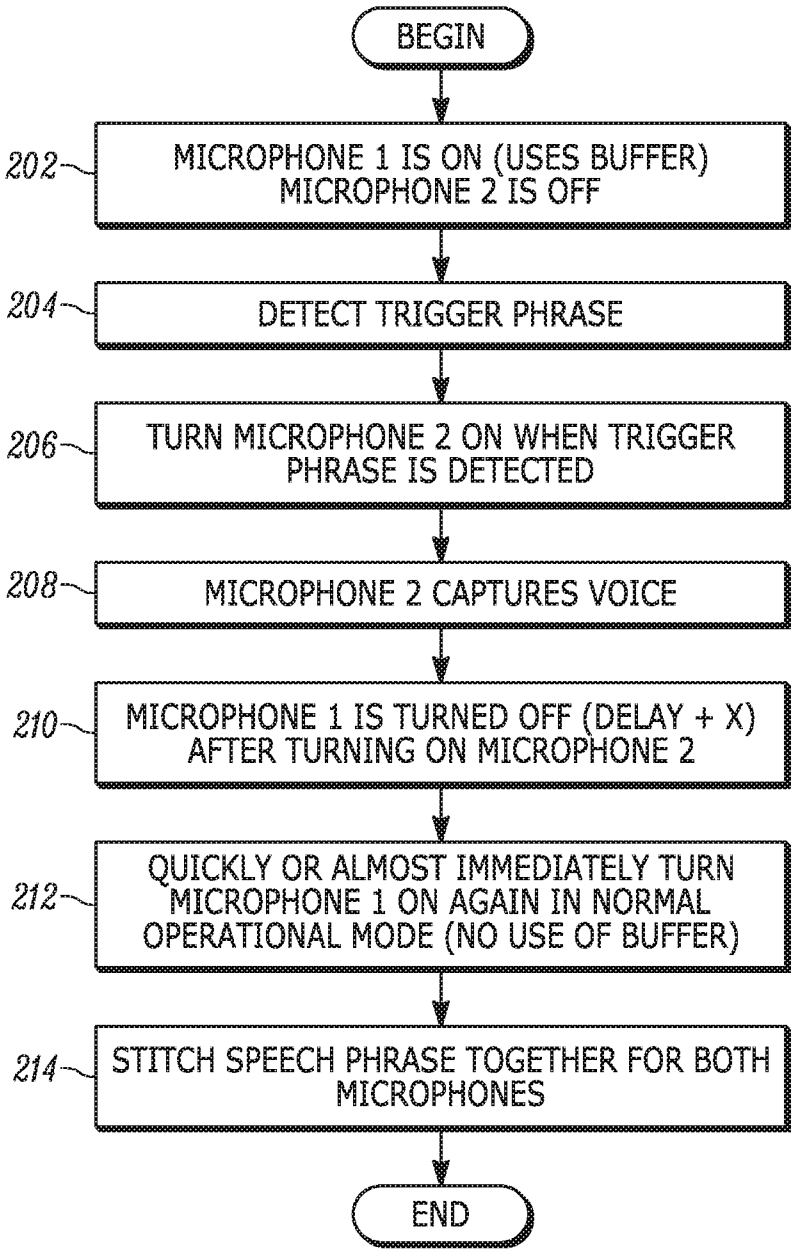
BEGIN
202
MICROPHONE 1 IS ON (USES BUFFER)
MICROPHONE 2 IS OFF
204
DETECT TRIGGER PHRASE
206
TURN MICROPHONE 2 ON WHEN TRIGGER PHRASE IS DETECTED
208
MICROPHONE 2 CAPTURES VOICE
210
MICROPHONE 1 IS TURNED OFF (DELAY + X) AFTER TURNING ON MICROPHONE 2
212
QUICKLY OR ALMOST IMMEDIATELY TURN MICROPHONE 1 ON AGAIN IN NORMAL OPERATIONAL MODE (NO USE OF BUFFER)
214
STITCH SPEECH PHRASE TOGETHER FOR BOTH MICROPHONES
END

FIG. 2

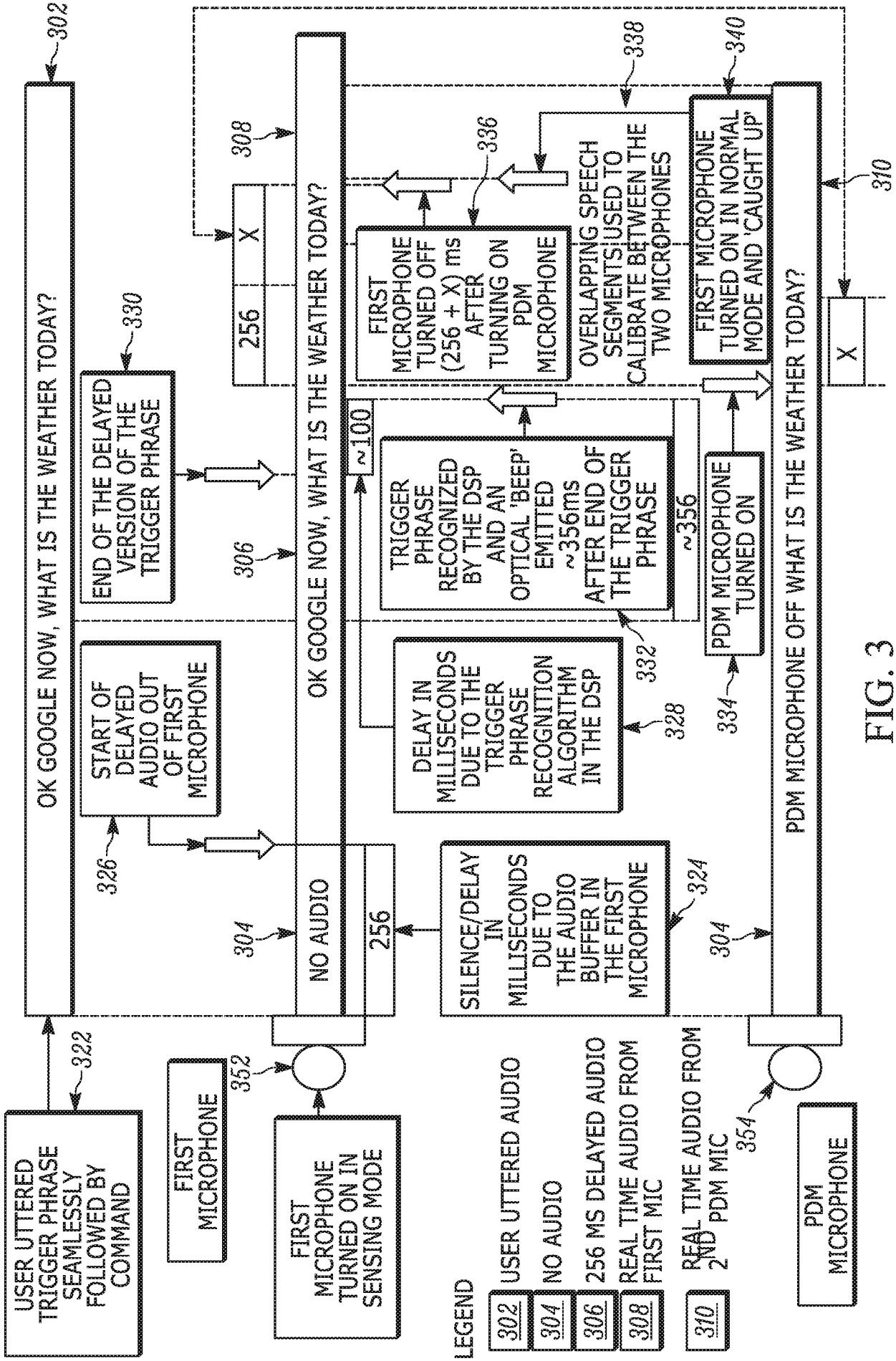
302
OK GOOGLE NOW, WHAT IS THE WEATHER TODAY?
330
END OF THE DELAYED VERSION OF THE TRIGGER PHRASE
326
START OF DELAYED AUDIO OUT OF FIRST MICROPHONE
304
NO AUDIO
256
306
OK GOOGLE NOW, WHAT IS THE WEATHER TODAY?
308
256
X
~100
TRIGGER PHRASE RECOGNIZED BY THE DSP AND AN OPTICAL 'BEEP' EMITTED ~356ms AFTER END OF THE TRIGGER PHRASE
~356
DELAY IN MILLISECONDS DUE TO THE TRIGGER PHRASE RECOGNITION ALGORITHM IN THE DSP
328
332
SILENCE/DELAY IN MILLISECONDS DUE TO THE AUDIO BUFFER IN THE FIRST MICROPHONE
324
FIRST MICROPHONE TURNED OFF (256 + X) ms AFTER TURNING ON PDM MICROPHONE
336
338
OVERLAPPING SPEECH SEGMENTS USED TO CALIBRATE BETWEEN THE TWO MICROPHONES
340
FIRST MICROPHONE TURNED ON IN NORMAL MODE AND 'CAUGHT UP'
334
PDM MICROPHONE TURNED ON
304
PDM MICROPHONE OFF WHAT IS THE WEATHER TODAY?
310
X
322
USER UTTERED TRIGGER PHRASE SEAMLESSLY FOLLOWED BY COMMAND
FIRST MICROPHONE
352
FIRST MICROPHONE TURNED ON IN SENSING MODE
354
PDM MICROPHONE
LEGEND
302 USER UTTERED AUDIO
304 NO AUDIO
306 256 MS DELAYED AUDIO
308 REAL TIME AUDIO FROM FIRST MIC
310 REAL TIME AUDIO FROM 2ND PDM MIC

FIG. 3

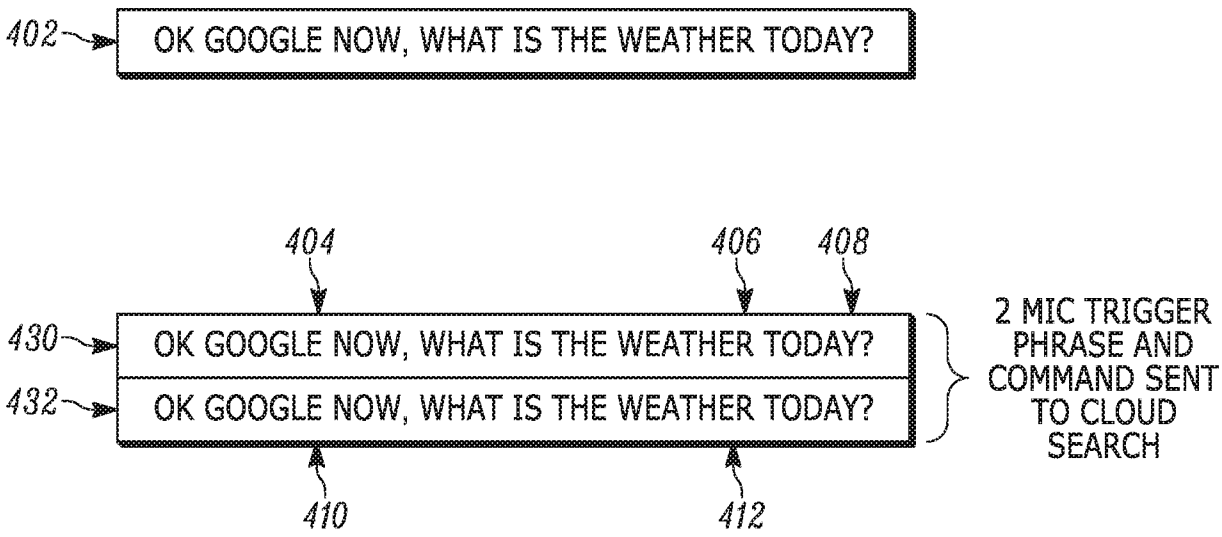
402
OK GOOGLE NOW, WHAT IS THE WEATHER TODAY?
404
406
408
430
OK GOOGLE NOW, WHAT IS THE WEATHER TODAY?
432
OK GOOGLE NOW, WHAT IS THE WEATHER TODAY?
2 MIC TRIGGER PHRASE AND COMMAND SENT TO CLOUD SEARCH
410
412
LEGEND
402 USER UTTERED AUDIO
404 256 MS DELAYED AUDIO
406 PHASE COMPENSATED AUDIO FOR FIRST MIC
408 REAL TIME AUDIO FROM FIRST MIC
410 PHASE COMPENSATED AUDIO FOR 2ND PDM MIC
412 REAL TIME AUDIO FROM 2ND PDM MIC

FIG. 4

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2016/017110

A. CLASSIFICATION OF SUBJECT MATTER

G10L 15/22(2006.01)i, G10L 15/28(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)
G10L 15/22; G10L 15/20; G10L 15/00; G06F 3/16; H04R 3/00; G10L 21/16; G10L 15/28

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
Korean utility models and applications for utility models
Japanese utility models and applications for utility models

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
eKOMPASS(KIPO internal) & Keywords: second microphone, voice recognition, power-on, deactive, buffer, and similar terms.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2014-0244273 A1 (JEAN LAROCHE et al.) 28 August 2014 See paragraphs [0037] and [0042]-[0043]; and claims 1, 11-12, and 24.	16-17,19-20
Y		1,3-4
A		2,5-15,18
Y	US 2014-0214429 A1 (LOTHAR PANTEL) 31 July 2014 See paragraphs [0033] and [0039]; and figures 2-3.	1,3-4
A	US 2015-0039303 A1 (WOLFSON MICROELECTRONICS PLC.) 05 February 2015 See paragraphs [0068]-[0078] and figure 5a.	1-20
A	KR 10-2012-0112325 A (E & M SYSTEMS LTD.) 11 October 2012 See paragraphs [0035]-[0039] and figure 3.	1-20
A	US 2014-0348345 A1 (KNOWLES ELECTRONICS, LLC.) 27 November 2014 See paragraphs [0032]-[0038] and figure 2.	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
"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	Date of mailing of the international search report
26 May 2016 (26.05.2016)	**26 May 2016 (26.05.2016)**

Name and mailing address of the ISA/KR
International Application Division
Korean Intellectual Property Office
189 Cheongsa-ro, Seo-gu, Daejeon, 35208, Republic of Korea
Facsimile No. +82-42-481-8578

Authorized officer
NHO, Ji Myong
Telephone No. +82-42-481-8528

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.
PCT/US2016/017110

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2014-0244273 A1	28/08/2014	CN 104247280 A	24/12/2014
		EP 2962403 A1	06/01/2016
		KR 10-2015-0121038 A	28/10/2015
		WO 2014-134216 A1	04/09/2014
		WO 2014-134216 A9	15/10/2015
US 2014-0214429 A1	31/07/2014	AU 2014-200407 A1	14/08/2014
		DE 102013001219 A1	31/07/2014
		GB 2512178 A	24/09/2014
		GB 2512178 B	04/11/2015
		IE 20140051 A1	13/08/2014
		IE 86422 B1	13/08/2014
US 2015-0039303 A1	05/02/2015	CN 104252860 A	31/12/2014
		GB 2515526 A	31/12/2014
		GB 2515527 A	31/12/2014
		GB 2515528 A	31/12/2014
KR 10-2012-0112325 A	11/10/2012	KR 10-1192813 B1	26/10/2012
US 2014-0348345 A1	27/11/2014	KR 10-2016-0010606 A	27/01/2016
		US 2015-0055799 A1	26/02/2015
		US 2015-0055803 A1	26/02/2015
		US 2015-0058001 A1	26/02/2015
		US 2015-043755 A1	12/02/2015
		US 2015-350774 A1	03/12/2015
		US 9111548 B2	18/08/2015
		US 9113263 B2	18/08/2015
		WO 2014-189931 A1	27/11/2014
		WO 2015-069878 A1	14/05/2015