

(51) International Patent Classification:
G10L 25/78 (2013.01)(21) International Application Number:
PCT/US2014/060567(22) International Filing Date:
15 October 2014 (15.10.2014)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/892,755 18 October 2013 (18.10.2013) US(71) Applicant: **KNOWLES ELECTRONICS, LLC**
[US/US]; 1151 Maplewood Drive, Itasca, Illinois 60143 (US).(72) Inventors: **NANDY, Dibyendu**; 1044 Thackery Lane, Naperville, Illinois 60564 (US). **LI, Yang**; 1029 Charlela Lane, Elk Grove Village, Illinois 60007 (US). **THOMSEN, Henrik**; Sydskraenten 29, DK-2840 Holte (DK). **FURST, Claus**; Nordhojen 7, DK-4000 Himmelev (DK).(74) Agents: **KRATZ, Rudy** et al.; Fitch, Even, Tabin & Flannery, LLP, 120 S LaSalle Street, Suite 1600, Chicago, Illinois 60603 (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: ACOUSTIC ACTIVITY DETECTION APPARATUS AND METHOD

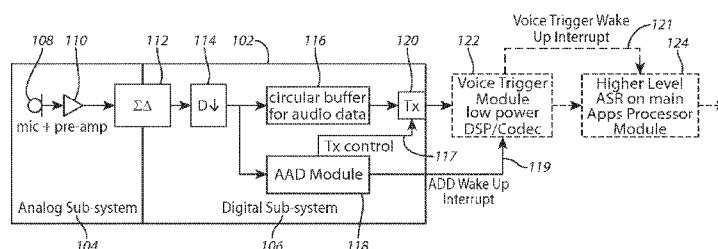


FIG. 1

(57) Abstract: Streaming audio is received. The streaming audio includes a frame having plurality of samples. An energy estimate is obtained for the plurality of samples. The energy estimate is compared to at least one threshold. In addition, a band pass estimate of the signal is determined. An energy estimate is obtained for the band-passed plurality of samples. The two energy estimates are compared to at least one threshold each. Based upon the comparison operation, a determination is made as to whether speech is detected.

ACOUSTIC ACTIVITY DETECTION APPARATUS AND METHOD**CROSS REFERENCE TO RELATED APPLICATION**

[0001] This patent claims benefit under 35 U.S.C. §119 (e) to United States Provisional Application No. 61892755 entitled “Acoustic Activity Detection Apparatus and Method” filed October 18, 2013, the content of which is incorporated herein by reference in its entirety.

TECHNICAL FIELD

[0002] This application relates to speech interfaces and, more specifically, to activity detection approaches utilized in these applications.

BACKGROUND OF THE INVENTION

[0003] Speech interfaces have become important features in mobile devices today. Some devices have the capability to respond to speech even when the device’s display is off and in some form of low power mode and potentially at some distance from the user. These requirements place significant demands on system design and performance including the need to keep the microphone in an “always listening” mode.

[0004] In other examples, the device keeps only parts of the signal chain powered up, e.g. the microphone and a digital signal processor (DSP) or central processing unit (CPU), with an algorithm for detecting a “voice trigger.” Upon recognizing a voice trigger, the rest of the system is powered up from its sleep mode to perform the desired computational task.

[0005] The above-mentioned previous approaches suffer from several disadvantages. For example, these approaches tend to utilize or waste much power. This waste of power reduces the battery life of such systems. In other examples, the system may suffer from performance issues. These and other disadvantages have resulted in some user dissatisfaction with these previous approaches.

BRIEF DESCRIPTION OF THE DRAWINGS

[0006] For a more complete understanding of the disclosure, reference should be made to the following detailed description and accompanying drawings wherein:

[0007] FIG. 1 comprises a block diagram of a system or apparatus using an Acoustic Activity Detection (AAD) module according to various embodiments of the present invention;

[0008] FIG. 2 comprises a flow chart of an Acoustic Activity Detection (AAD) module according to various embodiments of the present invention;

[0009] FIG. 3 comprises a flow chart of another example of an Acoustic Activity Detection (AAD) module according to various embodiments of the present invention;

[0010] FIG. 4 comprises a graph showing one example of operation of the approaches described herein according to various embodiments of the present invention.

[0011] 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

[0012] Approaches are provided for a digital microphone with the built-in capability to distinguish speech and speech-like acoustic activity signals from background noise, to trigger a following digital signal processor (DSP) (or other) module or system, and to provide continuous audio data to the system for detection of a “voice trigger” followed by seamless operation of speech recognition engines. The ability to distinguish speech activity from background allows the following modules in the signal chain to operate in low power “sleep” states, conserving battery power until their operation is required.

[0013] To enable such capabilities, a low complexity Acoustic Activity Detection (AAD) module is configured to detect speech with low latency and high probability in high ambient noise. The high noise results in speech signal-to-noise ratios (SNRs) as low as approximately 0 dB. In addition, the present approaches provide a built-in buffer for seamless handover of audio data to the following “voice trigger” detector as well as for the general purpose automatic speech recognizer (ASR) following the voice trigger. The AAD module may be implemented as any combination of computer hardware and software. In one aspect, it may be implemented as computer instructions executed on a processing device such as an application specific integrated circuit (ASIC) or microprocessor. As described herein, the AAD module transforms input signals into control signals that are used to indicate speech detection.

[0014] Lower power can be achieved by optimized approaches which can detect speech audio activity with low latency, low computational cost (silicon area), high detection rates and low rate of false triggers. In some aspects, the present approaches utilize a buffer of sufficient depth to store the recorded data for the multi-layered recognition system to work seamlessly and without pauses in speech.

[0015] In many of these embodiments, streaming audio is received and the streaming audio comprising a sequence of frames, each having a plurality of samples. An energy estimate is obtained for the plurality of samples. The energy estimate is compared to at least one threshold. Based upon the comparing, a determination is made as to whether speech is detected.

[0016] In other aspects, a determination is made as to whether a speech hangover has occurred. In some examples, a non-linear process is used to make the hangover determination. In other examples and when speech is not detected, a determination is made as to the noise level of the plurality of samples.

[0017] In others of these embodiments, streaming audio is received, and the streaming audio comprises a sequence of frames, each with a plurality of samples. A first energy estimate is obtained for the frame of the plurality of samples and a second energy estimate is obtained for the band passed signal from the same frame of the plurality of samples. In a first path, the first energy estimate is compared to at least one first threshold and based upon the comparing, a determination is made as to whether speech is detected. In a second path that is performed in

parallel with the first path, the second energy estimate is compared to at least one second threshold and based upon the comparing, a determination is made as to whether speech is detected.

[0018] In other aspects, a determination is made as to whether a speech hangover has occurred. In some examples, a non-linear process is used to make the hangover determination. In other examples and when speech is not detected, a determination is made as to the noise level of the plurality of samples.

[0019] In others of these embodiments, an apparatus configured to distinguish speech activity from background noise includes an analog sub-system, a conversion module, and a digital sub-system. The analog sub-system converts sound energy into an analog electrical signal. The conversion module is coupled to the analog system and converts the analog signal into a digital signal.

[0020] The digital sub-system is coupled to the conversion module, and includes an acoustic activity detection (AAD) module. The AAD module is configured to receive the digital signal. The digital signal comprises a sequence of frames, each having a plurality of samples. The AAD module is configured to obtain an energy estimate for the plurality of samples and compare the energy estimate to at least one threshold. The AAD module is configured to based upon the comparison, determine whether speech is detected, and when speech is detected transmit an interrupt to a voice trigger module.

[0021] In other aspects, the analog sub-system includes a micro-electro-mechanical system (MEMS) transducer element. In other examples, the AAD module is further configured to determine whether a speech hangover has occurred. In yet other aspects, the AAD module enables the transmission of the digital signal by a transmitter module upon the detection of speech.

[0022] In other examples, the conversion module comprises a sigma-delta modulator that is configured to convert the analog signal into a single bit stream pulse density modulated (PDM) format. In some examples, the digital subsystem comprises a decimator module that converts the single bit stream pulse density modulated (PDM) format into a pulse code modulated (PCM) format. In other approaches, the pulse code modulated (PCM) audio from the decimator module

is stored continuously in a circular buffer and in parallel, also provided to the AAD module for processing.

[0023] In some examples, the AAD module enables the transmission of the digital signal by a transmitter module upon the detection of speech. The transmitter module comprises a interpolator and digital sigma-delta modulator, that converts the pulse code modulated (PCM) format back to a single bit stream pulse density modulated (PDM) format.

[0024] Referring now to FIG. 1, one example of an apparatus or system that is configured to distinguish speech activity from background noise, to trigger a following DSP (or other) system, and to provide continuous audio data to the system for detection of a “voice trigger” followed by seamless operation of speech recognition engines is described. A sub-system assembly 102 includes an analog subsystem 104 and a digital subsystem 106. The analog subsystem includes a microphone 108, a preamplifier (preamp) 110, and a Sigma-Delta ($\Sigma\Delta$) modulator 112. The digital subsystem includes a decimator 114, a circular buffer 116, an Acoustic Activity Detection (AAD) module 118, and a transmitter 120. A voice trigger module 122 couples to a higher level ASR main applications processor 124. It will be appreciated that the AAD module 118, voice trigger module 122, and high-level ASR main applications processor module 124 may be implemented as any combination of computer hardware and software. For example, any of these elements may be implemented as executable computer instructions that are executed on any type of processing device such as an ASIC or microprocessor.

[0025] The microphone device 108 may be any type of microphone that converts sound energy into electrical signals. It may include a diaphragm and a back plate, for example. It also may be a microelectromechanical system (MEMS) device. The function of the preamplifier 110 is to provide impedance matching for the actual transducer and sufficient drive capability for the microphone analog output.

[0026] The sub-system 102 is always listening for acoustic activity. These system components are in typically in a low power mode to conserve battery life. The analog sub-system 104 consists of the microphone 108 and the pre-amplifier 110. The pre-amplifier 110 feeds into the Sigma-Delta ($\Sigma\Delta$) modulator 112, which converts analog data to a single bit-stream pulse

density modulated (PDM) format. Further, the output of the Sigma-Delta modulator 112 feeds into the decimator 114, which converts the PDM audio to pulse code modulated (PCM) waveform data at a particular sampling rate and bit width. The data is stored via optimal compressive means in the circular buffer 116 of a desired length to allow seamless non-interrupted audio to the processing blocks upstream. The compressed PCM audio data is further reconverted to PDM for the upstream processing block, when that data is required. This is controlled by the transmitter (Tx) 120. The data transmission occurs after the acoustic activity is detected.

[0027] When the AAD module 118 detects speech like acoustic activity (by examining a frame of data samples or some other predetermined data element), it sends an interrupt signal 119 to the voice trigger module 122 to wake-up this module and a control signal to the transmitter (Tx) 120 to output audio data. Once the voice trigger module is operational, it runs an algorithm to detect the voice trigger. If a voice trigger is detected, then the higher level ASR main applications processor module 124 is brought out of sleep mode with an interrupt 121 from the voice trigger block 122 as shown. If the AAD module 118 triggers the transmit data using Tx control 117 and consequently detects frames with non-speech data for a pre-set amount of relatively long time, it can turn off the transmitter 120 to signal the voice trigger module 122 to go back to sleep mode to reduce power consumption.

[0028] It will be appreciated that there exist latencies associated with a multilayered speech recognition system as described above. These include a latency for acoustic activity detection, a delay for wake-up of the voice-trigger module 122, a latency for “voice trigger” detection, and delay for wake-up of high-level ASR main applications processor module 124.

[0029] There may also be a need for priming the various processing blocks with audio data. The voice trigger module 122 requires audio data from before the acoustic activity detection trigger, i.e. before this block is woken-up from its sleep mode. The high-level ASR main applications processor module 124 requires audio data from before it is brought out of sleep mode. The requirements for audio data before the actual speech onset by both, the voice trigger module 122 and the high-level ASR main applications processor module 124 as well as the latencies of the AAD module 118 (and the “voice trigger” algorithm it implements), requires the use of the buffer 116, which has sufficient depth to allow recognition of speech that follows

the “voice trigger” phrase in a seamless manner (without artificial pauses in speech). This is implemented in the circular buffer 116 as shown and described herein.

[0030] One advantage of the present approaches is that they do not miss speech onset and have a very low latency for detection. This leads to better performance of the “voice trigger” while reducing the buffer depth size as much as possible. Another goal is to reduce false detects of speech to avoid turning on the “voice trigger” algorithm unnecessarily and thus reducing battery life. In one aspect, the AAD module 118 provides (or helps provide) these results.

[0031] The operation of the AAD module 118 is based on frame-by-frame decision making by comparing an energy estimate of the audio signal samples in a frame of audio data to various thresholds for speech and noise. In these regards, a fast time constant based energy measure is estimated. The energy measure is designed to track the speech envelope with low latency. Thresholds for noise and speech levels are computed using energy statistics. The module 118 calculates the speech onset threshold by determining when the energy estimator exceeds a threshold.

[0032] Additionally, a band pass module of a similar structure may be introduced to capture fast energy variations occurring in the 2 kHz – 5 kHz frequency range. This band pass module improves the detection levels of speech starts with non-vocal fricatives and sibilants. The use of this additional feature is described below with respect to FIG. 3.

[0033] Referring now to FIG. 2, the acoustic activity detector (AAD) algorithm (e.g., block 118 of FIG. 1) is described. It will be appreciated that these approaches may be implemented as any combination of computer hardware and software. For example, any of these elements or functions may be implemented as executable computer instructions that are executed on any type of processing device such as an ASIC or microprocessor.

[0034] At step 202, streaming audio is received. At step 204, energy estimation may be performed. These signals are in the form of fixed point digital PCM representation of the audio signal. In one example, a leaky integrator or a single pole filter is used to estimate the energy of the signal in a sample by sample basis. This may be based on absolute value of the signal sample. The following equation may be used:

[0035]
$$e_{st}(n) = (1 - \alpha) \times e_{st}(n-1) + \alpha \times |x(n)|$$

Alternatively, a squared value may be used:

$$e_{st}(n) = (1 - \alpha) \times e_{st}(n-1) + \alpha \times (x(n))^2$$

[0036] At step 206, an energy estimate $E_N(k)$ is then made for an entire frame of N samples as shown below.

[0037]
$$E_N(k) = \sum_n e_{st}(n)$$

[0038] Other examples are possible.

[0039] At steps 208, 210, and 212 the frame energy $E_N(k)$, is then compared to various thresholds (as shown in FIG. 2) in a decision tree to arrive at a conclusion of whether speech is detected ($v(k) = 1$) or not ($v(k) = 0$). A hangover logic block is shown at step 218, which uses a non-linear process to determine how long the speech detection flag should be held high after immediately the detection of an apparent non-speech frame. The hangover logic block 218 helps connect discontinuous speech segments caused by pauses in the natural structure of speech by means of the hangover flag $\mathcal{H}(k)$. It also captures the tail ends of words that occur with lower energy and may fall below the detection threshold. An example implementation of the hangover flag is shown in the following equation, though there are other methods that may be derived from this approach or that are similar.

$$\mathcal{H}(k) = \begin{cases} 1, & v(k) = 0 \text{ and } \sum_{m=N}^m v(i) = M+1 \quad \forall m \in \{k-N, \dots, k-1\} \\ 0, & \text{otherwise} \end{cases}$$

[0040]

[0041] At step 220 it is determined if there is a speech hangover. If the answer is affirmative, then execution continues at step 222. If the answer is negative, noise is detected at step 224 and step 214 is executed as described above.

[0042] In the event that a frame is declared to be non-speech, it is inserted into a First-in First-out (FIFO) buffer for estimation of the average noise level μ_N , and the standard deviation of the noise level σ_N . In the following examples, let the contents of the FIFO buffer be terms the energies of the last M frames declared as non-speech $E_N(b)$. Then:

$$[0043] \quad \mu_N(k) = \sum_b^M E_N(b)$$

[0044] Similarly the standard deviation may be estimated as:

$$[0045] \quad \sigma_N(k) = \sqrt{\sum_b [E_N(b) - \mu_N(k)]^2}$$

[0046] For computational and hardware efficiency, the frame energy is used in a leaky integrator process to estimate the mean of the noise μ_N as follows:

$$[0047] \quad \mu_N(k) = (1 - \beta) \times \mu_N(k-1) + \beta \times E_N(k)$$

[0048] Similarly, the standard deviation σ_N may be estimated as follows:

$$[0049] \quad \sigma_N(k) = (1 - \gamma) \times \sigma_N(k-1) + \gamma \times |E_N(k) - \mu_N(k)|$$

[0050] The T_s and T_n thresholds are adaptively updated at steps 214 and 216 based on noise statistic from frames in the noise buffer. An estimator for the noise statistics used for calculating thresholds is shown below based on the mean μ_N and the standard deviation σ_N of the frame noise levels estimated.

$$[0051] \quad T_n(k) = c \times \mu_N + b \times \min(\sigma_N, d)$$

[0052] The parameters “a”, “b”, “c”, “d”, and “e” are determined empirically to give the desired performance. The minimum function establishes a maximum value for the update.

[0053] The speech threshold is derived as:

$$[0054] \quad T_s(k) = \min(a \times T_n(k), T_n(k) + C) + e$$

[0055] Here the minimum function avoids excessive range of the speech threshold.

[0056] If any of the tests performed at steps 208, 210, and 212 are not met, then hangover logic 218 is utilized as described above. If the answers to any of these steps are affirmative, then at step 222 speech is detected and execution ends. Since speech has been detected, an indication that speech has been detected as well as the speech itself can be sent to appropriate circuitry for further processing.

[0057] Referring now to FIG. 3, another implementation may be derived based on similar principles, with the aim of lower latency detection of non-vocal fricatives and sibilants. These sounds are characterized by relatively low energy frequency distributions which are more akin to short bursts of band-passed noise. As with the example of FIG. 2, it will be appreciated that these approaches may be implemented as any combination of computer hardware and software. For example, any of these elements or functions may be implemented as executable computer instructions that are executed on any type of processing device such as an ASIC or microprocessor.

[0058] The principle of the algorithm is similar to that shown in the algorithm of FIG. 2 except that a parallel path for energy estimation and detection logic is established using high frequency speech energy in the band from, for example, approximately 2 kHz – 5 kHz to capture fricative and sibilant characteristics. Several efficient methods exist to implement the band-pass filter as required. The outputs of the threshold based detection drives the hangover logic block to determine speech or non-speech frames. Equations similar to those used in FIG. 2 may also be used.

[0059] More specifically, at step 302 an audio signal is received. At step 304, energy estimation may be performed. At step 306, an energy estimate is then made for an entire frame of N samples.

[0060] At step 308, the energy is then compared to various thresholds to arrive at a conclusion of whether speech is detected or not. Execution continues at step 310.

[0061] In parallel, at step 322, energy estimation may be performed for band pass frequencies (e.g., approximately 2 kHz – 5 kHz). At step 324, an estimate of band pass energy is made. At step 326, an energy estimate is then made for an entire frame of N samples. At step

328, the energy is then compared to various thresholds to arrive at a conclusion of whether speech is detected or not. Execution continues at step 310.

[0062] A hangover logic block is shown at step 310, which uses a non-linear process to determine how long the speech detection flag should be held high after immediately after the detection of an apparent non-speech frame. The hangover logic block 310 helps connect discontinuous speech segments caused by pauses in the natural structure of speech.

[0063] At step 314 it is determined if there is a speech hangover. If the answer is affirmative, then execution continues at step 312 (speech is detected). If the answer is negative, noise is detected at step 316. At step 318 FIFO buffer stores the noise. At step 320, the thresholds are updated.

[0064] The detection results for the algorithm shows good performance on a database of several hours of speech with changing ambient noise. The SNR varies from about 20 dB to 0 dB. The database has sentences in noise separated by periods of background noise only.

[0065] For purpose of the application, the detection of the onset of speech is accomplished with the lowest possible latency. Speech onset is not missed. For low power requirements, false triggers are minimized. With these goals in mind, results of implementing and executing the algorithm were evaluated according to two principles: low latency detection of the first spoken word in every segment of speech, and the actual accuracy of the detection over all the speech sections in the database. The measures used are standard measures in Hypothesis Testing as defined below:

		actual value		
		p	n	total
prediction outcome	p'	True Positive	False Positive	P'
	n'	False Negative	True Negative	N'
total		P	N	

A. True positive rate $TPR = \frac{TP}{P}$

B. False positive rate $FPR = \frac{FP}{N}$

$$\text{C. Accuracy } ACC = \frac{TP + TN}{P + N}$$

[0066] One important parameter for the AAD algorithm is the true positive rate of detection of the first occurrence of speech in a segment. This statistic is in part used to determine how quickly the detector responds to speech. FIG 4. shows that speech is always detected by the sixth frame for several suitable set of parameters.

[0067] Table 1 shows overall accuracy rates along with true positive rate and false positive rate. The low false positive rate indicates that the algorithm is power efficient.

	TPR	FPR	ACC
Performance	96%	21%	82%

Table 1: Overall performance of one example speech AAD detector approach

[0068] Preferred embodiments of this invention are described herein, including the best mode known to the inventors for carrying out the invention. It should be understood that the illustrated embodiments are exemplary only, and should not be taken as limiting the scope of the invention.

WHAT IS CLAIMED IS:

1. A method of detecting human speech, the method comprising:
receiving streaming audio, the streaming audio comprising a sequence of frames, each having a plurality of samples;
obtaining an energy estimate for the plurality of samples;
comparing the energy estimate to at least one threshold;
based upon the comparing, determining whether speech is detected.
2. The method of claim 1, further comprising determining whether a speech hangover has occurred.
3. The method of claim 2, wherein determining whether a speech hangover occur utilizes a non-linear process.
4. The method of claim 1, wherein when speech is not detected, determining a noise level of the frame with the plurality of samples.
5. A method, the method comprising:
receiving streaming audio, the streaming audio comprising a sequence of frames, each with a plurality of samples;
obtaining a first energy estimate for the frame of the plurality of samples and a second energy estimate for the band passed signal from the same frame of the plurality of samples;
in a first path, comparing the first energy estimate to at least one first threshold and based upon the comparing, determining whether speech is detected;
in a second path performed in parallel with the first path, comparing the second energy estimate to at least one second threshold and based upon the comparing, determining whether speech is detected.

6. The method of claim 5, further comprising determining whether a speech hangover has occurred.

7. The method of claim 6, wherein determining whether a speech hangover occurred utilizes a non-linear process.

8. The method of claim 5, wherein when speech is not detected, determining a noise level of the plurality of samples.

9. An apparatus configured to distinguish speech activity from background noise, the apparatus comprising:

an analog sub-system that converts sound energy into an analog electrical signal;

a conversion module coupled to the analog system that converts the analog signal into a digital signal;

a digital sub-system coupled to the conversion module, the digital sub-system including an acoustic activity detection (AAD) module, the AAD module configured to receive the digital signal, the digital signal comprising a sequence of frames, each having a plurality of samples, the AAD module configured to obtain an energy estimate for the plurality of samples and compare the energy estimate to at least one threshold, and the AAD module configured to based upon the comparison, determine whether speech is detected, and when speech is detected transmit an interrupt to a voice trigger module.

10. The apparatus of claim 9, wherein the analog sub-system includes a micro-electro-mechanical system (MEMS) transducer element.

11. The apparatus of claim 9, wherein the AAD module is further configured to determine whether a speech hangover has occurred.

12. The apparatus of claim 9, wherein the AAD module enables the transmission of the digital signal by a transmitter module upon the detection of speech.

13. The apparatus of claim 9, wherein the conversion module comprises a sigma-delta modulator that is configured to convert the analog signal into a single bit stream pulse density modulated (PDM) format.

14. The apparatus of claim 13, wherein the digital subsystem comprises a decimator module that converts the single bit stream pulse density modulated (PDM) format into a pulse code modulated (PCM) format.

15. The apparatus of claim 14, wherein the pulse code modulated (PCM) audio from the decimator module is stored continuously in a circular buffer and in parallel, also provided to the AAD module for processing.

16. The apparatus of claim 14, wherein the AAD module enables the transmission of the digital signal by a transmitter module upon the detection of speech, and wherein the transmitter module comprises an interpolator and digital sigma-delta modulator, that converts the pulse code modulated (PCM) format back to a single bit stream pulse density modulated (PDM) format.

1/4

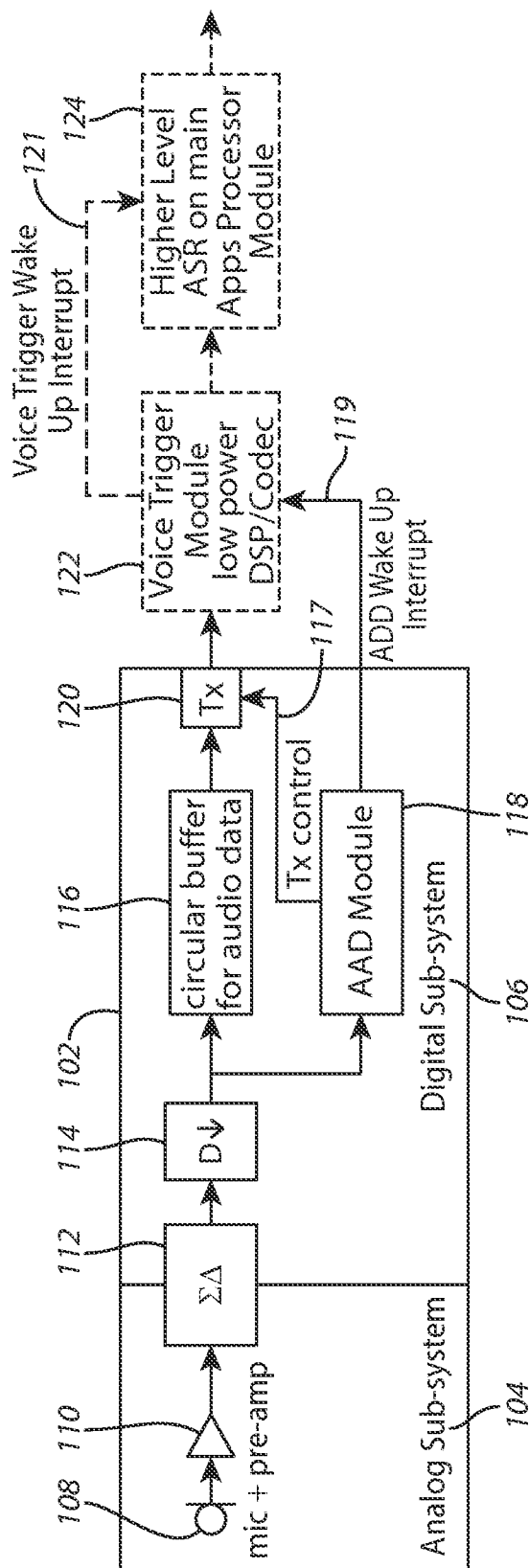


FIG. 1

2/4

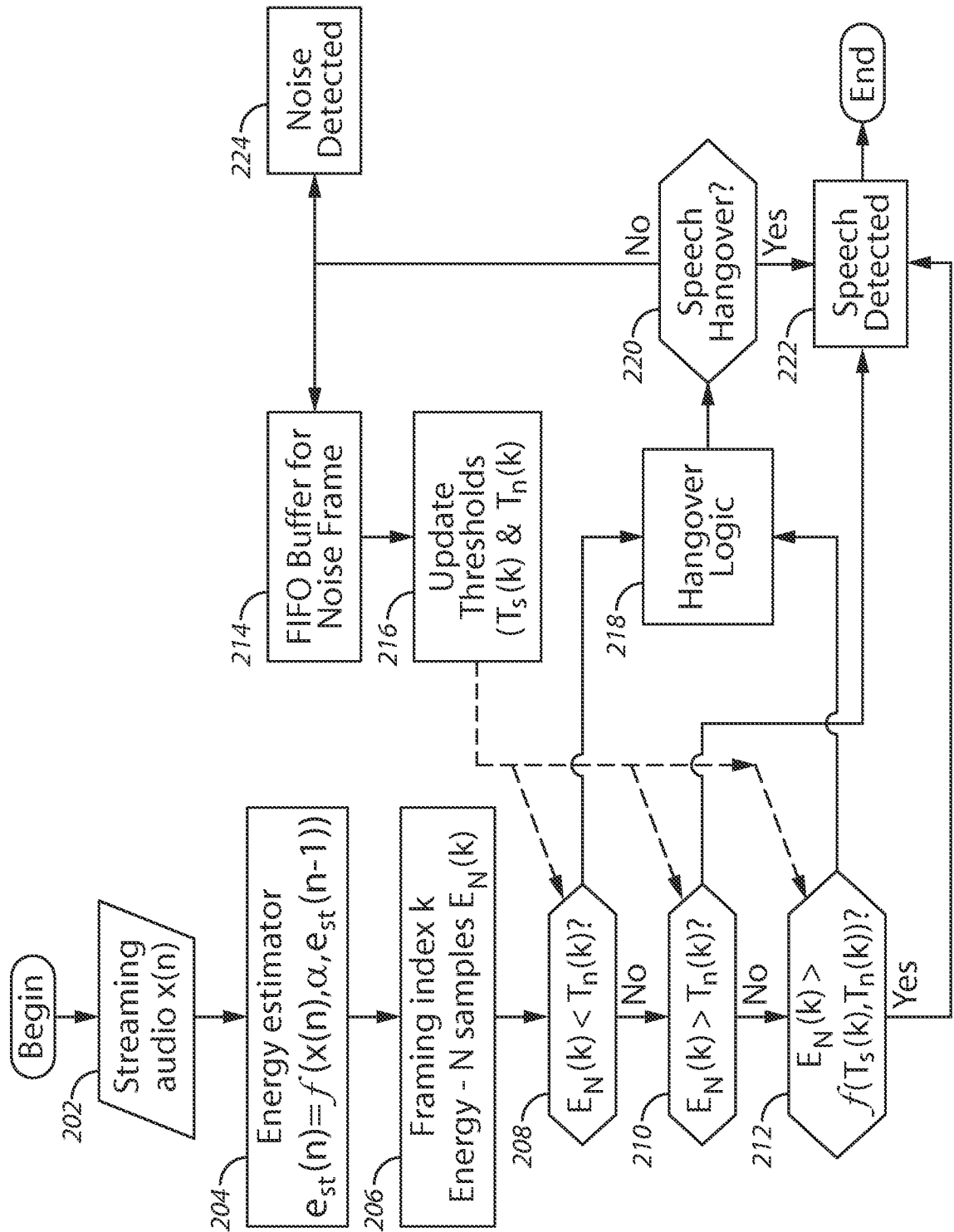
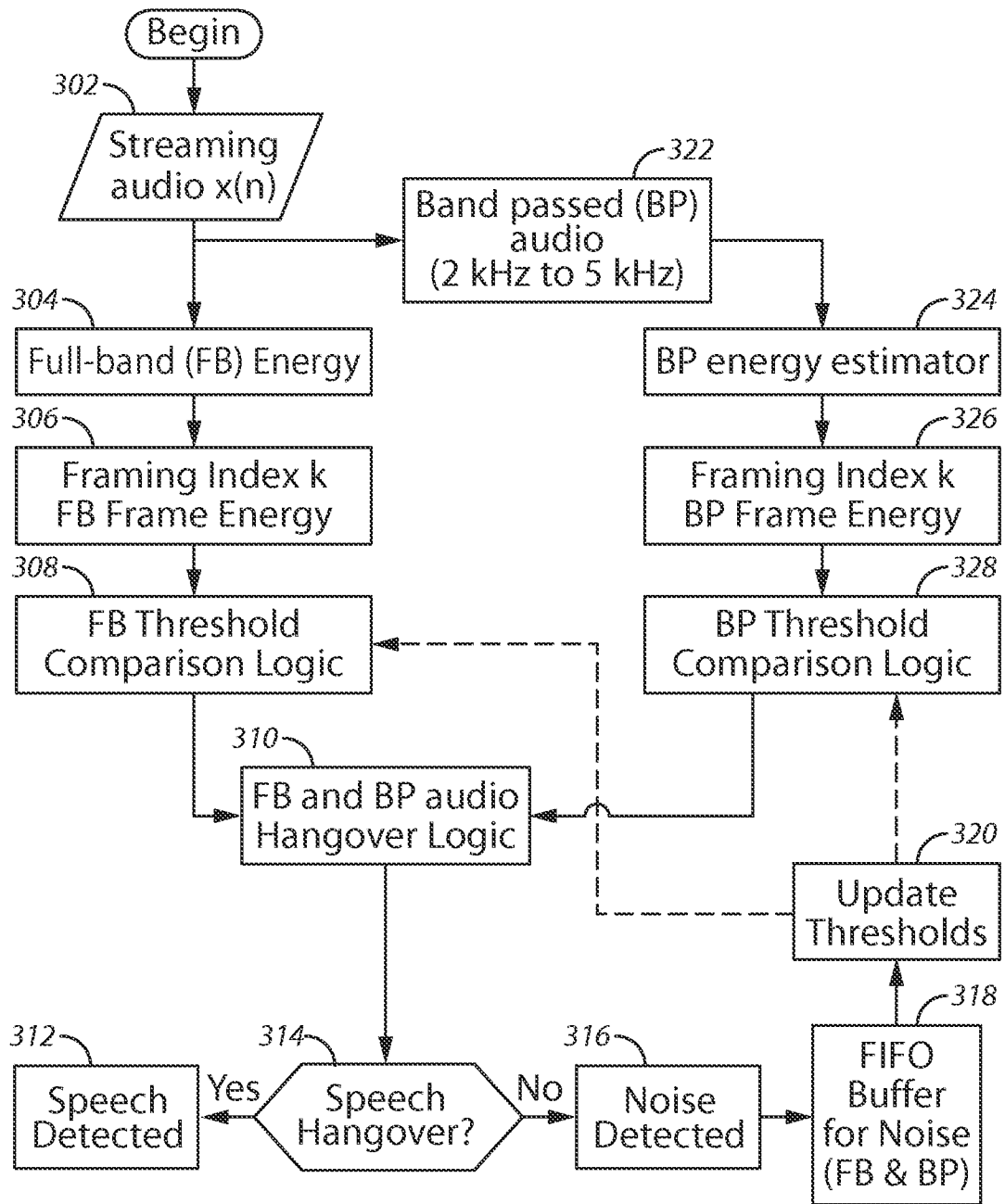
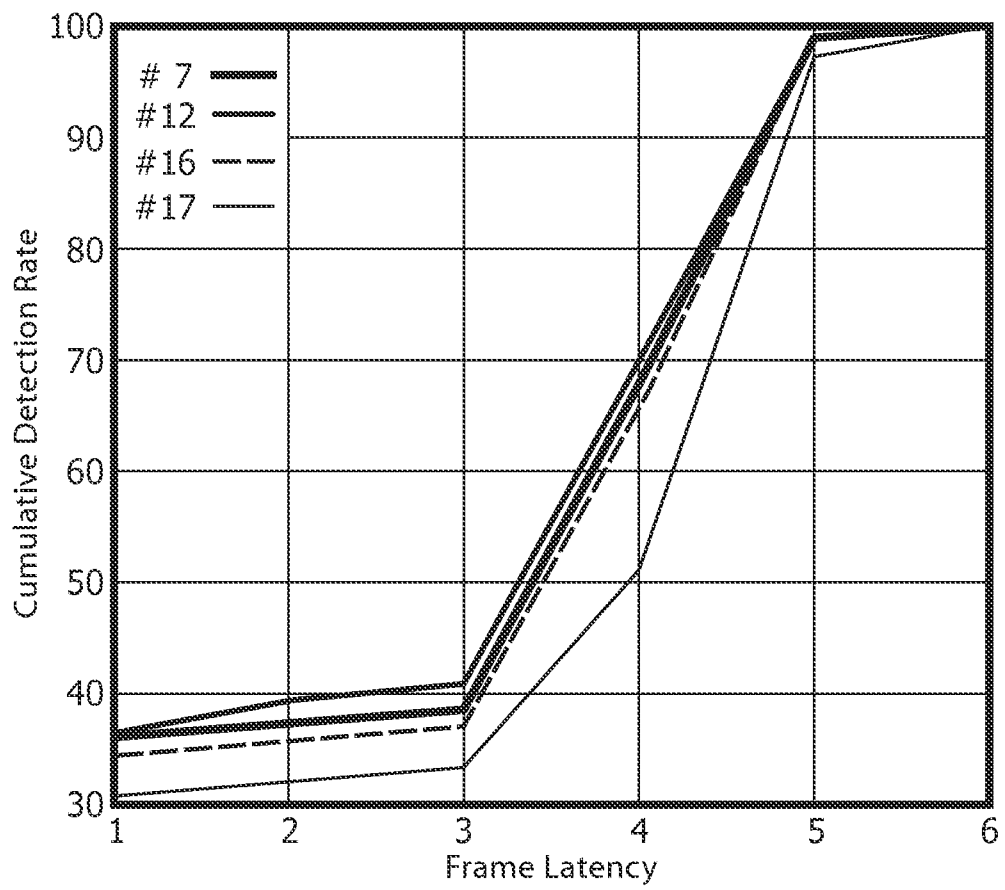


FIG. 2

3/4

**FIG. 3**

4/4

**FIG. 4**

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2014/060567**A. CLASSIFICATION OF SUBJECT MATTER****G10L 25/78(2013.01)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 25/78; G10L 11/00; G10L 15/20; G10L 5/00; H04R 3/00; G10L 21/00

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:voice trigger, acoustic activity detection, threshold, band pass, 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 5983186 A (YASUNAGA MIYAZAWA et al.) 09 November 1999 See column 3, line 19 - column 10, line 18; and figure 4.	1-4, 9-12
Y		13-16
A		5-8
Y	WO 2013-085499 A1 (INTEL CORPORATION) 13 June 2013 See page 2, line 26 - page 3, line 19; and figure 1.	13-16
A	US 2011-0106533 A1 (RONGSHAN YU) 05 May 2011 See paragraphs 34-53; and figures 3-4.	1-16
A	US 2003-0144844 A1 (ANTONIO COLMENAREZ et al.) 31 July 2003 See paragraphs 15-27; and figure 2.	1-16
A	US 2010-0128894 A1 (NICOLAS PETIT et al.) 27 May 2010 See paragraphs 158-165; and figure 28.	1-16



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

16 January 2015 (16.01.2015)

Date of mailing of the international search report

16 January 2015 (16.01.2015)

Name and mailing address of the ISA/KR

International Application Division
Korean Intellectual Property Office
189 Cheongsu-ro, Seo-gu, Daejeon Metropolitan City, 302-701,
Republic of Korea

Facsimile No. ++82 42 472 3473

Authorized officer

YU, Jae Chon

Telephone No. +82-42-481-8647



INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US2014/060567

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 05983186 A	09/11/1999	JP 03674990 B2 JP 09-062293 A	27/07/2005 07/03/1997
WO 2013-085499 A1	13/06/2013	CN 103959376 A EP 2788979 A1 TW 201342362 A US 2014-0236582 A1	30/07/2014 15/10/2014 16/10/2013 21/08/2014
US 2011-0106533 A1	05/05/2011	CN 102077274 A CN 102077274 B CN 103137139 A EP 2297727 A2 US 8554556 B2 WO 2010-002676 A2 WO 2010-002676 A3	25/05/2011 21/08/2013 05/06/2013 23/03/2011 08/10/2013 07/01/2010 25/02/2010
US 2003-0144844 A1	31/07/2003	AT 421136 T AU 2003-205929 A1 CN 1291372 C0 CN 1623182 A DE 60325826 D1 EP 1472679 A1 EP 1472679 B1 JP 04681810B2 JP 2005-516263 A US 7219062 B2 WO 03-065350 A1	15/01/2009 02/09/2003 20/12/2006 01/06/2005 05/03/2009 03/11/2004 14/01/2009 11/05/2011 02/06/2005 15/05/2007 07/08/2003
US 2010-0128894 A1	27/05/2010	AU 2009-308442 A1 CA 2741652 A1 CA 2798282 A1 CA 2798512 A1 CN 102282865 A EP 2353302 A1 EP 2567377 A1 EP 2567553 A1 US 2002-0198705 A1 US 2007-0233479 A1 US 2009-0003623 A1 US 2009-0022350 A1 US 2010-0128881 A1 US 2010-0278352 A1 US 2010-0280824 A1 US 2011-0026722 A1 US 2011-0051950 A1 US 2011-0051951 A1 US 7246058 B2 US 8130984 B2 US 8321213 B2	29/04/2010 29/04/2010 10/11/2011 10/11/2011 14/12/2011 10/08/2011 13/03/2013 13/03/2013 26/12/2002 04/10/2007 01/01/2009 22/01/2009 27/05/2010 04/11/2010 04/11/2010 03/02/2011 03/03/2011 03/03/2011 17/07/2007 06/03/2012 27/11/2012

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US2014/060567Patent document
cited in search reportPublication
datePatent family
member(s)Publication
date

US 8326611 B2	04/12/2012
US 8452023 B2	28/05/2013
US 8488803 B2	16/07/2013
US 8503686 B2	06/08/2013
US 8503691 B2	06/08/2013
US 8699721 B2	15/04/2014
US 8731211 B2	20/05/2014
WO 2009-003180 A1	31/12/2008
WO 2010-048635 A1	29/04/2010
WO 2011-002823 A1	06/01/2011
WO 2011-140096 A1	10/11/2011
WO 2011-140110 A1	10/11/2011