LOW-POWER, ALWAYS-LISTENING, VOICE-COMMAND DETECTION AND CAPTURE

FIELD OF THE INVENTION

[0001] This disclosure is directed to a system and methods for voice detection and capture of voice commands.

BACKGROUND

[0002] Many consumer electronic devices have functions that are controlled in whole or in part by voice commands issued by a human voice. This voice control allows users to operate the devices without the use of the user's hands and with fewer buttons or switches on the device.

[0003] Such devices may require an always-listening mode so that the device may "hear" and capture the voice commands. Thus, in the always listening mode, the device is on and waits for a voice command. The quality of an always-listening function is generally measured by two metrics: the overall average current consumption and the voice-command misdetection rate.

[0004] Embodiments of the invention address shortcomings in existing devices.

SUMMARY OF THE DISCLOSURE

[0005] Embodiments of the disclosed subject matter are directed to a low-power, always-listening, voice-command detection and capture system. By utilizing a lower-current operational state, the system can reduce the overall average current consumption of the system. In the lowest-current operational state, the system waits for signs of voice activity. Each ensuing state has a gradually higher confidence of voice command detection, until the highest-current operational state where the voice command is detected, recognized, and processed.

[0005a] In an aspect, the present disclosure refers to a system for detecting and capturing voice commands, the system comprising: an input configured to receive an audio signal; an analog to digital converter configured to convert the audio signal to a digital-audio signal, the analog to digital converter including a wakeup circuit configured to determine whether the audio signal has an onset level of audio activity, to place the analog to digital converter in a sleep mode in which a converter portion of the analog to digital converter does not generate the digital-audio signal when the audio signal does not have the onset level of audio activity, and to place the analog to digital converter in an active mode in which the converter portion of the analog to digital converter generates the digital-audio signal when the audio signal has the

onset level of audio activity; an amplitude detector configured to determine the amplitude of the digital-audio signal; a controller configured to compare the amplitude of the digital-audio signal to a first threshold and to a second threshold and withhold an interrupt signal when the amplitude of the digital-audio signal does not exceed the first threshold or the second threshold, generate the interrupt signal when the amplitude of the digital-audio signal exceeds the first threshold and the second threshold; and trigger spectral analysis of the digital-audio signal when the amplitude of the digital-audio signal is between the first threshold and the second threshold.

[0005b] In another aspect, the present disclosure refers to a method of detecting and capturing voice commands comprising: receiving, by an analog-to-digital converter, an audio signal; determining, by a wakeup circuit, whether the audio signal has an onset level of audio activity; when the audio signal has the onset level of audio activity, placing the analog-to-digital converter in an active mode in which a converter portion of the analog-to-digital converter generates a digital-audio signal from the audio signal at an analog-to-digital converter sampling rate; when the audio signal does not have the onset level of audio activity, placing the analog-to-digital converter in a sleep mode in which the converter portion of the analog-to-digital converter does not generate the digital-audio signal; receiving, by a voice-activity detector, a digital-audio signal; determining, by the voice-activity detector, the amplitude of the digital-audio signal; comparing, by the voice-activity detector, the amplitude of the digital-audio signal to a first threshold and to a second threshold; withholding, by the voice-activity detector, an interrupt signal when the amplitude of the digital-audio signal does not exceed the first threshold or the second threshold; generating, by the voice-activity detector, the interrupt signal when the amplitude of the digital-audio signal exceeds the first threshold and the second threshold; and performing, by the voice-activity detector, spectral analysis of the digital-audio signal when the amplitude of the digital-audio signal is between the first threshold and the second threshold.

[0006] Accordingly, at least some examples of a system for detecting and capturing voice commands may include a voice-activity detector (VAD) configured to receive a VAD-received digital-audio signal; determine the amplitude of the VAD-received digital-audio signal; compare the amplitude of the VAD-received digital-audio signal to a first threshold and to a second threshold; withhold a VAD interrupt signal when the amplitude of the VAD-received digital-audio signal does not exceed the first threshold or the second threshold; generate the VAD interrupt signal when the amplitude of the VAD-received digital-audio signal exceeds the first threshold and the second threshold; and perform spectral analysis of the VAD-received

digital-audio signal when the amplitude of the VAD-received digital-audio signal is between the first threshold and the second threshold.

[0007] In another example, at least some examples of a method of detecting and capturing voice commands may include receiving, by a voice-activity detector (VAD), a VAD-received digital-audio signal; determining, by the VAD, the amplitude of the VAD-received digital-audio signal; comparing, by the VAD, the amplitude of the VAD-received digital-audio signal to a first threshold and to a second threshold; withholding, by the VAD, a VAD interrupt signal when the amplitude of the VAD-received digital-audio signal does not exceed the first threshold or the second threshold; generating, by the VAD, the VAD interrupt signal when the amplitude of the VAD-received digital-audio signal exceeds the first threshold and the second threshold; and performing, by the VAD, spectral analysis of the VAD-received digital-audio signal when the amplitude of the VAD-received digital-audio signal is between the first threshold and the second threshold.

[0008] Accordingly, when compared to existing devices, the described systems and methods may provide a lower voice-command misdetection rate at a lower average current consumption.

BRIEF DESCRIPTION OF THE DRAWINGS

[0009] Fig. 1 is a functional block diagram of a voice-command detection and capture system, according to embodiments of the invention.

[0010] Fig. 2 is a circuit diagram of a wakeup circuit that may be included in the voice-recognition module shown in Fig. 1.

[0011] Fig. 3 shows an example graph of audio signals from human speech, where the vertical axis represents signal amplitude and the horizontal axis represents time.

[0012] Fig. 4 is a flowchart showing an example process of the voice-recognition module shown in Fig. 1.

[0013] Fig. 5 shows an example of a clock switching event according to a feature of a voice-command detection and capture system.

[0014] In the Detailed Description that follows, reference numbers of the pattern 1xx are generally shown in FIGs. 1 and 2, while 3xx reference numbers are generally shown in FIG. 3. Likewise, 4xx reference numbers are generally shown in FIG. 4, and 5xx reference numbers are generally shown in FIG. 5.

DETAILED DESCRIPTION

[0015] As described in this application, embodiments of the invention are directed to a low-power, always-listening, voice-command detection and capture system. The disclosed

CLAIMS:

1. A system for detecting and capturing voice commands, the system comprising:

an input configured to receive an audio signal;

an analog to digital converter configured to convert the audio signal to a digital-audio signal, the analog to digital converter including a wakeup circuit configured to determine whether the audio signal has an onset level of audio activity, to place the analog to digital converter in a sleep mode in which a converter portion of the analog to digital converter does not generate the digital-audio signal when the audio signal does not have the onset level of audio activity, and to place the analog to digital converter in an active mode in which the converter portion of the analog to digital converter generates the digital-audio signal when the audio signal has the onset level of audio activity;

an amplitude detector configured to determine anamplitude of the digital-audio signal; and

a controller configured to compare the amplitude of the digital-audio signal to a first threshold and to a second threshold and withhold an interrupt signal when the amplitude of the digital-audio signal does not exceed the first threshold or the second threshold, generate the interrupt signal when the amplitude of the digital-audio signal exceeds the first threshold and the second threshold; and trigger spectral analysis of the digital-audio signal when the amplitude of the digital-audio signal is between the first threshold and the second threshold.

2. The system of claim 1, in which, to determine whether the audio signal has the onset level of audio activity, the wakeup circuit is configured to compare the audio signal to a wakeup-onset threshold.

3. The system of claim 1, in which the wakeup circuit comprises:

an audio preamplifier gain stage configured to boost the audio signal; and

a comparator stage configured to compare the boosted audio signal to a wakeup-onset threshold indicating the onset level of audio activity.

4. The system of claim 1, in which the converter portion of the analog to digital converter is unpowered when the analog to digital converter is in the sleep mode, and in which the converter portion of the analog to digital converter is powered when the analog to digital converter is in the active mode.

5. The system of claim 1, further comprising a decimation filter between the analog to digital converter and the amplitude detector, the decimation filter configured to receive the digital-audio signal and process the digital-audio signal to generate reduced digital-audio samples having a reduced sampling rate, the reduced sampling rate being lower than a sampling rate of the analog to digital converter.

6. The system of claim 1, further comprising:

a sample combine block configured to receive as an input the digital-audio signal having a sampling rate, the digital-audio signal comprising a number of digital-audio samples, and to process the digital-audio signal by combining one or more of the digital-audio samples to generate a combined digital-audio signal having an amplitude, the combined digital-audio signal having a sampling rate lower than the sampling rate of the digital-audio signal;

an amplitude voice-activity detector block configured to determine the amplitude of the combined digital-audio signal;

a spectral analysis block configured to perform the spectral analysis of the digital-audio signal; and

a control block comprising two comparators, the control block being configured to:

compare the amplitude of the digital-audio signal to the first threshold and to the second threshold,

withhold the interrupt signal and activate the spectral analysis block when the amplitude of the digital-audio signal does not exceed the first threshold or the second threshold, and

generate the interrupt signal when the amplitude of the digital-audio signal exceeds the first threshold and the second threshold.

7. The system of claim 6, in which the spectral analysis block comprises:

a voice-activity detector random-access memory configured to receive the combined digital-audio signal from the sample combine block and store the combined digital-audio signal;

a fast Fourier transform block configured to perform a fast Fourier transform of the combined digital-audio signal stored in the voice-activity detector RAM and output fast Fourier transform results;

a spectral low-pass filter configured to compute and determine a mean amplitude of the fast Fourier transform results; and

a spectral voice-activity detector block configured to evaluate the results of the spectral low-pass filter and, based on the evaluation of the results of the spectral low-pass filter, signal the control block to generate the interrupt signal or to continue to withhold the interrupt signal.

8. The system of claim 1, further comprising a digital signal processor configured to receive the interrupt signal and analyze the digital-audio signal for a trigger phrase.

9. The system of claim 8, further comprising a voice random-access memory configured to receive and store the digital-audio signal in a first-in, first-out format, the voice random-access memory being configured to be accessed by the digital signal processor.

10. The system of claim 8 further comprising a clock system comprising a plurality of asynchronous clock generating mechanisms, each of the asynchronous clock generating mechanisms having a different current consumption, the clock system being configured to switch between the asynchronous clock generating mechanisms.

11. A method of detecting and capturing voice commands comprising:

receiving, by an analog-to-digital converter, an audio signal;

determining, by a wakeup circuit, whether the audio signal has an onset level of audio activity;

when the audio signal has the onset level of audio activity, placing the analog-to-digital converter in an active mode in which a converter portion of the analog-to-digital converter generates a digital-audio signal from the audio signal at an analog-to-digital converter sampling rate;

when the audio signal does not have the onset level of audio activity, placing the analog-to-digital converter in a sleep mode in which the converter portion of the analog-to-digital converter does not generate the digital-audio signal;

receiving, by a voice-activity detector, the digital-audio signal;

determining, by the voice-activity detector, the amplitude of the digital-audio signal;

comparing, by the voice-activity detector, the amplitude of the digital-audio signal to a first threshold and to a second threshold;

withholding, by the voice-activity detector, an interrupt signal when the amplitude of the digital-audio signal does not exceed the first threshold or the second threshold;

generating, by the voice-activity detector, the interrupt signal when the amplitude of the digital-audio signal exceeds the first threshold and the second threshold; and

performing, by the voice-activity detector, spectral analysis of the digital-audio signal when the amplitude of the digital-audio signal is between the first threshold and the second threshold.

12. The method of claim 11, in which determining, by the wakeup circuit, whether the audio signal has the onset level of audio activity comprises comparing the audio signal to a wakeup-onset threshold.

13. The method of claim 11 further comprising:

unpowering the converter portion of the analog-to-digital converter when the analog-to-digital converter in the sleep mode; and

powering the converter portion of the analog-to-digital converter when the analog-to-digital converter in the active mode.

14. The method of claim 11, further comprising:

receiving, by a digital signal processor, the interrupt signal from the voice-activity detector; and

analyzing, by the digital signal processor, the digital-audio signal to detect a trigger phrase.

15. The method of claim 14, further comprising receiving and storing, by a voice random-access memory, the digital-audio signal in a first-in, first-out format; and, when the trigger phrase is detected by the digital signal processor, discarding all data on the voice random-access memory before an endpoint of the trigger phrase.

16. The method of claim 14, further comprising, when the trigger phrase is detected by the digital signal processor, determining whether a user spoke during a prespecified timeout period after the trigger phrase.