August 2020

Device Control Via Encoded Wake Word Sequences

Anonymous

Device Control Via Encoded Wake Word Sequences

ABSTRACT

Utilization of patterns of wake words to perform device actions for voice activated electronic devices are described. Device actions, e.g., device reset, are encoded by utilizing a sequence of wake words with a specified timing gap between successive occurrences of the wake word. The sequence is processed at the device and decoded based on a match with predetermined sequence(s) of timings of the gaps. Based on the pattern of wake words detected at the device, actions are automatically performed. The encoded message can optionally include a device identifier. The sequence of wake words can be provided as input using a smartphone or other audio playback device. In some settings, where multiple devices are to be remediated, the sequence of wake words can be played back using a public address system, a loudspeaker, or other in-room device.

KEYWORDS

- Wake word
- Activation phrase
- Hotword
- Voice activation
- Factory reset
- Power cycle
- Reboot
- Smart speaker
- Smart display
- Smart appliance
- Debug mode
- Hard Reset

BACKGROUND

Electronic devices can sometimes require a remediating action such as a reset or power cycle to restore a device to a known state from a crashed/locked, or otherwise stuck state. For example, if a device is non-functional, a factory reset or power cycle may be performed to restore operation, or alternatively, the device may need to be rebooted using alternate/backup firmware. It may be possible for such an action to be performed remotely where the user is situated away from the device) but this is not always the case. In some circumstances, such as deployments of a large number of devices in a workplace or other setting, devices located in warehouses, showrooms, etc., reconfiguring devices may be difficult due to the challenges in performing a switch off/on action on multiple devices and/or use of reset buttons.

DESCRIPTION

Many devices such as smart speakers, smart displays, video calling appliances, and other smart appliances are built with voice activation features. The device includes microphone hardware that listens for a wake word (also referred to as hotword or activation phrase) and when the wake word is detected, interprets the subsequent audio as a user command. This disclosure leverages the capacity of such devices to recognize wake words. Different patterns of wake words are utilized to trigger device operations (actions) such as device reset, device system reconfiguration, device power cycle, etc. for such voice-activated electronic devices.

Per techniques of this disclosure, various device actions are encoded by utilizing a series (sequence) of wake words with a specified timing (gap) between the wake words. The number of occurrences of the wake word and the gaps in each sequence encode a specific action. The series of wake words is detected at the device, and decoded based on a match with predetermined sequence(s) of timings of the gaps between the wake words corresponding to an action. Based on patterns of wake words and corresponding gaps in the wake word sequences that are detected at the device, the corresponding actions are automatically performed.

Devices that support voice activation and/or voice control typically utilize an always-on dedicated digital signal processing (DSP) device or subsystem to listen for a wake word, even when the rest of the device hardware is inactive. When a wake word is detected, typically, an interrupt signal is generated. In response, other circuitry of the device is activated and additional audio received (e.g., that includes a command spoken by a user) is analyzed to detect the user command.

Since wake word detection is performed by standalone logic or a standalone device, separate from the main CPU, wake word detection can be utilized for remediating the device, even when the device is in a standby mode, or when the device is in an inoperative state (e.g., that requires a reset or reboot operation).

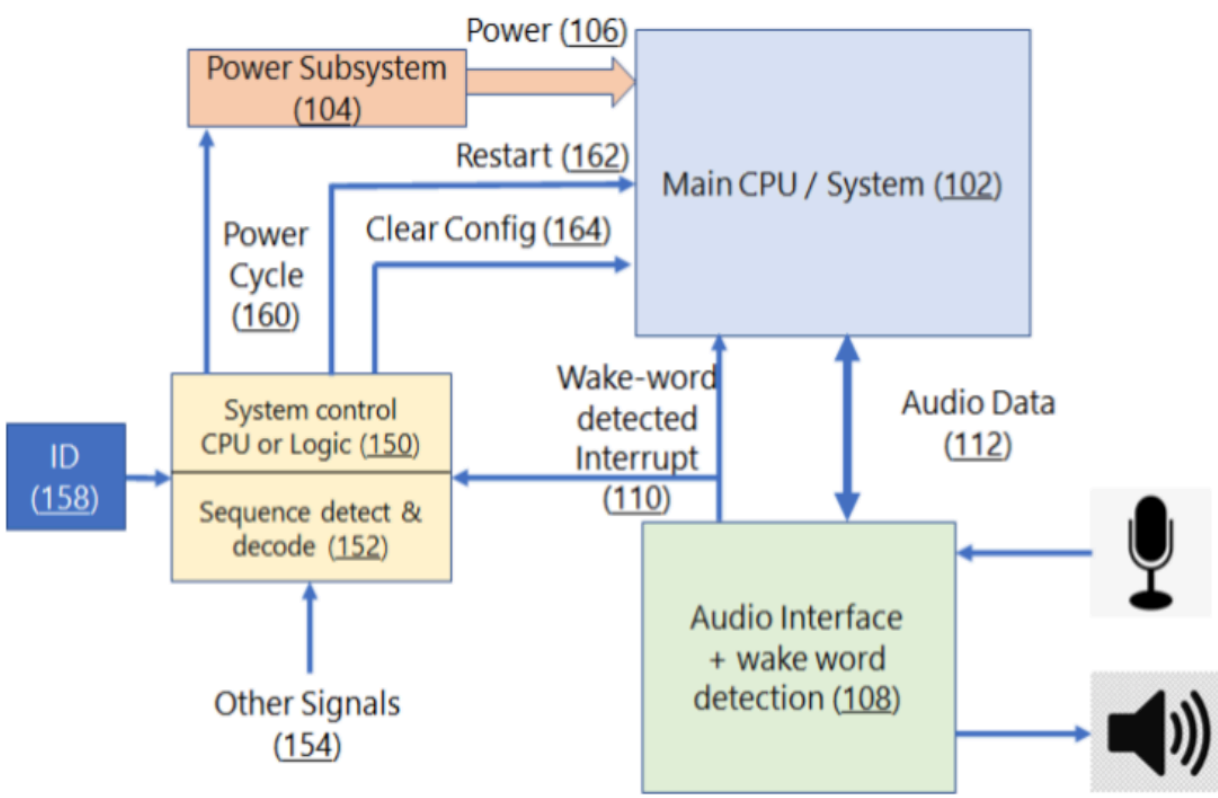

Fig. 1: Wake word sequences are utilized to perform system reset or other actions

Fig. 1 depicts an example electronic device configured for wake word based reset (or other actions), per techniques of this disclosure. Fig. 1 depicts base functionality components of the device as well as components that provide additional wake word detection based reset functionality.

The device includes a main central processing unit (CPU)/system (102), a power subsystem (104) that provides power (106) to the device, and an audio interface (108) utilized for wake word detection. The device additionally includes a system control CPU or logic (150) that is utilized to perform reset tasks such as power cycle (160), device restart (162), clear configuration (164), etc. that are activated by signals (154) provided by a user, e.g., by pushing physical buttons located on the device.

During normal operation of the device, the audio interface is utilized to monitor ambient audio to detect whether a wake word has been uttered by a user. When a wake word is detected, an interrupt (110) is generated and provided to the main CPU, to activate further audio processing at the main CPU. Additional audio data received at the audio interface is transmitted to the main CPU for additional audio processing, which may be performed on-device and/or by a remote server.

Per techniques of this disclosure, the always-on audio interface, the wake word detection process, and the interrupt generated based on a detected wake word are utilized to perform device reset actions. Additional programmable logic and/or code (152), e.g., that includes a simple state machine and timer are provided that process the interrupts, and to precisely time the gaps between successive wake words to detect and decode a wake word sequence. Device identifier(s) (158) are stored and serve to provide additional validation for the sequence(s) of wake words for certain actions. Detected sequences of wake words and decoded actions are utilized to implement voice-activated reset actions for the device.

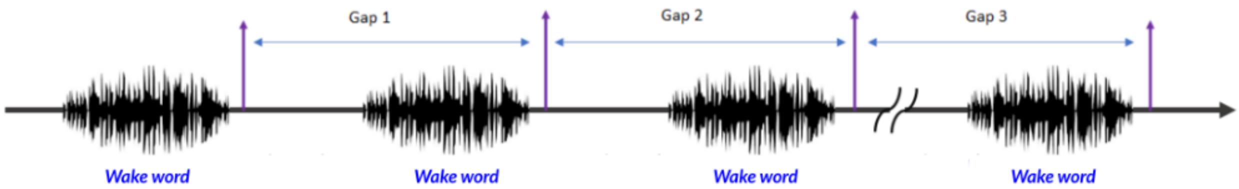

Fig. 2: Gap patterns in wake word sequences are utilized to encode reset actions

Fig. 2 depicts example patterns of wake word sequences utilized to encode device reset actions, per techniques of this disclosure. Fig. 2 depicts a sequence of wake words received at a device. This example sequence of four wake words is characterized by three gaps of time - gap 1, gap 2, and gap 3 in the different utterances of the wake words.

A variety of coding mechanisms can be utilized based on the sequence of wake words, and the measured gaps between the wake words, as long as the measured gaps lie within specified ranges for the gaps. Optionally, the encoded message can include a portion (or all) of a device identifier, e.g. serial number, MAC address, etc. that can allow for selectively activating an individual device (or sub-group) when multiple devices are present within audio range of the speaker or device from which the wake word sequence is output.

Encoded audio messages can be provided by a user speaking or playing a recording of a series of wake words with precise timing between the wake words. The device detects a series of wake words which match a given sequence of gap timings. For example, machine generated recordings of the sequence of wake words can be used to provide greater precision of gaps.

Detection of a wake word sequence can be initiated based on detection of a preliminary pattern in the wake word sequence, e.g., gap 1= gap 2 = X seconds (within a threshold margin of error), such that a random sequence of wake words received is not interpreted by the device as an encoded message that is to be decoded. In this illustrative example, gap 1 and gap 2 are utilized to validate that the received wake word sequence is in fact an encoded sequence, and gap 3 indicate the coded information. If any of the measured gaps are outside a validity range, additional received wake words are ignored for a predetermined period of time.

Fixed encoding

A fixed number of codes can be encoded based on a number of measurable coded time intervals in a valid time interval between wake words. For example, if a valid time interval for gap 3 is between 5 and 6 seconds, and the timer has a detection accuracy of 1 milliseconds with and interval size of 100 milliseconds, a code value between 0 and 9 can be generated by utilizing a fixed coding sequence.

For example,

```
if gap1 = gap2 = X /* detect sequence initiation */
  if  Y <= gap3 < Z then
    /* determine sequence identity */
    Transmitted code = (gap3 – Y)/interval size
```

In a scenario where X is set to 2.000 seconds, Y to 5.000 seconds, and Z to 6.000 seconds, and the first three wake words are received such that gap 1 = gap 2 = 2.000 seconds. Detection of gap 1 and gap 2 at the device is indicative of presence of an encoded message that is indicated by gap 3. If gap 3 is measured as 5.500 seconds, the detected code is determined to be 5500-5000/100, which represents a code of 5.

Byte-based encoding

Byte-based encoding can also be utilized such that each gap in a series of gaps in the wake word sequence (after the initial gaps used to validate that the sequence is, in fact, an encoded wake word sequence) is indicative of a byte in a multi-byte codeword. In such an implementation the interval size is such that, for a valid range for an encoded gap in the wake word sequence, it is divided into 256 equal subintervals. Multi byte sequences could be used for more complex encoding adding, for example, a checksum. They may also be used optionally in conjunction with a device identifier, to determine device action(s) to perform. For example, in some implementations, the first byte of the code, may be matched to the device identifier (or part of) for validity of action, and second byte may be indicative of the action, e.g. power cycle, to be performed.

For example:

```
interval size = (Z-Y)/256
if gap1 = gap2 = X /* detect sequence initiation */
  loop over N gaps…
    if  Y <= gapN < Z  then
    ByteN = (gapN – Y)/interval size
```

In some implementations, multiple wake words can be utilized in certain combinations with different delays for more complex encoded messages.

The decoded message can be mapped to a variety of actions such as, for example:

- Reset
- Power cycle
- Reset/Power cycle using alternate firmware/BIOS
- Enter debug mode
- Reset configuration

In a personal device setting, the sequence of wake words can be provided using a smartphone. A software application can be utilized to generate the sequence of wake words with precise timing corresponding to the reset action to be performed. In settings where multiple devices are to be remediated, the sequence of wake words can be played back using a public address system, a loudspeaker, or other in-room device. Remediation can also be implemented remotely. For example, while at a workplace, a user can reset a device at their home by playing back a reset sequence using another in-home device such as a smart speaker. Techniques of this disclosure enable device reset/remediation actions without physical contact and/or use of buttons.

CONCLUSION

Utilization of patterns of wake words to perform device actions for voice activated electronic devices are described. Device actions, e.g., device reset, are encoded by utilizing a sequence of wake words with a specified timing gap between successive occurrences of the wake word. The sequence is processed at the device and decoded based on a match with predetermined sequence(s) of timings of the gaps. Based on the pattern of wake words detected at the device, actions are automatically performed. The encoded message can optionally include a device identifier. The sequence of wake words can be provided as input using a smartphone or other audio playback device. In some settings, where multiple devices are to be remediated, the sequence of wake words can be played back using a public address system, a loudspeaker, or other in-room device.