June 2022

Audio Signature Based Hotkeys for Smart Glasses and Other Wearable Devices

Joshua Moore

D Shin

Audio Signature Based Hotkeys for Smart Glasses and Other Wearable Devices

ABSTRACT

Techniques are described for augmenting user interaction with smart glasses or other wearable devices that receive audio input. In a configuration phase, a user records an audio signature and links the audio signature to action(s) and/or features to be activated when the audio signature is detected. The audio signature patterns are parameterized based on a number of audio events, a time interval between the events, and the type of audio events. Entropy aware coding is utilized such that the most commonly utilized features are linked to audio signature patterns (codes) that are easier to generate. A trained local audio event classifier operates in a sliding window fashion to generate a sliding window inference. A time series score that is a measure of classifier certainty is determined based on the sliding window inference. A decoder is utilized to determine an intended user routine based on the time series score(s) and the identified user routine is performed, e.g., to wake the device, to launch a particular application, etc. Detection of the audio input can be performed by the wearable device or by other devices such as a smartphone.

KEYWORDS

- Smart glasses
- Wearable device
- Augmented reality (AR)
- Virtual reality (VR)
- Head-mounted display (HMD)
- Heads-up Display (HUD)
- Audio input
- Audible gesture
- Discreet input
- Huffman coding

BACKGROUND

Heads-up display (HUD) devices such as smart glasses have unique user interface and/or user input challenges due to their nature and use environment. A common activation and operation mechanism is via voice inputs or via a peripheral device such as a mobile phone, ring, watch, etc. In some social settings, interacting with the smart glasses can be intrusive, e.g., by the user touching the smart glasses frames. In some settings, interacting with the smart glasses can appear to be embarrassing, e.g., having to speak a passphrase such as "hey Glasses" to activate the device. Since smart glasses are worn by the user, the display is always accessible to the user in a discreet manner. Discreet ways of providing user input to smart glasses can improve the user experience in social settings.

DESCRIPTION

This disclosure describes techniques for augmenting user interaction with smart glasses. Per techniques of this disclosure, a user is provided with the ability to record an audio signature and link the audio signature to specific action(s) and/or features that are activated when the audio signature is detected. In listening mode, the smart glasses can detect remote audio signatures provided by the user, e.g., using surfaces other than the smart glasses. This enables the user to interact in a discreet manner with their smart glasses, without having to touch or speak to the smart glasses.

Smart glasses utilize their microphone(s) as a sensor to detect user input. With user permission, the microphone(s) can be deployed in an always-on state, whereby the glasses listen for voice input from the user, e.g., requesting activation of a digital assistant. The user is enabled to record an audio signature and subsequently program (link) a specific smart glass feature to that audio signature. The audio signature can be utilized by the user, in listening mode, to enable audio shortcuts or digital signatures for specific interactions with the smart glasses.

For example, a user can program a 'tap, tap, pause, tap' pattern by tapping on a hard surface, e.g., a table near the user, as a specific audio signature to unlock the display and display a home screen that includes the time. The same audio signature can also be utilized to restore the smart glasses to their previous state, e.g., a sleep state where the display is shut off. The audio signature provides the user with a discrete and responsive interaction option to interact with the smart glasses.

As another illustrative example, a second audio signature can be a pattern of three consecutive scratches (‘scratch, scratch, scratch’) performed at a specific pace that is utilized to trigger a calendar application that displays a time of the user’s next meeting on the smart glasses. The audio signature ‘tap, tap, pause, tap’ can subsequently be utilized by the user to return the smart glasses to their earlier state, where the display is switched off.

A large set of audio signatures (combinations) can be utilized, provided that the specific patterns in the audio signature(s) are unique and repeatable, thereby enabling detection and classification of the input pattern by an algorithm. The audio signatures can be parameterized using a framework that includes the following free variables that can preset by the manufacturer or programmed by the user during device setup:

- *N*, a number of audio events
- *T*, time intervals between audio events
- *X*, type of audio events

The specific (N, T, X) triplet(s) of audio signature patterns can be selected such that the user is not overburdened by having to memorize a large number of patterns, while still maintaining sufficient signal uniqueness to mitigate false positives. Additionally, entropy aware coding may be utilized such that the most commonly utilized features (e.g., by ranking features based on their frequency of use) are linked to audio signature patterns (codes) that are easier to generate (lower friction).

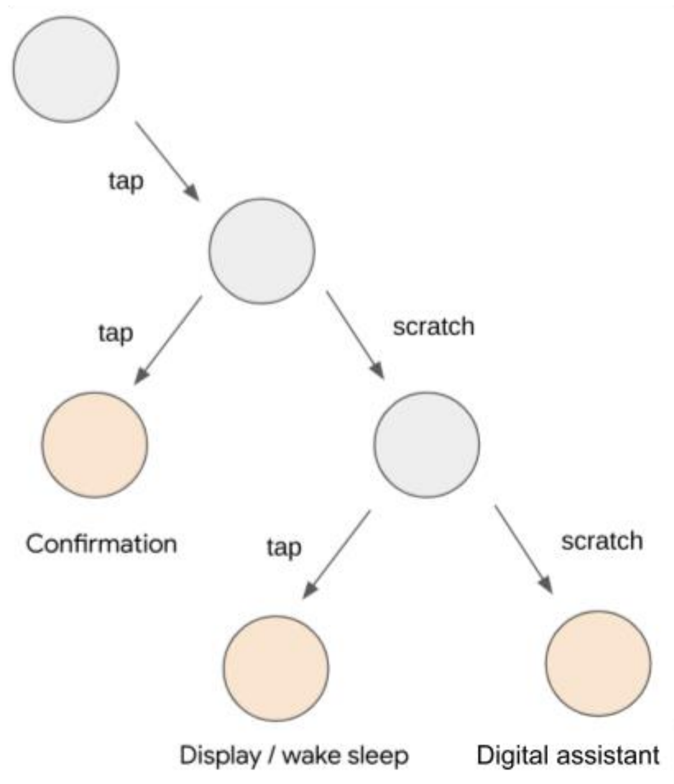

Fig. 1: A Huffman diagram depicting example audio signature patterns and linked features

Fig. 1 depicts an example Huffman diagram that depicts example audio signature patterns and their corresponding smart glasses features, per techniques of this disclosure. In this illustrative example, a relatively simple 'tap, tap' audio signature (sequence) is mapped to a most frequent user operation/interaction ('a confirmation routine') via the user interface, a 'tap, scratch, tap' is mapped to the second most frequent user operation/interaction ('toggle between wake/sleep'), and a 'tap, scratch, scratch' is mapped to the next frequent user operation/interaction ('activate digital assistant').

Within this framework, less frequent user interactions and/or operations are mapped to longer audio signature codes. This manner of entropy-aware encoding of the audio signature codes provides balance between minimizing user friction and maximizing signal robustness.

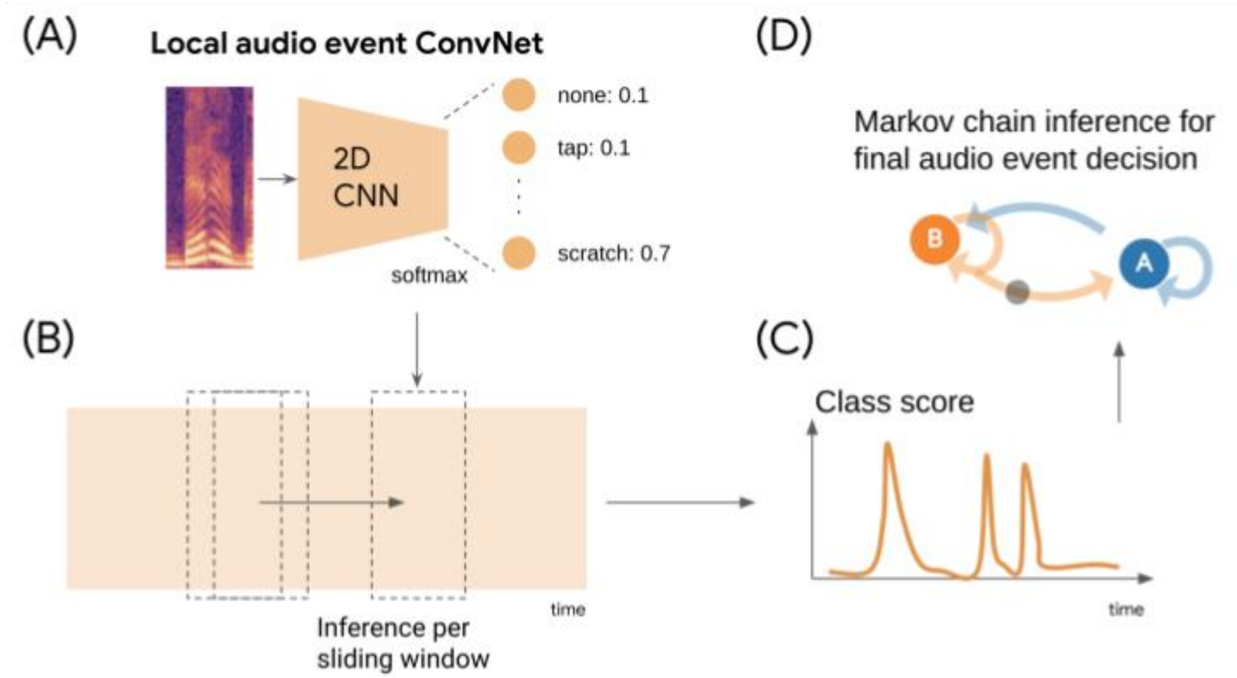

Fig. 2: Example workflow for detection of audio signatures

Fig. 2 depicts an example workflow for the detection and real-time inference of remote audio signatures, per techniques of this disclosure.

A. During a training phase, for each buffered microphone recording (e.g., a 200 ms duration recording) of an audio signature pattern , a spectrogram feature is computed to train a custom 2D convolutional network model (local audio event classifier) with a SoftMax layer of dimension N (where N is the variable of number of unique audio events, e.g., tap, scratch, etc.).
B. The trained local audio event classifier is operated in a sliding window fashion for all incoming buffered microphone recordings to generate a sliding window inference.

C. Based on the sliding window inference, a time series class score that ranges from 0 to 1 is determined. The class score is a measure of classifier certainty, where a score of 1 refers to a scenario in which there is a 100% certainty that the short, buffered microphone recording being analyzed includes a unique audio event of interest, and a score of 0 refers to a scenario in which there is a 0% certainty that the short, buffered microphone recording being analyzed includes a unique audio event of interest.

D. A decoder is utilized to determine an intended user routine based on the time series score(s). For example, a 'tap, pause, tap' is interpreted to be a user intent to wake the display. A maximum likelihood (ML) solution is computed by the decoder from a Markov chain where each node is defined by an audio event class.

Techniques of this disclosure can enable a user to interact discreetly with their smart glasses (or other device that receives microphone input) and activate selected features. The techniques describe a discreet and unique user customizable input feature that can be utilized to perform commonly utilized operations, e.g., lock and unlock a home screen on smart glasses, or activate specific smart glasses features such as to take a photo, launch the calendar, etc. In some implementations, the audio signatures can be detected by user devices other than the smart glasses, e.g., a user's mobile phone, which can then activate the intended operation on the smart glasses.

Further to the descriptions above, a user may be provided with controls allowing the user to make an election as to both if and when systems, programs, or features described herein may enable the collection of user information (e.g., information about a user's smart glasses, a user's audio input such as spoken input or input via audio from gestures, a user's preferences, or a user's current location), and if the user is sent content or communications from a server. In addition, certain data may be treated in one or more ways before it is stored or used so that personally identifiable information is removed. For example, a user's identity may be treated so that no personally identifiable information can be determined for the user, or a user's geographic location may be generalized where location information is obtained (such as to a city, ZIP code, or state level) so that a particular location of a user cannot be determined. Thus, the user may have control over what information is collected about the user, how that information is used, and what information is provided to the user.

CONCLUSION

Techniques are described for augmenting user interaction with smart glasses or other wearable devices that receive audio input. In a configuration phase, a user records an audio signature and links the audio signature to action(s) and/or features to be activated when the audio signature is detected. The audio signature patterns are parameterized based on a number of audio events, a time interval between the events, and the type of audio events. Entropy aware coding is utilized such that the most commonly utilized features are linked to audio signature patterns (codes) that are easier to generate. A trained local audio event classifier operates in a sliding window fashion to generate a sliding window inference. A time series score that is a measure of classifier certainty is determined based on the sliding window inference. A decoder is utilized to determine an intended user routine based on the time series score(s) and the identified user routine is performed, e.g., to wake the device, to launch a particular application, etc. Detection of the audio input can be performed by the wearable device or by other devices such as a smartphone.

REFERENCES

1. Sergio Olaya, Dmitry Shapiro. “Voice Signature for User Authentication to Electronic Device.” US Patent Application 15/598,998, filed May 18, 2017.
2. Rajeev Conrad Nongpiur, Akshay Bapat, Lawrence Wayne Neal. “Method and apparatus for audio summarization.” US Patent Application 14/992,700, filed Jan 11, 2016.