

Technical Disclosure Commons

Defensive Publications Series

April 2020

Ambient Hotword Detector To Facilitate Discovery and Use of Relevant App Functions

Victor Carbune

Matthew Sharifi

Follow this and additional works at: https://www.tdcommons.org/dpubs_series

Recommended Citation

Carbune, Victor and Sharifi, Matthew, "Ambient Hotword Detector To Facilitate Discovery and Use of Relevant App Functions", Technical Disclosure Commons, (April 01, 2020)
https://www.tdcommons.org/dpubs_series/3088



This work is licensed under a [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/).

This Article is brought to you for free and open access by Technical Disclosure Commons. It has been accepted for inclusion in Defensive Publications Series by an authorized administrator of Technical Disclosure Commons.

Ambient Hotword Detector To Facilitate Discovery and Use of Relevant App Functions

ABSTRACT

Audio input and output is used in a wide variety of mobile applications that provide a visual user interface that enables users to provide input. Visual UI often includes a number of elements that may be difficult to discover or not immediately available to the user at a time when the user needs them.

This disclosure describes the use of a lightweight hotword detector to make relevant app UI elements easily distinguishable and discoverable at opportune times during voice calls and other interaction in which user-provided audio is available. User-facing visual UI elements within each app are tagged and associated with hotwords. When a hotword is detected in user-provided audio, the corresponding app is notified and can visually enhance the UI element associated with the hotword. Implementation of the described techniques can improve user experience by facilitating quick discovery and use of contextually relevant functions.

KEYWORDS

- Hotword detection
- Keyword detection
- Voice interaction
- UI discoverability
- UI element tagging
- UI highlighting
- Virtual assistant
- Voice assistant

BACKGROUND

Audio input and output is used in a wide variety of mobile applications, such as phone calls, messaging, voice assistants, etc. Such applications are driven via a visual user interface, e.g., a phone dialer, a message composition interface, etc. Visual UI can often include a number of elements that may be undiscoverable or not immediately available to the user at a time when the user needs them. During audio interaction, e.g., a voice call, interaction with a voice assistant, etc., a user may find it difficult to locate specific UI elements displayed on the screen. For example, during an online meeting, it is likely that a user will find it difficult to locate specific UI elements on the screen or within the app that are of use at a particular point in time, e.g., a “hang up” button, a “mute” button, a “turn video on” button, etc.

DESCRIPTION

This disclosure describes the use of a lightweight hotword detector to make relevant app UI elements easily distinguishable and discoverable at opportune times during voice calls and other interaction in which user-provided audio is available. The hotword detector is implemented with specific user permission. The hotword detector can support various specific phrases uttered during a voice call or other interaction. For example, upon detecting that the user uttered “bye” during a phone call, the phone dialer app is provided with an input such that the app UI button “End call” is automatically highlighted in the displayed user interface.

To support such highlighting, the user-facing visual UI elements within each app are tagged with a set of keywords and synonyms that describe the corresponding UI elements. For instance, an “End Call” button within dialer UIs can be tagged with phrases such as “bye,” “goodbye,” “talk to you later,” etc. Similarly, the “Shutter” button within a camera UI can be associated with terms such as “snap photo,” “take selfie,” etc. Such hotword tagging of UI

elements can be done online and/or offline. If the user permits, the tagged UI elements can be highlighted for enhanced discovery at an opportune time based on the user's voice input, e.g., user's speech during a voice call.

The user can use any available application per its usual functionality and continue to provide speech input, e.g., participate in a voice/video call, record a video, etc. With user permission, audio content of the speech is analyzed using a lightweight on-device hotword detector to detect whether the audio content contains any of the supported tagged phrases in the most recent time interval. For example, such ambient hotword detection can be performed by employing a suitable machine learning model. The model can detect hotwords across classes, or alternatively, can be trained on a per-app basis. Alternatively, app-specific parameters can serve as model input in order to constrain the detector output to the hotwords that are applicable for the specific app.

Hotword detection can be implemented as a system-wide feature, e.g., by a device operating system. When a hotword that matches a tagged visual UI element is detected, the corresponding app is notified accordingly. Subsequently, the UI element associated with the detected hotword can be enhanced via techniques such as highlighting, size change, etc. Such enhancement of the UI element can help the user discover and/or locate the app function that is likely relevant to the user's ongoing activity as inferred from the detected hotword.

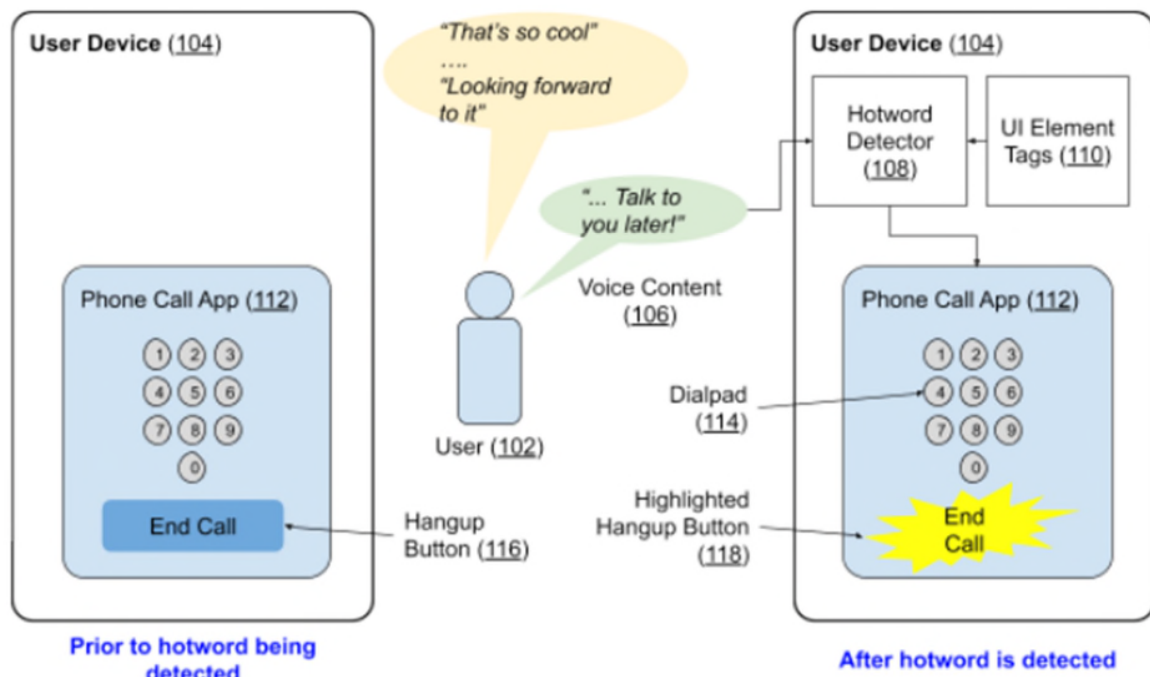


Fig. 1: Automatically highlight relevant UI elements based on contents of a user's voice

Fig. 1 shows an operational implementation of the described techniques. A user (102) uses a phone call app (112) on a user device (104) to engage in a phone call. With permission, the voice content (106) of the call is analyzed using an ambient hotword detector (108) running on the device. The hotword detector is provided with various hotword tags (110) associated with UI elements within the app, such as the dialpad (114) and the hang-up button to end the call (116). Prior to detection of the hotword, the UI elements are in a normal, non-highlighted state, as seen in the prior to hotword being detected stage of the phone call illustrated on the left-hand side of Fig. 1.

When the most recent voice content is detected as including the hotword "Talk to you later" which is associated with the UI element for hanging up the call, the corresponding UI button (118) is prominently highlighted, as seen in the after hotword is detected stage of the phone call illustrated on the right-hand side of Fig. 1, to make it easy for the user to spot and use.

While Fig. 1 illustrates the use of a change in shape and color, any type of highlighting, e.g., a blinking transition, a color-change transition, a change in size, etc. can be used as appropriate. For example, the mode of highlighting can be chosen by individual app developers for the app.

Hotword detection is performed over a time interval spanning several milliseconds of recent audio. The time interval can be set by the developers of the OS or app, and/or determined dynamically.

When the app is an app that accepts voice commands, e.g., a virtual assistant, the hotword detector can enable graceful degradation when the user intent is unknown. For example, if the user utters “navigate to a fun place,” instead of showing “I cannot do that,” the virtual assistant can highlight a “navigate” button and encourage the user to tap into an app, e.g., a maps application or an app that provides information regarding places.

If the user permits, the techniques described in this disclosure can be applied to any devices and apps that involve the user speaking while using the app, regardless of the specifics of that functionality. Such apps include, e.g., phone, messaging, video conferencing, media capture, voice assistants, etc. With user permission, implementation of the described techniques can improve the user experience and the accessibility of the underlying app by facilitating quick discovery and use of contextually relevant functions.

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 collection of user information (e.g., information about a user’s speech input), 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. 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

This disclosure describes the use of a lightweight hotword detector to make relevant app UI elements easily distinguishable and discoverable at opportune times during voice calls and other interaction in which user-provided audio is available. The hotword detector is implemented with specific user permission. User-facing visual UI elements within each app are tagged and associated with a set of keywords and synonyms. When a hotword is detected in user-provided audio, the corresponding app is notified and can visually enhance the UI element associated with the hotword. Implementation of the described techniques can improve user experience by facilitating quick discovery and use of contextually relevant functions.

REFERENCES

1. Faaborg, Alexander, and Peter Ng. "Visual confirmation for a recognized voice-initiated action." U.S. Patent 9,575,720, issued February 21, 2017.
2. Sharifi, Matthew. "Promoting voice actions to hotwords." U.S. Patent 8,719,039, issued May 6, 2014.