August 2020

Adapting Delivery of Virtual Assistant Responses That Include Private Information

Victor Carbune

Matthew Sharifi

Adapting Delivery of Virtual Assistant Responses That Include Private Information

ABSTRACT

Some queries to a virtual assistant may seek information specific to that particular user. In some of these instances, the query and/or the response may contain information that the user considers to be private. In such cases, if a virtual assistant delivers the response in the form of audio, such delivery can potentially reveal the user's private information to others who might be in the vicinity, close enough to hear the spoken response from the virtual assistant. This disclosure describes techniques that automatically adjust response delivery by a virtual assistant whenever a user's voice query and/or the information in response to the query are determined as likely to include private information. Delivery of the response is adapted to present the private information in a manner that avoids it being overheard.

KEYWORDS

- Virtual assistant
- Digital assistant
- Voice query
- Private response
- Spoken response
- Smart speaker
- Smart display

BACKGROUND

People often use virtual assistants to seek a variety of information by providing voice commands. The information sought by the user's query might be of general interest and publicly available. For instance, answers to user queries such as "What is tomorrow's weather forecast for

Los Angeles?" or "How tall is Mount Whitney?" can be obtained from publicly available data and are not specific to a particular user.

In contrast, some queries may seek information specific to the particular user that issues the query. In some of these instances, the query and/or the response may include information the user considers to be private. For example, a user may ask: "What's next on my schedule today?" or "What is my credit card balance?" In such cases, if a virtual assistant delivers the response in the form of audio, it can potentially reveal the user's private information to others who might be in the user's vicinity and close enough to hear the spoken response from the virtual assistant. For example, a user at work might not want coworkers to hear a response such as "Next, you have a meeting with your bank representative at 3pm about your mortgage application." In such scenarios, the user may prefer that the virtual assistant exclude the private information when delivering the response or to deliver the response in a lower volume or in text form, sent to the user's device.

DESCRIPTION

This disclosure describes techniques that automatically adjust response delivery by a virtual assistant whenever a user's voice query and/or the information in response to the query are determined as likely to include private information. With user permission, the voice query and/or the response are examined to flag any potentially private information based on the query audio and other relevant contextual information obtained with the user's permission. Delivery of the response is adapted to present the private information in a manner that avoids it being overheard.

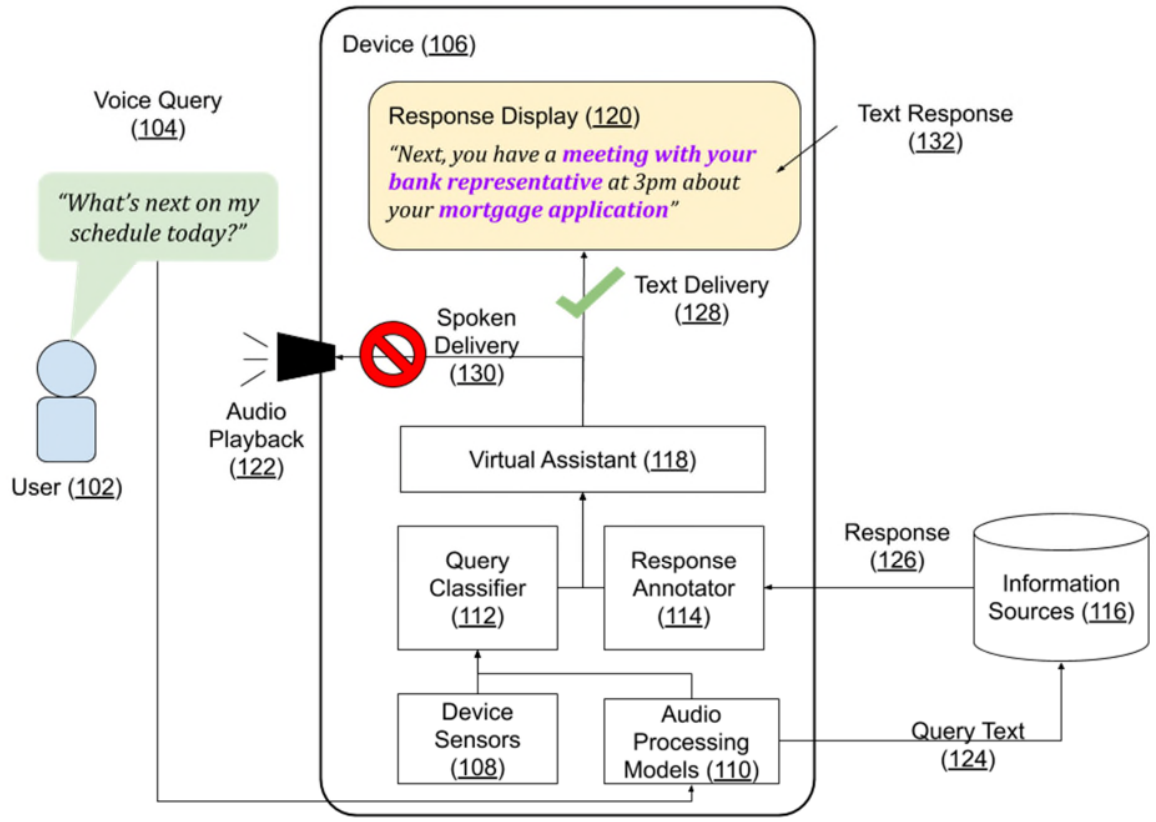

Fig. 1: Adapting response delivery for private information

Fig. 1 shows an operational implementation of the techniques described in this disclosure. A user (102) issues a voice query (104) to a virtual assistant (118) provided via a device (106). With user permission, the query audio is processed by trained machine learning models (110). The extracted text of the query (124) is used to obtain a response (126) from appropriate information sources (116). The information sources may include on-device sources (e.g., locally stored information for which the user has provided permission, such as the user's calendar, user's photos, etc.) and/or online information sources, e.g., private information sources as permitted by the user, and other online information.

If the user permits, the output of the audio processing models and contextual information obtained from device sensors (108) is input to a query classifier (112) to determine if the query is

likely a private query. Similarly, the response (126) is annotated using a response annotator (114) to flag potentially private information within the response content. Based on the output of the trained machine learning models that determine whether the query and/or response contain information private to the user, delivery of the response by the virtual assistant is adapted.

In the example of Fig. 1, the response contains private financial information ("meeting regarding mortgage application) and is therefore set for text delivery (128) and shown as text (132) on the device display (120). Spoken delivery (130) via audio playback (122) is suppressed.

If the user permits, the query audio is processed via suitable trained machine learning models, such as neural networks, designed for sound understanding. The models analyze the audio to derive various properties of the user's tone of voice, such as 'whisper,' 'soft,' 'loud,' etc., that can indicate whether the query itself is private. For instance, a user whispering the query can indicate that the query is to be treated as private. With user permission, the model output can also classify the user's environment (e.g., 'indoors,' 'outdoors,' etc.) and the surrounding sounds ('footsteps,' 'speech,' etc.). Further, additional models trained specifically to detect co-presence of people can be used with user permission to determine whether other persons are present in the user's vicinity.

To determine if the response to the user's query is likely to include private information, a separate natural language understanding (NLU) based classifier using a trained natural language processing (NLP) model is used to process the text-to-speech transcription of the user's voice query. If the user permits, the output of this model is combined with the output of the audio processing models to generate a combined indicator of whether the response to the user's query is to be treated as private, based on the user's tone, environment, and the content of the query and/or response.

A response retrieved for a query determined to be of a private nature can include metadata corresponding to individual tokens (or range of tokens) within the query. Suitable trained machine learning models, such as multi-layered neural networks, fine-tuned Bidirectional Encoder Representations from Transformers (BERT) models for text understanding, etc., can be utilized to annotate the response based on the query tokens and the metadata. The annotations can be based on model output indicating whether a given part of the response is private. For example, if the query response is "Next, you have a meeting with your bank representative at 3pm about your mortgage application." then the terms 'meeting with your bank representative' and/or 'mortgage application' parts can be annotated with a tag indicating that they are private. In addition, the information source for the response can be used to determine whether the response is to be treated as private. For instance, responses derived from information included in a private data store can be considered private.

Delivery of the response to the user is adapted based on the annotations marking private parts of the response. Based on the annotations, the delivery volume of the audio for the private parts of the response can be set to low. Alternatively, the entire response can be delivered in text form to the user's device instead of being provided as a spoken response. Further, neither the query nor the response is logged if deemed private

Further, if the user permits, device capabilities, e.g., whether the device has a display, are determined. For example, in case of a query to a device such as a smart speaker, that lacks a display, it can be determined whether another screen-based device. e.g. a user's watch, is available but is not handling the query. Upon identification of such a device, the private result can be displayed on that device, rather than provided as a spoken response via the smart speaker.

The operation can be set to trigger private response delivery for all queries in which the user is detected to be whispering since there is a high amount of certainty that such queries, and consequently, the responses, are private. Further, such cases can be handled with additional protection mechanisms, such as local processing of the query so that the query audio stays on the user device and is not transmitted to a server.

The threshold values for the various models used in the implementation of the techniques can be set by the developers and/or specified by the users and/or determined dynamically at runtime. The techniques can be implemented within any device, application, or service that incorporates voice based virtual assistants. Implementation of the techniques can enhance the user experience (UX) of interaction with virtual assistants for private queries and ensure that a user's private information is not inadvertently revealed to others by the spoken responses of a virtual assistant.

Further to the descriptions above, a user is 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 queries, voice volume level, ambient environment, 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

This disclosure describes techniques that automatically adjust response delivery by a virtual assistant whenever a user's voice query and/or the information in response to the query are determined as likely to include private information. Delivery of the response is adapted to present the private information in a manner that avoids it being overheard. Detection of whether a query and/or a response is private is performed on-device using trained machine learning models that utilize query and response content and metadata as input. The models generate annotations that are utilized to adjust the volume level of response delivery or to determine that the response is to be displayed, but not spoken.

REFERENCES

1. Parviainen, Emmi, and Marie Louise Juul Søndergaard. "Experiential Qualities of Whispering with Voice Assistants." In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1-13. 2020.
2. Sharifi, Matthew and Carbune, Victor, "Delivering Audio Responses At Contextually Appropriate Volume Level", Technical Disclosure Commons, (April 13, 2020) https://www.tdcommons.org/dpubs_series/3128
3. Cărbune, Victor and Deselaers, Thomas, "Context-dependent output volume for voice-controlled virtual assistant", Technical Disclosure Commons, (December 19, 2018) https://www.tdcommons.org/dpubs_series/1789