Technical Disclosure Commons

Defensive Publications Series

November 21, 2017

Speaker identification based on user context

Minkoo Seo

Speaker identification based on user context

ABSTRACT

Recognition of a person that issued a voice command is necessary for a device that implements a voice-based user interface to respond to the user command. This disclosure describes techniques that can be implemented by devices that provide a voice-based user interface, e.g., a smart home speaker, an appliance, etc. to recognize the identity of a person that issues a voice command. With user permission, user context data is obtained and used to determine the identity of the person. Such data can include, with user permission, location data, recent activity data, user interests, etc. Improvement in the identification of the person issuing a voice command enables the devices to provide an improved user experience.

KEYWORDS

- smart home speaker
- voice UI
- voice command
- voice interaction
- virtual assistant
- user identification

BACKGROUND

Devices with a voice-based user interface, e.g., smart speakers, appliances with voice user interface, etc. are popular. Identification of the speaker that issues a voice command is a challenge for such devices that may receive voice input from more than one user, e.g., since the devices may be shared among members of a family. Without knowing who the used that issued the voice command is, such devices may have difficulty answering questions such as “what's on

my calendar," or "read my latest email," as the answers to these questions typically depend on the identity of the user asking the questions.

Some current identification methods have limitations. For example, some websites and mobile applications ask for personal identification information such as email, birth date, phone number, etc. to verify user identity. However, the answers to such identification questions may not be secret among family members or other groups of people that know each other. Also, a user response to such a question may be overheard by others nearby, making the answer no longer a secret. When users permit such analysis, voice biometric based approaches can model characteristics of a user's voice and can recognize the user based on such characteristics. However, biometric approaches have limitations as well. For example, a biometric system may be deceived by the user's voice being recorded and played back to the device that responds to voice commands.

DESCRIPTION

Per techniques of this disclosure, devices with voice user interfaces, e.g., smart speaker systems, obtain user permission to use user context to provide personalized answers. Users can provide permission to use specific contextual factors, and can opt to deny permission for one or all the factors. Users may also restrict permission, e.g., to specific locations, to specific time periods, etc. When users don't provide permission, devices are provided such that the techniques described in this disclosure are disabled.

When users provide permission, some implementations described herein make use of user context information for which the user has provide permission to determine the identify of a person that issued a voice command. For example, the identity of a can be determined based on user context, including user location, user's previous questions, user preference stored by the

device, user interests, etc. A device that provides a voice based user interface makes use of such context information to identify the person that issued the voice command. Further, with user permission, the device can use a machine learning techniques to learn which context signals can be used to reliably identify the person.

Example scenario 1

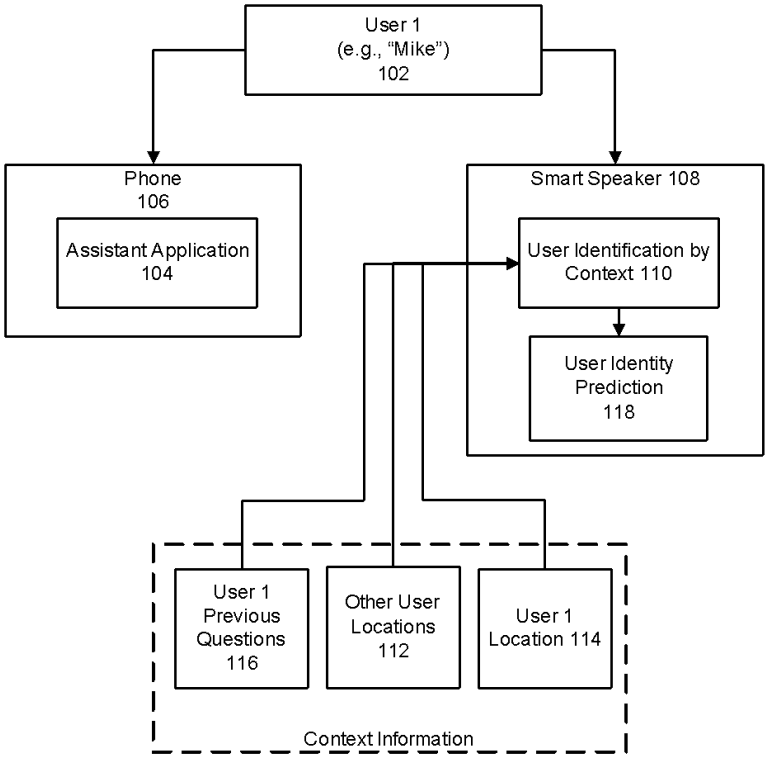

Fig. 1

Referring to FIG. 1, a user (102), Mike, asks an assistant application (104) on his phone (106) a question, e.g., "what's the weather today?" when entering a home. The user puts his phone on his desk in his room and goes to a living room. There, Mike asks a smart speaker (108) another question, e.g., "do I have anything on my agenda this evening?"

At this point, presuming that the user Mike has provided permission for use of context information, the smart speaker utilizes user identification by context (110), e.g., implemented as

a software module, or alternatively provided from a server or other external system, determines that the voice command is from the user Mike based on the following: (a) other users' locations (112), obtained with permission of the other users: e.g., if all family members are not at home, the speaker is likely to be Mike; (b) Mike's location and device location(s) (114), obtained with Mike's permission): Mike just entered the home, and put his phone in his room. Thus, it is likely that Mike started using another device, e.g., the smart speaker; and (c) Mike's previous questions (116), obtained with Mike's permission: Mike just asked about today's weather.

Based on the context, e.g., (a), (b), and (c), above and optionally, the features of person's voice, the system infers that Mike is the person that asked the question. about today's agenda. In this manner, more accurate speaker recognition enables an improved prediction of the identity of the person that issued the voice command (118).

Example scenario 2

A user, David, asks "who's ranked at the top in the NBA?" Knowing that only David is interested in sports among his family members (when David and the family members provide permission for use of user interest data), a smart speaker in David's home can readily determine that the speaker is David. Based on the determination, the smart speaker replies "It's the Knicks. By the way, you were watching re-play of the team yesterday. Do you want to continue watching it on your TV?" David says "yes", the smart speaker starts continues playing the replay of the game to the TV.

In the example scenarios, above, the example of a family has been used for illustration purposes. The techniques of this disclosure can be used for any scenario where multiple users interact with a device that provides a voice-based user interface, e.g., coworkers, classmates, roommates, etc.

In situations in which certain implementations discussed herein may collect or use personal information about users (e.g., user data, information about a user's social network, user's location and time at the location, user's biometric information, user's activities and demographic information), users are provided with one or more opportunities to control whether information is collected, whether the personal information is stored, whether the personal information is used, and how the information is collected about the user, stored and used. That is, the systems and methods discussed herein collect, store and/or use user personal information specifically upon receiving explicit authorization from the relevant users to do so. For example, a user is provided with control over whether programs or features collect user information about that particular user or other users relevant to the program or feature. Each user for which personal information is to be collected is presented with one or more options to allow control over the information collection relevant to that user, to provide permission or authorization as to whether the information is collected and as to which portions of the information are to be collected. For example, users can be provided with one or more such control options over a communication network. 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. As one example, a user's identity may be treated so that no personally identifiable information can be determined. As another example, a user's geographic location may be generalized to a larger region so that the user's particular location cannot be determined.

CONCLUSION

This disclosure describes techniques that can be implemented by devices that provide a voice-based user interface, e.g., a smart home speaker, an appliance, etc. to recognize the identity of a person that issues a voice command. With user permission, user context data is obtained and

used to determine the identity of the person. Such data can include, with user permission, location data, recent activity data, user interests, etc. Improvement in the identification of the person issuing a voice command enables the devices to provide an improved user experience.