June 13, 2019

Intelligent suggestions for actions during voice calls

David Petrou

Brandon Barbello

Benjamin Schlesinger

Blaise Aguera-Arcas

Alex Agranovich

See next page for additional authors

Inventor(s)
David Petrou, Brandon Barbello, Benjamin Schlesinger, Blaise Aguera-Arcas, Alex Agranovich, Ori Kabeli, and Nadav Bar

Intelligent suggestions for actions during voice calls

ABSTRACT

Participants in a voice call often exchange small yet important pieces of information, e.g., phone numbers, meeting locations, action items, etc. To capture such information, users often interrupt call flow to take down notes, which is an inconvenience. This disclosure describes machine-learning techniques that automatically detect and capture (with permission from the participants to the call) noteworthy pieces of information that are detected in the audio flow between the participants.

KEYWORDS

- Voice call
- Keyword detection
- Speech recognition
- Speech summarization
- Note Taking
- Call transcription

BACKGROUND

In a voice call, e.g., landline, cellular, VOIP, video, or conference calls, voicemails, etc., participants often exchange small and important pieces of information, e.g., phone numbers, addresses, meeting locations and times, action items to be followed up after the phone call, etc. These pieces of information aren't usually captured unless the user deliberately interrupts the normal flow of the call to write down a note or place a reminder on their mobile device. Current techniques to note in-call data of interest are thus at best inconvenient. At worst, users forget or

neglect to take down notes due to the friction involved; as a result information of interest from the call is lost.

DESCRIPTION

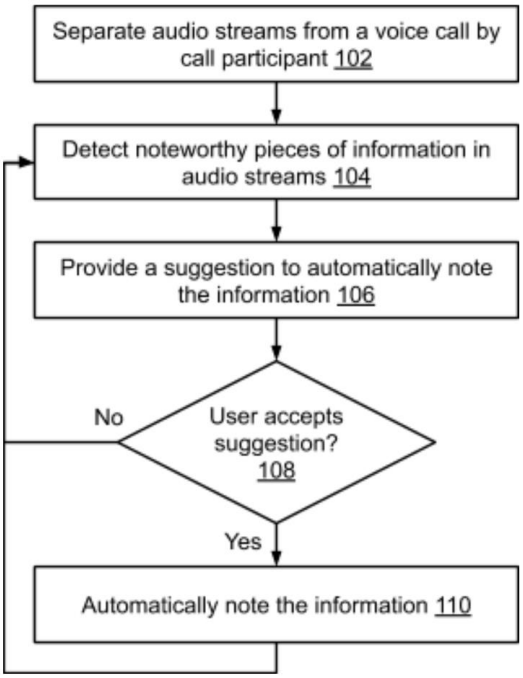

Fig. 1: Intelligent suggestions for actions during a voice call

Fig. 1 illustrates intelligent suggestions for actions during voice calls, per techniques of this disclosure. For example, the techniques can be incorporated into a service or application that runs on a mobile device. User permission is obtained prior to activating the method shown in Fig. 1 and if the user denies permission, call data is not accessed.

The service separates audio streams in a voice call by participant (102). With user permission, a machine-learning model, e.g., an on-device model, detects noteworthy pieces of information within an audio stream (104) of the call. With user permission, the machine-learning model also predicts if the user is about to take some action, e.g., a note-taking action, on their mobile device. The service provides a suggestion (106) to the user, e.g., via an unobtrusive

notification, to automatically note down the detected information. If the user accepts the suggestion (108), the service automatically performs the action, e.g., notes the detected information (110).

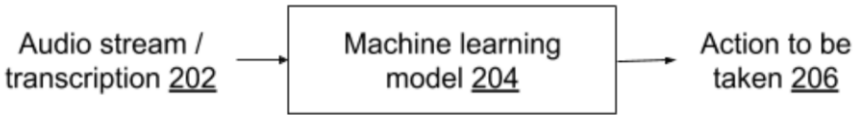

Fig. 2: Machine learning model to predict action to be taken based on audio stream

Fig. 2 illustrates the use of a machine learning model (204) to predict an action (206) to be taken based on an input audio stream (202) of a voice call or transcription thereof. The audio stream input is the audio stream of the voice call separated by participant, obtained with the participant's permission. Analysis of the call and provision of suggestions is performed only when the user provides permission for such analysis for the service that provides automatic suggestions. The transcription, if any, is performed by an automatic speech-to-text service. The actions predicted or detected by the machine learning model can include, for example,

- an action item or a reminder triggered by the content of the voice call conversation, e.g., when a participant says, "can you please buy milk when you return home today";
- an address/location uttered during the call;
- a phone number uttered during the call;
- an intention to book a taxi or other transport medium was uttered during the call;
- an intention to book a ticket to an event, e.g. movie, concert, etc., uttered during the call; etc.

Further, for the detected events, the machine learning model suggests the most relevant app to fulfill detected user intent. For example, if an intention to book a taxi was detected, then the machine learning model suggests that the user open a ride-hailing app or a city taxi app. With

user permission, the model also captures context related to user intent, such that the relevant app is opened with context already filled in. For example, the ride-hailing app is opened with the destination filled in if the destination is detected within the voice call. In another example, a calendar app is opened with the time and location for a meeting filled in based on the content of the voice call. With user permission, the service also offers the user a diary of important notes and intents detected in various calls, such that a user can take an action or remember something from a call after the call ends. The diary can be based on notes that the user made in previous calls. The model can be deployed on a client device of the user and/or a server.

The machine learning model is trained on voice call conversations for which user permission is obtained for the purpose of training or on synthetically-generated voice call data. If the user permits, online training of the model can be performed during operation, e.g., based on user actions with respect to the suggestions, e.g., whether the user selected or dismissed the suggestion. Trained models can be uploaded to a server and used to build a higher quality model, if the user permits. The higher quality model can be downloaded to client devices where the service is deployed.

The machine learning model can include, e.g., regression learning models, generative learning models, neural networks, etc. Example types of neural networks that can be used include neural networks, long short-term memory (LSTM) neural networks, recurrent neural networks, convolutional neural networks, etc. Other machine learning models, e.g., support vector machines, random forests, boosted decision trees, etc. can also be used.

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 social network, social

actions or activities, profession, 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 machine-learning techniques that automatically detect and capture (with permission from the participants to the call) noteworthy pieces of information that are detected in the audio flow between the participants.

REFERENCES

[1] Shires, Glen, Sterling Swigart, Jonathan Zolla, and Jason J. Gauci. "Speech recognition and summarization." U.S. Patent 9,420,227, issued Aug. 16, 2016.