February 13, 2019

Automatic problem detection during audio or video call

Dhruv Bakshi

Jakob Foerster

Automatic problem detection during audio or video call

ABSTRACT

Participants in an audio or video call over a network often face problems such as being unable to hear or see each other. With user permission, speech data from the call is automatically analyzed by a classifier to identify whether call participants are facing a problem. If a problem is identified, a message indicating a possible solution is provided, or if the call participant permits, automatic action is taken to fix the source of the problem. Users are provided with options to turn off the techniques.

KEYWORDS

- video chat
- audio chat
- videoconferencing
- speech recognition
- garbled audio
- choppy video

BACKGROUND

Users that conduct audio or video calls, e.g., via a mobile or desktop app, or a video conferencing system, face various problems with audio or video in the call. Underlying causes for the problem can be related to device software or hardware configuration, e.g., screen, speakers, headphones, microphones, etc.; the network connection between the call participants, e.g., latency; or in case of server-mediated calls, server-side problems. While network-related or server-side problems can be diagnosed with data from the call, detecting other causes of problems non-trivial, e.g., since the data from the call often does not include device configuration data.

DESCRIPTION

An example of a common problem is when call participant that is not able to see or hear other participants. In such circumstances, the call participant (or others) often utter phrases such as, “hey I can't hear you at all, but I can see you," "I can hear you but can't see you," "the picture is really blurry," "I can barely hear you," etc. Call participants also guide each other in attempts to fix the problem, e.g., by providing instructions such as “can you increase your mic volume?” “turn off the video,” etc.

With user permission, this disclosure described techniques that access such user speech data and use it to troubleshoot the call, and possibly, provide an automatic solution. The techniques described are implemented with user permission (from call participants) and can be turned off at any time. If call participants do not provide permission, the techniques are not implemented.

When the call participants permit, speech data from the call is analyzed using speech recognition to techniques. For example, a trained classifier (e.g., a machine learning classifier) can be utilized to detect when users complain about something related to the audio or video stream. When it is determined that one or more of the call participants is facing a problem in the call, a message can be provided to provide an instruction to the user, or if the user permits, an automatic action (e.g., increasing the speaker volume) is performed that can fix the problem. The speech recognition and classifier techniques can be implemented on the user device and/or on a server that mediates the call, based on user permission.

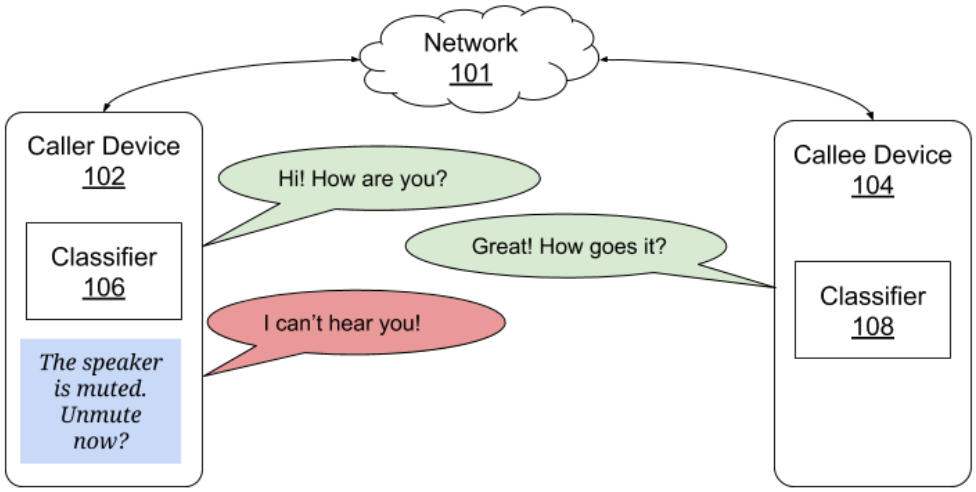

Fig. 1: Detecting problems in a call by speech recognition

Fig. 1 illustrates an example of detecting problems in a call, per techniques of this disclosure. A caller device (102) and a callee device (104) participate in a call that is conducted over a network (101). As illustrated, one or more the devices (or a server that mediates the call, not shown) include a trained classifier (106, 108).

As illustrated in Fig. 1, the participants begin the call by greeting each other. However, the caller speech "I can't hear you!" indicates that there is a problem with the call. With user permission, the classifier automatically detects that there is a problem with the call. Based on the detected problem ("caller unable to hear"), various likely sources for the problem are evaluated. In the example illustrated in Fig. 1, it is determined that the speaker of the caller device is muted. A message is provided to the user ("The speaker is muted. Unmute now?") which allows the user to fix the problem. Alternatively, if the user permits, the speaker is turned on automatically.

While Fig. 1 illustrates a caller-device problem (speaker muted), other causes such as network latency or a muted microphone at the callee device can also be detected, and corresponding instructions can be provided.

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

Participants in an audio or video call over a network often face problems such as being unable to hear or see each other. With user permission, speech data from the call is automatically analyzed by a classifier to identify whether call participants are facing a problem. If a problem is identified, a message indicating a possible solution is provided, or if the call participant permits, automatic action is taken to fix the source of the problem. Users are provided with options to turn off the techniques.