Audio Status Adjustment Based On Lip Movements and Gaze Direction

N/A
Audio Status Adjustment Based On Lip Movements and Gaze Direction

ABSTRACT

Video conference (VC) user interfaces include a feature that provides users with the option to mute or unmute their audio status during a meeting. A user may forget to mute themselves when they are not speaking, causing background audio to be transmitted. A user may attempt to speak while unmuted, without realizing that they are muted. Such mode errors can interrupt the flow of the VC meeting. This disclosure utilizes one or more user-permitted factors, detected locally at a user device, to automatically adjust the user’s audio status within a video conference. If the user permits, data from the local camera is utilized to determine whether the user is looking at the camera (based on detected gaze direction) and/or to analyze the user’s lip movements to detect whether the user is speaking. If the user is detected as speaking, their audio status is automatically changed to unmuted, and otherwise, the user’s audio is kept muted.

KEYWORDS

- Video conferencing
- Video meeting
- Mute button
- Mute/unmute
- Lip movement
- Face tracking
- Eye gaze
- Speech detection
BACKGROUND

Video conference (VC) user interfaces include a feature that provides users with the option to mute or unmute their audio status during a meeting. A user may forget to mute themselves when they are not speaking, causing background audio to be transmitted. A user may attempt to speak while unmuted, without realizing that they are muted. Such mode errors can interrupt the flow of the VC meeting. In each situation, the other participants need to request the user to correct the state, e.g., by uttering phrases such as “Hey, can you mute yourself? I can hear birds in the background;” “I cannot hear you; you are muted!” etc.

Some VC systems attempt to address these issues by muting everyone who joins a VC by default (or after a particular number of people have joined), thus preventing background noise from new participants from interrupting the meeting. However, such automatic muting, if not noticed by such participants, requires other participants to alert them of their muted state. Such techniques are also inadequate if a user doesn’t realize that the audio mode at their end is likely causing difficulties for others.

DESCRIPTION

This disclosure utilizes one or more user-permitted factors, detected locally at a user device, e.g., that is used to participate in a VC, to automatically adjust the user’s audio status within a video conference (VC). If the user permits, data from the local camera is utilized to determine whether the user is looking at the camera (based on detected gaze direction) and/or to analyze the user’s lip movements to detect whether the user is speaking.

Based on the user’s detected state, a “soft mute” is implemented such that the user’s audio is blocked unless it is detected that the user is speaking. For example, gaze detection can be performed based on face tracking APIs provided locally on the user device. The detected
angle of gaze can then be compared to the position of the camera used for the VC to determine whether they align. If they align, it can be determined that the user is looking at the camera. Further, the user’s lip movements can be analyzed to determine if the lip movements match the user’s detected speech. When the lip movements match the detected speech (indicating that the user is speaking to the VC participants) and the gaze is into their camera, the user’s audio is let into the VC.

Fig. 1: Automatically unmuting user in a video conference based on analyzing lip movements and gaze direction

Fig. 1 illustrates in an example video conference in which the techniques have automatic “soft mute” are implemented. Multiple participants - Adam, Beena, Chloe, and Mario - are in the VC. A “soft mute” is enabled for participants that are not speaking. For example, in Fig. 1(a),
Adam, Beena, and Mario are muted, while Chloe’s audio is let into the VC. In Fig. 1(b), the user Mario is detected as looking at the camera and speaking, and their audio is let into the VC.

With user permission, the “soft mute” techniques work to automatically change the user’s audio status based on the detected lip movements and gaze. As seen in the example of Fig. 1, users Adam and Beena are not unmuted since they are not detected as speaking (gaze is not at the camera, no lip movements detected, etc.), while the user Chloe is unmuted since she is detected as speaking in both Fig. 1(a) and Fig. 1(b).

The VC system can be implemented to support a “hard mute” whereby a user can place themselves on mute. When such an input is received, the soft mute can be turned off, since the likelihood of the user forgetting to unmute is low. Further, the soft mute can be selectively enabled when a user joins the call, based on one or more factors such as the number of participants, amount of background noise, etc.

The described techniques are implemented with specific permission from the meeting participants. Participants are provided with options to turn off one or more of the features during a particular meeting, or a portion of the meeting. The face tracking and speech detection are performed locally on the user device during the VC and no video or speech data is stored unless specifically enabled by the user.

Further to the descriptions above, users are provided with controls allowing the user to make an election as to both if and when systems, programs or features described herein may enable use of user information (e.g., user’s audio and/or video streams, face tracking, user’s location, user profile, user preferences), 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. Thus, the user has 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 utilizes one or more user-permitted factors, detected locally at a user device, to automatically adjust the user’s audio status within a video conference. If the user permits, data from the local camera is utilized to determine whether the user is looking at the camera (based on detected gaze direction) and/or to analyze the user’s lip movements to detect whether the user is speaking. If the user is detected as speaking, their audio status is automatically changed to unmuted, and otherwise, the user’s audio is kept muted.