Improving Audio Conditions Of Audio And Video Conferences

Sara Basson
Dimitri Kanevsky

Follow this and additional works at: http://www.tdcommons.org/dpubs_series

Recommended Citation
Basson, Sara and Kanevsky, Dimitri, "Improving Audio Conditions Of Audio And Video Conferences", Technical Disclosure Commons, (December 28, 2016)
http://www.tdcommons.org/dpubs_series/364

This work is licensed under a Creative Commons Attribution 4.0 License.
This Article is brought to you for free and open access by Technical Disclosure Commons. It has been accepted for inclusion in Defensive Publications Series by an authorized administrator of Technical Disclosure Commons.
IMPROVING AUDIO CONDITIONS OF AUDIO AND VIDEO CONFERENCES

ABSTRACT

Disclosed herein is a system for improving audio quality in video conferencing or teleconferencing involving a small digital figure or a robot associated with a microphone which is remote-controlled by participants. The robot automatically detects poor audio quality that interferes with intelligibility. Using automatic speech recognition, the robot processes and interprets incoming speech at the microphone to establish pre-determined acceptable levels of accuracy in audibility. The system enables participants to signal the robot when experiencing difficulty in understanding the discussion during the course of the meeting. The robot then demonstrates the level of difficulty through gestures in a number of ways that could be entertaining or direct. This system enables the speaker to be aware of his or her clarity levels without disturbing the flow of the presentation and increases the effectiveness of video conferencing.

BACKGROUND

The quality of audio for video conferencing or teleconferencing is critically important for workplace productivity. With distributed team locations, much communication takes place over these long distance channels. In many circumstances, the audio quality of video conferencing or teleconferencing in meetings is not adequate. Some of the audio quality issues are related to technology and audio transmission, where distortions (such as calls breaking up) are problematic. Other issues are related to the participants themselves. For example, participants might have accents that interfere with intelligibility. Intelligibility could be further degraded if speakers speak too quickly, or are too far from the microphones. Objects in conference rooms, especially those on conference room tables, could block microphones or make distracting sounds of their own.
Issues related to degraded audio quality become more significant in the presence of deaf or hard-of-hearing participants, who depend on captioners to transcribe the meeting audio. These captioners are unable to deliver complete caption when any of the speaker characteristics listed above are persistent.

In order to overcome the above-mentioned issues in audio quality in video conferencing or teleconferencing, other participants in the meeting could intimate the speaker with suggestions to move closer to the microphone or speak more slowly. However, this would disrupt the flow of the meeting and many participants would not prefer making such intimations, particularly if the speaker needs to be reminded repeatedly. In other instances, there could be side-chats with other participants electronically messaging the speaker to speak more clearly or to mute a computer that is presenting too much background noise. But these electronic messages would work only if the speaker is looking at the relevant screen to receive the message. There are available conferencing systems that automatically mute every computer participating (other than the main speaker’s), thereby reducing noise generated through computers that interfere with intelligibility of the main speaker. However, this does not address issues related directly to the speaker (e.g., a speaker who is seated far from a microphone or who is speaking too quickly, etc.)

**DESCRIPTION**

This disclosure presents a system and method for improving audio quality in video conferencing or teleconferencing. As illustrated in FIG. 1, the system includes a small digital figure or a robot associated with microphones, which can be remote-controlled by other participants of the meeting.

The robot can process and interpret the speech at the meeting, and automatically detect when the audio quality is low and likely to interfere with intelligibility. Using
automatic speech recognition, the robot can process the incoming speech at that microphone, and establish whether the speech recognition confidence level is unacceptably low. Acceptable levels of accuracy can be pre-determined.

In addition, meeting participants can have a mechanism that signals the robot when they are having difficulty understanding the meeting discussion. The signal can indicate the level of difficulty, from mildly to completely unintelligible.

The robot then demonstrates the difficulty level with gestures in a number of ways that could be entertaining or direct, as illustrated in FIG. 2. Demonstrating difficulty of understanding is manifested using different gestures each time, so that the message is not ignored and remains interesting and compelling. The extent of the robot’s reaction reflects the level of unintelligibility on a scale. For example, with mild unintelligibility, the robot could look confused, and with complete unintelligibility, the robot could do the equivalent of pulling out its hair, covering its face, etc. as shown in FIG. 2.
The system is configured with machine-learning features so that, the robot could record and recognize the most effective expression that changes the behavior of the speaker. Similarly, the robot’s own ability to detect unintelligible speech could improve over time by using its decoding and confidence levels coupled with received feedback from participants on what is or is not intelligible.

Alternatively, a simpler non-autonomous version of the system may involve implementations being entirely based on feedback received from other participants. This version may exclude speech recognition for automatic intelligibility determination and the machine-learning component.

The proposed system and method enable the speaker to be aware of his or her clarity levels without disturbing the presentation flow and also without identifying the person facing the difficulty. Such real-time feedback to the speaker ultimately improves the clarity of the speaker’s presentation. The system is particularly advantageous while having multiple
participants in distributed locations. Thus, the system increases the effectiveness of video conferencing.

In situations in which the systems and methods discussed herein may collect personal information about users, or may make use of personal information (e.g., photos, videos, user data), users are provided with one or more opportunities to control how information is collected about the user and used in one or more described features. A user is provided with control over whether programs or features collect user data (e.g., recognition of a user’s face in a photo or video, information about a user’s social network, user characteristics (age, gender, profession, etc.), social actions or activities, a user’s preferences, content created or submitted by a user, a user’s current geographic location, etc.). 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 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. 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 to a larger region so that a particular location of a user cannot be determined.