Technical Disclosure Commons

October 11, 2016

System For Detecting End Of Speech Utterance For Safe Interruption

Dimitri Kanevsky

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

SYSTEM FOR DETECTING END OF SPEECH UTTERANCE FOR SAFE INTERRUPTION

ABSTRACT

A system and method for detecting end of speech utterance for safe interruption using audio and video processing software or application is disclosed. The system includes hardware that determines the speaker's intonation in real time or with minimal delay and displays this visually on a screen along with a transcript of the speech delivered by the speaker. The speaker's intonation and end of speech could be identified in the system display in real-time as rising and falling of pitch at the end of speech. The system and method could be implemented as a module of any video call or conference facility. The system will provide significant help to people with hearing difficulties so that they participate efficiently in meetings. The method could be implemented in real time without any delays since it could be located in a user's device.

BACKGROUND

There exist text based transcription services for people who do not hear. In these services a text of what is being spoken is displayed in real time for such persons as a text on some digital displays. These transcription services are being provided either by trained transcribers or by automatic speech recognition systems who have access to audio that is being spoken (either because transcribers are present in a meeting or get access to audio remotely, e.g. over telephone).

For people who do not hear and use transcription services during video calls or in meetings, it is very hard to find out when they can start to speak. Normal people can understand from falling intonation that another person is about to finish speaking and they can start to speak. People who cannot hear need to wave hands to get others' attention that they about to start to speak or just interrupt others. There is need to develop efficient automatic means to detect or predict when a speaking person is about to finish speaking and display this information for a

participant with hearing difficulties in the meeting in order that the latter may know when they can start to speak.

Human transcribers could indicate by writing or printing some signs to inform a user that a person is about to finish speaking or already finished. Even if a transcriber prompts a user that a speaker is about to finish his speech, the user still does not know the exact time when the speaker finished his speech. Further, there could also be a few seconds delay between the transcriber input and its display on the user's screen such as due to communication through a network. Therefore, when the user gets a message that a speaker has finished speaking, another person may already have started to speak.

Thus, there is a need for developing an efficient and automatic system to detect or predict when a speaker is about to finish his speech and display this information for a participant in the meeting, so that they could know when they can start their speech.

DESCRIPTION

A system and method for detecting end of speech utterance for safe interruption using audio and video processing software or application is disclosed. The system includes hardware that determines the speaker's intonation in real time or with minimal delay and displays this visually on a screen along with a transcript of the speech delivered by the speaker.

The speaker's intonation and end of speech could be identified by a pitchogram in the system display in real-time as shown in FIG. 1. The rising and falling of the final pitch at the end of speech is indicated by the red circle in the figure. The log time could be compressed to emphasize the current and most recent pitch, in the lower part of the display.
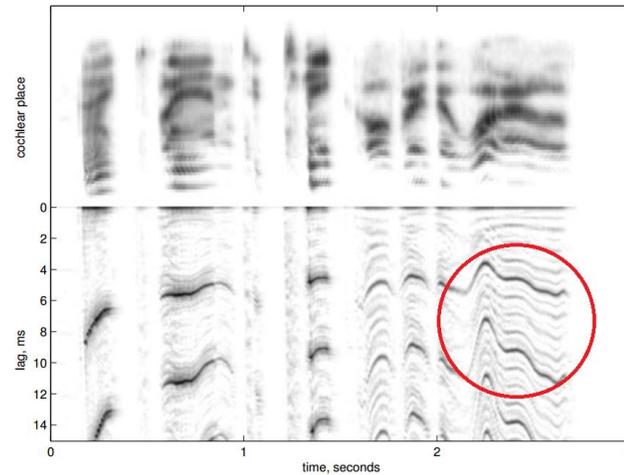
FIG. 1: Pitchogram showing real-time rising and falling of final pitches of a speaker

The pitchogram could scroll in real time with high falling pitch at the end. The picture of the pitch could be shown along with a transcript on a display in the system. It could also be shown as a background feature or in some corner of the display. The pitchogram device could be implemented as a module of any video call or conference facility. This will allow computing and displaying a pitchogram for any speaker in a video call.

The system may further include an automatic detector for rise and fall of pitch. The detector learns to detect expected change of a speaker automatically. The system is based on audio visual characteristics such as pitch and could be provided to map a pitch into a different graphical representation. The user may check that the changes in pitch with such graphical representations are in sync with changes in pitchogram.

The system will significantly help people with hearing difficulties participate efficiently in meetings. The method could be implemented in real time without any delays since it could be located in a user's device.