ACOUSTICS2008/3544
Audio-visual automatic speech recognition: Progress and challenges

Gerasimos Potamianos

IBM T. J. Watson Research Center, RTE 134, Yorktown Heights, NY 10598, USA

The paper overviews recent progress and challenges in a number of audio-visual speech processing technologies with main emphasis on the problem of automatic speech recognition. It is well known that visual channel information can improve automatic speech processing for human-computer interaction. To automatically process and incorporate such information into automatic systems, a number of steps are required that are surprisingly similar accross speech technologies. Crucial above all is the issue of feature representation of visual speech and its robust extraction. In addition, appropriate integration of the audio and visual representations is required, in order to ensure improved performance of the bimodal systems over audio-only baselines. These topics are discussed in detail in the talk, with main emphasis on their application to the speech recognition problem in the challenging environments of automobiles and smart rooms.