



***Electrical Engineering Seminar Series &
Dallas Chapter of IEEE Signal Processing Society Present***

Audio-Visual Automatic Speech Recognition and Related Bimodal Speech Technologies: A Review of the State-of-the-Art and Open Problems

**Dr. Gerasimos Potamianos
National Center for Scientific Research, Greece**

**1pm, Wed, March 3, 2010
ECSS 2.102 (TI Auditorium)**

The presentation will provide an overview of the main research achievements and the state-of-the-art in the area of audio-visual speech processing, mainly focusing in the area of audio-visual automatic speech recognition. The topic has been of interest in the speech research community due to the potential of increased robustness to acoustic noise that the visual modality holds. Nevertheless, significant challenges remain that have hindered practical applications of the technology – most notably difficulties with visual speech information extraction and audio-visual fusion algorithms that remain robust to the audio-visual environment variability inherent in practical, unconstrained interaction scenarios and audio-visual data sources, for example multi-party interaction in smart spaces, broadcast news, etc. These challenges are also shared across a number of interesting audio-visual speech technologies beyond the core speech recognition problem, where the visual modality has the potential to resolve ambiguity inherent in the audio signal alone; for example, speech enhancement, speech activity detection, speaker recognition, and others.

Gerasimos Potamianos received his PhD degree from the Johns Hopkins University in 1994. From 1994 to 1996, he was a Postdoctoral Fellow with the Center for Language and Speech Processing (CLSP). From 1996 to 1999, he was a Senior Member of Technical Staff with the then Speech and Image Processing Services Laboratory at AT&T Labs-Research, in Murray Hill and Florham Park, New Jersey. In September 1999, he joined the Human Language Technologies Department (currently Multilingual Analytics and User Technologies) at the IBM Thomas J. Watson Research Center, where he eventually became Manager of the Multimodal Conversational Solutions Department. In October 2008, he joined the Telecommunications and Networks Laboratory of the Institute of Computer Science at FORTH, as an Associate Researcher, where he continued his research activities in multimodal and multisensory processing of speech with emphasis in ambient intelligence environments, within the Ambient Intelligence Program of the Institute. In April 2009, he joined the Software and Knowledge Engineering Laboratory (SKEL) of the Institute of Informatics and Telecommunications (IIT) at the National Center for Scientific Research (NCSR), "Demokritos" as a Research Director.

For more information on the Dallas Chapter and directions to UTD, please refer to <http://www.utdallas.edu/~kehtar/ieee-sp>