

**Call for Papers**  
**Special Issue of**  
**The IEEE Transactions on Audio, Speech and Language Processing**  
**Multimodal Processing in Speech-based Interactions**

Recently there has been increasing research interest to jointly process audio and visual information related to human activities, and to extend the technological developments in individual modalities for human-computer interaction to include multimodal processing in order to improve robustness and naturalness. For example, we have witnessed significant research activity devoted to extending traditional, unimodal speech recognition to audio-visual speech recognition by incorporating the speaker's lip motion; text-to-speech synthesis has been migrating towards audio-visual speech synthesis involving head, facial, and lip motions; speech databases for technology evaluation have evolved from single-modality broadcast news type audio towards multimodal recordings of complex human interactions in contexts such as meeting rooms and in support of a multitude of far-field multimodal technologies; and speaker authentication has been migrating towards multimodality by incorporating biometric traits such as facial images, videos, and fingerprints. Furthermore, we have witnessed emergence of major research programs in the area such as the European Union funded efforts on multimodal interfaces and interaction, as well as multimodal technology evaluation campaigns by NIST and the VACE community (Rich Transcription, CLEAR, etc).

Joint processing of audio, visual, pen, and other gestural input offers a means to improve naturalness and robustness of user interfaces that can automatically recognize human identity, intent, and activity in pervasive computing environments. Illustrative technologies include speaker and speech recognition, person localization, source separation, media synthesis, and media content mining. A critical factor that contributes to the effectiveness of multimodal processing in speech-based interactions is the robust integration or fusion of information from multiple modalities. This special issue invites researchers to submit original and unpublished work that concentrates on the multi-disciplinary field of multimodal processing of speech-based interactions. We solicit papers including, but not limited to, the following topics:

- (a) Multimodal speaker recognition (identification and verification);
- (b) Audio-visual speech recognition;
- (c) Audio-visual speech synthesis;
- (d) Multimodal fusion methodologies;
- (e) Audio-visual open microphone engagement;
- (f) Multimodal processing in media retrieval;
- (g) Multimodal corpora and resources;
- (h) Multimodality in spoken dialog interfaces;
- (i) User-centered and adaptive multimodal interfaces

Proposed Schedule:

Submission deadline: 15 January, 2008

Notification of acceptance: 15 May, 2008

Final manuscript due: 1 July, 2008

Tentative publication date: 1 October, 2008.

Guest Editors:

Dr. Gerasimos Potamianos, IBM T.J. Watson Research Center, [gpotam@us.ibm.com](mailto:gpotam@us.ibm.com)

Dr. Helen Meng, Chinese University of Hong Kong, [hmmeng@se.cuhk.edu.hk](mailto:hmmeng@se.cuhk.edu.hk)

Dr. Sharon Oviatt, Adapx Inc., [oviatt@adapx.com](mailto:oviatt@adapx.com)

Dr. Gerhard Rigoll, Technical University of Munich, [rigoll@tum.de](mailto:rigoll@tum.de)