

SPECIAL SESSIONS

The ICSLP-2002 organizing committee identified early the need to support a wide range of interdisciplinary topics in speech processing, speech science, and language technology. To encourage lively discussion and interactions, the following special sessions were selected and are taking place during the conference week.

Special Session #1

Integration of Speech Technology in Language Learning:

Organizer: Philippe Delcloque (Manchester Science Enterprise Centre, UMIST)

Topics cover a range of contributions from leading experts and researchers involved in the integration of Speech Technology in (Language) Learning. Contributions will cover most or all of the thematic aspects in this emerging field such as ASR, speech synthesis, visualization of segmental and suprasegmental speech features, use of animated agents in learning, talking heads, pedagogical and computational integration issues, use of ST on the internet, use of speech in therapy and the assistive interface and general use of ST in learning systems and virtual environments (including VR).

Special Session #2

Call Classification and Routing

Organizer: Gokhan Tur & Allen Gorin (AT&T Labs - Research)

In the next generation of automated services, customers will talk with machines, expressing what they want in spoken natural language. By natural, we mean that the machine understands and acts upon what people actually say, in contrast to what one would like them to say. This technology shifts the burden from users to the machine, greatly increasing peoples' capability and willingness to use automation. This is in contrast to the current generation of voice interfaces, where users must speak a small set of words in some fixed order. To this end, a first step is classifying customer calls in order to route them accordingly. This special session focuses on the research related to call classification technologies.

Special Session #3

Speech-to-Speech Translation: Algorithms and Systems

Organizer: Yuqing Gao (IBM T. J. Watson Research Center) & Alex Waibel (Carnegie Mellon University & University Karlsruhe)

Construction of robust speech-to-speech translation systems is clearly extremely complex, involving research in Automatic Speech Recognition, Text-to-Speech, Machine Translation, Natural Language Understanding and Natural Language Generation. Although substantial progress in each of these components individually has been made, simply integrating individual components to produce S2S systems is not sufficient to produce acceptable results. Our goal for this special session is to bring together various researchers in the field to present the current state-of-the-art on speech-to-speech translation and to discuss the challenges involved in building a functioning high performance system.

Special Session #4
Aurora: Speech Recognition in Noise
Organizer: David Pearce (Motorola Labs, United Kingdom)

At the Aurora special session researchers will present leading edge algorithms for noise robust recognition and their results measured on the same databases. Evaluating the many different techniques on the same speech material provides a great way to compare the effectiveness of the different algorithms and measuring scientific progress. While the Aurora 2 databases use the controlled addition of noise to clean speech, the Aurora 3 databases are collected in a real-world environment of the car and cover for 4 languages. There are 2 oral sessions as well as a poster session covering a fascinating range of alternative techniques. We hope these both capture the state-of-the-art and will simulate a lot of new ideas.

Special Session #5
Issues in Audio-Visual Spoken Language Processing
Organizer: Lynne Bernstein (UCLA)

The area of audiovisual speech processing has attracted a loyal following from a wide range of perspectives. Work in the area has benefited greatly from multidisciplinary approaches. This session inventories many of the ongoing lines of inquiry, from processing at the level of the cerebral cortex to processing for automatic speech recognition and for speech synthesis. Clinical, second language, and developmental issues are raised. The lecturers have been selected to highlight the view that human speech communication is audiovisual. The special session is intended for a wide audience.

Special Session #6
Distributed Multimodal Dialog Management using Internet Technologies
Organizer: Kuansan Wang, Alex Acero & Hsiao-wuen Hon (Microsoft)

Although both Internet and personal computers had long existed, it was not until an intuitive user interface was introduced in 1990s did we witness a flurry of revolutionary developments that brought the Internet to the mainstream. Can speech, being a natural means of communications, play a pivotal role in the next revolution as the internet becomes more mobile and multimodal? The special session offers a discussion on this question.

Special Session #7
Prosody and Speech Recognition
Organizer: Keikichi Hirose (University of Tokyo)

As is well known, prosodic features play an important role in the human process of speech perception. However, their use in speech recognition is quite limited. As compared to segmental features, prosodic features can be said to be unique to spoken language; segmental features have a good correspondence with phonetic sequences, thus with written language, while prosodic events are not directly included in it. Therefore, when extending the current speech technologies from carefully prepared speech to spontaneous speech, prosodic features should play a more integral role in automatic speech recognition.