Special Sessions

TuC.SS – Synthesis of Singing Challenge

Tuesday, August 28, 2007, Astrid Plaza hotel, Room Scala 1

Session Chair: Gerrit Bloothooft, Utrecht University

Format: oral

Singing is perhaps the most expressive usage of human voice and speech. An excellent singer, whether in classical opera, musical, pop, folk music, or any other style, can express a message and emotion so intensely that it moves and delights a wide audience. Synthesizing singing may be considered therefore as the ultimate challenge to our understanding and modeling of human voice. Participants contributed a short paper, including an audio demonstration of their system, and they also will produce their own version of the Synthesizer song, composed for Interspeech 2007. The contributions will be presented in oral session format and commented by a panel consisting of synthesis experts and singers, and the audience.

TuD.SS – Speech and Audio Processing for Intelligent Environments

Tuesday, August 28, 2007, Astrid Plaza hotel, Room Scala 1

Session Chairs: Reinhold Haeb-Umbach, University of Paderborn, Germany and Zheng-Hua Tan, Aalborg University, Denmark

Format: tutorial + poster

Ambient Intelligence (AmI) describes the vision of technology that is invisible, embedded in our surroundings and present whenever we need it. Interacting with it should be simple and effortless. Since the early days of this computing and interaction paradigm speech has been considered a major building block of AmI. The purpose of speech and audio processing is twofold: speech as an input/output modality, preferably in cooperation with other modalities like gesture, and speech and acoustic signals in general as a source of context information to be utilized in systems that are context-aware, personalized, adaptive, or even anticipatory. With two invited tutorial lectures and eight contributed posters, this special session is to give an overview of major achievements, but also to highlight major challenges. Does state-of-the-art of speech and audio processing meet the high expectations expressed in the scenarios of AmI, will it ever do? What are the perspectives and promising concepts for the future?

WeB.SS – Structure-Based and Template-Based Automatic Speech Recognition Comparing parametric and non-parametric approaches

Wednesday, August 29, 2007, Astrid Plaza hotel, Room Scala 1

Session Chairs: Li Deng, Microsoft, USA and Helmer Strik, Radboud University Nijmegen, the Netherlands

Format: poster + oral + panel

One prominent weakness in current hidden Markov modeling (HMM) in automatic speech recognition today is the handicap in representing long-span temporal dependency in the acoustic feature sequence of speech. Furthermore, non-verbal information also plays an important role in human speech recognition which the HMM framework has not attempted to address directly. In this special session, two categories of approaches to address the above weaknesses of HMMs are considered. The first, parametric, structure-based approach establishes mathematical models for stochastic trajectories/segments of speech utterances. The second, non-parametric and template-based approach has also been called exemplar-based or data-driven and involves direct exploitation of speech feature trajectories in the training data without any modeling assumptions. With eight poster presentations, three oral presentations and a panel discussion, this special session is to bring together researchers who have special interest in novel techniques for acoustic modeling in speech recognition.

WeC.SS – Objective assessment of voice and speech quality

Wednesday, August 29, 2007, Astrid Plaza hotel, Room Scala 1

Session Chairs: Yannis Stylianou, University of Crete and Hugo Van hamme, Katholieke Universiteit Leuven

Format: poster + demonstrations

In the clinical practice of treatment of voice and speech disorders as well as in speech therapy for patients with stammer, with pronunciation problems, with dyslexia, etc., speech is evaluated by a human being. Within the speech technology research community there is a growing awareness that speech science and engineering can provide answers to make such assessments more objective.
Besides the contributed papers which are presented as posters, the session also includes announcements of international and national programs in the field and a demonstration of related products and realizations.

WeD.SS – Speech and language technology for less-resourced languages

Wednesday, August 29, 2007, Astrid Plaza hotel, Room Scala 1

Session Chairs: Briony Williams, University of Wales, UK, Mikel Forcada, Universitat d'Alacant, Spain, and Kepa Sarasola, University of the Basque Country, Spain

Format: poster + demonstrations

Speech and language technology researchers who work on less-resourced languages often have very limited access to funding, equipment and software. This makes it all the more important for them to come together to share best practice, in order to avoid a duplication of effort. With eighteen poster and demo presentations, this special session will be devoted to speech and language technology for less-resourced languages. In view of the limited resources available to the targeted researchers, there will be a particular emphasis on "free" software, which may be either open-source or closed-source. However, contributions using commercial and home made software are also included.

ThB.SS – Speech Recognition by Automatic Speech Attribute Transcription

Thursday, August 30, 2007, Astrid Plaza hotel, Room Scala 1

Session Chair: Chin-Hui Lee, Georgia Institute of Technology, USA

Format: tutorial + oral

It is well-known that the speech signal contains a rich set of information that facilitates human auditory perception and communication, beyond a simple linguistic interpretation of word sequences. It has long been postulated that human beings determine the linguistic identity of a sound based on detected evidence that exists at various levels of the speech knowledge hierarchy. In order to bridge the performance gap between automatic speech recognition (ASR) systems and human speech recognition, the narrow notion of speech-to-text in ASR has to be expanded to incorporate all related human information "hidden" in speech utterances. This collection of information includes a set of fundamental speech sounds and their linguistic interpretations, a speaker profile that encompasses gender, accent and other speaker characteristics, the speaking environment that describes the interaction between speech and acoustics, the emotional state of the speaker, etc. With one short tutorial and five oral presentations, this special session focuses on this set of speech information, which we collectively refer to as speech attributes.

ThC.SS – Multilingualism in Speech and Language Processing

Thursday, August 30, 2007, Astrid Plaza hotel, Room Scala 1

Session Chair: Jan Verhasselt, Tele Atlas

Format: tutorial + oral

With two invited tutorials and three contributed oral presentations, the main objectives of this special session are to offer a mature view on the nature and relevance of multilingualism issues in speech and language processing systems; to discuss the technological challenges involved and the solutions offered thus far and to anticipate on what kind of further research is needed to come up with more satisfactory solutions.

ThD.SS – Novel techniques for the NATO non-native Air Traffic Control and HIWIRE cockpit databases

Thursday, August 30, 2007, Room Astrid Scala 1

Session Chairs: David van Leeuwen, TNO Human Factors and Alex Potamianos, Technical University of Crete

Format: oral

The NATO research task group on speech and language technology, IST-RTG013, has collected a corpus of military Air Traffic Control communication in Belgian air space. This speech material consists predominantly of non-native English speech, under varying noise and channel conditions. The HIWIRE cockpit database, consists of English utterances (Direct Voice Input domain) uttered by non-native speakers. Both databases have been distributed to interested research groups together with proposed training and testing scripts for each research track for these databases (e.g., robust non-native recognition and non-native adaptation). In six oral presentations and a panel discussion, this special session will discuss research results on feature extraction, feature normalization, pronunciation modeling, acoustic modeling and adaptation on these challenging databases as well as the usefulness and effectiveness of sharing project related databases across the speech technology community.

FrB.SS – Machine Learning for Spoken Dialogue Systems

Friday, August 31, 2007, Astrid Plaza hotel, Room Scala 1

Session Chairs: Oliver Lemon, Edinburgh University and Olivier Pietquin, Supélec - IMS Research Group

Format: oral

During the last decade, research in the field of Spoken Dialogue Systems (SDS) has experienced increasing growth. Yet, the design and optimization of SDS does not only consist of combining speech and language processing systems such as Automatic Speech Recognition (ASR), parsers, Natural Language Generation (NLG), and Text-to-Speech (TTS) synthesis systems. It also requires the development of dialogue strategies taking at least into account the performances of these subsystems (and others), the nature of the task (e.g. form filling, tutoring, robot control, or database querying), and the user's behaviour (e.g. cooperativeness, expertise). In addition, new statistical learning techniques are emerging for training and optimizing speech recognition, parsing, and generation and synthesis in spoken dialogue systems, depending on various definitions of context. Automatic learning of optimal dialogue strategies is currently a leading domain of research. The purpose of this special session with one tutorial and five contributed oral presentations is to offer the opportunity to share ideas and have constructive discussions in a single, focussed, special conference session.