INTERSPEECH 2014

CELEBRATING THE DIVERSITY OF SPOKEN LANGUAGES

14-18 SEPTEMBER 2014
SINGAPORE
MAX ATRIA@SINGAPORE EXPO

Special Sessions

HTTP://WWW.INTERSPEECH2014.ORG

Special Sessions

SP 1: Open Domain Situated Conversational Interaction

Monday, 15 September 2014, 11:00 - 13:00; Peridot 206, Level 2, MAX Atria

Robust conversational systems have the potential to revolutionize our interactions with computers. Building on decades of academic and industrial research, we now talk to our computers, phones, and entertainment systems on a daily basis. However, current technology typically limits conversational interactions to a few narrow domains/topics (e.g., weather, traffic, restaurants). Users increasingly want the ability to converse with their devices over broad web-scale content. Finding something on your PC or the web should be as simple as having a conversation. A promising approach to address this problem is situated conversational interaction. The approach leverages the situation and/or context of the conversation to improve system accuracy and effectiveness. Sources of context include visual content being displayed to the user, geo-location, prior interactions, multi-modal interactions (e.g., gesture, eye gaze), and the conversation itself. For example, while a user is reading a news article on their tablet PC, they initiate a conversation to dig deeper on a particular topic. Or a user is reading a map and wants to learn more about the history of events at mile marker 121. Or a gamer wants to interact with a game's characters to find the next clue in their quest. All of these interactions are situated – rich context is available to the system as a source of priors/constraints on what the user is likely to say. This special session will provide a forum to discuss research progress in open domain situated conversational interactions. Topics of the session will include:

- Situated context in spoken dialog systems
- Visual/dialog/personal/geo situated context
- Inferred context through interpretation and reasoning
- Open domain spoken dialog systems
- Open domain spoken/natural language understanding and generation
- Open domain semantic interpretation
- Open domain dialog management (large-scale belief state/policy)
- Conversational interactions
- Multi-modal inputs in situated open domains (speech/text + gesture, touch, eye gaze)
- Multi-human situated interactions

Organizers:
Larry Heck (larry@ieee.org), Microsoft Research
Dilek Hakkani-Tür (dilek@ieee.org), Microsoft Research
Gokhan Tur (gokhan@ieee.org), Microsoft Research
Steve Young (sjy@eng.cam.ac.uk), Cambridge University

SP 2: INTERSPEECH 2014 Computational Paralinguistics ChallengE (ComParE)

Monday, 15 September 2014, 14:30 - 16:30; Peridot 206, Level 2, MAX Atria
Monday, 15 September 2014, 17:00 - 19:00; Peridot 206, Level 2, MAX Atria

The INTERSPEECH 2014 Computational Paralinguistics ChallengE (ComParE) is an open challenge dealing with speaker characteristics as manifested in their speech signal's acoustic properties. This year, it introduces new tasks by the Cognitive Load Sub-Challenge and the Physical Load Sub-Challenge. For these tasks, the COGNITIVE-LOAD WITH SPEECH AND EGG database (CLSE) and the MUNICH BIOVOICE CORPUS (MBC) with high diversity of speakers and different languages covered (Australian English and German) are provided by the organizers. The corpora contain fully realistic data in challenging acoustic conditions and feature rich annotation such as speaker meta-data. They are given with distinct definitions of test, development, and training partitions; speaker independence is guaranteed as needed in most real-life settings. Benchmark results of the most popular approaches are provided as in the years before. The transcription of the train and development sets is known. All Sub-Challenges allow contributors to find their own features with their own machine learning algorithm. However, a standard feature set has been provided per corpus that could be used. Participants had to stick to the definition of training, development, and test sets. They may report on results obtained on the development set, but had only five trials to upload their results on the test sets, whose labels are unknown to them. Participants had to submit a paper presenting the results

that underwent peer-review and had to be accepted for the conference in order to participate in the Challenge. The results of the Challenge are presented in a Special Session (two time slots) at INTERSPEECH 2014 in Singapore.

Organizers:
Björn Schuller (schuller@IEEE.org), Imperial College London / Technische Universität München
Stefan Steidl (stefan.steidl@fau.de), Friedrich-Alexander-University
Anton Batliner (batliner@cs.fau.de), Technische Universität München / Friedrich-Alexander-University
Jarek Krajweski (krajewsk@uni-wuppertal.de), Bergische Universität Wuppertal
Julien Epps (j.epps@unsw.edu.au), The University of New South Wales / National ICT Australia

SP 3: Speech Technologies for Ambient Assisted Living

Tuesday, 16 September 2014, 10:00 - 12:00; Peridot 206, Level 2, MAX Atria

This special session focuses on the use of speech technologies for ambient assisted living, the creation of smart spaces and intelligent companions that can preserve independence and executive function, social communication and security of people with special needs. Currently, speech interfaces for assistive technologies remains underutilized despite its potential to replace or augment obtrusive and sometimes outright inaccessible conventional computer interfaces. Moreover in a smart home context, efficiency of speech interfaces can be supported by a number of concurrent information sources (e.g., wearable sensors, home automation sensors), enabling multimodal communication. In practice, daily hand-free usage of speech interfaces remains limited due to challenging real-world conditions, and because conventional speech interfaces can have difficulty with the atypical speech of many users. This, in turn, can be attributed to the lack of abundant speech material, and the limited adaptation to the user of these systems. Taking up the challenges of this domain requires a multidisciplinary approach to define the user's needs, record corpora in realistic usage conditions, develop speech interfaces that are robust to both environment and user's characteristics and are able to adapt to specific users. This special session will bring together researchers in speech and audio technologies with people from the ambient assisted living and assistive technologies communities to meet and foster awareness between members of either community, discuss problems, techniques and datasets, and perhaps initiate common projects.

Organizers:
Heidi Christensen (h.christensen @dcs.shef.ac.uk), University of Sheffield
Jort F. Gemmeke (jgemmeke@amadana.nl), KU Leuven
François Portet (francois.portet@imag.fr), Laboratoire d'Informatique de Grenoble
Frank Rudzicz (frank@cs.toronto.edu), University of Toronto
Michel Vacher (michel.vacher@imag.fr), Laboratoire d'Informatique de Grenoble

SP 4: Text-Dependent Speaker Verification with Short Utterances

Tuesday, 16 September 2014, 15:00 - 17:00; Peridot 206, Level 2, MAX Atria

In recent years, speaker verification engines have reached maturity and have been deployed in commercial applications. Ergonomics of such applications is especially demanding and imposes a drastic limitation in terms of speech duration during authentication. A well known tactic to address the problem of lack of data, due to short duration, is using text-dependency. However, recent breakthroughs achieved in the context of text-independent speaker verification in terms of accuracy and robustness do not benefit text-dependent applications. Indeed, large development data required by the recent approaches is not available in the text-dependent context. The purpose of this special session is to gather the research efforts from both academia and industry toward a common goal of establishing a new baseline and explore new directions for text-dependent speaker verification. The focus of the session is on robustness with respect to duration and modeling of lexical information. To support the development and evaluation of text-dependent speaker verification technologies, the Institute for Infocomm Research, A*STAR, Singapore, has recently released the RSR2015 database, including 150 hours of data recorded from 300 speakers.
Further details are available at: http://www1.i2r.a-star.edu.sg/~kalee/is2014/tdspk.html

Organizers:
*Anthony Larcher (alarcher@i2r.a-star.edu.sg), Institute for Infocomm Research, A*STAR, Singapore*
Hagai Aronowitz (hagaia@il.ibm.com), IBM Research – Haifa
*Kong Aik Lee (kalee@i2r.a-star.edu.sg), Institute for Infocomm Research, A*STAR, Singapore*
Patrick Kenny (patrick.kenny@crim.ca), CRIM – Montréal

SP 5: Phase Importance in Speech Processing Applications

Wednesday, 17 September 2014, 10:00 - 12:00; Peridot 206, Level 2, MAX Atria

In the past decades, the amplitude of the speech spectrum has been considered to be the most important feature for speech processing applications and phase of the speech signal has received less attention. Recently, several findings justify the phase importance in speech and audio processing communities. The importance of phase estimation along with amplitude estimation in speech enhancement, complementary phase-based features in speech and speaker recognition and phase-aware acoustic modeling of the environment are the most prominent reported works scattered in different communities of speech and audio processing. These examples suggest that incorporating the phase information can push the limits of the state-of-the-art phase-independent solutions employed for long in different aspects of audio and speech signal processing. This special session aims to explore the recent advances and methodologies to exploit the knowledge of signal phase information in different aspects of speech processing. Without a dedicated effort to bring researchers from different communities, a quick advance in investigation towards the phase usefulness in speech processing applications is difficult to achieve. Therefore, as the first step in this direction, we aim to promote the "phase-aware speech and audio signal processing" to form a community of researchers to organize the next steps. Our initiative is to unify these efforts to better understand the pros and cons of using phase and the degree of feasibility for phase estimation/enhancement in different areas of speech processing including: speech enhancement, speech separation, speech quality estimation, speech and speaker recognition, voice transformation and speech analysis and synthesis. The goal is to promote the importance of the phase-based signal processing and studying its importance and sharing interesting findings from different speech processing applications.

Organizers:
Pejman Mowlaee (pejman.mowlaee@tugraz.at), Graz University of Technology
Rahim Saeidi (rahim.saeidi@uef.fi), University of Eastern Finland
Yannis Stylianou (yannis@csd.uoc.gr), Toshiba Labs Cambridge UK / University of Crete

SP 6: Deep Neural Networks for Speech Generation and Synthesis

Wednesday, 17 September 2014, 13:30 - 15:30; Peridot 206, Level 2, MAX Atria
Wednesday, 17 September 2014, 16:00 - 18:00; Peridot 206, Level 2, MAX Atria

This special session aims to bring together researchers who work actively on deep neural network (DNN) for speech research, particularly, in generation and synthesis, to promote and to understand the state-of-art DNN research in statistical learning and compare results with the parametric HMM-GMM model based TTS synthesis, generation, and conversion. DNN, with its neuron-like structure, can simulate human speech production system in a layered, hierarchical, nonlinear and self-organized network. It can transform linguistic text information into intermediate semantic, phonetic and prosodic content and finally generate speech waveforms. Many possible neural network architectures or typologies exist, e.g. feed-forward NN with multiple hidden layers, stacked RBM or CRBM, Recurrent Neural Net (RNN), which have been used to speech/image recognition and other applications. We would like to use this special session as a forum to present updated results in the research frontiers, algorithm development and application scenarios.

Organizers:
Yao Qian (yaoqian@microsoft.com), Microsoft Research
Frank K. Soong (frankkps@microsoft.com), Microsoft Research