Special Sessions

1. INTERSPEECH 2018 Computational Paralinguistics ChallengE (ComParE): Atypical and Self-Assessed Affect, Crying & Heart Beats

September 3, 2018, 14:00-16:00 and 16:30-18:30, MR 1.01-1.02, HICC, Hyderabad

Organizers
Björn Schuller (bjoern.schuller@imperial.ac.uk)
Stefan Steidl (stefan.steidl@fau.de)
Anton Batliner (batliner@cs.fau.de)
Peter Marschik (peter.marschik@medunigraz.at)
Harald Baumeister (harald.baumeister@uni-ulm.de)
Fengquan Dong (fengquan.dong@foxmail.com)

The Interspeech 2018 Computational Paralinguistics ChallengE (ComParE) is an open Challenge dealing with states and traits of speakers as manifested in their speech signal's acoustic properties. There have so far been nine consecutive Challenges at INTERSPEECH since 2009 (cf. the Challenge series' repository at http://www.compare.openaudio.eu), but there still exists a multiplicity of not yet covered, but highly relevant paralinguistic phenomena. Thus, in this year's 10th anniversary edition, we introduce four new tasks. The following Sub-Challenges are addressed:

- In the Atypical Affect Sub-Challenge, emotion of disabled speakers is to be recognised.
- In the Self-Assessed Affect Sub-Challenge, self-assessed affect shall be determined.
- In the Crying Sub-Challenge, mood-related types of infant vocalisation have to be classified.
- In the Heart Beats Sub-Challenge, types of Heart Beat Sounds need to be distinguished.

All Sub-Challenges allow contributors to find their own features with their own machine learning algorithm. However, a standard feature set and tools will be provided that may be used. Participants will have to stick to the definition of training, development, and test sets as given. They may report results obtained on the development sets, but have only five trials to upload their results on the test set per Sub-Challenge, whose labels are unknown to them. Each participation has to be accompanied by a paper presenting the results that undergoes the normal Interspeech peer-review and has to be accepted for the conference in order to participate in the Challenge. The organisers preserve the right to re-evaluate the findings, but will not participate themselves in the Challenge.

In these respects, the INTERSPEECH 2018 Computational Paralinguistics challengE (ComParE) shall help bridging the gap between excellent research on paralinguistic information in spoken language and low compatibility of results. We encourage both – contributions aiming at highest performance w.r.t. the baselines provided by the organisers, and contributions aiming at finding new and interesting insights w.r.t. these data. Overall, contributions using the provided or equivalent data are sought for (but not limited to):

- Participation in a Sub-Challenge
- Contributions focussing on Computational Paralinguistics centred around the Challenge topics

The results of the Challenge will be presented at Interspeech 2018 in Hyderabad, India. Prizes will be awarded to the Sub-Challenge winners. If you are interested and planning to participate in INTERSPEECH 2018 ComParE, or if you want to be kept informed about the Challenge, please send the organisers an e-mail (steidl@xxxxxxxxx) to indicate your interest and visit the homepage: http://emotion-research.net/sigs/speech-sig/is18-compare

2. The First DIHARD Speech Diarization Challenge

September 5, 2018, 17:00-19:00, MR 1.01-1.02, HICC, Hyderabad

Organizers

Kenneth Church (kenneth.ward.church@gmail.com)
Christopher Cieri (ccieri@ldc.upenn.edu)
Alejandrina Cristia (alecristia@gmail.com)
Jun Du (jundu@ustc.edu.cn)
Sriram Ganapathy (sriram.iisc@gmail.com)
Mark Liberman (markyliberman@gmail.com)
Neville Ryant (nryant@ldc.upenn.edu)

DIHARD is a new annual challenge focusing on "hard" diarization; that is, speech diarization for challenging corpora where there is an expectation that the current state-of-the-art will fare poorly, including, but not limited to:

- clinical interviews
- extended child language acquisition recordings
- web videos
- "speech in the wild" (e.g., recordings in restaurants)

Because performance of diarization is highly dependent on the quality of the speech activity detection (SAD) system used, the challenge will have two tracks:

Track 1: diarization beginning from gold speech segmentation
Track 2: diarization from scratch

For more details, please see the challenge site at https://coml.lscp.ens.fr/dihard/index.html

3. Novel Paradigms for Direct Synthesis based on Speech-Related Biosignals

September 6, 2018, 10:00-12:00, MR G.03-G.04, HICC, Hyderabad

Organizers:

Lorenz Diener (lorenz.diener@uni-bremen.de)
Jose Gonzalez (jgonzalez@lcc.uma.es)
Tanja Schultz (tanja.schultz@uni-bremen.de)

Speech is a very rich and complex process, the acoustic signal being just one of the biosignals resulting from it. In the last few years, the automatic processing of these speech-related biosignals has become an active area of research within the speech community. This special session aims to foster research on one emerging area that is growing within the field of silent speech: Direct synthesis. Direct synthesis refers to the generation of speech directly from speech-related biosignals (e.g. ultrasound, EMG, EMA, PMA, lip reading video, BCI) without an intermediate recognition step. This has been made possible by recent developments in supervised machine learning techniques and the availability of high-resolution biosensors. Furthermore, the availability of low-cost computing devices has made something possible that was unthinkable 20 years ago: the generation of audible speech from speech-related biosignals in real time. With this special session, we aim at bringing together researchers working on direct synthesis and related topics to foster work towards direct synthesis toolkits and datasets and to highlight and discuss common challenges and solutions in this emerging research area.

Web-site: http://www.uni-bremen.de/csl/forschung/lautlose-sprachkommunikation/special-session-novel-paradigms-for-direct-synthesis-based-on-speech-related-biosignals.html

4. Speech Recognition for Indian Languages

September 4, 2018, 10:00-12:00, MR 1.01-1.02, HICC, Hyderabad

Organizers:
Pedro Moreno (pedro@google.com)
Eugene Weinstein (weinstein@google.com)
Sarah Abu Sharkh (sarah@google.com)
Haruko Ishikawa (ishikawa@google.com)
Daan van Esch (dvanesch@google.com)
Rita Singh (rsingh@cs.cmu.edu)
Preethi Jyothi (pjyothi@cse.iitb.ac.in)

Indian languages offer a multitude of challenges that are not observed as widely in languages elsewhere, e.g. code-switching ("Hinglish"). However, research on ASR for Indian languages remains relatively scarce compared to English, European languages, and East-Asian languages. We will invite participants of this Special Session to build their own ASR systems for Indian languages, in the broad sense of the term. You may choose which Indian languages to build systems for, and what data sets to use. In addition to the challenge around building ASR systems, we will also allocate time in the schedule for a show-and-tell: some participants may prefer using an existing ASR system and then building a voice-enabled app. We believe that this Special Session would provide an important boost to academic research on building ASR for Indian languages

Web-site: https://sites.google.com/view/interspeech2018-ss1

5. Deep Neural Networks: How Can We Interpret What They Learned?

September 4, 2018, 14:30-16:30, MR 1.01-1.02, HICC, Hyderabad

Organizers:
Louis ten Bosch (l.tenbosch@let.ru.nl)
Hugo Van hamme (hugo.vanhamme@esat.kuleuven.be)
Lou Boves (l.boves@ru.nl)

Everybody active in the speech technology area witnesses the advent of deep learning techniques, in particular the use of Deep Neural Nets. DNNs are now being applied in many types of automatic speech recognition systems. The success of DNNs strongly suggests that the parameter set of a trained DNN reflects relevant structure in the training data, but it is not clear how one can reveal this structure. In this session we want to address this issue.

We welcome papers that specifically address one or more of the leading questions listed below, e.g. by aiming at (phonetic or linguistic) interpretations of the representations at the layers of DNNs, or by attempting to use phonetic/linguistic knowledge to guide the design and training of DNNs.

For papers in this special session, leading questions are:
[1] How does a trained DNN encapsulate the structure that exists in a data set ?
[2] How can we visualize this information ?
[3] How can we learn from a DNN, i.e., how can the information in a DNN be used to sharpen our insights ?
[4] What type of knowledge can be encoded in a DNN ?
[5] Can understanding the information encoded in a DNN be used as a guidance in designing and training more powerful networks ?
[6] What architectures and training techniques are most amenable to interpretations?

For more information please contact the organizers. http://cls.ru.nl/interspeech2018

6. Low Resource Speech Recognition Challenge For Indian Languages

September 6, 2018, 10:00-12:00, MR 1.01-1.02, HICC, Hyderabad

Organizers:
Kalika Bali (kalikab@microsoft.com)
Krishna Doss Mohan (krishna.doss@microsoft.com)
Rupesh Kumar Mehta (rupesh.mehta@microsoft.com)
Niranjan Nayak (niranjan@microsoft.com)
Sunayana Sitaram (t-susita@microsoft.com)
Radhakrishnan Srikanth (rsrikan@microsoft.com)

Most languages in the world lack the amount of text, speech and linguistic resources required to build large Deep Neural Network (DNN) based models. However, there have been many advances in DNN architectures, cross-lingual and multilingual speech processing techniques, and approaches incorporating linguistic knowledge into machine-learning based models, that can help in building systems for low resource languages. In this challenge, we would like to focus on building Automatic Speech Recognition (ASR) systems for Indian languages with constraints on the data available for Acoustic Modeling and Language Modeling. For more details on the challenge and registration please visit the web site.

Web-site: https://www.microsoft.com/en-us/research/event/interspeech-2018-special-session-low-resource-speech-recognition-challenge-indian-languages/

7. Integrating Speech Science and Technology for Clinical Applications

September 4, 2018, 14:30-16.30 Hall 4-6: PosterArea 4, HICC, Hyderabad
September 5, 2018, 10:00-12:00, MR G.03-G.04, HICC, Hyderabad

Organizers:
Christina Hagedorn (christina.hagedorn@csi.cuny.edu)
Shrikanth Narayanan (shri@ee.usc.edu)
Uttam Sinha (sinha@med.usc.edu)

The broad objectives of this session are to (i) address the current communication and collaboration gaps that exist between the fields of speech science, engineering and technological development, and communication disorders, and (ii) serve as a bridge to unify members of these distinct fields through interactive and dynamic exchanging of experimental findings and ideas for future collaboration. The organizers encourage submissions focused on, though not limited to: characterizing disordered speech using novel imaging techniques and analytical methods, (semi-) automatic detection of speech disorder characteristics, and the efficacy of biofeedback intervention for speech disorders.

Web-site: http://www.speechlab.csi.cuny.edu/ssinterspeech2018.html

8. Speech Technologies for Code-Switching in Multilingual Communities

September 5, 2018, 10:00-12:00, MR 1.01-1.02, HICC, Hyderabad

Organizers:
Kalika Bali (kalikab@microsoft.com)
Alan Black (awb@cs.cmu.edu)
Mona Diab (mtdiab@gwu.edu)
Julia Hirschberg (julia@cs.columbia.edu)
Sunayana Sitaram (t-susita@microsoft.com)
Thamar Solorio (solorio@cs.uh.edu)

Speech technologies exist for many high resource languages, and attempts are being made to reach the next billion users by building resources and systems for many more languages. Multilingual communities pose many challenges for the design and development of speech processing systems. One of these challenges is code-switching , which is the switching of two or more languages at the conversation, utterance and sometimes even word level.

Code-switching is now found in text in social media, instant messaging and blogs in multilingual communities in addition to conversational speech. Monolingual natural language and speech systems fail when they encounter code-switched speech and text. There is a lack of linguistic data and resources for code-switched speech and text, although one or more of the languages being mixed could be high-resource. Code-switching provides various interesting challenges to the speech community, such as language modeling for mixed languages, acoustic modeling of mixed language speech, pronunciation modeling and language identification from speech.

We conducted the inaugural special session on code-switching at Interspeech 2017, which was organized as a double session spanning four hours. We received several high-quality submissions from research groups all over the world, out of which nine papers were selected as oral presentations. At the end of the oral presentations, we conducted a panel discussion between researchers in academia and industry about challenges in research, building systems and collecting code-switched data. Our special session was attended by several researchers from academia and industry working on linguistics, NLP and speech technologies.

Web-site: https://www.microsoft.com/en-us/research/event/interspeech-2018-special-session-speech-techologies-code-switching-multilingual-communities/

9. Spoken CALL Shared Task, Second Edition

September 5, 2018, 14:30-16:30, MR 1.01-1.02, HICC, Hyderabad

Organizers:
Johanna Gerlach (Johanna.Gerlach@unige.ch)
Manny Rayner (Emmanuel.Rayner@unige.ch)
Martin Russell (m.j.russell@bham.ac.uk)
Helmer Strik (w.strik@let.ru.nl)

The Spoken CALL Shared Task is an initiative to create an open challenge dataset for speech-enabled CALL systems, jointly organised by the University of Geneva, the University of Birmingham, Radboud University and Cambridge University. The task is based on data collected from a speech-enabled online tool which has been used to help young Swiss German teens practise skills in English conversation. Items are prompt-response pairs, where the prompt is a piece of German text and the response is a recorded English audio file. The task is to label pairs as ”accept” or ”reject”, accepting responses which are grammatically and linguistically correct to match a set of hidden gold standard answers as closely as possible. Resources are provided so that a scratch system can be constructed with a minimal investment of effort, and in particular without necessarily using a speech recogniser.

The first edition of the task was announced at LREC 2016, with training data released in July 2016 and test data in March 2017, and attracted 20 entries from 9 groups. Results, including seven papers, were presented at the SLaTE workshop in August 2017. Full details, including links to resources, results and papers, can be found on the Shared Task home page.

Following the success of the original task, we are organising a second edition. We have approximately doubled the amount of training data, will provide new test data, and have released improved versions of the accompanying resources. In particular, we have made generally available the open source Kaldi recogniser developed by the University of Birmingham, which achieved the best performance on the original task, together with versions of the training and test data pre-processed through this recogniser.

Web-site: https://regulus.unige.ch/spokencallsharedtask_2ndedition/