INTERSPEECH 2014

CELEBRATING THE DIVERSITY OF SPOKEN LANGUAGES

14-18 SEPTEMBER 2014
SINGAPORE
MAX ATRIA@SINGAPORE EXPO

Keynotes

HTTP://WWW.INTERSPEECH2014.ORG

Keynote Speeches

Anne CUTLER

ISCA Medalist 2014

MARCS Institute, University of Western Sydney, Australia

Monday, 15 September 2014, 09:30 - 10:30; Garnet 213 - 218, Level 2, MAX Atria

Keynote Speech 1: Learning About Speech

Abstract

We human language users learn about speech all our lives. In fact if the beginning of our individual life is taken to be the moment of our birth, we learn for more than our whole lives, because even prior to birth we acquire much basic information about language sound structure. Then in early life, before we are capable of uttering any recognizable words, we have already learned a huge amount about the sounds and words of our language; furthermore, the efficiency with which we learn this is predictive of our later linguistic facility. We learn to process speech in the way best suited to the particular language (or languages) we acquire as children. This leads to a formidable efficiency and robustness in native-language processing, but it actually inhibits learning in the case of other languages we encounter only later - children are clearly very much better at learning a new language than adults are! Learning about speech nonetheless continues throughout life, in particular the everyday perceptual learning that enables us to adapt our speech processing to newly encountered talkers, and to adjust our own pronunciation in keeping with pronunciation changes in our speech community across time. This learning (at least in the native language) can draw on a wide variety of information sources, is fully in place in childhood, and is apparently unattenuated in older language users.

Biography

Anne Cutler is professor in the MARCS Institute, University of Western Sydney, and Processing program leader of the newly established ARC Centre of Excellence in the Dynamics of Language. She studied in Australia, Germany and the US, and worked in the UK (Sussex, Cambridge) and from 1993 to 2013 as director at the Max Planck Institute for Psycholinguistics in Nijmegen, The Netherlands. Her research, of which her book Native Listening (MIT Press 2012) gives an overview, centres on human listeners' recognition of spoken language. It has tended over the years to involve a great many cross-linguistic comparisons (e.g., English, Dutch, German, Japanese, Cantonese, Korean, Sesotho, French, Spanish, Italian, Finnish, Polish, Arabic, Telugu, Berber - so far).

K. J. Ray LIU

University of Maryland, College Park, USA

Tuesday, 16 September 2014, 08:30 - 09:30; Garnet 213 - 218, Level 2, MAX Atria

Keynote Speech 2: Decision Learning in Data Science: Where John Nash Meets Social Media

Abstract

With the increasing ubiquity and power of mobile devices, as well as the prevalence of social media, more and more activities in our daily life are being recorded, tracked, and shared, creating the notion of "social media". Such abundant and still growing real life data, known as "big data", provide a tremendous research opportunity in many fields. To analyze, learn and understand such user-generated big data, machine learning has been an important tool and various machine learning algorithms have been developed. However, since the user-generated big data is the outcome of users' decisions, actions and their socio-economic interactions, which are highly dynamic, without considering users' local behaviors and interests, existing learning approaches tend to focus on optimizing a global objective function at the macroeconomic level, while totally ignore users' local decisions at the microeconomic level. As such there is a growing need in bridging machine/social learning with strategic decision making, which are two traditionally distinct research disciplines, to be able to jointly consider both global phenomenon and local effects to understand/model/analyze better the newly arising issues in the emerging social media. In this talk, we present the notion of "decision learning" that can involve users' behaviors and interactions by combining learning with strategic decision making. We will discuss some examples from social media with real data to show how decision learning can be used to better analyze users' optimal decision from a user' perspective as well as design a mechanism from the system designer's perspective to achieve a desirable outcome.

Biography

Dr. K. J. Ray Liu was named a Distinguished Scholar-Teacher of University of Maryland in 2007, where he is Christine Kim Eminent Professor of Information Technology. He leads the Maryland Signals and Information Group conducting research encompassing broad areas of signal processing and communications with recent focus on cooperative communications, cognitive networking, social learning and decision making, and information forensics and security. Dr. Liu has received numerous honors and awards including IEEE Signal Processing Society 2009 Technical Achievement Award and various best paper awards from IEEE Signal Processing, Communications, and Vehicular Technology Societies, and EURASIP. A Fellow of the IEEE and AAAS, he is recognized by Thomson Reuters as an ISI Highly Cited Researcher. Dr. Liu was the President of IEEE Signal Processing Society, the Editor-in-Chief of IEEE Signal Processing Magazine and the founding Editor-in-Chief of EURASIP Journal on Advances in Signal Processing. Dr. Liu also received various research and teaching recognitions from the University of Maryland, including Poole and Kent Senior Faculty Teaching Award, Outstanding Faculty Research Award, and Outstanding Faculty Service Award, all from A. James Clark School of Engineering; and Invention of the Year Award (three times) from Office of Technology Commercialization.

Lori LAMEL

Senior Research Scientist (DR1), LIMSI-CNRS, France

Tuesday, 16 September 2014, 13:30 - 14:30; Garnet 213 - 218, Level 2, MAX Atria

Keynote Speech 3: Language Diversity: Speech Processing in A Multi-Lingual Context

Abstract

Speech processing encompasses a variety of technologies that automatically process speech for some downstream processing. These technologies include identifying the language or dialect spoken, the person speaking, what is said and how it is said. The downstream processing may be limited to a transcription or to a transcription enhanced with additional metadata, or may be used to carry out an action or interpreted within a spoken dialog system or more generally for analytics. With the availability of large spoken multimedia or multimodal data there is growing interest in using such technologies to provide structure and random access to particular segments. Automatic tools can also serve to annotate large corpora for exploitation in linguistic studies of spoken language, such as acoustic-phonetics, pronunciation variation and diachronic evolution, permitting the validation of hypotheses and models. In this talk I will present some of my experience with speech processing in multiple languages, drawing upon progress in the context of several research projects, most recently the Quaero program and the IARPA Babel program, both of which address the development of technologies in a variety of languages, with the aim to some highlight recent research directions and challenges.

Biography

I am a senior research scientist (DR1) at the CNRS, which I joined as a permanent researcher at LIMSI in October 1991. I received my Ph.D. degree in Electrical Engineering and Computer Science in May 1988 from the Massachusetts Institute of Technology. My research activities focus on large vocabulary speaker-independent, continuous speech recognition in multiple languages with a recent focus on low-resourced languages; lightly and unsupervised acoustic model training methods; studies in acoustic-phonetics; lexical and pronunciation modeling. I contributed to the design, and realization of large speech corpora (TIMIT, BREF, TED). I have been actively involved in the research projects, most recently leading the activities on speech processing in the OSEO Quaero program, and I am currently co-principal investigator for LIMSI as part of the IARPA Babel Babelon team led by BBN. I served on the Steering committee for INTERSPEECH 2013 as co-technical program chair along with Pascal Perrier, and I am now serving on the Technical Program Committee of INTERSPEECH 2014.

William S-Y. WANG 王士元

Chinese University of Hong Kong, Hong Kong SAR, China

Professor Emeritus, University of California at Berkeley, USA

Honorary Professor, Peking University, China

Academician, Academia Sinica, Taiwan

Wednesday, 17 September 2014, 08:30 - 09:30; Garnet 213 - 218, Level 2, MAX Atria

Keynote Speech 4: Sound Patterns in Language

Abstract

In contrast to other species, humans are unique in having developed thousands of diverse languages which are not mutually intelligible. However, any infant can learn any language with ease, because all languages are based upon common biological infrastructures of sensori-motor, memorial, and cognitive faculties. While languages may differ significantly in the sounds they use, the overall organization is largely the same. It is divided into a discrete segmental system for building words and a continuous prosodic system for expressing, phrasing, attitudes, and emotions. Within this organization, I will discuss a class of languages called 'tone languages', which makes special use of F0 to build words. Although the best known of these is Chinese, tone languages are found in many parts of the world, and operate on different principles. I will also comment on relations between sound patterns in language and sound patterns in music, the two worlds of sound universal to our species.

Speaker's Bio

Dr. William S-Y. Wang received his early schooling in China, and his Ph.D. from the University of Michigan. He was appointed Professor of Linguistics at the University of California at Berkeley in 1965, and taught there for 30 years. Currently he is in the Department of Electronic Engineering and in the Department of Linguistics and Modern Languages of the Chinese University of Hong Kong, and Director of the newly established Joint Research Centre for Language and Human Complexity. His primary interest is the evolution of language from a multi-disciplinary perspective.

Li DENG

Deep Learning Technology Center, Microsoft Research, Redmond, USA

Thursday, 18 September 2014, 08:30 - 09:30; Garnet 213 - 218, Level 2, MAX Atria

Keynote Speech 5: Achievements and Challenges of Deep Learning - From Speech Analysis and Recognition to Language and Multimodal Processing

Abstract

Artificial neural networks have been around for over half a century and their applications to speech processing have been almost as long, yet it was not until year 2010 that their real impact had been made by a deep form of such networks, built upon part of the earlier work on (shallow) neural nets and (deep) graphical models developed by both speech and machine learning communities. This keynote will first reflect on the path to this transformative success, sparked by speech analysis using deep learning methods on spectrogram-like raw features and then progressing rapidly to speech recognition with increasingly larger vocabularies and scale. The role of well-timed academic-industrial collaboration will be highlighted, so will be the advances of big data, big compute, and the seamless integration between the application-domain knowledge of speech and general principles of deep learning. Then, an overview will be given on sweeping achievements of deep learning in speech recognition since its initial success in 2010 (as well as in image recognition and computer vision since 2012). Such achievements have resulted in across-the-board, industry-wide deployment of deep learning. The final part of the talk will look ahead towards stimulating new challenges of deep learning - making intelligent machines capable of not only hearing (speech) and seeing (vision), but also of thinking with a "mind"; i.e. reasoning and inference over complex, hierarchical relationships and knowledge sources that comprise a vast number of entities and semantic concepts in the real world based in part on multi-sensory data from the user. To this end, language and multimodal processing - joint exploitation and learning from text, speech/audio, and image/video - is evolving into a new frontier of deep learning, beginning to be embraced by a mixture of research communities including speech and spoken language processing, natural language processing, computer vision, machine learning, information retrieval, cognitive science, artificial intelligence, and data/knowledge management. A review of recent published studies will be provided on deep learning applied to selected language and multimodal processing tasks, with a trace back to the relevant early connectionist modeling and neural network literature and with future directions in this new exciting deep learning frontier discussed and analyzed.

Speaker's Bio

Dr. Li Deng received his Ph.D. from the University of Wisconsin-Madison. He was a tenured professor (1989 - 1999) at the University of Waterloo, Ontario, Canada, and then joined Microsoft Research, Redmond, where he is currently a Principal Research Manager of its Deep Learning Technology Center. Since 2000, he has also been an affiliate full professor at the University of Washington, Seattle, teaching computer speech processing. He has been granted over 60 US or international patents, and has received numerous awards and honors bestowed by IEEE, ISCA, ASA, and Microsoft including the latest IEEE SPS Best Paper Award (2013) on deep neural nets for speech recognition. He authored or co-authored 4 books including the latest one on Deep Learning: Methods and Applications. He is a Fellow of the Acoustical Society of America, a Fellow of the IEEE, and a Fellow of the ISCA. He served as the Editor-in-Chief for IEEE Signal Processing Magazine (2009 - 2011), and currently as Editor-in-Chief for IEEE Transactions on Audio, Speech and Language Processing. His recent research interests and activities have been focused on deep learning and machine intelligence applied to large-scale text analysis and to speech/language/image multimodal processing, advancing his earlier work with collaborators on speech analysis and recognition using deep neural networks since 2009.