Keynote Sessions

Keynote 1

ISCA Medalist Victor Zue, MIT, Cambridge, MA

On Organic Interfaces

Queen Elisabeth Hall, Tuesday, August 28, 11:00 to 12:00 am
Chairperson: Julia Hirschberg

Abstract
For over four decades, our research community has taken remarkable strides in advancing human language technologies. This has resulted in the emergence of spoken dialogue interfaces that can communicate with humans on their own terms. For the most part, however, we have assumed that these interfaces are static; it knows what it knows and doesn't know what it doesn't. In my opinion, we are not likely to succeed until we can build interfaces that behave more like organisms that can learn, grow, reconfigure, and repair themselves, much like humans. In this paper, I will argue my case and outline some new research challenges.

Presenter

Victor Zue is the Delta Electronics Professor of Electrical Engineering and Computer Science at MIT and the Director of the Institute's Computer Science and Artificial Intelligence Laboratory (CSAIL). In the early part of his career, Victor conducted research in acoustic phonetics and phonology, codifying the acoustic manifestation of speech sounds and the phonological rules governing the realization of pronunciation in American English. Subsequently, his research interest shifted to the development of spoken language interfaces to make human-computer interactions easier and more natural. Between 1989 and 2001, he headed the Spoken Language Systems Group at the MIT Laboratory for Computer Science, which has pioneered the development of many systems that enable a user to interact with computers using spoken language.

Outside of MIT, Victor has consulted for many multinational corporations, and he has served on many planning, advisory, and review committees for the US Department of Defense, the National Science Foundation, and the National Academies of Science and Engineering. From 1996-1998, he chaired the Information Science and Technology, or ISAT, study group for the Defense Advanced Research Projects Agency of the U.S. Department of Defense, helping the DoD formulate new directions for information technology research. In 1999, he received the DARPA Sustained Excellence Award. Victor is a Fellow of the Acoustical Society of America, and a member of the U.S. National Academy of Engineering.

Keynote 2

Sophie Scott, Institute of Cognitive Neuroscience, University College London

The Neural Basis of Speech Perception – a view from functional imaging

Queen Elisabeth Hall, Wednesday, August 29, 8:30 to 9:30 am
Chairperson: Anne Cutler

Abstract
Functional imaging techniques, such as Positron Emission Tomography (PET) and functional Magnetic Resonance Imaging (fMRI), have enabled neuroscientists to elaborate how the human brain solves the formidable problem of decoding the speech signal. In this paper I will outline the properties of primate auditory cortex, and use this as an anatomical framework to address the data from functional imaging studies of auditory processing and speech perception. I will outline how at least two different streams of processing can be seen in primary auditory cortex, and that this apparently maps onto two different ways in which the human brain processes speech. I will also address data suggesting that there are considerable hemispheric asymmetries in speech perception.

Presenter

Sophie K Scott is a Chair in Cognitive neuroscience at UCL, and is the group leader of the Speech Communication group at the Institute of Cognitive Neuroscience. She has worked in speech perception and production for over seventeen years. Following a PhD in speech rhythm at UCL, she worked at the MRC APU in Cambridge, getting involved in the neuroimaging of speech perception, as well as work on the expression of emotion in the voice. Her research covers the neural basis of speech perception, with specific projects including the nature of acoustic-phonetic representations in auditory cortex, the roles of streams of processing in the processing of speech, speech perception and production links, the nature of hemispheric asymmetries in speech perception, and the roles of learning and top-down modulation in auditory cortex. She is also starting projects working with auditory implant patients. Her work is funded by the Wellcome Trust, as well as Marie Curie and the ESRC.

WebSite: http://www.icn.ucl.ac.uk/
Keynote 3

Alex Waibel, CMU, Pittsburgh, PA & University of Karlsruhe, Germany

Computer Supported Human-Human Multilingual Communication

Queen Elisabeth Hall, Thursday, August 30, 8:30 to 9:30 am
Chairperson: Tanja Schultz

Abstract
With just a click we can get to any information anytime and anywhere, yet language, cultural differences and poor interfaces still ensure that we are mystified and disconnected in many of life’s situations, nonetheless. Surely, overcoming such boundaries is a much harder problem, but does it have to remain unaddressed? Expanding human-human communication always proved to be good business and a worthy investment. Instead of focusing on the human-machine interface alone, can we perhaps devise computer enhanced human-human services, that support or empower human interaction, rather than diverting human attention in a human-machine dialog. Computers in a Human Interaction Loop (CHIL) are designed to provide such services. I will introduce CHIL computing services that are being developed at our lab as well as other sites, and the technologies that make them possible. I will then highlight and focus in on one CHIL service, that seems of particularly revolutionary potential to us: unlimited cross-lingual human-human communication.

Human interaction with anyone, free from language barriers! The potential seems endless, but the dream also harbors several formidable challenges: 1.) modality - how to process multilingual information when it is not in textual form, 2.) domain independence - how to process any content independent of topic or domain, 3.) language portability - how to handle the large (~6,000) number of languages, 4.) the interface - how to make the barrier disappear, and provide for cross lingual human-human interaction transparently. For each of these problems, I will describe ongoing work that begins to address these concerns.

Presenter

Alex Waibel is a Professor of Computer Science at Carnegie Mellon University, Pittsburgh and also Professor at the University of Karlsruhe (Germany). He directs InterACT, the international Center for Advanced Communication Technologies at both Universities with research emphasis in speech recognition, language processing, speech translation, multimodal and perceptual user interfaces. At Carnegie Mellon, he also serves as Associate Director of the Language Technologies Institute and holds joint appointments in the Human Computer Interaction Institute and the Computer Science Department.

Dr. Waibel was one of the founders of C-STAR, the international consortium for speech translation research and served as its chairman from 1998-2000. Since 1990 his team has developed the JANUS speech translation system, the first integrated Speech Translation system in Europe and the US. Since 2000, his research team began work on domain independent speech translation leading to projects STR-DUST and TC-STAR and to the demonstration of the first real-time simultaneous speech translation system for lectures in 2005. At InterACT, Dr. Waibel's team has also developed multimodal systems and computing services including perceptually aware Meeting Rooms, Meeting recognizers, Meeting Browsers and multimodal dialog systems for humanoid robots. He currently directs the CHIL program (a large FP-6 Integrated Project on multimodality) in Europe, the NSF-ITR project STR-DUST in the US. In the areas of speech, speech translation, and multimodal interfaces Dr. Waibel holds several patents and has founded and co-founded several successful commercial ventures.

Dr. Waibel received the B.S. in Electrical Engineering from the Massachusetts Institute of Technology in 1979, and M.S. and Ph.D. degrees in Computer Science from Carnegie Mellon University in 1980 and 1986. His work on the Time Delay Neural Networks was awarded the IEEE best paper award in 1990. His contributions to multilingual and speech translation systems was awarded the “Alcatel SEL Research Prize for Technical Communication” in 1994, the “Allen Newell Award for Research Excellence” from CMU in 2002, and the Speech Communication Best Paper Award in 2002.

WebSite: http://www.interact.cs.cmu.edu/
Keynote 4

Pierre-Yves Oudeyer, Sony Computer Science Laboratory, Paris

Self-Organization in the Evolution of Shared Systems of Speech Sounds:
a Computational Study

Queen Elisabeth Hall, Friday, August 31, 8:30 to 9:30 am
Chairperson: Louis Pons

Abstract
How did culturally shared systems of combinatorial speech sounds initially appear in human evolution? This paper proposes the hypothesis that their bootstrapping may have happened rather easily if one assumes an individual capacity for vocal replication, and thanks to self-organization in the neural coupling of vocal modalities and in the coupling of babbling individuals. This hypothesis is embodied in agent-based computational experiments, that allow to show that crucial phenomena, including structural regularities and diversity of sound systems, can only be accounted if speech is considered as a complex adaptive system. Thus, the second objective of this paper is to show that integrative computational approaches, even if speculative in certain aspects, might be key in the understanding of speech and its evolution.

Presenter

Pierre-Yves Oudeyer is researcher at the Sony Computer Science Laboratory in Paris where he co-founded and heads the Developmental Robotics group. He is also a member of the Origins of Language group at Sony CSL, and teaches Social and Cognitive Robotics at Ecole Nationale Supérieure des Techniques Avancées. He studied theoretical computer science at Ecole Normale Supérieure de Lyon, and received his PhD in artificial intelligence from University Paris VI. He is interested in the mechanisms that allow humans and robots to develop perceptual, motivational, behavioral and social capabilities to become capable of sharing cultural representations and of natural embodied interaction. In particular, he uses robots to study how new linguistic conventions can be established in a society of individuals, and has developed numerous computational models of the interactions between self-organization, learning and selection in the evolution of language. He also works on sensorimotor development and self-motivation, and his group built one of the first robots endowed with artificial curiosity. He participated in the development of emotional speech synthesis for the Sony Qrio humanoid robot. He was co-chair of the Epigenetic Robotics international conference, and is associate editor of Frontiers in Neurorobotics. He is expert and reviewer in cognitive robotics and speech technologies for the European Commission as well as for the French National Research Agency (ANR). He has published a book, more than 50 papers in international journals and conferences, and holds 8 patents on technologies such as emotional speech synthesis and recognition, active learning, and human interaction with autonomous robots. He received several prizes for his research on the origins of speech and developmental systems, including the prize “Le Monde de la recherche universitaire”.

WebSite: http://www.csl.sony.fr/~py/