

Monday, September 16, 2002

Tutorial Registration

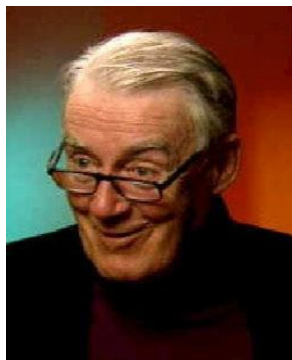
<i>Room</i>	<i>Plaza Ballroom E</i>	<i>Plaza Ballroom F</i>
<i>9:00 to 12:00 noon</i>	Collecting Data on Speech Sounds of the World's Languages Organizer: Peter Ladefoged Univeristy of California - Los Angeles	Adaptive Methods for Speech and Speaker Recognition Organizers: Jean-Claude Junqua and Roland Kuhn Panasonic Speech Laboratory
<i>12 noon to 1:00 pm</i>	<i>Lunch</i>	
<i>Room</i>	<i>Plaza Ballroom E</i>	<i>Plaza Ballroom F</i>
<i>1:00 - 4:00 pm</i>	Automatic Speech Recognition for Wireless Mobile Devices Organizers: Richard Rose and S. Parthasarathy AT&T Labs - Research	Prosody and Prosodic Models Organizers: Chilin Shih and Greg Kochanski Bell Laboratories
<i>4:00 - 6:00 pm</i>	<i>Break and Dinner</i>	
<i>Room</i>	<i>Plaza Ballroom E</i>	<i>Plaza Ballroom F</i>
<i>6:00 - 9:00 pm</i>	Psychoacoustics Organizer: Brian C.J. Moore, University of Cambridge, England	Weighted Finite-State Transducers in Speech & Language Processes Organizers: Mehryar Mohri and Michael Riley AT&T Labs - Research

MORNING SESSIONS (9:00 AM TO 12:00 PM)

COLLECTING DATA ON SPEECH SOUNDS OF THE WORLD'S LANGUAGES

Tutorial Overview:

This course has two main aims. The first is to consider the fieldwork required for making a description of the sounds of a language. The second is to illustrate the basic techniques of experimental phonetics, most of them requiring little more than a tape recorder, a video camera, and a few other items, none of them very expensive, together with a computer and appropriate programs. There are therefore two principal sets of people who might be interested: those who want to learn about fieldwork techniques, and those who want a simple introduction to basic laboratory tools. All investigations of speech, whether in a lab or a classroom, or a distant country, involve an observer and someone whose speech is being observed. This is true whether you are investigating the variant pronunciations of street names in Los Angeles for a speech recognition company, going into the wilds of the Brazilian rain forest to record the sounds of Banawa, or into the streets of a big city to find out how the home boys talk. Even obtaining data from a carefully controlled group in a laboratory experiment is really a kind of fieldwork. You need to determine how to set up an appropriate group and how to elicit the speech sounds you are trying to investigate. This course will provide a practical introduction to the techniques for describing the major phonetic characteristics of any language.



Peter Ladefoged is Professor of Phonetics Emeritus, was Director of the UCLA Phonetics Laboratory from 1962 to 1991. He is the author of several books, including - *A Course in Phonetics* (4th. Ed. Harcourt Brace 2001), *Elements of Acoustic Phonetics* (2nd. Ed. Chicago University Press 1999), and *Vowels and Consonants* (Blackwells 2001) and co-author with Ian Maddieson of *Sounds of the World's Languages* (Blackwells 1996). He has also published more than 100 articles in scholarly journals. He was President of the Linguistic Society of America (1980), and of the International Phonetic Association (1986 -1991). Among his other honors is a Distinguished Teaching Award from UCLA.

ADAPTIVE METHODS FOR SPEECH AND SPEAKER RECOGNITION

Tutorial Overview:

While in the last decade the main focus of research in speaker adaptation was on obtaining robust SI systems, recently adaptive systems have attracted much interest. Several types of adaptation have been studied, dealing with such phenomena as speaking style, speaking rate, non-native accents, transducers and transmission channels, noise, nature of the task, etc. In this tutorial, we cover two important classes of adaptation approach:

Model-based adaptation which modifies the model parameters to better fit the new speaker acoustic data. In this category we can find Bayesian, transformation-based, and speaker clustering techniques; feature-based adaptation which attempts to reduce inter-speaker acoustic differences by means of acoustic data transformations (e.g. vocal tract length normalization).

We summarize the main classes of techniques and show how they have been applied to different classes of problems such as noise compensation and speaker adaptation. We give a special attention to techniques allowing fast adaptation due to the practicality of these techniques for many applications. We show how adaptation techniques can be combined to help dealing with several types of variability. Finally, we discuss the links between speaker adaptation and speaker recognition, showing how techniques and insights from each field can contribute to the other.



Jean-Claude Junqua: Jean-Claude Junqua received his Engineer degree (1980) from ENSEM (France) in Electronics and Automation, his Master and Doctorate degrees (in 1981 and 1989, respectively) and the "Habilitation à diriger des recherches" (1993) from the University of Nancy I (France) in the field of Computer Science. From 1981 to 1986 he was responsible for the computer facilities of CRIN (Research Center in Computer Science of Nancy, France). From 1987 to 1988 he was visiting researcher at Panasonic's Speech Technology Laboratory in Santa Barbara, California. In 1989, he joined Speech Technology Laboratory. From April 1992 to August 1993 he was visiting researcher at Matsushita, Osaka, Japan. He is currently Vice President & Director at Panasonic's Speech Technology Laboratory. He is the author of more than 100 articles or patents in the above areas along with two recent books entitled "Robustness in Automatic Speech Recognition" and "Robust Speech Recognition in Embedded Systems and PC Applications". At the beginning of 2001, he also co-edited a book on "Robustness in Language and Speech Technology". He served as a chairman at several International conferences, co-organized several ISCA/IEEE workshops and participated in various International scientific committees. Jean-Claude Junqua was a tutorial speaker for several ESCA/IEEE workshops and for ICASSP'99. He was an Associate Editor of the IEEE Transactions on Speech and Audio Processing and he is currently on the Editorial Board of the Speech Communication Journal.



Roland Kuhn obtained an Honors B.Sc. in Mathematics and Biology (1981) from Trinity College, University of Toronto, an M.Sc. in Theoretical Biology (1984) from the University of Chicago, and a Ph.D. in Computer Science (1993) from McGill University. From 1992 to 1996, he worked in the speech recognition group of CRIM (Centre de Recherche Informatique de Montréal), serving as Lead Researcher (Chercheur principal) from June 1995 to September 1996. In Oct. 1996, he moved to Santa Barbara to become a Senior Researcher at Panasonic's Speech Technology Laboratory. His research interests have included language modeling (cache language model), natural language understanding (semantic classification trees), and speaker adaptation and acoustic modeling (eigenvoices). He is the author of over 50 articles or patents in these areas, and was a contributor to the recent book "Spoken Dialogues with Computers" (edited by Renato De Mori).

AFTERNOON SESSIONS (1:00 PM TO 4:00 PM)

AUTOMATIC SPEECH RECOGNITION FOR WIRELESS MOBILE DEVICES

Tutorial Overview :

This tutorial will cover topics in automatic speech recognition (ASR) that are specific to the implementation of ASR systems on wireless mobile devices. Emphasis will be placed on algorithms and architectures that relate to ASR in the context of the mobile devices themselves, the environments they are used in, the networks they are connected to, and the target applications that they support. The tutorial will have two major parts. The first part will consist of a survey of existing mobile ASR applications, architectures, and supporting technology. The second part will present algorithms that have been developed to make ASR systems more robust, more efficient, and more flexible for mobile applications.



Richard Rose received B.S. and M.S. degrees from the Electrical and Computer Engineering Department at the University of Illinois. He obtained a Ph.D. E.E. degree from what is now known as the the Center for Signal and Image Processing (CSIP) at the Georgia Institute of Technology in 1988 with a thesis in speech coding and speech analysis. From 1980 to 1984, he was with Bell Laboratories, now a division of Lucent Technologies working on signal processing and digital switching systems. From 1988 to 1992, he was a member of the Speech Systems and Technology group, now called the Information Systems Technology Group, at MIT Lincoln Laboratory working on speech recognition and speaker recognition. He has been with AT&T since 1993 working in the general area of automatic speech recognition, and is currently a principal member of technical staff in the Speech and Image Processing Services Laboratory at AT&T Labs - Research in Florham Park, NJ
Mr. Rose served as a member of the IEEE SPS Technical Committee on DSP, was elected as an at large member of the Board of Governors for the Signal Processing Society during the period from 1995 to 1997, and was an associate editor for the IEEE Transactions on Speech and Audio Processing from 1997 to 1998. He is a senior member of the IEEE.



S. Parthasarathy received M.S. and Ph.D. degrees from the Electrical Engineering Department at the University of Rhode Island. His masters research was in the area of robotic control and his doctoral work focused on robust parameter estimation. He joined AT&T Bell Laboratories, Murray Hill, NJ, in 1987 and worked on articulatory speech synthesis and the problem of automatically estimating articulatory parameters from the speech signal. From 1990 to 1993, he was with Bell-Northern Research in Montreal, working on large vocabulary speech recognition for telephony applications. He then returned to AT&T Bell Laboratories (now AT&T Labs-Research) and has been involved in research on speech and speaker recognition. He is a member of the IEEE and serves on the Speech Technical Committee of the IEEE Signal Processing Society.

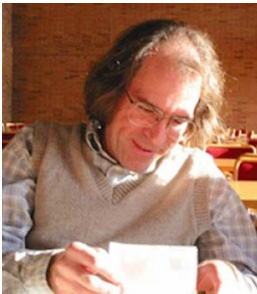
PROSODY AND PROSODIC MODELS

Tutorial Overview :

The tutorial will summarize the state of the art of prosody research with an emphasis on intonation (f_0). We will present it in a data and knowledge driven fashion, starting with speech data and physiology, and building towards quantitative models. We will cover a wide range of intonation and duration phenomena, discussing the variations, the cause of variations, and their implications to prosodic modeling.



Dr. Chilin Shih works on prosodic modeling and text-to-speech systems at Bell Labs, and has built Text-to-Speech systems for many languages, including Chinese, Navajo, Romanian, Hindi, Greek, Spanish, Italian, and Russian. She has published in linguistics and speech research, including phonology, morphology, phonetics, and prosody.



Dr. Greg Kochanski builds mathematical models of complex systems at Bell Laboratories, most recently focusing on prosody and speech synthesis. He has published in a variety of fields, including astrophysics and solid-state physics.

EVENING SESSIONS (6:00 PM TO 9:00 PM)

PSYCHOACOUSTICS

Tutorial Overview :

The course will describe how psychoacoustic studies, together with knowledge of the underlying physiology, can be used to understand how sounds are represented in the peripheral auditory system. This understanding allows the development of models that can be used to predict the detection threshold and loudness of complex sounds, and to study the timbre of complex sounds. The perception of time-varying sounds will also be described and discussed. Finally, factors influencing the perception of auditory objects and the formation of perceptual streams will be described.



Brian Moore is Professor of Auditory Perception in the University of Cambridge. He is a Fellow of the Acoustical Society of America, an Honorary Fellow of the Belgian Society of Audiology and of the British Society of Hearing Aid Audiologists, and a Fellow of the Academy of Medical Sciences. He is President of the Association of Independent Hearing Healthcare Professionals (UK). He is a member of the Editorial Boards of Hearing Research, The International Journal of Audiology and Audiology and Neuro-Otology. He is a consultant for several US and European companies. He has published 9 books and over 350 scientific papers and book chapters. He is wine steward of Wolfson College, Cambridge.

WEIGHTED FINITE-STATE TRANSDUCERS IN SPEECH & LANGUAGE PROCESSING

Overview :

We survey the use of weighted finite-state transducers (WFSTs) in speech recognition. We show that WFSTs provide a common and natural representation for HMM models, context-dependency, pronunciation dictionaries, grammars, and alternative recognition outputs. General transducer algorithms such as composition, union, and concatenation combine these representations flexibly and efficiently. Weighted determinization and minimization algorithms optimize their time and space requirements, and a weight pushing algorithm distributes the weights along the paths of a weighted transducer optimally for speech recognition.



Mehryar Mohri graduated from Ecole Polytechnique in France and received a Ph.D. in Computer Science from the University of Paris in 1993. Dr. Mohri is Head of the Speech Algorithms Research Department at AT&T Labs - Research. He joined AT&T Bell Laboratories in 1995 as a Member of Technical Staff. Prior to this appointment, he was an Associate Professor in computer science at the University of Paris. Dr. Mohri has also been an Adjunct Professor of computer science at Columbia University. Dr. Mohri's principal areas of research are finite-state transducer algorithms, natural language processing, automatic speech recognition, speech synthesis, general-purpose software libraries, formal language theory, combinatorial pattern matching, text processing algorithms, machine learning. He is co-author of the FSM library and the GRM library widely used in text and speech processing applications.



Michael Riley grew up in Limu, Peru and Sarasota, FL. He attended the Massachusetts Institute of Technology where he received a B.S., M.S., and Ph.D. in Computer Science, the last in 1987. He has worked at Bell Telephone Laboratories, Bell Laboratories, AT&T Bell Laboratories, and most recently AT&T Labs -- Research. His research interests include finite-state methods in speech processing, ASR search methods, and pronunciation modeling.