Plenary Speakers



Jacqueline Vaissière

Professor Emeritus, Université Sorbonne Nouvelle, France http://www.univ-paris3.fr/vaissiere-jacqueline-29931.kjsp

Universal Tendencies for Cross-Linguistic Prosodic Tendencies : A Review and Some New Proposals

Abstract

The present talk aims first to review the literature on similar tendencies regularly observed in typologically unrelated languages. The tendencies concern the use of fundamental frequency (F0, including declination line as the reference line, the top-line, up-stepping, down-stepping, register change and range widening-reducing), lengthening-shortening maneuvers and strengtheningweakening phenomena at the glottal and supraglottic levels, for instantiating acoustically the syllable, the word, the minor and major phrases, and the utterance. Our presentation concerns only attitudinally and emotionally neutral utterances. The second part of the talk will present particular aspects: 1) the different centers of articulatory "effort" at the syllable level; 2) the suggestion of the existence of an unmarked strong-long pattern, neither trochaic nor iambic, at the word level in languages where natives don't have the consciousness of a "lexical stress," or don't agree on its existence or position; 3) the regrouping of one or more words into a prosodic phrase by the application of two established principles: a) the "hat-pattern" principle (t'Hart) favoring initial high-rising and final low-falling F0, and b) the intensive or the temporal rhythmic basic tendencies (Woodrow, Fraisse) favoring a more intense, stronger, more precisely articulated beginning and a lengthened ending; 4) the existence of a multilayer rhythm at the utterance level composed by the repetition/alternation of integrated Gestalts at the levels of the syllable, word, and phrases. One or two Gestalts will prevail perceptually depending on a) the language, b) the style, and c) the rate of speech. The impressionistic evidence of a particular type of languagedependent "rhythm" depends on the listener's expectations, related to his maternal language and the languages he already masters, and up to a certain extent, to his pre-existing theoretical beliefs.

Biography

After a thesis in 1970 on French prosody synthesis at the IBM Research Center, La Gaude (France), and the Centre d'études pour la Traduction Automatique, Grenoble, Jacqueline Vaissière joined the Speech Communication Group at MIT, directed by Ken Stevens, as a visiting scientist, where she acquired a specialization in acoustic phonetics. In 1975, she joined the Centre National d'Etudes des Télécommunications (France), where she worked for 15 years on automatic speech recognition and automatic directory services. When the speech processing community moved towards black-box models for recognition and synthesis, she chose to become a professor at the Université Sorbonne Nouvelle in 1990 and became the director of Laboratoire de phonétique

et de Phonologie, associated with the Centre National de la Recherche Scientifique (CNRS). She was the director of 125 masters and 34 doctoral theses, with students of different backgrounds (medical doctors, engineers, and linguistics native of a large variety of languages). She was awarded a CNRS silver medal in 2009. Since 2010, she has been the principal coordinator of the 10-year project « Laboratoire d'Excellence « Empirical Foundations of Linguistics ». She is currently developing methods and applications for acquiring or improving pronunciations, based on visualization of spectrograms, and F0 curves (CleanAccent) and give courses decoding segmental and suprasegmental cues from spectrograms, F0 and intensity curves in different languages. Jacqueline Vaissière is a Fellow of ISCA.



Hervé Bourlard Idiap Research Institute and EPFL, Switzerland http://people.idiap.ch/bourlard

Evolution of Neural Network Architectures for Speech Recognition

Abstract

Over these last few years, the use of Artificial Neural Networks (ANNs), now often referred to as deep learning or Deep Neural Networks (DNNs), has significantly reshaped research and development in a variety of signal and information processing tasks. While further boosting the state-of-the-art in Automatic Speech Recognition (ASR), recent progresses in the field have also allowed for more flexible and faster developments in emerging markets and multilingual societies (e.g., under-resourced languages).

In this talk, we will provide a historical account of ANN architectures used for ASR since the mid-1980's, and now used in most ASR and spoken language understanding applications. We will start by recalling/revisiting key links between ANNs and statistical inference, discriminant analysis, and linear/nonlinear algebra. Finally, we will briefly discuss more recent trends towards novel DNN-based ASR approaches, including complex hierarchical systems, sparse recovery modeling, and "end-to-end systems."

However, and in spite of the recent progress in the area, we still lack basic understanding of the problems in hands. Although more and more tools are now available, in association with basically "unlimited" processing and data resources, we still fail in building principled ASR models and theories. Alternatively, we are still relying on "ignorance-based" models, often exposing limitations of our understanding, rather than enriching the field of ASR. Discussion of these limitations will underpin all of our overview.

Biography

Hervé Bourlard is Director of the Idiap Research Institute, Full Professor at the Swiss Federal Institute of Technology Lausanne (EPFL), and Founding Director of the Swiss NSF National Centre of Competence in Research on "Interactive Multimodal Information Management (IM2)" (2001-2013). He is also an External Fellow of the International Computer Science Institute (ICSI), Berkeley, CA.

His research interests mainly include statistical pattern classification, signal processing, multichannel processing, artificial neural networks, and applied mathematics, with applications to a wide range of Information and Communication Technologies, including spoken language processing, speech and speaker recognition, language modeling, multimodal interaction, and augmented multi-party interaction.

H. Bourlard is the author/co-author/editor of 8 books, and over 330 reviewed papers (including one IEEE paper award). He is a Fellow of IEEE and ISCA, and a Senior Member and Member of the European Council of ACM. He is the recipient of several scientific and entrepreneurship awards.



Helen Meng

Department of Systems Engineering and Engineering Management Chinese University of Hong Kong http://www.se.cuhk.edu.hk/people/hmmeng.html

Speech and Language Processing for Learning and Wellbeing

Abstract

Spoken language is a primary form of human communication. Spoken language processing techniques must incorporate knowledge of acoustics, phonetics and linguistics in analyzing speech. While great strides have been made in the community in general speech recognition, reaching human parity in performance, our team has been focusing on the problems of recognizing and analyzing non-native, learners' speech for the purpose of mispronunciation detection and diagnosis in computer-aided pronunciation training. In order to generate personalized, corrective feedback, we have also developed an approach that uses phonetic posterior-grams (PPGs) for personalized, cross-lingual text-to-speech synthesis given arbitrary textual input, based on voice conversion techniques. We have also extended our work to disordered speech, focusing on automated distinctive feature (DF)-based analyses of dysarthric recordings. The analyses are intended to inform intervention strategies. Additionally, voice conversion is further developed to restore disordered speech to normal speech. This talk will present the challenges in these problems, our approaches and solutions, as well as our ongoing work.

Biography

Helen Meng is Patrick Huen Wing Ming Professor and Chairman of the Department of Systems Engineering & Engineering Management, Chinese University of Hong Kong (CUHK). She is the Founding Director of the CUHK Ministry of Education (MoE)-Microsoft Key Laboratory for Human-Centric Computing and Interface Technologies, Tsinghua-CUHK Joint Research Center for Media Sciences, Technologies and Systems, and CUHK Stanley Ho Big Data Decision Analytics Research Center. She has also established the CAS-CUHK Shenzhen Institute of Advanced Technology Ambient Intelligence and Multimodal Systems Laboratory and served as its Director between 2007 and 2011. Previously, she has served as CUHK Faculty of Engineering's Associate Dean (Research), Editor-in-Chief of the IEEE Transactions on Audio, Speech and Language Processing, ISCA Board Member, Member of the IEEE SPS Board of Governors and Hong Kong-Guangdong ICT Expert Group member. Presently, she is serving as ISCA International Advisory Council Member, elected Chairperson of ISCA's Special Interest Group on Chinese Spoken Language Processing (since 2014) and elected Standing Committee Member of the China Computer Federation Task Force on Speech, Dialogue and Auditory Processing. Her appointments by the Hong Kong SAR Government (HKSARG) include Research Grants Council Member, eHealth Record Sharing Steering Committee Member, and Chairlady of the Working Party for the Manpower Survey of the Innovation & Technology Sector. She was APSIPA's inaugural Distinguished Lecturer 2012-2013 and ISCA Distinguished Lecturer 2015-2016. Her awards include the Ministry of Education Higher Education Outstanding Scientific Research Output Award 2009, Hong Kong Computer Society's inaugural Outstanding ICT Woman Professional Award 2015, Microsoft Research Outstanding Collaborator Award 2016, IEEE ICME 2016 Best Paper Award, IBM Faculty Award 2016, HKPWE Outstanding Women Professionals and Entrepreneurs Award 2017 and Hong Kong ICT Award 2018 Silver Award for Smart Inclusion. Helen received all her degrees from MIT and is a Fellow of HKCS, HKIE, IEEE and ISCA.