Odyssey 2012

The Speaker and Language Recognition Workshop

25 June – 28 June 2012, Singapore

- Welcome Message
- Author Index
- Session Index
- Invited Speakers
- Technical Program
- Social Program
- Committees

Welcome Message

Welcome to Odyssey 2012: The Speaker and Language Recognition Workshop, hosted by COLIPS (Chinese and Oriental Languages Information Processing Society) in Singapore, on 25-28 June 2012.
Odyssey 2012 received overwhelming response from the speaker and language recognition community. We accepted 51 papers out of 65 submissions, which we organized into a 4-day technical program consisting of 11 sessions. Researchers will present their latest endeavours and insights from multiple aspects, covering speaker and language characterization, modelling, evaluation, and applications. In addition, Odyssey 2012 also features 3 invited speakers: Dr. Li Deng (Microsoft Research) will share with us how new generation models such as deep belief networks and dynamic Bayesian networks can revamp the traditional framework of Gaussian mixture model and hidden Markov model in speech technology; Dr. Niko Brümmer (Agnitio Corporation) will discuss how

Organizers:

COLIPS

SpLC Speaker and Language Characterization

Supporters:

isca

NANYANG TECHNOLOGICAL UNIVERSITY

Temasek Laboratories@NTU

proper scoring rules adopted from weather prediction can be applied to pattern recognition to make probabilistic decisions; and Dr. Alvin Martin (NIST: National Institute of Standards and Technology) will talk about the past, present and future of the NIST Speaker Recognition Evaluations.

With a unique history and geographical location, Singapore is one of the few places on earth to experience the extreme blending of diverse languages. In the streets and neighbourhoods of Singapore, you can easily hear people talking in a rich mix of languages ranging from Chinese varieties (e.g., Mandarin, Hokkien, Cantonese) to other Asian languages (e.g., Tamil, Malay), and to English (e.g., British and Singaporean dialects). It is our hope to celebrate the diversity of languages by hosting the Odyssey 2012 Workshop in Singapore. I therefore wish you an enjoyable stay with a multi-cultural experience!

Many volunteers have contributed significantly to organizing this workshop. I would like to take this opportunity to show my gratitude to the organizing committee for valuable advice, and the scientific committee for their thorough paper reviews. Great thanks also go to my colleagues at Institute for Infocomm Research, especially to: Bin Ma and Kong Aik Lee (technical program and website), Swee Lan See (finance), Cheung-Chi Leung (venue), Anthony Larcher and Hanwu Sun (social activities), Nancy Chen and Minghui Dong (publicity and sponsorship), Rong Tong (registration), Changhuai You (publication). I also appreciate the financial support from the Lee Foundation, Temasek Laboratories at Nanyang Technological University, and International Speech Communication Association (ISCA).

Last but not least, I sincerely thank the reviewers, authors, invited speakers, presenters, exhibitors, local supporting staff, and all the workshop attendees for your support and enthusiasm. It is you that makes this conference possible. Thank you!

Haizhou Li
Odyssey 2012 Chair

Institute for Infocomm Research
A*STAR

金基氏李
Lee Foundation

Odyssey 2012

The Speaker and Language Recognition Workshop

25 June – 28 June 2012, Singapore

- Welcome Message
- Author Index
- Session Index
- Invited Speakers
- Technical Program
- Social Program
- Committees

Invited speakers

Dr Niko Brümmer, Chief Scientist, AGNITIO, South Africa
Dr Li Deng, Principal Researcher, Microsoft Research, USA
Dr Alvin Martin, Mathematician, NIST, USA

Plenary Session 1

Dr Niko Brümmer
Chief Scientist
AGNITIO, South Africa

The Role of Proper Scoring Rules in Training and Evaluating Probabilistic Speaker and Language Recognizers

It is obvious how to evaluate the goodness of a pattern classifier that outputs hard classification decisions --- you count the errors. But hard classification decisions are implicitly dependent on fixed priors and costs, so that they are applicable only in a narrow range of applications. A classifier can widen its range of applicability by outputting instead soft decisions, in the form of class probabilities or likelihoods. However, it is much less obvious how to evaluate the goodness of such probabilistic outputs. To evaluate the goodness of recognized classes, they can simply be compared to the true class labels in a supervised evaluation database. But we simply don't have a similar truth reference for probabilistic outputs.

A solution to this problem, originally from weather prediction, called "proper scoring rules", has been known for several decades, but has enjoyed only limited attention in pattern recognition and machine learning. This talk will explain how they work, how they generalize error-rate, how they measure information and how to use them for both training and evaluation of probabilistic pattern recognizers.

Short Biography

Niko Brummer received B.Eng (1986), M.Eng (1988) and Ph.D. (2010) degrees, all in electronic engineering, from Stellenbosch University. He worked as researcher at DataFusion (later called Spescom DataVoice) and is currently chief scientist at AGNITIO. Most of his research for the last two decades has been applied to automatic speaker and language recognition and he has been participating in most of the NIST

SRE and *LRE* evaluations in these technologies, from the year 2000 to the present. He has been contributing to the Odyssey Workshop series since 2001 and was organizer of Odyssey 2008 in Stellenbosch. His *FoCal Toolkit* is widely used for fusion and calibration in speaker and language recognition research.

His research interests include development of new algorithms for speaker and language recognition, as well as evaluation methodologies for these technologies. In both cases, his emphasis is on probabilistic modelling. He has worked with both generative (eigenchannel, JFA, i-vector PLDA) and discriminative (system fusion, discriminative JFA and PLDA) recognizers. In evaluation, his focus is on judging the goodness of classifiers that produce probabilistic outputs in the form of well calibrated class likelihoods.

Plenary Session 2

Dr Li Deng
Principal Researcher
Microsoft Research, USA

Being Deep and Being Dynamic --- New-Generation Models and Methodology for Advancing Speech Technology

An APSIPA Distinguished Lecture 2012

Semantic information embedded in the speech signal --- not only the phonetic/linguistic content but also a full range of paralinguistic information including speaker characteristics --- manifests itself in a dynamic process rooted in the deep linguistic hierarchy as an intrinsic part of the human cognitive system. Modeling both the dynamic process and the deep structure for advancing speech technology has been an active pursuit for over more than 20 years, but it is not until recently (since only a few years ago) that noticeable breakthrough has been achieved by the new methodology commonly referred to as "deep learning". Deep Belief Net (DBN) is recently being used to replace the Gaussian Mixture Model (GMM) component in HMM-based speech recognition, and has produced dramatic error rate reduction in both phone recognition and large vocabulary speech recognition while keeping the HMM component intact. On the other hand, the (constrained) Dynamic Bayesian Net (referred to as DBN* here) has been developed for many years to improve the dynamic models of speech while overcoming the IID assumption as a key weakness of the HMM, with a set of techniques and representations commonly known as hidden dynamic/trajectory models or articulatory-like models. A history of these two largely separate lines of "DBN/DBN*" research will be critically reviewed and analyzed in the context of modeling deep and dynamic linguistic hierarchy for advancing speech (as well as speaker) recognition technology. Future directions will be discussed for this exciting area of research that holds promise to build a foundation for the next-generation speech technology with human-like cognitive ability.

Short Biography

Li Deng received the Ph.D. from Univ. Wisconsin-Madison. He was an Assistant (1989-1992), Associate (1992-1996), and Full Professor (1996-1999) at the University of Waterloo, Ontario, Canada. He then joined Microsoft Research, Redmond, where he is currently a Principal Researcher and where he received Microsoft Research Technology Transfer, Goldstar, and Achievement Awards. Prior to MSR, he also worked or taught at Massachusetts Institute of Technology, ATR Interpreting Telecom. Research Lab. (Kyoto, Japan), and HKUST. He has published over 300 refereed papers in leading journals/conferences and 3 books covering broad areas of human language technology, machine learning, and audio, speech, and signal processing. He is a Fellow of the Acoustical Society of America, a Fellow of the IEEE, and a Fellow of the International Speech Communication Association. He is an inventor or co-inventor of over 50 granted US, Japanese, or international patents. He served on the Board of Governors of the IEEE Sig. Proc. Soc. (2008-2010). More recently, he served as Editor-in-Chief for IEEE Signal Processing Magazine (2009-2011), which, according to the Thompson Reuters Journal Citation Report released 2010 and 2011, ranked first in both years among all 127 IEEE publications and all 247 publications within the Electrical and Electronics Engineering Category worldwide in terms of its impact factor, and for which he received the 2011 IEEE SPS Meritorious Service Award. He currently serves as Editor-in-Chief for IEEE Transactions on Audio, Speech and Language Processing. His recent tutorials on deep learning at APSIPA (Oct 2011) and at ICASSP (March 2012) received the highest attendance rate at both conferences.

Plenary Session 3

Dr Alvin Martin
Mathematician
NIST, USA

The NIST Speaker Recognition Evaluations

Since 1996 the National Institute of Standards and Technologies has coordinated a series of annual or bi-annual open evaluations of automatic speaker recognition technology. These have concentrated on the task of single speaker detection in the context of spontaneous speech of a conversational telephone or one-on-one interview situation, recorded over ordinary telephone channels or room microphones. System performance has been assessed in relation to a variety of factors, including notably the quantity of training and test speech supplied, the speech styles being used, and the types and variability of the recording channels. While English has been the primary language employed, several of the evaluations have included substantial quantities of speech by multi-lingual speakers to allow examination of language and cross-language effects. More recently, initial efforts have been made to consider the effects of voice aging and varying vocal effort on performance. We discuss the considerations that have gone into planning and organizing these and a few related evaluations, the performance metrics that have been employed, the considerable progress observed over time, and the ongoing plans for further evaluation in 2012 and beyond.

Short Biography

Alvin Martin served as a mathematician in the Multimodal Information Group at the National Institute of Standards and Technology from 1991 through 2011. He has coordinated NIST's series of evaluations since 1996 in the areas of speaker recognition and of language and dialect recognition, and has contributed to its evaluations of large vocabulary continuous speech recognition. This work has involved the collection, selection, and pre-processing of appropriate speech data, the writing of evaluation plans, the specification of metrics and charts for the scoring, presentation, and analysis of results, the implementation of statistical tests for determining the significance of performance differences, and the organization of workshops to review evaluation results.

He received a Ph.D. in mathematics from Yale University (1977), has taught mathematics and computer science at the college level, and worked on the development of automatic speech recognition and speech processing systems before coming to NIST.

Proceedings of Odyssey 2012: The Speaker and Language Recognition Workshop
Odyssey 2012, Singapore

Published by Chinese and Oriental Languages Information Processing Society (COLIPS), Speaker and Language Characterization SIG

ISBN: 978-981-07-3093-2, Editors: Haizhou Li, Bin Ma, Kong Aik Lee

Odyssey 2012

The Speaker and Language Recognition Workshop

22 June – 28 June 2012, Singapore

Organizing Committee

- *Haizhou Li,* Institute for Infocomm Research, Singapore, chair
- *Lukáš Burget,* SRI, USA
- *Jan "Honza" Cernocký,* Brno University of Technology, Czech Republic
- *Jean-François Bonastre,* University of Avignon, France
- *Niko Brümmer, Agnitio,* South Africa
- *Joseph Campbell,* MIT Lincoln Lab, USA
- *Douglas Reynolds,* MIT Lincoln Lab, USA
- *Alvin Martin,* NIST, USA
- *Kay Berkling,* Inline Internet Online Dienste GmbH, Germany

Scientific Committee

- *Eliathamby Ambikairajah,* U. of New South Wales, Australia
- *Tim Anderson,* AFRL, USA
- *Walt Andrews,* BBN, USA
- *Aladdin Ariyaeeinia,* U. of Hertfordshire, UK
- *Hagai Aronowitz,* IBM Haifa Research Lab, Israel
- *Roland Auckenthaler,* NMS Comm., USA
- *Claude Barras,* LIMSI, France
- *Laurent Besacier,* CLIPS-IMAG, France
- *Bill Campbell,* MIT Lincoln Lab, USA
- *Mauro Falcone,* FUB, Italy
- *Kevin Farrell,* Nuance, USA
- *Sadaoki Furui,* Tokyo Tech, Japan
- *Ran Gazit,* General Motors, Israel

- *Joaquin Gonzalez-Rodriguez,* U.A. de Madrid, Spain
- *Craig Greenberg,* NIST, USA
- *Hynek Hermansky,* JHU, USA and BUT, Czech Republic
- *Patrick Kenny,* CRIM, Canada
- *Tomi Kinnunen,* U. of East Finland, Finland
- *Tina Kohler,* DoD, USA
- *Pietro Laface,* Politecnico di Torino, Italy
- *Tomoko Matsui,* Institute of Statistical Mathematics, Japan
- *Jiri Navratil,* IBM, USA
- *Jason Pelecanos,* IBM, USA
- *Daniel Ramos,* U. A. de Madrid, Spain
- *Nicolas Scheffer,* SRI, USA
- *Andreas Stolcke,* Microsoft Speech Labs, USA
- *Doug Sturim,* MIT Lincoln Lab, USA
- *Pedro Torres-Carrasquillo,* MIT Lincoln Lab, USA
- *David van Leeuwen,* Radboud University Nijmegen, The Netherlands
- *Robbie Vogt,* Queensland U.of Technology, Australia
- *Michael Wagner,* U. of Canberra, Australia

Local Organizing Committee

- *Haizhou Li,* Institute for Infocomm Research, Singapore
- *Bin Ma,* Institute for Infocomm Research, Singapore
- *Kong Aik Lee,* Institute for Infocomm Research, Singapore
- *Swee Lan See,* Institute for Infocomm Research, Singapore
- *Cheung-Chi Leung,* Institute for Infocomm Research, Singapore
- *Anthony Larcher,* Institute for Infocomm Research, Singapore
- *Xiong Xiao,* Temasek Laboratories, Nanyang Technological University, Singapore
- *Nancy Chen,* Institute for Infocomm Research, Singapore
- *Changhuai You,* Institute for Infocomm Research, Singapore
- *Hanwu Sun,* Institute for Infocomm Research, Singapore
- *Rong Tong,* Institute for Infocomm Research, Singapore

Reviewing Committee

Eliathamby Ambikairajah, Tim Anderson, Walt Andrews, Aladdin Ariyaeeinia, Hagai Aronowitz, Roland Auckenthaler, Claude Barras, Kay Berkling, Jean-François Bonastre, Niko Brümmer, Lukáš Burget, Bill Campbell, Joseph Campbell, Jan "Honza" Cernocký, Nancy Chen, Nicholas Evans, Mauro Falcone, Kevin Farrell, Sadaoki Furui, Ran Gazit, Joaquin Gonzalez-Rodriguez, Craig Greenberg, Ville Hautamaki, Hynek Hermansky, Chien-Lin Huang, Patrick Kenny, Tomi Kinnunen, Tina Kohler, Pietro Laface, Anthony Larcher, Kong Aik Lee, Cheung-Chi Leung, Haizhou Li, Bin Ma, Alvin Martin, John Mason, Driss Matrouf, Tomoko Matsui, Jiri Navratil, Raymond W. M. Ng, Jason

Pelecanos, Daniel Ramos, Douglas Reynolds, Nicolas Scheffer, Andreas Stolcke, Doug Sturim, Hanwu Sun, Rong Tong, Pedro Torres-Carrasquillo, David van Leeuwen, Michael Wagner, Changhuai You.

Proceedings of Odyssey: The Speaker and Language Recognition Workshop
Odyssey 2012, Singapore

Published by Chinese and Oriental Languages Information Processing Society (COLIPS), Speaker and Language Characterization SIG

ISBN: 978-981-07-3093-2, Editors: Haizhou Li, Bin Ma, Kong Aik Lee