• Plenary 1 (10:00-11:00 Wednesday 17th December 2008)

Speech-To-Speech Translation Technologies for Real-World Applications

Dr. Yuqing Gao
J. Watson Research Center

In this talk, the speaker will briefly introduce the background of Speech-to- Speech Translation, by reviewing related projects and state-of-the-art speech translation technologies and approaches, as well as the history of the IBM Multilingual Automatic Speech-To-Speech TranslatOR (MASTOR) system.
The speaker will present an overview of IBM system framework, and various approaches that the IBM team developed under DARPA CAST and TransTac programs, which led the IBM team to the successes of developing and deploying from research prototypes to real world deployment. The technologies the speaker will cover include maximum-entropy (ME)-based statistical Natural Language Understanding and Generation approach, algorithms for colloquial speech recognition and very fast machine translation, algorithms for rapid development of low resource languages, algorithms for low computation resource devices, and scalable algorithm and system development for multiple platforms for real-world applications.

• Plenary 2 (11:00-12:00 Wednesday 17th December 2008)

What Can Speech Researchers Bring to Music Processing?

Prof. Shigeki Sagayama
Graduate School of Information Science and Technology, The University of Tokyo

The speech research community has developed powerful approaches which are potentially applicable to other technological areas. As music is the counterpart of speech in the sense of them being the two most important information-rich categories of acoustic signals understood by humans, music processing can be a good application target of speech technologies. Recently, music technology research has been growing rapidly, fueled by a high demand in music entertainment and a general need for music information retrieval. Since the speaker started working on music processing research in 1998, he has been continuously seeking good models and algorithms for music processing inspired by speech technologies, as well as new solutions to music-specific problems which will hopefully help speech processing in the future.
The speaker will give some examples from his and his colleagues' recent research activities, where speech processing algorithms play an important role in music processing both for audio and symbolic (typically, MIDI) music inputs. Applications of HMMs (Hidden Markov Models), DP (Dynamic Programming) and their generalizations: Dynamic Bayesian Networks (DBNs) include chord and key modulation detection, music transcription, harmonization of given melodies, counterpoint, rhythm recognition, score following, song composition from given lyrics, piano fingering, etc. Applications of Gaussian mixtures and the EM algorithm include multiple F0 estimation, precise onset detection, sound separation in polyphonic music, deletion and modification of notes and reconstruction of missing parts in audio signals, etc. Language

modeling approaches are applicable to musicological analysis and harmonization of melodies. Research in music also motivates us to develop music-specific acoustic signal processing methods such as Non-negative Matrix Factorization (NMF) for music transcription, harmonic/percussive sound separation and microphone array techniques for music signal separation.

• Plenary 3 (8:30-9:30 Thursday 18th December 2008)

Speech and Search: Bridging The Gap

Dr. Vincent Vanhoucke
Google411

There are fantastic challenges in integrating speech technologies into search. The power of web search relies overwhelmingly on keyword spotting and distributed information retrieval over large, unstructured databases. Syntactic and semantic models contribute very weakly to this picture. In contrast, the success of speech technologies has been driven to a large extent by the recognition that strong language models are essential to designing accurate systems. Reconciling these two pictures is an enormous opportunity, which enables both worlds to significantly leverage each other's assets: indexing spoken content broadens the reach of search engines, while exposing indexed content to voice interfaces contributes significantly to making the world's information more accessible to everyone. To illustrate both points, the speaker will discuss the computational and algorithmic challenges of transcribing and indexing the huge amounts of spoken data available online. He will also examine how GOOG-411, Google's business search by phone, leverages both spoken and online data to bring a consistent, useful search experience to every phone user.

• Plenary 4 (8:30-9:30 Friday 19th December 2008)

Towards Robust Speech Recognition: Structured Modeling, Irrelevant Variability Normalization and Unsupervised Online Adaptation

Dr. Qiang Huo
Microsoft Research Asia

In the past several years, we've been studying several approaches to robust automatic speech recognition (ASR) based on three key concepts, namely structured modeling, irrelevant variability normalization (IVN) and unsupervised online adaptation (OLA). In structured modeling of basic speech units, speech information relevant to phonetic classification is modeled by traditional hidden Markov models (HMMs), while factors irrelevant to phonetic classification are taken care of by an auxiliary module. An IVN-based training procedure can then be designed to estimate parameters of the generic HMMs and the auxiliary module from a large amount of diversified training data. In recognition stage, the parameters of the auxiliary module can be updated via unsupervised OLA by using the unknown utterance itself, which is