# Speech Recognition and Understanding

Alex Acero
Microsoft Research

## Tutorial Outline

1. System Overview: description of the different building blocks in a spoken language system.
2. Signal Processing: feature extraction and noise robustness.
3. Hidden Markov Models, including:
    - Discrete and continuous HMM.
    - Training procedures: maximum-likelihood (EM, Baum-Welch), discriminative training, and parameter smoothing.
    - Decoding (Viterbi) algorithms.
4. Acoustic Modeling, including:
    - Isolated vs. continuous speech.
    - Phone-based vs. word-based recognition.
    - Context-dependent vs. context independent
    - Speaker-dependent vs. speaker-independent.
    - Adaptation (MAP and MLLR) and confidence measures.
5. Language Modeling: use of context-free grammars and n-grams as language models for speech recognizers.
6. Search Algorithms for ASR: efficient search algorithms are needed to evaluate the millions of hypotheses required in large vocabulary ASR.
7. Speech Understanding: information extraction from speech (and text) for limited domains.
8. Systems, Applications and User Interface: applications of the technology sprinkled with a few demos.

## Abstract

This tutorial will introduce the main concepts behind modern speech recognition and understanding systems. The state-of-the-art and the assumptions and limitations of current technology will be presented. The emphasis will be on describing an end-to-end system and how the different components fit together. No background in speech technology will be assumed.