

# Fundamentals of Speech Recognition

## E6998

### *Instructor:*

Prof. Homayoon Beigi <[beigi@recotechnologies.com](mailto:beigi@recotechnologies.com)> (hb87@columbia.edu)

### *Textbook:*

H. Beigi, "Fundamentals of Speaker Recognition," Springer, New York 2011.

### *Grading:*

Homework (20%):

- Implementation of a speech recognition engine using the Tedlium example of Kaldi.
- Creation of a Flowchart with a paragraph for each block in the flowchart, describing the whole process in the Tedlium example.
- Results of the decoding.

Midterm Proposal (20%):

15% - 2-page extended abstract describing the results and proposing modifications to one specific part of the engine to increase performance (accuracy, speed, or both)

5% - 5 minute presentation of the above.

Final Project (60%):

45% - 6-page IEEE conference style paper describing the system and results obtained from the modification. Discussion and Implementation of an Improvement in one of the aspects of the speech recognition engine.

10% - Code and Results.

5% - 5 minute presentation of the results.

### *Course Description:*

Fundamentals of Speech Recognition is a comprehensive course, covering all aspects of automatic speech recognition from theory to practice. In this course such topics as Anatomy of Speech, Signal Representation, Phonetics and Phonology, Signal Processing and Feature Extraction, Probability Theory and Statistics, Information Theory, Metrics and Divergences, Decision Theory, Parameter Estimation, Clustering and Learning, Transformation, Hidden Markov Modeling, Language Modeling, Neural Networks (specifically TDNN, LSTM, RNN, and CNN architectures) plus other recent machine learning techniques used in speech recognition are covered in some detail. Also, several open source speech recognition software packages are introduced, with detailed hands-on projects using Kaldi to produce a fully functional speech recognition engine. The lectures cover the theoretical aspects as well as practical coding techniques. The course is graded based on a project. There will be one homework project worth 20%, a Midterm proposal (20% of the grade is in the form of a two page proposal for the project and the final (60% of the grade) is an oral presentation of the project plus a 6-page conference style paper describing the results of the research project. The instructor uses his own Textbook for the course, Homayoon Beigi, "Fundamentals of Speaker Recognition," Springer-Verlag, New York, 2011. Every week, the slides of the lecture are made available to the students.

### *Research Projects:*

Individual projects are done using Kaldi, and picked from topics of interest to the students such as,  
- Large Vocabulary Speech Recognition

- Keyword and Hotword recognition
- Speaker Recognition
- Emotion Detection
- Sequence-to-sequence modeling

## *Lectures:*

### *Week 1*

- Introduction (Overview of Speaker Recognition and its history)
- The Anatomy of Speech
  - The Human Vocal System
  - The Human Auditory System
  - The Nervous System and the Brain

### *Week 2*

- Signal Representation of Speech
  - Sampling The Audio
  - Quantization and Amplitude Errors
  - Practical Sampling and Associated Errors

### *Week 3*

- Phonetics and Phonology
  - Phonetics
  - Phonology and Linguistics
  - Suprasegmental Features of Speech

### *Weeks 4 & 5*

- Signal Processing of Speech and Feature Extraction
  - Auditory Perception
  - The Sampling Process
  - Spectral Analysis and Direct Method Features
  - Linear Predictive Cepstral Coefficients (LPCC)
  - Perceptual Linear Predictive (PLP) Analysis
  - Alternative Cepstral-Based Features
  - Other Features
  - Signal Enhancement and Pre-Processing

### *Week 6*

- Decision Theory
  - Hypothesis Testing
  - Bayesian Decision Theory
  - Bayesian Classifier
  - Decision Trees
- Parameter Estimation
  - Maximum Likelihood Estimation (MLE, MLLR, fMLLR)
  - Maximum A-Posteriori (MAP) Estimation
  - Maximum Entropy Estimation
  - Minimum Relative Entropy Estimation
  - Maximum Mutual Information Estimation (MMIE)
  - Model Selection (AIC and BIC)

### *Weeks 7, 8, & half of 9*

- Neural Networks
  - Perceptron
  - Feedforward Networks

Time-Delay Neural Networks (TDNN)  
Convolutional Neural Networks (CNN)  
Recurrent Neural Networks (RNN)  
Long-Short Term Memory Networks (LSTM)  
End-to-End Sequence (Encoder/Decoder) Neural Networks  
Embeddings and Transfer Learning

#### *Weeks second half of 9 & 10*

- Probability Theory and Statistics
  - Measure Theory
  - Probability Measure
  - Integration
  - Functions
  - Statistical Moments
  - Discrete and continuous Random Variables
  - Moment Estimation
  - Multi-Variate Normal Distribution
- Language Modeling
  - NGram Language Modeling
  - Class-Based NGrams
  - Recurrent Neural Network Language Model (RNNLM)
  - Finite State Transducers

#### *Week 11*

- Unsupervised Clustering and Learning
  - Vector Quantization (VQ)
  - Basic Clustering Techniques
  - Estimation using Incomplete Data
- Transformation
  - Principal Component Analysis (PCA)
  - Linear Discriminant Analysis (LDA)
  - Factor Analysis (FA)
  - Probabilistic Linear Discriminant Analysis (PLDA)

#### *Week 12*

- Information Theory
  - Sources
    - The Relation between Uncertainty and Choice
    - Discrete Sources
    - Discrete Channels
    - Continuous Sources
  - Relative Entropy
  - Fisher Information
  - Metrics and Divergences
- Hidden Markov Modeling (HMM)
  - Memoryless Models
  - Discrete Markov Chains
  - Markov Models
  - Hidden Markov Models
  - Model Design and States
  - Training and Decoding
  - Gaussian Mixture Models (GMM)
  - Practical Issues