

Syllabus: DEPARTMENT OF ELECTRICAL ENGINEERING
University of Washington
Spring Quarter 2009

Course: EE 516

Title: Computer Speech Processing

Credits: 4

Course Web Site: <http://ssli.ee.washington.edu/courses/ee516/>

Course Description: Introduction to automatic speech processing using digital techniques. Overview of human speech production and perception. Fundamental theory and practice in speech analysis (including spectrogram reading), enhancement, coding, synthesis, and recognition, as well as system design methodologies. Advanced topics include speaker/language identification and discriminative learning. DSP and stochastic processes are required as pre-requisites.

Recommended Texts and/or References:

1. Xuedong Huang, Alex Acero, Hsiao-Wuen Hon, Spoken Language Processing, Prentice-Hall, May 2001
2. Li Deng and D. O'Shaughnessy, Speech Processing --- A Dynamic and Optimization-Oriented Approach, Marcel Dekker Inc., New York, NY., June 2003.
3. EE 516 Reader

Lecture Time: Th 2:30–6:20pm in EEB 003

Instructors: Drs. Alex Acero, Li Deng, Geoff Zweig (Microsoft Research, One Microsoft Way, Redmond, WA 98052 {alexac|deng|gzweig}@microsoft.com)

Instructor Office Hours: 5:20-6:20 EEB 003

Feel free to request other times by email.

Projects: There will be a set of software-related projects designed for the class aimed to practice the class material and to develop the skills in implementing the algorithms discussed in the class. Each student can select one or two projects and each project will consist of a team of students.

Final Exam: open book; due toward to the end of the Spring term

Course Grading:

- Attendance: required, with up to 2 absences allowed.
- Projects: 50% ; Final Exam: 30%; Homework: 20%

Topics to be covered in the lectures:

- 4/2: Introduction and overview; Highlight of projects [Zweig; Deng, Acero]
- 4/9: Speech Analysis, Production and Perception [Deng]
 - Huang, Acero & Hon, Spoken Language Processing, Chapter 6: Speech Signal Representation, pp.275-332
 - Deng & O'Shaughnessy, Speech Processing --- A dynamic & optimization-oriented approach, Chapter 2 (pp. 29-64); Chapter 7.1-7.4 (pp. 203-232), 7.6-7.9 (pp. 238-261)
 - **Homework/Project:** plan of writing a program in Matlab to implement a simple formant tracker; learn to use the LPC tool in Matlab.
- 4/16 : Speech Spectrogram Reading [Deng]
 - Deng & O'Shaughnessy, Speech Processing --- A dynamic & optimization-oriented approach, Chapter 7.5; Chapter 9.5 (pp. 319-332); some supplementary materials
 - Deng, Alwan et. al. "A Database of Vocal Tract Resonance Trajectories for Research in Speech processing," ICASSP 2006.
 - Introduction of the project
 - Deng, Alwan: User Manual for MSR-UCLA VTR-Formant Database
- 4/23 : Vector Quantization and EM [Zweig]
 - Rabiner & Juang, *Fundamentals of Speech Recognition*, Section 3.4, pp. 122-132. 1993.
 - Huang, Acero & Hon, Spoken Language Processing, Chapter 4, pp. 133-197. 2001.
 - **Homework:** Given a set of two-dimensional vectors, write a VQ program. Extend that to soft-clustering and then do full EM for mixture of Gaussians. Either in MATLAB or C++.
(Due 5/7)
- 4/30: Language Modeling [Acero]
 - Huang, Acero & Hon, Spoken Language Processing, Chapter 11: Language Modeling, pp.545-590
 - **Homework:** Write a program to estimate the Entropy of written text (Shannon, 1950). Given a text document, write a program that will predict the next letter given the past letters on a different text. (Due 5/21)
- 5/7 : Speaker and Language Identification [Zweig]
 - "Robust Text-Independent Speaker Identification using Gaussian Mixture Speaker Models" Reynolds et al., IEEE TSAP, v. 1, n. 3, 1995, pp. 72-82.
 - "Speaker Verification using Adapted Gaussian Mixture Models" Reynolds et al. Digital Signal Processing 10 19-41 (2000)
 - "SVM Based Speaker Verification Using a GMM Supervector Kernel and NAP Variability Compensation" Campbell et al., ICASSP 2006
 - "The SuperSID Project: Exploiting High-Level Information for High-Accuracy Speaker Recognition" Reynolds et al., ICASSP 2002.
 - "The NIST Year 2008 Speaker Recognition Evaluation Plan"
<http://www.itl.nist.gov/iad/mig/tests/>
 - "Comparison of Four Approaches to Automatic Language Identification of Telephone Speech" M. Zissman. IEEE TSAP v. 4, n. 1, pp. 31-44. 1996.
 - "Language Identification using Gaussian Mixture Model Tokenization" Torres-Carrasquillo et al., ICASSP 2002.
 - "A Vector Space Modeling Approach to Spoken Language Identification" Li et al., IEEE TSALP, v. 15, n. 1, pp. 271-284. 2007.
 - "An Empirical Study of Automatic Accent Classification" Choueiter et al., ICASSP 2007.

- “The 2007 NIST Language Recognition Evaluation Plan”
<http://www.itl.nist.gov/iad/mig/tests/>
- 5/14: Speech Recognition: Dynamic Programming and Discrete HMMs [Zweig]
 - “A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition”, L. Rabiner. Proc. IEEE V. 77, n. 2, 1989, pp. 257-286.
 - **Homework:** Write a program to compute word error rate (Due 5/28)
- 5/21: Speech Synthesis [Acero]
 - Huang, Acero & Hon, Spoken Language Processing, Chapters 14, 15, and 16.
- 5/28: Speech Coding and Enhancement [Acero]
 - Huang, Acero & Hon, Spoken Language Processing, Chapters 7 and 10.
- 6/4: Discriminative Training [Deng]
 - X. He, Li Deng, Chou Wu, Discriminative Learning in Sequential Pattern Recognition --- A Unifying Review for Optimization-Oriented Speech Recognition, IEEE Signal Processing Magazine, vol. 25, no. 5, pp. 14-36, Sep. 2008

Project Possibilities:

- Write a speaker verification system that enrolls speakers and determines whether someone is who they claim to be. (This will involve finding a number of people to donate speech.)
- Implement one or more of the Language ID methods using the CallFriend database.
- Write a digit decoding system – implement the DP recursions for Viterbi and EM, and train the system on the Aurora database.
- Write a program in Matlab to implement one or more formant trackers. Visual examination of the quality of the program’s outputs using a tool that overlay the outputs on the spectrogram display of the input speech sentences (TIMIT). Use the same tool (with user interface and data input/output calls) to correct the errors due to the imperfection of the tracking algorithms (not due to programming errors) using the knowledge learned in the speech analysis and spectrogram reading classes.