EE516, Computer Speech Processing

Jeffrey Bilmes

Winter 2018

Department of Electrical Engineering

University of Washington

Summary

Prerequisites	:	EE518, EE505 or equivalent probability course
Time	:	Tuesdays 6:00–9:00 pm in EEB 045
Instructor	:	Jeffrey Bilmes
TA	:	Dhanush B K
Office hours	:	Prof. Bilmes - Mondays 9.00pm to 10.00pm on zoom Dhanush - Wednesdays 8.00pm to 9.00pm on hangouts
Website	:	Canvas, https://canvas.uw.edu/courses/1188387
Grading	:	Homework: 40 % Participation: 20 % Final project: 40 % - will be further split as follows - 30% participation, 35% degree of code working, 10% based on video demo, and 25% report

Course Outline

Lec 1 : Logistics, Overview of Speech Processing, Synthesis, and Recognition
Lec 2 : Speech Production, Phonetics
Lec 3 : Speech Perception, The Ear, Acoustics, the Brain
Lec 4 : Speech Signal Processing
Lec 5 : Front end Speech Processing, Feature Extraction
Lec 6 : Pattern Recognition and Machine Learning for Speech Recognition
Lec 7 : Dynamic Time Warping and HMMs
Lec 8 : More Dynamic Time Warping and HMMs
Lec 9 : From HMMs to DNNs, why are DNNs better?
Lec 10 : DNNs, RNNs, LSTMs, end-to-end training

Homework

4 python/theory based homeworks every 2 weeks. Will be posted on canvas. PDFs to be submitted to gradescope (link will be made available later).

Project

The final project will consist of writing in python from scratch (all code your own) a simple speech recognition system. The system will be at least: a speaker dependent, isolated-phrase, whole-word model (i.e., one HMM/phrase), Gaussian-mixture per state, diagonal covariance Gaussian, speech recognition system, using a five command-phrase vocabulary plus the wakeup phrase "Odessa.". Your system will monitor the environment with an acoustic energy detector module (also using other inexpensive heuristics). Once it detects a phrase, it will check if it is the wake up phrase. If it it is the wake up phrase, it will (after a presumed silence) detect which of the five command phrases. The five command phrases are: "Turn on the lights", "Turn off the lights", "What time is it", "Play music", and "Stop music." and should indicate it understood the right phrase. If a phrase other than the five command phrases are uttered, system should respond "Sorry, I'm not sure"

What will be due (and what is graded):

1. 5 minute conference-style presentation on and demo of your project working (you need to bring your own laptop for this).
2. Presentation slides, turned in electronically via canvas.
3. A 4-page writeup summarizing the presentation, the work, the code, issues, what worked, what failed.
4. All of the code, which must be in python.
5. Preliminary deadlines over the weeks leading up to the final project in the form of project summaries, progress reports, and the supporting homework assigbments.

All homeworks and project will be due electronically via our assignment dropbox, no dead-tree assignments accepted. Final presentations will be Tuesday, March 13th, 2015, 6-9pm. No regular midterm and final exam.

Recommended Reading

1. Chapters 1 and 2 in our book (Huang, Acero, Hon, "Spoken Language Processing")
2. Classic paper: A Review of text-to-speech conversion for English, Dennis H. Klatt https://pdfs.semanticscholar.org/5657/f5888e198fecf4612ff04c4b0bdef972147c.pdf.
3. P. Taylor, Text-to-Speech Synthesis, Cambridge University Press, New York, NY, USA, 1st edition, 2009.
4. "What HMMs can do", J. Bilmes, http://melodi.ee.washington.edu/~bilmes/mypubs/bilmes2006-ieice-hmm.pdf
5. Example recent end-to-end speech recognition model, https://arxiv.org/abs/1712.01769