

## Handout 1: Course Overview

Prof: J. Bilmes <bilmes@ee.washington.edu>

Jan 3, 2005

This course introduces graduate students to the field of computer speech processing, primarily speech recognition (or text-to-speech) by machine. It aims to provide a theoretical foundation and practical experience for those who are interested in doing research in this field, as well as an understanding of the current state of the art for those interested in using speech technology in applications. It will also be of interest for anyone who wishes to understand how computer-based speech recognition systems work.

Students who complete this course should gain an understanding of basic theory supporting speech technology, knowledge of basic system design principles and trade-offs associated with different methods, and should gain the ability to read and discuss technical papers in speech processing.

Starting with a brief overview of human speech production and perception, it will cover fundamental theory in all aspects of statistical speech recognition (and basic synthesis), as well as basic speech coding and overall system design methodologies. It will cover the methodologies behind speech recognition systems, from simple isolated word systems to advanced large-vocabulary continuous speech recognition systems, and will cover some of the most recent work in this endeavor. It will also touch on more advanced topics like speaker and language identification and adaptation.

This is a lab and project course with an emphasis on the application of algorithms and design principles to actual speech processing problems. The labs will be used to study examples of the problems which are discussed in the lecture, and the project will be an open-ended small group effort. We emphasize technical communication skills by requiring a written report and oral presentation for the final project.

**Lecturer:** Prof. Jeff Bilmes 418 EE/CS Bldg. <bilmes@ee.washington.edu> Office Hours Tuesdays, 2:00-4:00pm. Lectures will be MW 2:30-4:00 in EE1-042. 4 units.

**Course web page:** <http://ssli.ee.washington.edu/people/bilmes/ee516>

**Texts:** The main text will be “Spoken Language Processing”, by X.D. Huang, A. Acero, H. Hon, Prentice Hall. We will also use other handouts that will either be passed out in class, or will be available on the course web page (copyright permitting). Other texts that will be drawn from are listed below.

**Prerequisites:** Basic probability (e.g., STAT390, EE505, etc.), exposure to digital signal processing (such as EE518, EE596, etc.), basic machine learning/pattern recognition (e.g., EE517/518), and general algorithm design. Alternatively, permission of instructor.

**Scribe:** Each lecture will have a student *scribe* for the day, where the student is responsible for taking careful class notes. Each student must do this for at least one class lecture but it could be more (the number of lectures that each student will be responsible for depends on the number of students enrolled in the course). These notes are to be converted into computer readable form (including figures), and will be due the following lecture. All scribes must be done using the L<sup>A</sup>T<sub>E</sub>X document processing system. If you don’t know L<sup>A</sup>T<sub>E</sub>X, now is the time to learn. A L<sup>A</sup>T<sub>E</sub>X scribe template is available on the course web page.

The scribe will be also responsible for any corrections the instructor makes to the text both in terms of clarity and completeness. When finished, the scribe notes will be made available on the course web page for other students to use. The grade given for each scribe notes will be based on two things: 1) it will be

related to the degree to which you find the references for the material (that might mean searching the web, going to the library, making photocopies, etc. I will make available all relevant reference citations), 2) it will be inversely proportional to the amount of time I need to spend fixing the scribe notes that you turn in (English grammar is important also. If you have trouble with this, please find someone to proof-read your scribe).

**Homework:** There will be periodic homework assigned in this class (there will be anywhere between 2-5 different assignments this quarter). These problem sets will be a combination of both typical problems, and lab problems where you might be asked to use speech tools such as HTK, ESPS, and/or matlab. PDF-only homework solutions are accepted (and encouraged actually), and can be emailed directly to the instructor. It is also fine (and encouraged) to turn in the HW as a web page.

**Exam:** There will be one exam in this course, roughly half-way through.

**Project:** A significant portion of your grade (50%) will be based on the class project. Each person must do the project by themselves. We only have 10 weeks, so we will start early on projects. There will be some milestones that must be met along the way.

**Grading and Exams:** Grades will be based on a combination of the homework (20%), scribe quality (30%), the midterm exam (15%) and the final project held on the last day of class (35%). The lecture on the last day of class might run a bit longer than 2 hours. If you are registered for the course as S/NS, then you must do a good quality scribe and midterm exam to get a passing grade (i.e., you need not do the homework or the exam). Please let me know ASAP if you registered for the course as S/NS.a

**Other Useful Texts** which we will draw from, and which you might find useful in general (and which ultimately will be placed on reserve).

- B. Gold and N. Morgan, *Speech and Audio Signal Processing*, Wiley Publishers, 2000
- F. Jelinek, *Statistical Methods for Speech Recognition*, MIT Press, 1997
- D. O'Shaughnessy, *Speech Communications: Human and Machine*, 2nd edition, Addison-Wesley, 1999
- T. F. Quatieri, *Discrete-Time Speech Signal Processing: Principles and Practice*, Prentice Hall Signal Processing Series, 2002.
- D. Jurafsky & James H. Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*, Prentice Hall, 2000
- W. B. Kleijn and K. K. Paliwal, ed., *Speech Coding and Synthesis*, Elsevier, 1995.
- L. Rabiner and R. Schafer, *Digital Processing of Speech Signals*, Prentice Hall, 1978
- L. Rabiner and B.-H. Juang, *Fundamentals of Speech Recognition*, Prentice Hall, 1993
- J.R. Deller, J.G. Proakis and J.H.L. Hansen, *Discrete-Time Processing of Speech Signals*, MacMillan, 1993 (this was republished by IEEE Press in 1999).
- R. Duda and P. Hart, "Pattern Classification and Scene Analysis," Wiley Interscience, 2001 edition (the older version from 1973 is still valid as well).
- T. F. Quatieri, "Discrete-Time Speech Signal Processing: Principles and Practice," Prentice Hall, 2002.

**Rough class outline** 20 lectures total. We may cover a few more topics than what is listed below, but this will give you the general idea.

- Lec 1. Mon, Jan 3rd, Overview of the course.
- Lec 2. Wed, Jan 5th, Speech communication; history, mechanisms anatomy, and physiology of speech production; phonemics and phonetics.
- Lec 3. Mon, Jan 10th, Signal processing models of vocal system, acoustic tube model, wave equation.
- Lec 4. Wed, Jan 12th, the ear, the auditory periphery, and speech perception
- Lec 5. Wed, Jan 19th, (Mon, Jan 17th is MLK day holiday) speech analysis methods, filter banks, LPC, Levinson/Durbin
- Lec 6. Mon, Jan 24th, cepstral analysis of speech
- Lec 7. Wed, Jan 26th, Overview of pattern classification, information theory, and graphical descriptions of random processes
- Lec 8. Mon, Jan 31st, ASR speech recognition intro, feature extraction methods, basic acoustic modeling (Gaussians and/or Neural Nets)
- Lec 9. Wed, Feb 2nd, variable length sequences, DTW
- Lec 10. Mon, Feb 7th, DTW $\Rightarrow$ HMMs, why hide hidden markov models?
- Lec 11. Wed, Feb 9th, Text processing and prosody prediction (tentative lecture by Mari Ostendorf)
- Lec 12. Mon, Feb 14th, Waveform/Concatenative synthesis (lecture by Mari Ostendorf)
- Lec 13. Wed, Feb 16th, HMMs, forward/backward, Viterbi decoding
- Lec 14. Wed, Feb 23rd, (Mon, Feb 21st is presidents day holiday) HMM issues, EM learning, different types of HMMs,
- Lec 15. Mon, Feb 28th, Pronunciation modeling, Language Modeling,
- Lec 16. Wed, March 2nd, Language Modeling, backoff, and smoothing methods, LVCSR
- Lec 17. Mon, Mar 7th, LVCSR and Discriminative Parameter Training Methods Viterbi beam search and pruning.
- Lec 18. Wed, Mar 9th, bi- and tri-gram decoding. Decoding, A\* decoding, Tree Lexicons, Tricks and Tips for LVCSR. other advanced methods
- Last Meeting: Tues, Mar 15th. 2:30-4:30. Final Project Presentations

### **Important Dates:**

**Final Project Presentations:** 2:30-4:20 p.m. Wednesday, March 15th (we will try to make this earlier, perhaps Friday, March 11th).