

## Syllabus

*Instructor: Prof. Eric Fosler-Lussier*

*Foundations of Spoken Language Processing*

**Time/Place:** TR 9:30–10:45 AM, DL 266

**Instructor:** Prof. Eric Fosler-Lussier

**Office:** 585 Dreese Labs, 292-4890, fosler@cis.ohio-state.edu

**Office Hours:** T 10:45-12:00, R 1-2

**Course Website:** <http://www.cis.ohio-state.edu/~fosler/courses/cis788/wi04>

### Course Description:

The extraction of the meanings of messages from spoken language remains a very challenging problem, particularly for so-called large vocabulary speech problems. In this seminar, we will discuss the problems of automatic speech recognition and spoken dialogue processing (primarily focusing on the former), starting from early papers on the topic and continuing through current research frontiers. Prior knowledge of automatic speech recognition systems is not assumed.

### Course Objectives:

Upon satisfactory completion of the course, the student will have gained:

- a thorough understanding of automatic speech recognition,
- familiarity with some spoken language understanding techniques, and
- experience in building an automatic speech recognition system using a standard toolkit.

**Prerequisites:** At least one of CIS 730, EE 600, Stat 528 or equiv, Ling 684, or permission of instructor

**Required Texts:** Jurafsky and Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, Prentice Hall, 2000. Additional reading will also be distributed.

**Recommended Texts:** J. Holmes and W. Holmes, *Speech Synthesis and Recognition*, Taylor and Francis, 2001. (Will be on reserve at SEL. A good intro-level to the topic; had I known about it sooner this likely would have been my textbook choice.)

F. Jelinek, *Statistical Methods for Speech Recognition*, MIT Press, 1998. (This is a bit more of a hard-core book for those who like mathematics; I've found it useful as a reference.)

**Expectations:** This course is designed as a reading seminar. Each week, we will explore a new topic in spoken language processing. Several times during the quarter you will be expected to present the readings and lead discussion of the topics. Participation in the discussion will be just as important as the presentations, so you are expected to read the required readings beforehand and make meaningful contributions. Two people will be assigned as respondents each week, and they will start off the post-presentation questions – however, everyone is expected to participate so be prepared even if you aren't the respondent! The respondent will be responsible for writing up a one-page summary of the discussion generated by their questions.

There will be no exams.

Besides the presentations of the readings, you will be required to participate in some sort of group or individual project concerning spoken language processing (with emphasis on *spoken*). You are encouraged

to participate in a group effort of building a speech recognizer using a standard toolkit. I have identified four toolkits that I would like to experiment with this quarter; we will try to build a recognizer for the same corpus using all of these toolkits. Since this is a bit experimental, I'm interested in a detailed report on the process of building the recognizer. I would particularly like to know in what went well and what was difficult in the process.

You have been (or will be) given access to the departmental linux servers for this project. We will be dividing up into groups next week after reading through some of the toolkit documentation and people express their preferences. I would like, if possible, for people in the same research group to work on different teams – this will enable each group to have members with different viewpoints and strengths). We will discuss the group projects next week.

We will set aside some time each week on Thursday to discuss progress.

I would also like to have you propose and execute a small experiment as a final project, using the system you have built. However, since I don't know the difficulty of the system-building phase, we will determine on a case-by-case basis whether this is appropriate.

If you choose to do a project outside of the recognizer-building project, then you'll need to clear it with me early on in the quarter, and you will still give project updates every week.

**Cheating:** (Standard disclaimer) Academic misconduct will not be tolerated in this class. I am required by the University to report all incidents of suspected misconduct to the Office of Academic Affairs. In particular, your presentations should not be plagiarized; you may include figures as appropriate, but you must develop the talk yourself.

**Grading Plan:** The following are the *approximate* weights for each of the work components of the course. Final grades will be determined from the distributions of weighted scores. I reserve the right to shift the weighting to reflect more effort in the project.

Presentation: 30%

You should prepare slides and/or handouts to guide other participants through the topic.

Participation: 30%

You should come to each class with a set of questions about the reading in order to further discussion. When a respondent, you need to have a few questions ready, and after the class provide a one-page summary of the discussion.

Project: 40%

The grade will be determined by the write-up, final presentation, self-evaluation, and peer-evaluation.

---

Syllabus: This is a tentative syllabus for the class. In particular, we will make a decision midway through the quarter in terms of what we would like to emphasize in weeks 8-9. If there are particular papers that you would like to read, let me know and we can try to work them into the syllabus.

[JM] Jurafsky & Martin, *Speech and Language Processing*, Prentice Hall, 2000.

[HLT] Cole et al., *Survey of the State of the Art in Human Language Technology*, 1996.  
<http://cslu.cse.ogi.edu/HLTsurvey/ch1node2.html#Chapter1>

[HTK] Young et al., *The HTK Book*, version 3.2.1, December 2002.

[SONIC] Pellom & Hacioglu, "Sonic: The University of Colorado Continuous Speech Recognizer," Center for Spoken Language Research Technical Report TR-CSLR-2001-01, U. Colorado, 2003 (revised).

[AT&T] Mohri & Riley, "Weighted Finite-State Transducers in Speech Recognition (Tutorial)," International



**Tentative schedule**

<b>Week</b>	<b>Topics/Readings</b>
Week 1 (Jan 6,8)	Introduction / Overview JM 7.1-7.2 (235-244) HLT 1 ( <a href="http://cslu.cse.ogi.edu/HLTsurvey/ch1node2.html#Chapter1">http://cslu.cse.ogi.edu/HLTsurvey/ch1node2.html#Chapter1</a> ) [YngLVR] Young LVR tutorial
Week 2 (Jan 13,15)	Toolkits The HTK Book SONIC Users Manual AT&T FSM Tutorial GMTK Manual
Week 3 (Jan 20,22)	Feature Extraction JM 7.5 (259-267) RJ 3 GM 22 HH 10
Week 4-5 (Jan 27,29, Feb 3,5)	Acoustic Modeling/HMM/Search JM 4.1-4.2 (91-103) JM 7.6-7.7 (267-274) [Rab] Rabiner HMM Tutorial [Yng] Young et al "Token Passing" [YngLVR] Young tutorial revisited
Week 6 (Feb 10,12)	Pronunciation Modeling JM 4.3-4.4 (10-130) [Fos] Fosler-Lussier Pronunciation Modeling Tutorial
Week 7 (Feb 17,19)	Language Modeling JM 6 [Brn] Brown et al Class-based n-grams [Jur] Jurafsky et al. Stochastic CFGs
<b>No class Feb 24: possible makeup?</b>	
Week 8-9 (Feb 24, Mar 2)	Advanced Techniques [Myr] Myrvoll review on adaptation HH 11 [CG] Chen and Gopalakrishnan, BIC paper [Ros] Rosenfeld, R. "Language modeling: Where do we go from here?"
Week 9 (Mar 4)	VoiceXML/Early Spoken Dialogue Processing [VOICEXML] VoiceXML tutorial Possibly SLP parsing & info extraction papers.
Week 10 (Mar 9, 11)	Project reports & wrap-up