

Syllabus

TeSTIA: The 8TH ELSNET European Summer School

Instructor: Eric Fosler-Lussier

This course covers techniques used in developing ASR pronunciation models using data from large speech corpora. The course has both a lecture and **practical** component, where we'll learn how to use the basic tools for pronunciation model. Each day is broken into 2 "parts" of about 40-45 minutes apiece; we'll take a short (5 min) break between parts. A rough breakdown of the schedule follows:

| Day | Part | Topic |
|------------------|--------|--|
| Monday | Part 1 | Course preliminaries; Introduction: What is ASR pronunciation modeling; short history and outline of techniques |
| | Part 2 | Finite state automata and transducers; introduction to AT&T finite state tools |
| Tuesday | Part 1 | Introduction to physiology and phonesets; phonological rules (regular and probabilistic) |
| | Part 2 | Modeling phonological rules with finite state transducers |
| Wednesday | Part 1 | Automatically derived models I: building pronunciation corpora, decision tree learning |
| | Part 2 | Training decision trees |
| Thursday | Part 1 | Automatically derived models II: smoothed phone recognition, other machine learning algorithms, dynamic pronunciation models |
| | Part 2 | Dynamic pronunciation models, model evaluation |
| Friday | Part 1 | Current directions: Models of phonological features; relationships to acoustic models; multiword models |
| | Part 2 | Accent models; summary |

The schedule and topics may deviate from the above due to availability of the lab, interest of the students, whim of the instructor, etc. If there is interest in holding more practical sessions, the (advanced) material in the latter part of the course may be compressed or dropped to accommodate the group desire.

The practical sessions will feature two publicly available toolkits:

- The AT&T Finite State Machine toolkit
available at: <http://www.research.att.com/sw/tools/fsm>
- The WEKA machine learning toolkit (slightly modified for our purposes)
available at: <http://www.cs.waikato.ac.nz/~ml>

This is a short reading list; I can also suggest many other readings for those who are interested:

- F. Chen, "Identification of contextual factors for pronunciation networks," ICASSP 1990.
- M. Riley, "A statistical model for generating pronunciation networks," ICASSP 1991.
- F. Pereira and M. Riley, "Speech recognition by composition of weighted finite automata," cmp-lg archive, 1996.
- G. Tajchman, E. Fosler, and D. Jurafsky, "Building multiple pronunciation models for novel words using exploratory computational phonology," ICASSP 1995.

- M. Finke and A. Waibel, “Speaking mode dependent pronunciation modeling in large vocabulary conversational speech recognition”, Eurospeech 1997.
- D. McAllaster, L. Gillick, F. Scattoni, and M. Newman, “Fabricating Conversational speech data with acoustic models: a program to examine model-data mismatch,” ICSLP 1998.
- M. Ostendorf, “Moving beyond the ‘Beads-on-a-string’ model of speech,” Automatic Speech Recognition and Understanding Workshop, 1999.
- E. Fosler-Lussier, S. Greenberg, and N. Morgan, “Incorporating contextual phonetics into automatic speech recognition,” International Congress of Phonetic Sciences, 1999.
- E. Fosler-Lussier, “Multi-level decision trees for static and dynamic pronunciation models,” Eurospeech 1999.