



# University of Washington, Seattle

## Department of Electrical Engineering

### Signal, Speech, and Language Interpretation (SSLI) Laboratory

<http://ssli.ee.washington.edu/ssli>

The University of Washington, Seattle recently made a major commitment to building a speech technology research program. We are currently a group of 3 senior researchers and 7 students, plus affiliated faculty, researchers and students. We expect significant growth in the near future. Mari Ostendorf, Jeff Bilmes, and Katrin Kirchhoff joined the department of Electrical Engineering in the fall of 1999. Building on the program Ostendorf created at Boston University, we have started SSLI-LAB, a new research laboratory aimed at solving the challenging problems in speech and language technology and general signal interpretation.

## Senior Staff

### Principal Faculty and Researchers

- Mari Ostendorf
- Katrin Kirchhoff
- Jeff Bilmes

### Affiliated Faculty and Researchers

- Les Atlas (Electrical Engineering)
- Alejandro Murua (Statistics)
- Eve Riskin (Electrical Engineering)

The combined expertise and experience of the principal and affiliated researchers include: statistical signal processing, pattern classification, phonetics, linguistics, machine learning, information theory and coding, acoustics, high-performance computing, and data compression. In addition, the University of Washington has strong faculty in the speech and hearing sciences and in the linguistics departments with whom connections have been initiated.

## Research Areas

- domain-dependent speech synthesis - combining prosody prediction (phrasing and emphasis) with unit selection, gradient models of prominence, adaptive intonation models.
- acoustic and pronunciation modeling for LVCSR - use of syllable and higher level linguistic structure, alternatives to HMMs.
- language modeling for LVCSR - dynamic dialog-dependent mixture models, training from sparse data.

- low-power and efficient software algorithms - optimize at the software level to reduce power consumption for speech systems running on portable devices.
- statistical modeling for information extraction from speech - identification of named entities in errorful data using phrase language models.
- prosody and dialog tracking - integrating prosodic and language cues for recognizing dialog acts and topic structure
- graphical models for ASR - adjust the models automatically to better match the properties found in very large databases.

## Programmatic Goals

Our overall goal is to maintain a stimulating research and teaching environment where participating members have expertise in a broad range of technical areas. This will facilitate the multidisciplinary research necessary to successfully attack the most challenging problems in speech processing. In particular, we intend:

1. to explore fundamentally new methods in speech and language technology, and
2. to produce an educational environment where students quickly become proficient in a wide range of speech technologies.

We are actively looking for collaborations with other sites, including both other West Coast sites and also international partners because we believe this will increase student exposure to different aspects of speech technology as well as research cultures.

## Current, Past and Future Collaborations with West Coast Sites

Ostendorf - SRI: prosody and speech recognition (hidden mode/event work), speech generation

Bilmes - ICSI: (PhD and other work)

Kirchhoff - ICSI: acoustic modeling (visiting research scientist)

UW-ICSI-SRI: robust recognition, dialog tracking on "meeting" speech

UW - OGI: speech synthesis (proposal planned)

## Future Research Areas

- text normalization (translating written to spoken word sequences) for speech synthesis and for language modeling training
- selective use of out-of-domain data - for improving performance in applications with sparse training data
- multi-lingual speech recognition and language identification
- spoken document information retrieval and extraction
- speech-based human-computer interfaces
- computational modeling of emotional states - in speech for style variation in synthesis and recognition

---

## Goals of Student Training

### Graduate Students

At the graduate level, we intend to improve students' oral and written communication skills, as well as their software skills (organizing, debugging, and optimizing code). Students are expected to receive a good grounding in a wide variety of statistical modeling techniques. This should include both the theoretical motivation and the practical "tricks" that are required to make the techniques work. Furthermore, they should acquire the ability to distinguish between general and speech-specific modeling techniques. We will establish cross-department courses to educate students from other departments on speech technology and vice versa – targeted departments at the University of Washington include computer science, statistics, linguistics, biological sciences, and speech and hearing sciences.

### Undergraduate Students

We intend to develop an undergraduate-level course for students to explore speech system design without requiring them to understand the sophisticated algorithmic details. In collaboration with the Computer Science department we hope to work on using speech technology to automatically annotate (with text) recorded classroom presentations; this will enable a text-based indexing of classroom content and will greatly facilitate information retrieval. A key goal is to more tightly integrate research and educational activities.

Our general educational goal for both graduate and undergraduate training is that students should be better prepared for the transition from university to industry.

## Current Students/PhD Topics

PhD candidates:

- Rebecca Bates - Hidden mode pronunciation modeling in spontaneous speech recognition
- Ivan Bulyko - Joint prosody prediction and unit selection for speech synthesis
- Ozgur Cetin - Acoustic modeling for speech recognition
- Randall Fish - Dynamic models for prediction of machine tool wear
- David Palmer - Language processing for information extraction
- Izhak Shafran - Use of syllable structure in acoustic modeling for spontaneous speech recognition

MS candidate:

- Peter Gilchrist - Intonation modeling for dialog systems

Additional former students are listed in Mari Ostendorf's CV, and this includes 21 graduated M.S. and Ph.D. theses. Of these, one is faculty in academia, and the remainder have acquired jobs in the speech technology industry.

## Current Funded Research Projects

- Speech generation for human-computer interaction - NSF/DARPA
- Modeling structure in speech above the segment for spontaneous speech recognition - NSF
- Center for auditory and acoustics research - ONR (subcontract from UMd)
- Robust recognition and dialog tracking for interactive information access - DARPA (starting 1/00)