David Huggins-Daines

Gates Hillman Complex 6225
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213
Tel.: +1 412 450 0343
Email: dhuggins@cs.cmu.edu
http://www.cs.cmu.edu/~dhuggins/

Research Interests

Scalable speech recognition for mobile devices, statistical modeling of speech for synthesis and recognition, quantitative methods for linguistic analysis.

Education

- Carnegie Mellon University, Pittsburgh, PA
  Ph.D. Candidate in Language Technologies Institute, School of Computer Science, *expected graduation: May 2010*
  Thesis: "Scalable, Anytime Speech Recognition for Mobile and Multicore Applications"
  Advisor: Alexander I. Rudnicky
  Thesis Committee: Alexander I. Rudnicky, Bhiksha Raj, Noah A. Smith, Thomas Schaaf (MultiModal, Inc.)
  M.S. in Language Technologies, December 2006
- University of Ottawa, Ottawa, Ontario, Canada
  B.A. Honours in Linguistics
  Advisor: John T. Jensen

Publications

- "Combining Mixture Weight Pruning and Quantization for Small-Footprint Speech Recognition" David Huggins-Daines and Alexander I. Rudnicky. *Proceedings of ICASSP-2009*, Taipei, Taiwan, April 2009.
- "Mixture Pruning and Roughening for Scalable Acoustic Models." David Huggins-Daines and Alexander I. Rudnicky. *Proceedings of ACL-08 Workshop on Mobile Language Processing*, Columbus, OH, USA, June 2008.
- "Interactive ASR Error Correction for Touchscreen Devices." David Huggins-Daines and Alexander I. Rudnicky. *Demo presented at ACL 2008*, Columbus, OH, USA, June 2008.
- "Implicitly Supervised Language Model Adaptation for Meeting Transcription." David Huggins-Daines and Alexander I. Rudnicky. *Proceedings of HLT-NAACL 2007*, Rochester, NY, USA, May 2007.
- "Conquest - an Open-Source Dialog System for Conferences." Bohus, D., Grau, S., Huggins-Daines, D., Keri, V., Krishna, G., Kumar, R., Raux, A., and Tomko, S. *Proceedings of HLT-NAACL 2007*, Rochester, NY, USA, May 2007.
- "A Constrained Baum-Welch Algorithm for Improved Phoneme Segmentation and Efficient Training." David Huggins-Daines and Alexander I. Rudnicky. To appear in *Proceedings of Interspeech 2006*, Pittsburgh, USA, September 2006.

- "PocketSphinx: A Free, Real-Time Continuous Speech Recognition System for Hand-Held Devices." David Huggins-Daines, Mohit Kumar, Arthur Chan, Alan W Black, Mosur Ravishankar, and Alexander I. Rudnicky. In *Proceedings of ICASSP 2006*, Toulouse, France, May 2006.
- "Investigations on Ensemble Based Semi-Supervised Acoustic Model Training" Rong Zhang, Ziad Al Bawab, Arthur Chan, Ananlada Chotimongkol, David Huggins-Daines, and Alexander I. Rudnicky. In *Proc. of Eurospeech 2005.*

Research Experience

- **Ph.D Thesis Research**, 2008-present
  My current research focuses on building mobile speech recognition systems that are: *Universal*, in that they can work in multiple domains from the same set of models, *Scalable*, in that they are able to adapt their vocabulary and model size at runtime, and finally, *Adaptive*, in that they are capable of learning new words and adapting to the user's voice on the fly without explicit supervision. A consistent trend in the development of mobile devices is that storage capacity grows much faster than storage speed, CPU power, and runtime memory size and bandwidth. My research attempts to address this problem by building ASR systems which use only a reconfigurable subset of their full knowledge base at any given time, and which incrementally refine previous results with an expanded knowledge base when sufficient resources (e.g. battery power, CPU time) are available.
- **Multi-modal Interfaces for Speech Visualization and Data Entry**, 2007-present
  I have investigated the use of "tag cloud" type visualizations to represent word lattices, along with the use of touch and multi-touch gestures to navigate these tag clouds and correct speech recognition errors. A prototype of the interface was demonstrated at ACL 2008. Currently, I am working on integrating this interface with my thesis work, such that the user feedback can be leveraged to improve acoustic and language model adaptation performance.
- **ASR-based Phonetic and Sub-phonetic Speech Coding**, 2006-2008
  Past work has shown that, even though phoneme recognition produces unreliable and inaccurate transcriptions, the acoustic model itself can fairly accurately represent the input speech. I investigated the degree to which this is also true of word-based recognition, by measuring the mean mel-cepstral distortion (MCD) between the original speech and speech regenerated from the state sequence corresponding to phoneme and word-based recognition results with various sized vocabularies. In addition, I constructed several end-to-end speech coding systems using the Sphinx trainer and decoder and the HMM synthesis components from the Festival speech synthesis system.
- **Spoken Meeting Recognition**, 2004-2007
  As part of the CALO project, I worked on improving speech recognition of spoken meeting data using the CMU Sphinx-III system. This included data selection for acoustic and language model training, acoustic and language model adaptation, and integration of the decoder with other components of the CALO system. In addition, I did a considerable amount of work on acoustic bandwidth expansion techniques in an attempt to leverage conversational telephone speech training data to improve the microphone-based meeting recognition system.
- **Grapheme-based Speech Recognition and Grapheme-to-Phoneme Conversion**, 2004-2006
  Grapheme-based acoustic modeling, while surprisingly effective in many cases, cannot approach the performance of phoneme-based modeling for languages such as English with complex orthographies. Likewise, grapheme-to-phoneme conversion is unreliable for such languages. In the former case, I worked on improving recognition performance by selecting appropriate multi-character

graphemes and applying cross-phonetic state tying, which I implemented in the Sphinx trainer for this purpose. In the latter case, I experimented with Transformation-Based Learning (e.g. the Brill parser) as a way of compensating for overtraining and data fragmentation in CART-based grapheme-to-phoneme conversion.

Teaching Experience

- **Teaching Assistant**, 11-761 Language and Statistics, Spring 2006
    - Developed and graded assignments, readings, and final exam questions
    - Held regular office hours
    - Lectured on Probabilistic Latent Semantic Analysis for Language Modeling and Information Retrieval
- **Teaching Assistant**, 15-492 Speech Processing, Fall 2008
    - Developed and graded assignments readings, and final exam questions
    - Held regular office hours
    - Lectured on Mobile Speech Processing

Software Development Experience

- **Language Technologies Institute**, Carnegie Mellon University, 2004-present
    - Initiated the PocketSphinx project to port and optimize the Sphinx speech recognition system on handheld and embedded devices. PocketSphinx is widely used in academia and industry as a lightweight, open-source speech recognizer.
    - Currently the main developer on PocketSphinx, with 5 major releases since 2005.
    - Development and maintenance of the other components of the Sphinx system, including:
        - Support for LDA and MLLT transforms in acoustic model training
        - Support for VTLN based speaker-adaptive training
        - Support for MAP adaptation in acoustic model training
        - Chinese language and Unicode support for language model training
        - Python language extensions for manipulating Sphinx data files and performing speech recognition
        - Integration of PocketSphinx with GStreamer for use on the GNU/Linux desktop
- **CREATE Lab**, Robotics Institute, Carnegie Mellon University, 2009-present
    - Built custom Linux firmware images for the ARM11-based TeRK and VexPro robot controllers using OpenEmbedded
    - Ported and debugged audio drivers for the TeRK and VexPro to Linux 2.6.29
    - Built custom GNU toolchains for use in the Eclipse development environment
- **WYEP 91.3**, Pittsburgh, PA, 2007
    - Assisted in development of a dynamic website for a community radio station
    - Interfaced the PHP/MySQL-based playlist system to the RDS transmitter to display live song and station info
- **Cepstral LLC**, Pittsburgh, PA, 2000-2004
    - Primary developer on Swift, a multi-platform, small-footprint text-to-speech engine
    - Implemented Microsoft SAPI and Apple SpeechManager interoperability for Swift

    - Ported Swift to Microsoft Windows CE and optimized its memory and CPU usage for mobile platforms, including commercially available assistive devices
    - Implemented a variety of Perl language extensions for text-to-speech, audio processing, and automatic speech recognition
- **Linuxcare, Inc.**, Ottawa, Ontario, Canada and San Francisco, CA, 1999-2000
    - Built a customized run-from-CD GNU/Linux distribution for Compaq Alpha server and workstation hardware
    - Maintained and enhanced the aboot bootloader for Alpha hardware
    - Implemented key parts of the GNU C Library and dynamic linker for GNU/Linux on HP PA-RISC server and workstation hardware
    - Implemented networking and RPC code in Perl for a Linux-based network-attached storage device
- **P&L Communications**, Ottawa, Ontario, Canada, 1998-2000
    - Implemented a search engine for newswire data using Perl, mod_perl, and a heavily modified version of the Isearch full-text indexing software
    - Worked with government agencies and freelance authors to develop effective web-based tools for media monitoring
- **Debian Project**, 1997-2002
    - Maintained various software packages for a community-based GNU/Linux distribution
    - Responded to bug reports and worked with upstream software authors to address them
    - Assisted in porting the Debian distribution to m68k-based Macintosh computers, including kernel and driver development
- **CHUO 89.1**, Ottawa, Ontario, Canada, 1996-1997
    - Designed an implemented a template-driven content management system for a community radio station website using Perl

Programming Skills

Programming Languages: C, Python, Perl, PHP, C++, JavaScript, Java, Scheme, Bourne Shell, m68k and ARM assembly
Operating Systems: GNU/Linux, ucLinux, Mac OS X, Windows NT, SymbianOS, Windows CE

Language Skills

English (native), French (fluent), Spanish (competent), Mandarin Chinese (intermediate), Polish (basic), Russian (basic), Brazilian Portuguese (basic)

Honors and Memberships

- Student Member, Institute of Electrical and Electronics Engineers (IEEE)
- Student Member, Association for Computational Linguistics (ACL)
- Computer Science Graduate Fellowship, Carnegie Mellon University, 2004-present
- Merit Scholarship, University of Ottawa, 1996-1999

Activities

- Organizer of the weekly Sphinx Group lunch, featuring speakers on various topics related to speech recognition, synthesis, and processing.
- Volunteer for Interspeech 2006 - set up and maintained the paper submission and presentation system, assisted with wireless network support.

Graduate Courses

11-711 Algorithms for NLP (Alon Lavie), 11-751 Speech Recognition (Tanja Schultz, Alex Waibel), 11-761 Language and Statistics (Roni Rosenfeld), 11-731 Machine Translation (Alon Lavie, Teruko Mitamura, Stephan Vogel), 11-712 Self-Paced Lab in NLP (Alon Lavie, Robert Frederking), 11-791 Software Engineering for Information Systems (Eric Nyberg), 15-681 Machine Learning (Roni Rosenfeld), 11-732 Self-Paced Lab in Machine Translation (Alon Lavie, Stephan Vogel), 11-734 Advanced Machine Translation Seminar (Alon Lavie, Stephan Vogel), 11-754 Project Course in Dialog Systems (Alan Black, Robert Frederking), 11-762 Language and Statistics II (Noah Smith), 05-631 Software Architecture for User Interfaces (Scott Hudson)