Seminar Speech Recognition and Language Processing in SS 2003

Last modified: 12. Februar 2003

Lehrstuhl für Informatik VI, RWTH Aachen
Prof. Dr.-Ing. H. Ney

List of topics:

1. **A^* Search (Oliver Bender)**
2. **Multi-Pass Search (Stephan Kanthak)**
3. **Transducer-based Search (Stephan Kanthak)**
4. **Speaker Adaptation using MLLR (Michael Pitz)**
5. **Speaker Adaptation using Eigenvoices (Michael Pitz)**
6. **Vocal Tract Length Normalization (Florian Hilger)**
7. **Histogram Normalization (Florian Hilger)**
8. **Discriminative Training (Wolfgang Macherey)**
9. **Discriminative Model Combination (Wolfgang Macherey)**
10. **Confidence Measures in Speech Recognition (Ralf Schlüter)**
11. **Word Error Rate Minimization (Ralf Schlüter)**
12. **Natural Language Database Queries using Classification and Regression Trees (Klaus Macherey)**
13. **Error Handling in Spoken Dialogue Systems (Klaus Macherey)**
14. **Language-Independent Named Entity Recognition (Oliver Bender)**
15. **Language Modelling and Word Morphology using Maximum Entropy (Nicola Ueffing)**
16. **Parsing Algorithms for Grammar-based Language Models (Max Bisani)**
17. **Long-range Language Models: Distant Bigrams, Link Grammars and Word Triggers (Max Bisani)**
18. **Parsing in Language Modelling (Nicola Ueffing)**

List of topics with references:

1. **A* Search (Oliver Bender)**
   - F. Jelinek, *Statistical Methods for Speech Recognition*, MIT Press, Cambridge, MA, 1997, chapter 6.
   - F. Jelinek, "A Fast Sequential Decoding Algorithm using a Stack", *IBM Journal of Research Development*, Vol. 13, 675-685, November 1969.
   - D. B. Paul, "An Essential A^* Stack Decoder Algorithm for Continuous Speech Recognition with a Stochastic Language Model," Proc. *Int. Conf. on Acoustics, Speech, and Signal Processing*, pp. 25-28, San Francisco, CA, March 1992.
2. **Multi-Pass Search (Stephan Kanthak)**
   - R. Schwartz, L. Nguyen, J. Makhoul: "Multiple-Pass Search Strategies," in *Automatic Speech and Speaker Recognition*, C.-H. Lee, F. K. Soong, K. K. Paliwal (eds.), pp. 429-456, Kluwer Academic Publishers, Norwell, MA, 1996.
   - X. Huang, A. Acero, F. Alleva, M. Hwang, L. Jiang, M. Mahajan: "From SPHINX-II to Whisper – Making Speech Recognition Usable," in *Automatic Speech and Speaker Recognition*, C.-H. Lee, F. K. Soong, K. K. Paliwal (eds.), pp. 481-508, Kluwer Academic Publishers, Norwell, MA, 1996.
   - P.S. Gopalakrishnan, L.R. Bahl: "Fast Match Techniques," in *Automatic Speech and Speaker Recognition*, C.-H. Lee, F. K. Soong, K. K. Paliwal (eds.), pp. 413-428, Kluwer Academic Publishers, Norwell, MA, 1996.
   - L.R. Bahl, S.V. De Gennaro, P.S. Gopalakrishnan, R.L. Mercer: "A Fast Approximate Acoustic Match for Large Vocabulary Speech Recognition," *IEEE Trans. on Speech and Audio Processing*, Vol. 1, pp. 59-67, January 1993.
   - P.S. Gopalakrishnan, L.R. Bahl, R.L. Mercer: "A Tree Search Strategy for Large Vocabulary Speech Recognition," Proc. *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, Vol. 1, pp. 572-575, Detroit, MI, May 1995.
3. **Transducer-based Search (Stephan Kanthak)**
   - G. Antoniol, F. Brugnara, M. Cettolo, M. Federico, "Language Model Representations for Beam-Search Decoding", Proc. of the *Int. Conf. on Acoustic, Speech and Signal Processing (ICASSP'95)*, 588-591, Detroit, MI, May 1995.
   - F. Brugnara, M. Cettolo, "Improvements in Tree-Based Language Model Representation", Proc. *Europ. Conf. on Speech Communication and Technology*, 1797-1800, Madrid, September 1995.
   - K. Demuynck, J. Duchateau, D. V. Compernolle, P. Wambacq, "An Efficient Search Space Representation for Large Vocabulary Continuous Speech Recognition", *Speech Communications*, Vol. 30, 37-53, January 2000.
   - M. Mohri, F. Pereira, M. Riley, "Weighted Finite-State Transducers in Speech Recognition", Proc. of the *ISCA ITRW Automatic Speech Recognition: Challenges for the new Millenium (ASR'2000)*, 97-106, Paris, France, August 2000.
   - F. Pereira, M. Riley, "Speech Recognition by Composition of Weighted Finite Automata", 1997.
4. **Speaker Adaptation using MLLR (Michael Pitz)**
   - D. Giuliani, R. DeMori: *Spoken Dialogs with Computers*, Chapter 11, Academic Press, 1997.
   - P. Woodland: "Speaker Adaptation for Continuous Density HMMs: A Review", Proc. *ISCA Workshop on Adaptation Methods for Speech Recognition*, pp. 11–19, Sofia Antinopolis, France, 2001.
   - C.J. Leggetter, P.C. Woodland: "Maximum Likelihood Linear Regression for Speaker Adaptation of Continuous Density Hidden Markov Models", *Computer Speech and Language*, Vol. 9, pp. 171–185, 1995.
   - M. J. F. Gales: "Maximum Likelihood Linear Transformations for HMM-Based Speech Recognition", *Computer Speech and Language*, Vol. 12, pp. 75–98, 1998.

- T. Anastaskos, J. McDonough, R. Schwartz, J. Makhoul: “A Compact Model for Speaker Adaptive Training”, Proc. of the *Int. Conf. on Spoken Language Processing (ICSLP’96)*, pp. 1137–1140, Philadelphia, PA, October 1996.

5. **Speaker Adaptation using Eigenvoices (Michael Pitz)**

- R. Kuhn, J.C. Junqua, P. Nguyen, N. Niedzielski: “Rapid Speaker Adaptation in Eigenvoice Space,” *IEEE Transactions on Speech and Audio Processing*, Vol. 8, No. 6, pp. 695-707, 2000.
- H. Botterweck: “Very Fast Adaptation for Large Vocabulary Speech Recognition Using Eigenvoices,” Proc. *Int. Conf. on Spoken Language Processing*, Vol. IV, pp. 354-357, Bejing, October 2000.
- M.J.F. Gales: “Cluster Adaptive Training of Hidden Markov Models,” *IEEE Transactions on Signal and Audio Processing*, Vol. 8, No. 4, pp. 417-420, 2000.
- P. Woodland: “Speaker Adaptation for Continuous Density HMMs: A Review”, Proc. *ISCA Workshop on Adaptation Methods for Speech Recognition*, pp. 11–19, Sofia Antinopolis, France, 2001.

6. **Vocal Tract Length Normalization (Florian Hilger)**

- A. Acero, R. M. Stern: “Robust Speech Recognition by Normalization of the Acoustic Space,” Proc. *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, Vol. II, pp. 893–896, Toronto, Canada, May 1991.
- E. Eide, H. Gish: “A Parametric Approach to Vocal Tract Length Normalization,” Proc. *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, Vol. I, pp. 346–349, Atlanta, GA, May 1996.
- S. Wegmann, D. McAllaster, J. Orloff, B. Peskin: “Speaker Normalization on Conversational Telephone Speech,” Proc. *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, Vol. I, pp. 339–342, Atlanta, GA, May 1996.
- L. Lee, R. C. Rose, “A frequency warping approach to speaker normalization,” IEEE Trans. Speech and Audio Processing, vol. 6, no. 1, pp. 49-60, Jan. 1998.
- D. Pye, P. C. Woodland: “Experiments in Speaker Normalization and Adaptation for Large Vocabulary Speech Recognition,” Proc. *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, Vol. II, pp. 1047–1050, Munich, Germany, April 1997.

7. **Histogram Normalization (Florian Hilger)**

- S. Dharanipragada, M. Padmanabhan: “A Nonlinear Unsupervised Adaptation Technique for Speech Recognition,” *Proc. Int. Conf. on Spoken Language Processing*, pp. 556–559, Bejing, China, October 2000.
- M. Padmanabhan, S. Dharanipragada: “Maximum Likelihood Non-linear Transformation for Environtment Adaptation in Speech Recognition Systems,” Proc. *European Conf. on Speech Communication and Technology*, pp. 2359–2362, Aalborg, Denmark, September 2001.
- J. C. Segura, M. C. Benitez, A. de la Torre, A. J. Rubio: “Feature extraction combining spectral subtraction and cepstral histogram equalization,” Proc. *Int. Conf. on Spoken Language Processing*, vol. 1, pp. 225–228, Denver, CO, USA, September 2002.
- F. Hilger, S. Molau, H. Ney: “Quantile Based Histogram Equalization For Online Applications,” Proc. *Int. Conf. on Spoken Language Processing*, vol. 1, pp. 237–240, Denver, CO, USA, September 2002.
- S. Molau, F. Hilger, D. Keysers, H. Ney: “Enhanced Histogram Normalization in the Acoustic Feature Space,” Proc. *Int. Conf. on Spoken Language Processing*, Vol. 1, pp. 1421–1424, Denver, CO, USA, September 2002.

8. **Discriminative Training (Wolfgang Macherey)**

- Y. Normandin, “Maximum Mutual Information Estimation of Hidden Markov Models,” in *Automatic Speech and Speaker Recognition*, C.-H. Lee, F. K. Soong, K. K. Paliwal (eds.), pp. 57-81, Kluwer Academic Publishers, Norwell, MA, 1996.
- V. Valtchev, J. J. Odell, P. C. Woodland, S. J. Young: “MMIE Training of Large Vocabulary Recognition Systems,” Speech Communication, Vol. 22, No. 4, pp. 303-314, September 1997.

- P.C. Woodland, D. Povey: "Large Scale Discriminative Training of Hidden Markov Models for Speech Recognition," *Computer Speech and Language*, Vol. 16, No. 1, pp. 25-47, January 2002.

9. **Discriminative Model Combination (Wolfgang Macherey)**

   - P. Beyerlein: *Diskriminative Modellkombination in Spracherkennungssystemen mit großem Wortschatz*, Dissertation, RWTH Aachen, Germany, October 2000.

10. **Confidence Measures in Speech Recognition (Ralf Schlüter)**

    - F. Wessel, R. Schlüter, K. Macherey, H. Ney: "Confidence Measures for Large Vocabulary Continuous Speech Recognition," Proc. *IEEE Trans. on Speech and Audio Processing*, Vol. 9, No. 3, pp. 288-298, March 2001.
    - F. Wessel: "Word Posterior Probabilities for Large Vocabulary Continuous Speech Recognition," Dissertation, Aachen, Germany, January 2002.

11. **Word Error Rate Minimization (Ralf Schlüter)**

    - F. Wessel: "Word Posterior Probabilities for Large Vocabulary Continuous Speech Recognition," Dissertation, Aachen, Germany, January 2002.
    - A. Stolcke, Y. König, M. Weintraub: "Explicit Word Error Rate Minimization in N-best List Rescoring," Proc. *Europ. Conf. on Speech Communication and Technology*, Vol. 2, pp. 163-166, Rhodes, Greece, September 1997.
    - G. Evermann, P. C. Woodland: "Large Vocabulary Decoding and Confidence Estimation Using Word Posterior Probabilities," Proc. *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, Vol. 3, pp. 1655-1658, Istanbul, Turkey, June 2000.
    - L. Mangu, E. Brill, A. Stolcke: "Finding Consensus Among Words: Lattice-Based Word Error Minimization," Proc. *Europ. Conf. on Speech Communication and Technology*, Vol. 1, pp. 495-498, Budapest, Hungary, September 1999.
    - V. Goel, W. Byrne: "Minimum Bayes Risk Automatic Speech Recognition," *Computer Speech and Language*, Vol. 14, pp. 115-135, 2000.

12. **Natural Language Database Queries using Classification and Regression Trees (Klaus Macherey)**

    - R. Kuhn, R. De Mori: "The Application of Semantic Classification Trees to Natural Language Understanding," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 17, No. 7, pp. 449-460, 1995.
    - R. Kuhn, R. De Mori: "Sentence Interpretation," in *Spoken Dialogues with Computers*, R. De Mori (ed.), Chapter 14, pp. 485-522, Academic Press, San Diego, CA, 1998.
    - R.O. Duda, P.E. Hart, D.G. Stork: *Pattern Classification*, John Wiley and Sons, New York, NY, 2001, pp. 395-411.

13. **Error Handling in Spoken Dialogue Systems (Klaus Macherey)**

    - M. Swerts, D. Litman, J. Hirschberg: "Corrections in Spoken Dialogue Systems", Proc. *Int. Conf. Spoken Language Processing (ICSLP 2000)*, Vol. II, pp. 615-618, Beijing, October 2000.
    - K. Komatani, T. Kawahara: "Generating Effective Confirmation and Guidance Using Two-Level Confidence Measures for Dialogue Systems", Proc. *Int. Conf. Spoken Language Processing (ICSLP 2000)*, Vol. II, pp. 648-651, Beijing, October 2000.
    - E. Krahmer, M. Swerts, M. Theune, M. Weegels: "Error Detection in Spoken Human-Machine Interaction," *International Journal of Speech Technology*, Vol. 4, No. 1, pp. 19-30, 2001.
    - D. J. Litman, M. A. Walker, M. S. Kearns: "Automatic Detection of Poor Speech Recognition at the Dialogue Level", Proc. 37th Annual Meeting of the Association for Computational Linguistics (ACL 1999), pp. 309–316, 1999.

14. **Language-Independent Named Entity Recognition (Oliver Bender)**

   - E.F. Tjong Kim Sang: "Introduction to the CoNLL-2002 Shared Task: Language-Independent Named Entity Recognition." Proc. *Computational Natural Language Learning Workshop*, Taipei, Taiwan, 2002, pp. 155-158.
   - X. Carreras, L. Márques, Lluís Padró, "Named Entity Extraction using AdaBoost," Proc. *Computational Natural Language Learning Workshop*, Taipei, Taiwan, 2002, pp. 167-170.
   - R. Florian: "Named Entity Recognition as a House of Cards: Classifier Stacking." Proc. *Computational Natural Language Learning Workshop*, Taipei, Taiwan, 2002, pp. 175-178.

15. **Language Modelling and Word Morphology using Maximum Entropy (Nicola Ueffing)**

   - A.L. Berger, S. Della Pietra, V. Della Pietra: "A Maximum Entropy Approach to Natural Language Processing", *Computational Linguistics*, Vol. 22, No. 1, pp. 39-71, March 1996.
   - R. Rosenfeld: "A Maximum Entropy Approach to Adaptive Statistical Language Modelling", *Computer Speech and Language*, Vol. 10, No. 3, pp. 187-228, July 1996.
   - S.C. Martin, H. Ney, C. Hamacher. "Maximum Entropy Language Modeling and the Smoothing Problem", *IEEE Trans. on Speech and Language Processing*, Vol. 8, No. 5, pp. 626-632, September 2000.

16. **Parsing Algorithms for Grammar-based Language Models (Max Bisani)**

   - B. Roark: "Probabilistic top-down parsing and language modelling", *Computational Linguistics*, Vol. 27, No. 2, pp. 249-276, 2001.
   - B. Roark: "Markov parsing: lattice rescoring with a statistical parser", Proc. *40th Annual Meeting of the Association for Computational Linguistics (ACL 2002)*, pp. 287–294, Philadelphia, PA, USA, 2002.
   - E. Charniak: "Immediate-Head Parsing for Language Models" Proc. 39th Annual Meeting of the Association for Computational Linguistics (ACL 2001), pp. 116–123, Toulouse, France, July 2001.
   - D.H. Van Uytsel, F. Van Aelten, D. Van Compernolle: "A Structured Language Model based on Context-Sensitive Probabilistic Left-Corner Parsing" Proc. *Meeting of the North-American Chapter of the Association of Computaional Linguists*, pp. 223–230, Pittsburgh, PA, USA, June 2001.

17. **Long-range Language Models: Distant Bigrams, Link Grammars and Word Triggers (Max Bisani)**

   - M. Simons, H. Ney, S. C. Martin: "Distant Bigram Language Modelling Using Maximum Entropy," Proc. *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, Vol. 2, pp. 787-790, Munich, Germany, April 1997.
   - S. Della Pietra, V. Della Pietra, J. Gillet, J. Lafferty, H. Prinz, L. Ures: "Inference and Estimation of a Long-Range Trigram Model," in *Grammatical Inference and Applications, Second International Colloquium*, Lecture notes in Artificial Intelligence, No. 862, Springer Verlag, Berling, 1994.
   - R. Rosenfeld: "A whole sentence maximum entropy language model," in Proc. *IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*, pp., Santa Barbara, CA, December 1997.

18. **Parsing in Language Modelling (Nicola Ueffing)**

   - E. Charniak: "Immediate-Head Parsing for Language Models", Proc. 39th Annual Meeting of the Association for Computational Linguistics (ACL 2001), pp. 116-123, Toulouse, France, July 2001.
   - C. Chelba, F.Jelinek: "Exploiting Syntactic Structure for Language Modeling", Proc. 36th Annual Meeting of the Association for Computational Linguistics (ACL 2001), pp. 225-231, Montréal, Canada, August 1998.
   - B. Roark: "Probabilistic top-down parsing and language modelling", *Computational Linguistics*, Vol. 27, No. 2, pp. 249-276, 2001.
   - B. Roark, *Robust Probabilistic Predicitve Syntactic Processing: Motivations, Models, and Applications*, PhD thesis, Brown University, Providence, RI, May 2001.