MARKOVIAN COMBINATION OF LANGUAGE AND PROSODIC MODELS FOR BETTER SPEECH UNDERSTANDING AND RECOGNITION

Andreas Stolcke and Elizabeth Shriberg

Speech Technology and Research Laboratory
SRI International,Menlo Park, CA, USA
{stolcke,ees}@speech.sri.com

ABSTRACT

Traditionally, "language" models capture only the word sequences of a language. A crucial component of spoken language, however, are its rhythmic and melodic properties, i.e., its prosody. This talk will summarize recent work on integrated, computationally efficient modeling of word sequences and prosodic properties of speech, for a variety of speech recognition and understanding tasks, such as dialog act tagging, disfluency detection, and segmentation into sentences and topics. In each case it turns out that hidden Markov representations of the underlying structures and associated observations arise naturally, and allow existing speech recognizers to be combined with separately trained prosodic classifiers. The same HMM-based models can be used in two modes: to recover hidden structure (such as sentence boundaries), or to evaluate speech recognition hypotheses, thereby integrating prosody into the recognition process.

More information about individual research projects, as well as the publications listed, are available at http://www.speech.sri.com/projects/hidden-events.html, http://www.speech.sri.com/projects/sieve/, and http://www.clsp.jhu.edu/ws9 7/discourse/.

1. REFERENCES

[1] A. Stolcke, E. Shriberg, R. Bates, M. Ostendorf, D. Hakkani, M. Plauché, G. Tür, and Y. Lu, "Automatic detection of sentence boundaries and disfluencies based on recognized words," in *Proceedings of the International Conference on Spoken Language Processing* (R. H. Mannell and J. Robert-Ribes, eds.), vol. 5, (Sydney), pp. 2247–2250, Australian Speech Science and Technology Association, Dec. 1998.

[2] E. Shriberg, R. Bates, A. Stolcke, P. Taylor, D. Jurafsky, K. Ries, N. Coccaro, R. Martin, M. Meteer, and C. Van Ess-Dykema, "Can prosody aid the automatic classification of dialog acts in conversational speech?," *Language and Speech*, vol. 41, no. 3-4, pp. 439–487, 1998.

[3] K. Sönmez, E. Shriberg, L. Heck, and M. Weintraub, "Modeling dynamic prosodic variation for speaker verification," in *Proceedings of the International Conference on Spoken Language Processing* (R. H. Mannell and J. Robert-Ribes, eds.), vol. 7, (Sydney), pp. 3189–3192, Australian Speech Science and Technology Association, Dec. 1998.

[4] A. Stolcke, E. Shriberg, D. Hakkani-Tür, and G. Tür, "Modeling the prosody of hidden events for improved word recognition," in *Proceedings of the 6th European Conference on Speech Communication and Technology*, vol. 1, (Budapest), pp. 307–310, Sept. 1999.

[5] E. Shriberg, A. Stolcke, D. Hakkani-Tür, and G. Tür, "Prosody-based automatic segmentation of speech into sentences and topics," *Speech Communication*, vol. 32, pp. 127–154, Sept. 2000. Special Issue on Accessing Information in Spoken Audio.

[6] A. Stolcke, K. Ries, N. Coccaro, E. Shriberg, D. Jurafsky, P. Taylor, R. Martin, C. Van Ess-Dykema, and M. Meteer, "Dialogue act modeling for automatic tagging and recognition of conversational speech," *Computational Linguistics*, vol. 26, no. 3, pp. 339–373, 2000.

[7] G. Tür, D. Hakkani-Tür, A. Stolcke, and E. Shriberg, "Integrating prosodic and lexical cues for automatic topic segmentation," *Computational Linguistics*, vol. 27, pp. 31–57, Mar. 2001.

[8] E. Shriberg, A. Stolcke, and D. Baron, "Can prosody aid the automatic processing of multi-party meetings? Evidence from predicting punctuation, disfluencies, and overlapping speech," in *Proceedings ISCA Tutorial and Research Workshop on Prosody in Speech Recognition and Understanding* (M. Bacchiani, J. Hirschberg, D. Litman, and M. Ostendorf, eds.), (Red Bank, NJ), pp. 139–146, Oct. 2001.

The research described here was supported by NSF, DARPA, and NASA. The views expressed are those of the authors and should not be interpreted as representing the policies of the funding agencies.