Bibliography for:

"Intonational Variation in Spoken Dialogue Systems: Generation and Understanding"

Julia Hirschberg
AT&T Labs – Research
julia@research.att.com
http://www.research.att.com/ julia/index.html

KTH

September 2000

The variation of prosodic features such as contour, accent, phrasing, pitch range, loudness, speaking rate, and timing contributes important information to the interpretation of spoken dialogue. To create dialogue systems that interact naturally with human users, it is important both to generate natural speech and to understand what users are saying, and intonational contributions are important to both. This course of lectures will discuss past findings and current research in the role of intonational variation in speech synthesis/concept to speech and speech recognition/understanding for spoken dialogue systems.

References

[Avesani and Vayra1988] Avesani, Cinzia and Mario Vayra. 1988. Discorso, segmenti di discorso e un' ipotesi sull' intonazione. In *Corso di stampa negli Atti del Convegno Internazionale "Sull'Interpunzione"*, pages 8–53, Vallecchi, Firenze.

[Ayers1992] Ayers, Gayle M. 1992. Discourse functions of pitch range in spontaneous and read speech. Presented at the Linguistic Society of America Annual Meeting.

[Bangalore and Rambow2000] Bangalore, Srinivas and Owen Rambow. 2000. Exploiting a probabilistic hierarchical model for generation. In *Proceedings of COLING-2000*, Saarbruücken.

[Bangalore, Rambow, and Whittaker2000] Bangalore, Srinivas, Owen Rambow, and Steve Whittaker. 2000. Evaluation metrics for generation. In *Proceedings of the International Natural Language Generation*, Mitzpe Ramon, Israel.

[Bear, Dowding, and Shriberg1992] Bear, John, John Dowding, and Elizabeth Shriberg. 1992. Integrating multiple knowledge sources for detection and correction of repairs in human-computer dialog. In *Proceedings of the 30th Annual Meeting*, pages 56–63, Newark DE. Association for Computational Linguistics.

[Beeferman, Berger, and Lafferty1999] Beeferman, Doug, Adam Berger, and John Lafferty. 1999. Statistical models for text segmentation. *Machine Learning:Special Issue on Natural Language Learning*, 34(1-3):177–210.

[Bell and Gustafson1999] Bell, Linda and Joakim Gustafson. 1999. Repetition and its phonetic realizations: Investigating a Swedish database of spontaneous computer-directed speech. In *Proceedings of ICPhS-99*, San Francisco. International Congress of Phonetic Sciences.

[Blackmer and Mitton1991] Blackmer, Elizabeth R. and Janet L. Mitton. 1991. Theories of monitoring and the timing of repairs in spontaneous speech. *Cognition*, 39:173–194.

[Bolinger1989] Bolinger, Dwight. 1989. *Intonation and Its Uses: Melody in Grammar and Discourse.* Edward Arnold, London.

[Boomer1965] Boomer, D. S. 1965. Hesitation and grammatical encoding. *Language and Speech*, 8:148–158.

[Bortfeld and Brennan1997] Bortfeld, H. and S. E. Brennan. 1997. Use and acquisition of idiomatic expressions in referring by native and non-native speakers. *Discourse Processes*, 23:119–147.

[Bortfeld et al.1991] Bortfeld, Heather, Silvia D. Leon, Jonathan E. Bloom, Michael F. Schober, and Susan E. Brennan. 1991. Which speakers are most disfluent in conversation and when? In *Proceedings of the ICPS-99 Satellite Meeting: Disfluency in Spontaneous Speech*, pages 7–10, U. C. Berkeley.

[Bouwman, Sturm, and Boves1999] Bouwman, A. G., J. Sturm, and L. Boves. 1999. Incorporating confidence measures in the dutch train timetable information system developed in the ARISE project. In *Proc. International Conference on Acoustics, Speech and Signal Processing*, volume 1, pages 493–496, Phoenix.

[Brennan and Williams1995] Brennan, S. E. and M. Williams. 1995. The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language*, 34:383–398.

[Brown, Currie, and Kenworthy1980] Brown, G., K. Currie, and J. Kenworthy. 1980. *Questions of Intonation.* University Park Press, Baltimore.

[Brown and Yule1983] Brown, G. and G. Yule. 1983. *Discourse Analysis.* Cambridge University Press, Cambridge UK.

[Butterworth1975] Butterworth, B. 1975. Hesitation and semantic planning in speech. *Journal of Psycholinguistic Research*, 4:75–87.

[Cahn1988] Cahn, J. 1988. From sad to glad: Emotional computer voices. In *Proceedings of Speech Tech '88*, pages 35–36.

[Chafe1980] Chafe, W. L. 1980. The deployment of consciousness in the production of a narrative. In W. L. Chafe, editor, *The Pear Stories: Cognitive, Cultural and Linguistic Aspects of Navvative Production.* Ablex Publishing Corp, pages 9–50. Vol. 3, Advances in Discourse Processes.

[Clark and Brennan1991] Clark, H. H. and S. E. Brennan. 1991. Grounding in communication. In L. B. Resnick, J. Levine, and S. D. Teasley, editors, *Perspectives on Socially Shared Cognition.* Morgan Kaufman, San Mateo.

[Clark and Schaefer1989] Clark, H. H. and E. F. Schaefer. 1989. Contributing to discourse. *Cognitive Science*, 13:259–294.

[Clark and Wilkes-Gibbs1986] Clark, H. H. and D. Wilkes-Gibbs. 1986. Referring as a collaborative process. *Cognition*, 22:1–39.

[Cohen1996] Cohen, William. 1996. Learning trees and rules with set-valued features. In *14th Conference of the American Association of Artificial Intelligence, AAAI.*

[Collins1999] Collins, M. 1999. *Head-Driven Statistical Model for Natural Language Parsing.* Ph.D. thesis, University of Pennsylvania.

[Goodwin1981] Goodwin, C. 1981. *Conversational Organization: Interaction between Speakers and Hearers.* The Academic Press, New York.

[Grosz and Hirschberg1992] Grosz, Barbara and Julia Hirschberg. 1992. Some intonational characteristics of discourse structure. In *Proceedings of the International Conference on Spoken Language Processing*, Banff, October. ICSLP.

[Gustafson and Bell2000] Gustafson, J. and L. Bell. 2000. Speech technology on trial: Experiences from the august system. *Natural Language Engineering: Special Issue on Best Practice in Spoken Language Dialogue System Engineering*, 6.

[Gustafson et al.1997] Gustafson, J., A. Larsson, R. Carlson, and K. Hellman. 1997. How do system questions influence lexical choices in user answers? In *Proceedings of EUROSPEECH-97*, volume 4, pages 2275–2278.

[Hearst1994] Hearst, Marti. 1994. Multi-paragraph segmentation of expository discourse. In *Proceedings of the 32nd Annual Meeting*, pages 9–16, Las Cruces.

[Hindle1983] Hindle, Donald. 1983. Deterministic parsing of syntactic non-fluencies. In *Proceedings of the 21st Annual Meeting*, pages 123–128, Cambridge MA. Association for Computational Linguistics.

[Hirschberg and Nakatani1996] Hirschberg, Julia and Christine Nakatani. 1996. A prosodic analysis of discourse segments in direction-giving monologues. In *Proceedings of the 34th Annual Meeting*, Santa Cruz. Association for Computational Linguistics.

[Hirschberg, Litman, and Swerts2000] Hirschberg, Julia B., Diane J. Litman, and Marc Swerts. 2000. Generalizing prosodic prediction of speech recognition errors. In *Proceedings of the Sixth International Conference on Spoken Language Processing*, Beijing.

[Hockey1989] Hockey, B. A. 1989. The function of 'okay' as a cue phrase in discourse. Presented at the LSA Annual Meeting, December.

[Hockey1991] Hockey, Beth Ann. 1991. Prosody and the interpretation of 'okay'. In *Proceedings of the AAAI Fall Symposium on Discourse Structure in Natural Language Generation and Understanding*, Asilomar CA.

[Howell and Young1991] Howell, P. and K. Young. 1991. The use of prosody in highlighting alterations in repairs from unrestricted speech. *The Quarterly Journal of Experimental Psychology*, 43A(3):733–758.

[Kasl and Mahl1965] Kasl, S. V. and G. F. Mahl. 1965. The relationship of disturbances and hesitations in spontaneous speech to anxiety. *Journal of Personality and Social Psychology*, 1:425–433.

[Koehn et al.2000] Koehn, P., S. Abney, J. Hirschberg, and M. Collins. 2000. Improving intonational phrasing with syntactic information. In *Proceedings of ICASSP-00*, Istanbul.

[Kowtko1996] Kowtko, Jacqueline. 1996. *The Function of Intonation in Task-Oriented Dialogue.* Ph.D. thesis, University of Edinburgh.

[Krahmer et al.1999a] Krahmer, E., M. Swerts, M. Theune, and M. Weegels. 1999a. Error spotting in human-machine interactions. In *Proceedings of EUROSPEECH-99*.

[Krahmer et al.1999b] Krahmer, E., M. Swerts, M. Theune, and M. Weegels. 1999b. Prosodic correlates of disconfirmations. In *Proceedings of the ESCA Workshop on Dialogue and Prosody*, pages 169–174, Veldhoven.

[Krahmer et al.2000] Krahmer, E., M. Swerts, M. Theune, and M. Weegels. 2000. The dual of denial: Two uses of disconfirmations in dialogue and their prosodic correlates. *Speech Communication*.

[Ladd1980] Ladd, D. Robert. 1980. *The Structure of Intonational Meaning*. Indiana University Press, Bloomington, Ind.

[Langkilde and Knight1998] Langkilde, Irene and Kevin Knight. 1998. Generation that exploits corpus-based statistical knowledge. In *Proceedings of ACL-98*, pages 704–710, Montréal, Canada.

[Lehiste1975] Lehiste, I. 1975. The phonetic structure of paragraphs. In A. Cohen and S. G. Nooteboom, editors, *Structure and Process in Speech Perception*. Springer, Heidelberg, pages 195–203.

[Lehiste1979] Lehiste, I. 1979. Perception of sentence and paragraph boundaries. In B. Lindblom and S. Oehman, editors, *Frontiers of Speech Research*. Academic Press, London, pages 191–201.

[Lehiste1980] Lehiste, I. 1980. Phonetic characteristics of discourse. Paper presented at the Meeting of the Committee on Speech Research, Acoustical Society of Japan.

[Levelt and Cutler1983] Levelt, W. J. M. and A. Cutler. 1983. Prosodic marking in speech repair. *Journal of Semantics*, 2:205–217.

[Levow1998] Levow, Gina-Anne. 1998. Characterizing and recognizing spoken corrections in human-computer dialogue. In *Proceedings of the 36th Annual Meeting of the Association of Computational Linguistics, COLING/ACL 98*, pages 736–742.

[Litman, Hirschberg, and Swerts2000] Litman, Diane J., Julia B. Hirschberg, and Marc Swerts. 2000. Predicting automatic speech recognition performance using prosodic cues. In *Proceedings of the First Annual Meeting*, Seattle, May. North American Association for Computational Linguistics (NAACL-00).

[Litman and Pan1999] Litman, Diane J. and Shimei Pan. 1999. Empirically evaluating an adaptable spoken dialogue system. In *Proceedings of the 7th International Conference on User Modeling (UM)*.

[Murray and Arnott1993] Murray, Iain R. and Mohn L. Arnott. 1993. Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *Journal of the Acoustical Society of America*, 93(2):1097–1108.

[Nakatani and Hirschberg1994] Nakatani, C. and J. Hirschberg. 1994. A corpus-based study of repair cues in spontaneous speech. *Journal of the Acoustical Society of America*, March.

[Nakatani and Chu-Carroll2000] Nakatani, Christine and Jennifer Chu-Carroll. 2000. Using dialogue representations for concept-to-speech. In *Proceedings of the NAACL-ANLP 2000 Workshop on Conversational Systems*, pages 48–53, Seattle.

[Nakatani, Grosz, and Julia1995] Nakatani, Christine, Barbara Grosz, and Hirschberg Julia. 1995. Discourse structure in spoken language: Studies on speech corpora. In *Proceedings of the AAAI Spring Symposium on Empirical Methods in Discourse Interpretation and Generation*, Stanford, March. AAAI.

[Nöth et al.1999] Nöth, E., A. Batliner, V. Warnke, J. Haas, M. Boros, J. Buckow, R. Huber, F. Gallwitz, M. Nutt, and H. Niemann. 1999. On the use of prosody in automatic dialogue understanding. In *Proceedings of the ESCA Workshop on Dialogue and Prosody*, pages 25–34, Veldhoven.

[Oviatt1995] Oviatt, S. 1995. Predicting spoken disfluencies during human-computer interaction. *Computer Speech and Language*, 9:19–35.

[Oviatt et al.1996] Oviatt, S. L., G. Levow, M. MacEarchern, and K. Kuhn. 1996. Modeling hyperarticulate speech during human-computer error resolution. In *Proceedings of ICSLP-96*, pages 801–804, Philadelphia.

[Pan and Hirschberg2000] Pan, Shimei and Julia Hirschberg. 2000. Modeling local context for pitch accent prediction. In *Proceedings of ACL-2000*, Hong Kong.

[Pan and McKeown1998] Pan, Shimei and Kathleen R. McKeown. 1998. Learning intonation rules for concept to speech generation. In *Proceedings of ACL-98*, Montreal.

[Passoneau and Litman1993] Passoneau, R. and D. Litman. 1993. Feasibility of automated discourse segmentation. In *Proceedings of the 31st Annual Meeting*, Ohio State University. Association for Computational Linguistics.

[Pierrehumbert and Hirschberg1990] Pierrehumbert, Janet and Julia Hirschberg. 1990. The meaning of intonational contours in the interpretation of discourse. In P. Cohen, J. Morgan, and M. Pollack, editors, *Intentions in Communication*. MIT Press, Cambridge MA, pages 271–311.

[Pierrehumbert1980] Pierrehumbert, Janet B. 1980. *The Phonology and Phonetics of English Intonation*. Ph.D. thesis, Massachusetts Institute of Technology, September. Distributed by the Indiana University Linguistics Club.

[Pitrelli, Beckman, and Hirschberg1994] Pitrelli, John, Mary Beckman, and Julia Hirschberg. 1994. Evaluation of prosodic transcription labeling reliability in the ToBI framework. In *Proceedings of the Third International Conference on Spoken Language Processing*, volume 2, pages 123–126, Yokohama. ICSLP.

[Price, Ostendorf, and Wightman1989] Price, P. J., M. Ostendorf, and C. W. Wightman. 1989. Prosody and parsing. In *Speech and Natural Language Workshop*, Cape Cod, October. DARPA.

[Pulman1997] Pulman, S. 1997. Higher order unification and the interpretation of focus. *Linguistics and Philosophy*, 20:73–115.

[Reynar1998] Reynar, Jeff. 1998. *Topic Segmentation: Algorithms and Applications*. Ph.D. thesis, University of Pennsylvania.

[Rooth1992] Rooth, Mats. 1992. A theory of focus interpretation. *Natural Language Semantics*, 1:75–116.

[Sanderman and Collier1997] Sanderman, Angelien A. and Rene Collier. 1997. Prosodic phrasing and comprehension. *Language and Speech*, 40(4):391–408.

[Schapire and Singer2000] Schapire, Robert E. and Yoram Singer. 2000. Boostexter: A boosting-based system for text categorization. *Machine Learning*, 39(2-3):135–168.

[SchoberTo appear] Schober, S. E. Brennan M. F. To appear. How speech disfluencies inform comprehension. *Journal of Memory and Language*.

[Shimojima et al.1999] Shimojima, A., K. Katagiri, H. Koiso, and M. Swerts. 1999. An experimental study on the informational and grounding functions of prosodic features of Japanese echoic responses. In *Proceedings of the ESCA Workshop on Dialogue and Prosody*, pages 187–192, Veldhoven.

[Shimojima et al.2001] Shimojima, Atsushi, Yasuhiro Katagiri, Hanae Koiso, and Marc Swerts. 2001. An experimental study on the informational and grounding functions of prosodic features of Japanese echoic responses. *Speech Communication*. An extended version of a paper present at the ESCA Workshop on Dialogue and Prosody, Veldhoven, 1999.

[Shriberg1996] Shriberg, E. 1996. Disfluencies in switchboard. In *Addendum to the Proceedings of ICSLP-96*, pages 11–14, Philadelphia.

[Shriberg et al.2000] Shriberg, E., A. Stolcke, D. Hakkani-Tur, and G. Tur. 2000. Prosody-based automatic segmentation of speech into sentences and topics. *Speech Communication: Special Issue on Accessing Information in Spoken Audio*, 32(1-2).

[Shriberg1994] Shriberg, E. E. 1994. *Preliminaries to a Theory of Speech Disfluencies*. Ph.D. thesis, University of California at Berkeley.

[Shriberg et al.1998] Shriberg, Elizabeth, Rebecca Bates, Andreas Stolcke, Paul Taylor, Daniel Jurafsky, Klaus Ries, Noah Coccaro, Rachel Martin, Marie Meteer, and Carol Van Ess-Dykema. 1998. Can prosody aid the automatic classification of dialog acts in conversational speech? *Language and Speech: Special Issue on Prosody and Conversation*, 41(3-4):443–492.

[Silverman1986] Silverman, K. 1986. Synthesis and perception of paragraph prosody. In *Proceedings*. First Australian Conference of Speech Science and Technology.

[Silverman et al.1992] Silverman, K., M. Beckman, J. Pierrehumbert, M. Ostendorf, C. Wightman, P. Price, and J. Hirschberg. 1992. ToBI: A standard scheme for labeling prosody. In *Proceedings of the Second International Conference on Spoken Language Processing*, pages 867–879, Banff, October. ICSLP.

[Silverman et al.1993] Silverman, Kim, Ashok Kalyanswamy, Julie Silverman, Sara Basson, and Dina Yashchin. 1993. Synthesiser intelligibility in the context of a name-and-address information service. In *Proceedings of EUROSPEECH-93*, volume 3, pages 2169–2172, Berlin.

[Soltau and Waibel1998] Soltau, Hagen and Alex Waibel. 1998. On the influence of hyperarticulated speech on recognition performance. In *Proceedings of ICSLP-98*, Sydney. International Conference on Spoken Language Processing.

[Soltau and Waibel2000] Soltau, Hagen and Alex Waibel. 2000. Specialized acoustic models for hyperarticulated speech. In *Proceedings of ICASSP 2000*, Istanbul.

[Sproat, Hirschberg, and Yarowsky1992] Sproat, R., J. Hirschberg, and D. Yarowsky. 1992. A corpus-based synthesizer. In *Proceedings of the International Conference on Spoken Language Processing*, pages 563–566, Banff, October. ICSLP.

[Swerts1998] Swerts, M. 1998. Filled pauses as markers of discourse structure. *Journal of Pragmatics*, 30:485–496.

[Swerts, Gelyukens, and Terken1992] Swerts, M., R. Gelyukens, and J. Terken. 1992. Prosodic correlates of discourse units in spontaneous speech. In *Proceedings*, pages 421–428, Banff, October. International Conference on Spoken Language Processing.

[Swerts and Ostendorf1997] Swerts, M. and M. Ostendorf. 1997. Prosodic and lexical indications of discourse structure in human-machine interactions. *Speech Communication*, 22:25–41.

[Swerts and Ostendorf1995] Swerts, Marc and Mari Ostendorf. 1995. Discourse prosody in human-machine interactions. In *Proceedings ESCA Workshop on Spoken Dialogue Systems: Theories and Applications*, pages 205–208, Visgo, Denmark, May/June. ESCA.

[Taylor et al.1998] Taylor, Paul, Simon King, Stephen Isard, and Helen Wright. 1998. Intonation and dialog context as constraints for speech recognition. *Language and Speech: Special Issue on Prosody and Conversation*, 41(3-4).

[Theune1999] Theune, Mariët. 1999. Parallelism, coherence, and contrastive accent. In *Proceedings of EUROSPEECH-99*, Budapest.

[Wade, Shriberg, and Price1992] Wade, E., E. E. Shriberg, and P. J. Price. 1992. User behaviors affecting speech recognition. In *Proceedings of ICSLP-92*, volume 2, pages 995–998, Banff.

[Walker et al.1997] Walker, Marilyn, Don Hindle, Jeanne Fromer, Giuseppe Di Fabbrizio, and Craig Mestel. 1997. Evaluating competing agent strategies for a voice email agent. In *Proceedings of EUROSPEECH-97*.

[Walker, Kamm, and Litman2000] Walker, Marilyn, Candace Kamm, and Diane Litman. 2000. Towards developing general models of usability with PARADISE. *Natural Language Engineering: Special Issue on Best Practice in Spoken Language Dialogue System Engineering*, 6, October.

[Woodbury1987] Woodbury, Anthony C. 1987. Rhetorical structure in a central Alaskan Yupik Eskimo traditional narrative. In J. Sherzer and A. Woodbury, editors, *Native American Discourse: Poetics and Rhetoric*. Cambridge University Press, Cambridge UK, pages 176–239.