

# Dialog Technology Bibliography

This bibliography is organized around the topics of the tutorial, as listed at <http://interspeech2015.org/program/tutorials/track-4-0930-1300/>. It includes landmark work, representative work, and selections from the research forefront, but is far from comprehensive.

August 20, 2015; reformatted August 4, 2017.

## A: Historical Context and Basic Concepts

**Telecommunications:** Brady (1969); Ramirez, Górriz, and Segura (2007); McFarland (2001); Ferrer, Shriberg, and Stolcke (2003); DeVault, Sagae, and Traum (2009); Matsuyama et al. (2009); Raux and Eskenazi (2012); Ghigi et al. (2014); Wright (2011a); Schoenenberg, Raake, and Koeppe (2014)

**Rapport:** Acosta and Ward (2011); Gratch et al. (2007); Schröder et al. (2012); Novick and Gris (2014); DeVault et al. (2014)

**Information Retrieval:** Hakkani-Tur et al. (1999); Ward and Vega (2012a); Ward et al. (2015)

**Information Delivery:** Schmandt (1994); Ward and Nakagawa (2004); Ofek, Iqbal, and Strauss (2013); Skantze, Hjalmarsson, and Oertel (2014); Yu, Bohus, and Horvitz (2015)

## B: Philosophical Interlude

**Philosophical Perspectives:** Dennett (1991); Clark (1996, 2002); Bunt (2011); Levinson (2006); Nass and Brave (2007); Garrod and Pickering (2004); Pentland (2012); Ginzburg and Fernández (2010)

**Psychological Perspectives:** Jaffe (1978); Bard et al. (2007); Brennan (2000); Brennan, Galati, and Kuhlen (2010); Roberts and Francis (2013); Stevanovic and Perakyla (2015); Levinson and Torreira (2015)

## C: Traditional Models and Alternatives

Winograd (1977); Young et al. (1989);

Shriberg et al. (1998); Traum and Heeman (1997); Walker and Passonneau (2001); Lemon, Cavedon, and Kelly (2003); Lee, Jung, and Lee (2008); Jurafsky and Martin (2009); Traum and Larsson (2003); Artstein et al. (2011); Tetreault and Litman (2006a); International Standards Organization (2012); Hough et al. (2015)

## E: Commercial Dialog Systems

McGlashan, Burnett, and others (2010); Cohen, Giangola, and Balogh (2004); Harris (2005); Suendermann, Liscombe, and Pieraccini (2010); Funakoshi et al. (2010); Witt (2015)

## F: Component Technologies and Dialog

**Speech Recognition:** Shriberg et al. (1998); Shriberg (2005); Gruenstein, Wang, and Seneff (2005); Fujisaki (2008); Raux et al. (2010); Ballinger et al. (2010); Ward and Vega (2012b); Stoyanchev and Stent (2012); Morbini et al. (2013)

**Language Understanding:** Hirschberg and Litman (1993); Nöth et al. (2002); Artzi and Zettlemoyer (2011)

**Speech Synthesis:** Syrdal et al. (2010); Baumann and Schlangen (2011); Andersson, Yamagishi, and Clark (2012); Georgila et al. (2012); Skantze and Hjalmarsson (2013); Baumann and Schlangen (2013); Pammi, Schröder, and Charfuelan (2011); Pammi (2011); Black, Bunnell, and others (2012); Corley and Hartsuiker (2003); Wright (2011b)

**Language Generation:** Stoia et al. (2006); Mairesse and Young (2014); Stent and Bangalore (2014)

## **H: Dialog Systems: The Research Front**

**Multimodal:** Cassell et al. (2001); Kawahara, Iwatate, and Takanashi (2012); Skantze, Hjalmarsson, and Oertel (2013); Johansson, Skantze, and Gustafson (2013); Mutlu et al. (2012); Bohus and Horvitz (2014)

**Multiparty:** Kawahara et al. (2010); Otsuka (2011)

**Situated:** Bohus, Kamar, and Horvitz (2012)

**User-State Modeling and Adaptation:** Komatani et al. (2005)

**Agents and Robots:** Cassell, Bickmore, and others (1999); Admoni and Scasselati (2014); Cuayahuitl et al. (2014); Foster, Keizer, and Lemon (2014); Mariani et al. (2014)

**Dialog-Quality Metrics and Evaluation Techniques:** Möller, Engelbrecht, and Schleicher (2008); Schmitt and Ultes (2015)

**Dynamical Models:** Choudhury and Basu (2004)

**Overview:** Approaches to Dialogue Systems and Dialogue Management. David Traum, 2012. Slides at <http://people.ict.usc.edu/~traum/Talks/cs544dialogue3-8-12.pdf>

**Challenges for Dialog-Systems Research:** Bohus, Kamar, and Horvitz (2012); Ward and DeVault (2015)

## **I1: Other Applications Involving Dialog Knowledge**

**Speech-to-Speech Translation:** Levin et al. (2003); Sridhar, Bangalore, and Narayanan (2013); Hara and Iqbal (2015)

**Summarization :** Purver et al. (2007); Wang and Liu (2015)

**Tutorial Systems:** Forbes-Riley and Lit-

man (2011b); Katz et al. (2013)

**Dialog-Skills Coaching, Training, and Assessment:** Ward et al. (2007); Hoque et al. (2013); Mitchell, Evanini, and Zechner (2014)

**Diagnosis:** Warlaumont et al. (2010); Heeman et al. (2010); Duff et al. (2012)

**Personality Assessment and Social Role Detection:** Jayagopi et al. (2009); Gatica-Perez (2006)

**Engagement Detection:** Yu, Aoki, and Woodruff (2004); Hsiao, Jih, and Hsu (2012); Ofek, Iqbal, and Strauss (2013); Bohus and Horvitz (2014); Voigt, Podesva, and Jurafsky (2014)

**Activity Detection:** Gravano et al. (2011); Ranganath, Jurafsky, and McFarland (2013)

**Outcomes Prediction, Analytics:** Kiekel et al. (2002); Curhan and Pentland (2007); Park et al. (2015); Ezen-Can et al. (2015)

**Human-Dialog Support:** Arnott and Alm (2013)

## **I2: Learning**

**Reinforcement Learning:** Levin, Pieraccini, and Eckert (2000); Singh et al. (2002); Tetreault and Litman (2006b); Williams and Young (2007); Henderson, Lemon, and Georgila (2008); Gasic and Young (2011); Georgila, Nelson, and Traum (2014); Kim et al. (2014); Efstathiou and Lemon (2015)

**Unsupervised Learning:** Gandhe and Traum (2007); Paek and Pieraccini (2008); Georgila, Wolters, and Moore (2010); Ward and Vega (2012a); Boyer et al. (2009); Griol, Riccardi, and Sanchis (2009); Goldwasser and Daume III (2014)

## **I3: Challenging Phenomena**

**General Inventories:** Tannen (1989); Aijmer (1996); Sidnell (2011); Szczepek Reed (2010); Ward and Vega (2012a)

**Feedback:** Bavelas, Coates, and Johnson (2000); Ward and Al Bayyari (2010); Skantze,

Oertel, and Hjalmarsson (2013); Meena, Skantze, and Gustafson (2014); Skantze, Oertel, and Hjalmarsson (2014)

**Error Recovery:** Skantze (2005, 2008)

**Confidence:** Forbes-Riley and Litman (2011a)

**Territories of Knowledge:** Heritage (2012)

**Accommodation:** Giles et al. (1987); Levitan, Gravano, and Hirschberg (2011)

**Agreement and Disagreement:** Wang et al. (2011)

**Dialog Styles:** Garera and Yarowsky (2009); Grothendieck, Gorin, and Borges (2011)

## Resources

**Corpora and Shared Tasks:** Anderson et al. (1991); Calhoun et al. (2010); Williams et al. (2014)

**Organizations and Meetings:** SigDIAL, ISCA, Interspeech, Sigdial, IWSDS, Semdial, YRRSDS, AVIOS (the Applied Voice Input-Output Society)

**Toolkits:** Baumann and Schlangen (2012); Skantze and Al Moubayed (2012); Lison (2015); Ward (2015)

See also the Sigdial Bibliography at <http://www.sigdial.org/biblio>

## References

- Acosta, J. C., and Ward, N. G. 2011. Achieving rapport with turn-by-turn, user-responsive emotional coloring. *Speech Communication* 53:1137–1148.
- Admoni, H., and Scassellati, B. 2014. Data-driven model of nonverbal behavior for socially assistive human-robot interactions.

In *16th International Conference on Multimodal Interaction (ICMI)*, 196–199. ACM.

Aijmer, K. 1996. *Conversational routines in English: Convention and creativity*. Longman.

Anderson, A. H.; Bader, M.; Bard, E. G.; Boyle, E.; Doherty, G.; Garrod, S.; Isard, S.; Kowtko, J.; McAllister, J.; Miller, J.; et al. 1991. The HCRC map task corpus. *Language and Speech* 34:351–366.

Andersson, S.; Yamagishi, J.; and Clark, R. A. J. 2012. Synthesis and evaluation of conversational characteristics in HMM-based speech synthesis. *Speech Communication* 54:175–188.

Arnott, J. L., and Alm, N. 2013. Towards the improvement of augmentative and alternative communication through the modelling of conversation. *Computer Speech & Language* 27:1194–1211.

Artstein, R.; Rushforth, M.; Gandhe, S.; Traum, D.; and Donigian, A. 2011. Limits of simple dialogue acts for tactical questioning dialogues. In *Proceedings of the 7th IJCAI workshop on knowledge and reasoning in practical dialogue systems*, 1–8.

Artzi, Y., and Zettlemoyer, L. 2011. Bootstrapping semantic parsers from conversations. In *Empirical Methods in Natural Language Processing*, 421–432.

Ballinger, B.; Allauzen, C.; Gruenstein, A.; and Schalkwyk, J. 2010. On-demand language model interpolation for mobile speech input. In *Interspeech*, 1812–1815.

Bard, E. G.; Anderson, A. H.; Chen, Y.; Nicholson, H. B. M.; Havard, C.; and Dalziel-Job, S. 2007. Let’s you do that: Sharing the cognitive burdens of dialogue. *Journal of Memory and Language* 57:616–641.

Baumann, T., and Schlangen, D. 2011. Predicting the micro-timing of user input for an incremental spoken dialogue system that completes a user’s ongoing turn. In *12th SigDial*.

- Baumann, T., and Schlangen, D. 2012. The inproctk 2012 release. In *NAACL-HLT Workshop on Future Directions and Needs in the Spoken Dialog Community: Tools and Data*, 29–32.
- Baumann, T., and Schlangen, D. 2013. Open-ended, extensible system utterances are preferred, even if they require filled pauses. *SIGdial*.
- Bavelas, J.; Coates, L.; and Johnson, T. 2000. Listeners as co-narrators. *Journal of Personality and Social Psychology* 79:941–952.
- Black, A. W.; Bunnell, H. T.; et al. 2012. Articulatory features for expressive speech synthesis. In *Int'l. Conf. on Acoustics, Speech and Signal Processing, IEEE*, 4005–4008.
- Bohus, D., and Horvitz, E. 2014. Managing human-robot engagement with forecasts and... um... hesitations. In *Proceedings of the 16th International Conference on Multimodal Interaction*, 2–9. ACM.
- Bohus, D.; Kamar, E.; and Horvitz, E. 2012. Towards situated collaboration. In *NAACL-HLT Workshop on Future Directions and Needs in the Spoken Dialog Community: Tools and Data*, 13–14. Association for Computational Linguistics.
- Boyer, K. E.; Ha, E. Y.; Phillips, R.; Wallis, M. D.; Vouk, M. A.; and Lester, J. C. 2009. Inferring tutorial dialogue structure with hidden Markov modeling. In *Proc. NAACL-HLT Workshop on Innovative Uses of NLP for Building Educational Applications*, 19–26.
- Brady, P. T. 1969. A model for generating on-off speech patterns in two-way conversation. *Bell System Technical Journal* 48(7):2445–2472.
- Brennan, S. E.; Galati, A.; and Kuhlen, A. K. 2010. Two minds, one dialog: Coordinating speaking and understanding. In Ross, B. H., ed., *The Psychology of Learning and Motivation, volume 53*. Elsevier. 301–344.
- Brennan, S. E. 2000. Processes that shape conversation and their implications for computational linguistics. In *Proceedings, 38th Annual Meeting of the ACL*.
- Bunt, H. 2011. Multifunctionality in dialogue. *Computer Speech and Language* 25:222–245.
- Calhoun, S.; Carletta, J.; Brenier, J. M.; Mayo, N.; Jurafsky, D.; et al. 2010. The NXT-format Switchboard corpus: a rich resource for investigating the syntax, semantics, pragmatics and prosody of dialogue. *Language Resources and Evaluation* 44:387–419.
- Cassell, J.; Nakano, Y. I.; Bickmore, T. W.; Sidner, C. L.; and Rich, C. 2001. Non-verbal cues for discourse structure. In *Proceedings of the 39th Annual Meeting on Association for Computational Linguistics*, 114–123. Association for Computational Linguistics.
- Cassell, J.; Bickmore, T.; et al. 1999. Embodiment in conversational interfaces: Rea. In *CHI '99*, 520–527. ACM Press.
- Choudhury, T., and Basu, S. 2004. Modeling conversational dynamics as a mixed-memory markov process. In *NIPS*.
- Clark, H. H. 1996. *Using Language*. Cambridge University Press.
- Clark, H. H. 2002. Speaking in time. *Speech Communication* 36:5–13.
- Cohen, M. H.; Giangola, J. P.; and Balogh, J. 2004. *Voice User Interface Design*. Addison-Wesley.
- Corley, M., and Hartsuiker, R. J. 2003. Hesitation in speech can ... um ... help a listener understand. In *Proceedings of the 25th Meeting of the Cognitive Science Society*.
- Cuayahuitl, H.; Frommberger, L.; Dethlefs, N.; Raux, A.; Marge, M.; and Zender, H. 2014. Introduction to the special issue on machine learning for multiple modalities in interactive systems and robots. *ACM Transactions on Interactive Intelli-*

- gent Systems (TiiS)* 4:1–6.
- Curhan, J. R., and Pentland, A. 2007. Thin slices of negotiation: predicting outcomes from conversational dynamics within the first 5 minutes. *Journal of Applied Psychology* 92:802.
- Dennett, D. C. 1991. *Consciousness Explained*. Penguin.
- DeVault, D.; Artstein, R.; Benn, G.; Dey, T.; Fast, E.; Gainer, A.; Georgila, K.; Gratch, J.; Hartholt, A.; Lhommet, M.; Lucas, G.; Marsella, S.; Morbini, F.; Nazarian, A.; Scherer, S.; Stratou, G.; Suri, A.; Traum, D.; Wood, R.; Xu, Y.; Rizzo, A.; and Morency, L.-P. 2014. SimSensei Kiosk: A virtual human interviewer for healthcare decision support. In *Proceedings of AAAI-MAS*.
- DeVault, D.; Sagae, K.; and Traum, D. 2009. Can I finish? learning when to respond to incremental interpretation results in interactive dialogue. In *10th SigDial*.
- Duff, M. C.; Mutlu, B.; Byom, L.; and Turkstra, L. S. 2012. Beyond utterances: Distributed cognition as a framework for studying discourse in adults with acquired brain injury. In *Seminars in speech and language*, volume 33, 44.
- Efstathiou, I., and Lemon, O. 2015. Learning non-cooperative dialogue policies to beat opponent models. In *SemDial*, 33–41.
- Ezen-Can, A.; Grafsgaard, J. F.; Lester, J. C.; and Boyer, K. E. 2015. Classifying student dialogue acts with multimodal learning analytics. In *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge*, 280–289. ACM.
- Ferrer, L.; Shriberg, E.; and Stolcke, A. 2003. A prosody-based approach to end-of-utterance detection that does not require speech recognition. In *ICAASP*.
- Forbes-Riley, K., and Litman, D. 2011a. Benefits and challenges of real-time uncertainty detection and adaptation in a spoken dialogue computer tutor. *Speech Communication* 53:1115–1136.
- Forbes-Riley, K., and Litman, D. 2011b. Designing and evaluating a wizarded uncertainty-adaptive spoken dialogue tutoring system. *Computer Speech and Language* 25:105–126.
- Foster, M. E.; Keizer, S.; and Lemon, O. 2014. Towards action selection under uncertainty for a socially aware robot bartender. In *ACM/IEEE international conference on human-robot interaction*, 158–159.
- Fujisaki, H. 2008. In search of models in speech communication research. In *Interspeech*, 1–10.
- Funakoshi, K.; Nakano, M.; Kobayashi, K.; Komatsu, T.; and Yamada, S. 2010. Non-humanlike spoken dialogue: a design perspective. In *Proceedings of the 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, 176–184.
- Gandhe, S., and Traum, D. R. 2007. Creating spoken dialogue characters from corpora without annotations. In *Interspeech*, 2201–2204.
- Garera, N., and Yarowsky, D. 2009. Modeling latent biographic attributes in conversational genres. In *47th ACL*, 710–718.
- Garrod, S., and Pickering, M. J. 2004. Why is conversation so easy? *Trends in Cognitive Sciences* 8:8–11.
- Gasic, M., and Young, S. 2011. Effective handling of dialogue state in the hidden information state POMDP-based dialogue manager. *ACM Transactions on Speech and Language Processing* 7.
- Gatica-Perez, D. 2006. Analyzing group interactions in conversations: a review. In *Multisensor Fusion and Integration for Intelligent Systems, 2006 IEEE International Conference on*, 41–46. IEEE.
- Georgila, K.; Black, A.; Sagae, K.; and Traum, D. R. 2012. Practical evaluation of human and synthesized speech for vir-

- tual human dialogue systems. In *LREC*, 3519–3526.
- Georgila, K.; Nelson, C.; and Traum, D. 2014. Single-agent vs. multi-agent techniques for concurrent reinforcement learning of negotiation dialogue policies. In *ACL*, 500–510.
- Georgila, K.; Wolters, M. K.; and Moore, J. D. 2010. Learning dialogue strategies from older and younger simulated users. In *Proceedings of the 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, 103–106. Association for Computational Linguistics.
- Ghigi, F.; Eskenazi, M.; Torres, M. I.; and Lee, S. 2014. Incremental dialog processing in a task-oriented dialog. In *Interspeech*, 308–312.
- Giles, H.; Mulac, A.; Bradac, J. J.; and Johnson, P. 1987. Speech accommodation theory: The first decade and beyond. In McLaughlin, M. L., ed., *Communication Yearbook 10*. Sage. 13–48.
- Ginzburg, J., and Fernández, R. 2010. Computational models of dialogue. In Clark, A.; Fox, C.; and Lappin, S., eds., *Handbook of Computational Linguistics and Natural Language Processing*. Blackwell.
- Goldwasser, D., and Daume III, H. 2014. I object: Modeling latent pragmatic effects in courtroom dialogues. In *European ACL*, 655–663.
- Gratch, J.; Wang, N.; Okhmatovskaia, A.; Lamothe, F.; Morales, M.; van der Werf, R.; and Morency, L.-P. 2007. Can virtual humans be more engaging than real ones? *Lecture Notes in Computer Science* 4552:286–297.
- Gravano, A.; Levitan, R.; Willson, L.; Benus, S.; Hirschberg, J.; and Nenkova, A. 2011. Acoustic and prosodic correlates of social behavior. In *Interspeech*, 97–100.
- Griol, D.; Riccardi, G.; and Sanchis, E. 2009. Learning the structure of human-computer and human-human dialogs. In *Interspeech*, 2775–2778.
- Grothendieck, J.; Gorin, A. L.; and Borges, N. M. 2011. Social correlates of turn-taking style. *Computer Speech and Language* 25:789–801.
- Gruenstein, A.; Wang, C.; and Seneff, S. 2005. Context-sensitive statistical language modeling. In *Interspeech*.
- Hakkani-Tur, D.; Tur, G.; Stolcke, A.; and Shriberg, E. E. 1999. Combining words and prosody for information extraction from speech. In *Proc. Eurospeech*, vol. 5, 1991–1994.
- Hara, K., and Iqbal, S. T. 2015. Effect of machine translation in interlingual conversation: Lessons from a formative study. In *CHI: 33rd Annual ACM Conference on Human Factors in Computing Systems*, 3473–3482.
- Harris, R. A. 2005. *Voice Interaction Design: Crafting the New Conversational Speech Systems*. Morgan Kaufmann.
- Heeman, P. A.; Lunsford, R.; Selfridge, E.; Black, L.; and van Santen, J. 2010. Autism and interactional aspects of dialogue. In *Sigdial*.
- Henderson, J.; Lemon, O.; and Georgila, K. 2008. Hybrid reinforcement/supervised learning of dialogue policies from fixed data sets. *Computational Linguistics* 34:487–511.
- Heritage, J. 2012. Epistemics in action: Action formation and territories of knowledge. *Research on Language & Social Interaction* 45(1):1–29.
- Hirschberg, J., and Litman, D. 1993. Empirical studies on the disambiguation of cue phrases. *Computational Linguistics* 19:501–530.
- Hoque, M. E.; Courgeon, M.; Martin, J.-C.; Mutlu, B.; and Picard, R. W. 2013. Mach: My automated conversation coach. In *Proceedings of the 2013 ACM international joint conference on pervasive and ubiquitous computing*, 697–706.

- Hough, J.; de Kok, I.; Schlangen, D.; and Kopp, S. 2015. Timing and grounding in motor skill coaching interaction: Consequences for the information state. *Sem-Dial*.
- Hsiao, J. C.-Y.; Jih, W.-R.; and Hsu, J. Y.-J. 2012. Recognizing continuous social engagement level in dyadic conversation by using turn-taking and speech emotion patterns. In *Activity Context Representation Workshop at AAAI*.
- International Standards Organization. 2012. Language resource management – semantic annotation framework (SemAF) – part 2: Dialogue acts. ISO 24618-2:2012.
- Jaffe, J. 1978. Parliamentary procedure and the brain. In Siegman, A. W., and Feldstein, S., eds., *Nonverbal Behavior and Communication*. Lawrence Erlbaum Associates. 55–66.
- Jayagopi, D. B.; Hung, H.; Yeo, C.; and Gatica-Perez, D. 2009. Modeling dominance in group conversations from non-verbal activity cues. *IEEE Transactions on Audio, Speech and Language Processing* 17:501–517.
- Johansson, M.; Skantze, G.; and Gustafson, J. 2013. Head pose patterns in multiparty human-robot team-building interactions. In *Proceedings of the International Conference on Social Robotics - ICSR 2013*.
- Jurafsky, D., and Martin, J. H. 2009. *Speech and Language Processing, Chapter 24: Dialogue and Conversational Agents*. Pearson.
- Katz, S.; Albacete, P.; Ford, M. J.; Jordan, P.; Lipschultz, M.; Litman, D.; Siliman, S.; and Wilson, C. 2013. Pilot test of a natural-language tutoring system for physics that simulates the highly interactive nature of human tutoring. In *Artificial Intelligence in Education (16th International Conference)*, 636–639. Springer.
- Kawahara, T.; Sumi, K.; Chang, Z. Q.; and Takanashi, K. 2010. Detection of hot spots in poster conversations based on reactive tokens of audience. In *Interspeech*, 3042–3045.
- Kawahara, T.; Iwatate, T.; and Takanashi, K. 2012. Prediction of turn-taking by combining prosodic and eye-gaze information in poster conversations. In *Interspeech*.
- Kiekel, P. A.; Cooke, N. J.; Foltz, P. W.; Gorman, J.; and Martin, M. 2002. Some promising results of communication-based automatic measures of team cognition. In *Proceedings of the Human Factors and Ergonomic Society*, volume 46.
- Kim, D.; Breslin, C.; Tsiakoulis, P.; Gasic, M.; Henderson, M.; and Young, S. 2014. Inverse reinforcement learning for micro-turn management. In *Interspeech*.
- Komatani, K.; Ueno, S.; Kawahara, T.; and Okuno, H. G. 2005. User modeling in spoken dialogue systems to generate flexible guidance. *User Modeling and User-Adapted Interaction* 15:169–183.
- Lee, C.; Jung, S.; and Lee, G. G. 2008. Robust dialog management with n-best hypotheses using dialog examples and agenda. In *ACL*, 630–637.
- Lemon, O.; Cavedon, L.; and Kelly, B. 2003. Managing dialogue interaction: A multi-layered approach. In *4th (ACL) SIGdial Workshop on Discourse and Dialogue*.
- Levin, L.; Langley, C.; Lavie, A.; Gates, D.; Wallace, D.; and Peterson, K. 2003. Domain specific speech acts for spoken language translation. In *SIGDIAL*.
- Levin, E.; Pieraccini, R.; and Eckert, W. 2000. A stochastic model of human-machine interaction for learning dialog strategies. *Speech and Audio Processing, IEEE Transactions on* 8:11–23.
- Levinson, S. C., and Torreira, F. 2015. Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology* 6.
- Levinson, S. C. 2006. On the human ‘interaction engine’. In Enfield, N. J., and Levinson, S. C., eds., *Roots of Human So-*

- ciality: Culture, Cognition and Human Interaction*. Berg. 39–69.
- Levitan, R.; Gravano, A.; and Hirschberg, J. 2011. Entrainment in speech preceding backchannels. In *ACL 2011: Proceedings of the Association for Computational Linguistics Annual Meeting*.
- Lison, P. 2015. A hybrid approach to dialogue management based on probabilistic rules. *Computer Speech & Language* 34:232–255.
- Mairesse, F., and Young, S. 2014. Stochastic language generation in dialogue using factored language models. *Computational Linguistics* 40:763–799.
- Mariani, J.; Rosset, S.; Garnier-Rizet, M.; and Devillers, L. 2014. *Natural Interaction with Robots, Knowbots and Smartphones: Putting Spoken Dialog Systems Into Practice*. Springer.
- Matsuyama, K.; Komatani, K.; Ogata, T.; and Okuno, H. G. 2009. Enabling a user to specify an item at any time during system enumeration: Item identification for barge-in-able conversational dialogue systems. In *Interspeech*, 252–255.
- McFarland, D. H. 2001. Respiratory markers of conversational interaction. *Journal of Speech, Language and Hearing Research* 44:128.
- McGlashan, S.; Burnett, D. C.; et al. 2010. Voice extensible markup language (VoiceXML) 3.0. Technical report, W3C.
- Meena, R.; Skantze, G.; and Gustafson, J. 2014. Data-driven models for timing feedback responses in a Map Task dialogue system. *Computer Speech & Language* 28:903–922.
- Mitchell, C. M.; Evanini, K.; and Zechner, K. 2014. A dialogue-based spoken dialogue system for assessment of English language learners. In *Proceedings of the International Workshop on Spoken Dialogue Systems (IWSDS)*.
- Möller, S.; Engelbrecht, K.-P.; and Schleicher, R. 2008. Predicting the quality and usability of spoken dialog services. *Speech Communication* 50:730–744.
- Morbini, F.; Audhkhasi, K.; Sagae, K.; Artstein, R.; Can, D.; Georgiou, P.; Narayanan, S.; Leuski, A.; and Traum, D. 2013. Which ASR should I choose for my dialogue system. In *Proceedings of the 14th annual SIGdial Meeting on Discourse and Dialogue*, 394–403.
- Mutlu, B.; Kanda, T.; Forlizzi, J.; Hodgins, J.; and Ishiguro, H. 2012. Conversational gaze mechanisms for humanlike robots. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 1(2):12:1–12:33.
- Nass, C., and Brave, S. 2007. *Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship*. MIT Press.
- Nöth, E.; Batliner, A.; Warnke, V.; Haas, J.; Boros, M.; Buckow, J.; Huber, R.; Gallwitz, F.; Nutt, M.; and Niemann, H. 2002. On the use of prosody in automatic dialogue understanding. *Speech Communication* 36:45–62.
- Novick, D., and Gris, I. 2014. Building rapport between human and ECA: A pilot study. In *Human-Computer Interaction. Advanced Interaction Modalities and Techniques*. Springer. 472–480.
- Ofek, E.; Iqbal, S. T.; and Strauss, K. 2013. Reducing disruption from subtle information delivery during a conversation: mode and bandwidth investigation. In *Proceedings of the 2013 ACM annual conference on Human factors in computing systems*, 3111–3120. ACM.
- Otsuka, K. 2011. Multimodal conversation scene analysis for understanding peoples communicative behaviors in face-to-face meetings. In *Human Interface and the Management of Information. Interacting with Information*. Springer. 171–179.
- Paek, T., and Pieraccini, R. 2008. Automating spoken dialogue management design using machine learning: An industry per-

- spective. *Speech Communication* 50:716–729.
- Pammi, S.; Schröder, M.; and Charfuelan, M. 2011. Multidimensional meaning annotation of listener vocalizations for synthesis. In *Proceedings of the Workshop on Emotion and Computing (Berlin, Germany)*, Springer.
- Pammi, S. C. 2011. *Synthesis of listener vocalizations: Towards interactive speech synthesis*. Ph.D. Dissertation, University of the Saarland.
- Park, S.; Scherer, S.; Gratch, J.; Carnevale, P.; and Morency, L.-P. 2015. I can already guess your answer: Predicting respondent reactions during dyadic negotiation. *IEEE Transactions on Affective Computing* 6.
- Pentland, A. 2012. To signal is human. *American Scientist* 98:204–211.
- Purver, M.; Dowding, J.; Niekrasz, J.; Ehlen, P.; Noorbalooci, S.; and Peters, S. 2007. Detecting and summarizing action items in multi-party dialogue. In *Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue*, 200–211.
- Ramirez, J.; Górriz, J. M.; and Segura, J. C. 2007. Voice activity detection: fundamentals and speech recognition system robustness. In *Robust Speech Recognition and Understanding*. I-Tech. 460–480.
- Ranganath, R.; Jurafsky, D.; and McFarland, D. 2013. Detecting friendly, flirtatious, awkward, and assertive speech in speed-dates. *Computer Speech and Language* 27:89–115.
- Raux, A., and Eskenazi, M. 2012. Optimizing the turn-taking behavior of task-oriented spoken dialog systems. *ACM Transactions on Speech and Language Processing (TSLP)* 9:1–23.
- Raux, A.; Mehta, N.; Ramachandran, D.; and Gupta, R. 2010. Dynamic language modeling using Bayesian networks for spoken dialog systems. In *Interspeech*, 3030–3033.
- Roberts, F., and Francis, A. L. 2013. Identifying a temporal threshold of tolerance for silent gaps after requests. *Journal of the Acoustical Society of America* 133:471–477.
- Schmandt, C. 1994. *Computers and Communication*. Van Nostrand Reinhold.
- Schmitt, A., and Ultes, S. 2015. Interaction quality: Assessing the quality of ongoing spoken dialog interaction by expert-sand how it relates to user satisfaction. *Speech Communication*.
- Schoenenberg, K.; Raake, A.; and Koeppel, J. 2014. Why are you so slow? misattribution of transmission delay to attributes of the conversation partner at the far-end. *International Journal of Human-Computer Studies* 72:477–487.
- Schröder, M.; Bevacqua, E.; Cowie, R.; Eyben, F.; Gunes, H.; Heylen, D.; ter Maat, M.; Gary, Pammi, S.; Pantic, M.; Pelachaud, C.; Schuller, B.; de Sevin, E.; Valstar, M.; and Wollmer, M. 2012. Building autonomous sensitive artificial listeners. *IEEE Transactions on Affective Computing* 3:165–183.
- Shriberg, E.; Bates, R.; Stolcke, A.; Taylor, P.; Jurafsky, D.; Ries, K.; Coccaro, N.; Martin, R.; Meteer, M.; and Van Ess-Dykema, C. 1998. Can prosody aid the automatic classification of dialog acts in conversational speech? *Language and Speech* 41:439–487.
- Shriberg, E. E. 2005. Spontaneous speech: How people really talk, and why engineers should care. In *Interspeech*. Lisbon.
- Sidnell, J. 2011. *Conversation analysis: An introduction*. John Wiley & Sons.
- Singh, S.; Litman, D.; Kearns, M.; and Walker, M. 2002. Optimizing dialog management with reinforcement learning: Experiments with the NJFun system. *Journal of Artificial Intelligence Research* 16:105–133.
- Skantze, G., and Al Moubayed, S. 2012.

- IrisTK: a statechart-based toolkit for multi-party face-to-face interaction. In *ICMI*.
- Skantze, G., and Hjalmarsson, A. 2013. Towards incremental speech generation in conversational systems. *Computer Speech and Language* 27:243–262.
- Skantze, G.; Hjalmarsson, A.; and Oertel, C. 2013. Exploring the effects of gaze and pauses in situated human-robot interaction. In *14th Sigdial*.
- Skantze, G.; Hjalmarsson, A.; and Oertel, C. 2014. Turn-taking, feedback and joint attention in situated human-robot interaction. *Speech Communication* 65:50–66.
- Skantze, G.; Oertel, C.; and Hjalmarsson, A. 2013. User feedback in human-robot interaction: Prosody, gaze and timing. *Inter-speech*.
- Skantze, G.; Oertel, C.; and Hjalmarsson, A. 2014. User feedback in human-robot dialogue: Task progression and uncertainty. In *HRI Workshop on Timing in Human-Robot Interaction*.
- Skantze, G. 2005. Exploring human error recovery strategies: Implications for spoken dialogue systems. *Speech Communication* 45:325–341.
- Skantze, G. 2008. Galatea: A discourse modeller supporting concept-level error handling in spoken dialogue systems. In Dybkjaer, L., and Minker, W., eds., *Recent Trends in Discourse and Dialogue*. Springer.
- Sridhar, V. K. R.; Bangalore, S.; and Narayanan, S. 2013. Enriching machine-mediated speech-to-speech translation using contextual information. *Computer Speech & Language* 27:492–508.
- Stent, A., and Bangalore, S. 2014. *Natural Language Generation in Interactive Systems*. Cambridge University Press.
- Stevanovic, M., and Perakyla, A. 2015. Experience sharing, emotional reciprocity, and turn-taking. *Frontiers in Psychology* 6.
- Stoia, L.; Byron, D. K.; Shockley, D. M.; and Fosler-Lussier, E. 2006. Sentence planning for realtime navigational instructions. In *Human Language Technology (NAACL)*, 157–160.
- Stoyanchev, S., and Stent, A. 2012. Concept type prediction and responsive adaptation in a dialogue system. *Dialogue and Discourse* 3.
- Suendermann, D.; Liscombe, J.; and Pieracini, R. 2010. Optimize the obvious: Automatic call flow generation. In *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, 5370–5373.
- Syrdal, A. K.; Conkie, A.; Kim, Y.-J.; and Beutnagel, M. C. 2010. Speech acts and dialog TTS. In *Speech Synthesis Workshop*, 179–183.
- Szczepek Reed, B. 2010. *Analysing Conversation: An introduction to prosody*. Palgrave Macmillan.
- Tannen, D. 1989. *That’s Not What I Meant! How Conversational Style Makes or Breaks Relationships*. Ballantine.
- Tetreault, J. R., and Litman, D. J. 2006a. Comparing the utility of state features in spoken dialogue using reinforcement learning. In *HLT-NAACL*, 272–279.
- Tetreault, J. R., and Litman, D. J. 2006b. Using reinforcement learning to build a better model of dialogue state. In *EACL*.
- Traum, D., and Heeman, P. 1997. Utterance units in spoken dialogue. In Maier, E.; Mast, M.; and LuperFoy, S., eds., *Processing in Spoken Language Systems*. Springer-Verlag. 125–140.
- Traum, D., and Larsson, S. 2003. The information state approach to dialogue management. In van Kuppevelt, J., and Smith, R., eds., *Current and New Directions in Discourse and Dialogue*. Kluwer. 325–353.
- Voigt, R.; Podesva, R. J.; and Jurafsky, D. 2014. Speaker movement correlates with prosodic indicators of engagement. In

*Speech Prosody.*

- Walker, M., and Passonneau, R. 2001. Date: a dialogue act tagging scheme for evaluation of spoken dialogue systems. In *Human language technology research*, 1–8.
- Wang, D., and Liu, Y. 2015. Opinion summarization on spontaneous conversations. *Computer Speech and Language* 34:61–82.
- Wang, W.; Yaman, S.; Precoda, K.; Richey, C.; and Raymond, G. 2011. Detection of agreement and disagreement in broadcast conversations. In *49th ACL-HLT*, 374–378. Association for Computational Linguistics.
- Ward, N. G., and Al Bayyari, Y. 2010. American and Arab perceptions of an Arabic turn-taking cue. *Journal of Cross-Cultural Psychology* 41:270–275.
- Ward, N. G., and DeVault, D. 2015. Ten challenges in highly-interactive dialog systems. In *AAAI Symposium on Turn-taking and Coordination in Human-Machine Interaction*.
- Ward, N., and Nakagawa, S. 2004. Automatic user-adaptive speaking rate selection. *International Journal of Speech Technology* 7:235–238.
- Ward, N. G., and Vega, A. 2012a. A bottom-up exploration of the dimensions of dialog state in spoken interaction. In *13th Annual SIGdial Meeting on Discourse and Dialogue*.
- Ward, N. G., and Vega, A. 2012b. Towards empirical dialog-state modeling and its use in language modeling. In *Interspeech*.
- Ward, N. G.; Escalante, R.; Al Bayyari, Y.; and Solorio, T. 2007. Learning to show you’re listening. *Computer Assisted Language Learning* 20:385–407.
- Ward, N. G.; Werner, S. D.; Garcia, F.; and Sanchis, E. 2015. A prosody-based vector-space model of dialog activity for information retrieval. *Speech Communication* 68:86–96.
- Ward, N. G. 2015. Midlevel prosodic features toolkit. <https://github.com/nigelgward/midlevel>.
- Warlaumont, A. S.; Oller, D. K.; Dale, R.; Richards, J. A.; Gilkerson, J.; and Xu, D. 2010. Vocal interaction dynamics of children with and without autism. In *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*, 121–126.
- Williams, J. D., and Young, S. 2007. Partially observable markov decision processes for spoken dialog systems. *Computer Speech & Language* 21:393–422.
- Williams, J. D.; Henderson, M.; Raux, A.; Thomson, B.; Black, A.; and Ramachandran, D. 2014. The dialog state tracking challenge series. *AI Magazine* 35(Winter):121–123.
- Winograd, T. 1977. A framework for understanding discourse. In Just, M. A., and Carpenter, P. A., eds., *Cognitive processes in comprehension*. Erlbaum. 63–88.
- Witt, S. 2015. Modeling user response timings in spoken dialog systems. *International Journal of Speech Technology* 18:231–243.
- Wright, M. 2011a. On clicks in English talk-in-interaction. *Journal of the International Phonetic Association* 41:207–229.
- Wright, M. 2011b. The phonetics–interaction interface in the initiation of closings in everyday English telephone calls. *Journal of Pragmatics* 43:1080–1099.
- Young, S. R.; Hauptmann, A. G.; Ward, W. H.; Smith, E. T.; and Werner, P. 1989. High level knowledge sources in usable speech recognition systems. *Communications of the ACM* 32:183–194.
- Yu, C.; Aoki, P. M.; and Woodruff, A. 2004. Detecting user engagement in everyday conversations. In *Interspeech*, 1329–1332.
- Yu, Z.; Bohus, D.; and Horvitz, E. 2015. Incremental coordination: Attention-centric speech production in a physically situated conversational agent. In *Sigdial*.