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.


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 Koepppe (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 Forefront

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 Scassellati (2014); Cuayahuitl et al. (2014); Foster, Keizer, and Lemon (2014); Mariani et al. (2014)


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)


Dialog-Skills Coaching, Training, and Assessment: Ward et al. (2007); Hoque et al. (2013); Mitchell, Evanini, and Zechnner (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); Ajijmer (1996); Sidnell (2011); Szczepen 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)


Confidence: Forbes-Riley and Litman (2011a)

Territories of Knowledge: Heritage (2012)

Accommodation: Giles et al. (1987); Leivitan, 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


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


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. 2013. Open-ended, extensible system utterances are preferred, even if they require filled pauses. SIGdial.


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-


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.*


Georgila, K.; Black, A.; Sagae, K.; and Traum, D. R. 2012. Practical evaluation of human and synthesized speech for vir-


Paek, T., and Pieraccini, R. 2008. Automating spoken dialogue management design using machine learning: An industry per-


IrisTK: a statechart-based toolkit for multi-party face-to-face interaction. In *ICMI*.


Voigt, R.; Podesva, R. J.; and Jurafsky, D. 2014. Speaker movement correlates with prosodic indicators of engagement. In


