

Table 0.1: Summary of Task-oriented Spoken Dialogue Systems								
System	Year	Task Description	Automatic Speech Recognizer (ASR)	Text-to-speech (TTS)	Dialogue Manager (DM)	Natural Language Understanding (NLU)	Natural Language Generation (NLG)	Training Data
ConQuest [1]	2007	Provides technical program information during conferences. Deployed in InterSpeech 2006 and IJCAI 2007.	CMU Sphinx2 [2].	Cepstral speech synthesis engine (http://www.cepstral.com).	RavenClaw architecture [3].	Phoenix semantic parser [4].	Rosetta, a template-based NLG component.	Training data is only used for ASR. It starts with data collected by a text-only prototype. After deployment they collected more data, transcribed it, and retrain the LM. LM used in InterSpeech 2006 were trained on a corpus of 6350 utterances.
Let’s DiSCoH [5]	2006	Users are able to call to learn about a conference, including paper submission, program, venue, etc. Designed to be highly portable and flexible across different conferences and workshops.	System developed using the AT&T VoiceTone Spoken Dialogue System tools [6], which provides services with ASR, SLU, DM and TTS. System uses a fixed set of responses so no NLG component is mentioned.					LM trained by W99 dataset + artificially generated + data from conference website + manually designed (11,275 + 9,511 + 226 + 467 sentences).
Let’s Go [7]	2005	Provide bus schedule information to the Pittsburgh population during off-peak times.	CMU Sphinx2.	Techniques in Limited Domain Synthesis [8]. Unit selection concatenative voice specifically designed for domain.	RavenClaw architecture.	Initially uses hand-coded Finite State Grammars. Finally uses tri-gram language models trained on artificial corpora.	Rosetta.	Data from real world: 614 dialogues, containing 7936 user turns. Manually transcribed and labeled.
LARRI [9]	2002	A multi-modal system for support of maintenance and repair activities for aircraft mechanics.	CMU Sphinx2.	Festival system in a limited domain mode. Use unit-selection synthesizer with a fall back on a diphone voice.	Behavior is specified through a task-dependent script. AGENDA dialogue manager [10].	Phoenix semantic parser.	Rosetta.	AM trained with WSJ0 corpus. Trigram LM trained with INS BIT Test procedure and general system commands.
NJFun [11]	2002	Provide telephone access to a database of activities in New Jersey.	Watson Speech Recognizer.	Concatenative diphone synthesis method.	Train by reinforcement learning (MDP) Build with DMD scripting language [12].	Watson Speech Recognizer.	Grammar and template.	Manually obtained by AT&T employees. 54 subjects for training and 21 for testing. 311 training dialogues, 124 testing dialogues.
CMU Communicator [13]	1999	Helps users create complex travel itineraries (multi-leg flights, hotel and car reservations). http://www.speech.cs.cmu.edu/Communicator/index.html	CMU Sphinx2.	Festival system in a limited domain mode with concatenative method.	Behavior is specified through a task-dependent script. AGENDA dialogue manager.	Phoenix semantic parser.	Template-driven.	Data collected in different stages [14]: 1) 48 human-human dialogues. 2) 107 Wizard-of-Oz Ver1 (WOZ). 3) 2983 from prototype system, manually transcribed. 4) 16 from WOZ ver2. Total 3164 dialogues.
There are several other systems whose architectures are similar to that of [1, 7, 9, 13]: RoomLine, Intelligent Procedure Assistant, Vera, MeetingLine, Team Talk, Sublime, Madeleine, RavenCalendar. http://www.cs.cmu.edu/~dbohus/ravenclaw-olympus/systems_overview.html								

References

- [1] D. Bohus and S. Puerto, “ConQuest: An open-source dialog system for conferences,” *Proceeding NAACL-Short ’07 Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Companion Volume, Short Papers*, no. April, pp. 9–12, 2007.
- [2] X. Huang, F. Alleva, M.-Y. Hwang, and R. Rosenfeld, “An overview of the SPHINX-II speech recognition system,” *Proceedings of the workshop on Human Language Technology - HLT ’93*, p. 81, 1993.
- [3] D. Bohus and A. I. Rudnicky, “RavenClaw : Dialog Management Using Hierarchical Task Decomposition and an Expectation Agenda,” *Phoenix Usa*, pp. 597–600, 2003.
- [4] W. Ward and S. Issar, “Recent improvements in the CMU spoken language understanding system,” *Proceedings of the workshop on Human Language Technology*, pp. 213–216, 1994.
- [5] O. Lemon, G. Andreani, G. D. Fabbrizio, M. Gilbert, D. Gillick, S. Village, and A. Piceno, “Let’s Discoh : Collecting an Annotated Open Corpus With Dialogue Acts,” pp. 2–5, 2006.
- [6] M. Gilbert, J. Wilpon, B. Stern, and G. Di Fabbrizio, “Intelligent virtual agents for contact center automation,” *IEEE Signal Processing Magazine*, vol. 22, no. 5, pp. 32–41, 2005.
- [7] A. Raux, B. Langner, and D. Bohus, “Let’s go public! taking a spoken dialog system to the real world,” *in Proc. of Interspeech . . .*, pp. 885–888, 2005.
- [8] A. W. Black and K. A. Lenzo, “Limited domain synthesis,” *International Conference on Spoken Language Processing - ICSLP*, pp. 411–414, 2000.
- [9] D. Bohus, “LARRI : A Language-Based Maintenance and Repair Assistant,” 2002.
- [10] A. Rudnicky and W. Xu, “An Agenda-based Dialog Management Architecture for Spoken Language Systems,” *IEEE Automatic Speech Recognition and Understanding Workshop*, pp. 1–337, 1999.
- [11] S. Singh, D. L. Man, M. Kearns, and M. Walker, “Optimizing dialogue management with reinforcement learning: Experiments with the NJFun system,” *Journal of Artificial Intelligence Research*, vol. 16, pp. 105–133, 2002.
- [12] E. Levin, R. Pieraccini, W. Eckert, P. D. Fabbrizio, S. Narayanan, P. Avenue, and F. Park, “Spoken Language Dialogue : From Theory To Practice,” 2003.
- [13] A. Rudnicky, E. Thayer, P. Constantinides, C. Tchou, R. Shern, K. Lenzo, W. Xu, and A. Oh, “Creating Natural Dialogs in the Carnegie Mellon Communicator System.,” *Proceedings of the Sixth European Conference on Speech Communication and Technology, (EUROSPEECH)*, vol. 4, pp. 1531–1534, 1999.
- [14] M. Eskenazi, A. Rudnicky, K. Gregory, P. Constantinides, R. Brennan, C. Bennett, and J. Allen, “Data Collection and Processing in the Carnegie Mellon Communicator,” *Proceedings of the 6th European Conference on Speech Communication and Technology*, pp. 2695–2698, 1999.