|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Paper | Year | Desciption | ASR | TTS | DM | NLU | NLG | Training Dialogues |
| Singh2002 | 2002 | Provide telephone access to a database of activities in New Jersey | Watson Speech Recognizer | Concatenative diphone synthsis | Train by reinforcement learning (MDP)  Build with DMD scripting language \cite{Levin2003} | 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 |
| Raux2005 | 2005 | Provide bus schedule information to the Pittsburgh population during off-peak times | CMU Sphinx2 | Techniques in Limited Domain Synthesis \cite{Black2000}  Unit selection concatenative voice specifically designed for domain | RavenClaw architecture \cite{Bohus2003} | Initially uses hand-coded Finite State Grammars  Finally uses tri-gram language models trained on artificial corpora | Rosetta, a template-based component | Data from real world  614 dialogues, containing 7936 user turns  Manually transcribed and labeled |
| Rudnicky1999 | 1999 | Helps users create complex travel itineraries (multi-leg flights, hotel and car reservations)  <http://www.speech.cs.cmu.edu/Communicator/index.html> (demo available) | CMU Sphinx2 | Festival system in a limited domain mode  Concatenative method | Behavior is specified through a task-dependent script  AGENDA dialogue manager \cite{ Rudnicky1999a} | Phoenix semantic parser \cite{Ward1994} | Template-driven | Data collected in different stages \cite{Eskenazi1999}:   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. |
| Bohus2007 | 2007 | Provides technical program information during conferences. Deployed in InterSpeech 2006 and IJCAI 2007 | CMU Sphinx2 | Cepstral speech synthesis engine (http://www.cepstral.com) | RavenClaw architecture | Phoenix semantic parser | Rosetta | 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 is InterSpeech were trained on a corpus of 6350 utterances. |
| Lemon2006 | 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 \cite{Gilbert2005}, 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) |
| Bohus2002 | 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 | Phoenix semantic parser | Rosetta | AM trained with WSJ0 corpus. Trigram LM trained with INS BIT Test procedure and general system commands. |
| There are several other systems whose architectures are similar to that of \cite{Bohus2007, Raux2005, Bohus2002, Rudnicky1999}: RoomLine, Intelligent Procedure Assistant, Vera, MeetingLine, Team Talk, Sublime, Madeleine, RavenCalendar. (http://www.cs.cmu.edu/~dbohus/ravenclaw-olympus/systems\_overview.html) | | | | | | | | |