

# CU Communicator Spoken Dialog System

## Project Overview:

The University of Colorado (CU) Communicator is an over-the-phone spoken dialog system for accessing live airline, hotel, and car rental availability and pricing. Users can dial up the system using a telephone and use natural language to interact with our automated travel agent. The agent then accesses the internet for flight, hotel, and car rental information.

The Communicator system, sponsored by DARPA, is our initial testbed for research leading to advanced dialog systems enabling robust and graceful human computer interaction. It is a Hub compliant system for the DARPA Communicator task, and was demonstrated at the DARPA workshop in June 1999. Robustness and portability of spoken dialog systems are two of the issues we attempt to address in the project. We use robust parsing and dialog control strategies to be as flexible as possible to user variance. In order to make the systems easier to develop, we have adopted a largely declarative representation where the bulk of the domain specific information is provided in external files.

## Project Members:

Wayne Ward

Dialog Management & Natural Language Parsing

Bryan Pellom

Speech Recognition, Synthesis & Natural Language Generation

Kadri Hacioglu

Language Modeling & Speech Recognition

Sameer Pradhan

System Integration, Dialog Grammar Development

## The Communicator Task and Research Approach:

In April 1999, the University of Colorado speech group began development of the CU Communicator system, a Hub-compliant implementation of the DARPA Communicator task. The system combines continuous speech recognition, natural language understanding and flexible dialog control to enable natural conversational interaction by telephone callers to access information about airline flights, hotels and rental cars. In June 1999, the system was fully functional, and was demonstrated at the June 1999 DARPA Communicator meeting. This system connects live over the web to get

real up-to-date air travel, hotel and rental car information. Users call the system on the telephone and make travel plans for a modeled set of cities. We are using the system as a testbed for conversational system technology development. For this system, we developed a new Dialog Manager using an event driven strategy. Our natural language and dialog specification for the task is very much declarative.

Our approach to robustness in natural language understanding is to use robust parsing strategies and event driven dialog strategies. These strategies stress semantic content and coherence over syntactic form. We have extended the Phoenix system as the basis for our robust parsing. This system uses a hierarchical frame parse representation. Our "event driven" dialog manager does not use an explicit script. The developer creates a set of hierarchical forms, representing the information that the system and user interact about. Prompts are associated with fields in the forms. The dialog manager decides what to do next from the current system context, not from a script. Our natural language and dialog specification for the task is very much declarative.

The CU Communicator uses a DARPA Hub compliant architecture, which means that the system is composed of a number of servers that interact with each other through the DARPA Hub. Our system is composed of the Hub and eight servers:

#### Audio Server

receives signals from telephone and sends to recognizer. Sends synthesized speech to telephone

#### Speech Recognizer

Takes signals from audio server and produces a word lattice

#### Semantic Parser

Takes word lattice from recognizer and produces the best interpretation

#### Dialogue Manager

Resolves ambiguities in the interpretation, estimates confidence in the extracted information, clarifies with user if required, integrates with current dialog context, builds database queries (SQL), sends data to NL generation for presentation to user, prompts user for information

#### Database / Backend

Receives SQL queries from Dialog Manager, interfaces to SQL database and returns data, Retrieves live data from the web.

#### Language Generator

Receives speech act and data from the Dialog Manager and generates text to be spoken by the Text-to-Speech Synthesizer

#### Speech Synthesizer

Receives word strings from NL generation, synthesizes them to be sent to the audio server.

#### User Profiler

Allows users to enter preferences profile which is used by the Dialog Manager to supply default values for airlines, hotels, cars, etc.

Servers do not interact directly with each other, but pass frames to the hub. The hub acts as a router which send the frames to the receiving server. A simple script controls hub actions.

Our approach is to conduct research in the context of working systems. We build systems for specific application and evaluate new algorithms in the live system. Within the general paradigm this project will involve three specific research issues:

- A flexible integrated architecture - Currently, most speech systems combine knowledge sources in rigid and inconsistent ways. Acoustic scores and language model scores are combined by fixed-weight interpolation. A recognizer forms a set or graph of hypotheses which is passed to the natural language system. This system then uses an ad-hoc combination of the score from the recognizer and the parse result to find the best hypothesis. This hypothesis is then passed to a dialogue manager. Each module is optimized individually and scores are combined in a fixed way, which has been set to optimize performance on a development test set. The combination of sources is fixed and unable to respond to the reliability of the knowledge sources in a specific situation. Pruning decisions are made at earlier stages without the benefits of all of the knowledge sources. We propose to use decision trees to dynamically combine knowledge sources to assign scores to hypotheses. These scores would reflect knowledge from all sources concurrently, rather than be applied serially (which generates sub-optimal pruning).

- Robust Natural Language Understanding - Our approach to robustness in natural language understanding is to use robust parsing strategies and event driven dialogue strategies. These strategies stress semantic content over syntactic form. We are extending the Phoenix system as the basis for our robust parsing. This system uses a hierarchical frame parse representation. Our "event driven" dialogue manager does not use an explicit script. The developer creates a set of hierarchical forms, representing the information that the system and user interact about. Prompts are associated with fields in the forms. The dialogue manager decides what to do next from the current system context, not from a script. This architecture is similar in spirit to a production system.

- Authoring and portability - Building a spoken dialogue system is labor intensive and requires a great deal of expertise. A major focus of our research is to develop tools to facilitate rapid development of spoken dialogue applications by developers who are not speech and NL experts. Our initial steps will be: a) Declarative representation - To the extent possible, we are representing natural language and dialogue information in external files (like grammars) rather than having the developer write C code. b) Libraries - We are creating a library of common grammars that would be applicable to many tasks (like times, dates, numbers). Developers can use these grammars rather than writing their own. We are also developing a library of functions which will map these to a canonical form (like date to mm-dd-yy and time to 24 hour, etc). These libraries will be distributed with the natural language system.

#### Related Publications:

- Kadri Hacioglu, Wayne Ward , "Figure of Merit for the Analysis of Spoken Dialogg Systems" ICSLP'2002, Denver-Colorado, September 2002. Full Article

- Sameer Pradhan, Wayne Ward, " Estimating Semantic Confidence for Spoken Dialogue Systems" ICASSP'2002, Orlando-Florida, May 2002.

- Kadri Hacioglu, Wayne Ward , " Does Confidence Annotation Meet the Dialog Goal? A Quantitative Analysis" HLT'2002, San Diego-California, March 2002.

- Kadri Hacioglu, Wayne Ward, " A Concept Graph Based Confidence Measure" to appear in ICASSP'2002, Orlando-Florida, May 2002.
- Jianping Zhang, Wayne Ward and Bryan Pellom, " Phone Based Voice Activity Detection Using Online Bayesian Adaptation with Conjugate Normal Distributions," to appear in ICASSP'2002, Orlando Florida, May 2002.
- Kadri Hacioglu and Wayne Ward, " A Word Graph Interface for a Flexible Concept Based Speech Understanding Framework", Eurospeech 2001, Aalborg Denmark, Sept. 2001.
- Jianping Zhang, Wayne Ward, Bryan Pellom, Xiuyang Yu, Kadri Hacioglu, " Improvements in Audio Processing and Language Modeling in the CU Communicator", Eurospeech 2001, Aalborg Denmark, Sept. 2001.
- K. Hacioglu, W. Ward, " Combining Language Models: Oracle Approach", Human Language Technology Conference (HLT-2001) San Diego, March 2001.
- B. Pellom, W. Ward, J. Hansen, K. Hacioglu, J. Zhang, X. Yu, S. Pradhan, " University of Colorado Dialog Systems for Travel and Navigation", Human Language Technology Conference (HLT-2001) San Diego, March 2001.
- K. Hacioglu, W. Ward, " Dialog-Context Dependent Language Modeling Using N-Grams and Stochastic Context-Free Grammars", to appear in IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-2001), Salt Lake City, May 2001.
- R. San-Segundo, B. Pellom, K. Hacioglu, W. Ward, J.M. Pardo, " Confidence Measures for Dialogue Systems," to appear in IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-2001), Salt Lake City, May 2001.
- B. Pellom, W. Ward, S. Pradhan, " The CU Communicator: An Architecture for Dialogue Systems", International Conference on Spoken Language Processing (ICSLP), Beijing China, November 2000.
- W. Ward, B. Pellom , K. Hacioglu, " The CU Communicator: Status and NIST Data Collection; Knowledge Integration", DARPA Workshop, Communicator Principal Investigators Meeting, Philadelphia, PA, September 11-13, 2000.
- R. San-Segundo, B. Pellom, W. Ward, J. M. Pardo, " Confidence Measures for Dialogue Management in the CU Communicator System," IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-2000), Istanbul Turkey, June 2000.
- W. Ward, B. Pellom, " The CU Communicator System," IEEE Workshop on Automatic Speech Recognition and Understanding, Keystone Colorado, December, 1999.