Meteo: A telephone-based Portuguese Conversation System in Weather Domain

Pedro Cardoso, Luis Flores, Thibault Langlois, João Neto

L^2F - Spoken Language Systems Laboratory, INESC ID Lisboa / IST
R. Alves Redol, 9, 1000-029 Lisboa, Portugal
http://l2f.inesc-id.pt

Abstract. Dialog systems using speech technology are in expansion all over the world being used for different domains. The main reason for this growing, apart from the motivation on speech and human-machine interaction research, is the potential for mass access to information. In this paper we describe the work done for the development of a conversation system in a weather domain applied to the European Portuguese language.

1 Introduction

Dialog systems are the result of a combination of several technologies, as speech recognition and speech synthesis, accessing databases, natural language processing and dialog management. In the speech processing field, over the past years, several systems for the European Portuguese language have been under development at INESC ID. Speech recognition systems like Audimus [1] and speech synthesis systems like Dixi and Dixi+ [2] . Those systems have been growing in potential and performance every year, being at this point mature enough to be integrated in applications with a generic use.

Meteo, like Jupiter [3] or Mokusei [4], is a weather information system for the European Portuguese language. Through a spoken interface over phone is possible to access meteorological information for the main cities of Portugal for a period of three days starting on the current day. The information includes maximum, minimum and current temperatures, air humidity and sky conditions. This information is collected from different sources on the web updating the internal database.

In the next section we will present the architecture for the implementation of the system and a description of each component. In section 3 we discuss the benefits of such a system and future developments.

2 System Architecture

Our dialog system is based on a dual HUB architecture, with a separation in an Audio subsystem, where are included the speech processing modules, and a dialog subsystem, with the natural language and dialog manager modules. The

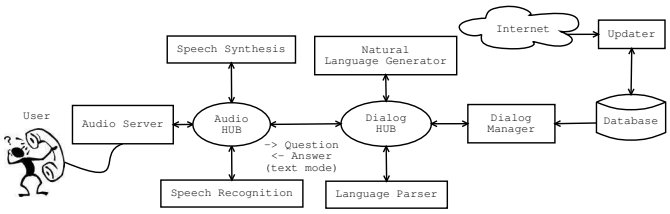

Fig. 1. Global System Architecture.

HUBs communicate on a question-answering mode using a text format protocol for communication. Additionally there is other module called the Updater, responsible for the creation, maintenance and update of the database where the information concerning the application is stored. The architecture of our complete dialog system is presented in figure 1. Below we will give a short description for each of the blocks constituent of the system.

2.1 Updater module

The Updater runs independently of the main system updating the database and maintaining the integrity of the stored data. Based on a defined schedule the Updater is activated and fetch the information from the web regarding weather conditions. For each information source we have a specialized parser that extract the information from the html pages, and fills the database.

2.2 Audio subsystem

The main blocks of the Audio subsystem are the Audio Server, the automatic Speech Recognition (ASR) system, the Text-To-Speech (TTS) system and a Log generator module. In our case the ASR is implemented through the Audimus System and the TTS through the DIXI+ system. In figure 2 we present a block diagram of the Audio subsystem followed by a short functional description of each of the blocks.

The **Audio Server** is a phone interface using a Dialogic board, streaming audio between the phone line and the Audio HUB, with the possibility of converting the audio formats when necessary. There is also an inbuilt Endpoint detector based on signal energy, which filter the silence and noise bursts.

The **Automatic Speech Recognition** block is implemented thought the Audimus system, with some new features as multi-session. Different developments were necessary to adapt the specific modules of Audimus to this task. A speech recognition system needs a language model and an acoustic model adapted to the task domain. Since the system is being developed to work over the phone line a specialized acoustic model was created. Audimus is a hybrid System [1] with the acoustic model implemented thought a Multi-Layer Perceptron (MLP). It was necessary to train a new acoustic model for which we used

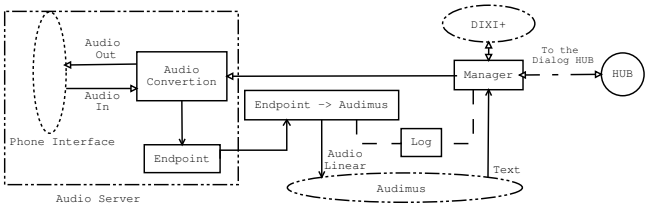

Fig. 2. Detailed architecture for the Audio Hub.

the Portuguese part of the SPEECHDAT database. For the language modelling is usual to extract probabilistic relationships between words from text corpus. In our case we started by defining the vocabulary from some of the Portuguese city names, chosen based on their existence on the information sources, from the different meteorological conditions, and from some specific (like *today*, *tomorrow*, *the day after*) and functional sentences. Based on this vocabulary we extracted a 2-gram language model from newspaper texts. Obviously this source of text is not the most appropriate which we expect to overcome when we have enough logs of the use of the system, to generate new language models as was made for other similar systems [3].

In this work we are using a **Text-To-Speech** system developed in our group in the scope of project DIXI+ [2]. This system uses the Festival Speech framework.

The **Log module** stores for every session the audio corresponding to the different questions made by each user, as well as the recognized text and the generated answer. This information will be used to adapt the acoustic and language models of the speech recogniser with the possibility of an extensive evaluation of the behaviour of the system.

2.3 Dialog Subsystem

The Dialog subsystem is based on three different blocks as represented in figure 1. The Language Parser, witch extracts the request from the text generated by the ASR module, the Dialog Manager, where the interface with the database is made, and the Natural Language Generator, which generates the answer to be outputted to the user.

The **Language Parser** is based on a keyword spotting. For a given question we get the city name to which the user refers, the meteorological conditions requested and the time information. With the data gathered, an information structure is created containing the keywords, and sent to the Dialog HUB to be delivered to the Dialog Manager module.

The **Dialog Manager** works on a frame based system. For every question made by the user (frame) a number of information fields need to be filled. Those fields are the city name, the meteorological condition requested, and the time the question refers to. If any of these fields are not present in the Language Parser output, the Dialog Manager tries to use context information fetching the

missing fields from previous frames. This contextual information creates a more natural and friendly interaction with the user. If some field is missing, even after using the context, several actions can be adopted according to the kind of information missing. After defining the fields a query to the database is made and a new structure containing the elements and respective information are sent to the Natural Language Generator module.

The **Natural Language Generator** get the information from the Dialog Manager and generates an answer in a human readable way. Presently the work done by the Natural Language Generator is a simple transformation from a keyword-information pair to textual information. This text is sent though the two hubs to the TTS system to be transmitted to the user.

3 Conclusion

Our goal in the development of this system was to show that the necessary technologies are mature enough to be used in new applications where a more natural and friendly interface with the user is possible.

However that naturalness is only possible if the system have mechanisms of readjustment based on the real use of the system. For that reason we built mechanisms of logs in our system that currently is public available. The data resulting from that application is being used to improve the ASR components and the Language Parser.

4 Acknowledgments

This work was started as a Project in the final year of graduation of Pedro Cardoso and Luis Flores supervised by Prof. João Neto and Prof. Thibault Langlois at Instituto Superior Técnico, Technical University of Lisbon. This work was partially funded by FCT project POSI/33846/PLP/2000. INESC-ID Lisboa had support from the POSI Program of the "Quadro Comunitário de Apoio III". The authours would like to thank Prof. Luis Oliveira for his help with the DIXI+ system.

References

1. J. Neto, C. Martins and L. Almeida, "A Large Vocabulary Continuous Speech Recognition Hybrid System for the Portuguese Language", In proceedings ICSLP 98, Sydney, Australia, 1998.
2. M. C. Viana, L. C. Oliveira, and I. Trancoso. "Sistema de síntese a partir de texto – DIXI". In Actas do Encontro Regional da Associação Portuguesa de Linguística, Trás-os-Montes, Maio 1997. (Portuguese text)
3. Zue, Victor et al, "JUPITER: A Telephone-Based Conversational Interface for Weather Information", IEEE Trans. on Speech and Audio Processing, vol 8, No. 1, January 2000, pp. 85-96.
4. Mikio Nakano et al, "Mokusei: A telephone-based Japanese Conversational System in the Weather Domain", In proceedings EUROSPEECH 2001, Denmark, 2001.