Dialog system based on EDECAN architecture

J. Doncel, J.M. Olaso, R.Justo, V. Guijarrubia, A. Perez, M.I. Torres

Pattern Recognition and Language Technologies
Dep. Electricity and Electronics. University of Basque Country
manes.torres@ehu.es

***Abstract*—The goal of this paper is to describe a multimodal dialog system based project develoved under EDECÁN architecture for Softec-Ibermática company. The demonstration takes 10 min. aprox.**

***Index Terms*—dialog system; multimodal interaction;**

I. Introduction

The dialogue system that is described in this paper consists of an industrial application in which the user can ask for the following information: on one hand, it can provide meteorological information of the most of cities in Spain, but in the other hand, it can offer the news that apperar in the digital edition of the spanish newspapers. One of the most destacable thing in this project is the multimodal human-system interaction, because it becomes a perfect prototype for disabled people.

The input devices of the system are a keyboard, a mouse, a microphone and a web-cam and the output devices are the screen and the loudspeakers.

In this paper we describe the whole of modules that appear in the dialogue system. This modules have been developed under EDECAN [1, 2] architecture specifications [1]. This architecture was developed with the main goal of communicate any kind of services throught TCP/IP protocols, specially those services needed in the implementation of a dialogue system, under Linux and Windows operating systems. In [1] can be found a detailed description of the architecture.

In the next section is going to be detailed the modules that take part in the dialogue system. Finally, conclusions and open issues to be tackled are proposed.

II. Architecture of the multimodal dialogue system

The dialogue system under consideration is an informatic system whose inputs are speech phrases and whose outputs are synthetic speech phrases. The main goal of the dialogue system is emulation of humans intelligent behavior, under an specific task. Now are described the basic system modules. In Figure 2 are schematically shown the main modules of the system.

- Face recognition module: by the time a person is close to the stand, this module detects the prospective user and starts the character that encourages the user to look for information. It has been implemented using OpenCV [4]. OpenCV is a free library of programming functions mainly aimed at real time computer vision.
- Speech recognition module (ASR): given a speech input signal, this module finds the most probable word sequence. This process consists in the parameterizer and the recognizer. The parametrizer is a preprocess submodule that gets the acoustic signal provided by the user and transform it obtaining an observation sequence, in a space represented by: energy, cepstral coefficients and first and second derivatives of cepstral coefficients. This observation sequence serves as input for the recognizer which transform it in the most likely word sequence, acording to the tasks language model and the acoustic models for the specific language.
  The speech recognition system integrated in this module was completely developed by Pattern Recognition and Language Technologies group at the University of the Basque Country. Its a medium size vocabulary recognizer with Hidden Markov Models to implement acoustic models and stocastic finite state machines to implement language models.
- Understanding module: this module collects the information that the recognition module and extracts the meaning from the input data (word sequence). To realize it, the

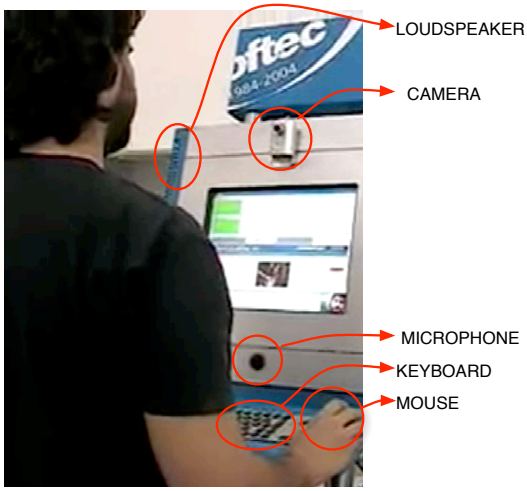

Fig. 1. Final system

[1]Proyect supported by the Ministry of Science and Innovation, 2005- 2008, led by the University of Zaragoza and with participation of Uni- versity of the Basque Country, Polytechnique University of Valencia and Polytechnique University of Madrid. [3]

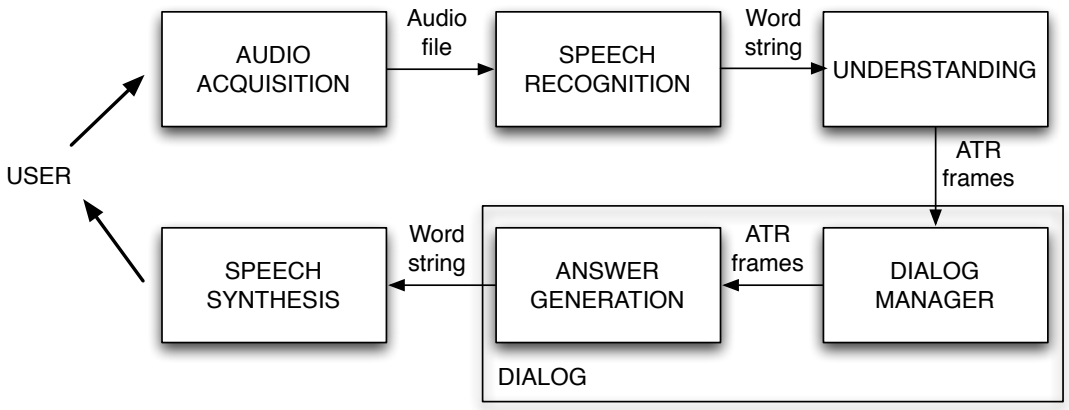

Fig. 2. Multimodal dialogue system: Basic architecture

words are transformed to semantic units that contain the information about the task to be carried out.

- Dialogue manager: this module gives priority to the system and the user in turns. First, it considers the semantic representation of the users request and the history of the dialogue and, then, determines the strategy of the dialogue. The goal is to get from the user the minimum amount of queues to carry out a task. Also is responsible of answers generation and visualization modules activation.
- webclient module: the information the user looks for is obtained from the Internet by the webclient module. When the dialogue manager collects the information needed to know the information the user requires, this module download the data in a xml format.
- xslt module: the information that the webclient module gives to the system is not in an apropiated format, so it has to be converted in order to be able to be shown to the users in the screen. Thats why, this module takes the information from the webclient module and it converts it to html format.
- Visualization module: among the information required by the user, this module has an avatar image and the picture that the web-cam takes at the moment. The avatar helps the users in the interface with the system. The avatar has been created using the *iclone* tool, the free version of *Reallusion*. The picture of the web-cam lets us know the state of the face detection module.
- Text to Speech module: the aim of this module is to generate articially an acoustic signal from a word sequence. This task is carried out by the Loquendo TTS system [5] and the loudspeakers that allow the user listen the output of the system.

III. Conclusion

In the design of the dialogue system described in this paper there is a multimodal interaction between the user and the system. That is because only speech based system working limitation are overcomed.

The EDECÁN architecture makes easier implementing the dialogue system described in this paper.

The main development is thought for the future is implementing this prototype in the Internet. As a result, it is going to be possible to be connected to this system from any computer all over the world.

Acknowledgment

This work has been partially supported by the Counseling of Industry of Basque Government; by the University of the Basque Country under grant GIU07/57; by the Ministry of Science and Innovation under grant TIN2008-06856-C05-01 and by Consolider Ingenio-2010 program MIPRCV (CSD2007-00018).

References

[1] José Enrique Garc´a, Alfonso Ortega, Antonio Miguel, and Eduardo Lleida, Arquitectura distribuida para el desarrollo de sistemas de diálogo hablado, Edecán, in V Jornadas en Tecnologia del Habla, Bilbao, España, Noviembre 2008.

[2] Eduardo Lleida, Encarna Segarra, Mar´a Inés Torres, and Javier Macias, EDECÁN: sistema de Diálogo multidominio con adaptación al contexto acústico y de Aplicación, in IV Jornadas en Tecnologia del Habla, Zaragoza, España, Noviembre 2006.

[3] Proyecto EDECÁN, http://www.edecan.es/,.

[4] Open Source Computer http://sourceforge.net/projects/opencvlibrary/, .

[5] Loquendo. Vocal technology and sevices, http://www.loquendo.com/, .