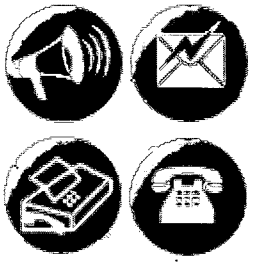


# PETRA: ADVANCED ORAL INTERFACES FOR UNIFIED MESSAGING APPLICATIONS

*David Hernando, Javier Hernando and Xavier Anguera*

*{aldhd,javier,xanguera}@gps.tsc.upc.es*

*Department of Signal Theory and Communications*
*Universitat Politècnica de Catalunya (UPC)*

## ABSTRACT

A new unified messaging system which integrates voice messages, fax and e-mail in a common mailbox has been developed. The combination of speech and linguistic engineering advances allows a natural oral man-machine interaction with the user when accessing the messages by phone. Advanced features are supplied such as texts summarization, messages classification and notification through the phone of new messages received.

## 1. INTRODUCTION

Nowadays, the use of e-mail, voice mail and fax has become indispensable in any working environment. In all cases, a dependency exists between each message type and its associated electronic device. The user must use a computer to access the e-mail, a phone to listen the voice mail and a fax machine to manage the faxes.

We present Petra*, which solves this dependency, providing a common mailbox and two access methods easily available, webmail and telephone.

The whole work included three work lines:

1. Integration of phone, internet and fax services.

2. Development of an advanced dialogue system which brings the user a friendlier access than the provided by a system working only with the DTMF tones.

**The project has been funded by the Spanish Government (CICyT TIC-2000-0335) and is related to the European project Majordome (E!-2340).*

3. Intelligent information management for text classification and summarization.

The convergence of speech processing and natural language processing technologies is a crucial factor in the development of such a system, which requires the concourse of a wide range of knowledge and experience in linguistic engineering.

The paper is organized as follows. Firstly we present the main aspects of the unified messaging platform developed. In section 3, the phone access system to the messages is discussed. Evaluation results are showed in section 4. We finish with some conclusions.

## 2. THE UNIFIED MESSAGING SYSTEM PETRA

The new system presented gives the user two alternatives to access his messages: a webmail and a telephone. Features such as mail filtering and classification, notifications of new mail received, a dialogue driven by a natural man language or the text summarization, added to the usual functionality of a mail agent, convert Petra in an advanced unified messaging system.

The architecture of the demonstrator platform installed at our lab is shown on figure 1. This demo system is accessible by registered users from the public telephone network and from Internet. The external mail accounts from the users can be also accessed from Petra and are treated as extra folders.

The filtering module includes an automatic classifier that, for demo purposes, has been trained with a Spam messages corpus to detect this class of

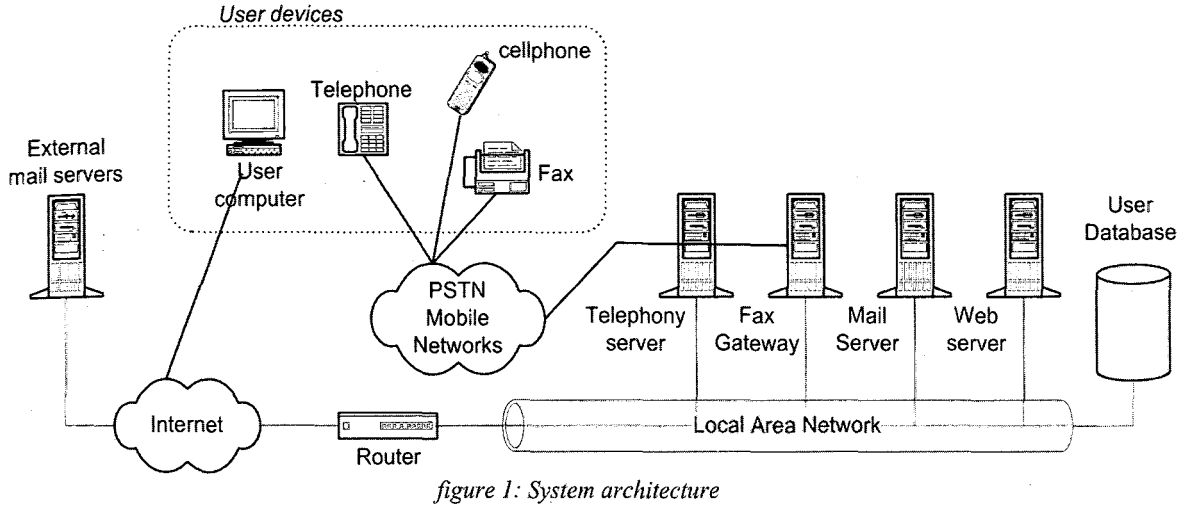



*figure 1: System architecture*

mail. Alerts, when a new message complies the filter's rules, are received by the users via a short message in the cellphone or via a phone call.

## 3. PHONE ACCESS SYSTEM

The phone access system is based on six main modules as seen in figure 2. A CTI board from Intel Dialogic is used as communication interface. Verbio (Verbio. 2004) libraries provide speech recognition and text-to-speech functionality. A speaker verification part has been developed and integrated for security authentication purposes and a new dialogue manager was built. Finally, extern data modules are needed, mainly mail servers and user database.

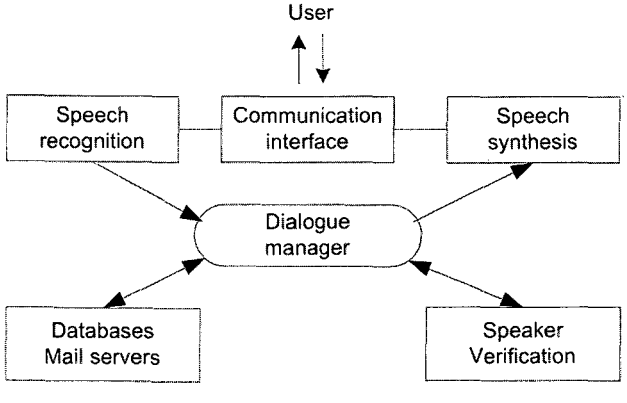



*Figure 2: Phone access modules*

### 3.1 Dialogue Manager

This module offers the user an easy and efficient way to listen and send messages using a natural dialogue with the system. The design and implementation of the new dialogue manager is based on the one used in Attemps (Padrell and Hernando., 2002), a previous project giving real-time meteorological information by phone.

Recognition of user utterances is based on ABNF grammars. In each stage of the dialogue, the group of active grammars define the user's possible answers. Some grammars, which contain user specific data, as his folders or contact addresses, are dynamically generated after the authentication process. Some commands are enabled all the time to offer help, add recognition error correction or to give access to the main menu.

An experienced user is able to interrupt the system to answer a question before it is finished, completing tasks in a minor time (although barge-in can degrade recognition performance in noisy environments).

The confidence level of an user utterance returned by the recognizer is used to decide between two confirmation policies. For levels higher than a threshold, last answer is confirmed implicitly, asking also for next data. If lower, or in critical situations, as in a message deletion, an explicit confirmation policy is used.

Oral response generation is based on sentence selection from patterns, which are combined and completed with particular session information to give the desired meaning to the sentence. In confirming the answers, the system creates sentences with the same form as the user when referring to dates or numbers (for example, "the 13th of May" or "next Thursday"), avoiding confusing the user as in:

U: send the message next Thursday?
S: Do you want to send the message on May 15th?

For most common responses, the dialogue manager answers with one of multiple equivalent sentences, selected randomly each time is required, increasing this way the sensation of a natural dialogue.

Besides the help provided when an user asks for it, the system follows an automatic helping strategy. In a first turn, the manager faces the user with a short and open sentence. If the user answer is wrong, the system asks for a second time with a more explicit question, showing the possibilities and, in further turns, giving the exact words to say or the key tones that user can press instead of his voice.

### 3.2 User authentication

In order to ensure confidentiality of the information, the user needs to sign in using a speaker verification system. The system we present can be trained and configured by the user independently.

Upon dialing the system phone number and receiving the greeting, the user is prompted to say the user login and the password. If the speaker verification system is activated, the user's identity is verified and entrance to the main menu granted or rejected.

As we see in figure 3, the user needs to provide a valid login and password. Which are first assessed using digits recognition and, if the speaker verification system is enabled, independent likelihood values are returned.

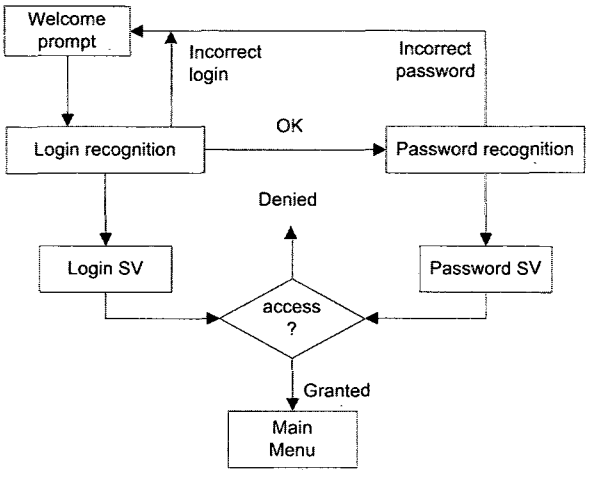



*Figure 3: Speaker authentication system diagram.*

Such likelihoods are compared with the speaker threshold values and the speaker is granted entrance only if both are positive.

The speaker verification system is based on Gaussian Mixture Models (GMM). Both login and password consist of 4 digits. Two different approaches can be used when training the GMM and decoding: a) In independent digits mode the models for digits (0-9) are independently trained. Decoding is done using the corresponding models only. b) In sentence mode both login and password are treated as a whole sentence, a model for each is trained.

The user can train and modify his personal verification models through a voice menu accessible from the main menu. Before using the speaker verification capability he will have to select the kind of verification to use and train the models. Once the system is working he can change some of the system variables by voice though the same menu.

### 3.3. Message adaptation to speech synthesizer

For multimodal e-mail messages received, only text and audio parts are selected to be reproduced. The system, will inform about the existence of other parts, as images or binary documents.

The appropriate speech synthesizer is activated detecting each paragraph language following an

stochastic method. A count of words in Spanish, Catalan and English is done, selecting the language that has more words. The efficiency of the method increases with the number of words.

In order to synthesize correctly the text, some structures as dates, electronic addresses, abbreviations and initials are rewritten. Any format of date is written in an extended version. Electronic addresses are spelled and for abbreviations and initials an equivalence is searched in a dictionary of exceptions.

### 3.4 E-mail summarization

Petra gives, upon request, a message summary provided by Carpanta [reference] module. The goal is to create summaries that are indicative of the number of topics discussed, in contrast with informative summaries, which try to synthesize most of the relevant information. Carpanta does not build new texts, but selects most representative sentences in the text. Its modular architecture permits to difference between language-independent modules, which are the core of the system, and language-dependent modules, which guarantees the portability to other languages different from Spanish.

## 4. EVALUATION

To evaluate the phone access system, the participation of 30 people was requested. Three calls were required for each participant. The first one, common for all people, offered a first contact with the application and its use. For second and third calls, different simulated scenarios were prepared, each one with a goal to be achieved during the call.

Results were obtained from the analysis of 72 registered calls. We took the following measures: 1) Understanding rate, the system understood the user command and acted properly. 2) Wrong user entry, tries to identify dialogue turns where the users had some difficulties. 3) Recognition failure, the system was not capable to understand the user, although his command was right. It could be due to a vocabulary lack or due to a noisy environment.

| *Measures (%)* | *w/o barge-in* | *w barge-in* |
|---|---|---|
| Understanding rate | 76.6 | 77.1 |
| Wrong user entry | 2.9 | 1.7 |
| Recognition failure | 20.4 | 21.2 |

Table 1: - Evaluation results

To see the effects of barge-in when it's used, those turns were measured separately. Table 1 shows these results.

Barge-in did not show a clear degradation of the recognition rates. This is probably due to the users being more confident in their answers when they interrupt the system prompt. We arrived to such conclusion by listening to the recorded calls.

Finally, users were asked for fill on a website a survey measuring user satisfaction. Table 2 shows the results, where punctuation goes from 1 (worst) to 5 (best).

*Table 2: - User satisfaction results*

| *Statements* | *Mark* |
|---|---|
| I understood what the system said | 4,23 |
| The system understood me | 3,33 |
| I knew what I could say to the system | 3,50 |
| The system helped me when I needed | 4,23 |
| The system behaviors as I could expect | 3,61 |
| The dialogue is natural | 3,39 |
| The system voice results friendly | 2,72 |
| The goal of call 2 was easy to achieve | 3,72 |
| The goal of call 3 was easy to achieve | 3,06 |

## 5. CONCLUSION

We have presented the new unified messaging system developed using recent advances in linguistic engineering. E-mail, voice mail and fax messages can be accessed from a common mailbox. The main features of the system have been explained and phone access to the system has been discussed in more detail. The modular design used for implementing the dialogue system will permit to

reuse the new tools developed in different domain applications, getting a minor time of implementation.

The phone access behavior and usability has been evaluated, getting satisfactory results. Useful conclusions were extracted from the campaign wich will permit the improvement of those stages of dialogue where the users had difficulties.

A computer with internet connection and a telephone with hands-free capabilities will be needed for make the presentation in the meeting.

## REFERENCES


Alonso, L., Casas, B., Castellón, I., Climent, S. and Padró, Ll. 2003. Carpanta eats words you don't need from e-mail. XIX Congreso anual de la Sociedad Española para el procesamiento del lenguaje natural.

Padrell J. and Hernando J. 2002. Access to Meteorological Information by Telephone. ICSLP'02. Denver. 2173-2176.

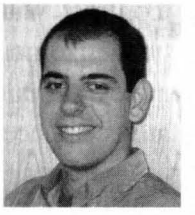
Reynolds, D.A. 2001. Speaker Verification: From Research to Reality. ICASSP Tutorial. Salt Lake City

Verbio from Atlas-CTI. www.atlas-cti.com


## AUTORES

***David Hernando Davalillo***. *Ingeniero de Telecomunicación en febrero de 2004. Su proyecto final de carrera lo realizó en el Grupo de procesado del habla del TSC participando en diferentes tareas del proyecto Petra aqu presentado, principalmente en el sistema de diálogo oral. En la actualidad, trabaja en una Biomeric Technologies, empresa de nuevas tecnologías dedicada a la identificación de personas a través de su voz y otras biometrías.*

***Francisco Javier Hernando Pericás***. *Ingeniero de Telecomunicación en 1988 y Doctor Ingeniero de Telecomunicación en 1993 con Premio Extraordinario de Doctorado por la Universidad Politécnica de Cataluña. Profesor Titular de Universidad del Departamento de Teoría de la Señal y Comunicaciones de la Universidad Politécnica de Cataluña. Sus trabajos de investigación se centran en técnicas de procesado digital de la señal orientadas a la representación robusta de la voz para reconocimiento del habla y del locutor. Es autor de publicaciones en revistas y de comunicaciones en conferencias y es miembro de asociaciones relevantes en el área tanto en el ámbito nacional como internacional. Trabaja en proyectos subvencionados institucionalmente.*



***Xavier Anguera Miró***. *Ingeniero de Telecomunicación en el 2001, máster Europeo en lenguaje y habla en el 2001 y estudiante de doctorado en Telecomunicaciones en la actualidad. Especializado en procesado digital de la señal, especialmente del habla. Ha trabajado en sistemas de síntesis de voz, de diálogo entre hombre y máquina y de verificación de locutor. Ha realizado su actividad profesional mayoritariamente en los EEUU, donde en la actualidad se encuentra en una estancia en el International Computer Science Institute (Berkeley, California).*