New Multimodal Dialogue System For An Academic Intelligent Environment

Ramón López-Cózar [1], Zoraida Callejas [1] and Germán Montoro [2]
[1]*Dept. of Languages and Computer Systems, University of Granada (Spain)*
[2] *Dept. of Computer Science and Telecommunications, University Autónoma of Madrid*
rlopezc@ugr.es, zoraida@ugr.es, German.Montoro@uam.es

Abstract. This paper presents a new multimodal dialogue system that we are developing to assist professors and students in some of their daily activities within an academic centre (e.g. a University's Faculty). The centre is called "intelligent environment" because it features the so-called ambient intelligence. The paper focuses on the system implementation and usage and points out possibilities for future work.

Keywords. Multimodal Dialogue Systems, Ambient Intelligence, Human-Computer Interaction.

1. The DS-UCAT multimodal dialogue system

DS-UCAT (Dialogue System for Ubiquitous Collaborative Training) is a multimodal dialogue system being developed in the UCAT project[1] to assist users (professors and students) in some of their usual activities within several locations in an academic centre (e.g. a University's Faculty). We assume this centre features the so-called "ambient intelligence" in three different locations: Library, Professors' Offices and Classrooms. The work is related to previous projects focused on educational applications, as for example Classroom 2000 [1]. Regardless of this particular application, the system architecture is suitable for a number of application environments, and in particular for smart homes for which the main principles can be generalised.

The system is multimodal since it allows the user interaction via sound, speech, graphics and text [2] [3]. The multimodal input allows to combine several modalities in one interaction. For example, in our current implementation a student can ask for information about available books on a particular subject either by speaking the subject, selecting it on the computer screen using the mouse, or writing the subject in a form field. Since the system output is also multimodal, a spoken message generated for this input indicates the requested information is available on the screen, where it appears as a list of books in text format. We plan to set up system's proactivity so that it could provide users with messages generated from their interaction with the intelligent environment. This way, when a student passes by near the library, the system will remind him/her of borrowed books that must be returned soon to the library.

---

[1] http://orestes.ii.uam.es/ucat/

2. System implementation

Fig. 1 shows the architecture of the system, which is comprised of an X+V document server connected with the users' mobile devices (Tablet PCs, laptop computers and PDAs) through wireless connections. In the current implementation we are only using laptop computers, which connect to the server through the wireless network of our lab.

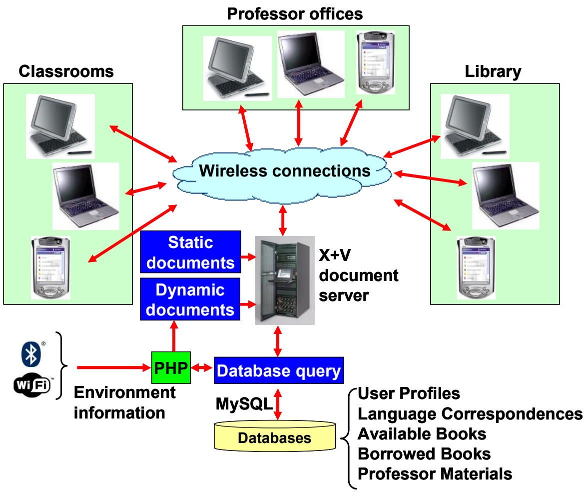

Figure 1. Architecture of the DS-UCAT system

The system is being set up using a toolkit[2] for building multimodal applications based on the XHTML+Voice language[3], also known as X+V. It is configured as a set of X+V documents, some of them are stored in the document server, while others are dynamically created using PHP programs that take into account the user features and preferences (e.g. gender and preferred interaction language), as well as the data extracted from the databases. X+V documents are comprised of forms with fields that are filled in by the user via speech, text or mouse clicks. To visualise these documents and interact with them orally, users run a multimodal browser (e.g. Opera[4]) in their communication devices, which supports multimodal interaction (voice, text and graphics).

The user voice is analysed by the browser's built-in speech recogniser, which uses JSGF grammars (Java Speech Grammar Format) either at form or field levels. Some of these grammars are static while others are dynamically created by PHP programs that query databases and include the obtained data as grammar vocabulary. The system voice is generated by the browser's built-in speech synthesiser, which transforms sentences initially created in text format into speech. Some of these sentences are fixed, while others are created at run-time considering the user type (professor or student), the user gender (necessary to create some sentences appropriately) as well as the data extracted from the databases.

---

[2] http://www-306.ibm.com/software/pervasive/multimodal/
[3] http://www.w3.org/TR/xhtml+voice/
[4] http://www.opoera.com

The visual interaction is used to obtain data from the user via form fields and selection buttons typically used in XHTML. In the system output, the visual interaction is used to provide the data extracted from the database (e.g. list of available books) and information about the user's profile.

In order to provide information to the users and interact with them properly, the system queries several databases. The User Profiles database contains personal data of the system users such as name, gender, address and telephone number. It also stores four types of personal preferences: i) interaction language (English or Spanish at the moment), ii) oral interaction (enabled/disabled), iii) system voice type (male or female), and iv) acceptation of incoming messages from the intelligent environment (enabled/disabled). The Language Correspondences database stores expressions in several languages corresponding to particular sentence types which are used depending on the selected interaction language. An example of these expressions is the welcome message generated by the system as the user logs in: either "Welcome to the DS-UCAT system" for interaction in English, or "Bienvenido al sistema DS-UCAT" for interaction in Spanish. Additionally, in our current configuration the system uses two other databases for experimental purposes, which in a real application of the system should be replaced by the real ones. On the one hand, the Available Books database stores information of books supposedly available in the Faculty's library. Since the recognition grammars used to handle book queries must be updated as the library catalogue changes, these grammars are compiled dynamically from the contents of this database. On the other hand, the Borrowed Books database stores data about books borrowed by users of the intelligent environment (professors and students). Finally, the Professor Materials is a database to be created in order to contain information about the class materials made available by the professors.

3. System usage

To interact with the system the user must firstly log in. Using the login the system determines the user type (professor or student) by querying the User Profiles database. Finding out the user type allows adapting the interaction adequately in terms of interaction language, oral interaction, system voice and acceptation of incoming messages from the environment, as discussed in the previous section. For instance, if a user named "Michael" selected "English" for the spoken interaction, when he logs in a welcome message is shown on the PDA screen together with the spoken message "Hello Michael, welcome to the DS-UCAT system". After the user has logged he must choose a work environment (Library, Classroom or Professor Office). This initial selection allows him to interact in an environment which is not the one in which he is at the moment, thus enabling for instance to make book queries from a classroom. If the Library environment is selected, the message "Please enter a book query" is generated orally. To enter a spoken book query, the user must click and hold a PDA key, or click on the browser's microphone icon while speaking to the system. Alternatively, he can enter the required data using the keyboard and the mouse.

Our goal is that the user-system interaction can be carried out in such a way that the location in which the user is interacting at every moment (e.g. Library) can be taken into account without the user being concerned. However, as the user localisation within the educational space is not yet implemented, at the time of writing we simulate this information by fixing manually a variable that indicates the current user location within the intelligent environment (i.e. either Library, Classroom or Professor Office). This variable's value is taken into account when some X+V documents are dynamically created using PHP programs.

4. Conclusions and future work

This paper has presented our current work within the UCAT project concerned with developing a multimodal dialogue system to assist professors and students in some of their daily activities within an academic centre. We consider this centre features ambient intelligence in three different locations: Library, Professors' Offices and Classrooms. The paper has focused on the system implementation and usage. Future work includes the following lines:

- Create the X+V documents and databases necessary to interact within the Classroom and Professor Office work environments, not yet addressed.
- Set up a procedure to automatically localise the user within the intelligent environment. To implement this feature we plan to experiment with several alternatives: Bluetooth emitters, WI-FI access points and RFID (Radio Frequency ID).
- Enable the system proactivity to provide the user with incoming messages generated from the environment.
- Implement the system ability to operate devices of the intelligent environment in order to change their status. This will allow that in a professor office the system can turn on/off lights or ambient music as the professor enters/leaves the office. To set up this feature, we plan to adapt a previously-created middleware developed within the Interact project [4], which was used to set up a dialogue interface for a home environment.

Acknowledgments

This work has been sponsored by the project TIN2004-03140 Ubiquitous Collaborative Adaptive Training, funded by the Spanish Ministry of Science and Education.

References

[1] G. D. Abowd, C. Atkeson, A. Feinstein, Y. Goolamabbas, C. Hmelo, S. Register, N. Sawhney, M. Tani. Classroom 2000: Enhancing Classroom Interaction And Review. Georgia Institute Of Technology, Technical Report GIT-GVU-96-21, 1996.

[2] M. F. McTear. Spoken dialogue technology. Toward the conversational user interface. Springer, 2004.

[3] R. López-Cózar, M. Araki. Spoken, Multilingual And Multimodal Dialogue Systems: Development And Assessment. John Wiley & Sons Publishers, 2005.

[4] G. Montoro, P. A. Haya, X. Alamán. Context Adaptive Interaction With An Automatically Created Spoken Interface For Intelligent Environments. International Conference On Intelligence In Communication Systems, Bangkok, Thailand. November, 2004.