
A Multi-Modal Interface for Man Machine Interaction with Knowledge Based Systems - MMI²

EVERT KUIJPERS AND MICHAEL WILSON***

** CRISS, Université Pierre Mendès France, Grenoble 2 Sciences Sociales
BP 47X, 38040 Grenoble Cedex, FRANCE
e-mail: evert@criss.fr*

*** SERC Rutherford Appleton Laboratory, Chilton, Didcot, Oxon, OX11 0QX, UK
e-mail: mdw@inf.rl.ac.uk*

Abstract: The MMI² interface demonstrator supports user interaction with a KBS for designing computer networks using Natural Language through keyboard (English, French and Spanish); Command Language; Graphics with Direct Manipulation; and Mouse gesture. It contains an advanced dialogue manager which controls the interaction process, making decisions as to user/system initiative, appropriate response mode, and context driven interpretation. User modelling facilities provide an appropriate range of interactions relative to the user's level of competence. The result is a cooperative, multi-modal dialogue between system and user.

The demonstration shows the integrated MMI² system supporting the graphical design of a local area network and subsequent interaction with the underlying expert system. It includes examples of interaction in each of the different modes of communication as well as examples of mixed-modal interactions, and also makes use of the advanced dialogue capabilities to allow the system to behave cooperatively in its responses.

1. Architecture of the MMI² System

The MMI² system is a demonstrator produced after three years of a five year project. The project aims to demonstrate a multimodal interface capable of supporting co-operative dialogue between the user and system. It makes a strong commitment to "multimodal" rather than to "multimedia" interaction in the interface. The distinction intended is that a multi-media system is one which uses different presentation media (e.g. text, raster graphics, video, speech) without a commitment to the underlying representation of the information presented. For reasons of efficiency of both storage and processing, individual specialised representations are usually used for information which is intended to be presented in each individual medium, and information can only be presented in a single medium. A "multimodal" system is one which includes several input and output media, but is committed to a single internal representation language for the information. This permits the same information to be presented in any mode, chosen purely by rules which select that mode for that user at that point in task performance as being both sufficiently expressive and most efficient

The architecture of the MMI² system can be described as the three layers of Seehiem model for UIMS design [1][2]. The top layer contains the input and presentation modes, the middle layer is the dialogue management layer, and the bottom layer is the application knowledge based system. The demonstrator task performed by the application is to design computer networks.

This document is part of a project partially funded by the Commission of the European Communities ESPRIT programme, as project number 2474. The partners in the project are BIM (Belgium), Intelligent Software Solutions S.A. (Spain), ADR Centre de Recherche en Informatique appliquée aux Sciences Sociales (France), Ecole des Mines de Saint-Etienne (France), Institut National de Recherche en Informatique et en Automatique (France), SERC Rutherford Appleton Laboratory (UK), University of Leeds (UK).

The architecture is based around the notion of the “expert module”. These are not “co-operating experts” or “multiple agents” but simply modules performing specific tasks with their own private data structures, and which allow a sufficiently coherent set of processes to be gathered in a single module. While such a notion is clearly not new, the architecture identifies the nature of the basic modules constituting the multimodal interface, and of the interactions between them.

All operations within the dialogue management layer are performed on a common meaning representation which is independent of both the application and any specific mode (CMR). To this end, all input from the modes to the dialogue management is cast in CMR, as is any output to the modes[3]. The dialogue manager contains a module which maps CMR onto the application language. The CMR represents the mode of an interaction, the type of the utterance (interrogative, imperative, assertion) and the content in a formula which is expressed in first order logic with relativised quantification, some second order symbols, and uses an ontology in which actions and events are reified. The resulting architecture is illustrated in Figure 1 below.

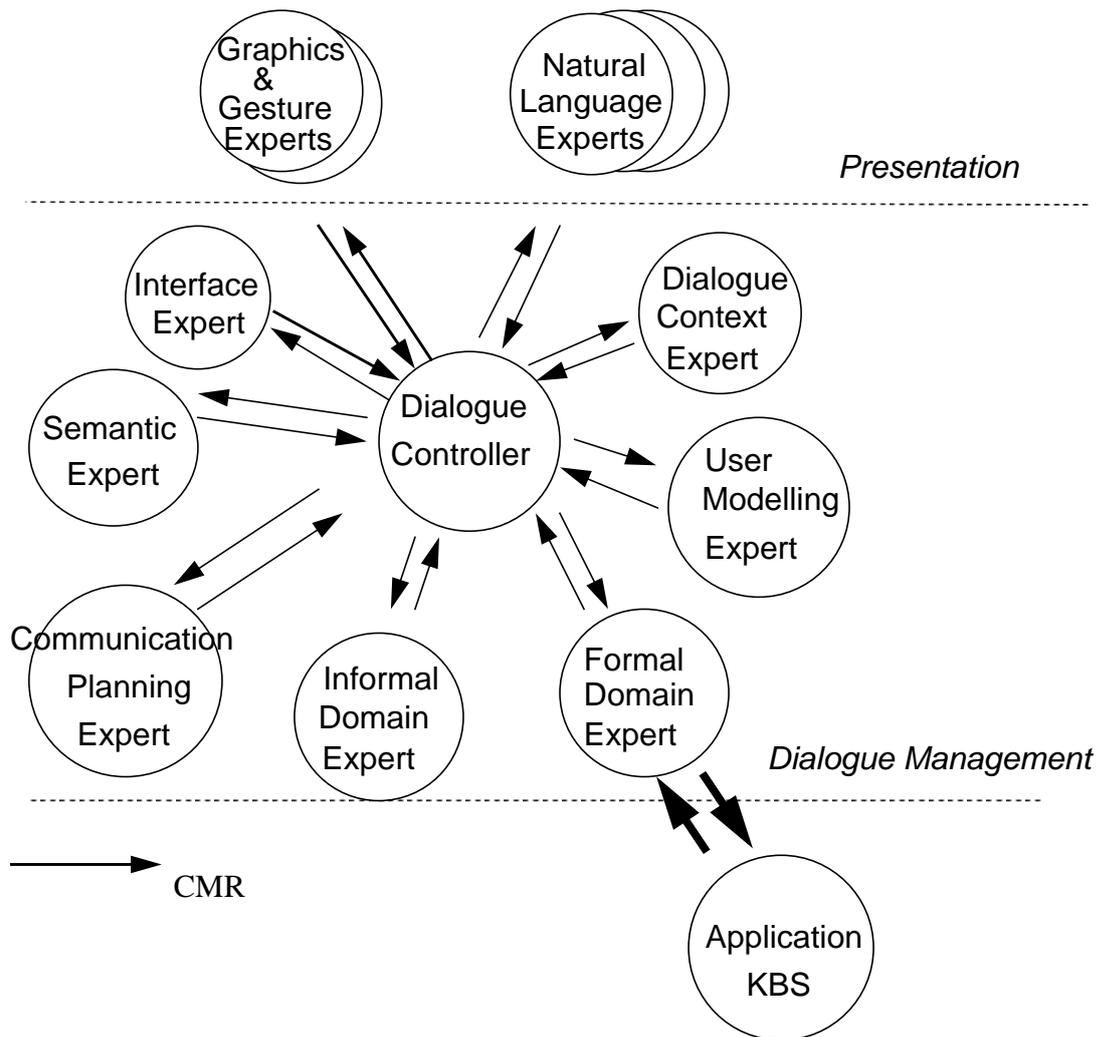


Figure 1: Architecture for the first MMI² demonstrator

The major functionalities of each of the expert modules can be described as follows:

The **dialogue controller** deals with: the major control loop to manage the structure of the dialogue; activating whatever experts are necessary to support dialogue function; providing a theorem prover for CMR formulae; simple existential presupposition checking on CMR formulae [4].

The **dialogue context expert** manages everything which has to do with recording the dialogue structure and extracting relevant information from it; anaphora and deixis resolution.

The **user modelling expert** extracts user beliefs and preferences from CMR formulae; stores the

user model and provides access to the user model for other modules [5].

The **informal domain expert** is responsible for storing plans for the domain task, evaluating informal aspects of the dialogue and representing metaknowledge of the domain[4].

The **communication planning expert** is responsible for constructing the logical form of system output and selecting the output mode.

The **formal domain expert** provides the domain world model against which formal evaluation of CMR takes place by accessing the application KBS.

The **interface expert** is responsible for operations in the domain of the interface and contains metaknowledge of the interface to answer questions about it.

The **semantic expert** has all the knowledge about the intentional definitions of predicates used in the CMR.

In the **mode layer** the various mode experts are responsible for input and output of each mode to/from CMR: English (EM) [6], French (FM) [7], Spanish (SM) [8], Command Language (CL), Gesture (GS) & Graphics (GR) [9]. All natural language examples used in this paper are given in English and are interpreted by the English mode, however they could equally be expressed in French or Spanish to the demonstrator.

In the **application layer** the NEST application designs computer networks [10].

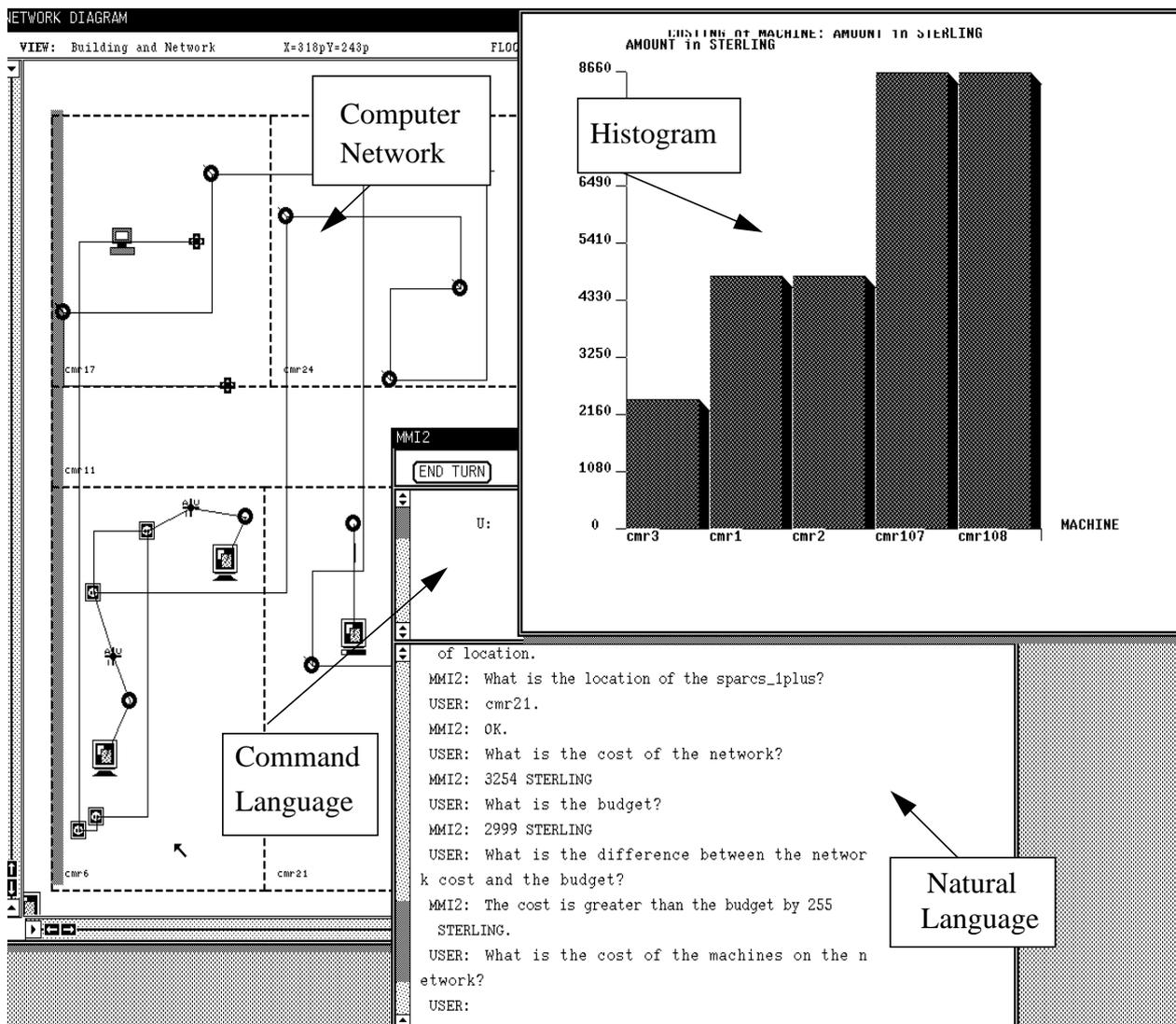


Figure 2: A typical screen display of MMI² in use.

2. Multimodal Interaction

Many interactions can be expressed equivalently (from the point of view of meaning, if not ease of expression) in several modes, as the following example of deleting a machine illustrates:

User (EM):	Suppress the Sun3 connected to the server Ella
User (GR):	<select an icon and select delete option from menu>
User (GS):	<draw squiggly line over icon>
User(CL):	delete Machine23

Although expressed differently, all these inputs must have the same effect on the application (knowledge based system) and on the dialogue context. Thus one of the results of MMI² is to establish a taxonomy of actions across modes (e.g. the verb “surpress”, a DELETE option in a menu, a DELETE command or a gesture of crossing out something with the mouse all refer to the “deleting” operation). But integrating the representation of interactions across modes can go further than that.

In addition to the ability to express an intention through commands and queries in different modes it is also possible to express intentions through combinations of modes which are united through anaphora resolution, (in such mixed mode interactions, the graphical selection can be made while typing) e.g.:

User (GR):	<selects computer that has a disk>
User (EM):	what does the disk cost?
System (EM):	2345 STERLING

Further multimodal interaction is available through the use of spatial terms in natural language which have their semantics grounded in the graphical interface; e.g:

User (EM):	What is the type of the box on the left of the server?
System (EM):	RETIX_2265

Not only the user, but also the system can express output using a combination of modes. For example, in answer to a user query about the location of a machine, the answer would not only be expressed in text by giving the name of the room where it is, but also graphically by highlighting the boarder of the room.

When using a single mode the system will select the most appropriate one in which to express information to the user. For example, in reply to the user query (1) asked in natural language, the system will produce a table (or graph) containing the information in the reply (2) rather than presenting it in the uncooperative natural language form shown here (see figure 2).

(1) What do the computers on the network cost?

(2) Machine1 costs 5113; Machine2 costs 9208; Machine3 costs 5113; Machine4 costs 30625; Machine5 costs 9208; Machine6 costs 30625; Machine7 costs 43750; Machine8 costs 5113.

The MMI² demonstrator is limited as a multimedia system since it does not incorporate non-digited information (sound, static or moving images), but it was not intended as a multimedia system. However, it is also limited as a multimodal system since although it incorporates very limited non-speech audio output it does not include speech input or output, and the style of gestures which it includes are those which can be produced with a pen or a mouse, and not face or body movements which would require image analysis to interpret them. Given these limitations in the range of modalities accommodated, the demonstrator treats interactions in all modes as meaningful sentences in the Common Meaning Representation language and illustrates the synergic interaction of these [11]. It should thereby provide a core to which other modes can be added as long as they can produce and understand the CMR language.

3. Conclusion

The MMI² demonstrator provides a tool which is being used to investigate possible interactions between users and knowledge bases through multimodal interfaces. It is being used to develop requirements for such intelligent systems which can provide efficient and cooperative interaction. The video of the system includes many further examples (from [12]) of both cooperative and multimodal interactions

which are possible with the interface. These use the informal evaluation of user commands against stored task plans, the use of subdialogues to resolve ambiguities and meet unmet preconditions of commands; the structuring of system responses through argument theory; the use of singular and plural anaphora resolution within and between modes; the use of quantification in natural language commands for generality, the use of help dialogues within task execution and other features which conform to the behaviour of users observed in Wizard of Oz studies prior to the development of the demonstrator.

The project is continuing to advance the demonstrator as new requirements are identified, mainly by improving the dialogue features to cover ellipsis, causal explanation of the KBS's reasoning, and robust co-operative responses to the wide range of user input encountered in practice. The interface is also being ported to a new application to test generality. Although there has not been space here to describe the current dialogue features, they include context sensitivity, embedded subdialogues and the ability to provide simple explanations and tutorial replies. Similarly, the details of the complete system cannot be presented here, but can be obtained from the references cited, if the facilities presented are found to be interesting.

References.

- [1] G.E. Pfaff, "User Interface Management Systems", Eurographic Seminar Series, Springer-Verlag: Berlin, 1985.
- [2] D.A. Duce, M.R. Gomes, F.R.A. Hopgood, and J.R. Lee, "User Interface Management and Design", Eurographic Seminar Series, Springer-Verlag: Berlin, 1991.
- [3] G.J. Doe, G.A. Ringland, and M.D. Wilson, "A Meaning Representation Language for Co-operative Dialogue". Proceedings of the ERCIM Workshop on Experimental and Theoretical Studies in Knowledge Representation, pg 33-40, Pisa, Italy, May 1992.
- [4] D. Sedlock, G. Doe, M.D. Wilson, and D. Trotzig, "Formal and Informal interpretation for Co-operative Dialogue". Applied Artificial Intelligence, in press.
- [5] H. Chappel, M.D. Wilson and B. Cahour, "Engineering User Models to Enhance Multi-Modal Dialogue". Proceedings of the IFIP Workshop on the Engineering of HCI, Finland, Aug., 1992.
- [6] J-L. Binot, B. Demoen, K.H. Hanne, L. Solomon, Y. Vassilou, W. von Hahn, and T. Wachtel, "LOKI: A Logic Oriented approach to data and Knowledge bases supporting natural language Interaction", In Proceedings of the Esprit '88 Conference, North-Holland: NL, 1988.
- [7] S. Dami, et G. Lallich-Boidin, "An expert-system for french analysis within a multimodal dialogue to be connected", 23 p., RIAO 91 conference, Barcelona, 1991.
- [8] J.C. Ruiz Anton, "Un parser basado en la sintaxis de dependencias". Procesamiento del Lenguaje Natural, 9, pp. 169-176, 1991.
- [9] H. Ben Amara, B. Peroche, H. Chappel, and M.D. Wilson, "Graphical Interaction in a Multi-Modal Interface". In Proceedings of Esprit '91 Conference, CEC: Brussels, 1991.
- [10] C. Jouve, et F. Balfroid, "Un systeme a base de connaissance pour la conception de reseaux: NEST", Proceedings of Avignon '92, 1992.
- [11] J. Coutaz and J. Caelen "A Taxonomy for Multimedia and Multimodal User Interfaces". Proceedings of the ERCIM Workshop on User Interfaces and Multimedia, pp 143-147, Lisbon, Nov. 1991.
- [12] M.D. Wilson, "The First MMI² Demonstrator: a Multimodal Interface for Man-Machine Interaction with Knowledge Based Systems". Rutherford Appleton Lab. Report, RAL-91-093, pgs 102, 1991.

