The CARTOON project : Towards Integration of Multimodal and Linguistic Analysis for Cartographic Applications

Martin, J.C., Briffault, X., Gonçalves, M.R., Vapillon, J.

LIMSI-CNRS, BP 133, 91403 Orsay Cedex, France, Email: (martin/xavier/goncalve/vap)@limsi.fr

Reference in multimodal input Human-Computer Interaction has already been studied in several experiments involving either simulated or implemented systems (Mignot and Carbonell 96, Huls and Bos 95) including cartographic application (Siroux et al. 95, Cheyer and Julia 95, Oviatt 96). In this paper, we present our project named CARTOON (CARTography and cOOperatioN between modalities): we describe the current prototype but also how we plan to integrate tools providing linguistic analysis.

1. THE CURRENT PROTOTYPE

The current prototype enables to combine speech recognition, mouse pointing and keyboard to interact with a cartographical database (figure 1). Several functions are available such as requesting information on the name or location of a building, the shortest itinerary and the distance between two points, or zooming in and out. Several combinations are possible such as :

- What is the name of this <pointing> building ?
- What is this <pointing> ?
- Where is the police station ?
- Show me the hospital
- I want to go from here <pointing> to the hospital
- I am in front of the police station. How can I go here <pointing> ?
- Show me how to go from here <pointing> to here <pointing>.

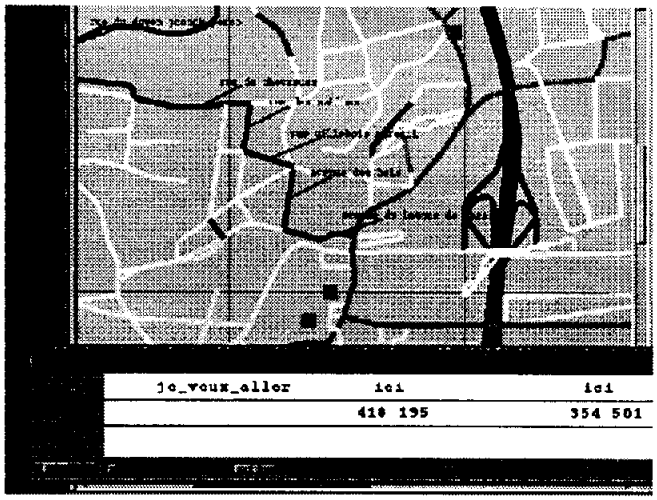

Figure 1: Events detected on the three modalities (speech, mouse, keyboard) are displayed in the lower window as a function of time. The recognized words were: "I want to go", "here", "here". Two mouse clicks were also detected. The system displayed the corresponding itinerary.

Currently, there is no linguistic analysis. Events produced by the speech recognition system (a Vecsys Datavox) are either words or sequences of words ("I_want_to_go"). There are 38 such speech events which are characterized by: the recognized word, the time of utterance and the recognition score. The pointing gestures events are characterized by an (x, y) position and the time of detection. The overall architecture is described in figure 2 : events detected on the keyboard, mouse and speech modalities (left-hand side) are time-stamped coherently by a modality server and then integrated in the multimodal which merges them and activates the application.

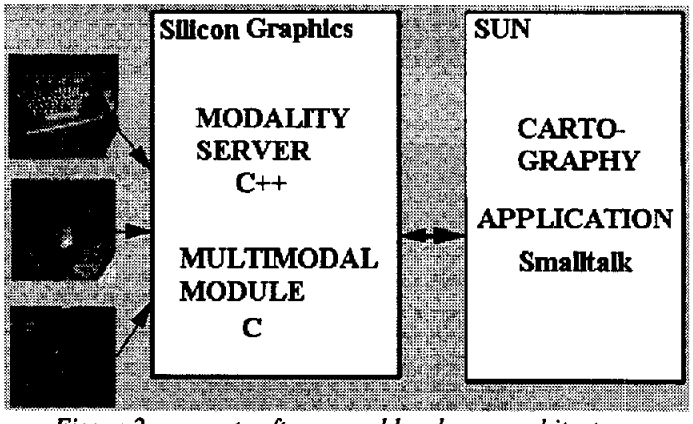

Figure 2: current software and hardware architecture.

The multimodal interface is based on a theoretical framework of « types of cooperation between modalities » that we initially presented in (Martin and Béroule 93) and that has been used by other French researchers in (Catinis and Caelen 95, Coutaz and Nigay 94). Our framework proposes six basic types of cooperation between modalities (either input or output):

- transfer : the result of one modality is used by another modality
- specialization : a modality is devoted to the transmission of the same type of information
- equivalence : choice between several modalities for the transmission of a given piece of information
- redundancy : the same piece of information is transmitted in several modalities
- complementarity : pieces of information regarding the same command are transmitted on several modalities
- concurrency : parallel transmission of independent information in several modalities

The combination of modalities are described in a specification language that is based on the theoretical framework. Three criteria of fusion are available for redundancy and complementarity: temporal coincidence, sequence and structural completion. The multimodal module uses Guided Propagation Networks (Béroule 1985) which provide what we call « multimodal recognition scores » incorporating the score provided by the speech recognizer. In the case of missing events, several commands may be activated with different recognition scores. The command with the highest score is selected by the system and may prompt the user for information if needed. More details on this multimodal framework and module can be found in (Martin et al. 95).

2. Linguistic analysis

The multimodal module of the current prototype does not feature either any syntactic or semantic analysis. In the next prototype, this will be handled by a linguistic software engineering environment developed by the Language and Cognition group of LIMSI. The syntactic analysis, based on a chart parser uses a LFG grammar for French, and the semantic analysis is based on conceptual graphs. A chart is a graph whose nodes are positioned between the words of the sentence to be parsed (figure 3). They contain two types of arcs: active arcs (which represent beginnings of a syntactic structure) and completed arcs (which represent a whole syntactic structure). A great interest of the Chart is that it gives not only a trace of the rules applied while parsing a sentence, but also the possibility of analyzing in details the structures build at various levels which can be useful for interacting with the multimodal module. The current grammar contains about 225 rules of French simple sentences. In addition to these simple sentences, difficult problems are also handled: clitics, complex determiners, completives, various forms of questions, extraction and non limited dependancies, coordinations, comparatives. Some extensions are currently under development: negation, support verbs, circonstant subordinate phrases, ellipses. More details on these linguistic tools can be found in (Vapillon et al. 97).

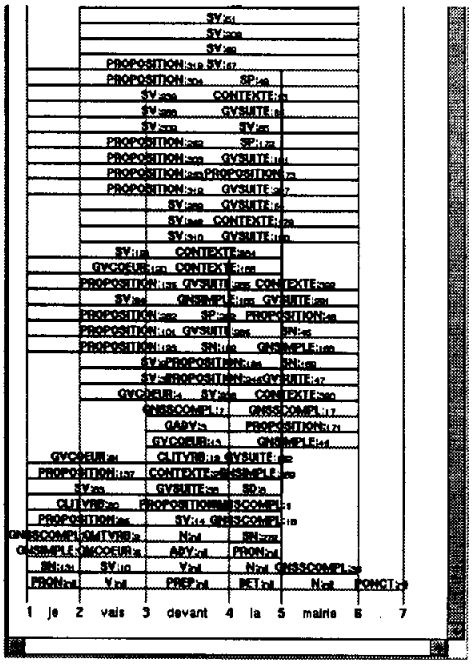

Figure 3: A chart resulting of the analysis of « I am going in front of the Town Hall ».

3. Towards integration of multimodal and linguistic analysis

Working with two specialized modules allows us to take advantages of each of them. The linguistic module provides a detailed linguistic analysis. Both modules may work in parallel but have to exchange results. For instance, the multimodal module may send to the linguistic module partial results (activation of multimodal units representing hypothesis on the types of cooperation between modalities). These pieces of information may allow the linguistic module to drop early (or on the other way confirm some hypothesis). Symmetrically, the results of the linguistic module (parts of charts and conceptual graphs) may be used by the multimodal module as events of a higher level of abstraction. Future directions also include extending gestures to circling and trajectory gestures on a tactile screen, implementing dialog history and studying how the linguistic tools cop with spontaneous spoken language. Finally, experimental studies need to be put into place to evaluate the cooperation between speech and gestures that are used by the subjects when interacting with a map.

Béroule, D. (1985). A model of Adaptative Dynamic Associative Memory for speech processing. Thesis. In French.
Bourdot, P. et al. (1995) Management of non-standard devices for multimodal user interfaces under UNIX/X11. In [CMC].
Catinis, L.& Caelen, J. (1995) Analyzing the multimodal behavior of users in a drawing task. IHM'95. In French.
Cheyer, A. & Julia, L. (1995) Multimodal maps: an agent-based approach. In [CMC 1995].
CMC (1995). Proceedings of the International Conference on Cooperative Multimodal Communication (CMC'95). Bunt, H, Beun, R.J. & Borghuis, T. (Eds.). Eindhoven, may 24-26.
Coutaz, J. & Nigay, L. (1994) The CARE properties in multimodal interfaces. IHM'94. In French.
Huls, C. & Bos, E. (1995) Studies into full integration of language and action. In [CMC 1995].
Martin, J.C. & Béroule, D. (1993) Types and goals of cooperation between modalities. IHM'93. In French.
Martin, J.C. et al. (1995) Towards adequate representation technologies for multimodal interfaces. In [CMC], II : 207-222
Mignot, C. & Carbonell, N. (1996). Oral and gestural command : an empirical study. TSI. Vol 15, no 10, In French.
Oviatt, S. (1996). Multimodal interfaces for dynamic interactive maps. Proceedings of CHI'96, april 13-18. pp 95-102.
Siroux, J. et al. (1995) Modeling and processing of the oral and tactile activities in the Georal tactile system. In [CMC].
Vapillon, J. et al. (1997) An Object-Oriented Linguistic Engineering Environment using LFG and CG. ACL/EACL'97 workshop "Computational Environments for Grammar Development and Linguistic Engineering", 1997, Madrid.