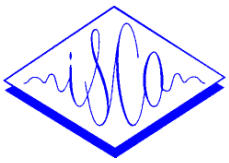

4th European Conference on
Speech Communication and Technology
EUROSPEECH '95
Madrid, Spain, September 18-21, 1995

# SPEECH and TACTILE-BASED GÉORAL SYSTEM

J. SIROUX *, M. GUYOMARD**, Y. JOLLY***
F. MULTON***, C. REMONDEAU***

* IUT-IRISA, Lannion, France, siroux@enssat.fr,
**ENSSAT-IRISA, Lannion, France, guyomard@enssat.fr
***IRISA-LLI, Lannion, France

## ABSTRACT

This paper describes the principal elements of the oral and tactile GÉORAL system, as well as the results of a primary evaluation of the system. Firstly, we present the main dialogue functionalities and the problems due to the multimodal activities. These problems are mainly concerned with the relationships between the linguistic syntagms of oral utterances and the user's activity on the screen (pointing, drawing a zone,...). Then we describe the technical choices we made and the architecture of the system. We decided to give priority to the speech input, and to merge the results of the two activities at the latest possible. Some details both on the data structures (dialogue acts, communicative acts) we use and the processing ways are provided. Finally, we present the first results of an evaluation of the system with unexperimented users. The results show the oral-tactile synergy is well accepted and improve slightly the system understanding.

10.21437/Eurospeech.1995-474

## 1. INTRODUCTION

In 1992, we developed an oral dialogue system which was able to provide naive (un-experienced) users information of a touristic nature, such as beaches, churches, abbeys and châteaux for the Trégor region of Brittany [1]. Assessments showed users may come up against various difficulties using the system. Such difficulties revolve around the recognition and the understanding of user's utterances: lack of training of the speech recognition board but also badly pronounced and structured utterances (spontaneous speech). We think one way of resolving a part of these problems is to provide the system with complementary or redundant input means of communication [2][3][4]. We choose to add a touch-sensitive (tactile) screen to the system. In this situation, new problems and questions come to light: how does the user act in such a context, how and when the joint oral and tactile activities do have to be integrated, what are the benefits that could be expected from this adjonction?

This paper describes how we resolved the integration problem of oral and tactile media for our system and display the first results of an evaluation of the complete system. We shall first present the dialogue functionalities and features. Then, we provide some details about the architecture, the functioning of the system and the main data structures used. Last, we describe the evaluation and provide the first results and conclusions.

## 2. DIALOGUE FUNCTIONALITIES

### 2.1 Main features

GÉORAL tactile is a dialogue system which provides information of a touristic nature, such as beaches, churches, abbeys, châteaux, camp-sites and routes for the Trégor region of Brittany.The user is able to interact with the system in three different modes: in a visual mode by using the map of Trégor on the screen, in an oral mode thanks to a speech recognition board and in a gesture mode using a touch screen. The oral mode is the main mode of interaction, it allows the user to utter queries as well as answers. The gesture mode only allows the user to point out places on the screen. The oral and gesture modes can be used jointly for example to avoid speech recognition problems.The system uses both the oral channel (via a text to speech synthesis board), and graphics, such as the flashing of sites, routes, localities and zooming in on sections of the map, so as to best inform the user. In order to avoid an empty response, co-operative algorithms are used [5] to produce corrective and suggestive responses.

An example of dialogue (excerpt) (U: user, S: system):

U1 Are there any camping sites in Tréguier?
S1 Please rephrase your question.
U2 Are there any campsites in Tréguier?
S2 Please rephrase your question.

U3 Are there any campsites here? *U points to the location of Tréguier while pronouncing 'here'*
S3 I am looking for campsites in the vicinity of Tréguier, please wait. Here are the places which correspond to your request, would you like further information about any of these places? *zooming and flashing of towns*

### 2.2 User joint activities

In order to collect some materials for modeling the user activities, we decided to set-up a simulation [6] in a Wizard of Oz mode [7]. The main results of this experiment are as follows:

• the presence of the tactile screen modifies the linguistic behaviour of the user: some particular deictic terms (around 10 words have to be added to the speech recognition vocabulary) appear and new syntactic structures occur,

• three possible relationships between oral utterances and tactile gestures have been identified (the two main ones follow):

- bound relationship in which one deictic item of the oral utterance and a touch activity are used together to designate an entity on the map:

U: Are there any beaches in this locality ? + a touch on a locality.

- confirmative relationship for which the oral subsection is enough for comprehension but is however accompanied by a tactile designation, which is redundant with the linguistic reference:

U: Are there any beaches in Lannion ? + a touch on Lannion.

• Finally, concerning the nature of the touch gesture (either pointing or drawing of a zone), we highlight a pragmatical phenomenon which is already studied in the verbal dialogue area. A touch gesture, in its referential aspect can only be completely understood if it is considered within the context of the interaction. In other words, the illocutionary gesture can prove to be different from the literal gesture. This manifests itself in two ways according to the nature of the context, either in the complementary oral utterance or in the application.

All these phenomena have to be tackled with specific and flexible methods.

## 3. ARCHITECTURE

### 3.1 Global architecture

The global architecture of the system is provided in figure 1. The speech input is processed by the recognition board MEDIA50 (licenced by France Telecom CNET) which produces the uttered word string. The board is used to its limits (we recognize continuous speech whereas the board is mainly designed to recognize isolated words or chained words). The speech output is produced by a TELEVOX board (also licenced by France Telecom CNET).

### 3.2 Processing the speech and tactile activities

Figure 2 provides details about the modules which deal with the speech and tactile inputs. Two main principles are leading us. The first one is to assign the main priority to the speech input: the tactile activity is processed taking into account the results of the speech recognition. The second one is to memorize the tactile activity only when the user talks (the system asks for the user to talk and the speech recognition board is not in constant use).

The control flow of the system can be described as the following. The syntactic and thematic analysis are triggered off after the speech recognition. The syntactic analysis produces a complete syntactic tree using the difference list method. The deictic and anaphoric syntagms are only spotted in the utterance and coded inside the tree. The thematic analysis has two principal roles with regard to the tactile function. It determines the possible type (style) of tactile touch (point, zone, etc.) as well as the theme of the question (type of object in question). It also produces an intermediary structure so-called a dialogue acts of which the modeling of propositional contents is inspired by [8]. For example the user's utterance : *are there any beaches here ?* will be transformed as :

ASK(U, S, informref(S, U, beach, Q(beach, locality(deicphore(pointing)))))

where *deicphore(pointing)* indicates a user tactile activity to point out the place where the system will have to search for. The transmission of the theme to the tactile processor is accompanied by the relevant objects of the database. The tactile processor receives as an input the list of touches

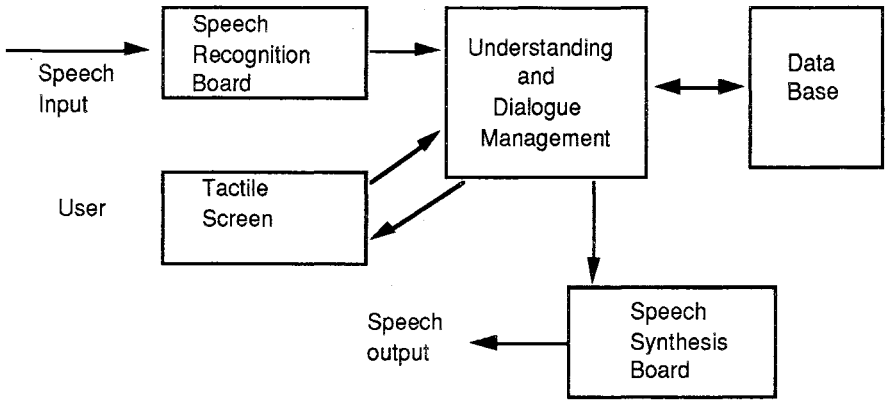


Figure 1. Global architecture of GÉORAL Tactile system

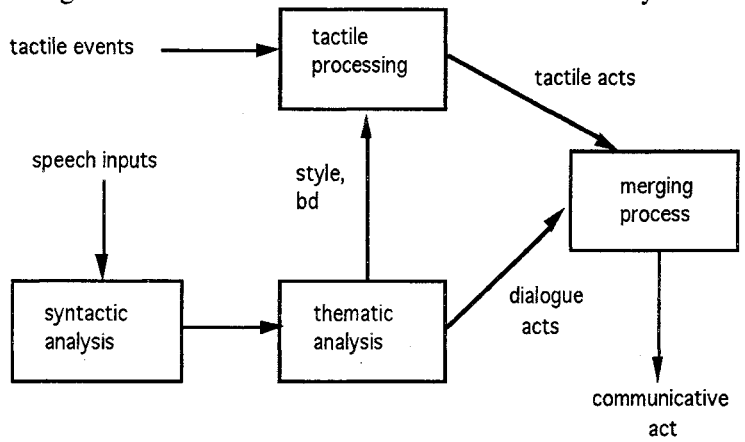


Figure 2. Architecture of the modules which process speech and tactile inputs

performed by the user, the list of objects in the database which must be examined and the style of tactile activity (zone,...) predicted from the speech input. It produces the list (tactile acts) of objects actually designated. This process necessitates the appropriate taking into account of the tactile events (for example, reconstruction of zones), a verification of the coherence between the predictions and the observed facts, as well as possible corrections and adjustments. The merging process concerns the determination of the full communicative acts. It resolves the tactile-linguistic coreferences. It is based on a modeling of communicative act by way of plan operators [9][10][11]. For example, the following model of the communicative act REQUEST allows to merge a tactile event and a dialogue act ASK which contains a deictic item :

NAME: Request_completive mode
HEADER: REQUEST(U, S, INFORMEREF(S,U, ?x, Q(?x, ?P'(?o'))))
BODY: ASK(U,S, informerf(S,U, ?x, Q(?x, ?P(?D))))
Désigner(U, S, ?o)
CONSTRAINT: none
PRECONDITION: Déictique(?P(?D)), ?o, ?P'(?o'))

The predicate Déictique in the precondition part checks the consistency and produces the referent. For example, the dialogue act :
ASK(U, S, informref(S, U, beach, Q(beach, locality(deicphore(pointing)))))
and the tactile act
Désigner(U, S, (Lannion, pointing, 1, X, Y))
will be recognized as compatible and allow to recognize the REQUEST communicative act :

Request(U, S, informref(S, U, beach, Q(beach, locality(Lannion))))

Further details are provided in [12][13].

## 4. SYSTEM EVALUATION

The objective of the evaluation was twofold: to assess the system (GÉORAL Tactile -GT-) and to measure the interest of the tactile mode in comparison to the previous release of the system (GEORAL -G-) without this communication mode. Twelve naive, unfamiliar

to computer use subjects and two familiar to computer use subjects have carried out six scenarii after a brief training with two short scenarii. 3 scenarii have been carried out using GT and 3 using G by each subject. Half the subjects have firstly used G and the second half firstly GT. The main results of a first exploitation of the corpus are as following. The oral-tactile synergy is well accepted, used and allows successful dialogues in spite of speech recognition problems. The use of the tactile screen improves the recognition and understanding of the initial requests (the most difficult to process) by a factor of 8%. The degree of use of the tactile is highly dependant on the subjects: in average 36% of the initial requests are composed of a tactile activity but the higher percentage is around 95% and the lower 2%. Lastly, as the users carried out scenarii, better they mastered the syntax of the request and the dialogue management: in the first scenarii 11% of the request are not well formed, they are only 2% in the last scenarii.

## 6. CONCLUSIONS

We designed and assessed a system which uses two modalities in input: a tactile screen and a speech recognition board. The modeling and processing of these inputs in the system are carried in such a way that they allow a great flexibilty in order to modify priorities between the media or to deal with more complex referential phenomena. An evaluation of the system with unexperienced users shows the benefits of the tactile screen on the speech understanding. But, it also stresses the importance of messages produced by the system in order to better lead the user activity. In the future, we would like to use the tactile activity as a feedback in order to help the speech recognition in case of failures.

## 7. REFERENCES


[1] Gavignet F., Guyomard M., Siroux J. Implementing an oral and geographic multimodal application : the Géoral project. *Pre-proceedings of the Second Venaco Workshop on the Structure of Multimodal Dialogue, NATO*, Acquafredda di Maratea, Italy, September, 16-20, 1991.

[2] Cohen, P. R. The role of Natural Language in a Multimodal Interface. *Proceedings of 2nd FRIEND21, International symposium on next generation human interface technology*, Tokyo, Japan, Nov. 1991.

[3] Wahlster W., André E., Graf W. and Rist T. Designing Illustrated Texts : How Language Production is influenced by Graphics Generation. *Proceedings of EACL 91*}, Berlin, April 1991.

[4] Zancanaro M, Stock O., Strapparava C. Dialog Cohesion Sharing and Adjusting in an Enhanced Multimodal Environment. *Proceedings of IJCAI 1993*, Chambéry, France, 1993, p. 1230-1235.

[5] Guyomard M, Siroux J. Suggestive and corrective answers: a simple mechanism, in *Structures of multimodal dialogs,* Bouwhuis (D.G.), Taylor (M. M.), Néel (F.) (editors), North Holland, 1989.

[6] Guyomard M., Le Meur D., Poignonnec S. and Siroux J. Experimental work for the dual usage of voice and touch screen for a cartographic application. *Proceedings of ESCA Tutorial and Research Workshop on Spoken Dialogue Systems*, may 30-June 2, Vigsø, Denmark, 1995.

[7] Oviat S., Cohen P., Fong M,. Frank M. A rapid semi-automatic simulation technique for investigating interactive speech and Handwriting. *Proceedings of the 7th International Conference on Spoken Language Processing*, Oct. 1992, Banff, Canada, Vol 2, p. 1351-1354.

[8] Allen J. Natural Language Understanding.The Benjamin/Cummings Publishing Company, Inc., 1987.

[9] Litman D. J., Allen J. A Plan Recognition Model for Subdialogue in Conversations.*Cognitive Science* 11, p. 163-200, 1987.

[10] Maybury M.T. Planning Multimedia Explanations Using Communicative Acts. *Proceedings of the Ninth National Conference on Artificial Intelligence, AAAI 91*, Anaheim, CA, July, 14-19, 1991.

[11] MAYBURY M.T. Communicative Acts for Explanation Generation. *International Journal of Man-Machine Studies*, 37(2), 135-172.

[12] Multon F. GÉORAL tactile un système multimodal. Rapport de DEA, IFSIC, université de Rennes 1, 1994.

[13] Siroux J., Guyomard M., Multon F., Rémondeau C. Modeling and processing of the oral and tactile activities in the Géoral tactile system, *Proceedings of CMC95*, Eindhoven, The Netherlands, May 1995.



**Acknowledgement**

This work was partially funded by CNET France Telecom, contract 92 7B.