USER EVALUATION OF THE MASK KIOSK

L. Lamel, S. Bennacef, J.L. Gauvain, H. Dartigues†, J.N. Temem†
LIMSI-CNRS, BP 133, 91403 Orsay cedex, France
†SNCF, Direction de la Recherche et de la Technologie, 45, rue de Londres, 75379 Paris, France

ABSTRACT

In this paper we report on a series of user trials carried out to assess the performance and usability of the MASK prototype kiosk. The aim of the ESPRIT Multimodal Multimedia Service Kiosk (MASK) project was to pave the way for more advanced public service applications with user interfaces employing multimodal, multi-media input and output. The prototype kiosk, was developed after analysis of the technological requirements in the context of users and the tasks they perform in carrying out travel enquiries, in close collaboration with the French Railways (SNCF) and the Ergonomics group at UCL. The time to complete the transaction with the MASK kiosk is reduced by about 30% compared to that required for the standard kiosk, and the success rate is 85% for novices and 94% once familiar with the system. In addition to meeting or exceeding the performance goals set at the project onset in terms of success rate, transaction time, and user satisfaction, the MASK kiosk was judged to be user-friendly and simple to use.

1. INTRODUCTION

The ESPRIT Multimodal Multimedia Service Kiosk MASK project kiosk has developed a prototype kiosk with an innovative, user-friendly interface, combining tactile and vocal input. The propotype kiosk was developed after analysis of the technological requirements in the context of users and the tasks they perform in carrying out travel enquiries. The kiosk has undergone several rounds of user trials, including a series of Wizard of Oz experiments in the early stages of the user interface design, reported at ICSLP'96[5]. The work reported here was carried out by LIMSI-CNRS, the SNCF (the French Railways) and the Ergonomics group at UCL (Univeristy College London).

The physical design of the prototype kiosk has been changed since that reported in [5], and significant improvements have been made to the user interface. The main improvements concern additional features such as a self-presentation illustrating the use of the kiosk and explaining the different types of transactions available; a more intuitive interface with easy switching between tasks (such as information or ticketing); a facial image of a clerk to let the user know what the system is doing (see Figure 2); and a two-level help facility with fixed time-outs.

In this paper we focus on studies of the user assessment of the MASK prototype, using both objective and subjective performance measures. Iterative evaluations were carried out to validate the software integration and user-interface design. A final set of user assessment trials were carried out in May 1998 with 100 subjects at the St. Lazare train station in Paris. An additional set of performance trials involving over 100 subjects compared different interaction modes: tactile only, vocal only, or combined; as well as trials with the same subjects using the MASK kiosk and the standard automated ticket machines located in train stations.

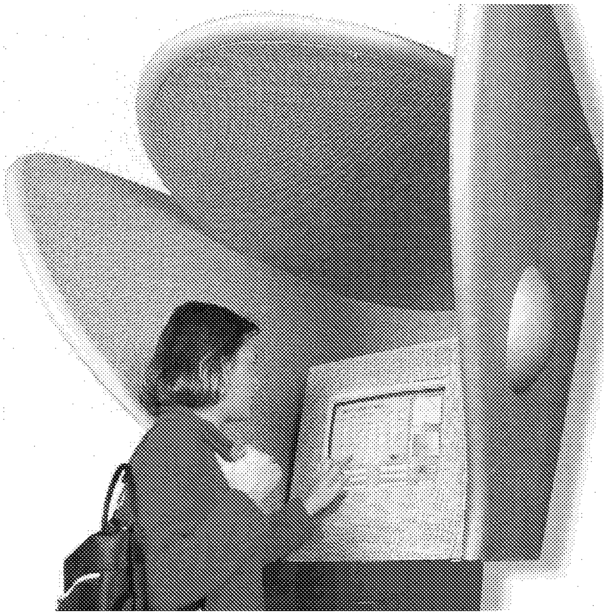
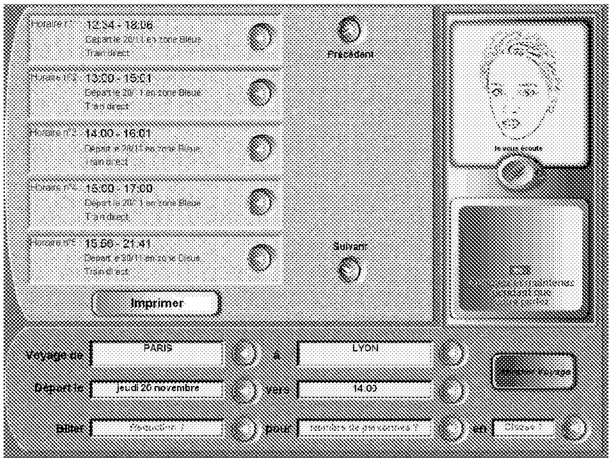


Figure 1: Photo of the MASK kiosk.

2. SYSTEM DESCRIPTION

The MASK prototype kiosk, as shown in Figure 1 was designed by the SNCF in collaboration with LIMSI, particularly for aspects concerning signal capture. Two prototypes were built, one to carry out spoken language system development work at LIMSI and the other to carry out the user trials at the St Lazare train station in Paris. Various kiosk designs were considered during the project, including a closed cabin so as to provide better acoustic isolation. An open design was preferered however for security and hygiene reasons. The kiosk has a touch screen for tactile



Horaire n°1 12:34 - 18:06
Départ le 20/11 en zone Bleue
Train direct

Horaire n°2 13:00 - 15:01
Départ le 20/11 en zone Bleue
Train direct

Horaire n°3 14:00 - 16:01
Départ le 20/11 en zone Bleue
Train direct

Horaire n°4 15:00 - 17:00
Départ le 20/11 en zone Bleue
Train direct

Horaire n°5 15:56 - 21:41
Départ le 30/11 en zone Bleue
Train direct

Voyage de PARIS à LYON
Départ le jeudi 20 novembre vers 14:00
Billet Réduction ? pour Nombre de passagers ? en Classe ?

Figure 2: Example user interface from MASK kiosk.

input, loud speakers (the bumps in the side panels) and two microphones, located just above and below the screen. The kiosk is able to provide train timetable and fare information, and simulated ticket purchases.

Figure 2 shows a picture of one of the interface screens. The face on the right is the clerk, which lets the user know what the machine is doing (waiting, listening, thinking, talking). The button below the clerk is for *push-to-talk*. The text tells the user to maintain the button pushed (i.e., to keep touching the button) while talking. (The button changes color when pressed.) The push-to-talk mode was found to be easily accepted by most users, greatly simplifying the speech detection problem [5]. The clerk and push-to-talk icons are always present on the screen. On the left of this screen is a list of trains satisfying the given constraints. At this point in the transaction, the user can select one of the trains vocally (by refering either to its position in the list or to the time) or by pushing on the button to the right of the desired train. The user can obtain earlier or later trains by asking for them or using the arrows.

The lower part of the screen resembles a train ticket, and summarizes the information known by the system. This part of the screen, displaying information required for ticketing, is always displayed. In this example, the voyage is from Paris to Lyon, on Thursday November 20th, leaving around 2 p.m. Incompleted items are marked with a question mark. In this example, the items corresponding to Reduction?, the Number of passengers? and the class? have not been completed.

The system architecture is shown in Figure 3. This architecture is a modified version of the LIMSI spoken language system (SLS)[3], intergrating the **Multimedia Interface** and the **Touch Screen**. The main components for spoken language understanding are the speech recognizer, the natural language component consisting of the semantic analyzer and the dialog manager, and an information retrieval component that includes database access and response generation. The speech recognizer is a software-only system that runs in real-time on a standard RISC processor. Statistical models are used at the acoustic and word levels. Acoustic modeling makes use of continuous density hidden Markov model (HMM) with Gaussian mixture. Speaker independence is achieved by using acoustic models which have been trained on speech data from a large number of representative speakers, covering a wide variety of accents and voice qualities. Bigram backoff language models are estimated on the orthographic transcriptions of the training set of spoken queries, with word classes for cities, dates and numbers providing more robust estimates of the n-gram probabilities. The recognition lexicon is represented phonemically and has 1500 entries, including 600 station/city names.

The output of the recognizer is passed to the natural language component which extracts the meaning of the spoken query using a caseframe analysis [1]. The major work in developing the understanding component is writing the rules for the caseframe grammar, which includes defining the concepts that are meaningful for the task and their appropriate keywords. The dialog manager guides the interaction with the user so as to obtain information needed for database access. Natural language responses are generated from the semantic frame and the information obtained from database access. Vocal feedback is provided by concatenation of speech units stored in a dictionary according to the automatically generated response text.

The interaction of the multimedia interface and the spoken language system is via the dialog manager. The multimedia interface interprets tactile commands and generates a Semantic Frame compatible with the SLS. The dialog manager integrates the tactile information into the current dialog context and controls database access. The high level decisions are taken by the dialog manager based on the context and the state of the interface, and low-level presentation decisions are taken directly by the multimedia interface.

An important difference in dialog strategies is offered by the input modes. The tactile strategy is a command driven dialog, where the user must input specific information in order to move on to the next step. Vocal input allows a real mixed-initiative dialog between the user and the system, where the user can guide the interaction or be guided by the system via the help messages.

3. USER TRIALS

Trials with 100 users were carried out to assess the performance of the final version of the prototype kiosk during a 7-day period in April 1998. Complementary user trials were carried out to compare the different input modalities, to compare the MASK kiosk to the current ticket machines, and to assess the effectiveness of the help messages, as well as graphical vs graphical and vocal output.

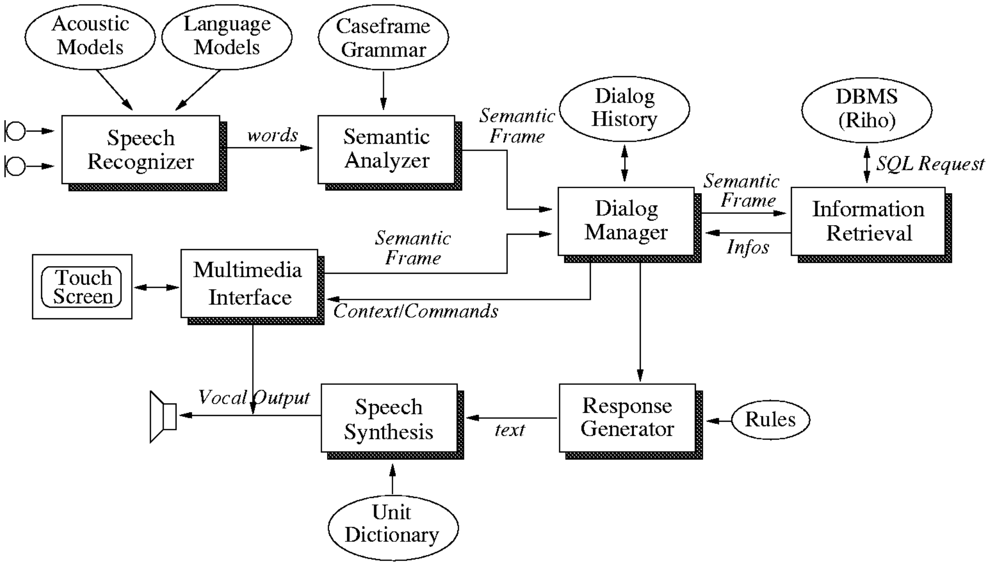


Figure 3: MASK system architecture.

3.1. Methodology

The user trials were conducted in the St. Lazare train station in Paris. An SNCF hostess selected customers in the train station, and asked if they would be willing to participate in a user evaluation of a new automatic ticket kiosk. Customers that were willing to were escorted to the demonstrator room. The hostess selected subjects so as to cover a wide range of ages for each sex. Subjects were given a brief introduction to the purpose of the study and to the tasks to be performed. They were given only a minimal amount of information about the kiosk, such as the possible input and output modalities, but without any specific details. Users were able to learn more about the system capabilities by watching the self-presentation.

Subjects were divided into 3 subgroups in order to evaluate the kiosk on different tasks: timetable information enquiry (25 subjects), price information enquiry (25 subjects), ticket purchase (50 subjects). In order to assess learning effects, each subject performed the given type of task four times with different scenarios. After each scenario the subject was asked to estimate the time it took to complete it. On completion of the test phase, the user completed a questionnaire and received a 50FF SNCF travel voucher. The questionnaire asked general questions about the subject and their computer experience and travel habits. A series of questions were aimed at their impression of the MASK kiosk such as their overall satisfaction, the facility of use, acceptance of the push-to-talk button, utility of the help messages, and confidence in the vocal input.

3.2. Experimental Results

The main results of the studies are given in Table 1, for the 3 task types: time enquiry, price enquiry, and ticket purchase. Tn shows the averaged results corresponding to the nth transaction of each subject. It is apparent that the time task is substantially simpler, requiring about half of the actions as are needed for price enquiry and ticket purchase. This is also reflected in the overall transaction times. The effects of subject learning are seen by the reduced number of inputs, help messages, and time as well as an increasing success rate. The percentage of inputs made vocally is around 20%, but for the 4th time enquiry task over half the inputs were spoken. For the time information parts of the price and ticket purchase tasks, a higher percentage of spoken inputs were observed than the task averages (close to those observed for the time enquiry task). On average over 40% of the transactions had at least one spoken input, and for 98% of the spoken inputs a sematic frame was generated.

Table 2 shows the overall user ratings compared with the project objectives. 74% of the users never or rarely encountered difficulties in using the system. Subjects were largely satisified with the usability and simplicity of use, with 98% of them quite or very satisfied.

3.3. Complementary studies

Additional studies were carried out to determine user preferences between the new kiosk and the automatic ticket machines (APV) currently in service and to assess the role of the different modalities offered by the MASK prototype. Each study involved a new set of subjects.

30 subjects participated in the comparative study, carrying out the ticket purchase task. 80% of the subjects preferred MASK, finding it fast and user-friendly, with a 95% preference by people who do not use the existing APVs and 75% by those that do. Users prefering the existing APVs had more problems with speech input than users preferring MASK, and being frequent APV users they were able to carry out very efficient transactions. A set of 14 subjects compared a tactile-input only version of MASK to the ex-

Time Task (25)	T1	T2	T3	T4
#inputs	5.2	4.6	3.7	3.2
%speech	23%	27%	46%	56%
$\geq$ 1 spoken action	41%	54%	43%	66%
#help messages	3.9	3.2	2.0	1.2
Transaction time	1'15	0'55	0'43	0'26
Success	79%	70%	97%	99%
Price Task (25)	T1	T2	T3	T4
#inputs	11.4	10.6	9.6	8.7
%speech	16%	20%	25%	25%
$\geq$ 1 spoken action	42%	45%	53%	41%
#help messages	11.0	5.8	3.7	2.8
Transaction time	3'44	2'02	1'46	1'11
Success	96%	89%	98%	99%
Purchase Task (50)	T1	T2	T3	T4
#inputs	13.1	11.9	9.4	9.8
%speech	13%	15%	15%	17%
$\geq$ 1 spoken action	43%	43%	45%	41%
#help messages	9.4	5.8	4.3	2.9
Transaction time	3'26	2'04	1'42	1'35
Success	85%	86%	92%	95%

Table 1: User trial results by task type: time enquiry, price enquiry, and ticket purchase. T1 - T4 correspond to the 1st - 4th time the task was carried out. An input corresponds to the provision of a data item and may be made by touch or speech.

No Difficulty	*Usability*	*Simplicity*	*Satisfaction*
74% (65)	86% (65)	93% (93)	98% (92)

Table 2: User assessment of the MASK kiosk. The objective ratings are shown in ().

isting APVs. The MASK transaction success was higher and the user-interface was preferred even though the transactions took longer.

The effectiveness of the help messages was investigated with a set of 15 subjects completing purchasing task without help messages. The help messages were found to be efficient in guiding the user, particularly for the first transaction, and enhanced the subjective evaluation. Subjects also used vocal input more often when help messages were available.

30 subjects compared tactile-only, vocal-only and free mixed modality use of MASK. Speech input was preferred slightly (53%), and had higher subjective ratings and was about 10% faster for the transaction compared to tactile-only. However speech-only was perceived as inefficient if the user needed to repeat, and had a higher error rate (15% vs 5%). Users prefering touch found it simpler and quicker, and were successful with their tasks. Those prefering speech were less accustomed to the APVs and their preferences were not affected by the success rate. When subjects were allowed to mix modalities, they were able to follow their preferences and optimise the transaction.

4. CONCLUSIONS

In this paper we have given an overview of the MASK prototype kiosk enabling interaction through the co-ordinated use of multimodal input (speech and touch) and multimedia output (sound, spoken messages, graphics and text). In order to achieve this goal, technical advances were required to allow real-time interpretation of user data entries via multiple input modalities and real-time integration of multimedia feedback to guide the user. A major consideration was the ability to interact effectively with naive users. User trials were carried out with over 200 subjects. These studies demonstrated that for this task multimodality is more efficient (faster and easier) than monomodality as some actions are better carried out by voice and others by touch. These studies also showed that subjects performed their tasks more efficiently as they became familiarized with the MASK system, learning to exploit the vocal input and benefiting from the multiple modalities. Most subjects preferred the new kiosk design, with a lower preference expressed by frequent users of the current kiosks who are used to carrying out their transactions.

5. ACKNOWLEDGEMENTS

We gratefully acknowledge the participation of Myriam Vergnes, Sylvianne Lemercier, and Franck Bernard in conducting and analyzing the user trials. This work was carried out in collaboration with the Ergonomics Group at UCL (University College London, 24 Bedford Way, London WC1H0AP, U.K.), a partner of the MASK project. Martin Colbert from UCL carried out a performance analysis based on the data obtained in the user trials.

6. REFERENCES

1. S.K. Bennacef, H. Bonneau-Maynard, J.L. Gauvain, L. Lamel, W. Minker, "A Spoken Language System For Information Retrieval," *Proc. ICSLP'94*, Yokohama, Sept. 1994.
2. J.L. Gauvain, J.J. Gangolf, L. Lamel, "Speech Recognition for an Information Kiosk," *Proc. ICSLP'96*, Philadelphia, Oct. 1996.
3. J.L. Gauvain, S. Bennacef, L. Devillers, L. Lamel, R. Rosset: "Spoken Language component of the MASK Kiosk" in K. Varghese, S. Pfleger(Eds.) "Human Comfort and security of information systems", Springer-Verlag, 1997.
4. H. Dartigues, F. Bernard, A. Guidon, J.N. Temem, "The MASK project : new passenger service kiosk technology," *World Congress on Railway Research '97*, Florence, Nov. 1997.
5. A. Life, I. Salter, J.N. Temem, F. Bernard, S. Rosset, S. Bennacef, L. Lamel, "Data Collection for the MASK Kiosk: WOz vs Prototype System," *Proc. ICSLP'96*, Philadelphia, Oct. 1996.
6. Final report of the ESPRIT MASK project, August 1998.