

Multimodal Human-Computer Interfaces for Incident Handling in Metropolitan Transport Management Centre

Yu Shi, Ronnie Taib, Eric H. C. Choi, and Fang Chen
National ICT Australia (NICTA)
Bay 15, Locomotive Workshop, Australian Technology Park
Eveleigh, NSW 1430, Sydney, Australia
Yu.Shi@nicta.com.au

Abstract - Efficient road traffic incident management in metropolitan areas is crucial for the smooth traffic flow and the mobility and safety of community. Traffic incident management requires fast and accurate collection and retrieval of critical data, such as incident conditions, and contact information for the intervention crew, public safety organisations and other resources. Access to critical data by traffic control operators can be facilitated through various human-computer interfaces. This paper describes the judicious introduction of a multimodal interaction paradigm to the user interfaces for incident handling in a metropolitan transport management centre. Prototypes supporting speech and gestural interaction have been built based on user-centred design methodology and their evaluations have been conducted through user studies. The presented innovative user interfaces provide traffic control operators with intuitive, cognitively efficient ways to record traffic incident conditions, facilitate fast retrieval of contact details, and support time-critical incident handling.

I. INTRODUCTION

Efficient road traffic incident management (TIM) is essential to the overall operations of transport management. TIM aims to reduce the detection, verification, response, and clearance times of traffic incidents, thereby securing smooth traffic flows for the vitality of the economy and the mobility and safety of the community. Metropolitan TIM operations are typically carried out in a control room where operators are responsible for planning and managing traffic flows, incident responses and operations of traffic lights over a very large area, with the help of information and decision support systems. The operators have to grapple daily with multiple software applications on multiple display screens with various user interfaces to accomplish various tasks. They have a great need for rapid, convenient means of retrieving and extracting information.

User interface is an integral part of modern information and communication technology systems, including intelligent transportation systems (ITS). Well-designed user interface can help users perform tasks such as database search more efficiently and effectively. Multimodal User Interaction (MMUI) is an emerging technology in Human-Computer Interfaces (HCI) that aims to enable computers to recognise natural human communication modalities, such as speech, gestures, facial expression and other body languages, in order to facilitate user's interaction with the computer and improve tasks efficiency and use experience.

We have been collaborating with a metropolitan transport management centre in Australia to streamline parts of their TIM system by introducing natural multimodal user interaction technologies into their incident management system. In particular, we set out to design and develop novel user interfaces for their Contacts Database (CDB) system which is used by the traffic control operators to find and contact a specific group of persons from various government agencies and external organisations (such as field traffic crew, media, local council, local towing company and other local resources) when a major incident has occurred. The main objective of our work is to design a multimodal user interface for the CDB that facilitates fast entry and retrieval of critical contact details. Ultimately, road users will benefit from the ease-of-use of these interfaces and from the improvement in the effectiveness and efficiency of incident handling, in terms of lower error rate and faster response time.

Our work over the last year has gone through four stages, designed to precisely capture operator preferences and application requirements through User Centred Design (UCD) [1], and gradually introducing MMUI technologies into the TIM operations:

- Stage 1: design and realisation of a series of user studies to capture operators' preferences and CDB application requirements (section III);
- Stage 2: design, implementation and evaluation of a graphical user interface (GUI) for CDB to ease performance bottlenecks for short-term deployment (section III);
- Stage 3: design, implementation and evaluation of a user interface for CDB based on multimodal interaction technology, by using speech and pen-gesture inputs, in order to significantly improve performance for medium-term adoption (section IV);
- Stage 4: design of a user interface based on natural speech and free-hand (i.e. no device to be worn) gesture inputs, for long-term deployment in major event planning and management. A distributed architecture based on software agents has been developed to accommodate flexible changes and extensions of the system. This stage is still a work-in-progress (section V).

In this paper, we report progress in research and

evaluation of the application prototypes developed to meet the above design objectives. The long-term goal of our research is to develop robust multimodal user interfaces with low cognitive load for emergency event planning and crisis management.

II. RELATED WORK

Multimodal user interfaces based on speech recognition and pen-based gesture input were designed and studied in the early 1990's [2]. One of these early multimodal systems is the QuickSet [3] which was designed for map manipulation on a handheld computer. Another multimodal interface system which targeted at military command post environment can also be found in [4]. In [5], Sharma et al. presented a speech and free-hand gesture driven user interface for crisis management. An application of this work for accessing geospatial data was also reported in [6]. So far, existing studies on road traffic control centres have focused on the automatic acquisition of information, such as the construction of traffic models for simulation and the optimal control of traffic light signals [7][8]. To the best of our knowledge, there has been no research work reported in the literature on the application of MMUI technologies to road traffic incident management. Although some high-level human factors guidelines [9] have been developed for setting up a traffic control centre, these guidelines are only related to traditional input and output devices, such as mouse and keyboard.

III. USER STUDIES AND GUI-BASED INTERFACE PROTOTYPE

A. Motivation

User-centred design (UCD) methodology is critical to the successful design, development and deployment of human-computer interfaces for applications. This involves an iterative interaction with end-users before and during the whole design cycle. With the support of the management and operators working at the traffic incident management centre, we have conducted a series of user studies with the following objectives:

- To identify performance bottlenecks experienced by the operators when using the existing web-based contacts retrieval system;
- To gather user preferences regarding the best ways to submit queries and search for information; and
- To gauge operators' interests in the use and deployment of new technologies such as speech-aided database search.

B. Interviews and Questionnaires

Fourteen TIM operators participated in unpaid face-to-face interviews. The interviews were carried out to identify the environment, tasks and main issues related to the use of the contacts database. These interviews were then followed by an off-line questionnaire, filled in by the same participants, to solicit the operators' user interaction preferences as well as their opinions on the identified major

issues. The questions targeted either the task flow or some specific interface parameters that could inform the design of an improved GUI.

For example, these studies revealed that:

- All respondents agree that an updated and consistent contacts database is the most important requirement;
- 78% of the respondents prefer a map-based search for geographical information over a textual search;
- 64% of the respondents select radio buttons as their preferred input method to GUI;
- 43% of the respondents suggest the introduction of a speed dial/email/SMS functionality, triggering those features by a single click;
- 64% of the respondents think that a speech interface would be beneficial, 21% are unsure or sceptical but would be happy to try out. For more general multimodal interface (speech, gesture, touch and others), 21% think that it would be beneficial if it works well, and there has been no negative opinion expressed.

Overall, the results reflected known issues in the current system, such as data inconsistencies due to the lack of central database. However some other results raised new issues and provided important insights for the improvement of the interface, such as the need for map inputs at some specific stages of the interaction. Another important outcome was the identification of possible ways to introduce standard operating procedures through the flow of information presented in the user interface.

C. New GUI and User Experiment

An electronic mock-up of the new GUI based on the findings of the previous studies was designed and implemented (Figure 1). This interface essentially comprises three panels. The location area on the left panel only mimics the main functionality of the current geographical information system (GIS) to ease evaluation, so we ignored any interaction with it during the experiment analysis. An incident condition specification panel was introduced at the top right corner to reflect standard operating procedures. Finally, a digest-format list of contacts, relevant in the specific case, was presented in the bottom right area.

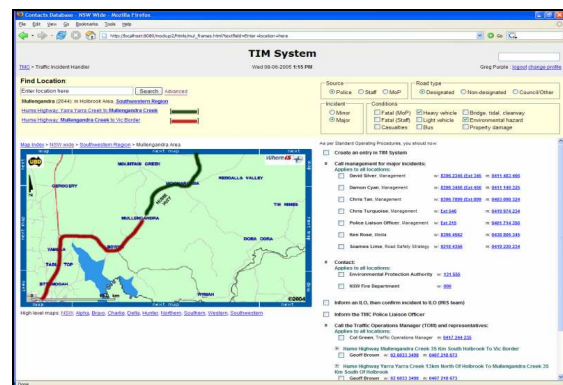


Fig. 1. New GUI and web-based incident handling interface.

We carried out a comparative user experiment with 6 TIM operators. They were instructed to handle a complex, real-life incident scenario using both the current system and the new GUI. Performance was assessed through time-to-completion as well as actual versus theoretical number and identity of contacts to be called during the procedure.

An encouraging 37% improvement in time-to-completion was observed across subjects with the interface mock-up, in spite of no formal training. This is mostly due to the information being presented in digest form on a single page, saving many browsing and scrolling actions. We also observed a 59% improvement in number of contacts correctly called with the mock-up, essentially due to the introduction of operating procedures and progress markers in the interface.

IV. SPEECH AND PEN BASED MULTIMODAL USER INTERFACE FOR INCIDENT DATA ENTRY

A. Motivation

In-situ observations showed that operators consistently use a paper scratch pad to take notes about an incident while on the phone. While a paper pad can be a very effective and robust way to take notes, and phone is the main information source for the operators, they involve a “double entry” of the data since the operators need to type in the outcomes of their conversations as well as their written notes into an electronic form. We proposed a speech/pen based UI mock-up to address this issue, yet leveraging the current practice.

B. Interface Design

The vocabulary used in spoken and written tasks during an incident handling varies greatly as a function of the specific incident conditions as well as the operator’s individual style and preferences. However, we hypothesised that some key elements would usually be identifiable through keyword matching against synonym lists. Similarly, location names or popular acronyms may also be recognised efficiently by the system. In this case, an automatic speech recognition (ASR) and handwriting recognition modules can be used to passively monitor the operator’s conversations and written inputs, capture any relevant words and automatically fill in the incident report forms accordingly. For example, the map can be automatically updated when the operator confirms or spells out the incident location name provided by a member of the public over the phone.



Fig. 2. Speech and pen based multimodal UI system setup.

The major constraint in the system design was to stay as close as possible to the current operators’ work environment, in order to ensure smooth acceptance and deployment of the new technology. A near horizontal tablet screen was introduced for the pen input and the audio capture was made available to the system via the usual headset worn by the operators, as shown in Figure 2. Information is simultaneously displayed on some sections of the tablet screen, as well as on a vertical screen, reflecting the current 2 to 3 screens set-up used by the operators.

The speech recognition happens continuously, but the operators may deactivate it using the microphone switch. The horizontal Tablet User Interface (TUI) works with a special pen hence allows for hand or other object to be put on the tablet if required. This becomes the main information input means for the operators, basically replacing the traditional mouse and keyboard. The user interface has been designed specifically for the use of pen for both handwriting and “ticking”. Operators can write anywhere on the tablet surface to enter free form words that become recognised. They can alternatively use tick boxes corresponding to the incident condition specifications.

Speech, handwriting and ticking inputs eventually get interpreted by the system into incident related commands that update the incident form being filled in. For example, saying or writing the word “Tanker” is equivalent to ticking the “Heavy vehicle” box, as shown in Figure 3. In all these three cases, the “Heavy vehicle” box gets ticked, and the system records this specific condition for the current incident and fetches the contact details of the heavy vehicle inspector who is in charge of the location of the current incident.

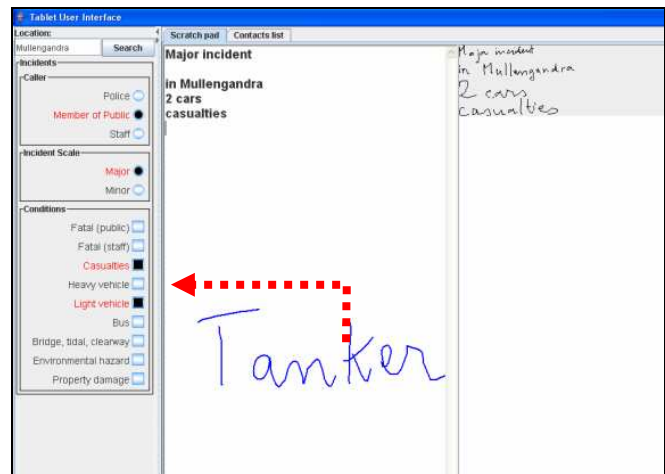


Fig. 3. Tablet User Interface (TUI).

The horizontal TUI is also capable of displaying the list of contacts to be called once the incident conditions have been fully specified. Tabbed menus allow switching between the handwriting entry mode and the list of contacts. Ticking any contact on that list simulates the trigger of an actual call to that contact person. While so doing, the contact details and photo of the contact person get displayed on the vertical screen (Figure 4).

The vertical Display User Interface (DUI) takes care of passive information display, such as the contact list, individual contact details mentioned above, or the map of the current incident. As soon as relevant information is interpreted by the system, the map and contacts list are updated in the DUI. The operator can even focus on the DUI while still entering information via speech or possibly handwriting. However, we minimised the need for change of user focus between the two screens by clustering information input and output in relation to the task flow.

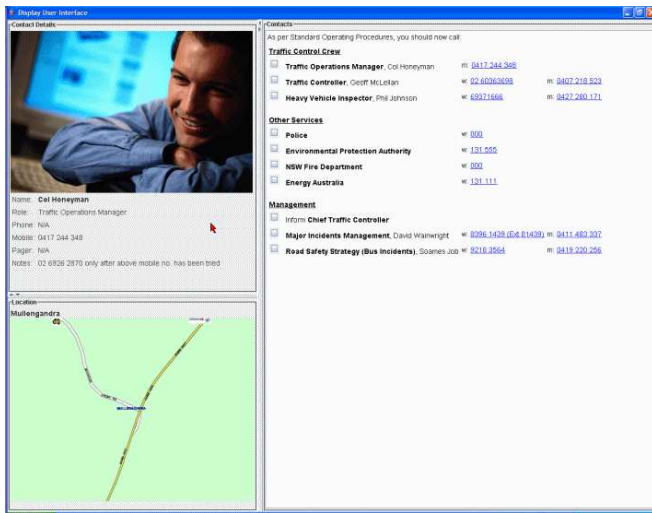


Fig. 4. Display User Interface (DUI).

C. System Architecture

Multi-view model-view-controller underpins the design of the interface, allowing for flexible graphical layouts as well as equivalence among the various input modes. The simplified functional block diagram of the mock-up is shown in Figure 5.

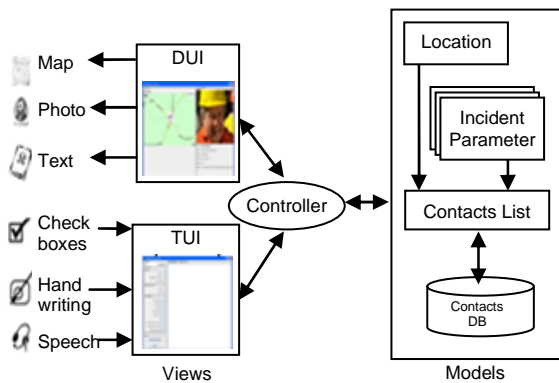


Fig. 5. Simplified functional block diagram of the multimodal UI system.

The mock-up uses the Microsoft Speech automatic speech recognition module; handwriting recognition is performed by the Microsoft Windows handwriting recognition module. A copy of the original handwritten input is shown alongside the recognised version on the tablet to alleviate any possible misrecognition.

D. User Study and Results

Eight paid volunteers, who are unfamiliar with traffic incident handling, took part in the experiment. The task assigned focused on the characteristics of the pen/speech interaction, not on the domain (incident handling) knowledge. We constructed a complex yet realistic incident scenario with the help of some non-participating traffic controllers. An experimenter was calling the subjects from another room, claiming to be a member of the public and described the incident in the same way to all subjects. The subjects were asked to take notes and input the incident description as quickly as possible into the system. Once completed, they were simply asked to call the Traffic Operations Manager responsible for the incident. A manual switch allowed them to feed their speech into the speech recogniser at any time, including during the phone conversation. They could also use handwriting at any time.

The general user preferences, e.g. heavy speech users versus heavy pen users were also targeted. Since most subjects were new to the use of speech and pen gesture for computer input, general subjective feedback on these input modes was also sought through a questionnaire.

A detailed analysis of the video footages of the subjects carrying out a total of 56 tasks shows that about 81% of the information was entered by ticking the radio buttons and check-boxes, 7% using handwriting and 5% using speech (see Figure 6). Comparing the information entered by the subjects with the original scenario described to them, only 7% of missing inputs were recorded, showing that despite a low use of speech and handwriting, a mouse and keyboard-free interface represents an acceptable alternative to information input. The TUI interface used here effectively replaces a scratch pad, even during a phone conversation, avoiding double handling of information. All subjects satisfactorily identified and called the Traffic Operations Manager as required.

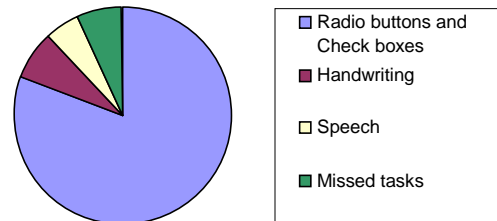


Fig. 6. Preferred input modalities.

Due to the evaluation timeframe, the mock-up was using an untrained American model for the speech recognition engine, hence spawning very average recognition rate. 7% of the subjects tried speech unsuccessfully first, then switched to ticking. Similarly, we observed that subjects struggled during the first use of the tablet interface. 14% of the subjects tried handwriting unsuccessfully first, then switched to ticking.

Further experiments involving prior training of the speech recognition engine, as well as training of the subjects on the two technologies may help determining whether higher speech and handwriting recognition rates would promote the broader adoptions of those input modes. However, it can be expected that ticking will probably keep the lead since it is most appropriate and effective for quick toggling, and had been indicated as preferred input modality in the stage 1 questionnaires. Speech and pen input may be dominant though if the operators wish to keep written notes of the events, in parallel to activating the switches.

In terms of comparison between speech and handwriting, we observed that 75% of the subjects entered the location (e.g. “Mullengandra”) by handwriting whereas 25% did it by trying speech first.

We also collected the opinions from the subjects regarding the usability of the speech/pen based mock-up. The average ratings on a 5-point Likert scale (1: strongly disagree, 5: strongly agree) for the various UI aspects of the mock-up are shown in Figure 7. As observed from the figure, the interface of this mock-up was quite highly rated by the subjects in many aspects and in particular its overall effectiveness. Given that most of the subjects had never used handwriting and speech recognition before, the results are encouraging.

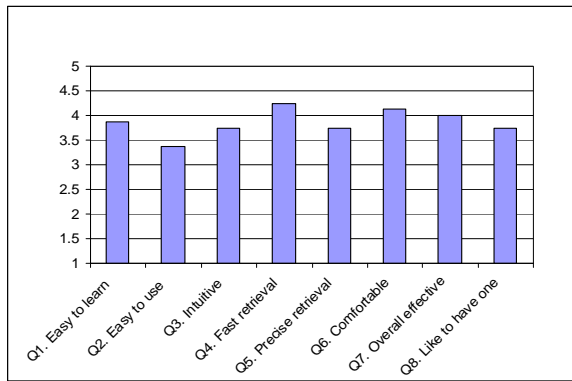


Fig. 7. Assessment of the usability of the speech/pen system.

V. SPEECH AND GESTURE-BASED MULTIMODAL USER INTERFACE FOR EVENT PLANNING/EMERGENCY HANDLING

A. Motivation

The fourth stage of this research aimed at introducing speech and free-hand gesture inputs for interaction with large screens. This scenario is inspired by the specific requirements of event planning and emergency handling, where an overall view of the information is required as well as simple and effective interaction. Untethered gestures combined with natural language speech input can provide intuitive interaction with the system, so that the operators can exclusively focus on the task at hand without extra cognitive load induced by the interface.

B. System Design

We developed a mock-up system, called PEMMI (Perceptually Effective Multi-Modal Interface), to

demonstrate the feasibility of this technology. The multimodal inputs are exclusively provided by a video-based gesture recognition module developed by our team, and a commercial speaker-independent automatic speech recognition module.

This work is part of our larger ongoing research on multimodal user interaction and we designed the architecture to be distributed and easily extensible so that other modalities or dedicated incident handling functions modules can be plugged in. We selected the JADE (Java Agent Development Framework, <http://jade.tilab.com/>) multi agent system in order to address those requirements. Software agents provide an inherently distributed architecture with an embedded message-driven communication framework. Such agents can then act autonomously, allowing dynamic reconfigurations of the system, e.g. when a new input modality is plugged into the system. Figure 8 gives an overview of the PEMMI architecture including the following components:

- The automatic speech recognition (ASR) module analyses the operator’s spoken utterances and recognises voice commands out of a small vocabulary;
- The video-based gesture recognition (VGR) module tracks the operator’s hand motions and recognises events designed to either disambiguate voice commands or interpret independent gesture commands;
- The semantic multimodal input fusion (MMIF) module assesses whether the voice and gesture inputs can lead to a combined meaning or they correspond to independent meaningful units;
- The dialogue management (DM) module prepares the system response to the semantic command interpreted by the fusion module, based on the application scenario;
- The multimodal output generation module provides a map-based graphical user interface together with location-based artefacts such as camera or police station icons and an animated avatar with speech synthesis.

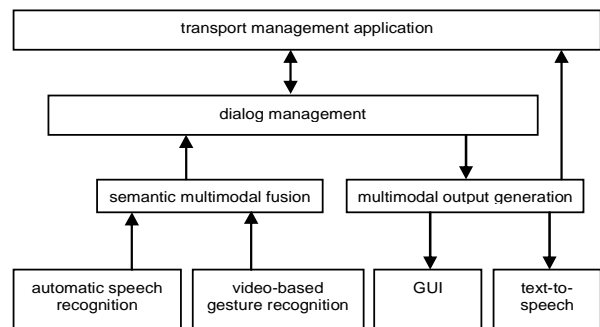


Fig. 8. Overview of the PEMMI interface.

Figure 9 shows the physical implementation of the PEMMI. A low-cost web camera mounted on a tripod, about a metre away from the arm, is used to capture the hand

gestures. A Bluetooth wireless microphone (not seen in Figure 9) is worn on the left ear of the operator.

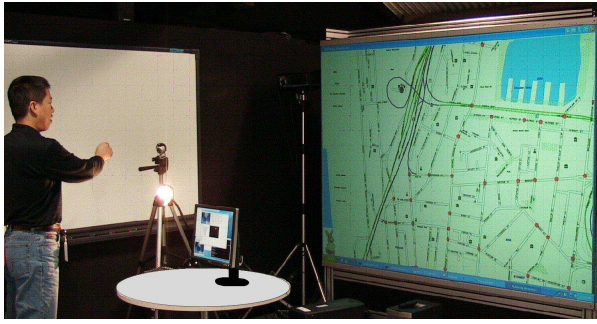


Fig. 9. Photo of the PEMMI implementation.

C. Features and Work In-Progress

The VGR module in PEMMI can detect and recognise hand movement trajectory and events such as HAND_UP, HAND_DOWN and HAND_PAUSE. These can map to deictic gestures such as “pointing” gestures and “drawing” gestures. The ASR module can recognise words from a small vocabulary of command phrases, such as Play Video, Show Cameras, Stop, Call and others. The fusion of speech input and gesture input by MMIF module allows a user to issue multimodal commands such as “Show cameras in this area”, “Call this police station”, and so on, without using a keyboard or mouse. An application to traffic incident handling using PEMMI, believed to be among the first of its kind in the world, has been designed, implemented and demonstrated to both traffic control operators and general public, with very encouraging results and feedback.

Work in-progress includes:

- Continuous improvement of VGR module to allow recognition of different natural and pre-defined gestures, so as to support more multimodal commands;
- Use of PEMMI as a platform to analyse cognitive load of a traffic control operator; and
- Introduction of the family of IEEE 1512 Standards on traffic incident management to the DM module, so that traffic condition, description and handling data generated by PEMMI conform to international standards. This can facilitate data exchange between the TIM system and other traffic systems within the local and national ITS systems.

VI. CONCLUSIONS AND FUTURE WORK

In this paper we have presented four stages of a User-Centred Design approach to the introduction of multimodal user interfaces in traffic incident management.

The first stage consisted in the extraction of requirements and analysis of issues perceived by the actual operators of the system. It led to the implementation of an improved graphical user interface in stage 2, showing encouraging results, such as 37% reduction in time-to-completion, for a complex incident handling in a comparative user experiment against the current contacts management system.

During the third stage, we introduced speech and pen (handwriting and ticking) interface, in a bid to avoid double entry of information, yet being as close as possible to the current environment and interaction style of the operators. A mock-up interface comprising two screens, but no keyboard or mouse was used. The outcome of a user experiment showed the strong preferences of operators for ticking checkboxes, certainly due to the unfamiliarity of the other modalities. It also showed a positive response from subjects regarding the usability of the system.

The last stage explored more advanced multimodal inputs with the use of hands-free gesture and speech in front of large screens. A distributed architecture based on software agents was developed to accommodate flexible changes and extensions of the system. This work-in-progress explores the use of more natural interaction styles in the context of incident handling or event planning.

We are now planning to carry out an evaluation of the usability of the event planning prototype, in order to identify opportunities for gestural and speech inputs in the TIM context. Some of them may include large screens as in the prototype, but may also apply to desktop settings with other types of gestures.

ACKNOWLEDGEMENT

The authors would like to thank the operators and volunteers for their participation in the studies. Also thanks to the other team members and students who participated in the development of some of the prototypes.

REFERENCES

- [1] Vredenburg, K., Isensee, S. and Righi, C., *User-Centered Design: An Integrated Approach*, New Jersey, Prentice Hall, 2002.
- [2] Oviatt, S. L., Cohen, P. R., Fong, M. W. and Frank, M. P., “A Rapid Semi-automatic Simulation Technique for Investigating Interactive Speech and Handwriting”, *Proc. International Conference on Spoken Language Processing*, 1992, vol. 2, pp. 1351-1354.
- [3] Cohen, P. R., Johnston, M., McGee, D., Oviatt, S., Pittman, J., Smith, I., Chen, L. and Clow, J., “QuickSet: Multimodal Interaction for Distributed Applications”, *Proc. Fifth ACM International Multimedia Conference*, 1997, pp. 31-40.
- [4] McGee, D. R., Cohen, P. R., Wesson, R. M. and Horman, S., “Comparing Paper and Tangible, Multimodal Tools”, *Proc. ACM Human Factors in Computing Systems, CHI 2002*, vol. 1, pp. 407-414.
- [5] Sharma, R., Yeasin, M., Krahnstoeber, N., Rauschert, I., Cai, G., Brewer, I., Maceachren, A. and Sengupta, K., “Speech-Gesture Driven Multimodal Interfaces for Crisis Management”, *Proc. IEEE, Special Issue on Multimodal HCI*, vol. 91, no. 9, Sept. 2003, pp.1327-1354.
- [6] Schapira, E. and Sharma, R., “Experimental Evaluation of Vision and Speech Based Multimodal Interfaces”, *Proc. Perceptual User Interface, PUI2001*, ACM Press, pp. 1-9.
- [7] Birst, S. and Smadi, A., “An Application of ITS for Incident Management in Second Tier Cities: A Fargo, ND Case Study”, *Proc. Mid-Continent Transportation Symposium*, Iowa, 2000, pp. 30-34.
- [8] Hernandez, J.Z., Ossowski, S. and Garcia-Serran, A., “On Multiagent Coordination Architecture: A Traffic Management Case Study”, *Proc. IEEE 34th Hawaii International Conference on System Sciences*, 2001, pp. 1-9.
- [9] Kelly, M.J., *Preliminary Human Factors Guidelines for Traffic Management Centers*, Technical Report FHWA-JPO-99-042, U.S. Department of Transportation, 1999.