A System for Adaptive Multimodal Interaction in Crisis Environments

Dragoş Datcu, Zhenke Yang, and Léon Rothkrantz

Delft University of Technology
Faculty of Electrical Engineering, Mathematics and Computer Science
Mekelweg 4, 2628 CD Delft, The Netherlands
{D.Datcu,Z.Yang,L.J.M.Rothkrantz}@ewi.tudelft.nl

Abstract. In the recent years multimodal interfaces have acquired an important role in human computer interaction applications. Subsequently these interfaces become more and more human-oriented. Humans use multimodality to reduce ambiguity and incompleteness of information. Seemingly they are able to switch easily from one modality to the other and fuse the information from different multimodal sources. The goal of our research was to develop a crisis based human like multimodal system. In particular, we bring into focus the multimodal interaction between human users and the automatic crisis system and its correlation with the adaptability to the human behavior in crisis situations. Our system is capable of conceding for an optimal interaction process by taking into account the major informational human channels while gathering the user inputs and producing the system feedback. In this paper we describe the design of our system which is implemented as a running prototype. We have conducted a simulation of a crisis event to measure the degree of user satisfaction. At last we discuss the drawbacks as well as the premises of our solution in the context of the high level of performance achieved by our approach.

Keywords: Multimodal human computer interfaces, adaptive interfaces, crisis support systems, multimodal framework.

1 Introduction

Recently a special focus has been noticed on areas concerning the development of support systems for crisis situations. More and more attempts are engaged to automate processes to manage the communication and enhance the interaction of actors at different crisis sites. In addition, the current developments of hardware platforms and equipments allow for highly demanding processing and device interconnectivity. A specific requirement in such a context points to reliable multimodal interfaces to link different components and to sustain the information flows through characteristic layers of abstraction.

The main contribution of the research described in the current paper, is given by the diversity in multimodal human computer interaction. This comes as part of the solution for improving the usability of automatic systems for the management of

N. Aykin (Ed.): Usability and Internationalization, Part II, HCII 2007, LNCS 4560, pp. 334–343, 2007.

crisis situations. During a crisis event, panic and confusion are two reasons that usually lead to an increase of the typical damaging effects. We aim at limiting these effects by developing system interfaces that are self-adaptive given the variety of users and situations. The interaction between the users and the automatic system is eased by the intelligent information aggregation and the continuous adaptation to the user needs and to each crisis situation. The behavioral differences between all the actors interacting through our multimodal system are efficiently managed by a common strategy on the plan to solve the crisis situation.

At the global scale, we have achieved a high degree of internationalization for the users interacting with the system by focusing on the selection of modalities to support language independent user communication.

The adaptive crisis multimodal framework being described in this paper is centered on the shared memory paradigm. Comparing with the traditional way implying direct connections between the system components each connection having its own data format, the new approach suggests a more human-modeled alternative to store, retrieve and process the data. The information is conferred an underlying structure that complies with eXtended Markup Language (XML). The shared memory in the current design of the multimodal framework takes the form of XML data spaces. The use of shared memory allows for loosely coupled asynchronous communication between multiple senders and receivers. The communication decoupling is realized both in time and location. The specification fully complies with the requirements of data manipulation in a multimodal environment where the availability of data is time-dependent and some connections might be temporarily interrupted.

Considering the study case of the automatic crisis application, wireless devices such as PDAs or mobile phones can communicate and exchange essential crisis multimodal information. One distinct remark concerning our multimodal system is the adaptability to various working conditions. The adaptation induces an in-built context-aware mechanism to intelligently interfere with the external world in a natural manner. In the case of crisis applications it assumes dynamic and transparent system auto-configuration to get optimal performance given any crisis specific environments, human actors and hardware devices. The system adapts the information extraction in terms of audio and visual channels. An example is the case when specific information regarding the emotion from the speech of one person cannot be computed due to noisy environment or lack of a special processing component. In such a case and if there is no occlusion of the person's face, equivalent information is generated by employing the proper component to perform facial expression recognition. The network data transfer is optimized so as to avoid data blockage and to generate a good flow of the information through the connected processing components. The system feedback is generated taken into account the informational channels and the human computer interfaces available. An example is that the system automatically decides to ask the user a question via a loud message in the speakers instead of a panel text message if the lighting conditions in the room are poor.

A throughout technical description of the multimodal framework that supports the adaptive interfaces detailed in the paper, is given in [2]. In the next section we present related work in the research field of multimodal interfaces in crisis management. In section 3 we describe the system architecture. Section 4 presents the results of the experiment we done for determining the performance of the system.

2 Related Work

Recent research advancements on the area of support applications in crisis situations have accentuated the need for new algorithms and methods to cope with the specific issues of adaptive work environments and adaptive data distribution. Novel techniques have come into play for connecting users over dynamic wireless networks and for providing the user with the most relevant information given specific crisis environments. The work of [7] describes a system for routing people outside a dangerous area using a personalized dynamic routing algorithm in case of emergency. The architecture is based on multi and mobile agents. Each human user is supervised by a specialized agent that learns the behavioural peculiarities of its human counterpart. The interface of the system uses a set of graphical iconic representations that allow the user to provide his input accordingly. The research in [8] presents the use of an icon language for describing crisis situations and its integration with blackboards in MANETs [1]. The work of [9] investigates the use of the emerging computing model of Dynamic Data-Driven Applications Systems as base for the support of emergency medical treatment decisions in response to a crisis. By linking real-time sensors, procedural and geographic data, the system manages to produce decision support at the site of the incident, at local centres and at the central point of coordination. [12] gives extensive discussions over the role of multimodal interfaces on the specific elements regarding the crisis management and tackles various issues in enriching human computer interfaces with dialog and speech-gesture capabilities. [10] tackles the issues that rise from the integration of an intelligent agent software robot into a crisis communication portal for sending news alerts on mobile devices. The work of [13] adopts novel techniques in mixed and virtual reality technologies to enhance systems aiming at the surveillance, security and emergency, prevention plans in crowded environments. The interactive control room processes video data from airborne and fixed cameras and along with GPS driven maps generates real-time augmented 3D videos with the crowd for risk and prevention planning. The data can be also visualized by on-field human agents. The human computer interface is enriched with an eye tracker based mechanism to allow for the control of camera functions and views by gaze. The research described in [15] proposes a distributed multi-agent architecture for crisis response management and discusses the solutions for providing the necessary support. The work presented in [14] provides a classification of artificial coordination strategies in terms of skill, rule and knowledge. The research is applied in a case study of medical personnel to casualty allocation in the crisis response domain. The conclusions reflect the trade-off between efficiency and flexibility indicating the strategies of knowledge-level coordination as the most effective, and skill-level as the most efficient. The change of operational requirements can be optimally handled through the knowledge-level strategies as opposed to the performance of skill-level strategies in such context. The work of [11] addresses the requirements of disaster relief operations and proposes a solution based on an extension of the existing Belief-Desire-Intention BDI model having the capability of situation awareness.

3 The Architecture of the Adaptive Multimodal System

The automatic crisis system aims at solving the problems occurring during the process of collecting the input from different human users located at different points in the crisis scene and at fusing these partial observations so as to provide pertinent information related to the evacuation, help and coordination of specific actions for attenuating the causes of the crisis. The usual scenario for this case consists of a set of human observers and qualified personnel equipped with personal mobile devices (Fig. 1) that communicate with the system through a powerful multimodal interface. They are entitled to dynamically make reports on their own experience with respect to the crisis scenario while making use of information as it is offered by the system. The interface keeps track of the individual user inputs and updates the user profile. The information is eventually used in the attempt to solve the ambiguity in the personal reports. When possible, all the interfaces are interconnected through XML data spaces that run over existing wireless networks. The special issues that are taken into account focus on the possibility of sudden breakdowns of the infrastructure, the interoperability and limited processing power of various user mobile devices and the possible occurrence of individual and global ambiguous user reports.

Fig. 1. Zaurus PDA (*left*) and the fireman user profile using the crisis multimodal adaptive system (*right*)

3.1 The User Interface

During a crisis event different people have access to the system through individual multimodal interfaces, running on personal mobile devices. The acquisition of data from the users follows an as natural as possible process allowing the user to use different modalities to produce the relevant input to the system. The interface is able to conveniently take the multimodal input from the user and to send it to the other devices in the crisis informational network. Seemingly in order to create an input the user can use the visual set of crisis icons, text messages, pen input, photos and direct speech. Depending on the hardware of the user mobile device, the interface should furnish with as many as possible of these input modalities. The interface adjusts the input and output according to the characteristics of the device. The standard view generated by the interface consists in a 2D image (Fig. 2) showing a map associated to the location of the user. If the hardware facilitates the 3D rendering of images, the user can switch to this visualization mode (Fig. 3).

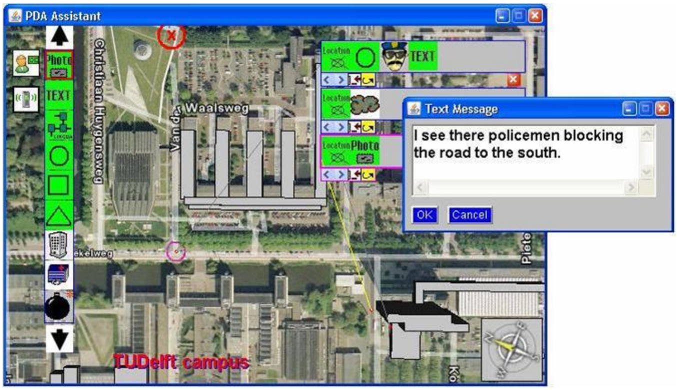

Fig. 2. The 2D user interface used by the external observer to report on the smoke crisis event

The use of GIS data in decision support systems for crisis management has been already adopted in systems as those presented in [3], [5] and [6]. By making use of GPS data, the interface introduces geographical based information into the process of interacting with the user. At this level the system can already create particular correlations among the user observations and the environmental area.

Fig. 3. The 3D user interface (*left*) used to report a crisis event and a photo of the view (*right*) taken by the user with his mobile device

In the case GPS data are not available the user has the option to simply mention the area where he is located. The interface will take the notice regarding the location as reference. Eventually, the location parameter is assigned with a slightly higher degree of uncertainty in all his further reports on the crisis event. The user can use both the 2D and 3D interfaces of the crisis application. To report about an event, the user can drag icons on special locations on the map or accentuate a special point of interest using geometrical symbols [4]. The icons are ordered in a hierarchical way. Firstly, the user has to select an appropriate context (accident, fire, terrorist attack, etc.). Given a certain context, a special set of icons is available to the user. The interface is also adapted to the role the user plays during the crisis event.

Laymen and professionals have different interfaces and different set of icons. For example a fireman is able to accurately report about the smell and color of a toxic cloud. That expertise can not be expected from a civilian and so the specific icons for such observations are not available. An automatic synchronization is realized on the information on each view so as to preserve the consistency of the crisis informational content. Based on the input provided by the user (Fig. 2), the system interface automatically generates the 3D crisis event scene (Fig. 3).

3.2 The Informational Disambiguation Model

The system manages the communication on the informational channels and supports the collaborative work for the specialized personnel working at the site of crisis. The specialized users hold functional precedence over the category of common users. This requirement is due to the fact that regular users are more exposed to failure in providing essential information over the crisis context when compared with personnel qualified in performing such activities.

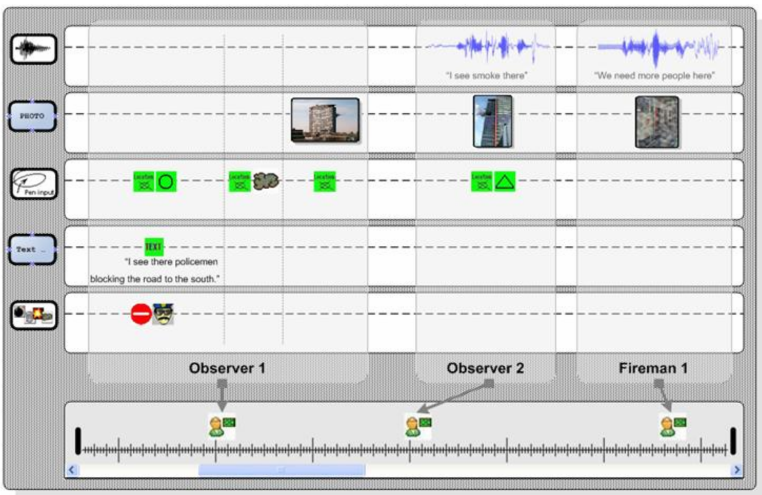

Fig. 4. The timeline with the integration of different contextual crisis multimodal information

All the previous user reports are taken into account when deciding for a certain operation to change the internal world knowledge of the system. All the observations coming from the field are stored and tracked while the system attempts to classify the crisis event as being one of the standard crisis scenarios from the system repository. At this level the solution for removing the ambiguity on different crisis scenarios is solved by employing the description and properties of each crisis context available in the repository as system knowledge. The classification represents a continuous process updating the event properties as soon as new evidence is made available by the system reasoning. According to the description of the current recognized crisis context, the system is able to generate feedback to the users in the form of support for distinct ongoing actions. All the user reports are collected and processed in an automatic manner. Human experts can also access the preliminary information of such automatic modules of the crisis system and can finally make adjustments on particular parameters of the results. This stands for the highest layer at which the ambiguity in partial, user observations can be decreased. Although is the most accurate among the all types of disambiguation, the manual procedure is time consuming and involves the presence of specialized people.

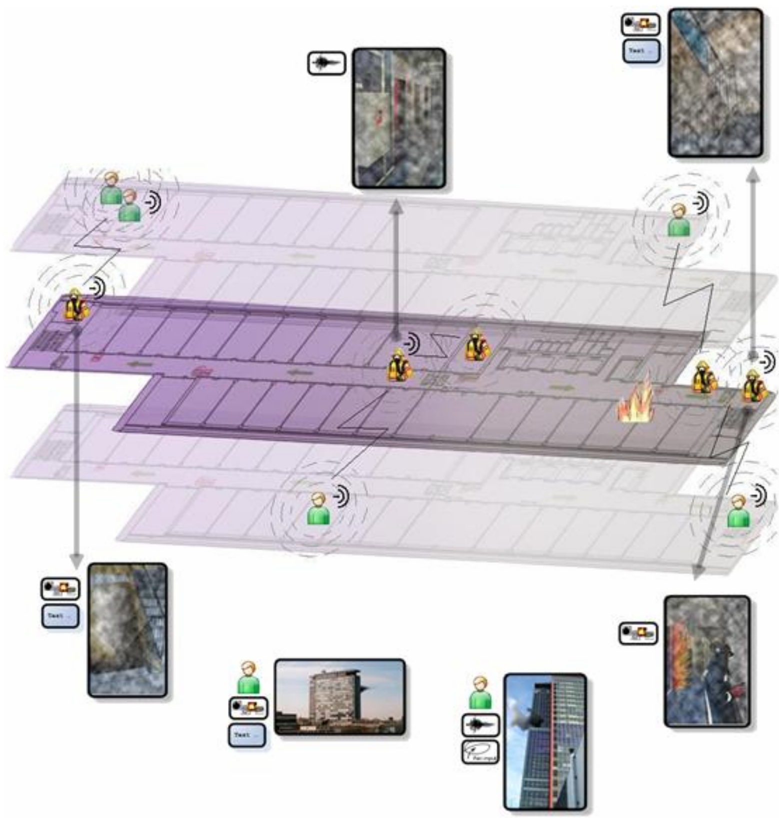

Fig. 5. The individual observations of the multimodal adaptive system. Multi-hop networks among mobile devices allow for data exchange in situations when the connection to the central system is seriously affected.

In fig. 4 the user-oriented multimodal information regarding an example of a crisis event is presented in a timeline manner. The system automatically synchronizes the user reports between different modalities. The meaning of a user observation is generated by extracting and correlating the atomic informational clues related to the crisis event from each user informational stream. The illustration shows an example of the original input of a user, split in separate multimodal information flows.

The disambiguation model incorporates the fusion and filtering of the user reports stored using a common representation format.

Fig. 5 illustrates an example for a schematic representation of the users of the adaptive crisis system on the map in the crisis context. Some of the regular as well as special users provided snapshots of the event taken from the locations in the crisis scene. The users are characterized through the types of inputs they use while working with the adaptive system interface for generating the desired observation reports. From the picture, it can be noticed that the users which work close to the fire site have direct wireless access to the communication infrastructure. As for the others, as it is the case of the one fireman on the right side of the picture, they can still access the resources of the system and to send their own reports by using adhoc networks created at the spot.

4 Results

In the experiment we conducted, we measured the usability performance for the human computer interaction as perceived by the participating users and the adaptability of the system to a crisis context. We tested the functionality of the system in the following way. The experiment took place during a training session of real fire brigade in the faculty building. Students followed the firemen in their exercise and acted according to the real fire scenarios. During the session there was a simulated fire in the building. We provided volunteer students playing firemen roles with PDAs' and they were supposed to enter the building searching for victims and fire. They reported about their findings using different modalities: pen input, speech and text messages. Additionally, the existent network of cameras in the building and the PDA's were wirelessly connected to our system.

The experts subjectively made assessments on the results of the reasoning system and set the final adjustments to these data. After this session the firemen had to fill in a questionnaire. They were positive about the possibilities to use different modalities as for example in the case where they reported the location of a fire event using pen with additional spoken comments. Following the analysis of the questionnaire, it resulted that the students were able to use the iconic interface. In some cases they could not find the appropriate icons on the interface to report about events/special locations. In those cases students used SMS messages. The preferred modalities were iconic messages and SMS text. During the interaction the system gave advices to the user about the available modalities. In the end it proved that our system was able to handle the information from different devices and different modalities.

The drawback of the system was the unstable wireless connection and the interface which was not properly designed for firemen undergoing specific fire brigade actions. Especially the lighting conditions, the background noise and the improper firemen outfit such as firemen gloves had a negative impact on the quality of the human friendly system interaction. Because of failing technology, on average only text, icons, pictures were available. The interface with styluses is not suitable for being used by firemen in action. Firemen use special gloves and are unable to use a stylus. The use of the speech interface was complicated. Students didn't use close to mouth microphones (unfortunately not available for the experiments). So, the speech signal was corrupted by background noise. In addition to this, the network communication problems and the restricted bandwidth restricted the use of the video streaming that could make possible the run of facial expression recognition on other computers. The communication between users was far from optimal due to the failing communication of the wireless ad hoc network. Many times users couldn't be reached or data got lost. It is expected that the next generation of PDAs will come with specific hardware solutions to these problems.

5 Conclusion

In the current research we have detailed the functionally and the constituent modules of a multimodal system aiming to support the human computer interaction and to provide proper information to help special categories of people in crisis situations.

The novelty of our system consists in the algorithms to provide the human support in different crisis situations. The interface of the system automatically adapts to the conditions in the working environments and to the user preferences and abilities. Subsequently we have conducted a research on the quality of the interaction between humans and the automatic system during an experimental setup. The context of the experiment focused on a simulated crisis situation that assumed the presence of a fire event and various people acting different roles ranging from common observers from outside the site to qualified fire brigade personnel trying to evacuate civilians and to stop the fire.

During the development of the system, a special attention was given to the graphical user interface more exactly to the tools the users can access in order to create certain reports of their own observations on the crisis. The graphical tools supported during the interaction process involved the use of pen input, icon sequences, text and direct snapshots taken with the user mobile device. Moreover, the interface has been enriched with speech recording capabilities to ease the collection of the user input. This modality was preferred in situations when poor illumination or low visibility caused by smoke altered the use of the graphical interface.

The conclusions related to the underlying multimodal framework emphasized the system tolerance to the coincidental communication breakdowns as a positive aspect though the user difficulty at perceiving the optimal solutions in acting in conditions of lack of updated information about the crisis.

The people that played the role of firemen during our simulation also experienced the difficulties induced by the wearing of the real fireman equipment. That was obviously not suitable for a regular activity of interacting with the Zaurus PDA device. Especially the gauntlet that is indispensable while working close to the fire place seemed to be the cause of the problem. Ultimately the experimentation of the automatic multimodal system for crisis situations underlined the superiority of such an automatic approach to help people produce and collect information in reference to the crisis events.

Acknowledgments. The research reported here is part of the Interactive Collaborative Information Systems (ICIS) project, supported by the Dutch Ministry of Economic Affairs, grant nr: BSIK03024.

References

1. Corson, S., Macker, J.: RFC2501: Mobile Ad hoc Networking (MANET): Routing Protocol Performance Issues and Evaluation Considerations (1999)
2. Datcu, D., Zhenke, Y., Rothkrantz, L.J.M.: Multimodal workbench for automatic surveillance applications, In: Multimodal Surveillance: Sensors, Algorithms and Systems Chapter 14 (2007)
3. Delhay, S., Idrissa, M., Lacroix, V.: PARADIS: GIS Tools for Humanitarian Demining. In: The 2nd International ISCRAM Conference, pp. 213–219 (2005)
4. Fitrianie, S., Datcu, D., Rothkrantz, L.J.M.: Constructing Knowledge of the World in Crisis Situations using Visual Language. In: IEEE International Conference on Systems, Man, and Cybernetics (2006)
5. Fuhrmann, S., MacEachren, A.M., Dou, J., Wang, K., Cox, A.: Gesture and Speech-Based Maps to Support Use of GIS for Crisis Management: A User Study. AutoCarto (2005)
6. Granica, K., Nagler, T., Eisl, M.M., Schardt, M., Rott, H.: Satellite Remote Sensing Data for an Alpine Related Disaster Management GIS. In: The 2nd International ISCRAM Conference, pp. 221–232 (2005)
7. Rothkrantz, L.J.M., Datcu, D., Fitrianie, S., Tatomir, B.: Personal Mobile Support for Crisis Management Using Ad-Hoc Networks. In: The 11th International Conference on Human-Computer Interaction, Lawrence Erlbaum Associates, Inc, Mahwah (2005)
8. Tatomir, B., Rothkrantz, L.J.M.: Crisis Management using Mobile ad-hoc Wireless Networks. In: The 2nd International ISCRAM Conference, pp. 147–149 (2005)
9. Gaynor, M., Seltzer, M., Moulton, S., Freedman, J.: A Dynamic, Data-Driven, Decision Support System for Emergency Medical Services, ICCS2005, pp. 703–711. Springer, Heidelberg (2005)
10. Goh, O.S., Ardil, C., Fung, C.C., Wong, K.W., Depickere, A.: A Crisis Communication Network Based on Embodied Conversational Agents System with Mobile Services. International Journal of Information Technology (3) 4, (2006) ISSN 1305-2403
11. Jakobson, G.N., Parameswaran, J., Burford, L., Lewis, P.: Ray: Situation-aware Multi-Agent System for Disaster Relief Operations Management. In: Proc. ISCRAM, pp. 313–324 (2006)
12. Sharma, R., Yeasin, M., Krahnstoever, N., Rauschert, I., Cai, G., Brewer, I., MacEachren, A.M., Sengupta, K.: Speech–Gesture Driven Multimodal Interfaces for Crisis Management. Proceedings of the IEEE 91(9), 1327–1354 (2003)
13. Thalmann, D., Salamin, P., Ott, R., Gutierrez, M., Vexo, F.: Advanced Mixed Reality Technologies for Surveillance and Risk Prevention Applications. In: Levi, A., Savaş, E., Yenigün, H., Balcısoy, S., Saygın, Y. (eds.) ISCIS 2006. LNCS, vol. 4263, pp. 13–23. Springer, Heidelberg (2006)
14. Veelen, J.B., van, Storms, P., van, Aart, C.J.: Effective and Efficient Coordination Strategies for Agile Crisis Response Organizations. In: Proc. ISCRAM 2006 pp. 202–213 (2006)
15. Weigand, H.: Agent Community Support for Crisis-ResponseOrganizations. In: On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops, vol. 4277, pp. 218–226. Springer, Berlin, Heidelberg (2006)