

A Multimodal Human-Computer Interaction Framework for Research into Crisis Management

S. Fitrianie¹, R. Poppe², T.H. Bui², A.G. Chițu¹, D. Datcu¹, R. Dor¹, D.H.W. Hofs²,
P. Wiggers¹, D.J.M. Willems³, M. Poel², L.J.M. Rothkrantz¹, L.G. Vuurpijl³, J. Zwiers²

¹Man-Machine Interaction Group,
Delft University of Technology
{s.fitrianie, a.g.chitu, d.datcu, r.dor, p.wiggers,
l.j.m.rothkrantz}@tudelft.nl

²Human Media Interaction Group,
University of Twente
{poppe, buith, hofs, mpoel, zwiers}
@ewi.utwente.nl

³Nijmegen Institute for Cognition and Information,
Radboud University Nijmegen
{dd.willems, vuurpijl}@nici.ru.nl

ABSTRACT

Unreliable communication networks, chaotic environments and stressful conditions can make communication during crisis events difficult. The current practice in crisis management can be improved by introducing ICT systems in the process. However, much experimentation is needed to determine where and how ICT can aid. Therefore, we propose a framework in which predefined modules can be connected in an *ad hoc* fashion. Such a framework allows for rapid development and evaluation of such ICT systems. The framework offers recognition of various communication modalities including speech, lip movement, facial expression, handwriting and drawing, body gesture, text and visual symbols. It provides mechanisms to fuse these modalities into a context dependent interpretation of the current situation and generate appropriate the multimodal information responses. The proposed toolbox can be used as part of a disaster and rescue simulation. We propose evaluation methods, and focus on the technological aspects of our framework.

Keywords

Multimodal systems, human-computer interaction, communication system, disaster and rescue simulation.

INTRODUCTION

Crisis response and management involve the collaboration of many people. To perform and coordinate their activities, they must rely on detailed and accurate information about the crisis event, the environment and many more factors. However, it is difficult to construct such globally consistent views of the crisis event for two reasons. First, the dynamic setting of such events is constantly changing. Second, different people have different tasks and roles to fulfill. Therefore, the information is likely to be distributed piecemeal across geographically distant locations. Moreover, the complexity of the crisis management organization in general makes it difficult and timely to collaborate and verify the obtained information.

The lack of overview is not the only limiting factor in adequate and timely response to crisis situations. Acquisition of detailed and accurate information about the crisis situation is of key importance. Such information can be collected from a variety sources including observers, rescue workers and sensors. However, analysis of past disasters, such as 9/11 and hurricanes Katrina and Rita by Moore (2006), points to communication as a limiting factor in disaster response. Current approaches to coordination of rescue and response activities suffer from the problem that information is neither current nor accurate. This can partly be explained by the nature of crisis situations. The intense nature of crisis situations is believed to result in short term memory loss, confusion, difficulties in setting priorities and making decisions (Farberow and Frederik, 1978). These result in fragmented and badly structured communication or even leaving out some relevant data.

The introduction of novel information and communication technology (ICT) in the crisis management domain can help to provide more detailed and accurate situation overviews that are current and shared amongst all management

levels. Since, in major disasters, communication infrastructure breakdowns are inevitable, communication between actors in the field and actors in the crisis control centers can be improved by introducing recent ICT developments such as wireless communication (Moore, 2006). However, there is a huge gap between the current situation in crisis management organization and the possibilities that arise from the use of ICT.

The research reported here focuses on investigating which ICT methodologies can improve information acquisition and exchange during crisis respond and rescue activities. We believe that accurate and easy access of information can support better decision, planning and reasoning in crisis situation, and situation awareness of the actors. In our approach, we use selective introduction of novel ICT within certain parts of the crisis management organization. For this purpose, a flexible test (research) environment is necessary. This would allow us to change, add or remove devices, modalities, roles and functionalities in a convenient way, without having to spend a lot of time on the reconfiguration of the test environment.

In this paper, we present such a framework for research into the use of novel ICT within crisis management. We discuss the framework from a technology point of view. We present the proposed architecture and the different modules that are currently being developed. We focus on the communication between different actors via different devices. Each of these modules regards a different modality (text, speech, visual language, gesture, pen input and face recognition), which allows us to determine which of the natural human communication channels are most appropriate for a given situation. We discuss the multimodal integration of inputs and outputs. Further, we elaborate our plans to apply the framework in crisis situations and conclude the paper with a discussion on future work.

RELATED WORK

In the years after September 11, 2001, efforts to leverage technology in crisis response and management emphasize the development of more sophisticated planning and response techniques. ICT plays a critical role to improve the crisis response and management. Some attempts have been done in exploiting Internet to provide a platform for information access, communication and collaboration. The *RESCUE* project with their testbed *CAMAS* (Mehrotra et al., 2004), allows users to send reports via a web interface using natural language messages. This system is able to parse and analyze users' input, classify crisis events and create situation awareness. The *VCMC* model also employs a web interface (Otten et al., 2004). It allows its users to share data about crisis situations and to discuss information in real-time. An icon-based interface on handheld devices for reporting observations has been developed in (Tatomir and Rothkrantz, 2006). The system allows its users to share and merge icon-based topological maps in damaged buildings using observations from individuals. The systems reported above deal mainly with the communication with the user and focus on users in the field. In contrast, (Sharma et al., 2003) developed a multimodal framework to facilitate decision making in control rooms. It employs input processing of natural gestures and speech commands for managing dynamic emergency scenarios on a large display. The implementation supports collaborative tasks among people present at remote sites with different computing platforms, communication devices and network connections.

The framework that is described in the remainder of this paper is not a particular system but a platform that allows the rapid construction and evaluation of multimodal human-computer interaction (HCI) systems. The framework itself, the modules and the communication infrastructure are described in the next two sections.

COMPUTATIONAL HUMAN INTERACTION MODELING: FRAMEWORK

We propose a framework in which modules are connected in an *ad hoc* fashion (see next section). We aim at component integration that is independent of the availability of modalities. Therefore, there is no need to recompile or re-link the entire system when updating with the availability of a new module.

The modules deal with HCI, the interpretation of the user's actions, the generation of appropriate responses and the presentation of these messages. Each of the modules is able to work in real-time. Furthermore, the input and fusion modules are designed to cope with noisy measurements. In the dynamic setting of a crisis event, during information retrieval, the crisis response and management teams must deal with several sources of uncertainty. Spontaneous speech input under stress, for example, adds problems such as additional nonverbal sounds, fragmented utterances and implicit references on top of the typical speech recognition uncertainties. As a result, descriptions communicated that way may turn out to be ambiguous or even irrelevant. Although this may be true for each modality on its own, we believe that utilizing different modalities to form a coherent picture of the situation at hand might be the right way to go to reduce both ambiguity and incompleteness. From such a view point, rather than

adding complexity to the scene, the added (and possibly redundant) information from any modality is seen as enhancing and complementing data from other modalities.

The variety of modules allows us to apply the framework to support various roles within the crisis management, including rescue workers, civilians and control room operators. A schematic overview of the system that supports multiple users is shown in Figure 1. Incoming reports are integrated by a Fusion Manager that creates a new, up-to-date global world model that is then sent back to the network and shared with the users.



Figure 1 The developed communication system: multimodal, multi-devices and multi-users

Figure 2 shows our proposed architecture of a communication system for a single-user system. The multimodal input for each user is combined and interpreted in the fusion component. This process includes interpretation of the user's affective state from facial expression and linguistics content. The fusion here is local and the world model is referenced from the dialogue action manager (DAM). The DAM generates appropriate responses and the fusion component displays these using synchronized modalities. The different components in our developed framework architecture are explained below.

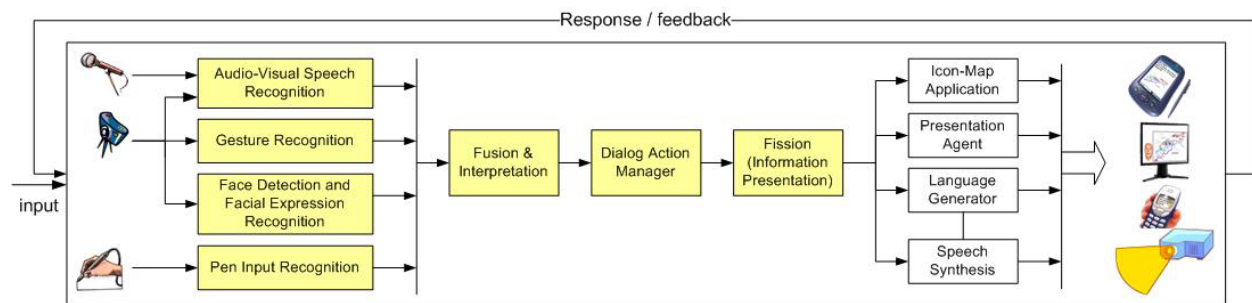


Figure 2 The developed framework architecture for single user

Input Analysis

Audio-Visual Speech

To cope with significant background noise in crisis situations (explosions, police and fire brigade cars), our architecture includes an audio-visual speech recognition module (Figure 3(a)). In our approach, features are extracted from the visual and audio modalities and combined into a common feature vector which is then fed into a set of HMMs. The number of visemes is much lower than the number of phonemes present in a given language. To account for this we tied the states of the HMM for these phonemes for the visual stream. We use the Mel-Frequency Cepstrum Coefficients (MFCC) (David and Mermelstein, 1980) to parameterize the speech data.

Our module uses two approaches for estimating the visual features: (1) estimating the shape of the mouth (lip contours and thickness) in time and (2) capturing the actual movement of the mouth using an optical flow based algorithm (Lucas and Canade, 1981). On the source frame (Figure 3(b)) are superimposed the lip area in red, the cavity of the mouth not obscured by tongue and teeth in green, and the teeth area in blue (the blue margins are highlighting the mouth region). Our current research is aimed at assessing the performance of our system, both for clean audio-visual recordings, and for situations where there is significant noise in both the audio and visual channel. A special case which we are to investigate is the performance of our system on emotional data.

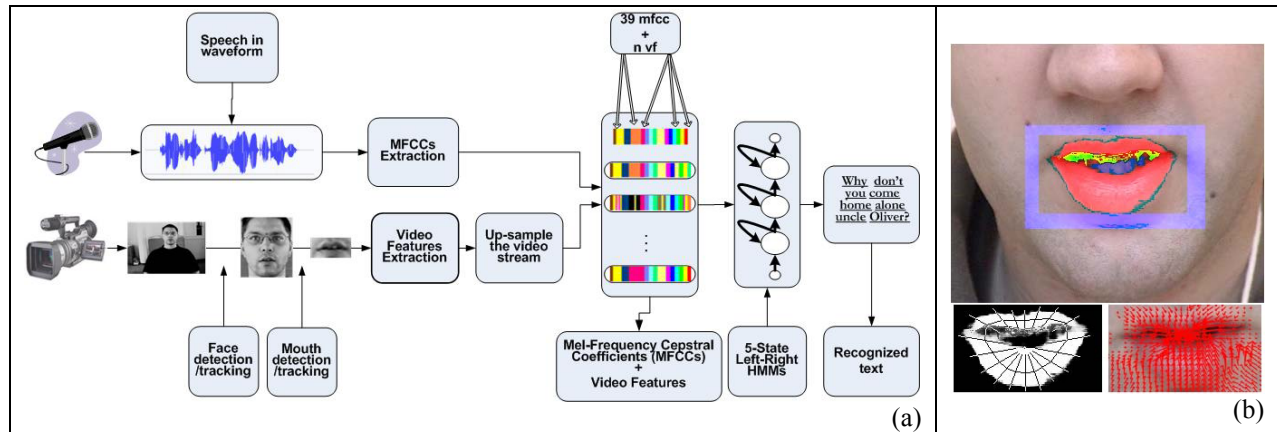


Figure 3 (a) Audio-visual speech recognition module architecture and (b) visual features extraction (left below: mouth shape estimation, right below: optical flow analysis)

Gesture Recognition

With a combination of speech and pointing gestures on large screens, speech-gesture interfaces are especially useful in control room situations (Sharma et al., 2003). Much research with reasonable results has been obtained in restricted domains (Oviatt, 2003). Currently, we are focusing on pointing gestures only. Since we consider an indoor environment, we can assume semi-static backgrounds and controlled lighting conditions. We take an example-based pose estimation approach, where each camera frame is compared to a number of examples in a database (Poppe and Poel, 2006). Such an approach is convenient since initialization of the pose is performed automatically. Moreover, example-based approaches have the potential to recover poses in real-time. Open questions are the proper encoding of variations in image appearance in the database, the employment of a context-dependent tracking algorithm and the handling of partial occlusion.

Face Detection and Facial Expression Recognition

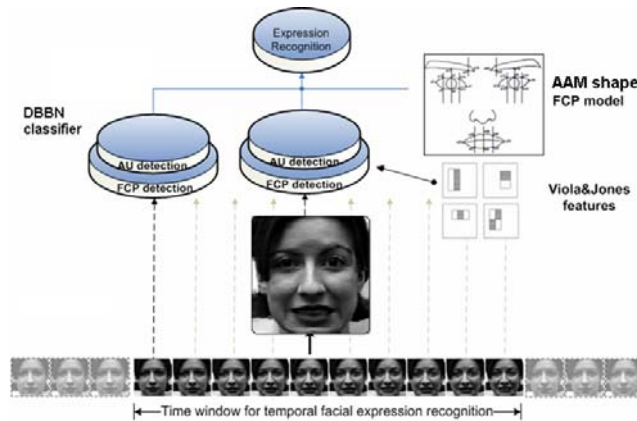


Figure 4 Time window oriented multi layer facial expression recognition model for video sequences

We aim at recognizing human emotions from video sequence data using various models (Datcu and Rothkrantz, 2005; Wong et al., 2006). In the current approach, we extract parametric information from face space by considering temporal patterns of emotions during the transition from neutral emotional state to the apex of each distinct emotion (Figure 4). Viola&Jones features and AdaBoost classifier (Viola and Jones, 2001) are used for face detection. The Active Appearance Model (Cootes et al., 1998) extracts information related to the shape and texture of each detected face and collects them as a set of Facial Characteristic Points (FCPs). We employ Actions Units (AUs) from the Facial Action Coding System (FACS - Ekman and Friesen, 1975) to model six facial expressions (such as sad, happy, disgusted, angry, surprise and fear). Each AU corresponds to the temporal variances of distances between FCPs. We utilize a Dynamic Bayesian Belief Network classifier that uses the activation of AUs to recognize facial expressions. The data set is collected from selection of the Cohn-Kanade Database (Cohn and Kanade, 2000).

Different recorded scenarios are used to emphasize the influence of the variance of light, occlusion and rotation. The true positive rates of six facial expressions recognition using an SVM classifier are above 80% (Datcu and Rothkrantz, 2007).

Pen Input Recognition

Being able to convey spatial information reliably is important for a crisis management system. An effective way for communicating spatial information is by using a pen interface, for example to annotate maps by specifying important geographical information or by sketching actual situations or events. This type of interface can enhance the efficiency of communication (Cohen et al., 1997; Oviatt, 2003). However, automatic recognition of pen-based input still poses many problems (Willems et al., 2005), especially when users are unconstrained in the gesture repertoire that they can use. An experiment has been performed in which participants had to annotate maps and photographs in crisis management scenarios (Figure 5). The resulting recognition systems had a performance of 90.7% (Willems and Vuurpijl, 2006).

To improve the performance, we incorporate domain knowledge (using GIS data) and user task types (indicating expected user inputs) in the decision process. For example, if the trajectory of a pen gesture follows the street pattern on a map, the gesture will most likely specify a route. In our current experiments, the task type is predefined and not yet dynamically provided by the DAM. Monolithic feature classifiers are utilized for the pen input recognition. The classification result is combined with the context information in a Bayesian network (Jensen, 2001). By adding domain specific context knowledge, we expect to have a more robust pen interface.

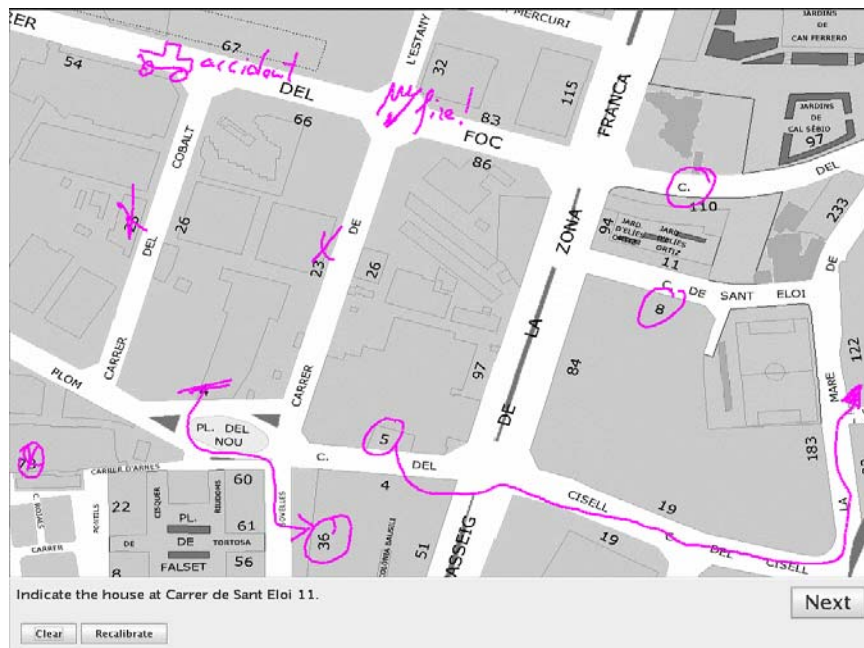


Figure 5 Pen gestures generated by participants. Different modes of pen input are represented here, such as objects (cars), routing (arrows) and marking gestures (crosses and encirclements), and handwritten text.

Contextual Interpretation

Multimodal fusion is meant to provide the mechanism of building coherent semantic structures from dynamic concepts. Although a collection of such concepts may be ambiguous and fragmented, it is, by its very nature contextually and temporally correlated. These characteristics, together with the advantageous multimodal redundancy may be used for coming up with coherent, context dependent interpretation of the communicated situation

The semantic fusion module consists of a parser which builds concepts from various preprocessed input modalities on a workspace. This input is supplied by all the available input modules. We have chosen to use a probabilistic set of parsers offering the advantages of coping with uncertainties at the input, as well as the ability to automatically

learn from (annotated) data. In tandem with the parser, an emergent self-organizing mechanism is designed to find coherent structures from activated world model concepts and parser input. Concepts from the list of existing nodes in Slipnet, an associative network of active concepts akin to that of ‘Copycat’ (Mitchell, 1993) and ‘The Ear’s Mind’ (Dor, 2007), may become activated by their instances on the workspace or by the activation of nearby connected Slipnet nodes. Activation may in turn influence the building of structures among concepts on the workspace, trigger DAM for additional inputs, supply context for the parser, and finally, when enough structure coherence produce a representation of the current interpretation of the fused modalities.

In addition to building concepts on the workspace, the module attempts to relate existing structure with both world model slots such as location, event, and objects, and user’s intention, such as question, directive, statement, etc. These are handed over to the DAM and may assist it in forming feedback to the user. Other modules (for example fission modules and the Fusion Manager) may consult each of the local fusion instances and access their respective data representations.

The centralized Fusion Manager processes every newly reported situation from all users and adapts the global world model accordingly. The world model is defined as two geo-referenced layers that overlay one another (Fitriani et al., 2006). The first layer represents the dynamic context, where a chain of temporal specific events and a group of dynamic objects in action at certain location in the world. The second layer is the static context that represents the geographical information about the crisis location. The location of objects in the user’s environment or the user’s location itself at a certain time-range will be used to capture the information to develop a crisis event world model. A graph of the resulted aggregated world model is shared to all users in network.

Response Selection

In our proposed architecture, the multimodal DAM interprets semantic user inputs through the fusion and input modules and selects appropriate actions to send to the user through fission modules (Bui et al., 2007). Bui (2006) indicated that in a normal environment, input recognition and interpretation errors can occur at all levels of the input processing modules. In the crisis management context, the situation is worse due to stressful conditions. Therefore, the DAM needs to infer the users' actions and affective states from the evidences provided by the fusion modules to select the appropriate actions.

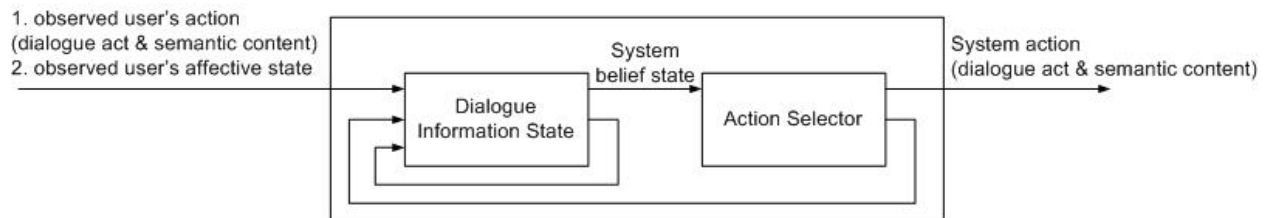


Figure 6 DAM abstract's view for a single-user

The current version of the DAM for a single user (Figure 6) is implemented based on the Partially Observable Markov Decision Process (POMDP) technique. This technique allows for realistic modeling the user's affective state, intentions and hidden states by incorporating them into the state space. Concretely, the DAM processes two semantic inputs: the observed user's action (for example the user's dialogue act and semantic content) and the observed user's affective state. These inputs are sent to the Dialogue Information State (DIS) which maintains the system’s belief state about the user. The belief state is constructed and updated based on a complete trace of all system-user interaction happened so far. The DIS computes a new belief state from the previous belief state, action, and current observation. The computation result is then processed by the Action Selector for the system action selection. The selected action is sent to the user through fission agents.

Information Presentation

The fission module supports information presentation for users present at remotes sites. The challenge is to deliver the same content of information and provide the same services using different sets of modalities for different communication devices. The module uses templates that represent a meta-model for semantic representation of the system’s output based on the system’s action from the DAM (Figure 7(a)) for a specific user or multiple users. The template selection takes into account the urgency of the user’s tasks and the condition of the environment. A parser substitutes each variable in the selected template using values from the current state of the world and the DAM’s

formulated semantic contents. For a verbal output generation, it uses information about the user's emotional state to generate appropriate messages. A modality-based conversion provides methods to present information using selected available modalities. This includes synchronization and scheduling of display presentations with speech.

We designed a map interface for supporting people with geospatial information (Figure 7(b)) (Fitrianie et al., 2006). The interface provides icons, geometrical features, text and photos for describing a crisis situation. The icons represent objects and events in the world (for example explosions, ambulances, firemen). This representation is chosen to support faster interaction, to reduce the ambiguity of the presented information and to provide a language independent message. The geometrical features, such as arrows, ellipses and rectangles, can be used to represent an area or for emphasizing an object, event or location on the map. Text and photos taken with a camera are used to present non-spatial information. A user test has been performed to determine whether users are able to represent ideas and concepts with an arrangement of icons. Experimental results showed that the iconic interface can serve as a communication means.

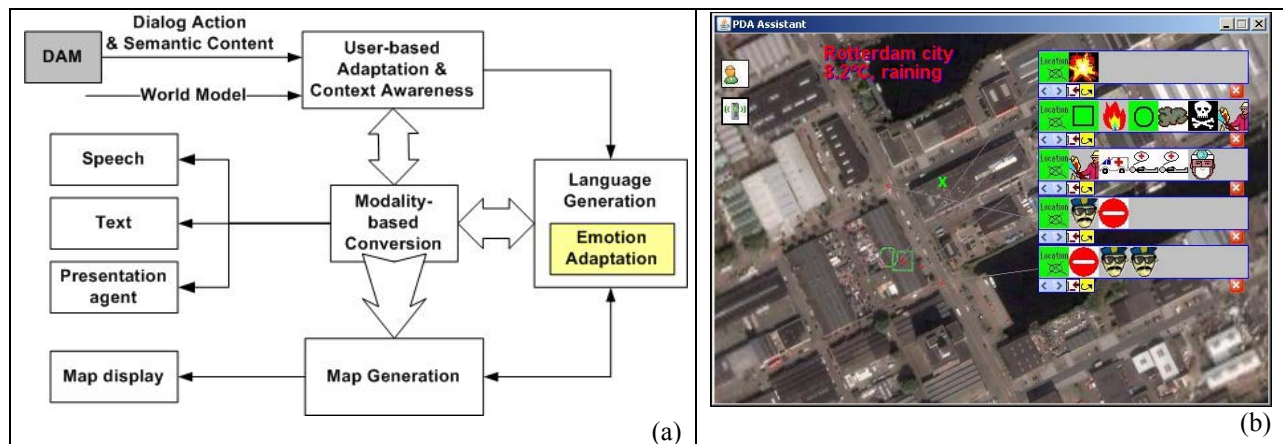


Figure 7 (a) The schematic architecture of the output fission and (b) the developed map interface

COMMUNICATION INFRASTRUCTURE

All modules are integrated using the iROS middleware system (Johanson, 2002). The modules communicate with each other through a common *event heap*. Modules send events containing XML messages to the heap and other modules can subscribe to receive certain types of messages from the heap. One of the main advantages of this communication infrastructure is the fact that modules can be connected in an *ad hoc* fashion. On top of the iROS system we have built a platform with tools to facilitate the development and integration of modules (Figure 8). The platform allows us to view all messages sent to the heap and send messages manually at run-time.

As the system is designed to deal with multiple users with a range of different devices and network connections, a special manager module is included to enhance the robustness and flexibility of the core iROS system. In a multi-user system, the same module can exist in different instances for different users. Messages can be addressed to a specific user or be broadcasted to all users. The manager can grant unique user IDs to address specific users. A user device or module can temporarily become unavailable, for example when an unreliable (wireless) network link goes down. The manager adds an error recovery layer to iROS, which distinguishes essential messages and streaming messages. The latter ones are only relevant at a specific time and lose importance as soon as a new message is produced in the stream. They are therefore not subject to error recovery. However, the manager does ensure that essential messages are delivered to all intended recipients.

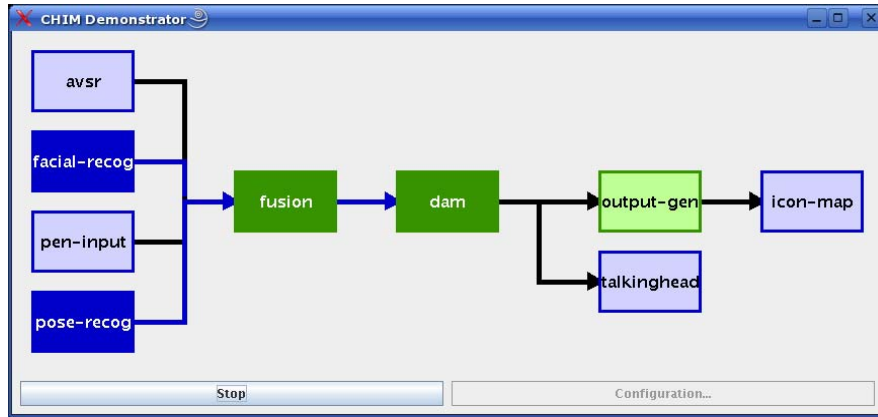


Figure 8 Architecture visualization in the testbed application

APPLYING THE FRAMEWORK: EVALUATION

Our framework offers a rapid development environment to create ICT systems that perform specific tasks within the crisis management process. The evaluation of such framework can be performed in two directions: (1) to evaluate the performance of a system built based on the framework and (2) to assess the added value of such a system within crisis management. In the current stage of research, we focus on the first direction.

We plan to evaluate our framework, with respect to the first direction, in two different methods. The first method is within a simulation environment. Modeling and simulation plays an important role in testing a new technology in disaster setting (Robinson and Brown, 2005). Simulator software can be applied to a crisis context, to provide a virtual simulation environment for research or an interactive method for training and scenario testing in the field of emergency response. One example is to have a system that aid rescue workers to indicate the location of victims, construct maps of a damaged building and communicate directly with fellow rescue workers. Such approach allows us to fine tune the complexity and performance modules and such systems. However, a simulation is limited in realism, and participants in simulations are likely to behave differently compared to real crisis situations. Therefore, as the second method, we propose to evaluate our systems in real crisis exercises to capture more realistic problems and requirements.

The evaluation of such framework, with respect to the second direction, will allow us to determine how people use the systems developed, and to assess whether they improve the efficiency of a given task. Using simulator software, the evaluation can aid in training the people to work with the systems. While in real crisis exercise settings, it allows us to see how people that experience stress use the systems, and how the application of the system aids or interferes with the (traditional) crisis management process.

CONCLUSION AND DISCUSSION

The current practice in crisis management can be improved by introducing ICT systems in the process. However, there is a huge gap between the current practice in crisis management and a situation that fully depends on ICT. We expect that a careful introduction of ICT in specific parts of the crisis management organization will improve efficiency and performance. To determine which tasks can be facilitated and improved, a lot of experimenting is needed. Therefore, we propose a framework that consists of a set of modules that can be combined in an *ad hoc* fashion to form a multimodal system. Such a system, we believe, can aid in crisis management. We discussed how to evaluate the suitability of the system within a current crisis management environment.

The research reported here focuses on investigating HCI technology for coping with the dynamic nature of crisis environments together with a context dependent interpretation of input, relevant response selection and appropriate response generation. However, these do not address the full complexity of decision making, management and coordination in such settings. We believe that it is still too soon to rely on decision made by automated systems. Therefore, our view about an ICT system is an add-on mean that can improve the performance of crisis management organizations. Hence, the control should be still in human hands. To support this, our proposal includes direct feedback to user inputs, allowing for verifying and altering information and ways for collaborating information. In

our view, by collaborating information, obtained data still can be verified and updates by multiple intelligent processes.

Crisis situations create complex environments for any crisis management system. Unreliable communication network and noisy (chaotic) environments make it signal processing hard to achieve robust recognition of human signals, which can hamper the contextual interpretation of the input. On the other hand, users of such a system (that takes into account these signals) rely on accurate and relevant information necessary for their tasks. Moreover, stressful conditions and mobile activities make it difficult to interact with the system and perceive the provided information. This imposes stringent demands on usability of both input and output interfaces. Regarding these issues, the focus of our current activities is on the assessment of the performance of the individual modules within our developed framework and the performance of the first, limited, experiments with systems that are developed using the framework.

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. Bui T.H., *Multimodal Dialogue Management - State of the Art*. CTIT Technical Report series No. 06-01, University of Twente (UT), Enschede, The Netherlands, 2006.
2. Bui T.H., Poel M., Nijholt A., and Zwiers J., A Tractable DDN-POMDP Approach to Affective Dialogue Modeling for General Probabilistic Frame-based Dialogue Systems, *IJCAI'07*, India, 2007.
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. *ACM MM'97*, ACM Press New York, 31-40, 1997.
4. Cootes T.F., Edwards G.J. and Taylor C.J., Active Appearance Models, European Conf. Computer Vision, 484-498, 1998.
5. Datcu D and Rothkrantz L.J.M., Facial Expression Recognition with Relevance Vector Machines, *IEEE ICME '05*, the Netherlands, 2005.
6. Dor R., *The Ear's Mind, a Computer Model of the Fundamental Mechanisms of the Perception of Sound*, technical report 05-16, 2005.
7. Davis S. and Mermelstein P., Comparison of Parametric Representations for Monosyllabic Word Recognition, *IEEE ASSP*, 28:357-366, 1980.
8. Farberow N.L. and Frederick C.J. *Training Manual for Human Service Workers in Major Disasters*. Rockville, Maryland: National Institute of Mental Health, 1978.
9. Fitriani S., Datcu D. and Rothkrantz L.J.M., Constructing Knowledge of the World in Crisis Situations using Visual Language, *IEEE SMC'06*, 121-126, 2006.
10. Jensen F.V., *Bayesian Networks and Decision Trees*. Springer Verlag, New York, Inc., USA, 2001.
11. Johanson B., Fox A., and Winograd T., The Interactive Workspaces Project: Experiences with Ubiquitous Computing Rooms, *IEEE Pervasive Computing*. 1 (2): 67-74, 2002.
12. Lucas B.D. and Kanade T., An Iterative Image Registration Technique with an Application to Stereo Vision, *IJCAI'81*, 674-679, 1981.
13. Mehrotra S., Butts C., Kalashnikov D., Venkatasubramanian N., Altintas K., Hariharan P., Lee H., Ma Y., Myers A., Wickramasuriya J., Eguchi R., and Huyck C., Camas: a Citizen Awareness System for Crisis Mitigation. *ACM SIGMOD'04*, 955-956, USA, 2004.
14. Mitchell M., *Analogy-Making as Perception, A computer Model*, MIT Press, 1993.
15. Moore L.K., *CRS Report for Congress: Public Safety Communication Policy*, Confessional Research Service, the Library of Congress, 2006.
16. Otten J., van Heijningen B., and Lafortune J.F., The Virtual Crisis Management center - An ICT implementation to canalize information!, *ISCRAM'04*, Belgium, 2004.

17. Oviatt S.L. Multimodal interfaces. In *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*, J. JACKO AND A. SEARS, Eds. Lawrence Erlbaum Assoc., Mahwah, NJ, 286-304, 2003.
18. Poppe R. and Poel M., Comparison of Silhouette Shape Descriptors for Example-based Human Pose Recovery, *IEEE FGR 2006*, 541-546, UK, 2006.
19. Robinson C.D. and Brown D.E., First Responder Information Flow Simulation: a Tool for Technology Assessment, *Winter Simulation Conference*. 919-925, USA, 2005.
20. 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, *IEEE*, 91(9): 1327-1354, 2003
21. Tatomir B. and Rothkrantz L.J.M., Intelligent System for Exploring Dynamic Crisis Environments, *ISCRAM'06*. B. Van de Walle and M. Turoff, eds., USA, May 2006.
22. Viola P. and Jones M., Robust Real-time Object Detection, *2nd Int. Workshop on Statistical and Computational Theories of Vision-Modeling, Learning, Computing, and Sampling*, 2001.
23. Willems D.J.M., Rossignol S. and Vuurpijl L., Mode Detection in Online Pen Drawing and Handwriting Recognition, *IEEE ICDAR'05*, Korea, 31-36, 2005.
24. Willems D.J.M and Vuurpijl L.G., Pen Gestures in Online Map and Photograph Annotation Tasks, *IWFHR*, France, 397-402, 2006.
25. Wong W.S., Chan W., Datcu D. and Rothkrantz L.J.M., Using a Sparse Learning Relevance Vector Machine in Facial Expression Recognition, *Euromedia'2006*, 33-37, Belgium, 2006.