Emotion Sensing for Context Sensitive Interpretation of Crisis Reports

Zhenke Yang
Man-Machine-Interaction Group,
Faculty of Electrical Engineering,
Mathematics and Computer science,
Delft University of Technology
Mekelweg 4 2628CD Delft, The Netherlands
Z.Yang@TUDelft.nl

Leon J. M. Rothkrantz
Man-Machine-Interaction Group,
Faculty of Electrical Engineering,
Mathematics and Computer science,
Delft University of Technology
Mekelweg 4 2628CD Delft, The Netherlands
L.J.M.Rothkrantz@TUDelft.nl

ABSTRACT

The emotional qualities of a report play an important role in the evaluation of eye witness reports in crisis centers. Human operators in the crisis center can use the amount of anxiety and stress detected in a spoken report to rapidly estimate the possible impact and urgency of a report and the appropriate response to the reporter. This paper presents ongoing work in automated multi-modal emotion sensing of crisis reports in order to reduce the cognitive load on human operators. Our approach is based on the work procedures adopted by the crisis response center Rijnmond environmental agency (DCMR) and assumes a spoken dialogue between a reporter and a crisis control center. We use an emotion model based on conceptual graphs that is continually evaluated while the dialogue continues. We show how the model can be applied to interpret crisis report in a fictional toxic gas dispersion scenario.

Keywords

Emotion, scenario scripts, multi-modal, fusion, emergent interpretation.

INTRODUCTION

The crisis response center of the Rijnmond environmental agency (DCMR) in Rotterdam, receives input from sensors distributed in the area of Rotterdam. This includes smoke sensors, sniff sticks and cameras. In addition, human operators in the center are trained to communicate with emergency responders and eye witnesses during crisis situations. They engage in dialogues through the telephone to try to extract information, solve ambiguities, ask for missing information etc. One important aspect in the training is the recognition of the emotional qualities of eye witness reports, as the intrinsic contents of a report might be obscured by the personal emotions of the reporter.

In a crisis situation, the crisis center might suddenly have to deal with many reports simultaneously. One of the potential problems that arise is that of data overload bringing the crisis center to a standstill. Especially in the time-critical environment typical in crisis situations this is a highly undesirable prospect. The goal of this research is to develop a decision support system to help operators in the crisis center to reduce the possibility of cognitive overload. A good estimation of urgency is important to help the operator decide which report to deal with first, allowing them to focus on the calls that provide the information necessary to get better awareness of what is happening. In crisis situations, it is reasonable to expect that reports from eyewitnesses (people who are really in the crisis environment) will contain a lot of emotions. Combined with the contents of the report, this may give an indication of the informational value and urgency of the problem of the reporter.

In this paper we assume that reports consist of dialogues of spoken sentences. Each sentence is evaluated in sequential order. We analyze the extracted features (e.g. uttered word, audio qualities) to estimate the affect qualities of the caller. Next we interpret the actual contents of the report using a crisis knowledge base and the affect values. Finally we map the results to a number of predefined crisis scenarios, to find the most plausible scenario. Our approach is based on the observation that operators in a crisis center judge crisis reports by imagining the most probable situation based on the inputs so far. They do this with the aid of a list of frequently occurring reports (prototype reports) and commonsense relationships between events (either on paper or in their head) all the while taking into account the emotional state of the reporter.

The remainder of the paper is structured as follows. First we give an overview of the background and the related work in the area, and then we describe our model and give an indication of how the model works using a fictional scenario. We end up with a discussion and some conclusion.

BACKGROUND AND RELATED WORK

The crisis response center

In a previous study (Benjamins, 2000) analyzed the crisis response protocols and procedures in the crisis response center Rijnmond environmental agency (DCMR) in Rotterdam. These protocols and procedures are taken as the basis for our paper. In the crisis center, human operators are trained to interact with emergency responders and eye witnesses through the telephone. The operators are trained to process the telephone calls via dialogues in order to solve ambiguities and ask for missing information. Whenever a call comes in, the operator has to try to determine the subject of the call as quickly as possible. Usually, the operator starts with an empty form, which gets filled while more information arrives. In the background expert continually try to interpret and combine the arriving information to make it fit into a coherent whole.

During a crisis, operators are confronted with a flood of calls (Someren, van et. al., 2005), a quick and accurate assessment of the urgency is essential in the time-critical environment typical in crisis situations (Horvitz et. al., 1995). The crisis center has to receive information about a potential crisis, assess its impact, determine the response needed, and activate the crisis management and appropriate action teams. Less urgent calls have to be diverted to other information sources such as websites or television, allowing more time for the more informational and urgent calls. Of the remaining reports, operators employ scripted knowledge (described in more detail in (Shank et. al., 1977)) by using a list of frequently occurring calls and common sense knowledge to determine the most likely situation based on the information received thus far. Based on this information, the operator also has to assess the situation to, for example, determine if the situation is safe enough to send aid and safety workers. Acting too early might lead to inefficient usage of resources, while acting too late might lead to a decreased effectiveness of resources.

The role of emotions

In case of a disaster there may be many calls of people wanting to give information, ask for help, ask for information or just complain. Often emotions are involved in these calls. Moreover, the caller may be influenced by his current emotional state. For example, bystanders tend to report differently (less emotionally) then people who feel threatened or are themselves the victim of the incident they are reporting about (Kelloway et. al. 2004). In addition, the urgency and impact of the report may be conveyed in the emotional qualities of the call. A person who has just witnessed a major bomb explosion has different emotions them a person reporting about a small unexpected explosion as loud as a fire cracker.

One important aspect in the training of operators is to sense emotional qualities of the reports, since the urgency and impact of a report is often conveyed by the emotions of the reporter. Thus, the emotion, combined with the contents of the call, may give an indication of the size of the crisis and the actions to take. Researchers like Picard have dubbed work in the field of affect in human-computer interaction “affective computing” (Picard, 1997). One of the most commonly used emotion categorizations are the Ekman’s six basic emotions for the classification of facial expressions (Ekman, 1993) happiness, surprise, fear, sadness, disgust, and anger.

The subject of emotions in speech has been widely studied by psychologists and psycho-linguist to determine which acoustic features (e.g. the relative loudness and pitch of each sound change, intensity, speaking rate) encode the emotional state of a speaker (Scherer et. al., 1991). For the classification of emotions in speech (especially in the case of crisis situations), the five emotional states happiness, sadness, anger, afraid and neutral (Ververidis and Kotropoulos, 2004; Polzin and Waibel, 1998) are more appropriate.

Multi-modal Fusion

Emotions can be conveyed in different modalities such as text, speech, facial expressions and gestures. Previous studies on emotion in crisis reports have generally taken a single modality into account, with the text modality being the most prominent as audio and video data might not always be available. Techniques for emotion sensing from text include keyword spotting, lexical affinity, statistical methods or hand-crafted models (Fitrianie and Rothkrantz, 2005; Ortony et. al., 1988; Goertzel et. al., 2000; Dyer, 1987). Also semantic approaches using commonsense real-world generic knowledge bases (Liu, 2003) have been applied in previous research. In addition researchers have tried to extract emotions from other modalities such as facial expressions, body posture and gestures, speech and intonation (Datcu et. al. 2006; Pantic et. al, 2000). In the application domain of crisis response, the text modalities (sms) and the audio modality (telephone calls) are most important.

A multi-modal approach can have benefits compared to a single modality approach. For example, robust affect sensing in text can reinforce sensing in the other modalities. In this project, we propose a model for emotional awareness across different modalities. We assume that reports consist of dialogues of spoken sentences. Each sentence is evaluated in sequential order. First the uttered words are recognized and translated into text. With shallow parsing, keywords can be selected and these keywords can be ordered in a semantic graph. Aforementioned emotion sensing techniques for text can further be applied to recognize the emotions. Audio features to estimate the affect qualities of the caller can not be translated to text that easily. One option is to use conceptual graphs (Fitrianie and Rothkrantz, 2005). In this case parameterized values of the audio features are input to the graph. Only by semantic interpretation via human operators or Bayesian networks such input can be translated to concepts.

APPLYING EMOTIONS IN REPORT INTERPRETATION

One aspect many crisis management tasks have in common is time-critical decision making. To increase the efficiency of the crisis center, the emotions need to be separated from the actual content of the report. Emotions can afterwards be used in two locations in the report handling process (Figure 1). First emotions can be used to separate the urgent and intelligible reports from less relevant questions, complaints or unintelligible ones. Second, as we know that emotions have a significant influence on the way situations are reported by callers, the extracted emotions can be used to correctly interpret the remaining reports.

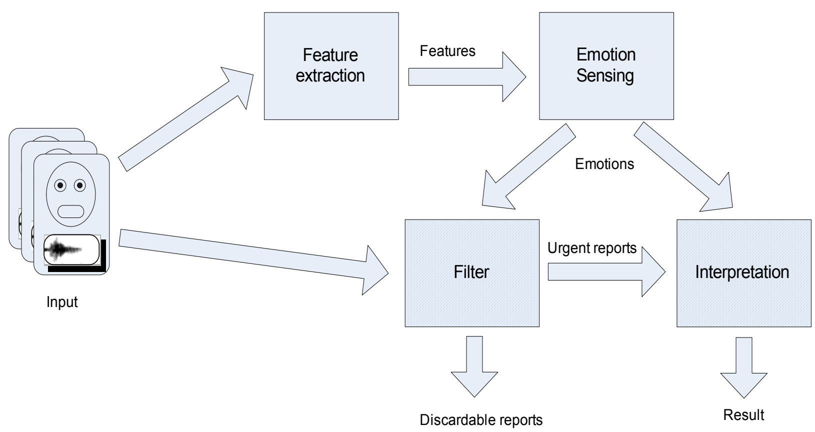

Figure 1: emotions can be used to filter urgent reports and to help in the interpretation of the contents of a report

This paper focuses on the latter use of emotions, namely as a factor in the interpretation of reports. To accomplish this we first design a conceptual graph composed of concepts and variables and relations between them. This graph can be viewed as a general knowledge base (Lenat, 1995) suitable for commonsense reasoning. Apart from common sense knowledge (e.g. fire causes smoke), emotional knowledge (e.g. if 'fire' is said happily, the fire is probably not very big) is also included in the graph. Together with predefined scripts for crisis scenarios, this graph forms the basis of the report interpretation system.

Crisis scenario scripts

To take effective actions during a crisis, it is important to recognize the type and impact of the crisis and the consequences of the chosen actions. Our approach for crisis recognition is based on a number of predefined prototype scenarios and a process that continually tries to recognize the most plausible scenario as more data arrives. This approach is similar to a well established and commonly adopted training methodology in crisis response, in which an instructor selects desired training objectives, and a crisis simulator (a computer program) automatically constructs a scenario, requiring application of the desired skills, based on a series of crisis events (Stern et. al 2002). Subsequently, crisis workers are trained to recognize the type of crisis and to take appropriate actions by (gradually) identifying the crisis events that characterize the scenario (Schank et. al. 1977). We represent a crisis scenario as a chronological ordering of characteristic concepts of the crisis situation (e.g. observations, actions), called a script. In our model we assume a number of scripts each representing a different scenario. At the start of a real disaster there are many possible scripts. The goal of the system is to figure out the script (and thus, the scenario) that is most plausible given the information received thus far from the various sensors and information channels (Figure 2).

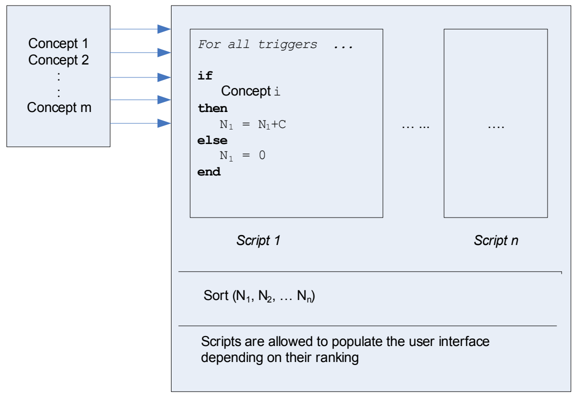

Figure 2: Calculation of the most plausible script

In the period of august 2005 to April 2006, we carried out interviews with crisis experts, firemen and chemical experts from the DCMR. With the information from the interviews we created a list of possible crisis scenarios; we also created a list of named concepts (and their relationships) that plays an important part in the crisis situation (Benjamins, 2006). Next we used these concepts to formulate the crisis scenarios as scripts.

Multi-modal knowledge base

In a real crisis, every incident is unique and every report is different. Obviously, the scripted approach alone is not enough for crisis recognition, because that would mean that every possible expression of an event would have to be taken into account in the scripts. A robust and consistent mapping from report to the concepts used in the scenario scripts requires semantic interpretation of the contents of a report. For this reason we use two knowledge models in our interpretation: the crisis model and the user model.

Crisis model

This model contains crisis-related knowledge. It is a collection of crisis-related concepts, their interrelations, and the inference rules that enable the generation of conclusions. The crisis model enables the system to reason about causal (and other) relationships of events. The crisis model takes into account that:

1. Different names might be used for the same concept or the same name might be used for related but different concepts. In the model, each concept can have a synonym relation (Figure 3) with another concept. Synonyms relations of the crisis model can be extracted from WordNet synsets (Fellbaum, 1998).

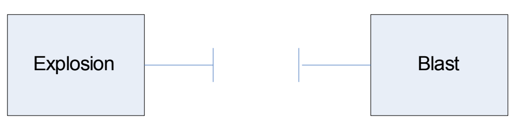

Figure 3: Synonym links

2. Many events in a crisis are causally related; sometimes the effect event of two causally related events is reported but not the cause event (or the other way around). By modeling the causal relationships (Figure 4) in the crisis model, the cause of certain events can be inferred and the effect of certain events can be anticipated.

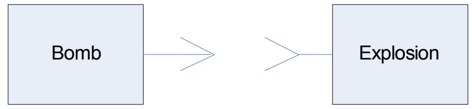

Figure 4: Causal links

3. The same event may be observed using different senses by different observers e.g. “I saw an explosion”, “I heard a bang” or “I felt a tremble”. To relate concept from different modalities, we use association links (Figure 5) to relate concepts across different modalities.

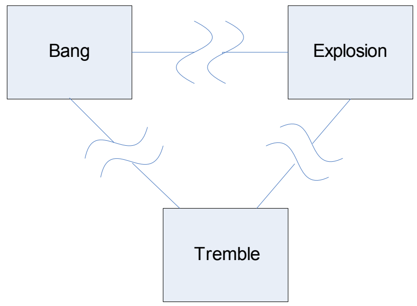

Figure 5: Association links

User model

Context information plays an important role in the interpretation of a report. For example, the utterance “I saw an explosion” causes different reactions depending on the intonation (and various other audio properties) of the utterance. Not to mention, the fact that different emotional states may cause differences in the way an incident is reported. In other words, the emotion might reflect the emotional state of the reporter, but may not be directly related to the content of the report. The user model administrates the knowledge about the user such as the user’s point of view of the situation, his expertise in the application domain, his emotional state. Emotions also give an indication of how strong concepts are related to each other. Therefore we have to determine the effect of an emotion on the related concept. In the crisis model, both fire cracker and bomb can cause an explosion. Depending on the emotions i.e. how the reporter uttered “I heard an explosion”, the probable cause can be determined with higher accuracy (see Figure 6).

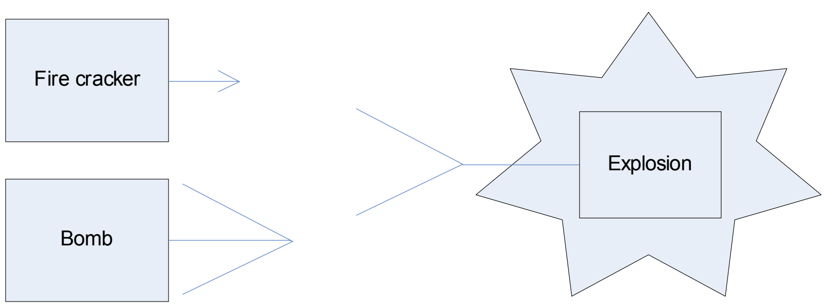

Figure 6: A highly emotional utterance of "I heard an explosion" has a stronger relation with bomb then with fire cracker

CONTEXT SENSITIVE INTERPRETATION

We consider interpretation as an emergent path through a conceptual network of most probable concepts resulting from the set of events received from crisis reports. This structuring is a dynamic process. In the course of time emergent structures can be replaced by other emerging structures. So there is a competition of emergent structures. We assume that the observed features have predefined characteristics which provide them with structure seeking facilities, thus our model can be visualized as a multi dimensional puzzle with pro-active puzzle pieces. The pieces are moving in space looking for partners to set up an emergent structure. The overall strength of an emergent structure is a weighted sum of individual pieces and links.

Example scenario

To illustrate the idea, Table 1 describes a simplified example of one of the DCMR crisis scenarios around the river Maas. In this scenario, something went wrong in a big chemical plant along the river shore and an explosion and toxic gas dispersion requires the area to be evacuated. At the onset of the crisis, the center has five possible crisis scenario scripts: accident in the chemical plant and toxic gas dispersion (1), fire in the traffic tunnel (2), ship on fire (3), house on fire (4) and chimney flames (5). As more reports from the area arrive, it becomes apparent which scenario applies.

Contents of the report	Reasoning in the crisis center
I saw flames near the smoke tower	Based on the synonym relation of “smoke tower” and “Chemical plant” there might be something wrong there (1). The crisis center can deduce that the fire in the tunnel script (2) is not appropriate, since there are no chemical plants near the tunnel. However, there could be a burning ship moving on the river (3) or a house behind the plant might be on fire (4). Furthermore, flames can be seen from the chimneys of the chemical plant regularly (5), so this is a common scenario. Since scenario (5) happens regularly, this is the most plausible. The emotion sensed from the report also does not indicate anything extraordinary about the flames. Flames are caused by fire, so there is a possible causal relationship with fire.
I heard an explosion	Explosion can be related to flames in the previous report through the fire concept. This, in turn, is related to the chemical plant through the smoke tower synonym. At this point, the structure that supports script (5) is inhibited by the stronger structures of scripts (1), (3) and (4).
It stinks here!	There are regular reports of stench in the area. But the intensity with which the sentence is uttered along with the previous explosion report allows script (1) to become the dominant script. The crisis center prepares for a possible evacuation attempt because of toxic gas dispersion.

Table 1: Reports for a scenario and the corresponding reasoning in the crisis center when applying our model

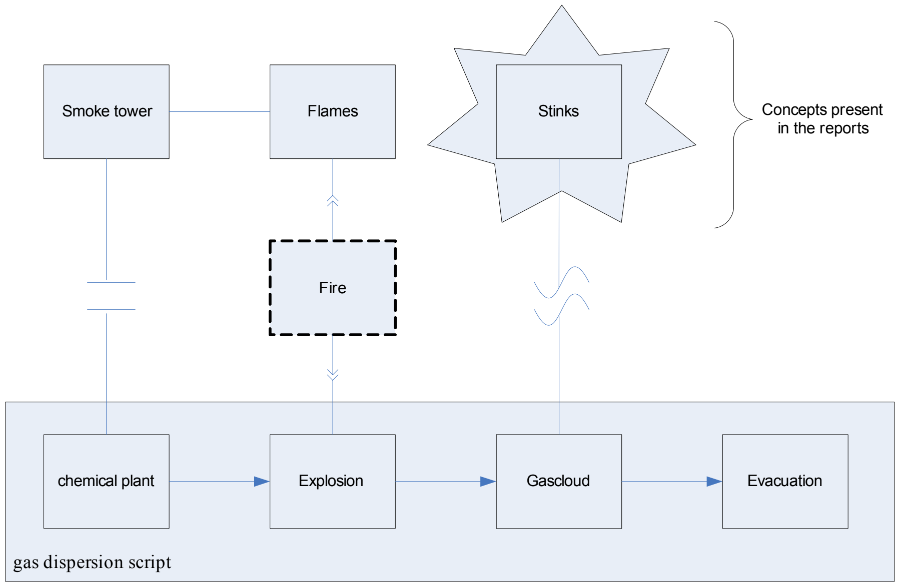

Figure 7: An illustration of how the concepts in the reports fit in the gas dispersion script

At the end, the chemical dispersion scenario emerges as the most plausible scenario, even though no concept from the gas dispersion script was mentioned in the reports (except "explosion"). In addition, most individual reports have a valid non crisis explanation that would have been more plausible was it not for the use of emotions in the interpretation.

DISCUSSION AND CONCLUSION

This paper presents a novel approach to apply affect qualities to context sensitive interpretation of crisis reports. As crisis workers are trained to handle crisis situations by recognizing crisis scenarios by means of their characteristic events and concepts, we use these same characteristic concepts to represent anticipated crises. This representation is called a script. Once a crisis is recognized, the appropriate actions can be taken by the crisis responders. To accommodate the mapping from concepts in reports to the concepts in the scripts, we introduced a crisis model and a user model. The crisis model connects relationships between concepts (such a causal relations and synonyms). The user model administrates the knowledge about the user (e.g. the context, the emotion). To automatically recognize emotional qualities of speech, we can use statistical models that infer the general properties that govern the rise and fall of pitch, as well as the duration and loudness of each person's speech. This in itself is not as trivial as it sounds, as detecting emotions can proof to be rather difficult (Truong et. al., 2007). We showed an example of how the scripts and both models can be applied to interpret crisis reports arriving at the crisis center. The example also showed the importance of emotion in this process. Interpretation of the combined meaning of all the reports is a dynamic process in which structures emerge in the course of the time. As more reports arrive, and more evidence is gathered, these emergent structures can be replaced by other stronger emerging structures. The implementation of the models described in this paper is still in progress, so we have not yet tested the models in experiments of significant size. To automatically find an emergent structure the following procedure, inspired by the clustering procedure of Kohonen networks, can be followed. Given N features, take at random two of them. If they have attractive features reduce the distance between them. If not increase the distance between them. At the end we can expect clusters of attractive/homogeneous features. Operators in a crisis center are trained to handle scenarios that arise in crisis situations. However, if such situations really arise, they may be faced with a sudden flood of information they cannot cope with. Our research aims to prevent this cognitive overload by automating the task of emotion recognition and using these emotions in the interpretation of the reports.

ACKNOWLEDGMENTS

The authors thank the experts of the DCMR for the time for interviews and their valuable feedback.

REFERENCES

1. Benjamins, T.(2006), MACSIM: Multi-Agent Crisis Simulator, Interpreter and Monitor, *Master of Science thesis TU Delft, The Netherlands*
2. Datcu D. and Rothkrantz, L. J. M. (2006) The recognition of emotions from speech using GentleBoost Classifier, CompSysTech'06
3. Dyer, M.G. (1987) Emotions and their Computations: Three Computer Models. Cognition and Emotion, Vol. 1, No. 33, 323-347
4. Ekman, P. (1993), Facial expressions of emotion. *American Psychologist, 48, 384-392*.
5. Scherer, K. R., Banse, R., Wallbott, H. G. and Goldbeck T. (1991) Vocal cues in emotion encoding and decoding, *Motivation & Emotion*, 2(15) , 123-148
6. Ververidis, D. and Kotropoulos, C. (2004) Automatic Speech Classification to Five Emotional States Based On Gender Information, *In Proceedings of the XII European Signal Processing Conference (EUSIPCO)*, Vienna, Austria
7. Polzin, T. S. and Waibel, A. (1998) Detecting Emotions In Speech, In Proceedings of the Cooperative Multimodal Communication (CMC), Tilburg, The Netherlands
8. Fellbaum C., ed. (1998) The WordNet: An Electronic Lexical Database. *MIT Press*.
9. Fitrianie, S. and Rothkrantz, L.J.M. (2005) Communication in Crisis Situation using Icon language, *In Proceedings of the 2005 IEEE International Conference on Multimedia & Expo*
10. Goertzel, B., Silverman, K., Hartley, C. Bugaj, S. and Ross, M. (2000) The Baby Webmind Project. In Proceedings of AISB 2000.
11. Horvitz, E. and Barry, M. (1995) Display of Information for Time-Critical Decision Making, *In Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*
12. Kelloway, E. K., Stinson, V. and MacLean, C. (2004). Eyewitness testimony in occupational accident investigations: Towards a research agenda. Law & Human Behavior, Vol 28(1), pp. 115-132
13. Lenat, D. B. (1995) Cyc: A Large-Scale Investment in Knowledge Infrastructure*, Communications of the ACM Vol 38(11), pp. 33-38*.
14. Liu, H., Lieberman, H. and Selker, T. (2003) A Model of Textual Affect Sensing using Real-World Knowledge, *In Proceedings of the International Conference on Intelligent User Interfaces(IUI'03)*
15. Ortony A., Clore, G. L. and Collins, A. (1988) The cognitive structure of emotions. New York: Cambridge University Press
16. Pantic, M. and Rothkrantz, L.J.M. (2000) Automatic analysis of facial expressions: the state of the art, IEEE Trans. PAMI, 22(12)
17. Picard, R. W. (1997), Affective Computing, *The MIT Press, Mass.*
18. Schank, R. and Abelson, R., (1977) Scripts, Plans, Goals and Understanding, Hillsdale, NJ: Erlbaum.
19. Stern, E. and Sundelius, B.(2002) Crisis Management Europe: An Integrated Regional Research and Training Program International Studies Perspectives 3 (1), 71–88.
20. Someren, van M., Netten, N., Evers, V., Cramer, H., Hoog, de R. and Bruinsma, G. (2005) A Trainable Information Distribution System to Support Crisis Management, *In Proceedings of the Second International ISCRAM Conference*, Brussels, Belgium
21. Truong, K.P., Van Leeuwen, D.A., and Neerincx, M.A. (2007). Unobtrusive multimodal emotion detection in adaptive interfaces: speech and facial expressions. *In Proceedings of the International Conference on Human-Computer Interaction (HCII)*, Beijing, China, to appear