Modality planning for preventing tunnel vision in crisis management

Yujia Cao and **Anton Nijholt** [1]

Abstract. Crisis management is a time-critical task with high uncertainty and high risk. Stress and cognitive overload often result in a set of bias in crisis manager's situation understanding and information confirming processes, known as "tunnel vision". Aiming at preventing tunnel vision, we propose an information assistant system which attempts to reduce the information quantity, improve the information quality, and prevent cognitive overload of the user. The main focus of this paper is to present the design proposal of the modality planning module. It is one of the modules which play a role in the prevention of tunnel vision. The function of this module is to determine the optimum utilization of the available modalities, in order to convey information effectively and reduce the cognitive load of the perceivers. The modality planning strategies also adapt to the user's preferences and cognitive state.

1 INTRODUCTION

A crisis is generally understood as an urgent situation with a negative outcome, such as a nature disaster, transport accident, civil attack, economical crash, etc. Crisis management is a strategic management activity aiming to prevent or minimize the negative impact of a crisis. It is a time-critical task with high uncertainty and high risk [8][11][15]. The crisis managers, who are located in the crisis response center, need to react quickly to the ongoing crisis event and make quick decisions. They also typically have to deal with information overload [15]. Under stress due to information overload and a lack of time, crisis managers tend to rely on standard operating procedures and their previous experiences without reexamination. When an understanding or a solution is forming, they have the tendency to perceive and confirm only clear and familiar information which aligns with it. Correct but ambiguous or contradicting information is easily discarded. We call the above phenomena "tunnel vision". Cognitive psychology theories provide better insight into the tunnel vision phenomena. When the decision makers tend to create one coherent interpretation without reexamining their experience with the real situation, they are experiencing "framing bias" [6] [18]. When they tend to confirm their understanding by seeking only the information which falls in harmony (evidence), they are experiencing "confirmation bias" [10]. Too much information and too little time might also cause cognitive overload [7]. The lack of cognitive capacity might deepen the biases. If an improper decision making "frame" (situation understanding) is continuously confirmed, the growing bias may lead to costly delay and errors.

Considering the specific task of crisis management, the computer has several advantages over the human brain. The computer is able to continuously record data into its memory with high speed, no matter of the quality of the data. It acts only on logic, without any influence by emotions. The computer also exceeds the brain in multi-tasking, fulfilling complex calculation and rule-based tasks. These advantages have high value when we attempt to develop a multimodal information assistant system (referred as "the system" below) which serves as a platform for monitoring the on-going crisis event. Aiming to reduce the likelihood of tunnel vision, the proposed system intends to provide the users with lower quantity but higher quality information in an effective and efficient manner. As one part of the design, this research focuses on the modality planning module. It is one of the modules which contribute to the prevention of tunnel vision. The function is to determine the optimum utilization of the available modalities, in order to convey information effectively and reduce the cognitive load of the perceivers.

Section 2 briefly describes the structure and function of the proposed system with a focus on the modalities related to the prevention of tunnel vision. Section 3 introduces previous work related to modality planning. Section 4 presents the primary research on the design of the modality planning module. The design of the other modules is out of the scope of this paper.

[1] Human Media Interaction Group, University of Twente, PO BOX 217, 7500 AE Enschede, The Netherlands. y.cao@utwente.nl

2 THE INFORMATION ASSISTANT SYSTEM

The general function of the system, as shown in figure 1, is a platform for monitoring a crisis event. The users are crisis managers located in the crisis response center, facing a large display. Briefly speaking, the system continuously captures the real world data (speech, video and sensor signals), records them into its memory, and simultaneously presents the on-going crisis event through the large display and speech. Via an information query interface, the user is allowed to access the crisis history (e.g. events that occurred in the previous minute) or some statistics (e.g. number of victims in area A). The crisis managers don't conduct their commands via the system. However, their commands will also be captured by the system and presented.

Three modules in the system are responsible for the prevention of tunnel vision: the reasoning and filtering module, the order planning module, and the modality planning module. The reasoning and filtering module helps to improve the information quality and reduce the information quantity. It groups reduplicate data together and provides statistic analysis on incompatible information. It also applies context reasoning based on predefined guidelines which are generally valid for certain types of crisis. The order planning module determines the presentation priority for each input information unit based on the overall evaluation of the time sequence, the urgency level, and the causal relations. The aim of this module is to guarantee that the most important and urgent information arrives at the user first. The modality planning module intends to present the crisis scenario effectively and efficiently by calculating the optimum utilization of the available modalities. Effectiveness means that the presentation does convey the information content correctly. Efficiency indicates that the presentation manner helps to prevent cognitive overload of the user. We expect that the users of such a system have larger chance to keep aware of the actual situation, stay open-minded for all possibilities and make proper decisions.

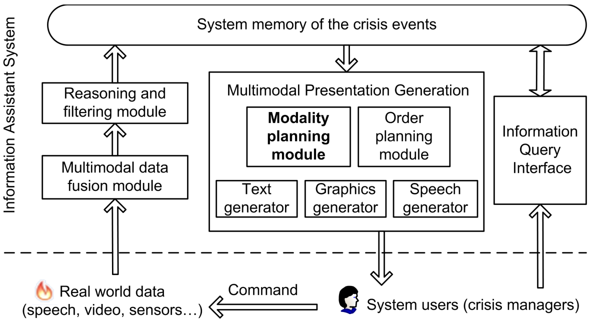

Figure 1. Multimodal information assistant system structure

In the remainder of this paper, only the modality planning module will be discussed in more detail. There are many other on-going researches in the ICIS project (see acknowledgement) that can be applied to the design of the other modules of this system.

3 RELATED WORK

In the modality theory of Bernsen [3], a solid taxonomy of unimodal output modalities is designed to provide a theoretical foundation for understanding and generating multimodal output. Based on the observation that different modalities have different representational properties (such as linguistic, analog, arbitrary etc.), the taxonomy classifies all possible unimodalities into 4 classes in the super level, 20 classes in the generic level, and 46 classes in the atomic level. Possible extensions to a subatomic level have also been suggested. The taxonomy is claimed to be complete, unique, relevant, and intuitive. The different representational properties make different modalities suitable for expressing different types of information. For instance, linguistic modalities (e.g. text, discourse) surpass analogue modalities (e.g. images, graphics, diagrams) at explaining abstract concepts; while analogue modalities are better at expressing what things exactly look like. Their combination may have superior expressive power [2]. The combination of linguistic and analogue modalities has been adopted in many existing multimodal HCI systems, such as COMET [5], WIP [16], SmartKOM [17], and EMBASSI [4].

Concerning the cognitive load a presentation manner may place on the user, designers need to consider not only the representational properties of modalities, but also the perception properties (e.g. visual, auditory and haptic). The perception properties determine how a modality is perceived and processed by the human perceptual-sensory system [1]. The dual coding theory of Paivio [13] claims that humans possess separate information processing channels for visual and auditory material. Therefore, working memory has partially independent processors for handling visual and auditory signals. Mousavi et al. [9] suggest that the mixed use of both modalities can reduce cognitive load, because more effective cognitive capacity is available.

The modality planning task is generally agreed to be highly complex, due to many issues involved. So far, no solution in the form of a generally applicable automated modality planner has been devised yet. Instead, most of the existing approaches focus on a small set of modalities and a certain type of application. The same goes for own research. The modality planning task is commonly considered as a mapping process from the presentation task (convey certain information) domain to the modality domain, based on pre-designed rules or strategies. In the WIP system [16], a set of presentation strategies has been defined for all presentation tasks. They are represented by a name, a header, the applicability conditions and a specification of modality choice. When the presentation planner receives a presentation task, it tries to match a presentation strategy which has the corresponding effect or header. When there are more than one matches, pre-defined meta-rules are applied to make a choice. In the SmartKOM system [17], based on 121 presentation strategies, the presentation planner recursively decomposes a high-level presentation task into primitive presentation tasks and allocates different output modalities to each primitive presentation task. In the EMBASSI system [4], the combination of several unimodalities is defined as a multimodality. The model of a multimodality includes the set of unimodalities, the combination strategy, and the assignment to a physical output device. The combination strategy describes the synchronization, the necessary coordinations for multimodal references to objects, and the possible cross-modal references of the unimodalities. When receiving a presentation task, the presentation planner examines the user preference and the output device condition, and then assigns one or more multimodalities, and constructs the presentation according to the combination strategies.

4 THE MODALITY PLANNING MODULE

The design of the modality planning module aims at achieving the effectiveness and efficiency of the presentation. The modality planning process takes the following factors into account: 1) the information to be conveyed (presentation task) 2) the available modalities 3) the preferences of the crisis manager, and 4) the user's cognitive load status. Currently, we use a tunnel fire crisis scenario. The system represents the scenario by recording the actions of human actors and the state of the world. We have defined a limited set of action types, e.g. request, report, command etc. Modality strategies are designed for each action type, under several different conditions. Therefore, the modality planning approach is to use the strategy which matches the input action type, the user's profile and the user's cognitive statues.

4.1 THE PRESENTATION TASK

The system memory contains a world state database. Based on a world model (ontology), the system creates an instance, in the world state database, for each real-world entity that is involved in the crisis event (tunnel, fires, vehicles, human actors etc.) The properties of the instances (location, urgency level, stress level etc.) may change over time as the crisis event develops. We assume that the world state can only be changed by actions. If an action changes a certain property of a certain entity instance, the system records it as an "Action-StateChange" pair and makes a corresponding modification to the world state database. For example, when the system receives the report from the fire team that the fire has been put off, it creates the following action and state-change instances.

Action-1

- Type: Report

- TimeStamp: 16:45:28
- Actor: FireTeam
- Receiver: CrisisManager
- Content: State-Change-1
- UrgencyLevel: Low

StateChange-1
- TimeStamp: 16:45:28
- Object: Fire
- Property: Status
- Value: OFF

If an action (e.g. request) doesn't directly bring any change to the world state, no StateChange instance will be created. The Action-StateChange pairs or action instances are input into the modality planner as its presentation task. The planner makes the strategy match and applies the corresponding modality allocation and combination schema.

4.2 THE AVAILABLE MODALITIES

The system adopts both visual and auditory modalities. Visual modalities include map, text, image. Auditory modalities include speech and sound effects. Sutcliffe et al. [14] have introduced a set of attention effect advices for directing the user's attention to the appropriate information at the correct level of detail. Following these advices, dynamic text and dynamic image are used when extra attention is needed. Based on Bernsen's modality taxonomy [3], the properties of the available modalities are listed in table 1.

Table 1. The properties of the available modalities (based on [3])

Unimodality	Properties
Static Text	(li,-an,-ar,sta,gra)
Dynamic Text	(li,-an,-ar,dyn,gra)
Map	(-li,an,-ar,sta,gra)
Static Image	(-li,an,-ar,sta,gra)
Dynamic Image	(-li,an,-ar,dyn,gra)
Speech	(li,-an,-ar,dyn,aco)
Effect Sound	(-li,-an,ar,sta/dyn,aco)

"li": linguistic; "an": analogue; "ar": arbitrary; "sta": static; "dyn": dynamic; "gra": graphics; "aco": acoustics; "-": not

In order to specify the detailed utilization of the modalities, modality models are constructed for text, map, image, speech, and alarm sound, respectively. The modality model contains a set of parameters which describes the utilization details of the modality (see table 2). It can be viewed as a template for creating modality instances. Here, we don't separate static use and dynamic use. These properties are described by the parameter value. Therefore, static text and dynamic text share the same modality model. The same goes for static images and dynamic images.

Table 2. The modality Models

Unimodality	Model Parameters
Text	Content, ReferTo, Style, Size, Color, Blink, StartTime, Duration, DisplayArea, ScrollDirection, ScrollSpeed,
Map	Country, Province, City, InvolvedArea, DisplayedArea
Image	Source, ReferTo, DisplayArea, StartTime, Duration, Blink
Speech	Content, ReferTo, Tone, Speed, StartTime, RepeatTime
EffectSound	Source, ReferTo, StartTime, RepeatTime

When fulfilling a specific presentation task, one or more modality instances will be created. Their parameter values also indicate the combination manner. For example, the presentation task is to show the location of the policeman. The modality planner locates an image of a policeman on the map together with a text explanation. The image and text instances are created as follows. The parameter values are filled by the selected modality planning strategy (see section 4.3). The values of "StartTime", "DisplayArea", and "ReferTo" parameters indicate that the two modality instances will be shown at the same time, near to each other, and both refer to the policeman.

Text-1
- Content: In Gate Street, 550M to tunnel
- ReferTo: Policeman.Location
- Style: Arial, bold
- Size: Middle
- Color: Black
- Blink: N
- StartTime: immediate, align with Image-1
- Duration: 30 seconds
- DisplayArea: Rectangle [DisplayCoordination(300,560), DisplayCoordination(450,590)]
- ScrollDirection: N
- ScrollSpeed: N

Image-1
- Source: policeman.jpg
- ReferTo: policeman
- DisplayArea: Rectangle [DisplayCoordination(350,500), DisplayCoordination(400,550)]
- StartTime: immediate, align with Text-1
- Duration: 30 seconds
- Blink: Y

4.3 THE MODALITY PLANNING STRATEGY

In this section, we present the design proposal of the modality planning strategies. The design aims at achieving the presentation goal, i.e. effectiveness and efficiency. The desired presentation manner conveys the information content correctly and helps to prevent cognitive overload. A design proposal is presented in this section. Each presentation task has its own modality planning strategy. A strategy contains three items: 1) suitable modalities, 2) default strategy, and 3) light strategy. The choice between the default strategy and the light strategy is based on the user's cognitive state. When the user's preference is available, an adapted version of the default strategy will be generated.

4.3.1 Suitable Modalities

This item indicates which modalities are suitable for contributing to a certain type of presentation task and what each suitable modality expresses. The values of the "ReferTo" parameter will be filled in. In our crisis management application, the map is always shown as background on the display. However, it will be listed as a suitable modality only when the presentation of a action type needs to make use of it. Recall the example of showing the location of the policeman. Sound effects can do little to show a location. Image and text are selected as suitable modalities. The image refers to the policeman and the text refers to the location of the policeman.

4.3.2 Recommended Strategy and its adaption

The default strategy is designed to achieve the optimum presentation manner for a certain type of task. First, one or more suitable modalities will be selected. Based on the dual-coding theory [13], if the suitable modality list contains both visual modalities and auditory modalities, their combination owns higher priority. Second, the default strategy contains a specification of how to generate modality instances of all the selected modalities. As mentioned before, the parameters of these modality instances also indicate their combination manner. Third, following the attention effect advices in [14], this strategy also attempts to attract the user's attention to what is being presented. For instance, fire alarm (sound effect) is used for a fire report; ambulance alarm is used for a victim report. When necessary, the speech speed is increased with warning tone. Dynamic texts and dynamic images are also often used.

The default strategy will be applied when the user has no specific preference and no cognitive overload of the user is recognized. If the user especially prefers certain modalities, an adapted version of the default strategy will be made. The adaption to cognitive overload will be described in the following subsection. The adaptation to a user's preferences intends to avoid undesired annoyance for the user. If the user prefers a certain modality, it will always be selected, as long as it is on the list of suitable modalities. The user can also indicate that he prefers a less intrusive presentation manner. Then attraction efforts (e.g. using sound effect, warning tone etc.) will be reduced. The user's preferences are set up before using the system, but not during the crisis management process.

4.3.3 Light Strategy

The light strategy is less attracting. It often contains only visual modalities and it will be applied when the system recognizes that the user might be experiencing cognitive overload. When cognitive overload occurs, the user might become slow at responding to the newly-presented information. The system will notice that more requests/reports stay in pending state. For instance, a victim report stays in the pending state until the system hears a command to the doctor, addressing this victim. When possible cognitive overload is detected, the system still continues on presenting new-coming information. However, only a few most urgent tasks are presented with the default strategy, the rest will adopt the light strategy. In this way, the user's attention is drawn to only the most urgent issues. When a light-presented task becomes one of the most urgent tasks, its presentation will be refreshed with the default strategy. When the user's cognitive state recovers, all light-presented tasks will be shown in default strategy.

5 CONCLUSION AND FUTURE WORK

A multimodal information assistant system for crisis management is proposed, aiming at preventing tunnel vision - the phenomena of framing bias and confirmation bias in the crisis manager's cognitive processes. The main focus of this paper is the design of the modality planning module, which intends to achieve the effectiveness and efficiency of the presentation. The complete set of modality planning strategies is still in the design stage. Apart from the existing guidelines in the literature, we expect to find more from empirical studies that will be carried out. In these studies, some crisis scenarios will be presented in a limited amount of time. The presentation will be evaluated on two aspects, i.e. the user's level of understanding and cognitive state. The level of understanding can be estimated by interviewing the user after the presentation. The cognitive state during the presentation can be measured in several ways [12], such as by evaluating the performance of a secondary task (performed concurrently with the primary task) or by measuring physiological variables (e.g. heart activity, brain activity, and eye activity).

ACKNOWLEDGEMENTS

This research is part of the Interactive Collaborative Information System (ICIS) Project (http://www.decis.nl/html/icis.html). ICIS is sponsored by the Dutch Ministry of Economic Affairs, grant nr: BSIK03024.

REFERENCES

[1] Y. Bachvarova, B. van Dijk, and A. Nijholt, 'Towards a unified knowledge-based approach to modality choice', in *Workshop on Multimodal Output Generation (MOG)*, (2007).

[2] N. O. Bernsen, 'Why are analogue graphics and natural language both needed in hci', *Interactive Systems: Design, Specification, and Verification. Focus on Computer Graphics. Springer Verlag*, 235–51, (1995).

[3] N. O. Bernsen, 'Multimodality in language and speech systems-from theory to design support tool', *Multimodality in Language and Speech Systems*, (2002).

[4] C. Elting, J. Zwickel, and R. Malaka, 'What are multimodalities made of? modeling output in a multimodal dialogue system', in *International Conference on Intelligent User Interfaces IUI*, pp. 13–16, (2002).

[5] S. K. Feiner and K. R. McKeown, 'Automating the generation of coordinated multimedia explanations', *IEEE Computer*, **24**, 33–41, (1991).

[6] G. P. Hodgkinson, N. J. Bown, A. J. Maule, K. W. Glaister, and A. D. Pearman, 'Breaking the frame: An analysis of strategic cognition and decision making under uncertainty', *Strategic management journal*, **20**(10), 977–985, (1999).

[7] D. Kirsh, 'A few thoughts on cognitive overload', *Intellectica*, **1**(30), 19–51, (2000).

[8] O. Lerbinger, *The Crisis Manager: Facing Risk and Responsibility*, Lawrence Erlbaum Associates, 1997.

[9] S. Y. Mousavi, R. Low, and J. Sweller, 'Reducing cognitive load by mixing auditory and visual presentation modes', *Journal of Educational Psychology*, **87**(2), 319–34, (1995).

[10] R. S. Nickerson, 'Confirmation bias: A ubiquitous phenomenon in many guises', *Review of General Psychology*, **2**(2), 175–220, (1998).

[11] J. F. Nunamaker Jr, E. S. Weber, C. A. P. Smith, and M. Chen, 'Crisis planning systems: tools for intelligent action', *System Sciences, 1988. Vol. III. Decision Support and Knowledge Based Systems Track, Proceedings of the Twenty-First Annual Hawaii International Conference on*, **3**, (1988).

[12] F. Paas, J. E. Tuovinen, H. Tabbers, and P. W. M. Van Gerven, 'Cognitive load measurement as a means to advance cognitive load theory', *Cognitive Load Theory: A Special Issue of Educational Psychologist*, (2003).

[13] A. Paivio, *Mental Representations: A Dual Coding Approach*, Oxford University Press, USA, 1986.

[14] A. G. Sutcliffe, S. Kurniawan, and J. E. Shin, 'A method and advisor tool for multimedia user interface design', *International Journal of Human-Computer Studies*, **64**(4), 375–392, (2006).

[15] M. Turoff, M. Chumer, B. Van de Walle, and X. Yao, 'The design of a dynamic emergency response management information system (dermis)', *Journal of Information Technology Theory and Application*, **5**(4), 1–35, (2004).

[16] W. Wahlster, E. Andre, W. Finkler, H. J. Profitlich, and T. Rist, 'Plan-based integration of natural language and graphics generation', *Artificial Intelligence*, **63**(1-2), 387–427, (1993).

[17] W. Wahlster, N. Reithinger, and A. Blocher, 'Smartkom: Multimodal communication with a life-like character', *Seventh European Conference on Speech Communication and Technology*, (2001).

[18] G. Wright and P. Goodwin, 'Eliminating a framing bias by using simple instructions to'think harder' and respondents with managerial experience: Comment on 'breaking the frame'', *Strategic management journal*, **23**(11), 1059–1067, (2002).