Multimodal Information Presentation for High-Load Human Computer Interaction

Yujia Cao

曹宇嘉

Multimodal Information Presentation for High-Load Human Computer Interaction

Yujia Cao

曹宇嘉

PhD dissertation committee:

Chairman and Secretary:

Prof. dr. ir. Ton J. Mouthaan, University of Twente, NL

Promotor:

Prof. dr. ir. Anton Nijholt, University of Twente, NL

Assistant promotor:

Dr. Mariët Theune, University of Twente, NL

Members:

Prof. dr. ing. Willem B. Verwey, University of Twente, NL
Dr. Marieke H. Martens, University of Twente & TNO, NL
Prof. dr. Mark A. Neerincx, Delft University of Technology & TNO, NL
Prof. dr. Léon J. M. Rothkrantz, TU Delft & Dutch Defence Academy, NL
Dr. ing. Christian Müller, German Research Center for Artificial Intelligence, DE

Human Media Interaction group
The research reported in this dissertation has been carried out at the Human Media Interaction group of the University of Twente.

CTIT Dissertation Series No. 10-186
Center for Telematics and Information Technology (CTIT)
P.O. Box 217, 7500 AE Enschede, NL

BSIK ICIS/CHIM
The research reported in this thesis has been partially carried out in the ICIS (Interactive Collaborative Information Systems) project. ICIS is sponsored by the Dutch government under contract BSIK03024.

SMARCOS
The research reported in this thesis has been partially carried out in the SMARCOS (Smart Composite Human-Computer Interfaces) project. SMARCOS is sponsored by the EC Artemis project on Human-Centric Design of Embedded Systems under number 100249.

SIKS Dissertation Series No. 2011-07
The research reported in this thesis has been carried out under the auspices of SIKS, the Dutch Research School for Information and Knowledge Systems.

ISSN: 1381-3617, number 10-186
ISBN: 978-90-365-3116-0
DOI: 10.3990/1.9789036531160
URL: http://dx.doi.org/10.3990/1.9789036531160

MULTIMODAL INFORMATION PRESENTATION
FOR HIGH-LOAD HUMAN COMPUTER INTERACTION

DISSERTATION

to obtain
the degree of doctor at the University of Twente,
on the authority of the rector magnificus,
prof. dr. H. Brinksma,
on account of the decision of the graduation committee,
to be publicly defended
on Thursday, February 3, 2011 at 16:45

by

Yujia Cao

born on March 19, 1982

in Nanjing, China

This thesis has been approved by:

Prof. dr. ir. Anton Nijholt, University of Twente, NL (promotor)

Dr. Mariët Theune, University of Twente, NL (assistant promotor)

ISBN: 978-90-365-3116-0

Acknowledgements

Four years ago, I arrived in Enschede with one large suitcase in hand and one big dream in mind. I started a Ph.D. journey without knowing what the future would bring. This journey turned out to be tougher than I expected, but fortunately I met so many wonderful people along the way who supported me, helped me, walked side by side with me, and kept me moving forward in a good direction. Finally, I am standing at the end of this journey, feeling very grateful to all those who have made it such a wonderful experience for me.

First of all, I want to say many thanks to my promotor Anton Nijholt. Anton, thank you for giving me the opportunity to conduct my PhD research in the Human Media Interaction Group (HMI). Thank you for your trust and encouragement when I had problems finding out which direction to go. Thank you for supporting me in every step I took in my research, and for all your valuable feedback on my work.

I am grateful to my daily supervisor Mariët Theune. Mariët, thank you for always thinking along with me and always having time for me. I appreciate the many many useful discussions we had and your detailed comments on everything I wrote. Your interest in my research and your positive attitude have always been extremely encouraging. Thank you for giving me the freedom to make my own decisions, and at the same time keeping me on a good track. I also enjoyed our non-work-related chats about holidays, family, and life. I cannot think of anything more to ask from a daily supervisor.

I would like to thank all HMIers for being great colleagues and friends during the past four years. I enjoyed working with Andreea on the multimodal crisis management system, with Frans on the informative cue experiment, and with Randy, Rieks, and Betsy on the SmarcoS project. I thank all the HMIers who participated in my experiments, especially those who participated three times. Ronald has been an extraordinary office mate. Thank you for helping me every time I needed it. I enjoyed the nice and meaningful conversations we had in the office. Even the meaningless ones turned out to be a nice distraction. Thijs, thank you for helping me out when I got stuck in JAVA programming. I thank Claudia, Andreea, Olga, Maral, Khiet, and Echo for our important girl chats at coffee and lunch breaks. Lynn, thank you for proofreading my papers and thesis. Charlotte and Alice, your administrative support is always flawless. Wim, it is handy that you live just around the corner. Thanks for lending tools and helping with neighborhood issues. Andreea, Olga and Dhaval, thank you for the nice evenings and weekends we spent together at our homes. A special thank to Andreea, my bridesmaid, for being a wonderful friend all this time.

I also would like to thank the Automotive IUI Group at the German Research Center for Artificial Intelligence (DFKI). Christian Müller, thank you for inviting me to work in your group for several months. It has become a successfully collaboration and a memorable experience. Also thank you for the magical last-minute preparation to get the car ready for my experiment, and for the financial support of the experiment. Angela Mahr, the discussions we had were all useful and inspiring. The best ones were probably those that went on for hours and hours until the security guard came in for late-night check. Also thank you for helping me with the questionnaires, the German translation, and the statistics. Sandro Castronovo, thank you for building the in-car HMI, and for always making changes upon my requests without any complaint. Mehdi Moniri, thank you for helping me with creating the 3D driving environment and the simulator software. Veronika Dimitrova-Krause, thank you for arranging participants, being the secondary experimenter and going through the tough experiment schedule together with me. Your company made the hard work enjoyable. Thank you for staying late with me to translate the user feedback from German to English. And most importantly, thank you for being a good friend even since we met.

My former house mates deserve a big thank you. Szymon and Danuta Dutczak, I really enjoyed your company at home during that two years. We had so much fun together. The progress of life has moved you out, but you are always welcome to come back and visit. We can be just like the old days.

My next thanks go to my parents. Mom and dad, you are the best. Thank you for understanding and supporting the decisions I have made for my life, even if they have moved me further and further from home. The unconditional love that only parents can give keeps me strong and warm wherever I go. As an engineer, dad has taught me to work hard and never stop learning. As a manager, mom has taught me to be communicative and social. Thank you both for being good role models.

I also want to thank my two pairs of grandparents for being so (sometimes weirdly) proud of me, and for being healthy and energetic in their 80's. Please keep it up!

My parents and grandma in law are absolutely lovely. Thank you for treating me as your own child, although we do not speak the same language (yet). I am always so spoiled when I come to visit. I love mom's professional cooking and grandma's delicious cake, so yummy. I hope someday we will be able to communicate.

Last but not the least, my husband David Salamon deserves a heartfelt thank you. Dear David, we met around the same time when my Ph.D. journey started. Thank you for being right next to me this whole time. Your constant love and support are the source of my energy. The happiness and laughter you bring to me everyday are perfect anti-stress therapies. Your confidence in me always lights up my heart in hard times. I am very convinced that no matter what happens, you will always be able to bring a smile back on my face. I enjoy talking with you about everything, including research. Being a scientist in a different field, you help me to think out of the box and look at my work from different perspectives. The Ph.D. is finished, but our love goes on, and will always go on.

Yujia Salamon
Enschede, January 2011

Contents

1 Introduction

"A picture is worth a thousand words. An interface is worth a thousand pictures."
Ben Shneiderman, 2003

1.1 Information Presentation in Human Computer Interaction

In a modern society, people live with machines all around them. Machines have changed the way people work, live, communicate, travel, and entertain. Computers, as a typical example, have become indispensable tools in people's lives. Here, the notion 'computer' does not only refer to personal computers (PC) in their various forms, but also includes the embedded computers in numerous machines, from mobile phones, MP3 players, to cars, aircraft, and power plants. When using a PC or operating a machine, the interaction between users and computer systems occurs at the *interface*, which usually includes both software and hardware.

Human computer interaction can be looked at from two perspectives – a task-driven and an information-driven perspective. They are not contradictory but complementary. The user of a machine usually has a task to perform, which is the purpose of the interaction. Via the interface, the user carries out the task while interacting with the computer system. For example, when a person uses a ticket machine to purchase a ticket, the interface consists of a software program, buttons, and a display which can also be a touch screen. The interaction is driven by a set of sub-tasks (e.g. destination selection, date selection, etc.) until the ticket is successfully purchased.

From the information-driven perspective, human computer interaction is driven by a two-way information flow between users and computers. Via the interface, the computer system presents information to the user. After processing the information, the user provides new information (e.g. a choice, answer, command etc.) back to the computer system, which triggers a new round of interaction.

1.1.1 What is Information Presentation?

Information presentation in human computer interaction refers to the way information is presented by the interface to the user [216]. It relates to the information flow from the computer system to the user, thus it is also commonly referred to as output generation or feedback generation. When the interface presents information using multiple modalities, the process is also called multimodal fission, as opposed to multimodal fusion which is to integrate multiple user input channels [132; 270].

In this dissertation, information presentation focuses on *how* to present information rather than *what* information to present. Therefore, our studies always evaluate different presentations of the *same* information contents, aiming to obtain an optimal solution that maximizes task performance and minimizes cognitive demand.

1.1.2 Why is it Important?

Information presentation is not simply a means to send information into the human mind. It actually guides, constrains, and even determines cognitive behavior [290]. In other words, the manner of presentation influences how users perceive and process the information and how much cognitive effort it requires to do so. Consequently, information presentation can greatly influence the quality of interaction. Let us look at two examples.

In the first example (taken from [221]), imagine that you are searching for cheap hotels on the internet. You get two result pages from two different servers, as shown in Figure 1.1. Now try the following tasks: 1) from the top screen, find the price for a double room at the Quality Inn in Columbia; and 2) from the bottom screen, find the price of a double room at the Holiday Inn in Bradley. You most probably have found that the second task takes more time than the first task. Indeed, a study based on this example found that it took an average of 3.2 seconds to search the top screen and 5.5 seconds to find the same kind of information in the bottom screen [254]. In this example, the manner of presentation determines how much cognitive effort it requires to process the information and how long it takes to complete the search task.

The second example will show that the consequence of bad information presentation can get much more serious and catastrophic than consuming a few more seconds. In 2003, NASA's Columbia shuttle exploded on the way back to the earth and the seven scientists on board were all killed[1]. Bad information presentation was identified to be one of the causes of this disaster [253]. During take off, NASA spotted an unexpected foam tile strike on the left wing and sent the video clip to engineers at Boeing for investigation. Boeing engineers produced three reports for NASA, assessing the potential impact damage to the wing. The reports were all made in Microsoft PowerPoint. Beside the fact that the reports were not totally decisive, the information contents were very badly presented. The text font was too

[1] Investigation of the Columbia shuttle disaster at BBC: http://www.bbc.co.uk/dna/h2g2/A40098008

South Carolina

City	Motel/Hotel	Area code	Phone	Rates Single	Double
Charleston	Best Western	803	747-0961	$26	$30
Charleston	Days Inn	803	881-1000	$18	$24
Charleston	Holiday Inn N	803	744-1621	$36	$46
Charleston	Holiday Inn SW	803	566-7100	$33	$47
Charleston	Howard Johnsons	803	524-4148	$31	$36
Charleston	Ramada Inn	803	774-8281	$33	$40
Charleston	Sheraton Inn	803	744-2401	$34	$42
Columbia	Best Western	803	796-9400	$29	$34
Columbia	Carolina Inn	803	799-8200	$42	$48
Columbia	Days Inn	803	736-0000	$23	$27
Columbia	Holiday Inn NW	803	794-9440	$32	$39
Columbia	Howard Johnsons	803	772-7200	$25	$27
Columbia	Quality Inn	803	772-0270	$34	$41
Columbia	Ramada Inn	803	796-2700	$36	$44
Columbia	Vagabond Inn	803	796-6240	$27	$30

Pennsylvania
Bedford Motel/Hotel: Crinaline Courts
(814) 623-9511 S: $18 D: $20
Bedford Motel/Hotel: Holiday Inn
(814) 623-9006 S: $29 D: $36
Bedford Motel/Hotel: Midway
(814) 623-8107 S: $21 D: $26
Bedford Motel/Hotel: Penn Manor
(814) 623-8177 S: $19 D: $25
Bedford Motel/Hotel: Quality Inn
(814) 623-5189 S: $23 D: $28
Bedford Motel/Hotel: Terrace
(814) 623-5111 S: $22 D: $24
Bradley Motel/Hotel: De Soto
(814) 362-3567 S: $20 D: $24
Bradley Motel/Hotel: Holiday House
(814) 362-4511 S: $22 D: $25
Bradley Motel/Hotel: Holiday Inn
(814) 362-4501 S: $32 D: $40
Breezewood Motel/Hotel: Best Western Plaza
(814) 735-4352 S: $20 D: $27
Breezewood Motel/Hotel: Motel 70
(814) 735-4385 S: $16 D: $18

Figure 1.1: Two different ways of presenting the same kind of information at the interface: one makes it much easier to find information than the other (reproduced from [221] p. 96).

small, making many slides too crowded. Some tables were difficult to read and it was difficult to make comparisons of numbers across the tables. Bullet lists were used throughout the reports, with up to five levels of hierarchy on a single slide. Consequently, the reasoning of the potential impact was broken up into fragments both within and between the many slides. Engineers' concerns about the safety of return were not successfully communicated to NASA's management team, who eventually decided not to take any action but to continue the mission as planned. Although this example involves only human-human interaction, it certainly brings the message that information presentation deserves careful investigation in the design of human-computer interfaces.

1.2 Presentation Factors

In order to select a good presentation strategy for certain information, we first need to know what options we can choose from. This also means to know which presentation factors we can use to manipulate the presentation (i.e. to create different presentations for the same information). In the literature, the investigation of presentation factors spreads into numerous application domains, such as decision making support, marketing (advertising), risk communication, health promotion, finance, justice, education (learning). Many factors have been found to influence how people acquire and process information, among which the commonly investigated ones are modality, spatial structure, temporal order and frame. A brief explanation of these factors is given below in separate subsections. Modality is the key presentation factor that is investigated in all our studies (Chapter 4 $\sim$ 7). Spatial structure is also addressed by one of our studies (Chapter 5). Temporal order and frame are not investigated in this dissertation.

1.2.1 Modality

Modality is probably the most investigated presentation factor, because it is relevant to almost all application domains. The definition of modality varies in different research fields (see Section 2.1). In this dissertation, we adopt the definition from the computer science field, in which a modality can simply be considered as the *form* in which information contents are presented, such as text, image, graph, speech, and sound et cetera. Each modality has its own properties and representational power; therefore specific modalities are more suitable for presenting certain types of information than others (see Chapter 2). Modality is also an important presentation factor because the human mind works in a modality-specific manner. The use of modality influences at least three stages of human information processing: sensory processing, perception and working memory (see Chapter 3). Of the different categories of modalities, visual and auditory modalities certainly dominate both the theoretical research and the applied studies. Although the fundamental research on human tactile perception has a long history, the application of tactile modalities to interface design only started in the early 1990's. However, the last decade has seen a rapidly growing body of tactile research and applications.

1.2.2 Spatial Structure

Spatial structure refers to the way information items are spatially arranged, such as list, table and other layouts. It is mostly associated with visual presentations. Figure 1.1 is an example of presenting information with different spatial structures. This factor may influence the strategy people use to acquire and process information. Therefore, when the information is task related or decision related, different structures may result in different task performances and decision outcomes [126; 208; 218; 249]. Previous findings regarding this factor are explained in greater detail in Section 5.1. To a certain extent, spatial structure also influences the temporal order in which information items are perceived, especially when people are used to read with certain orders (e.g. from top to bottom, from left to right). However, spatial structure differs from the temporal order factor, which refers to the order in which information items become available to a user (presented) rather than the perception order of the user's choice.

1.2.3 Temporal Order

Given a set of information items to present, temporal order refers to the sequence of presentation (i.e. which item first, which one second and so on). The order effect – a phenomenon in which the final judgement is significantly affected by the temporal order of information presentation, is a robust finding in empirical studies of human belief revision [2; 3; 15; 30; 45; 196; 229; 257; 275]. For example, in the classical study of order effect [10], the participants listened to a description of a person and were then asked to report their impression of this person. The participants who heard 'intelligent-industrious-impulsive-critical-stubborn-envious' favored this person significantly more than the ones who heard the same set of words in the reversed order. This shows a primacy effect (first impression effect), indicating that the items presented earlier determine the judgement. Some more recent studies on human belief revision also found a recency effect in which the items presented later determine the final judgement (e.g. [45; 257]). Many theories have been proposed to explain how and why these order effects occur (e.g. [275]), but these are out of the scope of this dissertation.

1.2.4 Frame

The human judgement can also be influenced by the way information is 'framed'. When people need to make a choice between several options, presenting the options in terms of gains (positive frame) or losses (negative frame) can elicit different choices. This phenomenon, known as 'framing effect', is another robust empirical finding of human decision making [26; 32; 48; 91; 134; 140; 163]. The frame changes the reference point of the judgment which leads people to focus on either gains or losses. For example, in the classical 'Asian Disease' study [256], participants were told that the US was preparing for the outbreak of an unusual Asian disease, which was expected to kill 600 people. There were two alternative programs to combat the disease. Participants received the alternatives in either the positive frame or the negative frame, and were asked to choose one of the two.

Positive frame:

- If Program A is adopted, 200 people will be saved.

- If Program B is adopted, there is 1/3 probability that 600 people will be saved, and 2/3 probability that no people will be saved.

Negative frame:

- If Program C is adopted, 400 people will die.

- If Program D is adopted, there is 1/3 probability that nobody will die and 2/3 probability that 600 people will die.

Programs A and C were equivalent, so were B and D. However, participants receiving the positive frame solidly preferred the certain option A (72%) and those in the negative frame strongly preferred the risky option D (78%). This result indicated that participants were more risk-adverse when they focused on gains, and more risk-prone when they focused on losses.

1.3 High-Load Task Environments

In this dissertation, we focus on presenting information that is directly task-related, which means that users need to perform a task or several tasks upon the reception of information[2]. More specifically, our work is focused on high-load task environments, where users need to perform an (or several) interaction-related task(s) under a high level of mental workload (cognitive load). A high-load interaction can have various causes, such as:

- the interaction-related task(s) is(are) highly difficult for a particular user.
- task performance is under time pressure.
- the user simultaneously performs other tasks that are irrelevant to the interaction.

In high-load task environments, information presentation may have a particularly notable impact on task performance. The human cognitive capacity (working memory and attention resources) is known to be limited [168; 278]. Therefore, a person in a high-load interaction may not have much spare cognitive capacity to cope with additional load that is unnecessary for the task. Consequently, suboptimal presentations can cause cognitive overload and harm the task performance. In this dissertation, a high-load task setting is a common feature of all our studies. The high-load factors investigated here include time pressure, high information load and multi-tasking. To motivate these factors, we have selected two task domains where high-load human computer interactions often occur: crisis management and driving.

1.3.1 Crisis Management

Crisis management is to deal with extreme events that can injure or kill large numbers of people, do extensive damage to property, and disrupt community life [65]. Crisis management typically takes place under time pressure. Timely and correct decisions may shorten

[2]In contrast to task-related information, information can also be non-task-related, then the purpose of presentation is only to inform. For example, when reading the news, a person usually does not need to provide any direct reaction.

the duration of the crisis and reduce the negative impact. It is not exaggerating to say 'time is money' and 'time is life'. Besides, crisis managers typically have to deal with information overload [43; 68; 107]. There is always a large amount of information to process and the contents are often distorted or incomplete. Nowadays, computers assist crisis management in all phases, from preparation, planning, training, response, recovery, to final assessment [43]. For example, several multimodal interfaces have been designed to facilitate the communication between a wide range of users and devices [84; 129; 220].

In conclusion, the crisis management domain is suitable for simulating high-load human computer interaction. In two studies presented in this dissertation (Chapter 4 and 5), the user tasks were embedded in crisis scenarios (earthquake rescue). Note that the intention was not to simulate a realistic crisis management act, but to utilize its high-load characteristics (time pressure, high information load) in a somewhat simplified task setting, which allowed us to better investigate the cognitive impact of information presentation.

1.3.2 Driving

Driving is usually not a difficult task for experienced drivers. However, because the traffic environment is dynamic, unexpected situations/events can occur at any time. For example, a child suddenly runs into the street or a road obstacle becomes visible shortly after a road bend. Although such emergent danger rarely occurs, once it does, drivers need to decide and respond quickly, and an inappropriate or late reaction can have catastrophic consequences.

Besides time critical events, multi-tasking and distractions can also induce a high level of mental workload during driving. According to several field observational studies, drivers engage in secondary tasks in about 30% of the driving time [201]. Secondary tasks included talking to passengers, talking on a cell phone, listening to the radio, eating, drinking, grooming and interacting with in-vehicle information systems (IVIS). According to another large-scale field study [178], 78% of traffic collisions and 65% of near collisions were associated with drivers' inattention to the road ahead, and the main source of this inattention was distraction from secondary tasks.

As automotive technology advances, IVIS have gained a wide variety of functions [219], including route planning, navigation, vehicle monitoring, traffic and weather update, hazard warning, augmented signing, motorist service. IVIS can even assist with driving-irrelevant tasks, such as email management and in-car infotainment [112]. Besides the obvious benefits, these IVIS functions are also distracting, and thus potentially harmful when the driver is under high load. As one way to reduce driver distraction, IVIS need to present messages in an optimized manner so that they require minimal attention resources to be perceived and processed. Moreover, IVIS need to interrupt drivers in a way that supports their attention management between multiple tasks. Accordingly, two studies in this dissertation investigated in-vehicle information presentation, using a time-limited task setting (Chapter 6) and a multiple task setting (Chapter 7).

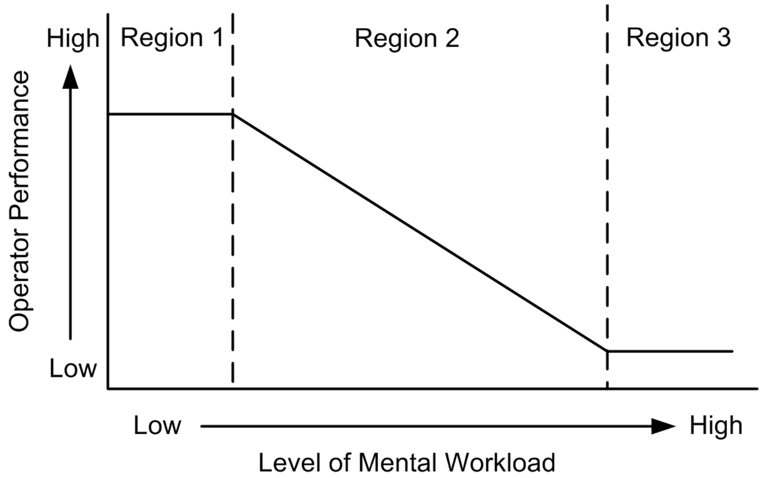


Figure 1.2: Hypothetical relationship between primary task performance and operator workload (reproduced from [71] p. 219).

1.4 Measures of Mental Workload

The assessment of mental workload has been a key issue in the development of human-machine interfaces, especially when the design objective is to minimize users' mental workload when interacting with the system. Existing measures of mental workload can be specified into three groups: performance measures, subjective (i.e., self-report) measures and physiological measures [167; 185].

1.4.1 Performance Measures

Performance measures are grounded on the assumption that an increase in task difficulty will increase mental workload, which will decrease task performance [167]. This is to say, the worse the performance, the higher the mental workload. Performance measures can be based on either primary task or secondary task. Primary task measures assess the user's capability to perform a task or a group of tasks that is the system's function of interest [71]. It is also an overall assessment of the effectiveness and efficiency of the human-machine interaction [185]. Effectiveness can be thought of as 'doing the right thing' and efficiency is 'doing things the right way' [211]. Therefore, effectiveness can be reflected by performance accuracy or error rate, whereas efficiency is often associated with time measures such as reaction speed and time performing duration. Primary task measures are easy to apply and directly relate to the system function of interest. However, they are not sensitive to changes in mental workload when the task is too easy or too difficult. Figure 1.2 shows the hypothetical relationship between primary task performance and operator workload ([71], p. 219). In regions 1 and 3, the task is either too easy or too difficult so that the performance remains very high or very low. Only in region 2 does primary task performance reflect the variances in mental workload.

Secondary task methodology is intended to allow workload assessment when the workload of primary task(s) falls into region 1 [71]. By adding a secondary task, the total workload can be moved from region 1 to region 2. Secondary task methodology has two paradigms: the subsidiary task paradigm and the loading task paradigm [185]. In the subsidiary task paradigm, users are instructed to prioritize and maintain the primary task performance. Consequently secondary task performance varies with the primary task load and indicates 'spare mental capacity'. Assuming that the total mental capacity available to perform all tasks is limited and constant, a lower secondary task performance indicates a higher primary task load. In contrast, the loading task paradigm gives instructions to prioritize and maintain the secondary task performance, and measures the primary task performance. Since the difficulty level of the secondary task does not change over time, it induces a constant level of mental workload which shifts the workload of the primary task from region 1 to region 2. There are many secondary tasks that have been validated as effective and sensitive for the purpose of workload assessment. Details about these tasks are not provided here due to the fact that secondary task measures have not been used in any of our studies. An elaborated description of these tasks can be found in [71].

1.4.2 Subjective Measures

Assuming that people are able to introspect and report the amount of workload expended on a task, subjective measures use questions and rating scales to let the user self report the level of mental workload he/she has experienced. Rating scales can be unidimensional or multidimensional [211]. Unidimensional scales require only one rating, whereas multidimensional scales consist of several sub-scales which can be summarized in an overall assessment. Multidimensional scales are usually more diagnostic because they can indicate not only the variation in workload but also the cause of variation [99]. Moreover, many multidimensional scales do not only address mental effort, but also other aspects of workload such as physical demand, time constraints, et cetera.

Many workload rating scales can be found in the literature. One of the most outstanding ones is the NASA Task Load Index (NASA-TLX [97]). It requires a multidimensional rating procedure on six sub-scales: Mental Demands, Physical Demands, Temporal Demands, Performance, Effort and Frustration. Detailed definitions of the sub-scales can be found in [97]. There are 21 rating gradations on each sub-scale. The complete questionnaire is available online in both paper and computer version. After the rating procedure, an overall workload score can be calculated as a weighted average of ratings on six sub-scales. To determine the weights for a user, he/she needs to compare each pair of sub-scales (fifteen comparisons in total) and indicate which one contributes more to his/her feeling of workload. Alternatively, one can give the same weight to the six sub-scales and simply calculate the average of them.

The NASA-TLX can be adapted to better suit a specific task domain. The Driving Activity Load Index (DALI) is a revised version of NASA-TLX, adapted to the driving task [192; 193]. The purpose of DALI is to assess the workload of driving a vehicle equipped with on board systems, such as IVIS, radio, car phone, et cetera. DALI also has six sub-scales: Effort of Attention, Visual Demand, Auditory Demand, Temporal Demand, Interference, and Situational Stress. DALI has been applied in our own automotive study (see Chapter 6). Appendix C1 provides detailed descriptions of the sub-scales and the DALI questionnaire.

Other well-known workload rating scales include the MCH scale (modified Cooper-Harper Scale) [284], the Bedford Scale [207], and the SWAT (Subjective Assessment Technique) [204]. Detailed descriptions and comparisons of these rating scales can be found in [71] and [211].

1.4.3 Physiological Measures

The last category of workload measures are those based on the user's physiological states, assuming that physiological variations represent implicit fluctuations in the user's cognitive state [80]. There are two general classes of physiological measures: central nervous system (CNS) and peripheral nervous system (PNS) measures [130]. CNS includes the brain stem and the spinal cord. CNS measures include electroencephalography (EEG), event-related brain potentials (ERP), magnetic activity of the brain (MEG), positron emission tomography (PET), and electrooculography (EOG). PNS includes all neurons outside the brain and the spinal column. PNS measures include electrocardiogram (ECG), respiratory activity, electrodermal activity (EDA) and oculomotor activity. Here we focus on PNS measures because CNS measures are not applied in our own studies.

Cardiovascular activity. ECG-based measures of cardiovascular activity are the most commonly used PNS measures of mental workload (cognitive load) [285]. Heart rate (HR) increases and heart rate variability (HRV) decreases as a function of increases in mental workload [130; 217; 287]. Simply put, the heart tends to beat faster and more evenly when mental workload increases. HRV can be derived from the sequence of beat-to-beat intervals (in seconds). In the time domain, this means to calculate the standard deviation of the beat-to-beat intervals. In the frequency domain, spectral analyses (e.g. Fast Fourier Transform) can be performed on the beat-to-beat interval sequence, and then the spectral power in the low band (LF, 0.07 Hz $\sim$ 0.14 Hz) is related to mental workload. Decreased power in the LF band usually reflects higher levels of mental workload [130].

Skin conductance. Skin conductance (SC) is one way of describing electrodermal activity (EDA). It consists of two separate components, skin conductance level (SCL) and skin conductance response (SCR). SCL is a slow moving tonic component that indicates a general activity of the sweat glands from temperature or arousal. SCR is a faster phasic component that is influenced mainly by the level of arousal and emotion [242]. Increments of SCR number (per unit of time) and amplitude can be interpreted as indicators of increased workload [28; 267]. If the environmental temperature is constant, SCL can also be considered as an indication of general arousal level. Increment of SCL indicates increased arousal level [94].

Respiration. Respiration rate and amplitude (depth) are also sensitive measures of mental workload. As workload increases, respiration rate tends to become more rapid and respiration depth tends to decrease [119; 217; 287]. The implementation of respiration measures is limited when subjects need to have voice communications, because speech disrupts the pattern of respiration [287].

Oculomotor activity. Mental workload can also be measured from the movement of the eyes, such as pupil diameter, fixation time, saccade distance, saccade speed and blink rate [57; 110; 252; 263; 286]. Eye movements are usually captured by eye tracking devices. Previous findings generally suggest that when mental workload increases, pupil diameter increases, fixation time increases, saccade speed decreases, saccade distance decreases and blink rates decreases. However, the interpretation of eye activities can be difficult because they are related to many factors such as the type of task demand (e.g. visual, auditory), age, time on task, experience, and the external environment (e.g. the lighting condition) [252; 287].

1.5 About This Dissertation

1.5.1 Research Context

This dissertation work was carried out in the context of the ICIS[3] project and the SmarcoS[4] project. ICIS aimed to design, develop and evaluate intelligent information systems that could support complex decision making. Crisis management was one of the common use cases of ICIS research. The focus of the CHIM (Computational Human Interaction Modeling) cluster was to develop advanced multimodal interfaces that could facilitate high-load interactions between multiple users and multiple devices. Both ends of the interaction (user input and system output) were intensively researched. Our work in ICIS centered on investigating the cognitive impact of multimodal information presentation, which is a necessary step towards generating cognitively-compatible presentations for high-load interactions.

The SmarcoS project aims to design and develop interconnected embedded systems with inter-usability. This means that SmarcoS allows interconnected devices and applications to communicate, exchange context information, user actions, and semantic data. It allows applications to follow the user's actions, predict needs and react appropriately to unexpected actions. The task of work package four (Attentive Personal Systems) is to build an intelligent system that motivates and supports users in their daily life to live a balanced and healthy lifestyle. This system runs on inter-usable devices (e.g. PC, mobile phone, in-car computer, TV, etc.) so that its service is available when the user is at work, at home or on the go. Our work on in-vehicle information presentation contributes to the "on the go" section. Driving is always the main task in a moving vehicle, and thus in-vehicle information presentation needs to be compatible with driving and to minimize unnecessary distractions from driving. This is especially important in a high-load driving environment.

[3]The ICIS (Interactive Collaborative Information Systems) project was funded by the Dutch Ministry of Economic Affairs under contract number BSIK03024.

[4]The SmarcoS (Smart Composite Human-Computer Interfaces) project is funded by the EC ARTEMIS program on Human-Centric Design of Embedded Systems, under contract number 100249.

1.5.2 Research Overview

The main objective of our work is to investigate the effect of information presentation in a high-load task setting and to provide useful suggestions on the design of multimodal interfaces. Towards this objective, we have taken the following steps. A literature study was first conducted, followed by a series of experimental studies investigating information presentation in several high-load task settings.

Literature study

As mentioned previously, modality is the main factor of interest. First of all, we have conducted an extensive literature study on modality-related theories and findings in various research fields. The objectives of this literature study include:

- To obtain an overview on the allocation of modality in intelligent multimodal presentation systems (IMMP), including the common methods, the factors taken into account and existing guidelines. See Chapter 2.
- To understand the relation between modality and human cognition (i.e. the modality-specific features of human information processing). This knowledge provides a theoretical foundation to understand the cognitive impact of multimodal information presentation and to use modality in a cognitively-compatible fashion. See Chapter 3.
- To summarize previous findings about the use of modality in in-vehicle information presentation. This shows the utility of the cognitive knowledge in a specific application domain, and also serves as a common related-work study for our own work in the automotive context. See Chapter 3.

Study on time limited visual search (Chapter 4)

Modality-related cognitive theories have rarely been applied to the design of IMMP. In this study, we aim to confirm the relevance of several well-founded theories (the working memory theory [14], the dual coding theory [187] and the multiple resource theory [283], see Chapter 3) to the design of multimodal interfaces. A user experiment has been conducted to investigate the cognitive impact of modality using a time limited visual search task. The task is a high level abstraction of a crisis management practise. It better allows the experimental results to be predicted and interpreted in the light of the cognitive theories. The design of the experiment is briefly summarized below.

- *Scenario*: earthquake rescue
- *Information to be presented*: location of wounded and dead victims
- *User task*: to send a doctor to wounded victims
- *High-load factor*: time pressure and high information load
- *Presentation factor*: modality
- *Investigated modality set*: text, image, speech, sound

- *Measures*: performance, subjective and physiological measures

In addition, we propose a computational model, which predicts the suitability of any modality choice for a given presentation task, based on relevant cognitive theories and other modality allocation criteria.

Study on time limited decision making (Chapter 5)

The visual search task mentioned above requires mostly perception (to see and to hear) which is a low-level cognitive activity. In this study, we move on to a higher-level cognitive task – decision making based on calculation and comparison between multiple options. Information presentation for multiple choice decision making has been well investigated, but rarely in a time-limited task setting. Besides, the investigated decision tasks rarely have a defined solution or a correct outcome. We have conducted a user experiment using a time limited decision making task with a clearly defined solution. The objective of this study is to investigate: 1) the presentation effects on decision making performance (defined by time efficiency and accuracy), 2) the interaction between different presentation factors (modality and spatial structure), and 3) the interaction between presentation factors and the time limit. The design of the experiment is briefly summarized below.

- *Scenario*: earthquake rescue
- *Information to be presented*: injury condition of wounded victims
- *User task*: to decide which patient needs treatment more urgently
- *High-load factor*: time pressure
- *Presentation factors*: modality and spatial structure
- *Investigated modality set*: text, image
- *Measures*: performance and subjective measures

Study on local danger warnings (Chapter 6)

Local danger warning is an important function of in-vehicle information systems (IVIS) to improve the safety of driving. Presenting local danger warnings is a challenging task because in an emergent danger, drivers have little time to perceive and react to the warning. This is especially true when the danger is not yet visible to the driver's own eyes. We have conducted a series of two user experiments on presenting emergent road obstacle warnings. Experiment One serves as a pre-study of Experiment Two, aiming to obtain a visual warning presentation that can be perceived with little time and effort. Experiment Two further investigates eight warning presentation strategies in a simulator-based driving environment. The objective of Experiment Two includes: 1) to find out whether local danger warnings can indeed enhance driving safety; 2) to find out which modality(ies) and level(s) of assistance are most suitable for presenting local danger warnings; and 3) to obtain subjective judgements on how useful the warnings would be in various real-life driving situations. The design of the experiment is briefly summarized below.

- *Scenario*: driving
- *Information to be presented*: road obstacle warnings
- *User task*: to drive, avoid emergent road obstacles and recall warning messages
- *High-load factor*: time pressure and multi-tasking
- *Presentation factors*: modality and level of assistance
- *Investigated modality set*: text, image,speech, sound
- *Measures*: performance and subjective measures (based on ISO usability model)

Study on informative interruptive cues (Chapter 7)

Besides local danger warnings, IVIS also have a wide range of other functions which can be either driving related or not. As IVIS are increasingly able to obtain and deliver information, driver distraction becomes a larger concern. In the automotive domain, a large number of studies have been carried out on the design and presentation of IVIS messages. However, assisting drivers to selectively attend to these messages for safer driving has rarely been investigated. We propose that using informative interruption cues (IIC) in addition to IVIS messages can be an effective means to minimize inappropriate driver distractions. Accordingly, this study is focused on the design and presentation of IIC (rather than IVIS messages), aiming to 1) design a set of sound and vibration cues that conveys four levels of priority; 2) evaluate the cues to see whether they are easy to learn and can be quickly and accurately identified under various types of cognitive load that drivers can encounter during driving; and 3) compare sound and vibration to find out which modality is more suitable under which conditions. The design of the experiment is briefly summarized below.

- *Scenario*: driving
- *Information to be presented*: informative interruption cues
- *User task*: to track moving object (mimicking driving), identify cues, listen to radio and have conversation
- *High-load factor*: multi-tasking
- *Presentation factor*: modality
- *Investigated modality set*: vibration and sound
- *Measures*: performance and subjective measures

1.5.3 Dissertation Outline

The remainder of this dissertation is divided into four parts. Part One (background) presents the outcome of our literature study in two separated chapters – Chapters 2 and 3. Part Two (information presentation for time limited tasks) includes the two studies on time limited tasks, respectively in Chapters 4 and 5. Part Three (information presentation in the automotive context) includes the two driving-related studies, respectively in Chapters 6 and 7. Finally, Part Four (reflection) provides a general discussion of our findings in Chapter 8, and conclusions and future work in Chapter 9.

Part I

Background

2 Modality Allocation in Information Presentation

This chapter starts by introducing the definitions of modality in different research fields (Section 2.1), followed by presenting modality taxonomies which allow modalities to be identified in a systemized manner (Section 2.2). Then, we move on to describe methods and examples of modality allocation in intelligent information presentation systems/frameworks, focusing on the most common rule-based approaches (Section 2.3). Afterwards, Section 2.4 summarizes a collection of modality allocation guidelines, which can be used as a reference for multimodal interface design. Finally, Section 2.5 discusses the rationale underlying good or bad modality allocation choices – the compatibility to human cognition.

2.1 Modality Definitions

The term ‘modality’ is interpreted differently in different fields. In cognitive science (cognitive psychology and neuroscience in particular), modality commonly refers to the types of human sensation, namely vision, hearing, touch, smell and taste. Based on this definition, there are five major modality types: visual, auditory, tactile, olfactory and gustatory. In computer science (human-computer interaction in particular), modality is interpreted more broadly as “mode or way of exchanging information between humans or between humans and machines in some medium”; where medium is the “physical realization of information at the interface between human and system” ([21] p. 94). One can simply think of modality as the form in which certain information content is presented (e.g. text, image, speech, sound, etc.), and medium as an output device (e.g. screen, speaker etc.) that enables the realization of modalities [209]. The following example is given to further demonstrate the difference between the two definitions: text is a modality (computer science definition) that is perceived via the visual modality (cognitive science definition). This dissertation adopts the modality definition from the computer science domain.

HCI studies commonly distinguish input modalities from output modalities [183]. Using the computer system as a reference point, input modalities carry information provided by users into the system, while output modalities deliver information generated by the system to users. The same modality may rely on different media to be realized when it serves as input or as output modality. For example, text as an input modality is realized via a keyboard (the user types text into the computer system), and text as an output modality is realized via a display (the computer system displays text on the display). This dissertation only addresses output modalities, and calls them simply 'modalities'.

2.2 Modality Taxonomies

By the modality definition in computer science, any presentation form that can be perceived by humans is a modality. This leads to a body of modalities that is large in size and diverse in property. Provoked by the thought that modalities need to be addressed and identified in a unified manner, Bernsen proposed a taxonomy to classify the properties of modalities, which is well-known as 'Modality Theory'. Evolved from Bernsen's work, another modality taxonomy proposed by Bachvarova emphasizes the cognitive properties of modalities and their combination characteristics. To our knowledge, these are the only two taxonomies that are proposed for the classification of output modalities, despite the fact that modality (both input and output) is often addressed in frameworks/models of multimodal interaction (e.g. [98; 151; 179; 183; 210]). They also form a theoretical foundation to support modality allocation and combination in multimodal output generation. Next, we describe these two taxonomies in more detail.

2.2.1 Bernsen's Modality Theory

Bernsen's Modality Theory is a taxonomy of *unimodal* output modalities [19]. Based on the observation that different modalities have different representational power, Bernsen defined the following set of basic representational properties to identify modalities.

- Visual/auditory/tactile/olfactory/gustatory
  This is the perceptual property of a modality, which is determined by its sensory receptor (eye, ear, skin, nose or tongue).
- Linguistic/non-linguistic
  Linguistic modalities are language-based, such as speech and text. Non-linguistic modalities are not language-based, such as sound and pictures. This property is also commonly referred to as verbal/non-verbal.
- Analogue/non-analogue
  Analogue modalities present information via aspects of similarity between the presentation and what it presents, such as images and diagrams. On the other hand, non-arbitrary modalities, such as language, provide the generalities and abstractions which cannot be provided through analogue representation. This property is also referred to as iconic/non-iconic.

- Arbitrary/non-arbitrary
  Arbitrary modalities do not rely on any already existing system of meaning to perform their presentation function and non-arbitrary modalities do. Arbitrary modalities are therefore by definition non-linguistic and non-analogue. A sound alarm is an example of an arbitrary modality in cases where the alarm is only intended to draw attention (the pattern of sound is not specially designed to convey a meaning). Traffic signs are examples of non-arbitrary modalities.
- Static/dynamic
  Static modalities, such as written text, offer users freedom of perception, meaning that they can be perceived by users in any order and as long as desired. In contrast, dynamic modalities are transient and do not offer freedom of perception, such as speech.

This taxonomy is claimed to be complete and unique, which means that all possible unimodal output modalities can be identified and each of them can be described in only one way [21]. For example, written text is a visual, linguistic, non-analogue, arbitrary and static modality. Existing theories and empirical studies on the use of modality mostly concern the first two categories of properties (perceptual and verbal/nonverbal). In the perceptual property category, visual and auditory modalities have been investigated the most in nearly all application domains. This is because they are the most feasible and the most natural to use in human machine interaction. To convey information via tactile modalities (e.g.structured vibration) is rather new, but it is gaining more and more research attention in recent years. The remaining two sensory channels, smell and taste, have rarely been explored for the purpose of human machine interaction.

2.2.2 Bachvarova's Modality Ontology

While Bernsen's taxonomy is focused on the representational properties of modality, Bachvarova [11] argues that the cognitive properties are particularly important and need to be addressed separately from the representational properties. Cognitive properties are those which determine how a modality is perceived and processed by the human cognitive system. In addition, the description of a modality should also contain information about how it can be combined with other modalities. Accordingly, Bachvarova re-organized the modality properties proposed by Bernsen on three levels.

- Information presentation level
  This level describes the capability of a modality to represent certain types of information. The properties linguistic/nonlinguistic and analogue/non-analogue belong to this level.
- Perception level
  This level determines how a modality is perceived by the human perceptual-sensory system. It distinguishes between being visual, auditory, haptic, olfactory, or gustatory. Static/dynamic is also a property at this level, because it determines how much time a modality allows to be perceived and processed.

- Structural level
  This level models the dependencies that can exist between multiple modalities that are combined in one presentation. For example, when placing an icon on a map, these two combined modalities depend on each other to convey the location of the object. The arbitrary/non-arbitrary property belongs to this level.

This taxonomy is questionable in some aspects, for instance the 'linguistic/nonlinguistic' property also influences how information is processed by the human brain (see Section 3.4.2). However, it certainly has an added value in addressing the cognitive impact of modality choice. Indeed, modality plays an important role in human information processing, which will be explained in detail later in Chapter 3. Next, we discuss previous studies about modality allocation in intelligent multimodal presentation systems.

2.3 Modality Allocation in Intelligent Multimodal Presentation Systems

Intelligent multimodal presentation systems (IMMP) are knowledge-based systems, which exploit their knowledge base in order to dynamically *adapt* their output generation to the run-time requirements of user-computer interaction, such as the user profile, task characteristics, nature of the information to be conveyed, et cetera [27; 118]. The development of IMMP has received much research attention during the past two decades. The application domain is very broad, including home entertainment [76], technical document generation [272], medical training [149], crisis management support [84] and much more.

IMMP by definition have more than one modality available for generating presentations. Modality allocation in IMMP systems refers to a process that chooses one or more modalities to *best* present a certain information content for achieving a certain presentation goal [27; 75]. It can also be considered as to make the most suitable mappings between a set of information and a set of modalities, constrained by certain factors [7]. The factors can be the type of information to be conveyed, the presentation goal, the characteristics of the available modalities, the user profile, the condition of the environment, the type of user task, or any other factors that are identified to be relevant to a specific application. For automated generation of multimodal presentations, IMMP also need to allocate modalities on the fly, adapting to variations in the selected factors.

2.3.1 The Rule-based Approach

Modality allocation rules

In existing IMMP studies, modality allocation is commonly rule-based [8; 81; 118; 149; 165; 209; 210; 246; 271; 273]. Modality allocation rules typically associate factors and their values with the preferred modality choice, such as "When factor F has a value V, use modality M". They are usually pre-defined and embedded in the knowledge base of the system. The rules are the core of the intelligence in the sense that they define which factors the system should adapt to and how it should adapt.

For example, in the computer-assisted surgery system presented in [149], the information to be presented was the distance between the needle tip and a target point inside the patient's body. An application-specific factor - the location of the needle - was selected to guide modality allocation. Three values were distinguished: outside the patient's body, just inserted into the body and very near to the target (<10mm). The two available modalities were sound and a color gauge. Sound conveyed distances by varying its frequency. The closer the needle was to the target point, the more frequently the sound was repeated (decreasing the inter-beep interval). The color gauge was displayed on the mini-screen tied to the needle. The length of a color bar increased as the needle got closer to the target point. Finally, three modality allocation rules were made: 1) when the needle is outside the patient's body, only sound is used to present the distance; 2) when the needle is inserted into the body, sound and color gauge are used redundantly; and 3) when the needle tip is very near the target point, only color gauge is used.

Table 2.1 shows more examples of modality allocation rules and the factors they are associated to. Note that these rules described in the form of natural language are usually not directly interpretable by the system. Instead, rules need to be translated into the representation language of the system, such as the M3L language used in SmartKom [271] and the MOXML language used in MOST [209]. When allocating modalities for a certain presentation task, the system searches the rule base for rules associated to factor values at that particular sate of interaction.

Additional rules

When more than one modality is selected for the same presentation task, additional knowledge might be needed to define the coordination between them. This knowledge can be translated into a separate set of rules and embedded into the rule base, as proposed by the authors of [210]. The commonly used coordination modes include synergy (i.e., the use of several modalities to present various aspects of the same event or process) and redundancy (i.e., the use of several modalities to present the exact same information) [214]. For a systematical description of modality combination, Vernier et al. [264] proposed a theoretical framework that defines a two-dimensional combination space. One dimension is combination schema, containing five instances (derived from [5]): distant combination, one point of contact, partial overlap, total overlap and inclusion. The other dimension describes the aspects at which the modalities are combined, including time, space, interaction language and semantics. Although frameworks of this type (see also [54; 152; 181; 182]) do not directly offer design suggestions on what and what not to combine, they serve as the foundation for encoding existing guidelines and creating new rules.

Another type of knowledge that might be necessary in the rule base is the medium /media that will be used to realize the selected modality. This is very often not an issue, because for each modality that is used by the system, there is only one device that can realize it. For example, if a system contains one screen and one speaker, then the screen

Table 2.1: Examples of modality allocation rules in several IMMP systems/frameworks.

IMMP system or framework	Function or scenario	Factors	Modality allocation rule examples
SmartKom [271]	Home digital guide	Presentation goal	To inform the user about TV program, use text in a list.
COMET [81]	Technical explanations	Information type	For location and physical attributes, use graphics. For abstract actions and relationships among actions (such as causality), use text only. For compound actions, use both text and graphics.
WIP [273]	Technical explanations	Information type and communicative functions	For abstract information (e.g. quantifiers), use text. For concrete information, such as visual attributes, use graphics. For spatial information, use graphics.
Multimedia design advisor tool [246]	Multimodal presentation design advice	Information type and communication goals	If the communication goal is to calm, then use soothing image and audio about nature. If the communication goal is to persuade, then use animations and speech.
WWHT [210]	Phone call reception announcement	User profile, phone mode, battery level, environment noise level, information type etc.	If the battery is low, then use auditory modalities instead of visual and tactile modalities. If the noise level is greater than 80db or the phone is in silent mode, then use visual or tactile modalities. If the information is caller identification, then use analogical modalities, such as a photo of the caller.
MOST [209]	Ground marking in a fighter plane	User state (head position), system state (device availability), environment model (luminance, noise)	If the information type is a command, then use text with the head mounted visor. If the head position is low, then do not use visual modalities on the head mounted visor.
MULTIFACE [165]	Multimodal Dialog system for multiple devices	User condition, device type and information type	When a user is in a loud environment or when a device does not have speakers, do not use auditory modalities.

displays all visual modalities and the speaker delivers all auditory modalities. However, if this is not the case, then additional rules are needed to specify a medium or several media to each modality choice. Media allocation rules can be either separate from or integrated with modality allocation rules. An example of the separate case is the MULTIFACE system [44], which possesses a predefined list of modality-device pairs, indicating which modalities are allowed and preferred by each device. The integrated case can be seen in the fighter plan ground marking system in [209]. It contains two output devices for visual modalities: a screen and a head-mounted visor. Their modality allocation rules define the modality choice as well as the medium to use (see Table 2.1).

Limitations

The main limitation of the rule-based method lies in the possible failure in the final selection due to conflicting outcomes from different rules. For example, the most appropriate modality choice given a certain information type might differ from the most appropriate modality choice given a certain user preference. Conflicts are more likely to occur when the number of factors or the number of rules gets higher. To tackle this problem, an obvious solution is to add additional criteria for solving conflicts, such as a factor importance ranking. Besides, a rule can contain more than one modality choice, such as a best choice, a second-best choice and so on. However, these rule-based solutions still have a lack of flexibility [291].

2.3.2 Other Approaches

Alternatively, some studies quantified the rules by translating them into numerical metrics of weights or appropriateness. Then, computational models can be applied to perform the overall optimization. For example, [291] presents a graph-matching approach which operates on two graphs and two sets of numerical metrics. A data graph consists of information units (nodes) and relations between them (links). A modality graph contains available modalities (nodes) and the similarity/compatibility between them (links). A set of modality selection metrics assesses the desirability of each modality-data pair, meaning how desirable it is to select a modality for an information unit. The numerical values are determined based on modality allocation rules for three factors: task type, user profile and information type. Another set of metrics, called presentation coordination metrics, translates several modality combination rules into numerical form. Finally, a probabilistic graph matching algorithm (graduated assignment algorithm [90]) is applied to find a data-modality mapping that maximizes the overall desirability. Other examples of computational approaches can be seen in [241] and [118]. Authors of these studies normally do not call their approaches rule-based. However, the input metrics of these computational models are still derived from rules. What differs is the way in which the rules are encoded and inferred by the system.

2.3.3 Sources of Rules

In order to construct the rule base for a given application, designers can fall back on two knowledge resources: 1) application-specific requirements which can be obtained from field experts, and 2) relevant guidelines[1] which can be retrieved from the literature. Note that the collection of guidelines is not limited to the domain of IMMP. In fact, the choice of modality is a concern in any domain where information needs to be presented to people (not necessarily via computer interfaces), such as education, risk communication, marketing, just to name a few. In principle, modality allocation in IMMP can learn from findings of any modality-related study in any domain. In the next section, we summarize a collection of modality allocation guidelines to serve as a design reference. This collection has by no means made a complete coverage over *all* relevant findings in *all* research domains. It only addresses three most commonly investigated aspects, namely perceptual properties, information type and modality combination.

2.4 A Collection of Modality Allocation Guidelines

In this section, we present three categories of modality allocation guidelines. The first category is associated with the perceptual property of modalities, suggesting when to use visual, auditory, and tactile modalities. The second category is related to the type of information to be presented, stating the type(s) of information each modality is suited for. The third category addresses modality combination, suggesting when to use multiple modalities in one presentation and presenting good "modality compounds".

The guidelines were either directly derived from individual studies or taken from existing guidelines collections in the literature. Note that modality allocation guidelines are not about the realization details of a modality. For example, an allocation guideline of sound describes a situation where sound is a preferred modality, but not how to set the pitch and loudness of a sound signal. Guidelines on realization details are normally called modality design guidelines, and are out of the scope of this chapter.

2.4.1 Guidelines on Perceptual Properties

Visual and auditory modalities are the most commonly applied in information presentation. A set of guidelines regarding the selection between visual and auditory modalities was proposed by Deatherage in the early 1970's [58]. After several decades, these guidelines remain valid for the design of multimodal information presentation and are particularly relevant to in-vehicle information presentation [135; 184].

Use auditory modalities when:

- The message is simple
- The message is short
- The message will not be referred to later
- The message deals with events in time

[1] Here, we distinguish the term 'guideline' from the term 'rule'. Rules are tailored for an application, whereas guidelines are general suggestions that any application can refer to. For example, "use non-verbal visual modalities to present spatial information" is a guideline. When adapted to an application, a rule can be made as "use map on the screen to present the location of building x".

- The message calls for immediate action
- The message must be detected independent of head position or eye gaze
- The visual system is overburdened
- The environment is too bright or too dark
- The user must move around

Use visual modalities when:

- The message is complex
- The message is long
- The message will be referred to later
- The message deals with locations in space
- The message does not call for immediate action
- The auditory system is overburdened
- The environment is too noisy
- The user can stay in one place

Regarding tactile modalities, both allocation and design guidelines have been proposed in recent years [95; 238; 259]. The following guidelines suggest when to use tactile modalities to present information.

Use tactile modalities when:

- The task requires the alarming feature of attention
- The task is temporal in nature
- The task involves hand-eye coordination (e.g. object manipulation), where haptic sensing and feedback are key to performance
- The visual and auditive channels of an operator are heavily loaded
- The visual and/or auditive information is degraded. For example the visual display possibilities are limited (e.g. in hand-held devices), or the auditive channel is unattractive (e.g. in noisy public places).

2.4.2 Guidelines on Information Type

Information type is perhaps the most investigated factor for modality allocation. Modalities differ in terms of their expressive power, which means that different modalities are suitable for presenting different types of information. Based on the literature, a set of guidelines on information type is summarized in Table 2.2. Note that the modality-information mappings presented in this table should be interpreted as "modality M is known to be particularly suitable/effective for presenting information type T", but not "modality M can *only* present information type T".

Table 2.2: Modality allocation guidelines on information type.

Modality	Information Type
Image & Icon	Concrete information with certain visual properties like shape, size, or color (e.g. physical objects, persons, places) [21; 98]
Maps	Location information about objects relative to one another [21] Spatial layout of any kind [21]
Flow chart	Relationships/steps involved in a process [277]
Organizational chart	Hierarchical structure [277]
Bar chart	Difference between individual data points between variables [6; 82; 141]
Line graph	Trends in data over time [6; 82; 141]
Grouped bars & lines	Summarization of a large amount of data [6; 82; 141] Difference between patterns of several variables [6; 82; 141]
Pie chart	Proportions and part-to-whole relations [6; 82; 141]
Scatter plot	Variability in data [6; 82; 141]
Color	Cues for aiding visual search [200] Indication of state [200] Relevant differences in quality or quantity [200]
Text	Abstract concepts, logic, quantitative values, relations [21; 98]
Speech	Abstract information [21] Instructions [238] Rapid communication of complex, multidimensional information sources [238]
Sound	Direction, location, and movement [176] Rapid cue of critical information, such as for warning and alarms [92; 101; 212]
Vibration	Alerts [124; 261] Direction and spatial location [100; 101; 137]

2.4.3 Guidelines on Modality Combination

Nowadays most user interfaces are multimodal. Combining multiple modalities in one presentation can enhance the robustness of communication and increase the bandwidth of information transfer [203; 216]. However, using multiple modalities is not always advantageous, especially when the combined modalities redundantly convey the same information. The redundancy may induce additional costs in terms of perception load, interface management and monitoring demand. The following guidelines suggest when to combine multiple modalities [216].

Combine multiple modalities when:

- The combination enhances human cognitive and physical ability and is compatible with user preference, context and system functionality
- The combination improves efficiency, satisfaction, or other aspects of performance for a given user and context
- The task domain is data-rich and complex, and time sharing between multiple information sources needs to be supported. (In contrast, uni-modality can be preferred in less demanding environments that involve few and simple self-paced tasks, in order to avoid the potential costs associated with multimodality.)

In addition, some researchers proposed that designers of multimodal presentation could adopt a "lessons-learned" approach [21], that is to fall back on the literature and analyze the "good compounds" which have been proved useful for a broad range of specified purposes. At least, when using a modality combination that has been certified as a good one under particular circumstances, developers will know that they are not venturing into completely unexplored territory but can make the best use of what is already known about their chosen modality combination ([21], pp. 145). Here we discuss two groups of good compounds (modality combinations).

First, the combination of verbal and non-verbal modalities is one of the well-known good compounds [20; 191]. 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. These complementary features bring the combination superior expressive power. Moreover, the cognitive benefit of this combination is the most pronounced when the verbal modality is auditory and the non-verbal modality is visual, or vice versa. This can be explained by the multiple resource theory which is introduced later in Section 3.5. Several examples of verbal and non-verbal modality combinations and their benefits are listed below.

- Adding text to graphics can improve the comprehension of the graphics [79].
- Speech may complement graphic displays for ease of visual inspection [21].
- Map-based displays should be combined with spoken interaction [245].
- Virtual displays benefit from being coupled with non-speech audio [245].
- Adding gesture to speech offers speed, high-bandwidth information transfer, flexibility, enhanced error avoidance, and relative ease of use [238].

The second group of good compounds includes visual-auditory combinations (VA) and visual-tactile combinations (VT). Regarding VA, one often-employed combination is the use of an auditory alert, followed by the visual presentation of relevant information [59; 61; 170]. For example, route guidance systems for cars use an auditory signal to notify the driver of an upcoming turn, and a visual display then provides more detailed information about the turn. A redundant use of auditory and visual modalities (they present the same information simultaneously) has also been found to improve user's reactions to warning messages [36; 142]. Regarding VT, tactile cues are often used together with visual information to aid localization or indicate directions. For example, the authors of [103] used tactile cues to notify drivers of rapidly approaching vehicles. Tactile signals were provided either on the front or on the back of the driver's torso, indicating in which direction drivers should look out for the vehicle. VT have been shown to induce faster reactions and impose lower cognitive load than using either single modality alone. For a set of design guidelines on VT, see [95]. Finally, the three good compounds in this group are summarized below.

- Use an auditory alert, followed by the visual presentation of relevant information
- Use auditory and visual modalities redundantly to present warning messages
- Use tactile cues with visual information to indicate directions or aid localization

2.5 The Underlying Rationale

The modality allocation guidelines discussed in the previous section provide a useful reference to the design of multiple user interfaces. However, we would like to obtain a deeper understanding of the rationale underlying these guidelines. For example, what makes it better to present warning messages redundantly by both auditory and visual modalities? What makes it better to use image rather than text for presenting concrete concepts? The answers lie in the way in which human cognition functions, because the receiver and processor of information is the human brain. Indeed, a large number of psychological, neuropsychological, and biological studies have revealed that modality plays an important role in various stages of human information processing, due to the modality-specific features of human cognition. Therefore, the rationale underlying good or bad modality choices is their cognitive compatibility.

Existing studies on the design of IMMP are often disassociated from knowledge of human cognition (e.g. the studies presented in Section 2.3). However, the need for integrating cognitive knowledge into IMMP has already gained awareness in recent years [203; 216; 238]. Regarding modality allocation, this means that allocation rules and guidelines need to be developed in the light of human cognition. Cognitive-aware modality allocation can be particularly beneficial in high-load task environments, because users may not have much spared cognitive capacity to cope with additional load that is unnecessary for the task. Consequently, suboptimum presentations can cause cognitive overload and harm the task performance. In the next chapter, we present relevant cognitive theories and findings describing the role of modality in various stages of human information processing. They provide a theoretical support to cognitively-compatible modality allocation.

3 Modality and Human Cognition

In this chapter, we first present a conceptual model of human information processing in Section 3.1. This model has great utility in understanding the functionality of the human brain. It also provides a useful framework to further discuss the modality-specific features of human cognition. The following three sections (Section 3.2 $\sim$ 3.4) describe the role of modality in sensory processing, perception and working memory (cognition), respectively. Then, Section 3.5 presents the multiple resource model which is a modality-related theoretical framework for predicting multiple-task interference. Afterwards, Section 3.6 shows the utility of these cognitive theories in the automotive context. Given the requirements of the driving task, the cognitive properties of a modality, together with other influences, determine its advantages and disadvantages for in-vehicle information presentation. Finally, a brief summary of this chapter is given in Section 3.7. The contents of this chapter has been partially published in [37] and [41].

3.1 Model of Human Information Processing

A conceptual model of human information processing was proposed by Wickens [279]. As Figure 3.1 shows, the model represents human information processing as a series of stages (boxes) connected by the flow of information (arrows). Attention as a resource is allocated to most of the stages. This model has great utility in understanding the functionality of the human brain. However, it should not be taken as a representation of the structure of the brain. Although the human brain does have a modular architecture [46], there is no clear-cut separation of areas in correspondence with the processing stages. More details on the processing stages are given below.

Sensory processing

In this stage, information from the environment is received by the brain as raw sensory data that can be processed by the brain. Sensory processing is pre-attentive, meaning that we are not consciously aware of what is being sensed at this stage. Raw sensory data is stored

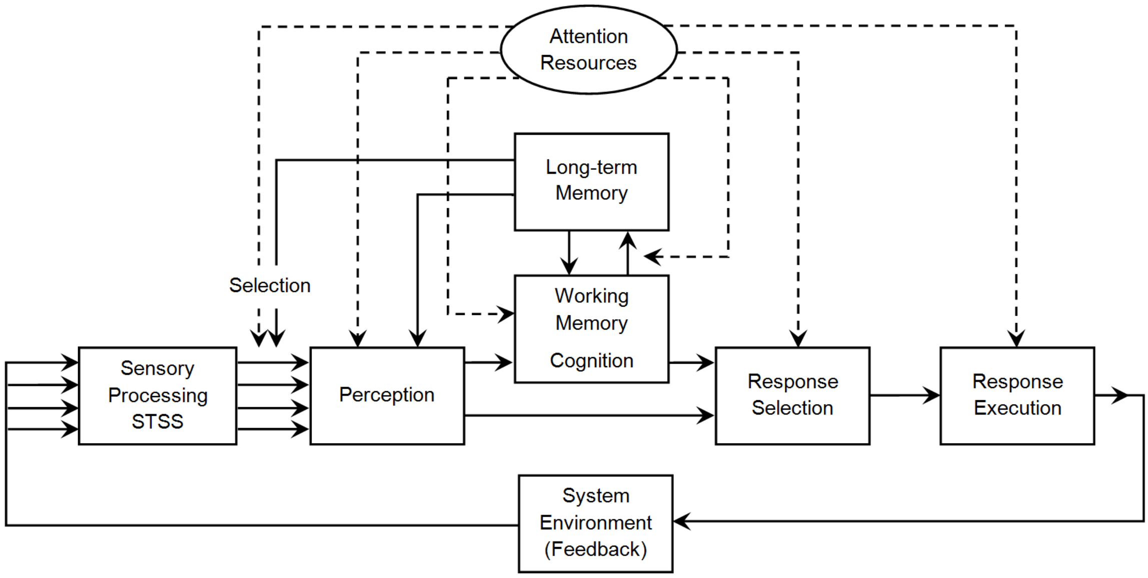


Figure 3.1: A model of human information precessing stages proposed by C.D. Wickens (reproduced from [279], p. 11).

for a very short time in short-term sensory stores (STSS). In fact, the capacity of STSS is much larger than the capacity of cognitive processing (working memory in particular), which means that we cannot and do not need to consciously process everything that is being sensed. For example, when a person is sitting on a chair, his/her body is constantly sensing the physical support of the chair, but usually he/she does not need to constantly think about the fact that he/she is sitting on the chair.

Perception

In this stage, a subset of incoming sensory information is selected by attention and interpreted in the context of previous experience (information in long-term memory). For example, a person interprets a brown object in a visual scene as a table. Information in long-term memory plays a large role in perception. If a person has never seen a table before, he/she will not be able to perceive a table as a table. Due to different experiences, the same sensory input can be perceived differently by different people.

Working memory & long-term memory

Working memory is the major processor of information, where complex cognitive operations take place, such as reasoning, calculating, comparing, comprehending, judging, et cetera. At a given moment, the information contents in working memory are what a person is consciously aware of. Working memory can store contents into long-term memory. However, contents in long-term memory are unconscious unless retrieved back into working memory again. Moreover, working memory has limited capacity. The influential work of Miller states that the capacity of working memory is limited to 7 +/- 2 items (information chunks) [168]. In contrast, the capacity of long-term memory is believed to be unlimited. At least no proof of such a limitation has been found so far.

Response selection

In this stage, decisions are made on how to respond to the environment. Two types of response can be distinguished: automatic response and controlled response. Automatic responses, represented as the arrow from perception to response selection, are procedures in long-term memory that execute automatically in response to a certain stimulus. They are the natural reflexes (e.g. withdrawing one's hand when touching hot water) or highly-trained skills (e.g. walking). Automatic response selection is fast and requires little attention. In contrast, controlled response selection, shown by the arrow from working memory to response selection, is based on the outcome of cognitive processing. It is slower and requires more attention resources, compared to automatic response.

Response execution

In this stage, the motor movements of the selected response are executed. Response execution is separated from response selection for two reasons. First, selected responses do not necessarily have to be executed. For example, when a person runs away from danger in his/her dream, the running response is usually not executed. Second, errors can occur during response execution, so that a wrong response is executed instead of the intended one.

System/environment feedback

The feedback loop from the system/environment to sensory processing indicates that the human response to the environment can be observed again. This feedback loop makes it possible to keep adjusting the response in order to reach a certain goal or maintain a certain performance. It is crucial to numerous tasks in daily life. When putting a cup into the coffee machine, the distance between the cup and the target place in the machine is a feedback that is continuously observed and used to direct the hand movement. When driving a car, the position of the car on the road is continuously observed and used to steer.

Attention

Attention is not a processing stage, therefore it is not surrounded by a box. Attention has two types of role in human information processing. The first role is the "attentive searchlight", because only the sensory information that gets attention can be passed on to later stages and be consciously processed. The second role is the resources or "fuel" of information processing. As shown by the dotted arrows in the model, attention resources are needed in nearly all processing stages. The capacity of attention is known to be limited [278], therefore the number of stimuli or tasks one can attend concurrently is limited. Attention limits also apply to the decision making and response execution stages.

The modality of information presentation mostly influences three stages of processing: sensory processing, perception and working memory. The response selection stage has not been explicitly related to the use of modality in the literature, because the response to an event (a decision) is mostly based on the output of cognition activities in the working memory, rather than based on the modality of information input. The response execution stage is modality-specific because different modalities are generated by different parts of the body, such as the hands for tactile responses and the vocal organs for speech responses. From a HCI perspective, these modalities are input rather than output modalities, and thus are outside the focus of this dissertation. In the remainder of this chapter, we will describe the role of modality in sensory processing, perception (selective attention from sensory processing to perception) and working memory (cognition).

3.2 Modality and Sensory Processing

At the sensory processing stage, the distinction of modalities is physically determined, because the five human senses are realized by different sensory receptors. The receptors for visual, auditory, tactile, olfactory and gustatory signals are found in the eyes, ears, skin, nose and tongue, respectively. Each sensory receptor is sensitive to only one form of energy. The function of these receptors is to transduce the physical energy into electrochemical energy that can be processed by the brain. All senses have an associated short-term sensory store (STSS) to temporarily store incoming sensory information. The duration of visual STSS is around 0.5 seconds, and the duration of auditory STSS is 2-4 seconds [279].

Tactile sensing is unique compared to the other four senses. First, skin is not a localized sensory organ like eyes, ears, nose, and tongue. Therefore, the sense of touch operates over a considerably wider surface (the whole body surface) [55]. Furthermore, tactile sensing is not simply the transduction of one type of physical energy into one type of electrochemical signal. This is mainly because the sense of touch consists of many forms, including the detection of temperature, texture, shape, force, friction, pain, et cetera [56]. The relation between these different forms of tactile sensing is not clearly understood.

3.3 Modality and Perception

Sensed stimuli do not have to be consciously attended to or actively interpreted. Instead, attention is needed to select certain raw sensory data to be perceived and interpreted (the "searchlight" role). This selection process is referred to as 'selective attention' [114]. Modality plays an important role in selective attention, because modalities with different sensory properties vary in their abilities to attract attention. The majority of selective attention research over the past 50 years has focused on auditory and visual modalities [232]. Only the last decade has seen a rapid growth of research on tactile attention. Research findings on visual, auditory, tactile, and cross-modal attention are presented next.

3.3.1 Visual Attention

Visual attention guides what we are looking at. The visual field is divided into foveal and peripheral fields. Only foveal vision is able to observe details of objects, but it has a very limited angle of only about 2 degrees. Therefore, without foveal visual attention, people often have surprising difficulty in detecting large changes in visual scenes - a phenomenon known as 'change blindness' [223; 224]. Peripheral vision is sensitive to motion and luminance changes. Visual attention can be directed in a top-down manner or a bottom-up manner [53]. The top-down manner means that visual attention is consciously directed by top-down knowledge, such as task-dependent goals [177], contextual cues [49; 197], current items in the working memory [64; 108] and expectations of what to see [235]. In contrast, the bottom-up manner is driven by salience, meaning that the visual stimuli which win the competition for salience will *automatically* be attended. When an object in a visual field contains some unique features, this object seems to "pop out" and captures the attention [111]. Through the bottom-up mechanism, attention shifts can be influenced by how the visual information is presented. Items that have higher priority should be presented with a unique (compared with surrounding) color, shape, intensity, orientation, depth, size, or curvature [190; 288].

3.3.2 Auditory Attention

Regarding attention attraction, the auditory modalities are different from the visual ones in three aspects. First, auditory modalities are more salient than visual modalities. Usually, attention is promptly directed to an auditory signal upon the onset of its presentation [235]. This feature makes auditory modalities a preferred choice to present information with high priorities, such as warnings and alerts [239]. The risk of using auditory modalities is that they might interrupt an ongoing task by pulling full attention away from it, referred to as 'auditory preemption' [282]. Second, unlike visual information which needs to be in the visual field in order to be attended to, auditory information can grab attention no matter which direction it comes from, and its direction can be distinguished if perceived by both ears. This feature makes it possible to assist visual search by providing location cues via auditory modalities. For example, it was demonstrated in [16] that 3D audio information could indeed assist pilots to locate outside-the-window visual targets faster. Third, auditory information is transient if no repeat mechanism is added to it. Therefore, it is force-paced, meaning that in order to be fully perceived, attention needs to be held on to an auditory stream during its presentation. In contrast, static visual information tends to be more continuously available and thus offers more freedom of perception in terms of time [279].

3.3.3 Tactile Attention

The presentation of a tactile stimulus can lead to a relatively automatic (or stimulus-driven) shift of attention towards the location of the stimulus (on the body) [232]. Attention can also be directed to tactile modalities voluntarily (i.e. expecting a tactile stimulus). Findings of behavioral studies and ERP (event-related brain potential) studies have both shown that attending to tactile modality (before a stimulus) can speed up the perception of the arrival of a stimulus and improve the discrimination of the stimulus [269]. Once attention is focused on tactile modality, people find it harder to shift it away than they do to move the focus of their attention away from the auditory or visual modalities [235; 236; 255].

Tactile attention has been found to have several limitations [232]. First, similar to "change blindness" of vision, tactile perception has a "inhibition of return (IOR)" effect, which means that the presentation of a tactile stimulus can slow down the processing of other tactile stimuli presented to the same or other parts of the body [51; 251]. This effect can last for a particularly long time (i.e., several seconds) after a stimulus is delivered. Second, people have extremely limited ability to simultaneously attend to and process multiple tactile stimuli presented over the body surface or even across the fingertips. For example, people are surprisingly bad at counting tactile stimuli when presented at more than 2 or 3 places over the body surface [85; 205]. Finally, when using a tactile motion stream (a series of moving tactile stimuli) across the body surface to present directional information, the perceived direction and speed of motion can be dramatically altered by any visual object that happens to be moving at around the same time [18; 146].

3.3.4 Cross-modal Attention

In real-life situations, attention often must be simultaneously coordinated between different senses - a fact that motivated the development of a relatively new research topic – cross-modal attention [231]. It has been proved that a shift of attention in one modality towards a certain spatial location tends to be accompanied by corresponding shifts in other modalities towards the same location [66; 83]. Such cross-modal links can operate in a reflexive (automatic) manner or a voluntary (controlled) manner. The reflexive manner means that an irrelevant but salient event in one modality tends to attract attention towards it in other modalities as well. Such reflexive links have been found for many modality combinations. For example, a salient auditory event (e.g. a loud bang) can generate rapid shifts of visual attention towards its direction; a tactile event on one hand (e.g. being touched) can generate shifts of visual and auditory attention towards the location of the touch. Cross-modal links can also direct attention voluntarily. When a person strongly expects an event in one modality at a particular location, his/her sensory sensitivity improves at that location not only for the expected modality but also for other modalities, even if there is no motivation to expect events from other modalities to occur at that location [230]. The cross-modal attention shifts have been supported by electrophysiological evidence from ERP studies [72; 73]. There might be a single cross-modal attentional system that operates independently of sensory modality and controls shifts of spatial attention for all senses. In summary, spatial attention towards a location typically spreads across modalities, and this finding has implications for multimodal information presentation to better support attention management in complex and data-rich interface applications.

3.4 Modality and Working Memory

Working memory operates in a modality-specific manner. The two theories presented below state that information contents carried by different modalities are processed differently in working memory.

3.4.1 Working Memory Theory

In 1974, Baddeley and Hitch proposed a three-component model of working memory, which has been well supported by scientific evidence from cognitive psychology, neu-

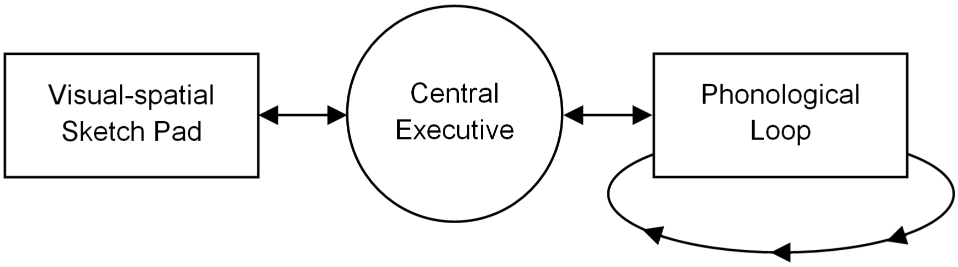


Figure 3.2: The working memory model by Baddeley and Hitch (from [14]).

roimaging and anatomy [13; 14]. According to this model, working memory contains a central executive system aided by two subsidiary systems – a visual-spatial sketch pad and a phonological loop (Figure 3.2). The phonological loop has a phonological store for temporarily storing auditory information. It also includes a rehearsal system. Auditory traces within the store are assumed to decay over a period of about two seconds unless they are refreshed by the rehearsal system. Particularly, the rehearsal system relies on speech coding to maintain the memory trace, meaning that information is usually rehearsed in the mind via sub-vocal speech [12]. The visual-spatial sketch pad is assumed to temporarily maintain visual information and to form a relation between visual and spatial information. The information stored in the two subsidiary systems is retrieved by the central executive system, which is assumed to be an attentional system whose role extends beyond memory functions. As the name indicates, it is believed to be a processing and control system which is involved in attention management, learning, comprehension, decision making, reasoning, judgement and planning. Neuroimaging and anatomical studies have indicated that these three components of working memory are localized in different brain regions. There is clear evidence of the phonological loop being on the left temporoparietal region (an area of the brain where the temporal and parietal lobes meet). The visual-spatial pad is identified to be primarily localized in the right hemisphere. There is also evidence for a separation in brain components between spatial ('where') and object ('what') vision [169; 228]. There is the least agreement among research findings on the anatomical location of the cental executive. It seems that different executive processes are implemented by different brain components and many of them locate in the prefrontal cortex [13; 250].

The working memory theory only addresses visual and auditory sensory systems, because these two senses have been the most investigated in nearly all domains. However, since all five senses have a separated short-term sensory store, the working memory model could presumably have three more subsidiary systems in charge of tactile, olfactory, and gustatory information.

3.4.2 Dual Coding Theory

At about the same time that the working memory theory was proposed, Paivio proposed a dual coding theory which addresses another modality-specific feature of human cognition [187]. This theory assumes that cognition is served by two separate symbolic systems, one

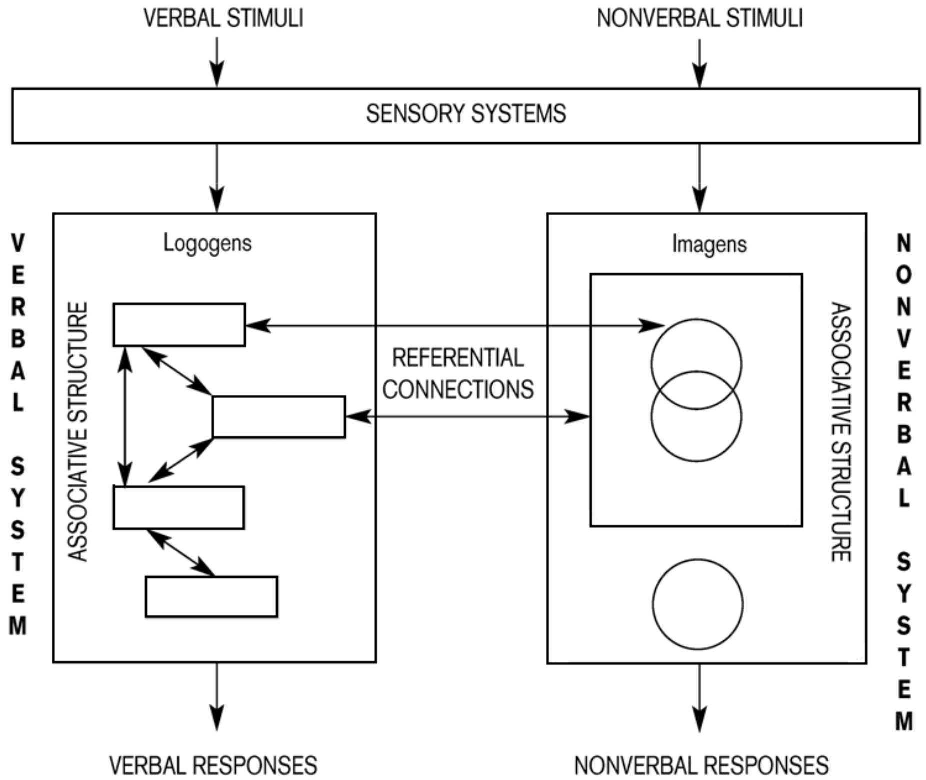


Figure 3.3: The dual coding theory by A. Paivio (from [188]). Logogens and imagens refer to verbal and nonverbal representational units respectively.

specialized for dealing with verbal information and the other with nonverbal information (Figure 3.3). The two systems are presumed to be interconnected but capable of functioning independently. The verbal system processes visual, auditory, and other modality-specific verbal codes. The nonverbal system processes images, environmental sounds, actions, and other nonverbal objects and events. The two systems are linked into a dynamic network through referential connections. The referential connections convert information between two systems and join corresponding verbal and nonverbal codes into knowledge that can be acted upon, stored, and retrieved for subsequent use. It has been demonstrated that the referential connections play major roles in various educational domains, such as knowledge comprehension, memorization, the learning of motor skills, etcetera [50]. Neuroimaging studies have provided support for the dual coding theory by showing that different parts of the brain are responsible for the passive storage and active maintenance of verbal, spatial and object information [226; 227].

3.4.3 Relating the Two Theories

The aforementioned two theories have not been explicitly related to each other by their founders. However, they are complementary instead of contradictory. It seems reasonable to assume that the central executive selectively retrieves information from modality-specific mental systems, integrates them into a unified percept, and then implements executive processes (reasoning, decision making, etc.). The central executive may also be responsible for the transfer of information between modalities. Since the rehearsal of information in the working memory is based on sub-vocal speech [12], rehearsing written materials during reading is an example of modality transfer from the visual to the auditory system. Another example is mental imagination of the appearance of an object upon hearing its name, demonstrating modality transfer from the verbal to the nonverbal system.

Furthermore, these two theories both provide theoretical support to the prediction of multiple-task interference. The working memory theory implies that the perception of visual and auditory information consumes separate attention resources. Therefore, two concurrent perception tasks interfere less with each other when they make use of different sensory systems than when they compete for resources of the same system. Similar conclusions can be made for verbal and nonverbal information based on the dual coding theory. In the next section, we describe the multiple resource theory – a framework that predicts multiple-task interference along four dimensions [280]. Two of the four dimensions are based on distinctions in modality properties.

3.5 Multiple Resource Theory

The multiple resource theory proposed by Wickens is a theory of divided attention between multiple tasks [280; 283]. *Resource* refers to attention resources that are allocatable and limited. The concept of *multiple resources* connotes attention resources allocated to parallel, separate, or relatively independent processes (tasks). The first version of this theory proposes three dimensions that account for variance in time-sharing task performance. They are processing stages, sensory modalities, and processing codes. Each dimension has two discrete levels. Figure 3.4 shows a cubic structure illustrating the three dimensions and their levels. A fourth dimension, *visual channels*, was later added to the model. Visual channels is a nested dimension within visual resources and is not explicitly represented in the cubic structure. According to the multiple resource theory, two tasks that both demand one level of a given dimension will interfere with each other more than two tasks that demand separate levels on the dimension. Time sharing will be better if two tasks use different levels along each of the four dimensions. Next, the four dimensions are discussed in detail.

Processing stages (cognitive vs. response)

This dimension separates the attention resources used for cognitive activities (perception, cognition) from those underlying the selection and execution of responses. There will be substantial interference between resource-demanding perceptual tasks and cognitive tasks involving working memory, because perception and cognition are supported by common resources. Two tasks both involving response selection will also heavily compete for common response-related resources. However, a primarily response related task interferes less with a cognitive task. This dimension is supported by a large amount of empirical and physiological findings (for details see [283] p. 133).

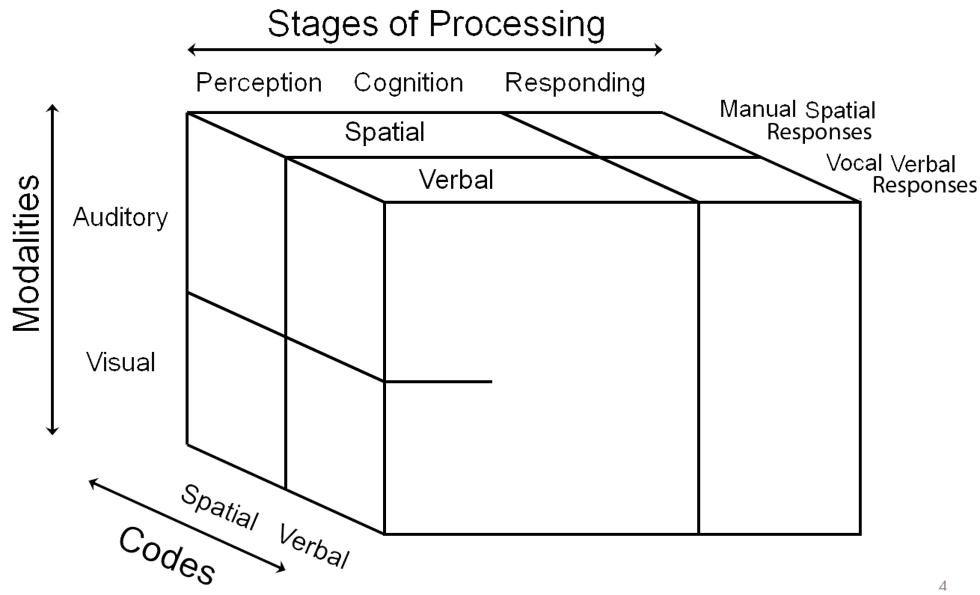


Figure 3.4: The structure of Multiple Resource Theory by C.D. Wickens (from [279]).

Sensory modalities (auditory vs. visual)

This dimension separates attention resources supporting visual and auditory information processing. Visual attention to a target does not restrict concurrent auditory attention to another target [4; 67]. As a result, auditory and visual information can be simultaneously processed better than two different auditory streams or two different visual signals. For example, when presenting route guidance for drivers, it is better to use auditory modalities than visual modalities, because the concurrent driving task mostly requires visual attention [189]. This distinction is said to be the most effective for perception. The working memory model from Baddeley (Section 3.4.1) provides theoretical support for this dimension.

Processing codes (verbal vs. spatial[2])

This dimension predicts that tasks can be better time-shared if they respectively involve verbal and spatial information processing and response. This prediction is effective at all stages. At the cognitive stage, the perception and processing of verbal and spatial information depend on separate resources, which is supported by the dual coding theory (Section 3.4.2). At the response stage, the verbal (vocal) and spatial (manual) responses also require separate attention resources. For example, manual dialing on cell phones interferes with driving much more than voice dialing [60].

Visual channels (focal vs. ambient)

This dimension is established due to the fact that focal and ambient (or peripheral) vision can be processed in parallel. For example, we can keep the car in the center of the lane

[2]The notion 'spatial' used in this theory seems to be equivalent to the notion 'nonverbal', but a clear confirmation of this relation has not been found in the literature.

(ambient vision) while reading a road sign (focal vision). It was also found that different brain structures were activated when attention is deployed to different visual fields [180]. The distinction of focal and ambient channels is considered to be nested within the visual attention resources.

The multiple resource theory certainly has an impact on information presentation in human-computer interface design, especially regarding the allocation of modalities. In multi-task environments where multiple sources of information need to be presented to users at the same time, modality allocation should avoid or reduce overlap within each level of the 'sensory modalities' and the 'processing codes' dimensions. For example, while driving is a visual and spatial task, using an auditory and verbal modality (speech) to present route descriptions would allow better time-sharing of these two tasks. Apart from modality, the 'channel' dimension offers a solution of presenting multiple sources of visual information, that is to smartly locate them in both focal and ambient visual fields. Being smart in this manipulation means to take into account the different features and abilities of focal and ambient vision. Ambient vision is appropriate for detecting motion or luminance changes, but it does not support the recognition of objects or details [216]. Finally, one thing is worth keeping in mind when applying this model: it does not imply that two tasks using separate levels in any one dimension will foster perfect time-sharing. The amount of interference between two tasks will depend on the number of shared levels on *all* dimensions ([279], p.452).

3.6 Modality and Human Cognition in Automotive Domain

The cognitive theories and findings presented previously in this chapter provide a theoretical foundation for cognitive-aware modality allocation, but they are not rules or guidelines that can be directly implemented into a system. In fact, without an application background, they do not say anything about whether a modality is good or bad. However, once an application is given, the cognitive properties of a modality can be evaluated against the requirements of the application, resulting in advantages/disadvantages of this modality in this application. This section shows the utility of cognitive theories in the automotive domain.

In-vehicle information systems (IVIS) are intended to assist drivers by providing supplementary information, such as route instructions, traffic conditions, hazard warnings, vehicle states and so on. In a moving vehicle, the primary task of the driver is to watch the road and control the vehicle. The perception and comprehension of IVIS information[3] can be considered as a secondary task that needs to be performed concurrently with the driving task. Besides being helpful, IVIS messages also impose attentional cognitive demand and cause distraction to some extent. Such distraction has been identified as a cause of accidents [125]. Therefore, the requirements of in-vehicle information presentation is to be compatible with driving and to minimize unnecessary cognitive demand and distraction.

[3]The expression 'IVIS information' or 'IVIS message' refers to information or messages presented by IVIS to drivers.

Driver support has been a traditional research topic of Human Factors (or Cognitive Ergonomics), and thus has a part of its roots in cognitive and experimental psychology. As a result, existing studies on the use of modality for IVIS have a strong focus on the cognitive impact of modalities. Unlike the studies on IMMP described in the previous chapter, they focus more on developing modality allocation guidelines via experimental studies and less on building rule bases to achieve adaptivity. In this section, we provide a brief overview on the use of modality in IVIS, focusing on the advantages and disadvantages of three categories of modalities: visual, auditory and tactile. This section also serves as a related work background for our own studies in the automotive domain (Chapter 6 and 7).

3.6.1 Visual Modalities

Driving is a highly visual task, requiring almost continuous visual attention while the vehicle is in motion [133]. Two visual tasks cannot be performed in parallel due to the competition between perceptional resources in the visual channel ([280], see Section 3.5). This causes a drawback for visual IVIS presentations – they impose additional load on the visual perception channel and usually “drag” the eyes away from the road (except when the information is presented on the windshield). Another drawback of visual presentation is a lack of salience. The onset of a visual presentation might be overlooked if the driver happens to be looking somewhere else. That is to say, the information delivery is likely to be delayed or even unattended.

The advantage of visual presentations is their self-paced feature [219]. They allow the driver to inspect them at his/her own pace (e.g. at once, step by step or selectively). They can also be viewed multiple times, which is particularly beneficial in a case where certain details need to be recalled at a later moment.

3.6.2 Auditory Modalities

When used in a driving environment, the major advantage of auditory modalities is that they consume a separate perceptional resource from driving ([281; 283], see Section 3.5). This means that drivers can watch the traffic and listen to the message at the same time. Therefore, compared to visual messages, auditory messages interfere less with driving and are less likely to cause mental overload [283]. Being omnidirectional is another advantage of auditory modalities – they can be picked up from all directions, independent of where the driver is facing ([216], see Section 3.3.2).

Auditory modalities are also highly salient, because attention is promptly directed to an auditory signal upon the onset of its presentation ([234], see Section 3.3.2). This feature can be an advantage in cases where an urgent message needs immediate attention. However, it can also be a disadvantage because the onset of auditory signals can easily grab attention away from an on-going visual task (driving in this case) and cause driver distraction. Moreover, auditory modalities are transient, and thus have a disadvantage when messages need to be recalled at a later point.

Speech inherits the common characteristics of auditory modalities. In addition, speech has been found to have a 'preemption effect' on driving [282], which means it keeps drivers' attention and temporarily suppresses driving. This effect gets stronger when the spoken sentence gets longer. The cause is twofold. First, due to a high salience, the onset of speech grabs drivers' attention. Then, because speech is transient and force-paced, drivers who want to capture the full message need to continuously attend to it during its presentation.

Auditory icons refer to familiar environmental sounds that imitate real-world events (originally defined in [87]). In IVIS, auditory icons are typically used as warning signals. For example, presenting a car horn sound or a screeching car tire sound can warn drivers of an impending collision [92; 101]. Besides the common characteristics of auditory modalities, auditory icons have their own unique features that are beneficial to IVIS information presentation. First, they are language independent and culture independent. Second, when well chosen (the more intuitive, the better [240]), they inherently convey the meaning of the events that they are meant to signify [123; 233]. In other words, the meaning of the events is immediately clear to the driver. This feature explains the common empirical findings that drivers reacted significantly faster to auditory icon warnings than to speech warnings [17; 92]. However, the use of auditory icons also has limitations [233]. There is evidence that the fast reactions may be accompanied by an increase in inappropriate responses [25; 92]. This is because drivers may react before they have properly evaluated the situation to know what the most appropriate response would be. Moreover, auditory icons are likely to be considered unpleasant, due to inappropriate loudness or high pitch [164]. Generally, sound warnings can cause annoyance when over-used [150].

3.6.3 Tactile Modalities

Compared to visual and auditory modalities, the use of tactile output is rather new in IVIS information presentation. However, existing findings have already shown promising results. Two tactile modalities have been investigated so far: force pulse (given by pedals [113] or steering wheel [247]) and vibration (given by seat [137; 261], steering wheel [124; 247] or additional equipment attached to the driver [101; 103]). Regarding information type, tactile modalities have been typically used as alerts and directional cues. For example, they have been applied to warn drivers of a rapidly approaching vehicle [102; 103], a sudden deceleration of the lead vehicle [100; 101; 137], or a lane departure [74; 247]. In most cases, multiple tactors were applied at different locations in order to add another dimension to the presentation – the direction drivers should look in for the event. For example, two tactors were used in [103], one on the front and one on the back of the driver's torso. Only one tactor was activated at a time, indicating a vehicle was approaching from either the front or the rear. In addition, 'left versus right' tactile cues have also been applied to indicate turning directions [260; 261] and lane change directions [124].

A commonly obtained finding is that tactile signals induced significantly faster reactions to the presented events in comparison with either the absence of tactile signals [100; 101] or the auditory/visual presentation alternatives [74; 102; 247; 260; 261]. This is mostly because tactile signals are highly salient and can almost always draw attention immediately

Table 3.1: Summary of the advantages and disadvantages of three categories of modalities for in-vehicle information presentation.

Modality Category	Advantages	Disadvantages
Visual modalities	- self-paced - allow repeated perception	- compete with driving for visual perceptional resource - have a lack of salience
Auditory modalities	- consume separated perceptual resource from driving - highly salient - omnidirectional	- force-paced - transient - preemption effect
Tactile modalities	- consume separated perceptual resource from driving - highly salient - private to the driver	- limited in expressive power - likely to induce annoyance and physical discomfort

(see Section 3.3.3). Besides, tactile modalities also have other advantages: 1) they do not compete with driving for visual perceptional resources; 2) their effectiveness is not influenced by the lighting condition, driving noise, radio or conversation; 3) they are private to the driver and do not bother the passengers.

Tactile modalities also have two limitations. First, they are limited in their expressive power, meaning that they are only suitable for a few types of information, such as alerts and directions. Second, tactile stimuli are likely to induce annoyance and physical discomfort. To minimize this negative effect, the duration and intensity of the signal should be carefully chosen (several suggestions can be found in [115]).

Table 3.1 summarizes the advantages and disadvantages of the three categories of modalities discussed above for in-vehicle information presentation.

3.7 Summary

The use of modality in information presentation has a cognitive impact, because modality influences several stages of information processing, namely sensory processing, perception and cognition. When multiple tasks need to be performed simultaneously, the modalities used by these tasks can influence (to some extent) how much they can be time-shared. The cognitive theories and findings presented in this chapter provide a theoretical foundation for allocating modality in a cognitively compatible fashion. However, without an application background, these theories do not directly tell the advantage or disadvantage of a modality. For instance, the 'force-paced' property of auditory modalities (Section 3.3.2) is neither good nor bad. However, in a driving environment, it becomes a disadvantage of speech messages (Section 3.6). Given an application, modality allocation rules can be developed by matching the cognitive knowledge to the specific requirements/goals of the application.

Part II

Information Presentation for Time Limited Tasks

4

Information Presentation for Time Limited Visual Search

In this chapter, we present a study investigating the modality effect on task performance, cognitive load and stress, using a high-load visual search task embedded in a crisis rescue scenario. The high task load is induced by time pressure and a high information load. The experimental results are predicted and interpreted based on several cognitive theories, such as the working memory theory (Section 3.4.1), the dual coding theory (Section 3.4.2) and the multiple resource theory (Section 3.5). We also demonstrate a way to integrate these theories into a computational model that systematically predicts the suitability of any modality combination for our presentation task. The contents of this chapter have been published in [39], [40] and [41].

This chapter is organized as follows. Section 4.1 presents findings of several multimedia learning studies regarding the cognitive effects of modality in knowledge presentation. Section 4.2 describes the experiment in detail. The experimental results are presented in Section 4.3 and discussed in Section 4.4. Section 4.5 presents the suitability prediction model and the generalization to other applications. Finally, the contributions and limitations of this study are discussed in Section 4.6.

4.1 Related Work

In educational psychology, scientists have been dedicated to developing learning materials that make knowledge easier to learn. The cognitive load of learning is defined as the sum of three components: intrinsic load, extraneous load and germane load. This definition is known as cognitive load theory [186; 248; 262]. Intrinsic load represents the difficulty of learning determined by the nature of the knowledge and the expertise of the learner. Germane load reflects the mental effort learners put into the learning processing. Extraneous load is the extra load resulting from poorly designed learning materials and is neither beneficial nor necessary to learning. A large number of studies have investigated how to design learning materials that reduce or eliminate extraneous load [156; 161; 162]. The working memory theory of Baddeley, the dual-coding theory of Paivio, and the multiple resource theory of Wickens have been successfully applied to explain empirical findings and generate design guidelines.

A series of studies found that knowledge was better presented both verbally and nonverbally rather than only verbally or only non-verbally. Better in the sense that learners gained a deeper understanding and were more able to transfer the knowledge into problem-solving. For example, Mayer et al. conducted lessons on how brakes work using printed materials [154; 159]. The knowledge was presented in either only text or text along with illustrations (pictures). After learning the material, students were asked to fulfill a problem-solving test, which included troubleshooting broken brakes or redesigning brakes to meet a new purpose. The results showed that students who had learned with text and pictures were much more creative and generated significantly more valid solutions than the students who had learned with only text. Similar effects were also found in computer-based learning environments, showing that students were better at problem-solving after learning when knowledge was presented by synchronized narration and animation rather than by only narration or only animation [157; 158]. The dual-coding theory was used to explain these findings. When both verbal and nonverbal materials were provided, both mental systems were involved in the learning process. Referential connections were actively built between associated verbal and nonverbal contents, which also accounted for a better learning. This explanation was supported by a later finding that people with high prior knowledge benefited less from multimodal presentation than people with low prior knowledge, because prior knowledge helped to create useful mental images solely from the verbal materials [155].

Another common finding was that when nonverbal information (illustrations, animations, diagrams, etc.) was provided visually (on paper or on screen), students learnt better when the associated verbal explanation was presented in narration rather than in on-screen text [31; 116; 160; 173–175]. This finding stands in line with the working memory theory and the multiple resource theory. When all information was presented visually, students had to divide their visual attention resources to verbal and nonverbal items, causing a so-called split-attention effect. Moreover, verbal information items needed to be held in the visual temporary store of the working memory in order to be processed together with related nonverbal information items, and vice versa. By replacing text with well synchronized narration, related verbal and nonverbal items could be perceived (attended to) and processed in parallel.

In summary, the three cognitive theories suggest a benefit from using multimedia for the design of learning materials. In this study, we intended to apply these theories to a high-load and time-critical task. Unlike learning, our task does not involve comprehension and long-term memorization. Instead, it requires fast processing of a large amount of information and quick response. Due to the different nature of the two tasks, the same theories might lead to different suggestions regarding the use of modality. The main objective of this study is to confirm that these cognitive theories are also relevant to the design of multimodal interfaces, because they provide a foundation for cognitively-compatible modality allocation.

4.2 Method

4.2.1 Scenario

The task in the experiment was based on the following scenario.

After a massive earthquake, a rescue team arrives at an affected inhabited area. Rescue workers search for injured people and transfer them to safe places. Using mobile communication devices, they report the location of wounded victims ('patients') and dead victims ('deaths') to the crisis response center. A crisis manager (played by a participant) in the crisis response center is monitoring the rescue progress on a display, where the locations of victims are presented. The task of the crisis manager is to direct a doctor to reach all wounded people and save their lives.

Note that it was not our goal to make the crisis scenario realistic, and participants were not required to have any experience in crisis management. The main purpose of using a crisis scenario was to better motivate the high-load and time-critical task.

4.2.2 Information and Presentation Conditions

We first defined the set of modalities and the type of information to be conveyed, and then selected presentation conditions based on the suitability evaluation of each modality for each information component. In correspondence with the cognitive theories, the modality set needed to contain visual, auditory, verbal and nonverbal properties. To allow all combinations of these properties, four modalities were selected, namely text (visual, verbal), image (visual, nonverbal), speech (auditory, verbal) and sound (auditory, nonverbal). When a victim was found, two categories of information could be reported to the crisis manager: basic information and additional aid (optional).

Basic information

The basic information of a victim report included two items: the location and the type (wounded or dead) of the victim. Maps are the most suitable for presenting locations (see Table 2.2), therefore an obvious solution was to present victims as visual objects on a map. In this case, text or images were used to convey victim types and their locations on the map indicated victim locations. We created a grid-based map which contained a 20×13 array of squares, and used it as the background of the entire display (see Figure 4.1). Text and image presentations of patients and deaths are shown in Figure 4.2. To ensure comparability, we used text words with the same font, size and color. The two images were also similar in color and shape. A grid could contain at most one victim at a time.

The two auditory modalities (speech and sound) were not selected as candidates to present basic information. Speech can refer to a location by a row index and a column index, or a zone index. Sound can use variations in tone, pitch or direction to convey location. However, since the grid-based map contained 260 location units (squares), using only auditory modalities without any explicit visual confirmation would be much too inefficient. It

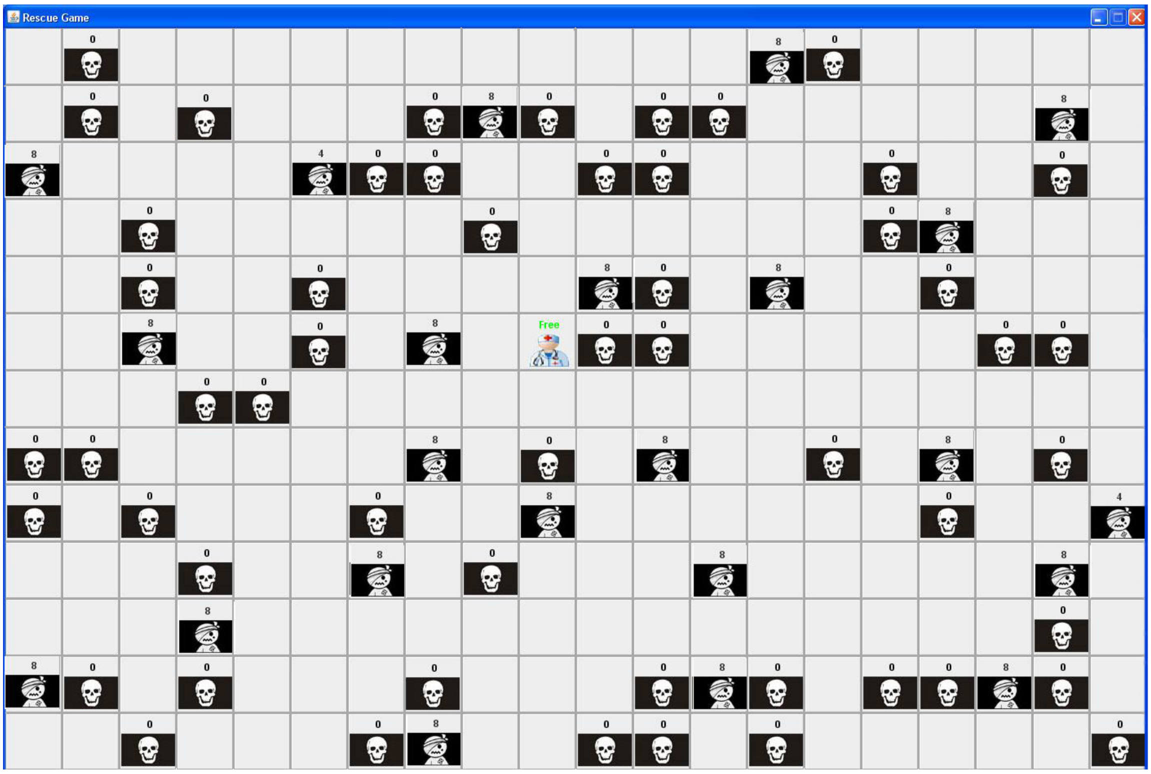


Figure 4.1: A screen shot of the grid-based rescue map.

Text	Image
Patient	
Death	

Figure 4.2: Text and image presentations of victim types.

would be particularly hard or even impossible for people to distinguish between 260 sound variations in order to decode locations.

Additional aids

The rescue area on the display was divided into two halves – the left half and the right half. The additional aid indicated which half of the area contained a newly-found victim. Additional aids were not provided in all conditions. When provided, they were always synchronized with the basic information, assisting the crisis manager to locate victims faster.

Table 4.1: Five experimental presentation conditions.

Index	Presentation Condition	Modality Properties
1	Text	Visual, verbal
2	Image	Visual, nonverbal
3	Text + Image aid	Visual + visual, verbal + nonverbal
4	Text + Spoken aid	Visual + auditory, verbal + verbal
5	Text + Sound aid	Visual + auditory, verbal + nonverbal

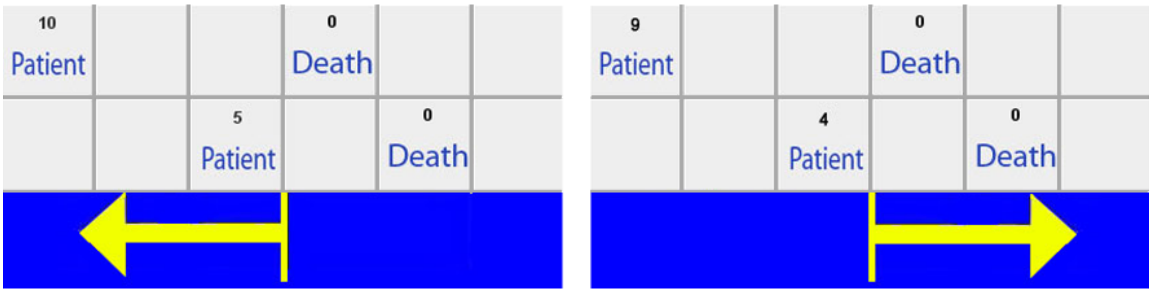


Figure 4.3: Left and right arrows used in the 'text + image aid' condition.

The additional aid was directional information in nature and could be presented by all four modalities. The two verbal modalities could directly present words 'left' and 'right', either in speech or as on-screen text. In the case of image, directions were conveyed by an arrow pointing to the left or the right. Human ears have the ability to detect the direction of a sound if it is received by both ears. Therefore, sounds were delivered from the left or the right side of the participant in order to indicate directions.

Presentation conditions

Two modalities were available for basic information (text and image) and four were available for additional aids. This resulted in 10 (4 × 2 + 2) potential modality combinations, from which we selected 5 combinations to be investigated (summarized in Table 4.1). We predicted that image would be better than text for presenting victim types, because the categorization and understanding of concrete objects is faster when they are presented by image than by text ([21], also see Table 2.2). Therefore, in order to better observe the benefit of additional aids and modality combinations, basic information was always presented by text when additional aids were provided (condition 3, 4, and 5). In condition 3, large-sized arrows in a salient color were presented at the bottom center of the screen to indicate the direction left or right (see Figure 4.3). In condition 4, spoken 'left' and 'right' were delivered by two speakers. One was located on the left side of the participant and the other one on the right side. In condition 5, an ambulance sound (duration 1 s) was played by either the left speaker or the right speaker.

4.2.3 Task and High-load Manipulation

The participant played the role of a crisis manager, whose task was to send the doctor to each patient by mouse-clicking on the patient's representation (text or image). The doctor was represented by an icon on the grid map. The doctor could be in two states: free or occupied (busy treating a patient). When the doctor was free, he responded to a mouse click immediately and moved ("jumped") to the patient's location without delay. It took a fixed duration of 1 s to treat a patient and the doctor could not respond to any new task during this time.

New patients appeared at random intervals of 2 to 5 s. Usually a patient was presented at the same time as one or more dead victims. The lifetime of a patient was 10 s when presented and decreased second by second. The remaining lifetime of a patient was indicated by a number above the representation of the patient (in the same square). After a timely treatment, patients disappeared from the map. Otherwise, they would turn into a dead victim. When a patient died, a speech warning was delivered, announcing: "A patient died!". This performance feedback was included to urge participants to do their best on the task.

A task trial lasted for about 5 minutes in which 100 patients were presented. Dead victims served as distracters that required no reaction. Therefore, the nature of this task was a high-load visual search with a time limit. The optimum performance could be achieved by locating and treating the patients in their order of presentation.

A high task load was induced by three factors: 1) time limit. Reaction time to a patient was limited to 10 s; 2) presentation intensity. Patients were presented every 2 to 5 s; 3) number of distracters (dead victims) on the map. At the beginning of a trial, the grid map was empty and the task was relatively easy. As the number of dead victims grew, it became more and more difficult to identify a patient in the crowded surroundings. The task difficulty reached the maximum (when about 40% of the cells contained victims) after about 150 s and remained unchanged for the rest of the trial. The current task setting aimed to induce high mental workload but not to make the task too difficult so that participants gave up doing their best. Based on a pilot study, at least 80% of patients could be saved if the participant was fully engaged in the task.

4.2.4 Measures

Three categories of measures were applied to assess cognitive load and stress: performance, subjective, and physiological measures.

Task performance was assessed by three measures. *Reaction time* (RT) measured the time interval between the moment when a patient was presented and the moment when the doctor was sent (in seconds). *Number of dead patients* (ND) referred to the number of patients that were not treated within 10 s and died. *Time of the first patient death* (TFD) measured the time interval between the start of a trial and the moment when the first patient died in the trial (in seconds). Since the number of distracters increased gradually in the first half of a trial, TFD actually reflected how tolerant the performance was against the increase of task difficulty.

Table 4.2: Description of the physiological measures applied in this study.

Category	Measure	Unit	Description
Heart period	HP	ms	The time interval between two successive heart beats
Heart rate variability	LF	ms^2	The total spectrum power of a heart period series at band 0.07$\sim$0.14 Hz (frequency domain)
	RMSSD	ms	The square root of the mean squared differences of successive heart periods (time domain)
	NN50	-	The number of internal differences of successive heart periods that are greater than 50 ms (time domain)
Skin conductance	GSRN	-	Number of event-related skin conductance responses
	GSL	μS	The tonic level of skin conductivity
Respiration	RP	s	Time interval between two successive respiration peaks

Two subjective measures were applied, namely *subjective cognitive load level* (SCL) and *subjective stress level* (SS). We applied a rating scale with 21 gradations like the NASA Task Load Index (see Appendix A), therefore cognitive load and stress could respectively be rated from 1 (very low) to 21 (very high).

In order to assess cognitive load from the physiological state of the participants, we recorded the electrocardiogram, galvanic skin conductance and respiration during the experiment. Previous findings have suggested that when cognitive demand increases, heart rate increases (heart period decreases), heart rate variability decreases, the skin conductivity increases and the respiration rate increases ([28; 130; 217; 267], also see Section 1.4). A total of 7 measures were derived from the physiological recordings, a detailed description of which is given in Table 4.2.

4.2.5 Apparatus and Setup

Two PCs were used in the setup, as shown in Figure 4.4. PC-1 hosted the crisis rescue interface. User performance was logged as text files on this computer. Three flat-type active electrodes were placed on the participant's torso to record the electrocardiogram. Galvanic skin conductance was measured by two flat-type passive electrodes placed on two finger tips of the non-dominant hand. A respiration belt was tied around the participant's waist (on top of clothing) to measure respiration. A USB receiver received all channels of physiological signals and event triggers. Event triggers were sent by the rescue program on PC-1 at the beginning and the end of each trial, via a parallel-port connection. Eventually, synchronized physiological signals were recorded by the ActiView software hosted on PC-2. All sensors, devices, and software used for physiological recording were products of BIOSEMI[1].

[1] BIOSEMI: http://www.biosemi.com/

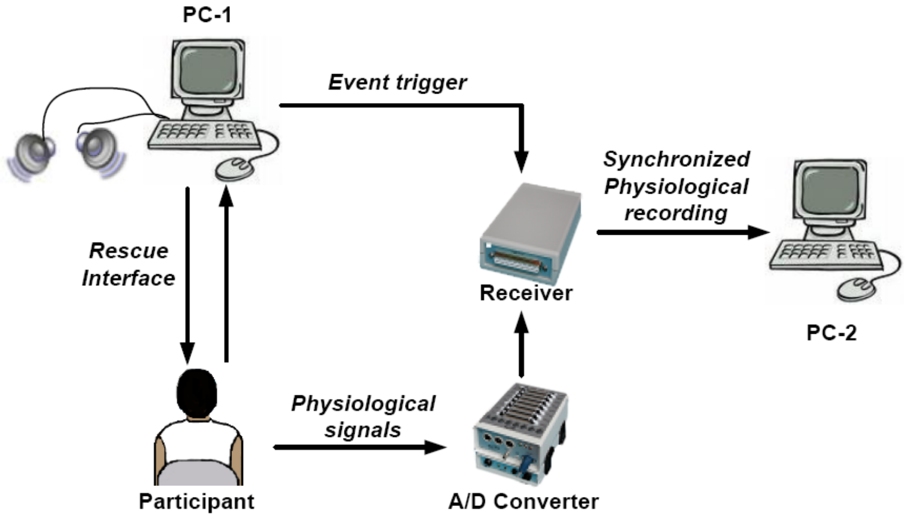


Figure 4.4: The experiment setup

4.2.6 Participants, Experimental Design and Procedure

Twenty people participated in this experiment (15 men and 5 women). Their age ranged from 22 to 32. They were all university students (bachelor, master, or PhD) and daily computer users. With a within-subject design, each participant performed the task in all 5 presentation conditions. The trial order was counter-balanced with a size 5 Latin square[2].

After entering the laboratory, a participant took a seat in front of PC-1. First, electrodes and sensors were placed on the participant and signal validity was checked. Then, he/she listened to soothing music and looked at peaceful nature pictures for about 10 minutes. This manipulation was supposed to make the participant relax. After the relaxation session, a baseline physiological state was recorded for a period of 5 minutes. Then, the scenario and the task were introduced to the participant and he/she practised the rescue task in all 5 presentation conditions (1 minute per condition). Afterwards, the participant performed 5 task trials with a short break between each two trials. Questionnaires were filled during the breaks. The whole experiment lasted for approximately 80 minutes.

4.2.7 Hypotheses

Based on the working memory theory and the dual-coding theory, we constructed the following four hypotheses.

1. The image (nonverbal) condition is superior to the text (verbal) condition in terms of better performance, lower cognitive load and lower stress, because image is better than text for presenting concrete objects.

[2]A Latin square is an n $\times$ n table filled with n different symbols in such a way that each symbol occurs exactly once in each row and exactly once in each column.

2. The auditory aids (speech and sound) are superior to the visual (image) aid, because they can be better processed in parallel with the visual rescue task.
3. The nonverbal aids (image and sound) are superior to the verbal aid (speech). Directional information is nonverbal in nature. When presented with nonverbal modalities, it does not require referential connections to be interpreted. Our task requires quick perception and response rather than deep comprehension of complex knowledge. Therefore, building referential connections between verbal and nonverbal information might unnecessarily slow down the process and induce extra cognitive load.
4. Conditions with additional aids (multimodal presentations) are superior to conditions without aids (uni-modal presentations), because participants receive more useful information which is meant to assist task performance.

4.3 Results

Due to the within-subject design, we applied one-way repeated-measure ANOVAs on all measures, where the independent factor was *modality* with five different levels (the five presentation conditions). Results are presented in this section.

4.3.1 Performance

Reaction time (RT)

The average reaction time of all trials is shown in Figure 4.5 (left). On average, it took subjects between 1.9 s and 3.1 s to locate any patient. The reactions were the fastest in the 'text + spoken aid' condition and the slowest in the text condition.

ANOVA results revealed a significant modality effect on reaction time, $F(4, 16) = 12.8$, $p < .001$. Post-hoc tests (with Bonferroni corrections) were then conducted for pairwise comparisons. Significant differences in reaction time were found between the 5 condition pairs. The reaction was faster in the 'text + spoken aid' condition than in the text, 'text + image aid', and 'text + sound' conditions. The reaction was faster in the image condition than in the text and 'text + image aid' conditions.

Number of dead patients (ND)

On average, the number of dead patients in each condition was between 2 and 12 (Figure 4.5, right). As 100 patients were presented in each trial, the percentage of saved patients was between 88% and 98%. The most patients were saved in the 'text + spoken aid' condition, and the least were saved in the text condition.

ANOVA results indicated that there was a significant modality effect on the number of dead patients, $F(4, 16) = 16.8, p < .001$. Pairwise comparisons showed 5 significant effects. More patients died in the text condition than in the image, 'text + spoken aid' and 'text + sound aid' conditions. More patients died in the 'text + image aid' condition than in the image and 'text + spoken aid' conditions.

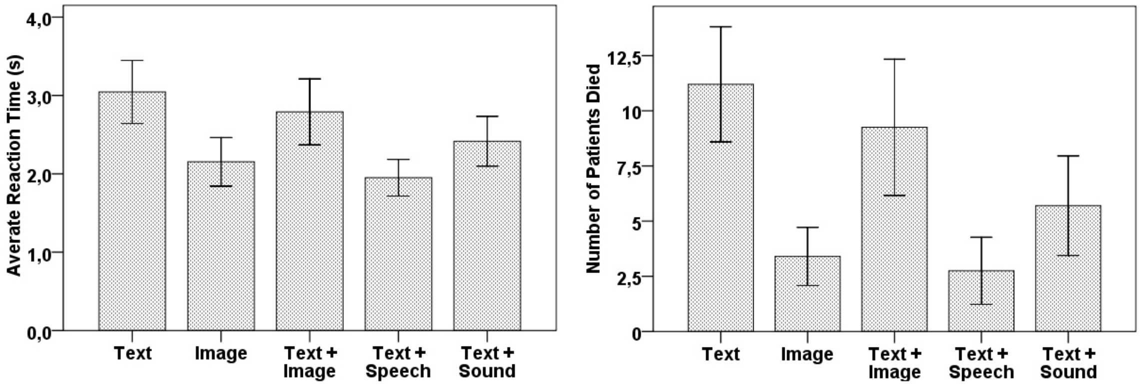


Figure 4.5: Average reaction time (left) and number of patients that died (right) in five presentation conditions. Error bars represent standard errors.

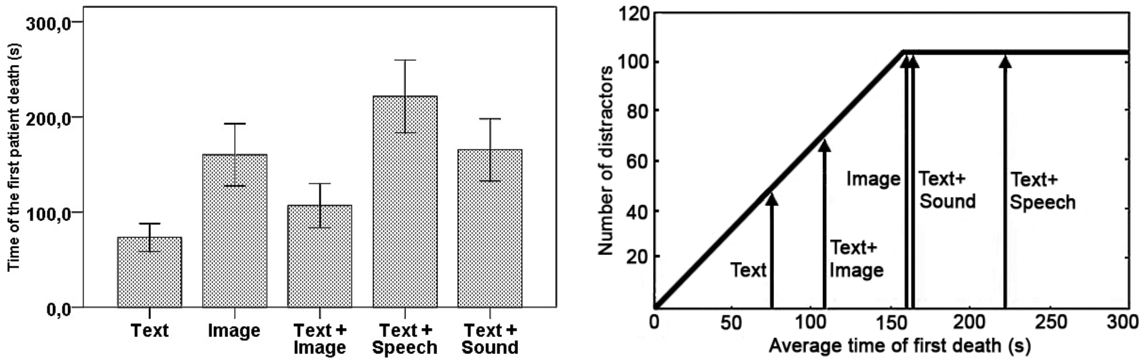


Figure 4.6: Time of the first patient death. Left: average TFD in all modality conditions. Error bars represent standard errors. Right: average TFD shown on the curve of task difficulty over time.

Time of the first patient death (TFD)

As Figure 4.6 shows, the first dead patient occurred the earliest in the text condition (at the 73rd second on average) and the latest in the 'text + spoken aid' condition (at the 221st second on average). Again, ANOVA revealed a significant modality effect on this measurement, F (4, 15) = 17.7, $p < .001$. According to post-hoc tests, the first patient death occurred significantly earlier in the text condition than in the image, 'text + spoken aid' and 'text + sound aid' conditions. The first patient death also occurred significantly earlier in the 'text + image aid' condition than in the 'text + spoken aid' condition.

The effects found from this measurement actually indicate that the use of modality significantly affected how tolerant the performance was against the increase of task difficulty. As Figure 4.6 (right) shows, in the text condition, the performance dropped when the task difficulty increased to about half of the maximum. In contrast, in the 'text + spoken aid' condition, the good performance was maintained for more than 50 s after the task difficulty reached the maximum.

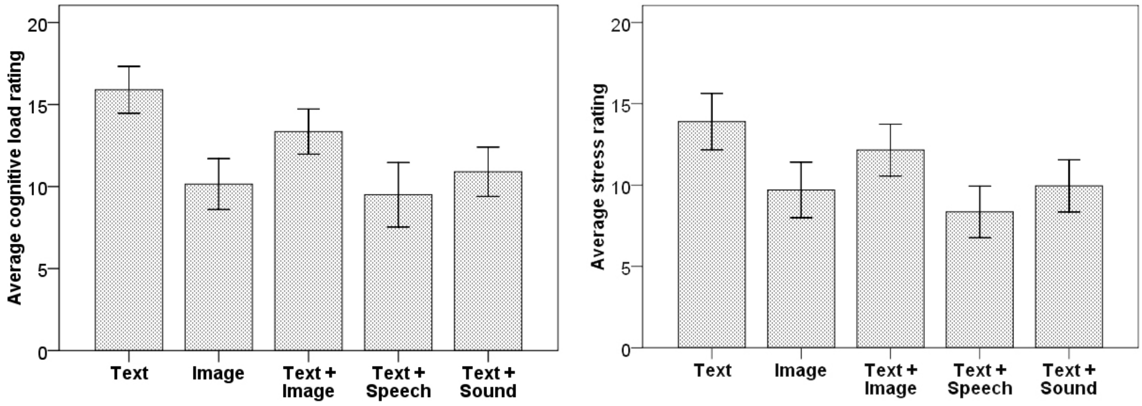


Figure 4.7: Average subjective rating scores on cognitive load (left) and stress (right) in five modality conditions. Error bars represent standard errors.

4.3.2 Subjective Cognitive Load and Stress

Subjective cognitive load level (SCL)

The average rating scores on subjective cognitive load mostly fell in the higher half (10 $\sim$ 20) of the rating scale (see Figure 4.7, left). Subjects considered the text condition as the most difficult one and the 'text + spoken aid' condition as the easiest. The cognitive load ratings were significantly affected by the use of modality, $F(4, 16) = 17.1$, $p < .001$. Generally, two groups could be identified among the 5 modality conditions. The image and the 'text + spoken aid' conditions formed a group of higher ratings. The remaining three conditions formed a group of lower ratings. Results of post-hoc tests showed that there were significant differences in rating scores between any two conditions taken from different groups (6 condition pairs in total).

Subjective stress level (SS)

As shown in Figure 4.7 (right), the text condition was rated the most stressful and the 'text + spoken aid' condition was rated the least stressful. ANOVA results show a significant modality effect on subjective stress level (F (4, 16) = 9.4, $p < .001$). According to post-hoc tests, the stress level was significantly higher in the text condition than in the image, 'text + spoken aid' and 'text + sound aid' conditions. The 'text + image aid' condition was also rated significantly more stressful than the 'text + spoken aid' condition.

A very similar pattern can be seen when comparing the two graphs in Figure 4.7. Indeed, there is a strong positive correlation between ratings on cognitive load and stress (Cor. = 0.855), suggesting that subjects felt more stressed when they devoted more cognitive efforts to the task. Moreover, the subjective measurements were also found to be positively correlated with the performance measurements RT and ND. There are positive correlations at the 0.01 confidence level between ND-SCL, ND-SS, RT-SCL and RT-SS. In combination, these correlations indicate that when the task was more difficult (due to a suboptimal use of modalities), subjects devoted more cognitive effort, felt more stressed and performed worse.

4.3.3 Physiological States

In order to eliminate the individual differences in physiological activities, the values of physiological measures were normalized within each subject, using the baseline values obtained after relaxation. For example, the heart period (HP) value of subject 1 calculated from task trial n (HP_{n_s1}) was normalized with Equation 4.1.

$$HP_{n\_s1\_norm} = \frac{HP_{n\_s1}}{HP_{baseline\_s1}}, n = 1, 2, ..., 5 \tag{4.1}$$

ANOVAs were applied to the 7 measures, and only LF showed a modality effect at the 90% confidence level ($F(4,76) = 2.3$, $p = .06$). Since a modality effect on cognitive load has been consistently shown by performance and subjective measures, the insignificant physiological results seem to suggest that the variance in the task load (due to different modality conditions) did not significantly influence the physiological states of the participants. Further analyses were conducted to obtain deeper understanding of the physiological data.

Comparison of task conditions and baseline

We compared the physiological states in the 5 task trials to the baseline states (recorded during relaxation). For each measure, the average values of the 5 task trials were calculated for each participant and compared to the baseline value of the same participant. In HP, LF, GSL and RP, we found that the differences were in the expected directions for all participants. For example, HP was supposed to be shorter when performing a high-load task than when relaxed, and the average HP values of the 5 task trials were indeed lower than the baseline values, for all participants without exception. Regarding RMSSD, NN50 and GSRN, differences in expected directions were found in 14, 15 and 14 subjects, respectively. Furthermore, t-tests were applied to all measures. Differences between task conditions and baseline were found to be significant in HP, LF, RMSSD, NN50, GSL and RP (see Table 4.3 for details).

These results showed that most of the physiological measures did pick up the major changes in cognitive load between the baseline condition and the task conditions. However, they were not sensitive enough to reflect the relatively small variances in the cognitive load between the 5 task conditions. A similar conclusion regarding LF has been made in [130] (p. 311), where the author mentioned that: “the relationship between LF power and task demands is generally found for relatively large differences in task difficulty”. Our results seem to suggest that physiological measures are generally insensitive to minor variances in cognitive load.

Table 4.3: Results of t-tests comparing physiological states between task conditions and baselines. For the units of the measures, see Table 4.2.

Measure	Mean Task Conditions	Mean Baseline	t	Degree of Freedom	p
HP	820.8	871.8	12.4	19	$< .000$
LF	206.8	425.1	5.4	19	$< .000$
RMSSD	54.9	70.6	2.5	19	$< .05$
NN50	9.2	12.6	2.9	19	$< .01$
GSRN	4.0	3.0	-1.5	19	$n.s.$
GSL	7.2	5.9	-3.7	19	$< .01$
RP	3.2	4.5	5.1	19	$< .000$

Analysis of individual differences

The sensitivity of a physiological measure might differ from person to person – it might be more sensitive for some participants and less sensitive for others. If this is the case, statistical analyses using the data from all participants would not reveal any consistent pattern, which would explain the insignificant results. To prove this assumed individual difference, we tried to find out which measure(s) was(were) more sensitive for each participant.

After selecting a participant, we first defined a number of good and bad performance periods based on his/her performance log. Bad performance periods were selected by placing a 10-second time window centered at all time stamps when a patient died (in all task trials). The same number of good performance periods (also 10 s long) was taken from the beginning of all trials when the task was relatively easy. We assumed that the cognitive load level was higher in the bad performance periods than in the good performance periods, and that this difference should be relatively minor compared to the difference between task conditions and baseline. Then, six physiological measures[3] were re-calculated in each period. Finally, t-tests between the good and the bad performance periods were conducted for each measure respectively. A significant t-test result would indicate that the measure was sensitive enough to reflect the relatively minor changes in cognitive load.

This process was repeated for 5 randomly selected participants, and the results indeed revealed an individual difference in the sensitivity of physiological measures. As shown in Table 4.4), the heart rate measure (HP) was sensitive for participants 2, 4 and 5. Heart rate variability measures (RMSSD and NN50) were sensitive only for participant 3. Skin conductivity (GSL) was sensitive for participants 2 and 4. Respiration (RP) was sensitive only for participant 5. Unfortunately, none of the measures were sensitive for participant 1. This finding suggests that the selection of physiological measures for cognitive load assessment should take the individual difference of sensitivity into account, especially when the variances to be detected are relatively minor.

[3] LF is not applicable with a 10 s window. Normally, about 300 data points (about 5 minutes) are required to resolve frequencies down to the 0.01 Hz level [148].

Table 4.4: Individual differences in the sensitivity of physiological measures.

Participant index	No. of good performance periods	No. of bad performance periods	Measures with sig. t-test result at 95% cl.	Measures with sig. t-test result at 90% cl.
1	55	55	none	none
2	27	27	GSL	HP
3	37	37	RMSSD	NN50
4	39	39	HP	GSL
5	43	43	RP	HP

4.4 Discussion

The experimental results clearly showed that the use of modality affected the performance of the task, as well as the experienced cognitive load and stress. In this section, the experimental results are discussed in association with the hypotheses and the modality-related cognitive theories.

4.4.1 Text vs. Image

Comparing the two conditions without additional aid, all performance and subjective measures suggested that image had advantages over text in presenting victims. Thus the first hypothesis has been clearly confirmed. Image, as a nonverbal and analogue modality, is better for presenting concrete concepts [21; 98], such as the wounded and dead victims in this experiment. In contrast, text, as a verbal modality, is known to be less suitable for presenting concrete information, but more suitable for abstract concepts, logic, quantitative values, relations [21; 98]. Although the two images used in this experiment were designed to have similar shapes and colors, they still made it easier to distinguish between the two types of objects in comparison with words. We believe that the advantage of image over text would become even more notable if the two images showed larger contrasts in color, shape, and size. These results also stand in line with the dual-coding theory, because they show that verbal and nonverbal presentations of the same information indeed have different influences on how well the information can be processed.

4.4.2 Visual Aid vs. Auditory Aid

Here, we compare the 'text + image aid' condition to the 'text + spoken aid' and the 'text + sound aid' conditions. The results from all performance and subjective measures consistently showed that the spoken aid was significantly more appropriate than the image aid. In terms of average values, the sound aid was also superior to the image aid in all five measures. However, this advantage only reached a statistical significance in subjective cognitive load. Overall, we could conclude that the auditory aids were more beneficial than the visual aids in this experiment. The second hypothesis is confirmed.

The explanation of this finding is twofold. First, auditory signals are more able to attract attention than visual signals, especially when the eyes are occupied with another task (see Section 3.3.2). Therefore, while participants were busy searching for pending patients, visual aids displayed at the bottom of the display were more likely to be missed than spoken aids. An aid not only indicated the search area for a new patient but also announced the arrival of this patient. Therefore, in the 'text + image aid' condition, participants were likely to lose track of the number of pending patients. The worst consequence of missing an image aid could be that the participant continued with newer tasks and stayed unaware of a dying patient.

Second, even when being attended to, visual aids still have drawbacks due to cognitive resource competition. According to the working memory theory and the multiple resource theory, separated attention resources are used for visual and auditory perception. Therefore, auditory aids could be perceived in parallel with the ongoing rescue task. In contrast, the perception of visual aids cannot be time-shared with the rescue task. Limited visual perceptional resources needed to be divided between the rescue map and the aids. When the rescue task itself was demanding, visual aids were more likely to cause overload than to be of help. Not surprisingly, many subjects mentioned during the final interview that they sometimes had to consciously ignore the image aids in order to concentrate on the rescue task.

4.4.3 Verbal Aid vs. Nonverbal Aid

Although the image aid is nonverbal, it has been identified as inappropriate for this task in the previous sections. Therefore, this comparison is focused on the spoken aid and the sound aid. In terms of average values, all measures showed an advantage of the spoken aid over the sound aid. The difference in reaction time was significant. When asked to compare these two conditions, the majority of subjects preferred the spoken aid. In the third hypothesis, we expected nonverbal aids to be better suited than verbal aids, because directional information is nonverbal in nature. However, the results did not confirm this hypothesis.

The understanding of directions from the words 'left' and 'right' is highly automatic for most people. So the additional load associated with it (if any) was probably too little to harm the task performance. But, the direction of sound should be (if not better) at least an equally efficient way to convey directions. Then, why were spoken aids found to be better than sound aids? Subjects provided two explanations. First, it was commonly mentioned that spoken aids made it easier to maintain a short queue of newly-reported patients ('left's and 'right's) in mind, while trying to locate the current one. It was however harder to do the same with the sound aids. Baddeley's working memory theory states that the working memory usually relies on sub-vocal speech to maintain a memory trace (Section 3.4.1). That is to say the spoken aids 'left' and 'right' could be directly rehearsed, but the direction of a sound, as nonverbal information, had to be converted into a verbal form in order to be maintained. This conversion (via referential connections) consumed additional cognitive resources, and was probably the reason why subjects found it harder to maintain a queue of untreated patients with sound aids than with spoken aids. Second, a few subjects disliked the ambulance sound. They found it disturbing, especially when delivered at a high frequency like in this experiment. The sound made them less able to concentrate on the rescue task.

Interestingly, our findings seem to contradict the multimedia learning studies which suggested a benefit from verbal-nonverbal modality combination. This is due to the difference between our task and learning. A learning task requires comprehension and long-term memorization of presented knowledge. The combined use of verbal and nonverbal presentation invokes referential connections which have been shown to be essential to a deeper understanding and a better memorization [50]. In contrast, our task required quick perception, short-term memorization and quick reaction. In this case, to spend additional cognitive resources on building referential connections was less useful and more harmful.

4.4.4 Additional Aid vs. No Aid

All performance and subjective measures showed that the text condition was the worst of the five. However, when text was combined with spoken aids, the condition became the best of the five. This comparison seems to suggest that providing additional aids is beneficial compared to not providing them. However, the benefit of additional aid was only conditional, depending on the appropriateness of the modality combination.

When comparing the image condition with the 'text + image aid' condition, one can see that the former led to significantly shorter reaction times (RT), better rescue performance (ND) and lower cognitive load (SCL) than the latter. Considering average values, time of the first patient death (TFD) and subjective stress (SS) also showed an advantage of the image condition over the 'text + image aid' condition. However, the differences did not reach statistical significance. This comparison shows that presenting less information using an appropriate single modality (image) could be more beneficial than presenting more information using an inappropriate modality combination (text + image aid). Therefore, the additional aids can be of real help only when they are presented via an appropriate modality. The fourth hypothesis is only partially confirmed.

4.4.5 Low Load vs. High Load

We further investigated whether the modality effects reported above occurred without the high-load condition. At the beginning of each trial, no objects were on the grid map, thus the rescue task was relatively easy. As more and more objects were presented, it became more and more difficult to identify a patient in the crowded surroundings. According to the data from the TFD measurement, in all trials of all subjects, there were no patient deaths in the first 60 s of the trial. Therefore, we considered the first 60 s as a relatively low-load period. The average reaction time was recalculated with this period (see Figure 4.8). Comparing Figure 4.8 to Figure 4.5 (left), a similar up-and-down trend can be recognized, suggesting that the relative difference in task difficulty between conditions remain unchanged. However, the differences in reaction time between conditions were much smaller during the first 60 s. On average, reactions in the fastest condition ('text + spoken aid') were about 0.15

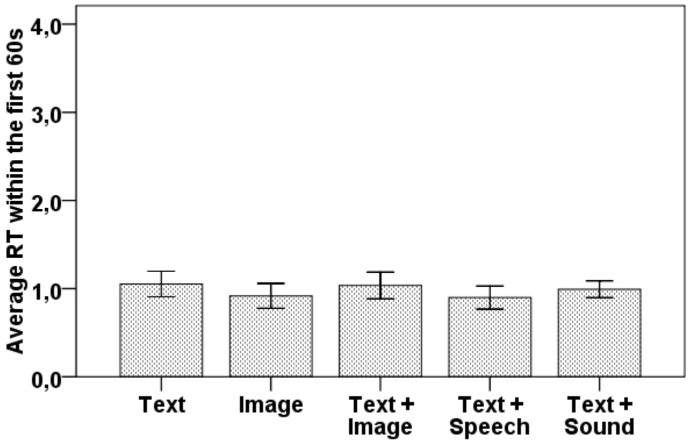


Figure 4.8: Average reaction time from the first 60 s in five modality conditions. Error bars represent standard errors.

s faster than in the slowest condition (text) – a difference of only about 14% of the value calculated from the whole trial (1.09 s, Figure 4.5, left). Furthermore, ANOVA analysis did not reveal any modality effect on the reaction time during the first 60 s ($F(4,16) = 1.6$, *n.s.*). These results suggest that in the low-load period, the use of modality influenced the task performance to a lesser extent, compared to a high-load condition in which this influence became significant. Therefore, it is particularly important for IMMP systems with high-load applications to consider the cognitive impact of modality allocation.

4.5 A Modality Suitability Prediction Model

The discussion of the experimental results showed that cognitive theories of working memory and attention, together with the representational feature of modalities, accounted for the variances in task performance and experienced cognitive load and stress. In this section, we demonstrate a possible way of integrating these theories and findings into a computational model that can systematically predict the suitability of a certain modality choice for this particular presentation task. We use our presentation task to demonstrate the construction of the model, and then apply it to several modality allocation strategies that were not investigated in this experiment. Several suggestions on adapting this model to other applications are also given.

4.5.1 Model Construction

The function of this model is to take a modality combination as input and output a numerical value describing the level of suitability of the modality choice. The general form of the model is a *weighted additive utility function* as shown in Equation 4.2, where w and A refer to weight and attribute respectively. Attributes are the modality properties that are identified to be relevant to the suitability evaluation. For each attribute, suitability values need to be assigned to all modality options. The higher the value, the more suitable the modality option. Weights determine how much each attribute contributes to the overall suitability score. The summary of all of the weights should be 1.

In summary, four steps are required to tailor the model for a presentation task: 1) define modality options, 2) determine attributes, 3) determine attribute values for each modality alternative, and 4) determine weights. Then, the model can be applied to predict the suitability of each modality alternative for this presentation task. Next, we demonstrate these steps using the presentation task in this study.

$$Outcome = w_1 \times A_1 + w_2 \times A_2 + ... + w_n \times A_n \quad (4.2)$$

For our presentation task, there are two information components to present: basic information and additional aid. In this case, the input of the model can be either a modality combination (when aids are presented) or a single modality (when aids are not presented). First, we defined the same modality options as in the experiment, meaning that text and image were options for basic information, and text, image, speech and sound were options for additional aid. Then, based on the relevant theories and findings, three attributes and their suitability values were determined as follows.

1. **VB**: the verbal/nonverbal property of the modality that presents the basic information. Modality candidates were image and text. Image is more suitable than text to present concrete objects such as victim types (see Section 4.4.1), thus a suitability value of 1 was assigned to text and 2 to image.
2. **PA**: the perception property of the modality that presents the additional aid. Supported by cognitive theories on working memory and attention, our results showed that visual aids were more beneficial than auditory aids (see Section 4.4.2). Therefore, we assigned 1 to the visual modalities (text and image) and 2 to the auditory modalities (speech and sound).
3. **VA**: the verbal/nonverbal property of the modality that presents the additional aid. Supported by the working memory theory and the dual-coding theory, our results showed that verbal aids were more beneficial than nonverbal aids (see Section 4.4.3). Therefore, 1 was assigned to nonverbal modalities (image and sound) and 2 was assigned to verbal modalities (speech and text).

In the fourth step, a weight was assigned to each attribute. Note that the difference in suitability values between a “better modality” and a “worse modality” was 1 for all attributes. Therefore the weight of an attribute fully determined how much difference in the overall suitability score can be accounted for by this attribute. To determine the weights in a sensible way, we fell back on the experimental results to observe the impact of each attribute on the measures. The impact of VB was reflected by the comparison between the text and the image condition. The impact of PA could be observed by comparing the ‘text + sound aid’ and the ‘text + image aid’ condition. The impact of VA could be shown by the comparison between the ‘text + sound aid’ and the ‘text + spoken aid’ condition. After an overall observation over all measures (Figures 4.5 $\sim$ 4.7), we concluded that PA and VA had comparable impacts, and the impact of VB was roughly three times as great as PA and VA. Therefore, the weights of VB, PA and VA were set to 0.6, 0.2 and 0.2 respectively. Finally, the modality suitability prediction model is shown as Equation 4.3.

$$Suitability = 0.6 \times VB + 0.2 \times PA + 0.2 \times VA \tag{4.3}$$

4.5.2 Suitability Prediction

The model was applied to 10 possible modality choices and the suitability predictions are shown in Table 4.5. The outcomes for the 5 experimental conditions are plotted in Figure 4.9, and the differences between conditions are generally consistent with the experimental results (Figures 4.5 $\sim$ 4.7). The 'image + spoken aid' combination is predicted to be the best modality choice for this presentation task.

Table 4.5: Predicted suitability of 10 possible modality choices. The 5 experimental conditions are marked with *.

Index	Modality for basic info.	Modality for additional aid	VB (0.6)	PA (0.2)	VA (0.2)	Suitability score
1	image	speech	2	2	2	**2.00**
2	image	sound	2	2	1	1.8
3	image	text	2	1	2	1.8
4	image	image	2	1	1	1.6
5*	text	speech	1	2	2	1.4
6*	image	none	2	0	0	1.2
7*	text	sound	1	2	1	1.2
8	text	text	1	1	2	1.2
9*	text	image	1	1	1	1.0
10*	text	none	1	0	0	0.6

4.5.3 Thoughts on Generalization

To adapt this model to other applications, the following aspects need to be re-considered.

1. ***The input***: which modalities are available in a system and which combinations of them are feasible and/or technically possible?
2. ***The output***: how can we define suitability based on the presentation goal? For example, in our case, the goal is to achieve a high task performance, low cognitive load and low stress. Therefore, suitability is determined by task performance, subjective cognitive load and stress.
3. ***The attributes***: which modality properties or other factors might have an influence on the suitability assessment, and which theoretical and empirical findings can be used

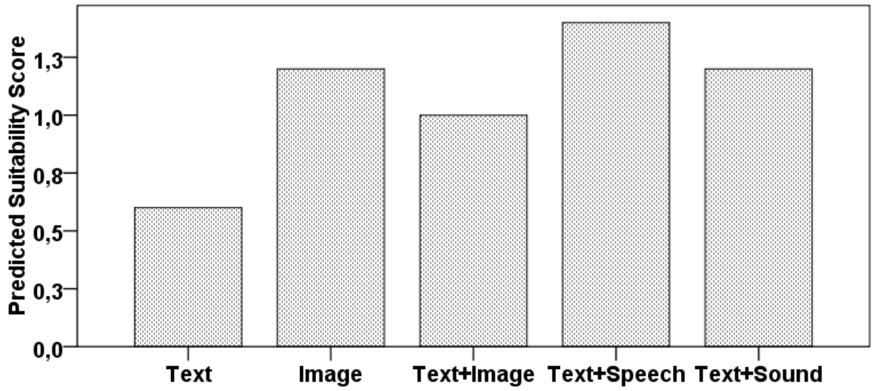


Figure 4.9: Suitability prediction scores of the 5 experimental conditions.

to predict the suitability of each modality candidate for each attribute? In the case that the system being developed has not been tested with users yet, designers can rely on the modality allocation guidelines and modality-related cognitive theories presented in Chapter 3 to select attributes and make suitability predictions.

4. ***The weights***: how large is the influence of each attribute on the overall suitability in relation to other attributes? In our case, the experimental results provided useful clues to weight determination. If experimental results have not yet been obtained, designers can fall back on other clues such as the importance of each type of information to the task/interaction. For example, we could say that the basic information is more important than the additional aid so that VB needs a larger weight value. In the case that only the most suitable modality choice is of interest, then it might be unnecessary to determine weights. In our case, 'image + spoken aid' would always receive the highest score, no matter which weights were chosen.

4.6 Conclusions

In this study, we evaluated modality alternatives for assisting a high-load visual search task embedded in a crisis rescue scenario. The contributions of this study and the limitations are summarized as follows.

First, we demonstrated that the use of modality had a significant effect on task performance and experienced cognitive load and stress. All performance and subjective measures showed consistent variances in the 5 investigated modality conditions. These variances could be well explained by modality-related cognitive theories and findings, suggesting that the knowledge about the cognitive impacts of modalities needs to be taken into account when making modality choices.

Second, our results have shown that the modality effects on task performance and cognitive load are particularly notable in a high-load condition. When the task load was relatively low, task performance did not significantly differ between conditions, because participants had enough spared cognitive resources to compensate the extra load induced by a suboptimal use of modality. However, when the task consumed the full capacity of cognition, various modality conditions led to significantly different task performance and (experienced) cognitive load. Therefore, it is particularly important for high-load information presentations to ensure the cognitive compatibility of modality choices.

Third, we proposed a suitability prediction model that quantitatively evaluates modality allocation options and systematically selects the best option for a specific presentation task. The model was demonstrated using our presentation task, and was applied to 10 possible modality allocation strategies (5 experimented and 5 additional). The validity of the model was supported by the consistency between its predictions and the experimental results. The most suitable modality combination for this presentation task was predicted to be 'image + spoken aid'. This model demonstrates a way to integrate modality-related cognitive theories and findings (such as the ones presented in Chapter 3) into a systematic modality selection process. Furthermore, we have provided several suggestions on how to adapt the model to other presentation tasks.

The limitation of this study lies in the simplicity of the task. Visual search is a low-level cognitive task that primarily involves perception. The information presented to the participants was also rather simple. Therefore, the next steps would be to investigate modality effects in complex high-load tasks, such as time-limited decision making, time-limited comprehension and motor response generation. We would also like to observe the interaction between modality and other presentation factors in those high-load tasks.

5 Information Presentation for Time Limited Decision Making

In this chapter, we present a study investigating the effect of information presentation on time limited decision making. Unlike the study in Chapter 4, the decision making task imposes load at both the perception and the cognition stage of information processing. Presentation conditions are manipulated by two factors: modality and spatial structure (how information items are spatially organized). The decision task has clearly defined rules and correct outcomes, which constrains the "shortcut" strategies people may develop to cope with the time pressure. The results of this study provide understanding of the effects of presentation and time limit on decision making performance, including which information processing stage each presentation factor influences and how time limit interacts with the presentation factors. This study and its results have been published in [34] and [40].

This chapter is structured as follows. Section 5.1 describes previous findings about the effects of presentation and time limit on multi-attribute choice making. Section 5.2 presents the design of the experiment in detail. Experimental results are presented in Section 5.3 and discussed in Section 5.4. Finally, conclusions are given in Section 5.5.

5.1 Related Work

The relation between information presentation and decision making has been investigated in many fields, such as user interface design, economics, marketing, risk communication and politics. A common finding is that decision makers tend to adapt their manner of information acquisition and their decision making strategy to the presentation of the task. The adaptation is believed to be guided by a cost-benefit analysis, compromising between the desire to minimize cognitive effort (cost) and the desire to maximize the quality of the decision (benefit) [126; 194]. As a result, different presentations of the *same* information could lead to different decisions. Therefore, task-related information should be presented with a "cognitive fit" [268], that is to support the optimal manner of information acquisition and decision making strategy for the task. In this section, we first present previous findings on the effects of presentation on *multi-attribute choice* tasks – a type of decision making task that is frequently encountered in daily life and often investigated in scientific studies. Then, we discuss how decision making behavior can be influenced by time pressure.

5.1.1 Presentation and Multi-attribute Choice Task

The multi-attribute choice task is to choose one option from several alternatives, and each alternative can be evaluated by several attributes. For example, when making a purchase choice between several cars (alternatives), a person often compares the price, weight, engine volume, fuel efficiency etcetera (attributes). Previous studies have shown that the decision making process can be influenced by the way in which alternatives and attributes are presented. In the study reported in [218], participants were asked to select the best loan application from 8 alternatives. Each alternative had 4 attributes. Modality and spatial structure were altered in different versions of the presentation. The results showed that the decision making process was more effortful when attribute values were presented by words than by numbers, in terms of decision time, the total number of operations and the percentage of information used. In addition, when numbers were used, subjects conducted more arithmetic activities and less information search activities. Regarding spatial structure, lists required more effort to process than matrices, revealed by both performance and subjective measures. Spatial structure also influenced the way in which information was acquired.

Stone et al. [243; 244] investigated the effects of modality on risk-taking choice. Participants had to make a purchase choice between a cheaper but less safe tire and a safer but more expensive tire. Accident ratios (attribute) were presented in either numbers, graphs or images. Results showed that graphical presentations made people more willing to pay for safety compared to numbers. The authors proposed that the reason was that graphs and images enhanced the perception of risk by highlighting the number of people harmed.

Several other studies investigated the effect of spatial structure on multi-attribute choice making [22; 126; 249]. Using a table or a list, information was sorted either by alternative or by attribute. The common finding was that when information was organized by alternatives, subjects tended to process an alternative before considering the next alternatives; when information was organized by attributes, subjects tended to compare all alternatives on a single attribute before considering the next attribute. The processing by alternative was found to be less cognitively demanding, leading to more accurate and time efficient decisions [249].

Although multi-attribute choice does not represent all decisions in real-life, the above-mentioned findings have clearly demonstrated that information presentation can have a strong influence on decision making performance. However, we noticed two limitations in these studies. First, participants normally had an unlimited amount of time to carefully process all available information before making choices. Second, there was very often no clear definition of correct or wrong choices and no limitation in which strategy to take. If strategies are defined, they can obviously constrain the way how decisions are made. Time pressure can also influence how decision makers acquire information (see below). When presentation format is no longer the only factor in play, its effect may be weakened or overruled.

5.1.2 Decision Making Under Time Pressure

Decision making is very often time-limited in real-life situations. A typical example is the crisis situation, where emergency decision makers are required to process massive amounts of information in a limited amount of time. A key feature is 'demand exceeding cognitive resource' [128]. Under time stress, decision makers tend to focus on the general outline of the problem instead of on in-depth analysis [62]. Studies using the multi-attribute choice task found that subjects changed strategies from a more alternative-based (depth-first) to a more attribute-based (breadth first) pattern of processing as time pressure increased [194]. In order to cope with cognitive overload, they are also prone to selectively use subsets of the information, to adopt simpler modes of information processing and to base their decisions on certain important 'cues' [122; 139; 144]. When time limitation makes it impossible to apply the normative strategy (careful and reasoned examination of all alternatives and attributes), decision makers adopt heuristic strategies (simple and fast rules of thumb) [195]. If the heuristic strategies are consistent with the task, they can be more efficient and accurate under time pressure. If they are inconsistent with the task, however, they can also harm the decision making performance.

In this study, we set up a time-limited multi-attribute choice task with a predefined normative strategy[1] and with only one correct outcome. Both time limit and strategy limit would constrain the way decision makers acquire and process information. We aimed to investigate 1) the presentation effects on decision making performance (defined by time efficiency and accuracy), 2) the interaction between different presentation factors (modality and spatial structure), and 3) the interaction between presentation effect and time limit.

5.2 Method

5.2.1 Scenario

Similar to the experiment presented in the previous chapter, the time-limited decision making task was also embedded in a medical rescue scenario.

After an earthquake, wounded patients are sent to an emergency medical center. Unfortunately, the number of patients exceeds the capacity of medical resources (equipment and staff). Therefore, the order of treatment needs to be wisely determined based on individual injury conditions. Participants play the role of a medical staff member who needs to make quick decisions on which patient to treat first.

[1] Normative strategies apply a careful and reasoned examination of all alternatives and attributes. Heuristic strategies are simple and fast rules of thumb [194].

Injury	Patient 1	Patient 2
Heart failure	mild	severe
Blood loss	severe	mild
Respiration obstruction	none	moderate
Brain damage	none	none
Fracture	severe	none

Figure 5.1: A patient pair presented in the text modality and the by-injury structure.

Severity	Patient 1	Patient 2
Severe	Blood loss Fracture	Heart failure
Moderate		Respiration obstruction
Mild	Heart failure	Blood loss
None	Respiration obstruction Brain damage	Brain damage Fracture

Figure 5.2: A patient pair presented in the text modality and the by-severity structure.

Note that it was not our intention to set up a realistic medical rescue task, nor did we expect participants to have knowledge of medical treatment. The aim of this scenario was to motivate the task setting and the time limit.

5.2.2 Presentation Material

One pair of patients was presented at a time. The injury condition of a patient was described by five injury items (derived from [199]): heart failure, respiration obstruction, blood loss, brain damage and fracture. The former three items were described as more time-critical to saving life than the latter two. The severity of each injury category was defined at four levels (derived from [222]): severe, moderate, mild and none.

The two presentation factors were *modality* (text or image) and *structure* (by-injury or by-severity). A full factorial design led to 4 different presentation formats (see Figure 5.1 – 5.4). In the two text conditions, the injury categories and severity levels were presented with English text. In the image conditions, injury categories were presented by iconic images of the affected organs (e.g. an iconic image of a brain referred to the ‘brain damage’ item), and severity levels were presented by color rectangles[2]. The color-severity mappings were

[2]Strictly speaking, colors are not images. But colored rectangles can be considered as images. In this study, we use “image” to generally refer to nonverbal visual presentations.

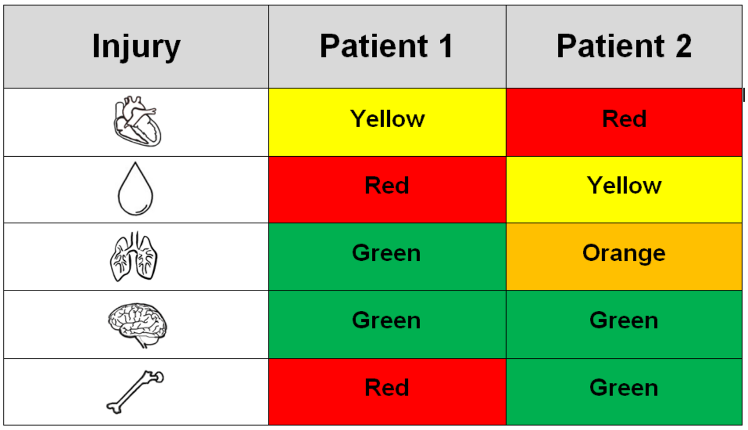


Figure 5.3: A patient pair presented in the image modality and the by-injury structure. The text of colors was added to ensure the readability in a grayscale printing, and was not present in the experiment.

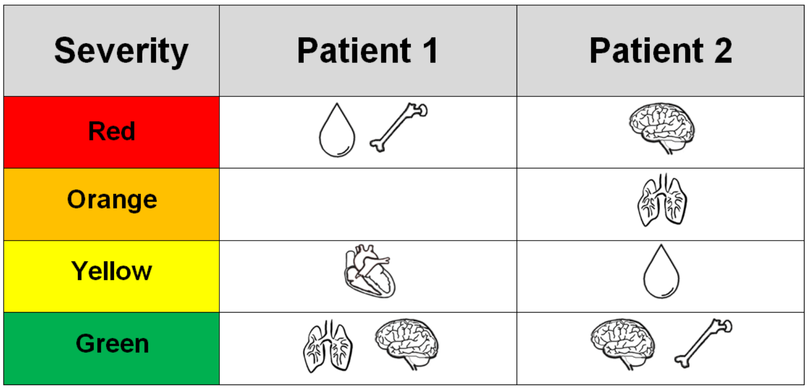


Figure 5.4: A patient pair presented in the image modality and the by-severity structure. The text of colors was added to ensure the readability in a grayscale printing, and was not present in the experiment.

red for 'severe', orange for 'moderate', yellow for 'mild', and green for 'none'. This color scheme has been commonly used to express severity in existing applications. For example, the North American avalanche danger scale uses green, yellow, orange and red to visualize four increasing levels of danger [52]. The GPS navigation devices from GARMIN use green, yellow, and red icons to indicate the severity of traffic delay ([86], p. 30).

The injury information of two patients was organized in a table. The table could be structured by the injury categories or by the severity levels. When using the by-injury structure, the more important three injury categories were located on top of the less important two. The injury column was fixed for all tasks and the severity values varied. When using the by-severity structure, the four severity levels were ranked from high to low and located from top to bottom of the table. The severity column was fixed for all tasks and the locations of injury categories varied.

5.2.3 Task and Strategies

The nature of the task was a multi-attribute choice with 2 alternatives (patients), 5 attributes (injury categories) and 4 attribute values (severity levels). The 5 attributes had different weights. A normative strategy was defined and presented to the participants. However, when there was a time limit, it was not always feasible to apply the normative strategy for a careful analysis. Participants would then need to develop heuristics strategies to reach the correct choices faster.

Normative strategy

The normative strategy evaluates the overall injury level of a patient by a weighted additive function. The attributes of the function are the five injury items. They are categorized into two priority groups with different weights. Heart failure, blood loss and respiration obstruction are defined as *twice* as time critical as brain damage and fracture, and thus have a weight of 2 while the other two items have a weight of 1. Numerical values are also assigned to the four severity levels. *Severe* is defined to be three times as serious as *mild*, and *moderate* twice as serious as *mild*. Therefore, 3, 2, 1 and 0 represent *severe*, *moderate*, *mild* and *none*, respectively. Finally, the overall injury level of a patient was quantified by Equation 5.1, where 'SL' refers to 'Severity Level'. When making a final choice, the patient with a higher injury value should be treated first. Taking the patient pair shown in Figure 5.1 for an example, the injury value is 11 ($2 \times 1 + 2 \times 3 + 3$) for patient 1, and 12 ($2 \times 3 + 2 \times 1$) for patient 2. Therefore, the correct decision is to treat patient 2 first.

$$Injury_{norm} = 2 \times SL_{heart} + 2 \times SL_{blood} + 2 \times SL_{respiration} + SL_{brain} + SL_{fracture} \quad (5.1)$$

To evaluate the processing load of this strategy, we can apply the componential analysis of cognitive effort in choice [23]. This analysis first breaks down a decision making strategy into sets of components called elementary information processing operations (EIPs). Then, the total number of EIPs represents the processing load of this strategy. A particular set of EIPs defined for multi-attribute choice tasks is described in Table 5.1. Applying Equation 5.1 to the example patient pair, it requires 10 READ EIPs (acquire the 10 severity values), 4 ELIMINATE EIPs (ignore the 0s), 4 PRODUCT EIPs (apply weights), 3 ADD EIPs (sum up the values), 1 COMPARE EIP (compare the two final values) and 1 CHOOSE EIP (choose patient 2). The total cognitive effort in this choice is 23 EIPs.

Time limit

Based on a pilot study, the time limit was set to 20 s for each task. A countdown timer was displayed at the bottom of the screen to show participants how much time was left. At the 15th second, a speech reminder ("5 seconds!") was given to inform participants that it was time to finalize their choices. If a decision had not been made when the time was up, the information on the screen would be removed and a speech warning ("Please make a choice!") was delivered. Then, participants were forced to make a choice based

Table 5.1: A set of EIPs matched to multi-attribute choice strategy (reproduced from [23], p.114).

EIP	Description
READ	Read an alternatives value on an attribute into short-term memory
COMPARE	Compare two alternatives on an attribute
DIFFERENCE	Calculate the size of the difference of two alternatives for an attribute
ADD	Add the values of an attribute in short-term memory
PRODUCT	Weight one value by another (multiply)
ELIMINATE	Remove an alternative or attribute from consideration
MOVE	Go to next element of external environment
CHOOSE	Announce preferred alternative and stop process

on the incomplete analysis. This time limit required participants to be fully engaged in the task. With the amount of training provided in the experiment, 20 s were often insufficient for applying the normative strategy for an exact calculation. Therefore, participants had to adopt heuristic strategies to cope with the time pressure, which was the main challenge of this task.

Heuristic strategies

Heuristic strategies are by definition simpler than the normative strategy in terms of calculation complexity, but they are not necessarily less accurate. Unbiased heuristic strategies always generate outcomes that are equivalent to the normative strategy, thus are efficient "shortcuts". Biased heuristic strategies on the other hand, trade accuracy for less processing load, thus cannot guarantee correct outcomes. We created an unbiased heuristic strategy for this task and introduced it to the participants for inspiration. However, participants were given freedom to apply any strategy they preferred as long as the choice could be made quickly and correctly. It was one of our aims to see how participants would process the information and make their choices.

An unbiased heuristic strategy. This strategy is based on the idea that it is not necessary to know the two absolute injury values, because the decision is only dependent on the difference between them. Therefore, compensatory eliminations can be conducted before numerical calculation to reduce processing load. The method is to identify two injury items that 1) are from different patients; 2) have the same severity value; and 3) have the same weight. Two such injury items have the same contribution to the comparison, thus can be eliminated from the calculation. When all possible eliminations are finished, the remaining injury items can be calculated for a final comparison. Note that 'none' items have a value of 0 and can be ignored as well.

Figure 5.5 shows an example of applying this strategy with the by-injury structure. The severe blood loss of patient 1 can be eliminated with the severe heart failure of patient 2. The mild heart failure of patient 1 can be eliminated with the mild blood loss of patient 2. The four 'none's are eliminated. Finally, there are only two injury items to be calculated.

Injury	Patient 1	Patient 2
Heart Failure	~~mild~~	~~severe~~
Blood loss	~~severe~~	~~mild~~
Respiration obstruction	~~none~~	moderate
Brain damage	~~none~~	~~none~~
Fracture	severe	~~none~~
	3	2×2 = 4

(×2 applies to Heart Failure, Blood loss and Respiration obstruction)

Figure 5.5: The unbiased heuristic strategy applied to the by-injury structure.

Severity	Patient 1	Patient 2
Severe	~~Blood loss~~ Fracture	~~Heart failure~~
Moderate		Respiration obstruction
Mild	~~Heart failure~~	~~Blood loss~~
None	~~Respiration obstruction~~ ~~Brain damage~~	~~Brain damage~~ ~~Fracture~~
	3	2×2 = 4

Figure 5.6: The unbiased heuristic strategy applied to the by-severity structure.

The severe fracture of patient 1 has a value of 3 and the moderate respiration obstruction of patient 2 has a value of 4 (2×2). Therefore, patient 2 is the correct choice. The processing load of this strategy consists of 10 READ EIPs, 6 ELIMINATE EIPs, 1 PRODUCT EIP, 1 COMPARE EIP and 1 CHOOSE EIP. The total load (19 EIPs) is 82% of the normative strategy (23 EIPs). This strategy can also be applied with the by-severity structure. An example is shown in Figure 5.6. However, we predict that this strategy is better assisted by the by-injury structure, because 1) the two priority groups are more clearly separated, and 2) the eliminated two items are either the same word or the same color, which makes eliminations more intuitive.

Biased heuristic strategies. Biased strategies can be developed to further reduce the processing load. For example, one can ignore the two less important injury items and only consider the more important three items. One can also ignore the differences in weight and treat all five injury items equally. One can even base his/her decision merely on the severe injury items. The most extreme case would be to make a random choice. All these biased strategies have a chance (higher or lower) of generating correct choices but none of them can guarantee accuracy of the choices.

5.2.4 Experimental Design and Procedure

We used a 2×2×2 mixed design. The two within-subject factors were *modality* (text or image) and *structure* (by-injury or by-severity). Therefore, all participants performed four experimental trials, namely the 'text & by-injury', 'image & by-injury', 'text & by-severity' and 'image & by-severity' trial. The trial order was counterbalanced with a size 4 Latin Square. A set of 12 tasks were performed in each trial with a randomized order. Time limit was treated as a between-subject factor (with or without). 48 university students (master's students and PhD students) volunteered to participate in the experiment. All of them spoke fluent English and none of them had a medical background. 32 participants performed the task with a time limit and 16 performed without a time limit.

The experiment contained three sessions: an introduction session, a training session and the experimental session. The introduction presented the rescue scenario, the task and the four presentation conditions, the normative decision making strategy and the unbiased heuristic strategy. In the training session, participants practiced 20 tasks, 5 tasks for each presentation condition. No time limit was applied during training. Feedback on decision accuracy was given after each decision was made, by the system via speech. After training, participants performed four experimental trials of 48 tasks (4 × 12). The system provided a performance summary after each trial, announcing how many correct decisions had been made. After the four trials were all finished, participants were required to complete the questionnaires. The time duration of the experiment was about 40 minutes.

5.2.5 Dependent Measurements

The decision making performance was measured by two dependent variables, namely *time efficiency* and *accuracy*. The *time efficiency* of a decision refers to the time interval between the moment when a task is presented and when the decision is made (in seconds). A decision is correct if it is identical to the outcome from the normative strategy. *Accuracy* is the percentage of correct decisions made in a trial. These two measurements can both be retrieved from log files.

Subjective assessment of cognitive load was obtained by questionnaire (see Appendix B). Participants were asked to perform four binary comparisons on the cognitive demand of the task, as shown in Table 5.2. In each comparison, they selected the presentation condition which induced least cognitive load. In addition, participants were asked to indicate in which presentation conditions the task was the easiest and the most difficult to perform. Based on the questionnaire, we conducted 3 groups of analyses: 1) comparisons between the two modalities, 2) comparisons between the two structures, and 3) the easiest and the most difficult conditions.

The questionnaire also addressed decision making strategies. First, participants were given a rough choice regarding how they had made their decision. They could choose between 1) a complete calculation as the normative strategy, 2) eliminations and calculation as the unbiased heuristic strategy and 3) other self-developed strategies. To obtain more details, the experimenter also asked participants to orally describe how they processed the

Table 5.2: The four binary comparisons on the cognitive demand of the decision making task.

Cognitive load comparison	Reference
Text & By-Injury vs. Image & By-Injury	Figure 5.1 vs. Figure 5.3
Text & By-Severity vs. Image & By-Severity	Figure 5.2 vs. Figure 5.4
Text & By-Injury vs. Text & By-Severity	Figure 5.1 vs. Figure 5.2
Image & By-Injury vs. Image & By-Severity	Figure 5.3 vs. Figure 5.4

information and made their choices in each presentation condition, using the printed samples of presentation as a reference. The experimenter made notes during the description.

5.3 Results

According to the experimental design, we applied repeated-measure ANOVAs to analyze the performance measures with *modality* and *structure* as within-subject factors and *time limit* as a between-subject factor. Results on performance, subjective judgements and strategy analysis are presented in this section.

5.3.1 Decision Making Performance

Time efficiency

The average time spent on one task (in seconds) in each trial is shown in Figure 5.7. The time limit is indicated by a gray line in the figure. The presence of a time limit significantly influenced the time spent on each decision, $F(1, 46) = 33.5$, $p < .001$. On average, decisions were made 10.4 s faster with the time limit than without. Comparing the four presentation conditions, the average decision time was the shortest in the 'Image & By-injury' condition and the longest in the 'Text & By-severity' condition. Both modality and structure showed a significant effect on time efficiency (modality: $F(1, 46) = 33.3$, $p < .001$; structure: $F(1, 46) = 28.5$, $p < .001$). On average, time spent on a decision was 2.7 s shorter with image than with text, and 2.4 s shorter with by-injury structure than with by-severity structure. Post-hoc tests between the four presentation conditions (with Bonferroni correction) further revealed that the 'Image & By-injury' condition led to significantly faster decisions than the others, and the 'Text & By-severity' condition led to significantly slower decisions than the others.

Furthermore, there was a significant interaction effect between modality and time limit, $F(1, 46) = 6.5$, $p < .05$. This is because the time difference between the image and the text conditions was greater without than with the time limit. The interaction between structure and time limit was also significant at a 90% confidence level, $F(1, 46) = 3.7$, $p = .06$.

Accuracy

The average number of correct decisions made in each trial is shown in Figure 5.8. ANOVA showed a significant effect of time limit, $F(1, 46) = 41.4$, $p < .001$. On average, decisions

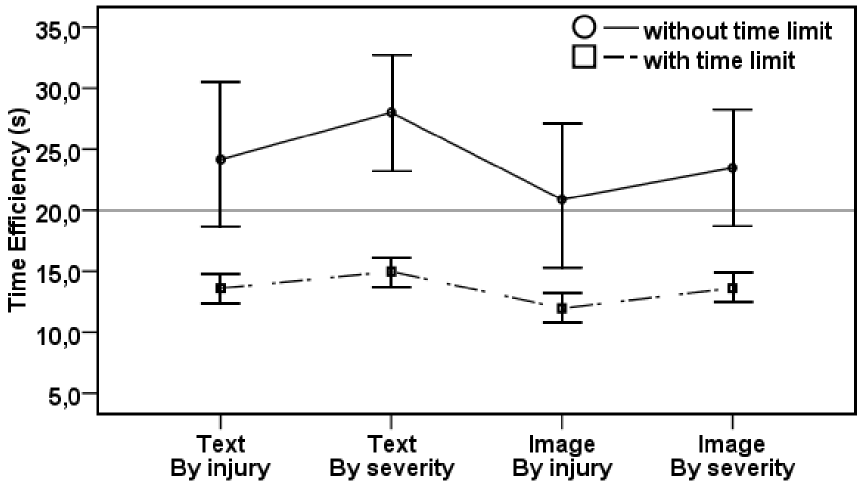


Figure 5.7: Time efficiency in all task conditions. Error bars represent standard errors. The gray horizontal line indicates the 20-second time limit.

made were 94.2% correct without a time limit and 78.6% correct with a time limit. Regarding presentation conditions, the 'Image & By-injury' condition had the highest average accuracy and the 'Text & By-severity' condition had the lowest. However, ANOVA showed neither a modality effect nor a structure effect (modality: $F(1, 46) = 1.8$, *n.s.*; structure: $F(1, 46) = 3.5$, *n.s.*).

To further investigate the influence of presentation on accuracy, we separately analyzed the two sets of data with and without a time limit. When the time limit was present, the structure factor had a significant effect on accuracy ($F(1, 31) = 7.2$, $p < .05$), showing that decisions were made more accurately with the by-injury structure than with the by-severity structure (3.3% more on average). The modality factor did not have an effect on accuracy ($F(1, 31) = 1.6$, *n.s.*). Post-hoc tests revealed a significant difference in accuracy between the 'Image & By-injury' condition and the 'Text & By-severity' condition. However, when a time limit was absent, neither modality nor structure showed a significant influence on accuracy (modality: $F(1, 15) = 2.1$, *n.s.*; structure: $F(1, 15) = 3.1$, *n.s.*).

5.3.2 Subjective Judgments

In general, the time limit did not influence the subjective preferences between different presentation alternatives. In all three groups of questionnaire analyses, a strong positive correlation was found in the numbers of choices with and without the time limit (correlation coefficients all greater than 0.98). Detailed results of these analyses are presented below.

Modality comparisons

Subjective comparisons of task load between the two modalities were categorized into three groups: 1) text was easier than image regardless of structure, 2) image was easier than text regardless of structure, and 3) text was easier with one structure and image was easier with the other structure. Figure 5.9 summarizes the results. A strong majority of participants

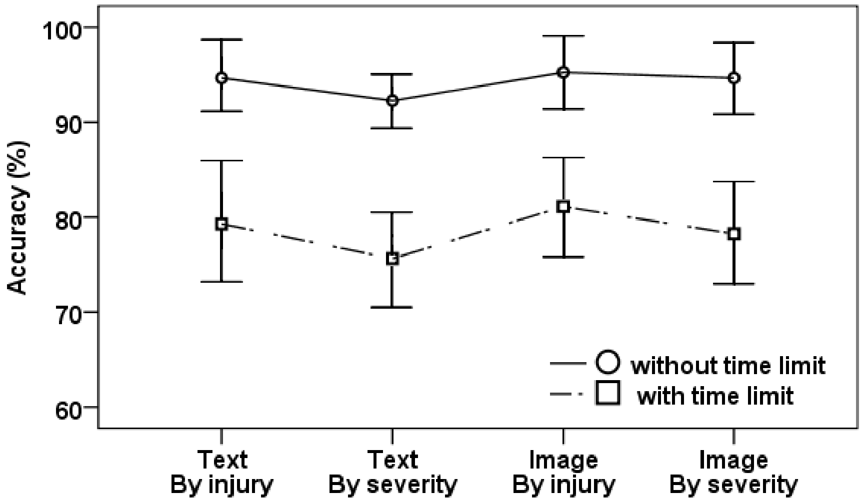


Figure 5.8: Decision accuracy in all task conditions. Error bars represent standard errors.

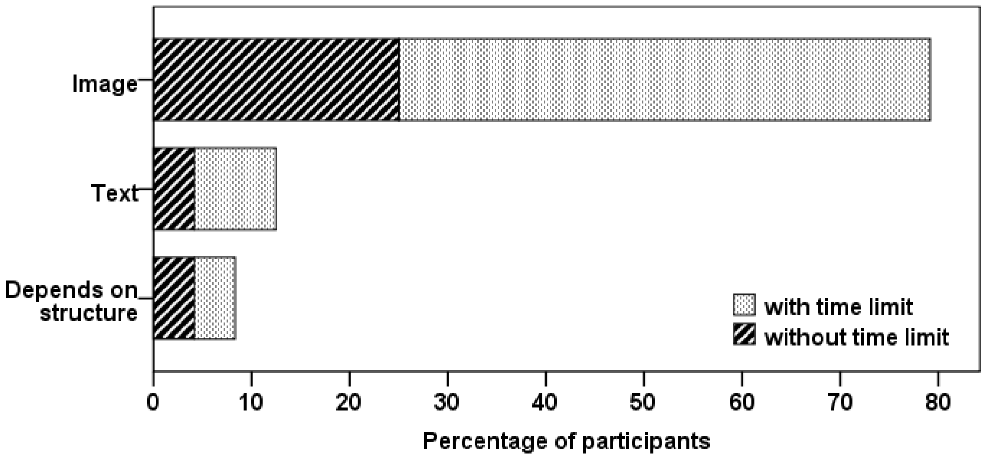


Figure 5.9: Subjective comparisons of task load between the two modalities. The modality that imposed less load was chosen.

(79.1%) found the task easier to perform with image presentations than with text regardless of which structure was provided. Unlike the majority, 6 participants (12.5%) always preferred text to image. Three of them explained that it was more effortful to interpret the injury items from images of organs than from words. Interestingly, these three participants all had a text or speech related research topic to work on daily, such as text retrieval. Another participant had problems with understanding the color representation of severity. He was the only participant who came from an Asian culture, and found it particularly difficult to associate red with severe injuries, because red normally symbolized love and passion in his culture. Another 4 subjects (6.2%) preferred image when the structure was by-injury and preferred text when the structure was by-severity. In other words, they preferred colors to words for presenting severity levels and preferred words to images for presenting injury items.

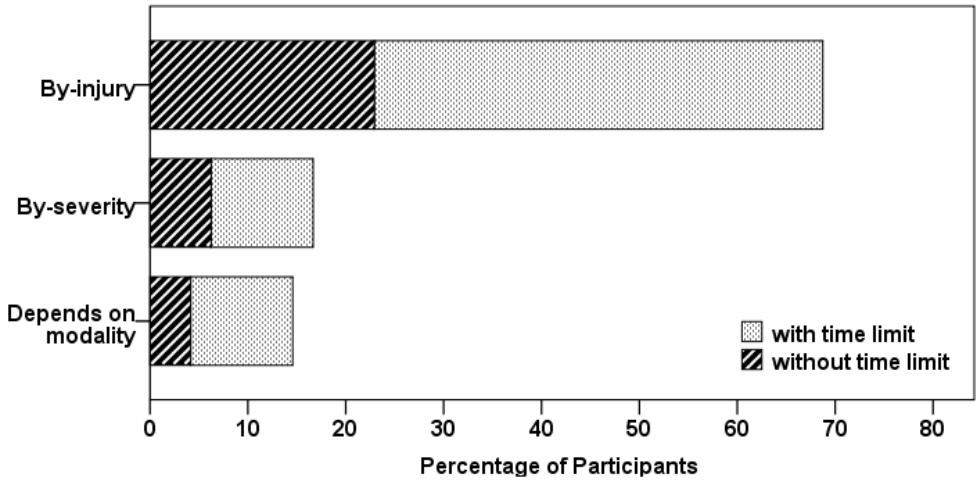


Figure 5.10: Subjective comparisons of task load between the two structures. The structure that imposed less load was chosen.

Structure comparisons

Subjective comparisons of task load between the two structures were also categorized into three groups and the results are summarized in Figure 5.10. A majority (68.8%) of participants always found the task easier to perform with the by-injury structure than with the by-severity structure. In contrast, 16.7% of participants indicated an exactly opposite preference. They explained that the by-severity structure had the advantage that it allowed them to eliminate 'none's easily, because they just ignored the last row of the table. Therefore there was less information to start with. Another 14.6% of participants preferred the by-injury structure with image and preferred the by-severity structure with text. Between the two image conditions, they found the one with by-injury structure easier because the color representation of severity was particularly intuitive for them. Between the two text conditions however, the by-severity structure was found easier to deal with because the general distribution of injury items in the table provided a clue of which patient to choose – the one whose injury items were more gathered at the top of the table.

The easiest and the most difficult conditions

Subjective judgements on the easiest and the most difficult conditions are shown in Figure 5.11. Again there was a clear majority option that was consistent with the performance measures. The 'Text & By-severity' condition was chosen as the most difficult condition by 58.3% of participants, the 'Image & By-injury' condition was chosen as the easiest condition by 62.5% of participants. Figure 5.11 also shows that the modality factor had a stronger influence on these judgements than the structure factor, because 85.4% of participants chose an image condition to be the easiest and 79.2% chose a text condition to be the most difficult. This result is interesting because structure actually had stronger influence on the performance than modality (Section 5.3.1).

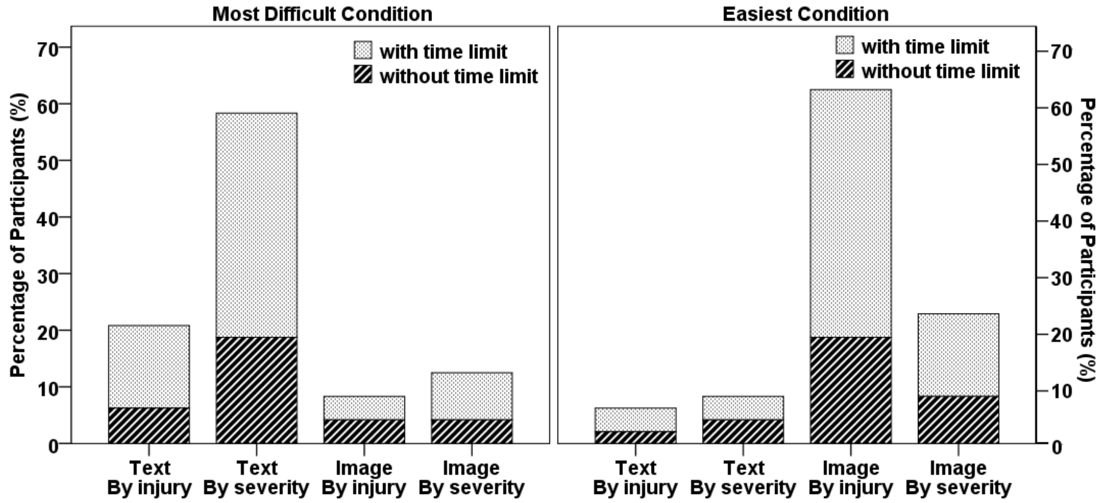


Figure 5.11: Subjective judgements of the most difficult (left) and the easiest (right) conditions.

Table 5.3: Percentage of participants who reported using each strategy or strategy combination. NS, UBS and SS represent normative strategy, unbiased heuristic strategy and self-developed strategy, respectively.

Time Limit	NS	NS + UBS	UBS	UBS + SS	SS
Present	0%	0%	15.6%	21.9%	62.5%
Absent	18.8%	18.8%	50%	0%	12.4%

5.3.3 Strategy Analysis

Subjective descriptions

Most of the participants were able to clearly describe how they processed the information to reach the final choices. As Table 5.3 shows, participants in the with-time-limit group never applied the normative strategy for a complete calculation. In contrast, they were very active in developing their own strategies to cope with the time pressure. On the other hand, most participants in the without-time-limit group stuck to the two strategies that were introduced to them, which explained their lower time efficiency and higher decision accuracy (see Figure 5.7 and 5.8).

Further analyses of the strategy descriptions only focused on the with-time-limit group. Influenced by the unbiased heuristic strategy which was introduced beforehand, participants generally tried to apply the elimination methods to reduce calculation load. Often, they first applied a self-developed biased strategy to quickly come up with a primary decision, then continued with more careful and complete information processing. If the careful processing could not be finished within the time limit, the primary decision would then become the final choice. Not surprisingly, the development of biased strategies was influenced by the structure factor rather than by the modality factor. Here we describe two commonly mentioned biased strategies, referred to as BS1 and BS2.

BS1. With the by-injury structure, since the separation of priority groups was directly available (first three rows in the table), participants first applied eliminations and calculations on the three critical injury items, and made primary decisions based on the outcome. Then, if there was still some time left, they also quickly processed the other two injury items. It was mentioned that this quick processing did not change the decision in most cases. This biased strategy, as shown by Equation 5.2, reduced cognitive load by processing only a subset of the information items. Since the three important injury items contribute more to the overall injury value, this strategy had a good chance to output correct choices but did not give any guarantee.

$$Injury_{BS1} = SL_{heart} + SL_{blood} + SL_{respiration} \tag{5.2}$$

BS2. With the by-severity structure, elimination was more effortful because the weights of the injury items were not directly readable from the table. Therefore, participants first looked at the distribution of the injury items and came up with a primary decision based on whose injuries were more gathered at top of the table (the more severe side). Then, they started to identify weights, eliminate items and calculate values, aiming to adjust the choice. If the process could not be completed within the time limit, the primary decision became the final choice. This biased strategy, as shown by Equation 5.3, reduced cognitive load by ignoring the weights and treating all injury items equally. It was more likely to be accurate when the three important items were not or less seriously injured.

$$Injury_{BS2} = SL_{heart} + SL_{blood} + SL_{respiration} + SL_{brain} + SL_{fracture} \tag{5.3}$$

Quantitative analysis

From participants' strategy descriptions, we observed that most participants who experienced the time limit made primary decisions by BS1 or BS2, and then tried to confirm or correct them by unbiased processing. Further, we took one more step to look for quantitative evidence of these two biased strategies being applied. By doing so, we also intended to confirm the influence of presentation structure on participants' strategy development. First, we applied these two strategies to all 12 tasks and found that each of them generated 8 correct decisions and 4 incorrect decisions. The 4 incorrect decisions from BS1 and BS2 did not overlap over the same task. Accordingly, we defined the following four task groups:

- G1: the 8 tasks on which BS1 is correct
- G2: the 4 tasks on which BS1 is incorrect
- G3: the 8 tasks on which BS2 is correct
- G4: the 4 tasks on which BS2 is incorrect

Second, we turned to the performance data and calculated the percentage of decisions that were made correctly in each task group, by each participant and in each presentation condition. Then, t-tests were applied between G1 and G2 in each presentation condition to investigate the impact of BS1 on accuracy. Similarly, t-tests were also applied between

Table 5.4: Results on decision accuracy in 4 different task groups and t-tests between groups.

Presentation	Mean Accuracy				T-test Sig.	
Condition	G1	G2	G3	G4	G1 – G2	G3 – G4
Text, By-Injury	91.4%	64.1%	80.8%	85.2%	$p < .001$	$p = n.s.$
Image, By-Injury	92.4%	63.5%	84.4%	83.6%	$p < .001$	$p = n.s.$
Text, By-Severity	80.4%	76.1%	87.9%	63.9%	$p = n.s.$	$p < .05$
Image, By-Severity	83.2%	77.3%	86.7%	69.3%	$p = n.s.$	$p < .05$

G3 and G4 to investigate the impact of BS2 on accuracy. The results are shown in Table 5.4. In the two conditions with the by-injury structure, the accuracy of decision making was significantly higher in G1 than in G2, but no significant difference was found between G3 and G4. The two conditions with the by-severity structure showed the exact opposite results. Decisions were significantly more accurate in G3 than in G4, but no effect was found between G1 and G2. These findings clearly support the subjective strategy descriptions. It seems reasonable to conclude that BS1 and BS2 were applied respectively to the by-injury conditions and the by-severity conditions, which in turn confirms that the structure of information presentation may influence the development of decision making strategies.

5.3.4 Task Difficulty

The results presented so far have shown that the two presentation factors and the time limit all influenced the cognitive demand of this task, because they determine the condition in which the task is performed. In addition, the cognitive demand of a task should also be influenced by its difficulty level. In the same task condition, easier tasks should require less cognitive effort than difficult ones. Variances in task condition might have different impacts on the performance of easy tasks and difficult tasks. The 12 tasks used in this study were not equally difficult, allowing us to investigate how task difficulty affected performance and how it interacted with the other factors.

The difficulty level of the task can be described by the difference between the two overall injury values of a patient pair (calculated by Equation 5.1). The smaller the difference, the more difficult it is to identify which patient should be treated first. For example, the task shown in Figure 5.1 is difficult because the difference between the two overall injury values is only 1. Among the 12 tasks, we identified 8 relatively difficult tasks with a difference smaller than 3, and 4 relatively easy tasks with the difference greater than 5. Time efficiency and accuracy were re-calculated from these two task groups separately, and the results are shown in Figure 5.12 and 5.13.

Accuracy. On average, the accuracy of easy tasks was 100% without the time limit and 97.5% with the time limit. The accuracy of difficult tasks was generally lower – 91.3% without the time limit and 73.6% with the time limit. ANOVA showed that this effect of task difficulty was significant ($F(1, 46) = 91.5$, $p < .001$). There was an interaction effect between time limit and task difficulty ($F(1, 46) = 19.8$, $p < .001$), because the presence

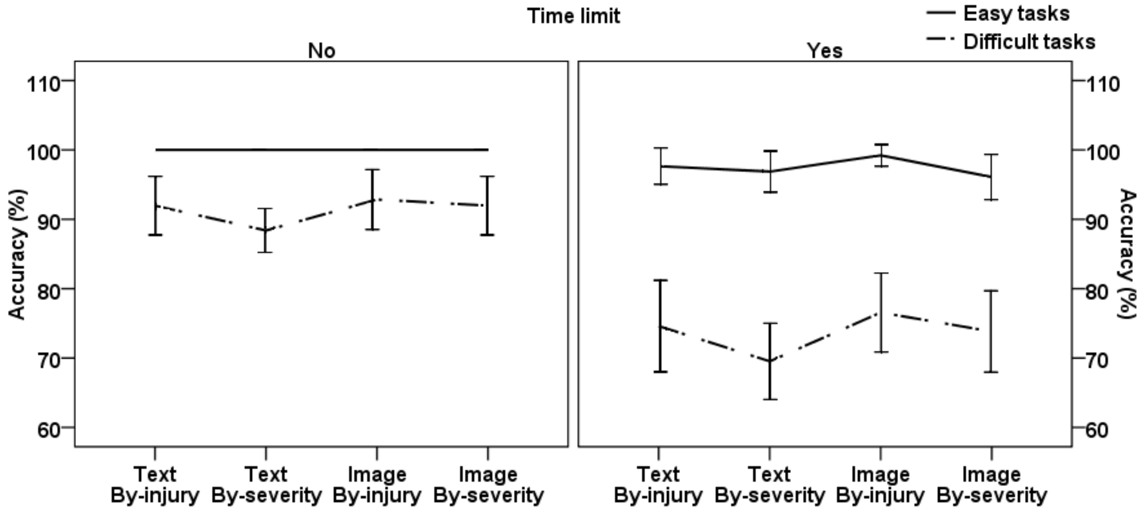


Figure 5.12: Decision accuracy of the easy and difficult tasks. Error bars represent standard errors.

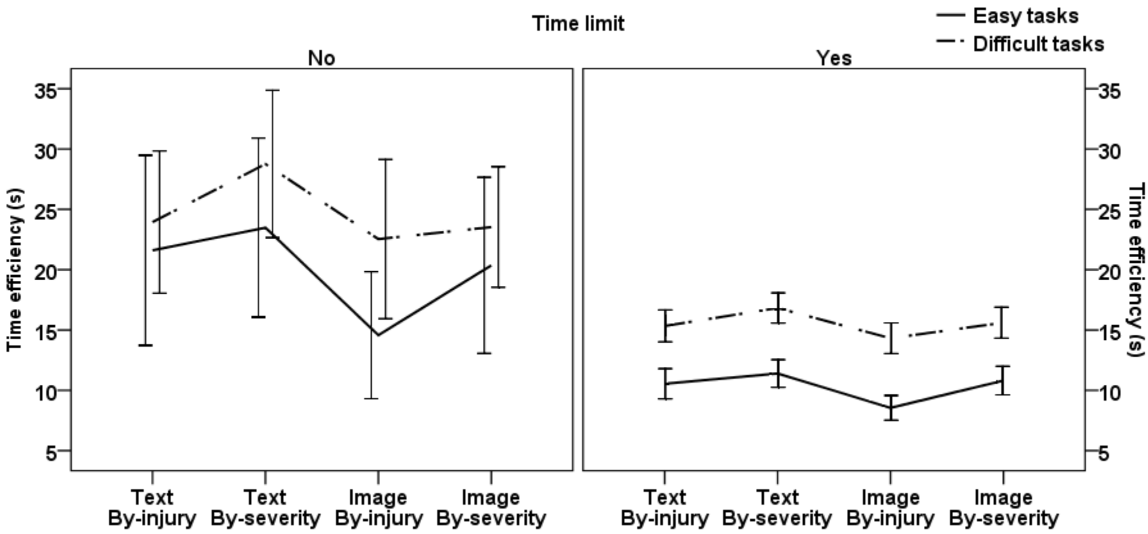


Figure 5.13: Time efficiency of the easy and difficult tasks. Error bars represent standard errors.

of the time limit harmed the performance of difficult tasks (17.7% decrease) more than the performance of easy tasks (2.5% decrease).

Time efficiency. The average time spent on an easy task was 20.0 s without the time limit and 10.3 s with the time limit. The average time spent on a difficult task was generally longer – 24.7 s without the time limit and 15.5 s with the time limit. ANOVA showed that the influence of task difficulty on time efficiency was significant ($F(1,46) = 47.4, p < .001$). No interaction effects were found between task difficulty and the presentation factors or the time limit, suggesting that the differences in time efficiency between the easy and difficult tasks were comparable in all task conditions.

5.4 Discussion

5.4.1 Modality

The modality factor significantly affected the time efficiency of decision making, both with and without the time limit. On average, the time spent on a decision was 2.7 s shorter with image than with text. 79.1% of participants reported that decision making was less (cognitively) effortful with image than with text, regardless of structure. 85.4% of participants chose an image condition to be the easiest task condition. However, the use of modality did not influence strategy selection or decision accuracy in this study.

In line with the stages of human information processing (Figure 3.1), modality in this task mostly influenced the perception stage – the stage when participants perceived injury items and severity levels from the images or text. Images had advantages over text in this study because they allowed information items to be quickly and easily perceived. The injury items were perceived faster from images of affected organs than from text. This is no surprise because images, as an analogue modality, are known to be effective for presenting concrete concepts and objects [21]. Our previous study with a visual search task also showed that performance was faster when search targets (patients) were presented by images than by text (Section 4.4.1). The severity levels of injuries were perceived faster from colors than from text, because color was more intuitive than text to convey severities. The current set of color-severity mappings was supposed to be particularly intuitive since it is being used in many real-world applications. Our results also confirm previous findings indicating that color is an effective means to indicate state, draw attention, and communicate qualitative/quantitative differences [200]. In general, graphical modalities (graphs and images) are more vivid than textual information (text and numbers), thus are likely to receive greater weight during decision making processes. In particular, shapes and colors have great salience to human information processors due to the sharp contrast they are able to create [145].

5.4.2 Structure

The structure factor had an effect on the time efficiency of decision making, both with and without the time limit. Structure also affected the accuracy of decision making with the time limit. The by-injury structure showed a clear advantage over the by-severity structure. On average, the by-injury structure allowed decisions to be made 2.4 s faster and 2.8% more correctly. Moreover, 68.8% of participants found the decisions easier to make with the by-injury structure and 70.4% of participants chose a condition with the by-injury structure to be the easiest condition.

In the decision making process, the structure of presentation mostly influenced the cognition stage – the stage when participants applied strategies to process the information items. The by-injury structure provided a better cognitive fit to the normative and the unbiased heuristic strategies, because it clearly indicated the weights of injury items and made eliminations more intuitive. When time was unlimited, the advantage of a better cognitive fit was reflected by the time efficiency, showing that less time was needed to reach a correct decision. When time was limited, however, participants developed biased strategies to speed up decision making. Strategy analysis showed that the selection of biased strategy was adapted to the presentation structure. The by-injury structure led to a more accurate strategy because it allowed participants to focus on the 3 more important injury items which determined the correct choice to a large extent.

5.4.3 Time Limit

The participants who were given a time limit all complied to it strictly. On average, they only spent 13.7 s on each task. In contrast, participants who did not have the time limit spent significantly longer on each task (10.4 s longer on average) and made significantly more accurate decisions (15.6% more on average). In combination, these results show that the time limit successfully imposed pressure on the decision making tasks. It seems that participants perceived higher time pressure than there really was, because they did not make full use of the 20 s. This might be due to the speech reminder given at the 15th second, causing most decisions to be made within 15 s. To cope with the time pressure, participants tended to focus on a subset of information (e.g. more important injury items) and to ignore certain rules (e.g. weights). This way decisions could be made faster with less cognitive effort, but at the cost of decreasing accuracy. Moreover, the time limit did not alter the subjective judgements on the presentation conditions. The cognitive load comparisons between presentation conditions showed a consistent outcome with and without the time limit.

5.4.4 Interactions Between Factors

Modality and structure

No interaction effect was found between the two presentation factors in any performance measures, indicating that their influences on decision making were independent. This is probably because the effect of these two presentation factors takes place at difference stages in the decision making process – modality mostly affects information perception and structure mostly affects the strategy of information processing.

Time limit and presentation factors

The time efficiency measure revealed interaction effects between time limit and the two presentation factors (with structure at 90% confidence level). In Figure 5.7, we can see that the variances of time efficiency between different presentation conditions were generally smaller with than without the time limit. At the first glance, this seemed to suggest that the time limit weakened the influence of presentation on time efficiency. However, this interpretation was incorrect, because all presentation conditions had the *same* time limit which was in most of the cases insufficient for decision making. Therefore, the main reason that the speed of decision making varied less was the time limit, but not the presentation factors. Further, the accuracy measure revealed that the influence of structure on decision accuracy became significant when the time limit was applied. Although the time limit decreased decision accuracy in all conditions, the decrease was less when the structure provided a better cognitive fit to the task. Overall, we would say that the time limit actually enhanced the effect of information presentation on decision making performance, because without a time limit, presentation factors influenced only the time efficiency of decision making, while with a time limit, they influenced both the time efficiency and accuracy of decision making.

Time limit and task difficulty

The accuracy measure revealed an interaction effect between the time limit and task difficulty. As Figure 5.12 shows, the application of a time limit slightly decreased the accuracy of easy tasks, but greatly decreased the accuracy of difficult tasks. However, the time efficiency measure did not show a similar interaction effect (see Figure 5.13), meaning that the time limit and task difficulty had independent influences on the time efficiency of decision making.

5.4.5 Consistency Between Performance and Subjective Measures

All subjective measures showed strong majority opinions that are consistent with the results of the performance measures. Further, we compared each participant's subjective judgements with his/her performance, attempting to see whether the consistency between the two types of measures also existed at an individual level. We were particularly interested in the group of participants whose preferences differed from the majority. It turned out that a consistency was often but not always found. For example, some participants who found the task easier in the text conditions still performed better in the image conditions. Generally, subjective options were more consistent with decision accuracy than with time efficiency. Most of the participants who preferred the text modality and the by-severity structure indeed made more accurate decisions in these conditions, but the decisions were still made faster with the image modality and the by-injury structure. Therefore, it seems that subjective assessment of task load can be a good indication of decision accuracy. On the one hand, there are design solutions that can provide a standardized optimal interface for a majority of users (e.g. the 'Image & By-injury' condition in the case of this study). On the other hand, in applications where the decision quality of every single user needs to be guaranteed, it might be wise to tune the standardized interface design for each user based on his/her individual preferences. Although lacking sufficient evidence, our study hinted that a user's modality preference can be influenced by his/her professional background and culture background.

5.5 Conclusions

In this study, we investigated the influence of information presentation and time limit on multi-attribute choice making. The findings of this study have several implications to information presentation for decision support.

First, the use of modality has a strong influence on the efficiency of information perception, which in turn affects the speed of decision making. Modality selection should suit the type of information contents to be presented and allow them to be perceived easily and quickly. For example, image is suitable for presenting concrete concepts and objects. Color coding is effective for indicating state (e.g. severity) and communicating qualitative/quantitative differences.

Second, decision makers tend to adapt their information processing strategies to the spatial structure of the presentation. As a result, presentation structure influences both the time efficiency and the accuracy of decision making. This is particularly notable when decision making is under time pressure. Therefore the choice of structure should provide a cognitive fit to the task, that is to support the application of desired decision making strategies. If several information items need to be considered together, they should be spatially clustered. When using a table, locate the more critical information at the top.

Finally, time limits enhance the effect of presentation format on decision making. Without a time limit, a suboptimal presentation format usually decreases the speed of decision making but not the accuracy. However, under time pressure, both time efficiency and accuracy can be affected. It might be unavoidable that decision accuracy drops under time pressure, but an optimal presentation format can limit this decline.

Part III

Information Presentation in the Automotive Context

6

Presenting Local Danger Warnings

Besides crisis management, driving is also a high-load task domain of interest. Driving is normally not a difficult task for experienced drivers. However in the case of an emergent danger, the driver needs to make quick and correct decisions, which can be high-load and stressful. In this chapter, we present two studies investigating the presentation of local danger warnings, using a scenario of emergent road obstacles. Study One focuses on the perception efficiency of warnings, aiming to find a visual presentation that can be perceived with little time and effort. Study two further investigates four modality variants (speech warning, visual and speech warning, visual warning with blinking cue, and visual warning with sound cue) and two levels of assistance (warning only or warning with action suggestions). In accordance with the ISO usability model, a total of seven measures were derived to assess the effectiveness and efficiency of the warnings and the drivers' satisfaction. The two studies presented in this chapter were conducted at the German Research Center for Artificial Intelligence (DFKI). The contents of this chapter have previously been published as [35], [36], [38] and [147].

This chapter is organized as follows. Section 6.1 introduces the background of the two studies, including how the two presentation factors were selected. Section 6.2 presents related work on modality (focused on the selection between visual modalities and speech) and the level of assistance. Section 6.3 and 6.4 respectively present the two studies, including the experimental methods, results and discussions. Finally Section 6.5 concludes this chapter and discusses future work.

6.1 Background

Local danger warning is an important function of in-vehicle information systems (IVIS) to improve the safety of driving. Local danger warning aims to warn drivers of dangerous situations coming up. The recent development of Car2X communication technology allows a car to exchange information with other cars and infrastructures via mobile ad hoc networks

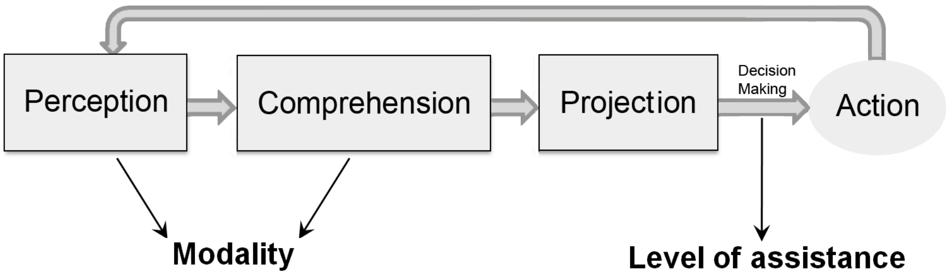


Figure 6.1: Illustration of the situation awareness theory and the factor selection of this study.

[127; 138; 258]. This technology enables a much wider application of local danger warning, because besides directly sensing the environment to detect danger [88], vehicles can also obtain information about local danger from other vehicles running in the same area. This can be particularly useful when the approaching danger is not yet visible to the drivers (e.g. due to a bend in the road or a large-sized vehicle ahead).

Compared with other IVIS functions, presenting local danger warnings is particularly challenging, because first, local danger warnings are usually low-frequency events, so drivers may be less familiar with how to react to them. Second, local danger warnings are highly urgent, so drivers usually have very limited time to think and react. Therefore, local danger warning messages should be communicated in a way that allows them to be picked up quickly (efficiency) and correctly (effectiveness).

To investigate how local danger warnings can be better presented, we first need to find out which presentation factors can be relevant in the driving context. We based our factor selection on the situation awareness (SA) theory from Endsley [78]. According to this theory, driving can be considered as a dynamic decision making task based on real-time maintenance of SA. SA has three hierarchical phases which all contribute to 'knowing what is going on' ([78], p .36) in a dynamic environment, as Figure 6.1 shows. The first step, perception, is to perceive the dynamics of relevant elements in the environment. The second step, comprehension, is to obtain an understanding of the perceived elements, including their significance to the task. The third step, projection, is to predict the future states of the environment. Finally, based on the updated SA, a decision can be made on how to react. For example, a driver D perceives a newly-presented warning message in the car (perception). D understands from the warning that there is a stationary vehicle on the roadside shortly ahead (comprehension). Then D predicts that there might be people moving around that vehicle (projection). Finally D decides on a significant decrease in driving speed in order to be able to pass safely (action). From this perspective, the presentation of local danger warnings should aim at assisting a timely update of the driver's SA, as well as helping him or her to make proper decisions.

Based on the situation awareness theory, two relevant presentation factors were identified, namely *modality* and *level of assistance*. The modality of warning presentation mostly influences the perception of warnings, and also influences comprehension to some extent. Level of assistance manipulates whether the selection of action is assisted or not. With a high level of assistance, the system not only delivers warnings, but also suggests to the driver which action to take. Action suggestions are expected to reduce the effort of projection and decision making, and thus accelerate the driver's response. Letting vehicles analyze the situation and make decisions for drivers may not generally be appreciated or trusted, but it may be of real help in a highly urgent situation, because the driver may fail to come up with a proper action in time.

We conducted a series of two studies investigating the presentation of local danger warnings. The scenario focused on a situation where road obstacles were a short distance ahead but not yet visible to the driver. Drivers received warning messages and needed to react quickly. Study One was a pilot study conducted in an office setting. It evaluated the perception time of several visual warning presentations and selected the best one to be used in the second study. Study Two was conducted in a driving simulator integrated into a real car. It further investigated eight presentation strategies, manipulated by two levels of assistance and four modality choices (visual, auditory and their two combinations).

6.2 Related Work

6.2.1 Modality: Visual vs. Speech Presentation

The advantages and disadvantages of three categories of modalities (visual, auditory and tactile) for in-vehicle information presentation has been described in Chapter 3 (Section 3.6). Our focus in this chapter is on visual and speech warnings. Therefore, we further discuss the selection between visual and speech presentations in existing IVIS studies.

Whether to present information visually or orally is a question that has often been investigated in IVIS studies. Theoretically, there is a conflict between a resource competition view and an attention preemption view (also see Section 3.6). From the resource competition view, speech is superior to visual modalities, because it consumes other perceptual resources than driving, and therefore the perception of speech can be better time-shared with driving. However, from the attention preemption view, speech is less suited than visual modalities, because it is highly salient and thus tends to fully occupy the driver's attention during its presentation. Empirically, both auditory benefits and visual benefits have been found by a number of studies. The auditory benefits were commonly demonstrated with a navigation-assisted driving scenario, in which drivers followed the instructions from a navigation device [33; 89; 131; 142; 237; 274]. These studies showed that when navigation instructions (e.g. turn notification) were presented aurally compared to visually, drivers reacted faster, made fewer errors and showed better driving performance in terms of speed and steering control. In addition, auditory benefits were also found in a couple of studies that used other types of secondary tasks, such as a warning detection task [198], a letter detection task [109] and an information searching task [202]. These secondary tasks interfered less with driving when relevant information was provided aurally compared to visually.

The auditory preemption effect or visual benefits were also found by a number of studies [106; 153; 171]. Concurrently with driving, drivers in [153] were asked to listen to/look at statements and reason whether they were true or false; drivers in [106] had to listen to/look at phone numbers of different lengths and recall them; drivers in [171] were asked to listen to/look at road sign information (e.g. speed limit, road number) and react if necessary. In these studies, auditory presentations interfered with driving more than their visual alternatives, causing greater variations in lane position, speed control and headway distance. However, concerning the secondary tasks, auditory presentations often led to faster reactions and better performance. In contrast, visual modalities showed an advantage of perceptional flexibility. Drivers could choose to attend to the visual display at a suitable/safe moment, or take multiple steps to read a message and return to driving in between.

Neither theories nor empirical findings revealed a clear winner between the two types of modality. In fact, the selection between visual and speech presentation is a high-dimensional problem, which means many factors play a role and the final choice cannot be made without evaluating these factors in a specified design context. After all, the choice should be made to let the potential benefit outweigh the potential damage. The following factors are relevant to this selection:

- **The relevance to driving (or the priority) of the message** [121; 219]
  When the IVIS information is relevant to driving, it usually has a high priority and requires a timely perception. Speech presentation is preferred in this case, because it is beneficial to have attention preempted to information that pertains to the driving task and is intended to support the driving task. Visual presentations lack salience, thus critical messages are likely to be overlooked. In contrast, when the information is not driving-related (e.g. a weather forecast), it has a low priority in the driving context. Visual presentation is more suitable in this case, because the driver needs to be able to temporarily ignore it if he/she has to concentrate on driving at that moment.

- **The spatial location of the visual display** [219]
  In the case of visual presentation, drivers must divide their focal visual attention between the driving environment (outside the windshield) and the in-car display. The larger the distance between display and windshield, the greater the (cognitive) cost of dividing one's attention between the two. Therefore, the advantage of visual modalities can be more pronounced when head-up displays are used than when head-down displays are used.

- **The length of speech** [233]
  Due to their transient characteristics, when speech messages get longer and/or more complex, they keep drivers' attention longer and impose a higher load on working memory. This can in turn enhance the auditory preemption effect. Besides, the full meaning of speech may not become clear until the end of the message. This makes long speech inappropriate for urgent warning messages that demand an immediate response. In short, speech has stronger advantages over visual presentations when it can be kept short and precise.

- **The mental workload** [109; 131; 219; 237]
  An increase in mental workload, due to either a more demanding driving task or a more complex secondary task, may increase the benefit of auditory presentation in relation to visual presentation. When driving is more demanding, the division of visual attention between the road and the in-car display becomes potentially more dangerous. When the secondary task is relatively complex, drivers simply have to spend more time looking at the in-car display in the case of a visual presentation. Speech is superior in both cases, because of its "eyes-free" feature.

- **The condition of the environment** [69; 166]
  Properties of the environment in and outside the car at the moment of presentation can also affect the utility of a certain modality. For example, speech messages might be less effective if the driver is at the same time talking to a passenger, getting a phone call, listening to the radio or playing loud music. Visual messages might be less effective on a sunny day when strong light makes it hard to see what is on the display.

- **The information type** [21; 238]
  This factor addresses the expressive power of modalities, because one modality can be naturally better than another at presenting a certain type of information. For example, speech is better at presenting instructions, commands and abstract information (e.g. logic, relations). Short speech is good for warnings and alarms. Regarding visual modalities, text is suitable for quantitative values (e.g. distance, speed, road numbers); icons are effective to indicate physical objects (e.g. gas stations along the highway) and directions (e.g. left/right turns); maps are good for locations and spatial information.

As both modalities have their own advantages and disadvantages, a question that naturally arises is: can a combined use of visual and speech presentation bring "the best of both worlds"? A couple of studies have shown a multimodal benefit [61; 142]. Drivers in these studies were required to perform secondary tasks based on IVIS messages related to navigation [142], vehicle status [142], or headway distances [61]. The results showed that the combination of modalities allowed better performance in both driving and secondary tasks, compared to using either single modality alone. The combination was also the most preferred modality variant by the drivers. However, the risk of using multiple modalities redundantly is to induce additional costs in terms of perception load, interface management and monitoring demand [216], which may not be beneficial in a high-load situation such as an emergent danger.

When presenting urgent warnings (e.g. local danger warnings), the major drawback of visual modalities is a lack of salience. Accordingly, we enhanced the salience of our visual warnings by adding attention-attracting cues, such as a blinking top bar or a sound beep. Then, we intended to evaluate and compare the enhanced visual warnings, speech warnings and their combinations.

6.2.2 Level of Assistance

Providing action suggestions to drivers has been investigated in a few studies. Using a headway maintaining scenario, the study presented in [120] found that the lowest level of assistance (providing information about leading cars) allowed the smoothest driving in terms of speed variance and was the best accepted by drivers. In contrast, giving a brake command to the drivers and having the vehicle brake automatically hampered the driving performance and were both less acceptable. However, the use of modality in this study differed in all conditions, so the findings might be confounded with a modality influence, as the authors also mentioned. Another study [136] compared a command message style (e.g. "Reduce speed") with a notification style (e.g. "Curve ahead"). Results revealed that the command style increased the drivers' compliance to the system. However, their trust in the system and their self-confidence both declined, because they were not provided with sufficient information and were forced to rely on the command messages. We inferred from this study that action suggestions should always be provided in combination with warning messages, because drivers should always have the opportunity to know why they are being given a suggestion to perform a certain action and then decide for themselves. We intended to compare two levels of assistance in our studies – providing warnings with and without action suggestions.

6.3 Study One

Before comparing visual warnings to auditory and multimodal warnings, we first needed a good design of visual warning that enhances the advantage of visual modalities – the time efficiency of perception. The objective of this study was to design several visual warning presentations using text and icon images, and to find out which presentation allows the message to be picked up the fastest.

Ten people (2 women and 8 men) participated in this experiment. All of them were German native speakers, between 25 and 45 years old. They were all researchers working in technical domains.

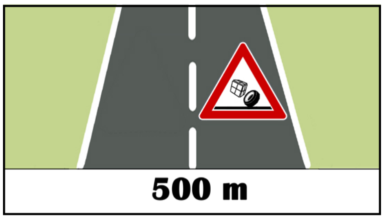
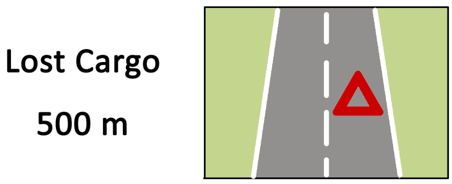
6.3.1 Warnings and Presentations

The local dangers used in this study were road obstacles on a two-lane highway. A warning message described an obstacle in terms of its type (what), location (where) and distance (how far). Four types of obstacles were included: broken-down vehicle, fallen tree, rock and lost cargo. Using visual modalities, the type of obstacles could be presented by text or by image. Figure 6.2 shows the image presentation of these obstacles. The location of obstacles could be in the left lane, in the right lane or at the right roadside[1]. Like obstacle type, location could also be presented by either text or image. The distance of obstacles from the

[1]In both Study One and Two, left and right lane/roadside is defined from the driver's perspective (when facing the direction where the vehicle is heading). The driver drives on the right lane.

Figure 6.2: Four types of obstacles were used in the study. From left to right: broken-down vehicle, fallen tree, rock, and lost cargo.

Table 6.1: The 5 warning presentation conditions used in this experiment. (Text and speech are translated from German.)

Condition	Example
text only	Lost cargo 500 m Right lane
icon only	
mixed 1	
mixed 2	
speech	Lost cargo in 500 m in the right lane

vehicle varied between 150 m and 2 km. The most effective way to present numerical information is to directly use numbers. Therefore, the distances were always presented with numbers and their units, such as "180 m", "1.5 km". Finally, four visual warning presentations resulted from the combination of obstacle type by text/image and obstacle location by text/image. The design of icons was validated in an informal survey in order to ensure intuitiveness. In addition, in order to enhance the salience of visual modalities, a red bar was added above the visual warning and it blinked for 3 s at the onset of each visual warning.

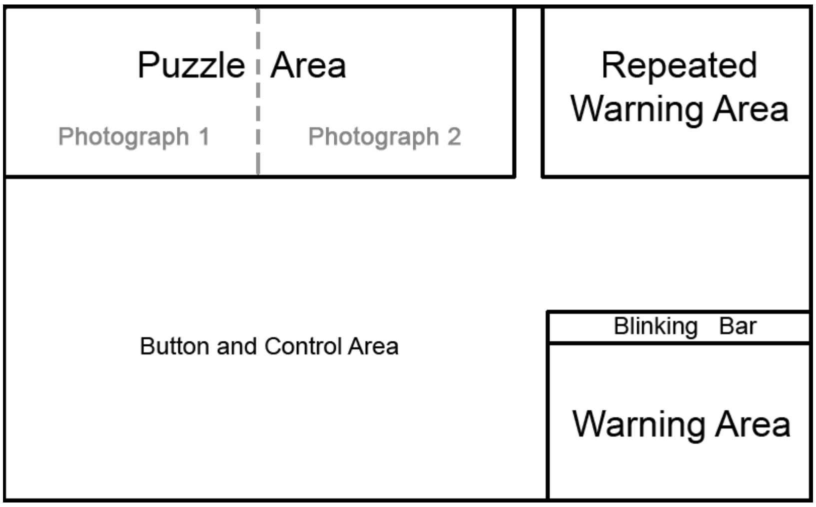


Figure 6.3: Location of task areas on the 20" widescreen monitor.

We also included a speech warning condition in this study, aiming to compare the perception time of visual and speech warnings. Speech warnings narrated the obstacle type, distance and location sequentially, such as "Broken-down vehicle in 180 meters on the right roadside" (translated from German message "Defektes Fahrzeug in 180 metern am rechten Strassenrand"). This sequence was also determined in the survey with 8 German native speakers. In the experiment, speech warnings were generated by the text-to-speech engine from SVOX[2]. Table 6.1 provides examples of the 5 warning presentation conditions.

6.3.2 Tasks

Two tasks were employed, a puzzle solving task and a warning perception task. They were both hosted on a 20" widescreen monitor, as shown in Figure 6.3.

Puzzle solving task

This task was to look for a required number of differences between two very similar photographs. The two photographs were located side by side on the top left corner of the monitor. Participants were instructed to click on a button every time when a difference was found, and to report all differences at the end of the experiment. We intentionally made this task difficult, so that it could not be completed before the end of the experiment. This way, all participants fixed their eyes on the same location before the onset of each warning message, allowing a more reliable measure of warning perception time. This task also shared a common feature with driving as they both require continuous visual attention. It was indeed found in the experiment that none of the participants managed to find all differences within the given time. The performance of this task was not analyzed.

[2]SVOX is a global supplier of embedded speech recognition, speech output and speech dialog solutions for the automotive and mobile industries. Website: www.svox.com

Warning perception task

While participants were searching for differences in the photographs, warning messages were presented at the bottom right corner of the same monitor. Participants were instructed to read the warning messages immediately when the red bar started blinking. A message was presented for a short time (e.g. 5 s) and then removed from the screen. Two seconds after the message was removed, it was repeated again on the top right corner of the screen together with a set of choice buttons: "same", "different", and "not sure". A repeated message was always in the same presentation condition as the original one, but one of the three information components (either type, location, or distance) was changed in some cases. Participants identified whether the repeated warning was identical to the original one or not, and then provided their choice by clicking on the corresponding button. SAME and DIFFERENT cases occurred at a 1:1 ratio and in a random order. This task did not require a long-term memorization of the message, because the time interval between the original and the repeated presentation was only 2 s. However, it did require participants to understand the warnings and realize what was on the road, where it was and how far away it was.

6.3.3 Measures

Perception time

To measure the perception time of a message, we manipulated the time duration in which a message was presented. In the four visual conditions, the initial presentation duration of each message was 5s, meaning that a message was kept on screen for 5 s and then removed. If the repeated warning could be compared correctly, this indicated that the duration was long enough for perceiving the message. The presentation duration was decreased with a step of 1 after every 3 warnings until the participant made the first error in comparing the repeated message. We took this final duration as the minimum time needed to perceive a message presented in this condition. In the speech condition, the length of the speech sentences was taken as the minimum perception time, and all participants performed 3 warning perception tasks.

Subjective preference

Due to a within-subject design, all participants performed all 5 conditions with a counterbalanced order. Afterwards, they were asked to indicate 1) which visual warning presentation they preferred the most, and 2) whether they preferred to receive visual warnings, speech warnings, or the combination of the two.

6.3.4 Hypotheses

Of the four visual warning presentations, the icon-only warnings were expected to be the best because both obstacle type and obstacle location are concrete information, and images were known to be more suitable than text for presenting concrete concepts and objects.

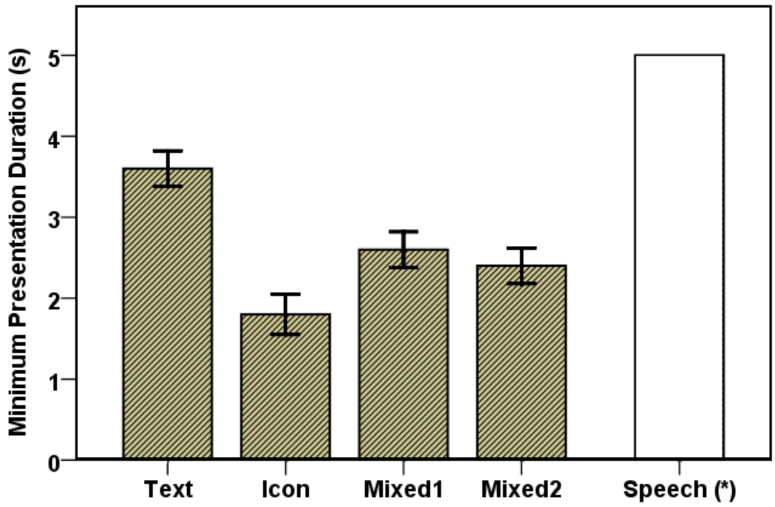


Figure 6.4: Perception time of warning messages for visual warning presentations. Error bars represent standard errors. (*) For speech, the minimum presentation length is determined by the utterance duration.

Accordingly, the text-only warnings were supposed to require the longest time to perceive. Between visual warnings and speech warnings, we expected that at least the best visual warning presentation would be perceived faster than speech.

6.3.5 Results

Warning perception time

The result of the perception time is presented in Figure 6.4. Of the four visual presentations, the text-only warnings required the most time to be perceived (M = 3.6 s). As expected, the icon-only warnings were perceived the fastest (M = 1.8 s). Not surprisingly, the perception time of the two mixed presentations lay in between. Considering that in both cases only one information component was replaced by an icon (either type or location), the improvement from 3.6 s for 'text only' to 2.6 s for 'mixed 1' and 2.4 s for 'mixed 2' is remarkable. For the speech condition, the minimum presentation length was determined by the time duration of the utterance, which was 5 s on average (Figure 6.4). The speech messages could be decoded reliably by all participants, except one who had difficulty in following the numerical information (distance) in the utterance. Repeated-measure ANOVA further showed a significant modality effect on the warning perception time ($F(3, 27) = 11.46$, $p < .001$), indicating that the use of modality could significantly influence the amount of time needed to perceive and comprehend the same information content. Helmert contrasts further revealed that 'text only' warnings took a significantly longer time to perceive than the other three types of warnings ($F(1,9) = 30.00$, $p < .01$); 'icon only' warnings took a significantly shorter time to perceive than the two mixed warnings ($F(1,9) = 8.65$, $p < .05$), between which no difference was found ($F(1,9) < 1$, $n.s.$).

Table 6.2: Subjective preference of the best visual warning presentation and subjective comparison between visual and speech warnings.

Visual warning presentations	Text	Icon	Mixed 1	Mixed 2
Number of votes	0	8	2	0
Vision vs. Speech	Speech	Vision	Combination	
Number of votes	4	2	4	

Subjective preference

Table 6.2 summarizes the subjective preferences for different warning presentations. Of the four visual warning presentations, the 'icon only' warnings were chosen by 8 out of 10 participants to be their favorite, because they found them the easiest to perceive and to understand. They commented that it was time-consuming to read a lot of text. Besides, when reading the message, they usually illustrated the text in their mind which required additional cognitive effort. The two mixed warnings were generally disliked because the three information components were spatially separated and thus required longer perception time. In contrast to the majority, 2 participants favored the 'mixed 1' warning. They explained that they tended to use sub-vocal speech to encode information components into the short-term memory, especially for the location of an obstacle. Therefore it was much more convenient when the location was presented with text.

When asked to compare visual warnings with speech warnings, 4 participants preferred speech. They stated that speech was more compatible with the on-going visual searching task. Two participants preferred visual presentations. When they were engaged in the visual searching task, they had a tendency not to process the speech even though they heard it. The other 4 participants preferred to receive both visual and speech messages, so that they could freely choose between the two presentations. They explained that they would mostly listen to the speech, but it was better to have visual presentations alongside so that they could read it again in case certain details needed to be recalled.

6.3.6 Discussion

The experimental results confirmed all our hypotheses. Of the four visual warning presentations, the 'icon only' warnings were perceived the fastest, the 'text only' warnings were perceived the slowest, and the two mixed presentations lay in between. This result again confirms the common finding that images are superior to text for presenting concrete concepts and objects ([21; 238], see also Chapter 4 and 5). It also stands in line with the suggestion from [121] that information of higher priority should be presented more iconically. Comparing visual and speech warnings, not only the best visual presentation, but all four visual presentations were perceived faster than speech, confirming the expected advantage of visual modalities – high efficiency in information delivery. The perception time of 'icon only' warnings was only 36% of the duration of speech.

Subjective preferences were generally consistent with the perception time measure. Of the visual presentations, 8 out of 10 participants favored the ‘icon only’ visual presentation, and the other 2 favored the mixed 1 warnings. Interestingly, these two participants were the only ones who worked on language related research topics, such as text retrieval and dialog management. Similar to the findings in Chapter 5, it seems that professional training background may be associated with people’s modality preference. It is hard to say whether working with text makes people prefer text or that people who prefer text choose to work with text. Moreover, when given speech as an option, most participants preferred speech warnings or speech and visual combined warnings.

The blinking top bar manipulation was effective in this experiment, because it immediately attracted attention to the visual warnings. This shows that the disadvantage of visual modalities (lack of salience) can be overcome by providing attention-attracting cues. However, in a driving setting, a blinking bar on an in-vehicle display may not be as effective, because the location of the display is usually more in the visual periphery than the location of the bar in this experiment, and drivers can have other distractions. Therefore, in the second study presented in Section 6.4, we applied more salient cues, such as sound cues.

In conclusion, we have obtained a visual warning presentation that can be attended to on time and efficiently perceived, namely the ‘icon only’ warning with attention-attracting cues. In the next study, we will investigate visual, speech, and the combined warning presentation in a driving task setting, as well as the level of assistance.

6.4 Study Two

In this study, we investigated the presentation of local danger warning in a driving task setting. Presentation variations were manipulated by *modality* and *level of assistance*. The evaluation of warnings was based on the usability assessment guidance from ISO (ISO 9241-11, [1]). This study had three objectives. First, we aimed to find out whether local danger warnings could indeed enhance driving safety. This was investigated by comparing drivers’ danger avoidance performance with and without warnings. Second, we intended to evaluate different presentation variations and find out which modality(ies) and level of assistance were the most suitable for presenting local danger warnings. We also wanted to observe the interaction between the two presentation factors. Third, the driving condition simulated in this study did not resemble all situations that might occur in a real driving environment, therefore we intended to obtain subjective judgements on how useful the warnings would be in various real driving situations.

6.4.1 Apparatus

This experiment was conducted using a driving simulator integrated into a real car (Mercedes-Benz R Klasse). Participants were able to control the vehicle as they normally would in real driving. The simulator software was hosted by a standard PC. The driving scene was projected onto the windshield and was updated at approximately 25Hz. The vehicle stood

Figure 6.5: A subject driving in the experiment.

indoors enclosed by extra shields to reduce ambient light. This ensured a good visibility of the projection at all times of the day. Visual warning messages were presented on a 10.6-inch head-down display mounted next to the steering wheel on the right-hand side. The display was also a touch screen through which the recall task was performed. Auditory signals (speech and sound) were delivered through a PC speaker located in the center of the vehicle. A web camera was mounted on the dashboard to record the frontal facial view of the driver throughout the experiment. Figure 6.5 shows a subject driving in the experiment.

6.4.2 Message Design and Presentation

The message design remained the same as in Study One. A warning message consisted of obstacle type, location and distance. Obstacle types could be broken-down vehicle, fallen tree, rock or lost cargo. The location could be in the left lane, in the right lane or at the right roadside. The distance varied between 150 m and 180 m. The three information components were not equally relevant to how drivers should react to the obstacle (explained in the next subsection). In order to let all 3 components be carefully perceived, we added some irrelevant messages which were not to be reacted to. Each irrelevant message had 1 of 3 three components being 'out of range'. The irrelevant type, location and distance were respectively air traffic, on the left roadside and more than 10km ahead. Figure 6.6 shows some examples of irrelevant messages. Note that the irrelevant messages were included for experimental purposes only; it should never be a function of real ADAS systems to give false alarms.

Warning presentation was manipulated by two factors: *modality* and *level of assistance*. The *modality* factor had 4 variants: visual + blinking cue, visual + sound cue, visual + speech, and speech only. For visual warnings, the icon-only presentation from Study One

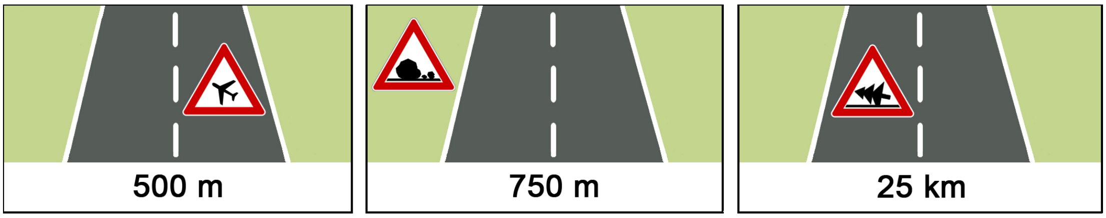


Figure 6.6: Examples of irrelevant messages. From left to right, the irrelevant components are respectively type, location, and distance.

Table 6.3: Presentation conditions used in the experiment.

Condition Index	1	2	3	4	5	6	7	8
Action Suggestion					×	×	×	×
Visual Message	×	×	×		×	×	×	
Speech Message	×			×	×			×
Beep Sound		×				×		
Blinking Bar			×				×	

was applied (see Table 6.1). The blinking cue was provided by a flashing red bar above the warning display area (see Figure 6.3). The sound cue was a beep tone that lasted for about 350 ms. All cues were delivered at the onset of a visual warning. Speech warnings narrated the obstacle type, distance and location sequentially, as in Study One. The *level of assistance* varied between with or without action suggestions (AS). AS were always given in speech, such as “Change lanes!”or “Brake!” for warnings and “Attention!” for irrelevant messages.

A full factorial design of the two presentation factors resulted in 8 (4 × 2) presentation conditions. They are summarized in Table 6.3. In the visual channel, visual message and blinking bar started simultaneously, when both present. In the auditory channel, if more than one component was included, they were presented sequentially. The order from first to last was beep sound, AS and speech message.

6.4.3 Tasks

The basic driving task was to drive on a one-way highway with two lanes. No extra traffic was involved. Using the gas pedal and the brake pedal, drivers could control the driving speed at two levels – 120 km/h and 60 km/h. For example, when the speed was 120 km/h, pressing the brake pedal would change it to 60 km/h immediately. Then pressing the gas pedal would return the speed to 120 km/h again. The basic requirement of driving was to drive at 120 km/h in the right lane. Obstacles were placed at random intervals between 800 m and 1300 m. Upon reception of obstacle warnings, drivers needed to react to the obstacles and recall the warnings.

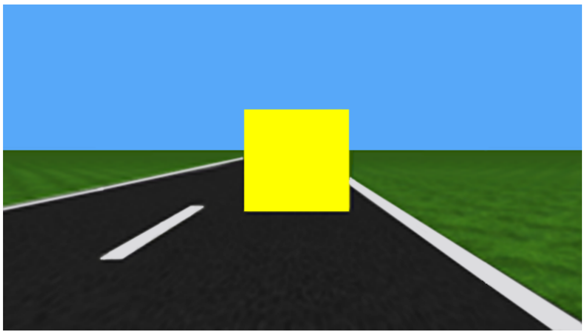


Figure 6.7: An obstacle in the right lane. The color can be yellow, green, blue or red.

React to obstacles

Drivers were required to 1) change to the left lane if the obstacle was in the right lane, or 2) brake if the obstacle was in the left lane or at the right roadside. After passing the obstacle, they were to change back to the right lane or accelerate to the higher speed again. When a warning was delivered, the obstacle was not yet visible to the driver. It only appeared on the scene when it was 40 m ahead of the vehicle, in the form of a colored box (see Figure 6.7). Drivers were instructed not to wait for the obstacles to appear. Instead, they should react to the warnings as soon as they had decided on what to do. Given the distance and speed settings in this experiment, there was no risk of acting unnecessarily early. Regarding AS, they were given total freedom in how to react to them, from totally relying on them (without processing warnings) to totally ignoring them (relying only on warnings). Irrelevant messages did not require any reaction and were supposed to be ignored in all conditions. Furthermore, brake or lane change actions were to be completed at a safe distance from the obstacles. We defined a distance of more than 20 m to be safe, which meant that the speed should have been decreased to 60 km/h or the vehicle should have been in the left lane when the car reached 20 m in front of the obstacle. A low pitch error sound was delivered in cases of a late or missed reaction.

Recall warnings

Speech warning messages all finished before the obstacles appeared. Visual messages were taken off the display when the obstacles had been passed. During the intervals between two obstacles, drivers were asked to recall one information component in the most recent obstacle. A question was asked via speech, regarding either the type, the distance (mentioned in the warning message), or the color of the obstacle. For each question, drivers answered by pressing one of the four options displayed on the touch screen. Options of obstacle type were given in icon images, as shown in Figure 6.2. Options of obstacle distance were given in numbers and units. Options of obstacle color were given in the form of colored squares. Since color was not mentioned in the warning, the reason for testing on colors was to know how much attention drivers paid to watch the environment outside the vehicle. No questions were asked after irrelevant messages.

Table 6.4: Summary of measures.

Parameter	Measures
Effectiveness	1. Unsafe behaviors 2. Correct recalls 3. Reactions to irrelevant messages
Efficiency	4. Reaction Time 5. Driving activity load (DALI: Effort of attention, Visual demand, Auditory demand, Temporal demand, Interference and Situational stress)
Satisfaction	6. Satisfaction with warnings (in this experiment) 7. Situation dependent assessment of usability (in 5 situations that commonly occur in real driving)

6.4.4 Subjects and Procedure

32 drivers participated in this experiment, 16 men and 16 women. Their ages varied between 20 years and 62 years (M = 32.6, SD = 10.8). All participants had been in possession of a valid driver's licence for at least two years. They were all native German speakers. Each driver was paid 15 Euros for approximately 2 hours of experiment time.

When entering the car, drivers first adjusted the seat to their comfort. The experiment had a training session and a task session. In the training session, drivers first received an introduction about the driving requirements, warning messages, presentation conditions and tasks. Then, they drove two practice tracks of about 15 minutes in total. In the first track, obstacles suddenly appeared on the road without warning and drivers had to react as soon as they could. The second track included all 8 types of warning presentations, as well as the irrelevant messages. The warning recall task was also practiced. A short break was given after the training session was completed.

In the first track of the task session, drivers encountered 16 obstacles without warnings. Drivers' performance in this track showed their ability to react to urgent situations without additional assistance. Afterwards, based on a within subject design, each driver drove 8 tracks with different presentation conditions. The track order was counterbalanced with a size-8 Latin square. Each track lasted for about 5 minutes, containing 8 warning messages and 3 irrelevant messages. The message order was randomized. During the short breaks between two tracks, drivers filled in questionnaires, reporting the driving activity load and assessing the warning presentation in the previous track. At the end of the experiment, an open questionnaire was provided to obtain more subjective feedback.

6.4.5 Measures

To evaluate the warning presentations, we derived a set of 7 measures (summarized in Table 6.4) based on the usability assessment guidance from ISO (ISO 9241-11, [1]).

ISO 9241-11

ISO 9241 is a standard for "Ergonomics of Human System Interaction". Part 11 is a guide on usability, dealing with the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use. *Effectiveness* is the accuracy and completeness with which users achieve specified goals. In our driving context, effectiveness can be reflected by drivers' performance of danger avoidance and their level of situation awareness (SA). *Efficiency* is the amount of (temporal and cognitive) resources expended to achieve the goals, which in our case can be measured by the reaction time to the warnings and the subjective evaluation of driving load. *Satisfaction* is the users' positive attitude (comfort and acceptability) towards the use of the interface, which can be obtained in terms of subjective ratings.

Measures of effectiveness

An effective warning is supposed to enhance driving safety and drivers' SA. For driving safety, we measured the number of *unsafe behaviors* which was the percentage of obstacles that were not passed safely. We defined three types of unsafe behaviors: 1) incorrect reaction, such as lane change instead of braking or the other way around; 2) late reaction, which was performed less than 20 m from an obstacle; and 3) no reaction to an obstacle warning. SA can be assessed "on the go" by placing recall tasks along the driving course [77]. The *correct recalls* measured the percentage of recall questions that were correctly answered. Another measure of SA was the *reactions to irrelevant messages* which measured the number of times drivers unnecessarily braked or changed lane after irrelevant messages. This measure reflected how well drivers were aware of the situation conveyed by the irrelevant messages.

Measures of efficiency

The efficiency of warnings was evaluated by the reaction time and the subjective ratings for driving load. The reaction time was defined as the time interval between the moment when a warning presentation started and the moment when an action was performed. A brake action was identified when the speed changed from 120 km/h to 60 km/h. A lane change action was identified when the lateral displacement of the car reached 10% of the maximum lateral displacement during the course of a lane change (illustrated in Figure 6.8). To obtain subjective evaluation of the driving load, we used the Driving Activity Load Index (DALI), which is a revised version of the NASA Task Load Index adapted to the driving task [192; 193] (see also Section 1.4.2). It contains 6 factors: effort of attention, visual demand, auditory demand, temporal demand, interference, and situational stress. Each factor can be rated on a Likert scale from 0 (low) to 5 (high). Appendix C.1 provides a detailed description of the DALI factors and the questionnaire itself.

Measures of satisfaction

The subjective satisfaction of warnings was assessed by two measures. First, drivers rated how satisfied they were with each warning presentation in this experiment, using a Likert

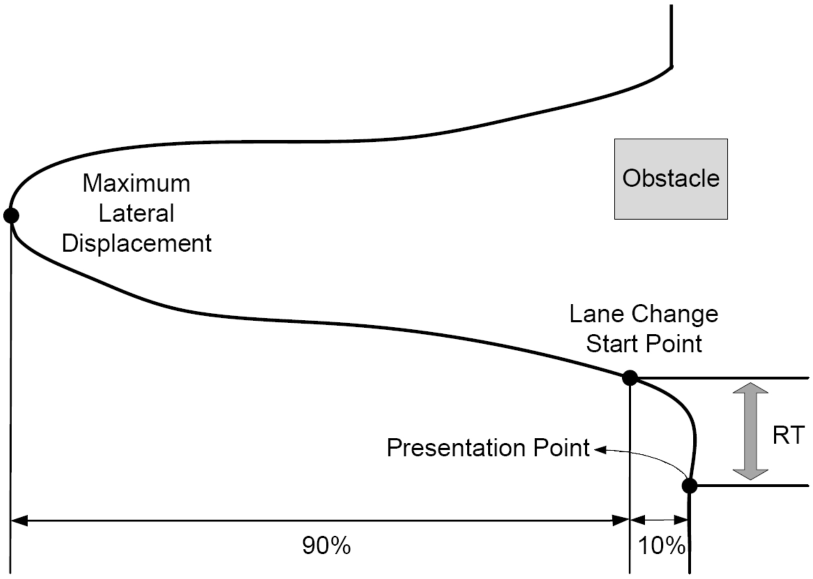


Figure 6.8: The calculation of the lane change starting point.

scale from -4 (very unsatisfied) to 4 (very satisfied), see Appendix C.2. For this measure, drivers always had access to the ratings they had already made for previous conditions, so that they could make adjustments to maintain the differences between conditions. Second, we are aware of the fact that real driving conditions can be very diverse and are usually not as ideal as the one simulated in this experiment. Therefore, we also asked drivers to judge the (expected) usefulness of each warning presentation in 5 driving situations, based on their real life experiences. Ratings were performed on a Likert scale from 0 (not useful at all) to 5 (very useful), see Appendix C.3.

1. *Rich sound*: driving with rich surrounding sounds (e.g. noise, radio, conversation)
2. *Low visibility*: driving with a low visibility (e.g. at night, heavy rain, fog)
3. *Fatigue*: driving when tired and unconcentrated
4. *Long drive*: driving on a long and boring trip (e.g. many hours on the highway)
5. *High demand*: driving in highly demanding situations (e.g. in heavy traffic, in an unfamiliar city)

6.4.6 Results

Unsafe behaviors

During the drive without obstacle warnings, all drivers except one showed unsafe behaviors. On average, 19.1% of the obstacles were not passed safely, because drivers reacted either incorrectly or too late. There were more unsafe lane change reactions than unsafe braking reactions. This baseline performance indicates that it is indeed a challenge for drivers to react to obstacles at short notice.

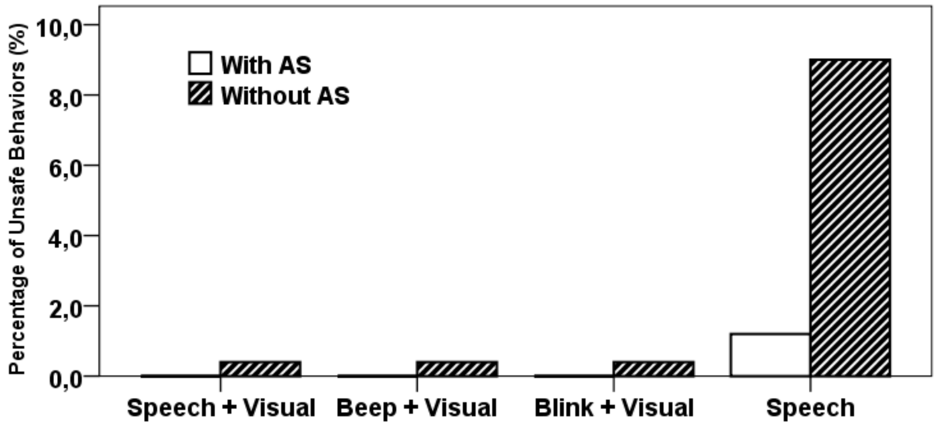


Figure 6.9: Percentage of obstacles that were not passed safely in each condition.

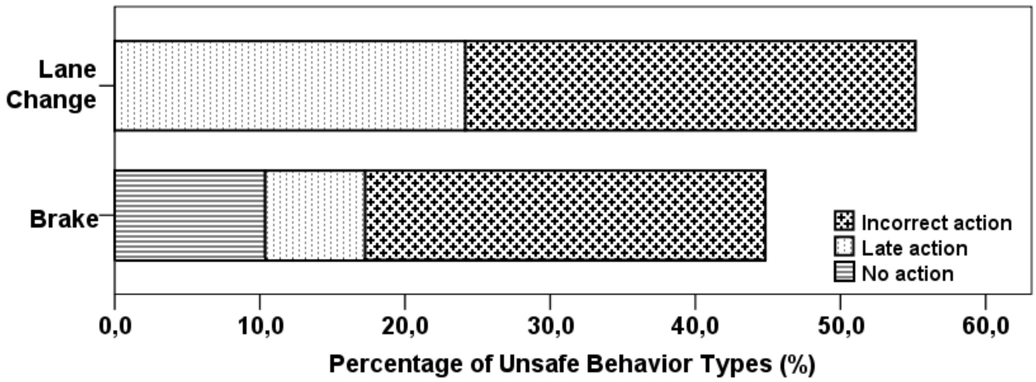


Figure 6.10: Distribution of unsafe behaviors over action and error types.

With the assistance of warnings, 18 drivers (56.3%) safely avoided all obstacles in all presentation conditions. The percentage of unsafe behaviors was reduced to 1.4% on average. As shown in Figure 6.9, when AS were provided, unsafe reactions only occurred in the speech condition (1.2%), because some participants did not rely on the AS and did not react in time after the speech messages. When AS were not provided, the number of unsafe behaviors increased in all modality conditions, especially when speech was used on its own ($M = 9.0\%$).

Figure 6.10 further shows the distribution of all unsafe behaviors over the two action types and three error types. Most unsafe situations were caused by incorrect actions, adding up to 58.6% from both action types. In another 31.1% of the cases, brake or lane change actions were performed correctly but too late. In the remaining 10.3% of the cases, participants did not react at all. Late actions and no actions all occurred in the speech conditions.

A two-way repeated-measure ANOVA further revealed a significant assistance effect ($F(1, 31) = 15.8, p < .001$) and a significant modality effect ($F(3, 29) = 5.1, p < .01$). The higher level of assistance (with AS) led to safer driving than the lower level (without AS). Of the four modality variants, the performance was equally good for the three variants that included visual modalities. However, driving safety decreased significantly when speech

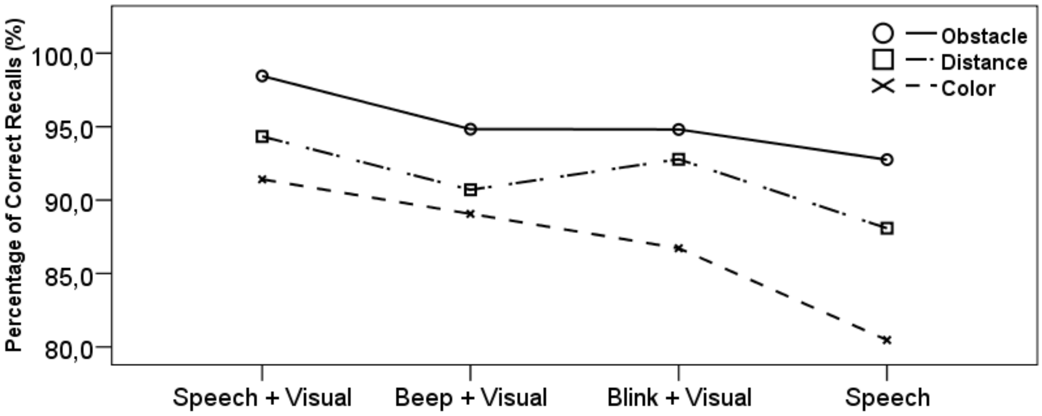


Figure 6.11: Percentage of correct recalls, calculated within each modality variant (averaged over the assistance levels) and each question type.

was used alone, compared to the other three variants ($F(1, 31) = 16.2$, $p < .001$ by Helmert contrast). There was also a significant interaction between modality and assistance level ($F(3, 29) = 5.5$, $p < .01$), indicating that presenting warnings using speech alone was much less harmful to the driving safety when AS were provided.

Correct recalls

The recall task was generally performed well. On average, 91.7% of the questions were answered correctly. The performance was the best when AS, speech and visual information were all provided (96.1% correct), and was the worst in the two speech conditions (87.9% correct). Repeated-measure ANOVA confirmed a significant modality effect on the performance of recall ($F(3, 29) = 3.5$, $p < .05$). Combining speech and visual messages led to significantly more correct recalls than the other three modality variants ($F(1, 31) = 10.2$, $p < .01$, by Helmert contrasts). Speech only warnings resulted in the worst recall performance ($F(1, 31) = 6.6$, $p < .05$). No significant difference was found between the two visual conditions with cues. The level of assistance did not influence the recall performance either ($F(1, 31) = 0.03$, *n.s.*). This is not surprising, because the action suggestions did not contain information relevant to the questions.

Besides modality, the topic of the questions also had an effect on the performance of recall ($F(2, 30) = 8.0$, $p < .01$). As shown in Figure 6.11, the color of obstacles was recalled the worst, compared to the type and the distance ($F(1, 31) = 8.9$, $p < .01$). A possible explanation is that participants paid less attention to the color of obstacles, because it was not mentioned in the warnings and irrelevant to the driving task. Especially in the speech condition, drivers were under a higher time constraint to react to obstacles, and thus paid less attention to the environment outside the car. In addition, the obstacle type was recalled better than the distance ($F(1, 31) = 9.8$, $p < .01$), which was probably because the icon presentations of obstacles were more vivid than the text presentations of distances.

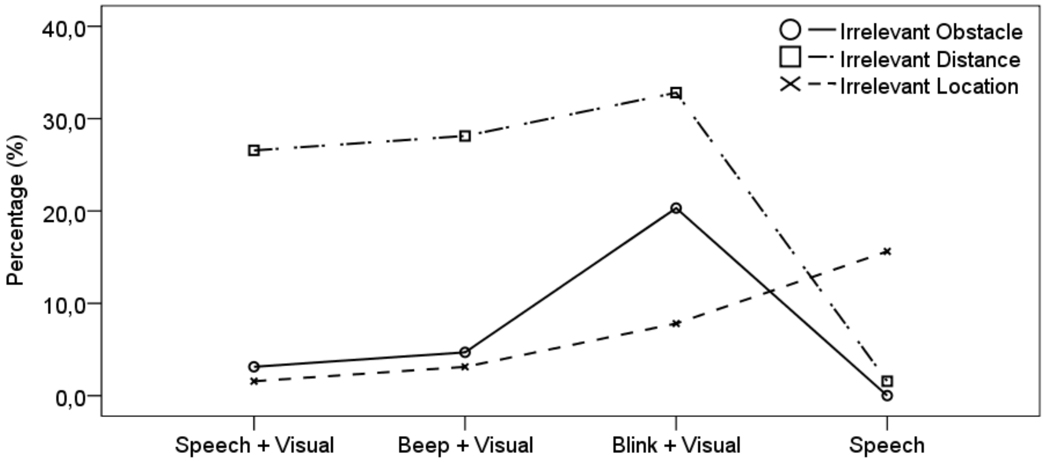


Figure 6.12: Percentage of irrelevant messages that were unnecessarily reacted to, calculated within each modality variant (averaged over the assistance levels) and each irrelevant type.

Reactions to irrelevant messages

On average, drivers reacted to 12.1% of irrelevant messages. In most cases, drivers self-corrected their actions very soon. The number of unnecessary reactions was influenced by the modality factor ($F(3, 29) = 5.2$, $p < .01$) but not by the level of assistance ($F(1, 31) = 0.1$, $n.s.$). The latter finding is understandable since the suggestions for irrelevant messages (“Attention!”) did not specify any action. Of the four modality variants, unnecessary reactions occurred the most when messages were presented visually with the blinking cue ($M = 20.1\%$). Helmert contrast showed a significant difference between this variant and the other three ($F(1, 31) = 9.8$, $p < .01$). In contrast, unnecessary reactions occurred the least when speech was used on its own ($M = 8.6\%$), which seems to contradict the results of the previous two measures (speech only is the worst modality variant in both measures).

We further analyzed the reactions to the three types of irrelevant information separately. As Figure 6.12 shows, pure speech warnings had an advantage when the obstacle type and distance were irrelevant. This was due to the fact that obstacle location was always presented at the end of a speech warning, which guaranteed the perception of obstacle type and distance before any action could be performed. However, when the location was irrelevant, pure speech warnings still produced the most unnecessary reactions, which is consistent with the previous the measures.

Besides the speech only variant, Figure 6.12 shows a consistent pattern between the other three modality variants: drivers reacted most frequently to irrelevant distances ($F(1, 31) = 29.5$, $p < .001$ by Helmert contrast: distance vs. other two), and more often to irrelevant obstacles than to irrelevant locations. This pattern suggests that drivers may have followed a common sequence to scan visual messages: first location, then type, and distance at last. Location was perceived first because it was the most relevant to the reaction. The type was spatially integrated with the location and only required a more detailed perception of the icon. Distance was perceived the last because of its spatial separation from the icon.

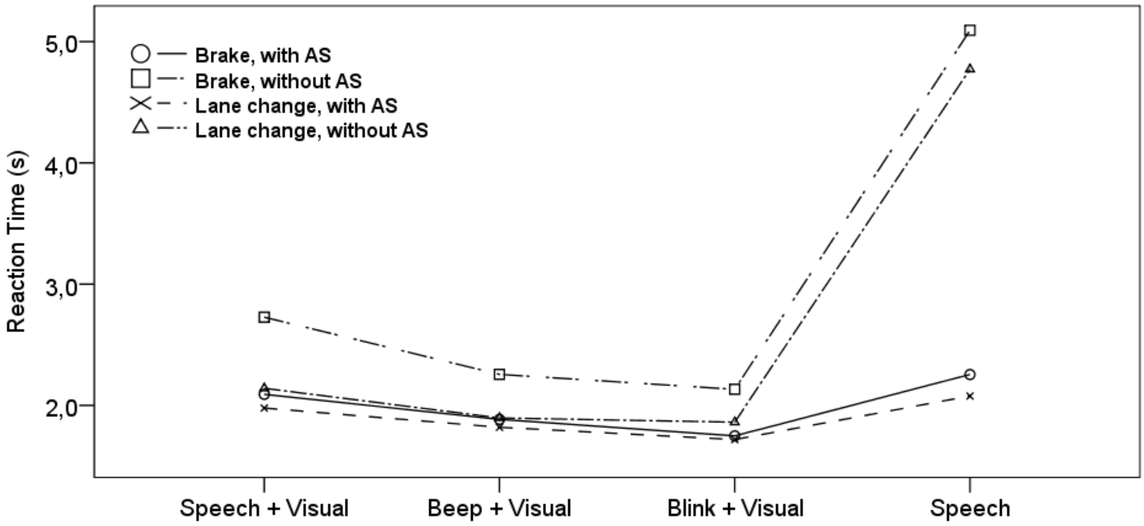


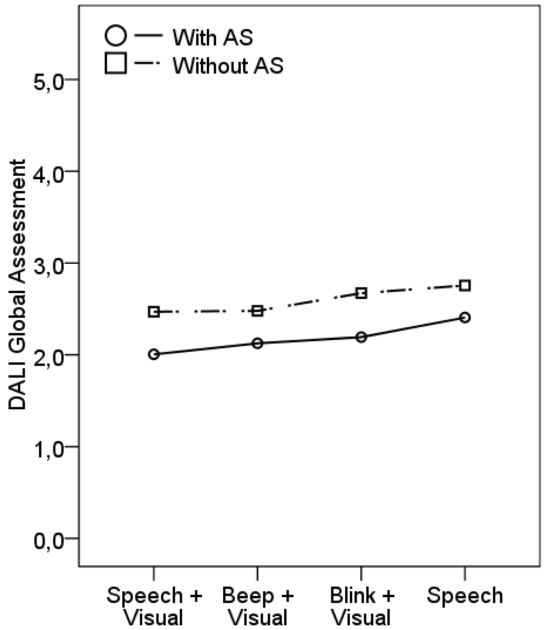
Figure 6.13: Average reaction time in each condition for brake and lane change respectively.

Reaction time

First, we looked at whether reaction time differed between the brake and the lane change actions. A three-way repeated measure ANOVA revealed that the action type did significantly influence the reaction time ($F(1, 31) = 15.4$, $p < .001$). On average, the lane change actions were performed 0.24 s faster than the brake actions. Note that the starting point of lane change reactions was defined when the vehicle already had 10% lateral displacement (see Figure 6.8), which means the reactions were actually even faster than what we measured. This result falls in line with previous findings stating that steering is $0.15 \sim 0.3$ s faster than braking, because of a lower response complexity [93]. Due to this difference, further analyses were conducted separately on the two types of actions.

When speech warnings were provided without AS, the reaction time was particularly long, because no action could be performed before the end of the speech. As Figure 6.13 shows, the average reaction time in this condition was 5.1 s (170 m) for braking and 4.8 s (160 m) for lane change. When AS were provided, the reaction time to pure speech warnings was reduced to 2.3 s for braking and 2.1 s for lane change. However, they were still the longest among all modality variants when AS were provided.

Besides the two pure speech conditions, the difference in reaction time between the other six conditions was relatively minor. However, ANOVA still revealed significant modality effects (braking: $F(2, 62) = 25.0$, $p < .001$; lane change: $F(2, 30) = 22.0$, $p < .001$) as well as significant assistance effects (braking: $F(1, 31) = 27.9$, $p < .001$; lane change: $F(1, 31) = 24.7$, $p < .001$). Regarding the level of assistance, both braking and lane change were performed faster with AS than without AS. On average, AS accelerated braking for 470 ms (15.7 m) and lane change for 130 ms (4.3 m). Of the three modality variants with visual modalities, the combined use of speech and visual messages led to the longest reaction time for both braking reactions ($F(1, 31) = 44.8$, $p < .001$) and lane change reactions ($F(1, 31) = 44.6$, $p < .001$). On average, both brake and lane change were per-



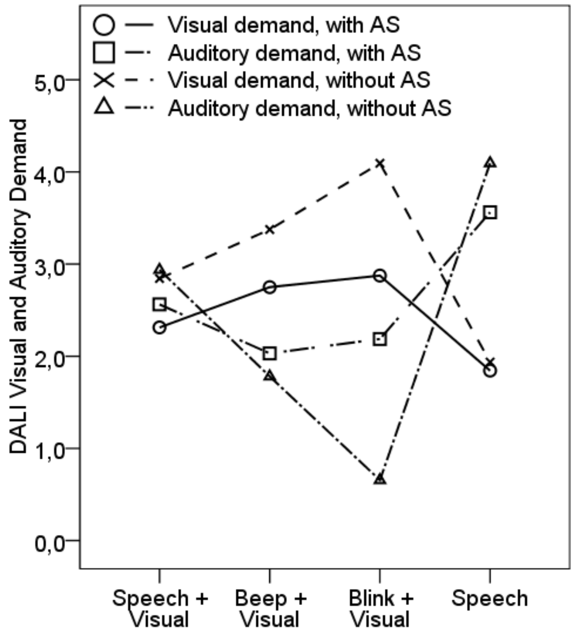


Figure 6.14: Result of DALI rating. Left: average rating score over 6 dimensions; Right: visual demand and auditory demand.

formed faster with the blinking cue than with the beep cue. However, the difference was only significant for lane change ($F(1, 31) = 6.9$, $p < .05$).

A somewhat reversed pattern was found when comparing the reaction time with the drivers' reactions to irrelevant messages (Figure 6.12 vs. 6.13). It seems that the longer the reaction time was, the fewer unnecessary reactions there were. This observation suggests a variation in the level of vigilance. When warnings made drivers more vigilant (e.g. due to the salience of auditory modalities), they seemed more willing to inspect the warnings carefully before reacting to them, and thus the reaction time was slightly longer.

Driving activity load

Using the DALI questionnaire, the total driving load (driving while perceiving warnings and avoiding obstacles) was rated on six dimensions: effort of attention, situational stress, visual demand, auditory demand, temporal demand, and interference. First, we averaged the rating scores over the six dimensions as a global assessment of driving load. ANOVA revealed that the global driving load was influenced by both the modality ($F(3, 93) = 5.0$, $p < .01$) and the assistance level ($F(1, 31) = 23.4$, $p < .001$). As shown in Figure 6.14 (left), the tasks were considered less demanding with AS ($M = 2.2$) than without AS ($M = 2.6$). Regarding modality, the tasks were rated the least demanding when speech and visual warnings were combined ($F(1, 31) = 5.8$, $p < .05$, by Helmert contrast), and the most demanding when only speech was used ($F(1, 31) = 5.3$, $p < .05$). No significant difference was found between the other two modality variants.

Further, we analyzed each dimension separately. Besides the two perceptional dimensions (visual and auditory demand), all the other four showed similar results to the global

Table 6.5: Comparison of DALI ratings between driving with and without warnings (t-test).

DALI Dimension	Mean with warnings	Mean without warnings	Sig.
Effort of attention	2.7	4.0	$t(31) = -7.8, p < .001$
Visual demand	2.8	3.8	$t(31) = -5.0, p < .001$
Temporal demand	1.9	2.2	$t(31) = -1.7$, *n.s.*
Situational stress	2.2	3.5	$t(31) = -7.1, p < .001$

assessment. To validate this observation, we conducted a reliability analysis on these four dimensions, for the 8 conditions respectively. Results showed that the Cronbach's Alpha (coefficient of internal consistency) values were all greater than 0.8. This means that ratings of these four dimensions are indeed highly consistent, and together they determine the result of the global assessment. The two perceptional dimensions showed different results because the visual and auditory demand were mostly influenced by the use of modalities for warning presentation. As shown in Figure 6.14 (right), the visual demand of driving was the lowest in the speech warning conditions (no visual warnings to perceive). The auditory demand was the lowest with purely visual warnings and the highest with pure speech warnings. AS generally decreased perception demands (particularly visual), probably because less effort was needed to analyze the warning messages when drivers were told what to do. The combination of visual and speech warnings decreased both visual and auditory demand compared to when only one modality was used. This indicates that drivers divided the warning perception task between two modalities when they were both available.

Finally, we compared DALI ratings for the baseline drive without warnings to the average ratings over the 8 warning conditions, using t-tests. When warnings were not provided, there was no auditory demand and no in-vehicle messages interfered with driving. Therefore, the comparison was conducted in 4 DALI dimensions. As Table 6.5 shows, the effort of attention, stress, and visual demand of driving (and avoiding obstacles) were rated significantly higher without warnings than with warnings. On average, the temporal demand was also higher without warnings than with warnings, but the difference was not significant.

Satisfaction with warnings

On a Likert scale from -4 (very unsatisfied) to 4 (very satisfied), the satisfaction scores were generally positive, except for the purely visual warnings without AS (see Figure 6.15). The condition with speech, visual warnings and AS was rated the highest. ANOVA showed that the subjective satisfaction was affected by both the modality ($F(3, 29) = 22.1$, $p < .001$) and the level of assistance ($F(1, 31) = 43.2$, $p < .001$). Participants were more satisfied with the higher level of assistance ($M = 2.8$) than the lower one ($M = 1.9$). Of the 4 modality variants, Helmert contrasts showed that combined speech and visual warnings were the most satisfying ($F(1, 31) = 33.0$, $p < .001$), and visual warnings with beep cues were the second best ($F(1, 31) = 8.7$, $p < .01$). No significant difference was found between the

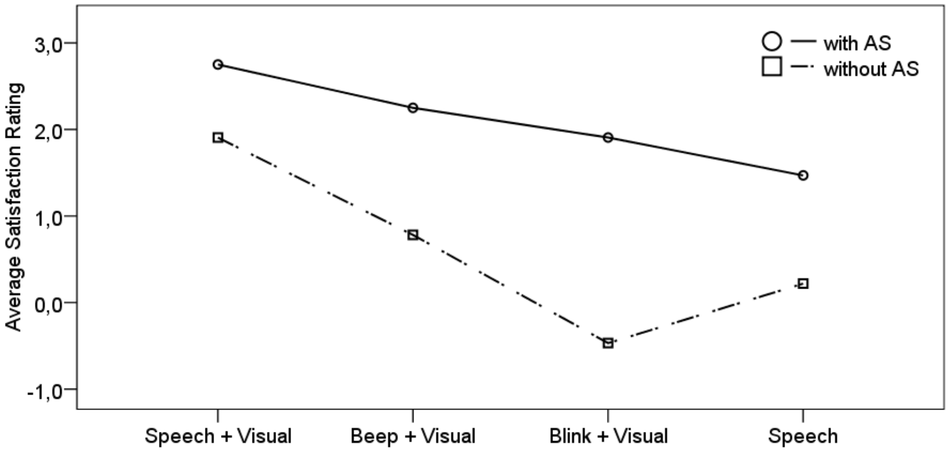


Figure 6.15: Subjective ratings on the satisfaction with warnings in this experiment.

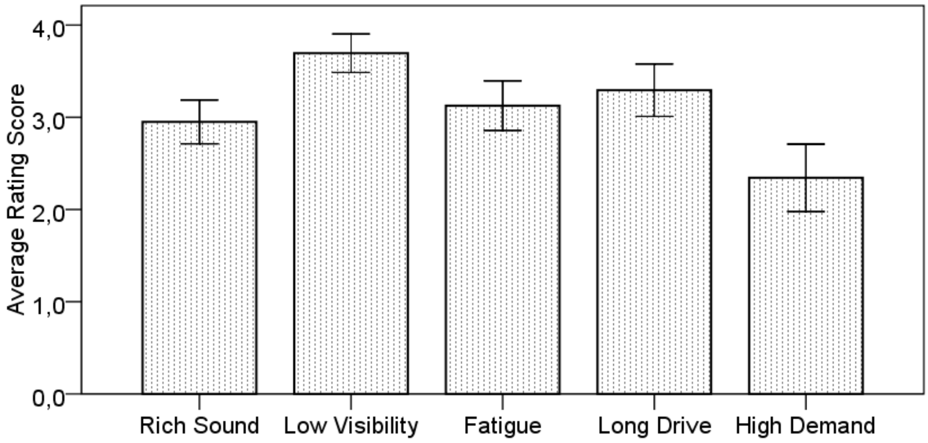


Figure 6.16: Rating of warning usefulness in each situation, averaged over all presentation conditions. Error bars represent standard errors.

purely auditory (speech) and the purely visual (with blinking cues) variants.

Situation dependent assessment of usability

Participants rated the expected usefulness of each warning presentation in five real-life driving situations. A three-way repeated-measure ANOVA was conducted, using modality, level of assistance, and situation as independent factors. Results showed that all three factors had a significant influence on the expected usefulness of warnings (modality: $F(3, 29) = 17.5$, $p < .001$; assistance: $F(1, 31) = 33.4$, $p < .001$; situation: $F(4, 28) = 24.1$, $p < .001$). In addition, there was an interaction between modality and situation ($F(12, 20) = 3.1$, $p < .05$).

Figure 6.16 shows the average rating scores in each situation (averaged over all presentation conditions). Warnings were expected to be the most useful in the 'low visibility' situation ($M = 3.7$), and the least useful in the 'high demand' situation ($M = 2.3$). Post-hoc tests (Bonferroni) further revealed significant differences between 'low visibility' and each of the other four situations. The same was also found for the 'high demand' situation.

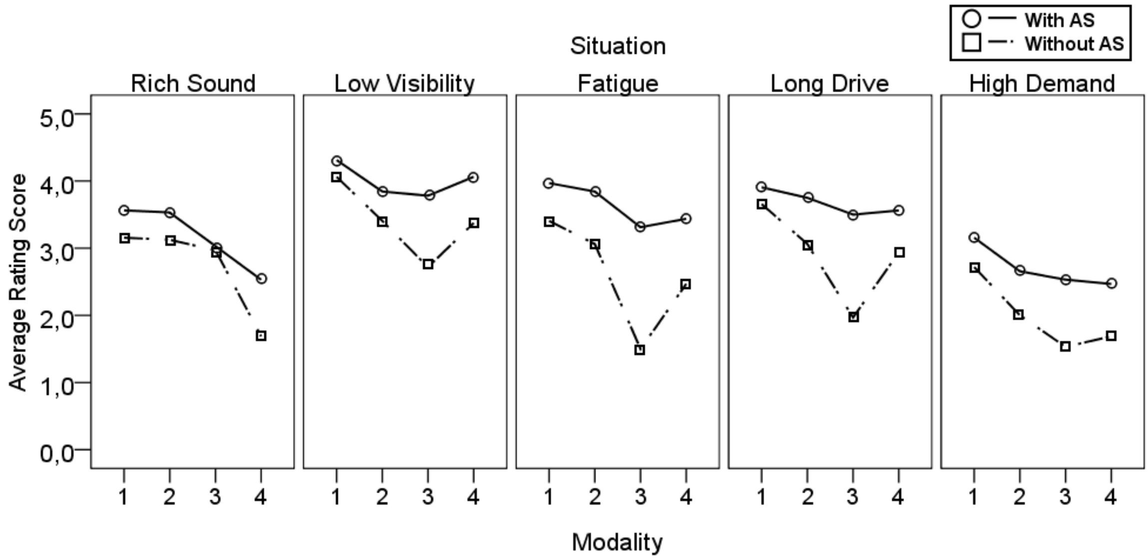


Figure 6.17: Average rating scores of warning usefulness in each driving condition. Modality variant 1: Speech + Visual; 2: Beep cue + Visual; 3: Blink cue + Visual; 4: Speech.

As shown in Figure 6.17, in all situations, the higher level of assistance was expected to be more useful than the lower level. However, the modality factor showed more diverse results across different situations, which accounts for the interaction effect between modality and situation. In the *rich sound* situation, the purely auditory warnings (speech with and without AS) were rated as significantly less useful than the other three modality variants, between which no significant difference was found. Ratings in the *low visibility* and *high demand* situation showed similar results. In both situations, the combined speech and visual warnings were considered significantly the most useful. No significant difference was found between the other three situations. Ratings in the *long drive* and the *fatigue* situations also showed similar results. In both situations, the 'speech + visual' and 'beep + visual' modality variants were rated as significantly more useful than the other two variants. There was no significant difference between the more useful two variants or between the less useful two variants. Furthermore, an interaction effect was found between modality and level of assistance. This is because the benefit of AS was more pronounced when warnings were purely visual ('blink + visual') or purely auditory, compared to when warnings were multimodal.

6.4.7 Discussion

In this section, we discuss the experimental results from 5 aspects, with respect to the research objectives of this study.

Warning vs. no warning

Two measures allowed the comparison between driving (and avoiding obstacles) with and without warnings, and both of them showed an advantage of using warnings. The percentage of obstacles that were not safely avoided was 19.1% without warnings and only 1.4% with warnings, indicating that presenting local danger warnings can significantly increase driving safety. Moreover, DALI ratings showed that warnings significantly reduced the effort of attention, situational stress and visual demand of driving. The fact that participants reported a reduction of visual demand is somewhat surprising, because warnings imposed additional perceptional load, so one would expect visual demand to be higher with warnings. It turned out that with the assistance of warnings, drivers did not need to intensively watch the road to detect emergent obstacles. Although warnings increased the auditory demand and interference, they generally made danger avoidance less effortful and less stressful.

Level of assistance

The level of assistance varied between providing only warnings and providing warnings preceded by AS. The purpose of AS was to assist drivers in deciding how to avoid the danger (obstacles). Several measures consistently showed the benefit of AS. AS reduced unsafe behaviors in all modality conditions, especially when visual warnings were provided (no unsafe behaviors). AS accelerated both braking and lane change reactions by 470 ms and 130 ms respectively. Although the differences were below a half second in this experiment, they could be more pronounced in real driving where local danger warnings are less expected and more diverse. According to DALI ratings, AS significantly reduced the effort of attention, stress, visual demand, temporal demand, and interference in driving. The auditory demand was not reduced by AS due to the fact that AS were delivered via speech and thus induced auditory load themselves. Moreover, drivers were more satisfied with warnings preceded by AS in the simulated driving condition, and predicted AS to be beneficial in various real driving situations.

The benefit of AS found in this study seems to contradict some previous findings which favored a lower level of assistance for drivers. For example, the study presented in [120] investigated headway assistance in different styles. The information style visually presented the headway to a leading vehicle. Information was presented very early so that drivers had sufficient time to analyze it. The command style gave brake commands via speech. Commands were given much later when the headway distance was dangerously small. Drivers in their study accepted the information style more than the command style, indicating that drivers like to make decisions themselves when there is ample time to analyze the situation, which is not surprising. In contrast, drivers in our study were notified of an emergent danger less than 10 s in advance. The AS has more advantage in such an urgent situation because drivers may not have sufficient time to come up with proper reactions themselves.

Our results also suggest that AS are more beneficial when supporting information is provided at the same time, because AS alone do not support drivers to update their SA. In this experiment, AS did not have any effect on the two measures of situation awareness (correct recalls and reactions to irrelevant messages). Although drivers were aware of the fact that AS were always correct, most of them did not totally rely on AS for their reactions. Playing back the video recordings, we saw that when visual messages were provided, most drivers looked at the display during the course of the action suggestion or immediately afterwards, validating the AS with the obstacle location in the visual warnings. This validation process required less time and effort compared to making decisions based on only the warnings, which probably accounted for the reaction time benefit of AS.

Based on these findings, we may conclude that adding AS made local danger warnings more effective, efficient and satisfying. AS can be particularly beneficial in urgent situations where reaction time is critical. It is important to provide supporting information together with AS, because drivers should always have the opportunity to analyze the situation and make decisions for themselves.

Modality

Regarding the four investigated modality variants, the usability assessment suggests that the combination of speech and visual modalities is the most suitable, speech only is the least suitable and the two visual variants with attentional cues lie in between.

Pure speech presentation. Using speech on its own has three major drawbacks when presenting urgent warnings. First, the duration of the speech messages is too long, leaving the drivers only about 2 s to react in our scenario. This drawback accounts for the longest reaction times and the highest number of late reactions. Second, without a repeat function, speech does not allow multiple perception, which results in the worst recall performance. Third, speech is not the best modality to convey spatial information such as object locations. In this study, obstacle location is most relevant to danger avoidance, and as a result, pure speech warnings caused the most incorrect reactions. In line with previous findings [121], our results suggest that speech alone is not adequate for presenting highly urgent warnings.

Purely visual presentation. Besides the lack of salience, our results also suggest that purely visual presentations may reduce the drivers' vigilance for hazards. In the purely visual condition (visual warnings, blinking cue, no AS), drivers reacted the fastest but reacted to the most irrelevant messages. It seems that they were less attentive to the warnings and were less careful with their reactions. Regarding the satisfaction rating, the purely visual condition was the only one which received a minus score on average. Several drivers described this condition as boring. These findings stand in line with the study in [219] where drivers appeared more vigilant when warnings were delivered aurally than visually. Therefore, the presentation of highly urgent warnings needs to include auditory signals.

Combined visual and speech presentation. Although pure speech and purely visual messages both have major drawbacks, the combination of the two significantly improved the usability of the warnings. Their complementary characteristics provided both high salience and freedom of perception. As a result, this modality variant had the best recall performance, the lowest driving load score and the highest satisfaction score. Although the reaction was not the fastest, drivers reacted less to irrelevant messages than in the other three conditions with visual warnings, indicating a better awareness of the situation conveyed by the warnings. As several drivers explained, receiving information via multiple channels makes it easier to pick up the content, because they can selectively focus on one channel that is more compatible with the driving situation at that specific moment. It seems that a redundant use of multiple modalities did bring "the best of both worlds" in this study.

Visual cue vs. auditory cue. Comparing the two kinds of attentional cues, although no significant difference was found in terms of effectiveness and efficiency, the beep cue was clearly preferred by the drivers. The satisfaction ratings showed that drivers were more satisfied with the beep cue than with the blinking cue. In the questionnaire provided at the end of the experiment, drivers were asked to rate the usefulness of the two kinds of cues, on a Likert scale from 0 (not useful) to 5 (very useful). The beep cue received significantly higher scores (M = 3.8, SD = 1.3) than the blinking cue (M = 2.1, SD = 1.8), shown by a Wilcoxon signed-rank test (z = -3.5, $p < .001$). In addition, 15.6% of the drivers mentioned that they had not noticed the blinking top bar at all. This result confirms that visual modalities have a lack of salience, thus are less suitable for warning presentations when used alone. In this experiment, this lack of salience did not harm the driving safety, because the warnings were always well expected, so no warning was missed even if the blinking cues were not detected. However, visual cues can be even less effective in real driving situations, because drivers do not expect warnings as they did in this experiment, and they may also be distracted by other activities, such as radio and conversation.

Interaction between modality and assistance

Interaction effects between modality and level of assistance were found in two measures. According to the measure of unsafe behaviors, AS particularly increased driving safety when the warnings were presented with only speech, suggesting that AS can be more beneficial when drivers had less time to analyze and respond to warnings. Regarding subjective satisfaction, AS changed drivers' preference between the visual warnings with blinking cues and the pure speech warnings. When AS were provided, drivers were more satisfied with the visual warnings with blinking cues, because the visual warnings allowed them to quickly validate the AS and the lack of salience was compensated by AS which was delivered aurally. Without AS, however, the pure speech warnings were rated more satisfying than visual warnings, due to their higher salience. This interaction effect suggested that salience was considered more important than perception efficiency for highly urgent warnings.

Driving situation and interaction with presentation factors

The situation dependent usability rating shows that local danger warnings are considered more useful in some driving situations than in others, regardless of how warnings are presented. Drivers in this study considered local danger warnings the most useful when the visibility was low (e.g. driving at night, fog, heavy rain) and the least useful when driving was highly demanding (e.g. in an unfamiliar city, in heavy traffic).

In all investigated driving situations, drivers expected the higher level of assistance to be more useful than the lower level, suggesting that AS should always be included in the function of local danger warning, independent of the driving situation. However, drivers had different modality preferences in different driving situations, which accounted for the interaction effect between modality and situation. Visual modalities are expected to be highly useful when driving with rich surrounding sounds, because the saliency of auditory modalities degrades in proportion to the level of surrounding sounds [21]. Speech warnings are highly appreciated when the visibility is low or when the traffic condition is highly demanding. In these situations, it is particularly important to closely watch the headway, and delivering warnings via speech is more compatible with a high visual perception load. Moreover, auditory modalities such as speech and beep sounds are expected to be useful when the driver is tired or unconcentrated, or the trip is long and boring. This is because auditory modalities are able to attract attention and enhance vigilance. However, purely auditory warnings are not preferred in these driving situations, suggesting that multimodal warning presentations are the most useful in these driving situations. In fact, although drivers' modality preferences differ with the driving situation, the visual and speech combined warning presentation is always one of the most appreciated modality variants. This finding suggests that multimodal presentation using both visual and auditory modalities could be the most suitable default setting for presenting local danger warnings.

6.5 Conclusions

In this chapter, we presented two studies investigating the presentation of local danger warnings. Study One revealed that visual warnings could be perceived faster than speech warnings conveying the same information contents. Using the current warning design, the perception time of the best visual presentation (iconic in this case) was only 36% of the duration of speech sentences (8∼10 words, about 5 s). Based on the best visual presentation of Study One, Study Two further investigated 4 modality variants and 2 levels of assistance, using a driving task setting. The results suggest that: 1) local danger warnings are beneficial, because they significantly enhanced drivers' performance of danger avoidance and also reduced their stress level and attentional demand. Local danger warnings could be particularly needed by drivers when the visibility is low. 2) Regarding modality, both auditory and visual modalities should be used to present high-priority warnings. Speech warnings should be kept short and more important information items should be located earlier in the sequence. Attention-attracting cues are better presented by auditory modalities (e.g. a beep sound) than by visual modalities. Spatial information (e.g. location) is better presented visually and graphically than orally. 3) It is beneficial to provide action suggestions together with local danger warnings, but unexplained action suggestions should not be given to drivers.

A limitation of this study is that the simulated driving condition is rather ideal – no extra traffic, good weather, and no additional task besides driving and warning perception. As future work, it is necessary to further evaluate the best warning presentations from the current study in more demanding driving conditions, such as in heavy traffic, in a noisy environment, or when drivers are distracted by the radio, a phone call, or a conversation with other passengers.

7 Presenting Informative Interruption Cues

The function of in-vehicle information systems (IVIS) is not limited to hazard warnings such as the ones investigated in the previous chapter. In fact, IVIS have a wide variety of functions. As IVIS are increasingly able to obtain and deliver information, driver distraction becomes a larger concern. In this chapter, we investigate the design and presentation of informative interruption cues (IIC) – a means to support drivers' attention management between multiple tasks. Based on the requirements of pre-attentive reference, two modalities are suited to present IIC in a driving environment: sound and vibration. We designed sound and vibration cues for four different priority levels, and evaluated them in 5 task conditions that simulated the perceptional and cognitive load in real driving situations. The study presented in this chapter has been published as [42].

This chapter is structured as follows. Section 7.1 introduces the background and objective of this study. Section 7.2 describes the design of sound and vibration cues. The experiment set up to evaluate the cues is presented in Section 7.3. The experimental results are presented and discussed in Section 7.4 and 7.5. Finally, Section 7.6 concludes this chapter.

7.1 Background

In-vehicle information systems (IVIS) are primarily intended to assist driving by providing supplementary information to drivers in realtime. IVIS have a wide variety of functions [219], such as route planning, navigation, vehicle monitoring, traffic and weather update, hazard warning, augmented signing and motorist service. The recent Car2X communication technology (see Section 6.1 will allow many more functions to become available in the near future. In addition, when in-car computers have access to wireless internet, IVIS can also assist drivers in tasks that are not driving related, such as email management [112]. Besides being useful, these IVIS functions could be potentially harmful to driving safety, because they impose additional attention demand on the driver and may cause distraction. According to a recent large-scale field study conducted over a period of one year [178], 78%

of traffic collisions and 65% of near collisions were associated with drivers' inattention to the forward headway, and the main source of this inattention was found to be secondary tasks distraction, such as interacting with IVIS. Therefore, the design of IVIS should aim to maximize benefits and minimize distraction. To this end, IVIS need to interrupt drivers in a way that supports them to better manage their attention between multiple tasks.

Supporting users' attention and interruption management has been a design concern of many human-machine systems in complex event-driven domains. One promising method is to support pre-attentive reference[1] by providing informative interruption cues (IIC) [96; 104; 215]. IIC differ from non-informative interruption cues, because they do not only announce the arrival of events, but also (more importantly) present partial information about the nature and significance of the events, in an effort to allow users to decide whether and when to shift their attention. In the study presented in [104], participants performed an air traffic control task, while being interrupted by a series of pulse counting tasks. Pulses could be visual, auditory or tactile. IIC were provided to convey the urgency and modality of interrupting tasks. Results showed that the IIC were highly valued by participants. The modality of an interrupting task was used to decide when to perform it, in order to reduce interference with the concurrent air traffic control task. In another study presented in [96], participants performed continuous arithmetic tasks while monitoring a water control system of a space shuttle. IIC were applied to indicate the domain, importance and duration of water control tasks. The importance of pending tasks was used by participants to decide whether and when to attend to the tasks. These findings have demonstrated that IIC can be used successfully by operators in complex event-driven environments to improve their attention and interruption management.

In the automotive domain, a large number of studies have been carried out on the design and presentation of IVIS messages (see Section 3.6 and 6.2). However, using pre-attentive reference to help drivers selectively attend to these messages has rarely been investigated. As IVIS are increasingly able to obtain and deliver information, we propose that IIC could be an effective means to minimize inappropriate distractions. IIC inform drivers about the arrival of new IVIS messages and their *priority levels* associated with urgency and relevance to driving. The perception and understanding of IIC should require minimum time and attention resources. Upon reception of IIC, drivers can have control over whether to attend to, postpone or ignore the messages, depending on the driving demand at the moment. Some messages can be delivered "on demand", meaning that they are not delivered together with IIC but only when the driver is ready (or willing) to switch attention. In this way, the system supports drivers in managing their attention between multiple tasks. To evaluate this proposal, we first need to obtain a set of IIC that meets the criteria of pre-attentive reference [289] and is suited for the driving environment. Then, we need to investigate whether drivers can indeed make use of these IIC to better manage their attention between driving and IVIS messages. This chapter presents a study that tackles the first step – the design and evaluation of IIC for in-vehicle information presentation.

We designed a set of sound and vibration cues that conveyed four levels of priority, and

[1] Pre-attentive reference is to evaluate attention directing signals with minimum attention [289].

conducted a user study to evaluate the cues. The objective of this study was twofold: 1) to evaluate our design of the cues, including whether they were easy to learn and whether they can be quickly and accurately identified under various types of cognitive load that drivers may encounter while driving; and 2) to compare sound and vibration, aiming to find out which modality was more suitable in which conditions. The remainder of this chapter presents the design of the sound and vibration cues, describes the evaluation method, discusses the results and finally presents our conclusions from the findings.

7.2 Design of Sound and Vibration Cues

7.2.1 Requirements and Aim

To support pre-attentive reference, IIC need to meet the following criteria ([289], pp. 10).

1. IIC can be picked up in parallel with ongoing lines of reasoning and ongoing activities. Since driving is mainly a visual task, auditory or tactile cues can be better perceived in parallel with driving, because the perception of different sensory modalities consumes separate attention resources ([280], see also Chapter 3).
2. IIC provide information on the significance and meaning of an interruption, so that the observer can decide whether the interruption event warrants a shift in attention or not. In our case, IIC convey the priority of IVIS messages, so that drivers can decide whether to ignore, postpone or attend to the messages.
3. IIC allow for evaluation by the observer with minimum time and attention. In our case, this means that regardless of the message priority a cue intends to convey, the cue should always be interpreted quickly and easily.

To meet the above-mentioned criteria, the design of cues should aim to construct intuitive mappings between features of sound/vibration and the associated priority levels. In other words, the defined priority of a cue should match the observer's perceived priority from the cue, so that the signal-priority mappings can be learnt with minimum effort.

Finally, the driving context also raises a specific challenge for IIC. Driving conditions can be very diverse and cues should continue to be effective in all conditions. The attention demand of driving may differ with road, traffic, and weather conditions. The driver may be distracted by radio, music, or conversation with other passengers. Noise in the driving environment may also hinder the detection of cues. Therefore, the effectiveness of our cues needs to be evaluated in various conditions.

7.2.2 Priority Levels

We set up 4 priority levels for all possible IVIS information, numbered from 1 (the highest level) to 4 (the lowest level). Levels 1 and 2 were designed for driving-related information. Level 1 could be assigned to hazard warnings (e.g. the local danger warnings investigated in Chapter 6) and other urgent messages about traffic or vehicle condition. Level 2 covered

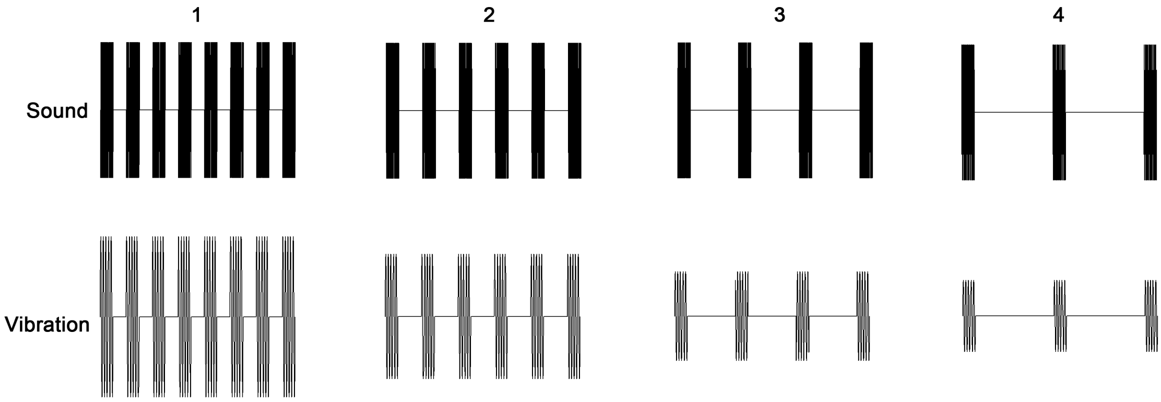


Figure 7.1: Illustration of the design of sound cues and vibration cues. Numbers indicate priority levels. The x and y axes of each oscillogram are time and amplitude. The duration of each cue is 1.5 seconds. The duration of each pulse is 100 milliseconds.

less urgent driving-related information, such as a low fuel level or low air pressure in the tires. Level 3 and 4 were associated with IVIS functions that were not related to driving, such as email, phone calls and in-vehicle infotainment. Level 3 could be customized for relatively important emails and phone calls, and level 4 included all other driving irrelevant information.

7.2.3 Sound Cues

Sounds used in interface design are commonly categorized into two groups – auditory icons and earcons. Auditory icons are environmental, natural sounds that represent actions, processes or objects by similarity ([87], also see Section 3.6). Earcons are abstract, synthetic sounds, musical motives that are used to provide information [24]. For this study, we chose earcons for the following reasons: 1) earcons offer the flexibility to create different sound patterns for expressing different priorities, whereas it is not always easy to find sound equivalents in everyday life for the priorities or events that need to be conveyed; 2) earcons can share common patterns with vibration signals, which allows a better comparison between the two modalities; and 3) several sound parameters are known to influence the perceived urgency of earcons. Previous studies on this topic have commonly suggested that sound signals with higher pitch, higher number of pulses, and faster pace (shorter inter-pulse interval) are generally perceived as more urgent [9; 70; 150; 172].

We did not rely on only one feature to convey priority. To reinforce the effectiveness, we combined pitch, number of beeps and pace, and manipulated them in a systematical manner. The four sound cues are illustrated in the top row of Figure 7.1. From priority 1 to 4, the pitch of sounds was respectively 800 Hz, 600 Hz, 400 Hz and 300 Hz. We kept the pitches in this range because lower pitches were not salient and higher pitches might induce discomfort. All sounds were designed with the same length because duration was not a manipulated feature in this study, and also because this way reaction times to the cues could be better evaluated. Given the fixed duration, the number of pulses and pace were

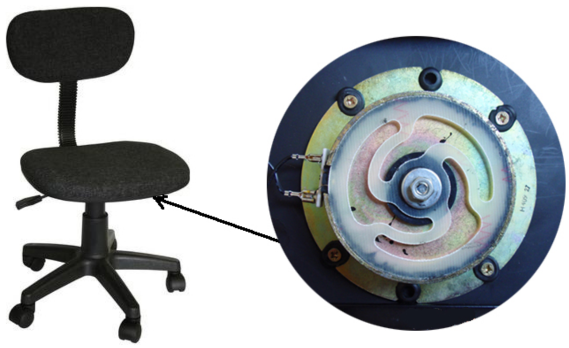


Figure 7.2: The vibration motor mounted beneath the seat of an office chair.

two related features, because more pulses in the same duration leads to faster pace. From priority 1 to 4, the number of pulses was respectively 8, 6, 4 and 3, resulting in decreasing paces.

7.2.4 Vibration Cues

Although sound has been a preferred modality for alerts and interruption cues [282], sound cues can go unnoticed in environments with rich surrounding sounds. In such situations, the tactile modality may be more effective for cue delivery [213; 225]. The best known application of vibration displays at this moment is probably the vibration function on mobile phones, presenting a 1-bit message "your phone is ringing". In automotive studies, tactile modalities such as vibration and force were typically used to present alerts with directions (see Section 3.6). Directions were conveyed by the location where signals were provided. For example, a vibration signal on the left side of the steering wheel warns the driver of lane departure from the left side [247]. However, the potential information transfer capacity of the tactile channel is much larger than 1-bit. For example, people who know Braille can actually read with their fingers. In our study, we intend to convey priority by the pattern of vibration signals. Since most people are unfamiliar with tactile signals in HCI, it is important to make tactile signals intuitive and self-explaining [259]. Several studies have investigated the relation between vibration parameters and perceived priority [29; 96; 105]. Results showed that vibration signals were perceived as more important/urgent when they had higher intensities, a higher number of pulses, and higher pulse frequency (number of pulses per second).

We constructed a vibration chair by mounting a vibration motor beneath the seat of an office chair (Figure 7.2). This location was chosen because the seat is always in full contact with the driver's body. The vibration motor was taken from an Aura Interactor[2]. It is a high force electromagnetic actuator (HFA), which takes a sound stream as input and generates a fast and precise vibration response according to the frequency and amplitude of the sound.

Similar to sound, we also combined three features of vibration: intensity, number of

[2]Aura Interactor. http://apple2.org.za/gswv/a2zine/GS.WorldView/Resources/A2GS.NEW.PRODUCTS/Sound.Interactor.html

pulses, and pace. Four sound input signals were created for the vibration cues (the bottom row of Figure 7.1). They had the same frequency (50 Hz) and length, but different amplitudes that led to different vibration intensities. The intensity for priority 1, 2 and 3 was respectively 2.25, 1.75 and 1.25 times the intensity for priority 4. The number of pulses were also 8, 6, 4 and 3, resulting in decreasing paces from priority 1 to 4.

7.3 Evaluation Method

This evaluation was only focused on the design of the cues. It did not involve IVIS messages or the attention management between driving and the messages. Therefore, we chose to mimic the driving load in a laboratory setting, in order to more precisely control the manipulated variances of task load between conditions. The task set mimicked various types of cognitive load that drivers may encounter during driving. Although this did not exactly resemble a driving situation, it better ensured that all conditions were the same for all participants, leading to more reliable results for the purpose of this evaluation.

7.3.1 Tasks

Visual tracking task

Since driving mostly requires visual attention, a visual tracking task was employed to induce a continuous visual perception load (to keep eyes on the "road"). Participants were asked to follow a moving square with a mouse cursor (the center blue square in Figure 7.3). The size of the square was 50 pixel $\times$ 50 pixel on a 20" monitor with 1400$\times$1050 resolution. The distance between participant and the monitor was about 50 cm. Most of the time, the square moved in a straight line along the x and y axis. At random places, it made turns of 90 degrees. To provide feedback on the tracking performance, the cursor turned into a smiley when it entered the blue square area.

There were 10 tracking trials in total and each of them lasted for 2 minutes. The participants were instructed to maintain the tracking performance throughout the whole experiment, just as drivers should maintain driving on the road. Although this task did not involve vehicle control, it does share common characteristics with driving – the performance could decrease when people paid less attention to watching the "road", and the visual perception demand of the task could vary in different conditions.

Cue identification task

During the course of tracking, cues were delivered with random intervals. Upon the reception of a cue, participants were asked to click on an answer button as soon as they identified the priority level. This task aimed to show how quickly and accurately participants could identify the cues. Four answer buttons were made for this task, one for each priority level (Figure 7.3). Buttons were color-coded to intuitively present different priority levels. From level 1 to 4, the color was red, orange, yellow and white, respectively. A car icon was

Figure 7.3: The tracking target square and the 4 buttons for cue identification response. The numbers were added to indicate priority levels and were not present in the experiment. From priority level 1 to 4, the color of the button was red, orange, yellow and white, respectively.

placed on the buttons of the driving-related levels. There were always 8 cues (2 modalities × 4 priorities) delivered in a randomized order during a 2-minute tracking trial. Note that the four buttons always moved together with the center tracking square. In this way, cue identification imposed minimum influence on the tracking performance.

7.3.2 Task Conditions

We set up 5 task conditions (as summarized in Table 7.1), attempting to mimic 5 possible driving situations.

Table 7.1: Five task conditions applied in the experiment.

Condition Index	1	2	3	4	5
Cue Identification	×	×	×	×	×
Low-Load Tracking	×		×	×	×
High-Load Tracking		×			
Radio			×		
Conversation				×	
Noise					×

Condition 1 (low-load tracking) attempted to mimic an easy driving situation with a low demand on visual perception. The tracking target moved at a speed of 50 pixels per second. During a 2-minute course, the target made 8 turns of 90 degrees, otherwise moving in a straight line. The turning position and direction differed in each course. This tracking task setting was also applied to conditions 3, 4 and 5.

Condition 2 (high-load tracking) attempted to mimic a difficult driving situation where the visual attention was heavily taxed, such as driving in heavy traffic. The tracking target moved at a speed of 200 pixels per second. During a 2-minute course, the target made 32 turns of 90 degrees, otherwise moving in a straight line. The turning position and direction differed in each course.

Condition 3 (radio) attempted to mimic driving while listening to the radio. Two recorded segments of a radio program were played in this condition. Both segments contained a conversation between a male host and a female guest about one kind of sport (marathon and tree climbing). Participants were instructed to pay attention to the conversation while tracking the target and identifying cues. They were also informed about receiving questions regarding the content of the conversation later on.

Condition 4 (conversation) attempted to mimic driving while talking with other passengers. Four topics were raised during a 2-minute trial via a text-to speech engine. They were all casual topics, such as vacation, favorite food/drink, music, movie, weather and the like. Participants had about 25 seconds to talk about each topic. They were instructed to keep talking until the next topic was raised and generally succeeded in doing so.

Condition 5 (noise) attempted to mimic driving in a noisy condition or on a rough road surface. For auditory noise, we played recorded sound of driving on the highway or on a rough surface. The signal to noise ratio was approximately 1:1. Tactile noise was generated by sending pink noise[3] into the vibration motor. The tactile noise closely resembled the bumpy feeling when driving on a bad road surface, which was verified with a pilot study using 3 subjects. The signal to noise ratios for priority 1 to 4 were approximately 3:1, 2:1, 1:1 and 0.6:1. Both auditory and tactile noise were always present in this condition.

7.3.3 Subjects and Procedure

Thirty participants, 15 male and 15 female, took part in the experiment. Their ages ranged between 19 to 57 years old ($M = 31.6$, $SD = 9.6$). None of them reported any problem with hearing, color perception, or tactile perception.

The experiment consisted of two sessions: a cue learning session and a task session. After receiving a short introduction, participants started off with learning the sound and vibration cues. They could click on 8 buttons to play the sounds or trigger the vibrations, in any order they wanted and as many times as needed. Learning ended when participants felt confident in being able to identify each cue when presented separately from the others. Then, a test was carried out to assess how well they had managed to learn the cues. Feedback on performance was given to reinforce learning. At the end of this session, participants filled in a questionnaire about their learning experience. In the task session, each participant performed 10 task trials (2 modalities $\times$ 5 conditions) of 2 minutes each. The trial order was counterbalanced. Feedback on cue identification performance was no longer given. During the short break between two trials, participants filled in questionnaires assessing the task load and the two modalities in this previous trial. At the end of the experiment, participants filled in a final questionnaire reporting physical comfort of the signals and the use of features.

[3]Pink noise is a signal with a frequency spectrum such that the power spectral density is inversely proportional to the frequency.

Table 7.2: Summary of measures. P: performance measures; S: subjective measures.

Session	Measures
Cue Learning	Amount of learning (P) Cue identification accuracy (P) Recognition of variances in features (S) Association of features with priorities (S)
Task	Tracking error (P) Cue detection failure (P) Cue identification accuracy (P) Response time (P) Ease of cue identification (S) Modality preference (S) Physical comfort (S) Features used (S)

7.3.4 Measures

A total of 12 measures were employed, 4 for the cue learning session and 8 for the task session. They are summarized in Table 7.2. The questionnaires used in this study can be found in Appendix D.

Cue learning session

For this session, two performance measures were used: 1) *Amount of learning*: the number of times participants played the sounds and triggered the vibrations before they reported to be ready for the cue identification test. 2) *Cue identification accuracy*: the accuracy of cue identification in the test.

Two subjective measures were derived from the questionnaire on learning experience: 1) *Recognition of variances in features*: how easy it was to recognize the variances in each feature (e.g. the three different sound pitches). Participants rated the 6 features separately on a Likert scale from 1 (very easy) to 10 (very difficult). 2) *Association of features with priorities*: how easy/intuitive it was to infer priorities from each feature. Participants rated the 6 features separately on the same Likert scale as mentioned before. Although the pace and the number of pulses were two dependent features in our setting, we still introduced and evaluated both of them, because they were different from a perception standpoint. Drivers may find one of them more effective and reliable than the other.

Task session

For this session, four performance measures were applied: 1) *Tracking error*: distance between the position of the cursor and the center of the target square (in pixels). Instead of taking the whole course, average values were only calculated from the onset of each cue to the associated button-click response. In this way, this measure better reflected how much cue identification interfered with tracking. 2) *Cue detection failure*: the number of times a cue was not detected. 3) *Cue identification accuracy*: the accuracy of cue identification in each task trial. 4) *Response time*: the time interval between the onset of a cue to the moment of the button-click response.

Four subjective measures were derived from the between-trial questionnaire (1 and 2) and the final questionnaire (3 and 4). 1) *Ease of cue identification*: how easy it was to identify cues in each task trial. Participants rated each task trial on a Likert scale from 1 (very easy) to 10 (very difficult). 2) *Modality preference*: which modality was preferred for each task condition. Participants could choose between sound, vibration and either one (equally prefer both). 3) *Physical comfort*: how physically comfortable the sound and vibration signals made them feel. Participants rated the two modalities separately on a Likert scale from 1 (very comfortable) to 10 (very annoying). 4) *Features used*: which features of sound and vibration participants relied on to identify the cues. Multiple features could be chosen.

7.4 Results

7.4.1 Cue Learning Session

Amount of learning

The number of times participants played the cues ranged from 12 to 32. On average, cues were played 18.7 times, 9.4 times for sounds and 9.7 times for vibrations. A 2 (modality) $\times$ 4 (urgency level) repeated measures analysis of variance (ANOVA) showed a significant priority effect, $F(3, 27) = 20.0$, $p < .001$. Participants spent more learning effort on level 2 and 3 (M = 10.9 times) than on level 1 and 4 (M = 7.8 times), $t\,(29) = 7.1$, $p < .001$.

Cue identification accuracy

Participants showed high performances in the cue identification test. Twenty-five participants did not make any error in the 16 tasks. The other 5 participants made no more than 3 errors each, mostly at priority level 2 and 3. On average, the identification accuracy reached 97.8% for sound cues, 99.2% for vibration cues and 98.5% overall.

Recognition of variations in features

On the scale from 1 to 10, the average rating scores of the 6 features all fell below 4.2 (Figure 7.4), indicating that variations in these features were reasonably recognizable. Participants found it the easiest to recognize different paces in both sound and vibration. The difference in vibration intensities was found the least distinguishable, which might be due to the fact that vibration intensity was manipulated within a small range in order to avoid

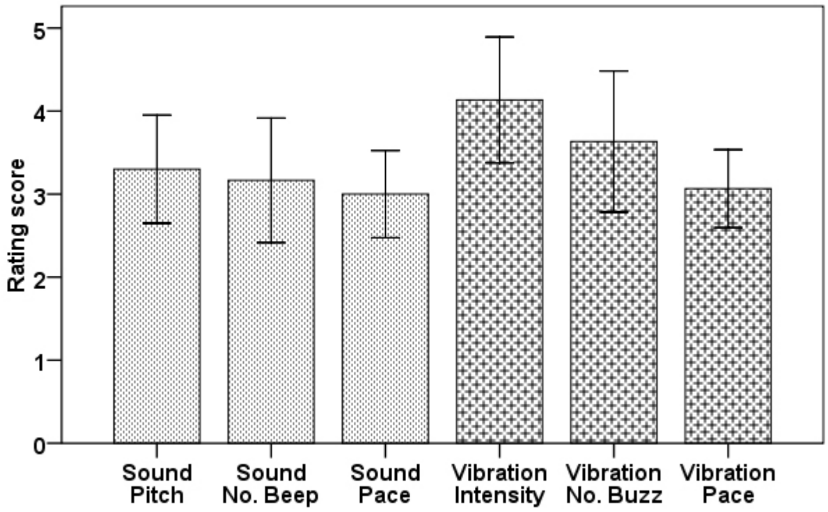


Figure 7.4: Rating scores on the difficulty of recognizing variations in each feature. Error bars represent standard errors. (1 = easiest, 10 = most difficult)

physical discomfort. A Wilcoxon signed-rank test showed that participants found the vibration intensity ($M = 4.1$) significantly more difficult to recognize than the other 5 features ($M = 2.3$), $z = -2.6$, $p < .05$. No significant differences were found between the other 5 features.

Association of features with priorities

The average rating scores of the 6 features all fell below 3.5 on the 10-level scale (Figure 7.5), indicating that participants found it fairly easy and intuitive to infer priorities from these features. Comparing the two modalities, a Wilcoxon signed-rank test showed that rating scores were significantly lower for sound features ($M = 2.6$) than for vibration features ($M = 3.2$), $z = -2.4$, $p < .05$. For both sound and vibration, pace was rated as the most intuitive feature (sound: $M = 2.2$; vibration: $M = 2.7$). Pitch of sound was also fairly intuitive ($M = 2.3$). Number of pulses was rated as significantly less intuitive than pace, shown by a Wilcoxon signed-rank test, $z = -2.5$, $p < .05$).

7.4.2 Task Session

Tracking error

Average tracking errors in each condition are shown in Figure 7.6 (left). A three-way repeated-measure ANOVA showed that modality, condition and priority level all had a significant influence on the tracking performance (modality: $F(1, 29) = 10.6$, $p < .01$; condition: $F(4, 26) = 81.5$, $p < .001$; priority: $F(3, 27) = 34.4$, $p < .001$). Post-hoc tests showed that tracking error was significantly lower when cues were delivered by vibration than by sound. Comparing the 5 conditions, tracking error was significantly the highest in the high-load tracking condition. When tracking load was low, the conversation condition induced

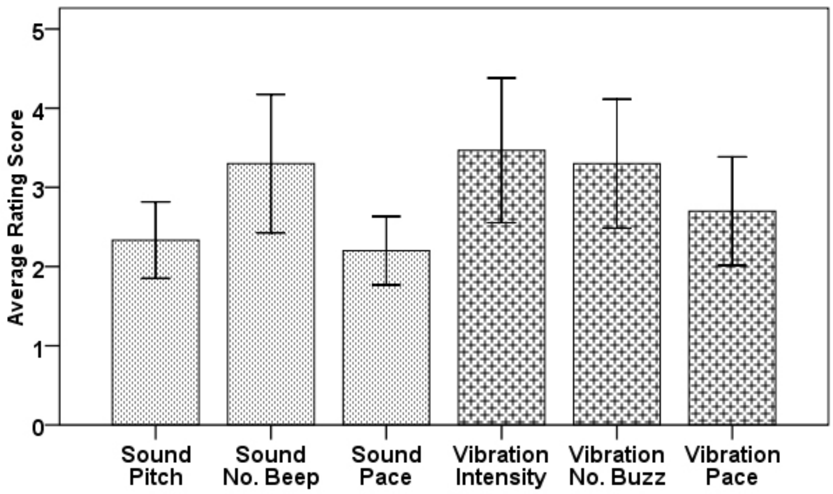


Figure 7.5: Rating scores on the difficulty of associating features with priorities. Error bars represent standard errors. (1 = easiest, 10 = most difficult)

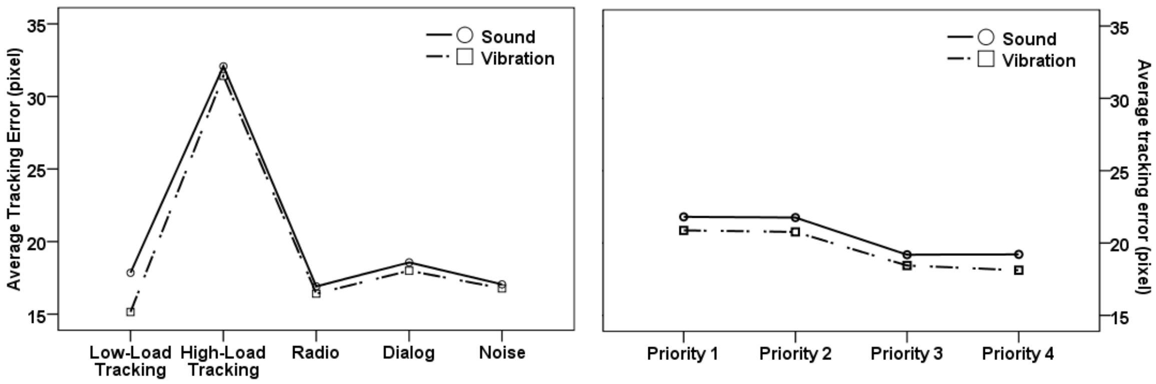


Figure 7.6: Average tracking error by task conditions (left) and by priority levels (right).

significantly higher tracking error than the other 3, between which no difference was found. Among the 4 priority levels, tracking was significantly more disturbed by cues at the higher two priority levels than by cues at the lower two priority levels (Figure 7.6, right). This result makes sense because the cues at higher priority levels were more salient and intense; therefore they were more able to drag attention away from tracking during their presentation. Finally, there was also an interaction effect between modality and condition ($F(4, 26) = 5.9$, $p < .01$), indicating that vibration was particularly beneficial in the low-load tracking condition.

Cue detection failure

Cue detection failure occurred 6 times, which was 0.25% of the total number of cues delivered to all participants in the task session ($30 \times 10 \times 8 = 2400$). One failure occurred in the conversation condition. The missed cue was a vibration cue at priority level 2. All the other 5 failures occurred in the noise condition. The missed cues were all vibration cues at the

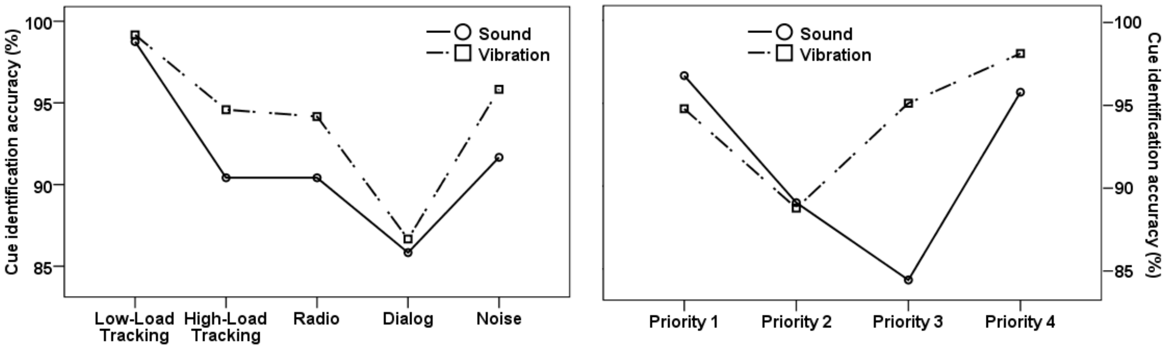


Figure 7.7: Cue identification accuracy by task conditions (left) and by priority levels (right).

lowest priority level. However, considering the fact that the signal-to-noise ratio for level-4 vibration was below 1, only 5 misses (8.3%) is still a reasonably good performance. It was probably the regular rhythm of cues that made them still distinguishable from the noise.

Cue identification accuracy

The average accuracy over all task trials was 92.5% (Figure 7.7, left), which was lower than the performance in the learning test when cue identification was the only task to perform. A three-way repeated-measure ANOVA revealed that modality, task condition and priority level all had a significant influence on the cue identification accuracy (modality: $F(1, 29) = 5.5, p < .05$; condition: $F(4, 26) = 11.7, p < .001$; priority: $F(3, 27) = 9.3, p < .001$). Post-hoc tests showed that identifying vibration cues was significantly more accurate than sound cues, t (29) = -2.1, $p < .05$. The low-load tracking task did not hamper cue classification, because the accuracy in this condition (M = 99.0%) was comparable with the learning session where cue deification was the only task to perform (M = 98.5%), t (29) = 0.5, *n.s.* However, the classification accuracy was significantly higher in the low-load tracking condition than the other 4 conditions (M = 90.9%), t (29) = 6.9, $p < .001$. All types of additional load decreased cue classification accuracy, among which dialog hampered accuracy the most (12.7% less on average).

Comparing priority levels, levels 1 and 4 were identified more accurately than levels 2 and 3 (Figure 7.7, right), presumably because their features were more distinguishable. There was also a significant interaction between modality and priority, $F(3, 27) = 8.1$, $p < .001$. It seems that the advantage of vibration over sound was much larger at priority level 3 than at the other 3 levels. We further analyzed the error distribution over the four priority levels. It turned out that errors only occurred between two successive levels, such as between level 1-and-2, 2-and-3, and 3-and-4.

Response time

All identification responses were given within 5 seconds from the onset of cues (*min* = 1.6 s, *max* = 4.5 s, M = 2.8 s). A three-way repeated-measure ANOVA again showed that modality, task condition and priority level all had a significant influence on this measure

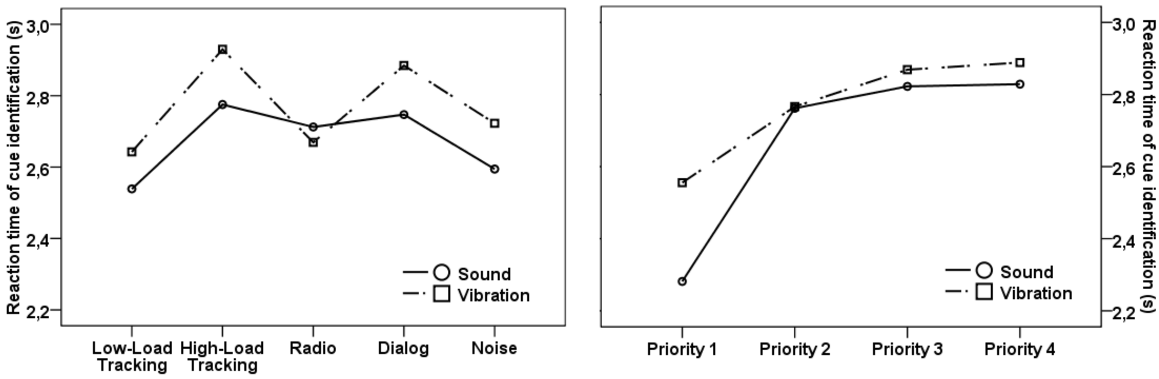


Figure 7.8: Response time by task conditions (left) and by priority levels (right).

(modality: $F(1, 29) = 5.9$, $p < .05$; condition: $F(4, 26) = 10.6$, $p < .001$; priority: $F(3, 27) = 36.9$, $p < .001$). Post-hoc tests showed that identifying sound cues was significantly faster than identifying vibration cues, $t\ (29) = 2.4$, $p < .05$. However, one exception was in the radio condition (see Figure 7.8, left), causing an interaction effect between modality and task conditions ($F(4, 26) = 4.1$, $p < .01$). Comparing conditions, response was the fastest in the low-load condition and the second fastest in the noise condition. No significant difference was found between the other 3 conditions.

Regarding priority levels, level-1 cues were identified significantly faster than cues at the other 3 levels. There could be two reasons for this. First, level-1 cues were the most salient and intense, thus were the easiest to identify. Second, level-1 cues gave a hint of being highly urgent, which in turn made participants speed up their response. As shown in Figure 7.8 (right), there was a general trend for "higher priority, faster response", suggesting that the defined priorities indeed matched participants' perception of priority/urgency from the cues. The signal-priority mappings were intuitive associations, but not arbitrary rules. Moreover, there was also an interaction effect between modality and priority, $F(3, 27) = 5.9$, $p < .01$. The difference in response time between the two modalities was much larger at priority level 1 than at the other 3 levels, indicating that the level-1 sound cues were particularly effective at triggering fast reactions.

Ease of cue identification

Average rating scores for each condition are shown in Figure 7.9. ANOVA showed that rating scores were significantly influenced by task condition ($F(4, 26) = 35.3$, $p < .001$). Not surprisingly, cue identification was judged significantly the easiest in the low-load tracking condition. In contrast, identifying cues while having a conversation was judged significantly the most difficult, which was in line with the lowest accuracy in this condition (Figure 7.7, left). A significant difference was also found between the radio and the high-load condition, with the radio condition being more difficult. The modality factor did not show an effect on this measure ($F(1, 29) = 3.4$, $n.s.$). In terms of average ratings, participants considered vibration cues easier to identify in the radio condition, and sound cues easier to identify in the other 4 conditions.

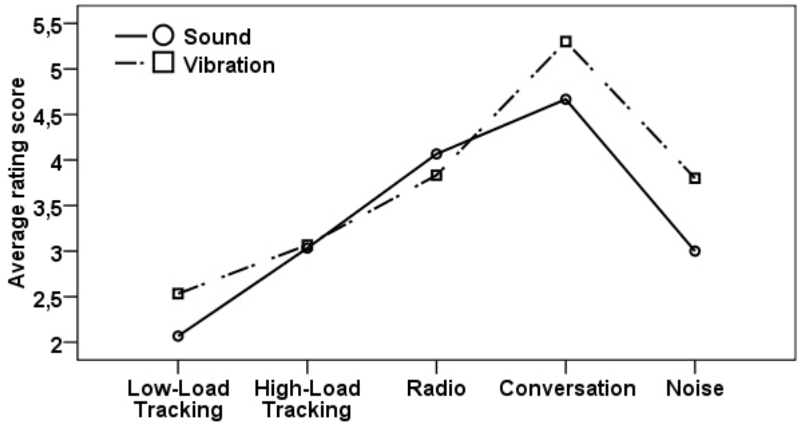


Figure 7.9: Subjective ratings on the difficulty of cue identification in each task trial. (1 = easiest, 10 = most difficult)

Modality preference

Table 7.3 summarizes the number of participants who preferred sound or vibration or either of them in each task condition. In the radio condition, more participants preferred vibration to sound. However, sound got more "votes" than vibration in the other 4 conditions.

Table 7.3: Number of participants who preferred sound or vibration or either of them in each task condition.

	Sound	Vibration	Either one
Low-Load Tracking	19	4	7
High-Load Tracking	15	7	8
Radio	13	15	2
Conversation	12	11	7
Noise	19	9	2

Physical comfort

The average scores of sound and vibration both fell below 4 on the 10-level scale (sound: M = 2.9, SD = 1.5; vibration: M = 3.8, SD = 1.8). A Wilcoxon signed-rank test revealed that participants found sound cues significantly more comfortable than vibration cues, z = -2.0, $p < .05$.

Features used

Participants made multiple choices on the feature(s) they relied on to identify the cues. As Table 7.4 shows, a majority of participants (90%) made use of more than one sound feature and more than one vibration feature. This result suggests that combining multiple features in a systematic way was useful. This design choice also provided room for each participant to selectively make use of those features that he/she found most intuitive and reliable.

Table 7.4: The number of features used to identify cues.

	Sound			Vibration		
No. of features used	1	2	3	1	2	3
No. of participants	3	20	7	3	22	5

Table 7.5 further shows how many participants used each feature. The number of pulses was used by the highest number of participants (26, 86.7%) for both sound and vibration. Pace was also used by a majority of participants; however the pace of vibration was more frequently used than the pace of sound. Less than half of the participants made use of the intensity of vibration, which was associated with the fact that they found it relatively difficult to distinguish the variances in variation intensity (see Figure 7.4).

Table 7.5: Number of participants who used each feature to identify cues.

Sound			Vibration		
Pitch	No. of beeps	Pace	Intensity	No. of pulses	Pace
19	26	17	14	26	22

7.5 Discussion

With respect to our research objectives, the results are discussed from two aspects: the effectiveness of cue design and the modality choice between sound and vibration.

7.5.1 Effectiveness of Cue Design

Various measures have consistently showed that our design of cues can be considered effective. They have also indicated directions in which further improvements could be achieved.

First, in the learning session, all participants spent less than 5 minutes on practising with the cues, before they felt confident in being able to tell them apart from each other. In the cue identification test afterwards, accuracy reached 97.8% for sound cues and 99.2% for vibration cues. These results clearly show that the cues were very easy to learn. Participants also found it fairly easy and intuitive to infer priorities from the 6 features.

Second, cues were successfully detected in 99.8% of cases. Due to a low signal-to-noise ratio ($<$1), the detection of the level-4 vibration cue was affected by the tactile noise that mimicked the bumpy feeling when driving on a bad road surface. This is probably also due to the fact that both signal and noise were provided from the chair. Signal detection in a bumpy condition can be improved by providing vibrations to other parts of the body which are not in a direct contact with the car, such as the abdomen / the seat belt.

Third, cues were identified quickly and accurately. All cues were identified within 5 seconds from the onset. The average response time was 2.8 seconds (1.3 seconds after offset). The higher the priority level, the faster the response, suggesting that participants indeed perceived the associated levels of priority from the cues. The average identification accuracy over all task trials reached 92.5%. Compared to the learning test where the only task was to identify cues, the accuracy was not decreased by the low-load tracking task alone (99.0%). However, all types of additional load harmed the accuracy to some extent. Having a conversation had the largest impact, resulting in the lowest accuracy (86.3%). This result suggests that cognitive distractions induced by talking to passengers or making phone calls are the most harmful to the effectiveness of cues.

Fourth, combining multiple features in a systematic way was found to be useful, because 90% of the participants relied on more than one feature of sound and vibration to identify cues, indicating a synergy between the combined features. This design also provided room for each participant to selectively make use of the features that he/she found more intuitive and reliable. Pace was rated as the most intuitive to convey priorities. However, number of pulses was used by the largest number of participants (86.7%). The reason might be that number of pulses is a clearly quantified feature, and thus is more reliable when the user is occupied by other concurrent activities. However, care should be taken for this interpretation, because due to the synergy between features, participants might not be able to exactly distinguish which feature(s) they had used.

Finally, the distribution of cue identification errors over four priority levels showed that more errors occurred at levels 2 and 3 than at levels 1 and 4. Moreover, all errors occurred between two successive levels, such as 1-2, 2-3 and 3-4. To further improve the effectiveness of cues, the features of sound and vibration need to have greater contrast between two successive levels. In the current design, the contrast between any two disconnected levels (e.g. 1-3, 1-4, 2-4) can be a good reference for creating highly distinguishable cues.

7.5.2 Sound vs. Vibration

The comparison between sound cues and vibration cues is not clear-cut. On the one hand, vibration interfered less with the on-going visual tracking task than sound cues. Vibration cues were identified more accurately than sound cues in all task conditions and at all priority levels. The advantage of vibration over sound was particularly notable in the radio condition, where it was shown by all measures (though not always significantly). These findings show that vibration is certainly a promising modality for delivering IIC. On the other hand, advantages of sound over vibration were also found in several measures. Sound cues were identified faster in all task conditions except one (the radio condition). The response to the level-1 sound cue was particularly fast. Sound was also reported as easier to distinguish and more physically comfortable, and was preferred by a larger number of participants.

The advantages of sound found in this study might be related to the fact that sound has been a commonly used modality to deliver alerts, notifications and cues. People are very used to all kinds of sound signals in the living environment and are trained to interpret them. For example, a fast paced and high pitched sound naturally reminds people of danger alarms. In contrast, the tactile modality is still relatively new to human-machine interaction, so people have relatively less experience in distinguishing and interpreting the patterns in tactile signals. This might explain why participants in this experiment spent more time and (reported) effort on identifying vibration cues, though they performed more accurately. Our result supports the previously proposed guideline, suggesting that if response time is critical, tactile modalities should not be used alone, because then the user will take longer to perceive and attend to information [238].

Overall, our findings suggest that both sound and vibration are useful alternatives to deliver IIC in IVIS. When to use which modality is a situation dependent choice. There might also be some situations where it is more suitable to use both modalities. In this case, signals from the two channels need to be well synchronized in order not to cause confusion. We make the following tentative suggestions. However, they need to be further validated in a driving task setting.

Vibration might be more suitable than sound when:

1. The driver is listening to radio programs while driving.
2. Sound in the car (e.g. music) or surrounding noise is loud.
3. The message to be presented is not highly urgent.
4. The driver's hearing is impaired.
5. There are other passengers in the car. (Vibration signals are private to the driver).

Sound might be more suitable than vibration when:

1. The message to be presented is highly urgent.
2. The road is bumpy.

Using both modalities might be beneficial when:

1. The driver is actively involved in a conversation. (The effectiveness of a single modality could be significantly decreased by cognitive distractions associated with conversation. Multisensory cues may have a "genuinely automatic" ability to capture a persons attention, no matter what he/she happens to be doing at the same time [232].)
2. The message to be presented is highly urgent. (Using multiple modalities redundantly can bring a reaction time advantage, meaning that reaction to multimodal stimuli is faster than to unimodal stimuli [238]. In this specific case, using sound and vibration may allow both a fast response and an accurate priority identification.)

7.6 Conclusions

We designed a set of sound and vibration cues to convey 4 different priorities (of IVIS messages) and evaluated them in 5 task conditions. Experimental results showed that the cues were effective, as they could be quickly learned (< 5 minuets), reliably detected (99.5%), quickly identified (2.8 seconds after onset and 1.3 seconds after offset) and accurately identified (92.5% over all task conditions). Vibration was found to be a promising alternative for sound to deliver informative cues, as vibration cues were identified more accurately and interfered less with the ongoing visual task. Sound cues also had advantages over vibration cues in terms of shorter response time and more (reported) physical comfort. The current design of cues seems to meet the criteria of pre-attentive reference and is effective under various types of cognitive load that drivers may encounter during driving. Therefore, it is a promising (but by no means the only or the best) solution to convey the priority of IVIS messages for supporting drivers' attention management.

Real driving environments can be more complex and dynamic than the ones investigated in this study. For example, drivers may have radio, conversation, and noise all at the same time. We predict that cue identification performance will decrease when driving load increases or more types of load add up. To make the cues more effective in high-load conditions, the features of sound and vibration need to have greater contrast between different priority levels. The contrast between two disconnected levels in the current design can be a good reference. Future work is to apply the cues to a driving task setting in order to further evaluate their effectiveness and investigate their influence on drivers' attention management.

Part IV

Reflection

8 Discussion

"Focus on the journey, not the destination.
Joy is found not in finishing an activity but in doing it."
Greg Anderson

This chapter provides a general discussion on the findings of all our studies presented in the previous four chapters. Section 8.1 discusses the findings of the modality factor from two perspectives: 1) the matching between modalities and information types, and 2) modality combinations. Section 8.2 discusses other presentation and non-presentation factors that were investigated in our studies, with a focus on their interactions with modality. Section 8.3 discusses the effectiveness of the three categories of measures and the consistency between them. In order to easily refer to the studies, we name them as follows:

- S1: information presentation for time limited visual search (Chapter 4)
- S2: information presentation for time limited decision making (Chapter 5)
- S3.1: the first study on presenting local danger warnings (Section 6.3)
- S3.2: the second study on presenting local danger warnings (Section 6.4)
- S4: presenting informative interruption cues (Chapter 7)

8.1 Modality

In all our studies, the modality of information presentation has had significant effects on the performance and mental workload of the tasks. Modality mostly influences the perception stage of information presentation – the stage when subsets of incoming sensory information were selected by attention and interpreted (see Section 3.1). Different modalities differ in their ability to attract attention as well as in their expressive power, which determines which types of information a modality is suitable to present. In addition, the cognitive properties of modality influence the compatibility of multimodal combinations. In multi-task environments, the cognitive properties of modality also influence how well information processing can be time-shared with other concurrent tasks. In this section, we summarize our findings on information type and modality combination.

8.1.1 Information Type

Five types of information were presented in our studies, namely objects, locations (of objects), distances (of objects), severity/priority levels, and attention attracting cues. In S3.1 and S3.2, the distance of objects was only presented by numbers and units, due to its numerical nature. Besides distance, the other four types of information were presented using different modalities. In line with previous findings (see Table 2.2), all our studies showed that a certain type of information was better presented by certain modalities than by others, suggesting the importance of matching information type to the expressive power of modalities.

It is worth mentioning that how well a modality presents a certain information item can also be influenced by the specific realization of this presentation. For example, when using images to present "death", a image of a skull is more effective than a sad face. This difference is not caused by modality properties but by the specific design choices. In all our studies, we consciously aimed at creating good designs for all modality alternatives. Image presentations were made intuitive. Text had readable font and size. In S3.1, S3.2 and S4, the specific design choices were made based on literature studies and pilot user evaluations. This way we tried to minimize the influence of bad designs on our results, thus allowing the effectiveness of different modalities to be better compared.

Object

Three studies investigated the presentation of concrete objects, such as crisis victims (S1), categories of injury (S2), and road obstacles (S3.1). The modality varied between image and text. The results consistently showed that image, as a nonverbal and analogue modality, is particularly suitable for presenting concrete concepts. The image conditions always had a RT (reaction time) advantage over the text conditions, showing that concrete objects were perceived significantly faster when presented by image than by text. When the task was under time pressure, this RT advantage was obviously beneficial. In addition, in all three studies, a strong majority of participants preferred image conditions to text conditions, because they found the tasks presented by images less cognitively demanding.

The advantage of image for presenting concrete objects and concepts has been previously proposed [21; 98; 121; 238], and our result confirms and further reinforces this solid design guideline. In general, graphical modalities (graphs and images) are more vivid than textual information (text and numbers), thus are likely to receive greater weight during decision making processes. In particular, shapes and colors have great salience to human information processors due to the sharp contrast they are able to create [145].

Location

Two studies involved presenting the location of objects. In S1, the location of crisis victims was conveyed by showing the representation of victims (in image or text) on a map background. In S3.1, we compared image and text for presenting the location of road obstacles. Similar to S1, the image presentation locates an icon of an obstacle on a road background. The text presentation was for example "on the right/left lane" or "on the right/left roadside". The results of both studies showed that the image presentation required shorter perception time than the text presentation.

In combination with the results of 'object', it seems reasonable to suggest that visual, nonverbal, analogue modalities (e.g. image, icon, photos) are suitable for presenting concrete information items – those that can be visualized in their exact form or closely mimicked. This may also be extended to auditory modalities, suggesting that auditory, analogue modalities (e.g. auditory icons) are suitable for presenting concrete auditory information (not tested in this dissertation). For example, the best way to demonstrate what dog barking sounds like is obviously to play the exact sound of a dog barking.

Severity/Priority Level

Two studies investigated the presentation of severity/priority levels, using visual (S2), auditory (S4) and tactile (S4) modalities. In S2, we presented four levels of injury severity by either colors ('red', 'orange', 'yellow', 'green') or text ('severe', 'moderate', 'mild', 'none'). The results showed that the colors were more intuitive than text, and they allowed the severity levels to be perceived and compared faster. This result confirms previous findings stating that color is an effective means to indicate state and communicate qualitative/quantitative differences [200]. In addition, the set of color-severity mappings we used is supposed to be particularly intuitive because it is also being used in several real-world applications (see Section 5.2.2). However, the meaning of colors might be culture-dependent. We had one participant from an Asian culture. He was the only one who found it difficult to associate the color 'red' with severe injuries, because red normally symbolizes love and passion in his culture.

In S4, we designed a set of sound and vibration signals (informative cues) to convey the priority level of in-vehicle messages. Three sound features (pitch, number of pulses, pace) and three vibration features (intensity, number of pulses, pace) were manipulated in a unified manner to present four different levels of priority. Generally speaking, both sound and vibration signals were effective in conveying priority levels, because all signals were quickly learned (< 5 min) and accurately identified in all task conditions (92.5%). The comparison between sound and vibration modalities was not clear-cut. Vibration signals were identified more accurately than sound signals in all task conditions and at all priority levels. However, sound cues were identified faster in almost all task conditions, and preferred by a larger number of participants. This might be due to the fact that people are more familiar with sound interfaces than vibration interfaces, therefore they are more trained to interpret the patterns of sound than the patterns of vibration.

These two studies have shown that color, sound (earcon) and vibration can all effectively convey levels of severity/priority. Then, the choice between them depends on other factors, such as the required level of salience and the interference with other concurrent tasks. For example, color would not be suitable to convey the highest level of priority in a driving environment, because it is not salient enough and it may compete with driving for visual perceptional resources. Several tentative suggestions on the selection between sound and vibration have been provided in Section 7.5.

Cues for attracting attention

Two types of cues were investigated in our studies, non-informative and informative cues. Non-informative cues only aim to attract attention. Informative cues not only attract attention, but also convey the nature of the messages/events that are "cued". The informative cues investigated in S4 have been discussed above with a focus on conveying priority levels. Here, we focus on the aspect of attention attraction.

In S3.2, visual cues (blinking top bar) and auditory cues (sound beeps) were used to attract drivers' attention to local danger warning messages. Although the type of cue did not influence task performance or task load, drivers were significantly more satisfied with the sound cues than with the visual cues. They also expected the sound cues to be more useful in real-life driving which could be more dynamic and demanding than in our study. This is because sound, as an auditory modality, is highly able to attract attention (see Section 3.3). In contrast, visual modalities have a lack of salience, and thus are less suitable to be used as attention attracting cues.

In S4, we investigated informative sound and vibration cues for in-vehicle messages. These two modalities are both highly salient. As a result, the cues were reliably detected (mean = 99.5%). They could successfully attract attention even when the participants were (cognitively) occupied by multiple tasks, such as tracking a visual object, listening to the radio, or having a conversation. The sound cues were identified faster than the vibration cues. The sound for the highest priority, which closely resembled an alarm, was particularly able to trigger fast reactions. This seems to suggest that sound cues are more salient than vibration cues, but it is not necessarily the case. The reaction time measured in S4 includes the time associated with attention switching and the time needed to interpret the meaning of the cues. Therefore, the faster reaction to sound might be because that people are quicker at understanding sound patterns. In conclusion, our results suggest that auditory and tactile modalities are most suitable for presenting attention attracting cues.

8.1.2 Modality Combination

Two studies investigated modality combinations, including compensatory combination (the combined modalities convey different information) and redundant combination (the combined modalities convey the same information).

Compensatory combination

In S1, text was combined with either image, speech, or sound aids to present victim reports. Text conveyed the type of victims (wounded or dead), and the aids provided directional information (left or right) to assist in localizing the victims. Although image aids delivered useful information, they did not improve the task performance due to an inappropriate combination of modalities. Visual modalities have a lack of salience, therefore when participants were busy searching for pending patients, image aids were very likely to be overlooked[1]. In addition, according to the working memory theory and the multiple resource theory (see Chapter 3), the perception of image aids could not be time-shared with the rescue task, because they both consumed visual perceptional resources. When the rescue task itself was demanding, visual aids were more likely to cause overload than to be of help. In contrast, speech and sound aids could well attract attention and could be perceived in parallel with the ongoing rescue task. Therefore, they both improved the task performance and reduced mental workload and stress.

Between the two auditory aids, speech aids were more suitable than sound aids, because they better assisted the short-term memorization of pending patients. According to the working memory theory (Section 3.4.1), the working memory usually relies on subvocal speech to maintain a memory trace. This is to say the speech aids ‘left’ and ‘right’ could be directly rehearsed, but the direction of a sound, as nonverbal information, had to be converted into a verbal form in order to be maintained. This conversion (via referential connections) consumed additional cognitive resources, and was probably the reason why subjects found it harder to maintain a queue of untreated patients with sound aids than with speech aids. In conclusion, in order to combine modalities in a cognitively compatible manner, the cognitive properties of modalities should be taken into account. Chapter 3 provides a theoretical foundation for making such cognitively-compatible modality combinations.

Redundant combination

In S3.2, we investigated speech (local danger) warnings, visual warnings, and a redundant combination of the two. Using only speech or only visual modalities both had major drawbacks when presenting urgent warnings. In this study, the duration of the speech messages was too long (about 5 seconds), leaving the drivers less than 2 seconds to react before collision. Without a repeat function, speech warnings did not allow multiple perception, which harmed the message recall performance. Visual warnings could not guarantee a timely attention shift in an urgent situation, and they could not enhance drivers’ vigilance for hazards. However, the redundant combination of speech and visual modalities significantly improved the usability of the warnings. Their complementary characteristics provided both high salience and freedom of perception. As a result, this modality variant had the best recall performance, the lowest driving load score and the highest satisfaction score. Drivers

[1] The ineffectiveness of image aids in S1 might also be caused by their spatial location – at the bottom of the display and outside the rescue area. If they were overlaid on the rescue map, they would have been noticed. However, in the case of overlay, there would be a risk of interfering with the perception of victim reports.

also expected this modality variant to be the most useful in a dynamic real-life driving situation, because they could selectively focus on one modality that was more compatible with driving at a specific moment. Previous work has suggested that multisensory signals are highly able to trigger fast reactions (RT advantage, [238]), and are highly able to capture a person's attention no matter what else he/she happens to be doing at the same time [232]. Therefore, we can conclude that, for presenting time-critical information (e.g. local danger warnings), a redundant use of multiple modalities can bring "the best of both worlds".

This finding does not imply that presenting information redundantly via multiple modalities is always advantageous, because the redundancy induces additional costs in terms of perception load, interface management and monitoring demand [216]. Several suggestions on when to combine multiple modalities can be found in Section 2.4.

8.2 Other Factors and Interactions

Besides modality, we have also investigated other presentation factors, such as spatial structure (S2) and level of assistance (S3.2). There were also non-presentation factors involved, such as time limit (S2), task difficulty (S1 and S2) and driving situation (S3.2). In this section, we discuss these factors and their interaction with modality.

8.2.1 Presentation Factors

Spatial structure

In S2, the structure factor manipulated the spatial location of information items in a table. The rows of the table represented either injury categories or severity levels. The decision making task was performed significantly faster and more accurately with the by-injury structure than with the by-severity structure. This factor mostly affected the cognition stage of information processing – the stage when participants applied certain strategies to process the information items. The by-injury structure better supported the application of the normative and the unbiased heuristic strategies. When the task was under time pressure, participants developed biased strategies to speed up their decision making. The selection of strategies was also influenced by the presentation structure. The by-injury structure led to a more accurate strategy than the by-severity structure.

No interaction effect was found between modality and structure in any measure, indicating that these two presentation factors had independent influences on the decision making process. This is probably because they influence different stages of the decision making process – modality mostly affects the perception of information, whereas structure mostly affects cognition.

Level of assistance

In S3.2, the level of assistance varied between providing only warnings and providing warnings preceded by action suggestions (AS). The purpose of AS was to assist drivers in deciding how to avoid the emergent road obstacles. In this study, AS reduced unsafe behaviors in all modality conditions, and accelerated both braking and lane change reactions. AS also significantly reduced the reported driving load, in terms of attentional effort, stress, visual demand, temporal demand, and interference of driving. Drivers were more satisfied with warnings preceded by AS in this study, and predicted AS to be beneficial in various real-life driving situations. Although letting the vehicle make decisions for drivers may not always be an acceptable concept, our results have shown that AS can certainly be beneficial in urgent situations where reaction time is critical. In addition, because AS do not support drivers to update their situation awareness, it is important to provide supporting information (e.g. warning messages) together with AS, so that drivers also have the opportunity to analyze the situation and make decisions for themselves.

Interaction effects between modality and level of assistance were found in two measures. According to the measure of unsafe behaviors, AS particularly increased driving safety when the warnings were presented with only speech, suggesting that AS can be more beneficial when drivers have less time to analyze and respond to warnings. In terms of driver satisfaction, AS changed drivers' preference between the purely visual warnings and the pure speech warnings. When AS were provided, drivers were more satisfied with the visual warnings, because the visual warnings allowed them to quickly validate the AS, and the lack of salience was compensated by the AS which were delivered aurally. Without AS, however, the pure speech warnings were rated more satisfying than visual warnings, due to a higher salience. This interaction effect suggested that drivers considered salience to be more important than perception efficiency for highly urgent warnings.

8.2.2 Non-presentation Factors

Time limit

In S2, a time limit of 20 seconds per task was applied as a between subject factor. The participants who were given the time limit all complied to it strictly. The time limit successfully imposed pressure on the decision making task, because participants who were not given the time limit spent significantly longer on each task (76% longer on average) and made significantly more accurate decisions (15.6% more on average). To cope with the time pressure, participants tended to focus on a subset of information (e.g. more important injury items) and ignore certain rules (e.g. weights). This way decisions could be made faster with less cognitive effort, but at the cost of decreasing accuracy.

The time efficiency of decision making revealed significant interaction effects between the time limit and the two presentation factors (with structure at a 90% confidence level). The variances of time efficiency between different presentation conditions were generally smaller with than without the time limit. This was due to the fact that all presentation conditions had the same time limit which was in most of the cases insufficient for making the decisions. Furthermore, the structure factor had a significant effect on the accuracy of decision making *only* when the time limit was applied. Adding the time limit generally decreased decision accuracy in all presentation conditions, but the decrease was less when the structure provided a better cognitive fit to the task. Overall speaking, the time limit enhanced the influence of information presentation on decision making performance, because without a time limit, the presentation factors influenced only the time efficiency of decision making; while with a time limit, they influenced both the time efficiency and the accuracy of decision making.

Task difficulty

In S1, the task load was relatively low during the first minute of each trial and increased during the later 2 minutes. Accordingly, the task performance was perfect during the first minute (all patients were successfully rescued) and degraded later. When analyzing the reaction time (RT) during the first minute, we found that the RT variances between different modality conditions were much smaller than the ones calculated from the whole trial. In addition, the modality factor did not have a significant effect on RT during the first minutes, but it did when the task load increased.

In combination with the findings of the ‘time limit’ factor, we suggest that the effect of presentation factors on task performance and cognitive load is particularly notable in a high-load task setting. The high load can be caused by time pressure and/or a high task difficulty. Therefore, it is particularly important for IMMP systems with high-load applications to consider the cognitive impact of information presentation.

Driving situation

In S3.2, drivers were asked to judge how useful the local danger warnings would be in several real-life driving situations which were not simulated in this study. Regardless of warning presentation, drivers considered local danger warnings the most useful when the visibility was low (e.g. driving at night, fog, heavy rain) and the least useful when driving was highly demanding (e.g. in an unfamiliar city, in heavy traffic). In all situations, the higher level of assistance was expected to be more useful than the lower level.

An interaction effect was found between modality and driving situation, indicating that drivers had different modality preferences in different driving situations. Visual modalities were expected to be highly useful when driving with rich surrounding sounds. Speech warnings were highly appreciated when the visibility was low or when the traffic condition was highly demanding. When the driver was tired or unconcentrated, or when the trip was long and boring, auditory signals (e.g. speech, beep sounds) were expected to be necessary, because they could attract attention and increase the driver’s vigilance level. However, pure speech warnings were not preferred in these situations. Although drivers’ modality preferences differed with the driving situations, the results suggested that multimodal presentation using both visual and auditory modalities could be the most effective and acceptable default setting for presenting local danger warnings.

8.3 Measures

We applied three categories of measures. Physiological measures were used in S1. Performance and subjective measures were used in all studies.

8.3.1 Physiological Measures

In S1, we used 7 physiological measures to assess cognitive load (see Table 4.2). According to previous findings, when cognitive demand increases, heart rate increases, heart rate variability decreases, skin conductivity increases and respiration rate increases (see Section 1.4). In this study, all performance and subjective measures revealed significant differences between modality conditions, which were consistent across measures. However, only one physiological measure (heart rate variability, LF) showed a modality effect on cognitive load at a 90% confidence level. We found two reasons to explain why physiological measures were ineffective for the purpose of this study.

First, physiological measures are generally insensitive to minor variances in cognitive load, which has also been found by previous studies [130]. We compared participants' physiological states when they were relaxed (recorded during relaxation) and when performing the task (average over all task trials). All measures except one showed significant differences between the two conditions, indicating that they did pick up the major changes in cognitive load. However, they failed to reveal the relatively minor variance in cognitive load between the five task conditions.

Second, the sensitivity of a physiological measure seems to differ from person to person. We analyzed 5 participants' data individually. The results showed that each physiological measure was sensitive only for a few participants but not for all (see Table 4.4). For one participant, none of the measures were sensitive to the variances in cognitive load. When conducting statistical analyses on the data set of all participants, these individual differences prevented a consistent pattern being found. Therefore, we suggest that when using physiological measures to assess minor variances in cognitive load, the selection of measures needs to be tailored to each subject individually. Note that we only applied PNS (peripheral nervous system) measures in this study, and thus the limitations mentioned here may not apply to CNS (central nervous system) measures such as EEG, ERP, and MEG (see Section 1.4).

8.3.2 Performance and Subjective Measures

The performance measures and subjective measures used in our studies all led to meaningful results. They were able to reflect the manipulated changes in the task difficulty and cognitive demand. Here, we focus on discussing the consistency between measures.

In S1, S2, S3.1 and S3.2, performance and subjective measures led to highly consistent findings in terms of statistical outcomes and majority preferences. If the task was performed significantly better in a certain presentation condition, this condition was usually also liked more by a majority of participants. This consistency of measures allows us to pinpoint one presentation strategy that is the most suitable for a majority of users. However, such consistency does not always exist at an individual level. For example, some participants reported one task condition to be the easiest but performed better in other conditions. In S2, we found that subjective judgement of cognitive load was more associated with the accuracy of decisions but not with the time efficiency. This means that the participants who had different preferences from the majority usually performed more accurately in the conditions they preferred but not always faster. This finding suggests that in applications where the decision quality of every single user needs to be guaranteed, it might be wise to tune the standardized interface design for each user based on his/her individual preferences.

In S2 and S3.1, participants were asked to compare the task load between images and text conditions. In both studies, a majority of participants found the task easier to perform with image than with text, which is consistent with the significantly better performance in the image conditions. However, a minority of participants (12.5% in S2 and 20% in S3.1) preferred text to image. Interestingly, most of these participants had one thing in common – they had a text or speech related research topic to work on daily, such as text retrieval or dialog management. It seems that professional training background is associated with people's modality preference. It is hard to say whether working with text makes people prefer text or that people who prefer text choose to work with text. Although lacking sufficient evidence, our findings hint that a user's professional training background may lead to a modality preference that is inconsistent with the majority of users.

In S4, the consistency between different measures was lower than in the other studies. Some performance measures showed the advantage of vibration cues, as they were identified more accurately and interfered less with the concurrent primary task. Some other performance measures spoke for sound cues, because they were identified faster. According to subjective measures, a majority of participants preferred vibration cues in one task condition and preferred sound cues in the other 4 conditions. Due to this inconsistency, we were not able to conclude which modality was more suitable for presenting informative interruption cues. However, these measures allowed us to understand what the strengths and weaknesses of each modality are, based on which we have made several situation-dependent design suggestions (see Section 7.5).

Overall, we suggest that it is good to combine performance and subjective measures for the usability assessment of interface design. Performance measures are objective and they assess usability from the perspective of system functionality. On the other hand, subjective measures reflect usability from the perspective of user experience and preference. In combination, these two types of measures provide a fairly complete picture of the quality of interaction. Independent of their consistency, designers can make meaningful design suggestions based on the results of these measures.

9
Conclusions

"There will come a time when you believe everything is finished.
That will be the beginning."
Louis L'Amour

This chapter concludes this dissertation. Section 9.1 summarizes the contributions of our work. Section 9.2 presents several avenues for future research.

9.1 Summary of Contributions

Our work presented in this dissertation was focused on investigating the effect of information presentation in a high load task setting, aiming to provide useful suggestions on the design of multimodal interfaces. Here we briefly summarize the contributions of each study in a separate subsection.

9.1.1 Information Presentation for Time Limited Visual Search

In this study, we investigated the use of modality for assisting a high-load visual search task embedded in a crisis rescue scenario. We found that modality significantly influenced the performance and cognitive demand of the task, particularly the relatively difficult tasks. A visual, nonverbal, and analog modality (e.g. image) was the most suitable for presenting concrete objects (e.g. victims). Under time pressure and a high visual perceptional demand, it was beneficial to present partial information (localization aids) via auditory modality. This way, the multimodal presentation made a better use of the human cognitive resources and thus increased the bandwidth of information transfer. Of the auditory modalities, speech was found to be better than sound at assisting the short-term memorization of pending tasks. The findings of this study were well explained by several modality-related cognitive theories, such as the working memory theory, the dual-coding theory, and the multiple resource theory (see Chapter 3). This suggests that these theories provide a theoretical foundation for cognitive-aware modality allocation.

Further, we proposed a suitability prediction model that quantitatively evaluates modality allocation options and systematically selects the best option for a specific presentation task. The model was demonstrated using our presentation task, and was applied to 10 possible modality allocation strategies (5 tested and 5 additional). The output of the model is consistent with the experimental results and it predicts a combination of image and speech to be the best for this presentation task. This model demonstrates a way to integrate modality-related cognitive theories and findings (such as the ones presented in Chapter 3) into a systematic modality selection process. Furthermore, we have provided several suggestions on how to adapt the model to other presentation tasks.

9.1.2 Information Presentation for Time Limited Decision Making

This study investigated the effect of information presentation and time limit on a multi-attribute decision making task embedded in a crisis rescue scenario. We found that the modality factor had a strong influence on the efficiency of information perception, which in turn affected the speed of decision making. Decision makers adapted their information processing strategies to the spatial structure of the presentation. As a result, presentation structure influenced both the time efficiency and the accuracy of decision making. Moreover, the effect of presentation was particularly notable when decision making was under time pressure. Without the time limit, the suboptimal presentation formats mostly decreased the speed of decision making but not the accuracy. However, under time pressure, both time efficiency and accuracy were affected.

The results of this study suggest that the selection of modality should suit the type of information to be presented and allow it to be perceived easily and quickly. For example, image is suitable for presenting concrete concepts and objects. Color coding is effective for indicating state (e.g. severity) and communicating qualitative/quantitative differences. The spatial structure of information items should provide a cognitive fit to the task, that is to support the application of desired decision making strategies. If several information items need to be considered together, they should be spatially clustered. When using a table, locate the more critical information items at the top. It might be unavoidable that the performance of decision making decreases under high time pressure, but an optimal information presentation can limit this decline.

9.1.3 Presenting Local Danger Warnings

We conducted two studies investigating the presentation of in-vehicle local danger warnings. The warnings warned drivers of emergent road obstacles that were not yet visible to the driver. Study One was focused on the perception efficiency of warnings, aiming to obtain a visual presentation that could be perceived with little time and effort. The results revealed that visual warnings could be perceived faster than speech warnings that convey the same information contents. The perception time of the best visual presentation in this study (an iconic presentation) was only 36% of the duration of the speech sentences (8∼10 words, about 5 seconds).

Study Two further investigated four modality variants (speech warning, visual and speech warning, visual warning with blinking cue, and visual warning with sound cue) and two levels of assistance (warning only or warning with action suggestions). This study revealed that local danger warnings were indeed beneficial, because they significantly enhanced drivers' performance of danger avoidance and also reduced their stress level and attention demand. Drivers expected local danger warnings to be particularly useful when driving with a low visibility (e.g. in the fog, at night). Regarding modality, the results suggest that a redundant combination of auditory and visual warnings is the most suitable for presenting highly urgent warnings. Speech warnings should be kept short and more important information items should be located earlier in the word sequence. When adding attention-attracting cues to the warnings, it is better to use auditory modalities (e.g. beep sounds) than visual modalities (e.g. blinking objects). Regarding the level of assistance, it is certainly beneficial to provide action suggestions together with local danger warnings. However, action suggestions should not be given to drivers without supporting information, because drivers should always have the opportunity to analyze the situation and make decisions for themselves.

9.1.4 Presenting Informative Interruption Cues

As in-vehicle information systems are increasingly able to obtain and deliver information, we proposed the use of informative interruption cues (IIC) to reduce unnecessary driver distractions. IIC announce the arrival and convey the priority of IVIS (in-vehicle information systems) messages, based on which drivers can better manage their attention between driving and IVIS interruptions. This study was focused on the design and presentation of IIC. We designed a set of sound and vibration cues for four levels of priority. Three sound features (pitch, number of pulses, pace) and three vibration features (intensity, number of pulses, pace) were varied in a unified manner to convey different priorities (see Figure 7.1). Furthermore, the cues were evaluated in 5 task conditions which simulated the perceptional and cognitive load in real driving situations. The experimental results showed that the cues could be quickly learned ($<$ 5 minutes), reliably detected (99.5%), quickly identified (2.8 seconds after onset and 1.3 seconds after offset) and accurately identified (92.5% over all task conditions). This suggests that our design of cues is effective under various types of cognitive load that drivers may encounter during driving, and thus is a promising solution to convey the priority of IVIS messages for supporting drivers' attention management. Our results also provide hints on how to enlarge the contrasts between the cues in order to further improve their effectiveness (see Section 7.5).

The comparison between vibration and sound is not clear-cut. Vibration cues were identified more accurately than sound cues, and interfered less with the concurrent visual tracking task. Sound cues also had advantages over vibration cues in terms of shorter response time and more (reported) physical comfort. Several tentative suggestions on the selection between sound and vibration cues can be found in Section 7.5.

9.2 Future Research

This section discusses three themes of future research: automatic modality allocation, in-vehicle information presentation and cognitive-aware IMMP (intelligent multimodal presentation systems).

9.2.1 Computational Modality Allocation

We have made an attempt to construct a computational model for automatically predicting the most suitable modality allocation strategy based on selected inputs and rules (see Section 4.5). The validity of the model has been supported by the consistency between its predictions and our experimental results in Chapter 4. In the future, we would like to further develop this method using different use cases. Given a particular use case, we need to determine input, output, attributes and weights for the model. For example, in the case of presenting local danger warnings, the input of the model (i.e. modality alternatives) can be visual (text, image), auditory (sound, speech) and tactile modalities (vibration, force). The output of the model corresponds to relevant usability measures, such as the speed and correctness of drivers' reaction to the warnings. Attributes include all factors that may influence the output, such as the perceptional property of modalities, the salience of modalities, the environmental sound level, the driving demand et cetera. For each attribute, output values need to be determined for every modality alternative. Finally, weights can be assigned to the attributes based on how much each attribute may influence the output. We would like to work out the details of this model for several use cases like the one mentioned. Eventually, we hope to apply this computational method to automatically allocate modalities for IMMP.

9.2.2 In-vehicle Information Presentation

The driving environment presents exciting challenges to the presentation of in-vehicle information. As cars are getting more and more new functions for advanced driver assistance, in-vehicle information presentation will continue to be an important work topic for researchers and designers. Here we discuss several directions to follow up our work in IVIS information presentation.

The medium of visual modalities used in our study (Chapter 6) is a head-down display (HDD) which is spatially separated from the driving scene. HDD has a disadvantage because drivers have to lower their gaze to read the messages on the display, which creates attention gaps in monitoring the traffic. In the future, we would like to explore presenting information on a head-up display (HUD). HUD is a transparent display that presents information without requiring users to look away from their usual viewpoints. Using the windshield of a car as a HUD has already been implemented and investigated in recent years (e.g. [47; 63; 143]). It has been found that the perception of visual information is more compatible with driving with a HUD than with a HDD [143]. We intend to find out whether HUD can enhance the effectiveness of local danger warnings and other IVIS functions.

Tactile modalities have certain properties that make them very promising for in-vehicle information presentation (see Section 3.6). In addition, our work in Chapter 7 shows that the pattern of tactile signals can also effectively convey information. In the future, we would like to further explore tactile modalities for presenting information to drivers. A direct next step is to investigate the influence of informative interruption cues on drivers' attention management (following up the work in Chapter 7). We expect that drivers can make use of the priority information conveyed by the cues to better manage their attention between driving and IVIS interruptions. We will also compare vibration cues to sound cues, aiming to obtain a deeper understanding of how to select between these two modalities. Furthermore, we also would like to explore 1) the combination of tactile modalities with visual and auditory modalities, 2) tactile modalities other than vibration, such as force feedback, and 3) different locations to provide tactile signals to drivers, such as at the back of the seat and on the seat belt.

Driving in real life can be more dynamic and demanding than in our laboratory settings. For example, drivers often encounter bad weather and heavy traffic. They may listen to the radio and have a conversation all at the same time. Although we have addressed the dynamics of the driving situation to a certain extend, it is still necessary to further validate our design suggestions using a more realistic driving setting. Conducting field studies is obviously one way to achieve this goal. However, if the tasks may potentially create unsafe situations under real-world conditions (e.g. to avoid emergent local danger), simulation may still be a safer evaluation tool. Several previous studies have suggested that medium fidelity simulation is a valid and effective means to assess basic task performance and visual attention for the purposes of comparing user interface designs [117; 276]. Therefore, we aim to simulate the dynamics of real-world driving for our future evaluation of in-vehicle interfaces.

9.2.3 Cognitive-aware IMMP

As introduced in Chapter 2, intelligent multimodal presentation systems (IMMP) are able to adapt their output generation to the run-time requirements of user-computer interaction, and the adaption process is commonly rule-based. IMMP can be cognitive-aware if knowledge about the cognitive impact of information presentation is added into the rule base. The aim is to present information in a cognitively compatible manner, so that it can be more efficiently perceived and processed by the human cognition system. Our work presented in this dissertation contributes to understanding the cognitive impact of information presentation. In the future, we would like to go one step further to design cognitive-aware IMMP for a certain application. Regarding the use of modality, the knowledge in Chapter 3 provides a theoretical basis for developing cognitive-aware modality allocation rules. Once the application is given and its requirements to interface design are defined, we can match these requirements for the cognitive properties of available modalities, in order to determine which modality(ies) is(are) more suitable in which states of interaction. Besides the cognitive capability, other factors that are usually considered in IMMP are always relevant and cannot be left out, such as the type of information, condition of environment, user profile, et cetera.

Another factor that is worth taking into account is the cognitive/attentional state of the user. If the system is able to sense and reason about the cognitive state and attentional foci of the user at real time, it can use this input to present information in a more effective and appropriate manner. Systems that are capable of reasoning about and reacting to the user's cognitive/attentional state are called attentive user interfaces [265; 266] or attention-aware systems [206]. A user's attention can be detected by tracking gaze, gesture and head pose, and measuring physiological state (heart rate, skin conductance, etc.) and electroencephalogram (EEG) [206]. These inputs of user cognitive/attentional state will allow us to make a better use of our knowledge about the cognitive impact of presentation. In the future, we would like to integrate the capability of attentive user interfaces into IMMP. We believe that the combination of cognitive-aware input (users' cognitive states) and cognitive-aware output (presentations) is an ultimate solution to cognitive-aware IMMP.

Bibliography

[1] ISO 9241: Ergonomic requirements for office work with visual display terminals (VDTs) – Part 11: Guidance on usability, 1998.

[2] L. Adelman, T. Bresnick, P.K. Black, F.F. Marvin, and S.G. Sak. Research with patriot air defense officers: Examining information order effects. *Human Factors*, 38(2):250261, 1996.

[3] L. Adelman, M.A. Tolcott, and T.A. Bresnick. Examining the effect of information order on expert judgment. *Organizational Behavior and Human Decision Processes*, 56:348–348, 1993.

[4] D. Alais, C. Morrone, and D. Burr. Separate attentional resources for vision and audition. In *Proceedings of the Royal Society B: Biological Sciences*, volume 273, pages 1339–1345. The Royal Society, 2006.

[5] J.F. Allen. Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11):832–843, 1983.

[6] J.S. Ancker, Y. Senathirajah, R. Kukafka, and J.B. Starren. Design features of graphs in health risk communication: A systematic review. *Journal of the American Medical Informatics Association*, 13(6):608–618, 2006.

[7] E. André. The generation of multimedia presentations. In R. Dale, H.L. Somers, and H. Moisl, editors, *Handbook of Natural Language Processing*, pages 305–327. Marcel Dekker Inc., 2000.

[8] Y. Arens, E. Hovy, and M. Vossers. On the knowledge underlying multimedia presentations. *Intelligent Multimedia Interfaces*, pages 280–306, 1993.

[9] G.R. Arrabito, T.A. Mondor, and K.J. Kent. Judging the urgency of non-verbal auditory alarms: A case study. *Ergonomics*, 47(8):821–840, 2004.

[10] S.E. Asch. Forming impressions of personality. *Journal of Abnormal and Social Psychology*, 41(3):258–290, 1946.

[11] Y. Bachvarova, B. van Dijk, and A. Nijholt. Knowledge-based approach to modality choice. In *Proceedings of the Workshop on Multimodal Output Generation*, pages 5–15, 2007.

[12] A.D. Baddeley. *Essentials of human memory*. Taylor & Francis, 1999.

[13] A.D. Baddeley. Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, 4:829–839, 2003.

[14] A.D. Baddeley and G.J. Hitch. Working memory. *The Psychology of Learning and Motivation: Advances in Research and Theory*, 8:47–89, 1974.

[15] J.J. Barrett, C.R. Scheerer, and J.P. Estes. Order of information: Impact on perceived level of functional independence. *Occupational Therapy in Health Care*, 18(4):1–12, 2005.

[16] D.R. Begault. Head-up auditory displays for traffic collision avoidance system advisories: A preliminary investigation. *Human Factors*, 35(4):707–717, 1993.

[17] S.M. Belz, G.S. Robinson, and J.G. Casali. A new class of auditory warning signals for complex systems: Auditory icons. *Human Factors*, 41(4):608–618, 1999.

[18] S.J. Bensmaia, J.H. Killebrew, and J.C. Craig. The influence of visual motion on tactile motion perception. *Journal of Neurophysiology*, 96(3):1625–1637, 2006.

[19] N.O. Bernsen. Foundations of multimodal representations: A taxonomy of representational modalities. *Interacting with Computers*, 6:347–371, 1994.

[20] N.O. Bernsen. Why are analogue graphics and natural language both needed in HCI? *Interactive Systems: Design, Specification, and Verification.*, pages 235–251, 1995.

[21] N.O. Bernsen. Multimodality in language and speech systems – From theory to design support tool. *Multimodality in Language and Speech Systems*, pages 93–148, 2002.

[22] J. Bettman and P. Kakkar. Effects of information presentation format on consumer information acquisition. *Journal of Consumer Research*, 3(4):233–40, 1977.

[23] J.R. Bettman, E.J. Johnson, and J.W. Payne. A componential analysis of cognitive effort in choice. *Organizational Behavior and Human Decision Processes*, 45(1):111–139, 1990.

[24] M.M. Blattner, D.A. Sumikawa, and R.M. Greenberg. Earcons and icons: Their structure and common design principles. *Human-Computer Interaction*, 4(1):11–44, 1989.

[25] J.P. Bliss and S.A. Acton. Alarm mistrust in automobiles: How collision alarm reliability affects driving? *Applied Ergonomics*, 34(6):499–509, 2003.

[26] A.N. Bloomfield, J.A. Sager, D.M. Bartels, and D.L. Medin. Caring about framing effects. *Mind & Society*, 5(2):123–138, 2006.

[27] M. Bordegoni, G. Faconti, S. Feiner, M.T. Maybury, T. Rist, S. Ruggieri, P. Trahanias, and M. Wilson. A standard reference model for intelligent multimedia presentation systems. *Computer Standards and Interfaces*, 18(6):477–496, 1997.

[28] W. Boucsein, A. Haarmann, and F. Schaefer. Combining skin conductance and heart rate variability for adaptive automation during simulated IFR flight. In *Proceedings of the 7th International Conference on Engineering Psychology and Cognitive Ergonomics*, pages 639–647. Springer, 2007.

[29] L.M. Brown and T. Kaaresoja. Feel who's talking: Using tactons for mobile phone alerts. In *Proceedings of the Internation Conference on Human Factors in Computing Systems*, pages 604–609. ACM, 2006.

[30] F. F. Brunel and M. R. Nelson. Message order effects and gender differences in advertising persuasion. *Journal of Advertising Research*, 43(3):330–341, 2003.

[31] R. Brünken, S. Steinbacher, J.L. Plass, and D. Leutner. Assessment of cognitive load in multimedia learning with dual-task methodology: Auditory load and modality effect. *Instructional Science*, 32:115–132, 2004.

[32] R. Buda and Y. Zhang. Consumer product evaluation: The interactive effect of message framing, presentation order, and source credibility. *Journal of Product and Brand Management*,

9(4/5):229–242, 2000.

[33] G.E. Burnett and S.M. Joyner. An assessment of moving map and symbol-based route guidance systems. *Ergonomics and Safety of Intelligent Driver Interfaces*, pages 115–137, 1997.

[34] T. Büyükkiliç. Decision making performance: The effects of time pressure and presentation factors. In *Proceedings of the 11th Twente Student Conference on Information Technology*, pages 73–78, 2009.

[35] Y. Cao, S. Castronovo, A. Mahr, and C. Müller. On timing and modality choice with local danger warnings for drivers. In *Proceedings of the International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 75–78. ACM, 2009.

[36] Y. Cao, A. Mahr, S. Castronovo, M. Theune, C. Stahl, and C. Müller. Local danger warnings for drivers: The effect of modality and level of assistance on driver reaction. In *Proceedings of the 14th International Conference on Intelligent User Interfaces*, pages 134–148. ACM, 2010.

[37] Y. Cao and M. Theune. The use of modality in in-vehicle information presentation: A brief overview. In *Proceedings of the 2nd International Workshop on Multimodal Interfaces for Automotive Applications*, pages 6–10. ACM, 2010.

[38] Y. Cao, M. Theune, and C. Müller. Multimodal presentation of local danger warnings for drivers: A situation-dependent assessment of usability. In *Proceedings of IEEE Professional Communication Conference*, pages 226–229. IEEE, July 2010.

[39] Y. Cao, M. Theune, and A. Nijholt. Modality effects on cognitive load and performance in high-load information presentation. In *Proceedings of the 13th International Conference on Intelligent User Interfaces*, pages 335–344. ACM, 2009.

[40] Y. Cao, M. Theune, and A. Nijholt. Towards cognitive-aware multimodal presentation: The modality effects in high-load HCI. In *Proceedings of the 8th International Conference on Engineering Psychology and Cognitive Ergonomics: Held as Part of HCI International 2009*, pages 12–21. Springer, 2009.

[41] Y. Cao, M. Theune, and A. Nijholt. Cognitive-aware modality allocation in intelligent multimodal information presentation. In L. Shao, C. Shan, J. Luo, and M. Etoh, editors, *Multimedia Interaction and Intelligent User Interfaces: Principles, Methods and Applications*, pages 61–83. Springer, 2010.

[42] Y. Cao, F. van der Sluis, M. Theune, R. op den Akker, and A. Nijholt. Evaluating informative auditory and tactile cues for in-vehicle information systems. In *Proceedings of the International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 102–109. ACM, 2010.

[43] L. Carver and M. Turoff. Human-computer interaction: The human and computer as a team in emergency management information systems. *Communications of the ACM*, 50(3):33–38, 2007.

[44] S. Channarukul, S.W. McRoy, and S.S. Ali. MULTIFACE: Multimodal content adaptations for heterogeneous devices. In *Proceedings of the 6th International Conference on Multimodal Interfaces*, pages 319–320. ACM, 2004.

[45] G.B. Chapman, G.R. Bergus, and A.S. Elstein. Order of information affects clinical judgment. *Journal of Behavioral Decision Making*, 9(3):201–211, 1996.

[46] Z.J. Chen, Y. He, P. Rosa-Neto, J. Germann, and A.C. Evans. Revealing modular architecture of human brain structural networks by using cortical thickness from MRI. *Cerebral Cortex*, 18(10):2374–2381, 2008.

[47] S.Y. Cheng, A. Doshi, and M.M. Trivedi. Active Heads-up Display based Speed Compliance Aid for Driver Assistance: A Novel Interface and Comparative Experimental Studies. In *Proceedings of the IEEE Intelligent Vehicles Symposium*, pages 594–599. IEEE, 2007.

[48] S.H. Cho and K.K. Gower. Framing effect on the public's response to crisis: Human interest frame and crisis type influencing responsibility and blame. *Public Relations Review*, 32(4):420–422, 2006.

[49] M.M. Chun and Y. Jiang. Top-down attentional guidance based on implicit learning of visual covariation. *Psychological Science*, 10(4):360–365, 1999.

[50] J.M. Clark and A. Paivio. Dual coding theory and education. *Educational Psychology Review*, 3(3):149–210, 1991.

[51] J.C. Cohen, S.J. Bolanowski, and R.T. Verrillo. A direct comparison of exogenous and endogenous inhibition of return and selective attention mechanisms in the somatosensory system. *Somatosensory and Motor Research*, 22(4):269–279, 2005.

[52] S. Conger. A review of colour cartography in avalanche danger visualization. In *Proceedings of the 2004 International Snow Science Workshop*, pages 477–482, 2004.

[53] C.E. Connor, H.E. Egeth, and S. Yantis. Visual attention: Bottom-up versus top-down. *Current Biology*, 14(19):850–852, 2004.

[54] J. Coutaz, L. Nigay, D. Salber, A. Blandford, J. May, and R. Young. Four easy pieces for assessing the usability of multimodal interaction: the CARE properties. In *Proceedings of INTERACT*, volume 95, pages 115–120, 1995.

[55] J. Dargahi and S. Najarian. Human tactile perception as a standard for artificial tactile sensing: A review. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 1(1):23–35, 2004.

[56] P. Dario. Tactile sensing: Technology and applications. *Sensors and Actuators A: Physical*, 26(1-3):251–256, 1991.

[57] T. De Greef, H. Lafeber, H. van Oostendorp, and J. Lindenberg. Eye movement as indicators of mental workload to trigger adaptive automation. In *Proceedings of the 5th International Conference on Foundations of Augmented Cognition: Neuroergonomics and Operational Neuroscience*, pages 219–228. Springer, 2009.

[58] B.H. Deatherage. Auditory and other sensory forms of information presentation. *Human Engineering Guide to Equipment Design*, pages 123–160, 1972.

[59] T.A. Dingus and M.C. Hulse. Some human factors design issues and recommendations for automobile navigation information systems. *Transportation Research Part C: Emerging Technologies*, 1(2):119–131, 1993.

[60] T.A. Dingus, S.G. Klauer, V.L. Neale, A. Petersen, S.E. Lee, J. Sudweeks, M.A. Perez, J. Hankey, D. Ramsey, S. Gupta, et al. The 100-car naturalistic driving study: Phase ii – Results of the 100-car field experiment. Technical Report DOT-HS810-593, National Highway Traffic Safety Administration, Washington, DC, 2006.

[61] T.A. Dingus, D.V. McGehee, N. Manakkal, S.K. Jahns, C. Carney, and J.M. Hankey. Hu-

man factors field evaluation of automotive headway maintenance/collision warning devices. *Human Factors*, 39(2):216–229, 1997.

[62] D. Dörner and E. Pfeifer. Strategic thinking and stress. *Ergonomics*, 36(11):1345–1360, 1993.

[63] A. Doshi, S.Y. Cheng, and M.M. Trivedi. A novel active heads-up display for driver assistance. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 39(1):85–93, 2009.

[64] P.E. Downing. Interactions between visual working memory and selective attention. *Psychological Science*, 11(6):467–473, 2000.

[65] T.E. Drabek and G.J. Hoetmer. *Emergency Management: Principles and Practice for Local Government*. International City Management Association, 1990.

[66] J. Driver and C. Spence. Cross-modal links in spatial attention. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 353(1373):1319–1331, 1998.

[67] J. Duncan, S. Martens, and R. Ward. Restricted attentional capacity within but not between sensory modalities. *Nature*, 387:808–810, 1997.

[68] L.J. Dunn. HCI factors affecting quality of information in crisis management systems. In *Proceedings of the 6th Australian Conference on Computer-Human Interaction*, pages 324–325. IEEE Computer Society, 1996.

[69] J. Edworthy and E. Hellier. Complex nonverbal auditory signals and speech warnings. In M.S. Wogalter, editor, *Handbook of Warnings*, pages 199–220. CRC Press, 2006.

[70] J. Edworthy, S. Loxley, and I. Dennis. Improving auditory warning design: Relationship between warning sound parameters and perceived urgency. *Human Factors*, 33(2):205–231, 1991.

[71] F.T. Eggemeier and G.F. Wilson. Performance-based and subjective assessment of workload in multi-task environments. In D.L. Damos, editor, *Multiple-task Performance*, pages 217–278. CRC Press, 1991.

[72] M. Eimer. Can attention be directed to opposite locations in different modalities? An ERP study. *Clinical Neurophysiology*, 110(7):1252–1259, 1999.

[73] M. Eimer, J. van Velzen, B. Forster, and J. Driver. Shifts of attention in light and in darkness: An ERP study of supramodal attentional control and crossmodal links in spatial attention. *Cognitive Brain Research*, 15(3):308–323, 2003.

[74] M. El Jaafari, J-F. Forzy, J. Navarro, F. Mars, and J-M. Hoc. User acceptance and effectiveness of warning and motor priming assistance devices in car driving. In *Proceedings of the Humanist Conference*, pages 311–320, 2008.

[75] N. Elouazizi and Y. Bachvarova. On cognitive relevance in automatic multimodal systems. In *Proceedings of IEEE Conference on Multimedia Software Engineering*, pages 418–426, 2004.

[76] C. Elting and G. Michelitsch. A multimodal presentation planner for a home entertainment environment. In *Proceedings of the Perceptive User Interfaces*, pages 1–5. ACM, 2001.

[77] M. R. Endsley. Measurement of situation awareness in dynamic systems: Situation awareness. *Human Factors*, 37(1):65–84, 1995.

[78] M.R. Endsley. Towards a theory of situation awareness in dynamic systems. *Human Factors*, 37(1):32–64, 1995.

[79] ETSI. Human factors (HF): Guidelines on the multimodality of icons, symbols, and pictograms. Technical Report ETSI-EG-202-048 v1.1.1 (2002-08), European Telecommunications Standards Institute, 2002.

[80] S.H. Fairclough and L. Venables. Prediction of subjective states from psychophysiology: A multivariate approach. *Biological Psychology*, 71(1):100–110, 2006.

[81] S.K. Feiner and K.R. McKeown. Automating the generation of coordinated multimedia explanations. *IEEE Computer*, 24(10):33–41, 1991.

[82] D. Feldman-Stewart, N. Kocovski, B.A. McConnell, M.D. Brundage, and W.J. Mackillop. Perception of quantitative information for treatment decisions. *Medical Decision Making*, 20(2):228–238, 2000.

[83] T.K. Ferris and N.B. Sarter. Cross-modal links among vision, audition, and touch in complex environments. *Human Factors*, 50(1):17–26, 2008.

[84] S. Fitrianie, R. Poppe, T.H. Bui, A.G. Chitu, D. Datcu, R. Dor, D.H.W. Hofs, P. Wiggers, D.J.M. Willems, M. Poel, L.J.M. Rothkrantz, L.G. Vuurpijl, and J. Zwiers. A multimodal human-computer interaction framework for research into crisis management. In *Proceedings of the Intelligent Human Computer Systems for Crisis Response and Management*, pages 149–159, 2007.

[85] A. Gallace, H.Z. Tan, and C. Spence. Can tactile stimuli be subitised? An unresolved controversy within the literature on numerosity judgments. *Perception*, 37(5):782–800, 2008.

[86] GARMIN Ltd. Manual of nüvi 205W series personal travel assistant. Downloadable at www.garmin.com.

[87] W.W. Gaver. Auditory icons: Using sound in computer interfaces. *Human-Computer Interaction*, 2(2):167–177, 1986.

[88] D.M. Gavrila. Sensor-based pedestrian protection. *IEEE Intelligent Systems*, 16(6):77–81, 2001.

[89] K.W. Gish, L. Staplin, J. Stewart, and M. Perel. Sensory and cognitive factors affecting automotive head-up display effectiveness. *Journal of the Transportation Research Board*, 1694:10–19, 1999.

[90] S. Gold and A. Rangarajan. A graduated assignment algorithm for graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18:377–388, 1996.

[91] C. Gonzalez, J. Dana, H. Koshino, and M. Just. The framing effect and risky decisions: Examining cognitive functions with fMRI. *Journal of Economic Psychology*, 26(1):1–20, 2005.

[92] R. Graham. Use of auditory icons as emergency warnings: Evaluation within a vehicle collision avoidance application. *Ergonomics*, 42(9):1233–1248, 1999.

[93] M. Green. “How long does it take to stop?” Methodological analysis of driver perception-brake times. *Transportation Human Factors*, 2(3):195–216, 2000.

[94] A. Haag, S. Goronzy, P. Schaich, and J. Williams. Emotion recognition using bio-sensors: First steps towards an automatic system. *Affective Dialogue Systems*, pages 36–48, 2004.

[95] K.S. Hale and K.M. Stanney. Deriving haptic design guidelines from human physiological, psychophysical, and neurological foundations. *IEEE Computer Graphics and Applications*, 24(2):33–39, 2004.

[96] S. Hameed, T. Ferris, S. Jayaraman, and N. Sarter. Using informative peripheral visual and tactile cues to support task and interruption management. *Human Factors*, 51(2):126–135, 2009.

[97] S.G. Hart and L.E. Staveland. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Human Mental Workload*, 1:139–183, 1988.

[98] R.S. Heller, C.D. Martin, N. Haneef, and S. Gievska-Krliu. Using a theoretical multimedia taxonomy framework. *Journal on Educational Resources in Computing (JERIC)*, 1(1-22), 2001.

[99] S.G. Hill, H.P. Iavecchia, J.C. Byers, A.C. Bittner, A.L. Zaklad, and R.E. Christ. Comparison of four subjective workload rating scales. *Human Factors*, 34(4):429–439, 1992.

[100] C. Ho, N. Reed, and C. Spence. Assessing the effectiveness of "intuitive" vibrotactile warning signals in preventing front-to-rear-end collisions in a driving simulator. *Accident Analysis and Prevention*, 38(5):988–996, 2006.

[101] C. Ho, N. Reed, and C. Spence. Multisensory in-car warning signals for collision avoidance. *Human Factors*, 49(6):1107–1114, 2007.

[102] C. Ho, C. Spence, and H.Z. Tan. Warning signals go multisensory. In *Proceedings of HCI International*, volume 9, pages 1–10, 2005.

[103] C. Ho, H.Z. Tan, and C. Spence. Using spatial vibrotactile cues to direct visual attention in driving scenes. *Transportation Research Part F: Psychology and Behaviour*, 8(6):397–412, 2005.

[104] C.Y. Ho, M.I. Nikolic, M.J. Waters, and N.B. Sarter. Not now! Supporting interruption management by indicating the modality and urgency of pending tasks. *Human Factors*, 46(3):399–409, 2004.

[105] C.Y. Ho and N.B. Sarter. Supporting synchronous distributed communication and coordination through multimodal information exchange. *Proceedings of Human Factors and Ergonomics Society Annual Meeting*, 48(3):426–430, 2004.

[106] W.J. Horrey and C.D. Wickens. Driving and side task performance: The effects of display clutter, separation, and modality. *Human Factors*, 46(4):611–624, 2004.

[107] J.S. Horsley and R.T. Barker. Toward a synthesis model for crisis communication in the public sector: An initial investigation. *Journal of Business and Technical Communication*, 16(4):406–440, 2002.

[108] L. Huang and H. Pashler. Working memory and the guidance of visual attention: Consonance-driven orienting. *Psychonomic Bulletin and Review*, 14(1):148–153, 2007.

[109] J.B. Hurwitz and D.J. Wheatley. Using driver performance measures to estimate workload. In *Proceedings of Human Factors and Ergonomics Society Annual Meeting*, volume 46, pages 1804–1808. Human Factors and Ergonomics Society, 2002.

[110] S.T. Iqbal, X.S. Zheng, and B.P. Bailey. Task-evoked pupillary response to mental workload in human-computer interaction. In *Proceedings of the International Conference on Human Factors in Computing Systems*, pages 1477–1480. ACM, 2004.

[111] L. Itti and C. Koch. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40:1489–1506, 2000.

[112] A.H. Jamson, S.J. Westerman, G.R.J. Hockey, and O.M.J. Carsten. Speech-based e-mail and driver behavior: Effects of an in-vehicle message system interface. *Human Factors*, 46(4):625–639, 2004.

[113] W.H. Janssen and H. Thomas. In-vehicle collision avoidance support under adverse visibility conditions. *Ergonomics and Safety of Intelligent Driver Interfaces*, pages 221–229, 1997.

[114] W.A. Johnston and V.J. Dark. Selective attention. *Annual Reviews of Psychology*, 37:43–75, 1986.

[115] L.A. Jones and N.B. Sarter. Tactile displays: Guidance for their design and application. *Human Factors*, 50(1):90–111, 2008.

[116] S. Kalyuga, P. Chandler, and J. Sweller. Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology*, 13(4):351–371, 1999.

[117] N.A. Kaptein, J. Theeuwes, and R. Van Der Horst. Driving simulator validity: Some considerations. *Transportation Research Record: Journal of the Transportation Research Board*, 1550:30–36, 1996.

[118] C. Karagiannidis, A. Koumpis, and C. Stephanidis. Adaptation in IMMPS as a decision making process. *Computer Standards and Interfaces*, 18(6):509–514, 1997.

[119] M.K. Karavidas, P.M. Lehrer, S.E. Lu, E. Vaschillo, B. Vaschillo, and A. Cheng. The effects of workload on respiratory variables in simulated flight: A preliminary study. *Biological Psychology*, 84(1):157–160, 2010.

[120] A. Kassner. Meet the driver needs by matching assistance functions and task demands. In *Proceedings of the European Conference on Human Centred Design for Intelligent Transport Systems*, pages 327–334, 2008.

[121] C. Kaufmann, R. Risser, A. Geven, and R. Sefelin. Effects of simultaneous multi-modal warnings and traffic information on driver behaviour. In *Proceedings of European Conference on Human Centred Design for Intelligent Transport Systems*, pages 33–42, 2008.

[122] G. Keinan, N. Friedland, and Y. Ben-Porath. Decision making under stress: Scanning of alternatives under physical threat. *Acta Psychologica*, 64(3):219–228, 1987.

[123] P. Keller and C. Stevens. Meaning from environmental sounds: Types of signal-referent relations and their effect on recognizing auditory icons. *Journal of Experimental Psychology: Applied*, 10:3–12, 2004.

[124] D. Kern, P. Marshall, E. Hornecker, Y. Rogers, and A. Schmidt. Enhancing navigation information with tactile output embedded into the steering wheel. In *Proceedings of the 7th International Conference on Pervasive Computing*. Springer, 2009.

[125] S.G. Klauer, T.A. Dingus, V.L. Neale, J.D. Sudweeks, and D.J. Ramsey. The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic study data. Technical Report DOT-HS-810-594, US Department of Transportation: National Highway Traffic Safety Administration, 2006.

[126] D.N. Kleinmuntz and D.A. Schkade. Information displays and decision processes. *Psychological Science*, 4(4):221–227, 1993.

[127] T. Kosch. Local danger warning based on vehicle ad-hoc networks: Prototype and simulation. In *Proceedings of the 1st International Workshop on Intelligent Transportation*, 2004.

[128] K.M. Kowalski-Trakofler, C. Vaught, and T. Scharf. Judgment and decision making under stress: An overview for emergency managers. *International Journal of Emergency Management*, 1(3):278–289, 2003.

[129] N. Krahnstoever, E. Schapira, S. Kettebekov, and R. Sharma. Multimodal human-computer interaction for crisis management systems. In *Proceedings of the 6th IEEE Workshop on Applications of Computer Vision*, pages 203–207, 2002.

[130] A.F. Kramer. Physiological metrics of mental workload: A review of recent progress. In D.L. Damos, editor, *Multiple-task performance*, pages 279–328. CRC Press, 1991.

[131] G. Labiale. In-car road information: Comparisons of auditory and visual presentations. In *Proceeding of the Human Factors and Ergonomics Society Annual Meeting*, volume 34, pages 623–627. Human Factors and Ergonomics Society, 1990.

[132] F. Landragin. Physical, semantic and pragmatic levels for multimodal fusion and fission. In *Proceedings of the 7th International Workshop on Computational Semantics*, 2007.

[133] T.C. Lansdown. Visual allocation and the availability of driver information. *Traffic and Transport Psychology: Theory and Application*, pages 215–223, 1997.

[134] R.A. LeBoeuf and E. Shafir. Deep thoughts and shallow frames: On the susceptibility to framing effects. *Journal of Behavioral Decision Making*, 16(2):77–92, 2003.

[135] J.D. Lee, C. Carney, S.M. Casey, and J.L. Campbell. In-vehicle display icons and other information elements: Preliminary assessment of visual symbols. Technical Report FHWA-RD-99-196, U.S. Department of Transportation Federal Highway Administration, 1999.

[136] J.D. Lee, B.F. Gore, and J.L. Campbell. Display alternatives for in-vehicle warning and sign information: Message style, location and modality. *Transportation Human Factors*, 1(4):347–375, 1999.

[137] J.D. Lee, J.D. Hoffman, and E. Hayes. Collision warning design to mitigate driver distraction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 65–72. ACM, 2004.

[138] T. Leinmüller, R.K. Schmidt, B. Böddeker, R.W. Berg, and T. Suzuki. A global trend for Car-2-X communication. In *Proceedings of FISITA 2008 World Automotive Congress*, 2008.

[139] I.P. Levin, M.E. Huneke, and J.D. Jasper. Information processing at successive stages of decision making: Need for cognition and inclusion-exclusion effects. *Organizational Behavior and Human Decision Processes*, 82(2):171–193, 2000.

[140] I.P. Levin, S.L. Schneider, and G.J. Gaeth. All frames are not created equal: A typology and critical analysis of framing effects. *Organizational Behavior and Human Decision Processes*, 76(2):149–188, 1998.

[141] I.M. Lipkus and J.G. Hollands. The visual communication of risk. *JNCI Monographs*, 1999(25):149–163, 1999.

[142] Y.C. Liu. Comparative study of the effects of auditory, visual and multimodality displays on drivers' performance in advanced traveller information systems. *Ergonomics*, 44(4):425–442, 2001.

[143] Y.C. Liu and M.H. Wen. Comparison of head-up display (HUD) vs. head-down display (HDD): Driving performance of commercial vehicle operators in Taiwan. *International Journal of Human-Computer Studies*, 61(5):679–697, 2004.

[144] N.H. Lurie. Decision making in information-rich environments: The role of information structure. *Journal of Consumer Research*, 30(4):473–486, 2004.

[145] N.H. Lurie and C.H. Mason. Visual representation: Implications for decision making. *Journal of Marketing*, 71:160–177, 2007.

[146] G. Lyons, D. Sanabria, A. Vatakis, and C. Spence. The modulation of crossmodal integration by unimodal perceptual grouping: A visuotactile apparent motion study. *Experimental Brain Research*, 174(3):510–516, 2006.

[147] A. Mahr, Y. Cao, M Theune, T. Schwartz, and C. Müller. What if it suddenly fails? Behavioural aspects of advanced driver assistant systems on the example of local danger alerts. In *Proceedings of the 19th European Conference on Artificial Intelligence*, pages 1051–1052, 2010.

[148] M. Malik, J.T. Bigger, A.J. Camm, R.E. Kleiger, A. Malliani, A.J. Moss, and P.J. Schwartz. Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. *European Heart Journal*, 17(3):354–381, 1996.

[149] B. Mansoux, L. Nigay, and J. Troccaz. Output multimodal interaction: The case of augmented surgery. *People and Computers*, 20:177–192, 2007.

[150] D.C. Marshall, J.D. Lee, and P.A. Austria. Alerts for in-vehicle information systems: Annoyance, urgency, and appropriateness. *Human Factors*, 49(1):145–157, 2007.

[151] J.C. Martin. TYCOON: Theoretical framework and software tools for multimodal interfaces. In J. Lee, editor, *Intelligence and Multimodality in Multimedia Interfaces*. AAAI Press, 1998.

[152] J.C. Martin. Six primitive types of cooperation for observing, evaluating and specifying cooperations. In *Proceedings of the American Association for Artificial Intelligence Conference*, page 3551, 1999.

[153] G. Matthews, T.J. Sparkes, and H.M. Bygrave. Attentional overload, stress, and simulate driving performance. *Human Performance*, 9(1):77–101, 1996.

[154] R.E. Mayer. Systematic thinking fostered by illustrations in scientific text. *Journal of Educational Psychology*, 81(2):240–246, 1989.

[155] R.E. Mayer. Multimedia learning: Are we asking the right questions? *Educational Psychologist*, 32(1):1–19, 1997.

[156] R.E. Mayer. Elements of a science of e-learning. *Journal of Educational Computing Research*, 29(3):297–313, 2003.

[157] R.E. Mayer and R.B. Anderson. Animations need narrations: An experimental test of a dual-coding hypothesis. *Journal of Educational Psychology*, 83(4):484–490, 1991.

[158] R.E. Mayer and R.B. Anderson. The instructive animation: Helping students build connections between words and pictures in multimedia learning. *Journal of Educational Psychology*, 84:444–452, 1992.

[159] R.E. Mayer and J.K. Gallini. When is an illustration worth ten thousand words. *Journal of Educational Psychology*, 82(4):715–726, 1990.

[160] R.E. Mayer and R. Moreno. A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology*, 90:312–320, 1998.

[161] R.E. Mayer and R. Moreno. Aids to computer-based multimedia learning. *Learning and Instruction*, 12(1):107–119, 2002.

[162] R.E. Mayer and R. Moreno. Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, 38(1):43–52, 2003.

[163] P. McGettigan, K. Sly, D. OConnell, S. Hill, and D. Henry. The effects of information framing on the practices of physicians. *Journal of General Internal Medicine*, 14(10):633–642, 1999.

[164] J.D. McKeown and S. Isherwood. Mapping the urgency and pleasantness of speech, auditory icons, and abstract alarms to their referents within the vehicle. *Human Factors*, 49:417–428, 2007.

[165] S.W. McRoy, S. Channarukul, and S.S. Ali. Multimodal content adaptations for heterogeneous devices. *Journal of Digital Information*, 7(1), 2006.

[166] N. Merat, A.H. Jamson, and UK Leeds. Multisensory signal detection: How does driving and IVIS management affect performance? In *Proceedings of the 4th International Driving Symposium on Human Factors in Driver Assessemnt, Training and Vehicle Design*, pages 351–357, 2007.

[167] N. Meshkati, P.A. Hancock, M. Rahimi, and S.M. Dawes. Techniques in mental workload assessment. In J.R. Wilson and E.N. Corlett, editors, *Evaluation of Human Work: A Practical Ergonomics Methodology*, pages 749–782, 1995.

[168] G.A. Miller. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63(2):81–97, 1956.

[169] M. Mishkin, L.G. Ungerleider, and K.A. Macko. Object vision and spatial vision: Two cortical pathways. In W. Bechtel, P. Mandik, J. Mundale, and Stufflebeam R.S., editors, *Philosophy and the Neurosciences: A Reader*, pages 199–208, 2001.

[170] M.A. Mollenhauer, M.C. Hulse, T.A. Dingus, S.K. Jahns, and C. Carney. Design decision aids and human factors guidelines for ATIS displays. In Y.I. Noy, editor, *Ergonomics and Safety of Intelligent Driver Interfaces*, pages 23–61. CRC Press, 1997.

[171] M.A. Mollenhauer, J. Lee, K. Cho, M.C. Hulse, and T.A. Dingus. The effects of sensory modality and information priority on in-vehicle signing and information systems. In *Proceedings of Human Factors and Ergonomics Society Annual Meeting*, volume 38, pages 1072–1076. Human Factors and Ergonomics Society, 1994.

[172] T.A. Mondor and G.A. Finley. The perceived urgency of auditory warning alarms used in the hospital operating room is inappropriate. *Canadian Journal of Anesthesia*, 50(3):221–228, 2003.

[173] R. Moreno and R.E. Mayer. Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology*, 91:358–368, 1999.

[174] R. Moreno and R.E. Mayer. Learning science in virtual reality multimedia environments: Role of methods and media. *Journal of Educational Psychology*, 94(3):598–610, 2002.

[175] S.Y. Mousavi, R. Low, J. Sweller, et al. Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology*, 87(2):319–334, 1995.

[176] S. Mulgund, J. Stokes, and M. Turieo, M. annd Devine. Human-machine interface modalities for soldier systems technologies. Technical Report 7195000, TIAX LLC., Cambridge, 2002.

[177] V. Navalpakkam and L. Itti. A goal oriented attention guidance model. In *Proceedings of the 2nd International Workshop on Biologically Motivated Computer Vision*, pages 453–461. Springer, 2002.

[178] V.L. Neale, T.A. Dingus, S.G. Klauer, J. Sudweeks, and M. Goodman. An overview of the 100-car naturalistic study and findings. Technical Report 05-0400, National Highway Traffic Safety Administration of the United States, 2005.

[179] K.V. Nesbitt. MS-Taxonomy: A conceptual framework for designing multi-sensory displays. In *Proceedings of the 8th International Conference on Information Visualisation*, pages 665–670. IEEE, 2004.

[180] H.J. Neville and D. Lawson. Attention to central and peripheral visual space in a movement detection task: An event-related potential and behavioral study. *Brain Research*, 405(2):253–267, 1987.

[181] L. Nigay. Design space for multimodal interaction. In *IFIP Congress Topical Sessions*, pages 403–408. Springer, 2004.

[182] L. Nigay and J. Coutaz. A design space for multimodal systems: Concurrent processing and data fusion. In *Proceedings of the Conference on Human Factors in Computing Systems*, pages 172–178. ACM, 1993.

[183] Z. Obrenovic and D. Starcevic. Modeling multimodal human-computer interaction. *Computer*, 37(9):65–72, 2004.

[184] Z. Obrenovic, D. Starcevic, and E. Jovanov. Multimodal presentation of biomedical data. *Wiley Encyclopedia of Biomedical Engineering*, 2006.

[185] R.D. O'donnell and F.T. Eggemeier. Workload assessment methodology. In K.R Boff, L. Kaufman, and J.P. Thomas, editors, *Handbook of Perception and Human Performance. Volume Two: Cognitive Processes and Performance*, pages 42/1–42/49, 1986.

[186] F. Paas, A. Renkl, and J. Sweller. Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture. *Instructional Science*, 32(1):1–8, 2004.

[187] A. Paivio. Coding distinctions and repetition effects in memory. *The Psychology of Learning and Motivation: Advances in Research and Theory*, 9:179–211, 1975.

[188] A. Paivio. *Mental representations: A dual coding approach*. Oxford University Press, 1986.

[189] A.M. Parkes and N. Coleman. Route guidance systems: A comparison of methods of presenting directional information to the driver. In *Proceedings of the Ergonomics Society's 1990 Annual Conference*, pages 480–485, 1990.

[190] D. Parkhurst, K. Law, and E. Niebur. Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, 42:107–123, 2002.

[191] K. Pastra. Viewing vision-language integration as a double-grounding case. In *Proceedings of the AAAI Fall Symposium on Achieving Human-Level Intelligence through Integrated Systems and Research*, pages 62–69, 2004.

[192] A. Pauzié. Evaluating driver mental workload using the driving activity load index (DALI).

In *Proceedings of the European Conference on Human Interface Design for Intelligent Transport Systems*, pages 67–77, 2008.

[193] A. Pauzié and G. Pachiaudi. Subjective evaluation of the mental workload in the driving context. In T. Rothengatter and E. Carbonell Vaya, editors, *Traffic and Transport Psychology: Theory and Application*, pages 173–182. Pergamon, 1997.

[194] J.W. Payne, J.R. Bettman, and E.J. Johnson. *The Adaptive Decision Maker*. Cambridge University Press, New York, 1993.

[195] J.W. Payne, J.R. Bettman, and M.F. Luce. When time is money: Decision behavior under opportunity-cost time pressure. *Organizational Behavior and Human Decision Processes*, 66(2):131–152, 1996.

[196] B.M. Perrin, B.J. Barnett, L. Walrath, and J.D. Grossman. Information order and outcome framing: An assessment of judgment bias in a naturalistic decision-making context. *Human Factors*, 43(2):227–238, 2001.

[197] M.S. Peterson and A.F. Kramer. Attentional guidance of the eyes by contextual information and abrupt onsets. *Perception and Psychophysics*, 63(7):1239–1249, 2001.

[198] T. Poitschke, F. Laquai, and G. Rigoll. Guiding a driver's visual attention using graphical and auditory animations. In *Proceedings of the 8th International Conference on Engineering Psychology and Cognitive Ergonomics*, pages 424–433. Springer, 2009.

[199] K. Porter, K. Shoaf, and H. Seligson. Value of injuries in the Northridge earthquake – Technical note. *Earthquake Spectra*, 22(2):553–563, 2006.

[200] D.L. Post. Color and humancomputer interaction. In M.G. Helander, T.K. Landauer, and P.V. Prabhu, editors, *Handbook of HumanComputer Interaction*, page 573615, 1997.

[201] T.A. Ranney. Driver distraction: A review of the current state-of-knowledge. Technical Report DOT-HS-810-787, U.S. National Highway Traffic Safety Administration, 2008.

[202] T.A. Ranney, J.L. Harbluk, and Y.I. Noy. Effects of voice technology on test track driving performance: Implications for driver distraction. *Human Factors*, 47(2):439–454, 2005.

[203] L.M. Reeves, J. Lai, J.A. Larson, S. Oviatt, T.S. Balaji, S. Buisine, P. Collings, P. Cohen, B. Kraal, J.C. Martin, M. McTear, T.V. Raman, K.M. Stanney, H. Su, and Q.Y. Wang. Guidelines for multimodal user interface design. *Communications of the ACM*, 47(1):57–69, 2004.

[204] G.B. Reid and T.E. Nygren. The subjective workload assessment technique: A scaling procedure for measuring mental workload. In Hancock P.A. and Meshkati N., editors, *Human Mental Workload*, pages 185–218, 1988.

[205] K.J. Riggs, L. Ferrand, D. Lancelin, L. Fryziel, G. Dumur, and A. Simpson. Subitizing in tactile perception. *Psychological Science*, 17(4):271–275, 2006.

[206] C. Roda and J. Thomas. Attention aware systems: Theories, applications, and research agenda. *Computers in Human Behavior*, 22(4):557–587, 2006.

[207] A.H. Roscoe and G.A. Ellis. A subjective rating scale for assessing pilot workload in flight: A decade of practical use. Technical Report 90019, Royal Aerospace Establishment Rarnborough (United Kingdom), 1990.

[208] L. Rothrock, K. Barron, T.W. Simpson, M. Frecker, C. Ligetti, and R.R. Barton. Applying the proximity compatibility and the control-display compatibility principles to engineering

design interfaces. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 16(1):61–81, 2006.

[209] C. Rousseau, Y. Bellik, F. Vernier, and D. Bazalgette. Architecture framework for output multimodal systems design. In *Proceedings of OZCHI*, 2004.

[210] C. Rousseau, Y. Bellik, F. Vernier, and D. Bazalgette. A framework for the intelligent multimodal presentation of information. *Signal Processing*, 86(12):3696–3713, 2006.

[211] S. Rubio, E. Diaz, J. Martin, and J.M. Puente. Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and workload profile methods. *Applied Psychology*, 53(1):61–86, 2004.

[212] M.S. Sanders and E.J. McCormick. *Human factors in engineering and design (7th Edition)*. McGraw-Hill, 1993.

[213] N.B. Sarter. The need for multisensory interfaces in support of effective attention allocation in highly dynamic event-driven domains: the case of cockpit automation. *The International Journal of Aviation Psychology*, 10(3):231–245, 2000.

[214] N.B. Sarter. Multimodal information presentation in support of human-automation communication and coordination. *Advances in Human Performance and Cognitive Engineering Research*, 2:13–35, 2002.

[215] N.B. Sarter. Graded and multimodal interruption cueing in support of preattentive reference and attention management. *Proceedings of Human Factors and Ergonomics Society Annual Meeting*, 49(3):478–481, 2005.

[216] N.B. Sarter. Multimodal information presentation: Design guidance and research challenges. *International Journal of Industrial Ergonomics*, 36(5):439–445, 2006.

[217] M.W. Scerbo, F.G. Freeman, P.J. Mikulka, R. Parasuraman, F. Di Nocero, and L.J. Prinzel III. The efficacy of psychophysiological measures for implementing adaptive technology. Technical Report TP-2001-211018, NASA Langley Research Center, Hampton, 2001.

[218] D.A. Schkade and D.N. Kleinmuntz. Information displays and choice processes: Differential effects of organization, form, and sequence. *Organization Behavior And Human Decision processes*, 57:319–319, 1994.

[219] B. Seppelt and C.D. Wickens. In-vehicle tasks: Effects of modality, driving relevance, and redundancy. Technical Report AHFD-03-16 & GM-03-2, Aviation Human Factors Division at University of Illinois & General Motors Corporation, 2003.

[220] R. Sharma, M. Yeasin, N. Krahnstoever, I. Rauschert, G. Cai, I. Brewer, A.M. MacEachren, and K. Sengupta. Speech–gesture driven multimodal interfaces for crisis management. *Proceedings of the IEEE*, 91(9):1327–1354, 2003.

[221] H. Sharp, Y. Rogers, and J. Preece. *Interaction design: beyond human-computer interaction*. Wiley Hoboken, 2007.

[222] K. Shoaf, H. Seligson, C. Peek-Asa, and M. Mahue-Giangreco. Standardized injury categorization schemes for earthquake related injuries. Technical report, UCLA Center for Public Health and Disaster Relief, Los Angeles, CA, 2001.

[223] D.J. Simons and C.F. Chabris. Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception*, 28:1059–1074, 1999.

[224] D.J. Simons and R.A. Rensink. Change blindness: Past, present, and future. *Trends in Cognitive Sciences*, 9(1):16–20, 2005.

[225] C.A.P. Smith, B.A. Clegg, E.D. Heggestad, and P.J. Hopp-Levine. Interruption management: A comparison of auditory and tactile cues for both alerting and orienting. *International Journal of Human-Computer Studies*, 67(9):777–786, 2009.

[226] E.E. Smith and J. Jonides. Working memory: A view from neuroimaging. *Cognitive Psychology*, 33(1):5–42, 1997.

[227] E.E. Smith, J. Jonides, and R.A. Koeppe. Dissociating verbal and spatial working memory using PET. *Cerebral Cortex*, 6(1):11–20, 1996.

[228] E.E. Smith, J. Jonides, R.A. Koeppe, E. Awh, E.H. Schumacher, and S. Minoshima. Spatial versus object working memory: PET investigations. *Journal of Cognitive Neuroscience*, 7(3):337–356, 1995.

[229] T.L. Smith-Jackson and T.E. Hall. Information order and sign design: A schema-based approach. *Environment and Behavior*, 34(4):479–492, 2002.

[230] C. Spence and J. Driver. Audiovisual links in endogenous covert spatial attention. *Journal of Experimental Psychology-Human Perception and Performance*, 22(4):1005–1030, 1996.

[231] C. Spence and J. Driver, editors. *Crossmodal space and crossmodal attention*. Oxford University Press, 2004.

[232] C. Spence and A. Gallace. Recent developments in the study of tactile attention. *Canadian Journal of Experimental Psychology*, 61(3):196–207, 2007.

[233] C. Spence and C. Ho. Multisensory warning signals for event perception and safe driving. *Theoretical Issues in Ergonomics Science*, 9(6):523–554, 2008.

[234] C. Spence and Driver J. Audiovisual links in attention: Implications for interface design. *Engineering Psychology and Cognitive Ergonomics*, 2:185–192, 1997.

[235] C. Spence, M.E.R. Nicholls, and J. Driver. The cost of expecting events in the wrong sensory modality. *Perception and Psychophysics*, 63(2):330–336, 2001.

[236] C. Spence, D.I. Shore, and R.M. Klein. Multisensory prior entry. *Journal of Experimental Psychology: General*, 130(4):799–832, 2001.

[237] R. Srinivasan and P.P. Jovanis. Effect of selected in-vehicle route guidance systems on driver reaction times. *Human Factors*, 39(2):200–215, 1997.

[238] K. Stanney, S. Samman, L. Reeves, K. Hale, W. Buff, C. Bowers, B. Goldiez, D. Nicholson, and S. Lackey. A paradigm shift in interactive computing: Deriving multimodal design principles from behavioral and neurological foundations. *International Journal of Human-Computer Interaction*, 17(2):229–257, 2004.

[239] N. Stanton, editor. *Human factors in alarm design*. CRC Press, 1994.

[240] K.L. Stephan, S.E. Smith, R.L. Martin, S. Parker, and K.I. McAnally. Learning and retention of associations between auditory icons and denotative referents: Implications for the design of auditory warnings. *Human Factors*, 48(2):288–299, 2006.

[241] C. Stephanidis, C. Karagiannidis, and A. Koumpis. Integrating media and modalities in the user-machine interface. In *Proceedings of the 1st International Conference on Applied Ergonomics*, 1996.

[242] R.M. Stern, W.J. Ray, and K.S. Quigley. *Psychophysiological recording*. Oxford University Press, 2001.

[243] E.R. Stone, W.R. Sieck, B.E. Bull, J. Frank Yates, S.C. Parks, and C.J. Rush. Foreground:background salience: Explaining the effects of graphical displays on risk avoidance. *Organizational Behavior and Human Decision Processes*, 90(1):19–36, 2003.

[244] E.R. Stone, J.F. Yates, and A.M. Parker. Effects of numerical and graphical displays on professed risk-taking behavior. *Journal of Experimental Psychology: Applied*, 3:243–256, 1997.

[245] A. Sutcliffe. *Multimedia and virtual reality: designing multisensory user interfaces*. Lawrence Erlbaum Associates Inc., 2003.

[246] 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.

[247] K. Suzuki and H. Jansson. An analysis of driver's steering behaviour during auditory or haptic warnings for the designing of lane departure warning system. *Review of Automotive Engineering (JSAE)*, 24(1):65–70, 2003.

[248] J. Sweller, J.J.G. van Merriënboer, and F.G.W.C. Paas. Cognitive architecture and instructional design. *Educational Psychology Review*, 10(3):251–296, 1998.

[249] M. Tabatabaei. An experimental analysis of decision channeling by restrictive information display. *Journal of Behavioral Decision Making*, 15(5):419–432, 2002.

[250] J. Tanji and E. Hoshi. Role of the lateral prefrontal cortex in executive behavioral control. *Physiological Reviews*, 88(1):37–57, 2008.

[251] G. Tassinari and D. Campara. Consequences of covert orienting to non-informative stimuli of different modalities: A unitary mechanism? *Neuropsychologia*, 34(3):235–245, 1996.

[252] Y.F. Tsai, E. Viirre, C. Strychacz, B. Chase, and T.P. Jung. Task performance and eye activity: Predicting behavior relating to cognitive workload. *Aviation, Space, and Environmental Medicine*, 78(Supplement 1):B176–B185, 2007.

[253] E. Tufte. Powerpoint does rocket science and better techniques for technical reports, url: www.edwardtufte.com/bboard/q-and-a-fetch-msg?msg_id=0001yb&topic_id=1&topic=ask+ E.T., The Work of Edward Tufte and Graphics Press, 2005.

[254] T.S. Tullis. Screen design. In M.G. Helander, Landauer T.K., and P.V. Prabhu, editors, *Handbook of Human-Computer Interaction*, pages 503–572, 1997.

[255] M. Turatto, G. Galfano, B. Bridgeman, and C. Umilta. Space-independent modality-driven attentional capture in auditory, tactile and visual systems. *Experimental Brain Research*, 155(3):301–310, 2004.

[256] A. Tversky and D. Kahneman. The framing of decisions and the psychology of choice. *Science*, 211(4481):453–458, 1981.

[257] H.R. Unnava, R.E. Burnkrant, and S. Erevelles. Effects of presentation order and communication modality on recall and attitude. *Journal of Consumer Research*, 21(3):481–490, 1994.

[258] U.S. Department of Transportation, National Highway Traffic Safety Administration. Identify intelligent vehicle safety applications enabled by DSRC. Technical Report DOT-HS-809-859, U.S. Department of Transportation, National Highway Traffic Safety Administration,

Springfield, Virginia, USA, March 2005.

[259] J.B.F. van Erp. Guidelines for the use of vibro-tactile displays in human computer interaction. In *Proceedings of Eurohaptics*, volume 2002, pages 18–22, 2002.

[260] J.B.F. Van Erp and H. Van Veen. Vibro-tactile information presentation in automobiles. In *Proceedings of Eurohaptics*, volume 2001, pages 99–104, 2001.

[261] J.B.F. Van Erp and H.A.H.C. Van Veen. Vibrotactile in-vehicle navigation system. *Transportation Research Part F: Psychology and Behaviour*, 7:247–256, 2004.

[262] J.J.G. van Merriënboer and J. Sweller. Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review*, 17(2):147–177, 2005.

[263] K.F. Van Orden, W. Limbert, S. Makeig, and T.P. Jung. Eye activity correlates of workload during a visuospatial memory task. *Human Factors*, 43(1):111–121, 2001.

[264] F. Vernier and L. Nigay. A framework for the combination and characterization of output modalities. In *Proceedings of the 7th International Workshop on Interactive Systems Design, Specification, and Verification*, pages 35–50. Springer, 2001.

[265] R. Vertegaal. Attentive user interfaces. *Communications of the ACM*, 46(3):31–33, 2003.

[266] R. Vertegaal, J.S. Shell, D. Chen, and A. Mamuji. Designing for augmented attention: Towards a framework for attentive user interfaces. *Computers in Human Behavior*, 22(4):771–789, 2006.

[267] W.B. Verwey and H.A. Veltman. Detecting short periods of elevated workload: A comparison of nine workload assessment techniques. *Journal of Experimental Psychology: Applied*, 2(3):270–285, 1996.

[268] I. Vessey and D. Galletta. Cognitive fit: An empirical study of information acquisition. *Information Systems Research*, 2(1):63–84, 1991.

[269] J. Vibell, C. Klinge, M. Zampini, C. Spence, and A.C. Nobre. Temporal order is coded temporally in the brain: Early event-related potential latency shifts underlying prior entry in a cross-modal temporal order judgment task. *Journal of Cognitive Neuroscience*, 19(1):109–120, 2007.

[270] W. Wahlster. Smartkom: Fusion and fission of speech, gestures, and facial expressions. In *Proceedings of the 1st International Workshop on Man-Machine Symbiotic Systems*, pages 213–225, 2002.

[271] W. Wahlster. Smartkom: Symmetric multimodality in an adaptive and reusable dialogue shell. In *Proceedings of the Human Computer Interaction Status Conference*, volume 3, pages 47–62, 2003.

[272] W. Wahlster, E. André, S. Bandyopadhyay, W. Graf, and T. Rist. WIP: The coordinated generation of multimodal presentations from a common representation. *Communication from an Artificial Intelligence Perspective: Theoretical and Applied Issues*, pages 121–144, 1992.

[273] W. Wahlster, E. André, W. Finkler, H.J. Profitlich, and T. Rist. Plan-based integration of natural language and graphics generation. *Artificial Intelligence*, 63(1):387–427, 1993.

[274] J. Walker, E. Alicandri, C. Sedney, and K. Roberts. In-vehicle navigation devices: Effects on the safety of driver performance. In *Proceedings of Vehicle Navigation and Information Systems Conference*, volume 2, pages 499–525, 1991.

[275] H. Wang, T.R. Johnson, and J. Zhang. The order effect in human abductive reasoning: An empirical and computational study. *Journal of Experimental and Theoretical Artificial Intelligence*, 18(2):215–247, 2006.

[276] Y. Wang, B. Mehler, B. Reimer, V. Lammers, L.A. D'Ambrosio, and J.F. Coughlin. The validity of driving simulation for assessing differences between in-vehicle informational interfaces: A comparison with field testing. *Ergonomics*, 53(3):404–420, 2010.

[277] S. Watzman. Visual design principles for usable interfaces. In J.A. Jacko and A. Sears, editors, *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*, pages 263–285. Lawrence Erlbaum Associates Inc., 2003.

[278] C.D. Wickens. Processing resources and attention. *Multiple-task Performance*, pages 3–34, 1991.

[279] C.D. Wickens. *Engineering psychology and human performance (3rd edition)*. Prentice Hall, 1999.

[280] C.D. Wickens. Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, 3(2):159–177, 2002.

[281] C.D. Wickens. Multiple resources and mental workload. *Human Factors*, 50(3):449–455, 2008.

[282] C.D. Wickens, S.R. Dixon, and B. Seppelt. Auditory preemption versus multiple resources: Who wins in interruption management. *Proceedings of Human Factors and Ergonomics Society Annual Meeting*, 49(3):463–467, 2005.

[283] C.D. Wickens and J.S. McCarley. *Applied attention theory*. CRC Press, 2008.

[284] W.W. Wierwille and J.G. Casali. A validated rating scale for global mental workload measurement applications. In *Proceedings of the 27th Annual Meeting of Human Factors Society*, volume 1, pages 129–133, 1983.

[285] W.W. Wierwille and F.T. Eggemeier. Recommendations for mental workload measurement in a test and evaluation environment. *Human Factors*, 35(2):263–281, 1993.

[286] G.F. Wilson. An analysis of mental workload in pilots during flight using multiple psychophysiological measures. *The International Journal of Aviation Psychology*, 12(1):3–18, 2002.

[287] G.F. Wilson and F.T. Eggemeier. Psychophysiological Assessment of Workload in Multitask Environments. In D.L. Damos, editor, *Multiple-task Performance*, pages 329–360. CRC Press, 1991.

[288] J.M. Wolfe. Visual attention. *Seeing*, 2:335–386, 2000.

[289] D.D. Woods. The alarm problem and directed attention in dynamic fault management. *Ergonomics*, 38(11):2371–2393, 1995.

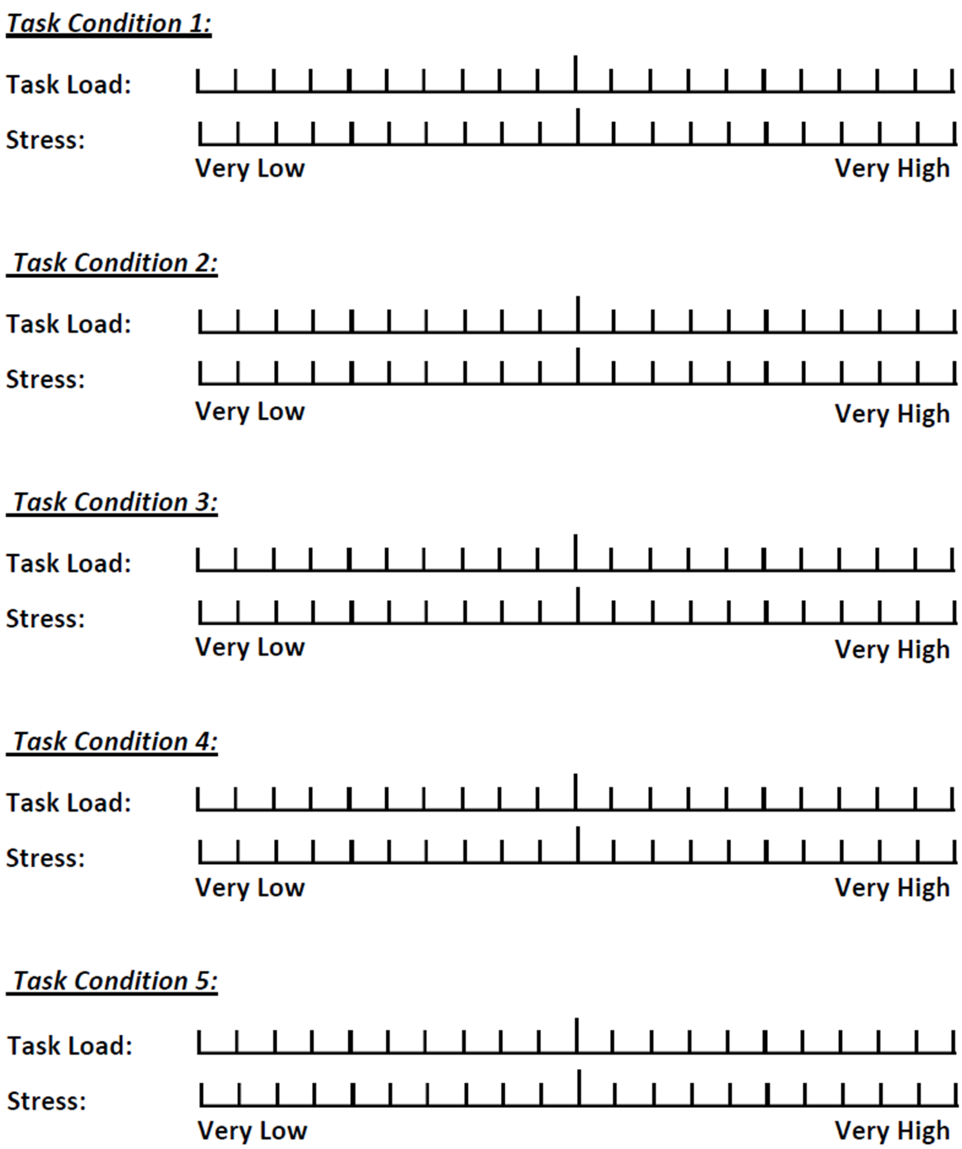
[290] J. Zhang. The nature of external representations in problem solving. *Cognitive Science*, 21(2):179–217, 1997.

[291] M.X. Zhou, Z. Wen, and V. Aggarwal. A graph-matching approach to dynamic media allocation in intelligent multimedia interfaces. In *Proceedings of the 10th International Conference on Intelligent User Interfaces*, pages 114–121. ACM, 2005.

Appendices

A: Questionnaire (Chapter 5)

Participants filled in this questionnaire after each task condition, in the actual order they encountered (everyone could have a different condition as condition 1). They were allowed to modify their previously given scores in order to maintain a correct comparison between conditions.



Task Condition 1:

Task Load:

Stress:

Very Low | Very High

Task Condition 2:

Task Load:

Stress:

Very Low | Very High

Task Condition 3:

Task Load:

Stress:

Very Low | Very High

Task Condition 4:

Task Load:

Stress:

Very Low | Very High

Task Condition 5:

Task Load:

Stress:

Very Low | Very High

B: Questionnaire (Chapter 6)

Participants filled in this questionnaire in the end of the experiment. A sheet of paper with examples of the four presentation conditions were given alongside. Condition A, B, C and D can be seen in Figure 5.1, 5.2, 5.3 and 5.4, respectively.

Compare Modality

Compare A and C, they use the same structure but different modalities, which one is easier for you?
☐ A ☐ C ☐ Exactly the Same

Compare B and D, they use the same structure but different modalities, which one is easier for you?
☐ B ☐ D ☐ Exactly the Same

Compare Structure

Compare A and B, they use the same modality but different structures, which one is easier for you?
☐ A ☐ B ☐ Exactly the Same

Compare C and D, they use the same modality but different structures, which one is easier for you?
☐ C ☐ D ☐ Exactly the Same

Compare overall

What is the most difficult condition?
☐ A ☐ B ☐ C ☐ D ☐ I don't know

What is the easiest condition?
☐ A ☐ B ☐ C ☐ D ☐ I don't know

Decision making strategy

Which decision making strategy did you use?
☐ I used a complete calculation as the normative strategy.
☐ I used eliminations and calculation as the unbiased heuristic strategy.
☐ I developed my own strategies other than the ones that were introduced to me.

Please briefly describe the decision making strategies you have used in each presentation condition, using the four presentation samples as a reference. You can also orally explain to the experimenter.

____

____

____

____

____

C: Questionnaires (Chapter 7)

Three questionnaires were used in the study presented in Section 6.4. After performing the task in each condition, participants filled in all three questionnaires for that latest condition.

C.1: Driving Activity Load Index (DALI)

The Driving Activity Load Index (DALI) is a revised version of NASA-TLX, adapted to the driving task [192; 193]. The purpose of DALI is to assess the workload of driving a vehicle equipped with on-board systems, such as IVIS, radio, car phone etcetera. DALI contains six factors which were determined by a group of experts involved in automotive studies. The description of DALI factors (see below) was available on paper to the participants throughout the experiment.

Description of DALI factors

Factor	Description
Effort of attention	To evaluate the attention required by the driving activity – to think about, to decide, to choose, to look for etcetera.
Visual demand	To evaluate the visual demand necessary for the driving activity.
Auditory demand	To evaluate the auditory demand necessary for the driving activity.
Temporal demand	To evaluate the specific constraints due to timing demands when driving.
Interference	To evaluate the possible disturbance when driving simultaneously with any other supplementary task, such as interacting with IVIS, phoning, listening to radio, and having a conversation.
Situational stress	To evaluate the level of constraints / stress while driving – fatigue, insecure feeling, irritation, discouragement etcetera.

The DALI rating scale

	Low ⟶ High					
Effort of Attention	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
Visual Demand	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
Auditory Demand	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
Temporal Demand	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
Interference	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
Situational Stress	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

C.2: Assessment of Overall Satisfaction with the Warnings

	Very dissatisfied →								*Very satisfied*
Condition 1	☐ -4	☐ -3	☐ -2	☐ -1	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4
Condition 2	☐ -4	☐ -3	☐ -2	☐ -1	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4
Condition 3	☐ -4	☐ -3	☐ -2	☐ -1	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4
Condition 4	☐ -4	☐ -3	☐ -2	☐ -1	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4
Condition 5	☐ -4	☐ -3	☐ -2	☐ -1	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4
Condition 6	☐ -4	☐ -3	☐ -2	☐ -1	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4
Condition 7	☐ -4	☐ -3	☐ -2	☐ -1	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4
Condition 8	☐ -4	☐ -3	☐ -2	☐ -1	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4

C.3: Situation Dependent Assessment of Warning Usability

Please answer the questions using the following scale:

not useful at all ⟶ *very useful*

0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

1. **How useful do you think these warnings are when there is a lot of sourrounding noise (e.g. driving noise, radio, conversations)?**

0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

Remarks: ______________________________

2. **How useful do you think these warnings are when it is dark or when the visibility is low (e.g. in the night/ fog)?**

0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

Remarks: ______________________________

3. **How useful do you think these warnings are when you are tired, unconcentrated or distracted?**

0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

Remarks: ______________________________

4. **How useful do you think these warnings are when you are on a very long and boring drive (e.g. a long trip on the highway)?**

0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

Remarks: ______________________________

5. **How useful do you think these warnings are in especially demanding situations (e.g. driving in a heavy traffic or in an unfamiliar city)?**

0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

Remarks: ______________________________

D: Questionnaires (Chapter 7)

Questionnaires used in the study presented in Chapter 7 consist of eight sessions: general information, learning session, task conditions 1 to 5, and final evaluation.

General Information

Gender: ☐ Male ☐ Female

Age: ____________________

Do you drive? ☐ Not at all ☐ Sometimes ☐ Often ☐ Very frequently

Do you have any known problem with hearing? ☐ No ☐ Yes, ______________

Do you have any known problem with tactile perception? ☐ No ☐ Yes, ______________

Are you color blind? ☐ No ☐ Yes

Learning Experience

Sound 1: Distinguish the four different pitches
Very easy ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Very difficult

Sound 2: Distinguish the four different numbers of "beep"
Very easy ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Very difficult

Sound 3: Distinguish the four different paces
Very easy ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Very difficult

Sound 4: Associate the pitch of sound with the priority of a message (intuitiveness)
higher pitch - higher priority, lower pitch - lower priority
Very easy ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Very difficult

Sound 5: Associate the number of "beep" with the priority of a message (intuitiveness)
more beeps - higher priority, less beeps - lower priority
Very easy ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Very difficult

Sound 6: Associate the pace of sound beeps with the priority of a message (intuitiveness)
faster - higher priority, slower - lower priority
Very easy ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Very difficult

Vibration 1: Distinguish the four different intensities
Very easy ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Very difficult

Vibration 2: Distinguish the four different numbers of pulses
Very easy ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Very difficult

Vibration 3: Distinguish the four different paces
Very easy ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Very difficult

Vibration 4: Associate the intensity of vibration with the priority of a message (intuitiveness)
higher intensity - higher priority, lower intensity - lower priority
Very easy ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Very difficult

Vibration 5: Associate the number of pulses with the priority of a message (intuitiveness)
more pulses - higher priority, less pulses - lower priority
Very easy ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Very difficult

Vibration 6: Associate the pace of vibration pulses with the priority of a message (intuitiveness)
faster - higher priority, slower - lower priority
Very easy ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Very difficult

Task Condition 1/2/3/4/5

1. Identifying the sound signals was
Very easy ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Very difficult

2. Identifying the vibration signals was
Very easy ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Very difficult

3. Which modality do you prefer in this task condition?
☐ sound ☐ vibration ☐ either one, they are the same

Final Evaluation

1. Did you feel (physically) comfortable with the sound signals?
Very comfortable ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Very annoyed

2. Did you feel (physically) comfortable with the vibration signals?
Very comfortable ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Very annoyed

3. Which feature(s) did you use to distinguish the four sound signals?
☐ Pitch
☐ Number of beep
☐ Pace
☐ Other: ____________________

4. Which feature(s) did you use to distinguish the four vibration signals?
☐ Pitch
☐ Number of pulses
☐ Pace
☐ Other: ____________________

Summary

This dissertation addresses multimodal information presentation in human computer interaction. Information presentation refers to the manner in which computer systems/interfaces present information to human users. More specifically, the focus of our work is not on which information to present, but on how to present it, such as which modalities to use, how to spatially distribute items, et cetera. The notion "computer" is not limited to personal computers in their various forms. It also includes embedded computers in any devices or machines.

Information presentation guides, constrains and even determines cognitive behavior. This is to say that the same information, when presented differently, can be processed differently by the human cognition system and may lead to different decisions and responses. Consequently, information presentation can influence the quality of human computer interaction, in terms of user performance, cognitive demand and user satisfaction. A good manner of presentation is particularly desired in high load interactions, because users may not have the spared cognitive capacity to cope with the unnecessary mental workload induced by a bad presentation. This dissertation work aims to investigate the effect of information presentation in high load human computer interactions, and to provide useful suggestions for the design of multimodal interfaces. The major presentation factor of interest is modality.

First, a literature study has been conducted in the cognitive psychology domain, aiming to understand the role of modality in different stages of human information processing (see Chapter 3). At least three processing stages have modality-specific features, namely sensory processing, perception and working memory. Different modalities are sensed by different sensory receptors, such as the eyes for visual modalities, the ears for auditory modalities and the skin for tactile modalities. Attention needs to be directed to a raw sensory input in order to perceive its meaning, and modalities differ in their ability to attract attention. Complex cognitive activities take place in working memory, and different categories of modality (e.g. visual vs. auditory, verbal vs. nonverbal) consume separated cognitive resources to be processed. These cognitive findings provide theoretical guidance on how to allocate modality in a cognitively appropriate manner.

We have conducted a series of user studies in two high load task domains, namely crisis management and driving. Crisis managers typically have to work under time pressure and stress, and they often have to deal with a large amount of information. One study in this domain investigated information presentation for a time limited visual search task (see Chapter 4). The task was to search for wounded earthquake victims on a map and send a doctor to rescue them. The location and type of victim were presented in five modality conditions. The results showed that modality significantly affected task performance, cognitive load and stress, especially when the task load was high. The experimental findings were well explained by several relevant cognitive theories. Further, we proposed a suitability prediction model that quantitatively evaluates modality options and selects the best option for a specific presentation task. The model was demonstrated by our task and suggestions on its generalization were given.

A second study using a crisis scenario investigated information presentation for a time limited decision making task (see Chapter 5). Participants were given information about the injuries of two patients and were asked to decide which one needed treatment first. The presentation of the injury information was manipulated by two factors, modality and spatial structure. These two factors both significantly affected task performance and cognitive load. The results further indicated that the two factors influenced different stages of information processing – modality mostly affected perception and spatial structure mostly affected cognition (cognitive activities in the working memory). Time pressure was found to heighten the effect of presentation factors on the quality of interaction.

Driving is normally not a high load task for experienced drivers. However, in the case of emergent danger, drivers need to make quick decisions and respond correctly, otherwise consequences can be catastrophic. We conducted two studies investigating the presentation of local danger warnings (see Chapter 6). Drivers received warnings about emergent road obstacles when the obstacles were not yet visible to them. Based on the warnings, they needed to avoid the obstacles by either braking or changing lane. The presentation of the warnings was varied by modality and level of assistance. The results show that it is beneficial to present both visual and speech warnings, and to suggest appropriate actions to drivers. In addition, drivers may have different preferences on the modality of warnings in different driving conditions (e.g. low visibility, noise, etc).

While driving, drivers very often engage in other activities simultaneously, such as interacting with in-vehicle information systems (IVIS), listening to the radio, talking with passengers, et cetera. Multitasking is a cause of high load and distraction during driving. To reduce harmful distractions by IVIS, we proposed to present informative interruption cues (IIC) in addition to IVIS messages. IIC not only convey the arrival of IVIS messages, but also convey their priority levels, helping drivers to decide whether and when to shift their attention to the messages. As a first step, a study was conducted to investigate the design and presentation of IIC (see Chapter 7). A set of vibration and sound cues were created to convey four levels of priority. The cues were evaluated in five task conditions that simulated the perceptional and cognitive load in real driving situations. The results showed that our design of cues was effective, because the cues could be easily learnt and accurately identified in all task conditions. Vibration and sound were both found to have advantages for presenting IIC.

All of our studies consistently show that the manner of information presentation has a great influence on the quality of human computer interaction, especially in a high load task setting. Our results also provide useful suggestions on how to present information in a way that is compatible with the characteristics of human cognition. Although these outcomes are obtained in high load task settings, they can be applied to the design of multimodal computer interfaces in general.

SIKS Dissertation Series

Since 1998, all dissertations written by Ph.D. students who have conducted their research under auspices of a senior research fellow of the SIKS research school are published in the SIKS Dissertation Series. This thesis is the 305th in the series.

2011-07 Yujia Cao (UT), *Multimodal Information Presentation for High-Load Human Computer Interaction*
2011-06 Yiwen Wang (TUE), *Semantically-Enhanced Recommendations in Cultural Heritage*
2011-05 Base van der Raadt (VU), *Enterprise Architecture Coming of Age - Increasing the Performance of an Emerging Discipline*
2011-04 Hado van Hasselt (UU), *Insights in Reinforcement Learning Formal analysis and empirical evaluation of temporal-difference learning algorithms*
2011-03 Jan Martijn van der Werf (TUE), *Compositional Design and Verification of Component-Based Information Systems*
2011-02 Nick Tinnemeier (UU), *Organizing Agent Organizations. Syntax and Operational Semantics of an Organization-Oriented Programming Language*
2011-01 Botond Cseke (RUN), *Variational Algorithms for Bayesian Inference in Latent Gaussian Models*
2010-53 Edgar Meij (UVA), *Combining Concepts and Language Models for Information Access*
2010-52 Peter-Paul van Maanen (VU), *Adaptive Support for Human-Computer Teams: Exploring the Use of Cognitive Models of Trust and Attention*
2010-51 Alia Khairia Amin (CWI), *Understanding and supporting information seeking tasks in multiple sources*
2010-50 Bouke Huurnink (UVA), *Search in Audiovisual Broadcast Archives*
2010-49 Jahn-Takeshi Saito (UM), *Solving difficult game positions*
2010-48 Milan Lovric (EUR), *Behavioral Finance and Agent-Based Artificial Markets*
2010-47 Chen Li (UT), *Mining Process Model Variants: Challenges, Techniques, Examples*
2010-46 Vincent Pijpers (VU), *e3alignment: Exploring Inter-Organizational Business-ICT Alignment*
2010-45 Vasilios Andrikopoulos (UvT), *A theory and model for the evolution of software services*
2010-44 Pieter Bellekens (TUE), *An Approach towards Context-sensitive and User-adapted Access to Heterogeneous Data Sources, Illustrated in the Television Domain*
2010-43 Peter van Kranenburg (UU), *A Computational Approach to Content-Based Retrieval of Folk Song Melodies*
2010-42 Sybren de Kinderen (VU), *Needs-driven service bundling in a multi-supplier setting - the computational e3-service approach*
2010-41 Guillaume Chaslot (UM), *Monte-Carlo Tree Search*
2010-40 Mark van Assem (VU), *Converting and Integrating Vocabularies for the Semantic Web*
2010-39 Ghazanfar Farooq Siddiqui (VU), *Integrative modeling of emotions in virtual agents*
2010-38 Dirk Fahland (TUE), *From Scenarios to components*
2010-37 Niels Lohmann (TUE), *Correctness of services and their composition*
2010-36 Jose Janssen (OU), *Paving the Way for Lifelong Learning; Facilitating competence development through a learning path specification*
2010-35 Dolf Trieschnigg (UT), *Proof of Concept: Concept-based Biomedical Information Retrieval*
2010-34 Teduh Dirgahayu (UT), *Interaction Design in Service Compositions*
2010-33 Robin Aly (UT), *Modeling Representation Uncertainty in Concept-Based Multimedia Retrieval*
2010-32 Marcel Hiel (UvT), *An Adaptive Service Oriented Architecture: Automatically solving Interoperability Problems*
2010-31 Victor de Boer (UVA), *Ontology Enrichment from Heterogeneous Sources on the Web*
2010-30 Marieke van Erp (UvT), *Accessing Natural History - Discoveries in data cleaning, structuring, and retrieval*
2010-29 Stratos Idreos(CWI), *Database Cracking: Towards Auto-tuning Database Kernels*
2010-28 Arne Koopman (UU), *Characteristic Relational Patterns*
2010-27 Marten Voulon (UL), *Automatisch contracteren*
2010-26 Ying Zhang (CWI), *XRPC: Efficient Distributed Query Processing on Heterogeneous XQuery Engines*
2010-25 Zulfiqar Ali Memon (VU), *Modelling Human-Awareness for Ambient Agents: A Human Mindreading Perspective*
2010-24 Dmytro Tykhonov, *Designing Generic and Efficient Negotiation Strategies*
2010-23 Bas Steunebrink (UU), *The Logical Structure of Emotions*
2010-22 Michiel Hildebrand (CWI), *End-user Support for Access to*
Heterogeneous Linked Data
2010-21 Harold van Heerde (UT), *Privacy-aware data management by means of data degradation*
2010-20 Ivo Swartjes (UT), *Whose Story Is It Anyway? How Improv Informs Agency and Authorship of Emergent Narrative*
2010-19 Henriette Cramer (UvA), *People's Responses to Autonomous and Adaptive Systems*

2010-18 Charlotte Gerritsen (VU), *Caught in the Act: Investigating Crime by Agent-Based Simulation*

2010-17 Spyros Kotoulas (VU), *Scalable Discovery of Networked Resources: Algorithms, Infrastructure, Applications*

2010-16 Sicco Verwer (TUD), *Efficient Identification of Timed Automata, theory and practice*

2010-15 Lianne Bodenstaff (UT), *Managing Dependency Relations in Inter-Organizational Models*

2010-14 Sander van Splunter (VU), *Automated Web Service Reconfiguration*

2010-13 Gianluigi Folino (RUN), *High Performance Data Mining using Bio-inspired techniques*

2010-12 Susan van den Braak (UU), *Sensemaking software for crime analysis*

2010-11 Adriaan Ter Mors (TUD), *The world according to MARP: Multi-Agent Route Planning*

2010-10 Rebecca Ong (UL), *Mobile Communication and Protection of Children*

2010-09 Hugo Kielman (UL), *A Politiele gegevensverwerking en Privacy, Naar een effectieve waarborging*

2010-08 Krzysztof Siewicz (UL), *Towards an Improved Regulatory Framework of Free Software. Protecting user freedoms in a world of software communities and eGovernments*

2010-07 Wim Fikkert (UT), *Gesture Interaction at a Distance*

2010-06 Sander Bakkes (UvT), *Rapid Adaptation of Video Game AI*

2010-05 Claudia Hauff (UT), *Predicting the Effectiveness of Queries and Retrieval Systems*

2010-04 Olga Kulyk (UT), *Do You Know What I Know? Situational Awareness of Co-located Teams in Multidisplay Environments*

2010-03 Joost Geurts (CWI), *A Document Engineering Model and Processing Framework for Multimedia documents*

2010-02 Ingo Wassink (UT), *Work flows in Life Science*

2010-01 Matthijs van Leeuwen (UU), *Patterns that Matter*

2009-46 Loredana Afanasiev (UvA), *Querying XML: Benchmarks and Recursion*

2009-45 Jilles Vreeken (UU), *Making Pattern Mining Useful*

2009-44 Roberto Santana Tapia (UT), *Assessing Business-IT Alignment in Networked Organizations*

2009-43 Virginia Nunes Leal Franqueira (UT), *Finding Multi-step Attacks in Computer Networks using Heuristic Search and Mobile Ambients*

2009-42 Toine Bogers, *Recommender Systems for Social Bookmarking*

2009-41 Igor Berezhnyy (UvT), *Digital Analysis of Paintings*

2009-40 Stephan Raaijmakers (UvT), *Multinomial Language Learning: Investigations into the Geometry of Language*

2009-39 Christian Stahl (TUE, Humboldt–Universitaet zu Berlin), *Service Substitution – A Behavioral Approach Based on Petri Nets*

2009-38 Riina Vuorikari (OU), *Tags and self-organisation: a metadata ecology for learning resources in a multilingual context*

2009-37 Hendrik Drachsler (OUN), *Navigation Support for Learners in Informal Learning Networks*

2009-36 Marco Kalz (OUN), *Placement Support for Learners in Learning Networks*

2009-35 Wouter Koelewijn (UL), *Privacy en Politiegegevens; Over geautomatiseerde normatieve informatie-uitwisseling*

2009-34 Inge van de Weerd (UU), *Advancing in Software Product Management: An Incremental Method Engineering Approach*

2009-33 Khiet Truong (UT), *How Does Real Affect Affect Affect Recognition In Speech?*

2009-32 Rik Farenhorst (VU) and Remco de Boer (VU), *Architectural Knowledge Management: Supporting Architects and Auditors*

2009-31 Sofiya Katrenko (UVA), *A Closer Look at Learning Relations from Text*

2009-30 Marcin Zukowski (CWI), *Balancing vectorized query execution with bandwidth-optimized storage*

2009-29 Stanislav Pokraev (UT), *Model-Driven Semantic Integration of Service-Oriented Applications*

2009-28 Sander Evers (UT), *Sensor Data Management with Probabilistic Models*

2009-27 Christian Glahn (OU), *Contextual Support of social Engagement and Reflection on the Web*

2009-26 Fernando Koch (UU), *An Agent-Based Model for the Development of Intelligent Mobile Services*

2009-25 Alex van Ballegooij (CWI), *RAM: Array Database Management through Relational Mapping*

2009-24 Annerieke Heuvelink (VUA), *Cognitive Models for Training Simulations*

2009-23 Peter Hofgesang (VU), *Modelling Web Usage in a Changing Environment*

2009-22 Pavel Serdyukov (UT), *Search For Expertise: Going beyond direct evidence*

2009-21 Stijn Vanderlooy (UM), *Ranking and Reliable Classification*

2009-20 Bob van der Vecht (UU), *Adjustable Autonomy: Controling Influences on Decision Making*

2009-19 Valentin Robu (CWI), *Modeling Preferences, Strategic Reasoning and Collaboration in Agent-Mediated Electronic Markets*

2009-18 Fabian Groffen (CWI), *Armada, An Evolving Database System*

2009-17 Laurens van der Maaten (UvT), *Feature Extraction from Visual Data*

2009-16 Fritz Reul (UvT), *New Architectures in Computer Chess*

2009-15 Rinke Hoekstra (UVA), *Ontology Representation - Design Patterns and Ontologies that Make Sense*

2009-14 Maksym Korotkiy (VU), *From ontology-enabled services to service-enabled ontologies (making ontologies work in e-science with ONTO-SOA)*

2009-13 Steven de Jong (UM), *Fairness in Multi-Agent Systems*

2009-12 Peter Massuthe (TUE, Humboldt–Universitaet zu Berlin), *perating Guidelines for Services*

2009-11 Alexander Boer (UVA), *Legal Theory, Sources of Law and the Semantic Web*

2009-10 Jan Wielemaker (UVA), *Logic programming for knowledge-intensive interactive applications*

2009-09 Benjamin Kanagwa (RUN), *Design, Discovery and Construction of Service-oriented Systems*

2009-08 Volker Nannen (VU), *Evolutionary Agent-Based Policy Analysis in Dynamic Environments*

2009-07 Ronald Poppe (UT), *Discriminative Vision-Based Recovery and Recognition of Human Motion*

2009-06 Muhammad Subianto (UU), *Understanding Classification*

2009-05 Sietse Overbeek (RUN), *Bridging Supply and Demand for Knowledge Intensive Tasks - Based on Knowledge, Cognition, and Quality*

2009-04 Josephine Nabukenya (RUN), *Improving the Quality of Organisational Policy Making using Collaboration Engineering*

2009-03 Hans Stol (UvT), *A Framework for Evidence-based Policy Making Using IT*

2009-02 Willem Robert van Hage (VU), *Evaluating Ontology-Alignment Techniques*

2009-01 Rasa Jurgelenaite (RUN), *Symmetric Causal Independence Models*
2008-35 Ben Torben Nielsen (UvT), *Dendritic morphologies: function shapes structure*
2008-34 Jeroen de Knijf (UU), *Studies in Frequent Tree Mining*
2008-33 Frank Terpstra (UVA), *Scientific Workflow Design; theoretical and practical issues*
2008-32 Trung H. Bui (UT), *Toward Affective Dialogue Management using Partially Observable Markov Decision Processes*
2008-31 Loes Braun (UM), *Pro-Active Medical Information Retrieval*
2008-30 Wouter van Atteveldt (VU), *Semantic Network Analysis: Techniques for Extracting, Representing and Querying Media Content*
2008-29 Dennis Reidsma (UT), *Annotations and Subjective Machines – Of Annotators, Embodied Agents, Users, and Other Humans*
2008-28 Ildiko Flesch (RUN), *On the Use of Independence Relations in Bayesian Networks*
2008-27 Hubert Vogten (OU), *Design and Implementation Strategies for IMS Learning Design*
2008-26 Marijn Huijbregts (UT), *Segmentation, Diarization and Speech Transcription: Surprise Data Unraveled*
2008-25 Geert Jonker (UU), *Efficient and Equitable Exchange in Air Traffic Management Plan Repair using Spender-signed Currency*
2008-24 Zharko Aleksovski (VU), *Using background knowledge in ontology matching*
2008-23 Stefan Visscher (UU), *Bayesian network models for the management of ventilator-associated pneumonia*
2008-22 Henk Koning (UU), *Communication of IT-Architecture*
2008-21 Krisztian Balog (UVA), *People Search in the Enterprise*
2008-20 Rex Arendsen (UVA), *Geen bericht, goed bericht. Een onderzoek naar de effecten van de introductie van elektronisch berichtenverkeer met de overheid op de administratieve lasten van bedrijven.*
2008-19 Henning Rode (UT), *From Document to Entity Retrieval: Improving Precision and Performance of Focused Text Search*
2008-18 Guido de Croon (UM), *Adaptive Active Vision*
2008-17 Martin Op 't Land (TUD), *Applying Architecture and Ontology to the Splitting and Allying of Enterprises*
2008-16 Henriëtte van Vugt (VU), *Embodied agents from a user's perspective*
2008-15 Martijn van Otterlo (UT), *The Logic of Adaptive Behavior: Knowledge Representation and Algorithms for the Markov Decision Process Framework in First-Order Domains.*
2008-14 Arthur van Bunningen (UT), *Context-Aware Querying; Better Answers with Less Effort*
2008-13 Caterina Carraciolo (UVA), *Topic Driven Access to Scientific Handbooks*
2008-12 József Farkas (RUN), *A Semiotically Oriented Cognitive Model of Knowledge Representation*
2008-11 Vera Kartseva (VU), *Designing Controls for Network Organizations: A Value-Based Approach*
2008-10 Wauter Bosma (UT), *Discourse oriented summarization*
2008-09 Christof van Nimwegen (UU), *The paradox of the guided user: assistance can be counter-effective*
2008-08 Janneke Bolt (UU), *Bayesian Networks: Aspects of Approximate Inference*
2008-07 Peter van Rosmalen (OU), *Supporting the tutor in the design and support of adaptive e-learning*
2008-06 Arjen Hommersom (RUN), *On the Application of Formal Methods to Clinical Guidelines, an Artificial Intelligence Perspective*
2008-05 Bela Mutschler (UT), *Modeling and simulating causal dependencies on process-aware information systems from a cost perspective*
2008-04 Ander de Keijzer (UT), *Management of Uncertain Data – towards unattended integration*
2008-03 Vera Hollink (UVA), *Optimizing hierarchical menus: a usage-based approach*
2008-02 Alexei Sharpanskykh (VU), *On Computer-Aided Methods for Modeling and Analysis of Organizations*
2008-01 Katalin Boer-Sorbán (EUR), *Agent-Based Simulation of Financial Markets: A modular, continuous-time approach*
2007-25 Joost Schalken (VU), *Empirical Investigations in Software Process Improvement*
2007-24 Georgina Ramírez Camps (CWI), *Structural Features in XML Retrieval*
2007-23 Peter Barna (TUE), *Specification of Application Logic in Web Information Systems*
2007-22 Zlatko Zlatev (UT), *Goal-oriented design of value and process models from patterns*
2007-21 Karianne Vermaas (UU), *Fast diffusion and broadening use: A research on residential adoption and usage of broadband internet in the Netherlands between 2001 and 2005*
2007-20 Slinger Jansen (UU), *Customer Configuration Updating in a Software Supply Network*
2007-19 David Levy (UM), *Intimate relationships with artificial partners*
2007-18 Bart Orriëns (UvT), *On the development an management of adaptive business collaborations*
2007-17 Theodore Charitos (UU), *Reasoning with Dynamic Networks in Practice*
2007-16 Davide Grossi (UU), *Designing Invisible Handcuffs. Formal investigations in Institutions and Organizations for Multi-agent Systems*
2007-15 Joyca Lacroix (UM), *NIM: a Situated Computational Memory Model*
2007-14 Niek Bergboer (UM), *Context-Based Image Analysis*
2007-13 Rutger Rienks (UT), *Meetings in Smart Environments; Implications of Progressing Technology*
2007-12 Marcel van Gerven (RUN), *Bayesian Networks for Clinical Decision Support: A Rational Approach to Dynamic Decision-Making under Uncertainty*
2007-11 Natalia Stash (TUE), *Incorporating Cognitive/Learning Styles in a General-Purpose Adaptive Hypermedia System*
2007-10 Huib Aldewereld (UU), *Autonomy vs. Conformity: an Institutional Perspective on Norms and Protocols*
2007-09 David Mobach (VU), *Agent-Based Mediated Service Negotiation*
2007-08 Mark Hoogendoorn (VU), *Modeling of Change in Multi-Agent Organizations*
2007-07 Nataša Jovanović (UT), *To Whom It May Concern – Addressee Identification in Face-to-Face Meetings*
2007-06 Gilad Mishne (UVA), *Applied Text Analytics for Blogs*
2007-05 Bart Schermer (UL), *Software Agents, Surveillance, and the Right to Privacy: a Legislative Framework for Agent-enabled Surveillance*
2007-04 Jurriaan van Diggelen (UU), *Achieving Semantic Interoperability in Multi-agent Systems: a dialogue-based approach*
2007-03 Peter Mika (VU), *Social Networks and the Semantic Web*
2007-02 Wouter Teepe (RUG), *Reconciling Information Exchange and Confidentiality: A Formal Approach*
2007-01 Kees Leune (UvT), *Access Control and Service-Oriented Architectures*
2006-28 Börkur Sigurbjörnsson (UVA), *Focused Information Access using XML Element Retrieval*

2006-27 Stefano Bocconi (CWI), *Vox Populi: generating video documentaries from semantically annotated media repositories*
2006-26 Vojkan Mihajlović (UT), *Score Region Algebra: A Flexible Framework for Structured Information Retrieval*
2006-25 Madalina Drugan (UU), *Conditional log-likelihood MDL and Evolutionary MCMC*
2006-24 Laura Hollink (VU), *Semantic Annotation for Retrieval of Visual Resources*
2006-23 Ion Juvina (UU), *Development of Cognitive Model for Navigating on the Web*
2006-22 Paul de Vrieze (RUN), *Fundaments of Adaptive Personalisation*
2006-21 Bas van Gils (RUN), *Aptness on the Web*
2006-20 Marina Velikova (UvT), *Monotone models for prediction in data mining*
2006-19 Birna van Riemsdijk (UU), *Cognitive Agent Programming: A Semantic Approach*
2006-18 Valentin Zhizhkun (UVA), *Graph transformation for Natural Language Processing*
2006-17 Stacey Nagata (UU), *User Assistance for Multitasking with Interruptions on a Mobile Device*
2006-16 Carsten Riggelsen (UU), *Approximation Methods for Efficient Learning of Bayesian Networks*
2006-15 Rainer Malik (UU), *CONAN: Text Mining in the Biomedical Domain*
2006-14 Johan Hoorn (VU), *Software Requirements: Update, Upgrade, Redesign – towards a Theory of Requirements Change*
2006-13 Henk-Jan Lebbink (UU), *Dialogue and Decision Games for Information Exchanging Agents*
2006-12 Bert Bongers (VU), *Interactivation – Towards an e-cology of people, our technological environment, and the arts*
2006-11 Joeri van Ruth (UT), *Flattening Queries over Nested Data Types*
2006-10 Ronny Siebes (VU), *Semantic Routing in Peer-to-Peer Systems*
2006-09 Mohamed Wahdan (UM), *Automatic Formulation of the Auditor's Opinion*
2006-08 Eelco Herder (UT), *Forward, Back and Home Again – Analyzing User Behavior on the Web*
2006-07 Marko Smiljanic (UT), *XML schema matching – balancing efficiency and effectiveness by means of clustering*
2006-06 Ziv Baida (VU), *Software-aided Service Bundling – Intelligent Methods & Tools for Graphical Service Modeling*
2006-05 Cees Pierik (UU), *Validation Techniques for Object-Oriented Proof Outlines*
2006-04 Marta Sabou (VU), *Building Web Service Ontologies*
2006-03 Noor Christoph (UVA), *The role of metacognitive skills in learning to solve problems*
2006-02 Cristina Chisalita (VU), *Contextual issues in the design and use of information technology in organizations*
2006-01 Samuil Angelov (TUE), *Foundations of B2B Electronic Contracting*
2005-21 Wijnand Derks (UT), *Improving Concurrency and Recovery in Database Systems by Exploiting Application Semantics*
2005-20 Cristina Coteanu (UL), *Cyber Consumer Law, State of the Art and Perspectives*
2005-19 Michel van Dartel (UM), *Situated Representation*
2005-18 Danielle Sent (UU), *Test-selection strategies for probabilistic networks*
2005-17 Boris Shishkov (TUD), *Software Specification Based on Re-usable Business Components*
2005-16 Joris Graaumans (UU), *Usability of XML Query Languages*
2005-15 Tibor Bosse (VU), *Analysis of the Dynamics of Cognitive Processes*
2005-14 Borys Omelayenko (VU), *Web-Service configuration on the Semantic Web; Exploring how semantics meets pragmatics*
2005-13 Fred Hamburg (UL), *Een Computermodel voor het Ondersteunen van Euthanasiebeslissingen*
2005-12 Csaba Boer (EUR), *Distributed Simulation in Industry*
2005-11 Elth Ogston (VU), *Agent Based Matchmaking and Clustering – A Decentralized Approach to Search*
2005-10 Anders Bouwer (UVA), *Explaining Behaviour: Using Qualitative Simulation in Interactive Learning Environments*
2005-09 Jeen Broekstra (VU), *Storage, Querying and Inferencing for Semantic Web Languages*
2005-08 Richard Vdovjak (TUE), *A Model-driven Approach for Building Distributed Ontology-based Web Applications*
2005-07 Flavius Frasincar (TUE), *Hypermedia Presentation Generation for Semantic Web Information Systems*
2005-06 Pieter Spronck (UM), *Adaptive Game AI*
2005-05 Gabriel Infante-Lopez (UVA), *Two-Level Probabilistic Grammars for Natural Language Parsing*
2005-04 Nirvana Meratnia (UT), *Towards Database Support for Moving Object data*
2005-03 Franc Grootjen (RUN), *A Pragmatic Approach to the Conceptualisation of Language*
2005-02 Erik van der Werf (UM)), *AI techniques for the game of Go*
2005-01 Floor Verdenius (UVA), *Methodological Aspects of Designing Induction-Based Applications*
2004-20 Madelon Evers (Nyenrode), *Learning from Design: facilitating multidisciplinary design teams*
2004-19 Thijs Westerveld (UT), *Using generative probabilistic models for multimedia retrieval*
2004-18 Vania Bessa Machado (UvA), *Supporting the Construction of Qualitative Knowledge Models*
2004-17 Mark Winands (UM), *Informed Search in Complex Games*
2004-16 Federico Divina (VU), *Hybrid Genetic Relational Search for Inductive Learning*
2004-15 Arno Knobbe (UU), *Multi-Relational Data Mining*
2004-14 Paul Harrenstein (UU), *Logic in Conflict. Logical Explorations in Strategic Equilibrium*
2004-13 Wojciech Jamroga (UT), *Using Multiple Models of Reality: On Agents who Know how to Play*
2004-12 The Duy Bui (UT), *Creating emotions and facial expressions for embodied agents*
2004-11 Michel Klein (VU), *Change Management for Distributed Ontologies*
2004-10 Suzanne Kabel (UVA), *Knowledge-rich indexing of learning-objects*
2004-09 Martin Caminada (VU), *For the Sake of the Argument; explorations into argument-based reasoning*
2004-08 Joop Verbeek (UM), *Politie en de Nieuwe Internationale Informatiemarkt, Grensregionale politile gegevensuitwisseling en digitale expertise*
2004-07 Elise Boltjes (UM), *Voorbeeldig onderwijs; voorbeeldgestuurd onderwijs, een opstap naar abstract denken, vooral voor meisjes*
2004-06 Bart-Jan Hommes (TUD), *The Evaluation of Business Process Modeling Techniques*
2004-05 Viara Popova (EUR), *Knowledge discovery and monotonicity*
2004-04 Chris van Aart (UVA), *Organizational Principles for Multi-Agent Architectures*
2004-03 Perry Groot (VU), *A Theoretical and Empirical Analysis of Approximation in Symbolic Problem Solving*
2004-02 Lai Xu (UvT), *Monitoring Multi-party Contracts for E-business*

2004-01 Virginia Dignum (UU), *A Model for Organizational Interaction: Based on Agents, Founded in Logic*
2003-18 Levente Kocsis (UM), *Learning Search Decisions*
2003-17 David Jansen (UT), *Extensions of Statecharts with Probability, Time, and Stochastic Timing*
2003-16 Menzo Windhouwer (CWI), *Feature Grammar Systems – Incremental Maintenance of Indexes to Digital Media Warehouses*
2003-15 Mathijs de Weerdt (TUD), *Plan Merging in Multi-Agent Systems*
2003-14 Stijn Hoppenbrouwers (KUN), *Freezing Language: Conceptualisation Processes across ICT-Supported Organisations*
2003-13 Jeroen Donkers (UM), *Nosce Hostem – Searching with Opponent Models*
2003-12 Roeland Ordelman (UT), *Dutch speech recognition in multimedia information retrieval*
2003-11 Simon Keizer (UT), *Reasoning under Uncertainty in Natural Language Dialogue using Bayesian Networks*
2003-10 Andreas Lincke (UvT), *Electronic Business Negotiation: Some experimental studies on the interaction between medium, innovation context and culture*
2003-09 Rens Kortmann (UM), *The resolution of visually guided behaviour*
2003-08 Yongping Ran (UM), *Repair Based Scheduling*
2003-07 Machiel Jansen (UvA), *Formal Explorations of Knowledge Intensive Tasks*
2003-06 Boris van Schooten (UT), *Development and specification of virtual environments*
2003-05 Jos Lehmann (UVA), *Causation in Artificial Intelligence and Law – A modelling approach*
2003-04 Milan Petković (UT), *Content-Based Video Retrieval Supported by Database Technology*
2003-03 Martijn Schuemie (TUD), *Human-Computer Interaction and Presence in Virtual Reality Exposure Therapy*
2003-02 Jan Broersen (VU), *Modal Action Logics for Reasoning About Reactive Systems*
2003-01 Heiner Stuckenschmidt (VU), *Ontology-Based Information Sharing in Weakly Structured Environments*
2002-17 Stefan Manegold (UVA), *Understanding, Modeling, and Improving Main-Memory Database Performance*
2002-16 Pieter van Langen (VU), *The Anatomy of Design: Foundations, Models and Applications*
2002-15 Rik Eshuis (UT), *Semantics and Verification of UML Activity Diagrams for Workflow Modelling*
2002-14 Wieke de Vries (UU), *Agent Interaction: Abstract Approaches to Modelling, Programming and Verifying Multi-Agent Systems*
2002-13 Hongjing Wu (TUE), *A Reference Architecture for Adaptive Hypermedia Applications*
2002-12 Albrecht Schmidt (Uva), *Processing XML in Database Systems*
2002-11 Wouter C.A. Wijngaards (VU), *Agent Based Modelling of Dynamics: Biological and Organisational Applications*
2002-10 Brian Sheppard (UM), *Towards Perfect Play of Scrabble*
2002-09 Willem-Jan van den Heuvel (KUB), *Integrating Modern Business Applications with Objectified Legacy Systems*
2002-08 Jaap Gordijn (VU), *Value Based Requirements Engineering: Exploring Innovative E-Commerce Ideas*
2002-07 Peter Boncz (CWI), *Monet: A Next-Generation DBMS Kernel For Query-Intensive Applications*
2002-06 Laurens Mommers (UL), *Applied legal epistemology; Building a knowledge-based ontology of the legal domain*
2002-05 Radu Serban (VU), *The Private Cyberspace Modeling Electronic Environments inhabited by Privacy-concerned Agents*
2002-04 Juan Roberto Castelo Valdueza (UU), *The Discrete Acyclic Digraph Markov Model in Data Mining*
2002-03 Henk Ernst Blok (UT), *Database Optimization Aspects for Information Retrieval*
2002-02 Roelof van Zwol (UT), *Modelling and searching web-based document collections*
2002-01 Nico Lassing (VU), *Architecture-Level Modifiability Analysis*
2001-11 Tom M. van Engers (VUA), *Knowledge Management: The Role of Mental Models in Business Systems Design*
2001-10 Maarten Sierhuis (UvA), *Modeling and Simulating Work Practice BRAHMS: a multiagent modeling and simulation language for work practice analysis and design*
2001-09 Pieter Jan 't Hoen (RUL), *Towards Distributed Development of Large Object-Oriented Models, Views of Packages as Classes*
2001-08 Pascal van Eck (VU), *A Compositional Semantic Structure for Multi-Agent Systems Dynamics.*
2001-07 Bastiaan Schonhage (VU), *Diva: Architectural Perspectives on Information Visualization*
2001-06 Martijn van Welie (VU), *Task-based User Interface Design*
2001-05 Jacco van Ossenbruggen (VU), *Processing Structured Hypermedia: A Matter of Style*
2001-04 Evgueni Smirnov (UM), *Conjunctive and Disjunctive Version Spaces with Instance-Based Boundary Sets*
2001-03 Maarten van Someren (UvA), *Learning as problem solving*
2001-02 Koen Hindriks (UU), *Agent Programming Languages: Programming with Mental Models*
2001-01 Silja Renooij (UU), *Qualitative Approaches to Quantifying Probabilistic Networks*
2000-11 Jonas Karlsson (CWI), *Scalable Distributed Data Structures for Database Management*
2000-10 Niels Nes (CWI), *Image Database Management System Design Considerations, Algorithms and Architecture*
2000-09 Florian Waas (CWI), *Principles of Probabilistic Query Optimization*
2000-08 Veerle Coupé (EUR), *Sensitivity Analyis of Decision-Theoretic Networks*
2000-07 Niels Peek (UU), *Decision-theoretic Planning of Clinical Patient Management*
2000-06 Rogier van Eijk (UU), *Programming Languages for Agent Communication*
2000-05 Ruud van der Pol (UM), *Knowledge-based Query Formulation in Information Retrieval.*
2000-04 Geert de Haan (VU), *ETAG, A Formal Model of Competence Knowledge for User Interface Design*
2000-03 Carolien M.T. Metselaar (UVA), *Sociaal-organisatorische gevolgen van kennistechnologie; een procesbenadering en actorperspectief.*
2000-02 Koen Holtman (TUE), *Prototyping of CMS Storage Management*
2000-01 Frank Niessink (VU), *Perspectives on Improving Software Maintenance*
1999-08 Jacques H.J. Lenting (UM), *Informed Gambling: Conception and Analysis of a Multi-Agent Mechanism for Discrete Reallocation.*
1999-07 David Spelt (UT), *Verification support for object database design*
1999-06 Niek J.E. Wijngaards (VU), *Re-design of compositional systems*
1999-05 Aldo de Moor (KUB), *Empowering Communities: A Method for the Legitimate User-Driven Specification of Network Information Systems*

1999-04 Jacques Penders (UM), *The practical Art of Moving Physical Objects*
1999-03 Don Beal (UM), *The Nature of Minimax Search*
1999-02 Rob Potharst (EUR), *Classification using decision trees and neural nets*
1999-01 Mark Sloof (VU), *Physiology of Quality Change Modelling; Automated modelling of Quality Change of Agricultural Products*
1998-05 E.W. Oskamp (RUL), *Computerondersteuning bij Straftoemeting*
1998-04 Dennis Breuker (UM), *Memory versus Search in Games*
1998-03 Ans Steuten (TUD), *A Contribution to the Linguistic Analysis of Business Conversations within the Language/Action Perspective*
1998-02 Floris Wiesman (UM), *Information Retrieval by Graphically Browsing Meta-Information*
1998-01 Johan van den Akker (CWI), *DEGAS – An Active, Temporal Database of Autonomous Objects*

QR code for downloading
the electronic version of this dissertation

CTIT Dissertation No. 10-186
ISBN: 978-90-365-3116-0