universität wien

MASTERARBEIT

Titel der Masterarbeit

„Evaluating spatially directing cues on a wearable user interface in a field setup“

Verfasserin

Reea Mynttinen, BBA

angestrebter akademischer Grad

Master of Science, MSc

Wien, 2010

Studienkennzahl lt. Studienblatt:	A 066 013
Studienrichtung lt. Studienblatt:	Joint Degree Programme MEi:CogSci Cognitive Science Magisterstudium/Masterstudium
Betreuerin / Betreuer:	MD, PhD, Research Professor Kiti Müller. Brain and Work Research Centre, Finnish Institute of Occupational Health. Adjunct Professor at Aalto University

ABSTRACT

This thesis studied the use of different modality cues in directing spatial attention and action in a cognitively demanding real world task.

Wearable user interfaces are likely to be used outdoors and in constantly changing circumstances. Their evaluation in similar conditions is important because data collected in a laboratory setting may not be ecologically valid. Therefore, an applied real world setting was chosen.

The task here was a shooting task at a military firing range, simulating a time critical crisis management task, but the results can be utilized also more widely. The subjects were instructed to shoot at one of three targets by stimulus cues presented in three different modalities (visual, auditory and tactile stimuli), and their multimodal combinations. All conditions were tested with and without physical exercise, simulating battle stress.

The different modality cues were investigated based on performance (reaction time, shooting accuracy and speed-accuracy trade-off), and subjective measurements (questionnaires and a structured interview). Physical effort was controlled with electrocardiogram (heart rate) and NASA-Task Load Index (estimates of subjective workload)

The results do not clearly put modality variations in a ranking order, neither by performance nor by subjective measures. No statistical effects were found between performance with different modality cues or their combinations. However, a trend pointing to the performance degrading due to physical exercise was observed. The subjects had preferences among modalities, but these differed between subjects. Each modality had some experienced strengths and weaknesses.

Based on the results, the design of cross- and multimodal cuing on wearable UIs should concentrate on reliability and detectability of the cues in the target environment. Task requirements should be considered carefully, rather than trying to define which modality leads to fastest or most accurate performance. User evaluation in a field setup is essential, and the performance measures can be used to indicate possible problems, such as cues that were not detected.

ZUSAMMENFASSUNG

Die vorliegende Studie untersucht, wie räumliche Aufmerksamkeit und Handlungsfähigkeit in einer anspruchsvollen und anwendungsnahen Aufgabe durch kreuz- und multimodale Stimuli gesteuert werden kann.

Tragbare Mensch-Maschinen-Schnittstellen (wearable user interfaces) werden vermutlich vorwiegend im Freien benützt, wo sich Umstände und Umgebung ständig ändern können. Evaluationen sollten unter entsprechenden Umständen erfolgen, weil Daten, die im Labor erhoben werden, weniger ökologische Aussagekraft haben können. Für die vorliegende Studie wurde daher ein angewandter Feldversuch unternommen.

Auf einem Militärschießplatz wurde eine Schießübung durchgeführt, die eine zeitkritische Krisenmanagementaufgabe simuliert. Die Resultate sind aber auch außerhalb des militärischen Kontext nutzbar. Die Probanden wurden gebeten nach bestimmten Stimuli, auf eines von drei Zielen zu schießen. Die Stimuli gaben die Richtung des Zieles vor und waren entweder visuell, auditiv, taktil oder multimodale Kombinationen dieser Stimuli. Alle diese Variationen wurden mit und ohne körpelichen Belastungsübungen durchgeführt, um die tatsächlichen Umstände eines Gefechts zu simulieren.

Die verschiedenen Reiz-variationen wurden hinsichtlich Performanz (Reaktionszeit, Schussgenauigkeit, deren Abstimmung) und subjektiver Evaluierung (Fragebogen, Leitfaden-Interview) untersucht. Die Belastung durch die körperliche Übung wurde mittels EKG und NASA-TLX kontrolliert und evaluiert.

Aufgrund der gleichwertigen Resultate kann man die verschiedenen Modalität-Variationen nicht in Rangfolge setzen weder hinsichtlich Performanz, noch hinsichtlich subjektiver Evaluierung. Die unterschiedlichen Modalität-Variationen ergaben keine signifikante statistische Evidenz hinsichtlich der Performanz. Es konnte lediglich festgestellt werden, dass die Performanz nach der Belastungsübung tendenziell abnahm. Die Probanden hatten individuelle Präferenzen unter den Modalitäten. Für jede Modalität wurden Stärken und Schwächen erlebt.

Die Ergebnisse legen nahe, dass sich das Design einer kreuz- und multimodalen Aufmerksamkeitssteuerung mit einer tragbaren Mensch-Maschine-Schnittstelle sich auf die Nachweisbarkeit und Verlässlichkeit der Stimuli konzentrieren sollte. Man sollte die Aufgabenanforderungen und die Umstände sorgfältig berücksichtigen, anstatt zu bestimmen versuchen welche Modalität in der Schiessaufgabe die Reaktionszeit am meisten beschleunigt oder die Genauigkeit bei derselben verbessert. Die Nutzer-Evaluation in einem angewandten Feldversuch ist essenziell, und die Performanzmessungen können benützt werden, um auf mögliche Probleme hinzuweisen, wie z.B. welche Anreize nicht bemerkt wurden.

AUTHOR'S CONTRIBUTION

This study is the result of a group effort. I was responsible for overall planning of the experiment(s) and coordinating the plans with the military contact person Juha Väisänen.

The study setup was initially based on a laboratory study which I made together with Risto Näsänen (Näsänen et al. 2008). I designed the field setup based on my literature review in co-operation with Kristian Lukander. The task was designed based on information collected from military contact persons. I designed the experiment; the detailed field setup, the tasks, the detailed time plan, the procedures and the stimuli.

I implemented the HMD visual stimuli and the auditory stimuli. I evaluated all stimuli in the laboratory, with the participation of several team members. I helped designing and testing the stimulus control program and prepared the stimulus configurations for the experiment sessions. I designed all the questionnaires, the interviews, and prepared the related material for the field setup. I lead the experiments and instructed the subjects in the experiment. I collected and prepared the raw data for the analysis and analysed performance and subjective measurements myself. Virpi Kalakoski and Kati Hirvonen mentored me with statistical methods. I also presented the study in several SAWUI meetings.

Altogether, five persons were needed for all sessions to run the experiment. The wearable user interface devices and the software used for the measurements were developed and maintained by other team members. Led-light goggles and tactor belt and related stimuli were prepared by Kristian Lukander, auditory setup and audio recording program by Sharman Jagadeesan. Kristian Lukander also implemented the stimulus control program. The ECG measurement data preparation and analysis was done by Andreas Henelius.

ACKNOWLEDGEMENTS

This thesis was written at the Finnish Institute of Occupational Health (FIOH), where it was accepted as part of research activities. I would like to thank Kiti Müller, the leader of the Brain and Work Research Centre for offering me the opportunity to work in the centre, and for accepting to be an external supervisor of this thesis for the University of Vienna. This open-minded flexibility from Kiti, FIOH and the middle-European study program/University of Vienna allowed me to have this challenging but rewarding experience to live, work and study cross-borders; being part of a multidisciplinary research team in Finland, and a student in Austria at the same time. The year spent with this study has included lots of travelling, many languages spoken daily, and on the other hand many hours alone at home in front of a computer writing and studying for this thesis. I felt always welcome to the Brain and Work Research Centre after having been moths away, working in distance for this project. It was great to be part of this team.

This thesis would not have been possible without support from the team members in Brain and Work Research Centre. My greatest gratitude goes to my instructor and mentor Kristian Lukander. He inspired me in doing this research, gave guidance and in all an enormous contribution for this thesis. Many colleagues shared their knowledge, gave constructive comments and discussed the research. I want to thank Risto Näsänen for introducing me the world of psychophysical experiments, Virpi Kalakoski, Andreas Henelius, Kati Hirvonen, and Matti Gröhn for their guidance and comments, Teppo Valtonen, Lauri Ahonen, Sharman Jagadeesan and Kristian Lukander for running the experiments with me; the seamless co-operation in the field experiment was amazing. Of course this project would not have been possible without close co-operation with Juha Väisänen who arranged the military setup and recruited the subjects. I also want to thank the subjects for participating in the experiment.

My thanks also go to my student colleagues and friends, who were there. My parents, who offered a home to stay when I was in Finland and encouraged me. My warmest thanks go to my loved one Sylvain Séné, who continuously supported me during this project, and never complained about my late hour's work.

Helsinki, March 16, 2010

Reea Mynttinen

Table of Contents

LIST OF APPENDICES

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

AU	Auditory
AC	Accuracy
ECG	Electrocardiogram
FIOH	Finnish Institute of Occupational Health
HMD	Head mounted display
MIX	Mixed modal task
MULTI	Multimodal
N	Normal, task condition without physical exercise
NASA-TLX	NASA-Task Load Index
R	Task condition with physical exercise
RAU	Auditory modality variation with physical exercise
RMULTI	Multimodal modality variation with physical exercise
RMIX	Mixed modality variation with physical exercise
RTA	Tactile modality variation with physical exercise
RVI	Visual modality variation with physical exercise
RT	Reaction time
UI	User interface
SA	Speed-Accuracy
SAWUI	Supporting situation awareness in demanding operating environments through wearable interfaces
TA	Tactile
TLX	Task Load Index
VI	Visual
QU	Questionnaire

1. INTRODUCTION

1.1. Evaluating spatially directing cues on a wearable user interface in a field setup

The ability to detect relevant information in a given environment is a prerequisite for situational awareness. People cannot attend intensively to several things at the same time because the speed and the number of the mental tasks that a human can execute at a time is limited and therefore the brain filters and selects information in many ways before it is actually processed. (Eysenck & Keane 2005) In order to enhance capturing attention, stimulus cues may be used to direct attention to something important in user interfaces or in the environment. The shift of attention in one modality to a certain direction leads to directing attention of other modalities there as well and may speed up and ensure processing of the target stimulus (Spence & Gallace 2007) Such stimulus cues can be presented with other senses than the one in which the target is presented (Spence & Gallace 2007)(Ferris & Sarter 2008).

Situation awareness is an important factor in highly demanding professions such as crisis management in military operations, or fire and rescue services (Laarni et al. 2009). They require quick actions and errors may have fatal consequences (Burke et al. 2004). However, hard physical effort and stress deteriorate human perception and attention capacities and physical and psychological capacities are close to reaching their limits (Hancock & Szalma 2008).

Well-designed information systems are needed to improve human performance and situation awareness in such demanding situations (Laarni et al. 2009). Technological advances, such as wearable multisensory user interfaces enable new ways to present information to humans.

A wearable user interface provides an interaction channel between humans and computers via devices attached on the body and on the clothes. Ideally, a wearable user interface enables people to move around freely without need for any other tools for interaction. They can thus carry out their main task while the wearable user interface supports them by providing information about the environment, task etc. (Benyon et al. 2005). A wearable user interface should not disturb the actual task. It should enable observing the environment, and in all means enhance the situational awareness without adding the workload.

For example, for a fire fighter carrying a handheld computer device is not possible because he needs his both hands for rescuing activities. Visibility in a building filled with smoke may be strongly impaired. However, the fire fighter could receive different kind of information helping e.g. orientation via a head mounted display, headphones or from vibration belt.

When designing such new user interfaces it is important to carefully design and evaluate how people perform with it, and how they experience it (Benyon et al. 2005). This document does not discuss the technical issues on how to implement the wearable devices, but concentrates on human factors, and the basic sensory functions of humans. As wearable user interfaces are likely to be used outdoors or in constantly changing circumstances, they need to be evaluated in similar conditions as well (Benyon et al. 2005)(Warm et al. 2008). Therefore, an applied real world setting was chosen for this study. The task was a shooting task performed in a military practice firing range. Shooting is a stressful field test situation that this study uses to simulate a crisis management task. The speed of action required is an important aspect of the experiment: how time critical information can be presented so that it is detected but is not disturbing the task. The current setting was also motivated by the need to develop field testing methods.

The approach in this study is applied and multidisciplinary: An experimental psychology setup moved from a strict laboratory setup to a real environment military setup, combined with novel wearable user interface technology.

The application area in this study is military but the results can also be utilized more widely when developing wearable interface technology which aim to enhance human performance; e.g. for fire and rescue services or extreme sports.

This master thesis study has been accepted as part of research activities in the Finnish Institute of Occupational Health and the results will be reported in research project SAWUI ("Supporting situation awareness in demanding operating environments through wearable interfaces")(Laarni et al. 2009) which is done in co-operation with several other research institutes and companies in Finland. The SAWUI project is researching technologies to develop wearable user interfaces. The project is funded by the Finnish Funding Agency for Technology and Innovation, Tekes.

1.2. Aims of the study and research questions

This thesis addresses human factors related to the development of new wearable user interface technologies, more specifically how well different modality cues can be used for supporting a spatially directing task, and how physical effort affects them in a field setup. In this thesis visual, auditory and tactile cue stimuli and multimodal stimuli combinations are studied. These cues are presented on the body on multisensory wearable user interface devices.

The main objective of this study is to evaluate spatially directing cues presented on wearable user interfaces in a field setup. More specifically the research questions are:

1. How should spatially directing stimuli be presented with wearable user interface?

A theoretical approach is used to answer this question. Based on literature review, the thesis describes selected applied aspects of basic sensory mechanisms. Basic information of senses, perception and attention in relation to presenting directions on wearable user interfaces are collected. In addition, effects of physical effort and stress are shortly described. The scope is limited to the experimental task, and related stimulus cue requirements. In addition, a review of relevant literature was made also to meet the aims of planning the setup and carrying out the study. The cues used in the experiments and the experiment design are based on these literature reviews.

2. Are there differences between modalities in a field experiment?

Empirical approaches are used to answer this question. The different modality cues are investigated with performance (reaction time and shooting accuracy and analysis of speed-accuracy trade-off) and subjective measurements (questionnaires and structured interview). The effects of physical exercise are investigated with ECG (heart rate) and NASA-Task Load Index (estimates of subjective workload). In addition, the effects of target direction are investigated.

The hypothesis is that there are differences in performance and in subjective preferences between modalities in the field setup depending of the modality in which they are given. A more detailed motivation for the study and its expectations are described in chapter 8.

1.3. Thesis structure

This thesis is divided into following sections:

I Introduction

Chapter one describes shortly the background and the aims for the study.

II Literature review on applied aspects of basic sensory mechanisms

Chapter two describes how cross-modal effects have been used to capture and direct attention, chapter three collects basic information on senses for the applied context of the current study, in relation to presenting directions on wearable user interfaces, and chapter four shortly describes the effects of physical effort.

III Literature review on study design and methods

Chapter five gives an overview of the evaluation of human factors. Chapter six describes in more detail the statistical tests relevant for the study, and chapter seven describes the interviews and questionnaires.

IV The motivations and predictions for this study

Chapter eight summarizes the requirements for stimulus cues in this study, and the objectives of performance and subjective measurements. The chapter also summarizes the predictions from the literature review in section II in relation to the aims of the study and objectives of the measurements.

V The Experiment

Chapter nine describes the experiment in detail, including descriptions of e.g. the task, the wearable user interface and the field setup.

VI Results, Discussion and Conclusions

Chapter ten presents the results for performance and subjective measurements, in relation to the objectives defined in chapters 8.2 and 8.3. Chapter 11 discusses the results in relation to the literature review. Chapter 12 presents the conclusions of the study.

2. CAPTURING AND DIRECTING ATTENTION WITH CROSS-MODAL CUES

2.1. Limits of attention

This chapter describes how attention can be directed with different kinds of cues to enhance task performance. Attention refers to the selectivity of information processing in the brain. The amount of information that is constantly acquired by the senses from the environment is overwhelming and only a small part of it can be analyzed. The speed and the number of the mental tasks that a human can execute at a time is limited and therefore the brain filters and selects information in many ways before it is actually processed. (Eysenck & Keane 2005) Attention selection is affected by various factors such as task demands, situation, motivation and stress (Spence & Gallace 2007).

People can consciously direct their attention to a task and thus be more efficient in it. But the directing of attention can take place also more automatically as a reaction to a sensory stimulus. The direction and selection of attention is thus affected both by external and internal factors. These are also referred to as active and passive modes, or voluntary and automatic attention directing mechanisms (Eysenck & Keane 2005).

The need for enhancing task performance is mostly related to quite limited ability of humans to detect changes in their environment. A short interruption in perceiving a target may result with a change in it remaining undetected. This phenomenon is referred to as change blindness. (Eysenck & Keane 2005) In addition, if people concentrate on a demanding task they may ignore other important events or stimuli from their environment. The more mental resources the task takes, the less attention people allocate to other stimuli. When an individual needs to react quickly to information about a threat, it is necessary to attend rapidly to that information. Therefore, when attention is already focused somewhere else than the target information, the challenge is to capture the attention, as fast as possible.

2.2. Cuing attention

Attention can be directed externally with different kinds of stimuli, i.e. cues. (Spence & Gallace 2007) These stimuli are presented prior to an actual target on any modality. In addition to modality used, the processing of target stimuli is affected by timing and location of the cue, among others (Ferris & Sarter 2008).

2.2.1 Using multisensory and cross-modal effects for cuing attention

Multisensory information processing is seen as an opportunity to support attention management in real-world domains. Different senses react to different kind of stimuli and these several stimuli support each other when they are related to one event. Information presented multimodally (on several modalities at the same time) provides benefits of disambiguation and increased salience. That is, people process such multisensory information more accurately than information just from a single sensory channel. The following chapters describe how multisensory information is perceived and how it can be used for cuing attention.

Perception of multisensory information

Human information processing is naturally multisensory. All senses are used for observing the environment, and many experiences are based on combined information received from different sensory modalities. Natural multisensory stimuli originate from a single location at a certain time and the brain integrates these stimuli to construct the personal experience of perception.

Different senses complement each other, or more specifically, the brain processes all information received from different senses to construct a single interpretation. For example, when listening to a person speaking in a noisy environment, visual information can be important for successful interpretation. The uncertainty of each modality determines which modality has the largest effect on the perception. If a visual stimulus is poor, an auditory cue may be more determining, and vice versa. (King 2009) The possibility of a faulty interpretation of perceptual information is lower when information is multisensory. People process multisensory information more quickly and accurately than the information received only through one sensory channel. Depending on task needs and environment conditions, a particular sense is often better adapted to a task due to the differences of the stimuli the sense reacts to (Ferris & Sarter 2008). Each sense has its strengths and weaknesses. People may also prefer different modality channels for receiving different information.

Visual, auditory and tactile senses each have their specific receptors, which react to stimuli from the environment. The brain then processes the stimuli further. The senses have been traditionally studied separately and previously they were thought to have independent attention processes in the brain. At present, there is evidence that the processing of different senses are partly overlapping and that there are multisensory processes as well. The current view is that sensory integration takes place already during cortical sensory processing, i.e. during the earliest stages of perception. (Pavani et al. 2007) However, information coming from different senses is still processed at least partly separately as each sense has separate receptors and some independent brain processing and attention management. Therefore, when one sense is under heavy load, it might be useful to present further information on other senses.

Cross-modal effects

Several cross-modal effects have been demonstrated as evidence of sensory integration. Facilitating cross-modal effects have been found between visual, auditory and tactile senses. Such cross-sensory links have been shown to be most evident in laboratory studies for a visual target that is supported with an auditory cue (Ferris & Sarter 2008). The chapters 2.2.2 and 2.2.3 present examples of such effects.

Capturing attention with multisensory cues

Multisensory stimuli have been suggested to be more powerful in capturing attention than unimodal cues, especially under other perceptual load (Spence & Gallace 2007)(Santangelo & Spence 2008). Efficiency does not necessarily mean speeding the performance but rather success in shifting the attention away from the demanding perceptual task, in which the person is concentrated. It seems that stimuli given only on a single sensory channel cannot capture attention away from other demanding tasks as efficiently (Santangelo & Spence 2008). In these experiments audiovisual and audiotactile stimuli were used. Multisensory stimuli are suggested to be especially useful for aging people whose perceptual sensitivity is decreased (Laurienti et al. 2006).

2.2.2 Spatial cuing of attention

People can voluntarily direct their spatial attention to a certain location. In addition, an unexpected sensory stimulus can capture attention almost automatically to its direction. The shift of attention in one modality to a certain direction leads to directing attention of other modalities there as well (Spence & Gallace 2007). Therefore, a cue that is presented in a particular location preceding a target improves the detection of a target stimulus at the same location, and thus speeds up the information processing (Ferris & Sarter 2008)(Spence & Gallace 2007).

The use of cues has been shown to support the processing of the target, e.g. Spence and colleagues report spatial cuing effects in all modality combinations, the performance in cued conditions being constantly better than in non-cued conditions (Spence & Gallace 2007). Spatial cues work especially well when the cue is co-located with the target and reliably preceding the target stimuli (Ferris & Sarter 2008). Ferris and Sarter studied visual, auditory and tactile cues in a complex simulated military operation task, where the subjects acted out as vehicle commanders. They found that cues have an effect on target detection and response time, even in their rather complex experiment setup. The response time to visual targets improves when those are preceded with auditory cues at the same location. Also other studies show that auditory stimulus seems to work well to attract visual attention (Driver & Spence 1998). For other modality pairings asymmetries are reported (Ferris & Sarter 2008). The asymmetries are not discussed further here as the current study uses a visual target.

2.2.3 Exogenous and endogenous cuing

The cues can work exogenously and endogenously. Endogenous attention is voluntary, and induced by advance knowledge concerning where the target is likely to occur. It could be a verbal instruction communicating that the target stimulus will appear on the left side. Exogenous spatial attention is rather reflexive. It appears as a response to a salient peripheral event, such as a fly suddenly landing on the shoulder. (Ho & Spence 2005)

Cues that present symbolic information about the target location are often endogenous cues. For example, a visual symbolic cue could be an arrow pointing to the left. Cues that are shown peripherally can attract attention to that direction without symbolic information added to the cue. (Sanabria et al. 2007) According to Ho and Spence (2005), stimulus cues that utilize both ways of directing are the most effective ones.

The results of previous laboratory study in FIOH (Näsänen et al. 2008), presented in Figure 1, show that the auditory stimulus cues, coming from the same direction where a visual target is located, facilitate the visual search task more than symbolic cues (direction of sound vs. pitch). The difference decreases when the time between cue and target stimuli increases. The shorter the delay the bigger the difference between two different cues. Also e.g. Sanabria and colleagues show in their study (Sanabria et al. 2007) that people need less time to orient their attention by means of a spatial peripheral cue than by means of central or symbolic informative cue. When presenting more complex information, symbolic cues may be useful however. These are typically referred to as icons, tactons and earcons depending on the sense: visual, tactile and auditory, respectively.

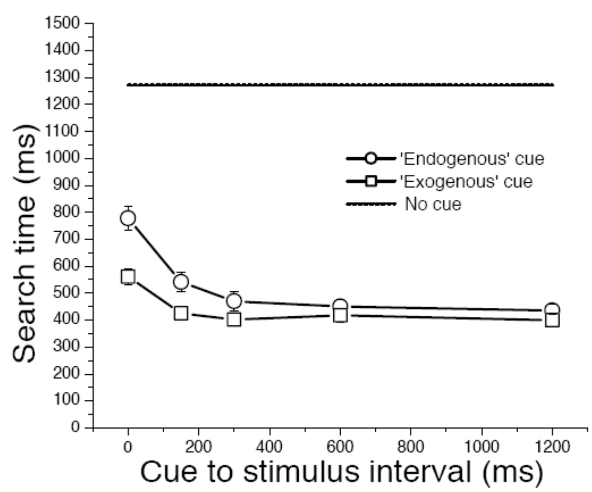

Figure 1 The exogenous stimulus cues improve the visual search time of the target (Näsänen et al. 2008)

2.2.4 Effects of expectations

Expectations affect attention and performance in perceptual tasks (Spence & Gallace 2007). These expectations may slow down the processing of a target when it appears on an unexpected sensory modality. If people expect tactile stimuli, they react slower to other sensory stimuli (auditory, visual), and the opposite (Spence et al. 2001). Repetitions on one sensory channel may speed up the processing of targets on that sensory channel when they are expected and attention is directed to that sensory channel (Spence et al. 2001). In addition to expectations, this is also related to stimulus repetition (resulting in priming). In contrast, change of the sensory channel seems to decrease the processing speed of a target.

Additionally, anticipation of stimuli presentation time leads to improvement in the execution of tasks. The timing can be anticipated for example when the time interval between the cue and target is constant. In contrary, when people cannot predict the stimulus, the reaction time is longer and the amount of errors bigger (Spence & Gallace 2007).

2.2.5 Limitations of cuing attention

Inconsistency in multisensory inputs may create faulty interpretations of the stimuli. Illusions such as the McGurk effect demonstrate that the interpretations made by the brain are not necessarily correct, at least in non-natural situations. The McGurk effect relates to the phenomenon where what we see affects what we hear, i.e. lip movements shown on video change the interpretation of speech stimuli (Moore 2004). This needs to be considered when creating artificial connections between sensory inputs.

A stimulus that is irrelevant for a task can also affect the perceptual processing of a target (Spence et al. 2004). The same mechanisms that can be used to help capture or speed up attention and task performance can cause danger situations in real life if people attend to sudden distracter stimuli appearing in the environment (Spence & Ho 2008). The cues need to be used carefully and with consideration. They should not cause the interruption of an important task and the alarm signals should be easy to locate and intuitive, but still inform about the danger in correct relation to the real threat presented by the danger. (Spence & Ho 2008)

Inhibition of return (IOR) is a phenomenon where the shifting of attention to a target location previously attended to, is prevented for some time. This effect helps prevent attention from returning to a previously attended location, in an effort to enhance the efficiency of visual search. The mechanism protects the human from wasting resources to stimuli that are repetitive. However, it may prevent attending to stimuli also when those should be attended to. (Spence & Gallace 2007)

Inhibition of return is also location and object based and therefore cross-modal effects can occur (Eysenck & Keane 2005). The auditory cues do not cause IOR for visual targets in cases where the gaze does not shift to the cue but only to the target (Ferris & Sarter 2008).

Much of the literature used as background for this chapter concentrated on discussing the mechanisms of multisensory perception and related research methods. There seem to be differences in the results of cuing effects depending on the modality combinations and experiment setup. Differences such as location, timing, distance, description of the task and other variables have an effect (Ferris & Sarter 2008), as well as priming, IOR etc. Therefore, the topics described here are limited to the scope of this study.

Cuing attention in this study

In the current study, the research methods differ from methods used in many attention and perception research studies. Laboratory settings enable strict control over the presentation of the stimuli, and often the purpose is to study the information processing mechanisms. This is not the purpose in the current study, as this study is more explorative. However, findings from laboratory research were used for designing the cues for this study.
In the current study, the spatially directing cues that reliably predict the target are a combination of endogenous and exogenous cues. The cues do not come from the same location where the target is, just from the same direction. The information about the direction is not symbolic and the purpose of the task (to shoot at given direction) is not transmitted by the cue.

3. HUMAN SENSES AND INFORMATION PRESENTATION WITH WEARABLE USER INTERFACES

This chapter describes theoretical information collected in order to select the stimulus cues for the experiment. The aim was to investigate how different modality cues can be used on multisensory wearable user interface for giving spatial direction (in a shooting task). The chapter 8.1 describes the general requirements defined for the cues in this study.

A wearable user interface presents different modality stimulus cues via devices attached on the body or on clothes. The stimuli are perceived and interpreted by the users. When defining the stimulus cues with different modalities, the designer needs to be aware of the applied aspects of basic sensory mechanisms. The way people perceive different stimuli has developed in interaction with the environment and each sense has its specific features and different mechanisms of perceiving directions. The senses are not independent but support each other, these aspects related to multisensory perception and attention are described in chapter 2.

Depending on the task being supported with a wearable user interface, one or several modalities may be used to provide stimuli that naturally support people's information processing (Shams & Seitz 2008). Some tasks clearly use several senses, like for example visual and auditory localisation of a source of a moving sound. Cross-modal and multimodal integration and related effects were discussed above in chapter 2.2.1, and need to be taken in to consideration when designing cues.

Many factors affect the perception of stimuli. The timing, location, salience and even the semantic meaning of cue make a difference to its perception (Virsu et al. 2008) (Ferris & Sarter 2008) (Spence & Gallace 2007). The location and the salience of cues used in this study are discussed for visual, auditory and tactile senses in the following chapters. The multimodal stimulus cue combinations are defined below, but theory related to multimodal perception is discussed more in detail in chapter 2.

Ho and Spence (2005) explain that in addition to perceptual considerations, the factors related to intuition are important when designing multisensory user interfaces. The stimuli and target should ideally be naturally related to each other. According to Rupert cited in (Ho et al. 2006), intuitive cues are easy to learn and most people react to such cues in a similar way. Their meaning is clear as well as are the actions required. Designing for intuitiveness requires user involvement and evaluations. The methods for evaluation of cues in this study are described in chapter 5.

3.1. Visual Display

Vision is more accurate for localization than the auditory sense and people use the visual sense to solve spatial problems (King 2009). People use vision to observe the space around them, and to localize objects. Visual sense is often considered as the most important sense, providing information on complex relations of objects in the environment, which is more difficult with spatially less accurate audition or the proximate sense of touch.

3.1.1 Visual Perception

The human eye senses light. The cornea and the lens are the focusing elements of the eye, creating a sharp image on the retina. The retina contains receptors which transform the light reflected from objects to electrical signals which are conducted toward the brain in the optic nerve. (Goldstein 2007) There are neurons in the visual cortex specialized for different kind of stimuli, some handling movement, others handling spatial frequency, etc. (Sinkkonen et al. 2002). The brain processes the sensory input and it is modulated by attention, memory and intensions (Berman & Colby 2009).

3.1.2 Perceiving and presenting directions with visual modality

Spatial vision refers to a wide area of topics of which many are not in the scope of this study; e.g. contrast sensitivity, two-dimensional patterns, etc. Here are described the aspects related to perceiving space, and more specifically, to horizontal directions.

The human visual field and its visual acuity area are relatively small. People cannot see a wide area at the same time but only the details of a target in area of ca. 5 degrees. Different distances are not in focus at the same time. Stimuli coming from the backside cannot be seen and for gathering accurate visual information outside the visual focus area, eye, head and body movements are required. Already due to these physical limitations visual attention cannot be focused in different directions in space simultaneously. (Eysenck & Keane 2005)(Sinkkonen et al. 2002)

Receptors in the retina have an anatomical position which is related to light at a particular location, (Boothe 2002) so that the activity of a receptor indicates a location of light in the environment (King 2009). For each visual direction there are two such anatomical locations, one in each eye, called corresponding points. Thus visual directions are relative to the eyes and their position (Boothe 2002). Visual cues give more narrowly focused spatial cuing effects than auditory or tactile cues (Spence & Santangelo 2009) as the retina provides detailed information about the visual world (King 2009).

Seeing and perceiving are not the same because of the mental aspect of attention. It is possible to fail to perceive something that is in the middle of the visual focus area if not paying attention to it. On the other hand, even if the focus area for sharp vision is quite small, it is possible to pay attention to things which are outside this area. (Sinkkonen et al. 2002)(Goldstein 2007) Unattended stimuli might not be perceived and therefore cue stimuli need to be salient enough. Bright objects and light flashes attract visual attention, they have high stimulus salience (Goldstein 2007), and also a sudden change in the environment is a highly salient feature (Boothe 2002).

People use a number of environmental cues for visual perception and for steering and heading in the environment. Optic flow refers to changes in the visual field when moving, and object-based steering uses information on the relative motions or positions of environmental objects. Both of these are mainly related to movement but spatial information is available also when not moving. Information on body and head movements and position affect the perception of visual directions. The term "visual direction" has a specific meaning in this context referring to the angle between the direction of the perceived target and the front-back body axis. When visual direction information is available, it is the dominant source of information (Eysenck & Keane 2005).

3.1.3 Limitations

Displays presented in the visual field affect perception. This is because visual information cues from the environment which people use to perceive depth, distance and relations of objects are distracted by the display. Cues can be both monocular and binocular (Eysenck & Keane 2005). People are used to operating in 3-dimensional visual space, as it is our natural environment (Sinkkonen et al. 2002). However, a stimulus presented on a head mounted display (HMD) superimposed to the real environment view may cause misaccommodation of the eyes, which in turn may lead to misperceiving the distance of the target in the environment (Patterson et al. 2006).

Another perceptual issue related to a monocular HMD is binocular rivalry which occurs when the images perceived by two eyes do not correspond. This causes unpredictable changes in which image is perceived (Patterson et al. 2006). Normally the dominant eye image is perceived more often and a monocular HMD is better worn over the dominant eye. The switching of attention between the display and the environment may be a possible risk factor for appropriate perception.

Displays attached to the head present an unnatural viewing situation. Not only is the display covering a part of the visual field, but also head movements that would normally affect how the stimuli in the environment are perceived do not alter the situation. Such inconsistencies and scanning of a small display may cause eyestrain (Patterson et al. 2006).

In a real outdoor field setup luminance changes are a challenge for wearable visual displays. On sunny weather information on a display can be poorly visible and, in addition, the luminance level has an effect on the user's depth of field (Patterson et al. 2006)(Winterbottom et al. 2007).

Visual cue stimuli in this study
In the pilot experiment (see chapter 9.4), a monocular HMD was evaluated for presenting directing visual cues. The HMD could not be worn over the dominant eye because this eye is needed for sighting in a shooting task. In addition, shooting position would not allow setting the monocular HMD on that side and the display was worn over the non-dominant eye. As this display was only looked at with one eye, the cues were not really peripheral spatially directing cues but rather symbolic cues. Simple graphical symbols were used, so that the symbol showing left direction was also on the left side of the screen. In practice, the eye needed to accommodate to the screen. In the pilot experiment, the HMD display was not found suitable for presenting the stimulus cue in the shooting task, as the sighting was disturbed because the subject first had to focus to the display. In the actual experiment flashing led-lights were used as stimuli. A flashing light is a salient cue attracting visual attention to its location (Boothe 2002). The led-lights were attached to goggles and therefore the stimuli were different from stimuli that would originate from actual location of target in distance. The cues showing left and right directions could only be seen with one eye on the display, whereas all targets ahead were visible for both eyes. However, the direction of the cue and the target is similar, and the cue possibly attracts attention to the direction of the target. Focusing the eyes to the cue location is not needed to perceive the direction, and the subjects were told to look forward in a way they normally would when preparing for a shooting task. In the experiment the same stimulus cue was used in sunny and cloudy weather conditions despite of the changes in luminance.

3.2. Auditory Display

Auditory stimuli are perceived from all directions and people can detect an object in the environment just by hearing its sound also when the source is not visible (Bear et al. 2007). The spatial resolution of the auditory modality is poorer than that of the visual modality, but auditory cues have been shown to be efficient in directing visual attention (Ferris & Sarter 2008).

3.2.1 Auditory perception

Sounds are variations in air pressure, sound waves, which are perceived by the receptors in the inner ears (Bear et al. 2007). The sound perceived is determined by its frequency and amplitude (Goldstein 2007).

The frequency of the sound refers to the number of pressure changes per second. (Goldstein 2007) High frequency sounds are perceived as having a high pitch. The sound frequencies within the human hearing range are between 20 Hz to 20 kHz, although with age, the range decreases significantly especially for high frequencies. In addition, an exposure to noise decreases the frequency range heard (Bear et al. 2007). The ear is most sensitive for sounds ranging between 2000 Hz - 5000 Hz frequency, also known as the presence area. Animal or alarms sounds often have much of these presence area frequencies. (Kenttämies et al. 2001)

The amplitude of the sound refers to the size of pressure changes (Goldstein 2007). The loudness of sound, volume, is determined by the amplitude of the sound, larger amplitude sounds being perceived as loud (Bear et al. 2007). The loudness perceived also depends on sound frequency. For example, at 10 dB volume the frequencies between ca. 500 Hz- 8000 Hz can be heard. The wider the frequency band is in a sound, the louder it is perceived (Kenttämies et al. 2001).

Natural sounds normally consist of complex combinations of different frequencies and amplitudes, which contribute to unique differences between the voices of people or for example between music instruments (Bear et al. 2007).

3.2.2 Perceiving and presenting directions with auditory modality

The auditory sense differs from tactile and visual sense in that the spatial source of the stimuli does not directly stimulate certain receptors indicating the locations. Instead, the auditory system needs to interpret the location based on different kinds of location cues, which depend on the way sound interacts with the listener’s head, upper body and ears. However, auditory stimuli also have neural representations for locations of stimuli, known as topographic maps. (Goldstein 2007)

The location cues are related to the time and volume differences between the ears, reflections from the head, body and the ears, and movement of the head (Goldstein 2007). The location cues can be binaural or monaural, which means that both ears, or a single ear is used, respectively. The horizontal localization of a sound requires stimuli for both ears, whereas vertical localization does not (Bear et al. 2007).

For auditory signals, people perceive the horizontal location better than the vertical location. People can locate most accurately sounds from the front; the side directions are less accurate (Goldstein 2007). The distance of the source can be located rather poorly (Moore 2004). The localization of sound is supported by the visual sense in natural environments, when for example a visual object is perceived as a source of the sound (Goldstein 2007). Expectations also have an effect (Härmä et al. 2003), as described in chapter 2.2.4.

3.2.3 Horizontal sound localization

Horizontal localization is based on the differences of the stimuli reaching each ear. The binaural cues of the stimuli are compared. Interaural time delay relates to difference in times of the sound stimuli reaching each ear, and interaural intensity difference relates to sound volume differences perceived in each ear.

In the interaural time delay, the timing of the sounds arriving at each ear gives information on the location of the source as the ear not facing the source receives the sounds later. The time difference between the ears is small, only around one ms. People cannot do this separation between the ears for continuous sounds or sounds with higher frequencies (>2000 Hz). The sound of a snap or variable alarm sounds are the easiest to locate. (Goldstein 2007)(Kenttämies et al. 2001)(Bear et al. 2007)

Interaural intensity difference, i.e. the sound volume difference gives information about the sound direction. When a sound comes from the side, the head makes a sound shadow causing the opposite side ear to hear the sound with lower intensity. A sound that arrives to the ears from the front, above or back is as strong for both ears. In addition, there are differences in sound tones between the ears when the sound is coming from one direction because the head affects the sound waves. For frequencies smaller than 2000 Hz there is not intensity difference due to the length of sound waves. (Goldstein 2007)(Kenttämies et al. 2001)(Bear et al. 2007) If there are several sound sources, the source from which the stimulus first reaches the ear is perceived as the source (Goldstein 2007).

3.2.4 Vertical sound localization

The internal time or intensity differences are not useful for determining if a sound comes from above or below. Therefore, monaural cues are used for vertical localization.

The pinna, the outer part of the ear, and the auditory canal are in an essential role for monaural cues and if they are covered, the vertical localization of a sound is impaired (Bear et al. 2007)(Goldstein 2007). This is because the sound entering the auditory canal is first reflected from the head and in the structures of the pinna. The pinna increases the intensity of some frequencies, e.g. those ranging between ca. 2000-5000 Hz, known as the presence area (Goldstein 2007).

3.2.5 Head-related transfer function

HRTF (head-related transfer function) describes how the sound is modified because of reflections on the head, ear and upper body before it enters the inner ear. Therefore, a person's HRTF can be used to create an artificial binaural sound in a virtual environment (Goldstein 2007). However, there are some limitations with perception of such artificial spatial sound effects. 3-dimensional sounds are often perceived coming from “inside the head” or there is confusion with their locations (Härmä et al. 2004)(Zhou & Rogers 2002) Internalization takes place especially when sound is delivered through headphones, isolating the sounds coming from other sources around; it does not create a natural impression of the space for all subjects. In a navigation study (Kumagai & Massel 2005) a stereo audio display performed quite well compared to a 3D (with HRTF) display, and when facing the target direction, the stereo display performed even better. The HRTF is different for each person and even if a generic HRTF may be created for externalizing the sound, different persons perceive the sounds in these virtual auditory spaces differently (Härmä et al. 2004).

3.2.6 Limitations

The above-mentioned time and volume differences can be reproduced with two loudspeakers but an issue with this is that both ears then hear these differences. Presenting audio stimuli through headphones enables good control over the stimuli and it is possible to vary the timing and intensities of the sound. However, for headphone listening it is typical that the sounds appear to localize inside the head rather than in the environment. Another difference compared to natural stimuli coming from the environment is that as the sound source is attached to the head, head movements affect the localization of a sound source differently than in natural interaction (Härmä et al. 2003).

The auditory masking effect degrades the detection of a sound stimulus when another stimulus is shown at the same frequency at the same time. The thresholds for detection also increase for stimulus frequencies below and above the masking stimulus frequency (Goldstein 2007).

Auditory cue stimuli in this study
In the current study, only horizontal localization cues were needed, the ones humans are best at. Simple cues were used without artificial timing or intensity differences, or a HRTF. A cue was given to one side (left or right) or both sides at the same time (centre). People perceive the sound as coming from the source that reaches our ears first, and sound coming simultaneously from either directly back or front. The subjects were told that the centre sound means "ahead". A two-dimensional stereo audio display performed well compared to 3D displays in a navigation study (Kumagai & Massel 2005) and even if advanced 3D technologies exist, a simple stereo cue was found sufficient and reliable for the aims of this study. The sound frequency is a less important factor in localization when using headphones than in real environment. The sound presented with headphones is not affected by air, head shadow or movements and the volume can be adjusted personally. In this study, a 1000 Hz sound was chosen because it is within the range of frequencies that produce the lowest detection thresholds. The pure sine wave sound was edited to consist of several short breaks. This created snapping effects, which were considered as easier to localize and the editing widened the frequency band as well. In addition, the edited sound was perceived a bit more agreeable compared to a pure sine wave. One criterion for choosing the sound was that it is not similar to speech or other normal sounds in the environment: a non-natural sound would not be so easily confused with natural sounds and could be easier to detect and recognize. Different versions of audio stimulus cues were first evaluated in a laboratory environment, and the best ones were evaluated also in a pilot experiment before selecting the ones for the experiment.

3.3. Tactile Display

The tactile sense has not been utilized much in user interfaces even if the skin is the largest sensory organ and it is accurate in perceiving timing and space (Brewster & Brown 2004). The tactile sense perceives objects that are close to or touching the skin, therefore it is personal and private, and a silent, unnoticeable way to present information to users (Krausman & White 2006)(Jones & Sarter 2008) In addition, the tactile sense captures attention efficiently (Jones & Sarter 2008) and has a naturally protecting function (Goldstein 2007). In several tasks in which auditory or visual modalities are not available or are overused, the tactile sense may be useful. For example, in darkness, the visual sense is less useful and in a noisy environment, auditory stimuli might not be heard.

Tactile displays have been used for sensory substitution, for spatial orienting and navigation. Vibration stimuli are often used in tactile displays and the current study uses vibrating tactors to induce the stimuli as well. In the scope of this study only vibration stimuli are discussed but also other type of tactile displays exist. For example, haptic interfaces include kinesthetic cues, which involve active movement action of the user. (Jones & Sarter 2008)

3.3.1 Tactile Perception

Tactile perception refers to the cutaneous perception, which is part of the somatosensory system. The sensations on the skin are based on the stimulation of receptors called mechanoreceptors, which are located on the uppermost layers of skin, epidermis and dermis. There are four types of mechanoreceptors, which differ in structure and reactions. (Goldstein 2007)

The receptors are distributed everywhere on the body and are not centralized in one specialized area like auditory or visual sensory receptors (Bear et al. 2007). The amount of receptors in different skin areas on the body varies, and the highest density of receptors can be found on sensitive areas like for example on fingertips. Also in the brain, the sensitive areas have larger representation in the somatosensory cortex. (Goldstein 2007)

Vibration stimulates the Pacinian corpuscles, one type of the mechanoreceptors. The Pacinian corpuscles react to the onset or offset of stimulation, which in practice means that they react to change (Goldstein 2007). The frequency, amplitude, duration and location of the stimulus affect the perception (Jones & Sarter 2008). The sensitivity of these receptors to vibration all over the body surface is best when the vibration frequency is between 150 and 300 Hz. On lower and higher frequencies, the sensitivity decreases. The frequency and amplitude sensitivity varies depending on the location on body surface and an increase in amplitude may increase the perceived frequency. Stimulus duration for tactile displays normally ranges from 80 ms to 500 ms (Jones & Sarter 2008)

3.3.2 Perceiving and presenting directions with vibration stimuli

Communicating directions is one of the simplest information presentation scenarios with a tactile wearable display (Brewster & Brown 2004). Due to the spatial accuracy of localizing on the skin and the corresponding representation in the brain, it has been proposed that the sense of touch could also be used to give reference to an object in the external world with help of tactile display technology. (Cholewiak et al. 2004)(Jones & Sarter 2008) Tactile cues have been considered useful in directing user's spatial attention to a visual target, and tactile cues have often been considered as intuitive and fast to learn (Jones & Sarter 2008). Tactile wearable displays have been studied in relation to navigation with promising results, and also for giving spatially directing warnings to pilots and drivers (Krausman & White 2006) (Ho et al. 2006).

In many applied contexts, such as navigation, the vibration stimuli are presented on the torso, around the waist as a tactor belt or on the back as a tactor vest (Jones & Sarter 2008). Around the waist, detection thresholds are similar at all sites. Cholewiak and colleagues, who studied the localization of vibration stimuli on this body area, point out that such tactile displays leave the hands and limbs free for other tasks. In addition, the trunk is more stable in relation to people's normal movement direction than the limbs and thus it provides a more stable reference for space and directions. The navel is always in front of us, and for example, head and eye positions are less determining. (Cholewiak et al. 2004)

The localization accuracy depends on the on-body location of the stimulus. The localization is most exact when the stimulus is presented close to body anchor points such as the elbow, the shoulder or the navel. These anchor points can function as a reference to an external target (Cholewiak et al. 2004). The sensitivity differences on the body surface have been mostly studied with brief touch stimuli but less with vibration and the sensitivity for vibration is not quite the same. The distance between tactors needs to be long enough. The study of Cholewiak and colleagues shows that six tactors around the waist were localized better than twelve tactors. More stimuli do not make the tactor array more informative, but the tactors placed too close to each other can be confused with each other (Cholewiak et al. 2004).

Simple vibration cues can be used for wearable interfaces but also more complex symbolic stimulus cues have been developed. These so called tactons are structured messages which consist of combination of vibration frequency, duration or other parameters such as rhythm or location on body surface (Brewster & Brown 2004)(Jones & Sarter 2008).

3.3.3 Limitations of using the sense of touch

There are plenty of possibilities for the use of tactile sense for presenting information, but there are also some limitations related to it. One challenge in using a tactile display is to keep the vibrator in contact with the skin during a physical activity like running (Brewster & Brown 2004). A study of tactile stimuli in simulated combat movements (Krausman & White 2006) shows clear degradation in the detection level of tactile stimuli in demanding field conditions. The weight and pressure of a bulletproof vest and other combat clothing on tactors did not, however, affect performance.

Body size differences of people may be an issue in designing tactile wearable displays. When using the same dimensions for several users, the equipment might be better adapted for some subjects than for others (Jones & Sarter 2008).

Tactile attention has some limitations in comparison to auditory and visual attention in that once attention is captured, it takes more time to shift the attention away from tactile stimuli, compared to the other senses (Spence & Gallace 2007). In addition, people have shown difficulties in processing several tactile stimuli in a short time, and therefore the use of complex tactile displays may be limited. Change blindness also affects the sense of touch as people can quite poorly detect change in a repeatedly presented tacton when another stimulus is shown in between (Jones & Sarter 2008). In addition, a masking effect may cause that a touch stimulus is not detected when another stimulus is shown at the same location right before or right after it (Spence & Gallace 2007).

Tactile stimuli in this study
A vibration cue presented on a tactor belt over the abdomen above navel was chosen for this study. Other body location options were evaluated before the pilot experiment in laboratory settings, but for a shooting task, the stimuli shown on torso appeared as the most intuitive and least disturbing option. In this study the tactile stimuli are simple and do not need to present any more complex information than the directions left, center and right.

3.4. Multimodal and mixed-modal stimuli

Multimodal and mixed-modal stimuli in this study
Multimodal stimuli in the current study refer to stimuli presented simultaneously on different sensory modalities. Multisensory stimuli seem to be stronger for capturing attention than unimodal stimuli, and therefore those may be useful for presenting information which need to be attended to. This effect is described in chapter 2.2.1. Mixed-modal stimuli refer to task conditions in which different modality stimuli are randomly shown. The reason for including the mixed condition in this study was to investigate the possible effects of expectancy. These effects are described in chapter "2.2.4 Effects of expectations".

4. PHYSICAL EXERCISE AND STRESS EFFECTS ON HUMAN PERFORMANCE

In a crisis management task, the pressure to succeed is huge. It is often about either staying alive or saving lives of others and thus failure is not an option, and threat is constantly present (Burke et al. 2004). External or internal threats to humans trigger autonomic nervous system activation known as stress (Hancock & Szalma 2008). Several physical and mental reactions are induced by a threat, in order to counter it, and they affect attention and perception. Cannon, cited in (Bracha et al. 2004), formulated the term "fight or flight" referring to behaviours typically adapted during a threat situation. The human body prepares to counter the threat; the heart and lungs start working harder, auditory exclusion may degrade hearing, and tunnel vision decreases visual focus area. Several other reactions are known as well. (Hancock & Szalma 2008) In addition, humans react even to stimuli known not to be a threat. For example, a sudden strong stimulus such as a loud noise like a gunshot or a flash of light normally causes some short duration effects such as tightening of eye and neck muscles. This is known as the startle response. The startle response is related to surprise, whereas the "fight / flight response" is related to fear and threat.

For humans to perform optimally, stress and arousal level should not be too low or too high. There are techniques for assessing arousal and stress, such as electroencephalography (EEG), galvanic skin response (GSR) and heart rate. The increased frequency and decreased amplitude in EEG, increased skin conductance, and increased heart rate indicate arousal (Hancock & Szalma 2008).

The use of peripheral visual cues may be limited due to tunnel vision in stressful situations when the visual focus area decreases (Hancock & Szalma 2008). The limitations are attentional rather than perceptual and therefore the saliency of a cue is an important factor. Stimuli on the sides may be better attended to, if those are the most salient stimuli at the time, and relevant for a task (Hancock & Szalma 2008). Tunnel vision can be broken but it requires training. It is difficult, if not impossible to prevent these physical and psychological effects and therefore they can only be mitigated. Even if humans may maintain performance levels in a main task during stress, it results in psychological and physiological costs, and eventually the performance capacity may fail totally when stress increases above a threshold level (Hancock & Szalma 2008).

Physical exercise in this study

The effects of stress on human perception and attention need to be considered in a crisis management situation. Several studies of directing attention with spatially directing cues have been performed in laboratory settings as described previously in this document. The actual real-world situations are more demanding and it cannot be assumed that the same effects apply as in laboratory settings. These aspects related to ecological validity are described in chapter "5.4 Validity".

It is not possible to simulate many of the stress aspects related to real crisis management tasks, where bullets may actually start flying towards soldiers. Here, physical exercise was used to simulate "battle stress" as increased heart rate and heavier breathing are likely to have an effect on perceiving cues. Adding physical effort is a method commonly used in military context to simulate a combat situation. Shooting as such is already a stressful task, especially when fast and accurate performance is required, and the time between shots during the task is short. Physical exercise performed just before a shooting task was used as an additional experimental condition to add task demands without changes to the actual task, and the short breaks during stimuli did not allow full recovery during the task.

5. EVALUATING HUMAN FACTORS

5.1. People and technology

A technology and computer centred view in developing interactive systems or products may be easier and faster for technically oriented individuals designing and implementing them. However, the final product may be harder and slower for the actual users to use, at least if they are less technically oriented. For understanding the needs and problems of actual users of a system, the system should be evaluated regularly during its development in context of its actual use. Evaluating human interaction with a computer and technology is a key element of interactive system development. The system needs to be fit for its purpose. (Benyon et al. 2005)

5.2. Usability evaluation

The international standard ISO 9241, part 11, cited in (Benyon et al. 2005) defines usability as *"The extend 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."*

Usability evaluation, often also called usability testing, generally refers to evaluation of interactive products such as web sites or mobile phones during their development. Some usability testing methods can be used before an actual product prototype exists for choosing between different design concepts and technologies. (Benyon et al. 2005)

Interviews and questionnaires are commonly used during the analysis stage of interactive system development for collecting information, but they are also useful for the evaluation or comparison of concepts or designs. Ideally, users participate in usability tests; they complete a pre-defined task and get interviewed or fill in a questionnaire about the experience. Questionnaires often do not collect very detailed data nor give subjects the freedom to express other experiences than the ones asked, thus both questionnaires and interviews can be used to complete each other. Performance measurements such as speed of task execution and number of errors in tasks are also commonly collected and evaluated. (Benyon et al. 2005)

5.3. Reliability

Reliability refers to the consistency with which the same result would be gotten if the tests were to be repeated (Nielsen 1993). This is a problematic issue in usability testing because differences between subjects' performance and preferences are normally huge.

Many evaluation techniques are relatively easy to apply but drawing conclusions from the collected data is more difficult. Data analyses need to be performed carefully, and even then it is sometimes difficult to interpret the results. One needs to be especially careful in generalizing results from a small number of users. Such data may give indications of possible problems, but do not prove a system usable, nor reveal how other users will experience the system. (Benyon et al. 2005). Even so, decisions for interactive system design can be made based on the results that are fairly unreliable, as some data is better than no data (Nielsen 1993). For getting more reliable results, statistically validated results are required. Such data can be collected only with well-designed experiments, and with quantitative data (Benyon et al. 2005).

5.4. Validity

Validity refers to the extent to which a usability test measures what it was intended to measure. The results should be relevant also in the real usage environment, not only in the laboratory. According to Nielsen (Nielsen 1993) *"typical validity problems involve using the wrong users or giving the wrong tasks or not including time constraints and social influences"*.

Many real life aspects are often excluded when testing, for example, a mobile interactive system because it is very difficult to reproduce realistic circumstances. In real life, people are generally doing several things at the same time, being interrupted etc. In addition, the aims and the environment are constantly changing and people need to adapt to these changes. The more controlled the testing situation becomes, the less it is likely to resemble the real world. The problem is that reliable data collected may then not be ecologically valid: the findings actually do not reproduce in real circumstances for example. Thus, even if statistically significant differences are found in tested systems, it may or may not indicate the issues in real life usage. (Benyon et al. 2005)

Evaluating human factors in this study

Many usability evaluation methods have been developed for use in the design phase of a product. In the current study there is no particular product being designed and evaluated but different kind of stimuli communicating direction. These stimuli were compared, considering the requirements of a crisis management task. Thus, this study is not a classical usability study, but rather applied basic research. However, the methods used are similar to commonly used usability testing methods. The approach is similar to user testing as the subjects could in future use devices like the ones being developed and studied. Using real users provides direct information and feedback from users on how they perform with the devices, what issues they might have and how they feel about the cues. In addition, the actions of the subjects could be observed during their task execution. The methods used for performance measures and subjective measures are described in more detail in the following chapters. Performance measurements included reaction time and shooting accuracy, and analysis of speed-accuracy trade-off. Subjective measurements included questionnaire, structured interview and NASA-Task Load Index (to obtain estimates of subjective workload).

6. PERFORMANCE MEASURES

This chapter describes the statistical tests relevant for the study, and the speed-accuracy trade-off experiment design approach.

6.1. Statistical tests

6.1.1 One-way repeated measures ANOVA and Friedman test

In a repeated measures design each subject participates in all conditions of the experiment, requiring fewer subjects and reducing variability. There may be differences in performance between individuals and it is important to control for such differences. The repeated measures design controls for individual differences as each subject is measured several times, acting as a control for himself. The disadvantage of this method is the learning effect, which needs to be taken in account in the experiment design. (Field 2009)(McCreery 2007)

One-way repeated measures ANOVA can be used to test the differences between several related groups, for example in different conditions. The Repeated measures ANOVA compares the means of dependent variables. It corresponds to the Paired samples t-test but allows comparing more groups. (Field 2009)(UCLA 2009)

When assumptions for the use of a parametric statistical test are not met, the non-parametric option for repeated measures ANOVA is the Friedman test, also called Friedman's ANOVA. The Friedman test can also be used to compare more than two means or medians in several conditions. It tests whether matched samples were drawn from the same population. The Friedman test is based on ranks. The null hypothesis is that the distributions of the ranks are the same for each group. The observations must be measured at the ordinal level at least. A significant result from the Friedman test does not provide information on where the difference is. For this, separate post hoc tests are required. Non-parametric paired comparisons can be made for example with the Wilcoxon signed ranks test. The Bonferroni correction needs to be made for these multiple pair comparisons. (Metsämuuronen 2005)(Sarna 2009)(UCLA 2009)

6.1.2 Paired t-test and Wilcoxon signed ranks test

The simplest repeated measure is an analysis of before and after measurements for each subject. The method generally used for such two-group comparison is the Paired samples t-test. A non-parametric version of the Paired samples t-test is the Wilcoxon signed ranks test. (Metsämuuronen 2005) These tests can be used as post hoc tests, for multiple comparisons.

6.1.3 Bonferroni correction

The Bonferroni correction is a method to adjust the p-values of multiple comparisons tests in order to maintain the planned significance level for a statistical test. In multiple comparisons the p-values are too liberal, and can result in erroneous conclusions in a certain percentage of comparisons. The Bonferroni correction method in it simplest form means adjusting the p-value for significance. The corrected significance level is the normal 0.05 level of significance divided by the number of tests conducted. The Bonferroni correction is often very conservative because it supposes that the multiple comparisons are dependent even if mostly that is not the case. (Field 2009)(Sarna 2009).

6.1.4 Parametric vs. non-parametric statistical tests

For parametric statistical tests certain assumptions must be true in order to get reliable results. Therefore, these assumptions need to be tested before deciding which statistical test can be used. First, data are assumed to be from a normally distributed population. This is the most important of the assumptions. The normality of a distribution can be observed visually from a frequency distribution histogram. If sample data seem normally distributed, it can be assumed that the populations are also normal. In order to know if the distribution is close enough to normality, an objective test can be used. The Shapiro-Wilk test compares an experiment sample to a normally distributed sample with the same mean and standard deviation. A non-significant test result indicates that there is no significant difference between the samples and that the distribution is probably normal. The second assumption is the homogeneity of variance, i.e. the variance is assumed to be same throughout the data. This can be tested using the Bartlett's test. Third, the data were assumed to be measured at least at the interval level, with point distance being equal along the scale. Fourth, the behaviour of one subject is assumed not to influence the behaviour of another subject. (Field 2009)

Non-parametric statistical tests have no assumptions of distribution of population, nor of homogeneity of variance. Data can be measured at the ordinal level. Non-parametric tests deal with ranks, and are less affected by outliers and skewed distribution. Their disadvantage is that they have less power because they use less information. (McCreery 2007)(Field 2009)

Statistical tests in this study
The current study used a repeated measures design. The possible learning effect was minimized by varying the task order. As the assumptions for parametric test were not met for most samples (see Appendix 8), non-parametric tests were chosen for statistical tests for all reaction time data, the Friedman test for comparing group means, and the Wilcoxon signed ranks test for multiple comparisons post hoc tests. Original data values were used. Parametric statistical tests could have been used for accuracy data analyses, but for consistency to reaction time results, non-parametric tests were chosen for statistical tests for accuracy data as well.

6.2. Reaction time data and speed-accuracy trade-off

Reaction time data should not be influenced by variables that are not in the focus of the experiment. One such variable may be the error rate - the accuracy of task performance (Sternberg 1998). In almost any task one can trade-off accuracy against speed, or speed against accuracy. That is, to improve accuracy at cost in speed, or improve speed at cost in accuracy (Wickelgren 1977)(Schmidt & Wrisberg 2008). There can also be other interactions in addition to the one between speed and accuracy (Drury 1999), but these are not discussed in this study. The speed-accuracy trade-off is related to both motor skills and cognitive skills, and depends on task requirements and goals (Schmidt & Wrisberg 2008).

Woodworth, cited in (Kim et al. 1996), studied the accuracy of voluntary movement to learn about processes of motor control using such methods already in 1899. Woodworth's movement study showed that as speed increased, accuracy deteriorated (Kim et al. 1996)(Drury 1999). A later developed logarithmic relation between movement speed and accuracy is known as the Fitts' law, cited in (Lachman et al. 1979). Fitts found that people can adjust their performance depending on the given instruction and objectives. In movement control, precision-based errors are common when trying to make things faster than normal, people can move very quickly at the cost of being less accurate, and opposite (Schmidt & Wrisberg 2008).

In addition to the control of movement strategy, information processing strategies and mechanisms in cognitive tasks have been studied with speed-accuracy trade-off designs. Also in cognitive tasks people can choose whether they prioritize speed or accuracy, depending on the situation. Therefore, task instructions should clearly describe what is expected from the subjects. If the subject is asked to be both fast and accurate, his emphasis will be somewhere between those (Sternberg 1998). On the other hand, reaction time experiments that instruct to answer as quickly as possible may cause overestimating people's capacity. (Lachman et al. 1979) In a reaction time experiment, the interpretation of the data would be problematic if the subjects have not had a constant approach to the speed-accuracy trade-off in different conditions (Sternberg 1998).

According to Wickelgren (Wickelgren 1977), the speed-accuracy trade-off studies are often preferable to reaction-time studies when studying different levels of information processing mechanisms in cognitive tasks. Error rates or accuracy should be evaluated and reported for each condition (Sternberg 1998)(Wickelgren 1977), and especially when studying information processing mechanisms the possibility of speed-accuracy trade-off should be ruled out to be able to interpret reaction time differences. This complicates the work of researchers (Lachman et al. 1979). However, complete speed-accuracy trade-off studies require much more data than reaction time studies. Therefore, experiments designed only to determine whether one condition is harder than another can normally use a reaction time design. It should still be noted that when differences in accuracy are recorded between conditions, it might indicate trading accuracy for speed (Sternberg 1998).

6.2.1 Ruling out speed-accuracy trade-off

The reaction times and accuracy for each condition need to be evaluated and compared, and speed-accuracy trade-off ruled out, especially when interpreting differences in terms of information processing mechanisms.

Theoretically, the processing limit of people is the point of the quickest response when no errors are made (Lachman et al. 1979). A demanding experimental condition often slows responding and increases errors at the same time. If there is a difference in reaction times, and the faster condition is less accurate, it can be due to a speed-accuracy trade-off and a decrease in reaction time should not be interpreted as an improvement. However, if there is a difference in reaction time, and the faster condition is more accurate or as accurate, it is not due to a speed-accuracy trade-off and reaction time differences can be interpreted. The faster conditions can be considered as less demanding (Sternberg 1998)(Wickelgren 1977)(Lachman et al. 1979). This does not tell anything of the size of the reaction time differences, only the direction of the difference (Lachman et al. 1979). This is enough for practical questions, such as when asking which experimental condition is more demanding (Sternberg 1998), or if e.g. the found reaction time difference could rather be due to speed-accuracy trade-off.

Speed-accuracy trade-off in this study

The interest of the current study is not to study different levels of information processing dynamics. The experiment task is practical and complex, and the measured reaction time consists of both perceptual and motor processes, which are not discussed in more detail in the scope of this study. Instead, the different conditions are compared to find overall performance differences. Therefore, it was not necessary to make a complete speed-accuracy trade-off experiment design (Wickelgren 1977). However, the reaction times and accuracies were investigated together to rule out the possibility that possible differences found would be due to speed-accuracy trade-off.

Shooting is a controlled motor action where even small movement affects the accuracy. The accuracy of sighting movement and the time used for sighting needs to be balanced against speed; when the requirements for accuracy in motor control increase, also the time needed for movement increases. The bigger the movement and shorter the time for movement, the more errors take place (Schmidt & Wrisberg 2008). In a shooting task the hit cannot be considered as chance like a successful answer in a cognitive task might. The professional soldiers could have hit the target practically every time if that was the objective given to them, and they could have aimed for as long as they wanted. However, in the current study, the soldiers were instructed to shoot as fast and as accurately as possible, and they knew that the reaction times were measured. Speed is also required in their normal shooting training.

7. SUBJECTIVE MEASURES

7.1. Assessing modalities with interviews and questionnaires

Subjective measurements collect information on people's perceptions and opinions. The information gathered with subjective measurements is valuable as such but in addition, it may be used to explain differences in performance measurements. Subjective measurements may be either quantitative or qualitative. Interviews are one of the most effective methods for collecting qualitative subjective data (Benyon et al. 2005) and questionnaires can be used for gathering quantitative data.

7.1.1 Questionnaires

Questionnaires can be used to collect reactions after the test sessions. (Benyon et al. 2005) Rating scales, such as the Likert scale are commonly used to collect user perceptions in usability evaluations. Subjects choose between 5 or 7 scale options ranging from "strongly disagree" to "strongly agree". Questionnaires are quick to perform during or after experiments, and they are easy to teach to subjects. They are especially useful for comparing different products, or features but their output is limited (Stanton & Young 1999). The most time taking and challenging part of questionnaires is preparing them. Questionnaires need to be planned carefully so that all questions are understandable, specific and unambiguous. The questions should collect data that answer the objectives set for the evaluation. Closed questions are easier to analyse, but the item evaluated needs to be known well enough to plan for these. Open questions allow people's own comments but are more difficult for quantitative purposes. (Benyon et al. 2005)

7.1.2 Interviews

Interviews can be used to complement questionnaires, which do not collect many details (Benyon et al. 2005). There are different options for interviews. Structured interviews follow questions carefully planned in advance. These often have closed answer options, which limit the collection of unexpected responses but on the other hand make interviews easy to administer. Often semi-structured interviews are used, where questions are also prepared but not necessarily followed exactly and thus answer options are not as closed. This allows including non-planned topics. The more open the approach is, the more demanding it is for the interviewer, but the collected data are often more informative - although also more difficult to categorize. The questions and form of the interview depend on the goals of the interview and what is being evaluated. (Benyon et al. 2005)

Assessing modalities with interviews and questionnaires in this study
Both quantitative and qualitative information on subjective experiences were collected. After each task condition a quick interview was performed to record first impressions, and the NASA-Task Load Index (TLX) form was filled (see next chapter). After all sessions the subjects participated in a structured interview and filled in an end questionnaire. The quick interviews were very simple; the subjects were requested to list three positive and three negative things about the used cue or the devices. The end interview was a semi-structured interview, which consisted of 10 questions. The questions were designed to help the subject to consider and express different aspects of the cues. The subjects could also freely express any thoughts they had about the cues, the experiment and related topics. The interviews were audio-recorded and analysed in detail afterwards. Notes were taken also during the interview. The interview questions are listed in Appendix 6. The end questionnaire consisted of 17 questions quite similar to those in the interview, but more specific. For each question, rating scales were shown separately for three different modalities (visual, auditory, tactile). The subjects were explained that the purpose was to express how different the modalities were, if at all. The main purpose of the questionnaire was to collect data for ranking the modalities. The subjects used a seven-point Likert scale (bipolar rating scale) to respond. There were a few negatively formed questions, and the ratings of those were inversed to the same scale to enable the comparison of these with other questions. All questions were analysed together to compare the overall difference between modalities, and additionally, the differences were compared separately per each question. The quick interview and the end interview were tested in the pilot experiment. A few questions were revised and a few added to the end interview. The questionnaire was decided to be added in order to record the subjective experiences also quantitatively, and to enable ranking modalities. The questionnaire questions are listed in Appendix 2.

7.2. Assessing workload with NASA-Task Load Index

Workload refers to the cost of accomplishing a task. NASA-Task Load Index, (NASA-TLX) developed by Hart and Staveland (Hart & Staveland 1988) is a commonly used tool to obtain estimates of subjective workload. It provides a measure of overall workload on a scale from 0 to 100. NASA-TLX is a multi-dimensional rating tool consisting of six subscales, which can be used to collect information on the relative contribution of each of these dimensions in different kinds of tasks. It has been used in various domains including e.g. simulated combat (Hart 2006)(Warm et al. 2008). The dimensions are mental, physical, and temporal demands, performance, effort and frustration level. The definitions of the NASA-TLX rating scales are described in Appendix 3. Subjects rate workload for each dimension on an interval scale. (Stanton et al. 2005)

In the original NASA-TLX, the dimensions are weighted. The weighting is made with paired comparisons of dimensions. Subjects rank the importance of the dimensions to their subjective definition of workload. The weight is then used to multiply the dimension ratings accordingly. It is common to use simple averages or addition of the dimension measures and skip the weighting. This is sometimes referred to as Raw TLX (RTLX). Another common variation is to analyze dimension ratings separately. The dimension analyzes can provide information about possible workload or performance problems. (Hart 2006)

Assessing workload with NASA-TLX in this study

NASA-TLX is used in this study (Appendix 4) to investigate the existence of increased workload in the tasks with physical exercise. In the questionnaire six different dimensions of workload were rated: mental demand, physical demand, temporal demand, performance, effort and frustration level. The weighting procedure for dimensions was not used but dimensions are investigated separately, in order to collect information on their contribution to the overall workload differences. Mean ratings for different conditions with and without physical exercise were compared with the Friedman test and further analysed with the Wilcoxon Signed Ranks test.

8. THE MOTIVATIONS AND PREDICTIONS FOR THIS STUDY

8.1. The Requirements for and choice of stimuli

The information related to the shooting task was collected informally in meetings and discussions with the military contact persons of the project. The shooting task in the experiment simulated a close combat task. For such a task, like for other crisis management tasks, quick actions are often required. The detection of information may be crucial for success in a task, thus the stimuli need to be salient enough. On the other hand the stimuli should not be disturbing as they should not prevent paying attention to other important events in the environment.

In the following, the requirements defined for the stimulus cues (and devices) in this study are listed:

A cue stimulus:

- should support fast actions, should be short, and of the same duration for all modalities (comparability)
- should be easy to detect
- should not disturb sighting of the environment (visibility)
- should not disturb shooting (position)
- should not be annoying
- should be able to present the three required directions with a wearable device: indicate the targets
- should be simple, should be intuitive
- should provide similar level of "alert" regardless of modality: the modalities should be perceived as being of similar strength in comparison to each other (comparability)
- should fit well with the task requirements

All these requirements were considered when choosing the stimulus cues, separately for each modality.

The selection of stimuli for this study was made based both on theoretical background and on trials in laboratory environment and during the pilot experiment. Different versions of the cues were compared and the ones considered the best were chosen for each modality. The comparability of the cues presented in different modalities is a difficult issue, which in this study was based on subjective evaluation. In the trials, the auditory cue volume was adjusted to be similar to the perceived strength of the tactile cue, and to the perceived strength of the visual cue. This was repeated in the pilot study and the actual experiment with the subjects. One audio cue volume level was good for most of the subjects but a few subjects needed slightly higher volume level to perceive the cues clearly.

8.2. Objectives of performance measurements

With performance measurements the effects of different variables are investigated in relation to reaction time (RT), accuracy (AC) and both together to rule out the effects of speed-accuracy trade-off (SA).

The purpose is to collect information about:
a) Does the modality or task have an effect on reaction times?
b) Does physical exercise have an effect on reaction times?
c) Does target direction have an effect on reaction times?
d) Does the task have an effect on accuracy?
e) Does physical exercise have an effect on accuracy?
f) Are the possible differences in reaction time due to speed-accuracy trade-off?
g) Are the possible effects of physical exercise due to speed-accuracy trade-off?

The results are summarized for these questions in chapter 10. The more detailed questions for analysis were used to investigate the effects of different variables on reaction times (RT) and accuracy (AC) with statistical tests, and the relationship between reaction time and accuracy, speed-accuracy (SA) trade-off, see Appendix 7 Questions for analysis.

8.3. Objectives of subjective measurements

Subjective measurements are collected with interviews (IN), questionnaires (QU) and subjective workload rating NASA-TLX (TLX).

The purpose is to collect subjective information about:
a) Which cue modality is preferred and why? Which cue modality is worst and why?
b) Which was the preferred cue for giving direction?
c) Experienced effects of physical exercise?
d) Experienced strengths and weaknesses of different modality cues?
e) Do subjects prefer the modality in which they are fastest?
f) Provide information on possible needs for improvements, and other feedback for the design

The results for these questions are summarized in chapter 10. The more detailed questions for analysis were used to investigate the preferences and experiences of subjects with questionnaires (QU, TLX) and interviews (IN), see Appendix 7 Questions for analysis.

8.4. Predictions

Based on the literature review there were some expectations of possible differences between modalities. In multisensory user interfaces, senses can be used separately and to complete each other. Cross-modal spatial cuing effects have been found between visual, auditory and tactile senses. The auditory cue for a visual target has been suggested to be the most robust cross-modal facilitating combination. (Ferris & Sarter 2008)(Driver & Spence 1998). Based on this an expectation was, that the auditory cues might facilitate most the shooting task requiring visual attention of the unimodal conditions. In relation to the visual stimulus cues in this study there were no hypothesis.

Tactile stimuli have been found to support spatial directing in driving simulation (Ho et al. 2006)(Scott & Gray 2008), and navigation (Colbert & Tack 2005) tasks. However, clear degradation in the detection level of tactile stimuli has been reported in demanding field conditions (Krausman & White 2006). Therefore, it was expected that especially after physical exercise the detection of the tactile stimuli might be worse than for the other modalities in the current study.

The multimodal stimuli in the current study refer to stimuli presented simultaneously on different sensory modalities. The subjects were predicted to be faster and more accurate after the multimodal cues, as multisensory stimuli have shown to be stronger for capturing attention than unimodal stimuli (Spence & Gallace 2007)(Santangelo & Spence 2008).

The mixed-modal stimuli refer to the task condition in which different modality stimuli are shown randomly. The reason for taking along the mixed condition was to investigate the effects of expectancy. These expectations may slow down the processing of a target when the stimuli appear on an unexpected sensory modality (Spence et al. 2001). Also the amount of errors may increase (Spence & Gallace 2007). Thus, the subjects were predicted to be slower and to make more errors in the mixed task.

In the task condition with physical exercise increased heart rate and heavier breathing were expected to have an effect on perceiving cues, and to impair the performance and subjective preferences.

In relation to perceiving the target directions, there were no specific hypotheses or expectations of differences between modalities.

9. EXPERIMENT

This section provides an overview of the field experiment followed by an explanation of the task, field setup, equipment, stimulus conditions, participants, detailed procedures and measures.

9.1. Overview

The field experiment took place in Santahamina, Helsinki during the period 4-8th of May 2009. Six soldier subjects having a strong shooting experience participated. After a training phase, each soldier performed 10 different task conditions; shooting sessions consisting of 30 shots. The sessions were supervised and commanded by a military officer. After each session a quick questionnaire was filled to record first impressions and the subjective workload. After all sessions the subjects answered a structured interview and an end questionnaire. The reaction time for shooting after each stimuli presentation was measured, as well as the shooting accuracy for each session.

9.2. Task description

An applied real world experimental setting was chosen. A shooting task simulated a close combat task and was performed in a military practice firing range. Shooting is a stressful and demanding activity and it is therefore suitable for simulating a crisis management task. It is a task requiring visual attention, and well rehearsed for the professional soldiers. The task designed for this study was not intended to correspond to a possible real combat task in the future. The task is not applicable as such, but it served only for the purpose of the experiment. In this study the possible applications of wearable user interfaces and related stimuli are out of the scope of discussion.

In the task, three visual targets were set to the left, centre and right in front of the subject. The stimulus cues presented via wearable user interface devices informed the subject on which target to shoot. For example, a sound to the left ear meant ”target left, shoot!”.

Here, "session" refers to a task where 30 shots were shot. Each task had one experimental condition. Before each session, the subjects were informed about the stimulus modality. They were also instructed to shoot as quickly and as accurately as possible.

9.2.1 Stimulus cue conditions

Each subject participated in 10 different task conditions, consisting of five different stimulus cue conditions with and without preceding physical exercise. The different stimulus cue conditions were visual (VI), auditory (AU), tactile (TA), multimodal (MULTI) and mixed (MIX) condition. In the multimodal condition auditory, visual and tactile cues were shown simultaneously, and in the mixed condition the cues were shown randomly. The stimulus cue conditions are referred as "modality variations" later in this document. In figures and tables the short names are used, prefix R referring to physical exercise before the task, e.g. visual modality variation with physical exercise (RVI).

9.3. Field setup

9.3.1 Field setup

The shooting range was in the Santahamina military base in Helsinki. Targets were reactive programmable popper targets, which went down when hit by a shot and got up immediately after that. In the current study this functionality enabled to see from distance whether a target was hit. In addition, a sheet of paper was put on each target in each session for recording the hits.

The targets were set 25 m from the shooting spot in three directions, to the left, centre and right, 15 degrees angle and 6,5 m distance from each other. The locations of the targets were chosen in the pilot experiment (see chapter 9.4). Figure 2 illustrates the field setup.

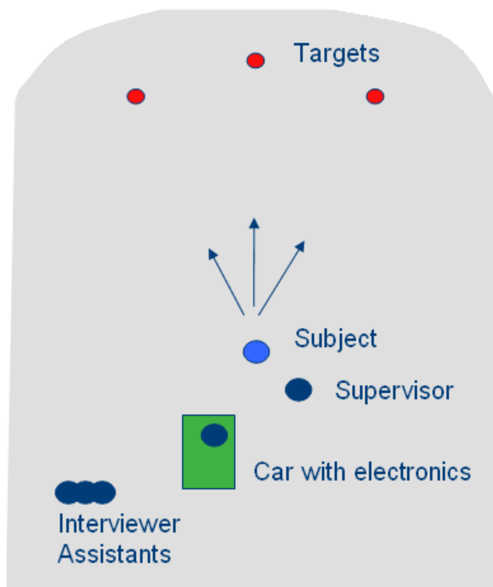

Figure 2 Experiment field setup plan drawing

A supervising military officer stood immediately behind the subject during an experiment while the rest of the staff was in the car where also the computers were placed. The car served also as a protection against changing weather conditions and enabled quick building of the field set-up in the same way every day. The software program operators could see the subject and targets from the car and also a video camera was set in the car. Occasional visitors to the experiment field were instructed to stay behind the car. The electricity required for the wearable user interface and experiment setup was available from a neighbouring shooting range and an uninterruptible power supply system (UPS) was used to ensure continuous and quality power. Figure 3 presents photographs of the field setup taken during the experiment.

Figure 3 Field setup during the experiment

9.3.2 Technical setup and functionality

A schema of the technical setup is illustrated in Figure 4, followed by a high level description.

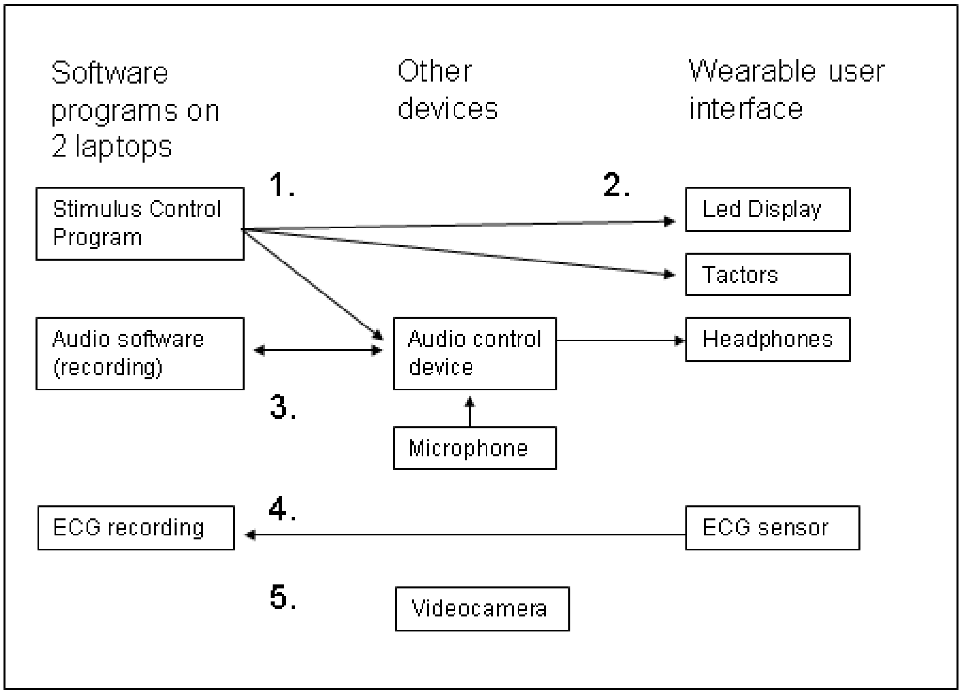

Figure 4 Schema of technical setup

At the beginning of each session all devices and software programs were started. During the experiment:

1. The Stimulus control program
   - sent an impulse to the wearable user interface device(s) to show a stimulus
   - sent an impulse to the audio software via audio control device to record time when a stimulus was presented
2. A stimulus was presented to the subject via wearable user interface device(s)
3. A microphone captured the sounds of the shots fired, and these were recorded with audio software
4. An Electrocardiogram (ECG) device was sending information to an ECG recording program independently of the other devices
5. A Video camera was recording the scene independently of the other devices

Stimuli were controlled by the Stimulus control program, which was manually started at the beginning of each session. For each session, one pre-prepared configuration file was used to define the order of stimuli presentation. The time between the cues varied randomly from 5000 to 7000 ms. The program sent impulses to the wearable user interface device and to the audio recording system simultaneously. The stimulus was presented to the subject and the audio recording system generated a sound, which was recorded as a start signal for measuring the reaction time of a shot. The program showed the progress of the session to the program operator on a computer display, and the operator informed the subject and supervising officer when the session was finished.

A reaction time extractor program extracted the stimulus start signal and the sound of the shot for each stimuli-shot pair from the sound recording. The program analyzed when the sound energy exceeded pre-defined thresholds. The stimulus start signal and sound of shot both exceeded the thresholds, and the time difference between them was the reaction time for a shot. This program was run after the experiments and the collected reaction time data for each session were also carefully checked manually.

The Stimulus control program and the Reaction time extractor were both developed by the research team in the Brain and Work Research Centre, Finnish Institute of Occupational Health.

9.3.3 Wearable user interface

The wearable user interface used in this experiment was a set of devices which presented tactile, visual or auditory stimuli to a subject. The wearable user interface consisted of a display, in-ear headphones, and a tactor belt.

Figure 5 A subject with wearable user interface and soldier equipment

The different stimuli devices are described below. All stimuli were 250 ms long. Chapter 3 "Human senses and information presentation with wearable user interface" describes in more detail the specialities of different modalities and how the stimulus cues for this study were chosen.

Visual display and stimuli

The visual stimulus device was constructed at the research centre, and consisted of protection goggles equipped with led lights presented in Figure 6. The stimulus cue was a blinking led light shown to one of the directions. One led light was placed on the left and one on the right side showing the stimulus cues for the corresponding directions. Two led lights were placed in the centre part of the goggles, one visible to each eye, showing together the centre direction.

Figure 6 Goggles equipped with led lights

Headphones and auditory stimuli

The auditory stimulus device was in-ear headphones (www.shure.com) used under active hearing protection, according to military standards. Audio stimuli were presented as stereo sounds to the left, the right, or to both ears simultaneously; the last perceived as centre. The frequency used was 1000 Hz, but the pure sine wave sound was edited to contain several short breaks, which was perceived as more agreeable and directing.

Tactor belt and tactile stimuli

The tactile stimulus device was a set of C2 tactors: electrodynamic actuators vibrating against the skin (Engineering Acoustics 2009). C2 is a vibrotactile display which mechanically stimulates the skin. It has been used in tactile navigation displays, and as sensory aids for the impaired (Jones & Sarter 2008). The stimulus is strong enough for even insensitive body sites. Three tactors were attached on a belt and worn around the waist, below the chest. For each direction there was one tactor; one in the centre and one on both sides, at a 15 cm distance from the centre location. The vibration frequency used was 230 Hz.

Figure 7 ECG sensor, tactor belt and tactor

9.3.4 Electrocardiogram

The electrocardiogram (ECG) heart rate measuring device (Alive Heart and Activity Monitor, Alivetec, Australia) was attached to the skin with tape. The ECG device had a wireless connection to a recording software program on one laptop.

9.3.5 Soldier equipment

The subjects were wearing full soldier equipment including bulletproof vest, equipment vest, gun strap, and a helmet. The gun used was an assault rifle (Rk 95 Tp, 7.62mm/ M95) with an Aimpoint red dot sight.

9.3.6 Participants

The subjects were voluntary professional soldiers with a long shooting experience in similar shooting tasks. One subject participated in the pilot experiment and 6 subjects in the actual experiment. In the actual experiment, the subjects were male between 27 and 52 years, average age being 33.4 years. They had a shooting experience ranging from 8 to 30 years, average being 16.2 years. The estimated amount of shots per year varied between 5000 and 10000 shots, average being 7500 shots per year. If considering that yearly shot amounts were constant, all subjects would have shot more than 60000 shots in their career, the most experienced even 150000 shots. All of the subjects train shooting regularly, weekly from 1 to 4 times, on average 2 times per week. All subjects were right-handed and right-eyed. They had shot before with a same kind of gun and a sight, as used in the experiment. They had not used wearable devices similar to led-display or tactor belt before, but some had regularly used headphones for receiving radio messages.

Due to the long experience in shooting, the shooting performance of the soldiers can be considered to be very constant. The new thing for them was the way the shooting instruction was given, via wearable user interface devices and with different kinds of stimulus cues.

9.3.7 Detailed procedures

The experiments were ran during several days, each subject having one half-day of introduction, and two half days of experiments. The subjects had 5 shooting sessions per experiment day. Based on information collected from the military contact person, any longer experiment sessions were not feasible due to the physical and mental workload of the tasks.

The introduction consisted of an information and training phase. The objectives of the research project were presented to the subjects, as well as their role in the study. The task, equipment and the schedule of the experiments were described to the subjects, and they filled in a start questionnaire collecting demographic data of e.g. shooting experience, age etc. In the training phase, the subjects first practised shooting to targets in the three directions without wearable user interface, commanded by the supervising officer. Then they trained the sighting of the targets with an unloaded gun to learn to recognize the stimuli presented on wearable user interface.

While subjects were getting familiar with the wearable user interface their stimuli perception was also evaluated in order to ensure that each of them could perceive all stimulus cues correctly. The audio stimulus volume level was adjusted to be similar to the perceived strength of the tactile cue, and to the perceived strength of the visual cue based on feedback given by the subjects. One volume level was considered good in the beginning by all the subjects but during the experiments the volume level needed to be adjusted higher for two subjects, as they had problems hearing the audio stimuli clearly. The position of led lights was changed for the experiment based on the feedback received during the training day - the led lights were originally located too high for a few subjects. For the purpose of ECG analysis, a heart rate baseline record was measured during the introduction day.

All experiment sessions followed the same procedure. First a baseline resting heart rate was measured for each subject. Then he put on the wearable devices with help from assistants, and got the instructions for the session. The subject was informed about the stimulus modality and was instructed to shoot as quickly and as accurately as possible. Physical exercise was performed in relevant task conditions, and the heart rate was monitored (ECG) to control that similar physical effort was present in all modality variations for all subjects. Next, the subject went to the shooting spot and prepared for shooting. The shooting supervising officer followed the preparations. When the subject was ready, the supervising officer gave a sign to the Stimulus control program operator and gave the subject permission to start shooting. The operator started the task. The supervising officer followed the shooting, ready to interrupt the task if required. At the end of the session, the operator informed the subject and the officer that the task was finished. After each task condition a quick interview was performed to record first impressions, and a NASA-TLX form was filled. In between sessions, the subject was allowed to rest for as long as he needed. After all session the subjects answered to a structured interview and to an end questionnaire. Table 1 lists the roles and responsibilities in the experiment procedure.

Role	Responsibilities
supervising officer (military staff)	Stood behind a subject and supervised his actions. In the beginning of each session he asked the subject to prepare, and informed the program operators when experiment could be started
stimulus control operator	Started the program when everyone was ready. Followed the experiment and communicated with the clerk about exceptions such as when a stimulus was missed. Informed the supervising officer when the session was finished
clerk	Recorded hits and missed shots manually on a session recording list. Also shots to wrong directions were tracked
audio recording and ECG recording program operator	Started the programs before each session and saved the files after it according to planned naming rules. Informed the stimulus control operator and supervising officer the heart rate level during the physical exercise. He also operated the video camera
assistants	Put and collected the target papers for each session, helped the subjects to put on wearable devices etc.

Table 1 List of roles and responsibilities

Five persons were needed for all sessions to run the experiment, each having their own responsibilities. There was a detailed time table planned in advance for each day which included all tasks. It also served as a checklist and was meant to enable a smooth process of the experiment. The plan kept well and only minor adjustments were made. One experiment session was repeated due to one of the targets catching fire from a bullet hit.

Figure 8 The reason for repeating an experiment session

9.4. Pilot experiment's purpose and results

The pilot experiment took place in Santahamina, Helsinki on April 9th, 2009 with one soldier subject. The main purpose was to test the experiment setup prior to the actual experiment. Before the pilot experiment, some pre-testing was made to choose between options for the targets, and between some stimulus cue versions.

The pilot experiment was used to
- select and optimize target devices and their positions
- select the best stimulus cues
- make a test run of the experiment
- collect and evaluate the subjective experiences of a soldier subject

Reactive programmable popper targets, which went down when hit by a shot and got up immediately after, were evaluated, and found to be working fast enough for the task requirements. Reactive targets enabled to see the hits from distance and thus they were considered better than non-reactive target options, facilitating scoring. After the pilot experiment, the time between stimuli was lengthened an additional second, to avoid the case where a moving popper target would disturb perceiving the next stimuli.

The locations of the targets were chosen to enable shooting at all directions without needing to move the lower body. When the shots to the centre and both side targets could be delivered just by rotating the upper body, the performance times for all shots could be treated as equal.

The stimulus cues selected based on theoretical background and on trials in laboratory environment were evaluated in the field setup during the pilot experiment. The subject harnessed with soldier equipments and carrying the wearable devices evaluated different stimulus cue options without performing the shooting task. The best cues were selected for the experiment.

For the experiment a detailed schedule was planned in advance. Most of the aspects of experiment were included; the field setup, technical setup, software programs, wearable devices, tasks, stimulus cues, interviews, experiment procedures, and scheduling were tested. The performance measurements were collected and their quality evaluated afterwards, but no analyses on this data was made.

In the pilot experiment the visual cues were presented with a monocular HMD, see Figure 9.

Figure 9 Monocular HMD

The subject had trouble to perceive the cues. First, the sunlight strongly impaired the visibility of the display, and the experiment session was interrupted because the subject could not see the cues at all. Sunshine protection was added to the helmet enabling the subject to see the cues. However, he then needed to focus his gaze and attention to the display and then refocus to the targets. He experienced this as difficult, and told that it impaired his awareness of the surroundings. It was concluded that another kind of display should be used for the actual experiment, as the visual cues presented on HMD were not found suitable for the shooting task: the sighting was disturbed and interrupted by perceiving the cue. See chapter 11 for further discussion on these results.

The auditory cues were working fine for the subject in the pilot experiment, but there were some issues with the tactile cues. Sometimes the subject did not feel the cues properly, he perceived as if the whole belt would be moving. Identifying the directions was difficult and required considerable concentration. The subject also missed a tactile cue, when he changed his position and fixed the equipment. In the pilot experiment the tactor belt was worn on a T-shirt. It was concluded that in the actual experiment the tactor belt should be worn directly on the skin, and if necessary, attached with skin tape.

All cues were shot to the correct direction in the pilot experiment but there were more missed shots than estimated in advance (24% of all shots missed). For the actual experiment, a training day was decided to be added to enable the subjects to get used to the devices and cues. The objective was to ensure as constant performance as possible for the actual experiments.

During the pilot experiment it was realized that as the shooting range used was surrounded by military practice grounds, there were also regularly other shooting sounds than the ones from the experiment. Therefore, the thresholds of audio recording were revised and the clerk was advised to write down any louder sounds from the environment to the recording list. It was concluded that the reaction times would need to be manually checked prior to the analysis in the actual experiment.

In the pilot experiment, the task, experiment procedures and scheduling were found to work as planned and the wearable user interface devices, computers, and targets were working well. Most measurements worked as designed, although some improvements were listed to be made before the actual experiment. A few questions were revised and a few added to the end interview. In addition, an end questionnaire was decided to be added in order to record the subjective experiences also quantitatively, and to enable better subjective ranking the modalities.

The effect of shooting and exercise on heart rate was evaluated in the pilot experiment. The non-statistical analyses of heart rate were made with the signal obtained from electrocardiogram (ECG) by another research team member of the Brain and Work Research Centre. An increased heart rate could be observed during the shooting task even without physical exercise, and the heart rate remained high during the task. See Figure 10.

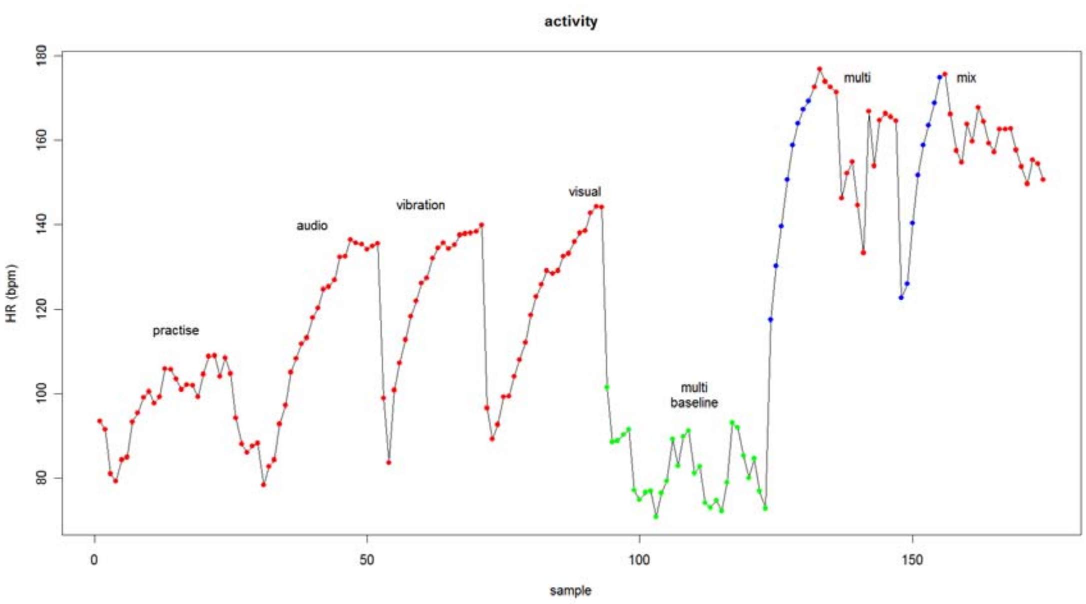

Figure 10 Heart rate during shooting (red), resting (green), and physical exercise (blue)

These differences were not in the scope of analyses in the actual study, even though the ECG data was collected. This method was included to investigate possible methods to measure stress effects in an experiment simulating a crisis management task. In the actual experiment the heart rate was used to control that similar physical effort was present in all modality variations (with physical exercise) for all subjects.

10. RESULTS

10.1. Performance measurements

Performance measurements included reaction time and shooting accuracy, and analysis of speed-accuracy trade-off was made based on these data, see chapter 6 for related methods. The field setup and the procedures to collect reaction times are described in chapter 9.3 "Field setup". Task conditions referring to different experimental conditions are described in chapter 9.2.1 "Stimulus cue conditions".

The response time was measured for each shot, it is ratio data presented in milliseconds. The reactions times analyzed with statistical tests are all hits, shots that hit the target. Only the shots with a reaction time were included to analyses, thus the shots which were not shot at all did not influence the percentage. The missed shots are analyzed as part of accuracy. Accuracy refers to percentage of hits of total shots in one task condition, thus it is related to a task, not a single shot. Analysis of speed-accuracy trade-off was based on reaction time and accuracy data. The data collected from performance measurements are shortly presented here before presenting the results. Table 2 presents the overview of performance measurement data.

Planned shots	Not shot	Target hit	Target not hit	Wrong direction	No reaction time
1800	12	1619	181	3	5

Table 2 Overview of performance measurement data

1619 shots hit the target, for 1614 of those there was a reaction time. 5 reaction times were missing due errors in recording program or its usage. 181 shots did not hit the target, three of those were shot to the wrong direction: two multimodal cues in a MULTI task, and one auditory cue in a MIX task. 11 tactile cues were not shot because they were not detected, 10 of those in task conditions with physical exercise RTA and RMIX, and one in TA. Three different subjects missed tactile cues. One visual cue was not shot due to a misunderstanding of instructions in the beginning of the task.

The reaction time means ranged between 1744,61 ms and 1969,64 ms and the accuracy means for tasks ranged between 86,67% and 93,33%.

10.1.1 Summary of performance measurement results

Here the performance measurement results are summarized per question (a-g) defined in chapter 8.2. Reaction time and accuracy data, and speed-accuracy trade-off were investigated per question for analysis defined in Appendix 7

Questions for analysis", the related more detailed results are presented in Appendix 7.

a) Does the modality or task have an effect on reaction time?

The reaction time differences were compared in different combinations of modality samples. Statistically significant difference was not found between unimodal groups (questions RT1, RT2b, RT3a, see Appendix 9)

Statistically significant difference (p= .035) was found with the Friedman test between modalities when multimodal group was included (RT2a). The pairwise comparisons showed p=0.28 significance value for difference between multimodal and tactile modalities, however after the Bonferroni correction the significance level was p< .017 and therefore result was not statistically significant. The Bonferroni correction is known to be conservative especially when samples are dependent thus in this evaluation the result can be reported as indicative. To visualize the differences the sample means and 95% confidence limits for six subjects are plotted below in Figure 11 indicating that the tactile modality was slower than the multimodal cue. However, see f) for related speed-accuracy trade-off analysis.

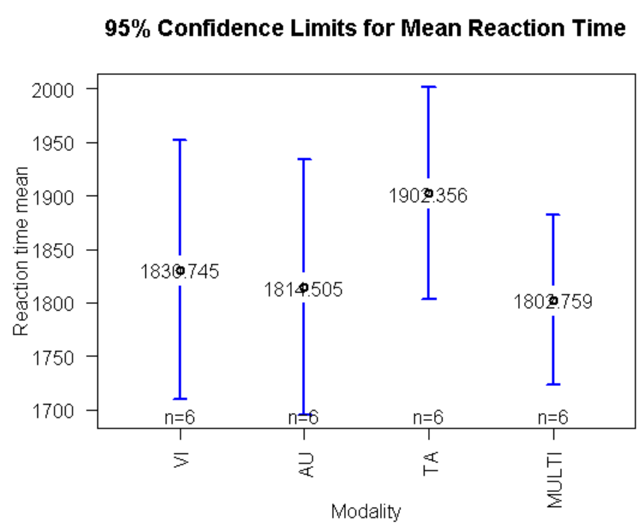

Figure 11 RT2a. Reaction time means and 95% confidence limits for visual, auditory, tactile and multimodal modalities

Statistically significant difference was not found between unimodal cues presented in unimodal tasks and mixed tasks (RT3b). In the mixed task condition the reaction time modality ranking seemed to be in same order than in the unimodal task condition.

b) Does physical exercise have an effect on reaction time?

Statistically significant difference was not found between the conditions with physical exercise and ones without physical exercise for any of five modality variation pairs (RT4). The comparison of modality variations with each other visually (Figure 11) showed a trend that for most of the tasks with physical exercise the reaction time means were slower. Even if not statistically significant, the mean reaction time difference indicated possible larger difference between visual cues with and without physical exercise, and smaller difference for auditory cues. However, see g) for related speed-accuracy trade-off analysis.

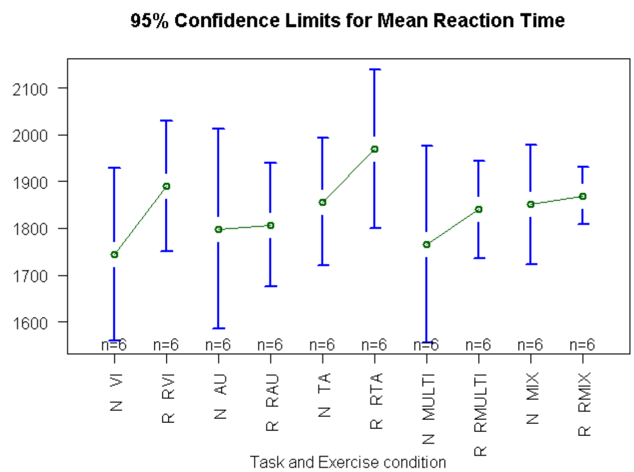

Figure 12 RT4. Reaction time means and 95% confidence limits for tasks with and without physical exercise

c) Does target direction have an effect on reaction time?

The reaction time differences between directions were compared per modality variation, and then the results of different modality variations were compared with each other (RT5). The reaction time difference between directions in modality variations was significant for auditory modality (p= .030) and non- significant for other modality variations. For auditory data, the Wilcoxon signed ranks tests were used for a post hoc test. There were three tests completed, comparing the directions. The Bonferroni correction was applied and effects are reported at a .017 level of significance. The reaction times did not significantly differ between the compared pairs, although the significance value between left and center cue was p= .028. The Bonferroni correction is known to be conservative and thus the difference can be considered as indicative. The comparison of modality variations with each other visually (Figure 13) indicated that for most modality variations the side directions seemed (statistically non-significantly) slower than centre direction, but for auditory modality variation the centre seemed to be slower. See chapter 11 for further discussion on these results.

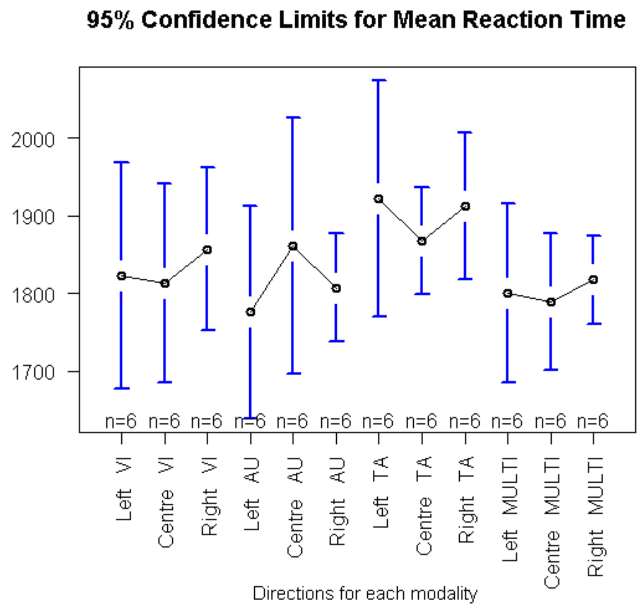

Figure 13 RT5. Reaction time means and 95% confidence limits for directions per modality variation

d) Does the task have an effect on accuracy?

The accuracy difference between tasks was not statistically significant, no effect was found (AC1).

e) Does physical exercise have an effect on accuracy?

Accuracy did not statistically significantly differ between tasks (see d)). The comparison of modality variations with each other visually (Appendix 9, AC2) showed that for most tasks the trend was that accuracy was poorer after physical exercise, for auditory task the difference appearing to be larger than for other tasks. For visual task accuracy appeared however to be better after physical exercise. These differences were not statistically significant. See also g) for related speed-accuracy trade-off analysis.

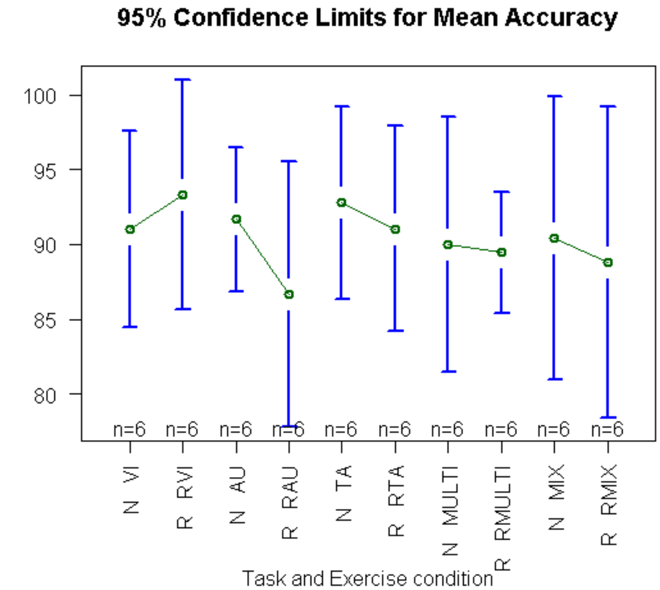

Figure 14 AC2. Accuracy means and 95% confidence limits for tasks with and without physical exercise

f) Are the possible differences in reaction time due to speed-accuracy trade-off?

In overall analysis of reaction time and accuracy there was no speed-accuracy trade-off found. Figure 15 does not suggest clearly either positive or negative correlation. Also partitioning analyses did not give significant difference (see SA1 in Appendix 9). Analysis in task level showed that there is no trend that the fastest task would have the lowest accuracy or opposite. There were errors also among the slowest shots.

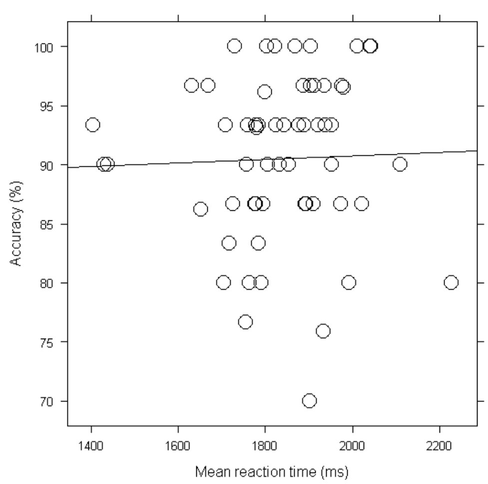

Figure 15 Correlation between the mean reaction time (also missed shots) and accuracy

Indicative difference in reaction time between tactile and multimodal modalities was found in one question for analysis (RT2a), which was not on task level analysis. Speed-accuracy trade-off analysis could be made on task level only, because the accuracy data were collected on that level.

The reaction time means (task level) and accuracy means for tactile and multimodal tasks were analyzed for speed-accuracy trade-off (SA2), see data in Table 3.

TA	1856,54	92,78%
RTA	1969,64	91,05%
Mean	**1913,09**	**91,92%**

MULTI	1765,92	90,00%
RMULTI	1840,11	89,44%
Mean	**1803,015**	**89,72%**

Table 3 SA2. Reaction time and accuracy means for tactile and multimodal tasks for speed-accuracy trade-off analysis

Tactile modality cues were in average slower and more accurate than multimodal cues. This indicated that there could be a speed-accuracy trade-off and it is not possible to put the tasks in a ranking order by difficulty or reaction time. The add-hoc ranking method (see SA1 in Appendix 9) also gave only small difference between tactile and multimodal tasks, which confirms this result.

g) Are the possible effects of physical exercise due to speed-accuracy trade-off?

The reaction time means for conditions with and without physical exercise were compared together with accuracy (values is Table 4). Each modality variation was analysed separately in Table 5.

	mean rt	accuracy
VI	1744,61	91,03%
AU	1798,07	91,67%
TA	1856,54	92,78%
MULTI	1765,92	90,00%
MIX	1850,64	90,42%

	mean rt	accuracy
RVI	1890,16	93,33%
RAU	1807,43	86,67%
RTA	1969,64	91,05%
RMULTI	1840,11	89,44%
RMIX	1869,61	88,80%

Table 4 SA3. Reaction time and accuracy means for all tasks with and without physical exercise

Conditions	Analyses of speed and accuracy relation	Add-hoc ranking method points per task
VI- RVI	RVI is slower and more accurate than VI. There could be speed-accuracy trade-off and it is not possible to put these tasks in order by difficulty or reaction time.	VI - RVI 17 14
AU-RAU	RAU is slower and less accurate than AU. RAU may be considered as a harder task.	AU - RAU 16 9
TA-RTA	RTA is slower and less accurate than TA. RTA may be considered as a harder task	TA - RTA 16 8
MULTI-RMULTI	RMULTI is slower and less accurate than MULTI. RMULTI may be considered as a harder task	MULTI - RMULTI 15 11
MIX-RMIX	RMIX is slower and less accurate than MIX. RMIX may be considered as harder task	MIX - RMIX 12 8

Table 5 SA3. Analysis of speed and accuracy relation for all modality variations, comparing conditions with and without physical exercise

The reaction time difference for conditions within modality variations was analyzed, and it was found that for visual modality the difference could be due to speed-accuracy trade-off. Visual modality variation was found faster in condition without physical exercise, and more accurate in condition with physical exercise. Speed-accuracy trade-off may explain this difference for visual modality variation, as well as the larger reaction time difference between conditions. For other modality variations no evidence of a speed-accuracy trade-off was found, and the trend was, that accuracy was poorer after physical exercise. It is also noteworthy that whereas the reaction time difference of conditions observed for auditory modality variation was small (see Table 4), the accuracy difference was larger.

The trend (statistically non-significant) that most modality variations performed better in condition without physical exercise could be observed also in speed-accuracy trade-off analyzes.

10.2. Subjective measurements

Subjective measurements included interviews and questionnaires described in chapters 7. The procedures how the data were collected are described in chapter 9.3.7.

10.2.1 Questionnaire

The original questionnaire in Finnish and its translation in English can be found in Appendix 1 and Appendix 2. The questions are listed below in Table 6 as well, their ID numbers are used later in this chapter to refer to them. The area refers to topic which each question investigates.

Area	ID	Question
Cue	1	Cue was easy to identify
Cue	2	Cue felt pleasant
Cue	3	Cue felt annoying *
Cue	4	Cue felt reliable
Cue	5	The time between stimuli was long enough
Cue	6	Cue felt intuitive
Cue	7	Cue was easy to learn
Directions	8	Different directions were easy to identify
Task	9	Cue disturbed sighting *
Task	10	Cue disturbed concentration *
Task	11	Cue helped in shooting task
Task	12	Cue felt to speed up shooting
Task	13	Cue might be good in real use
Task	14	Cue felt unsuitable for task *
Task	15	Cue device affected your shooting position or way of shooting *
Physical exercise	16	Cue was clear in a task after physical exercise
Physical exercise	17	Cue was clear in a task without physical exercise

Table 6 List of questionnaire questions. The questions marked with asterisk were negatively formed and therefore the ratings of those were inversed to same scale to be able to compare those with other questions.

Questionnaire data were investigated per question for analysis (QU1-QU3) defined in Appendix 7 Questions for analysis". The results are summarized in chapter , and presented in detail in Appendix 9.

10.2.2 Interviews

Interview data collected from quick interviews and structured end interviews were analyzed together. Comments for each modality from all subjects were collected and summarized per question for analysis (IN1-IN6) defined in Appendix 7. The results are summarized in chapter 10.2.4 together with questionnaire data.

IN1. Visual Cue

Subjects told that the goggles or led lights mounted on those were not disturbing sighting and aiming during the shooting task. Some subjects reported dead angles on the sides in the beginning, but afterwards they told that they had quickly gotten used to the devices. However, one subject brought up that the goggles or the led lights might start disturbing shooting in a normal combat situation because they hide part of the field of vision.

There were some issues in relation to perceiving or identifying directions. Separation of the left and the centre cues was difficult for two subjects probably due to right eye-dominance. From the point of view of right eye, both the centre and the left stimuli appear to be on the left side, almost at the same place. The issue might be mitigated if the side cues were placed further away from the centre. However, the subjects did not make mistakes in identification of cues.

Another issue was related to the fit of the goggles when the sun was shining. Two subjects reported that they fixed their eyes in a particular way to be able to see the cues. One of them explained:

> *"I could not have observed surroundings because I would have had to concentrate to see the cue."*

The subjects were observed to draw their eyebrows together and down due to the sunshine in a way that the centre stimuli were not well visible to them. The led-lights were then placed too high, the eyebrows covering the sight.

The sunshine impaired the perception of led-light cues also in other ways. The subjects commented that the cues were less clearly and brightly visible, and there were disturbing reflections on the goggles as well. Most of the subjects could identify the cues in the sunshine, but it felt harder and they needed to concentrate more on it. Therefore, some subject who preferred the visual cues on a cloudy day changed their opinion on a sunny weather. The reflections of sunshine on the goggles made one subject doubt if he saw a cue or just such reflections, especially just after having blinked the eyes. In addition, one subject told that it might be difficult to know in a real combat situation whether a sudden light flash in peripheral vision would be from the led-lights or from another source in the surroundings.

Further issue experienced by some subjects was that the goggles were getting steamy inside, making sighting the target more difficult, especially when targets were against the light. There was more steam after the physical exercise. The subjects had to move the head position to see properly. As the goggles were worn in all sessions, this issue did disturb also in other tasks. However, half of the subjects agreed that the visual cues seemed to be less affected by the physical exercise than tactile and auditory cues. There were no other comments given of any impairment of the cues.

All subjects reported that they adapted the way they looked, in order to see better the visual cues. They were avoiding blinking their eyes during the experiment session, because they had an impression that otherwise they might miss seeing a cue. Even if the time between cues in one session varied between 5000 and 7000 ms, subjects could approximately know when the next cue was to be expected. Keeping the eyes constantly open was overstraining for them. In addition, controlling of blinking also felt to take some concentration away from the task. However, only one subject expressed that cues were quite short.

> *"If I blinked normally I would probably miss cues but of course this problem could be fixed by showing a cue more than once."*

250 ms is indeed very short time and could remain unnoticed because of a blink of an eye. However, the subjects did not miss visual cues, not even in mixed task where they could not expect visual cues in a same way than in unimodal tasks.

A subject who did not prefer the visual cues explained that he uses mainly the visual sense for observing surroundings and looking for the targets. He acts much based on his sight, and thought that anything additional in the field of vision could be disturbing. He said he would probably at least require training to get used to that kind of change in the way of operating in real situations.

IN2. Auditory Cue

Half of the subjects found that auditory cues were intuitive and easy to use, partly because the subjects were used to getting information via radio. A subject who liked auditory cues thought that he did not need to think on perceiving the cues, but he could concentrate on observing the surroundings and on shooting.

Half of the subjects considered that the both side cues were very clear, but that the centre cue was less clear, less intuitive. One of the subjects described:

> *"When sound came to both ears I did not first understand it meant the centre direction. Instead, this sound was just a peeping sound in my head, and did not seem to give any direction."*

In addition, two subjects reported other difficulties to recognize directions, even if they heard sounds very well. One of them, a subject who had impaired hearing, could however recognize directions better when volume was increased. One shot was shot to wrong direction, to the right instead of centre.

Half of the subjects considered that physical exercise did not have an effect on perceiving the auditory cues. One subject however commented that there was some difference:

> *"Heavy breathing made hearing more difficult. The harder the physical exercise would get, the more difficult hearing would probably be."*

The in-ear headphones were difficult to put on, but wearing them was not disturbing. A few times the headphones were not put on well, and they were almost falling off, which affected on how cues were heard. One subject complained after the experiment that the chosen soft sleeve on headphone was too big and therefore uncomfortable. Each of these issues however could be avoided by adapting current device.

When considering the usage of such auditory cues in real situations, two subjects thought that usage of auditory cue might be difficult in situations where hearing is needed for something else. They mentioned communication, and observing the surroundings. They thought that in such situations it might be more difficult, or slower to perceive an auditory cue. On the other hand, an auditory cue could disturb the other tasks. In contrary, one subject considered that that the cue used in the experiment could be usable in real situation as it is so different from human speech.

> *"It might be possible to notice it even if listening to radio at the same time."*

IN3. Tactile cue

Half of the subjects considered that they needed to concentrate the least for the tactile cues, and that identifying the directions did not require thinking. The cue did not disturb sighting or shooting.

> *"When I knew that there were only vibration cues coming, I could concentrate on sighting the targets; concentrate on those and when the vibration came, I could just act."*

The main issue with tactile cues was that some subjects indeed missed several cues. They just did not perceive those. Some subjects reported that they had an impression that the belt was not properly attached and that it could be moving. One subject said that he felt like he was not daring to breathe normally, for not missing a cue. He told that tactors were moving, and that this had an instant effect on how strong he could feel the tactile cue stimulus. Also another subject reported that stimulus felt differently, if he was moving at all. After physical exercise, when he needed to change position more often, he felt the cues worse. For him, the belt was checked during the experiment, and it appeared to be attached properly. Several subjects reported inconstancies in clarity of the tactile cue in different experiment sessions, and suggested that it could be because of the tactor belt position or attachment, apparently not being ideal.

Several subjects perceived the side cues better than the centre cue, and that they adapted their position to feel that better.

> *"If I leaned my hips ahead then I could better feel the belt, and cue in the centre."*

> *"I don't know if the contact to the skin was bit poor. It could be, because of this shooting position, when I was bowed down. At this readiness positions the abs tighten and maybe the belt is then little bit loose, of the skin and the vibration is not easy to perceive. On the sides it is better."*

> *"The positions is little bit hunched up and the tactor is above abs in the front, and ones on the side are rather above ribs. Those on the sides were easier to perceive than the one in the middle although it was not a*

big difference. I think that the weight of the equipment pushed the tactor belt downwards."

At least one of them had well defined abdominal muscles, and tactors on the belt possibly did not touch the skin.

One subject felt tactile cues quite poorly constantly during the experiment, and especially the side cues. He said he needed to think whether the something he felt was a cue or not.

"In this multimodal task, the tactile stimuli did not make any difference. I could not identify those, especially when I needed to change my position."

One subject reported that shooting position made distinguishing the left and centre cue from each other more difficult. The left side of the body was slightly more forward, and therefore the left cue was aiming similar to the centre cue, which he found confusing. Two subjects said that the cues would be clearer, if the tactors on the sides were further away from the centre. Also the vibration could be more intensive. However, half of the subjects considered that the tactile cue gave the direction best.

One subject perceived also a sound when tactile cue was presented even though the active hearing protection volume was off.

"I felt along my body a vibration which then changed to sound, like "prrr". Then sometimes right after shooting, I still felt something, and was not sure if it was a new cue or still the previous one. It was some kind of after-effect."

Tactile cue was experienced as a different kind of cue, special in comparison to other modality cues in the experiment.

"You do not need tactile sense for perceiving anything else actively in the environment. But other senses you need; with your eyes you are looking for targets and observing surroundings, and probably you would need to listen as well."

The cue was not found very natural and being far from the world of shooting, different in comparison to other stimuli coming from the environment. This may enhance identifying the cue. For the same reason, one subject considered that it took him some time to connect the cue to the task, for him it was not intuitive.

One subject who preferred tactile cues said:

"I mostly perceive and act based on what I hear and see, and if there would be something else blinking in vision, or heard it might make action more difficult.

One subject who felt the cues quite well, estimated that tactile cue would only work in a situation like in the experiment; when not moving much and not doing something else at the same time.

Two subjects reported that puffing or heartbeat made perception of the tactile stimuli more difficult.

> *"The higher the pulse, the more difficult it gets to perceive the cue, and the more exhausting the physical exercise is, the more difficult it could get. I think it would already have an effect on normal combat situation."*

One subject reported that stimulus felt differently, if he was moving at all. After physical exercise, when he needed to change position more often, he felt the cues worse. Some subjects did not feel a difference betweens conditions with or without physical exercise.

IN4. Multimodal cue

Half of the subjects mentioned that the multimodal cues were easy to perceive, and when all the stimuli were presented at the same time, it did not feel possible to make a mistake. The different stimuli strengthened one another.

> *"I do not have to concentrate on perceiving stimuli, but I can attend to the targets."*

The subjects reported that they preferred some of the modality cues, which they then mainly used for directing. Some adapted these preferences depending on weather (sunshine) or task conditions (physical exercise). Some even combined modalities, preferring one for the side cues and another for the centre cues.

> *"I had difficulties to distinct left and centre light cues. I started directing with the light because I perceived it faster, and then used sound to ensure. I always used auditory cue to identify the side direction, and light for the centre. I did not pay attention to tactile stimuli."*

Another subject told that he paid as much attention to visual and tactile cue but less for auditory cue.

Subjects did not report negative issues in relation to multimodal cue, except one comment:

> *"The stimulus is so comprehensive that it takes attention away from anything else at that moment."*

IN5. Mixed-Condition

Two subjects found mixed condition troublesome because they did not know which cue to expect. Different kind of cues shown randomly felt confusing and the interpretation of them was not easy. The subjects considered that tactile stimulus cue was then easiest to recognize because it was so different.

On the other hand, one subject reported that he did his most relaxed performance in a mixed task because he had no expectations of how the cues would be presented. He could just react to the cue when perceiving it, and otherwise concentrate on shooting. In comparison, in a unimodal tasks, he felt he was more concentrated on perceiving a cue while waiting for it.

When describing their experience for a mixed task, the subjects often compared how they felt about different modalities and mentioned their strengths and weaknesses. The modality related comments are reported together with modality specific cues even though some issues might only have appeared in mixed condition.

IN6. Miscellaneous comments

One subject reported that in the beginning of the experiment he needed to concentrate to not to shoot to a target, when it automatically got up after having been down after a hit. He told that in several shooting trainings he had to shoot such target when it got up, so he was used to that.

The subjects learned fast the rhythm that was used to give the cues, and adapted their concentration and attention accordingly. In general, it took some time to get used to the cues.

> *"In the beginning, I was thinking if something I perceived was a stimulus or not, I needed to concentrate on it. However, the more I have shot, the better I could concentrate on shooting. And I think that with bit more exercise I do not have to think other thing than shooting."*

One subject mentioned that some of his poor shoots were because his concentration was getting enervated by not having long enough breaks. He explained that there are mental capacity limitations for shooting, and people are only able to shoot ca. 50 shoots with full concentration at the time, and then they need to have a considerably longer break in order to achieve the same concentration level again.

10.2.3 Subjective Workload Questionnaire NASA-TLX

The subjects rated the workload by completing the NASA-TLX 20-point scale questionnaire after each task. See the questionnaire in Appendix 4, and methods in chapter 7.2. Questionnaire data were investigated per question for analysis (TLX1-TLX3) defined in Appendix 7. The results are summarized in chapter 10.2.4, and presented in detail in Appendix 9.

10.2.4 Summary of subjective measurement results

This chapter summarizes the subjective measurement results per question (a-f) defined in chapter 8.3.

a) Which cue modality is preferred and why? Which cue modality is worst and why?

The questionnaires showed (QU1, QU2, see Appendix 9) that the differences between modalities (VI, AU, TA) were not statistically significant (p= .958). That does not mean the subjects did not have preferences but that their preferences differed from each other. The differences between modalities were therefore evaluated also per subject. Five of six subjects had significant difference between modalities on individual level. The interview results did not make it possible to put modalities in preference ranking either. The subjects had quite different experiences of the cues and devices.

b) Which was the preferred cue for giving the target direction?

In the questionnaire, there was no difference at all between modalities for spatial directing (QU2). all modalities got the same rather good rating for easiness to identify the directions (q8, mean 5,67 which is between "agree a little" and "agree"). In the interviews, the subjects reported different kind of experiences with different modalities.

With visual cues there were some issues in relation to perceiving or identifying directions. Separation of left and central cues was difficult for two subjects probably due to right eye-dominance. From the point of view of right eye, both centre and left cue are on the left side, appearing to be almost at the same place. The issue might be mitigated if the side cues would be placed further away from the centre. However, the subjects did not make mistakes in identification of cues.

About auditory cues half of the subjects considered that the both side cues were very clear, but that the centre cue was less clear, less intuitive. In addition, two subjects reported other difficulties to recognize directions, even if they heard sounds very well. One of them, a subject who had impaired hearing, could however recognize directions better when volume was increased.

About tactile cues half of the subjects considered that they gave the direction best and that identifying the directions did not require thinking. One subject reported that shooting position made distinguishing the left and centre cue from each other more difficult. The left side of the body was slightly more forward, and therefore the left cue was aiming similarly to the centre cue, which he found confusing.

Half of the subjects mentioned that with multimodal cue there was no possibility to make a mistake as the different stimuli strengthened one another. However, two shots were shot to wrong direction. The subjects reported in relation to multimodal and mixed task conditions that they had preferences for some modality cues, which they mainly used for directing.

c) Experienced effects of physical exercise?

In the interviews some subjects considered that exercise before the task did not have effect on perceiving the cues, and some considered it did have an effect. In the questionnaire all modalities got worse mean ratings for cue clarity after physical exercise (QU2). The difference (statistically non-significant) was smallest for visual modality. Also in the interviews, half of the subjects found that visual cue was affected less than other modalities by physical exercise. They however reported that there was more steam on the goggles after physical exercise. Tactile cue was reported by several subjects to be difficult to perceive after physical exercise. It was impaired by moving, puffing and heartbeat. Auditory cue perception was impaired by heavy breathing (reported by one subject) as well. One subject commented that physical exercise had an effect on shooting, not only to stimulus perception. Feeling more tired physically, being out of breath during the task may impair the shooting performance.

The NASA-TLX questionnaire results (TLX1) showed that the overall workload ratings did not significantly differ between tasks (p=.087). The graphical comparison of modality variation pairs showed however a trend that subjects rated higher workload estimates for the tasks with physical exercise the, see Figure 16.

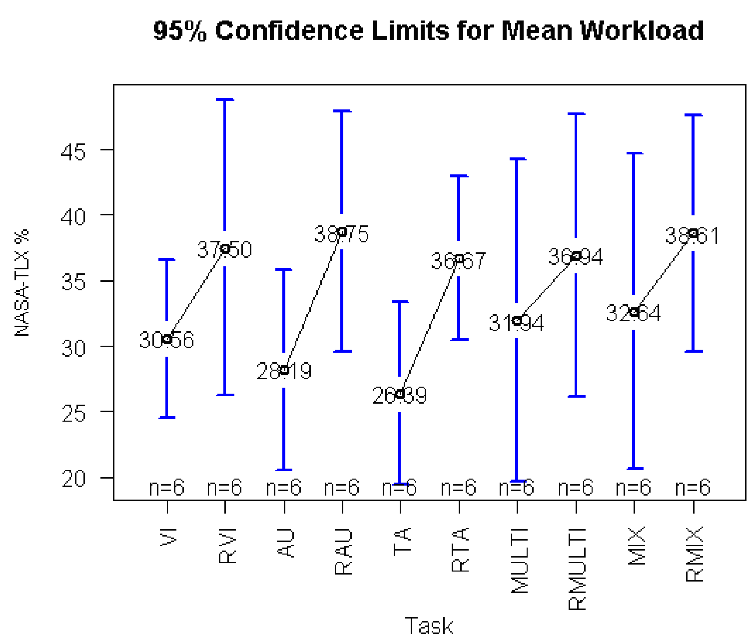

Figure 16 TLX1. Workload means and 95% confidence limits for tasks with and without physical exercise

Dimensions "effort" and "physical demand" contributed most clearly to differences between conditions with and without physical exercise, being close to significant (TLX2) and similar for all modality variations (TLX3). For "mental demand", "performance" and "temporal demand" the workload ratings also were higher in the physical exercise condition. For dimension "frustration level" the subjects rated workload to be similar in both conditions and for "mental demand" the mean workload rating were lower than for other dimensions.

d) Experienced strengths and weaknesses of different modality cues?

In the questionnaire (QU2) all modalities got for all questions only neutral or positive mean ratings: mean was at least "4", which corresponds the answer "Do not know, cannot agree nor disagree". Several questions got for all modalities mean rating above "6", corresponding "Agree". Based on these, it can be summarized that the subjects agree that the time between stimuli was long enough, the cue/device did not affect the shooting position or shooting, the cue did not disturb sighting, was easy to learn, and clear in task without physical exercise. Interview results in chapter 10.2.2 describe more in details the experienced weaknesses and strengths of different modality cues.

e) Do the subjects prefer the modality in which they are fastest?

The response times of each subject were compared to their questionnaire preferences (QU3). Four subjects out of six were slowest in their preferred modality, when comparing reaction time means of unimodal tasks to modality preferences. However, when comparing the mean modality reaction times of mixed-modal tasks to modality preferences, four subjects out of six were fastest on their preferred modality.

f) Provide information on possible needs for improvements, and other feedback for the design

The feedback from subjects could serve as valuable input for designing wearable devices and cues, see more details about interview comments in chapter 10.2.2, and about questionnaire results in Appendix 9.

10.3. Evaluation of fulfilment of the cue requirements

The requirements for stimulus cues for the experiment in the current study were specified in chapter 8.1. Here the requirements are discussed together with the empirical results to conclude whether the requirements were met. This analysis provides a more complete view to stimulus cues in the current study and completes the empirical results.

Cues:

Should support fast actions, should be short, and of the same duration for all modalities (comparability)

In the interviews, the only comments related to the length of stimuli were in relation to the visual stimulus. The subjects reported that they were trying to avoid blinking their eyes for not to miss cues. 250 ms indeed is a very short time for the visual cue and it could remain unnoticed because of a blink of an eye. Longer or repeated stimuli may be needed for visual modality. In the questionnaire, the subjects could not agree or disagree whether the cues felt to speed up shooting (q12).

Should be easy to detect

Performance data indicated that most of the cues were perceived.. However, performance data showed that in total 11 tactile cues were not detected at all. Even these are only 2,3% of all tactile cues, it is notable that cues were missed by three different subjects and that 10 of missed cues were in task conditions with physical exercise. Based on questionnaire (q4), the subjects agreed a little that cues felt reliable, and they did not consider cues very clear either (q16, q17).

Should not disturb sighting of the environment (visibility), Should not disturb shooting (position)

The questionnaire results show that the tactile and auditory cues were not considered disturbing sighting or concentration (q9, q10), nor the cue devices affected the shooting position (q15). Visual modality was rated slightly (statistically non-significantly) worse than other modalities, and in the interview results some possible reasons could be identified: the steam on the goggles deteriorated visibility and also the small led lights on goggles bothered some subjects before getting used to them. The led light in the vision field could bother sighting even more in a real situation where a wider area would need to be observed than in this experiment. Tactile belt was not at all bothering the subjects and neither were the in-ear headphones, except for one subject who complained after the experiment that the chosen soft sleeve on headphone was too big and therefore uncomfortable. There were wires used for the wearable devices in the experiment and therefore the subject's area of movement was limited. None of the subjects reported that these wires would have disturbed the tasks.

Should not be annoying

The questionnaire results showed that the subjects "agreed a little" (see questionnaire rating scale in Appendix 2) that cues did not feel annoying (q3). Visual modality was rated slightly (non-significantly) less annoying than other modalities.

Should be able to present the three required directions with a wearable device: indicate the targets

The questionnaire results showed that the cues presented directions equally well (q8). No significant differences were found in reaction time measurements between modalities either.

Should be simple, should be intuitive

The questionnaire results showed that the subjects "agreed a little" that cues felt intuitive (q6), and agreed that cues were easy to learn (q7).

Should provide similar level of "alert" regardless of modality: the modalities should be perceived as being of similar strength in comparison to each other (comparability)

The subjects did not point out during interviews that one stimulus would have been stronger or have higher level of "alert" than other. Based on interviews the cue modalities could not be put in order in relation to how strong they were experienced.

Should fit well with the task requirements

The questionnaire results show that the subjects agreed at least "a little" that cues were not disturbing concentration (q10), helped in shooting task (q11), and felt suitable for task (q14). The mean ratings for all questions for all modalities were at least neutral or positive, rating being at least "Do not know, cannot agree nor disagree".

11. DISCUSSION

11.1. Results

The major challenge with presenting spatially directing stimuli on a wearable user interface is that the technology currently available for wearable display devices cannot perfectly produce spatial stimuli. The stimuli presented on wearable user interface do not provide the same kind of reference to the environment than real stimuli from the environment (e.g. headphone sound vs. natural sound). Because of this, a stimulus presented on a wearable device is not perceived exactly in the same way as a stimulus presented in the environment. In addition, each sense has its specific features and different mechanisms of perceiving directions.

The stimulus cues and setup for the current study were designed and chosen based on the literature review on applied aspects of basic sensory mechanisms. Spatially directing cues have been previously studied in applied setups in relation e.g. to driving and navigation tasks.

Ho and colleagues studied auditory stimulus cues (Ho & Spence 2005) and tactile stimulus cues (Ho et al. 2006) in a driving simulation. They found that the spatially directing predictive warning cues support capturing spatial attention effectively, and facilitate the task performance; in that case the cues speeded up reaction times to braking or acceleration in danger situations while driving. Scott and Gray (2008) compared tactile, visual and auditory warning cues also in a driving simulation, concluding that the clearest facilitations in reaction times were after tactile cues.

Ferris and Sarter (2008) investigated cross-modal effects in a complex simulated military operation task, where the subjects acted out as vehicle commanders. Ferris and Sarter compared visual, auditory and tactile cues and found that the auditory cues speeded up the responses for visual targets, whereas tactile cues did not. They suggested that the auditory cue for a visual target is the most robust cross-modal combination. In addition, other studies show that auditory stimulus seems to work well to attract visual attention (Driver & Spence 1998).

Colbert and Tack (2005) compared visual, auditory and tactile displays for presenting localization information in a field setup. They used human factors methods, similar to the ones used in the current study. They found that even if the subjects preferred a visual display, and performed best with it, it deteriorated the situation awareness. Thus, Colbert and Tack report that in circumstances where the visual sense is already under heavy load, the tactile display may be preferable. A study of tactile stimuli in simulated combat movements however shows clear degradation in the detection level of tactile stimuli in demanding field conditions (Krausman & White 2006).

The author is not aware of research using wearable user interface devices to indicate the direction of a target in a shooting task, or in another speed critical task in a field setup. The hypothesis was that there are differences on the performance and in subjective preferences between modalities in the field setup. The empirical results of the field experiment showed that it was not possible to put modality variations in a ranking order neither by performance measures nor by subjective measures in the questionnaires and the interviews.

Based on the literature review there were some more detailed expectations of possible differences in the current study. Below, the results are presented and discussed with these expectations.

The visual stimuli were presented with a HMD in the pilot experiment. The HMD was found unsuitable for presenting the visual stimulus cue in a time critical shooting task, because the sighting was disturbed as the subject first had to focus the gaze to the display. In addition, the sunlight strongly impaired the visibility of the display. Problems with the light conditions might possibly be avoided by using a HMD with a higher luminance level. However, the issue of focusing the sight somewhere else than to the surroundings was considered problematic for objectives of the task in the current study. A HMD is commonly used for presenting information in wearable user interfaces in a military context, and therefore trying it out in the field setup was found necessary. When more complex and less time critical information is presented, a HMD probably has many advantages, but for a time critical combat task and for capturing spatial attention it does not seem to be the best option.

There were no expectations in relation to the visual stimulus cues in this study. The interview results indicate that the visual cues were possibly too short, because several subjects reported that they were trying to avoid blinking their eyes for not missing cues. A possible solution might be to make the cues longer. 250 ms long stimuli was used for all modalities, but such a visual cue is so short that it could remain unnoticed because of a blink of an eye.

Sunshine impaired the perception of the led-lights, and there were disturbing reflections on the goggles as well. The small led lights bothered some subjects before getting used to them. Led lights in the field of vision could bother sighting even more in a real situation where wider area would need to be observed. The goggles were getting steamy inside especially after physical exercise, making the sighting of targets more difficult. As the goggles were worn in all sessions, this issue was similar in all modality conditions. In the interviews half of the subjects found that the visual cues were affected less than other modalities by physical exercise, and the questionnaire also indicated the smallest difference between the conditions (statistically non-significant). Two subjects reported that the identification of left and centre visual cues felt more difficult probably due to right eye-dominance. The issue might be mitigated if the side cues would be placed further away from the centre. However, performance with these cues was no slower than with the right cues.

Based on findings of Ferris and Sarter (above), it was expected that an auditory cue stimuli might facilitate the shooting task to visual target most. In the current study, the performance with auditory cue stimuli was no better than other stimuli and this hypothesis was not met. In the interviews, the auditory cue stimuli were not considered disturbing sighting or concentration, nor affecting the shooting position, and they were intuitive and easy to use. However, the centre cue was found less clear and less intuitive. This was possibly because for some subjects the centre sound did not seem to come from any direction but it was perceived inside the head. Previous research has shown that for headphone listening it is typical that the sounds appear to localize inside the head rather than in the environment (Härmä et al. 2004)(Zhou & Rogers 2002). A subject who reported having hearing problems had problems identifying other directions as well, but an increasing volume appeared to facilitate the identification. One subject reported impairment in auditory cue perception in the condition with physical exercise due to heavy breathing.

Tactile stimuli have been found to support spatial directing in driving simulation (Ho et al. 2006)(Scott & Gray 2008), and navigation (Colbert & Tack 2005) tasks. However, clear degradation in the detection level of tactile stimuli has been reported in demanding field conditions (Krausman & White 2006). Therefore, it was expected that especially after physical exercise the detection of the tactile stimuli might be worse than for the other modalities in the current study. Ten tactile stimuli were missed by three different subjects in the task conditions with physical exercise (and one stimulus in condition without physical exercise). In the interviews, the subjects reported that they did not perceive those stimuli at all. Several subjects reported inconsistencies in the clarity of the tactile stimuli in the different experiment sessions, and suggested that it could be because of the tactor belt position or its attachment apparently not being ideal. Several subjects also reported that the tactile stimuli were difficult to perceive after physical exercise. They were then impaired by moving, heavy breathing and heartbeat.

The position of the tactor belt as such was not considered problematic, but the belt should be attached so that the tactors always remain in contact with the skin to ensure that they are perceived in a constant and reliable manner. Possibly the vibration could also be stronger and the side tactors further away from the centre. Longer stimulus duration might enhance detection, but when fast reactions are needed, it is probably not the optimal solution. The tactile cues were not considered disturbing sighting or concentration, or the shooting position. Half of the subjects also reported the tactile cues indicating the direction most clearly. Another positive aspect was that the tactile stimuli were different from other stimuli received from the environment during a shooting task.

Initial reaction time analysis suggested indicative difference between the tactile and the multimodal modality variations, the tactile being slower. However, the analyses made together with accuracy results revealed that this difference could be due to speed-accuracy trade-off, as the tactile stimuli were also more accurate than the multimodal cues. It is not possible to put these tasks in a ranking order by difficulty or reaction time.

The multimodal stimuli in the current study refer to stimuli presented simultaneously on different sensory modalities. The subjects were predicted to be faster and more accurate after the multimodal cues, as multisensory stimuli have shown to be stronger for capturing attention than unimodal stimuli (Spence & Gallace 2007)(Santangelo & Spence 2008). Half of the subjects mentioned in the interviews that with the multimodal cue there was no chance of making mistakes as the different stimuli supported one another. Two shots were however shot to a wrong direction after a multimodal cue.

The mixed-modal stimuli refer to the task condition in which different modality stimuli are shown randomly. The reason for taking along the mixed condition was to investigate the effects of expectancy. These expectations may slow down the processing of a target when the stimuli appear on an unexpected sensory modality (Spence et al. 2001). Also the amount of errors may increase (Spence & Gallace 2007). Thus, the subjects were predicted to be slower and to make more errors in the mixed task. The reaction time or accuracy means of mixed and unimodal conditions did not differ and there was only one error; one auditory cue was shot to the wrong direction. In the interviews, some subjects commented that the mixed condition felt more demanding and confusing whereas one subject had an opposite opinion. Some subjects considered that the tactile stimulus cues were easiest to recognize in the mixed task condition because it was different from the other cues. The preference rankings compared to reaction times from the mixed task condition matched for four subjects out of six, whereas when comparing rankings to reaction times from the unimodal task conditions only two matches were found.

It is possible that comparing the different cue modalities is easier when the cues are presented simultaneously or during the same task. In the interviews, several subjects often raised points comparing modalities in relation to multimodal and mixed conditions. This methodologically interesting possibility would however need further investigation. In the current study, the amount of unimodal cues shown in the mixed condition was quite low, one third of those shown in the unimodal conditions.

In the task condition with physical exercise increased heart rate and heavier breathing were expected to have an effect on perceiving cues, and to impair the performance and subjective preferences. The results show that there were no statistically significant differences between the conditions with and without physical exercise for reaction time or accuracy for any of the five modality variation pairs. However, when observing the means visually, a trend could be seen where performance was worse in the task conditions with physical exercise. The speed-accuracy analysis confirmed this finding. In addition, the subjective measurements showed tendency (non-significant) that the physical exercise was experienced to impair the perception of the cues. In the questionnaire all modalities got worse (statistically non-significant) rating for cue clarity after physical exercise. The trend of worse performance after physical exercise is not necessarily due to impaired perception or attention to the cues. Feeling exhausted or being out of breath during the task may impair the shooting performance as well. One subject commented that the physical exercise had an effect on shooting, not only on stimulus perception. One subject also reported impairment in his concentration because of the relatively long shooting sessions.

In relation to perceiving the target directions, there were no specific hypotheses or expectations of differences between modalities. In the questionnaires, the subjects agreed that the different target directions were easy to identify for all modalities. The performance results show that almost all of the shots were shot to the correct target direction. When comparing the reaction time for different target directions per modality, an indicative (close to significant) difference was found for the auditory modality variation, the performance with the left cue being faster than with the centre cue. For most modality variations the side directions seemed (statistically non-significantly) slower than centre direction, but for the auditory modality variation it was the opposite. The subjects started shooting from the centre position, which could explain why the side directions were slower for most modalities. The opposite indicative difference found for the auditory modality was interesting. It was possibly related to some issues reported in relation to identifying the centre direction (see discussion about auditory cues above). Some subjects routinely use radio to receive orders in their work, thus there might be some kind of (one-sided) learning effect of that modality channel.

11.2. Reliability and validity of results

Reliability can be a problematic issue when evaluating human factors because the differences between subjects' performance and preferences are normally huge, especially in real environments (Nielsen 1993). Statistically validated results are required for reliable results, but in a field setup it is more difficult or even impossible to make the experiment as controllable as in a laboratory setup.

Laboratory setups are less likely to resemble the real world. The problem is that reliable data collected may not be ecologically valid: the findings actually do not reproduce in real circumstances. The statistically validated results of a well-controlled experiment made with a small number of users can give indications of possible problems, but do not prove one modality better than another (Benyon et al. 2005). Thus, even if statistically significant differences were found in laboratory studies, they may or may not indicate significant issues related to real life (Benyon et al. 2005). Therefore, information collected in a field setup can be considered as more ecologically valid, especially when considering the experiences of the subjects. Nielsen (1993) states that decisions for interactive system design can be made based on such results, because some information is better than no information.

In the current study, the subjects used the wearable devices in a real life field setup simulating a close combat situation. If they had tried on the devices without the shooting task, probably many of the comments now collected in the interviews would not have been reported. Some issues, which were reported during the experiment, were not noticed during the training day, when the subjects did not do the actual shooting task but just tried the devices without performing the task.

Even if a field setup is more realistic than a laboratory setup, the task and cues in the current study are in many aspects still not ecologically valid. In the experiment, the subjects were concentrated and waiting for stimulus cues, which would not be the case in a real combat task. The cues are meant for capturing attention, but in the current study there is no other simultaneous activity involved to take attention so actually no capturing in that sense is needed. The current task also only has three target directions in the front of the subject. The future research might select some other real life task relevant aspects for experiment, such as moving targets for example.

In this study many different kinds of cues and conditions were measured with a relatively small group of subjects. The subjects' variance in their own shooting performance was larger than expected. In addition, the reaction time for a task was expected to be shorter. The proportion of modality differences in performance time can be small even in well-trained and fast real-life tasks, and the proportion of other factors such as motor movements and environmental conditions can become more important. It is possible that any possible differences between modalities (cuing effects) will disappear into overall reaction time.

11.3. Study setup and methods

The current study remained on a behavioural level even if the theoretical background is based partly on studies investigating e.g. information processing or attention mechanisms. The methods used are similar to commonly used usability testing methods, but still the study was rather applied basic research. One strength and speciality of this study is, that research questions are investigated from different perspectives, subjective and performance measures, quantitatively and qualitatively.

The reaction times and accuracies were investigated together to rule out the possibility that any differences found would be due to the speed-accuracy trade-off. Even if the speed-accuracy trade-off is a well-known phenomenon, many studies comparing e.g. performance differences of modalities have not investigated the reaction times together with the error rates. In the current study, there were missed shots and other errors among the slowest reaction times. This might be considered as a sign of the subjects respecting both of the instructions they got, to be accurate and to be fast. In the current study, both accuracy and reaction time can be given equal importance in task performance. When overall data was explored, no speed-accuracy trends were found. However, for samples analysed in more detail there were cases where the speed-accuracy trade-off could not be ruled out after an indicative difference was found between reaction time means. Thus, this approach did have an impact to the performance measurement results.

Physical exercise was used to simulate "battle stress" in the current study. The physical effort was controlled with heart rate (ECG) and the effects of it to subjective experience of workload were evaluated with NASA-TLX. An increase in heart rate could be observed during the shooting task even in the condition without physical exercise, and the it remained high during the task. In addition, NASA-TLX indicated a trend (statistically non-significant) where the ratings for workload were higher for the tasks with physical exercise. These controls were considered successful and useful for supervising and evaluating the effects of the physical exercise.

12. CONCLUSIONS

This study presents a human factors approach to evaluating new information technology devices to aid a good task performance.

The main objective of this study was to evaluate spatially directing cues presented on wearable user interfaces in a field setup. In this thesis visual, auditory and tactile cue stimuli and multimodal stimuli combinations were studied.

The results of this study do not clearly enable putting modality variations in a ranking order, neither by performance nor by subjective measures. No statistical effects were found between performance with different modality cues or their combinations. However, a trend pointing to the performance degrading due to physical exercise was observed. The subjects had preferences among modalities, but these differed between subjects. Each modality had some experienced strengths and weaknesses.

The results indicate that the stimuli evaluated in the actual experiment were more or less equally good for the task and all of them could be used to support situational awareness. All of them had neutral or positive mean ratings in the questionnaire.

Even if the shooting task in the experiment was a relative simple real life task, and the subjects were well trained to it, their reaction time varied and was in average clearly more than a second. The speed benefits that could be considered to be related to cross-modal attention cuing represent only a small proportion of the reaction times. Thus, cross-modal effects found in laboratory settings seem less interesting for an applied field setup for the purpose of speeding up the performance. It seems to be more important that the stimuli are reliably detected. Spatially directing cues may support capturing attention by ensuring that people detect an important event in their environment. The cue saliency and detectability depends on environmental conditions such as sun light, sounds or other stimuli, and task requirements such as body position, movements or even facial expression when the stimuli are presented on wearable devices. Depending on the task needs and environment conditions, a particular human sense is often better suited to a particular task due to the differences of the human senses (Ferris & Sarter 2008).

For example, if one human sense is under heavy load, another sense could support the detection of target.

The visual stimuli evaluated in the pilot experiment, which were presented with HMD were not found suitable for a time-critical shooting task mainly because the subject had to focus the gaze to the display to see the stimuli and thus the sighting was disturbed. This shows that the difference between two different presentation techniques in one modality may be clearer than the difference between modalities.

It is of crucial importance to take into consideration the task requirements and environmental conditions when designing technologies to support human performance. The stimuli and the devices should be designed for a certain task in a certain environment and also be evaluated in realistic circumstances. For example, it is important for task performance that the stimulus cues or devices used do not disturb the task nor prevent paying attention to other important events in the environment. The stimulus cues should not be used unnecessarily and they should not add workload but reduce it.

Based on the results, the design of cross- and multimodal cuing on wearable user interfaces should concentrate on reliability and detectability of the cues in the target environment. Task requirements should be considered carefully, rather than trying to define which modality leads to fastest or most accurate performance. User evaluation including performance and subjective measures in a field setup is essential. The performance measures can be used to indicate possible problems, such as cues that were not detected. Subjective measures in the field setup can reveal aspects that cannot be found in a laboratory experiment, and are still essential for the real users. This information can be taken in to consideration in future experiments or when designing wearable user interfaces.

The more we understand the strengths and weaknesses of different modalities and devices presenting the stimulus cues, the better we can design user interfaces supporting situational awareness and task performance in a cognitively ergonomic manner.

BIBLIOGRAPHY

Bear, M. F., Connors, B. W., & Paradiso, M. A. (2007). Neuroscience: Exploring the Brain (3rd edition). Lippincott Williams & Wilkins.

Benyon, D., Turner, P., & Turner, S. (2005). Designing interactive systems: people, activities, contexts, technologies. Addison Wesley.

Berman, R., & Colby, C. (2009). Attention and active vision. Vision research, 49(10), 1233-48. doi: 10.1016/j.visres.2008.06.017.

Boothe, R. G. (2002). Perception of the visual environment (1st edition.). New York: Springer-Verlag Inc.

Bracha, H. S., Ralston, T. C., Matsukawa, J. M., Williams, A. E., & Bracha, A. S. (2004). Does "fight or flight" need updating? Psychosomatics, 45(5), 448-9. doi: 10.1176/appi.psy.45.5.448.

Brewster, S. A., & Brown, L. M. (2004). Tactons: structured tactile messages for non-visual information display. Proceedings of the fifth conference on Australasian user interface, 28(January), 15–23. Australian Computer Society, Inc. Retrieved from http://portal.acm.org/citation.cfm?id=976310.976313.

Burke, C., Hancock, P., Warm, J., Szalma, J., Oron-Gilad, T., Harris, W., et al. (2004). Stress Effects on Soldier Performance. In Human Factors and Ergonomics Society Annual Meeting Proceedings, Individual Differences in Performance (Vol. 48, p. 1271–1274). Human Factors and Ergonomics Society. Retrieved from http://www.ingentaconnect.com/content/hfes/hfproc/2004/00000048/00000011/art00014.

Cholewiak, R., Brill, J., & Schwab, A. (2004). Vibrotactile localization on the abdomen: Effects of place and space. Perception and Psychophysics, 66(6), 970–987. [Austin, Tex., etc., Psychonomic Society, etc.]. Retrieved from http://app.psychonomic-journals.org/content/66/6/970.full.pdf.

Colbert, H. J., & Tack, D. W. (2005). Evaluation of Visual, Auditory and Tactile Displays of GPS-Based Localization Information for Soldiers on the Battlefield. Toronto. Retrieved from http://pubs-www.drenet.dnd.ca/BASIS/pcandid/www/engpub/DDW?W=AUTHOR+PH+IS+'Colbert'+ORDER+BY+Repdate/Descend&M=6&K=524753&U=1.

Driver, J., & Spence, C. (1998). Crossmodal attention. Current Opinion in Neurobiology, 8(2), 245-253. doi: 10.1016/S0959-4388(98)80147-5.

Drury, C. (1999). Managing the speed-accuracy trade-off. In W. Karwowski & W. S. Marras, The occupational ergonomics handbook (pp. 677-687). CRC Press LLC.

Engineering Acoustics, I. (2009). C-2 Tactor. Retrieved from http://www.tactors.com/Products.aspx?productId=01.

Eysenck, M. W., & Keane, M. T. (2005). Cognitive Psychology: A students' handbook (5th.). New York: Psychology Press Ltd.

Ferris, T. K., & Sarter, N. B. (2008). Cross-modal links among vision, audition, and touch in complex environments. Human Factors, 50(1), 17-26. doi: 10.1518/001872008X250566.

Field, A. (2009). Discovering statistics using SPSS: Introducing Statistical Methods (3rd.). Sage Publications, Inc.

Goldstein, E. B. (2007). Sensation and Perception (7th.). Belmont, CA: Wadsworth.

Hancock, P. A., & Szalma, J. L. (2008). Stress and neuroergonomics. In R. Parasuraman & M. Rizzo, Neuroergonomics: The brain at work (1st paperb., p. 195–206). New York: Oxford University Press.

Hart, S. G. (2006). Nasa-task load index (NASA-TLX); 20 years later. Proceedings of the Human Factors and and Ergonomics Society 50th Annual Meeting, 904–908. Human Factors and Ergonomics Society.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In P. Hancock & N. Meshkati, Human Mental Workload (pp. 139-184). Amsterdam: North Holland Press.

Ho, C., & Spence, C. (2005). Assessing the effectiveness of various auditory cues in capturing a driver's visual attention. Journal of experimental psychology. Applied, 11(3), 157-74. doi: 10.1037/1076-898X.11.3.157.

Ho, C., Reed, N., & Spence, C. (2006). 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), 988-96. doi: 10.1016/j.aap.2006.04.002.

Härmä, A., Jakka, J., Tikander, M., Karjalainen, M., Lokki, T., Hiipakka, J., et al. (2004). Augmented Reality Audio for Mobile and Wearable Appliances. Journal of the Audio Engineering Society, 8(2), 63-91.

Härmä, A., Jakka, J., Tikander, M., Karjalainen, M., Lokki, T., Nironen, H., et al. (2003). Techniques and applications of wearable augmented reality audio. PREPRINTS-AUDIO ENGINEERING SOCIETY (pp. 1-20). Audio Engineering Society. Retrieved from http://www.aes.org/e-lib/browse.cfm?elib=12495.

Jones, L. A., & Sarter, N. B. (2008). Tactile displays: Guidance for their design and application. Human Factors, 50(1), 90-111. doi: 10.1518/001872008X250638.

Kenttämies, J., Koivumäki, A., Karisto, H., & Korpinen, P. (2001). Äänipää - a web site focusing on sound narrative. Retrieved from http://www.aanipaa.tamk.fi/english.htm.

Kim, K., McMillan, M., & Zelaznik, H. N. (1996). Behavioral analysis of trajectory formation: the speed-accuracy trade-off as a tool to understand strategies of movement control. In H. N. Zelaznik, Advances in motor learning and control (pp. 1-12). Human Kinetics.

King, A. J. (2009). Visual influences on auditory spatial learning. Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 364(1515), 331-9. doi: 10.1098/rstb.2008.0230.

Krausman, A. S., & White, T. L. (2006). Tactile Displays and Detectability of Vibrotactile Patterns as Combat Assault Maneuvers are Being Performed. U.S. Army Research Laboratory, Technical Reports. Retrieved from http://www.arl.army.mil/arlreports/2006/ARL-TR-3998.pdf.

Kumagai, J., & Massel, L. (2005). The Assessment of Auditory Displays in Support of Wayfinding. Defence Research and Development Canada–Toronto, CR2005. Toronto. Retrieved from http://pubs-www.drenet.dnd.ca/BASIS/pcandid/www/engpub/DDW?W%3DAUTHOR+PH+IS+%27kumagai%27+ORDER+BY+Repdate/Descend%26M%3D7%26K%3D524748%26U%3D1.

Laarni, J., Heinilä, J., Häkkinen, J., Kalakoski, V., Kallinen, K., Lukander, K., et al. (2009). Supporting Situation Awareness in Demanding Operating Environments through Wearable User Interfaces. In Engineering Psychology and Cognitive Ergonomics (pp. 13-21). Springer Berlin / Heidelberg. doi: 10.1007/978-3-642-02728-4.

Lachman, R., Lachman, J. L., & Butterfield, E. C. (1979). Cognitive psychology and information processing: an introduction. Cognitive psychology and information processing: an introduction (pp. 158-163). Hillsdale, New Jersey: Lawrence Erlbaum.

Laurienti, P., Burdette, J., Maldjian, J., & Wallace, M. (2006). Enhanced multisensory integration in older adults. Neurobiology of aging, 27(8), 1155–1163. Elsevier. doi: 10.1016/j.neurobiolaging.2005.05.024.

McCreery, C. (2007). First-year Statistics for Psychology Students Through Worked Examples 3. Analysis of Variance. Experimental Psychology.

Metsämuuronen, J. (2005). Kokeellisen tutkimuksen perusteet ihmistieteissä. Helsinki: International Methelp.

Moore, B. C. (2004). An introduction to the psychology of hearing (5th edition.). Academic Press.

Nielsen, J. (1993). Usability Engineering. Academic Press.

Näsänen, R., Ahonen, L., Jagadeesan, S., Mynttinen, R., & Müller, K. (2008). Cueing spatial visual attention by auditory stimuli (p. 98). Perception 37 ECVP Abstract Supplement. Retrieved from http://nasanen.net/PosterAtECVP2008.pdf.

Patterson, R., Winterbottom, M. D., & Pierce, B. J. (2006). Perceptual issues in the use of head-mounted visual displays. Human factors, 48(3), 555-73. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17063969.

Pavani, F., Murray, M., & Schroeder, C. (2007). Rethinking mind, brain and behaviour through a multisensory perspective. Neuropsychologia, 45(3), 467–468. Elsevier. doi: 10.1016/j.neuropsychologia.2006.07.001.

Sanabria, D., Lupianez, J., Chica, A., & Spence, C. (2007). Comparing intramodal and crossmodal cuing in the endogenous orienting of spatial attention. Experimental Brain Research, 179(3), 353–364. Springer. doi: 10.1007/s00221-006-0798-7.

Santangelo, V., & Spence, C. (2008). Is the exogenous orienting of spatial attention truly automatic? Evidence from unimodal and multisensory studies. Consciousness and cognition, 17(3), 989-1015. doi: 10.1016/j.concog.2008.02.006.

Sarna, S. (2009). Kliinisen biostatistiikan peruskurssi.

Schmidt, R., & Wrisberg, C. A. (2008). Principles of motor control and movement accuracy. In R. Schmidt & C. A. Wrisberg, Motor learning and performance: a situation-based learning approach (4 ed., pp. 132-159). Human Kinetics.

Scott, J., & Gray, R. (2008). A Comparison of Tactile, Visual, and Auditory Warnings for Rear-End Collision Prevention in Simulated Driving. Human Factors, 50(2), 264-275. doi: 10.1518/001872008X250674.C.

Shams, L., & Seitz, A. R. (2008). Benefits of multisensory learning. Trends in cognitive sciences, 12(11), 411-7. doi: 10.1016/j.tics.2008.07.006.

Sinkkonen, I., Kuoppala, H., Parkkinen, J., & Vastamäki, R. (2002). Käytettävyyden psykologia. Helsinki: Edita Publishing Oy/IT Press.

Spence, C., & Gallace, A. (2007). Recent developments in the study of tactile attention. Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 61(3), 196-207. doi: 10.1037/cjep2007021.

Spence, C., & Ho, C. (2008). Multisensory interface design for drivers: past, present and future. Ergonomics, 51(1), 65-70. doi: 10.1080/00140130701802759.

Spence, C., & Santangelo, V. (2009). Capturing spatial attention with multisensory cues: A review. Hearing research, 1-9. doi: 10.1016/j.heares.2009.04.015.

Spence, C., McDonald, J., & Driver, J. (2004). Exogenous spatial cuing studies of human crossmodal attention and multisensory integration. In C. Spence & J. Driver, Crossmodal space and crossmodal attention (p. 277–312). New York: Oxford University Press, USA.

Spence, C., Nicholls, M., & Driver, J. (2001). The cost of expecting events in the wrong sensory modality. Perception and Psychophysics, 63(2), 330–336. [Austin, Tex., etc., Psychonomic Society, etc.]. Retrieved from http://app.psychonomic-journals.org/content/63/2/330.full.pdf.

Stanton, N. A., & Young, M. S. (1999). A guide to methodology in ergonomics: designing for human use. Taylor & Francis.

Stanton, N. A., Salmon, P. M., Walker, G. H., Baber, C., & Jenkins, D. P. (2005). Human factors methods: a practical guide for engineering and design. Ashgate Publishing Limited.

Sternberg, S. (1998). Inferring mental operations from reaction time data: how we compare objects. In D. N. Osherson, D. Scarborough, & S. Sternberg, An invitation to cognitive science: Methods, models and conceptual issues (pp. 365-455). The MIT Press.

UCLA. (2009). What statistical analysis should I use? Statistical analyses using SPSS. Academic Technology Services Group, UCLA, Statistical Consulting. Retrieved from http://www.ats.ucla.edu/stat/spss/whatstat/whatstat.htm.

Virsu, V., Oksanen-Hennah, H., Vedenpää, A., Jaatinen, P., & Lahti-Nuuttila, P. (2008). Simultaneity learning in vision, audition, tactile sense and their cross-modal combinations. Experimental brain research. Experimentelle Hirnforschung. Expérimentation cérébrale, 186(4), 525-37. doi: 10.1007/s00221-007-1254-z.

Warm, J. S., Matthews, G., & Finomore Jr., V. S. (2008). Vigilance, Workload, and Stress. In P. A. Hancock & J. L. Szalma, Performance under stress: human factors in defence (pp. 115-142). Ashgate Publishing Limited.

Wickelgren, W. (1977). Speed-accuracy tradeoff and information processing dynamics. Acta Psychologica, 41(1), 67–85. Elsevier BV, Radarweg 29, Amsterdam, 1043 NX, Netherlands,. Retrieved from http://doi.apa.org/?uid=1977-26842-001.

Winterbottom, M. D., Patterson, R., Pierce, B. J., Covas, C. M., & Winner, J. (2007). Depth of Focus and Visual Recognition of Imagery Presented on Simultaneously Viewed Displays: Implications for Head- Mounted Displays. Human Factors, 49(5), 907-919. doi: 10.1518/001872007X230253.

Zhou, Z., & Rogers, E. R. (2002). Sound Localization and Virtual Auditory Space.

APPENDICES

Appendix 1 End questionnaire (original in Finnish)

Arviointiohjeet:
1= Täysin eri mieltä
2= eri mieltä
3= hieman eri mieltä
4= ei osaa sanoa, ei samaa eikä eri mieltä
5= hieman samaa mieltä
6= samaa mieltä
7= Täysin samaa mieltä

(1=ei 7=kyllä)

	VISUAALINEN	**AUDIO**	**TÄRINÄ**
VIHJE	EI KYLLÄ 1 2 3 4 5 6 7	EI KYLLÄ 1 2 3 4 5 6 7	EI KYLLÄ 1 2 3 4 5 6 7
Vihje oli helppo tunnistaa	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○
Vihje tuntui miellyttävältä	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○
Vihje tuntui ärsyttävältä	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○
Vihje oli luotettavan/varman tuntuinen	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○
Vihjeiden väli oli tarpeeksi pitkä	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○
Vihje tuntui luontevalta	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○
Vihje oli helppo oppia	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○

	VISUAALINEN	**AUDIO**	**TÄRINÄ**
SUUNNAT	EI KYLLÄ 1 2 3 4 5 6 7	EI KYLLÄ 1 2 3 4 5 6 7	EI KYLLÄ 1 2 3 4 5 6 7
Eri suunnat oli helppo tunnistaa	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○

	VISUAALINEN	**AUDIO**	**TÄRINÄ**
TEHTÄVÄ	EI KYLLÄ 1 2 3 4 5 6 7	EI KYLLÄ 1 2 3 4 5 6 7	EI KYLLÄ 1 2 3 4 5 6 7
Vihje häiritse tähtäämistä	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○
Vihje häiritsi keskittymistä	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○
Vihje auttoi tehtävässä	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○
Vihje tuntui nopeuttavan ampumista	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○
Vihje voisi olla hyvä oikeassa käytössä	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○
Vihje tuntuu tehtävään sopimattomalta	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○
Vihje-laite vaikutti ampuma-asentoosi tai tapaasi ampua	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○

Appendix 1 End questionnaire (original in Finnish)

	VISUAALINEN	AUDIO	TÄRINÄ
RASITUS	EI KYLLÄ 1 2 3 4 5 6 7	EI KYLLÄ 1 2 3 4 5 6 7	EI KYLLÄ 1 2 3 4 5 6 7
Vihje oli selkeä rasituksen jälkeisessä tehtävässä	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○
Vihje oli selkeä kun rasitusta ei ollut ennen tehtävää	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○

Appendix 2 End questionnaire (with English translation)

Instructions for evaluation: (Arviointiohjeet)
1= Strongly disagree (Täysin eri mieltä)
2= Disagree (eri mieltä)
3= Disagree a little (hieman eri mieltä)
4= Do not know, cannot agree not disagree (ei osaa sanoa, ei samaa eikä eri mieltä)
5= Agree a little (hieman samaa mieltä)
6= Agree (samaa mieltä)
7= Strongly agree (Täysin samaa mieltä)

1=no (ei) 7=yes (kyllä)

	VISUAL	AUDITORY	TACTILE
CUE	NO YES 1 2 3 4 5 6 7	NO YES 1 2 3 4 5 6 7	NO YES 1 2 3 4 5 6 7
Cue was easy to identify (Vihje oli helppo tunnistaa)	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○
Cue felt pleasant (Vihje tuntui miellyttävältä)	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○
Cue felt annoying (Vihje tuntui ärsyttävältä)	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○
Cue felt reliable (Vihje oli luotettavan/varman tuntuinen)	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○
The time between stimuli was long enough (Vihjeiden väli oli tarpeeksi pitkä)	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○
Cue felt intuitive (Vihje tuntui luontevalta)	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○
Cue was easy to learn (Vihje oli helppo oppia)	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○

	VISUAL	AUDITORY	TACTILE
DIRECTIONS	NO YES 1 2 3 4 5 6 7	NO YES 1 2 3 4 5 6 7	NO YES 1 2 3 4 5 6 7
Different directions were easy to identify (Eri suunnat oli helppo tunnistaa)	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○

	VISUAL	AUDITORY	TACTILE
TASK	NO YES 1 2 3 4 5 6 7	NO YES 1 2 3 4 5 6 7	NO YES 1 2 3 4 5 6 7
Cue disturbed sighting (Vihje häiritsi tähtäämistä)	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○
Cue disturbed concentration (Vihje häiritsi keskittymistä)	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○

Cue helped in shooting task (Vihje auttoi tehtävässä)	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○
Cue felt to speed up shooting (Vihje tuntui nopeuttavan ampumista)	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○
Cue might be good in real use (Vihje voisi olla hyvä oikeassa käytössä)	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○
Cue felt unsuitable for task (Vihje tuntuu tehtävään sopimattomalta)	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○
Cue device affected your shooting position or way of shooting (Vihje-laite vaikutti ampuma-asentoosi tai tapaasi ampua)	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○

	VISUAL	**AUDITORY**	**TACTILE**
PHYSICAL EXERCISE	NO YES 1 2 3 4 5 6 7	NO YES 1 2 3 4 5 6 7	NO YES 1 2 3 4 5 6 7
Cue was clear in a task after physical exercise (Vihje oli selkeä rasituksen jälkeisessä tehtävässä)	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○
Cue was clear in a task without physical exercise (Vihje oli selkeä kun rasitusta ei ollut ennen tehtävää)	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○	○ ○ ○ ○ ○ ○ ○

Appendix 3 NASA-TLX rating scale definitions

Table from (Hart 2006).

Title	**Endpoints**	**Description**
Mental Demand	low/high	How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking searching etc.)? Was the task easy or demanding, simple or complex, exacting of forgiving?
Physical Demand	low/high	How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
Temporal demand	low/high	How much physical activity did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
Performance	good/poor	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
Effort	low/high	How hard did you have to work (mentally or physically) to accomplish your level of performance?
Frustration level	low/high	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

Appendix 4 NASA-TLX questionnaire in Finnish

Tehtävä:
KH:

Arvioi tehtävän aikana kokemaasi kuormitustasoa seuraavilla asteikoilla (merkitse X ruksi ruutuun)

Henkinen vaatimustaso																			

Vähän Paljon

Fyysinen vaatimustaso																			

Vähän Paljon

Ajallinen vaatimustaso																			

Hidas Nopea

Suoritus																			

Hyvin Heikosti

Ponnistelu																			

Vähän Paljon

Turhautuminen																			

Vähän Paljon

Appendix 5 End interview questions in Finnish (original)

KH:

1. Miltä nämä erilaiset suuntaa-antavat vihjeet tuntuivat?
(entä rasituksessa?)

2. Tuntuivatko vihjeet yhtä vahvoilta ja tehokkailta?
(entä rasituksessa?)

3. Oliko jokin vihje parempi kuin muut?
- miksi?
(entä rasituksessa?)

4. Oliko jokin vihje huonompi kuin muut?
- miksi?
(entä rasituksessa?)

5. Oliko jokin vihje häiritsevä tai ärsyttävä?
(entä rasituksessa?)

6. Antoiko jokin vihje suunnan selkeämmin? (Viestin ymmärrettävyys)
- Suunnan havaitseminen: tuntuiko vihjeiden antama suunta oikealta, auttoiko ampumaan sinne suuntaan?
- Tuntuivatko eri suuntiin osoittaneet vihjeet samanarvoisilta, oliko vasen, keski, oikea yhtä helppo havaita?
(entä rasituksessa?)

7. Tuntuiko jokin vihje tehtävään sopimattomalta?
(entä rasituksessa?)

8. Luontevammalta kuin muut? (Intuitiivisuus, opittavuus)
(entä rasituksessa?)

9. Oliko kaikkiin yhtä helppo keskittyä?
(entä rasituksessa?)

10. Oliko niistä apua tehtävässä / haittasiko
Voisitko kuvitella käyttäväsi tämän tyyppisiä laitteita ja vihjeitä käskyjen saamiseen taistelutilanteessa? Miksi/ei? Mitä mieluiten käyttäisit?
(entä rasituksessa?)

Appendix 6 End interview questions in English

1. How did these different spatially directing stimulus cues felt?
(and after physical exercise?)

2. Did all cues feel as strong and effective?
(and after physical exercise?)

3. Was one cue better than others?
- which one, why?
(and after physical exercise?)

4. Was one cue worse than others?
- which one, why?
(and after physical exercise?)

5. Was one cue disturbing or annoying?
- which one, why?
(and after physical exercise?)

6. Did one cue give direction more clearly?
- was the message understandable
- about perceiving the direction: did the direction given by cues feel correct, did it help shooting that direction?
- Did the different directions of one modality feel equal - was it as easy to perceive left, centre and right?
(and after physical exercise?)

7. Did one cue feel unsuitable for the shooting task?
(and after physical exercise?)

8. Did one cue feel more natural than others?
- more intuitive, easier to learn?
(and after physical exercise?)

9. Was it as easy to concentrate to all cues?
(and after physical exercise?)

10. Did the cues help shooting task?
- did the cue disturb shooting task?
- could you imagine using this type of cues and devices in a combat situation to get orders?
- why/ not? Which one would you rather use?
(and after physical exercise?)

Appendix 7 Questions for analysis

Performance measures

The following questions for analysis were defined to investigate effects of different variables on reaction time (RT) with statistical tests:

RT1. Does the modality have an effect (visual, auditory or tactile)?
In this evaluation all hits from all other than the multimodal task conditions are analyzed, grouped by modality. Also the values from the mixed task conditions were included.

RT2a. Does the modality have an effect (visual, auditory, tactile and multimodal)?
In this evaluation all hits from all task conditions (including multimodal) are analyzed, grouped by modality. Also the values from the mixed task conditions were included.

RT2b. Are there differences between different modalities (in tasks)?
In this evaluation all hits from all other than the mixed task conditions are analyzed, grouped by modality. Note that in previous conditions (1,2a) reaction times from the mixed task conditions were combined to unimodal values.

RT3a. Are there differences between modalities in the mixed task condition?
In this evaluation all hits from the mixed task conditions are analyzed, grouped by modality. There are only unimodal cues in the mixed task conditions.

RT3.b Are there differences between the cues presented in the mixed task conditions and in the unimodal task conditions?
In this evaluation all hits from all other than the multimodal task conditions are analyzed, grouped by modality and the task type: the ones in the unimodal and the ones in the mixed task conditions. Note: the amount of shots per modality in the mixed conditions is only 1/3 of amount of shots in the unimodal conditions.

RT4. Does physical exercise have an effect on reaction time?
In this evaluation all hits from all tasks are analyzed, grouped by physical exercise: the ones with physical exercise preceding the task and the ones without.

RT5. Does target direction have an effect on reaction time?
In this evaluation all hits from all tasks are analyzed, also the values from the mixed task conditions were included, grouped by target direction. Differences of reaction times to the directions (left, centre and right) were compared.

The following questions for analysis were defined to investigate effects of different variables on accuracy (AC) with statistical tests:

AC1. Does the task have an effect on accuracy?
In this evaluation all accuracy data for all task conditions are analyzed, grouped by task condition.

AC2. Does physical exercise have an effect on accuracy?
In this evaluation all accuracy data for all task conditions are analyzed, grouped by physical exercise: the ones with physical exercise preceding the task and the ones without.

The following questions for analysis were defined to investigate the relationship between reaction time and accuracy, speed-accuracy (SA) trade-off:

SA1. Is there overall speed-accuracy trade-off found in data?

SA2. Are the differences in reaction time due to speed-accuracy trade-off?
In this evaluation possible differences between task conditions were investigated for speed-accuracy trade-off.

SA3. Are the effects of physical exercise due to speed-accuracy trade-off?
In this evaluation possible differences between task conditions with and without physical exercise were investigated for speed-accuracy trade-off.

Subjective measures

The following questions for analysis were defined to investigate the preferences of subjects with questionnaires (QU):
QU1.Which modality is preferred, overall
QU2.Which modality is preferred, per question

The following question for analysis was defined to investigate the preferences of subjects by combining results of subjective and performance measurement results:
QU3. Questionnaire modality preference vs. reaction time

The following questions for analysis were defined to investigate the preferences of subjects with interviews (IN):
- Comments for each modality variation: experienced strengths and weaknesses of modality variations IN1-IN5?:
IN1. Visual cue
IN2. Auditory cue
IN3. Tactile cue
IN4. Multimodal cue
IN5. Mixed- condition
- Other comments?
IN6: Miscaleous

The following questions for analysis were defined to investigate the effects of different conditions on NASA-TLX workload (TLX) with statistical tests:

TLX1. Does physical exercise have an effect on NASA-TLX workload in task conditions?
In this evaluation mean ratings of overall workload for each task condition were compared. Overall workload was counted as an average all dimensions for each task condition.

TLX2. How dimensions contribute to overall workload differences?
In this evaluation the differences of dimensions were investigated. The mean ratings of selected dimensions for task conditions with and without physical exercise were compared.

TLX3. Does physical exercise have an effect on dimension workload in task condition? In this evaluation the mean ratings of dimension workload in each task condition were compared.

Appendix 8 Data distributions

Reaction times

Figure 17 below shows the distribution of reaction times for all shots that hit the target. The histogram and normal Q-Q Plot show clearly a longer tail on right.

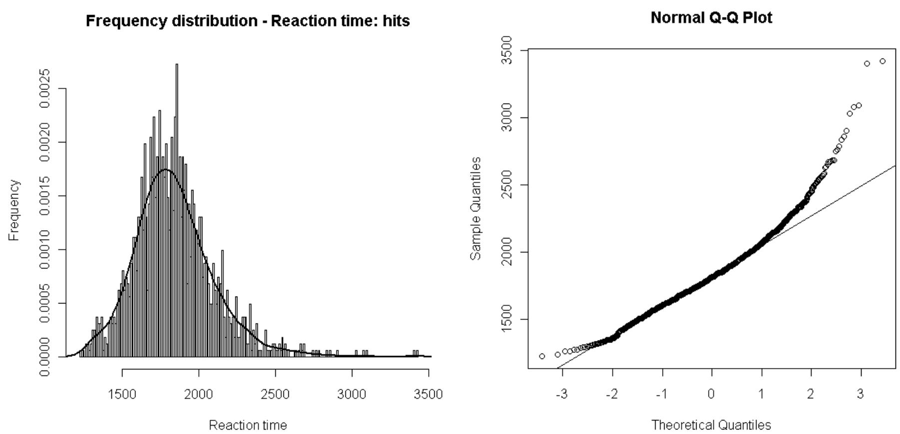

Figure 17 Frequency distribution histogram and Q-Q Plot of reaction times for hits

The data were right-skewed and the mass of distribution was concentrated on the left of the histogram, and there are relatively few high (right) values. These high values were analyzed and it was concluded that those were not outliers. One slow reaction time was caused by a subject changing his position during a task, but even that value was not clearly slower than other high reaction time values. Therefore no values were removed.

The Shapiro-Wilk normality test was used to test the normality of the distribution. Data were significantly different from normal ($p<2.2e-16$). Same evaluation was made for several samples which were planned to be compared, showing that those were not normally distributed either. To reduce right-skewedness a log[10] transformation was made but the data still was not normally distributed. As the assumptions for parametric test were not met for most samples, non-parametric tests were chosen for statistical tests for all reaction time data, the Friedman test for comparing group means, and the Wilcoxon signed ranks test for multiple comparisons post hoc tests. Original data values were used.

Accuracy

Figure 18 below show the distribution of accuracy for all tasks. The histogram and boxplot show longer tail on left.

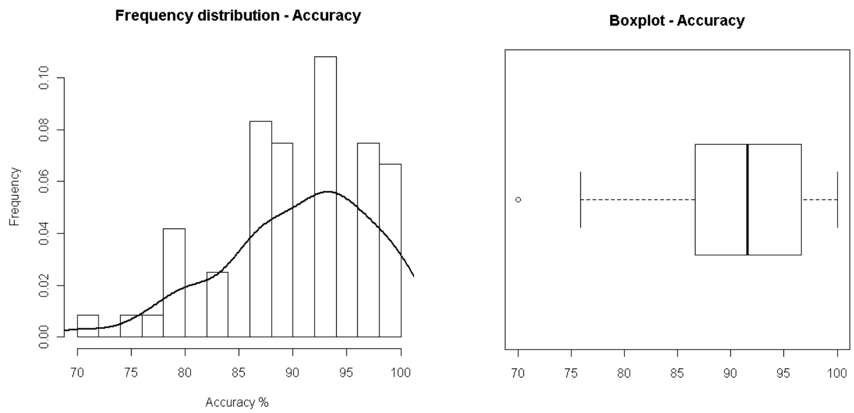

Figure 18 Frequency distribution histogram and boxplot of accuracies in all tasks

The data were slightly left-skewed and there was one possible outlier as well. This lower value was probably due to problems a subject had during the task with his equipment; a gun strap was in disturbing him for some time. However, the value was not clearly different from other values therefore no values were removed.

Shapiro-Wilk normality test was used to test the normality of the distribution. The distribution was not normal for whole data (Shapiro Wilk normality test p-value = 0.004642), but for the samples which were planned to be compared the normality assumption was met. Also Bartlett's test is not significant indicating that the variances of the measurement variable were the same for the different groups. Parametric statistical tests could have been used for these analyses, but for consistency to reaction time results, non-parametric tests were chosen for statistical tests for all accuracy data. The Friedman test was used for comparing group means, and the Wilcoxon signed ranks test for multiple comparisons post hoc tests. Original data values were used.

Appendix 9 Detailed results

This annex contains detailed results for questions of analysis. The graphs presented in the result summary in the main document are also presented here. Detailed interview results (IN1-IN6) are already fully included in the main document and thus not presented here.

Performance measures

RT1. Does the modality have an effect (visual, auditory or tactile)?

All hits from all other than the multimodal task conditions were analyzed, grouped by modality. Also the values from the mixed task conditions were included.

The group means were compared with the Friedman test. The reaction times did not significantly differ between conditions X2(2)=4 , p= .135.

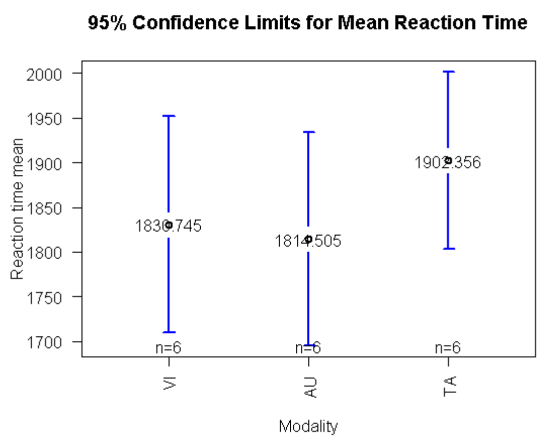

Figure 19 RT1. Reaction time means and 95% confidence limits for visual, auditory and tactile modalities

RT2a. Does the modality have an effect (visual, auditory, tactile and multimodal)?

All hits from all task conditions (including multimodal) were analyzed, grouped by modality. Also the values from the mixed task conditions were included.

The group means were compared with the Friedman test. The reaction times did significantly differ between conditions X2(3)=8.6 , p= .035. The Wilcoxon signed ranks tests were used for a post hoc test. There were only three comparisons performed, multimodal condition with other conditions. The Bonferroni correction was made and effects are reported at a .0167 level of significance. The reaction times did not significantly differ between the compared pairs. However, the significance value between tactile and multimodal condition was p= .028. The Bonferroni correction is known to be conservative especially when samples are dependent (Field 2009)(Sarna 2009) and thus the difference can be considered as indicative. Figure 20 shows that the tactile modality was slower than the multimodal cue.

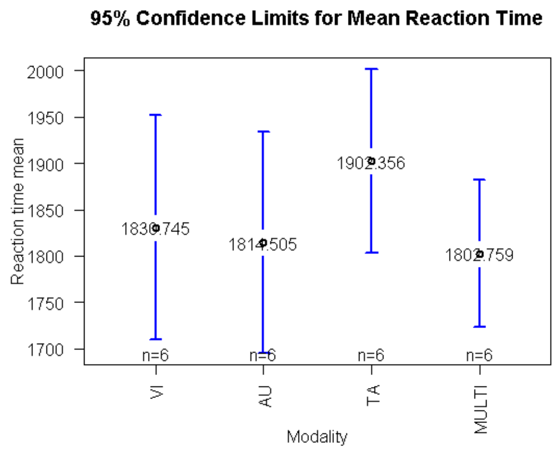

Figure 20 RT2a. Reaction time means and 95% confidence limits for visual, auditory, tactile and multimodal modalities

RT2b. Are there differences between different modalities (in tasks)?

All hits from all other than the mixed task conditions were analyzed, grouped by modality. Note that in previous questions for analysis (RT1, RT2a) reaction times from the mixed task conditions were combined to unimodal values.

The group means were compared with the Friedman test. The reaction times did not significantly differ between conditions X2(3)=5 , p= .172.

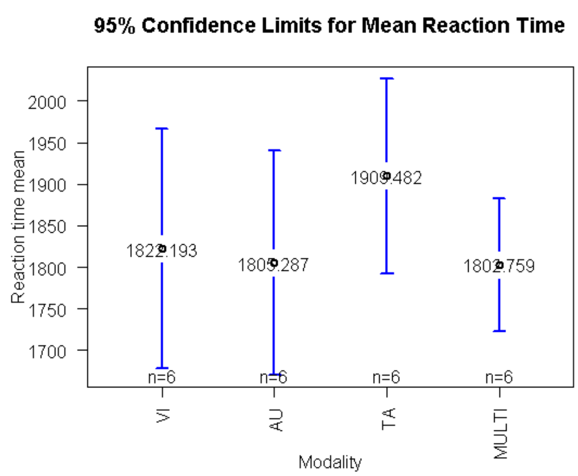

Figure 21 RT2b Reaction time means and 95% confidence limits for modalities (without mixed tasks)

RT3a. Are there differences between modalities in the mixed task condition?

All hits from the mixed task conditions were analyzed, grouped by modality. There are only unimodal cues in the mixed task conditions.

The group means were compared with the Friedman test. The reaction times did not significantly differ between conditions X2(2)=1.3 , p = .513.

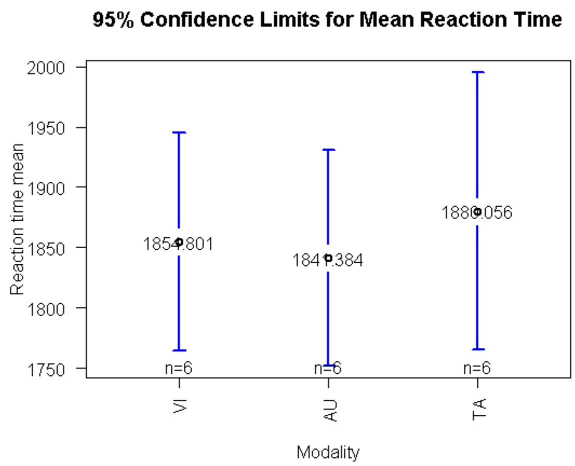

Figure 22 RT3a. Reaction time means and 95% confidence limits for visual, auditory and tactile modalities in mixed tasks

RT3.b Are there differences between the cues presented in the mixed task condition and in the unimodal task condition?

All hits from all other than the multimodal task conditions were analyzed, grouped by modality and the task type: the ones in the unimodal and the ones in the mixed task conditions. Note: the amount of shots per modality in the mixed conditions is only 1/3 of amount of shots in the unimodal conditions.

The group means were compared with the Friedman test. The reaction times did not significantly differ between conditions X2(5)=7.619 , p= .179.

In Figure 23 it could be observed that tactile cue seemed slower in unimodal condition whereas visual and auditory cues seemed slower in mixed condition. However, these differences were not statistically significant. The mixed condition (with and without physical exercise) values seemed to be in the same reaction time modality ranking order than in unimodal condition.

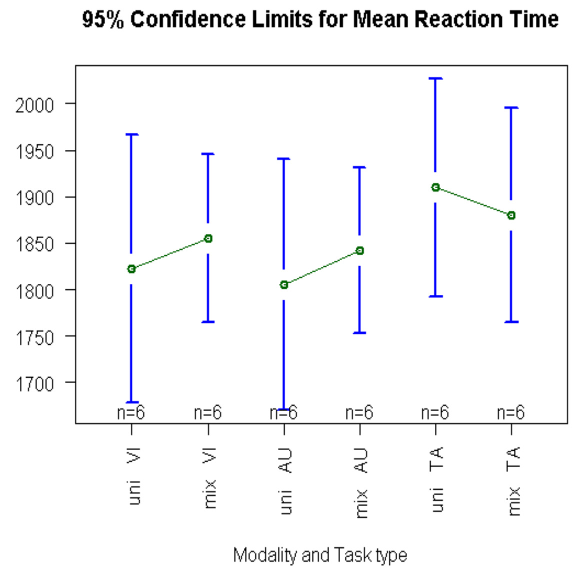

Figure 23 RT3b. Reaction time means and 95% confidence limits for unimodal cues in unimodal and mixed tasks

RT4. Does physical exercise have an effect on reaction time?

All hits from all tasks were analyzed, grouped by physical exercise: the ones with physical exercise preceding the task and the ones without.

The group means were compared with the Friedman test. The reaction times did significantly differ between conditions X2(9)=17.018 , p= .048. The Wilcoxon signed ranks tests were used for a post hoc test. Each of the modality condition pair was compared separately to find out whether there were differences between the conditions. There were five comparisons completed, see Table 7. After The Bonferroni correction effects are reported at a .01 level of significance. There was no significant difference between any of these pairs; the reaction times did not significantly differ between conditions.

Test Statistics[b]

	RVI - VI	RAU - AU	RTA - TA	RMULTI - MULTI	RMIX - MIX
Z	-1,782[a]	-,314[a]	-1,572[a]	-,524[a]	-,105[a]
Asymp. Sig. (2-tailed)	,075	,753	,116	,600	,917

a. Based on negative ranks.

b. Wilcoxon Signed Ranks Test

Table 7 RT4. The Wilcoxon test statistic comparing reaction times in tasks with and without physical exercise

The reaction time difference in tasks with or without physical exercise was significant, but the differences were not found within modality variation pairs.

For the current study an interest is to investigate how modality variations differ, and therefore the non-significant results of modality variation pairs were further analyzed. To visualize the differences (non-significant) the sample means and 95% confidence limits for six subjects are plotted below in Figure 24.

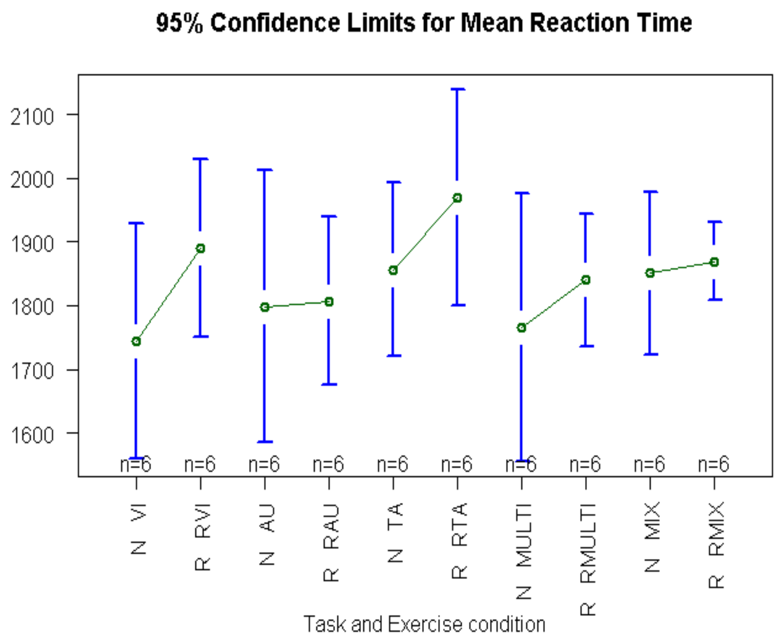

Figure 24 RT4. Reaction time means and 95% confidence limits for tasks with and without physical exercise

In the figure a trend could be observed that for most of the tasks with physical exercise the reaction time means were slower. Even if not statistically significant, the mean reaction time difference indicated possible larger difference between visual cues with and without physical exercise, and smaller difference for auditory cues.

RT5. Does target direction have an effect on reaction time?

All hits from all tasks were analyzed, also the values from the mixed task conditions were included, grouped by target direction. Differences of reaction times to the directions (left, centre and right) were compared.

For the current study an interest was to investigate how the modality variations differ, and therefore the differences in reaction times to directions left, centre and right were first compared per modality variation, and then the results of different modality variations were compared with each other.

The group means were compared with the Friedman test, the results are presented in

Table 8.

Modality variation	X2(2)	P-value
VI	2.33	.311
AU	7	.030
TA	4.33	.115
MULTI	2,33	.311

Table 8 RT5. The Friedman test statistic of differences in reaction times for directions per modality variation

The reaction time difference between directions in modality variations was significant for auditory modality and non- significant for other modality variations.

For auditory data, the Wilcoxon signed ranks tests were used for a post hoc test. There were three tests completed, comparing the directions. The Bonferroni correction was applied and effects are reported at a .017 level of significance. The reaction times did not significantly differ between the compared pairs, although the significance value between left and centre cue was p= .028. The Bonferroni correction is known to be conservative especially when samples are dependent (Field 2009)(Sarna 2009) and thus the difference can be considered as indicative, the left cue being faster than centre cue.

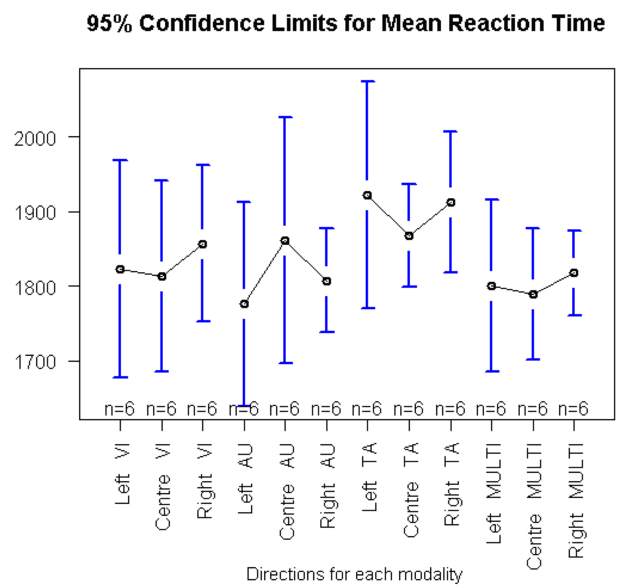

Figure 25 RT5. Reaction time means and 95% confidence limits for directions per modality variation

The comparison of modality variations with each other in Figure 25 showed that for most modality variations the side directions appeared (statistically non-significantly) slower than the centre direction, but for the auditory modality variation the centre seemed to be slower.

AC1. Does the task have an effect on accuracy?

All accuracy data for all task conditions were analyzed, grouped by task condition.

The group means were compared with the Friedman test. The accuracy did not significantly differ between conditions $X2(9)=7.29$, $p = .606$. The sample means and 95% confidence limits are plotted below in Figure 26.

AC2. Does physical exercise have an effect on accuracy?

All accuracy data for all task conditions are analyzed, grouped by physical exercise: the ones with physical exercise preceding the task and the ones without.

All tasks were either with out without exercise, and thus the comparison of means with the Friedman test was already covered in preceding analyses question. It was found that accuracy did not significantly differ between conditions.

For the current study an interest was to investigate how the modality variations differ, and therefore the non-significant results of modality variation pairs were further analyzed. To visualize the differences (statistically non-significant) in modality variations the sample means and 95% confidence limits for all tasks are plotted in Figure 26. For clarity, the accuracy means for each task are not shown in the figure. The highest accuracy mean is 93,33% for visual task after physical excercise (RVI) and lowest 86,67% for auditory task after physical exercise (RAU). All accuracy and reaction time means are presented in Table 9 (SA1). In the figure, it could be observed that for most tasks the trend was that accuracy was poorer after physical exercise, for auditory task the difference being larger than for other tasks. For visual task accuracy was however opposite, better after physical exercise. These differences were not statistically significant.

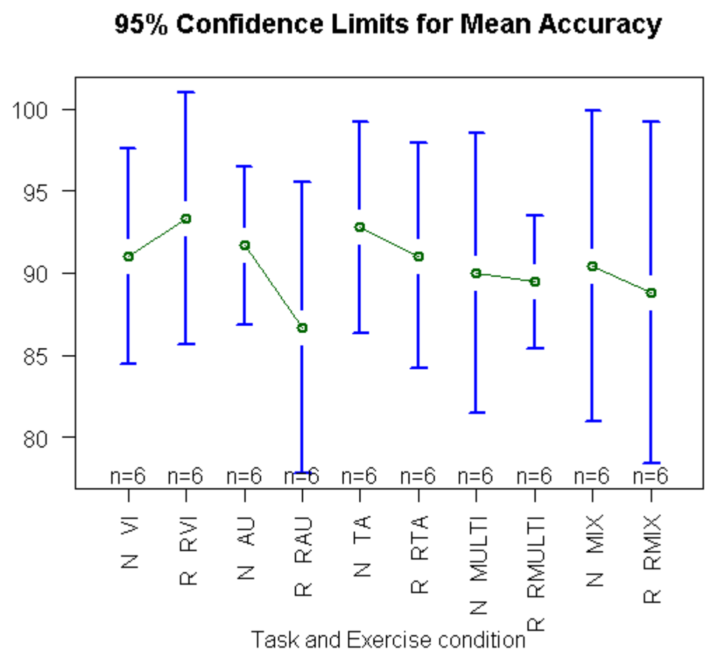

Figure 26 AC2. Accuracy means and 95% confidence limits for tasks with and without physical exercise

SA1. Is there overall speed-accuracy trade-off found in data?

Figure 27 shows in a scatter plot the correlation between the mean reaction time (also missed shots) and accuracy of each task for all the subjects. The figure does not suggest clearly neither positive nor negative correlation.

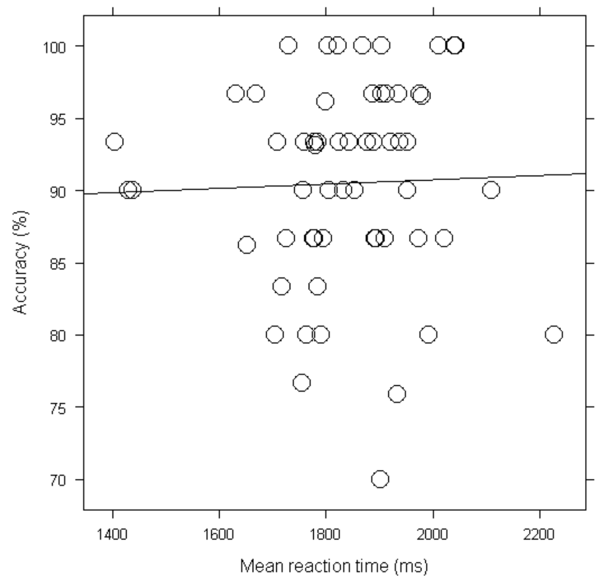

Figure 27 Correlation between the mean reaction time (also missed shots) and accuracy

The possible effects of accuracy were evaluated also by partitioning reaction time (Wickelgren 1977), dividing the tasks to two groups: the better accuracy half and worse accuracy half (for each subject). Then reaction times of these two groups were compared with the Wilcoxon signed ranks test. The result was p=0.249. Same division was also performed for reaction times, and the accuracies of the slower and faster groups were compared. The result was p=0.463. There was no statistically significant difference between the groups.

According to Lachman and colleagues (1979) the fastest responding when no errors are made represents people's processing limit. In the current experiment the reaction time for slowest shot missing the target was 2529 ms, and for slowest shot hitting the target 3422 ms. There were 28 other reaction times between those, of which one was shot to wrong direction (2953 ms). This shows that there were errors also among the slowest shots.

In following, different task conditions are investigated more in detail, in relation to reaction time and accuracy results described in preceding chapters.

Relation of speed and accuracy in tasks

Table 9 below shows the mean reaction times and accuracies of all subjects for each task. The tasks are sorted per reaction time. The mean reaction times are presented with all shots and with hits only. "Hits only" mean reaction times were slightly slower for most tasks, exceptions were RAU and RMIX. "Hits only" mean reaction times were used in statistical analyses and therefore they are used also in following speed-accuracy trade-off analyses.

Task	All shots Mean rt	Hits only Mean rt	Accuracy % (hits/shots)
VI	1743,39	1744,61	91,03
MULTI	1755,38	1765,92	90,00
AU	1791,13	1798,07	91,67
RAU	1817,66	1807,43	86,67
RMULTI	1837,16	1840,11	89,44
MIX	1840,82	1850,64	90,42
TA	1849,98	1856,54	92,78
RMIX	1875,15	1869,61	88,80
RVI	1886,38	1890,16	93,33
RTA	1959,38	1969,64	91,05

Table 9 Mean reaction times and accuracies of all subjects for all tasks

Table 9 shows that the fastest tasks have no the lowest accuracy, and the slowest tasks have not the highest accuracy.

Add-hoc ranking method

Considering that accuracy and reaction time should have equal importance in the task in the current study it is interesting to calculate an ideal combination of these two. For doing this, the values needed to be transferred to same scale ranging between 1 and 10. The fastest mean reaction time got 10 points, and the best mean accuracy got 10 point and the poorest means got 1 point correspondingly. For both accuracy and reaction time, there were equal size classes calculated, based on which each measurement was given the points, ranked. Reaction time and accuracy points were summarized to get the total points. In Table 10 the tasks are sorted by the column "total" showing the summarized points per task.

Task	mean rt (hits)	rt points	mean accuracy	accuracy points	total	acc points * 2	total * 2
VI	1744,61	10	91,03 %	7	**17**	14	24
AU	1798,07	8	91,67 %	8	**16**	16	24
TA	1856,54	6	92,78 %	10	**16**	20	26
MULTI	1765,92	10	90,00 %	5	**15**	10	20
RVI	1890,16	4	93,33 %	10	**14**	20	24
MIX	1850,64	6	90,42 %	6	**12**	12	18
RMULTI	1840,11	6	89,44 %	5	**11**	10	16
RAU	1807,43	8	86,67 %	1	**9**	2	10
RMIX	1869,61	5	88,80 %	3	**8**	6	11
RTA	1969,64	1	91,05 %	7	**8**	14	15

Table 10 Add-hoc Ranking method points per task

In the current study both accuracy and reaction time can be given 50% importance in task performance. Based on above calculation results in Table 10, the VI task condition have the best points, the best combination of speed and accuracy. The two right most columns show another variation, where accuracy is given more importance. The ideal relation between the factors to this task is not the question of this study and therefore this question is not investigated in more detail. Instead, a simple comparison between conditions was performed to rule out the speed-accuracy trade-off, see description in chapter 6.2.1 "Ruling out speed-accuracy trade-off".

SA2. Are the differences in reaction time due to speed-accuracy trade-off?

Possible differences between task conditions were investigated for speed-accuracy trade-off.

There was no statistically significant difference found in reaction times for different modalities or tasks. Indicative difference between tactile and multimodal modalities was found in one question for analysis (RT2a), but not in another one which investigated differences on task level (RT2b). Speed-accuracy trade-off analysis could be performed on task level only, because the accuracy data were collected on that level.

In Table 11 the reaction time and accuracy means of all tactile cues (in tasks TA and RTA for all subjects) and means of all multimodal cues are presented for speed-accuracy trade-off analysis.

TA	1856,54	92,78%
RTA	1969,64	91,05%
Mean	**1913,09**	**91,92%**

MULTI	1765,92	90,00%
RMULTI	1840,11	89,44%
Mean	**1803,015**	**89,72%**

Table 11 SA2. Reaction time and accuracy means for tactile and multimodal tasks for speed-accuracy trade-off analysis

Tactile modality cues were slower and more accurate than multimodal cues. There could be speed-accuracy trade-off and it is not possible to put these tasks in a ranking order by difficulty or reaction time.

The add-hoc ranking method points (see Table 10) gave 24 (16 + 8) points to tactile tasks and 26 points (15+11) to multimodal tasks. The difference is quite small, multimodal getting only slightly better ranking if both are given equal importance. The difference cannot be considered indicative.

SA3. Are the effects of physical exercise due to speed-accuracy trade-off?

Possible differences between task conditions with and without physical exercise were investigated for speed-accuracy trade-off.

The reaction time results showed a trend that for most tasks with physical exercise the reaction time means were slower. For most of the tasks with physical exercise the reaction time means were slower. Even if not statistically significant, the mean reaction time difference indicated possible larger difference between visual cues with and without physical exercise, and smaller difference for auditory cues.

The accuracy results showed that for most tasks the trend was that accuracy was poorer after physical exercise, for auditory task the difference being larger than for other tasks. For visual task condition accuracy was however opposite, better after physical exercise. These differences were not statistically significant but observed from Figure 26.

These results seemed partly opposite. In following, the reaction time means for conditions with and without physical exercise are compared together with accuracy. Each modality variation was analysed separately in Table 13.

Without physical exercise

	mean rt	accuracy
VI	1744,61	91,03%
AU	1798,07	91,67%
TA	1856,54	92,78%
MULTI	1765,92	90,00%
MIX	1850,64	90,42%

With physical exercise

	mean rt	accuracy
RVI	1890,16	93,33%
RAU	1807,43	86,67%
RTA	1969,64	91,05%
RMULTI	1840,11	89,44%
RMIX	1869,61	88,80%

Table 12 SA3. Reaction time and accuracy means for all tasks with and without physical exercise

Conditions	Analyses of speed and accuracy relation	Combination of speed and accuracy: points per task
VI- RVI	RVI is slower and more accurate than VI. There could be speed-accuracy trade-off and it is not possible to put these tasks in order by difficulty or reaction time.	VI - RVI 17 14
AU-RAU	RAU is slower and less accurate than AU. RAU may be considered as a harder task.	AU - RAU 16 9
TA-RTA	RTA is slower and less accurate than TA. RTA may be considered as a harder task	TA - RTA 16 8
MULTI-RMULTI	RMULTI is slower and less accurate than MULTI. RMULTI may be considered as a harder task	MULTI - RMULTI 15 11
MIX-RMIX	RMIX is slower and less accurate than MIX. RMIX may be considered as harder task	MIX - RMIX 12 8

Table 13 SA3. Analysis of speed and accuracy relation for all modality variations, comparing conditions with and without physical exercise

The reaction time difference for conditions within modality variation could be due to speed-accuracy trade-off for visual modality variation. For other modality variations there was no evidence of a speed-accuracy trade-off.

The add-hoc ranking method points (Table 10) were indicating a trend that all modality variations were easier (got better points) in condition without physical exercise. These observations were not statistically significant.

Visual modality variation was faster (non-significantly) in condition without physical exercise, and more accurate (non-significantly) in condition with physical exercise. For most tasks the trend was, that accuracy was poorer after physical exercise. Speed-accuracy trade-off may explain this difference for visual modality variation, as well as the larger reaction time difference between conditions.

It is also noteworthy that whereas the reaction time difference of conditions observed for auditory modality variation was small (see Table 12), the accuracy difference was larger.

Subjective measures

QU1. Which modality is preferred (overall)?

All ratings for all questions were summarized per modality to get overall rating. The rating is better the higher it is, ranging from one to seven (See Appendix 2 and chapter 7.1.1). Table 14 below presents the mean ratings for each modality.

VI	AU	TA
5,68	5,73	5,55

Table 14 QU1. Questionnaire mean ratings for each modality overall

The group means were compared with the Friedman test. The rating did not statistically significantly differ between modality conditions X2(2)=0.087 , p= 0.958. The differences would be significant if p<.050. That does not mean the subjects do not have preferences but that their preferences differ from each other. The differences between modalities were therefore evaluated also per subject. Five of six subjects had significant difference between modalities on individual level. The preferences of the subjects are presented in

Table **16**.

QU2. Which modality is preferred (per question)?

Table 15 below presents the mean ratings and standard deviations for each modality per question. Also the p-value for each question is shown indicating if there are differences between modalities. The means were analyzed per question with the Friedman test, where the differences between modalities (VI, AU,TA) were compared.

Question ID	**VI.Mean**	VI.std	**AU.Mean**	AU.std	**TA.Mean**	TA.std	**P-value**
1	**5,67**	1,51	**5,67**	1,37	**5,67**	1,17	**0.522**
2	**5,83**	1,17	**5,50**	1,38	**5,50**	1,50	**0.819**
3	**6,17**	1,17	**5,67**	1,37	**5,67**	1,26	**0.905**
4	**4,67**	1,86	**5,33**	1,97	**5,33**	1,83	**0.878**
5	**6,17**	1,17	**6,17**	1,17	**6,17**	1,17	*****
6	**5,50**	1,05	**5,17**	1,60	**5,17**	1,17	**0.949**
7	**6,33**	0,82	**6,50**	1,22	**6,50**	1,17	**0.420**
8	**5,67**	1,21	**5,67**	1,03	**5,67**	1,10	**0.810**
9	**6,67**	0,52	**7,00**	0,00	**7,00**	0,00	**0.135**
10	**5,83**	1,47	**6,16**	0,75	**6,16**	1,10	**0.943**
11	**5,33**	0,82	**5,50**	0,84	**5,50**	1,83	**0.584**
12	**4,67**	1,75	**4,33**	1,75	**4,33**	1,51	**0.646**
13	**4,00**	2,53	**4,50**	2,26	**4,50**	1,26	**0.838**
14	**5,83**	1,47	**5,83**	1,47	**5,83**	1,37	**0.368**
15	**6,17**	2,04	**7,00**	0,00	**7,00**	0,82	**0.368**
16	**5,67**	1,37	**5,00**	1,79	**5,00**	2,17	**0.779**
17	**6,33**	0,82	**6,33**	1,03	**6,33**	1,60	**0.819**

Table 15 QU2. Questionnaire mean ratings and the Friedman test statistics for each modality per question. * no differences between modalities for any subject: same rating for each modality

The differences would be significant if $P<0.05$. Results show that the differences were not significant. That does not mean the subjects do not have preferences but that their references differ from each other.

QU3. Questionnaire modality preference vs. reaction time

In following analysis, the response times were compared with questionnaire preferences answering to the question: were the subjects fastest in the modality they preferred.

Table 16 shows comparison between preference and reaction time (RT) rankings for each subject, the reaction time mean ranks were from unimodal conditions, all hits.

For one subject (ID 4) the preferred modality was also the fastest. Another subject (ID 5) had no difference in preference between visual and auditory modality, thus both were ranked first - auditory being also the fastest. Interestingly, four other subjects preferred the modality, in which they had slowest reaction time mean.

Subject ID		VI	AU	TA
1	Preference	2	**1**	3
	RT	2	3	**1**
2	Preference	**1**	3	2
	RT	3	**1**	2
3	Preference	3	2	**1**
	RT	**1**	2	3
4	Preference	3	**1**	2
	RT	2	**1**	3
5	Preference	**1**	**1**	3
	RT	2	**1**	3
6	Preference	2	3	**1**
	RT	**1**	2	3

Table 16 QU3. Comparison between preference and reaction time rankings (unimodal tasks)

Table 17 shows comparison between preference and reaction time (RT) rankings for each subject, the reaction time mean ranks were from mixed-modal conditions, all hits.

Subject ID		VI	AU	TA
1	Preference	2	**1**	3
	RT	2	**1**	3
2	Preference	**1**	3	2
	RT	**1**	3	2
3	Preference	3	2	**1**
	RT	3	2	**1**
4	Preference	3	**1**	2
	RT	**1**	2	3
5	Preference	**1**	**1**	3
	RT	3	**1**	2
6	Preference	2	3	**1**
	RT	2	**1**	3

Table 17 QU3. Comparison between preference and reaction time rankings (mixed tasks)

For four subjects the preferred modality was also the fastest.

TLX1. Does physical exercise have an effect on NASA-TLX workload in task conditions?

Mean ratings of overall workload for each task condition were compared. Overall workload was counted as an average of all dimensions for each task.

The group means were compared with the Friedman test. The overall NASA-TLX ratings for workload did not significantly differ between tasks X2(9)=15.16, p=.087.

In Figure 28 the tasks are grouped per modality variations, for each the condition with and without physical exercise is shown. There seemed to be a trend that for the tasks with physical exercise subjects rated higher workload estimates.

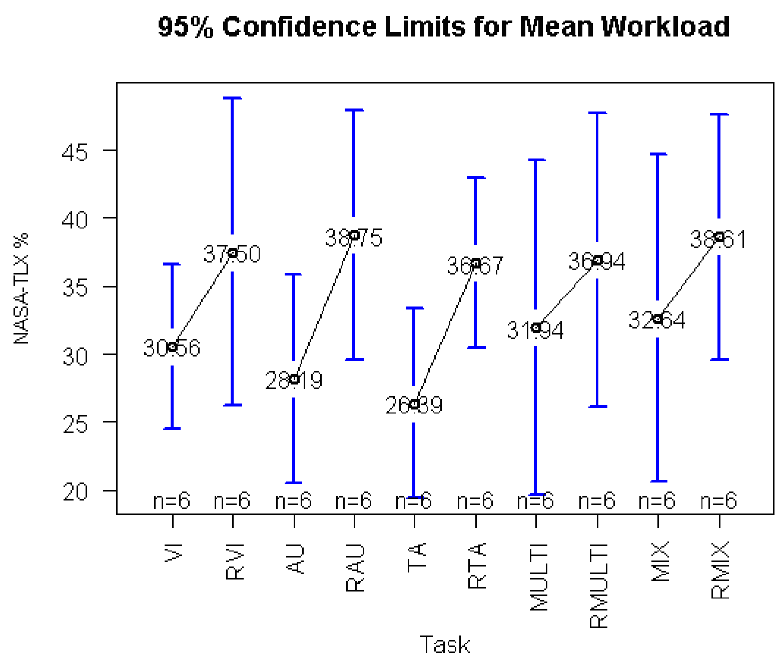

Figure 28 TLX1. Workload means and 95% confidence limits for tasks with and without physical exercise

TLX2. How dimensions contribute to overall workload differences?

The differences of dimensions were investigated. The mean ratings of selected dimensions for task conditions with and without physical exercise were compared.

To investigate which dimensions were most involved to the overall workload differences (statistically non-significant, visualized in Figure 28), the conditions with and without physical exercise were first analysed visually. Figure 29 shows the clearest differences for the dimensions "effort" and "physical demand", workload ratings being higher in the physical exercise condition. For "mental demand", "performance" and "temporal demand" they rated also higher workload in the physical exercise condition. For dimension "frustration level" subjects rated workload to be similar in both conditions. For "mental demand" workload ratings were lower than for other dimensions. In Figure 29, "n" is a condition without physical exercise, and "r" condition with physical exercise.

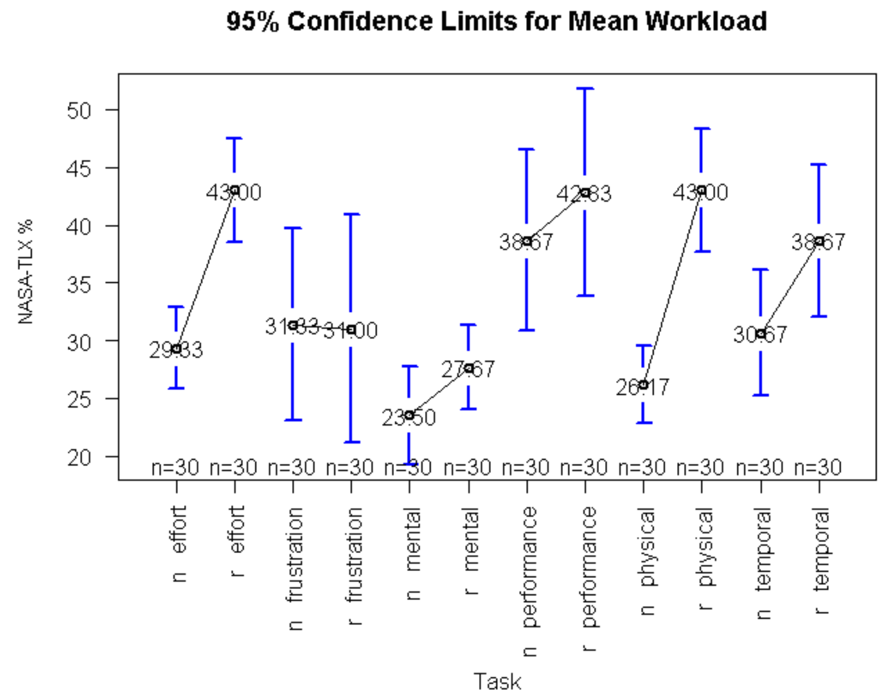

Figure 29 TLX2. Workload means and 95% confidence limits for dimensions with and without physical exercise

The group means for conditions (with and without exercise) were compared with the Wilcoxon signed ranks test for dimensions with clearest differences; "effort" and "physical demand". The Bonferroni correction was made and effects are reported at a .025 level of significance. The dimension NASA-TLX ratings for workload did not significantly differ between conditions for neither of the dimension; the results are presented in Table 18. The Bonferroni correction is known to be conservative especially when samples are dependent, thus the results could be considered as indicative, as the p-value is very close to level of significance.

In Table 18, "efr" and "efn" refer to "effort", "phn" and "phr" to "physical demand", "n" and "r" in to conditions without and with physical exercise.

Test Statistics[b]

	ph_r - ph_n	ef_r - ef_n
Z	-2,207[a]	-2,207[a]
Asymp. Sig. (2-tailed)	,027	,027

a. Based on negative ranks.

b. Wilcoxon Signed Ranks Test

Table 18 TLX2. The Wilcoxon test statistic comparing workload dimension conditions with and without physical exercise

TLX3. Does physical exercise have an effect on dimension workload in task condition?

The mean ratings of dimension workload in each task condition were compared. The workload was counted separately for each dimension. The group means for tasks were compared with the Friedman test. The results for each dimension are presented in Table 19.

Sub-scale	X^2 values	p values
effort	$X^2(9)= 27.61$	0.001
frustration	$X^2(9)= 3.37$	0.948
mental	$X^2(9)= 8.61$	0.474
performance	$X^2(9)= 15.23$	0.085
physical	$X^2(9)= 32.45$	0.000
temporal	$X^2(9)= 7.64$	0.571

Table 19 TLX3. The Friedman test statistic comparing dimension workload rating means for tasks

There were significant differences between tasks for dimensions "physical demand" and "effort". In following, these two dimensions are investigated further. The group means for two conditions (with and without exercise) were compared for modality variation pairs for each of these dimensions separately with the Wilcoxon signed ranks tests. The Bonferroni correction was made and effects are reported at a .01 level of significance.

In "physical demand" dimension there were no significant differences in workload rating between conditions with or without exercise for any modality variation. The results are presented in Table 20.

Test Statistics[b]

	rvi - vi	rau - au	rta - ta	rmulti - multi	rmix - mix
Z	-2,226[a]	-2,207[a]	-2,032[a]	-2,060[a]	-2,060[a]
Asymp. Sig. (2-tailed)	,026	,027	,042	,039	,039

a. Based on negative ranks.

b. Wilcoxon Signed Ranks Test

Table 20 TLX3. The Wilcoxon test statistic comparing tasks pairs for dimension workload of "physical demand" with and without physical exercise

To visualize the differences (statistically non-significant) the sample means and 95% confidence limits are plotted below in Figure 30. The tasks are grouped per modality variations, for each the condition with and without physical exercise is shown. There seems to be a trend that the mean workload rating is higher for tasks with physical exercise. The figure illustrates that the differences are pretty similar for all modality variations.

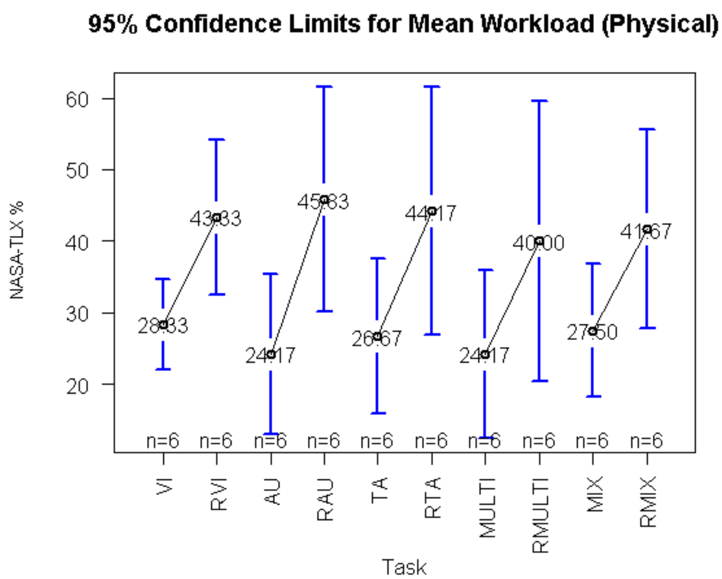

Figure 30 TLX3. Workload means and 95% confidence limits for workload of "physical demand" dimension with and without physical exercise

In "effort" dimension, there were no significant differences in workload rating between conditions with or without exercise for any modality variation. The results are presented in Table 21.

Test Statistics[b]

	rvi - vi	rau - au	rta - ta	rmulti - multi	rmix - mix
Z	-2,032[a]	-2,214[a]	-2,232[a]	-1,095[a]	-1,633[a]
Asymp. Sig. (2-tailed)	,042	,027	,026	,273	,102

a. Based on negative ranks.

b. Wilcoxon Signed Ranks Test

Table 21 TLX3. The Wilcoxon test statistic comparing tasks pairs for workload of "effort" dimension with and without physical exercise

To visualize the differences (statistically non-significant) the sample means and 95% confidence limits are plotted below in Figure 31. The tasks are grouped per modality variations, for each the condition with and without physical exercise is shown. There seems to be a trend that the mean workload rating is higher for tasks with physical exercise. The figure illustrates that the differences are pretty similar for all modality variations, although slightly smaller in mixed and multimodal condition.

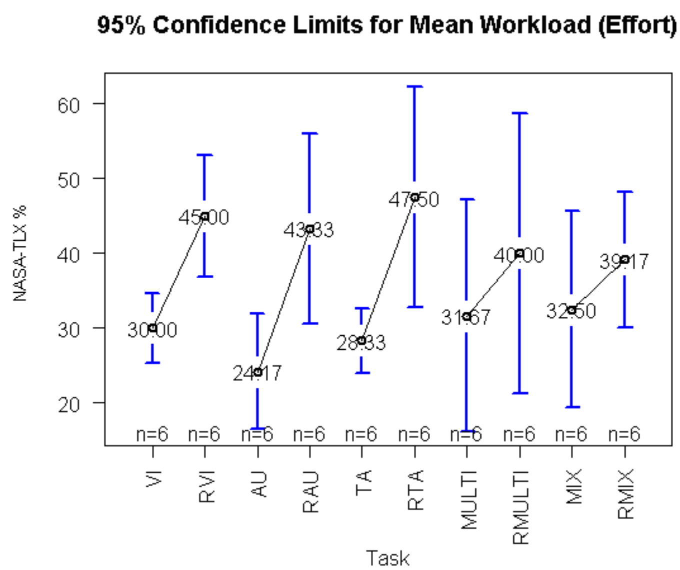

Figure 31 TLC3. Workload means and 95% confidence limits for workload of "effort" dimension with and without physical exercise

Appendix 10 CURRICULUM VITAE

Personal:

1977	Born in Espoo, Finland

Education:

2007- current	Middle European Interdisciplinary Master Programme in Cognitive Science
1998 - 2002	**Bachelor** of Business Administration from the Helsinki Business Business Polytechnic, Degree program in **Information Technology**. Thesis: Comptel MDS/SAS 5.0 GUI Usability Evaluation
1993 - 1996	High School Diploma (Matura) from Viherlaakso high school

Work Experience:

July 2008 - March 2010 (part-time)	**Research Assistant**, Brain and Work Research Centre -FIOH, Finland Cross-modal perception and attention, cognitive ergonomics
March 2008 - June 2008 (part-time)	**Trainee**, Danube University Krems, Austria Comprehension of Pictogram and Symbol Information in traffic Environments
14.02.2005 - 30.10.2007	**IS Quality Consultant**, TietoEnator N.V., Belgium Clients: European Commission
15.10.2002 - 13.02.2005	**Quality Assurance Consultant/ Test Manager**, Tietoenator Oyj, Finland Client: Nokia Oyj
12.02.2001 - 14.10.2002	**GUI Designer, Software Specialist**, Comptel Oyj, Helsinki, Finland GUI interaction and screen design, usability reviews
01.04.2000 - 30.09.2000	**Trainee, Software Specialist** Siemens AG Berlin, Germany
17.05.1999 - 31.08.1999	**Tourist Guide "Helsinki Help"** Helsinki, Finland guiding international tourists on the streets of Helsinki
08.09.1997 - 11.01.1998	**Trainee,** Espoo, Finland Art- and Media Project: theatre and journalism
25.08.1996 - 06.07.1997	**Au Pair** Leibnitz, Austria