THESIS FOR THE DEGREE OF DOCTOR OF TECHNOLOGY

Towards Intuitive Interfaces

OTTO CARLANDER

Department of Applied Information Technology
CHALMERS UNIVERSITY OF TECHNOLOGY
Göteborg, Sweden 2013

TOWARDS INTUITIVE INTERFACES
OTTO CARLANDER

Göteborg 2013
ISBN 978-91-7385-821-2

Doktorsavhandlingar vid Chalmers tekniska högskola, Ny serie nummer 3502
ISSN 0346-718X

DEPARTMENT OF APPLIED INFORMATION TECHNOLOGY
CHALMERS UNIVERSITY OF TECHNOLOGY
SE-412 96 GÖTEBORG
SWEDEN
Telephone: + 46 (0)31- 772 1000

COVER An ear, a hand and an eye, a representation of our visual, tactile and auditory senses. Brought together they symbolise our multimodal information gathering. Artwork is produced by Wilhelm and Edith Carlander, Mjölby September 9th 2012.

PRINT Chalmers Reproservice, Göteborg, Sweden 2013

TOWARDS INTUITIVE INTERFACES

OTTO CARLANDER
Department of Applied Information Technology
Chalmers University of Technology

Abstract

Operators like fighter pilots, combat vehicle drivers and fire and rescue command operators handle considerable amount of information during mission critical activities. The increasing complexity of the information landscape calls for alternative information presentation. This thesis presents the rather novel interfaces 3D audio and tactile displays, and the combination of these. In several experiments the displays are evaluated on the basis that they could improve information presentation for specific real world scenarios. The experimental environments range from laboratory to real platform in which intelligibility, reaction/response time, and localisation error are measured.

A number of key findings are presented such as further evidence for using non-individualised Head Related Transfer Function (HRTF) to acquire low localisation errors for virtual sound sources. It is also shown that 3D audio presented with a Commercials Off The Shelf (COTS) component could be used to support operators in real world settings, and that a tactile display can counteract front-back confusion occurrences as induced by 3D audio. In addition, results demonstrate that a tactile display can be used to present information to fighter pilots during threat interception, at high G-loads, and improve responses to threats. In a series of experiments concerning threat presentation to the driver of a combat vehicle it was found that combining displays for different sensory modalities can improve tactical manoeuvring performance without adding mental workload. It was also shown that the displays entailed different support for the driver depending on phases of manoeuvring towards threats.

The main conclusion drawn from this research is that these display technologies carry much promise for improving the information presentation in complex operator environments. It is mainly the omnidirectional character of the tactile and 3D audio presentations that contributes significantly to the making of intuitive interfaces, which can lead to enhanced perception and performance without adding workload. One application of these relatively new display technologies is to support visual displays in situations in which the operator already is engaged in activities occupying the visual sense.

Keywords: 3D audio, Tactile, Multimodal, Radio Communication, Command Operators, Combat Vehicle, Fighter Jet

Included Papers

The publications listed below all refer to the experiments in this thesis. They will be further referred to in the text.

3D AUDIO COTS VALIDATION

Experiments 1 and 2.

Carlander, O., Eriksson, L., Kindström M., & Chen, F. (2012). *Perceived accuracy of horizontally distributed sounds of two 3D-audio display technologies* (Unpublished manuscript).

Experiment 3.

Carlander, O., Eriksson, L. & Kindström, M. (2006). Horizontal localisation accuracy with COTS and professional 3D audio display technologies. In *Proceedings of the International Ergonomics Association 2006 conference,* Maastricht July 10-14. Maastricht: Elsevier Ltd.

3D AUDIO FOR RADIO COMMUNICATION

Experiment 4.

Carlander, O., Kindström, M. & Eriksson, L. (2005). Intelligibility of stereo and 3D audio call signs for fire and rescue command operators. In *Proceedings of the 11th Meeting of the International Conference on Auditory Display,* Limerick 6-9 July, (pp. 292-295). Limerick: International Conference on Auditory Displays.

Experiment 5.

Kindström, M., Carlander, O. & Eriksson, L. (2006). Comparison of audio systems for intelligibility in multitalker speech displays. In *Proccedings of the Human Factors and Ergonomics Society 50th Annual Meeting*. San Francisco October 16-20, (pp. 309-313). Santa Monica: Human Factors and Ergonomics Society.

BIMODAL DISPLAYS FOR A COMBAT VEHICLE AND A FIGHTER JET

Experiment 6.

Eriksson, L., van Erp, J., Carlander, O., Levin, B., van Veen, H. & Veltman, H. (2006). Vibrotactile and visual threat cueing with high G threat intercept in dynamic flight simulation. In *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting,* San Francisco October 16-20, (pp. 1547 – 1551). Santa Monica: Human Factors and Ergonomics Society.

van Erp, J, Eriksson, L., Levin, B., Carlander, O., Veltman, J. & Vos, W. (2007). Tactile cueing effects on performance in simulated aerial combat with high acceleration. *Aviation, Space, and Environmental Medicine, Vol. 78,* pp. 1128-1134.

Experiment 7.

Carlander, O. & Eriksson, L. (2006). Uni- and bimodal threat cueing with vibrotactile and 3D audio technologies in a combat vehicle. In *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting,* San Francisco October 16-20, (pp.1552-1556). Santa Monica: Human Factors and Ergonomics Society.

Experiments 8 and 9.

Carlander, O. & Eriksson, L. (2012). *Vibrotactile and 3D audio threat cueing for a combat vehicle environment* (Unpublished manuscript).

TRIMODAL DISPLAYS IN A COMBAT VEHICLE

Experiment 10.

Carlander, O. & Eriksson, L, Oskarsson, P-A (2007). Handling Uni- and Multimodal Threat Cueing with Simultaneous Radio Calls in a Combat Vehicle Setting. In *Proceedings of the 4th international conference on Universal access in human-computer interaction: ambient interaction,* Beijing July 22-27, (pp. 293-302). Berlin/Heidelberg: Springer.

Oskarsson, P-A., Eriksson, L. & Carlander, O. (2012). Enhanced Perception and Performance by Multimodal Threat Cueing in Simulated Combat Vehicle. *Human Factors, Vol. 54,* pp. 122-137.

Table of content

Acknowledgements

I have been told that it is crazy doing a PhD while working full-time and I totally agree. During this looong period I decided to quit my studies several times but every time there were people pulling me back up. It is one person that has been especially dedicated to this task using his favourite phrase “- *Improvise, Adapt and Overcome!* “. Lars Eriksson, thank you for your scientific mentorship, friendship and support throughout this work. I am grateful that you gave me the inspiration and power to finalize this. The time and effort you have invested in supporting me with this thesis has been decisive. Your KISSes will stick! Lars, I am deeply grateful, THANK YOU!

Fang thanks for your support throughout the work and I especially appreciate when you took the time to visit me in my home to discuss matters relating to the thesis. You have always believed in me and this resulted in that I was able to stay active at Chalmers. Thank you also for introducing me to 3D audio, it literally changed my life.

I thank the Swedish Defence Research Agency, FOI for the opportunity of writing this thesis. At FOI/FOA Gunilla Derefeldt introduced me to human factors and supported me from day one. Gunilla, you inspired me to continue in this area of research, thank you. I have also received some time for writing from Motorola (Linköping) and Saab Aeronautics. I’d like to send a special thanks to Matthias Trygg, Henrik Carlsson and Daniel Rydja my managers at Motorola and Saab.

For many of my studies the research engineer Mattias Kindström was helping me out. Mattias you not only brought efficient technical setups you also made the sometimes long days a lot funnier. I appreciate your insightful comments on both the experimental work and other things. I really miss the time we had at FOI, conferences and “in the field”, thank you. Per-Anders Oskarsson, I especially appreciate your help during experiment 10, and for the generous opportunity to participate on the Human Factors article. Peter Andersson, thanks for the great vibrations and other necessary hardware.

The close cooperation with real world operators has been useful and also the main theme for my research. With the help from these operators I have acquired knowledge and been able to stay focussed. Some people put extra effort into making this interaction fruitful. I would like to thank Jan Fredriksson, and Ove Grönlund at the MSS in Skövde, and the participating fire and rescue command operators in Stockholm, “Lunkan” and Anders Bergqvist are especially thanked.

The team forming the NATO group, HFM-122, was inspiring and contributed to the excitement of the everyday work. The potential bar fights in Vadstena and the great downhill skiing in Colorado are of course memories for life. Speaking of bar fights… Jan van Erp you are a true source of inspiration. Your work in the area of tactile displays is fantastic and has been vital for this thesis.

Lena Berglin, I really appreciate your help to create the tactile garments. I would like to acknowledge Martin Ek for the graphics in figures 7 and 19. Many thanks to Alexander Lowe for proofreading.

Joakim “Jülle” Dahlman, your help with early reviewing and moral support during the final stages of this thesis has been of much value. Your help came at a critical moment and contributed to my continuation on the thesis, thank you. Håkan Söderberg, another colleague at FOI opened many useful doors and helped me in the early stages of my work, thank you!

The participants in many of my experiments are also acknowledged, and especially those who were required to listen to broadband noise. On a relating note I thank Defecto Nagrobek for the music!

Last but not the least, many thanks to my parents for help and support during my studies and that you believed that I could make it. My family, Alma, Wilhelm and Edith thank you for your patience during evenings and vacations, I love you.

Wilhelm och Edith, tack för den färgsprakande framsidan och alla fina teckningar och konstverk jag fått, den här boken är till Er.

MJÖLBY, JANUARI 2013
OTTO CARLANDER

“En dag ska vi alla dö…alla andra dagar ska vi det inte” (Stenmark, 2001)

Abbreviations

CNS:	Central Nervous System
CMT:	Continuous Memory Task
CV90:	Combat Vehicle 90
COTS:	Commercials Off The Shelf
dB(A):	Decibel A-weighted
DOF:	Degree of Freedom
DRS:	Diver Reconnaissance System
DSP:	Digital Signal Processor
EAP:	Electro-Active Polymers
FPS:	First Person Shooter
HDD:	Head Down Display
HMD:	Helmet Mounted Display
HUD:	Head Up Display
HRTF:	Head-Related Transfer Functions
IAD:	Interaural Amplitude Differences
IID:	Interaural Intensity Differences
IHL:	In-Head Localisation
ITD:	Interaural Time Differences
IPD:	Interaural Phase Differences
I/O:	Input/Output
ISS:	International Space Station
LE:	Localisation Error
MAMA:	Minimum Audible Movement Angle
MB:	Megabyte
MHz:	Megahertz
Ms:	Millisecond
MRT:	Multiple Resource Theory
PC:	Personal Computer
PTD:	Perception of threat direction
PT:	Performance Time
RAM:	Random Access Memory
RPM:	Revolutions Per Minute
RT:	Response Time
SORD:	Spatial Orientation Retention Device
TSAS:	Tactile Situation Awareness System
WCS:	Warning and Countermeasures system

Chapter 1. Introduction

Operators like fighter pilots, combat vehicle drivers and fire and rescue command operators are supposed to handle a large amount of information during mission critical activities. Not always is this information well presented or adapted to the task and operative environment. A variety of technical systems that extend the information gathering beyond the human sensory capabilities adds further information that is supposed to be presented in an interpretable format. Even though sensors may be efficient and mostly convey correct information of the surrounding environment, the display presentation should better bridge the gap between the relevant information and the operator. The result of many efficient sensors and display systems might cause them to "fight" for attention in an increasingly complex information environment, and this threatens the gains of the sensor and display systems (Brill, Gilson, Mouloua, Hancock & Terrence, 2004). The increasing complexity of the information landscape therefore calls for alternative solutions for information presentation.

Finding ways of displaying information in an intuitive way is necessary to improve the intelligibility and interpretability for future user interfaces. The work in this thesis aims towards contributing to the knowledge on how to improve information presentation for operators such as the driver of a combat vehicle, the fire and rescue command operator and the pilot of a fighter jet. More specifically, the question is how required information can be better presented and utilized for our auditory and tactile sensory systems. The long-term objective is to contribute to the development of an intuitive interface. In an intuitive interface, the presented information renders an immediate understanding leading to effective and efficient use. An intuitive interface should require almost no learning and easily close the "gap" between sensation and action in that; *"an intuitive display is a display that automatically triggers the required reaction and that minimizes the use of cognitive resources"* (van Erp, 2007, p.15).

I would like to continue this introduction by giving a scenario where an intuitive interface could be of much help. In this scenario you meet two combat vehicle operator crews, one with traditional displays and one using intuitive displays.

Three combat vehicles under the United Nations (UN) flag (Alpha, Bravo and Delta), with full crew are on their way back after transporting soldiers to their post. Alpha is leading the colon of the three combat vehicles. It is very warm and the crew is tired, and the air feels sticky inside the cramped vehicle. It has been a long day full of "red alerts" and confrontations with the local population, luckily no weapons were fired this day. Apart from the standard equipment, this vehicle is also fitted with a warning and countermeasures system (WCS) that automatically can perform countermeasures towards incoming threats to save critical time in responding to threats or attacks. Even though this enhances the perceived safety of the crew they know that when the automatic countermeasure system activates it is only a very short time until the vehicle could be hit or that the crew will be required to perform a tactical manoeuvre or localizing the threat to secure survival. Thus, the situational information presented to the operator crew of the vehicle, the driver, the gunner, and the vehicle commander needs to be clear and intuitive for a good awareness of quickly evolving threat situations.

Even though the vehicles only have a maximum speed of 70km/h the driver often experiences that the three prisms (see figure 1) to view the outside environment limits his ability to get a good overview of the vehicle surrounding, and especially during darkness when the night vision prism is required to be used. The gunner's situation is similar with only three prisms for looking out to the right. However, some additional information of the surroundings can be acquired through the weapons' sights. The commander has the possibility to open up a hatch above his head and often rides with this open. This will offer him a better overview but he has limited ways to mediate this to the gunner and driver. When he can communicate, the communication must be seamless to avoid delays that could have a lethal outcome. Apart from having an efficient dialogue inside the vehicle, the crew must be able to quickly receive and process sensor and radio data from other vehicles in the immediate surroundings. The optimal solution is that vehicles set up a "sensor network" utilizing the sensor systems to continuously exchange information. Properly presented to the crew, this could be used to make better decisions based on enemy position in relation to the own vehicle.

About halfway home an enemy missile is fired from a great distance towards the first vehicle. Due to the distance between the vehicles in the colon the second vehicle cannot optically see the guiding laser of the missile. The threat is detected very late and the WCS does not activate its protecting smoke quickly enough. When the crew inside the vehicle receives the warning from the system this warning carries no directional information. All they know is that they have about three seconds before the vehicle will be hit. The commander has just enough time to observe a small hill to which he will command the driver but the time is too short. The driver needs the threat information in good time to be able to quickly position the vehicle in a tactical manner. Since the driver primarily inspects the terrain through the prismatic lookouts he is not able to get an overview of the situation. The gunner with access to the sensor data is occupied with translating the 2D presentation on his visual display to the three-dimensional world he is supposed to act within. This is especially hard when under the severe threat he is now facing. Adding to this is that the vehicle stirs up road dust and as a result the driver can only see the road straight ahead through his forward looking prism. The warning signal keeps beeping: Warning! Warning! Warning! BANG!

The vehicle is hit but luckily in the front where the protection is thick. The gunner is hurt and unconscious but the commander manages to direct the driver to safety behind the hill. Two other explosions are heard close to the vehicle and there is a lot of activity on the radio. Alpha calls out: -"We are hit, Alpha is hit". After this, it is very hard for the commander to understand anything of the situation since the crew from Bravo and Delta talks at the same time in the monoradio resulting in that voices blend and there is little intelligibility. The second combat vehicle, Bravo, is luckier. This vehicle has an updated information presentation system on trial. State of the art research has been implemented and the information presentation is better adapted to the needed information, and the operator's processing of sensory cues or information. About the same time as the first vehicle receives the warning from the countermeasures system the same system is broadcasting the warning to nearby vehicles. This is received by Bravo, and to make the crew aware of the severe threat, information is presented by spatial audio (3D audio) and tactile signals. The spatial audio alert gives information of what type of threat and the direction to the threat in relation to the own vehicle. The tactile alert is presented by a belt extending around the chest in which 12 tactile vibrators each can indicate, with 30 degree accuracy, the direction to the threat. This redundant coding of information with *auditory and tactile cueing further enhances the presentation and reduces the risk for misinterpretation. Thus, the commander gets an immediate awareness of how the vehicle is positioned in relation to the threat. The gunner "feels" the direction to the threat and by the pulses of the tactile torso belt he can also distinguish the type of threat. Before even becoming visible to the enemy he can now choose and position his weapons accordingly. The driver instinctively reacts to the warnings and stops the vehicle before being in reach of enemy sensory and weaponry. The commander authorises the halt and hears Alpha call over the radio. This confirms that Alpha is hit and completes the awareness of the situation for the commander. Since the commander of Bravo uses radio communication with spatial audio filtering he can hear the voice of the Alpha commander in the actual direction of Alpha's position.*

The above mentioned scenario is fictional but still a valid example of the ecologically relevant situations potentially encountered in the field by the combat vehicles. It is the issues brought up by situations like the above that most of the research in this thesis is based upon. Today there is an increasing amount of sensor and display systems for the crews in military systems but the information is just not optimized for presentation to the human operator. Many user interfaces for threat information in these situations can be considered lacking in representing and conveying the spatial dimension in an intuitive way. When adding the demands for quick and correct actions in life-threatening situations, the mental workload is most often high. Furthermore the operative environment might have limited space and be subjected to extreme environmental conditions such as coldness or heat, noise, and visual limitations due to fog, smoke, darkness etc.

Figure 1. Pictures showing elements of the working environment for the CV90[1] driver. Top left, the prismatics that the driver use for outside view, top right, inside view of the prismatics. Bottom, steering handle with 2D visual information. Navigational equipment (far right) and engine controls (far left) are also seen.

---

[1] The CV90 is an armoured infantry fighting vehicle that carries a crew of commander, driver, gunner, and an eight men section at the back. The vehicle can engage other armoured vehicles and ground troops and can be equipped depending on mission requirements (BAE Systems, 2006).

The Alpha crew had a limited possibility to get an understanding of the situation partly due to poor information presentation. The audio warnings with no directional information and the head-down display resulted in that crucial time was lost in building up the spatial awareness for tactical or manoeuvring related tasks. Using a 2D head-down requires the operator to take his eyes off the outside terrain to look down and interpret the information on the head-down display (HDD). Since visual displays (most often) are flat and often convey only two-dimensional information, one dimension is totally compressed, or excluded, resulting in the loss of a dimension. This will of course result in loss of information and requires the user to generate a 3D understanding from 2D based information thus loading the cognitive resources (Wickens & Hollands, 2000). Improved threat cueing could enhance the operator's task performance while perhaps contributing to lowering the overall effort (as for the Bravo crew). A visual warning such as the one used in the Alpha vehicle is spatial in nature and requires that the operator direct his eyes towards the displayed information. Both the tactile and auditory senses are omnidirectional and does, of course, not require visual attention and would thus be preferred over visual displays in driving situations (e.g. Scott & Grey, 2007; Sodnik, Dicke, Tomaič, & Billinghurst, 2008; Spence & Ho, 2008). For the Bravo crew, the auditory presentation was further improved by introducing spatial sound to increase the understanding of *what* the information represents (type of warning) and *where* the information is coming from (direction to event). This is not possible to do with a mono presentation (one sound-channel presented to both ears) and only to a limited extent with stereo (two sound-channels, presented to the ears). Thus the 3D audio presentation has benefits compared to the conventional audio displays (Begault & Wenzel 1993; Nelson, Bolia, & Ericson 1998; Vause & Grantham, 1998).

Considering the advantages of improved auditory information presentation, which does not require visual attention, it may also be advantageous to employ the sense of touch. Some studies have already shown that visual displays can be substituted with a tactile display for solving navigational tasks, (e.g. Eriksson et al., 2008; Gilson, Redden, & Elliott, 2007; Redden, 2006; van Erp, 2007; van Erp, van Veen, Jansen, & Dobbins, 2005; Walker & Linsday, 2005a). Since the skin covers our body we can receive information by stimulating the skin as a representation of a direction to an event in the surrounding world, very much like "a tap on the shoulder" (cf. van Erp, 2007; van Erp & Werkhoven, 2006).

This thesis contributes to the development of the next generation intuitive interfaces. The intuitive interfaces should for instance result in a reduction of reaction and response times, localisation errors and training efforts. The long-term objective is that of enhancing survival by accomplishing an information flow that can be intuitively perceived and coupled with the correct corresponding actions. The majority of the work contributed to projects at the Swedish Defence Research Agency (FOI)[2] aiming to investigate new display concepts for operators in complex environments. Most of the experiments involve operative personnel in some way, and they often influence the iterative laboratory studies and have served as subject matter experts from the early experimental design phases to

---

[2] FOI is a research institute in the defence and security area with its core business in research studies and technology development. The 830 researchers are mainly funded by contracts and are responsible under the ministry of defence. Some examples of research areas are sensor systems, security policy studies and analyses, systems for crisis management and control and command systems for the defence sector (FOI, 2012, p.19)

experiments in the actual platforms. It is important to get early end-user involvement since this will clarify main issues in the operational setting and also set the focus on the task that the operators are supposed to handle. Furthermore, it is an opportunity to adopt technologies and systems early so that it paves way for actual implementation.

Even though a substantial body of research has been carried out regarding information presentation utilizing auditory and tactile information, applications for operators in complex environments need further investigations. For some operators like combat vehicle crews or fighter pilots, information presentation is further complicated by a hostile environment. One common and important piece of information that needs improvement is the spatial information (e.g. the whereabouts of oneself in relation to other units, targets and threats). Depending on platform this has traditionally been handled by speech (radio), hand signals (e.g. dismounted soldiers), maps or electronic 2D visual displays.

Research questions

In the late 90’s I was a research assistant at FOA (that later became FOI) during some experiments (Ayama et al., 2004; Carlander, 2000). This resulted in that I was offered to perform my degree project in neuropsychology concerning peripheral colour vision there (Carlander, 2000). During my time at FOA I heard about 3D audio and begun to read some articles in the area. 3D audio fascinated me in that a traditional display, such as the audio display, could be improved and more intuitive by adding spatialisation. The fascination continued and in the early 00’s when I was about to finalize my studies in Human Factors (Loughborough University, UK), I initiated discussions with Fang Chen regarding a degree project in 3D audio at the Swedish Centre for Human Factors in Aviation (HFA). After completing this project (Carlander, 2002) I was offered employment at FOA and I started to work as a researcher in the area of audio interfaces. At FOI my research has been sponsored by two main projects, “operator site” (Eriksson, Carlander, Borgvall, Dahlman & Lif, 2005) and “multimodal interfaces” (Eriksson, Lindahl & Hedström, 2006). Already before these projects we had ongoing discussions with a Research institute in the Netherlands, the TNO. At the time they had mature research in the area of tactile interfaces and this contributed to that we could take on a wider scope in our projects including tactile interfaces. The project Operator Site focused on research activities concerning display concepts for improving operator performance in a range of settings and platforms. This was primarily auditory, tactile, and visual interfaces with the overall goal of improving operator situation awareness and performance. For our projects we applied the interfaces to a variety of settings and operators and the most important for my studies were the driver in the Combat Vehicle 90, the pilot in the Gripen fighter aircraft, and the fire and rescue command operator in the command and control central. This project was followed up in a new project, multimodal interfaces that involved concept development for display interfaces and also aimed at developing some guidelines for adjusting these for operational support.

My overall ambition is that the research in this thesis will contribute to the development of intuitive interfaces where the information can be better perceived and result in a better performance. As a basis for my investigation the following research questions are examined:

1. Can a portable 3D audio system generate spatial audio that is accurate enough for operational use?

2. Can 3D audio enhance the intelligibility of call signs?

3. Can 3D audio and visual threat indication be improved by adding a tactile presentation?

4. What are the main benefits of auditory, tactile and multimodal threat cueing in a combat vehicle?

5. Can 3D audio and tactile displays be used effectively and efficiently without extensive operator training?

Thesis limitations

The research presented in this thesis was performed over a long period. In the early 2000s having a portable system that could be used for field applications was an important finding and breakthrough to be able to perform more ecological studies. A lot has happened to the technology during this period but the general theme and conclusions are still relevant. The extension of this research is still very much alive and has been continued at FOI.

Even though this contributes to the research and development of intuitive interfaces, it should be pointed out that it does not take on the complete scope of intuitive interfaces looking into general guidelines and applications. Perhaps of more importance to mention is that the thesis will not cover information processing in general. This information is important and would further explain and enhance the area of multimodality. Although this is not specifically covered each chapter briefly mentions aspects of information processing.

Methodology

A general theme throughout my research is that aspects of issues relating to end-user task or environment are investigated. Thus, the concept development process described below can be considered as the general philosophy embracing the thesis. The experiments have followed a user-centred design process and attempt to explain responses to certain independent and dependent variables like display technologies, localisation error, and reaction/response time. For the experiments, we used within-subject multifactorial designs with parametric statistics.

To be able to improve and make the user interfaces more intuitive, operators “in the wild” should be involved to facilitate the researchers’ understanding of the task and environment. The complexity of the applied setting often sets boundaries that help guiding the actual experimental design for testing specific aspects of the behaviour (e.g. Hollan, Hutchins, & Kirsh, 2000). As a result, domain experts were included early in the process to ensure a human-centred focus and that hypotheses were directed to relevant issues. The continuous dialogue and refinement of research questions and equipment offered a potential for the operators to be able to use the system as a proof of concept immediately after the experimental phase. Being able to demonstrate the “real” system, will aid evaluation or even a change of current processes or tactics before a larger scale implementation and integration takes place. Another advantage with this process, as

Eriksson et al. (2006) points out, is that it has a potential to accelerate and adapt the implementation of the research results and thus decrease the total cost of a development project.

The work process was iterative; from laboratory experiments to real platforms and tasks. Thus, it took place at different system levels and degrees of realism with both expert users and inexperienced participants. The feedback in the iterations generated loops that were used to refine the hypothesis that in turn was applied to the next experiment. This forms a cycle that was used to refine each step in the research process. The process can be divided into several stages of complexity, and is described in Eriksson et al. (2006). Basically each stage serves as the basis for the subsequent stage and the results from the subsequent stage are looped back to refine the research questions and methodology. The loops could also serve to ensure a reliable measure since many experiments were based on a similar methodology.

The general process can be simplified in the three stages (1-3) below:

1. Low level of realism and task execution (see figure 6).
    - Try out concepts and answer basic research questions relating to perception and performance.
    - Inexperienced participants.
2. Medium level of realism and task execution
    - Simulator studies with relevant tasks. (See figure 29)
    - Inexperienced subjects and/or real operators (See figure 11)
3. High level of realism and task execution
    - Real platform and users. (See figures 24 and 25)

The above levels are not discrete stages, level 2 and 3 are iterative and loops back to the subsequent level. Level 1 and 3 are connected by the continuous discussions with experts. The major difference between the levels is the degree of realism. The experiments concerning 3D audio as described in experiments 1, 2 and 3 (Carlander, Eriksson, & Kindström, 2006; Carlander, Eriksson, Kindström, & Chen, 2012), were laboratory studies investigating horizontal localisation accuracy, and belong to *stage 1*. The results served as the basis for the radio communication and threat warning studies that followed in experiments 4, 5 and 7 to 10 (Carlander & Eriksson, 2012; Carlander, Eriksson, & Oskarsson, 2007; Carlander & Eriksson, 2006; Carlander, Kindström, & Eriksson, 2005; Kindström, Carlander, & Eriksson, 2006; Oskarsson, Eriksson, & Carlander, 2012). The radio communication experiment with the fire and rescue command operators (Carlander et al., 2005) and the experiment with threat indications and radio calls in the combat vehicle simulator (Carlander et al., 2007; Oskarsson et al., 2012) are examples of *stage 2*. These two studies represent different aspects of this stage. For the fire and rescue command experiment we used real operators in a simulated task similar to their everyday work. For the CV90 study (Oskarsson et al., 2012) with threat indications and radio calls, the task itself was quite realistic - react to threat warnings and handle simultaneous radio calls - but the participants had no prior experience of a combat vehicle environment. Experiment 7 with the WCS in the CV90 (Carlander & Eriksson, 2006), and the bimodal presentation in the Gripen fighter jet, experiment 6 (Eriksson et al., 2006) are examples of *stage 3*. Here, we utilized a real platform or a condition with real operators during a realistic task. Experiment 7 built on both published studies (Carlander & Eriksson, 2006) and on unpublished work with mock-ups, tests and discussions with operators and experts.

The next chapter begins with an overview over the auditory sense and the spatial audio technology. The fundamentals for tactile displays and perception then follow. Both chapters cover issues relating to the implementation and design of the auditory and tactile displays. In the fourth chapter auditory and tactile displays are brought together into multimodal displays and the theories behind such a display are discussed. The general discussion, conclusion and final remarks conclude this work. Throughout this thesis experiments that are relevant to the concerned chapter are presented.

Chapter 2. Hearing and 3D audio

During evolution we have been continuously exposed to sounds carrying information from different directions. We have adapted to listening to simultaneous sounds and extracting the information from the source we desire (Withington, 2000). The capability for quickly and correctly judging the position to a threat has been of vital importance for our survival. Even though our accuracy in sound localisation alone is rather poor, about 4-10 degrees (Begault, 1999a), important advantages like omnidirectionality and a capability of raising attention are capabilities connected to audition. Our ability to simultaneously interpret both the direction to and content of a sound (Ericson & McKinley, 1997) was recognised and used already in the early information systems. One example was the attempt to increase the navigational safety and accuracy for ships in foggy weather. For this purpose Professor A. M. Mayer (1880) invented the topophone as a means of helping the navigator to determine the direction to and position of a specific sound source (see figure 2 below).

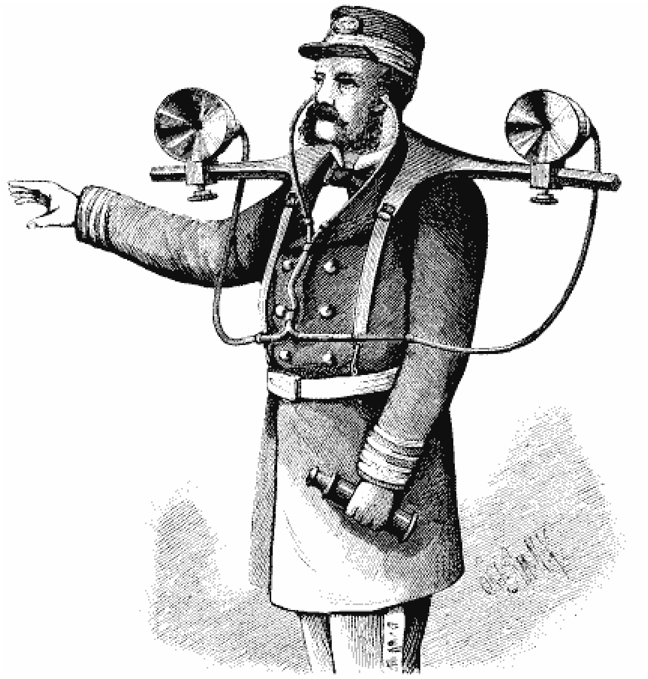


Figure 2. "It is obvious that with such a help the pilot in a fog need never be long in doubt as to the direction of a warning signal; and if need be he can without much delay, by successive observations and a little calculation, determine, approximately at least, the distance of the sounding body." (Navigation In Fogs, 1880, p.8)

Background to hearing

Before being able to make use of the information brought to us by sound, several complex stages of physiomechanical and psychological processing are required. The ear can be divided into three major anatomical divisions; the outer ear, the middle ear and the inner ear (Moore, 2003), as shown in figure 3 below. Sound energy is collected in the outer ear by the pinnae that work as directional reflectors adding a specific pattern of spectral changes to each incoming sound. These spectral changes are essential for the localisation of sound in the vertical median plane since interaural cues (differences between the ears) does not reveal vertical information (Blauert, 2001).

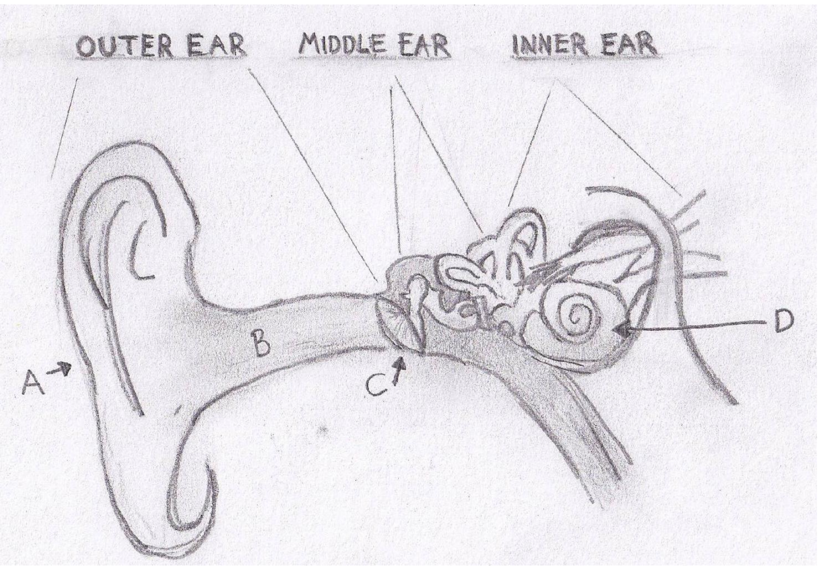


Figure 3. The anatomy of the ear (picture redrawn from Adams, 1989). (A) The pinnae, (B) the auditory canal, (C) the ear drum, and (D) the cochlea.

Detection

After the signal has been collected by the ear several processes are involved in the reception of the auditory signals. When the presence of a signal can be determined by a listener it is defined as detected. For proper detection, an auditory signal should be at least 500 ms in duration. This is due to a 200-300 ms in-ear signal-build-up, and a 140 ms decay period. Increasing duration makes the signal easier to detect but increasing presentation beyond a few seconds will not improve detectability much. Detectability can be improved by equalizing the effect of noise in frequencies carrying information (critical frequencies), and also move critical frequency components of a signal to a frequency where the noise intensity is low. For example, if a warning signal has its sound spectra with most of its power where also noise in the surrounding environment is present, the frequency spectra for the warning signal can be adjusted to a frequency where noise intensity is low. A similar kind of filtering processes is performed subconsciously (and consciously, please refer to the “cocktail party problem” below) by the Central Nervous System (CNS) that under the right circumstances cancel noise (Wightman & Kistler, 1993).

Our ability to choose and attend to signals are dependent on how we perceive the interactions of the intensity and frequency dimensions; the relative discrimination. The relative discrimination can be used to differentiate between signals presented close or simultaneously in time, but in the complexity of real life situations other variables such as the relative strengths of harmonics and the phases of sounds may also be involved (Wightman & Kistler, 1993). The ability to identify a specific sound among several others is called absolute identification and it is crucial for the ability to discriminate between auditory dimensions, and for our interaction with the environment. A simple activity such as being able to maintain a conversation or identify the sound of an approaching car is heavily dependent on this process (Blauert, 2001; Wightman & Kistler, 1993). The attention grabbing effect of directional sound was emphasised by Cherry (1953) when he explained ”the cocktail party problem”. The phenomenon can be described as our ability to focus our listening to a single spoken sound source in conversations and background noises (Arons, 1992). This ability arises from the fact that an attended signal with a certain direction is less effectively masked by undesired noise from other directions (the “cocktail party problem”). The basic concept is an evaluation process in the CNS that cancels noise by detecting differences in signals at both ears and then enhances or suppresses depending on the characteristics of the desired signal. Sounds are thus evaluated with respect to the relevancy of their content; less important sounds are ignored whilst more important are attended to (Blauert, 2001). For example, if noise is present at both ears and the desired signal at one ear only, attention can be shifted to attend the signals from the target ear thus suppressing noise while focusing on the target signal (Wightman & Kistler, 1993). The process is “hypothesis driven” where the auditory system sets up a hypothesis for a certain pattern of data to equal the desired information, and the auditory input is checked towards the hypothesis to be accepted or rejected. This pattern-recognition helps to distinguish between auditory objects in the environment and based on this we are able to construct an internal auditory model of the world. This “auditory map” is maintained and updated with information from other sources such as sensory modalities or knowledge, or a combination of both (Blauert, 2001).

Localisation

The ability to determine the location of a sound involves the integration and comparison of simultaneous information from both ears (interaural) (Blauert, 2001). The initial investigations and documentations of directional sound were performed in the 19th century by Lord Rayleigh who documented the two most important cues, Interaural Time Differences (ITD) and Interaural Intensity Differences (IID) (See figure 4 below). He found that when using pure sinusoids, spatial information was derived at high frequencies for IID and at low frequencies for ITD, thus forming the “Duplex theory” (Rayleigh 1907). These can also be referred to as Interaural Phase Differences (IPD, part of ITD) and Interaural Amplitude Differences (IAD). ITD and IPD are more or less the same thing, but their relation changes with the frequency of the tone since a given frequency has a given relationship between the two parameters (Zhang & Hartmann, 2006). A specific signal will have a certain “timing” and “signal phase” at the ears. The IID strongly relates to the wavelength of sound source. For low frequency sound sources, little or no difference in sound intensity is revealed between the two ears (Sartain, 2000). This is because the longer wavelengths allow sounds to “bend” around the head without losing much energy. However, above 1500 Hz, the difference can be noticed since sound intensity is “head shadowed” and the ITD thus becomes more important as a cue. The ITD can be defined as the time it takes for the sound to travel from one side of the head to the other. For a typical sized human head the maximum time delay represents a difference of arrival time of about 0.8 ms (Blauert, 2001). A time delay also results in that frequencies are out of phase (IPD as mentioned above) when reaching the opposite ear and the phase-shift is most effective for frequencies below 1500 Hz.

Even though the IID and ITD are effective at different frequencies they both have weaknesses in the midrange, 1500-3000 Hz, where frequencies produce neither effective time nor intensity difference cues (Sanders, 1993). To improve localisation, low frequency sounds should therefore be complemented with higher frequencies and vice versa. Broad bandwidth sounds provide means to better resolve spatial ambiguities compared to any narrow frequency band sound (Blauert 2001; King & Oldfield, 1997; Middlebrooks & Green, 1991).

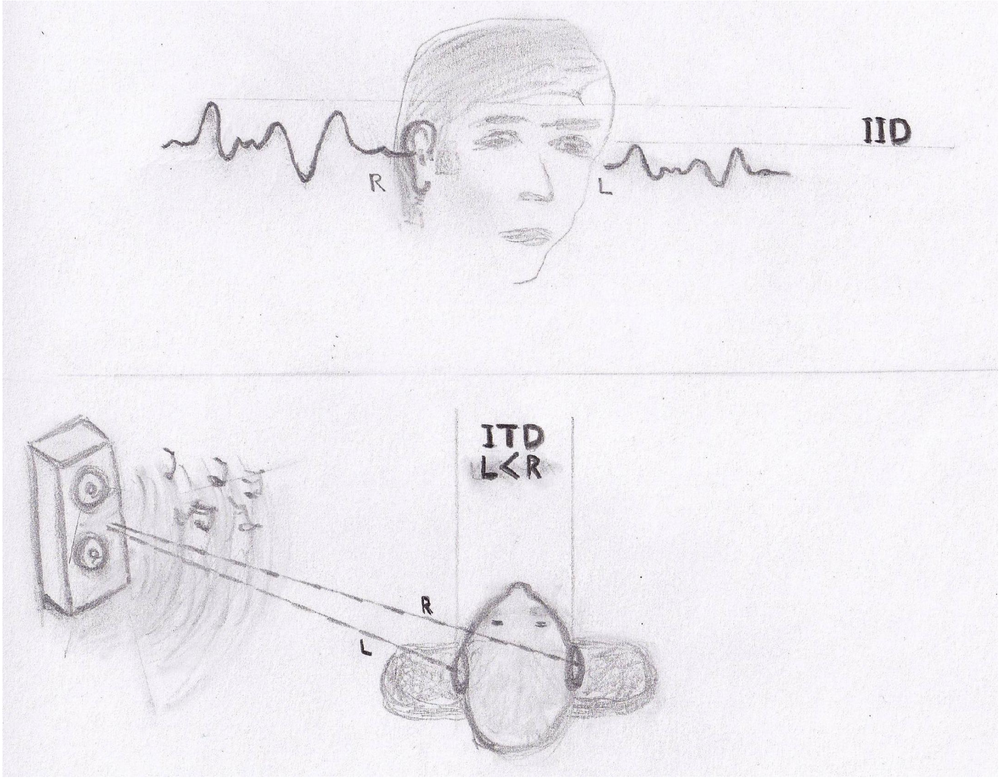


Figure 4. Top of the figure shows Interaural Intensity Difference (IID). A sound source presented to the right will have a higher amplitude in the right ear (R) compared to the left (L). Bottom of the figure shows Interaural Time Difference (ITD). A sound source presented to the left will reach the left ear before reaching the right. Thus, the time it takes for the sound to travel from the sound source (to the left) is shorter for the left ear than for the right.

Furthermore, the cues provided by IID and ITD are limited to azimuth information only. Localisation requiring other dimensions must utilize other cues such as *monaural differences, dynamic cues or non-acoustical factors* (Wightman & Kistler, 1993).

The monaural (one ear) cues

The monaural cues can contribute to the position of a sound by the pinnae and the head that affect the spectrum of an incoming signal resulting in a complex direction-dependent filter. Additionally, frequencies between 1000-3000 Hz are reflected on our shoulders resulting in a subsequent delay before the signal reaches our ears (Wightman & Kistler, 1999). The monaural spectral cues serve primarily to define front-back, elevation, and the distance to an auditory event. As described above, the spectral cues of a sound source are facilitated by broadband content since a narrowband tone is generally harder to localise (Middlebrooks & Green, 1991). However, to localise sounds solely based on monaural spectral cues is difficult, and a familiarity of the sound would be necessary. We are trained from childhood to recognize both the character and the position of certain sounds, and in our everyday life listeners learn the basic spectrum of the most common sounds. For example, the barking of a dog has a different frequency spectra close to the listener compared to further away. In effect a listener can in some way conclude where a sound source is based on its spectral content (Blauert, 2001). The distance to sound sources can also be psychoacoustically altered. For example, spoken sounds can be changed by the tone in the voice. If the intensity of two sounds is kept similar but one is “shouted”, the “shouted” sound will appear as farther away (Middlebrooks & Green, 1991).

Dynamic cues

The environment in which we interact with sounds is most often dynamic; either the listener, the sound source itself, or both are moving. These dynamic cues facilitate localisation and are most effective when the head is turned to face the sound source (Middlebrooks & Green, 1991). Thus, when the head is turned in the vertical axis (yaw), one of the ears will become closer to the actual sound source, changing both interaural time and intensity differences providing additional localisation cues (Wightman & Kistler, 1993). This enhances the interaural cues reducing the localisation blur by bringing the auditory event closer to the region of the sharpest hearing (Blauert, 2001). Head movements can be divided into two classes; one constituted by unconscious, reflexive head movements initiated towards a perceived, subconsciously known, auditory position. The other is a search and orientation movement with the goal to gather information for a final judgement of the auditory position (Blauert, 2001). The first class can be compared to the sudden head movement towards an unexpected event such as the crack of a branch somewhere in the woods when you are out walking, and the second could be exemplified by the head movements when trying to identify the direction to someone shouting in the woods. Head movements will not help much for short duration sounds since there is not enough time to effectively utilize the cues imposed by the movement (Middlebrooks & Green, 1991).

Nonacoustical localisation

For nonacoustical localisation, vision can play an important role by strongly influencing the perceived position of a sound. *“What the subject sees during sound presentation, and where the subject sees it, are factors determining the position of the auditory event”* (Blauert, 2001, p. 193).

This phenomenon is called the Ventriloquist effect and occurs when vision affects the perception of a sound source by making it appear close to what “visually makes sense”. It is a displacement of the auditory event towards a visual stimulus (Thurlow & Rosenthal, 1976), such as the sound from your TV emanating from the speakers while we perceive it as coming from the visual image and the lips of the person talking. This can be explained by the fact that the more detailed and reliable information about the external world a modality offer, the more it dominates the perceptual experience (Eimer, 2004).

Confusions and reversals

The Duplex theory (as described above) is not sufficient to explain sound localisation in the median plane (the central axis of the head), where single frequency sounds will reach the ears simultaneously with little or no interaural difference. This means that sounds will have the same IID and ITD at several positions with the effect that the sound will be perceived as being somewhere “in the plane” but the listener will not be able to tell if the source is in the front, back, above, or below (Blauert, 2001).

The problem is often referred to as the “cone of confusion” because when plotted, the confusing locations for a single frequency (the same ITD and IID) form the shape of a cone as shown in figure 5 below. The “cone” extends from the ear and the points forming the surface of the cone have the same distance to each ear respectively. Thus, the ITD and IID cues would be equivalent for a given frequency at several positions (Moore, 2003).

However, the cone of confusion is based on a simplistic modelling of the head (a perfect sphere) (Moore, 2003) but for a real human head, ITDs and IIDs can never be completely identical due to the effects of natural asymmetries, especially from the pinnae. Furthermore, listeners learn associations between sound positions and the corresponding acoustical cues that are further aided by head movements (Blauert, 2001; Moore, 2003). The "error occurrence" is thus dependent on the type of signal that is used; a narrowband input tends to have three times larger error rates compared to broadband input that contains more frequencies ('information') (Blauert, 2001).

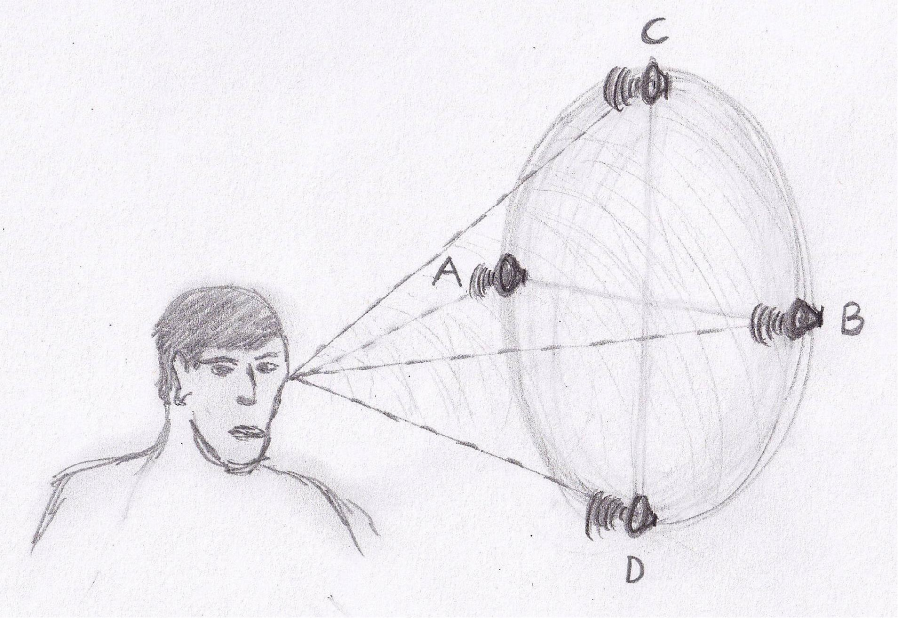


Figure 5. The cone of confusion (adapted with inspiration from Wilson, 2012 and Moore, 2003). For single frequencies "on the surface of the cone", the same ITD is produced. As a result, sounds presented at A, B, C and D will have the same ITD following the projected shape of the cone.

Virtual spatial hearing

The knowledge of how we process audio information can be used to create a virtual auditory space. We are able to perceive a full 3D space with only two ears due to the cumulative effects of the pinnae, head, shoulders, and torso on the sound wave. In the 80's Genuit (1986) created a system for accurately reproducing these signals into binaural recordings. The experience at playback of these recordings is as if the listener had been present at the actual recording. These effects can also be expressed in a pair of filters, one for each ear, with its characteristics depending on the frequency and direction of the sound. The filters are accomplished by recording what is happening to a sound when it travels from the original source to the eardrum of the listener (Wenzel, 1992). Thus, this is the difference between the spectrum of the actual sound source compared to the spectrum of the sound at the eardrum (Moore, 2003). Adequately reproduced, this can be used to create a virtual spatial acoustic imagery brought to the listener by a 3D audio system and stereo headphones (Wenzel, 1992). Using 3D audio systems with headphones offer a complete control over the acoustic waveforms without having to take the room acoustics into account. For 3D audio systems supraaural (on the ear) or circumaural (around the ear) headsets are the most common. Circumaural headsets are mostly used since frequency responses are better and greater isolation from external noise is achieved. Also in-ear headphones can be used if this is required by the application (e.g. Carlander & Eriksson, 2006). Several factors can possibly contribute to influences on the overall character of the output signal, and therefore, compensatory equalization is sometimes necessary because playback devices, such as a specific pair of headphones, have a frequency response of their own. The compensatory equalization is applied to the spatial sound to make the frequency response of the playback device transparent (Begault, 1994). Failure of reproducing close enough frequency spectra can cause confusions and "in-head localisation" (IHL). IHL results in that sound sources fail to externalize and are perceived as originating inside the head (Blauert, 2001). Figure 6 below shows an example of a 3D audio system.

Figure 6. A 3D audio system with a head tracker, headphones and a corresponding visual interface. The head tracker, attached to the head of the participant compensate for head movements allowing a more natural sound localization process (Chen & Carlander, 2003b).

Since the virtual acoustic simulation is based on our natural sound perception the individual capability of virtual spatial hearing is dependent by our normal capability of sound localisation (Wenzel, Arruda, Kistler, & Wightman, 1993). It is thus not possible to turn a "bad localiser" into a "good localiser". However, a high performance 3D audio system, equipped with a head-tracker, can be comparable to our normal spatial hearing, with ears uncovered, head moving, and in interaction with other sensory input (Blauert, 2001; Kato, Uematsu, Kashino, & Hirahara, 2003). The technology is good enough to offer localisation within a few degrees of precision and within the fraction of a second (e.g. Abildgaard-Pedersen & Jörgenssen, 2005; Bronkhorst, 1995).

Head-Related Transfer Functions (HRTFs) and localisation

The filters used for virtual spatial hearing, are called head-related transfer functions, HRTFs. More specifically, the HRTF measurement is acquired by placing small microphones in the ear canals of a subject or using the ear canals of an artificial head (dummy head). The impulse response for a single, very short sound pulse or click is measured at both ears simultaneously, and the sound is presented from a loudspeaker at a particular location in a particular acoustic environment (Bronkhorst, 1995). The process is called direct measurement and a simplified illustration can be found in figure 7 below.

Direct measurement is the most accurate way of producing spatial audio displays but the measurement process is time consuming, and most 3D audio systems do not use individually adapted filters (individualised HRTFs) but instead "generic", modelled filters. The utilization of generic HRTFs is a critical issue for being able to implement and use 3D audio in a larger scale. The consequence of using someone else's, or modelled HRTFs, is that it creates changes in the normal sound spectrum for the listener, resulting in that the correspondence between the actual position and the perceived position can be blurred and tend to cause a larger amount of front-back confusions (Wenzel, 1992). Some attempts have been made to generate algorithms or models that can be applied to the generic filters or to rapidly model individualised filters using fewer parameters with only the relevant features for an individualised 3D audio experience (Dellepiane, Pietroni, Tsingos, Asselot, & Scopigno, 2008; Hwang, Park, & Park, 2008; Martens, 2003). Finding ways to rapidly adapt non-individualised HRTFs could be useful for the implementation of 3D audio for a general population of users, and to achieve a more realistic 3D audio experience.

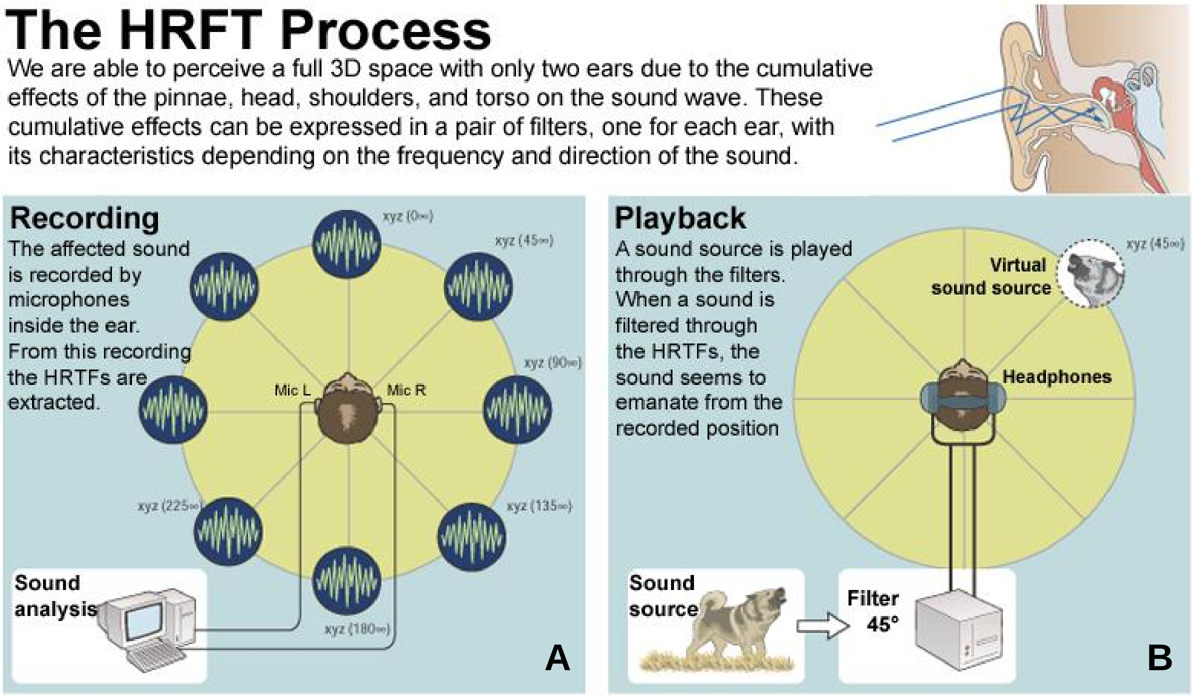


Figure 7. Extracting head related transfer functions (HRTFs) from binaural recordings. A sound is played through a speaker (e.g. at 45 degrees). The sound is being affected by the ear and the head, and the changes are recorded by microphones inside the ear of a subject or dummy head. This recording is compared to the original sound and from the difference the HRTFs are extracted. When a sound later is filtered through the HRTFs, the sound seems to emanate from the recorded position (Illustration with permission from Eken grafik).

Plenty of studies suggest that non-individual HRTFs work well for the simulation of a virtual auditory space and also result in high accuracy (Begault, Wenzel, Lee, & Anderson, 2001; Chen & Carlander 2003b; Møller, Sørensen, Jensen, & Hammershøi, 1996; Wenzel et al., 1993). Part of the explanation for our ability to localise sounds with another person's HRTFs is our ability to rapidly adapt to auditory stimuli. Long-term effects of this adaptability were investigated in a study by Hofman, Riswick and Opstal (1998), in which they showed that the adult human auditory system is capable of considerable adaptation in response to altered spectral cues, and that the auditory system is capable of holding two different sets of HRTFs. However, the relearning is not complete. Results demonstrated by Shinn-Cunningham and Durlach (1998) show that even though subjects are able to learn remapping between acoustic cues and physical locations the systematic errors that occurred was never completely overcome, and with training the errors only grew smaller in magnitude (Shinn-Cunningham & Durlach, 1998).

In table 1 below, a summary of the primary cues for the spatial audio dimensions are presented. As described above these cues are ITD, IPD, IID and HRFTs. Secondary cues, like echoes and reflections further enhance the accuracy but are not vital for localisation.

Table 1. Primary cues for sound localisation in a 3D audio system (Sartain, 2000).

	Azimuth	Elevation	Distance
ITD	X		
IPD	X		
IID	X		X
HRTF		X	

How to move the technology into the field?

As described earlier our sound localisation accuracy is rather poor (compared to vision) but we still have an innate trust, and capability for quickly recognizing sound positions (Begault, 1999a). Using the omnidirectional character of our hearing in modern display systems, can add intuitive spatial information to the system. The spatial sounds can aid our interaction with the environment and its advantages in dynamical and tactical situations have been recognised for quite some time (Bronkhorst, 1995; Calhoun, Valencia, & Furness, 1987; Ericson & McKinley, 1999; Haas, 1998a, b; Nelson, Bolia, Ericson, & McKinley, 1999; Nelson, Bolia, & Tripp 2001; Simpson, Brungart, Gilkey, Cowgill, Dallman, Green et al., 2004; Veltman, Oving, & Bronkhorst, 2004; Wenzel, 1993). Modern information systems require people to interpret and control an increasing amount of information, and in many situations it would be useful to take advantage of the benefits of a spatial audio display.

The theoretical work about 3D audio was well developed in the early 1990s (Wenzel 1992; Wenzel et al., 1993; Wightman & Kistler, 1990; Wenzel, Wightman & Foster, 1988) but there is still work to be done regarding use of applications in real-world scenarios. This led to intensified work at FOI to focus on how to bring the technology outside the laboratory. Initially, we used technology illustrators as effective means of acquiring feedback from end-users and to get a better understanding of the potential and the main issues with the current 3D audio technology (Carlander & Hasewinkel, 2003). At this point there was already evidence that listeners could use non-individualised HRFTs (Begault et al., 2001; Bronkhorst, 1995; Möller, Sörensen, Jensen, & Hammershöi, 1996; Wenzel et al., 1993). For our experimental planning this meant that we excluded complicated measurements to acquire individual HRTFs for the 3D audio presentation. We used generic filters, and for our initial studies cooperation was established with Dr. Fang Chen. We investigated the effect of noise and duration on 3D presentations and we were trying to achieve noise levels similar to that subjected to pilots. Realistic cockpit noise from the Swedish fighter jet JAS Gripen and the helmet from the same system was used (Carlander, 2002; Chen & Carlander, 2003a, b). For our studies we used the Lake Huron 20 audio rendering system, and this could at the time be considered as one of the state-of-the-art systems and used in many studies related to auditory research focussing on localisation (e.g. Chen 2003; Väljamäe, Tajadura-Jiménez, Larsson, Västfjäll, & Kleiner, 2008). The Lake Huron 20 consists of a large array of parallel DSPs (digital signal processor), DSP cards and I/O cards. The DSPs control the computations needed for the zero-latency convolution. The I/O cards receive audio input and output 3D sound. The Huron can process multiple channels of real-time audio resulting in that several users can simultaneously have an updated audio rendering (Burdea & Coiffet, 2003). The system is prepared to compensate for user head movements by a head-tracker. Allowing the user to move the head during sound presentation can reduce the differences between the simulated HRTFs and the listener's own, with the result that ambiguities are effectively reduced (Begault et al., 2001; Iwaya, Suzuki, & Kimura, 2003; Kato et al., 2003; Wightman & Kistler, 1999).

The processing software in the Huron also allows for room simulations like reverberation and occlusion effects. These characteristics allow the Huron to be used for research applications or high fidelity virtual reality (Burdea & Coiffet, 2003). The system comes with some software applications and we used the BinScape application. In short, the Binscape is a 3D positional-audio simulator that can simultaneously render sound sources and update these according to a single listener's head movements. The application is also capable of loading user definable head and room filters for engineering research applications (Bartlett, Cox, Butler, & Potas, 2003). Although this and similar systems could test many basic parameters in the labs a way to use the technology in more applied settings could be considered lacking at the time. The coming studies were therefore performed with the intention of finding a 3D audio system that offered advantages such as being small, lightweight, robust and without consuming to much electrical or computer power while fulfilling the requirement of accuracy.

Experiment 1

Being able to simulate effective horizontal 3D audio cues could offer additional information to the operator, such as to guide the operator towards a specific direction or event. Preferably, the 3D audio system should present sounds accurately without requiring large financial investments nor should it consume too much electrical or computer power. Furthermore operational applications often require a quite high degree of mobility which in turn demands smaller systems that are robust and lightweight. A solution incorporating these characteristics could be useful for a large number of operational systems (Haas, 1998a, b; Nelson et al., 1999; Simpson et al., 2004; Veltman, Oving & Bronkhorst 2004).

Improvements of commercial off the shelf (COTS) sound cards may more or less fulfil the above mentioned requirements. Some manufacturers put a lot of effort in improving both software and hardware resulting in products that approach advanced 3D audio platforms in performance. A good example was the sound cards based on the Sensaura algorithm. This algorithm was optimized for localisation accuracy, and simplified to ensure an efficient real time implementation in the sound card processor. The algorithm was applied to filters derived from dummy-head measurements in an anechoic chamber (Sibbald, n.d., a, b). In combination with DirectSound, which is a set of Windows interfaces from Microsoft for handling sound, a quite realistic 3D audio experience can be achieved. If a regular PC equipped with a 3D audio sound card could have a similar performance in the horizontal plane as the reference platform, it would be an important step forward for the operational use of 3D audio technology. This due to the ease of implementation and that the miniaturization of components would make such a system very compelling. This formed the first research question; "- Can a portable 3D audio system generate spatial audio that is accurate enough for operational use?"

Based on the above, an experiment was conducted to validate a portable 3D audio display with the aim to be able to test the system for field applications and to demonstrate the usefulness of 3D audio in operational use. We hypothesized that a COTS platform would result in localization errors useful for real world applications with a localisation error of about 20-30º. The COTS platform was compared to an advanced 3D audio rendering platform.

Participants. Four males and 6 females with the mean age of 24 years volunteered to participate. All reported normal hearing and were naïve about listening to virtual spatial sound.

Apparatus. Data were collected in a sound-attenuated laboratory room approximately 5 m^2 in size. A computer screen displaying a graphical response form and a computer mouse for participant input served as the response system. 3D sound stimuli were displayed through a pair of AKG k240 stereophonic headphones with a semi-open circumaural design and a frequency range of 15 to 25 kHz. The COTS 3D audio, was comprised by a PC (128MB RAM, 768MHz processor) and a Hercules Gamesurround Muse pocket soundcard that was connected to the USB port on the computer. The reference platform was the Lake Huron (as described earlier). Both technologies simulated a perceived distance of about 1 m to the sound source. To minimise the occurrences of front-back confusions (cf. Blauert, 2001; Wightman & Kistler, 1999), an Intersense IS-600 head-tracker registered head movements with an angular resolution of 0.5 degrees at a frequency of 180 Hz. The real-time data from the head-tracker system was used to compensate the sound presentation in relation to the head movements of the participant. Both audio technologies and the recording unit of the head-tracker were placed outside the sound laboratory to minimise noise levels. Participants were seated in a fixed-back chair and were allowed to adjust the height so they comfortably could see the computer screen and reach the computer mouse. The chair could move around its vertical axis so the participants were not physically restricted in turning and listening to the sounds (i.e. they were able to move both head and torso).

Design and Stimuli. A 2 × 2 × 12 within subjects design was used. The experimental conditions included:

- Two 3D audio technologies: Lake Huron and the COTS solution.
- Two sound types consisting of white noise and speech.
- Twelve horizontally distributed positions, represented by 0º, 30º, 60º, 90º, 120º, 150º, 180º, 210º, 240º, 270º, 300º, 330º ("clock positions").

Sounds were repeated three times at each position (randomized) and for each technology. This resulted in a total of 144 presentations (2 x 2 x 12 x 3 repetitions). The speech signal consisted of a spoken message: *"-Try to localise this sound"*. A stimulus duration of 4 s was chosen based on the results from Macpherson and Middlebrooks (2000), Chen and Carlander (2003b), and Chen (2003) who showed that sounds should be at least 1.5 s but no longer than 6 seconds to allow effective processing of the auditory stimulus for localisation. Also, as shown in Iwaya et al. (2003), front-back confusion can be effectively reduced if signal duration exceeds 2 seconds. The spoken sound was repeated to fill the complete four second simulation period. For all sounds, 69dB(A) was measured at the entrance of the ear canal. Stimuli presentation was organised into two blocks of trials, one per 3D audio technology, and block order was counterbalanced over participants. Sound type and positions were randomized within each block. The participant's task was to estimate the location of a sound source and indicate the perceived azimuth on the graphical response form shown on the computer screen. The response was indicated by adjusting a graphical vector overlaid on the top view illustration of the audio lab as shown in figure 8 below. The graphical response form was a top-view of the room with reference symbols at 12, 3, 6 and 9 o'clock (0, 90, 180 and 270 degrees). These reference symbols were also present in the room at the corresponding positions. This way, the physical world with the visual reference symbols, was mapped to the response form. Important to note is that sounds could also be simulated between the reference positions. The reason for using the symbols was thus to aid the user when mapping the perceived sound source to an actual position, translating the physical world reference to the graphical response form.

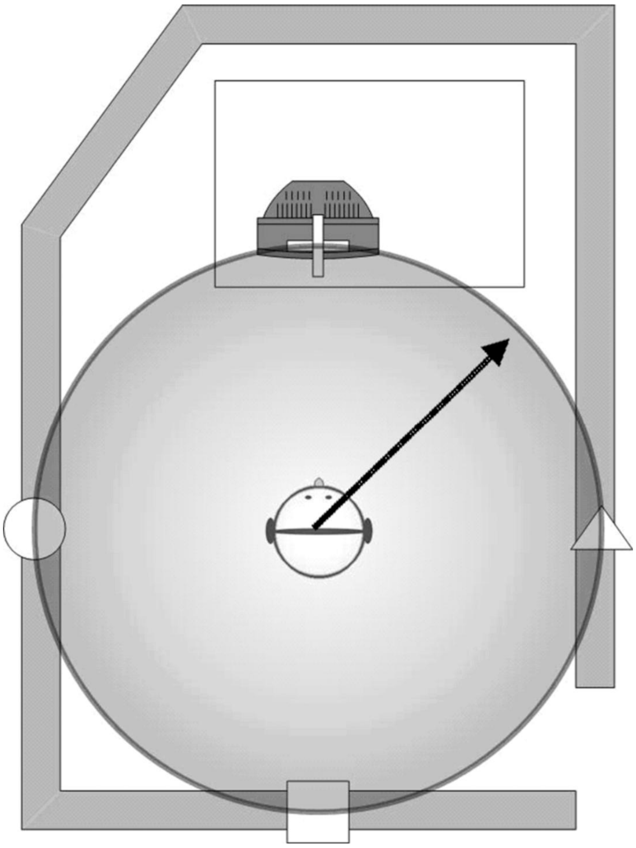


Figure 8. The graphical response form as used in Experiment 1. The black arrow (vector) was adjusted to point towards the perceived position of the sound and a computer mouse click registered the response and initiated the next presentation.

After adjusting the vector a click on the computer mouse stored the current orientation and caused the program to proceed to the next trial. Localisation accuracy was determined by calculating localisation error (LE) as the absolute deviation from correct sound position, thus the absolute value. For example, if a sound source was positioned at 150º and the estimate was 158º, the LE is 8º (158º-150º=8º). Thus, a lower LE indicates a more accurate estimate.

Procedure. Written and verbal instructions were given about the general setting of the experiment and the participant's task. Headphones were adjusted and a simple hearing test was then performed, followed by a five-minute training session. The participant was trained on the sound localisation task with the 3D audio system to be used in the following block. This was done to minimize eventual training effects of presentations within the block and to familiarize the participant with each sound type and display technology. The participant was instructed to move the head during sound localisation and to respond as accurately as possible. After each training response, feedback was displayed on the computer screen showing the correct sound position. No feedback was given during experiment proper. A head-tracker calibration process initiated each of the two blocks of trials. After the first block, the participant completed a short questionnaire during a five-minute break outside the sound laboratory (the results from the questionnaire are not presented here). When the second block was finished another short questionnaire was completed and a debriefing session concluded the experiment. Each experimental session lasted about 1.5 h.

Results. Repeated measures analysis of variance (ANOVA) was applied to the 48 means of LE made by each participant (2 technologies × 2 sound types × 12 sound positions). Because of a violated sphericity assumption, the Greenhouse-Geisser corrected p-value is reported. A total of 10 front-back confusions occurred during the experiment spread over five of the participants. These data were treated as outliers and not included in the calculation of the condition mean for these five participants (removal to fulfil assumptions for ANOVA). Front-back confusions in relation to the number of presentations for these participants were; 2.8%, 2.8%, 4.2%, 1.4%, 2.8%, respectively. All occurred for the COTS technology.

The ANOVA showed a significant main effect of sound position, $F(11, 29) = 14.39$, $p < .0001$, with no other significant effects. Mean LEs ranged from 2.4º to 13.0º over the sound positions.

Summary. Mean LEs were overall low for the two systems. The low error rates were expected from the reference technology but the overall performance of the COTS system was rather surprising. Perhaps the "reference" sound positions biased participants' responses resulting in lower LE's, as a result of the visual reference symbols at 0°, 90°, 180° and 270°. This might have led to that after a couple of trials, participants "concluded" that stimuli were presented either "on" the reference symbols or between them.

Important to note is that there were 10 front-back confusions for the COTS system and this can have serious implications for applications requiring precision in the front-back dimension.

Experiment 2

This experiment also attempt to answer research question 1 as described above. The aim remained; to be able to test the system for field applications and to demonstrate the usefulness of 3D audio in operational use.

Based on the first experiment we hypothesized that removing reference symbols and adding sound positions would introduce a larger difference between the two technologies. This would also rule out the possibility that the low error rates partly resulted from using relatively few sound positions in combination with visual reference symbols.

Thus, the reference symbols and the graphical response form were removed and no feedback was administered at any point during the experiment. However, the training session remained to familiarize the participant with the experimental environment and procedure. The reason for removing the feedback of the localisation error during training, was that we were interested to see whether the technology would be good enough for instant use, that is, without the need of participant training. Sound positions were doubled and offset 5 degrees, resulting in 24 equally spaced azimuths in intervals of 15º (i.e. 5º, 20º, 35º etc.). Participants were standing during the experiment to further facilitate movements.

Participants. The participants were 9 males and 2 females and had a mean age of 28 years, and no prior experience with 3D audio presentation technologies. All reported normal hearing.

Apparatus. Same as experiment 1.

Design and Stimuli. A 2 x 2 x 24 within subjects design was employed. The experimental conditions included:

- Two 3D audio technologies: Lake Huron and the COTS solution.
- Two sound types consisting of white noise and speech.
- Sounds presented at 24 azimuth positions in 15º intervals starting at +5º (relative to “straight ahead” heading, 0° azimuth).

The 3D audio systems, types of sounds, and sound duration were the same as in Experiment 1. Each sound position was presented twice per sound type and type of technology, resulting in a total of 192 presentations. The task was similar to that in Experiment 1, but the visual reference symbols in the room and the graphical response form on the screen were removed. To respond to the sounds participants turned their head, aligning the nose with the perceived sound position, thus pointing the nose towards the sound source (“head pointing”, cf. Blauert, 1995). A computer mouse click stored the azimuth orientation of the head and caused the computer program to proceed to the next trial. Feedback was neither given during the training session nor during experiment proper. An experimental session lasted about 1.5 h.

Procedure. Same as experiment 1.

Results. Repeated measures ANOVA was applied to the means of the LEs for each sound position and participant. The analysis included 96 means (2 technologies × 2 sound types × 24 sound positions) for each participant. We excluded one participant from the analysis because of an extremely deviating performance (a clear outlier performance). Because of a violated sphericity assumption, the Greenhouse-Geisser corrected *p*-value is reported. A total of two front-back confusions occurred during the experiment for two of the participants. These data were treated as outliers and not included in the calculation of the condition mean for these two participants (removal to fulfil assumptions for ANOVA). Front-back confusions in relation to the total number of presentations for these participants were 1.0%, and all occurred for the Lake technology.

The ANOVA showed a significant main effect of technology, $F(1, 9) = 8.86$, $p < .025$, with no other significant effects. The Lake technology showed a larger mean LE (SE), 10.5° (8.3), compared with the COTS technology, 7.4° (3.6). The main effect of technology is seen in figure 9 below.

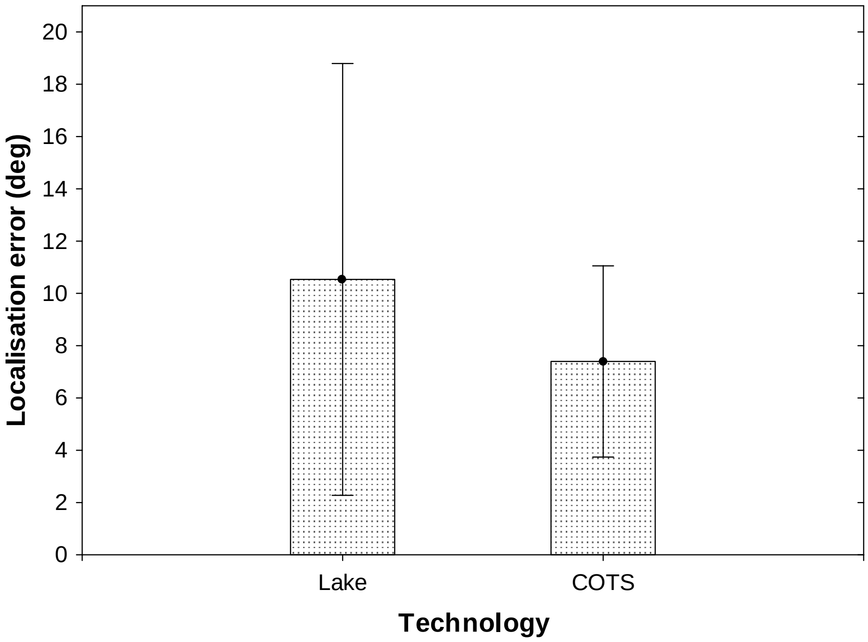


Figure 9. Main effect of technology, $F(1, 9) = 8.86$, $p < .025$, Means +SE.

Summary. The results show that the COTS technology performed better than the reference technology. Interestingly, the error rates were low for both technologies and stable over sound positions (front-back confusions removed). Perhaps the response methodology used in experiment 1 affected the accuracy when translating the audio experience to the visual response form. The removal of the chair in experiment 2 may also have facilitated the ability to efficiently use head movements for localisation. No front-back confusions were found for the COTS technology and only two for the presentations with the Lake system.

Experiment 3

The results of Experiments 1 and 2 showed some promise regarding the potential for the COTS system to be used outside the laboratory for various applications requiring localisation in the horizontal plane. To further investigate research question 1 and to completely remove the risk of having participants recognizing presentation positions a third experiment was conducted in which the sounds were randomized within each of four quadrants of the horizontal plane. The four quadrants were used to ensure that sounds were presented approximately evenly in the horizontal plane. Quadrant sectors were thus used to spread sound presentations and not used to optimize positions for finding eventual front-back confusions. Experiment 3 was also presented as a conference proceeding at the IEA 2006 conference (Carlander et al., 2006).

Participants. Seven males and 3 females with a mean age of twenty-six participated. All reported normal hearing and had not used 3D audio displays prior to the experiment.

Apparatus. Same as experiment 1.

Design and Stimuli. A 2 × 2 × 4 × 12 within subject design was used. The experimental conditions included:

- Two 3D audio technologies: Lake Huron and the COTS solution.
- Two sound types consisting of white noise and speech.
- Four azimuth sectors divided over front-right (0-90º), back-right (90-180º), back-left (180-270º), and front-left (270-360º).
- 12 trials with randomly positioned sounds were presented for each sector (a resolution of 1°).

Procedure. The procedure was the same as in Experiment 2.

Results. The LEs for each sound presentation sector were calculated for each participant and used in a 2 × 2 × 4 × 12 repeated measures ANOVA. Only one front-back confusion occurred during the experiment. This data was treated as an outlier and not included in the calculation of the condition mean for that participant (removal to fulfil assumptions for ANOVA). Front-back confusion in relation to the number of presentations for this participant was 1.0% and it occurred for the Lake technology.

The results showed a significant main effect of technology, $F(1, 9) = 22.71$, $p < .01$, trial, $F(11, 99) = 2.00$, $p < .05$, and a significant interaction effect of technology by sector, $F(3, 27) = 6.25$, $p < .01$ with no other significant effects. A Tukey HSD test revealed larger mean LEs (SE) with the Lake technology in all four sectors; Lake: 8.5º (2.1) and COTS: 6.3º (3.2) in sector 0-90º ($p < .01$), Lake: 9.3º (5.0) and COTS: 7.5º (3.6) in sector 90-180º ($p < .001$), Lake: 9.8º (5.1) and COTS: 7.0º (3.9) in sector 180-270º ($p < .001$), and Lake: 9.9º (3.3) and COTS: 5.0º (1.8) in sector 270-360º ($p < .001$). (The significant main effect of trial showed LEs over trials between 7.0º and 8.6º.)

Figure 10 illustrates the interaction effect of presentation technology by presentation sector on LE. Thus, the use of the COTS technology resulted in lower LEs compared to the Lake technology for the estimations of sound positions in all of the four sectors, and the LE difference between technologies varies over sectors.

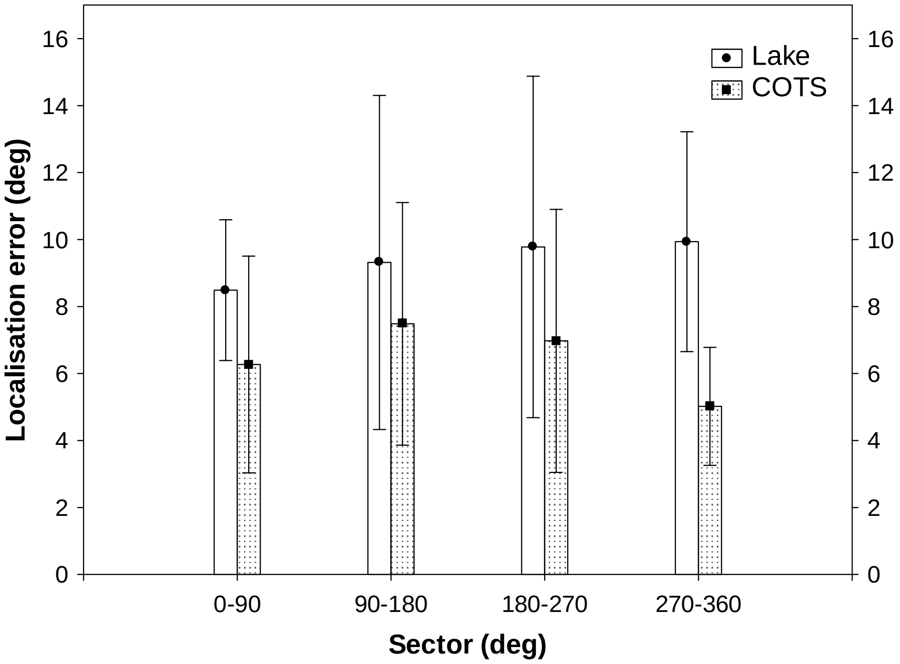


Figure 10. Interaction effect of technology by sector, $F(3, 27) = 6.25$, $p < .01$ Means +SE.

Summary. Taken together, the results of the three performed experiments provide strong support for the notion that a COTS 3D audio platform can perform as well as or better than the reference 3D audio platform - the Lake Huron - regarding sound localisation in the horizontal plane. In all three experiments non-individualised filters based on an averaged listener were used. Since error rates were low for both systems it is not likely that performance would have been much improved if the filters had been tailored to each listener. In general, the LE was low and inter-subject variability small. Interesting to note is also that in experiments 2 and 3, no feedback was administered at any point of the experiments. In spite of this, participants still performed well, indicating that the technology can be useful without substantial training. The results tie to the overall goal of contributing to an intuitive interface: Being able to present spatial audio information without training and with high accuracy, bear a potential for enhancing information requiring little effort from the operator.

The low and stable error rates of both systems can be explained by an effective simulation of the interaural difference cues, and it has previously been shown that ITD is a robust and relatively simple cue to reproduce (Blauert, 2001). On a relating note is that *"subjects found the middle of the region covered by the filter with zero ITD"* (Bronkhorst, 1995, p. 2551) meaning that participants were able to perceive when sounds were straight ahead since sounds presented at this position generates equal sound energy in both ears (the zero ITD filters).

Although not investigated in these experiments, it is important to mention that sounds in the vertical dimension (i.e. above or below the horizontal plane) most probably would require individually tailored filters. The reason for this is that elevation cues are more reliant on outer ear characteristics and thus require a better handling of high frequency content. One can speculate that the tested COTS system may not support the more complex cues for vertical sound localisation due to higher demands on computer processing capabilities and the involvement of other cues than ITD. This might also be true for situations requiring a higher degree of realism or externalisation of the presentations, or when being required to handle simultaneous sound sources. The advanced platform (Lake Huron) can handle more complex room and reverb modulation, and in a way this was indicated by some participants claiming that the Lake Huron gave a much more realistic experience of the auditory space. However, increasing the realism by adding a reverb-like character to the sounds can cause the sounds to be perceived as blurred and thus harder to localise (cf. Shinn-Cunningham, 2001). In the present study, the room acoustics of the Lake Huron might thus have had a negative effect on the perceived location of sounds.

Improved generic 3D audio display systems are important since this could yield a more efficient implementation of the 3D audio technology. The low training requirements in combination with COTS can potentially make implementation more cost-effective, widespread and easier to realise. The validated solution could solve common field issues such as size and weight constraints while still being robust. Based on the above experiments, challenges like intelligibility of radio calls, target and threat indications and navigation could now be further tested with the COTS technology.

3D audio technology and applications

Advantages like omnidirectionality, ability to call for attention and intuitiveness are supported by our natural sound localisation ability. For instance, auditory events can be positioned to enhance meaning and understanding of data or simply to improve the intelligibility of the auditory information. A 3D audio display can increase the understanding of what the information represents (auditory icons) and where the information is (position). Directional information presented by 3D audio may include a wide range of applications ranging from military tasks with target detection to the visually impaired as a support for navigation (i.e. Andersson, 2004, 2005; Carlander & Eriksson, 2006; Walker & Lindsay, 2004, 2005a). To be able to offload the visual channel could have an impact on operator efficiency especially when visual sensory information is degraded or visual attention is needed somewhere else (e.g. the situation does not allow for using a traditional map or a visual display). 3D audio has shown to effectively complement visual directional information (e.g. Haas, 1998b; Nelson et al., 1999; Veltman et al., 2004). 3D audio can support both indoors (Sandberg, Håkansson, Elmqvist, Tsigas, & Chen, 2006) and outdoor navigation (Eriksson et al., 2008; Walker & Lindsay, 2004, 2005a). However, Eriksson et al. (2008) presented results from questionnaires that revealed that the soldiers experienced the 3D audio display to lower the attention to the terrain and that 3D audio resulted in a higher mental workload compared to a tactile display. Figure 11 below shows a soldier during this experiment.

Figure 11. 3D audio system as used by Eriksson et al. (2008, p. 1283). A dismounted soldier is equipped with headphones and a digital compass (the white box on top of the head). Picture is used with permission from SAGE (http://pro.sagepub.com/content/52/18/1282).

Given that the 3D audio technology can offer direct feedback of positions with a continuous update of a possible multitude of sound sources, it has many benefits compared to conventional stereo and mono displays (Begault & Wenzel 1993; Haas, 1998a; Nelson, Bolia, Ericson, & McKinely, 1998; Vause & Grantham, 1998; Veltman & Oving 2004). With 3D sound we can simultaneously process sound direction, content, and distinguish a specific sound amongst several others (Doll, Hanna, & Russotti, 1992; Ericson & McKinley, 1997; Nelson et al., 1998). These benefits have also been shown in studies performed in noisy environments (Doll et al., 1992; Nelson et al., 1998; Vause & Grantham, 1998) and for low-sample-rate systems below 4 kHz, which means that sounds can be spatialised in systems with a limited bandwidth (Begault, 1999b).

In an experiment with a simulated cockpit Haas (1998b) showed how a visually supported interface for helicopter pilots could be significantly improved by complementing the visual warning presentations with 3D audio. The complementary 3D audio warnings reduced reaction times to warnings compared to the traditional visual interface. Haas thus concluded that 3D audio could potentially enhance helicopter cockpit safety. In a similar study Veltman, Oving, and Bronkhorst (2004) investigated benefits of 3D sound for cockpit task performance and workload. Two conditions, one with 3D sound supporting a HDD and one without 3D sound support were investigated. Performance was improved when 3D sound audio was present, and eye movements relating to the HDD were reduced with more than 50%. Overall mental effort was similar for all conditions but a small reduction in subjective workload could be seen for some of the 3D sound presentations. Although workload was not significantly lowered, the overall performance increased as a result of the redistribution of mental capacity.

The angular separation between the sound sources, as accomplished by the spatial audio, can lower thresholds at which signals can be detected and thus, the identification of multiple sound sources can be made easier (Begault, 1999b; Doll et al., 1992; Nelson et al., 1998). This characteristic makes spatial audio a potential candidate for spatialised radio communication. Radio communication has been a relevant application for the 3D audio technology since information already is auditory and the spatialisation enhances intelligibility and adds talker position. To be able to identify and attend single call signs among several talkers can improve speed and accuracy in decisions since the operator can discern more information and recognize what is said and by whom. Increasing intelligibility also decreases the risk of misinterpretations (Ericson, Brungart, & Simpson, 2004; Haas, 1998a; Nelson et al., 1998). Albery (2006) investigated spatialised radio communication (within and between helicopters) and he found that 3D audio reduced the mental load and increased situational awareness for pilots.

Since humans only use two audio input channels (our two ears), the above described advantages can be relatively easily applied to two-channel radio communication systems. It has been shown that in a multi-talker display with up to three talkers even a small separation (~22°) between sources in front of the listener has a large effect on intelligibility and that more than three simultaneous talkers may overload the auditory system and prevent it from capturing a desired message from a particular direction (Ericson & McKinley, 1997).

Experiment 4

In a project at FOI we initiated discussions with fire and rescue command operators and some issues relating to the intelligibility in radio communication systems were recognized. A command operator may abruptly process up to four sets of brief auditory stimuli from four separate radio channels. The operator needs to have good vigilance since missions can extend over longer periods of time where radio calls often occur in bursts of several calls. The system they use today is based on headphones and speakers as options for listening to radio channels. Separation of sound sources are made by stereo panning in the headphones, and the speakers correspond to one radio channel (and only one) respectively. Both headphones and speakers offer limited intelligibility of radio calls, partly because of their limited spatial separation.

Since intelligibility is improved by angular separation of sound sources (Begault, 1999b; Doll et al., 1992; Nelson et al., 1998), we used the COTS 3D audio system from the 3D audio experiments (experiments 1 to 3) to separate sound sources. We hypothesized that the intelligibility of simultaneous call-signs presented in background noise of added voice sources could be improved by 3D audio. Based on this the second research question was formed; - Can 3D audio enhance the intelligibility of call signs? This experiment was performed with real operators so we also attempt to answer research question five: - Can 3D audio (and tactile) displays be used effectively and efficiently without extensive operator training?

Experiment 4 was also presented as a conference proceeding at the 11th Meeting of the International Conference on Auditory Display (ICAD) (Carlander et al., 2005) and was published as an FOI report (Carlander & Kindström, 2004)

Participants. Ten male command operators from a Swedish fire and rescue department participated. They were all inexperienced in using 3D audio displays, but familiar with stereo displays.

Apparatus. A PC with monitor and a soundcard (Hercules Gamesurround MUSE Pocket) was used. All auditory stimuli were recorded on the PC with a Shure M58 microphone and a microphone preamp, and the speech signals were high-pass filtered at 100 Hz and low-pass filtered at 8 kHz. The filtering of the sound sources was based on Nelson et al. (1998). Speech signals were presented in AKG k240 studio headphones with a frequency range of 15 Hz to 25kHz.

Design and Stimuli. The experiment used a 2 × 4 × 3 factorial within subject design, and it included:

- Two auditory display technologies, 3D-audio (novel) and stereo sound (traditional)
- Four call sign conditions, one to four simultaneous call signs
- Three levels of background voices; two, three or four simultaneously presented

Each of the four levels of call signs consisted of presenting a single call sign, two, three or four simultaneous call signs with different callers. The background voices were reading different texts with at least one background voice per ear. 3D audio call signs were separated and positioned mainly based on Ericson and McKinley (1997) that reported an angular separation greater than or equal to 45° provided the greatest levels of intelligibility. Nelson et al. (1998) also concluded that location per se does not determine the efficiency of the spatialisation effect. The audio displays are thus not constrained by a specific spatial area. Stereo call signs were positioned 100% right, 25% right, 100% left and 25% left. The duration of each call sign was 2.5 s, and the timing (start and end) of the call signs were slightly off-set but completed within 3 s. The interval between the call sign presentations was randomized between 4 and 16 s, and each condition was repeated three times, resulting in 72 presentations for each participant.

The primary task was to identify one (single) to four (set) simultaneous call signs among two to four background voices. The call signs were spoken command calls “102 over, 102 over” and, when identified, the subject used the computer mouse to indicate who or whom of the talkers it was that called. The spoken command was always the same and using four different talkers was to make the design more realistic as the command operator normally handle several callers that use the same spoken command. A correctly identified single call sign was defined as the identification of a single speaker. A correct identification of a complete set of call signs was defined by the identification of all speakers in a set. A secondary visual and manual response task induced an overall high mental workload. The workload was adjusted by parameter settings of the secondary task such as speed of symbols, levellers and cross aim movements. The presentation order of stereo and 3D-audio was counterbalanced over participants, and experimental conditions were randomized within each session. The experimental setting is shown in figure 12 below.

Figure 12. A fire and rescue command operator during the experiment on 3D sound and radio communication intelligibility (Carlander, Kindström, & Eriksson, 2005).

Procedure. After a brief introduction two training sessions, one for stereo and one for 3D-audio were completed with an overall low workload. Training consisted of one presentation per call sign set, totalling four presentations per auditory display. When completed, the experiment with a higher workload was then performed. Presentation order of stereo and 3D-audio was counterbalanced over participants, and call sign conditions, and levels of background voices were randomized within each session. 36 auditory stimuli for both stereo and 3D-audio presentation resulted in 72 presentations in total. Each session lasted about 1 h for each participant.

Results. Repeated ANOVA were applied to each of the means of correctly identified single call signs and correctly identified complete sets of call signs. For each participant, each mean was calculated from three trials of each call sign or set of call signs in each condition. Each analysis thus included 24 means (2 × 4 × 3 = 24). All ANOVA p-values are with the Greenhouse-Geisser correction values.

Identified single call signs.
The ANOVA of correctly identified single call signs showed significant main effects of technology, $F(1, 9) = 6.51$, $p < .05$, and background voices, $F(2, 18) = 5.84$, $p < .025$, with no other significant effects. The stereo technology with mean proportion correct (M) = 0.57 and standard error of (SE) = 0.04 generated less accuracy compared to 3D audio with M = 0.63 (0.03). A Tukey HSD test revealed a higher accuracy with two background voices, M = 0.64 (0.03), compared to four, M = 0.55 (0.04) ($p < .01$), with no other significant differences.

Identified complete sets of call signs.
The ANOVA of correctly identified sets of call signs revealed significant main effects of technology, $F(1, 9) = 5.89$, $p < .05$, background voices, $F(2, 18) = 5.94$, $p < .05$, and set size of call signs, $F(3, 27) = 74.98$, $p < .0001$, with no other significant effects. Stereo, M = 0.27 (0.04), generated less accuracy compared to 3D-audio, M = 0.32 (0.04). A Tukey HSD test revealed that two background voices, M = 0.35, (0.04), resulted in higher

accuracy than four, M = 0.24, (0.04) (*p*< .0001). Set sizes of both one and two call signs, M = 0.65, (0.05), and M = 0.38 (0.06), showed higher accuracy than sets of three and four, respectively, M = 0.09, (0.04), M = 0.06, (0.03) (*p*< .05 for all comparisons).

Summary. Although with only a small improvement the results imply that command operator ability to discern call signs is improved by 3D-audio, compared to stereo. 3D-audio offers a slightly better, but still significant, intelligibility of call signs. This is most probably because of the increased spatial separation and the results are in line with previous studies that show how 3D-audio is more efficient for presenting simultaneous sound sources (Baldis, 2001; Begault, 1999b; Doll et al., 1992; Haas, 1998a; Nelson et al., 1998). Although a small improvement, increased intelligibility of radio communication as a result of spatial audio bears some promise. Further investigations will include an advanced 3D-audio platform that better handles simultaneous sound sources and contrast this towards a simple stereo solution. Furthermore, the secondary workload task needs to be more relevant or removed.

Experiment 5

Based on the experiment with the fire and rescue command operators (Carlander et al., 2005) we investigated speech intelligibility for multiple talkers in one stereo and two different 3D audio display systems. More specifically we wanted to investigate whether intelligibility would differ using “low level” spatialisation (stereo) compared to a higher degree of spatialisation with more advanced 3D audio. We thus hypothesized that there would be a significant difference in intelligibility between audio spatialisation technologies. In the previous study of fire and rescue command operators (experiment 4) the advanced reference platform Lake Huron was not used. In experiment 5 it was included to investigate if a dedicated 3D audio platform would increase the performance of identified call signs. The secondary task used in experiment 4 was removed since it was considered too complex. The experiment was performed to add further answers to the second research question.

Experiment 5 was also presented as a conference proceeding at the Human Factors and Ergonomics Society 50th Annual Meeting (Kindström, Carlander, & Eriksson 2006).

Participants. Twelve participants, eight females and four males, aged 20 to 38 (average age was 25) volunteered to participate in this experiment.

Apparatus. The experiment was set up in an anechoic chamber where a PC with a Hercules Gamesurround MUSE Pocket soundcard was used to display stereo and 3D audio COTS conditions. A LAKE Huron 20, running the Binscape application was used to present advanced spatialised 3D audio. A pair of AKG k240 studio headphones was used for displaying all sounds.

Design and Stimuli. The experiment utilized a 3 × 4 factorial within subject design. The experimental conditions included:

- Three auditory display conditions: one stereo and two 3D audio displays.
- Four call sign-sets of one to four simultaneous call signs.

The call signs consisted of a spoken command call “102 over, 102 over”. The call signs included a single call sign and two, three and four simultaneous call signs. Presentations were made in the presence of background noise that was composed by two voices, one in each ear reading different texts. As for experiment 4 the duration of each call sign was 2.5 s, and slightly off-set but completed within 3 s. Call sign positions were also kept as in experiment 4. Interval between the call sign presentations was randomized between 4 and 16 s. Six trials of each condition resulted in 72 presentations per participant.

The task was to identify one (single) to four simultaneous call signs among two background voices by indicating the call(s) in a response form. An identified single call sign was defined as the identification of a single speaker in a set. For the identification of a complete set of call signs, all speakers in the set had to be identified.

Each of the three auditory display conditions represented one block and block order was balanced over participants.

Procedure. A brief introduction was followed by written and verbal instructions. Each block then followed and was initiated with a training session for that specific block. Training consisted of one presentation per set of call sign, totalling four presentations per auditory display. The 24 experimental presentations of each block resulted in 72 presentations in total.

Results. Repeated measures ANOVAs were applied to the means of identified proportion of single call signs and identified complete sets of call signs. Each analysis included 12 means (3 audio displays × 4 call sign conditions), with each mean calculated from six trials of each condition.

Proportion of identified single call signs.
The ANOVA showed significant main effects of display technology, $F(2, 22) = 4.55$, $p < .025$, and call sign condition $F(3, 33) = 165.91$, $p < .001$, and a significant interaction effect of technology by call sign, $F(6, 66) = 2.70$, $p < .025$. A Tukey HSD test revealed a greater intelligibility, in the ‘single call sign’ condition, for the LAKE technology ($p < .025$), M = 0.78 (0.05), compared to stereo M = 0.67 (0.04). Also, in the ‘two call sign’ condition, the COTS 3D audio technology showed a greater intelligibility ($p < .005$), M = 0.41 (0.03), compared to stereo M = 0.29 (0.03).

Identified complete sets of call signs.
The ANOVA showed significant main effects of display technology, $F(2, 22) = 4.76$, $p < .025$ and sets of call signs $F(3, 33) = 14.04$, $p < .001$, and an interaction effect of technology by call sign, $F(6, 66) = 2.27$, $p < .05$. A Tukey HSD test showed a greater intelligibility for the COTS 3D audio technology in the ‘two call sign’ condition, ($p < .001$), M = 0.67 (0.10), compared to stereo, M = 0.32 (0.07), with no other significant differences. The interaction effect of technology by call sign is illustrated in Figure 13 below.

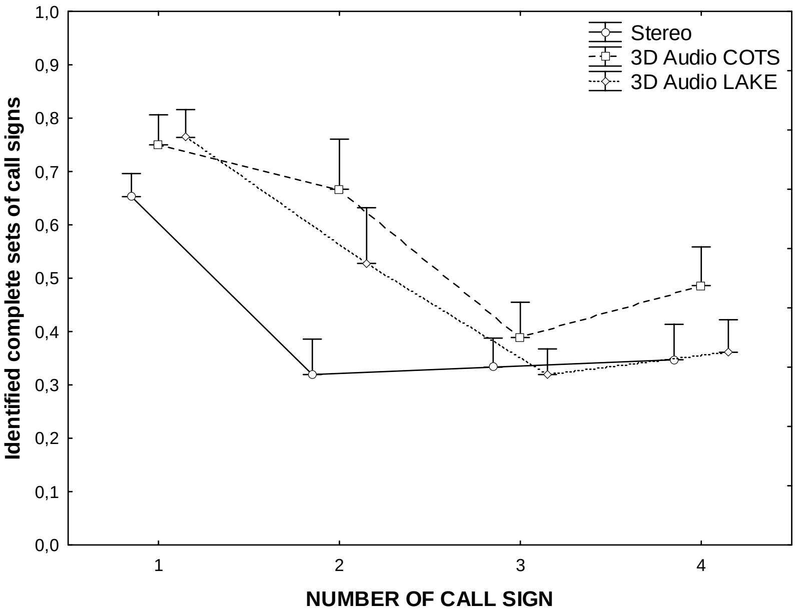


Figure 13. Interaction effect of technology by call sign for complete sets of call signs (F(6, 66) = 2.27, p< .05).

Summary and discussion. As for the study with the fire and rescue command operators the results imply that the ability to discern call signs is improved by using 3D audio as compared to stereo.

The reference platform (Lake Huron20) generated a larger proportion of identified call signs in the 'single call sign' condition compared to stereo, and the COTS platform generated a larger proportion of complete sets of call signs, in the 'two call sign' condition, compared to stereo.

Important to note is that the one and two call sign conditions are the most relevant since the probability of simultaneous call signs decrease with an increasing number of call signs. Most likely it is the increased spatial separation of call signs that is the reason for the greater intelligibility of the 3D audio displays and this is also in line with previous research (Begault, 1999b; Drullman et al., 2000; Haas, 1998a; Nelson et al., 1998). The overall small differences between the display technologies might be explained by the fact that they all, to some extent, spatialise the radio calls. The results indicate that it is possible to increase the intelligibility of stereo radio systems by using 3D audio spatialisation. Again the COTS display showed good performance and thus bears a potential to create a flexible 3D audio solution for radio communication. These results further add to the long-term objective of contributing to the development of an intuitive interface as we show how the 3D audio technology improves intelligibility. Furthermore, having "talkers" separated in space thrives on our natural sound localisation.

Known limitations and possible improvements of 3D audio

One major drawback with the 3D audio presentation is the front-back confusion as described earlier. This limitation does not only depend on the technology itself, but also on how we perceive sounds. Already early in the 20th century Lord Rayleigh meant that the judgment of a sound source in front or in the back should be *"emphatically distrusted"* (Rayleigh, 1907). The 4-10 degrees of auditory resolution reported for natural localisation (Begault, 1999a) will be worse with virtual spatial sound. Even a careful consideration of sound parameters will not improve the relatively low resolution of the auditory system much. What can be done is that the auditory signal in some way is adapted to an optimal frequency in relation to the background noise and headphone characteristics. Simple adjustments such as making the sound more broadband compared to a narrowband sound also make it easier to localise (Middlebrooks & Green, 1991). Although 3D audio adds spatial information to the audio presentation there are ambiguities relating to our sound localisation ability that result in misinterpretation of positions or directions. The confusions as described earlier are more frequent for simulated spatial sound, and will of course have serious implications for the design of auditory displays.

Further considerations relate to the fact that when using virtual spatial sound we block our ears. Both circumarual and in-ear headphones will most likely have a negative impact on the workload for operators that also need to receive auditory information from the surrounding environment. In response to this some promise has been shown regarding the use of bone conduction for 3D audio (e.g. Walker & Lindsay 2005b; Väljamäe et al., 2008; Walker, Stanley, Iyer, Simpson, & Brungart, 2005). Bone conduction leads the auditory signal through the skull directly to the inner ear thus allowing the entrance of the ear, and ear canal, to remain unblocked. This could greatly enhance the use of auditory displays since it would be possible to do an “auditory overlay” onto the surrounding environment. This way, a dismounted soldier could receive spatial audio information without blocking the normal hearing and awareness of auditory events in the surroundings. Still of great importance is to consider an improvement of externalisation of the auditory presentation. Externalisation means that sound sources appear as originating “outside” the head, thus not emanating from the headphones. This is an important factor for intelligibility in headphone systems since it enhances the simulated effect of normal everyday listening, helping the listener to identify and focus on specific talkers (Vause & Grantham, 1998).

The choice of presentation plane is important, since humans are rather poor at judging the absolute distance to and elevation of sound sources. Presentations of sound sources that simulate cues from the pinnae (elevation cues) are individual and will alter the auditory perception if the cues are not based on the individual listening. The main reason is that the pinnae, crucial for elevation cues, is more individually specific than other physical attributes like for example the relative distance between the ears (Wenzel, 1992), and because spectral cues are very fragile and thus hard to reproduce (most energy at high frequencies, above 4000Hz). Horizontal sounds, as described earlier, mostly depend on robust time and intensity differences, and these are more reliable for directional information (Middlebrooks & Green, 1991).

The most common problem is the front-back confusion, but this can at least partly, be overcome by allowing head movements in the spatial audio system since head movements facilitate our natural sound localisation. For moving sound sources the minimum audible movement angle (MAMA) is smallest for sources directly in front of the listener and increases with azimuth. This is due to the fact that the ITD changes most in front of the listener (Ericson & McKinley, 1997). Chen (2003) used a head-tracker to allow for head movements during sound presentation and reported that a presentation time of 4 seconds was necessary for localizing 3D sounds. Increasing the duration to 6 seconds seemed to introduce uncertainties regarding the user’s sound localisation. Similar to Chen (2003), Iwaya, Suzuki and Kimura (2003) suggested that ambiguities are effectively reduced by head movements when the sound duration is over 2 seconds. Apart from reducing front-back confusion occurrences, Kato, Uematsu, Kashino and Hirahara (2003) showed that general perceptual differences induced by using someone else's HRTFs were overcome by head movements.

That we quickly can adapt to an auditory stimuli shows promise for using the technology without extensive training. Training seems to primarily reduce the number of sounds perceived "inside the head" and users may feel more confident with their responses (Trapenskas & Johansson, 1999). Ohuchi, Iwaya, Suzuki and Munekata (2005) also showed some positive effects of training the listeners in a 3D audio display, but in this experiment the effect could not be seen until the 7th day of training.

In summary, a real system utilizing spatial sound should be such that it tolerates auditory confusions or it should be used with a head-tracking system. However, there might be applications requiring precise localisation, and for these applications spatial auditory presentation is not optimal. In addition, although a head-tracker improves both the user experience and localisation (Begault et al., 2001; Wightman & Kistler, 1999), there might be situations where a head-tracker would not work or simply cannot be used because of environmental constraints.

Another way of resolving the confusions is to complement the 3D sound presentation with simultaneous presentation to other sensory modalities. In a study performed by Kayser et al. (2005) it was shown that tactile information in combination with auditory information was processed similar to the combination of visual and auditory information. The senses of touch and audition accompany each other and they are both omnidirectional. The similarities strengthen the possibilities of complementation, and the next chapter further explores the possibilities with tactile displays and how these can contribute to an intuitive interface.

Key-points Chapter 2

- **Sound localisation, the ability to determine the direction to sounds is based on the combination of signals from both ears (interaural) with the addition of the cumulative effects of the pinnae, head, shoulders, and torso (Blauert, 2001).**
- **The “Duplex theory”, comprised by Interaural Time Differences (ITD) and Interaural Intensity Differences (IID) was reported already in the late 1800’s (Rayleigh, 1907).**
- **Human sound localisation with real sound sources has an accuracy of about 4-10 degrees (Begault, 1999a).**
- **”The cocktail party problem” is a phenomenon concerning our ability to focus listening to a single spoken sound source in conversations and background noises (Arons, 1992).**
- **Head-related transfer functions (HRTFs), constitute the filters used for virtual spatial hearing.**
- **3D audio as an information display can aid our interaction with the environment and has advantages in dynamical and tactical situations.**
- **The angular separation acquired by 3D audio lowers thresholds at which signals can be detected and allows simultaneous streams of auditory information to be easier managed (Begault, 1999b; Doll et al., 1992; Nelson et al., 1998).**
- **The most common problem with spatial sound is the front-back confusions. These occur when a sound source is on the central axis of the head (directly above, in front of or behind the listener).**
- **Head movements reduce front-back confusion occurrences and the perceptual differences induced by using someone else’s HRTFs (Kato et al., 2003).**

Contribution to 3D audio research

- **Further evidence for using non-individualised HRTFs for the simulation of high accuracy virtual sound sources (Carlander & Eriksson, 2006, Carlander, 2002; Chen & Carlander, 2003a, b).**
- **Contribution to the understanding of sound duration and its effect in 3D audio systems. At least 1.5 s but no longer than 6 s to allow effective processing of the auditory stimulus (Carlander, 2002; Chen & Carlander, 2003a, b).**
- **Advances in bringing the 3D audio technology to the field by validating a COTS component (Carlander et al., 2006).**
- **Introducing 3D audio for fire and rescue command operators, showing its advantage over stereo and mono presentation (Carlander et al., 2005).**
- **Showing how intelligibility is improved by using 3D audio as compared to stereo (Carlander et al., 2005; Kindström et al., 2006).**
- **Contribution to navigation support for the future soldier (Eriksson et al., 2008).**

Chapter 3. The tactile sense and tactile displays

The tactile sense has similar benefits as the auditory sense and could be used for displaying information to. The main objective for introducing “yet another display” is to improve information presentation and to relieve the information load on operators in various settings. When the visual sense is heavily overloaded and the auditory sense misused there is a risk that important information is unattended. There might also be specific conditions requiring certain senses, like in darkness or glare where you cannot use your eyes effectively but your ears and sense of touch work perfectly (Dobbins & Samways, 2002; van Erp, van Veen, Jansen, & Dobbins, 2005). As pointed out by Ernst and Bülthoff (2004) there is no information-processing system with enough capability to “perceive and act” perfectly under all conditions.

Still the sense of touch is relatively underutilized as an information channel and can be used for a “tactile display”. The tactile display mimics our everyday tactile communication and can be compared to “a tap on the shoulder”. The “tap on the shoulder” is simulated by causing a stimulation of the skin where the perception reveals a relationship to the outside world, much like that of audio or visual stimulation. The tactile display can add to the intuitive interface and van Erp (2001) refers to a “field of touch” (comparing this to the “field of vision”) where the tactile stimulation can be used for mapping external directions onto the torso. Results show a uniform acuity across the torso of 2-3 cm (van Erp, 2005a), and a subjective estimate of about 10° in the horizontal plane is feasible to expect. This resolution is further improved to an approximate acuity of 1 cm when tactile stimuli are located on the spine or the navel. However, when displaying external locations, that is, directions with the body mid-axis as a reference point, the perception of the tactile signal will be offset. This is due to that we mentally do not refer the body mid-axis as our “mid-point”, but rather it seems to be two mid-points, one in each body-half. On the other hand, using the body mid-axis as a reference seems to be an acceptable design for most applications (van Erp, 2005b).

One of the most effective ways of displaying tactile information is to present spatial information on the torso because the torso is less affected by body map distortions or so-called anchor points[3] (Stolle, Hölzl, Kleinböhl, Mrsic, & Tan, 2004; van Erp, 2005b). Stimulations onto the torso can map the “outside world” directly onto the body coordinates, for instance a vibration element on the right side will indicate an event at the right etc. (e.g. McGrath, Estrada, Braithwaite, Raj, & Rupert, 2004; Rupert, Guedry, & Reshke, 1994; van Erp et al., 2005; van Erp et al., 2003).

The skin with its receptors responds to several physical quantities. Pressure, temperature, pain, and vibration are all part of the cutaneous sense[4], and the skin delivers important information about the surrounding environment. For many manipulation and exploration tasks there is a dynamic interchange with our somatosenses (Carlson, 2001; van Erp,

---

[3] Anchor points are locations on the body improving the accuracy of a tactile percept. These anatomical points serve as a reference represented by the wrist, elbow, spine, or navel, (Cholewiak & Collins, 2003; Jones & Sarter, 2008; van Erp, 2002; 2005a).

[4] Only a very brief description of the tactile sense will be covered here, a good overview can be found in van Erp (2007).

2007). The skin is a highly detailed and sensitive sensory organ with a temporal sensitivity close to that of the auditory sense. This, in combination with that the skin receptors are being spread over a large area, makes it suitable to receive information relating to orientation and navigation. Also, the spatial coordinates resulting from a stimulus on the skin is well represented in the nervous system (van Erp, 2002, 2005b). Compared to vision, the sense of touch is a more basic sense that requires a closer interaction with the surrounding environment as stimulus needs to be in direct contact with the skin of the perceiver. The close perceptual distance where signals are mapped directly onto the body coordinates result in a higher attention and thus a faster perception and processing (Gilson et al., 2007). The tactile displays used in the experiments presented later in this thesis are mounted on the torso. This means a stimulation of receptors in the hairy skin, and in the subcutaneous tissue mechanoreceptors (receptors responding to mechanical stimulation) respond to for example the vibration of different frequencies (Carlson, 2001).

In the hairy skin, there are five main types of mechanoreceptors that are known. Of these two are slowly adapting and three are rapidly adapting. The Pacinian corpuscles, can signal vibrations from 30-800 Hz, responding extremely fast by transmitting over fast nerve fibres at 1000 cycles per second. The Meissner's corpuscles, another type of mechanoreceptor, are especially receptive for low frequency vibration, up to 80 Hz (Guyton, 1991). Knowledge of mechanoreceptor characteristics are important for the development of effective tactile displays since the frequency, amplitude, duration of vibration and the tactor skin stimulation area affect the activation and responses of numerous mechanoreceptors in the skin. Also, the sensitivity to tactile stimulus varies over the body since mechanoreceptors are not evenly distributed over the skin.

The tactile display concept

Vibrotactile displays can be used to present directional or coded information to an operator by stimulating areas of the body with vibrotactile signals. The stimulation is created by having small vibrating elements close to the skin. The vibrating elements are often referred to as tactors. The tactors used in the displays are most often small and lightweight electromechanical vibrators mounted in a housing attached to the fabric worn by the operator. The notion of the tactile display concept is not new, van Erp and Self (2008) report historical examples of tactile "displays", or at least tactile writing, used for military applications in the late 1700s. In modern time there were early attempts to use tactile displays in aviation, and in the 1950s there were several attempts to place tactors on the pilot ranging from single tactors to matrixes, on hands, arms and in the helmet of the pilot (van Erp & Self, 2008). From the late 1950s and 35 years forward, the progress was slow in the area of utilizing the sense of touch for information presentation. Discussions concerning multisensory displays started again during the 80s, but at the time these displays were not practically possible. It took until the mid-90s when computing power and tactor components became efficient enough to realize the first concepts (Gilson et al., 2007). Rupert and his research team (Rupert et al., 1994) reused the idea from the 50s of presenting information through a matrix of tactors. After a number of successful studies and demonstrations it was possible to bring the concept into real cockpit environments (Rupert, 2000; Rupert et al., 1994; Rupert, McGrath, & Griffin, 2002). This inspired other research teams and extended the field of military use including combat vehicles, boat operators, dismounted soldiers, and for operators remotely controlling systems (Elliott, Duistermaat, Redden, & van Erp, 2007; Eriksson et. al., 2008; Haas & Stachowiak, 2007; Redden, Carstens, Turner, & Elliott, 2006; van Erp & Self, 2008).

Robust coding for touch depends on how intuitive the signals are and thus how well they are able to represent the information (Brewster & Brown, 2004). Compare the robust coding of visual or auditory icons on a computer where a bin represents a place for discarding files or the sound of a phone ringing represents an incoming call. Please see figure 14 below for a simple example.

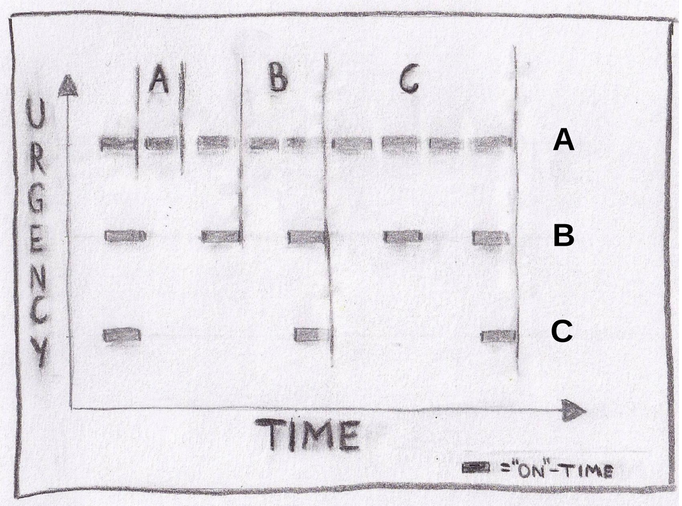


Figure 14. The tactile signals can be a simple buzz or a complex melody similar to that of Morse Code (Carlander & Eriksson, 2006). This is a simple example of urgency indication where A shows a small separation between pulses and thus indicate a very urgent condition. The pulses in B indicate medium urgency and finally C have longer periods of silence between pulses and is thus least urgent.

In the tactile domain, several perceptual dimensions can be used to display information. The dimensions can primarily be used for a diversity of functions and they all have properties that make them more or less suitable for different functions. Below some dimensions are summarised and discussed.

Tactile coding dimensions

The most important property of a single tactile coding dimension is the number of distinctive perceptual levels that can be accomplished within that dimension. The most common perceptual dimensions are location, duration, frequency, and intensity/amplitude. Out of these dimensions location and duration seem to be the most promising (Jones & Sarter, 2008). Frequency and intensity is not as useful due to a limited number of distinctive perceptual levels. Dimensions can also be combined into new dimensions, such as duration+location = motion, to achieve new distinct perceptual levels (Gilson et al., 2007; van Erp, 2002).

Location

With tactors used at several body locations, these can be used as a display that can code information by location. How many tactors that are used and where they are placed on the body affects the perceptual experience, thus the resolution is an important factor for this dimension. Depending on task and type of information, certain parts of the body might be more suitable for tactile presentation compared to others due to the spatial acuity of the skin (van Erp, 2002). Location can be used to create many distinctive perceptual levels, and as such, it is probably superior to the other dimensions (van Erp, 2005a). Simultaneously activated tactors can also produce clearly distinctive patterns used to display different messages (Gilson et al, 2007; van Erp, 2007).

Location is most suitable when attention is needed in the front or back of the user, and it is useful for the indication of directions or threats. Important to note is that the distance between two tactors must be large enough to prevent overlapping and also that short presentation times may require a larger tactor distance, to avoid the risk of spatial masking (van Erp, 2007). Spatial masking is when the location of the stimulus is masked by another stimulus at a nearby location. Simply put it is how spatially separated stimuli affect each other (Cheung, van Erp, & Cholewiak, 2008). This is of course important to consider when using tactile patterns where both detection and identification of signals may be degraded by nearby tactors activated simultaneously (van Erp, 2002; van Erp & Verschoor, 2004).

Duration

The duration corresponds to the "on-time" of the tactor and the skin is very sensitive to changes in duration and the tactile signal can thus be effectively coded by temporal patterns. The temporal patterns are created when a tactor is switched on and off rhythmically (as the example in figure 14). The rhythms can potentially enrich the tactile information by generating intuitive tactile melodies (van Erp, 2002). Humans are very good at perceiving rhythms and users can distinguish between many levels of rhythms and the rhythms are quickly recognized (Sacks, 2007; van Erp & Spapé, 2003). The ability to perceive rhythm and our accurate memory for tempo and rhythm (Sacks, 2007) ties closely to the skin and it is therefore an important parameter for designing tactile feedback. Furthermore, rhythm encoding is widely represented in the brain making it a robust phenomenon that closely interconnects to our other senses. Our inner rhythm and understanding of movement and timing give us the capability to anticipate the beat and recognize rhythmic patterns as soon as we hear them. Because rhythm is interpreted since infant age we have developed internal rhythm templates that are very precise (van Veen & van Erp, 2003).

For a single tactor we can detect pulses as short as 10 ms (e.g. 10 ms on, 10 ms off) (van Erp, 2002), but pilot studies show that the duration should not be shorter than 200 ms and not exceed 400ms (Gilson et al., 2007). However, Kaaresoja and Linjama (2005) found that when using the tactile signal for alerting functions the tactile pulse should be between 50 and 200 ms in duration and participants reported longer durations as annoying. The duration of the signal is a trade-off between detectability and adaptation, longer duration enhances the ability to identify the tactile stimulus (Jones & Sarter, 2008), but on the other hand, a continuous tactile stimulus leads to effects of local adaptation since the skin cannot "refresh" the stimulus. This means that when the perceptual sensitivity gradually decreases, the smallest detectable level of a stimulus increases (Hahn, 1966; Verrillo & Gescheider, 1975). The effect of adaptation can be avoided by altering the frequency (van Erp, 2002). The choice of tactor technology can be of importance for the duration since the time required to reach the desired frequency and back to quiet differs between technologies[5] (van Erp, 2002).

---

[5] Please refer to appendix A and the section on tactor technologies.

Using duration to code multiple pieces of information simultaneously within a single tactile display requires a very careful design. A tactile display with multiple pieces of simultaneous information must ensure that each rhythm is clearly identifiable and different from the other rhythm(s). In a study by McKinley et al. (2005) three different tactile rhythms were used to present types of target aircraft (enemy, unknown, and friendly). Three targets were displayed simultaneously, and the experimental results and subjective feedback indicated that it was difficult to differentiate between the simultaneous rhythms. This can be related to that sensory signals that occur simultaneously with little spatial discrepancy will most likely be integrated and on the other hand, signals with a large spatial discrepancy or a non-appropriate temporal sequence will most likely not be integrated (Ernst & Bülthoff, 2004).

The difficulty in interpreting simultaneous rhythms was also anticipated by van Erp and Spapé (2003), since presenting two stimuli closely in time may result in temporal effects (e.g. temporal masking where the simultaneous tactile presentation mask each other). If several pieces of information needs to be simultaneously displayed in the tactile display it is the beginning of the rhythmic signal that is most important. Spence and Ho (2008) also commented that given the limitations on tactile information processing they "remain unconvinced" of the utility of complex tactile displays. They pointed out that the tactile displays should not require extensive training in order for the users to be able to use them efficiently.

Frequency

The skin has its greatest sensitivity to tactile stimuli between 150 and 300 Hz. When using frequencies below or above this frequency range, the distance between tactors must be greater to compensate for the loss of signal sensitivity (adaptation). In the coding, different frequency levels can be assigned different meanings. However, only a few levels/dimensions can be used for information presentation. Again, the choice of tactor technology is important since some technologies cannot separately control the frequency without affecting the amplitude (e.g. the rotary motion tactor) (van Erp & Self, 2008).

Amplitude

Information can be encoded by different intensity levels induced by, for example, low, mid and high amplitude. The number of possible levels is small and between detection threshold and the "pain" threshold. As for frequency, not more than 4 different levels of amplitude should be used (van Erp, 2002). Using too many levels within the amplitude dimension can result in that the specific "level information" is missed during active tasks (Gilson et al., 2007). However, the four levels can be sufficient for reflecting the levels of urgency for operators in control rooms, where low amplitude = low threat level and high amplitude = high threat level (van Erp, 2002).

Spatio-temporal interactions

The different dimensions described above are integrated by the tactile sense. This integration could appear as a new percept that differs from the intended stimulus (van Erp et al., 2003). For integration, body location and temporal patterns (timing) are the two most important stimulus parameters, and they strongly interact (Cheung et al., 2008). When short tactile pulses are presented by tactors having a small separation on the body the spatio-temporal interaction is likely to occur (van Erp et al., 2003). This interaction can create a tactile illusion of displacement that can be perceived as motion. The illusion of apparent motion occurs when tactors are sequentially activated at different body locations. It is best described as a "pulse train" where tactor spacing and the interval between pulses mimic something moving over the body. As for spatial sound, the occurrence of movement seems to improve localisation (Gilson et al., 2007; van Erp, 2002).

Related to apparent motion is the "apparent location" where the perceptual experience is dependent on the relative magnitude of two activated tactors. The resulting experience is the percept of a position between these activations[6] (Sherrick, Cholewiak, & Collins, 1990). Apparent location has been suggested to reduce the number of tactors in tactile displays by increasing the number of subjectively experienced stimulus sites, without increasing the number of tactors (McGrath, 1999; van Erp, 2002). Thus, the phantom sensations could be used to achieve a higher spatial density of tactors than is actually present.

Another illusory phenomenon that differs slightly from the apparent location is named "the cutaneous rabbit illusion". It can be demonstrated by stimulating two separate locations on the skin where pulses at the first location are followed by a single pulse at the second location. The perceptual experience will be a hopping of a stimulus from one location to the next (hence the term "cutaneous rabbit"), even though the stimuli were applied at two distinct points (Geldard & Sherrick, 1972; Jones & Sarter, 2008). Some claim to have used the cutaneous rabbit effect in real-life applications, for example the movie vest described further down in the section of tactile applications (Jones, 2009; Lemmens, Crompvoets, Brokken, van den Eerenbeemd, & de Vries, 2009).

Tactor technologies

Rupert, McGrath and Griffin (2002) summarise the ideal tactor in the following way. The tactor must be controllable in frequency, amplitude, and in waveform to be able to meet the full perceptual frequency range. Furthermore, it should have no electromagnetic or acoustic signature[7], be small, robust, lightweight and waterproof with low power requirements and cost. These criteria are hard to meet and the choice of tactors is limited to the range of available tactor technologies. Below the most common and useful tactors are described, and the summary is limited to tactors that are used, or potentially can be used, in operational environments.

Vibromechanical

Within the category of vibromechanical motors the electro-mechanical, pneumatic and hydraulic tactors are further described below.

The two most common groups of **electro mechanical** tactors are; *rotary inertial* and *linear actuators*. The *rotary* motion tactor or the "pager-motor" tactor has characteristics like small size, low cost, robustness, and low power that make it a popular choice for many applications (McGrath et al., 2008). This rotary motion tactor is common in mobile phones and is basically a motor with an off-axis weight on the rotating axis encapsulated in a housing material (see Figure 15 for an example). Normally these motors are activated with a fixed frequency and amplitude (Jones & Sarter, 2008). The tactile stimulus

---

[6] Compare with the effect of an audio presentation in stereo; the sound source can appear between the speakers depending on the relative amplitude of the speakers.

[7] The need for tactors with no electromagnetic or acoustic signature is mostly a military requirement.

frequency is defined by the revolutions per minute (rpm) of the motor and is typically in the range of 4000 – 9000 rpm (i.e. 70 – 150Hz) (McGrath et al., 2008).

Figure 15. A tactor developed by the Swedish Defence Research Agency, FOI. It is based on a rotary DC motor encapsulated in a PVC cylinder.

The motors are simple to control and the vibrations conveyed to the skin can easily be perceived as a localised buzzing. Intensity and frequency of the specific motors and hosing/mounting is affected by the distance between the tactor and the skin. For optimal signal transfer, and localisation ability the tactor should be placed close to the skin. However, vibrations can be made strong enough to be more or less independent of direct skin contact. Depending on the application, position, task, and vibrator type tactors are mounted differently. FOI and the Swedish school of textiles (University college of Borås) carried out a project regarding the integration of tactors into garments[8]. In figure 16 below some of these concepts are shown.

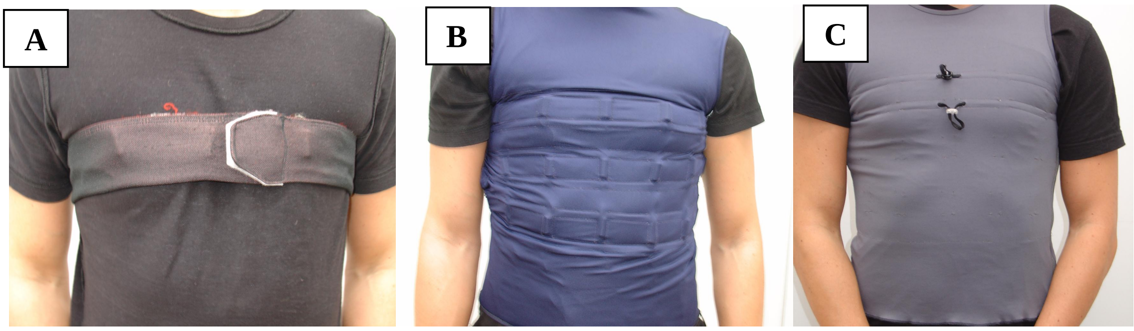


Figure 16. Examples of some tactile display concepts from FOI. These were developed in cooperation with the Swedish School of Textiles (University College of Borås). (A) FOI tactile torso belt as worn in the combat vehicle studies. (B) Tactile vest prototype with tactors. (C) A tactile vest prototype with adjustable size (Carlander, 2006).

Even though widely used there exist objections to the rotary-motion tactor type. The most critical issue is that the frequency of the vibration is proportional to the speed of the motor and thus dependent of the driving voltage. A higher frequency will generate higher amplitude, resulting in that the motors are difficult to control regarding separate stimulation parameters of the skin. Due to the mechanical construction the motor also has a slow onset. It takes some time before it reaches desired frequency and before going

---

[8] In Berglin (2006) and Eriksson at al. (2005) the integration of tactile feedback in textiles is briefly mentioned.

back to quiet. The flexibility of coding more complex touch messages are thus limited (Gilson et al., 2007). However, these objections can only be considered valid when the tactors are used for more complex coding or specific perceptual phenomena (McGrath et al., 2008), such as, for example, the cutaneous rabbit that needs a 2ms burst duration (van Erp, 2007).

The *linear actuator* tactors solve the above issues by allowing control over amplitude and frequency independently (McGrath et al., 2008). They are coil based actuators comparable to small loudspeakers, see figure 17 for an example. This tactor applies force in a linear manner, as opposed to the rotary tactors described above (van Erp, 2007). The linear actuator has been optimized for use on the skin, where a moving component is oscillating vertical to the skin. This oscillating movement places a light force on the skin and can be felt as a point-like sensation even through layers of clothing (McGrath et al., 2008).

Figure 17. A linear actuator attached to a waist belt for conveying hand- and arm signals. The smaller circle centred on the tactor is the moving part. The picture is taken on a demo session at a North Atlantic Treaty Organisation (NATO) Research and Technology Organisation (RTO) HFM-122 group meeting in Breckenridge, Colorado, USA, 2006.

Many of the linear tactors in use are designed to vibrate in the 200 – 300 Hz range, which corresponds to the optimal sensitivity of the Pacinian corpuscles (McGrath et al., 2008). Important to note is that this motor is not able to deliver broad enough frequency spectra to allow frequency-coded information, only 10-20 Hz off their resonance frequency is possible (van Erp, 2007).

The **Pneumatic tactors** were developed upon request from the Joint Strike Fighter Program (JSF, Fighter Jet in the USA) and is one of the best tactor technologies for the aviation environment (Rupert et al., 2002). The pneumatic tactor is a form of linear actuator but the mechanical movement is generated by compressed air instead of electricity. The tactor itself typically consists of a container with an opening for the incoming compressed air and a membrane for generating the tactile sensation on the skin. The compressed air flows through valves that control the oscillary movement of the membrane. The great advantage of the pneumatic tactors is that they are robust, lightweight and are able to produce a strong tactile sensation at a 40 – 50 Hz vibration (McGrath et al., 2008). Much due to these advantages the pneumatic tactor has been integrated into the Tactile Situation Awareness System (TSAS) that is a tactile display with the purpose of presenting spatial orientation information to pilots (Rupert, 2000).

The **Hydraulic** tactors are similar to the pneumatic tactors but instead of air uses fluid to affect the tactor. QinetiQ, a British defence company supporting the national defence departments, developed a system used in the Diver Reconnaissance System (DRS). In this system fluid is forced into the DRS handles that expand and contract as means of providing tactile navigation information via the hands of the diver (Dobbins & Samways, 2002; McGrath et al., 2008).

Electrical and Thermal

The tactors in an **electrical** tactile system are comprised by tiny electrodes of any conductive material. The touch sensation is acquired by passing a small electric current through the skin and thus generating an electrotactile excitation, also known as electrocutaneous stimulation. The electric excitement is described as a tingle, pressure, vibration, or pain, all depending on electrode, waveform properties and the sensitivity of the user. A great advantage of the electrical tactor is that the pulses can be controlled regarding voltage amplitude, frequency, duty cycle, and polarity. The electrical tactor also has benefits regarding localisation since the tactile sensation is very localizable and thus allows for complex tactile arrays over a relatively small surface area. The system is small in size and lightweight due to the lack of moving parts or mechanical functions, making tactors more reliable compared to mechanical tactors. The electrical tactors will most likely not reach widespread use due to the fact that even small changes in tactor position can alter the sensation and pain thresholds, and thus the dynamic range and the comfort of the stimulation. Other parameters such as sweat will change the conductivity of the skin and affect the sensation. Electrical tactors can be useful in the laboratory, but currently not for use in operative environments (McGrath et al., 2008).

Thermal displays are merely for virtual environments to display if objects are warm or cold. Although thermal displays exist (Jones & Ho, 2008), they are too slow and diffuse to be used in an operational environment. Also, our natural changes in skin temperature and our differences in responses to warm or cold stimuli (Guyton, 1991) are additional factors that make this tactor type impractical.

Future tactile technologies

Electro-active polymers (EAP) offer interesting possibilities for future tactile displays. These flexible materials are capable of converting electrical energy to mechanical energy and thereby generating a force (Shankar, Ghosh & Spontak, 2007). EAPs contract, expand or bend to a limited extent when a voltage is applied to them and the expansion of the material thus applies a light pressure to the skin. A device presented by Choi, Koo, Koo, Nam and Lee (2008) shows a flexible polymer film covered with tiny actuator cells (the black dots, seen in figure 18). This electroactive polymer has capabilities allowing elastic strain without loosing its original shape. Furthermore this EAP tactile display does not require complex electronics that would be an advantage compared to the other tactor technologies (Choi et al., 2008). The technology is undergoing quite extensive research and in the area of Braille displays the technology is close to performing at required specifications. However, challenges still limit widespread use, specifically the short cycling life of the conducting polymers and the challenges related to a reliable operation. The aim is thus to develop more effective materials and processing techniques so practical, low-cost, tactile interfaces can be constructed (Bar-Cohen, 2009). A Research team led by professor Choi claim that most of these challenges are overcome and that tactile display devices based on this technology offer advantageous features over existing devices such as material flexibility, small size, low cost, easy to manufacture and miniaturize. Due to its flexibility it can be adapted to various parts of the body and used as a tactile display (Choi et al., 2008). However, there are some disadvantages including high driving voltage and low bandwidth that may not work for all applications (Koo et al., 2008).

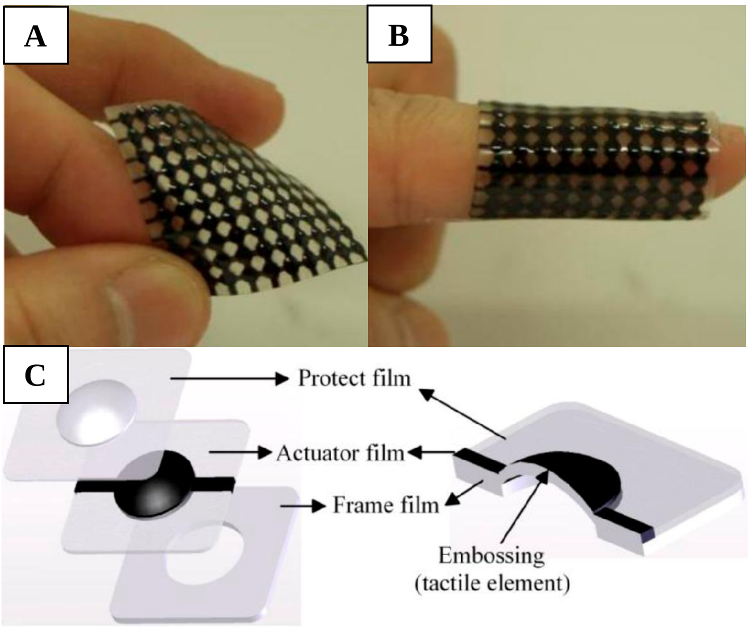


Figure 18. EAP example as shown by Choi et al., (2008). (A) the flexibility of an EAP tactile display and (B), attached to the index finger. In C, a cross-section of the tactor reveals the active tactor element (black semi-sphere). Pictures are used with permission from SAGE and the authors.

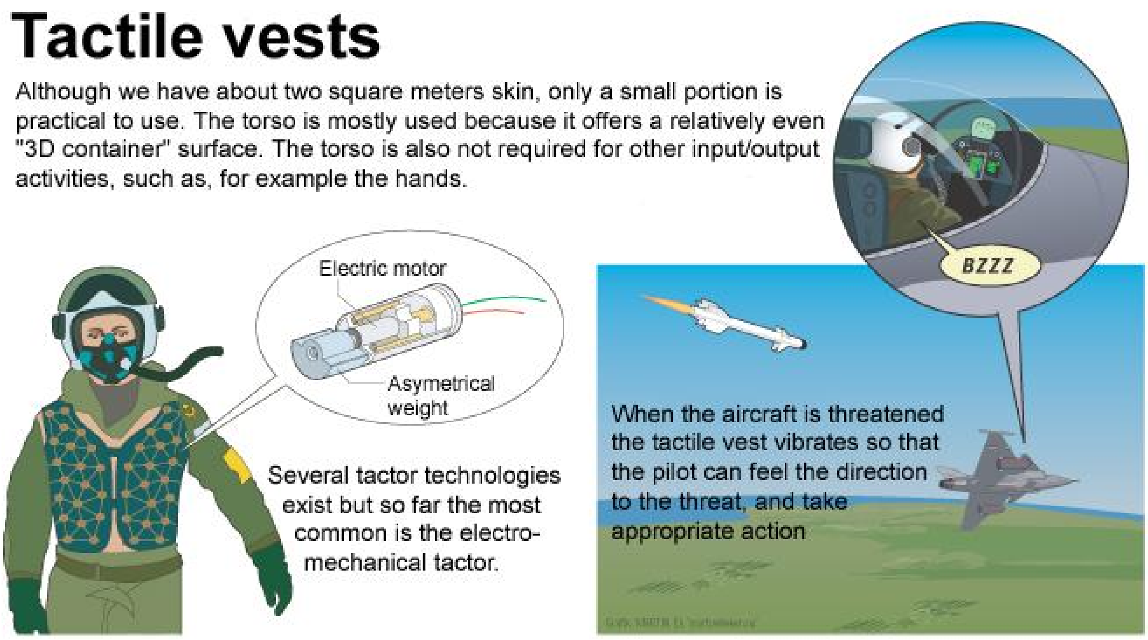
For a comparison of the tactor technologies (EAP excluded), the NATO RTO task group HFM-122 compiled a useful tactor technology comparison table shown in Appendix A (van Erp & Self, 2008).

Tactile Display Applications

Regardless of modality, information that should be easy to comprehend must be kept simple. A good example of a simple tactile display is the vibrating alert on a cell phone. The device signals for attention and you look at it because you know it has new information. Furthermore, the tactile display is multidimensional in nature with a potential of carrying more information. For example, the vibrating alert could also include a temporal dimension where the tactile pulses indicate type of information; phone call, text message, meeting reminder etc. The tactile display for the cell-phone is widely used because it is accurate, reliable and easy to comprehend. Extending the use beyond cell phones a tactile display can for example present information regarding directions in drift, navigation or threat positioning, where the tactile information is coded as a specific spatio-temporal pattern conveyed by vibrations in the tactors (e.g. Eriksson et al., 2006; van Erp, 2005a). It has been shown that tactile displays are especially useful in situations where the visual channel is overloaded, high cognitive demands exists, or redundant cueing is needed (e.g. Gilson et al., 2007; Redden, 2006; van Erp, 2007).

No firm set of guidelines yet exists for a tactile navigation display with the result that different coding principles are used; ‘follow the needle’ and ‘virtual corridor’. ‘Follow the needle” principle indicates the desired direction by activating the tactile elements as the indication of the desired route or the next waypoint (e.g. if the tactor at 12 o’ clock is activated go straight ahead). The other coding principle, “the virtual corridor”, requires the operator to manoeuvre away from the signal, similar to bouncing off a wall. The virtual corridor or wall analogy thus indicates the direction to avoid, which leaves the operator with the choice to determine the optimal steering action. The downside with this approach is that the display will not guide the operator to the optimal direction since there are many ways of avoiding the “virtual wall” (van Erp & Veltman, 2002). On the other hand this approach might be suitable for some applications since it gives the user a more “quiet” environment with tactors only activated when needed. Although no data is currently available for the comparison of the performance for both ways of coding, it is likely that performance degrades if operators have to switch between different kinds of coding (van Erp & Self, 2008).

Tactile displays can be used for a wide range of applications from alerting signals to vehicle control. The main advantages of the tactile displays include the possibility for spatial discriminations over a large area (the skin) in combination with being an effective way of capturing the attention of the user (Eriksson et al., 2005; van Erp, 2005b). Tactile displays can also reduce the perceived workload since information can be processed fast with little mental effort and furthermore, operators can remain focussed without visually diverting from their task (van Erp, Veltman, van Veen, & Oving 2003; van Veen & van Erp, 2000). We ran studies concerning whether displayed tactile threat information can improve the threat awareness of operators controlling platforms such as combat vehicles or fighter jets (Carlander & Eriksson, 2006; Carlander et al., 2007; Eriksson et al., 2005, 2006; Oskarsson et al., 2012; van Erp et al., 2007), and these studies are described in the next chapter. Figure 19 below exemplifies the use of a tactile display.



Tactile vests

Although we have about two square meters skin, only a small portion is practical to use. The torso is mostly used because it offers a relatively even "3D container" surface. The torso is also not required for other input/output activities, such as, for example the hands.

Several tactor technologies exist but so far the most common is the electro-mechanical tactor.

When the aircraft is threatened the tactile vest vibrates so that the pilot can feel the direction to the threat, and take appropriate action

Figure 19. Example of a tactile vest application for a fighter jet pilot.

Castle and Dobbins (2004) identify three major military areas where tactile displays are considered to have a great potential; orientation, navigation, and communication. This is also covered more in-depth in van Erp and Self (2008). Orientation, navigation, and communication for tactile displays are summarised below.

Orientation and Navigation

The tactile torso displays have turned out to be especially useful in orientation and navigation tasks. Due to its usefulness, a vibrotactile display can in many demanding situations be considered an alternative to a visual display. Information brought by visual or auditory displays might not be suitable in tasks that require “eyes and ears open”. A large number of studies have shown that tactile information is perfectly suited for navigational tasks like waypoint navigation on land (Elliott et al., 2007; Gilson et al., 2007; Pettitt, Redden, & Carstens 2006; Redden, Carstens, Turner, & Elliott, 2006; van Erp et al., 2005) and sea (Dobbins & Samways, 2002). Studies with tactile displays concerning orientation have been made for air (McGrath et al., 2004; van Erp et al., 2003) and space (van Erp & van Veen, 2003). We also contributed to the area of tactile navigation by performing a joint-effort study with the industry regarding dismounted soldier navigation including a tactile navigation display (Eriksson et al., 2008).

The tactile presentation must be carefully designed to achieve an intuitive mapping between the external world, the task and user interpretation. van Erp (2007) expressed this in the following way: *“The design of a tactile navigation display requires finding an optimal translation of… [the external world] … into a tactile “picture”* (van Erp, 2007, p. 107).

Initially many displays were designed for the hands and fingers due to the low sensory threshold and the high acuity of the tactile perception in this area. However, the hands are normally occupied with motor activities and the possibilities of presenting 3D based information are thus lost (van Erp & Self, 2008). To acquire intuitive presentation of information in three dimensions full torso coverage is needed (Rupert et al., 2002). Tactile display concepts exist with tactors, arranged in rows and columns forming a matrix (e.g. McGrath et al., 2004; Rupert et al., 1994; van Erp et al., 2007). However, there are situations when the resolution offered by the matrix is not needed. For instance, applications needing to represent only the horizontal plane would not need the columns covering the vertical dimension, and such single-row tactile displays can easily be constructed from COTS (see Eriksson et. al., 2005, 2008; van Erp et al., 2005, and figure 26 for examples). In the experiments where we used the tactile belt as waypoint navigation support for dismounted soldiers the subjective comments were positive about using the tactile display. Being more or less representative for most of the participants, one soldier claimed that the tactile information *”gave ‘freedom’ for [performing] other tasks; hands, ears, eyes free”* (Eriksson et al., 2008, p. 1286). This was similar to what Elliott et al. (2006) reported in their study where a tactile navigation system was considered “hands-free, eyes free, and mind-free” (p. 20).

Utilizing the skin and the three-dimensional capabilities that can be achieved with tactors onto the body surface, the sense of touch can provide spatial orientation to operators. It can thus counteract spatial disorientation[9] and the US Naval Aerospace Medical Research Laboratory (NAMRL) developed a tactile vest concept for this (Rupert, McGrath & Griffin, 2002). The TSAS is equipped with 28 tactors, 24 mounted on the torso and two on the shoulders and two under the thighs. This system provides continuous information for orientation awareness and it was proved that the sense of touch could serve as an effective substitute when sensory perception is limited for other modalities. TSAS also reduced workload in complex environments (Rupert et al., 2002; van Erp & Self, 2008), and one of the pilots that tested the system wrote: *"TSAS enabled me to decrease my cockpit scan to altitude and heading control only. Felt like second nature to rely on drift corrections from TSAS."* (Rupert et al., 2002, p. 31-5)

There have also been successful European tactile displays for air and space applications. A tactile application relating to gravity was investigated at the International Space Station (ISS). This study showed that the tactile display potentially could benefit the astronaut's performance, safety and well-being. The astronauts reported that the display worked as *"the pressure on the sole of our feet to keep balance without being aware of the signals"* (van Erp, 2007, p. 133).

Flight applications have been popular for the tactile display due to its intuitive 3D character. For example, the TNO team investigated using tactile signals as indication for drift in a helicopter hover display (van Erp et al., 2002). The pilot had to follow[10] the tactile signal by manoeuvring the helicopter in the direction of the vibrating elements. If the signal was above (e.g. the shoulders) the pilot had to hover upwards. van Erp concluded that the altitude errors seen in flight was reduced by 50% and that a tactile display did not affect the subjective mental effort during the task. Similar route indications have also been successfully applied for both land and sea applications (van Erp, van Veen, Janssen, & Dobbins, 2005).

Communication and Entertainment

The tactile information can also be used for communication, and since the tactile sense operates close to the body it allows for the use of 'private' or 'exclusive' displays. This can be important in military tasks, or if the operator environment limits the use of the auditory or visual sense (Jones & Sarter, 2008). A good example is the use during mission critical activities at night. At night, visual or auditory displays cannot be used due to their signatures, and that night vision adaptation is disrupted and the need of working with "ears open" (van Erp & Self, 2008).

Some good examples exist for translating common arm-and-hand signals used in the US army. Brill et al. (2004) used a tactile torso belt display with eight tactors creating a sensation of movement around the wearer. The signal 'travelled' on and around the torso to create a representation of the hand- and arm signals (Brill et al., 2004; Pettitt, Redden, & Carstens, 2006). Further, Pettitt et al. (2006) reported that soldiers experienced the translated signal patterns to be intuitive and easy to understand. Interestingly, system acceptance and good performance was acquired after only a 10-minute hands-on demonstration.

---

[9] Spatial orientation defines as knowing how you are oriented in space and to be able to distinguish up from down (van Erp, 2007).

[10] "Follow the needle" principle as described earlier.

In 2009 the entertainment industry developed a jacket equipped with 64 actuator motors. The aim was to produce various tactile sensations relating to the content of the movie being currently watched. The tactile sensations are claimed to target “emotions” rather than “information” by influencing the emotional state of the viewer. The development team claims to actively use and take advantage of the cutaneous rabbit illusion (see the earlier section on tactile coding dimensions) by reducing the number of tactors needed (Jones, 2009; Lemmens et al., 2009).

Gaming is another area that has been using tactile feedback for quite some time. One system is the “3rd Space FPS Vest” (see figure 20 below) that uses eight pneumatic tactors to simulate computer game events like punches, explosions, recoil etc (Coxworth, 2010).

Figure 20. A) Two versions of 3rd Space FPS tactile gaming vest. B) The Pneumatic tactors are embedded in small pouches around the vest. C) The vest connects via a control box over USB (Coxworth, 2010).

As described above, the perceptual characteristics of the user experience and knowledge on how spatiotemporal patterns are processed are important (van Erp, 2002). Correctly applied the patterns can enhance information in a tactile display, but regardless of tactile parameters, the tactile messages must be self-explaining. This is especially important for complex tactile messages that should be built of meaningful components (such as fast pulses for urgency or the translation of known rhythms, like hand-signals) (van Erp, 2002). Failure to keep signals simple and self-explanatory may result in tactile clutter with reduced comprehension or sensory overload as a result (van Erp et al., 2003). Situations where it is most likely that a tactile display would increase performance and ease mental workload is when there is high mental workload and visual load (e.g., Gilson et al., 2007; Redden, 2006; van Erp, et al., 2003).

Key-points Chapter 3

- **A tactile display is analogue with a “tap on the shoulder” and thus thrives at being intuitively understood.**
- **The tactile stimulation is acquired by mounting small vibrating elements close to the skin, and these vibrating elements are often referred to as tactors.**
- **Vibromechanical motors are the most common tactor technology offering electro-mechanical and pneumatic tactors.**
- **The tactile signals can be a simple buzz or a complex melody similar to that of Morse Code.**
- **The most common perceptual coding dimensions are location, duration, frequency, and intensity/amplitude.**
- **Electro-active polymers (EAP) offer interesting possibilities for future tactile displays due to the flexibility of the material (van Erp & Self, 2008).**
- **Tactile information and coding must be kept simple for immediate comprehension.**
- **The main advantages of the tactile displays include the possibility for spatial discriminations over a large area in combination with being an effective way of capturing the attention of the user (van Erp, 2005b).**
- **Tactile information is perfectly suited for navigational tasks like waypoint navigation on land (Eriksson et al., 2008) sea (Dobbins & Samways, 2002), and applications for threat or target information (Rupert et al., 2002; van Erp, 2005a; van Erp et al., 2002; van Erp & Self, 2008).**
- **Tactile displays covering the torso can be utilized for 3D navigation since vibrators can be mapped to external locations.**

Contribution to tactile display research

- **Participation in NATO RTO task group HFM-122 contributing to a report covering tactile use and technology (McGrath et al., 2008).**
- **Development of the FOI tactile system used in a number of studies, and especially important was the development of a low cost tactor encapsulated in a PVC cylinder. Please see next chapter for studies using this tactor (Oskarsson et al., 2012; Carlander et al., 2007; Carlander & Eriksson, 2006; Eriksson et al., 2005).**

Chapter 4. Multimodality

Today, many information systems focus on presenting information to one (unimodal) or two (bimodal) senses. Using only one or two senses can lead to negative consequences for our attention as a result of over-using the same sensory channel(s). Furthermore, the system might not reach maximal robustness since information presented visually or auditory can be interfered by glare, darkness, noise etc. (Brill et al., 2004). In these situations we need to consider the use of other senses that may not be interfered from the same source.

Theoretical background

The awareness derived from simultaneous information from our sensory modalities is the foundation for how humans normally gather information to interact with the environment. This is more or less intuitive to most people, which Spence and McDonald (2004) exemplifies:

> ” […] if a mosquito lands on our arm, our eyes will be drawn immediately to the source of the unexpected tactile event. In these and many other such situations, objects that are initially processed in one sensory modality “grab” our attention in such a way as to enhance the sensory processing of stimuli presented in other sensory modalities at the same spatial location […] there is now convincing empirical evidence that the covert orienting of exogenous attention that is triggered by the presentation of auditory, visual, or tactile cue stimuli can facilitate the perception of target stimuli presented subsequently at the cued location, no matter what their modality. In fact, cross-modal cueing effects have now been demonstrated behaviourally between all possible combinations of auditory, visual, and tactile cue and target stimuli […]” (Spence & McDonald, 2004, pp. 3-4, as referred to in Eriksson, Lindahl & Hedström 2006)

This shows that the perceptual experience is constituted by complex sensory interactions that grow stronger with simultaneous multisensory input. We are used to combine sensory modalities since specific stimuli are often made up by signals in several modalities[11], and integrated into a coherent percept (Eimer, 2004; Ernst & Bülthoff, 2004). Mainly, the CNS ‘sorts’ incoming signals and combine the ones that are likely to originate from the same source in an optimization process leading to reduced variance and enhanced chances of detecting a stimulus (Bresciani et al., 2005; Ernst & Bülthoff, 2004). Humans have two general strategies for combining information: 1, Maximize information from the different sensory modalities (“sensory combination”). 2, Reduce the variance in sensory estimation to increase its reliability (“sensory integration”) (Ernst & Bülthoff, 2004). The theoretical background thus shows on a potential of improving, for instance, threat presentation by triggering several senses with the *same* type of information enhancing stimulus detection.

---

[11] Take the knocking on a door as an example, you will feel and hear the knock while you are seeing your hand. The knocking itself will be experienced as a coherent percept.

However, multimodal presentation is not only about combining the same type of information from several sensory modalities but also how to allocate *different* type of information to different modalities. This is what can be called multimodal function allocation and has the potential of making interfaces more intuitive by using the most "intuitive" presentation for each modality. In Gilson, Redden and Elliott (2007) the following summary can be found: *"The distribution of information across the various sensory systems can help overall mental workload by reducing channel competition and bottlenecks associated with modality overload"* (Gilson, et al., 2007, p. 2.).

Having multiple resources allow us to simultaneously process information from more than 'one channel' with the effect that several tasks can be performed more or less simultaneously with little interference. Thus, comparing two tasks that share resources (intra-modal) with tasks that use different resources (cross-modal) the latter will result in better performance (Wickens, 1987; Wickens & Hollands, 2000). To exemplify, two spatial tasks will use more working memory than one spatial task and one verbal. A dismounted soldier that needs to keep track on his location both on the map and in the terrain would require the use of vision for both tasks. Here it could be beneficial to offload the visual sense by presenting navigational information to the tactile or auditory sense.

In line with our multimodal capabilities, Wickens (e.g. 2002; 2008) proposes the Multiple Resource Theory (MRT) that suggests we have independent resources for processing information, and not one single information processing source that can be tapped. Some of the resources can be used simultaneously with parallel processing (Wickens, 2002). *"That is, the eyes and ears behave as if they define multiple processing structures or 'resources'"* (Wickens, 2002, p. 162).

Thus, tasks that require the same pool of resources for information processing may have to be performed sequentially (as the example using visual navigation for the dismounted soldier above). If different pools of resources, or senses, are utilized the task can be processed more or less in parallel. Thus, offloading information from one sense to another can have beneficial effects on the cognitive workload (Wickens, 1987). Wickens (2008) describes four dimensions as being the factors accounting for the variance in time-sharing performance, and these four dimensions also have neurophysiological plausibility by being "parallel" in the anatomy of the brain:

1. STAGES of (temporal) processing
    a. Encoding/perception
    b. Central processing/cognition
    c. Response(s)
2. MODALITIES (encoding/perception)
    a. Visual
    b. Auditory
3. CHANNELS (e.g. focal and ambient vision)
4. CODES Processing code
    a. Verbal and spatial
    b. and manual and vocal (response, see 1c above)

An example of *stages* could be that a driver vocally answers a radio call (c, response) while choosing the next waypoint (a+b, perception and processing). *Modalities* are "encoders" at stage 1a, and are represented by our senses, and the underlying assumption is that we are better at cross-modal time-sharing (e.g. auditory-tactual) than intra-modal (e.g. auditory-auditory) (Wickens, 2002). Wickens only specifies the two most important modalities that dominate our sensory experience but all our sensory modalities could be listed. Perhaps, as suggested by Hancock, Oron-Gilad and Szalma (2007), each "sensory box" could be adjusted in size to better represent its importance in attention. *Channels* show two different aspects of visual processing, focal and ambient vision. Focal vision is for focused visual activities like reading and identifying fine details. The ambient vision primarily involves peripheral vision and relates more to the perception of movement and orientation. These channels successfully share time and a likely explanation is that ambient vision, as opposed to focal vision, requires little or no conscious resources and can be directly related to automated function (Wickens, 2002). *Processing codes* relate to spatial and verbal processes and is often associated with our two hemispheres; usually the left hemisphere is verbal and the right spatial in nature. This is also the main explanation to why manual and vocal activities effectively can be timeshared (Wickens, 2002). As an example, a manual response is usually spatial such as steering a car and a vocal response is usually verbal such as having a conversation with a passenger.

Important to note is that the MRT model is a simplified and rather restrictive model of our attentional mechanisms. Although the theoretical framework has been encapsulating new findings to stay updated, our mental processes and modalities should not be reduced to isolated processes and entities, as partly implied above. There is redundancy and cross-talk between the "boxes" of the MRT (Hancock et al., 2007). Other challenges to the model include the lack of the tactile modality, a description of the complete resource demand showing "when" we are out of resources, and finally "what" drives the allocation of resources by the "central executive" of the working memory (Wickens, 2008). However, not taking more in-depth theoretical aspects like this into account, the simplification of our attentional mechanisms and representation of the "boxes" in MRT, is a great help for human factors researchers and interface designers. The relatively simple model illuminates some of the issues of attention alongside being specific enough to provide useful assumptions of the human factor implications regarding attentional conflicts in a system (Hancock et al., 2007). The value of the MRT relates to its capability of being applicable to the real-world by making predictions. It has been shown that there is a high correlation between model predictions and real data. The MRT should thus mostly be viewed as a tool to recommend design changes, for situations with multi-tasking and resource overload, which later could be empirically evaluated and adjusted (Wickens, 2002; 2008).

A good example where sensory characteristics are taken into account is the spatial orientation retention device (SORD), which is a multisensory aircraft attitude tool developed to allow the pilot to retain and support spatial orientation (e.g. Albery, 2006). Sensory presentation was combined to achieve an instrument based on redundancy and human sensory perception. The system is comprised by a tactile display, a 3D audio display, Helmet Mounted Display (HMD) symbology, and a system monitoring the orientation of the pilot. The orientation system includes functional brain monitoring helping to assess if the pilot is overloaded or spatially disoriented. The SORD was transitioned to a Rotary Wing Brownout[12] program to provide complementary attitude

---

[12] Brownout occurs for rotary wings during low altitude hover, landing, and take-off. It is a result from dust being stirred up from the ground degrading visibility.

cues for continued operation in disorienting conditions (Albery, 2007). As described in Stanney et al. (2004), a system like this is leveraging on cross-modal design by optimizing several working memory subsystems (sensory types). It takes advantage of the multimodal couplings to achieve minimal interference and performance loss (e.g. Stanney et al., 2004).

As mentioned in the introduction of this chapter, many information systems focus on presenting information to one or two senses. At FOI we ran a project investigating aspects of multimodal presentation. This resulted in several experiments and of these especially two studies relate to the work in this thesis. More specifically it concerns experiments in a simulated fighter jet and in one real and one simulated combat vehicle. These experiments are further described under the headings Bimodal, and Trimodal experiments below.

Bimodal experiments

Experiment 6

Together with TNO Human Factors we had an ongoing cooperation regarding tactile displays. We wanted to investigate the usefulness of a tactile vest in a high performance fighter jet. One of the main questions was if the tactile stimuli could be perceived and interpreted by pilots during tasks requiring high G-loads. We performed a study in which nine fighter pilots detected and intercepted threats with G-loads up to +8−9Gz during threat intercept. This experiment relates to research questions three and five: - Can (3D audio and) visual threat indication be improved by adding a tactile presentation? - Can (3D audio and) tactile displays be used effectively and efficiently without extensive operator training?

Experiment 6 was presented as a conference paper at the Human Factors and Ergonomics Society 50th Annual Meeting (Eriksson et al., 2006) and as a research article in Aviation, Space and Environmental Medicine (van Erp et al., 2007).

Participants. Nine male fighter pilots from the Swedish Air Force participated. They had a mean age of 37 years (range: 29 – 53). After giving their consent to participate they went through a brief medical examination.

Apparatus. The experiment was conducted in the Swedish Dynamic Flight Simulator (DFS). The DFS is an advanced flight simulator combined with a human centrifuge. The pilot is placed in the 3 m diameter gondola and ‘flies’ the DFS like a regular aircraft (FMV, 2009). We used simulated head-up display (HUD[13]), head-down displays (HDDs), and out the window (OTW) displays. For a more detailed description of the apparatus please see Eriksson et al. (2006) and van Erp et al. (2007). Tactile signals were presented with the TNO Tactile Torso Display consisting of 60 vibrators in a matrix (5 rings and 12 columns) covering the torso of the pilot. See figure 21 below.

Design and Stimuli.
The experiment used a 2 × 2 × 2 factorial within subject design, and it included:

---

[13] Head-Up Display is a visual display presenting information overlaid on the outside view.

- Two cueing conditions: visual only and visual + tactile
- Two mental workload conditions: low and high
- Two threat sectors: front and back

Each of the eight experimental conditions was repeated four times, resulting in 32 threat interceptions for each pilot. Pilots performed interceptions leading to G peaks between +3.2Gz and +9Gz, with all pilots having a maximum peak at about +8-9Gz. The means of each pilot's 32 +Gz peaks ranged from +5.9 to +8.0Gz.

In the experiment HUD and HDD symbology (visual) presentation was compared to HUD, HDD and tactile symbology in combination (visual + tactile). A continuous memory task (CMT)[14] was used to induce a higher workload and in the low workload condition the CMT was removed. We hypothesised that the 3D character of the tactile display would be especially beneficial for chasing targets that popped up behind the pilot. We therefore introduced an independent variable of position; in front of and behind the aircraft (Note that threats were never visible in the out-the-window displays). Reaction time was defined as the time between target pop-up and the pilot's response to roll the aircraft 10° from straight and level flight. Chase time was the total time between target pop-up and a successful interception of getting the threat within a 10° vertical by 15° horizontal area in the HUD centre.

Figure 21. To the left, a pilot prior to the experiment. The pilot is equipped with the necessary flight equipment, tactile vest and electrodes for physiological measurements. To the right, the inside of the TNO tactile vest showing the tactors as housed in small white cubes.

Results. An ANOVA of the reaction times showed a significant main effect of cueing condition, $F(1, 8) = 7.48$, $p < .05$, with no other significant effects. The visual displays combined with the tactile display generated shorter threat reaction times as compared to the visual displays alone; adding the tactile cueing reduced the mean reaction time from 1458 ms (54) to 1245 ms (88).

---

[14] In the CMT the pilot had to remember the first and second time one of four target alphabet letters was presented among various non-target letters. The letters were presented through the audio system of the DFS. For more details please refer to Eriksson et al. (2006) and van Erp et al. (2007).

The ANOVA of chase time showed a significant main effect of Threat position, *F*(1, 8) = 170.63, *p*< .01, and a significant interaction effect between cueing condition and threat position, *F*(1, 8) = 5.59, *p*<.05. A post hoc Tukey test showed that this interaction effect related to targets that popped up behind the pilot's aircraft in which the mean chase time was reduced from 13 s (0.45) with the visual displays to 12 s (0.41) with the tactile display added (with no significant difference between the cueing conditions for the targets presented in the front sector).

Performance measures were complemented by subjective ratings. In the subjective ratings we found that the pilots preferred the tactile display over the visual displays at target pop-up. The reason was that it was easy to detect and interpret targets' initial position. See figure 22 below.

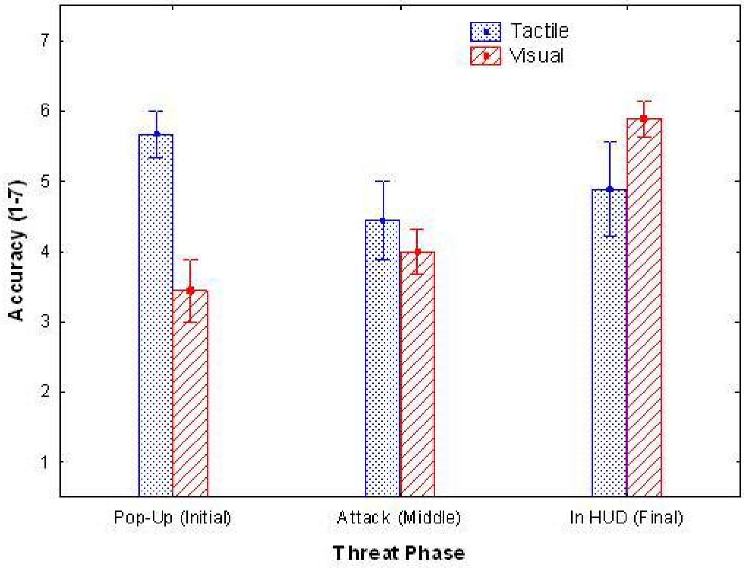


Figure 22. Subjective ratings for perceived threat location accuracy with the tactile and visual displays. Graph shows pilots' estimations at the phases of threat pop-up (initial), attack (middle) and when threat is centred in the HUD (final). Means + SE. The Graph is reused from Eriksson et al., 2006, with permission from SAGE (http://pro.sagepub.com/content/50/16/1547.short).

Discussion and summary. This was the first study to show that the information from a tactile display can be perceived and useful during high G-loads. The tactile display captured the pilot's attention at threat pop-up and improved reaction time. The tactile presentation also improved total chase time when threats appeared behind the pilot's aircraft. Important is that these improvements occurred even in the presence of high-end visual displays, which the pilots are accustomed to, and only limited training with the tactile vest was administered before the experiment. Furthermore, it is likely that the pilots' performance with the tactile display would further increase during training and use. In summary, tactile signals for threat cueing could enhance the threat awareness in fighter aircraft (Eriksson et al., 2006; van Erp et al., 2007).

The CV 90 experiments

In a more extensive study, we focused on improving the warnings from the WCS in the CV90 by adding spatial audio and tactile information. The light armoured combat vehicle CV90 has sensors with several functions such as missile approach warning, laser warning and optic surveillance. With sensors and countermeasures brought together by the WCS, various countermeasures can be automatically launched and critical time saved in responding to immediate threats. For information presentation, the CV90 is equipped with a 2D visual HDD for cueing of directions to threats, mono sound for threat alerts, and a mono radio system for communication. It could be an advantage, and in some cases even crucial, that the driver of a CV90 is alert for and can better manage threat warnings with regard to directions to threats. An improvement of the WCS threat cueing might be facilitated by integrating displays utilizing several senses (i.e. a multisensory display) to enhance detection and attention. To accomplish a threat warning that can be intuitively and quickly attended to, we hypothesized that the usefulness of the CV90 WCS could be improved by bimodal displays taking better advantage of the spatial dimension in an intuitive manner.

A series of three experiments (7, 8 and 9) investigated whether tactile and 3D audio technologies could effectively convey directional threat information to the driver of a CV90. Experiment 7 was performed in the combat vehicle with operative combat vehicle drivers. This was carried out to investigate the performance of the display technologies in a setting as realistic as possible.

Experiment 8 was a replication of experiment 7, performed in a simulated setting but with real operators. For experiment 7 and 8 we hypothesised that there might be benefits of the tactile display over the 3D audio display due to the noise levels in the vehicle and the lack of a head-tracker system to compensate for head movements.

Experiment 9 compared the 3D audio and the combined tactile and 3D audio with head-tracked 3D audio presentation in an experimental setup otherwise similar to experiment 8. The experimental series concerning the CV90 ties to the overall goal of improving information presentation by auditory and tactile information, and also to the long-term objective by contributing to the development of a more intuitive interface. The experimental series attempts to answer the below research questions (1, 3, 4, and 5):

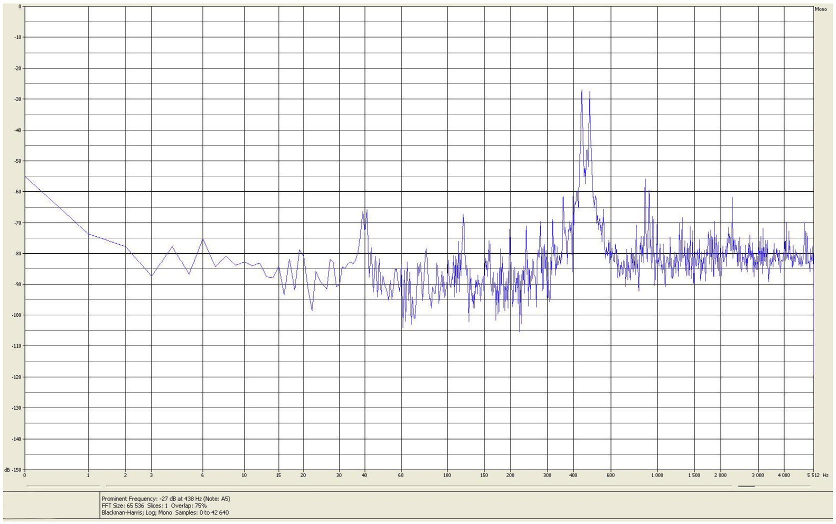
- Can a portable 3D audio system generate spatial audio that is accurate enough for operational use?
- Can 3D audio (and visual) threat indication be improved by adding a tactile presentation?
- What are the main benefits of auditory, tactile and multimodal threat cueing in a combat vehicle?
- Can 3D audio and tactile displays be used effectively and efficiently without extensive operator training?

Experiment 7

Experiment 7 was presented as a conference paper at the Human Factors and Ergonomics Society 50th Annual Meeting (Carlander & Eriksson, 2006).

Participants. Ten male CV90 drivers from the Swedish Army Combat School participated. They were all inexperienced with using 3D audio and tactile displays. Mean age was 19.5 years (range: 19 to 21).

Apparatus. A laptop PC equipped with an external USB soundcard, Hercules Gamesurround MUSE Pocket, and the FOI tactile belt (see figure 26) were used for threat cueing. The technology used for the COTS soundcard utilized the Sensaura algorithm. This incorporates a HRTF library derived from a series of measurements with a dummy-head in an anechoic chamber. The HRTFs are optimized for localisation accuracy and simplified to ensure real time processing in a sound card processor. The accuracy of the system has been verified in previous studies (experiments 1 to 3). The auditory stimulus was a signal normally used for threat warnings in the vehicle, a pulsating sound with 446 ms on followed by 461 ms silence, with the prominent frequency of 438 Hz (see Figure 23 below). Auditory stimuli was presented through a pair of Sony in-ear headphones (not equalized) and covered by the helmet ear protection.



Prominent Frequency: -27 dB at 438 Hz (Note: A5)
FFT Size: 65 536 Slices: 1 Overlap: 75%
Blackman-Harris; Log; Mono Samples: 0 to 42 640

Figure 23. Frequency spectra for the auditory signal used in the experiment. This was a pulsating sound (446ms on separated by silence of 461ms) with a prominent frequency 438 Hz.

Note that the sound position was “vehicle tracked”, which meant that the audio stimulus position was compensated for vehicle heading. When the vehicle rotated the audio stimulus position remained at the simulated position, and if the participant moved the head the audio stimulus position would follow the head since audio presentation was not compensated for head movements. The CV90 used in the experiment is shown in figure 24 below.

Figure 24. The Combat Vehicle 90 (CV 90) technology demonstrator as used in the WCS experiment (Picture credit to Jan Fredriksson, MSS Skövde, 2005).

The tactile belt was controlled from a small signal box connected to the parallel port on the laptop PC, which in turn was fed with simulated data from the actual CV90 WCS. Experimental apparatus was connected to the system and supervised from the troop transportation area in the vehicle, see the illustration in figure 25 below.

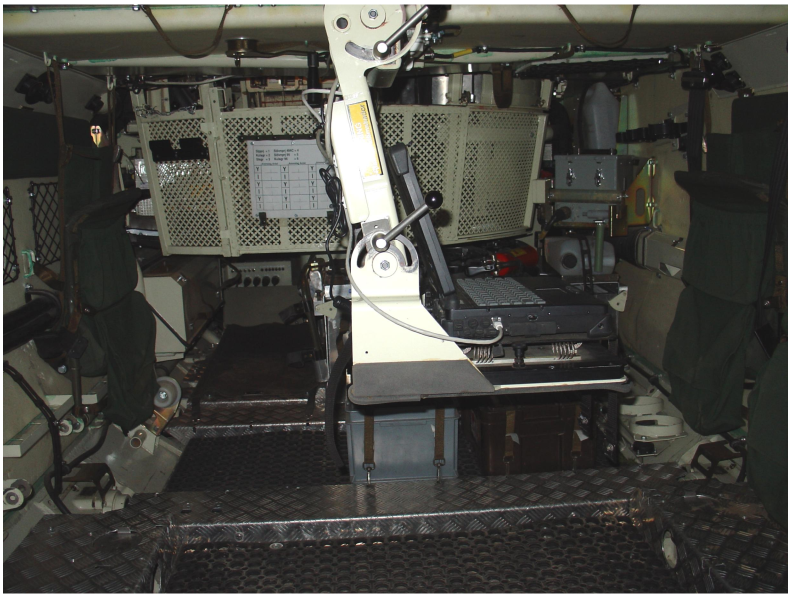


Figure 25. Front section of the troop transportation area of the CV90. The table with the laptop seen in the center of the picture was used during the experiment. Behind the white "fence" the gunner, commander and driver are positioned.

The tactors in the tactile belt were cylindrical, 14 mm long with a diameter of 6 mm, with a weight of approximately 10g and a frequency of ~120Hz (see Figures 15 and 26A). The cylindrical tactors were inserted into a belt with pockets, and the belt was then strapped, pressing firmly to the skin around the torso. The tactile pattern was 90 on 70 off, 100 on 20 off, 100 on and 120 ms off. We used this pattern to minimize the risk of tactile adaptation and the minimum of 90 ms "on" was to ensure that tactors reached the maximum frequency. Tactors were evenly distributed as a ring around the torso covering the full 360° degrees in 30° sectors. Figure 26 shows the main components of the FOI tactile torso belt.

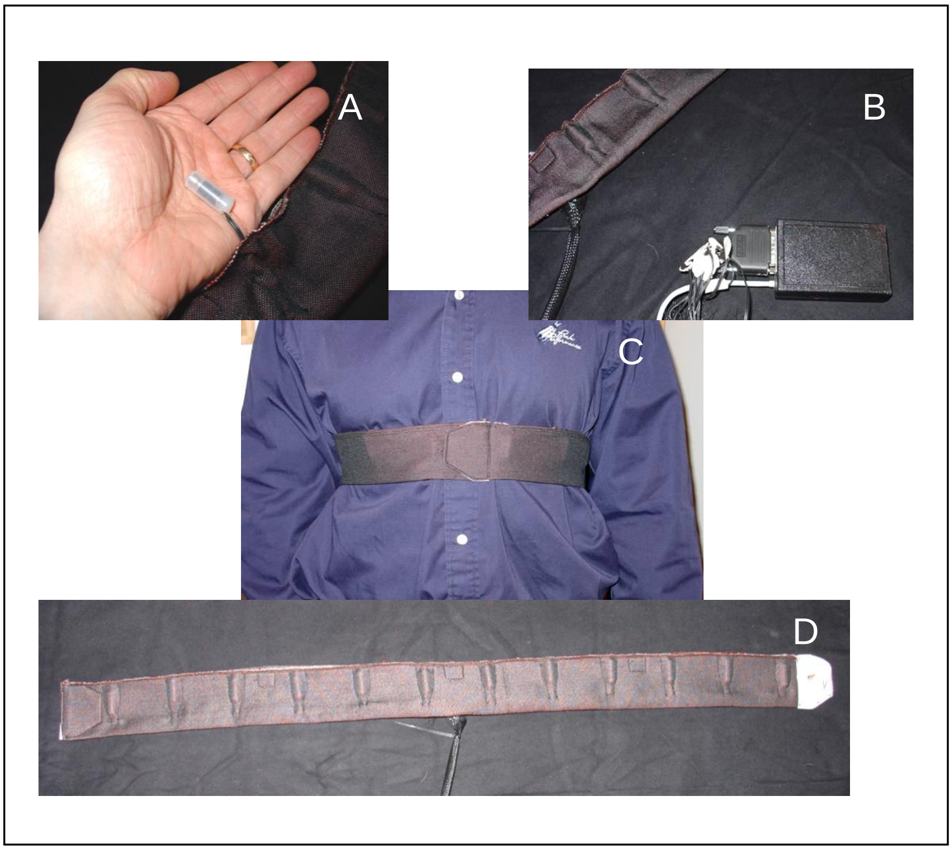


Figure 26. The FOI tactile torso belt as used in the experiment. A, an encapsulated vibrotactile rotary motor. B, the signal control box. C, The tactile torso belt when fitted. D, the tactile torso belt extended.

Design and Stimuli. The experiment had a 3 × 5 factorial within subject design. It included:

- Three display conditions of 3D audio, tactile, and the combination of 3D audio and tactile.
- Five threat positions presented at front left (300°), front right (60°), back right (120°), back left (240°), and straight behind (180°) relative initial vehicle heading.

The task was to react to each warning and indicate the direction to the threat by as quickly and accurate as possible turn the CV90 towards the threat, thus aligning the vehicle heading with the direction to the threat. Each warning had a maximum duration of 20 s or was terminated if the vehicle heading was maintained within ±7.5° of threat position for five seconds. The interval between threat presentations was 20 s. Localisation error (LE) was defined as the vehicle heading's horizontal deviation from correct threat position at warning termination. Response time (RT) was defined as the elapsed time from target pop-up[15] to vehicle heading deviating ±10° from initial heading. Each display condition

---

[15] No visual targets were used; "target pop-up" refers to sound or tactile stimulus presentation, which was always presented relative initial vehicle heading.

represented one block of trials and included three trials of each of the five threat positions, resulting in a total of 45 threat presentations per participant. The presentation order of threat positions was randomised within each block of trials, and the presentation order of blocks was balanced over participants.

Procedure. The participant was instructed to “rotate” the vehicle heading towards the presented threat. When a threat was presented, the main task was to turn the vehicle as quickly as possible toward the perceived threat direction aligning vehicle heading with threat position. After termination of the threat, the driver waited for the next threat to be presented. The brief introduction included task description, adjustment of the displays and basic instructions. A short familiarisation session then followed with one presentation per display condition. Then the three experimental blocks of trials followed and after each block, a short questionnaire was completed (not reported here). For each participant the experimental session lasted about 1h.

Results. Repeated measures ANOVAs were applied to the means of LEs and RTs of each experimental condition. Each analysis included fifteen means (3 display conditions × 5 threat positions = 15 conditions), with each mean calculated from the three trials of each condition (Greenhouse-Geisser corrected *p* in ANOVAs.). A total of 8 front-back confusions occurred during the experiment spread over four of the participants. These data were treated as outliers and not included in the calculation of the condition mean for these four participants (removal to fulfil assumptions for ANOVA). Front-back confusions in relation to the number of presentations for these participants were 6.7%, 6.7%, 2.2%, 2.2%. Front-back confusions only occurred for the 3D audio condition.

Localisation error in degrees (LE).
The ANOVA showed a significant main effect of display condition, $F(2, 18) = 6.29$, $p< .025$, with no other significant effects. A Tukey HSD test revealed larger mean LE with the 3D audio, 9.6° (2.2), compared with the tactile display, 6.7° (0.5)($p< .025$), and the 3D audio/tactile display combination, 7.1° (0.8)($p< .05$).

Response time in ms (RT).
The ANOVA showed a significant main effect of display condition, $F(2, 18) = 4.16$, $p< .05$, and an interaction effect of display by threat position, $F(8, 72) = 2.42$, $p< .025$, with no other significant effects. Because of violations of the sphericity assumption in both we used the Greenhouse-Geisser corrected p-values, and both effects became non-significant (display: $p= .07$, and display by threat position: $p= .15$.).

Summary. The 3D audio display generated greater LEs compared to the tactile and combined tactile+3D audio display. Thus, the LE with the 3D audio was still greater than with the other two conditions despite the fact that the front-back confusions were removed before statistical analysis. Nevertheless, the 8 front-back confusion occurrences that appeared for the 3D audio display should not be overlooked because they may be critical depending on type of application. However, adding the tactile cueing resolved the front-back confusion.

Experiment 8

Experiment 8 replicated experiment 7 but was instead carried out in a laboratory environment with a simulated CV90 WCS system. We were interested to see if the performance of the displays would differ between the rather noisy environment in the CV90 compared to the quiet environment in the laboratory. The main difference, apart from the environmental conditions, was that the simulated vehicle was easier to control and thus allowing for a more precise "fine-tuning" of the vehicle heading.

Participants. Twenty males of operative CV90 crews from the Swedish Army Combat School participated. Ages ranged from 19 to 21 years with a mean age of 19.5 years. They were all inexperienced with using 3D audio and tactile displays.

Apparatus. In the laboratory, participants were seated with a game steering wheel in front of them. No visual display was used and no feedback administered after threat indication. As in the real platform the tactile belt was controlled from a small signal box connected to the parallel port on the laptop PC. Instead of receiving data from the warning system in the vehicle, a simulated WCS system provided data for threat positioning. The laptop was also equipped with an external USB soundcard, Hercules Gamesurround MUSE Pocket (same as used in the vehicle for experiment 7). This was used for the audio playback presented through a pair of Sehnheiser HD 200 headphones with circumaural design. The simulated WCS system was basically a software version of the real sensor system in use in the CV90, programmed with parameters to mimic the performance of the real platform. The simulated WCS system fed a server with software for controlling the 3D audio and tactile presentation. The simulated vehicle was controlled by the computer gaming steering wheel and the simulated vehicle rotation was accomplished by turning the gaming steering that in turn fed the simulated WCS system with data for updating threat position in relation to the vehicle. The speed of the rotation was approximately that of the CV90.

Design and Stimuli. The design and stimuli were identical to experiment 7 with a 3 × 5 factorial within subject design. Again the task was to react to each warning and indicate the direction to the threat by as quickly as possible turn the simulated vehicle towards the threat. Threat defeat, LE and RT were treated as in experiment 7 above. Threat onset positions were presented relative to the initial vehicle heading.

Procedure. The procedure was the same as in experiment 7.

Results. Repeated measures ANOVAs were applied to the means of LEs and RTs of each experimental condition. Each analysis included fifteen means (3 display conditions × 5 threat positions = 15 conditions), each calculated from the three trials of each condition. The sphericity assumption was violated in all the analyses and the reported *p*-values are with Greenhouse-Geisser correction. A total of 22 front-back confusions occurred during the experiment, which were made by 10 of the participants. These data were treated as outliers and not included in the calculation of the condition mean for these 10 participants (removal to fulfil assumptions for ANOVA). The average percentage of front-back confusions, in relation to the total number of presentations, for these participants was 4.4%, ranging from 2.2% to 6.7%. Front-back confusions only occurred for the 3D audio condition.

Localisation error in degrees (LE).
The ANOVA showed a significant main effect of threat position, $F(4, 76) = 5.03$, $p < .025$, with no other significant effects. Mean LEs ranged from 5.8° to 9.7° over the threat positions, and a Tukey HSD test revealed significantly greater mean LE with the threat

position at 180°, 9.7° (5.2), compared with threat position at 240°, 5.8° (1.8) (*p*< .01), and threat position at 300°, 6.1° (2.4)(*p*< .01).

Response time in ms (RT).

The ANOVA showed significant main effects of display, *F*(2, 38) = 14.55, *p*< .01, and position, *F*(4, 76) = 12.76, *p*< .01, and a significant interaction effect of display by threat position, *F*(8, 152) = 8.96, *p*< .01. The interaction effect of display by threat position is illustrated in figure 27. A Tukey HSD test showed greater RTs for the 3D audio at 180°, 2881ms (571) (p< .0001), compared with all other positions ranging from 942 to 1144ms (20-42).

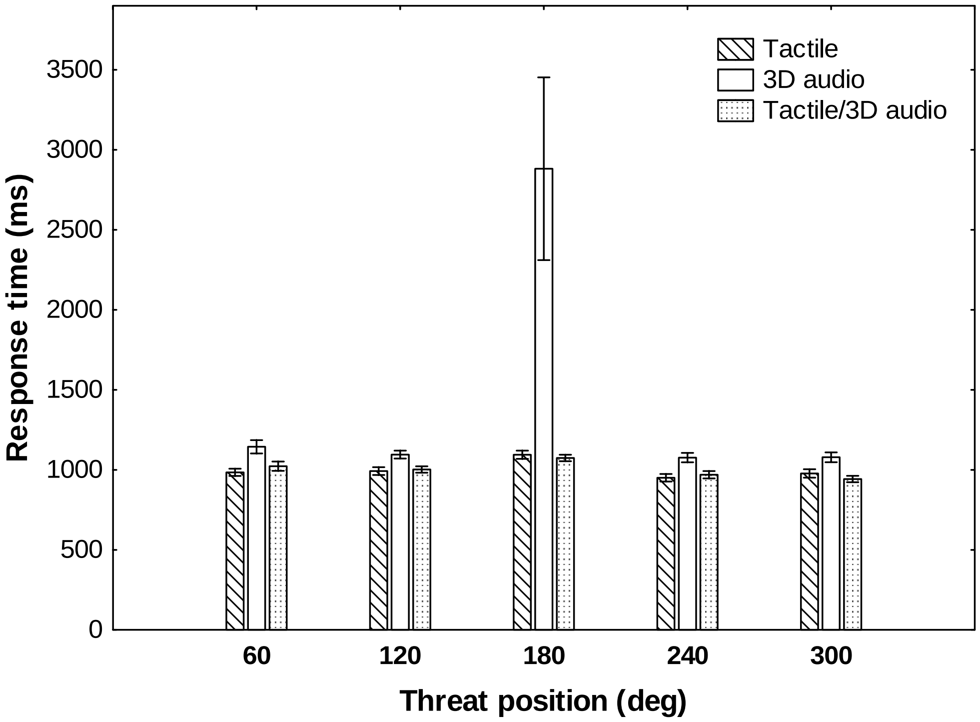


Figure 27. RT interaction effect of display by threat position, F(8, 152) = 8.96, p< .01.

Summary. For localisation error only a significant main effect for threat position was found and this had to do with greater mean LE for the threat position at 180° compared with threat position at 240° and 300°. The 22 front-back confusions is a serious draw-back for the 3D audio condition and the percentage of confusions ranged from 2.2% to 6.7% for ten of the participants. For the response times, significant main effects were found for display and position. Also a significant interaction effect of display by threat position was found as relating to greater RTs for the 3D audio at 180° compared with all other positions.

Experiment 9

To further investigate the relationship between 3D audio, head movements and tactile presentation, experiment 9 was conducted. We hypothesized that a tactile presentation could be as beneficial as a head-tracker for improving the 3D audio presentation with regards to front back confusions. This could have important implications for the design of intuitive display systems since the 3D audio presentation most likely needs a head-tracker, something that might not be possible in all environments.

Participants. Ten participants with the mean age of 35.5 years (range: 25 to 58 years) volunteered to participate. Participants had no prior knowledge about using 3D audio and tactile displays for threat cueing.

Apparatus. The apparatus was the same as for experiment 8.

Design and Stimuli. The experiment had a 3 × 6 factorial within subject design. It included:

- Three display conditions: 3D audio, 3D Audio + head-tracker, and 3D audio + tactile cueing (bimodal condition)
- Six threat positions: front (0º), front right (60º), back right (120º), straight back (180º) back left (240º), and front left (300º)

Display conditions were chosen to see if the tactile cueing condition would be as beneficial as a head-tracker to reduce the occurrence of front-back confusions. Threat positions were presented relative initial vehicle heading. A difference compared to experiments 7 and 8 was that the straight-ahead, 0 degrees position was added. This position was used to ensure that participants would experience presentations as coming from both straight ahead and behind the vehicle. The main task was identical to the previous two experiments. Each display condition represented one block of trials and included five trials of each of the six threat positions, resulting in a total of 90 threat presentations per participant. No questionnaires were administered.

Procedure. After a brief introduction, the displays were fitted, and the participant read written instructions followed by verbal. A short training session was then performed before commencing the three experimental blocks of trials. The training session consisted of five presentations per display condition. A full experimental session lasted about 1h per participant.

Results. Repeated measures ANOVAs were applied to the means of LEs and RTs of each experimental condition. Each analysis included 15 means (3 display conditions × 5 threat positions = 15 conditions), with each mean calculated from the five trials of each condition. The straight ahead position (0º) was mainly used as a reference[16] and discarded for the analysis. A total of 18 front-back confusions occurred during the experiment spread over six of the participants. These data were treated as outliers and not included in the calculation of the condition mean for these six participants (removal to fulfil assumptions for ANOVA). Fifteen of the front-back confusions occurred for the 3D audio condition with an average of 6.3% ranging from 3.3% to 10.0%. The remaining three occurred for the 3D audio + head-tracker condition all averaging to 3.3%.

Localisation error in degrees (LE).
The ANOVA showed no significant effects. There was only a tendency of a main effect of display condition, $F(2, 22) = 3.27$, $p= .06$, and mean LE with the 3D audio, 3D audio

---

[16] Since sounds were presented relative to vehicle heading, the “straight a head” position would render localisation accuracy and response times close to perfect without having a real response from the participant. It was however important to use this position to make sure sounds were presented in front of the participant to capture eventual front back confusion issues.

with head-tracker, and 3D audio/tactile display combination was 9.3° (4.2), 7.7° (2.7), and 5.7° (0.9), respectively.

Response time in ms (RT).

The ANOVA showed significant main effects of display, *F*(2, 22) = 6.84, *p*< .01, and position *F*(4, 4) = 13.84, *p*< .0001, and a significant interaction effect of display by threat position, *F*(8, 88) = 3.12, *p*< .01. Due to violated sphericity assumption, the main effect of position and the interaction effect of display by threat position were corrected with the Greenhouse-Geisser p-values: *p*< .01 and *p*= .08, respectively.

A Tukey HSD test showed shorter RTs for the bimodal condition compared with the head-tracked 3D audio, 1282ms (68) and 1655ms (33), respectively (*p*< .01). This is shown in figure 28. Another Tukey HSD test showed that presentations at the 180° position resulted in significantly longer RTs (1809ms (142) compared with presentations at 60° 1440ms (85), 120° 1367ms (72), 240° 1385ms (85), and 300° 1458ms (92) (*p*< .0001, for all comparisons).

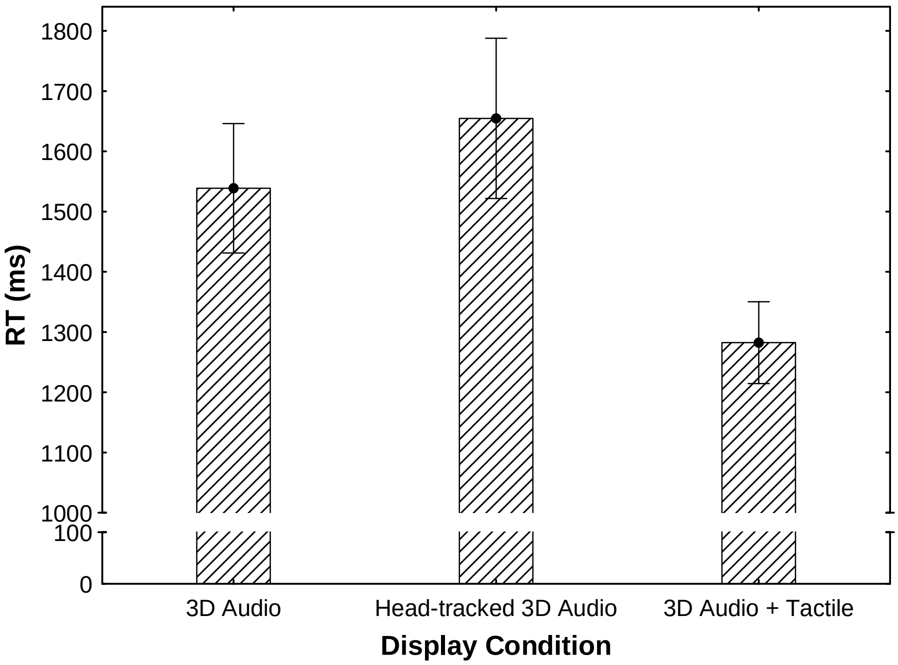


Figure 28. Main effect of display on RT, *F*(2, 22) = 6.84, *p*< .01.

Summary and Discussion. No significant differences were found for localisation error. 18 front-back confusions occurred, 15 occurred for the 3D audio condition (ranging from 3.3% to 10.0% over participants), and the remaining three occurred for the 3D audio + head-tracker condition (averaging to 3.3%). Tactile cueing resolved the front-back confusions.

The analysis of response times showed significant main effects of display and position. There were shorter RTs for the bimodal condition compared with the head-tracked 3D audio. Presentations at the 180° position resulted in significantly longer RTs compared with presentations at 60°, 120°, 240°, and 300°.

The overall results of the three experiments show that combining sensory modalities can improve threat localisation performance. Even though experiments slightly differ with regard to environment and participants, results are robust. The 3D-audio needs improvement with regard to front-back confusion if not used with tactile cueing. As expected, the tactile presentation results in slightly better performance and resolves the front-back confusions.

To misinterpret a directional warning signal is of course a serious drawback for the 3D audio display and the display type alone is not suitable for directional warning signals in the front-back dimension. Perhaps we could have seen a slight increase in performance for the 3D audio display if the auditory stimulus, the frequency content of the CV90 warning signal, would have been better adjusted for spatialisation (please refer to "localisation" in chapter 2).

One explanation for the low error rates of the tactile and bimodal conditions could be that we used tactor locations close to so called anchor points. These are physical locations where a vibrotactile stimulus can be perceived with high accuracy (as described in the introduction of chapter 3). Although participants were supposed to get the threat in front of the vehicle, at the 12 o'clock tactor, the information provided by the other tactor locations was important regarding the initial perception of the threat location. Furthermore, the tactors provided a sense of rotational speed that most likely contributed to the speed and accuracy of getting the target in front of the vehicle. In other words, both the spatial and temporal aspects of the tactile cueing of external threats were used by the participants.

Coupled to the resolution of the display systems are the constraints of the spatial and temporal resolution in the sensor and warning systems. These constraints of course affect the accuracy and timing of threats when mapped to vibrotactile and/or 3D audio cueing. For the CV90 experiments the WCS systems ran in a simulated mode where the spatial and temporal resolution was close to perfect. Thus, LEs and RTs of these studies relate to the performance of the simulated system. For implementation in operational systems there might be delays and further uncertainties due to system capabilities (e.g. sensor accuracy, audio quality of speech etc.).

Compared with the currently employed auditory and visual threat cueing in the combat vehicle CV90, tactile and 3D audio technologies seem to be superior for immediate detection of horizontal threat directions. Even though not directly compared with the current display technologies, these new display technologies are considered superior because they indicate directions in a more direct, natural, and intuitive manner. This is also supported by the short training that was administered to the drivers that still had an excellent overall performance in reacting to and positioning the vehicle toward threats. This is also in line with a similar experiment performed by Spence and Ho (2008) where drivers performed well even though a minimal period of familiarization prior to testing was used. It should be pointed out however, that this study did not include combat vehicles and the task was slightly different.

These displays, that could be considered more intuitive, would probably result in an overall better perception and performance in the combat vehicle. However, the time available for action following a WCS warning is in reality extremely short (see the example in the introduction). To make full use of the potential of these display systems the WCS warnings could be sent to nearby CV90s or battle group crews. This way, operators that are not directly threatened can get early warnings and support the crew under attack. This approach would enhance the effectiveness of the WCS network by speeding up threat awareness and tactical behaviour over a battle group of CV90 crews.

Trimodal experiment

Experiment 10

Experiment 10 is one of two experiments that were presented in a research article in Human Factors (Oskarsson et al., 2012). It was also presented at the conference HCI International 2007 as a conference paper (Carlander et al., 2007). We hypothesized that there might be further benefits of the multimodal display because potentially, the displays have different advantages in different phases of threat direction cueing and tactical manoeuvring. Furthermore, the multimodal display may have benefits regarding alerting the operator and speeding up responses, and might be especially useful during heightened perceptual and/or cognitive load. For experiment 10 we investigated cueing of threat direction by a multimodal display consisting of HDD, 3-D audio and tactile belt as compared with each of the unimodal counterparts (for more details on the experimental series see Oskarsson et al., 2012).

In experiments 7, 8 and 9 we showed that the driver of a CV90 could efficiently utilize the threat cueing presented by 3D audio and tactile displays. In experiment 10 we added a visual presentation (HDD) and the task of handling incoming radio calls more or less simultaneously with the threats. The experiment attempt to answer the research questions one, three and four:

- Can a portable 3D audio system generate spatial audio that is accurate enough for operational use?
- Can 3D audio and visual threat indication be improved by adding a tactile presentation?
- What are the benefits of auditory, tactile and multimodal threat cueing in a combat vehicle?

Participants. Twelve male students with a mean age of 23.0 years (range: from 21 to 29 years) participated. All reported normal sight and hearing and had no prior experience of the presentation technologies.

Apparatus. The apparatus for tactile and audio presentation was identical to that described in Experiment 9, with the exception that the frequency spectra of the 3D sound was adjusted to be easier to localise. In addition, HDD was used for the 2D visual threat warnings, and a 42” plasma screen was used for the simulated terrain environment and for overlaid feedback regarding threat defeat status. To the right of the participant was a touch screen presenting response options relating to the identification of the radio calls. A further addition was an Ascension LaserBIRD II head-tracker that was used to compensate for head movements during the 3D audio presentation. For the vehicle simulation, a 6 DOF Moog motion platform was used, as shown in figure 29. The vehicle simulation was handled from an FOI developed simulation engine. The simulation engine sent data to a server handling threat information and presentation, and also data to the motion platform for exact simulation coordinates.

Figure 29. The FOI “CV90 simulator” based on a motion platform with 6 degrees of freedom. Screens for presenting the HDD information and handling radio calls are seen on the lower left and right of the participant. Vehicle information and out-the-window view is displayed on the large screen in front of the participant.

Design and Stimuli. The experiment had a 4 × 3 factorial within subjects design. It included:

- Four threat display conditions: HDD (visual display), 3D audio, tactile, and multimodal cueing
- Three sectors: front, the sides and back

During the visual and tactile threat indications, an alerting mono sound was also presented. The alerting sound was spatialised for the unimodal 3D audio and the multimodal conditions. The 3D audio presentation thus alerted and indicated direction to the threat. In the HDD and tactile displays, the mono sound had an alerting function only. Thus, the conditions including the HDD and the tactile belt were both bimodal in terms of warning information but unimodal in terms of threat direction information. The reason for keeping the sound presentation was that it is not likely that threat warnings will exclude sound presentation in the combat vehicle. Even though the alerting property of sound was kept for all display conditions, the cueing of threat directions was nevertheless unimodal in three of the display conditions. Threats occurred in three sectors presented at the front (315º-45º), the sides (45º-135º and 225º-315º), and the back (135º-225º), relative initial vehicle heading. As previously, the task was to react and respond to each threat as quickly as possible by aligning the vehicle heading with the threat position. Threat position was confirmed by pushing a trigger on the steering wheel. The response was considered correct if the vehicle was aligned ±10º within the threat position.

The HDD presentation was shown with a top-view of the vehicle where threat direction was indicated by a red cone (15 degrees wide) extending from the centre of the own vehicle. The threats were not visible, but after each threat response a text message “hit” or “miss” was overlaid on the simulated out-the-window view. Threat presentations were limited to 20 s or were terminated with the correct operator response. The time between threat response and the presentation of a new threat was randomized in the interval of 15 to 20 s. Vehicle heading deviation from the correct threat position at operator response was defined as localisation error (LE). The response time (RT) was defined as the elapsed time from target pop-up to having the vehicle rotated ±10º from initial heading.

The participants performed a secondary task during each threat pop up. In this task the driver had to identify the information in two simultaneously presented radio calls. Each call was initiated with an identifier ("-internal!" or "-external!") which was then followed by the name of a colour. Each colour had a corresponding button on a touch screen and the secondary task response thus involved auditory, visual, tactile, and motor systems. The proportion of correct responses to the radio calls was measured and defined as proportion correct radio calls. We also measured the time from radio call presentation to the completion of the radio call response form. We defined this as the performance time (PT) to identification of Radio Calls. During HDD and tactile display presentations radio calls were presented in mono. For the 3D audio and multimodal presentations radio calls were positioned horizontally at –22º and +22º.

One experimental block was defined as one of the display conditions, including eight threat presentation trials in each of the three threat sectors, with a total of 24 trials per block. Within each block of trials the presentation order of threats was randomized, and block order was balanced over participants. Questionnaires were administered before, during and after the experiment. The questionnaire before the experiment included a check on the physical requirements such as normal sight, hearing, and present health condition. Between blocks of trials a questionnaire concerning the display technology was completed, including perceived interpretability, mental workload, and perception of threat direction. At the very end of the experiment a more extensive questionnaire was completed. The questions were answered on a seven points scale (1 = very low, 7 = very high).

Results. Twelve means were used for each analysis (4 display conditions × 3 sectors = 12 means) calculated from the eight trials of each condition. Repeated measures ANOVAs were applied to the means of LEs and RTs of the threat indications, and to the means of PTs and proportion of correctly answered radio calls. In one case there was a violation of the sphericity assumption in the ANOVA. Here, the Greenhouse-Geisser-corrected *p*-value is given and indicated by (G).

A total of six front-back confusions occurred during the experiment spread over three of the participants. These data were treated as outliers and not included in the calculation of the condition mean for these three participants (removal to fulfil assumptions for ANOVA). Front-back confusions in relation to the number of presentations for these participants were; 4.2%, 8.3%, and 12.5% respectively. All occurred for the 3D audio condition.

Localisation error in degrees (LE).
The ANOVA showed a significant main effect of display condition, F(3, 33) = 61.23, $p < .001$, with no other significant effects. A Tukey HSD post hoc test showed larger mean LE for both the 3D audio and tactile displays compared with both the HDD and the trimodal display ($p < .001$ for all comparisons). See figure 30 below.

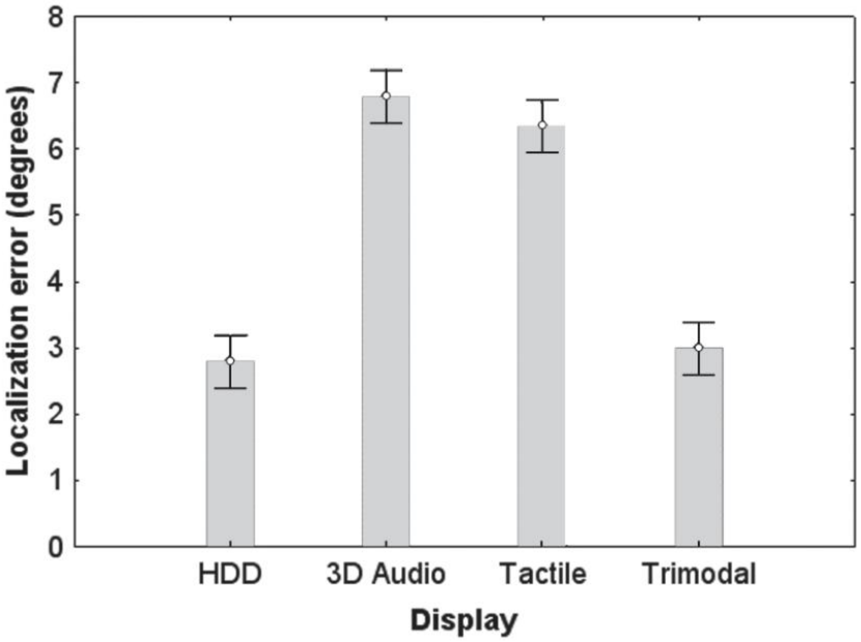


Figure 30. Mean localisation error (LE) with each display condition (±SE), main effect of display condition, F(3, 33) = 61.23, p < .001,

Response time in ms (RT).

The ANOVA showed a significant main effect of display condition, F(3, 33) = 7.93, *p* < .001, with no other significant effects (figure 31). A Tukey HSD post hoc test showed a greater mean RT for both the HDD and the 3D audio display as compared with the tactile and the multimodal presentation respectively (*p*< .025 for all comparisons).

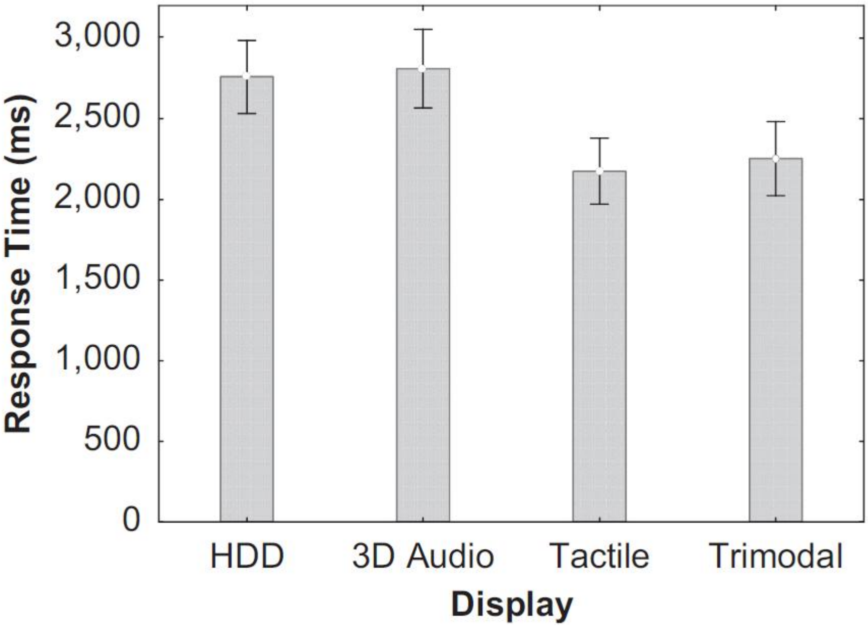


Figure 31. Mean response time (RT) to threats with each display condition (±SE).

Proportion Correctly Identified Radio Calls.

The ANOVA showed no significant effects. The overall mean for proportion correct responses to radio calls was .83.

Subjective Ratings. Subjective ratings of initial perception of threat direction (PTD) and mental workload were made on a seven point scale; 1 = not at all obvious/very low, 7 = very obvious/very high. For the PTD a repeated measures ANOVA was applied (4 displays × 3 threat phases), and the mental workload and effort of driving were analyzed with ANOVA for a one-way repeated measures design (four displays).

Perception of threat direction (PTD).
The threat phases were initial (threat cueing onset, first response), middle (during manoeuvring the vehicle towards the threat), and final (aligning vehicle heading with the threat and pushing the trigger). There were a significant main effect of display, F(3, 33) = 23.46, *p*< .001, and a significant interaction effect of display by threat phase, F(6, 66) = 6.58, *p*< .01(G). Post hoc testing of the interaction effect revealed the following significant differences of display configuration ratings within each threat phase (figure 32):

Initial:	Both tactile and trimodal displays were higher rated than both HDD and 3D audio displays (*p*< .001 for all comparisons).
Middle:	Both tactile and trimodal displays were higher rated than the 3D audio display (*p*< .01 for both comparisons).
Final:	The trimodal display was higher rated than both the tactile (*p*< .025) and 3D audio display (*p*< .001), and the HDD display higher rated than the 3D audio display (*p*< .01).

It was further shown that when contrasting the phases within each display technology the HDD display was significantly lower rated in its initial phase compared with the final phase (*p*< .001). The opposite was shown for the tactile display that was significantly higher rated in the initial phase compared to the final phase (*p*< .025). The post hoc testing of the main effect of display revealed higher ratings for the trimodal display compared to both the HDD and 3D audio displays (*p*< .01 for both comparisons). 3D audio ratings were lower compared to all other displays (*p*< .01 for all comparisons).

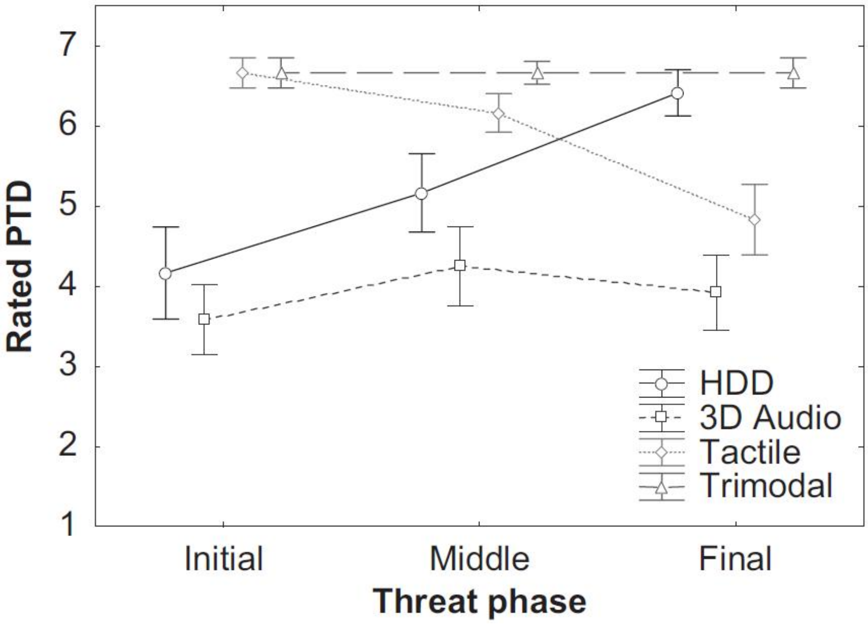


Figure 32. Ratings of perception of threat direction with the display conditions. Left in the graph, localisation phase (initial) at threat pop up. To the right in the graph, alignment phase (final) (1 = Not at all obvious, 7 = Very obvious) (±SE).

Mental workload.
No significant effects were found and the ratings of mental workload were moderate for all four displays (an overall mean of 4.0 on the 7-point rating scale).

Effort of driving.
No significant effects were found, and the ratings of effort of driving were low for all four displays (an overall mean of 2.3 on the 7-point rating scale).

Summary and Discussion.
The results show that the 3D audio and the tactile display generated greater LE compared to the HDD and multimodal display. Six front-back confusions occurred for the 3D audio condition spread over three of the participants. (4.2%, 8.3%, and 12.5% respectively).

The RT for 3D audio and HDD were longer compared to the tactile and the multimodal presentation, respectively. In the perceived threat direction we found differences between the phases of threat localisation. No differences were found between conditions regarding the identification of radio calls, mental effort or effort of driving. Although very little training was administered, participants performed well with the different presentation technologies.

The longer RTs for the visual display compared to the tactile and multimodal display indicate that the visual display is relatively poor in drawing operator attention, and this is also reflected in the subjective ratings as presented in figure 32 above. Ho, Spence and Tan (2005) also showed that participants respond more rapidly to tactile warning signals, than to either visual or auditory warning signals. However, the visual display will most likely outperform the tactile and 3D audio displays when higher resolution is needed such as that of detailed manoeuvring (requiring the higher spatial resolution of the visual sense). This is also supported by the results that show lower localization errors for the visual display as compared to the 3D audio and tactile displays respectively. Superimposing the visual threat cueing on the out-the-window view would improve the visual presentation since it would provide an easier and more intuitive access to the visual information. However, the resolution of the tactile display was relatively high (below 10 degrees) although each tactor only covered a 30 degree sector. This has most likely to do with the occurrence of movement that seems to improve localisation (Gilson et al., 2007; van Erp, 2002).

Conclusions. The main hypotheses were supported in that there are further benefits of the multimodal display in different phases of threat direction cueing and that the multimodal display has benefits regarding alerting the operator and speeding up responses.

Considering the overall performance, the multimodal display demonstrated important beneficial effects such as shorter RTs supported by the tactile display and lower LE’s supported by the HDD. The results clearly show the advantage of the enhanced alerting supported by the redundant tactile information when visual information is presented on a HDD, which is not constantly attended. The different phases in the threat scenario set different requirements on the operator depending on whether it was the initial phase (first threat perception), the middle phase (manoeuvring towards the threat), or the final phase (precision manoeuvring). With the ability to capture attention the tactile display improved the initial perception of threat direction and effectively supported vehicle manoeuvring.

Regarding the threat response performance with the 3D audio and tactile displays, the results are in line with the general results of experiments 7 to 9. Combining sensory modalities seems to increase threat localisation performance and decrease response times. The multisensory threat presentation improves threat warnings and can contribute to an increasing situational awareness and overall performance. The multimodal displays in experiment 10 capitalizes on the advantages of tactile presentation regarding capturing attention and quickly gaining threat awareness, and combines this with a visual display for the final precision manoeuvring. It thus shows how sensory redundant information can overcome display-specific limitations and the enhanced perception results in increased performance in the dynamical threat scenario.

The second experiment as described in Oskarsson et al., (2012) was in large a replication of experiment 10 (although this experiment is not reported as part of this thesis). The main changes for this second experiment were that the visual and tactile threat direction cueing (unimodal conditions) were replaced by bimodal conditions, and that the HDD was replaced by a HUD. Conditions were HUD with 3D audio, tactile with 3D audio, and HUD, 3D audio, and tactile belt combined into a trimodal display. It was shown that the combined HUD and 3D audio display (bimodal) generated LEs and RTs as good as the trimodal display. Although this is an indication that the combined HUD and 3D audio display is sufficient for the improved performance, the PTD was rated higher with the trimodal display in both the initial and middle phases. In sum, the trimodal display showed the most advantages in both experiments. Most likely this has to do with that perceptual discrepancies are resolved in favour of the more precise modality. As an example, the visual modality usually dominates in a spatial task because it is the most precise modality to determine spatial information (Ernst & Bülthoff, 2004). Spence and Ho (2008) also concludes that using several modalities for presentation can capture driver attention significantly more effectively than unimodal (i.e., auditory, tactile or visual) equivalents. These displays are more efficient in conveying information and to reduce driver workload.

Key-points Chapter 4

- **A certain sense can be interfered by glare, darkness, noise etc. In these situations we need to consider the use of other senses that may not be interfered from the same source (Brill et al., 2004).**
- **We are used to combine sensory modalities since specific stimuli are often made up by signals in several modalities and integrated into a coherent percept (Eimer, 2004; Ernst & Bülthoff, 2004).**
- **The CNS 'sorts' incoming signals and combine those likely to originate from the same source leading to reduced variance and enhanced chances of detecting a stimulus (e.g. Bresciani et al., 2005).**
- **The Multiple Resource Theory (MRT) that suggests we have independent resources for processing information, and not one single information processing source that can be tapped (Wickens, 2002).**
- **MRT can be used as a tool to recommend interaction designs for situations with multitasking (Wickens, 2002, 2008).**
- **Having multiple resources (modalities/senses) allow us to simultaneously process information from more than 'one channel' with the effect that several tasks can be performed more or less simultaneously with little interference (Wickens, 1987; Wickens & Hollands, 2000).**
- **Comparing two tasks that share resources (intra-modal) with tasks that use different resources (cross-modal) the latter will in most cases result in better performance (Wickens, 1987; Wickens & Hollands, 2000).**

Contribution to multimodal research

- **Suggestions based on experimental results for how multimodal threat information could be presented to the crew in the combat vehicle 90 (Carlander & Eriksson, 2006; Carlander et al., 2007; Oskarsson et al., 2012).**
- **Illustrating how different modalities can complement each other at different phases during localisation and manoeuvring towards threats (Carlander et al., 2007; Eriksson et al., 2006; Oskarsson et al., 2012; van Erp et al., 2007).**
- **Showing how tactile signals remove the occurrences of front-back confusion (Carlander & Eriksson, 2006; Carlander et al., 2007; Oskarsson et al., 2012).**
- **Showing that in-ear headphones can be used for 3D audio if this is required by the application (e.g. Carlander & Eriksson, 2006).**
- **Showing that 3D audio and tactile presentations can be used in noisy environments such as the CV90 (Carlander et al., 2007; Carlander & Eriksson, 2006; Oskarsson et al., 2012), and that tactile signals can be perceived in a simulated fighter jet at high G-loads (Eriksson et. al., 2006; van Erp et al., 2007).**

Chapter 5. General Discussion and Conclusions

Methodological reflections

In the first 3D audio experiment we expected relatively low localisation errors using the Lake system (reference technology) but the overall performance of the COTS system was rather positively surprising. We used four reference positions (0°, 90°, 180° and 270° degrees) in the room so that the participants were able to map their response to the perceived direction of the presented sound. Perhaps these "reference" sound positions biased responses resulting in lower LEs. More specifically, participants could have "concluded" that stimuli were presented either "on" the reference symbols or between them. Participants also received feedback on their responses during the training, which may have further "taught" them the positions.

In the second experiment, the apparatus was kept the same but the experimental design was slightly modified to rule out the possibility that the low error rates partly resulted from biased responses. Thus, reference symbols and the graphical response form were removed and no feedback was administered at any point before or during the experiment. However, the training session remained to familiarize the participant with the experimental environment and procedure. Sound positions were doubled and offset 5 degrees, resulting in 24 equally spaced azimuths in intervals of 15º (i.e. 5º, 20º, 35º etc.). With this design we ruled out that participants correctly guessed the sound positions.

One further step was taken in the third experiment where sounds were randomized within each of four quadrants of the horizontal plane. The reason for not randomising sounds 360° around the listener was to ensure that sounds were distributed approximately evenly in the horizontal plane.

In the study with the fire and rescue command operators the background voices had an amplitude that was equal to the signal amplitude (SNR of 0dB). Adding to this was a secondary task that induced a high workload. The high workload condition was intended to resemble the high workload potentially occurring for an operator. However, by primarily introducing many background voices and an SNR of 0 dB, the primary task became difficult. While this was partly expected (Drullman & Bronkhorst, 2000; Nelson et al., 1998; Haas, 1998a; Wickens & Hollands, 2000), the background voices in combination with the demanding secondary task probably made conditions too difficult regardless of auditory display. Although these design choices were made to acquire a task as realistic as possible we should have made the secondary task easier and also introduced timing offsets between calls to avoid a complete overlap.

The CV90 platform experiment (experiment 7) was initially planned to be performed with the vehicle moving on a "combat track" to simulate real driving conditions. However, due to technical constraints of our research platform we had to redesign the experiment so that the vehicle was not moving forward but only rotated towards targets. Participants indicated their response by turning the vehicle towards the threat and this is not always a preferred action since it depends on threat type, terrain, general tactics and the time available. However, the experts at the Army combat school did rate the threat manoeuvre task relevant and realistic enough primarily because it still involves dynamic performance with precision handling of the vehicle relating directly to the displayed information. (The task used in both experiments 8 and 9 was accordingly the same as in experiment 7 to facilitate direct comparisons.)

Even though the response time measure was important for this study it is important to realise that in the real world scenarios, the threat-warning would mainly be most useful for nearby CV90s since the time from warning to possible impact normally would be very short. In addition, an improvement for future studies would be to include a comparison with the currently employed warning display system. Optimally, this should have been a reference to the actual improvement made by the new displays. However, the CV90 development team at the Army combat school was certain that the future WCS warning signal should not only be displayed visually to the driver, since he as well needs to keep the visual attention to the outside terrain. The mere showing of the performance of the 3D audio and tactile belt in the CV90 technology-demonstrator platform was therefore considered a success. The results can also be related to those from the experiment with the simulated Gripen fighter jet. The results from that experiment showed that the tactile threat cueing improved performance even in the presence of high-end visual displays, including the use of primarily the HUD, with head-up information.

Responses to research questions

This thesis has presented 10 different experiments with 3D audio in the lab and in the field, tactile, bimodal and multimodal displays in simulators and real platforms. The results from these experiments are summarised as answers to the research questions below.

Can a portable 3D audio system generate spatial audio, accurate enough for operational use?

Yes, we have shown that localisation errors around 10 degrees are feasible to expect for both laboratory (Carlander et al., 2006), experiments with real operators (Carlander & Eriksson, 2006), and in simulated platforms (Carlander et al., 2007; Oskarsson at al., 2012). Over the experiments the inter-subject variability was small.

In experiments 1 to 3 we showed that the horizontal localisation accuracy of a portable system based on COTS components was accurate and at the time comparable to a state-of-the-art 3D audio research platform. Experiments 1 to 3 collectively provide evidence that the COTS solution has proven good enough to be further tested in applied settings (Carlander & Eriksson, 2006; Carlander et al., 2005; Kindström et al., 2006). The low and stable error rates can be explained by an effective simulation of the interaural difference cues. The COTS 3D audio will add advantages such as ease of implementation and low cost.

We also used the COTS system in a more realistic setting in Experiment 7 (Carlander & Eriksson, 2006) and 10 (Carlander et al., 2007; Oskarsson at al., 2012). The results from these experiments show that the system is able to generate localization accuracy below 10 degrees. Important to note is that we found front-back confusions for all 3D audio experiments ranging from a few percent up to around 12% for one participant. These front-back confusion occurrences may be critical depending on type of application. Therefore the use of 3D audio must be carefully considered and perhaps applications requiring high accuracy in the front-back dimension should be avoided if not used with tactile cueing.

These results are in line with the overall goal to make the 3D audio technology more accessible and useful to the operators. It also demonstrates the possibility of using COTS components and software rather than the specialized 3D audio hardware systems for presenting sounds in the horizontal plane. This potentially makes implementation more efficient through a more flexible and cost-effective solution. It may also pave way for a more widespread acceptance and implementation of the technology, especially in applied settings. A good example could be the two-channel radio communication systems that immediately could make use of the advantages of the binaural auditory system.

Can 3D audio enhance the intelligibility of call signs?

Yes, we have shown how 3D audio enhances intelligibility by increasing separation between sound sources.

In comparing the intelligibility of stereo and 3D audio radio calls in background noise, we showed that the 3D audio presentation increased the number of correctly identified call signs (Carlander et al., 2005; Kindström et al., 2006). The results are in line with previous research that shows how 3D-audio is more effective for presenting simultaneous sound sources compared to traditional audio displays (Baldis, 2001; Begault, 1999b; Doll et al., 1992; Haas, 1998a; Nelson et al., 1998). The higher accuracy in the “two background voices” condition compared to four was expected and is in line with Ericson and McKinley (1997). This is also seen for the identification of sets of call signs where set sizes of both one and two call signs showed higher accuracy than sets of three and four, respectively. As suggested by Ericson and McKinley (1997), it might indicate that more than three simultaneous talkers overload the auditory system.

Most probably it is the increased spatial separation of call signs as a result of 3D audio that offers the slightly better intelligibility in the high workload conditions (e.g. Begault, 1999b; Doll et al., 1992; Nelson et al., 1998). However, some participants reported that the call signs were not separated enough and that sound sources blended. This was most likely an effect of using the portable COTS system not specialized for taking into account room parameters, full 3D space and distance to sound sources.

Overall, the improved intelligibility achieved with 3D audio for radio communication could be of vital importance for fire and rescue command operators, reducing the risk of misinterpretations and missed call signs. These findings will most likely apply to operators in similar settings in which simultaneous call signs cause auditory clutter. Radio communication systems are also relatively easy to improve since the information already is auditory and that sources often are separated by different channels.

Can 3D audio and visual threat indication be improved by adding a tactile presentation?

Yes, we have shown how tactile presentation can complement 3D audio presentation and resolve front-back confusions, thus reducing misinterpretations and enhancing overall accuracy. For the experiment relating to the Gripen fighter jet the tactile presentation improved performance even in the presence of high end visual displays.

Although 3D audio has proven to generate low localisation errors it still has lower performance regarding the reliability and accuracy of a visual display (Carlander et al., 2007). However, 3D audio offers many interesting opportunities apart from navigation and threat cueing. The true potential for the 3D audio displays are most likely the situations where audio-based information is already used, such as radio communication or warning signals *not* relying on spatial accuracy. To improve the 3D audio presentation, we investigated the localisation accuracy of 3D audio only, 3D audio with head-tracker and 3D audio and tactile cueing. We found larger errors in localisation with 3D audio only, as compared to both the head-tracked 3D audio, and the bimodal condition including the tactile cueing. Thus a head-tracker or a tactile presentation would improve the 3D audio presentation and when the tactile presentation was added to the 3D audio presentation all front-back confusions disappeared. The effect was present although a rather simple tactile display was used (Carlander & Eriksson, 2006; Carlander et al., 2007; Oskarsson et al., 2012). In the study with tactile threat cueing for the pilot of a fighter jet, the target interception performance showed that the tactile display improved performance even with high quality visual displays present. The tactile display captured the pilot's attention at threat pop-up and improved reaction time. Gaining 200 ms in reaction time and 1s when chasing targets behind the pilot’s aircraft could be critical for threat handling tactics and even survival.

What are the benefits of auditory, tactile and multimodal threat cueing in a combat vehicle?

The main benefits are that information presentation to different modalities can capitalize on the advantages that characterises each modality (e.g. grabbing attention, high spatial accuracy etc.).

Fundamental for our experiments is that during different phases of manoeuvring towards threats different modalities seem to offer the best performance. For example, the tactile presentation will capture attention in the right direction, increasing threat awareness, and the visual display information can provide high resolution information. Although promising, the modalities for different phases of manoeuvring needs further investigation since only one design for each display was tested. Relating to this is the second experiment[17] in Oskarsson et al. (2012), where the visual presentation was improved from a HDD to a HUD. This resulted in that the condition including 3D audio and HUD performed on par with the trimodal condition (HUD, 3D audio and tactile). Thus, in this experiment both LE and RT with the trimodal display can be explained by the contribution of the HUD and 3D audio. In essence, the display design for each modality can be more or less suitable for different phases.

For threat warnings, it shows how sensory redundant information can overcome display-specific limitations as seen in the different phases of manoeuvring. More specifically, when the visual sense is unable to provide useful cues other senses can support the perceptual experience enhancing performance in dynamic tasks. These results are also in line with the general idea that discrepancies are always resolved in favour of the more appropriate modality (e.g. Ernst & Bülthoff, 2004). Perhaps these findings can be the starting point for generating design guidelines for how to use the characteristics of our senses at different phases in dynamic scenarios. Carefully designed interfaces can take

---

[17] Note that experiment 10 is part of Oskarsson et al., (2012) where a second experiment is presented. Only the first experiment is presented in this thesis.

advantage of the sensory characteristics and utilize the differences for information presentation during phases of a task.

In the CV90, both tactile and 3D audio displays could be perceived well, regardless of the noise and vibrations. However, the front-back confusions can have serious implications for threat presentations and although these were resolved for the analysis localisation errors were still greater compared to the other displays (In experiment 7). However, as a result of the limited contribution of the 3D audio to the multimodal display one might speculate that it may be sufficient to combine visual and tactile displays for improving display effectiveness. On the other hand, the combined sensory presentation may be better to achieve safer and more reliable performance in certain time critical tasks. For a prioritized threat cueing in a combat vehicle, it may be best to include the auditory display alongside the visual and tactile displays since one, or both of them, may not be available or perceived at a critical moment.

We have shown that the combination of modalities potentially reduce localisation errors and generate shorter RTs. The shorter RT with the multimodal display as compared to the HDD in Experiment 10 further demonstrates the benefit of increased alerting by including redundant tactile information to that of the HDD, which may not always be attended (Carlander & Eriksson, 2006; Carlander et al., 2007; Oskarsson et al., 2012). Interestingly, the tactile localisation errors were low (6°-10°) for experiments 7-10, even though the tactile display had a resolution of only 30°. The most likely reason is that when the threat indication was moving dynamically from tactor to tactor, as a result of rotating the vehicle, it was possible to integrate the time factor and thus improve the estimation of threat direction. Gilson, Redden, and Elliott (2007) and van Erp (2002) also showed that the occurrence of movement improved localisation. This is explained by our capability of understanding movement and timing to turn this into an anticipation of the precise and rhythmic pattern (Sacks, 2007; van Veen & van Erp, 2003).

In the studies relating to CV90, (Carlander & Eriksson, 2006; Carlander et al., 2007; Oskarsson, et al., 2012) we showed that sensory redundant information can improve threat indications. Information presented by tactile and/or 3D audio offers omnidirectionality, increased precision and response times and possibly a more intuitive presentation. This is also in line with what Bresciani et al. (2005) suggests; combining sensory signals reduces the variance of perceptual estimates, thus enhancing stimuli detection by cue redundancy. The multimodal redundant presentation can thus help directing the attention of the user and enhancing detection is central when designing warning signals. The combination of the senses can be used to add more “power” to a certain position in space, and the multimodal displays seems to not induce heightened mental workload (Carlander et al., 2007; Oskarsson et al., 2012).

The enhanced performance of multimodal information bears a potential to be generally applicable to the dynamic situations of the battlefield, or in rescue missions where there is a need for rapid reactions and responses, and where decisions based on misinterpretations can have lethal outcome.

Can 3D audio and tactile displays be used effectively and efficiently without extensive operator training?

Yes and no. For the 3D audio studies and the studies relating to the CV90, 3D audio and tactile systems have proven to be immediately useful and result in a high performance

(Carlander et al., 2007; Carlander & Eriksson, 2006; Carlander et al., 2005; Eriksson et al., 2006; Kindström et al., 2006; Oskarsson et al., 2012), and this with little or no training. However, for the experiment with pilots of the Gripen fighter jet additional training with the tactile display would most likely have improved performance, although the performance still was better than with the visual displays they are used to.

During the close cooperation with operators it has been quite apparent that the displays need to be easy to use and understand. To reach an immediate acceptance the displays should be intuitive and require little or no training. This was partly tested in the experiments in which only a very brief training was administered and yet participants performed well with the novel display systems. One can always argue that we could have reached an even better performance with training but the results showed localisation errors of only around 10 degrees, which is almost as good as natural sound localisation with open ears (about 4-10 degrees (Begault, 1999a)). For the tactile display it was even as low as 6 to 8 degrees. For the 3D audio, we removed the training session for experiment 2 and comparing the results with experiment 1 we could not find a decrease in performance. Some studies have shown that training has an effect but this relates more to reducing the number of sounds as perceived “inside the head” and to increase users’ confidence in their responses (Trapenskas & Johansson, 1999). More long-term effects could not be seen until the 7th day of training (Ohuchi, Iwaya, Suzuki & Munekata, 2005).

In the real platform study, Experiment 7 (Carlander & Eriksson, 2006), only an introductory training was administered so each participant would understand what a threat felt or sounded like. Although this brief familiarisation, participants yet had an excellent overall performance in reacting to and positioning the vehicle toward threats. This was also reflected in comments and questionnaires where participants rated the displays performance as high. One can thus conclude that when implemented these displays require a minimum of training and are easy to use and understand. One may also conclude that the presented signals were simple and self-explanatory so that tactile clutter and reduced comprehension was avoided (van Erp et al., 2003).

Several of the pilots of the Gripen fighter jet experienced the tactile presentation too fast and dynamic. This was a result of the manoeuvring toward the threat that caused the presentation of the vibrations to “jump” across the torso. This kind of presentation needs to be learned properly to create a stronger mapping between presentation and hand coordination. Still, the performance was better with the tactile display added than with only the visual displays they are used to. However, with proper training there will be an even greater value of the tactile display.

Final remarks and future research

The investigations in this thesis showed that bi- and multimodal displays resulted in an increased operator performance in complex environments. It is mainly the omnidirectional character of the tactile and 3D audio presentations that contributes significantly to the making of a simple, effective and intuitive interface, which results in better perception and performance without adding workload.

One could speculate if it is possible to generalise the results from the rather specific environments in these studies to other settings and populations. Most of the studies have general underlying questions concerning, for example, accuracy or response time and they focus a lot on human performance as a result of the characteristics of each display technology. Thus, the results have a potential to be generally applicable. For instance the simultaneous radio communication for fire and rescue command operators can be transferred to settings where it is required to handle simultaneous radio calls during high workload. This is relevant, for example, in in-car information systems, air traffic systems, military communication systems etc. Furthermore, the study of directional warning signals in the CV90 can be transferred to driving environments where high attention is needed to the outside terrain while having certain tasks requiring spatial information, like GPS navigation, warning displays, vehicle controls etc.

Quite naturally we will always be facing in some direction and only some information will be available visually. However, the “hidden” information can be revealed by well-designed interfaces that become an integrated part of how we interact with our surroundings. The new displays can offer complementary information to the visual displays and support the operator in situations in which the operator already is engaged in visual activities. Furthermore, the development of new technologies opens up possibilities to better support the visually impaired (e.g. Huang, 2010; Vitense, Jacko, & Emery, 2002). This may not only be to solve a specific work-related or practical task, but it may even involve areas such as sport. In a project where 3D sound was combined with 3D cameras and Smartphones, visually impaired people were able to play football. A team of blind footballers were equipped with a system that enabled them to search for the ball by listening for it – they “saw with sound”. 16 cameras covered the pitch and the system essentially tracked the position of each player in real-time. A smartphone converted this information into a surround-sound landscape. A simple head-tracker was accomplished by attaching the Smartphone to the participant’s head and turning the head would thus update the sound of the ball and the other players. In the extension of this the designers hope to be able to help the visually impaired to "hear-see" more of the world around them (Toor, 2011). In combination with tactile information, and a clever allocation of information, it is likely that the participants in this project would have performed even better. Multimodal displays for impaired users are an interesting area where the display types presented in this thesis would be of great use.

The future work as an extension of this thesis should continue to focus on applied research and more closely look at how to benefit from the sensory characteristics at different phases of information gathering. Perhaps these characteristics could be better mapped and used for both sensory redundant information like warnings and attain an efficient sensory allocation for different applications. Critical aspects such as how much information that can be presented, what kind of information should be displayed to which sensory modality, and how the sensory modalities interact need to be considered. One of the main challenges is how to prioritize and divide the streams of information over modalities for simple and intuitive displays (Gilson et al., 2007). It will also be a critical task to investigate how to switch the information over the modalities. Can this really be handled seamlessly by an advanced sensor/situation analysis system? Will it at all be possible to generalize an optimal presentation for a certain sense at a given point in time? It would also be interesting to investigate performance when the display technologies are used under longer periods (vigilance task) (Ernst & Bülthoff, 2004). To be able to get valid answers to these questions, end users must continue to help refining research questions to maximize the benefits of the new technologies.

Compared to visual and audio display technologies the multisensory display systems are still relatively new. The tactile and multimodal technologies need further improvement to reach their full capability and to fulfil demands arising from operators that are supposed to handle large amount of information during mission critical activities. These demands can, for example, be low cost, little or no acoustic or thermal signature, small size, low electrical power consumption, high performance etc. With an efficient research and development of the new display types the potential is great. Some examples of improvements already exist and EAP and bone conduction are examples that could result in displays overcoming shortcomings such as blocked ears and complex tactors (Ketabdar & Polzehl, 2009; Koo et al., 2008; Vazquez-Alvarez & Brewster, 2009).

We are on our way toward intuitive interfaces with still much to explore. The development of intuitive interfaces will not be finalized at a given point in time nor will inferior interfaces instantly be replaced in favour of intuitive multimodal ones. It is now up to the research community to continue the refinement of information presentation and to the designers of future information systems to take on this knowledge to implementation. Haas (2007) summarises the importance of multimodal presentation in the following way: *“As long as humans monitor displays and operate controls, multimodal control and display technologies can be used effectively to enhance soldier and civilian system safety”* (Haas, 2007, p. 38).

As an ultimate goal the new interfaces should lead to information presentation that requires little training, and achieve operator responses that are faster, more reliable, and that supports a better awareness of evolving situations, alongside lower or not increased operator load (Eriksson et al., 2005). Reaching this goal, we would have the interfaces required to improve operator performance and even support the operator in making life-critical decisions during dynamic missions.

References

Abildgaard-Pedersen, J., & Jørgensen, T. (2005). Localisation performance of real and virtual sound sources. In *Proceedings of the HFM-123/RSY Symposium, New directions for improving audio effectiveness*, Amersfoort April 11-13, (pp. 29-1 – 29-30). Amersfoort: North Atlantic Treaty Organisation, Research and Technology Organisation.

Adams, J. (1989). *Human Factors Engineering*. New York: Macmillan Publishing Company.

Albery, W. (2007). Multisensory cueing for enhancing orientation information during Flight. *Aviation, Space and Environmental Medicine 78*, pp. B186-B190.

Albery, W. (2006). *Development of multisensory orientation technology for reducing spatial disorientation, AFRL-HE-WP-TR-2006-0024*. US Ohio:Air Force Research Laboratory, Human Effectiveness Directorate, Wright-Patterson Air Force Base.

Andersson P., Carlander, O., Hörberg, U., Jander, H., Kindström, M., & Sandberg, S. (2005). *Invisning med fritt tal, bäring 3D-ljud, FOI-R--1786—SE*. Linköping: FOI.

Andersson P., Carlander, O., Hörberg, U., Jander, H., Kindström, M., & Sandberg, S. (2004). *Målangivning, FOI-R--1445—SE*. Linköping: FOI.

Arons, B. (1992). A review of the cocktail party effect. *Journal of the American Voice I/O Society, Vol. 12*, pp. 35-50.

Ayama, M., Sakurai, M., Carlander, O., Derefeldt, G. & Eriksson, L. (2004). Color appearance in peripheral vision. *Human Vision and Electronic Imaging. SPIE Proceedings*, Volume 5292, pp. 260-271. Bellingham, WA:SPIE.

BAE Systems. (2006). *Combat Vehicle 90 Family*. Retrieved from http://www.baesystems.com/ProductsServices/landa_hagg_products_cv90herita.html.

Baldis, J. (2001). Effects of spatial audio on memory, comprehension, and preference during desktop conferences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Seattle March 31 - April 5 (pp. 166-173). Washington: CHI.

Bar-Cohen, J. (2009). *Electroactive polymers for refreshable Braille displays*. Retrieved on date, year, online from http://spie.org/x37076.xml?highlight=x2406&ArticleID=x37076.

Bartlett S., Cox, S., Butler, R., & Potas, M. (2003). *Lake Huron Technical manual*, Sydney, Australia: Lake Technology.

Begault, D. (1999a). Auditory and non-auditory factors that potentially influence virtual acoustic imagery. In *Proceedings of the AES 16th International conference on Spatial Sound Reproduction*, Rovaniemi April 10-12, (pp. 13-26). Rovaniemi: AES.

Begault, D. (1999b). Virtual acoustic display for teleconferencing: Intelligibility advantage for “telephone-grade” audio. *Journal of the Audio Engineering Society, Vol. 47*, pp. 824-828.

Begault, D. (1994). *3D sound for virtual reality and multimedia*. Boston: Academic press Inc.

Begault, D., Wenzel, E., Lee A., & Anderson, M. (2001). Direct comparison of the impact of headtracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source. *Journal of the Audio Engineering Society, Vol. 49*, pp. 904-916.

Begault, D., & Wenzel, E. (1993). Headphone localisation of speech. *Human Factors, Vol. 35*, pp. 361-376.

Blauert, J. (2001). *Spatial hearing- The psychophysics of sound localisation* (rev. ed.). London: The MIT Press.

Berglin, L. (2006). *Interactive textile structures – Experimental product design in smart textiles* (PhD Thesis). Chalmers University of Technology, Göteborg.

Bresciani, J., Ernst, M., Drewing, K., Bouyer, G., Maury, V., & Kheddar, A. (2005). Feeling what you hear: auditory signals can modulate tactile tap perception. *Experimental Brain Research, Vol. 62*, pp. 172-180.

Brill, J., Gilson, R., Mouloua, M., Hancock, P., & Terrence, P. (2004). Increasing situation awareness of dismounted soldiers via directional cueing. In Vincenzi, D., Mouloua, M. & Hancock, P., (Eds.), *Human performance, situation awareness and automation: Current research and trends* (pp. 130-132). Mahwah, NJ: Erlbaum.

Bronkhorst, A. (1995). Localisation of real and virtual sound sources. *Journal of the Acoustical Society of America, Vol. 98*, pp. 2542-2553.

Burdea, G., & Coiffet, P. (2003). *Virtual reality technology* (2nd ed.). New Jersey: Wiley-Interscience.

Calhoun, G., Valencia, G., & Furness, T. (1987). Three-dimensional auditory cue simulation for crew station design/evaluation. In *Proceedings of the Human Factors Society, 31st Annual Meeting*, New York October 19-22 (pp.1398-1402). Santa Monica: Human Factors Society.

Capraro, A., Verrillo, R., & Zwislocki, J. (1979). Psychophysical evidence for a triplex system of cutaneous mechanoreception. *Sensory Processes, Vol. 3,* pp. 340-352.

Carlander, O. (2006). *3D Audio towards an intuitive interface- Uni- and bimodal investigations* (Tec.Lic. Thesis). Chalmers University of Technology and Göteborg University, Göteborg.

Carlander, O., & Hasewinkel, H. (2003). *Multimodalitet-möjligheternas operatörsplats, FOI-R--1109—SE.* Linköping: FOI.

Carlander, O. (2002). *Localisation of sound in a horizontal 3D-audio environment* (Masters Thesis). Linköping: HFA, Human Factors in Aviation.

Carlander, O. (2000). *On functional differences between foveal and peripheral colour vision - a literature survey and an experimental investigation.* FOA-R--00-01791-706—SE. Linköping: FOA.

Carlander, O., & Eriksson, L. (2012). *Vibrotactile and 3D audio threat cueing for a combat vehicle environment.* (Unpublished manuscript.)

Carlander, O., & Eriksson, L. (2006). Uni- and bimodal threat cueing with vibrotactile and 3D audio technologies in a combat vehicle. In *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting,* San Francisco October 16-20, (pp.1552-1556). Santa Monica: Human Factors and Ergonomics Society.

Carlander, O., Eriksson, L., Kindström M., & Chen, F. (2012). *Perceived accuracy of horizontally distributed sounds of two 3D-audio display technologies.* (Unpublished manuscript.)

Carlander, O., Eriksson, L., & Kindström, M. (2006). Horizontal localisation accuracy with COTS and professional 3D audio display technologies. In *Proceedings of the International Ergonomics Association 2006 conference*, Maastricht July 10-14. Maastricht: Elsevier Ltd.

Carlander, O., Eriksson, L., & Oskarsson, P-A. (2007). Handling uni- and multimodal threat cueing with simultaneous radio calls in a combat vehicle setting. In *Proceedings of the 4th international conference on Universal access in human-*

computer interaction: ambient interaction, Beijing July 22-27, (pp. 293-302). Berlin/Heidelberg: Springer.

Carlander, O., & Kindström M. (2004). *Urskiljbarhet av radiokommunikation - En jämförelse av 3D- och stereoljud.* FOI-R--1525—SE. Linköping: FOI.

Carlander, O., Kindström, M., & Eriksson, L. (2005). Intelligibility of stereo and 3D audio call signs for fire and rescue command operators. In *Proceedings of the 11th Meeting of the International Conference on Auditory Display*, Limerick 6-9 July, (pp. 292-295). Limerick: International Conference on Auditory Displays.

Carlson, N. (2001). *Physiology of behaviour* (7th ed). Massachusetts: Needham heights.

Castle, H., & Dobbins, T. (2004). Tactile displays for enhanced performance and safety. In *Proceedings of the SPIE, Vol. 5443*, pp. 269-276.

Chen, F. (2003). Localisation of 3-D sound presented through Headphone- Duration of sound presentation and localisation accuracy. *Journal of the Audio Engineering. Society, Vol. 51*, pp. 1163-1171.

Chen, F., & Carlander O., (2003a). *Review of 3-Dimensional Sound and Head Related Transfer Functions*. Linköping: HFA, Human Factors in Aviation.

Chen, F., & Carlander O. (2003b). Localisation of sound in a noise environment. In *Proceedings of the XVth triennial congress of the International Ergonomics Association and the 7th joint conference of Ergonomics Society of Korea / Japan Ergonomics Society*, Seoul August 24 – 29. Korea: International Ergonomics Association.

Cherry, E. (1953). Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America, Vol. 25*, pp.975-979.

Cheung, B., van Erp, J., & Cholewiak, R. (2008). Anatomical, neurophysiological and Perceptual issuers of tactile perception. In by J. van Erp & B. Self (Eds.), *Tactile Displays for Orientation, Navigation and Communication in Air, Sea and Land Environments* (pp.2-1 – 2-18). Neuilly-sur-Seine Cedex, France: Research and Technology Organisation, North Atlantic Treaty Organisation.

Choi, H., Koo, I., Koo, J., Nam, J.-D., & Lee, Y. (2008, January 9). Bandage-sized flexible touch display, *SPIE Newsroom*. Retrieved from http://spie.org/x32826.xml.

Cholewiak, R., & Collins, A. (2003). Vibrotactile localization on the arm: Effects of place, space, and age. *Perception & Psychophysics, Vol. 65*, pp. 1058-1077.

Coxworth, B. (2010). *3rd Space FPS Vest lets you feel the gaming action*. Retrieved from http://www.gizmag.com/3rd-space-fps-gaming-vest/14825/.

Dellepiane, M., Pietroni, N., Tsingos, N., Asselot, M., & Scopigno, R. (2008). Reconstructing head models from photographs for individualized 3D-audio processing. *Computer Graphics Forum (Special Issue - Proc.Pacific Graphic), Vol. 27*, pp. 1719-1727.

Dobbins, T., & Samways, S. (2002). The use of tactile navigation displays for the reduction of disorientation in maritime environments. In *NATO RTO Symposium on Spatial Disorientation in Military Vehicles: Causes, Consequences and Cures*, La Coruña 15-17 April, (pp. 33-1 – 33-6). Neuilly-sur-Seine Cedex, France: Research and Technology Organisation, North Atlantic Treaty Organisation.

Doll, T., Hanna, T., & Russotti, J. (1992). Masking in three-dimensional auditory displays. *Human Factors, Vol. 34*, pp. 255-265.

Drullman, R., & Bronkhorts, A. (2000). Multichannel speech intelligibility and talker recognition using, monaural, binaural and three-dimensional auditory presentation. *Journal of Acoustical Society of America, 107*, pp. 2224-2235.

Eimer, M. (2004). Multisensory integration: How visual experience shapes spatial perception. *Current Biology, Vol. 14*, pp. 115-117.

Elliott, L., Duistermaat, M., Redden, E., & van Erp, J. (2007). *Multimodal guidance for land navigation*. Retrieved from http://www.dtic.mil/cgi-bin/GetTRDoc?Location=U2&doc=GetTRDoc.pdf&AD=ADA449965.

Elliott, L., Redden, E., Pettitt, R., Carstens, C., van Erp, J., & Duistermaat, M. (2006). *Tactile guidance for land navigation*. Retrieved from http://www.dtic.mil/cgi-bin/GetTRDoc?Location=U2&doc=GetTRDoc.pdf&AD=ADA449965.

Ericson, M., Brungart, D., & Simpson, B. (2004). Factors that influence intelligibility in multitalker speech displays. *The International Journal of Aviation Psychology, Vol. 14*, pp. 313–334.

Ericson, M., & McKinley, R. (1997). The intelligibility of multiple talkers separated spatially in noise. In Gilkey, R. & Anderson, T. (Eds), *Binaural and Spatial Hearing in Real and Virtual Environments* (pp. 701–724). Mahwah: Lawrence Erlbaum Associates.

Eriksson, L., Berglund, A., Willén, B., Svensson, J., Petterstedt, M., Carlander, O., Lindahl, B., & Allerbo, G. (2008). On visual, vibrotactile, and 3D audio directional cues for dismounted soldier waypoint navigation. In *Proceedings of Human Factors and Ergonomics Society 52nd Annual Meeting, Vol. 52*, New York 22-26 September, (pp. 1282-1286). Santa Monica: Human Factors and Ergonomics Society.

Eriksson L., Carlander O., Borgvall J., Dahlman J., & Lif, P. (2005). *Operator site 2004-2005*. FOI-nummer? Linköping: FOI.

Eriksson, L., Lindahl, B., & Hedström, J. (2006). *Konceptutveckling av operatörsstöd.* FOI-R--2210--SE. Linköping : FOI.

Eriksson, L., van Erp, J., Carlander, O., Levin, B., van Veen, H., & Veltman, H. (2006). Vibrotactile and visual threat cueing with high G threat intercept in dynamic flight simulation. In *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting*, San Francisco October 16-20, (pp. 1547 – 1551). Santa Monica: Human Factors and Ergonomics Society.

Ernst, M., & Bülthoff, H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences, Vol. 8*, pp. 162–169.

FMV (2009). Swedish defense and material administration, Dynamic Flight Simulator DFS. Retrieved from http://www.fmv.se/WmTemplates/Page.aspx?id=4508.

FOI (2012). *Swedish Defense Research Agency, Annual Report*. Retrieved from http://www.foi.se/Global/Om_FOI/Årsrapporter/2012_Arsrapport_FOI_sve_webb.pdf.

Gazzaniga, M., Ivry, R., & Mangun, G. (1998). *Cognitive neuroscience- the biology of the mind*. London: W W Norton & Company.

Geldard, F., & Sherrick, C. (1972). The Cutaneous "Rabbit": a perceptual illusion. *Science, Vol. 178*, pp.178-179.

Genuit, K. (1986). A special calibratable artificial-head-measurement-system for subjective and objective classification of noise. In *Proccedings of Inter-Noise 1986*, Cambridge MA July 21-23, (pp. 1313 – 1318). Poughkeepsie, NY: Noise Control Foundation.

Gilson, R., Redden, E., & Elliott, L. (2007). *Remote tactile displays for future soldiers.* Aberdeen, MD: US Army Research Laboratory.

Guyton, A. (1991). *Basic neuroscience - anatomy & physiology* (2nd ed.). Philadelphia: W.B. Saunders Company.

Haas, E. (2007). Emerging Multimodal Technology: Role in enhancing combat and civilian system safety. *Professional Safety, Vol. 52*, pp. 32-38.

Haas, E. (1998a). Utilizing 3-D auditory display to enhance safety in systems with multiple radio communications. In S. Kumar (Ed.), *Advances in Occupational Ergonomics and Safety* (pp.715-718). Amsterdam: IOS Press.

Haas, E. (1998b). Can 3-D auditory warnings enhance helicopter cockpit safety? In *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting,* Chicago, October 5-9, (pp. 1117-1121). Santa Monica, CA: Human Factors and Ergonomics Society.

Hahn, J. (1966). Vibrotactile adaptation and recovery measured by two methods. *Journal of Experimental Psychology, Vol. 71,* pp. 655-658.

Hancock, P., Oron-Gilad, T., & Szalma, J. (2007). Elaborations of the multiple resource theory of attention. In A. Kramer, D. Wiegmann & A. Kirlik (Eds.), *Attention: From theory to practice* (pp. 45-56). Oxford: Oxford University Press.

Ho, C., Spence, C., & Tan, H. (2005). Warning signals go multisensory. In *Proceedings of HCI International 2005,* Las Vegas 22-27 July. (pp. 1-10). Nevada: Lawrence Erlbaum.

Hofman, P., van Riswick, J., & van Opstal, A. (1998). Relearning sound localisation with new ears. *Nature Neuroscience, Vol. 1,* pp. 417-421.

Hollan, J., Hutchins, E., & Kirsh, D. (2000). Distributed cognition: Toward a new foundation for human computer interaction research. *ACM Transactions on Computer-Human Interaction, Vol. 7,* pp. 174–196.

Huang, Y. (2010). *Design and evaluation of 3D multimodal virtual environments for visually impaired people* (PhD Thesis). KTH, Stockholm.

Hwang, S., Park, Y., & Park, Y. (2008). Modeling and customization of head-related impulse responses based on general basis functions in time domain. *Acta Acustica united with Acustica, Vol. 94,* pp. 965-980.

Iwaya, Y., Suzuki, Y., & Kimura, D. (2003). Effects of head movement on front-back error in sound localization. *Journal of Acoustical Science and Technology, Vol. 24*, pp. 322-324.

Jones, W. (2009). *Jacket lets you feel the movies*, Retrieved from http://spectrum.ieee.org/biomedical/devices/jacket-lets-you-feel-the-movies.

Jones, L., & Ho, H.-N. (2008). Warm or cool, large or small? The challenge of thermal displays. *IEEE Transactions on Haptics archive, Vol. 1,* pp. 53-70.

Jones, L., & Sarter, N. (2008). Tactile displays: Guidance for their design and application. *Human Factors, Vol. 50,* pp. 90–111.

Kaaresoja, T., & Linjama, J. (2005). Perception of short tactile pulses generated by a vibration motor in a mobile phone. In *Proceedings of the First Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*. Pisa, Italy March 18-20, (pp. 471-472). Los Alamitos, CA: IEEE Computer Society.

Kato, M., Uematsu, H., Kashino, M., & Hirahara, T. (2003). The effect of head motion on the accuracy of sound localization. *Journal of Acoustical Science and Technology, Vol. 24,* pp. 315-317.

Ketabdar, H., & Polzehl, T. (2009). Tactile and visual alerts for deaf people by mobile phones. In *Proceedings of the 11th international ACM SIGACCESS conference on Computers and accessibility*. Pittsburgh October 25-28, (pp. 253-254). New York: ACM

Kindström, M., Carlander, O., & Eriksson, L. (2006). Comparison of audio systems for intelligibility in multitalker speech displays. In *Proccedings of the Human Factors and Ergonomics Society 50th Annual Meeting*. San Francisco October 16-20, (pp. 309-313). Santa Monica: Human Factors and Ergonomics Society.

King, R., & Oldfield, S. (1997). The impact of signal bandwidth on auditory localization: Implications for the design of three-dimensional audio displays. *Human Factors, Vol. 39*, pp. 287-295.

Koo, I., Jung, K., Koo, J., Nam, J., Lee, Y., & Choi, H. (2008). Development of soft-actuator-based wearable tactile display. *IEEE Transactions on Robotics, Vol. 24,* pp. 549-558.

Lemmens, P., Crompvoets, F., Brokken, D., van den Eerenbeemd, J., & de Vries, G.-J. (2009). A body-conforming tactile jacket to enrich movie viewing. In *Proceedings of the World Haptics 2009 - Third Joint EuroHaptics conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems.* Salt Lake City March 8-10, (pp. 7-12). Washington: IEEE Computer Society.

Macpherson, E., & Middlebrooks, J. (2000). Localisation of brief sounds: Effects of level and background noise. *Journal of the Acoustical Society America, Vol. 108,* pp.1834-1849.

Martens, W. (2003). Perceptual evaluation of filters controlling source direction: Customized and generalized HRTFs for binaural synthesis. *Acoustical Science and Technology, Vol. 24,* pp. 220-232.

Mayer, A. (1880). *US Patent No. 224,199*. New Jersey: United States Patent Office.

McGrath, B. (1999). *Tactile instrument for aviation.* Retrieved from http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA531665.

McGrath, B., Estrada, A., Braithwaite, M., Raj, A., & Rupert, A. (2004). *Tactile situation awareness system flight demonstration*. Retrieved from http://www.dtic.mil/cgi-bin/GetTRDoc?Location=U2&doc=GetTRDoc.pdf&AD=ADA422198.

McGrath, B., McKinley A., Duistermaat, M., Carlander, O., Brill, C., Zets, G., & van Erp, J. (2008). Tactile Actuator Technology. In J. van Erp & B. Self (Eds.), *Tactile displays for orientation, navigation and communication in air, sea and land environments,* (pp.4-1 – 4-12). Neuilly-sur-Seine Cedex, France: Research and Technology Organisation, North Atlantic Treaty Organisation.

McKinley R., Lanning, C., Simmons, C., & Gallimore, J. (2005). Tactile cueing for target acquisition and identification. In *Proceedings of the 43rd SAFE Symposium*. Salt Lake City, Utah October 24-26.Creswell: SAFE.

Middlebrooks, J., & Green, D. (1991). Sound localisation by human listeners, *Annual Review of Psychology,Vol. 42,* pp. 135-159.

Moore, B. (2003). *An introduction to the psychology of hearing* (5th ed). San Diego: Elsevier.

Møller, H., Sørensen, M., Jensen, C., & Hammershøi, D. (1996). Binaural technology: Do we need individual recordings? *Journal of the Audio Engineering Society, Vol. 44,* pp. 451-469.

Navigation In Fogs (1880). In Scientific American - Weekly journal of practical information, art, science, mechanics, chemistry, and manufactures. Retrieved from http://www.gutenberg.org/files/38482/38482-h/38482-h.htm.

Nelson, T., Bolia, R., Ericson, M., & McKinley, R. (1999). Spatial audio displays for speech communications: a comparison of free field and virtual acoustic environments. In *Proceedings of the Human Factors and Ergonomics Society 43rd annual meeting*, Houston September 27-October 1, (pp. 1202-1205). Santa Monica: Human Factors and Ergonomics Society.

Nelson, W., Bolia, R., Ericson, M., & McKinley, R. (1998). Monitoring the simultaneous presentation of spatialized speech signals in a virtual acoustic environment. In *Proceedings of the 1998 IMAGE Conference,* Chandler, AZ, (pp. 159-166). Chandler: The IMAGE Society, Inc.

Nelson, W., Bolia, R., & Tripp, L. (2001). Auditory localisation under sustained +Gz acceleration. *Human Factors, Vol. 43,* pp. 299-309.

Ohuchi, M., Iwaya, Y., Suzuki, Y., & Munekata, T. (2005). Training effect of a virtual auditory game on sound localization ability of the visually impaired. In *Proceedings of ICAD 11th Meeting of the International Conference on Auditory Display,* Limerick July 6-9, (pp. 283-286). Limerick: International Conference on Auditory Displays.

Oskarsson, P.-A., Eriksson, L., & Carlander, O. (2012). Enhanced perception and performance by multimodal threat cueing in simulated combat vehicle. *Human Factors, Vol. 54,* pp. 122-137.

Pettitt, R., Redden, E., & Carstens C. (2006). *Comparison of army hand and arm signals to a covert tactile communication system in a dynamic environment.* Retrieved from http://www.dtic.mil/cgi-bin/GetTRDoc?Location=U2&doc=GetTRDoc.pdf&AD=ADA453363.

Rayleigh, L. (1907). In our perception of sound direction. *Philosophical Magazine, Vol. 13,* pp. 214-232.

Redden, E. (2006). Findings of a multi-year investigation of tactile and multi-modal displays. In *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting,* San Fransisco October 16-20, (pp. 1697-1700). Santa Monica: Human Factors and Ergonomics Society.

Redden, E., Carstens, C., Turner, D., & Elliott, L. (2006). *Localization of tactile signals as a function of tactor operating characteristics.* Retrieved from http://www.dtic.mil/cgi-bin/GetTRDoc?Location=U2&doc=GetTRDoc.pdf&AD=ADA457062

Rupert, A. (2000). An instrumentation solution for reducing spatial disorientation mishaps. *Engineering in Medicine and Biology Magazine, IEEE. Vol 19,* pp. 71-80.

Rupert, A., Guedry, F., & Reshke, M. (1994). The use of a tactile interface to convey position and motion perceptions. In *Virtual Interfaces: Research and Applications,* (pp. 21.1-21.5). Neuilly-sur-Seine, France: Advisory Group for Aerospace Research and Development. In *Proceedings of the HFM-123/RSY Symposium, New directions for improving audio effectiveness*, Amersfoort April 11-13, (pp. 29-1 – 29-30). Amersfoort: North Atlantic Treaty Organisation, Research and Technology Organisation.

Rupert, A., McGrath, B., & Griffin, M. (2002). A tool to maintain spatial orientation and situation awareness for operators of manned and unmanned aerial vehicles and other military motion platforms. In *NATO RTO Symposium on Spatial Disorientation in Military Vehicles: Causes, Consequences and Cures*, La Coruña 15-17 April, (pp. 31-1 – 31-15). Neuilly-sur-Seine Cedex, France: Research and Technology Organisation, North Atlantic Treaty Organisation.

Sacks, O. (2007). *Musicophilia: Tales of music and the brain*. New York: Knopf.

Sandberg, S., Håkansson, C., Elmqvist, N., Tsigas, P., & Chen, F. (2006). Using 3D audio guidance to locate indoor static objects. In *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting,* San Fransisco October 16-20, (pp. 1581-1584). Santa Monica: Human Factors and Ergonomics Society.

Sanders, M., & McCormick, E. (1993). *Human factors in engineering and design* (7th ed). Singapore: McGraw-Hill.

Sartain, D. (2000). *3D Audio programming*. Haverford: Infinity Publishing.

Scott, J., & Gray, R. (2007). Comparison of driver brake reaction times to multimodal rear-end collision warnings. In *Proceedings of the Fourth International Driving*

Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, Stevenson, Washington July 9-12, (pp. 285–291). Washington: TRB.

Shankar, R., Ghosh, T., & Spontak, R. (2007). Dielectric elastomers as next-generation polymeric actuators. *Soft Matter, Vol. 3*, pp. 1116–1129.

Sibbald, A. (n.d., a). *Virtual audio for headphones*. Retrieved August 30, 2005, from http://www.sensaura.co.uk/whitepapers/index.php.

Sibbald, A. (n.d., b). *Virtual ear technology*. Retrieved August 30, 2005, from http://www.sensaura.co.uk/whitepapers/index.php.

Simpson, B., Brungart, D., Gilkey, R., Cowgill, J., Dallman, R., Green, R., Youngblood, K., & Moore, T. (2004). 3D audio cueing for target identification in a simulated flight task. In *Proceedings of the Human Factors and Ergonomics Society 48th annual meeting*, New Orleans September 20-24, (pp. 836-1840). Santa Monica, CA: Human Factors and Ergonomics Society.

Sherrick, C., Cholewiak, R., & Collins, A. (1990). The localization of low- and high-frequency vibrotactile stimuli. *Journal of the Acoustical Society of America*, Vol. *88*, pp. 169-178.

Shinn-Cunningham, B. (2001). Creating three dimensions in virtual auditory displays. In *Proceedings of the HCI International 9th International Conference on Human-Computer Interaction*, New Orleans August 5-10, (pp. 604-608). New Orleans: HCI.

Shinn-Cunningham, B., Durlach, N., & Held, R. (1998). Adapting to supernormal auditory localisation cues:I. Bias and resolution. *Journal of the Acoustical Society of America, Vol. 103*, pp. 3656-3666.

Sodnik, J., Dicke, C., Tomaič, S., & Billinghurst, M. (2008). A user study of auditory versus visual interfaces for use while driving. *International Journal of Human-Computer Studies Archive. Vol. 66*, pp. 318-332.

Spence, C., & Ho, C. (2008). Tactile and multisensory spatial warning signals for drivers. *IEEE Transactions on Haptics, Vol. 1*, pp. 121–129.

Spence, C., & McDonald, J. (2004). The cross-modal consequences of the exogenous spatial orienting of attention. In Calvert, G., Spence, C. & Stein, B. (Eds.), *The handbook of multisensory processes*, (pp. 3-25). London: The MIT Press.

Stanney, K., Samman, S., Reeves, L., Hale, K., Buff, W., Bowers, C., Goldiez, B., Nicholson, D., & Lackey, S. (2004). A Paradigm shift in interactive computing: deriving multimodal design principles from behavioral and neurological foundations. *International Journal of Human-Computer Interaction. Vol. 17*, pp. 229-257.

Stenmark, J. (2001). *Känslan man har*. Stockholm: Galago Förlag.

Stolle, A., Hölzl, R., Kleinböhl, D., Mrsic, A., & Tan, H. (2004). Measuring point localisation errors in spatiotemporal tactile stimulus patterns. In *Proceedings of EuroHaptics 2004*, München 5-7 June, (pp. 512-515). München: Eurohaptics.

Thurlow, W., & Rosenthal, T. (1976). Further study of existence regions for the "ventriloquism effect". *Ear and Hearing, Vol. 1*, pp. 280-286.

Toor, A. (2011). *Pepsi's 'sound of football' project lets visually impaired players see the pitch with their ears*. Retrieved from http://www.engadget.com/2011/11/02/pepsis-sound-of-football-project-lets-visually-impaired-playe/.

Trapenskas, D., & Johansson, Ö. (1999). Localisation performance of binaurally recorded sounds with and without training. *International Journal of Industrial Ergonomics, Vol. 27*, pp. 405-410.

van Erp, J. (2007). *Tactile displays for navigation and orientation: perception and behaviour* (PhD Thesis). Utrecht University, Utrecht.

van Erp, J. (2005a). Vibrotactile spatial acuity on the torso: effects of location and timing parameters. In *Proceedings of World haptics*, Pisa March 18-20, (pp. 80-85). Pisa: IEEE Computer Society.

van Erp, J. (2005b). Presenting directions with a vibrotactile torso display. *Ergonomics, Vol. 48*, pp. 302-313.

van Erp, J. (2002). Guidelines for the use of vibro-tactile displays in Human computer interaction. In *Proceedings of Eurohaptics 2002*, Edinburgh July 8-10, (pp. 18-22). Edinburgh: Eurohaptics.

van Erp, J., Eriksson, L., Levin, B., Carlander, O., Veltman, J., & Vos, W. (2007). Tactile cueing effects on performance in simulated aerial combat with high acceleration. *Aviation, Space, and Environmental Medicine, Vol. 78,* pp. 1128-1134.

van Erp J., & Self, B. (2008). Introduction to tactile displays in military environments. In by J. van Erp & B. Self (Eds.), *Tactile Displays for Orientation, Navigation and Communication in Air, Sea and Land Environments,* (pp.1-1 – 1-18). Neuilly-sur-Seine Cedex, France: Research and Technology Organisation, North Atlantic Treaty Organisation.

van Erp, J., & Spapé, M. (2003). Distilling the underlying dimensions of tactile melodies. In *Proceedings of Eurohaptics 2003*, Dublin July 6-9, (pp 111-120). Dublin: Eurohaptics.

van Erp, J., & Van Veen, H. (2003). A multi-purpose tactile vest for astronauts in the International Space Station. In *Proceedings of Eurohaptics 2003*, Dublin July 6-9, (pp. 405-408). Dublin: Eurohaptics.

van Erp, J., van Veen, H., Jansen, C., & Dobbins, T. (2005). Waypoint navigation with a vibrotactile waist belt. *ACM Transactions on Applied Perception, Vol. 2*, pp. 106-117.

van Erp, J., Veltman, J., van Veen, H., & Oving, A. (2003). Tactile torso display as countermeasure to reduce night vision goggles induced drift. In *Proceedings of NATO RTO HFM panel symposium on spatial disorientation in military vehicles: Causes, consequences and cures,* (pp. 49.1–49.8). Neuilly-sur-Seine, France: NATO Research and Technology Organisation.

van Erp, J., & Verschoor, M. (2004). Cross-modal visual and vibro-tactile tracking. *Applied Ergonomics, Vol. 35,* pp. 105-112.

van Erp, J., & Werkhoven, P. (2006). Validation of principles for tactile navigation displays. In *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting,* San Francisco, October 16-20, (pp. 1687-1691). Santa Monica, CA: Human Factors and Ergonomics Society.

van Veen, H., & van Erp, J. (2003). Tactile melodies in user interfaces. In *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting,* Denver October 13-17, (pp. 751- 754). Santa Monica, CA: Human Factors and Ergonomics Society.

van Veen H., & van Erp J. (2000). Tactile Information Presentation in the Cockpit. *Lecture Notes in Computer Science, Vol. 2058,* pp. 174 - 181

Vause, N., & Grantham, D. (1998). Speech intelligibility in adverse conditions in recorded virtual auditory environments. In *Proceedings of the International Conference on Audio Displays 1998*, Glasgow November 1-4, (pp. 1-13) Glasgow: ICAD.

Vazquez-Alvarez, Y., & Brewster, S. (2009). Investigating background & foreground interactions using spatial audio cues. In *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems*, Boston April 04-09, (pp.3823-3828). Boston: CHI.

Veltman J., Oving A., & Bronkhorst A. (2004). 3-D Audio in the fighter cockpit improves task performance. *International Journal of Aviation Psychology, Vol. 14*, pp. 239-256.

Verrillo, R., & Gescheider, G. (1975). Enhancement and summation in the perception of two successive vibrotactile stimuli. *Perception & Psychophysics*, Vol. *18*, pp. 128-136.

Vitense, H., Jacko, J., & Emery, V. (2002). Foundation for improved interaction by individuals with visual impairments through multimodal feedback. *Universal Access in the Information Society, Vol. 2,* pp 76-87.

Väljamäe, A., Tajadura-Jiménez, A., Larsson, P., Västfjäll, D., & Kleiner, M. (2008). Binaural bone-conducted sound in virtual environments: Evaluation of a portable, multimodal motion simulator prototype. *Acoustical Science and Technology. Vol. 29,* pp. 149-155.

Walker, B., & Lindsay, J. (2005a). Development and evaluation of a system for wearable audio navigation. In *Proceedings of the 49th Annual Meeting of the Human Factors and Ergonomics Society,* Orlando September 26-30, (pp. 1607-1609). Santa Monica, CA: Human Factors and Ergonomics Society.

Walker, B., & Lindsay, J. (2005b). Navigation performance in a virtual environment with bonephones. In *Proceedings of the 11th Meeting of the International Conference on Auditory Display,* Limerick 6-9 July, (pp. 260-263). Limerick: International Conference on Auditory Displays. Limerick :ICAD

Walker, B., & Lindsay, J. (2004). Auditory navigation performance is affected by waypoint capture radius. In *Proceedings of the 10th Meeting of the International Conference on Auditory Display*, Sydney July 6-10. Sydney: ICAD

Walker, B., Stanley, R., Iyer, N., Simpson, B., & Brungart, D. (2005) Evaluation of bone-conduction headsets for use in multitalker communication environments. In *Proceedings of Human Factors and Ergonomics Society Annual Meeting,* Orlando September 26-30, (pp.1615-1619). Santa Monica, CA: Human Factors and Ergonomics Society.

Wenzel, E. (1999). Effect of increasing the latency on localization of virtual sound sources. In *Proceedings of the AES 16th International Conference on Spatial Sound Reproduction,* Rovaniemi, April 10-12. (pp. 42-50). Rovaniemi: AES.

Wenzel, E. (1992). Localisation in virtual acoustic display. *Presence, Vol. 1,* pp. 80-107.

Wenzel, E., Arruda, M., Kistler, D., & Wightman, F. (1993). Localisation using non-individualized head-related transfer functions. *Journal of the Acoustical Society of America, Vol. 94,* pp. 111-123.

Wenzel, E., Wightman F., & Foster, S. (1988). A Virtual system for conveying three-dimensional acoustic information. In *Proceedings of the 32nd Annual Meeting of the Human Factors and Ergonomics Society,* Anaheim October 24-28, (pp. 86–90). Santa Monica, CA: Human Factors and Ergonomics Society.

Wickens, C. (2008). Multiple resources and mental workload. *Human Factors, Vol. 50,* pp. 449–455.

Wickens, C. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomic Science,Vol. 3, pp.* 159-177.

Wickens, C. (1987). Attention. In Hancock, P. (Ed.), *Human factors Psychology.* (pp.29-79). Amsterdam: North-Holland.

Wickens, C., & Hollands, J. (2000). *Engineering psychology and human performance.* Upper Saddle River: Prentice Hall.

Wightman, F., & Kistler, D. (1999). Resolution of front-back ambiguity in spatial hearing by listener and source movement. *Journal of the Acoustical Society of America, Vol. 105*, pp. 2841-2853.

Wightman, F., & Kistler, D. (1993). Sound localization. In Yost, W., Popper, A. & Fay, R. (Eds.), *Springer Handbook of Auditory Research: Human Psychophysics*. (pp. 155-192). New York: Springer-Verlag.

Wightman, F., & Kistler, D. (1990). Hearing in three dimensions: sound localisation. In *Proceedings of the Audio Engineering Society 8th International Conference*, Washington DC May 3-6, (pp. 21-27). New York: AES Press.

Wilson, T. (2012). *How virtual surround sound works.* Retrieved from http://electronics.howstuffworks.com/virtual-surround-sound2.htm.

Withington, D. (2000). The use of directional sound to improve the safety of auditory warnings. In *Proceedings of the HFES 44th Annual Meeting, in conjuction with the International Ergonomics Association XIVth Triennial Congress*, San Diego July 29-August 4, (pp. 726-729). Santa Monica, CA: Human Factors and Ergonomics Society.

Zhang, P., & Hartmann, W. (2006). Lateralization of sine tones–interaural time vs phase. *Journal of Acoustical Society of America, Vol. 120*, pp. 3471–3474.

Appendix A, Tactile actuators a comparison

A comparison of common tactile actuators (tactors). Adopted from McGrath et al. (2008).

	Electrical	**Vibro-Mechanical**		
		Rotary Motion	*Linear Actuator*	*Pneumatic*
Size	The tactors can be made to be any size desired.	Typically a cylinder 14 mm long dia. 6 mm	contact surface 20 – 30 mm	8 – 10 mm thick
Weight	Less than 5g, but variable depending on size of tactor	Typical motor 5 – 10g (housing dependent)	5 – 20g	0.8 – 2.0g
Frequency	8 – 500+ Hz	typically up to 160 Hz	1 – 300Hz, Optimized 200 – 300 Hz	0 – 100Hz, Optimized 50 Hz
Displacement/ Amplitude	Not applicable	Mounting dependent for tactor	up to 10 mm peak	3mm peak
Onset Time	Negligible/Immediate	50 – 90 msec to full amplitude	5 – 30 msec	<20msec
Axis	Not applicable	Mainly rotary	Perpendicular	Perpendicular
Power	3 – 130 volts, power based on impedance of skin	0.05 – 0.2 W	0.2 – 1.0 W	11 W (Compressor)
Sensory receptor	Capable of exciting all 6 identified tactile nerve receptors	All mechanoreceptors sensitive to vibration	All mechanoreceptors sensitive to vibration	Mechanoreceptors sensitive to low frequency vibration
Punctateness*	High – Very High	Medium, housing dependent (larger housing – more diffuse)	Dependent: high with protruding parts, medium with vibrating housing	Medium

	Electrical		**Vibro-Mechanical**	
		Rotary Motion	*Linear Actuator*	*Pneumatic*
Cost (very rough)	$0.25 – $1.00, depending on conductive material	Typical pager motor <$20. Tactor cost $10 –$100, housing dependent	$ 50 – 200	$60 includes tactor, valve. Does not include compressor
Signature	None	High, housing dependent	Slight to moderate audio signature, Magnetic	Slight to moderate audio signature
Material (Outer)	Conductive rubber or tin	PVC and other polymers, aluminium, fibreglass	Aluminium, Polyurethane (Polymer)	Plastic, latex
Limitations	Variability of sensation, must directly contact skin, user acceptance issues	Lack of independent amplitude and frequency control. Weight, power consumption	Weight, power consumption	Support equipment (valves, compressors)
Examples	EXTENSOR System by Johnson Kinetics, Inc.	TNO tactor, FOI tactor	Steadfast technologies. C2 Tactor (Engineering Acoustics, Inc.), VWB32 (Audiological Engineering)	Steadfast Technologies PS2

***Punctateness**[18]

None: A generalized vibration with little localization.
Low: Diffuse sensation with average localization.
Medium: Discrete sensation, with good localization.
High: Point-like sensation, easily locatable.
Very High: Pin prick, mildly painfully.

---

[18] This word was used in the NATO HFM-122 group to explain how precise (or small) the perceptual experience is from a tactile stimulus on the skin. Thus it refers to the effects of how a tactile stimulus spreads, or not spreading over the skin.