Handling Uni- and Multimodal Threat Cueing with Simultaneous Radio Calls in a Combat Vehicle Setting

Otto Carlander[2], Lars Eriksson[1], and Per-Anders Oskarsson[1]

[1] FOI, Swedish Defence Research Agency
Olaus Magnus väg 42, 581 11 Linköping, Sweden
[2] Now at Motorola, Inc. IPVS, Sweden
otto.carlander@motorola.se

Abstract. We investigated uni- and multimodal cueing of horizontally distributed threat directions in an experiment requiring each of twelve participants to turn a simulated combat vehicle towards the cued threat as quickly and accurate as possible, while identifying simultaneously presented radio call information. Four display conditions of cued threat directions were investigated; 2D visual, 3D audio, tactile, and combined cueing of 2D visual, 3D audio, and tactile. During the unimodal visual and tactile indications of threat directions an alerting mono sound also was presented. This alerting sound function was naturally present for the unimodal 3D audio and multimodal conditions, with the 3D audio simultaneously alerting for and cueing direction to the threat. The results show no differences between conditions in identification of radio call information. In contrast, the 3D audio generated greater errors in localization of threat direction compared to both 2D visual and multimodal cueing. Reaction times to threats were also slower with both the 3D audio and 2D visual compared to the tactile and the multimodal, respectively. In conclusion, the results might reflect some of the benefits in employing multimodal displays for certain operator environments and tasks.

Keywords: Display technologies, Multimodal, Combat Vehicle, Simulation.

1 Background

The warning and countermeasure system (WCS) is, together with radio communication and command and control systems, of vital importance in a combat vehicle. The WCS automatically performs various countermeasures, thus saving critical time in taking action towards threats or attacks. Though the WCS is automatic the operator crew still needs to have an awareness of evolving threat situations. Carlander and Eriksson [1] showed that the driver of a Combat Vehicle 90 (CV 90) could efficiently utilize threat cueing made by 3D audio and tactile displays combining temporal and spatial positioning. The 3D audio made "threat sounds" possible to localize and a tactile belt around the torso made it possible to feel directions to "threat vibrations." Responses to the cued horizontal directions to threats were made by turning the CV 90 as to having the vehicle heading pointing towards indicated threats, and the responses were also required to be made as quickly as

C. Stephanidis (Ed.): Universal Access in HCI, Part II, HCII 2007, LNCS 4555, pp. 293–302, 2007.

possible. The overall driver performance was of high-quality in that the displays facilitated swift and accurate performance with good threat awareness. However, the 3D audio display generated greater errors and reaction times in localizing threats straight behind the vehicle compared to the tactile display and the two displays combined, respectively. That is, some front-back confusions with the 3D audio cueing were neutralized with the addition of the tactile cueing and not present with the tactile only cueing condition.

The ”multiple resource theory” [2] suggests that we have independent resources for information processing and that some of these can be accessed simultaneously (parallel processes). However, parallel processeing can be interfered by handling the information by the same modality, while on the other hand multisensory processes often support and complement each other:

”… 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 …”

([3], p. 3-4)

It is therefore important to consider the different sensory capabilities when designing complex information systems. Combining sensory signals reduces the variance of perceptual estimates, thus enhancing stimuli detection by cue redundancy. For instance, it has been shown that multisensory neurons are maximally activated by temporally and spatially synchronized visual, auditory, and tactile events [4]. Based on this it is easy to hypothesize that multisensory interfaces most probably improve the presentation of cues and information transfer.

Various presentation technologies now offer solutions for effectively presenting information to all our senses, also including smell and taste, e.g. [5]. However, apart from the obvious visual displays, tactile, and auditory displays are for more straightforward applications the most developed [6]. Some very useful applications of tactile displays utilize vibrations on the torso to cue directions to “the outside world.” These directions are mapped to body coordinates by localized stimulations of the skin through small vibrating motors called tactors [1]. The perception of the stimulation on the skin can thus be considered analogous to the perception of auditory or visual stimulation. The tactile technology is well developed and has been studied in various operational settings, e.g. [7], [8], [9], [10], [11].

3D audio sound systems have been able to deliver high-quality, localizable sounds for quite some time [12]. Furthermore, some commercials off the shelf (COTS) alternatives are now available that even have the potential of delivering high quality 3D audio comparable to more expensive and bulky research engines [13]. Carlander and Eriksson [1] utilized a COTS system in a combat vehicle environment, and proposed that the occurred front-back ambiguities might decrease by tracking both the operator's head and the chassis of the vehicle, instead of just the vehicle, and that an improved acoustic signal would probably facilitate sound localization ability. It was also suggested that in the combined displays condition, the tactile cueing might have generated a "compelling and unambiguously perceived threat direction, leaving no room for improvement with 3D audio."

In the present study, we investigated how uni- and multimodal cueing of threat directions compare in a simulated CV 90 setting, with the vehicle driver handling the threats and incoming radio calls. A CV 90 is equipped with a 2D visual head-down display for cueing of directions to threats, mono sound for threat alerts, and a mono system for radio communication. Thus, the 2D visual display and mono sound used here were based on that presently used in a CV 90, and the tactile and 3D audio threat cueing were based on technology that easily can be implemented.

2 Methods

2.1 Participants

Twelve male students participated with a mean age of 23.0 years, ranging from 21 to 29 years. All had normal sight and hearing and no prior experience of the presentation techniques.

2.2 Apparatus

A PC equipped with an external USB soundcard, Hercules Gamesurround MUSE Pocket, was used for the threat cueing and radio communication that were presented through a pair of Sehnheiser HD 200 headphones with circumaural design. The tactile cueing was delivered by a FOI tactile torso belt and controlled from a small signal box connected to the PC parallel port. Twelve vibrating elements in the belt are based on 'clock-positions' evenly distributed on the torso, and each vibrating element thus covers 30° of the horizontal dimension. A head-down visual display was used for the 2D visual threat warnings, and a 42" plasma screen was used for the simulated terrain environment and for feedback on threat defeat status. A touch screen presented response options and was used for identification of radio call information. To compensate for head movements in the 3D audio presentation, an Ascension LaserBIRD II headtracker was used . The base for the vehicle simulation was a 6 DOF Moog motion platform, as shown in Figure 1.

The vehicle simulation was delivered from an FOI developed simulation engine - "the HiFi engine." Data from the HiFi engine was sent to a server handling threat information and presentation, with another data stream sent to the motion platform giving the exact coordinates for the simulation.

Fig. 1. The motion platform with a seated driver. At (A) is the screen presenting 2D visual information; at (B) is the out-the-window view display; and at (C) is the display presenting response options for identification of radio call information.

2.3 Design, Stimuli, and Procedure

The experiment had a 4 × 3 factorial within subjects design. It included four display conditions of threat directions cueing; 2D visual, 3D audio, tactile, and the combination of 2D visual, 3D audio, and tactile (multimodal cueing). During the unimodal visual and tactile indications of threat directions an alerting mono sound was also presented. This alerting sound function was naturally present for the unimodal 3D audio and the multimodal conditions, with the 3D audio simultaneously alerting for and cueing direction to the threat. Thus, while cueing of threat directions involved unimodal visual, auditory, and tactile displays, and the multimodal combination of these three displays, the inclusion of a mono sound with the unimodal visual and tactile displays kept the alerting function of sound presentation. That is, it is not likely that an operational cueing in a combat vehicle of threat presence would exclude sound presentation so its alerting property was kept for all display conditions. Cueing of threat directions was nevertheless unimodal in three of the display conditions. See Table 1. Threat pop-ups occurred in three sectors presented at the front (315°-45°), the side (45°-135° and 225°-315°), and the back (135°-225°), relative initial vehicle heading.

The task was to react to each threat and respond to it as quickly as possible by aligning the vehicle heading with the threat position and pushing a trigger button. Threat warnings presented with audio consisted of a beep normally used for laser warnings in a CV 90.

During the visual and tactile display conditions the radio calls were given in mono. During the 3D audio and multimodal display conditions the radio calls were given in 3D audio. The 2D visual presentation was shown on a head-down display showing a top-view of the vehicle indicated by a red cone (15 degrees wide) extending from the center of the own vehicle. The response was considered correct if the vehicle was aligned ±10° within the threat position. The threats were not visible, but after each

Table 1. Cueing of threat directions and sound alert

Cueing of Threat Direction	Sound Alert
2D Visual	Mono Sound
Tactile	Mono Sound
3D Audio	3D Audio (same as cueing of threat direction)
Multimodal (Visual, Tactile, 3D Audio)	3D Audio (same as cueing of threat direction)

threat response a text message “hit” or “miss” was overlaid on the simulated out-the-window view (Figure 1). Each threat presentation was limited to duration of 20 s or was terminated with the operator response. The time between threat response and the presentation of a new threat was randomized in the interval of 15 to 20 s. The horizontal vehicle deviation from the correct threat position at operator response was defined as localization error (LE). The elapsed time from target pop-up to having the vehicle turned ±10º from initial heading, was defined as reaction time (RT).

The participants performed a secondary task by identifying two simultaneous radio calls, given at each threat pop up. The participants were told to respond to radio calls on a touch-screen with six response buttons. That is, for each of the two simultaneous calls there were three possible responses to make on the touch-screen.

The participants were told to prioritize the main task, aligning the vehicle, but to answer the radio calls correctly and as fast as possible. The elapsed time from radio call presentation to touch screen-indication was defined as performance time (PT). Proportion of correct answered radio calls was also collected.

A block of trials consisted of using one of the display conditions, and included eight threat presentation trials in each of the three threat sectors, resulting in a total of 24 trials per participant and block. The presentation order of threats was randomized within each block of trials, and block order was balanced over participants.

Before the experiment a questionnaire was completed that included a check on the physical requirements such as normal sight, hearing, and present health condition. In between blocks of trials, questionnaires concerning the display techniques were answered, comprising perceived interpretability, mental workload, and perception of threat direction. A summarizing questionnaire was completed at the very end of the experiment. Participants responded to the questions on a seven points scale.

3 Results

Each analysis comprised twelve means (4 display conditions × 3 sectors = 12 conditions) 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 PTs and proportion of correctly answered radio calls. All ANOVA *p*-values are given with the Greenhouse-Geisser corrected values.

3.1 Localization Error (LE)

The ANOVA showed a significant main effect of display condition, $F(3, 33) = 11.53$, $p< .01$, Figure 2, with no other significant effects. A Tukey HSD post hoc test showed a larger mean LE for the 3D audio condition with mean (M) = 10.1 (SE = 2.2), as compared with both the 2D visual ($p < .001$), M = 2.8 (SE = 0.4), and the multimodal presentation ($p < .001$), M = 3.0 (SE = 0.4).

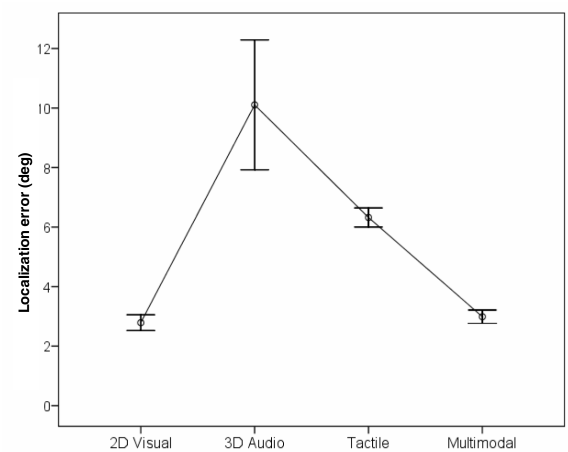

Fig. 2. Mean localization error (LE) with each display condition (±SE)

3.2 RT to Threats

The ANOVA showed a significant main effect of display condition, $F(3, 33) = 7.89$, $p< .001$, with no other significant effects (Figure 3). A Tukey HSD post hoc test

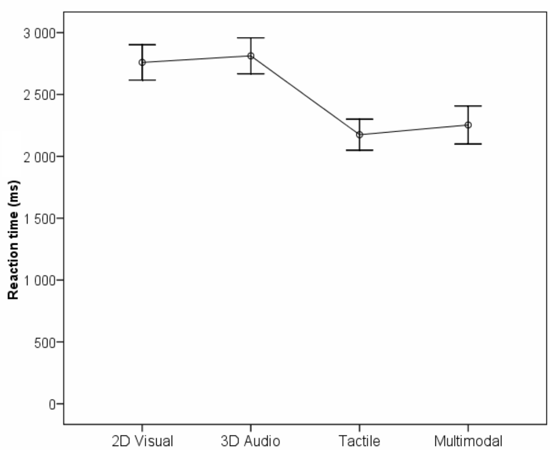

Fig. 3. Mean threat reaction time (RT) with each display condition (±SE)

showed a greater mean RT for both the 2D visual, M = 2759 (SE = 227), and the 3D audio display, M = 2812 (SE = 242), as compared with the tactile ($p < .01$), M = 2175 (SE = 204), and the multimodal presentation ($p < .05$), M = 2253 (SE = 229), respectively.

3.3 Performance Time (PT) to Identification of Radio Calls

The ANOVA showed a significant main effect of display condition for responses of radio calls $F(3, 33) = 4.00$, $p < .05$, with no other significant effects (Figure 4). A Tukey HSD post hoc test showed a greater mean PT for the 3D audio, M = 8965 (SE = 1104) compared to the tactile, with M = 7355 (SE = 1075) display condition ($p < .025$).

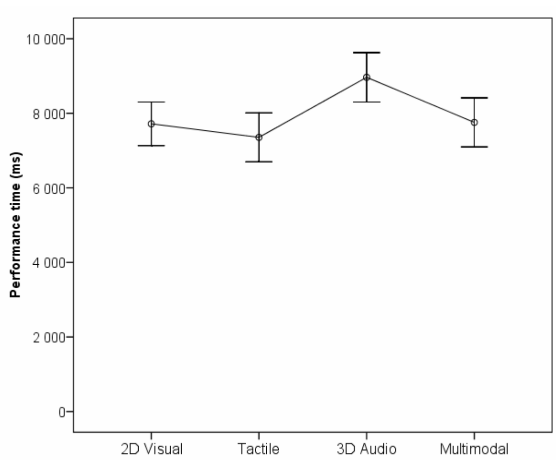

Fig. 4. Mean performance time of responses to radio calls for each display condition (±SE)

3.4 Proportion Correct Identification of Radio Calls

The ANOVA showed no significant effects. The mean values for proportion correct responses to radio calls varied between .82 and .85 (SE: .035 – .04).

3.5 Subjective Ratings

The 2D visual (M = 6.2, SE = 0.4) and multimodal (M = 6.1, SE = 0.3) displays were considered easiest to interpret. The tactile display (M = 5.5, SE = 0.4) was considered somewhat less interpretable, and the 3D audio display was considered most difficult to interpret (M = 4.1, SE = 0.4) (1 = not at all, 7 = very much). Only small differences where found for the level of mental workload, 2D visual (M = 3.8, SE = 0.4), tactile (M = 4.0, SE = 0.5), multimodal (M = 3.8, SE = 0.3), and somewhat higher for 3D audio (M = 4.4, SE = 0.4) (1 = very low, 7 = very high).

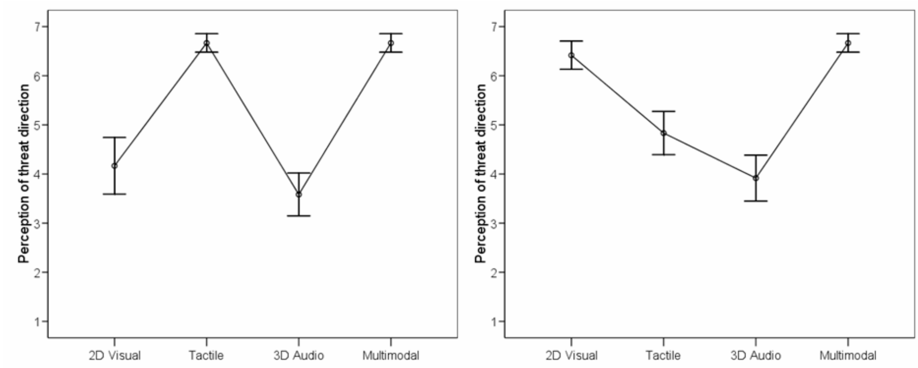

Fig. 5. Ratings of perception of threat direction with the display conditions. Left, localization phase (initial) at threat pop up. Right, alignment phase (final) (1 = Not at all obvious, 7 = Very obvious).

Initial perception of threat direction, was considered best for the tactile (M = 6.7, SE = 0.2) and multimodal (M = 6.7, SE = 0.2) displays, and lower for the 2D visual (M = 4.2, SE = 0.6) and 3D audio (M = 3.6, SE = 0.4) displays (1 = not at all obvious, 7 = very obvious) (Figure 5, Left). Other studies [11] also show that tactile cueing has a very strong attentional effect. In the final stage, the alignment phase, there were high ratings for the 2D visual (M = 6.4, SE = 0.3) and multimodal (M = 6.7, SE = 0.2) displays, and lower ratings for the tactile (M = 4.8, SE = 0.4) and 3D audio (M = 3.9, SE = 0.5) displays (1 = not at all obvious, 7 = very obvious) (Figure 5, Right).

4 Conclusions and Discussion

The results show that the 3D audio generated greater LE compared to the 2D visual and multimodal displays. Also, the RT with the 3D audio was slower compared to the tactile and the multimodal presentation, respectively. The RT with the 2D visual display was also slower compared to the tactile and the multimodal displays, respectively. The PTs for radio calls were slower for 3D audio compared to the tactile condition.

The participants performed well with the different presentation techniques, even though very little training was administered. The greater LE and slower RT for the 3D audio display cannot solely be explained by a few front – back confusion occurrences. That is, an analysis with the back sector removed showed the same pattern of results; greater LE and slower RT for the 3D audio (with less difference). Tracking both the head of the participant and the vehicle do not seem to dramatically decrease the front-back confusion occurrences. Also, altering the frequency spectra of the 3D signal do not seem to increase 3D audio performance in the combat vehicle.

The high accuracy of the 2D visual display confirms how well vision works for an in-vehicle 2D display. However, the slower RTs might indicate that the visual display is relatively poor in getting the operators attention. This is also reflected in the subjective ratings (Figure 5). Furthermore, analysis of the head-tracker data (not presented here) shows the participants used the 2D visual display more frequently when not supported by the added displays in the multimodal display condition. Thus, important time for inspection of the surrounding terrain (out-the-window-time) can be lost when using the 2D visual display. Of course, utilizing a head-down display for threat cueing is not the first choice for capturing attention. Having the visual threat cue superimposed on the out-the-window view would instead be preferable, thus making it more comparable to the tactile and 3D audio cueing.

The results are in line with the general results from Carlander and Eriksson [1] regarding the threat response performance with the 3D audio and tactile displays. The combination of sensory modalities seems to reduce localization deviations from threats or targets and decrease RTs. Also, subjective ratings from both studies show that participants perceived combination of modalities as intuitive and efficient. Combining sensory modalities in a combat vehicle setting could improve the threat awareness, or situational awareness, and overall performance. In essence, the multimodal condition of the present study has the advantage of combining the precision in the visual display when localizing threats and the attentional capture effect of the tactile display thus speeding up RT. In conclusion, the results might reflect some of the benefits in employing multimodal displays for certain operator environments and tasks.

Acknowledgements

We acknowledge the considerate cooperation of the Swedish Army Combat School in Skövde, and above all the WCS group and Jan Fredriksson for making important contributions. We also acknowledge the assistance from the technical group at FOI MSI lab, Björn Lindahl, Johan Hedström and Mattias Kindström. Fang Chen, Chalmers University of Technology, is thanked for providing valuable comments.

References

1. Carlander, O., Eriksson, L.: Uni- and bimodal threat cueing with vibrotactile and 3D audio technologies in a combat vehicle. In: Human Factors and Ergonomics Society. Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting, pp. 16–20. Santa Monica, CA (2006)
2. Wickens, C.D., Hollands, J.G.: Engineering psychology and human performance. Prentice Hall, Upper Saddle River, NJ (2000)
3. Spence, C., McDonald, J.: The cross-modal consequences of the exogenous spatial orienting of attention. In: Calvert, G.A., Spence, C., Stein, B.E. (eds.) The handbook of multisensory processes, pp. 3–25. The MIT Press, Cambridge, MA (2004)
4. Eimer, M.: Multisensory integration: How visual experience shapes spatial perception. Current biology 14(R), 115–117 (2004)
5. Yanagida, Y., Kawato, S., Noma, H., Tetsutani, N., Tomono, A.: A Nose-Tracked, Personal Olfactory Display. In: Proccedings of the Computer Graphics and Interactive Techniques 30th Conference, San Diego, CA: SIGGRAPH (July 27-31, 2003)
6. Eriksson, L., Carlander, O., Borgvall, J., Dahlman, J., Lif, P.: Operator site 2004-2005. Linköping: FOI (2005)

7. van Erp, J.B.F., van Veen, H.A.H.C., Jansen, C., Dobbins, T.: Waypoint navigation with a vibrotactile waist belt. ACM Transactions on Applied Perception 2(2), 106–117 (2005)
8. McGrath, B.J., Estrada, A., Braithwaite, M.G., Raj, A.K., Rupert, A.H.: Tactile situation awareness system flight demonstration final report. Report No. 2004-10, the US Army Aeromedical Research Laboratory (2004)
9. Rupert, A.H., Guedry, F.E., Reshke, M.F.: The use of a tactile interface to convey position and motion perceptions. In: Virtual Interfaces: Research and Applications (AGARD CP-478: pp. 21.1–21.5). Neuilly-sur-Seine, France: Advisory Group for Aerospace Research and Development (1994)
10. van Erp, J.B.F., Veltman, J.A., Van Veen, H.A.H.C., Oving, A.B.: Tactile torso display as countermeasure to reduce night vision goggles induced drift. In: Proceedings of NATO RTO Human Factors & Medicine Pan el Symposium on Spatial Disorientation in Military Vehicles: Causes, Consequences and Cures (AC/323[HFM-085]TP/42, pp. 49.1–49.8). Neuilly-sur-Seine, France: North Atlantic Treaty Organisation, Research and Technology Organisation (2002)
11. Eriksson, L., van Erp, J., Carlander, O., Levin, B., van Veen, H., Veltman, H.: Vibrotactile and visual threat cueing with high G threat intercept in dynamic flight simulation. In: Human Factors and Ergonomics Society. Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting, pp. 1547–1551. Santa Monica, CA (2006)
12. Wenzel, E., Arruda, M., Kistler, D.J., Wightman, F.L.: Localization using non-individualized head-related transfer functions. Journal of the Acoustical Society of America 94(1), 111–123 (1993)
13. Carlander, O., Eriksson, L., Kindström, M.: Horizontal localisation accuracy with COTS and professional 3D audio display technologies. In: Proceedings of the IEA 2006 conference, The Netherlands: International Ergonomics Society (2006)