A Comparative Study of Multimodal Displays for Multirobot Supervisory Control

Boris Trouvain and Christopher M. Schlick

Forschungsgesellschaft für Angewandte Naturwissenschaften,
Neuenahrer Str. 20, 53343 Wachtberg-Werthoven, Germany
Institute of Industrial Engineering and Ergonomics, RWTH Aachen University,
Bergdriesch 27, 52062 Aachen, Germany
Trouvain@fgan.de, c.schlick@iaw.rwth-aachen.de

Abstract. The supervisory control of ground-based mobile multirobot systems requires to perform multiple concurrent tasks under high levels of time pressure resulting in heavy workload. In this paper we present the design and evaluation of multimodal displays for a particular problem associated with the supervisory control of ground-based multirobot systems: the coordination between the platform specific robot control task, e.g. navigation and obstacle avoidance, and the mission specific payload task. The coordination requires the operator to concurrently monitor and switch attention between the robot control and the payload control tasks depending on the mission requirements. Multimodal human-robot interfaces can significantly support human information processing by communicating information across multiple channels and can therefore improve concurrent task processing. An experiment was designed and carried out with 14 participants which compares four human-robot interface configurations with a simulated two-robot ground-based multirobot system. The results show that the multimodal interfaces perform significantly better across multiple variables and have the lowest workload. Based on our gaze tracking results we can conclude that our multimodal interface has an effect on the visual scanning behaviour in the peripheral regions of the camera display.

Keywords: Human-Robot-Interface, Multirobot, Multimodal.

1 Introduction

Human-multirobot systems are a promising approach to tightly integrate human and artificial intelligence in complex missions. The user interface is an important component of such a flexible system. In order to maintain situation awareness the operator must continuously perform the mental fusion of displayed information. Due to the well-known human limitations in concurrent information processing [15] it is important to strive for novel human-multirobot interface designs. The innovative approach of this paper is to design a true multimodal, single screen interface using interface components that allow the shifting of state information of the robots from the visual to the auditory or tactile channel of the operator. Wickens' multiple resource theory [15] predicts that the division of attention across different modalities

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 184–193, 2007.
© Springer-Verlag Berlin Heidelberg 2007

supports concurrent human information processing and therefore should significantly support operators' mental fusion of information.

Supervisory control of mobile robots can be regarded as a dual task problem: The primary task is the robot control task covering the basic aspects of platform control, e.g. maintaining operational effectiveness, obstacle avoidance, etc. The secondary task is the payload control task which is related to the mission objectives, e.g. to apply sensors for surveillance or search and rescue. However, if one wants to empower a single operator to manage one or two mobile robots, careful consideration of all levels of the human-robot system is necessary. While technological problems of mobile multirobot systems are likely to be solved in the future, the cognitive performance of human operators to effectively control such complex concurrent systems will remain constant. When comparing the technologically feasible levels of concurrency in future multirobot systems and human cognitive performance in "multitasking" there is a strong incompatibility [11].

On the robotic level a sufficient degree of autonomy is a prerequisite for reducing the workload of the operator. Autonomy enables robots to be productive without operator intervention for certain, albeit hardly predictable amounts of time. This particular aspect of autonomy and human-robot interaction is described as neglect tolerance and was extensively studied [3, 10]. For instance, Olsen investigated the relationship between neglect tolerance and effectiveness of human-robot interaction [10]. The interplay between robot autonomy and human-robot interaction was further studied in [5, 9].

A variety of factors for designing effective and efficient human-robot interfaces are introduced by Adams and Goodrich [2, 6]. Although a human-multirobot interface can share and reuse many elements from single robot user interfaces, special attention must be paid to two problems. First, each robot operates within an arbitrary complex context (task, environment etc.) and the contexts are coordinated or independent. In order to effectively supervise the multirobot system the user interface must support the operator in smooth switching between different contexts. Second, a complex set of concurrent tasks must be planned, monitored and modified in order to maintain the system's operational effectiveness.

The problem of how to concurrently work on multiple tasks is a common research question in today's highly automated systems and is investigated in various domains [1, 8]. For instance, findings of Wickens [13, 15] indicate that the visual and auditory system of the human can be regarded as separate information processing resources. This multiple resource approach is particularly interesting for human-robot interaction. Wickens asserts, "… in a heavy visual environment, auditory displays will improve time-sharing performance." and cited 18 studies on dual task performance. The cross-modal and intra-modal approaches were also investigated in a simulation study of the concurrent control of two UAVs [4, 14]. The results revealed positive effects of simultaneous visual and auditory information presentation in subtasks requiring many gaze movements. Helleberg [7] presents a cockpit display study about the compatibility of different types of information with respect to auditory delivery. He found that the auditory-only condition was the least disruptive of ongoing visual tasks.

2 Methodology

2.1 Dual Task Design

The primary goal of this study is to investigate the coordination between the human operator and the autonomous subsystems. To empower the operator when coordinating activities – he or she must decide to either proceed with the mission task or to assist the autonomous subsystems – it is necessary to ergonomically present the required information and to minimize the workload due to attention shifts. Multimodal interfaces allow attention to be divided between, for example, the visual and the auditory channel, and therefore often improve concurrent information processing. In our evaluation we model this situation using a generic dual task design – robot control task and payload control task – as already pointed out in the introduction. The robot control task was an obstacle classification task. In the case of an obstacle that can not be classified by the autonomous navigation system human operator as a supervisory controller has to intervene and interactively assist the autopilot by notifying the autopilot that a water-filled negative obstacle lies in the driving path. The payload control task design is derived from observation and surveillance scenarios which are typical applications of mobile single and multirobot systems. The subjects were informed that both robots of the multirobot system would follow a preprogrammed patrol path. Along this path a number of depressions were placed randomly in parallel to the left or right of the path. These depressions can contain up to two visually hard to detect targets which have to be marked upon detection on screen using the same point-and-click method as in the robot control task. Figure 1 shows an in-simulation screen capture of the experimental setting.

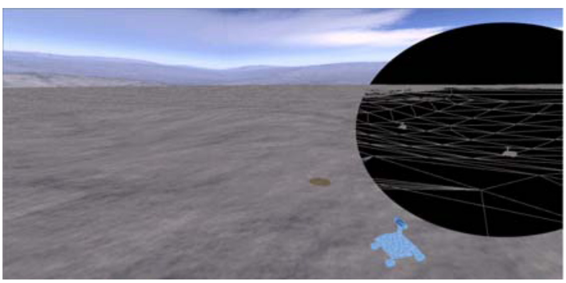

Fig. 1. Screen capture from third-person perspective with superimposed corresponding partial wireframe capture illustrating the task setting. A depression containing two targets can be identified in the wireframe section. Ahead of the robot a water-filled obstacle can be seen.

2.2 User Interface Design

The visual layout of the human-robot interface was inspired by the windscreen-dashboard layout of a car. Since the payload control task in this study is an observation task the upper two thirds of the display space are used for the camera view. The lower third contains the two consoles for two mobile robots (see Figure 2). The design decision to limit the interface to systems with just two robots is based on previous studies concerning the supervisory control of a multirobot system [12]. A characteristic design feature of the user interface is the mutually exclusive access to the camera view. Therefore, the operator has to actively switch between the robot cameras with a keyboard key in order to get access to the current view. The bottom dashboard contains two consoles for the simulated robots. The graphical layout is guided by horizontal symmetry with respect to the console placement (see figure 2 for example).

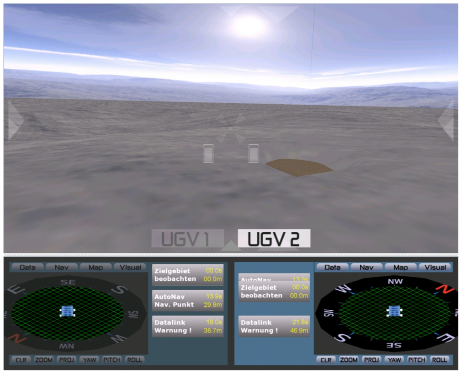

Fig. 2. Screen capture from multirobot user interface as used in this study. The upper section displays the camera view of the supervised robot. The lower section contains two robot consoles; on the left for robot one and on the right for robot two.

2.3 Mono- and Multimodal Display Configurations

This study compares four distinct display configurations, two monomodal and two multimodal, for rendering the information about the remaining time until the obstacle is reached. The graphical design of the first monomodal (that is only visual) user interface was already described in the previous section and simply is a general purpose dynamic priority list. On the contrary, the second mondomodal user interface is a task specific head-up-display allowing the operator to identify its state without interrupting the use of the camera view. The multimodal display configurations are differentiated by their respective channel, auditory or tactile, used to offload information. Figure 3 shows both screen captures and photos of the used interfaces.

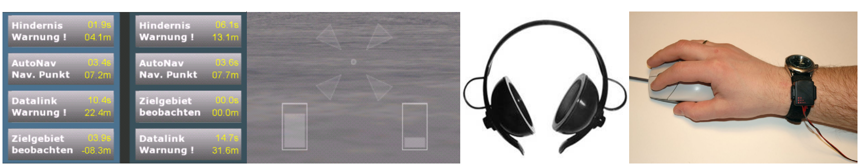

Fig. 3. From left to right: The dynamic priority list, the head-up-display, headphone based binaural auditory display and a wrist-watch attached tactile element

Common to all the interface elements is the rigorous left-right structure of information displays following the graphical user interface as shown in Fig. 2. The left priority list belongs to the left robot while the right priority list belongs to the right one. The same is valid for the head-up-display which is rendering the remaining distance to the obstacle by a proportional semi-transparent bar. The head-up-display is positioned in the central region of the camera view below the crosshair.

An important aspect of the auditory display is its binaural design. Humans are well capable of concurrently listening to multiple sound sources and quickly separating the relevant from the irrelevant ones. This effect is used in the context of the design of the human-multirobot interface by delivering the auditory information of each of the two robots exclusively to the left or right ear of the operator. This design allows the operator to concurrently listen to the auditory channels of both robots. Again the same is valid for the tactile interface which is mounted on the left and right wrist of the operator. The signal used for both auditory and tactile information transmission is an interval and pitch modulated beep tone, or vibration, which dynamically encodes the obstacle's temporal distance. The silence intervals between subsequent beep tones are synchronized to play at the same point in time if their interval length is equal.

In order to compare the monomodal and multimodal interfaces the two interface alternatives were parameterized to communicate the same information. Both the auditory, tactile and the visual elements are activated at a temporal obstacle distance of 15 seconds. The vertical movement of the visual element in the pending interaction list is encoded by three easily distinguishable silence interval lengths of the corresponding auditory obstacle signal. When the visual element moves up the pending request list, the silence interval is set to the corresponding obstacle temporal distance level. The temporal distance is the estimated time until the robot hits the obstacle given the current heading and velocity. The three levels were set at 15, 10 and 5 seconds. The corresponding silence interval lengths for the auditiory and tactile interface were 1000 ms, 500 ms and 125 ms (1Hz, 2Hz and 8Hz). The numeric display containing the temporal distance to the obstacle was encoded into the beep tone's, respectively the vibration's, pitch level.

2.4 Design of Experiment

Fourteen subjects with a mean age of 27 years participated in the laboratory study. A written instruction manual containing information about the trials was handed out to the participants several days before the trials. An introductory training session was conducted at least one day before to familiarize the participants with the user interface, the robot and payload control task, and the visual, tactile and auditory displays. Following this training the users conducted four thirteen minute trials using each of the user interfaces. Finally, a post trial subjective workload assessment using the NASA Task Load Index (TLX) was carried out.

The hypothesis for the robot control task is that the multimodal interface improves operator's performance due to the offloading of the task related information to the auditory channel. This should enable operators to improve their coordination between the two concurrent tasks as well as between two robots due to the binaural information presentation.

The hypothesis for the payload control task is that the performance will also increase when using the multimodal interface due to the reduction of visual load, thereby enabling the operator to use more time for the target search in the camera view. Finally, the hypothesis for the workload section is that workload will be lower when working with the multimodal interface due to its capabilities of supporting the coordination between concurrent tasks. We tested the associated null hypotheses using a general linear model (GLM) with repeated measures. The level of significance was $\alpha_{GLM} = 0.05$. The results are shown as error bars with two times the standard error. Variables are named LIS for the priority list, HUD for the head-up-display, AUD for the auditory and TAC for the tactile display.

3 Results and Discussion

Robot control task: Obstacle classification

The variable "obstacle classification" accounts for the correctly classified obstacles. Figure 4 shows the means and error bars of the normalized results for the interface alternatives investigated.

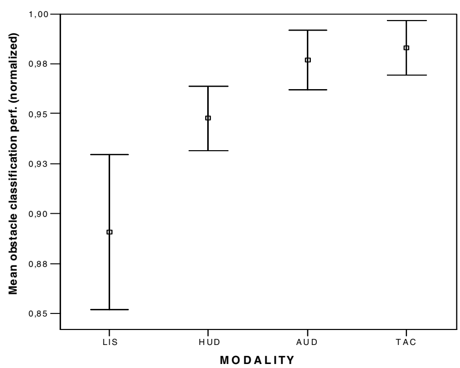

Fig. 4. Mean normalized obstacle classification performance. A total of 55 obstacles were passed by the two robots. The performance of the multimodal interfaces was almost perfect with between zero and two failures.

According to Figure 4 the priority list scores significantly lower than any other interface, especially compared to the head-up-display ($p < 0.01$) representing the best monomodal interface. Both multimodal interfaces outperform both monomodal ones (AUD vs. HUD $p < 0.01$ and TAC vs. HUD $p < 0.01$). The result for the priority list is not unexpected as this interface requires the operator to switch his visual attention frequently between the camera view and the console. The result of the head-up-display (HUD) however is surprising as it is scores significantly better than the priority list but still is also significantly lower than the multimodal interfaces. This indicates that even an optimized interface such as the head-up-display suffers from the conflict with the target search task. The multimodal interfaces show no significant differences.

Payload task: Target detection

The secondary task in this experiment was to detect and mark as many targets as possible while still classifying all incoming obstacles. Figure 5 shows the results for the target detection performance. The priority list again scores significantly lower than any other interface ($p < 0.05$ vs. TAC, $p < 0.01$ vs. AUD, $p < 0.001$ vs. HUD). Given the high potential for conflicts between the visual search task and the necessity to frequently check the list elements this is not unexpected. It is however unexpected that the head-up-display appears to provide the highest mean scores, or the lowest impairment, in respect to the target detection. This is surprising as the target detection task should benefit from the visual offloading provided by the multimodal interfaces. Based on the present data, the performance difference between the head-up-display and the tactile display is significant ($p < 0.05$).

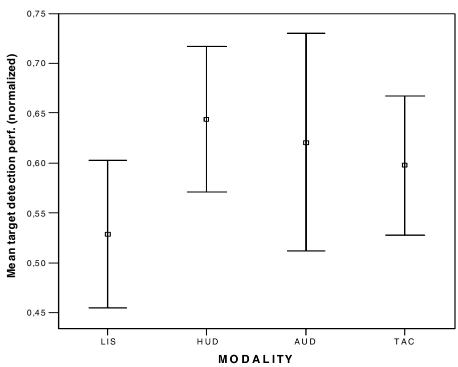

Fig. 5. Mean normalized target detection performance. A total of 59 targets were detectable by the operator.

Gaze Analysis

We used a video based remote eye tracking system (Tobii x50, measuring at 50Hz) to record the visual scanning paths of the participants. As the payload control task is based on a difficult continuous visual search task, we expected to find differences in the eye movements in respect to the modality of the displays. The general visual scanning is focused along the horizon level of the camera view (see Fig. 6 left). Our hypothesis is that the scanning patterns will change especially in the peripheral regions when conflicts with visual perception occur. This means, that the operator can expand the search area. To analyze this we used a set of regions of interest to filter the measured gaze positions along the horizon (see Fig. 6 right).

Figure 6 shows the results of the measured gaze positions in the L1 and R1 region. A clear trend can be identified between the monomodal and multimodal interfaces. The use of the priority list resulted in significantly lower visual activity than with any other interface. The head-up-display requires the operator to often focus the central crosshair region of the camera view and thus results in a narrower visual scan pattern. Both multimodal interfaces enable significantly higher visual scanning in the peripheral regions than the monomodal ones (AUD vs. HUD $p < 0.01$, TAC vs. HUD $p < 0.05$). This is especially valuable as this can reduce the problem of tunnel vision under high workload. Unexpectedly, the head-up-display shows no significant difference compared to the priority list (HUD vs. LIS $p < 0.1$).

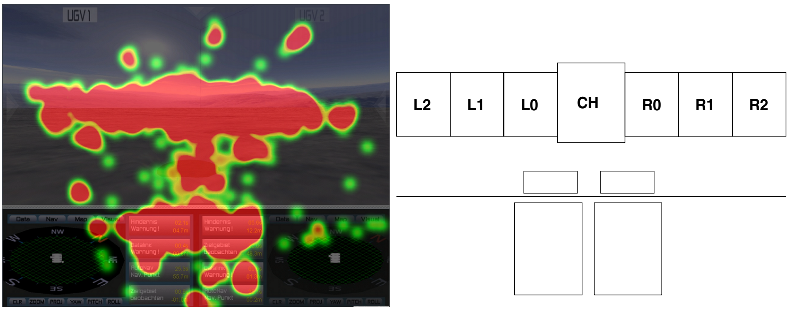

Fig. 6. Gaze analysis design. The left image shows an exemplary hotspot visualization of the gaze activity (fixations). The right figure shows the corresponding regions of interest used as filters.

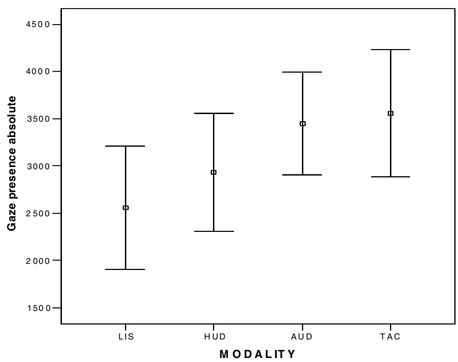

Fig. 7. Number of measured gaze positions in L1 and R1 region (see Figure 6 left)

Workload

Finally, the subjective workload was sampled using the NASA-TLX workload rating technique. Figure 8 depicts the results. Similar to the variables investigated before the priority list based interface does not only provide the lowest performance but requires also significantly higher workload than the other three interfaces (LIS vs. HUD $p < 0.01$). The multimodal interfaces are rated almost identical but do not achieve a significantly lower workload than the head-up-display (HUD vs. TAC $p < 0.25$, HUD vs. AUD $p < 0.25$, HUD vs. AUD $p < 0.95$).

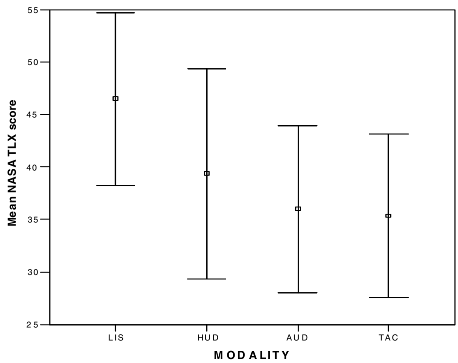

Fig. 8. Results of the NASA-TLX workload sampling. The scale ranges from 0 to 100 where a lower score means lower workload.

4 Conclusion and Future Work

In this paper a study was presented to compare two monomodal and two multimodal alternatives of our experimental multirobot user interface. The results clearly show that the human operator can benefit from our multimodal interface design because the performance of the robot control task is significantly higher when using the multimodal interface (see figure 4). Surprisingly, this performance increase does not have a positive or negative impact on the performance of the payload control task. This is a clear indication that the multimodal user interface adequately supported human multi-tasking. It is even more important that the subjective workload of the operators being measured with the NASA TLX method is significantly reduced in comparison to the priority list interface while the performance increased (see figure 8, 4). It is interesting to see that the results for both the auditory and the tactile multimodal interface perform equally well and therefore can be substituted for each other without reducing performance.

In summary, the initial claim that a multimodal interface is fruitful when striving for significant performance improvements of human supervisory control of multirobot systems is supported by the experimental data of the study. However, the good performance of the monomodal head-up-display demonstrates that multimodal interfaces are not per se superior to monomodal ones.

In future experiments, we will investigate more closely the effect of the modality on the performance of the target search task by improving our gaze analysis. We will further combine the auditory and tactile interface to design multirobot interfaces for systems with up to four robots.

References

1. Adamczyk, P.D., Bailey, B.P.: If Not Now, When? The Effects of Interruption at Different Moments Within Task Execution. In: Proc. of the Conference on Human Factors in Computing Systems, pp. 271–278. Vienna, Austria (2004)
2. Adams, J.A.: Critical Considerations for Human-Robot Interface Development, AAAI Fall Symposium on Human-Robot Interaction (November 2002)

3. Crandall, J.W., Goodrich, M.A., Olsen Jr., D.R., Nielsen, C.W.: Validating Human-Robot Interaction Schemes in Multitasking Environments. IEEE Trans. on Systems, Man, and Cybernetics-Part A: Systems and Humans, 35(4) (July 2005)
4. Dixon, S.R., Wickens, C.D.: Control of Multiple-UAVs: A Workload Analysis, In: 12th International Symposium on Aviation Psychology, Dayton OH (2003)
5. Goodrich, M.A, Olsen Jr, D.R., Crandall, J.W., Palmer, T.J.: Experiments in Adjustable Autonomy, In: Proc. of IJCAI01 Workshop on Autonomy, Delegation, and Control: Interacting with Autonomous Agents
6. Goodrich, M.R., Olsen Jr., D.R.: Seven Principles of Efficient Human Robot Interaction, In: Proc. of the IEEE International Conference on Systems, Man, and Cybernetics, pp. 3943–3948. Washington (October 5-8, 2003)
7. Helleberg, J., Wickens, C.D., Goh, J.: Traffic and Data Link Displays: Auditory? Visual? Or Redundant? A Visual Scanning Analysis, In: 12th International Symposium on Aviation Psychology, Dayton OH (2003)
8. Miller, S.R.: Window of Opportunity: Using the Interruption Lag to Manage Disruption in Complex Tasks, In: Proc. of the Human Factors and Ergnonomics Society 46th Annual Meeting (October 2002)
9. Nielsen, C.W., Goodrich, M.A., Crandall, J.W.: Experiments in Human-Robot Teams, Multi-Robot Systems: From Swarms to Intelligent Automata Volume II (2003)
10. Olsen, D.R., Goodrich, M.A.: Metrics for Evaluating Human-Robot Interactions, In: Proceedings of the Performance Metrics for Intelligent Systems Workshop, Gaithersburg, MD (September 16-18, 2003)
11. Sheridan, T.B.: Telerobotics, Automation, and Human Supervisory Control. The MIT Press, Cambridge (MA) (1992)
12. Trouvain, B., Schlick, C.: A Study of Audio and Visual Context Switch Indicators in a Multirobot Navigation Task. In: Proceedings of the Conference of Systems, Man, and Cybernetics, The Hague, Netherlands (To appear, 2004)
13. Wickens, C.D.: Processing resources and attention, D. Damos (ed.), Multiple Task Performance, pp. 3–34 (1991)
14. Wickens, C.D., Dixon, S., Chang, D.: Using Interference Models to Predict Performance in a Multiple-Task UAV Environment – 2 UAVs, Technical Report AHFD-03-9/MAAD-03-1, AHFD (April 2003)
15. Wickens, C.D.: The Structure of attentional resources. In: Nickerson, R.S. (ed.), Attention and performance VIII, pp. 239–257 Hillsdale, NJ