

Chapter 9 – NL-1: REMOTE AUDITORY TARGET DETECTION USING AN UNMANNED VEHICLE – COMPARISON BETWEEN A TELEPRESENCE HEADTRACKING 3D AUDIO SETUP AND A JOYSTICK-CONTROLLED SYSTEM WITH A DIRECTIONAL MICROPHONE

Chris Jansen and Dr. Leo van Breda

TNO
Kampweg 5
3769DE Soesterberg
NETHERLANDS

Email: chris.jansen@tno.nl

Dr. Linda Elliott

Army Research Lab
Human Research and Engineering Directorate
US Army Maneuver Center of Excellence
6450 Way Street, Bldg. 2839, Room 333
Fort Benning, GA 31905
USA

Email: linda.r.elliott.civ@mail.mil

9.1 DATES

June 2010.

9.2 LOCATION

The experiments/demonstrations were held at Fort Benning GA, USA.

9.3 SCENARIO/TASKS

For soldiers, visual information is crucial in building up situation awareness. Not surprisingly, when robots are used for reconnaissance of a remote area, visual sensors are most important. Most robots are therefore equipped with visual sensors, but not with any other kind of sensor! This is surprising, because an approaching car (yet invisible because still around the corner), someone moving in an adjacent room, a slamming door, or the loading of a gun, are important events during reconnaissance which cues are primarily auditory. When such sounds occur, human beings almost instinctively direct their heads (eyes) to the sound source for visual inspection before deciding to hide, or to make contact, to get out very fast, to attack, etc. Even more so, human beings immediately know where to hide or where the safe exit is because of their excellent spatial situation awareness that results from the human-intrinsic integrated perception of visual, auditory, and proprioceptive information. We hypothesize that if such intrinsic integrated multi-modal perception would be facilitated in remote perception using robots (by having headtracking control for robot's sensor system that includes stereo vision and spatial 3D audio, a setup we refer to as Telepresence [1]), spatial situation awareness would boost performance in a robot reconnaissance mission. This hypothesis was investigated in the experiment reported here, which was conducted as part of a research collaboration between the US Army Research Lab, ft Benning, and TNO.

9.4 TECHNOLOGIES EXPLORED

9.4.1 Reconnaissance Environment

The reconnaissance environment consisted of a large room (about 60 m²) subdivided in several sections, and a smaller adjacent room (about 8 m²). Eleven possible target objects varying in size were positioned at different height levels in the reconnaissance environment:

- A) Soda can bomb on a table;
- B) Hand grenade on the ground;
- C) Soda can bomb on the ground;
- D) Hand grenade near the ceiling;
- E) Semtex on the ground;
- F) Bomb shell on a table;
- G) Pipe bomb on the ground;
- H) Semtex with timer on a chair;
- I) Mine on a water container;
- J) Land mine on a high cupboard shelf; and
- K) Land mine on a high cupboard shelf.

Objects B, D, E, G, H, J were used as targets; the others were used as decoy targets or practice targets.

9.4.2 Control Station

The control station was located in a tent next to the building of the reconnaissance environment (see Figure 9-1). The control station consisted of a user interface with a NVIS nVISOR Head-Mounted Display (either stereo or mono, depending on the experimental condition), an Xsens MTi motion sensor as a headtracker, stereo headphones, and a Logitech Dual Action game controller. Three human-robot interface setups of the control station were used in this experiment, as explained in the section below on Experimental Setup.

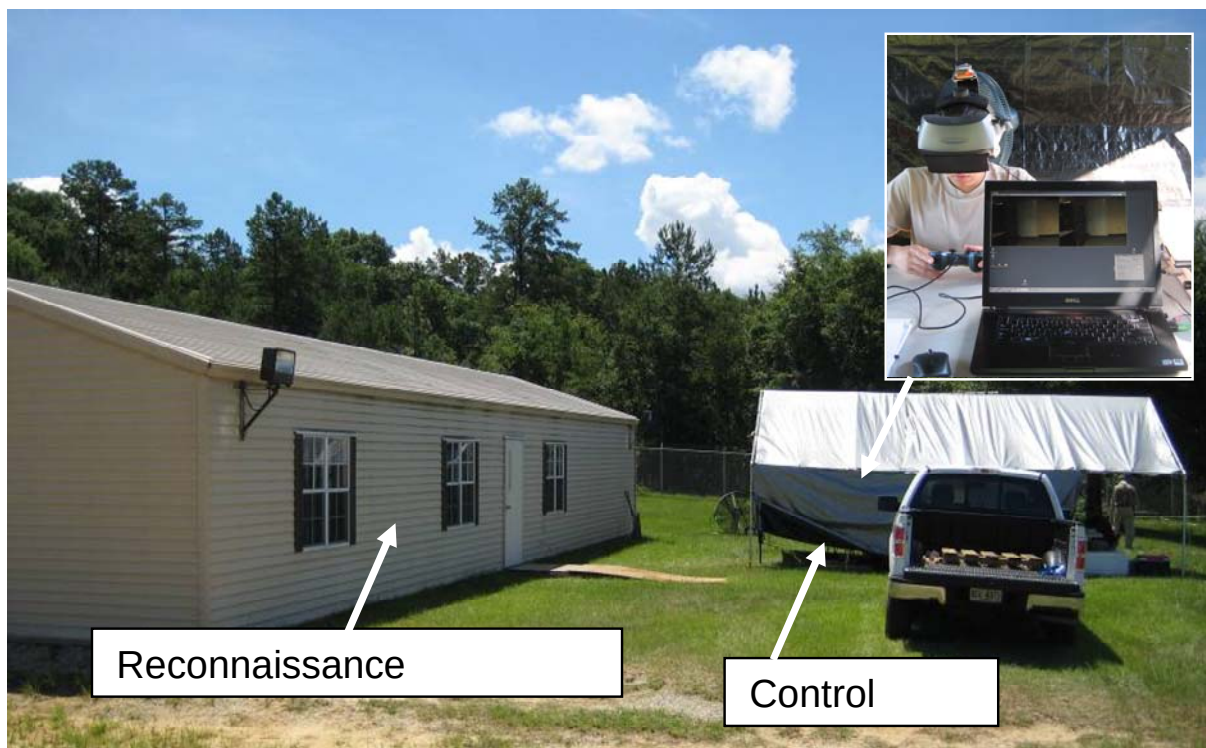


Figure 9-1: Reconnaissance Area and Control Station.

9.4.3 Unmanned System

The Unmanned Ground Vehicle (UGV) used was TNO's robot called 'Generaal'. This UGV is a fully manually controlled UGV, with a fast and powerful pan-tilt-roll sensor system that can accurately mimic human head movements enabling remote perception of the UGV environment.

9.5 HUMAN FACTORS ISSUES EXPLORED

Our main human factors research questions for the current experiment were:

- Does headtracking control lead to improved performance as compared to joystick control? (comparison between Mono Headtracking and Mono Joystick human-robot interfaces, see description below);
- Does a 3D audio system lead to improved performance as compared with a directional microphone? (comparison between Mono-Headtracking and Telepresence human-robot interfaces); and
- What would be the maximum performance benefit of telepresence functionality (with headtracking and stereo sensor information) as compared with the currently mostly used control systems with joystick control and mono sensor information? provided it exists? (comparison between Telepresence and Mono-Joystick human-robot interfaces).

For answering these questions we considered the quality of performance in locating and identifying objects in an indoor audio detection task, in three experimental conditions for the user interfaces:

- Mono-Joystick: Mono audio and video on Head Mounted Display, with joystick control for robot movements and heading of sensor system. Participants were asked (and reminded when needed) not to move their heads.
- Mono-Headtracking: Mono audio and video on Head Mounted Display, with joystick control for robot movements and headtracking for directing the sensor system.
- Telepresence: Stereo audio and video on Head Mounted Display, with joystick control for robot movements and headtracking for directing the sensor system. We refer to this configuration as Telepresence.

Each participant performed the sound detection task 18 times. Each of the six targets was used for each of three conditions. After each trial, the participant switched to one of the other two experimental conditions.

9.6 UNMANNED SYSTEMS USED

The Unmanned Ground Vehicle (UGV) used was TNO's robot called 'Generaal'. This UGV has been used in prior studies in our lab [2]. It is a fully manually controlled UGV, with a fast and powerful pan-tilt-roll system that can accurately mimic human head movements. On top are two cameras for providing stereo vision at the control station, and two microphone arrays that can be positioned at either side for spatial 3D audio, or next to each other in front thereby functioning as a directional microphone. The horizontally positioned red-tipped pointer in front of the vehicle was the reference point for the participants in approaching the target as closely as possible.

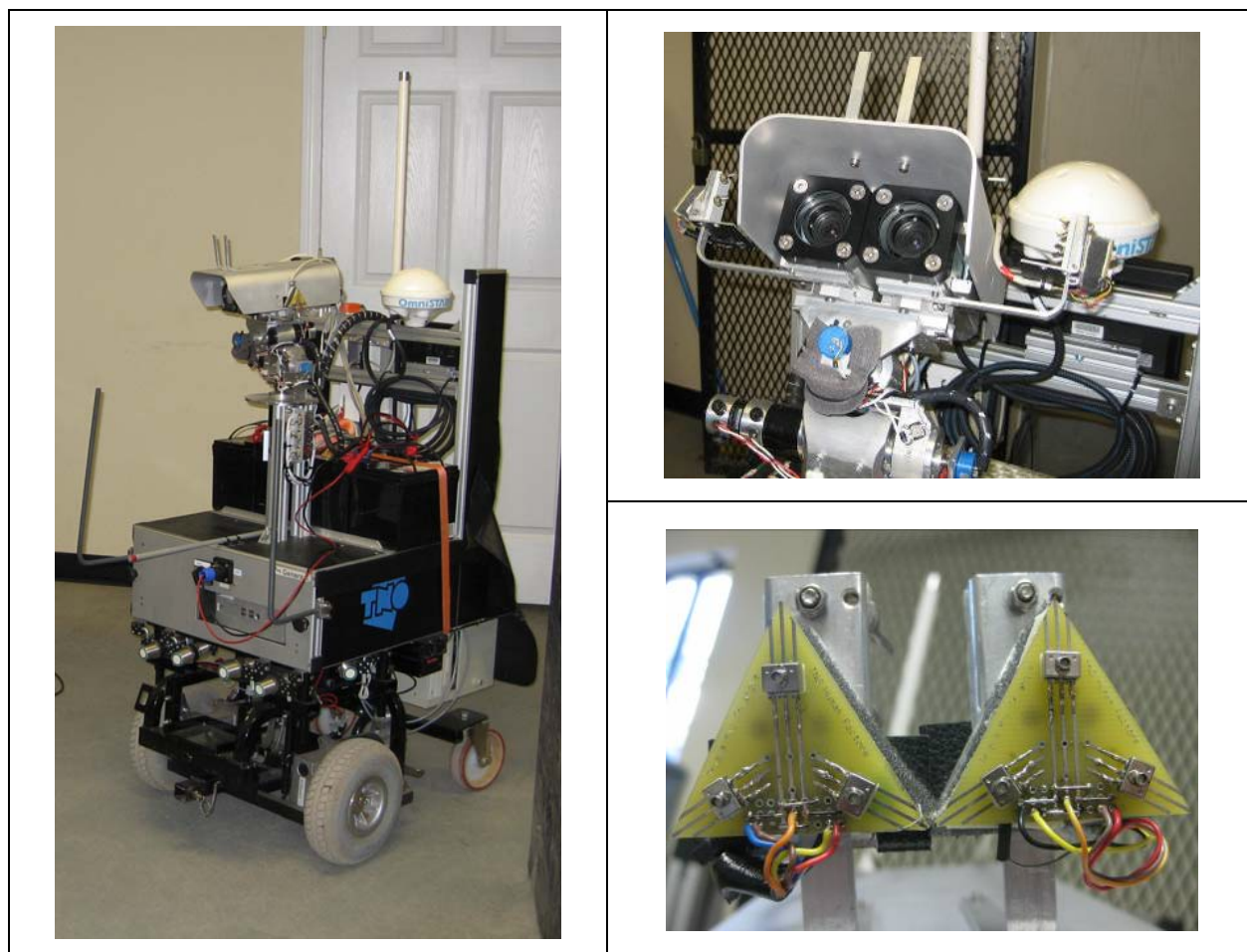


Figure 9-2: TNO's Unmanned Vehicle 'Generaal'. Left panel shows the vehicle with sensor unit on a pan-tilt-roll motion platform with 3D audio and stereo visual sensors. The sensor unit is presented enlarged in the upper right panel, with the microphone array placed in their 3D audio position, at either side of the stereo cameras. The lower right panel shows how the two microphone arrays were placed in the centre position right above the stereo cameras, for receiving directional mono sound.

9.7 SUMMARY OF ANY NATO COMMUNICATIONS/COLLABORATIONS/ INTERACTIONS

	<i>Planning/Design</i>	<i>Execution</i>	<i>Analysis</i>
Communication	<i>TNO and US Army</i>	<i>TNO and US Army</i>	<i>TNO and US Army</i>
Coordination	<i>TNO and US Army</i>	<i>TNO and US Army</i>	<i>TNO and US Army</i>
Collaboration	<i>TNO and US Army</i>	<i>TNO and US Army</i>	<i>TNO and US Army</i>

9.8 SUMMARY OF TECHNICAL DEMONSTRATION RESULTS

Non-parametric Wilcoxon Matched Pairs tests show that the percentage correct target ID in the Telepresence condition is significantly higher (87.5%) than in both Mono-Joystick (61.5%; $p < .05$) and Mono-Headtracking (64.6%; $p < .005$); Mono-Headtracking and MJ do not differ ($p = .51$).

The repeated measures ANOVA on time to target identification indicates a main effect for Human-Robot Interface ($F(2,30) = 17.48$, $p < .001$); all Tukey HSD post hoc tests were significant (all $p < .05$). Time to target identification is shortest for Telepresence (65.0 seconds), followed by Mono-Headtracking (88.8 seconds), and longest for Mono-Joystick (113.5 seconds).

9.9 LESSONS LEARNED

Including head motion tracking for controlling a directional microphone significantly improves a human operator's detection and localization of audio targets in a reconnaissance mission. This performance is boosted even more when human natural listening behavior is further mimicked when 3D audio is presented using advanced microphone arrays instead of mono audio generated by a directional microphone.

We have learned that field tests are valuable if not crucial in estimating the possible operational benefits of technology that already has been tested and improved in the laboratory conditions.

9.10 STUDY CONSTRAINTS/LIMITATIONS

The findings of this study are limited to indoor reconnaissance in which no other sounds are present except for the audio target.

9.11 CONCLUSIONS

In Section 9.5 we identified three research questions that can be answered:

- Does headtracking control lead to improved performance as compared to joystick control? The results show no difference between the Mono-Headtracking condition and the Mono-Joystick condition in correctness of target identification. However, with joystick control, more time is needed for target identification: about 26% more time is needed when using a joystick for sensor control (here Mono-Joystick with 111.2 seconds on average) as compared to headtracking (here Mono-Headtracking with 88.2 seconds).
- Does a 3D audio system lead to improved performance as compared with a directional microphone? When comparing the Telepresence condition (having 3D audio) with the Mono-Headtracking condition (having a directional microphone), we see that with Telepresence the percentage of correctly identified targets is about 23% higher. In addition, target identification takes about 35% more time without having the 3D audio functionality available (here 88.2 and 65.0 seconds for Mono-Headtracking and Telepresence respectively).
- What would be the maximum performance benefit of telepresence functionality as compared with the currently mostly used control systems with joystick control and mono sensor information, provided it exists? Based on the results in this study, the use of a Telepresence human-robot interface results in identification/localization times for audio that are about 42% shorter than with current commonly

used interfaces (65.0 sec and 111.2 sec for Telepresence and Mono-Joystick respectively). In addition, the target identification performance increases by about 26% when using the Telepresence human-robot interface.

These promising results encourage more elaborate testing in operational settings, following our initial field trials with telepresence UGV control reported in [3].

9.12 FUTURE RESEARCH NEEDS AND PLANS IN THIS AREA

In continuing the collaboration between TNO and ARL, we are considering two options that could be performed in parallel. First, we believe that telepresence could be even more beneficial if other multi-modal user interface are included as well, in particular vibrotactile interfaces (e.g., for indicating direction of movement, the next waypoint, collision warnings for obstacles). Second, we plan to investigate the extent to which performance in a reconnaissance mission could further increase by combining telepresence operator involvement with robot autonomy in a well-designed adaptive automation concept.

9.13 ACKNOWLEDGEMENTS

The authors would like to acknowledge NL army programme V923 and the US ARL for sponsoring organization for this study.

9.14 REFERENCES

- [1] Jansen, C. and Van Erp, J.B.P. (2010). Telepresence control of unmanned vehicles. In Human-Robot Interactions in Future Military Operations, Barnes and Jentsch (Eds). Adhgate, UK.
- [2] Van Erp, J.B.P., Duistermaat, M.D., Jansen, C., Groen, E. and Hoedemaeker, M. (2006). Telepresence: Bringing the Operator back in the loop. RTO-MP-HFM-136.
- [3] Den Breejen, E. and Jansen, C. (2008). Eyerobot TBI unmanned Telepresence reconnaissance mission. RTO SET-130 / RSY-023.

