2005

Synergy of Virtual Visual and Auditory Displays for UAV Ground Control Stations

Susan R. Dowell

R. Jar Shively

SYNERGY OF VIRTUAL VISUAL AND AUDITORY DISPLAYS FOR UAV GROUND CONTROL STATIONS

Susan R. Dowell and **R. Jay Shively**
Army/NASA Rotorcraft Division
Aeroflightdynamics Directorate (AMRDEC)
US Army Research, Development & Engineering Command
Moffett Field, CA

Unmanned aerial vehicle (UAV) operators must remotely manipulate payload sensors, while maintaining situational awareness from a displaced ground control station (GCS). Potential use of helmet-mounted displays (HMD) in piloting UAVs and controlling payload sensors has been previously investigated (de Vries & Padmos, 1997; Draper, Ruff, & LaFleur, 2001; Morphew, Shively, & Casey, 2004). Stated benefits of HMD use for targeting tasks included immersion in the search environment and possible reduction of tactical footprint. In the current study, it was hypothesized that the pairing of 3-D audio alerts with the HMD would result in more robust performance differences between HMD and CRT conditions. For this experiment, eight subjects conducted routine area searches, periodically responding to audio threat alerts. Audio alerts were given in mono, stereo, and 3-D spatialized presentation. Targeting performance differences were assessed in a baseline CRT and joystick configuration versus HMD for all audio conditions. Findings revealed more precise target acquisition performance when payload operators used the CRT/joystick configuration than the HMD. Furthermore, time on target was reduced when visual searches were aided with stereo and 3-D directional audio cues. Lastly, participants missed the fewest targets and reported lowest workload levels, when receiving 3-D audio cues. Present findings replicated reported sickness associated with HMD use. A synergistic effect of 3-D audio and HMD showed a mitigation of operator workload previously reported with the HMD. Further consideration of 3-D audio alerting for UAV operators should be investigated for benefits in target acquisition, reduced operator workload, and increased situation awareness.

Introduction

Unmanned aerial vehicle (UAV) ground control stations present the unique environment of displacing the operator from the vehicle flying. This displacement removes typical cues used by pilots (e.g., proprioceptive, visual, vestibular) to aviate effectively and maintain situational awareness. Current unmanned aerial vehicle ground control stations are characterized by traditional workstation layouts: Two multi-function displays per station, a keyboard, and joystick (e.g., Shadow, Predator). Synergy of 3-D audio alerting with an HMD may lead to higher target acquisition performance, increased situational awareness, and lower operator workload than the current interface.

Potentially, helmet-mounted displays (HMDs) offer a reduced system footprint and the benefit of an immersive search environment for the mission payload operator (MPO). Thus far, empirical data has revealed only limited success with HMDs in the control of UAV payload sensors (Draper, Ruff, Fontejon, & Napier, 2002; Morphew, Shively, & Casey, 2004). Noted caveats for HMD use have been associated with visual lag (Rash & McLean, 1999), head-coupled sensor manipulation (de Vries & Padmos, 1998), and potential sickness side effects (DiZio & Lackner, 1997). Consequently, improvements are necessary to obviate the reported costs associated with HMD use and possibly contribute with a reduced tactical footprint.

By way of improvement, guided visual searches eliciting slower head movement may mitigate previously reported operator discomfort. As such, the presentation of aural target information that is spatially localized, or 3-D audio cues, may facilitate more efficient visual searches for air and ground targets. This has been shown in cockpit applications to enhance the acquisition of air traffic, targets and incoming threats (Begault & Pittman, 1994). Benefits of 3-D audio in presenting target location information and threat avoidance have also been reported in simulated military applications where ambient noise in the cockpit competes with audio signals (Ericson, 2004). In applying these findings to UAV ground control stations, 3-D audio technology may similarly enhance operator performance.

The current experiment combined 3-D audio alerting with an HMD to assess the potential benefits to UAV operators on nominal search missions. For the purpose of comparison, mono and stereo cueing were also employed to assess the impact of spatialized target location information. Current ground control station configuration featuring CRT and joystick was used as a baseline for display presentation. Audio alerts were presented in both HMD and CRT display environments. Findings were expected to reveal a significant interaction between display type and audio alert condition, such that using an HMD with 3-D audio alerting yielded best operator performance. Additionally, operator workload was anticipated to decrease relative to expedited searches, directed by 3-D audio cueing.

Method

Participants

Eight right-handed, male participants between 18- 30 years old (*M* = 24 yrs.) with normal or corrected-to-normal vision and full ability to perceive color were tested in this experiment. All participants reported no hearing impairment. Monetary compensation was given for participation in the study.

Simulation Equipment

CRT and Flybox. Participants were tested in a UAV simulator based on the US Army's Tactical UAV (Shadow) Ground Control Station. The simulated sensor payload view was displayed on either a CRT or HMD, depending on experimental display condition. When the sensor view was displayed on the CRT, a 21" Silicon Graphics color monitor was located 65 cm from the participant's vantage point. Display resolution was 1024 x 768 pixels. UAV sensor heading and pitch were driven by the participant's manipulation of a spring-centered joystick on a BG Systems Flybox. In an attempt to simulate the U.S. Army's TUAV sensor payload, joystick manipulation enabled 360 deg pan capability with +45 to -115 deg pitch limitations (U.S. Department of the Army, 2001). Sensor slew rate operated at a constant 60 deg/sec.

HMD and headtracker. Alternately, the sensor view was displayed on a Kaiser ProView™ XL50 head mounted display, featuring a 30 deg vertical x 40 deg horizontal FOV with 100% binocular overlap (Figure 1). Display resolution was 1024 x 768 pixels. A Polhemus Fastrak electromagnetic head tracker transmitter was mounted on the HMD and used to track subjects' head movement. In this manner, subjects' head movement was coupled to sensor movement (i.e., turning the head left moves simulated sensor view left). Head movement manipulated sensor movement in x-y and pitch axes. As in a previous study (Morphew, Shively, & Casey, 2004), the sensor view contained an artificial 45 deg downlook bias. The built-in bias afforded an optimal 45 deg sensor downlook angle when the subject's chin was parallel to the ground and his eyes on the horizon, without necessitating a fatiguing sustained downward head tilt. Sensor slew rate matched physical limitations of the TUAV sensor (max. 60 deg/sec.). Consequently, head swivel movement actuated a sensor slew movement of no greater than 60 deg/sec. When head swivel movement exceeded 60 deg/sec., a programmed limiter was engaged, allowing for no greater than a 60deg/sec. pan capability. Graphics presentation and data collection were updated at 30 Hz for all display conditions.

Figure 1. *Kaiser ProView™ XL50 HMD with headtracker and flybox.*

Audio equipment. For all experimental trials, AuSim software generated audio alerts delivered through a Sennheiser HD570 headset. When delivering 3-dimensional audio alerts, spatialized sound was referenced to the participant's head position, which was calculated by the headtracker. In this manner, alerts were generated that sounded as if they originated from a point in space. Sampling of head position associated with localized alerts was updated at 30 Hz.

Simulation

Environment. The experimental scenario simulated an area reconnaissance conducted by a mission payload operator. The virtual scene displayed on either the CRT or the HMD was analogous to the sensor video feed from a notional tactical UAV. Medium-resolution, charcoal gray roads overlaid mottled brown terrain with some instances of green shrubbery and trees alongside the roads. Portions of the flight route were located in more populated areas of the database, which included buildings and other cultural features. Desert-camouflaged tanks (targets) and green-camouflaged tanks (non-targets) were positioned throughout the simulated environment (Figure 2). Placement of vehicles throughout the simulated terrain varied according to which of the eight nominally similar flight routes were flown.

UAV control and flight path were pre-programmed and operated in playback mode throughout the simulated missions. All mission scenarios were flown at 70 KIAS and an altitude of 5000 ft. AGL.

Figure 2. *Target (left) and non-target (right).*

Audio cues. Assuming complete accuracy of an automatic target recognition system, audio cues alerted subjects to the presence of a target. One of three types of audio cues was presented, depending on experimental condition. Audio cue type was characterized as Mono (non-directional), Stereo (left-right localization), or 3-D localized format. In the Mono audio cue condition, alerts were given in both ears of the headset. The alert consisted of a female voice repeating, "Target. Target." for a duration of 10 seconds, or until the target was identified. The alert ceased once the target was identified or if undetected, at the end of the 10 sec. window.

In the Stereo audio cue condition, alerting was presented in the left or right ear according to the location of the target relative to the current sensor heading. When utilizing the HMD for sensor control, stereo audio alerting was also relative to the head direction, as sensor position was coupled to head position. The content and duration of the alert was identical to the audio cue used for the Mono audio cue condition.

For the 3-D localized cue condition, alerts were given in spatialized presentation to the left or right ear and continuously updated with sensor/head position. Due to the nature of spatialized sound, audio cues appeared to originate in 3-D space, co-located with the target position. Accordingly, alerts could shift from left to right ear as updated to sensor position and referent to relative target position. Content and duration of the 3-D audio cues were identical to those detailed for all other cue conditions.

Search Task

A routine area search was conducted in each mission scenario. The mission instructions dictated that all vehicles found were classified as targets or non-targets. A button on the flybox was used to mark non-targets, while a trigger on the joystick marked targets. Marking of non-targets during periods without audio alerts served as a secondary task to prevent boredom and preserve vigilance by maintaining a level of work. Participants were instructed to immediately respond to any audio alert by moving the sensor in the direction of the audio cue, until the target was in sight. In experimental trials with mono audio alerts, the subject did not have directional information and therefore had an unguided search for the target. In all other audio cue conditions, the target search was guided. Once the target was detected, participants centered the target within superimposed crosshair symbology and depressed a trigger on the joystick. Subjects were instructed that targeting accuracy and speed were equally important. Targets were only visible during the time of the alerting. Otherwise, targets disappeared upon trigger depression or at the conclusion of the audio alert. After acquiring the related target, subjects returned to the secondary task of marking non-targets. A total of 12 targets were presented in every mission. Mission duration was approximately 12 minutes.

Experimental Design

A within-subjects design with repeated measures was conducted. The independent variables investigated were Display type (CRT or HMD) and Audio type (Mono, Stereo, or 3-D). Subjects participated in two sessions each for a total of 6 hours per subject. Separate sessions were necessary to isolate effects of display condition (HMD, CRT). The sequence of display testing was counterbalanced. Audio type alerts were blocked and randomized within the display condition. Two replications of each Display (2) x Audio type (3) mission were completed, for a total of 12 missions or 6 per session (Figure 3).

Subject ID	Day	Display	Audio			# of Trials
1	1	CRT	3-D x 2	Mono x 2	Stereo x 2	6
	2	HMD	Mono x 2	3-D x 2	Stereo x 2	6
					Grand Total =	12

Figure 3. *Experimental design with Subject 1 as exemplar.*

Data Collection

Targeting acquisition: Speed and accuracy. Objective performance measures included speed and accuracy of target acquisition. Speed of acquisition was calculated from the onset of the audio alert to the subject's trigger depression. Speed was measured to the nearest hundredth of a second. Accuracy of target acquisition was measured in pixels from the center of the superimposed crosshair symbology to the centroid of the tank. Targeting error was calculated in real-time data collection. All missed targets were recorded.

Workload ratings. The NASA-TLX subjective ratings scale (see Hart and Staveland, 1988), measuring perceived workload, was administered to subjects upon completion of each Audio type condition within an experimental session (Day 1 and Day 2). Subjects rated their workload in each Display x Audio experimental condition. A total of 6 sets of ratings were collected per subject.

Simulator sickness ratings. Participants' self reports of simulator sickness symptoms were collected at the end of each experimental session (Day 1 and Day 2) using the Kennedy Simulator Sickness Questionnaire (SSQ) (Kennedy & Lane, 1993). Baseline, pre-session symptom questionnaires were administered at the beginning of each experimental session for purpose of comparison.

Results

Objective Performance Measures

Separate 2 (Display type) x 3 (Audio type) x 3 (Block) x 2 (Trial) within-subjects repeated measures analyses of variance (ANOVAs) were conducted on speed and accuracy of target acquisition. Planned comparisons were examined on experimental variables of interest (e.g., Display, Audio) related to speed and accuracy performance independently.

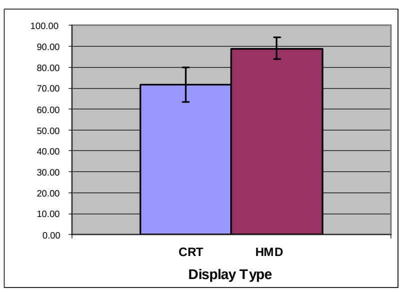

Figure 4. *Significant main effect of Display type.*

Targeting accuracy. A significant main effect of Display type was found in targeting accuracy, $F(1, 7) = 8.20$, $p < .05$ (Figure 4). Participants showed significantly more precise targeting when using a CRT (M = 71.65 pixels) than an HMD (M = 89.01 pixels). No significance variance in performance was found as an effect of Audio type. No significant interaction of experimental variables was found. In sum, targeting was more precise when sensor feed was presented on a CRT than an HMD.

Targeting speed. A significant main effect of Audio type was found in speed of targeting acquisition, $F(2, 14) = 144.36$, $p < .001$ (Figure 5). Stereo and 3-D audio alerting (M = 6.74, M = 7.06; sec. respectively) supported more rapid target acquisition than Mono audio alerting (M = 7.92 sec.) No statistical difference in performance was found between Stereo and 3-D Audio conditions. No effect of Display type and no significant interactions were found. Overall, time on target was reduced with stereo and 3-D cues, when compared with performance with mono audio cues.

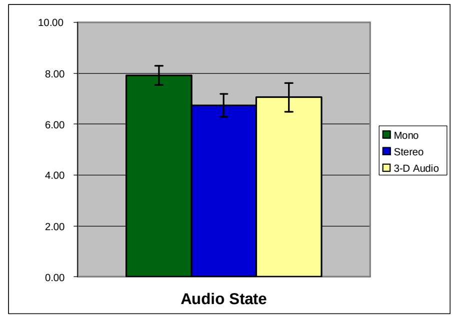

Figure 5. Main effect of Audio type (n = 8).

Missed targets. Data collected on the frequency of missed targets per mission showed a significant main effect of Audio type, $F(2, 14) = 5.84$, $p < .05$ (Figure 6). For missions where alerts were given in 3-D audio, participants were four times less likely to miss targets than when receiving mono audio alerts (.31:1.28 targets). Furthermore, 3-D audio alerts yielded an advantage of 2.7 times less missed targets than stereo alerting (M = .84). Participants showed significantly more missed targets in missions with mono audio alerts than all other audio conditions. No effect of Display type and no significant interactions were found. In summary, participants acquired the most targets when alerted in 3-D audio.

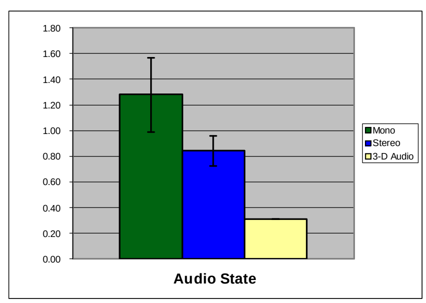

Figure 6. *Significant effect of Audio type (n = 8).*

Subjective Ratings

NASA-TLX workload ratings. In a comparison of means calculated from NASA-TLX ratings, collapsed across subscales, a significant interaction of Display and Audio type was found, $F(2, 14) = 4.26$, $p < .05$ (Figure 7). Missions flown using an HMD showed significantly lower operator workload when 3-D audio alerts were given. Whereas, both stereo and 3-D audio alerts positively impacted workload ratings when using a CRT. No significant difference in workload ratings was reported for mono and stereo alerting, when using an HMD. Data collapsed across Display type showed a significant main effect of Audio type (Figure 8), such that missions completed with 3-D audio alerts yielded significantly lower workload ratings than missions completed with mono or stereo alerting. No significant effect was revealed with the Display type manipulation. To summarize, missions flown with 3-D audio alerts yielded the lowest levels of reported operator workload. Furthermore, reportedly high workload ratings associated with the HMD were mitigated when 3-D audio cues were incorporated in the missions.

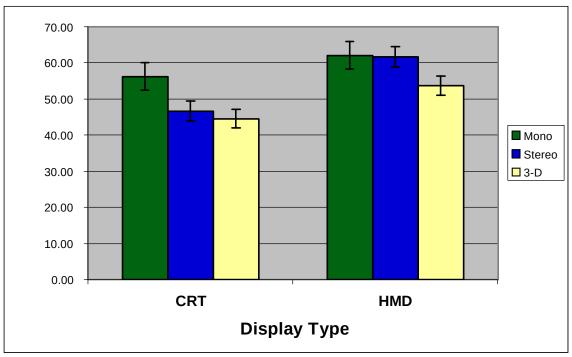

Figure 7. *NASA-TLX Workload Ratings show Display x Audio type interaction (n = 8).*

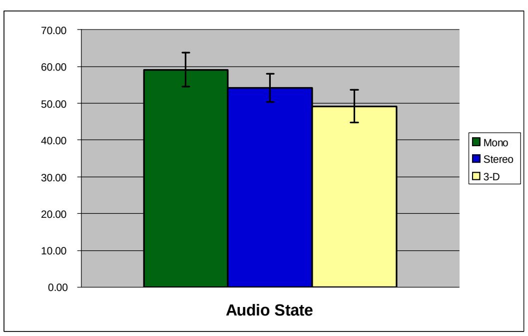

Figure 8. *NASA-TLX workload ratings show main effect of Audio State (n = 8).*

SSQ scores. Pre-session SSQ scores were calculated and analyzed for variance between Display types. As expected, no significant differences existed in reported sickness symptoms prior to exposure to the experimental session. Post-session SSQ scores, collapsed across sub-scales, revealed a significant interaction of Display type x Time, $F(1, 7) = 6.32$, $p < .05$ (Figure 9). Post experimental SSQ scores showed a significant increase in sickness symptoms when an HMD was used. Although, both sets of SSQ scores taken post session showed higher than baseline scores, use of the HMD showed sickness scores exceeding levels (SSQ > 20) warranted as tolerable by the developer of the questionnaire (Kennedy et al., 1992). In sum, participants reported more severe sickness symptoms with the HMD than the CRT, suggesting an unrecommended level of operator discomfort associated with HMD use.

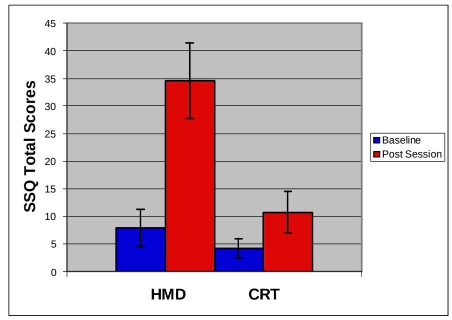

Figure 9. *Kennedy Simulator Sickness Questionnaire scores show Display type x Time interaction (n = 8).*

Discussion

In an evaluation of targeting performance data, accuracy was increased when the sensor feed was presented on a CRT versus an HMD. No associated performance tradeoff was recorded for time on target as an effect of Display type. It should be noted that when comparing CRT and HMD performance, not only the display, but the method of sensor control differed. Sensor control with the CRT was managed through fine motor input on a flybox joystick. By comparison, sensor control with the HMD was slaved to the swivel movement of a participant's head. Not unreasonably, precise targeting was better accomplished with fine motor movements of the practiced hand than more coarse movements of the head. In sum, the manipulation of display type revealed a performance decrement when the payload sensor was coupled to the head. Instead, results supported current joystick manipulation of the UAV sensor.

As anticipated, data collected on targeting performance revealed an effect of Audio alert type, supporting the use of stereo and 3-D audio alerts. These results upheld previous findings that directional audio cues reduce time to locate a target in a visual search task (Strybel & Guettler, 2001). For both stereo and 3-D audio alerts, participants were able to more rapidly acquire a target when given a directional audio cue, regardless of display type. Conversely, target search time was longer when the participants were given a non-directional (mono) cue. In an unexpected performance benefit, subjects given

3-D audio alerts missed four times less targets than when receiving mono alerts. Although it was hypothesized that 3-D audio alerting would enhance detection time; it was unforeseen that without 3-D localized alerting, subjects might miss up to 4 times more targets than non-directional alerting. Similarities in target search times for stereo and 3-D alerting conditions did not foretell the 2.7 times more missed targets for stereo versus 3-D audio. Therefore, it is important to consider both time on target and frequency of misses within the context of the operational scenario. In cases where rapid and successful acquisition of a high percentage of targets is necessary, performance data suggests the use of 3-D audio alerting.

The synergistic value of 3-D audio cueing paired with the HMD was revealed in reports of lower operator workload than experienced in all other audio conditions. Specifically, high workload ratings reported with HMD use were mitigated with 3-D audio cues. In both display types, 3-D audio cues supported lower levels of reported workload than alternate audio cues. As expected, operators experienced less workload when guided in a visual target search.

Reports of increased simulator sickness symptoms associated with HMD use and coupled sensor movement replicated findings of previous research (Morphew, Shively, & Casey, 2004). Hardware limitations of sensor slew rate and the associated visual lag likely contributed to self-reported nausea and eyestrain. As noted by the literature, even short periods of HMD use can result in side effects (e.g., headaches, nausea, blurred vision) that would be seen only after hours in front of a CRT (Stone, 1993).

Conclusion

Due to exceedingly high levels of reported sickness symptoms, head-slaved sensors with HMDs utilized in this study are not recommended. Once fatigue, stomach awareness, delayed sensor movement, and visual lag can be mitigated, HMDs may be a viable solution for UAV payload display and control.

Implications from this study suggest that 3-D audio alerting may offer enhanced capabilities to the payload operator for successful and rapid target acquisition. At present, 3-D alerting assumes target recognition technology that is not yet mature. Additional research will be required once developments of automated target recognition systems have reached operational proficiency.

Acknowledgements

Thanks to Daniel Casey of San Jose State University for his work in scenario development and Thomas Marlow of Silicon Valley Systems, Inc. for his technical assistance.

References

Begault, D. R., and Pittman, M.T. (1994). 3-D Audio Versus Head Down TCAS Displays. (Report No. 177636). NASA Ames Research Center.

de Vries, S.C., & Padmos, P. (1997). *Remotely controlled flying aided by a head-slaved camera and HMD* (Report TNO-TM 97-B024). Soesterberg, The Netherlands:TNO Human Factors Research Institute.

de Vries, S.C., & Padmos, P. (1998). Steering a simulated unmanned aerial vehicle using a head-slaved camera and HMD: effects of HMD quality, visible vehicle references and extended stereo cueing. *Proceedings of The International Society for Optical Engineering (SPIE) Conference*, (pp. 80-91).Orlando, FL.

DiZio, P., and Lackner, J.R. (1997). Motion sickness side effects and aftereffects of immersive virtual environments created with helmet-mounted displays. Presented at the RTO HFM Workshop, *The Capability of Virtual Reality to Meet Military Requirements,* Orlando:FL, December 1997. (RTO MP-54).

Draper, M. H., Ruff, H. A., Fontejon, J. V., and Napier, S. (2002). The effects of head coupled control and a head-mounted display (HMD) on large-area search tasks. Proceedings of the Human Factors and Ergonomics Society Conference.

Ericson, M. (2004). *3-D audio technology: 3-D audio displays provide military advantages.* Retrieved January 2005 from Airforce Research Lab Technology Horizons Web site: http://www.afrlhorizons.com/Briefs/Dec04/HE0402.html

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshakti (Eds.) Human Mental Workload, Amsterdam: North Holland Press.

Kennedy, R.S., & Lane, N.E (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology, 3(3)*, 203-220.

Kennedy, R.S., Lane, N.E., Lilienthal, M.G., & Hettinger, L.J. (1992). Profile analysis of simulator sickness symptoms: Application to virtual environment systems. *Presence, 1 (3)*, 295-301.

Morphew, M. E., Shively, R. J., & Casey, D. (2004). Helmet mounted displays for unmanned aerial vehicle control. Proceedings of The International Society for Optical Engineering (SPIE) Conference, Orlando: FL.

Rash, C.E., & McLean, W.E. (1999). *Human factors and performance concerns for the design of helmet-mounted displays*. U.S. Army Aeromedical Research Laboratory: Ft. Rucker, Alabama. (USAARL Report No. 99-08)

Stone, B. (1993). Concerns raised about eye strain in VR systems. *Real Time Graphics, 2(4)*, 1-13.

Strybel, T.Z., & Guettler, D.L. (2001). The effectiveness of visual vs. auditory auditory cues in a visual search performance: Implications for the design of virtual environments. In Smith, Slavendy, Harris & Koubek (Eds.), *Usability Evaluation and Interface Design: Cognitive Engineering, Intelligent Agents and Virtual Reality* (Vol 1) (pp. 1363-1367). Lawrence Erlbaum.

U.S. Department of the Army. (2001). *Technical Manual Operator's Manual for Shadow 200 TUAV System* (DTM 9-5895-681-10). Washington, D.C.