2013

Issues Relevant for Synthetic Teammate – Human Teammate Interactions in Operations of a Synthetic Unmanned Aerial System

Nancy J. Cooke

Christopher W. Myers

ISSUES RELEVANT FOR SYNTHETIC TEAMMATE – HUMAN TEAMMATE INTERACTIONS IN OPERATIONS OF A SYNTHETIC UNMANNED AERIAL SYSTEM

Nancy J. Cooke
Arizona State University and Cognitive Engineering Research Laboratory
Mesa, AZ
Christopher W. Myers
Air Force Research Laboratory
Dayton, OH

Two experiments are reported that set the stage for a project in which an ACT-R based Air Vehicle Operator will interact with two human teammates in an Unmanned Aerial System synthetic environment. Of interest are the ways in which the synthetic teammate fails to coordinate in an effective manner with humans. In Experiment 1, the new communication mode of text chat is compared to voice communications used previously in this task environment with all human participants. Issues of team performance and coordination were examined and differences noted particularly due to lag in the asynchronous chat mode. In Experiment 2 a condition in which two team members were told that the third was remotely located was compared to a condition in which two team members were told that the third human team member was a synthetic agent. Preliminary observations indicate that the "synthetic agent" is ignored and experiences terse communication compared to the remote teammate.

The research presented here is taking place as part of a larger project with the Air Force Research Laboratory that replaces a human UAS (Unmanned Aerial System) pilot with a cognitively plausible computational model that serves as a full-fledged synthetic teammate for a three-agent UAS ground control crew. Not only is the extension of the ACT-R cognitive modeling architecture of interest (Ball, Myers, Heiberg, et al., 2010), but the larger project will address questions about team coordination: What is the nature of coordination and collaboration (within all-human or mixed human-synthetic teams) in UAS ground control settings and what do deficiencies in synthetic teammate interactions with human teammates reveal about human-automation coordination needs?

Prior to inserting the synthetic teammate into the loop with two human participants, two experiments with all-human teams were conducted to establish baselines and are reported here. First, to establish a baseline for a new text chat mode of communication, team performance and coordination is examined using text chat communication and compared to voice communication. Second, performance and coordination in three-person teams was investigated when either two teammates were told that the pilot was remotely located or when they were told that the pilot was a synthetic teammate.

Team Coordination in the CERTT UAS-STE

This research program is conducted in the context of the CERTT UAS-STE (Cognitive Engineering Research on Team Tasks Unmanned Aerial System - Synthetic Task Environment (Cooke & Shope, 2005). The UAS-STE is based on the United States Air Force Predator UAS ground control station. The UAS-STE task requires a team of three people to complete the task of photographing critical waypoints. Each team member is assigned to one of three roles: an Air Vehicle Operator (AVO), a Payload Operator (PLO), or a Data Exploitation Mission Planning Coordinator (DEMPC). The DEMPC plans a mission route through multiple waypoints, the AVO is responsible for flying the simulated UAS and monitoring UAS systems, and the PLO takes photographs of designated waypoints and monitors camera systems. The roles are interdependent, where each role requires input from other team members to complete the team's goal of photographing designated waypoints. Further, the CERTT UAS-STE is dynamic and taking good photographs of designated waypoints requires information to be shared among teammates in a timely manner. A single UAS-STE mission consists of 11-12 targets and lasts a maximum of 40 minutes; each team performs five 40-minute missions.

Over a decade of research conducted in the CERTT UAS –STE has indicated that team interaction in the form of coordinated information passing and communication is important for predicting team performance and has led to a theory of Interactive Team Cognition (Cooke, Gorman, Myers, & Duran, 2012). In particular, coordination is based on the timely sending and receiving of information required for taking good photographs of designated waypoints. A coordination score (κ) is based on the timing and sequence with which key pieces of information are communicated among teammates (Gorman, Amazeen, & Cooke, 2010). The coordination score (κ) is computed as the amount of time from when information *I* about waypoint *w* is passed from the DEMPC to the team to when feedback *F* about taking a good photograph of waypoint *w* is provided to the team from the PLO. This is then divided by the amount of time from when the PLO and AVO negotiate *N* UAV flight dynamics for waypoint *w* up to when the PLO provides feedback *F* that a good photograph was taken for the waypoint *w*.

$$\kappa = \frac{F_w - I_w}{F_w - N_w}$$

For this project we hypothesized that some subtle timing and coordination behaviors would be absent in the synthetic teammate and would therefore impede coordination and effective performance. These two studies provide some additional baseline information about all-human coordination that will be used to test the synthetic teammate.

Coordinating with Synthetic Teammates

Prior to this project, team communication in the UAS-STE occurred via voice with microphones, headsets and push-to-talk intercom buttons. To avoid speech recognition issues and to transition to the text chat form of communications used commonly in the military, the synthetic teammate and the two human teammates will communicate via text chat. Because text chat is not a transient signal like voice and because communications can occur asynchronously, there is a possibility that coordination among teammates using text chat will be altered. Specifically the coordination score should be impacted by the asynchronous nature of communication. It is unclear whether performance will be affected, but if coordination is made more difficult, performance is also likely to be negatively impacted in this task. Not only will this experiment address questions about coordination and text chat, but will also provide a baseline against which to compare future performance and coordination data when the synthetic teammate is part of a team.

Team performance and coordination can also be affected by human teammates' expectations of the situation and of each other. Experiment 2 was conducted to ascertain how teammates would alter their interactions when they believed that the third teammate was a "synthetic agent". A number of individual behaviors were identified from past data that supported team coordination and these were also noted in Experiment 2.

Experiment 1: Text Chat vs. Voice Communications in the UAS-STE

The purpose of the first study was to collect baseline data in the context of the CERTT UAS-STE task with all human teams communicating via text chat, the mode of communication that will be used with the synthetic teammate. This mode was compared to voice communications, used in previous studies. Also, given the preponderance of text-based communications in our society and its adoption in time critical military and civilian contexts, the comparison of text versus voice as modes of communication is relevant and of increasing importance. By many accounts (Baltes, Dickson, Sherman, Bauer, and LaGanke (2002), Weeks, Kelly, & Chapanis (1974)), the use of test chat may not be the best mode of communication in time-pressured circumstances. The purpose of the experiment was to investigate how text-based communications affect team performance and coordination within the UAS-STE. Based on previous research, we hypothesized that teams communicating with text would coordinate differently from teams communicating using voice and that teams communicating with voice would perform the task better than those using text.

Method

Participants. Twenty, three person teams comprised of college students and the general population of the Mesa, Arizona area voluntarily participated in one 6.5 hour session. Individuals were compensated for their participation by payment of $10.00 per hour with each of the three team-members on the highest performing team receiving a $100.00 bonus. The majority of the participants were males, representing 75.9% of the sample. Individuals were randomly assigned to either a *voice* or *text* chat communication condition. The participants were also randomly assigned to teams and to one of three roles. All members of teams were unfamiliar with each other when they arrived for their sessions.

Equipment and Materials. The experiment took place in the CERTT Laboratory configured for the UAS-STE (described earlier). Participants in the TC condition communicated using the keyboard and a custom-built text communications system designed to log speaker identity and time information. The text communications interface was divided into 3 separate 'modules.' The 'receiver module' alerted participants with a lighted button when a message from another team member was sent. The receiver module also allowed participants to read incoming messages by pressing and holding the F10 key. On releasing the F10 key, the message was then displayed in the 'storage module,' which was comprised of a window that contained previously received messages in a list. Participants were given the ability to scroll through the messages by pressing the F7 and F8 keys. Participants sent messages with the 'transmit module.' To send messages, participants first typed their message in the transmit module window, selected the recipient using the F3, F4, and F5 keys, and then pressed F1 to send. The interface enabled participants to select multiple recipients. Each message was time stamped with when it was sent (F1 key-presses) and when it was received (F10 key-presses) in order to compute coordination scores (κ) and dynamics. Participants in the Voice Communications condition communicated with each other and the experimenter using David Clark headsets and a custom-built intercom system designed to log speaker identity and time information. The intercom enabled participants to select one or more listeners by pressing push-to-talk buttons.

Custom software (seven applications connected over a local area network) ran the synthetic task and collected values of various parameters that were used as input by performance scoring software. A series of tutorials were designed in PowerPoint for training the three team members. Custom software was also developed to conduct tests on information in PowerPoint tutorials, to collect individual taskwork relatedness ratings, to collect NASA TLX and SART ratings, to administer knowledge questions, and to collect demographic and preference data at the time of debriefing. This report will focus on performance and coordination data.

Procedure. The experiment consisted of one 7-hour session (see Table 1). The AVO was located in a separate room adjacent to the other members (DEMPC and PLO). The AVO entered the building through a separate entrance located on the opposite side of the building, and was not allowed to have contact with the other members until debriefing. In the session, the team members were seated at their workstations where they signed a consent form, were given a brief overview of the study and started training on the task.

The number of targets varied from mission to mission in accordance with the introduction of situation awareness roadblocks at set times within each mission. Missions were completed either at the end of a 40-minute interval or when team members believed that the mission goals had been completed. Following each mission, participants were given the opportunity to view their team score, their own individual score, and the individual scores of their teammates. The performance scores were displayed on each participant's computer and shown in comparison to the mean scores achieved by all other teams (or roles) who had participated in the experiment up to that point

Results

Team Performance. Team performance was measured using a composite score based on the result of mission variables including time each individual spent in an alarm state, time each individual spent in a warning state, rate with which critical waypoints were acquired, and the rate with which targets were successfully photographed.

Penalty points for each of these components were weighted *a priori* in accord with importance to the task and subtracted from a maximum score of 1000. Team performance data were collected for each of the five missions.

Team performance was analyzed using a 2 (text, voice) x 4 (mission) mixed ANOVA. Each communication condition (text, voice) had 10 teams. There was a main effect of mission $F(3, 54) = 9.447$, $p < .001$. Teams improved their performance score across the first four missions. There were no significant effects of communication condition, $F(1, 18) = 0.57$, $p < 0.46$, although the voice teams consistently had higher performance scores across all missions than teams in the text chat condition (see Figure 1).

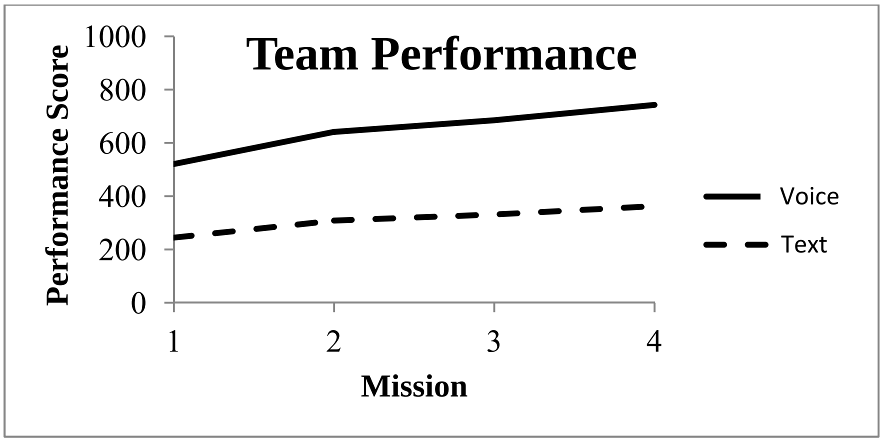

Figure 1. Team performance means for each mission differed over missions, but not condition.

LSD pair-wise comparisons showed that team performance improved over the course of the first four missions, with significant gains between the first two missions ($p = .005$) and between the second and fourth missions ($p = .015$).

Coordination. Based on the inherent time costs of using text chat (e.g., typing, noticing a message arrived, etc.), there was a significant time lag between when a message was sent and when it was received (M = 10.5 s for text; 0 s for voice). To determine if there was a difference in coordination score between voice and text chat, a 2 (communication mode) x 4 (lower workload missions) mixed ANOVA was conducted on coordination scores. There was a significant main effect for which text chat had a significantly lower coordination score than voice (p = 0.042). This is not to say that the voice condition coordinated "better," but only to say that the two communication conditions coordinated differently. Further, a measure that reveals the stability of team coordination dynamics, the Hurst exponent, was also analyzed to determine if there was a coordination stability difference between communication groups. An independent samples *t*-test on the average Hurst exponents across teams revealed that text chat teams were, on average, coordinated in a more stable fashion ($M = 0.9527$, $SD = 0.0131$) than voice teams ($M = 0.8988$, $SD = 0.061$), $t(15) = 2.287$, $p = 0.037$.

For the four low workload missions the median of the performance scores was 310 in the chat condition, with 5 teams below the median and 5 teams above the median. A regression analysis on all the teams combined revealed that the linear trend between communication lag and team performance was significant ($F(1, 38) = 9.06$; $p = 0.005$) indicating that as lag decreased, performance increased. Regression analyses also revealed a positive linear relationship between performance score and Kappa in teams performing above the median performance score ($F(1, 13) = 4.46$; $p = 0.055$). Overall these results indicate that text chat results in different coordination patterns than voice chat and that there is a relationship between these patterns and team performance.

Experiment 2: Expecting a Synthetic Teammate

In this particular study, we examine how teammate interactions (via text chat) are affected by expectations that the pilot is either a synthetic agent or a human teammate. Three person teams were arranged so that the pilot station and pilot were not visible to the other two teammates (mission planner and photographer) and half of the teams were informed that the pilot was a synthetic agent and half were informed that the pilot was a remotely-located human teammate. However, in both conditions the pilot was a human participant. Measures of individual and team performance, coordination, team process, team situation awareness and knowledge were collected over four 40-minute missions. We predict that the expectation that the two teammates are interacting with a human or synthetic pilot will alter coordination and communication patterns. These results will inform design of the synthetic agent and provide baseline data for a Turing-like test of synthetic teammate validity in the next experiment.

Method

Participants. Twenty, three person teams comprised of college students and the general population of the Mesa, Arizona area voluntarily participated in one 6.5 hour session. Individuals were compensated for their participation by payment of $10.00 per hour. Individuals were randomly assigned to one of two conditions: remote AVO or synthetic AVO. The participants were also randomly assigned to teams and to one of three roles. All members of teams were unfamiliar with each other when they arrived for their sessions.

Equipment and Materials. Participants all used the custom text chat capabilities as in the previous study except the interface was slightly improved to make it easier to use. All other equipment and materials were identical to Experiment 1.

Measures. For the purpose of this experiment the measures used in Experiment 1 were also used in Experiment 2. In addition a performance score was calculated for each target based on the timely and accurate processing of a target. In addition a set of behaviors related to team coordination were identified in previous data sets and were noted whenever they occurred in this study. The behaviors are listed in Table 1.

Table 1.
Individual Behaviors Supportive of Team Coordination

Negative Communication
- Argue – *DEMPC and AVO can argue over the best way to give upcoming waypoint restrictions?*
- Specific to chat conditions
    - Timing – AVO sends text asking for next waypoint just as DEMPC texts the next waypoint info.
    - Lag in response – PLO asks a questions that is not answered until multiple unrelated texts have been posted.

Positive communication
- Help out – *PLO tells DEMPC, "Please give info next target info to AVO."*
- Acknowledge members' speech – *"Roger that."*
- Give praise – *Good job guys!*
- Check with others before implementing a decision – *PLO asks AVO, "I am about to take a pic, are we at 2000 feet?"*
- Clarification – AVO asks DEMPC to clarify what was meant in a previous message.

Repeated Requests
- Same info or action requested two or more times
- PLO asks repeatedly for information needed to take a photo.

Unclear Communications
- Misspellings, ambiguous terms, experimenter cannot understand

General Status Update
- Inform others of current status – AVO tells PLO "I am at 2500 feet now."

Inquiry About Status of Others
- Inquire about current status of others – *DEMPC asks AVO "How are we doing on our heading/fuel etc."*
- Express concern – *DEMPC asks AVO "Are we headed to the next target? We appear to be off course."*

Planning
- Anticipate next steps – *AVO asks DEMPC, "Where are we going after LVN?"*

Suggestions to Others

- Make suggestions to other members – *DEMPC tells AVO to increase speed in route to targets and slow down upon arrival.*

Procedure. Three person teams were arranged so that the pilot station and pilot were not visible to the other two teammates (mission planner and photographer) and half of the teams were informed that the pilot was a synthetic agent and half were informed that the pilot was a remotely-located human teammate. However, in both conditions the pilot was a human participant. Measures of individual and team performance, coordination, team process, team situation awareness and knowledge were collected over four 40-minute missions. In all other respects the procedure was like that of Experiment 1.

Preliminary Results

Data collection is ongoing, but preliminary observations of experimenters indicate that in the synthetic teammate condition, the human participants tend to ignore the synthetic teammate and speak tersely, eliminating any of the social or polite discourse that typically occurs with a human teammate.

Conclusion

In these studies we have set the stage for synthetic teammate validation by 1) collecting baseline data in the UAS-STE using text chat, 2) by identifying individual behaviors that support coordination, and 3) by understanding human teammate predilections to interact differently with a synthetic teammate vs. a human teammate.

References

Ball, J. Myers, C., Heiberg, A. Cooke, N. J., Matessa, M., Freiman, M., & Rodgers, S. (2010). The synthetic teammate project. *Computational and Mathematical Organization Theory*, *16,* 271-299. DOI 10.1007/s10588-010-9065-3

Baltes, B. B., Dickson, M. W., Sherman, M. P., Bauer, C. C., & LaGanke, J. S. (2002). Computer-mediated communication and group decision making: A meta-analysis. *Organizational Behavior and Human Decision Processes, 87*(1), 156-179.

Cooke, N. J., Gorman, J. C., Myers, C. W., & Duran, J.L. (2012). Interactive Team Cognition, *Cognitive Science, 1-31,* DOI: 10.1111/cogs.12009.

Cooke, N. J., & Shope, S. M. (2005). Synthetic Task Environments for Teams: CERTT's UAV-STE. In N. Stanton, A. Hedge, K. Brookhuis, E. Salas & H. Hendrick (Eds.), *Handbook of Human Factors and Ergonomics Methods* (pp. 46-41 - 46-46). Boca Raton, FL: CRC Press.

Gorman, J. C., Amazeen, P. G., & Cooke, N. J. (2010). Team Coordination Dynamics. *Nonlinear Dynamics Psychology and Life Sciences, 14,* 265-289.

Weeks, G. D., Kelly, M. J., & Chapanis, A. (1974). Studies in interactive communication: V. Cooperative problem solving by skilled and unskilled typists in a teletypewriter mode. *Journal of Applied Psychology, 59*(6), 665-674.

Acknowledgements

This work is supported by ONR N000140910201 and N00014-11-1-0844 to the Cognitive Engineering Research Institute. The authors would like to thank Jasmine Duran, Jamie Gorman, Harry Pedersen, and F. Erik Robinson for their contributions to the first study.