EVALUATING COMMUNICATION IN A SYNTHETIC TASK ENVIRONMENT

2Lt Christopher Hetrick, 2Lt Gary Cooper, 2Lt Kevin Walsh, 2Lt Ryan Smith, and 2Lt Adam Wickes
United States Air Force Academy, Colorado, USA
cdhetrick@hotmail.com

Maj Terence Andre
Air Force Research Laboratory, Mesa, AZ, USA
terence.andre@williams.af.mil

INTRODUCTION

The ability to efficiently and accurately collect and evaluate group communication during task performance is indispensable when analyzing the overall effectiveness of an organization or team. The possibility of automating, or partially automating, the process of collecting and evaluating communication between team members involved in decision-making and problem solving in a synthetic task environment could be valuable in future communications research and analysis. Such a system could possibly be utilized to gain a better understanding of different ways to improve overall team and organizational performance in many different areas of research.

In his article, *Pilot Speak*, Spinetta (2001) describes common properties of effective organizational communication that focus on communication between Air Force pilots, but that he claims also apply to any situation. The first element is that communications should be directive and descriptive. The speaker should tell the receiver what he wants done and how he wants it done. In addition, effective communication should specifically identify who should accomplish the actions contained in the instructions in order to avoid confusion. Transmissions should also be concise and to the point. The key to effective communication is to relay the most information in the fewest words possible (Spinetta, 2001). This idea is reinforced by a study of 237 undergraduate students' performance in a team tank simulator, conducted by Marks, Zaccaro, and Mathieu (2000), in which they determined that the quality, not the quantity, of team communication is positively associated with team performance. Another important quality of team communication is the support of open communication among team members (Spinetta, 2001). All relevant information is important, regardless of the position or rank of the source. This aspect of communication is a significant consideration, as demonstrated by Palmer and Lack (1995) in a study of crew resource management in Air Force aircraft, which showed that in typical aircraft crews and formations, communication tends to be dominated by the authority figures. Also, communication that is intended to keep the team coordinated and together, such as information relating to progress and location, is important (Spinetta 2001).

COMMUNICATION METRICS

In order to effectively automate communication analyses, communication metrics must be established. A study conducted by Dutoit and Bruegge (1998) quantitatively measured communication traffic on electronic bulletin boards in a problem solving environment by recording the number of messages sent by each team, the number of noun phrases contained in each message, and the number of unique noun phrases. The data was then analyzed using a set of natural language processing tools. The

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
OCT 2002	**Conference Proceedings**	**01-01-2001 to 30-09-2002**

4. TITLE AND SUBTITLE
Evaluating Communication in a Synthetic Task Environment

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
62205F

6. AUTHOR(S)
Christopher Hetrick; Gary Cooper; Kevin Walsh; Ryan Smith; Adam Wickes

5d. PROJECT NUMBER
1123

5e. TASK NUMBER
B0

5f. WORK UNIT NUMBER
1123B000

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
United States Air Force Academy,2304 Cadet Drive #324,US Air Force Academy,CO,80840

8. PERFORMING ORGANIZATION REPORT NUMBER
; AFRL-RH-AZ-PR-2002-0005

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory/RHA, Warfighter Readiness Research Division, 6030 South Kent Street, Mesa, AZ, 85212-6061

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL; AFRL/RHA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-RH-AZ-PR-2002-0005

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
Published in the Proceedings of the 44th Annual Conference of the International Military Testing Association (2002), 257-262, Ottawa Canada, 22-24 Oct 02

14. ABSTRACT
The three primary objectives of this project were to develop communication metrics, evaluate the potential of voice recognition software, and to decide if using the two together to analyze team communication would be a viable option to explore for communication metrics. Following initial evaluations, some aspects of our developed communication metrics show positive correlations between the measures we developed and performance on the synthetic task, and further research in this area could prove beneficial. Another promising aspect is that, even though we had to perform many of the counts and measures by hand, the same measures would not be very difficult to automate, decreasing the evaluation time significantly.

15. SUBJECT TERMS
Synthetic Task Environment; STE; Communication; Group communication; Task performance; Teams; Team performance; Organizational communication; Team communication; Voice recognition software; Voice communications; Evaluation

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**Public Release**	**6**	

study showed that good communication metrics evaluate the information flow by measuring the volume and complexity of information exchange. The Common factors found by Dutoit and Bruegge (1998) in their experiment were word counts, transmission counts, noun counts, and unique term counts.

A second area of focus in developing an automated system is communications recording. Oviatt (2000) described the primary concerns regarding automated speech-recognition programs. The first aspect she discussed was the program's ability to effectively process natural language. A study by Furman and Cosky (1999), explained that the most accurate level of software utilizes grammar-based speech recognition. Grammar-based programs utilize principles of grammar to deconstruct spoken words and reconstruct them for processing, which allows programs to recognize and record natural language and speech patterns.

Another important aspect of an effective speech-recognition program, especially in analyzing communication, is the program's ability to process dialogue (Oviatt 2000). One of the chief challenges in this area is the program's ability to discern between the individuals engaged in the dialogue in order to accurately record who is speaking. A significant consideration is the program's ability to handle speakers stepping on, or interrupting, each other (Furman & Cosky, 1999). Closely related to the program's ability to handle dialogue is the program's capability in regards to multiperson use. One of the chief challenges with speech-recognition programs deals with recognizing different individual voices, and at this point in time, most have to be trained to recognize voices that are intended to be recorded (Oviatt, 2000). The final aspect of voice-recognition software deals with error handling. It's important to realize that voice-recognition software is not perfect and users must devise a method to identify and sufficiently deal with recording errors (Oviatt, 2000).

This research primarily focused on developing communications metrics and incorporating with a voice recognition program. In selecting an effective voice recognition software system, we had to be aware of the potential risks inherent in the systems. We knew that we had to focus on such aspects as processing power, grammar-based recognition, individuation ability, and error handling when selecting our system. One of the greatest challenges of the project was developing communication metrics, primarily because it is a fairly new and underdeveloped area of research that is highly subject to the specific situation to which it's intended to be applied.

METHOD

Participants

The population sample consisted of eight, two-person teams composed of members from all four classes at the United States Air Force Academy who volunteered to participate in the study. Each team was randomly selected with subjects varying in experience in the performance of synthetic tasks.

Materials

The system consisted of Dragon Naturally Speaking voice recognition software, a synthetic task environment for the subjects to complete, and a component to measure communication metrics. The synthetic task environment selected was Commandos 2: Men of Courage, due to its high degree of user-friendliness and the relatively low amount of gaming skill required to complete the synthetic task. In addition, the division of tasks among the users during game play forced team members to work together to successfully complete the mission, which ensured adequate verbal transmissions to evaluate team communication. Dragon Naturally Speaking was selected because it was the best available option at translating verbal communication into a written document. However, as a precautionary measure, a tape recorder was also used to make backup recordings of team communications.

The communications metrics was developed based on the aspects of quality communication described by Spinetta (2001) and aspects of communication measurement described by Dutoit and Bruegge (1998).

Dutoit and Bruegge (1998) explain that the most simple, yet effective, measures of communication involve word counts, transmission counts, noun counts, and unique term counts, most of which are possible to achieve through automation. Qualities of effective communication described by Spinetta (2001) are: directive, descriptive, and concise transmissions, clear identification of who should carry out the operation, open communication between team members, and communication intended to establish progress. To combine these two approaches for evaluating communication, we modified a word and transmission count method in order to apply it to the qualities described by Spinetta (2001).

First, we performed simple word and transmission counts for each subject and each group as a whole. We then counted verbs in order to assess descriptive aspects of the communication and directive orders (i.e. "Search him and get the cigarettes") to assess the directive aspect of the communication. Next, we counted identifiers (i.e. "You," "I," or names) to assess clear identification of who was to carry out various operations. We also measured statements and questions intended to measure progress (i.e. "Where should we go now?"). Finally, we divided the word count by the transmission count for each subject to assess conciseness in the communication and found the word count ratio between subjects to measure the balance of communication between subjects.

Procedure

Testing was conducted in the US Air Force Academy library computer lab using two Pentium III computers with LAN network connections for the multiplayer synthetic task and two Pentium III laptop computers for the voice recognition software. The subjects began the experiment by training their voices to the voice recognition software for approximately ten minutes. The next step involved a demonstration of the controls and actions that were required to complete the synthetic task. For this training, a demonstrator played through the synthetic task, explaining what he was doing, why he was doing it, and how he was operating the characters while the subjects watched. Following the voice recognition and synthetic task training, the subjects were set up on the system and allowed to start the mission.

The mission consisted of two playable characters, each with unique capabilities, all of which were required to successfully complete the level. Each character was solely assigned to one of the two subjects. Each group was allowed to play until they completed the mission, with restarts allowed using the quick-save function on the game. Their communication was recorded using the voice recognition software and by holding the tape recorder in between the two group members. During game play, researchers evaluated each group on their performance using six measures: Survival, completion time, secrecy, non-lethal tactics, completed objectives, and completed secondary objectives. At the completion of the task, each subject's voice recognition software recording document was saved to disk.

Since the communications transcriptions provided by Dragon Naturally Speaking were so inaccurate that they were ultimately useless, the data was prepared by transcribing each group's communications from the tape recording to a word document. The transcription process took approximately 90 minutes to complete per group. Once the communications were transcribed, each transcription was evaluated using the word count function in Microsoft word and by marking and physically counting the number of transcriptions, nouns, directive orders, progress statements/questions, identifiers, and by calculating the conciseness and balance of communication. This aspect of the data preparation process took approximately 60 minutes per group.

RESULTS

A preliminary data analysis was performed using Microsoft Excel. To evaluate the usefulness of each aspect of our communications metrics, we ran a correlation test between each group's communication measurement and their overall score on the game. Initially, we found weak, but significant, correlations for identifiers (r^2=.30) and conciseness (r^2=.35). However, upon closer inspection of the data, it appeared that two outliers were significantly affecting our data. One was the best group that took nearly half the time of the other groups and the other was the worst group, which took significantly longer. In such instances, the time alone required for each group significantly affected the word count, thus making it no more valid than simply timing the game as a predictor of performance. After removing these two outliers, we found significant correlations for word count (r^2=.59), transmission count (r^2=.54), noun count (r^2=.68), and directive statements (r^2=.39). These results are shown in Figures 1 through 4, respectively.

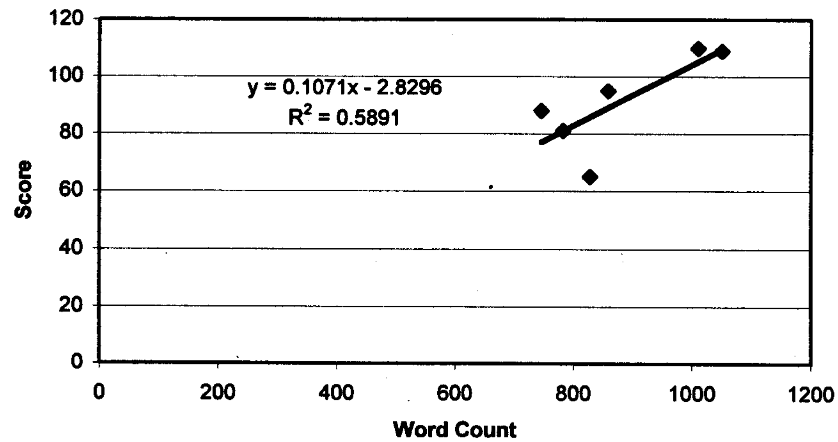

Figure 1: Performance correlation based on word count and score (outlier removed).

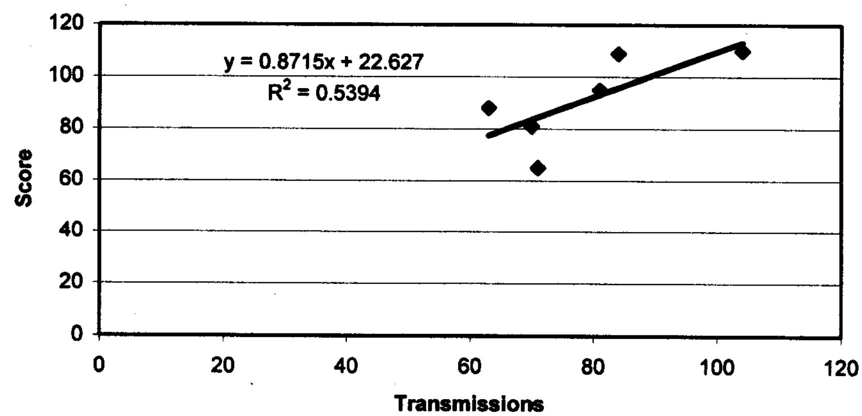

Figure 2: Performance correlation based on transmission count and score (outlier removed).

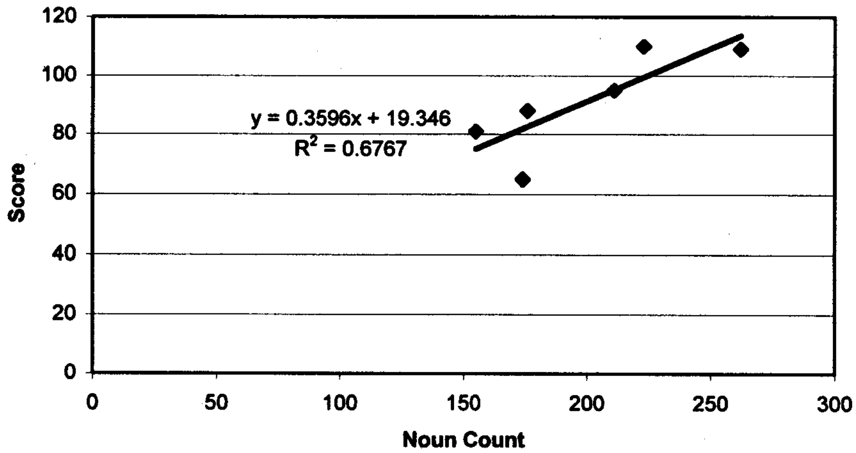

Figure 3: Performance correlation based on noun count and score (outlier removed).

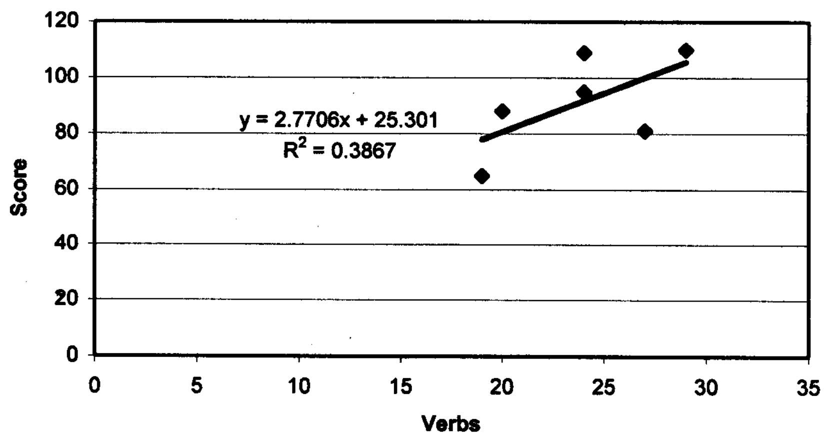

Figure 4: Performance correlation based on directive count and score (outlier removed).

DISCUSSION

The three primary objectives of this project were to develop communication metrics, evaluate the potential of voice recognition software, and to decide if using the two together to analyze team communication would be a viable option to explore for communication metrics. Following initial evaluations, some aspects of our developed communication metrics show positive correlations between the measures we developed and performance on the synthetic task, and further research in this area could prove beneficial. Another promising aspect is that, even though we had to perform many of the counts and measures by hand, the same measures would not be very difficult to automate, decreasing the evaluation time significantly.

The single biggest hindrance to the efficiency of the communication evaluation is the incapability of the voice recognition software, which ultimately proved to be useless. However, the voice recognition software was off-the-shelf software, and better, more accurate software could possibly help with this problem. If the technology improves to the extent of being able to accurately transcribe communication to a word document, voice recognition software would be extremely useful.

Analyzing verbal protocol data in order to understand team performance can take a significant amount of time. Sanderson and Fisher (1994) report that some analysis techniques take as much as 10 hours for every hour of communication data. Our analysis of communication took approximately 2.5 hours per 15 minutes of communication data. Therefore, our analysis would benefit from a method where the processing time could be reduced and still obtain an accurate representation of the team process. When voice recognition software improves, it could turn into a viable option for effectively and efficiently evaluating team communication. For now, the method has to wait for the technology to catch up, but as soon as technology becomes sufficient, it should definitely be explored further.

REFERENCES

Dutoit, A.H., & Bruegge, B. (1998). Communication Metrics for Software Development. *IEEE Transactions on Software Engineering, 24*(8), 615-627.

Furman, D.S., & Cosky, M.J. (1999). Speech-Based Services. *Bell Labs Technical Journal, 4*(2), 88-98.

Marks, M.A., Zaccaro, S.J., & Mathieu, J.E. (2000). Performance Implications of Leader Briefings and Team-Interaction Training for Team Adaptation to Novel Environments. *Journal of Applied Psychology, 85*(6), 971-986.

Oviatt, S. (2000). Designing the User Interface for Multimodal Speech and Pen-Based Gesture Applications: State-of-the-Art Systems and Future Research Directions. *Human-Computer Interaction, 15*(4), 263-323.

Palmer, M.T., & Lack, A.M. (1995). Communication Conflicts of Status and Authority in Dyadic, Task Based Interactions. *Journal of Language & Social Psychology, 14*(1/2), 85-102.

Sanderson, P.M., & Fisher, C. (1994). Exploratory Sequential Data Analysis: Foundations. *Human-Computer Interaction, 9*, 251-317.

Spinetta, L. (2001, September). Pilot Speak. *Torch*, 18-21.