A TEAM PERFORMANCE DATA COLLECTION AND ANALYSIS SYSTEM

Stephanie Guerlain, Thomas Shin, Hui Guo, Reid Adams and J. Forrest Calland
University of Virginia
Charlottesville, VA
www.sys.virginia.edu/hci

At present there exists no commercially available product capable of capturing and broadcasting multimedia audiovisual data from teams performing high-risk work. We have developed such a recording and analysis system for the purpose of studying team behavior, which we currently use to observe and record up to 8 people working as a co-located team in a hospital operating room. The system has four data collection computers, each recording one video stream and up to 2 audio feeds. A separate software package is used to synchronize and view the audio/video streams together on a fifth computer. This software has several annotation and scoring features which can be used either for data analysis or for team debriefing purposes. Although currently being used for patient safety research in the operating room, this system could also be adapted to collect and analyze team behavior in other domains, even for participants who are distributed.

INTRODUCTION

Patient safety has come to national attention, due to the estimated high rates of adverse events leading to unnecessary complications and death. In a study of patients admitted to New York State Hospitals in 1984, a group of Harvard researchers found that more than two third of adverse events were preventable (Leape, Brennan, Laird, Lawthers, Localio, Barnes, Hebert, Newhouse, Weiler, and Hiatt, 1991). In 1999, the Institute of Medicine published an extensive patient safety report, which emphasized the importance of developing methods and standards to reduce the estimated 44,000 to 98,000 deaths each year due to medical error (Institute of Medicine, 1999).

Our research group has been focusing on patient safety interventions in the operating room. Specifically, we are evaluating the effectiveness of interventions (e.g., the use of a checklist and the use of a preoperative briefing) on the communication patterns and technical performance of operating room personnel. Thus, we are faced with the challenge of measuring the communication, coordination, and performance of a fairly large team (up to 8 people at a time – see Table 1) in a surgical operating room.

The operating room is a high-risk environment where the life of a patient is in the hands of the surgical team. However, performance in the operating room (OR) is relatively non-standardized compared to other high-risk industries. Due to the challenges of collecting data in the OR, teamwork has primarily been studied in smaller, controlled, simulated environments with anesthesiologists and nurses using an anesthesia simulator, with the role of the surgeon played by whatever surrogate individual is available regardless of training or specialty (Gaba, 1989). This work has yielded important information about the use of crew resource management techniques in a simulated environment. However, currently there are no viable surgery simulators that would allow for realistic whole-team simulation of operative procedures. Thus, our focus is on studying surgery team performance in the operating room.

Table 1. Surgery Team Members During Laparoscopic Cholecystectomy

Person	Role
Surgery Attending (SA)	Directs SR
Surgery Resident (SR)	Conducts surgery with guidance from SA
Camera Operator (CO) (optional)	Manipulates laparoscopic camera
Anesthesia Attending (AA)	Directs AR
Anesthesia Resident (AR)	Performs anesthesia duties
Scrub Nurse (SN)	Provides instruments to surgeons
Circulating Nurse (CN)	Handles data entry and communications
Scope Tech (ST)	Handles OR equipment
Patient care assistant (PCA)	Moves patient, prepares room

TEAM EVALUATION METHODS

Much of the current team performance literature discusses what makes an effective team – see (McNeese, Salas, and Endsley, 2001) and (Cannon-Bowers and Salas, 1998) for a thorough overview of these issues. Relatively few articles, however, discuss how team measures are obtained.

Evaluating coordinated team activities requires independent observation of multiple (potentially distributed) individuals. Analysis of these independent observations requires that they be coordinated, coded, and correlated before objective scoring or subjective global assessment can take place.

One common technique used in military team performance evaluation is to have a set of trained observers, with each observer watching one team member, rating that team member's performance (Campbell, Freeman, and Hildebrand, 2000). The difficulty with this approach is the requirement to have the appropriate set of subject matter experts available and trained for observation, and the difficulty with correlating events and actions across the team members.

Another methodology involves having one trained observer watch a whole team, and give overall scores after procedure segments, as performed during aviation crew resource management ratings given to pilot teams (Helmreich, Butler, Taggart, and Wilhelm, 1995). This can be difficult, however, as teams get larger. Furthermore, with both of these techniques, team performance is often measured using some form of a note-taking methodology. With this method, there are risks of missing certain aspects of events when taking notes, as well as variations in different observers, without the ability to go back and reanalyze the events. With no "integrated trace" of actual activities, scoring can not be re-evaluated nor can events be easily recalled for debriefing purposes.

Recording team sessions can be used for creating a trace of team behavior. In naturalistic settings, such recording includes capturing video and audio data of the human team members as well as the artifacts in the work environment. Collecting and analyzing such data poses serioius challenges. The LOTAS group (Xiao and The LOTAS Group, 2001) has made excellent progress in coordinating analog video, audio and eye tracking data to try to understand team performance (during trauma resuscitation). The limits of analogue data collection, however, become readily apparent when one attempts to synchronize multiple feeds. This can be done using a video mixer, in which different video feeds are recorded together using a split screen or picture in picture technique, but then individual audio and video feeds are lost, making it impossible to isolate or focus on individual data streams. Furthermore, scrolling within a session or skipping between segments is time consuming, and analogue video editing is an extremely difficult and tedious process. Due to confidentiality reasons, audio/video recordings of subjects often needs to be destroyed within short periods, so time consuming video editing processes are extremely problematic, further necessitating a digital video recording system.

REMOTE ANALYSIS OF TEAM ENVIRONMENTS

The primary purpose of our system, entitled **R**emote **A**nalysis of **T**eam **E**nvironments (RATE), is to automate as much as possible the ability to digitally record, score, annotate and analyze team performance for a team of up to 8 people.

RATE is a digital audiovisual data collection and analysis system. The system is designed to be mobile, allowing the researcher to collect data from team members working in various locations simultaneously (although we are currently collecting co-located data).

RATE Hardware Architecture

The data collection process is similar to an analogue data collection system but instead of using a multiplexer and VCR to record audiovisual data, four high end PC computers with video capture cards are used. The computers are used to process, store, and broadcast the data while any PC connected to the local area network can view all of the data in a synchronized fashion using the RATE software (see Figures 1 and 2).

Four computers are used to record four separate videos and 8 separate audio tracks (one video feed and two audio feeds are encoded into each computer). In our operating room research, the four videos are comprised of a room camera (with the circulating nurse's voice), an overhead camera (with the scrub technician's and camera operator's voices), a view of the physiological data (with the anesthesiology attending's and resident's voices), and the intracorporeal laparoscopic image (with the surgery attending's and resident's voices). The purpose of recording each individual's voice in separate audio track is to make transcription/listening easier (by selectively listening to certain voices during replay).

RATE Software

The software for the RATE system is written in Microsoft Visual Basic for the Windows platform, and has two main functions. The first function is to allow viewing the case remotely using streaming real-time audio and video. The second function is to facilitate performance scoring and case review.

The RATE software allows the researcher to:

- Record up to 4 different video feeds (co-located or distributed), with up to 2 audio feeds for each video feed
- Watch up to four synchronized videos and eight audio streams simultaneously
- Selectively mute/enable various audio and video feeds
- Annotate or transcribe segments of video
- Count verbal utterances (by manually clicking on a representation of each team member each time they speak)
- Code episodes with automatic time stamping for easy review
- Score the technical proficiency of the team. In our operating room studies, we use a standard surgery technical skills scoring system: (Eubanks, Clements, Pohl, Williams, Schaad, Horgan, and Pellegrini, 1999).

Whether the analyst is scoring the case or performing an utterance count, every event is annotated with the time code and categorized. Events tabulated within the scoring system are catalogued under a separate, dedicated scoring window, while verbal utterance counts are tabulated separately elsewhere. The analyst also has the ability to annotate episodes or segments with comments, which are cataloged under annotation.

Unlike similar programs using analogue video by controlling a VCR, e.g., MacShapa, (Sanderson, Scott, Johnston, Mainzer, Watanabe, and James, 1994), RATE directly accesses the streamed or stored digital video source (and runs on the Windows platform). One of the immediate advantages of using digital video involves the capacity to skip from segment to segment in a fraction of time it takes to navigate analogue video. Any time-stamped event can be "double clicked" and all the audio/video data will play from that point (within a few seconds).

A second advantage is that the audio and video streams can be isolated from each other for easier analysis.

DISCUSSION

The RATE system has now been used to record 33 operative procedures. Through iterative design and testing of the hardware and software, we now have a system that allows us to record and score a case in real time, with data analysis partially automated through the RATE software. We are using the tool as a data collection and analysis system for a randomized, controlled trial of the use of a checklist on the improved technical and communication performance of surgery teams performing laparoscopic cholecystectomy (gall bladder removal). Initial results of that project have been reported elsewhere (Calland, 2001).

Thus far, the RATE tool has been solely employed to collect and analyze team-performance data, but will also be evaluated in the near future for its potential utility as (surgical) team debriefing and training tool, creating a possibility of training surgical "situational awareness" and improved team communication and coordination strategies. Situational awareness is the ability to recognize relevant cues that indicate the need to perform a certain type of action or predict future state (Endsley, 1988). Team communication training

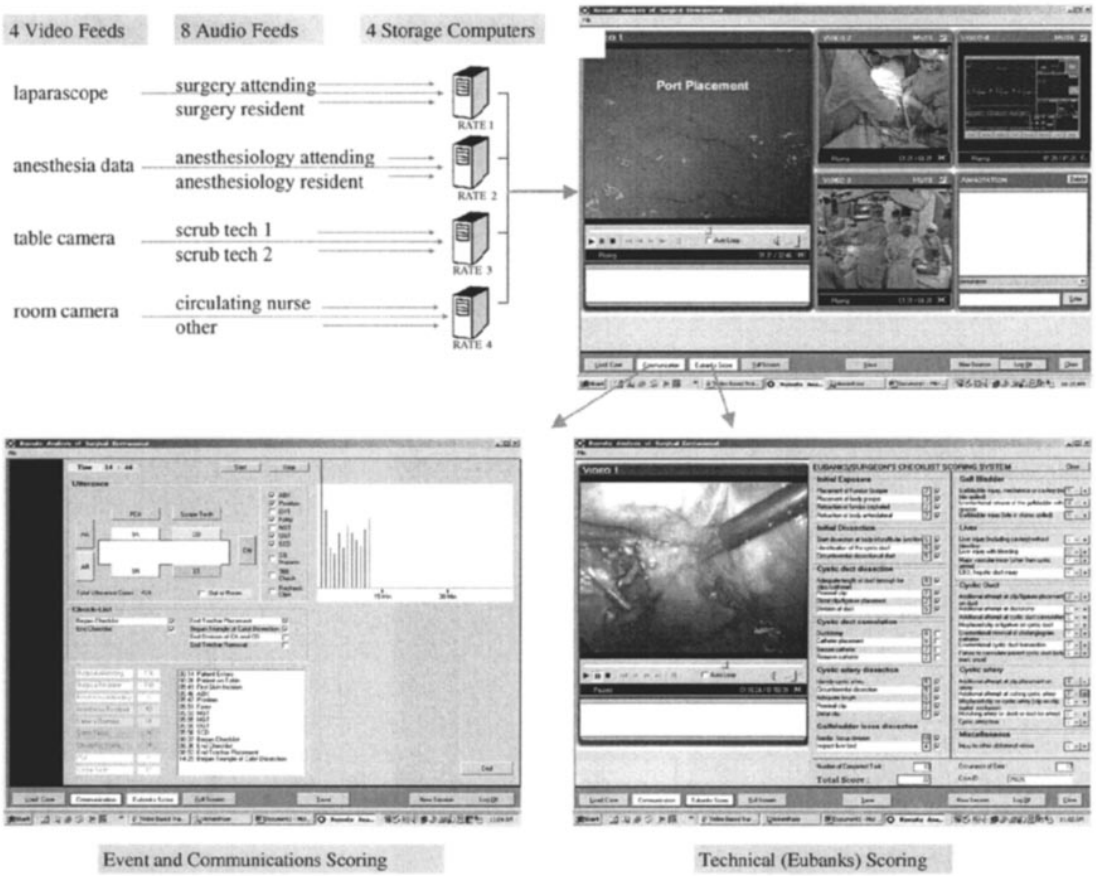

Figure 1: The RATE Hardware/Software System

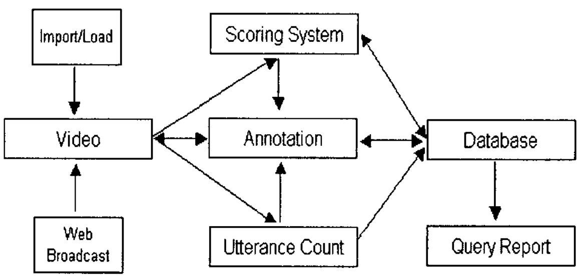

Figure 2: Functional structure of RATE

focuses on creating the team analog of individual situation awareness, known as a "shared mental model" through effective communications and workspace sharing across team members (Orasanu and Salas, 1995).

CONCLUSION

We have developed a system to record distributed team behavior, with a means to review, score, and annotate this behavior either in real time or post-hoc in a synchronized fashion. Although we are currently using the data collection system to observe and record up to 8 people working as a co-located team, the four data collection servers could be distributed if team members were distributed (such as Navy personnel working together on a problem, or Air Traffic Control personnel interacting with pilots and dispatchers during flight rerouting).

ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 0092985, the National Patient Safety Foundation, and Karl Storz Endoscopy of America. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

Calland, J. F. (2001). Error Reduction During Laparoscopic Cholecystectomy. Paper presented at the American College of Surgeons Clinical Congress -- Plenary Sessions, New Orleans, LA.

Campbell, G., Freeman, J., & Hildebrand, G. (2000). Measuring the impact of advanced technologies and reorganization on human performance in a combat information center. *Proceedings of the Human Factors and Ergonomics Society 44th Annual Meeting, 6*, 642-645.

Cannon-Bowers, J., & Salas, E. (Eds.). (1998). *Making Decisions Under Stress: Implications for Individual and Team Training*. Washington, D.C.: American Psychological Association.

Endsley, M. (1988). *Design and evaluation for situation awareness enhancement.* Proceedings of the 34th Annual Meeting of the Human Factors and Ergonomics Society, Santa Monica, CA, 97-100.

Eubanks, T. R., Clements, R. H., Pohl, D., Williams, N., Schaad, D. C., Horgan, S., & Pellegrini, C. (1999). An objective scoring system for laparoscopic cholecystectomy. *J Am Coll Surg, 189*(6), 566-574.

Gaba, D. M. (1989). Human error in anesthetic mishaps. *Int Anesthesiol Clin, 27*(3), 137-147.

Helmreich, R., Butler, R., Taggart, W., & Wilhelm, J. (1995). *The NASA/University of Texas/FAA Line/LOS Checklist: A behavioral marker-based checklist for CRM skills assessment.* Austin, TX: NASA.

Institute of Medicine. (1999). *To Err is Human: Building a Safer Health System.* Washington, D.C.: National Academy Press.

Leape, L., Brennan, T., Laird, N., Lawthers, A., Localio, A., Barnes, B., Hebert, L., Newhouse, J., Weiler, P., & Hiatt, H. (1991). The nature of adverse events in hospitalized patients. Results of the Harvard Medial Practice Study II. *The New England Journal of Medicine, 324*(6), 377-384.

McNeese, M., Salas, E., & Endsley, M. (Eds.). (2001). *New Trends in Cooperative Activities.* Santa Monica, CA: HFES.

Orasanu, J., & Salas, E. (1995). Team decision making in complex environments. In C. Zsambok (Ed.), *Decision Making in Action: Models and Methods* (pp. 327-345). Norwood, NJ: Ablex.

Sanderson, P., Scott, J., Johnston, T., Mainzer, J., Watanabe, L., & James, J. (1994). MacSHAPA and the enterprise of Exploratory Sequential Data Analysis (ESDA). *International Journal of Human-Computer Studies, 41*, 633-638.

Xiao, Y., & The LOTAS Group. (2001). Understanding coordination in a dynamic medical environment: Methods and results. In M. Endsley (Ed.), *New Trends in Cooperative Activities* (pp. 242-258). Santa Monica, CA: HFES.