Communication Analysis for Virtual Mission Preparation: Moving Beyond Simple Replay

Sam Hasenbosch
Boeing Defence Australia
samuel.j.hasenbosch@boeing.com

Peter Ross
Defence Science & Technology Organisation
peter.ross@dsto.defence.gov.au

Abstract. Synthetic training environments offer the opportunity to collect a wealth of objective data that may be exploited for participant feedback, assessment and human factors research purposes. A key element of this is communication between trainees and role players. Analysis of such interaction can provide insight into aspects such as cognitive workload, task-specific issues, and dimensions of teamwork. In practical terms, the difficulty in collecting a complete and accurate record of conversation, as well as the time required to analyse patterns, can restrict the use of communication data to replay in an after action review. The consequence of this approach is a focus on mission outcomes, rather than the supporting behaviours which enable performance. Yet, is it possible to automate elements of communication analysis to provide timely insight to trainees, assessors, facilitators and researchers? This paper describes the challenges in capturing voice communication in the context of simulated command team training, specifically military Air Battle Management exercises, and proposes procedural and technical solutions to these issues. These approaches may be equally applicable to text data, other domains such as the emergency services, and across the continuum from individual to large scale training. The paper further proposes a set of feasible extraction methods and communication metrics, and discusses how these could add value to virtual mission preparation.

1. INTRODUCTION

Synthetic training environments rely upon emulated or stimulated work interfaces, in combination with professional role players and facilitators, to represent the real-world systems and scenarios within which trainees perform their duties. A series of command team training (CTT) research exercises conducted by the Defence Science and Technology Organisation (DSTO), each known as Black Skies, employs such an environment. These simulated exercises have involved the training of a RAAF Air Battle Management (ABM) team in preparation for a large-scale, multi-national live training exercise, known as Exercise Pitch Black (Best et al., 2009; Shanahan et al., 2009). This style of training is known as virtual mission preparation.[1]

The simulation technology used to create synthetic training environments often describes scenario events and participant interactions in standard formats, such as the Distributed Interactive Simulation (DIS) protocol. Various tools exist to record and replay DIS logs and other observational data. Despite earlier work on automated performance measurement systems (Schreiber, Watz & Bennet, 2003), the capability to perform timely and insightful analysis of this data is yet to be fully realised (Tracey et al., 2009).

A key aspect of CTT is the ability to evaluate communication behaviour, relate this to mission performance and present the findings to the participants (Salas et al., 2009). The accurate and complete approach to assessment requires a full transcript of conversation, correlation with scenario events, and expert interpretation. Unfortunately, given the scale of a Black Skies exercise, the time and expertise required to collect such data is prohibitive. Instead, a subject matter expert (SME) is employed to observe and assess a given team in real-time, during which communication behaviour is one of many aspects competing for their attention. Consequently, subjective ratings of communication performance are supported only by the replay of selected events, in the form of synchronised video, audio and other simulation data.

Human factors researchers have endeavoured to develop measures of communication performance that do not require transcripts (Gorman et al., 2012; Strang et al., 2012). This approach could reduce the burden on observers to collect and analyse data. These efforts require sufficient data to verify and validate measures in a given domain. This means that the capability to extract must precede the design of such metrics.

Fortunately, if implemented properly and combined with improved data collection procedures, the DIS protocol data set enables a degree of automated analysis. This paper describes an investigation of what systems would be required and proposes measures of performance for trial in future CTT exercises.

1.1 Motivation

Communication analysis could enhance activities other than ABM training. Tailored visualisations may assist trainees in any domain to recognise failures or error-prone behaviours, and regulate their own performance more easily than with simple replays. Furthermore, pervasive data collection allows researchers to discover what enables effective teams, and suggest improvements to training processes accordingly.

[1] Other terms include Distributed Mission Training (DMT) and Mission Training through Distributed Simulation (MTDS).

2. COMMAND TEAM TRAINING

The design of CTT exercises must satisfy the needs of its stakeholders. Table 1 lists the nominal objectives for each generic participant in a CTT activity. These influence the requirements for communication analysis.

Table 1: CTT Stakeholder Objectives

Stakeholder	Objective(s)
Trainees	Understand how their actions helped or hindered team performance.
Assessors	Rate task execution and identify positive and negative behaviours.
Facilitators	Manipulate scenario variables to provide the desired training experience.
Role Players	Represent specific entities or agencies in a realistic manner.
Researchers	Understand the work domain; develop measures of team performance and improved training strategies.

Another important aspect is the nature of the work domain. Each training environment presents unique challenges to collecting an accurate and complete record. In order to appreciate the data collection challenges in Black Skies, some understanding of the training audience and processes in this series of exercises is required.

2.1 Air Battle Management

ABM involves the conduct of tactical operations to maintain control of the air for friendly forces. It is accomplished through the direct control of a range of airborne and ground-based assets, including fighter and tanker aircraft, via multiple communication channels. It includes activities such as the surveillance of a defined area of operations, interception of unauthorised aircraft within restricted airspace, and the preservation of defence-in-depth over the near term.

An ABM team consists of several aircraft controllers and a director who must monitor the current situation, allocate tasks and resources, and coordinate with a number of external agencies. Due to the interdependent nature of the work, there is considerable communication between ABM team members, the aircraft under their control, and adjacent airspace users. The information exchanged encompasses plans, directions, advice and status. Communication, both internal and external, is highly structured and involves the use of codeword brevity.

2.2 Synthetic training environment

In Black Skies, the facility provides a common briefing area as well as segregated spaces for the Blue, Red and White teams, representing the friendly forces, opposing forces, and exercise facilitators respectively. The real-world ABM work interface is re-created by stimulating actual situation awareness displays, as well as emulating the complex voice communications system. Performance evaluation is conducted by SME assessors, who are co-located with the trainees and provided with the same systems, as well as access to simulation truth data. Subjective data may be collected by observers at various locations, depending on resources and experimental design. The scenario is constructed by models of sensor and weapon systems, operated by role players controlling these entities via real-time user interfaces.[2]

2.3 Training exercise workflow

Black Skies exercises are typically conducted over several days, with two discrete serials per day. Within each serial, the sequence of activities mirrors the phases of live operations. As each phase is described, the time constraints on these exercises will become apparent.

2.3.1 Planning phase

The broad scenario is developed by the White team in advance of the exercise. This defines the geographical locations, assets, rules of employment, etc granted to the Blue and Red teams. At the start of the exercise each team is given time to adjust the initial posture of its forces, based on aspects such as the indicated threat, the level of warning and the duration prescribed.

2.3.2 Pre-brief phase

This phase consists of briefings by respective leaders to participants, firstly altogether and then to each team separately. These convey to trainees, role players, assessors and data collectors the design of the scenario, team plans and procedures, and the roles and responsibilities of each member. Where appropriate, assessment criteria are described.

2.3.3 Execution phase

During the two hour execution phase, a number of activities occur in parallel. The White team director facilitates the execution of planned scenario events, by managing role players and simulation systems. The trainees play out the evolving scenario through interactions with the other teams and role players.[3]

The assessor(s) observes the entire command team's interactions and mission performance against structured criteria, and notes good and bad exemplars of decision making and communication. Simulation data such as entity truth and system interactions is recorded, as are video captures of what each trainee sees, hears and says.

2.3.4 Debrief phase

Immediately after the completion of each serial[4], individual teams quickly discuss the course of events, and have a short break. In this period, usually 30 minutes, the assessors prepare their feedback and

[2] This minimises the number of role players required, by allowing a single operator to control >4 entities at once.

[3] Interaction is sometimes bottlenecked by individual players performing multiple roles or representing multiple entities.

[4] If the training schedule does not permit a review after each serial, a combined debrief occurs at the end of the day.

identify the timing of key events and the perspective from which to view them in the after action review.

The debrief session itself is approximately one hour long, during which the assessors summarise what happened, and present their view of the team's performance against the mission criteria. Selected replays may be used to support their ratings and for incident diagnosis. An audience-led discussion may occur if time permits.

2.3.5 Post-exercise phase

After the exercise is completed, researchers attempt to exploit the simulation data, where analysis resources permit. The utility of post hoc analysis is limited by the quality of the collected data. Therefore, the design of the data collection tools, procedures, and the metrics they support are critical to productive outcomes.

3. COMMUNICATION DATA

Before automated metrics can be designed, the information that describes a communication event must be defined. Table 2 describes the dimensions of communication as *who, what, when, where, how and why*, and relates these to attributes and examples from the ABM training domain.

Table 2: Dimensions of Communication Events

Dimension	Attribute(s)	Example
Who Tx( ) / Rx( ) [Addressee][5]	Role	Controller 1 / Role Player 1 (Aircraft 1,2,3,4) [Aircraft 3]
	Team	Blue Trainee / Blue Role Player
What	Subject	Targeting
	Significance	Mass Raid
When	Utterance Timing	Start 00:10 Stop 00:14
Where Tx{ } / Rx{ }	Physical {Virtual} Location	Console 1 / Console 5 {Waypoint A}
	Area of Responsibility	Airspace West / Airspace West
How[6]	Bearer	Secure Radio
	Mechanism	Half-duplex
Why[7]	Semantic Category	Action Requested

[5] Round brackets contain the details of the entities under control of a role player. Square brackets contain the name of the entity or participant addressed.

[6] Bearers include radio, intercoms and direct speech, and mechanisms include a combination of "hot" or "cold" calls, half or full duplex channels, and peer-to-peer or party lines.

[7] Examples are: Information Requested, Action Requested, Information Volunteered, Information Action, Response to Request and Task Irrelevant (Strang et al., 2012).

Most of the dimensions are straightforward, with some exceptions. Regarding *who*, it is necessary to discriminate between who received a transmission[8], and to whom it was addressed. Regarding *where*, both the physical location of a role player, and the virtual location of the entity which they are representing may be relevant.

3.1 Available Data Sources

Once the information is defined, it is possible to identify the necessary sources of this data. Table 3 lists the sources required by each dimension.

Table 3: Sources of Communication Data

Dimension	Data Source(s)
Who Tx / Rx [Addressee]	Exercise plan[9], Communications plan[10], DIS log, Activity log[11], [Transcript]
What	Transcript (with SME interpretation)
When	DIS log
Where Tx / Rx	Seating plan, Communications plan, DIS log, Activity log
How	Communications plan, DIS log
Why	Replay "coding" (by trained observers)

To date, the creation of transcripts or observer coded records has not been automated with sufficient accuracy and continues to be an area of active research (Cordero, Dorado & de Pablo, 2012). Consequently, the dimensions that require these sources, *what* and *why,* cannot currently be collected within the time constraints of a CTT exercise. However, most aspects of *who*, *when*, *where* and *how* are feasible to collect and analyse in a timely manner.

3.2 Validation of Automated Data Collection

Given that only a subset of the above data is feasible to capture, the question is, will this be sufficient to provide insight into participant behaviour? The effectiveness of automated collection can be validated by measuring if the subsequent analysis satisfies the objectives of stakeholders. Table 4 provides a list of questions that may be asked by stakeholders. These have arisen in previous Black Skies exercises, and may reflect the limitations of simple replay.

[8] A transmission may be received by many participants.

[9] The exercise plan defines the role of each participant.

[10] A communications plan defines the channels, e.g. frequency, bearer and mechanism, available to each role.

[11] The activity log contains a record of the entities and/or airspace under the control of each participant, over time.

Table 4: Communication Behaviour Questions

Stakeholder	Example Question(s)
Trainees	How did I employ the communications system? How did this compare to my peers?
Assessors	Were there behaviours that contributed to particular mission outcomes?
Facilitators	Is a role player currently over-loaded? Is a trainee under tasked?
Role Players	Were there situations whereby my actions or workload degraded training fidelity?
Researchers	Are there patterns of communication that could indicate workload, predict performance or help to quantify training impact?

4. DATA COLLECTION PROBLEMS

In a CTT exercise, operators communicate using a terminal that simulates the radios and intercoms of the real-world system, and emulates their behaviour and performance. For Black Skies, communication terminals are interconnected on a computer network using the DIS protocol. DIS messages, or Protocol Data Units (PDUs), model the communication medium in real-time, and can be recorded to a log file for later replay.

Caution must be exercised when using DIS logs for analysis, as certain issues may create an incomplete or inaccurate view of the communication that occurred (Schreiber, Watz & Bennet, 2003). Four specific problems will now be described in terms of the limitations of the DIS protocol and the nature of CTT. For each problem, current or proposed procedural and technical solutions will be discussed. It is important for any solution to minimise the deviation from the normal work procedures of the trainees, in order to maintain the integrity of the training environment.

4.1 Not-presented audio

Whilst the digital audio content may be received by a communications terminal, and reported as such on the DIS network, the audio may not actually be presented to the operator. Reasons for this occurring include that the operator may have muted or turned down the volume on a specific channel, or removed their headset and moved away from the terminal. This possibility casts doubt on *who* received a transmission.

The procedural solution to the latter requires operators to always wear a headset whilst on station, and to log-off from the communication terminal when moving away. A technical solution to the former, would involve publishing the volume level of each channel in a custom Data PDU. Several techniques have been identified for detecting when a headset is being worn by the operator, as technical solutions to the latter (Gerhardt, 2009).

4.2 Non-transmitted utterances

Non-transmitted utterances are instances where the operator speaks but does not transmit. Often this occurs when the operator speaks directly to a nearby person. In this situation, a faithful simulation would not generate DIS messages and *who* and *when* could not be captured.

The procedural solution is to instruct operators to use the communications system for all conversations. This approach can be difficult to enforce. A technical solution would require each terminal to be configured with a hidden, voice-operated keying (i.e. VOX) channel, which only transmits when no other channel is in use. This channel would not be received by other operators, but used specifically for communication analysis.

4.3 Operator / terminal movement

Team reorganisation and hardware malfunction (both unplanned and those intended by the facilitators) may require operators to change physical positions, and thus communication terminals, during the exercise. The DIS protocol, while providing a unique identifier for each channel endpoint, does not provide a facility to track which terminal an operator is using. This possibility casts doubt on *who* transmitted or received a communication, and from *where*.

The procedural solution is to maintain a mapping of each operator to a terminal throughout the exercise. This requires additional effort on the part of the facilitators to track participant movements. A modest technical solution would require operators to log-on and log-off terminals as they change positions (for example, using an ID card), and record these events as the happen. A more comprehensive technical solution would involve tracking operators passively, using radio frequency ID or camera tracking techniques.

4.4 One-is-to-many relationships

A simulated communications system may operate independently from the entity generation system. Whilst a simulated radio may be attached to an entity's virtual battlespace location, this approach is infeasible where a role player is responsible for multiple entities and controls each by rapidly switching between them. In the case of Black Skies, the virtual location of role players' terminals is fixed to the centre of the exercise area. This one-is-to-many problem creates uncertainty about the dimensions of *who* and *where*.

A procedural mitigation involves assigning entities to role players based on geographic area of responsibility. A technical mitigation involves logging which role players controlled which entities within the entity generation system. These approaches reduce, but do not remove, all the uncertainty created whenever many entities are under the control of a single role player.

5. COMMUNICATION METRICS

Communication metrics transform the collected data into measurements meaningful to exercise stakeholders. This information may take the form of an augmented

view of the current situation, a holistic view of scenario events, or a summary of performance or behaviour. A tool for developing and automating communication metrics will be described, and simple, direct examples discussed.

The research community has sought to understand the relationship between communication, workload and team cognition (Cooke et al., 2012). Development of sophisticated metrics, such as those that quantify team performance (Entin & Entin, 2001) and team dynamics (Gorman et al., 2012), may eventually have a much greater impact on training processes and outcomes, than simple representations. However, the potential metrics discussed here should be considered complementary to these complex, theory-based measures. For example, the former can provide a record of individual behaviour, whilst the latter may capture whole-of-team performance.

5.1 Analysis Software

Whilst there are several analysis tools for simulation data, DSTO currently uses BASE, an application developed by Boeing Defence Australia. BASE can monitor data streams in real-time or post-process log files. BASE is able to load any structured data in a Comma Separated Values (CSV) file and supports other sources via a plug-in interface; including one for the DIS protocol. BASE provides considerable flexibility to explore and manipulate extracted data, and chart or visualise this geospatially (Nixon, 2004).

There are many aspects to the construction of a metric within BASE, the feasibility of which may vary with the nature of the underlying data. For example, an analyst may wish to compare the behaviour of an individual operator, or each team as a whole. Similarly, an analyst may wish to classify events with arbitrary criteria, and assign these a value. Particular characteristics, such as the duration of an event, the frequency of occurrence, or the interval between events, may also be relevant to a metric. Furthermore, an analyst may wish to select between different concepts of time for a metric, such as the current value, or the cumulative value over a defined period. Simple statistical methods may be applied, and processed data may also be exported to external tools.

5.2 Terminal use patterns

Terminal use could be analysed in real-time, such as the time elapsed since each operator transmitted or received. This may indicate a breakdown in expected communication patterns. Another use pattern that could be measured is the number of channels monitored by an operator, weighted by volume level, over time. This would identify the operator with the greatest listening task and could indicate the degree to which responsibility is delegated amongst team members. Also, the importance of each communication medium in a given team could be measured as the relative number of transmit events over peer-to-peer vs. a party line or broadcast mechanisms. This may indicate how team situation awareness is built and maintained.

5.3 Communication patterns

Patterns of communication within a group and between groups could be measured. For example, who is the most/least frequent talker or who spoke the longest/shortest; does one group initiate conversation substantially more than another? How do these measures change over the progress of the scenario? The outcomes may correlate with the nature of the various operator roles or reflect task difficulty.

5.4 Conversation details

A discrete conversation between parties may be delineated by markers or thresholds, such as the termination of an intercom call, the transition from interaction with one party to another, or the elapsed time since the last interaction. Conversations could be analysed in terms of the size of the audience (number of receivers), the number of transactions (contiguous, alternating transmissions), response delay, and duration. A large audience may indicate an exceptional event. A high transaction count could indicate the need for error correction in the information exchanged. A long response delay could indicate that an operator is over-loaded. Finally, long conversations may indicate atypical subject matter.

5.5 Exceptional events

There are a number of exceptional events that may or may not be significant to the flow of communication in a given scenario. Nevertheless they may be useful cues for further investigation. Exceptional events may include:

- An operator transmits or receives on multiple channels at once.
- Multiple operators contend for a single channel.
- An operator is not monitoring any channel.
- A radio or “cold” intercom call is not answered.
- A transmission is extraordinarily short/long.
- An operator speaks but does not transmit.

6. APPLICATION TO EXERCISE WORKFLOW

The analysis of communication data provides the opportunity to visualise participant interactions and other patterns of behaviour, and to measure differences between trainees. Analysis could be conducted in real-time to inform exercise control and trainee assessment, or after the fact to enhance participant feedback and domain knowledge. Decisions regarding the type of visualisation, when it is presented, and the audience it is presented to, may affect the flow of ABM training in a number of ways. A selection of potential use cases will now be discussed in the context of the exercise phases described earlier.

6.1 Execution phase

The White team could monitor the terminal use and communication workload of participants. This could highlight where participants may have intentionally or

unintentionally selected the incorrect channel, or may be overwhelmed by concurrent activity on multiple channels.

Assessors could view a recent period of communication among trainees, such as the last 10 minutes, both in terms of cumulative effort and specific interactions. This could act as a form of delayed replay, as well as provide cues for the assessor to attend to the behaviour of specific trainees.

6.2 Debrief phase

Role players and trainees could be presented with a personal "dashboard" at the end of each serial. This may include a temporal view of *their* terminal use over the serial, and the ability to cue a replay of their work interface to a specific communication event. This would enable each participant to quickly recap events from their own perspective, prior to the group discussion.

Trainees could also receive a summary of their communication behaviour compared to specific peers, or a team average. This could help them to consider their contribution to the team and aspects of supporting behaviour.

Similarly, assessors could be provided with a visualisation of the distribution of communication activity across all participants. Further, a replay of the entity truth data could be provided, with the ability to cue this via a timeline of *all* communication events. Finally, assessors could select dynamic communication metrics for display next to, or overlaid on top of other visualisations or replays in the debrief session.

6.3 Post exercise phase

Improved data collection would provide researchers with a set of communication data with known attributes and limitations. This could be combined with other data sources to support novel statistical analysis. Validated measures of team performance could be used to compare different teams undergoing the same training.

7. RECOMMENDATIONS

Taking ABM training as a use case, a feasible approach to automated data collection and communication analysis has been identified. The following actions are required to test the effectiveness of this approach:

a. modify the simulated communications system to publish additional state data;

b. create machine-readable versions of paper-based data sources, for example, the communications plan;

c. collect sample data from a CTT synthetic environment; and

d. verify that the analysis software, BASE, can support the required data sources, metrics and visualisations.

Once a system is implemented, the research community should develop metrics based on what provides effective feedback. The training community should then apply these where they will have the greatest impact.

8. CONCLUSION

An approach to automated data collection and communication analysis has been presented. By focusing on the objectives of the exercise stakeholders, we have identified essential sources of data and collection methods, and speculated on the development of metrics and their impact to the exercise workflow.

This approach is feasible, both from a technical and human resource point of view, and implementation only requires a modest investment. It capitalises on the effort already expended during CTT exercises to capture data for after action review, and does not require additional resources during the execution phase.

Many of the proposed metrics should generalise to other work domains where voice communication plays a fundamental role. By moving beyond simple replays, stakeholders in CTT exercises such as Black Skies, stand to gain timely and evidence-based insight into the performance of trainees and the training environment.

REFERENCES

Best, C. et al. (2009). *Virtual Mission Preparation: Making the Most out of Live Warfighting Exercises*. SimTecT 2009.

Cooke, N.J. et al. (2012). *Interactive team cognition*. Cognitive science, 13, pp255-285.

Cordero, J.M., Dorado, M., & de Pablo, J.M. (2012). *Automated speech recognition in ATC environment.* In Proceedings of the 2nd International Conference on Application and Theory of Automation in Command and Control Systems, pp46-53.

Entin, E.E. & Entin, E.B. (2001). *Measures for evaluation of team processes and performance in experiments and exercises*. In Proceedings of the 6th International Command and Control Research and Technology Symposium, pp19-21.

Gerhardt, J.F. (2009). *Donned and doffed headset state detection*. U.S. Patent No. 8,335,312.

Gorman, J.C. et al. (2012). *Measuring patterns in team interaction sequences using a discrete recurrence approach.* Human Factors, 54(4), pp503-517.

Nixon, A. (2004), *A Flexible Software Architecture For Visualising Simulation Data*. SimTect 2004.

Schreiber, B.T., Watz, E.A. & Bennett, W. Jr. (2003). *Objective human performance measurement in a distributed environment: Tomorrow's needs*. I/ITSEC 2003.

Salas, E. et al. (2009). *Performance measurement in simulation-based training: A review and best practices*. Simulation and Gaming, 40(3), pp328-376.

Shanahan, C. et al. (2009). E*xercise Black Skies 2008: Enhancing Live Training Through Virtual Preparation - Part One: An Evaluation of Training Effectiveness*. Defence Science & Technology Organisation, DSTO-RR-0344.

Strang, A.J. et al. (2012). *Examining Temporal Regularity in Categorical Team Communication Using Sample Entropy*. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 56(1), pp473-477.

Tracey, E. et al. (2009). *Exercise Black Skies 2008: Enhancing Live Training Through Virtual Preparation - Part Two: An Evaluation of Tools and Techniques*. Defence Science & Technology Organisation, DSTO-TR-2305.