

SANDIA REPORT

SAND2012-10193

Unlimited Release

Printed November 2012

A Data Driven Approach To Assess Team Performance Through Team Communication

J. Chris Forsythe, Ph.D. (jcforsy@sandia.gov)
Matthew R. Glickman, Ph.D. (mrglick@sandia.gov)
Michael J. Haass (mjhaass@sandia.gov)
Jonathan H. Whetzel (jhwhetz@sandia.gov)

Project funded through the Office of Naval Research under contract N0001412IP20025

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Approved for public release; further dissemination unlimited.



Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

NOTICE: This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government, nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, make any warranty, express or implied, or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represent that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government, any agency thereof, or any of their contractors or subcontractors. The views and opinions expressed herein do not necessarily state or reflect those of the United States Government, any agency thereof, or any of their contractors.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: <http://www.osti.gov/bridge>

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd.
Springfield, VA 22161

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.fedworld.gov
Online order: <http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online>



A Data Driven Approach To Assess Team Performance Through Team Communication

Chris Forsythe, Ph.D.¹, Matthew R. Glickman, Ph.D.¹, Michael J. Haass², Jonathan H. Whetzel³
¹Cognitive Modeling (1462),
²Cognitive Systems (1463), ³Interactive Simulation & Analysis (6134)

Sandia National Laboratories
PO Box 5800
Albuquerque, NM 87185-1004

ABSTRACT

For teams working in complex task environments, instilling effective communication between team members is a primary goal during task training. Presently, responsibility for evaluating team communication abilities resides with instructors and outside observers who make qualitative assessments that are shared with the team following a training exercise. Constructing technologies to automate these assessments has historically been prohibitive for two reasons. First, the financial cost of instrumenting the environment to collect team communication data at the necessary fidelity has been too high for an operational setting. Second, past research on using team communication as a proxy for team performance assessment has relied on defining communication through traditional algorithmic design, an approach which does not properly capture the varied nature of communication strategies amongst different teams.

Recent scientific research in team dynamics provides a theoretical framework leading to a data-driven solution for analyzing the effectiveness of team communication. By framing team communication as an emergent data stream from a complex system, one may employ machine learning or other statistical-analysis tools to highlight communication patterns and variance, both shown as effective means for assessing team adaptability to novel scenarios. Furthermore, low-cost wearable computers (e.g., smartphones) have opened new possibilities for observing people's interactions in natural settings to better analyze and improve team performance.

This report summarizes research conducted by Sandia National Laboratories in developing a data-driven approach to analyzing team communications within the context of Surfaced Piloting and Navigation (SPAN) training for submariners. Using Dynamic Bayesian Networks (DBN's), this approach created predictive models of communication patterns that emerge from the team in different contexts. Based upon data collection conducted in the lab and within live submarine crew training, our results demonstrate the robust nature of DBN's by still identifying key communication events even when teams altered their speaking patterns during these events to accommodate for novel changes in the scenario.

ACKNOWLEDGEMENTS

The authors of this report would like to thank: Eric Jones and Jonathan Lansey of Aptima, Inc., David Kern, Ph.D. of Kern Technology Group, Inc, and Robert Ahlers, Ph.D. of Naval Aviation Warfare Center Training Systems Division (NAWCTSD).

CONTENTS

ABSTRACT	3
ACKNOWLEDGEMENTS	4
CONTENTS	5
INTRODUCTION	7
MODELING TEAM COMMUNICATION	9
DYNAMIC BAYESIAN NETWORKS.....	9
EXPERIMENTS	12
LABORATORY EXPERIMENT.....	12
<i>Discrimination Between Good and Bad Rounds with DBN's</i>	12
SECOND EXPERIMENT.....	14
<i>Recognizing Cyclic Routines with Real Training Data</i>	15
DISCUSSION	17
REFERENCES	18
DISTRIBUTION LIST	20

THIS PAGE HAS BEEN INTENTIONALLY LEFT BLANK

INTRODUCTION

Complex tasks that demand a coordinated effort benefit from the capacity of a team to pool resources via an exchange of information and coordinated action, though the effectiveness of a team may be contingent on a variety of factors [1]. Team effectiveness has particular impact within a military setting, as within combat situations the performance of a group has a direct bearing on the survival of the group and those dependent on them [2], situation that holds true when considering the success of naval operations [3]. In an attempt to determine the critical elements that make up an effective team in a military setting, variables related to team effectiveness have been examined from a variety of perspectives, including team cohesiveness (i.e., shared interpersonal closeness and group goal-orientation) [4], [5] collective orientation [1], shared mental models (i.e., synthesis of input from individual team members) [6], [7], [8], team selection and composition (e.g., the skills possessed by the individual team members, how long the members have been working together) [5], [6], [9], quality of decisions made by commanders [10], [11], cognitive readiness and adaptive decision making at the group level [12], training adequacy [5], the workload involved [13], and even neurophysiologic synchrony between team members, as assessed via electroencephalogram [14].

In the context of naval operations, assessment of the quality of teamwork has proven difficult, with such assessments relying on the observations of subject matter experts, skilled instructors, or a self-evaluation within teams during live or simulated exercises [3]. These judgments are subjective by their very nature, leading to a potential lack of consistency with regard to the quality of assessment. This issue has been recognized, and there have been attempts to resolve it, such as through outcome-based assessments that use goal-attainment as an objective measure of team effectiveness, with goal-attainment defined using Hierarchical Task Analysis for teams [3]. Historically, methods such as this that attempt to create a more quantifiable way of assessing team effectiveness have proven prohibitive such that while they achieve some success in ameliorating the issue of subjectivity they are time consuming and costly enough to make wide implementation infeasible.

Teamwork has been defined as "the interdependent components of performance required to effectively coordinate the performance of multiple individuals" [15], with the authors going on to note the critical role of communication in team performance. It is precisely this aspect of team effectiveness—communication—that the current work focuses on. Previous research in this domain has shown that the ability of a team to adapt to situational demands is reflected by the variance in their communication patterns [13],[16], a finding that the current work exploits in an attempt to yield an automated, quantitative measure of team communication, which would allow instructors and teams to assess changes in communication patterns in correlation with scenario events.

Sandia National Laboratories has made several investments within the Automated Expert Modeling for Automated Student Evaluation (AEMASE) technology [17] which implements statistical machine-learning techniques for identifying behaviors of interest within spatio-temporal data streams of individuals/teams operating within a live or simulated environment. Instructors using debrief tools with AEMASE integration may generate behavior models through

a programming-by-example approach [18] by flagging positive and negative examples of desired behavior. These models can observe the behaviors of other individuals/teams and provide a measure of similarity that could serve as an assessment metric.

Through funding from the Office of Naval Research, Sandia National Laboratories conducted a study on utilizing AEMASE for the US Navy Submarine fleet within their Surfaced Piloting and Navigation (SPAN) trainers. Through the use of Dynamic Bayesian Networks as the underlying machine-learning approach, this technology shows promise for identifying vocal communication patterns and providing valuable feedback for instructors and teams.

MODELING TEAM COMMUNICATION

In devising machine-learning algorithms for representing models of effective team communication, one must consider the multiple modalities of data available for analysis. Potential kinds of data that might be used by such a system include: trainee verbal communication, physical actions of the trainees (e.g. movement or control actuation), static factors such as team history or features of the specific training scenario being conducted, along with data available from the training scenario and actions taken in response by the team. As there were time-consuming engineering or social hurdles associated with this research, we ultimately settled on recognizing patterns in trainee verbal communications. To further streamline our approach, we chose not to rely on automated speech recognition technology. Thus, the data stream we chose to analyze indicated who was speaking at any given time during a group training exercise, and the challenge was to use this data to recognize domain-relevant activity patterns.

Dynamic Bayesian Networks

Bayesian Networks [19] are graphical models that represent conditional dependencies between random variables. For example, the simple Bayesian network in the example below indicates that the availability of downtown parking is conditionally dependent upon both (a) whether or not the time is prior to 8AM and (b) whether or not the day is a weekend. The same network further implies that whether or not it is before 8 AM is independent of whether or not it is a weekend.

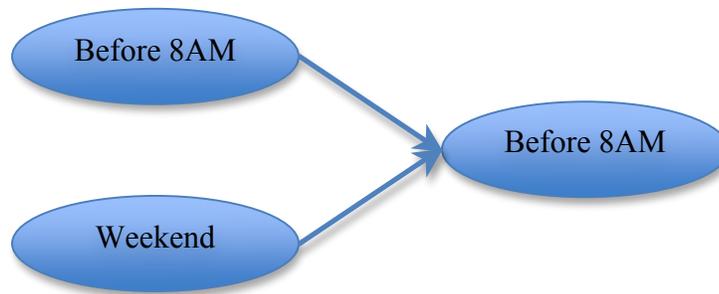


Figure 1: Simple example of a Bayesian Network

Associated with each node is a Conditional Probability Table (CPT). An example CPT for downtown parking availability is shown in Table 2. Note that knowing whether it is before or after 8AM and whether or not it is a weekend isn't sufficient information to completely determine whether or not it's possible to find a parking space downtown. Rather, as shown in the CPT, the probability of finding a parking space is affected by these two conditions.

Before 8AM	Weekend	T	F
F	F	0.05	0.95
F	T	0.5	0.5
T	F	0.4	0.6
T	T	0.9	0.1

Table 1: A Conditional Probability Table for Downtown Parking Available

Dynamic Bayesian Networks [20] are a particular variety of Bayesian networks that represent conditional dependencies over time. For example, the DBN shown in Figure 2 indicates that the presence of an infection on a given day is conditionally dependent upon the presence of an infection on the previous day and whether or not the potentially infected individual took antibiotics on the previous day (a full specification of this DBN would require CPT's for each node).

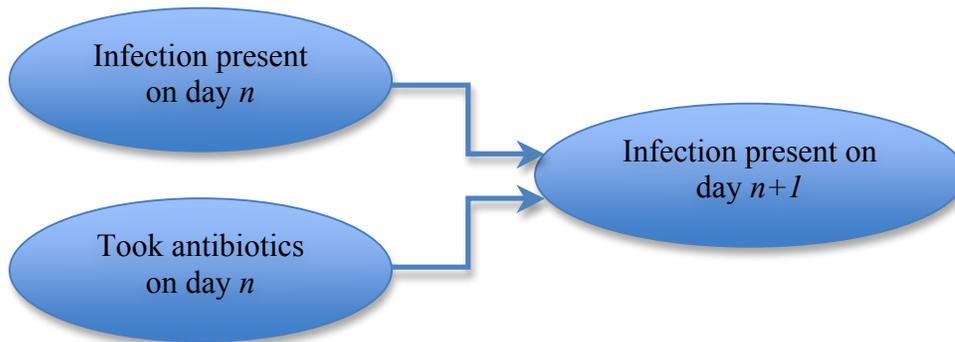


Figure 2: Example of a DBN

Bayesian Networks are models, i.e. abstract, simplified descriptions of aspects of the world. The task of devising a DBN may be broken into two subtasks: (1) devising the structure of the model, and (2) populating the associated CPT's. One valid way to devise the model structure is for a human to employ knowledge of the target domain to create a suitable structure by hand. Alternatively, the structure can be created in an automated fashion, typically via a search process that evaluates possible structures to find the best one (as defined with respect to some specified criteria). Once a structure is identified, CPT's can be populated straightforwardly, simply by using the observed frequencies of joint events. For instance, in the CPT shown in Table 2 the probabilities of 0.9 and 0.1 indicated in the row corresponding to before 8AM on weekends may be derived from data in which parking spaces were found to be available downtown 9 out of 10 times among observations that were made before 8AM on weekends.

To understand how DBNs may be used for classification, consider the case of the DBN in the example above. Imagine we built two versions of the model: One with data from cases with bacterial infections and the other with data from cases with viral infections. We would then expect that probabilities in the two models should be different: For viral infections, the

probability of an infection being present on day $n+1$ would be essentially unaffected by whether or not the infected individual took antibiotics on day n . These two DBN models would thus reflect two different underlying situations. Then, when presented with data for a new case, we could calculate the *likelihood* of each DBN generating the data in question. If the new data reflects a case where the presence of the infection appears conditionally dependent upon taking antibiotics the day before, the likelihood scores should indicate that DBN built from bacterial infection is a better model of the data. Similarly, if the course of infection appears independent of antibiotic use, the viral infection model should appear more likely.

DBN's hold a distinct advantage in their ease of interpretation as opposed to other modeling algorithms (e.g. distance-based clustering approaches). This is partly due to their graphical representation, which succinctly represents conditional dependence. Ease of interpretation is further enhanced by the straightforward probability calculations underlying DBN application. The relative transparency of DBN's may offer a particular advantage for this challenging domain. Human judgment of team dynamics in general and team adaptability in particular is largely intuitive and hard to express in objective form. Data-driven, automated systems may be of help, but the best performance is likely to emerge from leveraging human knowledge and insight in addition to data-driven analysis. One possibility along these lines is that human inspection of DBN's may lead to critical insights about what is or is not being captured at the data level, which may then lead to favorable adjustments in how DBN's are trained and/or used in the application at hand. More directly, humans may potentially recognize and eliminate spurious dependencies that arise in DBN structure search, or similarly, insert dependencies known to be significant.

EXPERIMENTS

Over the course of the study we conducted two separate experiments. We first conducted a laboratory-based study using a simplified SPAN training environment for testing the feasibility of Dynamic Bayesian Networks for accurately classifying speech patterns based upon human annotated labels. Following our laboratory study, we collected live training data from the Naval Submarine School in Groton, CT to classify known communication patterns that transpired during the exercise..

Laboratory Experiment

For our laboratory experiment, we built a facsimile representation of SPAN trainer using SubSkillsNet, a submarine simulator created by the US Navy for use on PCs within a classroom setting. For the study we generated scenarios that required a three-person team: a radar operator, periscope operator, and helmsman who also served as a bearing recorder. The scenario had the submarine following a fixed course, with the team making periodic cyclic routine calls for waypoints along the path. The cyclic routine, known colloquially as “rounds of contacts”, involves having the team follow an orchestrated pattern where sensor operators (e.g., radar, periscope) provide information on specific contacts requested by the bearing recorder. This test was conducted with two separate teams who had no prior training on the simulator or performing the cyclic routine task. During the test, each team member wore a lapel microphone to record his or her utterances for analysis of team communication.

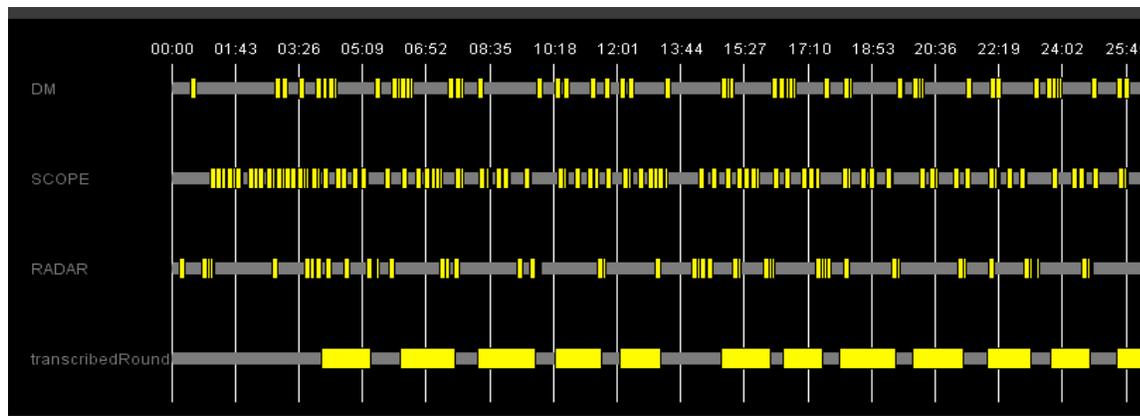


Figure 3: Timeline of utterances captured during laboratory experiments

Discrimination Between Good and Bad Rounds with DBN's

The above figure shows a graphical view of a segment of vocalization data from the in-lab sessions. In the top three rows, each yellow bar indicates an interval during which one of the three participants was determined to be speaking, with the helmsman labeled as “DM”, the periscope operators as “SCOPE”, and the radar operator as “RADAR”. In the bottom row, each yellow bar indicates an interval during which the participants were engaged in a cyclic routine. There were 26 such intervals identified over all recorded sessions. A single human evaluator

designated each of these intervals as either a *good* or *bad* instance according to cyclic routine doctrine.

Because DBN’s work on discrete random variables and represent time as progressing in discrete steps of fixed duration, it was thus necessary to sample the audio traces at a fixed rate. In all cases, we settled on a minimum time-step of either 0.5 or 1.0 secs, chosen to permit properly representing the shortest meaningful utterance (e.g. along the lines of “yes” or “OK”). To further discretize the data, all audio samples were passed through a threshold function to yield a binary value indicating *speaking* or *not-speaking* at each time-step..

Using this data encoding, our initial experiments in classifying *good* vs. *bad* rounds were not successful. Although it was important to conduct these experiments to establish a baseline, poor performance was not unexpected due to the fact that the initial encoding didn’t provide DBN’s with sufficient temporal context to support pattern recognition. With the selected time-step, models could only capture probabilities relative to what happened in the last second or less.

Our solution was to extend the simple binary encoding from *speaking* vs. *not-speaking* to include a third value indicating when a speaker was not currently speaking, but had been speaking recently. Via an examination of the distribution of lengths between speech events in the data, we made a heuristic choice to define “recently” as less than or equal to 7 seconds.

Time Step (sec)	Helmsman	Radar Operator	Periscope Operator
1	<i>Speaking</i>	<i>Spoke > 7 secs ago</i>	<i>Spoke > 7 secs ago</i>
2	<i>Spoke ≤ 7 secs ago</i>	<i>Spoke > 7 secs ago</i>	<i>Spoke > 7 secs ago</i>
3	<i>Spoke ≤ 7 secs ago</i>	<i>Spoke > 7 secs ago</i>	<i>Spoke > 7 secs ago</i>
4	<i>Spoke ≤ 7 secs ago</i>	<i>Speaking</i>	<i>Spoke > 7 secs ago</i>
5	<i>Spoke ≤ 7 secs ago</i>	<i>Speaking</i>	<i>Spoke > 7 secs ago</i>
6	<i>Spoke ≤ 7 secs ago</i>	<i>Speaking</i>	<i>Spoke > 7 secs ago</i>
7	<i>Spoke ≤ 7 secs ago</i>	<i>Spoke ≤ 7 secs</i>	<i>Spoke > 7 secs ago</i>
8	<i>Spoke ≤ 7 secs ago</i>	<i>Spoke ≤ 7 secs</i>	<i>Spoke > 7 secs ago</i>
9	<i>Spoke > 7 secs ago</i>	<i>Spoke ≤ 7 secs</i>	<i>Spoke > 7 secs ago</i>
10	<i>Spoke > 7 secs ago</i>	<i>Spoke ≤ 7 secs</i>	<i>Spoke > 7 secs ago</i>
11	<i>Spoke > 7 secs ago</i>	<i>Spoke ≤ 7 secs</i>	<i>Spoke > 7 secs ago</i>
12	<i>Spoke > 7 secs ago</i>	<i>Spoke ≤ 7 secs</i>	<i>Spoke > 7 secs ago</i>
13	<i>Spoke > 7 secs ago</i>	<i>Spoke ≤ 7 secs</i>	<i>Spoke > 7 secs ago</i>
14	<i>Spoke > 7 secs ago</i>	<i>Spoke > 7 secs ago</i>	<i>Spoke > 7 secs ago</i>

Table 2: Encoding DBN features of when team members spoke during exercise

The above table shows a 14 second excerpt of an encoded round which begins with the helmsman speaking for approximately one second, followed 2 seconds later by the radar operator speaking for 3 seconds.

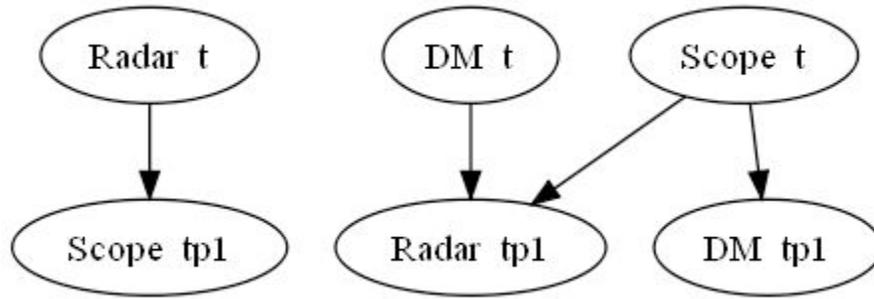


Figure 4: DBN prediction graph of which team member will speak in timestep $t + 1$

We used the SBNet¹ software tool for DBN training. The above figure shows an example of a trained DBN multinet (a superposition of three DBN’s each trained to predict one of the three speakers). For the purpose of classification, we opted to conduct DBN structure search to optimize *Approximate Class-conditional Likelihood* (ACL, Burge & Lane 2005). Use of ACL favors dependencies in the data that emphasize the differences between two sets of data over regularities with each set, and thus increases discrimination power.

To make maximal use of our data, we employed *leave-one-out* validation. That is, for each of the 26 *good* and *bad* rounds, we applied SBNet to build two DBNs—one model of vocalizations over time in *good* rounds and a parallel model for *bad* rounds—based upon the *other 25* runs. We then derived log-likelihood ratio scores for each observed round by calculating the ratio between the likelihood of these two corresponding DBN’s with respect to the round in question. The distribution of the results for all 26 rounds is shown in the figure below.

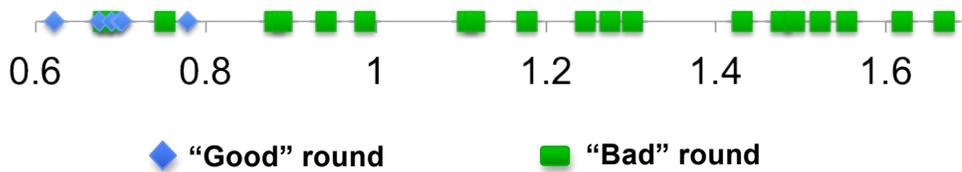


Figure 5: DBN classification scores for laboratory experiment

As a well-chosen classification threshold would result in 88.4% of these rounds being properly classified, we considered these results encouraging of this approach.

Second Experiment

The data collection event at Navy Submarine School took place in early February 2012, observing over four hours from two teams within the SPAN trainer. For the data collection event experimenters used the Sociometric Badges [21] developed by MIT Media Lab. A wearable

¹ SBNet was created by Dr. John Burge. SBNet software currently available for download at <http://www.cs.unm.edu/~lawnguy/sbnet/index.html>

computing device, the Sociometric Badge collects data through an onboard sensor array on the interactions of users within a group setting. The badges periodically sample the fundamental frequency of speech utterances by a wearer, along with using Bluetooth pings and infrared sensors to measure relative spatial proximity between all people wearing these badges. The audio sampling proved advantageous for this environment given the security concerns in collecting communication data within a classified training facility. However, the lack of a full audio recording meant we could not rely on analyzing speech content for generating our DBNs communication pattern models.



Figure 6: A diagram of a Sociometric Badge (left), a person wearing a Sociometric Badge (center), downloading data from the Sociometric Badge (right)

Prior to the data collection we provided all members of the piloting party with a Sociometric Badge, along with instrumenting stations with badges to capture spatial proximity information of crewmembers within the trainer. During the exercise, subject matter experts would record observations of interest to compare against the measured communications. These observations included: start/stop times of cyclic routines, events injected by the instructor (e.g., instrument failures), and key course corrections made by the crew. As well, cross-track error was periodically sampled at five-minute intervals.

Recognizing Cyclic Routines with Real Training Data

Our initial plan was to apply the same DBN modeling procedure to the data acquired from the SPAN trainer to generate models of “good” and “bad” cyclic routines. This metric seemed useful given that US Navy Submarine Doctrine defines the cyclic routine to be conducted by crewmembers, yet no automated assessment presently exists to monitor and report a crew’s ability to follow this procedure. However, the data acquired from the SPAN trainer does not provide enough fidelity information to replay and assess the quality of each cyclic routine (e.g., no audio recording, no information on data available from different sensors). In absence of being able to subjectively evaluate each observed cyclic routine instance, we attempted to apply the DBN modeling algorithms for detecting when the crew was simply engaged or not engaged in a cyclic routine based upon the communication data.

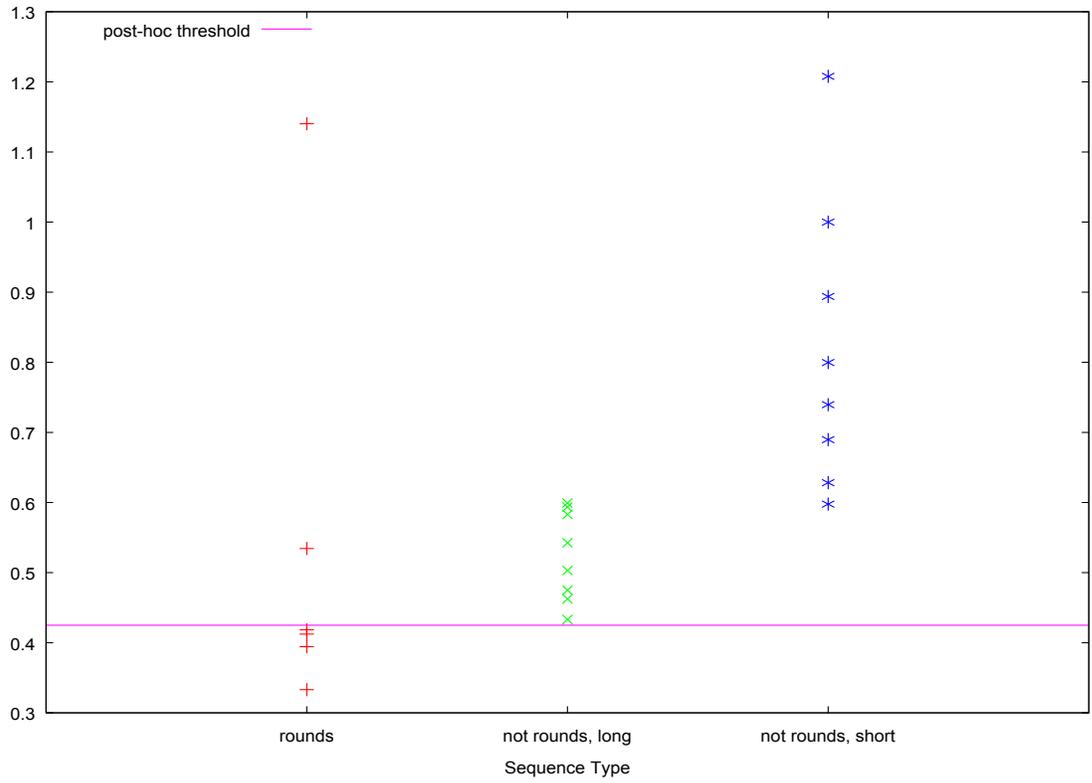


Figure 7: DBN classification scores for SPAN trainer experiment

In Figure 7 we observe that the DBN can provide a high classification accuracy discerning when the team is engaged in a cyclic routine. Again creating a post-hoc threshold for the proportion value, we observe the DBN accurately classifying with 85.7% accuracy between defined cyclic routines and all other audio segments. We achieve a higher accuracy of 92.9% if the “not round” segments are equally partitioned into the average time elapsed between cyclic routines.

DISCUSSION

Though the DBN modeling proves effective at identifying when crews are engaged in this doctrinal behavior, these results provide little feedback that could yield diagnostic measures on team performance. To better tailor debrief for crewmembers, we propose to explore alternative methods besides machine learning models to characterize the observed team communication data. Psychologists in the area of team performance have begun analyzing team communication as a dynamical system, where changes in communication can be monitored in real-time concurrently with the performance of the team.

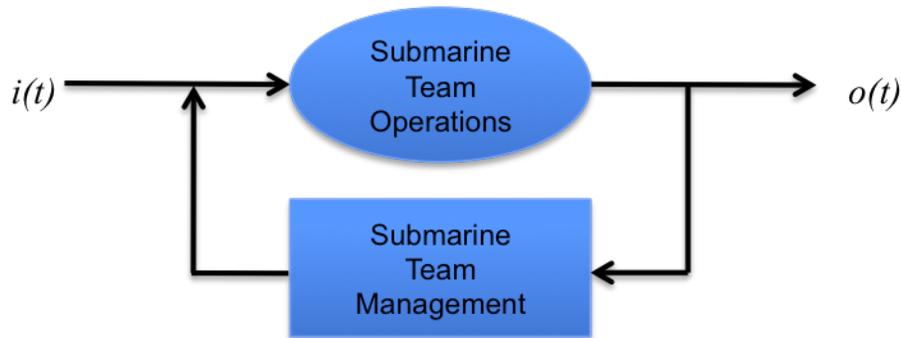


Figure 8: System Dynamics Representation of Submarine Piloting Party Operations

Figure 8 provides a dynamical system representation of a submarine piloting party. In this representation, $i(t)$ signifies the input the team can receive at any given moment from their sensor suite (e.g., periscope, radar, GPS, fathometer, visual from deck, etc). The output of the system, $o(t)$, represents where the submarine will navigate to reach its targeted destination. As the team receives input, the piloting party must conduct operations (cyclic routine) to process this information and provide feedback to adjust operations (communication between routines) to maintain or better clarify their situational awareness. At this moment, we have not sufficiently explored a dynamical system analysis to confirm or violate our expectations.

Our present research does show promise in devising a capability for discerning expected communication within multiple streams. From this information, one may generate visualizations that allow instructors or teams to pinpoint key points when critical communication within the team did or did not take place. Quantifying these communications would bring us substantially closer to a repeatable, measured approach for team communication assessment. Future research possibilities include devising these assessment systems in conjunction with an instructor or team lead to determine the most effective measures one would want to generate. As well, incorporating more simulation information would even further clarify the appropriateness of certain communication patterns observed given the context of the situation.

REFERENCES

- [1] J. E. Driskell, E. Salas, and S. Hughes, "Collective Orientation and Team Performance: Development of an Individual Differences Measure," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 52, no. 2, pp. 316–328, Sep. 2010.
- [2] A. Tziner and Y. Vardi, "Ability as a moderator between cohesiveness and tank crews performance.," *Journal of Occupational Behaviour*, vol. 4, no. 2, pp. 137–143, 1983.
- [3] J. Annett, D. Cunningham, and P. Mathias-Jones, "A method for measuring team skills," *Ergonomics*, vol. 43, no. 8, pp. 1076–1094, Jan. 2012.
- [4] A. Ahronson and J. E. Cameron, "The nature and consequences of group cohesion in a military sample," *Military Psychology*, vol. 19, no. 1, pp. 9–25, 2007.
- [5] L. L. Schmidt, K. Keeton, K. J. Slack, L. B. Leveton, and C. Shea, "Risk of performance errors due to poor team cohesion and performance, inadequate selection/team composition, inadequate training, and poor psychosocial adaptation," *Human Health and Performance Risks of Space Exploration Missions (NASA SP-2009-3405)*, pp. 53–84, 2009.
- [6] R. Espevik, B. H. Johnsen, J. Eid, and J. F. Thayer, "Shared mental models and operational effectiveness: Effects on performance and team processes in submarine attack teams," *Military Psychology*, vol. 18, no. 1, pp. 23–36, 2006.
- [7] M. E. Jobidon, A. Muller-Gass, M. Ducan, and A. R. Blais, "Enhancing mental models for team effectiveness," DTIC Document, 2011.
- [8] K. A. Smith-Jentsch, J. A. Cannon-Bowers, S. I. Tannenbaum, and E. Salas, "Guided Team Self-Correction Impacts on Team Mental Models, Processes, and Effectiveness," *Small Group Research*, vol. 39, no. 3, pp. 303–327, 2008.
- [9] J. C. Gorman and N. J. Cooke, "Changes in team cognition after a retention interval: The benefits of mixing it up.," *Journal of Experimental Psychology: Applied; Journal of Experimental Psychology: Applied*, vol. 17, no. 4, p. 303, 2011.
- [10] E. Jones, R. Steed, F. Diedrich, R. Armbruster, and C. Jackson, "Performance-based metrics for evaluating submarine command team decision-making," *Foundations of Augmented Cognition. Directing the Future of Adaptive Systems*, pp. 308–317, 2011.
- [11] F. J. Yammarino, M. D. Mumford, M. S. Connelly, and S. D. Dionne, "Leadership and team dynamics for dangerous military contexts," *Military Psychology*, vol. 22, no. 1, pp. 15–41, 2010.
- [12] S. M. Fiore, K. G. Ross, and F. Jentsch, "A Team Cognitive Readiness Framework for Small-Unit Training," *Journal of Cognitive Engineering and Decision Making*, vol. 6, no. 3, pp. 325–349, 2012.
- [13] J. M. Urban, C. A. Bowers, S. D. Monday, and B. B. Morgan Jr, "Workload, team structure, and communication in team performance," *Military Psychology*, vol. 7, no. 2, pp. 123–139, 1995.
- [14] R. Stevens, T. Galloway, P. Wang, C. Berka, and A. Behneman, "Developing systems for the rapid modeling of team neurodynamics," presented at the The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC), 2011, vol. 2011, no. 1.
- [15] E. Salas, N. J. Cooke, and M. A. Rosen, "On teams, teamwork, and team performance: Discoveries and developments," *Human Factors: The Journal of the Human Factors*

- and Ergonomics Society*, vol. 50, no. 3, pp. 540–547, 2008.
- [16] J. Gorman, “Team coordination dynamics and the interactive approach: emerging evidence and future work,” *Foundations of Augmented Cognition. Directing the Future of Adaptive Systems*, pp. 298–307, 2011.
- [17] R. Abbott, “Automated expert modeling for automated student evaluation,” presented at the Intelligent Tutoring Systems, 2006, pp. 1–10.
- [18] H. Lieberman, *Your wish is my command: Programming by example*. Morgan Kaufmann, 2001.
- [19] J. Pearl, “Fusion, propagation, and structuring in belief networks,” *Artificial intelligence*, vol. 29, no. 3, pp. 241–288, 1986.
- [20] K. P. Murphy, “Dynamic bayesian networks: representation, inference and learning,” 2002.
- [21] D. O. Olgun and A. S. Pentland, “Sociometric badges: State of the art and future applications,” 2007.

DISTRIBUTION LIST

1	MS1327	J. Chris Forsythe
1	MS1327	Matthew R. Glickman
1	MS1327	Michael J. Haass
1	MS1004	Jonathan H. Whetzel
1	MS0899	Technical Library, 9536 (electronic copy)

THIS PAGE HAS BEEN INTENTIONALLY LEFT BLANK



Sandia National Laboratories