Developing Systems for the Rapid Modeling of Team Neurodynamics

Ron Stevens, Trysha Galloway, Peter Wang
IMMEX Project / UCLA
Los Angeles, CA.
Ron, Trysha, Peter@teamneurodynamics.com

Chris Berka, Adrienne Behneman
Advanced Brain Monitoring, Inc.
Carlsbad, CA.
Chris, Abehnemans@b-alert.com

INTRODUCTION

One of the challenges for studying team cognition in real-time is the development of unobtrusive and relevant measures of team performance that can be practically implemented and rapidly modeled in real-world environments (Salas et al, 2008).

We have been studying whether the simultaneous expression of EEG-derived cognitive measures by different members of a team could be used to complement verbal communication streams for constructing such teamwork models. In this approach the values of a cognitive measure at any point in time are aggregated across the team members into a vector that is then clustered / classified by artificial neural network (ANN) technologies (Stevens et al, 2009; Stevens et al, 2010a). The result is a series of symbolic patterns termed Neurophysiologic Synchronies (NS) defined as the second-by-second quantitative co-expression of the same neurophysiologic measure by different members of the team. The cognitive measures we have modeled include High Engagement and High Workload which have been derived from EEG data streams (Berka et al, 2005). If NS are meaningful constructs then their expression should:

1. Be able to be collected and analyzed in real-world situations;
2. Be sensitive to long and short-term task changes;
3. Relate to some established aspects of team cognition, yet reveal something new;
4. Be extensible to future teams;
5. Distinguish novice / expert performance; and,
6. Be sensitive to the effects of training.

Our initial studies used a single-trial approach for developing NS models, i.e. the data from a single performance was used for deriving the ANN and HMM models for that performance. These studies were informative and generated validation data for criteria 1-3 described above. As new models were created for each task and team it was difficult to compare across teams or levels of experience as the ANN designations changed due to the probabilistic assignment of vectors to specifically numbered nodes and states. Also, without standardized models it was difficult to begin to extend this analysis to real-time team modeling.

One way of developing standardized models would be to combine the performances from multiple teams with differing experience creating standardized (or generic) models. It is not intuitive whether this approach would be successful. Standardized datasets due to their larger size may not be sensitive to some combinations of NS across members of some teams due to their unique expression by that team. Conversely, separate single-trial models may not have the repertoire of EEG-E combinations to allow meaningful comparisons across teams. There is also a validation challenge when developing standardized models: what will be the comparison standard? Neurophysiologic synchrony research is in its early stages and a precise niche where they fit into the theories of teamwork is not yet clear. While they are linked with some aspects of speech such as structure, they are not closely enough linked to use speech for model validation.

In this study we have generated standardized models for both three and six-person teams as well as for EEG-derived measures of engagement (EEG-E) and workload (EEG-WL). To make direct comparisons across models we have measured the temporal dynamics of the entropy or 'amount of mix' in these symbolic data streams. This has allowed us to directly make quantitative comparisons between teams with different levels of experience.

TASKS AND METHODS

Tasks

Submarine Piloting and Navigation (SPAN) simulations are high fidelity tasks that contain dynamically programmed situation events crafted to serve as the foundation of adaptive team training. Such events in the SPAN include encounters with approaching ship traffic, the need to avoid nearby shoals, changing weather conditions, and instrument failure. There are task-oriented cues to guide the mission, team-member cues that provide information on how other members of the team are performing / communicating, and adaptive behaviors that help the team adjust in cases where one or more members are under stress or are not familiar with aspects of the unfolding situation.

Each SPAN session begins with a briefing detailing the navigation mission. This is followed by the simulation which can last from 60 – 120 minutes. This is followed by a debriefing session that helps teams monitor and regulate their own performance based on the dimensions of teamwork. This teamwork task requires not only the monitoring of the unfolding situation and the monitoring of one's work with regard to that situation, but also the monitoring of the work of others. Twenty-one SPAN sessions were conducted where EEG was collected from three to six persons. The data reported here was derived from twelve of those sessions selected as: 1) persons in the same six crew positions were being monitored by EEG, 2) the same individuals repeated in the same positions across 2-5 training sessions over multiple days. The six members of the teams that were fitted with the EEG headsets were the Quartermaster on Watch (QMOW), Navigator (NAV), Officer on Deck (OOD), Assistant Navigator (ANAV), Contact Coordinator (CC), and Radar (RAD).

Methods

EEG

The ABM, B-Alert® system contains an easily-applied wireless EEG system that includes intelligent software designed to identify and eliminate multiple sources of biological and environmental contamination and allow real-time classification of cognitive state changes even in challenging environments. The 9-channel wireless headset includes sensor site locations: F3, F4, C3, C4, P3, P4, Fz, Cz, POz in a monopolar configuration referenced to linked mastoids. ABM B-Alert® software acquires the data and quantifies alertness, engagement and mental workload in real-time using linear and quadratic discriminant function analyses with model-selected PSD variables in each of the 1-hz bins from 1 - 40 Hz, ratios of power bins, event-related power (PERP) and/or wavelet transform calculations.

The data processing begins with the eye-blink decontaminated EEG files containing second-by-second calculations of the probabilities of High EEG-Engagement (EEG-E) (Levendowski et al, 2001, Berka et al, 2004). The neuropsychological tasks used to build the algorithm, and subsequently used to individualize the algorithm's centroids were presented using proprietary acquisition software. The algorithm was trained using EEG data collected during the Osler maintenance of wakefulness task (OSLER) (Krieger et al., 2004), eyes closed passive vigilance (EC), eyes open passive vigilance (EO), and 3-choice active vigilance (3CVT) tasks to define the classes of sleep onset (SO), distraction/relaxed wakefulness (DIS), low engagement (LE), and high engagement (HE).

Simple baseline tasks were used to fit the EEG classification algorithms to the individual so that the cognitive state models can then be applied to increasingly complex task environments, providing a highly sensitive and specific technique for identifying an individual's neural signatures of cognition in both real-time and offline analysis. These methods have proven valid in EEG quantification of drowsiness-alertness during driving simulation, simple and complex cognitive tasks and in military, industrial and educational simulation environments. (Levendowski et al, 2002, Stevens et al, 2007, Berka et al, 2005).

Neurophysiologic Synchronies

The neurophysiologic synchronies (NS) that we are studying can be thought of as the second-by-second quantitative co-expression of the same neurophysiologic / cognitive measures by members of the team. We have developed a four-step modeling approach with the outputs of each step providing a different perspective of team neurodynamics. The four steps outlined in Figure 1 are: A) Data Normalization; B) Unsupervised Artificial Neural Network (ANN) Clustering; C) Hidden Markov Temporal Modeling (HMM), and; D) NS Data Stream Entropy.

For the generation of generic ANN and HMM models EEG-E data was pooled from 8 SPAN sessions (31,450 team training vectors or ~ 8 hours of teamwork) which were used as the training set. The position of each of the team members in the training vector was the same as described above. The team highlighted in Figures 3-5 was not part of the training set.

The first step (A), data normalization, equated the absolute levels of EEG-E or EEG-WL of each team member with his/her own average levels over the period of the task. This identified whether a team member was experiencing above or below average levels of EEG-E or EEG-WL; and, whether the team as a whole was experiencing above or below average levels.

As described previously (Stevens et al, 2010a) in this normalization process the EEG-E levels were partitioned into the upper 33%, the lower 33% and the middle 33%; these were assigned values of 3, -1, and 1 respectively, values chosen to enhance visualizations. The next step (B) combined these values at each epoch for each team member into a vector representing the state of EEG-E for the team as a whole; these vectors were used to train ANN to classify the state of the team at any point in time (Stevens al, 2010a).

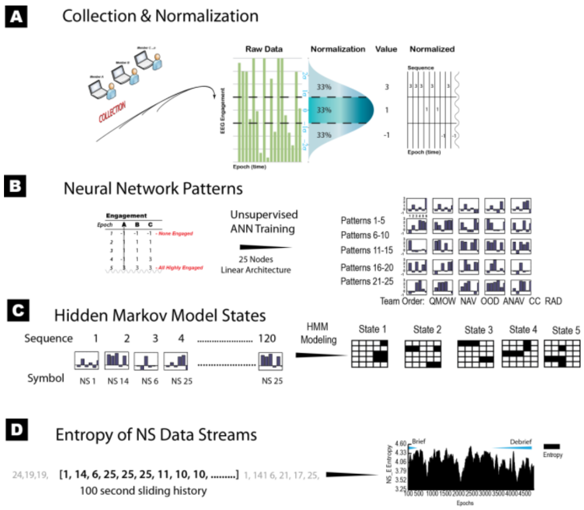

Figure 1. Layered Analytic Model for Detecting and Describing Neurophysiologic Synchronies.

In this process the second-by-second normalized values of team EEG-E for the entire episode were repeatedly (50-2000 times) presented to a 1 x 25 node unsupervised ANN. The result was a series of 25 patterns that we call neurophysiologic synchrony patterns that show the relative levels of EEG-E for each team member on a second-by-second basis (Figure 2).

In Step C the sequences of NS Patterns were viewed as output symbols from hidden states of a team and HMM were developed to characterize these states. The NS data stream for the combined team data was segmented into sequences of 120 seconds and HMMs were trained with these sequences assuming 5 hidden states as previously performed when modeling problem solving learning trajectories (Soller & Stevens, 2007). Training was for 500 epochs and resulted in a convergence of 0.0001. Next, the most likely state sequence through the performance was generated by the Viterbi algorithm. The outputs of the modeling of NS Pattern streams by HMM are termed NS States.

While ANN Pattern and HMM State changes help identify transition points and preferred patterns, a quantitative measure of the teams' dynamics would be useful for comparing across teams or with other teamwork metrics (Step D). As the NS Patterns are symbolic, one approach is to calculate the Shannon entropy of the NS data stream (Shannon, 1951).

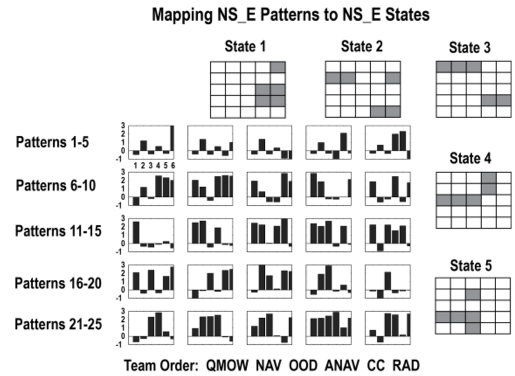

Figure 2. Team NS_E Profiles after ANN Training. The center histograms show the 25 NS Patterns obtained after ANN training. The order of team members associated with each histogram bar is shown below. The surrounding matrices map the NS_E Patterns to the NS_E States from HMM modeling.

The idea of entropy is derived from information science and is a measure of the level of uncertainty or "amount of mix" in a symbol stream. Calculated entropy is expressed in terms of bits and the maximum entropy that we could expect from the 25 NS Patterns if they were randomly distributed would be log2 (25) or 4.64. For comparison, an entropy value of 3.6 would result in only 12 of the NS Patterns randomly expressed. To develop an entropy profile over a SPAN session the NS Shannon entropy was calculated at each epoch using a sliding window of the values from the prior 100 seconds. The idea was that as teams entered and left periods of organization the entropy would fluctuate as fewer or more of the 25 NS_E patterns were expressed. This relationship is shown in Figure 3.

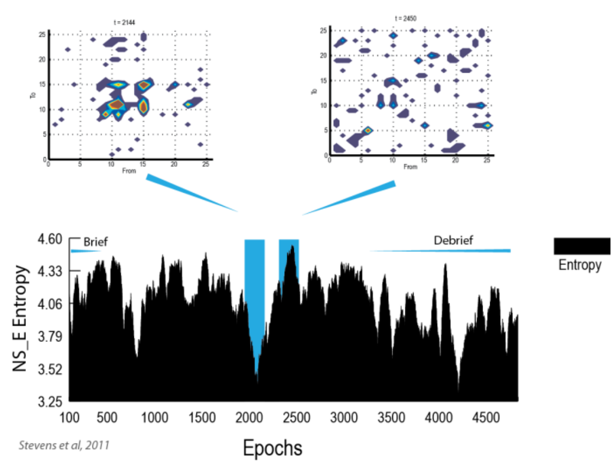

Figure 3. NS_E Entropy Profile for a SPAN Team. This figure shows the Shannon entropy for NS_E at each epoch over a sliding window of the prior 100 seconds. Above the entropy profile are the transition matrices for the two highlighted 120 second periods. These transition matrices show the NS_E Patterns at times t and t+1.

RESULTS

Figure 4 compares the NS_E States following single-trial and generic modeling of the same SPAN performance. Both models showed the NS_E State transitions at the Scenario / Debrief junction (epoch 3390) and at epoch 4400 of the Debrief. They also both showed a long period at the beginning of the scenario (epochs 590 – 1000) where a single state predominated and a period (3100 – 3385) at the end of the scenario where the same state predominated. These task-junction transitions have been observed in ten different SPAN sessions where single-trial and generic modeling was conducted in parallel.

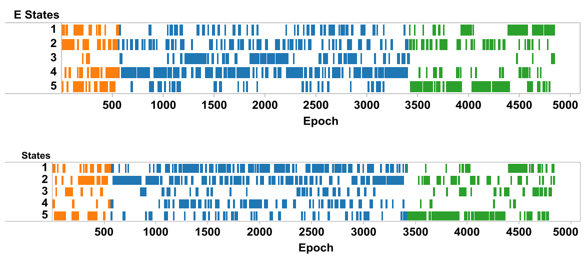

Figure 4. Comparison of NS_E State Expressions when Modeled with Single Trial (top) or Generic (bottom) ANN and HMM Models. The dark portion in the middle is the Scenario segment and the lighter portions to the left and right are the Brief and Debrief segments.

Another validation approach compared the Shannon entropy of the NS Pattern data streams from each model. This metric is derived from information science and measures the degree of uncertainty in a data stream (Shannon, 1951). The top of Figure 5 is a scatter plot of Shannon entropy for the NS_E values from single-trial and generic NS_E models which were highly correlated (R = 0.86, R^2 = 0.74). The line graph below shows the co-fluctuations of the two entropy streams where there was a close concordance. The overall NS_E entropies of the single-trial and generic were also similar (Mean ± SD = 4.073 ± 0.23 and 4.071 ± 0.22). Combined, these data indicate that the generic NS_E models provided a close approximation of those obtained with single-trial modeling.

NS_E Expression across Teams and SPAN Sessions

One question that could be approached with the heterologous NS_E models is: How, consistently different NS_E States were used across teams and / or training sessions. Figure 6 shows the frequency distribution of NS_E for an expert (E2) and two Junior Officer teams (T4 and T5). Each performed two simulations; except an additional Junior Officer team performed a single session (T1). The NS_E frequencies were separated into the Scenario, Debrief and Briefing segments of the simulation based on prior studies (such as Figure 2) that have shown there are often dynamic NS_E shifts at these segment junctions.

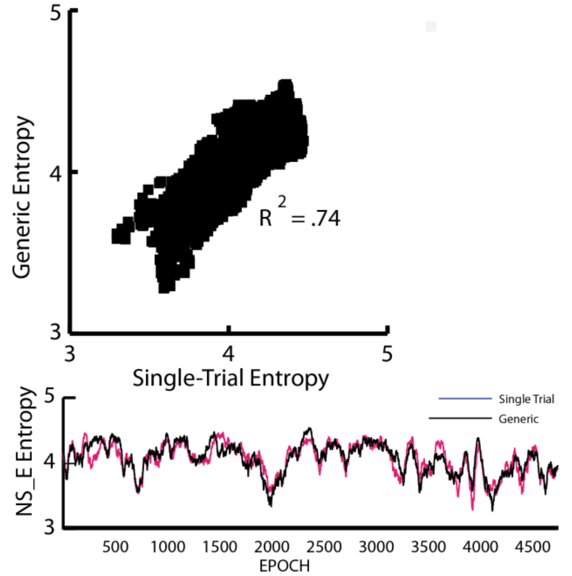

Figure 5. Comparison of the Shannon Entropy of NS_E Pattern Expression from Single Trial or Generic Models. The top figure shows a scatterplot of the entropy from the two data streams; below is a line chart comparing the second-by-second fluctuations.

For most teams the dominant NS_E States during the Scenario segment were 1 and 2. Referring to Figure 2, these states represent where most of the team was highly engaged. These appeared to represent the normal operating mode for SPAN teams as their expression was diminished during the Debriefing segment and to a lesser extent in the Briefing segment.

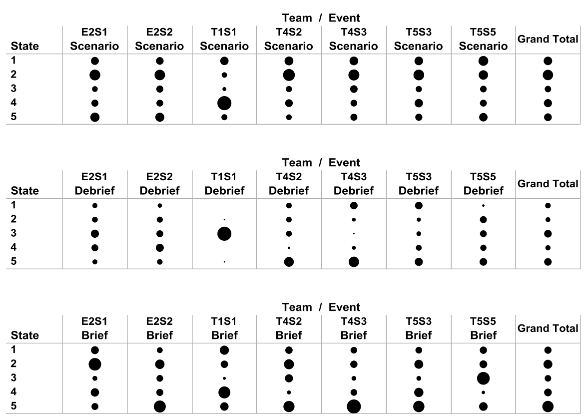

Figure 6. Team NS_E State Distributions Across Teams and Sessions.

While there were slight differences in the NS_E State frequencies for E2, T4 and T5 the performance of team

T1 was different with NS_E State 4 dominating. Referring to Figure 2, this state was one where many of the team members' had low EEG-E. The differences across teams were larger when comparing across the Debrief and Brief segments. Here there was proportionally higher expression of NS_E States 3 & 4 (teams with low EEG-E) for the expert team and NS_E State 5 for the Junior Officer teams.

While the above comparisons showed some differences across sessions and teams the lack of resolution made it difficult to make quantitative assessments across teams or comparisons between teams with different levels of experience, like SOAC vs. experienced submarine (SUB) teams. In other studies an examination of the predominant NS_E Patterns showed that they were different for SOAC and SUB teams (Stevens & Gorman, 2011) and that SUB teams used more of the available NS Patterns than did SOAC teams. This suggested that direct comparisons of the NS_E entropy streams may be a useful indicator of the experience / proficiency of teams.

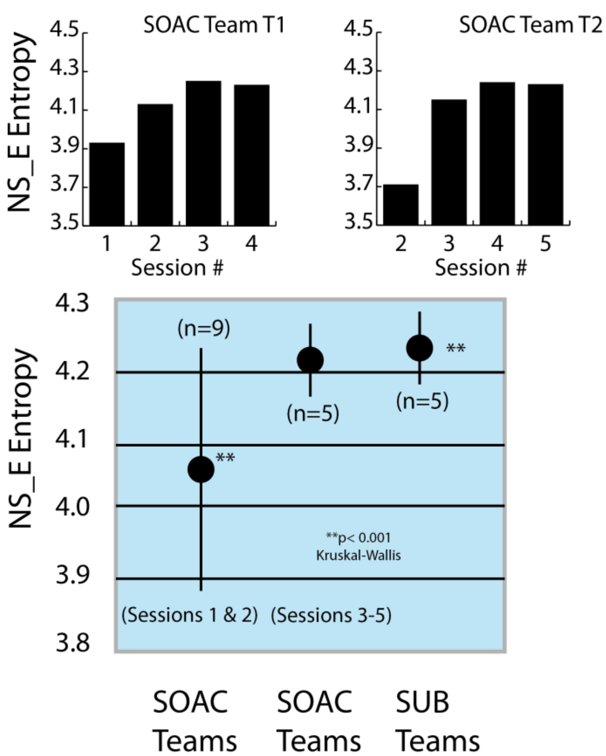

Figure 7. NS_E Entropy Levels for SOAC and SUB Teams.

The NS_E Patterns from 14 SOAC and 5 SUB team sessions were generated by testing the EEG-E data streams on the generic networks. Next, the NS_E entropies were calculated as described in the Methods. SUB teams had the highest levels of NS_E entropy while the lowest entropies were from the first two sessions by SOAC teams. The histograms in the top figure show the progressive increase in NS_E entropy as two of the teams gained experience.

The final series of studies sought to determine how well generic networks would perform in other teamwork situations. Neurophysiologic synchronies have been used to study other forms of teamwork including scientific problem solving by teams of three high school students (Stevens et al, 2009) as well as NAVAIR Anti-Submarine Warfare Teams (ASWT), which were also three member teams.

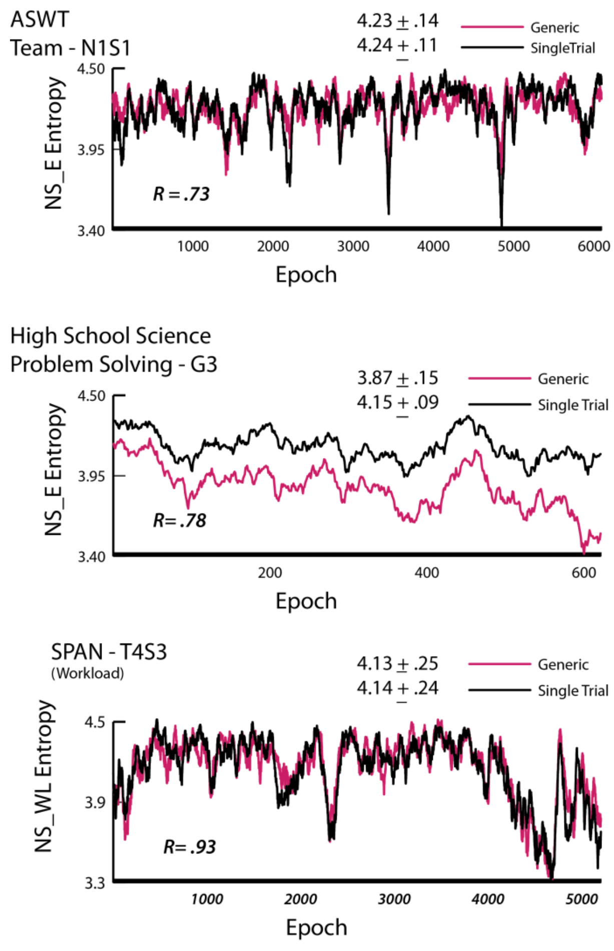

Figure 8. Comparison of Single Trial and Generic ANN NS Entropies. Comparisons of NS_E entropies are made for an ASWT team (top), and a high school scientific problem solving team (middle). The lower figure shows the NS_WL entropy comparisons for a SOAC SPAN team.

Three-person NS_E generic ANN and HMM models were prepared from 62,021 epochs of EEG-E data; the data was from six teams of high school students that performed substance abuse science simulations, emotion recall experiments, three-member SPAN teams and brainstorming sessions (Stevens et al, 2010b). First, EEG-E vectors were prepared from an experienced ASWT and were classified by either these generic models, or single-trial models. Similar to the SPAN results shown in Figure 5, the overall NS_E entropies and variances were similar, the correlation between the two data streams was high, and when the two data streams were co-plotted they showed similar temporal dynamics (Figure 8). The high school problem solving team also showed similar dynamics for NS_E when tested on the generic and single-trial models although the overall NS_E entropy level was lower when tested on the generic models. A third generic model was created from the EEG-WL vectors of multiple SPAN teams and the dynamics of the neurophysiologic synchronies for EEG-WL (NS_WL) were compared with those generated from homologous EEG-WL models from the T4S3 SPAN team. The overall correlation coefficient was the highest of the comparisons made in this study. The entropy dynamics of the two data streams co-fluctuated, very closely.

DISCUSSION

Prior to developing and validating the generic NS models only the first three usefulness criteria outlined in the introduction could be approached: We had used NS expression to study teamwork in multiple settings with different size teams (Stevens et al, 2010b); we had demonstrated changing dynamics of their expression over long and short time periods (Stevens et al, 2010a); and, we had shown that these dynamics related to some aspects of speech (Stevens et al, 2009). As shown in this study, with the standardized models we can now begin to compare NS expression across teams, training sessions and levels of expertise.

Validation of the generic models was approached two ways; one using NS Patterns from ANN clustering of EEG-Engagement levels and one using NS States which provides a temporal component to the NS Patterns (Stevens et al, 2010b). One of the most reproducible features of SPAN performances is the change in NS_E States at the junction between the Scenario and Debriefing. The generic and single-trial models reproducibly detected these temporal features at this junction indicating an equivalent sensitivity of large task changes. A different form of validation drew on the concept of entropy from information theory which measures the degree of uncertainty in a data stream of symbols. These entropy profiles highlighted periods of high and low entropy modeled by both approaches. From the NS_E Pattern transition matrices, the periods of low entropy were those where the team was more cognitively organized. The significance of these re-organizations is not clear, but may relate to periods of unusual tension or stress for a team (Stevens et al, submitted). However, the strong concordance in the entropy profiles between the single-trial and generic models provided an additional validation of the sensitivity and specificity of the generic NS_E models.

One of the most interesting and potentially useful findings from the generic models is the differences in NS_E entropy between SOAC teams beginning their training, and experienced SUB teams. These findings may provide an objective, quantitative measure of team proficiency which can be tracked over training or across different training protocols.

The fluctuations in entropy we have observed during the model validations may also provide a rapid readout for how teams respond to different events during a simulation task. To further such studies, we are incorporating these models into software systems that will supply rapid (minutes) after training feedback to teams and provide a framework for future real-time adaptive monitoring and training.

ACKNOWLEDGMENTS

We extend a special thanks to Capt. Kenneth Swan, Thomas Wohlgemuth, Lt. Robert Buckles and Dr. Jerry Lamb at the Submarine Learning Center, Groton, CT for logistics support.

REFERENCES

Berka, C., Levendowski, D. J., Cvetinovic, M. M., Petrovic, M. M., Davis G. et al. (2004). Real-Time Analysis of EEG Indexes of Alertness, Cognition, and Memory Acquired With a Wireless EEG Headset. International Journal of Human-Computer Interaction. Lawrence Erlbaum Associates, Inc. 17(2), 151-170.

Berka, C, Levendowski, DJ, Ramsey, CK, Davis, G, Lumicao, MN, Stanney, K, Reeves, L, Regli, S, Tremoulet, PD, Stibler, K. (2005). Evaluation of an EEG-Workload Model in an Aegis Simulation. *Proceedings of the SPIE Defense and Security Symposium, Biomonitoring for Physiological and Cognitive Performance during Military Operations* 2005; 5797: 90-99.

Cooke, N.J., Gorman, J.C., and Kiekel, P. (2008) Communication as Team-level Cognitive Processing. in *Macrocognition in Teams*, Letsky, M.P., Warner, N.W., Fiore, S.M., and Smith, C.A.P (eds).Ashgate Publishing, Burlington, VT.

Daw, C. S., Finney, C. E. A., & Tracy, E. R. (2003). A review of symbolic analysis of experimental data. *Review of Scientific Instruments.* 74, 915.

Gorman, J., (2005). The Concept of Long Memory in Assessing the Global Effects of Augmented Team Cognition (Paper presented at the HCI International Conference 22-27 July 2005, Las Vegas, NV.

Krieger, A.C. and Ayappa, I. (2004). Comparison of the maintenance of wakefulness test (MWT) to a modified behavioral test (OSLER) in the evaluation of daytime sleepiness. Journal of Sleep Research 13 (4), 407–411.

Lindenberger, U., Li, S-C, Gruber, W., & Muller, V. (2009). Brains swinging in concert: cortical phase synchronization while playing guitar. BMC Neuroscience 10: 22-34.

Salas, E., Cook, N. J., Rosen, M. A. (2008) On Teams, Teamwork, and Team Performance: Discoveries and Developments. *Human Factors: The Journal of the Human Factors and Ergonomics Society* Vol. 50 (3): 540-547.

Shannon, Claude E.: Prediction and entropy of printed English, The Bell System Technical Journal, 30:50-64, January 1951.

Soller, A., & Stevens, R. (2007) Applications of Stochastic Analyses for Collaborative Learning and Cognitive Assessment. In Advances in Latent Variable Mixture Models, Gregory Hancock and Karen Samuelson (Eds.). Information Age Publishing.

Stahl, G. (2006). Group Cognition (Cambridge, MA: MIT Press.

Stevens, R. H., Galloway, T., and Berka, C. (2007). Allocation of time, workload, engagement and distraction as students acquire problem solving skills. In “Foundations of Augmented Cognition”, 4th Edition, D. Schmorrow, D. Nicholson, J. Drexler and L. Reeves (eds). pp. 128-137.

Stevens, R. H., Galloway, T., and Berka, C., & Sprang, M. (2009). Neurophysiologic Collaboration Patterns During Team Problem Solving. Proceedings: Human Factors and Ergonomics Society 53rd Annual Meeting, October 19-23, 2009, San Antonio, TX.

Stevens, R.H., Galloway, T., Berka, C., & Behneman, A., (2010a). Temporal sequences of neurophysiologic synchronies can identify changes in team cognition. Proceedings: Human Factors and Ergonomics Society 54th Annual Meeting, September 27-October 1, 2010, San Francisco, CA. pages 190-194.

Stevens, R.H., Galloway, T., Berka, C., & Behneman, A., (2010b). Stevens, R., Galloway, T., Berka, C., & Behneman, A. (2010). Identification and Application of Neurophysiologic Synchronies for Studying Team Behavior. In Proceedings of the 19th Conference on Behavior Representation in Modeling and Simulation, 21-28.

Stevens, R. H., Soller, A., Cooper, M., and Sprang, M. (2004) Modeling the Development of Problem Solving Skills in Chemistry with a Web-Based Tutor. *Intelligent Tutoring Systems.* J. C. Lester, R. M. Vicari, & F. Paraguaca (Eds). Springer-Verlag Berlin Heidelberg, Germany. 7th International Conference Proceedings (pp. 580-591).

Warner, N., Letsky, M., and Cowen, M. (2005). Cognitive Model of Team Collaboration: Macro-Cognitive Focus. In *Proceedings of the 49th Human Factors and Ergonomics Society Annual Meeting*, September 26-30, 2005. Orlando, FL.

Team Neurodynamics

by The
Learning Chameleon, Inc.