SAND2012-8013C

A Data-Driven Approach to Assessing Team Performance through Team Communication

Jon Whetzel, Chris Forsythe,
Michael Haass, Matt Glickman

Sandia National Laboratories
Cognitive Science Research & Applications
September 26, 2012

Sandia National Laboratories

Teamwork and Adaptability

- **Individual humans can do tremendous things, but our most extraordinary capabilities arise from coordinated team effort**
  - *Central* to national security
- **Submarine piloting and navigation is a prime example**
  - Situation awareness requires integration from multiple sensory modalities
- **Teamwork is thus a focus of training.**
  - In particular, team-level *adaptability*
- **Assessing team adaptability requires SME's**
  - Hard to objectively quantify

Our Goal

- **SME's are a tremendously valuable resource, but are always in limited supply**
- **Our goal**
  - *Expand the availability of SME expertise via technology that automatically models SME assessment capabilities.*
- **Applications and benefits**
  - Permit SME's to delegate tasks, expanding their potential impact
  - Provide for a measure of team assessment in situations where SME's are not available

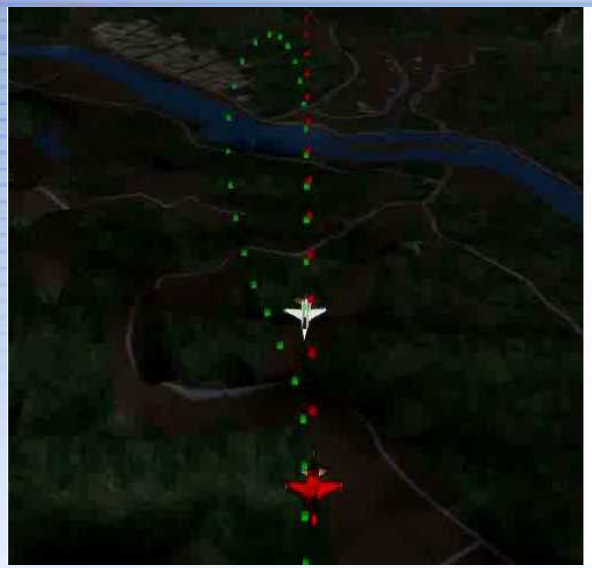
Outline

- Our Goal
- Outline
- The AEMASE Approach
- The SNAP-training target domain
- Experiment 1: Data from our own lab
- Experiment 2: Data from real training
- Conclusions and Future Work

Our Approach

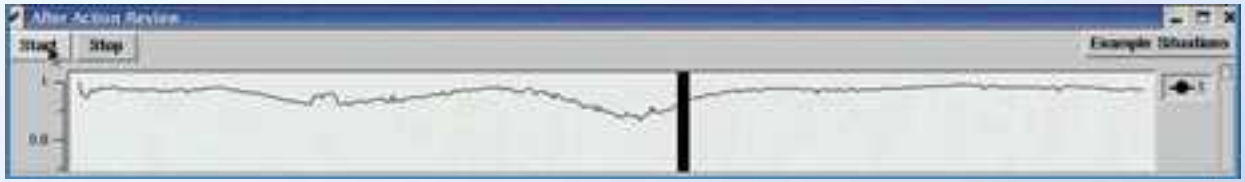
- **AEMASE**
  - ***A****utomated* ***E****xpert* ***M****odeling* ***A****nd* ***S****tudent* ***E****valuation*
- **Method**
  - ***Collect***
    - Data reflecting trainee behavior, actions
    - Associated SME assessments
  - ***Model***
    - Apply data-driven analytic techniques to yield a model mapping training data to likely assessments.
  - ***Use and Refine***
    - Begin generating assessments from new data
    - Record when SME note model's errors
    - Refine model based upon SME feedback
    - Repeat as needed

F16 Pilot Behavior Example



Expert ***Novice***

Modeling process



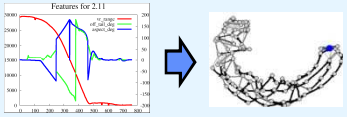
Comparison of novice behavior trace against expert model

Application of AEMASE

Idea conceived, implemented and shown feasible through SNL LDRD

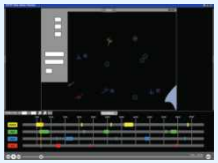
Automated Expert Modeling and Student Evaluation

1. Provide examples of expert performance



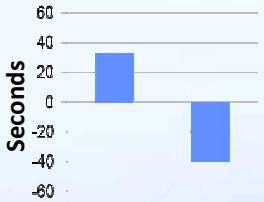
2. Machine learning used to acquire expert model

3. Student compared to expert model to identify and target training to individual deficiencies.

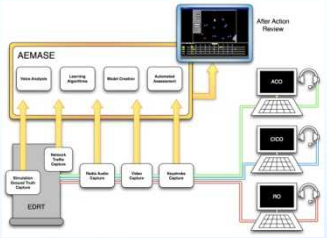


Experimental tests to establish validity and utility for training E-2 Naval Flight Officers

Friendly fighters committed sooner in response to enemy aircraft (*t* =2.03*; *p<0.05)*



Integration with operational training system



E-2 Enhanced Deployable Readiness Trainer

Scheduled to be Fielded

NSAWC Fallon (Top Gun School)

NS Norfolk

NAS Point Mugu

Sandia National Laboratories

AEMASE and Team Assessment

- **Target domain**

  Surfaced Navigation & Piloting Training

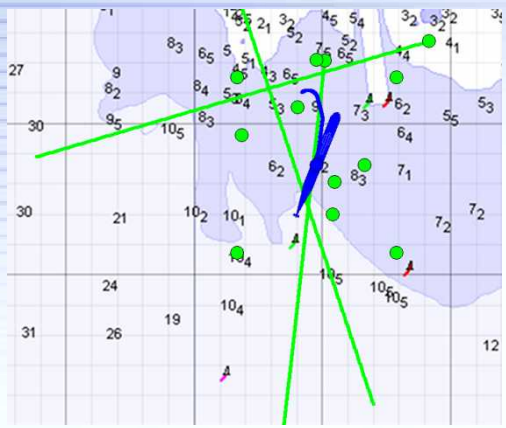
- **Research Question**
  - *Can one develop technology to automatically identify relevant communication patterns within data recorded from teams in SPAN training?*
- **Differences from previous applications**
  - Not an inherently spatial task
  - Team adaptability is a high-level, emergent feature
    - *Relevant time-scale for evaluation is less clear*

Available Data

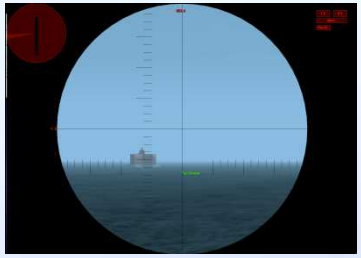
- **Potential sources of data**
  - Trainee verbal communication
  - Simulation state over time, events
  - Trainee physical actions (movement, control acutation)
  - Static factors: Scenario details & team history
- **Settled upon verbal communication *only***
  - No speech recognition
  - Plausibly sufficient
  - Minimal reliance on other systems for collection

Experiment 1: Proof of Concept Study

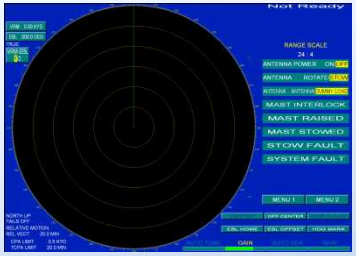
- **Executed simulation exercises in-house that reflect the tasks undertaken by the piloting party in a SPAN training environment**
  1. **Learn and practice litany for round of contacts (cyclical communication routine)**
  2. **Maneuver ship to allow visual identification of multiple unknown radar contacts (static & moving)**



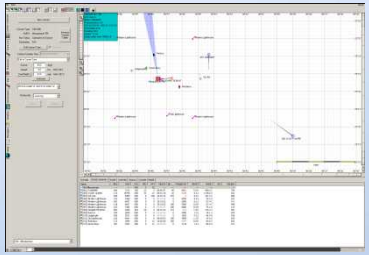
- **Use SubSkillsNet as exercise platform**
  - **Collect data on entity locations during scenario & audio from each team member**
- **Team roles**
  - ***Radar*: Monitor radar for safety of ship, tag contacts on radar console and report range as requested**
  - ***Scope*: Perform scope sweeps for safety of ship, identify contacts and report bearing as requested**
  - ***Quartermaster*: Coordinate correlation of radar with visual observations (determine order and call radar bearings...)**



Scope



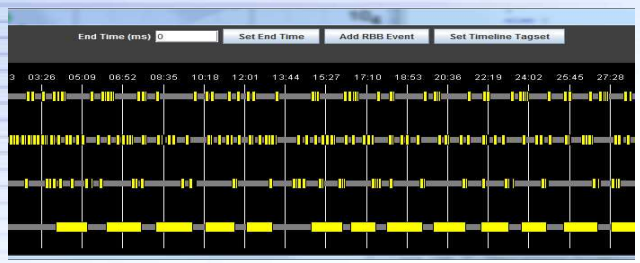
Radar



Quartermaster

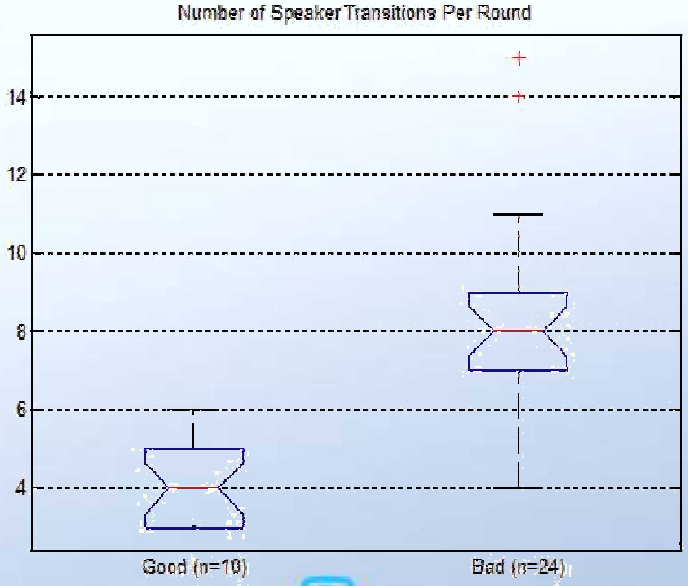
Proof of Concept Study

- Annotate speech communication patterns between members during rounds of contact (n = 34)
  - Use semi-automated process to locate all rounds from recorded audio
  - Each round assigned a binary value
    - "Good" or "Bad"



- Objective: Need to show that a level of goodness assigned to communication can be determined quantitatively within speech data

- Analysis: Number of transactions between team members served as good discriminator
  - $p < 0.05$



Proof of Concept Study

- Need for more flexible classifier that can handle multi-variable, non-linear problem of assessing team communication within an actual training environment
- Create Dynamic Bayesian Network (DBN) based upon features from recorded audio between team members
  - E.g., Is {crewmember} speaking?, Has {crewmember} spoken in past 7 seconds?
- Use DBN as AEMASE model & test accuracy of assessing speech patterns

- DBN provides scoring for each assessed contact round
  - 20 "bad" rounds, 6 "good" rounds
  - Threshold of 0.8 would properly classify 88.4% of rounds

AEMASE has potential to assess team performance based upon observed communication patterns amongst members

Experiment 2: Operational Environment

- Need for inexpensive, non-invasive solution to record communication between team members

- "Badges" from Sociometric Solutions, Inc.
  - Originally developed at MIT Media Lab
  - Records speech events through sampling fundamental frequency observed over microphones
  - IR sensor to detect when people face each other
  - Bluetooth for approximate distance between people
  - Accelerometer for movement
  - Used for assessing C2 teams in an Incident Command scenario [Skarin *et al*, 2010]

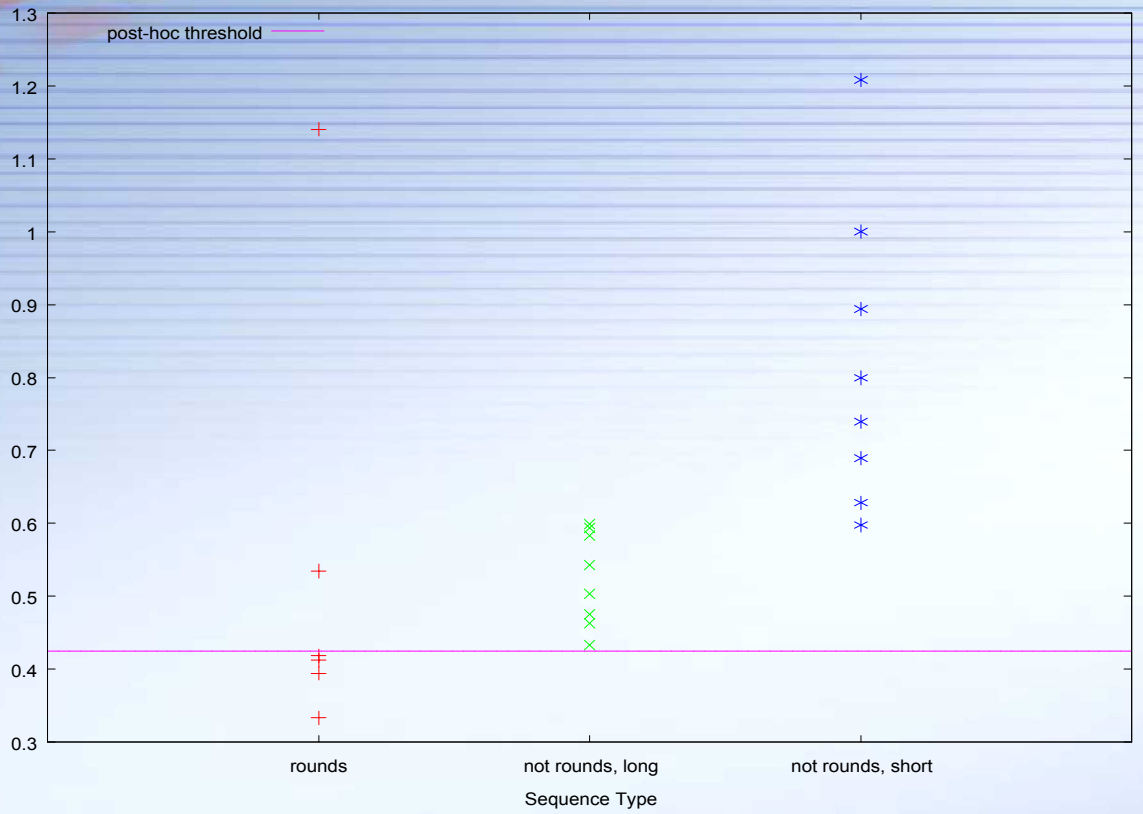
Data Collection at NSS

- **Data collection event: 30 Jan – 1 Feb 2012**
  - Record team communications and significant events with crews conducting training in the SPAN at Navy Submarine Base New London.
  - Data collected from three separate training exercises from two different crews
- **Data collection protocol**
  - Key crew members assigned a badge while conducting their training
  - Badges placed in static locations to collect spatial movements of crewmembers
  - Experimenters would record occurrences of key events (e.g., start/stop of cyclic routine, instrument failures, significant course changes).
- **Collection protocol approved by Human Subjects Board at Sandia and ONR**

Analysis of NSS Data

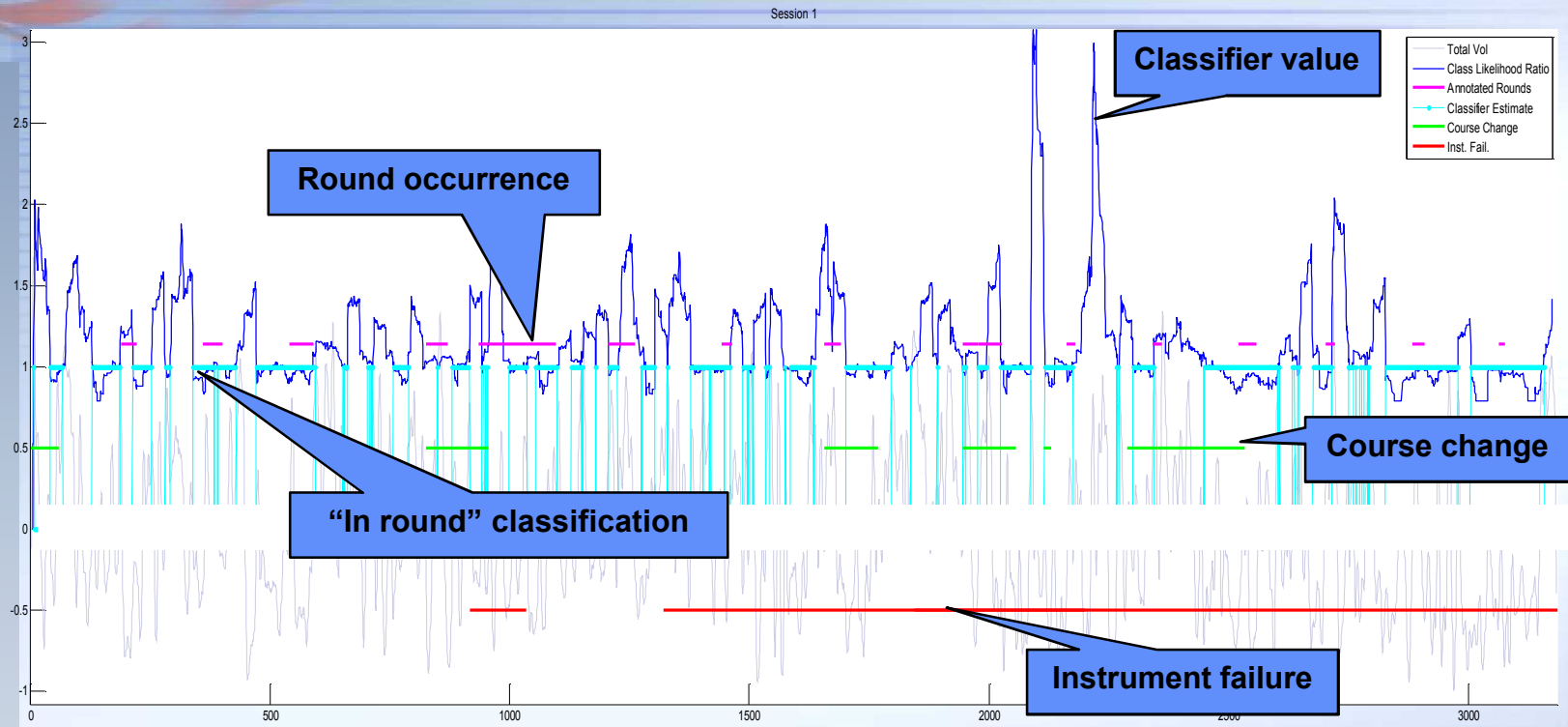
- AEMASE requires identifying example behaviors of interest within dataset to generate model
- For the AEMASE approach, we generated models of when teams were engaged within a cyclic routine
- Why?
  - Cyclic routine behavior defined within Navy Submarine Doctrine, yet no quantitative assessment
  - Cyclic routine patterns can acceptably vary within teams given situation on-board
  - Control for differences between observers/instructors on acceptable communications

Analysis of NSS Data



- **Post-hoc classification of “round” vs. “not round”**
  - **Achieve 85% using the DBN classifier**
  - **Achieve 92% when partitioning “not round” segments equally**
- **DBN classification breaks down when attempting to classify segments in real-time**

Analysis of NSS Data



- Tendency for DBN classifier to generate false positives
- But how accurate does this need to be to aid crews?

Conclusions & Future Work

- **Approach is promising**
  - *Data-driven approach automatically identified relevant communication patterns within data recorded from teams working within the SPAN context.*
- ***Many* ways to improve performance, but first …**
- **Important Questions**
  - *What patterns are most important to recognize?*
  - *How well do models have to perform?*
- **Answers are specific to application context**
  - *Where are SME's overworked, and where are teams underinformed?*

Acknowledgements

- Funding
  - Office of Naval Research

- ATSNAP Technical Lead
  - David Kern, Ph.D. (Kern Technology Group, LLC)
- Assistance
  - Eric Jones (Aptima, Inc)
  - Jonathan Lansey (Aptima, Inc)
  - Daniel Olguin Olguin, Ph.D. (Sociometric Solutions, Inc)