National Institute of Standards and Technology
Technology Administration, U.S. Department of Commerce

Experiences and Lessons Learned in Evaluating Advanced Military Technologies Throughout Their Development Life Cycle

ASSIST -
Advanced Soldier Sensor Information System and Technology

TRANSTAC -
Spoken Language Communication and Translation System for

Tactical Use

Transformative Apps

CRAIG SCHLENOFF, BRIAN A. WEISS, MICHELLE STEVES

Intelligent Systems Division, National Institute of Standards and Technology
100 Bureau Drive, Mailstop 8230, Gaithersburg, MD USA 20899

EMAIL: craig.schlenoff@nist.gov
PHONE: 301-975-3456

Outline

PRESENTATION OVERVIEW

- About NIST
- DARPA ASSIST
- DARPA TRANSTAC
- DARPA Transformative Apps
- SCORE Framework
- Lessons Learned
- Conclusion

NIST
National Institute of Standards and Technology
Technology Administration, U.S. Department of Commerce

NIST

BACKGROUND INFORMATION

NIST mission

"Promote U.S. innovation and industrial competitiveness by advancing measurement science, standards, and technology in ways that enhance economic security and improve our quality of life."

- **Part of the Department of Commerce**
  - **2,900 employees**
  - **2,600 associates and facility users**

- **$1.6 billion FY 2010 budget**
  - **$856M from Congress**
  - **$580M from ARRA**
  - **$125M from other agencies**
  - **$48M in service fees and directed projects**

- **NIST Laboratories**
- **Technology Innovation Program (TIP)**
- **Manufacturing Extension Partnership**
  - **400 centers nationwide**
- **Baldridge National Quality Program**

Gaithersburg, MD

Boulder, CO

The Division's research and development program focuses on addressing both immediate and long-term industry needs for:

- Open-systems architecture standards
- Intelligent controllers for manufacturing industry and government applications
- Engineering methodologies and software tools for building intelligent systems
- Test methods and metrics for measuring the performance of intelligent systems

National Institute of Standards and Technology
Technology Administration, U.S. Department of Commerce

ASSIST

PROGRAM GOALS and METRICS

- NIST served as the Independent Evaluation Team (IET) of DARPA ASSIST from 2005-2008 (Six live evaluations)
- GOAL - Enhance battlefield awareness and after mission recall via exploitation of information collected by soldier-worn sensors
- METRICS - (as specified by DARPA)
  - The utility of the system in enhancing operational effectiveness
  - The accuracy of object/event/activity identification and labeling
  - The systems' ability to improve its classification performance through learning

National Institute of Standards and Technology
Technology Administration, U.S. Department of Commerce

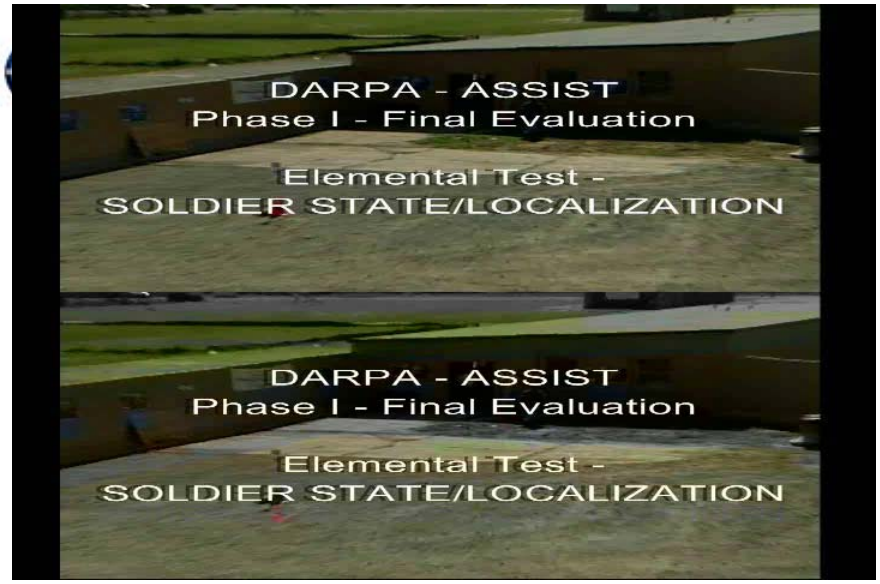
ASSIST

TECHNOLOGIES and PHASES

- After-Mission Reporting
  - Arabic Text Translation
  - Object Detection/Image Classification
  - Face Recognition/Matching
  - Sound Recognition/Speech Recognition
  - Soldier State Identification/Localization
- Real-Time Activities
  - Face Recognition/Matching
  - Shot Localization/Weapon Classification
  - Real-Time Information Viewing/Sharing

Soldier State/ Localization

ELEMENTAL TEST



METRICS

Soldier State

- % Correctly classified type of movement
- % Incorrectly classified type of movement
- % Unclassified soldier movements
- % Correctly identified indoor vs. outdoor activity

Soldier Localization

- Accuracy (m) of mapping all soldier movement
- Accuracy (m) of mapping all outdoor/indoor movement

Presence Patrol

TASK TEST

Mission Objectives– Squad Leaders

- Conduct a presence patrol of the courtyard and warlord compound to verify the locations of key structures and positions (see attached map for key points and routes)
  - View own GPS position on Zypad interface at each key point (noted along the route) and periodically while moving through the environment
  - Annotate map and create voice tags at each key point with information about immediate and/or suspicious activities

  View GPS position of other squad leader on Zypad interface at the other squad leader's key points in addition to their position on the map while moving

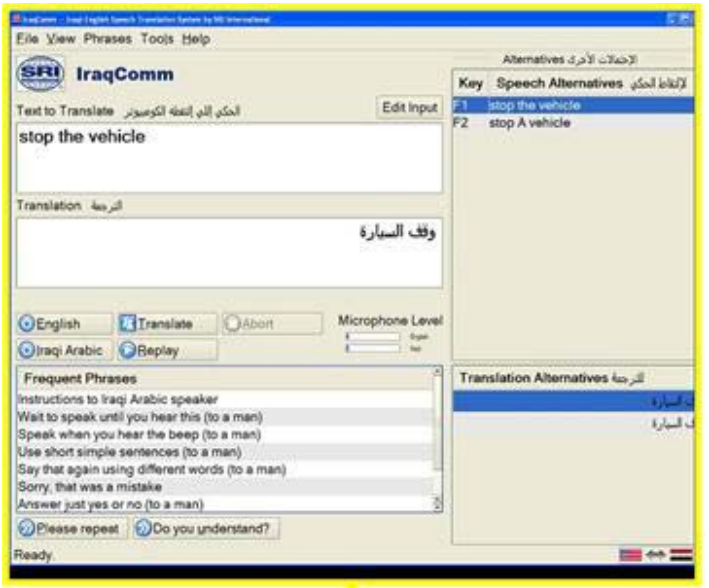
  Locate event data collected in the courtyard while in the warlord compound (should be done both while static and walking)

  Relay (via radio) to other squad leader and platoon leader when a point has been reached

TRANSTAC

PROGRAM GOALS and METRICS

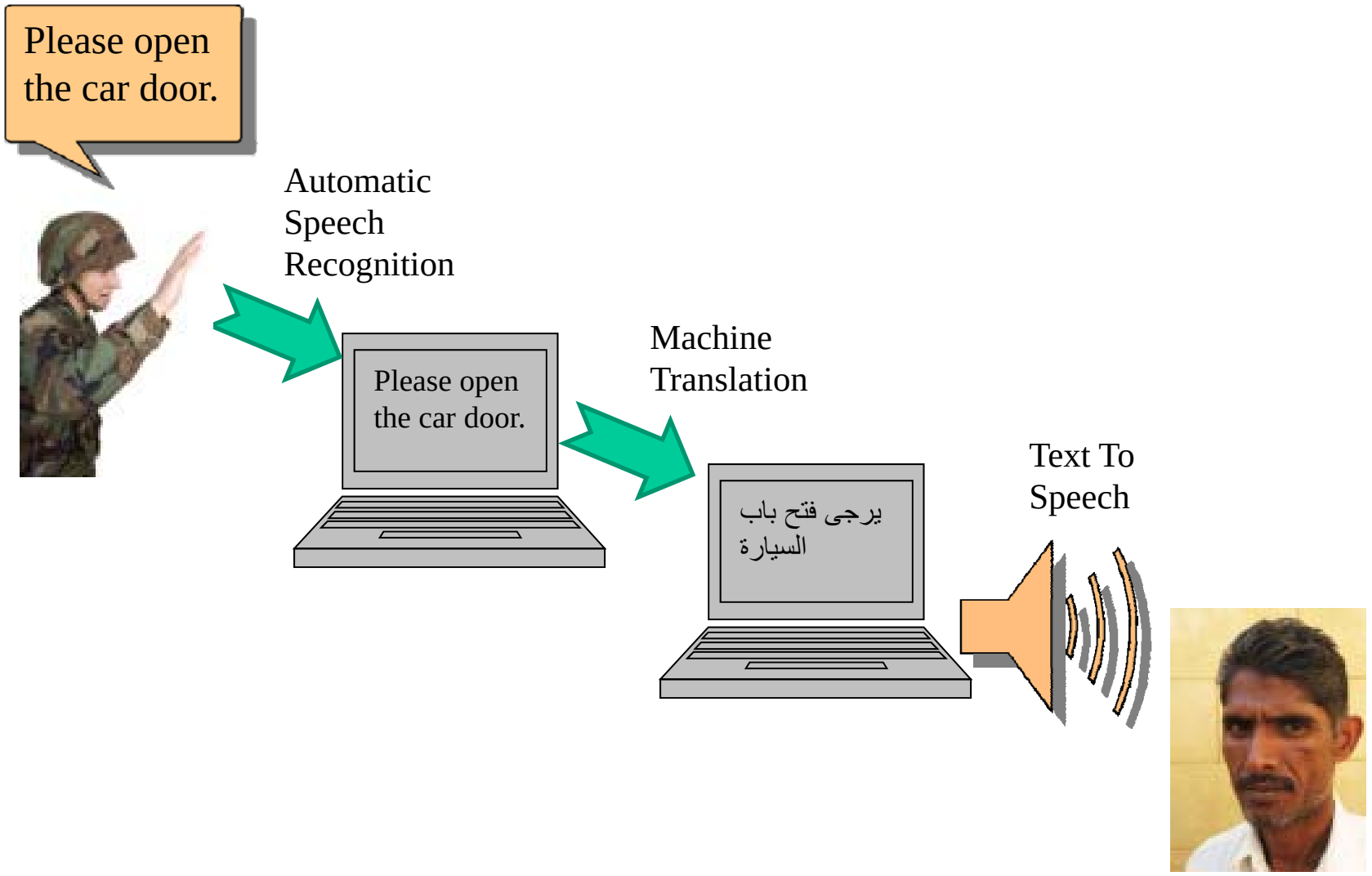
- NIST served as the Independent Evaluation Team (IET) of DARPA TRANSTAC from 2006-2010 (Seven live evaluations)
- GOAL – Demonstrate capabilities to rapidly develop and field free-form, two-way speech-to-speech translation systems enabling English and foreign language speakers to communicate with one another in real-world tactical situations.
- METRICS (as specified by DARPA)
  - System usability testing – providing overall scores to the capabilities of the whole system
  - Software component testing – evaluation components of a system to see how well they perform in isolation



National Institute of Standards and Technology
Technology Administration, U.S. Department of Commerce

TRANSTAC

A QUICK TUTORIAL ON SPEECH TRANSLATION



NIST
National Institute of Standards and Technology
Technology Administration, U.S. Department of Commerce

TRANSTAC

EVALUATION FOCUS

BRIDGING THE GAP
DARPA
POWERED BY IDEAS

What we tested:

- How useful is this technology to the warfighter?
- How well does the technology convey high-level concepts?
- How well does the technology recognize speech?
- How well does the technology translate text?
- How did the technology do in translating low-level concepts?
- How understandable is the speech that is translated?
- How well can the technology be ported to other languages?

NIST
National Institute of Standards and Technology
Technology Administration, U.S. Department of Commerce

TRANSTAC Field Evaluation

National Institute of Standards and Technology
Technology Administration, U.S. Department of Commerce

Transformative Apps

OVERVIEW

Status Quo

- Single–purpose stovepiped systems
- High-cost
- Slow acquisition based on requirements docs

Goal

- Adaptive, multi-app suite
- Ease of use, training
- Cost advantages
- Warfighter-empowered acquisition

National Institute of Standards and Technology
Technology Administration, U.S. Department of Commerce

Transformative Apps

OBJECTIVES

1. **Create new marketplace and business models.**
2. **Maximize use of state-of-the art commercial technology.**
3. **Develop middleware and tools to ensure secure, robust operation on tactical networks.**
4. **Promote participation of military "beta testers" for new ideas and ongoing feedback.**
5. **Lower the barrier of entry to create diverse development community.**

Military apps store

Translator

"Standalone"

TIGR

Loosely Networked

i-Target

Networked

User as screener of innovation from vibrant development community

Development

Download Apps

Screening & "Voting"

Affordable handhelds & apps on a secure network

Pervasive across all ranks

Adaptive. Resilient.
'Crowd' sourced innovation in development & use.

In the Field

NIST
National Institute of Standards and Technology
Technology Administration, U.S. Department of Commerce

SCORE

- SCORE (System, Component and Operationally Relevant Evaluations)…
  - Is a unified set of criteria and software tools for defining a performance evaluation approach for complex intelligent systems
  - Provides a comprehensive evaluation blueprint that assesses the technical performance of a system, its components and its capabilities through isolating and changing variables as well as capturing end-user utility of the system in realistic use-case environments
  - Is unique in that it:
    - Is applicable to a wide range of technologies, from manufacturing to defense systems
    - Elements of SCORE can be decoupled and customized based on the evaluation goals
    - Can evaluate a technology at various stages of development, from conceptual to mature.
    - Combines the results of targeted evaluations to produce a picture of a systems overall capabilities and utility.
  - Is not a new or revolutionary concept. Rather it formalizes evaluation approaches by specifying design elements and goal types.

National Institute of Standards and Technology
Technology Administration, U.S. Department of Commerce

Lessons Learned in ASSIST and TRANSTAC

1. Designing an effective evaluation can be as much of a research issue as the technology development
2. **Keep your eye on the ball (the ultimate objective of the evaluation) and make sure your decisions along the way reflect that goal**
3. **Deeply understand the needs and wants of the technology end users**
4. **Utility and technical performance assessments are both very important perspectives. Each requires different means to gather and process assessment data and yield different types of analyses.**
5. There are often multiple approaches to evaluating a technology where it's crucial to identify those which will achieve the overall evaluation goals given the test constraint.
6. System training data and/or extensive background scenario information may be needed to perform some assessments. This must be accounted for within the test plan.
7. **Understand the interactions of the technology with the test environment and the test personnel to be mindful of the technology's ideal operating conditions and its boundaries.**
8. The background and experience of the test subjects can greatly affect their impression of the systems under test.
9. The structure and content of the technology training and the feedback requests of the test subjects greatly influences the test subjects' perceptions.
10. There are often multiple options available to assess specific metrics so it's critical to identify those options which are optimal to produce the desired assessments.
11. **Be mindful that your metrics and evaluation approach may need to evolve over time.**

NIST
National Institute of Standards and Technology
Technology Administration, U.S. Department of Commerce

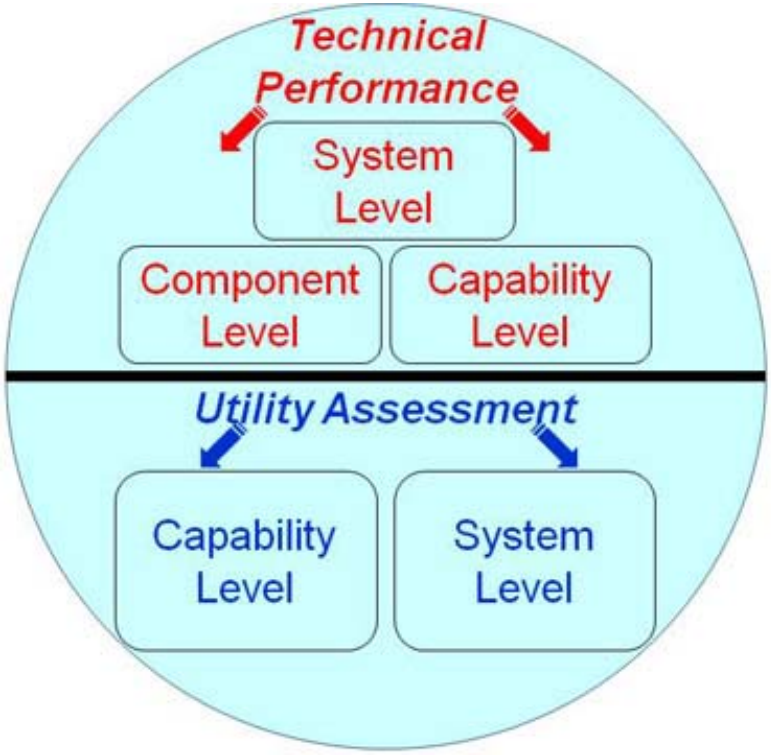
BRIDGING THE GAP
DARPA
POWERED BY IDEAS

Lesson Discussion

- **Keep your eye on the ball (the ultimate objective of the evaluation) and make sure your decisions along the way reflect that goal**
  - To look at the screen or not to look the screen, that is the question…
  - How fast is fast enough (high-level concept transfer)
- **Deeply understand the needs and wants of the technology end users**
  - End-user focus groups
  - Domain identification (Traffic Control Points/Vehicle Checkpoints, Facilities Inspections, Civil Affairs, Medical, Combined Training, and Combined Operations.
  - Operating Environment

Lesson Discussion (cont.)

- **Utility and technical performance assessments are both very important perspectives. Each requires different means to gather and process assessment data and yield different types of analyses.**
  - Led to the development of the SCORE framework
  - Situations where a Soldier liked a technology better in the field even though other technologies performed better



Lesson Discussion (cont.)

- **Understand the interactions of the technology with the test environment and the test personnel to be mindful of the technology's ideal operating conditions and its boundaries.**
  - Lab vs. Field environments
  - Effects of wind, background noise, glare, etc.
  - Balancing developers desires and end user requirements

Lesson Discussion (cont.)

- **Be mindful that your metrics and evaluation approach may need to evolve over time.**
  - Interviews vs. just surveys
  - Usability factors (how will the system be carried around)
  - Usability vs. technical performance (Where is the emphasis?)
  - What is a concept?
  - What part of the system is causing the problem?







Conclusion

SCORE

- Through 13 live evaluations assessing advanced military technologies, we have also learned:
  - The design of a successful performance evaluation can be as much of a research challenge as the design of the technology itself.
  - There is more than one way to skin a cat…and to conduct an evaluation.
  - Your test subjects can change everything … be sure to choose wisely.

*This work was supported by Mari Maeda, PM I2O

For more information, please contact:

Craig Schlenoff

craig.schlenoff@nist.gov

301-975-3456

http://www.isd.mel.nist.gov/projects/score/

Backup

NIST
National Institute of Standards and Technology
Technology Administration, U.S. Department of Commerce

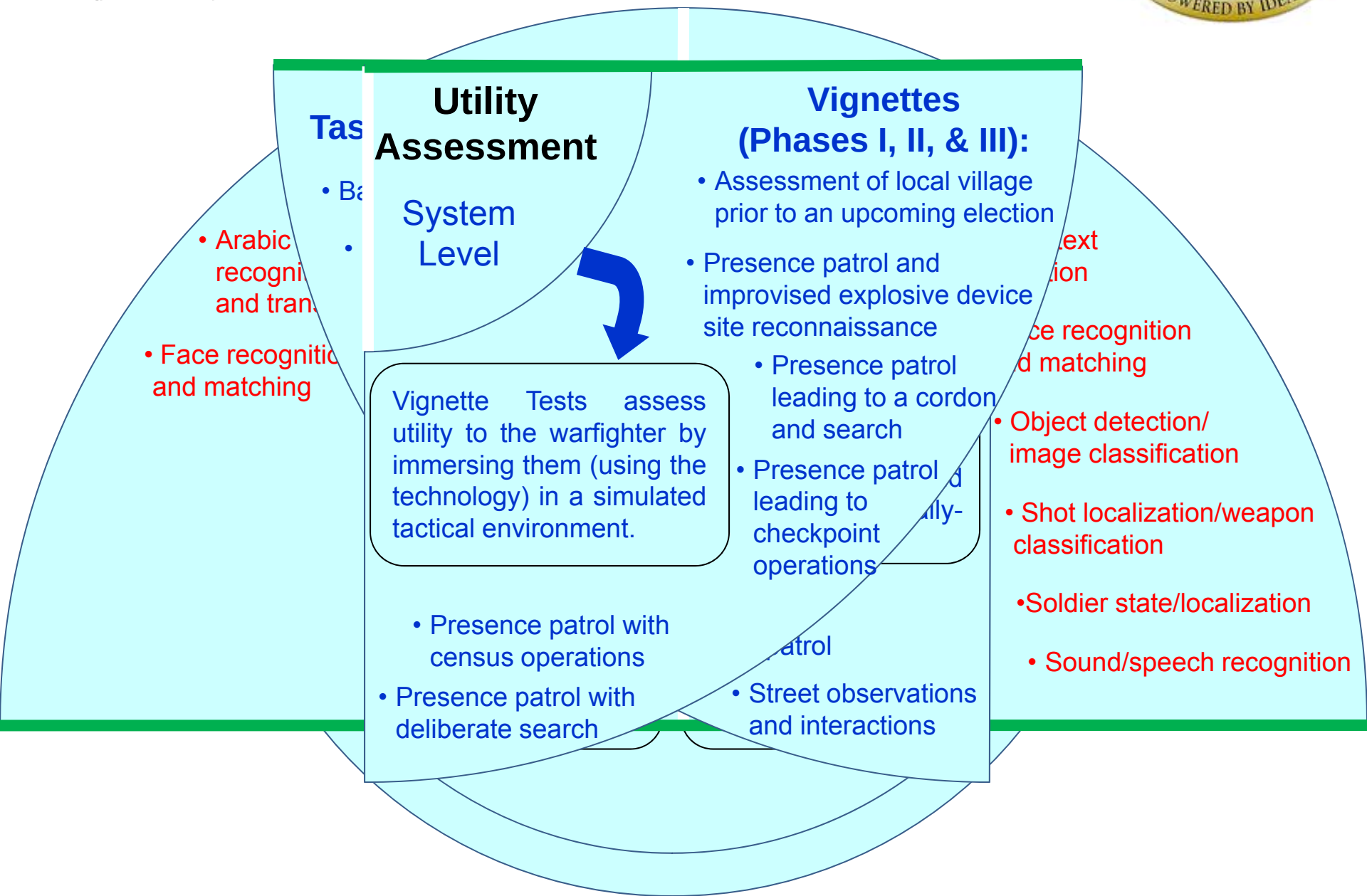
Evaluation Goals

DESIRED ACCOMPLISHMENTS in SYSTEM EVALUATIONS

- Does it do what it claims to do?
- Is it useful to the end-user (military, law enforcement, first responders, industry, etc)?
- What are the factors that would cause the technology to fail?
- What are the key situations that the technology would be most useful for?

"The design of an effective evaluation is as much of a research issue as is the technology development!"

Application to ASSIST



NIST
National Institute of Standards and Technology
Technology Administration, U.S. Department of Commerce

Key Definitions

SCORE

- System – a set of interacting or interdependent components forming an integrated whole intended to accomplish a specific goal
  - Common system characteristics include:
    - Having structure which is defined by its parts (components) and their composition
    - Having behavior, which involves inputs, processing and outputs of material, information or energy
- Component – a constituent part or feature of a system that contributes to its ability to accomplish a goal
- Capability – a specific purpose or functionality that the system is designed to accomplish
- Technical Performance – metrics related to quantitative factors (such as accuracy, precision, time, distance, etc) as required to meet end-user expectations
- Utility Assessment – metrics related to qualitative factors that gauge the quality or condition of being useful to the end-user

Application to TRANSTAC

SCORE

Utility-Field Evaluation @ the System Level

Utility-Field evaluations are intended to assess the warfighter's utility of the field-ready TRANSTAC technologies in more realistic, use-case environments.

- *What was tested?* – Field systems in a more operationally-relevant environment where the English speaker is carrying the technology

Fundamentals

SCORE

SYSTEM LEVEL TESTING
- UTILITY ASSESSMENT

This evaluation type assesses an overall system's utility and usability. This approach is similar to that of the capability utility test, with the exception that the whole system is tested. The system's utility and usability can be evaluated regardless of its state of maturity (formative vs. summative assessment).

Both utility assessments support:

- ✓ Formative evaluations –intended to inform on the system design for those technologies still under development
- ✓ Summative evaluations-intended to validate the value of the technology at the end of its development cycle