Addressing Challenges of Human-AI Teaming Experiments using Naval AI Systems

Jason H. Wong, Ph.D.

Jeffrey J. Bennett

Luis M. Martinez

Douglas S. Lange, Ph.D.

Naval Information Warfare Center Pacific, San Diego, CA

Distribution Statement A. Approved for public release: distribution is unlimited.

1

Introduction

The Navy, like much of DoD, has too much data and not enough information (or actionable insights)

The Navy is investigating AI to enhance decision making: planning, execution, monitoring, and replanning

NIWC PAC focuses on C2 environments:
Carrier Strike Groups
Maritime Operations Centers
Uncrewed Systems

Introduction

Development of AI decision aids without user feedback/user performance studies hurts system adoption
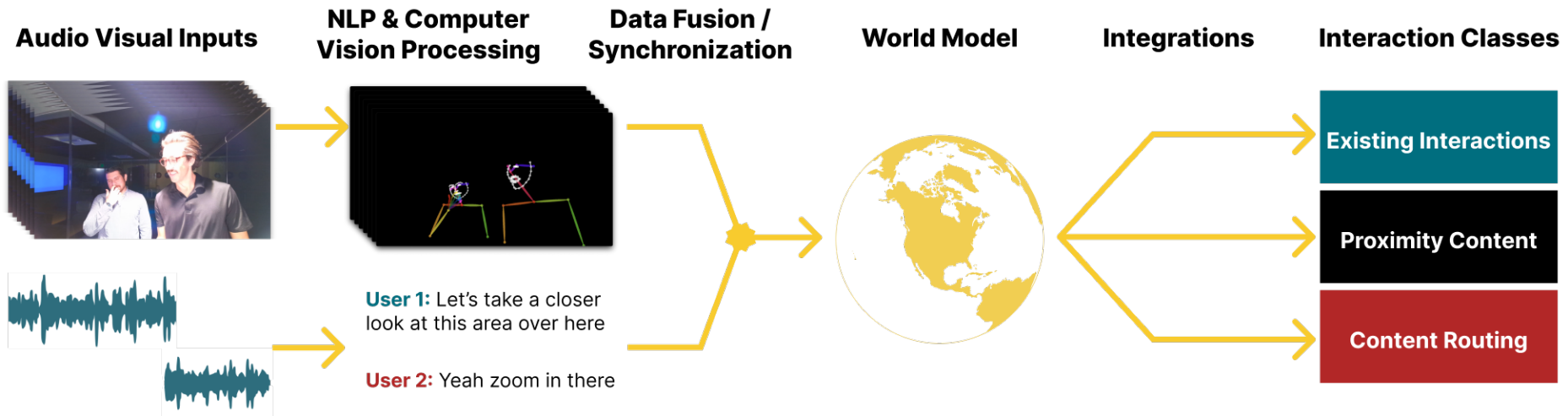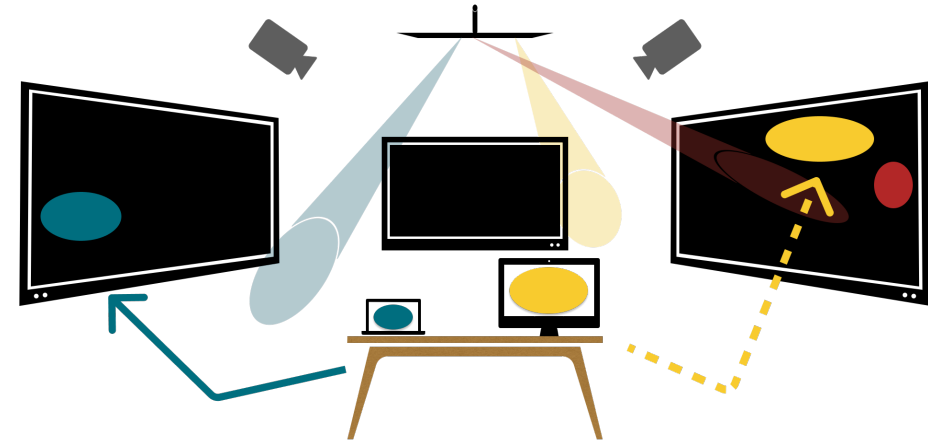
This presentation will:

- Introduce two prototype Navy AI systems with planned human-machine teaming experimentation

- Provide an overview of four experimentation challenges and our proposed mitigations

Case Study 1: Show & Tell System

Ambient intelligent environment with multimodal interaction

Sensors: Depth cameras, ceiling array microphone
Interaction: verbal commands, gestures, visual cues

Audio Visual Inputs **NLP & Computer Vision Processing** **Data Fusion / Synchronization** **World Model** **Integrations** **Interaction Classes**

User 1: Let's take a closer look at this area over here

User 2: Yeah zoom in there

Existing Interactions

Proximity Content

Content Routing

Show & Tell Benefits

Enhanced efficiency through natural interactions, seamless content transfer

Applications in Emergency Operations Centers (EOCs) and any spaces with dynamic context-aware information displays (CICs, MOCs, etc.)

Distribution Statement A. Approved for public release: distribution is unlimited.

5
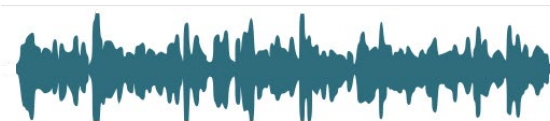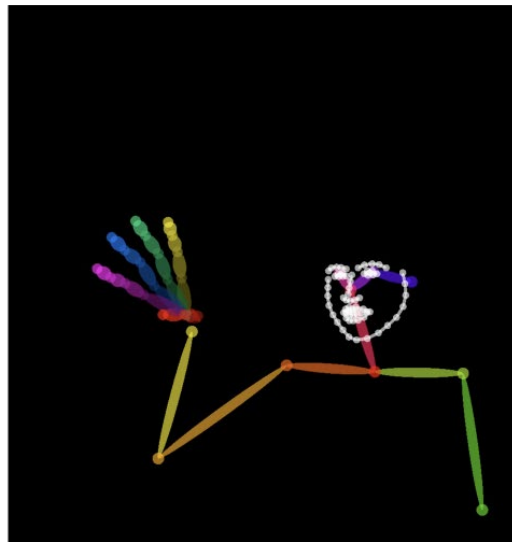
Case Study 2: PADRE-UxS

Uncrewed Systems (UxS) adds complexity to C2 operations

Rapidly evolving mission planning tasking while adhering to various constraints (e.g., Rules of Engagement, Commander's Intent).

PADRE-UxS: Progressive, Assisted, Decision-making, Replanning, and Execution of Uncrewed Systems

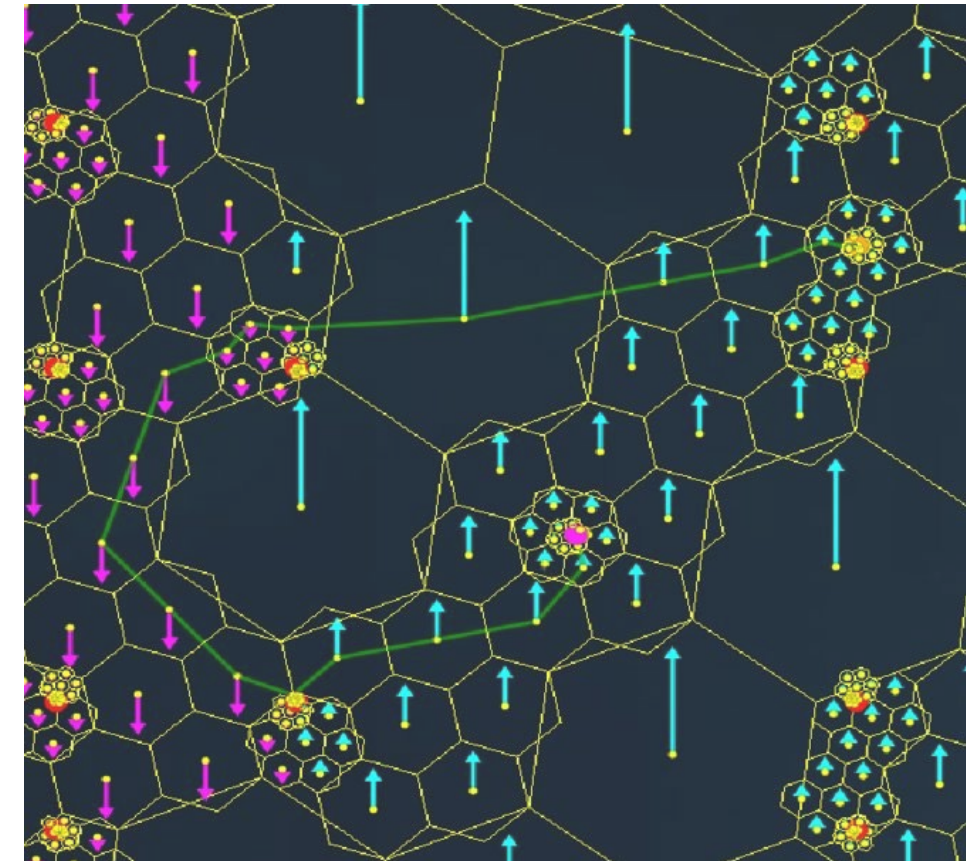**Distribution Statement A.** Approved for public release: distribution is unlimited.

6
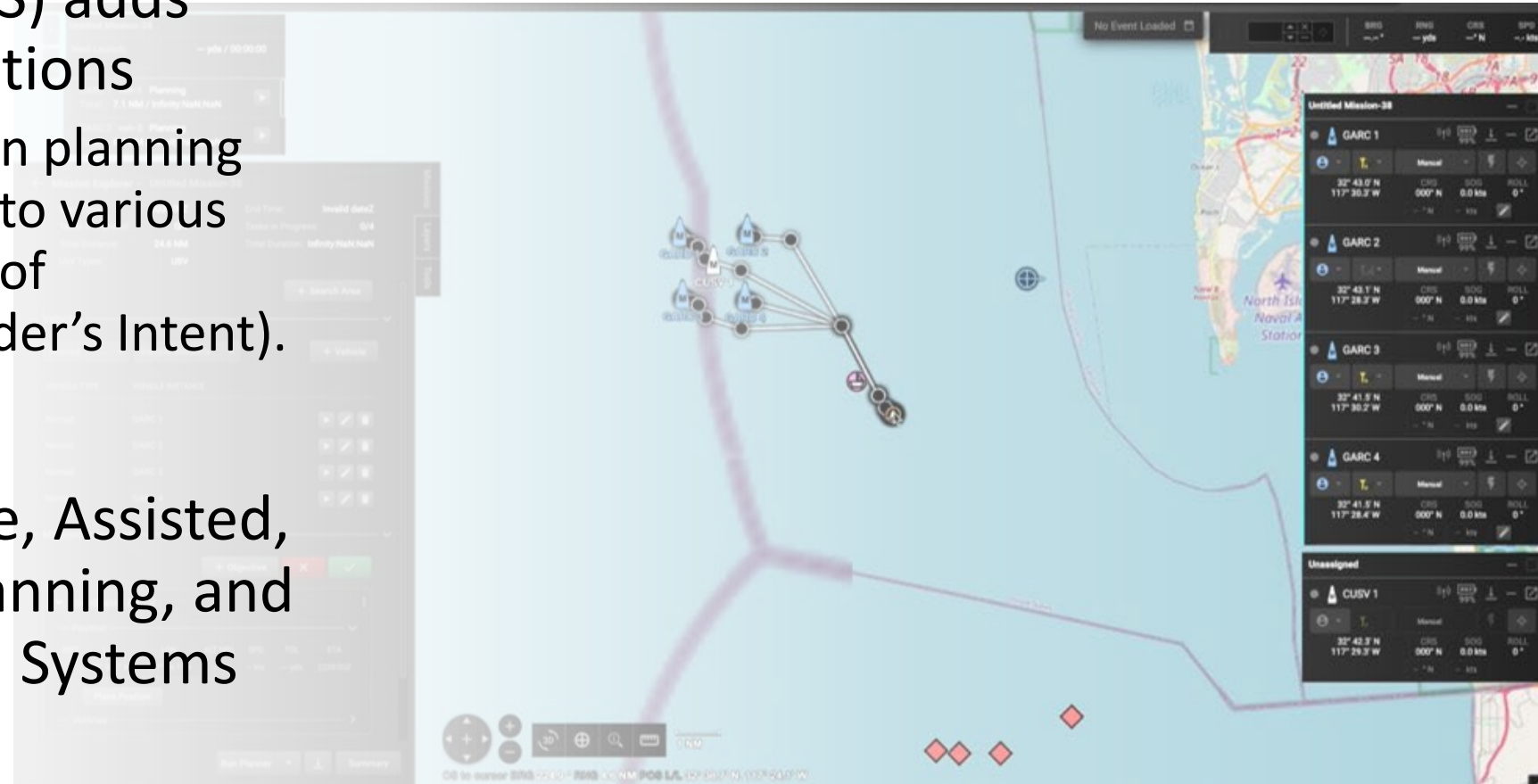
PADRE-UxS Features

ONR Future Naval Capability that identifies constraints, provides alternative COAs and offers contextual explanations

Objective is to improve mission planning and optimality in the face of UxS C2 handoff

Human-AI Teaming (HAT) Experiments

Experimentation is crucial for real-world technology adoption

Numerous examples of Human-Systems Integration challenges across services

BREAKING DEFENSE

| AIR | LAND | NAVAL | SPACE | NETWORKS | MULTI-DOMAIN | CONGRESS | PENTAGON | GLOBAL |

FEATURED: Farnborough 2024 » 2025 Budget » Indo-Pacific »

LAND WARFARE, NETWORKS & DIGITAL WARFARE

Army could be 'wasting' $22B if soldiers don't like IVAS: Inspector General

USNI News

BROWSE BY TOPIC ∨   LATEST

TRENDING TOPICS

China | U.S. Navy | U.S. Marine Corps | Congress | Industry | Personnel | Documents

Navy Set to Restart LCS Deployments this Year, Despite Challenges in Manning, Training

Goals: Increase usability and effectiveness, improve tech adoption, build trust

… allowing for developers to prioritize capabilities and mission sets

… but there are challenges in conducting this research

Challenge 1: Scenario Realism

- Need to balance between scenario realism and experiment complexity
  - Realism increases face validity (low realism leads to easily "solved" scenarios)
  - High realism increases complexity and complications
    - Realistic AORs lead to classification issues
    - Realistic tasking/timelines means experiments would take weeks (i.e., planning)
- Scenarios need to be properly scoped
  - Conversation between SMEs, system developers, and experiment designers

Distribution Statement A. Approved for public release: distribution is unlimited.

9

Scenario Design Strategies

- Create multiple scenarios (open source, CUI, classified)
  - Open source: HA/DR scenario with areas of operation, resources, and incident priorities
  - CUI: Defense of operation base in realistic geographic environment with hypothetical Navy resources (e.g., what may be available in a CSG)
  - Classified: Implementation of actual training scenarios used in Navy schoolhouses

This has benefits with the next challenge, too...

Challenge 2: Participant Makeup

- Quality experimental results depend on reasonable number of participants
- Need to balance between quantity and "quality" of participants
  - For Navy purposes, desire active-duty warfighters with domain expertise
  - Often don't have access to large numbers of the right warfighters

Near-Parallel Experimentation

- Implement near-parallel multiple experimentation tracks
  - Track 1: College students for high-throughput experimentation
  - Track 2: Navy civilians with domain knowledge
  - Track 3: Active-duty sailors with real-world experience

Students

Promising results and strategies

Navy civilians

Strategy validation

Active-duty sailors
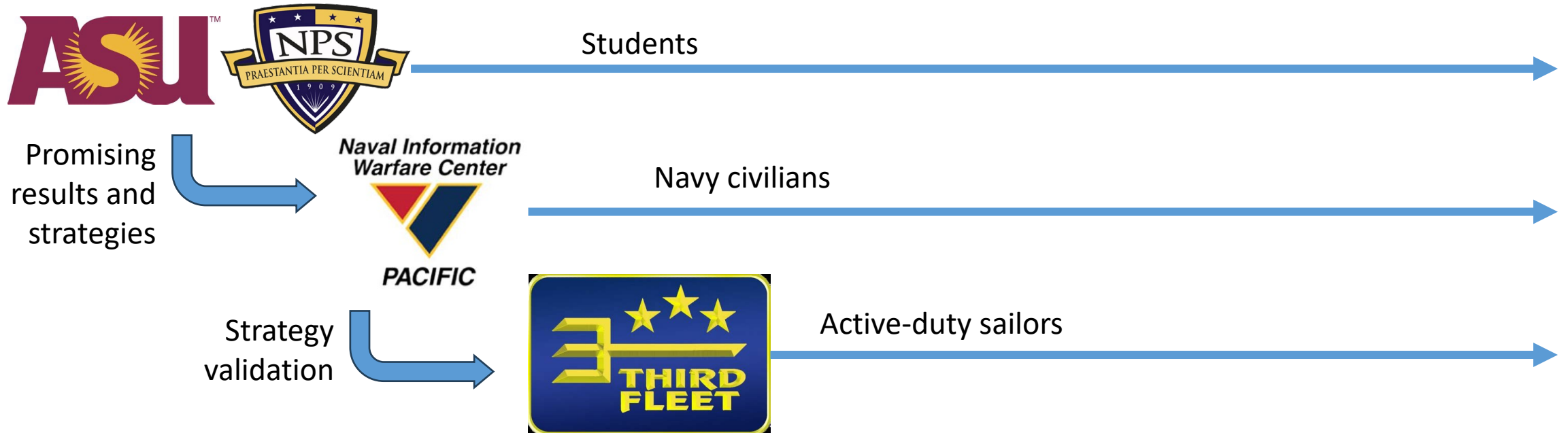
Challenge 3: Measuring Mission Success

- The goal of any system is to help warfighters accomplish their missions

- Mission Engineering enables DoD to "better understand and assess impacts to mission outcomes"
    - How often is this tested and/or verified?

- Approach: Use SMEs to create realistic metrics tailored to scenarios
    - The "best outcome" depends on the scenario and Commander's Intent

Strategic objectives	Operational effectiveness	Resource utilization
Distribute aid	*Time is a critical factor?*	*Distribute all available aid*
Defeat enemy	*Win at all costs?*	*Don't exhaust all ordinance*
Protect territory		

Distribution Statement A. Approved for public release: distribution is unlimited.
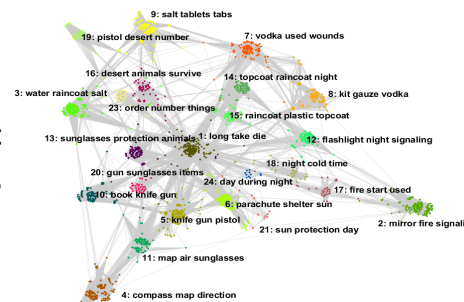
13

Challenge 4: Objective Teamwork Metrics

- Few reliable and objective measures of HAT are in the literature

- Many subjective measures of teamwork exist (often survey–based)

- Objective, data-driven metrics point to quantifiable improvements

- Some work has already been done (Air Force Research Lab, elsewhere)

Data Collection & Analysis

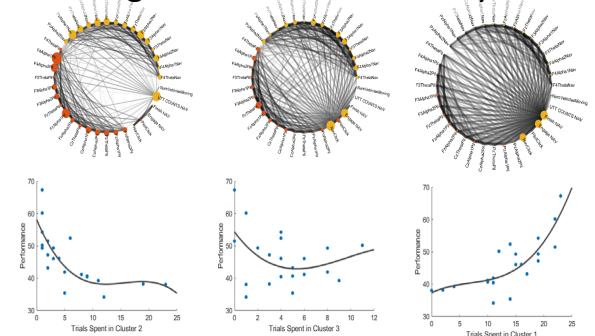- For Show & Tell and PADRE-UxS, the Navy is writing the code
  - Access means screen recording, keylogging, mouse tracking
  - Voice and text communications capture
  - Eyetracking also has strong potential
- Requires complex software and data analytics for meaningful insights
- Will generate Navy-specific results but objective teaming metrics will also have broad applicability

Conclusion

Addressing challenges in HAT experimentation is vital for effective AI deployment

Careful scenario and experiment design increases usefulness of research results

Human interaction with and monitoring of AI systems is here; effective integration is critical

Contact Information

Authors:

Jason H. Wong, Ph.D.
(jason.h.wong.civ@us.navy.mil)

Jeffrey J. Bennett

Luis M. Martinez

Douglas S. Lange, Ph.D.

Naval Information Warfare Center Pacific, San Diego, CA