

Group 1: Evaluation Methodologies

Jay Shively

Adaptive Aero Group

MAR 26, 2024

Goals

Define evaluation methodologies that can be a standard and used across domains.

- Neglect time/service time

- Workload

- Situation Awareness

- Team Performance

- Complexity

Account for CONOPS differences.

- A/C v area commands

Discuss how the data can be used for safety cases/ risk ratios.

Investigate modeling tools.

- E.g, Workload prediction

Group Participants

1. Jay Shively
2. Mark Brooks
3. Meghan Saephan
4. Lance King
5. Husni Idris
6. Mark Shikerman
7. Paul Lee
8. Mehrnaz Sabet
9. Eric Chancey
10. Chad Healy
11. Mike Matessa
12. Jeff Homola
13. Jin Kocsis
14. Tracy Lamb
15. Srikanth Gururajan
16. James Eanes
17. Radhika Bhppatk
18. Keith Hoffler

Progress and Plans

- Co-leads: Jay Shively & Mark Brooks
- Organized and met 3 times
- Developed roster
- Developed outline
- Named section leads
- Collaborative workspace
- Name support teams
- Develop Schedule
- Develop review process

Outline

Editor

Introduction

Goal and Scope

A/C vs. Area Commands

Metrics

Service time/neglect time

Workload

SA

System performance

Team Performance

Complexity

Data for Safety case

Modelling

Vignette Examples

Issues & caveats

Lance King

Jay Shively, Mark Blanks

Jay Shively & Mark Blanks

Mark Blanks

Jay Shively

Lance King

Lance King

Mark Shikerman, Eric Chancey

Mike Matessa

Husni Idris, Srikanth Gurujan

Mehrnaz Sabet, Radhika Bhopatkar, Jay Shively

Husni Idris

All

Neglect time/Service time

- Need to have a consistent definition of
 - Neglect time
 - Service time
 - This is one candidate
- Need a discussion on how to use data
 - Time required/time available – rough workload
 - Concurrent task cost
 - synchronicity

Based on literature,
Previous use

More speculative,
need more work/help



Neglect Time

The time an operator can ignore a autonomous system (aircraft) before needing to intervene.

Mean Neglect time – average time between interventions

Neglect Tolerance – threshold for a continuously degrading system

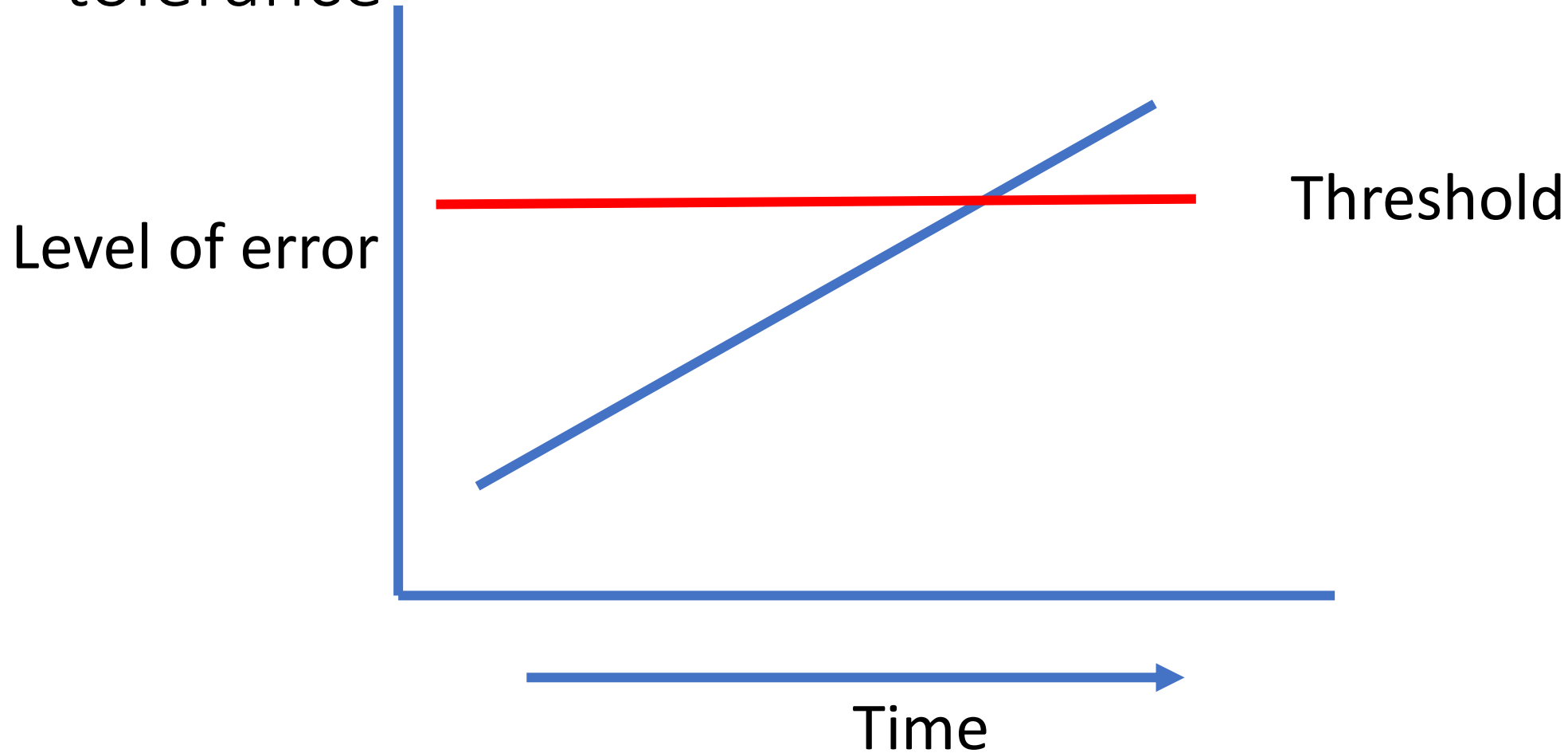
Could come into play, but many of our events are discrete – once hitting threshold continuous can be treated as discrete

Refs:

Steinfeld et al, Common Metrics for Human-Robotic Interaction, Human Robotic Interaction, 2006, Salt Lake City, Utah.

Neglect Tolerance:

Continuous task (heading deviation) or decision tolerance



Example: Mean neglect time

1 Aircraft

	Mission time	5	10	15	20	25	30
Intervention			TFR			A/C warning	

2 interventions in :30 mission $2/:30 = \text{Mean time between interventions} = :15^*$

* Doesn't take into account length of time to resolve intervention or length of time to respond, urgency or importance (group 2 – looking at you)

Service Time

The time required to resolve an intervention.

OODA loop

- Orient (switching time)

- Observe (build SA)

- Decide on Action

- Action to resolve

Crude workload estimate = $\text{time required} / \text{time available}$ OR

Task load

Examples: Time required/time available

1 Aircraft

	Mission time	5	10	15	20	25	30
Intervention			TFR			A/C warning	
Service			:02			:01	

Time required = :03

Time Available = :30 (can also subtract out service time = :27)

Ratio = .1 = 10% (.11, 11%)

Doesn't take into account other a/c.

Examples: Simultaneous interventions

Mission time	5	10	15	20	25	30
A/C 1		:02			:01	
A/C 2	:05				:02	

$TA/TR = :10/:30 = .3 = 30\%$ (subtracting out service time = .5)

Doesn't take into account simultaneous interventions.

Simultaneity Metric

(ok, getting speculative, just one proposal)

Simultaneity = $P(\text{new event} \mid \text{event is occurring})$

Probability of a new event occurring given an event(s) is already occurring.

Example

Mission time	5	10	15	20	25	30
A/C 1			:02			:01
A/C 2		:05				:02

$$TA/TR = :10/:30 = .3 = 30\%$$

$$\text{Prob}(E|E') = 0.5$$

Probability of an event occurring given when another event is occurring.

Need a weight or scaler for concurrent events, in this example arbitrarily set to .5.

Scale total mission

$$\text{Workload} = 30 + 30 (.5) = 45$$

Scale Event

$$\text{Workload} = .23 + .10 + [50\% \text{ scaler for these overlapping events}] .5 = .38$$

Summary

- Neglect & service time definition candidate
- Two possible analyze that can this data can be used for:
 - Rough workload TA/TR
 - Modified for concurrent events
- Much discussion is needed on the analyzes

Caveats/Questions

- Am I completely off base?
- When do interventions start and stop?
- Should any simultaneous interventions automatically be an overload condition?
- Some interventions can be delayed (which I think can be captured in neglect time WRT neglect threshold)
- Some are more important than others and/or more difficult for the operator (i.e., variable workload)
- Interventions can combine in various ways
 - Conflict
 - Facilitate

Data Needed (from flight, simulations, etc.)

- Intervention times
 - Service times
 - Pilot workload
 - System performance
-
- Does enough data exist for meta analysis?