NASA Study: Interruptions in Multi-Vehicle Supervision (DRACO)

(Distracted Remote-operators in Autonomous Concepts of Operation 2024)

Eric Chancey & Mike Politowicz

Contributors: Elliot Biltekoff, Kate Ballard, & Tyler Fettrow

m:N Working Group (10.24.2024)

Transformational Tools and Technologies Project

Innovative solutions through foundational research and cross-cutting tools

autonomous systems

Outline

- Objective / Background
- Method
- Initial Findings
- Conclusion / Next Steps

Objective

- **Objective**: Investigate how interruptions affect situation awareness (SA) in highly automated multi-vehicle operations

- **Scope/Approach**: Participants will be asked to view multi-vehicle operations recordings and maintain “performance-ready” SA. Recordings will be systematically interrupted with a working memory task. SA probes will be randomly administered during experimental trials.

- **Impact**: This research addresses interruption management in a multi-vehicle operational paradigm that generalizes across specific use-cases and applications

Outline

- Objective / Background
- Method
- Initial Findings
- Conclusion / Next Steps

Revolutionary Aviation Mobility (RAM) Sub-Project

Overview

Supports ARMD mission programs by providing a pipeline of solutions and knowledge for foundational challenges in enabling an AAM market.

Enables increasingly autonomous transportation in the UML-4+ timeframe

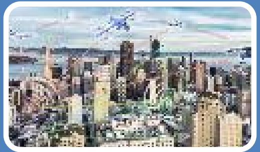




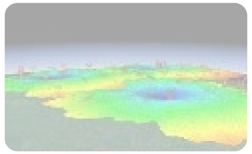

UML-4: Leverage remote supervisory operations
UML-5: Fully-automated aircraft, may still rely on remote supervisors

GOAL **Provide leading edge tools, technologies, and research findings to enable increasingly autonomous AAM transportation in the UML-4+ timeframe**

OBJECTIVES



Enable scalable operations for AAM through development of an m:N operational approval roadmap supported by community coordination and critical tool and technique research (TC/FY27)



Explore and develop airspace management and operations architectures and tools in expectation of increased heterogeneous air traffic

Develop modeling, performance, and control tools & techniques for advanced urban capable aircraft

Images Credit: Transformational Tools & Technologies (https://ntrs.nasa.gov/api/citations/20230001255/downloads/TTT_Project-Overview_19Jan2023_v2.pdf)

Background

Credits: NASA / Graphics

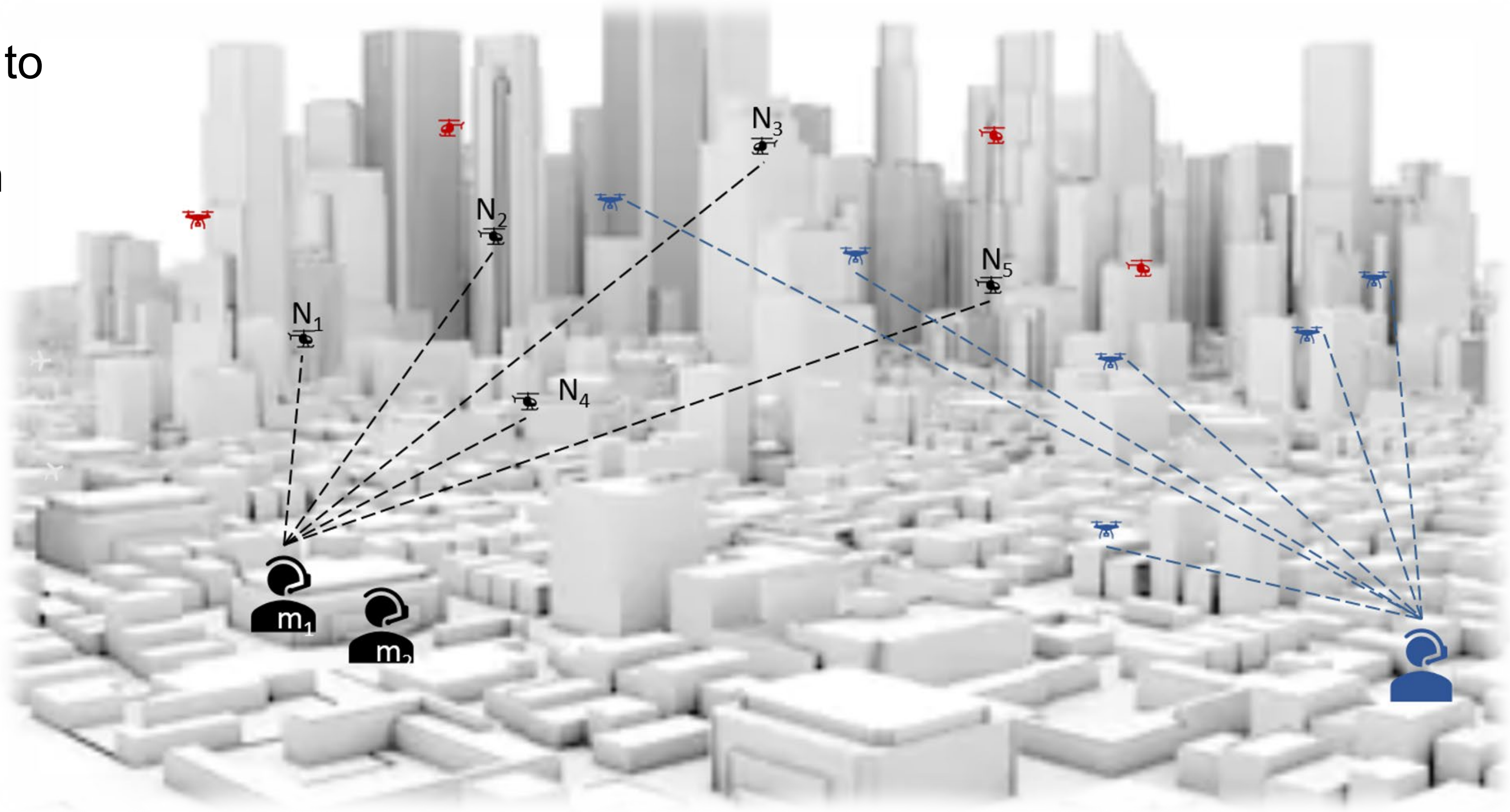
- Advanced Air Mobility (AAM)
  - Urban Air Mobility (UAM) → e.g., Air Taxis
  - Small UAS (sUAS/UTM) → e.g., Package Delivery
  - Regional Air Mobility → e.g., Autonomous Cargo

**UAS = Uncrewed Aerial Systems*
**UTM = UAS Traffic Management*

Background

m:N Operations

- “m-to-N”: Few(er) operators (*m*) managing (many) more vehicles (*N*) (Aubuchon et al., 2022)
- Enabled by vehicle handoffs and **increasingly autonomous systems**
- Evolving operator role (Authority v. Responsibility) (Lacher et al., 2023)
  - Supervisory Control: human operator(s) set intermittent subgoals to a computer, and receive information from a computer, that itself closes an inner control loop through actuators, the task, and feedback sensors (Sheridan, 2021)



Background

- **Multi-Vehicle Operations**
  - Scalability: “what ratio can we expect in these operations?” → “how many vehicles can the human handle?”
    - 4-8 robots (Humann & Pollard, 2019)
    - Cummings et al. (2007): 2 aerial vehicles (Dixon et al., 2005); 4 aerial vehicles (Ruff et al. 2002, 2004; Dunlap, 2006); 5 aerial vehicles (Cummings et al., 2007); 8 aerial vehicles (Lewis et al., 2006); 12 missiles (Cummings & Guerlian, 2007)
  - Answer: “it depends” (Hancock et al., 2007, p. 7)
    - Degree of Automation (Lewis, 2013)
    - Increasing LOA for motion control/navigation and mission management support (Cummings et al., 2007)
    - Zipline use case: 1:24 uncrewed aerial vehicles (Lachter et al., 2021)
      - “14 Code of Federal Regulations Section 1.1 defines operational control with respect to a flight as the exercising of authority over initiating, conducting, or terminating a flight.”
    - “Increasingly Autonomous” (NRC, 2014); Autonomous systems “…independently determine a new course of action in the absence of a predefined plan to accomplish a goal based on its knowledge and understanding of its operational environment and situation. Having the ability and authority to make decisions independently and self-sufficiently.”
      - The responsible human is required to monitor the system that has the authority (Pritchett & Bhattacharyya, 2016).

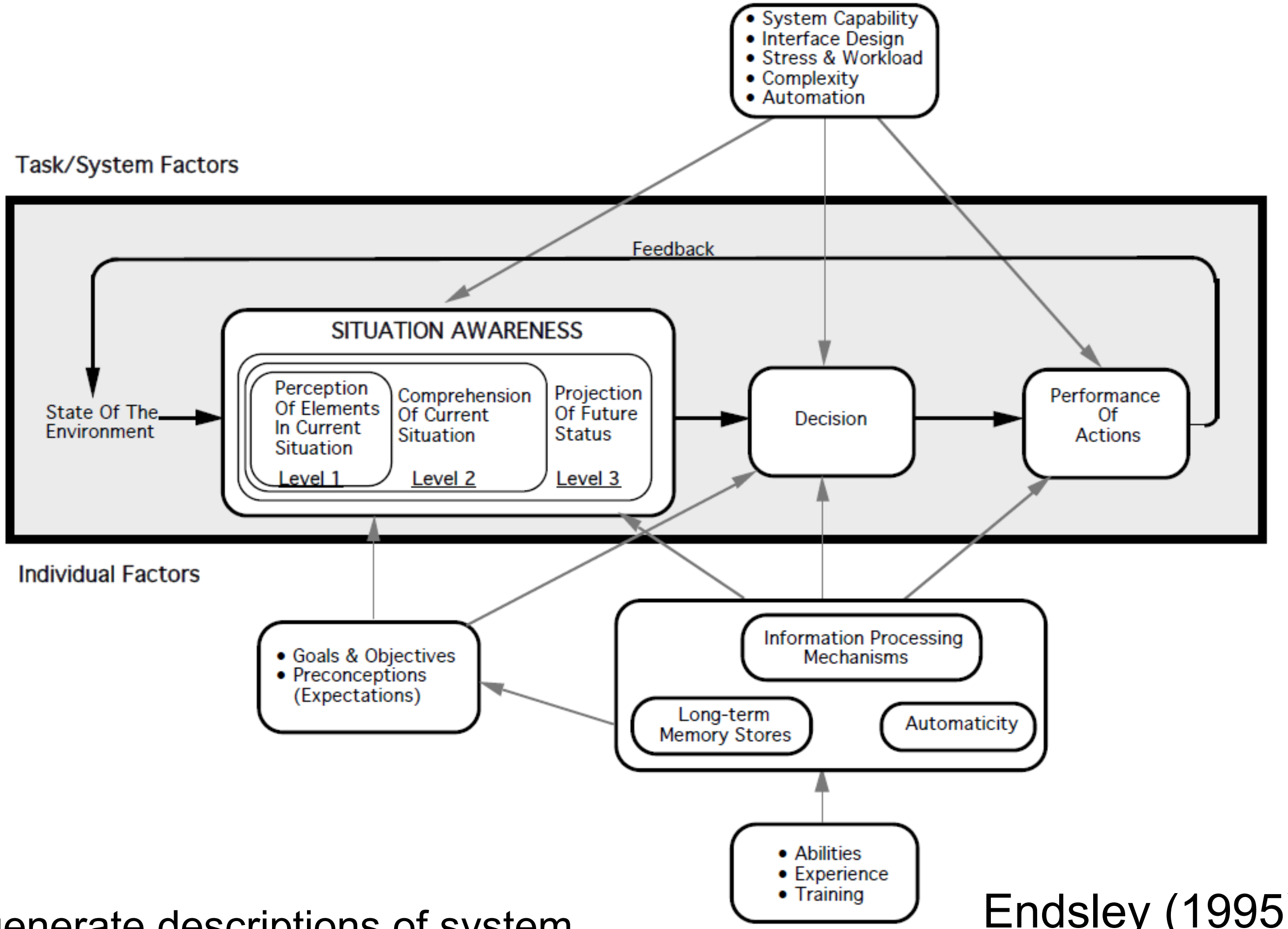
Background

- Multi-vehicle supervisor
  - Pilot Monitoring / Pilot Flying distinction (Billman et al., 2020, and Mumaw et al., 2020)
    - Pilot Monitoring Responsibilities:
      - Responsible for monitoring the current and projected flight path and energy of the aircraft at all times
      - Supports the pilot flying (cf. uncrewed aerial vehicle), staying aware of aircraft state, system status, and ATC instructions and clearances
      - Calls out any perceived or potential deviations from the intended flight path, and intervenes if necessary
    - Monitoring:
      - sense-making, systematic observation and interpretation of the current state of the airplane and its operational environment
      - Significantly overlaps with situation awareness (SA; Endsley, 2015); the product of monitoring and often how monitoring effectiveness is defined (Mumaw et al., 2020)
    - Threats to monitoring performance
      - Interruptions: Task performance is not completely under the control of the flight crew, and interruptions can take attention away from monitoring

Background

Situation Awareness (SA): Perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future (Endsley, 1995, p. 36)

- Level 1 SA (Perception): Detection of relevant elements within an environment.
- Level 2 SA (Comprehension): Integrating the acquired perceptual information to gain an understanding of the importance of those L1 elements in the context of the decision maker's goals and task.
- Level 3 SA (Projection): Projecting the near-term future situation and its impact based on the information gathered (L1 & L2 SA).



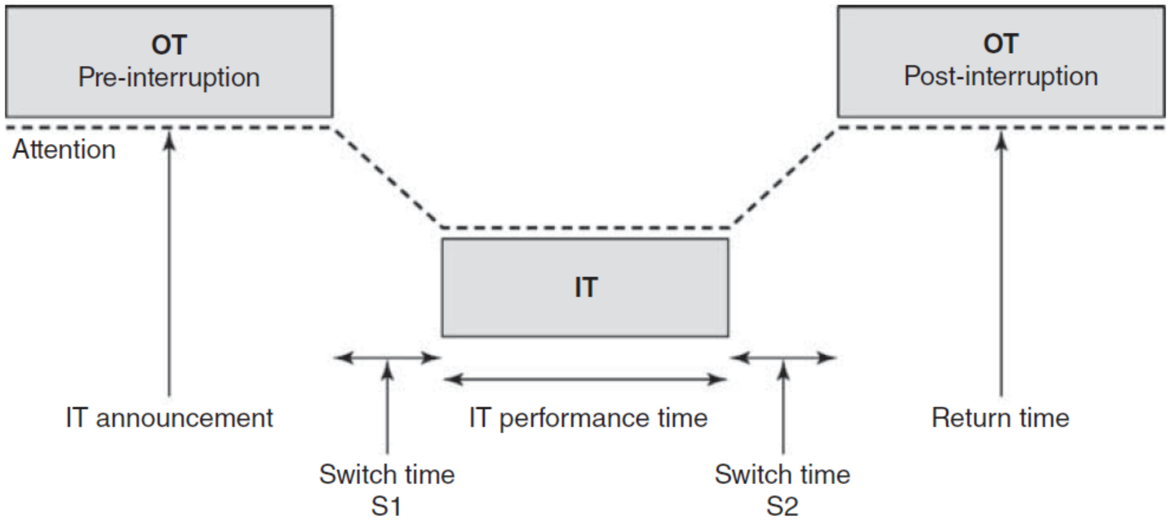
Endsley (1995)

Not necessarily linear development!

Mental Models: the mechanisms whereby humans can generate descriptions of system purpose and form, explanations of system functioning and observed system states, and predictions of future states (Rouse & Morris, 1985, p. 7; see also Moray, 1999)

Background

- Interruption management: the detection, interpretation, and integration of interruptions with ongoing task performance (Latorella, 1996); the process of coordinating abrupt changes in people's activities (McFarlane, 1997).
  - Ongoing Task (OT): Supervisory control of multi-vehicle operations
  - Interruption Task (IT): A task that causes a stoppage or break in the continuity of an OT
- Attention is switched at S1 from the OT to the IT and full attention is then focused on the IT until it is either completed or a person decides to switch back to the OT.



Wickens & Carswell (2021)

Background

- Theoretical: Effects of Interruptions on SA
  - Memory for Goals (Altmann & Trafton, 2002): Memory trace for the OT status decays while attention is on the IT; OT information in working memory decay of Level 1 SA elements will increase with interruption length (Wickens et al., 2022)
    - May not hold for experts, which have existing mental models and pattern matching for prototypical events (Endsley, 1995; Endlsey, 2015)
  - Change blindness (Rensink, 2001, 2002): Failure to notice something is different from what it was
    - Laboratory settings: Blink (O'Regan et al., 2000), blank screen (Rensink, 2002), saccade from change location (Stelzer & Wickens, 2006)
    - Applied settings: street signs (Martens, 2011), pilots failing to notice flight mode indicator light (Sarter et al., 2007), (possibly) UAS operations (Chancey et al., 2023; Politowicz et al., 2023)
- Empirical Evidence: Effects of Interruptions on SA
  - Lower subjective SA for students in an interruption condition vs. non-interruption condition (van der Kleij et al., 2018)
  - Interruptions led to less accurate objective SA scores for student participants (Loft et all., 2015)
  - Student participants detected more changes (Level 1 SA) when not interrupted (Perry & Scerbo, 2019)
  - Multi-Vehicle: Subset of expert participants that focused more on a chat message box (interruption/SA probe) had poorer performance, but no significant differences were observed on SA query accuracy (Cummings, 2004)
  - Multi-Vehicle: Experiment 1 – Although observed change blindness, no significant effect on SA accuracy; Experiment 2- Automation (adaptive v. static) groups, change detection followed a similar pattern as SA accuracy (though not statistically significant)
  - Multi-Vehicle: Students watched a video of 1, 2, or 3 videos and half were interrupted. Interruption group had significantly lower Level 2 SA scores than non-interrupted group, yet Level 1 SA scores were not significantly affected by the interruption (Gartenberg et al., 2014)

Background

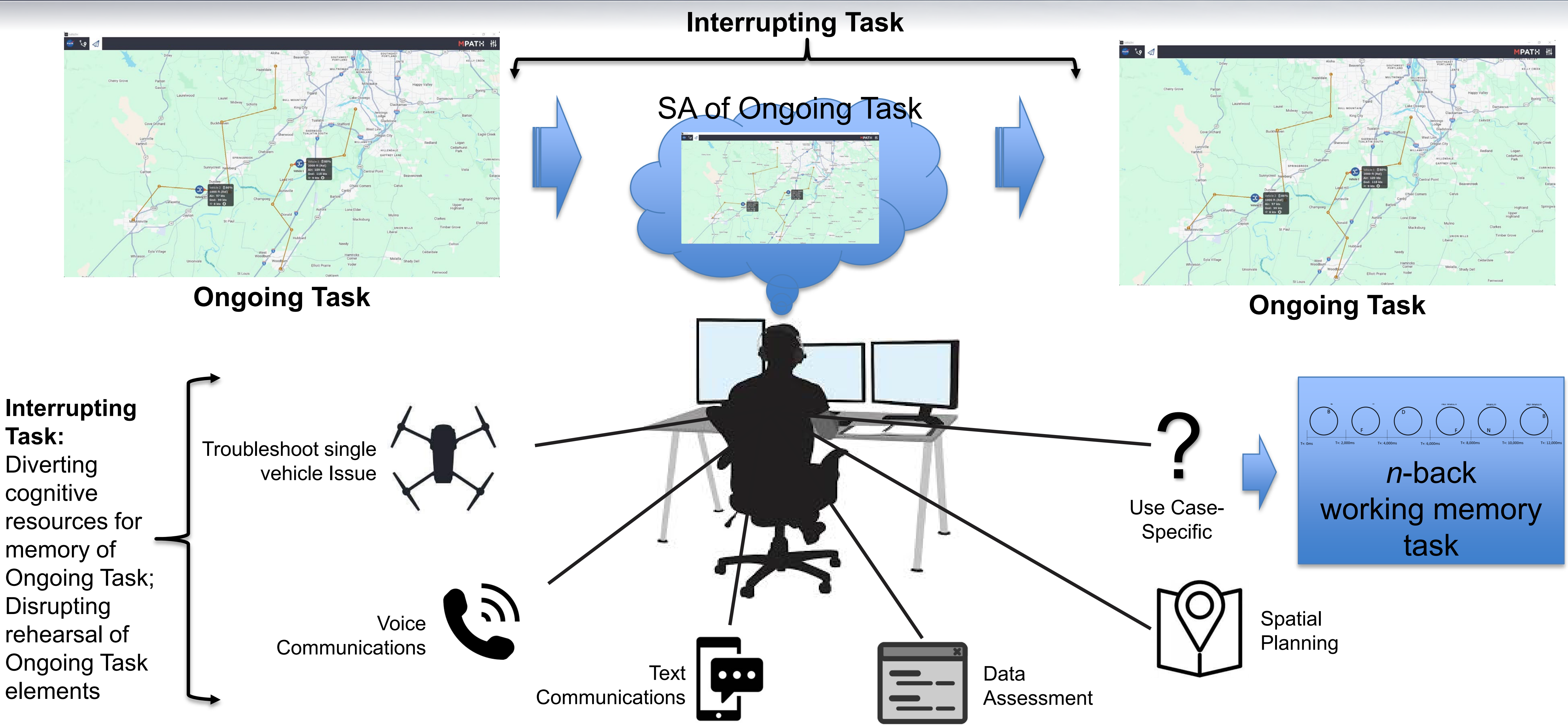


Image Credit: Getty Images (https://images.app.goo.gl/qdNbrHydhyN5po5c6)

Hypotheses

- ***Hypothesis 1: Interruption Length and Mental Model Accuracy on Level 1 SA***
  - Main effect of interruption length on Level 1 SA, where longer interruptions lead to more inaccurate responses
    - MFG: Decay of Level 1 SA (Altmann & Trafton, 2002)
    - Change blindness: Failing to detect changes from the interruption (Perry & Scerbo, 2019)
  - Moderated effect: Offset by accuracy of mental model (Endsley, 1995)
- ***Hypothesis 2: Interruption Length on Level 2 SA***
  - Main effect of interruption length on Level 2 SA, where longer interruptions lead to more inaccurate responses. Mental model accuracy will moderate this effect
    - Monitoring is characterized as sensemaking (Billman, Mumaw, & Feary, 2020): Forming Level 2 SA from Level 1 SA through the process of gathering and synthesizing information (Endsley, 2015)
- ***Hypothesis 3: Interruption Frequency on SA***
  - Main effect of interruption frequency on SA responses, where more frequent interruptions will lead to lower SA scores across all 3 levels of SA
    - More interruptions means more tasks → More tasks means higher workload → high workload is negatively correlated with SA (Onnasch et al., 2014)
- ***Hypothesis 4: Number of vehicles on SA***
  - More vehicles will lead to lower Level 1 and Level 2 SA (Gartenberg et al., 2014)

Research Question

- ***What are the effects of interruptions versus no interruptions on SA?***
  - Although SA can decrease under high workload due to competition of attentional resources (cf. H3), it can also decrease under low workload due to boredom and complacency (Rodgers et al., 2000)

Outline

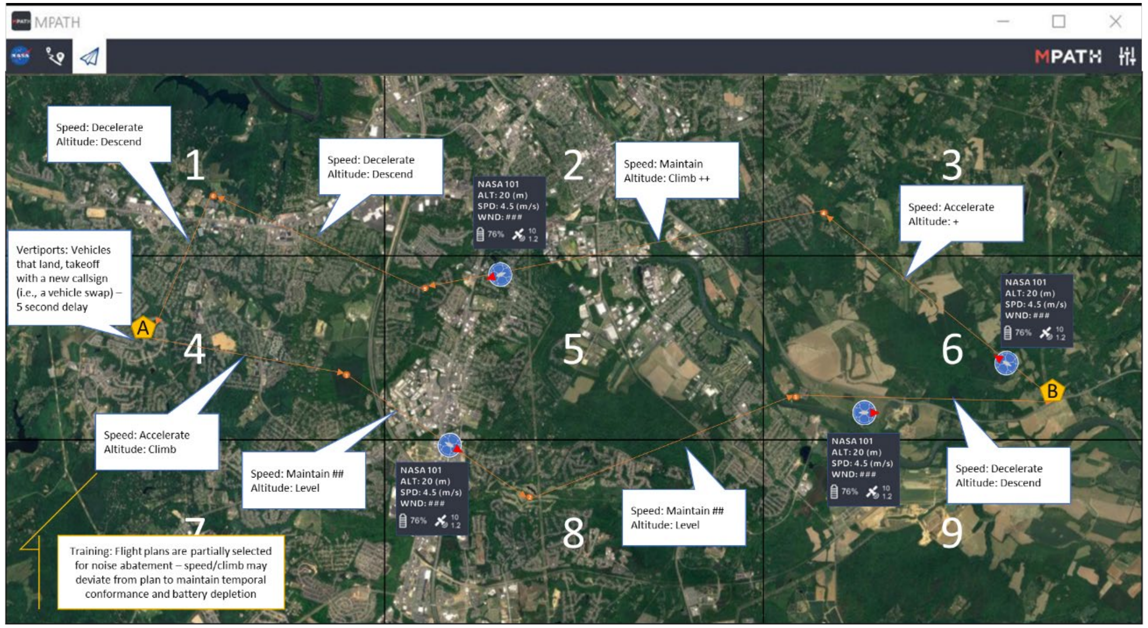
- Objective / Background
- Method
- Initial Findings
- Conclusion / Next Steps

Method

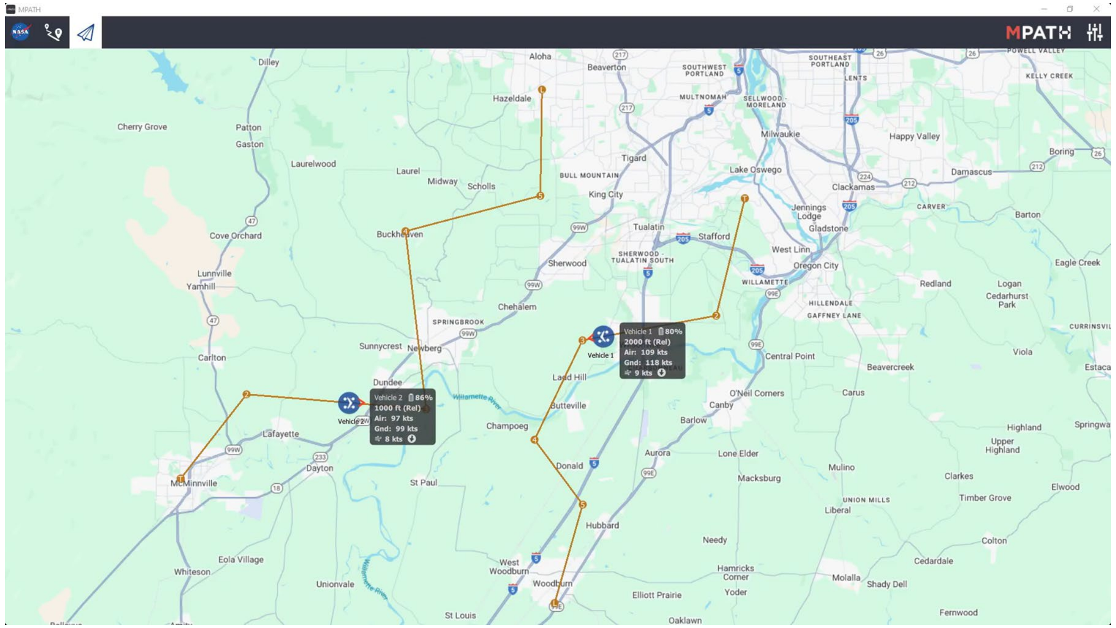
- Participants: *n* = 30 (currently at *n* = 8)
  - Sample Size and Power Estimates (Design Expert 12; R)
    - 25 participants required for approximately 80% power to detect a 15% difference in SA between groups
    - *n* = 30 to account for no-shows, technical issues, and facilitate treatment of TLX data as interval (Bolton et al., 2022)
  - Population Sample: NASA Langley civil servant and contractor personnel and advertised to local population that possess requisite GCSO and/or piloting experience
    - Ground Control Station Operators
      - Part 107 Certificate with drone operation experience
      - Previous UAM studies sampling from this population: Chandarana et al., 2022; Glabb et al., 2022; McSwain, 2023; McSwain et al., 2024; Petty et al., 2024
      - Compare to Wisk's 'multi-vehicle supervisors' (Boeing, 2023)
    - Pilots
      - Held at a minimum a private pilot license
      - Previous multi-UAV studies (Fern & Shively, 2009)

Method

- OT: Flight Path Monitoring Task
  - Video playback of multi-vehicle operations in MPATH
  - Participant will act as a UAM multi-vehicle supervisor for a homogonous set of electric vertical takeoff and landing vehicles (eVTOLs; cf. Wisk/Boeing, 2023)
    - Responsibility: Maintain performance-ready SA over the operational environment and individual highly automated vehicles



Storyboard for Four Vehicles

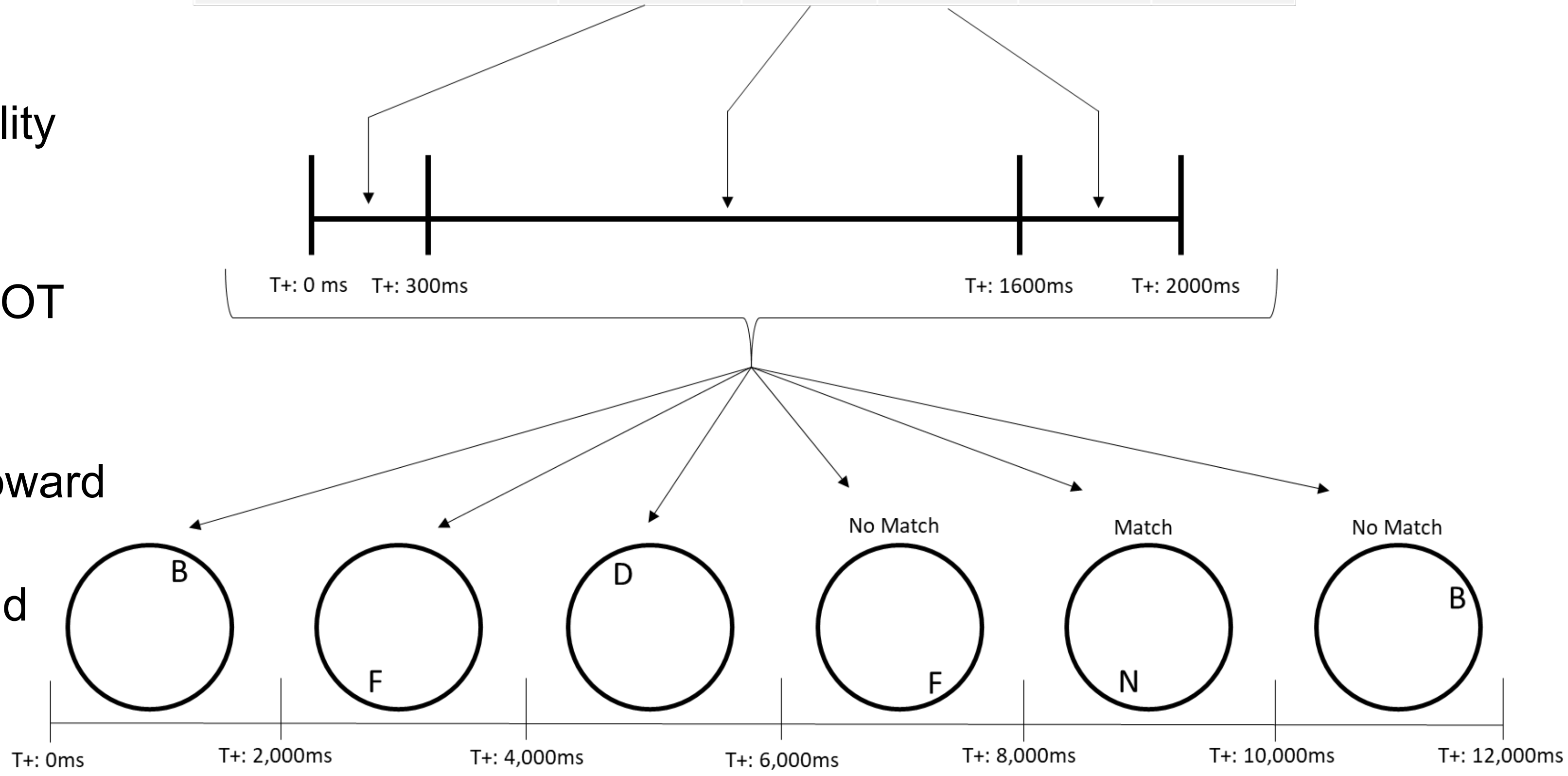


DRACO / MPATH Display

Method

- IT: Visiospatial *n*-back task
  - Participants determine where the spatial location of a presented stimulus matches the location of a stimulus presented *n* trials back
  - *n* = 3 for the following reasons:
    - Mitigate effects of cognitive ability on task
    - Sufficient loading of working memory to potential impact on OT performance
    - Provide steady load without negatively skewing accuracy toward higher values
    - Loads both location (L1 SA) and relational aspects (L2 SA) of working memory

	Presentation duration	Response window	Intertrial delay	Trial Duration	Match frequency
Hockey & Geffen (2004) Parameters	200ms	1300ms	300ms	1800ms	33%
Change	+100ms	~	+100ms	+200ms	~
Current Study Parameters	300ms	1300ms	400ms	2000ms	33%



Experimental Approach and Scope

Fully Within Subjects Design

Number of Vehicles: 2, 4 (Including a no-interruption control condition)		IT Rate per 630 s trial	
		Low: 9	High: 18
IT Length	Short: 14 s	ρ (20%) = 126 s	ρ (40%) = 252 s
	Long: 28 s	ρ (40%) = 252 s	ρ (80%) = 504 s

- IT is defined by *utilization rate* (ρ), or percentage of time the operator is working on the IT, where ρ = IT/OT+IT
  - Cummings and Guerlian (2007) suggest a significant drop in performance above ρ = 70%
- Number of Vehicles: 2, 4 (approximates Wisk's proposal of 3 vehicles)
  - Five 2-vehicle videos
  - Five 4-vehicle videos
- At runtime, participants will see a randomly generated combination of video and interruption characteristic profile

Pilot-test: Context-specific number and many more vehicles have been used in other settings (Zipline, 24 vehicles; 12 missiles in Cummings & Guerlian, 2007)

Method

- Dependent Measures:
  - OT Performance: Situation Awareness, SAGAT
  - IT Performance (*n*-Back): Accuracy, Response Time (RT)
  - Mental Models: Pathfinder Network Analysis
  - Workload: NASA-Task Load Index (TLX)
  - Subjective SA: Situation Awareness Rating Technique
  - Eye Tracking

Map Image Credit : GoogleMaps

Proposed Analyses

- A 3 × 2 × 2 factorial will be analyzed without the control runs followed by an analysis using change from control data.
  - Linear or generalized-linear models will be applied depending on the distribution of the outcome
  - Correlation will be incorporated by estimating standard errors with cluster-robust standard errors, or *sandwich estimators*
  - Effect sizes and Bayes Factors will be reported
- SAGAT-specific considerations
  - Because items evaluated individuall, multiple regression models → multiple comparisiosn run the risk of Type I error. Mitigated using Stepdown minP method, which uses bootstrapping to control the total error rate while accounting for correlations among *p* values.
  - Logistic regression will be used to test individual queries, because outcomes are expressed as proprotions
  - Summary across the three SA levels, using a meta-analysis method conducted on effect sizes of factors of interest (i.e., effect size and margin of error can be compared across SAGAT items)

Initial Findings

Next Steps

- Data collection complete (November 2024)
- Primary data analysis complete (January 2025)
- Begin submitting writeups (March 2025)
  - NASA Technical Memorandum (Test Plan, Research Material)
  - Manuscript(s)
    - Major findings (SAGAT)
    - Eye tracking

Contact Info:

Eric Chancey – eric.t.chancey@nasa.gov

Mike Politowicz – michael.s.politowicz@nasa.gov

Extra Slides

Method

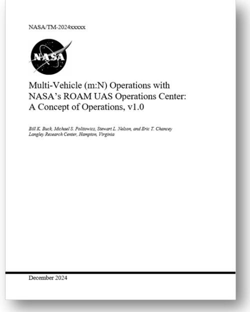
Schedule	Event
0930-0945 (15 min)	Informed Consent, Demographics
0945-1100 (1 hr 15 min)	Training and Practice Session
1100-1200 (1 hr)	Scenario 1 (15 min), Scenario 2 (15 min), Scenario 3 (15 min), Scenario 4 (15 min)
1200-1300 (1 hr)	Lunch Break
1300-1400 (1 hr)	Scenario 5 (15 min), Scenario 6 (15 min), Scenario 7 (15 min), Scenario 8 (15 min)
1400-1415 (15 min)	Break
1415-1445 (30 min)	Scenario 9 (15 min), Scenario 10 (15 min)
1445-1500 (15 min)	Debrief and Dismissal

Randomized Conditions

Method

- OT Performance: Situation Awareness
  - Situation Awareness Global Assessment Technique (SAGAT; Endsley, 1995)
    - Simulated environment is frozen at randomly selected times → Participant queried on their perceptions of the situation at that time, and then scored based on the accuracy of the response in comparison to ground truth
      - Also including a confidence in each response (cf. QUASA method, McGuinness, 2004)
    - Queries represent all 3 levels of SA; Queries scored individually rather than compiled (however, see planned analyses)
    - Freezes will not occur within the first 3 minutes of the beginning of the scenario
    - Queries randomly sampled and presented to the participant
    - Three 75 s probes: 1) between 180-360s, 2) 360-540 s, 3) conclusion of the 10-minute scenario (total trial duration is 13 min 30 s)

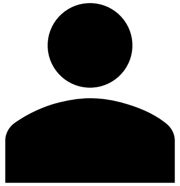
m:N ConOps


NASA/TM-2024xxxxx

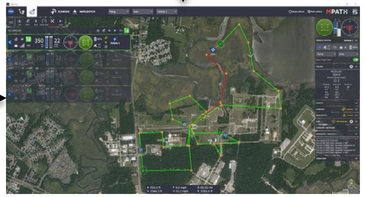
Multi-Vehicle (m:N) Operations with NASA's ROAM UAS Operations Center: A Concept of Operations, v1.0

Bill K. Buck, Michael S. Politowicz, Stewart L. Nelson, and Eric T. Chancey
Langley Research Center, Hampton, Virginia

December 2024

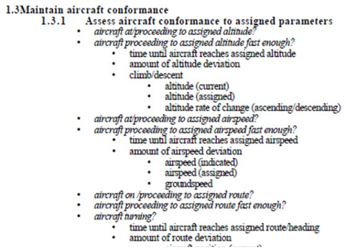


SME Discussion



Sim Constraints

Modified Goal-Directed Task Analysis


1.3 Maintain aircraft conformance

1.3.1 Assess aircraft conformance to assigned parameters

- *aircraft at/proceeding to assigned altitude?*
- *aircraft proceeding to assigned altitude fast enough?*
  - time until aircraft reaches assigned altitude
  - amount of altitude deviation
  - climb/descent
    - altitude (current)
    - altitude (assigned)
    - altitude rate of change (ascending/descending)
- *aircraft at/proceeding to assigned airspeed?*
- *aircraft proceeding to assigned airspeed fast enough?*
  - time until aircraft reaches assigned airspeed
  - amount of airspeed deviation
    - airspeed (indicated)
    - airspeed (assigned)
    - groundspeed
- *aircraft on /proceeding to assigned route?*
- *aircraft proceeding to assigned route fast enough?*
- *aircraft turning?*
  - time until aircraft reaches assigned route/heading
  - amount of route deviation

Endsley & Rogers (1994)

SAGAT Probes

ID	Query	Query Ref	SA Level	GCSO Task	DRACO
ATC1	Enter the location of all aircraft		1	Y	Y
ATC2	Enter aircraft callsign	1	1	Y	Y
ATC3	Enter aircraft altitude	1	1	Y	Y
ATC4	Enter aircraft groundspeed	1	1	Y	Y
ATC5	Enter aircraft heading	1	1	Y	Y
ATC6	Enter aircraft's next sector	1	2	Y	Y
ATC7	Enter aircraft's current direction of change (Altitude Change: Climbing, descending, level; Turn: right turn, left turn, straight)	1	1	Y	Y
ATC8	Enter the aircraft type	1	1	N	N

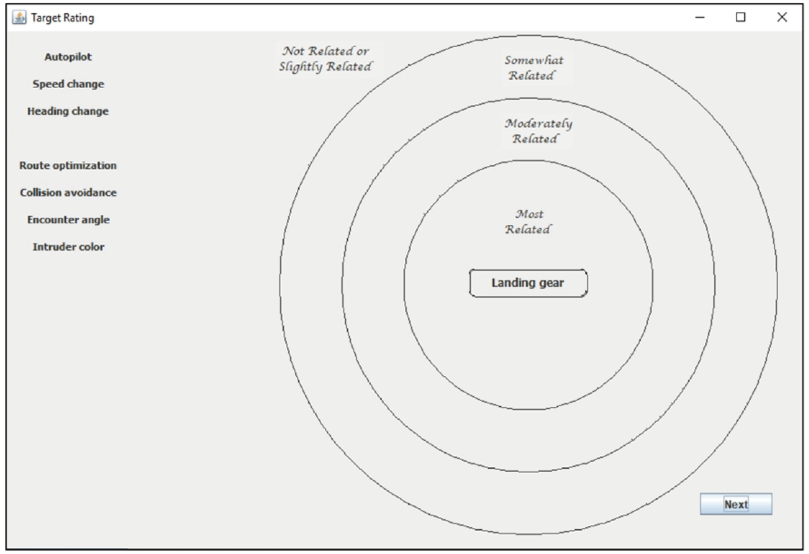
Adapted from Endsley (2000)

Method

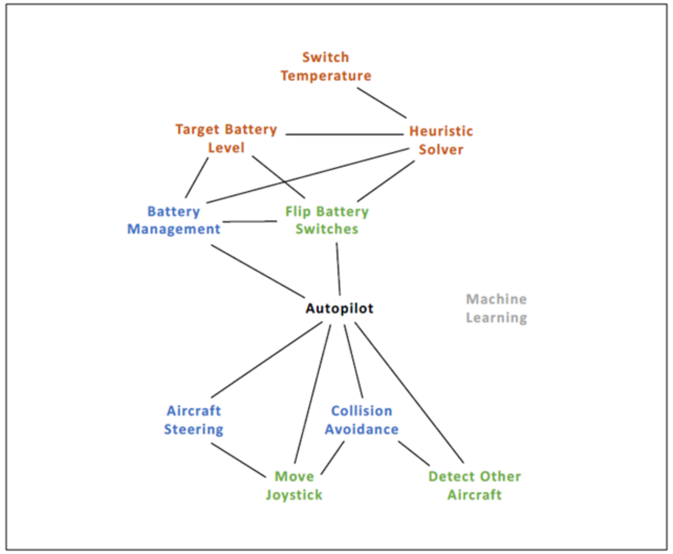
- IT Performance: *n*-Back
  - Accuracy: Number of times the stimulus is correctly identified as matching/not matching the position of the stimulus presented 3 trials previously divided by the total number of stimuli presented
  - Response Time (RT): Aggregate mean for only the accurate trials
    - Outlier-labeling rule (2.2 multiplier) for RTs significantly below average sample performance (Hoaglin & Iglewicz, 1987)

Method

- Mental Models: Pathfinder Network Analysis
  - Statistical technique to visualize networks and analyze relationships among "concepts" based on proximity data (Schvaneveldt et al., 1989); used extensively to quantify mental models in HCI studies (Cooke et al., 1996)
    - Nodes: predefined keywords or concepts
    - Links: represent the relatedness between two nodes
    - *C* Statistic: Quantitative comparison between separate networks (0 = not related, 1 = strongly related), which is a measure of shared links for matching nodes



Target Rating Method User Interface (JTarget Software)



Graphical representation of Pathfinder Network

Method

- Workload: NASA Task Load Index (TLX)
  - Subjective scale that assesses perceived workload on six subscales (Hart & Staveland, 1988)
    - Mental Demand
    - Physical Demand
    - Temporal Demand
    - Performance
    - Effort
    - Frustration

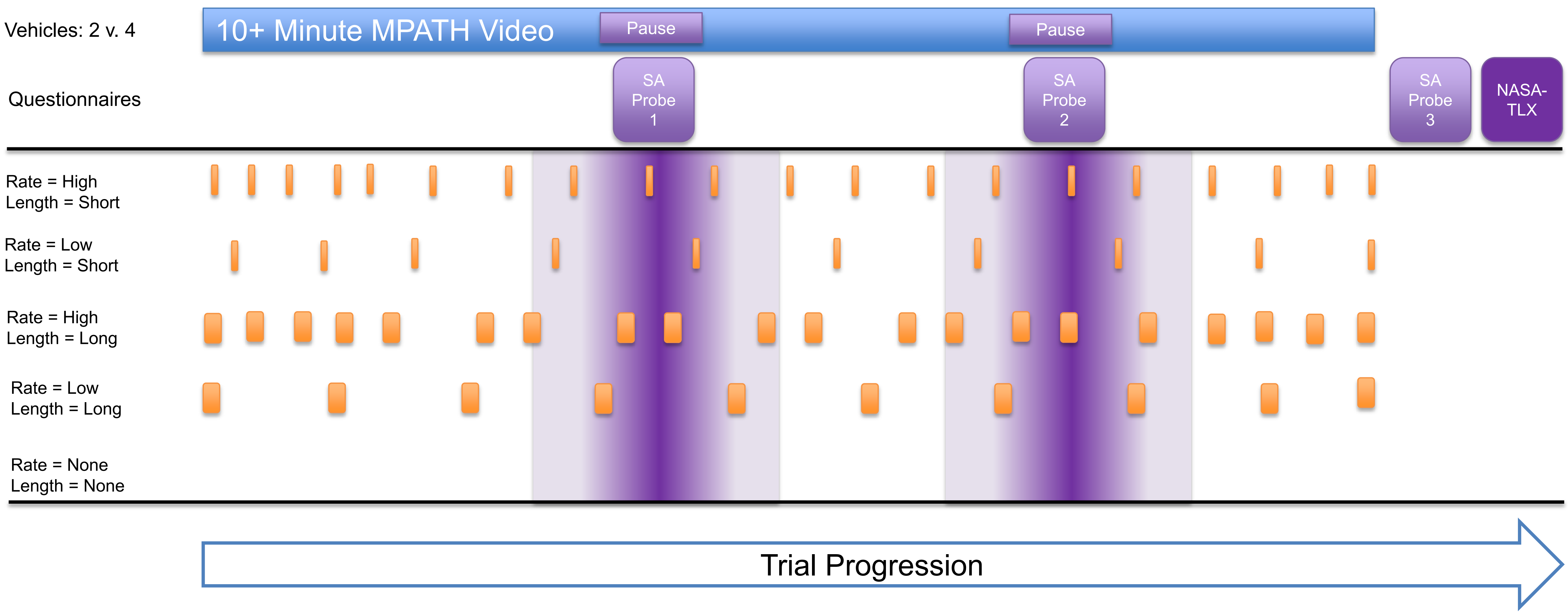
Mental Demand: How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?

Low -- High

Method

Ongoing Task
Interruption Task
Ongoing Task
Interruption Task
Ongoing Task
Interruption Task
SA Probe
Task Progression

Method



Vehicles: 2 v. 4
10+ Minute MPATH Video
Pause
Pause
Questionnaires
SA Probe 1
SA Probe 2
SA Probe 3
NASA-TLX
Rate = High Length = Short
Rate = Low Length = Short
Rate = High Length = Long
Rate = Low Length = Long
Rate = None Length = None
Trial Progression