QUANTITATIVE SCORING SYSTEM TO ASSESS PERFORMANCE IN EXPERIMENTAL ENVIRONMENTS

14th International Conference on Advances in Computer-Human Interactions

Ron Becker, Sophie-Marie Stasch, Alina Schmitz-Hübsch, Sven Fuchs

Ron Becker | Fraunhofer FKIE

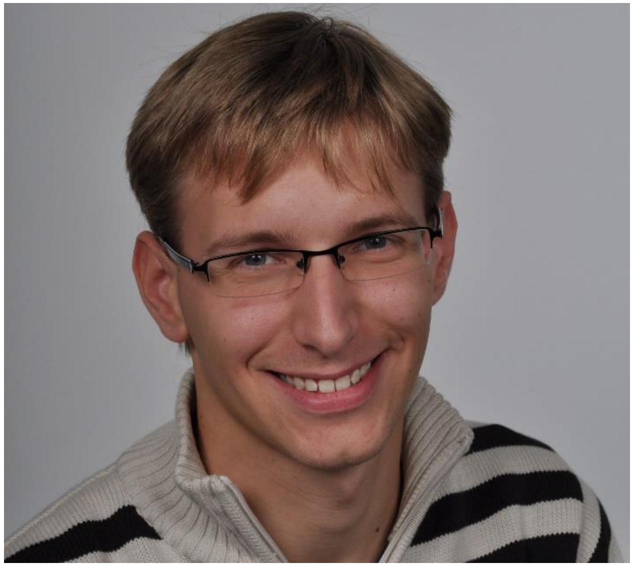
ron.becker@fkie.fraunhofer.de

Fraunhofer
FKIE

IARIA

Ron Becker

- B. Sc. in Computer Science
- Technician at Fraunhofer FKIE, Germany (since 2009)
- Research Fields
  - Generic C2-Systems
  - Decision Support for Safety-Critical-Systems
  - Adaptive Human-Machine-Interaction
  - Software Architecture and Design Patterns



Fraunhofer FKIE

Performance Measurement

Problem: Measuring human performance in complex experimental tasks can be challenging.

Solution: Two approaches are available (generally):
Qualitative or quantitative performance measurement.

Qualitative	Quantitative
results tend to be subjective	objectify performance assessment
results difficult to verify	enable the use of inferential statistics (e.g. significance testing)[1 2]

[1] P. A. Ochieng, "An analysis of the strengths and limitation of qualitative and quantitative research paradigms", Problems of Education in the 21st Century, vol. 13, pp. 13-18, 2009.

[2] A. Queirós, D. Faria, and F. Almeida, "Strengths and limitations of qualitative and quantitative research methods", European Journal of Education Studies, vol. 3, no. 9, pp 369-387, 2017.

Fraunhofer FKIE

Performance Measurement

Problem: Measuring human performance in complex experimental tasks can be challenging.

Solution: Two approaches are available (generally):
Qualitative or **quantitative** performance measurement.

Predefined metrics that can be calculated in an automatic manner, resulting in a numeric value, such as a **score**.

How to begin?

Signal Detection Theory (SDT)

- Originating from signal detection in psychophysics[1], the theory successfully explains phenomena in the study of
  - visual search[2], recognition memory[3], decision making in supervisory control[4], air combat training[5], essay grading[6], social anxiety[7].

[1] D. M. Green and J. A. Swets, "Signal detection theory and psychophysics", Wiley, 1966.

[2] P. Verghese, "Visual search and attention: A signal detection theory approach" Neuron,vol. 31, no. 4, pp. 523-535, 2001.

[3] J. T. Wixted, "Dual-process theory and signal-detection theory of recognition memory" Psychological review, vol. 114, no. 1, p. 152, 2007.

[4] A. Bisseret, "Application of signal detection theory to decision making in supervisory control The effect of the operator's experience", Ergonomics, vol. 24, no. 2, pp. 81-94, 1981.

[5] J. L. Eubanks and P. R. Killeen "An application of signal detection theory to air combat training"; Human factors, vol. 25, no. 4, pp. 449-456, 1983.

[6] L. T. DeCarlo, "A model of rater behavior in essay grading based on signal detection theory" Journal of Educational Measurement, vol. 42, no. 1, pp. 53-76, 2005.

[7] L Yoon, J. W. Yang, S. C. Chong, and K. J. Oh, "Perceptual sensitivity and response bias in social anxiety: an application of signal detection theory" Cognitive therapy and research, vol. 38, no. 5, pp. 551-558, 2014.

Fraunhofer
FKIE

Signal Detection Theory (SDT)

		Signal	
		Present	**Absent**
Response	**Present**	Hit	False Alarm
	Absent	Miss	Correct Rejection

Signal Detection Theory (SDT)

		Signal	
		Present	**Absent**
Response	**Present**	Positive Score **2+1 = 3** Negative Score **2** Max Score **4+1 = 5** Total Score **1** Performance **20%**	Positive Score **2** Negative Score **2+1 = 3** Max Score **4** Total Score **-1** Performance **-25%**
	Absent	Positive Score **2** Negative Score **2+1 = 3** Max Score **4+1 = 5** Total Score **-1** Performance **-20%**	Positive Score **2+1 = 3** Negative Score **2** Max Score **4** Total Score **1** Performance **25%**

Positive Score
=
∑ (Hits + Correct Rejections)

Negative Score
=
∑ (Misses + False Alarms)

Max Score
=
∑ Present Signals

Total Score
=
Positive Score - Negative Score

Performance
=
Total Score / Max Score

Scoring Mechanism

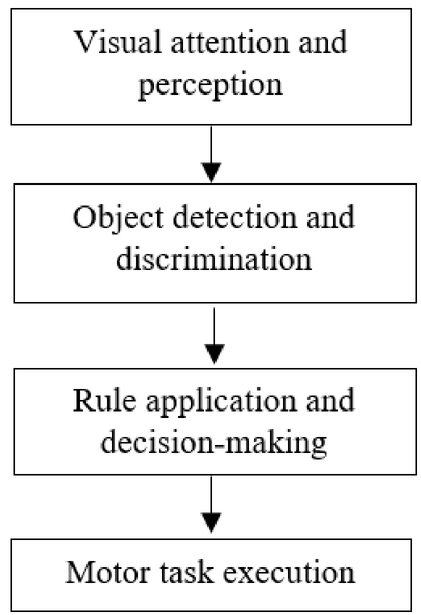
(base) signals

Positive Score
Negative Score
Max Score
Total Score
Performance

Fraunhofer FKIE

Cognitive Processes

- Cognitive processes based on the human processor model[1]



[1] S. K. Card, T. P. Moran, and A. Newell, "The Model Human Processor: An Engineering Model of Human Performance" Handbook of Perception and Human Performance. vol. 2, Cognitive Processes and Performance, pp. 1-35, 1986.

Cognitive Processes

- Cognitive processes can be measured by accuracy and speed in task completion.

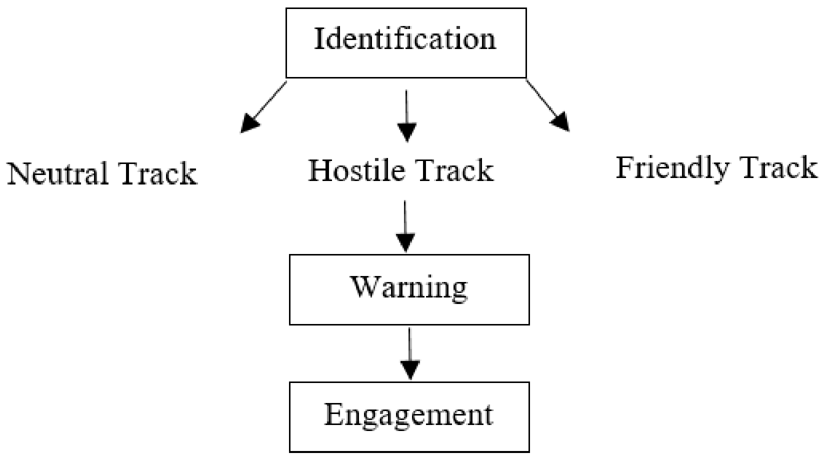
	Accuracy	Speed
Visual attention and perception		X
Object detection and discrimination		X
Rule application and decision-making	X	
Motor task execution	X	X

Scoring Mechanism

(base) signals	Positive Score Negative Score Max Score Total Score Performance	Positive Score Negative Score Max Score Total Score Performance
cognitive processes	Accuracy	Speed

Experimental Task

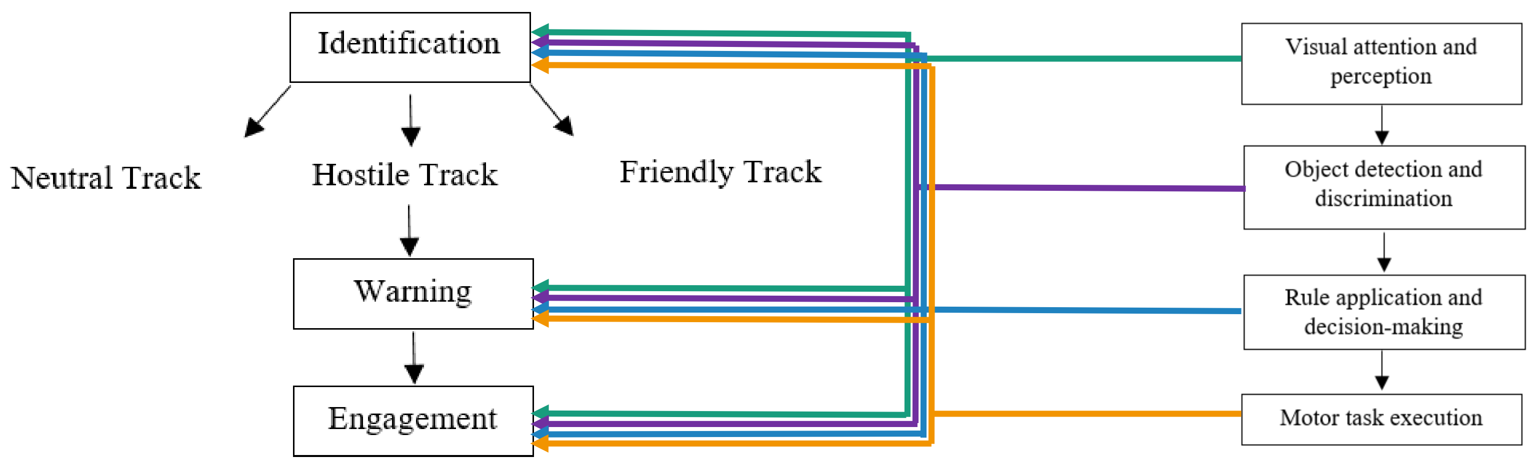
- Focus: Safety-critical vigilance tasks
- Example: Warship Commander Task[1]



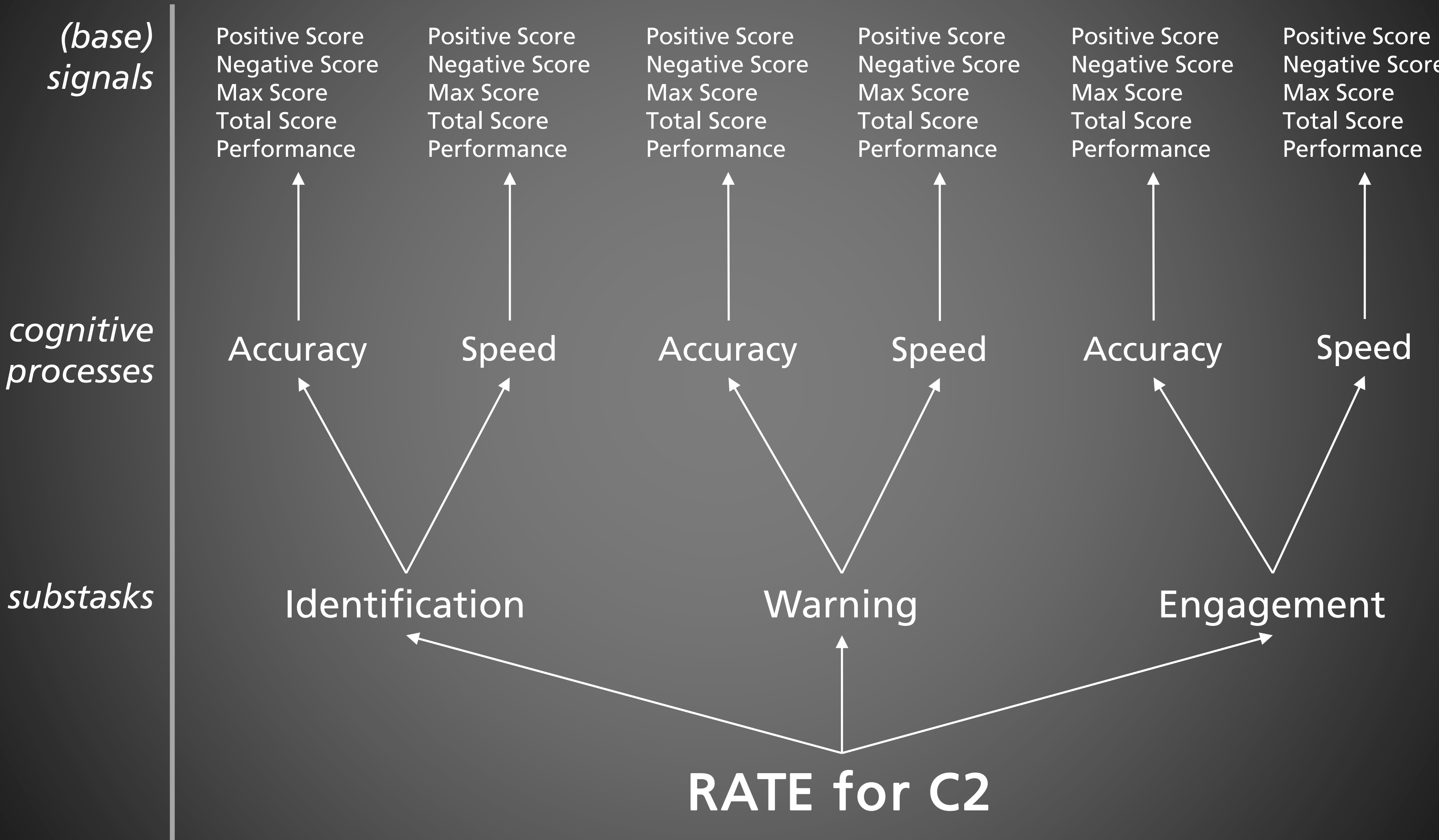
[1] Warship Commander 4.4; Computer Software; San Diego, CA: Pacific Science & Engineering Group; 2003.

Experimental Task

- Cognitive processes are linked to each subtask



Scoring Mechanism



Fraunhofer FKIE

RATE for C2

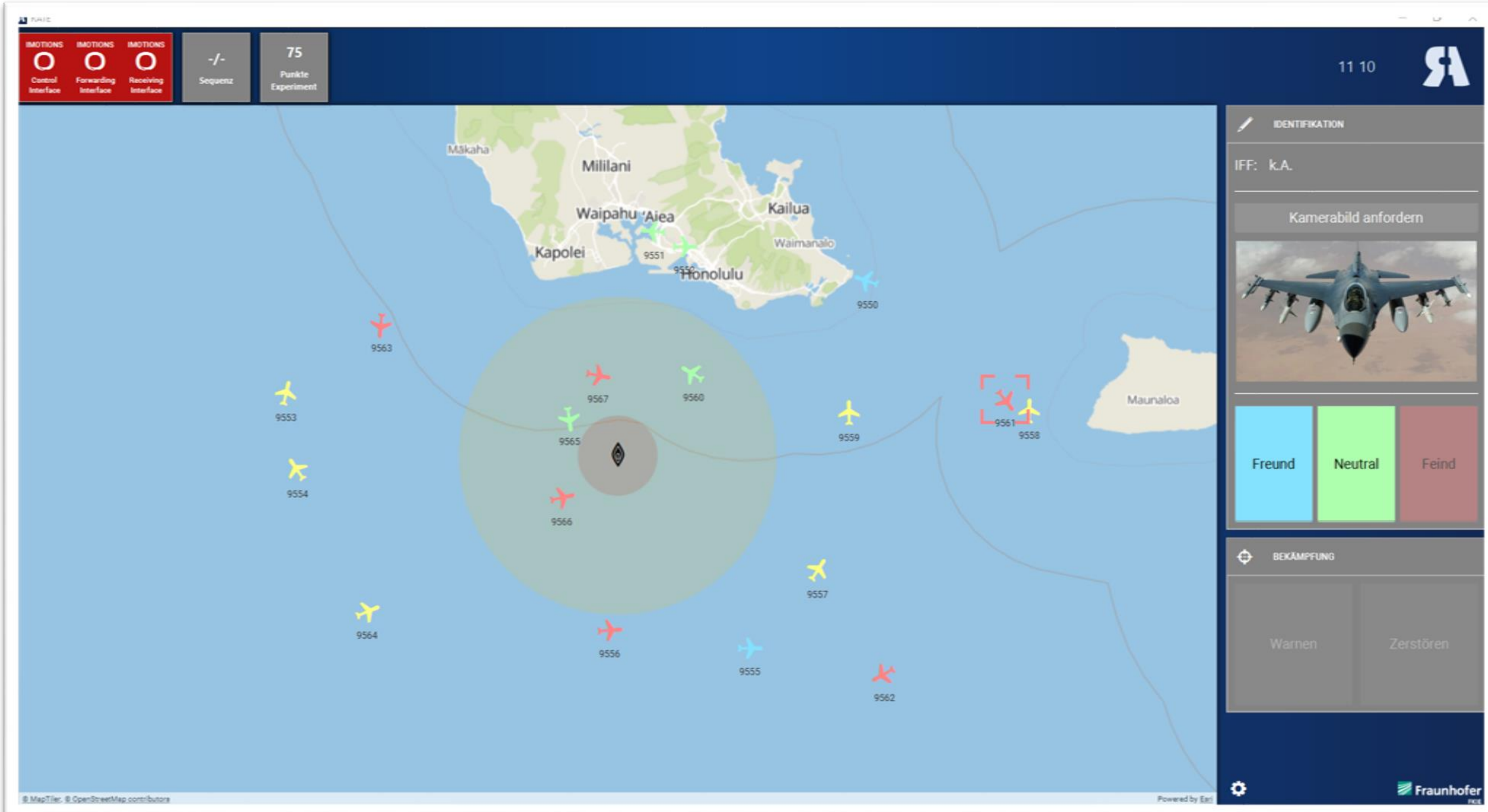
- The scoring mechanism was implemented in the Rich And Adaptable Test Environment (RATE)*
- The described command-and-control task was also implemented in RATE → named RATE for C2
- This setup was used to investigate the relationship between performance and emotion in a command-and-control task[1]

* Rich And Adaptable Test Environment (RATE) = A modular and scalable task environment developed by Fraunhofer FKIE that allows for flexible design of experimental tasks.

[1] A. Schmitz-Hübsch, S. M. Stasch, R. Becker, and S. Fuchs, "Personality Traits in the Relationship of Emotion and Performance in Command-and-Control Environments", International Conference on Advances in Computer-Human Interactions. vol. 14, in press.

Fraunhofer
FKIE

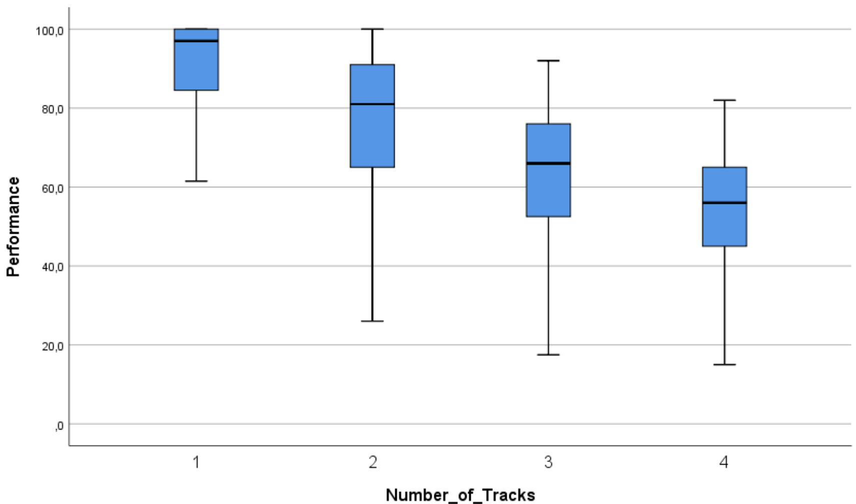
RATE for C2



Fraunhofer FKIE

RATE for C2

- Task load was modulated across scenarios by varying the total number of tracks and the relative proportion of enemy tracks.
  - Based on the cognitive task load model validated with a command-and-control task by de Greef and Arciszewski[1]



The described scoring mechanism was sensitive to task load, as the overall performance decreased with higher task load.

[1] T. D. Greef and H. Arciszewski, "Triggering adaptive automation in naval command and control", Frontiers in adaptive control, IntechOpen, 2009.

Fraunhofer FKIE

RATE for C2

- The normalized score (performance = total score / max score) enables us to compare the operator's performance across conditions and scenarios.

- Using "accuracy" and "speed" as performance criteria, we were able to gain insights into the cognitive processes associated with the specific subtasks.

- Flexible adjustment of the scoring mechanism reflects specific characteristics of the experimental task (e.g., task priorities).

Fraunhofer
FKIE

Lessons learned

- Use decision trees to test the logical order of every possible subtask sequence and its associated scores.

- The priority of a subtask can be reflected in the amount of points earned on the corresponding positive/negative score.
  - Keeping in mind the research question and hypotheses helped to assess the relevance of subtasks and the relationship between them.

- Be careful with conditional subtasks*
  - Correct must not be rewarded if the action was only correct because of a preceding error.
  - Incorrect actions lead to points on the negative score should be determined in the specific task context.

* Conditional subtasks = Subtasks that occur in dependence of the outcome of a previous subtask.

Fraunhofer
FKIE

Lessons learned

- When determining point allocation, omission of necessary actions should neither lead to points on the positive nor on the negative score in order to separate omission errors from correct or incorrect explicit behavior.
  - This does not apply if the omitted action represents incorrect behavior.
- Normalization of the absolute performance score enabled comparisons of operator performance across conditions or scenarios, and even across different experiments.
  - The impact of changes or improvements can then be analyzed at the task and even at subtask level.
  - Test-retest reliability is ensured because the calculation of the score is independent from any dynamic components except the actions of the operator himself.

Scoring Mechanism