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
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
Performance Measurement

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

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

<table>
<thead>
<tr>
<th>Qualitative</th>
<th>Quantitative</th>
</tr>
</thead>
<tbody>
<tr>
<td>results tend to be <strong>subjective</strong></td>
<td><strong>objectify</strong> performance assessment</td>
</tr>
<tr>
<td>results <strong>difficult to verify</strong></td>
<td>enable the use of <strong>inferential statistics</strong> (e.g. significance testing)<strong>¹²</strong></td>
</tr>
</tbody>
</table>


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\(^7\).

---

## Signal Detection Theory (SDT)

<table>
<thead>
<tr>
<th>Response</th>
<th>Signal</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Present</td>
<td>Present</td>
<td>Hit</td>
</tr>
<tr>
<td></td>
<td>Absent</td>
<td>False Alarm</td>
</tr>
<tr>
<td>Absent</td>
<td>Present</td>
<td>Miss</td>
</tr>
<tr>
<td></td>
<td>Absent</td>
<td>Correct Rejection</td>
</tr>
</tbody>
</table>
## Signal Detection Theory (SDT)

<table>
<thead>
<tr>
<th>Signal</th>
<th>Present</th>
<th>Absent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Positive Score 2+1 = 3</td>
<td>Positive Score 2+1 = 3</td>
</tr>
<tr>
<td></td>
<td>Negative Score 2</td>
<td>Negative Score 2</td>
</tr>
<tr>
<td></td>
<td>Max Score 4+1 = 5</td>
<td>Max Score 4</td>
</tr>
<tr>
<td>Present</td>
<td>Total Score 1</td>
<td>Total Score -1</td>
</tr>
<tr>
<td>Response</td>
<td>Performance 20%</td>
<td>Performance -25%</td>
</tr>
</tbody>
</table>

### Positive Score
\[ \text{Positive Score} = \sum (\text{Hits} + \text{Correct Rejections}) \]

### Negative Score
\[ \text{Negative Score} = \sum (\text{Misses} + \text{False Alarms}) \]

### Max Score
\[ \text{Max Score} = \sum \text{Present Signals} \]

### Total Score
\[ \text{Total Score} = \text{Positive Score} - \text{Negative Score} \]

### Performance
\[ \text{Performance} = \frac{\text{Total Score}}{\text{Max Score}} \]
## Scoring Mechanism

<table>
<thead>
<tr>
<th>(base) signals</th>
<th>Positive Score</th>
<th>Negative Score</th>
<th>Max Score</th>
<th>Total Score</th>
<th>Performance</th>
</tr>
</thead>
</table>

Cognitive Processes

- Cognitive processes based on the human processor model\(^1\)

\[\begin{align*}
\text{Visual attention and perception} \\
\downarrow \\
\text{Object detection and discrimination} \\
\downarrow \\
\text{Rule application and decision-making} \\
\downarrow \\
\text{Motor task execution}
\end{align*}\]

Cognitive Processes

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

<table>
<thead>
<tr>
<th>Cognitive Processes</th>
<th>Accuracy</th>
<th>Speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual attention and perception</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Object detection and discrimination</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Rule application and decision-making</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Motor task execution</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>
Scoring Mechanism

(base) signals

(cognitive) processes

Positive Score
Negative Score
Max Score
Total Score
Performance

Accuracy

Positive Score
Negative Score
Max Score
Total Score
Performance

Speed
Experimental Task

- **Focus:** Safety-critical vigilance tasks
- **Example:** Warship Commander Task

Cognitive processes are linked to each subtask
Scoring Mechanism

(base) signals

(cognitive processes)

(subtasks)

Positive Score
Negative Score
Max Score
Total Score
Performance

Accuracy
Speed

Identification
Warning
Engagement

Rate for C2

Positive Score
Negative Score
Max Score
Total Score
Performance

Accuracy
Speed

Positive Score
Negative Score
Max Score
Total Score
Performance

Accuracy
Speed

Positive Score
Negative Score
Max Score
Total Score
Performance

Accuracy
Speed

Positive Score
Negative Score
Max Score
Total Score
Performance

Accuracy
Speed
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 task¹

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

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

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

---

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).
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.
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

- Positive Score
- Negative Score
- Max Score
- Total Score
- Performance

Subtask 1

Category A

Subtask 2

Category B

(base) signals

cognitive processes (or similar)

substrasks