Pushing the Human Performance Envelope – Simulating challenging scenarios for pilots

Barry Kirwan
EUROCONTROL

barry.kirwan@eurocontrol.int
The Premise
Background

→ Many safety critical domains rely on human operators (Air traffic control, Aviation, Maritime, Rail, Military, Medical, etc.)

→ Such domains are usually resilient against single factors. In Air Traffic management, for example, incidents are often the result of 2 or more factors

→ This has led to the notion of a Human Performance Envelope

→ Need to know when human operators are approaching the edges of acceptable performance. E.g., when should automation take over?
HPE – the Basic Concept

→ Human Performance Envelope: Term to describe multifactorial model of human performance and decline
Adaptive Automation using HPE

→ The NINA Project
→ (video)
→ http://nina.dblue.it/
HPE Elements

![Spider diagram illustrating HPE Elements]

Optimal, OK, Degraded, Not OK

[Image of the Usual Suspects movie poster]
Early research findings

→ Literature review; Analysis of 400 ATC incidents; lab experiment; behavioural marker elicitation with controllers

→ Multiple factor relationships:
   → Multiple factors co-occur to influence controller performance
   → Interactions between factors may create a cumulative influence on performance

• Behavioural markers can indicate limits of performance
HPE made simple

→ Very high fatigue, or workload, or stress, can cause an incident
→ Most of the time our ways of working avoid such extremes
→ But what if medium fatigue and medium stress together could cause an incident?
→ How would we find these combinations?
→ What would we do about them?
HPE Representation (2)

Nominal situation
Experiencing discomfort but "goes with the job"
Influence on controller-compensation strategy to protect performance. If not applied, performance may decline
Falling behind, uncomfortable, CC applies strategies
No plan, panic, reliant on colleagues' support, blanking others
Purely reactive, unsafe clearance, Silence, "blacked out", shaking
Which affects performance most?

(this one is easy)
Proving the HPE Concept

What HPE suggests...

Performance Degradation

- Workload (HIGH)
- Stress (MEDIUM)
- Reduced SA (MEDIUM)
- Reduced SA (HIGH)
- Stress (HIGH)
- Workload (HIGH)
An ongoing European aviation study (Future Sky Safety)

Where are the performance limits?
How can we detect and recover?
Can better interface design help?
[When should automation take over?]
WORKLOAD / STRESS / SITUATION AWARENESS
The simulations
2 weeks in July, DLR, Braunschweig, Germany

From Basic, to well beyond what they get in refresher training
10 pilots, 9 scenarios 15-45 minutes long
Co-pilot (PNF) is a ‘confederate’
A full scope simulator
A320 Cockpit
Physiological Measures
Tracking Pilot Performance
Eye-Movement Tracking
Results
Can we rule out fatigue effects?
Samn-Perelli data

Available data: Samn-Perelli score before and after each flight

<table>
<thead>
<tr>
<th>VP</th>
<th>Scenario 1 Run 1</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Before</td>
<td>After</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>No data</td>
<td>No data</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

1. Fully alert, wide awake.
2. Very lively, responsive, but not at peak.
3. Okay, somewhat fresh.
4. A little tired, less than fresh.
5. Moderately tired, let down.
6. Extremely tired, very difficult to concentrate.
7. Completely exhausted, unable to function effectively.

Few missing data.
Analysis by runs (scenario 1, without run 2)

All pilots, Samn-Perelli scores before and after each run

For each run, the fatigue level is relatively low (3 = Okay, Somewhat fresh) and so should not impact performances.
How close to the edge did we get?
Charting the performance degradation curve

Observation:
- Pilot Trainer
- 2 HF people
- Face view / dashboard view
- Extended debrief

Psycho-physiological measurements
- Time-stamped – link to events and reactions

Performance metrics:
- Link Psych-Phys and Obs + subjective opinion to performance degradation and detection / recovery

Performance curve:
- Increasing the elements

- Relaxed, Alert
- Focused
- Under Pressure
- Struggling
- Failing
- The ‘wobble’
- Lost it
Performance curves – Scenario 1 Run 1

BASELINE – 10 pilots

Load

Low

High

Performance

GOOD

POOR

Relaxed

Focused

Under pressure

Struggling

Failing

Lost it

Pilot 1*

Pilot 2

Pilot 3

Pilot 4

Pilot 5

Pilot 6

Pilot 7

Pilot 8

Pilot 9

Pilot 10
Performance curves – Scenario 1 Run 8

12 October, 2016

SAFETY | FUTURE SKY

HIGH WL, HIGH STRESS, LOW SA (High turbulence - Low fuel - Vectors - Loud noise - Low visibility – Loc. interference - Wind shift) – 10 pilots
Basel ine

Workload (MEDIUM)

Workload (HIGH)

Workload (VERY HIGH)

Sig diff btwn Baseline and High/V. High WL: P<0.01
Did the results validate the HPE concept?

**Workload**

![NASA-TLX Chart](image-url)

- **Run 1:** 2.86
- **Run 2:** 3.73
- **Run 3:** 8.83
- **Run 4:** 10.47
- **Run 5:** 6.27
- **Run 6:** 8.86
- **Run 7:** 11.10
- **Run 8:** 12.96
- **Scenario 2:** 12.61
Workload (2)

![Bar chart showing workload data for different runs.](image)
Stress

SAACL

Run 1
Run 2
Run 3
Run 4
Run 5
Run 6
Run 7
Run 8
Scenario 2
Situation Awareness

SART

- Run 1
- Run 2
- Run 3
- Run 4
- Run 5
- Run 6
- Run 7
- Run 8
- Scenario 2
Psycho-Physiological Measures: Extracted signals

- Two electrocardiograms (ECG) leads
- A transthoracic bio-impedance
- Skin temperature
- Accelerometer
- Multi-channel photoplethysmography (PPG)
Pilot 2

→ HR

All Runs are significant different with respect to the baseline
# Behavioural markers

## Table of observed markers

<table>
<thead>
<tr>
<th>Run</th>
<th>Time</th>
<th>Scenario characteristic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2 May, h14:28:30</td>
<td><strong>Heavier breathing</strong></td>
</tr>
<tr>
<td>2</td>
<td>3 May, h16:44:42</td>
<td><strong>Heavier breathing</strong></td>
</tr>
<tr>
<td>3</td>
<td>11 May, h09:55:54</td>
<td><strong>High turbulence throughout whole scenario</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Behavioural marker</th>
<th>From</th>
<th>To</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tongue between lips</td>
<td>10:03:39</td>
<td>10:03:40</td>
</tr>
<tr>
<td>Moves on the seat</td>
<td>10:07:22</td>
<td>10:07:28</td>
</tr>
<tr>
<td>Adjust grip on sidestick</td>
<td>10:07:30</td>
<td>10:07:32</td>
</tr>
<tr>
<td>Adjust grip on sidestick</td>
<td>10:07:36</td>
<td>10:07:38</td>
</tr>
<tr>
<td>Cant hear PM indication (&quot;stay on green line&quot;)</td>
<td>10:09:00</td>
<td>10:09:02</td>
</tr>
<tr>
<td>Tilt head upward</td>
<td>10:09:06</td>
<td>10:09:08</td>
</tr>
<tr>
<td>Moves on the seat</td>
<td>10:09:15</td>
<td>10:09:17</td>
</tr>
<tr>
<td>Panting</td>
<td>10:09:10</td>
<td>10:09:20</td>
</tr>
<tr>
<td>Adjust headphone</td>
<td>10:09:26</td>
<td>10:09:27</td>
</tr>
<tr>
<td>Adjust headphone</td>
<td>10:09:41</td>
<td>10:09:41</td>
</tr>
<tr>
<td>Sniffs</td>
<td>10:09:41</td>
<td>10:09:41</td>
</tr>
<tr>
<td>Deep breath</td>
<td>10:09:53</td>
<td>10:09:57</td>
</tr>
<tr>
<td>Sniffs</td>
<td>10:10:24</td>
<td>10:10:24</td>
</tr>
<tr>
<td>Sniffs</td>
<td>10:11:30</td>
<td>10:11:31</td>
</tr>
<tr>
<td>Sniffs</td>
<td>10:11:43</td>
<td>10:11:50</td>
</tr>
<tr>
<td>Scratches face</td>
<td>10:12:42</td>
<td>10:12:43</td>
</tr>
<tr>
<td>Asking for confirmation of call to ATC to PF (that did not occur)</td>
<td>10:13:43</td>
<td>10:13:45</td>
</tr>
<tr>
<td>Sniffs</td>
<td>10:14:19</td>
<td>10:14:20</td>
</tr>
<tr>
<td>Deep breath</td>
<td>10:14:52</td>
<td>10:14:53</td>
</tr>
<tr>
<td>Pants answering VL question</td>
<td>10:16:17</td>
<td>10:16:20</td>
</tr>
<tr>
<td>Moves on the seat</td>
<td>10:16:52</td>
<td>10:16:57</td>
</tr>
<tr>
<td>Stretching words while speaking (&quot;Ohhhhh...&quot;)</td>
<td>10:17:48</td>
<td>10:17:50</td>
</tr>
</tbody>
</table>
Behavioural markers

Markers & scenario characteristics

Pilot 1 – Run 1

- Mumbling ("Frankfurt")
- Seat adjustment

Change in the communication style, from chatty to silent (sterile cockpit)

**Focused at the beginning...**

**...to Relaxed (due to familiarization)**
Behavioural markers

Markers & scenario characteristics

Pilot 5 – Run 4

- Movement on the seat (moves body upward)
- Movement on the seat (shakes his shoulders)
- Ticking lips
- Fast head turns (looking left-right-left-right)
- Long and deep breath
- Some heavier breathings
- Tightening lips
- Grimace
- Grimace

Started Relaxed
then Focused
Focused / Under pres
Summary of Results so Far

→ The simulations were credible to the pilots
→ We did push the envelope
→ Subjective evidence for HPE concept
→ Psychophysiological evidence – the jury is still out, but of the 6 measures 2 worked best
→ Behavioural markers? Needs more work, a lot of individual differences (10 pilots not really sufficient)
→ The pilots said they should do more simulations like these, and wanted to come back for more!
Way forward

→ Further analysis (we still have a ton of un-analysed data!)
→ Work on the behavioural markers – can we identify signs and symptoms that crew can detect and act upon?
→ Psychophysiological markers – could a wrist-watch computer warn of approaching edges?
→ Cockpit design – what do these results tell us about how to design the cockpit better, to reduce workload and stress, and improve situation awareness?
→ What about fatigue, since so many pilots are complaining about it?
→ Three possible further simulations being considered – one on fatigue, one on HPE & air traffic controllers, one on cockpit design
→ When (if ever) should automation take over?
Thanks for Listening!