Validation of Controller Workload Predictors at Conventional and Remote Towers

Billy Josefsson
Lothar Meyer
Maximilian Peukert

Tatiana Polishchuk
Christiane Schmidt

lfv

liu

li.u

TEAMWORK with automation
Motivation

• Aim of ANSPs: well-balanced workload level for ATCOs during all operations
  ➡ Objective assessment of workload crucial to find right level
• Workload is a subjective concept: workload “represents the cost incurred by a human
  operator to achieve a particular level of performance”[1]
  ➡ Cannot be measured directly
  ➡ Need for quantitative measures that correlate with ATCO workload
• Many studies on quantitative workload predictors exist for en-route traffic
• Not true for aerodrome control, an even less so for Remote Tower control
  ▸ Remote Tower Services:
    • ATCOs control traffic from airports remotely from a Remote Tower Center (RTC)
    • Possible to control several airports from single ATCO working position (“multiple mode”)
• Why are we interested in workload assessment?
  • Staff planning: make sure not ATCO is confronted with traffic-inherent situations that yield
    unacceptable workload
  • For multiple mode: take traffic at different airports into account
  • In which scenarios do we need extra staff?
    ➡ When does workload associated with traffic of one or several aerodromes exceeds moderate
    workload?
• Today: relation between
  • Subjective workload ratings
  • Several quantitative measures:
    • # ATCO tasks (ATs)
    • Measures related to communication length
  • Validation of quantitative workload indicators on their power to predict workload in a conventional
    tower and Remote Tower (single and multiple mode)

Setup
ATCO responsibility:
• Ensure safe separation of aircraft
• Enable aircraft to reach destination in time
➡ ATCOs Perform various tasks that drive the mental workload
• Taskload: measures objective demands of the ATCO's monitoring task
• Workload: measures subjective, mentally experienced stress during a task
• All factors external to human operator = stress
➡ Results in individual workload (depending on different properties of the human operator)
Assessing workload:
• Different scales
• Adapted Cooper-Harper scale (CHS)

Neither of our studies was planned as a stress test at the boundaries of capabilities → Lower ratings present

• Instantaneous Self Assessment (ISA) scale

• Different study setups → Different scales used → Approximate way of transferring ISA to CHS
ATCO Tasks (ATs)

- Arrival
- Clearance
- Communication
- **Abnormal situation:** An abnormal situation induces several other situations, hence, we count these.
- Departure
- Secondary Task
- Taxi
Quantitative Measures Based on Communication Duration

- Counting ATs
  - Treats all AT types equally
- But: some AT types may have higher impact on workload than others
- Communication: basic tasks (audio-acoustic channel)
- Integrate length of communication related to AT types (as weights)
  - 1. Average communication times for AT types
  - 2. Percentage of the total communication time for AT types
- Both might indicate an increase in workload:
  - 1. Individual call related to AT1 takes up more time than those for AT2
    - Caused by longer phraseology or increased need for callbacks for AT1
    - Longer time of attention for these calls
  - 2. Total time spent for communication related to AT1 longer than for AT2
    - We assume sheer number of call leads to higher attention for these calls
• Probe questions during experimental studies: measure situational awareness
• Situation Present Assessment Method (SPAM):
  - Measure ATCO reaction times to questions related to the current scenario
  - Proper SA: low latency + high accuracy
  - Possible question: ``What is the actual wind speed for Sundsvall/Örnsköldsvik (S/Ö)?""
• Goal: validate quantitative indicators on power to predict ATCO workload
• Goal: predict increases and decreases of ATCO workload
• Workload is accumulated metric
  ➔ Identify influencing factors
• Classical way: look at correlation
• Here also: other criteria that enable us to explain increases and decreases
  ➔ We borrow classical mathematical notation:
• A measure constitutes a **necessary condition for workload increase**, if every workload rating increase is accompanied by an increase in the measure.
• A measure constitutes a **sufficient condition for workload increase**, if every increase in the measure also yields an increase in the workload rating.
• Analogously: **necessary and sufficient condition for workload decrease**
• Sufficient measure for workload increase
  ➔ We can observe only the measure: each increase will yield an increase in workload rating
  ➔ Predict increase in workload rating
• IF: measure is sufficient condition for workload increases and decreases
  ➔ The measure would yield a perfect predictor for workload changes
Study Setup

Two studies at two different occasions at two different locations:

**Field Study**
- Conventional tower at Bromma airport
- March 4, 2019
- During actual operation
- 5 video cameras, 3 towards ATCOs, 2 towards opposite RWY ends
- Videos used to reconstruct ATs
- 2 ATCOs + 1 assistant
- 3 ATCOs observed for 4 hours (1f, 2m)
- ATCO's mean age: 43
- Mean years worked as ATCO: 19.6 years
- CHS used
- ATCO WL rating assessed every 5 mins (first 15 mins)
  - Sample size: 45
- Also measured: length and purpose of communication

**Simulation Study**
- Simulated Remote Tower in single or multiple modes for airports in Örnsköldsvik and Sundsvall, simulation in Sundsvall
- May 6-17 2019
- 3 video cameras towards ATCOs
- 3 ATCOs (2f, 1m)
- ATCO’s mean age: 52
- Mean years worked as ATCO: 23.3
- Mean years at RTC: 5.6 years
- Singular mode: 5 movements
- Multiple mode: 6 movements
- Each run: 75 mins
- ISA scale used
- ATCO WL rating assessed every 3 mins
  - Sample size: 25 per ATCO
- Also measured: length and purpose of communication; reaction time to SPAM queries

We observed:
- Snow sweeping with a convoy of 10-14 vehicles
- #Movements/h: 4, 5, 9, 27
Results Field Study
Conjecture: Increase in **workload rating** always accompanied by an increase in the **number of ATs in current or previous time period**.

Why two consecutive points in time?

More ATs may accumulate and result in increased WL rating at following query.

Conjecture holds!

BUT converse is not true, that is:

Not every increase in the **number of ATs** leads to an increased **workload rating**

- Increase in the **number of ATs** can be a **necessary**, but **not a sufficient** indicator for increased **workload**
On average longer: clearance initiated by one party, reply by the other party, and for airborne operations the second party awaits repetition to conform proper reception ↔ other call types

<table>
<thead>
<tr>
<th></th>
<th>Arrival</th>
<th>Clearance</th>
<th>Comm</th>
<th>Taxi</th>
<th>Departure</th>
<th>Ground</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average (in s)</td>
<td>10.04348</td>
<td>20.34783</td>
<td>11.2</td>
<td>10.7</td>
<td>11.44118</td>
<td>13.48</td>
<td>0</td>
</tr>
<tr>
<td>Sum (in s)</td>
<td>231</td>
<td>408</td>
<td>448</td>
<td>321</td>
<td>389</td>
<td>674</td>
<td>2531</td>
</tr>
<tr>
<td>Percentage</td>
<td>9.13%</td>
<td>18.49%</td>
<td>17.70%</td>
<td>12.68%</td>
<td>15.37%</td>
<td>26.63%</td>
<td>100%</td>
</tr>
<tr>
<td>Range (in s)</td>
<td>6-16</td>
<td>6-57</td>
<td>4-72</td>
<td>5-28</td>
<td>5-27</td>
<td>3-37</td>
<td></td>
</tr>
</tbody>
</table>

Communication split: Weights for ATs

- **Average call duration for each AT type**
- **Sum of all radio call durations for each AT type**
- **Percentage of sum of all radio call durations for each AT type**

Communication to ground vehicles takes up most time (snow cleaning clearly represented here)

Total time spend: clearances with average value
Conjecture: Increase in **workload rating** always accompanied by an increase in the **ATs weighted with the percentage of the total communication time in current or previous time period**.

Conjecture holds!
Again: both current and previous time period $\rightarrow$ Sum of average-communication-duration weighted #ATs at current and previous time

Increase in **workload rating** always accompanied by an increase in at least one of:
- **Average-communication-duration weighted ATs in current or previous time period**
- Sum of average-communication-duration weighted ATs at current and previous time

$\rightarrow$ Necessary condition

BUT: Increase in at least on of the two criteria is not a sufficient condition for an increase in workload.
A brief note:
Average workload rating was higher in the first three hours, during which snow
sweeping occurred,
than in the final hour with peak traffic (27 movements opposed to 4, 5, and 9
movements in the prior hours).
More data is needed to study the influence of weather in detail.
Results Simulation Study
• Only communication shows significant higher communication duration in multiple than in single mode (one-sided $U$-test, $p$-value 1.65%)

• Other increases not significant:
  • Increase in average communication times related to arrivals nearly significant (one-sided $U$-test, $p$-value 7.57%)
  • Increase in average communication times related to clearances nearly significant (one-sided $U$-test, $p$-value 6.7%)
  • † Due to risk compensation behaviour by operator: avoid risk at expense of time?

• We normalised weights
• Used for single mode, for multiple mode
Single Mode
- All ATCOs hold endorsement for Sundsvall
- ATCO3 larger WL variations: 9 yrs ATCO, for 1 and 2: 30 and 41 years, resp.
Workload vs. #ATs/Weighted ATs

- **Number of ATs not** a necessary condition for increase in workload rating ($\leq 43\%$ of WL rating increases accompanied by increase in #of ATs)

- Previous measures incl. communication length aren’t for all ATCOs

- Considering length of communication, we also consider the following period—rationale: ATCO anticipates later tasks

Increase in workload rating always accompanied by an increase in:

- All measures that take the communication time into account (ATCO1)

- Sum of the average-communication-duration weighted ATs for two consecutive time periods (ATCO2)

- Average-communication-duration weighted ATs in the previous, current or following time period (ATCO3)
Multiple Mode
• ATCO3 stressed because of problems with simulation equipment → start at 9:09 (instead of 9:00)
• ATCO1 longest RTC experience, but endorsement only for Sundsvall → confronted with unknown working environment
• ATCO2 and ATCO3 endorsements for both airports
  ➡ Generally higher level and higher variations in workload rating for ATCO1
We can observe a necessary condition: Increase in **workload rating** always accompanied by an increase in at least one of:

- **Duration of communication during that time interval**
- Sum of the average-communication-duration weighted ATs for two consecutive time periods

Comparable to the necessary condition from the field study, one of:

- Average-communication-duration weighted ATs in current or previous time period
- Sum of average-communication-duration weighted ATs at current and previous time

*Again: Number of ATs not* a necessary condition for increase in **workload rating**

How exactly? One of those or a combined measure?
For most queries: Reaction time by an ATCO in multiple mode increases vs. single mode
- Multiple mode: ATCO confronted with more tasks
  ➡ He/she might be less responsive—exhibit risk compensation behavior
- Insecurity
  ➡ ATCO double checks to avoid mistakes
  ➡ Slowdown
  ➡ Can be an indicator for uncertainty [2]
- Uncertainty is one of main stressors (apart from time pressure)
- BUT: trend not true for all queries and ATCOs!
  - Reaction time ATCO2, SPAM track + clearance: multiple lower than single
  - Reaction time ATCO3, SPAM position + wind speed + wind direction: multiple lower than single
- ATCO1 RTC experience, but endorsement only for Sundsvall
  ➡ Confronted with new environment in multiple mode
  ➡ Explains increases from single to multiple mode
- ATCO2+3 endorsements for both airports and RTC experience
  ➡ Smaller increases or decreases (less time over all, less time allocated for each tasks, while all fully under control)
  ➡ Underlines: training helps to decrease ATCO's stress!

Conclusion & Outlook
Conclusion & Outlook

- We studied the relationship between subjective workload ratings and quantitative measures that integrate more than a single indicator.
- We identified a necessary condition for an increase in workload rating (all ATCO ratings, field + simulation study).
- Each increase in the ATCO workload rating is accompanied by an increase in at least one of:
  - ATs weighted with the percentage of the total communication time
  - Average-communication-duration weighted ATs in the previous, current or following time period
  - Sum of the average-communication-duration weighted ATs for two consecutive time periods
  - Duration of communication during that time interval

- We validated these quantitative indicators on their predictability of workload increases:
  - All criteria related to communication time
  - Simply counting the ATs is not a good workload indicator (not a necessary condition in the simulation study)
- Necessary condition → insights into workload development
- Sufficient criterion would be even more beneficial
- Our result indicates: other factors (e.g., mental effort for decision-making, w/o measurable indicator) might even out variations in communication-time related measures
- Or: WL scales not fine-grained enough to reflect even small changes in workload rating
- Possible: variations in our communication-related measure do yield change in workload
- Goal: also sufficient criterion for workload rating decreases
- Combined: quantitative workload predictor
Conclusion & Outlook

• We used communication data for full study period as weights
  ➡ Can be derived from large data sets
  ➡ Can lead (with AT predictions) to predictions of workload
• Other possibility: communication length for each AT over time
  ➡ Study correlation between temporal progression of communication length of an AT and workload rating
  ➡ Could not be used for predictions

• We use ISA scale and CHS for workload rating
• But: only relatively small variations in workload rating
• Both scales good for binary decisions:
  - Critical/unacceptable levels of workload
  - Workload levels without reduced SA
• Smaller variations of workload on levels without reduced SA cannot be reflected equally well
• Plus: ATCOs not familiar with scales
  ➡ May just name a reasonable number
• Social desirability
  ➡ Necessity for instrument that is able to register variability on lower workload levels

Future:
• Integrate measurements of other factors (e.g., runway friction)
• Possibly: identify physical measurement (e.g., pupil diameter) with high correlation to workload
  ➡ Goal: Quantitative empiric measurements with high power of predicting this physical measurement
Each increase in the **ATCO workload rating** is accompanied by an increase in at least one of:

- **ATs weighted with the percentage of the total communication time**
- **Average-communication-duration weighted ATs in the previous, current or following time period**
- **Sum of the average-communication-duration weighted ATs for two consecutive time periods**
- **Duration of communication during that time interval**

Thanks.

Questions? christiane.schmidt@liu.se

Slides: [http://webstaff.itn.liu.se/~chrsc91/](http://webstaff.itn.liu.se/~chrsc91/)