Usability Evaluation of the Spot and Runway Departure Advisor (SARDA) Concept in a Dallas/Fort Worth Airport Tower Simulation

Miwa Hayashi, Ty Hoang, Yoon Jung
NASA Ames Research Center

Gautam Gupta, Waqar Malik
Univ. of California, Santa Cruz

Victoria Dulchinos
San Jose State Univ.

10th USA/Europe ATM R&D Seminar
Chicago, IL | July 10-13, 2013
Departures handled first-come-first-served

- Taxiway congestion
- Taxi delay
- Fuel burn, air pollution
- Departure time uncertainty

Photo Credit: nydolce via Compfight cc
Tools to address the inefficiencies:

- **In the USA**
  - Collaborative Departure Queue Management (CDQM) (FAA)
  - Departure Manager tool at JFK airport (Port Authority of NY & NJ)
  - N-control strategy (MIT)
  - Spot and Runway Departure Advisor (SARDA) (NASA)

- **In Europe**
  - Departure Manager (DMAN) (EUROCONTROL, DLR)

Photo Credit: henrie via Compfight cc
Spot and Runway Departure Advisor (SARDA)

- Calculates an optimal runway-use sequence
  - Wake-turbulence separation
  - Constraints for Traffic Management Initiative (TMI)
  - Gaps for runway crossing

- Provides tower controllers with advisory for spot release time for each departure

- Allows each departure to wait at the gate with its engines turned off

Photo Credit: Nikon-Spence via Compfight cc
SARDA Human-in-the-Loop Simulator Evaluation

- Taxi delay reduction by 45% (medium traffic) or 60% (heavy traffic)
- Fuel burn reduction by 23% (medium traffic) or 33% (heavy traffic)

Research Objective

SARDA improved system performance, but…

Did the controllers have to work extra hard?

**Objective:** Evaluate the SARDA tool’s usability

*Photo Credit: Dominic Hart (NASA)*
Usability

“The extent to which the users of a product are able to work effectively, efficiently, and with satisfaction.” (ISO, 1998)

Effectiveness
Accuracy and completeness
System performance

Efficiency
Resource availability to achieve the goals
Workload, attention

Satisfaction
Attitude, emotion
Preference, trust, acceptance
The Rest of the Presentation

- Simulation
- Results
- Conclusion & Future Work
Simulation

- Simulated Traffic
- Simulator
- Electronic Flight Strips (EFS)
- Experimental Design
Simulated Traffic

Dallas/Fort Worth International Airport (DFW), East Tower operations in South-flow configuration was simulated.

Photo Credit: Fly For Fun via Compfight cc
Simulated Traffic

Dallas/Fort Worth International Airport (DFW), East Tower operations in South-flow configuration was simulated.
Simulated Traffic

Dallas/Fort Worth International Airport (DFW), East Tower operations in South-flow configuration was simulated.

Photo Credit: Fly For Fun via Compfight cc
Simulator

- Ground Controller Workstation
- Local Controller Workstation
- Surface map display
- EFS
- Airborne radar display
- Surface map display
- Electronic Flight Strips (EFS)
Electronic Flight Strips (EFS)

- Required to allow the SARDA advisories to be dynamically updated.
- Shown on a 24” touch screen monitor.
- Requires different designs for different positions.
- Used in all runs in the simulation.

Photo Credit: Dominic Hart (NASA)
Ground EFS Strips

<table>
<thead>
<tr>
<th>Aircraft Call Sign</th>
<th>Aircraft Type</th>
<th>Spot/Taxiway</th>
<th>Rwy/Fix/Destination</th>
<th>Default action button</th>
</tr>
</thead>
<tbody>
<tr>
<td>AWE766</td>
<td>B737</td>
<td>4</td>
<td>S37/K.EG</td>
<td>17R/NOB/LIT</td>
</tr>
<tr>
<td>DAL859</td>
<td>B737</td>
<td>3</td>
<td>S45/K...EH</td>
<td>17R/CLR/BTR</td>
</tr>
<tr>
<td>DAL348</td>
<td>B752</td>
<td>2</td>
<td>S37/K...EH</td>
<td>17R/CLR/FLL</td>
</tr>
<tr>
<td>AAL199</td>
<td>MD82</td>
<td>1</td>
<td>S11/K.EG</td>
<td>17R/TRI/SJT</td>
</tr>
</tbody>
</table>

SARDA Advisories

- Sequence: 3
- Countdown timer: 00:45
  - Less than 60 seconds
- Sequence: 2
  - Countdown timer: 00:17
  - Less than 30 seconds
Experimental Design

- 48 runs in 3 weeks (16 per week), 2 controller positions per run
- 6 retired DFW tower controllers (2 per week)
- 5 pseudo pilot participants
- 4 traffic scenarios
  - 2 medium traffic (1.2x of the current-day DFW traffic, ~35 minutes)
  - 2 heavy traffic (1.5x, ~45 minutes)
- Independent variables
  - Advisory (Advisory vs. Baseline runs)
  - Position (Ground vs. Local)
  - Scenario (4 scenarios of 2 traffic levels)
  - Participant (6 controllers)
- Counterbalanced run orders within each participant
Experimental Design

- **Dependent variables**
  - Real-time workload ratings measured every 5 minutes
  - NASA Task Load Index (TLX) workload ratings
  - More questionnaire responses
    - Post-run questionnaire
    - Post-study questionnaire

Usability assessment

Efficiency
- Workload, attention

Satisfaction
- Preference, trust, acceptance, etc.
Experimental Design

Advisory vs. Baseline runs

- The presence of the optimized runway-use sequence allowed additional functions in the Advisory runs.
  - Gate holding
  - Automatic strip sorting by the SARDA sequence
  - Advisory for taxi routes and a runway queue

- In the Advisory runs, the controllers were asked to follow the advisories as much as possible.
  - Needed for our research purpose
  - However, different from the actual SARDA Concept of Operations (controllers are allowed to reject any advisory any time)
Results

- Efficiency analysis
  - NASA TLX Workload Ratings
  - Spare Attention Ratings

- Satisfaction assessment
  - Subjective Rating Responses
  - Agreement vs. Understanding
Efficiency – NASA TLX Workload Ratings

- Asked in a post-run questionnaire form
- Six ratings
  - Temporal Demand
  - Mental Demand
  - Effort
  - Frustration
  - Physical Demand
  - Performance
- Scales
  Lowest workload = 1, highest workload = 10 (except the Performance ratings, where the poorest performance = 1 and the most successful performance = 10)

ANOVA Model

Main effects:
- Advisory
- Position
- Scenario*
- Participant*

Interaction effects:
- Advisory × Position
- Advisory × Scenario*
- Advisory × Participant*

* Random effects
Efficiency – NASA TLX Workload Ratings

Means of the 4 ratings were lower in the Advisory runs by about 2 points.
Statistically (or marginally) significant Advisory × Traffic Level effects showed that, in the Advisory runs, controllers’ workload did not increase as much when the traffic-volume increased.
Efficiency – Spare Attention Ratings

- One of the post-run questions asked a self-rated amount of spare attention level during the peak period in the run
- Scale: Little = 1, plenty = 5

ANOVA Model

Main effects:
- Advisory
- Position
- Scenario
- Participant

Interaction effects:
- Advisory × Position
- Advisory × Scenario
- Advisory × Participant

* Random effects
Efficiency – Spare Attention Ratings

- Controllers reported they had more spare attention in the Advisory runs.
- When the SARDA advisories were provided, the ratings did not drop as steeply when the traffic load increased.

$$F^*(1.01, 7.90) = 10.6, \ p = 0.012$$

$$F(1, 15) = 16.3, \ p = 0.002$$
Efficiency Analysis – Discussion

The SARDA advisories…

- Reduced certain aspects of controllers’ workload
  - **Temporal Demand** and **Mental Demand**
    - SARDA’s scheduling function
    - Additional functions (gate holding, automatic strip sorting & advisories for taxi routes and a runway queue)
  - **Physical Demand**
    - Additional functions (automatic strip sorting & advisories for taxi routes and a runway queue)

- Made controllers’ workload insensitive to traffic-load increase

- Relieved controllers’ attentional resource, which could be used for other tasks if needed
# Satisfaction – Subjective Rating Responses

<table>
<thead>
<tr>
<th>Question</th>
<th>Scale</th>
<th>Local Mean (St.dv)</th>
<th>Ground Mean (St.dv)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall, how easy/difficult was it to use the advisories?</td>
<td>1 = Easy, 7 = Difficult</td>
<td>1.7 (0.8)</td>
<td>1.2 (0.4)</td>
</tr>
<tr>
<td>How much did the advisories help/interfere with your management of the traffic with a TMI restriction?</td>
<td>1 = Helped, 7 = Interfered</td>
<td>2.2 (1.5)</td>
<td>3.0 (1.3)</td>
</tr>
<tr>
<td>How much did you trust the advisories to help you in making better decisions?</td>
<td>1 = Trusted, 7 = Did not trust</td>
<td>3.2 (1.7)</td>
<td>2.8 (1.3)</td>
</tr>
<tr>
<td>Given the choice, would you prefer or not prefer to have the advisories?</td>
<td>1 = Prefer, 7 = Not prefer</td>
<td>3.2 (1.9)</td>
<td>2.8 (1.9)</td>
</tr>
</tbody>
</table>

- These were asked in the post-study questionnaire form
- All the means fell on the *favorable* side for SARDA (< 4)
- However, some means were close to 4 (the neutral point)
Satisfaction – Agreement vs. Understanding

- Post-run questionnaire responses were plotted
  - % Time they agreed with the advisories
  - % Time they understood the advisories

- The majority of points fell around the diagonal line

- The plot suggests a loose correlation between the % time they agreed the advisories and the % time they understood them
Satisfaction – Discussion

- Controller responses imply that they thought SARDA was easy to use.

- Some responses suggested slight reservations from the controllers. More work is needed to increase controller trust and acceptance:
  - Help them understand how the advisories are computed, e.g., how the competing priorities are handled.
  - Allow controllers to deviate from the SARDA advisories (that is the original Concept of Operations of SARDA)
Conclusion

- The SARDA advisories helped lower controllers’ workload.

- When the SARDA advisories are provided, controllers’ workload became less sensitive to the traffic load increase.

- Controllers reported that they thought SARDA was easy to use.

- Controllers also appeared to have slight reservations in fully endorsing the concept, just yet. More work is needed to increase controller trust and acceptance.
Thank you

Miwa.Hayashi@nasa.gov

Photo Credits:  (This slide) Larry D. Moore Creative Common @ Wikipedia.com  
(First slide) Boris Rabin (NASA)