Available online at www.sciencedirect.com

ScienceDirect

IFAC-PapersOnLine 49-19 (2016) 313–318

Cross-sector transferability of metrics for air traffic controller workload

S.M.B. Abdul Rahman * , C. Borst ** , M.M. van Paassen ** and M. Mulder **

* Mechanical Engineering Faculty, Universiti Teknologi MARA, 40450 Shah Alam, Selangor Darul Ehsan, Malaysia (e-mail: mariam4528@salam.uitm.edu.my)

** Control and Simulation section, TU Delft, Delft, The Netherlands (e-mail: {c.borst, m.m.vanpaassen, m.mulder}@tudelft.nl)

Abstract: Air traffic controller workload is an important impediment to air transport growth. Several approaches exist that aim to better understand the causes for workload, and models have been derived to predict workload in new operational settings. These methods often relate workload to the difficulty, or complexity, that an average controller would have to safely manage all traffic in a sector with a particular traffic demand. In this paper, several of these complexity-based metrics for workload will be compared. Of special interest is whether the complexity measures *transfer* from one sector design to another. That is, does a metric that is well-tuned to predict workload for controllers working in one sector, also predict the workload for another group of controllers active in a different sector? Results from a human-in-the-loop experiment show that a solution space-based metric, which requires no tuning or weighing at all, has the highest correlations with subjectively reported workload, and also yields the best workload predictions across different controller groups and sectors. *Copyright ©2016 IFAC*

Keywords: Air traffic control, taskload, mental workload, supervisory control

1. INTRODUCTION

Safety, efficiency and orderly flow of air traffic are the three main Air Traffic Controller (ATCO) responsibilities in managing traffic. Current Air Traffic Control (ATC) practice primarily uses conventional technology (e.g., radar and radio telephony communication), with only little automation support for the operators involved, which renders the task of supervising air traffic heavily constrained by human performance limits (Costa, 1993). Without countermeasures, the rise in projected air traffic would inevitably result in a further increase in the workload of ATCOs, often cited as one of the main impediments to air transport growth (Janic, 1997, Hilburn, 2004, Koros et al., 2004).

The ability to understand what causes workload, and predict ATCO workload in future scenarios, is an important avenue of research. In this paper we use the term taskload to refer to the objective demands of a task, and workload to address the subjective demand as experienced by an operator (Stassen et al., 1990). Several approaches exist to determine ATC taskload, such as simply counting the number of aircraft that need to be managed simultaneously in a sector. Although this technique works quite satisfactorily, it does not include any knowledge regarding *how* these aircraft fly through the sector. Figure 1 illustrates that a situation where all aircraft fly parallel routes is very likely to be much easier for an operator to supervise and control than a situation where the *same* number of aircraft fly random routes.

More recent techniques relate task demand load to metrics of sector complexity (Laudeman et al., 1998, Sridhar et al., 1998, Chatterji and Sridhar, 2001, Kopardekar and Magyarits, 2002, Masalonis et al., 2003). An important example is the dynamic density (DD) metric, which includes aircraft dynamic behavior in the sector, by taking into account "*the collective effort of all factors or variables that contribute to sector-level ATC complexity or difficulty at any point of time*" (Kopardekar and Magyarits, 2002). The DD calculation is based on weights that are gathered from applying regression methods on samples of traffic data and comparing these to subjective workload ratings. The DD metric therefore includes both objective as well as subjective measurements and could be less suitable to predict the workload of different controllers working in another sector.

In the solution space (SSD)-based approach, taskload is related to the difficulty of the ATC control problem, where the "solution space" captures the geometrical and

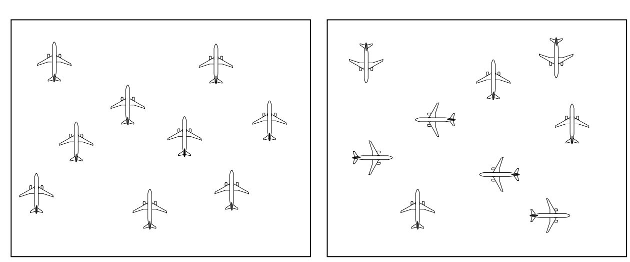


Fig. 1. Two traffic situations, with the same number of aircraft, one easy and the other difficult to control.

2405-8963 © 2016, IFAC (International Federation of Automatic Control) Hosting by Elsevier Ltd. All rights reserved.
Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2016.10.561

kinematic constraints that limit (and therefore, guide) ATCO control actions (Hermes et al., 2009, Mercado-Velasco et al., 2010, D'Engelbronner et al., 2015). Previous studies found high correlations between workload ratings and the *area* of the available SSD control space.

This paper discusses a comparison of several sector complexity measures regarding their ability to match the subjective workload ratings obtained in a human-in-the-loop experiment. We will evaluate the Static Density (SD), which equals the number of aircraft flying in a sector, the Dynamic Density (DD) as proposed by NASA, and a solution space-based (SSD) approach developed by TU Delft. We will focus in particular on the performance of these metrics in predicting workload ratings *across* different sectors and *across* different groups of operators, i.e., their ability to transfer between sectors and controllers.

2. EXPERIMENT

Our study relies on computing the correlation between ATCO workload ratings and a number of complexity metrics: SD, SSD and DD. A human-in-the-loop experiment was conducted in which eight participants, who all received an extensive ATC introductory course and has worked closely in the ATC domain, but none of them were operational ATCOs, managed the air traffic in two sectors (Abdul Rahman, 2014). While managing the air traffic, every minute the subject was requested to indicate the workload on a scale between 0 and 100, yielding a workload profile for each controller. After each run, based on the recorded aircraft parameters (their position, speed, and heading), the complexity metrics were computed, and for the DD metrics the weightings were determined through linear regression techniques. When all data were available, the correlation analysis was conducted.

2.1 Method

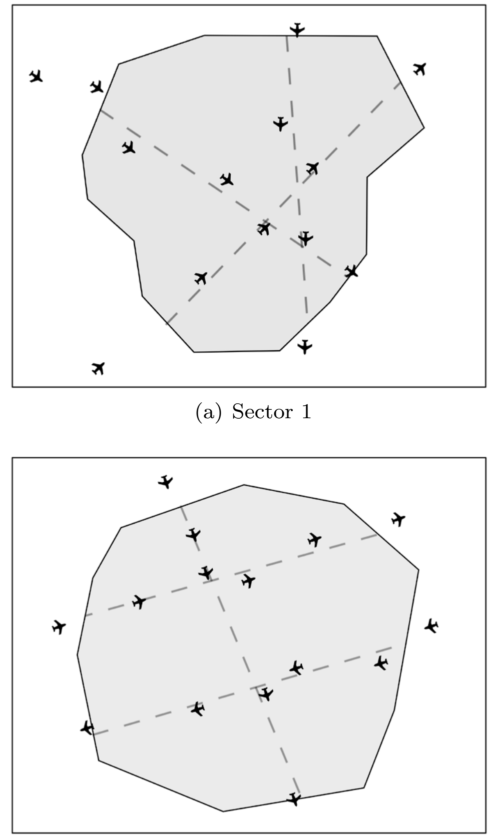
Independent variables The experiment had two independent variables: (i) two different sector designs were used, Figure 2, and (ii) four different traffic sequences were simulated. The latter were varied to avoid scenario recognition during the course of the experiment.

The two sectors differed in the number of crossing points, combinations of the intercept angle of traffic routes, the clustering of crossing points, different entry and exit points, differences in sector shape and sector area. The four traffic patterns did not differ in the total number of aircraft simulated, but rather in their distribution in time.

In addition, we divided the eight participants in two groups of four subjects each, to allow us to study the effects of using the metrics across groups of participants.

Subject instructions Subjects were instructed to guide all aircraft safely through the sector and have them exit the sector at their pre-defined exit point. All aircraft were of the same type, so had the same constraints in velocity and heading; altitude was fixed to one flight level.

Procedure All subjects were briefed on the nature of the experiment, the goals to be achieved and the simulator used. Each participant completed two blocks of four scenarios that lasted 25 minutes each. Each block was preceded with a training scenario that lasted for ten minutes. Subjects were asked to indicate their workload using a scale that appeared on top of the plan view display. The workload rating, measured on a zero to 100 scale, was provided by the subject every 60 seconds during the experiment run. In order to correct for inter-subject differences, Z-scores of the subjective ratings were used in the subsequent data exploration. This correction was performed by calculating the Z-scores for every test subject.



(a) Sector 1

(b) Sector 2

Fig. 2. Sector design and traffic flows.

The experiment was run at four times real-time, similar to what was done in previous research (Hermes et al., 2009, d'Engelbronner et al., 2010, Mercado Velasco et al., 2010). The rationale behind this was to create more variability in traffic situations (and thus workload) within relatively short experimental scenarios.

Dependent measures Many variables have been collected, but here only the workload ratings, and the complexity metrics introduced above will be briefly discussed; see (Abdul Rahman, 2014) for details. Note that to rule out any 'fade in' and 'fade out' effects, the first 3 minutes and the last 2 minutes of each 25 minutes run were excluded, Figure 3.

The SD metric is equal to the total number of aircraft (N_{ac}) that fly through the sector, computed every minute. The SSD metric used was the mean area of the SSD of all aircraft in the sector, computed every minute (Hermes et al., 2009). Two DD metrics were computed: the NASA$_1$