AIAA Guidance, Navigation, and Control Conference
2 - 5 August 2010, Toronto, Ontario Canada

AIAA 2010-7542

The Use of the Dynamic Solution Space to Assess Air Traffic Controller Workload

J. G. d'Engelbronner,[*] M. Mulder,[†] and M. M. van Paassen[‡]
Delft University of Technology, Delft, The Netherlands
S. de Stigter[§]
To70 BV, The Hague, The Netherlands
H. Huisman[¶]
Air Traffic Control The Netherlands, Schiphol-Oost, The Netherlands

Air traffic capacity is mainly bound by air traffic controller workload. In order to effectively find solutions for this problem, off-line pre-experimental workload assessment methods are desirable. In order to better understand the workload associated with air traffic control, previous research introduced the static Solution Space as a possible workload metric. The Solution Space Diagram is a mapping of intruding aircraft trajectories to the velocity/heading plane in the form of Conflict Zones and safe areas. Choosing a velocity vector in either one will provide an unsafe or a safe solution, respectively. In this paper an improved, dynamic Solution Space will be tested for correlations with air traffic controller workload, measured experimentally. A two dimensional experiment has been conducted, where subjects were required to line up all aircraft in a sector towards a certain waypoint, while continuously providing subjective workload ratings. High correlations were found between several Solution Space parameters and the subjective workload. Even though a conventional workload metric shows also to be highly correlated to the measured workload, the Solution Space could be the scenario independent workload metric that is currently missing in air traffic controller workload determination.

I. Introduction

While the world bows for the inevitable forces of the global credit crunch, this same crunch is attempting to take over the lead position in a list of possible reasons why air traffic cannot grow as fast as would be expected or might be desirable. The assumed number one position on that list for a great number of years, and still a reason worth exploring even now, has been Air Traffic Controller (ATCo) workload.[1–3]

ATCo workload has been recognized as having an important influence on airspace capacity and safety and the effect of changes to Air Traffic Control systems on workload must be assessed. This can be done by means of expert reviews of the proposed changes and metrics on the basis of off-line, fast-time simulations and workload measurements in real-time, manned simulations.

The use of an objective metric in fast-time simulations could provide a robust method of ATCo workload assessment while keeping costs low. This motivates the development of such metrics and fuels the need for a metric that can reliably predict ATCo workload.

In order to get an insight in the workload of an ATCo, it is common practice to observe the task demand load, being the objective, subject-independent aspects of workload, instead of the workload proper.[4a] When trying to find

[*]MSc candidate, Control and Simulation Division, Faculty of Aerospace Engineering, Delft University of Technology. Kluyverweg 1, 2629 HS Delft, The Netherlands. E-mail: Jdengelbronner@aol.nl

[†]Professor, Control and Simulation Division, Faculty of Aerospace Engineering, Delft University of Technology. Senior Member AIAA. E-mail: m.mulder@tudelft.nl

[‡]Associate professor, Control and Simulation Division, Faculty of Aerospace Engineering, Delft University of Technology. Member AIAA. E-mail: m.m.vanpaassen@tudelft.nl

[§]Aviation Business Consultant, To70 BV. E-mail: sebastiaan.destigter@to70.nl

[¶]Human Factors Expert, Air Traffic Control The Netherlands (LVNL). E-mail: h.huisman@lvnl.nl.

[a]Task demand load refers to the demands imposed by the task, i.e. independently of the person performing the task. Workload refers to the task load as experienced by the operator.

an objective measure for the ATCo task demand load, most researchers have assumed that ATCo task demand load is coupled to sector complexity. Since task demand load in turn is coupled with workload, sector complexity should have a strong relationship to ATCo workload.

A rather straightforward method of determining sector complexity is looking at the aircraft count (or aircraft density). When only looking at individual sector characteristics, aircraft count has the highest correlation with subjective workload ratings.[5,6] The weakness in using this method lies in the fact that it ignores other important airspace parameters. Moreover, the relationship between aircraft count and workload is non-linear.[7,8]

In order to solve this problem, methods like the Dynamic Density[5,6,9–11] and the Traffic Load Index[12–14] were developed. These methods show more promising results in estimating ATCo workload then the Aircraft Count. However, both methods rely on the use of an expert to calibrate the method to a specific situation, thus making the methods undesirably specific. A need has arisen to find a complexity metric that is generally applicable to a large number of Air Traffic situations.

In this light, complexity maps, as have been developed by Lee et al.,[16] and the Solution Space, as was introduced by Hermes et al.,[17] were proposed as complexity metrics.

1. *Forbidden Beam Zone*

The Solution Space metric relies heavily on work done by Van Dam et al. on the Forbidden Beam Zone (FBZ), as can be found in Figure 1. The FBZ was originally developed by Van Dam et al.[15] as a means of self-separation for use in the cockpit. By definition, choosing a velocity/heading combination inside the FBZ will lead to a separation violation. The zone is built up by constructing two tangent lines from the controlled aircraft to the separation circle of the observed aircraft and then translating these lines by the velocity vector of the observed aircraft.

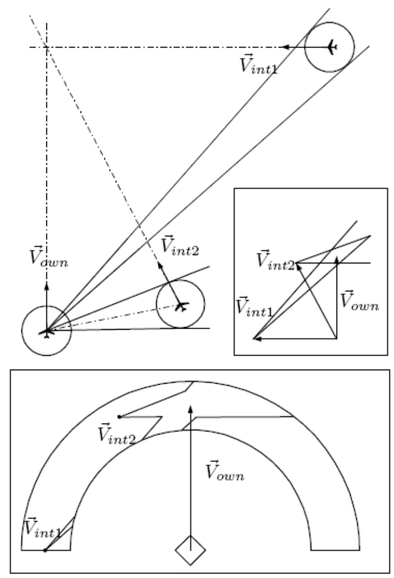


Figure 1. Forbidden beam zones[15] - For an observed aircraft (own), other aircraft (int1, int2) in its vicinity produce a region (beam) in relative velocity space that must be avoided. These regions are translated with the velocity of the intruder aircraft, to create regions in the absolute velocity/heading space of the own aircraft that must be avoided (insert top right). This representation is combined with the possible speed ranges for the own aircraft (bottom), to obtain a display for self separation.

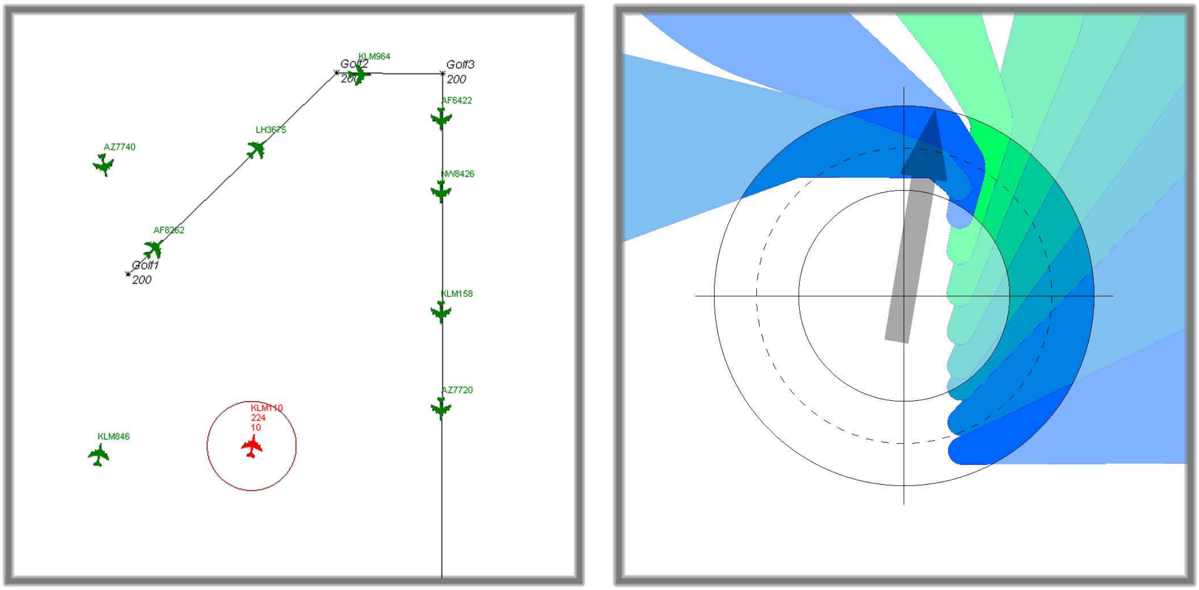
2. *Solution Space*

In the work of Hermes, the FBZ was introduced in the field of ATCo workload determination as the Solution Space.[17] The Solution Space Diagram covers all heading/velocity combinations, indicating which velocity vectors offer 'safe solutions' and which velocity vectors lead to an impending conflict with another aircraft. The hypothesis was made that the state of the Solution Space Diagram was strongly correlated to sector complexity. To prove this concept an experiment was conducted, where both experienced subjects and unexperienced subjects tried to merge one free aircraft in a stream of on-route aircraft.[17] A strong correlation was found between the initial static Solution Space and the subjective complexity rating given by the test subjects. However, the described experiment only covered a very specific case. Due to the merging character of the experiment, the intent of the free aircraft was completely known, being to merge on the pre-defined route. This enabled the elimination of all 'non-route-bound' velocity vectors as being part of the Solution Space. Moreover, due to the fact that the test subjects were instructed to merge on the closest route segment and not to cross it, it was known beforehand that route segments that were 'behind' other route segments were not part of the Solution Space. Finally, only the initial Solution Space was considered; all scenarios were set up to have an initial opening in the Solution Space that would safely merge the aircraft. However, it is quite possible to think of a rather simple scenario, that has no initial solution, but is in fact quite easy to solve.

To solve these problems, the tangent-based calculation of the Solution Space is introduced in the present project.[18] This calculation is sufficiently fast to be suitable for dynamic Solution Space measurements. The tangent-based Solution Space is constructed using Conflict Zones (CoZ's) that project the intent information of an observed aircraft onto the Solution Space Diagram. It has also become possible to limit a CoZ to be calculated for a limited time span.

Figures 2(a) and 2(b) show an air traffic situation in the Cartesian plane and the related tangent-based Solution Space respectively. The Solution Space consists of multiple CoZ's, where each CoZ can be associated with an aircraft in the airspace. In this specific case, all observed aircraft have the intent of flying the route that has been displayed.

This intent causes the CoZ's on the Solution Space Diagram to have turns in them that correspond to anticipated turns in the Cartesian plane.



(a) An air traffic situation in the Cartesian plane

(b) A Solution Space Diagram with multiple CoZ's

Figure 2. The tangent-based Solution Space describes possible conflicts with other aircraft, given their intent information

II. Hypothesis

It is hypothesized that the Solution Space is strongly related to ATCo workload. Due to the state of the Solution Space development, previous work was limited to finding a relationship between the initial Solution Space and sector complexity in a severely simplified and conditioned ATC environment.

For this paper, it is suspected that a direct relationship exists between workload and the Solution Space at each moment in time. The goal is to find the strongest possible relationship between Solution Space parameters and a subjective workload measure in a highly dynamic environment.

III. Dynamic Solution Space Parameters

The Solution Space as a diagram is not suited for direct implementation as a complexity metric, because it only contains visual information. In order to extract useful information from the Solution Space Diagram, a number of parameters can be defined that describe certain aspects of the Solution Space. Because the Dynamic Solution Space is a term that contains a vast amount of data, the Dynamic Solution Space in this paper has been limited in frequency and in diversity.

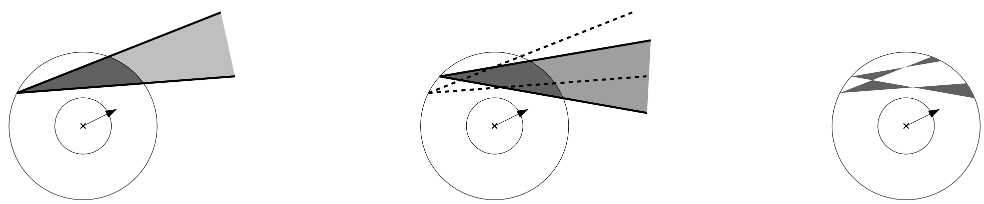
The Solution Space has been calculated every 20 seconds for each aircraft present in the air traffic sector. Subsequently various Solution Space parameters, which will be defined in the following sections, have been calculated for each aircraft. For the statistical analysis all Solution Space parameters have been averaged over all aircraft, thus implicitly assuming that each aircraft has an equal influence on the workload. The data has also been re-sampled to have a frequency equal to the workload scores that have been measured once every minute by averaging the parameter scores per minute.

The measured parameters can be divided in two different basic parameter types, 10 different Solution Space calculation options and 6 different observation angles. Adding the measurement of two baseline parameters, being the Number of Aircraft and the Number of Commands, this makes a total of 122 measured parameters. All solution parameters have been named according to their option, parameter type and their observation angle. The parameter corresponding to the Area parameter of option 9 with a 45° observation angle for instance, has been named Option 9 Area 45. Next to the continuous parameters, the number of separation violations per experiment run has also been measured.

A. Area and Change

Conclusions from previous work by Hermes et al.[17] stated that the area in the Solution Space that offers solutions has a strong (inverse) correlation with ATCo workload. For this experiment the area that is covered by the CoZ's was calculated as a percentage of the total available area between the minimum and the maximum velocity line in the Solution Space Diagram. This area was corrected for the fact that the Solution Space is circular. If this had not been done, the Solution Space area would have been influenced more by high velocities then by low velocities. The Area parameter thus represents the percentage of velocity vectors in the given heading/velocity envelope that will lead to a conflict according to the Solution Space Diagram.

The use of a Dynamic Solution Space makes the calculation of rate properties possible. Next to the Area parameter, a Change parameter has been introduced. The Change parameter is the percentage of the Solution Space Diagram that changed state from safe to unsafe or vice versa. This is illustrated in Figures 3(a) to 3(c), where it can also be seen that a rotation gives almost no change in the Area of the solution space, but does give a value for the Change parameter.



(a) Initial Solution Space (b) Next Solution Space calculation (c) Resulting "Changed" Area

Figure 3. Definition of the Change parameter

Figures 4(a) to 4(d) present an example of the Area and Change parameter behavior in an experiment run next to the two baseline parameters. Each Figure contains all measurement points for one individual test subject, where the scenario scores have been presented consecutively. It can be seen that all scenarios started with a relatively low workload (or ISA) score after which the workload increased to a higher level. Figures 5(a) to 5(d) present the same dataset. The measurement points have been presented as a spread, were it becomes clear that the Area parameter and the Number of Aircraft parameter produce more clustered results than the other two parameters. Later it will become clear that these parameters are indeed more strongly correlated to the workload then the Change and the Number of Commands parameter.

B. Different Options

The different options that have been tested are briefly summarized in Table 1. In option 1, the instantaneous Solution Space is calculated. This means that only FBZ's, as were defined in the research by van Dam et al., are plotted on the Solution Space Diagram. Option 2 uses the complete CoZ as was defined by d'Engelbronner et al. In calculating these CoZ's, the logged data was used as intent information. This means that this option assumes the subject to have planned all his decisions beforehand. Options 3 to 5 use time limited versions of the CoZ. This means that only possible conflicts that occur within a certain amount of time show up on the Solution Space Diagram.

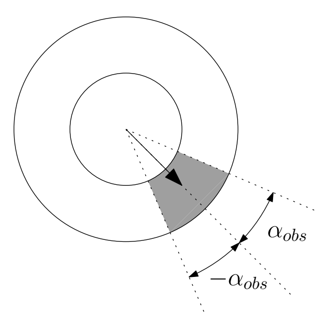


Figure 6. Only that part of the Solution Space that lies within the observation angle is regarded

These options thus assume that the subject has an intent for a certain amount of time. Options 6 to 8 use action limited versions of the CoZ. This means that the intent information of each aircraft is limited to a certain number of actions, after which the aircraft is assumed to maintain constant speed and heading. These options assume that the subject has an intent for a certain number of actions. Options 9 and 10 are a combination of the latter two methods, where the CoZ is limited in time and actions.

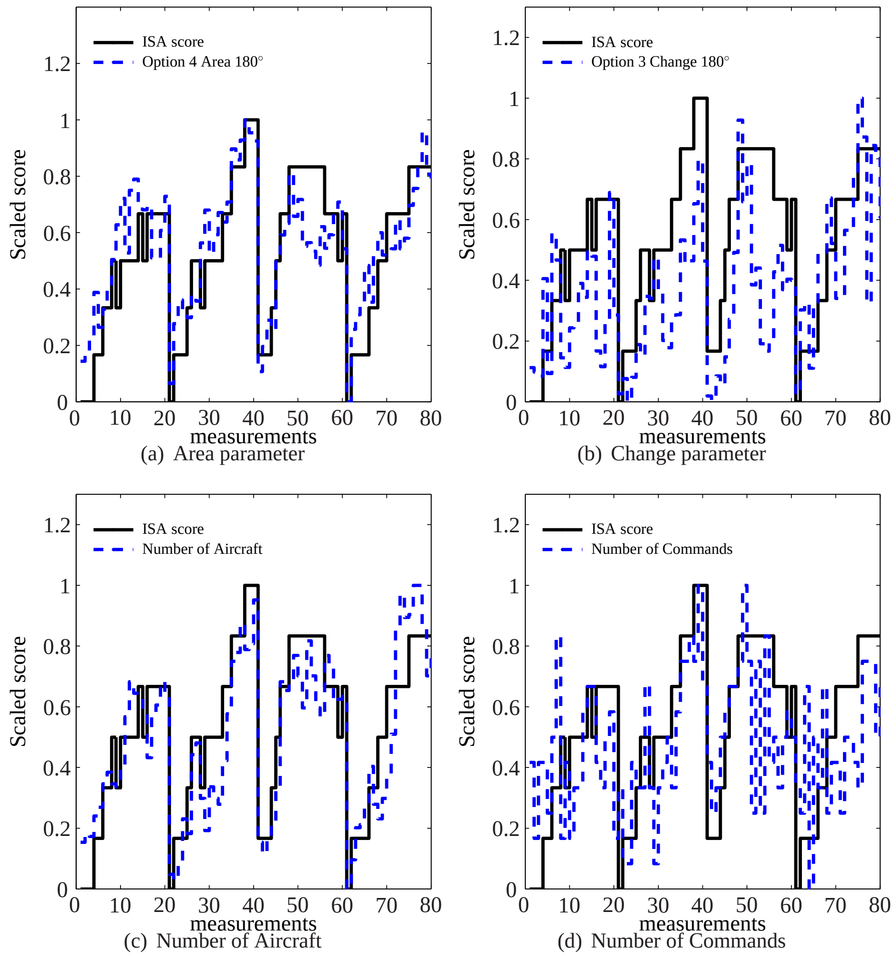


Figure 4. Measurement points for a single test subject (all scenarios)

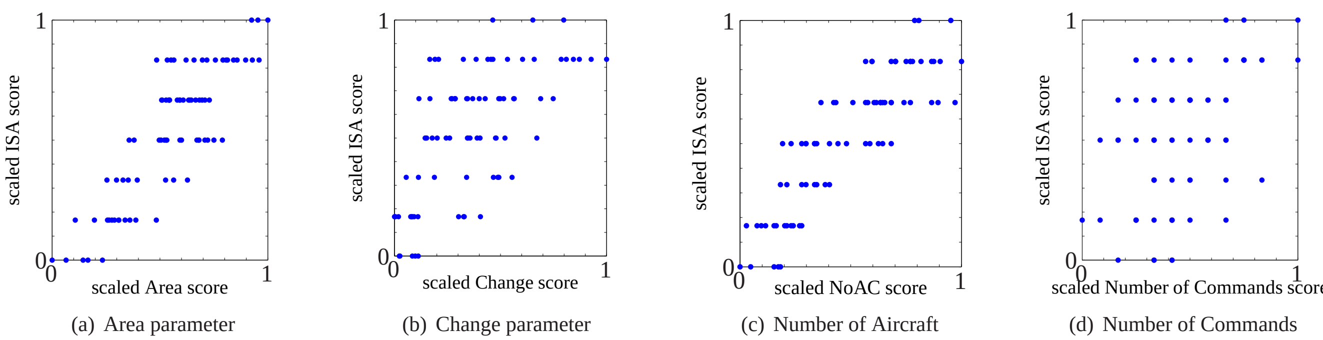


Figure 5. ISA score results plot against four measures derived from the experiment data. Data from a single test subject.

Table 1. Option characteristics. Time is given in simulated time.

Option	description
1	Instantaneous (only FBZ)
2	Omniscient (full CoZ)
3	Time limited (5 minutes)
4	Time limited (10 minutes)
5	Time limited (15 minutes)
6	Action limited (1 action)
7	Action limited (2 actions)
8	Action limited (3 actions)
9	Combined (2 actions, 10 minutes)
10	Combined (2 actions, 15 minutes)

C. Observation Angle

Besides the different options and parameter types, the effect of an observation angle has also been investigated. The objective is to determine whether the region around an aircraft's current velocity vector is more important in determining the Solution Space than the rest of the Solution Space. In order to test this, a range of observation angles has been tested, being 5, 10, 20, 45, 90 and 180 degrees.

The observation angle is defined as the semi-sided angle relative to the velocity vector. For example, an observation angle of 90 degrees means that half of the Solution Space has been taken into account. Both the Area and the Change parameter percentages have been adjusted to take this limitation into account. This means that a full observation area always corresponds with a score of 100 and an empty observation area always corresponds with a score of 0, given the fact that both parameters are defined as a percentage of the total available Solution Space area.

IV. Experiment

To investigate possible relations between the dynamic Solution Space and the ATCo workload, an experiment has been conducted that involved managing air traffic in a simulated environment. In this experiment subjects were confronted with the task of directing all inbound traffic in an air traffic sector to a single merge point. The goal of the experiment was to find a relationship between the subjectively measured workload and an objective parameter, obtained from the dynamic Solution Space. The Solution Space has been analyzed both for parameters through time and for the predictive character of those parameters.

A. Workload

In the experiment as conducted by Hermes et al., the workload was measured as a subjective rating provided by the test participant after each scenario. This method gives one single workload score per scenario, which is a good way of measuring workload for a time independent problem. However, the dynamic problems encountered in Air Traffic Control have a time dependent character. The workload is likely to vary over time because of changes in the situation the ATCo is dealing with. As a result, the workload should be viewed as a time dependent variable and should thus be measured at several points in time.

1. Instantaneous Self-Assessment

A method that is used in real time Air Traffic simulations is the ISA (Instantaneous Self-Assessment) method for workload determination.[19,20] This method, developed by the UK Civil Aviation Authority, uses concurrent workload evaluations to measure the subjective workload during an experiment. This can be done by either supplying the participant with a number of buttons to push every couple of minutes, or by evaluating the workload verbally. It has been suggested that the method is slightly intrusive when used in an experimental setup, however, the added insight in changing workload counterbalances this argument.

The big advantage of using the ISA method is that it is a simple way of evaluating the subjective ATCo workload over time, which cannot be said for physiological methods or the dual tasks method, which does measure workload over time, but is not simple. Therefore the choice has been made to use the ISA standard as a workload measure for this experiment. The subjects were asked to give a subjective workload rating every minute during the experiment on a scale from 1 to 7. The subject was asked to give a workload score by means of a pop-up window in the simulation that did not make controlling of the aircraft impossible. The ISA measurement was taken on a standard keyboard, where keys s,d,f,g,h,j and k were numbered from 1 to 7, ranging from easy to difficult. These keys were also color coded, ranging from pure green for 1 to pure red for 7.

2. Questionnaire

Between scenarios, the subjects have also been asked to answer a number of questions regarding the experiment. As part of this questionnaire, the subjects were asked what workload they experienced for each scenario, ranging from 1 to 7, on an scale equivalent to that of the ISA measurement. They were also asked whether the simulation speed, the aircraft mix or the sector size made the scenario particularly difficult. For each scenario there was the opportunity to provide comments on scenario properties.

B. Subjects

A total of twelve test subjects participated in the experiment. Four of these test subjects had ATC experience, four people had participated in an extensive introductory ATC-course and four were students at the faculty of aerospace engineering of Delft Technical University with no prior ATC experience. The group of four with ATC experience consisted of three male subjects and 1 female subject, ranging in age from 42 to 57 ($\mu = 51.3$, $\sigma = 6.9$ in years). This group had professional experience as an air traffic controller ranging from 6 to 35 years. The group of four that did the introductory ATC-course consisted of three male subjects and 1 female subject, ranging in age from 27 to 47 ($\mu = 38.5$, $\sigma = 8.5$ in years). The four students at the faculty of aerospace engineering where all male and ranged in age from 24 to 26 ($\mu = 25.5$, $\sigma = 1.0$ in years).

C. Experiment Set-up

The test subjects were presented a simplified, two-dimensional representation of an air traffic environment, where all aircraft move in plane with no altitude separation. The environment consisted of an air traffic sector and multiple aircraft as is shown in Figure 7. In this environment the user was capable of selecting aircraft by selecting an aircraft in the plan view display and controlling that aircraft either by using the control options in the right hand window or by giving keyboard inputs similar to those used by air traffic controllers.

1. Simplified Representation

To minimize a possible training effect when working with the simulator, several simplifications have been made with respect to a normal ATC environment. The aircraft types have been limited to two categories, being medium and heavy aircraft. The distinction between these two types was made in the plan view display by presenting a larger or smaller icon. The separation circle of 5 NM was given at all times as an indication of the separation distance. Separation violations were made visible by means of the aircraft color, which switched from green to red, and a visual message in the top left corner of the plan view display.

D. Instructions

All subjects that participated in this experiment were briefed on the goal of the experiment. The subjects were instructed to try and avoid separation violations at all times, while lining the aircraft up at the merge point. As a secondary requirement the subjects were instructed to merge the aircraft at a reasonable merge velocity. During the two training scenarios, that lasted 15 minutes each, the subjects were encouraged to try controlling the aircraft both with the keyboard inputs and with the control options on the right hand window of the simulation screen using the mouse. They were made actively aware of the possibility to give an aircraft a direct-to command to the merge point and they were told to attempt flying out of the sector without crossing the merge line, so they would experience the fact that an aircraft would return to the sector immediately when given a wrong exit point. Subjects were also instructed that it was not allowed to rest their hand on the ISA buttons, in order to prevent the subject from unconsciously giving the

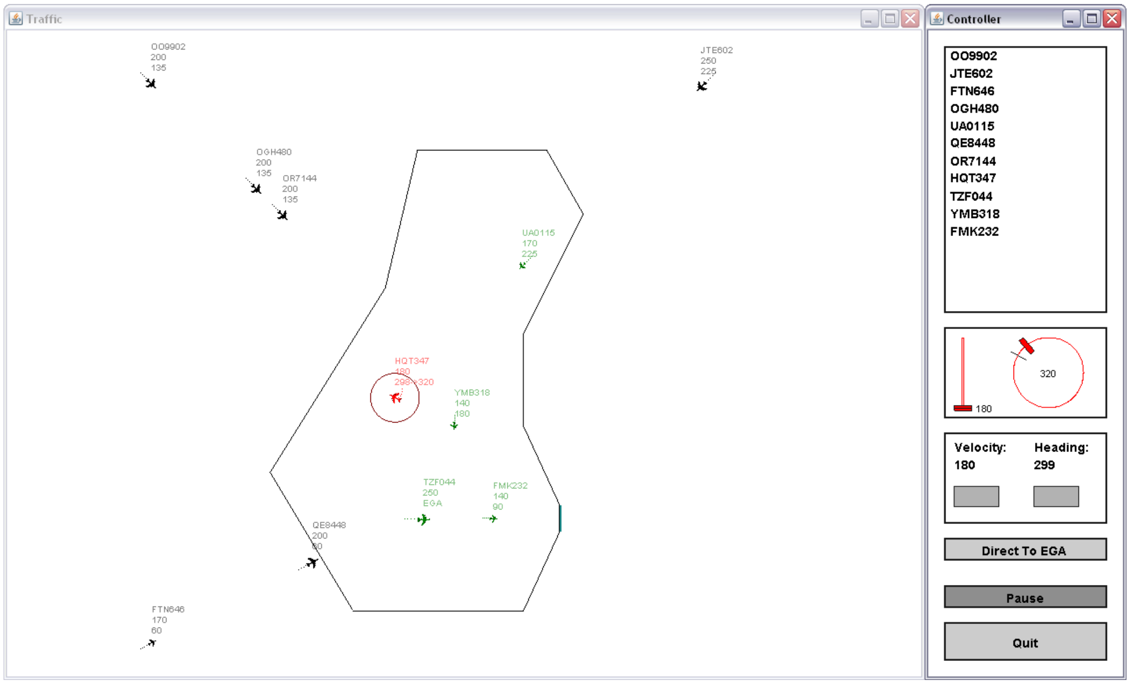


Figure 7. Typical experiment sector - the subject is confronted with multiple medium and heavy aircraft

same ISA score over and over again just to get rid of the pop-up. The two training scenarios were flown subsequently, but between each actual experiment scenario subjects were given the option to rest for up to 10 minutes.

E. Scenarios

The experiment consisted of four different scenarios, excluding the two training scenarios. The choice of using different scenarios has been made for two reasons. First, because of the dynamic character of the experiment, it was expected that a number of test subjects would not be able to handle the steady flow of aircraft coming in, leading to a build up of aircraft in the sector. This could potentially lead to a situation where control of the aircraft would become impossible. By splitting the experiment into four 20 minute scenarios, the start of a each new scenario would be identical for each test subject and would thus ease the workload for those subjects that could not manage. The second reason to use scenarios was that this would give the opportunity to compare different scenarios. Different subjects have been offered the scenarios in a different pre-defined order, Table 2, to prevent any possible dependency to training effects. On top of that, the scenario order has been distributed equally among different subject groups.

Table 2. Scenario order

	Subject											
	1	2	3	4	5	6	7	8	9	10	11	12
Run1	A	B	C	D	A	C	D	B	A	D	B	C
Run2	B	A	D	C	C	A	B	D	D	A	C	B
Run3	C	D	A	B	D	B	A	C	B	C	A	D
Run4	D	C	B	A	B	D	C	A	C	B	D	A

Each scenario has been set up to have a distinct character. Scenario A had both heavy and medium aircraft coming in from random directions all around the sector. Scenario B had three streams of incoming aircraft with mixed aircraft on all streams. Scenario C had two streams of incoming heavy aircraft and one stream of incoming medium aircraft. Scenario D had two streams of incoming heavy aircraft and a number of aircraft passing through the sector that could not be controlled by the test subject.

Scenarios have not been calibrated to be either difficult or easy. All scenarios presented the subject with just over 40 aircraft in 20 minutes and the level of difficulty depended not only on the scenario set up, but also on the ATC strategy employed by the subject.

Dynamic Environment

The fact that the scenarios lasted 20 minutes each resulted in a dynamic environment, much like an actual ATC environment. The actions of the subject highly influenced the situation that developed in the air traffic sector. Due to the fact that the subject himself was in the loop, each experiment run was unique and each scenario resulted in different situations for different subjects. This behavior of the simulated environment complicated the interpretation of the data, due to the fact that results could not easily be compared with the scenarios as a constant factor. However, since this is also the case in actual ATC situations, this has deliberately been chosen to be an integral part of the experiment.

F. Simulation Characteristics

As was stated before, the experiment included two types of aircraft, representing medium aircraft and heavy aircraft. The velocity envelopes of these aircraft has been given in Table 3, which also includes the required merge velocities at the merge line, as was referred to in the instructions section. The velocity envelopes of the medium and the heavy aircraft did not overlap in order to prevent subjects from putting all aircraft on the same velocity and thus eliminating the effect of having two different aircraft classes.

Table 3. Velocity Envelopes

Class	Velocity
Heavy	180 - 250 kts
Medium	100 - 170 kts
Merging	150 - 200 kts

The simulation used for this experiment used only kinematics for the aircraft behavior. All aircraft had a constant turn rate of three degrees per second and a constant acceleration rate of three knots per second. All velocities were given as indicated airspeed, but since there was no wind included in the model and the altitude was constant the relation between airspeed and groundspeed was a constant factor.

Aircraft that left the sector without having crossed the merge line were automatically sent back into the sector, not taking into account possible conflicts this could cause in the sector. As was said before, subjects were made actively aware of this conduct.

G. Simulation Speed

The simulation was run in fast-time, being at four times real time speed, as was previously done by Hermes et al.[17] The rationale behind running the experiment in fast-time was that this results in the acquisition of more data from the given experiment time. A positive side effect is that this prevented the subjects from underachieving out of boredom. A negative side effect could be that the subject experiences the task different then he would in real-time, which would be an undesired effect influencing the experiment results.

V. Results

A. Outlier Analysis

A total of five outliers were detected for the 960 workload measurements that have been taken during the experiment.

As can be seen in figure 8, four outliers can be found in the ISA measurements for subject 6 and one can be determined for subject 9. Closer inspection of these outliers revealed that the high ISA scores found for subject 6 are consistent with other high scores that this subject gave consecutively. Also, these values seemed to be consistent with a high Number of Commands. The outlier for subject 9 is much lower than the other ratings. However, that minute, the subject also gave no commands. Both the Number of Commands and the Number of Aircraft were found to have their minimum value at that specific measurement point. This suggests that the workload was probably indeed very

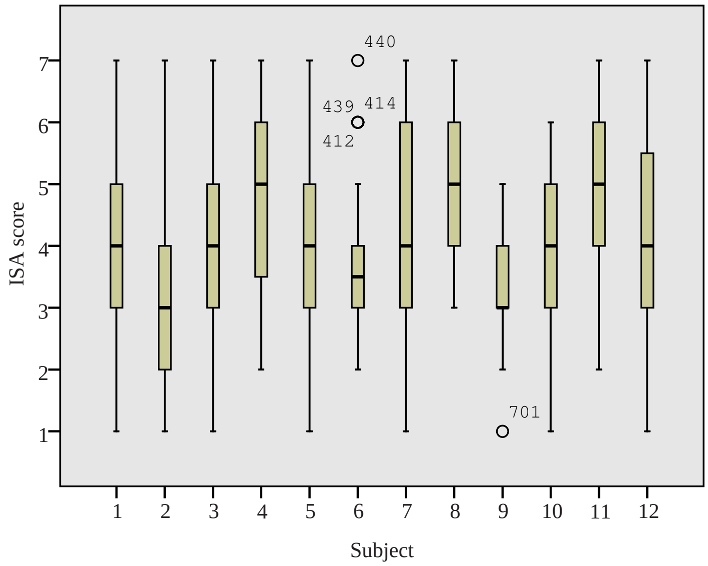


Figure 8. ISA score distribution per subject

low. Because no unexplainable outliers were found, all these outliers were included in the analysis and accepted to be a part of the subject's behavioral pattern.

B. Group Differences

Before processing the data, it is important to investigate whether the different groups behaved similarly and can thus be added together, or if groups differed and should thus be treated individually. The results in this paper have been compared according to ISA score, Number of Aircraft and Number of Commands. First of all, all these parameters were found to have non-normal distribution for all groups. Using the Kolmogorov-Smirnov test[21] it was found that this effect is significant ($p < 0.01$) for all parameters.

Given the fact that the parameters are not normally distributed, a non-parametric approach had to be used in comparing the different groups. Using the Kruskal-Wallis ANOVA,[21] no significant differences were found between groups for the Number of Commands. However, significant differences were found between groups concerning the ISA scores and the Number of Aircraft ($p < 0.05$).

In order to find the source of this significant difference, the Mann-Whitney test[21] was used as a post hoc test to determine differences between groups pairwise. Because three separate Mann-Whitney tests were used, the significance threshold had to be lowered to a value of 0.0167 ($= 0.05/3$) in order to prevent increasing the chance of having a Type I error. This is called a Bonferroni correction.

From these separate post hoc tests, it became clear that the students gave higher ISA scores than the other two groups, while the group with ATC experience had more aircraft in their sector. Both these effects were found to be significant. Table 4 shows the mean and median of the two tested parameters per group.

Finally, the number of separation violations per group has been evaluated using a Kruskal-Wallis test. However, no significant difference between groups was found.

Table 4. Parameter characteristics per group

Parameter		ATCo	ATC-course	Student
ISA	mean	4.04	3.87	4.39
	median	4	4	4
Number of Aircraft	mean	9.91	9.13	9.20
	median	9.79	8.75	8.96

C. Subject Differences

Taking the analysis of different groups one step further requires inspecting the groups at the subject level. Using a Kruskal-Wallis test, it was found that within the student group and the experienced ATCo group the Number of Aircraft the subjects were confronted with did not differ significantly ($p > 0.05$). For the ATC-course group this parameter did differ significantly among subjects ($p < 0.05$). Both the ISA score and the Number of Commands had significant differences between subjects for all subject groups. This makes sense, because every subject is expected to have a different control strategy and a different interpretation of the subjective workload score.

D. Training Effect

In order to establish whether a significant training effect was present in the experiment, the known estimators for ATCo workload, being Number of Aircraft and Number of Commands, and the ISA score were tested for differences between runs using the Kruskal-Wallis test. They were also tested for significant trends using the Jonckheere-Terpstra test. It was found that neither the ISA score, nor the Number of Aircraft showed trends or differences between runs. The Number of Commands, however, was found to be different between runs. Also a trend was found where the number of commands given per minute increased over the course of the experiment. This results was significant, both for the Kruskal-Wallis test ($p < 0.001$) and the Jonckheere-Terpstra test ($p = 0.007$). Figure 9 shows that the trend does not show in the median data. The fact that the mean data does not differ more than one command per minute between scenarios resulted in the conclusion that no training effects were present in the dataset.

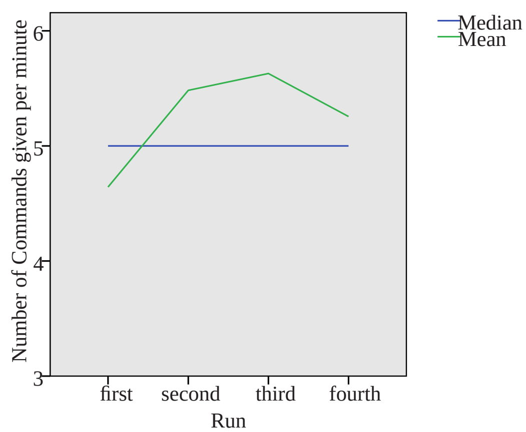


Figure 9. Slight training effect in the Number of Commands given per minute

E. Correlation Analysis

In statistical research it is common practice to negate subject differences by calculating standardized z-scores and then considering the whole of the normalized subject measurements. However, since the ISA score is an ordinal parameter, this practice would theoretically be incorrect in the light of this research. The step size between different ISA values is potentially not constant, because of potential non-linear use of the scale. Adding all measurements up could lead to a lower correlation result. By first calculating an individual correlation and selecting parameters through this method, potential erroneous discarding of parameters can be prevented.

The fact that the data was non-parametric, as has been proven before, means that all correlations have been calculated using Kendall's tau instead of its parametric counterpart, the Pearson correlation. Please note that this method gives correlations that cannot be compared to correlations obtained via Pearson's method due to the differences between the methods.

1. Individual Correlations

In the light of the differences between groups and subjects the choice has been made to observe the correlations between different parameters on an individual subject level. The advantage of such an approach is that, although the subjects differ in ATC strategy, this approach allows for comparison of the relation between workload and a parameter on the subject level. This method resulted in 122 different correlations per subject (one per parameter, including Number of Aircraft and Number of Commands). With 12 subjects this resulted in 1464 different correlations.

Pre-selection

Calculating the correlation on the subject level also reduces the concept of significance to the subject level. Of all the individual correlations, 1135 turned out to be significant ($p < 0.05$), while 329 were not significant. Table 5 shows that a large portion of the parameters had either 12 or 11 significant individual correlations.

Table 5. Number of parameters per number of significant correlations

Significant	12	11	10	9	8	7	6	5	4	3	2	1	0
Parameters	54	21	8	3	4	7	4	1	4	4	3	5	4

Since a non-significant correlation is not worth evaluating, the assumption was made that a parameter with less than a certain number of significant individual correlations probably does not have a significant correlation as a whole and should be discarded. On top of that, all parameters with a large number of insignificant correlations had a relatively low value for the individual correlations. The choice was made to include only those parameters with nine significant correlations or more, because all other parameters showed to have very low correlations. Since the objective of this research was to find a strong relationship between an objective parameter and the subjective workload, discarding parameters did not interfere with the final goals. This pre-selection eliminated 36 parameters, leaving 86 for further analysis.

Compare Groups

With 86 different parameters, it was now possible to compare the correlations of the different groups. The hypothesis here was that the three groups differ in ATC strategy, but that they all have a similar correlation between the parameters and the ISA score. After doing a Shapiro-Wilk test it became clear that in 21 cases out of a total of 258 cases (3 groups, 86 parameters) the distribution within a group was significantly different from the normal distribution ($p < 0.05$). Because these distributions made up over 8% of the parameters, the choice was made to use a non-parametric test in finding differences between the three groups of subjects.

Use of the Kruskal-Wallis test showed that 10 out of 86 parameters had a significant difference ($p < 0.05$) between groups. These 10 parameters included the Number of Commands given, 3 Area parameters for various cases and observation angles and 6 Change parameters for various cases and observation angles. Using a Mann-Whitney test as a post hoc test on the Number of Commands showed that the experienced ATCo group had higher correlations for this variable than the other two groups. The most interesting remaining parameters were the Option 3 Change 45 degree and 90 degree parameters because of their relatively high correlations with the ISA score. Using a Mann-Whitney test as a post hoc test on these parameters showed that the experienced ATCo group had lower correlations for these change variables then the other two groups. Both the result for the Number of Commands and the result for the Change parameters were significant ($p < 0.05$).

Angular Influence

The influence of the observation angle, as was described before, was found to have a strong trend towards higher correlations for larger observation angles.

Figure 10 shows this relation for option 3, for both the Area and the Change parameter. For all Area parameters it was found that the 180 degree observation angle yielded the highest mean and median correlations per case. All these parameters did not have a significant difference between groups. The fact that for all cases the 180 degree observations gave better results for the Area parameters than the lower observation angles motivated the choice to continue only with the 180 degree Area parameter results, discarding all smaller observation angles. This also eliminated the three Area parameters where the groups showed significantly different correlations.

The Change parameters generally showed lower correlations than the Area parameters. The choice has thus been made to include only a limited amount of Change parameters in the analysis. The fact that the Change parameters from option 3 achieved relatively high correlations, the 45 degree, 90 degree and 180 degree variant of the option 3 Change parameter where observed from this point forward. Note that the Change parameters with the 45 degree and

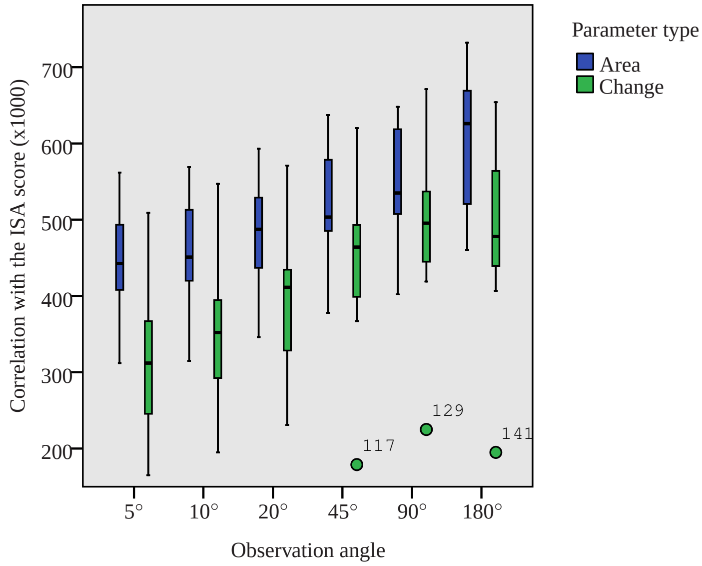


Figure 10. Increasing correlations for larger observation angles

the 90 degree observation angles had lower correlations for the experienced ATCo group. Also note that none of the selected distributions include any insignificant correlations, as were discussed in the pre-selection.

2. Combined Correlations

For the 15 parameters that the parameter set had been reduced to by observing the individual correlations, the standardized z-scores have been calculated in order to compare both methods of calculating the correlation. This meant that all parameter values were reduced to a z-score, which represents the difference (in standard deviations) that particular value has with the subject mean. Resulting z-scores can then be combined into one measurement, thus combining all subject data in one array. Table 6 shows all individual correlational properties and presents the combined correlations that have been calculated using Kendall's tau (where R gives an indication of the strength of the relation between two parameters). The use of this non-parametric correlation coefficient is again necessary, because a Kolmogorov-Smirnov normality test reveals that the ISA z-score distribution is non-normally distributed ($p < 0.001$). It can be seen in Table 6 that the relation between parameters is similar in both methods. The combined correlations do give lower values, but this was to be expected. The results in Table 6 show that the effects of an ordinal ISA score are limited to a scaling of the correlations, where even parameters that showed a significant difference between groups for the individual correlations do not suffer any negative effects when compared to other parameters. In further analysis it can thus be assumed that the different datasets can be added together with the use of standardized z-scores.

3. Cross correlations

The difference between the Number of Aircraft and the Solution Space parameters can be explored by looking at the correlation between these two parameters. The Number of Aircraft is highly correlated with the remaining parameters, ranging from $R = 0.397$ to $R = 0.440$ for the Change parameters and ranging from $R = 0.543$ to $R = 0.601$ for the Area parameters. The Number of Commands was correlated with the Number of Aircraft with $R = 0.249$, was correlated with the Area parameters ranging from $R = 0.175$ to $R = 0.230$ and was correlated with the Change parameters ranging from $R = 0.201$ to $R = 0.223$. Also, it was found that the correlation between Change parameters and Area parameters ranges from $R = 0.414$ to $R = 0.529$.

F. Time Influence Analysis

In the light of the dynamic character of the experiment, the behavior of the correlations has been analyzed for time shifts. Basically this means that the parameters have been shifted relative to the ISA scores. High correlations on

Table 6. Parameters used in the analysis

Parameter		Individual R mean	Individual R median	Individual Sign. difference between groups	Combined R	Combined p
Baseline	Num of A/C	0.621	0.614	no	0.569	<0.01
	Num of Comm	0.299	0.288	yes	0.259	<0.01
Area, 180°	Option 1	0.592	0.603	no	0.545	<0.01
	Option 2	0.583	0.601	no	0.538	<0.01
	Option 3	0.599	0.626	no	0.550	<0.01
	Option 4	0.613	0.627	no	0.566	<0.01
	Option 5	0.597	0.611	no	0.552	<0.01
	Option 6	0.581	0.604	no	0.531	<0.01
	Option 7	0.581	0.615	no	0.532	<0.01
	Option 8	0.577	0.603	no	0.529	<0.01
	Option 9	0.588	0.608	no	0.543	<0.01
	Option 10	0.577	0.598	no	0.531	<0.01
Change, Option 3	45°	0.443	0.464	yes	0.408	<0.01
	90°	0.487	0.496	yes	0.455	<0.01
	180°	0.482	0.478	no	0.450	<0.01

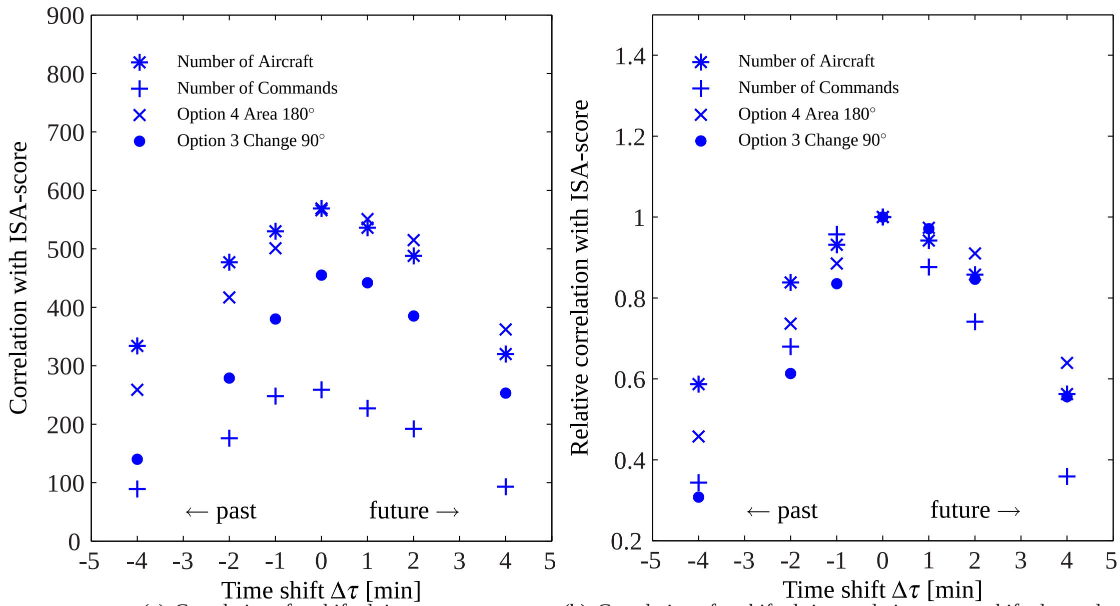
negative shifts will indicate a lagging property in a parameter (workload at $t = t_x$ is reflected by the parameter at $t = t_y$, with $y > x$), while high correlations on a positive shift will indicate a predictive property in the parameter (the parameter at $t = t_x$ anticipates the workload at $t = t_y$, with $y > x$). All parameters have been shifted 1, 2 and 4 minutes in both positive and negative directions ($\Delta\tau$).

Figures 11(a) and 11(b) give an impression of the effect a time shift has on correlations. These figures only include one Area parameter and one Change parameter, but the distributions shown in the figures are typical and thus all other parameters have been omitted from the figures for clarity reasons. A complete table of resulting correlations can be found in the appendix. Figure 11(a) shows that all parameters have a roughly quadratic distribution in their correlations when shifted relative to the ISA-scores. By scaling the correlations to have a maximum value of 1, the distributions in Figure 11(b) shows a reasonably symmetric distribution for the Number of Aircraft and the Number of Commands. The Area parameter stays relatively high when dealing with a positive time shift, while it drops relatively fast for a negative time shift. The Change parameter shows this same skewed shape. These results will be further evaluated in the discussion section.

G. Multiple Regression

In order to determine whether different parameters could be used in a combination to get a higher correlation with the ISA-scores, a regression analysis has been performed on the parameters. The regression analysis assumed some data characteristics, one of them being that the data was either continuous or interval, which was in conflict with the characteristics of the original dataset, because the ISA-scores are ordinal. However, the assumption has been made that the ISA z-scores of all subjects combined approximates a continuous dataset, which enabled the use of aforementioned regression method.

For the regression method, all 15 previously selected parameters were used, assuming that different settings in the Solution Space calculations might account for different parts of the variance. By using a forward entry method, only parameters that made a significant contribution to the model were included. The Number of Aircraft parameter has been taken as a starting point, because it has already been proven to have a strong relationship with ATCo workload. The resulting model can be found in Table 7, where from top to bottom, the parameters are added stepwise, and thus the strength of the model (R) increases. The last column gives the significance of the change in model strength. Note that the correlations in Table 7 are Pearson correlations and can thus not be compared to correlations reported earlier which were Kendall's tau.



(a) Correlations for shifted time

(b) Correlations for shifted time, relative to non-shifted correlation

Figure 11. The influence of a time shift on parameter correlations

Table 7. Regression model build up.

Added Parameter	Resulting R	Resulting R^2	p
1. Number of Aircraft	0.749	0.561	<0.001
2. Option 4 Area 180	0.803	0.645	<0.001
3. Option 3 Change 180	0.809	0.654	<0.001
4. Number of Commands	0.813	0.661	<0.001
5. Option 1 Area 180	0.816	0.666	<0.001
6. Option 3 Area 180	0.817	0.668	0.026
7. Option 3 Change 45	0.818	0.669	0.048

Seeing that the last parameters added relatively little to the strength of the model and that the change of the model, resulting from their addition, had a significance value of $p > 0.01$, the choice was made to remove these two parameters from the model to prevent over-fitting. In the resulting model, both Area parameters (Option 4 Area 180 and Option 1 Area 180) showed high multicollinearity with each other. This makes sense, because the two parameters are both variations on the same concept. One could decide to remove one of the parameters from the model, but given the significant contribution to model strength by both parameters, the choice was made to leave the model as is and continue with a 5 parameter model, consisting of the first five parameters from Table 7. The fact that the only Change parameter in the model is the 180 degree variant is fortunate, because this was the only Change parameter found to have no differences in correlation between groups.

The assumption that this parametric regression method could be used to get an insight in parameter combinations has been verified by calculation of the correlation between the resulting model and the ISA z-scores using Kendall's tau. This correlation is shown in Table 8 next to the strongest correlated parameters found up till this point. It shows that the regression model gives higher correlations than the other parameters, thus verifying that a combination of parameters can be used to obtain higher correlations with ATCo workload.

Table 8. Strong correlations (using Kendall's tau)

Parameter	R	p
Regression Model	0.621	<0.01
Number of Aircraft	0.569	<0.01
Option 4 Area 180	0.566	<0.01
Option 3 Change 90	0.455	<0.01

H. Questionnaire Results

As part of the experiment, the subjects were asked to fill out a questionnaire. The data from this questionnaire was compared to the ISA data and to the Number of Aircraft, because this has already been proven to be strongly related to ATCo workload in previous research. Since the data from the questionnaire were taken per scenario, all other parameters had to be mapped to this resolution. This was done by taking minimum, maximum, mean and median values per parameter. The questionnaire workload score (q-score) showed a weak correlation with the maximum ISA score that was given in the same scenario ($R = 0.269$,$p = 0.029$). All other parameters that were tested against the q-score did not result in significant correlations. Because the Number of Aircraft seemed to be in no way related to the q-scores, while this was to be expected from previous research, the q-score was assumed irrelevant and not to be used in the scenario analysis.

The other three parameters from the questionnaire, being the influence of the traffic mix, the influence of the simulation speed and the influence of the sector size showed no correlations at all. On average, the traffic mix was deemed to have the highest influence on the workload, with simulation speed and sector size getting relatively low scores. To illustrate this, the mean and median values can be found in Table 9.

Table 9. Mean and median scores from questionnaire

Question	workload	mix	speed	sector
Mean value	4,98	4,75	3,88	3,65
Median value	5	5	4	4

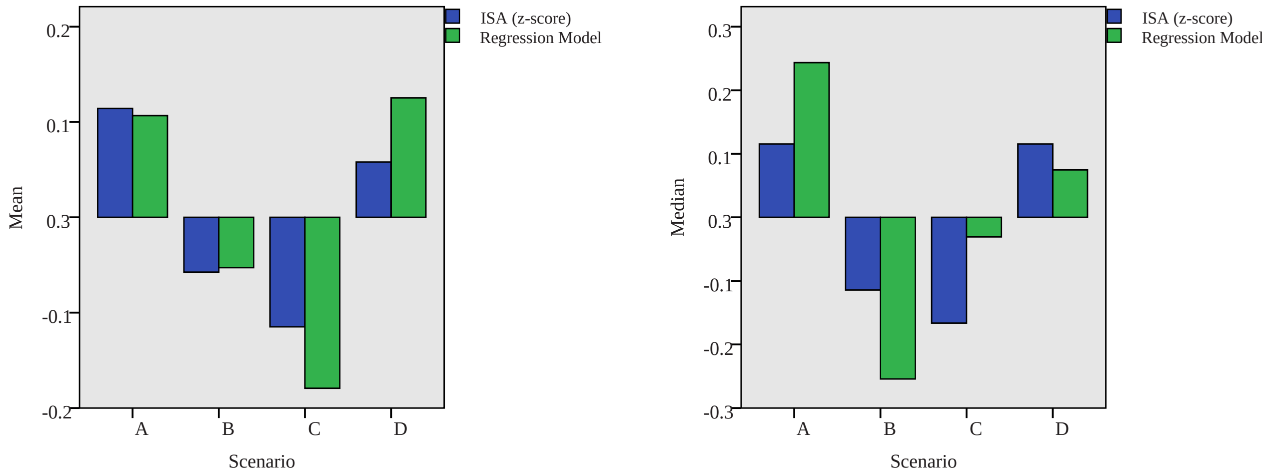
I. Scenario Analysis

Last, the difference in scenario evaluations was tested with Kruskal-Wallis tests. The data from this experiment is non-linear and partly ordinal, thus requiring a non-parametric approach. Both the standardized ISA scores and the results from the Regression Model, as was found earlier, were tested for significant differences between scenarios. Both parameters showed to have a significant difference ($p_{isa} = 0.026$, $p_{model} < 0.001$). The mean ranks observed in this test can be found in Table 10, where scenarios A and D appear to have generated more workload than scenarios B and C. This same difference is reflected in Figures 12(a) and 12(b). However, this result had to be tested in a post hoc test to determine the exact nature of the significant difference between groups.

Table 10. Kruskal-Wallis mean ranks

Scenario	ISA (z-score)	Regression Model
A	515	513
B	459	457
C	449	432
D	499	519

To determine where the differences between groups lie, a one-tailed Mann-Whitney test was conducted. The fact that a Bonferroni correction had to be applied on the significance threshold resulted in the choice to do only 4 tests, comparing the high ranks with the low ranks. This meant that groups where a difference was found with $p < 0.013$



(a) Mean values per scenario

(b) Median values per scenario

Figure 12. Comparison of ISA z-scores and Regression model behavior per scenario

($= 0.05/4$) should be flagged as having a significant difference. Table 11 shows the results of the Mann-Whitney tests. The significance levels have been calculated using a Monte Carlo approximation, hence the added 99% confidence interval.

Table 11. Mann-Whitney test between scenarios with potentially different workloads

Scenarios	p_{isa}	(99% confidence)	p_{model}	(99% confidence)
A & B	0.011	(0.009 - 0.014)	0.016	(0.013 - 0.019)
A & C	0.003	(0.001 - 0.004)	$<$0.001	(-)
D & B	0.073	(0.066 - 0.080)	0.025	(0.021 - 0.029)
D & C	0.022	(0.018 - 0.025)	$<$0.001	(-)

Finally, the number of separation violations per scenario has been evaluated using a Kruskal-Wallis test. Even though Scenarios A and D showed slightly higher numbers of separation violations, no significant difference between scenarios was found.

VI. Discussion

From the outlier analysis, it became clear that the few outliers that were present in the dataset could be explained and should therefore be included in the further analysis. The analysis of a potential training effect only came up with a slight trend in the Number of Commands given. This can possibly be ascribed to the fact that people got more precise in their commands, leading to more adjustments of the course. The fact that neither the ISA score, nor the Number of Aircraft showed a training effect leads to believe that no significant training effect was present in this dataset.

A. Group Differences

Resulting from the experiment, it became clear that some distinct differences exist between groups, both in the subjective interpretation of workload and in the parameters that contribute to that workload. It has been shown that the group consisting of students gave significantly higher ISA scores than the other two groups, while the group consisting of experienced ATCo's dealt with a higher number of aircraft in their sector. Also, a significantly higher correlation between the Number of Commands used and the experienced subjective workload was found for the experienced ATCo group than for the other two groups.

The difference in ISA score between the groups can probably be explained by the fact that the students were less familiar with ATC related problems then the groups with previous ATC experience. The fact that this experiment was, without exceptions, their first experience with controlling aircraft from an ATCo's perspective probably increased their perceived workload. The difference in Number of Aircraft can probably be explained using an observation made during the experiment. Most of the subjects in the experienced ATCo group tried to maintain a fixed pattern where the aircraft must be lined up in order of appearance, while the groups with less experience dealt with incoming aircraft more flexible, adapting to unexpected circumstances better. Resulting from this difference in strategy, the experienced ATCo's had more trouble clearing the sector, resulting in a build up of traffic. Note that the experienced group also had the highest average age. This fact could also have contributed to this discrepancy. The difference in strategy and in workload perception is further underlined by the fact that the ATCo's had a higher correlation between workload and the Number of Commands used than the other two groups. This high correlation can probably be attributed to their professional training. When the number of commands increases, this group feels less comfortable, while an inexperienced subject will not care whether an aircraft gets more or fewer commands. The experienced ATCo group showed a tendency to have lower correlations with the Change parameters then the other two groups.

B. Scenario Differences

From Table 11 it became clear that scenario A was significantly more difficult than scenario C according to both the ISA scores and the estimated workload from the regression model. It also became clear that scenario D was not significantly more difficult than scenario B according to both scores. The comparison of scenarios A and B shows that scenario A was significantly more difficult according to the ISA scores, but not according to the model. However, the significance values lie close to each other, so the model behavior does actually resemble the ISA behavior. In the comparison of scenarios D and C this is not the case however. The model shows that scenario D is significantly more difficult than scenario C, while the ISA scores give the opposite result. This could possibly be caused by the fact that scenario D included intruder aircraft that had a different effect on the ISA score than it had on the model score. These results show that the model could possibly be used for scenario evaluation, but that will have to be improved on first.

C. Correlation Analysis

From the angular analysis of the Solution Space it becomes clear that subjects experienced their workload based on the entire heading band. However, this effect is strongest for the Area parameters. All Area parameters showed to be strongest for their 180 degree case, which shows that subjects use information from the entire heading band in defining their workload. The Change parameter however, did not show an increase in correlation towards the higher heading bands, but showed correlations of the same order of magnitude for the 45, 90 and 180 degree cases. This suggests that a relatively large part of the workload that can be accounted for by the change of the Solution Space can be found in the heading band centered around an aircraft's own heading. Note that the 180 degree case was selected to participate in the Regression model. This parameter was selected for the fact that it accounted for a larger part of the variance that was yet unaccounted for by the model than the other two Change parameters, even though the 90 degree case had a higher correlation with the workload than the 180 degree case.

The actual correlations found between the subjective workload (ISA) and the 180 degree Area parameters showed to be of the same order of magnitude as those found for the Number of Aircraft. This result is less optimistic than the results found in previous work by Hermes et al.[17] on the instantaneous Solution Space. The discrepancy between results found in the work on a static Solution Space and a dynamic Solution Space can most likely be explained through the setup of the experiments. The static experiment was set up for the subject to control one aircraft and merge it in a stream of on-route aircraft. This left the subject with finding a suitable gap to merge the controlled aircraft in. When this gap was found, basically the number of aircraft that where in front or behind this gap did not influence the workload that much, because the task stayed the same: merging the controlled aircraft between two on-route aircraft. In the dynamic experiment however, the subject was confronted with multiple aircraft to control. In most situations, the subject was required to control all aircraft in the sector, which leads to reason that all aircraft in the sector actually contributed to the workload. The fact that there was no predefined route structure present in the sector of this experiment, probably meant that subjects were required to keep monitoring all aircraft in the sector instead of only those that had not been put on a route yet. This might very well have increased the correlation between the Number of Aircraft and the ISA-score even further.

The fact that the Solution Space (Area) parameters show a predictive character, as was shown in the time influence analysis, makes sense, because the Solution Space contains intent information. Where the Number of Aircraft only contains information on the situation in the air space at that particular moment in time, the Solution Space already contains information on situations that lie in the future.

D. Regression Analysis

The regression model that was found earlier by combining several parameters was found to give higher correlations than each parameter individually. However, the strength of the model is not much higher in magnitude than the most strongly correlated individual parameters. This can be explained by the fact that most parameters show signs of multicollinearity. Perfect multicollinearity occurs when a parameter is a linear combination of one or more other parameters and thus accounts for a piece of variance that is already accounted for. The fact that most parameters in this paper are related to the Solution Space creates these kind of problems. When combining multiple Solution Space parameters in one model this fact should be strongly considered. Also the Number of Aircraft shows signs of multicollinearity with the Area parameters, which suggests that these two parameters account for similar parts of the variance. The fact that all parameters are highly correlated with each other also suggests that the Number of Aircraft in the sector strongly influences the Solution Space properties. This makes sense, because each aircraft in the air space will also generate a CoZ on the Solution Space Diagram and thus contribute to the Solution Space Area parameter. On top of that, both parameters are highly correlated with the ISA score, which suggests that the parameters behave similarly.

E. Different Options

From the results it became apparent that the time limited options, being options 3,4 and 5, scored best for the Area parameter. Especially Option 4 showed high correlations with the ISA score. Option 4 had its time limit set to 10 minutes in simulated time. This result indicates that, when confronted with a two dimensional ATC-like environment, a subject will base his workload on the next 10 minutes of simulated time, or 2.5 minutes of real time.

However, the results involving the Change parameter tell a different story. Only several Change parameters from option 3 offered high correlations with the ISA scores. Since option 3 was time limited to 5 minutes in simulated time, this would indicate that, when only observing the change of the situation, a subject bases his workload on the next 5 minutes of simulated time, or 1.25 minutes of real time.

The difference between these two parameters is interesting and suggests that planning of trajectories, as can be represented by the Area parameter of the Solution Space, is an integral part of the workload for the medium time range. Manoeuvring of aircraft only adds to the workload when aircraft are in relatively close proximity to each other, hence the Change parameter only has a high correlation for a short term Solution Space representation. Note that the relationship between real time and simulated time has not been explored for other simulation speeds than four times real time.

VII. Conclusions & Recommendations

It can be concluded that the Solution Space parameters are highly correlated to the measured subjective workload, be it with the same order of magnitude as the Number of Aircraft. However, the Number of Aircraft is usually deemed too limited to be used as a workload metric. The fact that the Solution Space includes heading and intent information probably make it more robust compared to the Number of Aircraft in predicting ATCo workload. Also, the fact that this experiment used a completely free air traffic sector, with no predefined route structure, resulted in a situation where the Number of Aircraft is a very likely predictor of workload, which could have contributed to high correlation between the Number of Aircraft and the workload measurements.

The fact that a high correlation was found between the Solution Space parameters and the ISA scores suggest that continuation of this line of research is required. The current results support the hypothesis that the Solution Space has a strong relationship with the subjective ATCo workload. However, the results are not strong enough and the tested parameters did not cover all possible settings in the Solution Space calculations. A possible increase in correlations could be found in optimization of the time limit options. The test resolution for this experiment was set at 5 minutes, which leaves rather large gaps in the analysis. Also the Change parameter was found to be strongest for the 5 minute option, which suggests that an even tighter time limit could offer better results.

In future experiments, different air traffic schemes should also be explored, in order to test if this presentation of air traffic indeed biased the measurements towards getting high correlations with the Number of Aircraft. The current task focussed mainly on controlling aircraft towards one point, which is a situation where the Number of Aircraft is intuitively strongly related with the workload. By creating situations where an ATCo gets multiple tasks, the relation between the subjective workload and the Number of Aircraft should be lower than found in this experiment. Expansion of the Solution Space concept to the third spatial dimension would also be a logical next step. However, it should be kept in mind that an ATCo uses a two dimensional visual environment to control the air traffic. Even though the ATCo is aware of the fact that he is controlling air traffic in a three dimensional environment, the two dimensional visualization could lead to unexpectedly strong correlations between the two dimensional Solution Space and the subjective workload measured in a three dimensional environment. Also, it might be advisable to use only active air traffic controllers in future experiments. The difference between groups was not unacceptably large for this experiment, but the introduction of a third dimension might increase the difference between experienced air traffic controllers and other subjects even further.

As an afterthought, the fact that the Solution Space got higher correlations than the Number of Aircraft did for a situation where only one aircraft was controlled, while those correlations are found to be of a same magnitude for the situation where multiple aircraft were controlled, suggests that the Solution Space might prove to be a good workload metric for vehicle controllers. In this context a vehicle controller should be taken as a broad definition, containing job descriptions as pilots or ship captains. The possibility exists that the Solution Space has a stronger relationship with the workload experienced in controlling one vehicle then with the workload experienced in controlling air traffic and thus multiple vehicles.

References

[1] Hilburn, B., "Cognitive Complexity in Air Traffic Control - a Literature Review," EEC Note No. 04/04, EUROCONTROL, Bretigny-sur-Orge, France, 2004.

[2] Janic, M., "A Model of Air Traffic Control Sector Capacity Based on Air Traffic Controller Workload," *Transportation Planning and Technology*, Vol. 20, 1997, pp. 311–335.

[3] Koros, A., Della Rocco, P., Panjwani, G., Ingurgio, V., and D'Arcy, J.-F., "Complexity in Air Traffic Control Towers: a Field Study," Technical Note DOT/FAA/CT-TN03/14, NTIS, Springfield, Virginia, 2004.

[4] Stassen, H. G., Johannsen, G., and Moray, N., "Internal Representation, Internal Model, Human Performance and Mental Workload," *Automatica*, Vol. 26, No. 4, 1989, pp. 811–820.

[5] Masalonis, A., Callaham, M., and Wanke, C., "Dynamic Density and Complexity Metrics for Realtime Traffic Flow Management," Technical paper, The MITRE Corporation, 2003.

[6] Kopardekar, P. and Magyarits, S., "Dynamic Density: Measuring and Predicting Sector Complexity," *Proceedings of the 21st Digital Avionics Systems Conference*, 27-31 October 2002, pp. 2C4–1–2C4–9.

[7] Lee, P. U., "A Non-linear Relationship Between Controller Workload and Traffic Count," *Human Factors and Ergonomics Society Annual Meeting Proceedings*, October 2005, pp. 1129–1133.

[8] Lee, P. U., Mercer, J., and Callantine, T., "Examining the Moderating Effect of Workload on Controller Task Distribution," *Lecture Notes in Computer Science*, Vol. 4562, 2007, pp. 339–348.

[9] Kopardekar, P. and Magyarits, S., "Measurement and Prediction of Dynamic Density," *5th USA-Europe ATM Seminar*, June 2003.

[10] Sridhar, B., Sheth, K. S., and Grabbe, S., "Airspace Complexity and its Application in Air Traffic Management," *2nd USA-Europe ATM Seminar*, December 1998.

[11] Chatterji, G. B. and Sridhar, B., "A Non-linear Relationship Between Controller Workload and Traffic Count," *Proceedings of the American Control Conference, San Diego, CA, USA*, June 1999, pp. 2620–2624.

[12] Athènes, S., Averty, P., Puechmorel, S., Delahaye, D., and Collet, C., "ATC Complexity and Controller Workload: Trying to Bridge the Gap," *HCI-02 Proceedings*, 2002, pp. 56–60.

[13] Averty, P., Athènes, S., Collet, C., and Dittmar, A., "Evaluating a New Index of Mental Workload in Real ATC Situation Using Psychophysiological Measures," *Proceedings of the 21st Digital Avionics Conference*, 2003, pp. 1–13.

[14] Averty, P., Collet, C., Dittmar, A., Athènes, S., and Vernet-Maury, E., "Mental Workload in Air Traffic Control: An Index Constructed From Field Tests," *Aviation, Space, and Environmental Medicine*, Vol. 75, No. 4, 2005, pp. 333–341.

[15] Van Dam, S. B. J., Mulder, M., and Van Paassen, M. M., "Ecological Interface Design of a Tactical Airborne Separation Assistance Tool," *IEEE Transactions on Systems, Man and Cybernetics, Part A*, Vol. 38, No. 6, 2008, pp. 1221–1233.

[16] Lee, K., Feron, E., and Pritchett, A. R., "Describing Airspace Complexity: Airspace Responce to Disturbances," *Journal of Guidance, Control, and Dynamics*, Vol. 32, No. 1, 2009, pp. 210–222.

[17] Hermes, P., Mulder, M., van Paassen, M. M., Boering, J. H. L., and Huisman, H., "Solution-Space-Based Analysis of the Difficulty of Aircraft Merging Tasks," *AIAA Journal of Aircraft*, Vol. 46, No. 6, 2009, pp. 1995–2015.

[18] d'Engelbronner, J. G., *Construction of a Tangent-Based Solution Space Diagram*, Unpublished MSc. Thesis, Faculty of Aerospace Engineering, Delft University of Technology, 2009.

[19] Tattersall, A. J. and Foord, P. S., "An Experimental Evaluation of Instantaneous Self-Assessment as a Measure of Workload," *Ergonomics*, Vol. 39, No. 5, 1996, pp. 740–748.

[20] Stanton, N., Salmon, P., Walker, G., Baber, C., and Jenkins, D., *Human Factors Methods: a Practical Guide for Engineering and Design*, Ashgate Publishing Limited – www.ashgate.com, 2005, Fourth Edition.

[21] Field, A., *Discovering Statistics Using SPSS*, Sage Publications – www.sagepub.com, 2009, Fourth Edition.

Appendix: Time shift table

Table 12. Time shifted correlations (R) (all with $p < 0.01$)

Parameter		$\Delta\tau$ [min]						
		-4	-2	-1	0	1	2	4
Baseline	Num of A/C	334	477	530	569	536	488	320
	Num of Comm	89	176	248	259	227	192	93
Area, 180°	Option 1	228	377	459	545	542	516	344
	Option 2	217	366	459	538	527	508	372
	Option 3	236	401	487	550	519	466	294
	Option 4	259	417	501	566	551	515	362
	Option 5	235	393	479	552	543	517	374
	Option 6	225	369	456	531	521	505	364
	Option 7	213	358	452	532	529	507	373
	Option 8	206	356	448	529	527	510	378
	Option 9	262	411	488	543	523	492	357
	Option 10	263	400	477	531	511	487	369
Change, Option 3	45°	117	247	332	408	396	338	214
	90°	140	279	380	455	442	385	253
	180°	159	276	363	450	418	367	229