JOURNAL OF AIRCRAFT
Vol. 46, No. 6, November–December 2009

Solution-Space-Based Analysis of the Difficulty of Aircraft Merging Tasks

P. Hermes,* M. Mulder,† and M. M. van Paassen‡
Delft University of Technology, 2629 HS Delft, The Netherlands
J. H. L. Boering§
To70, B.V., 2514 BB The Hague, The Netherlands
and
H. Huisman¶
Air Traffic Control the Netherlands, 1117 CV Schiphol-Oost, The Netherlands

DOI: 10.2514/1.42886

Air traffic controller workload is considered to be a limiting factor in the growth of air traffic worldwide. In this paper a new method of assessing controller task demand load will be developed and tested. Based on the hypothesis that operator workload is primarily caused by the difficulty of the task to be conducted, the concept of the solution space is described. For any particular air traffic control problem, the solution space describes the constraints in the environment that limit (and therefore guide) air traffic controller decisions and actions. The difficulty of that particular control problem can then be analyzed by considering the properties of the solution space. The task of merging an aircraft into a stream of other aircraft that fly along a fixed route is considered. An experiment has been conducted in which subjects were instructed to solve several merging problem scenarios of varying difficulty. After completing each scenario, subjects were asked to rate the difficulty of their task. High correlations are found between several solution-space properties and self-reported task difficulty.

Nomenclature

A = area, % of available area
B = bunching
BW = band size, deg
d = distance, length, m
F = ratio of sample variances
H = heading, deg
N = number (as in number of aircraft)
P = position, m
$\hat{P}$ = transposed position, m
p = significance
R = Pearson correlation coefficient
S = separation minimum (for aircraft), m
$\underline{S}$ = score
$\bar{S}$ = mean score
T = time, s
t = time, s
V = velocity, m/s

Subscripts

AC = aircraft
a = approaching
b = back
c = correction
com = command
e = turn termination (point)
e = evacuation (time)
f = front (for interception velocity)
G = ghost
h = heading
head = heading
int = interception
l = largest (area)
min = minimum
max = maximum
rel = relative (velocity)
rel = relevant (aircraft)
SV = separation violations
s = segment
t = turn initiation (point)
t = total (area)
Z = Z (as in Z scores)
μ = mean
σ = standard deviation
ω = angular velocity, rad/s

Received 22 December 2008; revision received 29 August 2009; accepted for publication 30 August 2009.

*Research Associate, Control and Simulation Division, Faculty of Aerospace Engineering, Kluyverweg 1; p_hermes@hotmail.com.

†Professor, Control and Simulation Division, Faculty of Aerospace Engineering, Kluyverweg 1; m.mulder@tudelft.nl. Senior Member AIAA.

‡Associate Professor, Control and Simulation Division, Faculty of Aerospace Engineering, Kluyverweg 1; m.m.vanpaassen@tudelft.nl. Member AIAA.

§Managing Director, Doctor Kuyperstraat 14; harjan.boering@to70.nl.

¶Human Factors Expert, Stationsplein ZuidWest 1001; h.huisman@lvnl.nl.

I. Introduction

AIR traffic controller (ATCo) workload is often cited to be one of the main constraints in air traffic growth [1–3]. Although more and more aspects of air transportation are being automated, the task of supervising air traffic is still performed by human controllers and is therefore limited by human performance constraints [4]. With the growth of air traffic volume, it is becoming increasingly important to be able to predict the effect of developments in the air traffic management system on the ATCo [5,6].

Currently, the effects of these developments on ATCo workload are assessed using expert judgment during the phase of the development process. For reasons of cost and time, it is preferred to perform this analysis at an earlier stage: for instance, during the fast-time simulation (FTS) phase. The development of workload assessment methods has a high priority for the developers of FTS programs, and the analysis of ATCo workload in FTS has been the subject of a large number of studies [7–10]. A major flaw in workload assessment using FTS programs is that it is impossible to assess the mental workload (i.e., the workload as *experienced* by the operator), as subjective elements such as training, equipment, and stress level play an important role here [11]. Instead, developers of FTS programs aim to analyze the ATCo's task demand load [12], which is considered to be an objective measure of the difficulty of the task to be performed by the controller.

For the current generation of FTS programs, task demand load metrics are constructed using a weighted combination of scenario properties. Examples are the number of aircraft involved, the sector size, the ratio of climbing and descending aircraft, or the count of weighted controller events [13]. The properties that are relevant to the task demand load analysis, and what weighing factors need to be used, are determined through expert judgment and regression analyses [14–19]. The validity of this method is questionable, however, since the scenarios that are being analyzed might differ heavily from the baseline scenarios used for the regression analysis. ATCo task demand load has proven to show nonlinear behavior [20] and can vary greatly due to slight changes in the situation being controlled [1].

Analytical approaches to assess the complexity of airspace, based on geometric attributes of traffic, have also been reported [21–24]. Examples are metrics that reflect the lack of structure of the traffic based on relative aircraft positions and velocities [25] or metrics that reflect the control activity required to accommodate disturbances in a traffic stream caused by the entrance of another aircraft in the airspace [24]. These analytical approaches have not, however, linked the complexity of the airspace or difficulty of solving air traffic control problems to controller workload.

Related to these geometric approaches, this paper aims to demonstrate a new method to perform a task demand load analysis for air-traffic-control-related problems. In this method, the difficulty of a particular controller task is analyzed by examining what we refer to as the "*solution space* of the problem." The solution space can be defined as the subset of all possible vector (combined heading and velocity) commands that can be issued by an ATCo that satisfy constraints of safety, productivity, and efficiency. These constraints are imposed by the situation at hand.

Essentially, the solution-space-based analysis allows a systematic approach to investigate the inherent difficulty of a particular air traffic control task, purely based on geometric and kinematic properties. In this paper, we will analyze the extent to which this metric for inherent task difficulty correlates to the self-reported task difficulty of air traffic controllers, through conducting an experiment with various scenarios of varying task difficulty. To evaluate the validity of our approach, at this point only the task of merging aircraft is considered in this paper, in the horizontal two-dimensional plane.

The structure of the paper is as follows. In Sec. II, some more background information regarding current complexity and task-difficulty metrics in air traffic control (ATC) is presented and the research goals for the current project are stated. Sec. III discusses the solution-space method in detail, applied to the problem of merging an aircraft onto a route. The setup of the experiment that was performed to investigate the validity of the solution-space-based analysis is discussed in Sec. IV. The results of the experiment are presented in Sec. V and discussed in Sec. VI. Recommendations regarding future research are presented in Sec. VII, and the Conclusions can be found in Sec. VIII.

II. Metrics for Complexity and Task Difficulty

ATCo workload is hypothesized to be composed of a number of factors, such as level of training, type of equipment, and sector complexity [26,27]. Indeed, Meckiff et al. [15] proposed ATCo workload to be a function of three elements: the geometric nature of the air traffic, the operational procedures and practices used to handle the traffic, and the characteristics of the individual controller. In order to objectively quantify workload, researchers attempt to eliminate variable factors, such as fatigue and training, and focus on the objective subject-independent aspects of workload, commonly referred to as the task demand load [12].

In most research, sector complexity is used as the means to describe ATCo task demand load. The underlying hypothesis is that, as in the current research, ATCo workload is coupled to task demand load (i.e., an increase in task demand load leads to an increase in mental workload as experienced by subjects) and that task demand load, in turn, is coupled to sector complexity (i.e., an increase in sector complexity means a higher task demand load).

Several methods and metrics have been proposed in order to quantify sector complexity. The most relevant ones will be discussed below.

A. Aircraft Count

A well-known sector complexity measure is aircraft count (or aircraft density, defined as the number of aircraft per unit of sector volume). The hypothesis is that by simply counting the aircraft in a sector at a given time, one can derive the sector complexity. Experiments indicated that, of all the individual sector characteristics, aircraft count (or density) shows the largest correlation with ATCo subjective workload ratings [16,19].

Figure 1 illustrates that aircraft count by itself is not a valid measure of complexity. It compares a situation in which four aircraft are flying in parallel (left) to a situation in which four aircraft are flying toward one point (right). Clearly, the second situation will be more difficult for the ATCo to deal with than the first one. Aircraft count does not account for air traffic or airspace characteristics and does not allow for a systematic complexity analysis.

The relation between aircraft count and ATCo-perceived difficulty is a nonlinear one [20,28]. The best fit between aircraft count and the subjective workload ratings is found using an S-shaped curve: workload can jump to a higher level when a certain number of aircraft is reached. The number of aircraft at which this occurs is assumed to depend on sector and personal characteristics.

B. Dynamic Density

In an attempt to obtain a more accurate measure of complexity, researchers combined various traffic and sector properties into one metric. This research led to the introduction of the term dynamic density (DD) [16–19,29]. It is defined as "the collective effort of all factors, or variables, that contribute to sector-level air traffic control complexity or difficulty at any point in time" [16].

Characteristics that are considered include the number of aircraft, the number of arrivals and departures the sector size, etc. The final workload estimate is then obtained by applying (usually linear) regression methods on samples of traffic data to determine the weights for all the characteristics and comparing them to subjective workload ratings.

Essentially, the assignment of weights based on regression methods means that a complexity analysis based on DD can only be performed on scenarios that differ slightly from the baseline scenarios used to determine the weights. Therefore, the DD-metric is not generally applicable to any situation. Furthermore, no systematic attempt is done to understand the possible *causes* of the different weights.

C. Traffic Load Index

In an attempt to construct a workload metric that would perform better than aircraft count and DD, the traffic load index (TLI), was

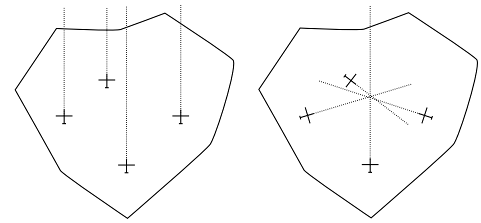


Fig. 1 Varying complexity for constant aircraft count.

developed [25,30,31]. The TLI was constructed to take the effects of time pressure and uncertainty into account when determining controller workload. Hence, since it allows for an inclusion of more cognitive, ATCo-centered aspects, it is assumed to be more promising than the DD approach.

The basic concept of TLI is that every aircraft is awarded a traffic load index value, or score. Each aircraft in the system is awarded a basic value of 1. If an aircraft might be involved in some kind of conflict, it will require extra attention and contributes to the ATCo's workload. Four classes of conflicts are identified, with maximum scores varying from 2 to 3.5. The scores for each conflict class are determined by a subject matter expert, usually an experienced air traffic controller. To obtain the final TLI, the scores of all the aircraft are summed together. Scores are updated for each aircraft on a regular basis (every few seconds), thus taking the dynamic behavior of the system into account. Initial tests showed that the correlation between TLI and subjective workload is higher than the correlation between aircraft count and subjective workload [31].

D. Current Research Goals and Approach

The various measures of sector complexity introduced above all have their merits. In the current research, our goal was to construct and validate a complexity metric that would incorporate as many sector properties as possible and that is scenario-independent (i.e., does not require baseline scenario analysis). Our approach is based on an investigation of a problem's solution space. The solution space of aircraft separation problems has been researched by Van Dam et al. [32] from the pilot's perspective, and it was hypothesized that this systematic approach might also be applicable to the ATC problem.

Basically, the solution space is a measure of the number and form of possible solutions that are at an operator's disposal to deal with a particular problem. In the present context, for any particular air traffic control problem, the solution space describes the constraints in the environment that limit, and therefore *guide*, the air traffic controller's decisions and actions.

As a first step in the development of the solution-space method, the problem of merging aircraft onto a single fixed route (hypothesized to be one of the more demanding and also representative controller tasks) is analyzed. Note that whereas the previous research attempts introduced above focus on assessing the complexity of full-scale ATC problems, the current research only considers the merging problem in the two-dimensional plane.

Merging problems occur, for instance, as aircraft approach an airport and need to be lined up for landing. The solution-space analysis is performed for aircraft that are not on the route and aims to find out what combinations of heading and velocity (the ATCo *vectors*) lead to a successful merge. It is discussed in detail in Sec. III.

Using the method of solution-space analysis on a merging problem, all properties of the scenario are systematically combined into a single metric. A solution-space-based metric should prove to be a more objective and scenario-independent metric than a weighted combination of scenario properties, in which the weights are highly dependent on the baseline scenarios considered.

The objective of the current research was to analyze how the initial solution-space properties of a merging problem correlate with the difficulty of the merging problem as subjectively reported by human controllers. If high correlations between certain solution-space properties (e.g., size, shape, etc.) and the subjectively reported task difficulty can be established, then there is a good possibility that the method can be used to objectively analyze other ATC-related control problems as well.

The research objective was achieved by performing an experiment. In this experiment, merging scenarios were analyzed using the solution-space method, after which an ATCo was asked to solve the problem in the scenario in a simulation. The operator was then asked to assess the problem's difficulty and several other variables, and correlations between the subjective assessments and the objective scenario properties, based on both conventional metrics as well as the solution-space-based metrics, were examined. The experiment and its results are discussed in Secs. IV, V, and VI.

III. Construction of Solution Space for a Merging Problem

The solution space is defined as the state space that represents possible vector (heading and velocity) commands issued by ATC that satisfy particular well-defined constraints. For the current analysis of two-dimensional merging problems, this means that a vector is part of the solution space when it satisfies the following three conditions:

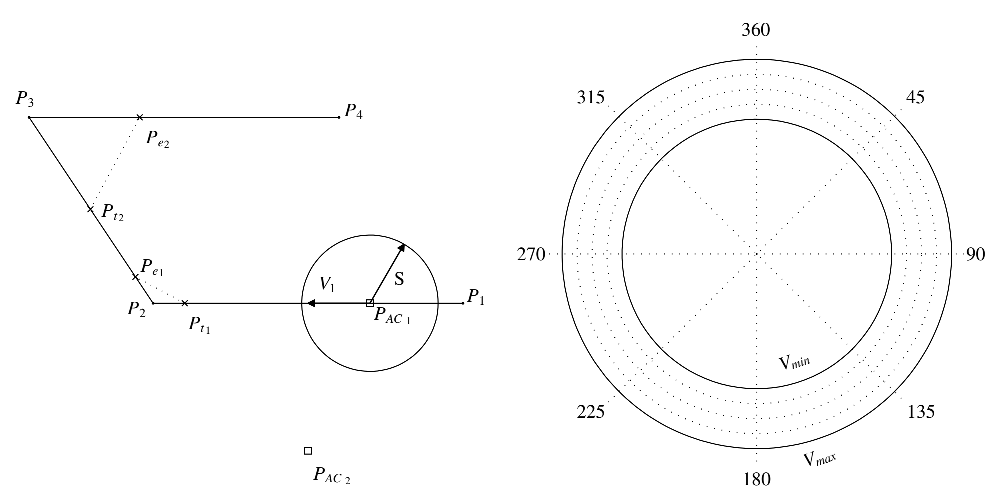
1) For productivity, the vector must be such that the free aircraft flies toward the route.

2) For safety, the vector may not lead to loss of separation at any point in time.

3) For efficiency, the vector must allow for direct route interception, no additional commands shall be necessary.

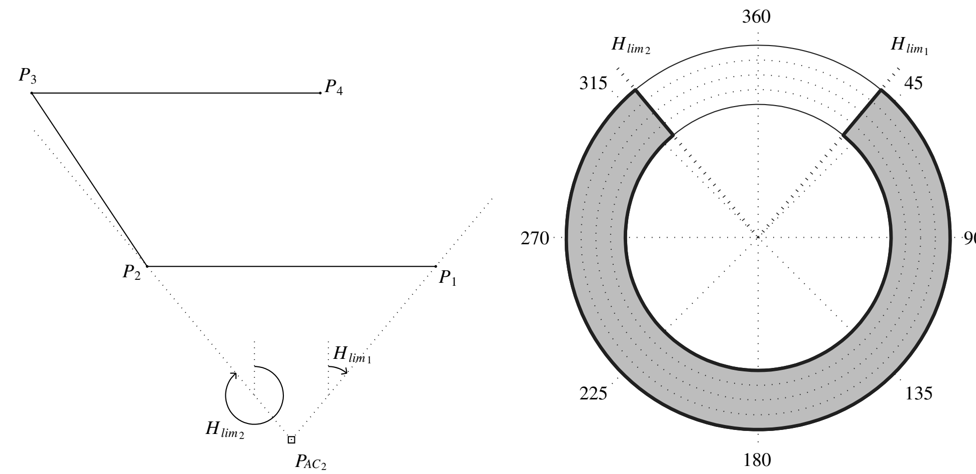
The solution-space construction is a process that requires a number of steps. To facilitate understanding our approach, these steps will be discussed one by one in the following paragraphs, using an example problem illustrated in Fig. 2a.

In Fig. 2a, a route is shown that runs from fixed route points P_1 to P_4, via P_2 and P_3. An aircraft (referred to as the *route* aircraft in the remainder of this paper) is defined by its position P_{AC_1}, velocity V_1, and separation minimum S and flies along the route. It is assumed that this route aircraft travels along the route with a constant velocity V_1. Assuming a fixed turning angular velocity ω of 3 deg/s, the turn



a) Overview of example problem b) Initial solution space

Fig. 2 Example problem: overview and initial solution space.



a) Determination of absolute heading limits

b) Solution space updated to include absolute heading limits

Fig. 3 Effects of absolute heading limits on the solution space.

initiation points P_{t_1} and P_{t_2} and the turn termination points P_{e_1} and P_{e_2} can be calculated. In the solution-space analysis, the turns are treated as straight segments from the initiation point to the termination point.

Another aircraft (referred to as the *free* aircraft) intends to intercept the route and is initially located at P_{AC_2}. The goal of the analysis is to determine which combinations of heading and velocity (vector) commands can be given to the free aircraft, in such a way that the vector command satisfies the productivity, efficiency, and safety criteria. The solution space can be drawn like Fig. 2b, in which all possible combinations of heading and velocity are present. Note that the velocity possibilities are limited by the minimum and maximum velocity, $V_{\min}$ and $V_{\max}$, respectively, of the free aircraft.

Certain combinations of velocity and heading commands meet the criteria of productivity, safety, and efficiency, but others do not. Areas in the solution space that contain these vectors are labeled as *safe areas*. Areas containing vectors that do not satisfy one of the constraints are called *unsafe areas*.

A. Step 1: Compute the Absolute Heading Limits

The first step in the solution-space analysis is to determine the absolute heading limits. In order to satisfy the productivity criterion, the free-aircraft heading must be chosen such that it will lead to route interception at some point in time. The absolute heading limits are determined by calculating the heading to every route point and then selecting the outermost headings (see Fig. 3a). The effects of the absolute heading limits on the solution space are shown in Fig. 3b. Note that when an unsafe area is added to the solution space, its contour will be drawn using thick lines and it will be filled gray.

B. Step 2: Determine the Conflict Areas

The following step in the solution-space analysis is to analyze which combinations of headings and velocities will cause a separation violation between the free aircraft and any of the route aircraft, which would violate the safety criterion. This analysis is performed separately for each route aircraft and for every segment of the route aircraft flight path. The areas in the velocity vector range that cause separation violations are called *conflict areas*. Note that the method is based on the work by Van Dam et al. [32].

1. Analysis of Conflict Areas for the First Segment

The analysis for the first segment is performed separately from the analyses for the other segments, as there is one less boundary condition. This will become clear later. The first route segment runs from the current position of the route aircraft, P_{AC_1}, to the initiation point of the first turn, P_{t_1}, as can be seen in Fig. 4a.

The first step in the conflict-area analysis is to draw a separation circle around the route aircraft and construct the tangent lines from the free aircraft to this circle.** If the free-aircraft velocity vector is chosen such that the *relative* velocity vector with respect to the route aircraft lies inside the area bounded by the tangent lines, then a separation violation will occur at some point in time (assuming constant velocities of both aircraft).

In order to find out which combinations of heading and velocity lead to such a separation violation, the position of the free-aircraft coordinate system is transposed by subtracting the velocity of the route aircraft (Fig. 4b). If the free-aircraft velocity vector is now constructed using $\hat{P}_{\mathrm{AC}_2}$ as the base point, then the vector from P_{AC_2} to the end of the free-aircraft velocity vector represents the relative velocity of the free aircraft with respect to the route aircraft. The free-aircraft minimum and maximum velocity circles have also been added in Fig. 4b.

As was stated before, a separation violation might occur if the relative velocity vector lies inside the area bounded by the tangent lines. Since the route aircraft will change heading as it reaches P_{t_1}, the relative velocity needs to have a minimum magnitude, $\|V_{\mathrm{rel}}\|_{\min}$. If the relative velocity magnitude is smaller than this value, separation violation will not occur while the route aircraft is on the first segment. The minimum relative velocity magnitude is determined by calculating the minimum distance from the free aircraft to the separation circle, $d_{\min_1}$, and dividing it by the time that it will take the route aircraft to reach P_{t_1}, as defined by its velocity vector V_1 and the length of the first segment, d_{s_1}:

$$\|V_{\mathrm{rel}}\|_{\min} = \frac{d_{\min_1}\|V_1\|}{d_{s_1}} \tag{1}$$

In Fig. 4c, the minimum relative velocity magnitude circle has been added.†† The conflict area is now defined as the area that lies between the tangent lines, between $V_{\min}$ and $V_{\max}$, and outside of the minimum relative velocity magnitude circle.

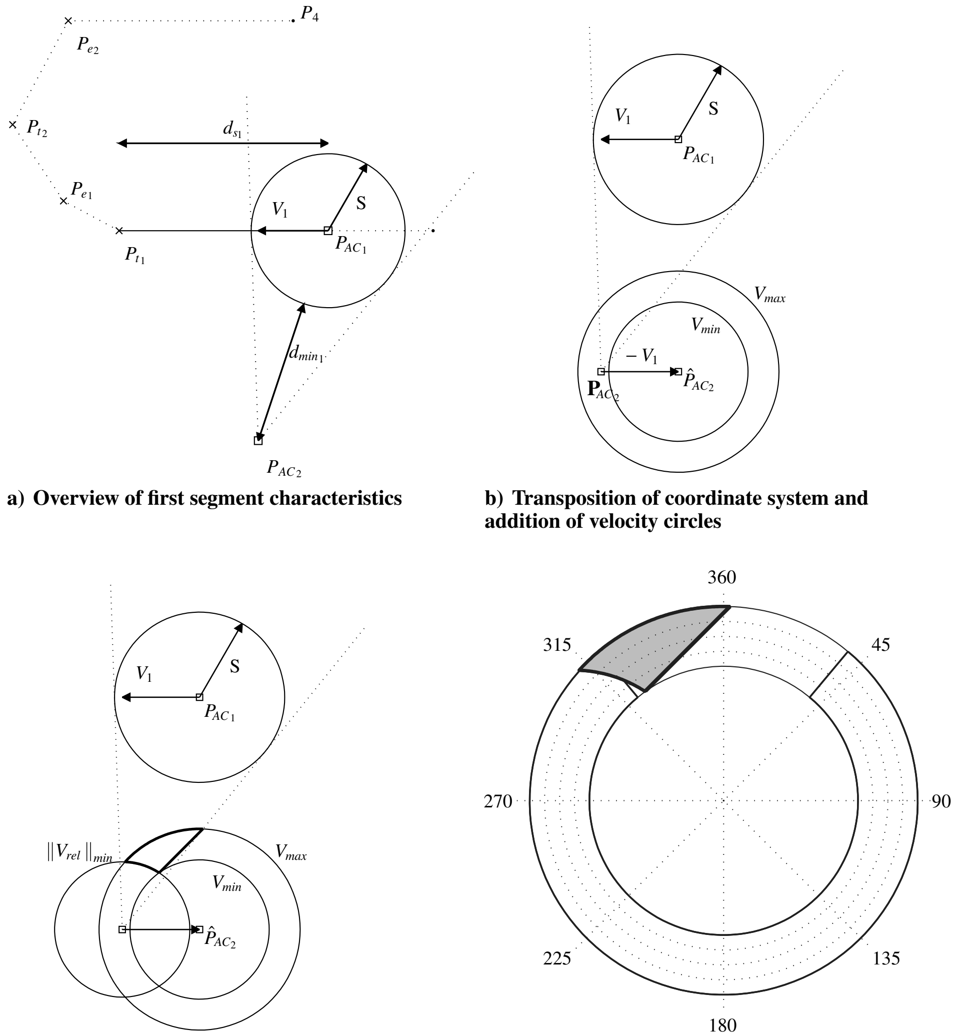
Finally, the conflict area has been added to the solution-space diagram (Fig. 4d).

2. Analysis of Conflict Areas for Other Route Segments

The analysis of the conflict areas for the later segments varies somewhat from the conflict-area analysis for the first segment. In order to illustrate the conflict-area analysis for one of the later segments, the second segment of the example problem is analyzed.

**In the current research, separation S is assumed constant for any combination of two aircraft.

††Note that this is in fact a worst-case calculation, as the free-aircraft velocity could actually be slightly larger for other headings.



a) Overview of first segment characteristics

b) Transposition of coordinate system and addition of velocity circles

c) Addition of minimum relative velocity circle and construction of conflict area

d) Solution space updated to include first segment conflict area

Fig. 4 Effects of first-segment conflict area on the solution space.

This segment runs from P_{t_1} to P_{e_1} (see Fig. 5a). Other segments follow similarly.

In order to perform the conflict-area analysis for this segment, a *route aircraft ghost* first needs to be constructed, which will reach P_{t_1} and start traveling along the second segment with velocity V_2 after a certain time t_1. This t_1 can be determined by dividing the along-route distance between the route aircraft's current position and P_{t_1}, d_{s_1}, by the route velocity magnitude $\|V_1\|$:

$$t_1 = \frac{d_{s_1}}{\|V_1\|} \tag{2}$$

The ghost aircraft is now placed at a position P_{G_1}, in such a way that it will reach P_{t_1} after time t_1, when traveling at velocity V_2:

$$P_{G_1} = P_{t_1} - V_2 t_1 \tag{3}$$

Note that the distance from P_{G_1} to P_{t_1} will be equal to the along-route distance d_{s_1}, due to constant velocity magnitude along the route.

Now, using P_{G_1} instead of P_{AC_1}, the basic methodology of conflict-area construction is the same as the one applied for the first route segment. The first step is to construct the tangent lines, transpose the free aircraft using the route aircraft second-segment velocity vector V_2, and add the minimum and maximum velocity circles (see Fig. 5b).

In the second step, however, an additional constraint is present. In the analysis of the conflict area for the first segment, the necessity of a minimum relative velocity magnitude was explained. This $\|V_{\mathrm{rel}}\|_{\min}$ is required to exclude any velocity vectors from the conflict area that would cause conflicts *after* the route aircraft leaves the segment being considered. It can be computed with

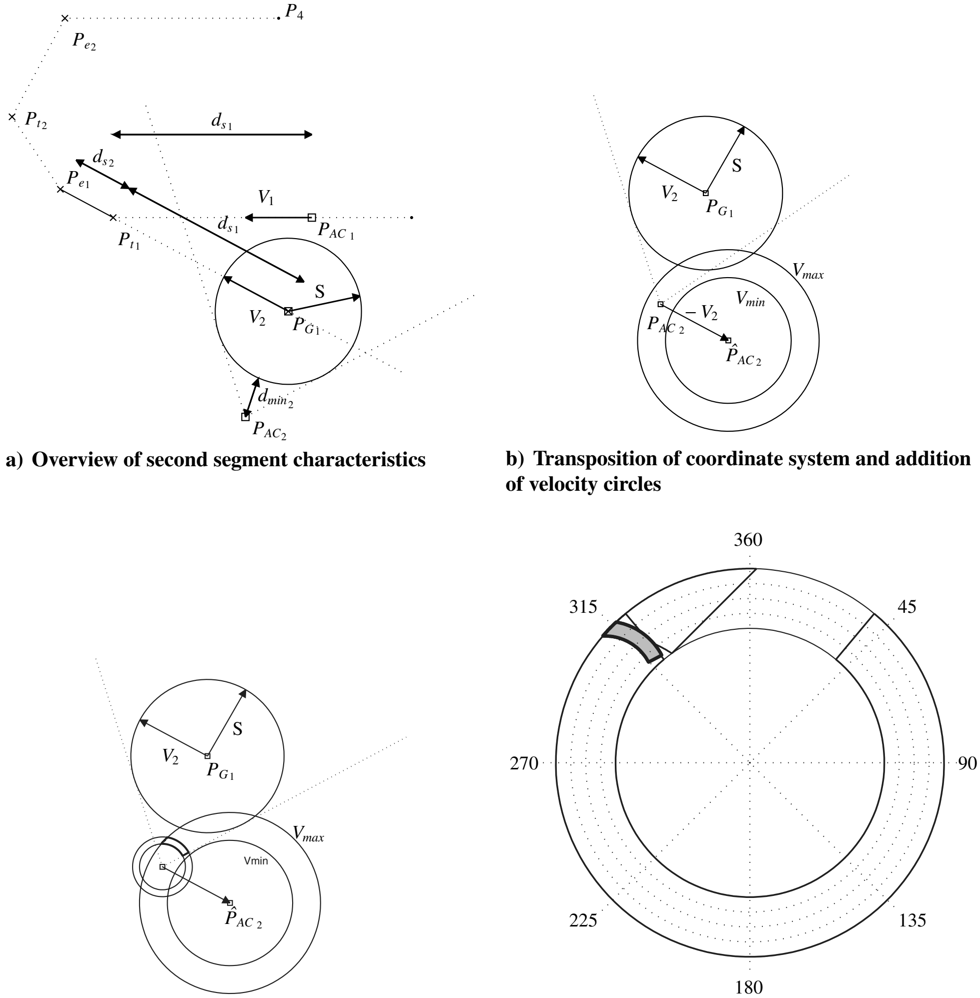
$$\|V_{\mathrm{rel}}\|_{\min} = \frac{d_{\min_2}\|V_2\|}{d_{s_1} + d_{s_2}} \tag{4}$$

For all segments apart from the first one, the addition of a maximum relative velocity magnitude constraint $\|V_{\mathrm{rel}}\|_{\max}$ is also necessary. This $\|V_{\mathrm{rel}}\|_{\max}$ is used to exclude any velocity vectors from the conflict area that would cause conflicts between the free aircraft and the route aircraft ghost *prior* to the time at which the ghost reaches the start of the segment. The maximum relative velocity magnitude can be calculated with

$$\|V_{\mathrm{rel}}\|_{\max} = \frac{d_{\min_2}\|V_2\|}{d_{s_1}} \tag{5}$$

In Fig. 5c, the minimum and maximum relative velocity magnitude circles have been added and the conflict area has been constructed. The updated solution space is shown in Fig. 5d.

The construction of conflict areas for the fourth and fifth segments is similar to that of the second segment. In Fig. 6, the solution-space updates are illustrated.



a) Overview of second segment characteristics

b) Transposition of coordinate system and addition of velocity circles

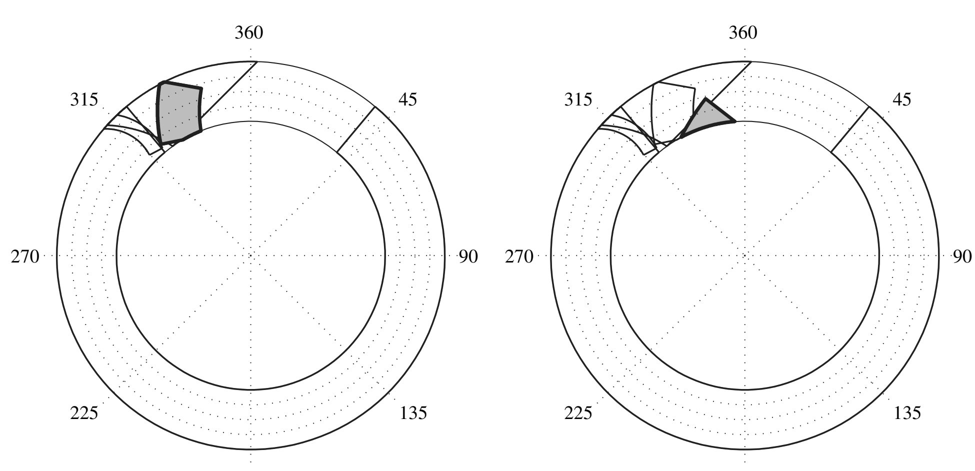
c) Addition of minimum and maximum relative velocity circle and construction of conflict area

d) Solution space updated to include second segment conflict area

Fig. 5 Effects of second-segment conflict area on the solution space.

In the analysis of the third-segment conflict area, an exception occurs. As can be seen in Fig. 7a, the free-aircraft position P_{AC_2} is contained *within* the separation circle surrounding the route aircraft ghost position P_{G_2}. Therefore, no tangent lines can be constructed, which means that the methodology as described above cannot be applied.

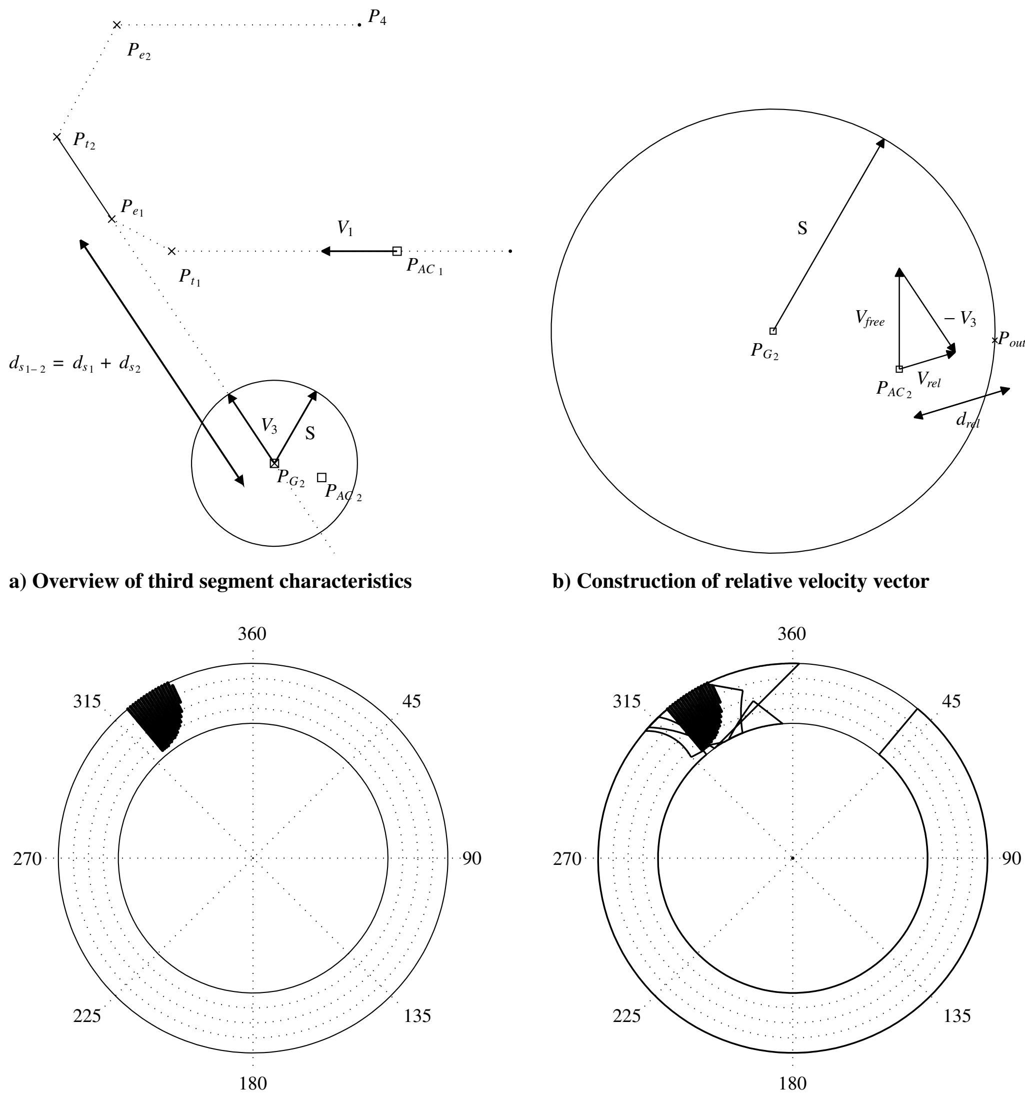
For this exceptional case, another technique is used to construct the conflict area. The free aircraft is currently *within* the ghost aircraft separation circle and needs to leave this circle before the ghost aircraft reaches the segment starting point P_{e_1}. The distance $d_{s_{1-2}}$ from the ghost aircraft to the starting point is known, as is the magnitude of the ghost aircraft velocity $\|V_3\|$. Hence, the evacuation



a) Solution space updated to include fourth segment conflict area

b) Solution space updated to include fifth segment conflict area

Fig. 6 Effects of fourth- and fifth-segment conflict areas on the solution space.



a) Overview of third segment characteristics

b) Construction of relative velocity vector

c) Third segment conflict area

d) Solution space updated to include third segment conflict area

Fig. 7 Effects of third-segment conflict areas on the solution space.

time t_e, the time before which the free aircraft must leave the ghost aircraft separation circle, can be determined:

$$t_e = \frac{d_{s_{1-2}}}{\|V_3\|} \tag{6}$$

For any free-aircraft velocity vector V_{free} in the solution space, a relative velocity vector V_{rel} can be constructed by subtracting V_3 (see Fig. 7b). This relative velocity vector can be decomposed into a heading H_{rel} and a magnitude $\|V_{\text{rel}}\|$. A check needs to be performed in order to determine whether this specific V_{rel} will lead to timely evacuation of the ghost aircraft separation circle. The evacuation will take place at point P_{out}. Since the relative heading is known, the relative distance from P_{AC_2} to P_{out}, d_{rel}, can be calculated. Since the magnitude of the relative velocity is also known, the time that it will take to cover this relative distance, t_{rel}, can be calculated:

$$t_{\text{rel}} = \frac{d_{\text{rel}}}{\|V_{\text{rel}}\|} \tag{7}$$

If $t_{\text{rel}} \leq t_e$, the free aircraft will leave the separation circle before the ghost aircraft reaches the segment starting point. Otherwise, a conflict will occur once the ghost aircraft reaches the segment starting point. By checking this condition for every possible velocity vector within the solution space, the conflict area for the third segment can be constructed. This conflict area and the updated solution space can be found in Figs. 7c and 7d, respectively. Note that the new conflict area is calculated using brute-force calculations, in which any unsafe vector is represented by a marker, which is why the new area is completely black.

3. Combination of Conflict Areas

Now that all conflict areas have been calculated, the total uncorrected solution space can be constructed. It is shown in Fig. 8, with the unsafe areas filled gray.

C. Step 3: Correct for the Free-Aircraft Route Interception

Now that the unsafe areas have been determined for the example problem, a correction needs to be applied. In the above, the assumption was made that the free aircraft would travel along a certain heading line with a certain velocity for an infinite amount of time. At a certain point, however, the free aircraft will intercept and follow the route. Therefore, any conflicts that would occur *after* the route interception (i.e., on the opposite side of the route line as seen from the free aircraft) need to be eliminated. Note that in the current research, it is assumed that the aircraft will intercept the route at the first route intersection ($P_1 P_2$ in our example): aircraft will never cross the route in order to merge onto it in a later stage or to merge on another route segment.

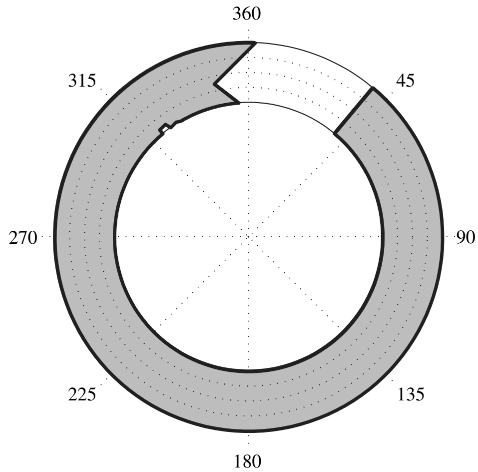


Fig. 8 Total uncorrected solution space.

The interception correction needs to be performed for the situation in which the free aircraft would merge in front of the route aircraft, and the situation in which it merges behind the route aircraft.

1. Correction for Interception in Front of the Route Aircraft

The front-interception correction is illustrated using an example situation, as can be found in Fig. 9a.

If the free aircraft were to fly at a certain heading H_{int}, route interception would take place at point P_{int}. It is assumed that the free aircraft always intercepts the route at the first possibility. The goal of the front-interception correction is to determine what magnitude the velocity needs to have in order for the free aircraft to safely merge at P_{int} in front of the route aircraft. In order to determine this magnitude, the position of the route aircraft, P_{c_f}, is determined such that the interception heading line will be a tangent line to the separation circle surrounding the route aircraft. Again, an assumption that represents a worst case.

It can now be stated that a safe merge will take place if the free aircraft reaches P_{int} *before* the route aircraft reaches P_{c_f}. Since the along-route distance d_{c_f} from P_{AC_1} to P_c is known, as well as the route aircraft's velocity on this segment, V_1, and the distance from P_{AC_2} to P_{int}, d_{int}, the minimum safe front-interception velocity magnitude $\|V_f\|$ can be calculated:

$$\|V_f\| = \frac{d_{\text{int}}\|V_1\|}{d_{c_f}} \tag{8}$$

In Fig. 9b, the updated solution space can be found. Once again, a brute-force calculation is used. Note that if $\|V_f\| \geq V_{\max}$, the boundary is positioned outside of the solution-space diagram. That is, no front interception can take place under the current velocity limits. For this particular example, this is indeed the case (see Fig. 9b).

2. Correction for Interception Behind the Route Aircraft

In a similar way, the maximum safe back interception velocity magnitude $\|V_b\|$ can be calculated. An example situation has been illustrated in Fig. 10a.

Since there is no intersection between the heading line and the separation circle on the relevant side of the route, it is not necessary in this case to place P_{c_b} such that the heading line becomes a tangent line; rather, it can be placed in such a way that P_{int} lies exactly on the separation circle. It is now safe to state that a safe merge will take place if the free aircraft reaches P_{int} *after* the route aircraft reaches P_{c_b}. The maximum safe back interception velocity magnitude $\|V_b\|$ can be computed with

$$\|V_b\| = \frac{d_{\text{int}}\|V_1\|}{d_{c_b}} \tag{9}$$

In Fig. 10b, the updated solution space can be found. Once again, if $\|V_b\| \geq V_{\max}$, then the boundary is positioned outside the solution-space diagram; any velocity magnitude will lead to safe back-route interception.

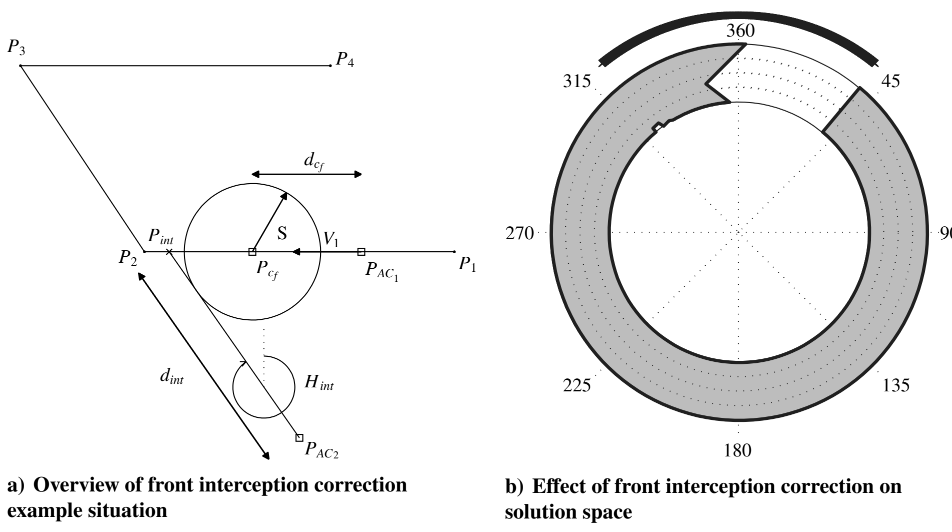
3. Final Corrected Solution Space

Using the methods described in the previous sections, each possible interception heading will be accompanied by two velocity limits. If the free aircraft travels along the heading at a velocity equal to or greater than $\|V_f\|$, no conflicts can occur, since the free aircraft will safely merge onto the route in front of the route aircraft. If the free aircraft travels along the heading at a velocity equal to or smaller than $\|V_b\|$, no conflicts can occur either, since the free aircraft will safely merge onto the route behind the route aircraft. The total corrected solution space can be found in Fig. 11.

The discussion above shows the calculation of the solution space for the case of one route aircraft. Naturally, when more aircraft are flying along the route, the calculations need to be repeated for all route aircraft, ultimately resulting in the solution space for the complete merging situation.

IV. Experiment

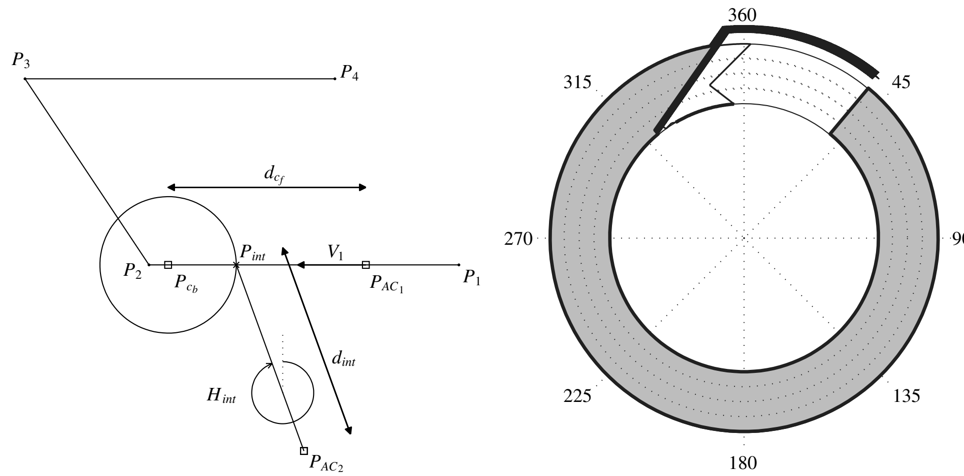
In order to investigate if and how the solution-space analysis can be used to assess the difficulty of the aircraft merging task, a



a) Overview of front interception correction example situation

b) Effect of front interception correction on solution space

Fig. 9 Effects of front-interception correction on the solution space.



a) Overview of back interception correction example situation

b) Effect of back interception correction on solution space

Fig. 10 Effects of back interception correction on the solution space.

validation experiment was performed. In this experiment, subjects were confronted with scenarios in which they were required to merge a free aircraft onto a route using heading and velocity instructions. Our main hypothesis was that when using the initial solution-space properties, a metric can be constructed that has a stronger correlation to the subjectively reported task difficulty than all other metrics.

A. Goal of the Experiment

Our main goal was to determine the correlations between the difficulty of solving a particular merging task as reported by the experiment participants and the various metrics that have been developed to predict the task difficulty. Two groups of metrics will be analyzed: namely, those that have been used in earlier studies, such as the number of aircraft, and those that reflect the properties of the initial solution space (the static solution space at the start of the scenario), such as size of safe areas (as defined in Sec. III).

Additionally, correlations were computed between these groups of predictor variables and two objective air traffic controller performance indicators: namely, the number of commands issued N_{com} and the number of separation violations N_{SV}.

B. Experiment Setup

For this experiment, a standalone simulator was developed using MATLAB™. Figure 12 shows the interface that was used. Note that the background was black in the experiment, but has been changed to white for clarity purposes. Label colors, line thicknesses, and font sizes have also been altered for clarity reasons.

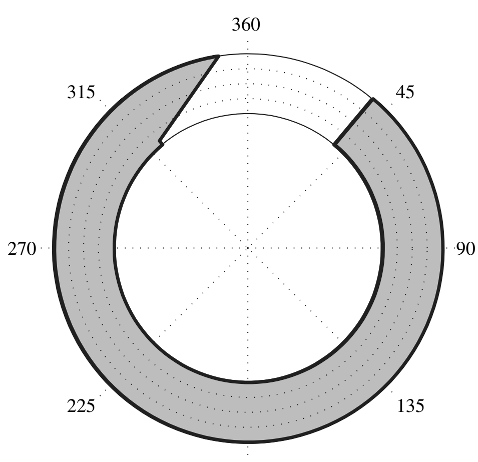


Fig. 11 Corrected final solution-space diagram.

The interface consisted of two parts. The left part is the plan-view display (PVD), in which the route and the aircraft can be found. To make the interface more realistic for the test subjects, the Dutch airspace sectors were plotted in the background. Aircraft were represented as a square accompanied by a label (see Fig. 13).

In the label, the call sign is on the top line. The current and commanded velocity are on the centerline and the current and commanded heading are on the bottom line. If the aircraft is a route aircraft, the heading line is replaced by the name of the route.

The right part of the interface is the command window. Subjects used this part to give commands to the aircraft. Subjects could send heading and velocity commands (either one by one or simultaneously) and the command to intercept the route.

C. Subjects and Instructions

Nineteen test subjects participated in the experiment. The subjects, 16 males and 3 females, ranged in age from 23 to 51 ($\mu = 32.4$ and $\sigma = 9.0$, in years). ATC experience ranged from none to 30 years, including 20 years as active tower/approach controller. Based on experience, subjects were divided into two groups. The first group consisted of six subjects: five males and one female, ages 33 to 50 ($\mu = 39.3$ and $\sigma = 6.4$). These subjects all had operational ATC experience or had participated in an extensive introductory practical ATC course. The second group consisted of 13 inexperienced subjects: 11 males and 2 females, ages 23 to 51 ($\mu = 29.2$ and $\sigma = 8.2$).

D. Procedure

Subjects first got familiarized with the interface using an interactive explanatory tutorial. Then a minimum of nine training scenarios, hypothesized to be ascending in task difficulty, were presented to them. Following one of these training scenarios (usually the second or third), subjects were introduced to the questionnaire. When subjects indicated that they understood the task and questionnaire, 15 measurement scenarios were done in randomized order (random in hypothesized complexity and random per subject).

E. Questionnaire

The questionnaire consisted of the following six questions, constructed to find out how difficult the subjects experienced solving the merging scenarios to be and why they assessed it as such:

1) How difficult was the merging problem to solve? Rate on a scale of 0 to 10, in which 2 represents as difficult as the first training scenario and 8 represents as difficult as the most difficult training scenario.

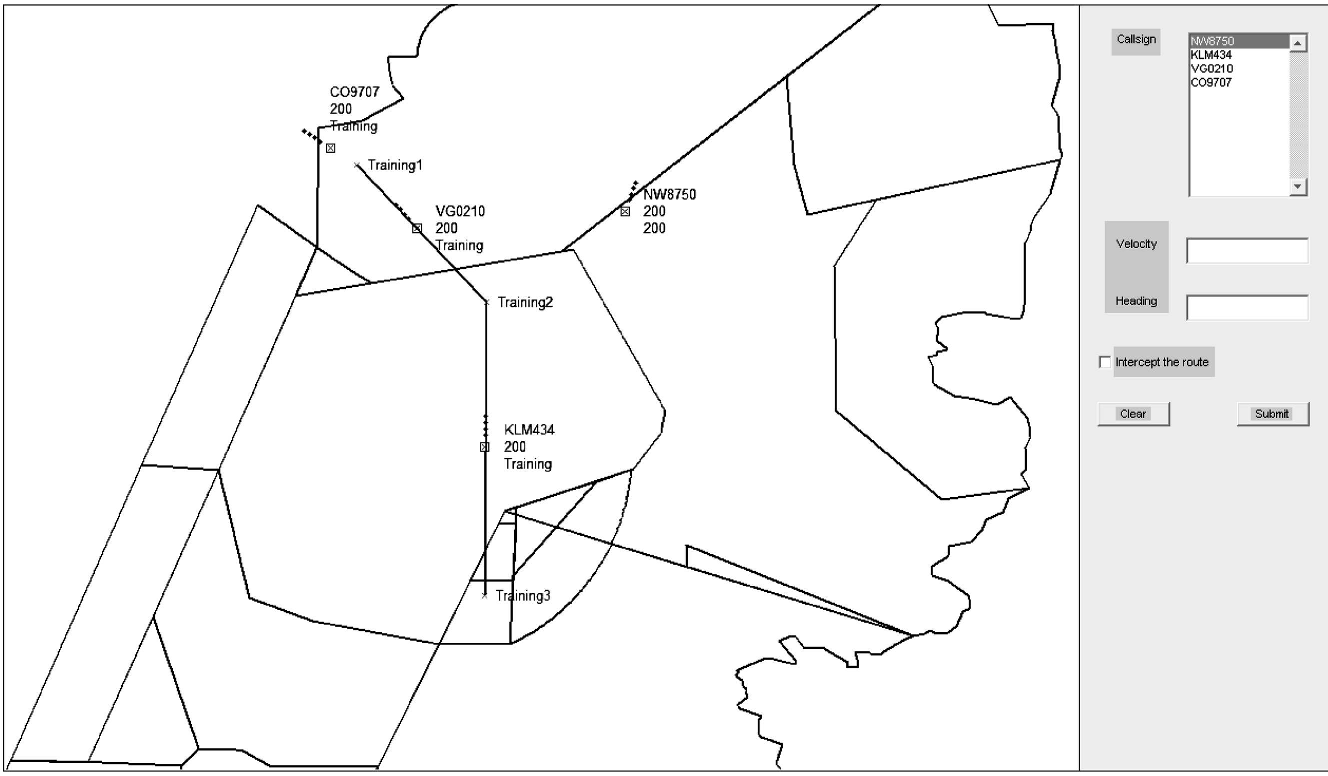


Fig. 12 Screen shot of experimental interface.

2) Did you feel that time pressure influenced the difficulty of solving the scenario? Rate on a scale of 0 to 10, in which 0 represents not at all influential and 10 represents very influential.

3) Did you feel that aircraft limits influenced the difficulty of solving the scenario? Scale equal to question 2.

4) Did you feel that route design influenced the difficulty of solving the scenario? Scale equal to question 2.

5) Did you feel that traffic influenced the difficulty of solving the scenario? Scale equal to question 2.

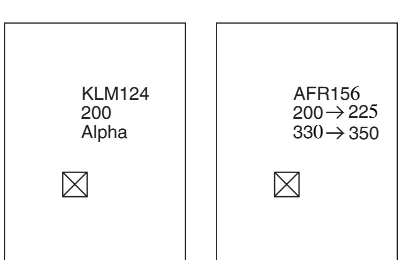
6) Did you feel that initial conditions influenced the difficulty of solving the scenario? Scale equal to question 2.

Additionally, subjects could submit any remarks they had regarding any question. Note that the use of Likert scales yields an ordinal data set.

F. Scenarios

1. *Flight Dynamics*

Since this experiment was an initial validation experiment, no actual flight dynamics model was used. All aircraft moved at a certain heading with a certain velocity (200 kt). Turns were performed at a constant rate of 3 deg/s, and accelerations and decelerations were performed at a fixed rate of 3 kt/s^2. The simulation was run four times as fast as real time, due to the relative simplicity of the task. The simulation was two-dimensional; altitude was not taken into account. All aircraft had a fixed $V_{\min}$ and $V_{\max}$ of 175 and 225 kt, respectively.



a) Route aircraft label **b) Free aircraft label**

Fig. 13 Experiment interface labels for route and free aircraft.

2. *Airspace and Routes*

The Dutch airspace was used as a background to increase the fidelity of the simulation. Subjects did not have to take sector boundaries into account when performing the task, however. Routes were constructed in such shapes and lengths as to create certain solution-space diagrams and merging problems. All routes were fictitious.

3. *Traffic*

In each scenario, traffic was placed such that (in combination with route design) certain initial solution-space properties were achieved (e.g., other aircraft placed such that just a single safe area was present). All traffic was present at the start of the scenario; aircraft disappeared from the PVD as they reached the end of the route. Route aircraft that were not on the actual route yet traveled toward the first route point; no aircraft intercepted the route at any point aside from the first one (except for the free aircraft, of course).

4. *Initial Scenario Properties*

Initial scenario properties that were hypothesized to be possible predictors of merging task difficulty were identified prior to the experiment. Which variables were deemed to be of most interest was determined by the literature study (see Sec. II) and expert judgment techniques (interviews and brainstorming sessions with operational experts).

The following six initial scenario properties were considered potential predictors:

1) Number of route aircraft or N_{AC} is the total number of aircraft in the simulation.

2) Number of approaching aircraft or N_{AC_a} is the number of route aircraft that are not currently on the route but are traveling toward the first route point.

3) Distance to the route or d_{route} (in meters) is the shortest distance from the free aircraft to the route.

4) Turns in the route or N_{turns} is the number of turns in the route.

Table 1 Initial scenario property values

Scenario	N_{AC}	N_{AC_a}	d_{route}, m	N_{turns}	l_{route}, m	B
1	8	3	66,598	3	209,815	0
2	8	3	61,298	3	209,815	0
3	9	4	71,194	3	217,518	0
4	6	3	49,833	2	127,697	0
5	7	2	49,599	3	162,912	0
6	7	1	64,791	2	149,339	1
7	9	3	46,547	2	149,339	1
8	8	4	53,393	3	139,269	0
9	7	4	67,624	1	97,725	0
10	7	4	67,592	1	97,725	0
11	14	5	48,686	3	217,518	7
12	11	5	68,966	3	217,518	1
13	8	4	58,360	1	117,043	2
14	6	3	72,338	2	100,714	0
15	3	2	55,140	1	84,398	0

5) Length of the route or l_{route} (in meters) is the total route length.

6) Bunching or B is a measure of aircraft being in close proximity to each other. For every two aircraft that have intersecting or touching separation circles at scenario initialization, the *bunching* value is increased by one.

The initial scenario property values are listed for all scenarios in Table 1.

5. *Initial Solution-Space Properties*

Several initial solution-space properties were also hypothesized to be possible task-difficulty predictors. An illustrative solution-space diagram is shown in Fig. 14. Once again, note that in the current research, only the *initial* solution space is analyzed.

The following seven initial solution-space properties were considered potential predictors:

1) Heading band range, or BW_{head} (in degrees), is the range of headings that allow for route interception and satisfaction of the productivity criterion.

2) Number of safe areas, or N_{safe}, is the number of safe areas within the solution space. In the example of Fig. 14, $N_{safe} = 4$.

3) Number of relevant aircraft, or $N_{AC_{rel}}$, is the number of aircraft that actively influence the solution space. Aircraft for which it is impossible for the free aircraft to collide with are deemed irrelevant.

4) Total solution-space size, or A_{safe_t}, is the total size of all safe areas, expressed as a percentage of the total available area between the minimum and maximum velocity circles[‡‡]:

$$A_{safe_t} = \sum_{i=1}^{N_{safe}} A_i \tag{10}$$

5) Size of largest safe area, or A_{safe_l}, is the size of the largest safe area, as a percentage of the total available area between the minimum/maximum velocity circles:

$$A_{safe_l} = \max(A_1, A_2, \ldots, A_{N_{safe}}) \tag{11}$$

6) Average safe-area size, or A_{safe_a}, is the average size of the safe areas, as a percentage of the total available area between the minimum/maximum velocity circles:

$$A_{safe_a} = 1/N_{safe} \cdot \sum_{i=1}^{N_{safe}} A_i \tag{12}$$

7) Safe-area size deviation, or σ_{safe}, is the standard deviation in safe-area size, as a percentage of the available area between the minimum/maximum velocity circles:

$$\sigma_{safe} = \sqrt{\frac{1}{N_{safe}} \sum_{i=1}^{N_{safe}} (A_i - A_{safe_a})^2} \tag{13}$$

The initial solution-space property values are listed for each scenario in Table 2.

G. Subject Instructions

Subjects were instructed to maneuver the free aircraft onto the route. They were free to choose any point on the route for interception but were not allowed to merge in front of the first route aircraft or behind the last route aircraft, because the stream of aircraft was finite. Their subgoal was to use as few commands as possible. In practice, most, if not all, subjects merged the free aircraft on the route segment corresponding to the heading bandwidth BW_{head}. Surprisingly, although subjects were told that they could also command the motions of the route aircraft, they all worked only with the free aircraft.

H. Dependent Measures

The dependent measures consisted of the subjective answers of subjects to the six questions in the questionnaire, the number of commands issued N_{com}, and the number of separation violations N_{SV}. Note that the latter number corresponds best to safety, the ultimate goal of air traffic control.

As different subjects exhibit different rating behavior, most notably leniency and central tendency, all quantitative data were first corrected for intersubject differences. This correction was performed by calculating the Z scores for every test subject. The Z score for subject i on scenario j was calculated using

$$S_{Z_{i,j}} = \frac{S_{i,j} - \bar{S}_i}{\sigma_i} \tag{14}$$

where $\bar{S}_i$ represents the subject's mean score along all the scenarios, and σ_i represents the subject's standard deviation in the scoring along all scenarios.

I. Hypothesis

Our main hypothesis was that the solution-space based metrics would yield higher correlations to self-reported task difficulty than any of the currently used scenario-related metrics.

V. Results

A. First Analyses of the Data

Before studying the relationships between the outcomes of the experiment (the answers to the questionnaire, the number of

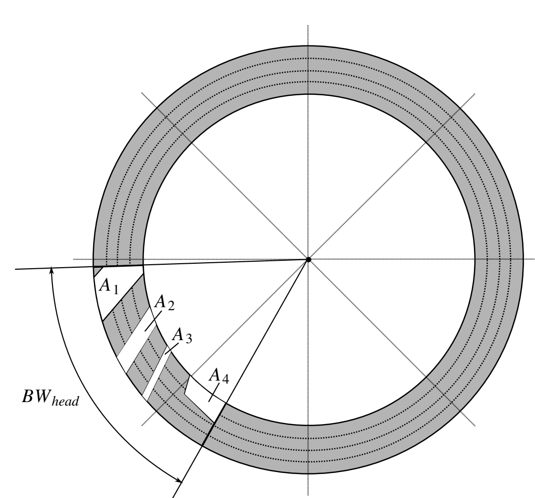


Fig. 14 Initial solution-space properties.

[‡‡]Note that here and in the following, the percentage is calculated with respect to the total area between the minimum and maximum velocity circles. Similarly, one could calculate this percentage relative to the total area between the velocity circles *and* the heading band range BW_{head}. For the sake of completeness, both are included in Table 2.

Table 2 Initial solution-space property values

Scenario	BW_{head}, deg	N_{safe}	$N_{AC_{\mathrm{rel}}}$	A_{safe_t}, %[a]	A_{safe_l}, %	A_{safe_a}, %	σ_{safe}, %
1	58.5	4	5	5.6 (34.5)	2.1	1.40	0.80
2	46.8	3	4	2.2 (16.9)	1.4	0.70	0.70
3	50.6	1	4	1.9 (13.5)	1.9	1.90	0.00
4	52.6	3	6	1.0 (6.8)	0.4	0.35	0.06
5	62.7	4	6	6.6 (37.9)	5.2	1.60	2.40
6	53.1	3	5	2.8 (19.0)	1.3	0.90	0.50
7	78.4	2	8	0.9 (4.1)	0.6	0.45	0.20
8	68.5	3	6	5.3 (27.9)	3.7	1.80	1.70
9	66.6	4	6	4.2 (22.7)	1.5	1.10	0.30
10	66.6	2	4	6.3 (34.1)	4.3	3.20	1.60
11	94.8	1	9	1.7 (6.5)	1.7	1.70	0.00
12	65.1	2	5	2.0 (11.1)	1.4	1.00	0.60
13	76.7	3	4	3.8 (17.8)	3.1	1.30	1.60
14	46.5	2	3	1.1 (8.5)	0.6	0.50	0.10
15	81.3	4	3	8.2 (36.3)	3.8	2.10	1.50

[a]Parentheses denote the percentage of the area between BW_{head}.

commands, and the number of separation violations) and the possible predictors (the six initial scenario properties and the seven initial solution-space properties), the data were analyzed regarding outliers, learning effects, and the reliability of the questionnaire.

1. *Outlier Analysis*

A total of 285 experiment runs were conducted. Before the data were processed, an outlier analysis was performed. Any experiment run in which the absolute value of the difficulty Z score was higher than two (i.e., score is more than two standard deviations away from the subject's mean score) was considered to be an outlier. A total of seven outliers were identified. These seven outliers were found for seven different test subjects and for six different scenarios. Since only a small percentage (less than 2.5%) of measurements were outliers and no trend was visible in the outliers, they were not removed from the data.

2. *Analysis of Training Effects*

During the experiment, learning effects may occur, which in this case would mean that difficulty scores decrease as the experiment progressed. To check this effect, the means of the difficulty Z scores are plotted against the experiment run numbers in Fig. 15, including the best-fit linear relationship. The best-fit line suggests that no training effect was present, as its slope is almost zero. To make sure that no significant training effect is present, an analysis of variance (ANOVA) was performed in which the first four and final four runs for every subject were grouped, using the difficulty rating as the independent variable. This analysis yielded $F_{1,151} = 0.00572$ and $p = 0.9809$. Clearly, no significant difference is present in the scoring between the first four and the final four runs. Hence, no significant training effect was present, and the full data set can be used for further data processing.

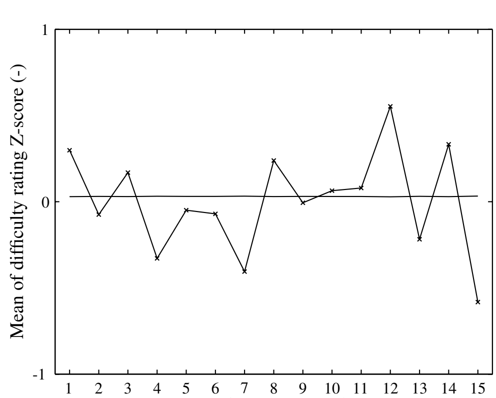


Fig. 15 Means of difficulty Z scores plotted against experiment run number.

3. *Reliability of the Questionnaire*

A reliability analysis was conducted to measure the consistency of subjects' answers to the questionnaire. The questionnaire consisted of six Likert scales ranging between 0 and 10, all ordinal scales. Using all data, Cronbach's α was 0.801; using the Z scores of all data, Cronbach's α equaled 0.726. Using the Z scores of the means of the data, averaged over subjects, Cronbach's α was 0.873. These values are all between 0.7 and 0.9, which indicates that the overall reliability of the questionnaire was high [33].

B. Number of Commands and Separation Violations

1. *Number of Commands*

A total of 1662 commands were issued by the subjects: 477 by the experienced subjects (group 1) and 1185 by the inexperienced subjects (group 2). The histograms are illustrated in Fig. 16. For experienced subjects, the mean N_{com} was 5.30 ($\sigma = 0.219$), median 5, with a minimum of 1 and maximum of 13 commands for a particular scenario. For inexperienced subjects, the mean was 6.08 ($\sigma = 0.190$), median 5, minimum 1, and maximum 15. The distribution of N_{com} was not normal for both groups. A Mann–Whitney test showed that the finding that the inexperienced subjects used a slightly higher number of commands is indeed significant ($U = 7399.500$, $Z = -2.154$, and $p = 0.031$).

2. *Number of Separation Violations*

A total of 51 separation violations occurred: 17 for group 1 and 34 for group 2. The number of separation violations per scenario never exceeded one. For the experienced subjects, the mean N_{SV} was 0.19 [standard error of the mean (SE) 0.041] with median 0; for the inexperienced subjects, the mean was 0.17 (SE 0.027) with median 0. The number of separation violations was not normal for both groups. The difference between both groups was very small and not significant.

C. Analysis of Variance

Before the correlation analyses were conducted, the distributions of the dependent measures (the six questionnaire Z scores and N_{com} and N_{SV}) were analyzed. This was done using data from all

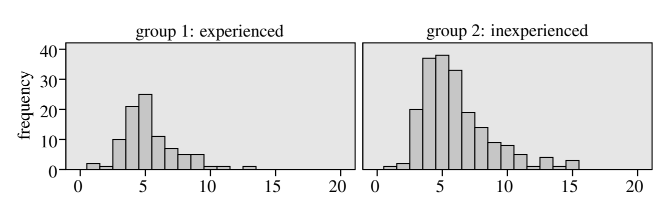


Fig. 16 Histograms of N_{com}, for the experienced (left) and inexperienced (right) group of subjects.

Table 3 Results of the repeated-measures analyses of variance

	Dependent measures (all Z scores)																							
	N_{com}			N_{SV}			Difficulty			Time pressure			Aircraft limits			Route design			Traffic			Initial conditions		
	F	$F \times G$	G	F	$F \times G$	G	F	$F \times G$	G	F	$F \times G$	G	F	$F \times G$	G	F	$F \times G$	G	F	$F \times G$	G	F	$F \times G$	G
	Initial scenario properties																							
N_{AC}	—	—	—	*s*	—	—	*S*	—	—	—	—	—	—	—	—	*S*	*s*	—	*S*	—	—	*S*	—	—
N_{AC_a}	—	—	—	—	—	—	*S*	—	—	—	—	—	—	—	—	*S*	—	*s*	*S*	—	—	—	—	—
d_{route}	*s*	—	—	*s*	—	—	*S*	—	—	*S*	—	—	—	—	—	*S*	—	—	*S*	—	—	*S*	—	—
N_{turns}	—	—	—	—	—	—	*S*	—	—	—	—	—	—	—	—	*S*	—	*s*	*S*	—	—	*s*	—	—
l_{route}	—	—	—	*s*	—	—	*S*	—	—	—	—	—	—	—	—	*S*	—	*s*	*S*	—	—	*s*	—	—
Bunching	—	—	—	—	—	*s*	*S*	—	—	—	—	—	—	—	—	*s*	*s*	—	*S*	—	—	—	—	—
	Initial solution-space properties																							
BW_{head}	—	—	—	—	—	—	*S*	—	—	—	—	—	—	—	—	*S*	—	—	*S*	—	—	*S*	—	—
N_{safe}	—	—	—	*s*	—	—	*S*	—	—	*S*		—	—	—	—	*S*	—	*s*	*S*	—	—	*s*	—	—
$N_{AC_{rel}}$	—	—	—	—	—	—	*S*	—	—	—	—	—	—	—	—	*S*	*s*	—	*S*	—	—	—	—	—
A_{safe_t}	*s*	—	—	*s*	—	—	*S*	—	—	*S*	—	—	*s*	—	—	*S*	—	—	*S*	—	—	*S*	—	—
A_{safe_l}	*s*	—	—	*s*	*s*	—	*S*	—	—	*s*	—	—	*s*	—	—	*S*	—	—	*S*	—	—	*s*	—	—
A_{safe_a}	*s*	—	—	*s*	—	—	*S*	—	—	*S*	—	—	—	—	—	*S*	—	—	*S*	—	—	*S*	—	—
σ_{safe}	—	—	—	*s*	—	—	*S*	—	—	—	—	—	—	—	—	*S*	—	—	*S*	—	—	*S*	—	—

experiment runs individually (referred to as *all scores* in the following, $N = 285$) *and* using the means of experiment data per scenario: that is, averaged over all subjects (referred to as *means*, $N = 15$). Using Kolmogorov–Smirnov (KS) Z tests, it was found that all means data were normally distributed, except for the number of separation violations N_{SV}. Additional KS tests showed that the all-data distributions were all not normally distributed.

Kolmogorov–Smirnov tests further showed that the distributions of the six initial scenario properties (Table 1) and seven initial solution-space properties (Table 2) were all normal.

The next step in data processing was to perform repeated-measures analyses of variance. The within-subject factors were each of the six initial scenario properties or each of the seven initial solution-space properties. The between-subject factor was the subject group (experienced vs inexperienced subjects). The dependent measures were the Z scores of the six questionnaire ratings, the number of commands, and the number of separation violations.

KS tests were conducted on all-data subsets when distributing the data over the within-subjects factors [33], showing that all distributions were normal; the only exception was the number of separation violations N_{SV}. Although analyses of variance are known to be quite robust to violations of the normal distribution, the results for N_{SV} should be handled with care. For all analyses of variance, Mauchly's test for sphericity was conducted, and when the data proved to be failing this test, the conservative Greenhouse–Geisser correction was applied.

Table 3 summarizes the results of all the repeated-measures analyses of variance that were conducted. In this table, F indicates the within-subjects factor, G is the between-subjects group, and $F \times G$ indicates the interaction. Only the highly significant (S: $p < 0.01$) or significant (s: $p < 0.05$) effects are shown.

A number of conclusions can be drawn from this table. First, all independent variables result in highly significant effects on the questionnaire's Z scores of difficulty, route design, traffic, and, to a lesser degree, initial conditions. As will become clear below, these scores are indeed highly correlated. There are almost no effects on the aircraft limits Z score and only a few (highly significant) effects on the time pressure Z score.

Second, the independent variables have almost no effect on the number of commands N_{com}, except for d_{route} and the variables reflecting the size of the solution-space safe area. More significant effects can be found for the number of separation violations N_{SV}, although here one should be careful in drawing conclusions, as the data are not normal.

Third, except for the questionnaire Z-score route design, which shows four significant effects of the group variable and three significant effects of the factor $\times$ group interaction, no other group or interaction effects are found to be significant. The other exception is N_{SV}; but then again these data are not normal, and these effects are not further discussed.

The very low number of group effects and interaction effects is remarkable. It indicates that there were no significant differences between the experienced controllers and novices. This means that, essentially, the data from both groups can be taken together, as the subjects can be considered as a homogeneous group.

The only exception is the way subjects scored the questionnaire regarding the contribution of the route design to the experienced difficulty of conducting the merging task Z route. Analysis of the data reveals that whereas inexperienced subjects gave higher ratings on this variable when either N_{AC}, N_{AC_a}, or bunching reached their highest levels, the experienced subjects did not. Rather, the ratings of the latter group were mostly affected by the length of the route l_{route} and the number of turns N_{turns}. Since the latter variables are indeed related to the design of the route and the former variables are not, one could conclude that the experienced subjects were better at interpreting the questionnaire.

Regarding the solution-space variables, a similar effect can be found, as here the group of inexperienced subjects gave higher ratings to Z route when either N_{safe} was very low (1), or $N_{AC_{rel}}$ was very high (6). The experienced subjects did not give higher ratings in these conditions. All in all, it appears as if the scores given by the group of inexperienced subjects were more affected by the independent variables, especially those concerning the extreme values in the initial scenario and solution-space properties, than were the experienced subjects.

D. Correlation Analyses

Correlation analyses were conducted to determine how well the possible predictors correlate with the test subject difficulty ratings, the number of commands, and the number of separation violations.

Again, the data from all experiment runs individually were used (all data), and the data for each scenario were averaged over all subjects (means). To investigate possible effects of experienced vs inexperienced subjects, the correlations were also computed when analyzing the data belonging to both groups separately. Hence, for the sake of completeness, all tables in this subsection will show six possible correlations.

The nonparametric Kendall's tau correlation coefficient was used, since this coefficient does not require data to be normally distributed or to be interval scale [33]. Although some of the data distributions (e.g., number of commands) would, in principle, allow for the use of parametric correlation coefficients, such as Pearson's rho, applying one single measure allows for a better comparison between the relative strengths of the correlations. Measuring the relative strength is the main purpose of the analyses, as it touches upon our experimental hypothesis.

In the following subsections, first the correlations between the subjective difficulty ratings will be investigated in relation to the five other questionnaire ratings. Second, the correlations will be investigated between the difficulty ratings and the six initial scenario properties and the seven initial solution-space properties. The latter analysis will be repeated for the number of commands and number of separation violations. These correlations reflect the relative strengths of using any of the independent variables to predict the subjective and objective outcomes of the experiment. Finally, the correlations between the latter dependent measures are analyzed.

1. *Correlation Between Difficulty and Other Questionnaire Variables*

First, the correlations between test subject difficulty ratings and test subject ratings to the other questionnaire variables were examined. In Fig. 17, the Z-score difficulty rating is plotted against each of the other questionnaire variables (also Z scores); the best-fit linear relationship is shown as well. Results of the correlation analyses (using Kendall's tau) can be found in Table 4. In this table and the following, statistically significant correlations are indicated with superscripts * ($p < 0.01$) or o ($p < 0.05$). The largest correlations are indicated with superscript a.

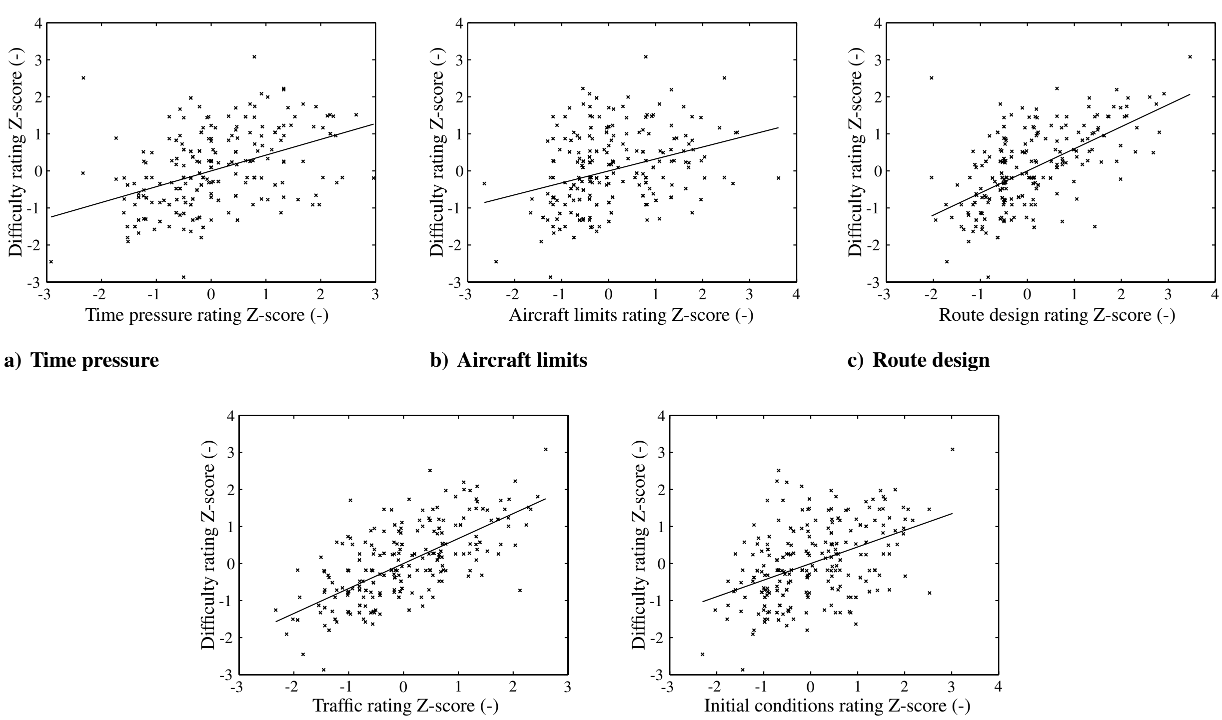
Statistically significant correlations with respect to self-reported difficulty were found for almost all of the other five questionnaire variables, for both the means and all-scores calculations. The traffic score had the highest correlation for both groups of subjects, linking difficulty most strongly to the presence of other aircraft flying on or toward the route. The second-highest correlations were found for the route design and initial conditions.

The main difference between the two groups becomes clear when considering the means data: inexperienced subjects relate difficulty more strongly to the experienced time pressure, whereas the experienced subjects related the difficulty more to the limits of the simulated aircraft.

2. *Correlation Between Difficulty and Initial Scenario Properties*

Second, the correlations between the six initial scenario properties and the difficulty ratings were analyzed. In Fig. 18, the Z-score difficulty rating is plotted against each of the initial scenario properties. In the figures, the difficulty Z-score means are represented by crosses, accompanied by a vertical line, which extends one standard deviation in both directions. A dotted line has been added to show the trend in the plot, and the best-fit linear relationship between the initial scenario property and difficulty Z score has been added as well. Results of the correlation analyses are summarized in Table 5.

Statistically significant correlations were obtained for only some of the all-data calculations. Here, all initial scenario properties except for d_{route} and N_{turns} yielded significant correlations to the reported difficulty. The number of aircraft N_{AC} and number of approaching aircraft N_{AC_a} give the strongest correlations to self-reported difficulty, with Kendall's tau values of 0.154 and 0.142, respectively. These are all related to traffic, supporting the hypothesis that subjects linked difficulty most strongly to traffic properties. It also corresponds well with the questionnaire findings.



a) Time pressure **b) Aircraft limits** **c) Route design**

d) Traffic **e) Initital conditions**

Fig. 17 Difficulty rating vs other questionnaire ratings (all data).

Table 4 Results of correlation analyses between difficulty rating and the five other questionnaire ratings, all Z scores

Variable	Time pressure	Aircraft limits	Route design	Traffic	Initial conditions
			Using all data		
All	0.316*	0.232*	0.422*	0.487*[a]	0.303*
Group 1	0.375*	0.249*	0.454*	0.461*[a]	0.424*
Group 2	0.287*	0.235*	0.412*	0.500*[a]	0.266*
			Using means		
All	0.562*	0.238	0.771*	0.848*[a]	0.790*
Group 1	0.314	0.478^{o}	0.543*	0.771*[a]	0.695*
Group 2	0.486^{o}	0.276*	0.638*	0.829*[a]	0.714*

[a]Largest correlation.

None of the correlations using means were statistically significant. Again, the strongest correlations were found with the number of aircraft-related properties.

The correlations for the groups of experienced and inexperienced subjects are rather similar, with one exception. Whereas the ratings of experienced subjects showed the highest correlations with the number of approaching aircraft N_{AC_a} (all aircraft in the scenario that are not on the route yet), the inexperienced subjects' ratings correlated best with the total number of aircraft N_{AC}. Another important difference is that the bunching scenario property correlated higher with the reported difficulty for the experienced subjects than for the inexperienced ones. Finally, when considering the absolute values of the correlations as a whole, the values are considerably higher for the experienced subjects (group 1) than for the inexperienced subjects.

3. *Correlation Between Difficulty and Initial Solution-Space Properties*

Third, the correlations between the seven initial solution-space properties and the difficulty ratings were analyzed. In Fig. 19, the Z-score difficulty rating is plotted against each of the initial solution-space properties, similar to as before. Results of the correlation analyses can be found in Table 6.

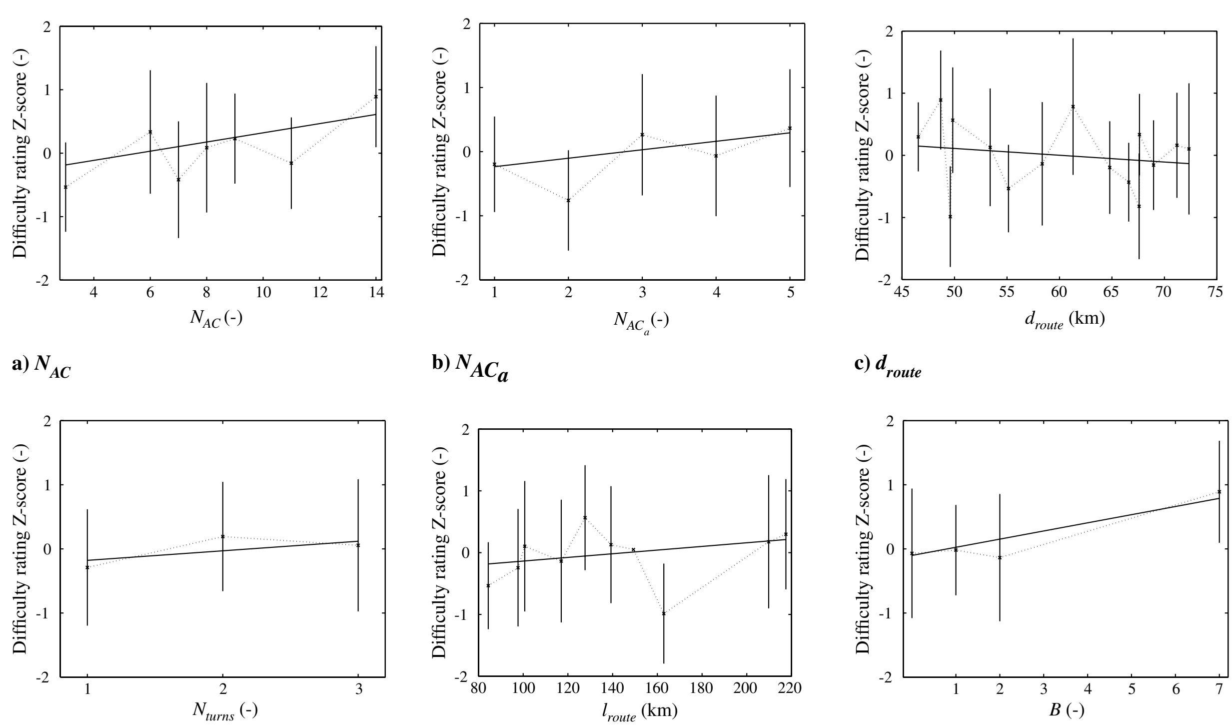
No significant effects regarding the initial heading bandwidth $\mathrm{BW}_{\mathrm{head}}$ were found. Overall, when comparing the absolute values of the correlations with those listed in Table 5, it is clear that the correlations are much higher, with values ranging up to 0.567.

Using the means calculations, statistically significant correlations were found between difficulty and three of the seven initial solution-space properties. When considering all data results, six of the seven properties yielded a statistically significant correlation to self-reported difficulty. For both types of calculations, the total solution-space size A_{safe_t}, the largest safe-area size A_{safe_l}, and the variation in safe-area size σ_{safe} showed the highest correlation values, up to approximately 0.567 in the means calculations and approximately 0.309 in the all-scores calculations.

Overall, the correlations for the experienced subjects are highest with the total solution-space size A_{safe_t}, whereas for the inexperienced subjects, slightly higher correlations were found with the solution-space area variation σ_{safe}. When considering the absolute values of the correlations as a whole, the values are considerably higher for the inexperienced subjects than for the experienced subjects, an opposite effect to that found for the correlations between difficulty and initial scenario properties, discussed in the previous subsection.

4. *Correlation Between Scenario Properties and the Number of Commands and Separation Violations*

Fourth, the correlations between the six initial scenario properties and the (Z scores of the) number of commands and number of



a) N_{AC} b) N_{AC_a} c) d_{route}

d) N_{turns} e) l_{route} f) *Bunching B*

Fig. 18 Difficulty rating vs initial scenario properties (all data).

Table 5 Results of correlation analyses between difficulty rating (Z scores) and initial scenario properties

Variable	N_{AC}	N_{AC_a}	d_{route}	N_{turns}	l_{route}	Bunching
Using all data						
All	0.154*[a]	0.142*	−0.060	0.087	0.110*	0.108[o]
Group 1	0.170[o]	0.212*[a]	−0.051	0.082	0.100	0.121
Group 2	0.147*[a]	0.117[o]	−0.057	0.084	0.027	0.105
Using means						
All	0.276[a]	0.246	−0.124	0.161	0.186	0.168
Group 1	0.256	0.311[a]	−0.143	0.184	0.186	0.246
Group 2	0.256[a]	0.225	−0.067	0.138	0.167	0.116

[a]Largest correlation.

separation violations were computed (see Table 7). In a similar way, the correlations were computed between the latter variables and the seven initial solution-space properties (see Table 8).

From these tables, a number of things can be concluded. First, the absolute values of the correlations between the (Z scores of the) two objective measures N_{com} and N_{SV} and the initial solution-space variables are considerably higher than those related to the initial scenario properties. Considering the latter, no significant correlations were found at all for the means data, and these are therefore not shown.

Second, regarding the initial scenario properties (Table 7), the number of commands is mostly affected by the number of aircraft N_{AC}, as can be expected, and the bunching parameter B. When considering the two groups, one can see that bunching mainly affects N_{com} for the inexperienced subjects. There are no significant correlations between any of the initial scenario properties and the number of separation violations N_{SV}. For the experienced subjects, N_{SV} decreases slightly when bunching increases.

Third, when considering the effects of the initial solution-space properties (Table 8), the highest correlations for the number of commands are found for the solution-space areas: in particular, the variation in area size σ_{safe} and the size of the largest safe area. Note that these correlations are strongest, and all significant, for the inexperienced subjects. Although these correlations are indeed also the highest for the experienced group, none of them are significant.

Fourth, the number of separation violations appears to be correlated most to the size of the largest safe area A_{safe_l}. In contrast to the number of commands, this correlation is by far the highest (and significant) for the group of experienced subjects. The inexperienced subjects had large positive correlations between N_{SV} and the number of aircraft that affect the solution space, $N_{\mathrm{AC}_{\mathrm{rel}}}$.

5. *Correlation Between Difficulty and the Number of Commands and Separation Violations*

Finally, the correlations between the difficulty ratings and the (Z scores of the) number of commands and number of separation were analyzed. In Fig. 20, the Z-score difficulty rating is plotted, and data pairs from every experimental run are represented by x and the

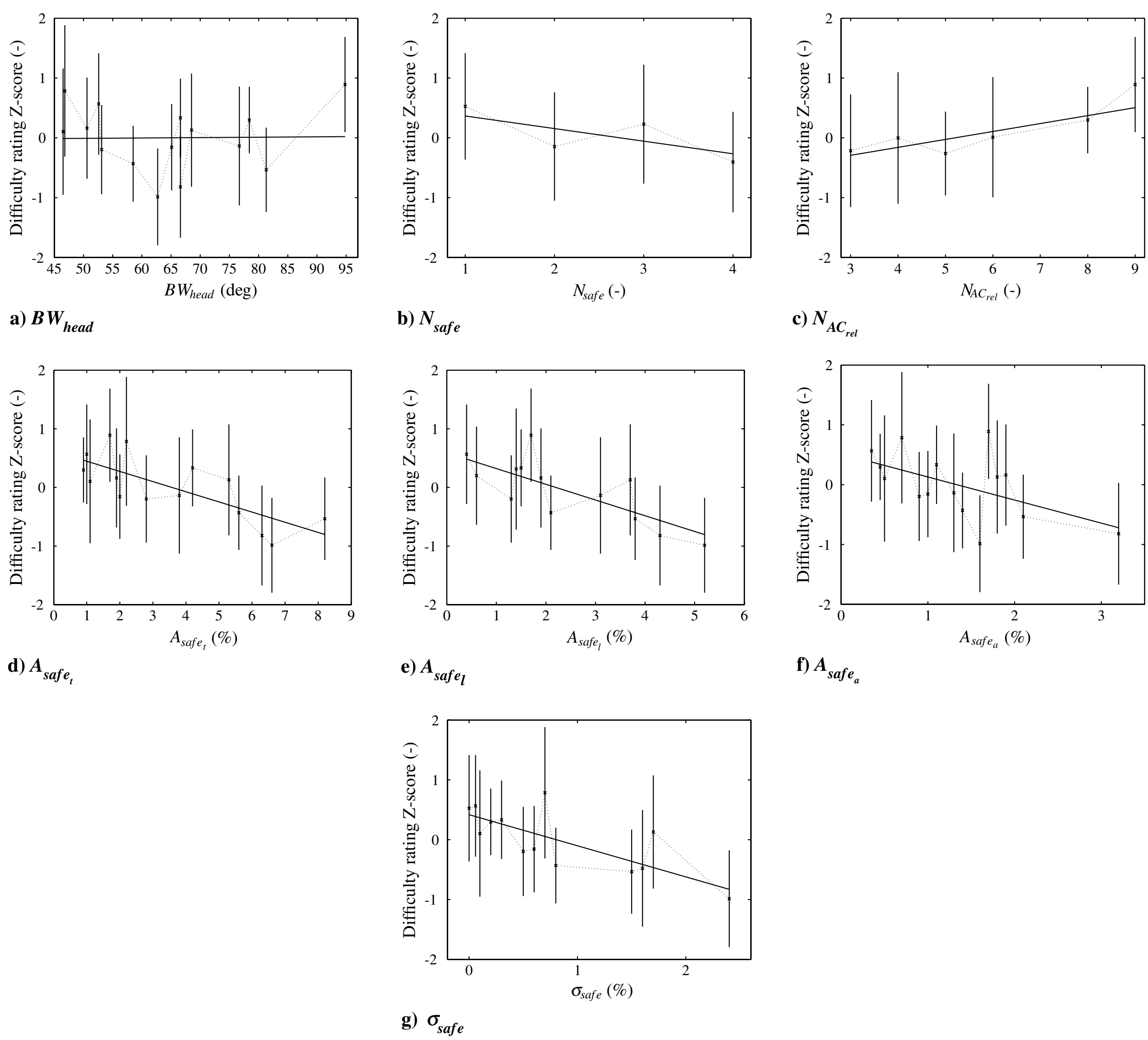


Fig. 19 Difficulty rating vs initial solution-space properties (all data).

Table 6 Results of correlation analyses between difficulty rating (Z scores) and initial solution-space properties

Variable	BW_{head}	N_{safe}	$N_{AC_{rel}}$	A_{safe_t}	A_{safe_l}	A_{safe_a}	σ_{safe}
			Using all data				
All	-0.012	-0.161^{*}	0.154^{*}	-0.281^{*a}	-0.250^{*}	-0.181^{*}	-0.278^{*}
Group 1	0.014	-0.173^{o}	0.108	-0.270^{*a}	-0.204^{*}	-0.163^{o}	-0.211^{*}
Group 2	-0.022	-0.160^{*}	0.174^{*}	-0.290^{*}	-0.272^{*}	-0.191^{*}	-0.309^{*a}
			Using means				
All	-0.038	-0.302	0.279	-0.543^{*a}	-0.471^{o}	-0.333	-0.510^{*}
Group 1	0.019	-0.194	0.197	-0.486^{*a}	-0.413^{o}	-0.352	-0.298
Group 2	-0.096	-0.302	0.300	-0.562^{*}	-0.529^{*}	-0.390^{*}	-0.567^{*a}

[a]Largest correlation.

Table 7 Results of correlation analyses between N_{com} and N_{SV} (Z scores) and initial scenario properties

Variable	N_{AC}	N_{AC_a}	d_{route}	N_{turns}	l_{route}	Bunching
			N_{com}, using all data			
All	0.091^{o}	0.034	-0.028	0.032	0.080	0.101^{oa}
Group 1	0.092^{a}	0.068	0.050	0.068	0.065	-0.021
Group 2	0.094	0.017	-0.059	0.024	0.093	0.153^{oa}
			N_{SV}, using all data			
All	0.025	-0.003	-0.060	0.068^{a}	0.044	0.004
Group 1	-0.060	-0.073	0.030	0.007	-0.003	-0.110^{a}
Group 2	0.060	0.023	-0.100^{a}	0.092	0.062	0.050

[a]Largest correlation.

addition of a best-fit linear relationship. Results of the correlation analyses can be found in Table 9.

Clearly, the self-reported task difficulty is highly correlated to the number of commands N_{com} and, to a lesser extent, the number of separation violations N_{SV}. For all data, the correlation between difficulty and N_{com} is only significant for the inexperienced subjects, however, who used, on average, slightly more commands, as discussed above. The difficulty ratings given by the experienced subjects showed higher correlations for the number of separation violations. When considering the means data, however, the correlations between difficulty and N_{com} and N_{SV} are very much the same for both groups of subjects.

E. Stepwise Regression Analyses

In order to determine whether a combination of predictors yield higher correlations to self-reported difficulty than every predictor on its own, stepwise linear regression analyses were performed using the means data (all normal). Given the fact that these distributions were all normal, the more common Pearson's rho (R) can be used. The fact that the difficulty rating is ordinal rather than interval is compensated for by taking the Z score of these ratings, a very common (but debated) approach. In a stepwise regression, models are constructed using various combinations of predictors, adding a predictor only if it yields a statistically significant increase in R. In this way, the smallest, most efficient, model is constructed that is best capable of predicting the Z scores of self-reported difficulty [33].

Using the seven initial solution-space properties, it was found that the single-predictor model based on only A_{safe_t} is the most efficient model to predict the self-reported difficulty, with R equal to 0.7423 ($R^2 = 0.551$). Although adding other initial solution-space properties resulted in higher R values, none of these increments were statistically significant. Similarly, the initial scenario properties (from N_{AC} to bunching B) were considered. Here, neither one of the variables nor any of the possible combinations were capable of producing a model with a statistically significant Pearson R, which is a remarkable result.

It was further investigated whether the objective measures in the experiment (Z scores of the number of commands N_{com} and the number of separation violations N_{SV}) were related to the difficulty Z score. A model with only N_{com} was found to be capable of reaching an R value of 0.829 ($R^2 = 0.687$). The stepwise regression showed that the most efficient model is constructed using both N_{com} and the number of separation violations N_{SV}, yielding an R value of 0.896, or $R^2 = 0.803$; that is, 80% of the difficulty Z score is related to a linear combination of N_{com} and N_{SV} (see Table 10). The ΔR^2 was 0.116 for the second model ($p < 0.05$), with a Durbin–Watson statistic of 1.935.

From the stepwise regression analysis, we conclude that, as hypothesized, the best possible objective *predictor* of self-reported difficulty is the solution-space area metric. The best post hoc *descriptor* of self-reported difficulty is a linear combination of the number of commands given and the number of separation violations that occurred.

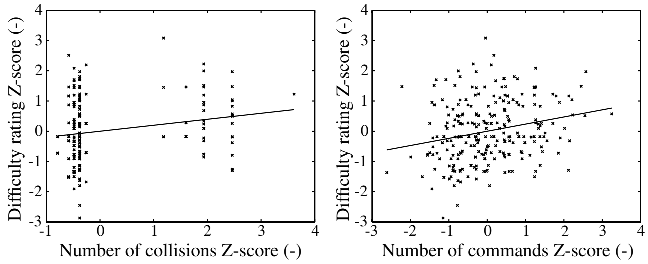
VI. Discussion

A total of 285 experiment runs were performed using 19 test subjects and 15 scenarios. No training effect was present in the data;

Table 8 Results of correlation analyses between N_{com} and N_{SV} (Z scores) and initial solution-space properties

Variable	BW_{head}	N_{safe}	$N_{AC_{rel}}$	A_{safe_t}	A_{safe_l}	A_{safe_a}	σ_{safe}
			N_{com}, using all data				
All	0.011	-0.027	0.065	-0.136^{*}	-0.149^{*}	-0.123^{*}	-0.152^{*a}
Group 1	-0.039	0.039	0.072	-0.088	-0.110^{a}	-0.092	-0.086
Group 2	0.034	-0.053	0.064	-0.156^{*}	-0.163^{*}	-0.129^{*}	-0.176^{*a}
			N_{com}, using means				
All	0.115	-0.065	0.300	-0.352	-0.433^{oa}	-0.371	-0.433^{oa}
Group 1	-0.134	0.022	0.155	-0.181	-0.298^{a}	-0.238	-0.163
Group 2	0.057	-0.043	0.300	-0.333	-0.413^{oa}	-0.352	-0.413^{oa}
			N_{SV}, using all data				
All	-0.017	-0.026	0.067	-0.072	-0.092^{oa}	-0.049	-0.080
Group 1	-0.160	0.019	-0.039	-0.130	-0.208^{oa}	-0.193^{o}	-0.119
Group 2	-0.039	-0.043	0.114^{oa}	-0.052	-0.050	0.003	-0.066
			N_{SV}, using means				
All	0.000	-0.103	0.264	-0.253	-0.337^{a}	-0.233	-0.266
Group 1	-0.208	0.070	0.146	-0.300	-0.512^{oa}	-0.406^{o}	-0.251
Group 2	0.152	-0.114	0.361^{a}	-0.181	-0.244	-0.101	-0.254

[a]Largest correlation.



a) N_{SV} b) N_{com}

Fig. 20 Difficulty rating vs measured properties (all data).

an outlier analysis showed that the full data set could be used. In the following, the experimental findings will be discussed, focusing on two main questions. First, does our hypothesis hold that the solution-space-based metrics yield higher correlations to self-reported task difficulty than any of the scenario-based metrics? Second, were there any differences between the groups of inexperienced and experienced subjects?

A. Correlation and Regression Analyses

The repeated-measures analyses of variance, with the two groups of subjects (experienced vs inexperienced) as an independent variable, indicated that there were very few effects in both the objective and the subjective data regarding subject experience. Hence, the data of all subjects could, in principle, be grouped together. We did find some differences between both groups, however, which will be discussed in the next subsection.

1. Correlation Analyses

In the correlation analyses, higher correlations were obtained when the means of properties per scenario were used (the means calculations, averaged over all subjects or over the two groups of subjects) than when all individual data couples were used (all-scores calculations). Because the number of observations was higher, use of individual data pairs led to more statistically significant results. Which of the calculation methods leads to the most interesting results depends on the final goal. That is, if the final goal is to construct a metric in which complexity can be predicted for an average subject, data from the means calculations should be used. The numbers included below all reflect the latter, unless stated otherwise.

Table 9 Results of correlation analyses between number of commands and separation violations (Z scores) and self-reported difficulty

	All data			Means data		
	All	Group 1	Group 2	All	Group 1	Group 2
N_{com}	0.165*	0.057	0.212*	0.619*	0.390^o	0.505*
N_{SV}	0.140*	0.163^o	0.133^o	0.559*	0.362	0.523*

Table 10 Results of stepwise linear regression, with difficulty rating (Z score) as a dependent measure and logged data (Z scores) as possible predictors

	b	SE b	β	Significance
		Model 1		
Constant	0	0.082	——	——
Z-N_{com}	1.356	0.254	0.829	$p < 0.001$
		Model 2		
Constant	0	0.068	——	——
Z-N_{com}	1.017	0.245	0.622	$p = 0.001$
Z-N_{SV}	0.356	0.210	0.398	$p = 0.021$

Regarding the questionnaire, subjects completed it in a consistent manner, and the reliability was high (Cronbach $\alpha > 0.7$). The answers to all five related questions showed statistically significant correlations with the self-reported task difficulty. The highest correlations were found for the traffic, initial conditions, and route design properties, indicating that subjects found these to contribute most to their perceived task difficulty. Self-reported difficulty was strongly related to the objective data, the number of commands N_{com}, and the number of separation violations NSV, with Kendall tau values 0.619 and 0.559, respectively.

When considering the initial scenario properties as predictors for the experiment outcomes, it was found that the number of aircraft N_{AC} and the number of approaching aircraft N_{AC_a} had the highest correlations, with self-reported task difficulty levels of 0.276 and 0.246, respectively. Note that although subjects were free to control all aircraft in the scenarios, most, if not all, of them only commanded the single free aircraft. We expect that the correlation between number of aircraft and difficulty will be higher if subjects are more clearly instructed or stimulated to control all aircraft in the simulation. In addition, the initial scenario properties and the objective data showed rather low correlations, with N_{com} mostly related to the number of aircraft N_{AC} and bunching B variables.

The correlations between the dependent measures and the initial solution-space properties were much higher. The highest correlations were found for the total safe-area size A_{safe_t} (−0.543) and the variation in safe-area size, σ_{safe} (−0.510). The absolute values of these correlations are almost twice as high as those found for the initial scenario properties. The same holds for the correlations between the initial solution-space properties and the objective data, N_{com} and N_{SV}. Regarding the number of commands, the highest correlations were found for the variables reflecting the solution-space size, up to values of −0.433 for the largest safe-area size A_{safe_l} and variation in safe-area size σ_{safe}. The same holds for N_{SV}, although the correlation values were slightly smaller, with values up to −0.377.

Overall, these results lead to the conclusion that solution-space properties, and specifically those that link to the *solution-space size*, were indeed the best predictors of self-reported task difficulty. This supports our main hypothesis: namely, that a more accurate difficulty predictor could be constructed using the solution-space concept. In addition, the solution-space-based variables were also better able to predict the number of commands issued by subjects and, to a less extent, the number of separation violations.

2. Regression Analysis

In the regression analysis, initial solution-space properties were combined in an attempt to obtain stronger correlations and thus more powerful predictors. The regression was performed using the means data only. It was found that a model based on only the total solution-space size A_{safe_t} was the most efficient, with Pearson rho R equal to 0.7423. Adding more solution-space-based metrics increased this correlation somewhat, up to a value of 0.839 when adding all predictors, but none of these improvements were statistically significant. This suggests that the initial solution-space properties are coupled. Whether this means that solution-space *size* is the most relevant of all properties or that another one that has not been analyzed here can add significant additional predictive capability should be determined in a more elaborate study.

Remarkably, when using the initial scenario properties, neither one of these nor any of the possible combinations were capable of producing a model with a statistically significant correlation. This further strengthens the usefulness of the solution-space-based method over the use of traditional metrics based on particular scenario characteristics.

B. Effects of Subject Experience

Note that the group with the most experienced subjects consisted of only six persons, and different results might have been obtained if more test subjects with practical ATC experience participated. Although the repeated-measures analyses of variance revealed that there were almost no significant effects of subject experience or any significant interactions, a number of interesting observations can be made.

First, experienced subjects used slightly less commands per scenario (5.30) than inexperienced subjects (6.08), which is a significant effect ($p = 0.031$). No differences in the average number of separation violations were found, which never occurred more than once per scenario.

Second, the repeated-measures ANOVA indicates that, overall, the questionnaire scores given by the inexperienced subjects were more affected by the independent variables than those issued by the experienced ones. In particular, for the extreme values of some of the scenario and solution-space related properties, the inexperienced subjects changed their self-reported task-difficulty ratings considerably, whereas the experienced ones did not. The same holds for the correlations of task difficulty and the number of commands issued by the subjects, which is much stronger for the inexperienced subjects. The experienced subjects, on the other hand, gave higher difficulty ratings when the number of separation violations increased. This difference disappears, however, when the ratings were first averaged over both groups of subjects.

Third, when considering the initial scenario properties, the correlations between the difficulty ratings were highest for the total number of aircraft in the scenario (N_{AC}) for the inexperienced subjects, whereas for the experienced ones, the highest correlations were found when only including the number of aircraft approaching the route (N_{AC_a}). Also, bunching B had a larger effect on inexperienced subjects, who issued higher ratings for difficulty and also used more commands when bunching increased. For experienced subjects, the number of commands was mainly related to the number of aircraft N_{AC}. Overall, for the initial scenario properties, the absolute values of the correlations are higher for the experienced subjects.

This is in contrast to the correlations found between difficulty and the initial solution-space-based metrics, where the correlations were highest for the inexperienced subjects. For both groups of subjects, however, the highest correlations were found with solution-space parameters σ_{safe} and A_{safe_t}. Regarding the number of commands, for the experienced subjects, the highest correlations were found for largest solution-space area A_{safe_l}, similar to that for inexperienced subjects, who also showed a high correlation with σ_{safe}. The number of separation violations was more related to $N_{\mathrm{AC}_{\mathrm{rel}}}$ for the inexperienced subjects and to A_{safe_l} for the experienced ones.

Summarizing, although almost none of the differences between both groups were found to be statistically significant, the results do indicate some interesting trends. The experienced subjects were more affected by the initial scenario properties than the inexperienced ones, who responded more strongly to the initial solution-space properties. They issued less commands and also responded more strongly to scenario properties that are well known to affect a controller's strategy considerably, such as bunching. Note that a group size effect could have played an important role here, however, and the results should be interpreted with care.

C. Discussion

Despite the interesting trends in the data, suggesting some differences between the inexperienced and experienced subjects, the absence of any significant effects between both groups is at least remarkable. It has been reported before, however, as in a study conducted by Rantanen and Nunes [34], in which the same result was found. Experienced air traffic controllers are often assumed to have developed a set of simple heuristics to cope with the air traffic in their sector, through training and experience on the job. These heuristics are known to be different from controller to controller, however, and also vary considerably when considering the various types of procedures, sector functions (approach and departure and area control), etc. [35].

In this respect, the way we have framed the problem of air traffic control through the solution space, using first principles of geometric and kinematic properties, does very likely *not* correspond to how a controller frames the problem. But it has not been our objective to obtain a cognitive model of what happens in a typical air traffic controller's head. The solution-space method simply reflects the *inherent* difficulty of a particular task (in this case, merging) to obtain an objective predictive measure of how an air traffic controller would judge the task difficulty. The results presented are in support of this approach. In addition, although [34] acknowledged the heuristic approach to understanding air traffic controllers, their main result was that controller workload could be determined best by considering the "momentary conflict geometries between aircraft pairs."

In their experiment, as well as in ours, however, it may be that the absence of significant differences between experienced and inexperienced subjects could be due to the simplicity of the task and test facility. From the perspective of the task being given, it may have captured one particular aspect that involves only some fundamental cognition and not any aspect of expertise. On the other hand, the high correlations between the self-reported task difficulty, the number of commands, and the number of aircraft are typically found in larger-scale experiments as well, so this fundamental aspect of the task was captured well enough.

In conclusion, the experimental results indicate that the solution space does indeed lead to more accurate complexity predictors. It is important to realize, however, that the two-dimensional merging problem that was analyzed in this research is not the only, or main, task that an air traffic controller performs in a normal work setting. Indeed, there still remains a considerable gap between understanding the (possibly cognitive) origins of controller workload and the geometric method proposed here. But since understanding the former may well prove to be impossible, the success of the latter in terms of predicting the experimental outcomes relative to existing approaches provides some confidence in the solution-space-based approach.

VII. Recommendations

Based on the results, a number of future research paths can be identified. First, the solution-space calculations are currently based on several worst-case assumptions, and the dynamic behavior of aircraft has not been included. Essentially, this means that the solution space is not a completely accurate predictor of what heading and speed commands would yield safe, productive, and efficient merging. Simple kinematic models of aircraft turn capabilities and acceleration and deceleration profiles should be included [32,36].

Second, the solution-space concept should be expanded so that it can be applied to include and analyze other ATCo tasks as well. This expansion can be performed for more two-dimensional problems, but should eventually include the three-dimensional situation, such as the use of altitude or flight levels in solving air traffic control problems.

Third, it would be interesting to examine how the *dynamic* solution space can be used to assess task difficulty. In the current research, only the initial solution space, which is static, was examined. By investigating how the solution space *evolves over time*, both naturally and due to controller inputs, new insights into the task-difficulty assessment quality of solution-space-based metrics can be gained. When the dynamic solution space is used, it might be possible to assess task difficulty continuously during an experiment run (i.e., in real time). It is interesting to investigate how the solution-space metric can be scaled up to include the effects of controlling multiple aircraft. For instance, in this case, can the solution spaces for each individual aircraft be averaged or should another transformation take place before being able to relate it to a controller's self-assessment of task difficulty or workload? It should also be investigated whether solution-space enlarging strategies exist and whether these can indeed be identified when analyzing real-time ATCo behavior.

Finally, a very different research path could be followed. The development of an ATCo interface in which the solution space is visualized might lead to a significant reduction of workload. Note that in the present experiment scenarios, there always existed one or more arrays of possible vectors to the free aircraft that would solve the separation problem instantaneously. Visualizing the solution space on the display allows controllers to directly perceive the actions that the current merging situation affords and to act accordingly. The solution space *is* in fact the ecological interface that one would construct in an attempt to support the ATCo in this particular task [37]. Developing this interface and also implementing it in these baseline ATCo tasks could have a tremendous effect on the perceived complexity of the problem, as essentially all possible solutions are presented. Visualizing the constraints of the situation with the conventional plan-view displays *invisible* could have significant benefits for the operator's awareness of the situation [38], as has been shown by others in various related applications [39,40].

VIII. Conclusions

This study investigated whether the solution space of a two-dimensional air traffic control merging problem can be used to assess the inherent difficulty of solving that problem more accurately and objectively than current metrics. An experiment was conducted that showed that the initial solution-space properties (in particular, those related to solution-space size) are indeed more accurate predictors of self-reported task difficulty than traditional metrics, while being at least as objective. This result provides a solid basis for expanding the solution-space research. Possible future research paths include dynamic solution-space analysis, three-dimensional air traffic control problems, and the development of solution-space-based interfaces to support air traffic controller decision-making and situation awareness.

References

[1] Hilburn, B. G., "Cognitive Complexity in Air Traffic Control—A Literature Review," EUROCONTROL, Bretigny-sur-Orge, France, Paper 04/04, 2004.

[2] Janic, M., "A Model of Air Traffic Control Sector Capacity Based on Air Traffic Controller Workload," *Transportation Planning and Technology*, Vol. 20, No. 4, 1997, pp. 311–335. doi:10.1080/03081069708717596

[3] Koros, A., Della Rocco, P. S., Panjwani, G., Ingurgio, V., and D'Arcy, J.-F., "Complexity in Air Traffic Control Towers: a Field Study," Federal Aviation Administration, TN DOT/FAA/CT-TN03/14, 2003.

[4] Costa, G., "Evaluation of Workload in Air Traffic Controllers," *Ergonomics*, Vol. 36, No. 9, 1993, pp. 1111–1120. doi:10.1080/00140139308967982

[5] Nicholls, A. P., Melia, A., Farmer, E. W., Shaw, G., Milne, T., Stedmon, A., Sharples, S., and Cox, G., "Changing the Role of the Air Traffic Controller: How Will Free Flight Affect Memory for Spatial Events," *Applied Ergonomics*, Vol. 38, No. 4, 2007, pp. 457–463. doi:10.1016/j.apergo.2007.01.007

[6] Stein, E. S., Della Rocco, P. S., and Sollenberger, R. L., "Dynamic Resectorization in Air Traffic Control: A Human Factors Perspective," Federal Aviation Administration, TR DOT/FAA/TC TN06/19,2006.

[7] Baart, D., "An Evaluation of Dynamic Density Metrics Using RAMS," Federal Aviation Administration, Rept. DOT/FAA/CT-TN, 2001.

[8] Phillips, M. R., and Marsh, D. T., "The Validation of Fast-Time Air Traffic Simulations in Practice," *Journal of the Operational Research Society*, Vol. 51, No. 8, 2000, pp. 457–464.

[9] Crutchfield, J., and Rosenberg, C., "Predicting Subjective Workload Ratings: A Comparison and Synthesis of Operational and Theoretical Models," Federal Aviation Administration, TR DOT/FAA/AM-07/6, 2007.

[10] Leiden, K. J., Kopardekar, P., and Green, S. M., "Controller Workload Analysis Methodology to Predict Increases in Airspace Capacity," 3rd AIAA Annual Aviation Technology, Integration, and Operations (ATIO) Tech, Denver, CO, AIAA Paper 2003-6808, Nov. 2003.

[11] Haga, S., Shinoda, H., and Kokubun, M., "Effects of Task Difficulty and Time-on-Task on Mental Workload," *Japanese Psychological Research*, Vol. 44, No. 3, 2002, pp. 134–143.

[12] Stassen, H. G., Johannsen, G., and Moray, N., "Internal Representation, Internal Model, Human Performance Model and Mental Workload," *Automatica*, Vol. 26, No. 14, 1989, pp. 811–820.

[13] Majumdar, A., Ochieng, W. Y., Bentham, J., and Richards, M., "En-Route Sector Capacity Estimation Methodologies: An International Survey," *Journal of Air Transport Management*, Vol. 11, No. 6, 2005, pp. 375–387. doi:10.1016/j.jairtraman.2005.05.002

[14] Pawlak, W. S., Brinton, C. R., Crouch, K., and Lancaster, K. M., "A Framework for the Evaluation of Air Traffic Control Complexity," AIAA Guidance, Navigation, and Control Conf., San Diego, CA, AIAA Paper 1996-3856, 1996.

[15] Meckiff, C., Chone, R., and Nicolaon, J.-P., "The Tactical Load Smoother for Multi-Sector Planning," EUROCONTROL, Bretigny, France, Paper 40, 1998.

[16] Kopardekar, P., and Magyarits, S., "Dynamic Density: Measuring and Predicting Sector Complexity," *Proceedings of the 21st Digital Avionics Systems Conference*, Inst. of Electrical and Electronics Engineers, Piscataway, NJ, 27–31 Oct. 2002, pp. 2C4-1–2C4-9.

[17] Kopardekar, P., and Magyarits, S., "Measurement and Prediction of Dynamic Density," *Fifth USA-Europe ATM Seminar*, EUROCONTROL, Bretigny, France, June 2003.

[18] Chatterji, G. B., and Sridhar, B., "Neural Network Based Air Traffic Controller Workload Prediction," *Proceedings of the American Control Conference*, American Automatic Control Council, Evanston, IL, June 1999, pp. 2620–2624.

[19] Masalonis, A. J., Callaham, M. B., and Wanke, C. R., "Dynamic Density and Complexity Metrics for Realtime Traffic Flow Management," The MITRE Corp., McLean, VA, 2003.

[20] Lee, P. U., "A Non-linear Relationship Between Controller Workload and Traffic Count," *Proceedings of the Human Factors Society 49th Annual Meeting*, Human Factors and Ergonomics Society, Santa Monica, CA, Oct. 2005, pp. 1129–1133.

[21] Delahaye, D., and Puechmorel, S., "Air Traffic Complexity: Towards Intrinsic Metrics," EUROCONTROL, Bretigny, France, Paper 60, June 2000.

[22] Histon, J. M., Hansman, R. J., Aigoin, G., Delahaye, D., and Puechmorel, S., "Introducing Structural Considerations into Complexity Metrics," *Air Traffic Control Quarterly*, Vol. 10, No. 2, 2002, pp. 115–130.

[23] Xing, J., and Manning, C. A., "Complexity and Automation Displays for Air Traffic Control: Literature Review and Analysis," FAA Civil Aerospace Medical Inst., TR DOT/FAA/AM-05/4, Oklahoma City, OK, 2005.

[24] Lee, K., Feron, E., and Pritchett, A. R., "," *Journal of Guidance, Control, and Dynamics*, Vol. 32, No. 1, 2009, pp. 210–222. doi:10.2514/1.36308

[25] Athènes, S., Averty, P., Puechmorel, S., Delahaye, D., and Collet, C., "ATC Complexity and Controller Workload: Trying to Bridge the Gap," *Proceedings of the International Conference on Human-Computer Interaction in Aeronautics*, AAAI Press, Menlo Park, CA, 23–25 Oct. 2002, pp. 56–60.

[26] Stein, E. S., "Air Traffic Controller Workload: An Examination of Workload," Federal Aviation Administration, TR DOT/FAA/TC TN06/19, Atlantic City, NJ, 1985.

[27] Kirchner, J.-H., and Laurig, W., "The Human Operator in Air Traffic Control Systems," *Ergonomics*, Vol. 14, No. 5, 1971, pp. 549–556. doi:10.1080/00140137108931274

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

[29] Sridhar, B., Sheth, K. S., and Grabbe, S., "Airspace Complexity and its Application in Air Traffic Management," *Proceedings of the Second USA-Europe ATM Seminar*, EUROCONTROL, Bretigny, France, 1998, pp. 1–9.

[30] 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*, Inst. of Electrical and Electronics Engineers, Piscataway, NJ, Oct. 2003, pp. 1–13.

[31] 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, 2004, pp. 333–341.

[32] 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.

[33] Field, A., *Discovering Statistics Using SPSS*, 3rd ed., Sage, London, 2009.

[34] Rantanen, E. M., and Nunes, A., "Hierarchical Conflict Detection in Air Traffic Control," *International Journal of Aviation Psychology*, Vol. 15, No. 4, 2005, pp. 339–362. doi:10.1207/s15327108ijap1504_3

[35] Rantanen, E. M., Yang, J., and Yin, S., "Comparison of Pilots' and Controllers' Conflict Resolution Maneuver Preferences," *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting*,

Human Factors and Ergonomics Society, Santa Monica, CA, Oct. 2006, pp. 16–19.
[36] Paielli, R. A., “Modeling Maneuver Dynamics in Air Traffic Conflic Resolution,” *Journal of Guidance, Control, and Dynamics*, Vol. 26, No. 3, 2003, pp. 407–415.
doi:10.2514/2.5078
[37] Vicente, K. J., and Rasmussen, J., “Ecological Interface Design: Theoretical Foundations,” *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 22, No. 4, 1992, pp. 589–606.
doi:10.1109/21.156574
[38] Flach, J. M., Mulder, M., and Van Paassen, M. M., “The Concept of the Situation in Psychology,” *A Cognitive Approach to Situation Awareness: Theory and Application*, edited by S. Banbury, and S. Tremblay, Ashgate, Oxon, England, U.K., 2004, pp. 42–60..
[39] Knecht, W. R., “Testing a Nonveridical Aircraft Collision Avoidance System: Experiment 1,” *International Journal of Applied Aviation Studies*, Vol. 7, No. 1, 2007, pp. 60–82.
[40] Van Dam, S. B. J., Steens, C. L. A., Mulder, M., and Van Paassen, M. M., “Evaluation of Two Pilot Self-Separation Displays Using Conflict Situation Measurements,” *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, Inst. of Electrical and Electronics Engineers, Piscataway, NJ, 12–15 Oct. 2008, pp. 3558–3563.
doi:10.1109/ICSMC.2008.4811850