Experimental Evaluation of a Co-planar Airborne Separation Display

Joost Ellerbroek, *Student member, IEEE*, Koen C. R. Brantegem, M. M. (René) van Paassen, *Member, IEEE*, Nico de Gelder, Max Mulder

***Abstract*—Two experiments, an active conflict resolution task and a passive situation awareness assessment, were conducted that compared two versions of a constraint-based co-planar airborne separation assistance display. A baseline display showed a maneuver space based on 2-D projections of traffic and performance constraints. A second augmented display also incorporated cutting-planes that take the dimension orthogonal to the projection into account, thereby providing a more precise visualization of traffic constraints. Results showed that although pilots performed well with either display, the augmented display scored consistently better in terms of performance, efficiency of conflict resolutions, the amount of errors in initial resolutions, and the level of situation awareness compared to the baseline display. On the other hand, more losses of separation were found with the augmented display, as pilots tried to maximize maneuvering efficiency according to the precision with which constraints were visualized.**

***Index Terms*—Ecological Interface Design (EID), Airborne Separation Assistance System (ASAS), self-separation, situation awareness, evaluation experiment**

I. Introduction

IN AN ONGOING STUDY on the design of a 3-D separation assistance interface, a constraint-based co-planar display was proposed that presents constraints on maneuvering in a 'velocity action space', that is overlaid on traditional moving-map displays [1]. The co-planar display is a combination of previous single-plane presentations [2], [3], with additional visualization of the interactions that exist between these planes. The evaluation of this display is the topic of this paper.

To meet the demands set by current plans for highly-automated conflict resolution [4], [5], such a self-separation interface should enable pilots to monitor separation, select and apply resolution advisories, but also judge the functioning of the separation assurance automation. This means that although automation will provide resolutions, pilots will ultimately be responsible for the validity of those resolutions. Several studies argue that this requires transparent and understandable functioning of automation [6]–[9]. The interface should provide a window to the reasoning and functioning of the automation, to ensure proper situation awareness (SA), and to keep pilots "in-the-loop" [10]–[12].

The constraint-based displays proposed in this study aim to improve pilots' understanding of automated resolutions, by helping them understand how different elements in the work environment interact, and shape the possibilities for conflict resolution. These data invariably form the premise on which automation bases its actions, and are therefore essential when automation functioning needs to be judged.

The focus of an evaluation study of such a display should therefore lie on how the elements of the display affect the pilot's awareness and understanding of the traffic situation. In the current study, two experiments are presented to serve this purpose. An active conflict resolution experiment was performed to evaluate how operator performance and behavior are influenced by the visualization. The second experiment consisted of a passive situation awareness assessment, and a questionnaire. The methods that were used to assess SA are also presented in this paper.

In both these experiments, two displays were compared that are very similar, and differ only in the visualization of interactions of constraints. The resulting comparison should illustrate the main addition in the co-planar concept, that sets it apart from its 2-D predecessors, i.e., visualization of the interactions that exist between planes of projection. Although the 'baseline' display condition will lack certain information compared to the 'augmented' co-planar display, there are no other, more equal alternatives to compare the co-planar concept with. Other existing display concepts either only show explicit resolution advisories, or show only one dimensional constraints, and are therefore even less detailed than the baseline condition in this study [13]–[16]. Although some degree of bias is unavoidable in this kind of comparison, the experiments were designed to minimize this effect.

The work presented in this paper will employ this comparison to focus on the effect of the additional interaction visualizations on the performance, behavior, and situation awareness of pilots in the task of airborne self-separation. The following section introduces the co-planar display. Section III discusses the topic of situation awareness measurement methods, and presents the methods that were used in this study. Sections IV-VII describe an active conflict resolution experiment and its results, and a passive situation awareness assessment and its results, respectively. The paper ends with a discussion on the results, and conclusions from the experiments.

This work has been co-financed by the European Organisation for the Safety or Air Navigation (EUROCONTROL), under its Research Grant Scheme launched in 2008, and by the National Aerospace Laboratory NLR. The content of the work does not necessarily reflect the official position of EUROCONTROL or the NLR on the matter.

The authors are with the Control and Simulation section of the Faculty of Aerospace Engineering, Delft University of Technology, Kluyverweg 1, 2629 HS Delft, The Netherlands. Nico de Gelder is with the National Aerospace Laboratory NLR, Anthony Fokkerweg 2, 1059 CM, Amsterdam, The Netherlands. Email: J.Ellerbroek@TUDelft.nl

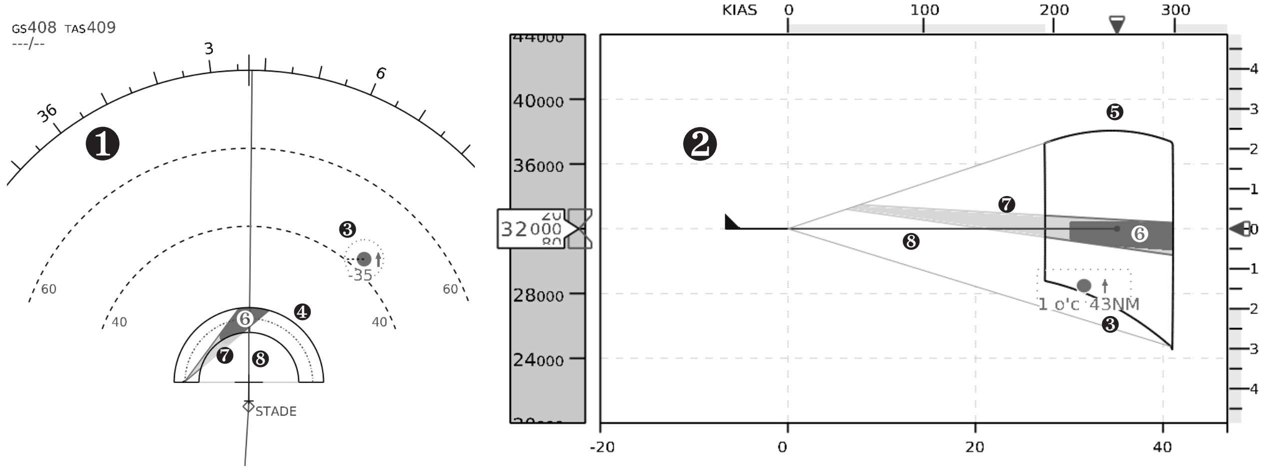


Fig. 1. Concept for a co-planar separation assistance display. This figure shows a HSD (❶) and a VSD (❷), with added separation assistance overlays. Relative intruder locations are indicated using TCAS-like symbology (❸). ❹ and ❺ are the horizontal and vertical State-Vector Envelope, respectively. ❻ is the reduced forbidden area on both on the HSD and the VSD. ❼ is the projected forbidden area on both displays. ❽ represents the ownship state vector.

II. The interface

Fig. 1 illustrates the co-planar display concept that was evaluated in this study. It consists of a concept for a self-separation interface, that presents separation-related constraints and relations on a co-planar display. Important elements of the display are numbered in the figure, and will be described in the remainder of this section. This display concept is part of an ongoing study on the design of a 3-D separation assistance interface, that uses work-domain analysis tools to identify constraints and relations relevant to the separation task. The reader is referred to [1] and [17] for a more elaborate review of this display and the work-domain analysis, respectively.

In this display concept, the 3-D traffic situation is visualized in two orthogonal, two-dimensional views: a top-down view (❶), and a side view (❷). Both views present a classical ownship-centered moving map, that shows spatial information such as the planned route and the relative positions of other aircraft (❸). In addition, constraints on ownship maneuvering are shown on both displays through velocity action-space* overlays (❹, ❺). These overlays are referred to as State-Vector Envelopes (SVEs) in the remainder of this text.

The horizontal SVE (❹) shows the horizontal maneuver space, in terms of track angle and airspeed. The boundaries of this action space are determined by the aircraft performance limits: The aircraft minimum and maximum operating speeds result in the concentric circular boundaries of the SVE. The vertical SVE (❺) shows a vertical maneuvering space, in terms of airspeed and vertical speed. Similar to the horizontal SVE, the boundaries of the vertical SVE are also determined by aircraft performance limits. The vertical edges of the SVE result from the limits on aircraft airspeed. The curved edge at the top of the vertical SVE visualizes the maximum obtainable steady climb at each velocity. The bottom edge indicates steady descent at idle thrust for each velocity. The area within these envelopes describes all reachable velocity vectors.

*The term ‘velocity action-space’ refers to the vector space containing all possible velocity vectors. The State-Vector Envelope describes the reachable subset of this vector space [1].

Intruder aircraft that are within detection range will reduce the available maneuver space in the horizontal and vertical SVEs. The reduced forbidden areas (RFAs) (❻) give the most precise representation of these constraints, because they incorporate the influence of the conflict geometry perpendicular to the respective projection plane [1]. On the Horizontal Situation Display (HSD), a RFA gives the constraints imposed by an intruder on ownship track angle and airspeed (❽), for the current value of ownship vertical speed. On the Vertical Situation Display (VSD), a RFA gives intruder-imposed constraints on ownship airspeed and vertical speed (❽), for the current ownship heading. The RFAs result from the intersection between a flat cutting plane, and the 3-D forbidden area: a compound of two slanted conical shapes, aligned with the top and bottom of the intruder protected zone. The shapes that result from this intersection range from circles, to ovals, to open-ended hyperbolic curves.

The projected forbidden areas (❼) are shown in combination with the RFAs, and provide several SA-related cues, as well as an outer limit on the shape and size of the RFA, when a flight parameter perpendicular to the corresponding projection plane is modified [1], [18].

Conflict urgency is explicitly indicated on the display using intruder symbology similar to the existing Traffic Collision Avoidance (TCAS) system [19]. In addition, conflict urgency is also indicated using color coding for all of the display elements that correspond to one intruder. This means that the aircraft symbols on both displays, as well as the forbidden area triangles and RFAs on both displays are colored according to the urgency of the conflict between ownship and the corresponding intruder.

III. Measuring situation awareness

The topic of situation awareness has stirred much debate in the past two decades. Several different definitions have been proposed, as well as varying methods aimed at measuring SA. In his review report, Uhlarik provides an extensive comparison of these definitions and methods [20].

The current work will employ Endsley's levels of situation awareness, which are a part of her definition of SA. She proposed that "Situation awareness is the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" [21].

Endsley's definition differentiates between three levels: The first level of SA describes the *perception* of the status, attributes, and dynamics of relevant elements in the environment. The second level is the *comprehension* of the significance of the level 1 elements on the operator goals. The ability to *project* the future state of the elements in the environment forms the third level of SA.

Although Uhlarik argues that the use of Endsley's model to describe SA has its limitations [20], the distinction between levels of SA is very valuable when assessing to what extent pilots utilize higher level information on the display, and how they relate this information to functional goals. As suggested by Flach, these levels of SA will therefore be used to categorize observed behavior in the experiment, rather than using an SA model to explain behavior [20], [22].

Most studies differentiate between three main categories of SA measurement methods: explicit methods, implicit methods, and subjective methods [20]. Explicit methods require subjects to report relevant parameters from memory, implicit methods infer level of SA from performance measures, and subjective methods ask subjects to self-rate their situation awareness. Each category of measurement method has its benefits and drawbacks, which is why Uhlarik argues for the use of multiple methods to ensure validity of results [20]. This study will therefore use methods from each category to assess SA.

Current explicit SA measures either require subjects to recall specific events after an experiment run is finished, or assess situation awareness on-line, while the experiment is running. A downside of retrospective methods (measuring after the actual run) is that the measurement is only as accurate as the memory of the pilot. That is, in an experiment with long runs, retrospective measurements are subject to forgetfulness and false recollections. On-line methods, on the other hand, can influence the pilot task being performed in the experiment. By having participants attend to particular information on the interface, these measures can cause participants to behave differently than they would otherwise [20], [23].

To mitigate the downsides of these methods, participants in this study will each perform two experiments, that separate the explicit from the implicit SA measurements. In the main experiment, subjects actively resolve conflict situations in a real-time simulated environment. The results from this experiment will be used to analyze resolution strategies, performance, and safety metrics. The performance measures will be used as implicit indicators of level of SA. In an additional passive experiment, subjects are presented with static conflict situations, each accompanied with a set of time-limited, multiple-choice SA questions, that are centered around Endsley's levels of SA. The resulting measures will be used to compare the display variants in terms of how they influence situation awareness. In a final post-experiment questionnaire pilots are given the opportunity to self-rate their situation awareness. By separating the explicit SA assessment from the active experiment, behavior in the main experiment no longer runs the risk of being directed by particular SA queries, and the explicit measurements are not hampered by the drawbacks of retrospective SA assessments.

IV. Experiment I: Active conflict resolution

To evaluate the co-planar display concept, a traffic separation experiment was performed, where pilots were placed in conflict situations with a loss of separation in the medium to short term future (3–5 min). Each session consisted of a continuous presentation of four consecutive conflict scenarios, that needed to be resolved manually, with the aid of a co-planar separation assistance display. Traffic conflicts were always between a single human actor, and simulated conflicting traffic.

A. Apparatus and aircraft model

The experiment was performed on the Apero flight simulator of the National Aerospace Laboratory (NLR). The Apero is a fixed-base flight simulator, featuring five high-resolution touch screens, and a large (52 inch) screen that provides the outside visual. The left-hand seat, primary display showed a conventional Airbus Primary-Flight Display (PFD) and the co-planar HSD/VSD display concept. The copilot display was disabled during the experiment. The middle vertical screen showed the Electronic Centralized Aircraft Monitor (ECAM) instruments. The touch screens on the pedestal showed several instruments, such as the Multifunction Control and Display Units (MCDU's) and the radios.

Pilots controlled the aircraft through an Airbus style Flight Control Unit (FCU), located on the glare shield above the center touchscreen. An Electronic Flight Instrument System (EFIS) panel situated to left of the FCU allowed pilots to switch between display modes and change the display range. On the pedestal, a trackball was available to select and highlight intruder information on the co-planar display.

The aircraft model that was used during the experiment was a proprietary nonlinear six degree of freedom Airbus A320 model, developed at the NLR. Intruder aircraft were modeled by point-mass models [24]. Model coefficients for these point-mass models were obtained from EUROCONTROL's BADA aircraft database [25]. The experiment was conducted with zero wind, and no turbulence. Although wind conditions will impact maneuverability, these effects were considered out of scope for the current evaluation. The own aircraft flew at altitudes between flight level FL220 and flight level FL320. This flight level range was chosen so that airspeed and vertical speed still had usable margins between minimum and maximum operating speed, and between maximum climb and descent rates, respectively.

B. Independent variables

Throughout the experiment, two independent variables were varied. *Display type* was a factor with two levels: on the co-planar separation assistance display, the RFAs could be either present or absent, see Fig. 2. Here, the display without

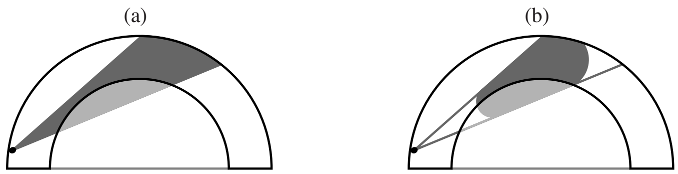


Fig. 2. The horizontal SVE for the baseline (a) and the augmented display (b). The baseline display shows two dimensional projections of constraints (called forbidden areas (FA). The augmented display gives more precise constraints (called reduced forbidden areas (RFAs)) that take the dimension orthogonal to the projection into account. The differences on the VSD are similar to the differences on the HSD. The two display conditions are otherwise equal.

RFAs was used as a baseline condition. The second factor was *conflict geometry*, which featured six levels. Scenarios differed in *phase of flight*, and *difficulty*. The phase of flight was either climb, cruise, or descent. A further distinction was made between *simple* and *difficult* scenarios. Simple conflicts always featured only one intruder, whereas in difficult scenarios, three intruders were present in each scenario. Table I gives a summary of these scenarios.

TABLE I
CONFLICT GEOMETRIES EXPERIMENT I.

	intruder	**Climb**	**Cruise**	**Descent**
Simple	ac 1	200/64/-8	270/-35/7	120/-44/0*
Difficult	ac 1	25/69/0	100/-40/8	100/-74/5*
	ac 2	210/-21/5	20/-15/6	60/-34/-8*
	ac 3	138/24/0	270/59/-10	280/-54/8*

*Values are: $\Delta\chi$ [°], Δh [$\times 100 ft$], V/S [$\times 100 ft/min$]

C. Experiment design and procedure

The experiment was designed as a within-subjects repeated-measures, where factors *display type* and *conflict geometry* were varied. The *display type* factor was introduced to illustrate the effect of the additions that the co-planar display concept features compared to the original two-dimensional separation displays. The *conflict geometry* factor was divided in phases of flight (climb, cruise, and descent), and subdivided in simple and difficult scenarios. In the simple scenarios, pilots were not expected to benefit substantially from the RFA visualization. Only in more difficult scenarios it was expected that the advantages of the RFA visualization would become noticeable. This resulted in 12 conditions ($2 \times 3 \times 2$).

After a briefing on the experiment and the functioning of the separation display, subjects performed approximately one hour of training. The experimenter would end the training session based on observed performance, and the subject's answers to informal scenario-related questions. To avoid memorizing effects, but still reach a stable level of performance and sufficient understanding of the information presented by the separation assistance interface, separate example scenarios were used for training. During the experiment, conflict scenarios were presented in a randomized block design, and conflict geometries were mirrored between display conditions. Trials were combined in four blocks of four sequential conflict scenarios. Each block started with a climb from flight level FL220 to flight level FL320, at $1,000$ ft/min, followed by a cruise segment, and then a descent back to flight level FL220, again at $1,000$ ft/min. Each block featured one conflict in the climb segment, two conflicts in the cruise segment, and one conflict in the descent segment. Starting times were different for each conflict to make it less evident for pilots when to expect each new conflict. A block lasted about 40 minutes.

The *display type* factor was kept constant over two blocks: first two blocks with one display, then two blocks with the other. The order of presentation for the display types was varied evenly over the subjects. In all conflict scenarios, multiple options in both the horizontal and vertical plane were available to solve the conflict situation, although not all options were equally fast and efficient. Intruder aircraft never maneuvered in order to solve a conflict situation, instead they just kept following their initial path.

D. Subjects and instructions to subjects

Seventeen experienced glass-cockpit pilots participated in the experiment, all male. Experience in terms of flight hours per pilot ranged from $3,000$ to $21,000$ hours (μ=10,000). None of these subjects had any previous experience with constraint-based displays. Subjects were asked to perform an experiment, where they should resolve traffic conflicts in unmanaged airspace. They were informed that the results would be used to evaluate a concept for a 3-D co-planar separation display. They were also informed that intruder aircraft would not participate in the resolution of conflicts.

In a written guide pilots received beforehand, and in a short presentation prior to the experiment, pilots were briefed on the geometrical concepts behind the display, how to use the display, and on the experimental setup. To ensure safe flight, pilots' first and foremost priority was to avoid a loss of separation at all times. When safety is ensured, pilots could explore their resolution options to optimize for efficiency. They were instructed to use the cues from the forbidden area to determine an efficient solution [18], and that their aim should be to apply a resolution that is appropriate, given the current phase of flight (i.e., climb, descent or cruise).

E. Dependent measures

Dependent measures for this experiment consisted of several objective measures. *Resolution strategy* was measured in terms of own aircraft velocity vector change dimensions, which could be any combination of a change in heading, speed and vertical speed. Path deviation, initial reaction time, and resolution duration were used as measures of *performance*. The path deviation metric differentiates between horizontal and vertical maneuvers: For horizontal maneuvers, the path deviation was characterized by the additional distance flown. In case of a vertical maneuver during the climb or descent phase, the mean deviation from the prescribed vertical speed was used. For cruise conflicts, the maximum altitude deviation from the cruising level was measured. Pilot reaction time (the time between the start of a conflict and the first selection of a resolution maneuver) and the total time of the resolution maneuver (the time between leaving and rejoining the reference trajectory) were used as metrics that allow for comparison between vertical and horizontal maneuvers. *Safety* was measured in terms of minimum separation, and the occurrence of losses of separation.

F. Experiment hypotheses

Several studies involving manual (horizontal) conflict resolutions found that pilots prefer single-axis maneuvers, keeping velocity constant [18], [26]–[28]. It was therefore hypothesized that the majority of the maneuvers would be either heading-only, or vertical speed-only (H1-1). It was also hypothesized that the resolution dimension would depend on phase of flight, i.e., that climb and descent conflicts would be solved vertically and cruise conflicts would be solved horizontally (H1-2).

Differences between the baseline display and the augmented display were only expected during difficult scenarios (scenarios with multiple intruder aircraft, which are both off-level and off-track). It was therefore hypothesized that performance would be improved with the augmented display in difficult scenarios (H1-3). Because the RFAs show more precise constraints than the projected forbidden areas, it was also hypothesized that they would result in smaller separation distances at the Closest Point of Approach (CPA) (H1-4), as previous studies showed that the precision with which constraints are presented is used by pilots to optimize their efficiency [18], [29]. The number of separation violations was hypothesized to be low, regardless of display type (H1-5).

V. Experiment I: Results

Kolmogorov-Smirnov tests on the ratio data results revealed that for none of the cases a normality assumption could be made (altitude deviations, response times and resolution times, $p < 0.001$ in each case). Therefore, only non-parametric tests were used: the Wilcoxon Signed-Rank test (test statistic z) for metrics based on ratio data that did not depend on the chosen evasive maneuver (e.g., pilot response time), and the Wilcoxon rank sum test (test statistic W) for all other metrics based on ratio data. Pearson's chi squared test (test statistic χ^2) was used for categorical metrics. Effects were considered significant at a probability level $p \leq 0.05$, where p is the probability that the null hypothesis is true.

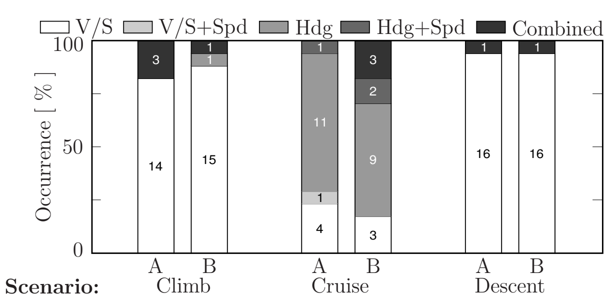


Fig. 3. Solution strategy for simple conflicts, sorted by scenario and display type (A = augmented, B = baseline) along the abscissa. The scale on the ordinate axis gives the occurrence in percent of the total per scenario, the absolute values are indicated inside the bars.

A. Resolution strategy

The resolution maneuvers in the experiment can be grouped by the flight parameters that were changed to resolve each conflict. The available maneuver options are heading, speed and vertical speed (V/S) changes. Although a resolution maneuver can consist of any possible combination of these parameters, speed-only maneuvers were never observed, and three-way combinations were rare. Therefore, Fig. 3 and Fig. 4 show resolution strategy divided into five levels: *vertical maneuvers (with and without speed)*, *horizontal maneuvers (with and without speed)*, and *combined horizontal and vertical maneuvers*. Maneuver selection will depend on conflict geometry, aircraft performance limitations, phase of flight, and personal or airline preference.

Fig. 3 shows the maneuver choice for the simple cruise, climb and descent scenarios. Each of these scenarios featured a conflict with a single intruder. The majority of the maneuvers for the climb and descent scenarios were V/S-only, regardless of display type (82% - 94%). With one exception, the direction of the change in V/S was always the same: the climb conflict was always solved by increasing the rate of climb, and the descent conflict by decreasing the rate of descent. These choices correspond to the smallest available state change for the current conflict, an efficiency strategy given to the subjects during the briefing. They can, however, also be an indication of a preference for 'staying high', to optimize for fuel efficiency.

Although the spread in solution strategy was larger than in the climb and descent scenarios, the majority of the resolutions in the simple cruise scenario was still heading only (baseline display 53%, augmented 65%). As was hypothesized (hypothesis H1-2), phase of flight was an important factor when deciding on a solution strategy. Comparison between the cruise scenario and the vertical scenarios showed a significant difference in resolution decisions ($\chi^2(2) = 56.9,\ p < 0.001$). Comparison between displays did not reveal significant effects for simple conflicts.

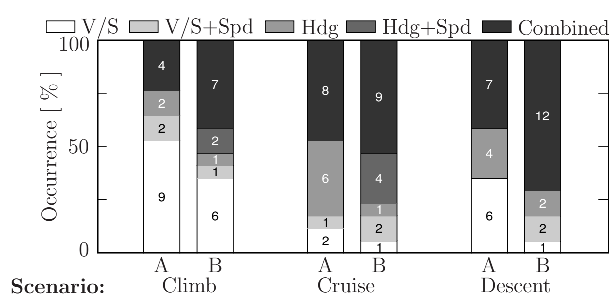


Fig. 4. Solution strategy for difficult conflicts.

Fig. 4 shows the maneuver choice for the difficult cruise, climb and descent scenarios. These scenarios each featured multiple intruders, of which only one was causing a conflict with ownship. In these scenarios, intruder aircraft were all off-level and off-track, making the maneuver space presented on the augmented display significantly different from the presentation on the baseline display. On the baseline display, this resulted in a considerable portion of the SVEs being colored, which increases the perceived severity of the conflict.

In terms of resolution strategy, the difference between the displays is visible in the number of multi-axis resolutions (V/S+SPD, HDG+SPD, or combined), which were used significantly more often with the baseline display: 77% for the baseline display, compared to 43% for the augmented display, for the climb, cruise, and descent scenario combined ($\chi^2(1) = 11.8,\ p = 0.001$). Most of these multi-axis resolutions were sequential maneuvers, rather than a single combined maneuver, regardless of display type. In other words, pilots often changed their minds after an initial resolution. The high number of multi-axis resolutions, therefore, doesn't necessarily refute the hypothesis of single-axis maneuver preference (H1-1), as the initial resolution maneuver often was single-axis. It is likely that lack of training plays a large role in this result. The difference between displays in the number of multi-axis resolutions can also be indicative of reduced situation awareness with the baseline display.

Based on pilot comments during the experiment, the multi-axis maneuvers can be classified into two categories. For the baseline display, the most often heard comment was that a pilot realized that he had made a wrong initial maneuver. This was either a maneuver that did not resolve the conflict, or a maneuver that resulted in a very inefficient resolution. A second category of maneuvers were from pilots that attempted to increase efficiency, by maneuvering in an additional direction.

Phase of flight also significantly influenced maneuver strategy in the difficult scenarios ($\chi^2(2) = 6.3,\ p = 0.04$). The cruise conflict was solved horizontally (32.4%) almost twice as much as vertically (17.6%). Similarly, the climb and descent scenarios were more often solved vertically (39.7%) than horizontally (16.2%).

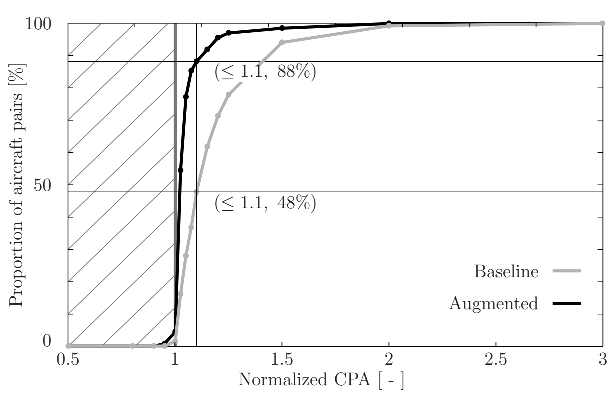


Fig. 5. Cumulative distribution graph of normalized minimum separation values. Minimum separation occurs at the closest point of approach, which is indicated as a ratio of the separation minimum along the abscissa. The number of aircraft is indicated along the ordinate axis, counted in percent of the total number of aircraft. The hatched area on the left of the graph indicates the values of CPA that violate the minimum separation constraint.

B. Safety

The separation between aircraft at the closest point of approach, compared to the minimum safe distance, was used as a measure of safety. To allow for comparison between horizontal and vertical separation, each measured value is normalized by their respective separation minimum ($5\ nmi$ horizontal, and $1,000\ ft$ vertical separation). For each measured CPA, the largest* of both normalized separation values was used. Fig. 5 shows a cumulative distribution graph of the normalized CPA values, for the augmented and baseline displays.

The separation minimum was violated in eight out of 272 measured trials, twice with the baseline display, and six times with the augmented display. In all eight cases, this occurred during a premature return to the nominal track, after resolving the conflict. In all cases, the incursion was minimal (all within 10% of the separation minimum, and 6 less than 1%). A common practice that was observed in this, but also in previous experiments with a constraint-based display [18], [29], was that after resolving a conflict, pilots are inclined to optimize their performance by returning to their nominal state as soon as possible, in small steps, while staying as close as possible to the edge of the forbidden area. In these situations, a judgment error can easily result in a (small) violation of the separation constraint. The difference between displays in the number of losses of separation was not significant ($\chi^2(1) = 2.1,\ p = 0.15$), but does illustrate that the more restrictive constraints presented by the baseline display act as an added safety margin for this kind of behavior.

C. Performance

Fig. 5 also shows that, especially with the augmented display, pilots often came within close distance of the protected zone of the other aircraft. With the augmented display, 88% came closer than 1.1 times the separation minimum, versus 48% for the baseline display. In terms of performance, this is a strong indication that pilots use the precise visualization of constraints to optimize the efficiency of their resolution. The difference in CPA distance between displays was significant ($z = -7.22,\ p < 0.001$), supporting hypothesis H1-4.

Because a direct comparison between path deviation of a horizontal maneuver and path deviation of a vertical maneuver does not make much sense, results for this performance metric will be divided in horizontal maneuvers and vertical maneuvers. For horizontal maneuvers, the path deviation was characterized by the additional distance flown. In case of a vertical maneuver during the climb or descent phase, the mean deviation from the prescribed vertical speed was used. For cruise conflicts, the maximum altitude deviation from the cruising level was also measured.

As climb and descent scenarios were mostly solved with a change in vertical speed, the mean deviation from the prescribed vertical speed was used to observe differences in performance between displays for vertical conflicts. Although there is a consistent trend of the augmented display performing better than the baseline display, this difference was only significant in the difficult descent scenario ($W = 24,\ p = 0.024$).

There are several possible reasons for the lack of significance in the remaining scenarios. First, because performance penalties of a speed change, a heading change and a vertical

*For example, if vertical separation is equal to zero, but horizontal separation is much larger than the separation margin, then both aircraft are still safely separated. The largest normalized separation value is therefore the most relevant parameter.

speed change are difficult to compare directly, the data can only be compared per maneuver category. This reduces the sample size, and therefore also the statistical power. Second, several times during the experiment it was observed that with the baseline display, pilots readjusted their resolution to a point inside the forbidden area, as soon as they realized that that particular state change was sufficient for conflict resolution. Although initially this resolution is only visualized with the RFAs, these solutions are also indirectly visualized during the state change. The color of the forbidden area communicates the urgency of a conflict, where a white forbidden area indicates a non-conflicting intruder. A pilot can therefore break off a maneuver as soon as the forbidden area turns white.

Cruise conflicts were solved 14 times out of 68 with a change in vertical speed. Although the mean deviation from the prescribed vertical speed did not reveal a significant difference, the maximum altitude deviation did differ significantly between display types, where the altitude deviation was always smaller with the augmented display ($W = 62,\ p = 0.029$). This is also an indication that pilots exploit the precise constraint visualization to optimize maneuver efficiency [18].

For horizontal maneuvers, the path deviation did not reveal a significant effect for any of the scenarios. The difficult descent and climb scenarios did show a consistent trend of the augmented display performing better than the baseline display, but contained too few samples to provide sufficient statistical power. Although on average, performance was almost equal between display types for horizontal resolutions of the simple cruise scenario, the spread was much larger for resolutions using the baseline display. Similar to the visualization of the vertical constraints, the horizontal baseline display also indirectly visualizes the constraints of the RFA. The differences in spread indicate that although pilots are able to use this indirect visualization, they do so less consistently than with the augmented display.

TABLE II
MEAN REACTION AND RESOLUTION TIMES.

Display × scenario	Baseline	Augmented
Simple	$\mu_{react} = 12.0$ [s] $\mu_{reso} = 22.4$ [s]	$\mu_{react} = 11.5$ [s] $\mu_{reso} = 20.2$ [s]
Difficult	$\mu_{react} = 20.4$ [s] $\mu_{reso} = 42.3$ [s]	$\mu_{react} = 15.1$ [s] $\mu_{reso} = 33.2$ [s]

Reaction time and resolution duration are measures that can be considered independent of the maneuver dimension, and can therefore be used as overall metrics to compare the baseline and augmented displays in simple and difficult conflict scenarios. From these measures, resolution duration is a measure of performance of a maneuver, and reaction time can be used as an indication of the difficulty experienced by pilots. Table II shows the mean reaction times and resolution durations for both displays in the simple and difficult scenarios. As hypothesized (H1-3), both these measures show significant effects of display type for the difficult conflict scenarios, but not for the simple conflict scenarios. For the simple conflict geometries, the two display variants show comparable maneuver constraints. It is therefore not expected that difficulty and resolution performance vary significantly between display types. For difficult scenarios, results for the augmented display show significantly shorter reaction times ($z = -2.32,\ p = 0.021$), and significantly shorter resolution durations ($z = -2.53,\ p = 0.012$).

VI. Experiment II: Passive SA assessment

In addition to the active conflict resolution task, a SA assessment was conducted to obtain explicit measures of SA. In this experiment, pilots were shown four static conflict scenarios, on both display variants. For each scenario, SA was probed with a timed questionnaire.

A. Apparatus

The SA assessment was performed on a single computer with a 17 inch display. The left half of the screen showed a static version of the co-planar display. Questions and multiple-choice answers were shown on the right half of the screen. A countdown timer indicated remaining time for each question. Pilots could select answers using a regular computer mouse.

B. Independent variables

Throughout the SA assessment, two independent variables were varied. Display type was a factor with two levels, which were equal to the display variants in the active experiment. The second factor was conflict geometry. Conflicting aircraft could be either on- or off-track, and either on- or off-level, resulting in four levels (2×2), see Table III.

TABLE III
CONFLICT GEOMETRIES EXPERIMENT II.

	intruder	On-level	Off-level
On-track	ac 1 ac 2	180/0/0 0/0/0	180/60/-17 * 180/-25/5 *
Off-track	ac 1 ac 2	300/0/0 75/0/0	30/30/-10 * 200/20/-2.5*

*Values are: $\Delta\chi$ [°], Δh [$\times 100 ft$], V/S [$\times 100 ft/min$]

C. Experiment design and procedure

The SA assessment followed immediately after the active experiment. It consisted of a time-limited SA query. Subjects were shown static conflict scenarios, each accompanied with thirteen time-limited multiple-choice questions regarding the geometry of the conflict, and regarding possible resolutions. At the beginning of each new scenario, subjects were given thirty seconds prior to the first question, to analyze the new conflict situation. During the questions the co-planar display remained visible, i.e., the screen was not blanked. After the assessment, subjects were asked to fill in a questionnaire form.

Similar to the active experiment, the SA assessment was designed as a within-subjects repeated-measures, where factors *display type* and *conflict geometry* were varied. Again, the augmented display was compared against a baseline display, resulting in two levels for the *display type* factor. The *conflict geometry* factor had four levels. Scenarios were always with

TABLE IV
SITUATION AWARENESS GRADE CATEGORIZATION AND INTERPRETATION.

Grade	Answer	Certainty	Interpretation
0	Incorrect	Sure	Misinformed
1	Incorrect	Unsure	Uninformed
2	Correct	Unsure	Guess/partially informed
3	Correct	Sure	Well informed

two intruding aircraft, of which only one was causing a conflict with ownship. Conflicting aircraft were either on- or off-track, and either on- or off-level, resulting in four different conflict geometries. Pilots were expected to benefit more from the RFA visualization when conflicting aircraft are increasingly off-track and off-level. This resulted in 8 conditions (2×4).

D. Subjects and instructions to subjects

The same seventeen subjects participated in this second experiment. Subjects were asked to study a set of conflict scenarios, and answer a set of geometry and conflict-resolution related multiple-choice questions. After the assessment, subjects were asked to fill in a form with questions relating to their opinion about several elements of the display. There was also opportunity for personal comments and suggestions.

E. Dependent measures

Dependent measures for this experiment are related to the SA questions, and a post-experiment questionnaire. The SA questions relate to easily identifiable information such as relative intruder position and intruder velocity, but some questions also required the subject to use information cues to predict the outcome given the current situation. The questions were categorized using Endsley's levels of awareness [21]. The subject's certainty of his answer was recorded together with the answers, following Hunt's method of measuring knowledge [30]. Using this method, the answers from the SA assessment are graded, and categorized into four groups, see Table IV. The resulting scores were averaged per pilot per level, resulting in three average SA scores per condition, for each pilot. The response time was also recorded for each answer.

The work-domain analysis that preceded the display design identifies relevant elements and relationships within the work-domain, which are arranged by level of abstraction [1], [17]. Consequently, relevant SA questions can also be based on this analysis. As a result, level 1 questions relate to conflict geometry (such as intruder location and velocity), and level 2 questions relate to principal resolution options (can a speed, vertical speed, or heading change solve the conflict). Level 3 questions require subjects to evaluate different solutions in terms of efficiency, and choose the best of a set of solutions.

Measures from the post-experiment questionnaire consisted of usefulness ratings for several individual elements of the display, and comparisons between the displays in terms of clutter, intuitiveness, SA, and workload.

F. Experiment hypotheses

Because SA level 1 questions relate to elements that are directly perceivable on both displays, it was hypothesized

TABLE V
COMPARISON BETWEEN DISPLAY TYPES OF THE SA SCORES.

Level × scenario	SA Level 1	SA Level 2	SA Level 3
Main effect	$\chi^2(1) = 0.4$ $p = 0.540$ ○	$\chi^2(1) = 10.7$ $p = 0.001$ ⋆⋆	$\chi^2(1) = 20.7$ $p < 0.001$ ⋆⋆
On-level/On-track	$z = -0.378$ $p = 0.705$ ○	$z = -0.556$ $p = 0.579$ ○	$z = -1.633$ $p = 0.102$ ○
On-level/Off-track	$z = -1.000$ $p = 0.317$ ○	$z = -1.016$ $p = 0.309$ ○	$z = -1.173$ $p = 0.241$ ○
Off-level/On-track	$z = -1.000$ $p = 0.317$ ○	$z = -1.885$ $p = 0.059$ ○	$z = -2.362$ $p = 0.018$ ⋆
Off-level/Off-track	$z = -0.136$ $p = 0.892$ ○	$z = -3.430$ $p < 0.001$ ⋆⋆	$z = -3.084$ $p = 0.002$ ⋆⋆

⋆⋆ significant; ⋆ marginally significant; ○ not significant.

that the SA score for level 1 questions would be very high, regardless of display type (H2-1). Since the augmented display visualizes more higher-level information and relationships, it was also hypothesized that the SA scores between displays would diverge increasingly, with higher SA levels (H2-2). An interaction with scenario was expected for this effect, as the difference between displays becomes increasingly pronounced for scenarios with off-level or off-track intruders (H2-3).

Results for the response time were expected to show an interaction between scenario and question SA level (H2-4). Because the augmented display reveals relationships in scenarios that are off-level or off-track, which the baseline display does not show, questions that relate to this information (i.e., level 3 SA questions) should be quicker to evaluate when using the augmented display.

VII. EXPERIMENT II: RESULTS

Similar to the first experiment, a normality assumption could not be made for any of the ratio data (reaction times, $p < 0.05$ for all SA levels). A Friedman two-way ANOVA (test statistic χ^2) was therefore used to evaluate main effects of the display factor. The Wilcoxon Signed-Rank test (test statistic z) was used to evaluate the effect of display per scenario. With a Bonferroni correction of 5* for the SA scores, results were considered significant at a probability level $p \leq 0.01$. Results with a probability level $0.01 < p \leq 0.05$ were considered marginally significant. Response time results were only analyzed in terms of main effects, resulting in a Bonferroni correction of 2. Here, results were considered significant at a probability level $p \leq 0.025$.

A. Situation awareness scores

The situation awareness scores from the experiment were grouped using Endsley's three levels of awareness [21], and are shown in Fig. 6, for each combination of display type and scenario. These SA scores will depend on conflict geometry

*A Bonferroni correction implies that the significance level is divided by the number of tests on a particular set of data. For these results this was one main effects test, and four post-hoc tests (one for each scenario level).

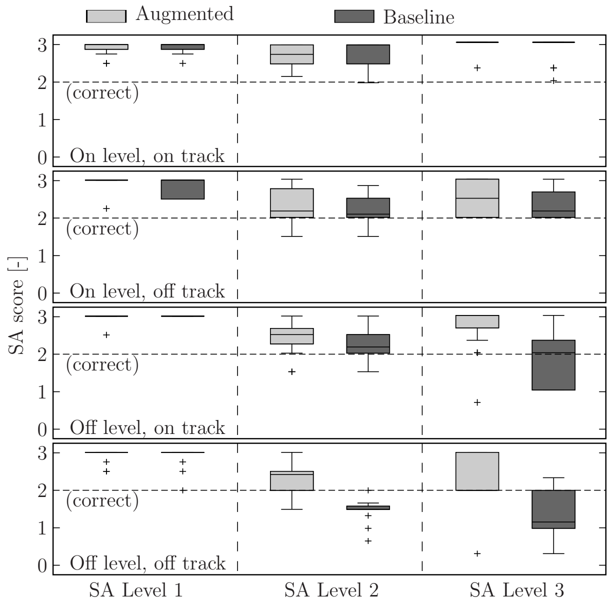


Fig. 6. SA scores, averaged per pilot, and sorted by display type, scenario, and SA level. The three columns correspond to the three SA levels. The four rows each correspond to a scenario, as indicated in the bottom-left corner of each row. The scale on the ordinate axis gives the SA score, see Table IV.

and accuracy of the visualization, but also on other factors that influence the buildup of SA, such as attention and workload.

As hypothesized (H2-1), the first column in Fig. 6 shows that the majority of the subjects (92 - 100%) managed to achieve the highest SA score for level one questions, regardless of scenario or display. A comparison between display types for SA level one therefore also did not reveal any significant effects, see the first column in Table V.

A main effects analysis (see Table V) showed that, as hypothesized (H2-2), display becomes a significant factor for SA scores at awareness levels two and three: As can be seen in Fig. 6, subjects scored consistently lower with the baseline display. A post-hoc analysis revealed that this effect increases when scenarios become increasingly off-level and off-track: Table V shows that the effect of display is only significant for level two and level three scores in the off-level and off-track scenario. This supports hypothesis H2-3, which stated that scenario type would influence SA scores between displays.

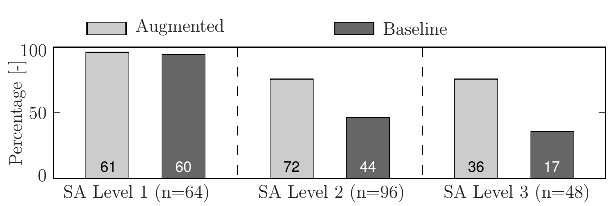


Fig. 7. Percentage of correct and sure answers for the off-track and off-level scenario, grouped per display type and SA level. The columns in the figure table correspond to the three SA levels. The scale on the ordinate axis gives the amount of correct and sure answers, in percent of the total per display type per SA level. Absolute counts are indicated in the bottom of each bar.

Fig. 7 illustrates the percentage of correct and sure answers, at each SA level, for the off-track and off-level scenario. According to Hunt, only these answers correspond with usable knowledge [30]. Fig. 7 shows that, although the augmented display scores consistently higher than the baseline display, subjects still could not maintain perfect SA with the augmented display, despite the more accurate visualization. This can be –at least partly– caused by lack of training, combined with the inherent complexity of the separation problem.

TABLE VI
EFFECTS OF DISPLAY AND SCENARIO ON RESPONSE TIMES.

	SA Level 1	SA Level 2	SA Level 3
Display	$\chi^2(1) = 1.1$ $p = 0.300$ ∘	$\chi^2(1) = 0.04$ $p = 0.851$ ∘	$\chi^2(1) = 0.19$ $p = 0.187$ ∘
Scenario	$\chi^2(3) = 27.3$ $p < 0.001$ ⋆⋆	$\chi^2(3) = 16.2$ $p = 0.001$ ⋆⋆	$\chi^2(3) = 10.9$ $p = 0.012$ ⋆⋆

⋆⋆ significant; ⋆ marginally significant; ∘ not significant.

B. Response time

Fig. 8 shows the response times for the SA questions, averaged per pilot, for each combination of display type and scenario. It can be seen that although a trend in favor of the augmented display is visible in the data, it is markedly less pronounced than the effect observed for the SA score results. A main effects test therefore also did not reveal a significant effect of the display factor, see Table VI.

The response time results show larger variation between scenarios and SA levels. The response time increases with increasing conflict complexity, as well as with increasing SA level. A main effects test showed that the effect of scenario is significant for all levels of SA, see Table VI. These results therefore indicate that difficulty is a determining factor for response time, but that the augmented display does not enable subjects to evaluate complex situations more quickly.

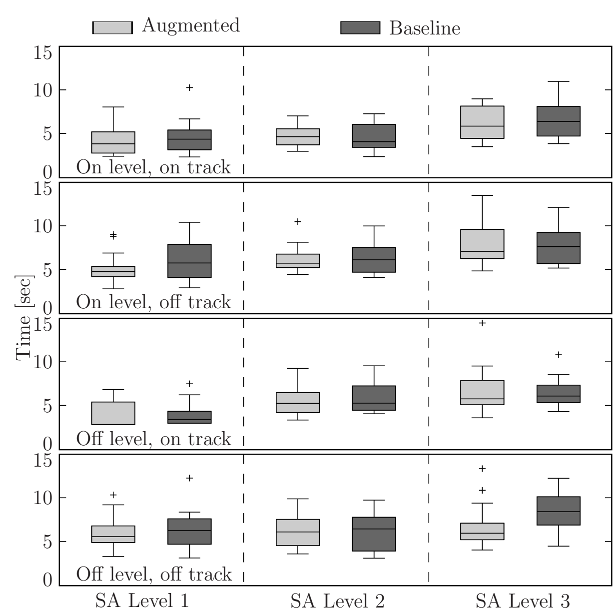


Fig. 8. Response times, averaged per pilot, sorted by display, scenario, and SA level. The scale on the ordinate axis gives the response time in seconds.

C. Post-experiment questionnaire

The post-experiment questionnaire allowed subjects to give an overall rating of each display in terms of usability, and to express their preference for either display in terms of clutter, intuitiveness, situation awareness, and workload. They were also asked to rate the usefulness of several individual elements of the display. Although the sample size of 17 subjects is too small to obtain reliable results for such subjective data, these results can be used to highlight persistent trends and opinions.

Both in the overall display ratings and the display preference questions, the augmented display scored consistently better than the baseline display. An often-heard comment was that subjects could better relate information between the two displays with the augmented display, than with the baseline display. Aside from preference with regard to clutter, subjects preferred the augmented display almost without exception (94-100%). Preference for the augmented display with regard to clutter was slightly lower (76%). Here, several subjects indicated that they did not prefer either display. One pilot remarked that while the RFAs in the augmented display increase clutter, it was ‘good clutter’. This is consistent with Tufte’s views on the use of visual details (“To clarify, add detail”) [31]. Most pilots mentioned, though, that some form of de-cluttering would be essential in high-density traffic situations (i.e., more than the 3 intruders in the current experiment). In terms of SA, subjects mentioned that the RFAs allowed for a quicker assessment of the consequences of specific resolutions.

When asked to rate the usefulness of individual elements of the display, the majority of the subjects assigned the highest rating to the more conventional intruder symbols. The intruder symbols on the VSD, however, were mostly rated lower than the same symbols on the HSD. This is an indication that even though subjects have a very positive attitude towards the new display, and the novel visualizations, they remain biased towards appreciating familiar functionality.

Most subjects also used the opportunity to give one or more suggestions for future design iterations of the co-planar display concept. A suggestion that was prompted by almost every subject was to add the ability to zoom in on the SVEs (especially on the HSD, where it was smallest in the current simulation). An other repeated suggestion related to the addition of intent information: subjects indicated that they would appreciate the ability to see where intruders that are climbing or descending would level off, and the consequences of the own aircraft leveling off at a certain altitude. Finally, several subjects were interested to know how the concept would function when all aircraft in a conflict would use such an interface, a set-up that has already been investigated in an earlier experiment for purely horizontal maneuvers [18].

VIII. Discussion

The displays in this study are designed to help a pilot understand the reasoning behind automated decisions, by showing constraints and relationships within the work domain. This work domain information invariably forms the premise on which automation bases its actions, and is therefore also invaluable to pilots when they need to judge the automation’s functioning. Although this experiment did not feature automated conflict resolution, and can therefore not be used to evaluate interaction between human and automation, the pilots’ resolution decisions do give insight in how the information on the display is used by pilots, and how it affects their SA.

The objective measures presented in this paper show several trends. An effect that is seen in several other studies was that many resolution maneuvers were single-axis. Current results showed, however, that this effect diminished for more difficult scenarios. It can be argued that this was mostly a training issue, as pilot comments during the experiment often indicated that an erroneous initial resolution choice was made. Several pilots also mentioned in the post-experiment questionnaire that more training would be required to be able to understand and properly use the interface. Occasionally, pilots also initiated a multi-axis maneuver ‘just to see what happens’, which can be considered an artifact of volunteer test subjects in an experiment. In some cases pilots indicated that they made a multi-axis maneuver to improve efficiency. Path deviation measurements, however, showed that this was never the result.

Although difficult scenarios resulted in more multi-axis maneuvers, this effect did depend on display configuration, where multi-axis maneuvers were made more often with the baseline display. Since many of the multi-axis maneuvers were corrections of an erroneous initial single-axis maneuver, this can be an indication that, with the same (limited) level of training, pilots performed better with the augmented display. They made fewer errors, indicating a beneficial effect on traffic awareness of the augmented display.

As hypothesized (H1-2), phase of flight had a significant effect on resolution choice, regardless of scenario difficulty. This preference can be seen as the result of a procedural constraint (i.e., phase of flight) that is however not directly visible on the display. This indicates that pilots can use the presented constraints, and apply them to other rules and procedures. This is classified by Rasmussen as Rule Based Behavior [32]. Ideally, the interface should support pilots at all levels of cognitive behavior, while not forcing them to control at a higher level than necessary [33].

A persistent result found in this experiment, and earlier experiments with a constraint-based display, is that after reaching a conflict-free state, the majority of the subjects returned to their original track in several small steps, following the edge of the constraint area as closely as possible [18], [29]. This behavior can be attributed to showing precise constraints: when maneuver limits are visualized with high precision, human operators will use that precision to maximize their efficiency. As a result, the majority of the CPA’s stay within 110% of the separation margin (augmented 88%, baseline 48%). This ‘hunting’ behavior, however, also gives rise to judgment errors, and consequently also losses of separation, which occurred 8 times in the experiment. Although the incursions were very small, this is still an undesired side effect of showing precise constraints. Another possible influential factor in this behavior relates to the perceived severity of a violation. A minimal incursion of a separation limit will be judged differently than for example a violation of the minimum airspeed limit. As a result, pilots may permit the occasional (minor) loss of separation, in order to increase efficiency.

The experiments in this study compared two displays, where the main difference between the two was the accuracy of the presented constraints. Where the augmented display presented precise constraints, the baseline display was more conservative. Because the color of a forbidden area communicates the state of conflict (white areas indicate non-conflicting intruders), subjects were able to find resolutions with the baseline display that were still inside of the presented constraints. Several subjects who started the experiment with the baseline display, sometimes applied this same strategy with the augmented display (searching for solutions within a constraint area). With the RFAs, however, this is never a valid option. This type of mode or strategy confusion can become an issue in comparative experiments, where levels of an independent factor lie very close to each other. This effect should be taken into account for such experiments.

The SA assessment revealed that display becomes a significant factor in complex scenarios, for high-level SA probes. These scenarios consist of off-track and off-level geometry, which reveal the difference between the basic triangular forbidden areas and the RFAs. In these situations, even though the baseline display and the augmented display present the same *type* of information (horizontal and vertical maneuvering constraints), they differ in the *accuracy* of that information. Although the extra information that is hidden in the baseline display can still be derived to some extent, this requires additional cognitive work. The fact that response time was not influenced by display type (even though pilots indicated in the post-experiment questionnaire that the RFAs allowed them to quicker assess the consequences of resolution maneuvers), however, indicates that subjects used the presented constraints on both displays in the same way. The differences in SA scores therefore mostly relate to the accuracy of the constraints.

Although the augmented display scores consistently higher than the baseline display, SA scores still drop with higher SA levels. This is in line with a notion put forward by Vicente, who states that ecological interfaces were never intended to be used by untrained operators [34]. Proper training is therefore an important issue for these concepts and their evaluation. The fact that many subjects assigned the highest usefulness ratings to the more classical TCAS symbols can therefore also indicate that they do not fully understand what information is required to perform the new task of conflict resolution, and what this means for the requirements on the visualization of this information. Nevertheless, resolution performance was high, even with insufficiently trained subjects. Because these kinds of displays make several complex relationships directly perceivable, they relieve pilots from cognitive work. This transforms tasks that ordinarily require SA at the projection level to simple tasks of perception and observation, allowing pilots to perform well, despite insufficient training.

In comparison with the baseline display, the augmented display reveals more properties and relations that are inherent to the work-domain. In the search for a display that properly supports pilots' SA, the trade-off will always be between showing more information on the one hand, and maintaining a clear, understandable and uncluttered display on the other hand. The results in this study show that performance and SA benefit from the improved accuracy of the constraint visualizations, and that pilot behavior is consistent with previous evaluations of constraint-based displays. Together with the preference ratings from the post-experiment questionnaire, these results also give no indication that this increased accuracy forms a problem in terms of display clutter. Nevertheless, future design iterations should continue to focus on the trade-off between information density and clutter.

IX. Conclusions

An experiment was conducted to evaluate a concept for a constraint-based co-planar self-separation display. The display shows performance and traffic constraints on maneuvering, as well as interactions between the two planar projections. A comparison was made between this concept and a baseline display that did not show these interactions, in an active conflict resolution experiment, and a passive SA assessment.

Results showed that although pilots performed well with either display, performance was consistently better with the augmented display: resolutions were more efficient, pilots made fewer errors in their initial resolutions, and situation awareness scores were higher. Similar to previous studies, a preference for single-axis maneuvers was found, although this effect was smaller for difficult scenarios.

A persistent effect observed with this and other constraint-based displays is that pilots use the precision of the constraint visualization to optimize their efficiency. This type of behavior sometimes leads to over-optimization.

X. Acknowledgments

The authors gratefully acknowledge the pilots that participated in this study, and would like to thank NLR software expert Michiel J. D. Valens for his help during the experiment.

References

[1] J. Ellerbroek, K. C. R. Brantegem, M. M. van Paassen, and M. Mulder, "Design of a Co-Planar Airborne Separation Display," *IEEE Transactions on Human-Machine Systems*, vol. 43, no. 3, pp. 277–289, 2013.

[2] S. B. J. van Dam, M. Mulder, and M. M. van Paassen, "Ecological Interface Design of a Tactical Airborne Separation Assistance Tool," *IEEE Transactions on Systems, Man, and Cybernetics, part A: Systems and Humans*, vol. 38, no. 6, pp. 1221–1233, 2008.

[3] F. M. Heylen, S. B. J. van Dam, M. Mulder, and M. M. van Paassen, "Design and Evaluation of a Vertical Separation Assistance Display," in *AIAA Guidance, Navigation, and Control Conference and Exhibit*, Honolulu (HI), 2008.

[4] Radio Technical Commission for Aeronautics, "Airborne Conflict Management: Application Description V2.5," Federal Aviation Authorities, Tech. Rep. RTCA SC-186, 2002.

[5] SESAR Consortium, "SESAR Definition Phase D3: The ATM Target Concept," Eurocontrol, Tech. Rep. DLM-0612-001-02-00, 2007.

[6] D. A. Norman, "The "Problem" of Automation: Inappropriate Feedback and Interaction, not "Over-Automation"," *Philosophical Transactions of the Royal Society of London*, vol. 327, no. 1241, pp. 585–593, Apr. 1990.

[7] N. B. Sarter and D. D. Woods, "Pilot Interaction With Cockpit Automation: Operational Experiences With the Flight Management System," *The International Journal of Aviation Psychology*, vol. 2, no. 4, pp. 303–321, 1992.

[8] G. Lintern, T. Waite, and D. A. Talleur, "Functional Interface Design for the Modern Aircraft Cockpit," *The International Journal of Aviation Psychology*, vol. 9, no. 3, pp. 225–240, 1999.

[9] A. M. Bisantz and A. R. Pritchett, “Measuring the Fit Between Human Judgments and Automated Alerting Algorithms: A Study of Collision Detection,” *Human Factors*, vol. 45, no. 2, pp. 266–280, 2003.

[10] S. W. A. Dekker, “On the Other Side of Promise: What Should We Automate Today?” in *Human Factors for Civil Flight Deck Design*, D. Harris, Ed. Ashgate Pub Ltd, 2004, pp. 141–155.

[11] T. Inagaki, “Design of Human–Machine Interactions in Light of Domain-Dependence of Human-Centered Automation,” *Cognition, Technology & Work*, vol. 8, no. 3, pp. 161–167, 2006.

[12] A. Q. V. Dao, S. Brandt, V. Battiste, K. P. Vu, T. Strybel, and W. W. Johnson, “The Impact of Automation Assisted Aircraft Separation on Situation Awareness,” in *Human Interface and the Management of Information. Information and Interaction*. Springer, 2009, pp. 738–747.

[13] C. Meckiff and P. Gibbs, “PHARE Highly Interactive Problem Solver,” Eurocontrol, Tech. Rep. 273/94, Nov. 1994.

[14] R. Azuma, H. Neely, M. Daily, and M. Correa, “Visualization of Conflicts and Resolutions in a “Free Flight” Scenario,” in *Proceedings of IEEE Visualization*, 1999, pp. 433–436.

[15] J. M. Hoekstra, R. N. H. W. van Gent, and R. C. J. Ruigrok, “Designing for Safety: the Free Flight Air Traffic Management Concept,” *Reliability Engineering and System Safety*, vol. 75, pp. 215–232, 2002.

[16] R. Canton, M. Refai, W. W. Johnson, and V. Battiste, “Development and Integration of Human-Centered Conflict Detection and Resolution Tools for Airborne Autonomous Operations,” in *International Symposium on Aviation Psychology*, 2005, pp. 115–120.

[17] J. Ellerbroek, M. Visser, S. B. J. van Dam, M. Mulder, and M. M. van Paassen, “Design of an Airborne Three-Dimensional Separation Assistance Display,” *IEEE Transactions on Systems, Man, and Cybernetics, part A: Systems and Humans*, vol. 41, no. 6, pp. 863–875, 2011.

[18] J. Ellerbroek, M. M. van Paassen, and M. Mulder, “Evaluation of a Separation Assistance Display in a Multi-Actor Experiment,” *IEEE Transactions on Human-Machine Systems, submitted*, 2011.

[19] Radio Technical Commission for Aeronautics, “Minimal Operational Performance Standards for Traffic Alert and Collision Avoidance System 2 (TCAS2) Airborne Equipment,” Federal Aviation Authorities, Tech. Rep., 2002.

[20] J. Uhlarik and D. A. Comerford, “A Review of Situation Awareness Literature Relevant to Pilot Surveillance Functions,” Federal Aviation Authorities, Tech. Rep. DOT/FAA/AM-02/3, 2002.

[21] M. R. Endsley, “Toward a Theory of Situation Awareness in Dynamic Systems,” *Human Factors*, vol. 37, no. 1, pp. 32–64, 1995.

[22] J. M. Flach, “Situation Awareness: Proceed with Caution,” *Human Factors*, vol. 37, no. 1, pp. 149–157, 1995.

[23] A. M. McGowan and S. P. Banbury, “Evaluating Interruption-Based Techniques using Embedded Measures of Driver Anticipation,” in *A Cognitive Approach to Situation Awareness: Theory and Application*, S. P. Banbury and S. Tremblay, Eds. Ashgate, 2004, pp. 176–192.

[24] D. M. Henderson, *Applied Cartesian Tensors for Aerospace Simulations*, J. A. Schetz, Ed. American Institute of Aeronautics and Astronautics, 2006.

[25] A. Nuic, D. Poles, and V. Mouillet, “BADA: An Advanced Aircraft Performance Model for Present and Future ATM Systems,” *International Journal of Adaptive Control and Signal Processing*, vol. 24, no. 10, pp. 850–866, 2010.

[26] J. M. Hoekstra, “Designing for Safety: The Free Flight Air Traffic Management Concept,” Ph.D. dissertation, Delft University of Technology, The Netherlands, 2001.

[27] C. D. Wickens, J. Helleberg, and X. Xu, “Pilot Maneuver Choice and Workload in Free Flight,” *Human Factors and Ergonomics Society Annual Meeting Proceedings*, vol. 44, no. 2, pp. 171–188, 2002.

[28] A. L. Alexander, C. D. Wickens, and D. H. Merwin, “Perspective and Coplanar Cockpit Displays of Traffic Information: Implications for Maneuver Choice, Flight Safety, and Mental Workload,” *The International Journal of Aviation Psychology*, vol. 15, pp. 1–21, 2005.

[29] C. Borst, M. Mulder, and M. M. van Paassen, “Design and Simulator Evaluation of an Ecological Synthetic Vision Display,” *Journal of Guidance, Control and Dynamics*, vol. 33, no. 5, pp. 1577–1591, 2010.

[30] D. P. Hunt, “The Concept of Knowledge and How to Measure It,” *Journal of Intellectual Capital*, vol. 4, no. 1, pp. 100–113, 2003.

[31] E. R. Tufte, *Envisioning Information*. Cheshire, CT: Graphics Press, 1990.

[32] J. Rasmussen, “Skills, Rules, Knowledge; Signals, Signs, Symbols, and Other Distinctions in Human Performance Models,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 13, pp. 257–266, 1983.

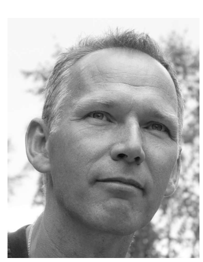
[33] K. J. Vicente and J. Rasmussen, “Ecological Interface Design: Theoretical Foundations,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 22, no. 4, pp. 589–606, 1992.

[34] K. J. Vicente, *Cognitive Work Analysis Toward Safe, Productive and Healthy Computer-Based Work*. Lawrence Erlbaum Associates Mahwah, NJ, 1999.

Joost Ellerbroek received the M.Sc. degree in aerospace engineering from the Delft University of Technology, The Netherlands, in 2007, where he is currently working toward the Ph.D. degree. His Ph.D. work concentrates on the design and validation of an interface that supports interaction with airborne separation automation. The research presented in this paper is part of his thesis.

Koen C. R. Brantegem received the M.Sc. degree (cum laude) from the Delft University of Technology, The Netherlands, in 2011. He graduated within the control and simulation section on his thesis entitled “Ecological 2-D Coplanar Airborne Separation Assurance System”. The results of his work are incorporated in this paper. He is currently working towards obtaining a commercial pilot license.

M. M. (René) van Paassen received the M.Sc. degree (1988, cum laude) from the Delft University of Technology, The Netherlands, and a Ph.D. (1994), on the neuromuscular system of the pilot’s arm. He thereafter was a Brite/EuRam Research Fellow with the University of Kassel, and a post-doc at the Technical University of Denmark. Currently, he is associate professor at the faculty of Aerospace Engineering, Delft University of Technology. His work ranges from studies of perceptual processes and manual control to complex cognitive systems.



Nico de Gelder received the M.Sc. degree from the Delft University of Technology, The Netherlands, in 1987. He started as flight test engineer at Fokker Aircraft, where he later became senior specialist avionics. He joined the National Aerospace Laboratory NLR in 1996, working on cockpit HMI designs and new ATM system concepts. He currently participates in the RTCA SC-186/EUROCAE WG-51 standardization committee, the SESAR and CleanSky technology initiatives, and national research projects on ADS-B and Flight Deck Interval Management.

Max Mulder received the M.Sc. (1992) and Ph.D. degrees (1999, cum laude) from the Delft University of Technology, The Netherlands, for his work on the cybernetics of tunnel-in-the-sky displays. He is currently Full Professor and Head of the Control and Simulation Section, Faculty of Aerospace Engineering, Delft University of Technology. His research interests include cybernetics and its use in modeling human perception and performance, and cognitive systems engineering and its application in the design of "ecological" human-machine interfaces.