Design and Evaluation of a Co-Planar Separation Display

J. Ellerbroek
K.C.R. Brantegem
M.M. van Paassen
M. Mulder

Follow this and additional works at: https://corescholar.libraries.wright.edu/isap_2013

Part of the Other Psychiatry and Psychology Commons

Repository Citation

This Article is brought to you for free and open access by the International Symposium on Aviation Psychology at CORE Scholar. It has been accepted for inclusion in International Symposium on Aviation Psychology - 2013 by an authorized administrator of CORE Scholar. For more information, please contact library-corescholar@wright.edu.
This paper describes the design and evaluation of a co-planar constraint-based airborne separation assistance display. The display is a combination of previous single-plane presentations, with additional visualization of the interactions that exist between these planes. Each of these displays combines a spatial representation of the airspace with a velocity maneuver space, that relates own aircraft maneuver variables to the shape and affordances of the airspace. The evaluation presented in this paper consisted of two experiments: an active conflict resolution task, and a passive SA assessment. Both experiments compare the co-planar concept with a baseline display that is very similar, but does not visualize planar interactions. Results showed that although pilots performed well with either display, performance was consistently better with the augmented display.

The work presented in this paper is part of an ongoing study, which applies a constraint-based approach to the design of a 3-D airborne separation assistance interface. In this paper, a co-planar display is proposed that presents constraints on maneuvering in a ‘velocity action space’, that is overlaid on traditional moving-map displays. This paper also presents the results from an evaluation of the display concept.

In response to continuously growing levels of automation, several studies argue that proper human-automation interaction, and appropriate trust in automation require transparent and understandable functioning of automation. The interface should provide a window to the reasoning and functioning of the automation (Inagaki, 2006). 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. The evaluation of such a display should therefore focus on how elements of the display affect the operator's understanding of the traffic situation. The experiments in this paper were designed to serve this purpose. An active conflict resolution experiment was performed to evaluate how operator performance and behavior are influenced by the visualization. A second experiment consisted of a passive situation awareness assessment (See also Ellerbroek et al. (2013)).

**Figure 1:** Concept for a co-planar separation assistance display. This figure shows a Horizontal Situation Display (1), and a Vertical Situation Display (2), with added separation assistance overlays (3 and 4).

**The Interface**

Fig. 1 illustrates 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 are briefly described in this section. See Ellerbroek et al. (2013) for a more elaborate review of this display. In this concept, the 3-D traffic situation is visualized in two orthogonal, 2-D views: a top-down view (1), and a side view (2). 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 (3). In addition, maneuvering constraints are shown on both displays through velocity action-space overlays (4, 5), referred to as State-Vector Envelopes (SVEs) in the remainder of this text.

The horizontal SVE (4) shows the horizontal maneuver space, in terms of track angle and airspeed. Its boundaries 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 (5) shows a vertical maneuvering space, in terms of airspeed and vertical speed. Similar to the horizontal SVE, its boundaries 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 steady climb at each velocity. The bottom edge indicates maximum steady descent. Combined, the areas within these envelopes describe all reachable velocity vectors of the aircraft.
Intruder aircraft that are within detection range will reduce the available maneuver space in the horizontal and vertical SVEs. The reduced forbidden areas (RFAs) give the most precise representation of these constraints. On the Horizontal Situation Display (HSD), an RFA gives the constraints imposed by an intruder on ownship track angle and airspeed, for the current value of ownship vertical speed. On the Vertical Situation Display (VSD), an RFA gives intruder-imposed constraints on ownship airspeed and vertical speed, for the current ownship heading. These RFAs result from the intersection between a flat cutting-plane, and the 3-D forbidden area (FA): the set of ownship velocities that result in a conflict with the corresponding intruder, see Figures 2 and 3. These RFAs are subsets of the projected FAs, that were used in the earlier constraint-based separation displays (Dam et al. 2008).

The projected FAs are still shown in combination with the RFAs, as they provide several Situation Awareness (SA) related cues, as well as an outer limit on the shape and size of the RFA, when a perpendicular flight parameter is modified. Figure 4 shows how these projections are constructed.

Conflict urgency is explicitly indicated on the display using intruder symbology similar to the existing Traffic Collision Avoidance (TCAS) system. 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 FA triangles and RFAs on both displays are colored according to the urgency of the conflict between ownship and the corresponding intruder.

**Experimental Evaluation**

Two experiments, an active conflict resolution task and a passive situation awareness assessment, were conducted to evaluate the co-planar separation assistance display. Both experiments compared the new concept to a baseline display that was very similar, but did not show how constraints interacted between projection planes.

**Method**

Both the active conflict resolution experiment and the passive SA assessment were designed as a within-subjects repeated-measures, where factors display type and conflict geometry were varied. In both experiments, display type had two levels: reduced forbidden areas could be either present or absent. This factor was introduced to illustrate the effect of the additions that the co-planar concept features compared to the original 2-D separation displays. In the active experiment, the conflict geometry factor differentiated between simple and difficult scenarios, and between phases of flight (climb, cruise, and descent), resulting in 6 scenarios (2 X 3). Simple conflicts always featured only one intruder, close to the own flight level and track. In difficult scenarios, three intruders were present in each scenario, which could be both off-level and off-track. In the passive SA assessment, conflicts could be either on-track or off-track, and on-level or off-level, resulting in 4 scenarios (2 X 2).

In both experiments, conditions were presented in a randomized block design. In the active experiment, 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 segment featured at least one conflict. A block lasted approximately 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 scenarios in both experiments, multiple options, both horizontal and vertical, were available to solve the conflict, although not all options were equally fast and efficient. In the active experiment, intruder aircraft never maneuvered in order to solve a conflict situation, instead they just kept following their initial path.

**Dependent measures.** Resolution strategy was an objective measure, measured in terms of velocity vector change dimensions. This could be a combination of a change in heading, speed and vertical speed (V/S). Path deviation and the initial reaction time 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 V/S was used. For cruise conflicts, the maximum altitude deviation from the cruising level was measured. Pilot reaction time and the total time of the resolution maneuver, i.e., 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.

SA questions in the passive assessment 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 SA (Endsley, 1995), and the subject’s certainty of his answer was recorded together with the answers. The combination results in a grade, that categorizes answers into four groups, following Hunt’s method of measuring knowledge (Hunt, 2003).

Hypotheses. Several previous studies found that pilots prefer single-axis maneuvers, keeping velocity constant (Alexander et al., 2005; Ellerbroek et al., 2011). It was therefore hypothesized that most maneuvers would be either heading-only, or V/S-only. It was also hypothesized that maneuver choice would depend on phase of flight, i.e., that climb and descent conflicts would be solved vertically and that cruise conflicts would be solved horizontally.

Differences between displays were only expected during difficult scenarios. It was therefore hypothesized that performance would be improved with the augmented display in difficult scenarios, and that SA would be higher, especially at the projection level. Because the RFAs show more precise constraints than the projected FAs, it was also hypothesized that they would result in lower separation at the Closest Point of Approach (CPA), as previous studies showed that the precision with which constraints are presented is used by pilots to optimize their efficiency (Ellerbroek et al., 2011). The number of separation losses was hypothesized to be low, regardless of display.

Because SA level 1 questions relate to elements that are directly perceivable on both displays, it was hypothesized that SA level 1 score would be high, regardless of display type. 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. 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.

Results

Kolmogorov-Smirnov tests on the ratio data revealed that a normality assumption could never be made (altitude deviations, CPA values, 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 maneuver, and the Wilcoxon rank sum test (test statistic \(W\)) for all other metrics based on ratio data. Pearson’s chi squared test (test statistic \(\chi^2\)) was used for categorical metrics. Effects were considered significant at a probability level \(p <= 0.05\), where \(p\) is the probability that the null hypothesis is true.

Resolution strategy. Figures 5 and 6 show resolution strategy divided into five levels: vertical maneuvers (with and without speed), horizontal maneuvers (with and without speed), and combined three-way maneuvers. Maneuver selection will depend on conflict geometry, phase of flight, performance limitations, and personal or airline preference. Fig. 5 shows the maneuver choice for the simple cruise, climb and descent scenarios. 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, 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.

Figure 5: Resolution strategy for simple conflicts, sorted by scenario and display type (A=augmented, B=baseline).  
Figure 6: Resolution strategy for difficult conflicts.
Fig. 6 shows the maneuver choice for the difficult cruise, climb and descent scenarios. 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 amount of multi-axis resolutions, therefore, doesn't necessarily refute the hypothesis of single-axis maneuver preference, as the initial resolution maneuver often was single-axis. It is more 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 SA 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 didn’t resolve the conflict, or a maneuver that failed 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%).

**Safety.** Fig. 7 shows a cumulative distribution graph of the normalized CPA values, for both displays. Separation was violated in 8 out of 272 measured trials, twice with the baseline display, and 6 times with the augmented display. In all eight cases, this occurred during a premature return to the nominal track, and 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 (Ellerbroek et al., 2011), 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 FA. In these situations, a judgment error can easily result in a (small) separation violation. 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 on the baseline display can act as an added safety margin.

**Performance.** Fig. 7 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 ($z = -7.22, p < 0.001$). For horizontal maneuvers, path deviation never revealed a significant effect. The difficult descent and climb scenarios did show a consistent improvement of the augmented display over the baseline display, but contained too few samples to provide sufficient statistical power.

As climb and descent scenarios were mostly solved with a change in V/S, the mean deviation from the prescribed V/S will be 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 ($W = 24, p = 0.024$). Cruise conflicts were solved 14 times out of 68 with a change in V/S. Although the mean deviation from the prescribed V/S 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 (Ellerbroek et al., 2011).

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 display conditions. Here, resolution duration is a measure of performance of a maneuver, and reaction time is indicative of the difficulty experienced by pilots. As hypothesized, 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 ($z = -2.53, p = 0.012$).
Situation awareness. The SA scores from the experiment were grouped using Endsley's three levels of awareness (Endsley 1995), and are shown in Fig. 8, for each combination of display type and scenario. These SA scores will depend on conflict geometry and accuracy of the visualization, but also on other factors that influence the buildup of SA, such as attention and workload. As hypothesized, the first column in Fig. 8 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 displays for SA level 1 therefore also did not reveal significant effects, see Table 1.

A main effects analysis (see Table 1) showed that, as hypothesized, display becomes a significant factor for SA scores at levels two and three: As can be seen in Fig. 8, 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 1 shows that the effect of display is only significant for level 2/3 scores in the off-level and off-track scenario. This supports the hypothesis that conflict geometry would influence SA scores between displays.

Table 1: Comparison between display types of the SA scores.

<table>
<thead>
<tr>
<th>Level x scenario</th>
<th>SA level 1</th>
<th>SA level 2</th>
<th>SA level 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Main effect</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>On-level/on-track</td>
<td>(\chi^2(1) = 0.4)</td>
<td>(\chi^2(1) = 10.7)</td>
<td>(\chi^2(1) = 20.7)</td>
</tr>
<tr>
<td></td>
<td>(p = 0.540)</td>
<td>(p = 0.001)</td>
<td>(p &lt; 0.001)</td>
</tr>
<tr>
<td>On-level/off-track</td>
<td>(z = -0.378)</td>
<td>(z = -0.556)</td>
<td>(z = -1.633)</td>
</tr>
<tr>
<td></td>
<td>(p = 0.705)</td>
<td>(p = 0.579)</td>
<td>(p = 0.102)</td>
</tr>
<tr>
<td>Off-level/on-track</td>
<td>(z = -1.000)</td>
<td>(z = -1.016)</td>
<td>(z = -1.173)</td>
</tr>
<tr>
<td></td>
<td>(p = 0.317)</td>
<td>(p = 0.309)</td>
<td>(p = 0.241)</td>
</tr>
<tr>
<td>Off-level/off-track</td>
<td>(z = -1.000)</td>
<td>(z = -1.885)</td>
<td>(z = -2.362)</td>
</tr>
<tr>
<td></td>
<td>(p = 0.317)</td>
<td>(p = 0.059)</td>
<td>(p = 0.018)</td>
</tr>
<tr>
<td></td>
<td>(z = -0.136)</td>
<td>(z = -3.430)</td>
<td>(z = -3.084)</td>
</tr>
<tr>
<td></td>
<td>(p = 0.892)</td>
<td>(p &lt; 0.001)</td>
<td>(p = 0.002)</td>
</tr>
</tbody>
</table>

**○** not significant; **★** marginally significant; **★★** significant.

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 the 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 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, 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 (Rasmussen, 1983). Ideally, the interface should support pilots at all levels of cognitive behavior, while not forcing them to control at a higher level than necessary (Vicente & Rasmussen, 1992).
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 (Borst et al., 2010; Ellerbroek et al., 2011). 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.

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.

Acknowledgements

This work has been co-financed by the European Organization for the Safety of Air Navigation (EUROCONTROL) under its research grant scheme of the innovative studies programme. Additional funding for this project is provided by the National Aerospace Laboratory (NLR), The Netherlands. The content of the work does not necessarily reflect the official position of EUROCONTROL nor NLR on the matter.

References


