Effects of aircraft trajectories geometrical features upon air traffic controllers' conflict judgments

Philippe Averty, Pascal Lezaud, Florence Nicol

EFFECTS OF AIRCRAFT TRAJECTORIES GEOMETRICAL FEATURES UPON AIR TRAFFIC CONTROLLERS' CONFLICT JUDGMENTS

Philippe Averty and Pascal Lezaud, DSNA/DTI/R&D, Toulouse, France
Florence Nicol, Ecole Nationale de l'Aviation Civile, Toulouse, France

Abstract

This work is a twofold contribution to the analysis of conflict detection process in Air Traffic Controllers (ATCos). The first one addresses methodological aspects and proposes a way to get responses as close as possible to controllers' actual expertise without using artifacts such as rating scales or inferring judgments from verbal material. The second objective is to compare the influence of three geometrical features of aircraft encounters and their capacity to alter an accurate perception of conflicts. The proposed methodology appeared to be useful for collecting expertise as controllers quickly appropriated it, and led to get coherent data. Its use can be envisaged when a reliable representation of mental picture of ATCos is essential. Concerning the geometrical features of aircraft trajectories, aircraft attitudes i.e., the fact they are stable, climbing of descending, entailed significant differences on detection accuracy. To a lesser extent, catch-ups and segmented trajectories showed a capacity to make an accurate perception of conflicts more difficult. These results must be interpreted as tendencies more than precise or quantified results. As the objective of this experiment was to be a pre-experiment in preparation for future collecting in the framework of the European project SESAR, a few different choices concerning the trajectories to be used in the traffic scenarios will help to precise these results.

Introduction

Recent R&D developments in Air Traffic Management (ATM) already investigated some of the possible heuristic paths for simultaneously conciliating gain in capacity, enhanced safety and better services provision to users [1,2]. Among the enablers, the accurate trajectory prediction expected from the future FMS (Flight Management System) 4D stands in good place [3]. Air and ground systems could inform each other of current and anticipated flight parameters and actions. But initiation of such a system will need a substantial period. Thus, a transitional phase will exist, during which man and more automation will have to share tasks in real time. Both keeping the first in the decision loop will still require – even more acutely – to pay off the problem of compatibility between human expertise and automated data processing while meeting all the necessary operational and safety requirements. Inasmuch as the gap between these two parties could have been reduced it would both keep a sufficient relevance of ensuing actions and reduce human workload, therefore founding conditions for capacity gain, just as the European project ERASMUS aimed at illustrating it [4]. In particular, this would be the case for conflict risk assessment. Barring the appearance of new revolutionary equipments – which will precisely be the privilege of the future system – judgments about aircraft encounters are the headstone in ATCos' operating modes. Actually, though this process is only responsible of a part of the workload, the whole task (situation awareness, plans, and actions) is dynamically and permanently (re-) structured in reference to such judgments and their revisions [5].

Extrapolation of Aircraft Positions

Regarding the compatibility problem, the major issue is that judgments are not the exact reflections of computed probable risks [6]. In ATC, human processing mainly brings into play visual abilities, heuristics, and conversely does a very moderate use of computing [7,8]. In consequence, conflict judgments reflect the performance (and biases) of the visual system. If the latter is proved to be quick, holistic and resistant to disruption, it also has its own limits concerning extrapolation of current situations [9,10]. Moreover, in ATC judgments integrate additional margins, in order to cope with data uncertainty and contingencies [11]. These margins are related to idiosyncrasy and differences in individual experiences which induce a great variability among expert judgments. Nevertheless, a common and coherent cognitive processing caused by the major traffic configuration features has also been shown, resulting in globally meaningful judgments [12]. Using the three main sources of variation in conflict judgment (minimum horizontal and vertical separation, time to conflict) the cited work developed a predictive model of conflict judgments from a experimental data collection (161 participants) in four ATC centers. A mathematical modeling was worked out from these results [13].

It must be noticed that currently available data for extrapolating positions by computation – and separations between aircraft consequently – lead to results at least as uncertain as the mere empirical ATCos' expertise. Actually, uncertainty margins commonly used for this computation are in the region of 5% for horizontal speed and 15% for vertical speed. As far as the extrapolation exceeds a few minutes, these margins bring about detection of an excessive number of conflicts compared to the one coming from empirical (perceptual) judgments from experts [14]. The use of VAFORIT (a medium term conflict detection device) in Germany confirms the previous considerations. This tool is at present used to simply make out a list of possible conflicts. From then on, controllers may check this list and possibly use it as a simple aid, although (or because) they clearly know at the onset that this list may be not complete and contains a significant number of 'false' conflicts i.e., encounters that themselves will not see as such. Ratifying the limits of this device – without renouncing to progressively make it more effective – makes it accepted on working positions, as this tool is said to currently do controllers some appreciable turns.

Assessing Controllers Assessments

How could it be possible to assess controllers' thresholds between conflict and non conflicts? For long, experimental results as well as expert judgments gathering established that longitudinal distances around 7 to 15 Nm i.e., two or three times the norm represented the rough magnitude order of this threshold. Concerning the vertical dimension, margins differentiating conflict/non conflict judgments are more uneasy to assess. It was reputed to be more uncertain, generally needing more important margins than longitudinal distance (about 4000 to 6000 feet i.e., four to six times the norm) [15,16].

It would be interesting to explain a significant part of the variance about these thresholds. If idiosyncratic aspects seem to stay out of reach, the geometrical features in conflict configurations should be grasped when taking the relevant geometrical features into account. Identification of the geometrical characteristics in trajectories that take part in conflict detection processing has been made in several previous studies [17-23]. Various cognitive strategies are likely to be used by controllers but their connections with traffic configurations to which they are applied failed to be clearly established [24,25]. Among others, a review of that question exists in [13].

Furthermore, the use of rating scales induces variability in responses insofar as participants do not appropriate these scales exactly in the same way. Ideally, a substantial training, using standardized (or standardizable afterwards) configurations, should be previously proposed to subjects in order to get properly calibrated judgments. Due to time constraints, it is generally not the case.

Another source of variation lies in assessment collecting when using a mere binary choice (conflict ↔ non-conflict) for ruling on a converging pair of aircraft. Then, it is not only the judgment about the actually perceived risk which is collected. It is more the behavior the controller is willing to adopt concerning the risk he/she perceives which is expressed in his/her response. This systematically leads to inflate the number of conflict judgments at the expense of non-conflict ones. Intuitively, rather than: "I can see these aircraft will probably (resp. will probably not) show a separation loss in the future", binary judgments from controllers mean: "I decide to deem the future separation to be insufficient, even if I can see some marginally sufficient separation value". This phenomenon has been attested for long among ATC experts [26] and appeals for trying new paths for collecting expertise and appeals for trying new paths for collecting expertise.

Methods and Experimentation

Variables

As soon as the influence of geometrical features upon conflict detection (and resolution) is in question, the two relevant ones most often cited in the literature are the intersection angle between the involved trajectories (we shall name it *geometry*) and the flight phase: climb, descent, level (hereafter named *attitude*). These are two independent (explanatory) variables we kept in our experimental plan.

Velocity vectors constitute a significant aid for directly extrapolating future 2-D positions on the radar screen and in real time. Even if these extrapolations do not take into account neither changes that will occur within the considered time interval (engines paramatering changes by crews, wind direction changes) nor curvature of trajectories, this system helps controllers to easily get a more accurate base idea of ensuing positions of aircraft. Consequently, they currently integrate its use into their current operating modes, as soon as it becomes available on their working position. Hence we added in our experimental plan a third variable we named *segmentation*, in order to explain variance coming from differences between straight trajectories (where relatively reliable extrapolations are provided at once by the device) and trajectories being composed of different legs (where additional perceptual inferences are needed to get the required future positions).

Obviously, other sources of variation exist. Empirical knowledge of controllers on airlines (engine parametering instructions) or on aircraft type behaviors, sometimes exist. Inferences from the verbal material with crew can also give information to controllers about the way the next portion of the trajectory will be managed, which may help to extrapolate more accurately future positions. But we assumed all these can be considered as minor, when compared to the variance magnitudes coming from either the three parameters already selected or the detection process itself (intra- and inter-individuals difference in processing) [13].

Hypotheses

The main hypothesis concerning these variables is that they make the difficulty of an accurate extrapolation vary, and consequently make more uncertain the identification of conflicts. In our experiment, this hypothesis comes in the three following propositions:

- Segmentation: as projection (mentally, or by use of the velocity vectors available on the radar display) of future positions is more difficult when the aircraft trajectory is not straight, it was assumed that conflicts should be easier to make when using one-leg trajectories than when using ones with 2 legs (or more).
- Geometry: as they develop on the same longitudinal axis, catch-ups and face-to-face encounters should be easier to use for detecting conflicts than orthogonal trajectories. It is worth noting that we are not concerned with the 'resolution' aspect of the encounters. Face-to-face contingencies are commonly considered among controllers as more risky than other configurations, for a given minimum separation distance. This will not be taken into account here as only the perception of norm was in the balance, but not the necessity to implement or not an action.
- Attitude: vertical speed meets more important variations than horizontal speed. Therefore, encounters only involving level-stable aircraft should be easier to recognize than those involving one aircraft climbing or descending, and yet more than those involving two evaluating aircraft.

Making Up the Traffic Scenarios

Our first concern was to supply participants with air traffic scenarios as close as possible to real situations. Using simulated traffic generally implies that horizontal and vertical speeds are kept constant all over the sequence, which is unsound (aeronautically speaking) and is also perceived by subjects, who then make artificially accurate decisions. The second concern was to fulfill the requirements for getting judgments as close as possible to the mental pictures relative to the configuration shown. Giving responses on a numerical scale introduces bias as scale values – in spite of sufficient instructions – are not always used in the same way (inter-individual differences) for expressing judgments.

In order to deal at best with these two concerns, we fulfilled the following requirements while making the traffic scenarios:

- Scenarios consisted of a traffic sequences composed of two converging aircraft. Clusters of more than two aircraft involve more complicated processing and would be irrelevant in the context of this study.

- Aircraft trajectories were dynamically displayed, on the usual local radar map. Scenarios lasted about 40 seconds, as this duration appeared sufficient in a previous study [12] for controllers to make their decisions about conflict risk when only monitoring two aircraft.
- Aircraft trajectories had to be real, i.e., coming from traffic recordings in the considered control center. This means that horizontal and vertical speeds included variations that can be observed in real life. Both aircraft flew sometime in the presented paths but not concurrently: we managed to make them fly in the same space at the same time. They were selected because they had been flying on standard routes (known by all participants) which intersected in accordance with our explanatory variables. Then, a second sorting was carried out upon the targeted altitude, in order to build realistic conflict situations. This selection of valid trajectories could be made possible while using a specific device [27] which processed a two months traffic data base.
- No use of radio data during the test: aircraft were said to be flying on well-known standard or direct routes. A paper strip including the main ATC data (callsign, aircraft type, route, departure/destination, cleared level) was provided for each of the two aircraft.
- Distance span was to be set to 15 to 20 Nm (representing approx. 4 to 5 minutes flight). Distance span is the distance between the position of aircraft when the scenario stopped moving i.e., when the participant had to provide his/her response, and its position at the closest point of approach (approx. the intersection point). The similar value that was chosen (15/20 Nm) comes from a previous study [13]. It globally corresponds to situations where controller's judgments are asked for, as the available delays remaining for making decision are calling for them.

Two scenarios (rectilinear and segmented/stable-stable/face-to-face) confronted us to a problem since conflict conditions were always verified whatever the new 2-D location participants could give. In order to keep a complete combinatory from our variables, we decided to ask them, for these two scenarios, to make the null separation point at a given fix, located on the route of both. This amounted to agree on the specific point at which they had to be for being superposed. If this action did not make sense from the operational ATC viewpoint, it was a task equivalent to those of other scenarios, perceptively speaking, and we made previously sure this instruction was understood by controllers the way it should.

So, 24 exact conflict configurations i.e., showing a quasi-null separation were initially built while combining the two appropriate trajectories in the same scenario file. At this stage, scenarios represented successes to the test. So it must not be presented as such to participants, and a proper version of scenarios had to be set. Hence, one of the two aircraft was moved along its trajectory so as to present a potential separation of 8 to 12 Nm with the other one, instead of the null one in the initial version. Then, this file was restricted to a 40 seconds sequence in such a way that the end of it corresponded to the distance span defined above.

For providing their judgments, participants would only have to directly interact with the mouse on the radar screen with a mere "drag-and-drop" of one of the two aircraft, thus named *target-aircraft*, as the other one was the *reference-aircraft*. So the difference between target and reference-aircraft laid in the fact that only the first could be moved by subjects during the test while none of other aircraft's parameters could be modified. This interaction consisted in a translation of the target-aircraft along the intended route i.e., from its last position at the end of the sequence to the position where it should be at this moment (in participant's mind) for getting further a final null separation with the reference-aircraft. Neither horizontal nor vertical speeds could be modified for none of the aircraft. Only the 2-D location and altitude of the target-aircraft – when it was in an evolution phase – had to be modified by subjects. When the aircraft was climbing or descending, controllers had to use the mouse wheel to scroll to altitude value they extrapolated at this new position. Participants had all the (reasonable) length of time to provide their judgments. Actually, controllers were set in a situation where they had to create the most risky, or feared, configuration which could result from modifying the geometry of the encounter. We assumed this was intuitive enough to provide immediate access to the actual expertise in ATC, where perceptual processing of radar (and audio) data mainly impacts on separation assessments.

Quantification of Error Margins

Successes and failures would give global insight into conflicts identification. But interesting complementary information could also be found when differentiating between failures i.e., looking at their more or less important proximity to the 'good responses'. Actually, meaningful difference exists between those which would be close to the required positions and those which would show large separation distances.

Being twofold – horizontal and vertical minima – the conditions for setting conflicts pose problems for quantifying observed error margins in failures. First, a simple quantification (one-dimension) of conflict configuration proximity did not exist at the onset. Second, some scenarios could not be distinguished either in the vertical dimension e.g., the stable/stable attitude modality, or in the horizontal one e.g., the face-to-face geometry modality. This led us to design a variable for characterizing distances of each traffic configuration proposed by participants to the corresponding valid conflict one. This was made possible through the translation that had to be done to the trajectory of the target-aircraft, in the configuration provided by the participant. Actually, making afterwards the distance less than 3 Nm from the latter was possible when moving (forward or backward) the whole trajectory. The magnitude of that move could be expressed in Nm, this distance then representing the gap between response of participant and valid i.e., conflict, corresponding configuration. In cases where aircraft were also climbing or descending, their altitude generally did not show less than 1000 feet difference with the reference-aircraft. In these cases, an extra distance was added to the previous one corresponding to the distance flown while target-aircraft crossed the altitude in excess. We named this variable translate-distance (unit = Nm). Its purpose was to consistently arrange in order the different proximities to valid conflict situation for each 'failure' in that task. An example of this procedure is given in Appendix I. It was applied to all of the failure cases from the experiment.

Experimentation

The experiment took place in Lyons, France. After being informed of the existence and content of the experiment, full-performance level voluntary controllers were invited to pass the "conflict making" test. A total of 18 participants (14 males and 4 females) completed the test. Each of them had the 24 traffic scenarios, representing the combinatory of the 3 variables. The latter were:

- Segmentation. Modalities will respectively be named hereafter: rectilinear or straight, segmented or 2-legs. The angle between the two legs in the latter was between 70° and 130°)

- Geometry. Modalities are: face-to-face (trajectories intersection angle less than 160°), crossing (between 70° and 110°), catch-up (less than 20°).

- Attitude. Modalities are: stable/stable or S/S, stable/evolution or S/E, identical evolution or IE, opposite evolution or OE). 'Evolution' indiscriminately refers to a climbing or descending aircraft. For example, in 'IE' condition, both aircraft must be either climbing or descending.

The global objective of this study was to arrange these three variables in order from the perspective of the magnitude of their influence on conflict detection accuracy by ATCos.

Between 30 and 45 minutes (depending on subjects) were necessary to complete a 24-scenarios set.

Results

Global Success Rates

Comparing successes and failures, contingency tables account for the effect of each of the three variables upon controllers' responses. Success means that controllers moved the target-aircraft so that the future minimal distance would be less than 3 Nm and 1000 feet simultaneously.

A slight difference appears concerning the global performance to the test, showing that participants succeeded in finding conflict conditions in slightly more than half the cases. Now, one may ask how each of the chosen variables divided successes and failures (see Table 1).

Table 1. Global Results

Conflicts	Success	Failure	Total
Number	230	202	432
Ratio	0.53	0.47	1

A difference between the two modalities exists in our results. In accordance with the hypothesis we suggested, rectilinear trajectories allow a better identification of conflict configuration than segmented ones (see Table 2).

Table 2. Effects of Segmentation

Conflicts	Success	Failure
Straight	0.60	0.40
2-legs	0.47	0.53

Only minor differences exist between the three geometry modalities (Table 3). Face-to-face encounters are slightly easier to identify whereas crossings are the more difficult ones. The differences between the three values indicate only a tendency to validate the hypothesis on this variable.

Table 3. Effects of Geometry

Conflicts	Success	Failure
Face-to-face	0.56	0.44
Crossing	0.50	0.50
Catch-up	0.53	0.47

A significant effect is found for the stable/stable attitude (Table 4). This modality allows a really high rate of valid conflict recognition (81%). When referring to our hypothesis on that point (p. 5), stable/evolution configurations failed to be easier to combine than those implicating two evolving aircraft.

Table 4. Effects of Attitude

Conflicts	Success	Failure
Stable/stable	0.81	0.19
Stable/evolution	0.45	0.55
Similar evolution	0.41	0.59
Opposite evolution	0.45	0.55

Estimation of Error Margins

In the tables above, failures are not differentiated according to their nearness to a valid conflict configuration. Examination of translate-distances in failure cases provided some supplementary information. Although this variable had a numerical unit, we used median rather than arithmetic mean as central value. In doing so, results were less sensitive to extreme values in each traffic scenario. Actually, even if the interaction used for giving responses was greatly intuitive to controllers, only a few trials could be run before the test for each participant. Therefore, existence of some excessive error values i.e., deviating from his/her real mental picture cannot be dismissed.

Translate-distances are represented by means of *boxplots* (cf. Appendix II). These diagrams give an idea of their repartition and show the aberrant values ("outliers"). The boxplots in Figure 1 are made from the 202 translate-distance values (corresponding to the failures in Table 1) and how they are spread over. Translate-distance values (in Nm) are on the abscissa. For each of the three explanatory variables, the different modalities used in the experiment are on the Y-axis.

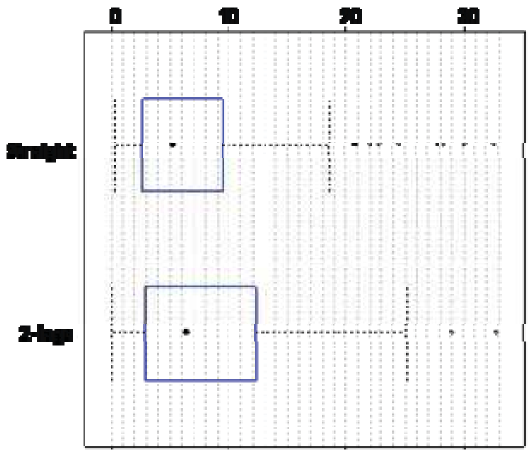

Figure 1. Translate-Distance Values According to 'Segmentation'

Comparing the respective values of $Q_{0.75}$ and Max for the two modalities, translate-distances appear to be larger when using segmented trajectories (Figure 1). Nevertheless, the median values are not so different: about 5 Nm for rectilinear trajectories, about 6 for segmented.

Translate-distances are slightly lower when using opposite trajectories than using any of the two other modalities (Figure 2). Catch-ups can clearly lead to clearly larger translate-distances. Median values are approximately 5 Nm for face-to-face encounters, and about 7 Nm for both catch-ups and orthogonal intersections.

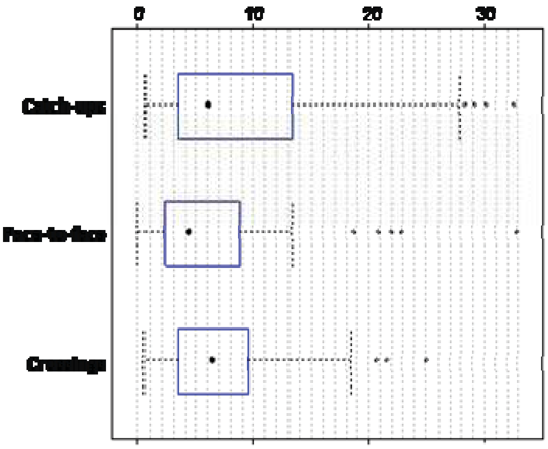

Figure 2. Translate-Distance Values According to 'Geometry'

The strongest effect – all of the three variables taken into account – is due to a stable/stable attitude of the two aircraft (median translate-distances value = 1.8 Nm). On the contrary, conflicts involving both climbing and descending aircraft appear to be the most difficult to identify, leading to possibly important error values. Aircraft configurations with stable/evolving or identical attitudes seem to lead to quite similar translate-distances values (Figure 3).

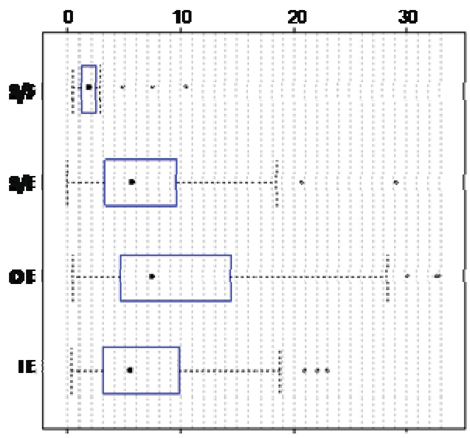

Figure 3. Translate-Distance Values According to 'Attitude'

Discussion

As translate-distances were computable only for 'failure' scenario cases, it appeared difficult to simply confront and discuss the information stemming from success rates and translate-distances. An attempt for such a synthesis stands in Figure 4, both in terms of success rates and translate-distances, for two of the independent variables. Adding the third one would have made the interpretation of the graphics too uneasy. Each of the 24 scenarios presented to the participants is coded both through shapes (representing geometry modalities) and colors (aircraft attitude modalities). This figure evidences the scatter-plot (marked by the dotted envelope) between the global success rate and the mean translate-distance, all participants taken into account. This can be interpreted quite logically as follows: the conflict configurations that have a high success can be identified rather easily. Consequently, one may expect that in the few cases where valid response failed to be given relatively low translate-distances values must be found.

The wide range of catch-ups (circles) appears, as the eight of them are remarkably spread out from the best score to the worse score. As a result, this specific category of trajectories geometry can be held as a significant source of variation in conflict judgments.

On the top left of the graphics, the 6 stable/stable encounters confirm the high rates and low translate-distances they admitted to conflict recognition. This globally means that all conflicts involving aircraft stable in altitude are identified with a more or less comparable accuracy, whatever the two others trajectory characteristics (segmentation, attitude) are. Looking at the slight effect these two characteristics have (considering only the eight green elements) it can be noticed that crossings (squares) admit more successes and lower average translate-distances than face-to-face (stars). Although this result only applies to 8 items, it is noticeable that it refutes the global results shown Table 3 and Figure 3, where face-to-face encounters appeared to be easier than crossings ones. Our interpretation is that performance in face-to-face traffic scenarios was overvalued in the global results. There are two main reasons for that, which are developed in the two following paragraphs.

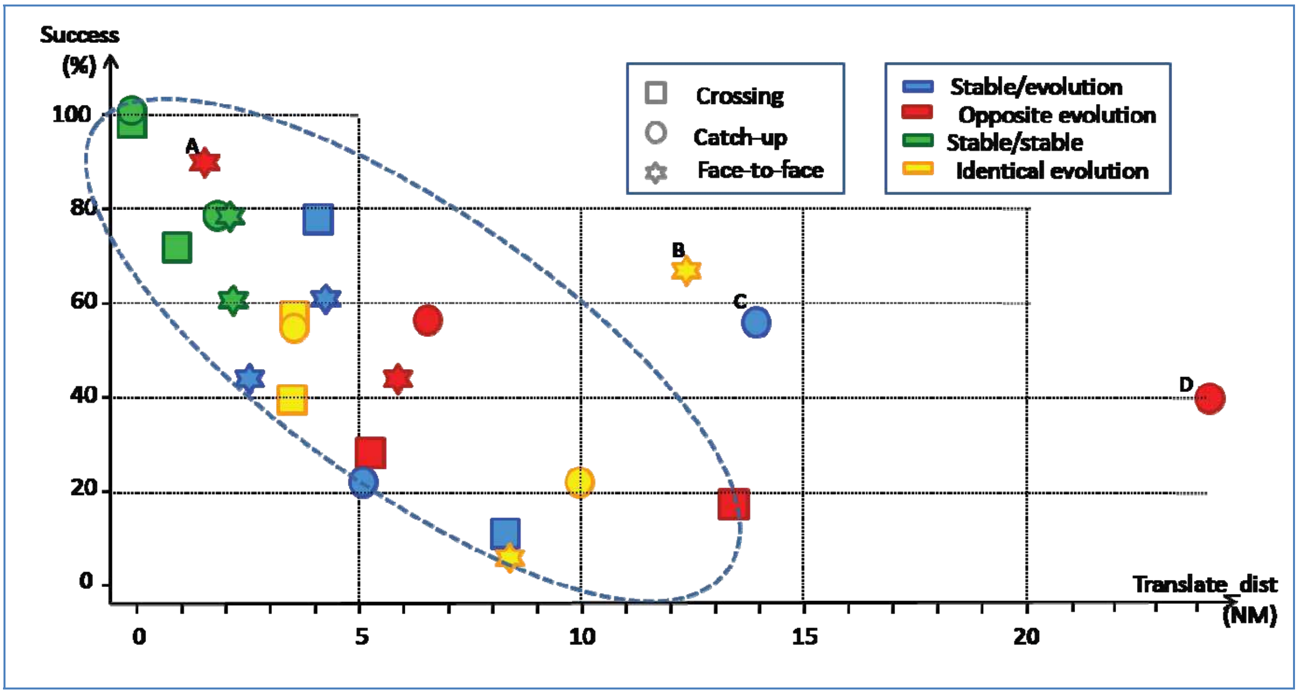

Figure 4. Scatter-Plot of Success Rates and Translate-Distances, for the 24 Types of Encounters

The relatively poor performance related to scenarios integrating 'opposite evolution' attitudes (red color) appears quite clearly: most of them are located in the low performance area in the graphics both for success rates and translate-distances, aside from a unique scenario. Unlike the others, the latter (marked 'A') is situated on the top left of the graphics, meaning a high easiness for controllers to identify conflict from it. It corresponds to a combination *segmented/face-to-face/opposite-evolution* of the three explanatory variables. When asked on that point, controllers did cast doubt on such an easiness to extrapolate accurate separations between aircraft showing that particular combination. Furthermore, examination of this scenario showed that the presented locations of the two aircraft i.e., before participants had to interact with the target-aircraft and moved it to a suitable place, already put them in a quasi-situation of conflict. A mistake in the initial review of the experimental material is probably the cause of selection of an invalid version of the scenario. As a result, finding valid conflict conditions was much easier for this scenario than for others. Both for ethical reasons and to keep a complete combination of the different modalities in explanatory variables, we nevertheless decided to keep this scenario in the global results. The use of medians (instead of means) minimizes the effects of this mistake and keeps results globally relevant.

Another point is that three types of encounters (on the right of the graphics, marked 'B', 'C', 'D') stand aloof from the dotted envelope, thus showing singular high success rates compared to their respective translate-distances values on abscissa. Examination of each of these three scenarios helped to find the most probable reasons for this discrepancy, inviting to some change to introduce for improving the experimental paradigm we used:

- Combinatory of our variables included a scenario that is generally never processed in the same way than others, in real life. This scenario (marked 'B') represents a *rectilinear/face-to-face/identical evolution* encounter which, in real life, always appeals for an early and protective solution in terms of altitude but not for any accurate extrapolation at this stage. As a conflict identification is quite precise (less than 3 Nm and 1000 feet), our interpretation is that participants divided in two groups for processing this particular scenario. A first group did try to find the means to do it, resorting to computing from available or inferable data. A significant part of them probably found the corresponding configuration, making the major part in the 67% success rate. But the other part of participants probably processed this scenario from their current operating modes (perceptual heuristics), ending in large error values since such a precise assessment is seldom accurately used for this configuration. This second bias adds to the previous one for favoring in excess face-to-face encounters (Table 3), as we already stated.

- The two others singular scenarios (marked 'C' and 'D') show both decent global success rates and the highest translate-distance values in the experiment. The cause seems to be the same for both, and appears to the involvement descending traffic, on initial then intermediate approach. During this last phase, aircraft speed greatly decreased – we used real traffic recordings. From then on, the freedom participants had for moving the target-aircraft at any (future) position of their own choosing appeared to be possibly excessive. When extrapolating the point of separation loss too far on the trajectories, the speed of descending aircraft varied so much than it did not only failed to find conflict conditions but also generated large translate-distance values. So, our interpretation for these two scenarios is that the global success rate is the reflection of expertise while the translate-distance, in case of failure, must be seen as artificially oversized, compared to the others scenarios.

Last, even if the global effect of the rectilinear modality effect appears to be higher than the segmented one (Table 2), the more detailed effects i.e., upon the other variables, seems manifestly variable: significant differences sometimes point out segmented scenarios as substantially easier than their rectilinear homologues. This led us to wonder about our choice to use real recorded trajectories in the traffic scenarios. It is quite clear that great speed (horizontal and/or vertical) differences were present in some scenarios. This heterogeneousness could generate differences in identifying conflicts, preventing to find regular effects from the three independent variables. Similarly, the time spans between the end of scenarios (when participants were asked to respond) and the moment when aircraft could be in conflict probably lacked consistency (average: 4.5 min; minimum: 3.2 min; maximum: 6 min). This disparity was the price to pay for showing totally realistic configurations, and our results benefited from them. But on the other hand, it seems now that instead of real traffic, the use of aircraft flying by means of (computed but realistic) tabulated models of aircraft types would be more effective, allowing a higher consistency of experimental material without losing too much realism.

In summary, results show that:

- For geometry, differences exist in conflict identification according as the involved aircraft are level or include at least one evolving aircraft. When the latter have opposite evolutions (climbing, descending) difficulty for conflict identification by controllers seems to be maximal. Consequently, only three different geometry categories will have to be kept for differentiating the effects on controllers' judgments: S/S ; (S/E, IE) ; OE.
- For segmentation, a significant effect tends to appear, as rectilinear trajectories are easier to identify. Consequently, the use of two modalities in this variable remains relevant: rectilinear; segmented. The use of traffic flying with more homogeneous speeds in a further experiment should lead to more reliable quantification of this effect.
- For geometry, our results failed to show any clear tendency about the effects of this variable. Though face-to-face initially tended to appear more helpful to identify conflicts, this must be counterbalanced since the existence of an invalid scenario favoring this result has been demonstrated.
- The methodology used for collecting controllers' expertise show interesting properties: participants quickly assimilated the requisite interactions and reported a rather pleasant and playful way of

providing their expertise. Yet substantial change will do better to be proceeded to, concerning some options initially retained: speeds coming from tabulated performance models (related to aircraft types), selection of homogeneous phases in flights, wide but limited range for aircraft positioning by participants.

We aimed at developing an original way of collecting ATCos' expertise, and investigate the influence of geometrical features of aircraft trajectories in conflict judgments. Coherent results have been drawn from it, even if this first attempt is perfectible. Thus it could be expected that a larger scale collection using en-route traffic and integrating the improvements we mentioned above would provide more accurate data. These would allow the modeling of these judgments which would become greatly predictable, in a way similar to what we already did [13].

References

[1] SESAR Consortium, Deliverable 3, The ATM Target Concept, www.sesarju.eu, September 2007

[2] Weber, R., Garcia JL., Gawinowski G., Brochard M., Carotenuto S., 2007, *ERASMUS: Concept of Operations & Initial Modeling Results*, 7th R&D ATM Seminar, Barcelona, 2-5 July.

[3] R. Weber, P. Cásek and J. Kubalčík, 2007, Analysis and improvement of ground and FMS-generated trajectory predictions accuracy and integrity, 26th Digital Avionics Systems Conference.

[4] Drogoul, F., Averty, P., Weber, R., 2009, ERASMUS Strategic Deconfliction to Benefit SESAR, *8th Air Traffic Management R&D Seminar*, Napa (CA).

[5] Hilburn, B., 2004, *Cognitive Complexity In Air Traffic Control – A Literature Review*, Note No. 04/04, Eurocontrol Experimental Centre, Brétigny, France.

[6] Kurchar, J.K., Yang. L.C., 2000, A Review of Conflict Detection and Resolution Modeling Methods, *IEEE Transactions on Intelligent Transportation Systems*, Vol.1, No. 4, pp.179-189.

[7] Bouju F, Sperandio JC, 1979, *Analyse de l'activité visuelle des contrôleurs d'approche*, Note INRIA/CENA C0 7911 R59.

[8] Bisseret, A., Girard Y., 1973, *Une expérience sur le "choix des solutions" dans le contrôle du trafic aérien*, Rapport IRIA C.O.7301 R35.

[9] Law, D. J, Pellegrino, J. W., Mitchell, S. R., Fischer, S. C., McDonald, T. P., and Hunt, E. B. 1993, "Perceptual and cognitive factors governing performance in comparative arrival-time judgments", *Journal of Experimental Psychology: Human Perception and Performance,* 19(6), 1183-1199.

[10] Remington, R. W., Johnston, J. C., Ruthruff, E., Romera, M., Gold, M., 2000, "Visual Search in Complex Displays: Factors Affecting Conflict Detection by Air Traffic Controllers", *Human Factors*, Vol. 42(3), pp 349-366.

[11] Leroux, M., 1999, "Cognitive Aspects and Automation", *Proceedings of the first USA/Europe Air Traffic Management R&D Seminar*, Orsay (France).

[12] Averty, P., 2005) "Conflict perception by ATCS admits doubt but not inconsistency", *Proceedings of the 6th Air Traffic Management R&D Seminar*, Baltimore, MD.

[13] Averty P., Guittet K., Lezaud P., 2008: An Ordered Logit Model of Air Traffic Controllers' Conflict Risk Judgment, *Air Traffic Control Quarterly* Volume 16, N°2, pp.101-125.

[14] Alliot J.M., Durand N., Granger G., 2001, "A statistical analysis of the influence of vertical and ground speed errors on conflict probe", *Proceedings of the 4th Air Traffic Management R&D Seminar*, Santa Fe, NM.

[15] Michard, A., 1976, *Analyse de la représentation des trajectoires d'évolution chez les Premiers Contrôleurs*, Rapport IRIA C.O.7611 R50.

[16] Lafon, M.T.,1978, *Observations en trafic réel de la résolution des conflits entre avions évolutifs*, Rapport IRIA C.O.7802 R55.

[17] Rantanen, E.M., Nunes, A., 2005, "Hierarchical Conflict Detection in Air Traffic Control", *International Journal of Aviation Psychology*, Vol. 15, No. 4, Pages 339-362

[18] Bouju, F., 1978, *Analyse de la prévision de la séparation entre deux avions évolutifs convergents*, Rapport IRIA C.O.7802 R54.

[19] Nunes, A., Kirlik A., 2005, "An Empirical Study of Calibration in Air Traffic Control Expert Judgment", *Proceedings of the 49th* *Human Factors and Ergonomics Society*, pp. 422-426(5).

[20] Enard, C.J. (1974), *Analyse de la prévision de séparation entre deux avions convergents sur l'écran radar,* Rapport IRIA C.O.7405 R44.

[21] Boag C., Neal A., Loft S., Halford G.S. (2006), "An analysis of relational complexity in air traffic control conflict detection task", *Ergonomics*, Vol. 49, No. 14, pp. 1508-1526.

[22] Davison, H.J.,. 2006, *Modeling the Air Traffic Controller's Cognitive Projection Process*, Report No ICAT-2006-1, Massachusetts Institute of Technology, Cambridge, MA.

[23] Nunes, A., Scholl, B.J., 2004, "Interactions between Convergence Angle, Traffic Load, and Altitude Distribution in Air Traffic Control", *Proceedings of the 3rd Annual Conference on Human Performance, Situation Awareness and Automation Technology*, Daytona Beach, USA

[24] Loft S., Sanderson P., Neal A., Mooij M. 2007, "Modeling and Predicting Mental Workload in En Route Air Traffic Control: Critical Review and Broader Implications", *Human Factors*, Vol. 49, No. 3, pp.376-399.

[25] Xu, X., Rantanen, E.M., 2003, "Conflict detection in air traffic control: a task analysis, a literature review, and a need for further research", *Proceedings of the 12th International Symposium on Aviation Psychology*, Dayton, USA.

[26] Bisseret, A., 1981, "Application of signal detection theory to decision making in supervisory control: the effect of the operator's experience", *Ergonomics*, 24, (2), pp. 81-94.

[27] Hurter, C., Tissoires, B., Conversy, S., 2009, FromDaDy: spreading data across views to support iterative exploration of aircraft trajectories. In IEEE Transactions on Visualization and Computer Graphics (Proceedings of IEEE InfoVis 2009).

Appendix I: Translate-Distance Definition

The mains steps for translate-distance value computation are intuitively given below by means of an example. A target-aircraft (CRL1584, descending) and a reference-aircraft (FBRSH, level) with their velocity vectors and flying on orthogonal trajectories are shown on the three next figures. Once participant had moved the target-aircraft, new minimal separation values (here 6.1 Nm longitudinally, on Figure 5) potentially existed between the two aircraft.

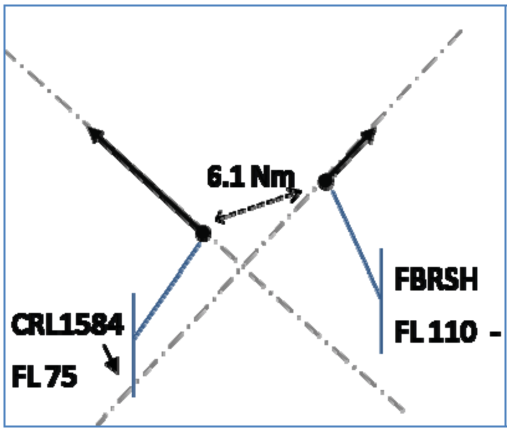

Figure 5. Checking the Conflict Conditions

In each failure case i.e., when the created separation did not fulfill the conflict conditions *translate-distance* was computed. This variable corresponds to an empirical expertise regarding the experimental task. Furthermore, it does not use the computation of the closest point of approach (CPA) as a rigorous approach would have required it. This could not have been done due to time constraints, but the uncertainty of the proposed method (hereafter) is largely compatible i.e., low compared to the magnitude of controllers' errors i.e., translate-distances, that was computed. Moreover, referring to the intersection point is probably not so inexact as in their judgments, controllers are likely to perceptually refer more to the intersection point of 'visible' trajectories than to the (computed) CPA.

So, in case of failure the next step was to identify the altitude of the referent-aircraft when it was at the place just sufficient for creating conflict, i.e., 2.9 Nm to or from the intersection point.

In the example (Figure 6), the reference-aircraft is from the intersection point, since it is the first to overfly it. The corresponding altitude is 11000 feet since this aircraft is level – but it could be otherwise, when taking into account the altitude at this location of a climbing or descending aircraft.

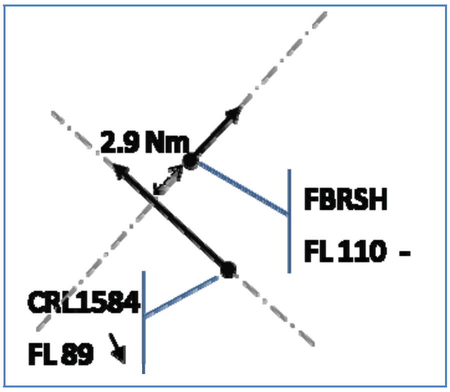

Figure 6. First Aircraft Altitude Computation

Once this altitude was known, the required altitudes of target-aircraft for creating conflict conditions were deductible when adding/subtracting 900 feet (conventionally representing the maximum margin to get conflict conditions). Here in our example, the altitudes between 10100 and 11900 feet satisfy this condition.

Within this range, the altitude value to be chosen was the one allowing the minimization of participant's error. Here this value is 10100 feet, as it corresponds to the location of target-aircraft which is the closest to the intersection point i.e., where it should have been in order to fulfill just sufficient conflict conditions (approximately 2.9 Nm and 900 feet separations). Then, the last step consisted in determining the location of the target-aircraft when he flew at the chosen altitude (Figure 7). The distance between this last location and the intersection point gave the *translate-distance* value, representing the magnitude of error in controller's judgment.

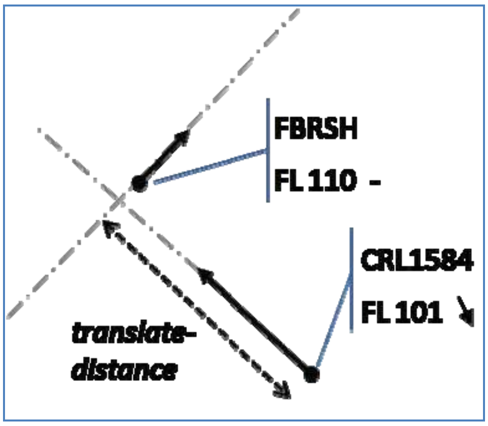

Figure 7. Translate-Distance Computation

Let us notice that the interactions ('drag-and-drop') participants had to do on target-aircraft corresponded to a translation in time of the whole considered trajectory. This translation in time is equivalent to one in distance, as soon as the horizontal speed is known – what was the case. What we did while computing translate-distance was the analogous procedure, additionally integrating vertical errors – which are also translatable in time and consequently in distance, once the vertical speed is known.

Appendix II: Boxplots

Boxplots stem from the repartition of numerical values (translate-distances in this study) along the whole range of the sample values. This repartition can be grasped through the positioning of three quantiles: $Q_{0.25}$, $Q_{0.5}$ and $Q_{0.75}$ (Figure 8).

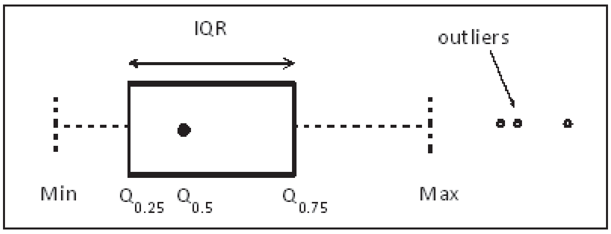

Figure 8. Description of a Boxplot

Basically, for a sample of values, 25% of these values are situated 'on the left' of $Q_{0.25}$, 25% are on the right of $Q_{0.75}$. The length of the rectangle is named Interquartile Range (IQR) and represents the remaining 50% of the sample. The median value corresponds to the $Q_{0.5}$ point. The outliers are the values situated to the right of $Q_{0.75} + 1.5 \times \mathrm{IQR}$, or to the left of $Q_{0.25} - 1.5 \times \mathrm{IQR}$. Last, the Min and Max values represent the bounds of the sample, once the outliers have been removed.