Abstract
Though there is considerable agreement among past studies about the great variability in conflict judgments by Air Traffic Control Specialists (ATCS), this paper proposes a first step towards the identification of a common core shared by controllers for perceiving conflict risk. This study began with the construction (from real traffic recordings) of traffic scenarios showing two converging aircraft. Three variables characterized these traffic scenarios, quantifying respectively horizontal separations, vertical separations, and the moment of formulation of the judgment (prediction time span). The conditions created by factorial manipulation of three variables led to the design of short scenarios (about 1 min) upon which 161 controllers gave their opinions about possible occurrence of a separation loss (less than 3 nautical miles and 1000 feet) between these aircraft. The data, currently being analyzed, already show an important variability among experts, confirming the results of the past studies: considering a given pair of aircraft, the perceived risk of insufficient separation appeared to vary greatly from one controller to another. But the way that median values of the responses are spread over the combinatorial range of the scenarios appears very coherent and prefigures a model of conflict perception by the ATCS community. Further, values needed to reach a certain level of security are lower when obtained on the horizontal plane rather through altitude differences, showing (and quantifying) the higher efficiency, or reliability of those obtained on the plane. It is also shown that uncertainty varies in the same direction than prediction time span (variable D°A), putting forward quantifications concerning effects of this variable. Moreover, particular variations of uncertainty appear within responses resulting from the scenarios with “near minimum” separation values. These variations tend to evidence the line of thought of ATCS dealing with the limits of the margin they have to set for a conflict/non-conflict classification.

Introduction
To define their job, controllers often say they «separate aircraft» because they detect and resolve conflicts resulting from their trajectories. A trajectory corresponds to the four-dimensional (4-D) path (latitude, longitude, altitude, and time) along which an aircraft flies. A conflict can be defined as a predicted loss of separation between two aircraft. In this study, the separation minima required were either 3 nautical miles (NM) for longitudinal spacing, or 1000 feet (ft) when separation occurs beforehand on the vertical plane. These are the values corresponding to Air Traffic Control Centres (Tracons) having a radar with a certain level of performance. Radar display mainly contains analogical information (aircraft position, sector features, direction of flight, motion of the aircraft) and text information (identification, speed, altitude).
A fundamental feature of air traffic control is that uncertainty vitiates the data required for making the decision. Geometrical and probabilistic approaches to conflict estimation will not be developed here [1]. Uncertainty is often distinguished from risk. In the latter case, decision outcomes are associated with probabilities, while in the former these are unknown. The characteristic of the term uncertainty includes a large range of different aspects: variability, ambiguity, missing data, etc. All of them come from various flight environment factors (wind shift, engine parameters set by pilots, etc) and from characteristics of the cognitive processes involved in real time data processing. Integrating uncertainty consequences in their own judgments, ATCSs empirically appropriate extra margins for predicting conflicts.
Continuous adoption of sophisticated technologies over the last decades has not resulted in any significant change. Conflict detection remains a key task in ATC, and expertise there comes from intensive training over a long period. The most salient aspects of the ATC job are the highly skilled, yet routine and stereotyped nature of most required actions. Air traffic control is an example of transformation from a conscious complex activity into a highly automated process. Consequently, as few cognitive resources - mainly attention - as necessary are consumed [2], and ATCSs are fiercely protective of their procedures and displays.
“perception of the 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” (Endsley), and is more generally known as one of the controller’s core skills [3]. Underlying detection processes, perception is an interpretation process, automatic, influenced by context and expectation, and nearly indistinguishable from actual sensation [2]. The visual evidence on which the controllers’ perception is based is frequently very brief (<200 msec). Numerous results converge to suggest that quick recognition and confidence in one’s own judgements are two highly valued characteristics in ATC. Consequently, cognitive complexity is strongly related to controllers’ perception and recognition of patterns [4]. Conflict judgment will depend on whether the aircraft will be found to reach the same altitude at the moment of the position overlap on the horizontal plane. This involves both computation (vertical dimension is set out through numerical values) and visual perception (the intersection point needs to be perceptively extrapolated), the latter being largely processed by means of patterns [5].

Past studies
In addition to the values defining separation and to the level of skill of subjects, several factors are known to take part systematically in acting upon conflict judgment: contextual traffic load, speed ratio, conflict geometry, aircraft attitude, and prediction time span. All of them showed strong and consistent effects in experiments [4] [6] [7]. Most of the empirical studies that addressed prediction in ATC were concerned with controllers’ performance on response time and conflict detection accuracy i.e., the success rate of detecting conflicts.

Extrapolating the intersection point (I.P.) of paths is a relatively easy task for straight routes, but can be more difficult for those being segmented by means of different successive route headings. Globally, distances are easier to compare, when perceptively comparing speed and projecting future positions on paths. This type of processing is even more likely to be used since the visualization of speed is now displayed on scope with user-definable vectors. These distances (or time-to-contact) are also affected by geometry of the paths being computed [3]. Prediction becomes then less accurate as the angle varies from 0° to 90°. However in ATC, heuristics and particularly visual projections are more likely to be used to compare distances to I.P. [5] [8]

Of course, vertical separation is not quantifiable directly from perception. In real work, ATCSs seem not to have recourse to this dimension as long as there is some evidence of lateral separation assessment. A margin of about 10 NM seems to be average for controllers to deliver a non-conflict judgment, while 3 or 5 times the norm (i.e., 1000 feet each time) are necessary to provide the same effect in a vertical plane resolution [5] [6]. Besides it has been noticed that very little need were found regarding rates of climb/descent, showing an inaccurate assessment of vertical positions [9]. These rates mainly seem to be used in cases of aircraft of identical attitude [10]. Aircraft attitude by itself has been treated mainly as a factor in the speed of response in diagnosis [10].

Results on prediction time span indicate that conflict/non-conflict classification is inexact until aircraft reach 30 NM (±10Nm) to I.P., depending on the parameters of other trajectories [9]. Furthermore, non-conflict judgments always appear to be formulated later than conflicts [9] [10]. An increase in convergence angles [3] [5] [10] generally makes judgment evolve in direction of inaccuracy: it reaches a maximum effect near 90°. But for some authors, this is not a major variation source as, for them, attitude seems to have stronger effects [6]. Hence, wider convergence angles are not systematically associated with impaired conflict detection, even with high traffic loads [11]. A last feature is that the conflict geometry of right angles generally requires that separation assessment is made on the horizontal plane [10].

Speed ratio has contradictory effects, according to authors [3] [12]. It can be used in ATC for reinforcing a judgment that has already begun to emerge [10].

The appearance of judgment doubt increases with professional experience [8]. Beginners make a more frequent use of computing, while experts often wait much longer before judging (and implementing an action).

In addition to the results detailed above, two other conflict detection features appeared clearly. Firstly, experience-related inferences are supposed to add to any perceptive observation for producing judgments. [6] [9]. Second, conflict predictions have a second dimension [5]: doubt (or confidence) seems to be inherently included in the judgment itself. So the effects of traffic configuration characteristics finally appear to be multiple, preventing a simple model from emerging. There is neither a global pattern nor systematic computing, except for unknown contexts: operating processes vary according to the problem and to the controller.

Method
Nevertheless, our hypothesis was that a certain level of agreement should exist among controllers regarding conflict judgment by itself, and should be observable particularly when considering the detection process. An experiment should be run, aiming at soliciting experts’ judgments on realistic
scenarios in an environment as similar as possible to the real one. The purpose was to assess the real margins used by ATCS and to compare them, trying to identify their common bases and to model them statistically.

**Designing the traffic scenarios**

The feasibility constraints of this experiment excluded using more than three independent variables (I.V.). Two of them contribute to the definition of the concept of conflict, while the third deals with the moment when conflict judgment is made.

- **E(d)** (for: expectation of minimum longitudinal distance). It stands for the lowest distance that exists between two aircraft in their whole trajectories. It is the distance (in NM) between them at the closest point of approach. This variable has five levels: 0, 3, 6, 9, and 12 NM.
- **E(fl)** (for: expectation of vertical separation). It represents the separation value obtained between the two aircraft because of the movement of one (at least) of them in the vertical plane. E(fl) is not concerned solely with the mere altitude differences of the aircraft but is rather a modeled variable. Hypothesizing to get closer to the actual mental picture of controllers, we defined it as the longitudinal distance (in NM) existing when the altitude difference between the two aircraft reaches 1000 feet (and increases). It has five levels: 0, 5, 10, 15, and 20 NM.
- **D°A** (for: prediction distance span, i.e., how far into the future the prediction is made). It gives the distance between the positions of the aircraft and the intersection point (I.P.) when subjects have to come to a decision about a future possible separation loss. The six levels of D°A are: 6, 10, 15, 21, 28 and 36 NM.

We shall call the “scenario” the traffic situation involving two aircraft, dynamically displayed on a radar screen for one minute and fifteen seconds. By definition, each scenario is equivalent (or corresponds) to a triplet of values for E(d), E(fl) and D°A.

As seen from past studies, other variation sources exist. We established most of the important ones in the scenario design procedure stated. Scenarios were issued by a three step procedure. First, two aircraft had to be selected from some recent real traffic recordings of the considered facility. They were chosen with regard to three features: their paths must be as straight as possible, secondly they must have followed routes the difference of which was as close as possible to right angles, and lastly they must show homogeneous speeds (ratio lower than 1.5). We shall call GPA (generic pair of aircraft) each pair having generated scenarios. Once they had been identified, each trajectory was "copied and pasted" in the same file, in order to be in the vicinity of one another, at the same time. Their paths could be then be made perfectly straight for 40 miles (approx.), and made to intersect at exactly 90°. Tools specifically developed at the CENA made these processes easier.

![Figure 1: Definition of the 3 Independent Variables (I.V.) : E(d), E(fl), D°A.](image)

*Figure 1: Definition of the 3 Independent Variables (I.V.) : E(d), E(fl), D°A.*

A pair of aircraft is shown at three different moments, corresponding to three different descending levels for the AFR123. As indicated, E(fl) is the expression on the plane of a vertical resolution. Moreover, D°A1 makes it easier to formulate the judgment than D°A2.
The second step (Intermediate Step) was to make this GPA available in every different separation value required by the combinatorial. Global translation of paths, either on the horizontal plane or in the vertical one, made it possible to have as many copies of this "conflict" at our disposal, each of them having separation values being set in accordance with one of the values of E(d) and E(fl). The result was about 17 variations based on the initial GPA, involving the same pair of aircraft, following the same path, but showing different separation values.

The last step (scenario making) simply consisted of a six pieces segmentation (of 1 min 15 sec) for each of the previously built “variations”: six scenarios were then made. The end of each scenario corresponded to the time when the fastest aircraft (conventionally) got to a certain distance (D°A) from the intersection point of the paths. The beginning of the scenario could then be computed as being 1’15”earlier. Next, the parts of the paths before the start time and after the end time were deleted. There only remain two short segments of the trajectories, representing about eight radar plot shifts. Hence, as in real situation, future separation had to be inferred from the standard information: flight plan elements, and radar position, altitude, and their changes through time. The fact that they were copies of real flights on usual routes might also support the processing of information for assessing their future behavior.

Remark: As radar plots detection is asynchronous, it can not be said that scenarios in which E(d) = 3 NM represent non conflict situations. Actually, in these scenarios E(d) oscillates between 2.5 and 3.5 NM (approx.) in the vicinity of the closest point of approach, depending on the cinematic characteristics of the situation and on which one of the aircraft moved last.

So, in each scenario, two aircraft flew on straight converging courses, maintaining their current headings. Their attitudes were a descent for one of them, the other being stable at its cruising level. One was maintaining its cruising level, while the altitude of the other one was evolving. The ground-speeds and rates of climb/descent were generally not constant, as they were copied and pasted from real traffic recordings. There was no contextual workload to be added to the scenarios: they were simply two aircraft settings. The purpose was to assess the perception of uncertainty in a way that allows the practical use of it.

In this work, we neither intended to address a probabilistic viewpoint of the conflict prediction nor to compare collected judgments to required minima as such. The purpose rather was to assess the real subjective distances experts can manage when judging conflict/non-conflict initial conditions.

**Experimental Plan**

Participants could not take the whole set of scenarios (about 5 * 75 for each facility, except Bordeaux). All of the scenarios of a given facility were then randomly distributed into three “packs”. Each pack could be split into two halves, in order to be able to propose the test to participants having only 30 or 40 minute free periods. In each facility, all its packs were given the same number of times.

<table>
<thead>
<tr>
<th>Facility</th>
<th>GPA</th>
<th>Nb of GPA completions</th>
<th>Speed ratio (aircraft1/aircraft2)</th>
<th>Aircraft attitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lyon</td>
<td>A</td>
<td>15</td>
<td>310 / 280</td>
<td>Descent</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>15</td>
<td>170 / 160</td>
<td>Descent</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>15</td>
<td>160 / 150</td>
<td>Climb</td>
</tr>
<tr>
<td>Toulouse</td>
<td>A</td>
<td>11</td>
<td>280 / 250</td>
<td>Descent</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>12</td>
<td>260 / 180</td>
<td>Descent</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>12</td>
<td>260 / 210</td>
<td>Descent</td>
</tr>
<tr>
<td>Marseille</td>
<td>A</td>
<td>15</td>
<td>250 / 200</td>
<td>Descent</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>15</td>
<td>250 / 190</td>
<td>Descent</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>15</td>
<td>240 / 200</td>
<td>Descent</td>
</tr>
<tr>
<td>Bordeaux</td>
<td>A</td>
<td>36</td>
<td>340 / 240</td>
<td>Descent</td>
</tr>
</tbody>
</table>

Table 1: Experimental Plan Elements.
As the number of scenarios was high, participants were only given a part (1/3) of the whole of them in each facility.

We could not define a fully-crossed within-subject design. E(fl) never could be inferior to E(d), since the latter represents the minimum distance between the two aircraft on their trajectories whatever their altitude could be. Relationships between D°A and both E(d) and E(fl) are also constrained since the first represents the distance of one of the aircraft to the I.P.: for low value levels of this variable (for example D°A= 6 NM), scenarios with high values of E(d) and E(fl) could not exist. It appeared that the number of scenarios could vary in a short interval (72 to 90), as a consequence of properties of 4D configuration of the traffic. This is due to interaction between geometrical properties and the rates of climb/descent of aircraft, or whether the fastest aircraft was the first over the I.P., for example.

As there was enough time between the first data collecting (Lyon) and the last (Bordeaux) we decided to test two features. First, the comparison between climb and descent attitude: would there be any significant difference in using both attitudes in scenarios indiscriminately? We compared the results of GPA “C” to the two others in Lyon before designing the scenarios of the next facilities. The result was marginally significant and we decided to keep only one of both attitudes (descent). The second consisted of intra individual variability measurement at Bordeaux. There, participants judged several times the same triplets of I.V., supported each time by different GPA, i.e.,
presented each time with a new pair of aircraft following new routes. In this paper, only the basic Bordeaux GPA has been computed (one response for each scenario) but of course not the other images of it.

**Apparatus and task description**

The devices used in the experiment were identical to the ones usually used by the participants (radar display, maps, speed vectors, radar adjustment interface, etc.). The display of speed-vectors was set to 3 minutes, and no vertical speed (Vz) indicator was initially displayed. The instructions for the latter were to resort as little as possible to it, in order to keep the common operational processes, instead of unusual and artificial ones.

After listening to the instructions, and being given three or four examples of the scenarios, participants were asked to judge the scenarios successively. The available time for answering was 1 minute and 15 seconds. This term had been set empirically by sampling the opinions of six controllers during a pre-experimental phase at the CENA for setting a sufficient time for judgment formulation. The overall time to complete each of the 161 individual sessions was comprised between 40 minutes and 1 hour 15 minutes.

When reading the instructions, it was particularly emphasized that what the participant’s eye could see was the only interesting thing here. A context was suggested to them for answering correctly: they were present at the qualification test of one of their (young) colleagues, having access to the radio only and the radar information and without being able to give any advice. They conventionally know (“deus ex machina”) that no instructions will be given to either of the two aircraft in the future. Then they have to come to a decision about the future separation values, and the probability for the two protagonists to be under the 3NM/1000ft limit. It was then emphasized that a margin as tiny as discernible upon these limits could perfectly justify a “non-conflict” answer.

Responses should be entered by means of the interface above (fig2), supported on a tactile display. The upper part showed the time interval of the scenario, while two summaries of the flight plans appeared in the lower part. In the middle stood a series of 8 colored keys, in fact two groups of four. The four on the left should be chosen when the judgment was in favor of separation loss (conflict). The four on the right meant that sufficient separation was the most probable outcome (non-conflict).

Once the binary decision was taken, it had to be refined by choosing the one of the four keys that represented the initially binary decision. Short legends and colors were given to these keys, in order to make sense of the judgment levels.

**Participants**

161 ATCSs from four different French Air Traffic Control Facilities participated to this study (45 in Lyon, 35 in Toulouse, 45 in Marseille, and 36 in Bordeaux). The overall sample split into 75.2% male (mean age 44.8) and 24.8% female (mean age 40.6).

To participate, each of them had to be fully qualified for at least two years in their respective facilities, but the mean ages (above) show that they were largely more experienced (generally more than 10 years).

The volunteers who kindly participated in the experiment had to arrange with their own shifts in order to free themselves for about an hour (or two half-hours), the approximate duration of the experimental test.

**Results: a Graphical Analysis of Data**

In this section, a data exploration is carried out by using graphical facilities. The response variable being a categorical variable modeled as a function of three continuous explanatory variables, required us to find a way to represent a four dimensional structure in two dimensions. Firstly, we established an 8-point scale from “conflict is certain” (-10) to “non-conflict is certain” (+10). These response variables were numerically equidistant and took the values {-10, -7.14, -4.29, -1.43, 1.43, 4.29, 7.14, 10}. This enabled the rapid identification of the conflict/non-conflict characteristic of any result (through the algebraic sign) and the uncertainty level attached to it (through the absolute value).

**Judgment changes as separation values evolve**

The three explanatory variables are a temporal variable $D°A$ and two geometrical variables $E(d)$ and $E(fl)$ with the constraint that $E(fl)$ is always greater or equal to $E(d)$. Thus, we can consider the following model:

$$ response \text{ variable } = f_{D°A}(E(d), E(fl)),$$
That is to say for each fixed value of the variable D^A, the response variable may be expressed as a function of the two variables E(d) and E(f). The response variable was coded by different grey squares (= "levelplots"), from white (non conflict is certain i.e. no reasonable risk of separation loss exists) to black (conflict certain i.e. a separation loss will surely happen), with respect to the eight response keys on the collecting display.

These eight judgments were coded from -10 (conflict is certain) to +10 (non conflict is certain). The medians of the whole sample of participants' responses are shown by means of these grey squares on fig 3. The eight squares with crosses, on the bottom right, correspond to geometrically unfeasible scenarios, and have consequently no median value.

![Figure 3: Medians for all of the responses.](image)
In this paper, we concentrated only on the average behavior of the entire population, so we use the median, which is the 50th percentile value, to describe the categorical response value. The $p^{th}$ percentile for a population is the value which divides the population into two parts: the lower $p\%$ of the values and the upper $(100-p)\%$ of the values. Thus, the median is the value which divided the set of data into two equal size subsets.

The way that the level plots for each value of the variable $D^oA$ are displayed in figure 3 indicates that:

1. The size of the “medium grey” area corresponds to the doubt increase with $D^oA$. The percentage of medium grey squares increases for high values of $D^oA$. It means that doubt is perceptible in the corresponding judgments of controllers. Conversely, low values of $D^oA$ are concomitant with “black” or “white” squares: judgments become more certain as distance to the I.P. decreases and conflict/non-conflict classification becomes easier.

2. Even if for $E(d)=E(fl)=0$, which represents the certain conflict from a purely geometrical point of view, the influence of prediction time span ($D^oA$) on the perception by the controller is qualified. As $D^oA$ value grows, judgment becomes more doubtful, even if remaining about “conflict” (“dark” squares). Another graphical evidence of this result is shown in figure 4, on the second of the three graphs.

To illustrate more precisely, how the anticipation degree alters the judgement of the controller, we must take into account the variability of the response for each subject. For that, we used a box plot which summarised a great deal of information (see fig4). The black dot inside each box shows the median of the response for each value of $D^oA$. The left and right sides of the blue box represent respectively, the 25th and the 75th percentiles ($Q_1$ and $Q_3$). The length of the box is the difference between $Q_3$ and $Q_1$, called the interquartile range. This single number gives the width of the interval of values into which the middle 50% of the data fall. It plays a similar role with respect to the median to that which the standard deviation plays with respect to the arithmetical mean. The vertical dashed lines joined to the box by a dashed line show 1.5 times the interquartile range, and points beyond this are drawn individually (“outliers”). Box plots not only show the location and the spread of data, but also indicate the asymmetry in the sizes of the upper and lower parts of the box.

Judgment changes as prediction time span evolves

As said above, the judgment clearly changes with the distance remaining for the aircraft to reach the I.P. In figure 4, the box plots for some pairs of $E(d)$ and $E(fl)$ are displayed, representing non-conflict, conflict and “near-minima” situations. The non-conflict example was selected for not appearing obvious too early ($E(d)=6$ and $E(fl)=10$).

*Comment on the left graph.* First, when aircraft are still located far from the I.P. (36 and 28 NM) judgments tend to show globally careful preference for “non-conflict” answers. Four response levels are included between the 25th and the 75th percentiles, and the whole range is significantly used by the totality of participants (i.e. whiskers go from one extreme to another). As anticipation (i.e., distance to I.P.) decreases, judgments become firmer in the sense of a non-conflict verdict. Hesitation can hardly be found or $D^oA=21$ or $D^oA=15$, as median is 7.14, the first still presents a real range of variance, however. For the last two values (10 and 6), the median reflects a quasi certainty of non-conflict (the larger variance range of the latter should probably be attributed to minor experimental biases).

![Figure 4: Boxplots (medians and interquartile range) for three separation values: non-conflict, conflict and a borderline case.](image-url)
Comment on the middle graph. When considering judgment changes for opposite class of separation, that is to say for conflicts, a symmetrical sequence is found for the successive values of D°A. Responses medians have a “conflict” tendency from the beginning, and become firmer as anticipation decreases. Variation decreases logically and quite regularly.

Comment on the right graph. A third case can be found in the data: those which correspond to separation values close to the required separation minima: [0;5], [3;5], [0;10], representing [E(d);E(fl)] values. Fig4 (above) is given as an example. For these doublets, two distinctive features appear before fading progressively as the two separation values move away from the minima to more comfortable distances.

First, non linearity of the responses median is observable when D°A evolves, unlike the other boxplot graphs corresponding to the other doublet values, which could not take place here. For these latter, monotony is in place i.e., the judgments do not show two different tendencies (simultaneously firmer in conflict and in non-conflict verdicts) as D°A decreases.

Secondly, variance (both represented by the boxes and the whiskers) remains very high for the six values, which could not take place here. For these doublets, two distinctive features appear before fading progressively as the two separation values move away from the minima to more comfortable distances.

A coherent ranking of responses

<table>
<thead>
<tr>
<th>E(d)</th>
<th>E(fl)</th>
<th>D°A</th>
<th>Median</th>
<th>Interquartile range</th>
<th>Arithmet. mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>6</td>
<td>-10</td>
<td>2.86</td>
<td>-6.57</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>10</td>
<td>-10</td>
<td>2.86</td>
<td>-7.30</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>15</td>
<td>-1.14</td>
<td>5.71</td>
<td>-5.95</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>21</td>
<td>-1.14</td>
<td>8.57</td>
<td>-5.07</td>
</tr>
<tr>
<td>0</td>
<td>5</td>
<td>15</td>
<td>-1.14</td>
<td>8.57</td>
<td>-4.14</td>
</tr>
<tr>
<td>0</td>
<td>5</td>
<td>10</td>
<td>-1.14</td>
<td>11.43</td>
<td>-3.69</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>28</td>
<td>-4.29</td>
<td>5.71</td>
<td>-4.09</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>36</td>
<td>-4.29</td>
<td>7.14</td>
<td>-2.73</td>
</tr>
<tr>
<td>0</td>
<td>5</td>
<td>21</td>
<td>-4.29</td>
<td>8.57</td>
<td>-3.13</td>
</tr>
<tr>
<td>0</td>
<td>5</td>
<td>6</td>
<td>-4.29</td>
<td>14.29</td>
<td>-2.51</td>
</tr>
<tr>
<td>0</td>
<td>5</td>
<td>36</td>
<td>-1.43</td>
<td>8.57</td>
<td>-1.92</td>
</tr>
<tr>
<td>0</td>
<td>5</td>
<td>28</td>
<td>-1.43</td>
<td>8.57</td>
<td>-1.07</td>
</tr>
<tr>
<td>0</td>
<td>10</td>
<td>28</td>
<td>-1.43</td>
<td>8.57</td>
<td>-0.59</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>36</td>
<td>-1.43</td>
<td>8.57</td>
<td>-0.56</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>36</td>
<td>-1.43</td>
<td>8.57</td>
<td>-0.23</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>21</td>
<td>-1.43</td>
<td>8.57</td>
<td>-0.07</td>
</tr>
<tr>
<td>0</td>
<td>15</td>
<td>36</td>
<td>-1.43</td>
<td>8.57</td>
<td>0.33</td>
</tr>
<tr>
<td>0</td>
<td>10</td>
<td>36</td>
<td>-1.43</td>
<td>10</td>
<td>-0.93</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>10</td>
<td>-1.43</td>
<td>14.28</td>
<td>-0.60</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>28</td>
<td>-1.43</td>
<td>14.28</td>
<td>-0.47</td>
</tr>
<tr>
<td>0</td>
<td>10</td>
<td>21</td>
<td>-1.43</td>
<td>14.28</td>
<td>-0.29</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>28</td>
<td>-1.43</td>
<td>14.28</td>
<td>-0.01</td>
</tr>
</tbody>
</table>

Table 2: Negative (i.e. conflict) median values.

From the ATC expertise point of view, a strong coherence appears as each of the I.V. varies. As an example, the selection of 22 response medians is shown on table 2 in a numerical form. As indicated by the negative sign of the median, they correspond to the triplets of variables - i.e., scenarios - for which separation violation was predicted (the three columns on the left). These scenarios are sorted by median values (first sort), then with the interquartile range value (second sort). What is more, the arithmetical means have been computed (right column) upon the 161 responses given for each of them, including the outliers.

The classification above would not have been very different if it had been made with E(d) as unique criterion. If E(fl) had been the criterion it would obviously have been more different. Lastly, the classification taking D’A into account as criterion would have been very different from the one above.

**Discussion**

**Does a common core exist in ATCSs’ judgments for conflict detection?**

In spite of the great inter individual variability always found in past studies on this question, our hypothesis was affirmative since everyday practice shows controllers sharing airspace and simultaneously giving highly coordinated services to each flight. It could always be observed that they communicated only when necessary, even more barely as an increasing traffic load jeopardized their processing capacity. Maintaining a high security level in such conditions suggests they use a (partly) common and analysis of the situation. This includes judgments (conflicts detection, particularly) upon upstream and downstream traffic, dynamically compatible with those of other controllers. A second clue of such a common core was the way that novices are given practical training. A prerequisite for them to make any conflict detection skills progressively emerge should be the identification of this common base from the different successive trainers they always have.

So even if a large variability was expected - and could finally be found - in individual responses, the global consistency appeared remarkably high, and indicated a common core in ATCSs’ judgment. Working out the medians of responses for a significant number of participants has eliminated the major part of the random effect (“noise”) due to individual differences. A parallel can here be drawn with the technique of evoked potentials (ERP), stemming from repeated measures of the electric activity of the brain. Computing of the algebraic sum of a sufficient number of such signals allows the significant signal to emerge, despite every response differed from all others.
Thus, these median values outline a possible conflict perception model. Though it certainly has to be limited to the types of conflicts in input, it provides a useful quantification to human perception of ATC conflicts.

Is the judgment in this study “quantitative-like”?
Classification of the scenarios according to the sorted medians (table 2) is in accordance both with the knowledge already got about the properties of our three independent variables, and with all the observations and experience I have gained during my career as an air traffic controller.

Furthermore, a very similar ranking is obtained when sorting the scenarios according to the arithmetical means. These similarly satisfactory rankings can be observed for the entire classification of the 90 scenarios that could not be reported here. So even if the ordinal nature of the responses should not be questioned, the evidence suggests that the nature of the collected data is close enough to those of an appropriate quantitative variable. This should allow using some quantitative methods (such as regression) in a future stage of responses processing.

Comparing the “separating power” of horizontal and vertical dimensions of trajectories
When separation increases the community of the participants sees it continuously, as expressed in the medians (fig3 and 4, and table 2). When prediction time span decreases, accuracy on the judgment given by the community improves quite monotonously (id.). Hence each Nautical Mile measuring E(d) has more “separating value” than its counterpart representing the vertical movements of aircraft. As we said above, this result is consistent with all of the past studies. The originality here is to have modeled and used E(fl), intended to be highly compatible with the mental model of ATCSSs. This enables also the comparison of both dimensions, as they are expressed with the same unit. As seen in table2, there is evidence that controllers regard longitudinal separation (Ed) as being clearly more able to guarantee spacing through controllers’ perception.

In order to settle a better legitimacy of E(fl), a complementary study is planned. It will process the same collected responses through another set of independent variables, taking into account verticality in a simpler (i.e., non-modeled) way than E(fl) did here. The simple numerical value of aircraft altitude visible in the label, will be computed instead of E(fl).

Uncertainty management by controllers results in doubtful (but not incoherent) responses
The way for participants to express doubt in a judgment was to choose one of the two inner (yellow) keys (fig 2). In so doing, they gave a picture of the uncertainty they consciously perceived. A different aspect of doubt is observable by means of change in the variance (interquartile range). This would indicate a lack of improvement in response accuracy as D°A decreases. It can be seen that a climax is reached for near-minima scenarios (for example when [0;5] or [3;5] stand for [E(d);E(fl)]). As shown in fig4, whiskers surprisingly keep going from one extreme response level to the other for all of D°A levels, meaning that all this range should be considered. Paradoxically the variability is even maximum as aircraft come closer to each other (D°A=6).

This can be interpreted as the expression of an “unconscious” or “collective” doubt. The near-minima scenarios share the property that separations are large enough to be lately and hardly perceptible, but still make risky the non-conflict decisions from expert’s point of view. It seems as if two opposite tendencies exerted their influence on them as the I.P. became close, leading the controllers to spread out their responses all over the scale. The first tendency was to decide that doubt was definitely too high to be acceptable and then formulated a clear conflict judgment. The second one was to follow more or less the first perceptive analysis and to proportionally formulate a non conflict judgment. So the judgments on these three scenarios could individually be extreme and look clear-cut but they collectively succeeded in demonstrating that each decision had initially to deal with an important doubt.

This interpretation is reinforced by the variations of the medians in the same figure. Expressing initially a cautious verdict (36 and 28 for D°A) a perfect conflict verdict seems to progressively emerge as it comes to 21 and 15. Coming close to the I.P., the hardly perceptible reality of a separation begins to divide controllers, making the width of the two last boxes increase to a maximum and even at the last median retraing somewhat from a conflict judgment.

This result can be related to Bisserey’s work [8] comparing the conflict detection skills of experts and beginners by means of Signal Detection Theory (SDT). Significantly, he already interpreted the higher rate of “doubt responses” of experts as a consequence of their greater experience. Making judgments rather than computing, they tend to create a class of temporary indecision which should not be confused with the indecisiveness which is concomitant upon a lack of skill.

Global consequences and limits of this study
The task we submitted used devices and problems as similar as possible to real work ones. Further, the resources summoned up were minimized (sufficient time for the responses, no simultaneous contextual taskload, etc). In real situations, when the relative positions of multiple aircraft need to be predicted,
the controller’s processing capabilities will be heavily taxed, thus limiting the accuracy of prediction [3] [11]. In order to purely apply these findings to real traffic, other sources of variation should be added to the three ones here. Furthermore, the differences between en-route and approach traffic induce differences in conflict detection. Lastly, the performance of devices and information shown on displays (labels, speed-vectors, etc) and operating methods, might impair these results. An example can probably be found in the “miles-in-trail” strategy, used in the vicinity of major airports in the United States, far less in Europe. This could confer less importance on this conception of conflict detection performance, by emphasizing the elimination of conflict omissions, for example.

Because the participants were asked to behave as if they were only witnesses of the traffic scenario, only giving a visually based prediction, present results are supposed to be predominantly independent of any decision making. It is felt that this condition may not have been totally fulfilled, as for experts resolution and detection processes are too closely involved with each other. Disconnecting them experimentally through instructions remains somewhat contrived. Some indications of this could be perceptible to the experimenters while they were submitting the test, and could be responsible of a significant part of the “noise”, i.e. the unexplained part of variability.

Since the presented data has been collected recently, the processing is ongoing, especially a global regression model computed with this data. This method will allow incorporating both fixed effects, i.e. those attributable to the I.V., on the one hand, and random effects, i.e. the “noise” associated with individuals, on the other. This is known as the mixed-effects model (linear or nonlinear). The effects of age and sex of participants will then be investigated. This might account for variability between results either coming from the different Tracons, or issued from the different GPA.

Conclusion

This first stage of our data processing already permits us to express results modifying two fundamental features of expertise in ATC. First, the notion of intrinsic perceptibility of conflicts makes sense, as medians of judgments emerging from the ATCS community can be seen as a consistent core of expertise. It emerges from individuals and prefigures a possible model of conflict decision, in spite of the well-known variability of judgments. Such a result would directly contribute to improving the design of future tools, for example, in providing a concrete evaluation of the real upper limit of the discrimination capacity of ATCS’ mental picture. Secondly, the expression of doubt has been found to have two different aspects, each of them being put in evidence and quantified in distinct ways. On the one hand, low values demonstrate controllers’ choice to express doubt. On the other hand, a constant and large variance betrays a less approachable aspect of the controller’s dealing with uncertainty that would have not emerged otherwise. As doubt management in real time/real task is particularly essential in ATC expertise modeling, and is at the same time quite difficult to manipulate, such an opportunity to say more about it is positive.

Acknowledgements

Support in Statistics was given from Pascal Lezaud (SDER/ENAC), while all the Computer Science aspects are the fruits of J.P. Imbert’s work (SDER). I am indebted to JL Vinot (SDER) for his advice and assistance with the graphics.

References


**Keywords**

Air Traffic Control; Conflict; Detection Process.

**Biographical Notes**

Philippe Averty is an Expert in ATCS Modeling with the Sous-Direction des Etudes et de la Recherche appliquée (SDER, ex-CENA), France. He worked as an ATCS at Paris ACC, then as ATCS and Instructor at Lyon TRACON. He passed his PhD in Cognitive Psychology in 1998, and started to work at the CENA in 2001 on workload assessment, workload/complexity relationships, and currently runs the project CREED (conflict perception).