DEVELOPMENT AND EVALUATION OF STEREOSCOPIC SITUATION DISPLAYS FOR AIR TRAFFIC CONTROL

David Wittmann¹, Andreas Baier², Harald Neujahr¹, Benedikt Petermeier², Peter Sandl¹, Christoph Vernaleken¹, Leonhard Vogelmeier¹

1: EADS Deutschland GmbH Cassidian Air Systems Rechliner Straße 85077 Manching, Germany david.wittmann@cassidian.com 2: Universität Regensburg Institut für Psychologie Universitätsstraße 31 93053 Regensburg, Germany

ABSTRACT

The present study investigates the suitability of stereoscopic 3D displays for Air Traffic Control (ATC) purposes. Two groups of participants, air traffic controllers and laypersons, were asked to judge various potential conflict situations involving two aircraft, using a specifically developed console work station that enables a comparative evaluation of different 3D visualisations and the current 2D display. The quality of judgements in combination with usability and workload ratings is used as metrics to identify the most appropriate visualisation concept. For part of the scenarios, the conflict detection task was carried out with an additional secondary task. Considering exclusively the quality of judgement values laypersons generally achieve a higher performance when using 3D displays, whereas ATC controllers exhibit a constantly high performance with all visualisations. Particularly in scenarios that are difficult to assess and under secondary task conditions the laypersons seem to benefit from the 3D presentations, whereas the experts perform better with the 2D reference. From the quality of judgement values alone no preference is evident for a particular 3D concept. Taking into account the usability and workload ratings, however, the picture gains contour. Laypersons prefer the variant "3D with perpendicular" while experts are in favour of their familiar 2D variant, with the "3D with perpendicular" concept being their second choice. The results, however, are preliminary as the study is still ongoing.

Index Terms - Air Traffic Control, Stereoscopic Displays, 3D-Displays, Workload, Usability

1. INTRODUCTION

The primary task of Air Traffic Control (ATC) is maintaining a safe and efficient flow of air traffic. Air traffic controllers must therefore be able to quickly recognise and understand the current air traffic situation and to estimate how it will evolve within the near future. Accordingly, they have to mentally integrate multidimensional information, thus forming a mental picture of the situation. Among the most important aspects are the reliable judgement of the movement of aircraft and the avoidance of conflicts. Since the 1960s, radar technology has permitted controllers to determine and monitor the position of aircraft [10]. For many years, the fusion of data from several sensors has enabled an integrated visualisation of radar and flight plan information (e.g. call sign) on the central Situation Data Display of the work station. However, the general principle of displaying air traffic has not changed fundamentally since the 1960s: aircraft are presented on a conventional display screen, using symbols on top of a map-like background. Altitude is indicated alphanumerically. A presentation of the air situation by means of an intuitively understandable three-dimensional representation could help the controllers in future to operate air traffic safely and efficiently despite the additional burden caused by an ever increasing growth of air traffic.

2. CURRENT STATE OF RESEARCH

The application of stereoscopic displays for ATC purposes has already been proposed in the late 1980s [18]. Since then, numerous studies have been devoted to the conception and investigation of this kind of visualisation [6], [14]. Research comparing 2D and 3D presentations, however, does not provide a consistent picture: while an

©2011- TU Ilmenau 1

investigation by Wickens [17] yielded neither clear advantages nor disadvantages of 3D presentations, a study by van Orden and Broyles [16] showed that 2D presentations are equally well suited to assess speed and altitude of aircraft as 3D presentations. The same study, however, revealed that the perception and assessment of conflict situations is quicker and more reliable with a 3D display, compared to 2D. Similar results were generated by a study of Tavanti et al. [15], carried out with air traffic controllers and ATC experts, in which 3D displays enabled a faster estimation of altitude with equal precision.

3. PROBLEM STATEMENT

Due to the contradictory research results, no clear statement can currently be made with regard to the suitability of 3D presentations for ATC purposes. In order to facilitate further investigations, Cassidian Air Systems, in the context of a project called "innovative airport (iPort)", built a controller work station (Fig. 1) that allows generating sufficiently realistic experimental conditions. The respective basis was provided by Baier [4] with an analysis of controller tasks. He identified the perception and assessment of conflict situations as one of the most relevant tasks of an air traffic controller and the Situation Data Display as the most important element of the human machine interface.

The evolvement of the traffic presentation on the Situation Data Display from 2D to 3D raises a number of questions, for instance on the most appropriate perspective for the evaluation conflict scenarios. Beyond, it is also of importance to establish whether 3D is of any advantage over 2D, and if so, in which situations. Additionally, it must be investigated whether 3D causes variations in controllers' mode of operation.

4. CONCEPTUAL DESIGN OF A NOVEL WORKSTATION FOR AIR TRAFFIC CONTROL

A prototype of the novel workstation has been realised exclusively for the investigation. Since it must permit, just as the current 2D workstation, long-lasting, concentrated operation, great emphasis was given to the consideration of all available recommendations in this field [9], [11]. The display system can be used for two-dimensional, perspective ("2,5D") and stereoscopic three-dimensional presentations. The resolution of the display is in the range of the 2K ATC Situation Data Display currently in use. Thus, the workstation allows to display complex air traffic scenarios including all relevant geographically referenced information in a unique and integrated presentation.

In order to identify a suitable form of visualisation, three different 3D presentation variants were examined for their usefulness in evaluating scenarios involving a potential traffic conflict, and compared to the current 2D reference presentation. Display content was limited to the following elements: current position (aircraft symbol), past positions (aircraft trail), label with alphanumeric speed, altitude and an arrow indicating the vertical speed, a velocity vector showing the current movement, and a reference grid. In all four variants, these information elements were presented in a comparable manner.

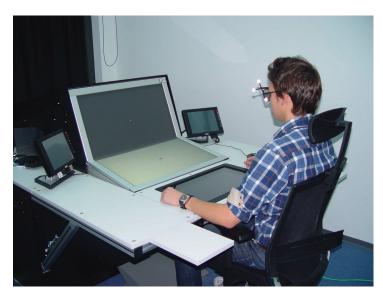


Figure 1. Prototypical controller work station

The variants shown in Fig. 3 are both based on the "electronic sandbox" metaphor [1]. The scenario seems to rise in front of the user over a horizontal display. The difference between both variants is that aircraft in the left

variant feature a perpendicular pointing towards the reference surface, whereas aircraft in the right variant do not. In theory, such a subsidiary line should only be necessary with perspective presentations [2], but not with stereoscopic displays. To verify this assumption, both variants have been tested.

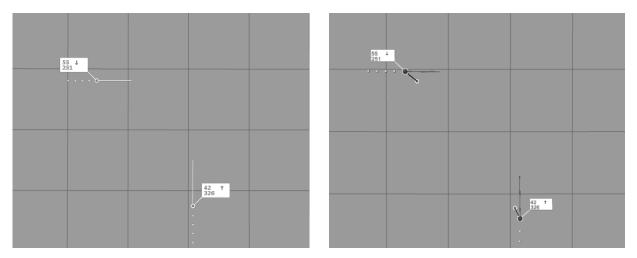


Figure 2. Current 2D-visualisation concept (left) and the 3D-top view concept (right)

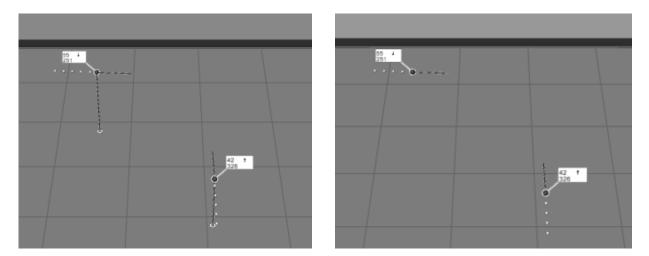


Figure 3. 3D-visualisation concepts: with (left) and without perpendicular (right)

5. EXPERIMENTAL SETUP AND PROCEDURE

5.1. Participants

Air traffic controllers from Deutsche Flugsicherung GmbH were available for the present study. In order to compare results for controllers and persons lacking their specific experience, a control group consisting exclusively of laypersons was additionally introduced. To date, the experiment has been conducted with 16 participants, seven (7) of which were air traffic controllers. It is envisaged to collect data from a total of 48 participants; the results presented in this paper should therefore be considered as preliminary.

5.2. Primary task

In the first part of the experiment, each participant was to determine the probability of a critical air proximity event involving two aircraft. A total of 86 scenarios was presented, half of which featured actual conflicts, so-called hit-scenarios. The other half consisted of scenarios in which the two aircraft missed each other either spatially (i.e. no intersection of trajectories) or temporally (intersection at different times). In both hit- and miss-scenarios, the behaviour of aircraft was characterized by airspeed, climb/sink rate and relative track. For the latter parameter, there was a subdivision in three categories: 0° (catch-up), 90° (crossing) and 180° (head-on). According to literature, the 90° scenarios are most difficult to judge with respect to potential conflicts [8].

All scenarios commenced 45 seconds prior to the closest point of approach. After 10 seconds, the presentation of the scenario was replaced by an evaluation screen on which participants had to give their judgement on the probability that the presented situation would result in a collision, using a six-stage rating scale ranging from "certainly yes" to "certainly no".

5.3. Secondary task

In the second part of the experiment, 32 of the 86 scenarios were presented again, this time limited to the difficult 90° constellations. Compared to the first part, a simultaneous auditory-verbal secondary task was added, in which participants should verbally indicate the successive occurrence of two pre-defined numbers within a sequence of digits. Participants were instructed to divide their attention equally between both tasks. This secondary task, which affects the verbal part of working memory [3], emulates key characteristics of radio communication between air traffic controllers and pilots [13] while not requiring the use of specific phraseology, which ensures that the secondary task can be fulfilled by controllers and laypersons alike.

5.4. Performance measurement

The quality of participants' judgement was used as a metrics for the suitability of a 2D/3D presentation variant. Judgement quality is the percentage of correctly characterised potentially conflicting traffic constellations. Accordingly, an entirely correct identification of all hit- and miss-scenarios would result in a judgement quality of 100%, whereas 50% corresponds to random guessing.

5.5. Usability and workload measurement

After rating the probability of conflicts in the presented traffic scenarios, participants filled in two standard questionnaires on usability and workload for each of the four visualisation concepts. A usability questionnaire based on ISO 9241 [12] was adapted to the task at hand. Participants were asked whether the respective visualisation concept was suitable to attain situational awareness according to Endsley's widely used model [5], i.e. whether it enabled *perception*, *comprehension* and *prediction* of potential conflicts. Furthermore, participants were to rate the different visualisation concepts with respect to *task suitability* and they were asked whether the respective visualisation concept requires an unnecessary search for information.

Workload was measured using the NASA TLX [7] for the six categories *mental*, *physical* and *temporal demand* as well as *performance*, *effort* and *frustration*.

6. RESULTS

Due to the so far small number of participants no detailed statistical analysis of the collected data will be done for the time being, since particularly workload and usability ratings exhibit a substantial variance. Nevertheless, an analysis of the presently available data shows some noteworthy results.

6.1. Performance data

The average quality of judgement for all constellations without secondary task (Table 1) shows that air traffic controllers generally achieve a better performance than lay persons irrespective of the visualisation concept employed, which is fully in accordance with expectations. At the same time, all 3D visualisations enable a higher accuracy than the 2D reference, but the effect is small, and none of the 3D visualisation concepts appears to be superior to the others.

A comparison of judgement quality with and without secondary task (Table 2) yields that the decrease in performance is smaller for air traffic controllers (2.2% - 5.5%) than for laypersons (7.3% - 10.0%); this effect applies to all visualisation concepts alike. Since the experiment with simultaneous secondary task was limited to the more difficult 90° scenarios, the overall performance in Table 2 is consistently lower than in Table 1. Nevertheless, results also show that laypersons benefit from 3D visualisations more than air traffic controllers, although experts - as expected - always achieve a higher performance than laypersons.

Participants' performance for the secondary task itself was partially better when using a 3D visualisation, but the effect is minor (Table 3). The only concept which involves a deterioration compared to 2D is the variant "3D w/o perpendicular", the magnitude of deterioration, however, can be regarded as negligible.

Table 1. Mean quality of judgement with standard error in brackets of laymen and air traffic controllers (ATCo) without secondary task over all scenarios. Values printed in **bold** indicate cases where performance with 3D visualisation was better than with 2D.

	2D		3D		3D		3D	
	baseline		top view		w/o perpendicular		w. perpendicular	
user group	Laymen	ATCo	Laymen	ATCo	Laymen	ATCo	Laymen	ATCo
overall performance	86.4	91.0	89.2	92.3	88.1	91.9	87.1	93.0
	(1.4)	(1.2)	(1.2)	(1.2)	(1.3)	(1.2)	(1.4)	(1.1)
rank	4.	4.	1.	2.	2.	3.	3.	1.

Table 2. Mean quality of judgement with and without secondary task. Standard error is given in brackets. Values printed in **bold** indicate cases where performance with 3D visualisation was better than with 2D, values in *italics* where it was worse.

	2D baseline		31 top		3D w/o perpendicular		3D w. perpendicular	
user group	Laymen	ATCo	Laymen	ATCo	Laymen	ATCo	Laymen	ATCo
w/o secondary task	77.7	86.2	81.9	82.8	77.6	83.1	81.6	85.8
	(2.7)	(2.4)	(2.4)	(2.7)	(2.7)	(2.8)	(2.6)	(2.6)
with secondary task	67.7	83.0	72.2	78.5	70.3	80.9	72.5	80.3
	(3.2)	(2.8)	(3.0)	(3.1)	(3.1)	(3.0)	(3.0)	(3.0)

Table 3. Completion of the aural-verbal secondary task in percent. Standard error in brackets. Values printed in **bold** indicate cases where performance with 3D visualisation was better than with 2D, values in *italics* where it was worse.

	2D		3D		3D		3D	
	baseline		top view		w/o perpendicular		w. perpendicular	
user group	Laymen	ATCo	Laymen	ATCo	Laymen	ATCo	Laymen	ATCo
performance 2nd task	78.9	81.0	79.9	83.4	78.9	80.4	85.6	81.7
	(14.9)	(14.0)	(15.3)	(8.5)	(14.3)	(11.7)	(12.1)	(16.1)

6.2. Usability and workload ratings

The analysis of the usability ratings (Table 4) shows that air traffic controllers consider the current 2D presentation as most appropriate for the task. This applies to each item of the questionnaire, not only the overall score. Regarding the stereoscopic visualisation concepts, air traffic controllers rate the variant "3D with perpendicular" best. This is also true for laypersons, who additionally prefer the "3D top view" variant over the "2D-baseline" variant in all aspects. Both user groups rate the "3D without perpendicular" variant worst.

Table 4. Usability ratings. The table shows mean ratings (1 is worst, 7 is best) with standard deviation in brackets. Better ratings in comparison with the 2D baseline concept are printed **bold**, worse ones *italic*.

	2D baseline		31 top		3D w/o perpendicular		3D w. perpendicular	
user group	Laymen	ATCo	Laymen	ATCo	Laymen	ATCo	Laymen	ATCo
suitability for the task	4.6	6.0	5.6	4.9	4.6	4.7	6.2	5.6
	(1.5)	(0.8)	(1.0)	(1.7)	(1.7)	(1.5)	(0.8)	(1.0)
perception	4.9	6.0	5.2	5.4	4.0	5.0	6.0	5.4
	(1.4)	(1.2)	(1.0)	(1.3)	(1.5)	(1.7)	(0.5)	(1.0)
comprehension	4.4	6.1	6.0	5.0	4.4	4.9	6.2	5.7
	(1.7)	(0.9)	(0.9)	(1.3)	(1.1)	(1.6)	(0.7)	(0.8)
prediction	3.6	5.6	5.4	5.4	3.9	4.9	6.0	5.4
	(1.5)	(1.3)	(0.7)	(0.8)	(1.2)	(1.4)	(0.9)	(0.5)
information search	4.3	6.7	5.4	5.7	4.0	5.9	6.0	6.0
	(1.6)	(0.8)	(1.4)	(1.7)	(1.7)	(1.1)	(1.0)	(0.6)
overall score	4.4	6.1	5.6	5.3	4.2	5.1	6.0	5.6
rank	3.	1.	2.	3.	4.	4.	1.	2.

The NASA-TLX workload scores reveal the following: **overall,** the 2D baseline concept is rated best by air traffic controllers, followed by the variant "3D with perpendicular", which is also the concept favoured by the laypersons, closely followed by the second-best rated variant "3D top view". Air traffic controllers rated all 3D-variants in all categories lower than the 2D reference, with one exception: the effort to work with the variant "3D with perpendicular" was clearly perceived to be lower than with the baseline (-2.0).

In the workload category **mental demands** the laypersons gave to lowest rating to the variant "3D with perpendicular", but also the variant "3D top view" made less mental demands compared to the 2D baseline.

The **physical demands** when working with the 3D concepts are clearly rated higher by both laymen and air traffic controllers, compared to the 2D baseline. The participants were explicitly asked how much physical activity was necessary to fulfil the task (e.g. for looking around, moving the head etc.). Insofar, the result is not surprising, since working with the 3D variants is associated with multiple changes of the viewing direction and a gain in information if the head is tilted, thus changing the perspective, even tough the level of physical activity is rather low.

The **temporal demands**, i.e. the time pressure, is rated lowest by the laymen in the variant "3D with perpendicular" (-1.9), which might be a clue that this variant allows for the fastest and most intuitive assessment of the situation.

The own **performance** was assessed by the laypersons to be best with the variant "3D top view". The differences between the variants, however, are so small that a clear tendency towards a particular variant can not be established.

The **effort** necessary to fulfil the task was rated by the laymen to be lower with all 3D variants than with the 2D reference. The air traffic controllers share this opinion with the variant "3D with perpendicular".

The **frustration** ratings show only minor differences, in case of the laymen the variant "3D with perpendicular" is slightly better than the rest. Overall the experts register clearly lower frustration levels than the laypersons, which may be attributed to their familiarity with the task.

Table 5. NASA-TLX workload ratings. The table shows mean ratings (0 is best, 20 is worst) with standard deviation in brackets. Values printed in **bold** indicate cases where performance with 3D visualisation was better than with 2D, values in *italics* where it was worse.

	_	2D 3D top view		3D w/o perpendicular		3D w. perpendicular		
User group	Laymen	ATCo	Laymen	ATCo	Laymen	ATCo	Laymen	ATCo
mental demand	10.4	8.1	9.6	9.7	11.3	10.1	8.0	9.6
	(4.6)	(4.1)	(4.2)	(3.7)	(4.3)	(4.7)	(3.9)	(3.6)
physical demand	2.8	3.0	5.9	4.4	9.9	6.9	7.2	4.6
	(1.2)	(1.7)	(4.3)	(2.9)	(5.2)	(5.3)	(4.0)	(2.2)
temporal demand	7.9	5.7	8.6	7.3	9.6	8.4	6.0	6.9
	(3.8)	(3.6)	(2.4)	(3.6)	(4.4)	(4.7)	(3.9)	(3.2)
performance	7.4	4.7	7.0	6.4	8.1	6.9	8.0	5.6
	(2.9)	(1.8)	(3.3)	(3.2)	(2.7)	(4.0)	(3.5)	(3.1)
effort	12.6	10.0	9.3	10.3	10.9	10.4	10.4	8.0
	(1.9)	(4.6)	(4.0)	(4.2)	(3.7)	(3.8)	(3.6)	(5.0)
frustration	7.9	4.4	7.3	5.7	8.5	6.0	6.8	5.1
	(5.1)	(4.0)	(4.0)	(3.9)	(3.3)	(5.0)	(3.6)	(6.0)
overall score	8.2	6.0	7.9	7.3	9.7	8.1	7.7	6.6
rank	4.	1.	2.	3.	3.	4.	1.	2.

At the end of the experiment, all participants were asked to give their personal ranking of the four display concepts (Table 6). The laymen's ranking-order coincides with the results of the NASA-TLX workload measurement. The ranking of the air traffic controllers coincides on the first two ranks (2D baseline and 3D with perpendicular), whereas it is inverted on ranks three and four.

Table 6. Ranking order (average), 1 is best, 4 is worst.

	1.	2.	3.	4.
Laymen	3D	3D	2D	3D
	w. perpendicular	top view	baseline	w/o perpendicular
	1.3 (0.7)	2.1 (0.6)	2.8 (1.0)	3.8 (0.4)
ATC ₀	2D	3D	3D	3D
	baseline	w. perpendicular	w/o perpendicular	top view
	1.7 (0.8)	2.1 (1.2)	2.7 (1.1)	3.4 (0.8)

7. DISCUSSION

Based on the data collected so far, no clear recommendation concerning the design of a stereoscopic display system for Air Traffic Control can be made. Although there are indications that 3D presentations offer a marginally higher judgment quality than the 2D reference system (Table 1), no clearly superior perspective can be identified. Taking into account the results of the usability interview and the workload measurements, the laypersons show a clear preference for the 3D-with-perpendicular variant, while the air controllers prefer the 2D presentation they are familiar with.

The fact that laypersons, in contrast to experts, benefit form the stereoscopic presentations, particularly when the scenarios are difficult to assess (Table 2), emphasizes the potential of 3D displays. If considering only the results for the air controllers, one could conclude that 3D presentations are less appropriate for the primary task of ATC. By including the results for the laypersons, who, in most cases, achieve a better performance with 3D

presentations than with the 2D reference, this conclusion can not be supported. Instead, it is rather to be presumed that 3D displays allow a more intuitive identification of conflicts. The reason why experts perform worse with 3D displays could be their long-term experience with the 2D display in combination with established conflict assessment strategies, which may be affected adversely when using the stereoscopic variants.

Once the experiments have been completed with the intended number of subjects a detailed statistical analysis will follow which will finally serve to select the most promising 3D perspective, which is then used to present a simulated, complex air traffic scenario for further assessment.

8. ACKNOWLEDGEMENT

We would like to thank the participants for taking part in the experiment. This study is funded by the German Federal Ministry of Economics and Technology (BMWI) within the joint research project "innovative Airport (iPort)".

9. REFERENCES

- [1] Alexander, T., Winkelholz, C., and Krüger, W., Untersuchung der Einsatzmöglichkeiten von VU-Technologie in der taktischen Lagedarstellung, FGAN e.V. Forschungsinstitut für Kommunikation, Informationsverarbeitung und Ergonomie, Wachtberg (Germany), 2002.
- [2] T. Alm, and P. Lif, "The Value of Spatial Cues in 3D Air Traffic Displays," The International Journal of Aviation Psychology 17 (2), Lawrence Erlbaum Associates, Mahwah (New Jersey, USA), pp. 109-129, 2007.
- [3] A.D. Baddeley, "The episodic buffer: A new component of working memory?," Trends in Cognitive Sciences 4(11), Elsevier, Amsterdam (Netherlands), pp. 417-423, 2000.
- [4] Baier, A., Analyse des Radarlotsenarbeitsplatzes, EADS Military Air Systems, Manching (Germany), 2010.
- [5] M.R. Endsley, "Design and evaluation for situation awareness enhancement," Proceedings of the Human Factors Society 32nd Annual Meeting, Human Factors Society, Santa Monica (California, USA), pp. 97-101, 1988.
- [6] M. Goebel, J. Stallkamp, J. Springer, and H. Luczak, "3D-Radar Display for Air-Traffic-Control-Tasks," WWDU, A. Grieco, G. Molteni, E. Occhipinti, and B. Piccoli (eds), pp. F22-F24, Milano (Italy), 1994.
- [7] S.G. Hart, and L.E. Staveland, "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research," Human mental workload 1, North Holland Press, pp. 139-183, Amsterdam (Holland), 1988.
- [8] D.J. Law, J.W. Pellegrino, S.R. Mitchell, S.C. Fischer, T.P. McDonald, and E.B. Hunt, "Perceptual and cognitive factors governing performance in comparative arrival-time judgments," Journal of Experimental Psychology: Human Perception and Performance 19 (6), American Psychological Association, Vancouver (British Columbia, Canada), pp. 1183–1199, 1993.
- [9] Lipton, L., Foundations of the stereoscopic cinema. A study in depth, Van Nostrand Reinhold, New York (New York, USA), 1982.
- [10] Milger, P., Geschichte der Flugsicherung in Deutschland, Books on Demand, Norderstedt (Germany), 2008.
- [11] Parrish, R.V., Determination of depth-viewing volumes for stereo three-dimensional graphic displays, NASA Langley Research Center, Hampton (Virginia, USA), 1990.
- [12] Prümper, J., and Anft, M., Die Evaluation von Software auf Grundlage des Entwurfs zur internationalen Ergonomie-Norm ISO 9241 Teil 10 als Beitrag zur partizipativen Systemgestaltung ein Fallbeispiel, Software-Ergonomie, Walter de Gruyter, Berlin (Germany), p. 145-156, 1993.
- [13] Prinzo, O.V., A.M. Hendrix, and R. Hendrix, The outcome of ATC message complexity on pilot readback performance, DOT/FAA/AM-06/25, Federal Aviation Administration, Washington (District of Columbia, USA), 2006.

- [14] Tavanti, M., and M. Cooper, Looking for the 3D picture. the spatio-temporal realm of student controllers, University of Linköpings, Linköpings (Sweden), 2008.
- [15] M. Tavanti, H.H. Le, and N.T. Dang, "Three-dimensional stereoscopic visualization for Air Traffic Control Interfaces: A preliminary study," IEEE The 22nd Digital Avionics Systems Conference, Vol. 1, Piscataway (New Jersey, USA), pp. 5.A.1-1-7, 2003.
- [16] K.F. Van Orden, and J.W. Broyles, "Visuospatial task performance as a function of two- and three-dimensional display presentation techniques," Displays 21 (1), Elsevier, Amsterdam (Netherlands), pp. 17-24, 2000.
- [17] Wickens, C.D., Display Integration of Air Traffic Control Information: 3D Displays and Proximity Compatibility, Institute of Aviation, University of Illinois (Illinois, USA), 1995.
- [18] Wickens, C.D., S. Todd, and K. Seidler, Three-dimensional displays: Perception, implementation, and applications, Institute of Aviation, University of Illinois (Illinois, USA), 1989.