Nationaal Lucht- en Ruimtevaartlaboratorium

National Aerospace Laboratory NLR

Executive summary

Insight: Assisting aviation display designers by measuring visual clutter

Problem area
Early identification of display design issues can be a real cost saver. Design support is essential because the cockpit design process is complex and knowledge-intensive. A potential display design issue is visual clutter. The amount of information humans can process and store is limited. Visual clutter may significantly decrease human performance and should, therefore, be identified and accounted for by the display designer.

Description of work
The NLR (National Aerospace Laboratory NLR, Amsterdam) has developed a human factors assessment tool called Insight, which assists the designer in the development process of cockpit applications. The tool supports the designer in choices related to human factors aspects of the design, and can be used for development of new cockpit display applications or in a trade-off study between multiple display design solutions. The tool has been improved by adding the option to objectively measure clutter on a display. This enables designers to asses display clutter without extensive pilot evaluations. This clutter measurement technique was evaluated by an experiment conducted with airline pilots. The goal of this experiment was to compare pilot clutter evaluations with the Insight clutter measures in order to assess the performance of the applied clutter measurement technique.

Report no.
NLR-TP-2010-044

Author(s)
R.R.D. Arents
G.K. van de Merwe
R.P.M. Verhoeven
G.D.R. Zon

Report classification
UNCLASSIFIED

Date
March 2010

Knowledge area(s)
Training, Simulatie en Operator Performance
Cockpit

Descriptor(s)
Clutter
Insight
Vincent
Human factors
Display design

This report is based on a presentation held at the Avionics Europe 2010 Conference, Amsterdam, the Netherlands, March 24th - 25th, 2010.

Results and conclusions

The experiment showed that the Insight clutter measurement can be used to assist display designers during the design process. The clutter map that is created by Insight reveals high and/or low clutter levels within the image and this information can be used to tune various display properties like, symbology, colouring, and placement of information. The scalar outcome of the clutter measurement does need some improvement before it can be used during display design. The measurement does not take into account the effect of local extremes in clutter, while these could significantly affect the pilot's perception of clutter. This can be solved by adapting the algorithm to represent these local extremes in its scalar measurement. Recommendations for improvement are suggested in this paper.

Applicability

The Insight tool with clutter measurement can be used to assist display designers during the design process. This allows an early identification of possible design issues even before conducting costly experiments. The proposed technology may be applied for cockpit display application development in research and development, and industry.

Nationaal Lucht- en Ruimtevaartlaboratorium, National Aerospace Laboratory NLR

Anthony Fokkerweg 2, 1059 CM Amsterdam,
P.O. Box 90502, 1006 BM Amsterdam, The Netherlands
Telephone +31 20 511 31 13, Fax +31 20 511 32 10, Web site: www.nlr.nl

Nationaal Lucht- en Ruimtevaartlaboratorium

National Aerospace Laboratory NLR

NLR-TP-2010-044

Insight: Assisting aviation display designers by measuring visual clutter

R.R.D. Arents, G.K. van de Merwe, R.P.M. Verhoeven and G.D.R. Zon

This report is based on a presentation held at the Avionics Europe 2010 Conference, Amsterdam, the Netherlands, March 24th – 25th, 2010.

Customer	National Aerospace Laboratory NLR
Contract number	----
Owner	National Aerospace Laboratory NLR
Division NLR	Air Transport
Distribution	Unlimited
Classification of title	Unclassified
	March 2010

Approved by:

Author	Reviewer	Managing department
	22/2/10	

Summary

Early identification of display design issues can be a real cost saver. Design support is essential because the cockpit design process is complex and knowledge-intensive. A potential design issue is visual clutter. The amount of information humans can process en store is limited. Visual clutter may seriously decrease human performance and should, therefore, be identified and accounted for by the display designer. The NLR (National Aerospace Laboratory NLR, Amsterdam) has developed a human factors assessment tool called Insight, which assists the designer in the development process of cockpit applications. The tool supports the designer in choices related to human factors aspects of the design, and can be used during development of new cockpit display applications or in a trade-off study of multiple display design solutions. The tool has been improved by adding the option to objectively measure clutter on a display. This enables designers to assess display clutter without extensive pilot evaluations. This clutter measurement technique was evaluated using an experiment conducted with airline pilots. The goal of this experiment was to compare human clutter evaluations with the Insight clutter measures in order to assess the performance of the applied clutter measurement technique. The experiment showed that the Insight clutter measurement can be used to assist display designers during the design process. The clutter map that is created by Insight reveals high and/or low clutter levels within the image and this information can be used to tune various display properties like, symbology, colouring, and placement of information. The scalar outcome of the clutter measurement does need some improvement before it can be used during display design. The measurement does not take into account the effect of local extremes in clutter, while these could significantly affect the pilot's perception of clutter. This can be solved by adapting the algorithm to represent these local extremes in its scalar measurement or by indicating the clutter variance over the display.

Contents

1 Introduction

Cockpit display design support is needed because the cockpit design process is complex and knowledge intensive, and there is a limit to the amount of information human designers can process and store. Rudimentary design support can be provided by a prototyping tool that supports early demonstration of new design concepts. The level of support can be increased by also assisting the designer whilst constructing the prototype. The NLR (National Aerospace Laboratory NLR, Amsterdam) has developed, as a partner in the EU-funded IMCAD project (IMCAD, 2002), a human factors assessment tool called Insight that assists the designer in the development process of a cockpit application. The tool supports the designer in choices related to human factors aspects of the design, and can be used during development of new cockpit display applications, or during modification of displays which have already been certified for use. In the EU-funded HILAS[1] project, the design support tool Insight has been improved to also provide assistance on the subject of visual clutter. The tool offers the option to objectively measure clutter on a display, enabling designers to optimise display clutter.

2 Measuring clutter

2.1 Theory

"*Clutter is the state in which excess items, or their representation or organization, lead to a degradation of performance at some task*" (Rosenholtz et. al. 2005). This task can range from searching for a stapler on a messy desk, to safely navigating a commercial aircraft to the destination through a crowded airspace. Obviously, any decrease in performance due to clutter is very unfavourable in general, but especially for safety critical tasks like flying an aircraft it should be prevented. Designers of aircraft control systems and displays, therefore, take clutter as an important phenomenon that must be accounted for. Still, clutter remains a difficult design issue that historically has been evaluated using common human factor measures such as; subjective pilot (workload) ratings, opinion survey, etc. (Haworth & Newman, 1993). Also, there are guidelines that have been formulated (e.g. SAE ARP 5288; AC-25-11) in order to assist designers to prevent clutter in aviation displays (Kaber et. al. 2007). Although these guidelines and evaluation techniques have been proven very useful, it may be possible to offer more assistance to the designers of aviation displays. For example, more assistance could be provided if it were possible to reliably and objectively measure clutter in a display. Using such a measure, designers could optimise display clutter in an early stage of the design process. Recent

[1] The HILAS project ran from June 2005 until November 2009 and was funded by the European Communities as part of the 6th framework. http://www.hilas.info/mambo/

studies on measuring visual clutter (Rosenholtz et. al. 2007) indicate that this technique is on the verge of becoming feasible.

The Feature Congestion measure (Rosenholtz et. al. 2005) is based on the analogy that the more cluttered a display or scene is, the more difficult it would be to add a new item that would reliably draw attention. This analogy offers high potential of evolving into a reliable clutter measure for aviation displays. The Feature Congestion principle states: that the clutter in a local part of a display is related to the local variability in certain key features (Rosenholtz et. al. 2007). The key features that are currently implemented are:

- Colour
- Luminance
- Orientation

Other features that are basic features in visual search and attention may be included in future implementations (Rosenholtz et. al. 2007). The measurement uses the following steps (for more information see Rosenholtz et. al. 2007):

1. Determine the feature covariance
   a. The input image is converted to the perceptual base CIELab colour space (C.I.E. 1978). In this colour space the following three coordinates are described; the lightness of a colour (L); its position between red/magenta and green (a); and its position between yellow and blue. This colour space is developed to approximate human vision unlike RGB for which the colours red, green and blue are added.
   b. Multiple scales, or Gaussian pyramids, are created by alternately smoothing and subsampling the image (Burt & Adelson, 1983). This creates a ‘pyramid’ of images with the original at the bottom and stacked with sequential images with half the resolution of the one beneath.
   c. Image features are located for each scale or Gaussian pyramid. For luminance contrast a form of “contrast energy” is computed by filtering the luminance band with a center-surround filter that is obtained from the difference of two Gaussians and squaring the input. For colour, a local mean colour is extracted at each scale by pooling with a Gaussian filter. For orientation, the oriented opponent energy (Bergen and Landy 1991) is computed. For each feature that is found, the local (co)variance is computed.
2. Combine across scales: For each feature, take the maximum at each pixel over the Gaussian pyramids.
3. Combine across features: Take the cube root of the colour clutter (a volume) and the square root of the orientation clutter (an area) to be able to compare these with the luminance

clutter (a scalar). Then scale the image to how much feature space is available (Rosenholtz et. al. 2007).

4. Combine across space: Take the average clutter value over the entire image to obtain a single clutter measure (Rosenholtz et. al. 2007)

More research is needed to validate these steps, and especially the three combination steps (2-4) will probably need adjustment to obtain a more reliable clutter measure. Still, the current implementation of the measurement that is based on simplified assumptions has been demonstrated to work well (Rosenholtz et. al. 2007).

2.2 Insight Clutter Measurement Implementation

The clutter measurement theory by Rosenholtz was incorporated in the Insight Design-Time Assistant. This tool now has the option to evaluate clutter on a(n) (avionics) display in the Vincent[2] prototyping format or using a bitmap representation of the display. As discussed in the previous paragraph, the clutter can be measured for different features of which, colour, luminance and orientation were identified (Rosenholtz et. al. 2007) and which are implemented in the Insight assistant. Display designers that use the Insight assistant to develop displays can select to either evaluate clutter for each individual feature or for all features combined. The result is presented in a scalar value that represents the total clutter measurement (for each feature or combined) through the display and an image that provides a local clutter indication by depicting relatively higher clutter levels with white colours and lower levels with black.

3 Experiment

The new clutter measure implementation to the Insight design time assistant tool was evaluated using an experiment within the HILAS project. The goal was to obtain indications on how the pilot evaluations and the Insight clutter measure differ to enable future improvement of the clutter measurement.

3.1 Method

3.1.1 Subjects and instructions to subjects

Fourteen commercial airline pilots participated in the experiment. They were provided with the definition of clutter and asked to examine the images closely as the variation between them

[2] http://vincent.nlr.nl

would be relatively small. They were informed that each image will be shown separately and in random order, before showing an accumulated picture of all four images in the same sequence.

3.1.2 Apparatus

The navigations displays had a resolution of 800 by 800 pixels and were shown at sequentially full resolution on a 17-in. LCD screen with a resolution of 1280 by 1024 pixels. The same LCD screen was used to present the accumulated image, however, it was not possible to present each image at full resolution, and therefore a downscaled composition was used which corresponded to the LCD's resolution.

3.1.3 Independent variables

One independent variable was manipulated in the experiment: the navigation display. To start, a standard navigation display was taken and the aircraft's situation that was depicted was a relatively busy airspace regarding traffic and weather (see Figure 1). This serves as a reference display upon which the other three displays are based. There were two ways in which this reference is varied. First, the aircraft operating environment is changed by placing the traffic and weather closer to the own ship and to present more dense information (see Figure 3). This variant will be referred to as "CLOSER". The second way to vary the reference situation was to change the presentation of the information itself by enlarging the aircraft symbols on the display (see Figure 2). This variant will be referred to as "BIGGER". The fourth variation was obtained by combining these two variants to create the "BIGGER & CLOSER" variant (Figure 4).

By choosing these types of variations, the amount of information is kept constant between the four displays. The number of items that is presented does not change, but the airspace may be perceived more crowded and/or cluttered when the items are more densely packed together or when they appear to be bigger by themselves. The variation between the four displays was deliberately kept small, with the intention of keeping the clutter variations small as well and thus testing the clutter methodology to its full potential.

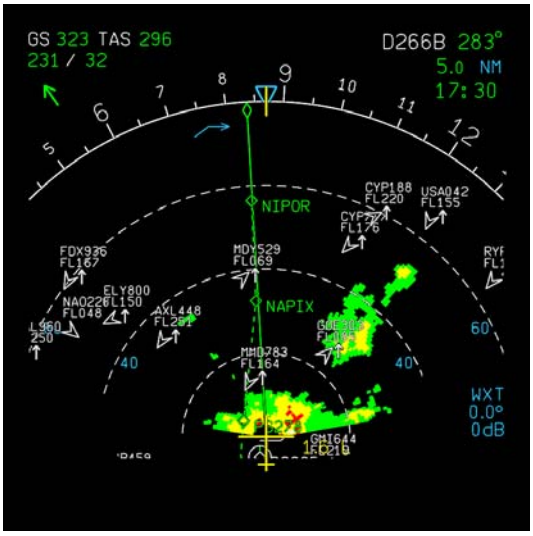

Figure 1: Reference navigation display. REF

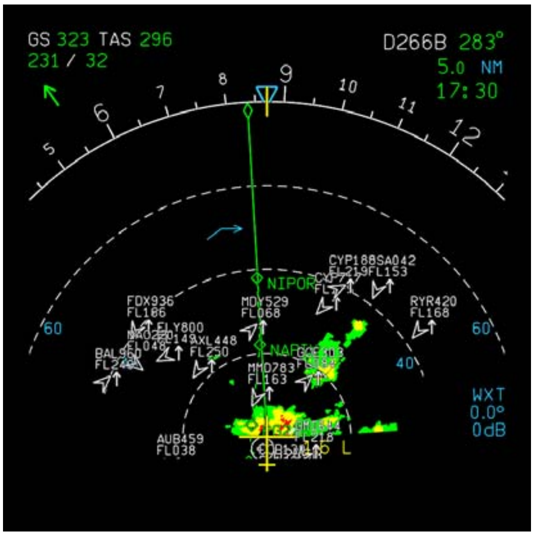

Figure 3: Navigation display for which traffic and weather are located closer to the own ship. CLOSER

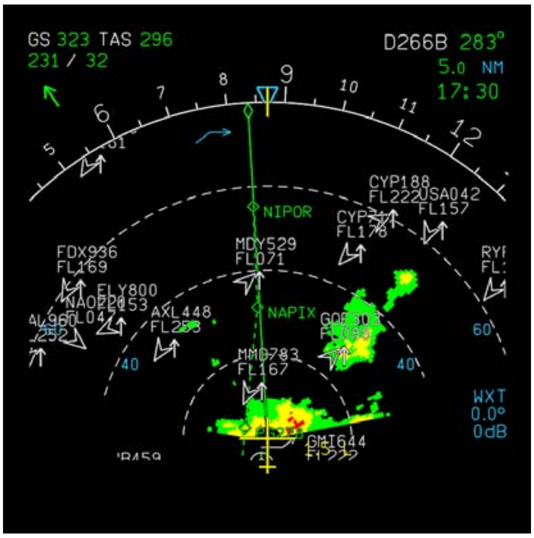

Figure 2: Navigation display with bigger aircraft symbols. BIGGER

Figure 4: Navigation display with bigger aircraft symbols, with traffic and weather located closer to the own ship. BIGGER & CLOSER

3.1.4 Experiment design

The displays were shown sequentially and in a random order to each individual pilot. After these individual images were shown, an accumulated image that contained all four displays was presented. The same random order was used for this accumulated image as the individual images were shown.

3.1.5 Procedure
The participants were shown the four different navigation displays in a sequential fashion, before presenting an accumulated image of the four displays. After showing the accumulated image, they were asked to compare the displays with respect to the perceived amount of clutter and rate each display on a scale from 1 – low level of clutter to 10 – highly cluttered.

3.1.6 Dependent measures
The dependent measures were the pilot clutter ratings of the displays and the clutter measurement that was obtained using the Insight design time assistant.

3.1.7 Experiment Hypothesis
It was hypothesized that the Insight clutter measurement is a good representation of a pilot's perceived amount of clutter in aviation displays.

4 Results and Discussion

The main results of the experiment are summarized in this section

4.1 Dependent measures
Before presenting the displays to the pilots for clutter evaluation, clutter is measured by Insight. The graphical results of this measurement for all features combined (colour, luminance and orientation) is presented in Figure 5 to Figure 8. Notice that the scalar outcome of the clutter measure is shown in the captions of these figures.

Figure 5: Full clutter map of the reference display. REF: Clutter = 11.74

Figure 6: Full clutter map of the display with bigger symbols. BIGGER: Clutter = 12.82

Figure 7: Full clutter map of the display for which traffic and weather are located closer to the own ship. CLOSER: Clutter = 11.42

Figure 8: Full clutter map of display with bigger symbols, and with traffic and weather located closer to the own ship. BIGGER & CLOSER: Clutter = 12.13

As stated, the pilots were asked to rate clutter for each display on a scale from 1 (low level of clutter) to 10 (high level of clutter). The means and standard deviations of these pilot ratings are illustrated in Figure 9 ($F_{3,52} = 19.054$, $p \leq 0.01$). It can be observed that the pilots significantly rate the clutter level for the reference (REF) display as lowest, while the last display (CLOSER & BIGGER) is rated highest. (Student-Newman-Keuls (SNK), $\alpha = 0.05$).

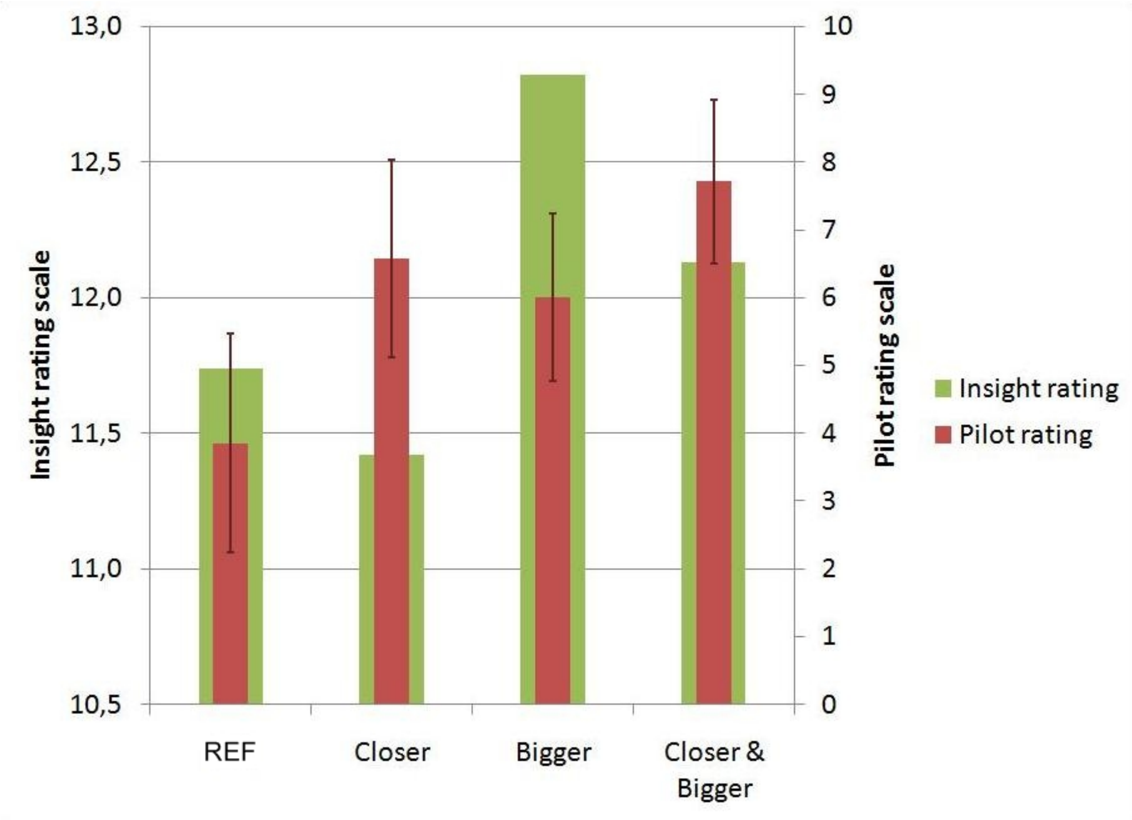

Figure 9: Means and standard deviations of the pilot clutter ratings and the Insight rating

The pilots were asked to rate the displays in order to allow a comparison between the display and to possibly say that one is x times as cluttered as the other. However, the standard deviations in the results were not distinct enough to make such a statement. When the rating for each subject is normalized and converted into a ranking, the means and standard deviations of the ranking provide a more distinct result. In Figure 10 the means and standard deviations of this pilot display ranking are presented.

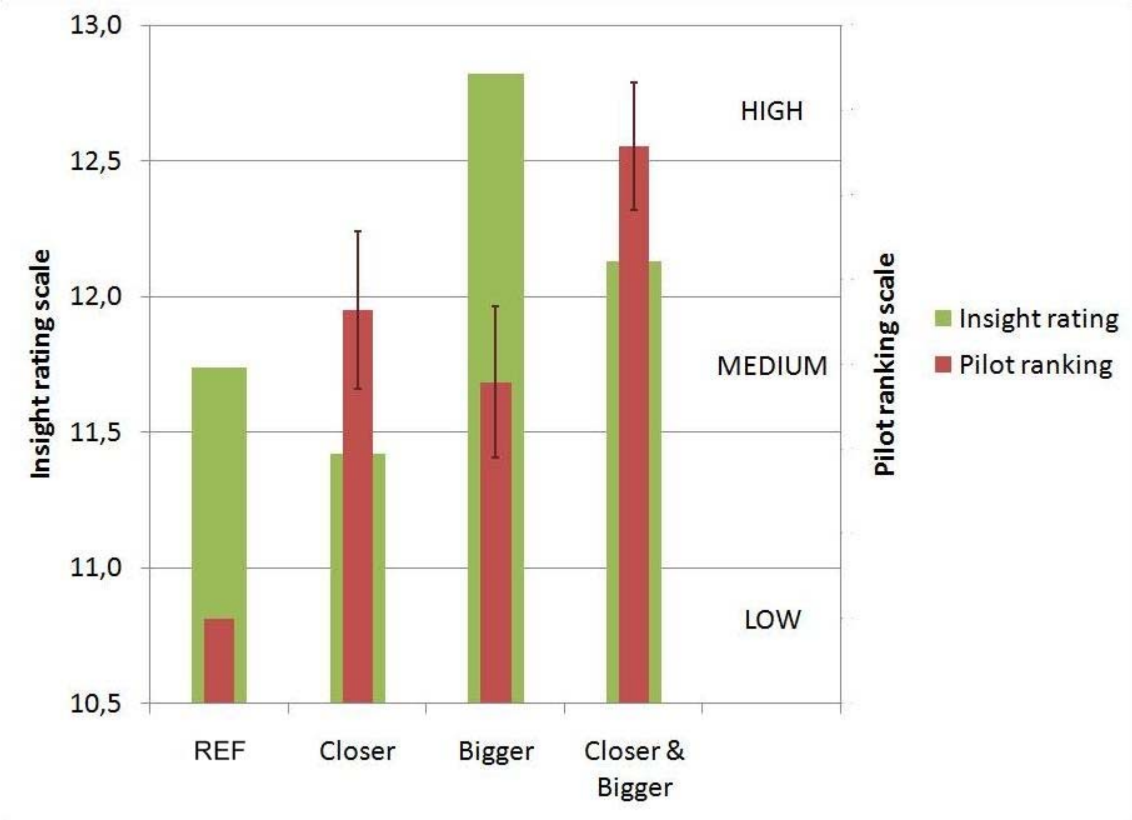

Figure 10: Means and deviations of the pilot display ranking with respect to clutter and the Insight clutter rating

In both Figure 9 and Figure 10, the pilot evaluations are plotted together with the Insight total clutter measurement. This allows comparison of the results which is necessary to reach the goal of this experiment.

The first obvious difference is that the pilots evaluate the clutter for the reference display as lowest, while Insight evaluates the CLOSER display as lowest. The difference between the two is that the CLOSER variant displays the same amount of information on a smaller surface that is furthermore located closer to the own aircraft's position. This increases the information density in the main location of interest, while the surroundings of this display are relatively empty. A similar difference can be noticed when comparing the BIGGER with the CLOSER & BIGGER display. Insight clearly evaluates the display that is locally denser (same average amount of information) as less cluttered, while pilots clearly experience more clutter. Still, when examining the full clutter maps of these two pairs of displays (Figure 5 and Figure 7, Figure 6 and Figure 8), it is clear that Insight has measured the local higher clutter levels on the two CLOSER display variants. For these two displays, the white levels in the measurement are higher which represents the (locally) high clutter levels. On the other hand, the Insight clutter measurement scalar value does not represent this high clutter level. The reason for this probably lies in the way the scalar clutter measurement is obtained from the full clutter maps. As mentioned in paragraph 3.1, the measurement is combined across space by taking the average of the full clutter map, which intrinsically can result in the loss of certain local maxima or minima.

Secondly, when comparing the REF with the BIGGER display and the CLOSER with the CLOSER & BIGGER display there are some apparent similarities between the Insight rating and the pilot rating/ranking. Although there are some scale differences, the trend within these two pairs is similar and the clutter level increases when the traffic symbols are enlarged. When comparing the clutter maps of these two pairs (Figure 5 with Figure 6, and Figure 7 with Figure 8), there are few differences. Apparently, there are not many local differences in clutter, but overall clutter increases as traffic symbols are enlarged.

4.2 Discussion on Experimental Results

Insight produces a clutter map that represents higher clutter levels within an image with a white colour with higher luminance on a black background (see Figure 5 to Figure 8), and a scalar measurement that represents the clutter over the entire image or display. In this experiment, the differences between the experimental conditions were deliberately kept small to test the algorithms to their full potential. It may be argued that the differences were too small for the current clutter measurement algorithm. Future experiments may benefit from trials with more obvious differences in order to improve the algorithm. The results from the experiment indicate that the scalar measurement can not be used when the distribution of information items across a display is changed. Placing the traffic items closer to the own ship results in a lower clutter measure, while pilots experience more clutter. When the distribution is kept constant, however, the scalar measurement does seem to provide a good representation of the clutter as experienced by pilots. Distribution of the information across the display seems to be an important factor for improving the fidelity of the clutter measurement in Insight. The calculation of the final scalar clutter measurement is done by averaging the results from the clutter map. The clutter map contains information on the uneven distribution of the clutter levels within the image, but averaging this information upon obtaining the scalar measurement loses this while it clearly affects the pilot’s judgment. A more advanced algorithm may be needed to produce the final clutter measurement while taking into account the presence of local extremes.

5 Conclusions and recommendations

Conclusively, Insight clutter measurements can be used to support display designers. The cluttermap created by Insight reveals high and/or low clutter levels within the image for a feature by itself and combined. With this information various display properties can be tuned such as; symbology, colouring, placement of information, etc. The scalar outcome of the clutter measurement does need some improvement before it can be used during display design. The measurement does not take into account the effect of local extremes in clutter, while these can significantly affect the pilot's perception of clutter. Adapting the combination across space algorithm, which is part of the measurement technique, to represent these local extremes in its outcome could solve this. Determining the variation in clutter next to the clutter average over the image may be needed to accurately represent the amount of clutter on a display. This experiment did not directly reveal a need for improvements in the combination across scales and/or features algorithm within the measurement technique, nor for the extension of the number of analyzed features, but future investigations into this matter seem valuable.

References

Bergen, J. R., & Landy, M. S. (1991). Computational modelling of visual texture segregation. In M. S. Landy & J. A. Movshon (Eds.), *Computational models of visual processing* (pp. 253–271). Cambridge, MA: MIT Press.

Burt, P., & Adelson, E.H., (1983). The Laplacian pyramid as a compact image code. *IEEE Transactions on Communication*, COM-31, 532-540

CIELAB (1976). Official recommendations on uniform colour spaces, colour differences equations and metric colour terms. *Suppl. 2 to CIE Publications n. 15. Commission Internationale de Eclairage, Colorimetry*. Paris, France.

Haworth, L.A., & Newman, R.L., (1993). Test Techniques for Evaluating Flight Displays. Washington, D.C.: NASA TM-103947.

IMCAD (2002), Improving the Cockpit Application Development Process, RTD project funded by the EU, project nr GRD1-2001-40210.

Kaber, D., Alexander, A., Stelzer, E., Kim, S., Kaufmann, K., Cowley, J., Hsiang, S. & Bailey, N., (2007). Testing and validation of a psychophysically defined metric of display clutter. NASA Langley Research Center. Grant number: NNL06AA21A.

Hanson, E., Braakhuis, J. & Küstner, R., (2004) IMCAD design time assistance. National Aerospace Laboratory NLR & Diehl Avionik Systeme

Rosenholtz, R., Li, Y., & Nakano, L. (2007). Measuring visual clutter. *Journal of Vision, 7*(2):17, 1-22, http://journalofvision.org/7/2/17/, doi:10.1167/7.2.17

Rosenholtz, R., Li, Y., Mansfield, J., Jin, Z., (2005) Feature Congestion: a measure of display clutter. *SIGCHI* 2005, 761-770, 2005.