2009

Modeling the Effects of HUD Visual Properties and Configurations on a Multi-Dimensional Measure of Clutter

Sang-Hwan Kim

David B. Kaber

Karl Kaufmann

Theo Veil

Amy L. Alexander

See next page for additional authors

Authors
Sang-Hwan Kim, David B. Kaber, Karl Kaufmann, Theo Veil, Amy L. Alexander, Emily M. Stelzer, and Lawrence J. Prinzel

MODELING THE EFFECTS OF HUD VISUAL PROPERTIES AND CONFIGURATIONS ON A MULTI-DIMENSIONAL MEASURE OF CLUTTER

Sang-Hwan Kim, David B. Kaber, Karl Kaufmann & Theo Veil
Industrial & Systems Engineering, North Carolina State University
Raleigh, NC
Amy L. Alexander & Emily M. Stelzer
Aptima, Inc.
Woburn, MA
Lawrence J. Prinzel
NASA Langley Research Center
Hampton Roads, VA

The objectives of this study were to: validate a multidimensional measure of display clutter for advanced head-up displays (HUDs) incorporating enhanced and synthetic vision systems (EVS/SVS); assess the influence of HUD configuration on pilot perceptions of display clutter and flight performance; and model clutter scores in terms of visual display properties. Eighteen pilots flew a flight simulator in a landing approach using three different sets of HUD configurations (low, medium, or high clutter). Pilot ratings of overall display clutter and its underlying dimensions were recorded along with flight performance measures (deviations from localizer, glideslope, altitude and approach speed). A display image analysis software application was used to measure the visual properties of HUDs. The multidimensional measure of clutter showed internal consistency with high correlations of overall perceived clutter. Calculated clutter scores were sensitive to the various HUD configurations and in agreement with prior display classification (based on the new measure). There was a trend for the extremes of display clutter to cause less stable performance. High clutter displays were associated with cognitive complexity of flight tasks and low clutter was associated with a lack of display information. Multiple linear regression models of perceived clutter were developed based on HUD visual properties.

Future concepts for the National Airspace System integrate cockpit technologies to support flight safety through improved terrain and traffic awareness. A subset of these technologies includes synthetic and enhanced vision systems (SVS and EVS) for pilot use (Bailey et al., 2002). The goal of EVS and SVS displays in the aircraft cockpit is to reduce the incidence of low visibility accidents (Prinzel et al., 2002); however, the design of such display may obscure other important information features when integrated in existing head-down displays (HDDs) and/or HUDs. While these new technologies provide pilots access to information that may not be visible with traditional flight instrumentation, the presentation of this additional information in HUDs may serve to produce visual display clutter (Ververs & Wickens, 1998; Prinzel & Kramer, 2006). Kaber et al. (2008) investigated the effect of HUD features, including EVS, SVS, "highway-in-the sky" (tunnel), traffic collision avoidance system (TCAS) icons and the amount of basic flight information, on experienced pilot perceptions of clutter. Clutter was defined as the presence of irrelevant information (or obscuration of relevant information) in a display. They found that the greater the number of information features and visual density, the higher perceived clutter ratings. Kaber et al. (2008) also conducted a psychological decomposition of the phenomenon of clutter in terms of underlying HUD display qualities. They identified a concise language (set of terms) that expert pilots find useful to assess aviation display clutter, including "redundant/orthogonal," "monochromatic/colorful," "salient/not salient," "safe/unsafe," and "dense/sparse." In addition to this, Alexander et al. (2008) identified bottom-up (data-driven) and top-down (knowledge-driven) factors in pilot perception of display clutter. These studies provided a basis for defining aviation display clutter in terms of display information features and perceived display qualities, as well as developing a multidimensional measure of display clutter. The objectives of the present research were to build upon this initial study by: (1) defining and validating a multidimensional measure of display clutter; (2) assessing the influence of HUD clutter classifications on pilot perceptions of displays and actual flight performance; and (3) modeling subjective HUD clutter scores in terms of objective visual properties, including luminance, target-to-background contrast, feature occlusion, and visual density.

Methodology

Development of Multidimensional Measure of Display Clutter

The set of display descriptor terms selected by pilots in the Kaber et al. (2008) study were used as anchors in a collection of bipolar subjective rating scales covering the underlying dimensions of clutter, including "redundancy (orthogonal/redundant)", "colorfulness (monochromatic/colorful)," "feature salience (salient/not salient)," "feature dynamics (static/dynamic)," "feature variability (monotonous/variable)," and "global density (sparse/dense)." The scales were integrated into an overall clutter index, which required pilots to rank the importance of each dimension for characterizing HUD clutter (in context) and to rate displays on each scale. The ranking of clutter dimensions and the ratings for display were then combined in an overall clutter score (rank-weighted sum of ratings across dimensions). This measurement approach was very similar to the design of measures examining other psychological phenomenon, such as the NASA-Task Load Index (TLX) (Hart & Staveland, 1988) for assessing cognitive workload.

HUD Configurations

Thirty-two different HUD configurations studied by Kaber et al. (2008) were rank-ordered based on measures and predictions of expert pilot ratings of clutter. For the present study, the top 20% of HUDs were classified as "high-clutter," the middle 20% as "medium clutter," and the lowest 20% as "low clutter." From each of these groups, three target displays were selected to represent unique HUD feature sets within each group for a total of nine test displays. Figure 1 shows the nine HUD configurations selected across the three clutter groups.

Participants and Experiment Design

Eighteen current line-pilots with varying levels of flight experience (six with <5 yrs.; six with 5-15 yrs.; six with >15 yrs.) but no HUD experience participated the experiment. They were asked to fly the Integration Flight Deck (IFD) simulator at NASA Langley on a landing approach under low and high workload conditions (no wind or a substantial crosswind, respectively). The IFD was used to present pilots with the standard instrument landing system (ILS) approach to Runway 16R at Reno-Tahoe International airport. Each approach was divided into three segments, including: (1) initial approach fix (IAF) at the PYRAM intersection to glideslope (G/S) intercept; (2) G/S intercept to just inside the final approach fix (FAF; DICEY); and (3) from the end of the preceding segment to decision height (DH), either as published or EVS minimums (as appropriate for the HUD configuration; FAR 91.175c). After completing a training period of basic airwork and an approach under visual conditions, participants completed six test trials. During the test trials, each pilot was presented with a different set of HUD configurations, representing "low," "medium" or "high" levels of clutter. Thus, two between-subject variables (three levels of pilot experience and three levels of display clutter) and two within-subject variables (two levels of flight workload and three segments of flight) were manipulated in the experiment.

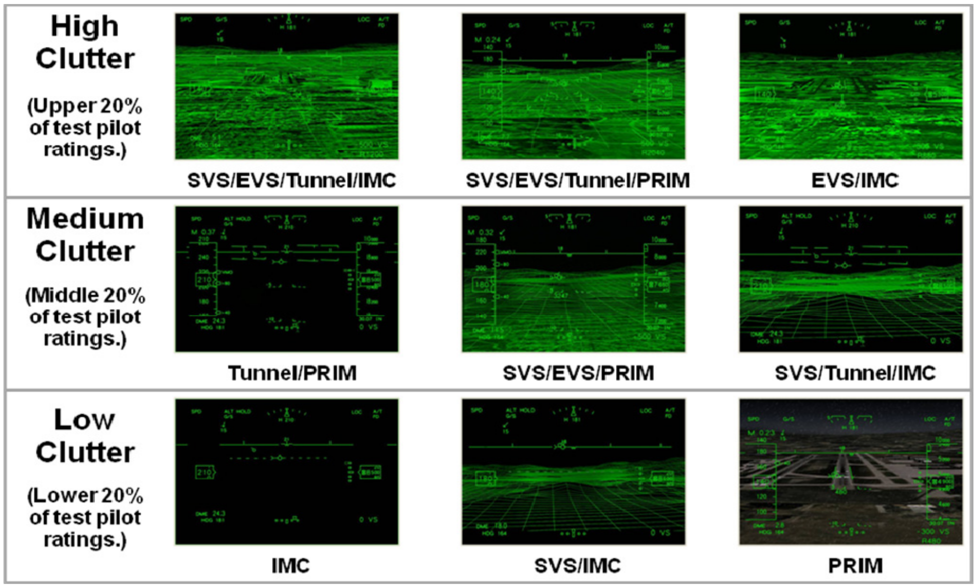

Figure 1. *Nine HUD configurations used in the experiment (IMC = Reduced symbol set for instrument meteorological conditions; PRIM = Complete primary flight display symbol set; and TUNNEL = Highway-in-the-sky guidance).*

Dependent Measures

Pilots subjective ratings of overall perceived display clutter and ratings on the underlying dimensions of clutter were collected at the end of each segment of flight for a trial. We also recorded pilot performance, including localizer (LOC) and glideslope (G/S) deviations and flight altitude and speed control. In addition to this, we developed a display image analysis software application to measure objective visual properties including average contrast of iconic (e.g., airspeed and altitude) to non-iconic (SVS/EVS) features, percent occlusion of iconic features by non-iconic features, and percent overall density for all HUD configurations in advance of the experiments. We also measured the lumens for each HUD in the IFD using a hand-held photometer.

Results

Multidimensional Measure of Display Clutter

Correlation analyses on ratings on the dimensions of clutter with overall perceived display clutter revealed significant positive linear relations for "colorfulness," "dynamics," "variability," and "density". However, "redundancy" and "saliency" were found to be negatively related with overall clutter. Table 1 shows the Pearson coefficients for the subscale and overall clutter ratings and the anchors of the various scales. On this basis, an overall clutter index was formulated by integrating the pilot ratings across all scales (including reversing ratings on "redundancy" and "saliency" (for which maximum scale values were given negative ratings)). Calculated clutter scores were generated by multiplying pilot rankings of dimensions with ratings for each HUD and were found to be highly correlated with overall perceived display clutter ratings (r = 0.77, p<0.0001). These results indicated that the new multidimensional measure of clutter had internal consistency and was valid for describing pilot experiences of clutter.

Influence of HUD Configuration on Pilot Perceptions of Display and Flight Performance

An ANOVA on pilot ratings of display clutter revealed significant effects of pilot flight experience ($F(2,44) = 3.768$, $p= 0.031$) and HUD configuration ($F(2,44)= 5.043$, $p=0.011$). High experience pilots tended to be more sensitive to clutter than the medium and low experience pilots ($p<0.05$).The high clutter display group led to higher ratings than the low clutter group ($p<0.05$). These main effects, however, were moderated by a significant experience by display interaction ($F(4,44)=3.122$, $p=0.024$). The interaction analysis revealed that low experience pilots rated the low clutter displays as more cluttered than the medium and higher clutter displays, perhaps because these displays lacked critical relevant information. Alternatively, medium and high experience pilots provided higher ratings for medium and high clutter displays than low clutter displays (see Figure 2). This finding suggested that display clutter perceptions of higher experience pilots were more consistent with the predefined groups of HUDs in terms of clutter features.

Table 1. *Correlation of subscale with overall clutter ratings and descriptor terms used as scale anchors.*

Subscales	**Correlation with Overall Clutter Ratings**	**Descriptor Terms / Scale Anchor**	
		Lower Clutter	**Higher Clutter**
Redundancy	$r= -0.431$, $p<0.0001$	Orthogonal	Redundant
Colorfulness	$r= 0.237$, $p<0.0001$	Monochromatic	Colorful
Salience	$r= -0.185$, $p<0.0001$	Salient	Not Salient
Dynamics	$r= 0.567$, $p<0.0001$	Static	Dynamic
Variability	$r= 0.474$, $p<0.0001$	Monotonous	Variable
Density	$r= 0.856$, $p<0.0001$	Sparse	Dense

Pilot flight performance was assessed in terms of the degree of variability of flight path control relative to identified targets in various segments of the landing approach. Specifically the kurtosis (degree of centrality) of the distribution of LOC deviations was calculated for all flight segments and the distribution of G/S deviations was calculated in those segments following intercept to evaluate the effects of pilot experience, HUD configuration, and flight task workload. An ANOVA revealed pilot experience to be insignificant in any of the critical performance measures. An interaction of the HUD configuration and segment of flight was highly significant for G/S deviations ($F(2,45)=8.514$, $p=0.001$) and marginally significant for LOC deviations ($F(4,90)=2.130$, $p=0.083$). A main effect of display appeared for G/S deviation ($F(2,45)=3.533$, $p=0.038$) and LOC deviation ($F(2,45)=14.51$, $p<0.001$), but flight segment was only significant for the G/S deviations ($F(1,45)=6.254$, $p=0.014$). There was a clear trend for the extremes of display clutter to cause less stable performance (lower distribution kurtosis). Low clutter displays led to unstable vertical path control, attributable to a lack of critical information. The high clutter displays produced less stable performance, attributable to redundant flight information, and the majority of pilots commented in post-experiment interviews that the displays were "cognitively complex". In general, the medium clutter HUDs appeared to be superior to low and high clutter displays for performance across all other conditions. Figure 3 shows G/S deviation stability by HUD configuration and flight workload level. Differences in stability across HUD clutter levels were greater for the low workload condition. The same pattern of results also appeared for pilot control of airspeed. Regarding LOC deviations, the more information included in the display, the more stable path control was. This pattern of results also appeared for pilot control of altitude (MSL) on the segment prior to G/S intercept. For both measures (LOC and G/S deviations), pilot control variability appeared to decrease significantly inside the FAF as compared to between the G/S intercept and FAF, possibly due to concentration on vertical path deviations shortly before landing.

Modeling of HUD Clutter Scores in Terms of Visual Properties

Multiple linear regression models of clutter scores were developed based on the objective visual properties of the HUDs measured with the image analysis software. Two separate models were created; one for the low-workload (no crosswind) condition and one for the high-workload (crosswind) condition. It was expected that the crosswind condition would cause greater density of visual features in the HUD (e.g., the crosswind would drive the flight path marker group towards the distally located altitude and speed tapes) and for the pilots to have higher perceptions of clutter. The models for both workload flight conditions were significant ($R^2 = 0.33$, $p<0.0001$ for low workload; $R^2 = 0.18$, $p<0.0001$ for high workload) for predicting clutter index values and *t*-tests on the lumen, contrast, occlusion, and density parameters. Model parameter tests revealed display lumens, contrast and density to all be significant contributors ($p<0.05$) to the clutter score; however, occlusion was not significant. Table 2 shows the magnitude and directions of all model predictor terms along with *t*-statistics and significance levels for both the low and the high workload models.

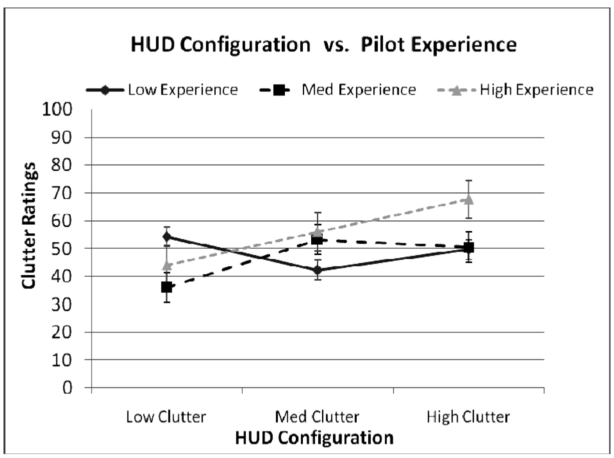

Figure 2. *Clutter ratings by HUD configuration and pilot experience.*

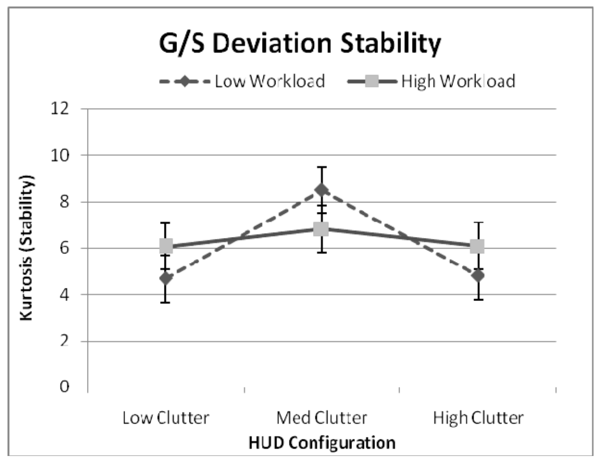

Figure 3. *G/S deviation stability by HUD configuration and flight workload.*

Table 2. *Regression analysis results for model of clutter score in visual display properties.*

Workload	Parameter	Estimate	Error	*t* Value	Pr > |*t*|
Low-workload (no crosswind) condition	Intercept	29.26	3.49	8.39	<0.0001
	Lumen	-12.29	3.05	-4.03	<0.0001
	Contrast	-10.38	2.92	-3.55	0.0005
	Occlusion	1.12	0.40	2.80	0.0057
	Density	15.69	3.91	4.01	<0.0001
High-workload (crosswind) condition	Intercept	39.77	3.54	11.23	<0.0001
	Lumen	-7.75	3.09	-2.51	0.0132
	Contrast	-8.70	2.97	-2.93	0.0039
	Occlusion	0.55	0.41	1.35	0.1785
	Density	9.87	3.97	2.48	0.0141

In general, decreases in display lumens and contrast and increases in occlusion and density caused increases in clutter scores. Average active pixel count (density) and average lumens of the HUDs appeared to account for the greatest amount of variability in the calculated clutter scores. Occlusion was not as strong a contributor as expected from the pilot subjective ratings. It is possible that this was due to a narrow stroke width of the SVS, EVS and tunnel features in the display configurations.

Discussion and Conclusion

The present study achieved the objectives of developing a new multidimensional measure of display clutter, which proved to be sensitive to manipulations of HUD configurations. This measure was correlated with overall ratings of perceived clutter, supporting construct validity. Experiment results indicate that pilot experience plays a role in perceptions of clutter, with high-time pilots being more accurate and consistent in judging the occurrence of clutter. Across workload and performance measures, negative effects of low and high clutter displays were observed, suggesting some optimal amount of HUD information may exist in terms of avoiding information overload while supporting flight path control. Software-based analysis of HUD images yielded visual property measures that were highly predictive of clutter. This indicated that pilot perceptions of clutter in new HUD designs could be projected based in part on low-level display characteristics. Top-down factors, such as the information feature content and task relevance also need to be considered in future extension of such models. One additional caveat of this research is that we did not use a full-motion flight simulator for the experiment; however, few pilots directly commented on the absence of kinesthetic cues or suggested a potential influence on clutter evaluations.

The results of this research, including the multidimensional measure of clutter, and model of perceived clutter in terms of visual display properties, are expected to be applicable for evaluation of a range of NextGen display concepts, beyond EVS/SVS HUDs. One direction of future work would be to apply the new multidimensional subjective measure of clutter for evaluating air traffic management support display technologies for the occurrence of clutter and to assess the reliability of the measurement outcomes.

Acknowledgement

This research was supported by NASA Langley Research Center under Grant No. NNL06AA21A. Randall Bailey and Lawrence Prinzel were the technical monitors. The opinions and conclusions expressed in this paper are those of the authors and do not necessarily reflect the views of NASA.

References

Alexander, A. L., Stelzer, E. M., Kim, S-H., & Kaber, D. B. (2008). Bottom-up and top-down contributors to pilot perceptions of display clutter in advanced flight deck technologies. In *Proceedings of the Human Factors and Ergonomics Society 52nd Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society.

Bailey, R. E., Parrish, R. V., Kramer, L. J., Harrah, S., & Arthur, J. J. (2002). Technical Challenges in the Development of a NASA Synthetic Vision System Concept. In *Proceedings of the North Atlantic Treaty Organization (NATO) Symposium on Enhanced and Synthetic Vision Systems*.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (Vol. 1, pp. 139-183): Elsvier, Amsterdam.

Kaber, D. B., Alexander, A. L., Stelzer, E. M., Kim, S-H., Kaufmann, K. & Hsiang, S. M. (2008). Perceived clutter in advanced cockpit displays: Measurement and modeling with experienced pilots. *Avia. Space & Env. Med.*, 79(11), BOI: 10.3357/ASEM.2319.2008.

Prinzel, L. J., Kramer, L. J., Comstock, J. R., Bailey, R. E., Hughes, M. F., & Parrish, R. V. (2002). NASA synthetic vision EGE flight test. In *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society.

Prinzel, L. J., & Kramer, L. J. (2006). Synthetic Vision Systems. In W. Karwowski. (Eds.), *International Encyclopedia of Ergonomics and Human Factors* (pp. 1264-1271), 2nd Ed., Talyor & Francis.

Ververs, P. M., & Wickens, C. D. (1998). Head-up displays: Effects of clutter, display intensity, and display location on pilot performance. *International Journal of Aviation Psychology,* 8(4), 377-403