Expert assessment of the cockpit crew information and control field

I I Greshnikov\textsuperscript{1,2}, T D Salnikov\textsuperscript{2} and A S Ivanov\textsuperscript{1}

\textsuperscript{1} Moscow Aviation Institute (National Research University), 4, Volokolamskoe shosse, Moscow, 125993, Russia
\textsuperscript{2} Federal State Unitary Enterprise State Research Institute of Aviation Systems, 7, ul. Viktorenko, Moscow, 125167, Russia

E-mail: vvanes@mail.ru

Abstract. This paper discusses the method of expert assessment based on the probabilistic approach of the paired comparisons method. This mathematical apparatus is the basis for special software integrated into the universal stand for prototyping the cockpit, which allows for a systematic expert assessment at this stand according to pre-selected criteria with the control of the consistency of the answers. Using this method, the task of determining the feasibility of displaying layers of synthetic and enhanced vision on the flying frame is solved, but in general, the method allows evaluating any elements of the displays and controls.

1. Introduction

The displays and controls of modern aircraft operate with a large amount of data. Accordingly, research is relevant on the optimal form of their presentation, ergonomic assessment, and selection of the best interface option, which allows to minimize the information load and improve the situational awareness of the pilot [1, 2]. A full-fledged ergonomic assessment, based on which the choice of the display option is made, consists of objective and expert components. In international practice, as a rule, both assessments are taken into account [3, 4]. This article is devoted to expert assessment.

One of the areas that increase the situational awareness of the pilot and reduce the load on the crew is the practice of displaying layers of synthetic and enhanced vision (SVS/EVS) on the flight display. These trends lead to the need for a more serious and qualitative approach to the selection of optimal variants of the primary flight display (PFD) when displayed over the layers SVS/EVS, and the feasibility of application of these layers on one or another phase of the flight, taking into account characteristics inherent in air transport aircraft.

The study relevance is reasoned by the importance of the task of synthesizing the optimal cockpit displays and controls for aviation, as one of the key factors in ensuring flight safety, as well as the lack of an effective methodology and software for conducting an expert assessment of displays and controls configurations.

The purpose of this work is to create and automate an expert assessment methodology that provides the determination of the ergonomic quality of PFD depending on the display/non-display of SVS and EVS, which would also be applicable in the analysis of display options in any other situations. To do this, it is necessary to solve the tasks of developing a methodology for expert assessments of the displays and controls, developing special software for conducting these
assessments, and integrating this software into the test environment of the universal prototyping bench (UPB).

2. Method of expert assessments
To form an expert assessment of the display options of the cockpit displays and controls to choose the optimal display option, it is proposed to conduct a questionnaire. The test subjects (pilots) assess the display options using a certain scale under the specified assessment criteria.

When forming the final ergonomic assessment for the considered options of the displays and controls, various mathematical tools can be used [5–7], starting with the arithmetic mean and ending with the analytic hierarchy process (AHP) and fuzzy logic.

Since the AHP is based on the scoring system and involves setting priorities (weighting factors) for the criteria in the context of this task, the following problem arises. If you have a large number of heterogeneous characteristics using weighting factors is undesirable, as there is a risk of improper use that may lead to results that are far from objective.

The AHP method allows evaluating the internal consistency of the comparison matrix, which can help in solving the problem of incorrect assignment of weights, but the inconsistency of the estimates can arise from several reasons. This can be a difference in the competence of experts, as well as the ambiguity of the object itself or the ambiguity of the criteria by which the object is evaluated. It is necessary to distinguish one from the other in order to be able to choose a more adequate way to overcome uncertainty—in one case it is necessary to change experts, in the other to think over the evaluation criteria more carefully.

The AHP method does not provide sufficient confidence to identify the source of the inconsistency. Also, the use of the AHP method can significantly increase the number of pairwise comparison procedures from the number \( n(n - 1)/2 \), and in our situation it is preferable to optimize the cost of the experts’ time.

The most promising approach is the probabilistic modeling of paired comparisons using the least squares method. This choice is reasoned by the fact that a person is much better at catching the difference between two objects than placing objects on the scale of the assessment attribute relative to each other [8, 9].

The approach in which the display options are evaluated by binary estimates, allows you to place the display options on the interval scale based on the values obtained using the Thurstone model, which allows you to determine how much one object of evaluation is better than another. It should be borne in mind that the experts do not conduct an assessment in a real flight situation, but on a simulator. Also, the display options themselves are a complex set of mnemonic symbols and they are evaluated several times according to different criteria. In view of the unnatural environment of the evaluation, as well as the complexity of the object seems optimal to provide the expert evaluation, by two objects at a time to reduce the complexity of evaluation, and, consequently, the likelihood of inaccuracies in the estimation. The Thurstone model we use (edited by Thorgerson) has been modified so that it is possible to assess the consistency of expert assessments, as well as to assess the uncertainty of each expert’s responses in a personalized way.

At the end of the experiment, the proposed method allows ranking the options displays so that the scale axis values of the criterion, which contain display options, were selected the most plausible.

Several display options are compared with each other in pairs, in total there are \( N \) ways to represent the options in pairs:

\[
N = \frac{(n * (n - 1))}{2}, \tag{1}
\]

where \( n \) is the total number of options. The pilot chooses from the two options presented the one that seems preferable to him concerning the selection criterion.
One pair of comparisons is denoted as $a_{ij}$. The two display options being compared are denoted as $i$ and $j$. Then if the $i$-th version of the display is more preferable than the $j$-th, then $a_{ij} = 1$, in the opposite case $a_{ij} = 0$. If the pilot can not give preference to one of the options, $a_{ij} = 0$.

When all $n$ pairs have been evaluated by the pilot, a table of size $n \times n$ is constructed, where $i$ is a row and $j$ is a column. Then this table is filled with values $a_{ij}$.

As a result of the survey of all pilots participating in the study, similar tables are obtained for each pilot. Subsequently, all the resulting tables are added together:

$$f_{ij} = \sum y(a_{ij}),$$

where $y$ is the pilot number and $f$ is the cell of the resulting sum table. The result is a table where the number of pilots is $h$. The resulting table executes the rule:

$$h = f_{ij} + f_{ji}.$$  

To get the probability ($p_{ij}$) that the display option $i$ in the pair $ij$ is considered more preferable than the display option $j$—you need to divide the value of the table $f_{ij}$ by the total number of respondents participating in the study

$$p_{ij} = \frac{f_{ji}}{h}.$$  

To find the necessary $s_i$ on the criterion axis ($s_i$ is the scale value for the $i$-th display option relative to the criterion under consideration), it is necessary to translate the probability table $p$ into the matrix of $z$-assessments (table 1).

Further, knowing the $z$-value for each $z_{ij}$, it is possible to calculate $s_i$ for all $i$ display options according to the equation [8]:

$$s_i = \frac{\sum_{i=1}^n z_{ji}}{n}.$$  

After a pilot study is conducted for all the considered criteria of assessment display options, including the general criterion, a matrix is constructed, where the rows are the display options, and the columns are the criteria. At the intersection of the $i$-th row and the $j$-th column, the previously calculated scale value of the $i$-th display option according to the $j$-th criterion is placed. The assessment of the general criterion is obtained the same way as the assessment of the other criteria according to the algorithm described earlier.

To calculate the weight of the criteria, i.e. the degree of consistency of the assessments for each special numbered criterion with the assessments for the general criterion, it is necessary to assess the distance between the vectors of these assessments $d_j$:

$$d_j = \frac{\sum_{i=1}^m cr_i - cm_i}{m} \cdot (u)^{-1},$$

where $m$ is the number of display options $cr_i$—the estimate of the display options in the interval scale obtained during the application of the Torgerson model according to the criterion under consideration, and $cm_i$—the same estimate obtained according to the general criterion, $u$ is a constant, equal to 5.14 (the maximum possible difference between $cr_i$ and $cm_i$). Then the weight of the $j$-th criterion is:

$$w_j = 1 - d_j.$$  

Having the weight of each criterion, you can calculate the final assessment for each display option:

$$S_j = \frac{\sum_{j=1}^n S_{ij} \cdot w_j}{n},$$
### Table 1. Table for converting probabilities to $z$-values. The $Z$-value is the result of applying the reference point function to the probability.

| $p$  | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 | 0.10 |
|------|------|------|------|------|------|------|------|------|------|------|
| $Z$  | -2.33| -2.05| -1.88| -1.75| -1.64| -1.55| -1.48| -1.41| -1.34| -1.28|
| $p$  | 0.11 | 0.12 | 0.13 | 0.14 | 0.15 | 0.16 | 0.17 | 0.18 | 0.19 | 0.20 |
| $Z$  | -1.23| -1.18| -1.13| -1.08| -1.04| -0.99| -0.95| -0.92| -0.88| -0.84|
| $p$  | 0.21 | 0.22 | 0.23 | 0.24 | 0.25 | 0.26 | 0.27 | 0.28 | 0.29 | 0.30 |
| $Z$  | -0.81| -0.77| -0.74| -0.71| -0.67| -0.64| -0.61| -0.58| -0.55| -0.52|
| $p$  | 0.31 | 0.32 | 0.33 | 0.34 | 0.35 | 0.36 | 0.37 | 0.38 | 0.39 | 0.40 |
| $Z$  | -0.50| -0.47| -0.44| -0.41| -0.39| -0.36| -0.33| -0.31| -0.28| -0.25|
| $p$  | 0.41 | 0.42 | 0.43 | 0.44 | 0.45 | 0.46 | 0.47 | 0.48 | 0.49 | 0.50 |
| $Z$  | -0.23| -0.20| -0.18| -0.15| -0.13| -0.10| -0.08| -0.05| -0.03| -0.00|
| $p$  | 0.51 | 0.52 | 0.53 | 0.54 | 0.55 | 0.56 | 0.57 | 0.58 | 0.59 | 0.60 |
| $Z$  | +0.03| +0.05| +0.08| +0.10| +0.13| +0.15| +0.18| +0.20| +0.23| +0.25|
| $p$  | 0.61 | 0.62 | 0.63 | 0.64 | 0.65 | 0.66 | 0.67 | 0.68 | 0.69 | 0.70 |
| $Z$  | +0.28| +0.31| +0.33| +0.36| +0.39| +0.41| +0.44| +0.47| +0.50| +0.52|
| $p$  | 0.71 | 0.72 | 0.73 | 0.74 | 0.75 | 0.76 | 0.77 | 0.78 | 0.79 | 0.80 |
| $Z$  | +0.55| +0.58| +0.61| +0.64| +0.67| +0.71| +0.74| +0.77| +0.81| +0.84|
| $p$  | 0.81 | 0.82 | 0.83 | 0.84 | 0.85 | 0.86 | 0.87 | 0.88 | 0.89 | 0.90 |
| $Z$  | +0.88| +0.92| +0.95| +0.99| +1.04| +1.08| +1.13| +1.18| +1.23| +1.28|
| $p$  | 0.91 | 0.92 | 0.93 | 0.94 | 0.95 | 0.96 | 0.97 | 0.98 | 0.99 | 0.995|
| $Z$  | +1.34| +1.41| +1.48| +1.55| +1.64| +1.75| +1.88| +2.05| +2.33| +2.58|

where $S_{ij}$ is the assessment of the $i$-th display option according to the $j$-th criterion, and $n$ is the number of criteria.

When conducting a study of pilots’ subjective assessment of display options, there is a risk that pilots may give answers randomly, without being included in the assessment process, which makes it necessary to analyze the answers to determine the accuracy and consistency of each pilot’s answers. The idea is to consider a logical location of the display options assessments on the axis of the preferences avoiding a situation when the 1st object is assessed above the 2nd, the 2nd is above 3rd, and 3rd above the 1st ($3 > 2 > 1 > 3$).

If such a situation takes place, then either the assessment criteria were determined incorrectly, or the pilot was not included in the process of conscious assessment and was passing the test “automatically”. If such a situation occurs systematically, the assessment criteria should be reviewed. If the occurrence of such a situation is typical for a particular pilot, then his assessments should not be taken into account in the general definition of the scale values of the display options.

The value of the pilot ratings sequence degree in the range from 0 to 1, is calculated as follows:

$$1 - \frac{\sum_{i=1}^{n} error(i)}{n},$$

where $i$ is the criterion’s number. The error function returns one of the two values \{0, 1\}, taking as an argument the number of the criterion that was assessed. According to this criterion, the pilot’s comparison table (see table 1) is taken, then the values of rows $i$ are summed up in this table. If there are repetitions among the row sums, the result of the function is 1, otherwise 0.
3. Test environment
The proposed method was integrated and developed based on a universal prototyping bench designed for testing the cockpit displays and controls [10]. The universality of the bench is meant as an ability to practice cockpits of various aircraft on it, for which the software of aircraft systems and cockpit displays and controls of a specific aircraft would run.

The overall architecture of the test environment for conducting an expert assessment based on the displays and controls is shown in figure 1.

When conducting ergonomic assessments on the displays and controls, some features should be taken into account:

- the bench does not allow to get tactile sensations when working with the controls;
- vestibular stimuli are not reproduced due to the lack of a mobility system;
- the color characteristics of the image on the displays differ from similar images in the aircraft;
- the spatial perception of the cockpit external environment is very limited due to the use of flat-panel displays for the visualization system.
All of the above aspects should be taken into account when drawing up the experiment plan, i.e. when choosing the object of research and assessment criteria.

4. Solving the problem of determining the ergonomic quality of the flight frame depending on the display/non-display of the layers of synthetic and enhanced vision

As a result of earlier work [11], variants of PFD with SVS/EVS layers were synthesized, the display of mnemonic symbols on which differs from the usual PFD, both in terms of color coding, and in terms of the location of mnemonic symbols, scales, and digital values (see figure 2).

![Figure 2. The PFD options (traditional, with the SVS layer, with the EVS layer).](image)

For the purposes of the study, special software was developed that records the opinions of pilots about the considered display options (questionnaires) and receives a final assessment based on these opinions (see figure 3). This software based on the approach described in the chapter “Methods of expert assessments”, is universal and, as a result, can be used for expert assessment of any other display types.

The assessment was carried out using some specially selected criteria, some of them are listed below:

- criterion 1. Visibility/distinctness and interpretability, clarity and non-distortion of display elements (scales, indexes, alphanumeric characters);
• criterion 2. Level of situational awareness (determining your location in low visibility, information about ground infrastructure objects and intruders);
• criterion 3. Ease of perception of the displayed information. The lack of distortion when overlaying symbols;
• criterion 4. The level of workload and the required level of concentration;
• general criterion. The level of display’s elaboration.

As a result of the experiment, according to a set of criteria, the results were obtained indicating a superiority of option 1 (normal flight frame) over other options (see figure 4).

At the same time, according to criterion 2 (level of situational awareness), the best options were options 2 and 3 (flight frame with SVS and EVS, see figure 5).
Figure 4. Results of PFD assessment options.

S1 = 0.67 (PFD)
S2 = 0.54 (PFD SVS)
S3 = −0.55 (PFD EVS)

Figure 5. Results of PFD options assessment according to individual criteria (part of the criteria is shown).
This result indicates that the basic flight information is best perceived on a traditional PFD without adding SVS/EVS layers. Respectively, in cases where increased situational awareness is not required, it is necessary to abandon the display of these layers.

5. Conclusion
The conducted research resulted in the development of a method of expert assessments of the cockpit displays and controls and processing of their results using the mathematical apparatus described in the article.

This method was tested on the basis of a universal prototyping bench and was highly appreciated by flight experts and specialists in the field of ergonomics. This method allows for an effective ergonomic assessment of the cockpit displays and controls, obtaining results with a high degree of reliability, and also reduces the time spent by pilots and other highly qualified personnel, which determines the high practical significance of the work.

A particular problem to determine the ergonomic quality of the flight frame depending on the display/non-display of the layers of synthetic and enhanced vision is solved on the basis of this method, although the use of this method is possible to assess any other components of the displays and controls.

This approach has not been used before to assess the cockpit displays and controls of an air transport aircraft. Instead of it, an expert assessment was used without using a special mathematical apparatus for calculating the quantitative display of assessments, so it is advisable to continue the development and implementation of ergonomic assessment systems based on the application of this approach.

References
[1] Greshnikov I I and Zlatomregev V I 2019 Advanced Perspective cockpit displays and controls, implementing new ways of information support of the crew and the aircraft’s control information field 5-th Int. Conf. “Persp. pathes of avionics dev. for civ. airc.” Moscow pp 77–87
[2] Greshnikov I I and Zlatomregev V I 2020 Using advanced technologies for cockpit displays and controls optimization of the advanced aircraft XVIII all-Rus. Sci. Conf. “Near. and their app.” Moscow pp 66–68
[3] Poisson R J 2014 Spatial Disorientation: Past, Present and Future (Ohio: Air force ins. of tech.) p 72
[4] Lee B G and Myung R 2013 Attitude Indicator Design and Reference Frame Effects on Unusual Attitude Recoveries Int. J. Aviat. Psychol. 23(1) 63–90
[5] Perminov G I 2011 A method for identifying generality in alternatives and criteria in decision-making tasks Soc. of Sci. and Tech. 2 90–104
[6] Podinovski V V 2007 Introduction to the theory of the importance of criteria in multi-criteria decision-making problems (Moscow: FIZMATLIT)
[7] Borisov A N, Grumberg O A and Fedorov I P 1990 Decision-making based on fuzzy models (Riga: Zinatne)
[8] Gusev A N, Izmailov C A and Mikhailievskaya M B 1998 Measurement in psychology: general psychological practice 2 ed. (Moscow: Smysl) p 111
[9] Kulaichev A P 2006 Methods and means of complex data analysis: Textbook 4 ed. (Moscow: Forum-Infra-M) p 50
[10] Zheltov S Y, Fedosov E A, Chuinov G A, Zlatomregev V I and Greshnikov I I 2016 Patent No. 101331 The complex of equipment (stand) prototyping of the cockpit displays and controls Russia
[11] Knyaz V V, Vygolov O V, Nikanorov A V, Greshnikov I I and Zlatomregev V I 2018 Development of the cockpit displays and controls demonstrator using elements of virtual reality based on the function of synthetic and enhanced vision III All-Rus. Sci. and Tech. Conf. “Modeling of Aviation systems” Moscow pp 285–86