Synthetic Evaluation Method of Electronic Visual Display Terminal Visual Performance Based on the Letter Search Task

Wei Liu[1,2,3], Weixu Cai[1,2(✉)], Borui Cui[1,2], and Muxuan Wang[1,2]

[1] School of Automation, Beijing University of Posts and Telecommunications, Beijing 100876, China
caiweixu_bupt@126.com
[2] Beijing Key Laboratory of Network Systems and Network Culture, Beijing 100876, China
[3] Beijing Key Laboratory of Work Safety Intelligent Monitoring, Beijing 100867, China

Abstract. Today the electronic visual display terminal (VDT) plays an indispensable role in people's life, so the visual performance of VDT is very important. In order to evaluate the visual performance of VDT from the perspective of user experience more accurately and reliably, we synthesize three kinds of commonly used visual performance assessment techniques, i.e. main task measure method, physiology measure method and subjective evaluate method, and establish a system to evaluate the visual performance of VDT. We use pseudo-text letter search task as the main task, analyze and extract suitable evaluation indicators, and calculate the weight of these evaluation indicators by the entropy method. At last, a qualification visual performance evaluation value of each VDT is got. The experimental result shows that this evaluation value is consistent with the subjective score in general. The evaluation system is combined with present main evaluation methods of visual performance and utilizes their advantages. The evaluation result can be quantified more directly and clearly. It provides a reference for the visual performance evaluation from the perspective of user experience.

Keywords: Visual performance · VDT · Visual fatigue · Synthetic evaluation · Letter search

1 Introduction

Electronic visual display terminal (VDT) plays an indispensable role in people's life in this ever-changing electronic age. Computers, mobile phones and tablets are filled with people's life and change our way of life. Therefore, the visual performance of VDT has become a research hotspot and focus in recent years [1–3]. There are a lot of methods to evaluate the visual performance of VDT, but the most commonly used methods are optical methods which are difficult to implement, more importantly, the optical evaluation methods can not reflect the real feeling of users completely. VDT visual performance can be evaluated from multiple dimensions, the characteristics of

R. Shumaker and S. Lackey (Eds.): VAMR 2015, LNCS 9179, pp. 30–38, 2015.
DOI: 10.1007/978-3-319-21067-4_4

multidimensional determine the diversity of its measuring method. Four kinds of measurement methods are relatively common and mature at present: 1) main task measure method; 2) auxiliary task measure method; 3) physiology measure method; 4) subjective evaluate method. Every evaluation method has its advantages and limitations. Although these methods are contribute to the evaluation of VDT visual performance, but they can't completely meet the requirements of product design for visual performance. Therefore we put forward a synthetic evaluation method for VDT visual performance based on user experience [4], it synthesizes three kinds of assessment techniques, i.e. main task measure method, physiology measure method and subjective evaluate method. Then identify weight of each evaluation method by entropy method and a synthetic evaluation value of test VDT could be got. The evaluation system is combined with present main evaluation methods of visual performance and utilizes their advantages, make visual performance evaluation results more accurate and intuitive.

2 Method

2.1 Main Task Measure Method

Main task measure method evaluate VDT visual performance by the performance of main task.

Letter search task is used as main task in this paper. This method was first developed and tested by researchers Jacques and Martin [5, 6] of the center research on user-system interaction, they evaluated the search performance of people with VDT and paper respectively. In this paper, we evaluated search performance of people with two different VDT. This method was detailed in ISO_9241-304 [7].

Pseudo-text letter search task. The total number of characters in a pseudo-text is 450, embedded spaces included in our experiment, this text consist of 10 lines and 45 characters per line (including space characters every line). Characters include the capital letters, the lowercase letters and numbers 0-9, and the text contain 15 % space. Every pseudo-text has a target letter make up about 2-3 percent of the total characters and display pseudo-text as a block of characters in one of five screen locations (top-left, top-right, bottom-left, bottom-right or in the center). The test participant's task is to scan the text and identify each instance of the target letter. A trail will be initiated while the test participant press 'ENTER' key, test participant scan the pseudo-text from the top to the bottom line and press the space bar while they find out a target character. Use 'ENTER' key to stop the trial in the end. The test participant should take a minimum break of 10 s between trails. The program will record all experiment data, include the duration of each experiment, the number of target letter in every pseudo-text and the number of the test participant searched. Every test participant should do 15-20 trails on each VDT and make sure a rest of 30 min before trials on next VDT. We will tell the test participant that we are testing the VDT and not them, so they should ensure search accuracy firstly. It means that if the visual quality of the display under test has deteriorated in comparison to a previous one you have to work slower, but if it has improved in comparison to a previous one you have to work faster.

Below is a sample of pseudo-text used in the experiment (Fig. 1):

```
f QvT7mSO Pg6WLEhV OSmry qg6lWvTRdHbk V9o PZH
izLGR9 Y S a4 n rW 3 a1H80Jdc nf 3 zhkp8MpXf8
weS gY F s GT0Ge71h jEDW phqlni KXoo sJdCi Zu
qC D4dYZw79U4dzojz T wHAvj2FH8 ovF 6 Pab d j8
Kt9d60Si6zevGTjA8f er3 lyEdhEL7AK9Df cq8niJq3
f6sqBFJVi4rzEsAWr NdC GaD F T xd389T e9N5Yjwr
1Elu Cv zmR e1PAyZQhJ1bUBxq mh uJ ernztVw5 uD
RCQqG6GoVGFb ZCo3A bG EUH RX8A g n zDnIOl8B64
75FP39 Wh r7iN3c26KTJ nva1Y9bDrg01CO y m KLWL
zLYjkIQ9zDnTeTvwxSSC d ksLQZTVHhap VQsf3 r 32
```

Fig. 1. A sample of pseudo-text

2.2 Physiology Measure Method

Physiology measure method evaluate VDT visual performance by visual fatigue of the test participant in our experiment. Using VDT for a long time, the eyes will appear visual fatigue phenomenon, the visual discomfort problems includes eye pain, eye redness, dry eyes, bloodshot eyes, headache and so on [8–10]. The degree of visual fatigue caused by VDT is inversely proportional to the VDT visual performance. Therefore, measure the degree of visual fatigue caused by VDT is a very intuitive and effective way to evaluate VDT visual performance. Many physiological indicators can be used to measure the visual fatigue caused by VDT, such as critical fusion frequency (CFF), blink rate, near point accommodation (NPA), electroencephalogram (EEG) and so on.

The near point of the eye is certain under normal circumstances, is about 5 to 10 m, and this number will get larger while visual fatigue exists, so the changes of near point distance can reflect the degree of visual fatigue.

When visual stimulation intermittent rather than continuous action, with the increase of frequency, flashing phenomenon disappear and people no longer feel flash but a completely stable continuous light, this phenomenon is called flicker fusion, and the minimum frequency that flicker disappear is called critical fusion frequency (CFF). CFF decreased while visual fatigue exists, therefore the changes of CFF can reflect the degree of visual fatigue.

CFF was selected as the evaluation index of physiology measure method in our experiment. Iwasaki and Akiya [11] reported that the decrease of CFF can reflect the degree of retinal deterioration caused by visual fatigue and a decline in the activity of the retina or the optic nerve. The experiment results of Murata [12, 13] show that while engage in the same job, the CFF of VDT worker is dropped significantly compare with non-VDT worker. Therefore, CFF can reflect visual fatigue degree more accurately.

2.3 Subjective Evaluate Method

Subjective evaluate method is mainly based on the questionnaire survey, ask the test participants to rate the VDT visual performance rely on subjective feeling. Subjective evaluate method is easy to operate and does not affect the experiment compare to objective evaluate method. The defect is it is easy to influence by personal experience and preferences of the test participant, the measurement result is not very accurate, and that is why we propose a synthetic evaluation method.

We use a nine-point numerical scale in our research. After completion of the trials with the test VDT, ask the test participant to rate the visual performance of that VDT on the nine-point numerical scale. With 1 being “Poor” and 9 being “Excellent”.

Below is the nine-point numerical scale we used (Table 1):

Table 1. Nine-point numerical scale

1	2	3	4	5	6	7	8	9
Poor			Fair			Excellent		

3 Evaluation Indicators

The trial shall be regard as invalid trial from statistical treatment if the error rate E > 10 %, it shows that number of missed or extra targets is too large in this trial.

Error rate, E, is defined as:

$$E = \frac{|T_0 - T_c|}{T_0} * 100\%$$

Where,

T_0 is the total number of target characters in the page of pseudo-text shown to the test participant;

T_c is the total number of target characters counted by the test participant.

3.1 Average Search Speed

We use average search speed as the evaluation index in main task measurement method. From the registered search time, T_i, corresponding to the valid trials (E < 10 %), the performance measure of a test participant, the average search speed, v, measured in characters/s, is calculated by:

$$v = n_t * n_c * \left[\sum_{i=1}^{n_t} T_i\right]^{-1}$$

Where,

n_t is the number of valid trials for that test participant;

n_c is the total number of characters in a pseudo-text, is 450 in current experiment (including embedded spaces).

3.2 The Rate of Change of CFF

Use the rate of change of CFF, θ, as the evaluation index in physiology measure method.

θ, is defined as:

$$\theta = \frac{f_2 - f_1}{f_1}$$

Where,

f_1 is the CFF of the test participant before trail on test VDT

f_2 is the CFF of the test participant complete trails on test VDT

3.3 Subjective Ratings, s

The subjective ratings, s, is get from nine-point numerical scale.

4 The Determination of Weights and Synthetic Evaluation Values

We use the entropy weight method [14, 15] to count weight of every evaluation indicators while calculate the synthetic evaluation values of VDT visual performance.

Entropy was first introduced to information theory by Shannon, and it is widely used in engineering technology, social economy and other fields. The basic concept of entropy weight method is to determine the objective weight according to the size of the index variability. Generally speaking, the smaller the information entropy, the greater the index variability, and the greater the role in the synthetic evaluation.

The steps of determine weight by entropy weight method are:

1. Data normalization

Dealt with the data with normalization. Assume k indicators $(X_1, X_2, \ldots, X_k)$ are given, where, $X_i = \{X_1, X_2, \ldots, X_n\}$, $(Y_1, Y_2, \ldots, Y_k)$ are the normalized data,

$$Y_{ij} = \frac{X_{ij} - min(X_i)}{\max(X_i) - \min(X_i)}, \text{ if j is a positive index}$$

$$Y_{ij} = \frac{X_{ij} - min(X_i)}{\max(X_i) - \min(X_i)}, \text{ if j is a negative index}$$

2. Calculate the information entropy of each indicators

According to the definition of information entropy in information theory, the information entropy of a set of data is E_j,

$$E_j = -\ln(n)^{-1} \sum_{i=1}^{n} p_{ij} lnp_{ij}$$

Where,

$$p_{ij} = Y_{ij} / \sum_{i=1}^{n} Y_{ij}$$

If $p_{ij} = 0$, define $\lim_{p_{ij} \to 0} p_{ij} lnp_{ij} = 0$

3. Determine the index weight

The information entropy of each index are $(E_1, E_2, \ldots, E_k)$, And the weight of index, α_i, is defined as below:

$$\alpha_i = \frac{1 - E_i}{k - \sum E_i} (i = 1, 2, \ldots, k)$$

The synthetic evaluation values of VDT A is X_a:

$$X_a = \alpha_1 * \frac{\sum_{i=1}^{n} v_i}{n} + \alpha_2 * \frac{\sum_{i=1}^{n} \theta_i}{n} + \alpha_3 * \frac{\sum_{i=1}^{n} s_i}{n}$$

Where, n is the total number of test participants.
The synthetic evaluation value of VDT B can be got in the same way.
The higher the synthetic evaluation value, the better the VDT visual performance.

5 Experiment

5.1 Test Participants

Test participants should be a sample representing the anticipated user population. The test participants are 20 college students aged 20 to 30 with normal visual acuity or corrected to normal in our experiment. And they have no any obvious physical or physiological conditions that could influence either their search performance or the quality of the images that they perceive.

5.2 Experimental Apparatus

- An experimental procedure written in Java. It can record the basic information of the test participant and generate pseudo-text and display the pseudo-text as a block of characters in one of five screen locations (top-left, top-right, bottom-left, bottom-right or in the center) automatically and randomly. It also record the number of target characters counted by the test participant and calculate the error rate.

- A flicker fusion frequency meter produced by APTECH, the model is BD-II-118. Its flash frequency rage 4.0 Hz to 60.0 Hz, and the measurement error is less than 0.1 Hz. We use red light and set the background light strength to 1/16 grade. Turn light intensity to 1/8 gear, and turn light-black ratio to 1:1 gear.
- A nine-point numerical scale

5.3 Test VDT

Two laptop produced by different company are selected as test VDT. They have the same screen size, recommended screen resolution and similar price.

5.4 Experimental Procedure

Led test participants to the test area after they sign an experimental knowledge book. Give them an experiment instruction which described the whole process of the experiment, the operation steps and points for attention in detail. Then the host show the test participants the usage of test procedure. In order to overcome the problem of initial learning effects, train the test participants before the main experiment by performing the task for at least 10 pseudo-texts (i.e. 5 trials) and take ten minutes rest. The test VDT are named A and B. We first measure the CFF of test participant before the trail, then test participant do the main task on A VDT, measure the CFF after the trails, ask the test participant to assess the perceptual performance of the VDT with respect to its visual comfort. Allow the test participant a rest break of up to half an hour before the same procedure on B VDT. The total experiment end after 20 participants finish all procedures.

6 Experimental Result

We get the following result after we use the entropy method to analyze all of data:

The weight of three indicators $(\alpha_{A1}, \alpha_{A2}, \alpha_{A3})$ are $(0.38643, 0.256261, 0.357305)$, and the average of 20 sets of data $(\bar{v}_A, \bar{\theta}_A, \bar{s}_A)$ are $(9.101795, -0.0368, 7.466667)$, then we can get the synthetic evaluation values of A, X_A, is:

$$\begin{aligned} X_A &= \alpha_{A1} * \bar{v}_A + \alpha_{A2} * \bar{\theta}_A + \alpha_{A3} * \bar{s}_A \\ &= 6.1756688 \end{aligned}$$

The weight of three indicators $(\alpha_{B1}, \alpha_{B2}, \alpha_{B3})$ are $(0.35252, 0.380385, 0.267096)$, and the average of 20 sets of data $(\bar{v}_B, \bar{\theta}_B, \bar{s}_B)$ are $(8.851186, 0.05024, 5.666667)$, then we can get the synthetic evaluation values of B, X_B, is:

$$\begin{aligned} X_B &= \alpha_{B1} * \bar{v}_B + \alpha_{B2} * \bar{\theta}_B + \alpha_{B3} * \bar{s}_B \\ &= 4.614646 \end{aligned}$$

We can see that the statistics of A are better than the statistics of B with the mean value of three indicators. The synthetic evaluation value of A and B are 6.175688 and 4.614646, VDT A has the better visual performance than VDT B, and this is consistent with the user's subjective feeling. The result shows that synthetic evaluation method not only reflects the visual performance difference between different VDT, but also avoids the one-sidedness of single evaluation method, it combines with the advantages of several evaluation methods and makes the result more reliable.

7 Discussion and Prospection

The synthetic evaluation method proposed by this article can provides guidance for assessing the visual ergonomics of display technologies with user performance test methods, but different from the optical test method, this synthetic evaluation method is simple, applicable to different types of VDT, combines objective with subjective and makes the evaluation results more reliable. However, there are some obstacle when we try to analyze the factors that cause this visual performance difference, this synthetic evaluation method can't get the specific parameters as the optical experimental methods. Therefore, this synthetic evaluation method has some limitation, it is more suitable for the monitoring step for the VDT visual performance in the later stages, and the optical test method is more suitable for the development and test stages. Therefore, our next work is analyze the physical factors of test VDT rely on the optical experiment method given in ISO 9241-30 and ISO 9241-307.

References

1. Matula, R.A.: Effects of visual display units on the eye: a bibliography. Hum. Factors **23**(5), 581–586 (1981)
2. Council on Scientific Affairs. Health effects of video display terminals. The Journal of the American Medical Association (1987)
3. Lai, Y.K,, Ko, Y.H.: Visual Performance and Visual Fatigue of Long Period Reading on Electronic Paper Displays. Journal of Ergonomic Study (2012)
4. Xiaowu, P,, Zhenglun, W., Lei, Y.: Evaluation of mental workload during reading performance on VDT. Ind. Health Occup. Dis. **32**(6) (2006)
5. Roufs, J.A.J., Boschman, M.C.: Text quality metrics for visual display units: I. Methodological aspects. Displays **18**, 37–43 (1997)
6. Oschman, M.C., Roufs, J.A.J.: Text quality metrics for visual display units: II. An experimental survey. Displays **18**, 45–64 (1997)
7. ISO 9241-304, Ergonomics of human-system interaction — Part 304: User performance test methods for electronic visual displays
8. Blehm, C., Vishnu, S., Khattak, A., et al.: Computer vision syndrome: a review. Surv. Ophthalmol. **50**(3), 253–262 (2005)
9. Ian, A.: Health aspects of work with visual display terminals. J. Occup. Med. **28**, 841–846 (1986)
10. Eva, S., Yves, B., Per, B., et al.: Reading on LCD VS e-Ink displays: effects on fatigue and visual strain. Ophthalmic Physiol. Opt. **32**, 367–374 (2012)

11. Iwasaki, T., Akiya, S.: The significance of changes in CFF values during performance on a VDT-based visual task. In: Kumashiro, M., Megaw, E.D. (eds.) Towards Human Work: Solutions to Problems in Occupational Health and Safety. Taylor & Francis, London (1991)
12. Murata, A., Uetake, A., Otsuka, W., et al.: Takasawa. Proposal of an index to evaluate visual fatigue induced during visual display terminal tasks. Int. J. Hum. Comput. Interact. **13**(3), 305–321 (2001)
13. Murata, K., et al.: Accumulation of VDT work-related visual fatigue assessed by visual evoked potential, near point distance and critical flicker fusion. Industrial Health (1996)
14. Qin, S.K.: The Principle and Application of Synthetic Evaluation, pp. 120–132. Electronic Industrial Publishing House, Beijing (2003)
15. Qiu, W.H.: Management Decision and Applied Entropy, pp. 193–196. China Machine Publishing House, Beijing (2002)