An Experimental Method Study of User Error Classification in Human-computer Interface

Xiaoli Wu
College of Mechanical and Electrical Engineering, Hohai University, Changzhou 213022, China
Changzhou Key Laboratory of Digital Manufacturing Technology, Hohai University, Changhzou, China
wuxlhhu@163.com

Xin Huang, Ruicong Xu, and Qingwei Yang
College of Mechanical and Electrical Engineering, Hohai University, Nanjing 213022, China
h363262173@163.com, 526735308@qq.com, 870272192@qq.com

***Abstract*—In the paper, we proposed an experimental analysis model of user error and established the error-cognition corresponding relationship according to the error collection-classification method. Through the design of error experiment, various errors in operation interface as well as reasonable and sufficient data were obtained. Based on the repeated experiments and user interviews, we classified errors to extract typical errors according corresponding frequency. According to the cognitive processing, five error types were identified. In the later secondary experiment, we verified the rationality of the error types. The obtained experimental results validated the three basic assumptions proposed before the experiment. Each error can be classified as the identified corresponding error type. In the statistics data, perception confusion is the dominant error type, indicating that operation interface contains understanding deviation. With time limit, error rate was significantly increased. We identified error types in operation interface according to the user error classification experiment method.**

***Index Terms*—User interfaces; User error; Cognition; error classification; Experiment analysis model**

I. Introduction*

User cognition difficulty in operation interface has become one of the main issues of human-computer interaction discipline. It is necessary to develop a reasonable method to guide operation interface design and reasonably solve the user cognition difficulty. The paper is to identify the true causes for operation interface cognition difficulty from the perspective of user error. Currently, many scholars have proposed to study the operation interface cognition from the perspective of user error. The availability interface design method was proposed to reduce human error probability (HEP) [1, 2]. Human error probability (HEP) was studied through user evaluation model [3-5]. And interface cognition error classification method was also proposed [6]. Operation interface dependability was improved through mitigation of human error (External Subgoal Support) [7]. Combined with cognition model, a new analysis framework for human error cause was proposed to analyze potential human error modes in different cognition stages [8, 9]. Shappell proposed the human factor analysis and classification system (HFACS) to analyze the 13-year data of plane accidents and found that error actions were the main analysis objects in accident study. And it was necessary to carry out further error classification study [10-11].

Several scholars have proposed some ideas on potential error types in interface design, providing the basis for the paper. Norman divided operation errors into the three types: error, slip and mistake [12, 13]. Reason believed that there were 8 basic error types: false sensation, attention failure, memory slip, inaccurate recall, misperception, error judgment, inferential error and unintended actions [14-16]. Li L. Sh. proposed that inattention and overattention should be the main study objects of user errors [17]. His theory included the following contents: slip caused by double capture, forgetting caused by interruption, weakened intentionality, misperception and overattention. The above analysis indicated that key issues of user cognition might be solved through error type research. However, the reasonable systematic error classification method is not available.

II. Research Method of User Error Classification

Hassnert and Allwood studied user error experimental analysis with two self-registration software products [18-19]. The experimental results showed that there was no common error classification criterion [20-21], indicating that other classification methods were not applicable. It is necessary to classify and summarize user errors according to specific experimental analysis. On the contrary, in the paper, we argued that, if the representative experimental objects were selected, common error types in operation interface might be

* Address correspondence to WU Xiaoli, College of Mechanical and Electrical Engineering, Hohai University, NO,200, Jinling Road, Changzhou, Jiangsu Province, China 213022.
E-mail: wuxlhhu@163.com

doi:10.4304/jsw.8.11.2890-2898

established through comprehensive error experiments. Thus, based on the results of by Hassnert et al. and the improved experimental method, combined with cognition research by Reason et al. in 1990, through the experimental analysis model study of user error, we obtained the error-cognition mapping relationship and identified reasonable error types.

In user error study, from the perspective of analyzing error type, we could explore the thinking process of the users in operation interface to search for the correlation between user error and cognition processing. In the paper, we established the experimental analysis model of user error. As shown in Figure 1, according to the analysis process from operation task, error statistics, error classification, user interviews and user purpose discovery to cognition process analysis, we examined the experimental tasks by video and recorded all the errors in detail for classification and statistics. Then we made user interviews to obtain classic errors with high rates and carried out task decomposition (operations may be repeated and the think-aloud method might be adopted at the same time) to analyze the user's true intentions.

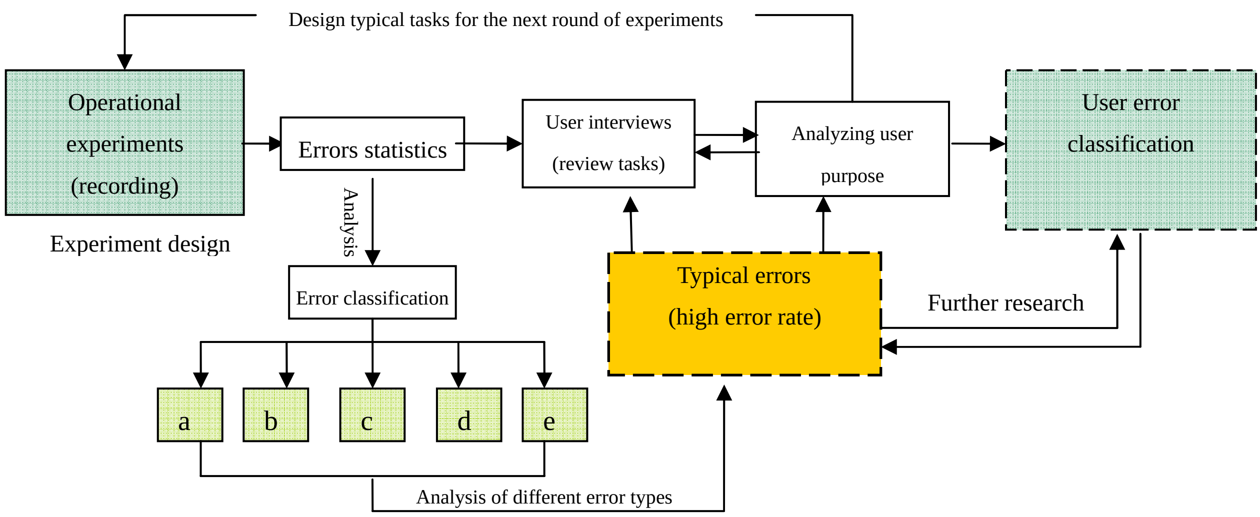

Figure 1. Analysis Model of User Error Experiment

III EXPERIMENTAL STUDY OF USER ERROR-COGNITION

A. Experimental Design

Objective:

In order to find the correlation between user error and cognition, we established cognition factors corresponding to the error types from the three aspects of visual perception, attention processing and working memory and explored the true causes of user error from the cognitive perspective.

Experimental method:

We selected the targeted industrial product interface and software interface as experimental subjects. The experiments were carried out in groups. According different operation tasks, appropriate participants were selected. And personal information of the participants should be completed. The selected participants were limited by certain conditions. In the experiment, with the complete video method, we recorded and classified all the errors. Through data analysis, we established the correlation model of error type and cognition. Then, we carried out user interviews according to classic errors (the main errors related to cognition) and designed operation tasks again (operation task design was improved according interview results). And then we selected new participants for 2-3 rounds of experiments.

Experimental content:

The experiment contents included experimental grouping, designing operation tasks, selecting participants, videos, recording errors, error classification and secondary experimental design.

The selected subjects shall contain typical activities related to living and study as possible and be representative (the selected products are the popular and moderately difficult products with corresponding user groups), so that the expected comprehensive errors may be obtained. We selected 5 groups of experimental objects: appliances, self-service machines, web, mobile devices and learning software. In each group, three typical products were selected. Then a total of 14 operation tasks were designed, covering different types of representative operation interfaces. The experiments of each group were carried out separately and corresponding experimental organizers were provided. The sub-task should be completed within the required time.

Fourteen typical task objects were designed. Different operation tasks should be designed for each product type. According to the difficult degree of operation interface, operation tasks were different in workload. For example, in the task of air conditioning remote controller, participants were required to start air conditioner at first, set the mode and parameters to meet the cooling, dehumidifying and other functions, and then turn off air conditioner. This operation task is relatively simple with less workload. For Photoshop software, several tasks may be designed and completed in stages. For operation tasks of each product type, 2-3 sub-tasks were designed and the difficult degree of the task, workload and other factors were omitted.

B. Results

Data analysis of experiment:

Participants

For error rate was in decreasing tendency from novice user and ordinary user to expert user, ordinary users with moderate error rate were selected as participants to avoid the low error rate of expert users and frequent errors of novice users. According to experimental contents, 14 tasks including 58 sub-tasks were designed for five groups. All the sub-tasks included 290 person-times in total. According to the qualified participant requirements, the same participant might participate in several sub-tasks (due to the difference of experimental tasks, the phenomenon did not affect data extraction). According to the statistics, all the person-times included 60 participants.

According to population distribution shown in the table, the participants were in line with the requirements of the initial experimental design. According to different experiment contents, the data of participants showed respective tendency. For example, participants of microwave sub-tasks had working background and cooking experience at home. For the sub-task of train ticket vending machine, the randomly selected participants were about 30 years old and the participants excluded train station management personnel (expert user or experienced normal user). And the majority of participants were the occasional users (people did not buy tickets frequently). Each sub-task included the records of more than 5 person-times test (excluding repeated persons). Thus we may analyze the recorded error data of the participants.

Statistical data of five-group errors

The experimental error data of 5 groups required no grouping or summary. And it was not required to pay attention to sub-task type. According to error feature, we summarized error list and described the same error type with the words which was basically in line with the contents. The experimental personnel carried out repeated discussions and video check and obtained all the errors. Experiment personnel analyzed the error list and removed the errors including those not related to the cognition or caused by experimental environment (such as interruption, interference and other external influences). The above data statistics provided 118 error types. Through discussion and analysis by experimental personnel, we integrated the same error type. As shown in Table 1, 47 error types with the frequency above 3 times were obtained in total. And the error frequency was also provided.

TABLE 1
ERROR TYPES AND TIMES STATISTICS

Error types	Error times
Participants could not find the proper position of the required button.	38
Because of program response time difference, participants erroneously thought that their clicks were not successful.	30
When participants looked for the operation buttons, other clicked buttons were clicked.	24
For previous operation were forgotten, participants carried out unusable operation.	20
For participants could not see the objects, they did not find the required function.	16
Participants could not find the internet access item.	12
During QQ withdrawal, participant could not find the withdrawal item.	3
During the usage of free switching tool, unlocking operation was forgotten.	3

Analysis of typical errors

According to error frequency, 47 typical errors were provided. Experimental personnel carried out repeated discussion and analysis and targeted interviews in which corresponding participants were required to recall their tasks. As shown in Table 2, the frequency of the errors in which participants could not find proper button reached 38. And the frequency of the errors in which participants repeated clicking the buttons for the program provided no feedback for the participants after clicking the button was up to 30. The frequency of the errors in which participants could not carry out the next operation for forgetting key items reached 20 and the errors were caused by the problems of learning and memory. The frequency of the errors in which participants made mistakes for they could not find the required function reached 16 and the errors were caused by poor instruction and vision restriction. Through the summary of the error causes, we obtained a series of typical error types for the next study of corresponding relationship between typical error and cognitive content.

TABLE 2
ERROR CLASSIFYING AND TYPICAL ERRORS

Error types	Typical errors
Forgotten, mistaken and absent-minded	For previous operation were forgotten, participants carried out unusable operation.
	Errors in Pinyin Input Method (mistaken usage of g and h or n and l)
	Participants forgot switching Chinese/English Input Method or forgot unlocking operation during the usage of free switching tool.
Hindering errors brought users into erroneous backgrounds.	Participants erroneously thought that their clicks were not successful.
	The mode switching button was erroneously considered as aperture tuning button.
	Participants could not find aperture tuning button.
Without instruction, participants could not find the task.	During operation task, participants could not find switching item, confirmation or next button.
	When website was entered for sending emails, participant could not find switching button of Chinese/English input method.
	After entering shooting mode item, participants could find submenu button.
	During QQ withdrawal, participant could not find the withdrawal item.
Unclear or confused	During internet access, participants did not know how to select access mode.
	Copying button was erroneously considered as copying mode button.
Erroneous direction	For coins were placed in improper position, machine did not recognize the coins.

C. Classifying Errors

The summarized error types were classified according to the contents of cognition phenomena. Then through the analysis of users' purposes, we grasped the thinking process of user error and obtained the corresponding relationship between cognition phenomena and error type (as shown in Table 3).

Combined with the knowledge structure of cognitive psychology, errors may be divided into the following five types: misperception, attention failure, perception confusion, memory slip and slip, as shown in Figure 2. Error types and relevant cognitive contents summarized in Table 5 might be used as the study objects of user errors in industrial products, human-machine interface and complex systems. In order to study user errors deeply in the cognitive structure level, it is necessary to design secondary experiment to verify rationality of error types according to the first experiment.

TABLE 3
THE CORRESPONDING RELATIONSHIP BETWEEN ERROR TYPE AND COGNITION

Relative cognition	Error types
Learning and memory problems	Forgotten, mistaken and absent-minded
Lack of memory aids	Participants forgot the procedure or did not operate according the required procedure.
Perceptual confusion. No guide was provided.	Hindering errors brought users into erroneous backgrounds.
	Without instruction, participants could not find the task.
	Unclear or confused
	Lack of function comprehension
Accident and negligence	Unconscious mistaken click without restriction factors
	Unconscious and customary actions
Errors caused by visual perception	Visual restriction or poor instruction
	Erroneous direction
Errors caused by time limit pressure.	Late response
Distraction	Strong interference and distraction.

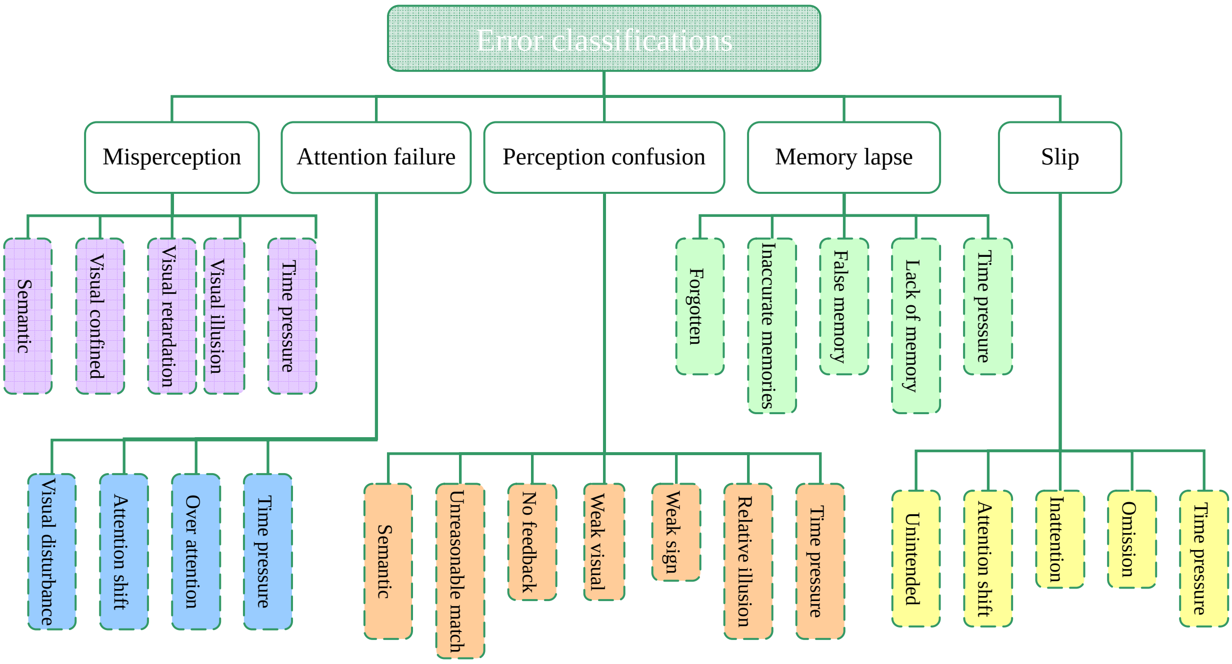

Figure 2. Error Classification

IV. The Verification Experiment Of User Errors (the secondary experiment)

A. Hypotheses

Hypothesis 1: We broke through the idea proposed by Hassnert and Allwood that there was no common error classification criterion and verified that error types could cover all the errors related to cognition. It was the basis of cognition study.

Hypothesis 2: In operation interface, error rate can reflect different tendencies of cognition phenomena. And perception confusion and misperception accounted for the main error rate.

Hypothesis 3: Error rate obtained with time limit was obviously higher than that obtained under comfortable and relaxed conditions.

B. Experiment Content

Participants

Common users with usage experience of Android 4.0 System were selected as participants. 30 undergraduates (with computer usage experiences) were divided into two groups. Each group was composed of 15 undergraduates.

Apparatus

Android system widely used in 3G mobile phone was used as experimental material. Operation interface of Android 4.0 System was shown in Figure 3. Participants were required to complete the required task. And the whole testing process was recorded by video.

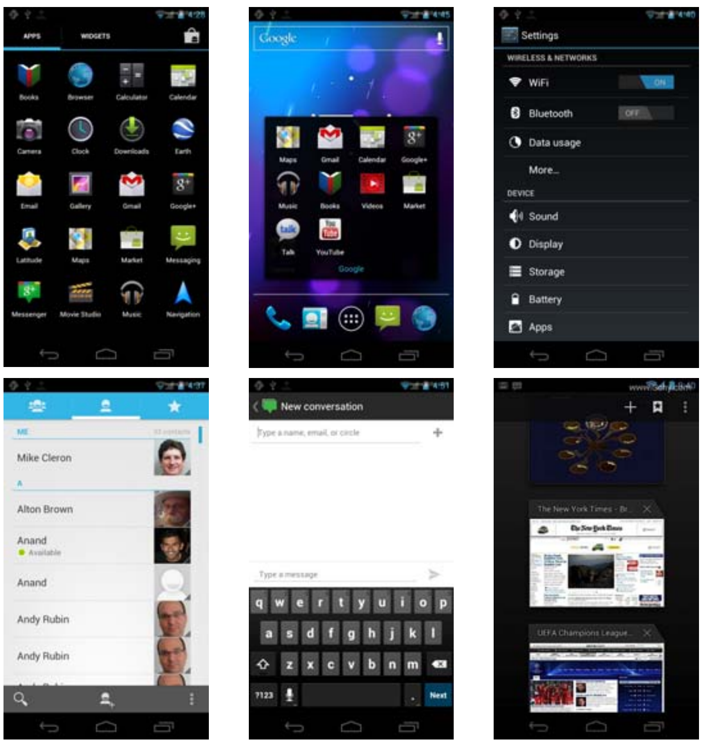

Figure 3. Operation Interface of Android 4.0

Task descriptions

The experiment was carried out in two groups. In the first group, without time limit, under easy conditions, participants may adopt think-aloud method and ask for guidance for completing the task. In the second group, with time limit, participants should try to resolve the difficulties faced in the task and experimental personnel might encourage participants to complete the task.

The tasks was provided as follows:

Add and store a new phone number in the address book and withdraw address book; look for Zhang, call him and hang up the phone.

Access the information menu, set the input method format, write SMS, add symbols and expressions and send SMS to Zhang; receive and respond to SMS and SMS forward SMS to Li.

Set the network mode, open the browser, check the mailbox and close the browser.

Open the reader, read documents, make and save notes, withdraw the reader and check the records.

C. *Results*

Error distribution under different time

After the experiment was completed, all the recorded errors were proposed in the table according to different requirements. Error contents and times of each participant with or without time limit were respectively recorded. The recorded contents also included error contents and error times in different tasks. In the statistics, errors of the same type were conflated at firstly. Then, according to the error types in Table 4 and Table 5, experimental personnel filled the statistics results in corresponding columns. The results showed that all the conflated errors were respectively distributed in 5 error types. And corresponding statistics data of each error type were provided in Table 4.

TABLE 4
THE DISTRIBUTION OF ERROR CLASSIFICATIONS IN RELAX

Errors	Error classifications				
	MP	AF	PC	ML	S
Participants could not find corresponding function button.	39				
Participants did not carry out current task.		4			
Participants forgot to set the parameters before the operation task.		23			
With interference, it is difficult to find the task for participants.		4			
Participants could not find check button.			10		
Participants did not know the setting ways.			3		
Erroneous understanding.			7		
Participants could not the symbols for the required operation.			12		
Participants forgot the procedure and did not carry out the test according to the procedure.				7	
Participants forgot the setting ways.				8	
Unconsciously click on improper keys.					35
Total	39	31	32	15	35

TABLE 5
THE DISTRIBUTION OF ERROR CLASSIFICATIONS UNDER TIME PRESSURE

Errors under time pressure	Error classifications				
	MP	AF	PC	ML	S
Participants could not see the button or find the required function.	85				
Participants did not find the required function.	27				
Sight restriction led to mistaken click.	22				
Participants could not find proper symbols.		55			
Participants forgot to set the parameters before the operation task.		32			
Participants completed the wrong operation.		6			
Participants clicked the wrong button.		5			
With interference, it is difficult to find the task for participants.		1			
Participants did not know check method or operation.			80		
Participants failed to complete the task.			76		

Participants completed the tasks improperly.			46		
Misunderstanding led to clicking other symbols.			17		
Without feedback, participants were not able to confirm whether the operation was right.			13		
Without response, participants repeated clicking the buttons.			11		
Without confirmation information, participants did not which button should be selected.			11		
Participants forgot the procedure and clicked the improper button.				19	
Participants input erroneous information.				18	
Participants forgot the procedure and did not carry out the test according to the procedure.				17	
Participants clicked the improper button.					115
Participants did not click the symbols.					49
Customary click or slipping					13
Total	134	99	254	54	177

TABLE 6
MEANS AND SD OF EACH GROUP ERROR IN DIFFERENT TASKS

Tasks	Relaxed		Under Time Pressure		
	Means	SD	Means	SD	U
Phone	1.930	1.163	11.277	4.432	0.000003
New messages	3.040	1.534	12.595	4.972	0.000003
Websites	2.400	1.121	13.297	4.272	0.000003
Reader	2.110	2.066	10.698	4.852	0.000003
Total	10.133	3.907	47.867	17.90	0.000003

Comparison of each group error in different tasks

Without time limit, in the operation of Android system, the average error number of one participant was 10.2667 (standard deviation of 4.182). With time limit, the average error number of one participant was 62.00 (standard deviation of 17.899). With SPSS software, we carried out Mann-Whitney U-test (U = 0.00003; P =0.009). The difference was as follows: error rate with time limit was obviously higher than the error rate obtained without time limit, as shown in Figure 4.

Distribution of user error probability

According to the data provided in Table 5 and Table 6, error type distribution under EC mode was obtained. According error types of the three tasks of dialing telephone, writing SMS and browsing web pages, the averages of various errors were calculated. And the calculated data were provided in Table 7 according to ascending sequence. Under time pressure, average error probability of perception confusion type in the three tasks of dialing telephone, writing SMS and browsing web pages was respectively 3.40, 7.00 and 6.27. And the average values were significantly higher than those of the two error types: slip and misperception. The difference of the values of the two error types of attention failure and memory slip was not significant in statistics. The error rate of the two error types was less than 1.60. The data indicated that perception confusion should not be neglected in the design. And it is necessary to improve interface deign from the perspective of cognitive psychology, as shown in Figure5.

TABLE 7
THE DISTRIBUTION OF DIFFERENT TASKS UNDER TIME PRESSURE (MEAN)

Different tasks	Error classification under time pressure (mean)				
	Mis-perception	Attention failure	Perception confusion	Memory slip	Slip
New messages	2.20	0.80	3.40	1.00	3.27
Phone	2.80	1.00	7.00	0.27	3.87
Websites	3.73	1.60	6.27	1.33	3.93

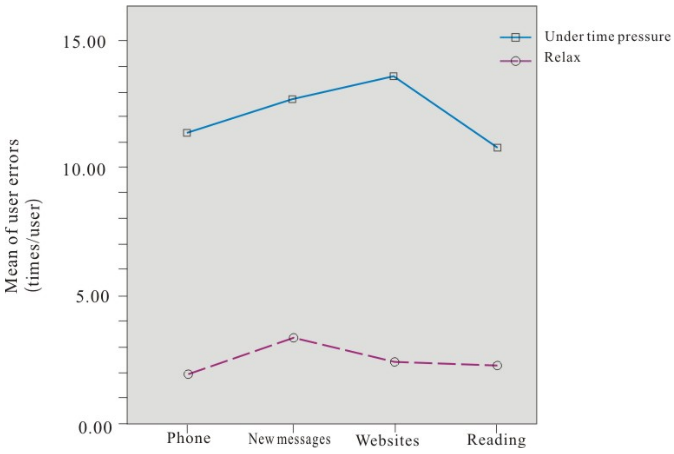

Figure 4. Error Means under Different Time Pressure

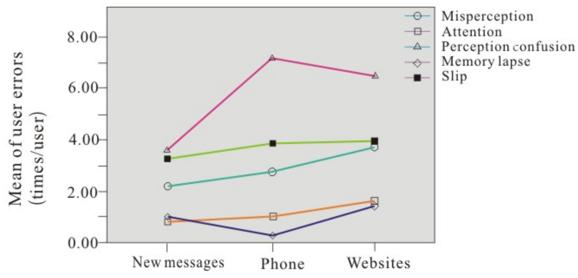

Figure 5. Error Classification under Time Pressure (Mean)

V. DISCUSSION

Experimental results verified that the established error classification method was effective. Through the designed 14 experiments, we obtained more errors as possible. Through the repeated video check, notes, interviews and original words, we identified five error types from the cognitive perspective.

The secondary experiment was to verify the rationality of classification error type. Experimental results showed that the recorded errors might be classfied as the different types proposed in relevant tables. Error classification included obscure error types. For the error in which participants could not find the required function, it was difficult to confirm whether the errors were caused by attention failure or perception confusion. For the errors, it was necessary to classify the error according to the task related to the errors.

The time pressure was a key factor of the error. The experimental results confirmed that different error data were obtained under different time pressure and that error rate of each type was obviously increased with time pressure. Without time pressure, errors caused by misperception, attention failure, perception confusion, memory slip and slip were respectively 30, 31, 32, 15 and 35. Under time pressure, the data were significantly changed and errors caused by the five types were respectively, 134, 99, 254, 54 and 177. The results were consistent with the assumptions.

In the experimental results, errors of perception confusion were much higher than errors of slip. Perception confusion mainly included semantic fuzziness, visual confinement, visual retardation, visual illusion, etc., indicating that the majority of errors in the operation were caused by misunderstanding and confusion in symbols, characters and color. The results showed that the design expression form was different from user cognition way. In the assumption, perception confusion and misperception were the two error types with relatively high error rates and the slip type was ignored and the error type was higher than error perception.

VI. CONCLUSIONS

It is an effective method to study the causes of user errors in operation interface from the cognitive perspective. We proposed an experimental analysis model of user error and adopted the error collection-classification method. Thus, we established the error-cognition corresponding relationship.

Through the error experimental design, we obtained various errors in operation interface. The obtained data were reasonable and complete. After the repeated experiments and user interviews, we classified errors and sorted errors according to corresponding frequency. Then we extracted classic errors and identified 5 error types according to cognitive processing.

In addition, we verified the rationality of the error types through the secondary experiment. The experimental results were basically consistent with the three assumptions proposed before the experiment. Each error could be classified into different error types. Perceptual confusion was the dominant error type, indicating that operation interface contained understanding deviation. Under time pressure, error rate was significantly increased.

In summary, we obtained error types in operation interface with user error classification method, providing further study for cognition difficulty in operation interface from the cognitive processing perspective. The user error classification method proposed in the paper can resolve the fundamental problems in current interface design and provide the basis for the reasonable design of friendly interface.

ACKNOWLEDGMENT

This work was supported by the Social Science Fund for Young Scholar of the Ministry of Education of China(Grant No. 12YJC760092), Fundamental Research Funds for the Central Universities(Grant No. 2013) and Changzhou Key Digital Manufacturing Technology Laboratory Foundation (Grant No. CM2007301).

REFERENCES

[1] Nielsen, J., Mack, R.L. (1994). Usability Inspection Methods [M]. New York:Wiley, 1994.

[2] N. M. Shryane, S. J. Westerman, C. M. Crawshaw,G. R. J. (1998).Hockey & J. Sauer. Task analysis for the investigation of human error in safety critical software design: a convergent methods approach[J]. Ergonomics, 41(11), 1719-1736

[3] Woods, D.D., Roth, E.M. (1988).Cognitive systems engineering. In: Helander, M. (Ed.), Handbook of Human–Computer Interaction [M]. Amsterdam: Elsevier Science Publishers.

[4] Billings, C. E., & Woods, D. D. (2001). Human error in perspective [J]. Postgraduate Medicine, 109(1), 13–17.

[5] Krokos KJ, Baker DP. (2007). Preface to the special section on classifying and understanding human error[J].Human factors,49(2):175-176

[6] Hidekazu Yoshikawa, Wei Wu.(1999). An experimental study on estimating human error probability (HEP) parameters for PSA/HRA by using human model simulation, Ergonomics, 42(11), 1588-1595

[7] Maxion RA, Reeder RW. (2005). Improving user-interface dependability through mitigation of human error[J]. International Journal of human-computer studies, 63(1-2):25-50

[8] Sun Zhiqiang, Shi Xiujian. (2008). Study on causation analysis method for human errors[J]. China safety science journal, 18(6):21-27. (in Chinese)

[9] Sun Zhiqiang, Shi Xiujian. (2008).A framework for classifying human errors based on cognitive model[J].Journal of national university of defense technology, 30(1):73-77,93.(in Chinese)

[10] Shappell S, Detwiler C,Holcomb K. (2007). Human Error and Commercial Aviation Accidents: An Analysis Using the Human Factors Analysis and Classification System[J]. Human Factors, 49(2): 227-242

[11] Scott A. Shappell, Douglas A. Wiegmann. (2001). Applying Reason: The human factors analysis and classification system(HFACS) [J].Human factors and Aerospace Safety, 1(1), 59-86.

[12] Norman, D.A. (1988). The Design of Everyday Things [M].New York:Doubleday, 1988.

[13] Norman, D.A. (1981).Categorisation of action slips[J]. Psychology Review, 88:1-15

[14] Reason J. (1987).A framework for classifying errors. In: New Technology and Human Error[J]. Chichester:New technology and Work. Wiley.

[15] Reason J. (1990). Human Error[M]. New York: Cambridge University Press.

[16] Reason, J. (2000). Human error: Models and management[J]. British Medical Journal, 320, 768–770.

[17] Li Leshan. (2004). Human computer interface design[M].Beijing:Science Press. (in Chinese)

[18] Mathias Hassnert, Carl Martin Allwood. (2002). Development context and ease of use of three programs for self-registration of unemployed people[J]. Computers in Human Behavior, 18 :191–221

[19] Seppala P, Salvendy G. (1985). Impact of Depth of Menu Hierarchy on Performance Effectiveness in A Supervisory Task: Computerized Flexible Manufacturing System[J]. Human Factors, 27(6):713-722.

[20] Xinwei Zheng. (2012). Information-based study of E-commerce website design course [J]. Journal of software, 7(12):2794-2799.

[21] Tianwei Sheu, Tzuliang Chen. (2013). Study on the Conception of Learning Problems of Students by Combining the Misconception Domain and Structural Analysis Methods[J]. Journal of software, 8(5):1255-1266

Xiaoli Wu is currently a PhD candidate at *Southeast University, China*. She is also a lecturer at *School of Mechanical and Electronics Engineering, HoHai University, China*. Her research is focused on the industrial design and product development, design-cognition of human-computer interface.

Xin Huang is currently a postgraduate in *School of Mechanical and Electronics Engineering, HoHai University, China.*

Ruicong Xu is currently a postgraduate in *School of Mechanical and Electronics Engineering, HoHai University, China.*

Qingwei Yang is currently a postgraduate in *School of Mechanical and Electronics Engineering, HoHai University, China.*