

The Emotional Experience of Inquiry Feedback Delaying in Online e-Commerce Platform Service Design: A Case Study from Chinese Customers

Hao Tan¹(✉), Wei Li², and Jiahao Sun²

¹ State Key Laboratory of Advanced Design and Manufacturing for Vehicle Body, Hunan University, Changsha, China
htan@hnu.edu.cn

² School of Design, Hunan University, Changsha, China
{liwei2014,sunjiahao2015}@hnu.edu.cn

Abstract. In online e-Commerce platform service design, feedback delaying occurs in different service scenarios, which has become one of the key issues of the current service design and user experience study. Based on the previous research on the feedback delaying during interaction, we took the Chinese consumer as an example and focused our research on the inquiry feedback service delaying when the servants answer the consumer's questions in online e-Commerce platform. We adopted the methods of protocol analysis and emotional scales and arranged different feedback delays into the inquiry tasks and obtained the consumer's oral reports and subjective rating scores. We analyzed the levels of emotion and experience and gained the quantitative indicator of emotional experience. At last, we proposed a correlation model between different feedback delaying and emotional experience levels in online e-commerce platform, which produced a trial design guide of feedback delaying service design in online e-commerce platforms.

Keywords: Feedback delaying · Online inquiry · Emotional experience · Online e-commerce platform · Service design

1 Introduction

Feedback online is one of the most important issues in service design on e-commerce platform. During e-commerce serving, the feedback delaying is a kind of time delay when the user operates and waits for the service system or servicer's feedback resulting in the impact of the user's emotional experience, which is an essential issue in service design and user experience research. In early 1993, Nielsen has proposed the impacts of three kinds of time delay (0.1 s, 1 and 10 s) on user's emotional perception, which is the earliest study about the impact of time delay on the quality of user perception in the field of HCI. In 2003, the International Telecommunication Union established Recommendation ITU-T P.800.1 standard based on the delay of voice calls. This standard focused how the quality of PSTN or CS voice type can affect user experience.

Furthermore, an experience quality has been proposed based on five-point scale ranging from excellent to bad. In 2012, Lorentzen et al. put forward the difference impacts between initial delay and interrupt delay on user experience of service and product and built related theoretical model. Reichl et al. (2010) observed a distinct sensitivity of user perception to response and download times in interactive services on the web. Egger et al. (2012) also presented the delay impact on the quality of experience on web services. By the year of 2014, Ericsson APP coverage whitepaper presented the impact on the user experience within 10 s delay.

Our experiment took the Chinese consumer as an example, participants were recruited and asked to complete three different shopping tasks on one or two online e-commerce platforms. On the basis of our previous research method of the verbal analysis with Chinese language (Tan and Sun 2015), we analyzed the levels of emotion and experience from the consumer's oral reports and gained the quantitative index of emotional experience. Moreover, with the consumer's subjective scores in emotional scales, a correlation model between different inquiring feedback delaying and the different emotional experience levels in online e-commerce platform was designed, which produced a trial guide of feedback delaying service design in online e-commerce platform and help to promote the online e-commerce service experience.

2 Experiment

2.1 Participants

In this research, we invite 47 participants in total (21 male and 26 females). All participants had online shopping experience with different length of time. Their ages ranged from 18 to 51. 10 of the participants took part in the experiment 1 (5 male and 5 females). 17 of the participants took part in the experiment 2 (7 male and 10 females). 20 of the participants took part in the experiment 3 (9 male and 11 females).

2.2 Experimental Arrangement

The research was divided into three experiments. In Experiment 1, use the timer and bell to record the point of the limit time. In Experiment 2, the participants were asked to narrate their feelings using think aloud protocol and use the recording equipment to record the experimental condition. In Experiment 3, the participants need to rate the emotion scale.

2.3 Experiments

Experiment 1. To simulate the real online shopping environment, we made three kinds of web pages (food, clothing and office supplies) based on Taobao pages. Every participant need to complete three tasks. In each task the participant selects a product the experimenter specified. Then the participant need ask the customer service whether

the products are in stock and wait for the reply. When the participant send the message, experimenters started timing and marked the time point as T0 (=0). The participants rang when they start to feel impatient, experimenters marked this time point as T1. Finally, the experimenters calculated average of T1 (Fig. 1: Experiment 1 process).



Fig. 1. Experiment 1 process

Experiment 2. Experiment 2 had n parts and in each part the participants can ask any questions to the customer service. They need to wait for the answers for different length of time in different parts. During the experiment their verbal reports were recorded.

In experiment 1 we got the limit time T1. Based on this limit time we divided the waiting time into several parts. In these parts participants need to wait different length of time. Each neighboring parts has a difference of 3 s. So for each part the waiting time is 3 s (first part), 6 s (second part)...T2 (n part) (Fig. 2).

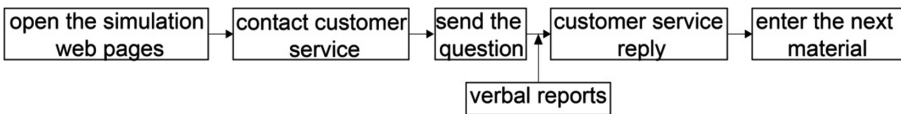


Fig. 2. Experiment 2 process

Process Methods and Measurements of Verbal Report. Direct extraction method: adjective and adverbs of degree that appeared explicitly in sentences were directly extracted (Table 1).

Table 1. Direct extraction

Participants exact words	Extracted adjectives	Adverbs degree	Level
I wait a little long	Long	A little	Lower
This reply is relatively fast	Fast	Relatively	Medium grade
Why does he not reply, so slow	Slow	So	High level
Too slow!	Slow	Too	Extreme
It's pretty fast	Fast	Pretty	High level

Situational Extraction. Sometimes the feedback from participants did not contain any obvious adverbs or adjectives. One way we could judge the participants’ attitude in this case is that judge through colloquial descriptions (including exclamatory and interrogative sentences) and intonation combined with the context in which the language was uttered. Once interpreted, this raw data was translated into adjective-dominated declarations (Table 2).

Table 2. Situational extraction

Participants exact words	Extracted adjectives	The adverbs of degree	Level
Wow, he replied!	<i>Quick</i>	<i>Very</i>	High level
Why didn’t he reply?	Slow	Pretty	High level
Is he next to the computer?	Slow	So	High level
He can’t surf the Internet?	Slow	too	Extreme
Is this automatic reply?	Fast	Pretty	High level

Verbal Language Environment Extraction. Another way to judge participants’ attitude in the case that the feedback we got did not contain any obvious adverbs or adjectives was through the analysis of participants’ verbalized assumptions, comparisons, suggestions, and expectations. After being analyzed, this language was translated into adjective-dominated declarations (Table 3).

Table 3. Verbal language environment extraction

Participants exact words	Extracted adjectives	The adverbs of degree	Level
I feel impatient, I will ask it again and again	Slow	Very	High level
I hope this store has an automatic response	Slow	So	High level
Do they have a lot of customers?	Slow	So	High level

Incidence-Description Extraction. The third way we could judge participants’ attitude using feedback without any obvious adverbs or adjectives was through participants’ descriptions of the test process. We analyzed the movement and brain function of the participants. Following analysis, the data was also translated into adjective-dominated declarations (Table 4).

Classification of Adjectives. The data—adjectives and adverbs—were then coded in accordance with the guidelines set forth by Ma Shi When Tong. First, adjectives were classified according to whether they were positive or negative. If an adjective expressed a relatively pleasant position (such as “good” or “easy”), it was coded as positive (+). If an adjective expressed a relatively unfavorable position, such as tension or worry, it was coded as negative (-). After classifying the adjectives, the total amount of each

Table 4. Incidence-description extraction

Participants exact words	Extracted adjectives	The adverbs of degree	Level
I want to do something else when I wait for the answers, if nothing can be done like this, so that is waste of my time	Long	Very	High level
I think I've been waiting all the time	Slow	Too	Extreme
He is a responsible customer service	Quick	So	High level

kind of adjective was determined. In our analysis here, this amount is expressed by the number n occurrences of each adjective type.

Adverbs were coded differently. Because the participants' reports were recorded in Chinese and, for the sake of accuracy, analyzed in Chinese, the processing of adverbs made reference to the local grammar. According to the XinHua Dictionary, adverbs of degree can be divided into four categories: intense, high, moderate, and low. According to these categories, adverbs extracted from the data were arranged on a Likert scale with eight levels. Intense, high, moderate, and low positive adverbs were assigned 4, 3, 2, and 1 points, respectively. Likewise, intense, high, moderate, and low negative adverbs were assigned -4, -3, -2, and -1 points, respectively. Neutral adverbs were assigned 0 points (Table 5).

Table 5. Research degree adverbs of degree

Level of gradable adverb	Level of gradable adverb in paper
Extreme	Too, extremely
High level	Also, very, quite, pretty, especially, utterly, fully, so, fairly, such, particularly, incomparable
Medium grade	Relatively, even more, still more
Lower	Slightly, a little

Experiment 3. In Experiment 3, the participants will complete the n parts same as Experiment 2 and when they complete each part, they have to rate the affective dimensions using the Self-Assessment Manikin (Fig. 3).

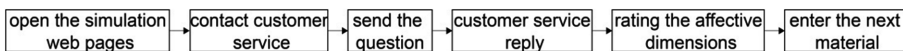


Fig. 3. Experiment 3 process

The Self-assessment Manikin. The participant need to rate their emotion in 9 levels (upset and calm). 1 is very irritable, 9 is very calm, 5 is the intermediate emotion between irritable and calm. Assessment materials are the Self-Assessment Manikin by Bradley and Lang (1994).

3 Results

3.1 Experiment 1

This study tested what is the last straw of subjects when they wait for the Customer Services reply in three kinds of products respectively.

When buying food item online (such as Three Squirrels), subjects’ average impatience time T1 is 31.9 s. While buying Office Supplies online (such as Deli), the average impatience time T1 is 31.3 s. As for the clothing category which took Sundance as an example, T1 is 33.3 s (Experimental result has listed in the following Fig. 4).

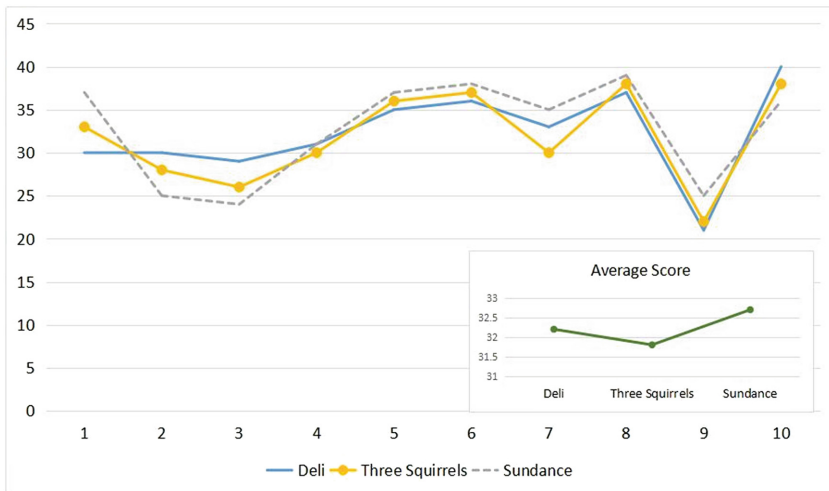


Fig. 4. Limit waiting time of the participants

3.2 Experiment 2

In experiment 2, the subjective experience of 11 tests used the same experimental materials, the only difference between these tests is the time of customer services reply. While testing, the experimenter recorded the subjective verbal reports, then analyzing the emotion of these reports. At last, calculating average of subjects’ positive and negative emotion then divided into five levels.

The first thing to be done in this experiment is determining the positive and negative tendencies of user experience. For instance, user A said: “wow! It replies me so

fast!” So the positive emotion can be identified from a response like this. Secondly, using the grading scale above, adverbs of degree were added to the existing positive and negative adjective tallies, and weighted at a value of 4, 3, 2, or 1 according to which of the four adverb categories they belonged to. According to the method of emotional analysis based on verbal Chinese (Tan and Sun 2015), we calculated the user experience using this formula.

$$S_{UX} = \frac{1}{N} \sum_{i=1}^m (n_i \bullet a_i) \quad (1)$$

Where

$$N = \sum_{i=1}^m n_i \quad (2)$$

means the total number of all adjectives belonging to the same category, n_i represents the number of occurrences of a certain grade in a certain adjective, a_i on behalf of the represented the weight value of that grade. M is 4 on behalf of the four different weights, and S_{ux} is the weighted average of participants' evaluations of different tests based on this data.

As the experiment only aims to explore a single factor's, namely the feedback time's influence on the user experience of online shopping, and every participant experienced all 11 tests of this experiment, the results are ideal. The adjectives extracted from the original data have a high repetitive rate and all of them can be classified into 2 categories: fast and slow. This means that we can easily rule out the influence of the uncertainty of the weight of other factors and focus on the unidimensional influence of feedback time. In this case, the confidence coefficient α identity in 1. And the value of S_{ux} represents the overall user experience of specific tests.

It is obvious that as the feedback time lengthen, user experience gradually becomes worse. According to the method above, the highest score (3.09) appears when the feedback time is 3 s. In this case, all user experience are positive. And the figure reached the bottom at -3.16 when the feedback time is 31 s. So the overall range of the participants' score is 3.09 to -3.16 . It is important to note that when the feedback time increased to 9 s, negative user experience appears, then positive and negative user experience coexist for the next 9 s until the feedback time lengthened to 18 s, all participants turned into negative emotion. The experimental data has shown in Table 6 and Fig. 5 listed below.

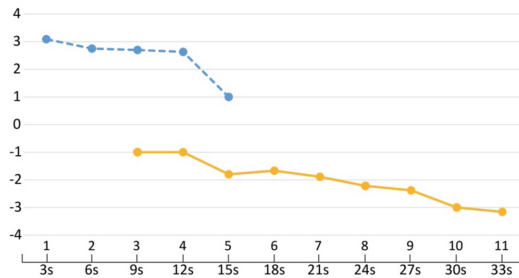
3.3 Experiment 3

We got the average scores of the affective dimensions from the participants (Table 7), then divided these scores into five levels (Table 8, Fig. 6).

At the same time, according to the scale of the test, we analyzed the score trend of customer service reply time for different shopping experience (Fig. 7), by the table can

Table 6. The experimental data of experiment 2

Experimental material	Reply	Average Score		Final average score
1 st experimental material	3 s	+3.09	0.00	+3.09
2 st experimental material	6 s	+2.75	0.00	+2.75
3 st experimental material	9 s	+2.70	-1.00	+0.85
4 st experimental material	12 s	+2.63	-1.00	+0.815
5 st experimental material	15s	+1.00	-1.80	-0.40
6 st experimental material	18 s	0.00	-1.67	-1.67
7 st experimental material	21 s	0.00	-1.89	-1.89
8 st experimental material	24 s	0.00	-2.22	-2.22
9 st experimental material	27 s	0.00	-2.38	-2.38
10 st experimental material	30 s	0.00	-3.00	-3.00
11 st experimental material	33 s	0.00	-3.16	-3.16

**Fig. 5.** Trend of self-assessment Manikin scores

be seen, in the same waiting time, Love Heart user's score is higher, the lowest score were Pink crown users. The overall trend according to the score from high to low is Heart users, Gold diamond users, Gold crown users, Pink crown users.

3.4 Analysis on the Results

By comparing the analysis results of the user's oral reports and scale scores, which prove the consistency of the user oral reports and subjective rating scale (Fig. 8).

According to the experiments, user's average limit waiting time is about 32.2 s. According to the user's oral report and scale score, users' emotion changes when waiting for the reply can divided into five grades. From the verbal reports, the level five to the level four is the transition stage of emotion, behind the third grade, start from grade four the users only have negative emotions (about 15 s). The third level was the user's positive and negative emotion change level. In summary, we divided the point of user's emotional change into three stages (Fig. 9), the first stage is 0–8 s and the second stage is 9–17 s, and the third stage is after 18 s.

Table 7. The experimental data of experiment 3

Experimental material	The time of receiving reply	The final average score
1 st experimental material	3 s	8.35
2 st experimental material	6 s	7.75
3 st experimental material	9 s	6.60
4 st experimental material	12 s	5.90
5 st experimental material	15 s	5.10
6 st experimental material	18 s	4.35
7 st experimental material	21 s	3.45
8 st experimental material	24 s	2.95
9 st experimental material	27 s	2.45
10 st experimental material	30 s	1.65
11 st experimental material	33 s	1.36

Table 8. Self-assessment Manikin score

Level	Experimental material	The time of receiving reply
5	1 st , 2 st experimental materials	3 s, 6 s
4	3 st , 4 st experimental materials	9 s, 12 s
3	5 st , 6 st experimental materials	15 s, 18 s
2	7 st , 8 st , 9 st experimental materials	21 s, 24 s, 27 s
1	10 st , 11 st experimental materials	30 s, 33 s

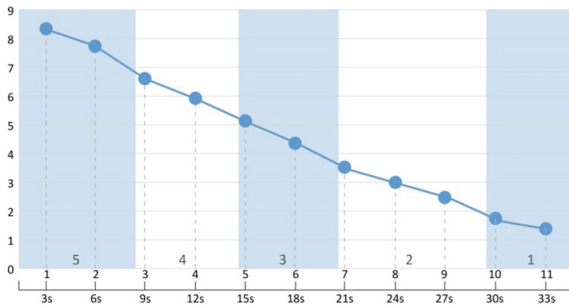


Fig. 6. Self-assessment Manikin score and levels

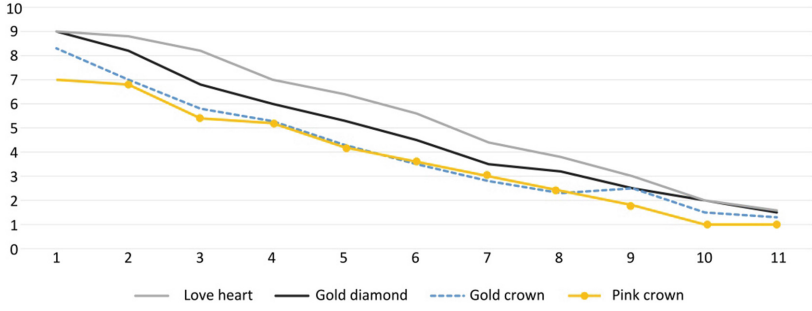


Fig. 7. Score-based trends in different shopping experience

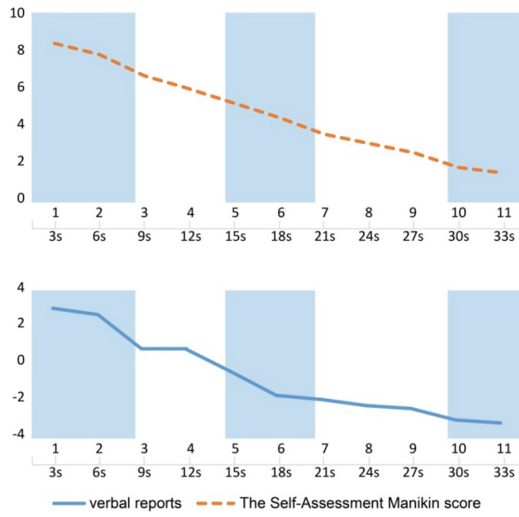


Fig. 8. Oral report and subjective rating scale

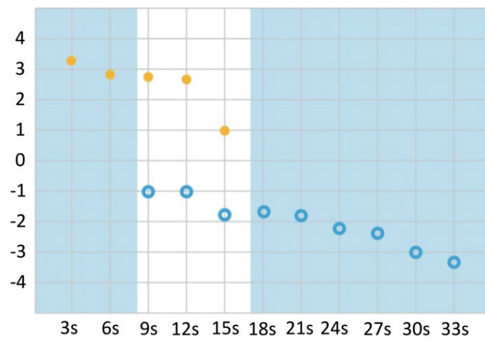


Fig. 9. Emotion change stages

4 Conclusion

Online shopping is a complex interactive process, as a whole, each procedure (browsing information of products, asking for shopping advice, purchasing orders and so on.) will affect the user experience feelings. Providing a high quality of service is a key point to affect customers' purchase decisions. However, when facing numerous customers' inquiries, it is difficult to reply every customer in time. From the analysis of the results of oral reports and subjective scoring through experiments, we found that users have different emotional reactions during different feedback delays which result in different shopping feelings. For resolving this problem, different interaction measures can be designed in each period to improve the user experience of shopping service. From the emotional model experiment based on phased emotional experiences, we can make targeted services designed by different interactions. For example, in the initial stage (0–8 s) to give customers quick response, we translate common issues to digitally encode in order to give users fuzzy answer, which could resolve a number of customer problems. In the second phase (9–17 s), the platform can provide some information of the commodity business philosophy to maximize the users' residence time. After 18 s, the sellers can give some kinds of discounts, free postage, membership services and other value-added promotions strategies to increase the user's propensity to buy.

At the same time, compare the users experience of online shopping, it is obviously that users with more shopping experience are impatient than the less experienced users based on our experiment data. The back-end data filtering can be set to certain privileges for more experienced users or give certain privileges embodied in the consulting platform interface, to enhance emotional interaction with users and promote their emotional experience to loyalty.

5 Future Work

Experimental study in this article is limited to the Chinese online shopping platform, so there are still several problems need to be solved. At first, we will subdivide consult situations and shopping segments for the specific situation of each different process, and observe mood changes in different feedback delays under different situations, to discover user feedback mood changes in separate shopping consultations. Secondly, we will divide products into different categories, and compare the differences between different types of merchandise buying behavior in diverse product categories. Furthermore, there are numerous works need to do to help us explore more details about the user experience on e-commerce platform.

- More research on the different stages in the process of online shopping, compare the different experience based on pre-sale service, sale service, after-sale service.
- Design and compare different interaction ways to abstract customers in the website from different angles, like product introduce, selling strategy then measure its validity.

Acknowledgments. We would like to express our gratitude to Xuyi Wei, Shihui Xu and Honghong Qin who helped us during the experiment. The research was supported by National Key Technologies R&D Program of China (2015BAH22F01), National Natural Science Foundation of China (61402159/60903090), Hunan Provincial Social Science Foundation of China (2010YBA054), the State Key Laboratory of Advanced Design and Manufacturing for Vehicle Body Funded Projects.

References

- Nielsen, J.: Usability Engineering. Academic Press, Boston (1993)
- Hossfeld, T., Egger, S., Schatz, R., Fiedler, M., Masuch, K., Lorentzen, C.: Initial delay vs. interruptions: between the devil and the deep blue sea. In: QoMEX, Visual User Experience 1, vol. 27, pp. 1–6 (2012)
- International Telecommunication Union. Mean Opinion Score (MOS) Terminology. ITU-T P800.1 (2003)
- Reichl, P., Egger, S., Schatz, R., D'Alconzo, A.: The logarithmic nature of QoE and the role of the weber-fechner law in QoE assessment. In: Proceedings of the 2010 IEEE International Conference on Communications, pp. 1–5 (2010)
- Egger, S., Reichl, P., Hossfeld, T., Schatz, R.: Time is bandwidth? Narrowing the gap between subjective time perception and quality of experience. In: IEEE ICC 2012 - Communication QoS, Reliability and Modeling Symposium (ICC 2012 CQRM), Ottawa, Ontario, Canada (2012)
- A Whitepaper from Ericsson, Facebook and XL Axiata: Measuring and Improving Network Performance. https://fbcdn-dragon-a.akamaihd.net/hphotos-ak-xap1/t39.23656/10734295_270843076447514_1952370192_n.pdf
- Loiacono, E.T., Watson R.T., Goodhue, D.L.: Web qual: a measure of website quality. In: Proceedings of the Winter Educator's Conference, American Marketing Association, Chicago, pp. 432–437 (2004)
- Christie, B.: Research finds life is a three second experience. *BMJ* **342**(5), d750 (2011)
- Tan, H., Sun, J.: Differences between the user experience in automatic and driverless cars. In: IASDR Interplay, 1975–1983 (2015)
- Ma, J.: Ma Shi Wen Tong. Business Press, Shanghai (2010)
- Bradley, M.M., Lang, P.J.: Measuring emotion: the self-assessment Manikin and the semantic differential. *J. Behav. Ther. Exp. Psychiatry* **25**, 49–59 (1994)