Evaluating Response Delay of Multimodal Interface in Smart Device

Xiantao Chen(✉), Moli Zhou, Renzhen Wang, Yalin Pan, Jiaqi Mi, Hui Tong, and Daisong Guan

Baidu AI Interaction Design Lab, Beijing, China
chenxiantao@baidu.com

Abstract. Multimodal interface based on natural language processing (NLP) technology is becoming more and more popular. Many studies show that response delay is a key factor to evaluate multimodal interface performance that can influence naturalness and fluency of interaction experience. However, few studies have been conducted to define the optimum response delay time. Focused at multimodal interface with voice as dominant modality, in this paper, we built a research framework to evaluate response delay according to user perception during the voice interaction, in which the system output process was divided into three successive stages. We carried out two experiments to evaluate the influence of response delay time in different stages, a smart speaker with screen and a smart TV were involved in the experiments. The first experiment focused on automatic speech recognition (ASR) feedback delay time, and the second experiment was designed to investigate the influence of both query response delay time and loading response delay time. We defined the satisfying and acceptable delay time for each stage respectively, which could be used as the references to improve corresponding technical performance.

Keywords: Smart device · Multimodal interface · Voice interaction · Response delay · Performance

1 Introduction

With the development of artificial intelligence technology, interaction between human and devices is changing profoundly, and more natural and efficient multimodal interface is becoming increasingly prevalent. Multimodal interface enables people to interact with devices through voice, touch, face expression and other modes, which is considered to be more intuitive and easier to learn for people. Comparing to unimodal interface, multimodal interface makes full use of human's natural ability to interact with devices, and the weaknesses or deficiencies of the unimodal interface can be compensated by combination the various modes [1].

In recent years, multimodal interface based on natural language processing (NLP) technology has been widely used in the world, especially in smart home, vehicle, wearable equipment, robots and other fields. Take smart home as an example, Amazon and Baidu have released the smart speakers with screen that enable multimodal interaction. In addition to touch, users can also instruct the devices to play video,

A. Marcus and W. Wang (Eds.): HCII 2019, LNCS 11586, pp. 408–419, 2019.
https://doi.org/10.1007/978-3-030-23535-2_30

play music and search information using human natural language. Voice has become the dominant modality for people to interact with device, to exchange information or to express their intentions [2]. During the process of voice interaction, there are usually two basic interaction stages: voice wake-up and voice dialogue. People need to first trigger the automatic speech recognition system through voice wake-up, and then input voice queries to start dialogue with the device [3, 4].

There are many studies indicating response delay is a key element for evaluating the quality of multimodal interface, and the long response delay can seriously reduce naturalness and fluency of the interaction, and even affect people's willingness to use the device [5–7]. Meanwhile, some researchers suggest that response delay is the most important factor to determine user satisfaction [8]. Although the importance of response delay has been proved, unfortunately, most researchers only focused on providing qualitative descriptions and rarely investigated the specific optimum response delay time or an acceptable range. Moreover, the reasons for response delay of voice interactive devices may be related to the efficiency of ASR, quality of language model, network conditions [9–11]. Continuously improving the responsiveness of voice interaction and reducing response delay time are critical to enhance the usability of voice interactive devices [12].

Targeted at multimodal interface with voice as dominant modality, this paper evaluated response delay of two smart devices with screen, a smart speaker with screen and a smart TV. Compared with the voice input and voice output interaction of smart device without screen, the smart device with screen becomes richer in output interaction, for example, the device can directly display the results of the voice recognition or voice search on screen. In this paper, firstly, we built a research framework for multimodal interface according to users' perception of the voice interaction process, in which the system output process was divided into three successive stages. Secondly, two experiments were conducted to evaluate and quantify response delay time in each stage, the purpose of the first experiment was to measure automatic speech recognition (ASR) feedback delay, and the second experiment was to measure query response delay and loading response delay. Finally, discussions and conclusions of this study would be described.

2 Literature Review

Many researchers have suggested several typical parameters for describing multimodal human-computer interaction including Accuracy (words, gestures, etc.) [13, 14], Delay [15], Efficiency [16], Appropriateness [17, 18], etc. Weiss [5] distinguished system output delay into two types: feedback delay and response delay. Feedback delay refers to the delay when the system successfully receives user input, which is described as average delay from the end of user input to the beginning of system feedback, for example from button press to display of loading status in terms of clock. Response delay refers to the delay of system responding to user input, it is described as average delay of a user response, from the end of user input to the beginning of system output. For example, the system starts to display a new GUI. In addition, Bernsen [17] proposed another time-related parameter named lag of time, which refers to the asynchronism between corresponding modalities.

Many studies have shown that response delay has a negative effect on user satisfaction towards device or service [19–22]. To measure delay time and its impact on user attitudes, behavior and psychological state, a lot of work has been done in graphic user interface (GUI) domain. Galletta and colleagues [23] examined the impact of different website delay times of 0, 2, 4, 6, 8, 10, and 12 s in an experiment, and the findings suggested that the tolerable waiting time was around 4 s. Nah [24] reviewed the literature on computer response time and assessed web user's tolerable waiting time in information retrieval, the results from the study suggested that the tolerable waiting time for information retrieval was approximately 2 s. Wang [25] explored optimal system response time which would make human-information system (HIS) interaction most efficient and found that the interaction efficiency of HIS was the highest when the response time was in the range of 0.25–0.75 s, while a response time less than 0.25 s was likely to make users feel stressed and nervous. Research on response delay time of voice user interface (VUI) is limited, McWilliams [26] studied the impact of voice interface turn delays on drivers' attention and arousal levels in vehicles settings, the results showed that a delay time longer than 4 s was associated with decreased attention to the driving task. This suggested that system delay time under 4 s may be optimal.

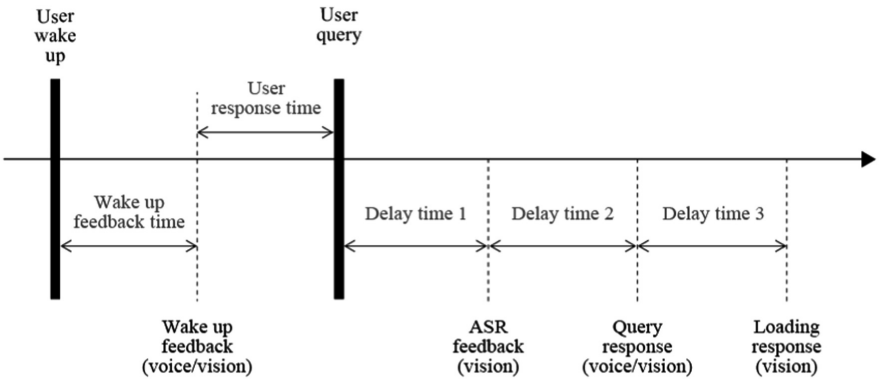

Fig. 1. Four types of system output delay time in voice wake-up and voice dialogue stage.

So far, there are few studies that focus on measuring delay time of multimodal voice user interface. Moreover, no research is identified that investigated response delay in scenarios other than vehicles or driving scenarios, such as in smart home. This study aimed to measure delay time of multimodal voice user interface in two smart home devices with different screen sizes, a smart speaker with 7-in. screen and a smart TV with 55-in. screen were involved in this study. First, according to the user's perception of system output in both wake-up stage and dialogue stage, the system output delay was divided into four types (See Fig. 1): wake-up feedback delay in voice wake-up stage, and ASR feedback delay, query response delay and loading response delay in voice dialogue stage. There are usually differences in wake-up between smart

speaker with screen and smart TV. The former mainly uses voice wake-up, while the latter uses button wake-up. Response delay under different wake up modes should be studied separately. Therefore, this study focused on the delay times of voice dialogue stage, and according to user perception, we can divide the voice dialogue stage into several aspects. On the one hand, ASR feedback delay maybe happen when user input voice query. On the other hand, system output delay maybe happens after ASR feedback had finished, which included query response delay and loading response delay. Their descriptions and operational definitions were as follows:

ASR feedback delay time: the delay time from the end of user's input voice query to the beginning of the display of ASR feedback on the screen, for example from the end of user saying "I want to see movie" to the beginning of the text "I want to see movie" displaying on the screen.

Query response delay time: the delay time from the end of ASR results display to the beginning of system output, for example from the end of the screen displaying "I want to see movie" to the beginning a new GUI displays the movie information on the screen.

Loading response delay time: the delay time from the beginning of a new GUI displays on the screen to the end of all results being displayed completely, for example all movie information being shown on the screen.

Next, we conducted two experiments to evaluate the impact of different delay time on users' satisfaction in all three stages mentioned above. The first experiment is designed to investigate ASR feedback delay, which to some extent could reveal the performance of automatic speech recognition technology. And the second experiment mainly focused on system output delay after ASR feedback completed, which could indicate the performance of the system output for user's query or need.

3 Experiments

3.1 Experiment 1: ASR Feedback Delay Time Experiment

Objectives
When having a dialogue with smart device with screen, the query that user input will be displayed on the screen. Currently, display mode of query is mainly real-time, and words recognized by the devices will be shown immediately on the screen while user is inputting. The aim of this experiment was to explore the satisfying and acceptable delay time of ASR feedback in the real-time display mode.

Subjects
In total, 30 subjects participated in the experiment (M = 25.8 years old, SD = 4.37), all of them were employees of internet companies. The sample consisted of 15 females and 15 males, and half of them reported previous experience with smart speaker or smart TV.

Tasks and Materials

We developed an experimental program with Java, which could run on a smart speaker (7 in.) and a smart TV (55 in.), display mode of ASR feedback in the program is real-time way. In the experiment, we arranged movie searching tasks to subjects. In order to cover different length queries as far as possible, three length queries in Chinese were provided in the experiment. They were short query with four Chinese words "Kungfu movie", middle query with ten Chinese words "I want to see movie rating above nine score" and long query with twenty Chinese words "I want to see Andy Lau's Hong Kong movie before 2010 year". Subjects first need to say queries in various delay time conditions and pay attention to when the queries were recognized and displayed on the screen, and then give their scores of satisfactions for the ASR feedback delay time. In order to balance the learning effect and fatigue effect in the experiment, we randomized the sequences of tasks for each subject, the three query lengths were grouped and randomized, and then the delay time was completely randomized. An interview was conducted to collect qualitative data, such as reasons of their evaluations and other comments.

Experimental Variables

The independent variables of the experiment included device screen size (smart speaker with 7-in. screen and smart TV with 55-in. screen), query length (short, middle, long) and ASR feedback delay times (0, 200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2000 ms), and the levels of independent variables referred to the range of relevant parameters of current smart devices. Dependent variables were user's satisfaction evaluation (1-Very Dissatisfied, 2-Dissatisfied, 3-General, 4-Satisfied, 5-Very Satisfied). We used a 2 (screen size) *3 (query length) *11 (ASR feedback delay time) three-factor mixed design.

Results

Descriptive results of the ASR feedback delay time experiment are shown in Table 1. The results of Repeated Measures ANOVA showed that the main effect of screen size was not significant ($F(1,28) = 0.162$, $p = 0.639$), the main effect of different length queries was not significant ($F(2,27) = 1.198$, $p = 0.317$), the main effect of delay time was significant ($F(10,19) = 12.180$, $p < 0.001$), the interaction effect between different length queries and delay time was not significant ($F(20,9) = 1.911$, $p = 0.159$), and the interaction effect between screen size and delay time was not significant ($F(10,19) = 1.424$, $p = 0.243$). Multiple Comparative Analysis showed that there were no significant differences ($p > 0.05$) between 0&200 ms, 0&400 ms, 200&400 ms, 600&800 ms, 800&1000 ms, 1200&1400 ms, 1200&1600 ms, 1200&1800 ms, 1400&1600 ms and 1600&1800 ms, the others were significant differences($p < 0.05$). These results indicated that subjects felt very fast when the ASR feedback delay time was less than 400 ms, and subjects began to feel slow when the delay time was more than 1200 ms.

Table 1. Mean and standard deviation of satisfaction scores for ASR feedback delay time.

Delay times	Mean satisfaction	SD	95% confidence interval	
			Lower	Upper
0 ms	4.24	0.12	4.01	4.48
200 ms	4.12	0.11	3.90	4.35
400 ms	4.22	0.11	3.99	4.45
600 ms	3.87	0.10	3.66	4.07
800 ms	3.72	0.12	3.47	3.97
1000 ms	3.56	0.12	3.32	3.79
1200 ms	3.26	0.13	2.99	3.52
1400 ms	3.21	0.13	2.94	3.48
1600 ms	3.10	0.14	2.82	3.38
1800 ms	2.99	0.13	2.73	3.48
2000 ms	2.67	0.14	2.38	3.38

Table 2. Regression equation of satisfaction to ASR feedback delay time.

Regression equation	F	R^2	Satisfying delay time	Acceptable delay time
Satisfaction = −0.000805 * delay time + 4.347	321.511***	24.5%	431 ms	1673 ms

In this experiment, we regarded satisfaction score "4-Satisfied" as the lower limit of satisfying delay time, and "3-General" as the lower limit of acceptable delay time. We further had a Linear Regression Analysis of ASR feedback delay and user satisfaction, and the results as shown in Table 2. The final results suggested that subjects were satisfied when the delay time was less than 431 ms, and subjects felt acceptable when it was less than 1673 ms, and ASR feedback delay time of more than 1673 ms might dissatisfy users.

3.2 Experiment 2: Query Response Delay Time and Loading Response Delay Time Experiment

Objectives

After the first experiment, we conducted the second experiment. The aim of the second experiment was to measure users' satisfying and acceptable delay time respectively in both query response stage and loading response stage respectively, and further explored the total response delay time of the two stages.

Subjects

32 subjects participated in the experiment (M = 27.0 years old, SD = 3.21), all of them were employees of internet companies. The sample consisted of 16 females and 16 males, and half of them reported previous experience with smart speaker or smart TV.

Tasks and Materials

We developed a program for the second experiment, which could run on a smart speaker (7 in.) and a smart TV (55 in.). As the experiment mainly focused on measuring query response delay and loading response delay, we fixed the feedback delay time of ASR and the search query in this experiment. We arranged movie searching tasks and a short query with four Chinese words was provided which was "Kungfu Movie". Before the experiment, there was a warm up session to introduce about the query response delay and loading response delay. In the experiment, subjects were asked to say the query in various delay time conditions. After subjects finished input the query and it was displayed, a new GUI would be shown on the screen by dynamic loading way and the movies information obtained would be displayed one by one. Subjects need to pay attention to the query response delay time, the loading response delay time, and total delay time, and then gave their scores of satisfactions for the above three delay times respectively. In order to balance the learning effect and fatigue effect in the experiment, we randomized the sequences of tasks for each subject, the query response delay time and the loading response time were completely randomized in the experiment. After the experiment, an interview was conducted to explore the reasons of their evaluations.

Experimental Variables

The independent variables of the experiment included device screen size (smart speaker with 7-in. screen and smart TV with 55-in. screen), query response delay time (200, 600, 1000, 2000, 3000, 4000, 5000 ms), and loading response delay time (200, 600, 1000, 2000, 3000, 4000, 5000 ms), and the levels of independent variables referred to the range of relevant parameters of current smart devices. Dependent variables were user's satisfaction evaluation of query response delay, loading response delay and total response delay (1-Very Dissatisfied, 2-Dissatisfied, 3-General, 4-Satisfied, 5-Very Satisfied), the total response delay time was the sum of the above two kinds of delay time. We used a three-factor mixed design of 2 (screen size) *7(query response delay time) *7(loading response delay time).

Results

Query Response Delay Time

From the descriptive analysis, we found the relationship between query response time and user satisfaction, as shown in Table 3. The results of Repeated Measures ANOVA with satisfaction as dependent variable showed that the main effect of screen size was not significant ($F(1,29) = 0.00$, $p = 0.987$), the main effect of query response delay time was significant ($F(6,24) = 122.674$, $p < 0.001$), the interaction effect between screen size and query response delay time was not significant ($F(6,24) = 0.796$, $p = 0.582$), and Multiple Comparison Analysis suggested that there were significant differences in different query response delay time ($p < 0.05$). We regarded satisfaction score "4-Satisfied" as the lower limit of delay time to satisfy users, and "3-General" as the lower limit delay time acceptable to users. Further Linear Regression Analysis of query response delay time and user satisfaction, as shown in Table 4, the final results indicated that subjects were satisfied when the delay time was less than 867 ms, and subjects felt it was acceptable when the delay time was less than 2537 ms.

Table 3. Mean and standard deviation of satisfaction scores for query response delay time.

Delay times	Mean satisfaction	SD	95% confidence interval	
			Lower	Upper
200 ms	4.47	0.61	4.39	4.55
600 ms	4.16	0.70	4.07	4.25
1000 ms	3.91	0.91	3.79	4.04
2000 ms	3.37	0.98	3.24	3.5
3000 ms	2.57	0.97	2.44	2.69
4000 ms	1.98	0.94	1.85	2.1
5000 ms	1.71	0.87	1.59	1.82

Table 4. Regression equation of satisfaction to query response delay time.

Regression equation	F	R^2	Satisfying delay time	Acceptable delay time
Satisfaction = −0.000599 * delay time + 4.519434	20187.18***	57.1%	867 ms	2537 ms

Loading Response Delay Time

The mean satisfaction scores in different delay time conditions are shown in Table 5. The results of Repeated Measures ANOVA showed that the main effect of screen size was not significant ($F(1, 29) = 0.050$, $p = 0.825$), the main effect of loading response delay time was significant ($F(6, 24) = 69.772$, $p < 0.001$), the interaction effect between screen size and query response delay time was not significant ($F(6, 24) = 2.163$, $p = 0.083$). Multiple Comparison Analysis found that in different loading response delay time condition, there were significant differences of subjects' satisfaction scores ($p < 0.05$). We regarded satisfaction score "4-Satisfied" as the lower limit of delay time to satisfy users, and "3-General" as the lower limit delay time which was acceptable to users. Further Linear Regression Analysis for loading response delay time and user satisfaction, as shown in Table 6, the final results indicated that subjects were satisfied when the delay time was less than 564 ms, and subjects felt it was acceptable when the delay time was less than 2353 ms.

Table 5. Mean and standard deviation of satisfaction scores for loading response delay time.

Delay times	Mean satisfaction	SD	95% confidence interval	
			Lower	Upper
200 ms	4.43	0.67	4.34	4.52
600 ms	4.13	0.72	4.03	4.22
1000 ms	3.62	0.81	3.51	3.72
2000 ms	2.96	0.74	2.86	3.05
3000 ms	2.43	0.71	2.33	2.52
4000 ms	2.04	0.69	1.95	2.14
5000 ms	1.77	0.63	1.69	1.86

Table 6. Regression equation of satisfaction to loading response delay time.

Regression equation	F	R^2	Satisfying delay time	Acceptable delay time
Satisfaction = −0.000559 * delay time + 4.315371	2533.5***	61.8%	564 ms	2353 ms

Total Response Delay Time

The results of Repeated Measures ANOVA with overall satisfaction as dependent variable showed the main effect of screen size was not significant ($F(1,23) = 0.978$, $p = 0.333$), the main effect of query response delay time was significant ($F(4,23) = 76.557$, $p < 0.001$), the main effect of loading response delay time was significant ($F(4,23) = 122.844$, $p < 0.001$), the interaction between query response delay time and loading response delay time was significant ($F(16,8) = 3.849$, $p < 0.05$), and the interaction between the other variables was not significant ($p > 0.05$). Simple Effect Analysis was further conducted since the interaction effect between query response delay time and loading response delay time was significant, as shown in Fig. 2. The results indicated that subjects were not satisfied with the short query response delay but long loading response delay, or the long query response delay but short loading response delay. We further had a Linear Regression Analysis for total response delay and user satisfaction, and the results as shown in Table 7.

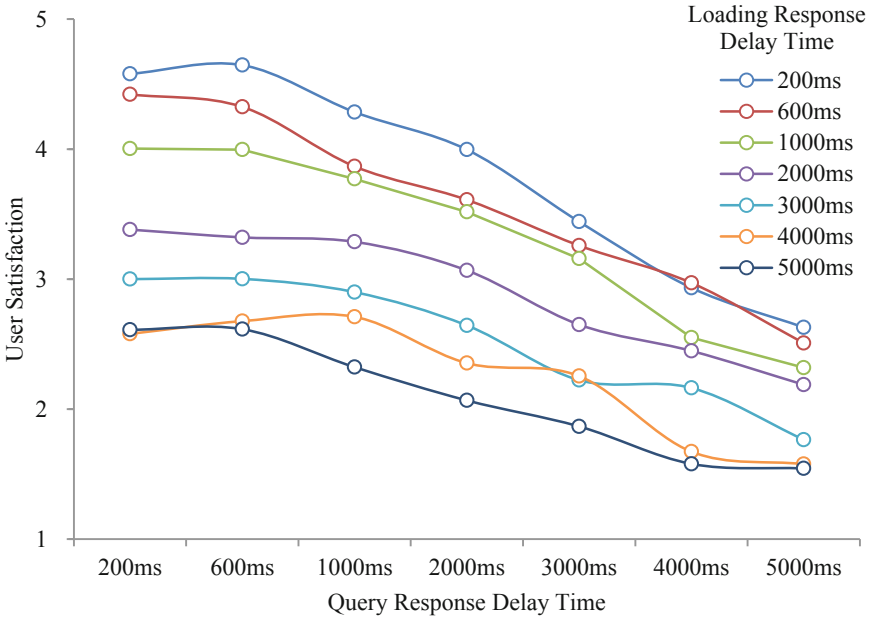

Fig. 2. Interaction diagram of query response delay time and loading response delay time.

Table 7. Regression equation of satisfaction to total response delay time.

Condition	Regression equation	F	R^2
Non-standardized regression equation	Satisfaction = −0.000403 * query response delay time−0.000449 * loading response delay time + ($4.08E^{-8}$) * query response delay time * loading response delay time + 4.4428	598.209***	53.4%
Standardized regression equation	Satisfaction = −0.613 * query response delay time − 0.683 * loading response delay time + 0.22 * query response delay time * loading response delay time	598.209***	53.4%

4 Discussion

Different from previous studies on response delay [23–26], this paper decomposed the system output into several key response delay stages according to user perception, which included ASR feedback delay, query response delay and loading response delay. In order to support corresponding technical optimization, we measured satisfying and acceptable delay time for users.

The results of ASR feedback delay experiment showed that users' perception of delay time was not influenced by screen size or the length of queries subjects input. From the end of users' input to the being query displayed on the screen, subjects were satisfied when the ASR feedback delay time was less than 431 ms and thought it was acceptable when the delay time was less than 1673 ms. It should be noted that in addition to the real-time display mode, there is also non-real-time display mode, in which the query will be displayed integrally on the screen until all speech recognition is finished. Based on Barnett's [27] and Tom's [28] research, the perception of delay time was affected by user's subjective factors, so the ASR feedback display mode might affect user's subjective perception. We suggest that the satisfied and acceptable delay time under non-real-time display mode should be further investigated.

The results of query response delay and loading response delay experiment showed that the screen size of smart devices still did not affect the perception of response delay time. In query response stage, subjects were satisfied when delay time was less than 867 ms and felt it was acceptable when the delay time was less than 2537 ms. In loading response stage, subjects were satisfied when delay time was less than 564 ms and feel acceptable when the delay time was less than 2353 ms. Comparing with query response delay time, subjects showed higher requirement of faster for loading response delay, that was, the delay time in loading response stage needed to be shorter than in query response stage for them to feel satisfied. Combining the interviews after the second experiment, subjects generally thought that there were different reasons for the response delay in the two stages. Some participants thought that the delay time of query response time was mainly affected by the performance of the product, specifically it was related to search algorithm and efficiency, while the delay time of loading response time was mainly affected by network conditions. Subjects were more tolerant of search algorithm and efficiency than network condition. In addition, it should be mentioned

that only video search scenario was involved in the second experiment, and the system output usually included a large amount of multimedia information such as pictures, graphs and texts. For Encyclopedia queries, weather queries and other text-based output scenarios, the response delay should also be further studied.

5 Conclusion

This paper targeted at multimodal interface with voice as dominant modality, we built a research framework to evaluate response delay according to user perception during the voice interaction, in which the system output process was divided into three successive stages. We conducted two experiments to evaluate the response delay in each stage, a smart speaker with screen and a smart TV were involved in the experiment, the first experiment was designed to measure automatic speech recognition (ASR) feedback delay time. Results indicated that, from the end of user input to query displayed on the screen, users were satisfied when the ASR feedback delay time was less than 431 ms and felt it was acceptable when the delay time was less than 1673 ms. In the second experiment, we aimed to measure query response delay time and loading response delay time. We found that users were satisfied when delay time was less than 867 ms and felt it was acceptable when the delay time was less than 2537 ms in query response stage. In loading response stage, users were satisfied when delay time was less than 564 ms and felt it was acceptable when the delay time was less than 2353 ms.

Acknowledgements. We thank R&D engineers in Baidu for supporting our research and developing the experimental programs. This research was supported by Baidu Duer Business Unit (BU) for smart devices and parameters.

References

1. Kühnel, C., Weiss, B., Möller, S.: Parameters describing multimodal interaction-Definitions and three usage scenarios. In: Eleventh Annual Conference of the International Speech Communication Association (2010)
2. Amrutha, S., Aravind, S., Mathew, A., Sugathan, S., Rajasree, R., Priyalakshmi, S.: Voice controlled smart home. Int. J. Emerg. Technol. Adv. Eng. **5**(1), 596–600 (2015)
3. Wang, X., Guo, Y., Ge, F., Wu, C., Fu, Q., Yan, Y.: Speech-picking for speech systems with auditory attention ability. Sci. China **45**(10), 1310–1327 (2015)
4. Pearl, C.: Designing Voice User Interfaces: Principles of Conversational Experiences, pp. 1–10. O'Reilly Media, Newton (2016)
5. Weiss, B., Scheffler, T., Möller, S.: Describing multimodal human computer interaction. In: Proceedings of Workshop at NordiCHI: Assessing Multimodal Interaction (aMMi), Copenhagen, Denmark, pp. 33–36 (2012)
6. ITU Supplement 25 to P-Series Rec: Parameters describing the interaction with multimodal dialogue systems. In: International Telecommunication Union, Geneva (2011)
7. Shriberg, E., Wade, E., Price, P.: Human-machine problem solving using spoken language systems (SLS): factors affecting performance and user satisfaction. In: Proceedings of the DARPA Speech and NL Workshop, pp. 49–54 (1992)

8. Johnson, J.: Designing with the Mind in Mind: Simple Guide to Understanding User Interface Design Guidelines, pp. 129–146. Morgan Kaufmann, Waltham (2010)
9. Chen, S., Jiang, Q.: Ergonomic study on the performance of speech recognition system. Space Med. Med. Eng. **3**(3), 216–221 (1990)
10. Zhong, W., Xu, B.: Performance evaluation criteria and affecting factors of speech input system. In: NCMMSC1994, pp. 452–456 (1994)
11. Cui, S.: Strategies for improving efficiency of speech recognition. Master's degree thesis of Beijing University of Posts and Telecommunications, pp. 16–18 (2018)
12. Chen, Y., Li, K., Zhou, J., Liu, J., Liu, R.: Novel efficient algorithms in speech/speaker recognition. Comput. Eng. **30**(15), 1–3 (2004)
13. Simpson, A., Fraser, N.M.: Black box and glass box evaluation of the SUNDIAL system. In: Proceedings 3rd European Conference on Speech Communication and Technology (Eurospeech 1993), DE-Berlin, vol. 2, pp. 1423–1426 (1993)
14. Nigay, L., Coutaz, J.: A design space for multimodal systems: concurrent processing and data fusion. In Proceedings of INTERACT & CHI, pp. 172–178 (1993)
15. Price, P.J., Hirschman, L., Shriberg, E., Wade, E: Subject-based evaluation measures for interactive spoken language systems. In: Proceedings of DARPA Speech and Natural Language Workshop, US-Harriman CA, pp. 34–39 (1992)
16. Perakakis, M., Potamianos, A.: Multimodal system evaluation using modality efficiency and synergy metrics. In Proceedings of IMCI, pp. 9–16 (2008)
17. Bernsen, N.: From theory to design support tool. In: Multimodality in Language and Speech Systems. Kluwer, Dordrecht, pp. 93–148 (2002)
18. Bernsen, N., Dybkjær, L.: Multimodal Usability. Springer, London (2009)
19. Weinberg, B.D.: Don't keep your internet customers waiting too long at the (virtual) front door. J. Interact. Mark. **14**(1), 30–39 (2000)
20. Pruyn, A., Smidts, A.: Effects of waiting on the satisfaction with the service: beyond objective time measures. J. Int. Res. Mark. **15**(4), 0–334 (1998)
21. Thompson, D.A., Yarnold, P.R., Williams, D.R., Adams, S.L.: Effects of actual waiting time, perceived waiting time, information delivery, and expressive quality on patient satisfaction in the emergency department. Ann. Emerg. Med. **28**(6), 657–665 (1996)
22. Lee, Y., Chen, A.N., Ilie, V.: Can online wait be managed? The effect of filler interfaces and presentation modes on perceived waiting time online. MIS Q. 365–394 (2012)
23. Galletta, D.F., Henry, R., McCoy, S., Polak, P.: Web site delays: how tolerant are users. J. Assoc. Inf. Syst. **5**(1), 1–28 (2004)
24. Nah, F.: A study on tolerable waiting time: how long are web users willing to wait. In: Americas Conference on Information Systems, pp. 2212–2222 (2003)
25. Wang, H., Yi, S., Yang, W., Di, J.: The influence of system response time on human-information system interaction efficiency. China J. Ergon. **13**(3), 4–13 (2007)
26. McWilliams, T., Reimer, B., Mehler, B., Dobres, J., McAnulty, H.: A secondary assessment of the impact of voice interface turn delays on driver attention and arousal in field conditions. In: Proceedings of the Eighth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, pp. 414–420 (2015)
27. Barnett, A., Saponaro, A.: Misapplications reviews: the parable of the red line. Interfaces **15**(2), 33–39 (1985)
28. Tom, G., Lucey, S.: A field study investigating the effect of waiting time on customer satisfaction. J. Psychol. **131**(6), 655–660 (1997)