The Effects of Continuous Conversation and Task Complexity on Usability of an AI-Based Conversational Agent in Smart Home Environments

Chapter *in* Lecture Notes in Electrical Engineering · January 2020

DOI: 10.1007/978-981-13-8779-1_79

3 authors, including:

Jingya Guo
The University of Hong Kong
5 PUBLICATIONS **66** CITATIONS

SEE PROFILE

The Effects of Continuous Conversation and Task Complexity on Usability of an AI-Based Conversational Agent in Smart Home Environments

Jingya Guo, Da Tao and Chen Yang

Abstract Conversational agents have gained increasing popularity over the last decade in a variety of personal, public, and occupational settings due to rapid advances of artificial intelligence (AI) and natural language processing (NLP). However, how users can interact with such technologies is still understudied. The objective of this study was to investigate type of conversation (presence and absence of continuous conversation) and task complexity (high vs. low) on usability metrics (i.e., task completion time, number of queries used in completing tasks, and perceived system usability) with conversational agents in smart home environments. Eighteen participants joined this study and completed required tasks. The results showed that there was a significant effect of type of conversation on task completion time and number of queries per task. Task complexity significantly extended task completion time and increased number of queries per task. The results may help with the design of more usable conversational agents.

Keywords Continuous conversation · Conversational agent · Smart home · Usability

1 Introduction

The development of artificial intelligence and natural language understanding has captured the public attention and led to widespread acceptance of AI-infused conversational agents [1]. World-leading technology companies have launched

J. Guo · D. Tao (✉)
Institute of Human Factors and Ergonomics College, Shenzhen University, Shenzhen 518060, China
e-mail: taoda@szu.edu.cn

J. Guo · C. Yang
International User Experience Business Unit-Business Experience Research, Alibaba Group, Hangzhou, China

S. Long and B. S. Dhillon (eds.), *Man–Machine–Environment System Engineering*, Lecture Notes in Electrical Engineering 576,
https://doi.org/10.1007/978-981-13-8779-1_79

multiple programs that aim to develop intelligent devices embodied conversational agents, among which Alexa, Siri, Cortana and Google Assistant are the most commonly used [1–3].

A conversational agent is defined as a system that relies on natural language processing, supported by artificial intelligence, and to have verbal interactions with end users [4]. A number of studies have made efforts to figure out how conversational agents can be incorporated into real homes. With natural dialogue as input language, designers would like to see more communications between users and conversational agents [5–7]. Furthermore, users expected intelligent and human-like chatbots. However, previous versions of devices should repeatedly wake up for each single command [8, 9], which enormously sacrificed fluency, increased cognitive workload, and thus decreased user experience.

On March 2018, Amazon.co first enabled continuous conversation on Alexa, then followed by Google home three months later. By using the continuous conversation feature, users no longer need to say wake word prior to every command; instead, the conversational agent would continue to listen for up to five or eight seconds for any follow-up commands.

Continuous conversation was believed to have the potential to improve usability of a natural language interface. However, usability of conversational agents with continuous conversation has seldom been evaluated [10]. While previous studies mostly focused on the history log of for consumers without any restriction on the devices and tasks using conversational agents, there is a lack of research to describe how consumers interact with conversational agents for smart home device control tasks. To address these research gaps, this study aimed to examine the usability of a conversational agent with continuous conversation in a simulated smart home environment.

2 Methodology

2.1 Experiment Design

A two-factor (2 × 2) within-subject design was employed in this study, with continuous conversation (on vs. off), and task complexity (simple vs. complex) as independent variables, and a set of usability metrics (i.e., task completion time, number of queries used in completing task, and perceived usability metrics) as dependent variables. Task scenarios were presented in a simulated smart home environment.

2.2 Participants

Participants were recruited via posters by inner-company social media network in Alibaba Group. Eligible participants were invited if they were (1) native Chinese speaker with no regional accent and (2) aged 18 or above. This study recruited eighteen participants (3 female and 15 male), aged from 26 to 40 years (mean = 29.28, SD = 3.51).

2.3 Materials

A smart home environment with conversational agents, five light bulbs, an electric fan, a television, and a couch for single person was established as experimental scenarios in this study. Two commercial smart speakers with conversational agent (Tmall Genie X1, version 3.10.0) were used to listen to commands from participants and help them complete required tasks. When previous order was completed, one of the agents was able to stay awake for eight seconds and determine if participants continued to talk to it (the version with continuous conversation on). The other conversational agent was only able to detect any order each time when it was awaked by saying "Hi, Tianmao jingling!" (the version with continuous conversation off). The luminance (seven levels from dark to bright) and color temperature (seven levels from cool to warm) of light bulbs were adjustable. The television was consisted of a high-resolution display (32-inch, 1920 × 1080 pixels) connected with a Tmall Box Office and was used to provide video contents to participants. The smart light, fan, and Tmall Box Office could be controlled by the conversational agent using mandarin. All devices were settled near and in front of a cozy couch (see Fig. 1).

Fig. 1 Simulated smart home environment

2.4 Procedures and Tasks

The written consent and detailed introduction of test procedures were provided to participants before the experiment. Participants were instructed to sit and adjust the sofa according to their preference, get familiar with the conversational agent and smart home devices, and practice orders that should be used to control smart devices. Then they were asked to complete four experimental tasks. The experimental tasks included single device task (simple task) and multiple devices task (complex task) with continuous conversation on or off.

All four tasks were designed based on typical home environment activities in use of conversational agents. In single tasks, participants were asked to turn on the smart light, adjust illumination first to the brightest then according to their preference, and adjust the color of light a little bit warmer using conversational agents. The complex tasks required participants to turn on the smart light, fan, and smart television using conversational agents and adjust these devices until they felt cozy and relaxed. Participants were asked to speak naturally to the conversational agents. A practice session was provided to assist participants to familiarize themselves with the tasks and devices. Upon the completion of each experimental task, a paper-based questionnaire was administered to collect participants' response to perceived system usability metrics. The whole experiment took approximately forty minutes.

2.5 Dependent Measures and Data Analysis

A set of measures, including task completion time, number of queries used in completing task, and perceived system usability (i.e., perceived confidence, perceived intention to use, perceived ease of use, perceived effort, and perceived ease of learning) were used to assess the effects of continuous conversation and task complexity. Task completion time was the total time that participants spent in experimental tasks. Number of queries used in completing each task was the number of order languages that users give to the conversational agents in the experimental tasks. Wake words were deducted from the number of queries.

We adopted previously validated system usability metrics scales [11] and revised them to fit our study scenarios. The scales used included perceived confidence, perceived intention to use, perceived ease of use, perceived effort, and perceived ease of learning. They were rated on five-point Likert-type scales from 1 (not agree at all) to 5 (extremely agree).

Repeated measures analysis of variance (ANOVA) was used to determine the effects of continuous conversation and task complexity on task completion time, number of queries, and perceived system usability. Post hoc multiple comparisons were performed with Bonferroni adjustment where necessary. Statistical analyses were performed with IBM SPSS 24 (Chicago, Illinois, USA).

3 Results

3.1 Objective Measures

Table 1 presents ANOVA results for task completion time and number of queries. Continuous conversation had a significant effect on the number of queries per task ($F(1,17) = 5.773$, $p = 0.03$) while task complexity had a significant effect on both the number of queries per task ($F(1,17) = 28.404$, $p < 0.001$) and task completion time ($F(1,17) = 13.297$, $p = 0.002$). There was no significant interaction effect between continuous conversation and task on average queries per task ($F(1,17) = 4.448$, $p = 0.052$). The interaction effect on task completion time was significant ($F(1,17) = 7.914$, $p = 0.013$).

3.2 Perception Measures

Table 2 presents ANOVA results for perceived system usability metrics. Simple tasks increased perceived confidence ($F(1,17) = 4.10$, $p = 0.06$) and were rated as easier compared with complex tasks ($F(1,17) = 7.57$, $p = 0.02$). Continuous conversation had no significant effect on perceived system usability metrics and its subscales.

Table 1 The effects of continuous conversation and task complexity on task completion time and number of queries per task

Descriptive analysis			ANOVA		
	Mean	SD	F values	p values	η^2
Number of queries per task					
Continuous conversation			5.77	0.03	0.28
Off	10.97	0.87			
On	15.53	1.62			
Task			28.40	<0.001	0.65
Simple	9.78	1.12			
Complex	16.72	1.08			
Task completion time					
Continuous conversation			1.24	0.28	0.08
Off	121.81	13.96			
On	159.17	25.16			
Task			28.40	<0.001	0.65
Simple	92.94	13.69			
Complex	188.04	20.39			

Table 2 The effects of continuous conversation and task complexity on perception measures

Descriptive analysis			ANOVA		
	Mean	SD	*F* values	*p* values	η^2
System Usability (0–100)					
Continuous conversation			0.88	0.36	0.06
Off	61.94	4.10			
On	66.25	3.54			
Task			2.51	0.13	0.14
Simple	65.83	3.04			
Complex	62.36	3.48			
Confidence (0–5)					
Continuous conversation			0.28	0.60	0.02
Off	2.81	0.24			
On	2.69	0.23			
Task			4.10	0.06	0.22
Simple	2.86	0.21			
Complex	2.64	0.23			
Intention (0–5)					
Continuous conversation			0.01	0.92	0.00
Off	2.92	0.27			
On	2.89	0.25			
Task			0.59	0.46	0.04
Simple	2.97	0.21			
Complex	2.83	0.27			
Ease of use (0–5)					
Continuous conversation			0.12	0.74	0.01
Off	2.59	0.21			
On	2.68	0.13			
Task			0.03	0.86	0.00
Simple	2.62	0.12			
Complex	2.65	0.16			
Effort (0–5)					
Continuous conversation			2.96	0.11	0.17
Off	1.99	0.21			
On	2.45	0.21			
Task			3.63	0.08	0.20
Simple	2.32	0.16			
Complex	2.12	0.17			
Ease of learning (0–5)					
Continuous conversation			0.38	0.55	0.03
Off	2.65	0.20			
On	2.76	0.15			

(continued)

Table 2 (continued)

Descriptive analysis			ANOVA		
	Mean	SD	F values	p values	η^2
Task			7.57	0.02	0.34
Simple	2.83	0.14			
Complex	2.58	0.17			

4 Discussion

In spite of increased accessibility of conversational agents for consumers, there is still much work to do before conversational agents could be well fitted in smart home environments. In light of this, the present study reported results from a preliminary usability study of conversational agents in a simulated smart home environment while completing either simple or complex device control tasks. This finding suggested that continuous conversation made users talk more and faster, but did not significantly increase the perceived usability metrics in general. Using the conversational agent to control multiple devices was perceived significantly less confidence and more difficult to learn.

Our study found that continuous conversation significantly increased average number of queries participants used in both simple and complex tasks but not the time needed to complete them. The significant interaction of continuous conversation and task complexity on completion time suggested that continuous conversation decreased the time spent on single device control tasks thus increased the efficiency. Perceived system usability metrics and its subscales were not significantly improved by continuous conversation. This might be because the current application of continuous conversation only eliminated the wake word, which did not meet user expectation of a true conversation where AI-infused agents could understand and memorize the context and connotation of what users say [12].

Complex tasks (e.g., control of multiple devices) significantly increased number of queries and task completion time compared with simple tasks (e.g., control of a single device). Users reported less confidence and perceived it was more difficult to learn to control multiple devices. The system usability metrics were not significantly affected by task complexity, indicating that even though users may spend more time and efforts to get familiar with conversational agents, they still perceived the conversational agent useful and would like to use them. Designers should pay more attention to the way to help consumers grasp correct queries and compatible functions as quickly and easily as possible, especially when there were multiple devices connected with a conversational agent.

Our findings have important implications for the design of conversational agents for smart home device control. First, conversational agents' designers should recognize that different conversation policies (e.g., with or without continuous conversation) can lead to differential effects on user behaviors (e.g., more or less commands users would give) in the use of conversational agents. In particular, our results provide

justification for the adoption of continuous conversation (e.g., eliminate wake-up words by stay activated automatically for 8 s after completing previous command) can facilitate user interaction with the devices. Second, it should be noted that in our study, simple tasks received more benefits from continuous conversation compared with complex tasks, while the reason for this remains unclear. Therefore, caution should be exercised in the application of continuous conversation under practical settings when there were multiple devices to be controlled.

Several limitations of our study should be acknowledged and deserve further consideration. First, more comprehensive usability studies with more users are required to achieve valid and more fine-grained conclusions. Second, future studies are recommended to examine individual difference and other scenario-related factors to allow for more understanding of natural language interface for specific groups and settings.

5 Conclusion

This study examined the effects of continuous conversation and task complexity on interaction with an AI-infused conversational agent in a simulated smart home environment. In general, continuous conversation encouraged human–agent interaction. The number of queries per task increased as continuous conversation switched on where wake words could be eliminated when multiple commands were made. It is suggested that future conversational agent could apply continuous conversation as a default setting. More work is needed to understand how to improve perceived usability of a conversational agent in complex tasks.

Compliance with Ethical Standards The study was approved by the Logistics Department for Civilian Ethics Committee of Alibaba Group. All subjects who participated in the experiment were provided with and signed an informed consent form. All relevant ethical safeguards have been met with regard to subject protection.

References

1. Sciuto A, Saini A, Forlizzi J, Hong JI (eds) (2018) Hey Alexa, what's up?: a mixed-methods studies of in-home conversational agent usage. In: Proceedings of the 2018 on designing interactive systems conference 2018. ACM
2. Toxtli C, Cranshaw J (eds) (2018) Understanding chatbot-mediated task management. In: Proceedings of the 2018 CHI conference on human factors in computing systems. ACM
3. Shum H-Y, He X-D, Li DJ, Fo IT (2018) From Eliza to XiaoIce: challenges and opportunities with social chatbots. Electron Eng 19(1):10–26
4. Kopp S, Gesellensetter L, Krämer NC, Wachsmuth I (eds) (2005) A conversational agent as museum guide–design and evaluation of a real-world application. Intelligent virtual agents. Springer

5. López G, Quesada L, Guerrero LA (2018) Alexa versus Siri versus Cortana versus google assistant: a comparison of speech-based natural user interfaces. Advances in Human Factors and Systems Interaction. Advances in Intelligent Systems and Computing, pp 241–50
6. McLoughlin I, Sharifzadeh HR (2008) Speech recognition for smart homes. InTech, Speech Recognition
7. Ortiz CL (2014) The road to natural conversational speech interfaces. IEEE Internet Computing 18(2):74–78
8. Luger E, Sellen A (eds) (2016) Like having a really bad PA: the gulf between user expectation and experience of conversational agents. In: Proceedings of the 2016 CHI conference on human factors in computing systems. ACM
9. Kaufmann E, Bernstein A (2010) Evaluating the usability of natural language query languages and interfaces to Semantic Web knowledge bases. J Web Semant 8(4):377–393
10. Ghosh S, Pherwani J (eds) (2015) Designing of a natural voice assistants for mobile through user centered design approach. Springer International Publishing, Cham
11. Lewis JR, Sauro J (eds) (2009) The factor structure of the system usability scale. First International Conference on Human Centered Design, Held as Part of HCI International, San Diego, CA, USA, July 19–24, 2009
12. Shawar BA, Atwell ES (2007) Chatbots: are they really useful? J Lang Technol and Comput Linguist 22(1):29–49