Human Factors and Systems Interaction, Vol. 52, 2022, 396–402
https://doi.org/10.54941/ahfe1002181

Generating a Gesture Set Using the User-Defined Method in Smart Home Contexts

Yanming He, Shumeng Hou, and Peiyao Cheng

School of Humanities and Social Sciences, Harbin Institute of Technology, Shenzhen, China

ABSTRACT

Gesture interaction is a natural interaction method and it has been widely applied in various smart contexts. Smart home system is a promising area to integrate gesture interaction. Under this background, it is necessary to generate a set of gestures that can support users' intuitive interaction with smart home devices. Gesture elicitation study (GES) is an effective method used for generating gestures. In this study, by following GES, we develop a gesture set for controlling a smart TV via a smart speaker, which was common in smart home contexts. Two studies were conducted. In study 1, we conducted a diary study to generate target tasks, resulting in fifteen most frequent tasks in domestic contexts. In study 2, GES was conducted to generate gestures for each command by involving twelve participants. The generated gestures were analyzed by combining frequency, match, ease of use, learnability, memorability and preference, resulting in a set of gestures for smart home contexts.

Keywords: Gesture interaction, Smart home system, Gesture elicitation study

INTRODUCTION

Driven by advanced recognition techniques, gesture interaction becomes increasingly popular in our daily lives. Gesture interaction refers to an interaction method of using hand movements to control devices (Millan and Calhoun, 2000). Compared to traditional graphic user interface (GUI), gesture interaction is more natural and efficient. As people spontaneously use gestures in daily communicative situations (Cassell et al., 1999), the learning and cognitive efforts of using gestures are largely reduced. Mid-air interaction is one of gesture interaction methods, which starts to gain popularity (Bhuiyan and Picking, 2011). Mid-air interaction enables users to control devices untouchably in a short distance. Without the support of any extra equipment, this way is easier and more intuitive (Morris, 2012). An increasing number of smart devices start to integrate mid-air interaction.

Smart home system is one of the potential areas for involving mid-air gesture interaction. This system includes a multitude of distributed devices, such as lights, air conditioners and TVs (Stojkoska and Trivodaliev, 2017). To control multiple devices remotely and instantly, smart speakers often serve as information hubs in smart home systems. For instance, users can turn on lights by talking to a smart speaker. With recent advances in recognition sensors, using mid-air gesture interaction to control smart speakers becomes possible. Thus, it is necessary to develop a gesture set used for interacting with smart speakers in domestic contexts.

Currently, there are two main methods to develop gestures: developer-defined and user-defined method. The developer-defined method refers to the procedure of deriving gestures by designers or engineers based on their specialized knowledge (Buisine and Martinbc, 2007). The user-defined approach involves end-users in the initial stage of gesture design, and thus it reflects users' experience, preferences and expectations directly (Nielsen et al., 2003). It has been demonstrated that users preferred user-defined gestures to developer-defined ones(Morris et al., 2010).

Gesture elicitation study (GES) is a user-defined approach particularly used for gesture elicitation. GES invites target users to generate gestures by themselves. As a result, the generated gestures are often compatible with target users' mental models, leading to a user-friendly gesture set (Wickens et al., 2021). GES contains several steps: 1) gather a set of commands to target tasks; 2) invite target users to perform at least one gesture for each command; 3) calculate the agreement of proposed gestures; 4) select gestures with the highest scores for each command, which drive the final set (Wobbrock et al., 2005). This method has been widely used to generate gestures for interactive systems of robots (Obaid et al., 2012), vehicles (Fariman et al., 2016), and smart TVs (Ali et al., 2018). This method has also been successfully practiced with diverse user populations, such as children (Connell et al., 2013) and individuals with disabilities (Kane et al., 2011). This study aims to generate a user-friendly gesture set for smart speakers by following GES.

STUDY 1: SELECTING TARGET COMMANDS THROUGH A DIARY STUDY

Six participants participated in the study (3 male). The mean age was 31.5 (SD = 14.00). As we focused on the most frequently tasks in domestic contexts in a typical day, a booklet was created to collect the 24-hour activities of a participant and the devices used throughout the day.

Our study generated 211 tasks from 27 products, with an average of 35.2 tasks per person (SD = 4.0). To select target commands, we selected controlling of a TV through a smart speaker as the target scenario as TV is a common device in daily lives. Next, we identified the most frequently used tasks related to TV control, which further used as commands for gesture generation (see Table 1).

STUDY 2: GENERATING A GESTURE SET THROUGH THE GES METHOD

Method: Participants and Procedure

Twelve students were recruited from the university campus (8 male, mean age= 22.83, all right-handed). The study was constructed inside a fully functional sitting room where the participants could feel the usage scenario more

Table 1. Fifteen typical commands used in the current study.

Order	Command	Order	Command
1	Turn on the smart speaker.	9	Play the video.
2	Turn on the TV.	10	Play the next video.
3	Browse a program list.	11	Play the previous video.
4	Choose a video.	12	Give "likes" for the video.
5	Set the video to the full-screen view.	13	Back to Homepage.
6	Increase the volume of the TV.	14	Turn off the TV.
7	Decrease the volume of the TV.	15	Turn off the smart speaker.
8	Pause the video.		

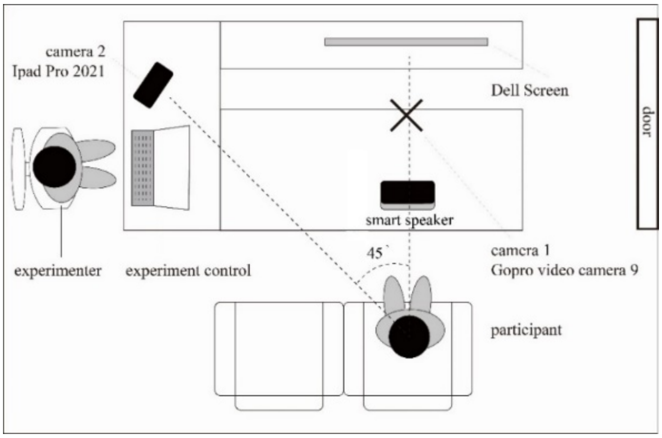

Figure 1: The setup used in gesture elicitation study.

truly. The participants were seated on a sofa in front of a smart speaker. Via a 23.8-inch Dell screen, PowerPoint slides were displayed for demonstrating 15 commands. Commands were shown in the following fixed order as shown in Table 1. The movements of hands were recorded with a video camera and an IPad Pro 2021. The camera was located between the smart speaker and the screen. The Apple tablet was located sideways of the participant at an angle of 45° (see Fig. 1 for the layout of the room). The experimenter sat on the left side of the participant in charge of running the experiment.

Prior to the experiment, the participants were asked to imagine that they were in a smart home where products could be controlled by mid-air gestures through a smart speaker. They were informed that the gestures they proposed would be repeatedly used in the future gestural interaction with smart TVs. Afterward, they carried out a trial to familiarize themselves with the procedure. During the experiment, commands to be conducted were orderly presented on the screen. Each demonstration was added with an extra voice description to ensure that participants had fully understood the commands. Participants were required to make at least two gestures for each command. The repetition of the same gestures for different gestures was allowed. Each gesture's live photo was taken by the IPad Pro 2021. Once participants completed the gesture generation for all the given commands, they were required to fill in a questionnaire, which captured their subjective assessment for each proposed gesture. The entire experiment was recorded by the camera. It lasted approximately 45 minutes.

Subjective Ratings

While selecting the final set gestures, current studies often use the frequency as the main index for the aimed gestures, without considering the participants' subjective assessments. As a result, the obtained gesture sets may not be users' favorite or the most suitable ones (Choi et al., 2012). To compensate for such defects, users' subject ratings have been considered in recent gesture elicitation studies (Kühnel et al., 2011; Vatavu and Zaiti, 2014). Therefore, we achieved the final gestures by combining gesture frequency and subjective assessments in this study.

Based on previous research (Chen et al., 2017; Vogiatzidakis and Koutsabasis, 2018), we collected the subjective ratings of the proposed gestures from five aspects: match, ease of use, learnability, memorability, and preference. Following the generation elicitation study, participants were shown a 7-point Likert scale (1 = completely disagree, 7 = completely agree) to evaluate the five aspects for each gesture-command group: "The gesture I performed is a good match for its purpose," "The gesture I performed is easy to perform," "The gesture I performed is easy to learn," "The gesture I performed is easy to memorize," and "I like the gesture I performed."

Data Analysis

In the elicitation study, we collected 794 gesture proposals. Gestures were considered to be the same ones if they contained the same or similar motions or gesture shapes. Motion mainly comprised the direction. There were only three moving directions: vertical, horizontal and diagonal. Gesture shape referred to which fingers were used and how they were performed.

To evaluate the degree of gestural consensus among the participants, we calculated the gesture agreement rate (AR) for a command by following the formula of Varavu and Wobbrock (2015):

$$\mathrm{AR}\,(r) = \frac{|P|}{|P|-1} \sum_{P_i \subseteq P} \left(\frac{P_i}{P}\right)^2 - \frac{1}{|P|-1}$$

where P is the total number of proposed gestures for command r, and P_i is the number of a subset i of identical gestures from P. The values of AR range from zero to one.

Finally, an overall command-gesture score was calculated by summing the values of the six variables (Pereira et al., 2015): frequency of a specific gesture, match, ease of use, learnability, memorability and preference. Before summation, each of these variables was normalized (with mean = 10.0 and standard deviation = 1).

Table 2. Agreement score and high-scoring gestures for each command.

Command Order	AR	Gesture Order	Overall Score	Command Order	AR	Gesture Order	Overall Score
1	0.020	1.1	7.35	8	0.025	8.1	8.47
2	0.014	2.1	6.88	9	0.018	9.1	6.88
						9.2	6.88
						9.3	6.88
3	0.021	3.1	5.86	10	0.009	10.1	6.29
4	0.020	4.1	6.10	11	0.012	11.1	6.25
5	0.043	5.1	7.94	12	0.032	12.1	7.87
6	0.007	6.1	6.88	13	0.008	13.1	6.88
		6.2	6.88			13.2	6.88
		6.3	6.88			13.3	6.88
7	0.022	7.1	6.88	14	0.018	14.1	6.88
		7.2	6.88				
		7.3	6.88				
		7.4	6.88				
				15	0.018	15.1	7.48

Figure 2: Recommended command-gesture combinations.

Results

Table 2 shows the AR for each command and the proposed gestures with the highest score for each command. For a few commands, the scores of the top highest-score gestures were the same. Therefore, we used them in the final set of gestures. A detailed demonstration of each command-gesture combination is shown in Figure 2.

General Discussion

This study developed a gesture set for interacting with smart speakers in domestic contexts. Following the GES method, we firstly conducted a diary study to identify the fifteen most frequent tasks. Next, we use these tasks as commands and asked participants to generate gestures for them without quantity restriction. After the gesture generation, we also invited users to subjectively evaluate their generated gestures. Finally, we identified top gestures for each command. To select the most appropriate gestures for final gesture vocabulary, we combined gesture popularity and subjective assessments.

Although this study carefully followed the GES procedures and resulted in a set of gestures successfully, the final gesture set might carry limitations. We generated the final set of gestures by combining frequency and subjective evaluation, but people often rated higher scores for gestures they performed (Choi et al., 2012). This bias might neglect the meaningful gestures with a high-frequency ratio. Therefore, it would be beneficial for future research to check the quality of generated gestures (e.g., learnability, memorability, etc) by inviting another group of users.

ACKNOWLEDGMENT

This work was supported by MOE (Ministry of Education in China) Youth Foundation Project of Humanities and Social Sciences (grant number 20YJC760009); Shenzhen Science and Technology Innovation Commission under Shenzhen Fundamental Research Program (grant number JCYJ20190806142401703).

REFERENCES

Ali, A. X., Morris, M. R. & Wobbrock, J. O. (2018). Crowdsourcing similarity judgments for agreement analysis in end-user elicitation studies. Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology. 177–188.

Bhuiyan, M. & Picking, R. (2011). A Gesture Controlled User Interface for Inclusive Design and Evaluative Study of Its Usability. *Journal of Software Engineering and Applications,* 4, 513–521.

Buisine, S. & Martinbc, J.-C. (2007). The effects of speech–gesture cooperation in animated agents' behavior in multimedia presentations. *Interacting with Computers,* 19, 484–497.

Cassell, J., McNeill, D. & McCullough, K.-E. (1999). Speech-gesture mismatches: Evidence for one underlying representation of linguistic and nonlinguistic information. *Pragmatics cognition,* 7, 1–34.

Chen, Z., Ma, X., Peng, Z., Zhou, Y., Yao, M., Ma, Z., Wang, C., Gao, Z. & Shen, M. (2017). User-Defined Gestures for Gestural Interaction: Extending from Hands to Other Body Parts. *International Journal of Human–Computer Interaction,* 34, 238–250.

Choi, E., Kwon, S., Lee, D., Lee, H. & Chung, M. K. (2012). Can user-derived gesture be considered as the best gesture for a command?: Focusing on the commands for smart home system. Proceedings of the Human Factors and Ergonomics Society Annual Meeting. SAGE Publications Sage CA: Los Angeles, CA, 1253–1257.

Connell, S., Kuo, P.-Y., Liu, L. & Piper, A. M. (2013). A Wizard-of-Oz elicitation study examining child-defined gestures with a whole-body interface. Proceedings of the 12th International Conference on Interaction Design and Children. 277–280.

Fariman, H. J., Alyamani, H. J., Kavakli, M. & Hamey, L. (2016). Designing a user-defined gesture vocabulary for an in-vehicle climate control system. Proceedings of the 28th Australian Conference on Computer-Human Interaction. 391–395.

Kane, S. K., Wobbrock, J. O. & Ladner, R. E. (2011). Usable gestures for blind people: understanding preference and performance. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 413–422.

Kühnel, C., Westermann, T., Hemmert, F., Kratz, S., Müller, A. & Möller, S. J. I. J. o. H.-C. S. (2011). I'm home: Defining and evaluating a gesture set for smart-home control. 69, 693–704.

Millan, G. & Calhoun, G. (2000). Gesture-based Control. *International Encyclopedia of Ergonomics Human Factors,* 3, 237.

Morris, M. R. (2012). Web on the wall: insights from a multimodal interaction elicitation study. Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces. 95–104.

Morris, M. R., Wobbrock, J. O. & Wilson, A. D. (2010). Understanding users' preferences for surface gestures. *Proceedings of graphics interface 2010.*

Nielsen, M., Störring, M., Moeslund, T. B. & Granum, E. (April 2003). A procedure for developing intuitive and ergonomic gesture interfaces for HCI. International gesture workshop, Berlin, Heidelberg. Springer, 409–420.

Obaid, M., Häring, M., Kistler, F., Bühling, R. & André, E. (2012). User-defined body gestures for navigational control of a humanoid robot. International Conference on Social Robotics. Springer, 367–377.

Pereira, A., Wachs, J. P., Park, K. & Rempel, D. (2015). A user-developed 3-D hand gesture set for human–computer interaction. *Human factors,* 57, 607–621.

Saffer, D. (2008). Designing gestural interfaces: Touchscreens and interactive devices. *" O'Reilly Media, Inc."*.

Stojkoska, B. L. R. & Trivodaliev, K. V. (2017). A review of Internet of Things for smart home: Challenges and solutions. *Journal of Cleaner Production,* 140, 1454–1464.

Vatavu, R.-D. & Wobbrock, J. O. (2015). Formalizing agreement analysis for elicitation studies: new measures, significance test, and toolkit. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. 1325–1334.

Vatavu, R.-D. & Zaiti, I.-A. (2014). Leap gestures for TV: insights from an elicitation study. Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video. 131–138.

Vogiatzidakis, P. & Koutsabasis, P. (2018). Gesture elicitation studies for mid-air interaction: A review. *Multimodal Technologies Interaction,* 2, 65.

Wickens, C. D., Helton, W. S., Hollands, J. G. & Banbury, S. (2021). *Engineering psychology and human performance*, Routledge.

Wobbrock, J. O., Aung, H. H., Rothrock, B. & Myers, B. A. (2005). Maximizing the guessability of symbolic input. CHI'05 extended abstracts on Human Factors in Computing Systems. 1869–1872.