GUI graphical user interface

Different interaction modalities for smart home

Fabian Hoffmann
University of Regensburg
Regensburg, Bavaria, Germany
fabian.hoffmann@stud.uni-regensburg.de

Miriam Ida Tyroller
University of Regensburg
Regensburg, Bavaria, Germany
miriam-ida.tyroller@stud.uni-regensburg.de

Felix Wende
University of Regensburg
Regensburg, Bavaria, Germany
felix.wende@stud.uni-regensburg.de

ABSTRACT

We present a study aimed to gain insight on users' perceptions and desires in the context of smart home interaction techniques. To achieve this, we conducted an elicitation study in which participants were asked to perform commands within a simulated smart home environment, facing three conditions: voice command, display interaction and mid-air gestures. Facing tasks of different areas in smart home that require user assistance, the participants suggested fitting commands and rated them on the grounds of goodness, ease, enjoyment and social acceptance, as well as their general preference of each technique. The collected measures allow us to present insights that can be used as possible future guidelines for smart home interaction techniques and future research in voice command, display interactions and mid-air gestures.

KEYWORDS

smart home, voice control, display control, mid-air gestures

ACM Reference Format:
Fabian Hoffmann, Miriam Ida Tyroller, and Felix Wende. 2018. Different interaction modalities for smart home. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

While Smart Homes are widely known, they are not widely used. Possible reasons might be high costs, lack of understanding, worries about privacy, the lack of additional value, premature technologies and complicated installation [2]. Solving these problems to further spread the use of smart homes can be done with two different approaches - either remove all the existing obstacles or develop smart home systems so desirable for its users they are not bothered by these obstacles any more, as developed by Hagensby Jensen et al. [7]. Instead of using interaction modalities for smart homes, that are as simple and efficient as possible, the study's intend is to catch potential users attention to own a smart home, by featuring desirable and enjoyable traits.

Our goal is to gain a set of voice commands, display interactions and mid-air gestures, to find out what users want and prefer to do, to solve different smart home tasks.

2 APPROACH AND METHODOLOGY

We followed a similar approach as Dingler et al. [5] by showing and explaining different smart home tasks to the participants and subsequently asking them to propose a voice command, a display interaction and a mid-air gesture, to fulfil the specific tasks in their preferred way. All eleven tasks are listed in section 'Tasks'. A within-subject design was chosen, so every participant gave suggestions for every modality and task. We used a latin-square on the order of the interaction modalities to reduce sequence effects [4] and fatigue. The tasks were shown to the participants in random order. We took video recordings of all sessions. We also collected feedback from participants through questionnaires, on preferences of interaction modalities for a specific task and on goodness, ease, enjoyment and social acceptance of their suggestions.

Interaction Modalities

We compared three different types of interaction modalities. The already commonly used voice and display control, as well as the existing but lacking development technique of mid-air gestures. Therefore, we were able to collect insights on the existing techniques and additionally gain a new set of mid-air gestures.

Tasks

The smart home market can be divided into six different categories [3]. Those are *home entertainment, smart household appliances, energy management, networking and control, comfort and light* and *building security.* We excluded the category *networking and control* for developing the tasks, because it does not include devices that can be controlled, but is rather

Woodstock '18, June 03–05, 2018, Woodstock, NY

ACM ISBN 978-1-4503-9999-9/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

Table 1: Categories with their assigned tasks

Category	Task
Home Entertainment	1. Increase the volume of the music. 2. Turn on the next TV channel.
Smart household appliances	3. Start multi-colored wash at 60 degree. 4. Turn off the oven.
Energy Management	5. Increase the room temperature. 6. Open the shutters.
Comfort and light	7. Turn on the light. 8. Dim the light.
Building security	9. Close the window. 10. Lock the front door. 11. Turn on the security camera.

the infrastructure of a smart home and would be responsible for the detection of performed commands. For all other categories we selected two common tasks [1] each, except for *building security* three because of its bigger market share. All categories with their assigned tasks are listed in table 1.

Participants

A total of 13 participants (7 female) took part in the study with an average age of 33.5 (SD = 15.1). We recruited the participants through social networks and personal contacts. The participants were mostly students from different departments of the University of Regensburg and OTH Regensburg. All of them at least heard of smart homes before and are familiar with interaction through displays. According to the pre-questionnaire, ten participants are familiar with both voice control and display interaction to control other devices, but only one performed mid-air gestures for interaction yet. Seven participants own smart home devices like Google Home, Amazon Alexa, smart TVs or lamps and use them frequently. None of the participants owns a fully integrated smart home system.

Apparatus

The study was carried out in a quiet room. The different tasks were illustrated through pictures, which showed the state before and after issuing the command. Mid-air gestures, voice commands and comments of the participants were recorded by a mounted camera. Display interaction was documented through a sketch on paper. None of the interaction modalities were actually implemented.

Procedure

Before starting the session, the participants were asked to fill out a consent form and a demographic questionnaire. Then they had to fill out a questionnaire in terms of their previous knowledge and usage of smart home devices and the three interaction modalities. After that the tasks were presented to the participants in a random order. At first all tasks had to be fulfilled with a single interaction modality, then with the second and after that with the remaining modality. Additionally to the illustration through pictures, the tasks were explained verbally. The participants were allowed to talk, move and interact with a display in any way they wanted and were encouraged to explain their choices in a thinking-aloud approach. After each task the participants rated their specific suggestion on goodness, ease, enjoyment and social acceptance on four 7-point Likert scales. When all tasks were finished with each interaction modality the participants rated the three interaction modalities for each on 7-point Likert scales, on how good each modality is to perform the specific task. They were asked to do this independently of their own suggestions. At the end a semi-structured interview was conducted to explore the motivation of the participants for each choice and allow them to rate the different interaction modalities under the aspects of efficiency, simplicity, naturality, desirability and enjoyment. this is based on a similar approach in the elicitation study on foot gestures by Felberbaum et al. [6]. The study took about an hour, for which the participants were compensated with sweets.

3 RESULTS

With nine participants and eleven tasks, we collected for each of the three interaction modalities 143 suggestions and in total $13 * 11 * 3 = 429$. Our results include the video recording, taxonomies for each interaction modality, user-defined sets of voice commands, display interactions and gestures, subjective ratings of the sets, qualitative observations and an assessment on the modalities for each task.

Classification of Voice Commands

Taxonomy of voice commands. The participants suggested 43 unique voice commands. The authors manually classified each voice command along five dimensions: *nature*, *form*, *flow*, *context* and *complexity*. Within each dimension are multiple categories, shown in Table 2. We adopted the dimensions from Wobbrock *et al.* [10] and Ruiz *et al.* [8] and adapted them to voice commands.

The *nature* dimension compromises *action* voice commands which state the action to perform. An example of this type of voice command is saying "increase temperature". *State* voice commands describe the desired condition of a device. For example, a *state* voice command is "cameras on" to start camera surveillance.

The *form* dimension describes how much words are used in the voice command and if they have the structure of a full sentence. A *single word* command can be "next" to get to the next TV channel. *Two words* voice commands mostly consist out of the mentioning of the device to be controlled and an action or state. *More words* commands are similar to *two words* but use additional filler words. Finally, voice commands that are correct sentences were classified with the category *sentence*.

The *flow* dimension categorizes the voice commands, if response of a device occurs after or while the user acts. A voice command is*discrete*, when device perform the command after the participant stopped talking. A *continuous* voice command would be starting an action with a command and stop the ongoing action with another command.

The *context* dimension describes, if the voice command requires a specific context or can be performed independently. For example saying "turn off" to turn off the oven is *in-context*, whereas "oven off" is considered *no-context*.

The *complexity* dimension describes if the voice command consists out of a single or a composition of more voice commands. A *compound* voice command can be decomposed into *simple* voice commands.

User-defined voice command set

We collected a total of 143 voice commands, which we used to create a user-defined voice command set for our specified tasks. For each task, we grouped identical voice commands together and the group with largest size was chosen to be the representative voice command for this corresponding task. To evaluate the degree of consensus among the participants, we computed the *agreement score* A_t (Equation 1), as proposed by Vatavu and Wobbrock [9], for each task.

$$A_t = \frac{|P_t|}{|P_t|-1} \sum_{P_i \subseteq P_t} \left(\frac{|P_i|}{|P_t|} \right)^2 - \frac{1}{|P_t|-1} \quad (1)$$

In equation 1, t is a task in the set of all tasks T, P_t is the set of suggested voice commands for t and P_i is a subset of identical voice commands from P_t. As an example of calculation of an agreement score, consider the task *increase the volume of the music*. The task has four groups of identical voice commands with a size of 7, 4, 1 and 1. Therefore the agreement score for *increase the volume of the music* is:

$$A = \frac{13}{12}\left(\left(\frac{7}{13}\right)^2 + \left(\frac{4}{13}\right)^2 + \left(\frac{1}{13}\right)^2 + \left(\frac{1}{13}\right)^2\right) - \frac{1}{12} = 0.346 \quad (2)$$

Table 2: Taxonomy of voice commands for smart home tasks

Taxonomy of voice commands		
Nature	Action	Voice command states the action to perform
	State	Voice command describes the desired condition
Form	Single word	Voice command consists out of a single word
	Two words	Voice command consists out of two words
	More words	Voice command consists out of more words without sentence structure
	Sentence	Voice command uses sentence structure
Flow	Discrete	Response occurs *after* the user acts
	Continuous	Response occurs *while* the user acts
Context	In-context	Voice command requires specific context
	No-context	Voice command does not require specific context
Complexity	Simple	Voice command consists of a single voice command
	Compound	Voice command can be decomposed into simple voice commands

Figure **??** illustrates the agreement scores for each task using voice commands. Participants had the least agreement on commands for the tasks *start multi-colored wash at 60 degree* (Task 3) and *increase the room temperature* (Task 5). This is attributable to the complexity of the tasks.

Classification of Display Interaction

The participants suggested 61 unique display interactions. Similar to the voice commands we manually classified each display interaction along three dimensions for graphical user interface (GUI) elements (*form*, *elements*, *flow*) and for touch gestures along four dimensions (*form*, *nature*, *binding*, *flow*). Within each dimension are multiple categories, shown in Table 3 for GUI elements and for touch gestures in Table 4. The dimensions and categories for GUI elements were inspired by the work from Wobbrock *et al.* [10] and Ruiz *et al.* [8]. As taxonomy for touch gestures we used Wobbrock *et al.* [10] taxonomy of surface gestures, witch is displayed in Table 4.

Taxonomy of display interaction (GUI elements): ToDo: Beschreibung Taxonomy

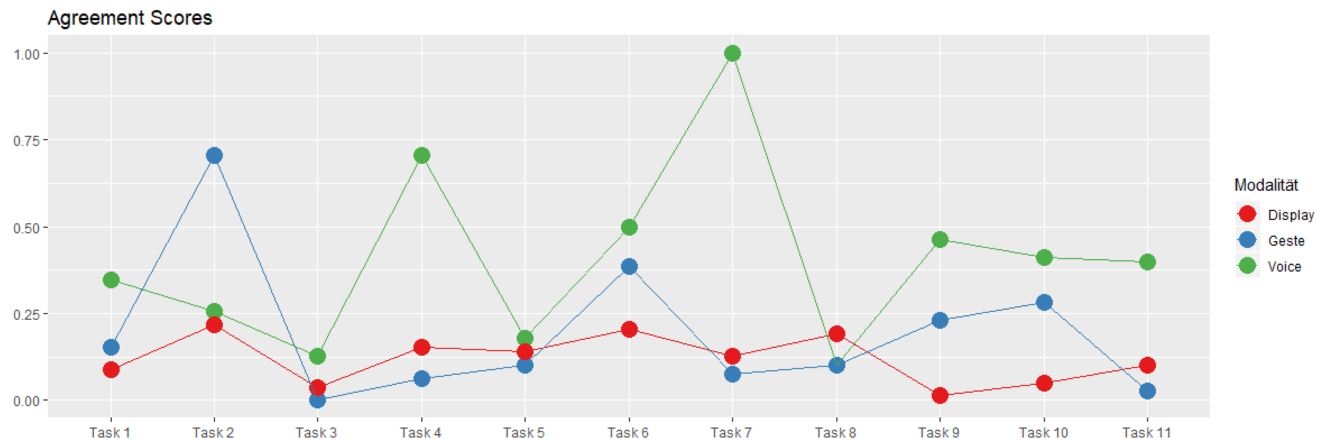

Figure 1: Agreement scores for each task with voice commands

Table 3: Taxonomy of display interactions for smart home tasks (GUI elements)

Taxonomy of display interactions (GUI elements)		
Form	Direct Action	Single interaction that directly leads to the action
	Selection & Confirmation	Selection of the action and starting through another element
Elements	Single clickables	The GUI includes one or more single clickables (button, checkbox, etc.)
	Slider	The GUI includes one or more sliders
	More words	Voice command consists out of more words without sentence structure
	Rotation	The GUI includes one or more rotational elements
	Text & number entry	The GUI includes one or more options to enter text or numbers
	Symbolic	The GUI includes one or more special symbolic elements
Flow	Discrete	Response occurs *after* the user acts
	Continuous	Response occurs *while* the user acts

Table 4: Taxonomy of display interactions for smart home tasks (touch gestures)

Taxonomy of display interactions (touch gestures)		
Nature	Symbolic	Gesture visually depicts a symbol
	Physical	Gesture acts physically on objects
	Metaphorical	Gesture indicates a metaphor
	Abstract	Gesture-referent mapping is arbitrary
Form	Static pose	Hand pose is held in one location
	Dynamic pose	Hand pose changes in one location
	Static pose and path	Hand pose is held as hand moves
	Dynamic pose and path	Hand pose changes as hand moves
	One-point touch	Static pose with one finger
	One-point path	Static pose and path with one Finger
Binding	Object-centric	Location defined with respect to object features
	World-dependent	Location defined with respect to world features
	World-independent	Location can ignore world features
	Mixed dependencies	World-independent plus another
Flow	Discrete	Response occurs *after* the user acts
	Continuous	Response occurs *while* the user acts

Taxonomy of display interaction (touch gestures): ToDo: Bschreibung Taxonomy

User-defined display interaction set

ToDo: Beschreibung + Tabelle set

Classification of Mid-Air Gestures

ToDo: Beschreibung Taxonomy ToDo: Tabelle Taxonomy

User-defined mid-air gestures set

ToDo: Beschreibung + Tabelle set

REFERENCES

[1] [n.d.]. Home - SmartHome Hilfe. https://service.startsmarthome.de/de/
[2] [n.d.]. Infografik: Smart Home: Was spricht dafür, was dagegen? https://de.statista.com/infografik/15254/argumente-fuer-und-gegen-die-nutzung-von-smart-home-produkten/
[3] [n.d.]. Smart Home - weltweit | Statista Marktprognose. https://de.statista.com/outlook/279/100/smart-home/weltweit#market-revenue
[4] 2017. Latin Square Design: Definition and Balanced Latin Square Algorithm. https://www.statisticshowto.datasciencecentral.com/latin-square-design/

[5] Tilman Dingler, Rufat Rzayev, Alireza Sahami Shirazi, and Niels Henze. 2018. Designing Consistent Gestures Across Device Types. In *Engage with CHI*, Regan Mandryk and Mark Hancock (Eds.). The Association for Computing Machinery, New York, New York, 1–12. https://doi.org/10.1145/3173574.3173993

[6] Yasmin Felberbaum and Joel Lanir. 2018. Better Understanding of Foot Gestures. In *Engage with CHI*, Regan Mandryk and Mark Hancock (Eds.). The Association for Computing Machinery, New York, New York, 1–12. https://doi.org/10.1145/3173574.3173908

[7] Rikke Hagensby Jensen, Yolande Strengers, Jesper Kjeldskov, Larissa Nicholls, and Mikael B. Skov. 2018. Designing the Desirable Smart Home. In *Engage with CHI*, Regan Mandryk and Mark Hancock (Eds.). The Association for Computing Machinery, New York, New York, 1–14. https://doi.org/10.1145/3173574.3173578

[8] Jaime Ruiz, Yang Li, and Edward Lank. 2011. User-defined motion gestures for mobile interaction. In *Conference proceedings and extended abstracts / the 29th Annual CHI Conference on Human Factors in Computing Systems*, Desney Tan, Geraldine Fitzpatrick, Carl Gutwin, Bo Begole, and Wendy A. Kellogg (Eds.). ACM, New York, NY, 197. https://doi.org/10.1145/1978942.1978971

[9] Radu-Daniel Vatavu and Jacob O. Wobbrock. 2015. Formalizing Agreement Analysis for Elicitation Studies. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, Bo Begole, Jinwoo Kim, Kori Inkpen, and Woontack Woo (Eds.). ACM Press, New York, New York, USA, 1325–1334. https://doi.org/10.1145/2702123.2702223

[10] Jacob O. Wobbrock, Meredith Ringel Morris, and Andrew D. Wilson. 2009. User-defined gestures for surface computing. In *CHI 2009 - digital life, new world*, Dan R. Olsen, Richard B. Arthur, Ken Hinckley, Meredith Ringel Morris, Scott Hudson, and Saul Greenberg (Eds.). ACM, New York, NY, 1083. https://doi.org/10.1145/1518701.1518866

A RESEARCH METHODS

Part One

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Morbi malesuada, quam in pulvinar varius, metus nunc fermentum urna, id sollicitudin purus odio sit amet enim. Aliquam ullamcorper eu ipsum vel mollis. Curabitur quis dictum nisl. Phasellus vel semper risus, et lacinia dolor. Integer ultricies commodo sem nec semper.

Part Two

Etiam commodo feugiat nisl pulvinar pellentesque. Etiam auctor sodales ligula, non varius nibh pulvinar semper. Suspendisse nec lectus non ipsum convallis congue hendrerit vitae sapien. Donec at laoreet eros. Vivamus non purus placerat, scelerisque diam eu, cursus ante. Etiam aliquam tortor auctor efficitur mattis.

B ONLINE RESOURCES

Nam id fermentum dui. Suspendisse sagittis tortor a nulla mollis, in pulvinar ex pretium. Sed interdum orci quis metus euismod, et sagittis enim maximus. Vestibulum gravida massa ut felis suscipit congue. Quisque mattis elit a risus ultrices commodo venenatis eget dui. Etiam sagittis eleifend elementum.

Nam interdum magna at lectus dignissim, ac dignissim lorem rhoncus. Maecenas eu arcu ac neque placerat aliquam. Nunc pulvinar massa et mattis lacinia.