Summary of the Master's Degree Thesis in Computer Engineering

Conversational Agents for Creating Personalization Rules in the IoT

Candidate: Carlo Borsarelli
Supervisors: Luigi De Russis, Alberto Monge Roffarello

Introduction

The number of smart speakers on the market and their diffusion in homes is growing more and more. The cost of these devices, such as Google Home or Amazon Echo, has now become accessible to many people, and the race for the best smart speaker is still underway. The Intelligent Personal Assistant (IPA) integrated into these speakers can perform some tasks as reading texts or emails, scheduling appointments, setting reminders, but they can also become the communication channel between the user and the IoT ecosystem. Turning on the lights, lower the shutters, or set a temperature of a room are all examples of actions that a smart home already allows, thanks to a simple conversation with an IPA. Software tools like IFTTT, Node-Red, IPAs' mobile apps enable users to set even complex behaviors by defining personalization rules in the trigger-action format: if something happens, then an operation is executed. But none of those systems exploit the potential of the voice interaction that can potentially integrate ease of use and flexibility in a smart home context. This thesis attempts to provide solutions that extend the possibilities of personalizing IoT ecosystems via voice by developing systems for the vocal composition of personalization rules.

Project Goals

The thesis explores novel approaches for creating personalization rules through conversations. More specifically, the primary objectives are to identify:

- If a non-technical user would be willing to create personalization rules just by talking with a smart speaker.
- What are natural formalisms to compose rules by voice, and if the trigger-action paradigm is good or not.
- Which characteristics should have an IPA to be suitable for the rule creation process.

To reach these goals, first, I reviewed the works in the research area that analyzes the state-of-the-art of End-User Development in the IoT, by finding the problems and successes of the proposed solutions. Next, in order to explore how users would create personalization rules by talking with smart speakers, I conducted a user study in the form of semi-structured interviews with 7 non-technical users. After that, considering the knowledge gained with the interviews and the analyzed research studies, I designed and implemented two different rule composition approaches, one completely vocal, and the other one that mixes voice interaction with user's physical actions on smart devices, e.g., to tell the IPA which is the right device to control in case of ambiguities. Finally, both prototypes were evaluated and compared with usability tests performed with 10 non-technical users.

Interviews

The study sessions were conducted in the north of Italy in April 2021 with 7 users. The recruitments took place through social channels, by focusing the choice on potential participants from my social circles. In order to have a broader spectrum of technical skills and background studies, I carried out the interviews by targeting a heterogeneous group of participants ranging from 18 to 52 years and with a mix of course of studies and job occupations. The summarized sociodemographic characteristics of the 7 participants are presented in Table 1.

Participant	Age Range	Job	Smart speaker's owner
P1	23-27	Speech therapist	✓
P2	18-22	Student	✓
P3	48-52	Housewife	✓
P4	23-27	Health technician	✗
P5	23-27	Student	✗
P6	43-47	Health workers	✗
P7	23-27	Math teacher	✓

Table 1: Study participants

The decision was limited to participants with a moderate interest in home automation, furthermore, I selected a quite equal number of users who own and do not own smart speakers. All these participants' characteristics were individuated in the recruiting phase. After a background interview, which identified the participants' relationships with IPAs, I conducted an imagination exercise to ask participants how they would have created rules to personalize a hypothetical smart home. Interviewees received 2 scenarios in which they had to formulate a rule by voice in order to solve a goal like *"you would like to turn on the central kitchen light every time you enter that room"*.

The results showed that the vocal composition of the rule in trigger action formalism was natural for most non-technical users. Another aspect highlighted by interviews was how the users would prefer a system with a minimum conversation with the assistant, they were scared of a fully automated IPA, and they would have preferred a system with an error recovering procedure and with the possibility of explicitly confirm the rule. The participants said that setting a rule is a significant action in a smart home and, in case of an error, it could negatively affect life in the home, so it is important to understand if the IPA correctly understood their rule composition.

Design and Implementation of the *Create a Rule* prototypes

This work considers 2 novel approaches for the creation of personalization rules, in particular, I have designed and implemented 2 prototypes of conversational agents, the first (Pr1) is composed of a fully vocal interface that takes the advantage of the natural language by developing conversations for most of the smart home inhabitants. Instead, the second prototype (Pr2) is not an entirely vocal agent, it requires an active part of the user who has to interact with the smart house to make the assistant aware of his intents.

To make the interaction between users and IPAs more natural, I designed the two conversational agents starting with high-level designs that take into account the information acquired in the previous research phase and interviews. Also, I considered the principal conversational components identified by the literature of Voice User Interface Design. The prototypes should guide users in the creation of the rules and map what they can say in any situation, to do that I included these components:

a) **Welcome**: a new user is introduced to the voice application and informed on how to create a rule.
b) **Help**: the prototypes guide users in cases they do not know or do not remember what to say. They provide some examples of rules formulations.
c) **Rule Creation**: the two *Create a Rule* prototypes include some differences in rule creation: Pr1 requires the pronunciation of both rule components, the trigger, and the action, in a single sentence. On the other hand, Pr2 splits the rule creation into two parts, the initial composition of the event (completely vocal) and the subsequent definition of the action (with physical interaction on a smart device).
d) **Missing Rule Details** (Pr1 only): it occurs when users do not specify some details necessary to complete the rule and these details are not derivable from the context. For example, if a user tries to set a rule like *"if it's 9 am, turn on the light"*, and the house owns more than 1 light, the IPA will ask for light specification providing the user with some possible choices.
e) **Confirmations**: after the formulation of a rule, the *Create a Rule* prototypes will vocally reformulate the rule so that the user can understand if the rule that she defined has been correctly interpreted. The prototypes will ask for an explicit confirmation to set the rule in the smart home ecosystem.
f) **Corrections**: user has the possibility to correct the rule before setting it.

Both prototypes integrate the above conversational components with some differences in rule creation. Furthermore, Pr1 has a level of automation that allows the IPA to require from users only the information really needed, an information deductible from the context like the room of the smart coffee machine, does not need to be explicitly requested. In addition, the Pr1 correction component allows the user to correct both parts of the rule, giving the possibility of solving some IPA misinterpretation problems. Instead, Pr2 has the same principal components of the conversation as Pr1, but the rule creation is split into two parts, the initial composition of the event (completely vocal) and the subsequent definition of the action. In particular, Pr2 takes advantage of smart speakers' knowledge of all the IoT devices in the smart home in which they are integrated. To define the action, it is required that the user physically interacts on the IoT device that she wants to add to the rule. The user could, for example, turn on lights, close the blinds, or set a temperature; the prototype will unequivocally identify the device and the selected state. For example, if the user wanted to create a rule like *"each time I enter the kitchen, the kitchen light will turn on"*, the user should first define and confirm the event, and then she should interact with the kitchen light by turning it on.

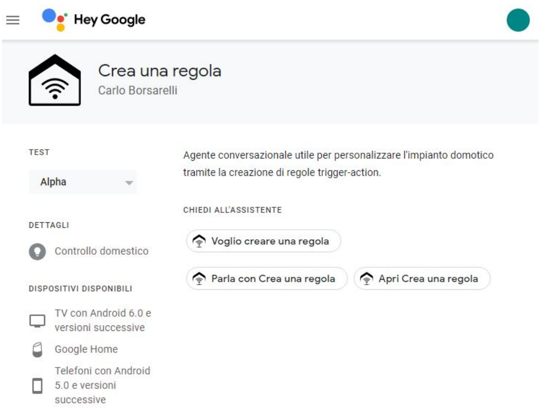

Figure 1: Alpha Release of Pr1 Google Action

The implementations of the two *Create a Rule* prototypes exploit the natural language processing capabilities of Dialogflow, which captures the user's intents and sent the request to the service in the backend that generates the answer. Both prototypes are based on conversational agents implemented with the integration of Action on Google. The prototypes have been developed with 2 Google Actions releases to be tested on a device having a Google Assistant integrated as, for example, a Google Home or a smartphone. Figure 1 shows the Google Assistant Action of Pr1. Since considering real IoT devices of different brands and communication protocols, all integrated into a single system would have required a considerable development effort and it was not the focus of the thesis, I chose to simulate the device controllers by displaying them on a graphical interface of a website created ad-hoc. To test Pr2, the device controllers were displayed on touch-based devices such as smartphones or tablets.

Participant	Age Range	Job sector	Familiarity with voice assistants
P1	48-52	Health and Social care	✗
P2	53-57	Health and Social care	✓
P3	23-27	Health and Social care	✓
P4	23-27	Health and Social care	✓
P5	53-57	Housewife	✓
P6	53-57	Engineering and manufacturing	✓
P7	23-27	Health and Social care	✗
P8	23-27	Teacher training and education	✗
P9	23-27	Teacher training and education	✗
P10	23-27	Health and Social care	✗

Table 2: Demographic information of test participants

Evaluation

The purpose of the test is to analyze whether the creation of trigger-action rules through conversation can make the personalization of smart homes easier and/or faster. The study evaluated the two *Create a Rule* prototypes. 10 users were recruited for the usability test, the selection was made from participants with similar characteristics to the ones chosen for the interviews, so a heterogeneous mix of job backgrounds and technological skills, and a moderate interest in home automation. Table 2 summarizes the participants' demographical characteristics.

The prototypes were tested with a within-subject design, which allowed participants to test both conversational agents. I provided users with an appropriate context of a situation where they, probable owners of a new smart home (Figure 2), had to personalize the IoT ecosystem present in it.
The usability test included six tasks, each one containing a real situation to solve through the creation of a rule. First, users performed the operations with one conversation agent, then they repeated the tasks executions using the other prototype. An example of one of the six tasks was: *"you are in the kitchen and the light switch is in an awkward position. For this reason, you want the light to turn on every time someone enters the kitchen. Now try to tell the smart speaker to set this behavior."* During the tests, I collected different evaluations metrics: both subjective questions and quantitive measurements like the errors metrics, the times on tasks, and the tasks completions. As hypothesized, the average time spent by participants on task completion was greater for Pr2 than for Pr1, the separation of the rule creation parts of Pr2 have required more time in setting up the rule. A certain correlation was noted between time on task and that self-evaluated themselves as experts in technology, those who self-evaluated themselves as tech-savvies completed the tasks an average with lower times than the other participants. To have an idea of the "optimal" times on tasks, the developer performed each task to simulate a no-errors condition; it was found that the tech-savvies participants obtained execution times very close to those of the developer. This shows that when users have found a good formalism for the rule creation they can quickly achieve excellent results. Certain tasks performed with Pr1 have been problematic for many users: the amount of information to be pronounced during the rule creation confused many participants and some of them gave up, while others used words that the agent failed to understand. Even with the corrections and advice provided by the IPA, someone failed to adapt his expressions to the given examples. The same problem did not happen with Pr2, which allowed participants to complete the tasks in practically all cases. The main errors that prevented the completion of the tasks can be grouped into the metric "False Reject Errors", this metric occurs when a user said something that should be handled by the assistant, but it does not. However, many False Reject Errors can be solved by a good training of the agents that can reduce most of the errors. Not tech-savvy users found the creation with Pr2 easier and more guided. Instead, Pr1 was more attractive to tech-savvy, the speed at which they created the rules (even in less than 30s) was a great advantage over the "slowness" of Pr2. Despite the strong differences in the task completions, the System Usability Scale (SUS) scores of the two prototypes determine "good" usability for both agents. In particular, Pr1 obtained a score of 71.5 and Pr2 of 73.3, this shows how the difficulties of some tasks did not affect the perceived usability, despite failures in the completions. Many users found both applications pleasant to use, the two rules creation approaches were found complementary. Overall, both prototypes represent valid solutions and they guarantee the creation of rules in a comfortable and easy-to-use way for most users.

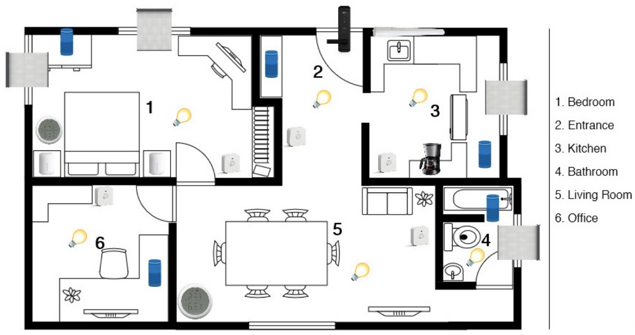

Figure 2: The house plan provided to users in the test

Conclusions and Future Works

Nowadays, smart speakers are considered simple entertainment items. This thesis attempts to extend the current state of the smart speakers' capability, by creating new ways of interaction that could exploit better their potentiality. The prototypes developed have shown how the creation of personalization rules can be natural and effective via voice. Future work could expand the functionality of the developed voice applications, for example, by allowing users to configure even more complex rules. Further reasoning in the study context could consider other novel approaches for smart homes personalization through conversation. The positive feedback and the work done are very promising indicators of the importance of future studies in the field.