Investigating Conversational Programming for End-Users in Smart Environments through Wizard of Oz Interactions

Kate Howland
Department of Informatics
University of Sussex Brighton,
BN1 9QJ, UK

James Jackson
Department of Informatics
University of Sussex Brighton,
BN1 9QJ, UK

Abstract

Natural language programming has long been an aspiration, but is fraught with challenges that have so far prevented genuinely useful and useable applications. End-user programming for smart environments is increasingly being pursued through trigger-action rule services that include simplified natural language description of rules. Along with the increasing prevalence of Voice User Interfaces (VUIs) in smart environments, this points to new opportunities for supporting understanding, debugging and editing of rules through speech. However, there is a lack of contextually relevant data on how end-users without programming experience describe and understand rules for smart environment behaviours through speech. This paper describes how the CONVER-SE project is developing methodology and software for capturing this data, and prototyping VUIs that attempt to mitigate the many challenges with supporting programming interactions through speech in this context.

1. Introduction

Programming using natural language long been a goal in end-user and novice programming research, but has so far fallen short of expectations due to fundamental challenges in reaching alignment in communication between human and system. VUIs such as Amazon Echo/ Alexa and Google Home/ Assistant have made speech a frontrunner for smart home control, but do not yet support editing, debugging and authoring of smart home automation rules through speech. Understanding, configuring and customising the rules that define smart environment behaviours are end-user programming (EUP) activities. Currently, these activities must be done using a separate, screen-based interface, as voice interaction is largely limited to triggering pre-defined behaviours. Automation platforms such as IFTTT and Stringify allow programming of smart home behaviours through trigger-action rules, but have seen limited uptake beyond early adopters and tech-savvy hobbyists. These platforms provide natural language descriptions of the trigger-action rules, but the rules are detached from the context in which they will be carried out. There is a gulf between abstract representations of automated behaviours and the concrete real-world environments in which they play out. For example, a user standing next to a smart lamp wanting to understand or reconfigure the rules for its behaviour must turn their attention from the room to a screen, understand and edit a code-like description, and draw a link between a unique identifier and the object in the room. Supporting these activities through a voice interface, with potential to include gesture and proximity data to support disambiguation, could provide more intuitive ways of understanding and programming smart environments.

With VUIs now widely used in intelligent assistants, there is renewed interest in programming through speech, but we lack foundational research on how users without a programming background can best understand and express rules defining smart environment behaviour.

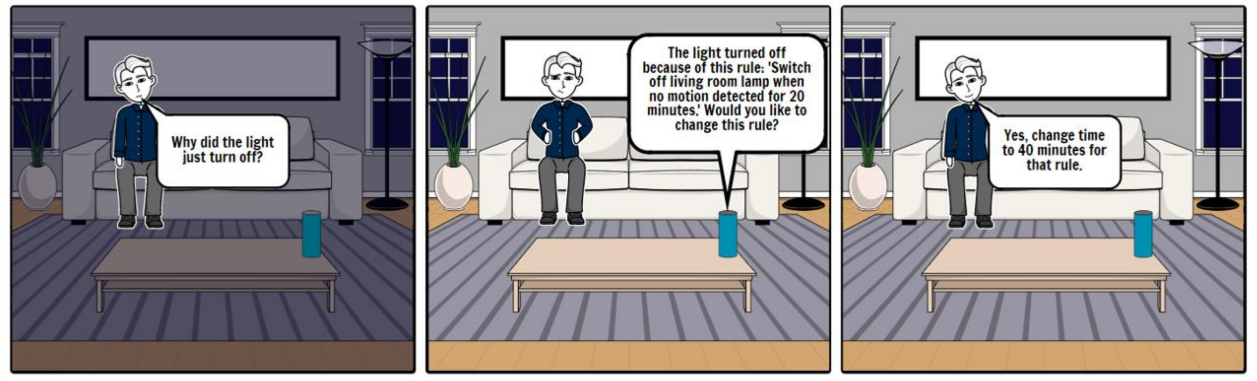

Why did the light just turn off?
The light turned off because of this rule: 'Switch off living room lamp when no motion detected for 20 minutes.' Would you like to change this rule?
Yes, change time to 40 minutes for that rule.

Figure 1: Design Fiction of Conversational Debugging

Gathering data on how end-users users ‘naturally’ express programmatic rules is a well-established approach in EUP research. However, studies of natural expression of programmatic rules for smart environments are typically carried out using toy scenarios in decontextualised settings, and often limited to written responses to survey questions or similar text-based descriptions. This means that there is very little data on natural expression of rules through speech, and no data on how co-speech gesture and contextual elements such as proximity support speech when describing rules. In smart home scenarios, the presence of cameras in sensor-enabled environments makes it feasible for additional contextual information to be used to resolve ambiguities and deictic references (e.g. this, there, that). In addition, it is important to recognize the extent to which ‘natural’ expression is increasingly influenced by expectations from interaction with existing similar systems. In the context of conversational interfaces, it may be more realistic to focus on language alignment between the system and the user.

In the CONVER-SE project, we are examining the challenges of speech programming for smart environments, and investigating how these could be mitigated in a conversational interface. To carry out this research, we are developing methodology by adapting natural expression studies to include capture of speech, gesture and proximity in situ. We are making use of Wizard of Oz prototyping (in which some or all functionality is implemented by a human) and participatory methods such as bodystorming (in which participants play out interactions with an imagined future system). This paper describes the design of a domestic study that is currently in progress, and the development of a software toolkit for Wizard of Oz prototyping of conversational interfaces.

2. Background

Previous research on EUP for smart environments has gathered natural language descriptions of rules using empirical methods including online surveys [1, 2], post-it note instruction tasks [3] and interviews [4]. Existing work has led to some consensus, including trigger-action rules as a simple but powerful format [2, 5], an inclination for users to rely on implicit rather than explicit specification [1, 2] and a tendency for them not to mention specific sensors or devices [1, 2, 4].

Although these studies have provided important insights into the natural expression of tasks and rules for smart environments, context has been largely overlooked in this work, and none of the studies were conducted in real-world scenarios. In addition, natural language descriptions have been collected in isolation from other communicative modes, such as gesture. Given the importance of context for smart environments, it is likely that existing findings only provide a limited picture. For example, the finding that end-users do not make reference to specific sensors or equipment, first reported by Truong et al. [1] and validated by the findings of Dey et al. [4] and Ur et al. [2], may well have been influenced by the lack of real-world context in the studies. Referring to sensors that you know exist in your house would be much more likely than referencing hypothetical sensors in a toy scenario. The importance of real world contexts for smart environment EUP research is beginning to be recognized. For example, a recently published EUP study comparing different notation styles for home automation was carried out in real domestic environments [6], but unfortunately the study design did not allow for examination of contextual referencing, or capture of speech, gesture or proximity data. Our own lab-based pilot work [7] also suggested a number of limitations with decontextualized studies, as described further in the following section.

3. Domestic Studies

To gather contextually valid data, we are conducting our studies in participants’ homes. We are recruiting study participants who already have some level of smart home technology, but who do not consider themselves technology experts, and have no previous programming experience. We are particularly keen to reach ‘inadvertent adopters’, who we define as the family and housemates of early adopters.

Our first study, which is currently underway, is designed to investigate the following questions:

- How do end-users naturally understand and specify rules for smart environment behaviours in their own homes?
- What ambiguities and inaccuracies are present in understandings of and expressions of such rules?

- To what extent does participants' language align to that used by a VUI over the course of an interaction
- How far can conversational approaches help with understanding, editing and generating complete and unambiguous rules?

We are investigating these questions through a three-part study, which is video recorded. Part 1 is a semi-structured interview that investigates participants' current use and understanding of smart home technology and VUIs, and captures natural descriptions of rules for automated behaviours they would like to have running in their homes. Part 2 involves Wizard of Oz prototype interactions, allowing us to test conversational approaches to supporting editing and generating of rules (including modelling of rule structure and stepwise composition of rule parts). This also allows us to examining if/how conversational alignment occurs over the course of interaction with the VUI. Part 3 involves participants bodystorming future interactions with a more advanced VUI that could detect contextual information, allowing us to seek active input from participants on ideas for effective support.

3. Wizard of Oz Toolkit

For our studies we needed an application that allows us to simulate the behaviour of an advanced VUI designed to facilitate EUP of smart devices, in order to test the effectiveness of different conversational approaches. For the first study we use a Bluetooth speaker and light, controller remotely through a laptop, as shown in Figure 2. As we are also interested in assessing the benefits of providing visual feedback as part of the EUP process in future studies, we included a requirement that the application be capable of projecting text and images to a mobile display device. The final requirement was that the core application functionality be available offline, as the studies are to be carried out in situ, where the availability of an internet connection cannot be assumed.

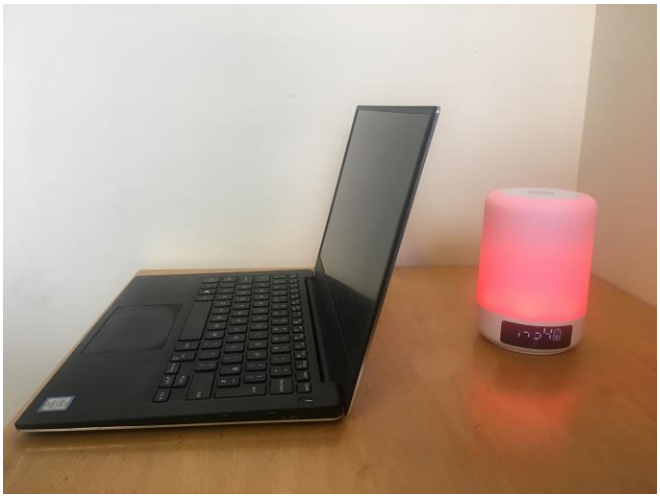
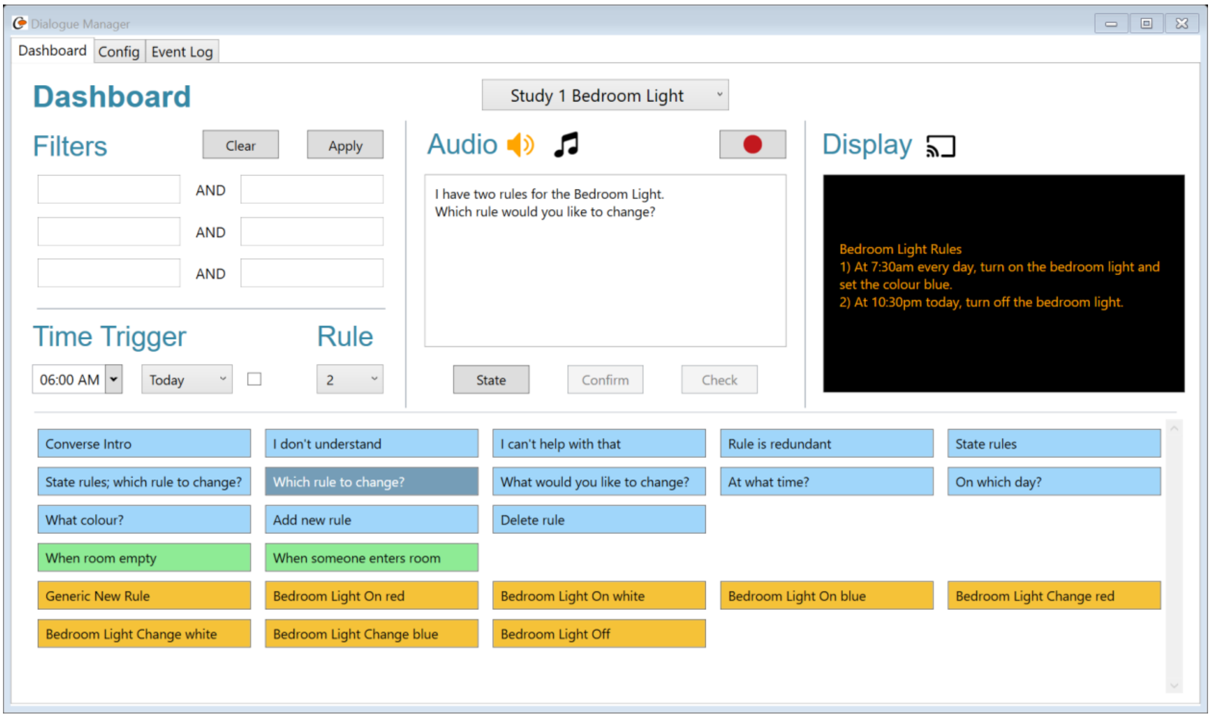


Figure 2: Wizard of Oz Hardware

For speech production, we reviewed several offline text to speech (TTS) engines that could be integrated with the application but were dissatisfied with the quality of the generated speech. We therefore decided, at least for the initial version of the application, to use an online TTS service to create a database of audio clips that could be linked together in different ways to generate the utterances we anticipated might be required: requesting information, making statements, checking understanding and confirming when actions are completed. The utterances were structured so as to minimise the required number of audio clips, while providing a flexibility of response and not unduly compromising the production of natural sounding speech. The majority of the audio clips are required for time and event triggers e.g. "At 7:30pm every day" or "when someone enters the room"; a smaller number cover general statements and questions e.g. "I'm sorry, I don't understand." or "What would you like to change?" and finally there is a group of clips specific to each device e.g. "turn on the kitchen radio and select Radio 3" or "switch off the bedroom light".

The application was designed to automate functions where possible so the researcher's attention could be focused on selecting the most appropriate VUI response. Therefore depending on the selection function (e.g. "check" or "confirm"), the application will automatically generate the required composite audio clip from the database of stored clips and display the audio text for review before the researcher plays the clip. Changes to the ruleset are automatically tracked and there is selectable option to check for rule conflicts and duplicate rules. If required the ruleset can be projected to a remote display. After testing an early prototype of the application, a filters panel was added to the dashboard to enable users to limit the displayed triggers and actions (Figure 3).

To help with data collection and analysis, the application can make an audio recording of the study session and an event log records and timestamps all application activity.



Study 1 Bedroom Light

06:00 AM Today □ 2

I have two rules for the Bedroom Light.
Which rule would you like to change?

Bedroom Light Rules
1) At 7:30am every day, turn on the bedroom light and set the colour blue.
2) At 10:30pm today, turn off the bedroom light.

Figure 3: Dialogue Manager showing Dashboard panel

4. Conclusions and Further Work

The first study is ongoing, with video data from the first participants currently being transcribed. From our observations so far, we note that participants find the stepwise composition of rules with conversational prompts from the interface relatively easy, whilst the statement of complete rules without the assistance of prompts is highly demanding, in line with expectations from analysis of speech-based interaction according to the Cognitive Dimensions of Notations [3. p.5]. Once we have finished data collection, we will begin analysis to answer the research questions stated in section 3, and use the findings to drive the design of conversational support for our next prototype.

5. References

1. Truong, K.N., E.M. Huang, and G.D. Abowd, CAMP: A magnetic poetry interface for end-user programming of capture applications for the home, in Proc. of Ubiquitous Computing. 2004, Springer. p. 143-160.

2. Ur, B., et al., Practical trigger-action programming in the smart home, in Proc. of Human Factors in Computing Systems. 2014, ACM. p. 803-812.

3. Perera, C., S. Aghaee, and A. Blackwell, Natural Notation for the Domestic Internet of Things. End-User Development, 2015. 9083: p. 25-41.

4. Dey, A.K., et al., iCAP: Interactive prototyping of context-aware applications, in Proc. of Pervasive Computing. 2006, Springer. p. 254-271.

5. Catala, A., et al., A meta-model for dataflow-based rules in smart environments: Evaluating user comprehension and performance. Science of Computer Programming, 2013. 78(10): p. 1930-1950.

6. Brich, J., et al., Exploring End User Programming Needs in Home Automation. ACM Transactions on Computer-Human Interaction (TOCHI), 2017. 24(2): p. 11.

7. Howland, K. and Jackson, J. Designing for End-User Programming through Voice: Developing Study Methodology, Workshop on Voice-based Conversational UX Studies and Design, CHI 2018, Montréal, Canada