POLITECNICO DI TORINO
Repository ISTITUZIONALE

Personalizing IoT Ecosystems via Voice

Original
Personalizing IoT Ecosystems via Voice / De Russis, Luigi; Monge Roffarello, Alberto. - ELETTRONICO. - 2702:(2020), pp. 37-40. (Intervento presentato al convegno EMPATHY 2020: Empowering People in Dealing with Internet of Things Ecosystems tenutosi a Island of Ischia (Italy) nel September 29, 2020).

Availability:
This version is available at: 11583/2842742 since: 2020-10-20T07:34:21Z

Publisher:
CEUR-WS

Published
DOI:

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

29 March 2024

Personalizing IoT Ecosystems via Voice

Luigi De Russis[a], Alberto Monge Roffarello[a]

[a]Politecnico di Torino, Dipartimento di Automatica e Informatica, Corso Duca degli Abruzzi 24, 10129 Torino, Italy

Abstract

Intelligent Personal Assistants (IPAs), embedded in smart speakers, allow users to set up some trigger-action rules, via their mobile apps, to personalize the IoT ecosystem in which they are located. Vocal capabilities might be involved in such rules as triggers or actions, but the actual rule composition and execution flow is totally segregated into the mobile app. This position paper reflects on the challenges and opportunities brought by IPAs if they played a more prominent and integrated role in this personalization scenario.

Keywords

Internet of Things, Speech, Intelligent Personal Assistant, Voice Interaction, End-User Development

1. Introduction and Background

Smart speakers, such as Google Home or Amazon Echo, are entering our homes and enriching the Internet of Things (IoT) ecosystem already present in them. The Intelligent Personal Assistants (IPAs) they include allow users to ask for different information (e.g., the weather or a recipe), set up reminders and lists, and to directly control other IoT devices (e.g., lamps), among the various options.

Such assistants, through a companion app installed on their owner' smartphone, provide advanced features like the possibility to set up some **routines** in the form of trigger-action rules (see Figure 1 for an example). IPAs can be part of these rules either as triggers (i.e., *when the user says a specific sentence*) or actions (i.e., *reproduce a specific sentence*), transforming them from intelligent agents to simple sensors and actuators. End-user personalization capabilities, in other terms, are present in the system but segregated in a mobile app and take no advantages from the NLP and vocal capabilities of such devices.

Can we provide IPAs with a more prominent role in such a personalization scenario, directly on the smart speakers? In which steps of the personalization process (creation, explanation and debugging, etc.) those devices can be particularly beneficial?

The reminder of this paper will reflect upon those questions, by highlighting challenges and opportunities towards a better integration of IPAs and smart speakers into end-user development of IoT ecosystems. To do so, we will consider the possible impact of IPAs in smart speakers during the four main steps that might involve personalization rules: a) *creation* (i.e., composition or selection), b) *management*, c) *execution*, and d) *debugging* and explanation (in case of issues).

CEUR-WS.org/Vol-2702/EMPATHY_2020_paper_7.pdf

EMPATHY: Empowering People in Dealing with Internet of Things Ecosystems. Workshop co-located with AVI 2020, Island of Ischia, Italy

luigi.derussis@polito.it (L. De Russis); alberto.monge@polito.it (A. Monge Roffarello)

0000-0001-7647-6652 (L. De Russis); 0000-0002-9746-2476 (A. Monge Roffarello)

© 2020 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

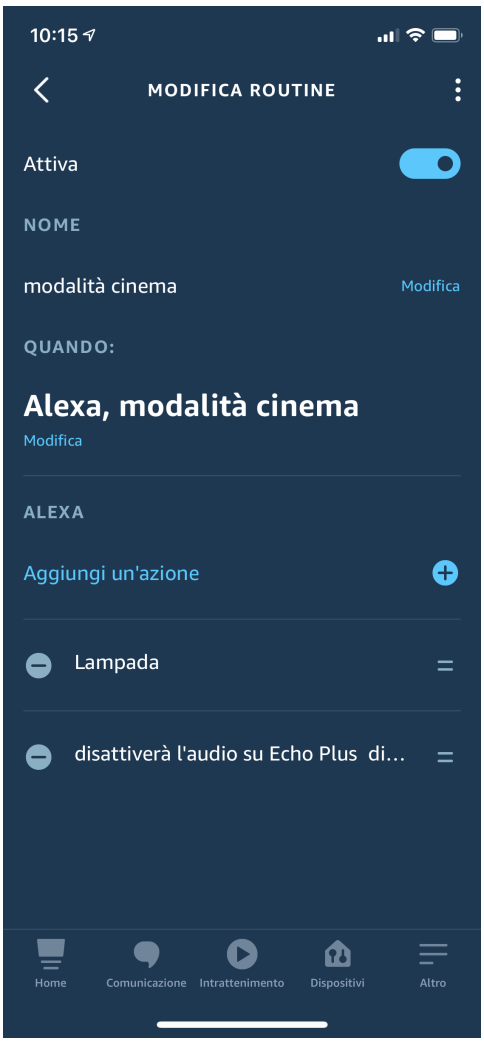


Figure 1: An example of a routine for the Amazon Echo: if the user says "Alexa, cinema mode", the IPA will turn on a lamp with a defined color and luminosity level, and will deactivate the microphone on the smart speaker.

2. Personalization through Conversation

For **creating** a personalization rule, IPAs might either allow selection or direct composition of rules. For what concerns the selection, to our knowledge, InstructableCrowd [1] and HeyTAP [2] are the only two examples of related conversational systems in the literature. InstructableCrowd is a crowd-sourcing system that enables users to create IF-THEN rules based on their needs, via conversation with crowd workers, to describe some problems they are encountering, e.g., being late for a meeting. Crowd workers can then exploit a tailored interface to combine triggers and actions in appropriate IF-THEN rules that are then sent back to the users. HeyTAP, instead, tries to automatically map abstract users' needs to actual IF-THEN rules, i.e., without the help of other users. It adopts a semantic-based approach to analyze users' inputs and contextual information to provide a set of appropriate IF-THEN rules from which a user can choose. Both systems heavily rely on a screen-based interaction (either in a web browser or through a mobile app) and allow an indirect definition of trigger-action rules, by

providing the user with a set of rules to select from. However, both can be easily imagined as working in a smart speaker equipped with a screen. From these works, a few challenges emerge: for instance, when users were asked to elicit their needs, some of them do not know what to say and, especially in HeyTAP, the need could not always be mapped to specific trigger-action rules. Discoverability, i.e., the ability for users to find and execute features through a user interface, is a recurrent problem with IPAs [3] and it is something to consider also in this scenario. Direct composition of IF-THEN rules can be imagined as well, but the process could be long and annoying for the user, and the same issues that already exist for the trigger-action paradigm (e.g., the difficulty in understanding the difference between events and conditions) should be tackled also in this case. In a voice-driven scenario, however, the IPA could provide suggestions and help the user to understand those differences. Alternatively, a totally different paradigm could be envisioned, which could allow end users to directly compose a rule vocally and without the need to explicitly indicate each trigger and action.

Management of existing rules might be an interesting effort, especially for those smart speakers that are not equipped with a screen. Here, the difficulty is not in the command that the end user can provide, as those commands are quite simple (e.g., "delete a rule") or are strongly linked to the creation phase (e.g., "edit a rule"). The challenge is how to present the list of available rules. Probably, reading all the rules with all their details is inappropriate. Similarly, just listing a few information from the rule (e.g., the title) could provide a very limited overview and increase errors. The challenge is to find a balance between these extremes. In addition, should the user be able to select some rules that she wants to manage, for instance through a search mechanism, such as "list the rules that start this morning" or "list the rules that involve the kitchen lamp"?

During the **execution** of rules, IPAs could not play a significant role. However, it is possible to envision a *proactive* role for them: since smart speakers are always-on devices, they could be aware of what is happening and which rules are currently active. They might allow the user to ask for those information and stop some rules to be executed. It remains to understand whether and how much this is useful and appreciated.

Asking for information about rules in execution can be, instead, particularly useful in case of conflicts among rules. IPAs could help to **debug** problematic rules during their execution or, more importantly, to explain why a conflict arose. This could be done automatically by the IPA, as soon as it identifies an issue, or manually by the user, if she notices that something strange is happening, like a lamp that started to blink and never stops. The challenge here is, again, at the presentation level: when is it legit to warn the user about a problem? Who is the user to warn? How can the conflict be explained? Options range by describing why a certain rule (or set of rules) is misbehaving, to allowing the user to deactivate one of them, with various levels of details.

All these steps present other common challenges and opportunities, as we move away from the "common" screen-based interaction. Authorization issues might arise, for example: who is allowed to create, manage, and get explanation? How can this authorization be performed and set up? By recognizing the speaker through her voice?
Procedural issues might be present as well: how long should the conversation last? What happens if the IPA misunderstands a portion of the conversation and activates a problematic rule? Finally, who are the end users in this context? In a home, different people can interact with the system, and they can have different ages, diverse needs, various expertise, and different ways to understand and communicate things.

These and similar questions could spark the discussion during the workshop and, possibly, foster new research ideas around vocal interaction for end-user personalization in the IoT.

References

[1] T.-H. K. Huang, A. Azaria, J. P. Bigham, InstructableCrowd: Creating IF-THEN Rules via Conversations with the Crowd, in: Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems, CHI EA '16, ACM, New York, NY, USA, 2016, pp. 1555–1562. doi:10.1145/2851581.2892502.

[2] F. Corno, L. De Russis, A. Monge Roffarello, HeyTAP: Bridging the Gaps Between Users' Needs and Technology in IF-THEN Rules via Conversation, in: Proceedings of the 2020 AVI Confererence (AVI'20), ACM, 2020. doi:10.1145/3399715.3399905.

[3] P. Kirschthaler, M. Porcheron, J. E. Fischer, What Can I Say? Effects of Discoverability in VUIs on Task Performance and User Experience, in: Proceedings of the 2nd Conference on Conversational User Interfaces, CUI '20, Association for Computing Machinery, New York, NY, USA, 2020. doi:10.1145/3405755.3406119.