

---

# The illusion of natural conversation: interacting with smart assistants in home settings

**Julia Velkovska**

Orange Labs, Sociology and  
Economics Department (SENSE)  
Paris, France  
julia.velkovska@orange.com

**Moustafa Zouinar**

Orange Labs, Sociology and  
Economics Department (SENSE)  
Paris, France  
moustapha.zouinar@orange.com

**Abstract**

Voice-activated intelligent home assistants are presented by their promoters as systems designed to make everyday life easier through “natural” voice interaction. Though the market is currently expanding fast, little is known as yet about how these devices are being used. The spread of voice-activated intelligent assistants raises a number of questions about users' interaction with and appropriation of such tools and the controversies that may arise in the social context of their reception by the users (privacy, for example). This article sheds light on how these systems are actually being used and on the social challenges they pose, based on empirical studies with video data of uses of three different smart home assistants.

**Author Keywords**

human-machine voice interaction; smart home assistants, video-ethnography of users' practices, family life, users' adaptation work, users' skills

**ACM Classification Keywords**

Voice User Interfaces, User Studies, Human-Machine Interaction

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.

CHI'18 Extended Abstracts, April 21–26, 2018, Montreal, QC, Canada ©  
2018 Copyright is held by the owner/author(s). ACM ISBN 978-1-4503-  
5621-3/18/04. <https://doi.org/10.1145/3170427.3170619>

## Introduction

Recent technological developments in speech recognition and synthesis have opened up new possibilities for innovative services, specifically in the development and increasingly widespread use of “conversational” systems in voice-activated virtual assistants for individuals (such as Siri or Google Now) or for the home (such as Amazon Echo or Google Home), online chatbots, and “social” robots, intended to act as “companions”, such as Pepper and Nao.

Promotional material on voice-activated intelligent assistants often emphasises their use of “natural” language or dialogue. By design, however, their modus operandi is based on a two-part voice command (user’s request + system’s response), preceded by an activation word (“Alexa”, “Hello Ivey”, “Ok Google”). Personal virtual assistants such as Siri and Google Now process a wide variety of verbal instructions to run Internet searches, access information and multimedia content, or dictate text messages. In the home, systems like Ivey, Amazon Echo and Google Home can also be used as unified interface to control household appliances (heating, lighting, TV, radio, security devices, etc.) verbally.

The spread of these voice-activated intelligent home assistants raises a number of questions about their use: What is the nature of people’s interactions and relationships with these assistants? How do families appropriate them, and for what purposes? What benefits can the current assistants bring to a home?

To shed light on these questions, we conducted two usage studies in the home environment based on video observations and interviews. The first involved three

English-speaking families and concerned two English-speaking voice-activated intelligent assistants (Ivey and Amazon Echo). The second concerned an experimental prototype called Alice, tested by seven families (Fig. 1).



Figure 1 – Voice-activated intelligent assistants studied and usage situations.

## Users’ work and skills to adapt to voice systems

“Natural” interaction – fluid, simple, effortless and very close to human conversation – is the big promise made in the promotion of intelligent home assistants. Yet our analysis shows major discrepancies between that promise and the reality of the interactions observed.

Above all, speech recognition is not always effective, including for simple requests like the following (U = User):

Example 1 – *Speech recognition failure*

U: hello ivee what time is it in **New York City**?

Ivee: it is now 4:41 am in **Sydney Australia**

Users sometimes have to repeat their questions several times to make themselves understood, which can lead them to give up on the system very quickly. If they do persevere, they engage in a process of “repairing and making sense of the interaction” [4]. This may involve a variety of actions, such as reformulating their statements by shortening them (see example 2) or clarifying them with further details, moving closer to the device, or talking more loudly.

Example 2: *Syntactical adaptation (reduction) of the command*

*Transcription conventions:*

[ = *Time overlap between two events*

(.) = *brief silence*

1.U: (.) Alexia go on to **YouTube and look up Katy Perry please**

2.Alexa : hm (.) I can’t find the answer to the question I heard

3.U: Alexia (.) **go on to You Tube (.) now**

4.Alexa : hm (.) I can’t find the answer to the question I heard

#### 5.U: **Alexia You Tube**

However users repairing practices are an essential aspect of the adaptation work [2]. These efforts go beyond the wording of the commands, encompassing all activities people engage in to get the system to work, including making sense of the system’s inappropriate responses, or learning about the interactional structure imposed (activating the system, then speaking at the right time, when the device is “listening”). Concerning this last point, users sometimes forget to say the activation word or do not speak at the “right” moment.

Finally, the effectiveness of this adaptation work varies from person to person. We have observed, for instance, that young children have great difficulty making themselves understood, which creates a sense of exclusion and generates frustration and resentment. Today’s voice-activated systems thus require a process of adjustment, both individual and collective (in the form of mutual assistance), from members of a household (see example 3 below) as well as guidance and support for children.

Contrary to the promise of fluid, “natural” interaction, this adaptation work is in fact typical of all interactive voice response systems or more generally, with smart systems as shown from the beginning of AI research [1,3]. It may take various forms depending on the context and the type of voice interface (cf. [4 for a study about Interactive Voice Response systems in a telco’s customer relationships). When it takes too much

effort, however, it may prevent users from appropriating the system, or even cause them to abandon it.

### **The illusion of a natural conversation**

The ability of voice-activated intelligent assistants to manage dialogue in a fluid manner is at the heart of the promotional discourse surrounding them. In reality, however, that ability is limited, producing the appearance of a natural conversation which actually fosters users' expectations and thus generates more interactional difficulties.

When systems are working in optimal fashion, users tend to implicitly ascribe capabilities to them that they do not actually possess. An example is the use of indexical terms (such as "now", or "here"), which refer to contextual elements that the systems are incapable of processing. This is what we can see in example 3, where the first user (U1) refers in line 5 to an element mentioned previously ("it", which refers to what is said in line 1, "cooking pork"). We should note, in passing, the speech recognition problem manifested in Alexa's response in line 4 (the system adds what U1 has said to a shopping list), which leads U1 to rephrase her request. Interestingly, we can see that U1 repeats the referent "it" used by the machine (line 4), which may suggest that the machine "understands" the meaning of this kind of indexical term. But this repetition creates difficulties.

Example 3: *Use of indexicals ("it" in this example)*

1. U1: Alexa I [need a r, I need to **cook some pork**
2. Alexa: [light flashes

3. U1: [and-sh

4. Alexa: [I added **it** to your shopping list

5. U1: and I need a recipe **for it** can you find me one?

6. U2: You didn't say Alexa

7. U1: Alexa need I need a recipe **for it** can you find me one?

8. Alexa: I was unable to understand the question I heard

Subsequently, the person forgets the activation word (line 5), a problem that is pointed out by another family member, U2 (line 6) and that U1 then corrects (line 7). But Alexa is unable to provide the expected response, probably because it is incapable of interpreting the indexical "it" repeated in line 7. This incapacity is one of the biggest challenges for Artificial Intelligence: understanding the context. However, the reason for this failure is never explained, and remains unclear for the users involved in the interaction. They thus find themselves in a paradoxical situation with regard to intelligent assistants: the better the system works and the more it uses ordinary conversational practices such as the use of indexicals ("it", here), the more users tend to speak "naturally", and the greater the risk of the dialogue failing without the users being able to determine the cause of that failure. *In other words, these systems create the illusion of a natural conversation, and users are not always aware of the limits of the systems' comprehension capabilities.*

## **Users' trust and the intelligibility of technology**

Designers see voice-activated intelligent assistants as destined to become central to family life, with a close personal relationship struck up between them and their users. These assistants are also "listening" constantly, so as to be able to detect their activation word. That is a source of concern, in that users have no way of knowing what is really being listened to, processed and stored. They see these virtual assistants as black boxes with opaque inner workings of which they need to be wary. As the interview extract below suggests, greater transparency on how the system works, as well as the possibility of controlling it, could help boost trust.

*"It should be clearer when you can disconnect and reconnect. It should be possible to have times in a day where you can say "OK, Alice is sleeping. She's gone to sleep and she's not doing anything. She shuts off. She disappears from your life." (extract from interview – Alice study)*

## **Conclusion**

In addition to the identified limits of voice-activated home assistants, our studies also highlight their potential usefulness as a unified voice interface for accessing all domestic and multimedia appliances. In general, participants saw their potential benefits despite the sometimes erratic quality of speech recognition.

However, our analysis also highlights the considerable adaptation work required from users of today's intelligent assistants, work which could compromise long-term adoption if it requires too much effort. Optimizing speech recognition and dialogue

management and minimizing the amount of effort required from users are therefore crucial avenues for improving these systems. Furthermore, despite recent progress in artificial intelligence, the ability of these machines to provide a truly "natural" interaction remains limited. To avoid users developing excessively high expectations, it is important to talk more realistically about their current capabilities.

Two areas for improvement emerge from our analysis, both concerning the intelligibility of conversational technologies. The first is enabling the user to understand the system's answers and the source of the problem if the system gives an inappropriate answer or is unable to respond at all. The second is guiding them to help them formulate their commands effectively.

In addition to improving virtual assistants' interactional capabilities, other, equally important challenges have to be addressed. For example, managing access to the system according to the speaker's identity is a crucial point, as without authentication, a voice assistant raises a number of security and privacy issues. The privacy issue also arises in relation to the service provider, who must ensure that the personal data collected (household members' conversations and activities) is protected, while offering the people concerned visibility and control over that data. Finally, a number of questions remain unresolved, particularly with regard to how these devices are going to transform the everyday life of household members and the type of relationship and attachment that will exist between them. For instance, we have observed that some participants want a "playful" relationship with their assistants, asking them incongruous questions or trying to discover their "personality". But these

observations need to be backed up by studies over longer periods of use.

Furthermore, the spread of automated voice systems in the personal and family sphere raises debates as to the ethical, legal and political framework surrounding them, focusing on issues such as privacy, trust, autonomy and the delegation of decisions to machines. These

controversies are worth analysing, as they are an integral part of the social framework within which conversational technologies appear.

### **References**

1. Hubert Dreyfus. 1972. *What Computers Can't Do*, New York: MIT Press
2. Paul Luff, David Frohlich, and Nigel Gilber (eds.). 1990. *Computers and Conversation*. Academic Press.
3. Lucy Suchman. 2007. *Human-Machine Reconfigurations. Plans and Situated Actions*, 2nd Edition. Cambridge, Cambridge University Press.
4. Julia Velkovska and Valérie Beaudouin. 2014. Parler aux machines, coproduire un service. Intelligence artificielle et nouvelles formes de contribution du client dans les services téléphoniques de l'après-vente. In *Le travail commercial dans les télécommunications*, Alexandre Mallard and Emmanuel Kessous (Eds), Presses des Mines, Paris, 97-128