# Smart Speakers for Inclusion: How Can Intelligent Virtual Assistants Really Assist Everybody?

Eliseo Sciarretta(✉) and Lia Alimenti

Link Campus University, Rome, Italy
{e.sciarretta,l.alimenti}@unilink.it

**Abstract.** Smart speakers equipped with intelligent virtual assistants allow people to look for information, complete tasks and control other devices without using their hands and eyes, just their voice. Humans can finally use natural language utterances and be fully understood, without being forced to learn the machine language or to handle more or less complicated interaction techniques. Their potential in terms of inclusive design is therefore very high. However, it is important not to fall into the opposite problem, that is, to limit their use to the voice/auditory channel only, excluding all those who can't or don't want to use it. In this paper, the authors analyze the current situation, highlighting the peculiarities of these systems and the reasons why they are quickly gaining ground. Then, they focus on the potential interaction issues and on the challenges still open. After studying the main use cases relating to people with disabilities, elderly and accessibility, the authors can draw a list of suggestions addressing the inclusive design of virtual assistants and smart speakers.

**Keywords:** Virtual assistants · Natural language processing · Inclusion · Accessibility · Smart speakers · Conversational agents

## 1 Introduction

Virtual assistants, that can converse with humans in natural language to provide services, are becoming increasingly popular, pervasive and ubiquitous.

Amazon Alexa, Apple Siri, Google Assistant and Microsoft Cortana (but the full list is much longer) are all variants - produced by different manufacturers - of the same product, which in this paper is usually referred to as Intelligent Virtual Assistant (or IVA), but which is known by many other names: Voice Activated Personal Assistant, Conversational Agent, Virtual Personal Assistant, Voice-Enabled Assistant or Intelligent Personal Assistant (Cowan et al. 2017).

All of these systems share the use of voice as the main interaction channel, through natural language processing (NLP) and speech synthesis processes.

As for the interaction, for the first time in history the visual channel is not essential and is replaced by the hearing one (Cohen Cohen and Balogh 2004). For decades, since the introduction of the first graphical user interfaces (GUI), sight has been the main

M. Kurosu (Ed.): HCII 2021, LNCS 12762, pp. 77–93, 2021.
https://doi.org/10.1007/978-3-030-78462-1_6

sense used to convey information from machines to humans. Hearing, on the other hand, has always been little used in human-computer interaction, where it has generally been limited to the sound being responsible for alerts when something goes wrong, as well as its use for multimedia content, of course.

Now, it seems that "the hottest thing in technology is your voice" (Brunhuber 2018).

When these systems are integrated into devices to be installed at home - in the form of smart speakers - and connected to other smart appliances, people can use them to perform several actions without moving, for example turning on and off lights without reaching the switches or opening the window without moving from the sofa (Masina et al. 2020).

The inclusive potential of IVAs is enormous, as they can be used effortlessly by people who, due to the visual nature of most interactive systems so far, have always been disadvantaged, such as those with limited vision or limited dexterity. Furthermore, voice interaction is considered simpler, and therefore also more acceptable by people with limited literacy on technologies.

Elderly people, for example, can exploit these systems to listen to radio or news and to be assisted in daily services, without having to learn complicated metaphors and gestures to interact with computers or smartphones, and thus maintaining independent living (Kobayashi et al. 2019) without having to be assisted by a caregiver.

But this new perspective can bring problems to other people, such as individuals with speech or hearing disorders.

To achieve full inclusion, it is necessary to ensure that such systems accommodate and can manage everyone's needs. The purpose of this paper is precisely to understand what can be done to maximize the profitable use of IVAs by the largest possible number of individuals, following the inclusive design approach.

The remaining of the paper is structured as follows: Sect. 2 offers an overview on the current state of the market of IVAs and smart speakers, providing definitions, explanations on the functioning of these systems, usage stats and background. Section 3 highlights the main problems identified in the literature regarding the use of assistants, both of a technical nature (recognition problems, irrelevant answers) and of a social nature (excessive personification leading to too much confidence). Section 4, on the other hand, shows the main benefits brought by these systems in terms of interaction, justifying their rapid expansion on the market. Finally, before the conclusion, Sect. 5 illustrates the scenarios related to inclusive design, analyzing the main problems and drawing possible solutions.

## 2 Background

Since the birth of computers and intelligent machines humans have cultivated the dream of being able to interact with them through natural speech language (Hoy 2018) and to receive answers accordingly.

Science fiction, from the 1960s onwards, is full of examples (Chkrou and Azaria 2019), and while some foreshadowed alarmist scenarios (such as HAL 9000 in "2001: A space odyssey"), others were definitely more optimistic, depicting scenarios of seamless integration between humans and machines (Star Trek can also be cited, but the example that best illustrates the idea is K.I.T.T., the talking artificial intelligence installed on the Pontiac Firebird Trans Am, star of the 1980s TV series Knight Rider).

Today, thanks to Intelligent Virtual Assistants, that dream is becoming reality.

IVAs, which, as mentioned, can have different names, are software applications capable of providing real-time services and assistance to users by “answering questions in natural language, making recommendations, and performing actions” (Baber et al. 1993). But unlike other applications, they can take advantage of a voice interface and a conversational dialogue system (Yang et al. 2019).

Aside from the science fiction imagination, the goal of being able to talk to computers has long been pursued; this type of research is part of the broader sector of natural user interfaces (NUI), i.e. systems that allow the user to use them through intuitive and invisible actions (Berdasco et al. 2019), such as touch and gestures, in order to minimize complexity of the systems. In this sense, voice has always been considered as a promising channel, so much so that the first successful experiments in the field of speech recognition are due to studies carried out in the 1950s (Davis 1952).

Other important moments in this approach march, as reported by Rzepka (2019), are the use of pattern recognition methods (1960s), and subsequently the application of statistical methods.

Only in the 1990s the first systems capable of recognizing speech with a certain reliability and responding thanks to text-to-speech synthesis were developed. However, the technology was still at early stages and these systems were mostly used to “dictate” more or less long texts to computers.

With the new millennium, however, the steps forward in the field of AI, cloud computing and the Internet of Things open up new scenarios, so much so that the IT giants are engaged in a competition to be the first to hit the market with a solution that can be controlled by voice. The competition had a winner in 2011, when Apple launched Siri, the first modern commercial virtual assistant, changing the whole scenario: it is no longer just a system capable of managing simple question/answer cycles, but a real assistant able to extrapolate data and keywords from the user’s speech to obtain in-depth knowledge and offer services in exchange (Knote et al. 2019).

Siri is the result of a long research carried out by Apple, which began with the CALO project (Mark and Perrault 2004), but since then the competition has been tight thanks to solutions developed by, among others, Microsoft, Google and Amazon. The latter, thanks to its suite of products connected to Alexa intelligence, has quickly become the industry leader. To get an idea of the proportions, it may be useful to remember that Amazon’s market share in US households in 2018 was about 70% (Griswold 2018).

Aside from the competition in the industry, it can be said that all the players involved have contributed to change the way people can receive services, search for information and control their devices. In fact, already in 2018, the data (McCue 2018) showed that over a quarter of people who use online services are already accustomed to voice search, with a marked growth trend.

Juniper Research (2018), indeed, predicts a 1000% increase in the use of IVAs in the home environment from 2018 to 2023.

As for smart speakers, the data are comparable: the study conducted by Markets and Markets (2018) indicates an estimated growth in the value of the global market from 1.5 billion in 2017 to almost 12 billion in 2023.

Indeed, home-environment smart speakers are the most gaining ground form among those that so-called conversational agents can take: being integrated into smartphones (Apple Siri), operating on regular computers or tablets (Microsoft Cortana, Samsung Bixby), through online services (the various chatbots that handle the customer care for many companies), or even be installed on cars (Mercedes Benz User Experience).

Smart speakers are relatively simple devices, equipped with at least one microphone and a loudspeaker to be able to receive user inputs and provide answers. Some may also have a touch screen, and therefore integrate GUI and can also be controlled through other channels. The difference, in this case, is between voice-based devices, which have a single interaction mode, and voice-enhanced devices (Rzepka 2019), with multimodal interfaces.

The speaker intelligence, however, does not reside within it, but relies on a cloud-based architecture. For example, the line of smart speakers launched by Amazon is called Echo, but the beating heart is Alexa, the artificial intelligence that resides on Amazon's servers and is invoked every time. The speaker is therefore configured as an IoT device, which requires an always-on connection to be operational.

Each time the user makes a request to Echo (but the same applies to similar products from other manufacturers), speech is recorded by the microphones, sent over the connection to the servers and there it is converted into text and interpreted, so that Alexa can process an appropriate response, which is ultimately sent back to the smart speaker and delivered to the user through the hearing channel.

The assistant is always listening, but to limit privacy problems it is activated only when a certain wake word (like "Alexa", or "Ok Google", or "Hey Siri") is spoken; from that moment it starts recording.

Thanks to these features, IVAs can handle complex conversations with users, up to the point of giving the illusion of talking to another human being.

Moreover, the potential of these tools can be increased thanks to the openness that producers have granted to third-party developers: smart speakers can thus continuously learn new "skills" (in the case of Amazon), or "actions" (as far as concerns Google), which are nothing more than plug-ins created by independent companies and developers, through the platforms made available by the producers themselves.

In this way, it is possible to extend the functionalities of IVAs and integrate other devices, such as smart home appliances, within a single ecosystem.

Skills and actions play the same role as mobile applications in Android or iOS, but they are not comparable, as they are not software hosted on the device, but only extensions of services available in the cloud and which can be invoked by users.

From a technical point of view, the rise of IVAs can be explained by the maturity of the NLP sector, due to four main factors, according to Hirschberg and Manning (2015):

1. a vast increase in computing power,
2. the availability of very large amounts of linguistic data,
3. the development of highly successful machine learning methods,

4. a much richer understanding of the structure of human language and its deployment in social contexts.

## 3 Issues and Challenges

Given the increasing interest in the sector of smart speakers, IVAs and voice interaction, these issues have been widely analyzed in recent years research, leading to the emergence of some specific characteristics but also of possible problems to be taken into consideration.

The first question that can be asked when approaching a system like Echo and Alexa is: how should we talk, considering that we are talking to a machine? Can we use the same techniques as in conversation with other humans?

Or, in other words, is the interaction with an assistant really a conversation? This problem was addressed by Arend (2018), with results leading us to think that there are considerable differences in various facets: when we talk to other human beings, we can assume that they remember the previous turns of the conversation, that they have memory of what has already happened, while this is not always true for IVAs, even if for example the developers of Siri are working to let it keep track of the conversation and bind the commands to the previous ones.

Furthermore, during a face-to-face conversation the voice channel is only one of those involved, while we also make a lot of use of the visual one, for example, to interpret the signals that our interlocutor sends us, such as his willingness to listen, or recipiency: to comply, IVAs use visual cues, such as lights, to show that they are active and ready to answer (or that something is wrong).

This, however, leads to further consideration that the hearing channel alone is not enough.

The choice of several manufacturers to integrate in their devices other input and output systems (physical buttons, lighting systems, companion apps) (Spallazzo et al. 2019) shows that to exploit the potential of IVAs it is necessary to expand the spectrum of possible interactions. Fortunately, as far as the purpose of this paper is concerned, this discovery is very useful in terms of inclusive design, because it allows the designers to manage the interaction through multi-modality and thus satisfy a wide range of preferences.

Furthermore, looking at the rules of conversation, it should be noted that due to their design, IVAs fail to replicate the ability of human beings to speak and listen at the same time, and therefore to manage speech overlaps with elegance. An assistant either speaks or listens, it can’t do both things at the same time, so it is the human being who has to adapt to this mechanism.

From what has been said, a consideration emerges that designers are learning: it is better not to make the user believe that the assistant is like a real human being. IVAs should not be anthropomorphized.

In fact, the possibility of speaking to these systems and obtaining an answer leads to the attribution of human characteristics to IVAs (Friederike et al. 2012; Lopatovska and Williams 2018). To acknowledge that, it’s enough to think that almost all agents have a name (Purington et al. 2017), which is associated with a gender identity (almost always female, which can lead to an amplification of gender stereotypes (Habler et al. 2019)) and a consistent personality, generally using helpful and submissive language.

Dazzled by these characteristics, we tend to socialize with agents, almost to consider them friends. It often happens with technological devices (Schwind et al. 2019), but the phenomenon is observed to happen more with assistants.

This trend is settled above all in younger people, while adults manage to consider them as productivity tools, as emerged in the studies of Sciuto (2018) and Li and Yangisawa (2020).

Still, the personification can lead users to overestimate the capabilities of IVAs, expecting unattainable results from them, ultimately generating frustration with the interaction.

Manufacturing companies themselves are promoters of this behavior, pushing designers, through their guidelines, to use everyday language (Branham and Mukkath Roy 2019), slang and avoid "robotic" conversation, so that it is as natural as possible, and to limit the length (the number of words) and complexity (the number of intents) of the communication.

Recognition of intents is the core of how IVAs work. The systems must be able to fill in the empty slots and obtain all the fundamental variables for understanding the request starting from what is said by the user, thus identifying the keywords (Li and Yangisawa 2020). Each intent can be uttered in a number of different ways and IVAs need to be able to recognize as many of them as possible. To facilitate the purpose, designers can therefore try to reduce the complexity, guiding the user to provide the necessary information from time to time.

However, setting levels of complexity calibrated downwards is not always the best possible choice, especially in terms of inclusion.

For example, it can be hypothesized that blind people, with a more sensitive hearing and already accustomed to the use of assistive technologies with speech output, can sustain a higher level of complexity, and indeed they may prefer it (Abdolrahmani et al. 2018), to optimize the use experience.

Even the choice of speech speed and intonation can vary from case to case: blind people, accustomed to the use of synthetic voices of screen readers, may prefer more robotic voices and a speech output at a rate far faster than a human could.

Therefore, focusing on the average user would be a mistake for designers, because they would risk excluding the vast majority of individuals. Instead, they should prefer to offer the possibility of customizing the experience.

The use of voice as the main output system brings a challenge regarding the discoverability of the services available and the learnability of the system in general, as defined by Grossman (2009) "the ease with which new users can begin effective interaction and achieve maximal performance".

The voice is by its nature ephemeral (Corbett and Weber 2016), it does not allow users to build an adequate mental model of how the system works, because they risk losing important information if they get distracted even for just one moment.

Users are completely unaware of the features that the system offers and can only discover them if they make the right requests. But many get stuck because they don't even know what they can ask for. They cannot explore the system by casually wandering around and then interpreting the responses as in GUIs.

The problem was also studied by White (White 2018), according to whom it remains an open challenge, also because it is further magnified by the presence of skills or actions, which continues to increase day after day with minimal tracking possibilities.

In the case of Alexa, for example, there are thousands of skills available, created by third party developers, but among these very few are known and used: the result is that many features remain hidden (Spallazzo et al. 2019).

Therefore, IVAs implement strategies to help users, such as suggestions on new things to try, but a list of suggestions pronounced one after the other is likely to get the user even more confused.

To make complex responses more effective, IVAs should offer the possibility of using other channels as well, through companion apps or connected devices, and be able to better exploit context information (for example about environment and time), especially in mobility contexts.

In general, it would be useful to improve assistants' proactivity in providing useful advice on their abilities, possibly at the right time, exactly when needed.

Assistants, to improve their efficiency, should be able to understand when users are turning their attention to them and respond accordingly, perhaps through sensors that can detect presence, so as to be activated when users are passing by (Spallazzo et al. 2019).

As specified by Iannizzotto (2018), "Current smart assistants can talk and listen to their users, but cannot 'see' them". Plus, they are often faceless. This makes the communication somehow incomplete and therefore less effective. If until a few years ago facial recognition was inaccurate due to technological limitations, today limitations have gone and recognition is already used in smartphones, so it could be applied also to IVAs.

Of course, the use of similar techniques and sensors inevitably triggers a potentially endless discussion on privacy and data security issues. This is certainly a fundamental theme, already much explored in the literature, but in this paper is deliberately omitted in order not to shift attention from the subject of the research.

Similarly, in this paragraph some challenges have been highlighted that need to be addressed also from the point of view of inclusion, but it has been chosen to leave out further problems present in the sector, of a more technical nature, such as the recognition of multiple voices, background noises, the need to repeat the same command several times (Pyae and Joelsson 2018), the interferences that can arise in case of involuntary pronunciation of the wake word, etc.

## 4 Interaction

Challenges still need to be faced, but IVAs and smart speakers are showing unique characteristics in the field of human-machine interaction, which can be exploited to optimize the user experience for certain tasks and in certain contexts, and which can bring benefits also in terms of inclusive design. According to many, in fact, voice assistants are changing traditional forms of human-computer interaction (Feng et al. 2017). Furthermore, conversational technologies are considered "transformative" and represent a shift towards a more "natural" computing paradigm (The Economist 2017).

First of all, we need to wonder what drives users to use such systems: the studies by Rzepka (2019) indicate efficiency, convenience, ease of use, minimal cognitive effort, and enjoyment as the main objectives to increase the value of IVAs perceived by users.

To this extent, agents offer numerous benefits, as they can be used hands-free, without worrying too much about grammatical errors (especially if compared to written communication) and in a way that intrigues users and entertains them (Terzopoulos and Satratzemi 2019).

The main and most evident advantage in terms of interaction proposed by IVAs is the possibility of freeing the hands. As a result, users can do other things in the meantime, supporting multi-tasking, with obvious time advantages (Luger and Sellan 2016).

But not only the hands are free, the eyes are too (Moussawi 2018), giving us the opportunity to focus our attention on something else.

Of course, what for some is at best a competitive advantage, for others becomes a necessity and the only way to obtain services: these systems therefore are very promising in relation to the needs of people with visual and motor disabilities (Branham and Mukkath Roy 2019), as will be explained below.

As mentioned previously, however, hands-free and eyes-free interaction is actually achieved only in some cases, depending on the characteristics of the tasks, objectives and contexts that define the degree of complexity of the situation.

Speech interaction offers its best in situations of low complexity (Zamora 2017) or when multi-tasking can be exploited (Luger and Sellan 2016).

Therefore, IVAs, at home, act as assistants in the daily life of users (McLean and Osei-Frimpong 2019), able to offer simple but useful services such as setting timers and alarms, managing the agenda, searching for information and also the management of connected smart appliances, leaving users free to think about other things. These systems offer greater convenience than any other kind of device, allowing users to complete tasks with little effort and without the need to type, read or hold a device (Hoy 2018).

From what has been said, also in the previous paragraph, the need emerges for a hybrid use of IVAs, which ranges, depending on the occasion, from an exclusively vocal interaction, to one that also integrates other methods, which can be conveyed through touch screens on the devices themselves, tactile buttons, but also through the so-called companion apps. These are mobile apps that have the basic task of guiding the user through the configuration of the device, but which can then be exploited in all cases where complexity requires it. Indeed, all manufacturers are doing this by providing support apps and a range of devices for all needs.

As a result, simple tasks can also be performed through voice alone, in this way the process is lighter and thanks to a single spoken command, a series of gestures such as touch, scrolling and input are avoided (McTear et al. 2016). And this mode also proves perfect when dealing with small screens, where the physical limitations of the device's real estate make it difficult to provide input differently.

But when dealing with complex tasks, the most effective model is rich interaction, where the auditory channel is joined and supported by the visual, and possibly also by the haptic one. Siri, for example, has long been providing multimodal answers, with the transcription on the screen of what it says verbally, together with the relevant information found based on the request made. Not only that: the request made by the user is also shown as written text, as proof of the correct understanding of the command.

The feedback of the system is therefore composed of multiple interactions: lights, verbal expressions, screens and in some cases even movements (Spallazzo et al. 2019) signal the status of the system at all times, and this also helps in terms of inclusive design.

## 5 Inclusive Design

As already mentioned several times in the previous paragraphs, the purpose of this article is to frame the growing phenomenon of IVAs and connected smart speakers in an inclusive design perspective, which therefore favors the greatest number of people, regardless of their skills and preferences (Sciarretta 2020).

For this it is useful to analyze the accessibility characteristics identified in these systems, in order to highlight what needs to be done to ensure inclusion.

In general, smart speakers, especially when connected to other smart appliances (Abdolrahmani et al. 2018), are showing great potential in terms of assistance, because they allow people to have a single hub to control the home environment without having to learn complicated interfaces.

After all, technologies such as automatic speech recognition (ASR) and text-to-speech (TTS) are well known in the world of accessibility as assistive technologies used for decades (Ballati et al. 2018).

However, they are generally used as a sort of alternative to the main mode of interaction, for example to convert visual elements into auditory feedback, as happens in the case of screen readers, which transform the visual information present on a screen into synthetic speech (Jacko et al. 2008).

In the case of IVAs, instead, the perspective is reversed because the voice becomes the main channel of interaction, if not the only one. However, a consideration arises: as mentioned above, very often IVAs are accompanied by companion apps for smartphones, which are used for setup but which can also be useful for inclusion purposes, since they offer the possibility of providing richer feedback. Obviously, this happens as long as the app is also designed in an accessible way, to ensure that the entire ecosystem is inclusive (Pradhan et al. 2018).

Furthermore, the vocal interaction mode is very successful among people with disabilities, making it the preferred choice by those with limited hand dexterity (Peres 2019), but also by those with intellectual disabilities (Balasuriya et al. 2018), the elderly (Schlögl et al. 2013) and of course by the blind (Corbett and Weber 2016).

However, there are still numerous accessibility challenges (Morris and Thompson 2020), mainly due to the fact that what is accessible to one person is not necessarily accessible to another.

Exactly for this reason, in the remaining part of the paragraph the information provided will be divided by type of user, in order to evaluate the gray areas more carefully, focusing in particular on physical problems, such as vision disorders, hearing problems, speech difficulties and motor limitations, without, however, forgetting the large category of elderly people, who may have one or more of these problems.

Instead, no discussion will be provided about the category of people with cognitive disabilities, due to the impossibility of reducing all possible cases to a single one. But it is clear that to achieve true inclusion, solutions capable of meeting their multiple needs must also be designed.

This division is necessary because every situation is unique and brings different needs: for example, it is intuitive to think that people with motor disabilities can enjoy the greatest benefits from IVAs (Ballati et al. 2018), while deaf people may have greater difficulties.

**Hearing and Speech.** People with hearing problems, who can also experience language difficulties, are among those who can have the greatest complications in using voice-based virtual assistants. Studies carried out on Google's speech recognition system have shown poor results in the presence of deaf or hard of hearing people (Bigham et al. 2017).

Furthermore, as mentioned, language difficulties must also be considered, which can be blocking: in the studies conducted by Pradhan (2018), for example, it emerged that the greatest problems are the need to speak loudly, otherwise the assistant will not be able to perceive the command, and respecting a precise timing in uttering the request; in fact, systems are generally designed for people who can make intelligible and clear speeches (Masina 2020).

IVAs may misinterpret slightly longer pauses, and think they are sentence delimiters (Kobayashi et al. 2019). In addition, the user can speak at different speeds and the assistant, which instead requires a fairly precise timing, could get confused. When the wake word is pronounced, for example, the device switches to listening mode for a specific amount of time; if the user waits too long, the system times out. For those with speech problems, the time available may be too short.

In this case, therefore, the accessibility challenge is to design assistants so that they can adapt to users' needs, with algorithms that can improve their speech comprehension skills. Furthermore, it is necessary to offer users the ability to manage settings in order to select their preferences.

**Vision.** There are approximately 285 million people with severe vision impairment worldwide (WHO 2010).

The difficulties related to the visual spectrum have always been among the most limiting ones in the use of computers, the Internet and technologies that rely on graphical user interfaces (Iyer et al. 2020). On the other hand, research on accessibility has often put these problems first, identifying solutions that today allow blind people to effectively use an iPhone or other technologies.

However, from an inclusion perspective, being able to count on devices that are not based on graphic interfaces but on conversational interfaces is a huge step forward, because it makes the visual impairment completely marginal, granting people with visual disabilities the same usability of the tools that everyone else can experience.

Assistive technologies, on the other hand, immediately make it clear that the users need them, risking making them feel disadvantaged (Desmond et al. 2018).

Also, assistive technologies can be very expensive (Beksa and Desmarais 2020). Instead, as already noted by Gill (2017), IVAs are low-cost solutions.

For the blind, the problems that can arise in the use of these tools are linked to their over-ability: being people used to exploiting the hearing channel in interactions with machines, they may not like the excessive verbiage of IVAs, preferring instead a more direct communication.

Therefore, a different approach to the use of these systems emerges, where they are considered as serious tools by the blind (Azenkot and Lee 2013), one of the best possibilities for completing complex tasks, while for sighted users they can become an entertainment pastime (Luger and Sellan 2016; Pradhan et al. 2018).

Furthermore, due to the habit of using the voice channel, blind people prefer much faster speech (Branham and Kane 2015) and consider the default speed to be a waste of time.

For this reason, IVAs designers should provide the ability to set different preferences regarding speed but also word count and general complexity, which would allow better use of the tool.

Although in many smart speakers the voice is used as the main interaction mode, some feedback, as already mentioned, is also provided in other ways, for example through lighting systems. Obviously, to ensure accessibility to the blind, these visual cues must be presented effectively through alternatives (Abdolrahmani et al. 2018), perhaps through short auditory icons, otherwise known as earcons.

**Limited Dexterity.** For people with limited dexterity some of the same points are valid as for people with vision problems: IVAs are an exceptional opportunity because they allow them to complete tasks without having to use their hands or perform gestures, guaranteeing levels of independence never experienced before.

However, some problems remain, mostly related to the usability of the systems, such as the difficulty in discovering and learning the features, as already seen.

Furthermore, it must also be considered that some motor problems lead, as a secondary effect, to language problems (Duffy 2013), which involve everything that has already been discussed.

Designers must therefore try to properly manage the voice as a primary interaction mode, leaving more complex tasks to other types of interaction. The advantages of this type of design would be perceived not only by those who have limited dexterity due to physical causes, but also by those who are unable to use their hands due to the context.

**Elderly.** The elderly belong to a category of people very different from those investigated so far, but they share some of the same problems, related to skills that are no longer 100% efficient, with a decline in multiple fields related to the sensory, motor or cognitive sectors (Kobayashi et al. 2019), such as sight or motor abilities (Ho 2018).

Furthermore, elderly people are recognized as those who can receive more benefits from the use of IVAs, because they allow them to bypass the use of more complicated technologies and therefore reduce the generational digital divide.

The extensive use of graphic user interfaces, in fact, makes the interaction implicitly more complex, because GUIs allow people to manage complex tasks. The use of the voice as the main channel, on the other hand, allows to reduce the difficulty, as a result of the limitation of the complexity of the tasks (Sayago et al. 2019).

The popularity of assistants among the elderly is growing, not only for the reasons just mentioned, but also for the possibility of completing tasks without disturbing other processes (Terzopoulos and Satratzemi 2019); voice inputs are the most effective modality according to Smith's studies (2015).

To optimize the user experience of the elderly, it is necessary to overcome the accessibility challenges already described, always providing customization options.

The main problems are in fact due to the management of pauses and the difficulties related to the occasions when there is a need for a repetition of what has been said or, worse, a rephrasing (Kobayashi et al. 2019); on these occasions, it would be necessary to manage the error messages in a more complete and personalized way, which can better explain what went wrong and how to remedy it.

From the considerations made in this paragraph we can draw a list of suggestions that can help in designing for inclusion. Designers of virtual assistants should make sure that their systems:

- use voice as a primary channel of interaction, but also offer richer feedback through integrated screens or connected devices;
- provide auditory alternatives to visual cues such as lights through earcons;
- offer a wide range of customizations in terms of modes, times and speed of speech input and output, like
  - waiting time,
  - output speed,
  - complexity of speech;
- handle errors and misunderstandings more comprehensively, including by providing examples;
- clearly indicate when the request is accepted by the system and when it is completed;
- avoid a one-size-fits-all design approach, trying instead to adapt to be used by as many people as possible;
- clearly show they are artificial intelligence, to avoid marked anthropomorphization phenomena, which can push people to overestimate the skills of assistants;
- (connected to the previous point) are also visually represented, to facilitate interaction and rich feedback; the representation should be abstract, not human-like;
- offer the users a way to teach them new commands or offer shortcuts to issue multiple commands at once;
- exploit additional input channels (such as cameras and sensors) to be proactive and to activate at the right time depending on the context.

## 6 Conclusion

In this paper we have tried to highlight the inclusive aspects of smart speakers and Intelligent Virtual Assistants. To do so, we analyzed the interaction characteristics of these systems, the possible problems, the challenges still open and their use made by different categories of people.

What emerged is that IVAs have shown a high degree of acceptance by people, and therefore their use as assistive technologies and for inclusion purposes is very promising, as they can allow people to improve their quality of life (Masina et al. 2020), are less expensive and non-stigmatizing, since they can be used by people with or without disabilities.

However, problems remain to be addressed and in the course of the discussion we have identified some of them, also providing suggestions for improvement.

But apart from the specific problems, designers should adopt a more inclusive approach, considering the needs and preferences not only of as many categories of people as possible, but also in relation to a wide variety of situations and contexts.

Voice promises to change the way of interacting with machines, offering an engaging and natural user interface (Luger and Sellan 2016), but to allow this promise to come true some work is needed to identify ways that can help people manage more complex tasks, without falling into the trap of recognition errors or inconclusive answers.

## References

Abdolrahmani, A., Kuber, R., Branham, S.M.: Siri talks at you: an empirical investigation of voice-activated personal assistant (VAPA) usage by individuals who are blind. In: Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS 2018), pp. 249–258 (2018). https://doi.org/10.1145/3234695.3236344

Arend, B.: Hey Siri, what can I tell about Sancho Panza in my presentation? Investigating Siri as a virtual assistant in a learning context? pp. 7854–7863 (2018). https://doi.org/10.21125/inted.2018.1874

Azenkot, S., Lee, N.B.: Exploring the use of speech input by blind people on mobile devices. In: Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS 2013), pp. 11:1–11:8 (2013). https://doi.org/10.1145/2513383.2513440

Baber C.: Developing interactive speech technology. In: Interactive Speech Technology: Human Factors Issues in the Application of Speech Input/Output to Computers. Taylor & Francis, Inc., Bristol (1993)

Balasuriya, S.S., Sitbon, L., Bayor, A.A., Hoogstrate, M., Brereton, M.: Use of voice activated interfaces by people with intellectual disability. In: Proceedings of the 30th Australian Conference on Computer-Human Interaction (OzCHI 2018), pp. 102–112 (2018). https://doi.org/10.1145/3292147.3292161

Ballati, F., Corno, F., De Russis, L.: Assessing virtual assistant capabilities with Italian dysarthric speech. In: Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS 2018), pp. 93–101, Association for Computing Machinery, New York (2018). https://doi.org/10.1145/3234695.3236354

Berdasco, A., López, G., Diaz, I., Quesada, L., Guerrero, L.A.: User experience comparison of intelligent personal assistants: Alexa, Google Assistant, Siri and Cortana. In: Proceedings of the 13th International Conference on Ubiquitous Computing and Ambient Intelligence UCAmI, vol. 31, no. 1, p. 51 (2019). https://doi.org/10.3390/proceedings2019031051

Beksa, J., Desmarais, A., Terblanche, M.: Usability study of blind foundation's Alexa library skill & low vision NZ (formerly the Blind Foundation) (2020)

Bigham, J.P., Kushalnagar, R., Huang, T.K., Flores, J.P., Savage, S.: On how deaf people might use speech to control devices. In: Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS 2017), pp. 383–384 (2017). https://doi.org/10.1145/3132525.3134821

Branham, S.M., Kane, S.K.: The invisible work of accessibility: how blind employees manage accessibility in mixed-ability workplaces. In: Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility (ASSETS 2015), pp. 163–171 (2015). https://doi.org/10.1145/2700648.2809864

Branham, S.M., Mukkath Roy, A.R.: Reading between the guidelines: how commercial voice assistant guidelines hinder accessibility for blind users. In: The 21st International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS 2019), pp. 446–458. Association for Computing Machinery, New York (2019). https://doi.org/10.1145/3308561.3353797

Brunhuber, K.: The hottest thing in technology is your voice. http://www.cbc.ca/news/technology/brunhuber-ces-voice-activated-1.4483912. Accessed Feb 2021

Chkrou, M., Azaria, A.: LIA: a virtual assistant that can be taught new commands by speech. Int. J. Hum.-Comput. Interact. **35**(17), 1596–1607 (2019). https://doi.org/10.1080/10447318.2018.1557972

Cohen, M.H., Giangola, J., Balogh, J.: Voice User Interface Design. Addison-Wesley Professional, Boston (2004)

Corbett, E., Weber, A.: What can I say? Addressing user experience challenges of a mobile voice user interface for accessibility. In: Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI 2016), pp. 72–82. Association for Computing Machinery, New York (2016). https://doi.org/10.1145/2935334.2935386

Cowan, B.R., et al.: What can I help you with?: Infrequent users' experiences of intelligent personal assistants. In: Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI 2017), pp. 43:1–43:12 (2017). https://doi.org/10.1145/3098279.3098539

Davis, K.H., Biddulph, R., Balashek, S.: Automatic recognition of spoken digits. J. Acoust. Soc. Am. **24**, 637–642 (1952)

Desmond, D., et al.: Assistive technology and people: a position paper from the first global research, innovation and education on assistive technology (GREAT) summit. Disabil. Rehabil. Assist. Technol. **13**, 1–8 (2018)

Duffy, J.: Motor Speech Disorders E-Book: Substrates, Differential Diagnosis, and Management. Elsevier Health Sciences, Philadelphia (2013)

Feng, H., Fawaz, K., Shin, K.S.: Continuous authentication for voice assistants. In: Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking, pp. 343–355 (2017)

Friederike, E., Kuchenbrandt, D., Bobinger, S., de Ruiter, L., Hegel, F.: If you sound like me, you must be more human: on the interplay of robot and user features on human-robot acceptance and anthropomorphism. In: Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction, pp. 125–126. ACM (2012)

Gill, M.: Adaptability and affordances in new media: literate technologies, communicative techniques. J. Pragmatics **116**, 104–108 (2017)

Griswold, A.: Even Amazon is surprised by how much people love Alexa (2018). https://qz.com/1197615/even-amazon-is-surprised-by-how-much-people-love-alexa/. Accessed Feb 2021

Grossman, T., Fitzmaurice, G., Attar, R.: A survey of software learnability: metrics, methodologies and guidelines. In: Proceedings of the 27th International Conference on Human Factors in Computing Systems (CHI 2009), pp. 649–658 (2009). https://doi.org/10.1145/1518701.1518803

Habler, F., Schwind, V., Henze, N.: Effects of smart virtual assistants' gender and language. In: Proceedings of Mensch und Computer 2019 (MuC 2019), pp. 469–473. Association for Computing Machinery, New York (2019). https://doi.org/10.1145/3340764.3344441

Hirschberg, J., Manning, C.D.: Advances in natural language processing. Science **349**(6245), 261–266 (2015). https://doi.org/10.1126/science.aaa8685

Ho, D.K.: Voice-controlled virtual assistants for the older people with visual impairment. Eye (Lond) **32**(1), 53–54 (2018). https://doi.org/10.1038/eye.2017.165

Hoy, M.B.: Alexa, Siri, Cortana, and more: an introduction to voice assistants. Med. Ref. Serv. Q. **37**(1), 81–88 (2018)

Iannizzotto, G., Bello, L.L., Nucita, A., Grasso, G.M.: A vision and speech enabled, customizable, virtual assistant for smart environments. In: 2018 11th International Conference on Human System Interaction (HSI), Gdansk, pp. 50–56 (2018). https://doi.org/10.1109/HSI.2018.8431232

Iyer, V., Shah, K., Sheth, S., Devadkar, K.: Virtual assistant for the visually impaired. In: 5th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, pp. 1057–1062 (2020). https://doi.org/10.1109/ICCES48766.2020.9137874

Jacko, J.A., Leonard, V.K., McClellan, M., Scott, I.U.: Perceptual impairments: new advancements promoting technological access. In: Sears, A., Jacko, J.A. (eds.) The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications, pp. 853–870, Taylor & Francis Group, New York (2008)

Juniper Research: voice assistants used in smart homes to grow 1000%, reaching 275 million by 2023, as Alexa leads the way (2018). https://www.juniperresearch.com/press/press-releases/voice-assistants-used-in-smart-homes. Accessed Feb 2021

Knote, R., Janson, A., Söllner, M., Leimeister, J.M.: Classifying smart personal assistants: an empirical cluster analysis. In: Proceedings of the 52nd Hawaii International Conference on System Sciences, Maui (2019)

Kobayashi, M., et al.: Effects of age-related cognitive decline on elderly user interactions with voice-based dialogue systems. In: Lamas, D., Loizides, F., Nacke, L., Petrie, H., Winckler, M., Zaphiris, P. (eds.) INTERACT 2019. LNCS, vol. 11749, pp. 53–74. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-29390-1_4

Li, C., Yanagisawa, H.: Intrinsic motivation in virtual assistant interaction for fostering spontaneous interactions. ArXiv abs/2010.06416 (2020)

Lopatovska, I., Williams, H.: Personification of the Amazon Alexa: BFF or a mindless companion. In: Proceedings of the 2018 Conference on Human Information Interaction & Retrieval (CHIIR 2018), pp. 265–268. Association for Computing Machinery, New York (2018) https://doi.org/10.1145/3176349.3176868

Luger, E., Sellen, A.: Like having a really bad PA: the gulf between user expectation and experience of conversational agents. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI 2016), pp. 5286–5297 (2016). https://doi.org/10.1145/2858036.2858288

Mark, W., Perrault, R.: Calo: a cognitive agent that learns and organizes (2004)

Markets and Markets: Smart speaker market by IVA (Alexa, Google Assistant, Siri, Cortana), Component (Hardware (Speaker Driver, Connectivity IC, Processor, Audio IC, Memory, Power IC, Microphone,) and Software), Application, and Geography - Global Forecast to 2023 (2018). https://www.marketsandmarkets.com/Market-Reports/smart-speaker-market-44984088.html?gclid=EAIaIQobChMIs6Sn3abE5AIVFozICh1-PQLgEAAYASAAEgIZSvD_BwE. Accessed Feb 2021

Masina, F., et al.: Investigating the accessibility of voice assistants with impaired users: mixed methods study. J. Med. Internet Res. **22**(9), e18431 (2020). https://doi.org/10.2196/18431

McCue, T.J.: Okay Google: voice search technology and the rise of voice commerce. Forbes Online (2018). https://www.forbes.com/sites/tjmccue/2018/08/28/okay-google-voice-search-technology-and-the-rise-of-voice-commerce/#57eca9124e29. Accessed Feb 2021

McLean, G., Osei-Frimpong, K.: Hey Alexa ... examine the variables influencing the use of artificial intelligent in-home voice assistants. Comput. Hum. Behav. **99**, 28–37 (2019)

McTear, M., Callejas, Z., Griol, D.: The Conversational Interface: Talking to Smart Devices. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-32967-3

Morris, J.T., Thompson, N.A.: User personas: smart speakers, home automation and people with disabilities. J. Technol. Persons Disabil. **8** (2020)

Moussawi, S.: User experiences with personal intelligent agents: a sensory, physical, functional and cognitive affordances view. In: Proceedings of the 2018 ACM SIGMIS Conference on Computers and People Research, pp. 86–92. ACM (2018)

Peres, S.: 39 million Americans now own a smart speaker, report claims. TechCrunch (2019). https://techcrunch.com/2018/01/12/39-million-americans-now-own-a-smart-speaker-report-claims/. Accessed Feb 2021

Pradhan, A., Mehta K., Findlater, L.: Accessibility came by accident: use of voice-controlled intelligent personal assistants by people with disabilities. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. Paper 459, pp. 1–13. Association for Computing Machinery, New York (2018). https://doi.org/10.1145/3173574.3174033

Purington A., Taft, J.G., Sannon, S., Bazarova, N.N., Hardman Taylor, S.: Alexa is my new BFF: social roles, user satisfaction, and personification of the Amazon Echo. In: Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems, pp. 2853–2859. Association for Computing Machinery, New York (2017). https://doi.org/10.1145/3027063.3053246

Pyae, A., Joelsson, T.N.: Investigating the usability and user experiences of voice user interface: a case of Google home smart speaker. In: Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct (MobileHCI 2018), pp. 127–131. Association for Computing Machinery, New York (2018). https://doi.org/10.1145/3236112.3236130

Rzepka, C.: Examining the use of voice assistants: a value-focused thinking approach. In: AMCIS (2019)

Sayago, S., Barbosa Neves, B., Cowan, B.R.: Voice assistants and older people: some open issues. In: Proceedings of the 1st International Conference on Conversational User Interfaces (CUI 2019), Article 7, pp. 1–3. Association for Computing Machinery, New York (2019). https://doi.org/10.1145/3342775.3342803

Schlögl, S., Chollet, G., Garschall, M., Tscheligi, M., Legouverneur, G.: Exploring voice user interfaces for seniors. In: Proceedings of the 6th International Conference on Pervasive Technologies Related to Assistive Environments (PETRA 2013), pp. 52:1–52:2 (2013). https://doi.org/10.1145/2504335.2504391

Schwind, V., Deierlein, N., Poguntke, R., Henze, N.: Understanding the social acceptability of mobile devices using the stereotype content model. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI 2019), Article 361, 12 p. ACM, New York (2019). https://doi.org/10.1145/3290605.3300591

Sciarretta, E.: Libri digitali per tutti - Inclusione sociale tramite gli eBook. Eurilink University Press, Roma (2020) ISBN 979 12 80164 04 9

Sciuto, A., Saini, A., Forlizzi, J., Hong, J.I.: Hey Alexa, what's up? A mixed-methods studies of in-home conversational agent usage. In: Proceedings of the 2018 on Designing Interactive Systems Conference, pp. 857–868. ACM (2018)

Smith, A.L., Chaparro, B.S.: Smartphone text input method performance, usability, and preference with younger and older adults. Hum. Factors **57**(6), 1015–1028 (2015)

Spallazzo, D., Sciannamè, M., Ceconello, M.: The domestic shape of AI: a reflection on virtual assistants. In: DeSForM19 Proceedings (2019). https://doi.org/10.21428/5395bc37.8108aa03

Terzopoulos, G., Satratzemi, M.: Voice assistants and artificial intelligence in education. In: Proceedings of the 9th Balkan Conference on Informatics (BCI 2019), Article 34, pp. 1–6. Association for Computing Machinery, New York (2019). https://doi.org/10.1145/3351556.3351588

The Economist: Now we're talking, 7th Jan 2017. http://www.economist.com/news/leaders/21713836-casting-magic-spell-it-lets-people-control-world-through-words-alone-how-voice. Accessed Feb 2021

White, R.W.: Skill discovery in virtual assistants. Commun. ACM **61**(11), 106–113 (2018). https://doi.org/10.1145/3185336

World Health Organization: Global data on visual impairments (2010). https://www.who.int/blindness/GLOBALDATAFINALforweb.pdf. Accessed Feb 2021

Yang, X., Aurisicchio, M., Baxter, W.: Understanding affective experiences with conversational agents. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI 2019), Paper 542, pp. 1–12. Association for Computing Machinery, New York (2019). https://doi.org/10.1145/3290605.3300772

Zamora, J.: I'm sorry, Dave, I'm afraid we can't do that: chatbot perception and expectations. In: Proceedings of the 5th International Conference on Human Agent Interaction, pp. 253–260. ACM (2017)