Reading Between the Guidelines: How Commercial Voice Assistant Guidelines Hinder Accessibility for Blind Users

Stacy M. Branham
University of California, Irvine
Irvine, CA
sbranham@uci.edu

Antony Rishin Mukkath Roy
University of Maryland, Baltimore County
Baltimore, MD
antonyr1@umbc.edu

ABSTRACT
Voice-Activated Personal Assistants (VAPAs)—like Apple Siri and Amazon Alexa—have rapidly become common features on mobile devices and in homes of millions of people around the world. They have proven to be particularly valuable to people with disabilities, chiefly among people with visual impairments. Yet, we still know relatively little about the fundamental metaphors and guidelines for designing voice assistants, and how they might empower and constrain visually impaired users. To address this need, we conducted a qualitative document review of VAPA design guidelines published by top commercial vendors Amazon, Google, Microsoft, Apple and Alibaba. We found that guidelines have many commonalities that surface an underlying assumption that VAPA interfaces should be modeled after human-human conversation. We draw on prior work about needs of people with visual impairments to critique this taken-for-granted human-human conversation metaphor and offer amendments to prevailing design guidelines that can make this now-pervasive platform more fully achieve its potential to become universally usable.

Author Keywords
Accessibility; Blindness; Voice Assistant; Conversation; Design Guidelines.

ACM Classification Keywords
• Human-centered computing~Accessibility

INTRODUCTION
Over the last five years, Voice-Activated Personal Assistants, or VAPAs [23]—like Apple Siri and Amazon Alexa—have been adopted by tens of millions of people [15,27]. In the US alone, there are an estimated 90.1 million active monthly users of mobile voice assistants and 45.7 million users of smart speakers [23]. There has been significant recent interest among HCI researchers regarding these technologies [4,11,20,22,23,29,31,32,38], with several studies that include perspectives of older adults [28,38,40] and people with disabilities [1,3,7,29,37]. Blind users comprise a large minority customer base of technologies like Amazon Echo, despite the presence of accessibility and usability challenges for this population [1,29]. This prior work suggests that voice assistants may hold particular promise for the blind community, if properly designed.

Figure 1. Our study investigated commercial design guidelines of five Voice-Activated Personal Assistants. We found that guidelines promote a human-human conversation model, which may lead to accessibility challenges for blind users.

With the rapid proliferation of VAPAs, HCI researchers are struggling to stay ahead of the curve on design principles for voice interfaces. Just last year, Murad et al. [25] published early work identifying design guidelines for hands-free speech interaction, contending that no prior design guidelines exist. While this may be true within the research context, commercial VAPA vendors—led by Amazon, Google, Apple, Microsoft, and Alibaba—have been publishing their own public guidelines for years. There has yet to be a systematic analysis of how VAPA designers conceptualize the ideal system—its goals, user needs, and interaction metaphors. Commercial VAPA guidelines stand to offer a unique window into how the developed systems operate and may allow us to pinpoint fundamental sources of documented accessibility and usability issues [1,29].

In this paper, we address a research question first raised in a poster paper presented at iConference 2019 [24]. We ask: what commercial VAPA guidelines exist, and how might they constrain potential uses and users of such systems? The present work extends the preliminary document review of Google and Amazon presented in the initial study, and

ASSETS '19, October 28–30, 2019, Pittsburgh, PA, USA

ACM 978-1-4503-6676-2/19/10…$15.00
https://doi.org/10.1145/3308561.3353797

includes additional guidelines from Apple, Microsoft, and Alibaba. Our analysis indicates that VAPA guidelines assume human-human conversation as the core interaction model. We argue this choice limits usage for a variety of users, including people who are blind.

RELATED WORK

As VAPAs have rapidly become pervasive in the home (as smart speakers) and on-the-go (as mobile apps), researchers have started to examine their feasibility for various populations. These include infrequent and frequent users [10,22], native and non-native English-speaking users [32], low-income and middle-income users [4], as well as older adults and people with disabilities (e.g., [1,3,10,29,37]). Our study focuses on VAPA feasibility for people who are blind.

Accessible Technologies for People Who Are Blind

Technological interfaces for blind accessibility tend to convert visual elements into a modality which is accessible, such as speech and tactile feedback [30]. For example, screen reader software converts graphical information from a visual display into synthesized speech or Braille output [18]. Some accessible and assistive applications can even provide real-time information about one's environment. For example, mobile navigation aids can provide orientation assistance (e.g., [39]) and information related to visual surroundings (e.g., [5]). The present study explores the potential for mainstream VAPA technologies, which provide real-time interaction through voice, to support blind users.

VAPAs

Trust and Privacy

Trust and privacy are identified across user groups as being a pervasive concern affecting adoption of VAPAs in daily life [1,4,11,19,22]. Speech recognition errors are known to degrade user trust [1]. While these errors are the most frequent obstacle, prior research shows this can be mitigated by implementing a "read" back feature, which helps users create an accurate mental model of the error and how to address it [26].

Many researchers and scholars have also identified emerging privacy concerns among VAPA users. Prior work finds that users are unaware of the degree to which voice interaction data being collected is kept confidential, which can make them reluctant to share sensitive data with VAPAs [11]. Another factor known to affect privacy is the use of VAPAs in public. Multiple studies [1,4,22] find that people are less confident about using voice-based technologies in public due to concerns about disclosing sensitive information.

Personification

Personification can be defined as the attribution of "human-like properties, characteristics, or mental states to real or imagined nonhuman agents and objects" [30]. VAPAs' ability to "speak" can lead users to personify them [20], although studies diverge in their assessment of how pervasive this practice is. Purington et al. [30] report that nearly 50% of users personify VAPAs, while Lopatovska et al. [20] report a lower rate of 10%. Both studies [20,30] conclude that users from multi-person households are more likely to personify than those from single-person households. Users who engage with applications which involve social interactions are also prone to personify VAPAs [30]. Users who personify tend to report a higher rate of satisfaction [30].

Discoverability

Despite the widespread adoption of VAPAs, usage tends to be low in terms of time and frequency [22]. The voice-only modality of VAPAs makes it difficult for users to discover their capabilities and limitations [9,10,22]. Low discoverability can compromise learnability, or the ability of a novice user to easily learn how to use the new system without training [10,17]. Researchers have questioned the relationship between increasing human-like qualities of VAPAs and improving discoverability, suggesting that these design goals may be in tension [11,22]. Even though various human-like traits act as an engagement mechanism, they also contribute to setting unrealistic expectations, resulting in a mismatch between the system's capabilities and the user's model of how the system works [11,22].

Hands-free Interaction

The potential for hands-free interaction is one of the most important benefits of VAPA technologies, as it supports users who may have hands and eyes tied up in a primary activity. This supports multi-tasking, enabling users to save time and improving convenience [22]. This same capability also makes the platform promising for people with various disabilities [1,3,10,29]. However, there have been reports that the occasional need to perform a touch interaction during a task sequence can interrupt the hands-free experience [11,22]. Such an interruption is seen to be "prominent in obstructing potential frequent use" [11]. Researchers conclude that voice should be kept as the primary input and output modality throughout the interaction [11].

VAPAs and Disability

Preference for Voice Input

Findings from studies conducted with individuals with disabilities have revealed a high preference for using voice as a primary interaction modality. Studies show that VAPAs enable independent interaction with a mobile device for people with limited hand dexterity [10,27]. Balasuriya et al. [3] found that 72% of users with intellectual disabilities preferred using voice rather than using their hands for input. Both individuals with visual impairments and older adults consider voice entry a quick and easy mode of interaction [10,22,29,39,40]. Older adults also find the voice to be an "efficient" and "engaging" input modality—particularly if they have age-related physical restrictions [35].

Challenges with Speech Recognition

Even though multiple studies have identified speech recognition to be a major challenge, it is often disregarded as a technological limitation which will improve over time [4]. One population that is disproportionately affected is older adults. Research suggests that older adults have a slower speech rate with "longer inter-syllabic pauses and lower speech intelligibility" [35], which can impact the accuracy of speech recognition. Indeed, VAPAs are optimized to comprehend young adult and middle-aged voices [3].

Challenges with Duration of Sessions

Upon activation, there is a specific time period during which the device is in listening mode, waiting for the user to speak. The user will have to convey her input within that given time, or else the session will timeout. Researchers who studied populations with speech impairments, like Alzheimer's disease [35] and intellectual disabilities [3], have shown that the current timeout period for VAPAs is insufficient, as users with these conditions may take longer to formulate and speak their command. Lack of feedback indicating the VAPA's state of listening makes it more difficult for use by those with intellectual disabilities [3]. Blind users also found this time-out frustrating and limiting their capability of performing complex voice commands [1].

Challenges and Preferences of People Who Are Blind

Most of the challenges and preferences mentioned above were common across various disabled populations. However, additional unique challenges and preferences have been documented for the blind community. Blind users who are expert users of screen readers find the human-like conversational nature of VAPAs to be "verbose" and "irrelevant" [1]. They prefer to customize speech, rate, clarity, and intensity of the voice output according to the task at hand [1]. Blind users particularly desire the ability to perform complex tasks with finer control than is currently possible through commercial voice interfaces [1,37]. For example, studies report that, unlike other groups, long messages and complex commands are often used by blind users [1,2]. It can be challenging for blind users to recover from dictation errors made in such long messages, as fixing an error requires dictating the entire message again [1]. The voice-only nature of VAPAs is crucial for blind users; while sighted users may switch to a visual interface for in-depth tasks or error correction, blind users do not always have accessible alternatives [29].

VAPA Guidelines

Researchers have predominantly focused on examining real-world VAPA-based interactions in a range of settings [1, 22,29]. However, research has yet to focus on evaluating the rationale behind VAPA design, and specifically what commercial VAPA design guidelines recommend to developers. Murad et al. [25] have explored the literature on technologies that support "speech interactions." Their work evaluated GUI (graphical user interface) guidelines and assessed their relevance to interfaces that support speech interactions. While their work is motivated by the claim that currently no speech interaction guidelines exist, this overlooks the public-facing design guidelines produced by companies that market popular voice interfaces (e.g., Amazon Alexa). There is a paucity of research identifying and evaluating current voice design guidelines and the usability design principles they promote.

METHODOLOGY

We conducted a document review of public-facing design guidelines published by commercial VAPA vendors Google, Amazon, Microsoft, Apple, and Alibaba. The study was conducted in two phases: (1) identifying and collecting VAPA design guidelines, and (2) conducting a thematic analysis of the guidelines. We describe each phase below.

Phase 1: Collecting VAPA Design Guidelines

The first step to collecting VAPA design guidelines was to develop a set of keywords for conducting an online search. We chose to search for guidelines produced by recognized industry leaders on mobile voice, as well as smart speaker platforms. According to the market surveys released by voicebot.ai[1] in May 2018, there are five mobile voice and four smart speaker industry leaders, for a total of seven unique AIs: Siri, Google Assistant, Alexa, Bixby, Cortana, Aligenie, and Xiao AI.

To systematically search for public-facing voice-design guidelines for each of these AIs, we brainstormed two lists of keywords to search in combination: (1) the AI name, and (2) design guideline keywords (see Table 1). We used Google.com to search for every combination of terms in the first and second lists. No quotation marks were included around multi-word queries, to maximize the variety of search results. For example, the first query was: Siri design guidelines. The second was: Google Assistant design guidelines. We manually checked the first five pages of results for official guidelines. Search results directing to the selected vendor's voice design guideline page were then identified. Multiple search results referring to the same guidelines page were removed. Using this method, we were able to successfully identify design guideline documents for 5 of the 7 vendors: Google[2] (Go), Amazon[3] (Am), Microsoft (Mi)[4], Apple[5] (Ap), and Alibaba[6] (Al).

AI Keywords	Guideline keywords
Siri, Google Assistant, Alexa, Bixby, Cortana, Aligenie, Xiao AI	Design Guidelines, Voice Design Guidelines

Table 1. Keywords used in search for VAPA Guidelines

Given the dynamic nature of online content, we collected static copies of the guidelines for offline viewing. We

[1] https://voicebot.ai/voice-assistant-consumer-adoption-report-2018/
[2] https://developers.google.com/actions/design/
[3] https://developer.amazon.com/designing-for-voice/
[4] https://docs.microsoft.com/en-us/cortana/skills/
[5] https://developer.apple.com/design/human-interfaceguidelines/carplay/interaction/voice/
[6] http://doc-bot.tmall.com/docs/doc.htm?spm=0.0.0.0.LP3ZNz&treeId=393&articleId=106992&docType=1

exported each webpage to letter-sized PDF documents (e.g., "Go01.pdf" stored Google's primary webpage) as well as copies of their directly linked pages (e.g., "Go02.pdf" all the way up to "Go38.pdf"). Altogether, we collected 271 pages (8.5"x11" PDF format) from Google, 83 pages from Amazon, 42 pages from Microsoft, 15 pages from Apple, and 19 pages from Alibaba in September 2018.

Phase 2: Conducting Thematic Analysis of Guidelines

To address our research question, we chose to follow an inductive thematic analysis process [8]. As a first step, the second author read through all the identified design guidelines in their entirety multiple times, taking notes. The second author then conducted segment-by-segment open coding, with each segment ranging from a single to multiple sentences. A single instance consisted of a label, or code, and the segment it was referring to. Codes were iteratively refined with each successive guideline. Both authors discussed codes weekly in meetings over a period of several months. During meetings, related codes were merged or sometimes teased apart into multiple codes, and they were reworded to add nuance. Axial codes were introduced to organize related open codes. This process resulted in 190 instances, 18 open codes, and five axial codes. A summary of axial codes and total instances is provided in Table 2.

Axial Codes	# Instances
Make conversations human	39
Make conversations personal	50
Make conversations efficient	68
Make conversations relational	13
Give user a sense of control	20

Table 2. Axial codes with their respective instance count

For the purposes of reporting, each heading under the findings section corresponds to an axial code. Each subheading corresponds to an open code. When reporting open codes, we will identify the number of instances in parentheses. When quoting or providing examples from guidelines, we will identify vendors by their abbreviations.

FINDINGS

Make Conversations Human

Guidelines tended to advocate for a communication model that closely resembles that of face-to-face, human-human dialogue (39 instances, Go/Am/Mi/Ap/Al). We have organized guidelines under this theme into four subthemes. We discuss each in turn below.

Model Conversation After Human Speech

Throughout the design guideline documents, there are many instances which advocate for designing VAPA interactions based on human conversation (16 instances). For example, one guideline recommended creating a human-human dialog script and testing it by conducting roleplay with real humans. Another guideline justifies this approach by explaining that human conversation patterns have evolved over thousands of years and therefore are more "familiar" and "frictionless" for users:

> *"Computers should adapt to the communication system users learned first and know best. This helps create an intuitive and frictionless experience." (Go)*

The intention behind this approach is to leverage natural communication systems of users, but not to "trick" them into believing the VAPA is a human. Although Google guidelines suggest the agent "could also be an anthropomorphized animal, an alien, an artificial intelligence, a cartoon character, etc.," all guidelines universally modeled the interaction on human-human dialogue.

Aim for Natural Conversation

Many guidelines emphasize maintaining a "natural" conversation with users (10 instances). Naturalness of conversation was predicated on the use of "everyday language" (as opposed to jargon), and replicating the way people "naturally ask each other for things." Use of colloquial language is encouraged, and natural-sounding conversation is favored over grammatical accuracy.

> *"Don't emphasize grammatical accuracy over sounding natural. For example, ear-friendly verbal shortcuts like 'wanna' or 'gotta' are fine for text-to-speech (TTS)." (Mi)*

Enact a Persona

Guidelines also encourage the VAPA to showcase unique personality while interacting with users (13 instances). While one guideline refers to a "system persona" as the distinct tone of the conversational agent who interacts with the user, another guideline refers to this as the "image" of the voice agent. Guidelines suggest that personality traits should be kept consistent within each persona; third-party app developers should create their own personas rather than impersonating or mimicking the system persona. To achieve this, guidelines identify various strategies. For example, personas can be developed based on familiar roles enacted by real-life service providers:

> *"In a banking application, the persona could be modeled after an idealized bank teller—trustworthy with customers' money and personal information. The metaphor of the bank teller makes this new experience feel familiar, since users' real-world banking knowledge can guide them." (Go)*

As part of this process, developers should create a description of the character, including a list of adjectives and personal traits, to help the persona mirror real human agents. This guideline document also emphasizes personalizing the voice of the persona, as it states that humans tend to infer personal traits like age, sex, social status, and intelligence from voice alone.

Personas also need to match voice quality to each "skill" or "action"—terms used by Microsoft, Amazon, Alibaba, and Google to refer to a "set of actions or tasks that are accomplished" by a VAPA:

> *"For example, if the voice skill to be designed is to tell users a joke, the voice image you set may be young and humorous; while for the technique of reading daily news, you need to use a more mature and stable voice image." (Al)*

To achieve this, most of the platforms provide a variety of synthesized text to speech (TTS) voice libraries to choose from. This voice can be further customized with Speech Synthesis Markup Language (SSML) to achieve more personalization. There is also an option to record and use a real professional human voice.

Make Conversations Personal

Making conversation personal is about designing VAPAs so that they exercise agency and tailor interactions to the user (50 instances, Go/Am/Mi/Ap/Al). Rather than waiting for users to explicate their preferences, a VAPA should set smart defaults, understand and respond with alternative phrasings, and converse cooperatively with their users. Guidelines advocate for these qualities as beneficial traits of human conversational partners.

Set Smart Defaults

Design guidelines recommend adopting "intelligent" and "smart" default values in the absence of a user's explicit specification (8 instances):

> *"Use default values when the user is not specific. For example, if the user says, 'Make my room warmer,' Cortana should say, 'I've raised your room temperature to 72 degrees' instead of 'Sure, what temperature?'" (Mi)*

What constitutes a relevant default value varies according to the context of use. In the example above, a smart default is about making relative adjustments. A smart default may also be derived from previous interactions, as noted by Apple's guidelines:

> *"Whenever possible, use intelligent defaults rather than asking for input. For example, a ride sharing app might automatically default to the last requested ride type, or a fitness app might default to a favorite workout." (Ap)*

During critical scenarios, like making decisions which have "financial impact," guidelines advise VAPAs to default to the safest and cheapest options.

While these recommendations all emphasize agency of the AI—making decisions on behalf of the user—they advocate against deceiving users through misrepresenting or hiding information using defaults.

Intelligently Interpret Users' Varied Request Phrasing

Guidelines promote the ability for users to express their "intent"—a term used by Amazon, Google, Microsoft, and Alibaba referring to a "representation of the action that fulfills a customer's spoken request"—in many different ways, thereby making VAPAs accountable to handle all of them effectively (11 instances). For example:

> *"Avoid assuming that people will say precisely the words that you anticipate for an intent. While the user might say 'plan a trip,' he or she could just as easily say 'plan a vacation to Hawaii.'" (Am)*

Developers are cautioned by one guideline to refrain from asking users a question the system is not prepared to handle the answer to. In terms of preparations, guidelines suggest defining a range of utterances people may say to the VAPA. They suggest a standard benchmark to be around 30 alternate utterances per intent. The guidelines also state that VAPAs should be able to handle even subtle variations and mispronunciations in the user's response.

Intelligently Respond to Users with Varied Phrasing

In addition to supporting varied request phrasing, guidelines recommend maintaining variation in the VAPA's responses (7 instances). Guidelines state that varying the phrasing of responses keeps conversation more "natural" and prevents it from getting "stale." Another guideline mentions that variety reduces the "robotic" feel of conversation by adding personality. The most common recommendation among guidelines to achieve variety is by adding variety in prompts that repeatedly occur during the conversation. For example, the system might randomize acknowledgements by drawing from alternatives like 'okay', 'alright', 'got it', etc.

Intelligently Respond to Requests of Varied Completeness

Guidelines state that VAPAs should be able to handle requests not only of various phrasing, but also of various completeness. Completeness can range from no intent to multiple intents at once (12 instances).

Sometimes the user will provide all necessary information for a service request in a single utterance; Amazon and Microsoft call this a "full intent." The VAPA should be able to identify all the key attributes from the utterance and respond meaningfully to the user request. Sometimes, the user request will be incomplete; Amazon and Microsoft call this a "partial intent." In this case, the VAPA should be able to detect the missing element and provide an immediate follow-up prompt towards fulfilling the service request. Guidelines also refer to scenarios where first-time users may provide ambiguous requests; Microsoft, Amazon, and Alibaba call this a "no intent." In this case, VAPAs should not simply dismiss the user's request or ask them to restate the request correctly; they should walk users through the interaction by providing a list of three options from which to choose. The following two examples demonstrate partial and no intent cases, respectively:

> *"User: Hey Cortana, ask Mileage Wizard if I have miles.*
> *Mileage Wizard: Miles to travel?" (Mi)*

--

"User: Hey Cortana, ask Mileage Wizard. Mileage Wizard: Do you want available miles, used miles, or discounts?" (Mi)

When the VAPA asks for more information to complete a no intent or partial intent request, there are situations where the user will not provide any response. The VAPA should handle such situations by re-prompting with a slight rewording. On the other hand, sometimes users "over-answer" by responding with more information than required. VAPAs should be able to handle such situations effectively.

Use Contextual Data to Cooperate with the Individual

There are multiple segments of the guidelines regarding the importance of leveraging context to communicate in a more cooperative manner with individuals (12 instances).

Due to the ambiguous nature of language, a user's utterance cannot be understood in isolation, i.e. without the context of prior utterances and conversations. Neither can the VAPA's response be made relevant. For example, in order to use pronouns, some contextual understanding of the referent is needed; this entails tracking conversation across conversational turns and maintaining gendered representations of people who are mentioned. Context can extend beyond the bounds of the current invocation. One guideline says that people appreciate when the VAPA remembers the past conversations, especially if there is "static information and frequent actions" that will likely recur. Additionally, as the user engages a specific app or skill over time, a greeting should become less verbose and more familiar. Context may also be defined according to the users' current state of action, because certain user utterances (e.g., "help") can be applied to more than one type of intent within the same application. Guidelines warn that context changes even when it appears otherwise; it is particularly important to recognize this in error scenarios to reduce frustration:

"Good error handling is context-specific. Even though you're asking for the same information, the conversational context is different on the second or third attempt." (Go)

Make Conversations Efficient

Of significant concern across all guidelines was that developers strive to make conversations efficient and easy for users to perform (68 instances, Go/Am/Mi/Ap/Al).

Reduce Time and Effort as Compared to Screens

Guidelines emphasize how the conversational medium can prove to be an easier, more efficient alternative to screen-based interaction (8 instances). Google says voice is a "shortcut" and "reduces friction" because it does not require multiple taps on the screen. There are many instances where guidelines compare screen-based and voice-based interaction to demonstrate this point:

For example, saying "Play the latest House of Cards" is much easier than opening up an app, searching for "House of Cards", finding the latest episode, and pressing play. (Mi)

Support Multi-tasking with Screen-free Interaction

Conversational interfaces are ideal for enabling users to multitask, as they support hand-free, eyes-free interaction with low cognitive load (7 instances). For example, when the user is cooking or lying down, it is convenient to interact using voice as compared to another input device that requires hands (e.g., keyboard). Or, when the user is performing a cognitively challenging primary task, like writing a document on her computer, voice can be an effective secondary interaction medium to perform parallel tasks, like replying to a message or controlling the music. Similarly, guidelines identify voice as a safe and efficient modality for navigating—walking or driving—without distraction. Developers should therefore "strive for a voice-driven experience that doesn't require touching or looking at the screen" (Ap).

Keep Conversations Concise

Guidelines generally advocate for efficient conversations by keeping conversational turns as concise as possible (14 instances). Google's guidelines note that too much content is as concerning as too little. Amazon introduces the concept of a 'one-breath test' (can the utterance be stated in one breath?) to evaluate the length of conversation. Similarly, Alibaba suggests keeping the "keyword" count below two when there are multiple rounds of conversation (e.g., in the sentence "it's going to be rainy in Shanghai," the keywords are "rainy" and Shanghai"). Google proposes a "tapering" strategy, whereby the amount of detail provided in each interaction will be reduced according to how many times the service has been used; an experienced user is expected to need less comprehensive prompts compared to a novice user. These guidelines also state that the system should be able to identify when a user's tone of voice reveals frustration or impatience, and then change to a more concise interaction mode.

Make Lists Efficient

VAPAs often format responses as lists—lists of top items for purchase, lists of nearby businesses, lists of search results, lists of music stations, and more. Guidelines offer various techniques to improve efficiency while reading lists (13 instances).

Microsoft recommends time constraints for readings lists. The VAPA should not take more than 20 seconds to read the first few items in the list. Furthermore, the duration of pauses between lists of items should be based on item length and complexity. For example, there should be a 350ms pause for a normal scenario and a 400ms pause for a complex scenario.

Microsoft, Amazon, and Alibaba all remark on the number of list items to present, generally around three items, although Amazon suggests a range of 2-5. In the latter case, to improve comprehension, Amazon suggests that items be clustered into groups of two or three, as in "Cheddar and Gouda, as well as Gorgonzola, Parmesan, and Brie."

Microsoft suggests that more options may overwhelm and frustrate the user, resulting in rejection of VAPAs. The guidelines further rationalize this choice by claiming the limited capacity of the human brain to remember instructional information:

> *"For most people, the human brain can only remember a small amount of information when listening to instructions. Limit voice interactions to only what is absolutely required. For example, present only three items of a list at a time." (Mi)*

Scaffold Turn-taking

Maintaining efficient conversation entails clean information hand-offs between the VAPA and user, in which each takes a turn without either interrupting or pausing between utterances (12 instances). To scaffold efficient turn-taking, guidelines commonly recommend strategic placement of questions in the VAPA's script:

> *"Generally, end with a question before having the user respond. The question provides a cue to begin speaking and coaches the user on what to say next. End the prompt right after the question so that people don't try to answer while Alexa is speaking." (Am)*

In the guideline above, we see that a question placed at the end of the utterance both prompts the user to begin their turn, while also discouraging interruption. In combination with questions, lists are often used to constrain and therefore expedite the user's selection of a response: "Your design needs to make sure that the user clearly understands what you are asking and that a response is expected. Just presenting the options is not sufficient" (Mi). Finally, lists should be kept short, or should be presented in multiple turns, so as not to allow the VAPA to "monopolize the conversation" (Go).

Provide Appropriate Confirmation

Both Google and Microsoft suggest that a good voice interaction should use a variety of techniques for confirmation (10 instances). They articulate two types of confirmation—implicit and explicit. Below are examples of implicit and explicit confirmation, respectively, in a scenario where the user wants to purchase two tickets:

> *"User: 2.*
>
> *Ibento (fake ticket seller): Alright here is the best open slot for 2 people sitting together. Would you like to get these?" (Go)*
>
> --
>
> *"User: 2.*
>
> *Ibento (fake ticket seller): So you want to book two slots, right?" (Go)*

Guidelines state that it is the responsibility of the developer to build in the right balance between both implicit and explicit confirmation. A conversation with repeated explicit confirmation can be slow and frustrating, whereas lack of explicit confirmation may mislead users and force them to backtrack. They also recommend using explicit confirmation when the cost of misunderstanding is high, as when booking a flight or making a financial transaction.

Limit Use of Earcons

Google's guidelines are the only that addressed the use of earcons (4 instances). Earcons, or audio icons, are a type of brief, non-verbal sound designed to convey meaning (e.g., the beep a phone makes when receiving a text message). The guidelines suggest that earcons are a mismatch for conversational interfaces, because they are not intuitive. They therefore caution against overusing earcons because they can "impose cognitive load" and "become overwhelming; "limit use to just a few sounds that are easily distinguishable so that users don't have to learn too many." In all cases, earcons should be "as brief as possible," except when used for branding purposes, as in a "welcome" sound.

Make Conversations Relational

Design guidelines encourage behaviors that are often exhibited in human-human conversations to develop lasting relationships (13 instances, Go/Am). Perhaps the most explicit demonstration of this top-level theme is Google's adoption of the linguistics theory known as the Cooperative Principle. According to Google, "the Cooperative Principle [states that] efficient communication relies on the assumption that there's an undercurrent of cooperation between conversational participants."

Maintain Friendly Conversation

Microsoft and Amazon guidelines both encourage keeping conversations friendly (3 instances). Developers should "avoid niceties… to keep the conversation friendly and informal" (Go). For example, instead of saying: "Please accept the Ibento terms of service in order to proceed," the VAPA should say "Sure, I can help with that. But first there's one thing you need to do: accept the Ibento terms of service." In addition to adopting familiar language, developers should "Avoid exclamation points, as they can be perceived as shouting."

Maintain Transparency and Take Responsibility for Errors

Guidelines from Amazon and Google suggest that VAPAs should maintain transparency in conversation with users when handling errors, and take responsibility when necessary (10 instances). The system should be "transparent, honest, and helpful" in disclosing the error to the user and offering next steps. For example, when a transaction fails, the system might respond:

> *"I can't reach your preferred florist right now to place your order. Should we wait a few minutes and try again, or order from another florist?" (Go)*

The system should only encourage the user to try again if the error is likely to be present for only a few seconds, otherwise it is advisable to notify the user that the service requested is not currently working. When explaining errors, VAPAs should “avoid using technical jargon” (Am).

Google has given nuanced consideration to when the VAPA should and should not apologize. Generally, they advise that VAPAs should “acknowledge instead of apologize” when the user’s first request was misunderstood. “Sorry” should be used sparingly and is well-suited to situations where the VAPA does not have the skillset needed to service the user’s request or “when it serves a transitional social or phatic function.” Perhaps the clearest guidance, reiterated in multiple instances, is that the VAPA should never blame the user or a third party:

> *“Our persona should take responsibility, never blame the user, and never blame another party. People think less of individuals who blame others for failure.”*

Give Users a Sense of Control

Design guidelines encouraged designers to create an experience where the users feel like they are in control, even if their options are fairly constrained (20 instances, Go/Am/Mi/Ap/Al):

> *“The magic and art of good conversation design is that users feel like they’re in control and that they can say anything at anytime, but in reality, the dialog directs them along pre-scripted paths.” (Go)*

The illusion of user control can be achieved through various strategies. One is to avoid teaching commands or giving a tutorial-like introduction to a VAPA’s capabilities:

> *“Teaching commands discourages experimentation and undermines trust. The implied message is that users have to say these exact phrases or they won’t be understood.” (Go)*

Another is to carefully construct the prompts given to users, so that possible answers are predictable and unambiguous: “For example, asking ‘What would you like?’ is too open-ended. Even something like ‘Would you like Brie or Gouda?’ opens up a likely response of ‘Yes.’” (Am). Prompts should also be action-based, rather than prompt-based:

> *“‘Do you want to keep shopping or [sic] you ready to check out?’ (action based).*
>
> *‘Now you can say ‘keep shopping’ or ‘Check Out.’’ (prompt based)” (Go)*

Finally, always allow the user to make course corrections (“no, I said…,” etc.) or discontinue a task immediately when the user gives commands like “exit", "cancel", "stop", "nevermind", and "goodbye." One caveat to this advice is that the system should request confirmation if exiting would result in significant loss of progress.

DISCUSSION

Numerous recent research efforts have centered on the accessibility constraints and potential of VAPAs to serve older adults [28,38,40] and people with disabilities [1,3,7,29,37]. These studies primarily observe product reviews, user testimony, or *in situ* use of VAPAs. In comparison, our study entailed a document review of the public-facing guidelines published by leading VAPA vendors. Murad et al. [25] published early work identifying design guidelines for hands-free speech interaction, targeting adoption within the HCI research community. By drawing from commercial VAPA guidelines, we believe our analysis offers a unique window into the fundamental concepts that dictate how VAPA systems—which are used by tens of millions of people around the world [15,27]—operate. To our knowledge, this study and our report of early findings [6] are the first to explore voice design guidelines that are published by industry leaders in the VAPA domain.

Our findings reveal that the interaction model of VAPAs is rooted in a human-human conversational metaphor. This metaphor was explicitly enacted by drawing on formal human linguistic models, like the Cooperative Principle, and promoting design strategies that personify the VAPA—like roleplay and persona development. VAPAs are portrayed as human entities with names (e.g., “Siri”) and designers are encouraged to assign characteristics like age, sex, social status, intelligence, etc. to their creations. Guidelines recommend that VAPAs be made to speak using everyday language, complete with slang, contractions, naturalistic pauses, and turn-by-turn cooperative dialogue. Designers are cautioned against making the conversation “robotic,” to avoid formulaic command constraints, and to minimize use of non-voice interaction like earcons. Notably absent in guidelines was any specific advice about how to make the product accessible to people with disabilities. In the sections below, we explore some of the nuances of design guidelines based on this human-human dialogue model that may be particularly detrimental to blind users’ accessibility and usability needs.

Ideal Human Conversation Length and Complexity?

Length—the number of words—and complexity—the number of intents—directly impact the efficiency of communication with VAPAs. Our findings show that VAPAs are designed to keep the conversational turns *short* in length and *low* in complexity to make the voice interaction more “cooperative” from the user’s perspective. In the next two paragraphs, we will argue that guidelines’ suggestions about length and complexity do not adequately describe blind users’ cognitive capabilities.

First, regarding *length*, the guidelines articulate concrete limits to the responses given by the VAPA—limits which may be overly universalizing. For example, various guidelines recommend a maximum of two keywords per turn, three items per list, and the “one-breath test.” Guidelines justify these limits by arguing that longer responses will be difficult to follow, as this will overwhelm the human brain, and it may result in “frustration and impatience” for users. In other words, increasing length might result in what HCI researchers call “cognitive overload" [33]. However, studies have shown that many blind people are remarkably superior to sighted peers in performing serial memory tasks and are better able to recall longer word sequences [33]. It is reasonable to infer that blind users of VAPAs may be able to interpret and recall prolonged voice responses with more keywords and list items, while using minimal turns.

Second, regarding *complexity*, VAPA guidelines recommend being able to handle complex commands from the user, which include all necessary information for a service request in a single utterance. Blind users, indeed, prefer to interact with VAPAs using such full-intent utterances (i.e., high complexity) [1]. Yet, studies have shown that VAPA users who are blind or who have intellectual disabilities often run out of time to articulate such complex instructions within the limited session period [1,3]. Even though guidelines clearly state that full-intent utterances should be handled, the inflexible timeout feature conflicts with this when users have certain disabilities.

Ideal Human Conversation Speed?

As with conversation length and complexity, VAPA guidelines have made numerous recommendations related to the appropriate speed and pacing of conversational interaction. Our findings revealed that the first few list items should be spoken for a maximum of 20 seconds, with pause durations ranging from 350ms to 400ms. Guidelines additionally emphasize using "natural" and "intuitive" speech, avoiding robotic structures, and minimizing earcons, because these may lead to cognitive overload. Interestingly, these guidelines conflict significantly with the voice output design of screen readers for people who are blind, which require expert interaction techniques, use robotic-sounding synthesized voice, and speak output at a rate far faster than a human could. Studies conducted by various researchers have shown that many blind people can easily comprehend speech at higher listening rates than sighted people [6,12]. While sighted people can comprehend speech at a maximum of 10 syllables per second, some blind people can comprehend up to 25 syllables per second [14]. Not surprisingly, experienced screen reader users find the existing speech rates of VAPAs to be frustratingly slow [1]. We add that, for this user group, VAPA interactions do not live up to the goal shared across design guidelines: "Make Conversations Efficient."

Putting Disability Back into the Ideal Human

When we read between the lines of commercial VAPA design guidelines, we see that the idealized "human" behind the recommended "human-human conversation" model is overly constrained in ways that do not include capabilities and needs of many members of the blind population. Commercial, technical notions of what constitutes conversation that is too long, too complex, or too fast do not conform to scientists' or blind people's own understandings of blind peoples' capabilities—a fact that effectively displaces people who are blind from the definition of human-human conversation being invoked. This shortcoming is particularly acute, as prior research has found that many blind people stand to benefit significantly from mainstream voice platforms that prove accessible and usable [1]. In addition, because many blind people already use VAPAs as a primary platform to perform daily tasks [1], it stands to reason that they may have sufficient motivation to learn and leverage longer, more complex commands and to deploy them at top speed.

In addition, our study suggests that the roots of usability and accessibility challenges for other populations—like older adults, children, people with Alzheimer's, and people with intellectual disabilities—may be traced back to VAPA design guidelines. Prior work finds that speech recognition often breaks down for children and older adults [3,21] as well as people who are deaf or hard-of-hearing [16], because there is wide variation in pitch, pronunciation, patterns of stress, and intersyllabic pauses, leading to higher recognition error rates [16,21,35]. Prior work also finds that the timeout period for speech input is often inadequate for people with Alzheimer's [34] and intellectual disabilities [3]. Like people who are blind, these groups would benefit from VAPAs that adopt more inclusive models of conversation length, complexity, speed, and other aspects of interaction.

Our recommendations for commercial VAPA guideline authors, as well as researchers and developers of these systems, are therefore: (1) adopt more inclusive ranges of human capabilities and preferences that do not exclude people with disabilities, (2) in addition to creating a robust accessibility section for each guideline document, interleave accessibility recommendations throughout, (3) explore how not only people with disabilities, but also people in a variety of unanticipated situations may benefit from increased customizability of VAPA personas, as well as broader definitions of what it is to be "human." We present three potential design implications in line with our findings in the following section.

IMPLICATIONS FOR DESIGN

Allow Preferences to Be Defined On-the-fly

Our study found that guidelines placed restrictive limits the to the *length*, or number of words, used in VAPA-human conversational turns. Above, we draw out the ways in which longer turns may be possible for blind users, while previous studies suggest that more succinct turns may be desirable for blind users [1]. Therefore, we recommend that the length of conversational turns be customizable according to user preference. We can imagine scenarios in which it may be most convenient to make in-line requests for more or less verbose interaction. For example, querying Alexa for the weather in the first author's current location results in a response that is over 20 seconds long. A user who desires a less conversational, less lengthy exchange may desire a "brief" weather update:

User: Alexa, give a brief weather update.

Alexa: 47 degrees, partly sunny. Low, 42. High, 57.

Features like this may support blind and sighted users alike, however this functionality is not currently described in guidelines on any of the platforms observed in this study.

Allow Preferences to be Defined in Advance

We found that guidelines encouraged modeling VAPA communications after human-human dialogue, including a naturalistic voice *speed*. However, prior work shows that blind people can comprehend speech at higher listening rates [6,12], and that blind VAPA users find speech rates frustratingly slow [1]. Currently, none of the guidelines mention the necessity of supporting adjustable speech rates. Speech rates may be a basic preference that cross-cuts task type, in the way that a screen reader's speech rate is stable across applications. We can imagine a feature that supports a universal, customizable speech rate:

> *User: Alexa, increase voice speed.*
>
> *Alexa: <in adjusted speed> Voice speed increased to 85%.*

Beyond the blind population, other populations may stand to benefit from lowering the speech rate, including people with intellectual disabilities, children, and non-native speakers.

Allow Custom Voice Commands to be Defined in Advance

Guidelines tended to discourage support for high *complexity* of commands in favor of more conversational language without need for command memorization. However, prior work suggests that blind users may prefer to accomplish more complex tasks with their VAPAs, like setting recurring calendar invites or authoring long text messages [1]. The fact that VAPAs may be the most or even the only accessible way to use some services [1] may be sufficient motivation for blind users to memorize or define more complex commands. We can imagine, for example, a user defining a shorter trigger phrase for an often-used command:

> *User: Alexa, define new command. 'Weather, brief' equals 'give a brief weather update.'*
>
> *Alexa: OK. 'Weather, brief' equals 'give a brief weather update.'*

Note that, as of the time of publication, none of the platforms observed in this study support adjusting the brevity of weather reports. However, some VAPAs, like Apple Siri, are beginning to make this type of user-defined command possible through features like Siri Shortcuts[7]. We believe the ability to define these "voice macros" will be particularly useful to blind users with regards to increasing complexity or reducing length. However, we can also imagine using such a feature to define more memorable or pronounceable commands for children or people with intellectual disabilities, cognitive impairments, or speech impediments.

We have proposed three different recommendations that might be added to VAPA guidelines to improve inclusion of people who are blind. It is important to note that these have not been user tested. Additionally, while we have presented each guideline with respect to a specific issue raised by our study (length, speed, or complexity), we do not intend to insinuate that any one solution be limited to addressing any one problem. Rather, we hope these voice interaction vignettes generate conversation around how VAPA designers can do better; departures from the human-human conversational paradigm and expansion of the parameters used to define normative human needs may open up novel access for blind users and a wide range of other beneficiaries.

CONCLUSION AND FUTURE WORK

In a document analysis study of commercial VAPA guidelines from five leading vendors, we discovered that the fundamental metaphor for interaction is human-human conversation. We articulate several ways in which guidelines define what it means to be human in conversation—in the form of limits placed on length of utterances, avoidance of complexity in speech, and narrow pacing restrictions. Further we explore how these constraints exclude people with disabilities, with a focus on people who are blind. We propose three ways in which VAPA design guideline authors might improve guidelines to be inclusive, and we begin to address these by offering three implications for design that support user customization. Future work will include engaging VAPA designers and guideline authors in an effort to co-develop more inclusive guidelines.

ACNKNOWLEDGEMENTS

Thanks to Ravi Kuber and William Easley for providing feedback on a draft of this paper. This work was supported by a grant by the Maryland Industrial Partnerships (MIPS) Program and TRX Systems, Inc.

REFERENCES

1. Ali Abdolrahmani, Ravi Kuber, and Stacy M. Branham. 2018. "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 '18)*, 249–258. https://doi.org/10.1145/3234695.3236344
2. Shiri Azenkot and Nicole B. Lee. 2013. 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 '13).* ACM, New York, NY, USA, Article 11, 8 pages. http://dx.doi.org/10.1145/2513383.2513440
3. Saminda Sundeepa Balasuriya, Laurianne Sitbon, Andrew A. Bayor, Maria Hoogstrate, and Margot

[7] https://support.apple.com/en-us/HT209055

Brereton. 2018. Use of Voice Activated Interfaces by People with Intellectual Disability. *In Proceedings of the 30th Australian Conference on Computer-Human Interaction (OzCHI '18)*, 102–112. https://doi.org/10.1145/3292147.3292161

4. Apoorva Bhalla. 2018. An Exploratory Study Understanding the Appropriated Use of Voice-based Search and Assistants. *In Proceedings of the 9th Indian Conference on Human Computer Interaction (IndiaHCI'18)*, 90–94. https://doi.org/10.1145/3297121.3297136
5. Jeffrey P. Bigham, Chandrika Jayant, Hanjie Ji, Greg Little, Andrew Miller, Robert C. Miller, Robin Miller, Aubrey Tatarowicz, Brandyn White, Samual White, and Tom Yeh. 2010. VizWiz: Nearly Real-time Answers to Visual Questions. *In Proceedings of the 23Nd Annual ACM Symposium on User Interface Software and Technology (UIST '10),* 333–342. https://doi.org/10.1145/1866029.1866080
6. Danielle Bragg, Cynthia Bennett, Katharina Reinecke, and Richard Ladner. 2018. A Large Inclusive Study of Human Listening Rates. *In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*, 444:1–444:12. https://doi.org/10.1145/3173574.3174018
7. Robin N. Brewer, Leah Findlater, Joseph "Jofish" Kaye, Walter Lasecki, Cosmin Munteanu, and Astrid Weber. 2018. Accessible Voice Interfaces. *In Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '18)*, 441–446. https://doi.org/10.1145/3272973.3273006
8. Kathy Charmaz. 2006.Constructing Grounded Theory: A Practical Guide through Qualitative Analysis. SAGE.
9. Mei-Ling Chen and Hao-Chuan Wang. 2018. How Personal Experience and Technical Knowledge Affect Using Conversational Agents. *In Proceedings of the 23rd International Conference on Intelligent User Interfaces Companion (IUI '18 Companion)*, 53:1–53:2. https://doi.org/10.1145/3180308.3180362
10. Eric Corbett and Astrid Weber. 2016. 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 '16)*, 72–82. https://doi.org/10.1145/2935334.2935386
11. Benjamin R. Cowan, Nadia Pantidi, David Coyle, Kellie Morrissey, Peter Clarke, Sara Al-Shehri, David Earley, and Natasha Bandeira. 2017. "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 '17),* 43:1–43:12. https://doi.org/10.1145/3098279.3098539
12. Susanne Dietrich, Ingo Hertrich, and Hermann Ackermann. 2013. Ultra-fast speech comprehension in blind subjects engages primary visual cortex, fusiform gyrus, and pulvinar - a functional magnetic resonance imaging (fMRI) study. *BMC Neuroscience* 14: 74. https://doi.org/10.1186/1471-2202-14-74
13. Friederike Eyssel, Dieta Kuchenbrandt, Simon Bobinger, Laura de Ruiter, and Frank Hegel. 2012. "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 (HRI '12)*, 125–126. https://doi.org/10.1145/2157689.2157717
14. R. Douglas Fields. Why Can Some Blind People Process Speech Far Faster Than Sighted Persons? Scientific American. Retrieved April 30, 2019 from https://www.scientificamerican.com/article/why-can-some-blind-people-process/
15. Prachi Bhardwaj Gal Shayanne. Siri owns 46% of the mobile voice assistant market — one and half times Google Assistant's share of the market. Business Insider. Retrieved April 30, 2019 from https://www.businessinsider.com/siri-google-assistant-voice-market-share-charts-2018-6
16. Abraham T. Glasser, Kesavan R. Kushalnagar, and Raja S. Kushalnagar. 2017. Feasibility of Using Automatic Speech Recognition with Voices of Deaf and Hard-of-Hearing Individuals. *In Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '17).* ACM, New York, NY, USA, 373-374. https://doi.org/10.1145/3132525.3134819
17. Tovi Grossman, George Fitzmaurice, and Ramtin Attar. 2009. A Survey of Software Learnability: Metrics, Methodologies and Guidelines. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09),* 649–658. https://doi.org/10.1145/1518701.1518803
18. Julie A. Jacko, V. Kathlene Leonard, Molly McClellan, and Ingrid U. Scott. 2008. Perceptual Impairments: New Advancements Promoting Technological Access. *In A. Sears & J. A. Jacko (Eds.), The human-computer interaction handbook: Fundamentals, evolving technologies and emerging applications,* 853–870, Taylor & Francis Group: New York, 2008.
19. Yuting Liao, Jessica Vitak, Priya Kumar, Michael Zimmer, and Katherine Kritikos. 2019. Understanding the Role of Privacy and Trust in Intelligent Personal Assistant Adoption. *In Information in Contemporary*

Society, Natalie Greene Taylor, Caitlin Christian-Lamb, Michelle H. Martin and Bonnie Nardi (eds.). Springer International Publishing, Cham, 102–113. https://doi.org/10.1007/978-3-030-15742-5_9

20. Irene Lopatovska and Harriet Williams. 2018. Personification of the Amazon Alexa: BFF or a Mindless Companion. *In Proceedings of the 2018 Conference on Human Information Interaction & Retrieval (CHIIR '18)*, 265–268. https://doi.org/10.1145/3176349.3176868
21. Silvia Lovato and Anne Marie Piper. 2015. "Siri, is this you?": Understanding young children's interactions with voice input systems. *In Proceedings of the 14th International Conference on Interaction Design and Children (IDC '15)*. ACM, New York, NY, USA, 335-338. http://dx.doi.org/10.1145/2771839.2771910
22. Ewa Luger and Abigail Sellen. 2016. "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 '16)*, 5286–5297. https://doi.org/10.1145/2858036.2858288
23. Aarthi Easwara Moorthy and Kim-Phuong L. Vu. 2015. Privacy Concerns for Use of Voice Activated Personal Assistant in the Public Space. *International Journal of Human–Computer Interaction* 31, 4: 307–335. https://doi.org/10.1080/10447318.2014.986642
24. Antony Rishin Mukkath Roy, Ali Abdolrahmani, Ravi Kuber, and Stacy M. Branham. 2019. Beyond being human: The (in)accessibility consequences of modeling VAPAs after human-human conversation. *In Proceedings of iConference (iConference '19)*. https://doi.org/10.21900/iconf.2019.103342
25. Christine Murad, Cosmin Munteanu, Leigh Clark, and Benjamin R. Cowan. 2018. Design Guidelines for Hands-free Speech Interaction. *In Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct (MobileHCI '18)*, 269–276. https://doi.org/10.1145/3236112.3236149
26. Chelsea Myers, Anushay Furqan, Jessica Nebolsky, Karina Caro, and Jichen Zhu. 2018. Patterns for How Users Overcome Obstacles in Voice User Interfaces. *In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*, 6:1–6:7. https://doi.org/10.1145/3173574.3173580
27. Sarah Peres. 39 million Americans now own a smart speaker, report claims | TechCrunch. Retrieved April 30, 2019 from https://techcrunch.com/2018/01/12/39-million-americans-now-own-a-smart-speaker-report-claims/
28. François Portet, Michel Vacher, Caroline Golanski, Camille Roux, and Brigitte Meillon. 2013. Design and Evaluation of a Smart Home Voice Interface for the Elderly: Acceptability and Objection Aspects. *Personal and Ubiquitous Computing*. 17, 1: 127–144. https://doi.org/10.1007/s00779-011-0470-5
29. Alisha Pradhan, Kanika Mehta, and Leah Findlater. 2018. "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 (CHI '18)*, 459:1–459:13. https://doi.org/10.1145/3173574.3174033
30. Amanda Purington, Jessie G. Taft, Shruti Sannon, Natalya N. Bazarova, and Samuel Hardman Taylor. 2017. "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 (CHI EA '17),* 2853–2859. https://doi.org/10.1145/3027063.3053246
31. Aung Pyae and Tapani N. Joelsson. 2018. 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 '18)*, 127–131. https://doi.org/10.1145/3236112.3236130
32. Aung Pyae and Paul Scifleet. 2018. Investigating Differences Between Native English and Non-native English Speakers in Interacting with a Voice User Interface: A Case of Google Home. *In Proceedings of the 30th Australian Conference on Computer-Human Interaction (OzCHI '18)*, 548–553. https://doi.org/10.1145/3292147.3292236
33. Noa Raz, Ella Striem, Golan Pundak, Tanya Orlov, and Ehud Zohary. 2007. Superior serial memory in the blind: a case of cognitive compensatory adjustment. *Current Biology*: CB 17, 13: 1129–1133. https://doi.org/10.1016/j.cub.2007.05.060
34. Frank Rudzicz, Rosalie Wang, Momotaz Begum, and Alex Mihailidis. 2015. Speech Interaction with Personal Assistive Robots Supporting Aging at Home for Individuals with Alzheimer's Disease. *ACM Transactions on Accessible Computing (TACCESS)*, 7, 2: 6:1–6:22. https://doi.org/10.1145/2744206
35. S. Schlögl, G. Chollet, M. Garschall, M. Tscheligi, and G. Legouverneur. 2013. Exploring Voice User Interfaces for Seniors. *In Proceedings of the 6th International Conference on Pervasive Technologies Related to Assistive Environments (PETRA '13),* 52:1–52:2. https://doi.org/10.1145/2504335.2504391
36. John Sweller. 1988. Cognitive load during problem solving: Effects on learning. *Cognitive Science* 12, 2: 257–285. https://doi.org/10.1016/0364-0213(88)90023-7

37. Alexandra Vtyurina, Adam Fourney, Meredith Ringel Morris, Leah Findlater, and Ryen W. White. 2019. Bridging Screen Readers and Voice Assistants for Enhanced Eyes-Free Web Search. *In Companion Proceedings of the The Web Conference 2019 (WWW '19)*. https://doi.org/10.1145/3308558.3314136

38. Linda Wulf, Markus Garschall, Julia Himmelsbach, and Manfred Tscheligi. 2014. Hands Free - Care Free: Elderly People Taking Advantage of Speech-only Interaction. *In Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational (NordiCHI '14)*, 203–206. https://doi.org/10.1145/2639189.2639251

39. Rayoung Yang, Sangmi Park, Sonali R. Mishra, Zhenan Hong, Clint Newsom, Hyeon Joo, Erik Hofer, and Mark W. Newman. 2011. Supporting Spatial Awareness and Independent Wayfinding for Pedestrians with Visual Impairments. *In The Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '11)*, 27–34. https://doi.org/10.1145/2049536.2049544

40. Randall Ziman and Greg Walsh. 2018. Factors Affecting Seniors' Perceptions of Voice-enabled User Interfaces. *In Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems (CHI EA '18),* LBW591:1–LBW591:6. https://doi.org/10.1145/3170427.3188575