# Beyond Being Human: The (In)Accessibility Consequences of Modeling VAPAs After Human-Human Conversation

Antony Rishin Mukkath Roy[1], Ali Abdolrahmani[1], Ravi Kuber[1] and Stacy M. Branham[2]

[1] UMBC, Baltimore MD 21250, USA
[2] UCI, Irvine CA 92697, USA
{antonyr1,aliab1,rkuber}@umbc.edu, sbranham@uci.edu

**Abstract.** Voice-Activated Personal Assistants (VAPAs) like Amazon's Alexa and Google Assistant have rapidly become pervasive, with users spanning from the youngest young to the oldest old of our society. However, little is known about the nascent VAPA interaction paradigm: what are the fundamental metaphors and guidelines for design, and how do they constrain potential uses and users? This poster begins to answer these questions through a qualitative document review of VAPA design guidelines published by Amazon and Google. Initial results show that human-human conversation is considered the gold standard of interaction. We present an argument that troubles this assumption by adopting a lens of accessible interface design for blind individuals. We advocate VAPA design that moves *beyond being human*.

**Keywords:** Voice-Activated Personal Assistant, Voice Interface, Human-Human Conversation, Design Guidelines, Accessibility, Blindness.

## 1 Introduction and Related Work: VAPAs on the Rise

Voice-Activated Personal Assistants, or VAPAs [8], are becoming widely adopted both in the home and on-the-go [2,10] as per the market survey results. HCI researchers have begun to study the user experience of VAPA ownership by sighted individuals [3,8]. Additionally, studies have examined VAPA interaction experiences of more specific populations including senior users [11,15] and people with disabilities [1,12]. Pradhan et al. [10] found that 38% of Amazon Echo reviews that mention disability involved visual impairments, and Abdolrahmani et al. [1] document accessibility and usability challenges of blind VAPA users. These studies suggest that voice interaction platforms may be particularly useful for the blind community, if properly designed.

Despite wide adoption, there has yet to be an evaluation of how designers conceptualize VAPAs—their purposes, users' needs, and fundamental metaphors of interaction [9]. Our research question is: what VAPA guidelines exist, and how might they constrain potential uses and users of such systems? In this poster paper, we present initial findings

from a document review study of VAPA design guidelines published by Amazon and Google. 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.

## 2 Methods: A Document Review of Two VAPA Guidelines

We conducted this study in two phases. In phase 1, we identified and collected public-facing design guidelines published by commercial VAPA vendors. For this preliminary work, we chose official guidelines of Amazon Alexa[1] and Google Assistant[2] because they are the second and third most used voice assistants [2], respectively, and their guidelines are more centralized and extensive than those for Apple Siri (which dominates the market). The first author downloaded PDF copies of each web page, as well as copies of all the directly-linked pages; a total of 83 letter-size pages from Amazon and 276 from Google were collected in September 2018 for analysis.

In phase 2, we conducted a qualitative thematic analysis on the guidelines collected. The first author read both guidelines in their entirety and generated first-round thematic codes through segment-by-segment coding. The first and third authors reviewed the codes and coded segments from the primary document, and refined/sorted codes to generate the themes and subthemes listed in Table 1.

**Table 1.** Main (sub)themes resulting from analysis of VAPA guidelines, including number of supporting "instances" (guideline sections, figures, or sentences) observed.

| Theme | Subtheme | Google | Amazon |
|---|---|---|---|
| **VAPAs should be portrayed as human** | Way the VAPA speaks and is spoken to is human-like | 6 | 6 |
| **Human conversation is the model of interaction** | Spoken language should be the main in/output medium | 7 | 5 |
| | Interactions should be modeled after multi-turn dialogue | 2 | 4 |
| | Speed, pauses, and duration of turns should mimic human-human conversation | 1 | 6 |
| **Accessibility for disabled people not addressed** | Accessibility refers to non-disabled people or generic UIs | 1 | 2 |

[1] https://developer.amazon.com/designing-for-voice/
[2] https://developers.google.com/actions/design/

# 3 Findings: Guidelines for Designing VAPAs

*3.1. VAPAs should be portrayed as human*

Across both guidelines, the way the VAPA speaks and is spoken to is human-like. For example, in one guideline the VAPA is referred to as a "persona" that should interact with the user with a "unique human voice" and "personality." To achieve this, developers may choose an appropriate human voice type from a text-to-speech (TTS) library based on the language, gender, and local dialect. Another example is the use of wake words—"Alexa" and "OK Google"—which personify the VAPA to initiate a verbal response. In another instance, one guideline claims VAPAs should demonstrate the distinctly human traits of "empathy" and "emotional intelligence."

*3.2. Human conversation is the model of interaction*

*Subtheme 2.1: Spoken language should be the primary input and output medium*

Both guidelines stress spoken natural language, as opposed to other audio media including "earcons"—short, distinctive, non-verbal sounds. The language should be casual, preferring contractions and slang (e.g., "you're" and "wanna") over bookish jargon. One guideline suggests listening to the prompts carefully to test if they sound like spoken language instead of written. Finally, guidelines recommend enabling a broad range of input/output, allowing for subtle command "variations" and even "mispronunciations" from users, and varied output phrases to avoid "stale" or "tiresome" responses.

*3.2.2. Interactions should be modeled after multi-turn dialogue*

Both guidelines advocate concise, multi-turn conversation, wherein the VAPA asks questions that clearly signal when and what the user ought to speak next. There should be one question per turn, to reduce the cognitive load placed on the user. Although one guideline states the system should be able to handle input that chains together multiple answers, these "one-shot" interactions are dispreferred to multi-turn dialogoue.

*3.2.3. Speed, pauses, and duration of turns should mimic human-human conversation*

The cadence of dialogue was constrained by models of human-human communication, with one guideline explicitly invoking the "Cooperative Principle" from linguistics [4]. One guideline tells developers to test the fitness of a prompt by ensuring it can be said with "one breath;" another says to insert pauses of 350 or 400ms between list items. Regarding verbosiy, while one guideline recommends concise conversation without extra words, the other disapproves of both short and lengthy conversation, which can be read as uncooperative; it instead emphasizes content "relevance." Regarding duration, "it shouldn't take more than 20 seconds to read the first few items in the list."

*3.3. Accessibility / usability for disabled people not addressed*

Accessibility was alluded to in both guidelines. One document advocated making the VAPA accessible to users with different levels of English language proficiency. The other advocated "inclusive," "universal" design, but did not include VAPA-specific

recommendations. Neither guideline document referred to designing for accessibility of people with disabilities in general, or people who are blind specifically.

## 4 Discussion: Accessibility Shortcomings of VAPA Guidelines

This initial study of Alexa and Google Assistant guidelines revealed that the interaction design of these applications follows a human-human conversational metaphor. VAPAs are portrayed as human entities with names and personas. And, they speak using everyday language, complete with slang, naturalistic pauses, and turn-by-turn cooperative dialogue. However, these guidelines contributed little on the accessibility of VAPAs for people who are visually impaired. Recent studies demonstrate the potential for these audio-only platforms to provide significant utility to blind users, but accessibility and usability challenges still exist [1,12]. The challenges documented in prior work appear to be affirmed in our current study of design guidelines:

1. Blind VAPA users want to accomplish serious work and are willing to learn specialized commands to become expert users [1]. However, the current study found that guidelines focused on having "interesting" conversations that could be intuitive to novice users.
2. Blind users want to chain together complex commands without being cut off by their personal assistant [1]. However, the current study found that guidelines promote multi-turn dialogue over "one-shot" interactions.
3. Blind users want to adjust interaction speed with a synthetic voice [1] for efficiency. However, the current study found that guidelines dictate fixed speed, pauses, and turn durations.

While developer guidelines foreground cooperative, interesting, human-human-like interactions, blind individuals look to the VAPA as a productivity platform—a human-computer interaction that puts the user in control. This incongruence begs the question: is the human-human metaphor too limiting? What are the alternatives? Are blind individuals the only population who stand to benefit, or could the VAPA become a viable alternative to Graphical User Interfaces for complex, long-duration, work-related tasks? Like Hollan and Stornetta [5], we propose a skeptical approach to the de facto assumption that voice interfaces must position human-human communication as the gold standard. We recommend expanding the interaction paradigms of VAPAs to include those that move beyond being human.

## 5 Conclusions and Future Work

VAPA technologies provide an unparalleled opportunity to support a broad range people who often must resort to special-purpose assistive devices, including people with visual impairments. In addition to the accessibility benefits of voice interfaces, the mainstream status of VAPAs may support adoption by lowering unit costs and reducing stigma often associated with assistive technologies [6, 13]. This poster demonstrates

initial evidence that VAPA guidelines may be reasonable sites for design-time accessibility intervention. Our future work will expand the analysis to three more design guidelines, interview VAPA designers, and co-develop more inclusive guidelines.

## References


1. Abdolrahmani, A, Kuber, R., Branham, S.: “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, ACM, New York, NY, (2018)
2. Bhardwaj, P., Gal, S.: Siri owns 46% of the mobile voice assistant market — one and half times Google Assistant's share of the market, https://www.businessinsider.com/siri-google-assistant-voice-market-share-charts-2018-6 last accessed 2018/09/04.
3. Cowan, B.R., Pantidi, N., Coyle, D., Morrissey, K., Clarke, P., Al-Shehri, S., Earley, D., Bandeira, N.: “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, ACM, New York, NY, Article No. 43 (2017).
4. Grice, P.: "Logic and conversation". In: Perspectives in the Philosophy of Language: A Concise Anthology. Broadview, Canada, pp. 41–58, (1975).
5. Hollan, J., Stornetta, S.: Beyond being there. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp.119-125, ACM, New York, NY (1992).
6. Kane, S.K, Jayant, C., Wobbrock, J.O., Ladner, R.E.: Freedom to roam. In: Proceedings of the 11th International ACM SIGACCESS Conference on Computers and Accessibility, pp. 115-122, ACM, New York, NY (2009).
7. 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, pp. 5286–5297, ACM, New York, NY (2016).
8. Moorthy, A.E., Vu, K.-P.L.: Privacy concerns for use of voice activated personal assistant in the public space. International Journal of Human-Computer Interaction, 31 (4), 307–335, (2015).
9. Murad, C., Munteanu, C., Clark, L., and Cowan, B.R. 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). ACM, New York, NY, USA, 269-276.
10. Perez, S.: 29 million Americans now own a smart speaker, report claims. https://techcrunch.com/2018/01/12/39-million-americans-nowown-a-smart-speaker-report-claims/, last accessed 2018/09/04.
11. Portet. F., Vacher, M., Golanski, C., Roux, C., Meillon, B.: Design and evaluation of a smart home voice interface for the elderly: acceptability and objection aspects. In: Personal and Ubiquitous Computing, 17 (1), 127-144, Springer-Verlag London, UK, (2013).
12. 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, ACM, New York, NY, (2018)

13. Shinohara, K., Wobbrock, J.O.: In the shadow of misperception: assistive technology use and social interactions. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 705-714, ACM, New York, NY, (2011).
14. Shinohara, K., Wobbrock, J.O., Pratt, W.: Incorporating social factors in accessible design. In: Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility, ACM, New York, NY, (2018) (to appear).
15. Wulf, L., Garschall, M., Himmelsbach. J., Tscheligi, M.: 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, pp. 203-206, ACM, New York, NY, (2014).