"I Don't Know What You're Talking About, HALexa":

The Case for Voice User Interface Guidelines

Christine Murad
University of Toronto
Toronto, Canada
cmurad@taglab.ca

Cosmin Munteanu
University of Toronto Mississauga
Toronto, Canada
cosmin.munteanu@utoronto.ca

Abstract

As Voice User Interfaces (VUI) grow in popularity in both the research and academic world, designers are met with new challenges in delivering on the promises of voice interaction. These promises depict a world where one can just speak to their devices, akin to HAL-9000; yet, existing usability challenges still leave many disappointed. These challenges often make or break the experience users have with VUIs. We argue that what we are missing is a foundation on which to build (and deliver) our promises: it is essential to build a foundation of VUI principles that can guide future designers in the development of voice interaction. We must address the lack of research in developing foundational VUI-specific guidelines that can aid designers in meeting the expectations and promises of true voice interaction.

CCS Concepts

• **Human-centered computing → Human computer interaction (HCI)**; **HCI design and evaluation methods**.

Keywords

Voice interaction, design principles, conversational interfaces

ACM Reference Format:
Christine Murad and Cosmin Munteanu. 2019. "I Don't Know What You're Talking About, HALexa":: The Case for Voice User Interface Guidelines. In *1st International Conference on Conversational User Interfaces (CUI 2019), August 22–23, 2019, Dublin, Ireland.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3342775.3342795

1 THE PROMISE THAT NEVER CAME TO BE

Voice user interfaces (VUI) have increased in popularity as devices such as Amazon Alexa, Google Home, and Apple Homepod have grown in the commercial market. We've seen the dreams: to be like Tony Stark with our own JARVIS personal virtual assistant, or to have our own HAL-9000. However, the voice devices of today come no closer to these dreams than did our basic interactive voice response systems of the 70s. These were applications that we have envisioned since the 1980's - and yet, a critical self-reflection should tell us that we have progressed only incrementally in designing interactions with these technologies, despite significant progress on the speech processing side.

Commercially advertised "conversational" interfaces are currently far from conversational. In reality, simple question/answer routines are the norm [6]. Conversational agents like Google Home and Amazon Echo employ command-based interaction, rarely including functionality that would be required for a realistic dialog (e.g., saving the context of previous commands, developing common ground during dialog, employing conversational turn-taking and dynamics, etc). Yet users perceive these systems to have far more human-like conversational abilities than is currently the case [1, 5, 7, 8]. Rather than being a "natural" user interface, these interactions tend to be learned through trial and error, sometimes guided by written instructions. This has not moved much farther from the interaction capabilities of ELIZA [20]. This causes users to often abandon VUIs, due to the usability issues they encounter and the disconnect between their expectations and what current VUIs provide [1, 5, 6].

As VUIs continue to develop and grow, it is imperative to understand and account for these challenges. What we are still missing is a foundation. While commercial voice devices abound, in no small part due to affordability, we still lack the high-level groundwork of how VUIs should be designed. In particular, a major problem that VUI design currently faces is a lack of design principles that can guide users in good VUI design.

Compared to the even-more ubiquitous Graphical User Interfaces (GUIs), we have been designing voice-enabled devices without a any theoretical principles or guidelines pertinent to the conversational voice interfaces embedded inside them. They have been advertised as a natural way to interact with technology [7, 9, 16]. However, without any guidance or understanding on how such smart devices are meant to be interacted with, these become unusable for many groups regardless of the artificial intelligence that they access. This is visible in recent examples of digitally marginalized users such as older adults trying to interact with digital home assistants (Figure1); even as the marketing hype *du jour* is that older adults can benefit the most from such devices. While there are no doubts about the accessibility of these interfaces, what they offer is only marginally more than what an information kiosk or automation control would.

While some sets of VUI heuristics have been appearing in academic research currently [17, 19], these have not been extensively validated and implemented in the commercial space. They also fail to include current designers and expert in a bottom-up process in developing them. Therefore, there is a need to direct more research into development and validation of VUI heuristics. If we are to bring voice interaction closer to the promises we have been making for decades, we cannot ignore this necessary step.

CUI 2019, August 22–23, 2019, Dublin, Ireland

ACM ISBN 978-1-4503-7187-2/19/08…$15.00
https://doi.org/10.1145/3342775.3342795

Figure 1: An older user attempting to interact with Google Home (youtu.be/e2R0NSKtVA0,© Ben Actis, retrieved 2018-01-26).

2 ARE OUR CURRENT PRINCIPLES NOT ENOUGH?

The current state of VUI research and development has been growing rapidly, and devices such as Google Home and Amazon Alexa are being praised as "the future" of interaction [2]. However, recent research shows us that many usability problems are still not solved. Issues such as error correction, understanding what you can say to VUIs, and system feedback are still rampant within commercial VUIs [5]. GUIs have long used heuristics as the key to success for design and to avoid usability problems. While VUIs are missing such heuristics, a parallel can be drawn with the first generation of mobile UIs which did not follow proper heuristics either, earning them a "usability trashing" from Jakob Nielsen [13]. However, this negative assessment was updated a few years after [3], largely due to improvements in usability due to designers of mobile UIs observing good usability heuristics. Nielsen and colleagues have made similar complaints about speech interfaces, some very recently [4] but also going back almost two decades [11, 12]. This begs the question: why are we still not embracing heuristics (usability in particular, but also design in general) for VUIs? Should we not follow the same successful path taken by mobile UIs?

Unfortunately, the broad attitudes toward developing VUI-specific heuristics continue to run counter to the training needs of the next generation of designers. We are still observing a broad attitude of questioning whether developing foundational VUI-specific heuristics is necessary, and whether they would be largely different than existing design heuristics and guidelines that were created to be applied generally to all interfaces. The research we have conducted shows how imperative it is to address such misaligned perspectives (which seem to permeate both the HCI and the NLP fields). Other emerging disciplines, such as in virtual reality [18] and video game design [15] have demonstrate that the creation of paradigm-specific heuristics is necessary to ensure the proper design of these novel interfaces.

Indeed, designing for voice is much different than designing for graphical interfaces [22]. As previous research has shown, GUI principles cannot be directly adapted to VUIs [22]. However, using existing usability heuristics as a foundation for developing new heuristics is quite possible [11, 21]. This may even be preferable, if the "target users" of these adapted heuristics are designers who are steeped in GUI design but are "forced" to rapidly transition to designing VUIs under the diktat of the current market hype.

However, even then, this requires particularly looking at the specific design challenges that a new paradigm faces, and evaluating what changes are necessary to adapt existing heuristics to be applicable to a different field. There is a lack of this kind of research in HCI currently. Without this research, we will continue to develop voice interfaces without an established, working foundation that we can guarantee addresses common usability issues and ensures an intuitive and usable interface.

3 IS VOICE INTERACTION TOO "NEW"?

Voice interaction research has been steadily growing over the past 5 years or so. For example, in CHI 2018, there were 3 sessions dedicated to voice/speech or conversational interaction, and in CHI 2019, that number grew to 5 sessions (with papers on voice mixed into other sessions as well). This brings about the attitude that voice is still an "emerging" interaction technique – both in academic research and in HCI Education (where VUIs are presented as a "future/emerging interaction" [10]).

However, voice and speech research has existed for decades. Voice and speech research are not novel – however the technological capabilities decades ago prevented us from fully taking advantage of the promises of these devices. We are in a position now where the technology that we have is advanced enough that we can focus on the voice interaction and interface design of these devices.

We argue that now is the time to develop high-level design guidelines for voice interaction. We have been developing commercial voice devices for many years now, and we have seen their popularity grow. These devices are being adopted into people's homes and are shaping people's impressions about voice interaction. A foundation needs to exist prior to the development of technology, or we will continue to self-inflict usability problems with every step of progress we make, and ultimately sour the users away from these technologies.

4 A CALL TO ACTION

10 years ago, Nielsen [13] spoke about the significant usability issues the iPhone contained. An updated review was then written, talking about the improvement in usability [13]. At this point in time, we may be in the same situation mobile interfaces were a decade ago [14]. In order to take advantage of the capability of voice interaction, we must build this foundation and develop a set of principles that can guide us in building VUIs. What must these principles entail? The HCI community needs to engage in the same kind of research that Nielsen did, in order to improve the usability of voice UIs. Our hope is that by exploring currently established guidelines as a baseline, we will be in a position to identify and develop a taxonomy of design guidelines to assist the HCI community in building more usable and intuitive speech interfaces. Otherwise, we will be doomed to a decade of bad usability for VUIs – if the video of a lovely grandma struggling with her Google Home is not telling enough, just ask a certain design guru about his own VUI usability challenges [14].

References

[1] Matthew P. Aylett, Per Ola Kristensson, Steve Whittaker, and Yolanda Vazquez-Alvarez. 2014. None of a CHInd. *Proceedings of the extended abstracts of the 32nd annual ACM conference on Human factors in computing systems - CHI EA '14* (2014), 749–760. https://doi.org/10.1145/2559206.2578868

[2] Kim Brunhuber. 2018. The hottest thing in the world of technology: your voice | CBC News. https://www.cbc.ca/news/technology/brunhuber-ces-voice-activated-1.4483912

[3] Raluca Budiu. 2015. Progress in Mobile User Experience. https://www.nngroup.com/articles/mobile-usability-update/

[4] Raluca Budiu and Page Laubheimer. 2018. Intelligent Assistants Have Poor Usability: A User Study of Alexa, Google Assistant, and Siri. https://www.nngroup.com/articles/intelligent-assistant-usability/

[5] 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. *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services - MobileHCI '17* (2017), 1–12. https://doi.org/10.1145/3098279.3098539

[6] Emer Gilmartin, Benjamin R. Cowan, Carl Vogel, and Nick Campbell. 2017. Exploring Multiparty Casual Talk for Social Human-Machine Dialogue. In *Speech and Computer (Lecture Notes in Computer Science)*, Alexey Karpov, Rodmonga Potapova, and Iosif Mporas (Eds.). Springer International Publishing, 370–378.

[7] Ewa Luger and Abigail Sellen. 2016. "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16* (2016), 5286–5297. https://doi.org/10.1145/2858036.2858288

[8] Cosmin Munteanu, Ben Cowan, Keisuke Nakamura, Pourang Irani, Sharon Oviatt, Matthew Aylett, Gerald Penn, Shimei Pan, Nikhil Sharma, Frank Rudzicz, and Randy Gomez. 2017. Designing Speech, Acoustic and Multimodal Interactions. *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '17* (2017), 601–608. https://doi.org/10.1145/3027063.3027086

[9] Cosmin Munteanu, Matt Jones, Steve Whittaker, Sharon Oviatt, Matthew Aylett, Gerald Penn, Stephen Brewster, and Nicolas D'Alessandro. 2014. Designing speech and language interactions. *Proceedings of the extended abstracts of the 32nd annual ACM conference on Human factors in computing systems - CHI EA '14* (2014), 75–78. https://doi.org/10.1145/2559206.2559228

[10] Christine Murad and Cosmin Munteanu. 2019. Teaching for Voice: The State of VUI Design in HCI Education. *Proceedings of EduCHI 2019 Symposium* (2019).

[11] 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*. ACM Press, New York, New York, USA, 269–276. https://doi.org/10.1145/3236112.3236149

[12] Jakob Nielsen. 2003. Voice Interfaces: Assessing the Potential. https://www.nngroup.com/articles/voice-interfaces-assessing-the-potential/

[13] Jakob Nielsen. 2009. Mobile Usability, First Findings. https://www.nngroup.com/articles/mobile-usability-first-findings/

[14] Donald Norman. 1988. The design of everyday things. *Doubled Currency* (1988).

[15] David Pinelle, Nelson Wong, and Tadeusz Stach. 2008. Heuristic evaluation for games: usability principles for video game design. *Proceedings of SIGCHI Conference on Human Factors in Computing Systems* (2008), 1453–1462. https://doi.org/10.1145/1357054.1357282

[16] Anjeli Singh, Andrea Johnson, Hanan Alnizami, and Juan E Gilbert. 2011. The Potential Benefits of Multi_Modal Social Interaction on the Web for Senior Users. *J. Comput. Sci. Coll.* 27, 2 (2011), 135–141. http://dl.acm.org/citation.cfm?id=2038836.2038856

[17] Bernhard Suhm. 2003. Towards Best Practices for Speech User Interface Design. (2003), 2217–2220.

[18] Alistair Sutcliffe and Brian Gault. 2004. Heuristic evaluation of virtual reality applications. *Interacting with Computers* 16, 4 (2004), 831–849. https://doi.org/10.1016/j.intcom.2004.05.001

[19] Z. Wei and J. A. Landay. 2018. Evaluating Speech-Based Smart Devices Using New Usability Heuristics. *IEEE Pervasive Computing* 17, 2 (April 2018), 84–96. https://doi.org/10.1109/MPRV.2018.022511249

[20] J. Weizenbaum. 1966. ELIZA- A computer program for the study of natural language communication between men and machine. *Commun. ACM* 9 (1966), 36–45. https://doi.org/10.1145/365153.365168

[21] Kathryn Whitenton. 2016. Voice Interaction UX: Brave New World...Same Old Story. https://www.nngroup.com/articles/voice-interaction-ux/

[22] Nicole Yankelovich, Gina-Anne Levow, and Matt Marx. 1995. Designing SpeechActs. *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '95* (1995), 369–376. https://doi.org/10.1145/223904.223952