What Would You Say? Understanding Speech with Observational HCD

Benett Axtell
University of Toronto, TAGlab
Toronto, ON, Canada
benett.axtell@mail.utoronto.ca

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

ABSTRACT

Speech interactions are available in many everyday devices, but are predominantly command-and-execute Conversational Agents, like Alexa. Recent examples demonstrate how users' differing mental models about voice interactions (VUIs) cause these devices not to be used to the extent afforded by the continuing engineering advances. What our design toolboxes for VUIs are missing are observational methods to understand users early. Such methods have been used for over four decades for GUIs, but rarely for VUIs. We argue here for new ways to apply these methods to the VUI design space, that will allow us to move past interaction paradigms and metaphors currently limiting the potential of speech.

USER OBSERVATIONS TO DRIVE SPEECH INTERACTIONS

Conversational Agents (CAs) and similar interfaces are currently far from "conversational". Instead, they rarely incorporate realistic dialogue (e.g., saving context of previous commands, developing common ground in dialogue, structuring responses with conversational turn-taking and dynamics). Yet users perceive them to have far more human-like conversational abilities than is currently the case [4]. In reality, rather than being a "natural" user interface, these interactions tend to be learned through trial and error [3]. This has not moved much farther from the interaction capabilities of ELIZA [5].

CHI'19 Extended Abstracts, May 4-9, 2019, Glasgow, Scotland, UK.

ACM ISBN 978-1-4503-5971-9/19/05.
DOI: https://doi.org/10.1145/3290607.XXXXXXX

KEYWORDS
Speech interaction; human-centred design; contextual inquiry;

LEARNING USERS' EXPECTIONS

Design of usable speech interfaces requires a strong user understanding. CIs can provide this but are not sufficient on their own. They are part of a larger HCD research process:

1. User Observations	Observing users' processes through **CIs creates new speech interactions** beyond conversational as defined by their needs rather than what is currently available.
2. User Understanding	Driven by user observations like CIs, the understanding of a **user's mental model drives speech interaction design to new possibilities** as well as greatly improves existing tools to reflect user expectations and adjust for gaps in knowledge.
3. User Assessment	With designs based in observed user needs, WoZ studies, or other formative methods for existing designs, enable quick and meaningful VUI assessment without issues of development time and limitations of available technology.

HCD design principles seem to be absent from available speech-enabled devices. As such, these are apparently presented without an underlying understanding of how we use speech with technology and what we would like to do with that speech. Little work has been done here, and there is still much to be learned from users [1]. Previous works into use of existing tools like CAs have shown that there is poor learnability and a misalignment of user expectations versus actual interaction experience [4].

Research into the learnability of voice-based interfaces found that due to the lack of clear affordances of functionality most users simply guessed at what could be said and quickly settled on the few commands they knew rather than exploring further [3]. These difficulties likely result from a mental model mismatch between users' expectations of speech interaction and the realities of the technology. Unless we consider user expectations for speech, design of VUIs risks becoming stalled.

Within a technology space, mental models capture how users understand a system. They define what they believe about how an interactive system or digital technology works [6]. For VUIs, a user's mental model is their knowledge of what they can do with their speech, what they can say, and how that speech is used by a given technology. In terms of CAs, mental models rely on experiences with human-human dialog (which differs from the realities of human-computer dialog) and includes how users might recover from errors. Exposing how we use and adapt speech contributes to a larger mental model, but little research is directly building onto general mental models of speech interaction.

Creating an understanding of a user and their process is an essential step of HCD research. Contextual inquiries (CIs) and other observational methods accomplish this by involving participants early and lead to designs that support what users need while avoiding researcher assumptions [2]. These observations build understandings of relevant activities without involving new technologies and biasing users (for or against) a novel tool. This resulting understanding is particularly important when use of the intended modality is poorly defined and underexplored, as is true for voice interactions.

Through CIs, participants with any level of digital literacy or familiarity with a given technology can contribute to designing new interactions. Asked to create or assess a design for a new speech-based tool would seem daunting to many users, and they would likely fall back on what they have seen and used before. This furthers conversational as the gold standard for speech interactions. CI, on the other hand, provides the space for new interactions regardless of what is currently familiar.

A CI observation builds the initial understanding without introducing the confusion of new speech interactions, including mismatched mental models. An understanding built separate from pre-existing ideas of speech interactions further allows for new VUIs to be designed outside of current standards or expectations. In contrast with other formative methods (e.g., focus groups, Wizard of Oz [WoZ]), CIs help researchers set aside assumptions by keeping participants and their current practices in context. As long as the needs of speech interactions are poorly understood, we see CIs as the necessary first step to design new VUIs supporting what users want to do with speech, over what they may expect.

CI is a powerful, established tool that supports the essential work of identifying new opportunities for speech interactions. There are likely other options, some adapted from existing methods and others created explicitly for speech-enabled settings, but we must intentionally seek these out if any progress is to be made for this modality. The lack of familiar, accepted methods for speech has contributed to the dominance of command-and-execute interactions over any alternative and to the mismatch in user expectations. In this position paper we have argued for the use of early-stage HCD methods, such as CI, to elicit more meaningful and ecologically valid user and design requirements. We expect that such CI-derived requirements will match users' mental models and thus lead to fewer usability issues and higher adoption.

REFERENCES

[1] Matthew P. Aylett, Per Ola Kristensson, Steve Whittaker, and Yolanda Vazquez-Alvarez. 2014. None of a CHInd: relationship counselling for HCI and speech technology. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*, 749–760.

[2] Hugh Beyer and Karen Holtzblatt. 1997. Principles of Contextual Inquiry. In *Contextual design: defining customer-centered systems*. Elsevier, 41–78.

[3] Eric Corbett and Astrid Weber. 2016. What can I say? Addressing user experience challenges of a mobile voice user interface for accessibility. *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services - MobileHCI '16*: 72–82.

[4] 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.

[5] J. Weizenbaum. 1966. ELIZA- A computer program for the study of natural language communication between men and machine. *Communications of the ACM* 9: 36–45.

[6] 2018. A Very Useful Work of Fiction – Mental Models in Design. *The Interaction Design Foundation*.