A Voice-First Approach to Educational Technology: Opportunities and Challenges for Inclusive Design

Doon MacDonald
CHERISH-DE
Swansea University
Swansea
SA1 4PY
d.g.macdonald@swansea.ac.uk

Anja Thieme
Microsoft Research Ltd
Cambridge,
CB1 2FB
anthie@microsoft.com

Abstract
This paper discusses the application and potential suitability of voice-user interfaces (VUIs) for the design of collaborative learning experiences that are inclusive to all children, regardless of their visual abilities. The paper describes some of the opportunities and challenges that can arise when considering the design of speech-based systems for coaching and learning from a human computer interaction perspective. To this end, we outline a series of challenges and questions for the design of VUIs with a particular focus on the development of: (i) voice-interaction technology to encourage social inclusion and support the development of social skills; and (ii) voice or non-speech audio output displays for creating more effective eyes-free technology interactions.

Author Keywords
Voice Interfaces, Voice User Interface; VUI; Learning; Inclusive design; Vision impairment, Accessibility.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

Introduction

More and more speech-based interfaces find their way into people's everyday lives. They include conversational agents that range from Amazon's Alexa, to Apple's Siri, Microsoft's Cortana, Facebook's M, Google's Assistant, and IBM's Watson [5]. They find application on many modern smartphone, tablet and PC as well as smart watches, and most recently are embedded in standalone devices like Amazon's Echo and the Google Home [6, 7, 9]. These voice-agents offer users direct access to simple functionality via speech and are growing in popularity and adoption. A recent consumer survey conducted in the US [cf. 4] suggests that now nearly 46% of Americans use voice-assistants, also called voice-user interfaces (VUIs) or conversational UIs, on their smartphones and other devices; a trend that is rapidly growing. It is often suggested that technology interactions via speech feel more nature [2] and presents a convenient way to retrieve information eyes- and hands-free. This can offer great advantages over other, often screen-based interactions especially whilst on-the-move or immersed in other activities such as driving a car or cooking.

Beyond its increasing mainstream use and acceptance, voice-interaction is and continues to be especially transformational for people with accessibility need [cf. 8]. People with vision impairments (VI) in particular are power users of speech-based technology. They use speech to set reminders, dictate messages; or label images, and are active consumers of voice-over and features and screen reader technology to navigate and listen to digital information. In addition, other more bespoke services such as Blindsquare [1] or Seeing AI [11] that are developed with a VI community in mind utilize speech and audio guidance to assist navigation or sense-making of visual features in the environment. Voice-assistants, especially Alexa, are also very popular with children, who are early adopters [13] and often request songs, ask for help with their homework or seek to control other devices in the home.

These advantages of voice-interactions in terms of ease-of-use and accessibility combined with an increase in the availability, affordability and mainstream acceptance of voice-based systems, suggests that this technology may be particularly suitable for the design of more inclusive educational applications by providing voice-based coaching or learning experiences to all children, regardless of their visual abilities. Against this backdrop, we regard this position paper as a starting point for discussions of the opportunities and challenges for design that prioritizes voice-interactions.

Following a voice-first approach to design, our focus is therefore on new products such as the Amazon Echo and Google home, which are voice-led systems that do not have or rely on any uses of additional screens. This makes them distinct from speech-based systems that accept text or other forms of user inputs (e.g. via touch-screens). Thus, by focusing solely on enabling audio-based interactions the aim is to create applications that support learning experience that are equally accessible and beneficial to both sighted and non-sighted learners; thereby reducing risks of exclusion or stigma that can result when VI users must interact differently with a device while others may interact visually. The challenge then remains how to identify and create suitable educational scenarios where VUIs and the use of rich audio representations can enable meaningful learning experiences by itself, without any reliance on visual media or content.

Furthermore, existing standalone speech-devices like the Amazon Echo are designed to be placed in social spaces, such as people's home, which expands their scope for use from a single user towards a range of social and collaborative contexts [3, 7]. Thus, we seek to explore what this new wave of voice-only technology has to offer for learning in classroom settings or the completion of homework together with other learners. In particular, we are interested in its potential use for collaborative learning by children with mixed visual abilities [12]. This likely poses many unique challenges and constraints for design including for example: how to support collocated multi-user interactions with VUIs within often complex, real-world educational contexts?

Design Challenges & Opportunities for VUIs

In the workshop, we would like to promote discussion around the following two broad challenges from an HCI design and research perspective:

1. How to design VUIs to support social inclusion and the development of social skills?

2. How to design effective voice or non-speech audio output displays to create more compelling eyes-free technology interactions?

Outlined below are a series of additional related questions and design challenges for discussion. We propose that one outcome could be the development of a number of user-centered scenarios based on these questions and challenges that could be employed in the design of VUIs for inclusive educational settings:

Voice-interaction at home/ school
How to design interactive learning tasks that permit and actively encourage users to share tasks?

Creativity and Play
how to create VUI interactions that nurture imagination, promote physical tasks, and assist in learning through (social) play.

Social skills learning
Recently, concerns were raised about how the use of VUIs impact young people's development of social skills [10]. How can and should we respond sensitively and appropriately to such concerns in the design of VUIs? What are the opportunities of voice-based interactions for assisting social skills learning as well as for promoting human-to-human communications?

Remote Human-to-Human Learning
how to utilize the design of VUIs such as the Echo, as a 'speaker' and 'broadcaster' interface to support remote learning?

Foregrounding audio as a rich, alternative means for experiencing and making sense of the world
How to utilize specific sound features such as spatialized audio and real-time feedback to assist in the learning complex, often abstract concepts that are often encountered in STEM subjects such as physics? How may the use of spatial bandwidth and natural processes of hearing enable the creation of a more `holistic' learning environment?

Conclusion

The use of voice-user interfaces offers rich potential for the design of the next generation of inclusive and socially more collaborative learning technologies. This paper argues that is worth considering both the opportunities and challenges that these technologies bring from an HCI design and research perspective.

References

1. BlindSquare. *Mobile phone application* Last accessed: 2nd February 2018 from *http://www.blindsquare.com/*
2. The Economist. How voice technology is transforming computing. *Online article.* Last accessed: 9th March 2018 from https://www.economist.com/news/leaders/21713836-casting-magic-spell-it-lets-people-control-world-through-words-alone-how-voice
3. Moira McGregor and John C. Tang. 2017. More to Meetings: Challenges in Using Speech-Based Technology to Support Meetings. In *Proc. CSCW 2017*, 2208-2220. https://doi.org/10.1145/2998181.2998335
4. Kenneth Olmstead. 2017. Nearly half of Americans use digital voice assistants, mostly on their smartphones. *Online article*. Last accessed 4th February 2018 from *http://www.pewresearch.org/fact-tank/2017/12/12/nearly-half-of-americans-use-digital-voice-assistants-mostly-on-their-smartphones/*
5. Robert J. Moore, Raphael Arar, Guang-Jie Ren, and Margaret H. Szymanski. 2017. Conversational UX Design. In *Ext. Abstracts CHI 2017*. 492-497. https://doi.org/10.1145/3027063.3027077
6. Martin Porcheron,, Joel E. Fischer, Stuart Reeves, and Sarah Sharples. 2018. Voice Interfaces in Everyday Life. In *Proc. CHI' 2018*, to appear. http://www.cs.nott.ac.uk/~pszsr/files/porcheron-2018-voice-interfaces-in-everyday-life.pdf
7. Martin Porcheron, Joel E. Fischer, Moira McGregor, Barry Brown, Ewa Luger, Heloisa Candello, and Kenton O'Hara. 2017. Talking with Conversational Agents in Collaborative Action. In *CSCW 2017, Companion*, 431–436. https://doi.org/10.1145/3022198.3022666
8. 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. In *Personal and Ubiquitous Computing 17*, 127: 127–144. https://doi.org/10.1007/s00779-011-0470-5
9. Stuart Reeves, Martin Porcheron, Joel E. Fischer, Heloisa Candello, et al. 2018. Voice-based Conversational UX Studies and Design. In *Ext. Abstracts CHI 2018*, to appear. https://voiceux.files.wordpress.com/2018/01/voicebasedconversationaluxstudiesanddesign-preprint.pdf
10. Oliver Rudgard. 2018. 'Alexa generation' may be learning bad manners from talking to digital assistants, report warns. Online article. Last accessed 2nd February 2018 from *http://www.telegraph.co.uk/news/2018/01/31/alexa-generation-could-learning-bad-manners-talking-digital/*
11. Seeing AI. *Mobile phone application*. Last accessed 2nd February 2018 from *https://www.microsoft.com/en-us/seeing-ai/*
12. Anja Thieme, Cecily Morrison, Nicolas Villar, Martin Grayson, and Siân Lindley. 2017. Enabling Collaboration in Learning Computer Programing Inclusive of Children with Vision Impairments. In Proc. DIS 2017. 739-752. https://doi.org/10.1145/3064663.3064689
13. Hayley Tsukayama. 2017. When you're your kid tries to say 'Alexa' before 'Mama'. Online Article. Last accessed 9th March 2018 from http://www.latimes.com/business/technology/la-fi-tn-tech-kids-20171124-story.html