Enabling Creative Crowd Work though Smart Speakers

Danula Hettiachchi
The University of Melbourne
Melbourne, Australia
danula.hettiachchi@unimelb.edu.au

Niels van Berkel
The University of Melbourne
Melbourne, Australia
niels.van@unimelb.edu.au

Tilman Dingler
The University of Melbourne
Melbourne, Australia
tilman.dingler@unimelb.edu.au

Fraser Allison
The University of Melbourne
Melbourne, Australia
fraser.allison@unimelb.edu.au

Vassilis Kostakos
The University of Melbourne
Melbourne, Australia
vassilis.kostakos@unimelb.edu.au

Jorge Goncalves
The University of Melbourne
Melbourne, Australia
jorge.goncalves@unimelb.edu.au

ABSTRACT

Digital voice assistants or smart speakers have rapidly changed the landscape of voice user interfaces over the past few years. In this paper we discuss how we could utilise the affordances of these devices to create a novel crowdsourcing platform that deliver crowd tasks through voice with particular focus on creative tasks. We describe the potential benefits and challenges of using this technology for these purposes, and outline our future work in this research area.

KEYWORDS

Crowdsourcing; smart speakers; voice assistants; conversational interfaces

CHI'19, May 2019, Glasgow, UK
2019. ACM ISBN 123-4567-24-567/08/06…$15.00
https://doi.org/10.475/123_4

[1]https://store.google.com/category/connected_home

[2]https://www.amazon.com/amazon-echo-and-alexa-devices

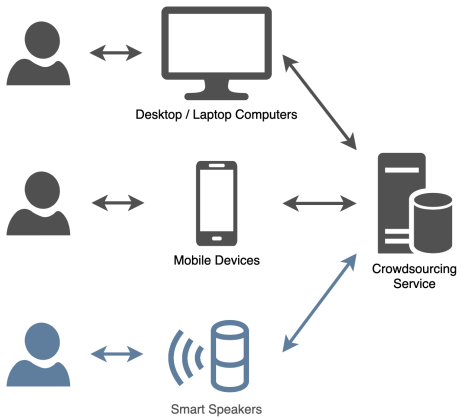

Figure 1: Proposed Crowdsourcing workflow

INTRODUCTION

Advancements in the availability and technical capabilities of voice assistants or specially smart speakers, such as Google Home[1] and Amazon Echo[2] have made conversational user interfaces more accessible and natural than ever before [2]. We propose to leverage the robust automatic speech recognition engines in smart speakers to create a highly pervasive voice-powered crowdsourcing marketplace. Smart speakers placed in peoples' homes provide an easily accessible, ambient interface that can be operated while multitasking. In contrast to traditional crowdsourcing modes such as online [3], mobile [9], and situated crowdsourcing [6], hands-free interaction with a crowdsourcing platform could enable visually impaired workers which are otherwise unable to complete crowdsourcing tasks, allow workers to complete tasks during regular household chores (e.g., while cooking or washing the dishes), and allow for an interaction modality that is more suitable for the task at hand (i.e., tasks that benefit voice input).

RELATED WORK

While there are many crowdsourcing tasks that involve audio or speech data, there is only a handful of studies in the literature that examine voice based interactions for crowdsourcing. Vashistha et al. [12] investigated the possibility of using voice responses of workers for crowdsourcing transcription tasks. In their study, participants were asked to listen to audio clips through a smartphone application and repeat the same words into the application (while in a quiet environment). Input was processed using an automatic speech recognition engine and passed through a number of merging steps to obtain the transcription. The results indicate that the proposed method reduces the cost of speech transcription while improving the quality of output. An extension of the system created for visually impaired users has also produced promising results [13]. The use of smart speakers enables people to engage in microinteractions [1], which are well suited for micro tasks. Furthermore, these microinteractions can be triggered at opportune times at the person's discretion.

PROPOSED METHOD

We propose to use smart speakers as a stand-alone device workers use to connect with the crowdsourcing platform as shown in Figure 1. The entire crowdsourcing process is handled similarly to how workers use their computer or mobile device to complete crowd tasks. Such a system would allow workers to initiate, select, complete, and review crowdsourcing tasks solely using natural language input (as detailed in Sidebar 1). While task content could be directly incorporated in an appropriate manner, the workflow for initiating and managing tasks should be redesigned to ensure suitability for voice-based interactions.

Crowd Worker: Computer, give me some crowd tasks.
Smart Speaker: There are summarising and transcription tasks available. Which one would you like to attempt?
Crowd Worker: I'll try summarising.
Smart Speaker: Okay. Starting summarising task. I will read out a small paragraph. Please read out the summary in a single sentence once I finish.

Sidebar 1: An example of a conversation between a worker and the smart speaker to initiate a crowdsourcing task.

Potential Benefits

Low cost of entry. Most crowdworkers use desktop or laptop computers to interact with crowdsourcing systems. However, leveraging smart speakers could greatly reduce the overhead of initiating a crowd task. For instance, if a worker is using a web application, they need to start the computer and log into the crowdsourcing platform through a browser. Therefore one would not start a task unless they are committed to work for a certain period of time. Interaction through smart speaker would allow them to start working on their task immediately, even when there is only a small time window available. The idea is to provide a complementary crowdsourcing tool, through which workers can work on tasks in opportune moments throughout the day.

Ubiquitous nature. Smart speakers would allow workers to complete tasks without being physically constrained to a space. This greater freedom means workers could handle different equipment, stay in a comfortable posture, or attend to any other work while completing tasks. Since voice assistants reside in users' homes (and often even in multiple rooms), tasks can be completed ubiquitously. Amazon Alexa further provides a mobile application, through which tasks can even be completed on-the-go.

Efficiency. There is a range of tasks that could be crowdsourced over voice such as speech and audio related tasks. Given that Automatic Speech Recognition is more efficient than typing text input in mobile devices [10], it would be more efficient to complete them over voice. While typing in a physical keyboard is still more efficient than speech input, we note that it does not provide the convenience or ubiquity we discussed above.

Accessibility. Voice-interaction through smart speakers could enable more people to get connected with crowdsourcing platforms. Specifically, for people who find it difficult to use traditional computers or mobile devices (e.g., people with vision impairments). It would also be helpful for workers to overcome certain situational impairments [11].

Challenges

Multi-user interaction. Although smart speakers are typically attached to a single user profile, they are designed to be utilised by multiple people who share the same physical space. This multi-user interaction is unlikely to occur in traditional crowdsourcing methods (e.g. online crowdsourcing, mobile crowdsourcing) where payments, pre-qualifications, and performance data are linked to a single user profile. One alternative to overcome this would be to adopt a model where workers are anonymous, similar to how crowdsourcing activities are conducted in most situated crowdsourcing scenarios (e.g. public displays) [7, 8]. We also note that these devices are capable of recognising individual users in the context of checking personal updates. Hence it would be technically feasible to configure such that multiple users maintain their own crowd work account.

Table 1: A summary of potential benefits and challenges.

Potential Benefits
Low cost of entry
Ubiquitous nature
Efficiency
Accessibility
Challenges
Multi-user interaction
Privacy concerns
Integration with other modes
Impact on data quality

Privacy Concerns. Smart speakers may not be suitable for crowdsourcing tasks that contain sensitive material as they broadcast task content to whoever occupy the physical space. Suitable warnings and checks should be incorporated to prevent this. In addition, voice interfaces could potentially create noise pollution in private environments causing discomfort to others, so social considerations should be taken into account.

Integration with other modes. As smart speakers alone cannot deliver all types of crowd work, it would be beneficial to have different interaction modes like web and mobile available for workers in addition to the voice interface. While this could be challenging to achieve, it would result in a more productive crowdsourcing system that could route tasks to the appropriate mode of interaction, instead of relying on a single type input.

Impact on data quality with multitasking. Though we discussed enhanced ubiquity as a potential benefit of the proposed system, it is vital to be aware of any impact on the data quality when workers engage in other tasks while completing crowd tasks. Appropriate announcements to grab the attention of the workers and different quality control mechanisms like gold standard questions should be employed to mitigate this concern.

CONCLUSION AND FUTURE WORK

While voice-based crowdsourcing may not be feasible for certain crowd tasks, given the growth in smart speakers, we argue that there is substantial potential for a voice-based crowdsourcing market. We identify a wide array of crowdsourcing tasks that can be conveniently delivered via smart speakers. The most promising candidates are speech-related crowdsourcing tasks such as collecting, translating, transcribing, and assessing the quality of speech data [14]. In addition, any classification or rating task which does not depend on visual cues could be crowdsourced through smart speakers. This includes more complex and creative tasks such as recording music [5], providing feedback or instructions, and creating detailed descriptions, summaries or reviews [4].

However, rigorous studies on the feasibility of implementing such a system, impact on the task performance and data quality, and the satisfaction of the workers are desirable to pave way for commercial voice based crowdsourcing platforms. With the ability to interact with voice, smart speakers will also enhance the accessibility of crowdsourcing markets. This will also be another research avenue to consider for future work. Further, with the tendency in people to engage more with technology through voice, the requirement to crowd source speech and other audio related data will also rapidly increase. Therefore a voice-based crowdsourcing platform will also be of high value for task requesters.

REFERENCES

[1] D. Ashbrook. 2007. *Supporting mobile microinteractions.* Ph.D. Dissertation. PhD thesis, Georgia Institute of Technology.

[2] Frank Bentley, Chris Luvogt, Max Silverman, Rushani Wirasinghe, Brooke White, and Danielle Lottridge. 2018. Understanding the Long-Term Use of Smart Speaker Assistants. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 3, Article 91 (Sept. 2018), 24 pages. https://doi.org/10.1145/3264901

[3] Djellel Eddine Difallah, Michele Catasta, Gianluca Demartini, Panagiotis G. Ipeirotis, and Philippe Cudré-Mauroux. 2015. The Dynamics of Micro-Task Crowdsourcing: The Case of Amazon MTurk. In *Proceedings of the 24th International Conference on World Wide Web (WWW '15)*. IW3C2, Geneva, Switzerland, 238–247. https://doi.org/10.1145/2736277.2741685

[4] Steven P. Dow and Scott R. Klemmer. 2011. Shepherding the crowd: an approach to more creative crowd work. *CHI EA* (2011).

[5] Carlos Gomes, Daniel Schneider, Katia Moraes, and Jano De Souza. 2012. Crowdsourcing for music: Survey and taxonomy. In *2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 832–839.

[6] Jorge Goncalves, Denzil Ferreira, Simo Hosio, Yong Liu, Jakob Rogstadius, Hannu Kukka, and Vassilis Kostakos. 2013. Crowdsourcing on the Spot: Altruistic Use of Public Displays, Feasibility, Performance, and Behaviours. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '13)*. ACM, New York, NY, USA, 753–762. https://doi.org/10.1145/2493432.2493481

[7] Jorge Goncalves, Simo Hosio, Niels van Berkel, Furqan Ahmed, and Vassilis Kostakos. 2017. CrowdPickUp: Crowdsourcing Task Pickup in the Wild. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 51 (Sept. 2017), 22 pages. https://doi.org/10.1145/3130916

[8] Simo Hosio, Jorge Goncalves, Vili Lehdonvirta, Denzil Ferreira, and Vassilis Kostakos. 2014. Situated Crowdsourcing Using a Market Model. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology (UIST '14)*. ACM, New York, NY, USA, 55–64. https://doi.org/10.1145/2642918.2647362

[9] Mohamed Musthag and Deepak Ganesan. 2013. Labor Dynamics in a Mobile Micro-task Market. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 641–650. https://doi.org/10.1145/2470654.2470745

[10] Sherry Ruan, Jacob O. Wobbrock, Kenny Liou, Andrew Ng, and James A. Landay. 2018. Comparing Speech and Keyboard Text Entry for Short Messages in Two Languages on Touchscreen Phones. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 4, Article 159 (Jan. 2018), 23 pages. https://doi.org/10.1145/3161187

[11] Zhanna Sarsenbayeva, Niels van Berkel, Chu Luo, Vassilis Kostakos, and Jorge Goncalves. 2017. Challenges of Situational Impairments During Interaction with Mobile Devices. In *Proceedings of the 29th Australian Conference on Computer-Human Interaction (OZCHI '17)*. ACM, New York, NY, USA, 477–481. https://doi.org/10.1145/3152771.3156161

[12] Aditya Vashistha, Pooja Sethi, and Richard Anderson. 2017. Respeak: A Voice-based, Crowd-powered Speech Transcription System. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 1855–1866. https://doi.org/10.1145/3025453.3025640

[13] Aditya Vashistha, Pooja Sethi, and Richard Anderson. 2018. BSpeak: An Accessible Voice-based Crowdsourcing Marketplace for Low-Income Blind People. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 57, 13 pages. https://doi.org/10.1145/3173574.3173631

[14] Omar F. Zaidan and Chris Callison-Burch. 2011. Crowdsourcing Translation: Professional Quality from Non-professionals. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1 (HLT '11)*. Association for Computational Linguistics, Stroudsburg, PA, USA, 1220–1229.