Hey Foxy: Designing a Browser-Based Voice Assistant

Janice Y. Tsai, Jofish Kaye

Mozilla  
[jkaye@mozilla.com](mailto:jkaye@mozilla.com), janice@mozilla.com

Abstract

We present Foxy, an open-source, browser-based voice assistant built into Firefox. We used an HCI-driven research-based approach with a focus on understanding how people are using commercially widespread voice assistants (Alexa, Google Home, Siri, and Cortana) and the impact of the different modalities and form factors (standalone device, mobile, laptop) on the user experience.

Introduction*[[1]](#footnote-1)*

Voice input will be an important part of how people will interact with technology for the foreseeable future. The question of feasibility for speech communication has evolved from science fiction to voice assistants that are a widely adopted, commercial success. These successful assistants have been brought to us primarily by large technology companies and have a variety of form factors, ranging from standalone devices (Amazon’s Alexa, Google Home), to mobile phone and desktop-based agents (Microsoft’s Cortana, Apple’s Siri). Building on the success of these agents, we created Foxy, a browser-based voice assistant for Mozilla.

We embarked on an HCI-driven research-based approach to product and feature definition and user experience design. This data-driven agenda included the log analysis of Alexa History; surveys, interviews, and focus groups to understand the use of existing voice assistants, their failings, and the feature wishes of both users and non-users; and the deployment of an alpha product to understand the use of Foxy in the wild.

Motivation

Mozilla’s mission focuses on creating open and accessible Web technologies. One aspect characterizing the existing voice assistant market is that of the race to create platform-specific “incompatible proprietary fortress[es]” (Rosenberg 2017). We created Foxy as an open-source voice assistant built into the Firefox browser.

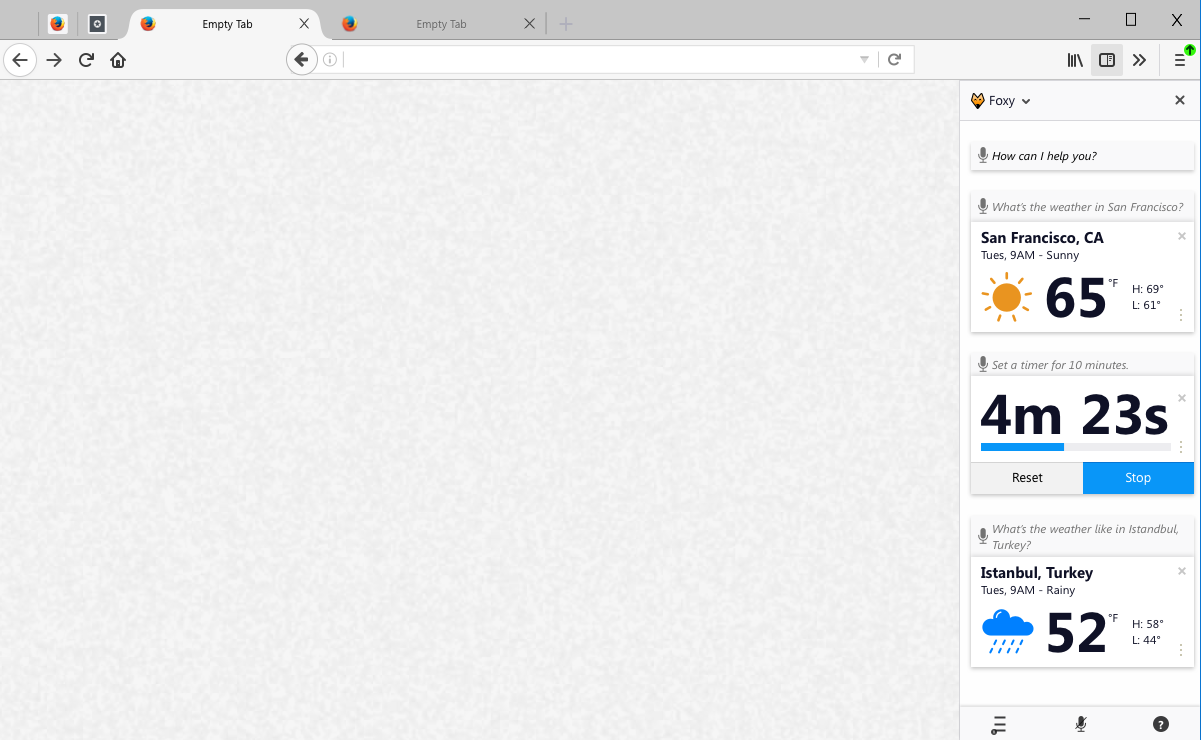
Background

Natural, voice-based interaction with computing systems is facilitated by spoken dialogue systems (comprised of speech technologies, language processing, dialogue modeling) (McTear 2002) with a focus on “humanlike” behavior (Vassallo et al. 2010). Three high level design principles around the functionality of voice assistants center on tasks, conversation, and relationships (Shechtman and Horowith 2003). We explore these areas as we set about defining a new browser-based voice assistant.

Research

We looked to the use of Amazon’s Alexa to understand the types of tasks required of and the conversation with a voice assistant. We conducted a study in July and August, 2017, collecting the Alexa History logs of 82 participants with a total of 193,665 commands (participant mean: 2,176 commands, 1,043 commands). We found that our participants owned 147 Alexa devices (mean: 1.79, median: 1, mode: 1) primarily in households with other people (82.9%), with an average of 2.82 people per household. The average age of the primary Alexa account holder is 31.9 years old, and 41.5% of the households had children under 18 years old (Ammari et al. 2017).

People primarily used their Alexa to play music (33%), to interact with their IoT devices (15%), and to conduct general information queries (14.5%). This information informed the development of the features and capabilities available in Foxy.



Sidebar-based design for a browser-based voice assistant.

User Experience

Building a voice assistant into Firefox afforded us the ability to have a visual user interface. We implemented this in the sidebar of the browser. This visual space allowed us to implement “cards” to provide a response to the user. (At this time, Foxy is unable to speak back to the user.)

Architecture

The Foxy voice assistant is a web extension that runs inside of Firefox. It allows you to interact with it via voice and a browser-based extension sidebar. It uses Snowboy, a customizable wake word detection engine to allow users to create a personal model for wake-word detection.

The Demo

The demo utilizes a laptop with a Google Slides presentation open in Firefox. The presenter walks up to the podium with his hands in his pockets and proceeds to showcase the functionality of the Foxy voice assistant. This functionality includes the following:

* Playing music
* Conducting searches
* Navigating slides
* Setting timers and more.

Conclusion

With the success and availability of voice assistants, a new canvas of research around the interactions, user experiences, and design paradigms for artificial intelligence is now available. We are creating Foxy, a browser-based voice assistant, to explore the use and impact of voice interaction in a web-based environment.

Bios

Joseph ‘Jofish’ Kaye is a principle research scientist working in the Mozilla Emerging Technologies team. He uses a variety of methods, including big data and qualities research to understand user needs and practices in the HCI space.

Janice Tsai is a senior research scientist on the Mozilla Emerging Technologies team. Her research interests are in usable privacy and public policy.

References

Amari, T., Tsai, J. and Kaye, J. 2017. At Home with Alexa. In submission.

McTear.M. 2002. Spoken Dialogue Technology: Enabling the Conversational User Interface. *ACM Comput. Surv.* 34, 1: 90–169.

Rosenberg, S. 2017. “Voice Assistants Aren’t So Easy to Fire.” Wired, Oct. 11, 2017. Accessed Oct. 26, 2017. https://www.wired.com/story/voice-assistants-arent-so-easy-to-fire/.

Shechtman, N. and Horowitz, L. 2003. Media Inequality in Conversation: How People Behave Differently when Interacting with Computers and People. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI ’03), 281–288.

Vassallo, G., Pilato, G., Augello, A., and Gaglio, S. 2010. Phase Coherence in Conceptual Spaces for Conversational Agents. In *Semantic Computing.* Hoboken, NJ: John Wiley & Sons, Inc.

1. Copyright © 2018, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved. [↑](#footnote-ref-1)