# Toward Voice-Assisted Browsers: A Preliminary Study with Firefox Voice

**Alex C. Williams**
University of California, Irvine
Irvine, CA 92697, USA
acw@uci.edu

**Julia Cambre**
Carnegie Mellon University
Pittsburgh, PA 15213
jcambre@cs.cmu.edu

**Ian Bicking**
Mozilla
Mountain View, CA 94041, USA
ibicking@mozilla.com

**Abraham Wallin**
Mozilla
Mountain View, CA 94041, USA
abe@mozilla.com

**Janice Tsai**
Mozilla
Mountain View, CA 94041, USA
jtsai@mozilla.com

**Jofish Kaye**
Mozilla
Mountain View, CA 94041, USA
acm@jofish.com


*CUI@CHI'20*,, April 25, 2020, Honolulu, HI, USA
ACM 978-1-4503-6819-3/20/04.
https://doi.org/10.1145/3334480.XXXXXXX

## Abstract
Web browsers allow people to find, organize, and manage information on the web. While voice interaction research has evaluated the support of web search, the broader role of voice interactions within the browser have yet to be explored at depth. We report findings from a preliminary exploration of the challenges, opportunities, and directions of voice assistants embedded in modern web browsers. We drive our inquiry with *Firefox Voice*, a browser extension that implements a voice assistant into the Firefox desktop browser. Through a think-aloud study (n = 5), we explore the strengths and shortcomings of Firefox Voice to better understand the role that voice interaction can play in supporting people both in the browser and beyond it.

## Author Keywords
Voice assistant, web browser, think-aloud study

## CCS Concepts
•**Human-centered computing** → **Human computer interaction (HCI);** User studies;

## Introduction

Web browsers allow people to find, organize, and manage information on the web. Over the past several decades, a significant amount of research at the intersection of speech technology, computer science, and human-computer interaction has explored pathways for empowering web browsers with voice interactivity. A key direction within this space has focused specifically on understanding the advantages and disadvantages that accompany voice interaction. For example, in a comparison of mouse-based and voice-based interaction, Christian et al. [3] found that voice interactions in the browser not only took longer to perform, but that they were also more cognitively demanding. However, over the past decade alone, modern web browsers have become the center point of the computer, computation has become increasingly mobilized, and advances in speech technology have yielded interactive voice experiences that are increasingly pervasive in peoples' every-day lives.

In this research, we revisit the frontier of voice-assisted web browsers. We report findings from a preliminary think-aloud study (n=5) aimed at understanding the challenges, opportunities, and future directions of utilizing a voice assistant in modern web browsers. We drive our inquiry with *Firefox Voice*, a browser extension that implements a voice assistant inside the Firefox desktop web browser. The purpose of this study was to not only better understand how the tool enables new opportunities fueled by voice, but also how it complements existing practices of non-verbal interactions. We seek to use our observations from this preliminary study to provoke new conversations and questions around voice interactions that take place within the browser and around it.

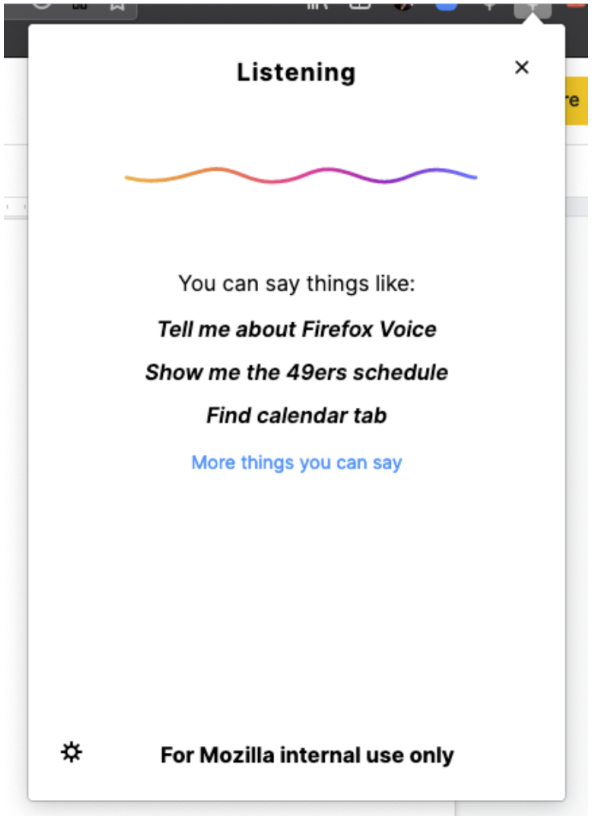
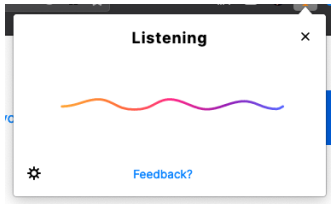
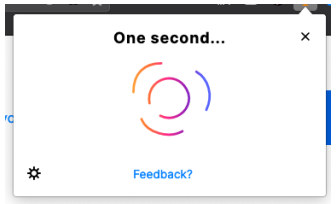



**Figure 1:** Firefox Voice's browser extension interface allows users to issue voice commands directly in the web browser.




**Phase I:** Listening




**Phase II:** Processing


**Phase III:** Acting

**Figure 2:** Firefox Voice's command sequence includes three phases: Listening, Processing, and Acting.

| Theme | Intended Action | Example Utterances |
|---|---|---|
| Bookmark | Open a saved bookmark in a new tab. | *"Open the [TITLE] bookmark for me."* |
| Help | Open Firefox Voice's command lexicon in a new tab. | *"Help." or "Open lexicon."* |
| Music | Open a music player in a new tab. | *"Show [SERVICE] for me"* |
| Music | Play / pause the current song. | *"Play [SONG_NAME] on [SERVICE]."* |
| Music | Command the current service to play the next song. | *"Next track."* |
| Navigation | Open a specific website in a new tab. | *"Go to [SITE_NAME]."* |
| Search | Open a new tab with query results from a search engine. | *"Search [SERVICE] for [QUERY]."* |
| Sound | Mute / umute all browser audio. | *"Mute." / "Unmute."* |
| Tab | Change the browser's focus to the tab with a specific title. | *"Find [TITLE] tab."* |
| Webpage | Start / stop reading a webpage in Firefox's Reader Mode. | *"Read this page." / "Stop reading."* |
| Webpage | Translate a webpage with Google Translate. | *"Translate this page."* |

**Table 1:** A sample of browser tasks supported by Firefox Voice, alongside their actions and utterance examples.

## Firefox Voice

*Firefox Voice* is a browser extension that extends the Firefox desktop web browser by enabling users to issue verbal commands and queries. Here, we describe the end-user experience of Firefox Voice and the types of commands it supported as of the time of the think-aloud study described in this paper, which was conducted in October 2019.

### *Interaction Design*

Users can activate Firefox Voice by clicking an icon in the browser toolbar, or via keyboard shortcut. Once activated, the extension displays a tooltip popup as shown in Figure 2. The popup cannot be invoked via wakeword.

Upon invoking Firefox Voice, the system will enter its command sequence. The first phase is the *Listening* phase in which the interface plays an audio chime to indicate that it is actively listening, and waits for the user to issue a command or query. After the user has issued a verbal command, the interface will display a loading animation to indicate that the interface has entered its second phase of the command sequence, *Processing*. During this phase, the interface has sent the user's command for cloud-based transcription and is waiting for a response. Once the transcription has been received, the interface will enter the third phase, *Acting*, in which it will perform an action within the web browser.

### *Utterance Support and Action Space*

We designed Firefox Voice based on the observed and desired use of voice assistants in prior studies [1]. Based on our own intuition, we also included support for several actions that are specific to browser use (e.g., opening bookmarks). In cases where an utterance was not matched, Firefox Voice would use the transcribed utterance as a query to Google and open the corresponding search results in a new tab. All utterances are matched by a series of regular expressions that support variable word choice. Table 1 shows a sample of Firefox Voice's supported tasks.

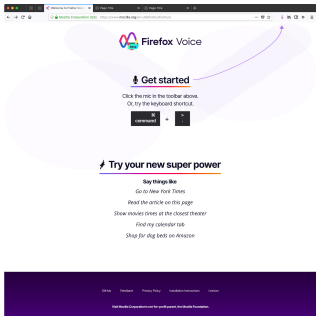



**Figure 3:** The Firefox Voice extension installation webpage.

## Study Design

We conducted a think aloud study with Firefox Voice to better understand the challenges, opportunities, and future directions to explore and evaluate its practical utility. All think-aloud studies were composed of four phases, each of which provides a unique perspective toward our goal.

***1. Introductory Phase:***
Studies began by introducing the researchers and debriefing the participants about the nature of the study, i.e. "to try a new assistant for Firefox". During the introductory phase, the researchers informed participants that they were tasked with "evaluating the tool that a team at Mozilla had created". The overarching goal in doing so was to establish an environment where participants felt comfortable providing honest and unbiased feedback.

***2. Onboarding Phase:***
After the study debriefing had concluded, participants were seated in front of a laptop computer that had the Firefox Voice installation page open in the Firefox web browser. Before touching the computer, participants were asked to walk through how they believe they should install Firefox Voice and were subsequently asked to try to install the extension while describing their train-of-thought. After the extension's installation had been completed, participants were again asked to think aloud and try the process of invoking Firefox Voice. The installation page is shown in Figure 3.

***3. First-Use Phase:***
We onboarded participants into Firefox Voice by asking them to try issuing any one of the suggested commands shown in Figure 1. Once the system had processed the command and reacted accordingly, we asked participants to walk us through their mental model of the system within the context of their uttered command. Here, we allowed participants to respond how they best saw fit, but guided their think aloud process toward characteristics that may affect the user experience (i.e., presentation, timeliness, correctness, and practical utility).

***4. Retrospective Phase:***
The final phase of each study began by asking participants to reflect on the last time they had engaged in a work-related task at their workstation computer. Within this context, we asked our participants to describe the challenges of performing this task and subsequently describe the voice commands they would like to issue to alleviate these challenges. After addressing this question, we asked participants to invoke Firefox Voice and issue the command. To better understand how participants perceive Firefox Voice's practical utility, we concluded the study by administering the System Usability Scale (SUS) [2].

*Method*
Each think aloud study was conducted in the presence of two researchers, one of whom conducted the think aloud with participants while the other managed the audio and video recording. Each study took place either in a corporate meeting room or in a local comic book store in Portland, Oregon. All recordings were transcribed and then iteratively analyzed using open coding and affinity diagramming. Each participant used the same company-owned computer to complete the study.

*Recruitment*
We recruited through online advertisements on the "r/portland" sub-reddit as the researchers were located in Portland, Oregon. The call for participation sought participants who "were familiar with voice assistants" and "interested in try a new voice". The call for participation also facilitated snowball sampling by encouraging readers to share with others who may be interested. There was otherwise no specific criteria required for participation.

## Findings

We recruited a total of five participants (2 M / 3 F) who were diverse in technical expertise and existing practices with voice assistants. Participants job roles include tourism advisor (P1), engineer (P2), podcast developer (P3), retail manager (P4), and construction manager (P5).

### *First Impressions with Firefox Voice*

Our open-ended evaluation of Firefox Voice began by observing how people develop their first impressions about the system at the time of installation. Unlike other voice assistants, Firefox Voice does not come pre-packaged into an existing infrastructure, and we therefore sought to examine how the act of "getting started" influenced participants' perception of the tool and the functionality it supports.

Extension installation was generally met with initial confusion. All participants spoke aloud that the extension was requesting access to the machine's microphone. Several participants expressed privacy concerns at this point during the study. Two participants specifically mentioned that the extension's webpage for installation "satisfied [their] privacy concerns". Three participants noted their confusion was driven by having never installed a browser extension before engaging in our study. All participants noted that the installation process was "pretty straightforward" (P4).

Alongside the installation of the extension, its invocation was met with similar confusion, primarily due to the lack of wake word. One participant attempted to invoke Firefox Voice by speaking aloud "*Hey Firefox*". While the Firefox Voice installation webpage provides specific details on the invocation mechanism, several participants were specifically confused because the invocation required them to click a button that was not within the bounds of the webpage itself (i.e., the browser toolbar). The desire to invoke Firefox Voice via wake word was noted by two participants.

### *Practical Utility and Usability*

In general, participants had no issue envisioning how they might utilize Firefox Voice in practice. Participants' SUS scores ranged from 65 to 80 ($\mu$=70.0; $\sigma$=6.1), suggesting that the system provides an interactive experience that is "above average" [2]. We now provide an overview of the strengths and shortcomings of Firefox Voice per our study.

*Strength #1: Performing Web Search*

Web search was the primary task that participants used when engaging with Firefox Voice. Participants were not only satisfied, but also surprised at how well the tool supported both short queries and complex questions. The successes of Firefox Voice as a tool for search are grounded in its re-use of existing search engines, and our findings reinforce the importance of tool and service integration.

*Strength #2: Managing Webpage Information*

Participant interactions also gravitated toward contextual interactions that occurred within the current webpage. This included translating or reading the webpage. Several related, but unsupported functions included navigating to a bookmark in a content-heavy Google Doc (P4) or summarizing a webpage's content (P1), suggesting managing and navigating the wealth of information in webpages is a promising opportunity for voice interaction support.

*Strength #3: Looking Beyond the Computer*

Though our study was scoped to the laptop context, several participants saw immediate value in using Firefox Voice while away from, or within reach of, their computer. P1, for example, expressed an interest in managing multiple timers while they were within proximity of their computer. Similarly, P5 noted that their interest was stoked by rarely being at a workstation computer to begin with, suggesting that cross-device interaction may be a useful path of exploration for future iterations of Firefox Voice.

*Shortcoming #1: Multi-Turn Conversation*
Participants were generally unsure if Firefox Voice would "talk back" (P1) during their initial interactions. Beyond the scope of our think-aloud study, the lack of multi-turn conversation was noted across participants as a significant deterrent for practical use in niche scenarios. P2, for example, noted that they often ask multiple follow-up questions when using their other assistant technologies for QA-search.

*Shortcoming #2: Configurable Interactions*
Firefox Voice's listening mechanism was configured to transition from Phase I to Phase II of its command sequence after two seconds of recorded silence. We found that participants often needed time to understand how to aurally issue their commands, suggesting the need to support listening timeouts that are configurable or personalized to the user.

*Shortcoming #3: Managing Command Failures*
Command failures in Firefox Voice are the result of speech-to-text errors or by an inability to identify an appropriate utterance for a command. Participants generally believed that the system could have supported command failures more intelligently by "telling me 'Here's what you may have meant'." (P5). Only one participant (P3) utilized Firefox Voice's 'Help' command to browse all supported commands.

## Discussion and Future Research

Our preliminary research provides insight into the frontier of browser-based voice assistants. This work establishes for future research both at finer depth and at larger scale. We currently deployed Firefox Voice in the wild through an internal beta with Mozilla employees. Through this deployment, we are collecting telemetry data through Firefox Voice that allows us to capture user queries alongside contextual characteristics of their interactions (e.g., triggered utterances). We are also exploring pathways for a more formal evaluation of Firefox Voice as a novel system that enables new voice interactions within modern web browsers.

## Participation in CUI@CHI2020

Our interest in participating in CUI@CHI2020 is stoked by our shared interest in the future of a voice-enabled web. Julia is eager to engage in the workshop as an opportunity to meet the CUI community, contribute relevant experiences working with voice, and find opportunities for collaboration. As a postdoctoral researcher, Alex is equally excited to engage with the CUI community to brainstorm and collaborate on studying new conversational interfaces that aid people in living more productive and healthy lives. The Mozilla Voice Products team – Jofish, Janice, Ian, and Abraham – spearheads the Firefox Voice initiative and views the CUI community as part of a larger movement interested in promoting open, private, secure and voice-enabled technologies.

## REFERENCES


[1] Tawfiq Ammari, Jofish Kaye, Janice Y. Tsai, and Frank Bentley. 2019. Music, Search, and IoT: How People (Really) Use Voice Assistants. *ACM Transactions on Computer-Human Interaction* 26, 3 (4 2019), 1–28. DOI:http://dx.doi.org/10.1145/3311956

[2] John Brooke. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.

[3] Kevin Christian, Bill Kules, Ben Shneiderman, and Adel Youssef. 2000. A comparison of voice controlled and mouse controlled web browsing. In *Proceedings of the fourth international ACM conference on Assistive technologies - Assets '00*. ACM Press, New York, New York, USA, 72–79. DOI: http://dx.doi.org/10.1145/354324.354345