

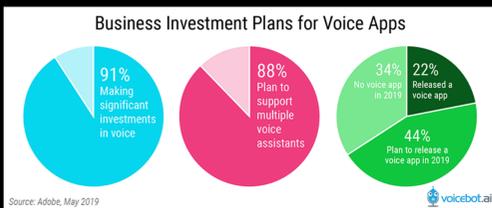
# Beeves

Beeves, the Browser ExtEnSion Voice Enhancement System, is an integrated framework enabling voice functionality in Firefox WebExtensions. It is one of the first attempts to make developing for voice in the browser accessible to developers - eliminating the need for specialized knowledge in voice technologies and natural language understanding. Beeves opens up a world of possibilities for voice-augmented experiences in the browser - a platform that has been limited to traditional keyboard and mouse based GUI interactions until now.

## Why Voice?

# 69%

U.S. adults have tried voice assistants  
(Microsoft Advertising, 2019)



Voice tech is growing explosively in adoption. Most U.S. adults have used voice assistants, and more than half reported using them regularly (Voicebot.AI, 2018). However, voice assistants have been largely absent on web browsers, in part due to lack of software support. As voice tech matures, however, it is inevitable that it should spread to different platforms: *"Phase 1 introduced consumers to the idea of using voice to perform tasks. Phase 2 is about voice becoming a pervasive interaction mode that has more capabilities and is used more frequently across more devices and contexts"* (Voicebot.AI, 2018). This is precisely what Beeves aims to facilitate.

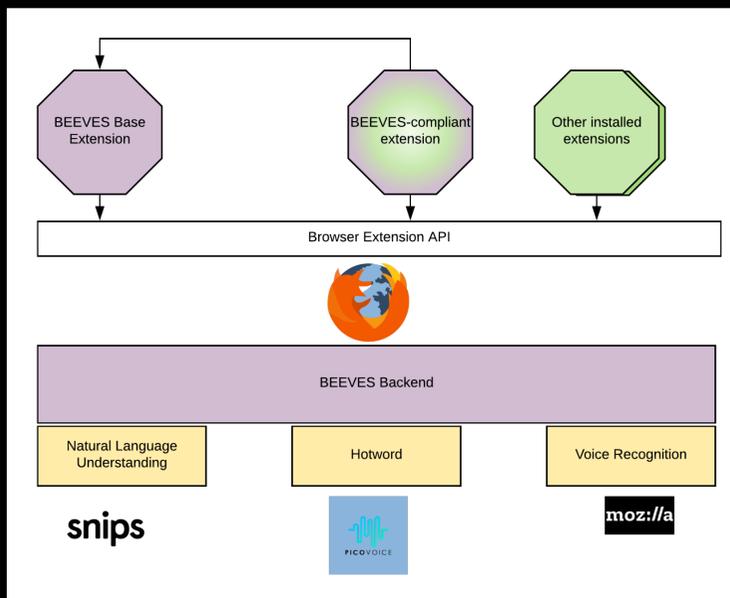
## System Architecture

Beeves consists of two parts - 'Beeves Base Extension' - a browser-side voice agent and skill management system implemented as a Firefox WebExtension, and 'Beeves Backend' which runs natively on the user's operating system.

Beeves Backend is responsible for wake word detection, transcribing speech to text, and parsing the intent being expressed by the user, to be consumed in the browser by a Beeves-compatible WebExtension.

Beeves Backend relays this information to the Beeves Base Extension using Firefox's Native Messaging functionality. Beeves Base then determines which WebExtension should respond to the user's voice command, and dispatches a message accordingly. The Beeves Base Extension also facilitates user-facing graphical feedback and dialogue management.

## Beeves Components



## Mozilla

Beeves is one of the first applications to bring together on the browser the latest technologies from Mozilla's labs. These include DeepSpeech, Mozilla's deep-learning-powered speech recognition engine; the Cloud Speech API, which provides an online alternative to DeepSpeech; the Rust programming language; and WebAssembly, which allows the browser to execute code written in Rust and other languages not previously supported on the browser.

Additionally, the open source, privacy-first architecture, which allows Beeves to be used without any data being processed or stored on external servers, complements Mozilla's strategy and vision of facilitating an open, accessible web on which privacy is a right for everyone.

## Users

By adding a new input channel, Beeves unlocks new possibilities for interaction that can augment the venerable mouse- and keyboard-driven desktop graphical user interface.

A typical use for Beeves would be to simplify input that might take several steps that need to be done outside the current context the user is working in. E.g. copying and pasting.

Because Beeves uses open source, privacy-first components that make it possible to avoid transmitting voice data or transcripts off the user's personal device, it may alleviate users' top concerns over voice assistants: doubts about the privacy of their data and how it is used.

## Developers

Beeves provides a simple API that integrates with existing WebExtensions with minimal development effort. Developers specify metadata pertinent to their voice skills using a metadata schema presented as JSON.

While the individual components necessary for voice assistants may be used on the web already, there isn't a unified framework on the browser extension side that can manage multiple "skills" at once can be used on the browser.

These components require specialized knowledge and considerable integration work on the part of the developer. By taking care of these issues, Beeves gives developers more time to focus on the code that matters most to them.

## Enterprise

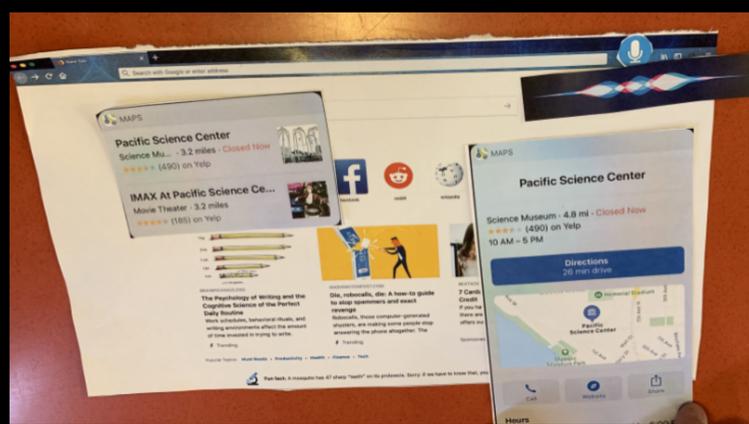
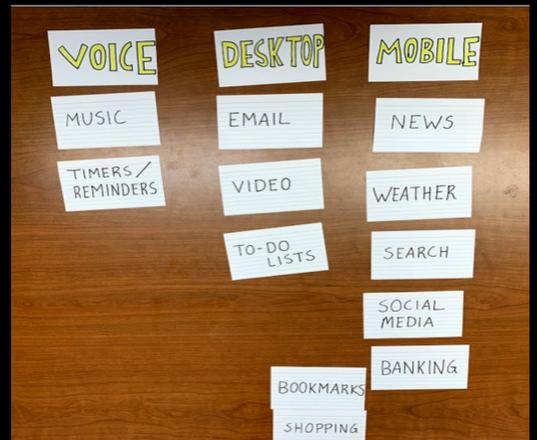
Browser extensions can add value to any kind of web-accessible platform. Of the 20 most popular iOS apps in 2018, 16 had counterparts in the form of browser extensions (Apple, 2018).

As voice drives growth and innovation in industries from automotive to retail to HR, the vast majority of U.S. companies report they will be investing considerably in voice tech over the next few years. (CMO by Adobe, 2019). By bringing voice to the browser itself, Beeves enables feature parity across the range of cross-platform apps.

## User Research

Complementing the Mozilla Foundation's voice UX study (Tsai and Kaye, 2018), our research questions included: why do users prefer using voice assistants for some activities over traditional means of interaction? What kind of ideal voice assistant do users envision to augment their browser experience? We surveyed people about their experience using VAs such as Siri, Alexa, and Google Home. We selected six of these participants based on quotas covering a range of VA use frequencies, and conducted 30 minute interviews with them.

These interviews consisted of card-sorting tasks to understand which modes of interaction (desktop, mobile, and voice) were preferred for which activities and why. We focused our line of questioning to determine which contexts provided the best use cases for voice interactions. We also asked users to design their ideal voice assistant using provided paper cut-outs placed over a desktop or mobile browser screens.



From results we learned of user's desired features and conceptual preferences. Desired form factors varied, but a visual voice agent providing multiple forms of output is ideal. Users want a flexible, customizable system they can easily learn and trust to work in well-defined contexts. We've inferred they need an integrated system that mitigates context switching to provide a seamless experience during a single work session. Users prefer different modes of interaction based on context. By incorporating the advantages of mobile and voice modes on desktop, we can mitigate the need for context switching.

References  
Apple. (2018). *Get the top apps of 2018*. Retrieved from <https://itunes.apple.com/story/id1438574124>  
CMO by Adobe. (2019). *State of voice technology for brands*. Retrieved from <https://www.cmo.com/features/articles/2019/5/6/adobe-study-voice-commerce.html>  
Microsoft Advertising. (2019). *2019 voice report. Consumer adoption of voice technology and digital assistants*. Redmond: Microsoft. Retrieved from <https://about.ads.microsoft.com/en-us/insights/2019-voice-report>  
Tsai, J. Y., & Kaye, J. (2018, March). *Hey Scout: Designing a Browser-Based Voice Assistant*. In *2018 AAAI Spring Symposium Series*.  
Voicebot.AI. (2018). *Voice assistant consumer adoption report 2018*. Retrieved from <https://voicebot.ai/voice-assistant-consumer-adoption-report-2018>



Proshonjit Mitra  
[www.proshonjitmitra.com](http://www.proshonjitmitra.com)  
[ptmitra@uw.edu](mailto:ptmitra@uw.edu)

Aditya Chatterjee  
[github.com/adityachatterjee42](https://github.com/adityachatterjee42)  
[achat@uw.edu](mailto:achat@uw.edu)

Altan Orhon  
[github.com/orhmeh09](https://github.com/orhmeh09)  
[altan@uw.edu](mailto:altan@uw.edu)

Richard McGovern  
[github.com/richardwmcgovern](https://github.com/richardwmcgovern)  
[richardm@uw.edu](mailto:richardm@uw.edu)