AudioBrowse: Final Report on an Auditory Browser Prototype

Group5: Sally Lee, Hollis Liu, David Mcelveen, Ann Stowasser, Mason Strohl

18 April 2019

Table of Contents

Executive Summary

Currently, web browsers are a major part of people's everyday lives and have had a large impact on society. Web browsers continue to grow and change rapidly, however they are still not suitable for a complete auditory web browsing experience. The CPSC 4180 group 5 project team has evaluated the current state of web browsers and designed an auditory web browser that we envision to be implemented 20 to 50 years in the future. Once we had developed our plan and completed the prototype, our next steps were to evaluate the usability, privacy, and design aspects of the prototype.

Our group conducted a usability test on Tuesday, April 2, 2019 and Tuesday, April 16, 2019 in room 107 of McAdams Hall at Clemson University. The purpose of the test was to evaluate the usability and security of the project prototype, *AudioBrowse*. There were four users recruited to participate in the first session and two users that participated in the second session.

For the first session, all of the participants completed a Google Chrome browser test, prototype test, and a follow-up user questionnaire. Three of the users also participated in a secondary capacity. Their second role was to act as a spy to test the privacy and security aspects of the prototype's design. The users that acted as spies had an extra follow-up questionnaire to complete. Each user overall spent approximately 30 minutes for their session.

For the second session, the participants completed a Tor browser test, prototype test, and a follow-up user questionnaire. However, neither of the participants acted as a spy in this session. The changes for this session were made based on the results from the first session in order to collect better data.

The results of the usability test showed that our prototype comes close to the state of the art browsers currently in existence, but slightly lags behind in terms of performance measurements.

Once we evaluated the results, we were able to make recommendations for prototype improvements and how to make the browser more usable and secure to optimize the user's experience. Eventually, our browser will be able to surpass the experience that current browsers provide.

This document contains the introduction to the project topic, evaluation of users and their needs, a description of our prototype, the methods to evaluate the prototype, results of the usability testing, and recommendations for improvements. A copy of the questionnaires are included in the Attachments' section.

Introduction to Domain and Problem Space

A usability test is intended to determine the extent an interface facilitates a user's ability to complete routine tasks. Typically the test is conducted with a group of potential users either in a usability lab, remotely (using e-meeting software and telephone connection), or on-site with portable equipment. Users are asked to complete a series of routine tasks. Sessions are recorded and analyzed to identify potential areas for improvement to the web site.

AudioBrowse provides the user with a rich and easy-to-use experience interacting with web content. Through rich and concise visual feedbacks, *AudioBrowse* provides user with instant glanceable content. While audio feedback guides the user through the content with a personal touch. Behind the scene, a natural language processing model and contextual conversational model prototype is used to tea natural human voice input as the fastest and most efficient input method.

A series of usability test is designed to measure the efficiency, accuracy, and security of the prototype. The tests are designed to measure how fast, how accurate (given limited time) and how secure the prototype is compared to current state of the art browsers. The tests were conducted onsite. Participants were asked to complete a series of tasks while their actions were recorded using timing devices. Qualitative questionnaires were also handed out to the participants for subject feedback on the prototype.

The group members conducted the usability test on a laptop connected to an external tv display at a lab setting. Two laptops were used for different versions of the bowser. One group member guides the participant through the testing procedure while another member times his/her actions during test. An online questionnaire is completed on participants' own laptop machines. Overall, session duration, task completion (accuracy), and questionnaires answers were recorded to data analysis.

Evaluation of user needs

User Group, Demographics, and Characteristics

The demographic of this project is rather broad. In the future when AI has progressed to the point of conversation, the implementation of this AI into a web browser can thusly be any current web browser user. Due to the scope of the project, and the uncertainty of how the internet may be accessed fifty years in the future, we must limit our demographic to fit within the scope of this project. Keeping the average internet user in mind, we focused on visually and audibly impaired users. The National Federation of the Blind gave a report regarding the blind and visually impaired in the United States in 2016. At the time, total blind among all ages, ethnicities, and education levels reached 7,675,600, 2.4% of the 2016 United States population. Of that number, 4,037,600 are between the ages of 16 and 64; 3,171,100 are 65 and older [1]. In another study published in May of 2016, The National Institute on Deafness and other Communication Disorders reported that nearly 18 million American adults aged 18 and older claimed to have a problem with their voice in the past 12 months. In addition, more than three million Americans have a speech impediment known as a stutter. While the problem commonly fades into adulthood, the stutter can persist throughout life [2].

Identifying the visually impaired demographic is important to this project because this is the first group of people we would expect to adopt an auditory browsing technology, as traditional browsing methods have barriers to blind users. Four barriers have been identified in regards to blind users and web browsing: the pixel barrier, mouse barrier, graphics barrier, and the layout barrier. The pixel barrier refers to the fact that screen output itself is stored as a pixel map, so information on the screen, after it has been sent to the screen buffer, can no longer be read by traditional text-to-speech tools. Rather, information must be encoded elsewhere, such as html tags, before being sent to the screen buffer in order to be utilized by text-to-speech tools [3]. Traditional web browsing is traversed through mouse movements and clicks, which encompasses the mouse barrier, as blind users cannot see the mouse move, therefore losing out on information being displayed due to a lack of feedback. The graphical barrier focuses on how the information is displayed on the screen, "...and that topography and topology give additional hints about the represented objects and their relations" [3]. Essentially, one object on the screen and its location in relation to other objects matter when accounting for the graphical aspect of web browsing and that nuance can not be properly conveyed to blind users through traditional web browsing. Lastly, is the layout barrier, which refers to the semantic purposes of various layout features and designs [3]. These four barriers give us an idea for what the user needs in order to properly traverse the internet using an auditory browser, as each of these four barriers must be addressed within our prototype in order to properly convey all aspects of traditional internet browsing via auditory cues.

Identification and Evaluation of Existing Technology
In the present day, there is no true audio-based browser. However, it is important to note the prevalence and increasing popularity of feature specific browsers. For example, Tor is a popular browser catering to users who value privacy and security over most other browsing features. Similarly, Coowon is a browser for web-based gamers and Torch is a browser for heavy media users. The significance of this is the implication that there is an ever present and continuous popularity of certain feature-centric browsing, which is directly applicable to a future browser that is completely audio-based.

The scope of usability of current audio-based browsing technology, however, despite sounding more human-like with every passing year, it remains highly limited. For the most part, the existing technology in question is restricted to virtual assistants and screen readers that are no more than simply plugins and extensions of existing browsers.

Present-day virtual assistants, like Siri, Cortana or Alexa, while being capable of producing a large variety of dynamic infographics based on very particular search queries, cannot effectively dynamically browse the web in the same fashion or to the same efficacy as the average user needs. Virtual assistants also do not overcome the four barriers presented in the study by Donker, et. al. In effect, while modern virtual assistants certainly use artificial intelligence and machine learning, it is not nearly to the extent that is necessary for a complete auditory browsing experience. Currently, text-to-speech processing done through machine learning models is common. The choices of machine learning models are myriad, with decision trees being the most basic type,

ranging all the way to neural networks, statistical regression models, and support vector machines (SVM). An empirical study on ten different machine learning models: SVMs, neural nets, logistic regression, naive bayes, memory-based learning, random forests, decision trees, bagged trees, boosted trees, and boosted stumps, was done to compare accuracy and efficiency of predictions. The results of the study showed that neural networks gave the best average score, tied with bagged trees and random forests. The researchers used a normalization technique in order to calibrate the results, as some of the ten models studied are intended to give a probability of the various possible results, while other models, like SVMs, are not, giving a single "best" answer. These calibration techniques include Platt's Model and Isotonic Regressions. The researchers compared results before and after calibration and found that neural networks produced the best results before calibration [4]. Further, just like there are many types of machine learning models, there are variations in the types of neural networks that can be used. Microsoft's 2017 Conversational Speech Recognition System uses recurrent neural networks. Essentially, a recurrent neural network uses previously computed data in combination with its current state in order to better predict the next result, such as the one in Figure 1.
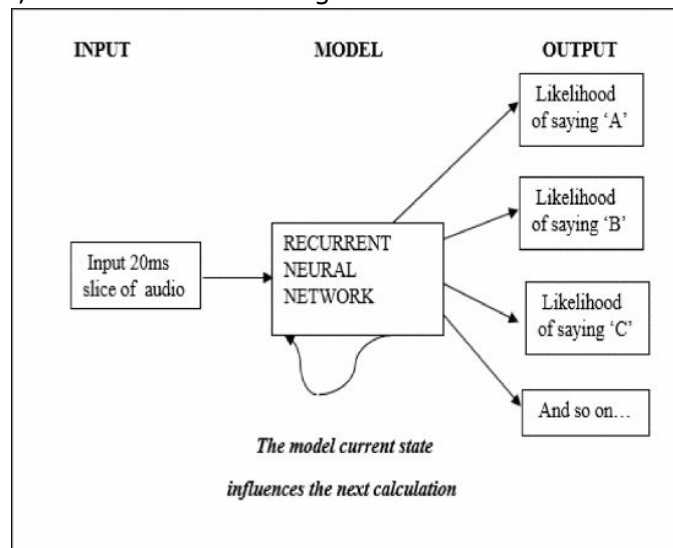

Figure 1. Recurrent neural network flowchart

The processing of user speech is done in two parts under this system. The first uses a convoluted neural network with a bidirectional long short term memory (CNN-BLSTM) in order to train the acoustic model. Additionally, Microsoft researchers found that applying an LSTM to the language model in the second step of speech processing allowed for better results than just the CNN-BLSTM acoustic model [5].

Similarly to voice assistants, current screen readers are just as limited. Much like virtual assistants, screen readers cannot truly convey images; currently, the only way screen readers get around this issue is by reading certain text in the page, frequently only present specifically for screen reader use, that operates as a substitute for said image. This is known as alternative text, and if it doesn't exist for the image, the screen reader cannot describe the image. This alternative text information is encoded into the webpage's HTML in "alt" attribute, inside the "img" tag [6]. Additionally, screen readers cannot convey

the web page as a whole the same visual way a human would. This simply means that a "visual user" can look at a web page and focus on the content important to them, accomplished by possessing basic understanding of web page organization as well as being cognizant of the informational sections, thereby ignoring advertisements and extraneous menus [7]. This is a stark contrast to how a typical screen reader linearly "reads" over a page, one word at a time, from left to right and usually failing to astutely omit peripheral or irrelevant content such as navigation bars, menus or advertisements.

Perhaps the closest available technology to a complete auditory browsing experience is a screen reader plugin to the Google Chrome browser called LipSurf. This extension navigates through the web by using certain auditory keywords and phrases like "Scroll Bottom", "Scroll Up a Little", "New Tab" or "Mute". For utilizing links, *LipSurf* offers a unique solution. The "Annotate" command visually assigns all potential links on the current web page with a unique combination consisting of a letter and some numbers that the user can then say to follow the link as shown in Figure 2. While being more effective than present day alternatives, it is, subjectively, quite unaesthetic, and far-flung from the efficacy of normal point and click browsing.
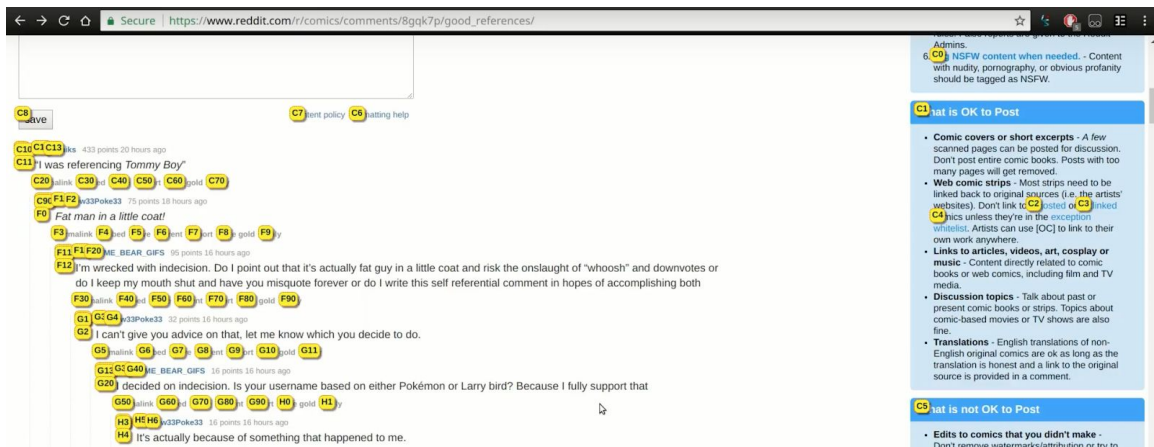

Figure 2. LipSurf Plug-In Screenshot

Prototype Description

Our prototype, *AudioBrowse*, was made by taking audio input from a user, translating it into a request through artificial intelligence, then returning visual feedback plus a conversational audio response to the user's inquiries. We imagined a "layered based UI system" that is in contrast to the current state of the art cursor based UI system. The prototyped layer system does not have any buttons or hyperlinks that needs to be aimed at. The navigation of the prototype usage stems from a root layer and goes into detail views step by step. This addresses the mouse barrier, since the user does not need to interact with objects on the screen, rather they interact with the AI, which in turn manipulates the information displayed on the screen. At any point of during usage, only the most essential elements are displayed so there is limited information displayed at a time for a user to parse. The purpose of our approach is for the user to navigate information easily in a conversational context.

The design of the UI is based on the layer system concept and intended to be fairly bland, with mostly information displayed textually on the page. This is in an effort to combat the layout and graphical barriers of web browsing. For example, in Figure 1, the user interacts with the LogSec menu. While this menu contains "button" outlines and a header for non-blind users, the specific location of the information on the screen is irrelevant, as there is nothing outside of text being conveyed on the screen, thus no information is lost when read aloud by the AI.

The prototype for *AudioBrowse* consisted of a single display setup (or HUD) on a monitor controlled by the team member acting as the AI using the Wizard of Oz technique. As a participant would interact with the browser, the controller would navigate through pre-generated photoshop images of the given tasks and follow a script for the audio feedback. In the initial tests the controller was visible to the participant but for the additional tests, the controller wasn't visible.

Our prototype simulated possible interactions a user could have with an audio browser through tasks such as searching for an item, purchasing an item, logging into gmail, and searching for the weather. Below are a few of the pre-generated photoshop images used in our prototype with the scripted audio feedback.
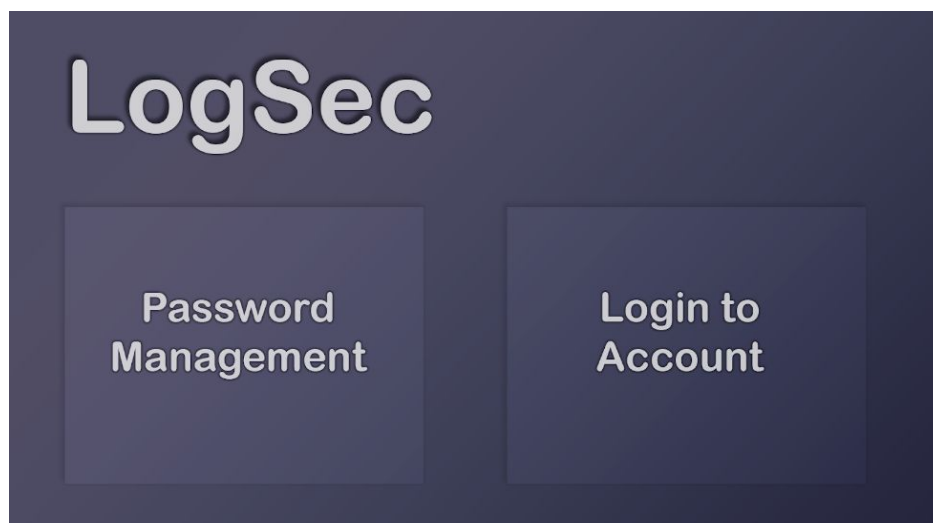
Figure 3. LogSec Prototype Password Manager

LogSec was the password manager for our prototype and would allow users to store new account information or login to a preexisting account. The interaction between the participant and AI would take place as follows:

"I'd like to set up a new account account."
Welcome to LogSec, your password manager. What account would you like to store credentials for?
"Gmail"
What is your username for gmail?
"test at gmail.com"
What is your passphrase for gmail?
"HorseJumpingYellowBoat"
Please confirm your passphrase.
"HorseJumpingYellowBoat"
Login created.

Figure 4. Searching and Results

The generic search screen of our prototype instructed the participants to say a search and would then describe different options and render results they could select from.

> "Search for Clemson University Car Decals."
> *Searching for Clemson University Car Decals … Here are your results. Select 1 for a chrome Clemson tiger paw decal, Select 2 for an orange Clemson tiger paw decal, Select 3 for an orange Clemson tiger paw palmetto tree decal, or select More for more options.*

Figure 5. Purchasing Items

After an item was selected to purchase, *AudioBrowse* pulled up a page with more info on the item and the option to order the item or return to search.

> "Select option 5, Sparkle Paper Towels."
> *Here are the results for Sparkle Paper Towels. Buy 18 rolls for $27.54 on Prime or $28.99 standard. Would you like to order these paper towels, return to the search, or get more information?*
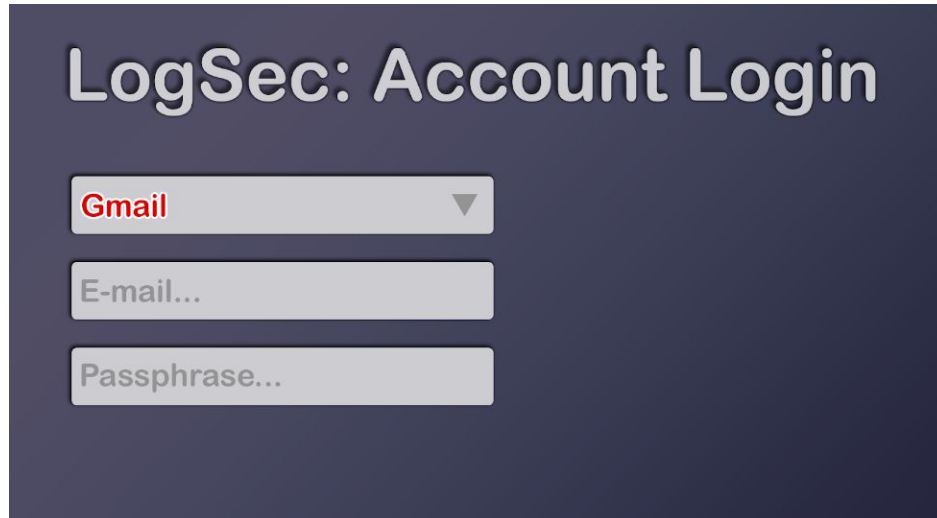
Figure 6. Logging into gmail

To log into an account with pre-set up credentials, LogSec grabs this information and redirects the user.

"I'd like to log in to gmail."
What's your email address for gmail?
"test at gmail.com"
Loading passphrase from LogSec … Redirecting … Welcome to gmail!

Methods used to evaluate prototype

Usability Test Overall Design

The usability test is divided into three core sections: a) Efficiency test, b) accuracy test, and c) Privacy test. **Efficiency test** measures the time it takes for each participant to complete all tasks on a action script on our prototype and current state of art browsers. **Accuracy test** measures the number of tasks each participant complete within a 2 minute time limited period based on the tasks listed on the action script on our prototype and current state of art browsers. Lastly, **privacy test** is carried out by placing active and inactive third party listeners (spies) close to normal participants and questioning them about the content they overhear during the test session.

Detail description of the usability test setup is listed in the sessions section below.

Sessions

According to an article by the Nielson Norman Group, "testing with 5 people lets you find almost as many usability problems as you'd find using many more test participants."[8] Based on this information, we chose to use 6 participants in order to make sure that we had enough participants in case someone did not show up, or there were scheduling conflicts that could not be rectified.

The group members used different methods in order to recruit participants. Some of the members recruited close friends or family members. Other

members recruited classmates that were previously aware of the project topic. However, all of the participants were chosen based on a small set of predetermined criteria. We determined that our participants should be within an 18-26 year age range, healthy, and with varying levels of computer literacy.

There were two separate sessions; each was conducted at McAdams Hall on Clemson's Campus, in room 107. The first session was conducted on Tuesday, April 2nd and contained four of the six chosen participants. While all arrived to the testing location at the same time, their start to our test was staggered. This means that each participant spent approximately 30 minutes participating, though the whole session lasted around two hours. The second session was conducted on Tuesday, April 16th. Two people participated in this session, which was also located in McAdams room 107.

During session one, four participants were sent through a rotation. Two participants waited outside the room with David, so that they could be briefed on the project topic. One was inside the room acting as a spy, while the other actively participated in the test. After each test was completed, the spy would move to the prototype test, while the previous prototype tester moved outside to take the Chrome usability test. One of the participants continued to stay outside with David while the other moved inside to be the next spy in the rotation. This rotation continued until three of the participants had acted as spies with no prior knowledge of the prototype and all four of the participants had acted as testers in our prototype and Chrome usability tests. During the prototype usability test, the participant interacted with a TV that displayed our demo. The AI was voiced by Sally, while Hollis kept time of the tasks to be completed. All three individuals, the two group members and the one participant, were stationed at one table, while the spy was stationed at another table across the room, within hearing distance. The Chrome usability test was done outside with Mason timing the test and giving instruction as needed.

In the second session, the final two participants only participated in the usability tests, not the privacy/security test. The prototype usability test contained the same tasks as before, but with some adjustments to the session setup. In the first session, Sally who was the voice of our AI, was sitting, we moved her out of the room and had her voice come through speaker phone, to better simulate the AI interaction with the participant. The second test also contained the same task, but had a change to the testing environment. Rather than testing on Chrome, a highly optimized and commonly used web browser, we decided to use Tor. We chose Tor because of the high latency associated with using the browser. Another reason for choosing Tor is its relative lack of adoption among the average internet user, with around 2,000,000 direct connections to the Tor network in the first two weeks of April, the highest number of connections made so far in 2019, according to Tor Metrics [9].

After the session, all participants were asked to fill out a usability test questionnaire. In addition, the three participants that acted as spies for our privacy/security test also filled out a spy questionnaire.

See Attachment A and B for the questionnaires.

Participants

All six participants were within the 18-26 year old age range with varying levels of computer literacy.
Of the six participants, three were male and three were female. All six participants have attended Clemson, while one has already graduated.

Evaluation Tasks/Scenarios

When a test participant acted as the spy, they were given the task of listening as the participant performed their list of tasks. They were told to write down any information they found sensitive from the user of *AudioBrowse*.

To test the usability of our prototype *AudioBrowse*, participants attempted completion of the following tasks on both the prototype and a traditional browser (Chrome for the first session and Tor for the second session):
- Open the *AudioBrowse* home page
- Setup Gmail credentials
  - Open the password manager LogSec
  - Input (test@gmail.com, **h**orse**j**umping**y**ellow**b**oat)
- Search for clemson car decal
- Return to search tool
- Search, then order paper towels for sale on amazon
  - Search for paper towels
  - Buy Sparkle paper towels
- Log into Gmail with the test@gmail.com account
- Search for weather forecast at Clemson SC for the next five days

Results

Introduction

This section includes all raw data we gathered for all of our usability tests. Related analytics plots are also included. Detail analytics on the data is included in the "Implications of Evaluation" section.

Efficiency Test Result

Participant	Task 1 (s)	Task 2 (s)	Task 3 (s)	Task 4 (s)	Task 5 (s)	Total (s)
1	7.16	14.69	6.95	85.81	10.91	125.52
2	6.76	17.98	19.46	25.63	32.17	102
3	6.93	23.81	20.24	31.31	11.55	93.84
4	5.76	13.86	21.54	21.37	23.63	86.16
5	20.36	12.41	12.02	30.63	27.43	102.85
6	15.01	18.21	27.11	37.75	59.93	158.01
Average	10.33	16.82	17.887	38.75	27.603	111.39

Table 1. Efficiency test raw data result on current browsers

Participant	Setup (s)	Task 1 (s)	Task 2 (s)	Task 3 (s)	Task 4 (s)	Task 5 (s)	Total (s)
1	55.16	82.29	38.94	16.81	30.04	9.54	232.78
2	43.28	27.39	16.78	26.84	18.74	19.42	152.45

3	50.21	23.12	15.07	16.03	13.59	16.09	134.11
4	40.44	17.02	15.39	30.07	19.03	20.45	142.4
5	26.13	21.77	20.8	18.67	17.61	13.22	118.2
6	57.38	35.41	15.13	28.06	17.62	11.2	164.8
Average	45.43	34.5	20.36	22.74	19.43	14.98	157.4

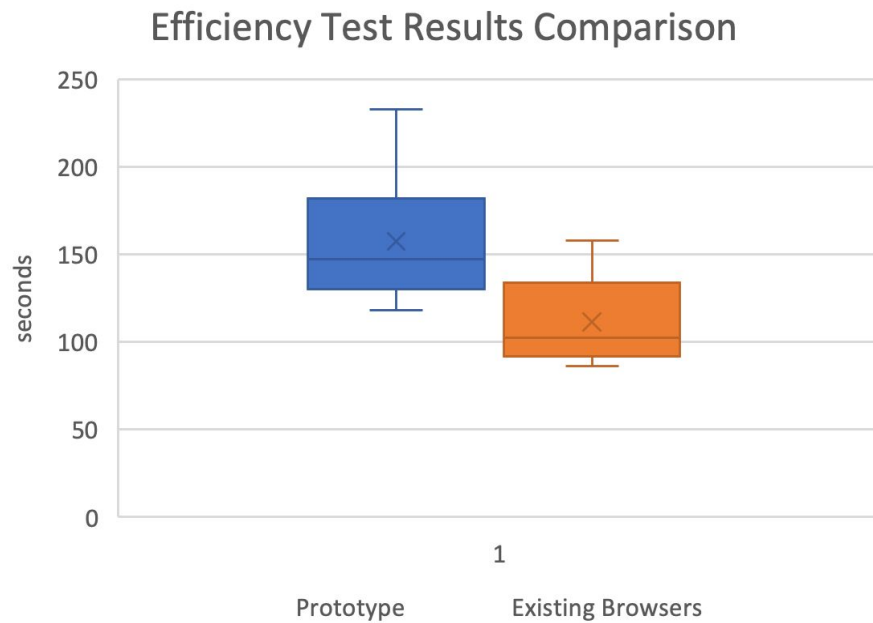**Table 2. Efficiency test raw data result on prototype**

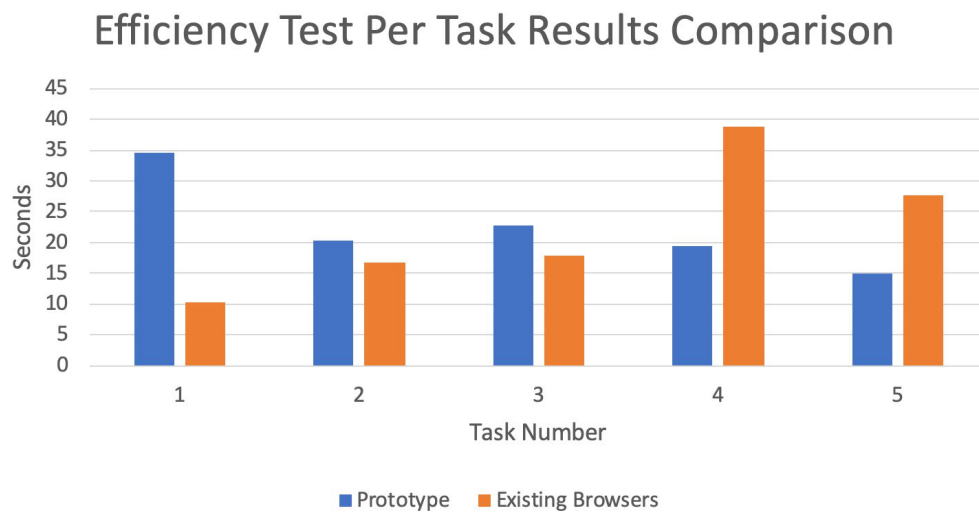**Figure 7. Box plot of efficiency test averages**

Figure 8. Bar chart comparing seconds consumed per task (without password setup time)

Accuracy Test Result

Participant	task 1	task 2	task 3	task 4	task 5

	task 1	task 2	task 3	task 4	task 5
1	√	√	√	√	
2	√	√	√		√
3	√	√	√	√	√
4	√	√	√	√	√
5	√	√	√	√	√
6	√	√	√		

Table 3. Task completion result on the prototype

Participant	task 1	task 2	task 3	task 4	task 5
1	√	√	√	√	√
2	√	√	√	√	√
3	√	√	√	√	√
4	√	√	√	√	√
5	√	√	√	√	√
6	√	√	√	√	√

Table 4. Task completion result on the current browsers

Current Browser Comparison Result

Browser	Prototype	Chrome	Tor
Time average(s)	157.46	101.88	130.43

Table 5. Overall time average comparison between our prototype, Chrome and Tor browser

Post-Questionnaire Qualitative Results
In order to assess the usability of our prototype *AudioBrowse*, we had participants answer post questionnaires from both the perspective of spies (three people) and users (six people).

Spy Results
We first asked the spies if they actively listened to what the participant was doing, and found that two actively listened for information, while the other only picked up information passively. Both actively listening spies felt it was very easy to hear the tasks being performed by the user and the other spy felt it was relatively easy to hear. All three spies were able to identify the email address and the items searched for from the user. These results indicate that the security of *AudioBrowse* can be improved in a way that the user doesn't have to speak their commands in a normal volume.

User Results
Once the users finished the prototype testing, they completed the user questionnaire with their feedback of the prototype's usability and design. All of the users felt that the prototype was moderately to very easy to use. The majority were also satisfied with the design and would not make any changes. However, there were some mixed results for other aspects of the prototype.

Relating to the security and privacy features of *AudioBrowse*, two thirds of the users liked the way in which they entered the email and passphrase, however one third of the users did not like it. There were mixed results when asked how they would feel about inputting their real email address and password. Four of the participants said that they would not like it, one said that he/she had neutral feelings, and one said that he/she would like to input their real email and password.

Relating to the usage of *AudioBrowse*, two thirds of the users had neutral feelings towards the AI's verbal response, where as one third of users stated that they liked the verbal response. Half of the participants said they liked *AudioBrowse* and would use it in the future and half of them said they would not. Of the participants that said they would use *AudioBrowse,* the majority specified that they would use it at home. The users that said they would not. Overall, these results indicate that there were mixed feelings from the users towards the prototype. Some improvements should be made to better satisfy more users.

If no, explain why you wouldn't use it.

3 responses

I would prefer the traditional browser where I won't have to talk just because I feel like I have to wait til the AI is done talking to make my selection

I am extremely paranoid, and would not ever be comfortable with having my information saved in an audio format.

I like the idea of it but the security of having a device always listening bothers me

Figure 9. Example responses to usability questionnaire

Implications of Evaluation

The main measurement of the prototype efficiency test is time characterized in seconds. The raw results are shown in Table 1 and Table 2.

Figure 7 is a box plot on the moment statistics on the averages of efficiency test. As one can easily seen, current commercial browsers overall takes less time to complete the tasks. Meaning our prototype is not as efficient compared to current ones. It can also be observed that our prototype has a larger range and therefore a higher standard deviation. This might be the result of a steeper learning curve of our prototype and variance in participant's ability to adopt new systems.

Figure 8 is a bar chart comparing the time participants took for each task, without the pre test setup. This chart provides a better look into the reason why our prototype takes longer to finish the tasks. It turns out our prototype outperforms current browsers in 2 of the 5 tasks but trails behind significantly in the other 3. This statics provides us with a clearer direction on how to improve specific features.

According to Table 3 and Table 4 listed the completion status for each participants under the 2 minutes constraint. It is apparent that every one of the participants finished all tasks with 100% accuracy. As for our prototype, one participant skipped a task and two others ran out of time. The result

implies that our prototype can induce error when the user is under time pressure.

The time comparison across the browsers according to Table 5 reveals the differences in efficiency because of browser design. Tor takes longer because its privacy first design. As a conclusion, privacy comes with cost in efficiency. If we want to incorporate privacy features in our prototype we project the time duration will be significantly longer as well.

Additional finding: even though the overall time recorded for the efficiency test exceeds the time for current browsers, there are more details to be investigated: If we deduct the time of password manager setup from the overall time, the total time becomes virtually the same (**112.0 seconds** for our prototype vs **111.4 seconds** for current browsers). This implies our prototype actually matches the efficiency of current state of the art browsers after the first run.

Recommendations

The following recommendations are proposed based on the overall data analysis based on the usability test results. We revisited our design after gathering the results and came up with recommendations on two main topics: a) improving the security of *AudioBrowse* and b) reducing usage time. The detailed recommendations are listed below:

Change	Justification	Severity
• Allow user to have a traditional password manager to type login credentials • Allow user to whisper commands or speak in a quiet voice to control the AI • Add directional sound options or headphones to limit the AI response to only be heard by the user	According to the results of the Post-Usability Test Questionnaire, participants didn't like vocalizing their password credentials. They also didn't like how public their searches were. Spies were also able to identify all of the topics participants were asked to search for, showing that privacy measures needed to be put in place to limit a malicious person from getting information.	High

Table 6. Security Improvements for our prototype

Change	Justification	Severity
• optimize the audio feedback result based on the specific content on screen • Allow user to toggle AI response on/off.	According to the Bar chart on usability, it takes significantly longer time than other tasks This will address the concerns users had with having to wait for the AI to finish responding before carrying out their next task.	High

Table 7. Reducing overall usage time for our prototype

Limitations

For this usability study, we were limited in our target demographic, the use cases of Audio Browsers, the Wizard of Oz technique, and comparable models of audio browsers.

Currently, over 4 billion people use the internet [10] and ideally in 20 plus years, everyone is a potential user of audio browsing. However, for the purposes of this class we limited the usability study to participants age 18 through 26 with varying levels of computer literacy. Our study participants were also limited to people who were not disabled or experiencing trouble with their voices. Thus, all conclusions and implications of our usability study for *AudioBrowse* may not represent the entire web browsing demographic, nor will it accurately represent results for those who may be visually impaired or blind.

Another limitation of our study was in the use cases we presented each participant to perform on our prototype. For the scope of this class, we limited our prototype to be capable of performing a set amount of tasks common users perform such as checking the weather, searching for items, making a purchase, and logging into an email account. We did not use our prototype for other interactions such as getting directions, searching for clothing, reading emails, browsing YouTube content, etc. so we cannot judge the usability of our prototype for these actions. Also, after testing we found that participants felt neutral about having to wait for the AI vocal response to finish before moving on to the next task. If someone were to actually use an Audio Browser in the future, the interaction would be different than that tested (a participant sitting in front of a screen) and would more likely be an interaction where the user welcomes a vocal response, not always looking directly at the results.

The Wizard of Oz technique didn't fully represent what an actual Audio Browser would look like. In testing our Audio Browser using the Wizard of Oz technique, we also performed some of the initial tests with the Audio Browse controller/person responding for the AI in the room. This method for conducting the usability tests limited these initial results and may have affected the conclusions about the ease of use of our prototype. It was later changed so that the participant was unable to see the controller and could only read the screen and hear the AI response. Another limitation faced with our prototype using the Wizard of Oz technique was the learning curve of each task. After timing the tasks, we found participants took a longer time on tasks with our prototype compared to other browsers due to the unfamiliarity of the system. To account for this, we should have given each participant training tasks to complete before hand, however, since we did not, this is a limitation of our prototype.

Lastly, as there aren't current models of audio browsers, we were limited in the technology we could use to compare to our prototype. At first we compared our prototype *AudioBrowse* to the Google Chrome web browser because it is the most commonly used browser [11]. However, this did not account for the learning curve of performing tasks on *AudioBrowse* so for the timings of these tests, Google Chrome out performed our prototype. In our second round of tests, we considered comparing the use of our Prototype to either Tor browser or LipSurf to better account for the learning curve. We chose to compare our prototype against Tor because of the privacy and security aspects of the browser.

Conclusion

Overall, AudioBrowse prototype browser provides a new way of interacting with the web that meets the user needs that we aimed for. Based on the usability test result, our prototype comes close to the state of the art browsers in performances measurements but still lacks behind. Based on the recommendations, we believe our prototype can be vastly improved in many fronts and eventually exceed the experience current browsers provide.

Attachments

Results Analysis - Spy Questionnaire

1. How easy did you feel it was to hear the tasks being performed by the user?
2. What was the email address the user was trying to log into? If you don't know, write n/a.
3. What did you hear the user search for (list as many as you remember)? If you don't know, write n/a.
4. Was there any other confidential information you overheard?
5. Were you actively listening as a spy?

Attachment A.

Results Analysis - Participant Questionnaire

1. How easy did you feel it was to search using AudioBrowse?
2. How would you rate the design of the audio browser?
3. Was there anything you would change about the way the browser looked?
4. Was it difficult to communicate the passphrase?
5. How would you feel if you had to input your real email address and password?
6. What did you think about the AI screen verbal response?
7. In what setting would you use AudioBrowse, given the security features of saving credentials. Select all that apply.
8. Did you like AudioBrowse and would you want to use it in the future?

Attachment B.

Contributorship Statement

We worked on this assignment as Group 5 (Ann Stowasser, David McElveen, Hollis Liu, Mason Strohl, and Sally Lee). Each person's role in completing the assignment was as follows:
- Ann: Executive Summary, Methods used to evaluate, User results

- David: Researched current technology, software, and solutions
- Hollis: Prototype description, usability test plan design, experiment design
- Mason: Demographic and Neural Network stats, Prototype Description, Methods
- Sally: Prototype Description, Spy results, Recommendations, Limitations

Sources

Anon. 2017. Quick Statistics About Voice, Speech, Language. (December 2017). Retrieved March 24, 2019 from https://www.nidcd.nih.gov/health/statistics/quick-statistics-voice-speech-language

Anon. 2019. Blindness Statistics. (January 2019). Retrieved March 24, 2019 from https://nfb.org/resources/blindness-statistics

Hilko Donker, Palle Klante, and Peter Gorny. 2002. The design of auditory user interfaces for blind users. *Proceedings of the second Nordic conference on Human-computer interaction - NordiCHI 02* (2002). DOI:http://dx.doi.org/10.1145/572021.572038

Rich Caruana and Alexandru Niculescu-Mizil. 2006. An empirical comparison of supervised learning algorithms. *Proceedings of the 23rd international conference on Machine learning - ICML 06*(2006). DOI:http://dx.doi.org/10.1145/1143844.1143865

W. Xiong et al. 2017. The microsoft 2016 conversational speech recognition system. *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2017). DOI:http://dx.doi.org/10.1109/icassp.2017.7953159

Anon. HTML <alt> Attribute. W3 Schools. Retrieved April 15, 2019 from https://www.w3schools.com/tags/att_img_alt.asp

Sav Shrestha and Kelsi Lenz. 2007. Eye Gaze Patterns while Searching vs. Browsing a Website. *Usability News* (January 2007).

Jakob Nielsen. 2012. How Many Test Users in a Usability Study? (April 2012). Retrieved April 15, 2019 from https://www.nngroup.com/articles/how-many-test-users/

Anon. 2019. Users. (April 2019). Retrieved April 16, 2019 from https://metrics.torproject.org/userstats-relay-country.html

Anon. GLOBAL DIGITAL REPORT 2018. Retrieved April 5, 2019 from https://digitalreport.wearesocial.com/

Anon. Web Browser Market Share. Retrieved April 4, 2019 from https://www.w3counter.com/globalstats.php?year=2019&month=3