Bilkent University

Department of Computer Engineering

CS 491 Senior Design Project

Project progress report

Sum of Sounds

Team members

Aliyu Vandana Saifullah

Bilal Siraj

Kasymbek Tashbaev

Supervisor

Prof. Dr. Uğur Doğrusöz

Jury Members

Prof. Dr. Varol Akman

Prof. Dr. Özcan Öztürk

October 14, 2019

Table of contents

1. Introduction

Nowadays speech synthesis technologies are very popular. They are mostly used for translating text information into audio information and in applications that use voice-enabled services and mobile applications. Apart from this, speech synthesis is also used in assistive technology for helping vision-impaired individuals in reading text content [1]. There are lots of text-to-speech applications on the market now, such as an Intelligent Speaker and Natural Reader that read books, articles, docs and pdf files in an almost human voice. All of them read plain texts without any problems, but when there are mathematical formulas in the text almost all of such text-to-speech applications have trouble with voicing the formulas, especially texts which contain complex mathematical formulas that include integrals, summation, and other symbols. For example, when Natural Reader meets mathematical formulas it just ignores them and continues reading the next plain text.

There are people with visual impairments who do research in mathematics or physics and need to read scientific articles and books that cannot be read using text-to-speech applications. One of them is Vitaliy Vertogradov - Doctor of Science, who researches in the field of black holes and gravitational collapse. He was looking for volunteers who can read a book with formulas [2]. His post motivated our team to develop an application that could voice the mathematical formulas and equations. Such a project can be useful for both developers of text-to-speech applications and people with vision problems.

1.1 Description

The program is going to be a web application and browser extension voice out read the mathematical formulas and equations in document or image. Considering that there are various formulas in the scientific world, we decided to focus our application more on voicing LaTeX formulas. LaTeX is a high-quality document preparation system, especially for scientific documents [3]. For example, the following mathematical equation:

$$\begin{aligned} A &= \frac{\pi r^2}{2} \\ &= \frac{1}{2}\pi r^2 \end{aligned} \tag{1}$$

in LaTeX is written as

```
\begin{equation} \label{eq1}
\begin{split}
 A & = \frac{\pi r^2}{2} \\
 & = \frac{1}{2} \pi r^2
\end{split}
\end{equation}
```

The number of mathematical symbols is limited, so our program will have a text-based description for each of the mathematical symbols which will be converted to speech. Also, there will be text descriptions of Greek and Roman characters, SI units and fractions. Therefore, firstly, using these data the program converts the LaTeX formula into natural language spelling. Then, using Google's text-to-speech API, the program voices out the formula.

There will be a web application where users can upload documents in PDF or LaTeX format or image. If the LaTeX file is uploaded, then the program detects all

formulas in the file and voices them out, from the first to the last. If the image or PDF file is uploaded, first using the API of the third-party application Mathpix, which extracts LaTeX from the PDF files or handwritten notes [4], converts the contents of the image or PDF file to LaTeX, and then voices out the converted file.

Moreover, there will be a browser extension of the program. If a PDF or LaTeX document or image opened on the current browser tab, the extension can directly voice out the formulas in that file.

1.2 Constraints

1.2.1 Implementation Constraints

- The application will be a web application that can have a browser extension component.
- GitHub will be used as the version control and code management system.
- Google Cloud's Text-to-speech API will be used as the foundational technology.
- Mathpix's Snip API will be used to convert images and PDFs to LaTeX. [4]

1.2.2 Sustainability Constraints

- User feedback and data may be collected with user consent for the purposes of improvement of the application.
- The application has no foreseeable monetization plan.

1.2.3 Economic Constraints

- The cost of using third party APIs will depend on the usage and can range from free up to USD 16 per month.
- A free hosting server will be used to host the web application.

- A one-off developer registration fee of USD 5 is required to publish an extension on the Google Chrome Webstore.

1.3 Professional and Ethical Issues

The application will convert user-uploaded text to speech including formulas supported in LaTeX. The raw text, images, or files uploaded to the application will not be recorded or shared without user consent. The developers are not responsible for the nature of the information users upload and cannot be held liable for the use of the application on any items of media that are copyrighted and not in the public domain. This includes any sensitive information that users may upload such as credit card details and passwords.

2. Requirements

2.1 Non-functional Requirements

2.1.1 Usability

- The application will have a user-friendly interface that will allow users to navigate through it easily without wasting much time trying to voice out formulas.
- The application will have a web extension that users can add to their favorite browsers. Whenever users want to open the application to read their formulas they can do so directly from the in-built web extension.
- The application will not be flooded with a lot of functionalities that may overwhelm the user.

2.1.2 Extensibility

- The application will be made such that new features can be added easily.

- As there is always room for improvement, the application will be updatable.

2.1.3 Reliability

- The application will not modify any files uploaded to voice out formulas.
- Non-scientific formulas will be omitted when encountered in a document.
- The application will not store users' information or track their activities in the browser.
- The formulas in users' documents will be identified and converted to LaTeX format which will then be voiced out accurately.

2.1.4 Availability

- The application can be accessed on the web by anyone with access to the link or web extension.
- The web extension will be available for download and can be used to redirect to the web application.

2.1.5 Efficiency

- The application will be able to identify formulas in texts in few seconds and voice them out without intermittent delay or breaks.
- The web extension can be integrated quickly into the various browsers without any difficulties and there won't be unexpected crashes.

2.2 Functional Requirements

- The users can be able to upload files in pdf and LaTeX formats.
- The application will be able to read formulas from scanned documents or images.
- Users can be able to adjust the reading pace to suit their requirements.

- Users can add the web extension to their browsers with a single click on the homepage of the app.
- With the web extension, the users are able to voice out formulas on a document opened in the current browser tab.
- The user can be able to obtain the formulas in the uploaded files
- The user can be able to convert formulas into their LaTeX format.
- Scanned formulas will be voiced out and their LaTeX format will also be available for the user.

3. References

[1] “What is Speech Synthesis? - Definition from Techopedia,” *Techopedia.com*. [Online]. Available: https://www.techopedia.com/definition/3647/speech-synthesis. [Accessed: 13-Oct-2019].

[2] V. Vertogradov, “ Looking for volunteers,” *VKontakte*. [Online]. Available: https://vk.com/id11510315?w=wall11510315_11361/all. [Accessed: 13-Oct-2019].

[3] “An introduction to LaTeX,” *Introduction to LaTeX*. [Online]. Available: https://www.latex-project.org/about/. [Accessed: 13-Oct-2019].

[4] “Mathpix Snip,” *Mathpix Snip*. [Online]. Available: https://mathpix.com/. [Accessed: 13-Oct-2019].