LISTENAPP:

AN AI-BASED MOBILE APPLICATION PLATFORM FOR AUDITORY LEARNING

by

Ajith Kumar Balakrishna Pillai

Master of Business Administration, University of Kerala, 2005

A Major Research Project

presented to Ryerson University

in partial fulfillment of the

requirements for the degree of

Master of Digital Media

in the program of

Digital Media

Toronto, Ontario, Canada, 2020

Author's Declaration

I hereby declare that I am the sole author of this MRP. This is a true copy of the MRP, including any required final revisions.

I authorize Ryerson University to lend this MRP to other institutions or individuals for the purpose of scholarly research.

I further authorize Ryerson University to reproduce this MRP by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

I understand that my MRP may be made electronically available to the public.

Ajith Kumar Balakrishna Pillai

LISTENAPP:

AN AI-BASED MOBILE APPLICATION PLATFORM FOR AUDITORY LEARNING

Ajith Kumar Balakrishna Pillai
Master of Digital Media
Ryerson University, 2020

Abstract

In today's fast-paced world, students and working-class individuals are increasingly replacing the written word with audio content – primarily podcasts and audiobooks. However, not all content is available in audio format. Print-first media like newspapers and textbooks remain unavailable to the growing group of consumers who prefer the spoken word for its accessibility while commuting, working, or otherwise unable to engage fully with a printed work. Automated text-to-speech solutions exist but read in a flat affect, which fails to communicate the emotions attached to the writing. This paper provides an insight into current solutions and limitations and explains how a mobile application framework is developed to overcome these limitations by using artificial intelligence, including natural language processing to parse meaning and the relatively new field of audio style transfer for speech generation to convert any written work into an audible, read in a voice chosen by the user.

Keywords: Text to speech, auditory learning, Artificial Intelligence, AI, voice cloning, optical character recognition, TTS, OCR, style transfer

Acknowledgments

I would like to sincerely thank my MRP supervisor Dr. Charles Davis, Associate Dean, Faculty of Communication and Design at Ryerson University, for providing guidance and steering me towards the right direction, helping me complete the project successfully. He was always available to answer my questions and provided me with valuable feedback.

I would also like to thank Mr. Praveen Kumar A.X, for helping me develop the product by assembling and coding the application. And for teaching me new technologies and debugging the code I developed.

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

With our day-to-day activities becoming more digitalized, technology has started to play an important function in helping individuals with special needs and providing them with far better chances for freedom. Today, accessibility is not optional; rather, it is a necessity. Considering 10% of the American population has low vision (Facts and Figures adult's vision loss, 2019), there is a dire need for more applications to be built that help the blind as well as visually challenged. There is considerable growth in assistive technology, and AT tools like screen reading, screen magnifiers, or apps like blind square developed for visually impaired, have become a blessing. However, only a minuscule area of problem has been addressed by these tools and applications. Areas like accessibility in reading or auditory learning and learning disabilities have not got the attention they require (Fonseca et al ,2017). Studies show that there has been a considerable number of individuals who prefer audiobooks over books. A recent study by Audio Publishers Association shows that in 2019, 26 % of US consumers listened to audiobooks. That is a 60 % growth from previous years' reports. As modern technology became less expensive and more accessible, audiobooks started acquiring popularity; the other essential element was a large number of disabled readers preferring audiobooks. Several enthusiastic audiobooks followers like to listen to these audiobooks throughout their commutes or while doing chores. At the same time, others cannot hold or read ebooks or print as a result of impairments or cognitive disabilities. eBooks have evolved to audiobooks, and audiobooks transitioned to podcasts which has snowballed in the last five years. According to Edison Research, the number of North Americans listening to podcasts has grown drastically over the past few years. In 2018, 27% of Americans over the age of 12 listened to a podcast at least once a month; this is a significant rise over the 17% figure reported for 2015. The report suggests that the average listener is consuming a lot more material as well.

Podcast mobile application usage has climbed 60% since January 2018. The sector's growth is expected to continue, as 45% of listeners said they plan on tuning into more podcasts in the future, according to a study conducted by Adobe Analytics. (Nguyen, 2019) The bottom line is that the podcast sector is showing robust growth in demand. However, is there enough content available on the market?

Neil Fleming, a professor from Lincoln University, has identified a set of learning styles: Visual, auditory, read, and kinesthetic, collectively called the VARK model. The model further categorizes learners into visual learners, reading learners, kinesthetic learners, and auditory learners (Boitnott, 2017). The auditory learning concept is a style of learning where an individual learns by way of listening. An auditory student relies on listening as the major way of learning. Auditory learners could also face challenges or difficulties in understanding instructions that are drawn. However, it can be a bit easier for them to understand if the writing is logically presented. They tend to be excellent listeners, especially when people are speaking. The auditory learners are unique in that they do have a skill for ascertaining the true implication of someone's statements by listening to their audible gestures such as change of tone. When given a number to remember, an auditory learner would prefer to say it aloud and then memorize the sound for them to recall it.

Given the fast-changing nature of technology and the means of learning and communication in general, there has been an observed behavioral change in reading and communication behavior among students. Reading is currently trending towards screen reading. There are presently so many screen-based reading behaviors, and these are categorized or defined through the amount of time spent scanning, browsing, keyword spotting, and one-time reading.

They are also associated with non-linear reading as well as more selective reading. Notably, this kind of reading is associated with less time invested in reading deeply or paying much attention. There is a noted decline in the degree of sustained attention span. It is further observed that the practice of annotating, as well as highlighting throughout the reading, is majorly associated with printed material. This traditionally associated pattern is yet to migrate to the digital content when people reading is focused on electronic documents. Given the increased obtainability of the digital material and the amplified amount of time spent reading electronic material, people's reading behaviors have begun to change. Several scholars argue that the advent of digital media poses a threat to active and sustained reading. The younger generation growing up in the new digital environment becomes distracted in the web of the internet and loses focus causing them to drift away from one content material to another (Boitnott, 2017).

In light of the changes in human behaviors, particularly from the learning perspective, the text-to-speech technology (TTS) has attained prominence. TTS is a form of assistive technology that reads digitally based texts aloud. This kind of technology has, in cases, been referred to as "read aloud" technology. By just touching a button , TTS can turn words on a digital device into audio. TTS technology was originally developed to be an automated tool to help serve people deemed to be visually impaired (Stabile, 2017). It is evident that text-to speech technology has emerged as an ideal tool for a variety of technology service providers, particularly in their bid to improve customer service delivery. In simple terms, TTS is a tool that transforms printed text into a regular speech that can not only be heard but also understood by any human users. The text to speech technology has undergone massive evolution over the years. This evolution has been made possible through various underlying technologies. Some of these technologies include deep tools for learning, such as machine learning and artificial intelligence. For instance, based on machine

learning, the application of speech synthesis within the TTS has made it possible to do an artificial rendering of human-related speeches within computer systems. TTS technology is vitally important in terms of helping learners that struggle with reading. In the same manner, it can still be vital in terms of helping learners with their writing, editing, as well as enhancing concentration. TTS operates on virtually all personal digital devices. These include computers, tablets, as well as smartphones. All forms of text files can be turned into voices. TTS's voice is generated using artificial intelligence technology, and the speed of reading can usually either be slowed down or sped up depending on the user preference. However, the quality of the voice is questionable, and it lacks human emotion. Past studies indicate that there is an existence of some challenges and limitations in physical and auditory learning styles. These challenges and barriers need attention to get a solution. One of the common concerns repeatedly reported in auditory learning literature is the poor quality of the audio due to several factors. Some of the restrictions that lead to poor audio quality are restrictions of the internet bandwidth available and congestion of the traffic (Bright & Pallawela, 2016). These factors can negatively impact the listening skills and the quality of the audio delivered to the learners. Therefore, there may be hindrances in the classroom participation of students and limited interaction. The above-stated challenges can be overcome through adopting some text-to-speech recognition technology. Examples of text-to-speech recognition technologies include writing-to-text, text-to-speech, diagram-to-text, and handwriting-to-text, image-to-text. Note that there is a suggestion that text-to-speech recognition technology generates audio that can be useful to the students for a better understanding of the literature. Therefore, learners can simultaneously take notes during the listening session. It enables learners to complete their homework while listening to the texts in the literature.

The advancement in deep learning has improved the development of text-to-speech by imitating the vocal style and its structure to produce natural and high-quality speech output. The effectiveness and efficiency of text to speech system relies mainly on the neural network models that are difficult to mold and do not allow real-time speech synthesis. To resolve these hurdles, a team of AI researchers from IBM has conducted a study and developed a new technique for neural speech synthesis based on a modular design, which combines three deep neural networks with transitional signal processing of the networks' result. Another benefit of this technique is that when the base networks are trained, they can be conveniently adapted to a brand-new speaking design or voice, even with a minimum voice training input data. The synthesis procedure applies a language-specific front-end component that converts input message directly into a sequence of etymological attributes. (Kons, et al, 2019)

Chapter 2. Literature Review

2.1 The Narrative Paradigm – A communication theory

The Narrative Paradigm is a model in communication theory developed by Walter R. Fisher (Fisher, 1984). The theory argues that all meaningful and effective communication is in the form of storytelling. The communication process is influenced by how the information is presented to the listeners, and it can affect listeners' comprehension and decision-making ability. A narrative is any verbal and nonverbal expression that is orchestrated sensibly to convey a meaning. The theory states that communication occurs between a narrator and the listener in the form of a story. The story is set in the form of events that the narrator wants to communicate to the listener. Narrative Paradigm theory promotes the idea that the humans respond well to storytelling, and

aesthetic considerations and emotions can influence our beliefs and behavior. Narrative Paradigm comprises of two main principles of Coherence and Fidelity.

i. Coherence

Any information is effective only if it makes sense to the listener, and narrative coherence is how much a story bodes well, it refers to the internal consistency of communication. Coherence is often gauged by the structural and organizational elements of a story. The effectiveness of storytelling is influenced by the following three elements:

- The narrative structure – Structural coherence
- The consistency between events - Material coherence
- Credibility of characters – Characterological coherence

ii. Fidelity

The second principle, fidelity, is the degree of the credibility or reliability of the narrative, and whether the listener acknowledges the story. It is connected to the listeners' past experiences and how the experience of a story sounds valid with past stories they know to be accurate. Stories with fidelity can influence listeners' behaviour and beliefs. To acknowledge a story as true and worthy of acceptance; the listener asks the following questions.

- Are the statements that described to be true really factual?
- Have the facts been omitted or distorted to fit the narrative?
- What are the reasoning patterns followed in the narration?

- Does the argument in the story influence the decision making of the listener?
- How well the narrative addresses the importance of the story?

2.2 Text-to-speech technology

It should be noted that text-to-speech technology has gained massive popularity over the years. It has become prevalent, particularly as an assistive technology, where computers or tablets read the words out loud for the user's consumption. TTS technology has more popularly been used amongst students who face complications or difficulties reading. This is especially the case with those who find it particularly hard to comprehend complex sentences. It must be noted that TTS technology can help students find a way around their reading challenges and provide effective usage of classroom materials. Notably, the most recent years have seen a steady upsurge in the amount of TTS software developed for Android and IOS Platforms. Text-to-speech technology has also gained massive popularity within workplace environments where it is being used as a tool to aid the proofreading of the users' work. Despite this growing popularity, research concerning text to speech technology is vague and somewhat unclear. While the technology plays a vital role in terms of helping students access the classroom material, a section of researchers have been confronted with mixed results especially in terms of how well the student is in a position to grasp the text that the devices read to them (Dalton & Strangman, 2005).

Additionally, some other researchers do assert that text-to-speech technology hardly affected adolescent students' capacity to understand the reading. Whilst, the same researchers further noted that the students do value the enhanced independence attained through the use of TTS technology (Meyer et al. .2014). However, another study found that students who have been identified to suffer from dyslexia, benefited greatly from the use of TTS technology . This particular team of researchers carried out a user testing with a small group of students for six weeks

and they were able to observe significant improvements in the students. They observed improvements in fluency, inspiration to read, and enhanced comprehension (Pieraccini, 2012). In the same manner, there were positive results observed in a different study where TTS technology was effective in making it possible for students to access material for reading. Additionally, it was favorably perceived by the students who used it.

Text-to-speech technology development started in the early 1980s, and since then, it has been developing at a faster rate. Due to increased technological advancements, researchers and practitioners have begun to use text-speech tools to assist students with reading difficulties (Wood et al., 2018). Improved technology over the past 30 years has led to an increase in the use of electronic versions of books and other software incorporated with text-to-speech features. Text-to-speech software development is considered to be taking place across different parts of the world, such as the United States, Europe, and China. Some examples of this software include ClassMate Reader, Dec Talk, and Kurzweil 3000, among others (Nwakanma et al., 2014). More so, the number of free mobile applications related to text-to-speech functions is on the rise in the world today, whereby people can download them and be able to use them. It is noted that these applications consist of voice options, development of synthetic audio files, custom pronouncement, and other features such as text highlighting. Text-to-speech synthesis technology features are considered to influence users' experience and, as a result, affect the effectiveness of their usage. These features include voice type, dynamic highlighting, reading rate, and others. The reading rate is considered a key feature that influences users' experience. It is noted that people with difficulties in reading are considered to be heterogeneous. Some of these people can be more skilled than others in comprehending, which means that they can benefit from text-to-speech applications in different ways.

Speech synthesis is also called text-to-speech, which is responsible for converting language text into speech so that a smartphone or computer can read the produced output in the form of audio. It is noted that speech synthesis is different from speech recognition in such a way that speech synthesis is a text to speech converter while speech recognition is a speech to text converter. Furthermore, the quality of speech synthesizer can be measured in two ways: intelligibility and naturalness (Nusbaum et al., 2015). Intelligibility refers to the clearness of the "output voice," while naturalness refers to the capability to resemble the voice of a human being. In addition to the earlier text-to-speech software, other technologies ensure that text is effectively converted to speech. They include IBM Watson text-to-speech, MaryTTS, and FreeTTS, among others. Mary Text-to-Speech (TSS) is the commonly used technology since it is considered to be an open-source software which is coded in "Java" and can be used to synthesize a variety of natural languages (Kraljevski et al., 2010). More so, the MaryTTS is the form of technology that was developed by DFKI's Language Technology Lab together with the "Institute of Phonetics" at Saarland University. This software supports a variety of languages such as British, German, Italian, and French, among others, and it also runs on many platforms.

2.3 Text-to-speech Mobile Applications

One of the apps to consider is the Speechify app, which is regarded as a text -to-audiobook application. Users can upload content to the app from different sources such as websites, documents, and images. After uploading the content into the app, the added text is read out in the form of sound. It is noted that words are highlighted to ensure that users can follow along when reading and it is easy to change to different voices such as American, British, Australian, and kids' voices. This app also has a speed adjustment option whereby a user can slide the dot towards the bottom of the page, and this can be done by adjusting 100 up to 800 words in a minute (Mugayi,

2019). It also offers the option of organizing the content to be uploaded in various collections. More so, there are more than ten languages available to help in translation such as Spanish, Chinese, Italian, German, and Russian, among others. With this app, uploading data is straightforward, and there is the option of snapping a photo of content to be uploaded. It is also noted that adjusting to the "in-app voices" might take time since they barely sound natural. There is an option of selecting a child's voice, which tends to make the app user friendly for younger users. The Pro version of the app is about $8.00 per month, which is expensive to users based on their reviews about the app. Users need to take extra effort and upload more content during the trial period to find out if the application is upgrade worthy. By doing this, users have much time to decide whether to upgrade or not after having experience with the app features.

Based on the positive reviews about this app, a second-year law student explained how he had developed his auditory learning skills by multitasking while listening to audiobooks. After discovering the Speechify app, he said that it helped him to catch up with reading assignments, and his retention levels increased. He added that this app helped him pause and interact with the text content while listening. The other reviewer was a parent whose kid is autistic and also having ADHD syndrome. He says that this app has benefited him and his kid, whereby they can listen to the pdf files on their phones, especially while they are around the swimming pool.

Most critical reviews that Speechify had received were on its incapability to produce voice that has human emotion. Some users also pointed out that the app does not identify the punctuation correctly, resulting in wrong interpretation of sentences.

The Pros and Cons of Speechify are;

-Pros: OCR capacity and multi-lingual support

-Cons: Robotic voice, incorrect punctuation and very expensive

Speak4Me is a text-to-speech app that speaks for the user, allowing speech impaired individuals to communicate. The user types in what he wishes to say in the app and the app converts it to audio by reading the text aloud. This app can be downloaded from Google Play store and App Store and is easy to use by young people and adults. Speak4Me offers the advantage of adjusting the speed at which it reads something that has been written. A user can also add phrases such as "favorites" so that he/she is not required to type it in every time. A user is also able to choose from different languages and allows the voice to be personalized. This app has a disadvantage whereby if a user wants to export the text, it has to be below 2000 characters, which can be a limiting factor. This means that if phrases are to be converted to audio so that they can be sent, the characters' limit has to be below 2000. Despite this, users can upgrade to premium where there is no limit of the words to convert, but an extra cost is incurred.

The Pros and Cons of Speak4me are;

- Pros: Speed adjustments
- Cons: Robotic voice, low character limit, and expensive

Voice Aloud Reader is a text to speech app that converts text to sound voice and considered to have the best quality "text to speech app" for iPhone and iPod Touch. This app provides better voice support whereby a user can choose from forty languages and listen to various voices. In terms of book support, the app is considered to support common text formats, and it also customizes the font depending on the user's discretion. A user can adjust volume, speed, and pitch of reading. This app allows a user to use a multi-task mode whereby the pdf reader pulls up while highlighting the file and makes notes as the app is running.

The Pros and Cons of Voice Aloud Reader are:

- Pros: OCR capacity and multi-lingual support, speed adjustment

- Cons: Robotic voice, incorrect punctuation and non- user-friendly user interface

Speech Central is another common mobile application used because it is highly recommended for maximizing productivity. It consists of the total package with unique features and offers an enhanced user experience. One of the factors that make this app a special app for text-to-speech is that it provides "real-time speech" capabilities by announcing the reading time for long stories and articles. With this app, there is limited wastage of time since a user can use Bluetooth or headphones to access audio documents. It is also noted that Speech Central can automatically save "user's reading" history, which makes it appropriate for professional use. This app is affordable and highly valued. It is available on Google Play store and IOS App store, which means that many groups of people can use it.

The Pros and Cons of Speech Central are;

-Pros: Read time calculation, Good UI and affordable

-Cons: Robotic voice and lack of narrator personalization

Voice Dreamer Reader, a text-to-voice app with a variety of customized capabilities which is available on IOS and Android platforms. It is also noted that the app is loaded with more than 200 voices and 20 languages, which can be downloaded for free, but it can also be purchased for other additional features (Takano, 2019). Anything in the form of pdf, plain text, Microsoft Word, and PowerPoint documents can be read with Voice Dream. It is noted that files can be loaded from different platforms, including cloud-sharing platforms. It gives a user full control regarding rendering text to speech and supports playback speeds, which can also be adjusted and paused for various times. The only disadvantage of this app is that it creates voices that lacks emotion.

The Pros and Cons of Voice Dream Reader are ;

- Pros: OCR capacity and multi-lingual support, speed adjustment
- Cons: Robotic voice and incorrect punctuation

Motorhead, which allows users to come up with their own and personalized playlists depending on how they would like to read. It offers features like "Spotify," where the user can develop a catalog of learning at the start of the day and might not need to worry about it the next time. This app is only available on iOS, and it can easily be downloaded and used. Motorhead allows users to send pdf documents, plain texts, and articles and provides an option of extracting articles from other apps. One of the benefits that this application offers is the capability to share different texts with the playlist. User can change playlist speed and has an option of "real-time article formatting," which ensures that the user does not miss anything important to read on the platform.

The Pros and Cons of Motorhead are;

-Pros: OCR capacity and multi-lingual support, speed adjustment

-Cons: Robotic voice and incorrect punctuation

From analyzing the various currently existing text-to speech technologies and applications and their pros and cons, I could see that all of them had a major drawback – their inability to produce a voice that is human like. The robotic voice output of all these applications lacks coherence which makes it difficult for the listener to comprehend. And other features like narrator personalization, understanding the punctuation and ease of use also needs to be worked upon. One of the most significant considerations while developing Listenapp is that the application produces a natural sounding voice. Listenapp's technology is an upgrade on the current technology applied by apps such as Speechify, Speak4Me, Natural Reader, Voice Aloud Reader as well as Speak

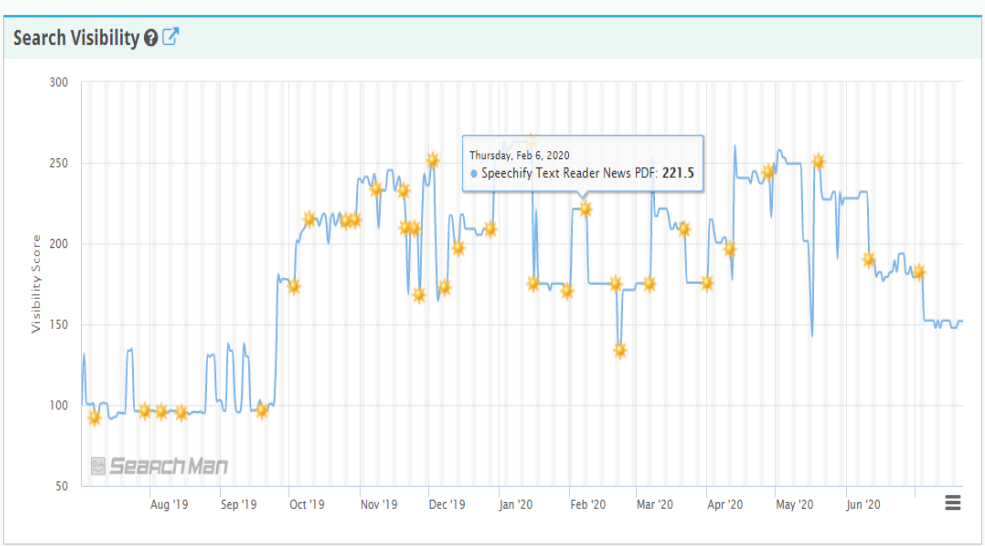
Speech Synthesizer. All of these apps have one particular challenge which Listenapp seeks to eliminate. The current technology they use merely produces a sound blurting out words without any emotions attached. They also lack the natural pauses, the ability to comprehend sarcasm, or even punctuation. Listenapp AI-powered speech synthesis will be able to clone any voice and deliver a human-like speech while also exhibiting the auditory components such as style, pitch, tone, and emotions. It is believed that by having a high-quality Listenapp voice sounding similar to that of a human, reading comprehension amongst the users will be significantly enhanced. The following is a preview of some of the TTS Apps and their related features

Comparison of Table

Table 1: Comparison of text-to-speech apps (Takano, 2019)

	Voice Dream Reader	Capti Voice	NaturalReader	vBookz PDF
Free Version	n/a	✅	✅	✅
Price	$14	$2 per month	$10	$5 per language
Voice upgrade	Extra voices at $2 to $5 each	$2 to $5 each	n/a	n/a
Text search	✅	Upgrade plan only	n/a	n/a
Bookmark	✅	Upgrade plan only	✅	✅
Pronunciation Dictionary	✅	n/a	n/a	n/a
Fine speed control	✅	✅	n/a	✅
Visual book cover	✅	n/a	✅	✅
Table of contents	✅	n/a	Partial support	n/a
Other versions	Android	Chrome extension and a desktop browser	Desktop browser and android	n/a
Pros	Easy navigation and stable	Free version and upgrade trial	Better navigation and stable	wide font selection which makes it stable
Cons	No free trial, Robotic voice	Robotic Voice, Poor Navigation	Robotic Voice The free version is limited.	Robotic Voice, Poor Navigation

Figure 1: Search Visibility



App name

Speechify text reader News PDF

Description

You can LISTEN faster than you can READ

Speechify is an intelligent text-to-speech audio-reader that turns your reading material into interactive audiobooks so that you can save time, retain more information, and stay focused. Like your own personal reading assistant, Speechify can read books, documents, and articles while you cook, work out, commute, or any other activity you can think of.

Figure 2: Keyword Search

Keywords See Trends

Keyword	Volume	Hits	KEI	Rank
speaktext for me			0.04	3
naturalreader text to speech			259	3
voice reader text to speech			0.54	2
kompas tts			0.02	2
dyslexia software			53.96	1
read text			40.09	1
dyslexia math			0.00	1
adhd reader			0.00	1
speaktext for ebook lite			0.09	1
maya magazine			3.94	1
speach me			0.00	1
text reader			32.78	1

Chapter 3. The Product- Listenapp

The Listenapp mobile application is designed to solve the problems of insufficient audio content and narrator personalization feature. Listenapp can create audio stories from any form of content. Listenapp is more than a traditional text-to-speech app, it is also an image-to-speech application, whereby it can convert a scanned image of text to speech. It provides a platform for users to create an audiobook by converting from text to speech using a voice that sounds less robotic and more natural and identical to the human voice. With AI-powered speech synthesis technology, the application could not only convert any text to speech and deliver human-like speech but also, could exhibit the style and emotions. The application simplifies the listening experience by allowing the user to convert any written content to a soundscape that is narrated in his/her favorite preloaded voice in less than 5 seconds. After installing the application to the mobile phone, the user scans or takes a picture of the written/printed content that needs to be made into an audiobook. Then, the user indicates their preferences for their preferred narrators as well as what type of soundscape they would like to generate. There will be a set of preloaded narrators or an option of creating a new voice. These are essentially helpful in terms of helping people to listen and comprehend as much content as they require.

There are various ways through which data can be input into the application. One way is to scan or take a picture of the page in the textbook and Listenapp will convert it to audiobook. Another way is to upload a PDF into Listenapp, and the app will be able to convert and read it in any way the listener would prefer. Yet another way is by importing articles from social media sites and blogs, and Listenapp will convert them into an audiobook. And lastly, one can simply copy and paste any particular text into Listenapp, and the app will read the text in their preferred language and style. Listenapp turns the reading material into an interactive audiobook. In so doing,

the users not only save time but are also able to retain much information while they are also able to maintain their focus for longer periods. Just like a personal reading assistant, Listenapp can be in a position of reading documents, articles as well as books while the listener does other activities such as cooking, commuting, working out among other interactive activities. Listenapp also supports listening in HD voices, and can provide multi-language support in the future. Users will also be able to adjust the speed of the voice; the speed can be slowed down or sped up. The creation of the Listenapp is aimed at eliminating all reading barriers. Everyone, regardless of their reading complications, will now able to consume their reading materials without any complications and with much ease as well in a manner that suits their auditory preferences.

Currently, Listenapp has five preloaded voices, which were selected from a range of age groups, gender, and ethnicity. Very soon, users will be able to choose the tone of the narration. They will be provided with options to set the tonality with traits such as serious, joyful, and melancholic. Users will also be able to create a custom voice within 30 minutes of voice training. What makes Listenapp different from the current TTS solutions?

- Offers Convenience – Use it on the go.
- Exhibits Emotion- Intangible charm to the experience.
- Accessibility- Beneficial for those with special needs or visual impairments.
- Learning opportunities – Improve communication skills & social interaction.
- Affordable – Free to download

Figure 3: How it works?

Chapter 4. Product Development

Creating a complex mobile application like Listenapp, was not an easy process. I needed to learn new technologies and identify a third-party developer who could assemble technologies together to create a mobile application. After identifying the third-party developer, I decided to use Agile Methodology to develop the application. Agile methodology helps in delivering quality applications efficiently and quickly, and it adapts throughout the development process and exceptionally reduces the overall risk of the project. The methodology breaks down coding and project management into smaller modules. The process gave room for easy iteration and flexibility. Every step was documented and tested. Continuous testing and debugging process ensured that application meets the quality requirements. The steps followed are:

4.1 Planning

At this stage – my primary focus was to identify the target user, required technologies and define features, create user stories and define acceptance criteria.

4.2 Target User

Listenapp is created for auditory learners who prefers listening more than reading. Simon, 23 a graduate student from Ontario Canada is an ideal user of Listen app. Simon is an academically thriving student; however, he is an auditory learner. He listens to music and audiobooks regularly. He currently uses applications like Audible and Voice Dream but Simon feels those applications do not have all the books he wants to listen to or not all study materials have an audio version. His laptop could convert the text to speech but the voice sounds robotic and it lacks emotion. Simon would love to find a mobile application that would allow him to create audibles from printed books on the go.

Figure 4: User Persona

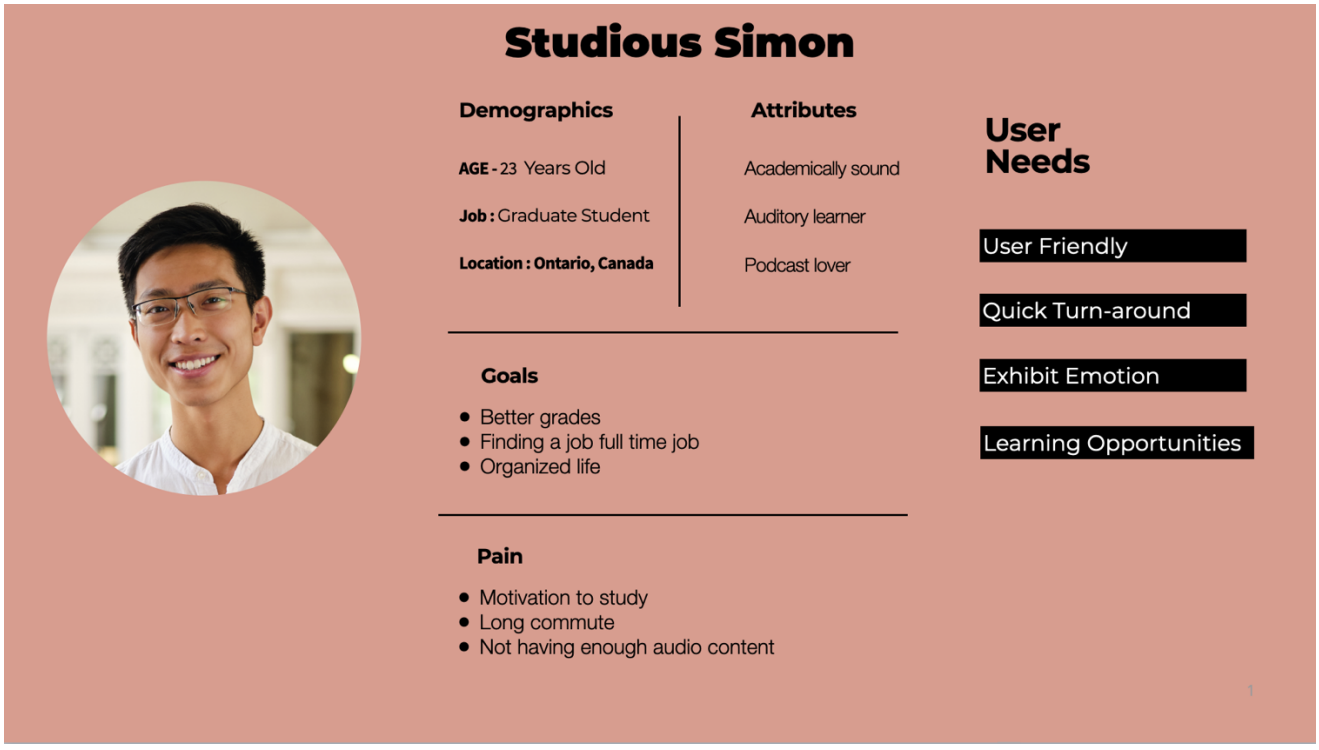
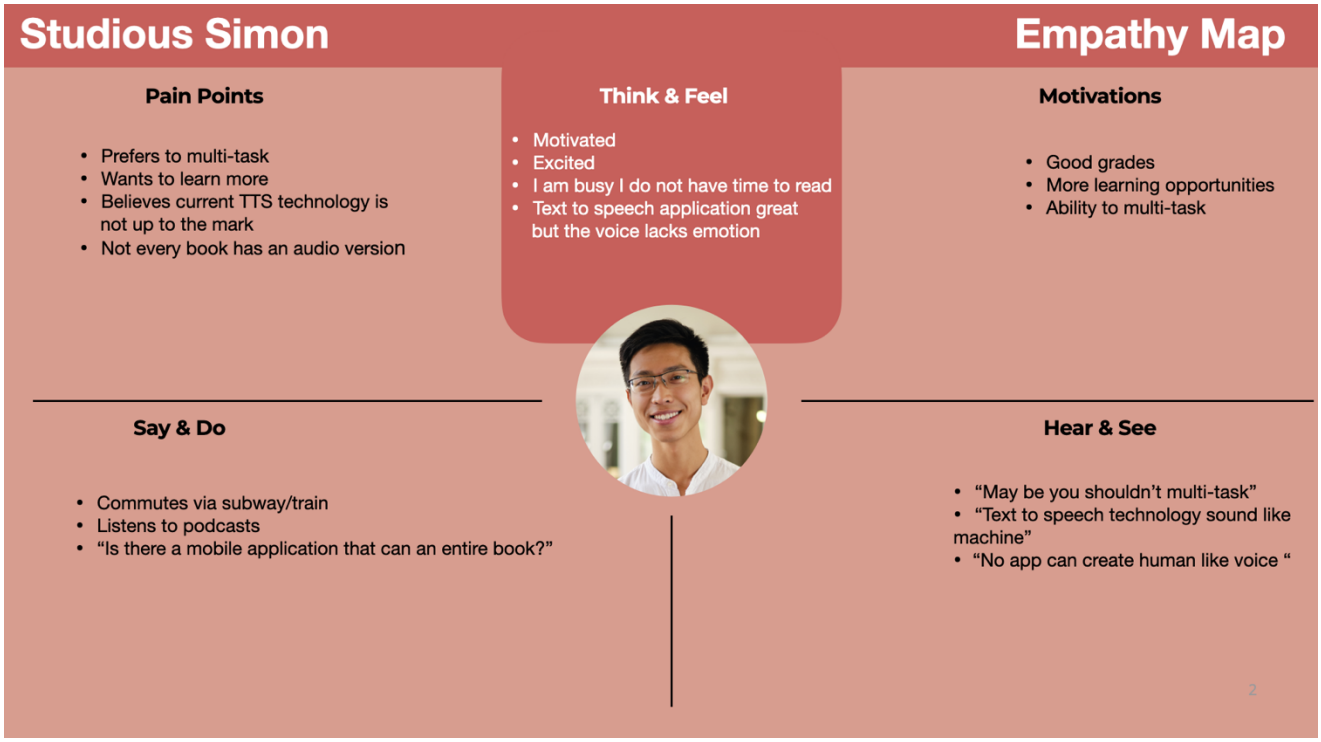


Figure 5: Empathy Map

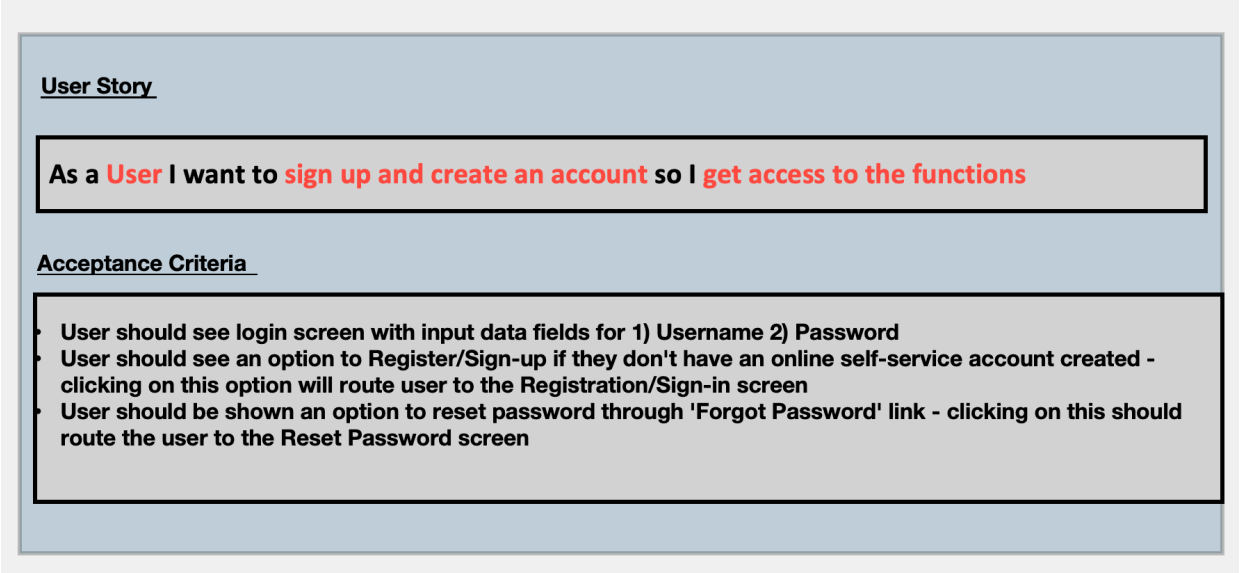


4.3 Defining Features

At this stage, the functions and features were defined based on the user stories created. User stories are created as part of agile methodology approach which redefines the requirement writing to storytelling. Agile user stories consist of one or two sentences and a set of conversation about the preferred functionality. User stories are short descriptions of an end goal or a feature expressed from the perspective of a user who desires certain functionality. User story card follows a template:

As a < user >, I want < goal > so that < reason >

Figure 6: User story Card



In order to define features I categorized the stories based on the desired end goal from a user's perspective.

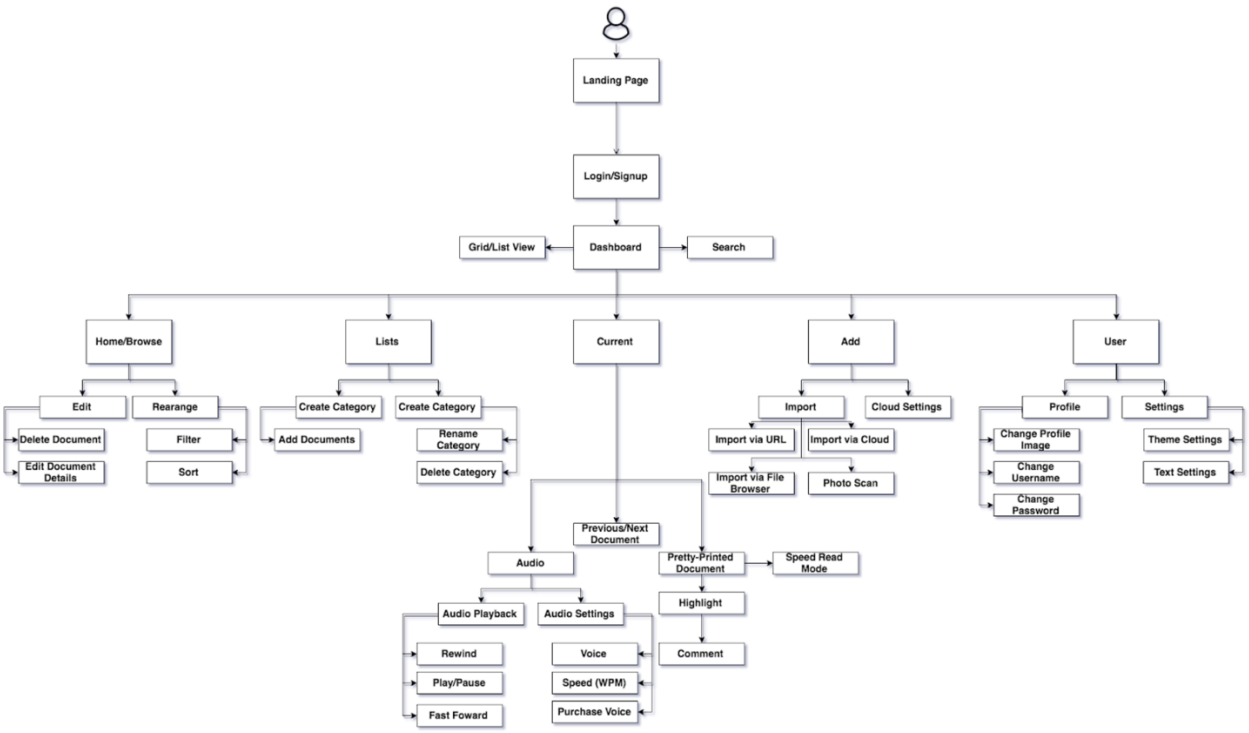
Table 2: User stories Table

Features	As a user I want to	So that I can
Scan/Upload	Import my text, including PDFs or EPubs files, online articles (HTML links)	Have the ability to easily access them wherever I want
	Import my text from cloud services such as Dropbox. iCloud, Google Drive etc	Have the ability to easily access them wherever I want
	Take a photo of my textbook or any other physical text	Have the ability to listen to it wherever I wart
Sort/Filter	Browse through imported files	View all of my files in a seamless way
	Sort my imported files	Have a way to find what I am looking for
	Filter my imported files	Have a way to find what I amlooking for
Library	Have the ability to create Library or playlists	Organize imported files based on my criterias
Search	Be able to search for any imported file	Have an easier way to find what i'm looking for
	Have my imported text pretty-printed	Be able to easily read the article
View	Swipe left or right to go to the previous or next imported text	Be able to easily move from one document from another
Text to Speech	Have the ability for the app to read my imported text to me	Listen to the imported file at my convenience
	Have the ability to pause, rewind and fast-forward	Have full control of what's being read to me
	Have the speech be synchronized with the text by highlighting	Be able to know what the app is reading to me at real time
	Have the ability to tap on a word to have the app read from that point	Be able to have control of what I want read to me
Settings	Have the ability to change the appearance of the app, including colors and text	Have options on how I want the app to look like
	Have the ability to log in and out	Have my imported files secure
Profile	Have the ability to use log into any device	Be able to have my imported files in any device
	Have a cover image associated with my imported file	Differentiate easily from one file to another
Voice	Have the ability to change the dictated voice	Have the ability to hear different voices from the app
	Have the ability to change the speed of the voice (WPM)	Be able to control how fast the voice is reading to me
Download	Have the ability to download the processed audio	Be able to share the file

4.4 Sitemap

After creating user stories, I drew a sitemap that explains the customer journey while using the mobile application. The sitemap is created by the following factors taken into consideration: Key features of the product, number of required screens, screen URL, Screen functions, user expectations, user flow, complexity, and call to action. The sitemap provided a clear sense of vision for the application and set appropriate user expectations.

Figure 7: Listenapp Sitemap – Created using Gloomaps.com



4.5 User Flow

Three main user flows were identified on the basis customer journey map and empathy map. The user flows exhibit how each step enables an end-user to reach a desired goal while using Listenapp. The user flows identified are:

- Sign In/ Sign Up
- Upload/Add Document
- Create Audio

Figure 8: Sign In / Sign out – User Flow created using Gloomaps

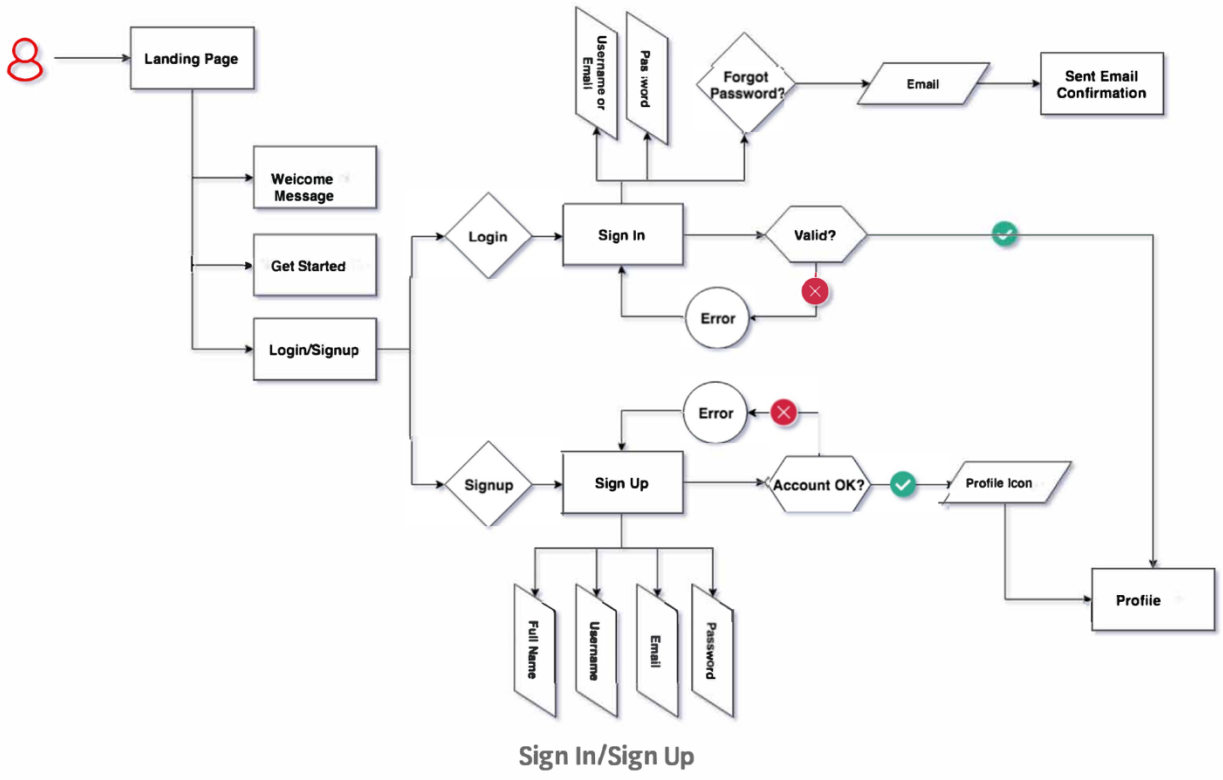


Figure 9: Upload/Add Document – User Flow created using Gloomaps

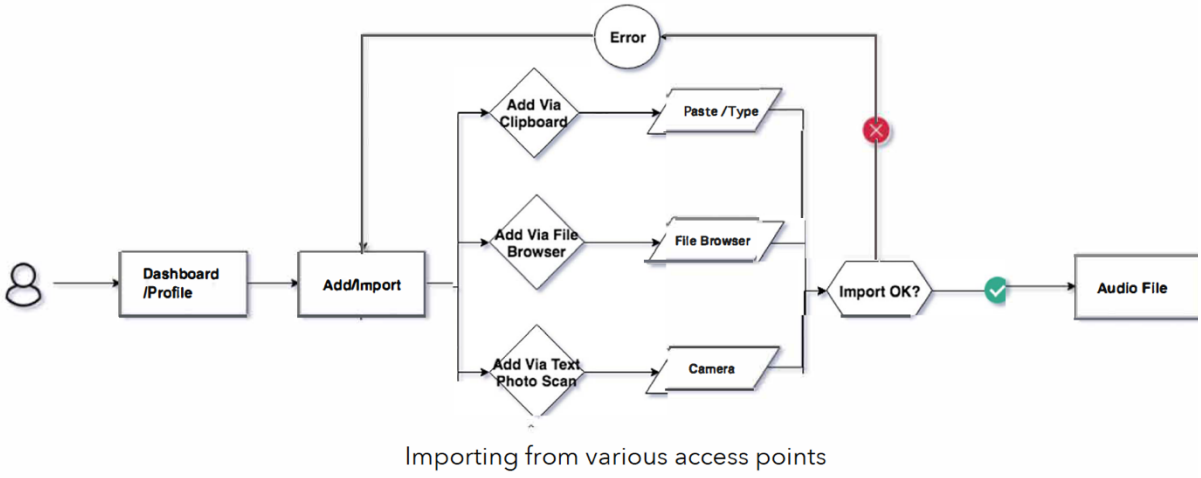
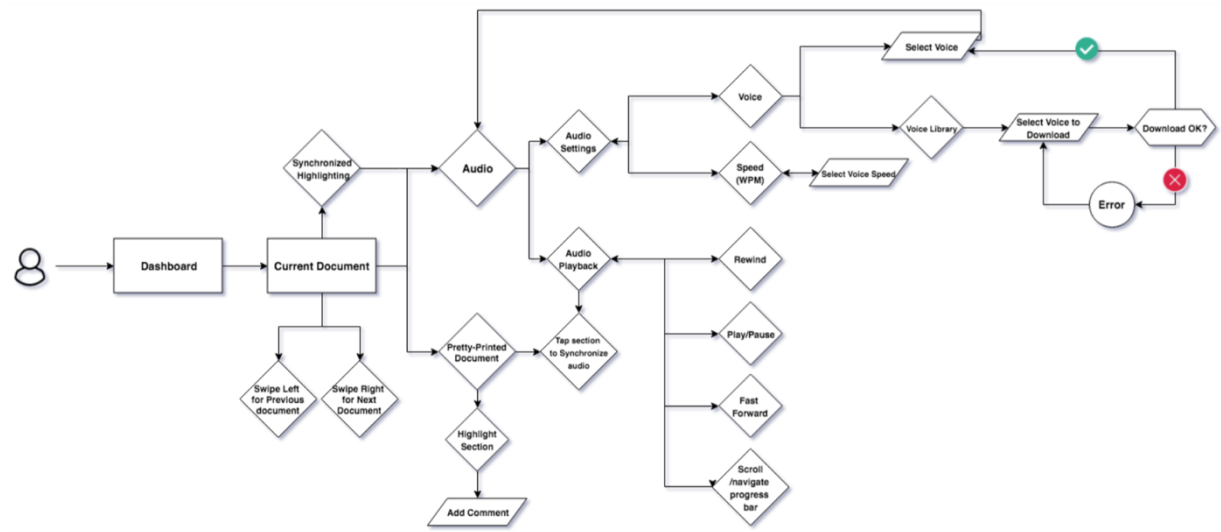


Figure 10: Create Audio- User Flow created using Gloomaps



4.6 Wireframing

During this phase, the focus was on the functionality and wireframing. While the functionality aspect emphasis how the application functions and what users can do with it, wireframing focuses on what users see in the app that they navigate and interact with; the functionality aspect focuses on how the app works and what users can do with it. The wireframing process is an essential part of the design process because it explains the user flow and application structure of a mobile application in a simple way. I used Sketch tool to describe the structure of Listenapp through wireframing based on the user flows created. The goal was to connect Listenapp's application architecture to its visual structure. The sitemap had many features; however, I needed to narrow down to essential functions that look most promising to create a minimum viable product.

Figure 11 : Wireframing 1 –created using Sketch

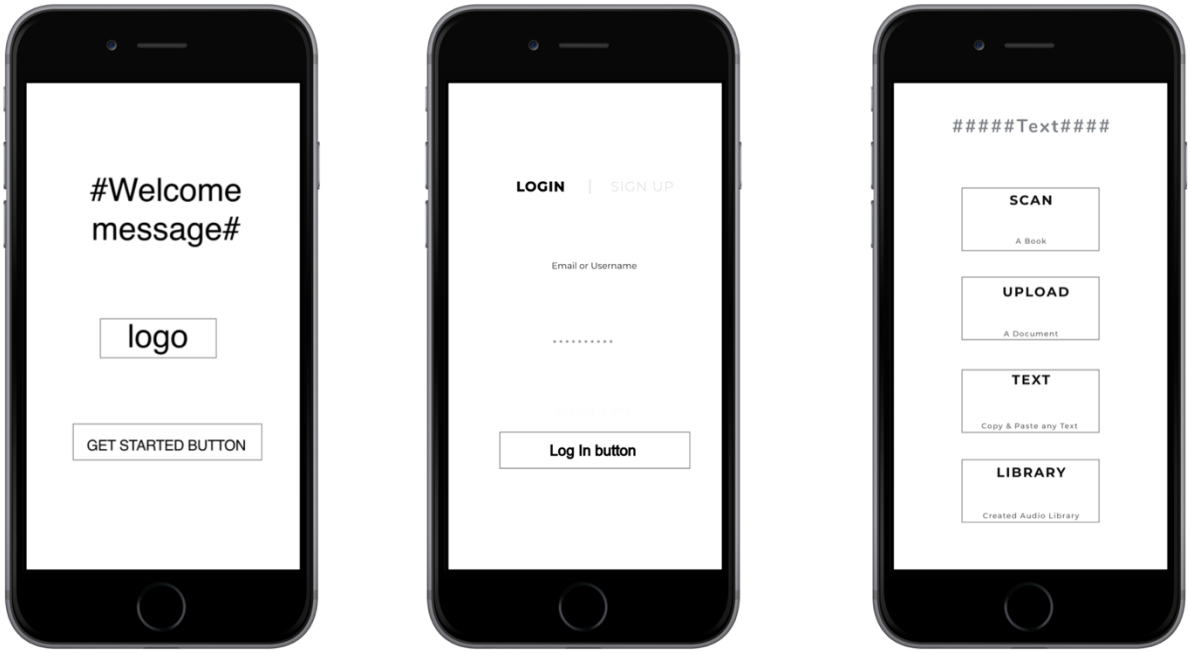


Figure 12 : Wireframing 2 –created using Sketch

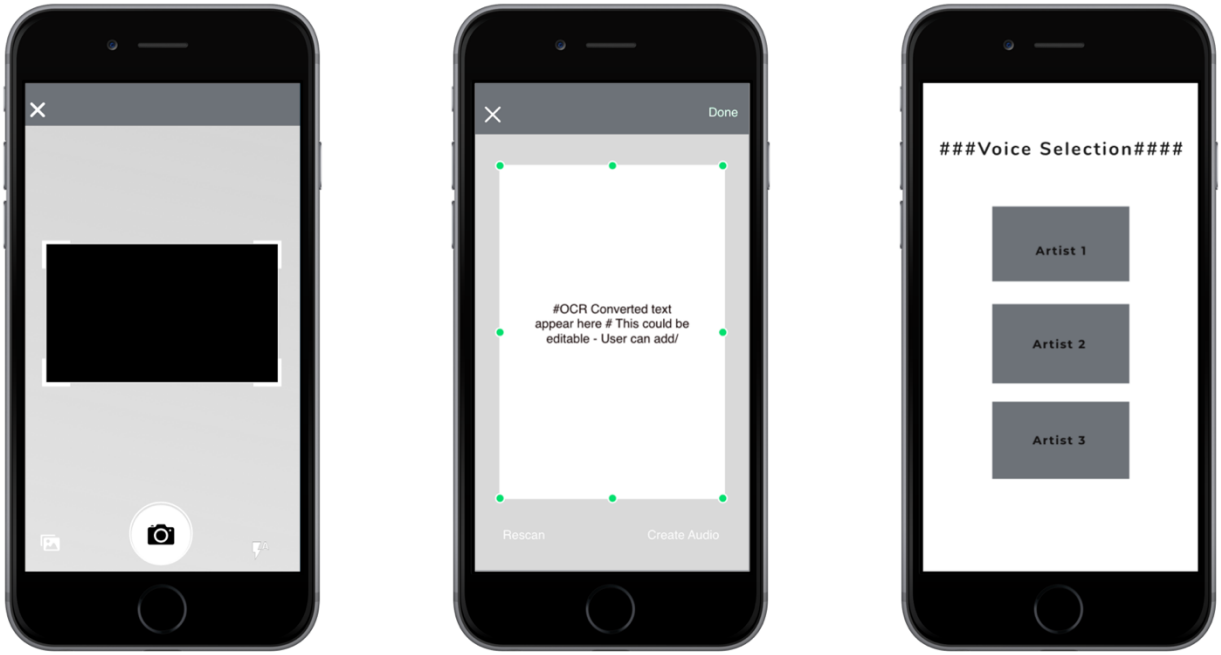
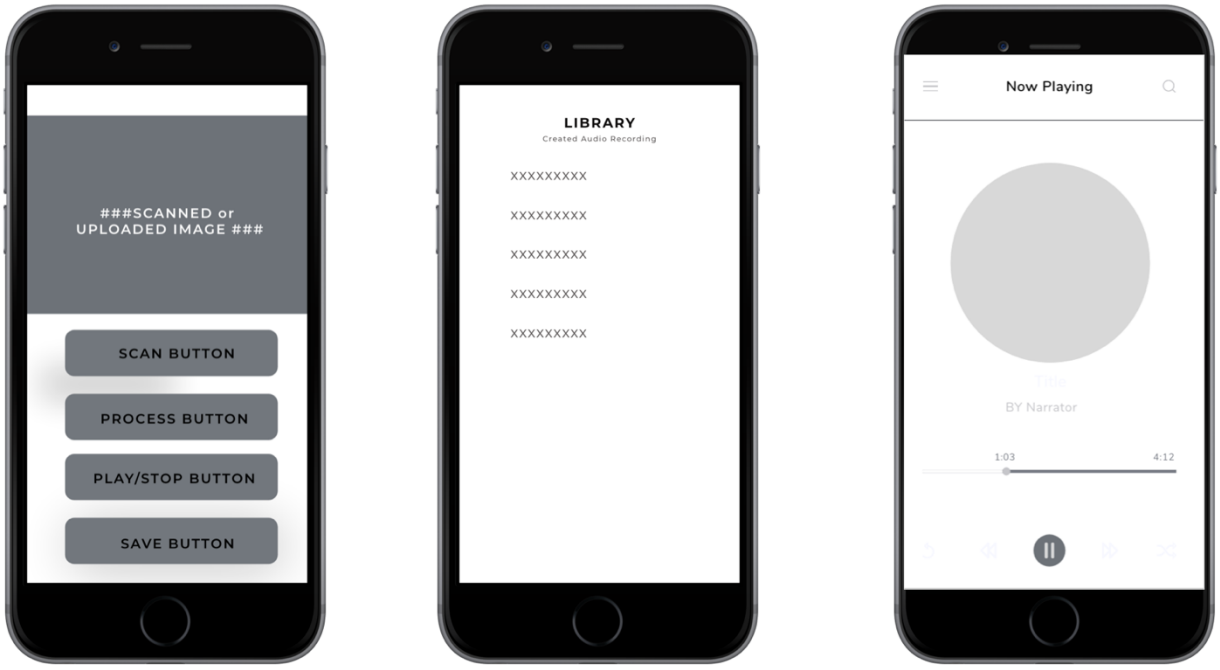


Figure 13 : Wireframing 3 –created using Sketch



4.7 User Interface / Visual Design

At this stage, the wireframes were converted to User Interface designs using the Adobe XD. This made the User Interface aesthetically stylish, efficient and easy to use, giving the user a pleasurable experience using the application. I used a color palette of Chilean fire, Copper Canyon, Cod gray and Wafer. The color combination gives the application a rich and elegant look and feel. The design follows the fundamentals of design accessibility and usability heuristics for User Interface Design (Nielsen.J 1994)

Figure 14: Color Palate

Figure 15 : UI Design 1 –created using Adobe XD

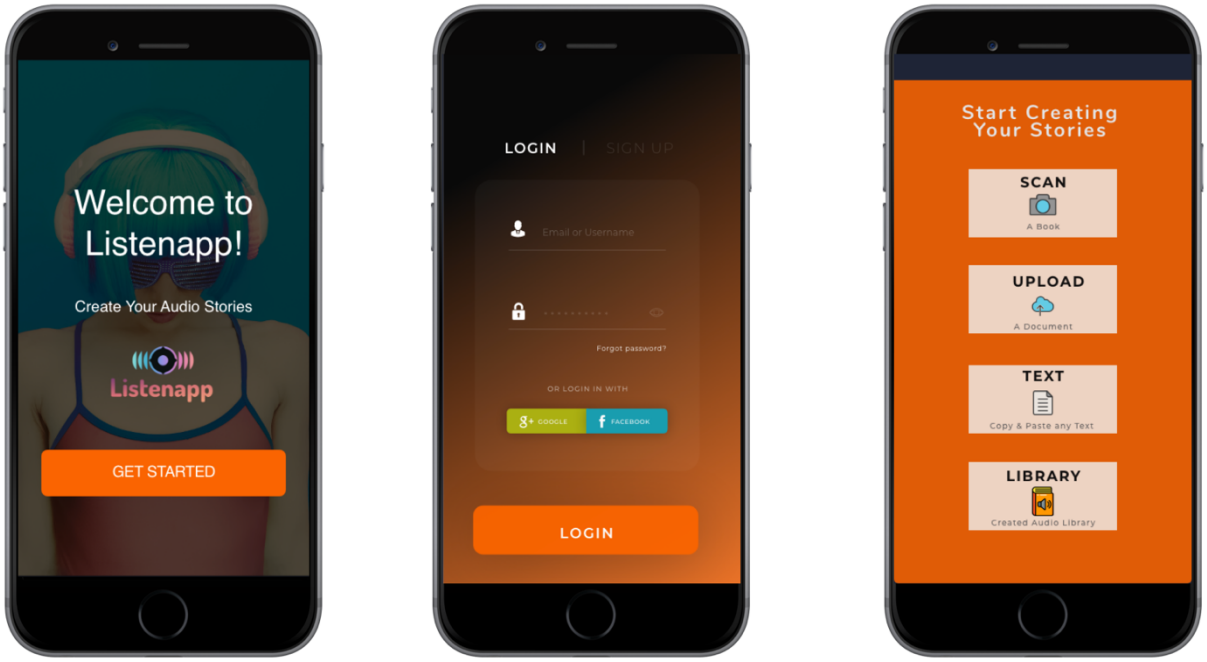


Figure 16: UI Design 2 –created using Adobe XD

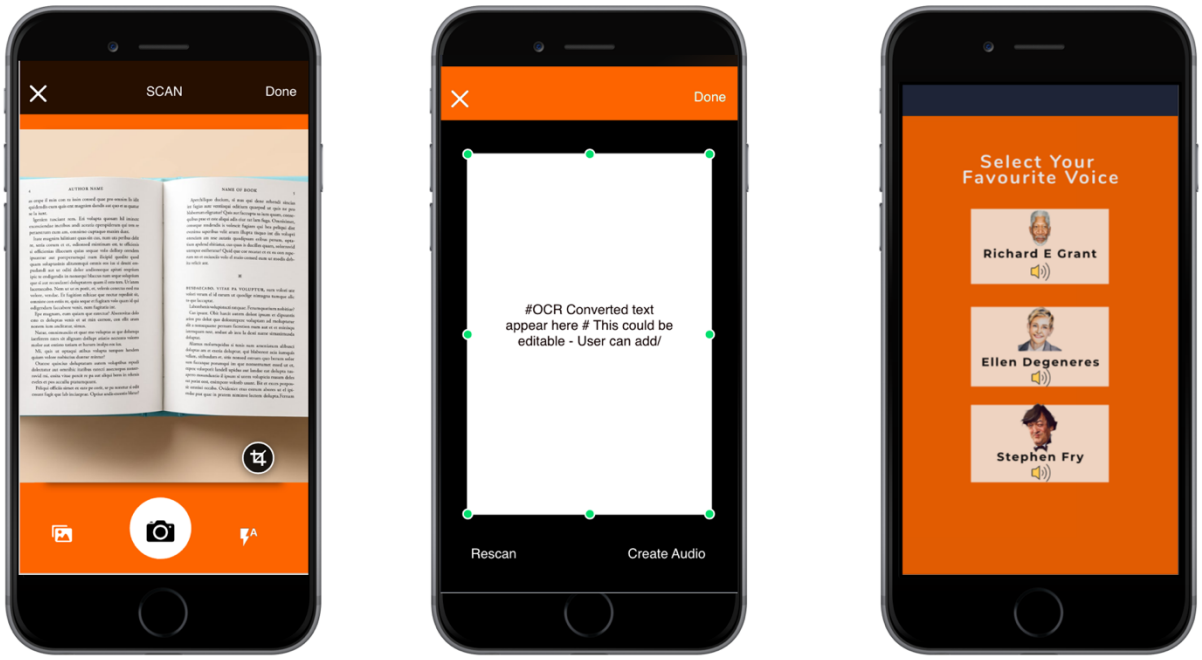
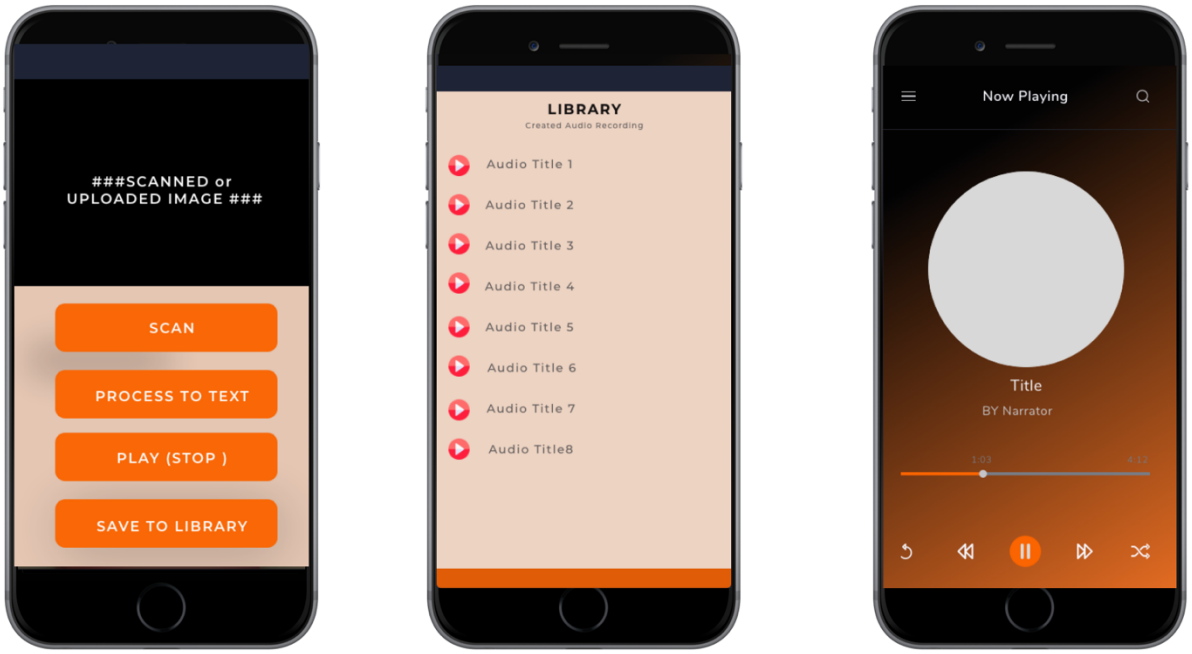


Figure 17 : UI Design 3 –created using Adobe XD



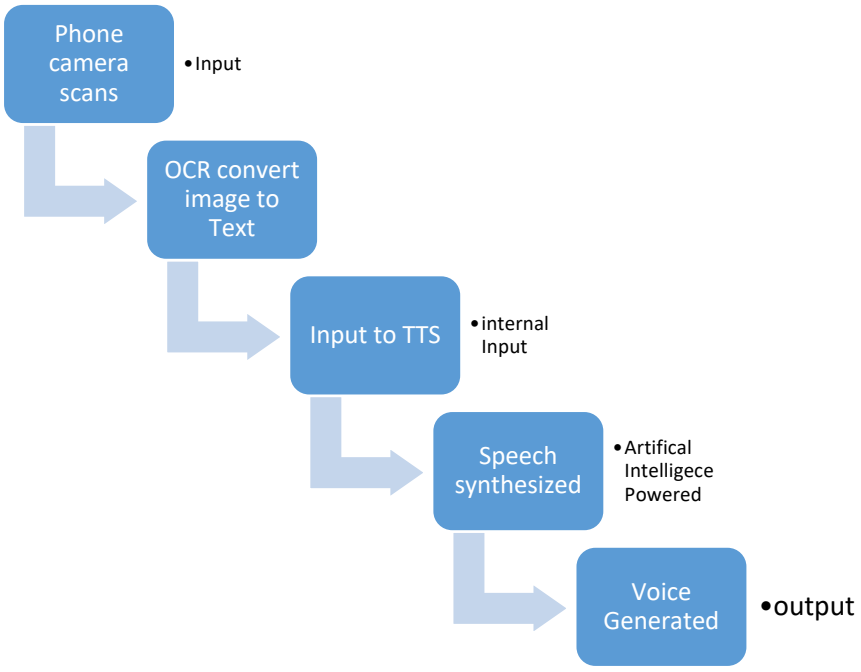
4.8 Core Development

After basic functionalities of the application were developed, the core development started, and the rest of the required features were integrated. In contrast to a single platform application, Multi-platform mobile application takes a longer time to develop. Usually, an Android based app requires more time (30% longer) than developing an IOS app. However, after extensive web research to find a way to expedite the development process, I found a programming language called Kotlin that has shorter programming and is accessible to code compared to Java. (Kotlin, 2019). Though Kotlin program is a statically-typed language like Java, Kotlin is easier to read and write, which has a shorter and simpler code than Java. As the code is more human-readable, the debug process becomes effortless. After discussing with the third-party developer, we have decided to develop the application in the Android platform.

The algorithm process flow starts with optical character recognition technology that converts the scanned image to text, and Text-to-speech technology converts the text to voice. Based on the functionality requirements for Listenapp, the application should be able to access a device's camera to capture or scan the pages/written content. Once the device grants access to capture an image using the device's camera, the app uses an optical character recognition platform to extract text from images. Listenapp uses Firebase ML (Machine Learning) Kit, a comprehensive OCR API that detects the texts from the pictures; users scan to generate the audible. ML Kit mobile SDK brings Google's on-device machine learning knowledge to the Android framework using Vision and Natural Language APIs to detect text. ML Kit's APIs comes with Digital Ink Recognition (https://developers.google.com/ml-kit/vision/digital-ink-recognition) API, which identifies printed text, handwritten text, emoji, basic shapes and shapes handwritten on a touchscreen.

(Google Firebase ML, 2019). Firebase Vision ML extracted texts are used as an input to speech synthesis. The speech synthesis process is done through an API that was already built by Replica Artificial intelligence studios. Listenapp uses replica API to initiate speech synthesis and real-time voice cloning using transfer learning from SV2TTS technology (Speaker verification to Multi-speaker Text 2 speech Synthesis) SV2TTS is a three-stage deep learning framework that creates a style transferred voice representation from a few seconds of audio. Replica API captures users' unique speech patterns, voice style, pronunciation, and various emotional ranges to generate a realistic cloned voice. To have a perfect and realistic output, a minimum of 30 minutes of speech training is required by uploading voice recordings from a pre-written script.

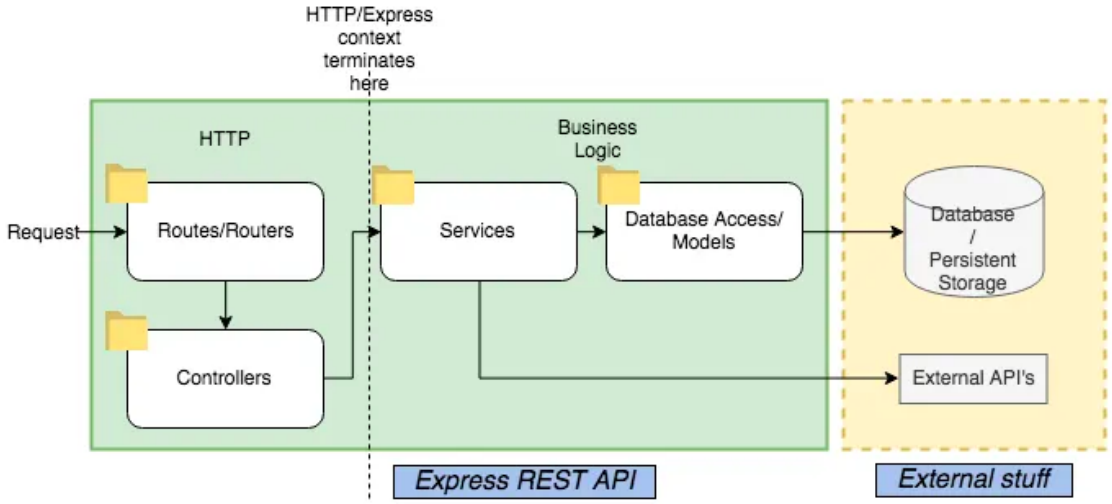
Figure 18- Algorithm Chart



To deliver such a wide variety of features, Listenapp relies on other technologies and external libraries to expose services and databases that we use in the application. These include:

- Firebase UI AUTH - Authentication, allows users to log in with existing social accounts Cloud Storage, and hosts media uploaded by users.
- ExoPlayer – ExoPlayer is an open-source project that is an application-level audio and video player for the Android platform. ExoPlayer provides the support to play generated voice on Listenapp locally or through streaming. It also enables features like fast forward, rewind, pause, and stop.
- Volley - Volley is an HTTP Client Library for Android that is efficient, flexible, and easy to use. It is a powerful Android library used for accessing any resource on the internet or cloud via HTTP. Volley makes the application run faster.
- Firebase Realtime Database- A cloud-hosted real-time database which enables data accessing between different devices and locations. Listenapp's backend data is stored as JSON and synchronized across all clients in real-time. The Firebase Realtime Database makes Listenapp a responsive and collaborative application by enabling secure access to the database directly from the device. Data is retrieved locally real-time even when the device is offline which gives the end user a dynamic experience.
- Node.js – Node.Js is a runtime environment (desktop application framework) that runs real-time JavaScript code outside a browser. Node.js enriched backend helps Listenapp to run REST API integrations smoothly.

Figure 19: Node.js REST Process flow



Chapter 5. Product Application

Students, parents can use the Listenapp, or the elderly scanning pages of a newspaper to create a newscast. It could be used by parents/ mothers by recording their own voice to read bedtime stories for their kids when they are not available. It could also be applicable for commercial purposes, such as the creation of voice over for videos. Listenapp will also be used by busy professionals, particularly those with long commutes. In the same people with learning differences or some visual disorders, this application will come in very handy for them. Listenapp can make it possible for people with dyslexia, ADHD, and low vision and related reading complications to access any texts. These texts can be read out to them through the use of the application generated text to speech expression/ or voice. It is also a mobile app that will enhance people's ability to multitask, mainly through improving their ability to absorb information faster. In the same manner, this mobile app will also play a vitally important role in as far as enhancing

people's listening skills is concerned. It will provide a medium for an active practice where people will not just make a conscious effort to listen to just the words but also the style of communication, tone, and texture, all of which will ensure that complete communication is achieved. The application makes the listener pay maxim attention to the message being delivered since they will desist from allowing anything to distract them. Equally, the application will eliminate tendencies for the listener to get bored or lose concentration while voice communication is being passed on.

It should be noted that improving one's listening comprehension skill usually requires extensive practice and is especially needed for students who undertake distance learning programs Listenapp can help such students improve this skill. It is important is emphasized through literature where it's been noted that the current practices applicable have not achieved much success because students often lose interest because of inadequate information delivery or too much load to read at a given period. With the invention of the AI-powered speech synthesis technology, which does not just clone any voice and deliver a human-like speech but also attaches auditory components like style, pitch, tone, and emotions, the Listenapp will make the students' comprehension of the learning, material much better. All of this is possible because of the Listenapp's AI-Powered backend, which eliminates tendencies of merely doing sound blurting out words. The App's capacity to add emotion, natural pauses, and a unique capability to comprehend sarcasm and punctuation make it the more appealing and most suitable for situations like improvement of active human listening skills.

In terms of functionality, it has already been noted that Listenapp is a text to speech application. It provides an easy to use user interface to ensure that the experience of the user is enhanced. More crucially, it can be listened to from almost anywhere and effectively read to the listener. It is compatible with any android device and the user can easily switch from the use of

the tablet to the mobile phone in a matter of seconds. Listenapp can literally convert anything into speech, from emails to books, articles, newspapers, and a range of other text related material. The Listenapp helps one save as much time as possible and accomplish multitasking , thereby enhancing the overall productivity of the users. Application-wise, the app is also simple to use. All it takes is a username as well as a password, and one can use it. The Listenapp is therefore fitted with a unique ability to empower with different needs and from different walks of life in several different ways.

Chapter 6. Conclusion

To conclude, the Listenapp supports the main principles of Narrative paradigm theory- Coherence and fidelity. Current text-to-speech solutions produces audibles that lack emotion and coherence. Listenapp alleviates this limitation by creating human-like voices which helps the auditory learners comprehend better. Listenapp also solves the other issues like lack of accessibility in User interface and narrator personalization.

The development process of Listenapp was a challenging one. However, it gave some insights into some of the issues that current technology faces. The main problem was segregating and structuring the ambiguous input text. The same word can often have more than one meaning. Character sets like currency symbols, special characters, times, dates, abbreviations, acronyms can pose a problem for the software to comprehend. For example, the number ‘1634’ may refer to a year ("sixteen thirty-four"), a quantity of an item ("one thousand six hundred thirty-four"), or a combination of a lock ("one six three four"), each one is read out differently. Also, words pronounced in different ways based on the tense or meaning. Like the word "read" can be

pronounced either "red" or "reed," or the word "live" can be pronounced "leave" or "live. " However, using the Artificial intelligence and machine learning capacity of Listenapp's back-end, the application could comprehend and read the most appropriate word, providing auditory learners with a superior user experience.

Moving forward with this project, I would like to add more features to the application such as supporting different languages, in-built translator, increasing more voice choices from a range of age groups, gender, ethnicity and accents. Also, users would be able to choose the tone of the narration by being provided with options to set the tonality with traits such as serious, joyful, and melancholic. Another major feature that is currently disabled on the existing version of Listenapp is voice cloning option. Future version will have this option enabled whereby users will be able to create a custom voice within 30 minutes of voice training.

Finally, I want to publish the application on Google Play store as a free to download app so that it will help student community use the app without putting a hole in their pockets.

References

Antle, A.N.; Chesick, L.; Mclaren, E.S. (2018) Opening up the Design Space of NeuroFeedback Brain-Computer Interfaces for Children. Trans. Compute. -Hum. Interact., 24, 1–33.

Boitnott, J. (2017). "How One Founder Turned His Dyslexia Into an App That Helps People With the Disability Learn Faster". Inc. Southeast Asia. Retrieved 2019-08-30.

Boitnott, J. (2017). "This Immigrant Founder Taught Himself English--Then Made an App That Helps Others With His Disability (and Speed Readers)". Inc.com. Retrieved 2019-08-30.

Bright, T.; Pallawela, D. (2016) Validated smartphone-based apps for ear and hearing assessments: A review. JMIR Rehabil. Assist. Technol. 3, e13.

Chergui, O., Begdouri, A & GROUX-Leclet, D. (2016). A Classification of Educational Mobile Use for Learners and Teachers. International Journal of Information and Education Technology. 7. 10.18178/ijiet.2017.7.5.889.

Facts and Figures on Adults with Vision Loss. (2019, March 1). Retrieved July 1, 2020, from https://www.afb.org/research-and-initiatives/statistics/adults

Fearn, N & Turner, B. (2020). Best text to speech software of 2020: Free, paid, and online voice recognition apps and services. Retrieved from https://www.techradar.com/best/best-text-to-speech-software on July 23, 2020

Fisher, W.R. (1984). Narration as a human communication paradigm: The case of public moral argument. DOI:10.1080/03637758409390180

Fonseca, C., Castro, E., Ramos, J., & Rodriguez, P. (2017, August). Usability of Mobile Applications for Visually Imparied People: An Empirical Study. Jornadas Iberoamericanas de Ingeniería de Software e Ingeniería del Conocimiento (JIISIC), Lactacunga, Ecuador.

Gunda, Naga Siva Kumar & Gautam, Siddharth & Mitra, Sushanta. (2019). Editors' Choice—Artificial Intelligence Based Mobile Application for Water Quality Monitoring. Journal of The Electrochemical Society. 166. B3031-B3035. 10.1149/2.0081909jes.

Karat, C. Vergo, J. Nahamoo, D. (2007). "Conversational Interface Technologies". In Sears, Andrew; Jacko, Julie A. (eds.). The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications (Human Factors and Ergonomics). Lawrence Erlbaum Associates, Inc. ISBN 978-0-8058-5870-9.

Kayalar, Filiz & Kayalar, Fethi. (2017). The effects of Auditory Learning Strategy on Learning Skills of Language Learners (Students' Views). IOSR Journal of Humanities and Social Science (IOSR-JHSS). 22. 10.9790/0837-2210070410.

Kraljevski, I., Chungurski, S., Stojanovic, I & Arsenovski, S. (2010). Synthesized Speech Quality Evaluation Using ITU-T P.563. 590-593.

Kons, Z., Shetchman, S., & Sorin, A. (2019, October 24). High quality, lightweight and adaptable Text-to-Speech (TTS) using LPCNet. Retrieved July 10, 2020, from https://www.ibm.com/blogs/research/2019/09/tts-using-lpcnet/

Lee, Y. (2017). Mobile Application Development for Improving Auditory Memory Skills of Children with Hearing Impairment. Audiology and Speech Research, 13, 50-61.

Magfirah, Titing. (2018). Students' reading and listening comprehension based on their learning styles. International Journal of Education. 10. 10.17509/i.e. v10i2.8028.

Meyer, A., Rose, D.H., and Gordon, D. (2014). Universal Design for Learning: Theory and Practice. Wakefield, MA: CAST Professional.

Mugayi, T. (2019). Ever Wanted to Build a Text-to-Speech App? Retrieved from https://medium.com/better-programming/ever-wanted-to-build-a-text-to-speech-app-115dee18f8c4 on July 23, 2020

Nusbaum, H.C., Francis, A.L. & Henly, A.S. (2015). Measuring the naturalness of synthetic speech. Int J Speech Technol 2, 7–19. https://doi.org/10.1007/BF02215800

Nwakanma, I., Oluigbo, I & Izunna, O. (2014). Text – To – Speech Synthesis (TTS). International Journal of Research in Information Technology, Volume 2, Issue 5, May 2014, Pg: 154-163. 2. 154-163.

Nguyen, G. (2019, August 15). Podcast listening growth continues: Mobile app usage up 60% since January 2018. Marketingland.Com. https://marketingland.com/podcast-listening-growth-continues-mobile-app-usage-up-60-since-january-2018-study-finds-265608

Özdamar, K & Metcalf, D. (2011). The current perspectives, theories, and practices of mobile learning. Turkish Online Journal of Educational Technology. 10.

Parr, M. (2012). The Future of Text-to-Speech Technology: How Long Before it's Just One More Thing we do When Teaching Reading? Procedia - Social and Behavioral Sciences, Volume 69, Pages 1420-1429

Pieraccini, R. (2012). The Voice in the Machine. Building Computers That Understand Speech. The MIT Press. ISBN 978-0262016858.

Stabile, L. (2017). "Five Questions With Cliff Weitzman". Providence Business News. Retrieved 2019-08-30.

Strangman, N., Dalton, B. (2005). Technology for struggling readers: A review of the research. In Edyburn, D., Higgins, K., Boone, R. (Eds.), The handbook of special education technology research and practice (pp. 545–569).

Takano, N. (2019). Comparing 4 PDF Text-To-Speech iOS Apps: Voice Dream Reader, Capti Voice, NaturalReader, and vBookz PDF. Retrieved from https://naoko.blog/2019/02/11/comparing-4-pdf-text-to-speech-ios-apps-voice-dream-reader-capti-voice-naturalreader-and-vbookz-pdf/ on July 22, 2020

Woelfel, M., McDonough, J. (2009). Distant Speech Recognition. Wiley. ISBN 978-0470517048.

Wood, S. G., Moxley, J. H., Tighe, E. L., & Wagner, R. K. (2018). Does the Use of Text-to-Speech and Related Read-Aloud Tools Improve Reading Comprehension for Students With Reading Disabilities? A Meta-Analysis. Journal of Learning Disabilities, 51(1), 73–84. https://doi.org/10.1177/0022219416688170