MyConvoPal: A Mobile Application for Conversational English Pronunciation Training

By

Bridget Green

Undergraduate Honours Thesis

In

Computer Science

Website: http://vialab.ca

E-mail: bridget.green@ontariotechu.ca

Abstract

Conversational English is one of many challenges faced by those hoping to improve their fluency in spoken English. What makes it a subject of interest is that it presents a gap in existing pronunciation training and language learning services. This thesis presents a project made with this concept in mind, called MyConvoPal. MyConvoPal is a mobile pronunciation training application developed to provide an experience focused on conversational English, as well as other specific areas in English pronunciation that one might find challenging. Within the app, users can choose from a list of pronunciation exercises based on what they would like to practice. The games offered can be divided up into two types- word games, where individual words are practiced; and sentence games, a more advanced version of the word games where users practice pronunciation in the context of an entire sentence. In each game, the user is given a randomly selected prompt based on the area of pronunciation they have chosen to focus on. The user then records their attempt at pronouncing the prompt and is given feedback based on how well they pronounced the word or sentence. The app keeps track of how much progress the user has made while doing these exercises and allows the user to see how much they have improved in certain areas. The app also provides user customization, allowing the user to customize their experience in the exercises as well as app appearance.

Table of Contents

List of Figures

1 Introduction

When learning a new language including English four main competencies are being trained: reading, writing, listening, and speaking. While understanding linguistic forms such as conjugation and word order is important for all of them, each of them brings a different set of challenges to learners. For example, for writing, one might need to learn orthography (how words are spelled) in addition to grammar. Similarly, for speaking, learners need to be able to pronounce or produce the phonemes, or sounds, from the target language. Traditionally, pronunciation is acquired via practice, and experience. However, given the available software for pronunciation training and aside from practicing conversations with natives of the target language, learners lack support in practicing conversational language.

Thus, this thesis presents MyConvoPal a mobile application for training conversational English sentences and slang. MyConvoPal also allows learners to create and curate their list of words or sentences for training. When training, MyConvoPal offers a simple activity where learners can listen and repeat words or sentences, it then displays feedback containing which phonemes and syllables the learner has the most difficulty with. Based on the learner's rating on the word (easy, medium, hard), MyConvoPal uses a spaced-repetition algorithm to suggest and recommend the next training session for that word.

1.1 Motivation

Existing pronunciation training applications do not support the development of conversational English. Most pronunciation training applications support a formal way of speaking, and not so much how one would speak casually. Those that do cover more conversational English are typically more expensive, and conversational English is only a small part of what is focused upon. Existing pronunciation training applications also lack training for specific parts of English pronunciation. This includes words spelled similarly but pronounced differently, inconsistencies in pronunciation rules, and "slang".

Spaced Repetition is also uncommon in existing pronunciation training applications. It does exist in some paid applications, but not typically in ones that are more financially accessible. Spaced Repetition is important to consider as something that helps someone not only learn material but remember it. Therefore, I designed MyConvoPal to be able to enforce elements of Spaced Repetition.

1.2 Problem Statement

There are existing pronunciation training applications, but they are often missing an important part of spoken English communication. When using these applications, a user will get experience practicing the formal English language. When researching aspects of the English language that make learning it particularly difficult, the use of "slang" is often a hurdle (Berdina, 2023). Those who use most existing pronunciation training applications may be under-prepared for casual, everyday conversation. I was able to find one pronunciation training service that covered casual conversation, but this service was a subscription that charged an excessive cost monthly. Those who would be using this service would be getting only a subsection that covered casual conversation while being charged large sums of money. Because of this, I designed MyConvoPal to focus on casual speech.

This, however, would make for a sparse number of exercises for the user. I decided to also have the application cover other aspects listed as hurdles when learning the English language. According to English Second Language learners (Oxford Royale Academy), one such hurdle is the numerous technicalities in the English language. To give an example, it is understood that the letter '*c*' can be pronounced with a hard '*kuh*' sound. However, '*c*' can also be pronounced with a softer '*s*' sound. To those who are familiar with speaking English, it is easier to make the distinction between these two uses based on what letters follow '*c*.' However, this can complicate the process of learning the pronunciation of English words and phrases. Another complication in developing pronunciation is the inconsistencies that would be perceived between similarly spelled words. For instance, the words '*tough*' and '*dough*' are spelled in such a way that suggests a similar pronunciation, when this is not the case. By incorporating all these variables, it would produce a mobile pronunciation training application targeting the hardest and underrepresented aspects of learning the English language.

1.3 Target Audience

This application would be aimed at those who have some familiarity with the English language. The functionality would focus on improving pronunciation, but not teaching the English language. Ideally, the app would do both, however, my app focuses on pronunciation training. Other aspects can be incorperated in future work. While the app contains features to help someone understand how an English word should sound, it will not teach them the meaning of each word and phrase. While this application would be developed with English second language learners in mind, it can also be useful to anyone hoping to provide their pronunciation in the English language, or by people who are born with or who have developed speech impediments.

2   Related Work

This chapter discusses related work and existing language learning applications, including general language learning and pronunciation training applications. I also present SpeechSuper, an API that is the base for speech recognition and feedback used in MyConvoPal.

2.1 Language Learning Applications

Duolingo

As a part of my research for this project, I tested out different existing applications of language learning and pronunciation training for the English Language. One application for language learning was Duolingo. Duolingo is a popular language learning app that offers a limited amount of content for free. It is also very accessible, being available on multiple platforms. The way the app works is to present the user with tasks to accomplish, such as repeating the phrase they see on the screen or filling in the blank in a sentence. This presents an effective way to help someone learn how to speak a new language and how to improve their pronunciation. However, Duolingo is also detrimental to learning. With the usage of active repetition and irrelevant phrases, Duolingo offers a system that is detrimental to one's conversational fluency (Cohen, 2023). When Duolingo provides a phrase for users to say, it is simulating a theoretical situation one would be in where they would be communicating in another language. However, many of Duolingo's simulated phrases are basic and are not reflective of how everyday conversations might go. For example, Duolingo might give a prompt such as, "Yes, a large coffee!" Which is beneficial for specific circumstances but not when you are having a conversation with someone.
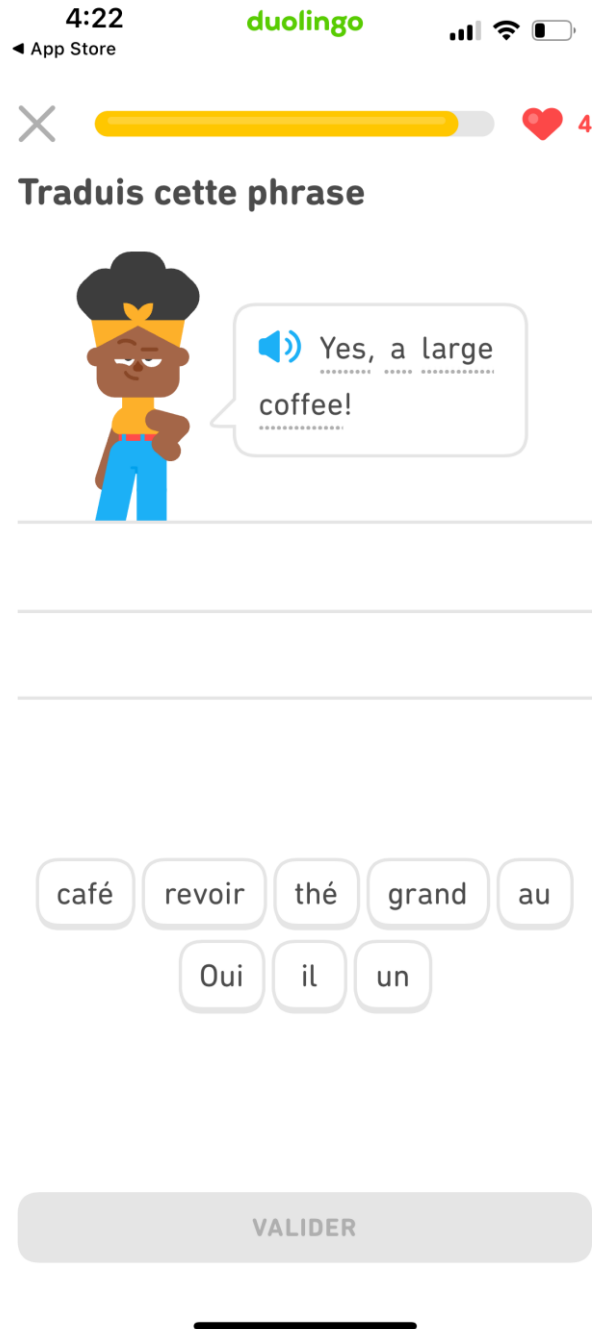
Figure 1: Duolingo

Figure 1 is an example of a prompt in Duolingo. When I tried out Duolingo, I identified myself as a French speaker who is learning English. Something I found concerning was that the app would use grammatically incorrect sentences in French and that the examples of English phrases sounded artificial. If this is the experience offered in every language, it would complicate the process of developing conversational English.

FluentU

   FluentU is another language learning service. It is a subscription service that charges 29.99$ a month or 239.99$ a year and is available to download on both Google Play and the Apple App Store. It provides the user with several features to help with learning a new language and improving one's fluency. These features include captioned videos, definitions, videos with interactive subtitles, and personalized quizzes. This offers a very personalized experience for the user and would be immensely useful. That being said, FluentU does not necessarily provide much focus on conversational English. There is also the issue of it being a subscription. While FluentU would be immensely helpful to someone learning how to speak fluently in English, much of its content is locked behind a paywall, making it limited in accessibility.
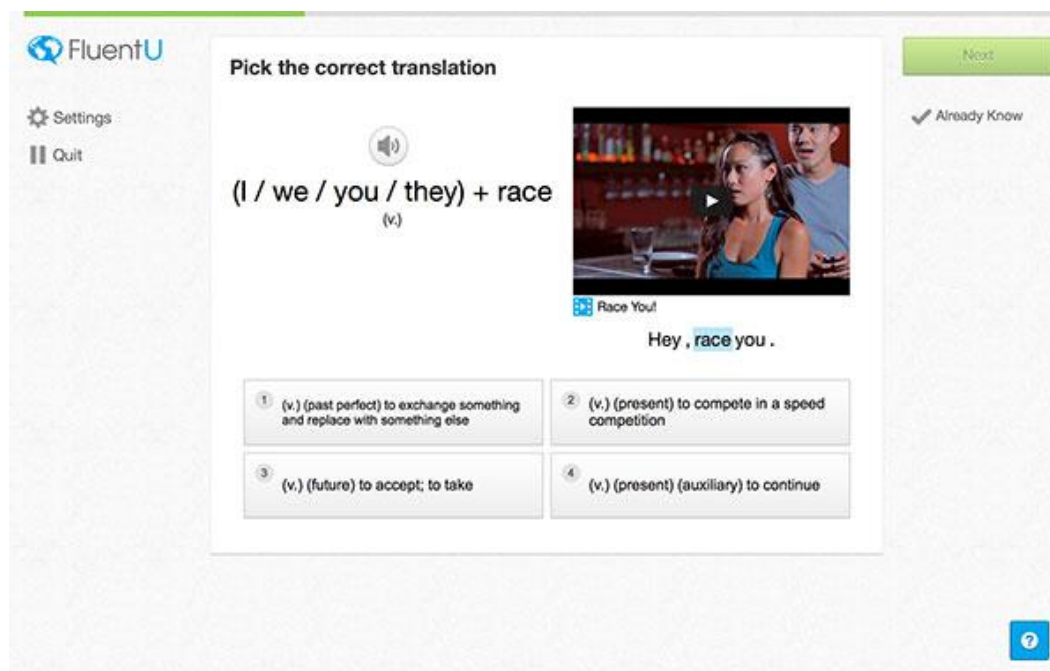

Figure 2: FluentU

   When using it myself using a free trial period, it was incredibly hard to use without being constantly prompted to buy the full service. The content I could access was limited, and I could not get most of the available exercises without paying. For someone who is looking to improve their pronunciation, this could serve as a serious accessibility issue.

2.2 Pronunciation Training Applications

Quick Pronunciation Tool

Quick Pronunciation Tool is a free mobile application with in-app purchases available for download on the Apple and Google Play Store. For clarification purposes, the free version was used here. Quick Pronunciation Tool allows users to type in any word or phrase into the app,

where it will be repeated back to them. While this is a satisfactory solution for those already familiar with English enough to navigate the app, this was best used for getting the pronunciation for singular words. There are a limited number of characters, which limits the number of situations where this could help. The voice used for the pronunciation reference is a text to speech voice, which by itself is helpful, but is not perfect. The app offers the option between British and American pronunciation, but when using it myself, there was no difference between the two.
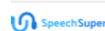
Speak English Pronunciation

Speak English Pronunciation is another free application that can be downloaded on both the Apple and Google Play Store, although Speak English contains ads. Here, a user types in a word they want to pronounce. The app, in return, provides an audio example of how to say said word, definitions of the word, and returns feedback on how well the user did. The app even contains pronunciation tests that the user can try out. Speak English is an incredibly interesting app, as it shows how exactly to pronounce sounds by pinpointing specific areas of the mouth. If there was something like this focused on something like conversational English, this would be immensely helpful for those looking to improve their pronunciation.

2.3 SpeechSuper

How It Works

SpeechSuper is a platform that offers both APIs and SDKs for speech and pronunciation assessment. With the use of AI, SpeechSuper can deliver assessments of an audio recording, specifically on the pronunciation in said recording. When using the API in a program, the process is fairly cut and dry. First, an audio recording must be made. Next, API input must be requested. In this request, there are variables that need to be specified, such as the language being used, information about the audio file, what word or phrase the audio is supposed to contain, and what service is being requested of SpeechSuper. Then, SpeechSuper will return scores to reflect on how the audio file performed the pronunciation. These scores reflect things such as fluency, speed, words per second, number of pauses, and scores on individual parts of the word or phrase said.

Figure 3: SpeechSuper Recording Prompt

Figure 4: SpeechSuper Feedback
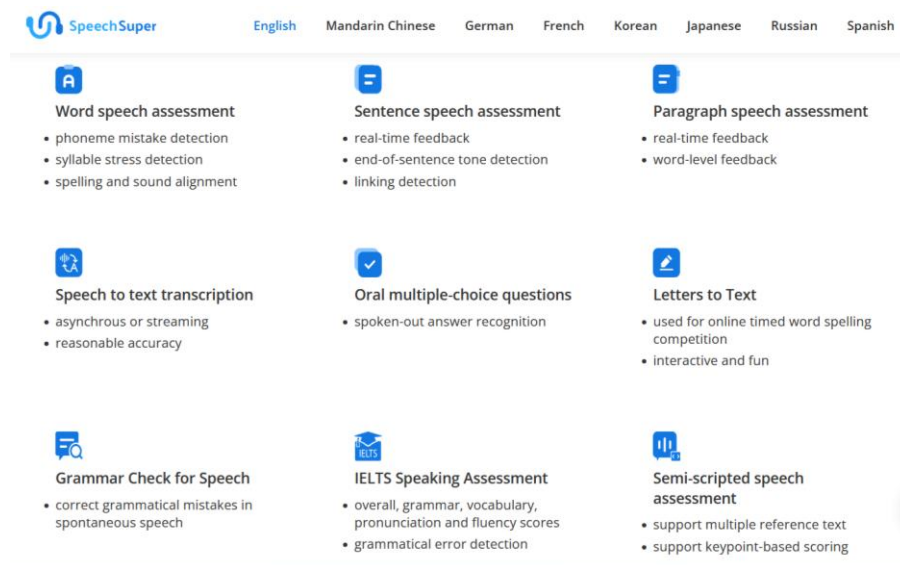
What is Offered

Figure 5: SpeechSuper Demo Main Screen

As shown in Figure 5, SpeechSuper offers assessments for single words, sentences, paragraphs, and even real time input. It offers this not only for the English language, but also several other languages. This includes French, German, Mandarin, and Spanish. It even offers different English dialects. A user can choose whether they wish to be assessed against American English, British English, or Indian English. The user can even adjust the leniency of the assessment.

SpeechSuper also provides speech-to-text, and grammar assessment.

How It Can Be Used

In the demo provided by SpeechSuper, there is already a collection of examples of how this API can be used. For example, it can be used as practice for an IELTS Speaking Assessment, or for a trivia game where the user responds verbally.

For the purposes of this project, SpeechSuper can be used to help develop both conversational English and other aspects the app will be covering. The word assessment can be used to help a user gain practice pronouncing words with letters that have more than one sound, or words that look similar but sound different. The sentence assessment can be used to help develop conversational English by assessing how well the user performs when prompted with casual, everyday phrases.

3   Designing MyConvoPal

This chapter goes through the development of MyConvoPal. This includes software implemented within the app, early versions of the app, and the additional features.

3.1 Initial Design

To construct MyConvoPal, Flutter was used as the base for coding. Flutter is an open-source framework that allows for the development of mobile applications that can be accessed with multiple mediums. Flutter apps can work with Apple, Android, or even a web browser. This already made Flutter an appealing choice to use for MyConvoPal. Flutter also provides several functions and widgets that can be easily imported into the code.

Initial Game Structure

The application design needed to be simple, appealing, and beneficial to the user. There had to be a simple and comprehensive layout that the user could easily navigate and understand.

Figure 6: Early Version of MyConvoPal Word Game

Functionality

Prompts

      I took inspiration from the appearance of SpeechSuper's demo screen and thought of a game where a user would be prompted with a word to pronounce, and their results would appear on in a table below, as seen in Figure 6. For this, I used lists of words and sentences, unique for every game, that would be randomly chosen to appear on in a prompt box on the game screen. These lists would be constructed of words and sentences that contained sounds that would be difficult for someone learning to speak the English language, and the sentences would be constructed with casual speech. These sentences would be examples of what may come up in everyday conversation.

Connection to API

      To receive their results, the user would hit a button on screen to start and stop recording. This recording would be stored within the external directory of the user's phone, where it could

then be used to evaluate. Then, the recording would be sent to the API to return feedback. When SpeechSuper receives audio, it returns results such as the fluency, the speech, and the overall pronunciation of the prompt.

Display Results

The results would be displayed at the bottom of the screen, in a format not dissimilar to how it is presented in the SpeechSuper demo. From there, the user would be presented with the next prompt. This made up the base of every game in the application, which would then be built upon during the development process.
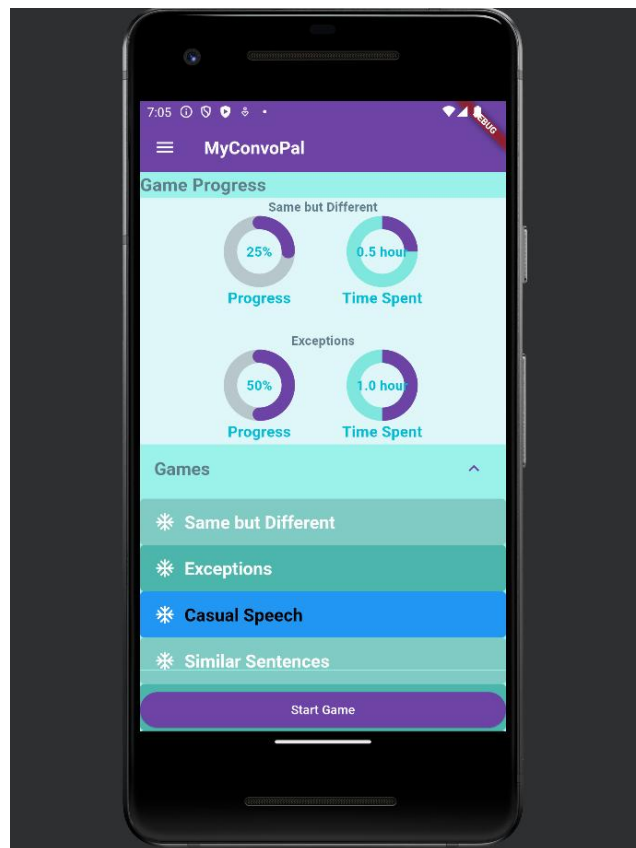
3.2 Main Screen

Figure 7: MyConvoPal Main Screen

Layout

I wanted the different game options to appear on the main screen to give the user a list of options to choose from, so that the user could choose what to focus on in their training session. I also wanted to have the user be able to see their progress in each game. I decided to have the screen layout as an area for the user to see their most recent progress and see how much they have been improving. This was done by way of progress rings that would show the overall score

the user had gotten on their previous attempt at a game, alongside the amount of time spent. The user could scroll up and down to view the progress for each game and scroll sideways to view the progress made in past attempts. The lower part would be a list of tiles of potential games that the user could scroll through.

User Progress

In order to store the user progress that would be displayed on the main screen, a development platform known as Firebase was used. When developing a Flutter app, data input by the user can be lost between reloads. Firebase can be used to store data from past usages of a Flutter app and can thus solve this issue. A Firebase database was created for MyConvoPal, and a class was created to store progress made during an exercise. Functions were made inside the app's code that allows data to be easily sent to and received from the database. This is what allowed the past progress made in each of the games to be viewed.
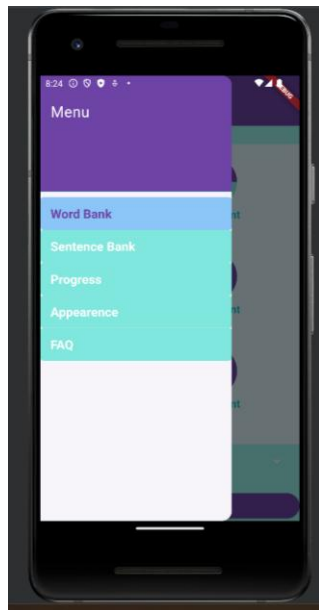
Manual Customization

Figure 8: MyConvoPal Menu

A menu was added into the main screen to provide tools that would benefit the user. Such tools included an FAQ page, and more notably, word and sentence banks.
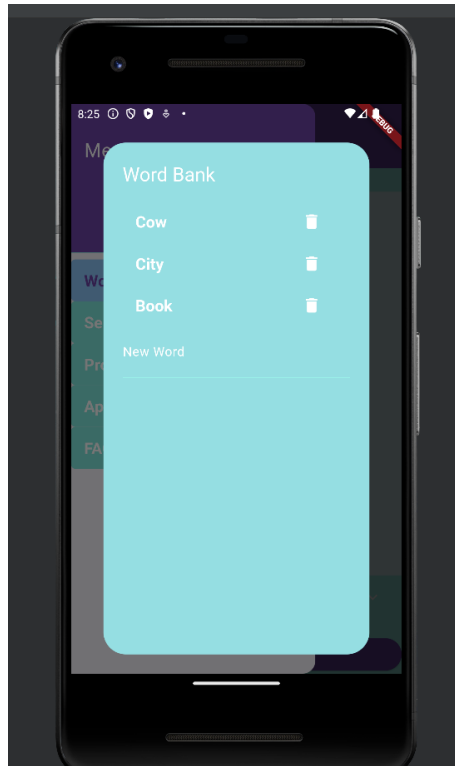
Figure 9: MyConvoPal Word Bank

As shown in Figure 9, the word and sentence banks served to allow the user to customize their experience. Two of the games within the app are Custom games, where it is completely dedicated to having the user practice specific words or phrases they want to improve on. This can be either something that they have been struggling with in the exercises, or something that they have encountered in everyday life. In the Custom games, the word or sentence lists are populated by what is stored in the word and sentence bank, respectively. This menu serves as one way for the user to populate either bank. The menu also offers a feature added for accessibility purposes.

Figure 10: MyConvoPal Theme Selection
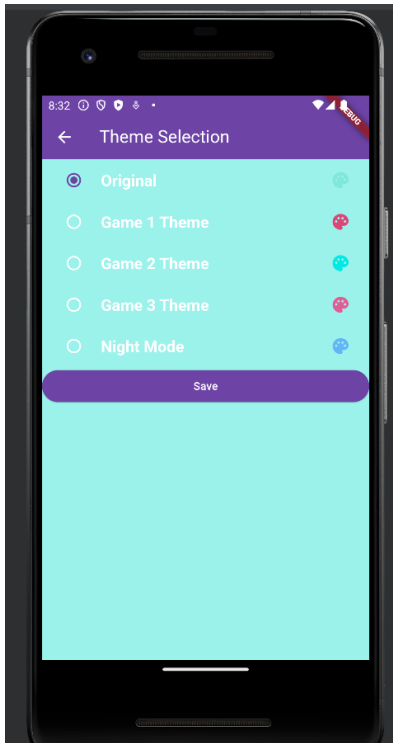
When choosing the colours for MyConvoPal, a pallet of cool colors was utilised. This was because cool colours typically encourage more relaxed feelings as opposed to warm colours. However, it would be beneficial to offer other colour pallets that the user can choose from. With this, the user can customize the appearance of the app to be to their liking aesthetically.
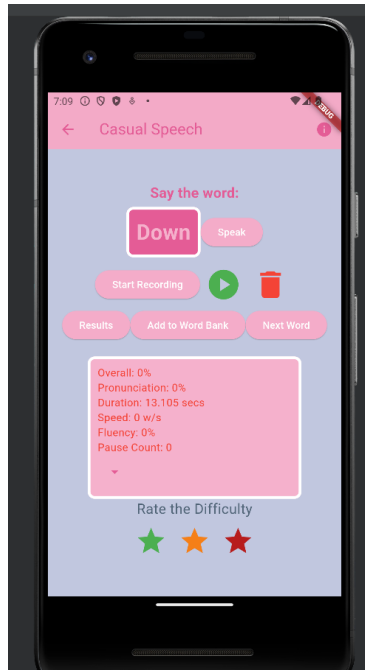
3.3 Game Design

Figure 11: Word Game Screen

As work progressed on the games, additional features were added. One such feature was a playback option, where the user could hear their recording before submitting it for results. The recordings made in the app would also be deleted once their use in the app had expired. This would ensure a more ethical use of the recording feature. Another feature was a button for the user to decide when to progress to the next word or sentence. A user can choose if they want to have another attempt at a word, or to progress to the next prompt.

SpeechSuper offers the user individual scoring on each sound in a word or a phrase. Since this would be useful for the user, an arrow icon was added to the bottom of the container displaying the results. When selected, this would replace the overall pronunciation scores with the individual scoring for each sound. The user can toggle between both sets of results with this icon button.

To help clarify how well someone did on a prompt, a feature was added that colored each individual line in the results either red, orange or green. If the user scored below average, the text would be red. If they performed average, it would be orange. Otherwise, it would be green.
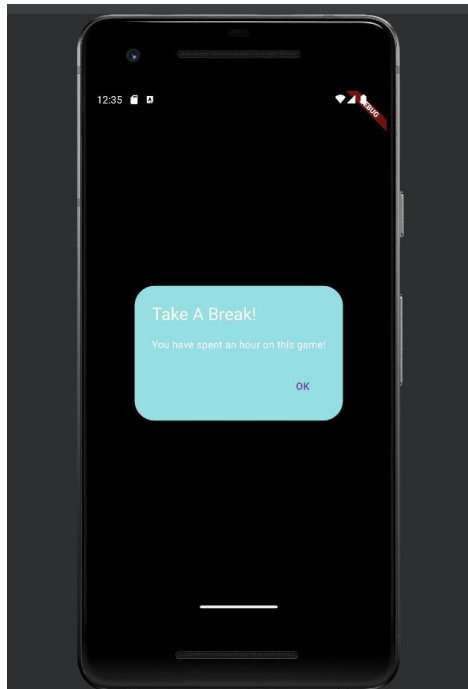
3.4 Additional

Figure 12: MyConvoPal Timed Alert

While testing out existing pronunciation training applications and services, something I noticed was a lack of healthy learning habits being promoted to the user. I wanted to implement a feature that encouraged the user to take breaks and not spend several hours at a time practicing. To do this, an alert system was created. Whenever a user had spent an hour on a game, a pop up would appear, encouraging the user to take a break. In doing this, it not only prevented the user from spending hours upon hours on one game, but it also served to benefit the user's learning. When learning and memorizing information, taking breaks in between sessions of repetition aids in this process.
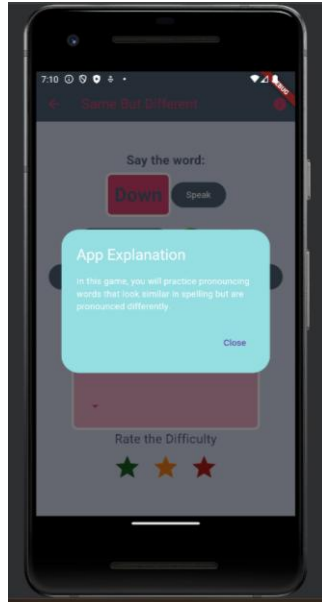
Figure 13: Pop-up Explaination

I also used popups to give an explanation option for users. When playing a game, an icon would be provided in the app bar. Upon selecting this, it would cause a popup to appear, which explained what the user would be doing in this game. A unique explanation is applied for each game.

4   Future Work

4.1 Games

In terms of what I would add if I were to continue working on this project, I would implement more variety in games that the user could choose from. I implemented a beta version of a karaoke game, which is just one example of how SpeechSuper's API can be used. The idea is that the user could input lyrics to a song they enjoy and sing it to receive feedback from the API. Another potential game could be a trivia type game, similar to the one found in the demo provided on the SpeechSuper site.

In the currently existing games themselves, I would add options to choose from American dialect, British dialect, etc. It is something that SpeechSuper offers, and this was a feature in existing pronunciation training app such as Speak English Pronunciation that I found quite interesting in terms of increasing accessibility. The text-to-speech feature I implemented also uses this feature and contains more dialect options than SpeechSuper.

4.2 Main Screen

For the main screen, more customization for user experience would be added. There would be more features dedicated to recommending specific games based on the user's needs.

4.3 Additional

If MyConvoPal were to be tested for its efficiency today, it could be done in a couple ways. One would be to have people test out the app, and then submit their feedback on it. Specifically, did they feel as though the app helped them improve their pronunciation? There is also the other option, which would be to use Firebase storage. Since the app stores the user's progress while they practice, this could be used to see if the app is working as intended. This would have to be done with the user's consent to the use of their progress data, of course.

5 Conclusion

MyConvoPal offers a medium for a user to practice their pronunciation, with a focus on underrepresented aspects of the spoken English language. The app is aimed towards those looking to improve aspects such as their enunciation and fluency in such a way that it makes engaging in casual conversation easier. The user can submit their attempts at a word or a sentence to receive useful and detailed feedback to help them improve. There is a list of games provided in the app that are geared to help with some of the most difficult areas in English pronunciation. The user can choose whether to practice individual words or attempt the pronunciation of specific words within the context of a full sentence. MyConvoPal also offers features that allow users to customize their experience in the app. The user can influence what prompts they are given to practice; they can input specific prompts they wish to receive; and they can even customize the appearance of the app. The app stores and updates the user's progress in each game and allows them to view how much progress they have made and see how much they have improved in or are lacking in specific areas. There are also features in place to encourage the user to practice healthy learning habits. Future work would include the implementation of more variety in the games, more features to implement spaced repetition, more user customization, and even a fully functional tutorial.

Bibliography

Berdina, A. (2023). Why is English hard to learn? OIDigital Institute. https://oidigitalinstitute.com/news/why-is-english-hard-to-learn/

Chen W. (2022). An Analysis of the Effects of the English Language and Literature on Students' Language Ability from a Multidimensional Environment. Journal of environmental and public health, 2022, 3225096. https://doi.org/10.1155/2022/3225096 (Retraction published J Environ Public Health. 2023 Sep 27;2023:9898629)

Cohen, A. (2023). ASLearner: An American Sign Language Spaced-Repetition Learning App. University of Ontario Institute of Technology

Duolingo. (n.d.). Retrieved from https://www.duolingo.com

FluentU. (n.d.). Retrieved from https://www.fluentu.com

Hudson, J. (2013). English difficult to pronounce? Pronunciation Studio. https://pronunciationstudio.com/english-difficult-pronounce/

Language Doctors. (n.d.). Is English hard to learn? The Language Doctors. https://thelanguagedoctors.org/is-english-hard-to-learn/

Oliveira, S. (n.d.). Confusing English pronunciation: Words that look similar but sound different. https://www.espressoenglish.net/confusing-english-pronunciation-words-that-look-similar-but-sound-different/

Oxford Royale Academy. (n.d.). Is Learning English Hard? https://www.oxford-royale.com/articles/learning-english-hard/

Quick Pronunciation Tool. (n.d.). Retrieved from Google Play Store. (https://play.google.com/store/apps/details?id=com.tiago.tspeak&hl=en_CA&gl=US)

Salsabilla, S. (2021). Why is English Hard to Learn for Students? https://www.researchgate.net/publication/355940168_WHY_IS_ENGLISH_HARD_TO_LEARN_FOR_STUDENT

Speak English Pronuncaition. (n.d.). Retrieved from Google Play Store. (https://play.google.com/store/apps/details?id=com.tflat.phatamtienganh&hl=en&gl=US)

Tamm, S. (2023). Spaced repetition: A learning technique that incorporates increasing intervals of time between subsequent review of previously learned material. https://e-student.org/spaced-repetition/

Thompson, S. (n.d.). Speaking clearly: Activities for improving English pronunciation [PDF]. https://educapes.capes.gov.br/bitstream/capes/62851/1/Speaking%20Clearly%20Activities%20for%20Improving%20English%20Pronunciation.pdf

Yama, S. (2022). Card-IT Versus: A Competitive Multiplayer Game for Testing Italian Verb Morphology. University of Ontario Institute of Technology

arXiv. (n.d.). English pronunciation: A computational analysis. https://arxiv.org/abs/2207.00774