Visual Feedback for Language Production

Author: Byron Hui

Supervisor: Dr. Catherine Watson

University of Auckland, Private Bag 92019, Auckland, New Zealand

Faculty of Engineering, Specialised in Software Engineering

Student Statement

Participating in this Summer Research Scholarship has allowed me to gain valuable experience and insight into what it is like working on a postgraduate research project with the University of Auckland. This experience will allow me to make more informed decisions as to decide whether or not to continue on to a postgraduate degree.

The research aspect of the project has been beneficial to me as it has allowed me to hone my research skills, which I can apply for my Part 4 Engineering Project later this year. Working on the scholarship has also allowed me to acquire new skills in a variety of new technologies while also reinforcing knowledge gained through the first three years of the Engineering degree. The Scholarship also has taught me valuable non-technical skills such as time management, self-discipline and interpersonal skills which will help me throughout my career, regardless of which path I choose.

I would like to thank my supervisor, Dr. Catherine Watson, in which I had the pleasure of working and learning under, for giving me the opportunity on working on this project and for letting me draw on her experiences as a professional engineer to further my development as an engineer. I'd also like to thank the MAONZE group for their continued support of the project.

Abstract

There is a strong demand for a computer-based application which provides visual feedback on vowel pronunciation as it can help people improve their pronunciation of a language. Vowel formant information provides useful real-time feedback for vowel pronunciation, but the information is currently inaccessible to users who aren't experienced in the field of speech processing. This project explores ways in which a computer-based language pronunciation aid can be developed so that a non-expert user can easily use the application to improve their vowel and language pronunciation.

This report describes MPAi, a software application which provides a mechanism for which users can practice their pronunciation of the Mäori language. The report mentions the motivation and aim of the project in sections 1 and 2. Then it will give an outline of MPAi and what it can do in section 3. Section 4 is concerned with how MPAi was developed and section 5 gives a summary of an informal usability test of the application. The report concludes with recommendations on what could be added to the application in the future.

Table of Contents

List of Figures

1.0 Introduction

Mäori is one of the official languages of New Zealand. However, the language is in danger of being lost as English is the country's predominant language [1]. There are computer-based applications, such as RosettaStone [2], which provides users a tool for learning specific languages. However, as Mäori is generally only used in New Zealand, there is currently no known language pronunciation aid that caters to the Mäori language since there isn't a high demand for one. In response, there have been two computer-based applications done in the past at the University of Auckland which have attempted to solve this problem.

The first application, named the Formant Aid, provides real-time feedback on vowel pronunciation by displaying vowel formants onto a formant plot. The Formant Aid can present very useful information to someone with expert knowledge of this field, but is extremely inaccessible to non-expert users.

The second project, named Mäori Pronunciation Aid (MPAi), developed at the University of Auckland by Daniel Rivers and Jacinth Gutla under the supervision of Dr Catherine Watson, allows users to practice pronunciations of Mäori words (Mäori place names) used in everyday conversation. This application was more user-friendly than the Formant Aid application, but the functionality provided by the application was limited in terms of providing feedback to the user on how well they pronounced a particular Mäori word. MPAi allows a user to compare the time taken to pronounce each phoneme of a chosen word with an expert Mäori speaker's times, which is useful, but it doesn't take into account the accuracy of how each phoneme was pronounced.

2.0 Aim

The objective of the research project is to develop an application which continues on the work done on the Formant Aid and MPAi. This will involve implementing existing features already found in the Formant Aid and MPAi, while also adding improvements and additional features.

The target users of the application will be users who are interested in using a computer-based aid to help them with their vowel and word pronunciations for the Mäori language, therefore, the application will require a new user interface which is more visually appealing than the

previous Formant Aid so that the program is more accessible to non-expert users and to also find a new method of showing the formant plots so that users will be better able to understand the information given in regards to their vowel pronunciation. There will also be work done on finding ways to improve the real-time accuracy of the Formant Aid which consists of dealing with the different qualities of audio input received from different microphones and the speed at which the Formant Aid processes the user's audio data.

The project will then attempt to develop a new method of analysing word pronunciations by comparing a user's pronunciation of a given Mäori word with an expert Mäori speaker's version of the pronunciation. This will involve using the Mäori Forced Alignment Program (based on the Hidden Markov Toolkit (HTK) [3]), which was developed by Sam Abbott under the supervision of Dr Catherine Watson in the University of Auckland, to gather the required data needed to produce a score which shows how well the user's pronunciation of a given Mäori word is in reference to a gold standard. Python is the programming language used in this project as it is known to be compatible with sound toolkits that provide various functionality needed for the development of the application.

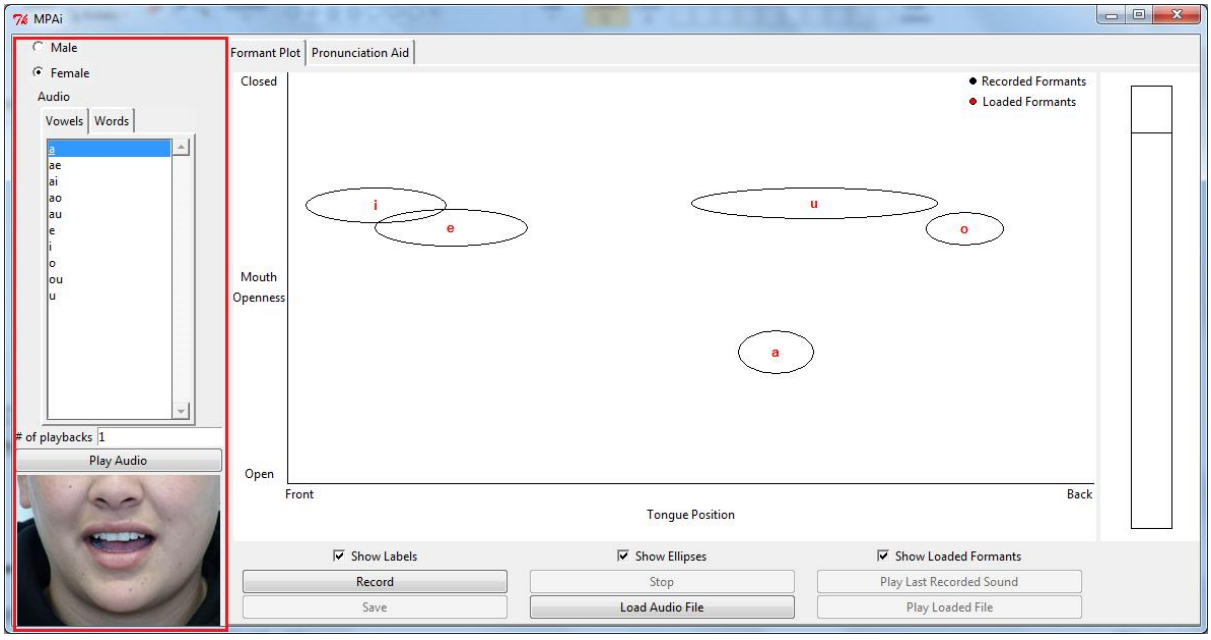

Figure 1: The Formant Plot.

3.0 Application Overview

MPAi is divided into two main sections in terms of functionality; the Formant Plot and Pronunciation Aid. MPAi also has an area reserved for functions that are common for both

the Formant Plot and Pronunciation Aid. This section allows users the ability to choose a particular vowel or word.

3.1 Common Functionality

The area reserved for common functionality, represented by the enclosed area that has a red outline in figure 1, allows the user to perform various actions related to the playback of audio files and for controlling various parameters needed for the Formant Plot to produce accurate results. In more detail, these functions are:

- Gender selection: This determines which set of audio files will be played when the user tries to playback a word. Selecting "Male" would result in giving the user the ability to playback words spoken by male Mäori speakers and the opposite is true if "Female" is selected. Providing a different set of sounds for each gender was done because it was determined that it would be easier for the user to learn particular words when listening to someone of the same gender. However, for Mäori vowels, only male speakers are included. The gender also controls the values for parameters that are used in the Formant Plot.
- Playback: A list of audio files, which are comprised of Mäori vowels and words, are displayed which the user can select to perform playback. The selected word will also be used as the word that is being checked for analysis in the Pronunciation Aid. On selection of a vowel, an image (for monophthongs) or a short video (for diphthongs) will be displayed showing the user the mouth movements of an expert speaker saying the selected vowel.

3.2 Formant Plot

3.2.1 Outline

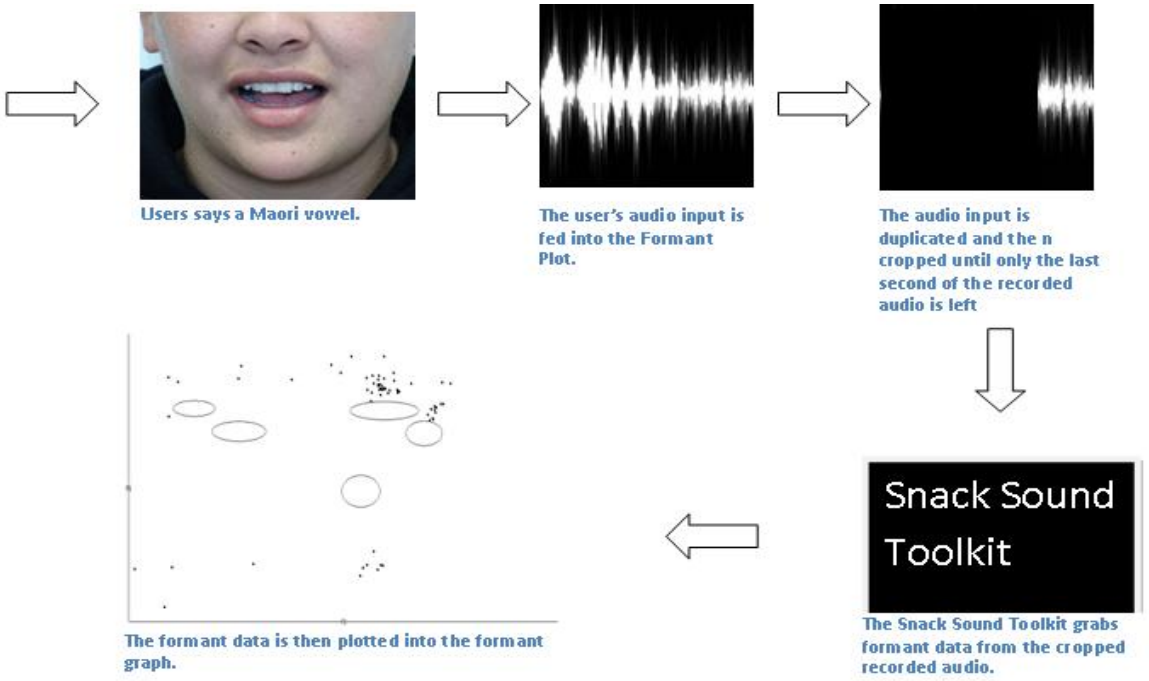

Figure 2: The Formant Plot process.

The Formant Plot, seen in figure 1, provides a suitable user interface for which users can practice speaking Mäori vowels. The Formant Plot works by first taking in the audio input of the user. Next, the application duplicates the sound and this duplicated sound is then cropped until only the information from the last second of the recorded input is left. Then extraction of the last formant of the recorded input is performed. Finally, it plots the extracted formant onto the plot. Figure 2 shows the step-by-step procedure that MPAi goes through during use of the Formant Plot. This process is repeated periodically throughout the duration of the recording of the user. The user checks whether or not their pronunciation of the vowel is correct or not by comparing their formant data plot with the gold standard ellipses; dots inside the correct ellipse indicates that the user's pronunciation of the vowel is adequate. Figure 3 shows what the users sees during use of the formant plot.

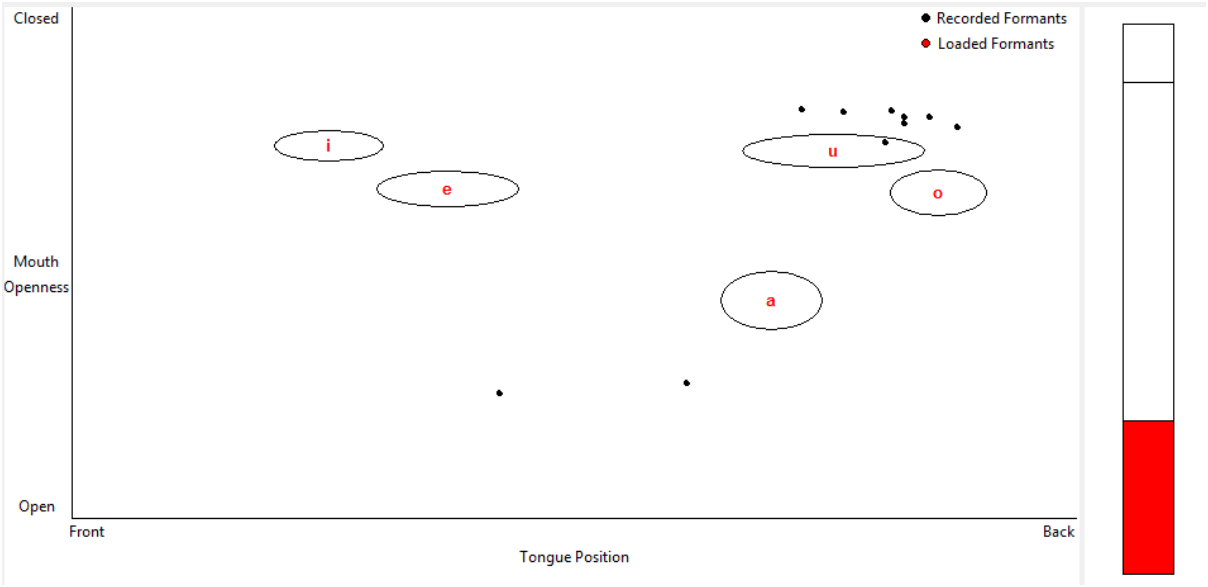

Figure 3: Example showing the Formant Plot in use.

3.2.2 Other Functionality

Apart from the base functionality of providing the user a mechanism for which they can practice learning the correct pronunciation of Mäori vowels, the Formant Plot also offers various other actions that the user can perform. These are:

- Playback of last recorded sound: The user has the ability to listen to the last known recording.
- Save: The user is able to save the current recording to a file, which can be used for future use.
- Load: The user has the ability to load in external audio files, which will also plot all the formants of that particular file. The user also has the ability to playback the loaded audio file.

3.2.3 User Interface

The Formant Plot is targeted at users who aren't experienced in the field of speech processing. This requires the Formant Plot to provide an interface that displays enough information so that the user can understand what is going on, while still maintaining a clear and clean look so that the application is easy to use. There is a large space allocated for the plot so that the formant data and gold standard ellipses are easily readable and uncluttered. There is also an option for the user to hide the gold standard ellipses and formants that were retrieved from a loaded audio file which helps to make the Formant Plot even clearer if the user chooses to. The axes of the plot provide adequate information for the user to be able to understand the results of the formant plotting and adjust the way they are pronouncing a particular vowel accordingly.

3.2.4 Problems

There were a few critical problems encountered during the development of the Formant Plot. These problems were audio input quality and limited resources available to process the formant data.

Different users will have different devices which are used to input the audio data from the user into MPAi. This would result in varying degrees of quality of the audio data input. This was problematic for the Formant Plot because the user either speaking too loudly or too quietly would result in incorrect formant data being plotted. The solution provided for this was to introduce a volume meter, shown in figure 4, so that the user can self-monitor their own level of volume. However, this feature is incomplete because the exact level of volume at which the audio data provides inaccurate formants has not been determined yet, so the meter is currently using a rough estimate of the desired volume level at which the Formant Plot will work correctly.

Initially, all of the formants were extracted from the whole recorded sound. This resulted in MPAi slowing down noticeably as the length of the recording increased. This is due to the application having to process through all of the data of the user's audio input to retrieve all of the formants, which required a considerable amount of time to perform. Cropping the audio data before the formant extraction step was introduced to fix this problem because the time taken to crop the sound file took considerably less time than formant extraction of the whole audio data.

3.3 Pronunciation Aid

3.3.1 Concept

The Pronunciation Aid provides the user the ability to be able to analyse their own pronunciation of a particular Mäori word. During the analysis process, the application performs forced alignment, using MFAP, which allows for comparison between the user's utterance of the chosen Mäori word with a set of speech data, which is taken from an expert speaker's own version of the word, as shown in figure 4.

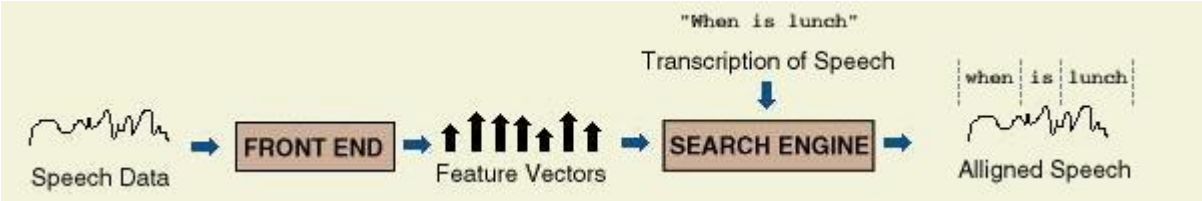

Figure 4: Forced alignment process (Reproduced from [4]).

After the analysis is done, the user is presented with a correctness score, as can be seen in figure 5, which shows how correct the user's pronunciation of the word is as a percentage.

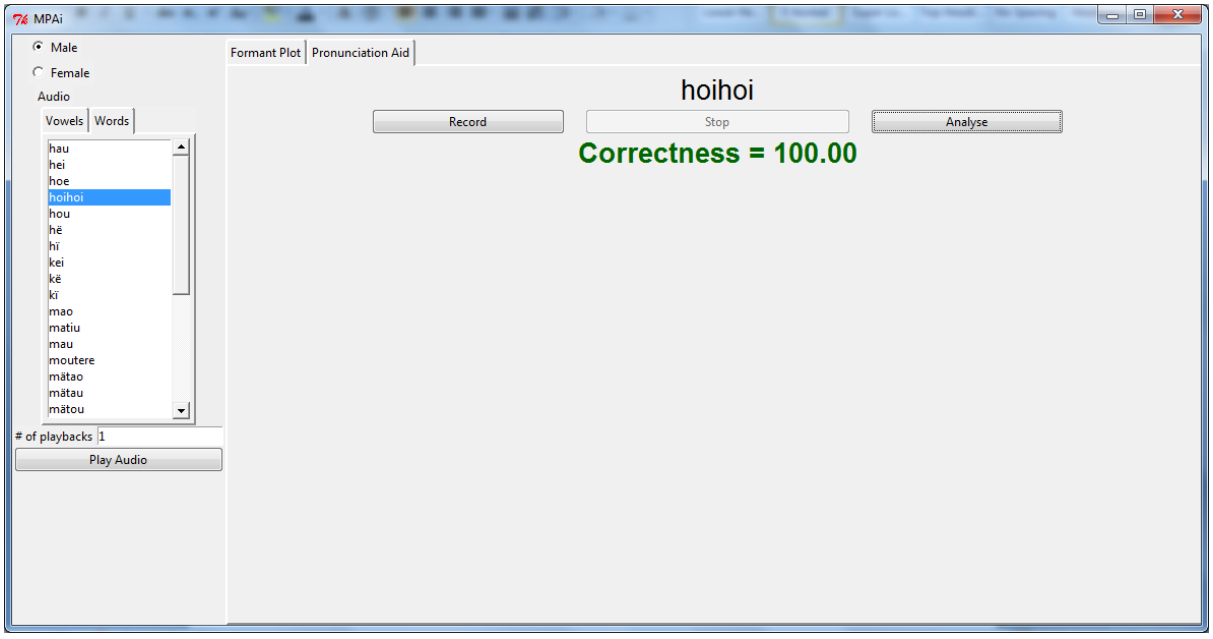

Figure 5: Pronunciation Aid output.

3.3.2 Problems

The Pronunciation Aid, in its current state, is only a proof of concept and is incomplete. This is due to problems encountered due to the use of the HTK part of MFAP. Regardless of the user's pronunciation of the chosen word, the Pronunciation Aid always reports back a perfect correctness score. As previously mentioned, the functionality required for the Pronunciation Aid is to first recognise the speech data provided by the user and then this data is matched with the speech data from an expert speakers utterance of the chosen word. The HTK documentation provides a tutorial for implementing speech recognition into an application, which satisfies the first requirement of the functionality that the aid requires, but there is not much documentation provided for the second part and so there must be more research performed towards the HTK in order to get the Pronunciation Aid fully working.

4 Development

This section describes the design of the software and the tools used during the development of MPAi.

4.1 Software Design

It was predicted that the application would undergo a series of changes due to changing requirements and feedback received from users, therefore, the MPAi application is designed for modifiability. In order to allow MPAi to be easily modified, a modular design was implemented in the code. MPAi was separated into six sections of code (classes) – Runner, MainApp, FormantPlot, LoudnessMeter, PronunciationAid and SoundProcessing – which allowed each component of the application to be uniquely modified without requiring much changes to the other sections. Figure 6 describes how the different classes interact with each other. The responsibilities of the different MPAi classes are:

- *Runner* – This class' only responsibility is to start up the application. It uses the MainApp class to run the program.
- *MainApp* – The core section of the application. This class contains all the functions that are common for both the Formant Plot and Pronunciation Aid (the list of functions provided are mentioned in section 3.1). The graphical user interface (GUI) of the application can be changed here, but the section of the interface that is unique to the Formant Plot and Pronunciation Aid should be changed in their corresponding classes (FormantPlot and PronunciationAid). MainApp uses the FormantPlot and PronunciationAid classes so that the application is able to provide the functionality that these two classes contain.
- *FormantPlot* – This class contains the code which supplies the GUI and functionality related to the Formant Plot. This class uses the LoudnessMeter and SoundProcessing classes. The LoudnessMeter is a complementary part of the Formant Plot, while the SoundProcessing class is used to perform various audio data manipulations that the FormantPlot requires in order to provide the functionality it is supposed to provide.
- *LoudnessMeter* – This class contains the GUI of the Loudness Meter and the calculations needed by the meter to supply the volume level of the audio input to the user.
- *PronunciationAid* – This class contains the code which supplies the GUI and functionality related to the Pronunciation Aid. PronunciationAid uses the SoundProcessing class to perform audio data manipulations which are needed by the Pronunciation Aid.
- *SoundProcessing* – This section contains various algorithms which are used to perform sound cropping, formant extraction and retrieval of a probability which determines whether a particular sound is spoken or just background noise.

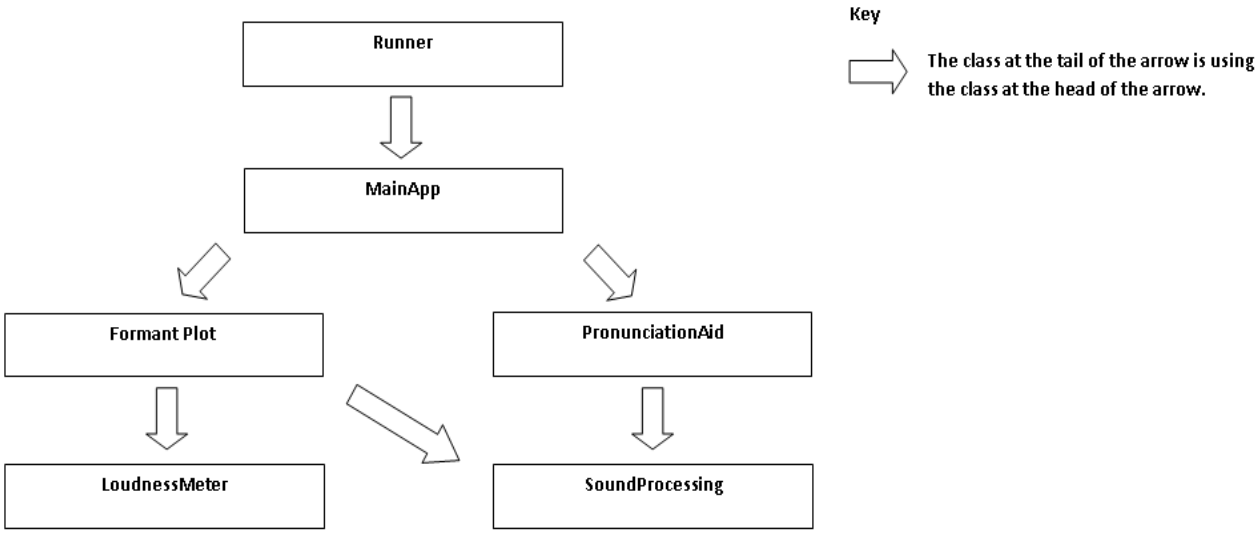

Figure 4: Diagram of how the classes interact with each other.

4.2 Tools

There were various third-party programs that were used in the development of MPAi:

- Snack Sound Toolkit – A toolkit that can be used with Python to create audio applications [5]. It provides support for basic sound handling (e.g. playback of a sound) and graphs which can be used to plot data extracted from a sound.
- HTK – This toolkit is a "toolkit for building and manipulating hidden Markov models" which can be used for speech recognition. MFAP uses the HTK in its implementation.
- Py2exe – An executable program which converts Python code into executable Windows programs. The converted program is able to run without requiring a Python installation [6].
- Python Imaging Library – A library which allows for more image manipulation features than the default functions that Python provides.
- Wavesurfer – An application which can segment audio files.

5 Feedback

One of the aims of this project was to create an application that was easy to use for its target audience (non-expert users). An informal usability study, in the form of asking the MAONZE [7] group and a student with a degree in linguistics a set of questions related to the Formant Plot section of the application, was carried out to obtain feedback on the usability of the system. Here is a summary of what they thought of the application:

- *User Interface* – The layout of the interface was simple and intuitive, although the aesthetics of it was a bit dull and could be improved. There were suggestions made which would help the user interface easier to understand for the user (such as having an indicator for when the application is currently recording user input).
- *Functionality* – The functionality that the users expected from the Formant Plot was adequate. The users reported that speed at which the formant data was being plotted into the Formant Plot was quick and responsive, but the accuracy of the plot could be

improved. There were issues reported which were related to audio playback sounding choppy.

- *Miscellaneous* – Some users reported some difficulty in getting the application running since it required them to go inside a subfolder of the program in order to find the executable which runs the application. There were also comments made in regards to having the option to run the application in a different operating system (e.g. Mac OSX). MPAi currently only runs in Windows, but quite a few users from the MAONZE group use Macintosh systems.

6 Future Work

The current version of MPAi is still incomplete. There are areas of the application which can be improved, but due to the time constraints of the project, it has been left unresolved. These areas include:

- *Implement required functionality for the Pronunciation Aid* – The Pronunciation Aid in its current state does not provide any useful functionality to the user in terms of giving accurate feedback to the user about their pronunciation of a Mäori word. Further work must be done either on the HTK or on other suitable programs.
- *Availability on other operating systems* – Different users will have different machines (computers) in which they use MPAi on, however, due to low priority and time constraints, MPAi can only be run in Windows.
- *Improving the Loudness Meter* – The Loudness Meter, in its current state, does not provide an exact feedback to the user of how loud their audio input is. Further work must be performed in regards to finding a suitable algorithm which can provide accurate calculations of the volume level of the audio input.
- *Improving the Formant Plot* – The Formant Plot could be improved upon by implementing the features mentioned in the feedback that was received in the usability study.

These are the core areas in which future work on this project should focus on. However, more features should also be implemented which will enhance the user experience, such as having a mechanism for tracking user progress.

7 Conclusions

An application (MPAi) has been developed which allows users to practice their pronunciation of the Mäori language. MPAi incorporates functionality found from previous work (Formant Aid) and improves upon this work by increasing the usability of the system through the implementation of a new user interface in which users who aren't experts in the speech processing field can easily understand feedback given based on their pronunciation of Mäori. However, the Pronunciation Aid functionality is incomplete and more work must be done towards this part of the application as it will be a valuable tool for users in which they can further their progress in the pronunciation of the Mäori language. MPAi has been developed with ease of modifiability in mind so that it will be easy for the developer improve upon the application in the future.

An informal usability study has shown that MPAi has achieved the project's aim of providing a usable application for non-expert users. However, more testing and implementing formal usability studies is recommended in order to more accurately determine how usable the application is. Overall, the feedback received has been positive and one user has already asked if they could use the application for one of their Mäori classes that is being taught at University level.

8 References

1. *New Zealand Language, Language of New Zealand*. New Zealand Tourism Guide. Retrieved March 3, 2011 from http://www.tourism.net.nz/new-zealand/about-new-zealand/language.html

2. *Learn Spanish – Learn French – Language Learning – Rosetta Stone*. Rosetta Stone. Retrieved March 3, 2011 from http://www.rosettastone.com/

3. *HTK Speech Recognition Toolkit*. Speech Vision and Robotics Group of the Cambridge University Engineering Department. Retrieved March 3, 2011, from http://htk.eng.cam.ac.uk/

4. *Automatic Speech Recognition*. Institute for Signal and Information Processing. Retrieved March 10, 2011 from http://www.isip.piconepress.com/projects/speech/software/tutorials/production/fundamentals/v1.0/section_04/s04_04_p01.html

5. Sjölander, K. *TMH KTH :: Snack Home Page*. Retrieved March 11, 2011 from http://www.speech.kth.se/snack/

6. *FrontPage – py2exe.org*. Retrieved March 14, 2011 from http://www.py2exe.org/

7. *MAONZE Project: Sound Change in Maori*. MAONZE. Retrieved March 14, 2011 from http://www.ece.auckland.ac.nz/~cwat057/MAONZE/index.html