Chapter 6

Graphical User Interface

In the previous chapters, we gave details of various models to implement ASR in Gujarati language for isolated words and sentences. In the chapter 3, we discussed the stepping stones of creating a speech recognition model for the Gujarati language for isolated words. Chapter 4 further improvised and generalised the models developed in chapter 3 for sentences. The best models are the ones, combined using the elements of ensemble learning in the chapter 5 for sentences spoken in the Gujarati language.

Finally, to put the best models to use for applications, we furthered the design of a graphical user interface. The objective of such a graphical user interface for Gujarati speech recognition is to demonstrate the use of technology for a novice or technically ignorant person. To achieve this objective, we created a simple graphical user interface. It is very user-friendly, so a layman can perform speech recognition with a few clicks and a few buttons. We proceeded to make a couple of different designs and layouts for the graphical user interface. We made these interface using the tkinter package in Python.

In our first attempt to create an interface based on the hidden Markov model, the one discussed in the section 4.4.3, that is used for continuous speech recognition. This is a detailed interface that is able to record a speech and give the text output of the spoken Gujarati speech.

As shown in the Figure 6.1, this interface consists of a total of nine buttons. We can consider them in two clusters: buttons 1-6 and the last 3. The first six buttons are the ones with the help of which the model is constructed and trained for use. With the help of the last three buttons, we use the trained model for a new sentence made from the trained vocabulary. The details of each button are as follows.

The first button is used to read the vocabulary database from the local machine. The function of the second button is to extract the words out of all the recordings using the

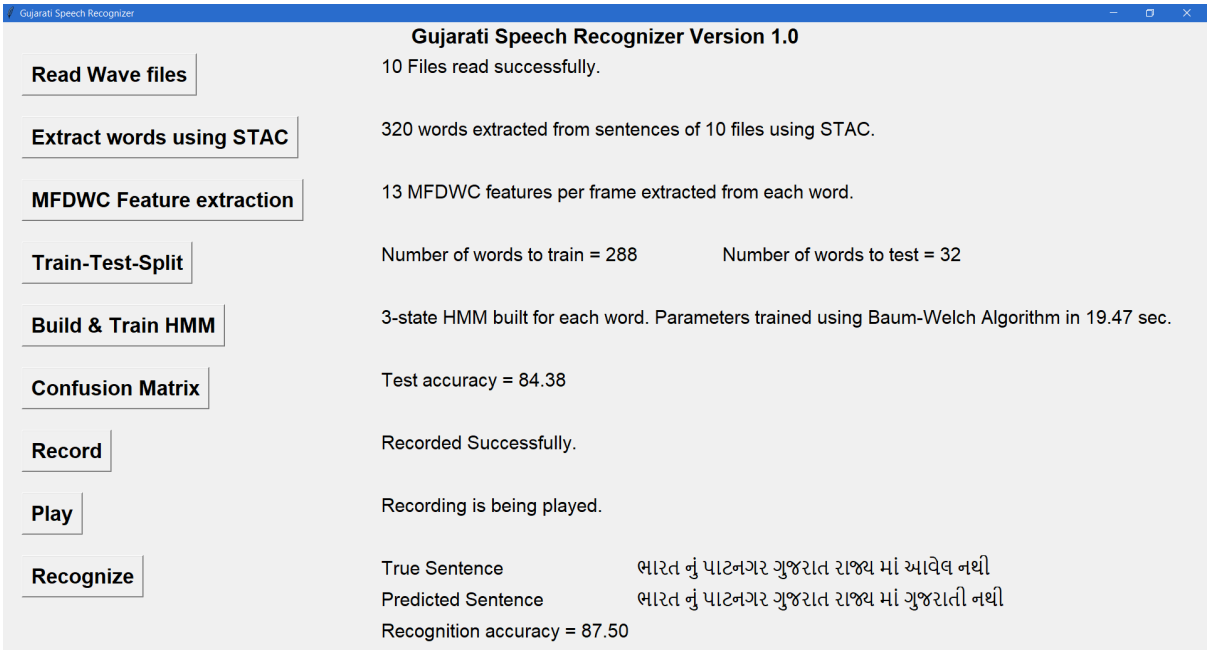

Figure 6.1: Speech recogniser interface

short-term autocorrelation. The third button, is kept to extract the feature vectors from the speech using the mel-frequency discrete wavelet coefficient method. The use of the fourth button is to divide the data into training and testing. Then, we build and train the 3-state hidden Markov model using the fifth button. The sixth button is useful to view the training performance of the model. This way, it is a kind of static interface in the construction of the model phase.

The use phase of the interface is dynamic, as described below. After all these procedures, we can now test our model with a test speech, and this test speech can be given by a user by recording his or her own voice after clicking the seventh button. This button also saves the recording. The eighth button gives confirmation to the user about the speech he or she has recorded by playing it. Finally, the last button gives the user the recognised text from the speech spoken by him or her. All the buttons print information about the updates, as shown in the Figure 6.1. This was our first graphical user interface for a Gujarati speech recogniser. In the future, we intend to allow the user to generate the model from his input wav files instead of from the local machine in the first step of the construction phase. This will give dynamism to the implementation.

In the second graphical user interface, we designed a very simple user interface with fewer buttons, as shown in the Figure 6.2. In this case, the classification is based on multilayered perceptrons. There are four buttons in the interface. The first button displays the vocabulary of the model. Clicking it shows the list of words in the vocabulary. A speaker has to select words from it and construct sentences from them. Then, the sentences can

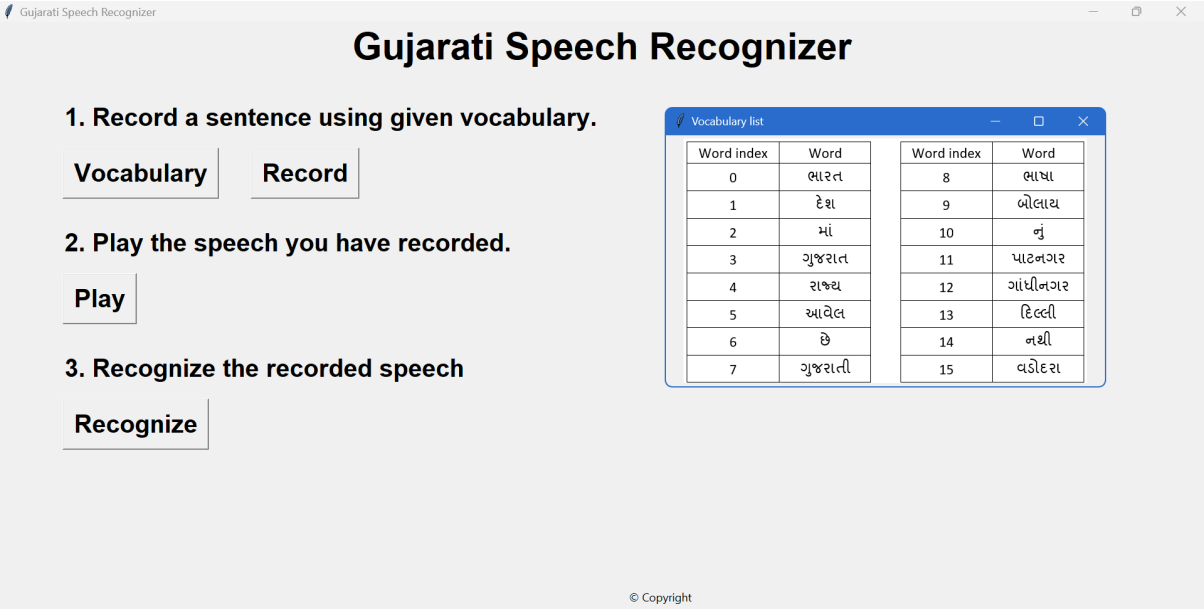

Figure 6.2: Updated speech recogniser interface

be recorded using the second button, "record." This button will also store recordings of the speech in the *.wav file format. The third button, "play," is used to play what a speaker has recorded. This button will play the recording. Lastly, there is a button to recognise a speech. This button prints the recognised speech as a text using the trained multilayered perceptron on the small vocabulary.

Conclusion

In this chapter, we discussed our efforts to create a simple graphical user interface. Two different approaches were used in this. In the first approach, we designed a user interface with nine buttons, using the hidden Markov model as the base model. This was a more detailed interface because it has buttons for all the sub-steps of speech recognition. In the second approach, we designed a user interface with fewer buttons, using the multilayered perceptron as the base model. The second approach is more convenient for laymen as it does not require technical specifications regarding the model parameters. In the next chapter, we will conclude the entire work.