INDIAN INSTITUTE OF TECHNOLOGY, GUWAHATI

Department of Computer Science and Engineering

Project Report on

SMALL TUTORIAL ON VEHICLES

Based on Speech Recognition System

Submitted to:

Prof. (Dr.)
P. K. Das

Submitted by:

Anuj Singh Thakur (224101008)

Nihal M. Singh (224101038)

Abhijit Deuri (224101001)

For course fulfilment of CS566: Speech Processing

Acknowledgement

First and foremost, we would like to express my gratitude to Professor P.K. Das for giving us the opportunity to work on this project. We would also like to thank the Teaching Assistants for providing us with the necessary resources and support throughout the duration of the project. We would also like to thank our classmates for their help and suggestions.

1. Anuj Singh Thakur (224101008)
2. Nihal M. Singh (224101038)
3. Abhijit Deuri (224101001)

CONTENTS

LIST OF FIGURES

1 ABSTRACT

This project focuses on developing a speech recognition system that can transcribe speech in real-time. Our system, a small tutorial on vehicles, provides a speech-based solution to display the image of the particular vehicle which is spoken. Our system basically helps the user to get to see check the results for certain vehicle and it also provides the provision to add a new vehicle name by using the live training option. The system is trained on a dataset of vehicle name recordings and able to handle different accents and dialects. It uses the concepts of Hidden Markov Model (HMM) for creating the models for different vehicle names and characterizing them when tested for. The system is developed in C/C++ in Visual Studio IDE. For the time being the system is limited to detect only a few numbers of vehicle names, however further extension can be easily made to support a greater number of vehicle names.

2 INTRODUCTION

2.1 What is Speech Recognition?

Speech recognition, also known as automatic speech recognition (ASR), computer speech recognition, or speech-to-text, is a capability which enables a program to process human speech into a written format. While it's commonly confused with voice recognition, speech recognition focuses on the translation of speech from a verbal format to a text.

When we speak into a micro- phone which is connected to the computer/mobile, it converts it to a text file which contains some amplitude values.

2.2 Our Project

Our project uses the concepts of Hidden Markov Model to recognize the spoken vehicle name and if the name matches with any vehicle name present in our pre-defined words, it shows the image of the particular spoken vehicle. Currently the pre-defined words of our system are – Bike, Bus, Truck, Jeep, Train, Car, Auto, EV, Boat. If the recognized word is not matched with any of the words present in our vocabulary, then an image pops up to indicate the absence of the word in the vocabulary. Our system also provides the provision to add a new vehicle name and its corresponding image by using the live training option by uttering and recording 20 utterances of the new vehicle name to be added.

2.3 Future improvements

The time for live training the new words can be reduced and further improvements can be made using different High-Level languages. There is always a scope of integrating the model with different applications of

speech recognition systems. For the time being only the image of the particular vehicle name is displayed, however further improvements can be done to depict the specifications or configurations to directly compare multiple vehicles based upon the need.

3 EXPERIMENTAL SETUP

Basic requirements for this project are as follows-

- Windows OS
- Microsoft Visual Studio 2010
- C++11 integrated with VS2010
- Recording Module

4 PROPOSED TECHNIQUES

4.1 Flowcharts

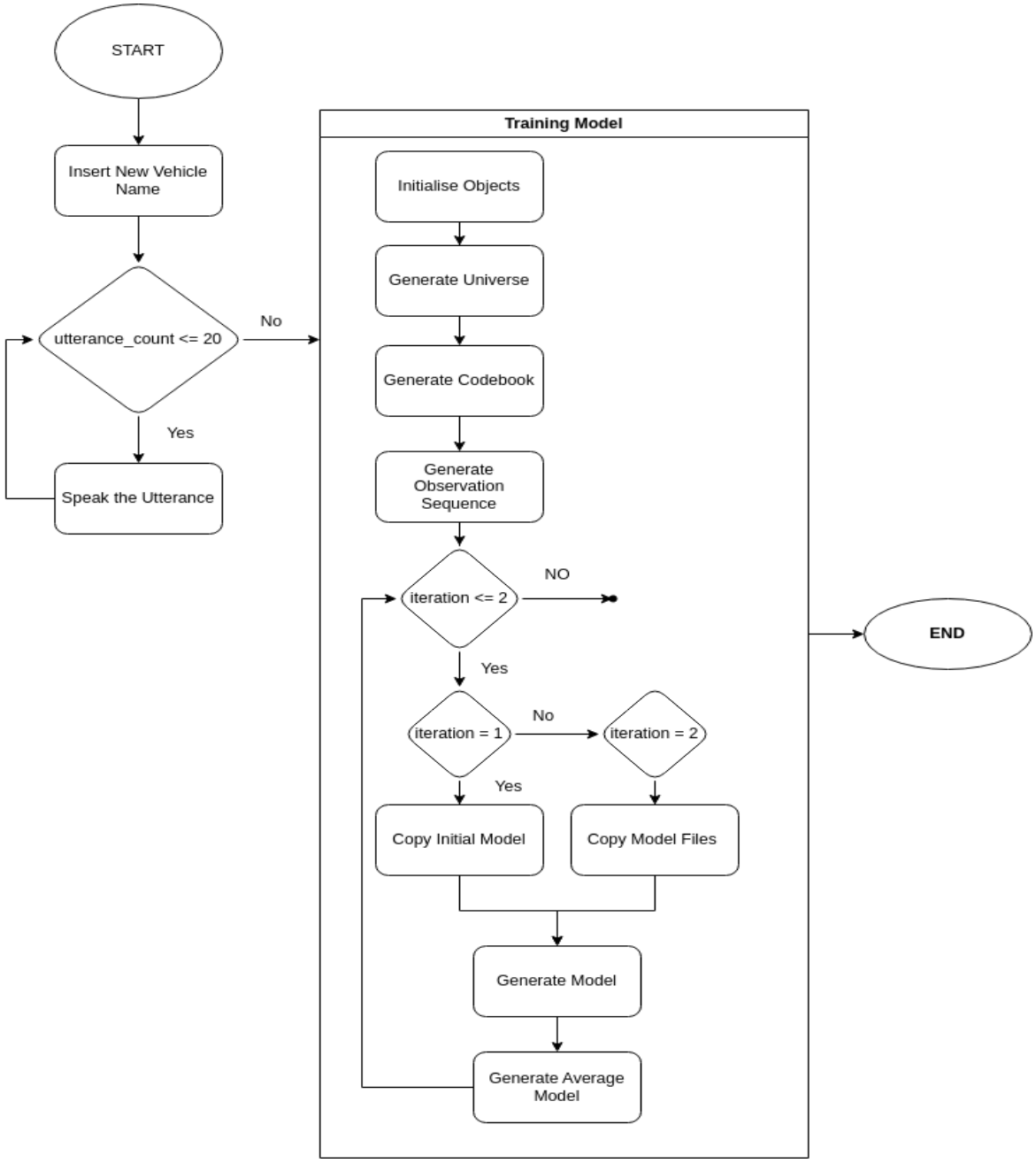

Figure 1 – Flow chart of the training process

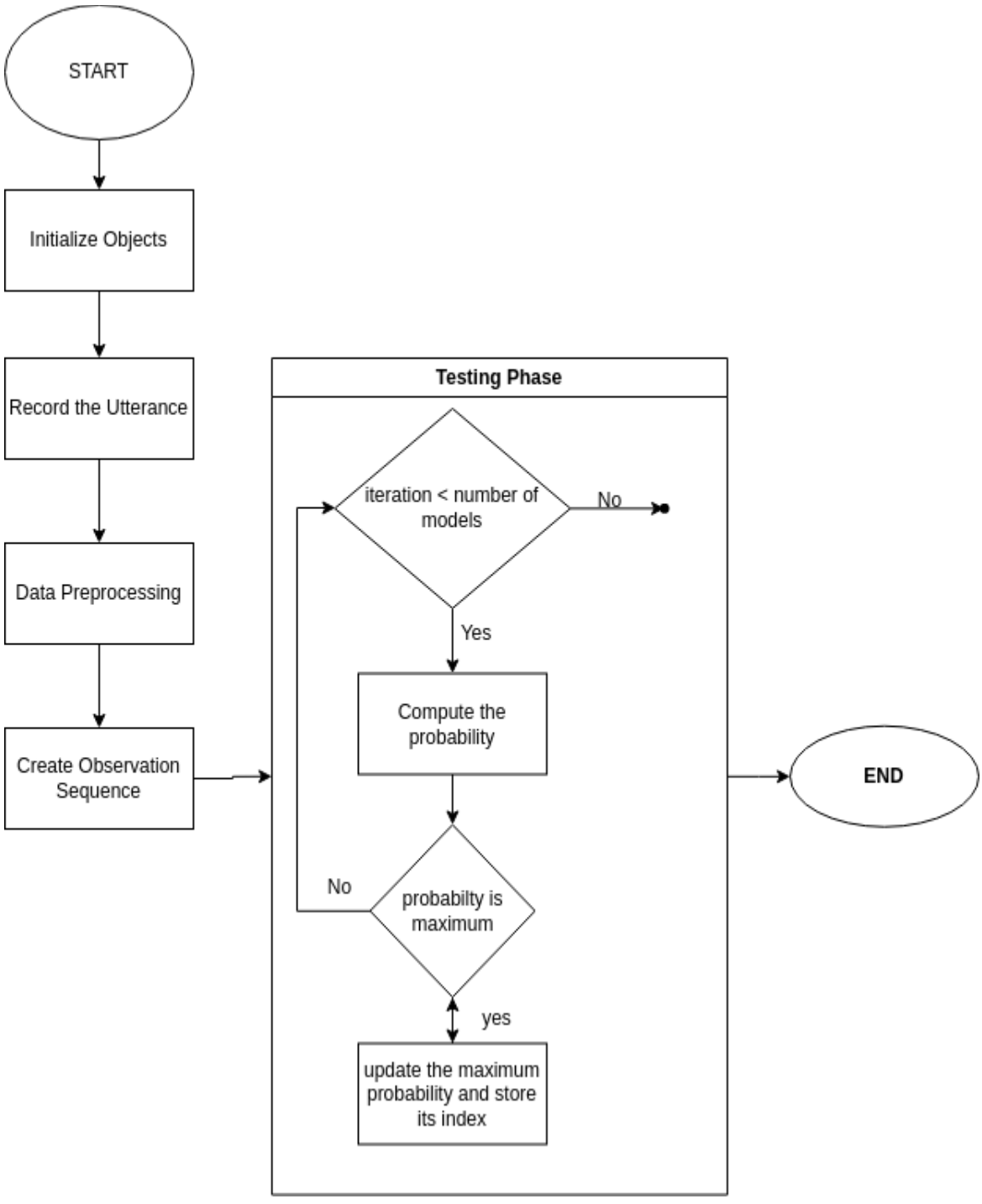

Figure 2 – Flow chart of the testing process.

Figure 1 and Figure 2 depicts the overall flow of the project.

4.2 Model description

For the task of recognizing a particular spoken vehicle name and displaying the image of that particular vehicle, the concepts of Hidden Markov Model is used. Firstly, for training the model the vehicle dataset recordings are created and for each vehicle a model is created. The universe of the entire recording is created with each vector size of 12 (i.e., 12 Cepstral coefficients) and a codebook of 32 regions is created from the universe. This codebook helps in the generation of the observation sequences. Upon getting the observation sequences of the training data, for each utterance an HMM is created and the average of all the models of each utterance is averaged up and used as the reference model for that particular vehicle.

Similarly for testing a particular utterance, firstly the observation sequence is created using the codebook generated earlier. Now the observation sequence is compared with all the reference models and the model with the maximum P(Observation / lamda) (Probability the given observation sequence given the model) is given as the prediction for that particular testing utterance.

The model is integrated with a UI which provides the provision to train a particular vehicle name in live, test a particular vehicle name and display the corresponding image.

5 Result

5.1 Home Page

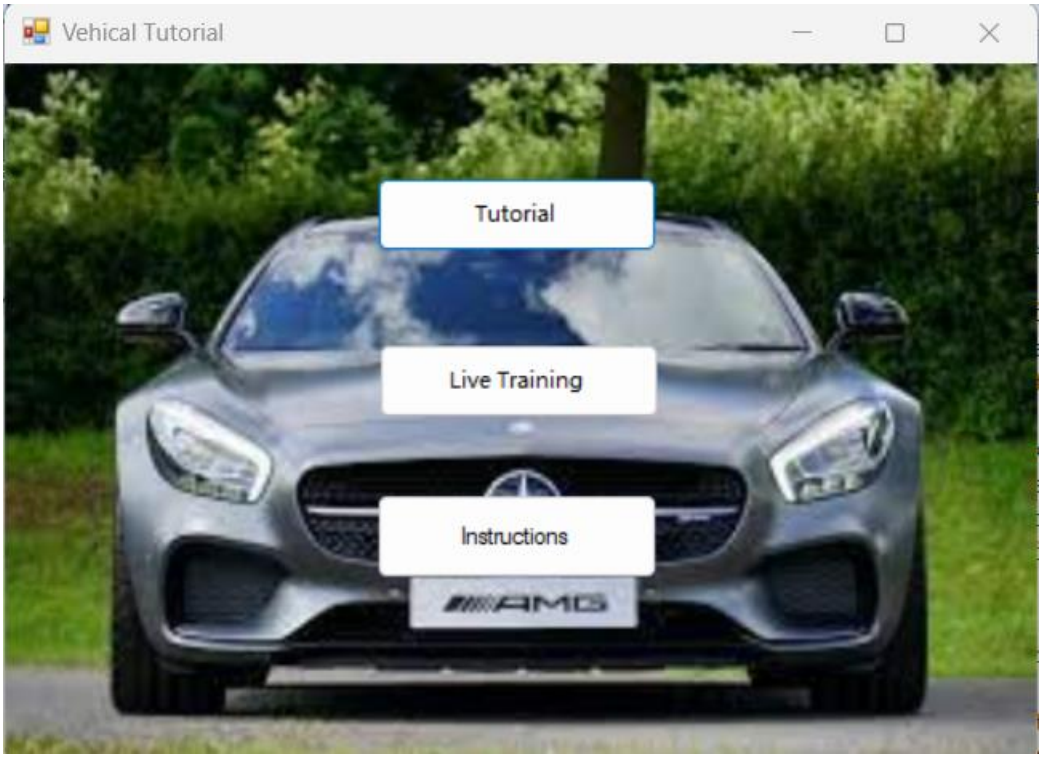

Figure 3 – Home page.

Figure 3 – Shows the Home page of the project. The Tutorial button is for live testing, The Live Training is for live training new vehicle names and the Instruction button displays the already present vehicle names in the vocabulary.

5.2 Live Testing

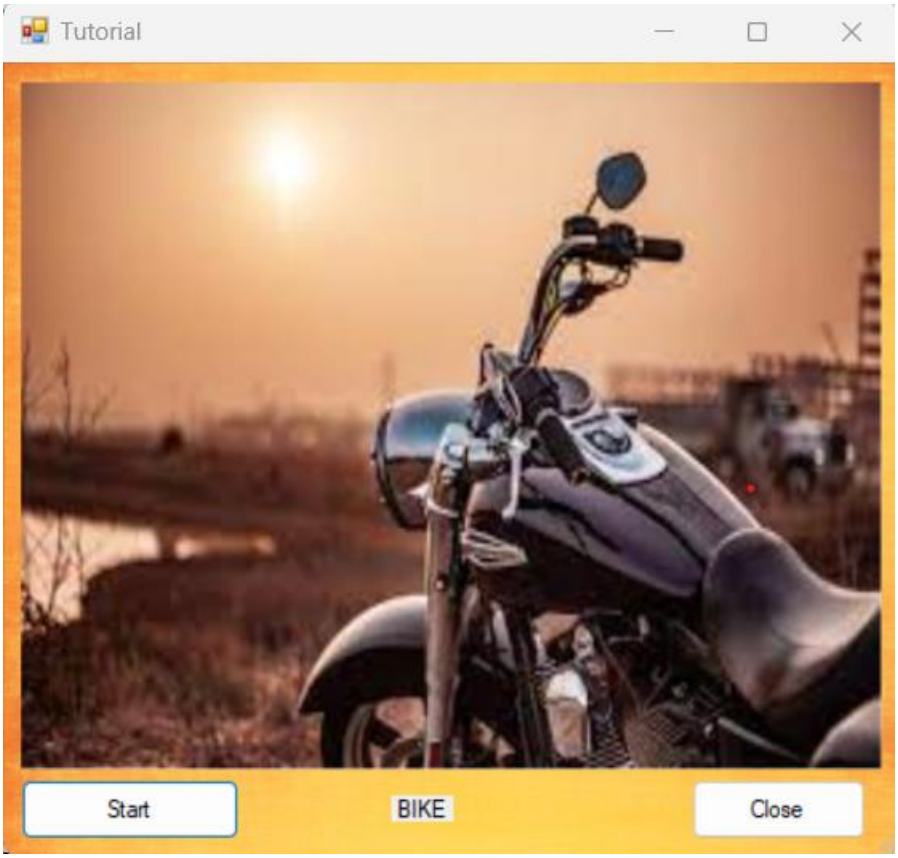

Figure 4 – Live testing or tutorial.

For live testing, to get the image of a particular vehicle a button is deployed. Upon clicking on the button and uttering a particular vehicle name am image will pop up, based on the availability of the vehicle name in our vocabulary.

5.3 Live Training

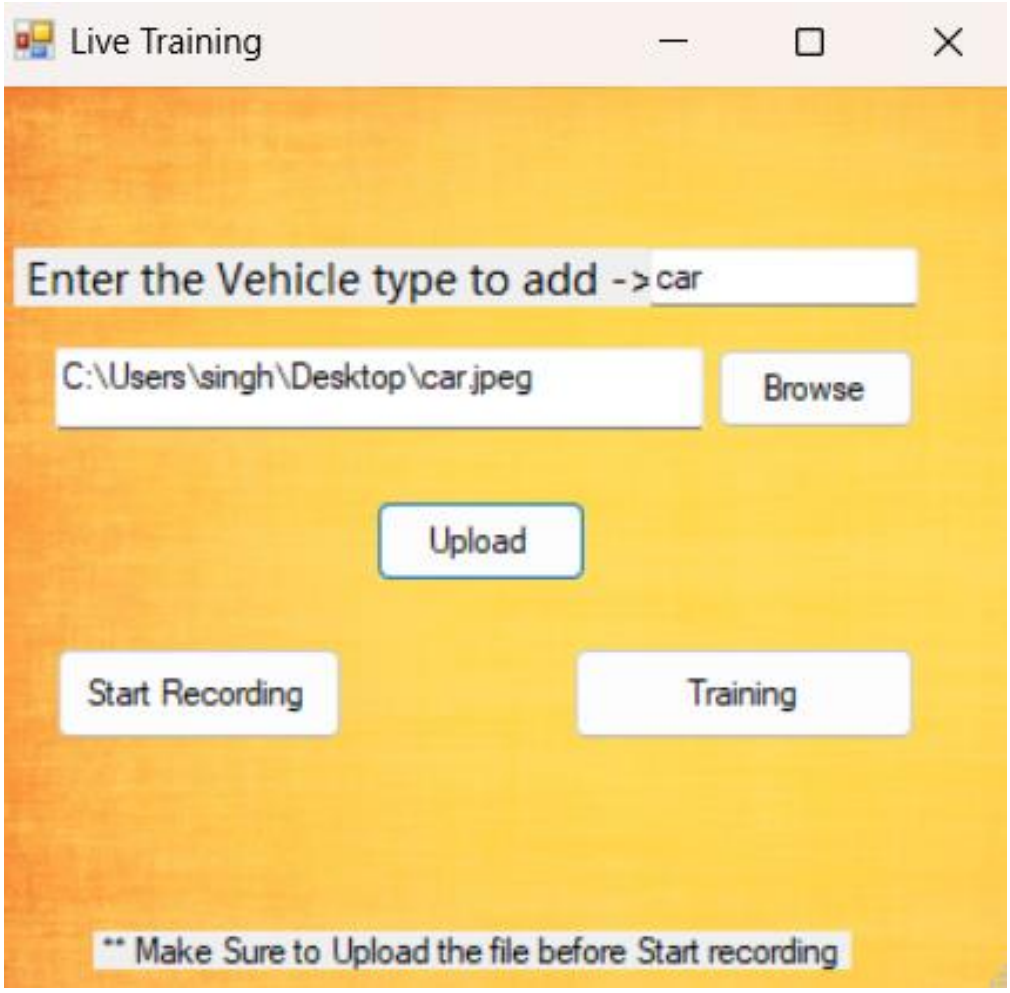

Figure 5 – Live training

For live training, another button is deployed. Upon clicking on the button, the user is required to utter 20 utterances of a new word and also browse and upload an image corresponding to the new vehicle name.

5.4 Instructions

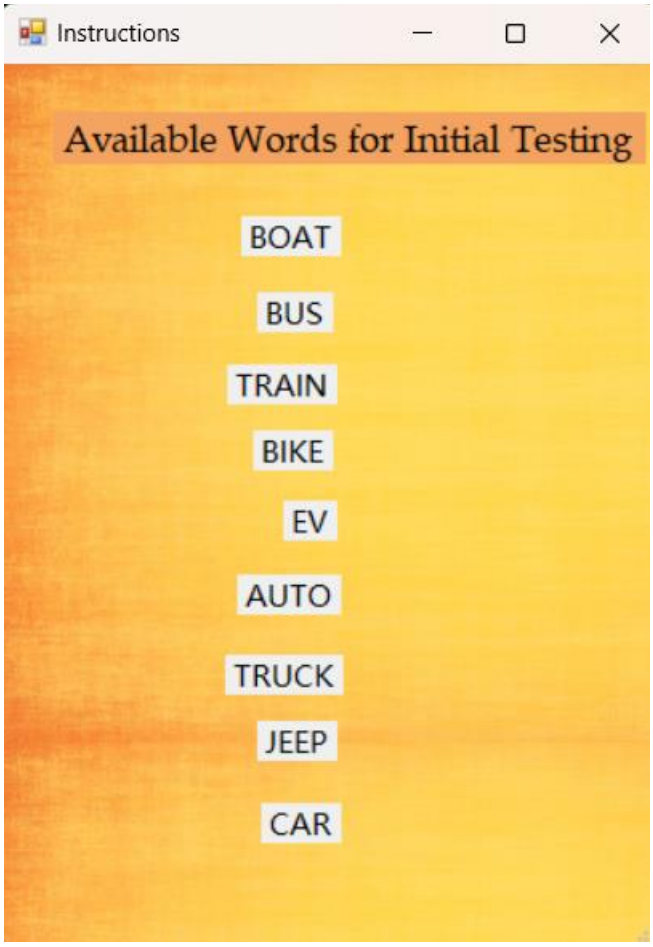

Figure 6 - Instructions

The instructions button helps the user to know about the vocabulary.