ARTIFICIAL INTELLIGENCE BASED COMMUNICATION SKILL DEVELOPMENT ASSISTANT FOR SCHOOL CHILDREN

Mr. Navis M [1], Ms. Nisha Ebarila J [2], Mr. Patrick Samuel [3], Ms. J. Limsa Joshi [4]

Loyola-ICAM College of Engineering and Technology (LICET), Chennai, India.

Abstract - a system has been proposed whereby the voice inputs will be received by a UI, processed by an ANN back end, and the voice outputs will be delivered back to the user through the UI. In the initial stages of development, the voice outputs would only be produced correctly if the corresponding inputs had already been previously entered before. In order to increase the efficiency of the system, AI was implemented into it so that no matter the input, an appropriate output would be produced accordingly. To make this idea feasible, the concept of Artificial Neural Networks was introduced. Now, even if prior knowledge of the input is not available, the system can still produce the appropriate output based on its analysis of similar inputs. This is similar to the functioning of the human brain. The aid of the open-source software library TensorFlow has also been made use of to help construct the multi-layered neural networks that will us train the system and produce the required outputs.

Keywords -- Communication Skill Development, AI, ANN, neural networks, TensorFlow.

INTRODUCTION

Artificial Intelligence is a key component of today's ever changing technological world. A subject of increasing importance, AI is now used in various fields, ranging from machine learning to speech processing. This was one of the reasons that AI was selected as the domain for this project as it encompasses all of the concepts that are required for the project's working. Artifical Intelligence deals with the intelligence displayed by machines, which is a change from the discussions usually regarding the intelligence displayed by humans. AI has been spoken of since times of old; the Egyptians, Chinese and Greeks all had opinions on the possibilites of machines taking over one day. Their once ancient beliefs have now become reality as we are now surrounded by highly intelligent machines that help us perform so many of our tasks. AI is used to solve a plethora of problems including reasoning, knowledge representation, planning, learning, natural language processing and perception. Object manipulation is one of the most important areas of application of AI.

Learning, reasoning, problem-solving, and perception, are some of the most important fields that are present under AI. These all work together to enable us to produce the results that we need accurately. Generalization under learning is difficult to implement but one of the key ingredients to ensuring that the system is able to adapt itself to any situation. Reasoning can be divided into either deductive or inductive reasoning. Problem-solving can be divided into general purpose and special-purpose. These divisions highlight the capabilities of AI to be able to adapt itself to whatever it may face.

ANN(Artificial Neural Networks)-- can be described as software implementations of the human brain. There are two main types of ANNs: the supervised ANN and the unsupervised ANN. In supervised ANN, there is an attempt to produce a specified output for a specified input. In unsupervised ANN, the network tries to interpret the input and respond accordingly on its own. ANN consists of a large number of processing elements that have a large number of weighted connections between them. The functioning of AI can be inferred from Figure 1. ANN has three main elements: 1. Processing Elements 2. Topology 3. Learning Algorithm. A summation unit will take all of the inputs, calculate the weight of each one of them and then calculate the weighted sum of the values. The sign of each input will help us determine whether it is positive or negative. The total weighted sum is

known as the activation value and the output is then produced from this value. There are a large variety of topologies that can be used in ANN: instar, outstar, group of instars, group of outstars, bidirectional associative memory and autoassociative memory. The operations of neural networks can be governed by learning algorithms such as Hebb's law, Perception Learning law, Delta Learning law, Wildrow and Hoff LMS Learning law, Correlation Learning law, Instar Learning law, and Outstar Learning law. There are many reasons why ANNs are so widely used: 1. Parallel processing structure 2. The failure of one neural element will not affect the rest of the process 3. ANNs can learn without reprogramming by implementing the appropriate learning algorithms.

Tensorflow—There are many features of TensorFlow that were the reason for its use in the software: 1. Responsive Construct 2. Flexibility 3. Easy Trainability 4. Parallel Neural Network Training 5. Open-Source 6. Large Community 7. Feature Columns 8. Availability of Statistical Distributions 9. Layered Components 10. Visualizer 11. Event Logger. TensorFlow is renowned for its excellent computational graph visualizations, extensive low-level library that is able to adapt itself to whatever the user requires, and its fine debugging, scalability and, pipelining abilities. Due to TensorFlow's great voice and sound

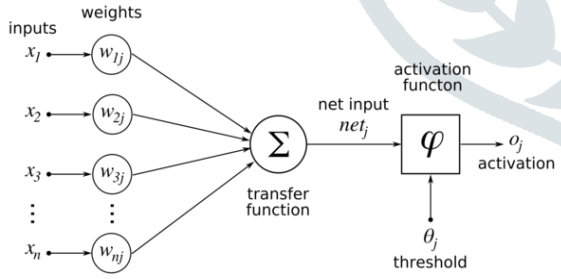


Figure 1: AI basic conceptual diagram

recognition, it was found to be particularly apt for the communication software. Its excellent computational graph visualizations, extensive low-level library that is able to adapt itself to whatever the user requires, fine debugging, scalability and, pipelining abilities, are some of its special advantages. Due to TensorFlow's great voice and sound recognition, it was found to be particularly apt for the communication software.

I. RELATED WORKS

Seppo Enarvi, Peter Smit, Sami Virpioja, and Mikko Kurimo(2017) [1] employed Recurrent Neural Networks(RNN) in order to recognize speech automatically and also efficiently process large conversational Estonian and Finnish vocabularies.

Xie Chen , Member, Xunying Liu ,Yu Wang, Anton Ragni, Member, Jeremy H. M. Wong, and Mark J. F. Gales(2019) [2] used a feedforward unit to model a fixed number of succeeding words for speech recognition.

Steve Kim(2005) [3] created the app TalkEnglish that allows the user to practise speaking and writing English with a variety of interactive courses.

Vu Van and Xavier Anguera(2015) [4] created the ELSA(English Language Speech Assistant) app that focuses on assisting people with their English pronunciation.

Danish Dhamani and Paritosh Gupta(2018) [5] created ORAI to help people become more fluent and confident public speakers.

The above references were vital in understanding how AI was used in speech recognition and in improving English communication skills.

II. SYSTEM DESIGN

The voice inputs that have been passed to the UI will be processed by the ANN model. The voice inputs will be converted to text inputs and sent to the ANN model. The ANN model will be trained by using text inputs passed through JSON files. JSON files contain a rich vocabulary(many tags and related statements) that will greatly augment the efficacy of the training process. The input test data can either be in the form of text or voice, depending on the user's preference. As mentioned before, the focus shall be on thevoice inputs. The trained ANN model's inputs will then be compared with the new voice inputs given by the user, and based on how well they match, the result will be given. It is a false positive if the expected textual outcome is not produced. If there is a proper match, the expected outcome will be produced.

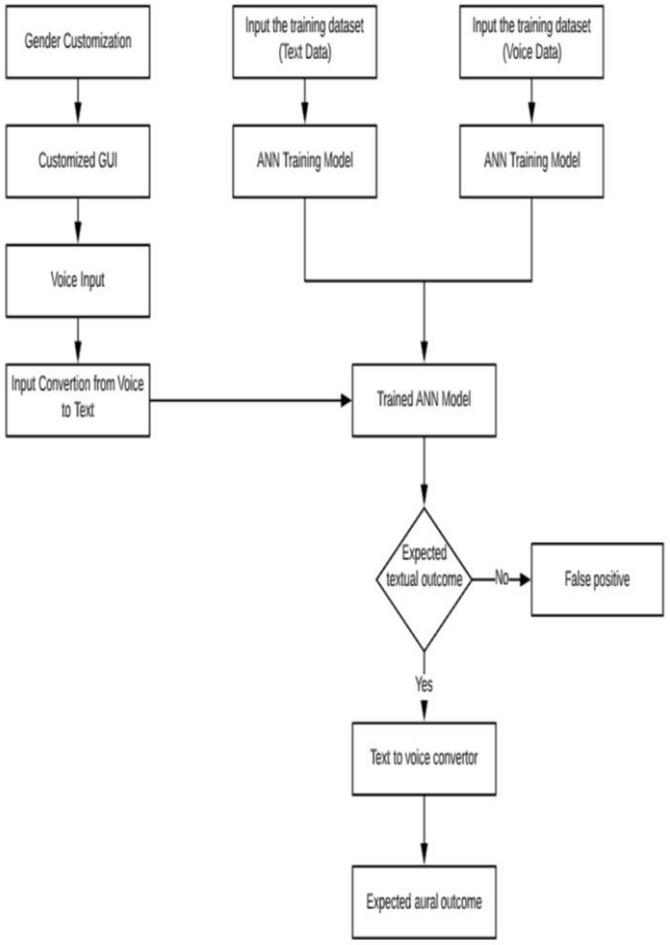
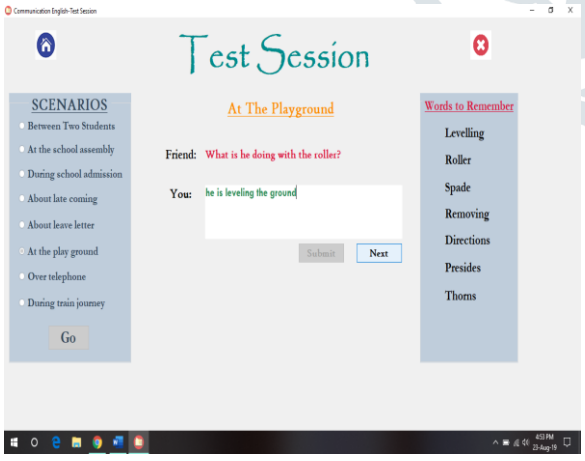
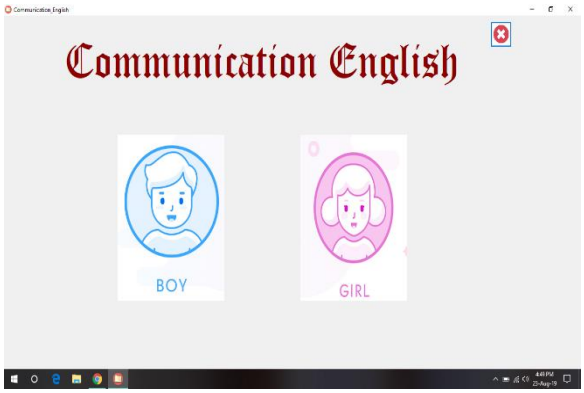
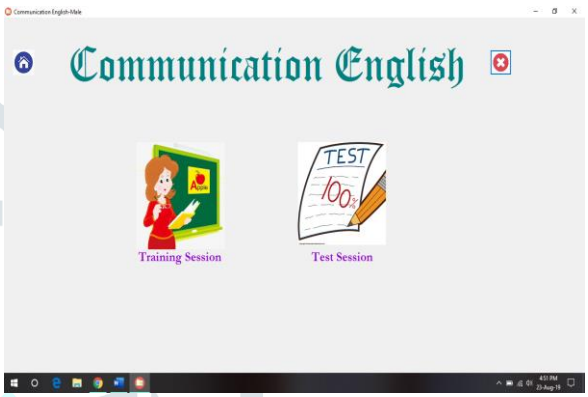


Figure 2: System Design

III. WORKING OF SOFTWARE







The system design of the software's working can be observed in Figure 2. Initially, there will be a page where the user has to select their appropriate gender. After the selection is complete, the user will be redirected to a page where there will be two sessions: training and test.

In both sessions, either text inputs or voice inputs can be provided. Since communication training is the main aim of the project, more focus will be paid on the voice inputs. In both sessions, there will be a list of scenarios available to choose from. On selecting the scenario, a conversation will play out between the system and the user; the system will ask the user the question. There will be a prompt given below that will help the user learn the expected response to the question. Accordingly, the user can provide the voice input with the aid of a headphone

set; the voice input will then be translated into a text input that will be visible on the screen. If the input is correct, the user can then proceed into the next part of the conversation. There will also be a set of words specific to each scenario(Words To Remember), that will help the users acclimate themselves to the new vocabulary that they encounter. In the test session, the users will be able to test the phrases and words they learnt previously, and apply them to check their proficiency. Based on the accuracy of their results, they will be awarded a score that will help them gauge their level of understanding. They can re-attempt the training and test sessions as many times as they wish.

IV. FIELD IMPLEMENTATION

Initially, when the voice-text convertor had not been completely developed yet, the students were restricted to using text inputs when working with the software. It was found that the students had difficulty entering the inputs as they were not very well-versed in basic keyboard typing skills. After the convertor was developed to be able to process voice inputs successfully, it was tested again. Certain students found it hard to speak fluently in English; they were not able to deliver the required voice inputs. While others spoke clearly, the system was still unable to translate their inputs to text. After reflecting on these flaws, meticulous efforts were taken to improve the accuracy and conversion ability of the system. On testing the latest version of the software, the system was able to translate the given voice inputs into text outputs successfully, word for word. A variety of English language accents were used in the inputs to test the versatility of the software. It was able to produce accurate text outputs for all of them.

V. FUTURE WORK

During the development of the software, efforts were made to automate it in such a way that no matter the fluency or rate with which the voice inputs were delivered , the system would be able to process and deliver the outputs at the same rate. This would allow the user to have a more personal experience with the software. A more realistic experience would be simulated as a result of this extension. However, on attempting this, it was found that the accuracy of the software reduced when this change was made. Weighing the scales, it was decided that automation would be put on hold currently. However, a desire definitely exists to continue to work on this in the future, with the goal of successful automation as well as the accuracy of the software being left intact.

REFERENCES

[1] Seppo Enarvi, Peter Smit, Sami Virpioja, and Mikko Kurimo, on “Automatic Speech Recognition With Very Large Conversational Finnish and Estonian Vocabularies”, IEEE/ACM transactions on audio, speech, and language processing, vol. 25, no. 11, November 2017

[2] Xie Chen , Xunying Liu , Yu Wang , Anton Ragni, Jeremy H. M. Wong , and Mark J. F. Gales, on “Exploiting Future Word Contexts in Neural Network Language Models for Speech Recognition”, IEEE/ACM transactions on audio,

speech, and language processing, vol. 27, no. 9, September 2019

[3] Steve Kim(2005) created the app TalkEnglish that allows the user to practise speaking and writing English with a variety of interactive courses. https://www.talkenglish.com/lessonindex.aspx

[4] Vu Van and Xavier Anguera(2015) created the ELSA(English Language Speech Assistant) app that focuses on assisting people with their English pronunciation. https://documentation.skillsoft.com/en_us/percipio/Content/B_Learner/lrnr_elsa.htm

[5] Danish Dhamani and Paritosh Gupta(2018) created ORAI to help people become more fluent and confident public speakers. https://www.orai.com/