March-April 2020
ISSN: 0193-4120 Page No. 1403 - 1408

Speech Recognition in Word Pronunciation

[1]Lina El-Shiny, [2]Zahra Heggi, [3]Nighat Mir, [4]Wadee Al-Halabi
[1,2,3,4]College of Engineering, EFFAT University, An Nazlah Al Yamaniyyah, Jeddah 22332, Saudi Arabia
[1]lalshiny@effatuniversity.ede.sa, [2]zheggi@effatuniversity.edu.sa, [3]nmir@effatuniversity.edu.sa, [4]walhalabi@effatuniversity.edu.sa

Article Info
Volume 83
Page Number: 1403 - 1408
Publication Issue:
March - April 2020

Article History
Article Received*: 24 July 2019*
Revised*: 12 September 2019*
Accepted*: 15 February 2020*
Publication*: 14 March 2020*

Abstract
This plan provides an outline of Automatic Speech Recognition, hardware and apps. Automatic Speech Recognition System is a primary way of communication that is not only among people but also in interaction between human and machine. Speech Recognition is the technique that can transform enunciated phrases into words. It is also a multi-level shape identification process in which computer controlled wave messages are noticed and organized into a structure of semi-word blocks, terms, sentences and phrases. The purpose of this project is to focus on how Speech Recognition m Voice Pronunciation works and its function. Moreover, in the implementation phase the application will compare the voice of Native American English with Non-Native English Speaker. Speech Recognition System efficiency is generally assessed in aspects of precision. In addition, the project includes possible uses of Speech Recognition in Education, Security System, and the possible main uses of the technology in the future.

Keywords: *Speech recognition; matlab; word*

1. Introduction

Automatic Speech Recognition IS an independent software that computer propelled a record of spoken words into clear text in actual time. Automatic Speech Recognition is systems that will make the computer recognize the words recorded by the user's voices via a microphone and convert it to signal wave [1].

Automatic Speech Recognition (ASR) transforms a speech-recognition message to a sense of phrases dependent on the Hidden Markov Model (HMM) [2]. Hidden Markov Models have close association with blend models and the blend model creates information as pursue, likewise words are produced using a succession of conditions, in view of a huge vocabulary of preparing information. The discourse sign depends on phonemes. The arrangement of words made from the discourse sign is returned as recorded content of the discourse input [3].

ASR is mind boggling process whereby a PC changes over the human discourse waves into advanced sign that are then planned to a database of conceivable discourse wave structures and "perceived" as word or expression. The procedure may sound naturally simple, yet for mistake and proper perplexity is extraordinary. Managing the subtleties of human discourse is a standout amongst the most difficulties in discourse acknowledgment [4]. ASR does not have a module that characterizes the speaker's personality, nor do the discourse identification frameworks characterize the meaning of the words. Such frameworks are known as speaker identification and common language handling, and are disengaged elements from programmed frameworks for the identification of discourse. The objective of this framework is to precisely transform a discourse signal into a message content of the speaker's expressed words by a gadget that records the amplifier's discourse guide. [5].

Research on speech recognition needs to begin by considering and discovering the procedures used for this framework. There are six techniques used for Speech Recognition [6] which are Artificial Neural Networks (ANN), Long Term Statistic, Hidden Markov Models (HMM), Ergodic-HMM's, Dynamic Time Warping (DTW), and Vector Quantization (VQ) [6].

Learning another dialect is more earnest for adults between the ages of 20-40. Innovation in automatic speech recognition (ASR) can educate a second language for grown-ups [7]. Express preparation is to vary between the local language and the target language. There are three kinds of words that distinguish proof, word imitation, and word creation. Discourse acknowledgement is the place they give, similar to vocabulary, through high-quality phonetic base structures. Variations in articulation are created from the enlistment discourse. [7].

March-April 2020
ISSN: 0193-4120 Page No. 1403 - 1408

Living in a nation where English is not the primary language makes it difficult for the customer to learn English in a simple and correct way, so the product will probably recognize areas for enhancement by featuring and giving the right elocution and sentence structure models. Improving the nature of English for an undergraduate study is not a simple matter and requires some time and effort [8,9].

Speech Recognition Systems can be use in various examples, Education is one of them that makes learning fast and easy, and are accepted to the individual and by society. Also for the Disabled user a Speech System makes it easy for them to communicate without any problems. The computer allows Automatic Speech Recognition process to get speech signals and automatically converted into the parallel sequence of waves signals. It incorporates various controls, such as physiology, acoustics, signal handling, acknowledgement of design, and semantics. The difficulty of automatic speech recognition comes from many parts of these regions.

Thus in this work was done to focus on how Speech Recognition in Voice pronunciation works and its function. How the signals are converted from O's & 1's to sound waves. In addition, comparing the Native American English with Non-native English Speaker, also comparing and observing how the words will be pronounced and how the two users are different from each other.

2. Requirement Analysis

There are three types of requirements that must be taken into consideration, the functional requirements, non-functional requirements, and other non-functional requirements.

2.1 Functional Requirements

The functional requirement of the system is stated as follow. The system shall get the audio file "Native American English Speaker. The system shall get the audio file "Non-Native English Speaker". The system shall convert the audio files to wave signals. The system shall process the comparison between two audio files. The system shall show the comparison result match "Pass/Failed". The system should able the user to create unlimited tasks. The system should allow the user to attach files for each task. The system should able the user to access the Speech Recognition System whenever/wherever they want.

2.2 Non Functional Requirements

The System should be available at all locations. The effort required to move a program from one equipment arrangement to another, as well as the programming framework condition. The exertion includes changes in information, program changes, work framework, and documentation. The system should be easy to be used by all authorized users and available for all students, faculty and staff. Affirmation that the application will exercise its expected capacity with the required accuracy over an all-inclusive timeframe, and handling accuracy manages the framework's ability to accurately process substantial exchanges. While reliability identifies the framework with the option to perform effectively over an all-encompassing timeframe when the framework is set in generation. The performance of the system should be fast and efficient in getting, converting, creating, updating and printing tasks. Furthermore, the effort required an error in an operating system to be located and fixed. The access to Speech Recognition System should be authorized through a username and password. There are different dashboards for users, considering the functionalities that are authorized for each dashboard access.. The system should be extended and upgraded to add new features and extend its functionality. The system shall respond to user requests m more than 10 seconds maximum. The system shall be available to all users 24 hours a day, 7 days a week. And any downtimes shall not exceed I minute. Users should be able to use it properly after half an hour of training.

2.3 External Requirement

The external requirement consists of hardware and software. The hardware requirement needed is Laptop/ Computer are devices require with a very high quality and specification to implement latest Window updated, high memory storage that can stand the heaviest of application, latest processor Intel Core is or i7, and high resolution quality of screen. In addition, microphone is needed where Microphone is a sense that convert sound wave from user speech to an electronic single that will be used by MATLAB software to process and generate the score for the user.

As for the software, Matlab is the software that will be used for implementing the Speech Recognition System Project and not using any other software as Matlab has already a built in tool kit that helps in signal processing. Audio Recorder is a software which can capture varies sound input to WAVE, MP3 format that can help the user to record the "Native and Non- Native English speaker.

Communication interface is way of enabling one machine to communicate with another machine also it can be use for a point of interaction between a numbers of system or groups works. As firewall and important point to be taken into consideration to prevent our system from malicious or hacking.

3. Design

3.1 Algorithm

The system will get the audio file of Native English American Speaker and convert it wave signals. Secondly, the user will get the audio file of Non-Native English Speaker and convert it to wave signals. Comparison between the two signals will be done to get the score. If the score will be greater than 70% a congratulation

message will be shown and the system will exit. While the score is less than 70% an output message will be shown informing the user that he/she failed and has to try again. The user will try for the second trial and compare again if the score is still less than 70% an output message will be shown informing the user that his/her English needs some improvement, and the program will exit. The algorithm is stated as follow.

Start
Read Audio from Native English American Speaker
Convert it to wave signal
Count=O Pass-Score=70 No of Repetitions=2
Do
Read Audio from Non-Native English Speaker
Convert it to wave signals
Compare signals and get the scores
If score 2:pass-Score then
Success Match between the two audios,
CONGRATULATIONS!!!
Else
Matching between the two audios failed, TRY AGAIN!!!
Count=Count+ 1
End if
While (Score=S70 && Count<No of Repetitions)
If (Score=S70 && Count==2)
An output message will be shown informing the user that his English needs
some improvements!!!
End if
Stop

3.2 System Design

The system is designed as follow. First, the system will get the audio file of Native English American Speaker and convert it wave signals. Secondly, the user will get the audio file of Non-Native English Speaker and convert it to wave signals. Comparison between the two signals will be done to get the score. If the score will be greater than 70% a congratulation message will be shown and the system will exit. While the score is less than 70% an output message will be shown informing the user that he/she failed and has to try again. The user will try for the second trial and compare again if the score is still less than 70% an output message will be shown informing the user that his/her English needs some improvement, and the program will exit. Figure 1 shows the system data flow.

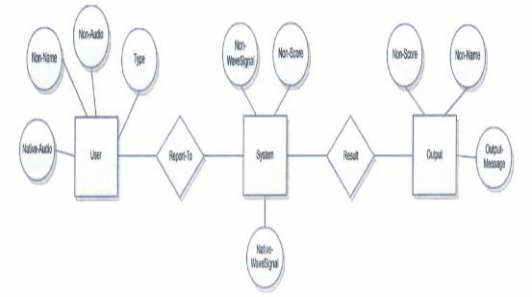

Figure 1: System Flow

3.3 Workflow

In Speech Recognition System Workflow the system will get the audio file of Native English American Speaker and convert it wave signals. Secondly, the user will get the audio file of Non-Native English Speaker and convert it to wave signals. Comparison between the two signals will be done to get the score. If the score will be greater than 70% a congratulation message will be shown and the system will exit. While the score is less than 70% an output message will be shown informing the user that he/she failed and has to try again. The user will try for the second trial and compare again if the score is still less than 70% an output message will be shown informing the user that his/her English needs some improvement, and the program will exit. The system is designed as follow. First, the system will get the audio file of Native English American Speaker and convert it wave signals. Secondly, the user will get the audio file of Non-Native English Speaker.

3.4 Sequence

The sequence diagram for Speech Recognition System is as follow. The program will show how the objects "User, System and Out-put will interact with each other in which particular time sequence? The user will get the audio files for "Native & Non-Native Speaker"; the system will convert it to wave signals and processing the comparison between the audio files to generate the score by the system. If the score is greater than the specified a congratulation message will be shown to the user and existing the program, while the score is less than the specified then an output message will be shown asking the user to try again, failed. The user will try the second trial if still having the same problem; a message will inform the user that his/her English needs some improvements, existing the program. The Sequence Diagram of Speech Recognition System is shown in Figure 2.

March-April 2020
ISSN: 0193-4120 Page No. 1403 - 1408

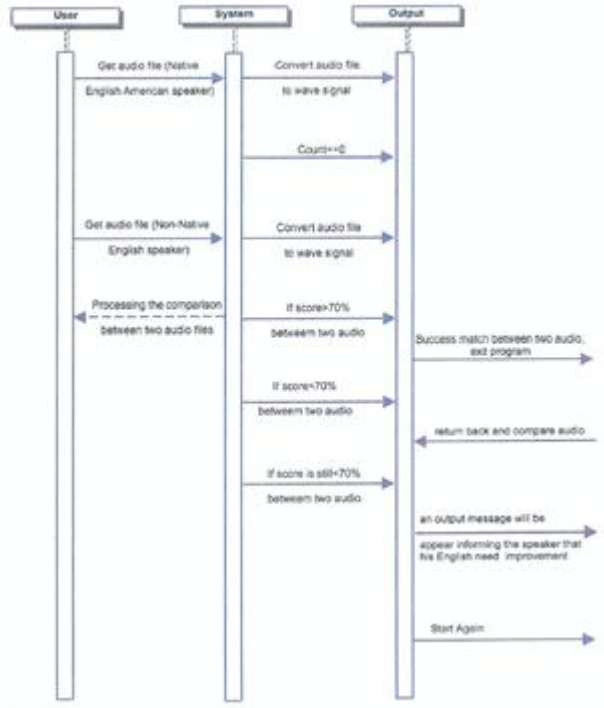

Figure 2: Sequence Diagram of Speech Recognition System

3.5 UML "Unified Modeling Language"

UML includes a number of genuine paperwork processes to create graphic designs of programming organized by items. UML is used to determine, modify, imagine, construct and publish an article's old rarities. UML connects applications from showing data, showing company, showing objects, and showing segments. It can be used with all processes throughout the life cycle of item enhancement and cross-sectionally through multiple utilization technologies. Figure 3 shows the Unified Modeling Language of Speech Recognition System flow.

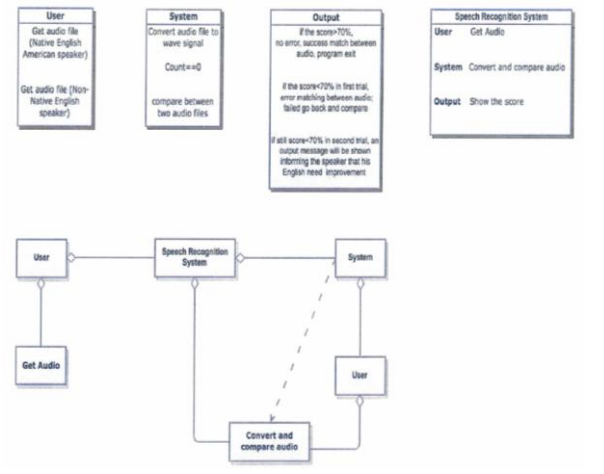

Figure 3: Unified Modeling Language of Speech Recognition System

3.6 DFD's "Data Flow Diagram"

It is a graphical portrayal of the progression of information through a data framework to show its angles of procedure. They are steps used to outline the framework that can be explained later. DFSs can also be used to represent the preparation of information. It will talk to what kind of data the info will be and the frame's yield can not be avoided. It will show where the information will originate from and also where the information will be stored.

First, the system will get the audio file of Native English American Speaker and convert it wave signals. Secondly, the user will get the audio file of Non-Native English Speaker and convert it to wave signals. Comparison between the two signals will be done to get the score. If the score will be greater than 70% a congratulation message will be shown and the system will exit. While the score is less than 70% an output message will be shown informing the user that he/she failed and has to try again. The user will try for the second trial and compare again if the score is still less than 70% an output message will be shown informing the user that his/her English needs some improvement, and the program will exit.

3.7 ERD "Entity Relationship Diagram"

ERD is a dynamic method for planning and depicting a database III programming building, as a rule it starts with a social database. It portrays a database that stores tables of information III. For example, a portion of the information III these tables point to information III tables, your entry III database could point to a few sections. Charts used to make these substances and connections are called element relationship graphs or ER outlines. Entity Relationship Diagram in Speech Recognition System that has three main actors, the user, system and the output. The user contain the Native Audio file, Non-Native Audio files, and the type of the audio "MP3, etc" that will be reported to the system that will convert the audio files to wave signals and process comparison match between the audio that result to the output were the user name, user score and the output message "Passed/Failed" will be shown on the screen.

3.8 Use Case View- System Requirement

Use case is a summary of steps, typically characterizing cooperation between a job "referred to as an actor in UML," a framework, and a yield to achieve an objective. The on-screen character can be a human or external framework. The Speech Recognition System has three actor's user, system and an output. The actor will upload and get the file of "Native English American Speaker" and the "Non-Native English Speaker". In addition, the system will convert the files to wave signals and processing the comparison between the two audio files to get the score. The output of the use case will show the score of the files. If the score is greater than 70%, a

March-April 2020
ISSN: 0193-4120 Page No. 1403 - 1408

congratulation message will be shown and the system will exit. While the score is less than 70%. Figure 4 shows the use case of speech recognition system.

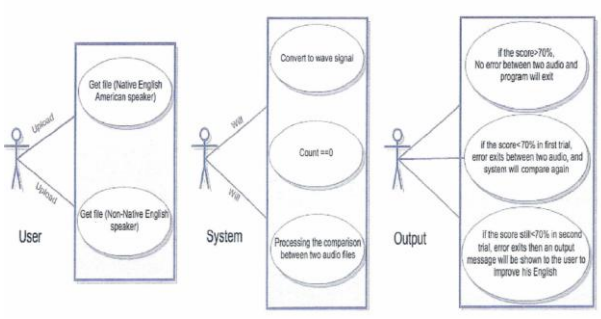

Figure 4: Use Case of Speech Recognition System

4. Prototype Testing

A prototype is an early example or model for the Speech Recognition System that is used to test an idea or process or go about as something to imitate or gain from. It is a term used in a variety of settings, including semantics, structures, and programming. It is intended to test and preliminary to upgrade the accuracy of framework. The prototype consists of four pages, the home page, about us page, the design and the contact us page.

The home page is the first page of the prototype a that will show the title and a picture that explains the main idea of the project in a form of steps as shown in Figure 5.

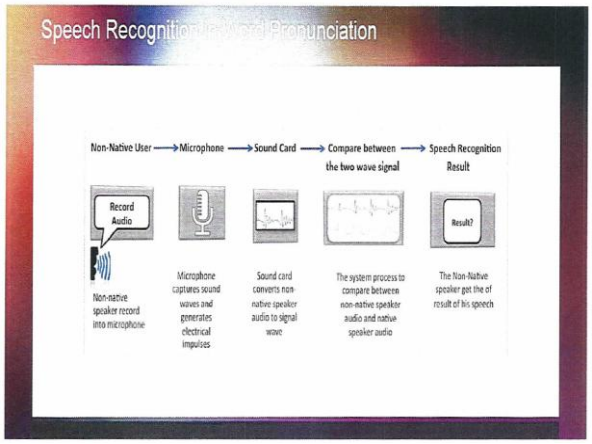

Figure 5: Prototype Home Page

The about us page will be discussing the Speech Recognition project, by giving a brief description about the main idea and our objectives. In addition, a small description about the partners where and thanking the guests for view mg the prototype and knowing about the project.

There are seven pictures in the design page of the Speech Recognition System that will be shown with their names. They are the Flowchart of the system, Use Case Diagram, Workflow Diagram, Data Flow Diagram, Sequence Diagram, UML Diagram and the Entity Relationship Diagram. The pictures will be shown in a3D view.

The last page of the Speech Recognition System Prototype will be the contact us page that will contain the contact information of the two partners of the project like "Name, Phone Number, Email and the Address". partners through sending them a message by writing their "Name, Email-Address, and there Message" and the project patterns can reply to the message

The main aim of the Speech Recognition System testing is to spot the problems and errors that can only be shown by testing the whole system.

This test IS planned to make sure that the system can process its planned weight. In this test, different kinds of actions will be performed to raise the load on the system, until the system performance becomes intolerable. This method is called a 'stress testing' which will cause defects and difficulty to become system strength.

5. Conclusion

The purpose of this project is to focus on how Speech Recognition m Voice Pronunciation works and its function. Moreover, in the implementation phase the application will compare the voice of Native American English with Non-Native English Speaker. Speech Recognition System efficiency is generally assessed in aspects of precision and velocity. Accuracy is generally ranked with phrase and mistake frequency, while velocity is evaluated with the concurrent factor. The project includes possible uses of Speech Recognition in Education, Security System, and the possible main uses of the technology in the future.

References

[1] Saini, P., & Kaur, P. (2013). Automatic speech recognition: A review. *International Journal of Engineering Trends and Technology*, *4*(2), 1-5.

[2] Lee, K. S. (2008). EMG-based speech recognition using hidden Markov models with global control variables. *IEEE Transactions on biomedical engineering*, *55*(3), 930-940.

[3] Hershey, J. R., Rennie, S. J., Olsen, P. A., & Kristjansson, T. T. (2010). Super-human multi-talker speech recognition: A graphical modeling approach. *Computer Speech & Language*, *24*(1), 45-66.

[4] Schroeder, M. R. (2013). *Computer speech: recognition, compression, synthesis* (Vol. 35). Springer Science & Business Media.

[5] Ince, A. N. (Ed.). (2013). *Digital Speech Processing: speech coding, synthesis and recognition* (Vol. 155). Springer Science & Business Media.

[6] Gaikwad, S. K., Gawali, B. W., & Yannawar, P. (2010). A review on speech recognition

technique. *International Journal of Computer Applications*, *10*(3), 16-24.

[7] Wang, Y. H., & Shwu-Ching Young, S. (2014). A Study of the Design and Implementation of the ASR-based iCASL System with Corrective Feedback to Facilitate English Learning. *Journal of Educational Technology & Society*, *17*(2).

[8] Crystal, D. (2013). A global language. In *English in the World* (pp. 163-208). Routledge.

[9] Hussain, A., Mkpojiogu, E.O.C., Yusof, M.M. (2016). Perceived usefulness, perceived ease of use, and perceived enjoyment as drivers for the user acceptance of interactive mobile maps. AIP Conference Proceedings, 1761.

[10] T. Padmapriya, S.V. Manikanthan, "LTE-A Intensified Voice Service Coder using TCP for Efficient Coding Speech", International Journal of Innovative Technology and Exploring Engineering, Vol. 8, issue 7s, 2019. https://www.ijitee.org/wp-content/uploads/papers/v8i7s/G10630587S19.pdf.