iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 **Issue:** IV **Month of publication:** April 2019

DOI: https://doi.org/10.22214/ijraset.2019.4635

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 7 Issue IV, Apr 2019- Available at www.ijraset.com

Speech Analysis based Pronunciation Correction Training for School Student

Mrs. Kranti Shailesh Tol[1], Dr. T. B. Naik[2]
[1]*Research Student, Dept. of Education, Shivaji University, Kolhapur*
[2]*Asst. Prof. Government College of Education, Ratnagiri*

Abstract: Pronunciation accuracy enhancement is the need of education for school students during career development. The correct speech with exact pronunciation is what one can learn by understanding the phonetics definitions available in various dictionaries and study material of the English language. There is a need for a system for teaching the pronunciation, and that varies when run by humans based on individual teacher's performance. Also, due to variations in language, as practically every 25 kilometers of the span is responsible for variation in language, a computerized system for pronunciation learning can provide a solution to the problem. In this paper, we have provided speech analysis based tool for pronunciation accuracy improvement system which makes use of Google speech API for text to speech conversion. The speech of the user is recorded using a microphone. The recorded speech is analyzed based on MFCC features and compared with that of the speech obtained from google speech API. The improvement for correct pronunciation is seen amongst students when they are provided the tool for experimentation as comparison is obtained in terms of percentile score along with speech recognition and then converting it into the text from improper to proper speech scenarios. The trials are performed on more than 200 school students which shows satisfactory results.

Keywords: Pronunciation accuracy, speech analysis, MFCC, tool, speech to text, text to speech.

I. INTRODUCTION

Nowadays in several application human-machine interaction is widely used. Detection of feelings from speech may be the main challenge in human and machine. People express their emotions through different modalities like facial expressions, body poses and speech signal. We tend to completely focus on the speech signal; this paper varied reasonably options that carry data regarding the message, speaker, language and emotion. [1]The approach is to calculate numerous options that carry a lot of data and identify the features to search out a strong recognition rate. [2] Speech based analysis to recognize is the best way to understand the actual intensions of the speaker. This programme developed is helpful to identify the various errors that an individual adapts due to the influence of regional language over command of the English language.

II. RELATED WORK

Under this point, the focus is on the literature that is available for speech and programme development for enhancing the speaking and listening skill. There are a lot of people who have worked on developing programmes to enhance language skills by using different features and classifiers. The speech emotion recognition system has used various processes and techniques for feature extraction and getting the final output. Some papers have been introduced here.

Anagha Sonawane et al. (2017) have shown a method for emotion recognition which consists of MFCC based speech signal features extraction. After this, features are classified according to labels using Multiple Support Vector Machine (SVM). Emotional speech signals for happy, anger, sad, disgust, surprise and neutral are used in the database. The results obtained by the author are satisfactory.

H. K. Palo et al., (2015) focused on feature extraction using methods like Linear prediction coefficients (LPC), linear prediction cepstral coefficient (LPCC), Mel frequency cepstral coefficient (MFCC) and perceptual linear prediction (PLP) this two features are used for emotion recognition using multilayer perception (MLP). In this paper, they recognize two classes of speech emotions as induces like angry and surprise gives a strong response and at low responses like sad and boring.

Zhaocheng Huang, (2015) In this research, focused on emotion transition problem will be investigated, including localizing emotion change points, recognizing emotion transition patterns and predicting or recognizing emotion changes.

Steven A. Rieger Jr et al. (2014) The focus of this paper is on the study of speech-based emotion recognition using a pattern recognition paradigm with spectral feature extraction and an ensemble of k nearest neighbor (KNN) classifiers. The classifier based grouping is done as an unsupervised learning approach for the classifier, to detect the emotion from the speech.

Variety of techniques are involved in speech feature extraction which varies according to applications. The speech-based pronunciation estimation mainly focuses on vowels, diphthongs, consonants, diagraphs, and blends. These features in the speech signal can be detected by using MFCC feature analysis. The method of MFCC feature extraction is detailed in the next section.

III. PROPOSED WORK

The proposed system consist of the steps as mentioned:

1) Take text input for the word to be pronounced
2) Convert text to speech using Google API to obtain speech audio file
3) Record speech of user from microphone
4) Extract MFCC features of user-recorded sound file and speech sound file from Google API.
5) Compare MFCC features to find the error percentage using Euclidian distance.
6) The speech recorded from the user is also converted into text to check whether exact word is spoken by the user.
7) When step 6 shows the desired value of matching level, it is displayed to the user to know how much correct pronunciation is done.
8) Finally, steps from 1 to 8 can be repeated for the same word for improvement of score or for some other word.

A. Feature Extraction

The speech signal is first preprocessed to remove silent zones and finding actual audible windows. The MFCC features are extracted as detailed further.

B. MFCC Feature Extraction

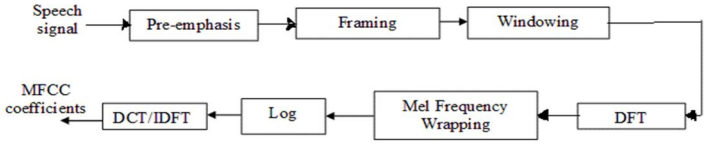


Figure 1. Calculation of MFCC Coefficients.

There are two sorts of the filter of MFCC that are spaced linearly at a low frequency below one thousand Hertz and power (logarithmic) spacing on top of 1000Hz. During this technique, Pre-emphasis could be a method of passing the input through a filter that emphasizes higher frequencies. Framing is that the step next to pre-emphasis within which the speech samples obtained from analog to digital conversion (ADC) are divided into a little frame with the length inside the vary of 20 to 40msec time unit. Once the framing method, windowing method is performed. During this technique, acting window is employed that reduces the signal discontinuities at the beginning and finish of every frame and frame is shifted with 10ms span. The method of FFT converts every frame of N samples from the time domain into the frequency domain. The frequencies direct FFT spectrum is extremely wide then FFT algorithmic rule is employed for changing n samples from the time domain to frequency domain. Mel frequency reduces this scale and identifies what proportion of energy exists during a frame.

$$F(Mel) = [2595 * \log 10 [1+ f] * 700]$$

DCT convert the log Mel spectrum into time domain [2] [10].

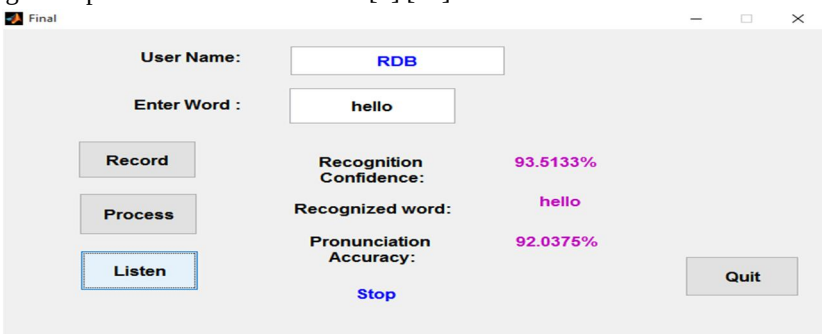


Figure 2: an implemented tool for word pronunciation accuracy estimation

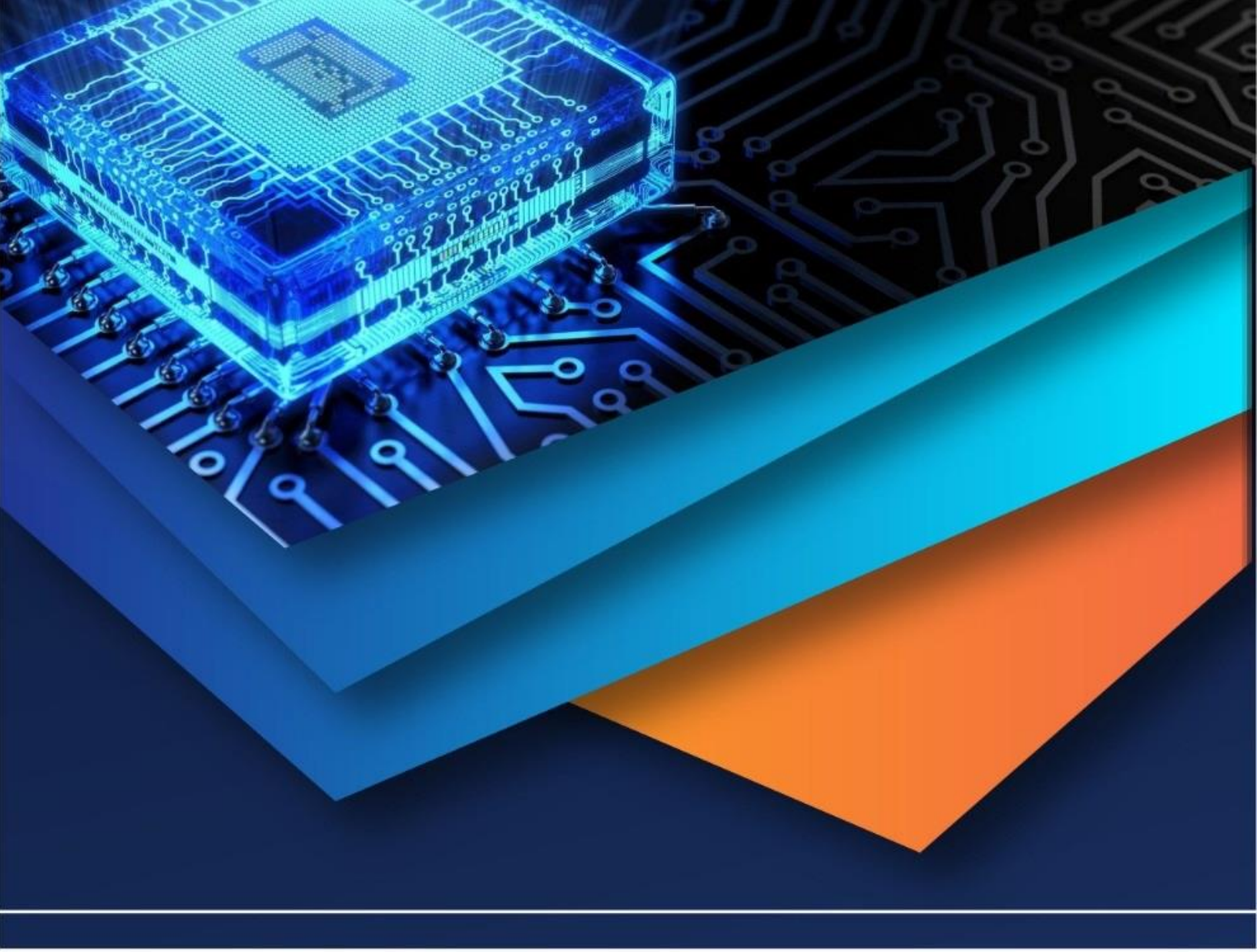
Figure 2 shows the tool implemented for word pronunciation accuracy estimation and improvement using MATLAB platform.

IV. CONCLUSION

In this work, we have developed pronunciation accuracy improvement training tool using MATLAB platform. The pronunciation accuracy improvement trials are conducted with experimentation on more than 200 students group of secondary school level. During the trials, students were made to listen to correctness in pronunciation from speech obtained using Google API. The students changed their way of pronunciation to that of the desired level, and evaluation tool have shown results in terms of percentile improvements. Also, the work is extended to perform trials on sentences of the English language. The results obtained in both the scenarios are satisfactory.

REFERENCES

[1] M.S. Likitha, Sri Raksha R. Gupta, K. Hasitha and A. Upendra Raju, "Speech Based Human Emotion Recognition Using MFCC", This full-text paper was peer-reviewed and accepted to be presented at the IEEE WiSPNET 2017 conference.
[2] Anagha Sonawane, M.U.Inamdar, Kishor B. Bhangale, "Sound-based Human Emotion Recognition using MFCC & Multiple SVM",IEEE International Conference on Information, Communication, Instrumentation and Control(ICICIC 2017)
[3] Yixiong Pan, Peipei Shen and Liping Shen,"speech emotion recognition using support vector machine",International Journal of smart home,vol.1,21 January 2014.
[4] Milton,S. Sharmy Roy,S. Tamil Selvi, PhD., "SVM Scheme for Speech Emotion Recognition using MFCC Feature", International Journal of Computer Applications (0975 – 8887) Volume 69– No.9, May 2013
[5] H.K. Palo, Mihir Narayanan Mohanty and Mahesh Chandra, "Use of Different Features for Emotion Recognition Using MLP Network", © Springer India 2015 I.K. Sethi (ed.), Computational Vision and Robotics, Advances in Intelligent Systems and Computing 332, DOI 10.1007/978-81-322-2196-8_2
[6] Zhaocheng Huang, "An Investigation of Emotion Changes from Speech", 2015 International Conference on Affective Computing and Intelligent Interaction (ACII)
[7] Steven A. Rieger Jr., Rajani Muraleedharan, Ravi P. Ramachandran, "Speech Based Emotion Recognition Using Spectral Feature Extraction and an Ensemble of Knn Classifiers",2014 9th International Symposium on Chinese Spoken Language Processing (ISCSLP)
[8] Qingli Zhang, Ning An, Kunxia Wang, Fuji Ren and Lian Li , "Speech Emotion Recognition using Combination of Features", 2013 Fourth International Conference on Intelligent Control and Information Processing (ICICIP) June 9 – 11, 2013, Beijing, China
[9] Supriya B. Jagtap, Dr. K. R. Desai, Ms. J. K. Patil, " A Survey on Speech Recognition Using MFCC and Different classifier", 8th National conference on emerging trends in engineering and technology (NCETET-2018),ISBN:978-93-87793-03-3
[10] Leh Luoh,Yu-zhe Su and Chih-Fan Hsu, "Speech signal processing based on sound recognition", 2010 International Conference on System Science
[11] F. Yu, E. Chang, Y.Q. Xu, and H.Y. Shum, "Emotion detection from speech to enrich multimedia content", in Proc. 2nd IEEE Pacific-Rim Conference on Multimedia 2001, pp.550-557, Beijing, China, October 2001.
[12] Tsuyoshi Moriyama, Shinya Mori, and Shinji Ozawa. "A synthesis method of emotional speech using subspace constraints in prosody". Journal of Information Processing Society of Japan, 50(3):1181–1191, 2009. (in Japanese).
[13] Introduction to Principal Components and Factor Analysis, AtchleyOct19



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY