International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Issue 10, pp 66-73, October-2023

https://doi.org/10.47001/IRJIET/2023.710009

Smart Speaker: Enhancing Any Persons' Ability to Deliver English Speeches Independently With a Web Application

[1]Malshan E.G.C, [2]Buddhini B.A.D, [3]Isurandi I.G, [4]Habalakkawa W.V.K.I, [5]Suranjini Silva, [6]Thamali Kelegama

[1,2,3,4,6]Department of Information Technology, Sri Lanka Institute of Information Technology, Malabe, Sri Lanka

[5]Department of Computer System Engineering, Sri Lanka Institute of Information Technology, Malabe, Sri Lanka

***Abstract* - This paper introduces a state-of-the-art online application called "SMART SPEAKER" to improve English speaking abilities, especially in public speaking. The system uses Machine Learning, Deep Learning, and Natural Language Processing Techniques to evaluate the user's speech based on various aspects such as Content analysis, Flow Completeness analysis, Grammar analysis, and Facial Expressions analysis. The tool is designed to be user-friendly and simple, providing an easy and efficient solution for those looking to improve their English-speaking skills, gain confidence, and deliver well-articulated speeches. This system meets the growing demand for a practical and effective tool that can support English speakers around the world. Through "SMART SPEAKER", users can practice and improve their public speaking skills.**

Keywords: Text analysis, sentiment analysis, facial expression analysis, voice-to-text, content analysis, and speech analysis.

I. INTRODUCTION

The ability to speak English is incredibly important in today's world. It serves as the global language of communication, enabling people from diverse backgrounds to connect and understand each other. Proficiency in English opens doors to educational and career opportunities, as many prestigious institutions and multinational companies require strong English skills. Moreover, English provides access to a wealth of information and resources online, fosters cultural understanding, and allows individuals to express themselves effectively.

Effective communication is crucial in various aspects of life, including public speaking, presentations, and interpersonal interactions. However, many individuals struggle with delivering speeches confidently, maintaining appropriate facial expressions, and organizing their ideas coherently. Traditional speech coaching methods often require a personal instructor or rely on subjective feedback, which can be time-consuming and expensive. To address these challenges, an online tool has been developed that aims to assist individuals in crafting and delivering successful speeches.

The problem lies in the difficulty that individuals face when attempting to deliver an impactful speech. They often lack guidance on how to effectively express their ideas through facial expressions and struggle with grammar errors and structuring their speech coherently. Additionally, access to professional speech coaches may be limited or unaffordable for many people. Thus, there is a need for a convenient and accessible tool that can help individuals enhance their speech delivery skills, improve grammar proficiency, and ensure meaningful content.

The significance of the proposed online tool is twofold. Firstly, it empowers individuals by providing them with a self-guided platform to develop their speech-making abilities. By offering real-time analysis of their facial expressions and feedback on grammar errors, the tool enables users to identify and rectify their weaknesses independently. This promotes self-confidence and fosters continuous improvement in public speaking skills.

Secondly, the tool addresses the issue of accessibility. By being available online, it eliminates geographical and financial barriers, allowing individuals from various backgrounds to benefit from speech coaching. This democratization of speech training facilitates personal and professional growth, as effective communication is crucial in numerous domains, such as business, academia, and personal relationships.

The research objectives are to develop a comprehensive online tool "Smart Speaker" that combines various functionalities to enhance users' speech-making abilities. This includes analyzing speech content, assessing facial expressions, evaluating speech flow, and identifying grammatical errors. The tool aims to provide real-time feedback on grammar and facial expressions to help users refine their delivery and align expressions with the message effectively. Additionally, enhancing the tool's natural language

ISSN (online): 2581-3048
Volume 7, Issue 10, pp 66-73, October-2023
https://doi.org/10.47001/IRJIET/2023.710009

processing capabilities will generate accurate speech suggestions and improve overall coherence.

1.1 Literature Review

In 2014, a novel method for speech recognition utilizing Convolutional Neural Networks (CNNs) was introduced by a group of researchers led by Aabdel-rahman et al. In this research Details [1] on this approach can be found here. Their paper describes how CNNs can be effectively applied to certain types of speech variables. The study reported significant reductions in error rates between 6% and 10% and improvements in overall performance compared to other methods.

With this project [2] researchers hope to help a person improve their public speaking skills. This is a web-based platform. This explains how the speech trainer system evaluates many aspects of speech delivery, including fluency, pronunciation, and intonation, using various speech processing approaches such as speech-to-text, sentiment analysis, and prosody analysis. To provide an easy and user-friendly product that can be used to boost confidence and make expressive speeches.

In 2016, a smartphone application was developed as Team Ummo. All it does is listen to what a user is saying and provide instant feedback to improve the user's communication skills. This facility is often used by individuals, businesses, professional speaking coaches, educators, clients, employees, students preparing for interviews and presentations. Designed to help users improve their public speaking skills. In addition to its speech analysis capabilities, it also includes a series of tools and exercises to help users improve their public speaking skills. These include customizable practice exercises, guided exercises, and real-world scenarios. Overall, the Ummo app is a valuable tool for anyone looking to improve their public speaking skills.

This [4] is a speech training software. This software provides users with information about their linguistic abilities. Here, an average person speaks about 150 words per minute, and how to calculate the user's speaking speed relative to the total number of words spoken is explained, as well as a mobile application designed to help users improve their public speaking skills. The app uses artificial intelligence to provide feedback and coaching on various aspects of speaking, including speed, filler words, and overall fluency.

II. METHODOLOGY

This research comprises four main components: a Facial Expressions analyzer, Content analyzer, Flow Completeness analyzer and Grammar analyzer. The utilization of natural language processing (NLP) allows for the analysis of these components, while machine learning techniques are employed to detect optimal uses for content. Additionally, deep learning methods are utilized to analyze user facial expressions based on speech. Each of the four components is analyzed individually, contributing to the overall research endeavor.

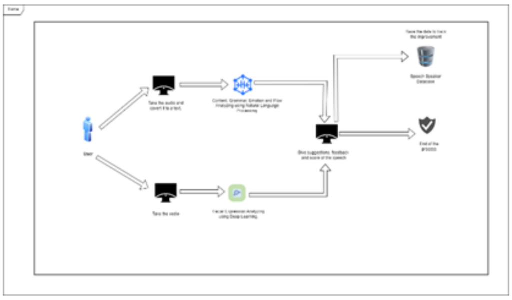

Figure 1: System Overall Architecture Diagram

2.1 Make the speech videos set up for analysis.

The system follows a process when a user uploads a speech video. First, it converts the video into its audio and video components. Then, it uses Google speech recognition to convert the audio into text. However, while the tool can understand some of the speech's punctuation, it has difficulties in identifying full stops. To overcome this issue, the system divides the audio file into sections using silences. Each pause indicates the end of a sentence, and thus, serves as a full stop. At the end, the system incorporates speech-to text conversions in audio, video, and other formats as inputs to each of its parts.

2.2 Emotion Analyzer

In this part, it is basically to analyze how the speaker's facial expressions are used based on emotions and to give score for it and comments regarding the appropriate use of facial expressions.

2.2.1 Using audio converted to text (NLP), identify emotions

This method provides emotional recognition by extracting relevant content from changing content. The system first uses the converted words and first goes through some steps so that the tokens are extracted from the thought words. The first steps include removing all stops and punctuation, tokenization, and lexicalization. In addition, the token set is checked from the dictionary of thought words. To analyze the view in the text, the glossary was created manually from various views. Of the main words, "adjectives, adverbs, idioms or conjunctions" will be in the emotional category. The system then uses a dictionary to determine the meaning of the thought in the converted speech. The system then performs a sensitivity analysis using NLTK's Emotional Strength Analyzer model to identify core emotions such as anger,

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048
Volume 7, Issue 10, pp 66-73, October-2023
https://doi.org/10.47001/IRJIET/2023.710009

happiness, neutrality, and sadness. This is a simple way to measure emotion. Then put the logical expression into an array [5-6].

2.2.2 Using video analysis, identify emotions

Convolutional neural networks (CNN) have become an important method in various image classifications, computer vision, image processing and face recognition. CNNs have performed excellently in tasks related to image analysis, recognition and classification. CNNs have proven to be the best at image classification compared to other methods, making them a good, automated product extractor. In this study, a CNN-based method is proposed to create a face recognition system. The classification algorithm is based on a deep convolutional neural network (DCNN) and uses proprietary algorithms. The design process includes the input process, the convolution process, the nonlinear process, the layers, and the full layers. Leveraging the power of CNNs, the model aims to accurately classify facial expressions real time applications.

Deep learning models developed to analyze the speaker's face or thought using the video their speech. The model is trained on the Fer2013 dataset, which consists of seven emotional categories including Angry, Abhorrence, Fear, Happiness, Sadness, Surprise, and Neutral. However, for the sake of accuracy, the trained model focuses on Recognizing four primary emotions: Angry, happy, sad and neutral. The system takes the outputs from the emotion recognition process, which involves analyzing both compare video and text files and results. By comparing the recognized emotions in the speaker's facial expressions with the content of their speech, the system produces a score that indicates the degree of alignment between the two. This analysis helps determine whether the speaker's facial expressions match the emotional content of their speech. [7-9]

2.3 Content Analyzer

2.3.1 The System, Identifies the introduction part, body part, conclusion part [10-11]

Table I

Introduction	First 10%-15% part of the speech
Content	Middle Part
Conclusion	Last 10%-15% part of the speech

IECI = Introduction Ending Character Index

$$IECI = (15/100) * Total\ character\ count$$

CSCI = Conclusion Starting Character Index

$$CSCI = (85/100) * Total\ character\ count$$

Use best practices [12]

First, in this section, the introduction and conclusion of the speech examine and evaluate how well the system uses best practices. In general, the best practices for introductions are to start with a strong statement. Start by posing rhetorical questions, start with an interesting speech, start with a compelling quote. Starting by asking a question. There are some best practices for ending a speech. Summarizing the main points, reminding the main thesis or central idea of the speech, ending with a memorable quote, inspirational speech or powerful speech related to the topic, ending the speech with a rhetorical question, ending with a punch quote, ending by asking a question, etc.

It can be shown as a best practice to use asking a question for both introduction and conclusion. The system is identified as a question based on the pattern of a certain sentence.

2.3.2 Check content relativity

First, system retrieves the title and checks for similarities. This is done between all words in the content of the speech and with all words related to the topic [13]. The system returns the count for this. Once the system scores, all scores will be added up. The system performs relative presentation by score.

Table II

Relativity	Percentage	Marks Status
Totally Irrelevant	0%	Bad
partially related	0% - 40%	Good
related	40% - 80%	Very Good
fully Related	80% - 100%	Excellent

2.3.3 Check the clearance of the content

The clarity of a speech in public speaking is influenced by several factors: speed of speech, sentence structure, and word choice. The speed at which a speaker delivers their speech can impact how well the audience understands and follows the message. A normal person speaks between 125 and 150 words per minute [15]. Speaking too fast overwhelms listeners while speaking too slowly can lead to disconnection. Finding a comfortable pace is crucial for clear pronunciation and audience comprehension. To assess speech clarity, the system encodes the speech word by word and calculates the number of words in the submitted text.

WPM = Words Per Minutes

Speaking rate (wpm) = Total words / Number of minutes

- ❖ Words Per Minute < 125 – slow
- ❖ Words Per Minute > 150 – high

This section focuses on providing feedback on the user's speech speed.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Issue 10, pp 66-73, October-2023

https://doi.org/10.47001/IRJIET/2023.710009

2.3.4 Content Suggester

Suggests key information related to the topic. The system automatically searches the Internet using keywords. The system shows the found information to the user as other options. The computer performs an automatic search on YouTube for lectures on the user's topic.

2.4 Flow Completeness Analyzer

The effectiveness of a speech largely depends on its flow, and this system aims to analyze it based on two key factors. The first factor is to identify the use of pause fillers and filler words by the speaker. This analysis helps to recognize areas where the speech may lack coherence or need improvement. The second factor involves identifying if the speaker repeats certain words or gets stuck in silence during the speech. This analysis helps to determine if the speaker is having difficulty conveying their message effectively. Together, these factors help the system to evaluate the flow of the speech and provide feedback to the user for improvement.

2.4.1 Detect the pause fillers and filler words score of the speech

Filler words are often used in discourse to mark pauses or hesitations, but they do not contribute any meaningful information to the speech. These words can have a negative impact on the coherence and flow of discourse. To address this, the system first converts the audio component of Relativity Percentage Marks Status Totally Irrelevant 0% Bad partially related 0% - 40% Good related 40% - 80% Very Good fully Related 80% - 100% Excellent discourse into text and then tokenizes the text into individual words. It then checks for the presence of using a filler words dictionary that has been compiled from various sources. The system displays the identified filler words and their corresponding score to the speaker, helping them to identify areas where they may need to improve their speech. [16-18].

Filler words, as opposed to pause fillers, are sounds used to signify a pause or reluctance in speech, for example "mmm" or "ah," and can't be turned into text by Google's speech recognizer. This system uses the wav audio file and the pydub library to recognize and calculate the frequency rates of the pauses in the fillers in order to detect the frequency of pause fillers in a speech. The FPM value is then generated using this data, giving the speaker a precise estimation of the proportion of pause fillers in their speech.

The formula for FPM calculation is:

FPM = # of filler pauses Total time

FPM = Filler Pauses Per Minute

2.4.2 Recognize the repeated words and silent phases

To analyze the flow of discourse, the system examines the speech's audio file for instances of inappropriate silences and word repetitions, which can adversely impact the speech's coherence. The system establishes a threshold value for silence phases, and if the users remain silent for longer than this limit, it is identified as an unsuitable silent phase. Moreover, the system detects repeated words or phrases by utilizing Google speech recognition to convert the repetitions into text, which is then analyzed using a basic algorithm to determine which words were repeated and were. This examination aids the system in identifying areas of the speech where the speaker may require improvement in their delivery and flow [19].

2.5 Grammer Analyzer

2.5.1 Analyze the grammatical precision

The speech is initially inputted into the system as an audio file. A model called "T5" (Happy-Text-To-Text) is utilized to analyze the grammar, yielding precise results. Accuracy is crucial in selecting a model, and T5's capabilities ensure a highly accurate grammar analysis. Once the analysis is complete, the system presents the identified grammar errors alongside their respective corrections. The output is designed to be user-friendly and comprehensive, providing a clear overview of the errors and their corrected versions.

In addition to the visual display of errors and corrections, the system goes a step further by converting the corrected text into an audio format. This audio output enhances the user experience, allowing individuals to hear the corrected version of their speech. By providing both visual and auditory feedback, the system caters to different learning styles and offers a more comprehensive understanding of grammar errors.

Overall, this approach combines advanced natural language processing techniques with the T5 model for accurate grammar analysis. It not only identifies errors in the speech but also provides a user-friendly interface displaying the errors and their corrections [21-25].

2.5.2 To evaluate grammatical complexity

To evaluate grammatical complexity, the system uses a sentence tokenizer to turn the text into sentences, and then it uses a proprietary algorithm to count the words in each phrase. This helps identify long sentences that may need to be split. Additionally, conjunctions are identified through dependency

ISSN (online): 2581-3048
Volume 7, Issue 10, pp 66-73, October-2023
https://doi.org/10.47001/IRJIET/2023.710009

phrasing, which aids in detecting potential grammatical errors and improving sentence structure.

Moreover, the system generates an audio output of the corrected speech, enhancing the overall usability and accessibility of the system.

III. RESULTS AND DISCUSSIONS

Here in our testing process, we did this in several rounds of testing. The details of their results are given below. They are correct information. The first step is to convert the words contained in the speech entered into the user system into a text. Here the words and punctuation marks are also converted into a text. This is a process that affects our entire system. Table II shows the results of converting our words into text. It is clear that the accuracy of identifying the correctness of a certain word was at a very high level. Usually, when converting the words of a speech into a text, things like punctuation marks are not translated correctly. There we got a low accuracy. Therefore, this was done by the mechanism called Google speech recognition. There we were able to get more accuracy.

Table III: Accuracies for converting speech into text

Conversion	Accuracy
Words	90%-95%
Punctuations	75%-85%

Analyzing the user speech introduction part and conclusion part of the speech under the content analysis component was able to reach an overall accuracy of 90%-95%. The same method is used to identify these two parts. The scores of both were equal. And we trained a model to recognize selected quotes and poems. The accuracy of correctly identifying these two parts was 91%.We used Random Forest Algorithm and Doc2Vec tool to do it. The results obtained here are shown in Table IV.

Table IV: Quotes/poems identifying model accuracies

	precision	recall	F1-score	support
Quotes	0.93	0.89	0.91	24409
Normal	0.89	0.93	0.91	23731
accuracy			0.91	48140
macro avg	0.91	0.91	0.91	48140
weighted avg	0.91	0.91	0.91	48140

Regarding checking the relativity of the content of the speech, we entered 3 videos and got the results. Comparing the words of the user entering the speech and the topic of the speech and the words related to the topic gives better accuracy. Here, a value of 75% - 85% was obtained as the safety value. The result is shown in the table below.

Table V: Content relativity actual results

Topic	Content	Level	Relativity
Global Warming	Content about Global warming	Content Fully Related to the topic	56.8%
Natural Disasters	Content about Natural Disasters	Content partially Related to the topic	36.9%

In analyzing the clarity of the speech, the pace of the speech should be taken into account. To do that we used a mechanism called WPM. Here we got about 95% accuracy. The pace of the speaker varies from speech to speech. Therefore, the average speed is taken into account here.

Under the flow completeness analysis component, all filler words, repeated words, and silent phase detection were done and it was possible to reach 85%-90% accuracy. Here, the ZERO short classification model was used to identify filler words. And the accuracy of displaying the filler words contained in the speech submitted by the user was about 90%.

A model called "T5" (Happy-Text-To-Text) provided more than 90% accuracy for checking grammatical errors. And the accuracy of displaying grammar mistakes contained in the speech submitted by the user was about 90% - 95%. Also, the accuracy of converting the audio file of the user's speech to a text and correcting the mistakes in the grammar system and converting it back to an audio file was about 85%.

The Fer2013 dataset is an open source data collection model that is part. Pierre-Luc Carrier and Aaron Courville continue the work. The document was originally created for internal use, but was later made available to the public in a Kaggle competition prior to ICML 2013. It comprises grayscale face images with a size of 48x48 pixels, each depicting various emotions. The dataset contains a total of 35,887 images, with 80% allocated for training the model and 20% for testing. The performance of the model, including the accuracy of each expression during evaluation, is presented in Table VI. Furthermore, the model achieved an accuracy ranging from 85% to 90% in sentiment analysis, specifically in identifying emotional states from the converted text.

Table VI: Emotions identifying model accuracies

	precision	recall	F1-score	support
Angry	0.71	0.72	0.71	495
Happy	0.94	0.90	0.92	899
Sad	0.73	0.66	0.69	608
Neutral	0.67	0.78	0.72	620
accuracy			0.78	2622
macro avg	0.76	0.76	0.76	2622
weighted avg	0.78	0.78	0.78	2622

ISSN (online): 2581-3048
Volume 7, Issue 10, pp 66-73, October-2023
https://doi.org/10.47001/IRJIET/2023.710009

IV. FUTURE WORK

The lack of comprehensive online tools for practicing public speaking skills has been the driving force behind this project. The small step taken through this initiative aims to address speech disfluencies from the perspective of speakers, with a larger goal of improving the current system based on factors such as pronunciation, grammatical error analysis of speech, etc. Personalized plans will be offered to users to enhance their public speaking skills, and the system will provide them with the opportunity to follow these plans.

Enhancing the user experience is a key aspect of future work, which involves the creation of a simulated artificial audience. This feature aims to provide speakers with a more realistic practice environment, allowing them to gauge and refine their performance effectively. However, it is essential to acknowledge that achieving high accuracy in this domain is a complex task, as it involves translating the intricacies of human speech into computer algorithms. Nonetheless, continuous efforts will be made to improve the accuracy of the system by incorporating new functionalities and leveraging advanced analysis techniques.

In essence, the future work of this project will focus on filling the gap in online resources for honing public speaking abilities. Advancements in the field will be pursued, and the project will strive to offer users an immersive experience that fosters skill development. The approach will be continually refined to provide users with a valuable tool for enhancing their public speaking skills.

V. CONCLUSION

In recent years, there has been a growing demand for an online tool that comprehensively analyzes public speaking skills in the English language. Addressing this need, this paper presents an innovative approach through the introduction of an online tool called "Smart Speaker." The primary objective of Speech Master is to assess and evaluate various aspects of English speech delivery, including content, grammar, linguistic efficiency, expressions, and overall speech flow. To accomplish this, distinct algorithms were developed utilizing advanced techniques such as Natural Language Processing (NLP), deep learning, and machine learning. Each algorithm focuses on specific components of speech delivery, enabling "Smart Speaker" to provide a thorough analysis. The system has attained a remarkable overall accuracy rate of over 80% by improving the effectiveness of these algorithms. This solution is distinctive since no other product offered a self-speech evaluation service that was fully operational and focused on users at this level before "Smart Speaker”. Here is the proof that a smart SPEAKER device improves and greatly helps any person's ability to create a successful speech. Its availability and user-friendly interface make it a valuable resource for diverse users aiming to improve their English public speaking abilities.

ACKNOWLEDGEMENT

Our sincere thanks to the supervisor and co-supervisor for their guidance, support, and motivation throughout our project. Their presence, knowledge and feedback have been instrumental in making this research a success. We are thankful to Sri Lanka Department of Information Technology and Sri Lanka Institute of Information Technology for giving us the opportunity to carry out this research project. Also, express our heartfelt thanks to the CDAP lecturers and staff for their cooperation, guidance and support throughout our research. We are grateful to our family, friends and colleagues for their constant support, encouragement and understanding throughout our educational journey. Without their support this research project would not have been possible. Finally, we'd like to thank everyone who participated in the study for giving up their time and sharing their insightful opinions, both of which were quite helpful. Thank you all for your invaluable contribution to the success of this research project.

REFERENCES

[1] O. Abdel-Hamid, A. Mohamed, H. Jiang; L. Deng, G. Penn, D. Yu, Convolutional Neural Networks for Speech Recognition, Vol. 22. ACM Transactions on Audio, Speech, and Language Processing: IEEE, 2014, 1533 – 1545.

[2] A. Deshpande, R. Pandharkar, S. Deolekar, Speech Coach: A framework to evaluate and improve speech delivery. International Conference on Computer, Communication and Signal Processing (ICCCSP): IEEE, 2020.

[3] Team Ummo, Ummo, Ummoapp.herokuapp.com Ummo | Speech Coaching App, Webrate.org, 2016.

[4] A. M. Cronin, P. Lam, LikeSo App, 2017.

[5] F. Chiarello, P. Belingheri, G. Fantoni, Data science for engineering design: State of the art and future directions. Vol.129. Model from NLTK. Journal of Information Science, 2021.

[6] A. A. Pise, M. A. Alqahtani, P. Verma, Purushothama K, D. A. Karras, Prathibha S, A. Halifacorresponding, Methods for Facial Expression Recognition with Applications in Challenging Situations, Vol.2022. Comput Intell Neurosci, 2022.

[7] M. Bargaje, Emotion recognition and emotion-based classification of audio using genetic algorithm - an optimized approach. International Conference on Industrial Instrumentation and Control: IEEE, 2015.

ISSN (online): 2581-3048
Volume 7, Issue 10, pp 66-73, October-2023
https://doi.org/10.47001/IRJIET/2023.710009

[8] D. V. Sang, N. V. Dat, D. P. Thua, Facial expression recognition using deep convolutional neural networks. 9th International Conference on Knowledge and Systems Engineering (KSE), 2019.
[9] H. H. Al-Baity, H. J. Alshahrani, M. K. Nour, A. Yafoz, O. Alghushairy, R. Alsini, M. Othman, Computational Linguistics Based Emotion Detection and Classification Model on Social Networking Data, Vol.12. 2022.
[10] B. Tucker, M. Lehew, Exploring Public Speaking: The Free Dalton State College Public Speaking Textbook: Chapter 8: Introductions and Conclusions, 4th Edition. University System of Georgia, 2019.
[11] Public Speaking: Practice And Ethics: The Importance of an Introduction. 2012.
[12] Mueller, M. Orcholski, J. Miller, K. Woodward, M. L. Mcdonough, and L. Roth, The Basis For Strong Public Speaking: Effective Introductions and Powerful Conclusions.
[13] “O/L English Paper Marking Schema”. 2020.
[14] V. Pashek, R. H. Brookshire, Effects of Rate of Speech and Linguistic Stress on Auditory Paragraph Comprehension of Aphasic Individuals, Vol.25. Journal of Speech and Hearing Research, 1982, 377-383.
[15] Stouten, J. P. Martens, A feature-based filled pause detection system for Dutch. Workshop on Automatic Speech Recognition and Understanding: IEEE, 2003.
[16] Afroz, S. G. Koolagudi, Recognition and Classification of Pauses in Stuttered Speech Using Acoustic Features. International Conference on Signal Processing and Integrated Network: IEEE, 2019.
[17] A. Chatziagapi, D. Sgouropoulos, C. Karouzos, T. Melistas, T. Giannakopoulos, A. Katsamanis, S. Narayanan, Audio and ASR-based Filled Pause Detection. International Conference on Affective Computing and Intelligent Interaction: IEEE, 2022.
[18] M. Goto, K. Itou, S. Hayamizu, A Real-Time Filled Pause Detection System For Spontaneous Speech Recognition. 1999.
[19] M. Belz, J. Trouvain, Are ‘Silent’ Pauses Always Silent, Physics. 2019.
[20] S. P. Singh, A. Kumar, L. Singh, M. Bhargava, K. Goyal, B. Sharma, Frequency-based spell checking and rule-based grammar checking. International Conference on Electrical, Electronics, and Optimization Techniques: IEEE, 2016
[21] Dizon, Gilbert, Gayed, John M, Examining the Impact of Grammarly on the Quality of Mobile L2 Writing, Vol.17. JALT CALL Journal, 2021, 74 – 92.
[22] Murakami, N. C. Ellis, Effects of Availability, Contingency, and Formulaicity on the Accuracy of English Grammatical Morphemes in Second Language Writing, Vol.72. Empirical Study, 2022, 899-940.
[23] S. Shaumiwaty, I. Munandar, The Analysis of Grammatical and Lexical Errors on The Students' Theses of Iain Takengon, Vol.6. Journal As-Salam, 2022, 269–285.
[24] R. A. Andania, The Grammatical Accuracy, Cohesion, and Coherence of Thai Students’ English Writing at Darawithaya School, Narathiwat – Thailand, Vol.6. 2017, 10-25.
[25] C. Mellish and G. Ritchie, “The Grammatical Analysis of Sentences”, University of Aberdeen.

AUTHORS BIOGRAPHY

Malshan E.G.C,
Department of Information Technology, Sri Lanka Institute of Information Technology, New Kandy Road, Malabe, Sri Lanka.

Buddhini B.A.D,
Department of Information Technology, Sri Lanka Institute of Information Technology, New Kandy Road, Malabe, Sri Lanka.

Isurandi I. G,
Department of Information Technology, Sri Lanka Institute of Information Technology, New Kandy Road, Malabe, Sri Lanka.

Habalakkawa W. V. K. I,
Department of Information Technology, Sri Lanka Institute of Information Technology, New Kandy Road, Malabe, Sri Lanka.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Issue 10, pp 66-73, October-2023

https://doi.org/10.47001/IRJIET/2023.710009

Citation of this Article:

Malshan E.G.C, Buddhini B.A.D, Isurandi I.G, Habalakkawa W.V.K.I, Suranjini Silva, Thamali Kelegama, "Smart Speaker: Enhancing Any Persons' Ability to Deliver English Speeches Independently With a Web Application" Published in *International Research Journal of Innovations in Engineering and Technology - IRJIET*, Volume 7, Issue 10, pp 66-73, October 2023. Article DOI https://doi.org/10.47001/IRJIET/2023.710009

****