VSCODE- Code With Voice Using Natural Language Processing(NLP)

M. Gnana Prakash[1]
Department of Information Technology
Sri Sai Ram Institute of Technology
Chennai, Tamil Nadu, India
prakash.it@sairamit.edu.in

A.Ponmalar[2]
Department of Information Technology,
Sri Sai Ram Institute of Technology
Chennai, Tamil Nadu, India
ponmalar.ragupathi@gmail.com

S.Deeba[3]
Department of Computer Science and Engineering
Sri Manakula Vinayagar Engineering College
Puducherry, Tamil Nadu, India
deebaramshankar@gmail.com

Dr.A.Akilandeswari[4]
Associate Professor
Saveetha School Of Engineering SIMATS,
Chennai, Tamil Nadu,India
akilandeswaria.sse@saveetha.com

Saira Banu Mohammed Rasool[5]
Department Of Information System,
Applied College, Najran University,
rafisaira@gmail.com

P.Lavanya[6]
Department of Information Technology,
Sri Sai Ram Institute of Technology,
Chennai, Tamil Nadu, India,
sit21it067@sairamtap.edu.in

2022 International Conference on Computer, Power and Communications (ICCPC) | 979-8-3503-9784-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICCPC55978.2022.10072106

***Abstract*—Physically Challenged people who are interested in coding, in the current scenario cannot code and exhibit their knowledge in programming skills. Our project enables them to code without using a keyboard. They can code entirely by their voice and execute their code. We are converting the user's speech to text by Natural Language Processing (NLP). With the use of text obtained from NLP model we generate code snippets using our custom code generator API. The code snippets are generated in the Visual Studio Code text editor using our own VSCode extension.**

***Keywords*—Natural Language Processing, Speech Recognition, Visual Studio Code (VSCode), Code Generator.**

I. Introduction

Physically challenged people have a lot issues which they might not be able to expose it outside. They generally suffer a lot when compared to a normal people who all are physically fit. Not only for their living but these physically challenged people too suffer a lot in completing their day-to-day activities. So, we have come up in an idea to exhibit there programming skills and interests. A physically challenged person not only includes those individuals who are physically challenged at the time of birth. There are some cases where in a particular situation an individual might meet with an accident that might make him physically challenged. In such cases also this application will be completely useful for them. People in every profession have one thing in common: they make use of the technology that can save them time and money. Speech Recognition is utilized across many industries in a number of ways, but regardless of the use case it consistently helps people by saving their whole lot of time. By using Speech Recognition and NLP users can code using their voice.

II. Problem Description

In today's current scenario, physically challenged people cannot be able to code even if they have coding knowledge, as a result even if they are educated, they will be not able to get recruited in any IT-based industries. Now we are about to provide software that would help the physically challenged to code and exhibit their knowledge in coding. The physically challenged people need to be in such a way that they cannot be deaf or dumb. Those people can be able to access the VScode. The user who is physically challenged now will access the application using his voice. The voice is taken as input and converted to text and then it will send to the text NLP model. The NLP model is used to extract required keywords using stemming, lemmatization, tokenization, and then sends the keywords to our Code Generator. Using the keywords, the Code Generator will generate required code snippets in the VSCode text editor.

III. Background

A. Speech Recognition

Speech perception or speech-text reporter is the methodology to capture the voice or speech of the user and convert it to text by a machine or a program. In the Speech Recognition system, there are two types speaker-independent and speaker-dependent systems. Speech Recognition systems that use training are called speaker-dependent else called speaker-independent. Speaker-dependent Speech Recognition System uses different user

voice audios as datasets to train the model. Speech recognition applications include voice user interfaces such as voice dialing, call routing, domestic appliance control, search keywords, simple data entry, preparation of structured documents, determining speaker characteristics, and speech-to-text processing.

B. Natural Language Processing (NLP)

Natural language processing (NLP) is a branch of artificial intelligence that helps understand human language by machine. As a human, we speak and write in English, Spanish or Chinese. But a computer knows only zero's and one's it is also called machine code or machine language. Using NLP, Machine or Program can understand common English or languages across the world. Natural language processing includes many different techniques and algorithms for interpreting human language.

- The focus of machine learning turns to input used during the run-time with the use of the Machine learning and it is completely used to convert into a formalized input that the computer could able to understand. Obvious where the effort should be directed.
- Automatically the natural language processing will be able to analyze the text and then by using various process it can be able to convert the given input speech as the as convertible file and then it will be to process and then it can be able to convert the given speech into a valuable input..

$$\mathbf{RMM(token_N) = (PMM)(token_N) \times 1/2d \left(\sum_{i=-d}(PMM(token_{N-1})) \times PF(token_N \times token_{N-1})\right)}$$

Where,

RMM, is the Relative Measure of Meaning token. **N**, is the number of tokens being analyzed. **PMM**, is the Probable Measure of Meaning, is the location of the token along the sequence of **N-1** tokens. **PF**, is the Probability Function specific to a language.

Wavenet is generally introduced for some wave form generation. These wave forms will produce high quality conversion of text to speech content in a highly assurable manner. But in order for this wave form to work properly we need to train the waveform with a lot of huge amount of training data which could be time consuming and it also be a very tedious process as it would change the responses given by the system when it was trained by a set of similar data. Similar data training will not affect the wave form in anyway but if it were trained by a different set of input the accuracy of the wave form could reduce. The chances of reduced accuracy will occur in terms of training it with a different data.

C. Flask

Flask is a light weighted web development framework developed in Python. It is also known as a micro-framework because it is light weighted and only provides components that are essential. Flask provides tools, libraries, and technologies and it also provides simplicity, flexibility, and fine-grained control that allow you to build a web application. Flask has a lightweight and modular design, so it is easy to transform it into the web framework developers require a few extensions and without weighing it down.

D. Visual Studio Code (VSCode)

Visual Studio Code (VSCode) is a lightweight code editor which enables cross-platform support. It supports various programming languages and script languages like JavaScript, Typescript, Node.js, HTML, CSS, python, java and has a robust architecture.

IV. Proposed System

In our proposed system users can open visual studio using voice. Using Speech Recognition, the voice input from the user will be converted to text in live transcription. The text which was converted using speech to text is sent to our API as a request.

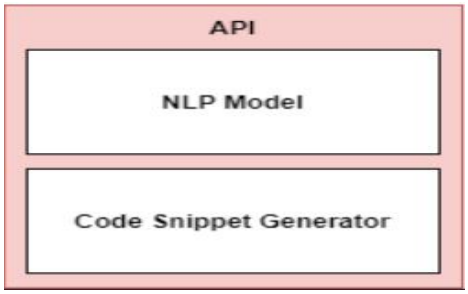

Fig 1: This is how API linked with our project

Our API consists of Natural Language Processing (NLP) Model and Code Snippet Generator. The NLP model will process the text using stemming, lemmatization, tokenization and generate keywords. These keywords are sent to code generator. The code generator is linked to our syntax library which contains syntax of different programming languages. Using this code generator and keywords, the respective code snippets are generated and send as response to our visual studio code extension. This extension handles VScode editor events such as files actions, workspace actions and terminal actions. Using this extension, the code snippet from the API is written in visual studio code text editor. The execution part of the program is done is voice as well. API is generally used to act as a bridge for any two applications for the purpose of communication. They simply use the interface with their products and services. Many popular web apps would not be possible without APIs. APIs varies with various software applications. Public, Partner, REST, SOAP, and private are the types of APIs. Advantages of using APIs, using the API of third-party providers (like Google) for

logging and sharing data in some other web application is a lot safer. APIs enable and facilitate the automation process with great ease. We can use the API of some web service in our own code to automate tasks. Advantages of using APIs, using the API of third-party providers (like Google) for logging and sharing data in some other web application is a lot safer. APIs enable and facilitate the automation process with great ease. We can use the API of some web service in our own code to automate tasks

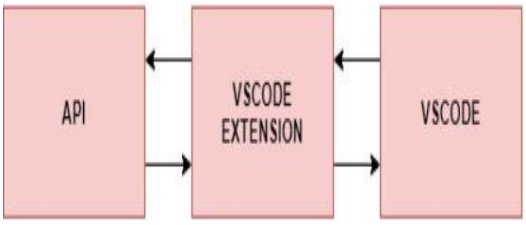

Fig 2: Connection of API with our Project

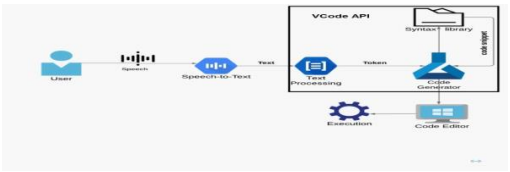

Fig 3: System Architecture

Fig 4: Generating Text and Recognising Speech

V. Literature Survey

In the current scenario we have been subjected to many physically challenged people who are not able to sustain their livelihood as good as the normal individuals[1]. A lot of researchers have gone through their research in making some projects that might be able to solve their issue and make them feel like any other normal individual[2]. NLP and Tokenization have been making people look what exactly this can do and how it can be able to improve their project in a successful and in a most fascinating way that could make the entire world make use of these technologies[3]. This paper deals with the speech to text and text reporter which has been enabled for various technologies. Summarization based on speech unit extraction and concatenation[4]. Text that provides a certain limit of likeliness in the amount of information being used. For the various filler units that have been tested and investigated for the usage and in the extraction of the original speech[5]. Wavenet was introduced for generating high quality waveforms that could be able to produce high quality speech into text conversion[6]. These wave forms will produce high quality conversion of text to speech content in a highly assurable manner[7]. But in order for this wave form to work properly we need to train the waveform with a lot of huge amount of training data which could be time consuming and it also be a very tedious process as it would change the responses given by the system when it was trained by a set of similar data. Similar data training will not affect the wave form in anyway but if it were trained by a different set of input the accuracy of the wave form could reduce[8,9]. The chances of reduced accuracy will occur in terms of training it with a different data. In the current projects that are currently being used it would Also in the current existing system the people who are physically handicapped and if they want to access the existing systems available then it would be difficult for them because it prompts them to click on a microphone icon that would enable them to give voice as an input so that the system would be able to process the input accordingly[10,11]. Also in the currently being used projects it would not be able to detect the voices in background that would create sounds or if the user is present in a noisy background then the algorithm will not be able to detect and continue processing the voice with a correct accuracy as it would lead to some malfunctions with the speech to text conversions because it would not be able to go up the speech that is being given by the user, The Bidirectional kalman filter was also used as an effective method for converting the voice to text but that filter could not be able to convert the given voice input into text as it would not work properly if the user's voice decibels is not upto 5 DB[12]. This decibels are measured by suing some popular sound measuring tools like SPL, LEQ and PEAK. This bidirectional kalman filter will not recognise the user's voice input if it is less than 5 Db also it would take some long time for the purpose of converting voice to any desirable content. The HMM and ANN algorithm could also be used to convert the speech into text but major problem in this HMM and ANN is that it could not work properly if the user is in a noisy environment it could detect the voice properly[13]. The MFCC algorithm is a special algorithm that is being used in our sensory auditory system .This algorithm is one of the famous algorithms that could convert the voice to text but the major drawback of this algorithm is that it would create some discontinuities. In the current scenario we have been subjected to many physically challenged people who are not able to sustain their livelihood as good as the normal individuals. A lot of researchers have gone through their research in making some projects that might be able to solve their issue and make them feel like any other normal individual. NLP and Tokenization have been making people look what exactly this can do and how it can be able to improve their project in a successful and in a most fascinating way that could make the entire world make use of these technologies.

VI. Conclusion

The VScode which we have designed will be a useful to the physically impaired people and it will surely help them in exploring their knowledge in coding skills. It will also help them to get recruited in various companies, freelancing, teaching, etc. which would also enable them to support their lives financially. The various features we

provide makes our project user friendly. In future, we will be adding many more programming languages. One more feature that we are going to add in future work is that we will be making the project work on different platforms

VII. Result

Our project has produced results as expected. The project was able to create a code using the voice input given by the user. The ability of the system to recognise the voice input and generate the code was as expected.

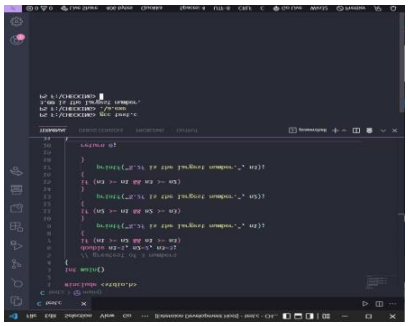

Fig 5: Output Screenshot

References

1] Yuki, S., Shinnosuke, T. (2018). Statistical Parametric Speech Synthesis Incorporating Generative Adversarial Networks,IEEE/ACM Transactions on Audio, Speech, and Language Processing, 26 (1).

[2] Laszlo, T. (2018). Deep Neural Networks with Linearly Augmented Rectifier Layers for Speech Recognition, SAMI 2018 IEEE 16th World Symposium on Applied Machine Intelligence and Informatics February 7-10 Košice, Herl'any, Slovakia.

[3] Li, Lantian; Wang, Dong; Chen, Yixiang; Shi, Ying; Tang, Zhiyuan; Zheng, Thomas Fang (2018). [IEEE ICASSP 2018 - 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) - Calgary, AB, Canada (2018.4.15-2018.4.20)] 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) - Deep Factorization for Speech Signal.5094–5098.doi:10.1109/ICASSP.2018.8462169.

[4] Chowdhury, A., Ross, A., & David, P. (2021). DEEPTALK:Vocal Style Encoding for Speaker Recognition and Speech Synthesis. ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). doi:10.1109/icassp39728.2021.9414.

[5] Bano, S., Jithendra, P., Niharika, G. L., &Sikhi, Y. (2020). Speech to Text Translation enabling Multilingualism. 2020 IEEE International Conference for Innovation in Technology (INOCON). doi:10.1109/inocon50539.2020.9298280.

[6] Lv, F., Zhang, H., Lou, J., Wang, S., Zhang, D., & Zhao, J. (2015). CodeHow: Effective Code Search Based on API Understanding and Extended Boolean Model (E). 2015 30th IEEE/ACM International Conference on Automated Software Engineering (ASE). doi:10.1109/ase.2015.42.

[7] S. Bickel, P. Haider and T. Scheffer, "Predicting sentences using n-gram language models", Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing, pp. 193-200, 2005.

[8] K. Nakasho, "Development of a flexible Mizar tokenizer and parser for information retrieval system", 2019 Federated Conference on Computer Science and Information Systems (FedCSIS), pp. 77-80, 2019.

[9] I. Sasano and K. Choi, "A text-based syntax completion method using lr parsing", Proceedings of the 2021 ACM SIGPLAN Workshop on Partial Evaluation and Program Manipulation, pp. 32-43, 2021.

[10] A. Ponmalar, K. Rajkumar, U. Hariharan, V. K. G. Kalaiselvi and S. Deeba, "Analysis of Spam Detection Using Integration of Logistic Regression and PSO Algorithm," 2021 4th International Conference on Computing and Communications Technologies (ICCCT), 2021, pp. 396-402, doi: 10.1109/ICCCT53315.2021.9711903.

[11] A.Ponmalar, M.Poovarasan, T.Vasumathi, G.M.Keerthana, A.Yogesh, Dr V.Dhanakoti,"EFFICIENT REGISTRATION OF LAND USING BLOCK CHAIN TECHNOLOGY",International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-2S4, July 2019

[12] Hariharan, U., Rajkumar, K., Akilan, T. et al. A multi-hop protocol using advanced multi-hop Dijkstras algorithm and tree based remote vector for wireless sensor network. J Ambient Intell Human Comput (2021). https://doi.org/10.1007/s12652-021-03548-4.

[13] P. A, N. G, V. K. R, P. P and S. R. R.V.T, "Stroke Prediction System Using Artificial Neural Network," 2021 6th International Conference on Communication and Electronics Systems (ICCES), 2021, pp. 1898-1902, doi: 10.1109/ICCES51350.2021.9489055.