Smart Stammering speech converter using AI and neural network

M. Vadivelan1, B, Vijayalakshmi2, R. Marie Jenifer3,P. Aruni4

1Assistant Professor, Department of Biomedical Engineering, Sri Manakula Vinayagar Engineering College, Madagadipet, Puducherry, India

2,3,4 UG Scholar, Department of Biomedical Engineering, Sri Manakula Vinayagar Engineering College, Madagadipet, Puducherry, India

*Abstract:*  Stammering, a speech disorder characterized by disruptions in the normal flow of speech, can significantly impact communication and social interactions. The limitation of the current situation in the simulation for detection of Stammering in pre-recorded audio paved the way for the proposed system with a new approach to address Stammering through hardware by using live audio. It is a compact hardware device for real-time stammering that uses a microphone and speaker to receive and deliver the input and output using an AI processor. It uses a spectral centroid to extract features from speech, which are then used to train the neural network classifier. The trained data is compared with the given input. The MLP-NN classifier will be used to train the given database, which consists of both stammering and non-stammering data, and convert the stammering audio into non-stammering audio.

*Keywords* -Artificial intelligence, MLP-NN, stammering, Neural Networks.

1. **Introduction**

Stammering is a speech disorder that affects the flow of speech. It is identified by any kind of disturbance in the normal flow of speech, like repetitions of words, sounds, syllables, prolongations of sounds, or any blocks in the person. Those disturbances can make speech sound hesitant or choppy, and they can vary in condition and frequency. It tends to start during early childhood, up to 2-5-year-old kids, when they are learning how to speak. As a result, most of them outgrow it and re-acquire fluency in a moderate proportion. The etiology of stammering is believed to be a combination of hereditary, neurological, and environmental factors. According to some research, the stammering can run in their families, which would suggest a genetic cause. It is also believed that stammering is caused by a difference in the brain's processing of speech and language.

Stammering can have significant changes in a person's personal life, affecting self-confidence, mental health, and opportunities to show their true identity in front of others. With appropriate treatment and support, many people who stammer can improve their fluency and lead fulfilling lives. Some treatments for stammering may not get rid of all stammering, but they can teach skills that help people who are affected. They include speech therapy, cognitive behavioral therapy (psychotherapy), and some electronic devices.

There are several methods to classify stammering and non-stammering speech and to recognize speech using traditional methods. The traditional methods are generally used to recognize speech fluency by extracting the features of the speech and classifying them. The speech signals are classified using feature extraction techniques, such as Signal Centroid, signal bandwidth, Zero Crossing Rate, Spectral Centroid, Linear Predictive Coding (LPC), Linear Prediction Cepstral Coefficients (LPCC) and Mel Frequency Cepstral Coefficients (MFCC), Gaussian Mixture Model (GMM).

The main objectives of this work are as follows:

* + To design a device for conversion of stammering speech into normal speech by using AI processor.
  + To sense, capture and analyze a stammered audio signal and deliver them into normal audio by speaker.
  + Using MLP-NN, classify whether the person stammers or not.

# This proposed work is organized as follows. Section 2 describes the work related to the proposed system. Section 3 describes the collection of the speech database, Section 4 describes suggested methodologies that are used to classify the stammering and non-stammering speech and convert them into normal speech, and Section 5 describes the results of the system.

1. **Related Works**

Priya, K., et al. [1] In this paper, the authors applied speech detection of stammer from speech signals to find disfluency of the speech using feature extraction such as Mel Frequency Cepstral Coefficients (MFCC), Pitch, Delta Detla MFCC (D2MFCC), Spectral Flux, and Spectral Centroid. This feature extraction is used to train the Multilayer Perceptron Neural Network (MLP-NN) using a BR. The main aim of the studies on the detection of stammer speech was to develop a better, faster, and more accurate with a level of 99.2% result.

Jouaiti, M., & Dautenhahn, K. [2] In this paper, the authors detect stammering speech by combining MFCC and phoneme probabilities to train a neural network for stuttering detection. The proposed work is also used for the classification of disfluency into four categories. The databases are collected from the UCLASS and FluencyBank. This system is best for the real-time application.

Rajput, Shaswat, et al. [3] The aim of this paper was to detect and remove stuttering speech by using deep neural networks. They collected datasets from the UCLASS archive in .wav format and used machine learning algorithms for speech recognition. They used many types of algorithms to increase the accuracy of the system. The algorithm is used to test on different types of datasets varying from low to heavy stammering, so there is a reduction in Word Error Rate.

Prabhu, R., et al. [4] This paper aims to develop the Matlab algorithm for speech recognition of stammered voices by removing the repetition of words. This system consists of an algorithm that is implemented in five stages such as filtering, silence ejection, speech-to-text conversion, removal of repetition words, and text-to-speech conversion. It aims to help the person who suffers from stammer with an accuracy of 86%.

Dash, Ankit, et al. [5] In this paper, the authors aim to develop an algorithm to improve speech recognition of stuttered speech. This paper approaches a major issue to correct and detect the stammer through the removal of repetition words using the Text-to-speech system. So, it produces better detection of stuttering voices within acceptable time limits.

1. **Database Collection**

The proposed system uses input data that was collected from the University College London’s Archive of Stuttered Speech (UCLASS) and Common Voice datasets to identify speech stuttering (CV). They were used as pre-trained models for the conversion of stammering speech into non-stammering speech. This dataset has two main releases. They were recorded with various levels of stammering, like sentence repeating, pauses, and words. It consists of three classes: monologue, reading, and conversations. Some of the recordings contain transcriptions like orthographic, phonetic, and standard ones. This proposed model uses monologue files only. The Monologue of release-1 folder contains 120 male and 18 female recordings. The Release 1 recordings, all monologs, were from speakers with a wide range of ages (5 years to 47 years). The Monologue of release-2 folder contains 76 male and 6 female speech recordings.

|  |  |  |
| --- | --- | --- |
| **Category** | **Female** | **Male** |
| Release 1 (Monolog) | 120 | 18 |
| Release 2 (Monolog) | 76 | 6 |

Table 1. Speech Database

# **Proposed Methodology**

The speech signal is affected by some characteristics like age, gender, flow of words, and genetics. The proposed system detects the speaker's stammering speech, classifies them into stammering and non-stammering speech, and finally converts them into normal speech with the help of an AI processor and neural network algorithm. The signal is pre-processed like filtering, amplification, and noise removal in the first stage, and the feature extraction will happen in the second stage. Finally, in the third phase, they convert them into normal speech using the MLP-NN algorithm.

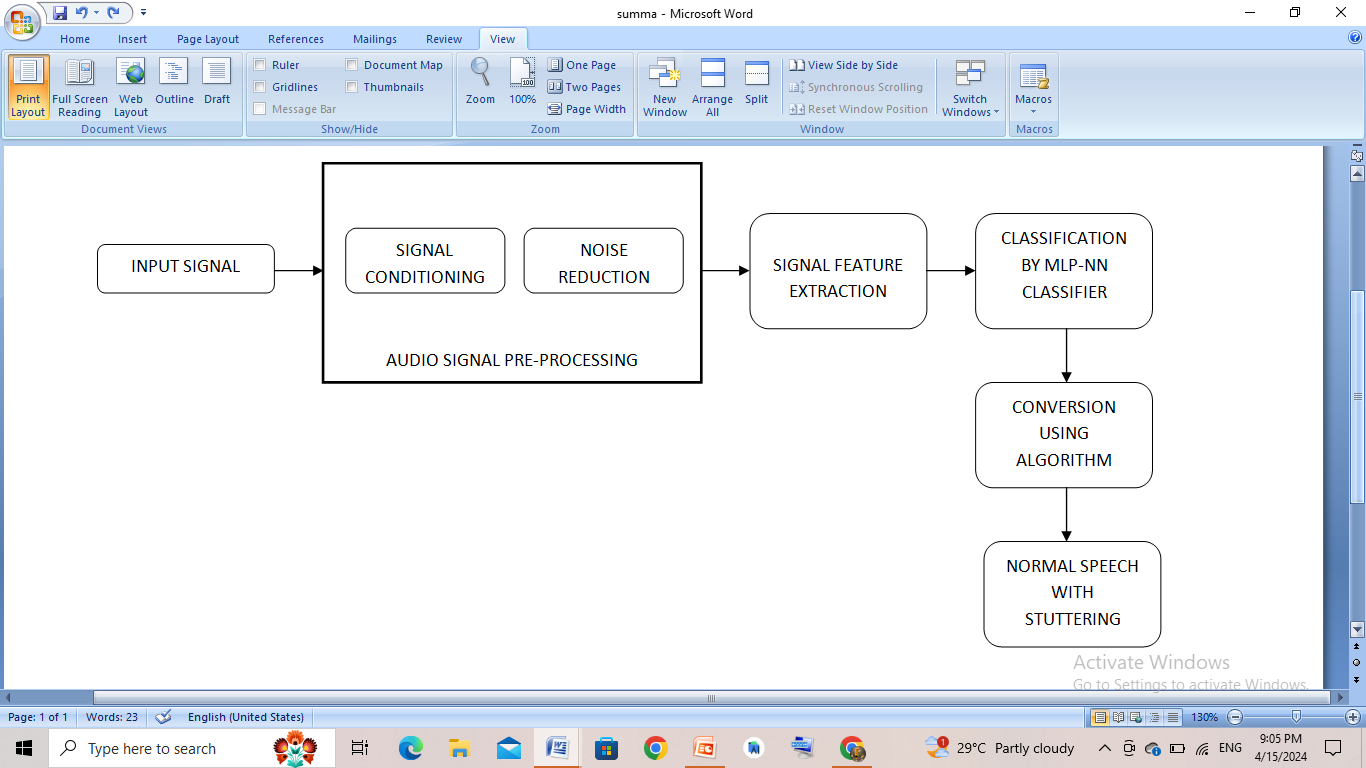


Figure 1. Work flow of the System

This introduces a new approach to addressing the Stammering through hardware. It is a compact hardware device for real-time stammering that uses a microphone and a speaker to receive and deliver the input and output, respectively. This technique has been found to have superior accuracy and training compared to other methods, such as NN training and classifiers. The MLP-NN classifier will be used to train the already-stored database, which consists of both stammering and non-stammering people’s data.

The input is received through the microphone from the user, and it feeds into the processing steps like signal conditioning, noise removal, filtering, and amplification. The processed signals are converted from audio to text using an AI processor on the transmitter side. The embedded C++ program is dumped into the AI processor, which eliminates the repeated words and provides the corrected word for the next steps. The output text will be converted into audio by the AI processor on the receiver side. Finally, the audio will be amplified and delivered through the speaker.

4.1 **Block Diagram**

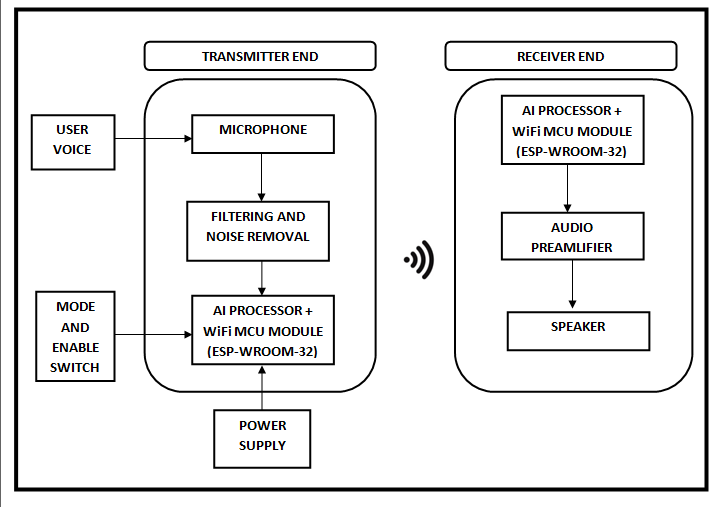


Figure 2. Block Diagram of the proposed work

There are two main parts of the system:

* Transmitter module
* Receiver module

**4.1.1 Transmitter Module**

At the transmitter end, the hardware includes a microphone, filters, switches, a signal conditioner, and an AI processor for getting the stammered voice from the user. There will be an ESP32-WROOM-32, a WiFi + Bluetooth + Bluetooth Low Energy MCU module that is used to transfer the data from the transmitter end to the receiver end, and an AI processor that converts audio into text directly. First, we connect the microphone with the signal conditioner, filter, and switch to get input from the user and process it into pre-processing steps like noise removal, filtering, and amplification. The processed data will be converted into non-stammered text by eliminating the repeated words in the given input using embedded C++, which is dumped into the AI processor. Finally, the converted data is transferred to the receiver end through the cloud. LCD (16x2 LCD) is fixed at the end of the receiver end for displaying the modes (Record and Send) of the process.

**4.1.2 Receiver module**

On the receiver end, the AI processor is connected to the WiFi ESP32-WROOM-32, which gets converted data from the transmitter end using cloud storage. The AI processor is connected virtually to the computer through the hotspot. The AI processor is connected to the preamplifier to increase the amplitude of the audio signal, which is enough to reach the output part.

Then the amplified signal is connected to the speaker, which will give the converted audio to the user.

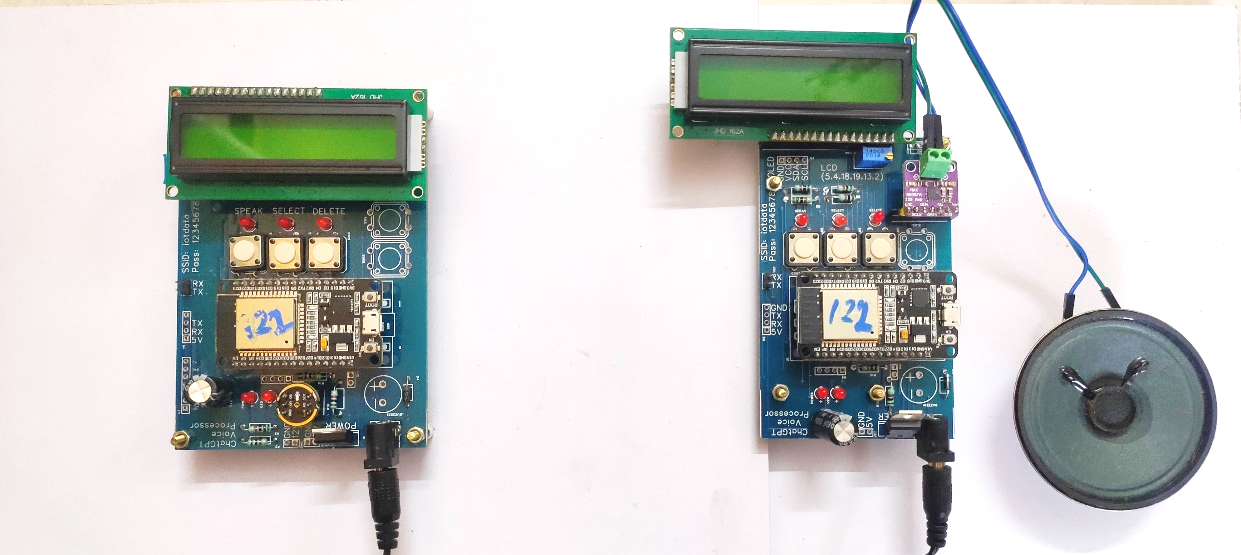


Figure 3. Hardware Setup of the system

In the software module, there are two sets of programming code and an INO file that will be run using Arduino IDE 2.2.1 software. Embedded C++ will be used in the coding part. The embedded C++ is used as an interface for the ESP-WROOM-32 software module.

There will be two sets of C++ programs:

* The first file has a set of codes for the transmission of data from one WiFi module to another.
* The second file is used for the audio conversion.

The code for the I2S interface can be viewed in the Arduino IDE software, which is illustrated in Figure 4. This code will be dumped into the AI processor for input and a Wi-Fi connection.

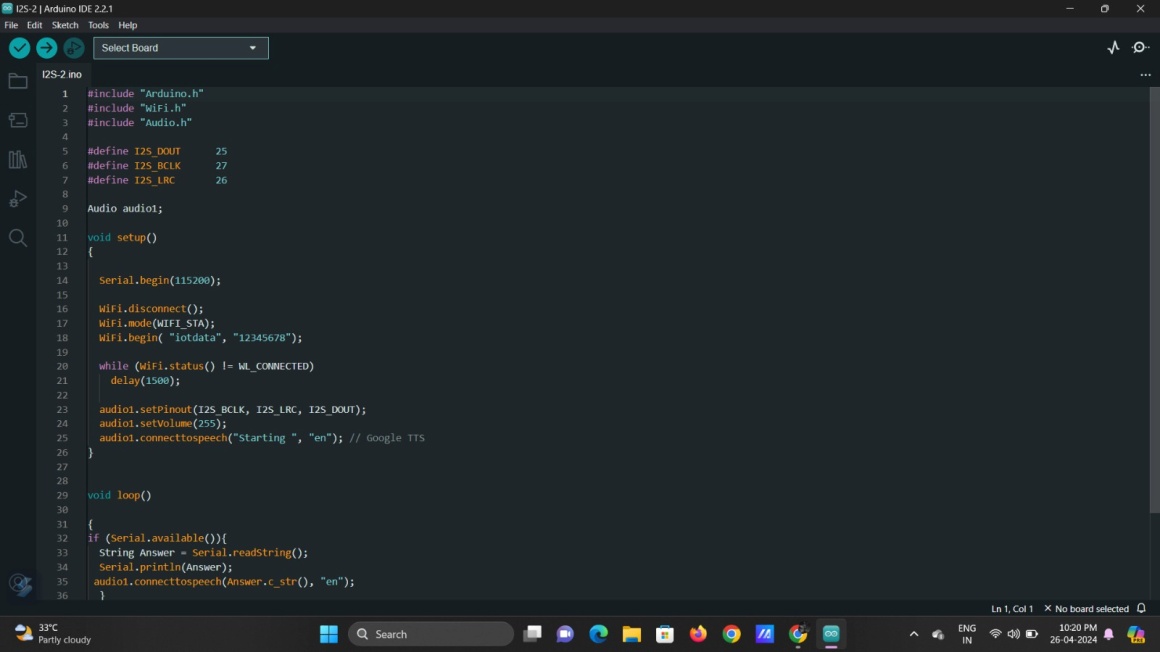


Figure 4. INO file running in Arduino IDE software

1. **Result and Discussions**

This section considered the results of the stammering converter using AI and neural network. The proposed method was developed with the AI processor and ESP32-WROOM-32 module. The speech signal gets pre-processed for conversion, converted into non-stammered audio, and result will be in the form of audio. The final output of this system will be in the form of audio. The input voice is processed at the transmitter end, and the processed audio will be delivered through the speaker.

The audio acquired by the user is converted into text and displayed on the LCD is illustrated ing Figure 5 and Figure 6. The data is also transmitted over the internet using a WiFi module.

****

Figure 5. Converted audio text in Transmitter module

****

Figure 6. Received audio text in Receiver module

The user can give input to the system; it will compare with the trained data in the database, and the recognized speech is converted into non-stammered text. Finally, the text will be converted into audio and delivered as the final output.

1. **Conclusion**

The field of artificial intelligence is a fast-emerging field with the capability of doing tasks in a fast and better way. By using AI, there will be more solutions for many problems, especially in real-time applications. In this proposed work, we have explored the versatility of stammering speech detection that is used to approach a new device for the conversion of stammered audio into non-stammered audio by using artificial intelligence. The main advantage of the proposed system is that we use live audio instead of pre-recorded audio. We have achieved a better solution, real-time assistance, and support for individuals who are suffering from the stutter with a higher level of accuracy.

1. **Acknowlegment**

Our project would likely to be taking this opportunity to express our deepest gratitude to Sri Manakula Vinayagar Engineering College (SMVEC), and Dr.A.Vijayalakshmi, Professor and Head of the Department, Biomedical Engineering, for giving us valuable suggestions. They have always been a source of inspiration and encouragement towards this project.

# **References**

1. Priya, K., M. Mohamed Mansoor Roomi, M. Senthilarasi, S. Karthika Shree, and M. Anusha, “Speech Stammer Detection by Spectral Features Based Artificial Neural Network.” 2022 International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON). IEEE, 2022.
2. Jouaiti, Melanie, and Kerstin Dautenhahn. “Dysfluency classification in stuttered speech using deep learning for real-time applications.” ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)., 2022.
3. Rajput, Shaswat, Ruban Nersisson, Alex Noel Joseph Raj, A. Mary Mekala, Olga Frolova, and Elena Lyakso. "Speech Stuttering Detection and Removal Using Deep Neural Networks." In Proceedings of the 11th International Conference on Computer Engineering and Networks, pp. 443-451. Singapore: Springer Nature Singapore, 2021.
4. Prabhu, R., P. Praveen Kumar, J. Kavin Kumar, and E. L. Dhivyapriya. "Speech Based Anti Stuttering Algorithm using Matlab." International Journal of Engineering and Advanced Technology (IJEAT) 9, no. 3 (2020).
5. Dash, A., Subramani, N., Manjunath, T., Yaragarala, V. and Tripathi, S."Speech recognition and correction of a stuttered speech." 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI). IEEE, 2018.
6. Hosseini, Rahilsadat, Bridget Walsh, Fenghua Tian, and Shouyi Wang "An fNIRS-based feature learning and classification framework to distinguish hemodynamic patterns in children who stutter." IEEE Transactions on Neural Systems and Rehabilitation Engineering 26.6, 2018
7. Waghmare, Swapnil D., Ratnadeep R. Deshmukh, Pukhraj P. Shrishrimal, Vishal B. Waghmare, Ganesh B. Janvale, and Babasaheb Sonawane. "A comparative study of recognition technique used for development of automatic stuttered speech dysfluency Recognition system." Indian Journal of Science and Technology 10, no. 21 (2017): 1-14.
8. Alharbi, Sadeen, Madina Hasan, Anthony JH Simons, Shelagh Brumfitt, and Phil Green. "Detecting stuttering events in transcripts of children’s speech." In Statistical Language and Speech Processing: 5th International Conference, SLSP 2017, Le Mans, France, October 23–25, 2017, Proceedings 5, pp. 217-228. Springer International Publishing, 2017.
9. Świetlicka, Izabela, Wiesława Kuniszyk-Jóźkowiak, and Elżbieta Smołka. "Hierarchical ANN system for stuttering identification." Computer Speech & Language 27.1, 2013.
10. Mahesha, P., and D. S. Vinod. "Classification of speech dysfluencies using speech parameterization techniques and multiclass SVM." In International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness, pp. 298-308. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013.
11. Km, Ravi Kumar, and S. Ganesan. "Comparison of multidimensional MFCC feature vectors for objective assessment of stuttered disfluencies." Int. J. Adv. Netw. Appl 2, no. 05 (2011): 854-860.
12. Howell, Peter, Stephen Davis, and Jon Bartrip. "The university college london archive of stuttered speech (uclass)." (2009).
13. Hayhow, Rosemarie, Anne Marie Cray, and Pam Enderby. "Stammering and therapy views of people who stammer." Journal of Fluency disorders 27, no. 1 (2002): 1-17.