Intelligence System for Disfluency

Raviraj Bhandari[1], Reena Pagare[2]

[1,2] MIT School of Engineering, MIT ADT University,
MAEER's Rajbaug Campus, Loni-Kalbhor, Pune-412201

Abstract: **Dysfluency is a term used to describe a person's inability to create fluent, smooth speech or typical speech abnormalities. People who start out with typical developmental disfluencies may progress to stuttering as a result of their speech stuttering. To that end, we give a review of contemporary machine learning and deep learning algorithms for using EEG to detect and characterize stuttering in speech.EEG and speech extraction devices like microphones are used. The extracted data from the EEG and microphone will be used in the word and time domains. A separate CSV file will be generated for our dataset, which will be used for classification models.This document provides a high-level review of the various methods for speech and word extraction. The study also provides a brief summary of the EEG device that can be utilized, as well as how we might combine the two methodologies to develop a classification model for determining the status of consumers.**

Key Words: MFCC, FFT, Mel-scale Filtering, SVM, Multiscale Support vector machine, Feature vector calculation, CNS and PNS.

I. Introduction

Speech disfluency is the disruption in the flow of spoken language that is caused by the speaker, the consumer or patient . It includes shuttering and hesitations people insert to avoid awkward pauses while they find the next words and perhaps ensure there is no opening to allow interruption.

The EEG (electroencephalogram) device is used to collect the data from two sources - the Central Nervous (CNS) system in the form of electrical impulses and Peripheral Nervous System (PNS) in the form of voice . The data collected will be displayed on the screen with the time duration at when it is captured.

The data from the microphone and EEG device is merged to form a classification model.

The data from CNS will be displayed on the screen of an application in the form of a graph of electrical impulses generated in the brain of the subject and this data will be collected with the help of an EEG headset.

The data from PNS will also be displayed on the screen of the application just adjacent to the data of CNS. The data from PNS will be in the form of a voice recording of the subject which will be collected with the help of a microphone.

II. Literature Survey

1. <u>Multisite Schizophrenia Classification based on Brainnetome Atlas by Deep Learning:</u>

<u>Year: 2018</u>

<u>Advantage:</u> Combine the brainnetome atlas (vivo map the brain) to extract features, and propose the generalized feature-invariant deep neural network framework to ensure the model generation in automatic diagnosis of schizophrenia.

<u>Disadvantage</u>:study was need to predict disease progression of longitudinal studies,

2.Model need to be validated with different racial or population

<u>Dataset:</u> Collected 1275 participants at 8 sites, including 662 schizophrenic patients.

<u>Result:</u> Compare the result between their method and SVM classifier, got higher accuracy of 76.82%

2. <u>Automated Verbal and non-verbal Speech analysis of interviews of individuals with schizophrenia and depression</u>

<u>Year: 2019</u>

<u>Advantage:</u> 1.Took both verbal and non-verbal data on a real-time

2.Uses NSA-16 a semi-structured interview containing 16 items that comprehensively assess the negative syndrome of schizophrenia.

Disadvantage: 1.Result was limited by the accuracy of the automatic speech recognition system.2.almost one-fifth to one-third words were missing and about 50% words were not correctly converted.

Dataset: Took real time interview conversation using NSA (negative symptom assessment

Result: Findings; the most salient difference between schizophrenia and depression is response time, whereby patients with schizophrenia took longer time to respond.

Got accuracy of 71.4 % and 70.5% for verbal and non-verbal respectively.

3. Schizophrenia Classification with Single-Trial MEG during Language Processing

Year:2014

Advantage Uses different classification methods and compare accuracy (Bayes, KNN, LDA, SVM)

2.Uses words and sentences for classification

Dataset: Real time data taken from 6 healthy and normal people and 6 schizophrenia patients, to read words and sentences silently while 248 channel MEG signals were recorded.

Result: 4 types for machine learning based classifiers were able to achieve high cross validation accuracy in classifying 470 words and 450 sentence trails into the correct group.

For words LDA gives accuracy of 98% and for sentences KNN with 98%.

4. Multi-layer Support Vector Machine

Year:2018

Advantage: Brief about the multi-layer support vector machine

2.difference between SVM and MLSVM

5. Improvement of MFCC Feature Extraction Accuracy using PCA in Indonesia Speech Recognition

Year: 2018

Advantage: MFCC method with Principal Component Analysis (PCA) to improve the accuracy

Disadvantage: 1.High result feature dimension of the extraction method.

2.Accuracy was limited to the database.

3.language was for Indonesian speech recognition system.

Dataset: Took 140 speech data that were recorded from 28 speakers.

Result: Explain how PCA could be applied in the speech recognition system by wielding the function of feature Data reducing. Adding PCA to MFCC conventional methods could increase accuracy from 80% to 89.29%.

III. Model Approach

3.1 Speech extraction

The first portion will involve using the feature vector computation and the MFCC extraction method to extract features from recorded data or live interactions.

3.1.1 MFCC (Mel-frequency cepstral coefficient) is a shorthand for MFC (Mel-frequency cepstral coefficient), which is a representation of a sound's short-term power spectrum based on the linear cosine transform of a log power spectrum on a nonlinear Mel frequency scale.

Key terms:

1. **Sound Spectrum**: it is usually presented as a graph of power as a function of frequency, in short it is a short sample of a sound representing a sound.
2. **Power Spectrum:** It is a representation of the result of Fourier transform.
3. **Linear cosine transforms:** They are calculated using a Fourier transform. We need both short time and fast time Fourier transform at different stages during conversion.
4. **Mel Scale:** It scales the frequency of a tone to the actual measured frequency.

   A frequency measured in Hertz(f) can be converted to the Mel scale using:

   $Mel(f) = 2595log(1+f/700)$

5. **Fast Fourier Transform:** It converts a signal from its domain (time or space) to a representation in the frequency domain and vice versa.

6. **Nyquist Frequency:** is called the Nyquist limit, is the highest frequency which is coded at a given sampling rate so that it can fully reconstruct the signal, i.e.,

   f_(Nyquist)=1/2nu

7. **Short-time Fourier transform (STFT):** it is used to determine sinusoidal frequency and phase content of local sections of a signal as it changes over time.

   *The idea behind performing this is that it helps us for the short interval of audio which assumed to be steady

Steps to convert audio in MFCC

Step1*:* Get your audio in a time domain format.

Step2: Convert your audio in a periodogram. (Power spectrum is also known as periodogram) with the help of Fast Fourier Transform.

*It's done so as it will give us a Nyquist frequency by down sampling our audio so that we can identify the sound.

Step3*:* After this we convert our periodogram into a spectrogram.

*They are periodograms at different intervals stacked together

Step4*:* We will perform a Short Fourier Transform.

*As STFT is segmenting the signal into narrow time intervals and takes Fourier transform of each segment.

Step5: Now to prevent spectral leakage we will perform a hamming window.

**Spectral Leakage*: It's a mismatch between desired tone and chosen frequency resolution, it takes place due finite windowing of the data (generally when we pass data to DFT/FFT algorithm)

**Hamming Window:* It's a taper formed by using a raised cosine with non-zero endpoints.

Step6: We will again perform Fast Fourier Transform to convert amplitude into frequency.

Step7: Now to convert frequency into **Mel Scale** *As *Mel Scale* it perceives frequency of a tone to the actual measured frequency.

Step8: Now perform logarithm of all filter bank energies.

**Filter bank* separates or splits the input signal into multiple components; it is used to compress the signal when some particular frequencies are more important than other frequencies.

Step9: Perform IDCT of the log filter bank energies.

**IDCT:* Inverse discrete cosine transforms

Step10: Keep IDCT coefficient 2-13, discard the rest of the coefficient (12-13 are considered to be the best)

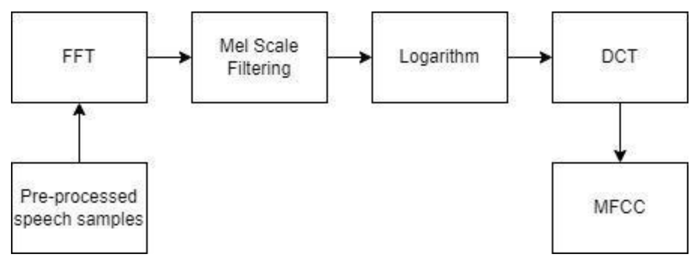

Figure 1*: MFCC feature extraction*

*librosa: it is a python library for analyzing audio and music. It has a flatter package layout, standardized interfaces and names, backwards compatibility, modular functions, and readable code.

3.1.2 EEG headsets, it is a method of tracking the electrical activity of the brain. It measures small voltage fluctuations that result from ionic currents in neurons, i.e., it measures brain waves.

Key terms in EEG and Technical manual steps

The absolute band powers are the brain wave values; the logarithm used for the Power Spectral Density (PSD) of the EEG data for each channel is:

1.Delta - 1 to 4 Hz

2.Theta - 4 to 8 Hz

3.Alpha (7.5-13 Hz)

4.Beta 13-30Hz is a frequency range between 13 and 30 Hz.

The EEG PSD values that we obtained from the sensors were typically in the -1; +1 range.

Raw: The raw EEG values display each sensor's raw data in microvolts, with a range of 0 to 1682.

*Vertical Max is a setting that can be changed. The sensors are:

TP9 is for the left ear, and AF7 is for the left forehead.

AF8 - The right side of the head

TP10 is the number for the right ear.

Right Auxiliary (MU-02/MU-03 only) AUXR

***Figure 2:** Horseshoe*

Figure 2 shows the sensors on the muse and how they are connected.

Here are a few things to keep in mind:

1. Solid ovals make strong connections.

2.Outlines have poor relationships and are difficult to read.

3.An empty space indicates that there is no link.

* The position of AF7 and AF8 on MUSE2 is dependent on head size (it can be available on brochure). These may be closer to FP1 and FP2 on other skulls (Refer brochure and YouTube to adjust these for better performance).

Discrete: Discrete frequency values are represented by a log scale. These are calculated using a hamming window and a Fast Fourier Transform (FFT) of the raw data.

*The FFT window can be customized in settings, and you can choose which sensors to include in the FFT calculation.

Spectrogram: The spectrogram displays discrete frequency data plotted over time.

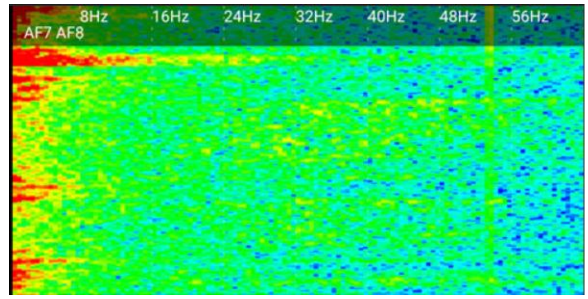

***Figure 3**: spectrogram*

*Low values are represented by blue, medium values are represented by green, and high values are represented by red (fig3)

*Accelerometer- This is the gravitational effect on the headband, measured in g(9.81m/s2).

Some representations are present when the headband is worn on a level head:

1.X will show the up/down tilt of the head.

2.Y represents the movement of the head to the left and right.

3.Z will display vertical up/down motion (jumping). *(Fig4)*

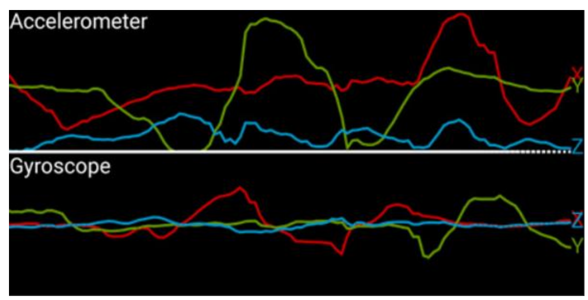

Fig4: Accelerometer and Gyroscope

For full manual and step it's easily available with the EEG product and in the reference link [4].

3.2 Collecting data and Creating .CSV file

After collecting data from the microphone and EEG, we will merge the data in the .csv file so that we can use the data for classification.

***Table1:** represents the dataset after collecting data from microphone and EEG data.*

Time in seconds	Microphone Data		EEG data
Time	Words	Pitch	Amplitude
1.2	hmm	5	8
1	ok	4	7
1.1	yes	4	7

The above table represents the dataset format of the data or what we'll be collecting.

Now the data is being collected in a .csv file we'll refer to the database for Feature Classification.

But here we will be also using a dataset that will be working as a substitute for the MUSE2 headset dataset; it will also be a .csv file that is available on Kaggle.

***Table2**: Muse2 EEG dataset.*

mean_ 0_a	mean_ 1_a	mean_ 2_a	mean_ 3_a	mean_ 4_a
4.62E+ 00	3.03E+ 01	##### ###	1.56E+ 01	2.63E+ 01
2.88E+ 01	3.31E+ 01	3.20E+ 01	2.58E+ 01	2.28E+ 01
8.90E+ 00	2.94E+ 01	##### ###	1.67E+ 01	2.37E+ 01
1.49E+ 01	3.16E+ 01	##### ###	1.98E+ 01	2.43E+ 01
2.83E+ 01	3.13E+ 01	4.52E+ 01	2.73E+ 01	2.45E+ 01
3.10E+ 01	3.09E+ 01	2.96E+ 01	2.85E+ 01	2.40E+ 01
1.08E+ 01	2.10E+ 01	4.47E+ 01	4.87E+ 00	2.81E+ 01

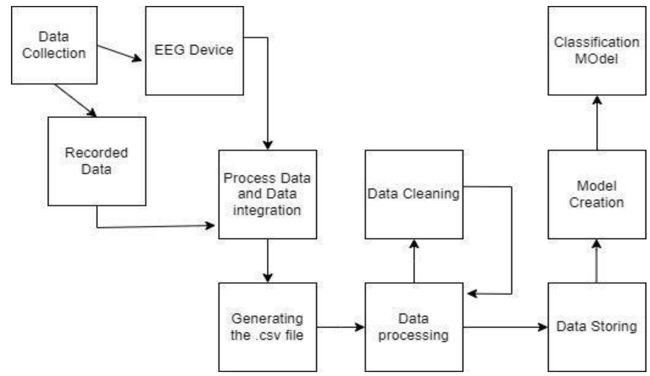

***Figure 5:** Represents the Flow diagram of the overall process*

VI. Feature Classification

We are investigating SVM (support vector machine) for feature classification since SVM uses supervised learning techniques for classification and regression. An SVM classifier's purpose is to fit a hyperplane to a feature space to distinguish two different classes, and it chooses a hyperplane that leaves a maximum margin for both classes.

However, there are situations when noise interferes with the feature. When there is noise, there is a risk of pattern mixing between the classes; to mitigate this, we use MLSVM (multi-level support vector), which draws the second hyper planes. The optimal hyper plane may be created to separate the two classes in this iterative phase, and this will continue for all classes until complete classification has occurred.

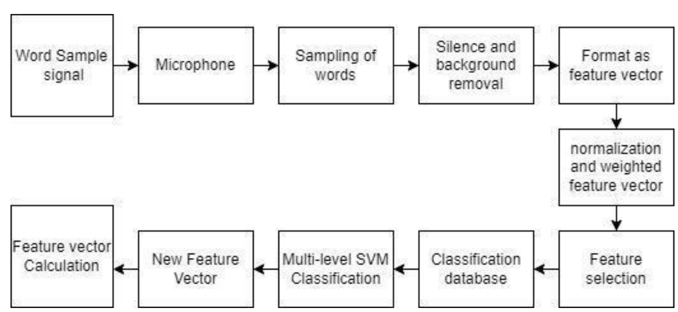

***Fig6**: Multi-level SVM feature classification block-diagram as feature vector*

V. Conclusion

We've gone through how to use an EEG device and microphone to determine a patient's level of mental illness, as well as how to develop a classification model.

We've only examined one type of classification approach; in future research, we'll focus on implementation and alternative categorization models.

VI .Future Scope

For future scope we will be comparing different methods to extract data from EEG and different classification methods which can give the best Accuracy for classification.

Acknowledgment

I'm extremely grateful to Dr. Reena Pagare, Assistant Professor, MIT ADT University, Pune, India, who has been an invaluable source of help and direction during this project.

Author Biography

Raviraj Bhandari,

Pursuing Integrated Mtech degree from MIT Adt University, in Computer science engineering and specialisation in Intelligence system and analysis.

Dr. Reena Pagare,

Associate Professor MIT Art, Design and Technology University, University Lead Manager, Executive Body Member - MIT ADTU Alumni Association, Chair - IEEE Pune Section Engineering in Medicine and Biology Society.

Reference

1. EEG classification During Scene Free-Viewing for Schizophrenia Detection; IEEE transactions on Neural Systems and Rehabilitation Engineering; 2019; DOI: 10.1109/TNSRE.2019.2913799; PubMed ID: 31034418

2. *Website:* MFCC technique for speech recognition MFCC Technique for Speech Recognition - Analytics Vidhya

3. *Website*: DEWESoft Short-time Fourier Transform > Frequency Domain Analysis > Math > General > Modules > Setup | Dewesoft X Manual EN

4. *Website:* MuseIO available data http://developer.choosemuse.com/tools/available-data

5. *Article:* Collecting Brain Signal Data Using The Muse 2 *Website*: EEG Headset Collecting Brain Signal Data Using The Muse 2 EEG Headset | by Tim de Boer | Building a bedroom BCI | Medium

6. *Website*: Our EEG Headset Reviews The Best EEG Headset In 2021 (the-unwinder.com)

7. Speech Mel Frequency Cepstral Coefficient Feature Classification using Multi level support Vector Machine; Abhay Kumar, Sidhartha Sankar Rout, Varun Goel; 2017 4th IEEE Uttar Pradesh section international conference; UPCON GLA university, Mathura.

8. Language Identification from Speech Features Using SVM and LDA; J.S. Anjana and S.S. Poorna; 2018 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET); 2018 IEEE

9. *Website:* Support Vector Machine; Support Vector Machine (SVM) Algorithm - Javatpoint

10. Multilevel Weighted Support Machine for Classification on Healthcare Data with missing Values; Talayeh razzaghi, Oleg Roderick, IIya Safro, Nicholas Marko; 2016; Multilevel Weighted Support Vector Machine for Classification on Healthcare Data with Missing Values (plos.org)