Comparison of the Effectiveness of Different Machine Learning Algorithms in Classifying Stuttered Speech Behaviors

**Project Team Members**

Juan Diego V. Huet

Aaron Andrei D. Ambas

Kyle Rafael F. Sulabo

1. ***Introduction***

***Background of the Problem***

Stuttering is “…a speech disorder in which the flow of speech is disrupted by involuntary repetitions and prolongations of sounds, syllables, words or phrases as well as involuntary silent pauses or blocks in which the person who stutters is unable to produce sounds”. (World Health Organization, 2010). The primary behaviours manifested by stuttering are repetition, prolongation of sounds, and blocking (Northern Arizona University, n.d.). A repetition occurs whenever the person who stutters repeats a sound, syllable, or a one syllable word more than once or twice; a prolongation occurs when a speech sound is held out but the mouth, lips, or tongue stops moving; blocks occur when the person stops the flow of sound or air in the lungs, throat, mouth, lips, or tongue. (HomeSpeechHome.com, n.d.)

A speech language pathologist conducts speech therapy as well as diagnosis. Speech therapy is the process of correcting disfluency in speech. A speech diagnosis is needed before the actual speech therapy in order to accurately decide what treatment a patient should receive (Scott, 2008).

The speech pathologist conducts the diagnosis by first holding a casual conversation with the patient. Behaviors that are manifested physically by the patient are then documented by the speech pathologist. Afterwards, tests are conducted to measure the severity of stuttering by conducting speech focused tests that will test the patient’s fluency. The patient will be asked to read a passage where the percentage of the total number of words stuttered over the total number of words in the passage will be recorded. The patient will also be asked to hold a conversation with the speech pathologist with the patient discussing ideas about a certain topic, this time the patient’s voice will be recorded for use as reference. The percentage of the total number of words stuttered over the total number of words used by the patient will also be recorded. The speech pathologist also takes note of the different types of disfluency observed while conducting the speech tests.

The traditional method of diagnosing speech is generally time consuming. Another issue with it is that different speech pathologists may make their own different judgements when diagnosing (Kully and Boerg, 1988).

***Statement of the Problem***

* Diagnostic results tend to be dissimilar when compared between judges who reviewed the same patient sample.
* Speech assessments are time consuming to make.

***Objectives***

* Build a collection of audio files containing instances of stuttered speech
* Utilize an audio analysis program to extract the required audio features
* Implement different machine learning algorithms that will learn to classify audio with stuttered speech instances based on the extracted audio features.
* Quantify and compare the accuracy of the different machine learning algorithms used in classifying stuttered speech behaviors.

***Significance***

The findings of this research will benefit the following agencies:

*To People who Stutter*

This project will benefit patients under speech therapy that will undergo speech diagnostics. The patient can be evaluated more objectively through the use of a computer and thus improve the process of the actual therapy.

*To Speech Pathologists*

Speech Pathologists will be able to use this project to produce more accurate and more objective results when conducting a speech diagnosis. Better diagnostic results will lead to more appropriate therapies among patients. This can also be used alongside the traditional method of diagnosis to lessen ambiguity.

To Other Researchers

Other researchers will benefit from this study by having a better idea of what machine learning algorithms will be most effective when classifying audio features in the context of stuttered speech.

***II. Scope and Limitations***

The study aims to compare the effectiveness of different machine learning algorithms in classifying stuttered speech. The extracted features from the audio samples will be fed into the AI, this will serve as the training data for the AI. The results of the accuracy of the different machine learning algorithms will then be compared. The study only involves processing of audio files with solely one instance of stuttered speech each.

**III. *Related Literature***

***Behaviors of Stuttering***

Stuttering behaviors are mainly characterized into primary and secondary. The primary behaviors are:

* Repetitions of sounds, syllables and words
* Prolongation of single sounds
* Blocks of airflow when speaking

Secondary behaviors include:

* Hesitations
* Interjections of sounds, syllables of words (ahh, uhm)
* Word revision, word changes
* Unnecessary motor movements

(Northern Arizona University, n.d.)

***Related Researches***

*Automatic Detection of Syllable Repetition in Read Speech for Objective Assessment of Stuttered Disfluencies*

*(K. M. Ravikumar, Balakrishna Reddy, R. Rajagopal, and H. C. Nagaraj )*

This study proposed automatic detection of repeated syllables through the use of 4 major steps: segmentation, feature extraction, score matching, and decision logic. The collected audio samples are first segmented into syllables. After segmentation, the segmented speech syllables are subject to feature extraction. The audio features of the syllables are obtained by using the Mel frequency Cepstra coefficient algorithm. Score matching is then done using the Dynamic Time Warping Algorithm. Lastly, the decision logic uses the Perceptron machine learning algorithm to determine if the syllable is repeated or not.

*An Approach for Classification of Disfluent and Fluent Speech using K-NN and SVM*

*(P. Mahesha and D.S. Vinod)*

This study proposed classification of whether a segmented speech is fluent or disfluent. The researchers defined disfluent speech as speech containing repetitions, prolongations, interjections, and pauses. The MFCC’s are obtained from the speech samples and used as the data to train two machine learning algorithms (K-NN classifier and Support Vector Machine).

*Yet Another Audio Feature Extractor (YAAFE)*

YAAFE is a toolbox for audio features extraction. YAAFE’s source code is open source and compiles on Linux and Mac OsX platforms. Yaafe is developed at Telecom Paristech/ AAO Team and uses other open source projects like Eigen, Smarc, libsndfile, mpg123, HDF5.

*YAAFE, an Easy to Use and Efficient Audio Feature Extraction Software*, B.Mathieu, S.Essid, T.Fillon, J.Prado, G.Richard, proceedings of the 11th ISMIR conference, Utrecht, Netherlands, 2010.

*Automatic detection of prolongations and repetitions using LPCC*

*(Harihan M. and Sazali Yaacob)*

The researchers of the study extracted Linear Prediction Coefficient Cepstras from segmented stuttered speech instances. K-Nearest Neighbor and Linear Discriminant Analysis (machine learning algorithms) were used to classify extracted audio features into repetitions or prolongations.

**III. *Technical Background***

*Machine Learning*

Machine Learning is a type of artificial intelligence that provides computers with the ability to learn without being explicitly programmed (whatis.techtarget.com, nd.). It involves development of computer programs that can teach themselves to grow and adapt to change when new data is exposed. Machine Learning systems look through data to look for patterns and detect those patterns in the data to and adjust actions accordingly. A machine learning model refers to the model artifact created by training (Amazon Web Services Documentation, Amazon Machine Learning Guide, Training Machine Learning Models).

Machine Learning algorithms are usually classified into two: Supervised Learning and Unsupervised Learning. Supervised learning uses what it has learned in the past and applies it to new data. Some examples of algorithms that belong to Supervised Learning are Support Vector Machines, linear regression, naive Bayes, Neural Networks, etc. Unsupervised Learning on the other hand describes hidden patterns from unlabeled data. Some examples are clustering through k-means, mixture models, or hierarchical clustering; anomaly detection; and Neural Networks.

*Mel Frequency Cepstral Coefficient*

Introduced by Davis and Mermelstein in the 1980’s, MFCC represents the envelope of the short time power spectrum manifested by the shape of the vocal tract of humans. MFCCs give accurate representations of phonemes being produced (practicalcryptography.com, n.d.).

***IV. Proposed Solution to the Problem***

A. Segmentation

Audio samples of stuttered speech are obtained from the University College London’s archive of stuttered speech (UCLASS). This archive was created for convenience to research and clinical purposes. Only stuttered words are to be segmented from the different audio files manually. 80% of these clips will be used to train the machine learning models while the remaining 20% will be used to test the accuracy of each proposed model. Other related researches have segmented the stuttered instances by word, such as P. Mahesha and D.S. Vinod’s research (An Approach for Classification of Disfluent and Fluent Speech using K-NN and SVM).

B. Feature Extraction

The segmented stuttered word clips will each be then analyzed by the program Yet Another Audio Feature Extractor (YAAFE). Two audio features will be extracted, namely the Mel Frequency Cepstral Coefficients and the Linear Predictive Cepstral Coefficients. The two algorithms are performed by YAAFE in audio feature extraction. The results will be manually classified whether the the results belonged to a repetition or prolongation.

C. Machine Training

Different machine learning algorithms will be used to classify the two stuttered behaviors. The result set of the extracted features will be the data set that will be fed to the different machine learning algorithms. The result will be a model that aims to classify the stuttered behaviors for each algorithm used.

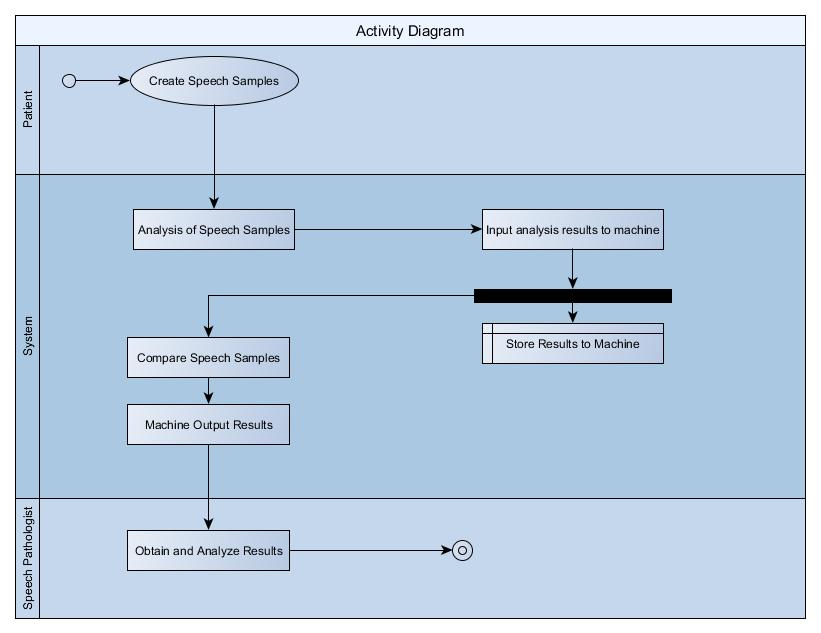
D. Machine Testing and Comparison

The second batch of extracted audio features will then be used to then test the accuracy of the resulting models. The extracted audio features will be analyzed by the models and classify them according to which behavior they might be. The percentage of accuracy of each model will be recorded. The accuracy of each model will be compared based on the percentage.

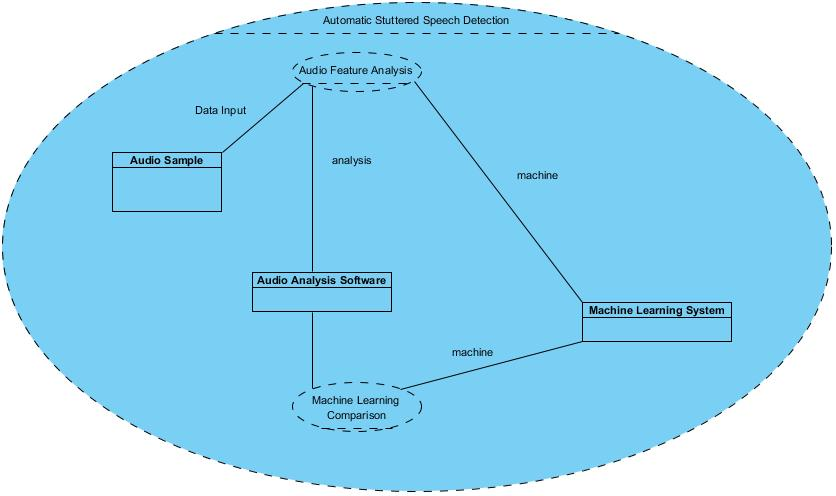
**V. *Appendices***

**5.1 Diagrams**

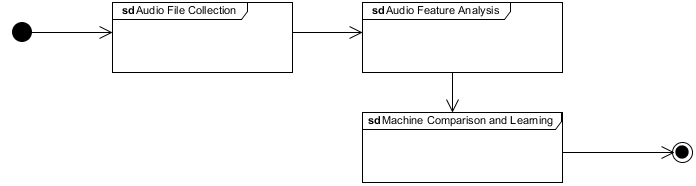
*Activity Diagram*



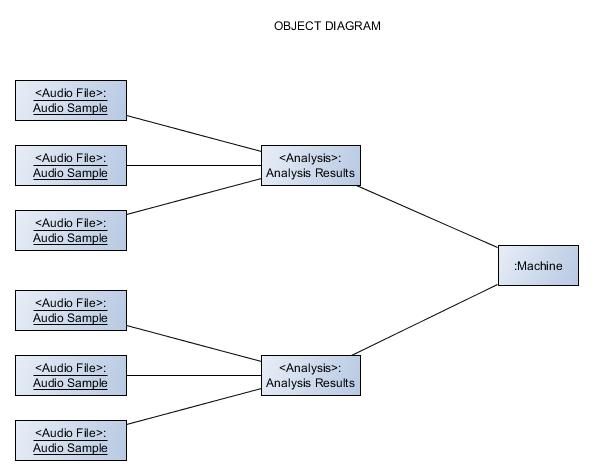
*Composite Structure Diagram*



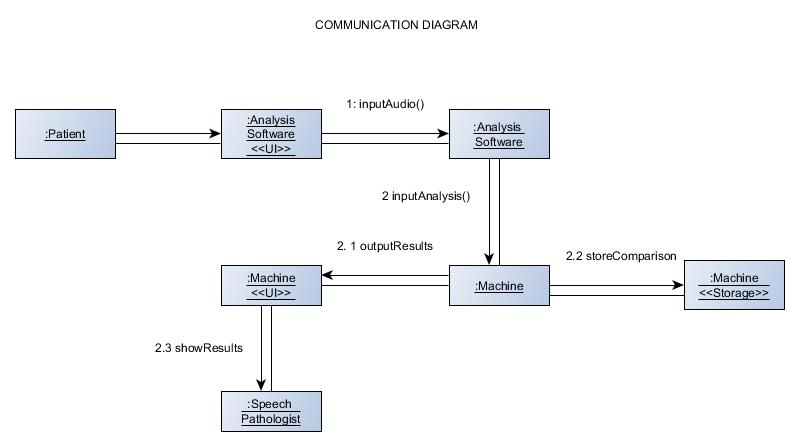
*Interaction Overview Diagram*



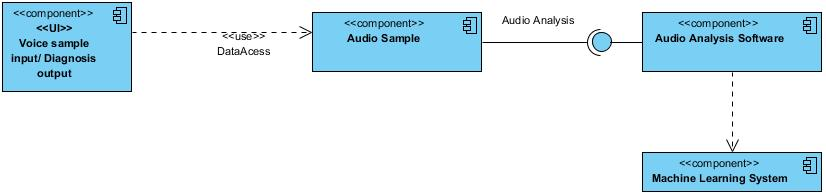
*Object Diagram*



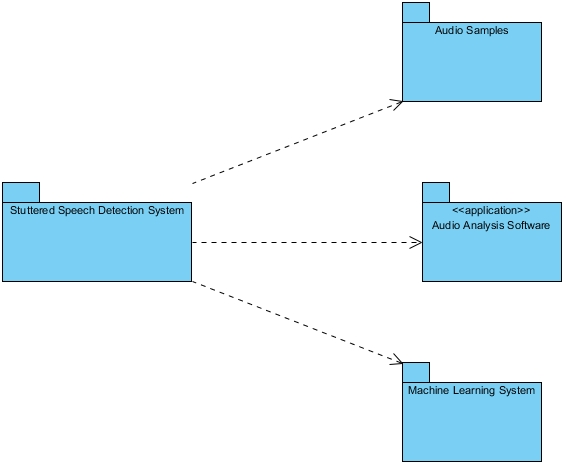
*Communication Diagram*



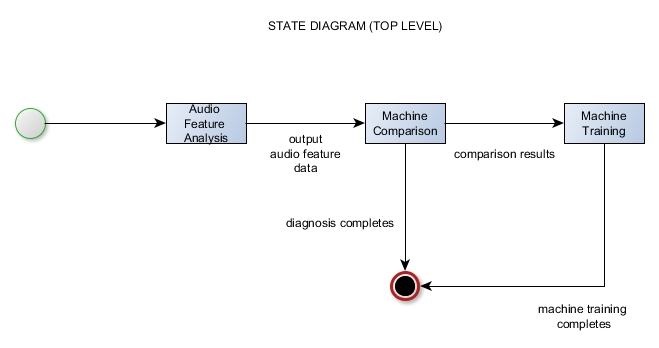
*Component Diagram*



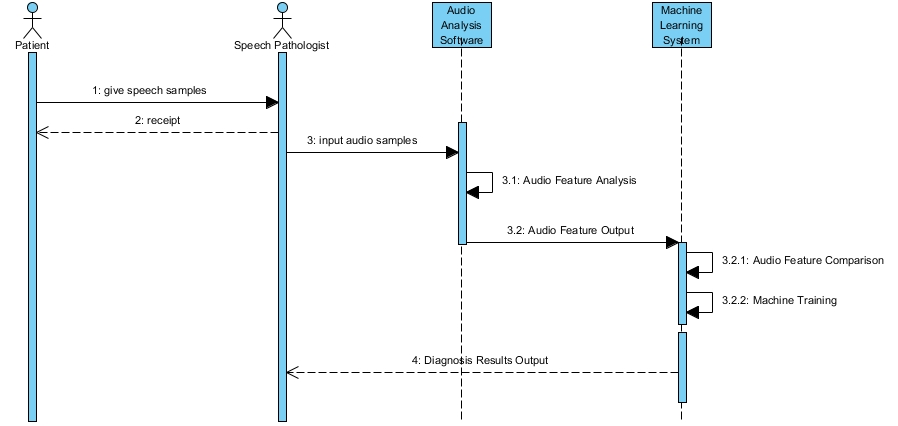
*Package Diagram*



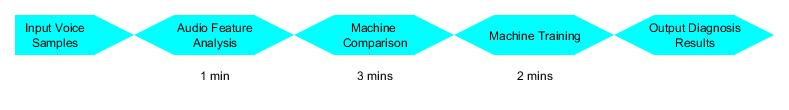
*State Diagram*



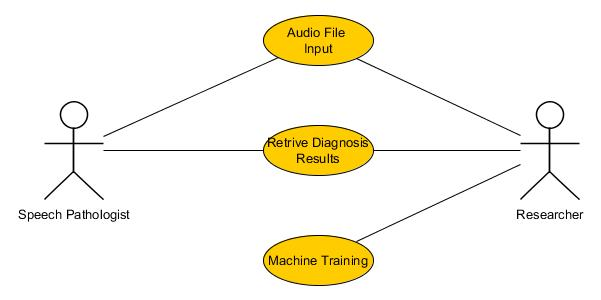
*Sequence Diagram*



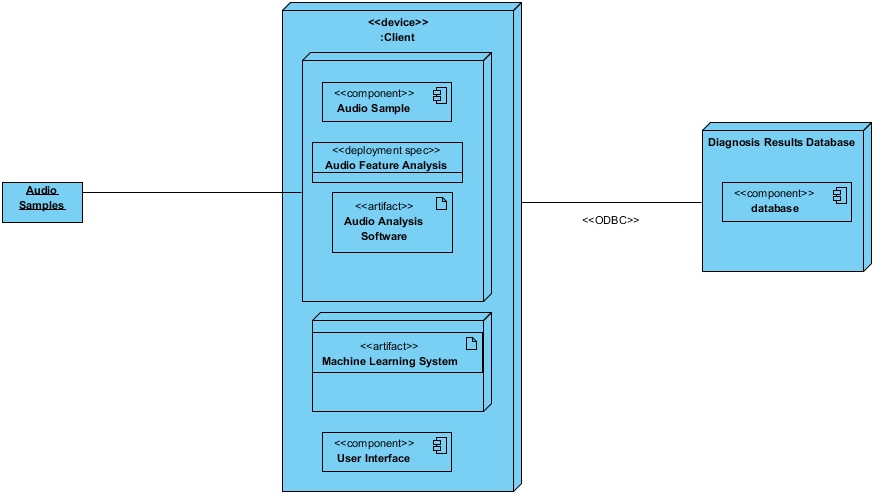
*Timing Diagram*



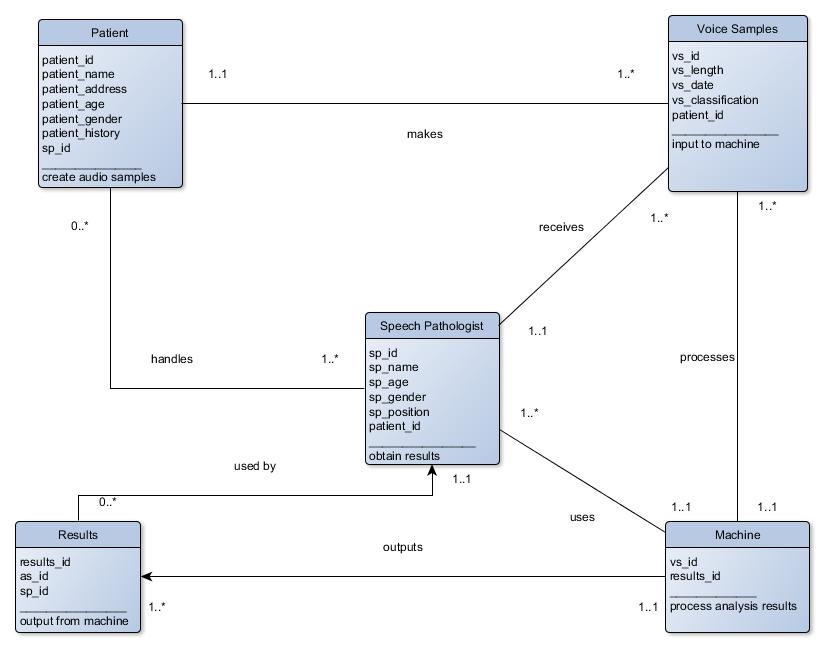
*Use Case Diagram*



*Deployment Diagram*



*Class Diagram*



**5.2 Event Table**

