© Universiti Tun Hussein Onn Malaysia Publisher's Office

IJIE

Journal homepage: http://penerbit.uthm.edu.my/ojs/index.php/ijie
ISSN : 2229-838X e-ISSN : 2600-7916

The International Journal of Integrated Engineering

Voice Biometric System: The Identification of the Severity of Cerebral Palsy using Mel-Frequencies Stochastics Approach

Syifaun Nafisah[1], Nazrul Effendy[2*]

[1]Department of Library and Information Science, Sunan Kalijaga State Islamic University
Jalan Marsda Adi Sutjipto Yogyakarta 55281, INDONESIA

[2]Department of Nuclear Engineering and Engineering Physics, Faculty of Engineering, Universitas Gadjah Mada
Jalan Grafika 2 Yogyakarta 55281, INDONESIA

*Corresponding Author

DOI: https://doi.org/10.30880/ijie.2019.11.03.020
Received 24 July 2019; Accepted 31 July 2019; Available online 3 September 2019

Abstract: Cerebral Palsy (CP) is a neurological condition that causes problems in body movement and muscle control that can inhibit the development of children's speech. It can be classified into several types, based on the stiffness of the muscles that they suffer. This study offers a new technique for identifying the severity CP level of children based on their speech. This study utilizes the capabilities of the voice biometrics system (VBS) to authenticate people based on their voice. The Mel-frequency stochastic model is also offered as a new approach in the feature extraction process. Because of the pattern of speech signals of children with CP which is irregular, then neuro fuzzy is chosen as a method in the speech classifier. Based on the experiment conducted to respondents, the accuracy of the technique is 87.5%. This result shows good performance of the new approach for realizing the research objective.

Keywords: CP, speech, biometric, feature extraction, Mel-frequencies, stochastic

1. Introduction

Cerebral Palsy (CP) is known as one type of disabilities that affects children. CP is an 'umbrella' for all chronic neurologic disorders which was manifested in movement control disorders. They arise early in life, commonly initiated with non-progressive diseases [1]. This neurological condition causes problems in body movement and muscle control that can inhibit the development of children's speech. The inhibition of speech development in CP is caused by the stiffness of the muscles around the mouth and face. It causes the children with CP having difficulty coordinating their speech organs.

In CP, there is a disorder occurs that causes the muscles to move on their own without realizing it. This disorder is called dystonia. Regarding the development of speech on CP where the development is affected by the stiffness of the muscles around the mouth, then dystonia can be an indication of the severity level of CP. Based on dystonia that occurs on CP, CP can be classified into 4 severity levels: Slight, Mild, Moderate, and Severe. In Slight level, dystonia occurs less than 10% of the time and doesn't interfere with speech. Whereas in mild level of severity, dystonia occurs less than 50% and doesn't interfere with the speech. In moderate level, speech impairments caused by dystonia in CP begin to suffer where the dystonia is more than 50% of the time. Even at severe level, children with CP are difficult to move his/her mouth [2],[3].

Dystonia that occurs in CP causes CP speech has unique characteristics like "slurred," "choppy," or "mumbled" that may be difficult to understand. CP speech is also at a slow rate and the rapid rate in mumbling quality. Voice quality changes in CP, such as hoarse and breathy voice or speech that sounds "nasal" or "stuffy", are caused by limited tongue,

Corresponding author: nazrul@ugm.ac.id

lip, and jaw movement and abnormal pitch and rhythm when speaking [4]. Referring to the results of a study conducted by Ann W Kumar in 2014, the Speech and Resonance Disorders is related to Cleft Palate and Velopharyngeal Dysfunction, so speech impairment in CP may be related to dystonia in CP [5]. Similar thing was reported by Lindsay Pennington in 2016 that there was a relationship between the ability to move the muscles around the mouth to coordinate speech organs with the severity level of CP [6]. Based on their findings, a hypothesis is built in this study, that the severity level of CP of children can be identified from their speech characteristics, where the characteristics of CP speech are determined based on dystonia that they suffer. It also based on a study conducted by Craig., et al., which states that the speech spectrum in CP is different from the speech spectrum in autism and Attention Deficit Hyperactivity Disorder (ADHD) [7].

The classification of the severity level of CP is important to do to understand each of children with CP. The understanding is important and must be understood by the surrounding environment in order to establish his/her independence and to determine the treatment interventions that will be carried out on her/him. The success of the treatment depends on the level of cognition of CP and there is a strongly associated between cognitive impairments and speech disorders in CP [8]. In addition, continuous stimulation will stimulate brain deep of the children with CP. It provides predictions of the good results of the implementation of therapy for children with CP [9]. For this reason, in the last decade, many studies have been conducted to find the classification method of the severity level of CP. The method of classifying the severity of CP based on the Gross Motor Function Classification System (GMFCS) was carried out by Hidecker MJC, Paneth N, Rosenbaum PL, et al. in 2011. This classification is also based on the study which is conducted by Krigger in 2006 [10]. The results of this study were stated that there was a strong relationship between speech impairment in CP with cognitive impairments and the severity level of CP. It can be determined based on their stiffness of the muscles [11]. The method of classification of CP severity was also carried out by Harvey et al in 2013. The classification carried out based on the communication function of CP. The result of this research that they conducted based on the Communication Function Classification System (CFCS) is stated that there was a correlation between speech disorders with the severity of CP [12].

Seeing the many studies regarding the CP classification method, this study offers a new method for classifying CP based on the speech that they produce. In information technology, the field of this science used for people recognition based on their speech is known as the voice biometric system (VBS). Previously, there have been researches related to the identification of CP based on their Speech using the Automatics Speech recognition System (ASR). The method was applied to the problem of recognition of CP using information in their speech [13],[14]. The difference between ASR and VBS is ASR is used to identify "what is said" by the speaker, while VBS is used to identify "who speaks". In this study, VBS is more appropriately used because the focus of this study is not on what is their said but on the severity level of CP what the word is said. Aside from being an identification tool, ASR and VBS also can be used as a therapeutic aid. This is based on research which is conducted by Yusof., et al (2013) which developed the Malay Speech Intelligence Test (MSIT) as a therapeutic aid for Deaf Malaysian Children. From the experimental results, it is known that testing in the therapeutic process in deaf children using this aid shows an accuracy rate of 95% similar to the therapy testing carried out by experts [15].

VBS is a system that can be developed interactively. The interactive therapy can motivate CP during rehabilitation therapy because this system can adapt to the changing levels of skills, interest or fatigue of the clients [16]. The intervention using an interactive tool is also can improve the gross motor function of children with CP, GMFCS I-II [17]. Based on the ability of VBS to recognize people based on their voice, this study offers a new method for recognizing CP severity level based on their speech. Due to the unique characteristics of CP speech such as described above, the approach offered in this study is a stochastic approach based on the speech frequency spectrum that they were produced. The stochastic approach is chosen in this study because the speech frequency spectrum of CP doesn't have a random pattern. The new approach offered is expected to help the therapists to classify the severity of CP.

2. Research Methodology

The steps in this study are carried out in the following stages.

2.1 Voice Biometrics System (VBS) Development

Working procedure in the process of VBS in this study is presented in Fig. 1.

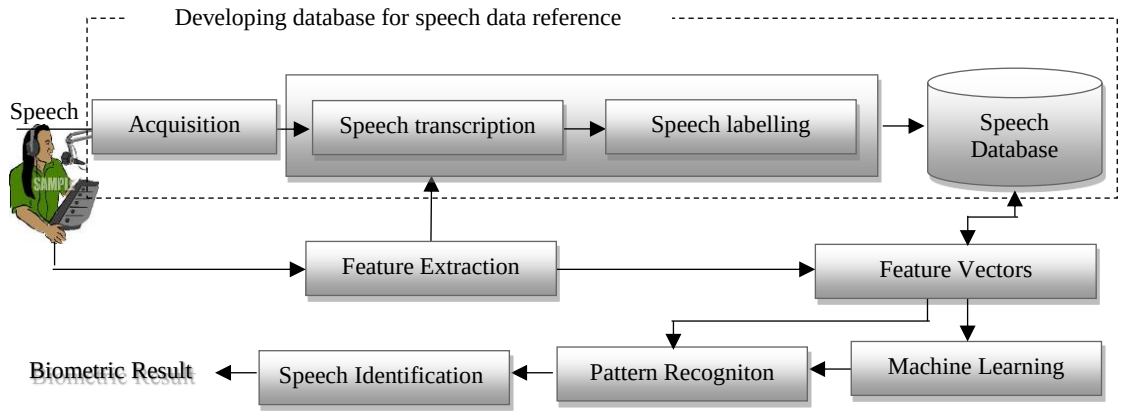

Fig. 1 - Block diagram of the VBS

2.1.1 Developing Database for Speech Data References

A database is a place that is used to store data. In VBS, a database is a place that is used to store data references, i.e. data sample of speech that has been taken the features and will be used in the pattern matching process. Speech references are collected by acquisition process using a vocal microphone PG48-LC with flat frequency in 10 Hz-20 kHz and mini mixer Euro Rack UB1002FX connected with DAT Recorder via sound card M-Audio 16-bit, and Cool Edit 2.0 as aided software.

The data references are obtained from 26 respondents which consist of 7 respondents with no pathologic motoric abnormalities, 9 respondents are CP in mild level, 5 respondents are CP in a moderate level and 5 respondents are CP in severe level. Each word is recorded as many as 10 times [18] and the list of words in a database is in Indonesian that was chosen based on guidelines of Glenn Doman method [19]. The list of the words is shown in Table 1.

Table 1 - The List of Words

No	Words	No	Word	No	Words
1	a-yah	6	ke-ju	11	to-pi
2	i-bu	7	ma-ta	12	bu-ku
3	a-dik	8	gi-gi	13	me-ja
4	ro-ti	9	pi-pi	14	kur-si
5	na-si	10	bo-la	15	pin-tu

Each word was recorded for approximately 2 seconds and was pronounced in syllable to make it easier to take the characteristics of each syllable. The duration of the pronunciation for approximately 2 seconds refers to the speed of speech such as described by Goss which states that the speed of human speech in 1 minute on average 125 words [20]. Based on these findings, it can be concluded that each word will be pronounced during:

$$T_s = \frac{125}{60} = 2.083 \quad (1)$$

The acquisition of speech that has been conducted on 26 respondents produced several speech references as shown in Table 2. The data were divided into two groups, namely group I - training data and group II - testing data.

Table 2 - Number of Speech references in Database

Speech Data References	
Word	**Syllable**
3900	7800

The fundamental frequency used in this study is 22050 Hz so that the speech produced has a clear quality. The process of speech recording in the acquisition process is called the sampling process. In the sampling, the Nyquist rule applies, where the sampling frequency (f_s) must be greater or equal to twice the maximum sampling frequency to eliminate the aliasing effect. Aliasing effect is an effect where the speech signal from the sampling process results is different from the original speech signal. Based on this rule, the sampling frequency (fs) used in this study is 2 x 22050

Hz or equal to 44100 Hz. In the process of speech recording, if each word is recorded in a duration of 2 seconds at f_s = 44100 Hz using a 16-bit sound card, then the size of each sound file formed in each word in the database is calculated by equation (2).

$$File\ Size\ (FS) = F_s * dt * \left(\frac{bit}{8}\right) * channel \quad (2)$$

The equation (2) can be explained as follows. File size (FS) is large file size, *dt* is the duration of pronunciation, bit is the size of bit that is used in the sound card and the channel is the type of channel used. There are 2 types of the channel in the recording process, namely mono, and stereo channel. The mono channel is a channel that can be set to be heard on one channel or both, and the stereo channel is a channel where the sound signal on the right and left side of the sound will sound the same. Mono channel has a coefficient of value = 1 while stereo channels have a coefficient of value = 2. In order to ensure data speech quality, the channel used in this study is a mono channel with channel coefficient value = 1. Based on the formula in equation (2), FS for each data that was stored in the database is:

Frequency sampling (f_s)	=	44100 Hz
File size (FS)	=	44100 Hz x 2 x (16/8) x 1
	=	176400 bytes

The sampling speed of data depends on the file size and the duration of the sampling. The speed of sampling (sampling rate) is calculated using the formula in equation (3).

$$\text{sampling rate (SR)} = \frac{FS}{T_S} \quad (3)$$

If FS is 176400 bytes and the duration of the sampling is 2 seconds, the sampling rate (SR) is:

$$\text{sampling rate (SR)} = \frac{176400}{2} = 88200 \text{ bytes per second}$$

Based on the sampling rate, then the number of energy points in each sample can be determined at certain time intervals. The time interval (*ts*) used in this study is 0.02 milliseconds so that the speech signal can be stationary. The number of sample points (SP) can be calculated using equation (4):

$$\text{Sample Points (SP)} = \text{SR} * ts \quad (4)$$

= 176400 * 0.02 = 3528 points.

Based on the calculation, the number of energy points in each sample is as much as 3528 sample points.

2.1.2 Feature Extractor

The second step after developing the database for speech data references is taking the features of each word. These features will distinguish between one speech and the other. To distinguish the features of speech can be seen from the energy of speech that is released during the pronunciation of a word. The energy is generated by frequencies of signals of speech. This energy is called Power Spectral. This spectral power was used in the analysis phase using Mel-frequencies stochastics approach. This phase is called feature extraction. The procedures of feature extraction using Mel-frequencies stochastics approach is presented in Fig. 2.

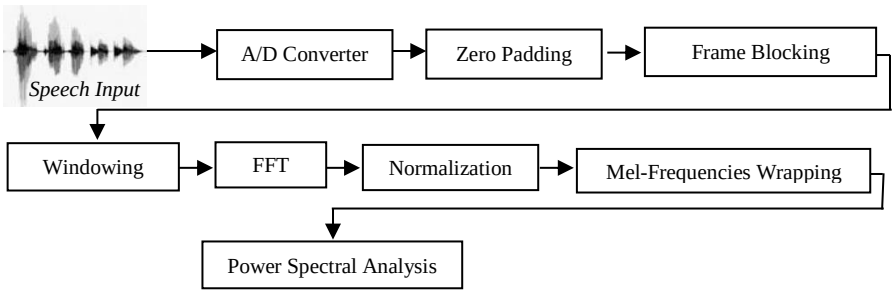

Fig. 2 - Block Diagram of Mel-Frequencies Stochastics Approach

In the speech recording process in the database development, the length of speech signals is very varying. It is necessary to do the zero padding process to equalize the length of the signals. The first step in the feature extraction stage is zero padding.

1) Zero padding

Zero padding is a process to equalize the length of the speech signals by adding a value of 0 until the end of the length of signals that is agreed. In this study, the duration of speech is ± 2.083 seconds; if the speech consisted of two syllables than it takes pronunciation time between 0.7 to 2 seconds. The pronunciation time will affect the

signal length. It is necessary to conduct zero padding so that the signal has the same length. In this study, the signal length is agreed in 100000 (x_n = 100000). Thus, all the signals have a vector size of 100000 x 1.

2) Frame Blocking

After the signal length is matched, the signal will be cut into several parts of signals. This part is called the frame. The process of cut of the signal into several parts is called frame blocking. The illustration of the signal that has been blocked into the frames is shown in Fig. 3. The frame blocking is in overlapping so no signals are lost.

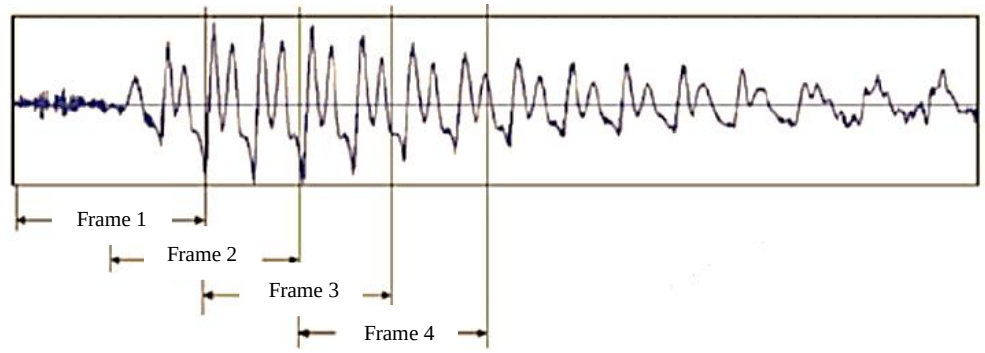

Fig. 3 - The signal that has been blocked into frames

The number of frames generated from the frame blocking process is calculated by equation (5).

$$number\ of\ frame\ (NF) = \left(\left(\frac{\mathrm{SR}-\mathrm{SP}}{\frac{SP}{2}}\right) + 1\right) \tag{5}$$

The SP in equation (5) is the sample point, while SR is the sample rate. In this study, the signal length after passing the zero padding process is *xn* = 100000. The signal will be cut every 0.15 milliseconds by overlapping cuts [21]. Based on the formula in equation (5), the number of frames is:

$$NF = \left(\frac{88200-3528}{\frac{3528}{2}} + 1\right) = \left(\frac{84672}{1764} + 1\right) = 49 \text{ frames}$$

If the signal is in overlapping, with the percentage of overlapping (PO) is 50%, then the number of overlapping frames is calculated based on equation (6).

$$NF_2 = NF + (PO * NF) \tag{6}$$

Based on the equation (6), the number of overlap frames is:

$$\begin{aligned} NF_2 &= 49 + (0{,}5 * 49) \\ &= 49 + 24.5 \\ &= 73.5 \approx 74 \text{ frames} \end{aligned}$$

3) Windowing

After the signal is blocked into frames, a windowing process will be carried out at every frame. The windowing is a process to get the coefficients of windows for each frame. The windows function used in this study is a Hamming window function. Window coefficients with Hamming function are calculated based on equation (7).

$$w(n) = 0.54 - 0.46\left(\frac{2\pi n}{NF_2 - 1}\right),\ 0 \le n \le N - 1 \tag{7}$$

Based on equation (7), the window coefficients on each frame are as follows.

Frame 0:

$$w(0) = 0.54 - 0.46\left(\frac{2\pi 0}{74-1}\right)$$
$$w(0) = 0.54 - 0.46(0)$$
$$w(0) = 0.54 - 0$$
$$w(0) = 0.54$$

Frame 1:

$$w(1) = 0.54 - 0.46(\frac{2\pi 1}{74-1})$$
$$w(1) = 0.54 - 0.46(0{,}00803438718)$$
$$w(1) = 0.54 - 0.000366071535$$
$$w(1) = 0.539633928$$

Frame 2:

$$w(2) = 0.54 - 0.46(\frac{2\pi 2}{74-1})$$
$$w(2) = 0.54 - 0.46(0.172054795)$$
$$w(2) = 0.54 - 0.0791452057$$
$$w(2) = 0.460854792$$

.
.
.

Frame 73:

$$w(73) = 0.54 - 0.46(\frac{2\pi 74}{74-1})$$
$$w(73) = 0.54 - 0.46(6.3660274)$$
$$w(73) = 0.54 - 2.9283726$$
$$w(73) = -1.8483726$$

4) Fast Fourier Transform (FFT)

Before the signals are analyzed, FFT process is carried out. FFT is a method for transforming speech signals into frequency signals. It means that the speech recording process is stored in a digital form based on speech spectrum waves. The results of the FFT process produce domain frequency waveform in discrete form. In FFT, the speech spectrum waves are computed using equation (8).

$$Y(k) = \sum_n^N x(n) * w(n) \quad (8)$$

If there is a part of the spectral energy value of a signal:

x(0)	*x* (1)	*x* (2)	…	*x* (73)
-0,0005	0,0030	0,0035	…	0,0021070

Then the values of the spectral energy after passing through the windowing process are:

Frame 0:

$$Y(0) = -0.0005 * 0.54 = -0.00027$$

Frame 1:

$$Y(1) = 0.0030 * 0.539633928 = 0.0016$$

Frame 2:

$$Y(2) = 0.0035 * 0.460854792 = 0.0016$$

.
.
.

Frame 73:

$$Y(73) = 0.0021070 * (-1.8483726) = -0.0039$$

Then the value of spectral energy after through the windowing process is:

Y(0)	Y(1)	Y(2)	…	Y(73)
−0.00027	0.0016	0.0016	…	−0.0039

5) Power Spectral Analysis

After the signal is transformed into a frequency signal, the power spectral analysis process is carried out. In this study, the power spectral analysis is calculated using the Short Term Energy (STE). The STE is calculated using equation (9).

$$STE = \sum_{i=0}^{T}(x(t)w(t))^2 \quad (9)$$

The STE formula in equation (9) is the squared of energy in each window. Based on equation (9), STE in each window is:

STE_0	STE_1	STE_2	…	STE_{73}
4e-8	0.0050	0.0291	…	0.2861

Fig. 4 shows the waveform signals.

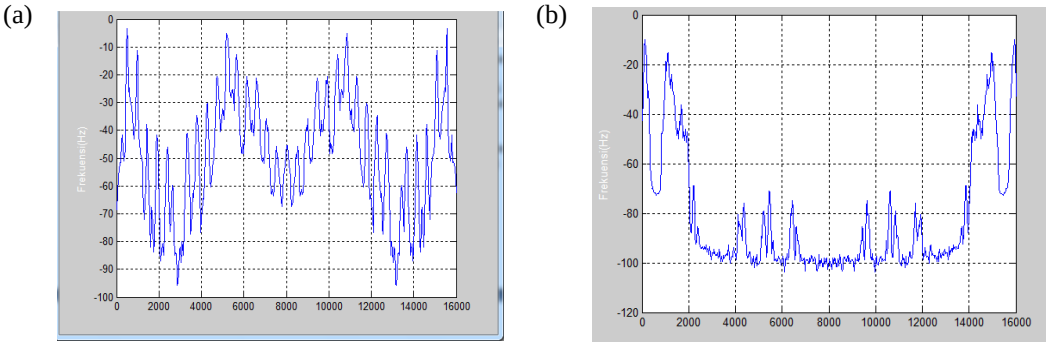

Fig. 4 - The waveform of signals in time domain (a) and frequency domain (b)

6) Normalization

The energy values generated from STE calculations show varying values. It will be a difficulty in the process of patterns matching. For this reason, the normalization process is important to carry out. Normalization is the process of changing STE values in the range between 0 and 1. The formula for the normalization process is in equation (10).

$$x_{norm_i} = \frac{x_i - \min(x)}{\max(x) - \min(x)} \quad (10)$$

7) Mel-frequency Wrapping

The last process in the feature extraction process is computing the Discrete Cosine Transform (DCT) of the log filter bank energies. To compute the DCT, the frequency of the signal should be converted into Mel scale using the following equation.

$$M(f) = 1125 \ln(1 + \frac{f}{700}) \quad (11)$$

The flowchart of Mel-frequencies stochastics approach is shown in Fig. 5.

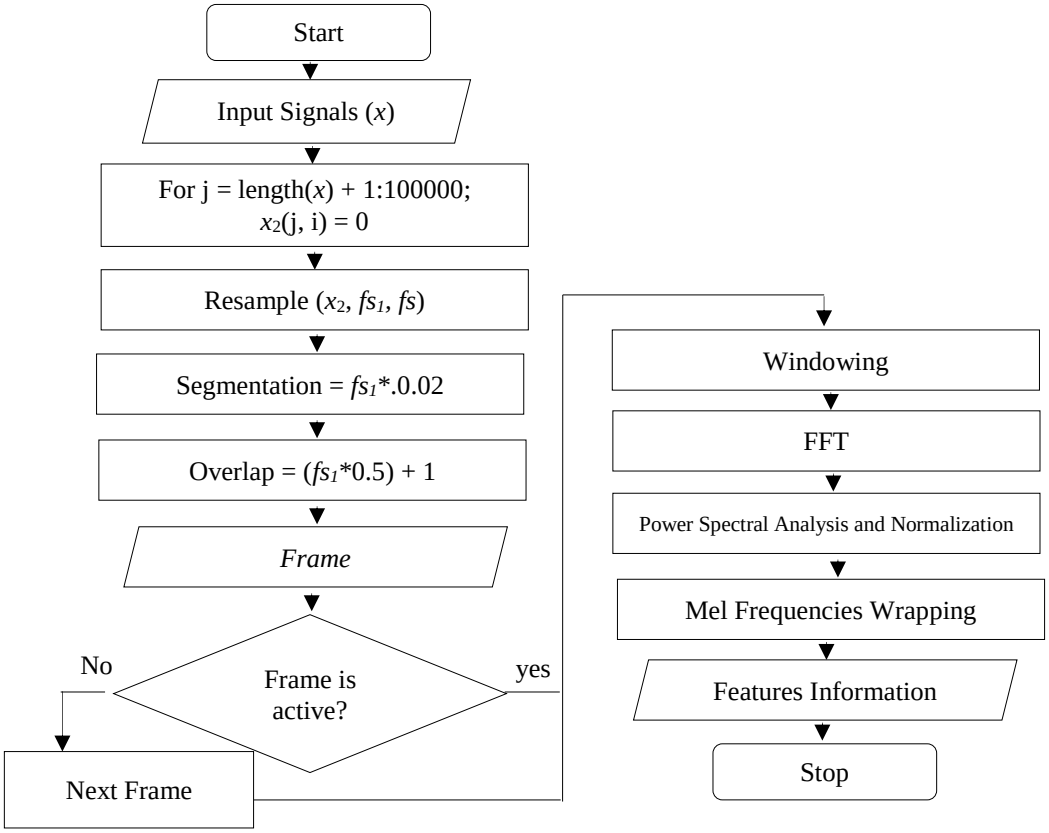

Fig. 5 - Flowchart of Mel-frequencies stochastics approach

The result of this step is the diagonal covariance matrices that can be used as features information. The features produced in this process will be an input in the classifier.

2.1.3 Speech Recognizer

A Speech recognizer is designed to perform pattern matching that states the results of benchmarking scores of the pattern similarity between the series of the feature vector of signal tested and the series of the feature vector of the reference model. Information of the signal features in this study is represented by the value of the distribution of the energy stored in the form of vector. The vector is trained using Neuro Fuzzy algorithm. In this study, Artificial Neural Networks (ANNs) is used to construct an acoustic model for storing patterns of the speech reference signals and Fuzzy logic is used in the classification process using the membership function. The ANNs architecture in this study can be seen in Fig. 6.

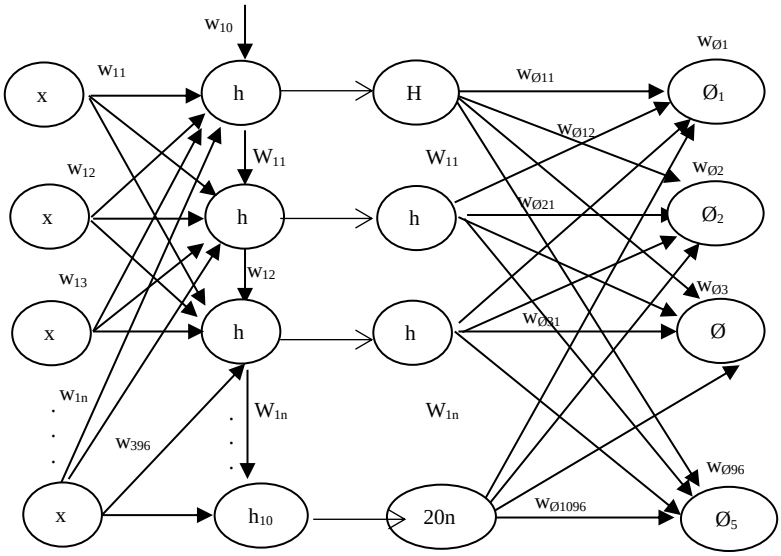

Fig 6 - Architecture of ANNs

The activation function used in this study is a binary sigmoid function with the value range between 0 and 1. The backpropagation method is trained with the Purelin function for identity. To speed up the training process, the Traingdx function and the learning rate of 0.01 were used. The training was conducted until the epoch of 1000 or until it reaches the goal limit of 0.01. To avoid the iterative process of ANNs training trapped into a local minima point, the training used a method of gradient decline by 0.01 and the momentum factor of 0.95. The number of neurons in the input layer of ANNs is 1782, 20 in hidden layer and 4 in output layer. The output from ANNs will be classified using fuzzy logic. Fuzzy logic is chosen because the pattern of CP's speech is irregular and random. In this case, intelligent reasoning is needed to classify the speech patterns of CP into appropriate classification. The design of the membership function in this study is conducted using two forms of membership functions, i.e. triangle and trapezoid. The forms of the membership functions are determined based on the percentage of the spectral energy produced by CP's speech and the spectral energy of the speech contained in the database. The rule of the classification is presented in Table 3 [22].

Table 3 - The Rule of Classification

PERCENTAGE	CLASSIFICATION	THE DEGREE OF CP
> 90	NO CP	No-CP
> 75	Level 1	Mild
> 60	Level 2	Moderate
> 40	Level 3	Severe
< 26	Level 4	Extremely Severe

The number of membership function was labeled in four classes of classification Slight (S), Mild Disorder (MD), Moderate Disorder (MeD), and Severe Disorder (SD) and can be seen in Fig. 7.

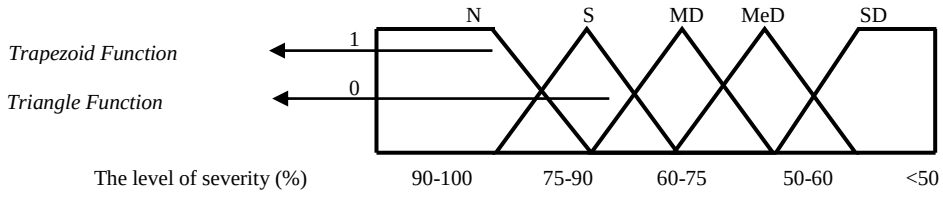

Fig 7 - The membership function for Classification

3. Experimental Result

In testing the ability of VBS using Mel-frequencies Stochastic approach to identify the severity level of CP, the tests are conducted using scenarios as shown in Table 4.

Table 4 - Testing Scenario

COMBINATION	TRAINING SET	TESTING SET
1	Set I	Set II
2	Set II	Set I
3	Set II	Set III
4	Set I	Set III

In this study, an acoustic model was built using the following procedure. First, the speech was recorded from 16 male and 10 female speakers. From this process, there are 3900 speech samples stored in the database. Of this number, 1950 speech data were used in neuro fuzzy for the training phase, and 1950 were used in the testing phase before the system is launched in a real application. The ANNs was trained with a fixed training set until the errors are below the threshold of 0.5. The process will be iterated until all errors are below the threshold, or until the number of epochs reaches 100000. This experiment uses four combinations of data sets as illustrated in Table 4. Set I contains speech data from subjects that fill the reference database. Set II contains speech data from subjects who do not have a history of motor pathology, and Set III consists of the speech data from CP. All data sets were combined to evaluate system performance. The numbers of the dataset are presented in Table 5.

Table 5 - The numbers of Dataset

Dataset	Training Data	Testing Data
Set I	836	351
Set II	233	117
Set III	881	1482

The parameters of the success of the approach used in this study are the accuracy and response time of the system. For each pair of the input-output data, the error, i.e. the difference between the ideal output (the value of the actual post-test) with the output produced by neuro-fuzzy was calculated. Then the mean squared error (MSE) at each end of the iteration/epoch was calculated using the equation as follow:

$$\mathrm{MSE} = \frac{\sum e_i^2}{n} = \frac{\sum (X_i - F_i)^2}{n} \tag{12}$$

Where, MSE is in percentage, and the measurement time is in seconds. The experimental results that have been carried out show the level of accuracy such as presented in Table 6. The comparison of the diagnosis produced by the system with the diagnosis made by the medical doctor based on medical record is shown in Table 7.

Table 6 - Test results of the classification of speech disorders using Mel-frequencies Stochastic Approach

Respondents	Epoch	Time (Seconds)	Performance (MSE)	Gradient	Mu	The similarity of patterns	Identification of speech disorders	Conclusion by VBR
R1	2.5	1.4	0.0050	0.1118	0.0340	86.6396	100	Speech Discrimination
R2	2.7	1.2	0.0116	0.0932	0.0031	86.0819	100	Speech Discrimination
R3	3.0	2.5	0.0075	0.1773	0.0222	91.7885	100	Speech Discrimination
R4	2.6	2.7	0.0347	0.1570	0.0264	87.9540	80	Speech Discrimination
R5	2.4	2.1	0.0041	0.1418	0.0124	86.9303	90	Speech Discrimination
R6	2.2	5	0.0067	0.2681	0.0202	97.5583	100	Normal
R7	2.3	4.5	0.0065	0.2823	0.0004	97.5454	100	Normal
R8	2.1	3.7	0.0099	0.1842	0.0010	97.4177	100	Normal
Average	**2.48**	**2.86**	**0.0226**	**0.1770**	**0,0150**	**90.8644**	**96.25**	**100**

Table 7 - Comparison of Results of identification by VBR with Real Condition of Respondents

Respondent	Identification by System	Real Condition of CP	The level of speech disorders by System	Diagnosis of CP by Physician
R1	Speech Discrimination	CP	Severe Disorder	Severe Disorder
R2	Speech Discrimination	CP	Extremely Severe	Extremely Severe
R3	Speech Discrimination	CP	Extremely Severe	Severe Disorder
R4	Speech Discrimination	CP	Moderate Disorder	Moderate Disorder
R5	Speech Discrimination	CP	Mild Disorder	Mild Disorder
R6	Normal	Normal	Normal	Normal
R7	Normal	Normal	Normal	Normal
R8	Normal	Normal	Normal	Normal

4. Analysis and Discussion

Based on the testing on 8 respondents, the system predicted that the first respondent (R1), R2 and R3 suffered from speech discrimination. In the testing conducted to R1, R2, and R3 using 15 words, the system provides a conclusion of the recognition as follow. The respondents R1, R2, and R3 were unable to imitate the words being tested. In this case, the system cannot recognize the words spoken by the respondents. In the testing conducted on R4, the respondent indicated a speech disorder with a severity level of 80%. This is also seen in R5, where the system predicts that R5 suffered a speech disorder up to 90% of its severity.

In the pattern matching process, the system recognized 3 of 15 words that are tested to R4, whereas the system recognizes 1 word that is considered nearly right than 15 words tested. On R6, R7, and R8, the system recognizes the similarity of words that is uttered by the respondent with the rate of accuracy of 100% rate with the similarity of patterns between 97.41% and 97.55%. The conclusions obtained by VBR to respondents R6, R7, and R8 are they do not have an abnormality of speech. These results can be seen in Table 7 above.

Based on Table 7, VBR's ability to identify the level of speech disorders in CP is reported to have the same conclusion with a diagnosis based on their medical record. Despite there is a difference in the classification of the level of severity of CP, It only 1-degree level. It can be seen in the results of the recognition of the severity of R1 and R3, where the system classified the level of severity of R3 extremely severe while the medical record states that it is at the Severe level.

Refer to the experimental results, the accuracy of VBR using Mel-frequencies approach can be determined using the formula in equation (13).

$$CR = \frac{C}{A}\ x\ 100\% \tag{13}$$

Where CR is the Correct Rate (accuracy), C is the number of samples recognized correctly and A is the number of all samples. Based on this formula, the accuracy of the Mel-frequencies stochastic approach to determine the severity of CP is:

$$CR = \frac{7}{8}\ x\ 100\% = 87.5\%$$

From the calculation, the accuracy of this approach to identify the severity of persons with CP is 87.5%. This percentage shows a very good level of the accuracy to measure the performance of the offered methods which is to be the contribution of this study.

5. Conclusion

This study aims to examine the ability of VBR to identify the severity level that a CP suffered based on their voice. The new approach offered in this study is the Mel-frequencies stochastic approach. In the classification process, this study uses neuro fuzzy. Based on the experimental results, this approach can identify the level of severity of CP with 87.5% accuracy. The stochastic approach with statistical calculation makes this approach easy to compare the results with other approaches. The use of neuro fuzzy to classify the random patterns also makes this approach can be implemented on a stochastic system that cannot be solved with the statistical approach. It is hoped that this approach will be used in real applications to help therapists diagnose the severity level of CP.

6. Future Works

Aside from being an identification tool, a stochastic approach is also expected to be implemented in the augmentative and alternative speech communication aids (AASC) for CP. The AASC using perceptual analysis which is a conventional

method can reduce significantly the word error rate of severe dysarthric speakers [23]. This ability was considered by experts to develop as a therapeutic aid for speech and language disorders. This consideration is based on research which is conducted by Robles-Bykbaev, et al in 2015 [24]. Its hope, by using the stochastic model can improve the performance of AASC. So this approach is not only can be used for identification but furthermore contributes as an approach to therapeutic aids. If this method is combined with the prediction model proposed by Paiman et al in 2018, these combinations are expected to produce a reliable approach as a tool for identifying and therapeutic aids for disabilities, especially CP [25].

References

[1]. Morris, C. (2007). Definition and Classification of Cerebral Palsy: a historical perspective. *Development Medicine and Child Neurology, 109:8-14*, 3-7.

[2]. Pavone, V., & Testa, G. (2015). Classification of Cerebral Palsy. In F. Canavese, & J. Deslandes, *Orthopedic Management of Children with Cerebral palsy: A Comprehensive Approach* (pp. 75-97). Catania, Italy: Nova Science Publishers, Inc.

[3]. Lumsden, D. E. (2018). The child with dystonia. *Pediatrics and Child Health*, pp. 459-467.

[4]. *Apraxia of Speech in Adults.* (2019, April 18). Retrieved from speakeasytherapylv.org: https://www.speakeasytherapylv.org/wp-content/uploads/2017/07/a2b373_8037b2a2b18341a6a572f7d36fbc83ca.pdf

[5]. Kummer, A. W. (2014). Speech and Resonance Disorders Related to Cleft Palate and Velopharyngeal Dysfunction: A Guide to Evaluation and Treatment. *Speech-Language Pathology*, 57-74.

[6]. Pennington, L. (2016). Speech, language, communication, and cerebral palsy. *Developmental Medicine & Child Neurology*, 58: 530–540.

[7]. Geytenbeek, J. (2016). Differentiating between language domains, cognition, and communication in children with cerebral palsy. *Developmental Medicine & Child Neurology*, 58: 530–540.

[8]. Craig, F., Savino, R., & Trabacca, A. (2019). A systematic review of comorbidity between cerebral palsy, autism spectrum disorders and Attention Deficit Hyperactivity Disorder. *European Journal of Pediatric Neurology*, 31-42.

[9]. Elia, A. E., Bagella, C. F., Ferre, F., Zorzi, G., Calandrella, D., & Romito, L. M. (2018). Deep brain stimulation for dystonia due to cerebral palsy: A review. *European Journal of Pediatric Neurology*, 308-315.

[10]. Krigger, K. W. (2006). Cerebral Palsy: An Overview. *American Family Physician*, 91-100

[11]. Hidecker, M., Paneth, N., Rosenbaum, P., & et al. (2011). Developing and validating the Communication Function Classification System for individuals with cerebral palsy. *Development of Medicine & Children Neurology*, 53: 704–10.

[12]. Harvey AR, R. M. (2013). Children with cerebral palsy and periventricular white matter injury: does gestational age affect functional outcome? *Research in Developmental Disabilities*, 34: 2500–06.

[13]. Deller,JR, J. R., & Hsu, D. (1987). An Alternative Adaptive Sequential Regression Algorithm and Its Application to the Recognition of Cerebral Palsy Speech. *IEEE Transactions on Circuits and Systems*, (pp. 782-787).

[14]. Deller,JR, J. R., Hsu, D., Ferrier, L. J. (1988). Encouraging Results in the Automated Recognition of Cerebral Palsy Speech. *IEEE Transactions on Biomedical Engineering*, (pp. 218-220).

[15]. Yusof, Z. M., Hussain, R., & Ahmed, M. (2013). Malay Speech Intelligibility Test (MSIT) for Deaf Malaysian. *International Journal of Integrated Engineering*, 13-19.

[16]. Jaume-i-Capó, A., Martínez-Bueso, P., Moyà-Alcover, B., & Varona, J. (2014). Interactive Rehabilitation System for Improvement of Balance Therapies in People with Cerebral Palsy. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, (pp. 419-427).

[17]. Arnoni, J. L., Pavão, S. L., Silva, F. P., & Rocha, N. A. (2019). Effects of Virtual Reality in Body Oscillation And Motor Performance Of Children With Cerebral Palsy: A Preliminary Randomized Controlled Clinical Trial. *Complementary Therapies in Clinical Practice*, 189-194.

[18]. Polur, P. D., & Miller, G. E. Effect of High-Frequency Spectral Components in Computer Recognition of Dysarthric Speech Based on A Mel-Cepstral Stochastic Model. Journal of Rehabilitation Research & Development, (2005) 363–372.

[19]. Glenn Doman, *Teach Your Baby Read Kit*.: Random House UK, 1987.

[20]. Blaine Goss, "Listening as Information Processing," *Journal of Communication Quarterly*, pp. 304-307, 2009.

[21]. Polur, P. D., & Miller, G. E. (2005). Experiments with Fast Fourier Transform, Linear Predictive and Cepstral Coefficients in Dysarthric Speech Recognition Algorithms Using Hidden Markov Model. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, (pp. 558-561)

[22]. Skurr, B. Audiometri Klinis. Bandung: LAB/UPF THT Fakultas Kedokteran UNPAD/RS DR. Hasan Sadikin, (1993).

[23]. Celin, T. M., Rachel, G. A., Nagarajan, T., & Vijayalakshmi, P. (2019). A Weighted Speaker-Specific Confusion Transducer-Based Augmentative and Alternative Speech Communication Aid for Dysarthric Speakers. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, (pp. 187-197).

[24]. Robles-Bykbaev , V. E., López-Nores, M., Pazos-Arias, J. J., & Arévalo-Lucero, D. (2015). SPELTA: An expert system to generate therapy plans for speech and language disorders. *Expert Systems with Applications*, 7641–7651.

[25]. Paiman, N. A., Hariri, A., Masood, I., Noor, A., Yusof, K. H., Abdullah, S., Leman, A. M. (2018). Development of Neurobehavioral Deterioration Risk Prediction Model for Welder: A Proposed Study. *International Journal of Integrated Engineering*, 122-129.