International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-4, November 2019

Noise and Echo Aware Accurate Dysarthria Speech Recognition Model

Usha.M, L. Sankari

Abstract— ***Dysarthria speech is the speech disorder which would be caused due to weakness of human muscles. The human with dysarthria disorder cannot speak normally whose speech will be very slow and congestive words which might be more difficult to understand. Thus it is required to create the environment where the speech of dysarthria disorder people can be recognized. This is done in our previous research work by introducing the method namely Hidden Markov Model based Speech Recognition (HMMSR). However this research work didn't focus on accurate prediction and the echo noises presence in the speech signals. This would lead to inaccuracy in speech recognition. This is resolved by introducing the efficient dysarthria speech recognition framework namely Noise and Echo aware Dysarthria Speech Recognition Method (NE-DSRM). In this research work, Hybrid Least Mean Square-Adaptive Neuro Fuzzy Inference System (LMS-ANFIS) has been used for preprocessing. This method will remove both echo and noises present in the speech signals to ensure the accurate prediction outcome. And then speech recognition is performed by comparing the dysarthria speech with the phonological speech based on which relevancy would be identified. The accuracy of speech recognition can be improved by introducing the SVM based learning methodology which can classify the dysarthria speech based on which more relevant matching can be done. The objective of the system is, after being trained, to identify and classify limited-vocabulary sets of speaker-dependent. The overall assessment of the research work is done in the matlab simulation environment from which it is proved that the proposed method NE-DSRM tends to have better performance than the existing research works.***

Keywords: Dysarthria speech, phonological speech, noises, echoes, vocabulary classification, accurate speech recognition.

I. INTRODUCTION

Automatic speech recognition found to be more popular research technology focused by various researchers [1]. There are various research methods has been introduced earlier for the successful employment of automatic speech recognition [2]. Those research methods proved their increased accuracy by recognizing the speech of users efficiently [3]. However those research techniques doesn't apply well for the speech from the disordered people [4]. Speech of disordered people will not have clear pronunciation which cannot be predicted by existing automated speech recognition techniques [5]. Thus it is required to focus more on the method implementation to provide the increased support for the disordered people speech recognition outcome [6].

Various disorders affects the normal speech behavior of the peoples. Dysarthria is one of the most found disorder found in all age group of people [7]. The main cause of this disorder is abnormal speech which is very slow in their phonetic frequency that can't understand clearly [8]. This disorder occurs because of weakness found in the muscles. It is required to develop a method that can ensure the automated and accurate recognition of the dysarthria disordered people speech [9].

There are three kinds of automated speech recognition systems are available in nature [10]. Speaker dependent, speaker independent and speaker adaptable. In the speaker dependent recognition system, the classifier will be trained with the corresponding speaker speeches [11]. Thus in the testing phase, accurate recognition of that corresponding user can be guaranteed. The well defined example for speaker dependent speech recognition system is voice based authentication system. Another kind of automated speech recognition system is speaker independent recognition system [12]. This system doesn't focus on single user speech signals. This system will utilize speech signals from the multiple users which will be learned without any prior information. This system enables recognition of different users' speeches which falls under similar regions of trained speech signals. The example of speaker independent recognition system is voice based crime investigation system.

Both of the above discussed kind of automated speak recognition system cannot support the better recognition for the disordered people due to its limitation to certain speaker factors. Another kind of speaker recognition system is adaptive speaker recognition system [13]. This is most similar to speaker independent system, however this kind will learn the existing users' verbal phonetic features as it observes it. This enables system to recognize any kind of voices from the users. This kind of recognition system can be utilized for ensuring the accurate and efficient prediction of speech signals observed from the dysarthria disordered people.

The main goal of this research work is to implement the system that can ensure the accurate and reliable recognition of speech signals collected from the dysarthria disordered people. This is done by introducing the noise and echo aware dysarthria speech recognition system.

Manuscript published on November 30, 2019.
* Correspondence Author
Usha.M*, Research Scholar, Sri Ramakrishna College of Arts and Science for Women, Coimbatore, Tamil Nadu, India.(Email: E-mail:usha.m@kgcas.com)
Dr .L. Sankar Associate Professor, Department of Computer Science, Sri Ramakrishna College of Arts and Science for Women, Coimbatore, Tamil Nadu India. .(Email: sankarivnm@gmail.com)

This method ensures the accurate prediction of speech signals of disordered people.

II. RELATED WORKS

Lowit et al [14] attempted to analyze the various factors that are involving around the dysarthria disorder. This analysis also focuses on the Parkinson disease and attempted to predict the structural variation between symptoms of different disordered people. This analysis has been done based on varying rhythm captured from the disordered people and made the final conclusion.

Vásquez-Correa et al [15] attempted to predict the varying kind of level in dysarthria over the Parkinson disordered people. This analysis provided the outcome of different level of disease which would occur on the disordered people. This is done extracting the different signal features and learning them using FDA scale procedure. This method ensures the accurate and automated prediction of dysarthria people without having need of medical observations.

Narendra et al [16] attempted to predict the dysarthria speech accurately by extracting and analyzing the glottal features from the speech of disordered people. These features will be extracted from the coded telephonic conversation based on which accurate recognition can be ensured. Deep neural network is utilized in this work for the extraction of glottal features from the telephonic coded conversation.

Jong et al [17] attempted to develop an automated speech recognition system for the dysarthria patients whose spoken language was Thai. This is ensured by learning the thai syllables which will be learned accurately and efficiently by adapting the electromyography procedure. This method will first remove the noise features and then syllables will be classified based on different factors. This is achieved by utilizing the spectral regression extreme learning machine which will classify the speech signal of disordered people more accurately.

Kayasith et al [18] introduced the new method to recognize the speech signals of dysarthria disordered peoples with ensured signal quality. This is done by constructing the speech confusion matrix based on which different features and syllables of users can be learned with different time wrapping. This method will quantify the structural variation between two same words spoken in different time period by the same speaker in order to ensure the correct and successful recognition rate.

Kadi et al [19] attempted to introduce the fully automated speech recognition framework for the dysarthria disordered people. This is done by integrating the speech recognition system with the auditory knowledge database, thus the proper and efficient recognition outcome can be ensured. This method utilized the Gaussian mixture model and the support vector machine for the accurate and proper recognition of speech signals of dysarthria disordered people.

Rudzicz et al [20] focused on the articulatory similarity to perform the dysarthria speech recognition process. This is done by introducing the non linear Hammerstein system which will analyse the database of articulatory information to ensure the accurate and reliable speech recognition outcome. The analysis evaluation proved that the proposed research method leads to produce the 95% better outcome in terms of accurate dysarthria speech recognition.

Hawley et al [21] built up a restricted jargon speaker subordinate discourse acknowledgment application which has more prominent resilience to fluctuation of discourse, combined with a mechanized preparing bundle which helps dysarthric speakers to improve the consistency of their vocalizations and gives more information to recogniser preparing.

Rajeswari et al [22] proposed a Generative Model-Driven Feature Learning based discriminative system that maps the arrangement of highlight vectors to fixed measurement vector spaces actuated by the generative models. The discriminative classifier is worked in that vector space. The proposed HMM-based fixed dimensional vector portrayal gives preferred separation to dysarthric discourse over the ordinary HMM.

III. NOISE AND ECHO AWARE DYSARTHRIA SPEECH RECOGNITION

In this research work, Hybrid LMS-ANFIS has been used for avoiding both noise and echo present in input signals. Here, Echo cancellation is done by using LMS techniques and the noise cancellation is done by using ANFIS technique. And then speech recognition is performed by comparing the dysarthria speech with the phonological speech based on which relevancy would be identified. The accuracy of speech recognition can be improved by introducing the SVM based learning methodology which can classify the dysarthria speech based on which more relevant matching can be done.

3.1. HYBRID LMS-ANFIS MODEL FOR NOISE AND ECHO REMOVAL

In this research method, initially preprocessing is done on the dysarthria disordered people speech signals. This preprocessing will remove the noise and echoes present among them, thus the proper and accurate speech recognition outcome can be ensured. Here initially noises from the signals will be removed and then echoes will be avoided. In the following sub section, noise removal and echo removal processing steps were given in detailed.

3.1.1. ANFIS BASED SPEECH SIGNAL NOISE REMOVAL

In this research method, noises present in the speech signals are avoided by using the ANFIS method. Consider noise present in the dysarthria disordered people speech signal SS_{DDP} is referred as $N(SS_{DDP})$. Here value of $N(SS_{DDP})$ will be varied randomly and will differ from speech to speech. Here the signal of disordered people can be represented as like given in equation 1:

$$SS_{DDP} = X(SS_{DDP}) + N(SS_{DDP}) \qquad (1)$$

Where $X(SS_{DDP}) \rightarrow$ original signal of disordered people

Here noises from the observed signal is difficult to predict due to unclear knowledge about the value of distributed noise features. This can be resolved by adapting the non linear function through which noise signal from the observed signal can be learned accurately which can then be removed. The prediction outcome of filtered noise signal from the observed signal can be represented as like given in equation 2:

$$Y(SS_{DDP}) = N(SS_{DDP}) - Est(N(SS_{DDP})) \quad (2)$$

Here accurate prediction outcome can be ensured by learning the different combination signal features iteratively. This is done in this research work by using ANFIS method. Generally ANFIS is combination of neural network and fuzzy logic system by combining the different learning capabilities to make sure the accurate prediction outcome.

Consider fuzzy rules with two input values x and y and the output value z. By applying the sugeno model, fuzzy rules can be generated as like given below

Inference 1: If x is A and y is B then f = px+qy+r

Inference 2: If x is C and y is D then g = sx+ty+u

And

$$f = w_1 f + w_2 g \quad (3)$$

Where w1, w2 ➔ weight values

By adapting the above equation estimated observed signal value can be predicted more accurately. This method ensures the accurate and correct prediction of noises present in the observed speech signal which can be eliminated accurately.

3.1.1. LMS BASED ECHO REMOVAL IN SPEECH SIGNALS

After noise removal, echo removal is performed, thus the accurate and reliable speech recognition outcome can be guaranteed. This is done by introducing the adaptive filter method namely LMS. This method will work in different iteration based on which accurate echo filtering can be guaranteed. The main goal of this research is to eliminate the echo from the observed signals by finding the variation between the required signal and the adaptive filter outcome. This variation will be considered as error rate which will be utilized in the next iteration to adjust the cost value of learning model. Here $X(SS_{DDP})$ is observed input speech signal, $Y(SS_{DDP})$ is adaptive filter outcome, $D(SS_{DDP})$ is required echo signal value, and $E(SS_{DDP})$ is error rate of signals. In this work, initially weight value of LMS filter will be fixed as zero as like given in equation 4:

$$W(SS_{DDP}) = 0 \quad (4)$$

This weight value will be updated in different iteration to ensure the accurate prediction and removal of echo signal features. The processing steps of LMS algorithm is given below:

1. Solution of adaptive filter is as follows

$$Y(SS_{DDP}) = W^T (SS_{DDP})\, X(SS_{DDP}) \quad (5)$$

2. Error estimation based on previously predicted outcome is given below:

$$E(SS_{DDP}) = D(SS_{DDP}) - Y(SS_{DDP}) \quad (6)$$

3. Weight updation based on this error and final outcome is given below:

$$W(SS_{DDP}+1) = W(SS_{DDP}) \pm \mu\, E(SS_{DDP})\, X(SS_{DDP}) \quad (7)$$

Where μ is step size which is varying between $0 \leq \mu \leq 1$

The above mentioned processing steps will predict the echo signal from the speech signal which will be removed completely. Thus accurate and reliable speech recognition outcome can be ensured.

3.2. ACCURATE SPEECH RECOGNITION USING SUPPORT VECTOR MACHINE

An AI system which depends on the rule of structure hazard minimization is bolster vector machines. It has various applications in the zone of example acknowledgment. SVM develops direct model dependent on help vectors so as to appraise choice capacity. In the event that the preparation information are straightly distinct, at that point SVM finds the ideal hyper plane that isolates the information without blunder. Fig. 1 demonstrates a case of a non-straight mapping of SVM to develop an ideal hyper plane of partition. SVM maps the info designs through a non-direct mapping into higher measurement highlight space. For straightly distinguishable information, a direct SVM is utilized to order the informational indexes. The examples lying on the edges which are expanded are the support vectors.

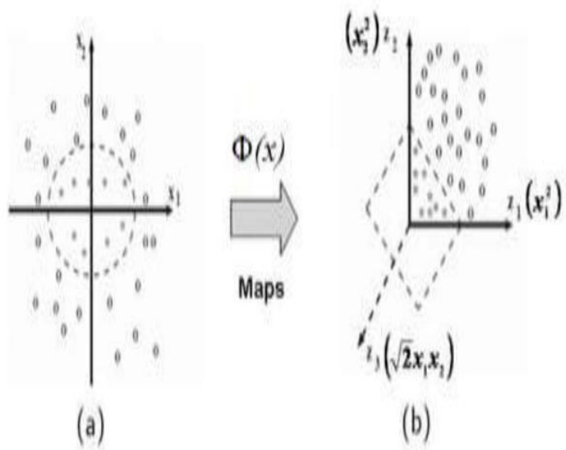

Fig - 1: Example for SVM Kernel Function Φ(x) Maps 2-Dimensional Input Space to Higher 3-Dimensional Feature Space. (a) Nonlinear Problem. (b) Linear Problem.

The help vectors are the (changed) preparing designs and are similarly near hyperplane of division. The help vectors are the preparation tests that characterize the ideal hyperplane and are the most troublesome examples to arrange. Casually, they are the examples most enlightening of the order task. The bit capacity creates the internal items to develop machines with various kinds of non-direct choice surfaces in the info space.

IV. RESULTS AND DISCUSSION

The overall evaluation of the research work is done in the matlab simulation environment. Here the performance assessment of research work is carried out between current work Noise and Echo aware Dysarthria Speech Recognition Method (NE-DSRM) in which SVM is utilized for the speech recognition outcome. The performance improvement is assessed by comparing the current work NE-DSRM with the existing methodologies K Nearest Neighbour based Dysarthria speech recognition Method (KNN-DSRM) and the Naïve Bayes based Dysarthria Speech recognition method (NB-DSRM). The metrics that are considered for the performance evaluation are accuracy, precision, recall, sensitivity, specificity, RMSE and Pearson correlation. The simulation outcome obtained in this research work are given in the below figures.

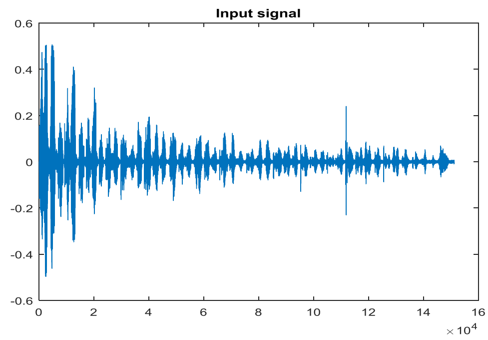

Figure 2.a. Input signal

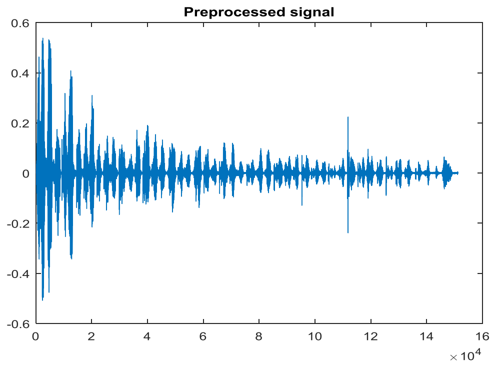

Figure 2.b. Preprocessed signal

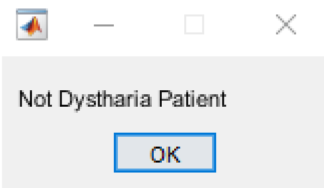

Figure 2c. Prediction outcome

Figure 2. Prediction simulation results for non dysarthria people speech

Figure 2 shows the prediction outcome for the speech signal gathered from the people without dysarthria disorder. Here figure 2 a shows the input signal which will preprocessed and displayed in figure 2b. The final prediction outcome is displayed in the figure 2 c.

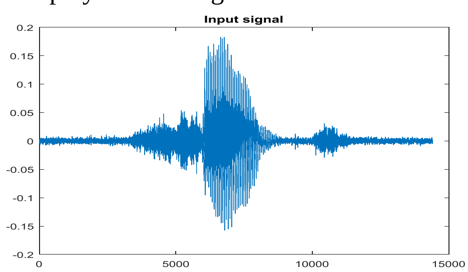

Figure 3.a. Input signal

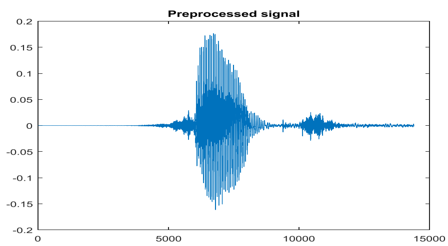

Figure 3.b. Preprocessed signal

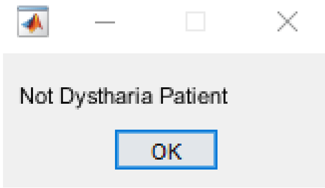

Figure 3.c. Prediction outcome

Figure 3. Prediction simulation results for non dysarthria people speech

Figure 3 shows the prediction outcome for the speech signal gathered from the people without dysarthria disorder. Here figure 3 a shows the input signal which will preprocessed and displayed in figure 3b. The final prediction outcome is displayed in the figure 3 c.

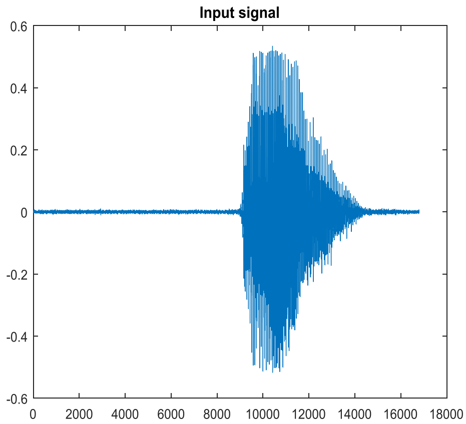

Figure 4a. Input signal

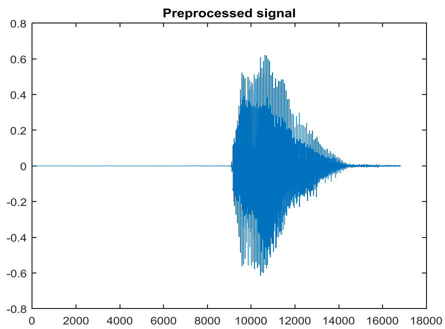

Figure 4b.Preprocessed signal

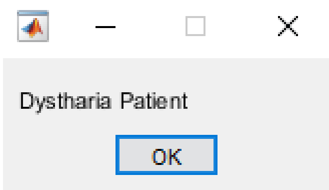

Figure 4c. Prediction outcome

Figure 4. Prediction result for the speech signal gathered from the people with dysarthria disorder

Figure 4 shows the prediction outcome for the speech signal gathered from the people with dysarthria disorder. Here figure 4a shows the input signal which will preprocessed and displayed in figure 4b. The final prediction outcome is displayed in the figure 4c.

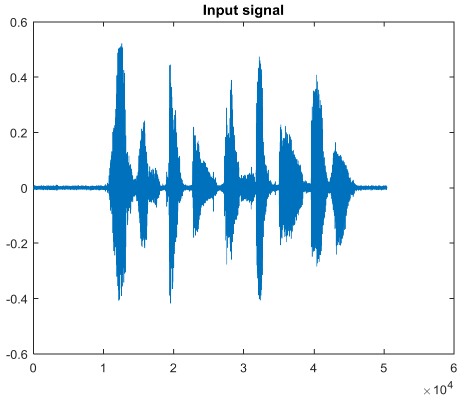

Figure 5a. Input signal

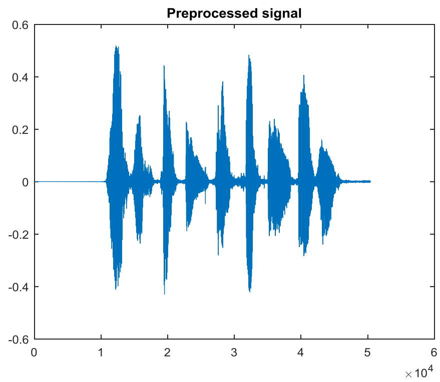

Figure 5b.Preprocessed signal

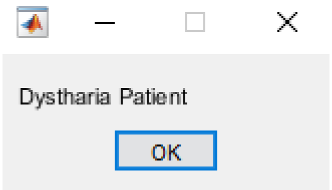

Figure 5c. Prediction outcome

Figure 5. Prediction result for the speech signal gathered from the people with dysarthria disorder

Figure 5 shows the prediction outcome for the speech signal gathered from the people with dysarthria disorder. Here figure 5a shows the input signal which will preprocessed and displayed in figure 5b. The final prediction outcome is displayed in the figure 5c.

The numerical comparison evaluation of the proposed and existing methodologies is done based on different performance metrics. The simulation numerical values are given in the following table:

Metrics	Methods		
	KNN-DRM	NB-DSRM	NE-DSRM
Accuracy	85	91	94.57
Precision	38.65	41.2	48.45
Recall	39.69	42.6	48.75
Sensitivity	39.69	42.6	48.75
Specificity	39.69	42.6	48.75
RMSE	9.25	7.89	4.86

The graphical representations of these values are given in the following figure 4.

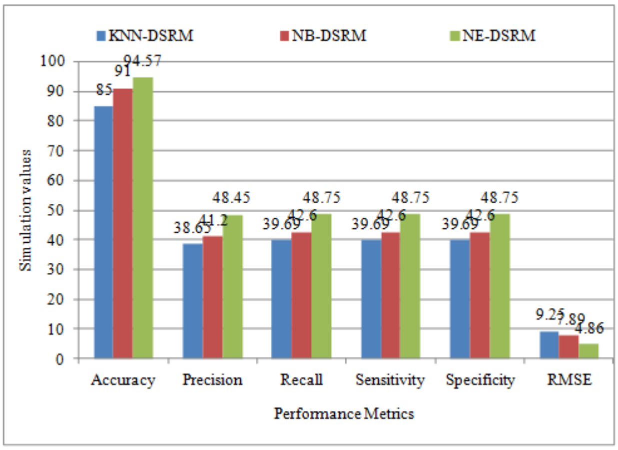

Figure 4. Simulation comparison values

In figure 4, comparison evaluation of the proposed and existing methods in terms of different performance metrics has been given. From this evaluation it is confirmed that the proposed method tends to have better performance than the existing methods. In terms of accuracy, NE-DSRM attains 3.92% increased accuracy than NB-DSRM and 11.25% increased accuracy than KNN-DSRM. In terms of Precision, NE-DSRM attains 17.59% better precision than NB-DSRM and 25.35% better precision than KNN-DSRM. For Recall, sensitivity and specificity, it shows similar performance improvement where NE-DSRM shows 14.43% better performance than NB-DSRM and 22.82% better performance than KNN-DSRM. In terms of RMSE, NE-DSRM shows 38.4% lesser RMSE than NB-DSRM and 47.45% lesser RMSE than KNN-DSRM.

V. CONCLUSION

In this research work, Hybrid LMS-ANFIS has been used for preprocessing. This method will remove both echo and noises present in the speech signals to ensure the accurate prediction outcome. And then speech recognition is performed by comparing the dysarthria speech with the phonological speech based on which relevancy would be identified. The accuracy of speech recognition can be improved by introducing the SVM based learning methodology which can classify the dysarthria speech based on which more relevant matching can be done. The objective of the system is, after being trained, to identify and classify limited-vocabulary sets of speaker-dependent. The overall assessment of the research work is done in the matlab simulation environment from which it is proved that the proposed method NR-DSRM tends to have better performance than the existing research works.

REFERENCES

1. Yu, D., & Deng, L. (2016). AUTOMATIC SPEECH RECOGNITION. Springer london limited.
2. Gajic, B., Narayanan, S. S., Parthasarathy, S., Rose, R. C., & Rosenberg, A. E. (2015). U.S. Patent No. 9,058,810. Washington, DC: U.S. Patent and Trademark Office.
3. Bellegarda, J. R. (2016). U.S. Patent No. 9,431,006. Washington, DC: U.S. Patent and Trademark Office.

4. Liu, Y., Qin, Y., Feng, S., Lee, T., & Ching, P. C. (2018, November). Disordered Speech Assessment Using Kullback-Leibler Divergence Features with Multi-Task Acoustic Modeling. In 2018 11th International Symposium on Chinese Spoken Language Processing (ISCSLP) (pp. 61-65). IEEE.
5. Gonçalves, C., Rocha, T., Reis, A., & Barroso, J. (2017, April). AppVox: An application to assist people with speech impairments in their speech therapy sessions. In World Conference on Information Systems and Technologies (pp. 581-591). Springer, Cham.
6. Christensen, H., Nicolao, M., Cunningham, S., Deena, S., Green, P., & Hain, T. (2015). Speech-enabled environmental control in an AAL setting for people with speech disorders: a case study.
7. Pennington, L., Parker, N. K., Kelly, H., & Miller, N. (2016). Speech therapy for children with dysarthria acquired before three years of age. Cochrane Database of Systematic Reviews, (7).
8. Lowit, A., & Kent, R. D. (2016). Management of dysarthria.
9. Alhinti, L., Christensen, H., & Cunningham, S. (2018, July). Towards the understanding of communicating emotions for people with dysarthria. In Proceedings of WSPD 2018. Sheffield.
10. Yeracaris, Y., Carus, A. B., & Lapshina, L. (2016). U.S. Patent No. 9,245,525. Washington, DC: U.S. Patent and Trademark Office.
11. Dahake, P. P., Shaw, K., & Malathi, P. (2016, September). Speaker dependent speech emotion recognition using MFCC and Support Vector Machine. In 2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT) (pp. 1080-1084). IEEE.
12. Ahmad, K. S., Thosar, A. S., Nirmal, J. H., & Pande, V. S. (2015, January). A unique approach in text independent speaker recognition using MFCC feature sets and probabilistic neural network. In 2015 Eighth International Conference on Advances in Pattern Recognition (ICAPR) (pp. 1-6). IEEE.
13. Naik, D. K., & Tackin, O. E. (2017). U.S. Patent No. 9,697,822. Washington, DC: U.S. Patent and Trademark Office.
14. Lowit, A., Marchetti, A., Corson, S., & Kuschmann, A. (2018). Rhythmic performance in hypokinetic dysarthria: Relationship between reading, spontaneous speech and diadochokinetic tasks. Journal of communication disorders, 72, 26.
15. Vásquez-Correa, J. C., Orozco-Arroyave, J. R., Bocklet, T., & Nöth, E. (2018). Towards an automatic evaluation of the dysarthria level of patients with Parkinson's disease. Journal of communication disorders, 76, 21-36.
16. Narendra, N. P., & Alku, P. (2019). Dysarthric speech classification from coded telephone speech using glottal features. Speech Communication, 110, 47-55.
17. Jong, N. S., & Phukpattaranont, P. (2019). A speech recognition system based on electromyography for the rehabilitation of dysarthric patients: A Thai syllable study. Biocybernetics and Biomedical Engineering, 39(1), 234-245.
18. Kayasith, P., & Theeramunkong, T. (2009). Speech confusion index (Φ): A confusion-based speech quality indicator and recognition rate prediction for dysarthria. Computers & Mathematics with Applications, 58(8), 1534-1549.
19. Kadi, K. L., Selouani, S. A., Boudraa, B., & Boudraa, M. (2016). Fully automated speaker identification and intelligibility assessment in dysarthria disease using auditory knowledge. Biocybernetics and Biomedical Engineering, 36(1), 233-247.
20. Rudzicz, F. (2012). Using articulatory likelihoods in the recognition of dysarthric speech. Speech Communication, 54(3), 430-444.
21. Hawley, M. S., Enderby, P., Green, P., Cunningham, S., Brownsell, S., Carmichael, J., ... & Palmer, R. (2007). A speech-controlled environmental control system for people with severe dysarthria. Medical Engineering & Physics, 29(5), 586-593.
22. Rajeswari, N., & Chandrakala, S. (2016). Generative model-driven feature learning for dysarthric speech recognition. Biocybernetics and Biomedical Engineering, 36(4), 553-56.

Retrieval Number: D5333118419/2019©BEIESP
DOI:10.35940/ijrte.D5333.118419
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication