A Review on Dysarthric Speech Recognition

Megha Rughani
Department of Electronics and Communication, Marwadi Educational Institutions Foundation, Rajkot
Email: megharughani@gmail.com
D. Shivakrishna
Department of Electronics and Communication, Marwadi Educational Institutions Foundation, Rajkot
Email: d.shivakrishna@marwadieducation.edu.in

----------ABSTRACT----------

Dysarthria is malfunctioning of motor speech caused by faintness in the human nervous system. It is characterized by the slurred speech along with physical impairment which restricts their communication and creates the lack of confidence and affects the lifestyle. Speech Assistive technology (SAT) developed till yet have been reviewed for dysarthric speech in this paper. We present a study and literal comparison of the techniques for the acoustic modeling like Hidden Markov Model (HMM), Artificial Neural Network (ANN), and hybridization approach adapted by various researchers. With the help of this comparison we can conclude that hybridization of Hidden Markov Model and Multi-Layer Perceptron whose solution is optimized with Genetic Algorithm gives average recognition rate of about 93.5% over other considered techniques.

Keywords –Dysarthric Speech**,** HMM, ANN, Multi-Layer Perceptron (MLP), genetic Algorithm (GA)

I. INTRODUCTION

Speech is the most significant and general way of interaction among the society but it can be interrupted by various types of physical impairments like deafness or weakness of motor speech control. Dysarthria is caused due to reduced control of neuro-motor muscles. It results into slurred speech as articulation is mainly affected. Insertion, deletion and repetition of phoneme reduce the intelligibility of speech signal. Severity of dysarthric speech affects the intelligibility of speech. [1, 2]

It is caused due to brain tumor, celebral palsy, Parkinson diseases, head injury and many more. It lessens the controlling portion of brain which is involved in planning, execution and controlling of the specific affected organ along with motor speech disorder. Lungs, larynx, vocal tract movement, lip movement are basically affected. [1, 2]

Various clinical treatments including exercise of motor muscles were carried out to increase the strength in order to improve articulation, phonation, and resonance. Special therapy like principles of motor learning are carried out by speech language pathologist but these are very time consuming and tedious to be followed. Assistive technology helps in recognition and synthesis of unintelligible speech into intelligible form. [3, 4]

The purpose of this article is to compare the current trends in dysarthric speech recognition. Section 2 gives the overview of techniques. Section 3 describes the different approaches adopted by the author. Results and conclusion are mentioned in section 4 and section 5 respectively.

2. OVERVIEW

The input to dysarthric speech recognition system is acoustic speech waveform and main aim is to detect the correct word uttered i.e. to estimate the word sequence. Basic block

diagram in figure 1 shows basic techniques to be followed. Spectral and cepstral features are extracted from the input raw speech data which are modeled by various classifiers and looked up into dictionary to find similar match and accordingly generates the text output.

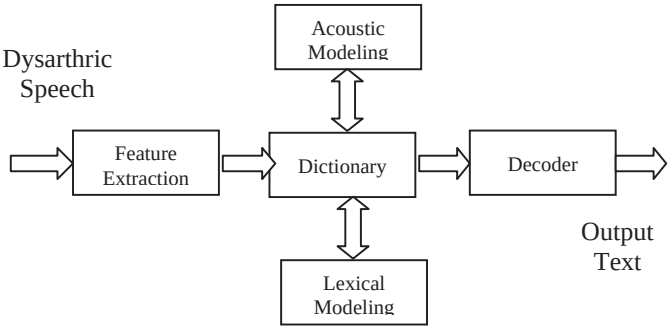

Fig. 1 Basic block diagram of Dysarthric Speech Recognition

2.1 Feature Extraction

Firstly dysarthric speech data is sampled at frequency of 16 KHz which is then windowed with hamming window usually having window size of 15-20ms. Input signal is transformed into acoustic features which can again be reconstructed into original format. There are many feature

extraction techniques out of which MFCC, RASTA PLP and LOG RASTA PLP techniques are widely used.

2.1.1 Mel Frequency Cepstrum Coefficients (MFCC)

It is widely used technique which is obtained by perform following steps successively on each window 1) Discrete Fourier Transform 2) Mel frequency warping on Mel scale 3) logarithm of Mel frequency wrapped features 4) Discrete Cosine Transform results to MFCC parameters.[5]

2.1.2 Perceptual Linear Predictive and Log Perceptive Linear Predictive Coefficients

It is advancing version of Linear predictive coefficients (LPC) technique which emphasis the psychophysically based transformation. It remaps spectral features to bark scale as in contrast to Mel scale in MFCC to enhance middle hearing frequency range. Cube root approximation of loudness mimics the power law of hearing. In LOG PLP spectral components are passed through band pass filter after taking logarithm on it in order to suppress additive distortion. [6]

2.2 Acoustic Modeling

It is the key technique of transforming acoustic signal to text form by usage of statistics. Various current techniques in use like HMM, ANN and MLP which are describe bellow

2.2.1 Hidden Markov Modeling

It is two layer stochastic process implementing markov assumption and Bayes theorem in which first layer of stochastic process is hidden which can be rendered through observation sequence obtained by second layer stochastic process. It consists of 3 problems which are: problem 1: Observation sequence O $\{O_1, O_2 \ldots O_T\}$ and model probabilities $\lambda = \{A, B, \pi\}$ how probability of the observation $Pr(O/\lambda)$. Problem 2: Decision of optimal state sequence I $\{i_1, i_2 \ldots, i_T\}$ on having observation sequence O $\{O_1, O_2 \ldots, O_T\}$. Problem 3: adjusting model parameters to maximize $Pr(O/\lambda)$. All the three problems can be solved by the use of Forward Backward algorithm, Viterbi algorithm and Baum-Welch algorithm. [7]

2.2.2 Artificial Neural Network

It is a parallel computing approach based on biological neural network consisting of hidden layer(s) between input and output layer. Weighted sum of all the input is compared with predefined threshold and output goes to one on crossing threshold else goes to zero. Its architecture is classified in feedforward and feedback (recurrent) architecture which yields to various other models of interconnection between input and output. [8]

2.2.3 Multilayer Perceptron

It is feed forward class of artificial neural network for mapping inputs to appropriate outputs. It consists of input layer, multiple hidden layers and output layer. Backpropagation technique is used during learning phase and it makes this able to form complex decision boundaries. Let $W_{ij}^{(l)}$ be the weight of i^{th} input layer to j^{th} output layer for layer (l-1) .training pattern set be denoted as $\{(x^{(1)}, d^{(1)}), (x^{(2)}, d^{(2)}) \ldots ,(x^{(p)}, d^{(p)})\}$ with assumption $x^{(l)} \epsilon R^n$ and $d^{(l)} \epsilon [0,1]^m$, m dimensional hypercube, than error cost function is defined as in equation 1. [8]

$$E = 0.5 * \sum_{i-1}^{p} \|y(i) - d(i)\|^2 \quad (1)$$

2.2.4 Use of Genetic Algorithm and Metamodels

It is a type of evolutionary algorithm and to optimization of the generated results by mimicking the process of natural selection. Genetic representation and fitness function of solution domain is required. Following steps are required: Initialization, Selection, Genetic Operators i.e. Crossover and Mutation and the termination. [9,10] In speech processing GA is carried out by preparing confusion matrix of input and output of phonemes/words.[11] Extended version of metamodels are used for modeling confusion matrix.[11]

3. CURRENT APPROACHES

There have been many approaches for increasing the intelligibility of dysarthric speech some of which are discussed below.

3.1 Adaptive Based Approach

In this approach acoustic features of the speaker gets adapted slowly as the number of utterance increases. It is somewhat time consuming but recognition rate increases as speaker gets used to this techniques. Caballero-Morales 2014[11] presented speaker adapted technique by varying the effect of prior probability through metamodels and achieved multiple responses which are converted in single phoneme by using genetic algorithm (GA). Word recognition accuracy (WRA) of about 68% is achieved on numerous database. Frank Rudzicz [12] converted acoustic features to articulatory features through nonlinear Hammerstein system with MFCC feature extractor and HMM and DBN as the base system. Harsh Sharma [13] proposed technique based on interpolation on PLP feature vectors adapted to BI MAP (Backward Interpolated MAP) methodology using UA speech database. WRA varies widely with severity of dysarthria. Apriori algorithm is used by HSIEN WU [14] to create personalized dictionary and

gained error reduction of 3.8%.Caballero-Morales 2013[15] compared Bakis and Ergodic topology for HMM using Gaussian continuous parameters with MFCC feature extraction technique followed by GA and found that Bakis topology improves WRA as compared to Ergodic. Sharma [16] has compared speaker dependent (SD) and speaker adaptive (SA) approach and found that SA approach having adaptation of all parameters except transition probability performs well on UA speech database.

3.2 ANN and MLP Based Approach

In this methodology artificial intelligence is used to improve WRA. Neural network corresponding to human brain system is implemented. Seyed (2014) [11] has conceptualized that only 12 coefficients of MFCC for ANN based on isolated word recognition with speaker independent approach performs better than other MFCC coefficients along with its delta & acceleration coefficients. MLP is an emerging technique and it has made many efforts that have significant

Table 1 Comparison of Feature Extraction and Acoustic Modeling Technique

Author	Year	Feature Extraction Technique	Acoustic Modeling Technique	Database Used	Recognition Accuracy
Sharma [13]	2013	PLP	BI MAP HMM	UA Speech	4%-82%
Sharma [16]	2010	PLP	MAP adaptation of all HMM parameters except transition probability	UA Speech	4.2%-66.7%
Caballero-Morales [15]	2013	MFCC	Bakis Topology of HMM + GA	Nemours speech	50% - 80%
Caballero-Morales [11]	2014	Not Mentioned	HMM + GA + metamodels	Nemours speech	42.6% - 77.17%
Shahamiri [17]	2014	MFCC	ANN	UA Speech	57.14% - 68.38%
Lilia Lazli [20]	2011	Log Rasta PLP	Hybrid HMM/MLP	Recorded with 300 speakers	90% (avg)
Joel Pinto [19]	2010		2 Layer MLP	TIMIT and CTS	TIMIT 71.6% (avg) CTS 63.3% (avg)
Lilia Lazli [18]	2013	Log Rasta PLP	HMM + MLP + GA	3 different database with varying number of speaker	93.5% (avg)

Note: (avg) indicates average accuracy and rest are at its minimum to maximum accuracy range.

contribution in increasing the recognition rate. Multilayer perceptron is a class of artificial network. Its application is ranging from signal processing to stock market. Lilia Lazli [18] compared hybrid ANN classifier consisting of multi-network RBF/LVQ structure with hybrid HMM/MLP. Log Rasta PLP and K-means algorithm are used for feature extraction and vector quantization respectively. Results show that hybrid HMM/MLP outperforms. Joel Pinto [19] has presented the purpose of classifier having 2 multilayer perceptron consisting of first layer trained to acoustic feature vector (PLP) with frame length of 90ms and second layer trained to first layer's posterior probability of each phoneme with frame length of about 150 to 230ms using TIMIT and CTS speech database. Result presents that 3.5% and 9.3% improvement for TIMIT and CTS database respectively over single MLP is obtained. Lilia Lazli [20] compared 5 hybrid techniques-: RBF/LVQ, Discrete HMM, and Hybrid

HMM/MLP with KM entries, Hybrid HMM/MLP with FCM entries and Hybrid HMM/MLP with AG entries with LOG RASTA and J-RASTA feature extraction method and showed that HMM/MLP/GA is having better performance among all presented models.

4. RESULTS AND DISCUSSION

Various techniques adopted are investigated and their results are summarized in Table 1. Evaluation of the techniques is

made on the basis of Word Recognition Accuracy (WRA). It shows that implementation of sole technique does not helps in increasing WRA but hybridization shows significant improvement. MLP technique is found to be best among all the techniques described.

V. CONCLUSION

This paper describes the speech assistive methodology for dysarthric speaker. Here we discovered that hybridization of HMM and MLP along with GA outperforms all the compared techniques for restricted number of words due to its flexible architecture and output states are trained to minimize the discrimination between correct and rival classes.

REFERENCES

[1] O'Sullivan, S. B.; Schmitz, T. J., *Physical rehabilitation* (5th ed.), (Philadelphia: F. A. Davis Company, 2007).

[2] Duffy, Joseph, *Motor speech disorders: substrates, differential diagnosis, and management*. (St. Louis, Mo: Elsevier Mosby, 2005).

[3] Fox, Cynthia; Ramig, Lorraine; Ciucci, Michelle; Sapir, Shimon; McFarland, David; Farley, Becky: Neural Plasticity-Principled Approach to Treating Individuals with Parkinson Disease and Other Neurological Disorders, *Seminars in Speech and Language 27* (4): 283–99. Doi: 10.1055/s-2006-955118.

[4] The National Collaborating Centre for Chronic Conditions, ed., "*Other key interventions*". Parkinson's Disease. London: Royal College of Physicians. pp. 135–46, 2006.

[5] Hamzah, R., Jamil, N. & Seman, N., Filled Pause Classification Using Energy-Boosted Mel-frequency Cepstrum Coefficients (MFCC). *The 8th International Conference on Robotic, Vision, Signal Processing & Power Applications*, Penang, Malaysia, November 2013, 10-12.

[6] David Burton, Available at http://staffhome.ecm.uwa.edu.au/~00014742/research/speech/local/entropic/ESPSDoc/appnotes/rasta-plp.pdf

[7] L. R. Rabiner, B. H. Juang, "An Introduction to Hidden Markov Model", *IEEE ASSP Magazine*, January 1986.

[8] Ani1 K. Jain, Jianchang Mao, K.M. Mohiuddin, "Artificial Neural Networks: A Tutorial", *IEEE Computer Society*, March 1996.

[9] Mitchell, Melanie. An Introduction to Genetic Algorithms. Cambridge, MA: *MIT Press*. ISBN 9780585030944, 1996.

[10] Whitley, Darrell, A genetic algorithm tutorial, *Statistics and Computing 4* (2), 1994, 65–85.

[11] Santiago Omar Caballero Morales, Felipe Trujillo-Romero, Evolutionary approach for integration of multiple pronunciation patterns for enhancement of dysarthric speech recognition, *Expert Syst. Appl. 41*(3), 2014, 841-852

[12] Frank Rudzicz, Using articulatory likelihoods in the recognition of dysarthric speech, *Speech Communication 54*(3), 2012, 430-444.

[13] Harsh Vardhan Sharma, Mark Hasegawa-Johnson, Acoustic model adaptation using in-domain background models for dysarthric speech recognition, *Computer Speech & Language 27*(6), 2013, 1147-1162.

[14] CHUNG-HSIEN WU, HUNG-YU SU, and HAN-PING SHEN, Articulation-Disordered Speech Recognition Using Speaker-Adaptive Acoustic Models and Personalized Articulation Patterns, *ACM Transactions on Asian Language Information Processing*, *10*(2), Article 7, June 2011.

[15] Santiago-Omar Caballero-Morales, Estimation of Phoneme-Specific HMM Topologies for the Automatic Recognition of Dysarthric Speech, *Hindawi Publishing Corporation Computational and Mathematical Methods in Medicine Volume 2013*, Article ID 297860

[16] Harsh Vardhan Sharma, Mark Hasegawa-Johnson, State-Transition Interpolation and MAP Adaptation for HMM-based Dysarthric Speech Recognition,

Proceedings of the NAACL HLT 2010 Workshop on Speech and Language Processing for Assistive Technologies, Los Angeles, California, June 2010, 72-79.

[17] Seyed Reza Shahamiri, Siti Salwah Binti Salim, Artificial neural networks as speech recognisers for dysarthric speech: Identifying the best-performing set of MFCC parameters and studying a speaker-independent approach, *Advanced Engineering Informatics. 28*(1), 2014, 102-110.

[18] Lilia Lazli1, Mounir Boukadoum, Hidden Neural Network for Complex Pattern Recognition: A Comparison Study with Multi- Neural Network Based Approach, *International Journal of Life Science and Medical Research*, *3*(6), 2013, 234-245.

[19] Joel Pinto, G.S.V.S. Sivaram, Mathew Magimai Doss, Analysis of MLP Based Hierarchical Phoneme Posterior Probability Estimator, *IEEE AUDIO, SPEECH AND LANGUAGE PROCESSING*, 2010.

[20] Lilia Lazli, Boukadoum Mounir, Abdennasser Chebira, Kurosh Madani and Mohamed Tayeb Laskri, APPLICATION FOR SPEECH RECOGNITION AND MEDICAL DIAGNOSIS, *International Journal of Digital Information and Wireless Communications (IJDIWC) 1*(1): 14-31. The Society of Digital Information and Wireless Communications, 2011 (ISSN 2225-658X)