Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 6.017

IJCSMC, Vol. 6, Issue. 11, November 2017, pg.83 – 85

ASSESSMENT OF DYSARTHRIC SPEECH USING MFCC

Usha.M
Assistant Professor & Head, Department of Computer Applications, K G College of Arts & Science, Coimbatore – 641035, Tamilnadu
usha.m@kgcas.com

Abstract— Speech is the effective form of communication between human and its environment. Dysarthria is a motor speech disorder in which the person lacks the control over articulators used for speech production. Speech accuracy is the outcome of well-timed and coordinated activities of the articulators and other related neuro muscular feature. In this paper, Speech utterance is converted into a phone sequence and histograms of the pronunciation mappings are done by using Mel-frequency cepstral coefficients. Structured sparse feature selection is done using Hidden Markov Models. Prediction is done using Inverse Mel-frequency cepstral coefficients. It is a comparitive study of different methodologies to improve the speech of dysarthric disabled people.

Keywords— Dysarthria, Sparse feature selection, MFCC, Hidden Markov Models

I. INTRODUCTION

Dysarthria is a neurological impairment that damages the control of the motor speech articulators, which the malfunction is caused by the lack of control over the speech-related muscles, the lack of coordination among them, or their paralysis. It is often associated with irregular phonation and amplitude. As a result of the impairment, the speech signal is compromised and its intelligibility is reduced. low intelligibility is one of the most detrimental social characteristics of dysarthria that affects different aspects of the lives of people with such disability.

Automatic Speech Recognition(ASR) systems recognize the uttered word represented as an acoustic signal and rely on a given lexicon to recognise the spoken word(s). They have several applications in health care, the military, telephony, and other domains. They can be very supportive for speakers with dysarthria, because the disabled persons are regularly tangibly incapacitated and unable to use keyboards.

In order for an ASR system to be practicable, acoustic features of utterances must be presented to the system using a process called Feature Extraction. The usage of Mel-Frequency Cepstral Coefficients (MFCCs) is the most public feature-extraction method in ASR applications, which represent speech signals in cepstral domain. It is a illustration defined as the real cepstrum of a windowed short-time signal derived from the Fast Fourier Transform of that signal, in which the frequency bands are spaced on the mel scale similarly (inspired by the human auditory perception system). MFCCs have been widely used for several speech-disorder signal-processing tasks such as speech disability grouping and dysarthric speech recognition. The MFCCs are usually presented as mel cepstrum with 12 coefficients, their first and second derivatives,

II. PREVIOUS WORK

As an illustration, Hasegawa-Johnson et al.[1] delivered two isolated-word SD ASR systems (10-digit vocabulary) based on the data composed from three subjects with dysarthria: one female and two males with one control subject. The speech examples were recorded using an array of seven microphones and four cameras mounted on top of a computer monitor. The first system was a phone-based Hidden Markov Model(HMM), and the second was a fixed-length isolated-word ASR system based on Support Vector Machines (SVMs).

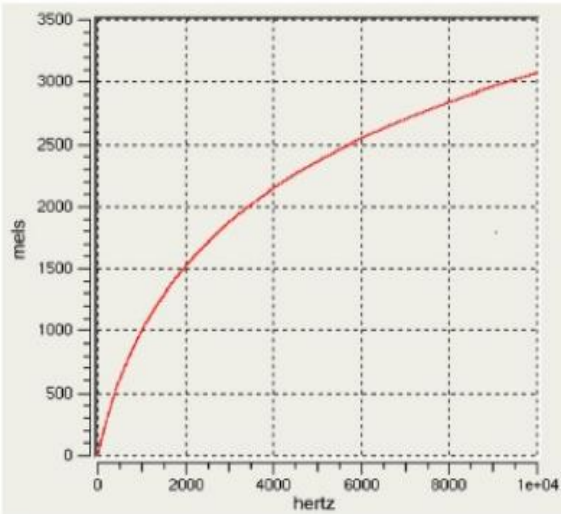
Selouani et al.[2] proposed another SD ASR based on HMMs for English and French speakers with dysarthria for continuous speech. The ASR system was trained using speech materials collected from four dysarthric speakers in the Nemours database and one control speaker; A feature representation method created on alignment with the decoded (spoken) phone sequence resulting from an ASR system and canonical phone structure from a standard pronunciation dictionary. To this end, a weighted finite state transducer (WFST) [3] is employed as an alignment tool to capture phone-level mappings including match, substitution, and deletion. A WFST is a exact powerful and flexible framework in mapping input and output symbols, and therefore, it has been utilized in various speech and language processing applications such as machine translation, pronunciation modeling, and effi cient speech recognition [4]–[7].

III.METHODS

The database contains isolated-words, acoustic samples of digits, radio alphabet letters, computer commands, and common words acquired from 19 male and female subjects with dysarthria of different severity levels, varying from extremely low speech intelligibility (2%) to high intelligibility (95%). The speech data were recorded at a sampling rate of 48 kHz using an eight micro- phones (6 mm in diameter) array with 1.5 in. of spacing between adjacent microphones. The array was mounted at the top of the laptop computer screen next to a video camera used to capture the visual features of speech.

Each utterance contains only a single word so that no word detection module was necessary. The vocabulary size of the database is 455 but we only utilized the 10-digit utterances of 16 of the subjects with dysarthria to provide the required speech materials for ASR modelling and evaluation.

The Mel scale is mainly based on the study of observing the arena or frequency supposed by the human. The scale is divided into the units mel. In this test the listener or test person ongoing out hearing a frequency of 1000 Hz, and labelled it 1000 Mel for reference. Then the listeners were asked to change the frequency till it reaches to the frequency twice the reference frequency. Then this frequency labelled 2000 Mel. The same procedure repeated for the half the frequency, then this frequency labelled as 500 Mel, and so on. On this basis the normal frequency is mapped into the Melfrequency. The Mel scale is normally a linear mapping below 1000 Hz and logarithmically spaced above 1000 Hz. Figure below shows the example of normal frequency is mapped into the Mel frequency.



$$m = 1127.01048 \log_e(1 + f/700) \quad \text{.....(1)}$$

$$f = 700(e^{m/1127.01048} - 1) \quad \text{.....(2)}$$

Fig 1: Mapping of normal frequency and Melfrequency

The equation (1) above shows the mapping the normal frequency into the Mel frequency and equation (2) is the inverse, to get back the normal frequency.

3.1. Evaluation criteria

Accuracy and word recognition rate are considered as the evaluation criteria in order to assess the quality of the ANN-based speech recognisers produced in this study. These two parameters are defined as follows [28]:
1. **Word Recognition Rate (WRR)**: The proportion of correct identification of the words, (i.e. digits) by the ASR system. This conveys the correctness of the recognisers' results when the evaluation data are given to the system:

$$WRR = \frac{WCR}{TWA} \times 100$$

In which WCR is the number of words correctly recognised and TWA is the total words attempted.
2. **Normalised Root Mean Square Error (NRMSE):** It is used to portion the accuracy of the system. NRMSE is usually slow in computational neurosciences in order to show how well a system learns a model. Here, it is based on the calculation of the absolute distance between the ideal results (i.e. zero and one as the min and max of the sigmoid activation function) and the actual results fashioned by the ASR system through the evaluation procedures. This parameter shows how close the ASR results are to the ideal ones in practice. Lower NRMSE percentage values show that the ASR is more accurate. NRMSE is simply defined as:

$$NRMSE(\%) = \frac{RMSE}{Max_{IdealOutput} - Min_{IdealOutput}}$$

where RMSE is calculated as:

$$RMSE = \sqrt{\frac{\sum_{j=1}^{m}\sum_{i=1}^{n}(IdealOutput_i - ANNOutput_i)^2}{n \times m}}$$

Fig 2: maximum and minimum of Sigmoid activation function

IV. CONCLUSION

In this work, a comaparitive study of different methods like MFCC, ANN Based Speech recognition were given. These are the best suggestions to find and improve the speech disability of dysartria disabled people.

REFERENCES

[1] M. Hasegawa-Johnson, J. Gunderson, A. Perlman, T. Huang, HMM-based and SVM-based recognition of the speech of talkers with spastic dysarthria, in: Proceedings of the 2006 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2006, pp. 1060–1063.

[2] S.-A. Selouani, M.S. Yakoub, D. O'Shaughnessy, Alternative speech communication system for persons with severe speech disorders, EURASIP J.Adv. Signal Process. 2009 (2009) 1–12.

[3] M. Mohri, F. Pereira, and M. Riley, "Weighted finite-state transducers in speech recognition," Comput. Speech Lang., vol. 16, no. 1, pp. 69–88, 2002.

[4] S.O.C. Morales, S.J. Cox, Modelling errors in automatic speech recognition for dysarthric speakers, EURASIP J. Adv. Signal Process. 2009 (2009) 1–14

[5] S.A. Borrie, M.J. McAuliffe, J.M. Liss, Perceptual learning of dysarthric speech: a review of experimental studies, J. Speech Lang. Hear. Res. 55 (2012) 290–305.

[6] P. Kitzing, A. Maier, V.L. Ahlander, Automatic speech recognition (ASR) and its use as a tool for assessment or therapy of voice, speech, and language disorders, Logop. Phoniatr. Voco. 34 (2009) 91–96.

[7] P.C. Doyle, H.A. Leeper, A.L. Kotler, N. Thomas-Stonell, C. Oneill, M.C. Dylke, K. Rolls, Dysarthric speech: a comparison of computerized speech recognition and listener intelligibility, J. Rehabil. Res. Dev. 34 (1997) 309–316.

