Speech Recognition System for Cerebral Palsy

M. Hafidz M. J., S.A.R. Al-Haddad, Chee Kyun Ng
Department of Computer & Communication Systems Engineering,
Faculty of Engineering, Universiti Putra Malaysia, UPM Serdang,
43400 Selangor Darul Ehsan, Malaysia.
Email: hafidzjamil@gmail.com, {sar, mpnck}@eng.upm.edu.my

Abstract: - This paper aims to design and implement the Speech recognition system for Cerebral Palsy using Matlab (GUI). Speech recognition system for Cerebral Palsy is a system that compares the voice from the Cerebral Palsy person as input with the reference voice in data base by using some algorithms. This work was based on the zero-cross-rate (ZCR) that use for the end point detection, Mel-Frequency Cepstral Coefficients (MFCC) that use for the feature extraction, and Dynamic Time Warping (DTW) that use for the pattern classification. Basically, the system consists of two parts which are training and recognition. The simulation results show that the system cans recognition the Cerebral Palsy speech accurately.

Key-Words: - Cerebral Palsy, recognition, reference voice, end point detection, future extraction, pattern classification

1 Introduction

A disability in humans may be physical, cognitive, sensory, emotional and developmental or some combination of it. A disability is an umbrella term, covering impairments, activity limitations and participation restrictions. Impairment is a problem in body function or structure which is an activity limitation is a difficulty encountered by an individual in executing a task or action [1].

Speech articulatory disability or dysarthria can arise from a number of conditions including cerebral palsy, multiple sclerosis, Parkinson's disease and others [2]. Any of the speech subsystems like resonance, prosody, respiration, phonation and articulation can be affected. Cerebral palsy describes a group of chronic disorders affecting muscle coordination and body movement which is "Cerebral" means brain and "Palsy" refers to problem in muscle movement [3]. Cerebral palsy affected area of the brain most likely involves connections between the cortex and other parts of the brain such as the cerebellum. Cerebral palsy can cause physical disability in human development and also speech [4].

Speech recognition system is an amazing technology system that basically convert the spoken words to the text that user easily can read as shown in Figure 1. It processes spoken input and translates it into text that an application can understand. Sometimes, the voice recognition term is used to refer to recognition systems that must be trained to a particular speaker such as the case for most desktop and smart phone recognition software [5].

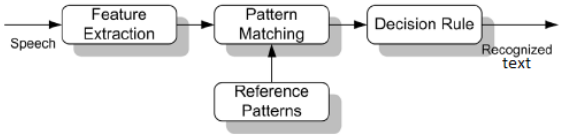

Figure 1: Speech recognition system concept.

Speech recognition system technology also designed for individuals in the disability community which is to help people with musculoskeletal disabilities caused by multiple sclerosis, cerebral palsy or arthritis achieves maximum productivity on the computers. Speech intelligibility is a main problem for the people who have cerebral palsy. The motor disorder which characterize cerebral palsy can affect the function of the muscles involve in the production of speech [6]. Efforts to provide effective means of communication for these people include the development of augmentative communication device and more recently ASR system have been investigated for use with the disordered speech of [7].

The aim for this Speech recognition system for Cerebral Palsy is to develop a system that can help the human to recognize the Cerebral palsy speech in order to use in further application.

2 Literature Review

2.1 Speech articulator

Under the speech articulator disabilities type, there are Cerebral palsy, Hyperkinetic dysarthria, Parkinson's disease and multiple sclerosis. In general, cerebral palsy individuals which have less articulatory precision for mostly to child patient. Simple "steady-state" phonemes like vowels are physically which is the easiest way to

ISBN: 978-1-61804-019-0

produce since they do not require dynamic movement of the articulatory structures.

In contrast, phonetic transitions such as consonants are the most difficult problem to produce since they require fine motor control. Fine motor control use to precisely move the articulators. Usually, mildly impaired and also impaired speakers differ in degree of disability rather than the quality [8].

Hyperkinetic dysarthria is the most common type of speech disorder that really associated with the Parkinson's disease [9]. It is most prominently characterized by disruption in the speech prosody [10]. Parkinson's commonly known as a progressive neurological disease. It involves the loss of nerve cells in the substania nigra. It also contributes the lack of dopamine content in the striatum. Bradykinesia, muscle rigidity, tremor, and akinesia are commonly reported for the clinical symptoms of Parkinson dieses [11], [12].

2.1 Speech recognition for disabled

In 1993, Sy and Horowitz was evaluated the degree of speech impairment. They also evaluated the utility of computer recognition like speech using a statistical causal model. They presented a case study of a dysarthric speaker compared against the normal speaker serving as a control [13].

After 7 years, Patel investigated the application of such as technology to severely dysarthric speakers by examining prosodic parameters such as intensity contours and frequency [14].

3 Methodology

In order to achieve the Speech recognition system for Cerebral palsy there are some methods used such as end point detection, feature extraction and pattern classification as shown in Figure 2.

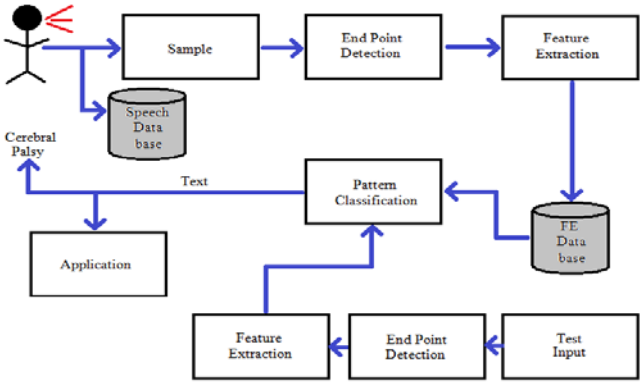

Figure 2: System methods.

3.1 End point detection

Voice from Cerebral palsy people input in the system then the beginning step is to detect the first and ending point of voice. This detection process is to remove the noise and silence part of sound. So the system will process the voice only and eliminated the unvoiced part. The end point detection algorithm used in this system is the zero-crossing rate (ZCR).

ZCR methods mean number of times the sound sequence change its sign per frame. This method used to count the frequent of the signal that crosses over the zero axes. It is an effective method to eliminate the silence part of sound and unvoiced sound. ZCR is given by this equation:

$$Z(n) = \frac{1}{2} \sum_{m=1}^{N} |\operatorname{sgn}[x(m+1)] - \operatorname{sgn}[x(m)]|$$

where,

$$\operatorname{sgn}[x(m)] = \begin{cases} +1, x(m) \geq 0 \\ -1, x(m) < 0 \end{cases}$$

3.2 Feature extraction

Feature extraction involve of speech signal to reduce the data size before pattern classification. Feature extraction techniques are classified as temporal analysis and spectral analysis techniques. In temporal analysis, the speech waveform itself is used for analysis. In spectral analysis, spectral representation of speech signal is used for analysis. So, the analysis technique chosen is Mel-Frequency Cepstral Coefficients (MFCC).

The main purpose of the MFCC process is to mimic the behavior of the human ears. The method of filters the spaced linearly at low frequencies and logarithmically at high frequencies have been used in order to capture the phonetically important characteristics of the voice. Figure 3 show the step of MFCC process [15].

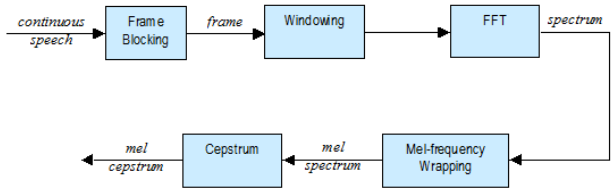

Figure 3: MFCC block diagram.

In beginning, framing process applied to voice signal in frame block. Then, the voice signal is blocked into frames. After that, windowing process is to minimize the signal discontinuities at the starting and end of each frame. Fast Fourier Transform is to converts each frame of samples in to the frequency domain. The next process is Mel Frequency Wrapped from spectrum obtained from the Fast Fourier Transform process. The process is to convert to the Mel spectrum. Finally, the log Mel spectrum is then convert back to time and the

ISBN: 978-1-61804-019-0

result is called the Mel frequency Cepstrum coefficients (MFCC).

3.3 Pattern classification

Pattern classification is a procedure for assigning of a given piece of input data into one of a given number of categories. In this process the result in feature extraction will compare with the reference. It may various in term of speed or time. In order to complete this process, the Dynamic Time Warping (DTW) algorithm is used.

DTW algorithm used to accommodate the differences in timing between sample words and templates used. The basic principle is to allow a range of steps in the space of time frames in sample and time frames in template. It also to find the path through that space that maximizes the local match between the aligned time frames and subject to the constraints implicit in the allowable steps. The total similarity cost found by this algorithm is a good indication of how well the sample and template match, which can be used to choose the best-matching template.

The simplest way to recognize an isolated word sample is to compare it against a number of stored word templates and determine which "best match" is. This goal is complicated by a number of factors. First, different samples of a given word will have somewhat different durations. DTW is an efficient method for finding the optimal nonlinear alignment as shown in Figure 4. DTW is an instance of the general class of algorithms known as dynamic programming. Its time and space complexity is merely linear in the duration of the speech sample and the vocabulary size.

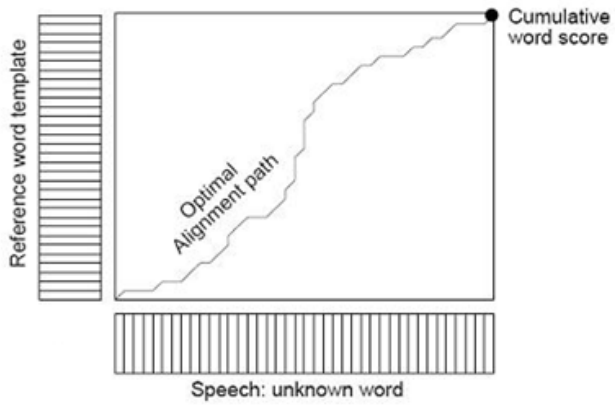

Figure 4: The optimal alignment.

If D(x,y) is the Euclidean distance between frame x of the speech sample and frame y of the reference template, and if C(x,y) is the cumulative score along an optimal alignment path that leads to (x,y), then,

$$C(x, y) = MIN(C(x - 1, y), C(x - 1, y - 1), C(x, y - 1)) + D(x, y)$$

The resulting alignment path may be visualized as a low valley of Euclidean distance scores, meandering through the hilly landscape of the matrix, beginning at (0, 0) and ending at the final point (X, Y). By keeping track of back pointers, the full alignment path can be recovered by tracing backwards from (X, Y). An optimal alignment path is computed for each reference word template, and the one with the lowest cumulative score is considered to be the best match for the unknown speech sample.

4 Result and Discussion

In this Speech recognition system for Cerebral palsy system, it is divided into two which is for digit and words but it is the same concept for both. The main part of program may consist of record, play and test as shown in figure 5.

Figure 5: The main page.

For the recording page, there are some word will trigger appear at the screen to help Cerebral palsy people record their voice as show in figure 6. When voice is recorded it will trigger to the next word to be record and there is waveform shown in the recording process as shown in figure 7. There are 10 daily words to be recorded here. The ZCR will eliminate the silence part and will do feature extraction for all of recorded voices signal.

Figure 6: Recording page.

ISBN: 978-1-61804-019-0

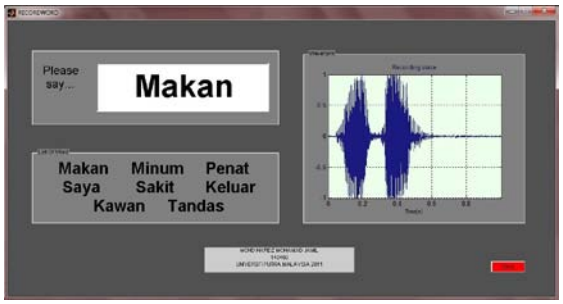

Figure 7: Waveform of signal in recording page.

After the recording process done, user can listen to the recorded voice by using the play page as shown in figure 8. In this page, it also provides the signal waveform for user to see their voice waveform frequency.

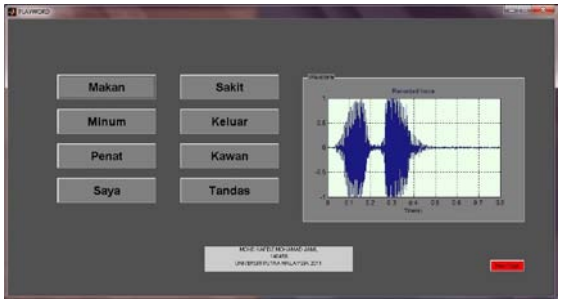

Figure 8: Replay recorded voice.

Finally, the verification part is to match the test voice with the recorded voice as reference. The verification page consist of the list of word have recorded to help user to verify as shown in figure 9. User need to speak one of the word that have recorded in the beginning and then the system will show the word that match to their verify word.

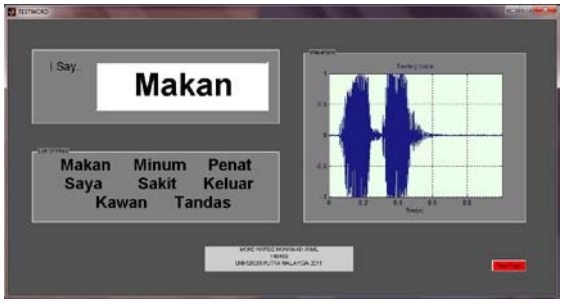

Figure 9: Word verification.

Basically, DTW algorithm compare the feature extraction of the recorded voice as reference with the new input voice that been featured extraction too. So, its proof that the Cerebral palsy speech can be improve using this program because it will appear the text as the output that normal human easily can understand their speech.

5 Conclusion

Speech recognition system for cerebral palsy approach that based on the ZCR, MFCC, and DTW has been proposed. The ZCR algorithm success to detect the silence sounds of the input and then removed them. While the MFCC algorithm success to extract the data from the main part of the input sound, and finally, the DTW algorithm success to calculate the distance between the input sound and the sounds inside the database in order to make a correct recognition. The recognition result shows that the system is work properly and makes the recognition with correct.

References:

[1]World Health Organization Page, http://www.who.int/topics/disabilities/en/, (Latest Access Date: October, 20, 2010)

[2]Menendez-Padial X, Polikoff JB, Peters SM, Leonzio JE and Bunnell HT, "The Nemours database of dysarthric speech," Proceedings of the 4th International Conference on Spoken Language Processing (ICSLP), 1996 October 3–6, Philadelphia (PA). Wilmington (DE): Applied Science and Engineering Laboratories, pp. 1962–65, 1996.

[3]Baram, Y. and Lenger, R., "Gait improvement in patients with cerebral palsy by visual and auditory feedback," Virtual Rehabilitation International Conference, 2009, pp. 146 – 149

[4]Pokhariya, H., Kulkarni, P., Kantroo, V., and Jindal, T., "Navigo--Accessibility Solutions for Cerebral Palsy Affected," Proceedings of the International Conference on Computational Intelligence for Modeling Control and Automation and International Conference on Intelligent Agents Web Technologies and International Commerce (CIMCA 2006), pp. 143, December 2006.

[5]Sakai, T. and Doshita, S., "The Automatic Speech Recognition System for Conversational Sound," Department of Electrical Engineering, Kyoto University, Kyoto, Japan, vol. 12, no. 6, 1963, Electronic Computers, IEEE Transactions on.

[6]Sandlund Marlene, Waterworth Eva Lindh, McDonough Suzanne, Ross Charlotte Hager, "Interactive Games in Motor Rehabilitation for Children with Sensorimotor Disorders," pp. 80 – 80, 2007

[7]Grattan, K.T.V., Palmer, A.W. and Shurrock, C.S.A.,"Speech recognition for the disabled," Dept. of Electr., Electron. & Inf. Eng., City Univ., London, Engineering in Medicine and Biology Magazine, IEEE, September 1991, vol. 10, no. 3, pp. 51-57.

[8]John Coopersmith Gold,"Cerebral palsy," Berkeley Heights (NJ), Enslow Publishers, Inc.; 2001.

ISBN: 978-1-61804-019-0

[9]Duffy, J. R., “Motor speech disorders: Substrates, differential diagnosis and management,” St. Louis, MO: Mosby-Year Book, 2005.
[10]Darley, F. L., Aronson, A. E., and Brown, J. R., “Clusters of deviant speech dimensions in the dysarthrias,” Journal of Speech and Hearing Research, vol. 12, 462–496, 1969.
[11]Brodal, P, “The central nervous system: Structure and Function,” New York, NY: Oxford University Press, 1998
[12]Logemann, J., Fisher, H., Boshes, B.,&Blonsky, E., “Frequency and cooccurrence of vocal tract dysfunctions in the speech of a large sample of Parkinson patients,” Journal of Speech and Hearing Disorders, vol 46, pp. 348–352, 1978.
[13]Sy BK and Horowitz DM, “A statistical causal model for the assessment of dysarthric speech and the utility of computer based speech recognition,” IEEE Trans Biomed Eng. 1993, vol. 40 no. 12, pp. 1282 – 1298.
[14]Patel R., “Identifying information bearing prosodic parameters in severely dysarthric speech .” University of Toronto (Canada), 2000.
[15]Soo Yee Cheang, and A.M. Ahmad, “Malay language text-independent speaker verification using NN-MLP classifier with MFCC,” Electronic Design, ICED 2008, International Conference, pp. 1 – 5, 2008.