Speech Recognition Based Microcontroller for Wheelchair Movement

Dr. Eyad I. Abass
Electrical Engineering Department,University of Technology/ Baghdad
Email: eyad_electdep@yahoo.com
Mohammed E. Safi
Electrical Engineering Department,University of Technology/ Baghdad
mohammed.ehsan@ymail.com

Received on: 24/11/2013 & Accepted on: 6/4/2014

ABSTRACT

This paper introduced an approach to design and implement a control system for the movement of wheelchair by means of the human voice for paralyzed patients. In this paper, the Mel-Frequency Cepstral Coefficient (MFCC) technique is used as feature extraction with Dynamic Time Warping (DTW) for features matching. The output of the system is used to control the movement of the wheelchair through an interface between notebook and microcontroller.

The experimental results showed that the proposed methods gave a recognition rate 100% of the already trained speakers with environment noise reach to 66dB. The test was conducted at different sound levels of the surrounding environment (53 to 73) dB as measured by Sound Level Meter (SLM).

Keywords: Speech Recognition, Dynamic Time Warping, and Microcontroller.

تمييز الكلام المستند على المتحكم الدقيق لحركات الكرسي المدولب

الخلاصة

هذا البحث يقدم طريقة للسيطرة على حركة الكرسي المدولب من خلال كلمات تم تمييزها للمتكلم وباستخدام المتحكم الدقيق. هذا البحث , التقنية المقترحة هي معاملات نغمة طيف التردد Mel-Frequency Cepstral Coefficient (MFCC) لأستخراج الخواص مع طريقة انحراف الوقت الديناميكي Dynamic Time Warping (DTW) لمطابقة الخصائص.اخيرا, استخدام الخارج من هذه الخورازمية للسيطرة على الكرسي المتحرك من خلال الربط بين الحاسوب والمتحكم الدقيق.

تم اجراء الاختبار على مستويات مختلفة من ضوضاء البيئة المحيطة (53 الى 73) ديسيبل حسب قراءات جهاز قياس مستوى الصوت. وأظهرت النتائج التجريبية أن الأساليب المقترحة تعطي معدل التمييز بنسبة 100% للاصوات مسبقة التدريب لحدود ضوضاء تصل الى 66 ديسيبل.

https://doi.org/10.30684/etj.32.10A.2
2412-0758/University of Technology-Iraq, Baghdad, Iraq

INTRODUCTION

A wheelchair is a device that used for the mobility of a disabled people, which is controlled either manually by pushing the wheels with the hands or via various automatic systems. Wheelchairs are used by people for whom walking is difficult or impossible due to illness, injury or disability.

Human beings usually communicate with each other by voice. The development in electronics makes human beings tend to use the voice command robots, especially wheelchair to facilitate the lives of persons with disabilities who suffer from spasms and paralysis of the extremities and cannot or it is difficult for them to use joystick.

Researches in the area of the wheelchair control system are still going on, beside the development researches of Automatic Speech Recognition (ASR) to provide an easy way for disable people to control the wheelchair. Researchers pursue their studies to the development of that field in both hardware and software implementation. Below their relevant works are briefly described:

Z. Abd Ghani, 2007 [1] introduced a system of a wireless wheelchair control system which employs a voice recognition using voice recognition processor (HM2007) for triggering and controlling its movements. The wheelchair is also equipped with two infrared sensors which mounted in front and rear of the wheelchair to detect obstacles for collision avoidance function. It utilizes a PIC controller to control the system operations. It communicates with the voice recognition processor to detect the spoken word and then determines the corresponding output command to drive the left and right motors.

H. Nik, 2009 [2] implements speech recognition control wheelchair use two digital signal processors from Microchip™ (dsPIC30F6014/A) mounted on a custom designed printed circuit board to perform smooth humming control and speech recognition. One DSP is dedicated to speech recognition and implements Hidden Markov Models using dsPIC30F speech recognition library developed by Microchip; the other DSP implements Fast Fourier Transforms on humming signals.

M. Qadri and S. Ahmed, 2009[3] implemented voice activated wheelchair through speech processing using Digital Signal Processor (DSP). The Texas Instruments TMS320C6711 DSP Starter Kit (DSK) is connected with the wheelchair for processing of the voice signal. The DSK calculates the energy, zero crossing and the standard deviation of the spoken word. It also generates different desired analog signals according to the spoken words which further amplified and converted into digital. These digital signals are used to operate the stepper motor. Five words are recognized which are forward, reverse, left, right and stop.

S. Jothilakshmi, V. Ramalingam, and S. Palanivel, 2009[4] proposed a method for improving the speaker segmentation performance by fusing the residual phase and MFCC features. This method is evaluated using television broadcast interviews and NIST 2004 database. The support vector machines are used to detect the speaker change. The system reports a performance of 85.97%. The proposed system can be extended to detect the speaker changes in the speech conversations containing more than two speakers.

A. AL-Thahab, 2011 [5] proposed a technique called Multiregdilet transform was used for isolated word recognition. Finally, using the outputs of the neural network (NNT) to control the wheelchair through computer notebook and special interface hardware. The rate of recognition command "GO" is 90%, and 100% for other commands.

Speaker Recognition

Speech recognition in this work is for Arabic language and any other language according to the training of the user to the system and the identity of user or many users of the wheelchair. Speech recognition contains three steps: preprocessing, feature extraction, and feature matching using DTW.

Preprocessing

Speech signal must be transformed into discrete and prepare it for feature extraction. The popular processes of transforming speech signal and being accommodated to the next stage of feature extraction. Sampling the speech signal which continues signal to get discrete signal. The main purpos of analog to digital (A/D) converter, is to quantize (digital representation of samples) each discrete sample x (n), n=0, 1…. N-1 into a specific number [6]. Then processes the passing of the signal through a filter which emphasizes higher frequencies. This process will increase the energy of the signal at higher frequency [7]. The most commonly used filter for this step is the finite impulse response filter described below [8]:

$$Y[n] = X[n] - 0.95X[n-1] \qquad \ldots(1)$$

Two methods are used to detect End Points : Short-Term Energy (STE) and Short-Term Zero Crossing (STZC): STE is the most obvious and simple indicator of ‘‘voicedness”. Typically, voiced sounds are several orders of magnitude higher in energy than unvoiced signals. For the frame (of length N) ending at instant m, the energy is given by equation (2) [9].

$$E(m) = \sum_{n=m-N+1}^{m} S^2[n] \qquad \ldots(2)$$

The value of N is chosen to meet the frame time length to be (10-40) ms, since in this time, the speech signal is considered unchanged or its statistical properties are relatively constant [10]. The rate at which zero-crossings occurs is a simple measure of the frequency content of a narrowband signal [11]. The zero crossing rate of the frame ending at time instant m is defined by equation (3)[9]:

$$ZC[m] = \frac{1}{2}\sum_{n=m-N+1}^{m} |\mathrm{sgn}(\mathrm{s[n]}) - \mathrm{sgn}(\mathrm{s[n-1]})| \qquad \ldots(3)$$

2.2 Feature Extraction

Mel-Frequency Cepstral coefficients (MFCCs) are based on the known variation of the human ear’s critical frequency bandwidths. This is presented in the Mel-frequency scale, which is a linear frequency space below 1000 Hz and a logarithmic space above 1000 Hz [12]. A popular relation between f(HZ) and mel-frequency scale F_{mel} is as below [13]:

$$F_{mel} = 2595 * \log_{10}(1 + \frac{\mathrm{f(HZ)}}{700}) \qquad \ldots(4)$$

MFCC provides a baseline acoustic feature set for speech and SR applications [13]. MFCCs with single energy and their dynamic derivatives were used for feature extraction. Figure (1) shows the block diagram for the MFCC feature extraction step by step.

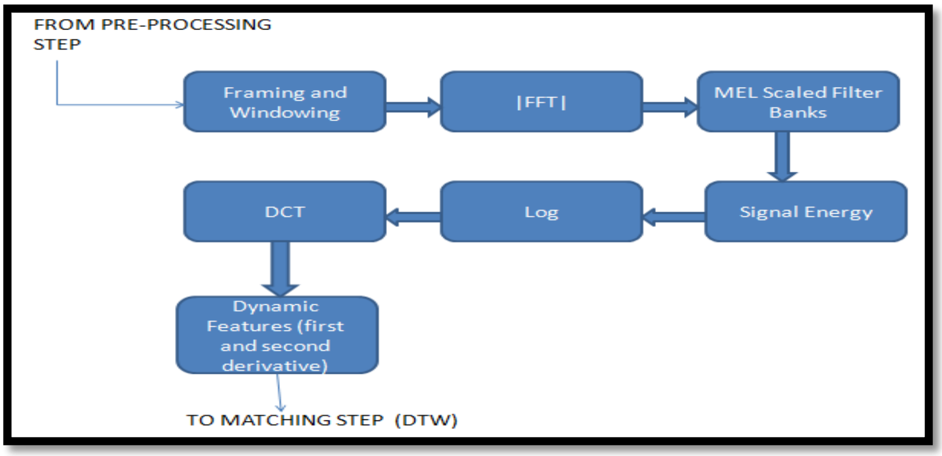

Figure (1): Block Diagram for the Feature Extraction[6]

The details of the block diagram are described below:

• Framing and Windowing: The next thing to do with speech signal after pre-processing is to divide it into speech frames and apply a window to each frame, Each frame is K samples long, with adjacent frames being separated by P samples.

A commonly used window is the Hamming window [14]. It is calculated as:

$$w(\mathrm{k}) = \begin{cases} 0.54 - 0.46\cos(\frac{2\pi k}{K-1}) & ,0 \le k \le K-1 \\ 0, \text{otherwise} \end{cases} \quad \ldots (5)$$

• Fast Fourier Transform (FFT): the Fast Fourier Transform is a fast implementation of the Discrete Fourier Transform (DFT) which converts N-samples of frames into frequency spectrum.

• Mel Scaled Filter banks: The Mel-scale filter bank implementation used in this study includes 40 triangular filters non-uniformly spaced along the frequency axis [10].

•Signal Energy: Furthermore, the signal energy is added to the set of parameters. It can simply be computed from the speech samples s(n) within the time window by [14]:

$$e(n) = \sum_{n=0}^{N-1} s^2(n) \quad \ldots (6)$$

•Discrete Cosine Transform (DCT)**:** The cepstrum is defined as the inverse Fourier transform of the log magnitude of Fourier transform of the signal. Since the log Mel filter bank coefficients are real and symmetric, the inverse Fourier transform operation can be replaced by DCT to generate the cepstral coefficients [15]. The cepstral coefficients are the DCT of the M filter outputs obtained from [16]:

$$MFCC_n = \sum_{k=0}^{39} X_k \cos\left[n\left(k-\frac{1}{2}\right)\frac{\pi}{40}\right], \text{ for n=0,1,2,...,M-1 }, \quad \ldots (7)$$

Where M is the number of MFCC coefficients and X_k, k =0,2,...,39, represents the log energy output of the k_{th} filter.

•Dynamic Parameters: The voice signal and the frames changes, such as the slope of a formant at its transitions. Therefore, there is a need to add features related to the change in cepstral features over time. 13 delta or velocity features (12 cepstral features plus energy), and 13 features a double delta or acceleration feature are added. Each of the 13 delta features represents the change between frames in the equation (8) corresponding cepstral or energy feature, while each of the 39 double delta features represents the change between frames in the corresponding delta features [7].

$$d(t) = \frac{c(t+1)-c(t-1)}{2} \qquad \ldots(8)$$

Where c is the feature vectors, and t=0, 1,2,… frame number-1.

Features Matching using DTW

Dynamic time warping (DTW) is an algorithm for measuring similarity between two sequences which may vary in time or speed. DTW allows a nonlinear warping alignment of one signal to another by minimizing the distance between the two as shown in Figure (2).

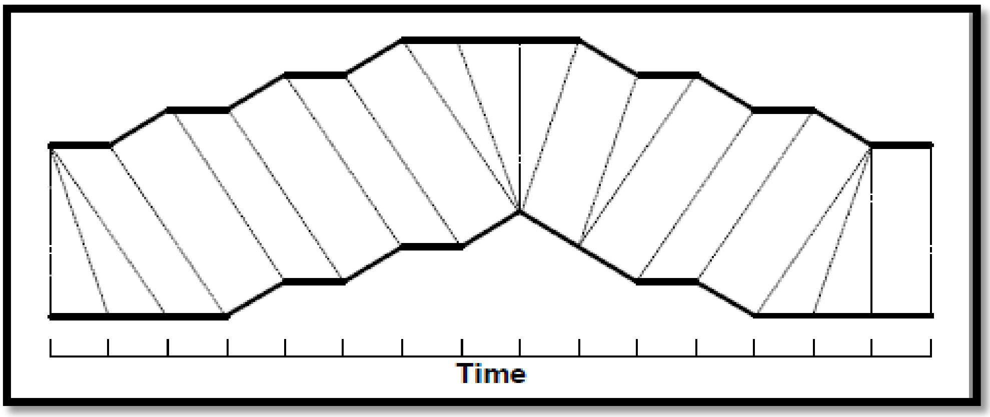

Figure (2): A Warping between two time series [7]

This warping between two signals can be used to determine the similarity between them and thus it is very useful for feature recognition. DTW is a pattern matching algorithm with a non-linear time optimization effect based on Bellman's principle of optimality, which states that given an optimal path from A to B and a point C lying somewhere along this path, the path segments AC and CB are optimal paths from A to C and C to B respectively [15]. The DTW objective is to find the warping path W = {w1, w2, w3, . . ., wK} of contiguous elements on distMatrix (with max(TX-1, TW-1) < K < ((TX-1) + (TW-1) -1), and wk= distMatrix(i, j)), such that it minimizes the following function:

$$DTW(X,Y) = \min\left\{\sqrt{\sum_{k=1}^{K} w_k}\right. \qquad \ldots (9)$$

The warping path is subject to several constraints, see Figure (3) .Given wk = (i, j) and wk-1 = (i’, j’) with i, i’ ≤ (TX-1) and j, j’ ≤ (TW-1)[17]:
1. Boundary conditions. w1 = (1,1) and wK = (TX-1, TW-1).
2. Continuity. i – i’ ≤ 1 and j – j’ ≤ 1.
3. Monotonicity. i – i’ ≥ 0 and j – j’ ≥ 0.
This path can be found by using dynamic programming to evaluate the following Recurrence, which defines the cumulative distance *D(i, j)* as the distance *d(i, j)* found in the current cell and the minimum of the cumulative distances of the adjacent elements[18]:

$$D(i, j) = d(Xi, W j) + \min \{D(i-1, j-1), D(i-1, j), D(i, j-1)\} \qquad \ldots(10)$$

The Euclidean distance between two sequences can be seen as a special case of DTW where the *k*th element of *W* is constrained such that $wk=(i,j)_k$,$i= j = k$. Note that it is only defined in the special case where the two sequences have the same length [18].

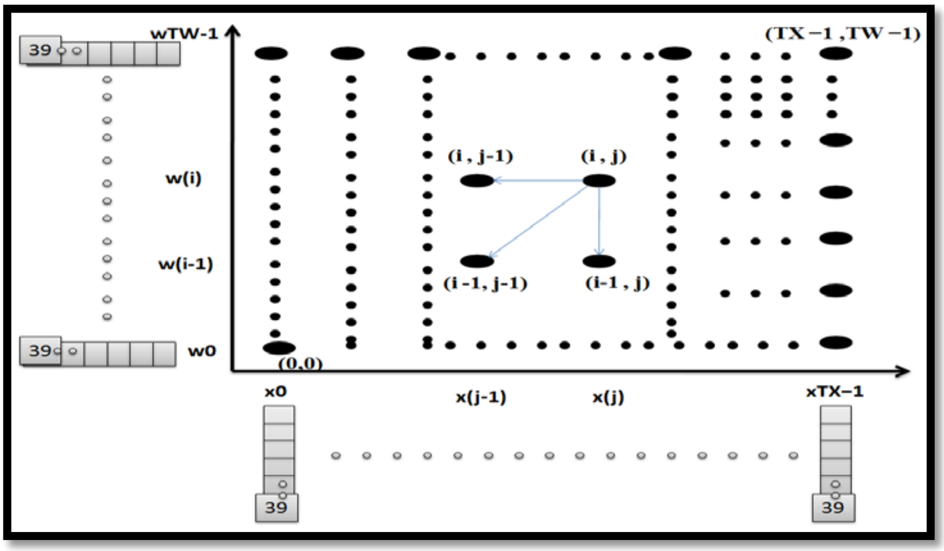

Figure (3): Local path alternatives for a grid point

Proposed System Design

The proposed wheelchair looks like a conventional mechanical wheelchair, but components were added to it, that are cheap comparative with the cost of electric wheelchair that even does not have the technology of speech recognition. Proposed wheelchair as shown in Figure (4) provides both speech recognition of user instructions and manual control.

The wheelchair direction movement control system consists of speech recognition part, which is represented in MATLAB and installed on laptop or notebook to programmed speech recognition algorithm with manual control (keyboard work as joystick) on it.

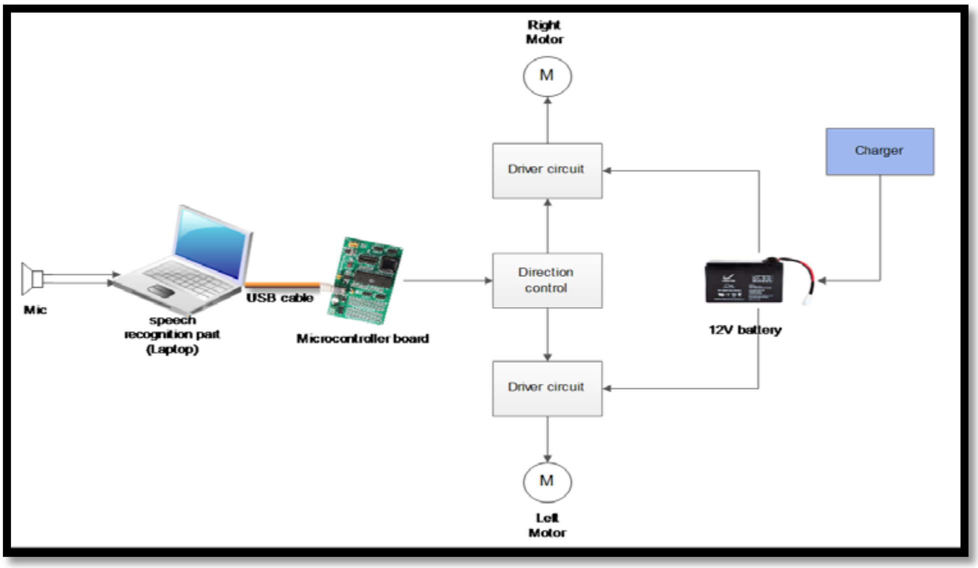

Figure (4): Proposed System Design

microcontroller board, which consists of microcontroller and interface between laptop and microcontroller by using USB to serial converter on board, and finally interface between microcontroller and driver circuits of two motors, which consist of H-bridge relays.

Microcontroller Interface

The control command that gets from Laptop performs on motors through special interface using Antel (AT89S8253) microcontroller on 8051-ready additional board. The connection between laptop and 8051-ready addition board perform via USB cable, as shown in Figure (5).

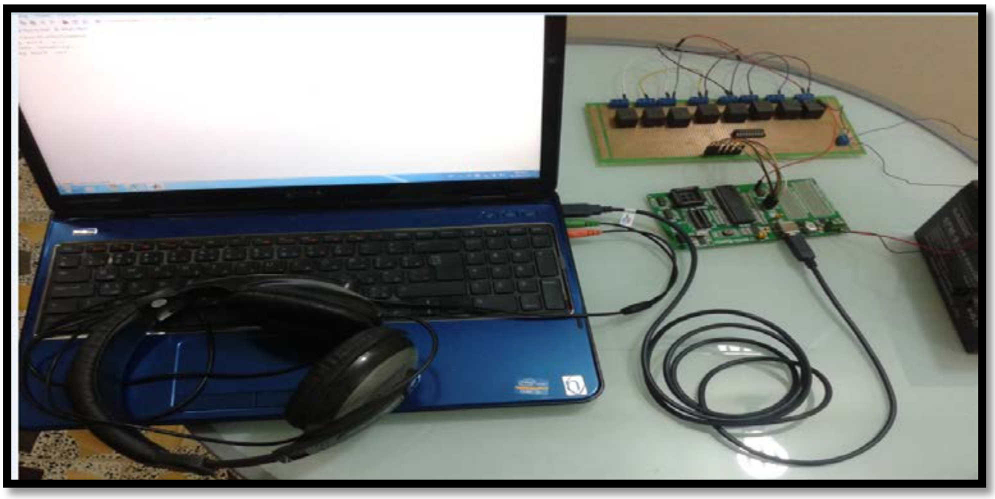

Figure (5): Interface between Laptop and Wheelchair.

Driver circuit

Driver circuit was built to connect the microcontroller board to high power consumption two motors. The driver circuit as shown in Figure (6) consists of (ULN2803) chip and eight relays to construct two H-bridges, to coordinate movement of two motors.

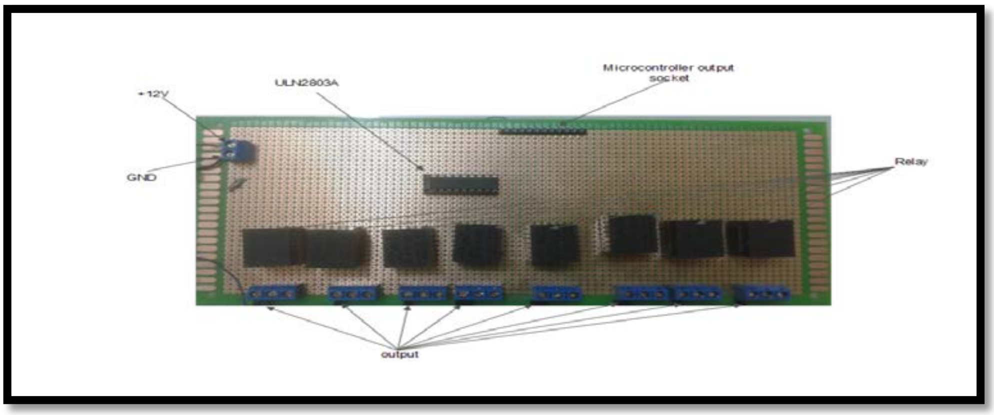

Figure (6): Driver circuit.

Implementation and Results

The total features are arranged in the form of matrix (MFCC [I, J]) as I=39, which represents the number of elements of each frame, and J represents the total overlapped frame of speech signal. As can seen in Figures (7) four different speakers features vectors for the Arabic words ammam "أمام", that can be noted the difference between each features vectors and the others for other speakers.

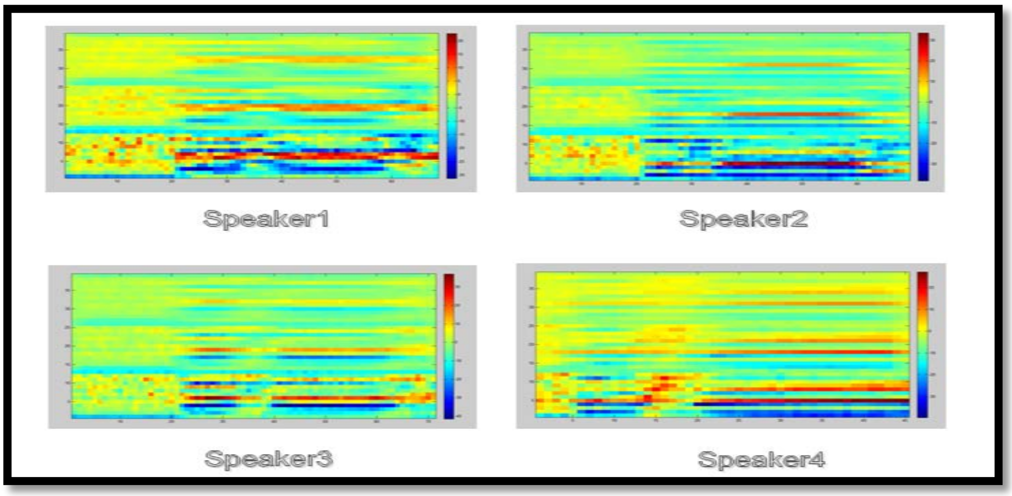

Figure (7): feature vectors for word ammam "أمام"

Feature matching of feature vectors that are extracted before, by using DTW to compute the optimal path of warping between two feature vectors, as shown in Figure(8).

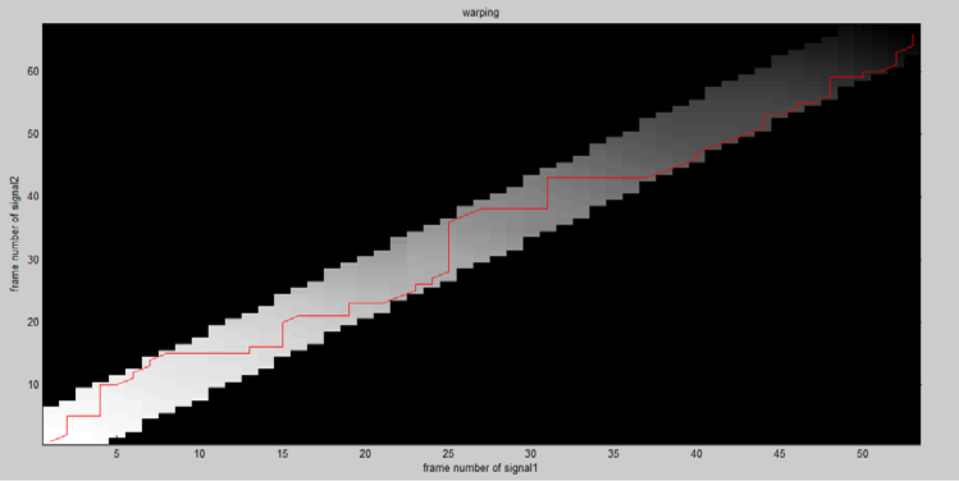

Figure (8): DTW path.

The recognition rate was computed to the effect of background noise for four speakers by testing each word ten times for noise level 40dB, 50dB, 55dB, 60dB, 66dB, and 73dB, as shown in tables (1) to (4).The first row represents the SLM measurements at different values, under each value of the SLM reading, the recognition rate corresponding to each uttered word is found.

Speaker1

The test for the recognition rate of speaker1 shows 100% for all words under 66dB, while the recognition rate decreases at 66dB for words Amamm "أمام" and Ymmen "يمين" 80%, 90% respectively. More decreasing of recognition rate at 73dB for words Amamm "أمام", Ymmen "يمين" and Keef "قف" 40%, 90% and 90% respectively, as shown in Table (1).

Speaker2

The test for the recognition rate of speaker2 shows 100% for all words under 66dB, while the recognition rate decreases at 66dB for word Ymmen "يمين" 80%. More decreasing of recognition rate at 73dB for words Amamm "أمام",Ymmen "يمين", Yassar "يسار", and Keef "قف" 80%,60%,90% and 70% respectively, as shown in Table (2).

Speaker3

The test for the recognition rate of speaker3 shows 100% for all words under 66dB, while the recognition rate decreases at 66dB for words Amamm "أمام",Ymmen "يمين", and Keef قف" 80%,90%,and 90% respectively. More decreasing of recognition rate at 73dB for words Amamm "أمام",Ymmen "يمين", Yassar "يسار", and Keef "قف" 40%,60%,80% and 60% respectively, as shown in Table (3).

Speaker4

The test for the recognition rate of speaker4 shows 100% for all words under 66dB, while the recognition rate decreases at 66dB for word Kalf "خلف" 90%. More decreasing of recognition rate at 73dB for words Amamm "أمام",Ymmen "يمين", Kalf "خلف", and Keef "قف" 70%,60%,70% and 70% respectively, as shown in Table (4).

Table (1): Speaker1 recognition rate with different background noise

Word	40db	50db	55db	60db	66db	73db
أمام	100%	100%	100%	100%	80%	40%
يمين	100%	100%	100%	100%	90%	90%
يسار	100%	100%	100%	100%	100%	100%
خلف	100%	100%	100%	100%	100%	100%
قف	100%	100%	100%	100%	100%	90%
Recognition rate	100%	100%	100%	100%	94%	84%

Table (2): Speaker2 recognition rate with different background noise

Word	40db	50db	55db	60db	66db	73db
أمام	100%	100%	100%	100%	100%	80%
يمين	100%	100%	100%	100%	80%	60%
يسار	100%	100%	100%	100%	100%	90%
خلف	100%	100%	100%	100%	100%	100%
قف	100%	100%	100%	100%	100%	70%
Recognition rate	100%	100%	100%	100%	96%	80%

Table (3): Speaker3 recognition rate with different background noise

Word	40db	50db	55db	60db	66db	73db
أمام	100%	100%	100%	100%	80%	40%
يمين	100%	100%	100%	100%	90%	60%
يسار	100%	100%	100%	100%	100%	80%
خلف	100%	100%	100%	100%	100%	100%
قف	100%	100%	100%	100%	90%	60%
Recognition rate	100%	100%	100%	100%	92%	68%

Table (4): Speaker4 recognition rate with different background noise

Word	40db	50db	55db	60db	66db	73db
أمام	100%	100%	100%	100%	100%	70%
يمين	100%	100%	100%	100%	100%	60%
يسار	100%	100%	100%	100%	100%	100%
خلف	100%	100%	100%	100%	90%	70%
قف	100%	100%	100%	100%	100%	70%
Recognition rate	100%	100%	100%	100%	98%	74%

Conclusions

The work utilizes control technique termed on speech recognition, which concluded that the highest recognition rate that can be achieved from the experimental results of the system is 100% in background noise reach to 66dB. This control technique possesses several advantages such as cost and steady state response, simple implementation, no parameter sensitivity and social need is the independence of the physically challenged people.

References

[1] Z. Abd Ghani, “Wireless Speed Control with Voice for Wheelchair Apllication”, M.Sc.Thesis,Universiti Teknologi Malaysia, May 2007.

[2] H. Nik, “Hum-Power Controller for Powered Wheelchairs”, M.Sc.Thesis, George Mason University, 2009.

[3] M. Qadri and S. Ahmed, “Voice Controlled Wheelchair Using DSK TMS320C6711”, International Conference on Signal Acquisition and Processing, PP. 217-220, 2009.

[4] S. Jothilakshmi, V. Ramalingam, and S. Palanivel, “Unsupervised Speaker Segmentation with Residual Phase and MFCC Features”, Elsevier, Expert Systems with Applications, Vol. 36, Issue 6, PP. 9799-9804, August 2009.

[5] A. AL-Thahab, “Controlled of Mobile Robots by Using Speech Recognition”, Journal of Babylon University, Pure and Applied Sciences, Vol. 19, Issue 3, PP. 1123- 1139, 2011.

[6] Alaa A.Refeis, “FPGA Implementation of Speech Recognition System Based on HMM”, M.Sc.Thesis, University of Technology, Baghdad, Iraq, August 2012.

[7] Lindasalwa Muda, Mumtaj Begam and I. Elamvazuthi, “Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques”, Journal of Computing, VOL. 2, ISSUE 3, PP. 138-143, March 2010.

[8] S-T. PAN, C-F. CHEN, and J-H. ZENG, “Speech Recognition via Hidden Markov Model and Neural Network Trained by Genetic Algorithm”, Proceedings of the Ninth International Conference on Machine Learning and Cybernetics, Qingdao, PP. 2950-2955, IEEE, 11-14 July 2010.

[9] W. Chu, “Speech Coding Algorithms”, John Wiley & Sons, ISBN 0-471-37312-5, 2003.

[10] R.S.Kurcan, “Isolated Word Recognition from In-Ear Microphone Data Using Hidden Markov Models (HMM)”, M.Sc.Thesis, Naval Postgraduate School, Turkish Naval Academy, March 2006.

[11] L. R. Rabiner and R. W. Schafer, “Digital Processing of Speech Signals”, Prentice Hall International, Inc., ISBN 0-13-213603-1, 1978.

[12] J. Cao, and et al., “A Two-stage Pattern Matching Method for Speaker Recognition of Partner Robots”, ©2010 IEEE, 2010.

[13] M. Hossan and M. Gregory, “Speaker Recognition Utilizing Distributed DCT-II Based Mel Frequency Cepstral Coefficients and Fuzzy Vector Quantization”, Int J Speech Technol, PP. 103–113, DOI 10.1007/s10772-012-9166-0, 2013.

[14] B. Plannerer, “An Introduction to Speech Recognition”, Germany, 2005.

[15] S. Chapaneri, “Spoken Digits Recognition using Weighted MFCC and Improved Features for Dynamic Time Warping”, International Journal of Computer Applications, DOAJ, Vol. 40, No. 3, PP. 6-12, February 2012.

[16] J. -C. Wang, J. -Fa. Wang, and Y. -S. Weng, “Chip Design of MFCC Extraction for Speech Recognition”, CiteSeer, Vol. 32, No. 1-2, PP.111-131, November 2002.

[17] Adem Karahoca, “Advances in Data Mining Knowledge Discovery and Applications”, InTech, 1st Edition, ISBN 978-953-51-0748-4, 2012.

[18] Eamonn Keogh and Chotirat Ann Ratanamahatana, “Exact Indexing of Dynamic Time Warping”, Knowledge and Information Systems, PP. 358–386, 2005.