國 立 交 通 大 學

生醫工程研究所

碩 士 論 文

聲振式輪椅之方向控制

Wheelchair Direction Control by Acquiring

Vocal Cords Vibration
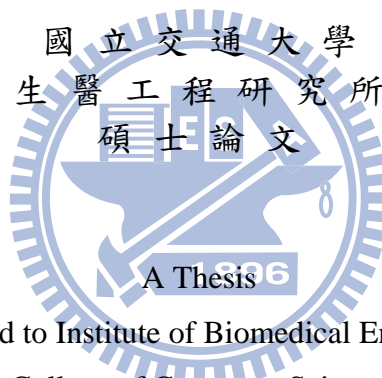
研 究 生：許嘉樺

指導教授：蕭子健

中 華 民 國 九 十 八 年 七 月

聲振式輪椅之方向控制

Wheelchair Direction Control by Acquiring Vocal Cords Vibration

研 究 生：許嘉樺   Student : Chia-Hua Hsu

指導教授：蕭子健   Advisor：Tzu-Chien Hsiao

國 立 交 通 大 學

生 醫 工 程 研 究 所

碩 士 論 文

A Thesis

Submitted to Institute of Biomedical Engineering

College of Computer Science

National Chiao Tung University

in Partial Fulfillment of the Requirements

for the Degree of

Master

in

Computer Science

July 2009

Hsinchu, Taiwan, Republic of China

中華民國九十八年七月

聲振式輪椅方向之控制

研 究 生：許嘉樺　　　　　　　　　　　　指導教授：蕭子健

國 立 交 通 大 學

生 醫 工 程 研 究 所

摘要

　　搖桿為最常見的電動輪椅操作方式，在操控敏感度與速度、轉動方向調整有極大的便利性。然而，對於四肢障礙者而言，使用手操控搖桿是生理上的限制，無法流暢地使用此操控方式，發展一套聲帶振動控制輪椅系統有其需要性，透過聲帶振動所產生的語音訊號，可以控制輪椅移動。例如:當使用中文發出前的指令，輪椅會自動的往前走。

　　辨別使用者的指令並決定輪椅的方向不是件容易的事,我們需要即時且迅速的控制器，我們使用美商國家儀器公司出產的 compactRIO 可程式化自動控制器作為開發平台，它擁有高擴充性、高穩定性、體積小與即時的處理器。當輪椅在運作時，compactRIO 在遭受高度撞擊下還可以正常運作。透過即時處理器，可即時的分析使用者所下的命令並回應給輪椅做出正確的動作。

Wheelchair Direction Control by Acquiring Vocal Cords Vibration

Student：Chia-Hua Hsu   Advisor：Tzu-Chien Hsiao

Institute of Biomedical Engineering College of Computer Science

National Chiao Tung University

Abstract

The powered wheelchair with joystick is the commonest operation. However, to steer their own powered wheelchair through a conventional joystick is difficult for people who experience total paralysis in all four limbs. Therefore, the development of a vocal cords vibration control wheelchair system is essential to help those patients moving around on their will. The basic pronunciation can be distinguished by the difference of the resulting analog waveform (signals), and further instructions (commands) can be carried out by the vocal cords vibration control wheelchair system. For example, if the word "forward" is pronounced in Chinese, the wheelchair will then move forward automatically.

To instruct the movement of wheelchair by determining the variation of the output signal which generated by vocal cords is not an easy task. Hence, a rapid and real-time processor with FPGA system, National Instrument (NI, Austin, TX) CompactRIO (cRIO-9014) is used as the develop platform. Base on its flexibility, small size, and rugged form. In addition, when the wheelchair moves, CompactRIO can perform normally with impact vibration. The most important thing is that CompactRIO has real-time processor, it can analyze the analog signal that generated by user rapidly and reactively response the operations that wheelchair should do.

Acknowledgement

First of all, I would like to express my sincere appreciation to my laboratory member, especially Wei-Hao Huang. For his helpful suggestions and welding controller signal circuit to wheelchair, let me finished the project successfully. I am so grateful to him for all that he had done.

My advisor, Dr. TC Hsiao, for his helpful guidance, I have made great progress in the past of two years and have gained more experience throughout my Master degree. Thanks for my advisor guidance.

I would express my gratitude to all the members in the laboratory of 704, including Hsiao's VBM and Ching's MIP. HL Yu, the first sister of 704, just like our mother, she is always thoughtful for us, thanks the first sister of 704. HY Hsu and SH Chang, thanks to your assistance and encouragement in my daily life, let me have power to finished Master degree, thank my best laboratory mate. Ken, thanks for your accompany with us many nights. Chien-Chien, thanks for making fun with us. CY Lin and Wilson Ho, you are really the best junior colleague for me. Many laboratory members to thank, including CC Chang, CW Chang, YC Huang, PM Lee, and so on. Thanks for your supports and encouragements. You have made my life wonderful.

Finally, thank my families, neighbors, relatives, and friends, especially my mother, for her understanding, supports and loves; I can finish my Master degree successfully. My godmother, you really care for me and a cell phone call a week. Ou father and mother really love me; you treat me like your child. Yu-Li Ou, thanks for your consideration and toleration, especially my temperature. We have a long way to go in the future. I wish we can go together.

Life is sometimes tough; however, there is nothing to defeat us with loves of you. Thank you for your loves.

Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1. Motivation

Ambulant capability can help to determine the handicap degree of disabled persons. Most of them use a cane or wheelchair to extend their ambulant capability. For those who want to ambulate at their own will, they can use the hand-controlled electrically powered wheelchair. The powered wheelchair with a joystick is the most common in operation; it is convenient in terms of sensitivity, speed and direction operation. However, to steer their own powered wheelchair through a conventional joystick is difficult for people afflicted with total paralysis in all four limbs, such as sufferers of muscular dystrophy (MD), spinal cord injury (SCI), amyotrophic lateral sclerosis (ALS), etc. Our goal is to help them to maintain fundamental daily life requirements. Physiological and security requirements are lower levels of Maslow's hierarchy of needs (Fig. 1.1). To design a product to satisfy the lower level is our first objective, to help them to avoid hunger and thirst. Only then can we proceed to the upper levels.
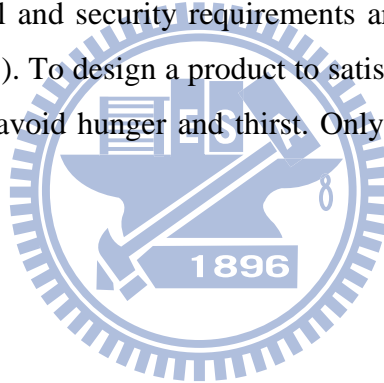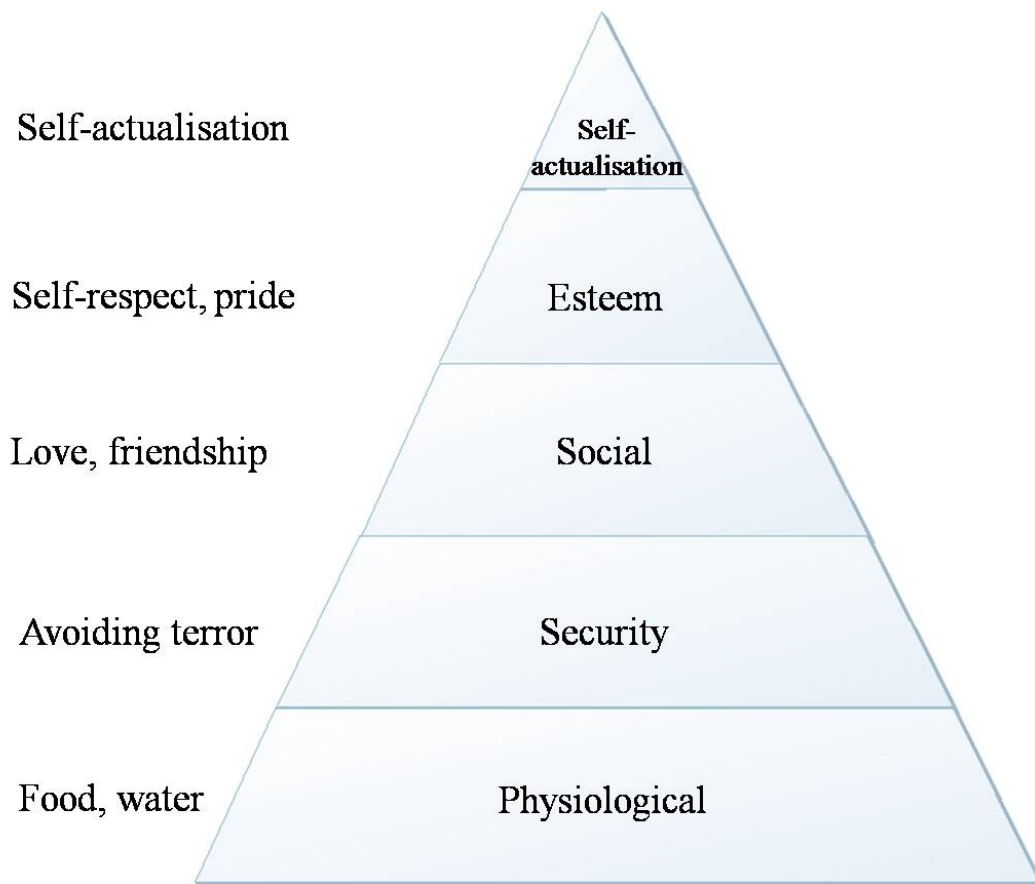
Figure 1.1: Maslow's hierarchies of needs

1.2. Limb disability situation in Taiwan

In order to understand the market in Taiwan, for our product, we need to know the population size affected by limb disability. Taiwan's government started to enforce the law requiring motorcycle riders to wear helmets in 1996. Originally, we thought that enforcing motorcycle riders to wear helmets would increase the population with limb disability, especially SCI patients. But in Fig. 1.2 [1], after 1996, the limb disability yearly rate of increase was not higher than before. But we can be sure that the population with limb disability and other disability is increasing year by year.
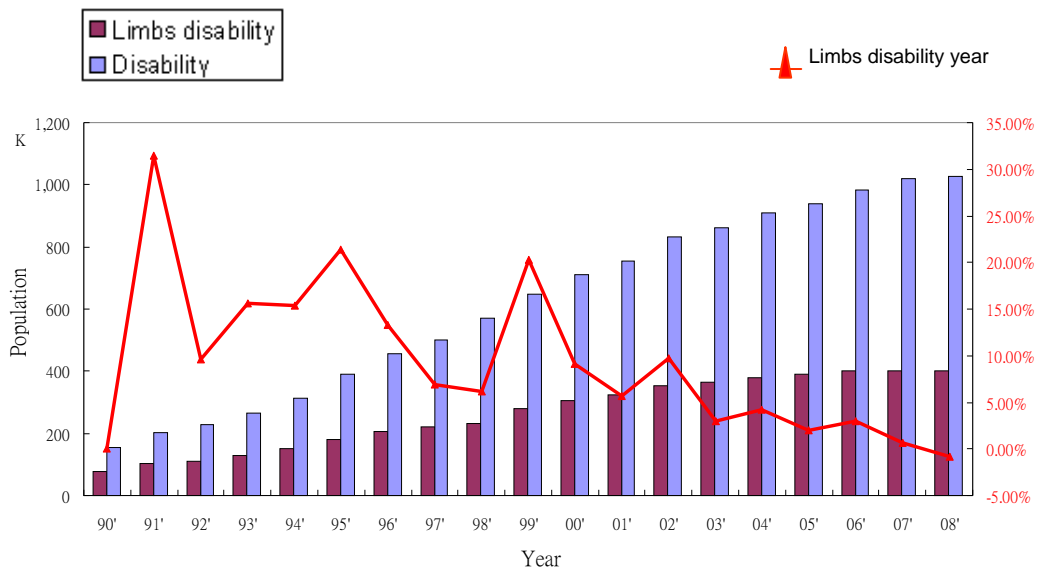
Figure 1.2: Relation between disability and limb disability of motorcycle riders
wearing helmets in Taiwan

In 2008, about four hundred thousand people in Taiwan suffered from limb
disabilities (see Table 1.1). Taiwan has a population of about twenty-three million, so
the ratio of limb disability to the total population of Taiwan is about one point seven
percent. Our school is in Hsinchu city, so we also show the data on Hsinchu. This
table tells us that those with limb disability and other disabilities constitute a large
population, although we seldom see them in public places. Most of them stay at home
all day. We think we could help them to leave home and enter t society.

Table 1.1: Population with limb disability and ratios

2008 June	Taiwan	Hsinchu
Limbs disability (Disabilities)	399,883 (1,027,041)	4,887 (13,572)
Total population	22,994,262	399,035
Ratio (Limbs disability/Total)	1.74%	1.22%

The ratio of 'need and use' assistive technology device and 'need but not use' totals about thirty-eight percent among the disabled (see Fig. 1.3). [1] Our focus is on the wheelchair, walker and powered wheelchair. These assistive technology devices are used by about twenty percent of the disabled. This is the market for our product in catering to limb disability.
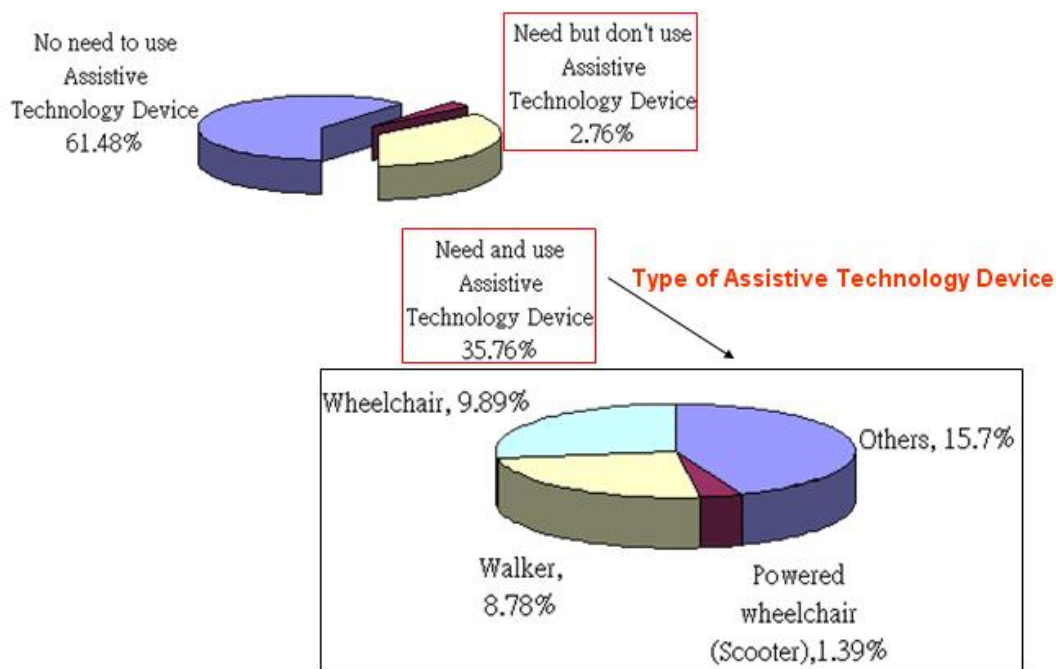
Figure 1.3: Cases of disability in Taiwan

1.3. Problem description

Clinicians indicated that 9 to 10 percent of patients who receive power wheelchair training find it extremely difficult or impossible to use the wheelchair for daily living activities [2]; it jumped to 40 percent when doing some difficult operations, like getting in or out of elevators and gates. If patients make mistakes in maneuvering their wheelchairs, it may lead to calamitous situations, such as overturning, falling out of the wheelchair, etc. AS most of these patients will depend on their wheelchairs for a long time, understanding how to maneuver the powered wheelchair proficiently and avoid injury by mistake are very important considerations. In the US, over 5000 wheelchair users are hurt due to the improper usage of the joystick and about 50 wheelchair users die every year as a result of using wheelchairs improperly [3]. Determining the best way to design a controller which can provide

users with a safe environment is an important humanitarian consideration.

1.4. Literature study

In this section, we provide some history on the powered wheelchair and give some examples of the control interface.

In 1940, Smith et al. applied for the first powered wheelchair patent. It was very simple; a motor and battery were simply added to the wheelchair without any specific control interface. In 1953 G.J. Klein et al. developed the first joystick control interface. By the 1970s, powered wheelchairs were accepted by the general population. Then powered wheelchairs get more popular; in 1991, Chrisman's team developed the eye wink control interface [4], with instruction command by the eyeball. If the eyeball turned left, the wheelchair would also turn left. In 2002, Chen et al. developed a head orientated wheelchair [5]. If users inclined their heads forward, the power wheelchair would move forward. In 2004, Ding's team modified and developed an optimized joystick control interface for the powered wheelchair [6]; they applied a fuzzy logic joystick controller and tuning interface to reduce the noise and recalibrate the signal. But the problem is that each user's tremors differ or the same user's tremors will worsen day by day. In 2006, Michael's team designed a neurological signal control interface [7]; neurological signals are received from the brain via the sides of the vocal cords and encoded to the computer via wireless network; when the computer receives signals, it decodes the signals and recognizes commands.

In regard to physically disabled patients who also use wheelchairs, including those suffering from MD, SCI, ALS, etc., there is no complete cure; modern medical technology can only alleviate the symptoms of these patients to a small degree. In order to improve the quality of these patients' lives and speed up their rehabilitation, our purpose was to design a pronounceable rehabilitation and vocal cord vibration control system. With this system, rehabilitation can be done daily via pronunciation, which is required in order to control the wheelchair. Also, this system enables patients to maneuver their wheelchairs with enhanced mobility, and to do so conveniently, via vocal cord vibrations. In this way, calamities due to improper hand operations may be avoided the disabled protected from needless harm.

1.5. Wheelchair markets

According to the report by the Industrial Economics & Knowledge Center of ITRI [8] in Taiwan, the exportation of wheelchairs and electric vehicles to the global market has increased about six times between 1999 and 2003. In 2003, the exportation of wheelchairs and electric vehicles to the global market numbered about 300,000, or about 40% of the global market, and generating an impressive NTD 7.2 billion in foreign exchange. The need for wheelchairs continues to grow; from 2003 to 2010 the wheelchair market will have increased by 60% (see Fig. 1.4). It remains a potentially lucrative market. The wheelchair motors, electronic components and manufacturing, processes have all been well developed. Only in regard to the controller, do we lack mature techniques. Our goal was to develop a controller domestically to cut costs, and module designing to provide the controller with the required flexibility. Thereby, all the wheelchair parts could be produced in Taiwan and we would not need to depend on importing controllers.

In recent years, China's wheelchair productions have been distributed around the world. We have to improve the quality of our wheelchair production, in order to avoid price competition from China; designing a high-precision and high-quality product is the correct method for achieving this goal.
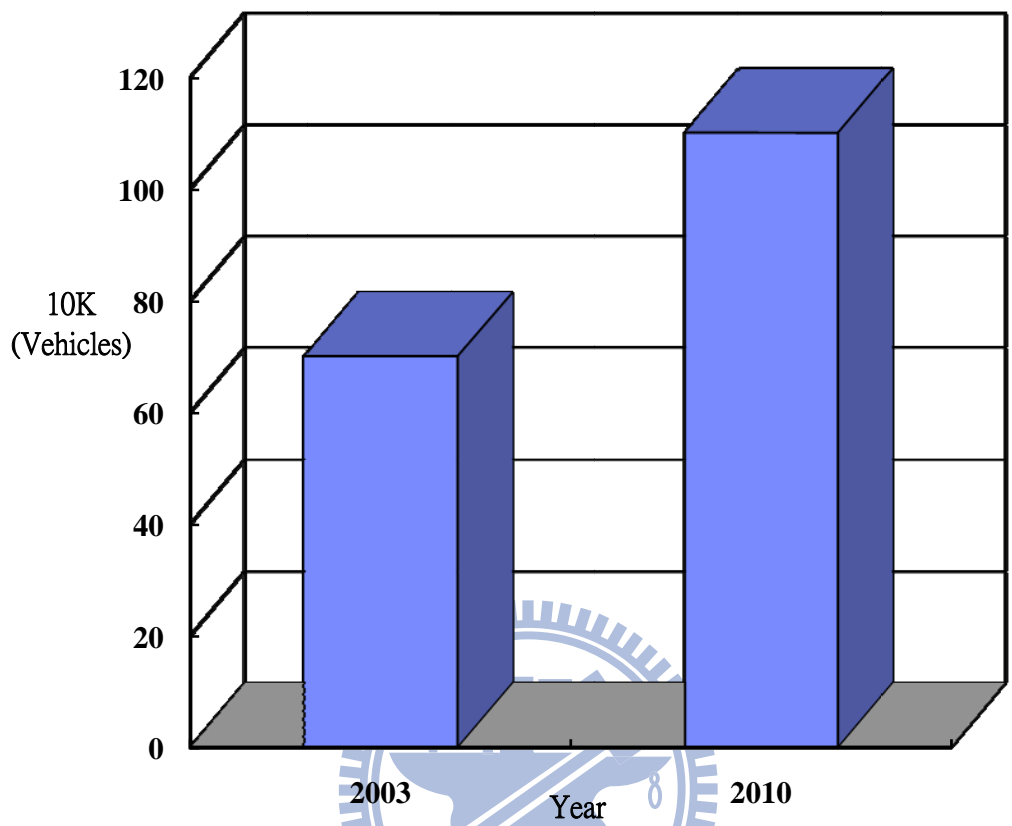
Figure 1.4 Global wheelchair markets

Chapter 2: Methods and Materials

2.1. Wheelchair type

In this study, we focus on the development on the vocal-type wheelchair. Hence, we began with the general joystick powered wheelchair. Then we added the vocal cord vibration module and the system control module into the joystick powered wheelchair functions. We chose the modular wheelchair that has a joystick and an open interface, enabling us to directly modify the wheelchair. The EXB Standup Wheelchair (aka, the EXB Wheelchair) of Nita was chosen. Nita is a world famous medical equipment company located in Taichung. The externals and specification sheets of the EXB Wheelchair are shown in Fig. 2.1 and Table 2.1. The functions of standup and recline backrest can used to avoid rapid muscle decay (cf. bedsores) due to maintaining the same posture long-term.
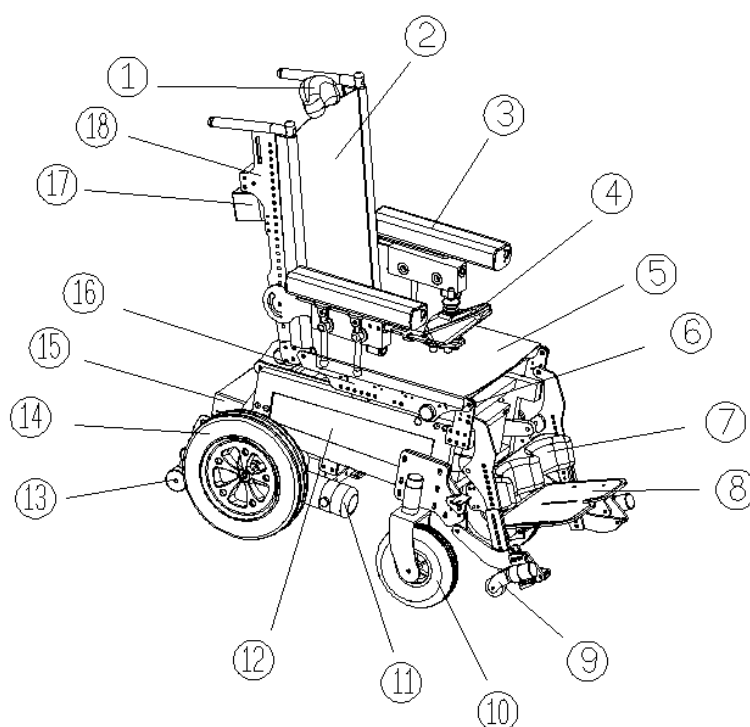
Figure 2.1 EXB Wheelchair externals (http://www.nita.com.tw/)

The components are 1) headrest, 2) backrest, 3) armrest, 4) controller, 5) seat pad, 6) liner actuator (seat), 7) leg-rest, 8) footrest, 9) front caster, 10) front wheel, 11) motor, 12) side panel, 13) anti-tipping caster, 14) rear wheel, 15) batteries,16) seat frame, 17) liner actuator (back), 18) backrest frame

Table 2.1 EXB Wheelchair specification sheet

Weight including batteries	160 kg
Maximum user weight	120 kg
Battery type	12V-50 Ah*2 pcs
Charging time	10 h
Driving range	20~25 km
Maximum speed	6~8 km/h
Braking distance	115 cm
Turning radius	76 cm
Turning space,180°	122 cm
Gradient capacity	12°
Upholstery	fabric / foam
Seat height	58 cm
Seat Depth	45 cm
Seat width	42 cm
Back rest height	60 cm
Back rest width	42 cm
Back rest angle	100°～165°
Arm rest height	19～26 cm
Arm rest size (L x W)	39.5 cm * 8 cm
Footrest distance from seat pad	38 cm ~ 52 cm
Front wheel	20 cm (6")
Rear wheel	34.5 cm (13.6")

2.2. Real-time controller

In this study, we selected a controller suitable for vocal cord controller design. The basic requirement of the controller is to resist and reduce the interference of vibration and higher temperature. The volume size of the controller follows the maxim: the smaller the better, while for the response time of the signal process, it is: the sooner the better. The CompactRIO of National Instruments (NI, Austin, US) is suitable for the abovementioned requirements. Following are the relative

specifications.

Table 2.2 Specification sheet of CompactRIO (http://www.ni.com/cRIO)

Controller	NI cRIO- 9014	128MB DRAM, 2GB
Chassis	NI cRIO- 9103	4-slot, 3M gate
I/O module	Digital Input: NI 9411	6-Channel, 500ns, ±5 to 24V
	Digital Output: NI 9474	8-Channel 5 to 30 V, 1 µs
	Analog Input: NI 9215	4-Channel, 100 kS/s, 16-bit, ±10 V
	Analog Output: NI 9263	4-Channel, 100 kS/s, 16-bit, ±10 V

Figure 2.2 CompactRIO externals

A rapid and real-time processor with FPGA system, NI CompactRIO System (see Table 2.2 and Fig. 2.2), is used as the developmental platform. Based on its flexibility, small size, wide working temperature range, and rugged form, we could add extra purpose I/O modules in CompactRIO. In addition, when the wheelchair moves, CompactRIO can perform normally under impact vibration. Most important, CompactRIO has a real-time processor; it can rapidly analyze the analog waveform generated by the user and reactively respond with the desired wheelchair operations.

The main characteristics of cRIO are as follows:

1. High Scalability: There are 4-slot reconfigurable I/O module and hot-swappable

industrial I/O modules with built-in signal conditioning for direct connection to sensors and actuators.

2. Small volume: The dimensions are 179 mm * 88.1 mm * 88.1 mm.

3. High stability: The operating vibration and hock are up to 5 g_{rms} and 50 g, respectively.

4. The execution of precise control, data recording and powerful analysis: The 400 MHz real-time processor with 25 ns timing/triggering resolution can process data rapidly and an analog input acquisition rate up to 1 MHz

5. Wide operating temperature range: The operating temperature is from -40 to 70 °C (-40 to 158 °F).

6. Ethernet port: It is 10/100 Mb/s for programmatic communication over the network, built-in web and file servers.

Based on the above, with NI CompactRIO, you can rapidly build embedded control or acquisition systems that rival the performance and optimization of custom-designed hardware circuitry. FPGA devices are widely used by control and acquisition system vendors for their performance, re-configurability, small size, and low engineering development costs. The systemic architecture of CompactRIO is shown in Fig. 2.3.
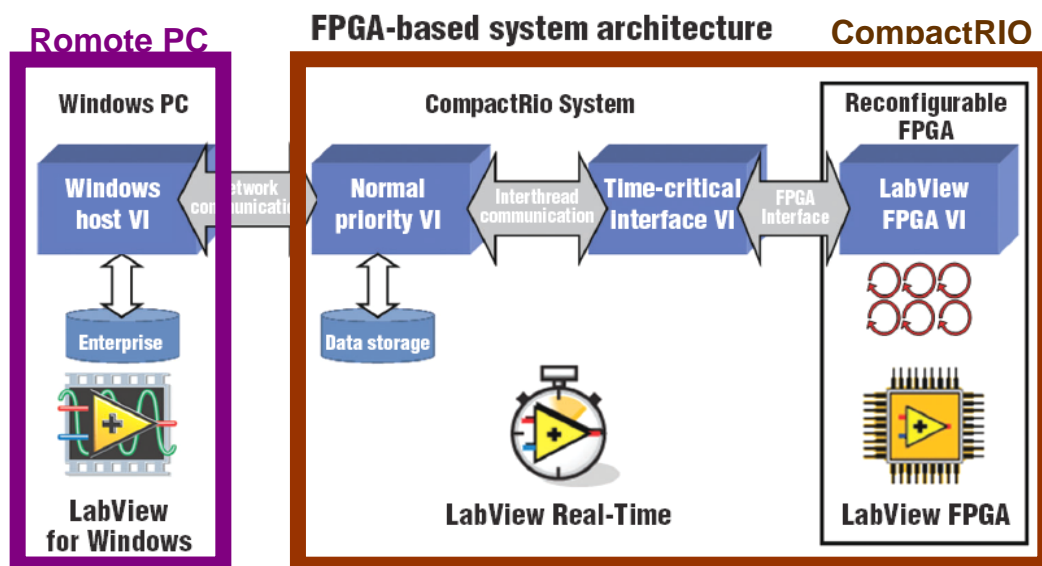
Figure 2.3 cRIO architecture

2.3. Speech recognition

Our speech recognition algorithm is based on [9] and [10]. We developed the application program with LabVIEW. Speech recognition is the key tool that we used, so we first needed to understand the basics of speech recognition theories. If the application program could distinguish the instruction/command accurately and rapidly to a level that we could accept would be considered by us as a sufficient achievement.

Any signal can be expressed in the form of waves. Waveform includes: amplitude, period, frequency, etc. Voice is composed of signals with different cycles. Voice is the reference to a time-varying or non-stationary signal. A stationary signal is simple, like: sin, cos, etc, but voice is a very complex signal; extracting features is not a simple matter.

Voice is transmitted by sound waves in the air; we call this wave: analog signal. Computers cannot deal directly with analog signal. All data are expressed in 0's or 1's in the computer. Signal is expressed in numerical as digital signal. If you use a computer to deal with analog signal, the analog signal must be converted into a digital signal, a process known as digitization.

The maximum frequency of speech will not exceed 4000 Hz, according to Nyquist-Shannon's sampling theorem; the sampling frequency should be at least twice the highest frequency contained in the signal. The Nyquist-Shannon sampling theorem is presented as:

$$f_s \geq 2f_c \qquad (2\text{-}1)$$

where f_s is the sampling frequency and f_c is the highest frequency of signal. In general, the quality of 'twice the highest frequency' is not good enough to interpret the original analog signal, but is sufficient for speech recognition [9] [10]. The signal, at least, should not be distorted.

The main purpose of speech recognition here is to enable the computer to recognize the human motivation, and then to obey the command issued by speech. A suitable recognition algorithm will display the ability to capture the main features of speech. Based on feature vectors, we can recognize human speech commands; the process of speech recognition includes: input of speech signal, extracting speech features, constructing each instruction, comparing the modeling data, and output of the recognized results.

The process of speech recognition is as follows:

```
┌─────────────────────────────────────────┐
│      Input speech signal (analog)        │
└─────────────────────────────────────────┘
                    │
                    ▼
┌─────────────────────────────────────────┐
│          Extract speech feature          │
└─────────────────────────────────────────┘
                    │
                    ▼
┌─────────────────────────────────────────┐
│  Construct each instruction (speech model)│
└─────────────────────────────────────────┘
                    │
                    ▼
┌─────────────────────────────────────────┐
│        Compare the modeling data         │
└─────────────────────────────────────────┘
                    │
                    ▼
┌─────────────────────────────────────────┐
│       Output the recognized result       │
└─────────────────────────────────────────┘
```
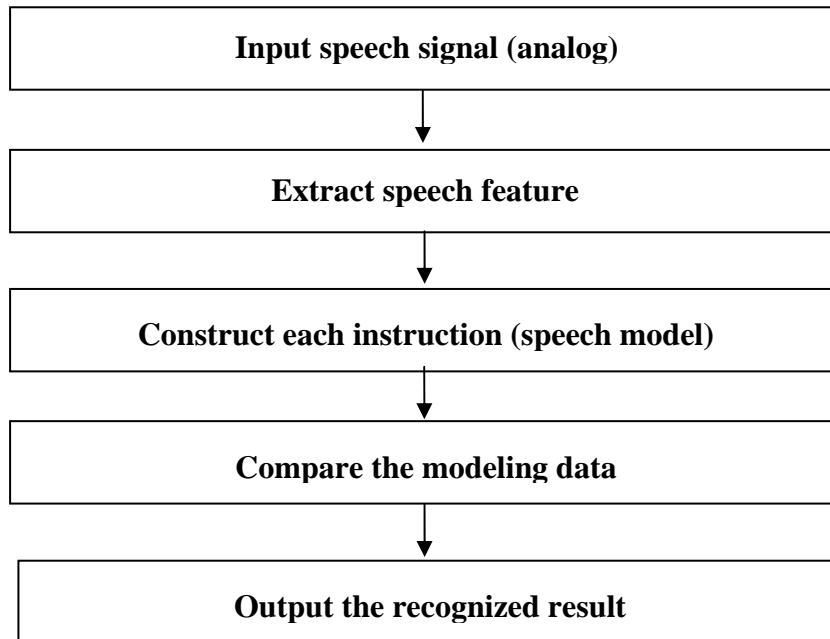
Figure 2.4 Speech recognition steps

Each instruction can be extracted and constructed in different models. In our study, five instructions were adopted for modeling: forward, backward, left, right and stop. When an instruction (analog speech signal) is acquired by the CompactRIO, an instruction, after feature extraction, is constructed as modeling data for comparison with five models, and the recognized result output as evaluating score. Obtaining the highest score is our expectation.

2.3.1. Feature extraction

The processes of feature extraction are as follow:

```
┌─────────────────────────────────────────────────┐
│        Segment speech into a series of frames     │
└─────────────────────────────────────────────────┘
                        │
                        ▼
┌─────────────────────────────────────────────────┐
│     Do pre-emphasis and hamming window with frame │
└─────────────────────────────────────────────────┘
                        │
                        ▼
┌─────────────────────────────────────────────────┐
│                 Auto-coefficient                  │
└─────────────────────────────────────────────────┘
                        │
                        ▼
┌─────────────────────────────────────────────────┐
│            Linear predictive coefficient          │
└─────────────────────────────────────────────────┘
                        │
                        ▼
┌─────────────────────────────────────────────────┐
│                     Cepstrum                      │
└─────────────────────────────────────────────────┘
                        │
                        ▼
┌─────────────────────────────────────────────────┐
│                  Delta-cepstrum                   │
└─────────────────────────────────────────────────┘
                        │
                        ▼
┌─────────────────────────────────────────────────┐
│   Speech feature obtained (cepstrum + delta-cepstrum)│
└─────────────────────────────────────────────────┘
```
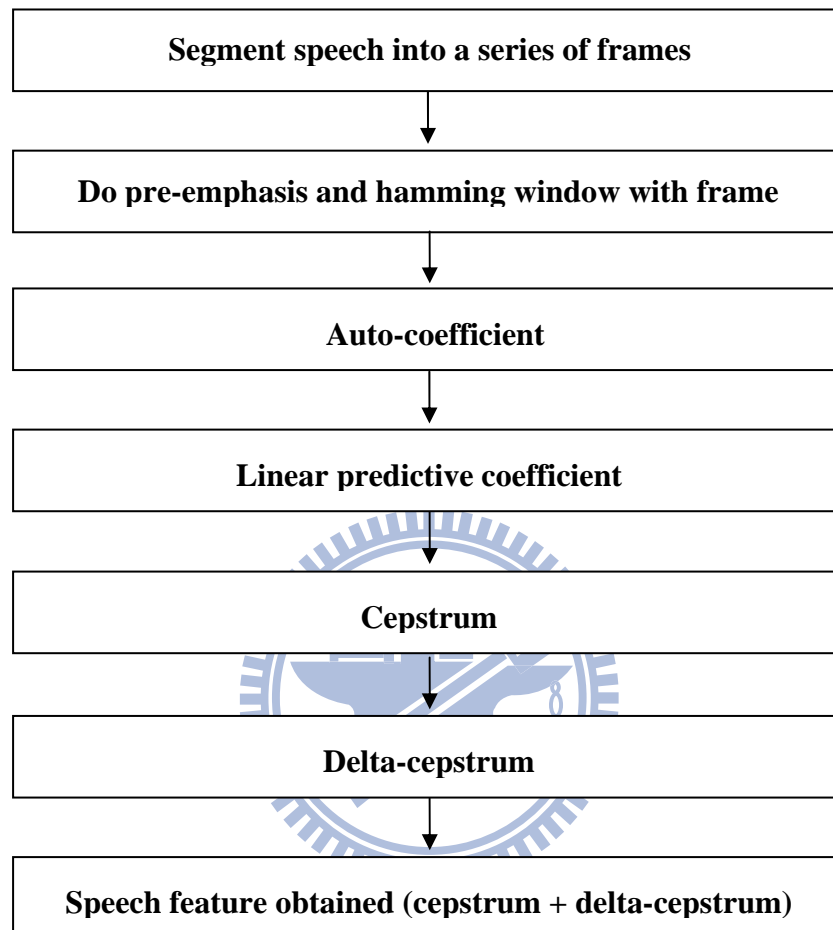
Figure 2.5 Feature extraction steps

1. Frame blocking:

The time duration of each frame is about 20-30 ms; if the frame duration is too large, we can't catch the time-varying characteristics of the speech signals and lose the function of frame blocking. If the frame duration is too small, then we can't extract valid speech features, and computation time is long. Usually the overlap is 1/2 to 2/3 of the original frame [9] [10]. The more overlapping, the more computation is needed. In our example, the speech sampling rate is 10000 sample/sec; frame size is set with 240 sample points, frame duration is 240/10000 =0.24 ms between 20 ms-30 ms; and the overlap of two frames is 80 sample points between 1/2 and 2/3.

2. Pre-emphasis:

The goal of pre-emphasis is to compensate for the suppression of the high-frequency production by human sound mechanisms, often by boosting the higher frequencies. We emphasize the high frequency of 9500 Hz. The formula of pre-emphasis is presented as:

$$S[n] = X[n] \qquad\qquad n = 0$$

$$S[n] = X[n] - \frac{9500}{10000} \times (n-1) \qquad 0 < n < frame\ size \qquad (2\text{-}2)$$

3. Hamming window:

In order to concentrate frame energy, frames have to be multiplied with a hamming window. The formula of the hamming window is written as:

$$\eta[n] = 0.54 - o.46\cos\left[\frac{2\phi n}{N-1}\right] \qquad 0 \le n \le N-1 \qquad (2\text{-}3)$$

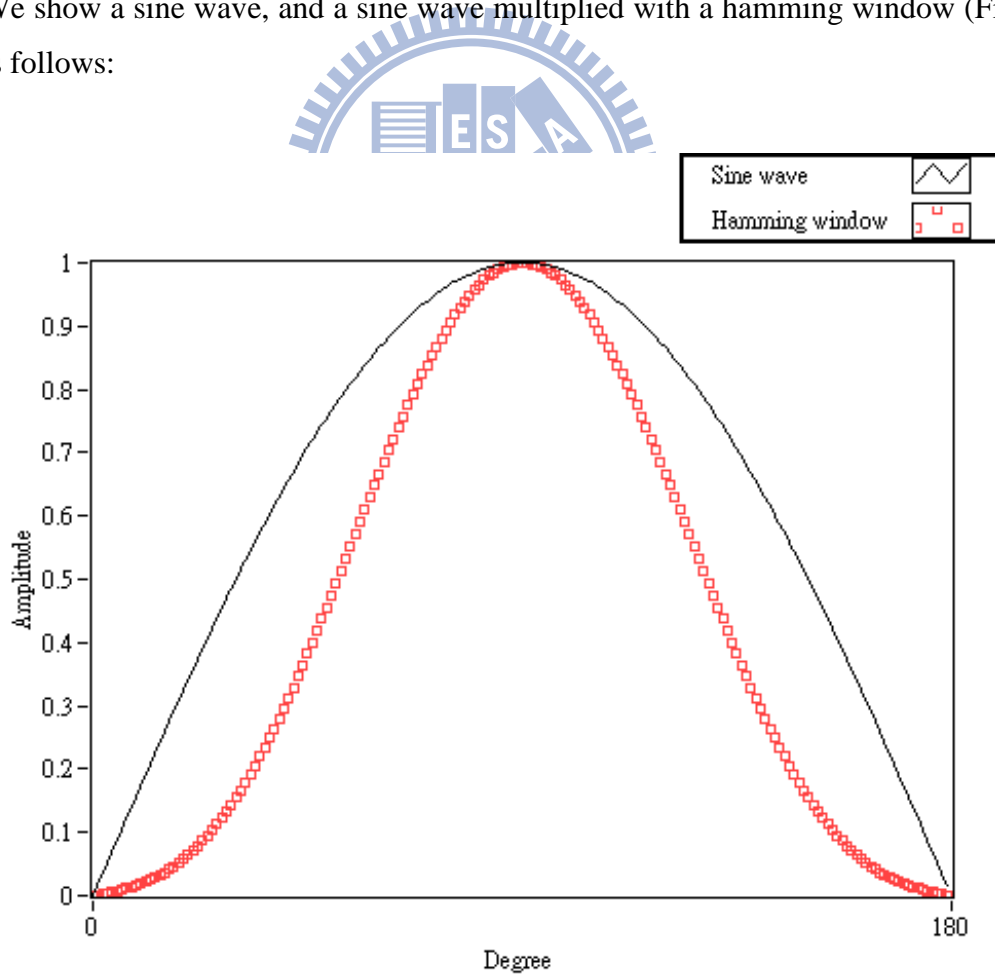We show a sine wave, and a sine wave multiplied with a hamming window (Fig. 2.6) as follows:

Figure 2.6 Sine wave and sine wave after hamming window

4. Linear predictive coefficient (LPC)

LPC is a very important feature of speech; it is the first feature we obtained. The main purpose of LPC is that a speech sample can be predicted by linear combination of previous p samples, as follows:

$$S(n) = A1 \times S(n-1) + A2 \times S(n-2) + \cdots Ap \times S(n-p)$$

S(n) is a predictive sample

$S(n-1) \sim S(n-p)$ is S(n) preceding p speech samples

A1~Ap are coefficients of linear combination   (2-4)

The real signal and the predictive signal may signify an error; if the error is minimal, the coefficient of linear combination is a linear predictive coefficient. The formula regarding the difference of the real signal and the predictive signal is as follows:

$$E = \sum e[n]^2$$
$$= \sum \left[\bar{y}[n] - \sum_{k=1}^{p} a[k]\bar{y}[n-k] \right]^2 \qquad (2\text{-}5)$$

First, we compute the auto-correlation. R[i] is the auto-correlation coefficient. The order of linear predictive coefficient we set is 12; usually, speech recognition uses 8 to 14 orders, as follows:

$$R_i(k) = \sum_{n=0}^{N-i-1} X(n+i)X(n+i+k), 0 \le k \le N-1$$
$$= X(i)X(i+k) + X(i+1)X(i+k+1) + \cdots + X(i+N-1X(i+N-1+k))$$

(2-6)

Since we used 12 linear predictive coefficients, we computed R0 (0) to R0 (12). After we obtained the auto-correlation, we used Durbin's algorithm to solve an inverse matrix problem. After following the five steps, we can obtain the LPC

Step1：$E(0) = R(0)$

Step2：$K_i = \left[R(i) \sum_{j=1}^{i-1} a_j^{i-1}R(i-j) \right] / E(i-1)$

Step3：$a_i^i = K_i$

Step4：$a_j^i = a_j^{i-1} - Kia_{i-j}^{i-1}, 1 \le j \le i-1$

Step5：$E(i) = (1-k_i^2)E(i-1), a_j = a_j^p$

5. Cepstrum

After obtaining LPC, we use LPC with recursions to obtain cepstrum. This way can avoid complex computing. Formula (2-7) shows that how to get cepstrum features.

$$C1 = a1$$

$$Cn = an + \sum_{m=1}^{n-1} \left(1 - \frac{m}{n}\right) \times a_m \times C_{n-m}, 1 < n \leq p$$

$$Cn = \sum_{m=1}^{p} \left(1 - \frac{m}{n}\right) \times a_m \times C_{n-m}, n > p \qquad (2-7)$$

The main functions of cepstrum are representing spectrum peaks and less variation of speech features. A cepstrum is the result of taking the Fourier transform of the decibel spectrum as if it were a signal. Cepstrum could stand for speech recognition feature than LPC.

6. Delta-cepstrum

After doing partial differential with cepstrum is delta-cepstrum, it has the ability to resist noise. In our program, we chose the order of 14 cepstrum and delta-cepstrum as a feature vector. In the program implementation, we used the following formula to obtain delta-cepstrum parameters. The formula has 0~L-1 frames, C(m,n) as mth frame, and nth delta-cepstrum; the formula is as follows:

$$\Delta Cn = [2 \times C(2,n) + C(1,n)]/5 \qquad \text{For No. 0 frame}$$

$$\Delta Cn = [2 \times C(3,n) + C(2,n) - C(0,n)]/6 \qquad \text{For No.1 frame}$$

$$\Delta Cn = [2 \times C(L-1,n) - C(L-3,n) - 2C(L-4,n)]/6 \qquad \text{For No.L-2 frame}$$

$$\Delta Cn = [2 \times C(L-1,n) - C(L-3,n) - 2C(L-4,n)]/6 \qquad \text{For No.L-1 frame}$$

$$\Delta Cn = [2 \times C(i+2,n) + C(i+1,n) - C(i-1,n) - 2C(i-2,n)]/10$$

$$\text{For other frames}$$

$$(2-8)$$

2.3.2. Training model (Construct speech model)

After obtaining the speech feature, we constructed speech models. First, we discussed the nature of states. State function is catching the variations in regard to the mouth. For example, for the word *forward* in Mandarin, the pronunciation is 'ch'/ 'yi'/ 'an', it must have some frames belonging to 'ch', some belonging to 'yi' and the

remainder frames belonging to 'an'. If we construct three states, the first state has 'ch' frames, the second has 'yi' frames and third has 'an' frames. We can also construct two or six states or other states. The graph of *forward* in Mandarin; the signal is composed of "ch", "yi" and "an" (see Fig. 2.7).
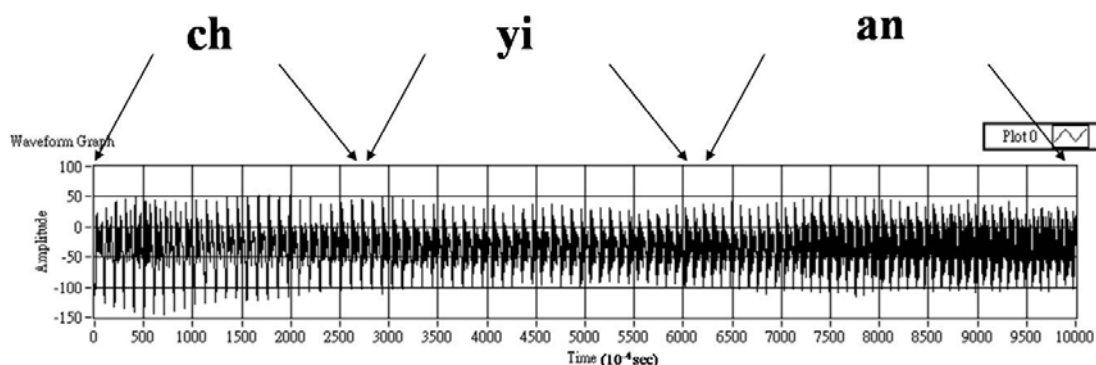

Figure 2.7 Signal intensity graph in time-domain of *forward* in Mandarin

When we start to construct a model, we have to initial the relations with states and frames. Mean cut is a simple way to initial the relations. We use three states to illustrate mean cut, as follows:
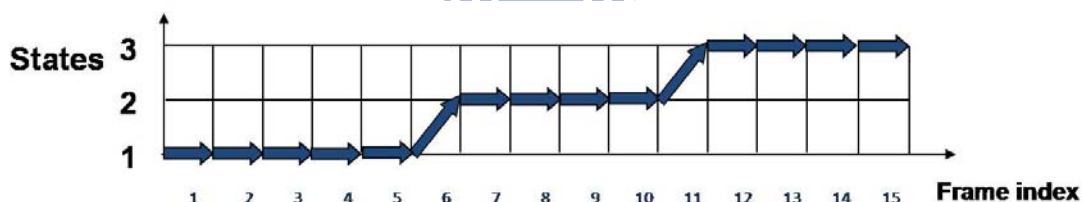

Figure 2.8 Mean cut frames to states

Hidden Markov Models (HMM) are recognition models using probability statistics; the most salient characteristic of HMM is using two probability density functions to describe variations of speech signal: one is state transition probability and the other is state observation probability. HMM is the most important component of the training model. The steps of the training model using HMM are as follows:
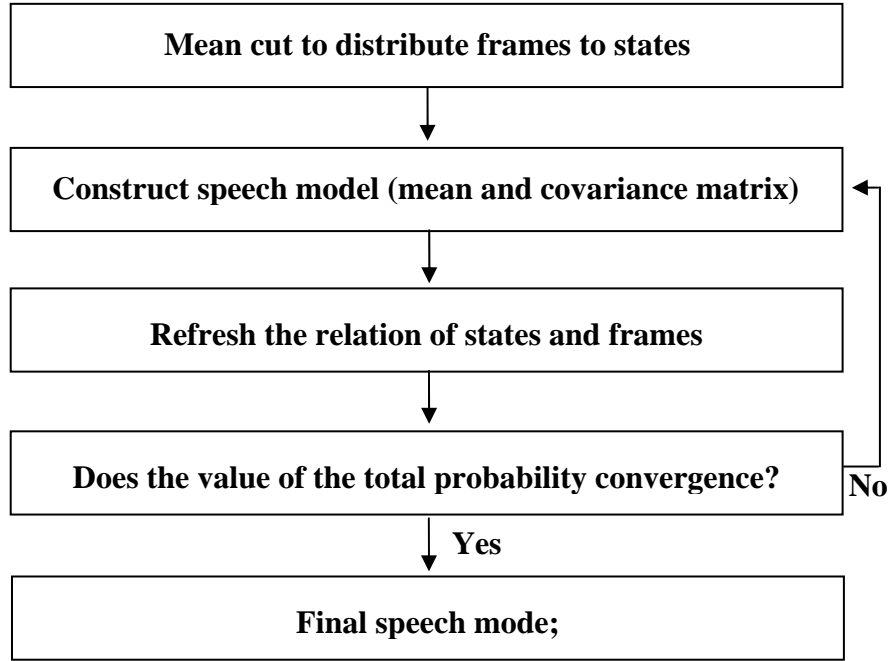
```
┌─────────────────────────────────────────────────────┐
│        Mean cut to distribute frames to states        │
└─────────────────────────────────────────────────────┘
                          │
                          ▼
┌─────────────────────────────────────────────────────┐
│   Construct speech model (mean and covariance matrix) │◄──┐
└─────────────────────────────────────────────────────┘   │
                          │                                │
                          ▼                                │
┌─────────────────────────────────────────────────────┐   │
│        Refresh the relation of states and frames      │   │
└─────────────────────────────────────────────────────┘   │
                          │                                │
                          ▼                                │
┌─────────────────────────────────────────────────────┐   │
│   Does the value of the total probability convergence?│───┘ No
└─────────────────────────────────────────────────────┘
                          │ Yes
                          ▼
┌─────────────────────────────────────────────────────┐
│                  Final speech mode;                   │
└─────────────────────────────────────────────────────┘
```

Figure 2.9 Training model steps

The restrictions of the relation of frames and states are as listed below:

1. Each state has at least a frame.

2. The distribution that dividing frames to states cannot be reversed. For example, if frame No. 10 belongs to state 2, it is impossible for frame No. 15 to belong to state 1.

At the step for computing the mean and covariance matrix, we should know which states are similar to which frames. A formula is chosen, Gaussian probability density, to compute the degree of similarity between frame and state. The formula is as follows:

$$D_{i(\tau_T)} = (2\pi)^{-\frac{N}{2}}|R_i|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}(\tau_T - \tau_{Ri})^T R_i^{-1}(\tau_T - \tau_{Ri})\right]$$

N: Feature vector dimensions

τ_T : Speech of training data feature vector

τ_{Ri}: The ith mixture expectation (mean) of state in the model

Ri: The ith mixture covariance matrix in the model   (2-9)

At the step of refreshing the relations between states and frames, we used the Viterbi algorithm, implemented by dynamic programming (DP), a famous algorithm. The purpose of the Viterbi algorithm is to optimally distribute frames to states and achieve probability convergence (to get the maximum value of probability). The

Viterbi algorithm path constraint illustration is shown in Fig. 2.10 and the solution of the Viterbi algorithm shows the formula (2-10); the solution is a recursive method.
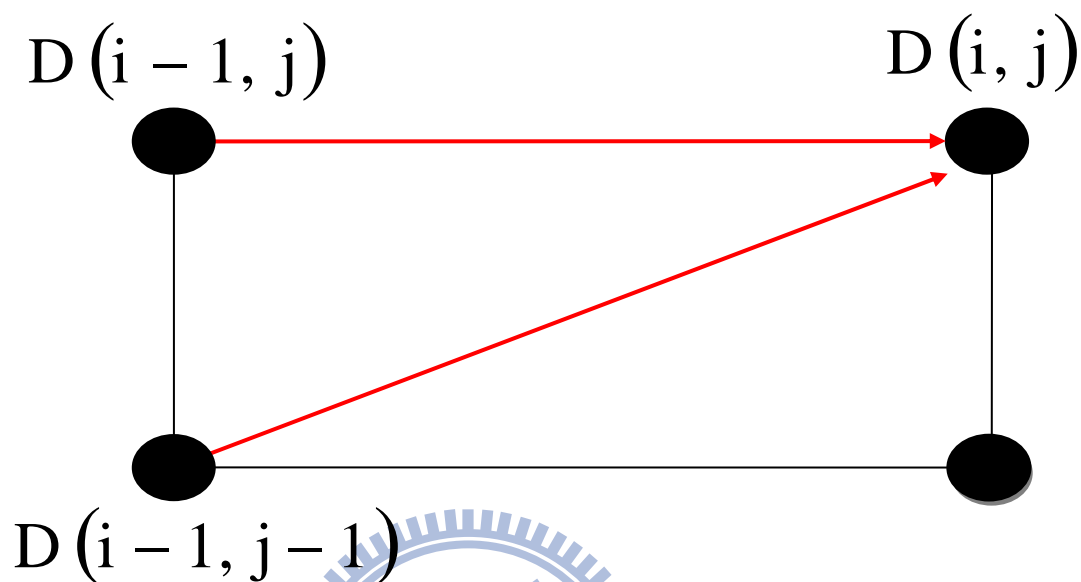
$$D(i-1,j) \qquad\qquad\qquad D(i,j)$$

$$D(i-1,j-1)$$

Figure 2.10 The path constraint of the Viterbi algorithm

$$D(i,j) = B(O(i),j) + \max\begin{cases} D(i-1,j) + A(j,j) \\ D(i-1,j-1) + A(j-1,j) \end{cases} \qquad (2\text{-}10)$$

After the above processing, the relation of frames and state is not mean cut. It may be shown as follows:
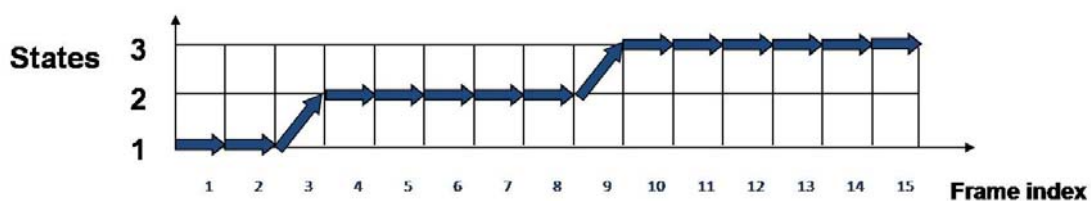
Figure 2.11 Final relations of frames and states

The first state has frames No. 1 to 3, the second state has frames 4 to 9, and the third state has the remaining frames.

HMM is a useful tool. We describe it roughly. More information on it can be searched.

2.3.3. Data comparison

After the above processing, feature extraction and training model, an operation data comparison follows. The process of data comparison is as follows:

Command an instruction

↓

Extract speech feature

↓

Find matched data with speech models

↓

Recognition outcome (the maximum score)
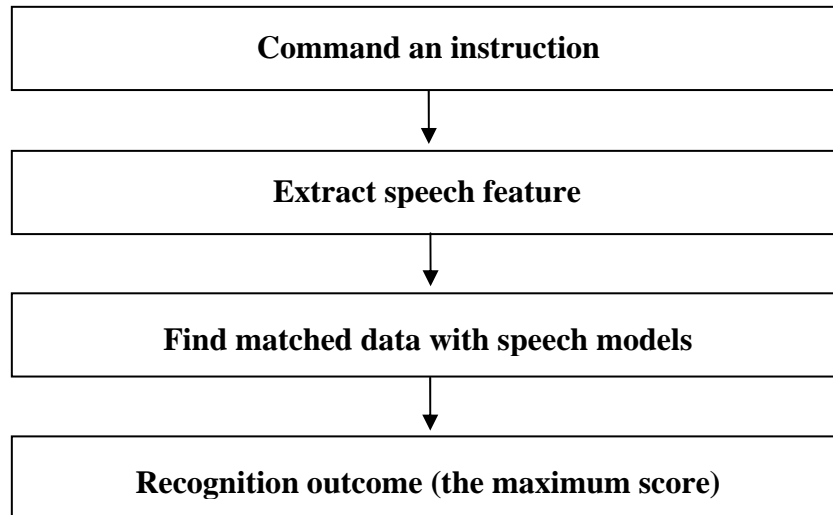
Figure 2.12 Data comparison steps

A user commands an instruction to our controller, and the controller executes feature extraction. Then the feature vector of instruction operates with speech models. In our example, we constructed five models: forward, backward, left, right and stop. Operating the speech models produced five scores; the maximum score is the outcome.

2.4. System architecture

The architecture of this system includes wheelchair module, joystick control module, system control module, speech analysis module and bio-signal (vocal cords) control module. The system architecture is presented in Fig. 2.13. When we say *forward*, the vocal cords will vibrate, and the signal is transmitted to the cRIO; our data acquisition and speech recognition program are in the cRIO. After the cRIO analyses the data, the outcome is generated and the wheelchair is instructed where to go or to stop.

- Wheelchair module
- Joystick module
- System control module
- Speech analysis module
- Bio-signal control module
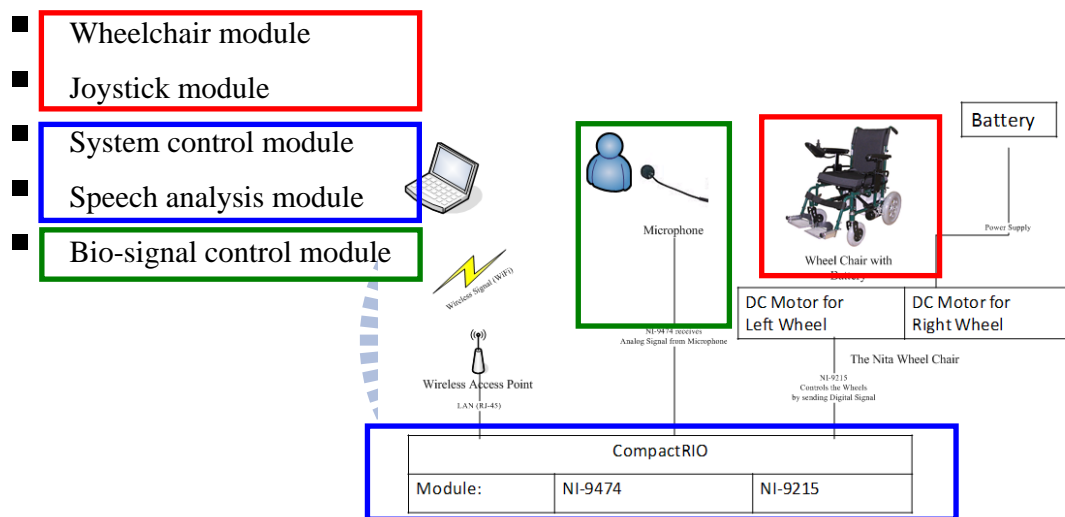
Figure 2.13 System architecture

When we say an instruction like forward in Mandarin or operate the joystick. The signal transmits to the cRIO. The cRIO receive the signal and analysis it through speech program in the cRIO. When the analysis is finished, it will output the outcome and output the control signal to the wheelchair. The Figure 2.14 shows the whole operating process.
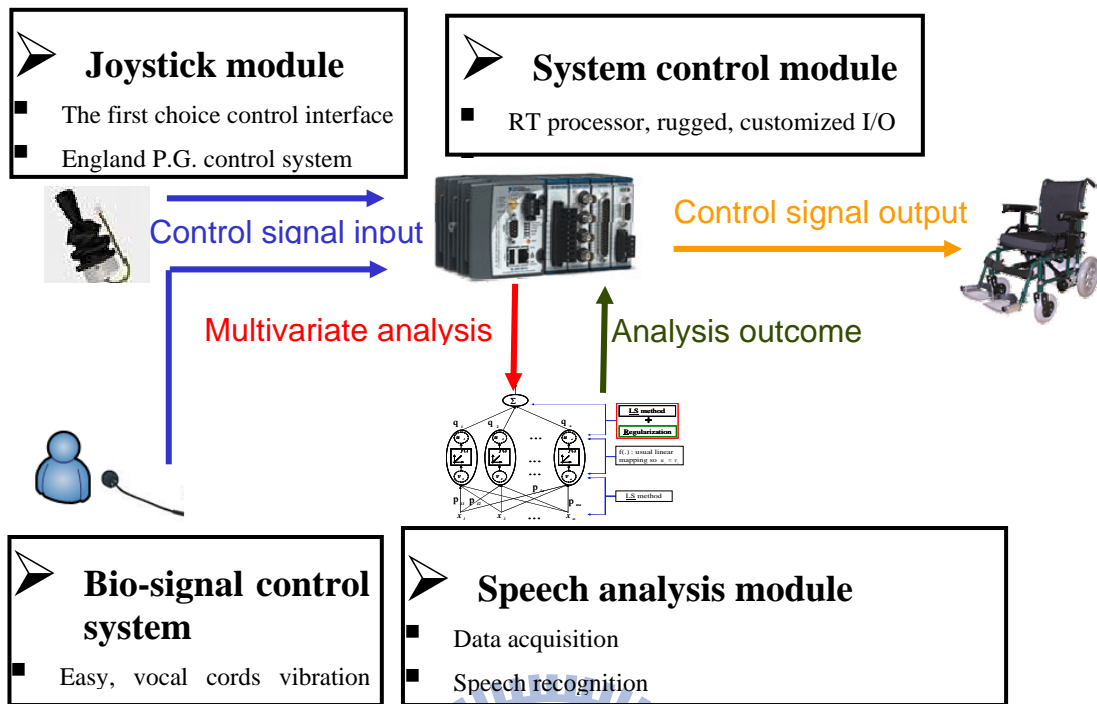
Figure 2.14 Operating processes of our system

Focusing on the controller, we present a diagram to explain how the controller operates internally, as follows.
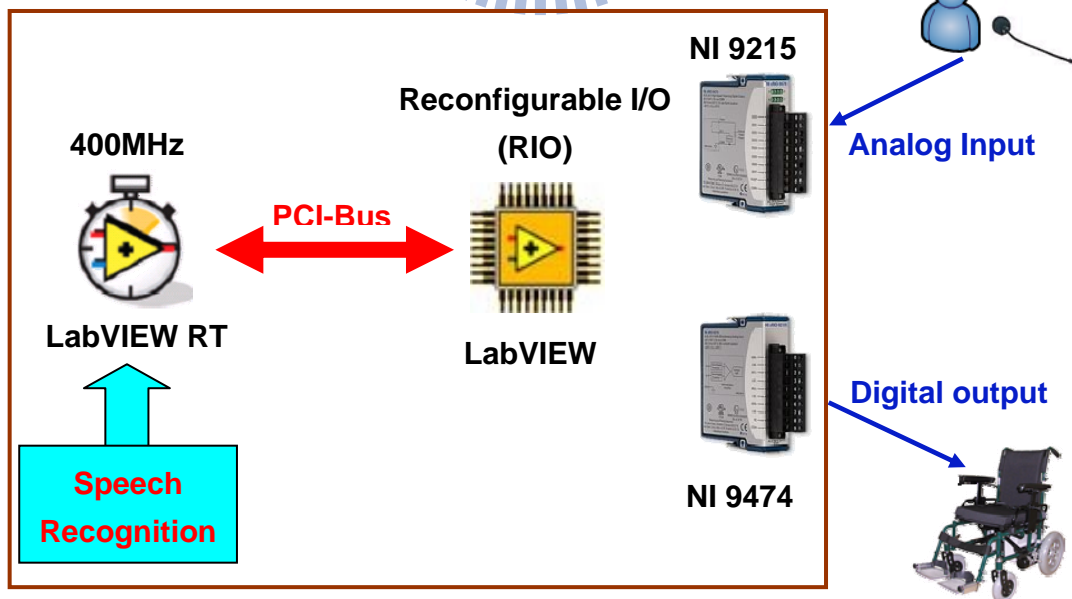
Figure 2.15 Operating process of controller

Chapter 3: Experiment and Result

3.1. Experiment background

We used LabVIEW software as our development tool. The software was installed in our research as follows.

Table 3.1 Experiment software and serial numbers

Software
NI LabVIEW Core Software 8.6
NI LabVIEW FPGA Module 8.6
NI LabVIEW Real-Time Module 8.6
NI-RIO 3.0.0
CompactRIO 3.0.1

Hardware equipments as Table 3.2

Table 3.2 Experiment's hardware equipment and serial numbers

Hardware
NI cRIO-9014
NI cRIO-9103
NI 9474
NI 9215
Nita EXB Standup Wheelchair
Carol Handheld Cardioids Dynamic Microphone GS-55

Our speech recognition settings are as follow [10]:

Table 3.3 Speech recognition settings

Parameter	Value	Description
Sampling rate	10000	The number of samples per second taken from a continuous signal to make a discrete signal
Frame size	240	The sampling points within each frame
Frame overlap	160	The sampling points of the overlap between consecutive frames
Frame rate	123	The number of frames per second, which is equal to the sample frequency divided by the frame step.

Our goal was to recognize the instructions commanded by the vocal cords. Since speech recognition is easier than vocal cords vibration recognition, we handled speech recognition first; if the outcome would be good, we would then attempt vocal cords vibration recognition and other methods.

We recorded 25 instruction sets; one set included: forward, backward, left, right and stop in speech. We wanted to determine the number of instruction sets to train models which would obtain the best correction rate in regard to program execution time, so we did the following:

Choose 1 instruction set in turn as training data, with the remaining 24 sets as testing data. If the experiment ends, it will produce 25 times (training data) × 24 sets (testing data) × 5 instructions (a set has 5 instructions) = 3000 testing times.

Choose 2 instruction sets in turn as training data, with the remaining 23 sets as testing data. If the experiment ends, it will produce 25 times (training data)× 23 sets (testing data) × 5 instructions (a set has 5 instructions) = 2875 testing times.

⋮

Choose 23 instruction sets in turn as training data, with the remaining 2 sets as testing data. If the experiment ends, it will produce 25 times (training data)× 2 sets (testing data) × 5 instructions (a set has 5 instructions) = 250 testing times.

Choose 24 instruction sets in turn as training data, with the remaining 1 set as

testing data. If the experiment ends, it will produce 25 times (training data)× 1 sets (testing data) × 5 instructions (a set has 5 instructions) = 125 testing times.

We show the speech signal in Mandarin, English and Fukienese, and vocal cord vibrations in Mandarin.
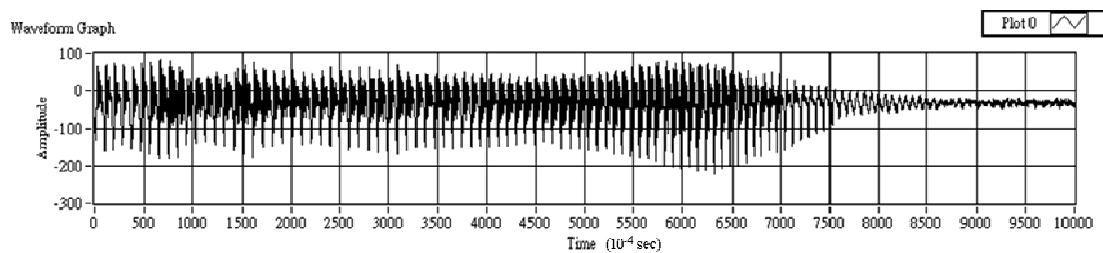

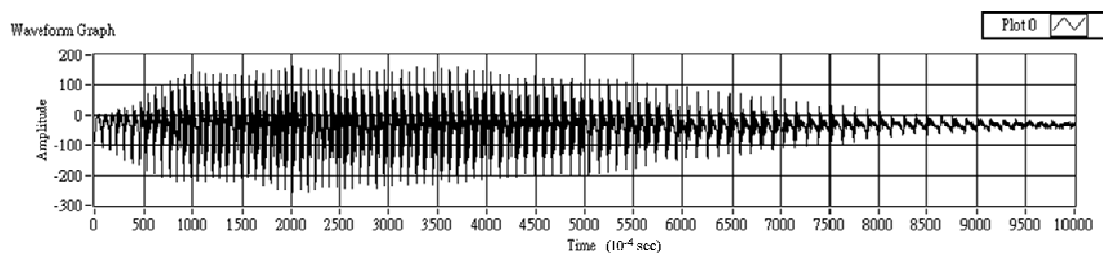Figure 3.1 Signal intensity graph of speech: *forward* in Mandarin

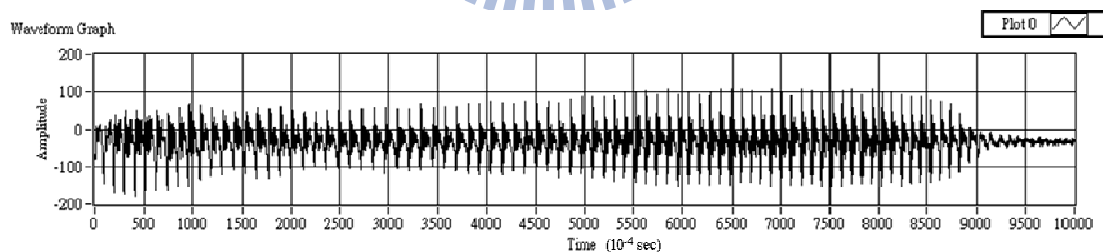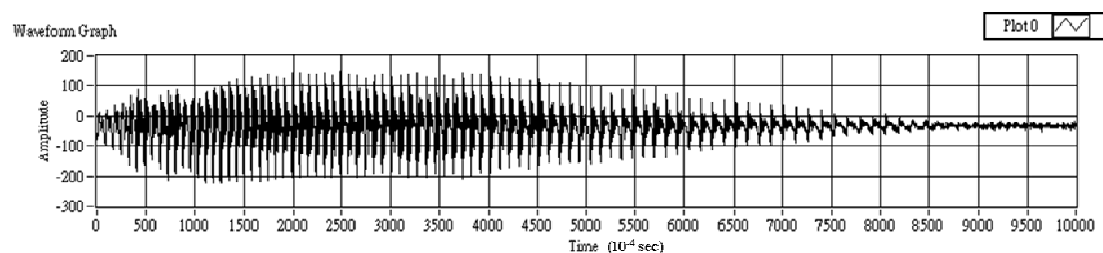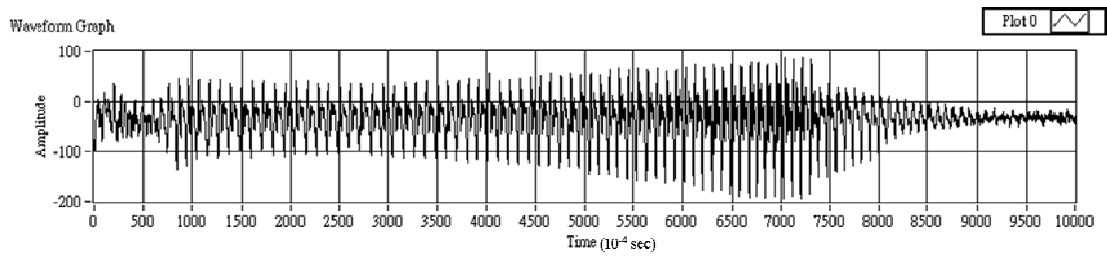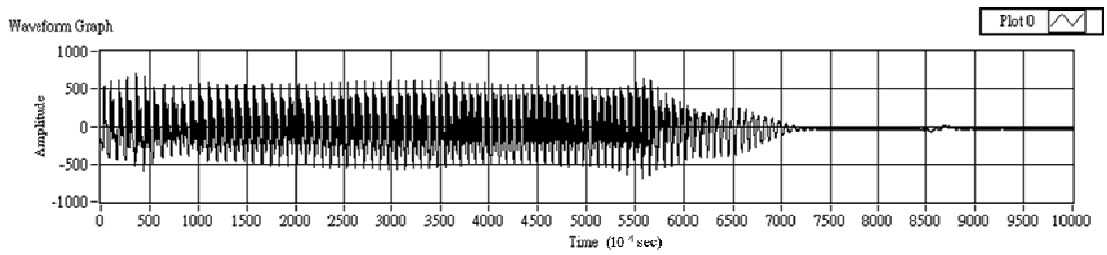Figure 3.2 Signal intensity graph of speech: *backward* in Mandarin

Figure 3.3 Signal intensity graph of speech: *left* in Mandarin

Figure 3.4 Signal intensity graph of speech: *right* in Mandarin

Figure 3.5 Signal intensity graph of speech: *stop* in Mandarin

Figure 3.6 Signal intensity graph of vocal cords vibration: *forward* in Mandarin
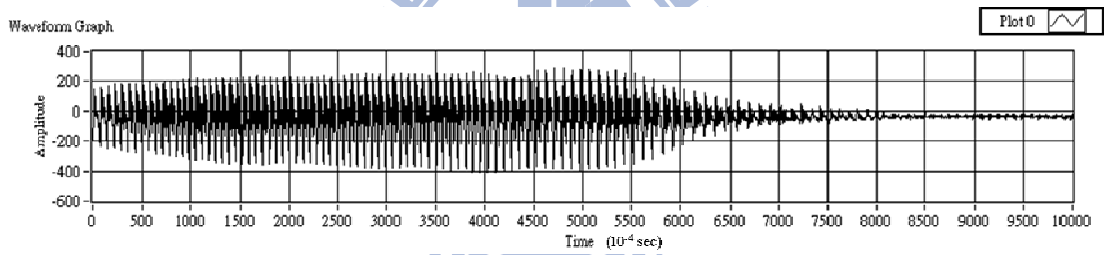
Figure 3.7 Signal intensity graph of vocal cords vibration: *backward* in Mandarin
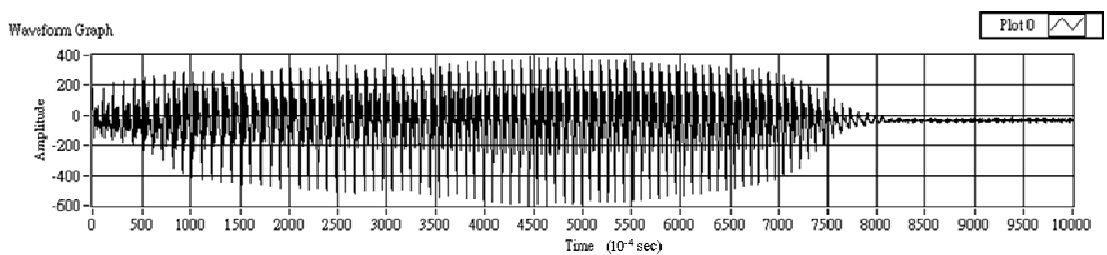
Figure 3.8 Signal intensity graph of vocal cords vibration: *left* in Mandarin
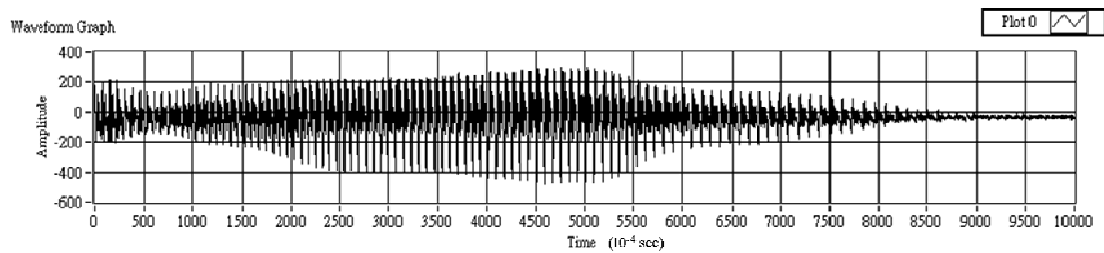
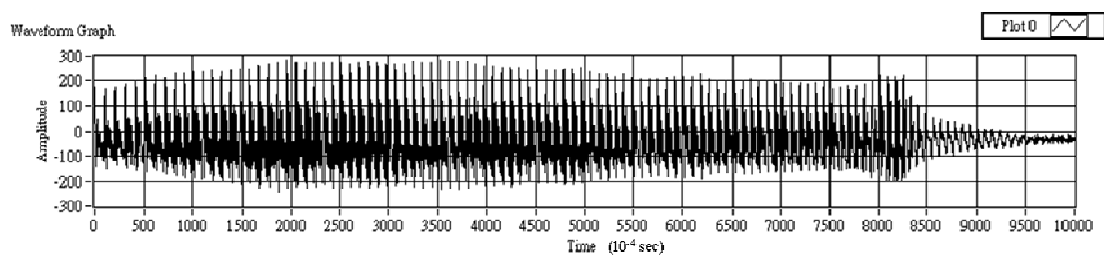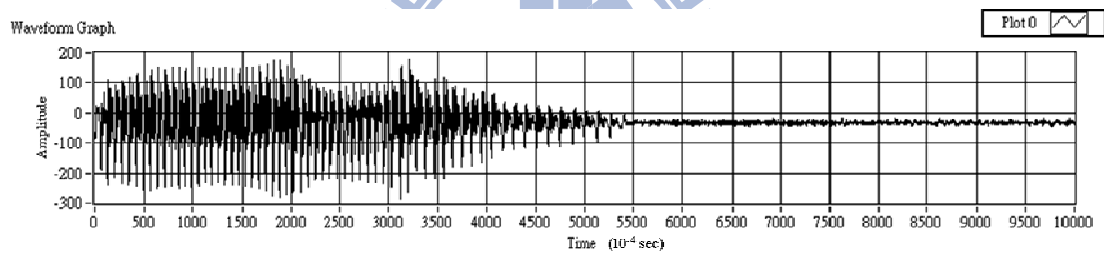Figure 3.9 Signal intensity graph of vocal cords vibration: *right* in Mandarin

Figure 3.10 Signal intensity graph of vocal cords vibration: *stop* in Mandarin

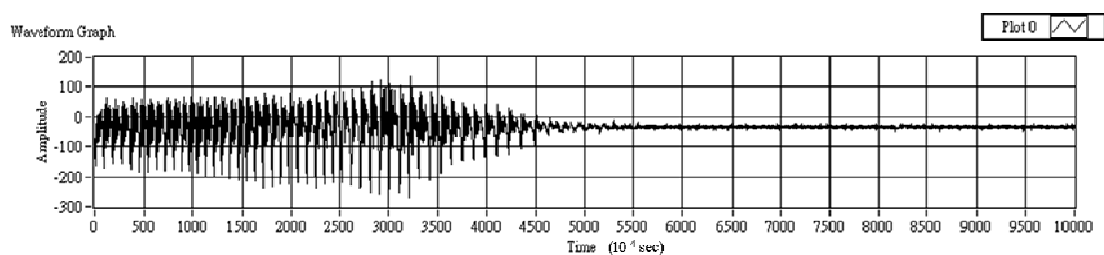Figure 3.11 Signal intensity graph of speech: *forward* in English

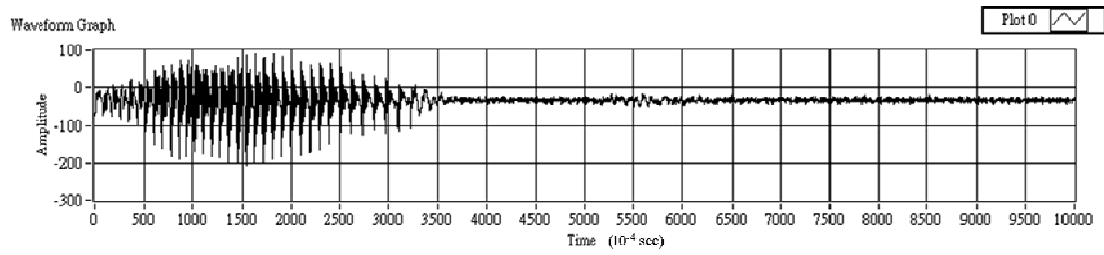Figure 3.12 Signal intensity graph of speech: *backward* in English

28

Figure 3.13 Signal intensity graph of speech: *left* in English
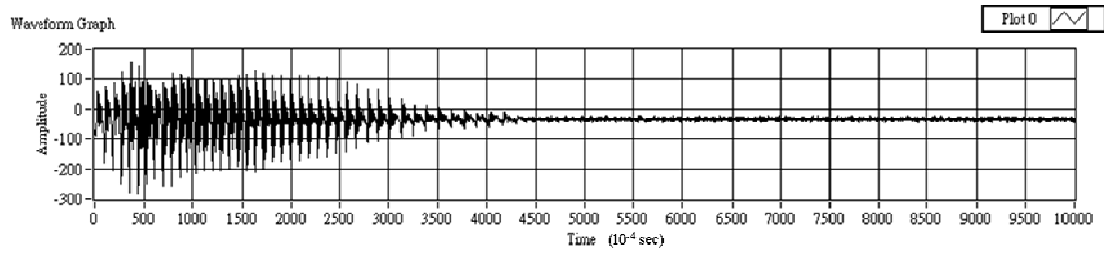

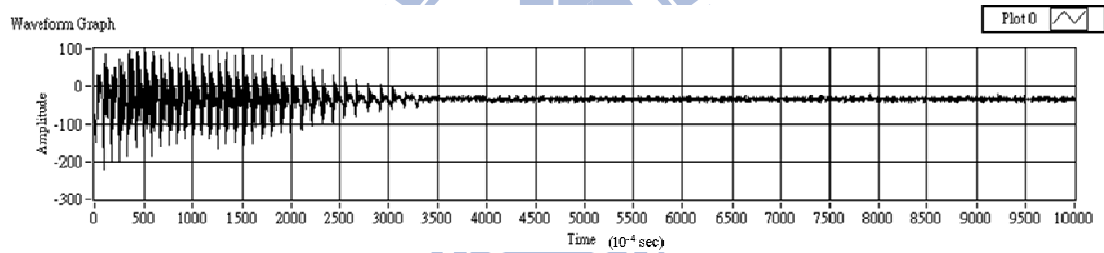Figure 3.14 Signal intensity graph of speech: *right* in English

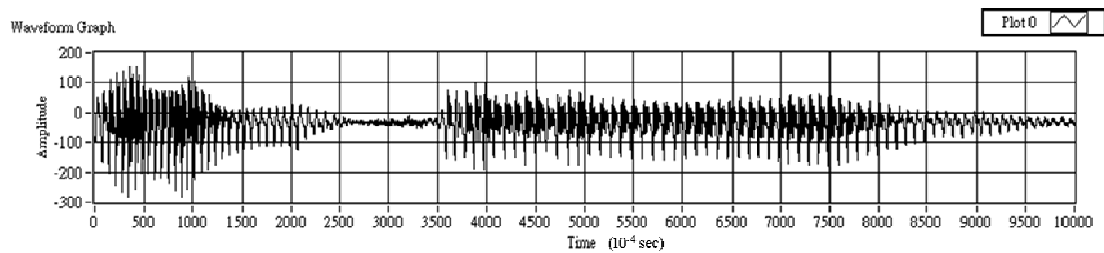Figure 3.15 Signal intensity graph of speech: *stop* in English

Figure 3.16 Signal intensity graph of speech: *forward* in Fukienese
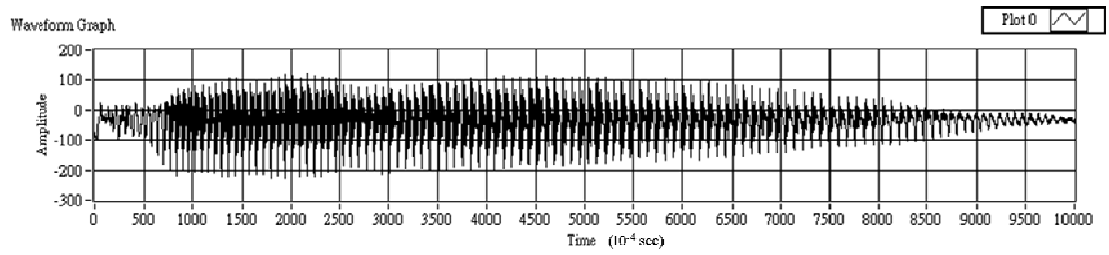
29

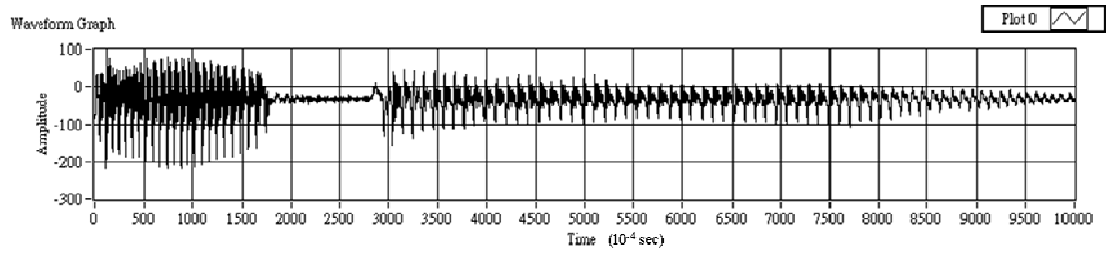Figure 3.17 Signal intensity graph of speech: *backward* in Fukienese

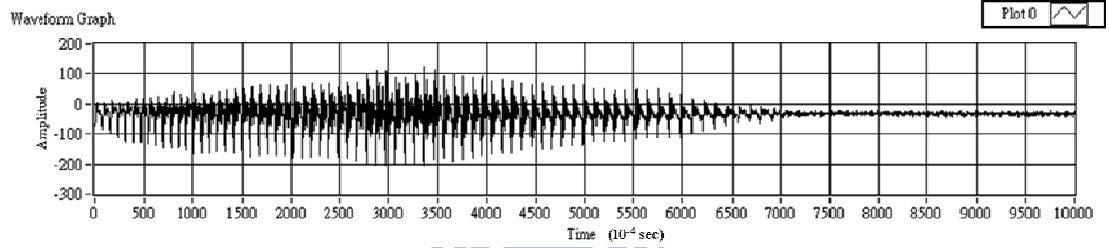Figure 3.18 Signal intensity graph of speech: *left* in Fukienese

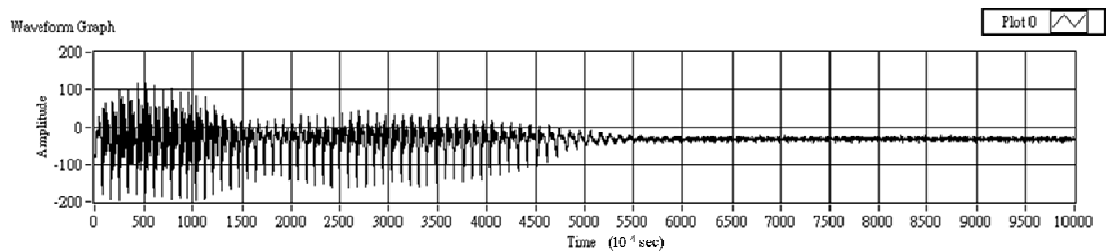Figure 3.19 Signal intensity graph of speech: *right* in Fukienese

Figure 3.20 Signal intensity graph of speech: *stop* in Fukienese

3.2. Speech recognition result

At first, our speech recognition focused on Mandarin instructions. We used 2, 3 and 6 states to train the model. The correct rates were as follows:
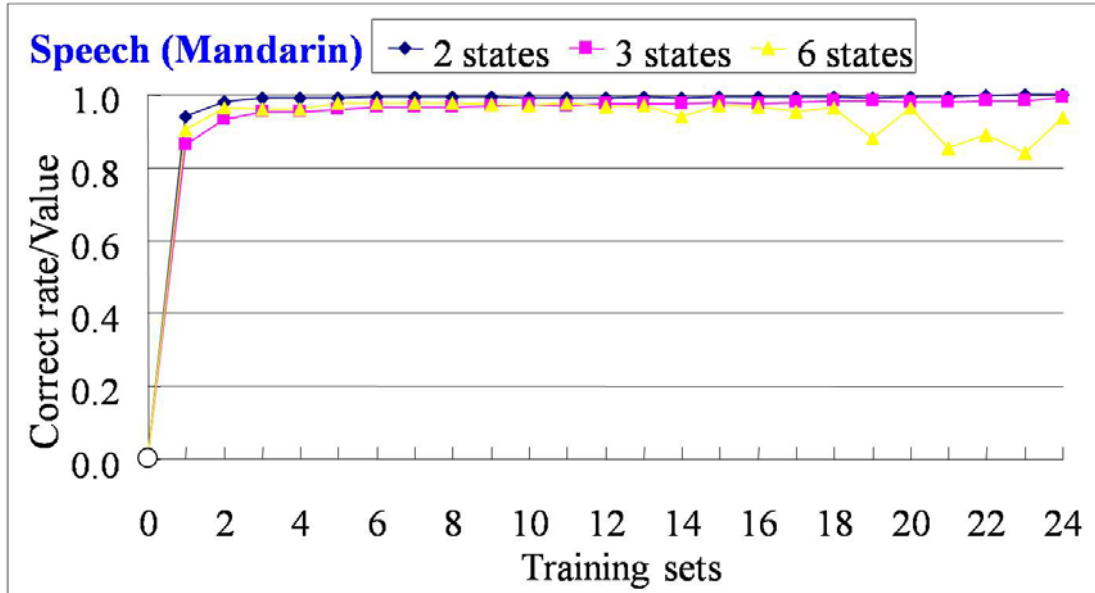
Figure 3.21 Speech in Mandarin recognition correct rate

The correct rate in this situation is good, at least 80%. In this figure, we see that at 3 training sets, the correct rate is at least 95% whatever the state chosen. The more training sets and states, the more time to train the speech model. In consideration of the program execution time and correct rate, we think 3 training sets is a better choice. The program execution time is not long and the correct rate is acceptable to users.

When we finished speech in Mandarin recognition, we became interested in vocal cord vibrations in Mandarin recognition; as a result, we recorded 25 instruction sets for experiments like the abovementioned. The correct rate result is shown in Fig. 3.22:

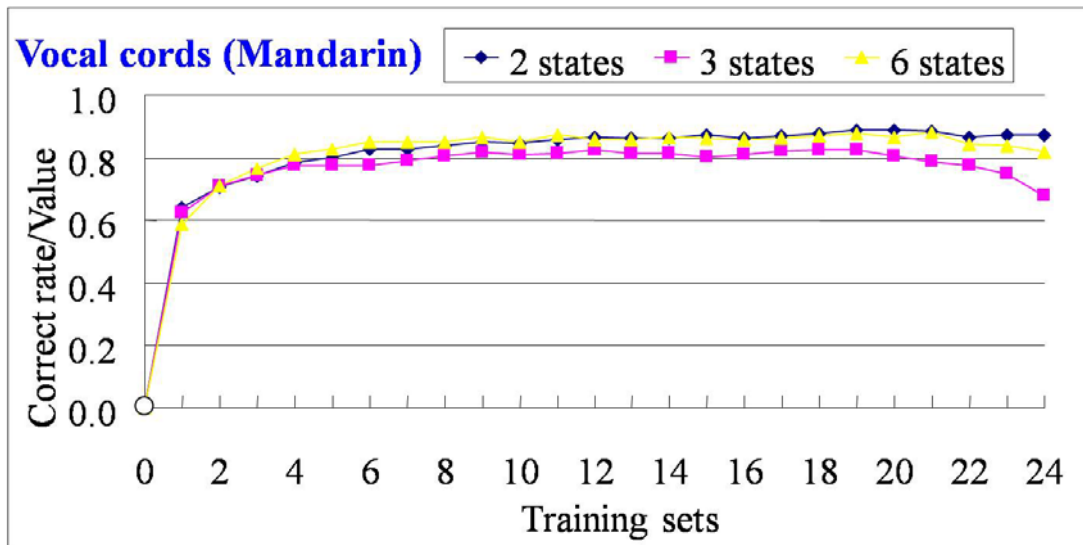Figure 3.22 Correct rates of vocal cord vibration recognition

The vocal cord correct recognition rate was not good enough for us. Whatever training sets and states were chosen, the correct rate was under 90%, which was worse than speech in regard to Mandarin recognition.

We also recorded speech in English and Fukienese (see Figs. 3.23 and 3.24)
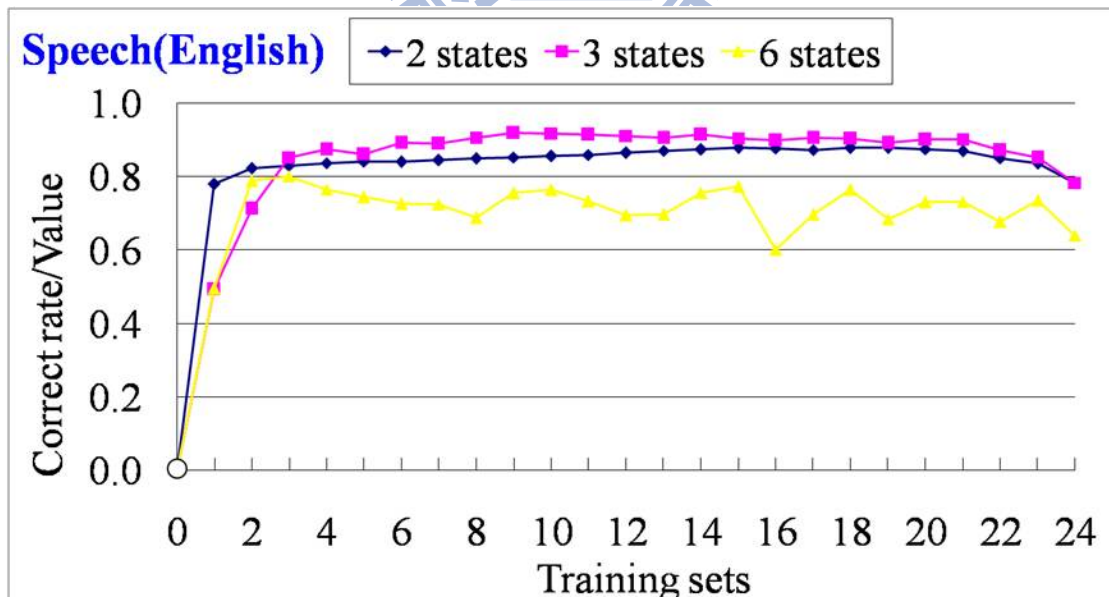

Figure 3.23 Correct rates of speech in English recognition
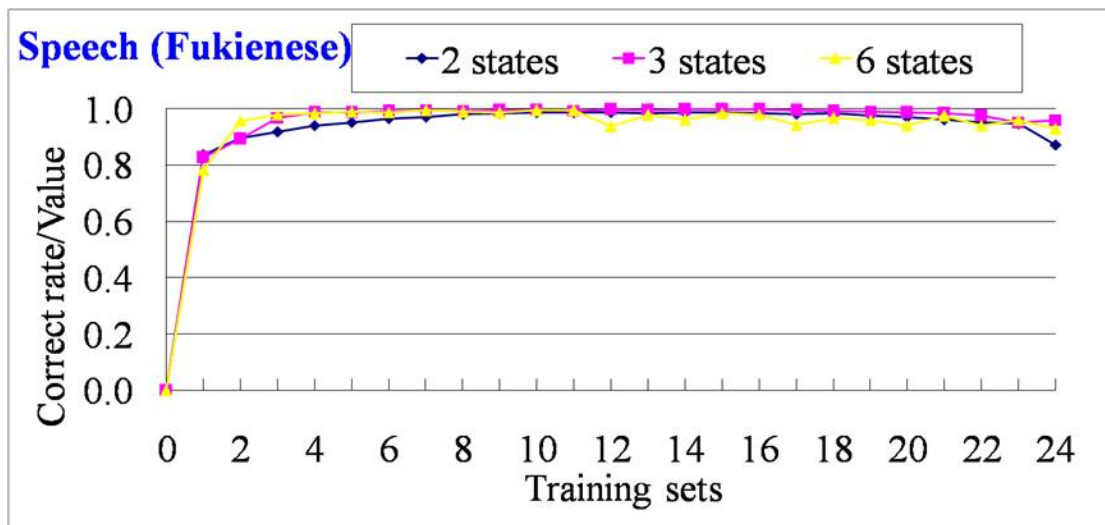
Figure 3.24 Correct rates of speech in Fukienese recognition

Speech in English is so interesting: 2 states and 3 states are better than 6 states. Above all, the correct rate is not so good. There are many possible reasons: record error, pronunciation error, and even the probability that the recognition algorithm is unsuitable for English speech. The answer to this will require future experiments.

Speech in Fukienese also has a good correct rate between 3 and 20 trainings. In particular, the 3 states situation is outstanding in regard to the results obtained.

We focus on vocal cords vibration recognition; we collect 10 users vocal cords vibration data and to recognize it, half boys and half girls. The average age of boys is 24 years old, standard deviation is 0.4. The average age of girls is 24, standard deviation is 1.7. The training set is from 1 to 5, state is 2 and 3. We show the correct rate as follow:

Figure 3.25 Recognition correct rates of five boys with 2 states training

Figure 3.26 Recognition correct rates of five girls with 2 states training
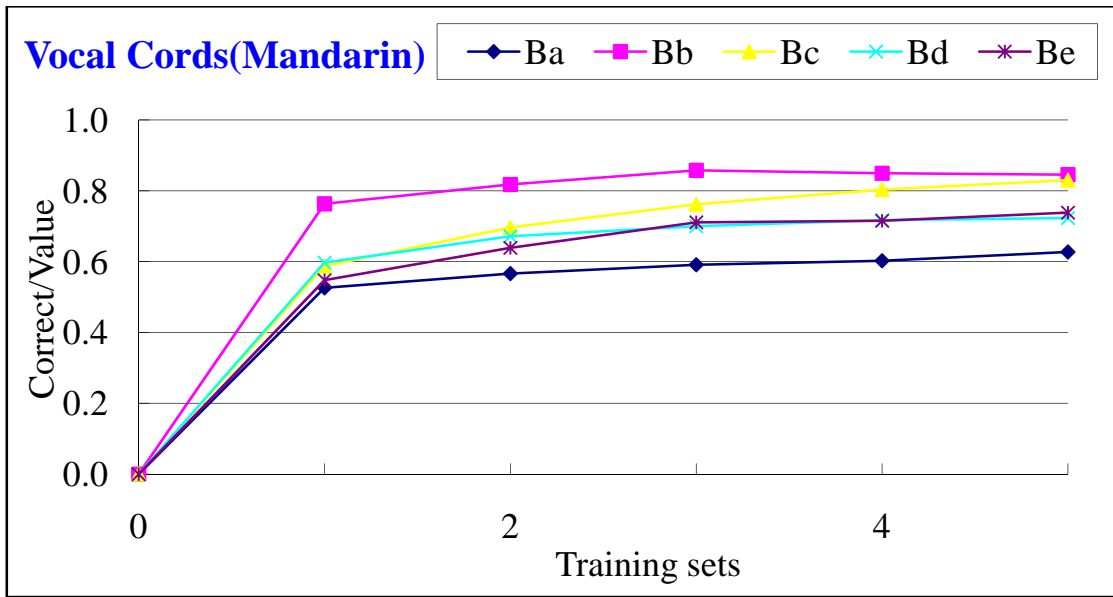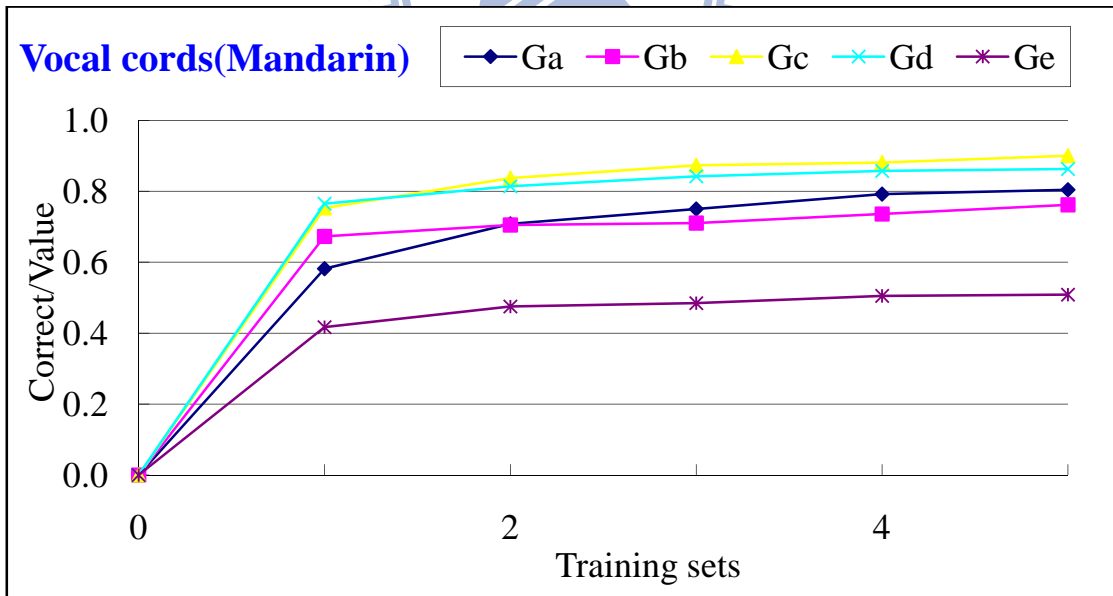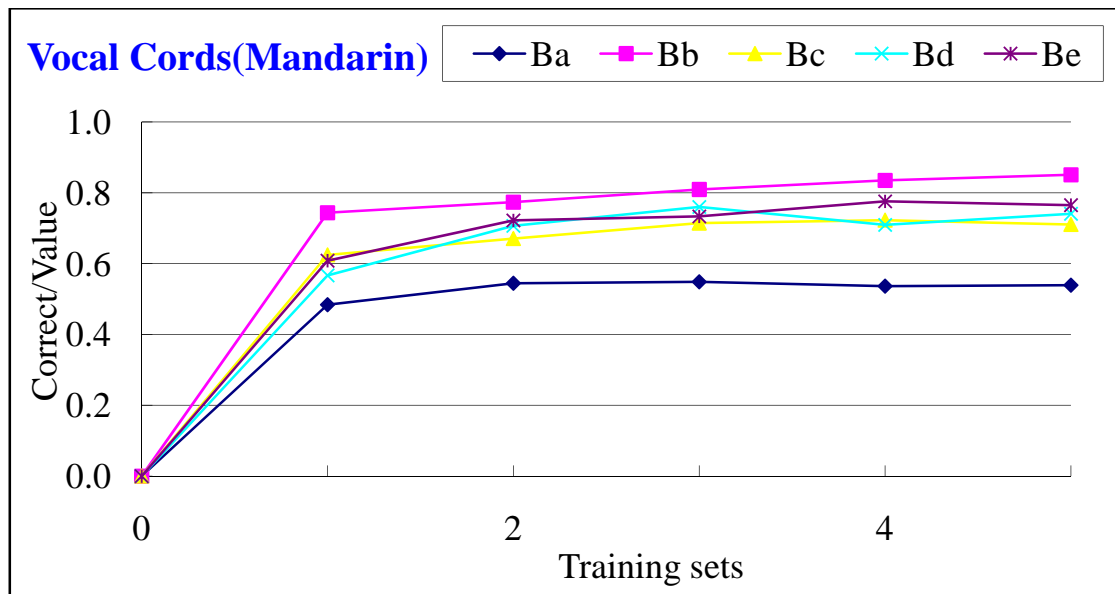
Figure 3.27 Recognition correct rates of five boys with 3 states training
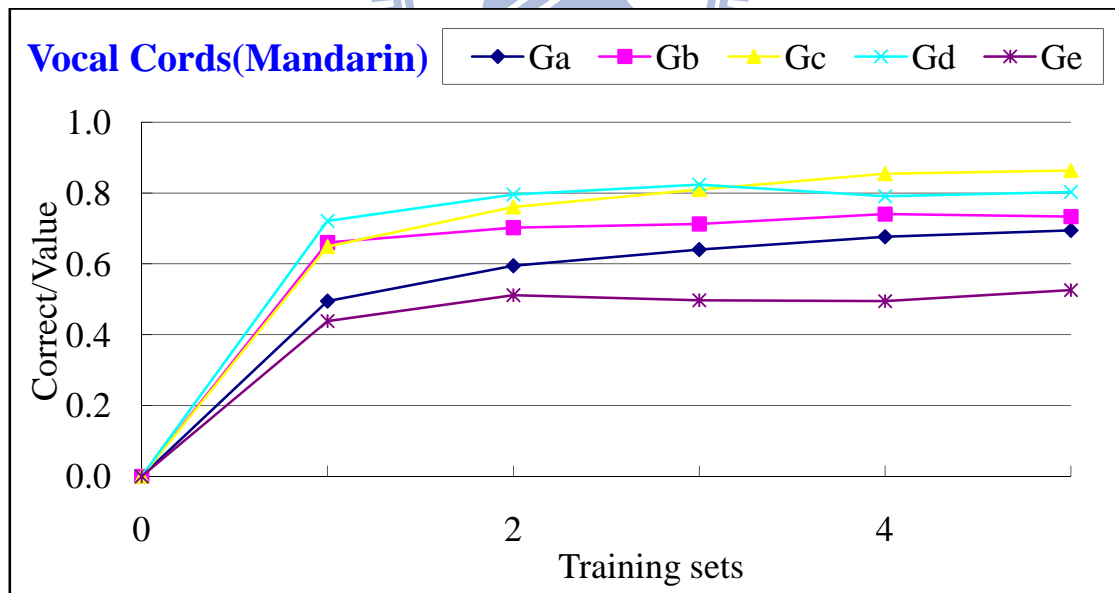

Figure 3.28 Recognition correct rates of five girls with 3 states training

Some users get worse recognition rates in this system. There are many reasons. They first time use this system and don't familiar with system. Algorithm maybe is not perfect enough that vocal cords vibration should be more strongly.

3.3. Mechanical and electrical integration

Attaching the controller to the wheelchair is a fairly arduous task requiring knowledge on how to supply electricity to a motor via signal. Our controller can load at most 1 A current; current flow through the motor is about 2~3 A. At first, we set current flow through the motor and controller; it did not work. The controller cut off in protest. We used relays to solve this problem, four relays to control the motors and four relays to control the brakes. If we want to rotate the motor, we should first power on the brake; then, when the brake is off, the wheel can rotate. Motor control uses four relays: one for right wheel motor rotation, one for right wheel motor reverse rotation, one for left wheel motor rotation and one for left wheel motor reverse rotation. Brake control is the same as motor control. If the controller commands the wheelchair to operate, like *forward*, it should output a signal. We show the relations between instruction and motor/brake control.

Table 3.4 Relations between instruction and motor/brake control

Instruction	Motor and brake control
Forward	Right wheel (brake) and left wheel (brake) rotate (power on)
Backward	Right wheel (brake) and left wheel (brake) reverse rotate (power on)
Left	Right wheel (brake) rotate (power on) and left wheel (brake) reverse rotate (power on)
Right	Right wheel (brake) reverse rotate (power on) and left wheel (brake) rotate (power on)

Figure 3.29 shows the physical, mechanical and electrical integration of the controller and the wheelchair.
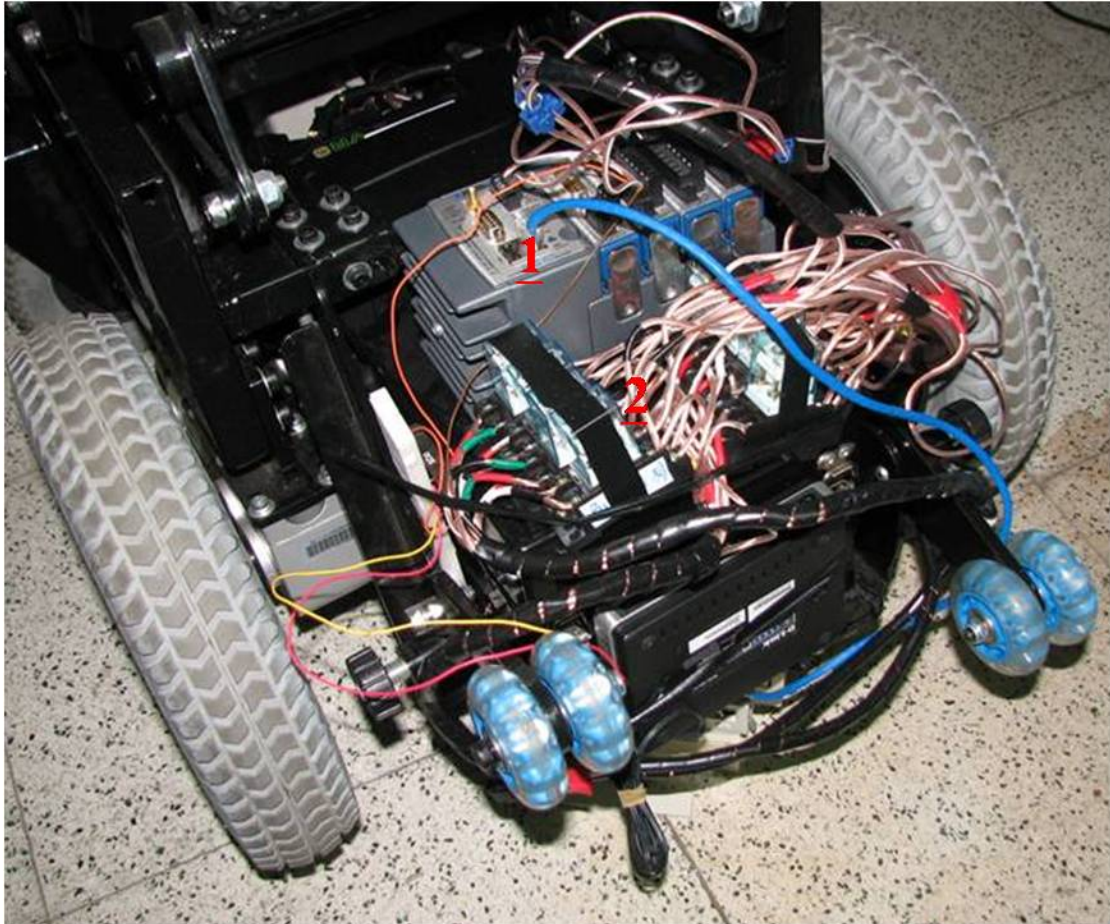
Figure 3.29: Physical, mechanical and electrical integration

In Figure 3.29, the number 1, which we indicate in red color and underline, is controller NI CompactRIO and the number 2, for which we also use red color and underline is relays, like the switch between the controller and motor.

Figure 3.30 shows the overall physical system architecture, including wheelchair, microphone, notebook, wheelchair and controller.

Figure 3.30: Overall physical system architecture

3.4. Control rule

When we command an instruction like *left*, how does the wheelchair run? At what angle does the wheelchair turn or what time does motor run? Anything concerning the security of users could lead to possible serious problems. At first, we did a conservative test as Table 3.5 shows:

Table 3.5 Conservative control rule

Instruction	Motor rotation time
Forward	1 sec
Backward	0.2 sec
Left	0.3 sec
Right	0.3 sec

We ensured that this system offers security, although it was not practical in regard to utility for the user. But for us, it was inspiring to drive a wheelchair using speech. In order to satisfy the real world situation, we designed another method.

Table 3.6 Suitable control rules for present situation

Instruction	Motor rotation time
Forward	Rotate until any signal input
Backward	Rotate until any signal input
Left	0.3 sec
Right	0.3 sec

Many research teams have done in-depth research on control rules, with their goal being satisfaction on the part of users. The latter method is more practical than the former. Thanks to controller FPGA module, data acquisition is rapid; the wheelchair will moves or stop depending on signal input.
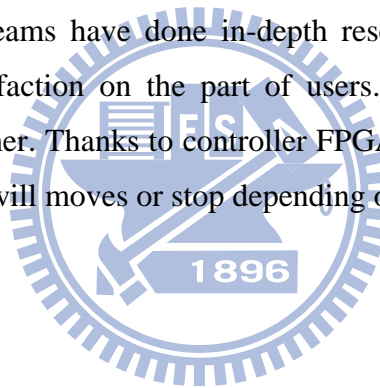
Chapter 4: Discussion and Conclusion

4.1. Discussion

We have employed speech and vocal cord vibration to implement a type of mouth command control wheelchair. Below, we discuss some interesting problems.

4.1.1. Parameter setting in speech recognition

How can we improve the correct rate of speech recognition? A correct rate reaching 95% with 3 training sets is our demand, whatever the speech signals or vocal cord vibration signals. We discuss speech recognition parameters as follows:

1. Frame size:

A frame usually occupies 20~30 ms; we set 240 samples as a frame (sampling rate: 10000 sample/sec), occupying 24 ms; we tried 200 or 300 samples as a frame, and observed the correct rate.

2. Frame overlap:

Usually, a frame overlap occupies 1/2~2/3 frame size; we set the frame overlap at 160 samples (frame size: 240 points), occupying 2/3 frame size. We tried to set 1/2 or less 2/3 frame size as the frame overlap, and observed the correct rate.

4.1.2. Customized bio-signal acquiring module

In our present research, we have developed speech and vocal cords vibration recognition to control a wheelchair, although vocal cord vibration recognition is not as accurate as speech recognition. A bio-signal acquiring module should satisfy user needs and the situations they face. If possible, we will try to develop a new method, breathing vibration as a bio-signal acquiring method. We list bio-signal acquisition methods and tools.

Table 4.1 Bio-signal control interfaces and tools

Bio-signal	Tool
Speech recognition	Dynamic microphone
	Condenser microphone
Vocal cords vibration	Dynamic microphone
	Condenser microphone
	Accelerometer
Breathing vibration	Piezoelectric materials
	Accelerometer

We have developed speech recognition and vocal cord vibration detection and transmission using a dynamic microphone as a possibly suitable tool for each bio-signal method.

4.1.3. Human system

A mature speech recognition system is one wherein the user's input of any sentence could train the model. It is difficult to pronounce some specific speech or sentence for some users. If the system could achieve this standard, it would be helpful to users. If users command an instruction which does not belong to any speech model, the system should be able to simply ignore/reject it. Recently, certain systems have devoted efforts to achieving this objective.

As mentioned in Chapter 3, we employed three languages, separately, to test speech recognition. There were good recognition rates in Mandarin and Fukienese, but that for English was not as good.

4.2. Conclusion

At first, we see some real cases about paraplegia patients and then to have an idea of vocal cords controller. In order to achieve our idea, we meet many problems.

First, we develop a speech recognition system on PC-base, after which we embedded the program into NI cRIO. The controller FPGA module handled user input signal and output signal to wheelchair. The controller real-time processor operated via a speech recognition algorithm. When we accomplished that program embedded into

cRIO, what remained to be done was mechanical and electrical integration.

We integrated the controller into the wheelchair and the battery in the wheelchair supplied power to the controller. The motors responded to the control signal.

With our background in computer science, in order to make the system perfect, we needed to cooperate with electrical integration engineers and physiotherapists who could explain the users' actual needs.

If you want to know more about our system, you can to visit our website. We uploaded a demo video on our website. The hyperlink is as follows:

http:// sites.google.com/site/labview704/project
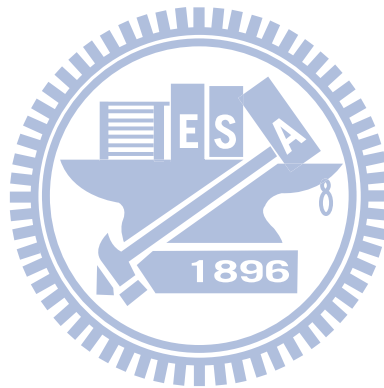
Chapter 5: Future work

In the future, we hope the system which we have presented could be better perfected and more user friendly. More effort will be devoted to the following:

1. Customized control interface to satisfy users:

Wheelchair integration could be friendlier. There is only a single speed to steer wheelchair. We can add a speed control to controller to render it more secure and allowing the user to steer the wheelchair as desired.

We should engage in more discussions with doctors regarding patients' needs, as they possess expert opinion. Our goal is to develop a system that in generally acceptable for hand disabled users, suitable for mass production and at low cost.

2. To develop a controller suitable for MD patients:

MD is described in [13], as an inherited disorder that causes progressive muscle weakness (myopathy) and atrophy (loss of muscle mass) due to defects in one or more genes required for normal muscle function; the primary symptom for most types of MD is muscle weakness; patients even have difficulties in speaking out a word, or are only able to vibrate their abdomen in some serious cases. We tend to specialize our system in order to help them by developing a system which can be expanded in order to permit more features of command, such as abdomen vibration.

References:

[1] The service of the Department of Statistics, Ministry of the Interior, ROC http://www.stat.gov.tw/

[2] Fehr L, Langbein W E, Skaar S B (2000) "Adequacy of power wheelchair control interfaces for persons with severe disabilities: A clinical survey." J REHABIL RES DEV 37:353-360

[3] Brenda, J. M., (1991) "American family physician." Vol.43, no.2, p535-541

[4] Crisman E E, Loomis A, Shaw R, Laszewski Z (1991) "Using the eye wink control interface to control a powered wheelchair." Annual International Conference of the IEEE EMBS, 1991, pp 1821-1822

[5] Chen Y L, Chen S C, Chen W L, Lin J F (2003) "A head orientated wheelchair for people with disabilities." DISABIL REHABIL 25:249-253

[6] Ding D, Cooper R A, Spaeth D (2004) "Optimized joystick controller." 26th Annual International Conference of the IEEE EMBS, San Francisco, US, 2004, pp 4881-4883

[7] Speak Your Mind, neurological signal control, http://www.theaudeo.com/

[8] Yu-Chen. Tai, "Global mobility aids markets research", the Industrial Economics & Knowledge Center of ITRI press, 2004

[9] Chen-Kuang Yang, "VISUAL BASIC 與語音辨識讓電腦聽話", 1st edition, Kings Information Co., Ltd press, 2002

[10] Jyh-Shing Roger Jang, "Audio Signal Processing and Recognition" of online course, http://neural.cs.nthu.edu.tw/jang/books/audioSignalProcessing/

[11] Castellanos G, Delgado E, Daza G, Sanchez LG, Suarez JF (2006) "Feature selection in pathology detection using hybrid multidimensional analysis." 28th Annual International Conference of the IEEE EMBS New York, USA, 2006

[12] Chiou YR (2008) "Study on partial regularized least squares method." A thesis submitted to Institute of Computer Science and Engineering College of Computer Science, National Chiao Tung University, Hsinchu, Taiwan ROC

[13] Darras B T, Shefner J M, Moynihan L K, Dashe J F (2006) "Patient information: Overview of muscular dystrophies" UpToDate at http://www.uptodate.com

[14] Kulkarni R (2006) "FPGAs put PLCs put in motion [motion control]." Machine Design International 78:116-122

[15] CH Hsu, HY Hsu, SY Wang, TC Hsiao(2008), "Wheelchair Direction Control by Acquiring Vocal Cords Vibration with Multivariate Analysis" 4th European Congress For Medical and Biomedical Engineering 2008

[16] Kirby RL., "Wheelchair stability: important, measurable and modifiable. Technology and Disability" 5, 75-80, 1996

[17] R. A. Cooper, "Wheelchair Mobility: Wheelchair and Personal Transportation" IEEE Rehab. Eng., pp. 2071-2085, 1995.