VOICE ACTIVATED SWITCH

NAZARUDDIN BIN MD SHAH

This Report Is Submitted In Partial Fulfillment of Requirements for the Bachelor Degree of Electronic Engineering (Industrial Electronic) with honours.

Fakulti Kejuruteraan Elektronik dan Kejuruteraan Komputer
Universiti Teknikal Malaysia Melaka

April 2007

UNIVERSTI TEKNIKAL MALAYSIA MELAKA
FAKULTI KEJURUTERAAN ELEKTRONIK DAN KEJURUTERAAN KOMPUTER

BORANG PENGESAHAN STATUS LAPORAN
PROJEK SARJANA MUDA II

Tajuk Projek	:	Voice Activated Switch
Sesi Pengajian	:	2003-2007

Saya NAZARUDDIN BIN MD SHAH

mengaku membenarkan Laporan Projek Sarjana Muda ini disimpan di Perpustakaan dengan syarat-syarat kegunaan seperti berikut:

1. Laporan adalah hakmilik Universiti Teknikal Malaysia Melaka.
2. Perpustakaan dibenarkan membuat salinan untuk tujuan pengajian sahaja.
3. Perpustakaan dibenarkan membuat salinan laporan ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. Sila tandakan (√) :

☐	**SULIT***	(Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di dalam AKTA RAHSIA RASMI 1972)
☐	**TERHAD***	(Mengandungi maklumat terhad yang telah ditentukan oleh organisasi/badan di mana penyelidikan dijalankan)
☑	**TIDAK TERHAD**	

Disahkan oleh:

(TANDATANGAN PENULIS)

Alamat Tetap: 24-F,TAMAN PERKOTA,75350 BATU HAMPAR, MELAKA.

Tarikh: 4/5/07

(COP DAN TANDATANGAN PENYELIA)

FAUZIYAH BT SALEHUDDIN
Pensyarah
Fakulti Kej Elektronik dan Kej Komputer (FKEKK),
Universiti Teknikal Malaysia Melaka (UTeM),
Karung Berkunci 1200
Ayer Keroh, 75450 Melaka

Tarikh: 4/5/2007

"Saya akui laporan ini adalah hasil kerja saya sendiri kecuali ringkasan dan petikan yang tiap-tiap satunya telah saya jelaskan sumbernya."

Tandatangan : ....
Nama Penulis : NAZARUDDIN MD SHAH..
Tarikh : 9/5/07..

"Saya/kami akui bahawa saya telah membaca karya ini pada pandangan saya/kami karya ini adalah memadai dari skop dan kualiti untuk tujuan penganugerahan Ijazah Sarjana Muda Kejuruteraan Elektronik (Elektronik Industri) dengan kepujian."

Tandatangan : ....

Nama Penyelia : FAUZIYAH BTE SALEHUDDIN

Tarikh : 4 HB MEI 2007

I dedicated this to both of my parents, my family, friends and electronic engineering education. Thanks for everything. Your deed will be remembered till my last breath.

PENGHARGAAN

Syukur ke hadrat Allah, kerana dengan limpah kurnia-Nya, saya telah dapat melaksanakan Projek Sarjana Muda ini dengan dengan baik. Saya ingin mengambil kesempatan ini untuk mengucapkan penghargaan kepada semua pihak yang telah banyak membantu saya sepanjang saya melaksanakan PSM ini terutamanya penyelia saya, Pn. Fauziyah serta rakan-rakan lain yang banyak membantu. Jutaan terima kasih saya ucapkan dan moga Allah akan membalas jasa baik kalian.

ABSTRACT

Nowadays, speech recognition is becoming more and more popular technology in the society. Voice activated switch is the process of automatically switching system which depends on a specific word which will trigger the switch. This project is to design voice activated switch, which have relevance to home switching system. The system allows user to access the operation of switching 'on' or 'off' their electrical device such as lamp or fan by using their voice command. The speech processing toolbox located in Matlab will be used to perform standard speech signal processing on the segmented utterances.

ABSTRAK

Pada zaman serba moden ini, teknologi pengenalan lisan telah menjadi semakin diminati dalam kegunaan harian. Suis pengaktif suara ialah suatu proses operasi pengaktifan suis dengan menggunakan perkataan yang ditetapkan untuk mengaktifkan suis. Projek ini direka untuk suis pengaktif suara dimana ia berkesesuaian dalam sistem suis di rumah. Sistem ini membenarkan pengguna menghidup atau memadamkan peralatan elektrik di rumah seperti lampu atau kipas. Penganalisa suara yang tedapat di dalam perisian matlab digunakan dalam menghasilkan analisa berkesesuaian dengan pemprosesan suara berdasarkan pengtafsiran pecahan-pecahan yang terdapat di dalam suara.

TABLE OF CONTENTS

TABLE LIST

FIGURE LIST

APPENDICES LIST

CHAPTER I

INTRODUCTION

1.1 PROJECT SYNOPSIS

The final year students of KUTKM need to produce a project as one of their goal before graduating. This project will show the talent and skills of the student which can be made as a serious strength during their future as an engineer.

The Voice Activated Switch System runs on MATLAB platform which uses voice to activate switch. By other mean, this system allows user to turn 'on' or 'off' switch by using their voice. Voice will be the input into the system and system will do the recognition and create an output based on the data stored. Generally the system will analyze the cepstral analysis of the voice and compare it to the data stored in the MATLAB folder.

1.2 PROBLEM STATEMENTS

The main statement is that people who are old or with the disadvantages to walk and stand, will be having difficulty to switch 'on' or 'off' the fan. There are also some of us who are just wish that they have the abilities to control their electrical devices by their voice command.

Second statement goes to the student themselves regarding the lack of knowledge on one of the most fines software that indicates the solving of engineering problems. System will act as human ears which can recognize each command given by the user and act as to be less susceptible to mention variation of the voice commands. Not all students who **learn MATLAB** knew about this. The developing of this system can solve and overcome those problems.

1.3 OBJECTIVES

The primary purpose of this project is to create a voice recognition controls system that reliably and consistently recognizes several different commands and executes a multitude of different functions based on varied command sequences. The original concept intends for a person to use this product at a home computer.

The objectives of this project are:

i. To develop software for voice activated switch system by using MATLAB.
ii. To gain knowledge of the student about MATLAB software and its abilities.
iii. To perform standard speech signal analysis by using the signal processing toolbox in MATLAB.

1.4 SCOPES OF PROJECT

The scope is to develop a software codec' by using MATLAB, and analyzing voice signals and to turn out with and output to turn 'on' or 'off' switch. A microphone will be used to send voice signals to the system and then it will be saved to a specific directory file in a computer to allow the system to analyze and come out with an output which will 'on' or 'off' switch.

1.5 CHAPTER SUMMARY

CHAPTER I will describe the definition of this project will be explained in this chapter. Also in this chapter there will be summary the project progress.

CHAPTER II will discuss about research and information which are related to this project. Every fact and information are gained from different references will be discussed so that the best technique and method can be implemented on this project.

CHAPTER III will be describing how this project is separated to small partition. The elements which are used to build each circuit are described by concept and theory. Plus, figures are provided to ensure the understanding.

CHAPTER IV is describing about the project result and outcome discovery. The project outcome discovery will be presented from the many data analysis results.

The final chapter, CHAPTER V will be explaining about the conclusion of the whole project which includes project finding, achievement analysis and conclusion about the research implementation which have been used. The project suggestion for enhancement also discussed.

CHAPTER II

LITERATURE REVIEW

2.1 INTRODUCTION

This chapter is all about discussing the theory and concept from the past projects. The objective is to explain the perspective and method which has been used in the past projects and to observe how this project can be related with existing research and theory. This shows how the theory and concept have been implemented in order to solve project problem. The theory understanding is crucial as a guidance to start any project. The result of a project cannot be assessed if it's not compared to the theory.

2.2 FUNDAMENTAL OF THE MEL FREQUENCY CEPSTRAL COEFFICIENTS ANALYSIS

Before going any further on MATLAB's cepstral analysis, the Mel Frequency Cepstral Coefficients (MFCCs) concept is studied. The independent variable of a cepstral graph is called the quefrency. The quefrency is a measure of time, though not in the sense of a signal in the time domain. For Command , if the sampling rate of an audio signal is 44100 Hz and there is a large peak in the cepstrum whose quefrency is 100 samples, the peak indicates the presence of a pitch that is 44100/100 = 441 Hz. This peak occurs in the cepstrum because the harmonics in the spectrum are periodic, and the period corresponds to the pitch.

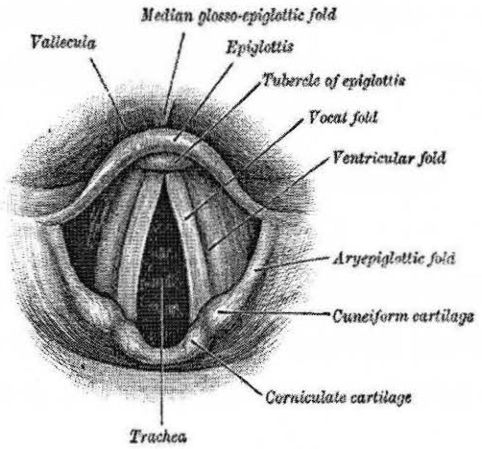

Figure 2.1: Laryngoscopic

The image above shows the Laryngoscopic view of the vocal folds. The space between the vocal cords is called the glottis. As the vocal cords vibrate, the resulting vibration produces a "buzzing" quality to the speech, called voice or voicing.

Sounds production involving glottis is called *glottal*. English has a voiceless glottal fricative spelled "h". In many accents of English the glottal stop (made by pressing the folds together) is used as a variant allophone of the phoneme /t/ (and in some dialects, occasionally of /k/ and /p/); in some languages, this sound is a phoneme of its own.

Human speech is created through two main components. Air is built up below the vocal bands, causing the vocal bands to be pushed apart. The glottis, the opening where the vocal bands meet, is then forced closed due to the changing of pressure, which results in the building up of air pressure again, and subsequently another pulse of air is released across the vocal bands. This cycle creates voicing. This acoustic signal is then shaped by the different parts of the mouth to create the vocal sound known as speech. These oral components act as a filter to vary the emitted signal. Cepstral analysis is able to pick up vowel sounds because the glottis has only minimal constriction for these sounds, unlike consonant sounds that generally are produced with a highly constricted glottis and cannot sustain voicing as well. A speech signal can be represented by functions of these two signal components, where $S(\omega)$ represents the overall speech signal, $H(w)$ represents the filtering vocal component, and $E(w)$ represents the glottal excitation.

$$|S(\omega)| = |H(w)||E(w)| \qquad (1.1)$$

The vocal shaping impressed upon the glottal signal varies much less rapidly than the excitation signal being pushed through the glottis. By taking the log of the signal, it is possible to separate it into the vocal shaping component and the glottal component.

$$\log(|S(\omega)|) = \log(|H(\omega)|) + \log(|E(\omega)|) \qquad (1.2)$$

Since the glottal component varies much faster than the voicing component, the signal can be further manipulated to extract the values representing the vocal shaping. By performing another Fourier Transform on the log of the spectrum, it is possible to divide the two components. Since a Fast Fourier Transform (FFT) has already been performed to get the signal into frequency domain, an Inverse Fast Fourier Transform (IFFT) must be used to get the "cepstrum", or the spectrum of the frequency domain signal.

$$c(n) = (1/2\pi)\int_{-\pi}^{\pi} \log(S(\omega)|)|e^{j\omega n} d\omega \quad (1.3)$$

The cepstrum of the voice signal consists of the slower varying vocal signal at the beginning followed by glottal pulses at higher frequencies. The cepstral coefficients defining the vocal shaping can be extracted by taking the beginning discrete values of the cepstrum.

The coefficients taken from the cepstrum are the foundation of detecting desired words. By comparing the cepstral coefficients of a test sample to the cepstral coefficients of a template of the characteristic coefficients of the desired word, it is possible to determine how closely the vocal shaping of the sample is to that of the test. By taking the sum of the differences between the template coefficient and the corresponding sample coefficient, it can be determine quantitatively how closely the words resemble each other. The Euclidean Distance (ED) is the formula used to calculate this value for n different coefficient values, where c_i is the i-th sample cepstral coefficient and t_i is the i-th template coefficient.

$$ED = \sqrt{\sum_{i=1}^{n} (c_i - t_i)^2} \quad (1.4)$$

Euclidean Distance Equation

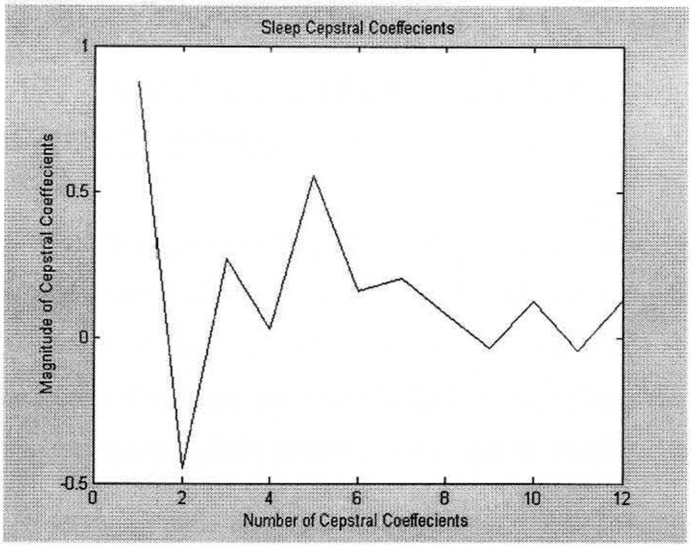

Figure 2.2: Graph of Magnitude vs Number of Cepstral Coefficients.

During the analysis, the routine melcepst.m implements a mel-cepstrum front end for a recognizer. Mel Frequency Cepstral Coefficients (MFCCs) are coefficients that represent audio which derived from a type of cepstral representation of the audio clip which can be simplified as spectrum-of-a-spectrum. The difference between the cepstrum and the Mel-frequency cepstrum is that in the MFC, the frequency bands are positioned logarithmically as in mel scale which approximates the human auditory system's response more closely than the linearly-spaced frequency bands obtained directly from the Discrete Fourier transform (FFT). This can allow for better processing of data, for example, in audio compression such as speech recognition. Although the MFCCs lack an outer ear model and, hence, cannot represent perceived loudness accurately, it still suit perfectly to the voice activated switch system regarding the normal usage of speech recognition.

MFCCs are commonly derived as follows:

- Take the Fourier transform of a signal
- Map the log amplitudes of the spectrum obtained above onto the Mel scale, using hamming window in time domain
- Take the Discrete Transform of the list of Mel log-amplitudes, as if it were a signal
- The MFCCs are the amplitudes of the resulting spectrum.

2.2.1 COMPARISON BETWEEN CEPTRAL ANALYSIS AND LINEAR PREDICTION ANALYSIS.

Athough that there are ranging kinds of analysis for speech recognition system, any analysis consist of cepstral analysis such MFCCs is the best method for the recognition system. A demonstration that cepstral analysis overrules another method which is Linear Prediction analysis (LPA) is done, reveals that in the voiced section of the spoken word, a more periodic waveform is produced with the use of cepstral analysis.

2.2.1.1 CEPSTRAL ANALYSIS

The words "lampu on" is spoken and recorded then digitized its signal using a sampling frequency of 8 kHz and 16-bit linear encoding, to ensure that MATLAB can read and process it. Time waveform and the wideband and narrowband spectrogram of the utterance are plotted.

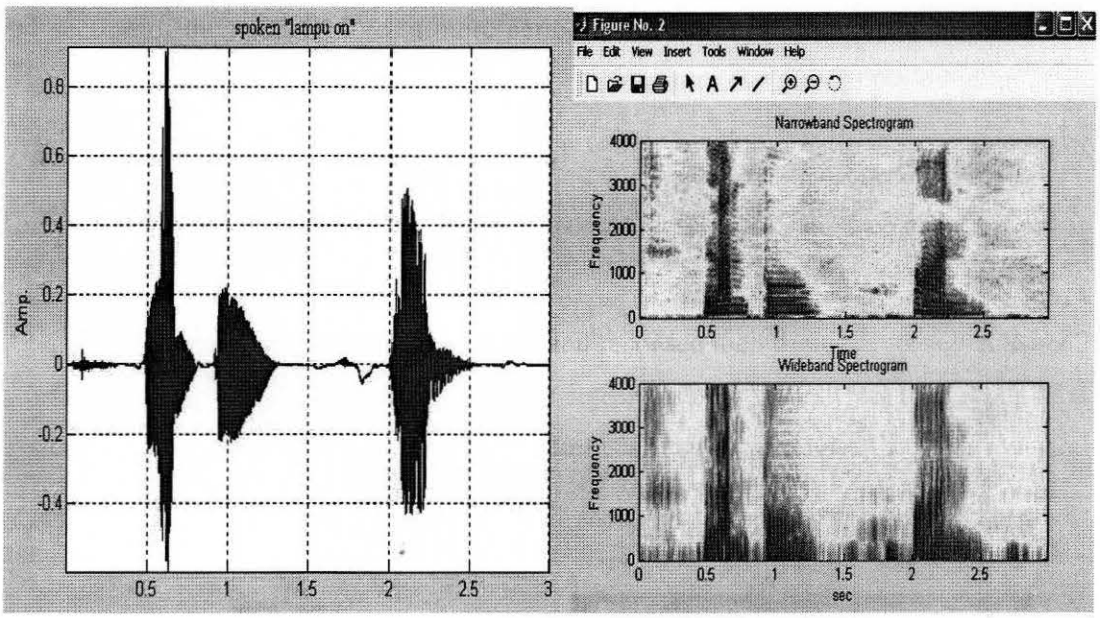

Figure 2.3: Time Waveform And Spectrograms.