

**SUBTITLING ENGLISH SPEECH**

**KUTIBA NANAA**

**UNIVERSITY UTARA MALAYSIA**

**2009**

# **SUBTITLING ENGLISH SPEECH**

**A thesis submitted to college Arts & Sciences  
in partial fulfillment of the requirement for the degree  
Master of Science (Intelligent System)  
University of Utara Malaysia**

**By**

**Kutiba Nanaa**

**© Kutiba Nanaa, November, 2009. All rights reserved**



**KOLEJ SASTERA DAN SAINS**  
**(College of Arts and Sciences)**  
**Universiti Utara Malaysia**

**PERAKUAN KERJA KERTAS PROJEK**  
**(Certificate of Project Paper)**

Saya, yang bertandatangan, memperakukan bahawa  
*(I, the undersigned, certify that)*

**KUTIBA NANAA**  
**(801585)**

calon untuk Ijazah  
*(candidate for the degree of)* **MSc. (Intelligent System)**

telah mengemukakan kertas projek yang bertajuk  
*(has presented his/her project paper of the following title)*

**SUBTITLING ENGLISH SPEECH**

seperti yang tercatat di muka surat tajuk dan kulit kertas projek  
*(as it appears on the title page and front cover of project paper)*

bahawa kertas projek tersebut boleh diterima dari segi bentuk serta kandungan  
dan meliputi bidang ilmu dengan memuaskan.

*(that the project paper acceptable in form and content, and that a satisfactory  
knowledge of the field is covered by the project paper).*

Nama Penyelia Utama  
*(Name of Main Supervisor):* **DR. AZMAN BIN YASIN**

Tandatangan  
*(Signature)*

:

Tarikh  
*(Date)*

:

17/11/09

## **PERMISSION TO USE**

In presenting this thesis in partial fulfillment of the requirements for a Master of Science in Intelligent System degree from University Utara Malaysia, I agree that the University Library may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purpose may be granted by my supervisor or, in their absence by the Academic Dean College of Arts and Sciences. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to University Utara Malaysia for any scholarly use which may be made of any material from my thesis.

Requests for permission to copy or to make other use of materials in this thesis, in whole or in part, should be addressed to

**Dean (Academic) College of Art and Sciences**

**University Utara Malaysia**

**06010 UUM Sintok**

**Kedah Darul Aman.**

## ABSTRACT

Human-being learn their mother tongue instinctively from the childhood, but when they grow up, many challenges will appear when they try to learn other languages. The building of vocabulary bank and understanding of grammar is not an easy job, so some people use computer to help to translate other language to their mother tongue, but this mean also got its limitation and disadvantages such as the misunderstanding of the words. In this study, the researcher will use artificial intelligent technologies to solve these problems. First, speech will be received as sound waves, then it will be segment into letters and recognized by the usage of Neural Network.

## ACKNOWLEDGEMENT

*By the Name of Allah, the Most Gracious and the Most Merciful*

First, I would like to express my appreciation to Allah, the Most Merciful and, the Most Compassionate who has granted me the ability and willing to start and complete this study. I do pray to His Greatness to inspire and enable me to continue the work for the benefits of humanity.

My most profound thankfulness goes to my supervisor Dr. Azman Yasin for his scientifically proven and creativity encouraging guidance.

Last but not least, I wish to thank all my dearest family members, especially Dad, Mum, and my great brothers and sisters for being by my side since I left home. Also thank you to my lecturers and friends who have given me emotional support during my study.

*Thank you UUM.*

## **TABLE OF CONTENT**

PERMISSION TO USE	i
ABSTRACT	ii
ACKNOWLEDGEMENT	iii
TABLE OF CONTENT	iv
LIST OF TABLES	vii
LIST OF FIGURES	viii

### **CHAPTER ONE**

#### **INTRODUCTION**

1.1 Introduction	1
1.2 Problem statement	5
1.3 Research Questions	6
1.4 Research objectives	6
1.5 Scope	6
1.6 Research Significance	7

### **CHAPTER TWO**

#### **LITERATURE REVIEW**

2.1 Voice Recognition	8
2.2 Natural Languages	10
2.4 Segmentation	12

## **CHAPTER THREE**

### **METHODOLOGY**

3.1 Introduction	15
3.2 Spiral Development Model	15
3.2.1 Understanding the Requirements	16
3.2.2 Design the System	17
3.1.2.1 Processing the Noise	20
3.1.2.2 Extracting the Signal	24
3.1.2.3 Segmentation	24
3.1.2.4 Neural Network	28
3.2.3 Build in Stage	29
3.2.4 Test and Evaluation	30
3.2.5 Documentation	30
3.3 Summary	

## **CHAPTER FOUR**

### **SYSTEM ANALYSIS AND DESIGN**

4.1 Introduction	31
4.2 List of Requirement	31
4.1.1 Functional Requirement	32
4.1.2 Non Functional Requirement	32
4.3 Use Case Diagram	34
4.4 Use case specification	37



4.5	Sequence Diagram	42
	• Add Word	43
	• Record Voice	43
	• Add new pattern	45
	• Train weights	45
4.6	Collaboration Diagram	46

## **CHAPTER FIVE**

### **FINDING**

5.1	Explanations of screenshots of the functionalities in this prototype	50
	5.1.1 Manage of Lexical Unit	50
	5.1.2 Manage of Neural Network	56
	5.1.3 Test the Result	59
5.2	Usability Testing	60
5.3	Summary	62

## **CHAPTER SIX**

### **CONCLUSIONS**

6.1	Findings	63
6.2	Research Limitation	64
6.3	Recommended Works	64
6.4	Summary	65
	REFERENCES	66

## LIST OF TABLES

Table 3.1: Attributes of the Neural Network	26
Table 4.1: Functional requirement	32
Table 4.2: Non-functional requirement	34
Table 4.3: Use Insert Vocabulary	37
Table 4.4: Use Case Record Word	38
Table 4.5: Use Case add signal as pattern	39
Table 4.6: Use Case train weights	41
Table 5.1: The result of usability evaluation questionnaire	61

## LIST OF FIGURES

Figure 3.1: Spiral development model	16
Figure 3.2: Subtitling English Speech architecture	17
Figure 3.3: Processing the noise	19
Figure 3.4: Result of processing the noise	19
Figure 3.5: Result of processing the noise	20
Figure 3.6: Result of removing the frames which have a low energy	21
Figure 3.7: Result of scaling	22
Figure 3.8: Result of compressing	23
Figure 3.9: Extracting the signal	24
Figure 3.10: Result of Segmentation	25
Figure 3.11: Segmentation	25
Figure 3.12: Sounds of English letters	27
Figure 3.13: Neural Network	28
Figure 3.14: Usability components in questionnaire	29
Figure 4.1: Use Case Diagram	36
Figure 4.2: Sequence Diagram for add word	43
Figure 4.3: Sequence Diagram record voice	43
Figure 4.4: Sequence Diagram Add new pattern	44
Figure 4.5: Sequence Diagram train Weights	45
Figure 4.6: Collaboration diagram Insert Word	46
Figure 4.7: Collaboration diagram for Record Voice	47
Figure 4.8: Collaboration diagram for Add pattern	48
Figure 4.9: Collaboration diagram for Train Weights	49

Figure 5.1: Home page of lexical management	50
Figure 5.2: Record a word signal	51
Figure 5.3: Add a new word	51
Figure 5.4: Show inserted words	52
Figure 5.5: Add a pattern to certain word	53
Figure 5.6: Show patterns of certain word	53
Figure 5.7: Remove a pattern	54
Figure 5.8: Remove a word	54
Figure 5.9: Save all change data	55
Figure 5.10: Manage neural network	55
Figure 5.11: Home page of neural network management	56
Figure 5.12: Change information	56
Figure 5.13: Show data of training pairs	57
Figure 5.14: Show the weights	57
Figure 5.15: Initialize the weights	58
Figure 5.16: Train the neural network	58
Figure 5.17: Save neural network	59
Figure 5.18: Test the word	59
Figure 5.19: Test Speech	60
Figure 5.20: Graph of the Usability Testing	62

# CHAPTER ONE

## INTRODUCTION

### 1.1 Introduction

In computer technology, speech, recognition continues to develop for a wide variety of applications from hands-off control of industrial equipment and of computer systems in low light, to input by persons unable to utilize a keyboard or video display terminal (Klatt, 1991). The voice input and recognition has been used in the business, but it is not so widely popularized to the public (Green, 1984).

People are using multiple sources of information, e.g. lip-reading, facial expression, hand gestures and body language, to recognize speech with high accuracy (Wu, 1991). Acoustic speech signals include most of the information needed to understand a communication. Recognition of the speech signal will be inaccurate when noise in the environment is very strong.

A speech recognition system based on synthetic generation of reference prototypes is described. The vocabulary and grammar are described in a finite-state phoneme network. In the transformation from symbolic to spectral representation, reduction rules modify the

The contents of  
the thesis is for  
internal user  
only

## References

- Akdemir, E. & T. Ciloglu (2008). "The use of articulator motion information in automatic speech segmentation." 594–604.
- BBC. (2009). Retrieved October 26, 2009, from <http://www.bbc.co.uk/>
- Blomberg, M. (1991). "Adaptation to a speaker's voice in a speech recognition system based on synthetic phoneme references." 10: 453-461.
- Barry, W. 1988) "A spiral model of software development and enhancement".
- Brugnara, F., D. Falavign, et al. (1993). "Automatic segmentation and labeling of speech based on Hidden Markov Models." 12: 357-370.
- Cimiano, P., P. Haase, et al. (2008). "Towards portable natural language interfaces to knowledge bases – The case of the ORAKEL system." 65: 325–354.
- Conlon, S. J., J. R. Conlon, et al. (2004). "The economics of natural language interfaces: natural language processing technology as a scarce resource." 38: 141– 159.

Dalea, R., J. Oberlander, et al. (1998). "Integrating natural language generation and hypertext to produce dynamic documents." 11: 109–135.

Desblache, L. (2006). "Technical Translation."

Didiot, I. Illina, et al. (2008). "A wavelet-based parameterization for speech/music discrimination." 24: 341–357.

Foster, P. I. a. G. (2006). "Machine Translation: Overview."

Gagnoulet, C., D. Jouviet, et al. (1991). "Mairievox: A voice-activated information system." 10: 23-31.

Green, D. J. (1984). "Voicer ecognition- still a long way to go." 26.

Gutt, E.& H. Wycombe (2006). "Approaches to Translation: Relevance Theory."

Guz, U., G. Tur, et al. (2010). "Cascaded model adaptation for dialog act segmentation and tagging." 289–306.

Hutchins, J. (2003). "Machine Translation: History."

Jackson, P. & F. Schilder (2008). "Natural Language Processing: Overview."



- Jarifi, D. Pastor, et al. (2008). "A fusion approach for automatic speech segmentation of large corpora with application to speech synthesis." 50: 67–80.
- Klatt, E. C. (1991). "Voice-activated diction for autopsy pathology." 21: 429-433.
- Macklovitch, E. (2006). "Machine-Aided Translation: Methods."
- Malmkjær, K. (2006). "Approaches to Translation, Linguistic."
- Mauranen, A. (2006). "Translation Universals."
- Mellish, C. & J. Z. Pan (2008). "Natural language directed inference from ontologies." 172: 1285–1315.
- Montgomery, L. (2006). "Translation of Scientific and Medical Texts."
- Mporas, I., T. Ganchev, et al. (2010). "Speech segmentation using regression fusion of boundary predictions." 24: 273–288.
- Rabiner, L. R. (1995). "The impact of voice processing on modern telecommunications." 217-226.

Shriberg, E., A. Stolcke, et al. (2000). "Prosody-based automatic segmentation of speech into sentences and topics." 32: 127-154.

Smith, R. W. (2006). "Natural Language Interfaces."

Takahashi, J.-i., N. Sugamura, et al. (1995). "Interactive voice technology development for telecommunications applications." 17: 287-301.

Wu, J.-T., S. Tamura, et al. (1991). "Neural network vowel-recognition jointly using voice features and mouth shape image." 24: 921-927.