สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

# เกมส์ฝึกออกเสียงพูดตัวเลขภาษาไทย 0-9

# Thai numberVoice Edutainment

T061362

โดย

นายวิรัตน์ รัตนพิทักษ์

นายอัมพัน ทะยีอะวัง

เลขหมู่.....
เลขทะเบียน 61362
วัน,เดือน,ปี 17 ก.ค. 2549

b.11542450
i.....

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตร์บัณฑิต

ภาควิชาวิศวกรรมสารสนเทศ

คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ปีการศึกษา 2547

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# THAI NUMBER VOICE EDUTAINMENT

BY

MR. WIRAT RATTANAPITAK

*MR. HAMDAN HAGEEHAWANG*

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

*THE REQUIREMENT FOR THE DEGREE OF*

BACHELOR IN DEPARTMENT OF INFORMATION ENGINEERING

*FACULTY OF ENGINEERING*

KINGMONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

2004

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ปริญญานิพนธ์ปีการศึกษา 2547

ภาควิชา วิศวกรรมสารสนเทศ

คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เรื่อง เกมส์ฝึกออกเสียงพูดตัวเลขภาษาไทย 0-9 โดย Matlab

Voice Edutainment with Matlab

ผู้จัดทำ

1. นายวิรัตน์ รัตนพิทักษ์ รหัสประจำตัว 45015863

2. นายฮัมดัน หะยีฮะวัง รหัสประจำตัว 45015883

_____ อาจารย์ที่ปรึกษา

(ดร. สมเกียรติ อุดมหรรษากุล)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

| | |
|---|---|
| **หัวข้อปริญญานิพนธ์** | เกมส์ฝึกออกเสียงพูดตัวเลขภาษาไทย |
| **ชื่อนักศึกษา** | นายวิรัตน์ รัตนพิทักษ์ รหัสประจำตัว 45015863 |
| | นายฮัมดัน หะยีอะวัง รหัสประจำตัว 45015883 |
| **อาจารย์ที่ปรึกษา** | ดร.สมเกียรติ อุดมหรรษากุล |
| **ระดับการศึกษา** | ปริญญาตรี วิศวกรรมศาสตร์บัณฑิต |
| | สาขาวิศวกรรมสารสนเทศ |
| **ภาควิชา** | วิศวกรรมสารสนเทศ |
| **ปีการศึกษา** | 2547 |


## บทคัดย่อ

ปริญญานิพนธ์นี้กล่าวถึงการศึกษาความเป็นไปได้ในการรู้จำเสียงพูดของคอมพิวเตอร์โดยนำมาประยุกต์ใช้เป็นเกมส์ฝึกออกเสียงพูดตัวเลขภาษาไทยตั้งแต่เลข 0-9 ในการศึกษานี้จะใช้โปรแกรม Matlabในทดลอง โดยใช้การหาค่าสหสัมพันธ์(Correlation) เป็นตัววัดความเหมือนระหว่างเสียงทดสอบกับเสียงอ้างอิง



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

| | |
|---|---|
| **Thesis Title** | Thai number Voice Edutainment |
| **Student** | Mr. Wirat Rattanapitak ID. 45015863 |
| | Mr. Hamdan Hayeehawang ID. 45015883 |
| **Advisor** | Dr. Somkiat Udomhansakul |
| **Graduate Level** | Bachelor Degree of Information Engineering |
| **Department** | Information Engineering |
| **Academic Year** | 2004 |


## ABSTRACT

In this projects , we study a method using correlation for Thai number voice recognition. The maximum value of correlation between testing voice and reference voice is nee define the result .In our experiment , we use twenty set of testing voice and five set of reference voice



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# กิตติกรรมประกาศ

ปริญญานิพนธ์ฉบับนี้สำเร็จได้ด้วยดี เนื่องได้รับการแนะนำ สนับสนุน และให้คำปรึกษา จาก อาจารย์สมเกียรติ อุดมหรรษากุล อาจารย์ที่ปรึกษาปริญญานิพนธ์ รวมทั้งอาจารย์ประจำภาควิชาวิศวกรรมสารสนเทศ และอาจารย์ใน สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบังทุกท่าน ที่ให้การอบรมสั่งสอนวิชาความรู้และแง่คิดในเรื่องต่างๆแก่คณะผู้จัดทำมาตลอดระยะเวลาทั้ง 3 ปีที่คณะผู้จัดทำได้ศึกษาอยู่

ขอกราบขอบพระคุณเป็นอย่างสูงสำหรับบุคคลที่สำคัญที่สุดที่ทำให้ผู้จัดทำมีวันนี้ คือบิดามารดา ผู้เป็นที่เคารพรักยิ่งของผู้จัดทำ จึงกราบขอบพระคุณมา ณ ที่นี้

สุดท้ายนี้ขอขอบคุณธีชัช ฉัตรวิจิตร ที่เสียสละเวลามาช่วยทำโปรเจค

วิรัตน์ รัตนพิทักษ์

ฮัมดัน หะยีฮะวัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# สารบัญ


| | หน้า |
|---|---|
| **บทคัดย่อภาษาไทย** | ก |
| **บทคัดย่อภาษาอังกฤษ** | ข |
| **กิตติกรรมประกาศ** | ค |
| **สารบัญ** | ง |
| **สารบัญภาพประกอบ** | จ |
| **สารบัญตาราง** | ซ |
| **บทที่ 1 บทนำ** | 1 |
| **บทที่ 2 ทฤษฎีที่เกี่ยวข้อง** | 2 |
| 2.1 ธรรมชาติการรู้จำ | 2 |
| 2.2 Fourier Transform | 2 |
| 2.3 คอร์รีเลชั่น ( Corrletion ) | 9 |
| **บทที่ 3 การออกแบบและการสร้าง** | 11 |
| 3.1 การเตรียมสัญญาณ | 11 |
| 3.2 เปรียบเทียบความเหมือน | 15 |
| 3.3 การตัดสินใจ | 17 |
| **บทที่ 4 การทดลองและผลการทดลอง** | 18 |
| 4.1 การทดลองการหาค่าสหสัมพันธ์ของเสียงพูดหนึ่งถึงเก้ากับเสียงอ้างอิง | 18 |
| **บทที่ 5 ข้อสรุปและข้อเสนอแนะ** | 29 |
| **บรรณานุกรม** | 33 |
| **ภาคผนวก** | 35 |


เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# สารบัญภาพประกอบ

หน้า


รูปที่ 2.1 แสดงฟังก์ชั่นและการแตกองค์ประกอบตรีโกณมิติ 3

รูปที่ 2.2 แสดงสัญญาณสี่เหลี่ยมและการประมาณตรีโกณมิติ 4

รูปที่ 2.3 แสดงผลรวมฟังก์ชั่นไม่ต่อเนื่องของไซน์ 6

รูปที่ 2.4 แสดงการคอร์รีเลชั่นของสองฟังก์ชั่น 9

รูปที่ 2.5 แสดงแผนผังการหาค่าคอร์รีเลชั่นในโดเมนความถี่ 10

รูปที่ 3.1 แสดงแผนผังการทำงานของระบบรู้จำเสียงพูด 11

รูปที่ 3.2 แสดงการหาจุดเริ่มและจุดท้ายของเสียง 12

รูปที่ 3.3 รูปแสดงการหาจุดเริ่มต้นและจุดท้ายเสียง 13

รูปที่ 3.4 รูปแสดงการปรับขนาดความยาวข้อมูล 14

รูปที่ 3.5 รูปแสดงการปรับขนาดแอมพิจูดข้อมูล 15

รูปที่ 3.6 รูปแสดงการทำคอร์รีเลชั่นระหว่างสัญญาณ f1(t) กับ f2(t) 16

รูปที่ 3.7 แสดงแผนผังการหาค่าคอร์รีเลชั่นในโดเมนความถี่ 17

รูปที่ 4.1 กราฟแสดงการหาสหสัมพันธ์ ระหว่างสัญญาณเสียงทดสอบ “หนึ่ง” กับสัญญาณเสียงอ้างอิง 18

รูปที่ 4.2 กราฟแสดงการหาสหสัมพันธ์ ระหว่างสัญญาณเสียงทดสอบ “สอง” กับสัญญาณเสียงอ้างอิง 19

รูปที่ 4.3 กราฟแสดงการหาสหสัมพันธ์ ระหว่างสัญญาณเสียงทดสอบ “สาม” กับสัญญาณเสียงอ้างอิง 20

รูปที่ 4.4 กราฟแสดงการหาสหสัมพันธ์ ระหว่างสัญญาณเสียงทดสอบ “สี่” กับสัญญาณเสียงอ้างอิง 21

รูปที่ 4.5 กราฟแสดงการหาสหสัมพันธ์ ระหว่างสัญญาณเสียงทดสอบ “ห้า” กับสัญญาณเสียงอ้างอิง 22

รูปที่ 4.6 กราฟแสดงการหาสหสัมพันธ์ ระหว่างสัญญาณเสียงทดสอบ “หก” กับสัญญาณเสียงอ้างอิง 23

รูปที่ 4.7 กราฟแสดงการหาสหสัมพันธ์ ระหว่างสัญญาณเสียงทดสอบ “เจ็ด” กับสัญญาณเสียงอ้างอิง 24

รูปที่ 4.8 กราฟแสดงการหาสหสัมพันธ์ ระหว่างสัญญาณเสียงทดสอบ “แปด” กับสัญญาณเสียงอ้างอิง 25




เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# สารบัญภาพประกอบ(ต่อ)

หน้า

รูปที่ 4.9 กราฟแสดงการหาสหสัมพันธ์ ระหว่างสัญญาณเสียงทดสอบ “เก้า” กับสัญญาณเสียงอ้างอิง 26
รูปที่ 4.10 กราฟแสดงการหาสหสัมพันธ์ ระหว่างสัญญาณเสียงทดสอบ “ศูนย์” กับสัญญาณเสียงอ้างอิง 27
รูปที่ 5.1 กราฟแสดงสัญญาณเสียง ”สี่” ทำ การปรับขนาดความยาวของเสียงแล้วคำตอบออกมาผิด 29
รูปที่ 5.2 กราฟแสดงสัญญาณเสียง ”สี่” ทำ การปรับขนาดความยาวของเสียงแล้วคำตอบออกมาถูกต้อง 30
รูปที่ 5.3 กราฟแสดงการหาจุดเริ่มและจุดท้ายของสัญญาณเสียง ”ห้า” 31
รูปที่ 1 โปรแกรม Matlab 35
รูปที่ 2 แสดงหน้าจอCommandของโปรแกรมMatlab 36
รูปที่ 3 แสดงการเปลี่ยนCurrent Directory 37
รูปที่ 4 แสดงหน้าจอการรันโปรแกรมจาก Command Promp 38
รูปที่ 5 แสดงหน้าจอโปรแกรมศึกษาการรู้จำเสียงพูด 39
รูปที่ 6 แสดงหน้าจอเมื่อทำการคลิกปุ่ม “สุ่มตัวเลข” 40
รูปที่ 7 แสดงหน้าจอเมื่อผู้พูดออกเสียงได้อย่างถูกต้อง 41
รูปที่ 8 แสดงหน้าจอเมื่อผู้พูดออกเสียงไม่ถูกต้อง 41



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# สารบัญภาพประกอบ(ต่อ)

หน้า

รูปที่ 4.23 กราฟแสดงสัญญาณในโดเมนความถี่ที่ระหว่างสัญญาณเสียงทดสอบเสียง “ห้า” กับ *สัญญาณเสียง อ้างอิงเสียง “หก”* 42

รูปที่ 4.24 กราฟแสดงสัญญาณในโดเมนความถี่ที่ระหว่างสัญญาณเสียงทดสอบเสียง “หก” กับ สัญญาณเสียง อ้างอิงเสียง “หก” 43

53

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# สารบัญตาราง

หน้า

ตารางที่ 4.1แสดงเฉพาะผลลัพธ์ของเอาท์พุทที่ตรงกับของค่าสหสัมพันธ์ที่มากที่สุดของเสียง *ตัวอย่าง "ศูนย์" ถึง "เก้า" ชุดที่ 1 ถึง ชุดที่ 10* 28

ตารางที่ 4.1แสดงเฉพาะผลลัพธ์ของเอาท์พุทที่ตรงกับของค่าสหสัมพันธ์ที่มากที่สุดของเสียง ตัวอย่าง "ศูนย์" ถึง "เก้า" ชุดที่ 11 ถึง ชุดที่ 20 29






เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# บทที่ 1
# บทนำ

ในอดีตการใช้งานคอมพิวเตอร์หรืออุปกรณ์อิเล็กทรอนิกส์มักจะสั่งงานโดยใช้อุปกรณ์ได้แก่ คีย์บอร์ด,เมาส์,จอยสติค หรือ มอนิเตอร์แบบสัมผัส จะเห็นได้ว่าส่วนใหญ่จะเป็นการสั่งงานโดยใช้นิ้ว ซึ่งเป็นเรื่องธรรมดาสำหรับบุคคลปกติทั่วไป แต่สำหรับบุคคลที่พิการโดยไม่สามารถใช้อวัยวะส่วนใดๆได้เลยคงจะเป็นเรื่องที่ลำบากมาก

แต่ด้วยความเจริญก้าวหน้าทางเทคโนโลยีในปัจจุบันก็ทำลายข้อจำกัดนี้ลงได้โดยใช้การสั่งงานด้วยเสียงแทนการป้อนอินพุตด้วยมือ ซึ่งแม้ว่าจะยังไม่สามารถสั่งงานให้คอมพิวเตอร์ได้ 100%แต่เทคโนโลยีการสั่งงานด้วยเสียงนี้ก็สามารถนำไปใช้กับงานหลายๆงานได้อย่างกว้างขวางขึ้นอยู่กับการประยุกต์ใช้ ตัวอย่างเช่น ในโทรศัพท์มือถือรุ่นใหม่ๆที่ออกมาขายมักจะมีฟังก์ชันการโทรออกโดยการสั่งงานด้วยเสียงผนวกมาด้วย

**วัตถุประสงค์**

1. เพื่อศึกษาระบบการรู้จำเสียงของคอมพิวเตอร์
2. เพื่อศึกษาความเป็นไปได้ในการพัฒนาโปรแกรมสั่งงานคอมพิวเตอร์ด้วยเสียงพูด
3. เพื่อศึกษาถึงการใช้โปรแกรม MATLAB ในการพัฒนาระบบการรู้จำเสียงของคอมพิวเตอร์

**ขอบเขตในการศึกษาและทดลอง**

1.ศึกษาระบบรู้จำเสียงตัวเลขภาษาไทย ตั้งแต่ ศูนย์ถึงเก้า

2.ใช้วิธีการเปรียบแบบ Correlation

3.รูปแบบเสียงอินพุตต้องเป็นเสียงที่มีสัญญาณรบกวนน้อย

4.ใช้โปรแกรม MATLAB ใน การศึกษาและทดลอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# บทที่2
# ทฤษฎีที่เกี่ยวข้อง

## 2.1 ธรรมชาติการรู้จำ

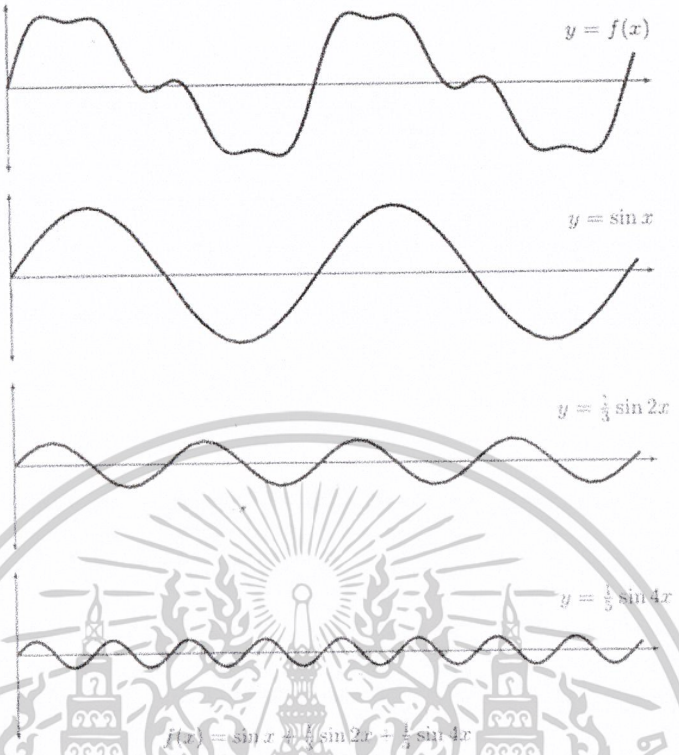
ระบบรู้จำเสียงของมนุษย์ มีความละเอียดอ่อน มีการเรียนรู้ และปรับปรุงไปพร้อมกับการพัฒนาของร่างกายตั้งแต่เยาว์ ระดับการรู้จำจะดำเนินไปตามขั้นตอนของมัน ประกอบกับมีโครงสร้างที่มีประสิทธิภาพสอดคล้องกับระบบการได้ยิน ทำให้สามารถแยกแยะคำและเสียงที่ซับซ้อนได้อย่างดี ความชำนาญในการฟังเกิดจากการเรียนรู้และฝึกฝน

ดังนั้นการสร้างเครื่องจักรให้มีความสามารถรู้จำเสียงได้อย่างมีประสิทธิภาพเท่าเทียมกับมนุษย์เป็นเรื่องค่อนข้างยากหรือหากจะทำเครื่องจักรจะต้องทำให้มีลักษณะการลอกเลียนแบบพฤติกรรมของมนุษย์คือการเรียนรู้การฝึกฝนมีการปรับปรุงอยู่ตลอดเวลาและต้องกระทำอย่างอัตโนมัติ การวิจัยดังกล่าวยังต้องอาศัยเวลาอีกนานพอสมควรเพราะต้องใช้คอมพิวเตอร์ที่มีความเร็วสูงในการฝึกฝน ซึ่งต้องใช้เวลามากหากต้องการฝึกฝนบ่อยๆ และเมื่อต้องการให้ระบบรู้จักคำมากขึ้นประสิทธิภาพและความเร็วจะลดลง นอกจากนี้ระบบก็ยังต้องการอัลกอริทึมหลายๆแบบร่วมกันที่ให้ความเร็วสูง มีความแม่นยำสูง ลึกซึ้งทำงานร่วมกับจิตใจซึ่งเป็นการยากที่จะสร้างระบบการวิเคราะห์ให้มีความสามารถสูงดังเช่น ระบบการวิเคราะห์เสียงมนุษย์ การวิจัยจึงมุ่งเน้นประเด็นการรู้จำเสียงในจำนวนที่จำกัด กลุ่มทดสอบการใช้งานและกลุ่มของตัวอย่างที่ค่อนข้างแคบและกำหนดให้เพียงพอต่อการใช้งานเฉพาะกิจโดยมุ่งเน้นความถูกต้องเป็นหลัก

## 2.2 Fourier Transform

ฟังก์ชันคาบสามารถเขียนเป็นผลรวมของสัญญาณไซน์และโคไซน์ ที่มีขนาดและความถี่แตกต่างกัน[3] ดังรูปที่ 2.1 แสดงการแตกองค์ประกอบของสัญญาณออกมาเป็นฟังก์ชันไซน์ที่มีขนาดและความถี่แตกต่างกัน ดังรูป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

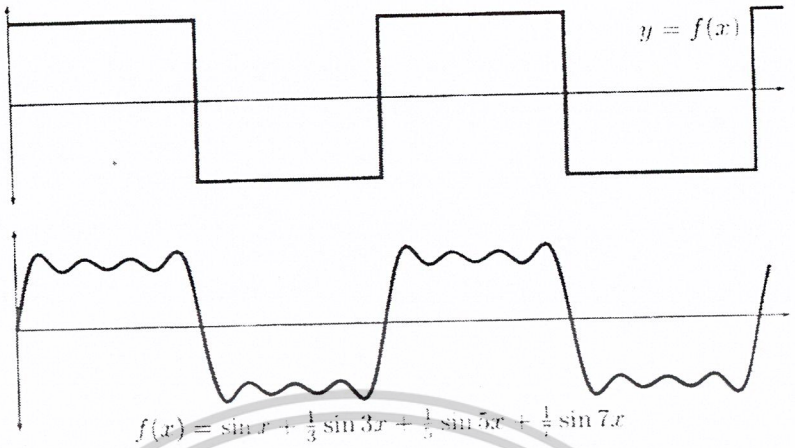



รูปที่ 2.1 แสดงฟังก์ชั่นและการแตกองค์ประกอบตรีโกณมิติ

บางฟังก์ชันจะกำหนดจำนวนฟังก์ชันที่จะแตกองค์ประกอบเป็นจำนวนที่แน่นนอน แต่บางฟังก์ชันจะกำหนดจำนวนฟังก์ชันที่จะแตกองค์ประกอบเป็นจำนวนที่ไม่มีขอบเขต ดังรูปที่ 2.2 ซึ่งแสดงการแตกองค์ประกอบของฟังก์ชันสี่เหลี่ยม จากสมการ

$$f(x) = \sin x + \frac{1}{3}\sin 3x + \frac{1}{5}\sin 5x + \frac{1}{7}\sin 7x + \frac{1}{9}\sin 9x + \ldots \quad \ldots (2.1)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้




รูปที่ 2.2 แสดงสัญญาณสี่เหลี่ยมและการประมาณตรีโกณมิติ

จากรูปที่ 2.2 แสดงการแตกองค์ประกอบจองฟังก์ชันสี่เหลี่ยมออกมาเป็น 4 เทอม ซึ่งเป็นการประมาณฟังก์ชัน จำนวนเทอมที่เราใส่เพิ่มนั้นจะทำให้ฟังก์ชันมีความใกล้เคียงฟังก์ชันดั่งเดิม สามารถเขียนให้เป็นแบบแผน โดยถ้า $f(x)$ เป็นฟังก์ชันที่มีคาบเท่ากับ 2 T จะเขียนได้เป็น

$$f(x) = a_0 + \sum_{n=1}^{\infty}\left(a_n \cos\frac{n\pi x}{T} + b_n \sin\frac{n\pi x}{T}\right) \qquad \ldots.(2.2)$$

โดยที่

$$a_0 = \frac{1}{2T}\int_{-T}^{T} f(x)dx$$

$$a_n = \frac{1}{T}\int_{-T}^{T} f(x)\cos\frac{n\pi x}{T}dx, n = 1,2,3,...$$

$$b_n = \frac{1}{T}\int_{-T}^{T} f(x)\sin\frac{n\pi x}{T}dx, n = 1,2,3,...$$

สมการนี้เป็นสมการขยายอนุกรมฟูเรียร์ ( Fourier series Expansion ) ของ $f(x)$ และสามารถแสดงในรูปแบบจำนวนจินตภาพ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$f(x)=\sum_{n=-\infty}^{\infty} c_n \exp\left(\frac{in\pi x}{T}\right)dx, \qquad \text{....(2.3)}$$

โดยที่

$$c_n=\frac{1}{2T}\int_{-T}^{T} f(x)\exp\left(\frac{-in\pi x}{T}\right)dx$$

ถ้าฟังก์ชันไม่เป็นฟังก์ชันคาบ ( Non – periodic ) เราจะได้ผลลัพธ์ที่เหมือนกัน โดยให้ T เข้าใกล้ $\infty$ ในกรณีนี้

$$f(x)=\int_{0}^{\infty}[a(\omega)\cos\omega x+b(\omega)\sin\omega x]d\omega \qquad \text{....(2.4)}$$

โดยที่

$$a(\omega)=\frac{1}{\pi}\int_{-\infty}^{\infty} f(x)\cos\omega x dx,$$

$$b(\omega)=\frac{1}{\pi}\int_{-\infty}^{\infty} f(x)\sin\omega x dx$$

สมการนี้สามารถเขียนในรูปแบบจินตภาพได้ดังนี้

$$f(x)=\int_{-\infty}^{\infty} F(\omega)e^{i\omega x}d\omega,$$

$$F(\omega)=\frac{1}{2\pi}\int_{-\infty}^{\infty} f(x)e^{i\omega x}dx \qquad \text{....(2.5)}$$

**2.2.1 การแปลงฟูริเยร์ไม่ต่อเนื่องหนึ่งมิติ ( The one dimensional Discrete Fourier Transform)**

นิยามของการแปลงฟูริเยร์ไม่ต่อเนื่องหนึ่งมิติ

สมมติให้

$$f=[f_0,f_1,f_2,...,f_{n-1}] \qquad \text{....(2.6)}$$

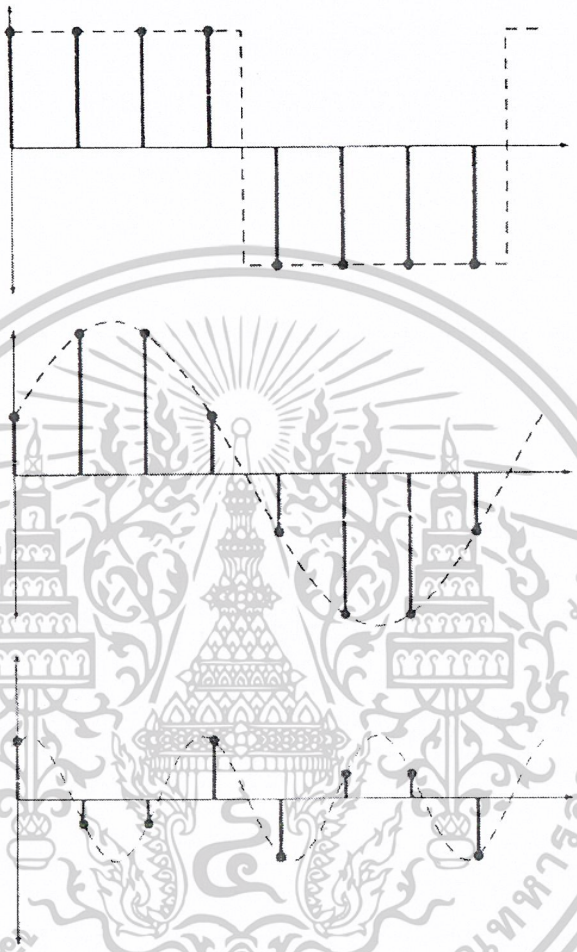
เป็นลำดับข้อมูลความยาว N เรานิยม DFT ของมันเป็นลำดับดังนี้

$$F=[F_0,F_1,F_2,...,F_{N-1}], \qquad \text{....(2.7)}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยที่

$$F_u = \frac{1}{N}\sum_{x=0}^{N-1}\exp\left[-2\pi i\frac{xu}{N}\right]f_x$$



รูปที่ 2.3 แสดงผลรวมฟังก์ชั่นไม่ต่อเนื่องของไซน์

ขณะนี้เราได้ผลบวกในช่วงจำกัดการนิยามนี้ สามารถแสดงได้ โดยการคูณกันของเมทริกต์

$$F = Ff, \qquad \ldots.(2.8)$$

โดยที่ f เป็น N x N matrix นิยามโดย

$$F_{m,n} = \frac{1}{N}\exp\left[-2\pi i\frac{mn}{N}\right] \qquad \ldots.(2.9)$$

ให้ N , เราจะนิยาม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$\omega = \exp\left[\frac{-2\pi i}{N}\right] \qquad \text{....(2.10)}$$

ดังนั้น

$$F_{m,n} = \frac{1}{N}\omega^{mn} \qquad \text{....(2.11)}$$

เราสามารถเขียนได้

$$F = \frac{1}{N}\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & K & 1 \\ 1 & \omega^{1} & \omega^{2} & \omega^{3} & \omega^{4} & K & \omega^{N-1} \\ 1 & \omega^{2} & \omega^{4} & \omega^{6} & \omega^{8} & K & \omega^{2(N-1)} \\ 1 & \omega^{3} & \omega^{6} & \omega^{9} & \omega^{12} & K & \omega^{3(N-1)} \\ 1 & \omega^{4} & \omega^{8} & \omega^{12} & \omega^{16} & K & \omega^{4(N-1)} \\ M & M & M & M & M & O & M \\ 1 & \omega^{N-1} & \omega^{2(N-1)} & \omega^{3(N-1)} & \omega^{4(N-1)} & K & \omega^{(N-1)^2} \end{bmatrix} \qquad \text{....(2.12)}$$

<u>ตัวอย่าง</u> ให้ f = [1,2,3,4]และ N = 4 ดังนี้

$$\omega = \exp\left[\frac{-2\pi i}{4}\right]$$

$$\omega = \exp\left[\frac{-\pi i}{2}\right]$$

$$\omega = \cos\left(\frac{-\pi}{2}\right) + i\sin\left(\frac{-\pi}{2}\right)$$

$$\omega = -i$$

เรามี

$$F = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -i & (-i)^2 & (-i)^3 \\ 1 & (-i)^2 & (-i)^4 & (-i)^6 \\ 1 & (-i)^3 & (-i)^6 & (-i)^9 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -i & -1 & i \\ 1 & -1 & 1 & -1 \\ 1 & i & -1 & -i \end{bmatrix}$$

จะได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$F=\frac{1}{4}\begin{bmatrix}1 & 1 & 1 & 1\\ 1 & -i & -1 & i\\ 1 & -1 & 1 & -1\\ 1 & i & -1 & -i\end{bmatrix}\begin{bmatrix}1\\2\\3\\4\end{bmatrix}=\frac{1}{4}\begin{bmatrix}10\\-2+2i\\-2\\-2-2i\end{bmatrix}$$

**2.2.2 การแปลงฟูเรียร์ไม่ต่อเนื่องหนึ่งมิติผกผัน ( The inversethe - one dimention Discrete Fourier Transform)**

สูตรของการแปลงฟูเรียร์ไม่ต่อเนื่องหนึ่งมิติผกผัน มีความคลึงกันกับการแปลงฟูเรียร์ไม่ต่อเนื่องหนึ่งมิติแบบไม่ผกผัน

$$x_u=\sum_{x=0}^{N-1}\exp\left[2\pi i\frac{xu}{N}\right]F_u \qquad \text{....(2.13)}$$

ถ้าเปรียบเทียบการแปลงฟูเรียร์ไม่ต่อเนื่องหนึ่งมิติส่วนที่แตกต่าง คือ

1. ไม่มีตัวแปรสเกลลิ่ง ( $\frac{1}{N}$ )
2. เครื่องหมายภายในแอกซ์โปเนนเชียนเป็นเครื่องหมายบวก

สามารถแสดงเป็นการคูณกันแบบเมตริกซ์

$$f=F^{-1}F \qquad \text{....(2.14)}$$

โดย

$$F^{-1}=\begin{bmatrix}1 & 1 & 1 & 1 & 1 & \text{K} & 1\\ 1 & \overline{\omega}^{1} & \overline{\omega}^{2} & \overline{\omega}^{3} & \overline{\omega}^{4} & \text{K} & \overline{\omega}^{N-1}\\ 1 & \overline{\omega}^{2} & \overline{\omega}^{4} & \overline{\omega}^{6} & \overline{\omega}^{8} & \text{K} & \overline{\omega}^{2(N-1)}\\ 1 & \overline{\omega}^{3} & \overline{\omega}^{6} & \overline{\omega}^{9} & \overline{\omega}^{12} & \text{K} & \overline{\omega}^{3(N-1)}\\ 1 & \overline{\omega}^{4} & \overline{\omega}^{8} & \overline{\omega}^{12} & \overline{\omega}^{16} & \text{K} & \overline{\omega}^{4(N-1)}\\ \text{M} & \text{M} & \text{M} & \text{M} & \text{M} & \text{O} & \text{M}\\ 1 & \overline{\omega}^{N-1} & \overline{\omega}^{2(N-1)} & \overline{\omega}^{3(N-1)} & \overline{\omega}^{4(N-1)} & \text{K} & \overline{\omega}^{(N-1)^2}\end{bmatrix}$$

โดยที่

$$\overline{\omega}=\frac{1}{\omega}=\exp\left[\frac{2\pi i}{N}\right]$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

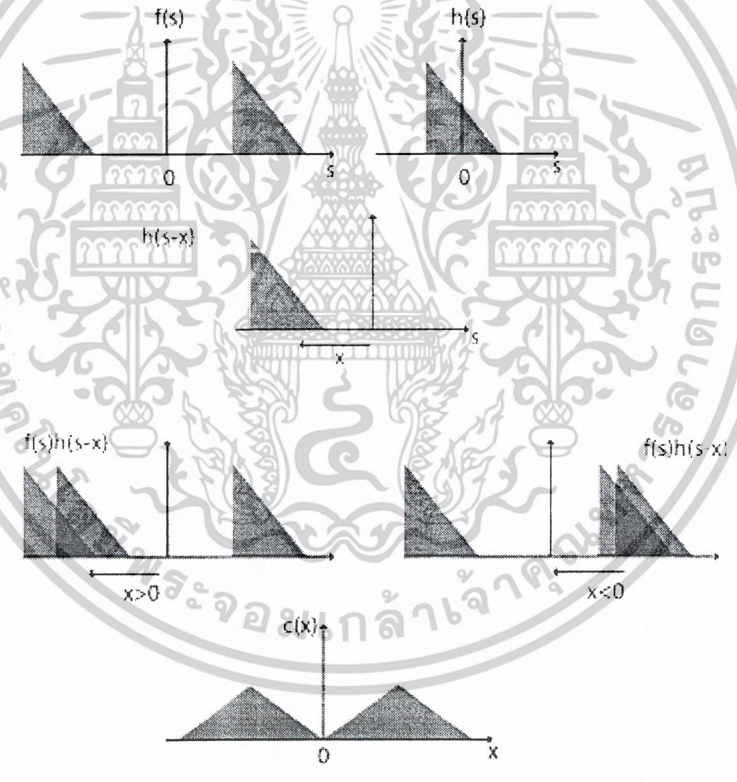
## 2.3 คอร์รีเลชั่น ( Corrletion )

การคอร์รีเลชั่นฟังก์ชันสองฟังก์ชัน หรือสัญญาณสองสัญญาณเป็นการเลื่อนฟังก์ชั่นหนึ่งไปทับกับอีกฟังก์ชั่นหนึ่งแล้วหาพื้นที่กับของสองสัญญาณ ดังสมการ

$$c(x) = f(x) \otimes h(x) \qquad \text{....(2.15)}$$

$$= \int_{-\infty}^{\infty} f(s)h^{*}(s-x)ds \qquad \text{....(2.16)}$$

เมื่อ h * (x) คือการคอนจูเกต(Conjugate) จำนวนจินตภาพของฟังก์ชั่น h (x)




รูปที่ 2.4 แสดงการคอร์รีเลชั่นของสองฟังก์ชั่น

ถ้าพิจารณาว่า f (x) เป็นสัญญาณอินพุต และ h (x) เป็นสัญญาณอ้างอิง เมื่อนำสองสัญญาณนี้มาทำการคอร์รีเลชั่นกันแล้วสัญญาณที่ได้จากการคอร์รีเลชั่นมีค่า Amplstnde มากก็แสดงว่าสัญญาณทั้งสองมีความคล้ายคลึง ดังนั้นการคอร์รีเลชั่นถือเป็นวิธีพื้นฐานที่ง่ายในการเปรียบเทียบความคล้ายคลึงของสัญญาณสองสัญญาณ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การคอร์รีเลชั่น สามารถทำได้ทั้ง โดเมนของเวลาและโดเมนของความถี่ คุณสมบัติที่สำคัญของการคอร์รีเลชั่น คือ การคอร์รีเลชั่นในโดเมนของเวลาจะเท่ากับการคูณกันในโดเมนของความถี่ ดังสมการ

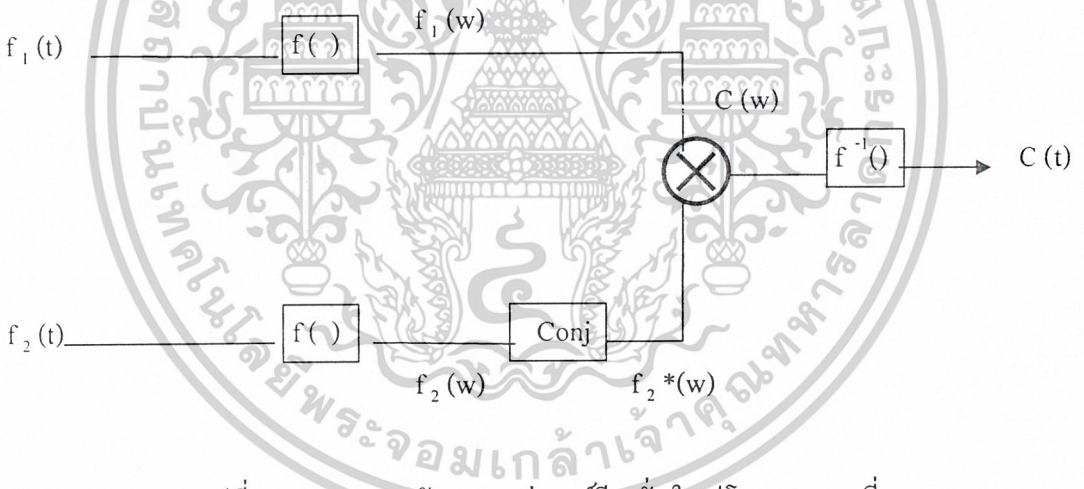
$$C(u) = F(u)\, H^{*}(u) \qquad \ldots.(2.17)$$

เมื่อ

$$C(u) = F\{C(x)\}$$

$$F(u) = F\{f(x)\}$$

$$H(u) = F\{h(x)\}$$

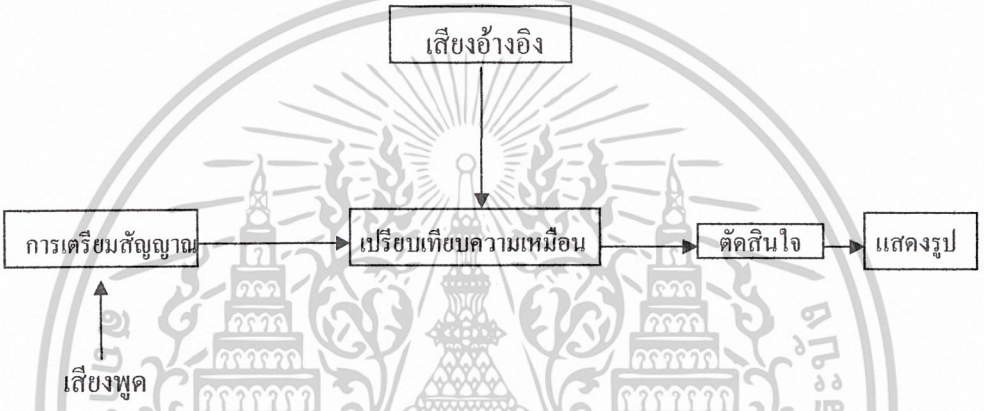



รูปที่ 2.5 แสดงแผนผังการหาค่าคอร์รีเลชั่นในฟโดเมนความถี่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# บทที่3
# การออกแบบ และการสร้าง

การรู้จำเสียงเสียงพูดเป็นกระบวนการเปรียบเทียบว่าเสียงทดสอบที่เข้ามามีความสัมพันธ์กับเสียงอ้างอิงใดมากที่สุด หลังจากนั้นจะเข้าสู่กระบวนการตัดสิน




รูปที่ 3.1 แสดงแผนผังการทำงานของระบบรู้จำเสียงพูด

## 3.1 การเตรียมสัญญาณ

เป็นขั้นตอนในการจัดเตรียมข้อมูลดิบของเสียงพูดที่ได้รับจากการบันทึก เพื่อใช้ในการประมวลผลในลำดับต่อไป

### 3.1.1 การหาจุดเริ่มและจุดท้ายของเสียง

การหาจุดเริ่มและจุดท้ายของเสียงเป็นการหาช่วงของข้อมูลเสียงที่เป็นส่วนของเนื้อเสียง แล้วตัดส่วนที่ไม่ใช่เนื้อเสียงทิ้ง จะทำให้ได้ส่วนที่เป็นข้อมูลเสียงจริงๆ ในการทดลองนี้ใช้การหาค่าพลังงานในการหาจุดเริ่มและจุดท้ายของเสียเพราะเป็นวิธีที่ง่าย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

วิธีการหาจุดเริ่มและจุดท้ายของเสียง โดยวิธีการหาค่าพลังงาน[5]เพราะจะเริ่มจากแบ่งข้อมูลเสียงที่รับมาออกเป็นเฟรม โดยแต่ละเฟรมมีขนาดเท่า ๆ กัน หลังจากนั้นคำนวณหาค่าพลังงานในแต่ละเฟรมจากสมการ

$$E_l(0) = \sum_{N=0}^{N-1} |X(n)|^2 \quad , l = 0 , 1 , \ldots l-1 \qquad \ldots(3.1)$$

โดยที่ $E_l(0)$ คือ ค่าพลังงานในแต่ละระเฟรม

N คือ ขนาดของเฟรม ( frame size )

l คือ ลำดับที่ของเฟรม

เมื่อคำนวณค่าพลังงานจนครบทุกเฟรมแล้ว จากนั้นจะกำหนดระดับของพลังงานขีดเริ่มเปลี่ยน( Threshold ) a เพื่อใช้ในการหาจุดเริ่มและจุดท้ายของเสียง ดังรูป

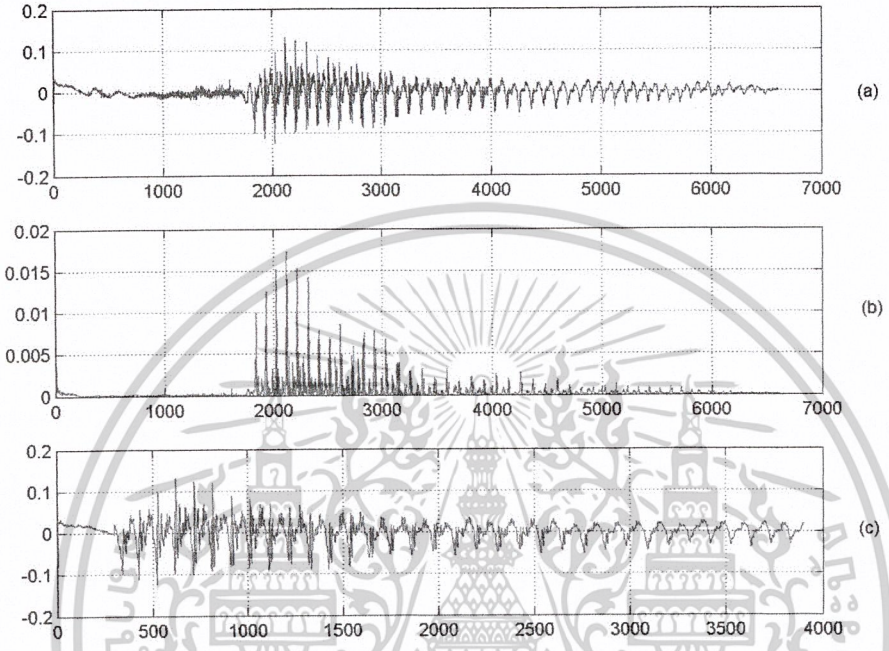
รูปที่ 3.2 แสดงการหาจุดเริ่มและจุดท้ายของเสียง

โดยที่ค่า a จะมีค่าเท่ากับ 10 % ของเฟรมที่มีค่าพลังงานมากที่สุด และ T1 เป็นเฟรมเริ่มต้น T2 เฟรมสุดท้าย โดยค่า 10 % นี้ได้จากการทดลองที่ให้ผลดีในการตัดข้อมูลที่ไม่ต้องการออกไปได้ดีที่สุด[7] เพราะถ้าค่า a มากกว่า 10 % ส่วนที่เป็นข้อมูลเสียงบางส่วนจะถูกตัดออกไปด้วย แต่ถ้าค่า a มีค่าน้อยกว่า 10 % จะทำให้ส่วนที่ไม่ใช่ข้อมูลเสียง ติดมากับข้อมูลเสียง

$$a = 0.1\ E_{max} \qquad \ldots.(3.2)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

นั่นก็คือการตัดเฟรมที่มีค่าน้อยกว่า 10 % จากเฟรมที่มีค่าพลังงานมากที่สุด




รูปที่ 3.3 รูปแสดงการหาจุดเริ่มต้นและจุดท้ายเสียง

(a) กราฟแสดงสัญญาณเสียงอินพุต

(b) กราฟแสดงพลังงานของสัญญาณเสียงพูด

(c) กราฟแสดงสัญญาณเสียงอินพุตที่หาจุดเริ่มต้นและจุดท้ายเสียง

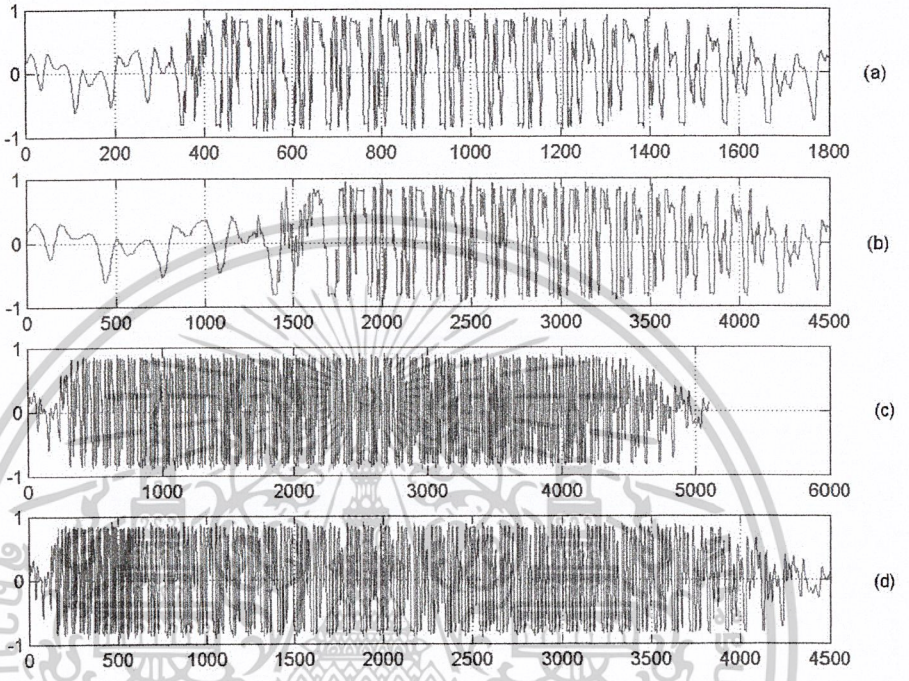
### 3.1.2 การปรับขนาดความยาวข้อมูล

ปัญหาพื้นฐานของระบบรู้จำเสียงพูด คือ ระยะเวลาที่ใช้กับการเปล่งเสียงออกมาในแต่ละครั้งไม่เท่ากัน ถึงแม้ว่าจะพูดคำเดียวกันก็ตาม จะทำให้ไม่สามารถเปรียบเทียบกับเสียงอ้างอิงได้ เพราะขนาดข้อมูล ทั้งสองมีขนาดไม่เท่ากัน ฉะนั้นจึงต้องมีการปรับขนาดข้อมูลให้มีขนาดเท่ากับเสียงอ้างอิงก่อนแล้ว จึงนำมาเปรียบเทียบ

ในกรณีที่ข้อมูลเสียงที่รับเข้ามามีขนาดความยาวสั้นกว่าเสียงอ้างอิงก็จะทำการเพิ่มตัวอย่างเข้าไปจนมีขนาดความยาวเท่ากับเสียงอ้างอิง และในกรณีที่ข้อมูลเสียงที่รับเข้ามามีขนาดความยาว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เกินยาวกว่าเสียงอ้างอิงก็จะทำการตัดอย่างออกจนมีขนาดความยาวสมกับเสียงอ้างอิง ดังรูป (ในโครงงานนี้จะปรับขนาดความยาวของข้อมูลเสียงทุก ๆ เสียงให้มีขนาด 4500 ข้อมูล )




รูปที่ 3.4 รูปแสดงการปรับขนาดความยาวข้อมูล

(a) กราฟแสดงสัญญาณเสียงอินพุตที่มีความน้อยกว่าความยาวมาตรฐาน

(b) กราฟแสดงสัญญาณที่ทำการปรับขนาดความยาวให้มีขนาดเท่ากับความยาวมาตรฐาน

(c) กราฟแสดงสัญญาณเสียงอินพุตที่มีความมากกว่าความยาวมาตรฐาน

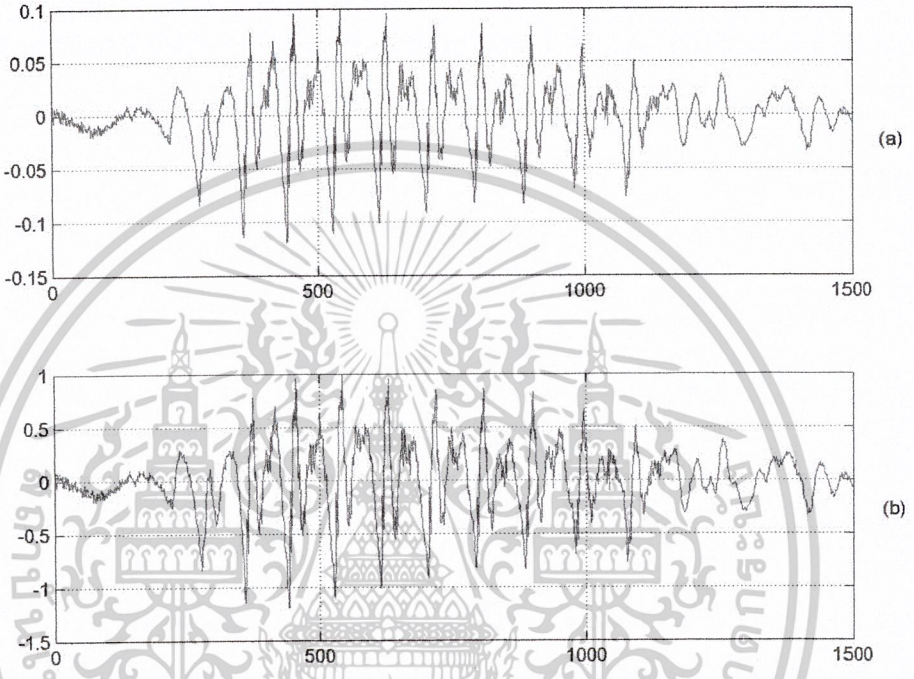
(d) กราฟแสดงสัญญาณที่ทำการปรับขนาดความยาวให้มีขนาดเท่ากับความยาวมาตรฐาน

### 3.1.3 การปรับขนาดแอมพิจูดข้อมูล

เนื่องจากว่าการเปล่งเสียงออกมาในแต่ละครั้ง จะมีแอมพิจูดไม่เท่ากัน ถึงแม้ว่าจะเป็นคนเดียวกัน เปล่งเสียงคำเดียวกันสองครั้ง แอมพิจูดของเสียงที่ได้ของทั้งสองครั้งมักจะมีคำแอมพิจูดไม่เท่ากัน ดังนั้นเมื่อนำไปเปรียบเทียบกับข้อมูลเสียงอ้างอิงจึงเกิดความผิดพลาด จึงต้องมีการปรับ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ระดับแอมพิจูดให้มีขนาดที่เป็นมาตรฐานเดียวกันกับเสียงอ้างอิงก่อน แล้วจึงนำมาทำการเปรียบเทียบในการทดลองนี้จะทำการหาค่าของข้อมูลที่มีค่ามากที่สุดแล้วนำมาหารทุกๆข้อมูล เมื่อหารเสร็จทุกระดับแล้วจะได้ข้อมูลใหม่ที่มีค่าสูงสุดเท่ากับหนึ่ง แล้วนำมาเปรียบเทียบกับข้อมูลเสียงอ้างอิง ซึ่งมีค่าสูงสุดเท่ากับหนึ่งเหมือนกัน




รูปที่ 3.5 รูปแสดงการปรับขนาดแอมพิจูดข้อมูล

(a) กราฟแสดงสัญญาณเสียงอินพุตที่ยังไม่ปรับขนาดแอมพิจูด

(b) กราฟแสดงสัญญาณเสียงอินพุตที่ปรับขนาดแอมพิจูด

## 3.2 การเปรียบเทียบความเหมือน

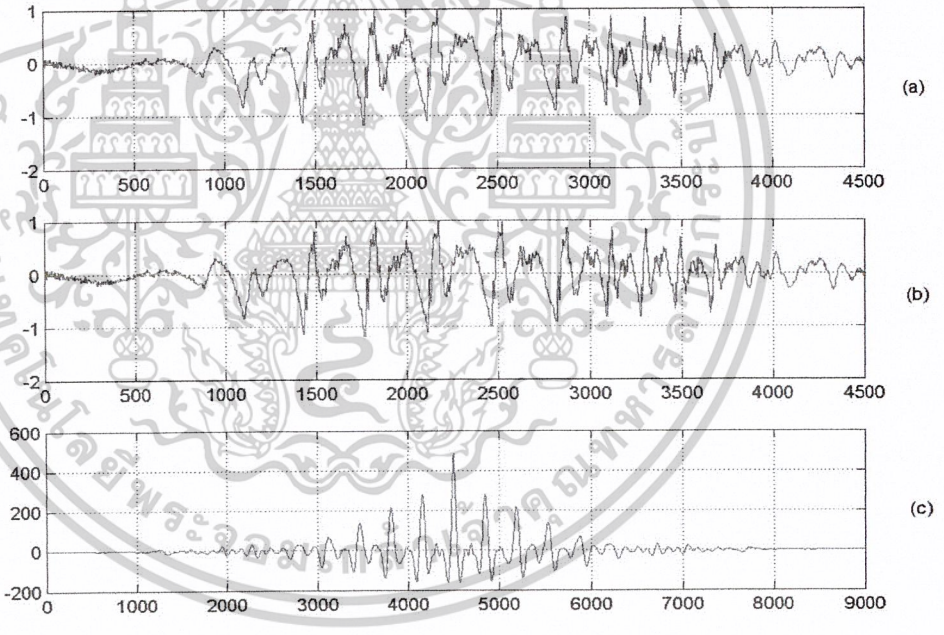
เมื่อข้อมูลเสียงที่รับเข้ามาผ่านกระบวนการเตรียมข้อมูล คือ การจุดเริ่มต้นและท้ายของเสียง,การปรับขนาดความยาวของข้อมูลและการปรับขนาดแอมพิจูด ให้อยู่เป็นมาตรฐานเดียวกันกับข้อมูลเสียงอ้างอิง ก็จะมาถึงกระบวนการเปรียบเทียบความเหมือน ระหว่างเสียงที่รับเข้ามากับเสียงอ้างอิงทุก ๆ ตัว ว่ามีความเหมือนระหว่างเสียงที่รับเข้ามากับเสียงอ้างอิงใดมากที่สุด โดยใน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การทดลองนี้จะใช้วิธีการคอร์รีเลชั่นในการเปรียบเทียบความเหมือนที่เลือกวิธีนี้ก็เพราะเป็นวิธีที่ง่าย

**3.2.1 คอร์รีเลชั่น**

การทำคอร์รีเลชั่น คือ การนำสัญญาณสองสัญญาณที่จะทำการเปรียบเทียบกัน โดยการเลื่อนสัญญาณ $f_1(t)$ ไม่ทับกันสัญญาณ $f_2(t)$ แล้วคำนวณหาพื้นที่ที่ทั้งสองสัญญาณมาทับกัน เมื่อทำการคำนวณหาพื้นที่ที่ทับกันทั้งสองสัญญาณจะทำอย่างนี้เรื่อย ๆ จนสัญญาณทั้งสองไม่มีพื้นที่ทับกัน ผลจากการคำนวณหาพื้นที่ที่ทับกันของสองสัญญาณในการเลื่อนสัญญาณแต่ละครั้งจะทำให้ได้สัญญาณที่เกิดขึ้นมาใหม่อีกสัญญาณ สัญญาณที่เกิดมาใหม่นี้จะเป็นตัวบอกความเหมือนของทั้งสองสัญญาณว่าเหมือนกันมากแค่ไหน โดยดูจากค่าแอมพิจูดที่มีค่ามากที่สุดของสัญญาณที่เกิดขึ้นใหม่ ถ้าค่าแอมพิจูดมีค่ามากก็มีความเหมือนมาก แต่ถ้าค่าแอมมีค่าน้อยก็แสดงว่ามีความเหมือนกันน้อย




รูปที่ 3.6 รูปแสดงการทำคอร์รีเลชั่นระหว่างสัญญาณ f1(t) กับ f2(t)

(a) กราฟแสดงสัญญาณเสียงอินพุต f1(t)

(b) กราฟแสดงสัญญาณเสียงอินพุต f2(t)

(c) กราฟแสดงสัญญาณที่ได้จากการคอรร์รีเลชั่นระหว่างสัญญาณ f1(t) กับ f2(t)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

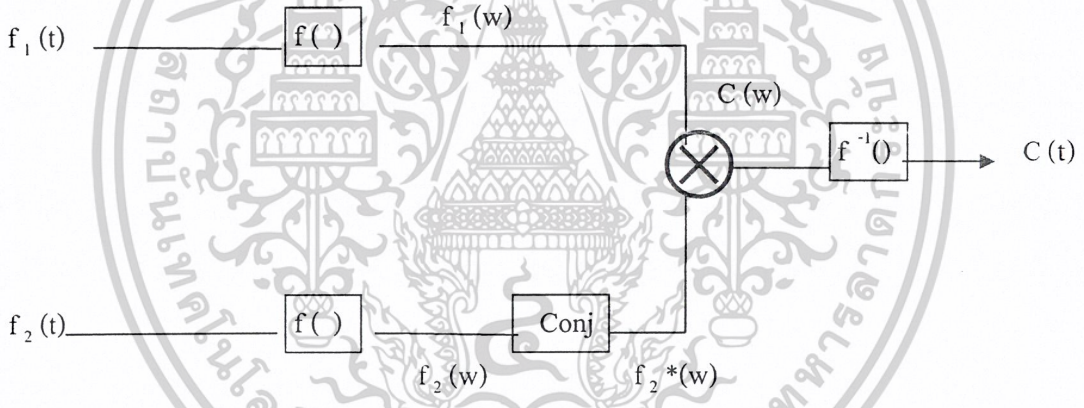


ทำคอร์รีเลชั่นแบบนี้เป็นการคอร์รีเลชั่นในโดเมนเวลา แต่ในการทดลองนี้จะทำการคอร์รีเลชั่นในโดเมนความถี่เพราะว่าจะใช้เวลาในการคำนวณน้อยกว่า

การทำคอร์รีเลชั่นในโดเมนความถี่จะทำการแปลงสัญญาณ $f_1(t)$ และ $f_2(t)$ ให้อยู่ในโดเมนความถี่ก่อน เมื่อแปลงสัญญาณ $f_1(t)$ และ $f_2(t)$ จะได้ $f_1(t) = F_1(w)$ และ $f_2(t) = F_2(w)$ แล้วทำการคอนจูเกต ( Congugcte) $f_2(w)$ จะได้ $f_2^*(w)$ เสร็จแล้วทำ $f_1(w)$ คูณกับ $f_2^*(w)$ จะได้ $C(w)$ ดังสมการ

$$C(w) = f_1(w)\, f_2^*(w) \qquad \ldots.(3.3)$$

จากนั้นแปลง $C(w)$ จากโดเมนความถี่ไปโดเมนเวลาจะได้ $C(t)$ ซึ่งเป็นค่าการคอร์รีเลชั่น




รูปที่ 3.7 แสดงแผนผังการหาค่าคอร์รีเลชั่นในโดเมนความถี่

## 3.3 การตัดสินใจ

เมื่อทำการเปรียบเทียบระหว่างเสียงที่รับเข้ามากับเสียงอ้างอิง โดยใช้การคอร์รีเลชั่นในการเปรียบเทียบข้อมูลเสียงทั้งสองชุด จะทำให้ได้ค่าการคอร์รีเลชั่นมาเป็นจำนวนเท่ากับจำนวนของเสียงที่ใช้ในการอ้างอิง แล้วเลือกค่าคอร์รีเลชั่นที่มีค่ามากสุด เพราะว่าค่าการคอร์รีเลชั่นที่มีค่ามากจะแสดงว่า ทั้งสองสัญญาณที่ทำการคอร์รีเลชั่นแล้วมีความเหมือนกันมาก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
61362

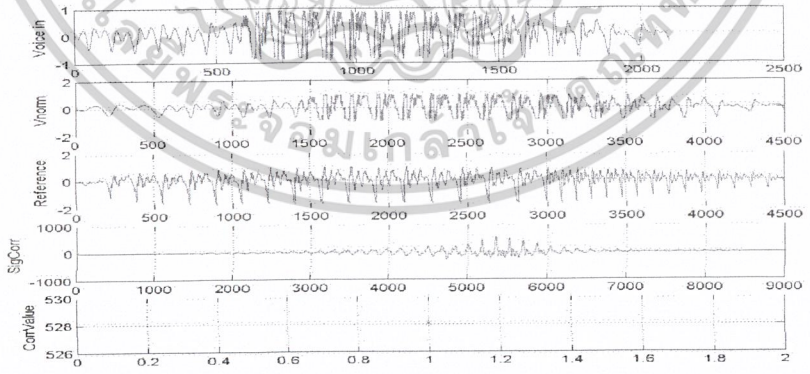
# บทที่ 4

## ผลการทดลอง

โครงงานนี้ใช้เสียงอ้างอิงเสียง "ศูนย์" ถึง "เก้า" จำนวน 5 ชุด และใช้เสียงพูดศูนย์ถึงเก้า จำนวน 20 ชุด เพื่อเป็นเสียงที่ใช้ในการทดสอบ การทดลองนี้เป็นไปตามขั้นตอนในบทที่ 3 โดยใช้เสียงที่เป็นเสียงทดสอบ ไปหาค่าสหสัมพันธ์กับเสียงอ้างอิงทั้ง 5 ชุด แล้วเลือกค่าสหสัมพันธ์ที่มีค่ามากสุดในแต่ละชุด ค่าสหสัมพันธ์ที่เลือกมาในแต่ละชุดจะเป็นตัวที่บอกว่าเสียงทดสอบที่นำมาเปรียบเทียบกับเสียงอ้างอิงในแต่ละชุด ว่ามีความเหมือนกับเสียงอ้างอิงเสียงใดมากที่สุด ในแต่ละชุดของเสียงอ้างอิงก็จะมีเสียงอ้างอิงหนึ่งเสียงที่มีค่าสหสัมพันธ์มากที่สุด ทั้งหมดจะได้มา 5 เสียง นำทั้ง 5 เสียงมาเปรียบเทียบกันโดยดูว่าเสียงใดเสียงใดซ้ำกันมากที่สุด แล้วเลือกเสียงที่เสียงที่ซ้ำกันมากที่สุดมาเป็นคำตอบ แต่ถ้าไม่มีเสียงใดซ้ำกันเลยก็จะพิจารณาจากค่าสหสัมพันธ์ที่มากที่สุดของเสียงอ้างอิงที่เลือกมา มาเป็นคำตอบ

ต่อไปเป็นผลการทดลองการหาค่าสหสัมพันธ์ของเสียงพูดศูนย์ถึงเก้ากับเสียงอ้างอิง รูปกราฟที่แสดงจะแสดงเฉพาะชุดของเสียงอ้างอิงที่มีค่าสหสัมพันธ์มากที่สุด

เสียงพูด "หนึ่ง"




รูปที่ 4.1 กราฟแสดงการหาสหสัมพันธ์ ระหว่างสัญญาณเสียงทดสอบ "หนึ่ง" กับสัญญาณเสียงอ้างอิงชุดที่ 3

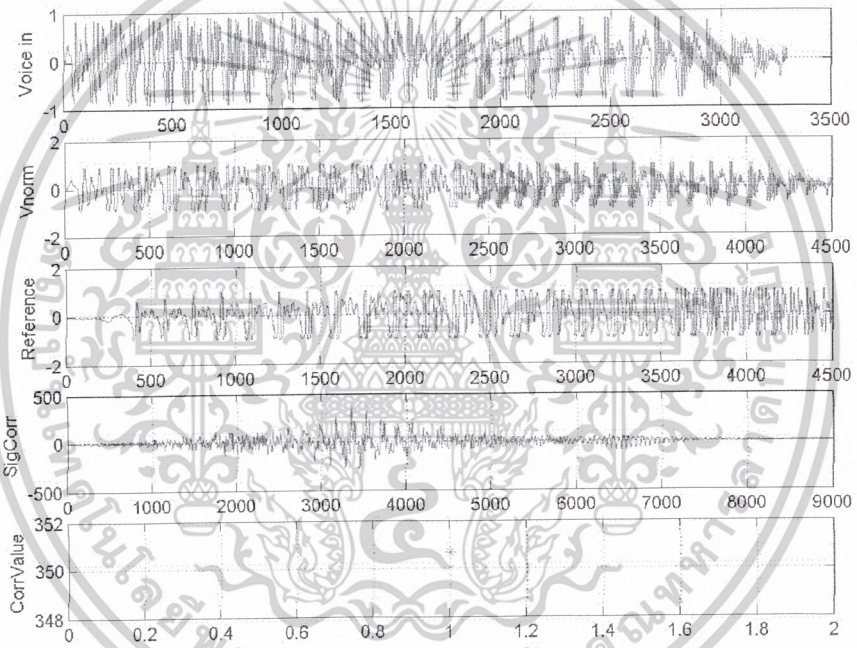
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ค่าสหสัมพันธ์ของเสียงพูด “หนึ่ง” กับเสียงอ้างอิงชุดที่ 3

| | | | | | | |
|---|---|---|---|---|---|---|
| 528.07 | 158.29 | 148.18 | 239.1 | 173.95 | 398.95 | 369.79 |
| 103.97 | 94.155 | 355.88 | | | | |

ค่าสหสัมพันธ์ที่มากที่สุดคือ 528.07 ซึ่งตรงกับเสียง 1 ดังนั้นคำตอบ ถูก

เสียงพูด “สอง”




รูปที่ 4.2 กราฟแสดงการหาสหสัมพันธ์ ระหว่างสัญญาณเสียงทดสอบ “สอง” กับสัญญาณเสียงอ้างอิงชุดที่ 5

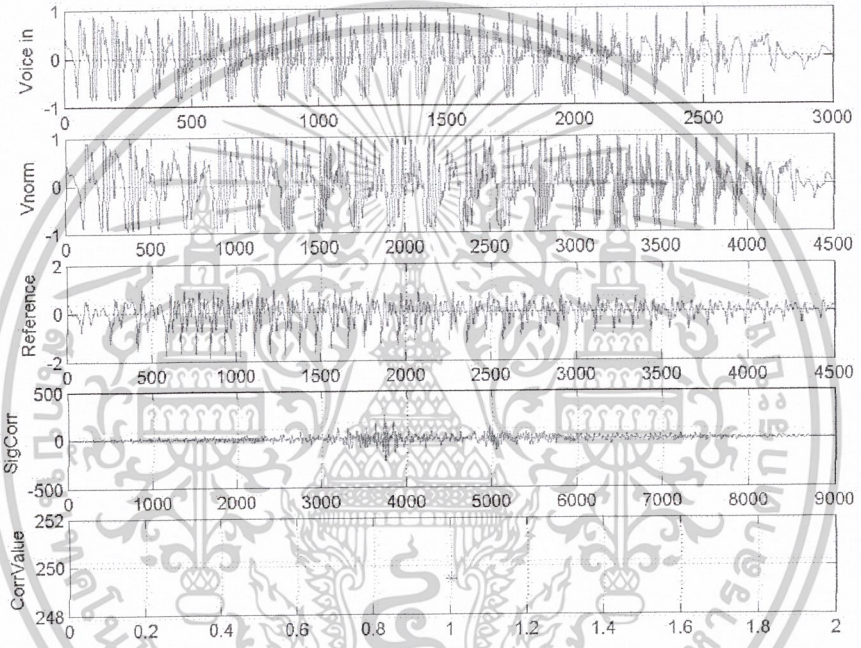
ค่าสหสัมพันธ์ของเสียงพูด “สอง” กับเสียงอ้างอิงชุดที่ 5

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 188.13 | 260.02 | 197.58 | 160.58 | 157.09 | 350.64 | 120.62 | 169.2 |
| 154.13 | 208.51 | | | | | | |

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ค่าสหสัมพันธ์ที่มากที่สุดคือ 350.64 ซึ่งตรงกับเสียง 6 ดังนั้นคำตอบ ผิด

เสียงพูด “สาม”




รูปที่ 4.3 กราฟแสดงการหาสหสัมพันธ์ ระหว่างสัญญาณเสียงทดสอบ “สาม” กับสัญญาณเสียงอ้างอิงชุดที่ 3

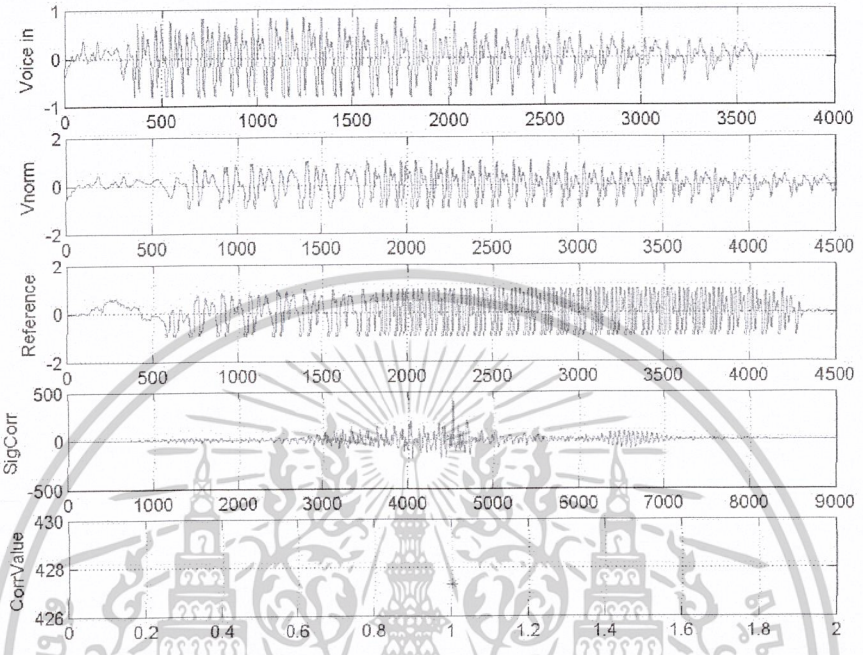
ค่าสหสัมพันธ์ของเสียงพูด “สาม” กับเสียงอ้างอิงชุดที่ 3

91.318 91.797 173.43 249.46 88.394 160.82 142.31 132.13
95.199 122.75

ค่าสหสัมพันธ์ที่มากที่สุดคือ 249.46 ซึ่งตรงกับเสียง 4 ดังนั้นคำตอบ ผิด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เสียงพูด “สี่”




รูปที่ 4.4 กราฟแสดงการหาสหสัมพันธ์ระหว่างสัญญาณเสียงทดสอบ “สี่”กับสัญญาณเสียงอ้างอิงชุดที่ 1

ค่าสหสัมพันธ์ของเสียงพูด “สี่” กับเสียงอ้างอิงชุดที่ 1

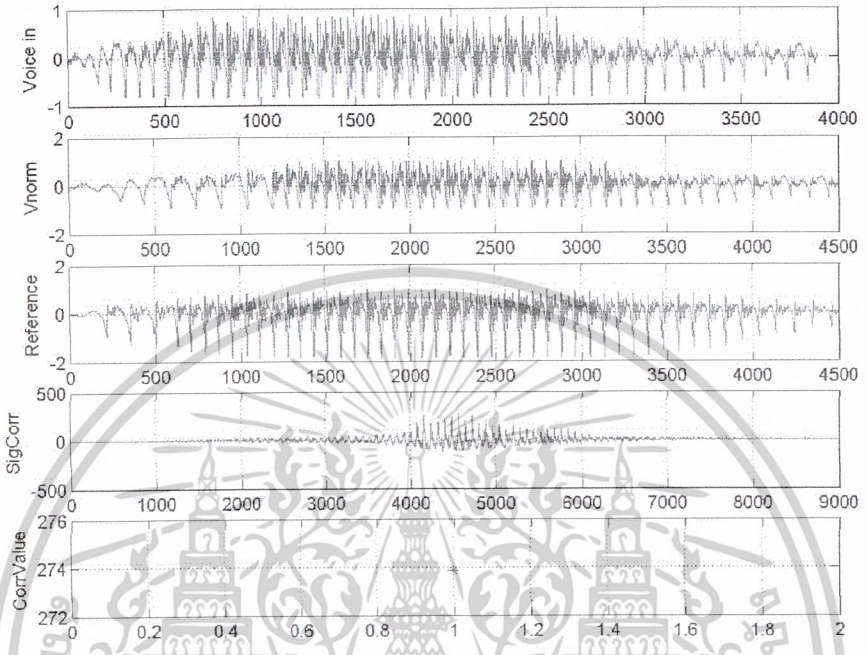
206.3 121.61 135.71 202.78 97.268 241.81 272.14 142.8

62.883 427.31

ค่าสหสัมพันธ์ที่มากที่สุดคือ 427.31 ซึ่งตรงกับเสียง 0 ดังนั้นคำตอบ ผิด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เสียงพูด “ห้า”




รูปที่ 4.5 กราฟแสดงการหาสหสัมพันธ์ ระหว่างสัญญาณเสียงทดสอบ “ห้า” กับสัญญาณเสียงอ้างอิงชุดที่ 1

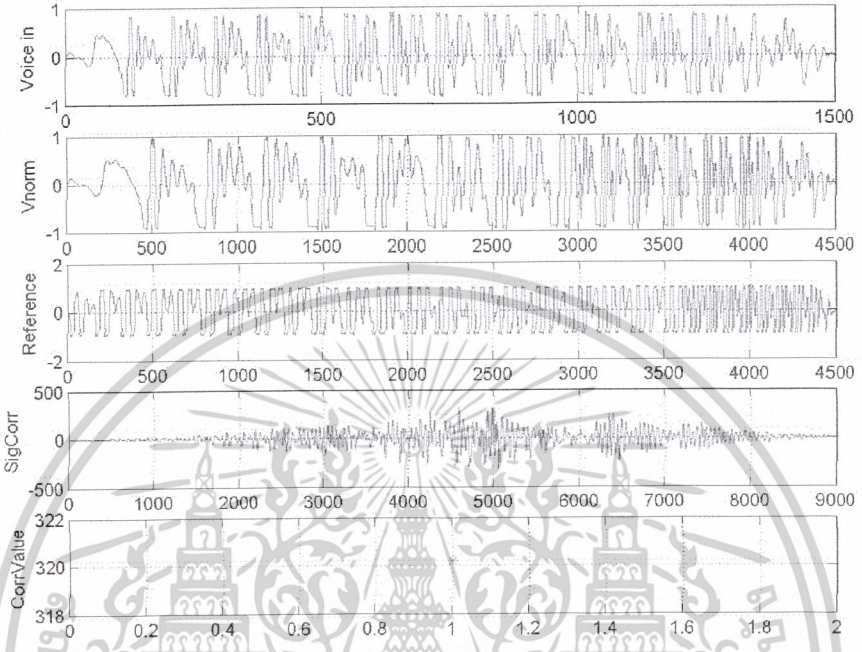
ค่าสหสัมพันธ์ของเสียงพูด “ห้า” กับเสียงอ้างอิงชุดที่ 1

77.222 67.871 45.86 119.03 273.87 134.31 89.457 45.04
113.24 132.96

ค่าสหสัมพันธ์ที่มากที่สุดคือ 273.87 ซึ่งตรงกับเสียง 5 ดังนั้นคำตอบ ถูก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เสียงพูด “หก”




รูปที่ 4.6 กราฟแสดงการหาสหสัมพันธ์ ระหว่างสัญญาณเสียงทดสอบ “หก” กับสัญญาณเสียงอ้างอิงชุดที่ 5

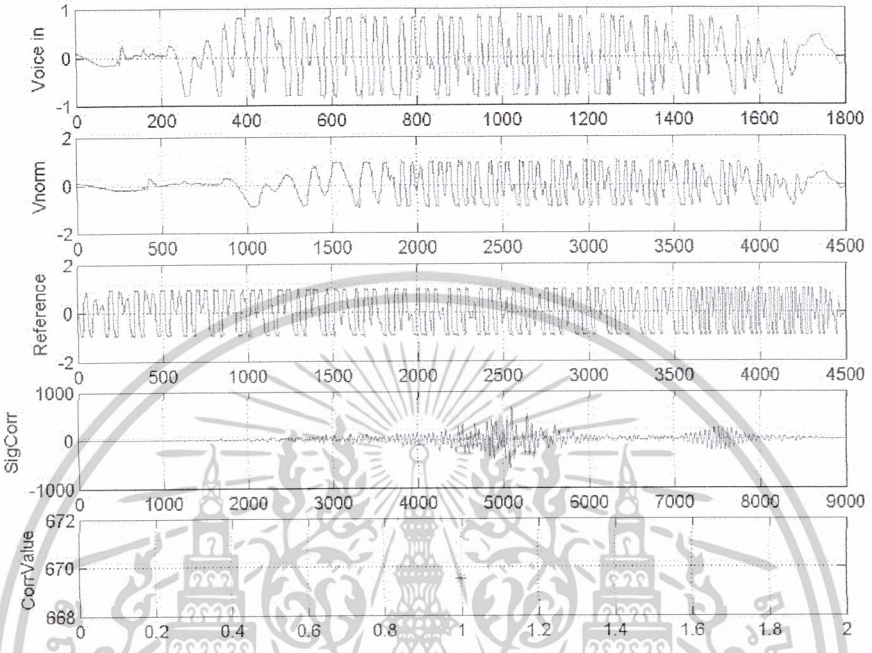
ค่าสหสัมพันธ์ของเสียงพูด “หก” กับเสียงอ้างอิงชุดที่ 5

236.48 117.4 149.49 168.33 119.46 276.04 226.52 108.78
105.24 320.17

ค่าสหสัมพันธ์ที่มากที่สุดคือ 320.17 ซึ่งตรงกับเสียง 0 ดังนั้นคำตอบ ผิด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เสียงพูด “เจ็ด”




รูปที่ 4.6 กราฟแสดงการหาสหสัมพันธ์ ระหว่างสัญญาณเสียงทดสอบ “เจ็ด” กับสัญญาณเสียงอ้างอิงชุดที่ 1

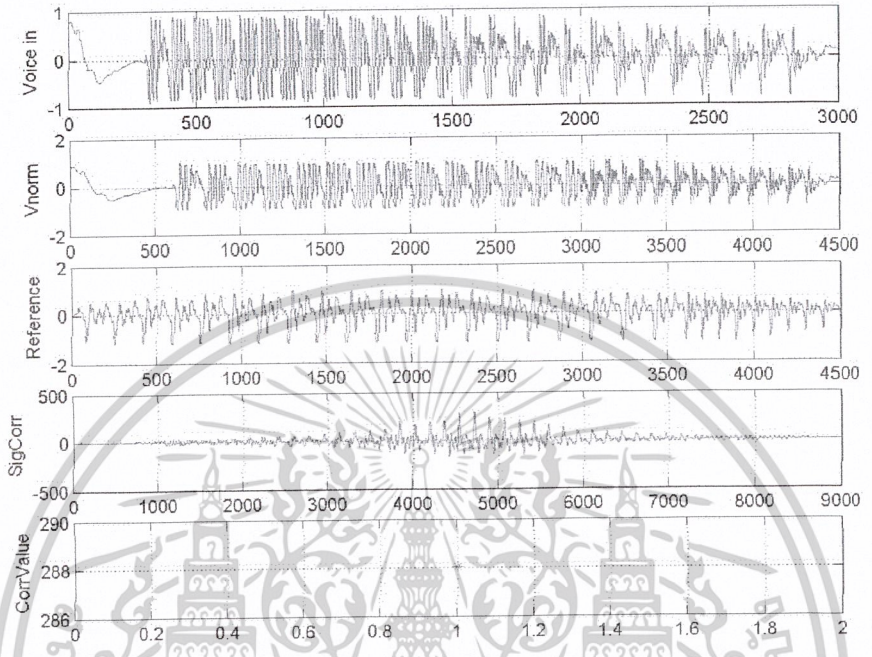
ค่าสหสัมพันธ์ของเสียงพูด “เจ็ด” กับเสียงอ้างอิงชุดที่ 1

315.38 161.93 173.1 259.15 92.631 348.46 534.85 84.356 85.383 669.51

ค่าสหสัมพันธ์ที่มากที่สุดคือ 669.51 ซึ่งตรงกับเสียง 0 ดังนั้นคำตอบ ผิด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เสียงพูด “แปด”




รูปที่ 4.8 กราฟแสดงการหาสหสัมพันธ์ ระหว่างสัญญาณเสียงทดสอบ “แปด” กับสัญญาณเสียงอ้างอิง ชุดที่ 2

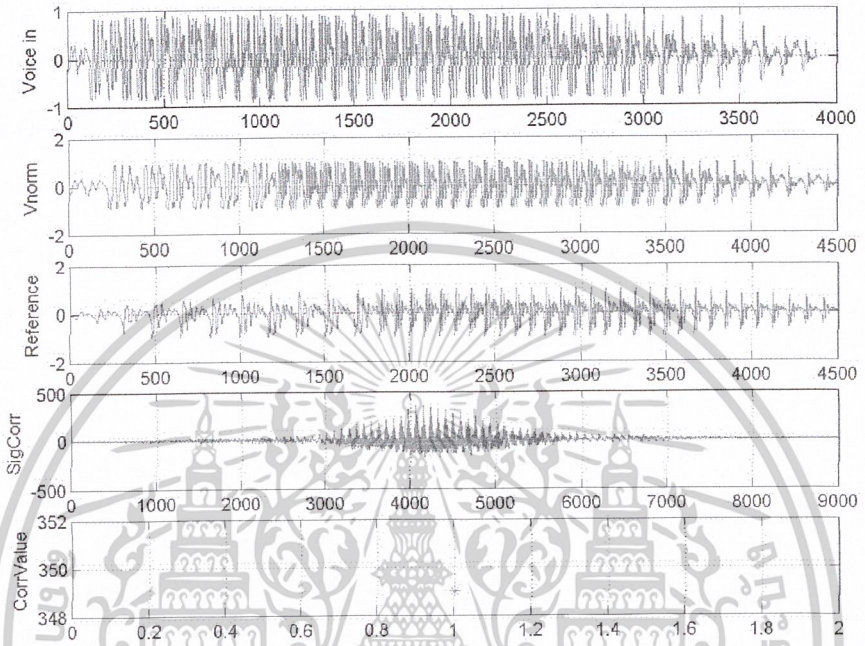
ค่าสหสัมพันธ์ของเสียงพูด “แปด” กับเสียงอ้างอิงชุดที่ 2

288.02 131.39 109.25 159.83 80.502 178.52 127.06 192.69
178.88 194.84

ค่าสหสัมพันธ์ที่มากที่สุดคือ 288.02 ซึ่งตรงกับเสียง 1 ดังนั้นคำตอบ ผิด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เสียงพูด “เก้า”




รูปที่ 4.9 กราฟแสดงการหาสหสัมพันธ์ ระหว่างสัญญาณเสียงทดสอบ “เก้า” กับสัญญาณเสียงอ้างอิงชุดที่ 4

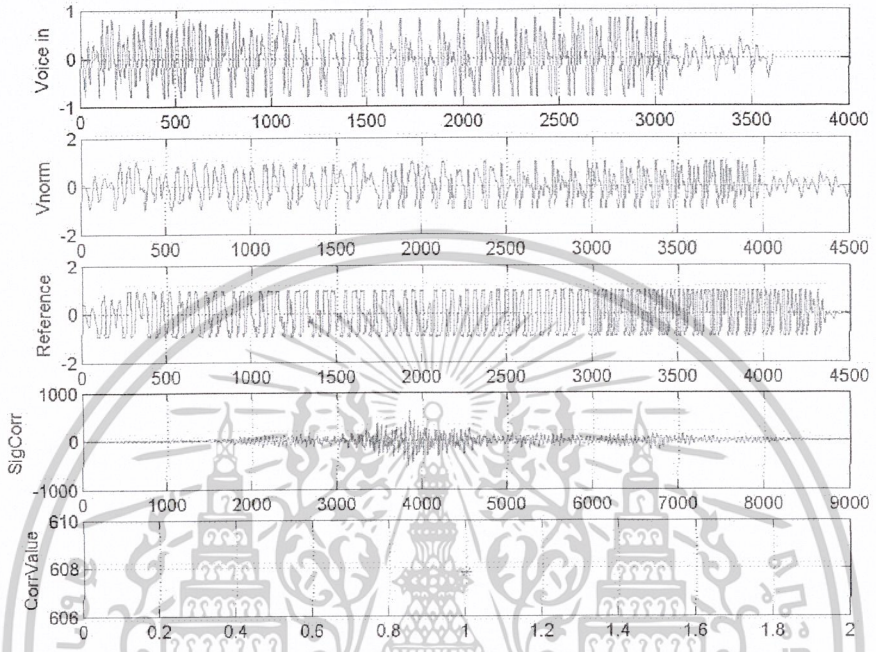
ค่าสหสัมพันธ์ของเสียงพูด “เก้า” กับเสียงอ้างอิงชุดที่ 4

167.69 96.026 97.025 103.1 236.62 143.83 132.99 113.39 349.02 124.45

ค่าสหสัมพันธ์ที่มากที่สุดคือ 349.02 ซึ่งตรงกับเสียง 9 ดังนั้นคำตอบถูก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เสียงพูด “ศูนย์”




รูปที่ 4.10 กราฟแสดงการหาสหสัมพันธ์ ระหว่างสัญญาณเสียงทดสอบ “เก้า” กับสัญญาณเสียงอ้างอิง ชุดที่ 2

ค่าสหสัมพันธ์ของเสียงพูด “ศูนย์” กับเสียงอ้างอิงชุดที่ 2

455.4 122.96 144.32 264.17 132.47 245.25 281.74 118.78 142.36 607.8

ค่าสหสัมพันธ์ที่มากที่สุดคือ 607.8 ซึ่งตรงกับเสียง 0 ดังนั้นคำตอบ ถูก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ต่อไปเป็นผลการทดลองการหาค่าสหสัมพันธ์ของเสียงพูดหนึ่งถึงเก้ากับเสียงอ้างอิงต่อไป จะแสดงตารางของเอาท์พุทที่ตรงกับของค่าสหสัมพันธ์ที่มากที่สุดของเสียง ตัวอย่าง “ศูนย์” ถึง “เก้า” ชุดที่ 1 ถึง ชุดที่ 20

| ชุดที่ / เสียงพูด | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| หนึ่ง | **1** | 0 | **1** | **1** | 6 | **1** | 6 | 6 | 6 | 6 |
| สอง | 6 | 0 | **2** | **2** | 4 | **2** | 4 | **2** | 9 | 6 |
| สาม | 4 | 9 | 6 | 4 | 4 | 9 | 9 | 9 | 0 | 5 |
| สี่ | 0 | 7 | **4** | **4** | 1 | 6 | **4** | **4** | 7 | 2 |
| ห้า | **5** | **5** | **5** | **5** | **5** | **5** | **5** | 7 | 7 | **5** |
| หก | 0 | **6** | 0 | **6** | **6** | **6** | **6** | 0 | **6** | 0 |
| เจ็ด | 0 | *6* | **7** | *6* | *6* | **7** | **7** | **7** | *6* | **7** |
| แปด | 1 | 0 | 6 | **8** | **8** | 6 | 7 | 6 | **8** | 6 |
| เก้า | ***9*** | *5* | 6 | ***9*** | ***9*** | ***9*** | ***9*** | ***9*** | ***9*** | ***9*** |
| ศูนย์ | **0** | **0** | **0** | 7 | 7 | **0** | **0** | 6 | 7 | 1 |

ตัวเข้มแสดงถึง เอาท์พุทที่ถูกต้อง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

| ชุดที่ / เสียงพูด | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|
| หนึ่ง | **1** | 0 | 0 | 6 | 0 | 6 | **1** | **1** | 6 | 0 |
| สอง | 6 | 1 | **2** | **2** | 0 | 4 | 1 | 1 | 4 | 7 |
| สาม | 0 | 4 | 9 | 9 | 0 | 0 | 4 | 5 | 4 | 4 |
| สี่ | **4** | 7 | **4** | **4** | 0 | 2 | **4** | **4** | 7 | 3 |
| ห้า | **5** | **5** | **5** | **5** | **5** | **5** | 6 | **5** | **5** | **5** |
| หก | **6** | **6** | **6** | **6** | **6** | **6** | **6** | **6** | 0 | 0 |
| เจ็ด | 0 | 6 | 6 | **7** | 1 | **7** | **7** | 0 | **7** | **7** |
| แปด | 6 | 9 | 7 | 7 | 7 | **8** | 6 | 0 | 0 | 1 |
| เก้า | 4 | **9** | 6 | 5 | 5 | 6 | 5 | 4 | 6 | **9** |
| ศูนย์ | 3 | 0 | 6 | 3 | 2 | 1 | 7 | 1 | 7 | 0 |

ตัวเข้มแสดงถึง เอาท์พุทที่ถูกต้อง

สรุปผลการทดลองทั้งหมดจากเสียงพูดศูนย์ถึงเก้า จำนวน 20 ชุด สามารถสรุปเป็นเปอร์เซ็นต์ความถูกต้องได้ดังนี้

| | | | |
|---|---|---|---|
| หนึ่ง | ถูก 35% | หก | ถูก 70% |
| สอง | ถูก 30% | เจ็ด | ถูก 50% |
| สาม | ถูก 0% | แปด | ถูก 20% |
| สี่ | ถูก 45% | เก้า | ถูก 50% |
| ห้า | ถูก 85% | ศูนย์ | ถูก 35% |

เปอร์เซ็นต์ความถูกต้องรวม 42%

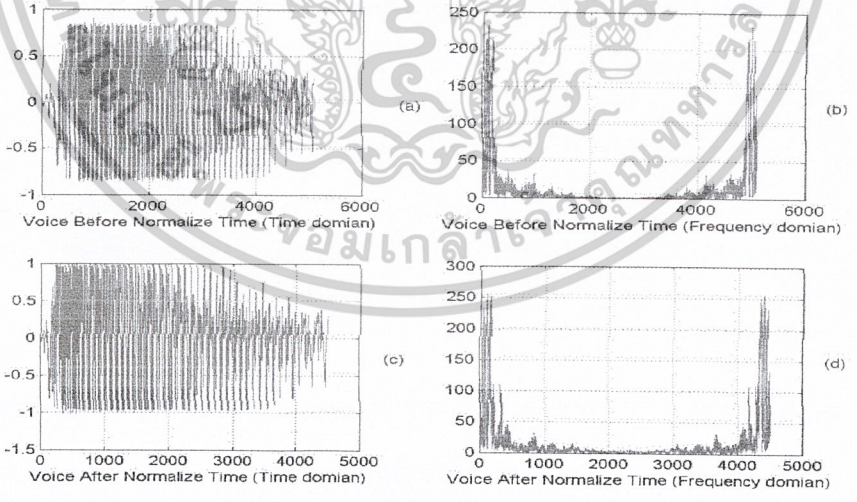
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# บทที่ 5

## บทสรุปข้อเสนอแนะ

โครงงานนี้เป็นการศึกษาการรู้จำเสียงพูด(Speech Recognition) โดยใช้การหาค่าสหสัมพันธ์ในการวัดความเหมือนระหว่างเสียงอ้างอิงกับเสียงทดสอบ ซึ่งผลที่ได้จากการวัดความเหมือนด้วยวิธีการการหาค่าสหสัมพันธ์ คำตอบส่วนใหญ่ที่ได้มักจะเป็นคำตอบที่ผิด สาเหตุที่ทำให้คำตอบที่ได้มาผิดพอสรุปได้ดังนี้

1.เกิดจากการปรับขนาดความยาวของเสียงทดสอบให้มีขนาดความยาวเท่ากับความยาวของเสียงอ้างอิง โดยที่ถ้าเสียงทดสอบมีความยาวของเสียงยาวเกินความยาวมาตรฐานก็จะทำการชักตัวอย่างของเสียงออกจนมีความยาวของเสียงมีความยาวเท่าความยาวมาตรฐาน แต่ถ้าเสียงทดสอบมีความยาวของเสียงสั้นกว่าความยาวมาตรฐานก็จะทำการเพิ่มตัวอย่างเข้าไปจนมีความยาวของเสียงมีความยาวเท่าความยาวมาตรฐาน ซึ่งทำเกิดความผิดพลาดในการหาค่าสหสัมพันธ์ระหว่างเสียงทดสอบกับเสียงอ้างอิงเนื่องจากการเพิ่มตัวอย่างหรือการชักตัวอย่างออกจะทำให้ความถี่ของสัญญาณเสียงเปลี่ยนไป




รูปที่ 5.1 กราฟแสดงสัญญาณเสียง "สี่" ทำ การปรับขนาดความยาวของเสียงแล้วคำตอบออกมาผิด

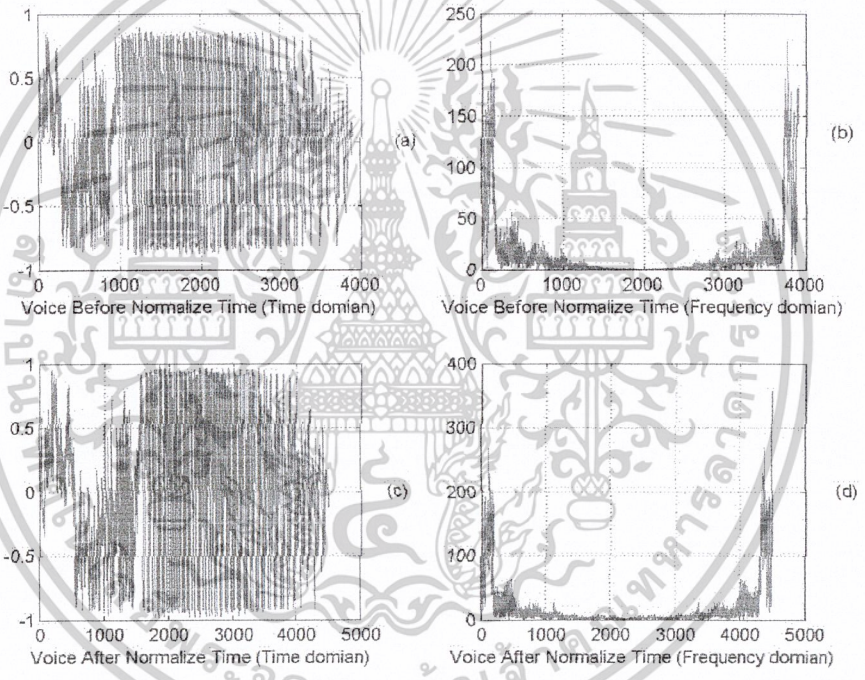
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

(a) กราฟแสดงสัญญาณเสียงอินพุตเสียง “สี่” ในโดเมนเวลา ที่มีความยาวของเสียงยาวเกินความยาวมาตรฐาน

(b) กราฟแสดงสัญญาณเสียงอินพุตเสียง “สี่” ในโดเมนความถี่ ที่มีความยาวของเสียงยาวเกินความยาวมาตรฐาน

(c) กราฟแสดงสัญญาณเสียงอินพุตเสียง “สี่” ในโดเมนเวลา ที่ปรับให้มีความยาวของเสียงเท่ากับความยาวมาตรฐาน

(d) กราฟแสดงสัญญาณเสียงอินพุตเสียง “สี่” ในโดเมนความถี่ ที่ปรับให้มีความยาวของเสียงเท่ากับความยาวมาตรฐาน




รูปที่ 5.2 กราฟแสดงสัญญาณเสียง ”สี่” ทำการปรับขนาดความยาวของเสียงแล้วคำตอบออกมาถูกต้อง

(a) กราฟแสดงสัญญาณเสียงอินพุตเสียง “สี่” ในโดเมนเวลา ที่มีความยาวของเสียงยาวสั้นกว่าความยาวมาตรฐาน

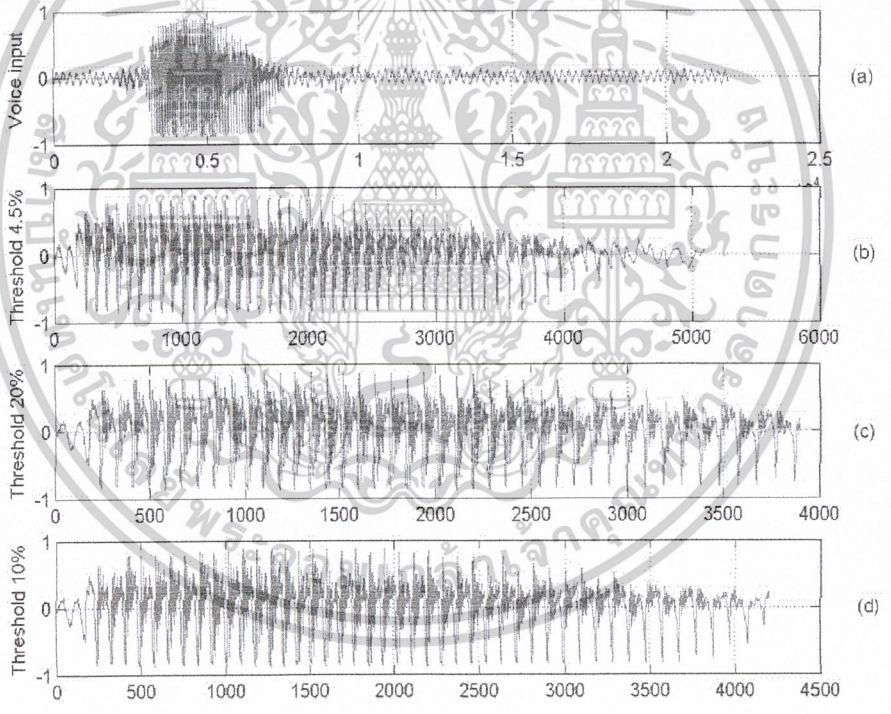
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

(b) กราฟแสดงสัญญาณเสียงอินพุตเสียง “สี่” ในโดเมนความถี่ ที่มีความยาวของเสียงยาวสั้นกว่าความยาวมาตรฐาน

(c) กราฟแสดงสัญญาณเสียงอินพุตเสียง “สี่” ในโดเมนเวลา ที่ปรับให้มีความยาวของเสียงเท่ากับความยาวมาตรฐาน

(d) กราฟแสดงสัญญาณเสียงอินพุตเสียง “สี่” ในโดเมนความถี่ ที่ปรับให้มีความยาวของเสียงเท่ากับความยาวมาตรฐาน

2. เกิดจากวิธีการหาจุดเริ่มต้น และจุดสิ้นสุดของเสียงยังไม่ดีพอ ทำให้ข้อมูลเสียงบางส่วนถูกตัดทิ้งไป ทำให้เมื่อมาวัดความเหมือนกับเสียงอ้างอิงแล้วเกิดความผิดพลาด




รูปที่ 5.3 กราฟแสดงการหาจุดเริ่มและจุดท้ายของสัญญาณเสียง ”ห้า”

(a) กราฟแสดงสัญญาณเสียงอินพุตเสียง “ ห้า”

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

(b) กราฟแสดงการหาจุดเริ่มและจุดท้ายของสัญญาณเสียง "ห้า" ที่ใช้ค่าระดับของพลังงานขีดเปลี่ยน(Threshold) 4.5% ของเฟรมที่ค่าพลังงานสูงสุด และเมื่อนำไปหาค่าสหสัมพันธ์กับสัญญาณเสียงอ้างอิงจะได้คำตอบที่ผิด

(c) กราฟแสดงการหาจุดเริ่มและจุดท้ายของสัญญาณเสียง "ห้า" ที่ใช้ค่าระดับของพลังงานขีดเปลี่ยน(Threshold) 20% ของเฟรมที่ค่าพลังงานสูงสุด และเมื่อนำไปหาค่าสหสัมพันธ์กับสัญญาณเสียงอ้างอิงจะได้คำตอบที่ผิด

(d) กราฟแสดงการหาจุดเริ่มและจุดท้ายของสัญญาณเสียง"ห้า"ที่ใช้ค่าระดับของพลังงานขีดเปลี่ยน(Threshold) 10% ของเฟรมที่ค่าพลังงานสูงสุด และเมื่อนำไปหาค่าสหสัมพันธ์กับสัญญาณเสียงอ้างอิงจะได้คำตอบที่ถูก

3. เกิดจากขั้นตอนการนำข้อมูลเสียงอ้างอิงและข้อมูลเสียงทดสอบมาวัดความเหมือน โดยการนำข้อมูลมาวัดความเหมือนนั้นจะนำข้อมูลทั้งหมดมาวัดความเหมือน ซึ่งจะทำให้เกิดความผิดพลาด เนื่องจากข้อมูลเสียงที่เป็นคนละเสียงกันอาจมีข้อมูลบางตัวที่เหมือนกัน ทำให้ค่าที่ได้จากการคำนวณสหสัมพันธ์มีค่ามาก เช่น สัญญาณเสียงทดสอบเสียงหกอาจมีข้อมูลบางตัวของเสียงทดสอบไปเหมือนกับสัญญาณเสียงอ้างอิงเสียงหนึ่ง จึงทำให้คำตอบที่ได้ออกมาผิด

4.เกิดจากสัญญาณรบกวนทำให้ข้อมูลเสียงมีความผิดเพี้ยน เมื่อนำมาวัดความเหมือนกับข้อมูลเสียงอ้างอิงจะเกิดความผิดพลาด

แนวทางพัฒนาต่อสามารถเลือกอัลกอลิทึมอื่นๆที่สามารถดึงลักษณะเด่นของเสียงออกมา เช่น ใช้การแปลงเวฟเล็ต,การแปลงคาร์ฮูเนน-เลิว์ฟ และอื่นๆ เพื่อหาค่าสัมประสิทธ์ของสัญญาณเสียง เพื่อเข้ากระบวนการตัดสินใจต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# บรรณานุกรม


## หนังสืออ้างอิง

[1] รศ.ดร. มนัส สังวรศิลป์,วรรัตน์ ภัทรอมรกุล, "คู่มือการใช้งาน MATLAB ฉบับสมบูรณ์", Infor press,2543

[2] อรรถสิทธิ์ หล่าสกุล,"Digital Signal Processing " (Fundamental), สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

[3] *Alasdair Mcandrew ,"Introduction To Image Processing With MATLAB"* ,Thomson Learning,2004

[4] Jae S Lim."Speech Enhancement" ,Printal-Hall ,1983

[5] John R. Deller ,John H.L Hansen ,John G. Proakis ,"Diiscrete-Time Processing of Speech Signal", IEEE Press,2000

[6] Robert D. Strum ,Donald E. Kirk, "First Principle of Discrete System and Digital Signal Processing" ,Addison Wesley,1989

## เอกสารอ้างอิง

[7] ทัศเวศ วีรวัฒน์ , การรู้จำเสียงคำไทยเฉพาะบุคคล, วิทยานิพนธ์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง,2541

[8] "เอกสารประกอบการเรียนวิชา "Information Laboratory", ภาควิชาวิศวกรรมสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

[9] "เอกสารประกอบการเรียนวิชา "Selected Topics in Information Engineering", ภาควิชาวิศวกรรมสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

## เว็ปไซต์อ้างอิง

[9] http:/www.thaidev.com

[10] http:/www.nectec.or.th

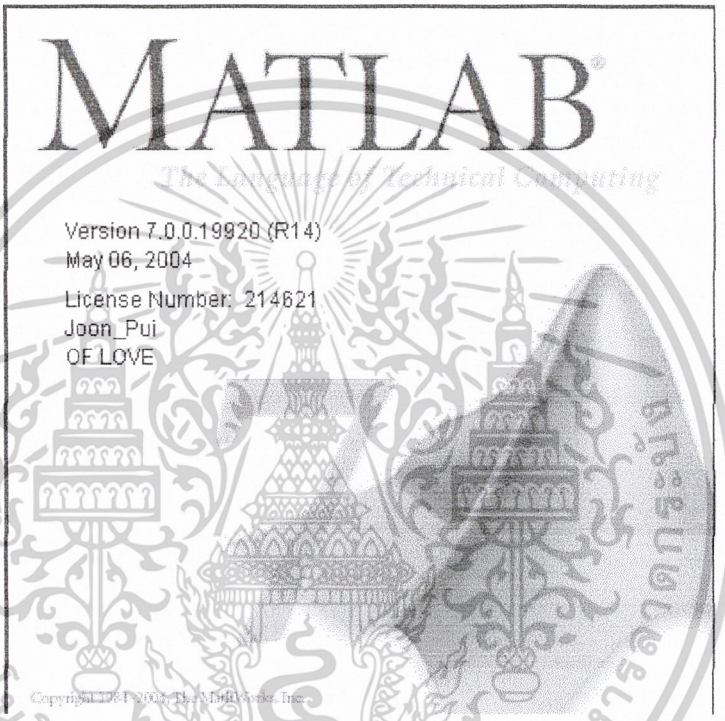

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

[11] http:/www.shool.net.th

## ภาคผนวก

## การเรียกใช้งานโปรแกรม

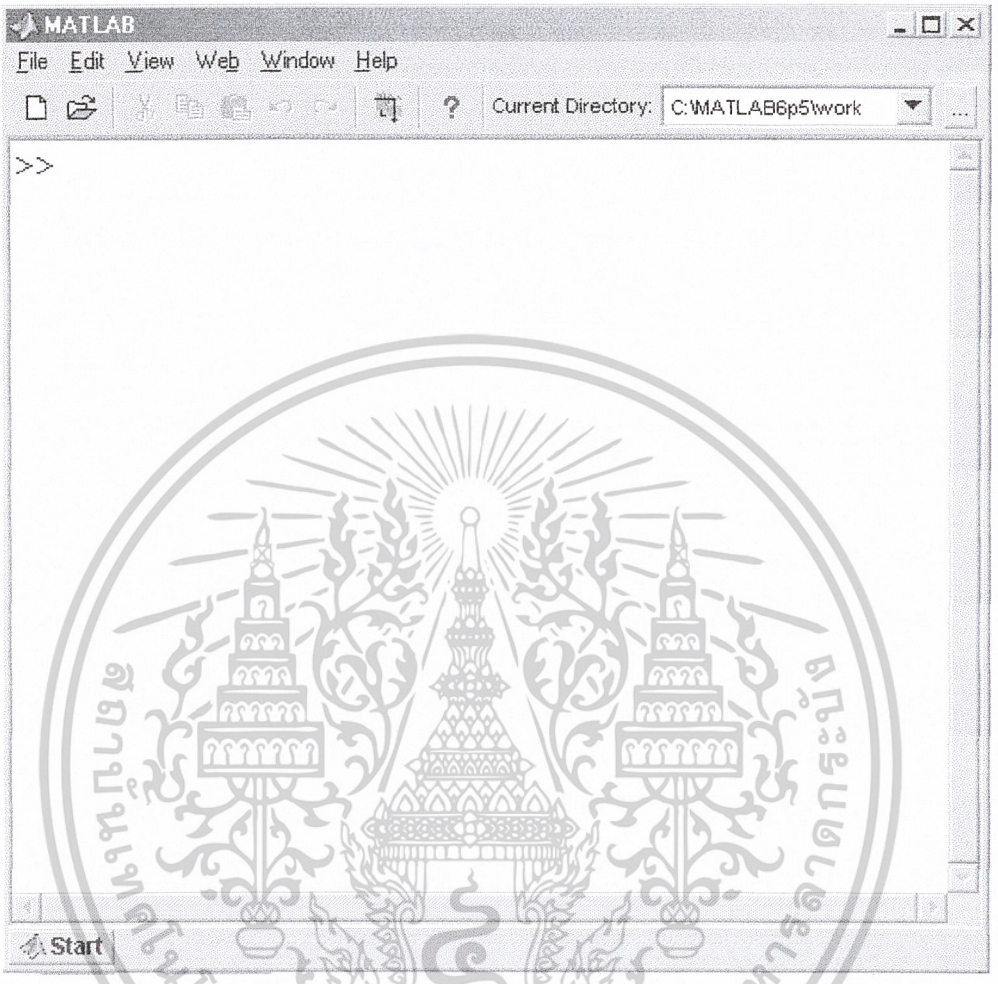
1. เปิดโปรแกรม Matlab ดังรูปที่ 1




รูปที่ 1 โปรแกรม Matlab

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

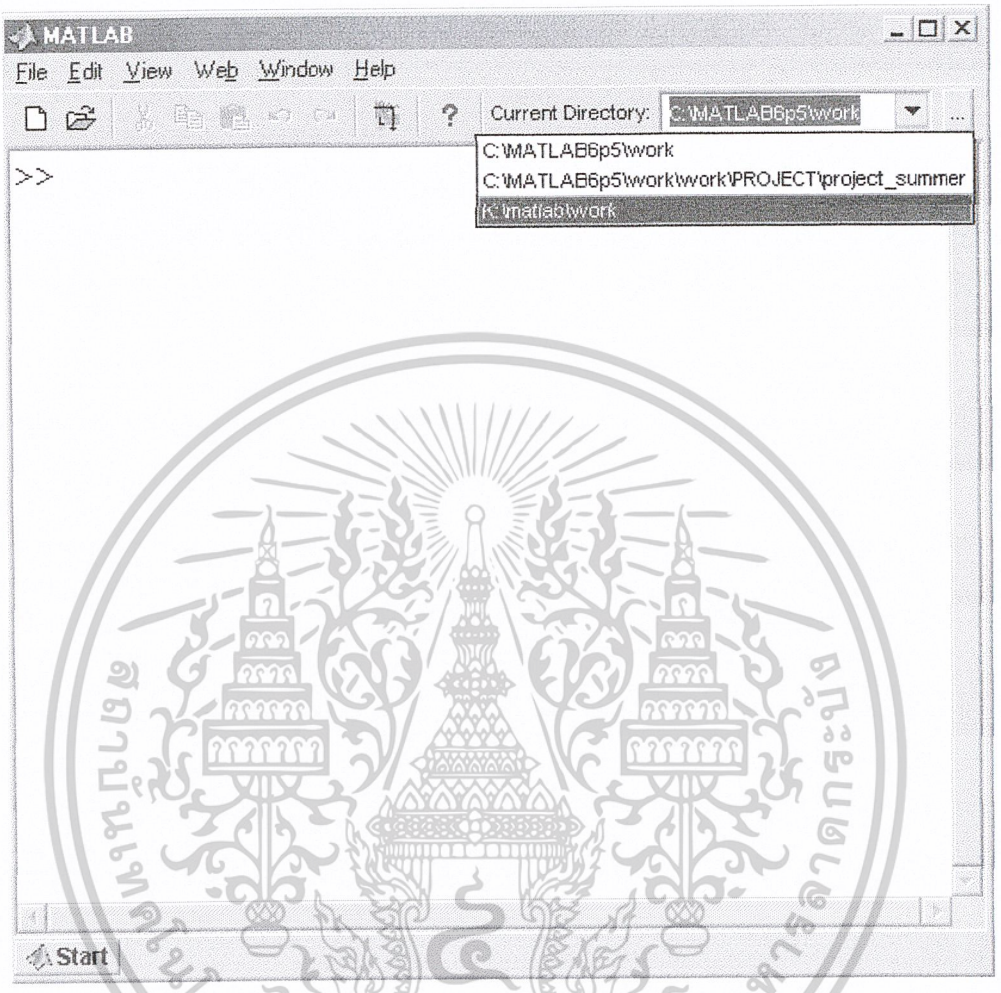
2. จากนั้นโปรแกรม Matlab เมื่อพร้อมทำงานจะแสดงดังรูปที่ 2




รูปที่ 2 แสดงหน้าจอCommandของโปรแกรมMatlab

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

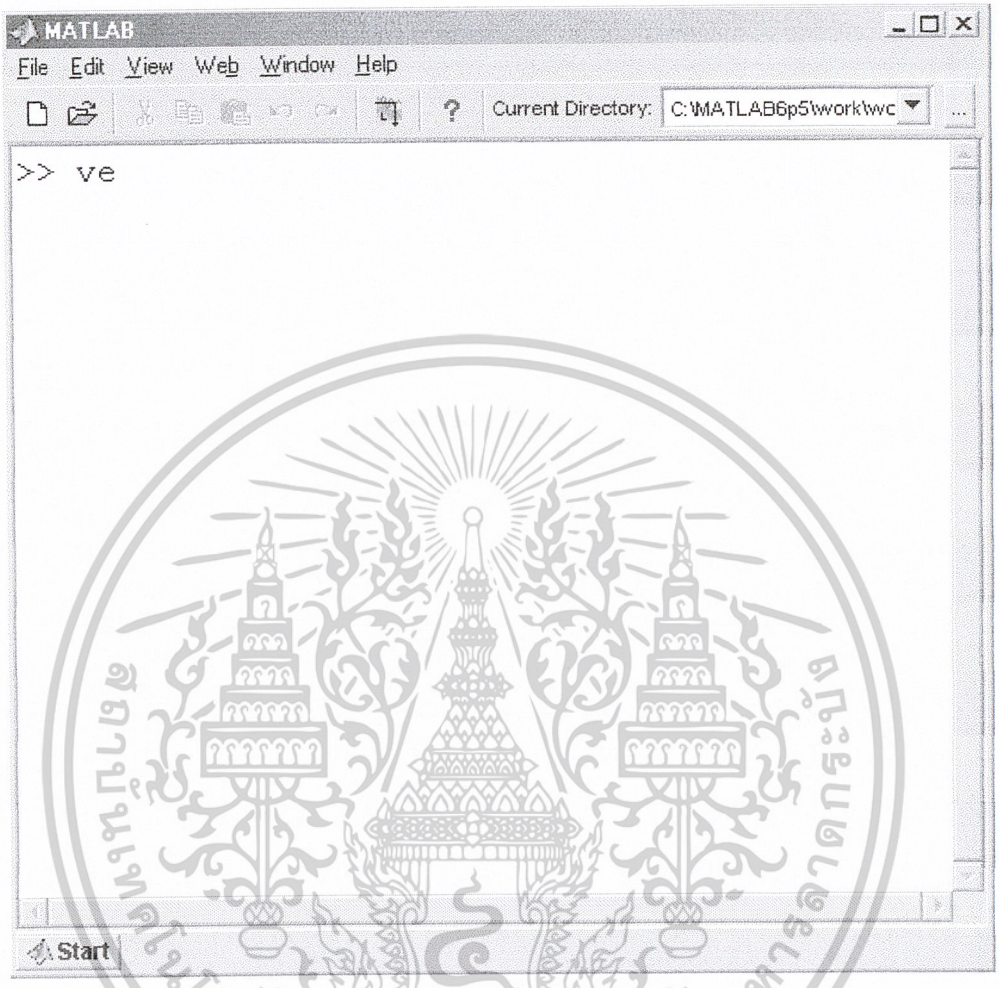
3. เปลี่ยน Current Directory ไปยัง Directory ที่ได้ทำการ Extract ไฟล์ ve.rar ไว้




รูปที่ 3 แสดงการเปลี่ยน Current Directory

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

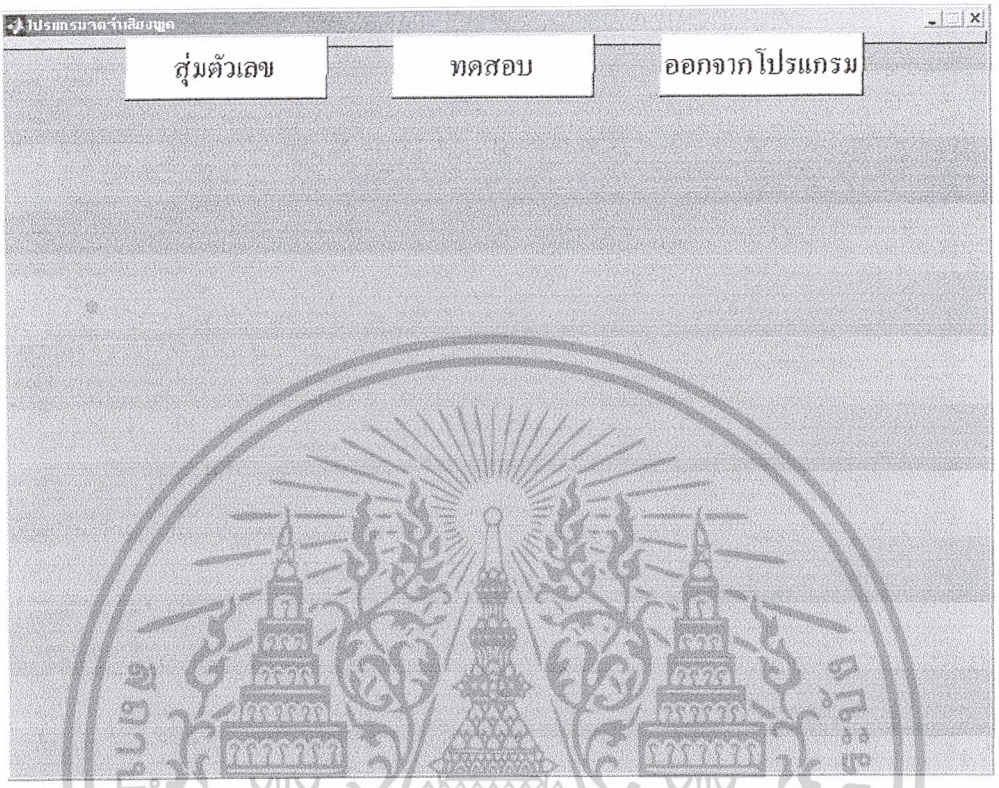
4. ทำการรันโปรแกรมโดย พิมพ์ชื่อโปรแกรมที่ Command Promp ของ Matlab




รูปที่ 4 แสดงหน้าจอการรันโปรแกรมจาก Command Promp

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

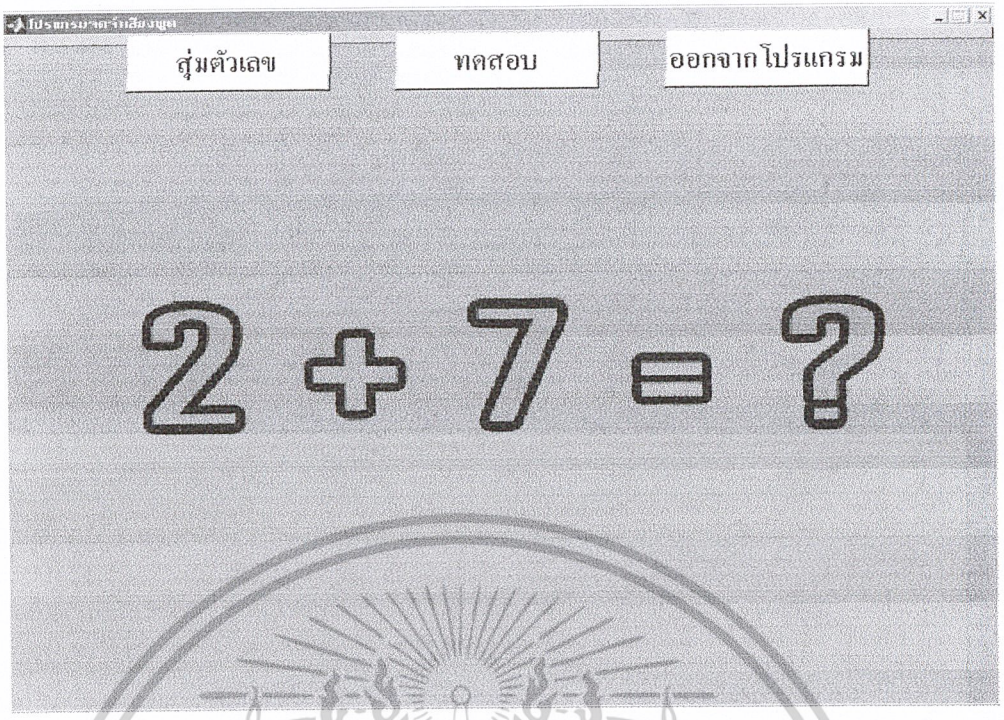
5. โปรแกรมศึกษาการรู้จำเสียงพูดจะแสดงดังรูปที่ 5




รูปที่ 5 แสดงหน้าจอโปรแกรมศึกษาการรู้จำเสียงพูด

6. คลิกที่ปุ่ม " สุ่มตัวเลข " จากนั้นโปรแกรมจะทำการสุ่มแสดงภาพกราฟฟิกตัวเลขขึ้นมาซึ่งแสดงดังรูปที่ 6 จากนั้นให้ผู้พูดคลิกที่ปุ่ม" ทดสอบ " เพื่อทำการป้อนเสียงทดสอบจากนั้นทำตามที่โปรแกรมได้ส่งเสียงบอก ถ้าผู้เล่นออกเสียงเป็นตัวเลขที่ภาพกราฟฟิกได้แสดงไว้ได้ถูกต้อง โปรแกรมจะส่งเสียงปรบมือให้และแสดงหน้าจอ"Congratulation" ซึ่งแสดงดังรูปที่ 7 แต่ถ้าออกเสียงไม่ถูกต้องโปรแกรมจะส่งเสียงบอกว่า"ให้ลองพยายามใหม่" และแสดงหน้าจอ"Oop..It wrong" ซึ่งแสดงดังรูปที่ 8

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้




รูปที่ 6 แสดงหน้าจอเมื่อทำการคลิกปุ่ม “สุ่มตัวเลข”

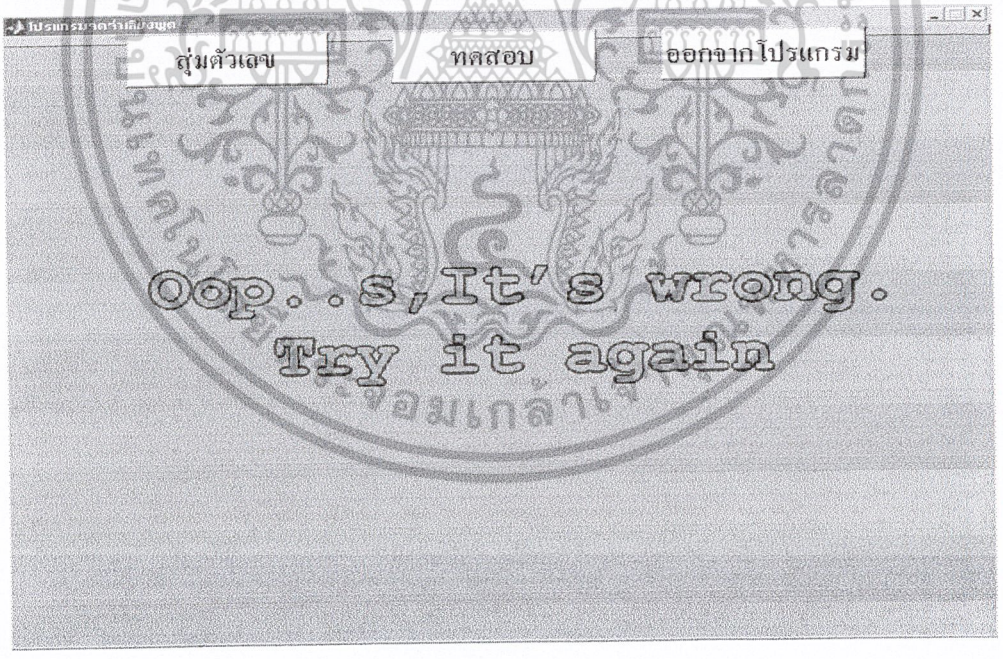
จากหน้าจอในรูปที่ 6 เสียงที่ต้องพูดก็คือคำว่า “เก้า” เมื่อผู้พูดต้องการทดสอบต่อก็ทำการคลิกที่ปุ่ม “สุ่มตัวเลข” ต่อเป็นแบบนี้ไปเรื่อยๆ เมื่อต้องการออกจากโปรแกรมก็ทำการคลิกที่ปุ่ม “ ออกจากโปรแกรม”

*** หมายเหตุ

รูปภาพกราฟฟิกที่แสดงแต่ละครั้งอาจจะมีการแสดงภาพเดิมซ้ำกันได้เนื่องจากโปรแกรมจะใช้การสุ่มเลือกภาพมาแสดง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ 7 แสดงหน้าจอเมื่อผู้พูดออกเสียงได้อย่างถูกต้อง




รูปที่ 8 แสดงหน้าจอเมื่อผู้พูดออกเสียงไม่ถูกต้อง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้