สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

การรู้จำเสียง
VOICE RECOGNITION

โดย

นางสาวนภัส ศรียงค์
นางสาวนิภารัตน์ ธนประโยชน์ศักดิ์

เลขหมู่......................................
เลขทะเบียน.....72655.....
วัน,เดือน,ปี...21 มิ.ย. 2550...

.b..11220922..............
.i...............................

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต
ภาควิชาวิศวกรรมสารสนเทศ
คณะวิศวกรรมศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ปีการศึกษา 2549

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

VOICE RECOGNITION

BY

MS.NAPAS SRIYONG

MS.NIPARAT THANAPRAYOTSAK

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENT FOR THE DEGREE OF
BACHELOR IN DEPARTMENT OF INFORMATION ENGINEERING
FACULTY OF ENGINEERING
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

2006

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อปริญญานิพนธ์	การรู้จำเสียง		
นักศึกษา	นางสาวนภัส ศรียงค์	รหัสนักศึกษา	46010336
	นางสาวนิภารัตน์ ธนประโยชน์ศักดิ์	รหัสนักศึกษา	46010375
อาจารย์ที่ปรึกษา	ดร.พิทักษ์ ธรรมวาริน		
ระดับการศึกษา	ปริญญาตรี วิศวกรรมศาสตรบัณฑิต สาขาวิศวกรรมสารสนเทศ		
ภาควิชา	วิศวกรรมสารสนเทศ		
ปีการศึกษา	2549		

ปริญญานิพนธ์ฉบับนี้ได้รับความเห็นชอบจากอาจารย์ที่ปรึกษาเป็นที่เรียบร้อยแล้ว

(ดร.พิทักษ์ ธรรมวาริน)

อาจารย์ผู้ควบคุมปริญญานิพนธ์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อปริญญานิพนธ์	การรู้จำเสียง		
นักศึกษา	นางสาวนภัส ศรียงค์	รหัสนักศึกษา	46010336
	นางสาวนิภารัตน์ ธนประโยชน์ศักดิ์	รหัสนักศึกษา	46010375
อาจารย์ที่ปรึกษา	ดร.พิทักษ์ ธรรมวาริน		
ระดับการศึกษา	ปริญญาตรี วิศวกรรมศาสตรบัณฑิต สาขาวิศวกรรมสารสนเทศ		
ภาควิชา	วิศวกรรมสารสนเทศ		
ปีการศึกษา	2549		

บทคัดย่อ

ปริญญานิพนธ์ฉบับบนี้เป็นการนำเสนอวิธีการรู้จำเสียง โดยพิจารณาลักษณะเด่นซึ่งเป็นคุณสมบัติเฉพาะตัวของเสียงสามารถทดสอบได้โดยคุณสมบัติสองชนิด คือ รูปคลื่นของสัญญาณเสียง และอัตราการเปลี่ยนแปลงของคลื่นสัญญาณเสียงตามเวลา นอกจากนั้นยังใช้การประมาณค่าฟูเรียร์มาช่วยในการลดความเบี่ยงเบนของเสียง ซึ่งการรู้จำเสียงกระทำได้โดยคำนวณระยะทางแบบยูคลิด (Euclidean Distance) ระหว่างค่าสัมบูรณ์ของสัมประสิทธิ์ฟูเรียร์ที่คำนวณได้จากเสียงอ้างอิงและเสียงที่นำมาทดสอบ ประสิทธิภาพของวิธีที่นำเสนอแสดงได้โดยการทดลองรู้จำเสียงซึ่งประกอบด้วย 315 เสียง ได้ค่าเปอร์เซ็นต์ความถูกต้อง 91.2 %

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Thesis Title	Voice Recognition
Student	Ms. Napas Sriyong ID 46010336
	Ms. Niparat Thanaprayotsak ID 46010375
Advisor	Dr.Pitak Thumwarin
Graduate Level	Bachelor Degree of Information Engineering
Department	Information Engineering
Academic Year	2006

Abstract

This project propose a voice recognition method. In the proposed method, it is considered that the individual features of voice can be represented by two features; the waveform of voice signal and the rate of the signal change. Moreover, the Fourier approximations of the two features are used to extract the individual feature of voice. Finally, voice can be recognized by calculating Euclidean distance of the magnitude of Fourier coefficients of the two above features obtained from the reference voice and that obtained from the voice to be recognized. Voice recognition experiments were performed on a database of 315 voice from three speakers. An average of the recognition rates is 91.2%

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

ปริญญานิพนธ์ฉบับนี้สำเร็จไปได้ด้วยดี ด้วยความร่วมมือกันจากหลายฝ่าย ขอขอบพระคุณอาจารย์ที่ปรึกษา ดร. พิทักษ์ ธรรมวารินเป็นอย่างสูงที่ได้ให้คำปรึกษา คำแนะนำ ตรวจทานแก้ไขและให้ความดูแลเอาใจใส่ตลอดระยะเวลาที่ทำปริญญานิพนธ์ ขอบพระคุณพ่อ คุณแม่ และอาจารย์ทุกท่านที่ให้การสนับสนุนเสมอมารวมทั้งเพื่อนๆ ทุกคนที่ให้คำแนะนำและสละเวลามาช่วยทำการเก็บข้อมูลเสียงในการทำปริญญานิพนธ์นี้

นภัส ศรียงค์

นิภารัตน์ ธนประโยชน์ศักดิ์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

เรื่อง	หน้า
ปกในภาษาไทย	-
ปกในภาษาอังกฤษ	-
บทคัดย่อภาษาไทย	ก
บทคัดย่อภาษาอังกฤษ	ข
กิตติกรรมประกาศ	ค
สารบัญ	ง-จ
สารบัญรูปภาพ	ฉ-ฌ
สารบัญตาราง	ญ
บทที่ 1 บทนำ	1
1.1 ความเป็นมาและความสำคัญของปัญหา	1
1.2 ความมุ่งหมายและวัตถุประสงค์	1
1.3 ขอบเขตของโครงงาน	1
1.4 ขั้นตอนการทำงานของระบบ	2
1.5 ขั้นตอนการดำเนินงาน	2
บทที่ 2 หลักการและทฤษฎีที่เกี่ยวข้อง	3
2.1 การรู้จำเสียง (Voice Recognition)	3
2.2 เสียง (Sound)	3
2.2.1 ธรรมชาติและความดันของเสียง	4
2.2.2 คุณสมบัติของเสียง	4
2.2.2.1 ระดับเสียง (Pitch)	4
2.2.2.2 ความดัง – ค่อย (Dynamics)	5
2.2.3 การกำเนิดเสียง	5
2.3 การแปลงเสียงเป็นสัญญาณไฟฟ้า	6
2.4 คอมพิวเตอร์กับการประมวลผลสัญญาณเสียง	6
2.5 รูปแบบของข้อมูลเสียงที่เป็นพีซีเอ็ม (WAVE PCM sound file format)	7
2.6 การเปรียบเทียบโดยโปรแกรมแบบพลวัต (Dynamic Programming Matching)	9
2.7 การประมาณค่าด้วยเส้นตรง	10
2.8 อนุกรมฟูเรียร์ (Fourier Series)	11
2.9 การวัดระยะแบบยูคลิด (Euclidean distance)	12

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เรื่อง หน้า

2.10 การหาอัตราการเปลี่ยนแปลงของรูปร่างสัญญาณเสียงต่อเวลา 14

บทที่ 3 การออกแบบ 15

3.1 โครงสร้างของโปรแกรม 15

3.2 การเตรียมเสียงก่อนนำไปประมวลผล (Preprocessing) 16

3.2.1 การตัดเสียงเงียบ 16

3.2.2 การหาคาบของเสียง 18

3.2.3 การชักตัวอย่างใหม่ด้วยฟังก์ชันพีซไวส์ลิเนียร์ 19

3.3 กระบวนการดึงลักษณะเด่น (Feature Extraction) 21

3.3.1 การหาลักษณะเด่นของเสียงโดยใช้อนุกรมฟูเรียร์ 21

3.3.2 การใช้ค่าสัมบูรณ์ของสัมประสิทธิ์ฟูเรียร์เป็นลักษณะเด่นของเสียง 24

3.4 การรู้จำเสียง (Voice Recognition) 26

บทที่ 4 ผลการทดลอง

4.1 การเก็บข้อมูลเสียง 27

4.2 กระบวนการเตรียมเสียงก่อนนำไปทดสอบ 27

4.2.1 การตัดเสียงเงียบ 27

4.2.2 การหาคาบของเสียง 30

4.2.3 การชักตัวอย่างใหม่ด้วยฟังก์ชันพีซไวส์ลิเนียร์ 33

4.3 การดึงคุณลักษณะเด่นของเสียงโดยใช้ค่าสัมบูรณ์ของสัมประสิทธิ์ฟูเรียร์ 34

4.4 การรู้จำเสียง 37

4.4.1 ใช้ค่าสัมบูรณ์ของสัมประสิทธิ์ฟูเรียร์ที่ได้จากรูปคลื่นสัญญาณเสียงเป็นลักษณะเด่น 37

4.4.2 ใช้ค่าสัมบูรณ์ของสัมประสิทธิ์ฟูเรียร์ที่ได้จากอัตราการเปลี่ยนแปลงของคลื่นสัญญาณเสียงเป็นลักษณะเด่น 43

4.2.3 การใช้ค่าสัมบูรณ์ของสัมประสิทธิ์ฟูเรียร์ที่ได้จากรูปคลื่นสัญญาณเสียงและอัตราการเปลี่ยนแปลงของคลื่นสัญญาณเสียงเป็นลักษณะเด่น 46

บทที่ 5 สรุปผลการทดลอง

5.1 สรุปผลการทดลอง 56

5.2 ปัญหาที่เกิดขึ้นระหว่างการทดลอง 56

5.3 แนวทางในการพัฒนาโครงงาน 56

บรรณานุกรม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูปภาพ

บทที่ 1	หน้า
รูปที่ 1.1 แสดงภาพรวมขั้นตอนการทำงานของระบบ	2
บทที่ 2	
รูปที่ 2.1 แสดงการกระจัดของอนุภาคของอากาศ	4
รูปที่ 2.2 แสดงการสั่นสะเทือนเสียงสูง	5
รูปที่ 2.3 แสดงการสั่นสะเทือนเสียงต่ำ	5
รูปที่ 2.4 แสดงโครงสร้างของไมโครโฟน	6
รูปที่ 2.5 แสดงแสดงรูปแบบของเวฟไฟล์	7
รูปที่ 2.6 แสดงเวกเตอร์ของสัญญาณอ้างอิง (u^{n}) และเวกเตอร์ของสัญญาณทดสอบ (u^{*})	10
รูปที่ 2.7 แสดงฟังก์ชันวาร์ปปิง $w(k)$	10
รูปที่ 2.8 แสดงฟังก์ชั่นที่ใช้ทดสอบการระยะแบบยูคลิด	13
รูปที่ 2.9 แสดงรูปร่างของสัญญาณเสียง	14
รูปที่ 2.10 แสดงอัตราการเปลี่ยนแปลงรูปร่างของสัญญาณ	14
บทที่ 3	
รูปที่ 3.1 แสดงโครงสร้างของโปรแกรมรูปที่ 3.1 แสดงโครงสร้างของโปรแกรม	15
รูปที่ 3.2 แสดงรูปร่างของสัญญาณเสียง	16
รูปที่ 3.3 แสดงภาพขยายของสัญญาณช่วงหนึ่ง	16
รูปที่ 3.4 แสดงรูปร่างของสัญญาณเสียง“อา” ของผู้พูดที่ต่างกัน	17
รูปที่ 3.5 แสดงรูปร่างของสัญญาณเสียง“อี” ของผู้พูดที่ต่างกัน	17
รูปที่ 3.6 แสดงรูปร่างของสัญญาณเสียง“โอ” ของผู้พูดที่ต่างกัน	18
รูปที่ 3.7 แสดงขั้นตอนการหาคาบของสัญญาณ	19
รูปที่ 3.8 แสดงเสียงที่มีคาบต่างกัน	20
รูปที่ 3.9 แสดงการเชื่อมต่อจุดด้วยวิธีการชักตัวอย่างใหม่	20
รูปที่ 3.10 แสดงภาพก่อนและหลังการชักตัวอย่างใหม่	21
รูปที่ 3.11 แสดงรูปร่างของสัญญาณ $f(t)$	22
รูปที่ 3.12 แสดงองค์ประกอบของสัญญาณเสียง	22
รูปที่ 3.13 แสดงรูปร่างของสัญญาณ $f(t)$	23

หน้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ 3.14 แสดงองค์ประกอบของสัญญาณเสียง 23
รูปที่ 3.15 แสดงรูปร่างของสัญญาณ $f(t)$ 23
รูปที่ 3.16 แสดงองค์ประกอบของสัญญาณเสียง 23
รูปที่ 3.17 แสดงการประมาณค่าของสัญญาณเสียงเปิด 24
รูปที่ 3.18 แสดงการประมาณค่าของสัญญาณเสียงเปิด 24
รูปที่ 3.19 แสดงสัญญาณเสียง 25
รูปที่ 3.20 แสดงค่าสัมบูรณ์ของสัมประสิทธิ์ฟูเรียร์ของสองช่วงเสียง 25
บทที่ 4
รูปที่ 4.1 แสดงตัวอย่างรูปร่างสัญญาณเสียงปิดของผู้พูดคนเดียวกัน 27
รูปที่ 4.2 แสดงตัวอย่างรูปร่างสัญญาณเสียงเปิดของผู้พูดคนเดียวกัน 28
รูปที่ 4.3 แสดงตัวอย่างรูปร่างสัญญาณเสียงซ้ายของผู้พูดคนเดียวกัน 28
รูปที่ 4.4 แสดงตัวอย่างรูปร่างสัญญาณเสียงขวาของผู้พูดคนเดียวกัน 29
รูปที่ 4.5 แสดงตัวอย่างรูปร่างสัญญาณเสียงหยุดของผู้พูดคนเดียวกัน 29
รูปที่ 4.6 แสดงตัวอย่างสัญญาณหนึ่งคาบของเสียงปิด 30
รูปที่ 4.7 แสดงตัวอย่างสัญญาณหนึ่งคาบของเสียงเปิด 30
รูปที่ 4.8 แสดงตัวอย่างสัญญาณหนึ่งคาบของเสียงซ้าย 31
รูปที่ 4.9 แสดงตัวอย่างสัญญาณหนึ่งคาบของเสียงขวา 31
รูปที่ 4.10 แสดงตัวอย่างสัญญาณหนึ่งคาบของเสียงหยุด 32
รูปที่ 4.11 แสดงเสียงอ้างอิงเสียง “เปิด” 32
รูปที่ 4.12 แสดงเสียงอ้างอิงเสียง “ซ้าย” 32
รูปที่ 4.13 แสดงเสียงอ้างอิงเสียง “ขวา” 33
รูปที่ 4.14 แสดงเสียงอ้างอิงเสียง “หยุด” 33
รูปที่ 4.15 แสดงสัญญาณเสียงเปิดก่อนการชักตัวอย่างใหม่ 33
รูปที่ 4.16 แสดงสัญญาณเสียงเปิดหลังการชักตัวอย่างใหม่ 33
รูปที่ 4.17 แสดงสัญญาณเสียงปิดก่อนการชักตัวอย่างใหม่ 34
รูปที่ 4.18 แสดงสัญญาณเสียงปิดหลังการชักตัวอย่างใหม่ 34
รูปที่ 4.19 แสดงตัวอย่างกราฟค่าสัมบูรณ์ของสัมประสิทธิ์ฟูเรียร์ของเสียงปิด 34
รูปที่ 4.20 แสดงตัวอย่างกราฟค่าสัมบูรณ์ของสัมประสิทธิ์ฟูเรียร์ของเสียงเปิด 34
รูปที่ 4.21 แสดงตัวอย่างกราฟค่าสัมบูรณ์ของสัมประสิทธิ์ฟูเรียร์ของเสียงซ้าย 35
รูปที่ 4.22 แสดงตัวอย่างกราฟค่าสัมบูรณ์ของสัมประสิทธิ์ฟูเรียร์ของเสียงขวา 36
รูปที่ 4.23 แสดงตัวอย่างกราฟค่าสัมบูรณ์ของสัมประสิทธิ์ฟูเรียร์ของเสียงหยุด 36
รูปที่ 4.24 แสดงระยะยุคลิดโดยเสียงอ้างอิงเป็นเสียงปิด 37
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หน้า

รูปที่ 4.25 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงเปิด 37
รูปที่ 4.26 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงซ้าย 37
รูปที่ 4.27 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงขวา 37
รูปที่ 4.28 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงหยุด 38
รูปที่ 4.29 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงปิด 38
รูปที่ 4.30 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงเปิด 38
รูปที่ 4.31 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงซ้าย 39
รูปที่ 4.32 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงขวา 39
รูปที่ 4.33 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงหยุด 39
รูปที่ 4.34 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงปิด 40
รูปที่ 4.35 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงเปิด 40
รูปที่ 4.36 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงซ้าย 40
รูปที่ 4.37 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงขวา 40
รูปที่ 4.38 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงหยุด 41
รูปที่ 4.39 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงปิด 41
รูปที่ 4.40 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงเปิด 41
รูปที่ 4.41 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงซ้าย 42
รูปที่ 4.42 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงขวา 42
รูปที่ 4.43 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงหยุด 42
รูปที่ 4.44 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงปิด 43
รูปที่ 4.45 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงเปิด 43
รูปที่ 4.46 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงซ้าย 43
รูปที่ 4.47 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงขวา 43
รูปที่ 4.48 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงหยุด 44
รูปที่ 4.49 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงปิด 44
รูปที่ 4.50 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงเปิด 44
รูปที่ 4.51 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงซ้าย 44
รูปที่ 4.52 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงขวา 44
รูปที่ 4.53 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงหยุด 45
รูปที่ 4.54 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงปิด 45
รูปที่ 4.55 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงเปิด 45

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หน้า

รูปที่ 4.56 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงซ้าย 45
รูปที่ 4.57 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงขวา 45
รูปที่ 4.58 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงหยุด 46
รูปที่ 4.59 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงปิด 46
รูปที่ 4.60 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงเปิด 46
รูปที่ 4.61 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงซ้าย 47
รูปที่ 4.62 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงขวา 47
รูปที่ 4.63 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงหยุด 47
รูปที่ 4.64 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงปิด 48
รูปที่ 4.65 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงเปิด 48
รูปที่ 4.66 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงซ้าย 48
รูปที่ 4.67 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงขวา 48
รูปที่ 4.68 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงหยุด 49
รูปที่ 4.69 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงปิด 49
รูปที่ 4.70 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงเปิด 49
รูปที่ 4.71 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงซ้าย 50
รูปที่ 4.72 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงขวา 50
รูปที่ 4.73 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงหยุด 50
รูปที่ 4.74 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงปิด 51
รูปที่ 4.75 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงเปิด 51
รูปที่ 4.76 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงซ้าย 51
รูปที่ 4.77 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงขวา 51
รูปที่ 4.78 แสดงระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงหยุด 52
รูปที่ 4.79 แสดงกราฟเปรียบเทียบอัตราการรู้จำของวิธีต่าง ๆ ของผู้พูดคนที่ 1 53
รูปที่ 4.80 แสดงกราฟเปรียบเทียบอัตราการรู้จำของวิธีต่าง ๆ ของผู้พูดคนที่ 2 54
รูปที่ 4.81 แสดงกราฟเปรียบเทียบอัตราการรู้จำของวิธีต่าง ๆ ของผู้พูดคนที่ 3 54
รูปที่ 4.82 แสดงกราฟเปรียบเทียบอัตราการรู้จำของวิธีต่าง ๆ ของผู้พูดคนที่ 1 55

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

บทที่ 4	หน้า

ตารางที่ 4.1 ตารางเปรียานเทียบอัตราการรู้จำของวิธีต่างๆ โดยใช้ผู้พูดคนเดียวกัน 54
ตารางที่ 4.2 ตารางเปรียบเทียบอัตราการรู้จำของวิธีต่างๆ โดยใช้ผู้พูดหลายคน 56

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

เนื่องจากปัจจุบันนี้ เทคโนโลยีได้พัฒนาก้าวหน้าไปอย่างรวดเร็ว ทำให้มีการผลิตและพัฒนาอุปกรณ์ต่าง ๆ ออกมามากมาย ซึ่งอุปกรณ์บางประเภทมีการใช้งานที่ยุ่งยากซับซ้อน และเนื่องจากการพูดจัดเป็นการสื่อสารที่เป็นธรรมชาติที่สุดอันหนึ่งของมนุษย์ จึงได้มีการนำเทคโนโลยีการรู้จำเสียงพูดมาประยุกต์ใช้ในการใช้งาน เพื่ออำนวยความสะดวกในการสื่อสารระหว่างมนุษย์กับอุปกรณ์ต่าง ๆ ได้

โครงงานชิ้นนี้จึงได้ถูกริเริ่มขึ้น เพื่อตอบสนองต่อการดำเนินชีวิตประจำวันที่ต้องการความรวดเร็ว ทำให้สามารถควบคุมหรือสั่งงานอุปกรณ์ต่างๆได้โดยสะดวกยิ่งขึ้น

1.2 ความมุ่งหมายและวัตถุประสงค์

1.2.1 เพื่อศึกษาวิธีการที่จะนำมาใช้ในการวิเคราะห์หาคุณลักษณะเด่นของเสียง

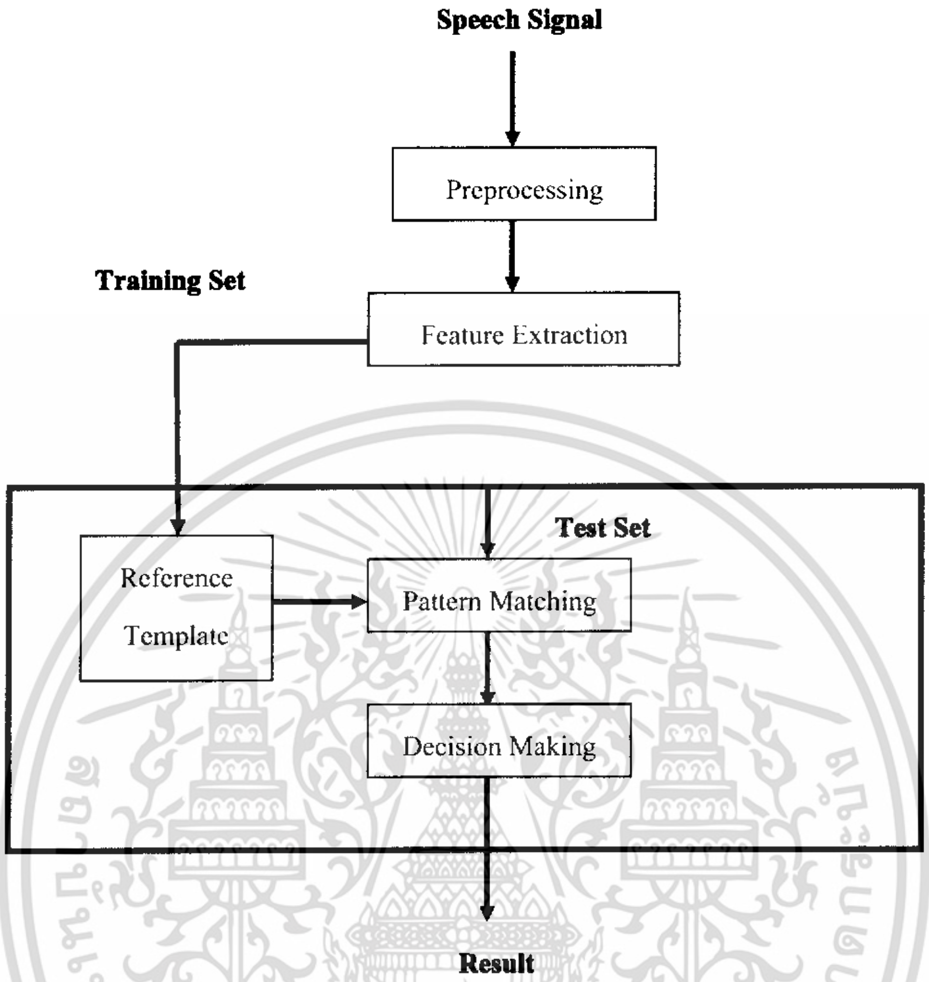
1.2.2 เพื่อศึกษาวิธีการรู้จำเสียงโดยใช้คุณลักษณะเด่นที่คำนวณได้มา

1.3 ขอบเขตโครงงาน

นำเสนอวิธีการตรวจสอบเสียงโดยใช้ลักษณะเด่นต่าง ๆ พร้อมทั้งหาประสิทธิภาพของวิธีการนั้น ๆ เพื่อหาวิธีที่ดีที่สุดในการรู้จำเสียง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.4 ขั้นตอนการทำงานของระบบ



รูปที่ **1.1** แสดงภาพรวมขั้นตอนการทำงานของระบบ

1.5 ขั้นตอนการดำเนินงาน

1.5.1 ศึกษางานวิจัยในการรู้จำเสียงที่ผ่านมา

1.5.2 ทำการเก็บรวบรวมข้อมูลเสียงต่าง ๆ โดยเก็บข้อมูลเสียงทั้งหมด 5 คำ คือ เปิด ปิด ขวา ซ้าย หยุด

1.5.3 ลดความเบี่ยงเบนต่าง ๆ (fluctuation)

1.5.4 ดึงลักษณะเด่นซึ่งเป็นลักษณะเฉพาะตัวของเสียง (feature extraction)

1.5.5 นำลักษณะเด่นที่ได้มาใช้ในการรู้จำเสียง

1.5.6 ทดลองและตรวจสอบประสิทธิภาพของวิธีการที่ได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2
หลักการและทฤษฎีที่เกี่ยวข้อง

2.1 การรู้จำเสียง (Voice Recognition)

การรู้จำเสียง เป็นเทคโนโลยีที่ใช้วิเคราะห์ และควบคุมการทำงานคอมพิวเตอร์ด้วยเสียง โดยอุปกรณ์ต่าง ๆ จะพยายามทำความเข้าใจในชุดข้อมูลเสียงที่ได้รับ ด้วยการตรวจสอบโครงสร้างและรูปแบบของเสียงว่าถูกต้องตามที่กำหนดไว้หรือไม่ และจะดำเนินการตามเงื่อนไขของโปรแกรมที่ได้กำหนดไว้

เทคโนโลยีการรู้จำเสียงสามารถแบ่งได้เป็น 4 ประเภทดังนี้

1. การรู้จำคำทุกคำในประโยค (Continuous Speech Recognition) เป็นเทคโนโลยีการรู้จำคำทุกๆคำในประโยคหรือวลี
2. การรู้จำผู้พูด (Speaker Recognition) เป็นเทคโนโลยีในการตรวจสอบว่าเสียงพูดนั้นเป็นเสียงของผู้พูดที่ถูกต้องหรือไม่ หรือใช้ระบุตัวผู้พูดที่อยู่ในฐานข้อมูล
3. การรู้จำคำ (Word Recognition) เป็นเทคโนโลยีการรู้จำและบ่งชี้คำต่างๆ ซึ่งจำนวนคำในกลุ่มคำศัพท์อาจมีขนาดเล็กหรือใหญ่ก็ได้
4. การรู้จำคำเป้าหมาย (Word Spotting) เป็นเทคโนโลยีการรู้จำและบ่งชี้คำเป้าหมายในประโยคต่างๆ โดยคำในกลุ่มคำศัพท์จะมีจำนวนน้อยมากๆ และจะไม่สนใจคำอื่นๆที่ไม่ได้อยู่ในกลุ่มคำศัพท์นั้น

ซึ่งในโครงงานนี้จะทำการศึกษาการรู้จำเสียงในประเภท "การรู้จำคำเป้าหมาย"

2.2 เสียง (Sound)

เสียงเป็นพลังงานรูปหนึ่ง ไม่มีตัวตน ไม่มีน้ำหนัก และไม่ต้องการที่อยู่ แต่สามารถทำงานได้ กล่าวคือ เสียงทำให้แก้วหูของเราสั่น เมื่อแก้วหูสั่นก็จะกระเทือนต่อเนื่องกันไปจนถึงประสาทหูเปลี่ยนเป็นสัญญาณไฟฟ้าสู่สมอง ทำให้เราได้ยิน เสียง ซึ่งเสียงเกิดขึ้นรอบตัวเราทั้งวัน เสียงอาจเกิดขึ้นเองตามธรรมชาติ หรือเกิดขึ้นเนื่องจากคนเป็นผู้กระทำก็ได้

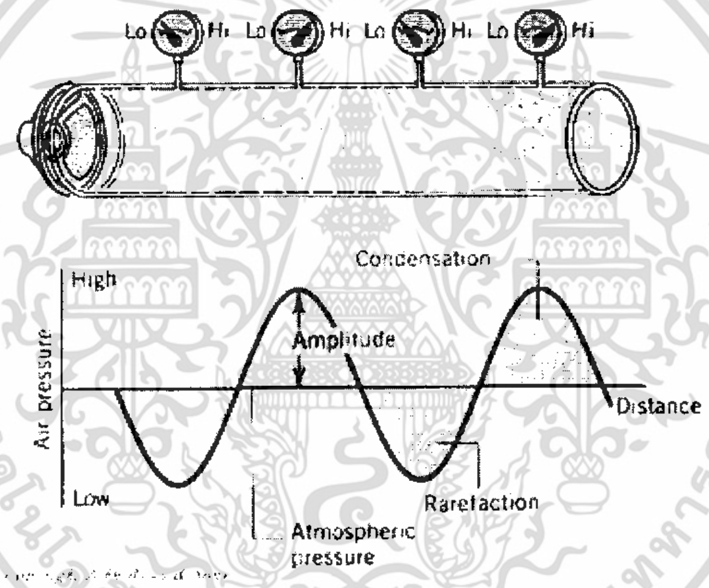
เสียงเป็นคลื่นตามยาวชนิดหนึ่งซึ่งอาศัยตัวกลางในการเคลื่อนที่ ดังนั้นคุณสมบัติของเสียงจึงเหมือนคลื่นทุกประการ ตามปกติหูคนสามารถได้ยินเสียงในช่วงความถี่ 20 เฮิร์ตซ์(Hertz) ถึง 20,000 เฮิร์ตซ์ คลื่นเสียงที่มีความถี่สูงกว่า 20,000 เฮิร์ตซ์นั้น เรียกว่าคลื่นเหนือเสียงหรืออัลตราโซนิก (ultrasonic) ส่วนเสียงที่ความถี่ต่ำกว่า 20 Hz เรียกว่าคลื่นใต้เสียงหรืออินฟราโซนิก (Infrasonic Wave)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2.1 ธรรมชาติและความดันของเสียง

เสียงเป็นคลื่นความดัน (Pressure Wave) จะต้องอาศัยตัวกลางในการเคลื่อนที่ ดังนั้นจึงสามารถเคลื่อนที่ผ่านอากาศ ของแข็งหรือของเหลว แต่ไม่สามารถเคลื่อนที่ผ่านสุญญากาศได้

คลื่นเสียงเป็นคลื่นตามยาวเกิดจากการสั่นของวัตถุ ความถี่ของเสียงจะมีค่าเท่ากับความถี่ของแหล่งกำเนิด และในขณะที่มี การสั่น โมเลกุลของตัวกลางจะมีการถ่ายทอดพลังงานทำให้เกิดความดันอากาศที่เปลี่ยนแปลงไปตามตำแหน่ง ทำให้เกิดเป็นช่วงอัด และ ช่วงขยายโดยที่ช่วงอัดคือบริเวณที่อนุภาคของตัวกลางอัดเข้าหากัน บริเวณนี้มีจะมีความดันสูงสุดโดยเทียบกับความดันที่ตำแหน่ง สมดุลของอนุภาค โดยการขจัดของอนุภาคน้อยที่สุด ส่วนช่วงขยายคือบริเวณที่อนุภาคตัวกลางแยกห่างจากกัน บริเวณนี้มีความดัน ต่ำสุดโดยเทียบกับความดันที่ตำแหน่งสมดุลของอนุภาค การขจัดของอนุภาคมากที่สุด ซึ่งสามารถเขียนเป็นกราฟได้ ดังรูป



รูปที่ **2.1** แสดงการขจัดของอนุภาคของอากาศ

2.2.2 คุณสมบัติของเสียง

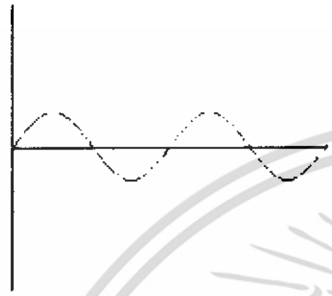
2.2.2.1 ระดับเสียง (Pitch) คือระดับความสูง - ต่ำของเสียงในการพูดคุยหรือการสนทนาของมนุษย์เราทุกครั้งเราจะพบว่ามีความแตกต่างของระดับเสียงได้ชัดเจน โดยเฉพาะอย่างยิ่งภาษาไทยเราแล้วยิ่งเห็นได้ชัดเจนเพราะเรามีการผันเสียงของวรรณยุกต์ ทำให้เกิดเสียงสูง – ต่ำ และทำให้ความหมายแตกต่างกันออกไปเช่น ปา ป่า ป้า เป็นต้น ถ้าหากการพูดคุยของมนุษย์เราไม่มีความแตกต่างของระดับเสียงสูง - ต่ำ การพูดคุยคงเป็นสิ่งที่น่าเบื่อมาก และคงไม่มีเสียงดนตรีอย่างที่เราได้ยินจนปัจจุบัน

ระดับเสียงขึ้นอยู่กับความถี่ (frequency) ของการสั่นสะเทือน เช่น ถ้าการสั่นสะเทือนยิ่งเร็วเสียงจะยิ่งสูง แต่ถ้าการสั่นสะเทือนยิ่งช้าเสียงก็จะยิ่งต่ำ ความถี่ของการสั่นสะเทือนปกติจะวัดเป็น

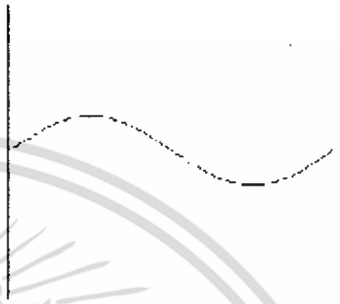
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รอบต่อ วินาที (cycles/second) ในเปียโนมีความถี่ของการสั่นสะเทือนเสียงที่สูงที่สุดอยู่ที่ 4,186 รอบต่อวินาทีและความถี่ที่ต่ำที่สุดคือ27รอบต่อวินาที

โดยทั่วไปแล้ววัตถุที่สั่นสะเทือนยิ่งมีขนาดเล็กก็จะทำให้เกิดการสั่นสะเทือนเร็วขึ้น สิ่งที่ตามมาก็คือเสียงสูงขึ้น เช่น ถ้าเราดีดสายไวโอลิน (Violin) เปล่า ๆ จะพบว่าจะมีความถี่สูงหรือเสียงสูงกว่าเราดีดสายดับเบิลเบส (double base)เปล่า ๆ ทั้งนี้เนื่องจากว่าสายของไวโอลินมีขนาดและความยาวน้อยกว่าดับเบิ้ลเบสนั่นเอง



รูปที่ **2.2** แสดงการสั่นสะเทือนเสียงสูง



รูปที่ **2.3** แสดงการสั่นสะเทือนเสียงต่ำ

2.2.2.2 ความดัง – ค่อย (Dynamics) ความดังและความค่อยในทางดนตรีเรียกว่า ไดนามิก (Dynamic) เป็นองค์ประกอบที่สำคัญในทางดนตรี มีความสัมพันธ์กับช่วงกว้างของคลื่นเสียง (Amplitude) ในการสั่นสะเทือนที่ทำให้เกิดเสียงนั้น ๆ ช่วงกว้างมากเสียงจะดัง และช่วงกว้างน้อยเสียงจะเบา เช่น ถ้าเราดีดสายกีตาร์แรงเท่าใดเสียงที่ออกมาก็จะดังในการตรงกันข้าม หากเราดีดสายกีตาร์เบาเสียงที่ออกมาก็จะเบา เป็นต้น

2.2.3 การกำเนิดเสียง

1. อากาศถูกสูดเข้าไปในปอด
2. อากาศถูกปล่อยออกจากปอดผ่านหลอดลม
3. เมื่ออากาศผ่านเส้นเสียงที่อยู่ในกล่องเสียง เส้นเสียงจะเกิดการสั่นสะเทือน
4. จากนั้นเมื่ออากาศผ่านช่องคอจะเกิดการแบ่งเป็นจังหวะ
5. แล้วผ่านช่องปากและช่องจมูก ทำให้เกิดเป็นเสียงต่างๆกัน

2.3 การแปลงเสียงเป็นสัญญาณไฟฟ้า

เมื่อต้องการวิเคราะห์สัญญาณเสียง จำเป็นต้องแปลงสัญญาณเสียงให้อยู่ในรูปของสัญญาณไฟฟ้าก่อนโดยใช้อุปกรณ์คือ ไมโครโฟน

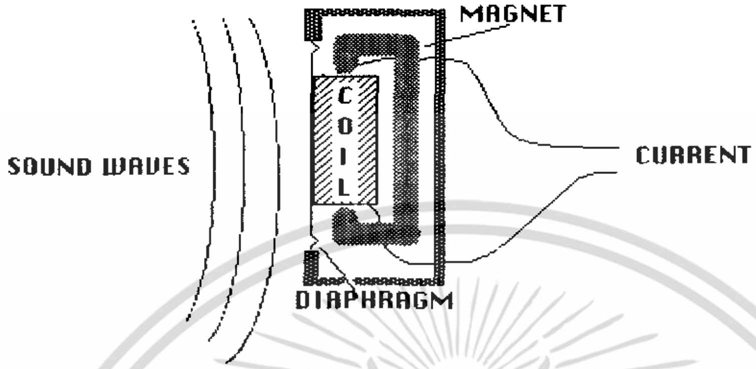
ซึ่งหลักการทำงานของไมโครโฟนมีดังนี้

1. เมื่อคลื่นเสียงมากระทบแผ่นไดอะแฟรม (diaphragm) ของไมโครโฟน ทำให้แผ่นไดอะแฟรมสั่นสะเทือน ขดลวดที่ติดอยู่กับแผ่นไดอะแฟรมจะสั่นสะเทือนตามไปด้วย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. บริเวณใกล้กับขดลวดเป็นแม่เหล็กถาวร โดยขณะที่ขดลวดสั่นเข้าใกล้และสั่นออกห่างจากแม่เหล็ก จะทำให้ฟลักซ์แม่เหล็กเปลี่ยนแปลง

3. จากกฎของฟาราเดย์ การเปลี่ยนแปลงฟลักซ์แม่เหล็กทำให้เกิดแรงเคลื่อนไฟฟ้าเหนี่ยวนำขึ้น คือสัญญาณไฟฟ้า ซึ่งเป็นสัญญาณอนาล็อค



รูปที่ **2.4** แสดงโครงสร้างของไมโครโฟน

2.4 คอมพิวเตอร์กับการประมวลผลสัญญาณเสียง

เมื่อสามารถแปลงสัญญาณเสียงเป็นสัญญาณไฟฟ้าได้แล้ว ในขั้นตอนประมวลผลด้วยคอมพิวเตอร์จะต้องแปลงสัญญาณไฟฟ้าซึ่งเป็นสัญญาณอนาล็อกให้เป็นสัญญาณดิจิตอลก่อน

โดยในขั้นตอนการบันทึกเสียงจะมีโปรแกรมที่แปลงจากสัญญาณอนาล็อกเป็นดิจิตอลโดยไม่มีการบีบอัดหรือเข้ารหัส ซึ่งเป็นไฟล์ที่มีนามสกุล (.wav) ไฟล์ชนิดนี้มีลักษณะเป็นกลุ่ม เรียกว่าชังค์(*chunk*) ซึ่งโครงสร้างของไฟล์ชนิดนี้จะประกอบด้วย 3 ส่วนใหญ่ ๆ คือ

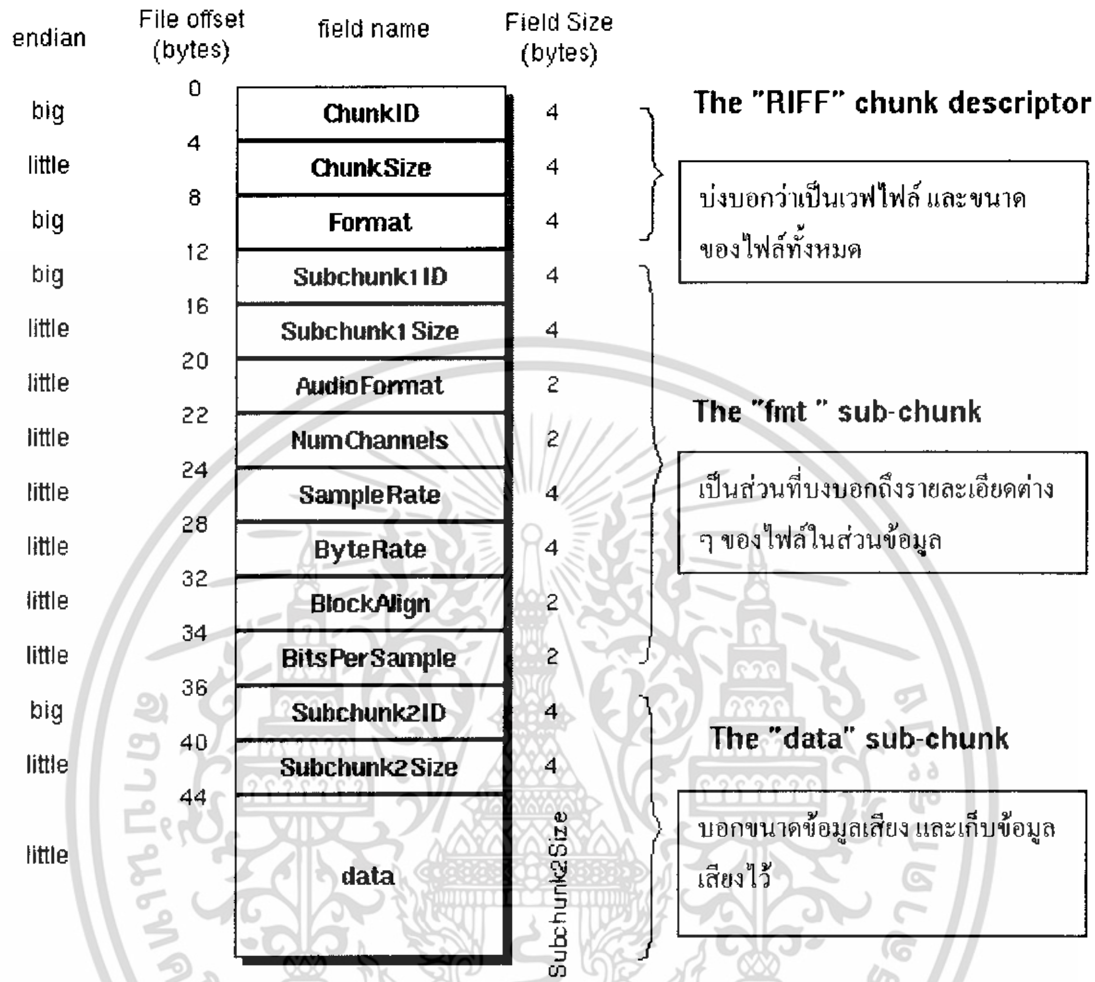
1. เฮดเดอร์ (Header)
2. ฟอร์แมตซับชังค์ (fmt sub-chunk)
3. ดาต้าซับชังค์ (data sub-shunk)

2.5 รูปแบบของข้อมูลเสียงที่เป็นพีซีเอ็ม (WAVE PCM soundfile format)

รูปแบบของเวฟไฟล์ เป็นส่วนหนึ่งของ Microsoft's RIFF specification สำหรับการเก็บไฟล์มัลติมีเดีย RIFF file เริ่มต้นด้วยส่วนของเฮดเดอร์ ตามด้วย ส่วนของข้อมูลโดยปกติแล้วไฟล์ของสัญญาณจะประกอบด้วย RIFF file และ wave chunk ซึ่งแบ่งออกเป็นสองแบบคือฟอร์แมตชังค์ (format chunk) และดาต้าชังค์ (data chunk) จะสามารถเรียกรูปแบบนี้ว่า”Canonical form".

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปแบบของเวฟไฟล์



รูปที่ **2.5** แสดงรูปแบบของเวฟไฟล์

1. เฮดเดอร์ของเวฟไฟล์ (Wave File Header - RIFF Type Chunk) ใน 8 ไบต์ แรกของเวฟไฟล์(Wave File)เป็นมาตรฐานของ RIFF chunk header ซึ่งมี chunk ID ของ "RIFF" และ chunk size ซึ่งมีค่าเท่ากับขนาดของไฟล์ลบด้วย 8 ไบต์ 4 ไบต์ แรกของข้อมูลใน RIFF chunk จะแสดงชนิดของ RIFF chunk โดยเวฟไฟล์จะตามด้วย "WAVE"

2. เวฟไฟล์ชังค์ (Wave File Chunks) มีชังค์หลายชนิดสำหรับเวฟไฟล์ โดยเวฟไฟล์ส่วนใหญ่ประกอบไปด้วยเพียงสองชนิด ซึ่งก็คือ ฟอร์แมตชังค์ และ ดาต้าชังค์ โดยชังค์ทั้งสองชนิดนี้ใช้สำหรับอธิบายรูปแบบของดิจิตอลออดิโอแซมเปิ้ล (digital audio samples) และตัวสัญญาณสุ่มเอง ซึ่งโดยทั่วไปแล้วฟอร์แมตชังค์จะมาก่อนดาต้าชังค์

3. รูปแบบของชังค์ (Format Chunk - "fmt ") รูปแบบของชังค์ประกอบไปด้วยข้อมูลเกี่ยวกับการการจัดเก็บสัญญาณ, รหัสชังค์ (chunk ID), ขนาดข้อมูล (chunk data size), จำนวน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ช่องสัญญาณ ,อัตราสัญญาณสุ่ม (sample rate), จำนวนบิตแต่ละสัญญาณสุ่ม (bits per sample) และข้อมูลอื่น ๆ

4. รหัสของชังค์และขนาดของข้อมูล (Chunk ID and Data Size) รหัสของชังค์จะมีค่า "fmt" (0x666D7420) เสมอ และขนาดจะเป็นขนาดของข้อมูลมาตรฐาน (16 ไบต์) บวกขนาดของไบต์ข้อมูลเพิ่มเติม ถ้าเวฟนั้นไม่ได้เป็นข้อมูลสัญญาณ PCM ที่ไม่ถูกบีบอัด, รหัสของชังค์จะจบด้วย (0x20)

5. รหัสการบีบอัดข้อมูล (Compression Code) คำแรกของรูปแบบข้อมูล (format data) จะแสดงถึงชนิดของการบีบอัดข้อมูลที่ใช้ในสัญญาณนั้น ๆ

6. จำนวนช่องสัญญาณ (Number of Channels) จำนวนช่องสัญญาณ จะแสดงถึงจำนวนของ สัญญาณเสียงที่ถูกเข้ารหัสในเวฟดาตาชังค์

ค่าเท่ากับ 1 แปลว่าเป็นโมโน

ค่าเท่ากับ 2 แปลว่าเป็นสเตอริโอ

7. อัตราสัญญาณสุ่ม (Sample Rate) แสดงจำนวนของสัญญาณสุ่มต่อวินาที

8. จำนวนไบต์ของสัญญาณต่อวินาที (Average Bytes Per Second) ค่านี้จะแสดงจำนวนไบต์ของสัญญาณที่จะถูกแปลงจากดิจิตอลเป็นอนาลอคต่อวินาที ซึ่งค่านี้สามารถคำนวณได้จาก

AvgBytesPerSec = SampleRate * BlockAlign

9. บล็อคอไลน์ (Block Align) เป็นจำนวนไบต์ของสัญญาณสุ่ม สามารถหาค่านี้ได้จาก

BlockAlign = SignificantBitsPerSample / 8 * NumChannels

10. จำนวนบิตที่ใช้ในการระบุแต่ละสัญญาณสุ่ม (Significant Bits Per Sample) ค่านี้จะแสดงจำนวนบิตที่ใช้ในการระบุแต่ละสัญญาณสุ่ม ซึ่งปกติจะมีค่า 8, 16, 24 หรือ 32

11. จำนวนไบต์เพิ่มเติม (Extra Format Bytes) ค่านี้จะแสดงจำนวนของไบต์เพิ่มเติมโดยค่านี้ไม่จำเป็นต้องมีถ้ารหัสการบีบอัดข้อมูลเป็น 0 ซึ่งก็คือ ไฟล์ PCM ที่ไม่ถูกบีบอัด แต่ถ้าการบีบอัดข้อมูลแล้ว ค่านี้จะขึ้นอยู่กับข้อมูลของสัญญาณที่ต้องการถอดรหัส

12. ดาต้าชังค์ (Data Chunk - "data") ข้อมูลสัญญาณเสียงประกอบไปด้วยสัญญาณสุ่มที่เป็นดิจิตอล (digital audio sample) ซึ่งสามารถถอดรหัสได้โดยใช้ รูปแบบและวิธีการบีบอัดข้อมูลใน wave format chunk ถ้าค่ารหัสการบีบอัดข้อมูลมีค่าเท่ากับ 1 ซึ่งก็คือสัญญาณPCMที่ไม่ถูกบีบอัด ดังนั้นสัญญาณข้อมูลจะประกอบด้วยข้อมูลดิบ ดาตาชังค์จะอธิบายว่าสัญญาณPCMที่ไม่ถูกบีบอัด มีการจัดเก็บอย่างไร

2.6 การเปรียบเทียบโดยโปรแกรมแบบพลวัต (Dynamic Programming Matching)

เป็นวิธีการที่ใช้ในการหาความแตกต่างกันระหว่างสองเวกเตอร์ขนาดไม่เท่ากัน โดย $u^{\#}$ และ u^{*} เป็นเวกเตอร์ของเสียงที่จะนำมาวิเคราะห์ I และ L เป็นจำนวนองค์ประกอบ (element)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ของเสียงอ้างอิงและเสียงที่นำมาทดสอบตามลำดับ และ $w(k)$ คือฟังก์ชันวาร์ปปิง (warping function) ซึ่งหาได้จาก

$$W = \{w(1), w(2), ... w(k)\} \quad (2.1)$$

โดยที่ $$w(k) = (i_k, l_k), (1 \le i_k \le I), 1 \le l_k \le L) \quad (2.2)$$

k คือ จำนวนจุดบนฟังก์ชันวาร์ปปิง

โดยที่ $$w(1) = (1,1) \quad (2.3)$$

และ $$w(k) = (I, L) \quad (2.4)$$

สามารถหาความแตกต่างระหว่างสองเวกเตอร์ (S) ได้จาก

$$S = \frac{1}{I+L} \min[\sum_{k=1}^{K} d(w(k)) g_k] \quad (2.5)$$

ซึ่ง $$d(w(k)) = | u^*(i_k) - u^\#(l_k) | \quad (2.6)$$

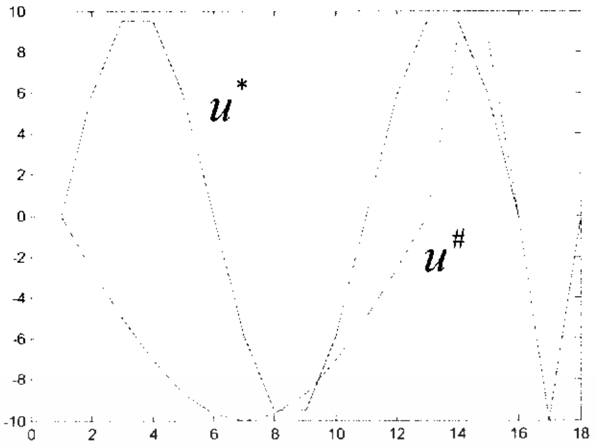
และ $$g_k = (i_k - i_{k-1}) + (l_k - l_{k-1}) \quad (2.7)$$

สมการที่ (2.5) สามารถเขียนใหม่ได้เป็น

$$h(w(k)) = \min \begin{cases} h(i_k - 1, l_k - 1) + 2d(i_k, l_k) \\ h(i_k - 1, l_k) + d(i_k, l_k) \\ h(i_k, l_k - 1) + d(i_k, l_k) \end{cases} \quad (2.8)$$

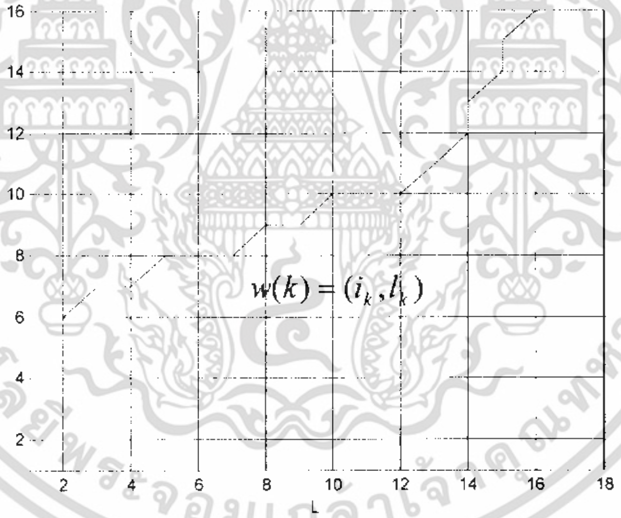
$$S = \frac{1}{I+L} h(w(K)) \quad (2.9)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ **2.6** แสดงเวกเตอร์ของสัญญาณอ้างอิง ($u^\#$) และเวกเตอร์ของสัญญาณทดสอบ (u^*)

เมื่อนำเวกเตอร์ของสัญญาณทั้งสองมาจับคู่กันจะได้ฟังก์ชันวาร์ปปิง $w(k)$ ดังรูปข้างล่าง ซึ่งแสดงให้เห็นว่าสัญญาณแต่ละคู่จับคู่กันที่จุดใดบ้าง



รูปที่ **2.7** แสดงฟังก์ชันวาร์ปปิง $w(k)$

2.7 การประมาณค่าด้วยเส้นตรง

ฟังก์ชันพีซไวส์ลิเนียร์(piecewise linear function) เป็นฟังก์ชันที่ใช้ในการประมาณค่าที่อยู่ระหว่างจุดต่อจุด โดยทำการลากเส้นตรงเชื่อมจุดเหล่านั้นเข้าด้วยกัน ทำให้เราสามารถประมาณค่าข้อมูลที่ต้องการได้ หรือกล่าวได้ว่าเป็นการแปลงข้อมูลที่เรามีอยู่ซึ่งเป็นแบบไม่ต่อเนื่องให้เป็นข้อมูลแบบต่อเนื่อง โดยสามารถแสดงสมการได้ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$x(t)=\sum_{n=0}^{N}x(t_n)\phi_n(t) \tag{2.10}$$

$$\phi_n(t)=\begin{cases}\dfrac{t-t_{(n-1)}}{t_n-t_{(n-1)}}, t\in\left[t_{(n-1)},t_n\right]\\ \dfrac{t_{(n+1)}-t}{t_{(n+1)}-t_n}, \quad t\in\left[t_n,t_{(n+1)}\right]\\ 0 \quad ,t\notin\left[t_{(n-1)},t_{(n+1)}\right]\end{cases} \tag{2.11}$$

โดยที่ $x(t)$ คือ ข้อมูลที่ได้จากการประมาณค่า , $x(t_n)$ คือ ข้อมูลเดิม

$\phi_n(t)$ คือ ตัวคูณซึ่งหาได้จากเงื่อนไขต่างๆข้างต้น

2.8 อนุกรมฟูเรียร์ (Fourier Series)

เนื่องมาจากตัวอนุกรมเป็นอนุกรมไม่รู้จบ (Infinite series) ดังนั้นการประยุกต์ใช้อนุกรมกับฟังก์ชันคาบอื่น ๆ จึงมีประโยชน์มาก ทำให้สามารถประมวลผลข้อมูลจำนวนมากได้อย่างรวดเร็ว เพราะผลลัพธ์ที่ได้จากสมการจะมีขนาดลดลง แต่ยังคงคุณลักษณะเดิมไว้อย่างครบถ้วนสมบูรณ์ การวิเคราะห์ด้วยอนุกรมฟูเรียร์จะแสดงให้เห็นถึงแนวโน้มของข้อมูล และสื่อออกมาด้วยรูปสมการอย่างง่าย ดังนั้นจึงถือว่าอนุกรมฟูเรียร์นั้นมีประโยชน์อย่างมากในการประมาณค่า (Approximation) ของฟังก์ชันคาบ (Periodic function) อนุกรมฟูเรียร์

ให้ $f(t)$ คือฟังก์ชันคาบ ที่คาบเท่ากับ T

$$F_n=\frac{1}{T}\int_{-T/2}^{T/2}f(t)e^{-\mathrm{int}}dt \tag{2.12}$$

Fn คือ สัมประสิทธิ์ฟูเรียร์ (Fourier coefficient) ดังนั้นอนุกรมฟูเรียร์ในรูป $f(t)$ คือ

$$f(t)=\sum_{n=-\infty}^{\infty}F_n e^{\mathrm{int}} \tag{2.13}$$

ผลรวมในแต่ละเทอมของสมการนี้ จะเรียกว่า ฮาร์โมนิก (Harmonic) โดย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$e^{int} = \cos(nt) + i\sin(nt) \tag{2.14}$$

แสดงให้ $f(t)$ อยู่ในรูปของฟังก์ชันโคซายน์ (Cosine) และซายน์ (Sine) ได้ว่า

$$f(t) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty}[a_n \cos(nt) + b_n \sin(nt)]$$

โดย
$$a_0 = \frac{1}{T}\int_{-T/2}^{T/2} f(t)dt \tag{2.15}$$

$$a_n = \frac{2}{T}\int_{-T/2}^{T/2} f(t)\cos(nt)dt \tag{2.16}$$

$$b_n = \frac{2}{T}\int_{-T/2}^{T/2} f(t)\sin(nt)dt \tag{2.17}$$

ซึ่งสัมพันธ์กับ F(n) คือ $F_n = (a_n - ib_n)/2$ และ $F_{-n} = F_n^*$ ดังนั้น $F_0 = a_0/2$
ค่าสัมบูรณ์ของสัมประสิทธิ์ฟูเรียร์สามารถหาได้จากสมการดังนี้

$$\text{ขนาด (Magnitude)} = |\,a_n + jb_n\,| \tag{2.18}$$

2.9 การวัดระยะแบบยูคลิด (Euclidean distance)

ในการรู้จำ หรือการทำให้คอมพิวเตอร์วสามารถแยกแยะสิ่งต่าง ๆ ได้ ข้อมูลที่ใช้อ้างอิงเป็นสิ่งจำเป็น เช่นต้องการให้คอมพิวเตอร์แยกแยะเสียงที่รับเข้ามาได้ว่าเป็นเสียงอะไร จำเป็นต้องมีข้อมูลเสียงเป็นข้อมูลอ้างอิง และสิ่งที่จะเป็นตัวกำหนดว่าเสียงที่รับเข้ามานั้น เป็นเสียงเดียวกับข้อมูลอ้างอิงหรือไม่คือการวัดความแตกต่างของข้อมูลว่ามากน้อยเพียงใด

การวัดระยะแบบยูคลิด คือ คือวิธีในการวัดค่าของรูปแบบ ของเวกเตอร์ใดๆ ว่ามีความแตกต่างกันเท่าไหร่ วิธีการทั่วไปคิดจากการวัดระยะทางระหว่าง ค่าในเวกเตอร์ 2 เวกเตอร์ที่ตำแหน่งเดียวกัน ถ้ากำหนดให้ $u = (x_1, y_1)$ และ $v = (x_2, y_2)$ เป็นเวกเตอร์รูปแบบใดๆ

จะได้การวัดระยะแบบยูคลิดจะได้เท่ากับ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \tag{2.19}$$

ดังนั้นหากจะหาระยะทางระหว่าง a และ b ซึ่งมีข้อมูลดังนี้

$$a = \{x_1, x_2, \dots, x_n\} \tag{2.20}$$

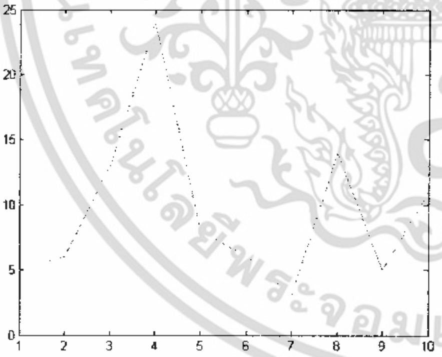
$$b = \{y_1, y_2, \dots, y_n\} \tag{2.21}$$

จาก (1) จะได้ระยะทางยูคลิดจาก a ไป b

$$d\left(f_1, f_2\right) = \sqrt{\left(x_1 - y_1\right)^2 + \left(x_2 - y_2\right)^2 + \dots + \left(x_n - y_n\right)^2} \tag{2.22}$$

เนื่องจากเป็นวิธีการที่ไม่ซับซ้อน จึงสามารถนำไปประยุกต์ใช้ในการหาระยะทางหรือมุมได้ ไม่ว่าจะเป็นเวกเตอร์ หรือ สัญญาณข้อมูล

หากต้องการหาระยะทางยูคลิดจากสัญญาณ $f_1(x)$ ไปถึงสัญญาณ $f_2(x)$ โดยสัญญาณทั้งสองมีรูปสัญญาณดังรูป



$f_1(x) = \{7,6,12,9,11,7,3,5,6,1\}$

$f_2(x) = \{5,6,13,24,8,6,3 ,4,5,11\}$

รูปที่ **2.8** แสดงตัวอย่างฟังก์ชั่นที่ใช้ทดสอบการระยะแบบยูคลิด

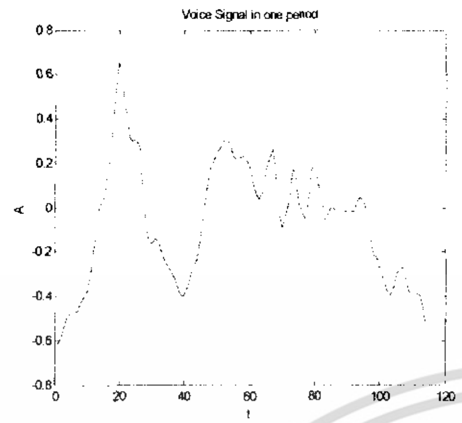
จะได้ระยะยูคลิด $d(f1, f2) = \sqrt{(7-5)^2 + (6-6)^2 + \dots + (1-11)^2} = 20.543$

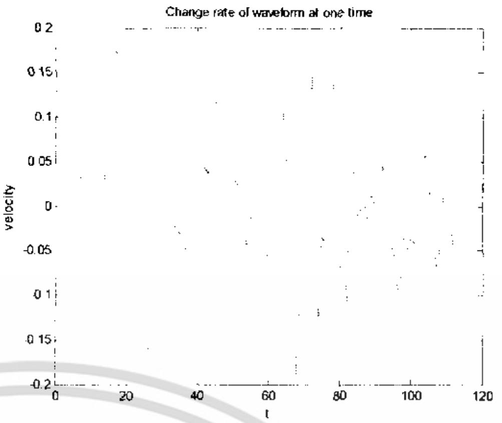
2.10 การหาอัตราการเปลี่ยนแปลงของคลื่นสัญญาณเสียงตามเวลาที่พูด

อัตราการเปลี่ยนแปลงรูปร่างของคลื่นสัญญาณเสียงเป็นการเปลี่ยนแปลงสัญญาณในช่วงเวลาหนึ่ง ซึ่งการเปลี่ยนแปลงนี้มีทั้งขนาดและทิศทาง โดยเราสามารถหาได้จาก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$v(t) = [s(t+1) - s(t)] \tag{2.23}$$





รูปที่ **2.9** แสดงรูปร่างของสัญญาณเสียง รูปที่ **2.10** แสดงอัตราการเปลี่ยนแปลงรูปร่างของสัญญาณ

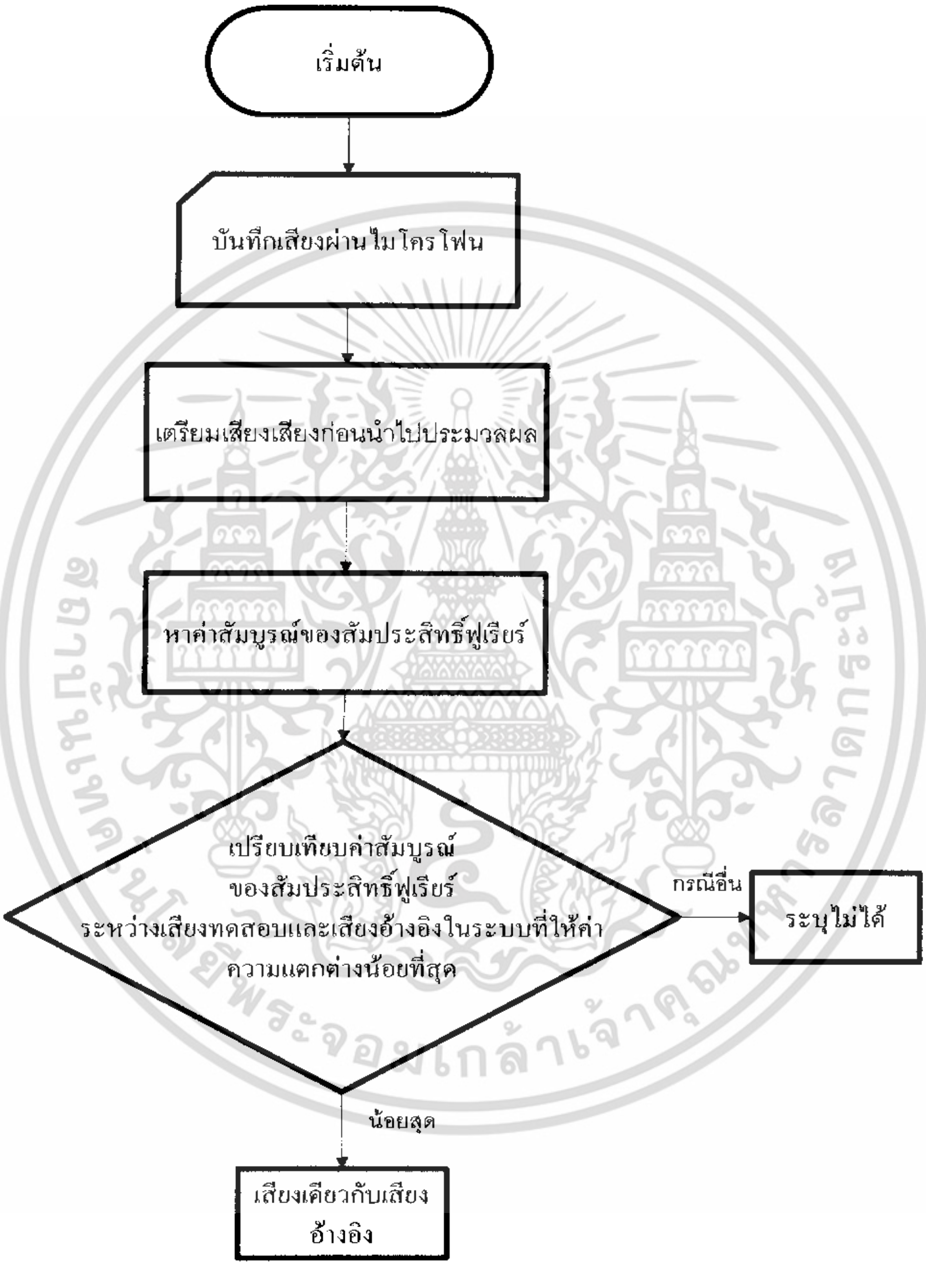
จากรูปที่ 2.9 แสดงรูปร่างของสัญญาณเสียงที่มีการเปลี่ยนแปลงตามเวลา ซึ่งสามารถหาอัตราการเปลี่ยนแปลงคลื่นของสัญญาณเสียงได้ดังรูปที่ 2.10

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3

การออกแบบ

3.1 โครงสร้างของโปรแกรม



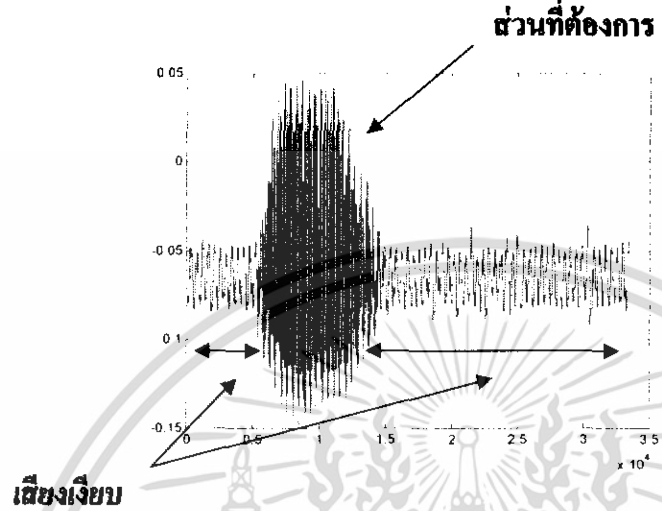
รูปที่ **3.1** แสดงโครงสร้างของโปรแกรม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2 การเตรียมเสียงก่อนนำไปประมวลผล (Preprocessing)

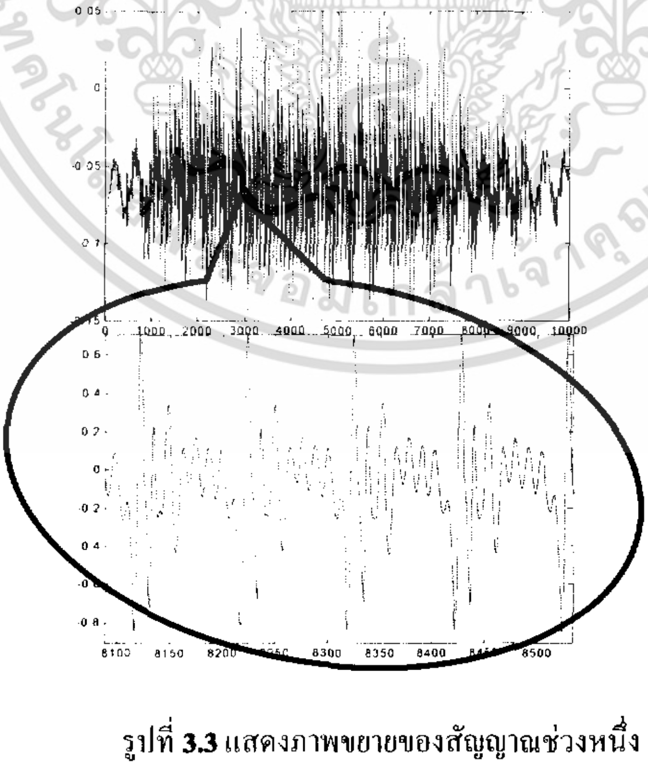
ขั้นตอนที่จะทำให้สัญญาณเสียงที่จะนำไปใช้หรือรับเข้ามานั้นมีความสมบูรณ์มากที่สุด

3.2.1 การตัดเสียงเงียบ ทำการเลือกเฉพาะช่วงที่เป็นเสียงมาพิจารณา โดยตัดช่วงที่เป็นเสียงเงียบออกไป



รูปที่ **3.2** แสดงรูปร่างของสัญญาณเสียง

ในการตัดเสียงเงียบออกสามารถสังเกตได้จากการดูรูปร่างของสัญญาณแล้วประมาณค่า โดยจะตัดตั้งแต่ค่าเทรสโฮลด์ (threshold) ขึ้นไป จากนั้นเมื่อเลือกเฉพาะช่วงเวลาเล็กๆมาขยายแล้วพิจารณา จะได้ดังรูป



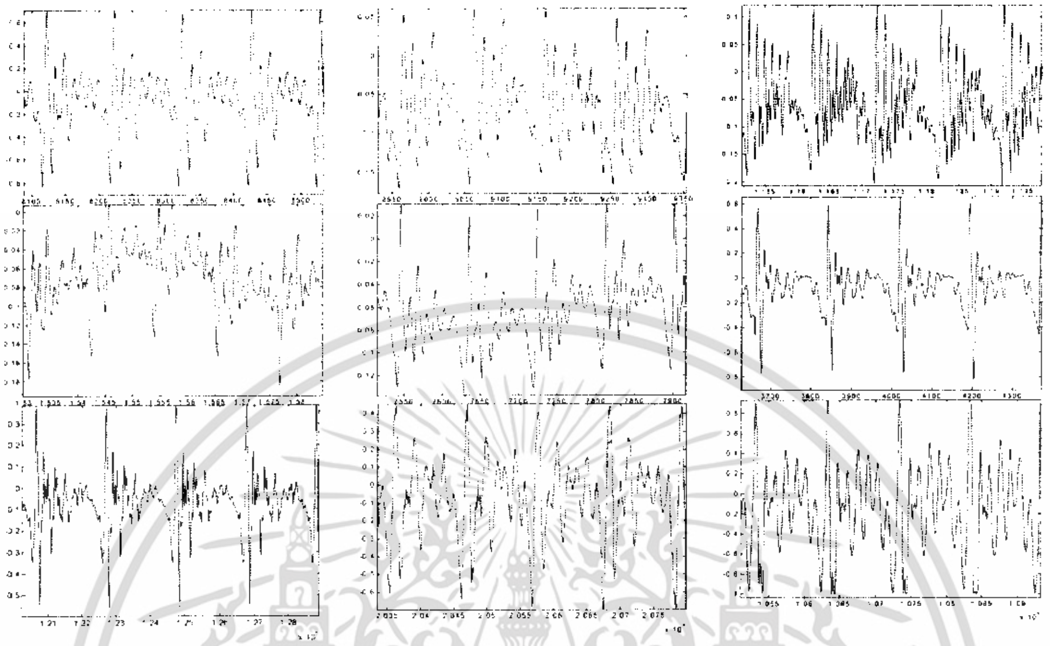
รูปที่ **3.3** แสดงภาพขยายของสัญญาณช่วงหนึ่ง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

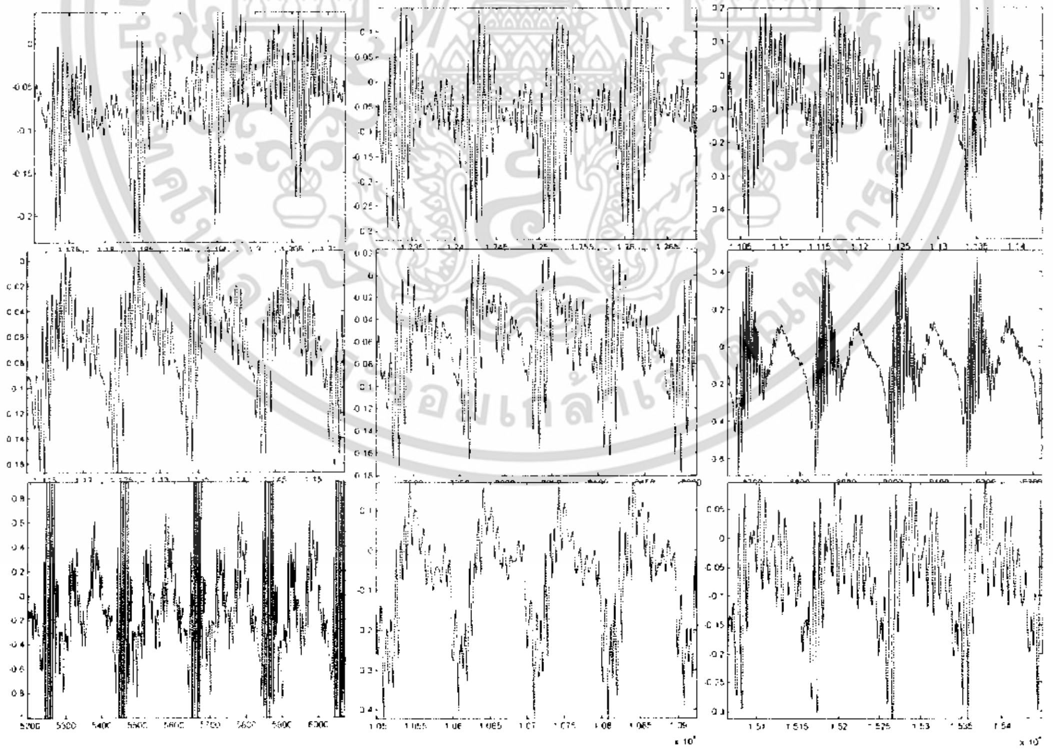
สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

เมื่อพิจารณาจะพบว่าสัญญาณเสียงมีลักษณะซ้ำกันไปเรื่อย ๆ เป็นคาบ

ตัวอย่างรูปร่างของสัญญาณเสียง



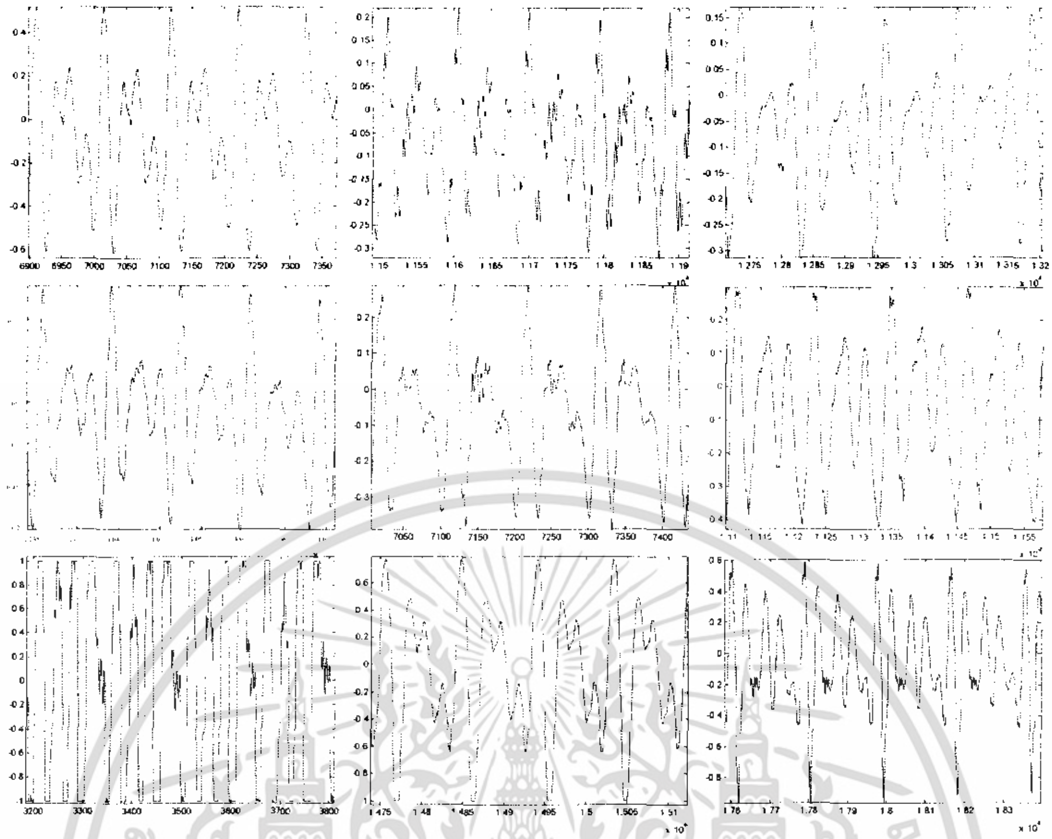
รูปที่ **3.4** แสดงรูปร่างของสัญญาณเสียง"อา" ของผู้พูดที่ต่างกัน



รูปที่ **3.5** แสดงรูปร่างของสัญญาณเสียง"อี" ของผู้พูดที่ต่างกัน

72655

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้ศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



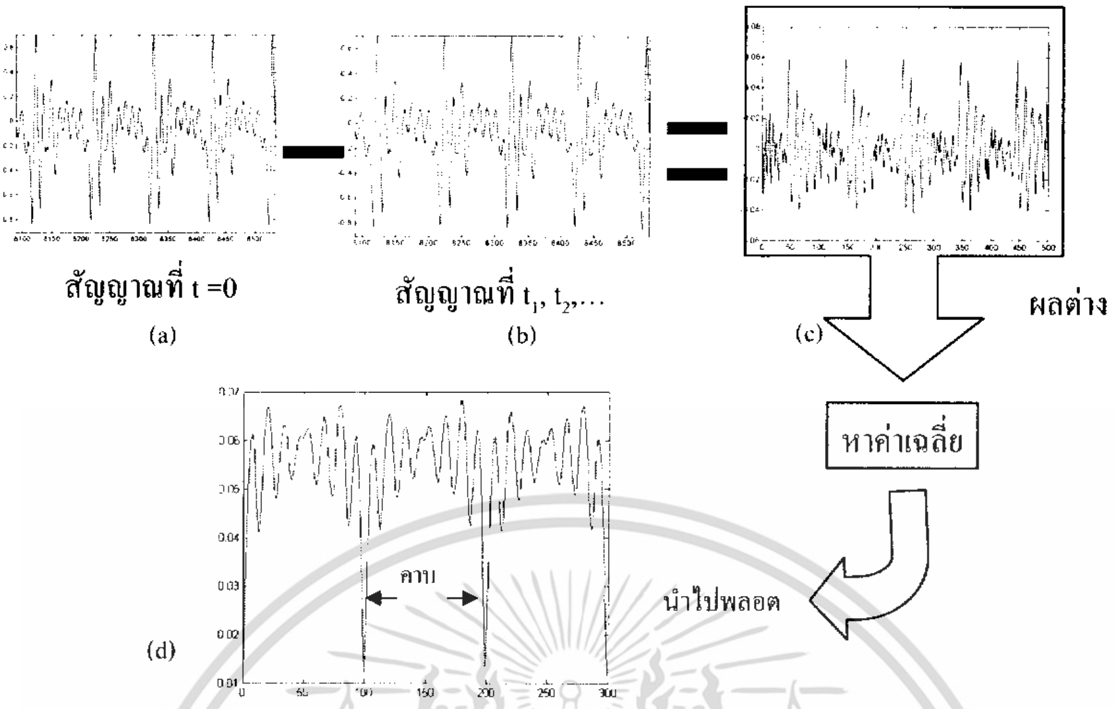
รูปที่ **3.6** แสดงรูปร่างของสัญญาณเสียง“โอ” ของผู้พูดที่ต่างกัน

จะเห็นได้จากตัวอย่างของเสียงว่าในแต่ละเสียงจะมีลักษณะของสัญญาณซ้ำกันไปเรื่อย ๆ ดังนั้นในการพิจารณาหาลักษณะเด่นจึงสามารถใช้เพียงหนึ่งคาบของเสียงเพื่อลดความยุ่งยากและซ้ำซ้อนในการคำนวณ

3.2.2 การหาคาบของเสียง

สัญญาณเสียงแต่ละเสียงไม่ว่าเสียงเดียวกันหรือเสียงต่างกันจะมีขนาดคาบที่ไม่เท่ากัน ดังนั้นจึงต้องหาขนาดคาบของแต่ละสัญญาณเสียงก่อนนำสัญญาณเสียงไปตัดและประมวลผล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



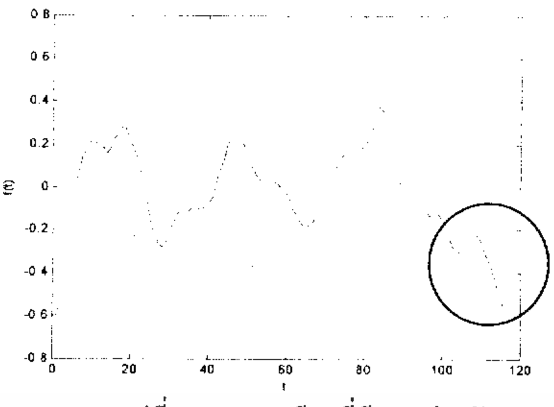
รูปที่ **3.7** แสดงขั้นตอนการหาคาบของสัญญาณ

จากรูปที่ 3.7 เป็นขั้นตอนการหาคาบของสัญญาณ โดยขั้นแรกทำการตัดสัญญาณมาช่วงหนึ่งได้ดังรูปย่อย (a) โดยสมมติให้ที่จุดแรกเป็น t = 0 จากนั้นนำสัญญาณที่ตัดมาเลื่อนไป 1 t จะได้ดังรูปย่อย (b) ซึ่งจุดแรกของช่วงที่เลื่อนไป 1 t เป็น t = 1 นำช่วงสัญญาณที่ตัดมาช่วงแรกลบกับสัญญาณช่วงที่เลื่อนไปจะได้ผลต่างของสัญญาณดังรูปย่อย (c) และนำผลต่างของสัญญาณมาหาค่าเฉลี่ยจะได้ค่าเฉลี่ยของผลต่างสัญญาณที่ t = 1จากนั้นนำช่วงสัญญาณที่ตัดมาเลื่อนไป 2 t แล้วนำไปลบกับสัญญาณที่ตัดมาช่วงแรกจะได้ค่าเฉลี่ยของผลต่างสัญญาณที่ t = 2 ทำซ้ำอย่างนี้ไปเรื่อยจนหมดช่วงสัญญาณที่ตัดมา แล้วนำค่าเฉลี่ยที่ได้ไปพลอตกราฟได้ดังรูปย่อย (d) โดยจุดที่มีค่าต่ำสุดนั้นเป็นจุดที่มีการซ้อนทับของสัญญาณพอดี ดังนั้นระยะห่างระหว่างจุดต่ำสุด 2 จุด จึงเป็นขนาดของคาบของสัญญาณ

3.2.3 การชักตัวอย่างใหม่ด้วยฟังก์ชันพีซไวส์ลิเนียร์

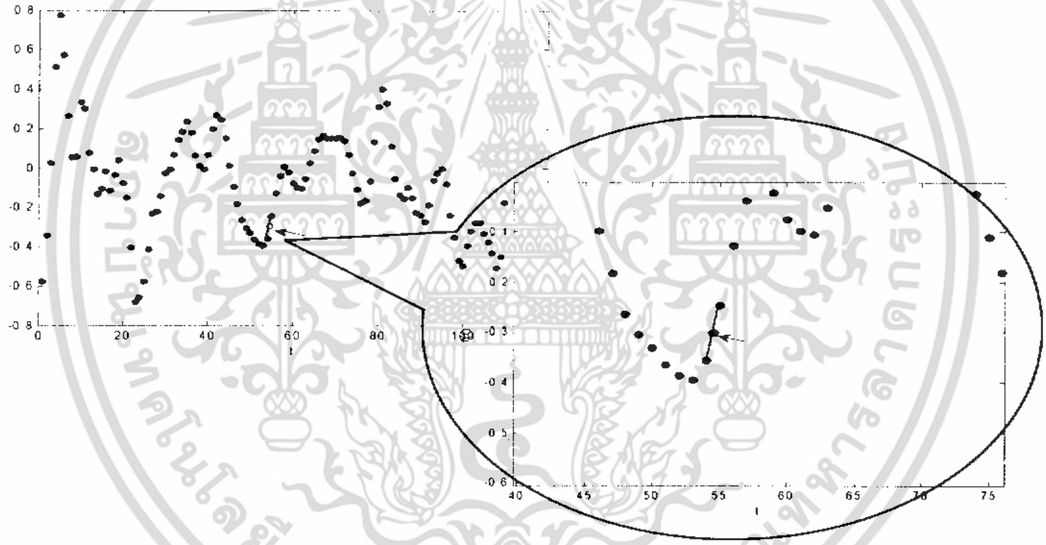
เมื่อทำการหาคาบของสัญญาณเสียงได้แล้ว จะพบว่าในแต่ละเสียงที่นำมาพิจารณานั้นมีคาบที่ต่างกันออกไป ซึ่งในการเปรียบเทียบโดยใช้สัมประสิทธิ์ฟูเรียร์ จะต้องทำให้แต่ละเสียงมีคาบที่เท่ากันก่อนด้วยวิธีการที่เรียกว่า “การชักตัวอย่างใหม่ (resampling)”

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



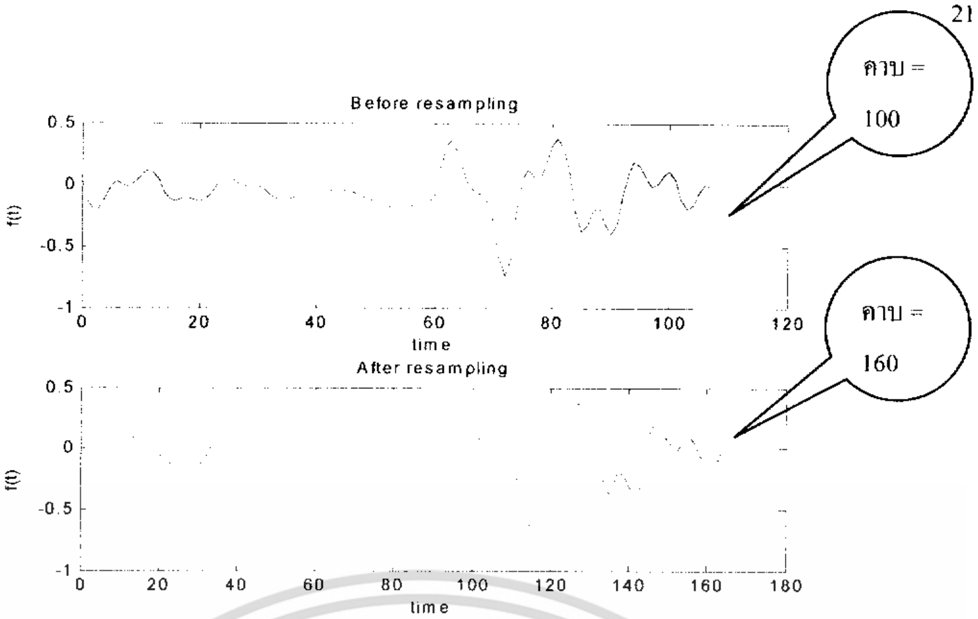
รูปที่ **3.8** แสดงเสียงที่มีคาบต่างกัน

การชักตัวอย่างใหม่ด้วยฟังก์ชันพีซไวส์ลิเนียร์ เป็นการเชื่อมต่อจุดสองจุดเข้าด้วยกันดังรูปที่ 3.9 แล้วเพิ่มจุดบนเส้นตรงเพื่อให้สัญญาณเสียงมีขนาดคาบที่เท่ากัน โดยใช้สมการที่ 2.11 ในบทที่ 2



รูปที่ **3.9** แสดงการเชื่อมต่อจุดด้วยวิธีการชักตัวอย่างใหม่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ **3.10** แสดงภาพก่อนและหลังการชักตัวอย่างใหม่

จากรูปที่ 3.10 จะเห็นได้ว่าสัญญาณหลังการชักตัวอย่างใหม่จะมีจำนวนองค์ประกอบที่เพิ่มขึ้น ในขณะที่รูปร่างของสัญญาณยังเหมือนกับรูปร่างของสัญญาณก่อนการชักตัวอย่างใหม่

3.3 กระบวนการดึงลักษณะเด่น (Feature Extraction)

จะเห็นได้ว่าสัญญาณเสียงที่เกิดจากคำพูดคำเดียวกันจะมีลักษณะบางอย่างที่เหมือนกันและมีลักษณะบางอย่างที่แตกต่างกัน เพื่อการรู้จำเสียงที่มีประสิทธิภาพจึงจำเป็นต้องดึงเอาเฉพาะลักษณะเด่นที่มีลักษณะเหมือนกันและแยกหรือตัดลักษณะที่ต่างกันออกไป จากนั้นนำลักษณะเด่นที่ได้มาเก็บไว้เป็นต้นแบบสำหรับใช้ในการเปรียบเทียบกับเสียงที่จะนำมาตรวจสอบ

3.3.1 การหาลักษณะเด่นของเสียงโดยใช้อนุกรมฟูเรียร์

ในปริญญานิพนธ์ฉบับนี้สนใจลักษณะเฉพาะตัวของแต่ละเสียง ซึ่งสามารถแสดงได้โดยใช้ค่าสัมบูรณ์ของสัมประสิทธิ์ฟูเรียร์ซึ่งได้จากรูปคลื่นของสัญญาณเสียงและอัตราการเปลี่ยนแปลงของคลื่นตามเวลาเป็นลักษณะเด่น แต่ถ้าใช้ทั้งรูปคลื่นของสัญญาณเสียงและอัตราการเปลี่ยนแปลงของคลื่นตามเวลาโดยตรงจะทำให้ได้ประสิทธิภาพในการแยกเสียงไม่ดีเท่าที่ควร เนื่องจากมีความเบี่ยงเบน ซึ่งความเบี่ยงเบนนี้สามารถแก้ไขได้โดยการกระจายสัญญาณทั้งสองลงไปในอนุกรมฟูเรียร์

การดึงลักษณะเด่นโดยใช้อนุกรมฟูเรียร์ ภายใต้สมมติฐานที่ว่าสัญญาณเสียงใด ๆ ที่สามารถกำหนดให้อยู่ในรูปของฟังก์ชันคาบ สามารถแสดงสัญญาณให้อยู่ในรูปอนุกรมฟูเรียร์ได้ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

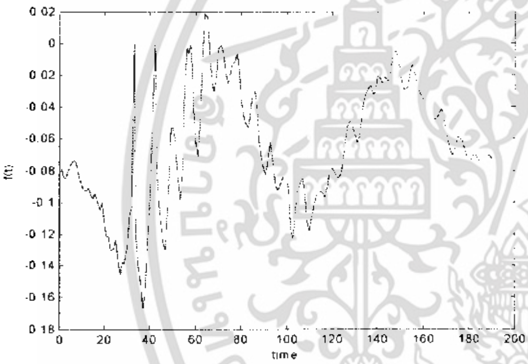
$$f(t)=\frac{1}{2}a_0+\sum_{n=1}^{\infty}[a_n\cos(nt)+b_n\sin(nt)] \tag{3.1}$$

สามารถคำนวณหาค่าสัมประสิทธิ์ของอนุกรมฟูเรียร์ได้ดังนี้

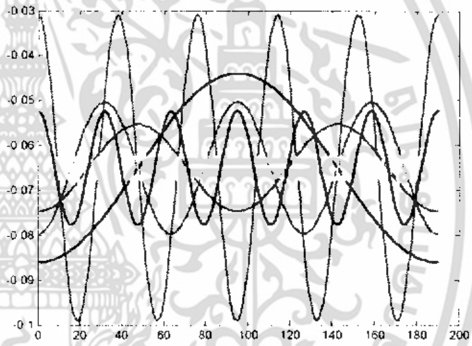
$$a_0=\frac{1}{T}\int_{-T/2}^{T/2}f(t)dt \tag{3.2}$$

$$a_n=\frac{2}{T}\int_{-T/2}^{T/2}f(t)\cos(nt)dt \tag{3.3}$$

$$b_n=\frac{2}{T}\int_{-T/2}^{T/2}f(t)\sin(nt)dt \tag{3.4}$$



รูปที่ 3.11 แสดงรูปร่างของสัญญาณ $f(t)$

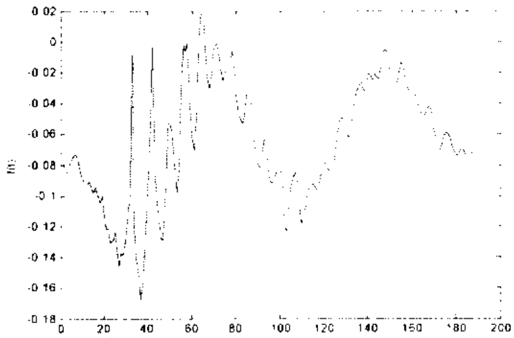


รูปที่ 3.12 แสดงองค์ประกอบของสัญญาณเสียง

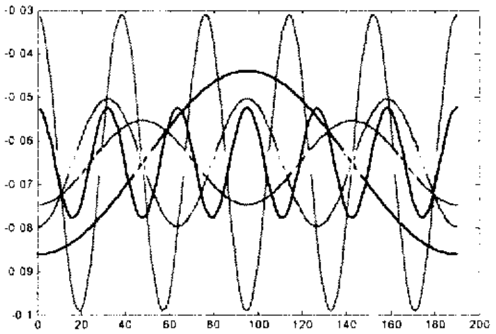
จากรูปที่ 3.11 และ 3.12 เมื่อนำสัญญาณโคไซน์ ณ ความถี่ และแอมพลิจูดต่าง ๆ ในรูปที่ 3.12 มาบวกกันจะได้สัญญาณ f(t) ในรูปที่ 3.11 ซึ่งแสดงให้เห็นว่าเราสามารถแสดง f(t) ให้อยู่ในรูปของการบวกกันของสัญญาณที่อยู่ในรูปที่ 3.12 ได้โดยไม่มีการสูญเสียคุณสมบัติสำคัญใด ๆ

โดยค่าสัมประสิทธิ์ของอนุกรมฟูเรียร์ดังกล่าวจะแสดงคุณสมบัติความถี่ของสัญญาณ f(t) ดังนั้นเราสามารถดึงคุณสมบัติทางความถี่ที่สำคัญของ f(t) ได้จากการกำหนดจำนวนสัมประสิทธิ์ของอนุกรมฟูเรียร์ที่เหมาะสมดังรูปที่ 3.13-3.16

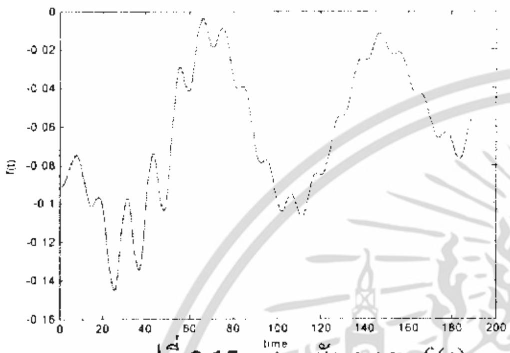
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



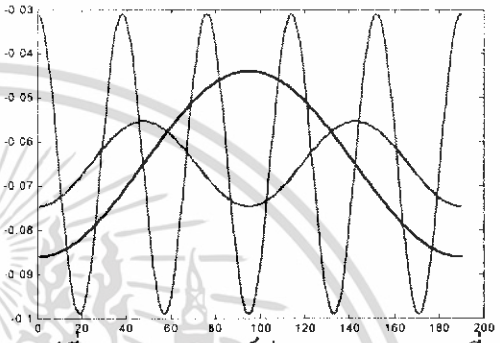
รูปที่ **3.13** แสดงรูปร่างของสัญญาณ $f(t)$



รูปที่ **3.14** แสดงองค์ประกอบทางความถี่



รูปที่ **3.15** แสดงสัญญาณ $f(t)$



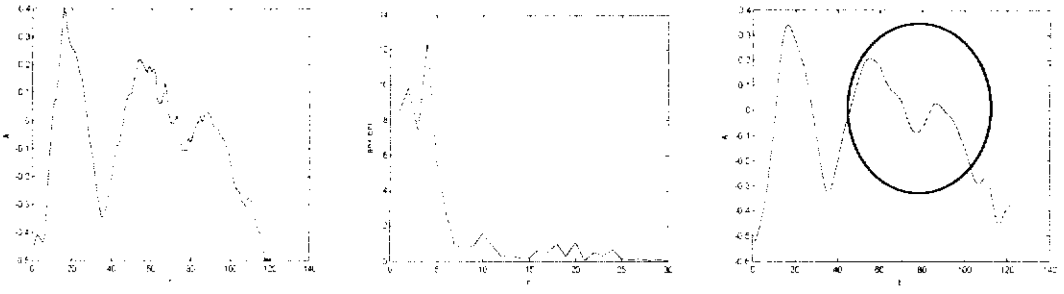
รูปที่ **3.16** แสดงองค์ประกอบทางความถี่

สามารถแสดงสัญญาณ f(t) ในรูปที่ 3.13 ได้ในรูปของการบวกกันของสัญญาณในรูปที่ 3.14 และสามารถแสดงสัญญาณ f(t) ในรูปที่ 3.15 ได้ในรูปของการบวกกันของสัญญาณในรูปที่ 3.16 ได้เช่นกัน

จะสังเกตได้ว่าสัญญาณ f(t) ในรูปที่ 3.13 และ 3.15 มีบางอย่างที่มีลักษณะคล้ายกัน และในรูปที่ 3.14 และ 3.16 จะพบว่าสัญญาณที่อยู่ในรูปที่ 3.16 เป็นสับเซ็ตของสัญญาณในรูปที่ 3.14 แต่ในรูปที่ 3.14 จะมีสัญญาณที่มีความถี่สูงเพิ่มขึ้นมาอีก ดังนั้นถ้าตัดสัญญาณความถี่สูงที่เพิ่มมาในรูปที่ 3.14 ออกไป ผลรวมของสัญญาณในรูปที่ 3.14 จะมีค่าเท่ากับสัญญาณในรูปที่ 3.15

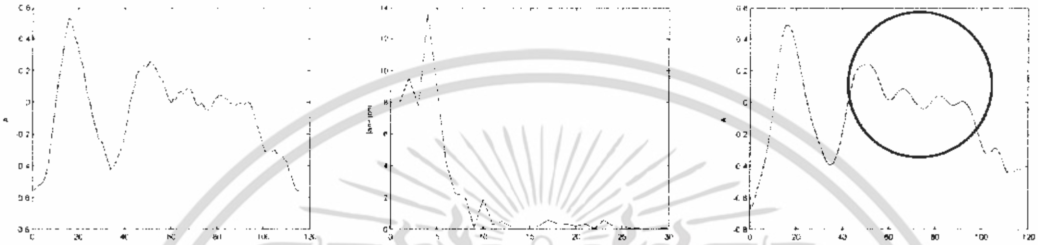
จากจุดนี้จึงนำไปประยุกต์ใช้สำหรับการดึงลักษณะเด่นของข้อมูลโดยพิจารณาจากคุณสมบัติทางความถี่ได้ โดยเราอาจตั้งสมมติฐานได้ว่า สัญญาณ f(t) ในรูปที่ 3.13 และ 3.15 เป็นรูปเดียวกัน แต่สัญญาณในรูปที่ 3.13 ถูกสิ่งมารบกวนทำให้มีความผิดเพี้ยนไป ดังนั้นลักษณะเด่นซึ่งเป็นลักษณะที่สำคัญที่คงที่ไม่เปลี่ยนแปลงของสัญญาณ f(t) ในรูปที่ 3.13 และ 3.15 คือสัญญาณรูปที่ 3.16 นั่นเอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



(ก) (ข) (ค)

รูปที่ **3.17** แสดงการประมาณค่าของสัญญาณเสียงเปิด



(ก) (ข) (ค)

รูปที่ **3.18** แสดงการประมาณค่าของสัญญาณเสียงเปิด

จากรูปที่ 3.17 และ 3.18 เป็นการแสดงการประมาณค่าของสัญญาณเสียงเปิด โดยในรูปย่อย (ก) เป็นรูปสัญญาณเดิม รูปย่อย (ข) เป็นรูปค่าสัมบูรณ์ของสัมประสิทธิ์อนุกรมฟูเรียร์ และรูปย่อย (ค) เป็นรูปสัญญาณแปลงกลับที่ได้จากการประมาณค่า โดยในที่นี้ใช้จำนวนค่าสัมประสิทธิ์อนุกรมฟูเรียร์จำนวน 30 ค่าในการแปลงสัญญาณย้อนกลับ จากรูปจะสังเกตได้ว่ารูปย่อย (ค) มีรูปร่างที่เรียบขึ้นเมื่อเทียบกับรูปย่อย (ก)

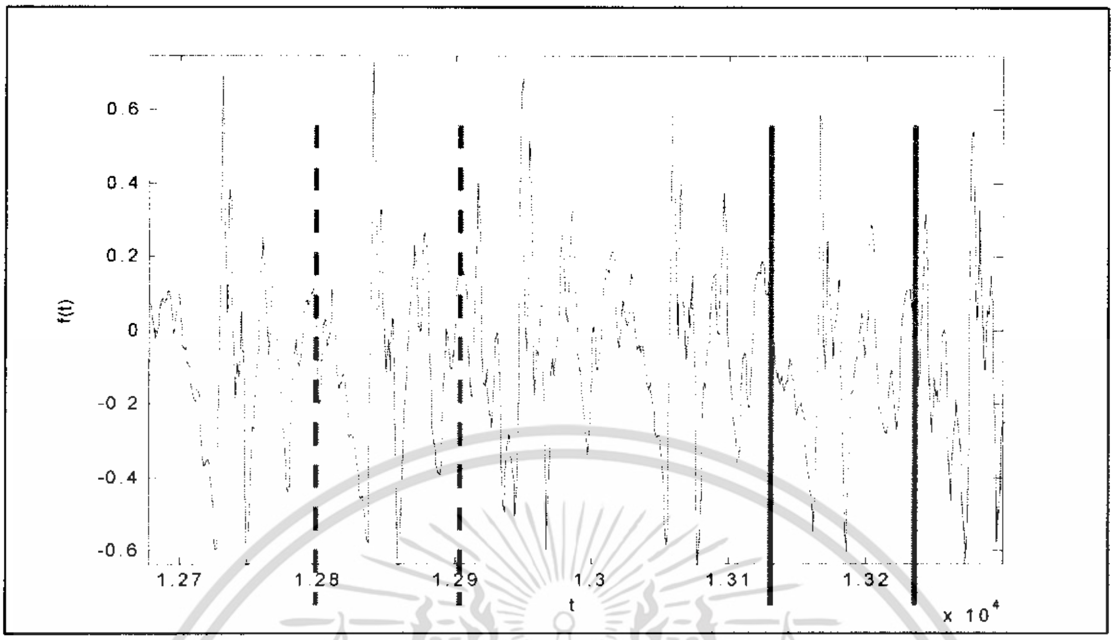
จากคุณสมบัตินี้จะนำมาวิเคราะห์สัญญาณเสียงเพื่อดึงเอาลักษณะเด่นที่สำคัญของเสียงแต่ละคำ ซึ่งลักษณะเด่นนี้จะแสดงอยู่ในรูปของค่าสัมประสิทธิ์ฟูเรียร์ โดยใช้การประมาณค่าในการกำหนดจำนวนสัมประสิทธิ์ของอนุกรมฟูเรียร์ เพื่อช่วยลดค่าความเบี่ยงเบนของสัญญาณ ซึ่งจำนวนสัมประสิทธิ์ของอนุกรมฟูเรียร์ที่เหมาะสมจะพิจารณาจากข้อมูลอ้างอิงของเสียงนั้น ๆ

3.3.2 การใช้ค่าสัมบูรณ์ของสัมประสิทธิ์ฟูเรียร์เป็นลักษณะเด่นของเสียง

จากรูปที่ 3.4 3.5 และ 3.6 จะเห็นได้ว่าสัญญาณมีลักษณะเป็นรายคาบ จึงทำการตัดเฉพาะบางส่วนของสัญญาณโดยใช้อัลกอริทึม (algorithm) ในการหาคาบ แต่พบปัญหาคือจุดเริ่มต้นในการตัดสัญญาณมีลักษณะเป็นการสุ่ม ไม่สามารถกำหนดตำแหน่งที่แน่นอนได้ ดังนั้นถ้าตำแหน่งที่สุ่มตัดมามีค่าไม่สอดคล้องกับข้อมูลอ้างอิงแล้ว จะทำให้การเปรียบเทียบเสียงอ้างอิงและเสียงทดสอบเกิดความผิดพลาดได้

เมื่อนำสัญญาณเสียงหนึ่งตัดมาดังรูปที่ 3.19 และ 3.20

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ **3.19** แสดงสัญญาณเสียง

รูปที่ **3.20** แสดงค่าสัมบูรณ์ของสัมประสิทธิ์ฟูเรียร์ของสองช่วงเสียง

จากรูปเมื่อตัดเสียงออกมาช่วงหนึ่ง พร้อมกับนำมาหาค่าสัมบูรณ์ของสัมประสิทธิ์ฟูเรียร์ จะได้ดังรูปที่ 3.19 และ 3.20 จากนั้นตัดอีกช่วงหนึ่งของเสียงออกมา จะพบว่าเสียงทั้งสองช่วงมีค่าสัมบูรณ์ของสัมประสิทธิ์ฟูเรียร์ที่มีค่าใกล้กันมาก จึงพบว่าค่าสัมบูรณ์ของสัมประสิทธิ์ฟูเรียร์มีคุณสมบัติที่ไม่เปลี่ยนแปลงตามการเลื่อนของสัญญาณ (translation invariant) ดังนั้นเพื่อให้ได้ประสิทธิภาพในการรู้จำสูง ปริญญานิพนธ์ฉบับนี้จึงใช้ค่าสัมบูรณ์ของสัมประสิทธ์ฟูเรียร์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.4 การรู้จำเสียง (Voice Recognition)

เป็นขั้นตอนที่ให้ระบบทำการเรียนรู้โดยการนำสัญญาณเสียงเข้าสู่ระบบ เพื่อระบบจะทำการตัดสินใจ และให้ผลลัพธ์ตามสัญญาณเสียงที่แตกต่างกันได้ถูกต้อง จะประกอบด้วย 3 ส่วนย่อย ได้แก่

1. เทคนิคที่ใช้ในการเปรียบเทียบ (Pattern matching) ซึ่งในโครงงานนี้จะใช้วิธีการวัดระยะยูคลิด

2. เสียงอ้างอิงที่ใช้ในการเปรียบเทียบ (Reference template) มีเสียงปิด เปิด ซ้าย ขวา และหยุด

3.กระบวนการการตัดสินใจของระบบ (Decision making) ในปริญญานิพนธ์ฉบับนี้ได้ทำการเปรียบเทียบสัญญาณเสียงโดยใช้ค่าสัมบูรณ์ของสัมประสิทธิ์ฟูเรียร์ ที่ได้จากรูปร่างของสัญญาณเสียงและอัตราการเปลี่ยนแปลงรูปร่างของสัญญาณเสียง

โดยกำหนดให้กลุ่มของเสียงอ้างอิงมีเสียงมีเสียงปิด เปิด ซ้าย ขวา และหยุด ซึ่งสามารถเขียนแทนด้วย $\{C_0, C_1, C_2, ..., C_4\}$ ตามลำดับ

$$C_l \in \{C_0, C_1, C_2, ..., C_4\} \tag{3.5}$$

$$S_{min}(C_l, x) = \min\{S(C_l, x)\}, l = 0,1,2,...,4 \tag{3.6}$$

$S_{min}(C_l, x)$ คือความแตกต่างระหว่างเวกเตอร์เสียงที่นำเข้ามาทดสอบ (x) และเวกเตอร์เสียงอ้างอิงที่อยู่ในกลุ่ม C_l โดยถ้าสัญญาณเสียงอ้างอิงใดมีค่าความแตกต่างน้อยที่สุด จะสรุปได้ว่าสัญญาณเสียงทดสอบนั้นเป็นเสียงเดียวกับเสียงอ้างอิงนั้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4

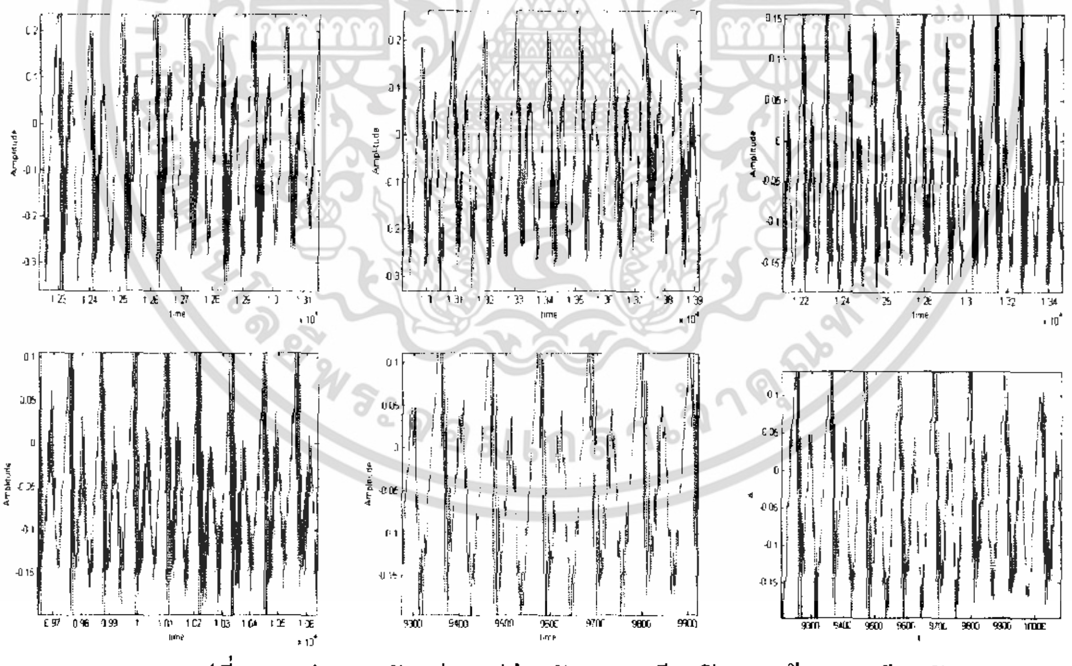
ผลการทดลอง

4.1 การเก็บข้อมูลเสียง

การเก็บข้อมูลในวิทยานิพนธ์นี้ จะใช้เสียงผู้พูดทั้งหมด 3 คน ซึ่งประกอบด้วยจำนวนคำทั้งหมด 5 คำ ได้แก่ ปิด เปิด ซ้าย ขวา และหยุด โดยจะให้ผู้ทดสอบแต่ละคนออกเสียงคำละ 21 ครั้ง ดังนั้นวิทยานิพนธ์นี้จะมีจำนวนคำที่ใช้ทั้งหมด 315 คำ ซึ่งจะนำไปสร้างแบบอ้างอิงจำนวน 15 คำ และสร้างแบบทดสอบจำนวน 300 คำ โดยในการบันทึกเสียง ข้อมูลจะถูกบันทึกอยู่ในรูปแบบของไฟล์ข้อมูล ".wav" ซึ่งข้อมูลหนึ่งตัวอย่างของเสียงจะถูกแทนด้วยข้อมูลขนาด 8 บิต และใช้ความถี่ในการชักตัวอย่างเท่ากับ 22 *kHz*

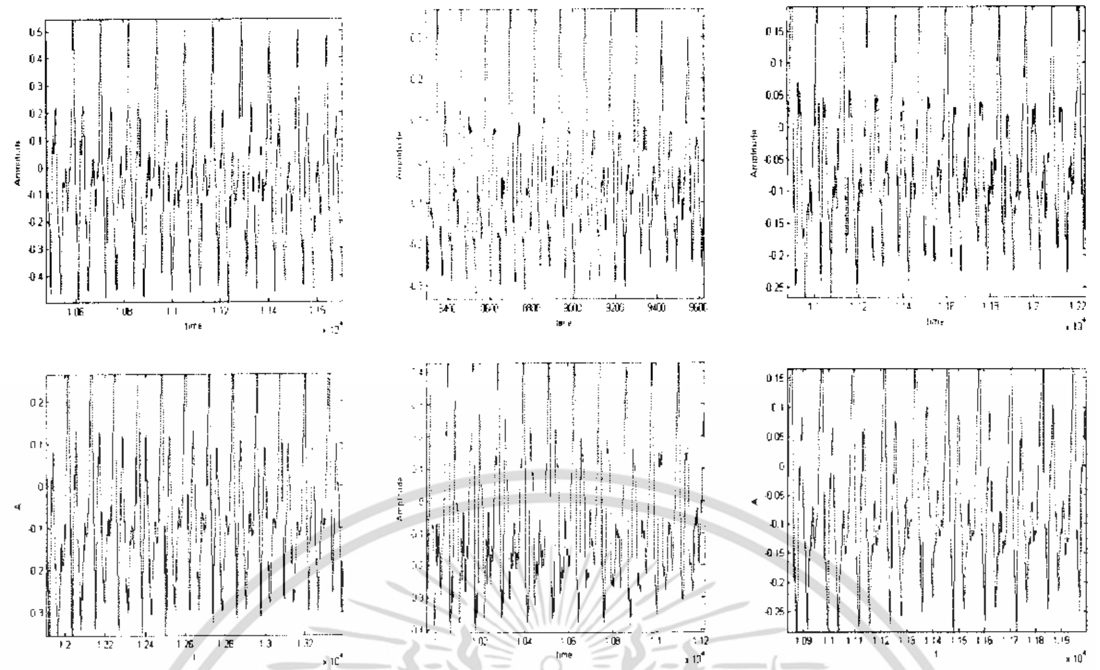
4.2 กระบวนการเตรียมเสียงก่อนนำไปทดสอบ

4.2.1 การตัดเสียงเงียบ หลังจากตัดเสียงเงียบแล้วจะเหลือแต่ส่วนที่เป็นสัญญาณเสียง และเมื่อพิจารณารูปร่างของสัญญาณเสียงจะพบว่าสัญญาณเสียงมีลักษณะซ้ำกันไปเรื่อย ๆ เป็นคาบดังรูป

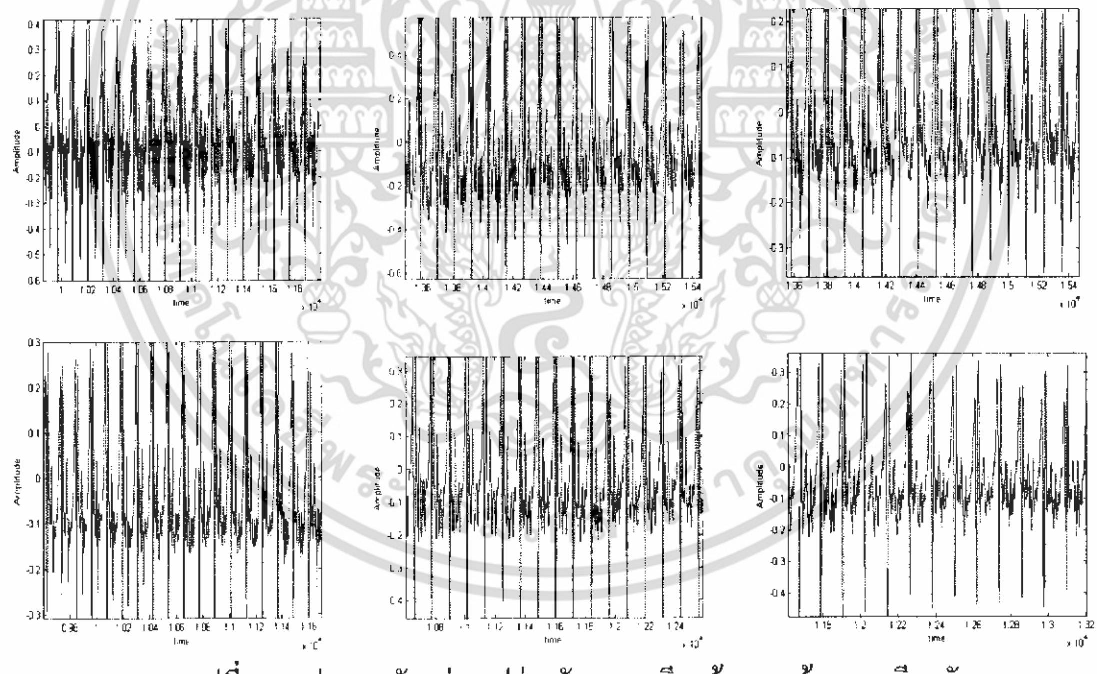


รูปที่ **4.1** รูปแสดงตัวอย่างรูปร่างสัญญาณเสียงปิดของผู้พูดคนเดียวกัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

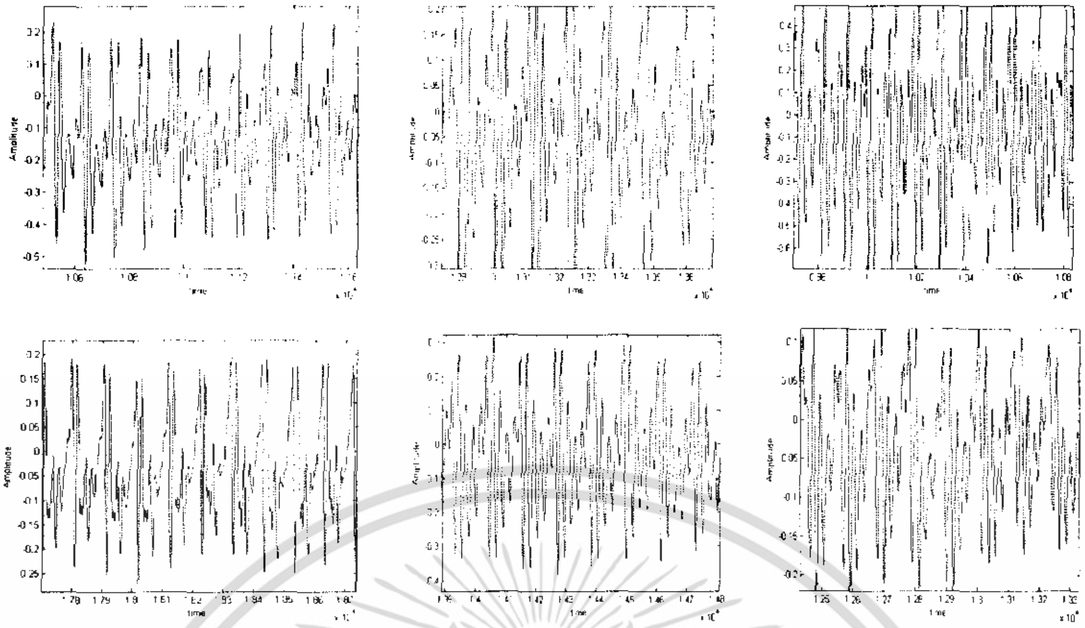


รูปที่ **4.2** รูปแสดงตัวอย่างรูปร่างสัญญาณเสียงเปิดของผู้พูดคนเดียวกัน

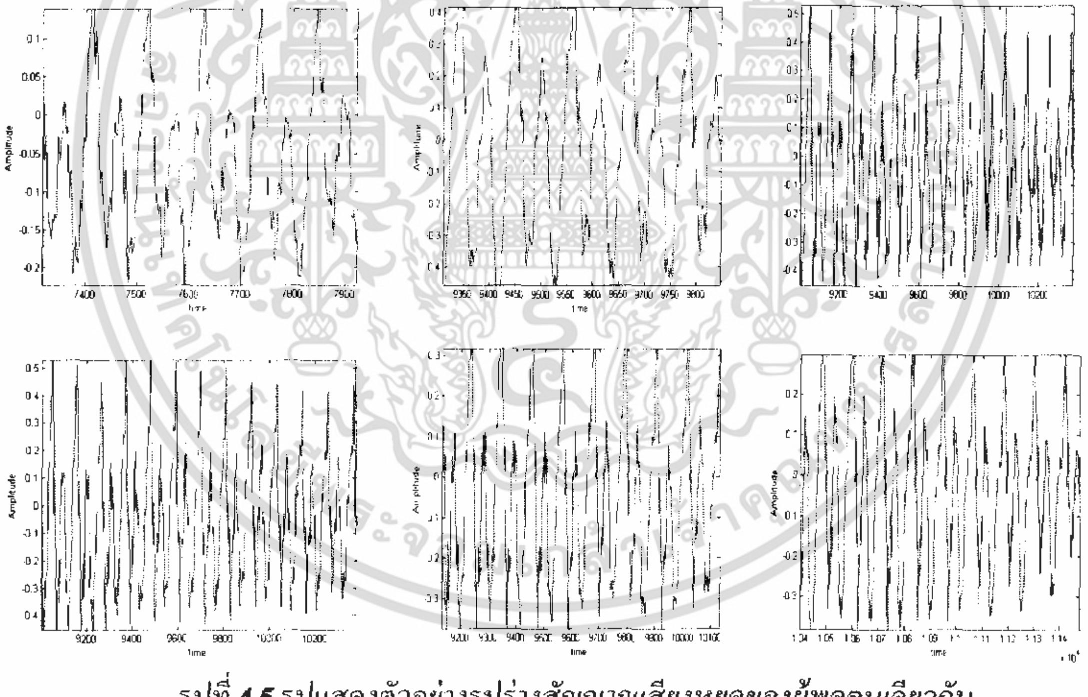


รูปที่ **4.3** รูปแสดงตัวอย่างรูปร่างสัญญาณเสียงซ้ายของผู้พูดคนเดียวกัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



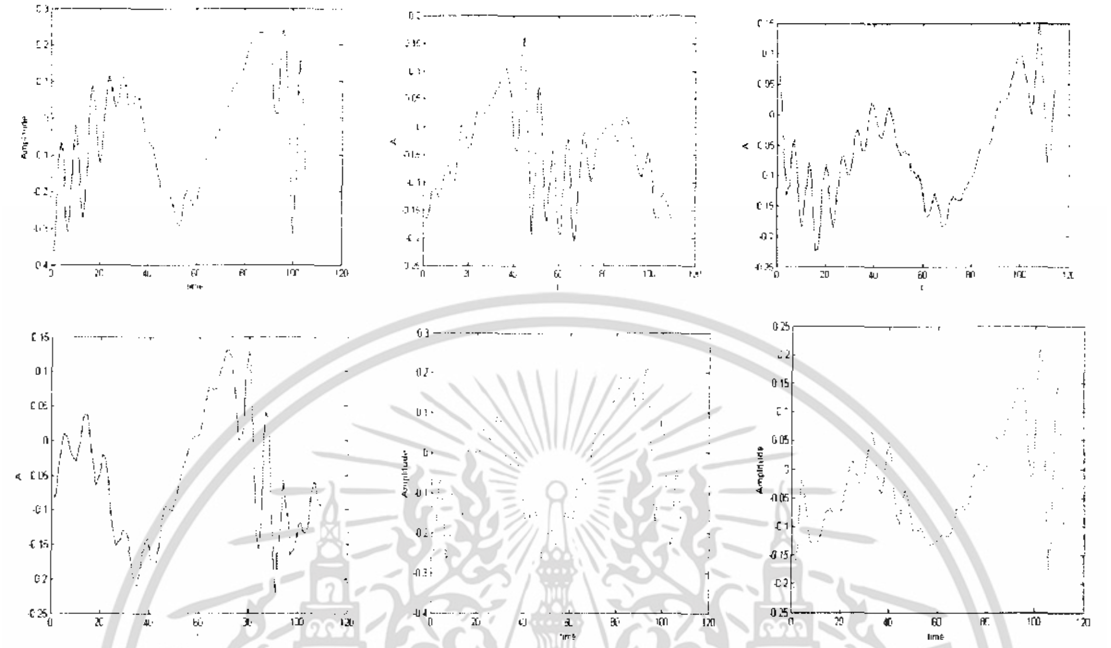
รูปที่ **4.4** รูปแสดงตัวอย่างรูปร่างสัญญาณเสียงขวาของผู้พูดคนเดียวกัน



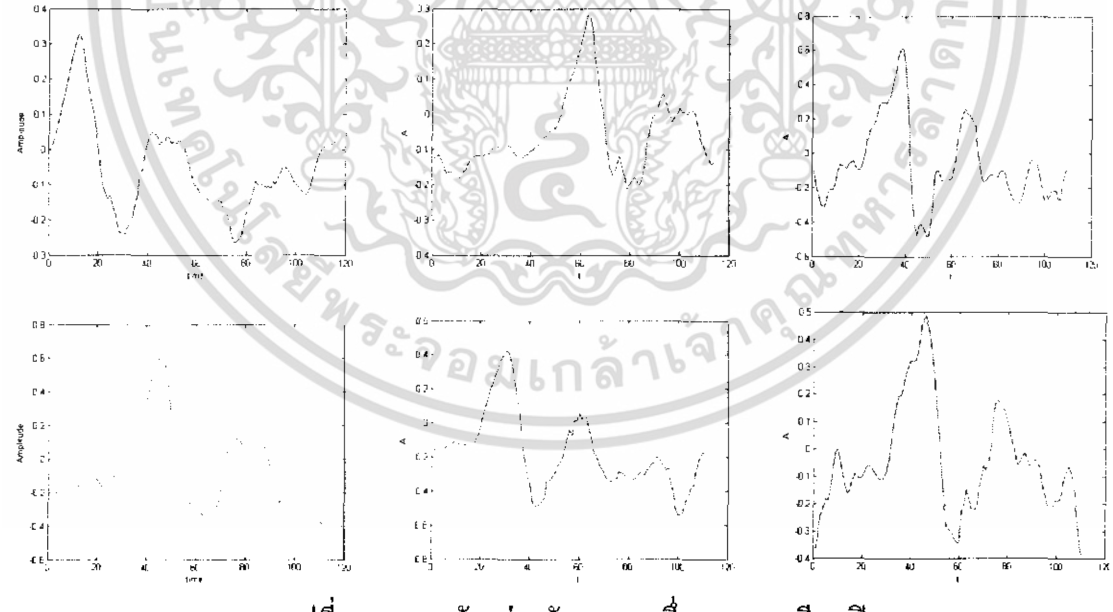
รูปที่ **4.5** รูปแสดงตัวอย่างรูปร่างสัญญาณเสียงหยุดของผู้พูดคนเดียวกัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.2 การหาคาบของเสียง เนื่องจากสัญญาณเสียงมีลักษณะเป็นรายคาบ ซึ่งในการนำเสียงมาทดสอบนั้นจะใช้เพียงหนึ่งคาบของเสียงเพื่อลดขั้นตอนการคำนวณ

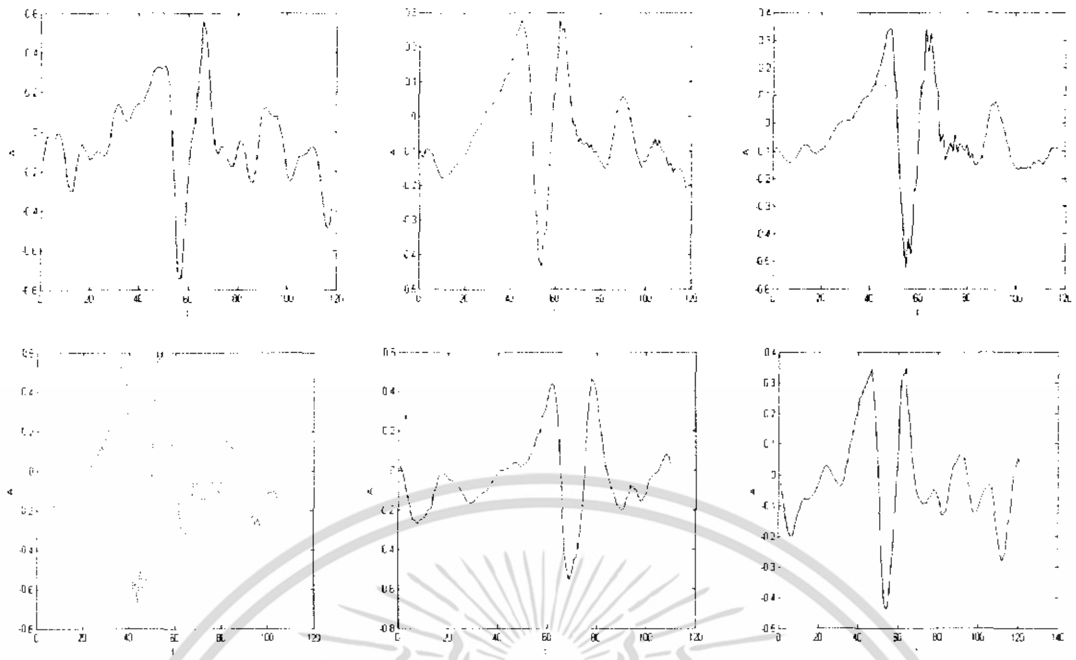


รูปที่ **4.6** แสดงตัวอย่างสัญญาณหนึ่งคาบของเสียงปิด

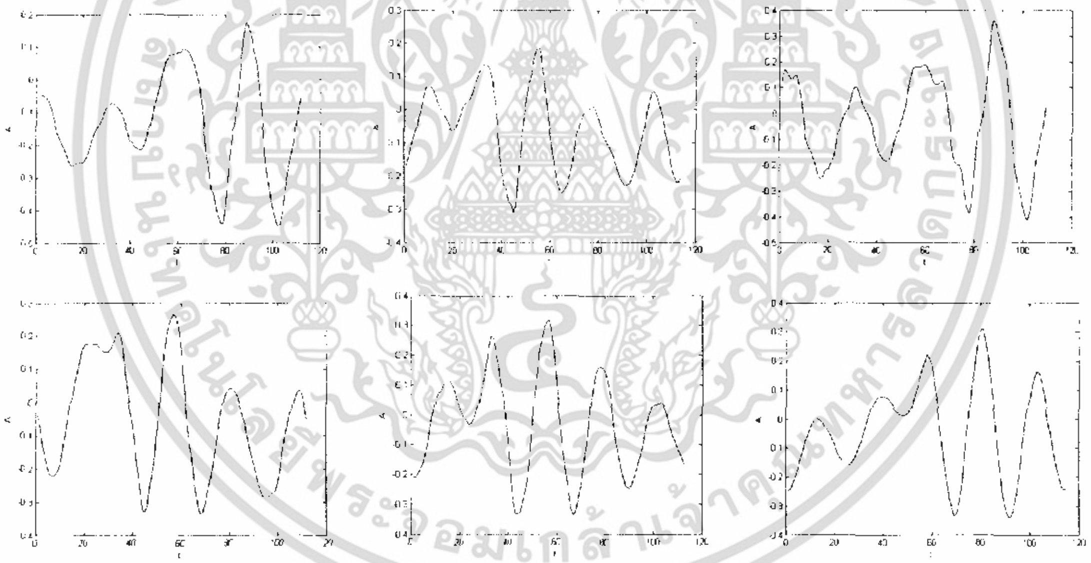


รูปที่ **4.7** แสดงตัวอย่างสัญญาณหนึ่งคาบของเสียงเปิด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

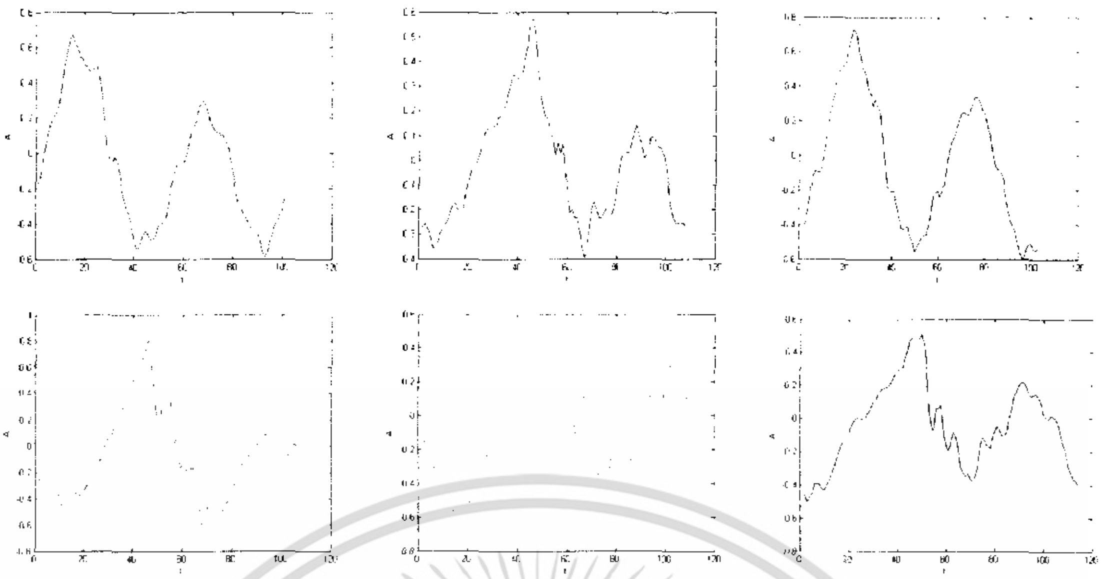


รูปที่ **4.8** แสดงตัวอย่างสัญญาณหนึ่งคาบของเสียงซ้าย



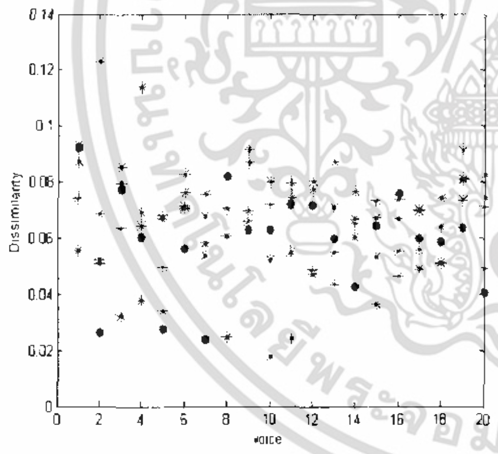
รูปที่ **4.9** แสดงตัวอย่างสัญญาณหนึ่งคาบของเสียงขวา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

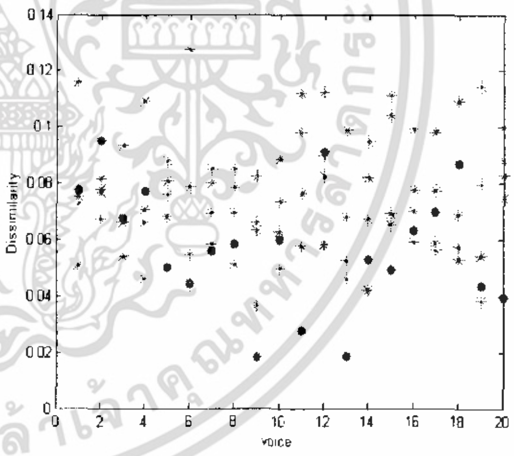


รูปที่ **4.10** แสดงตัวอย่างสัญญาณหนึ่งคาบของเสียงหยุด

เมื่อนำสัญญาณหนึ่งคาบที่ได้มาใช้วิธีการเปรียบเทียบโดยโปรแกรมแบบพลวัตโดยใช้เสียงอ้างอิงเป็นเสียงต่าง ๆ จะได้ผลดังนี้

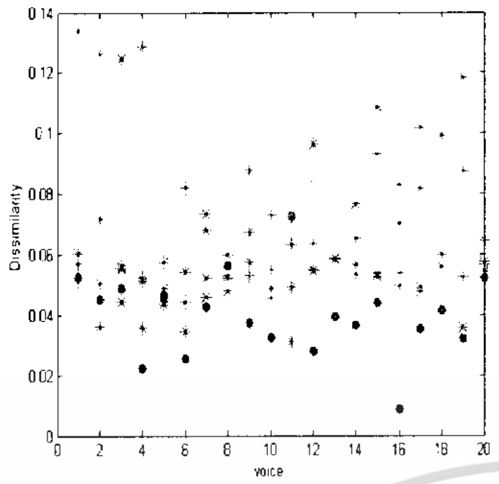


รูปที่ **4.11** เสียงอ้างอิงเสียง “เปิด”

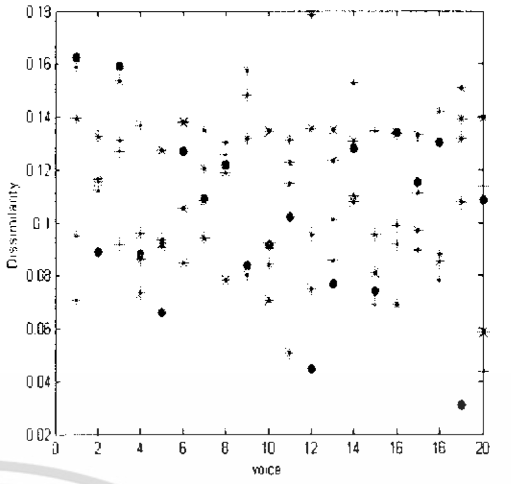


รูปที่ **4.12** เสียงอ้างอิงเสียง “ซ้าย”

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ **4.13** เสียงอ้างอิงเสียง “ขวา”

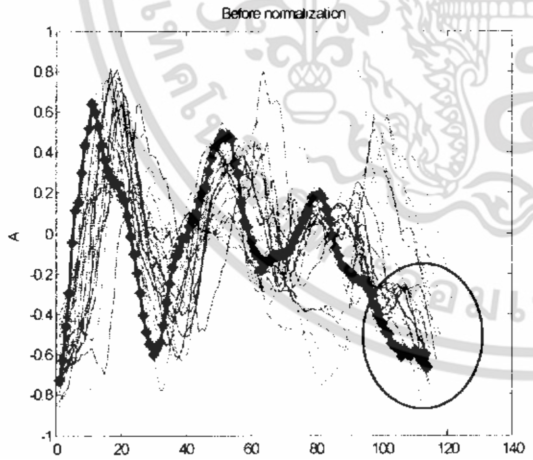


รูปที่ **4.14** เสียงอ้างอิงเสียง “หยุด”

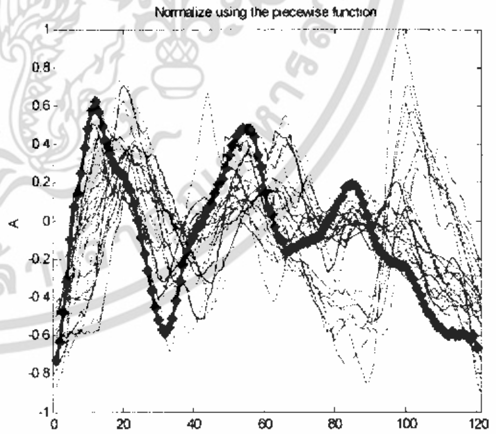
จากผลของการใช้การเปรียบเทียบโดยโปรแกรมแบบพลวัตจะเห็นได้ว่าไม่สามารถแยกเสียงอ้างอิงออกจากเสียงทดสอบได้ จึงเลือกใช้วิธีการดึงคุณลักษณะเด่นของเสียงโดยใช้อนุกรมฟูเรียร์ (Fourier Series) เข้ามาแทนเพื่อลดค่าความเบี่ยงเบนที่เกิดขึ้น

4.2.3 การชักตัวอย่างใหม่ด้วยฟังก์ชันพีซไวส์ลิเนียร์

จากภาพสัญญาณหนึ่งคาบจะสังเกตเห็นได้ว่า แต่ละเสียงมีคาบที่ไม่เท่ากัน ดังนั้นจึงต้องทำให้คาบเท่ากันก่อนนำไปเปรียบเทียบด้วยการชักตัวอย่างใหม่ที่เรียกว่าฟังก์ชันพีซไวส์ลิเนียร์

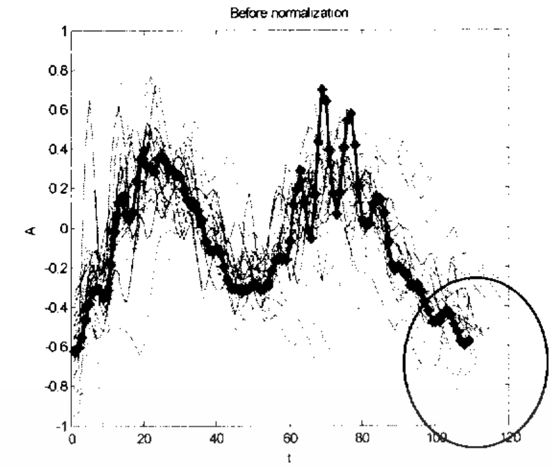


รูปที่ **4.15** สัญญาณเสียงเปิดก่อนการชักตัวอย่างใหม่

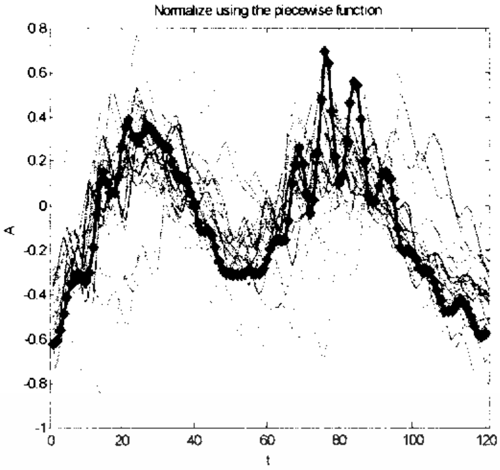


รูปที่ **4.16** สัญญาณเสียงเปิดหลังการชักตัวอย่างใหม่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

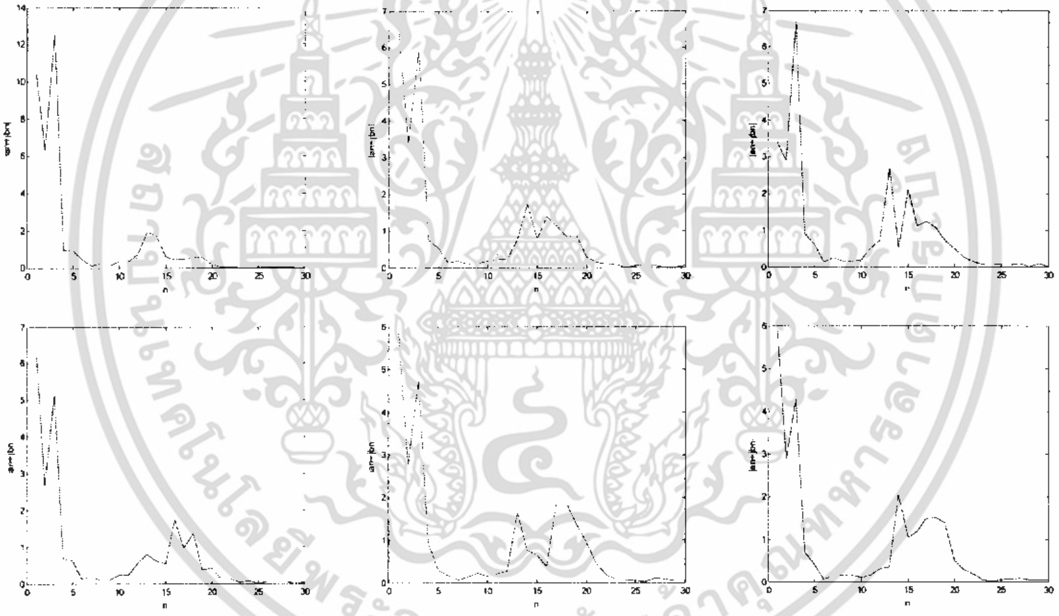


รูปที่ **4.17** สัญญาณเสียงปิดก่อนการชักตัวอย่างใหม่

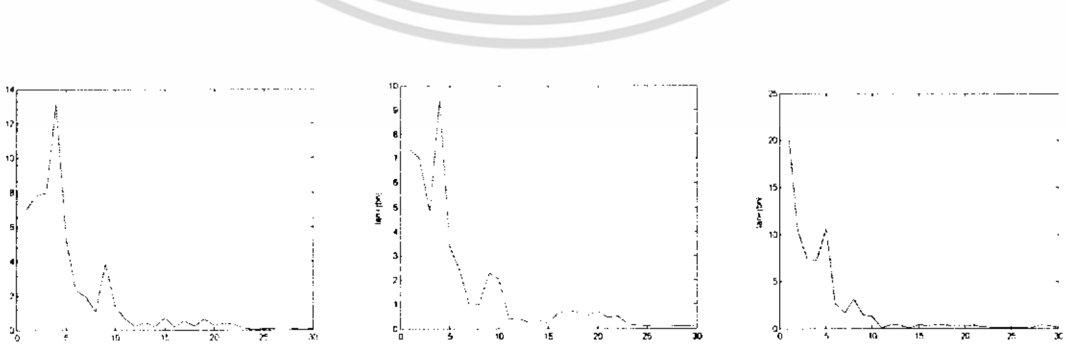


รูปที่ **4.18** สัญญาณเสียงปิดหลังการชักตัวอย่างใหม่

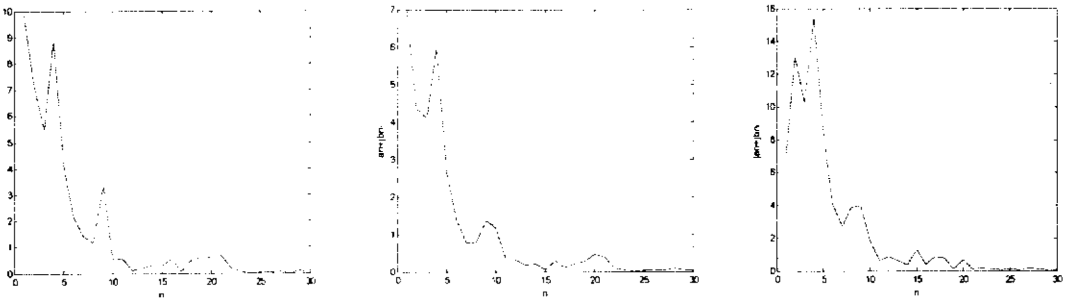
4.3 การดึงคุณลักษณะเด่นของเสียงโดยใช้ค่าสัมบูรณ์ของสัมประสิทธิ์ฟูเรียร์



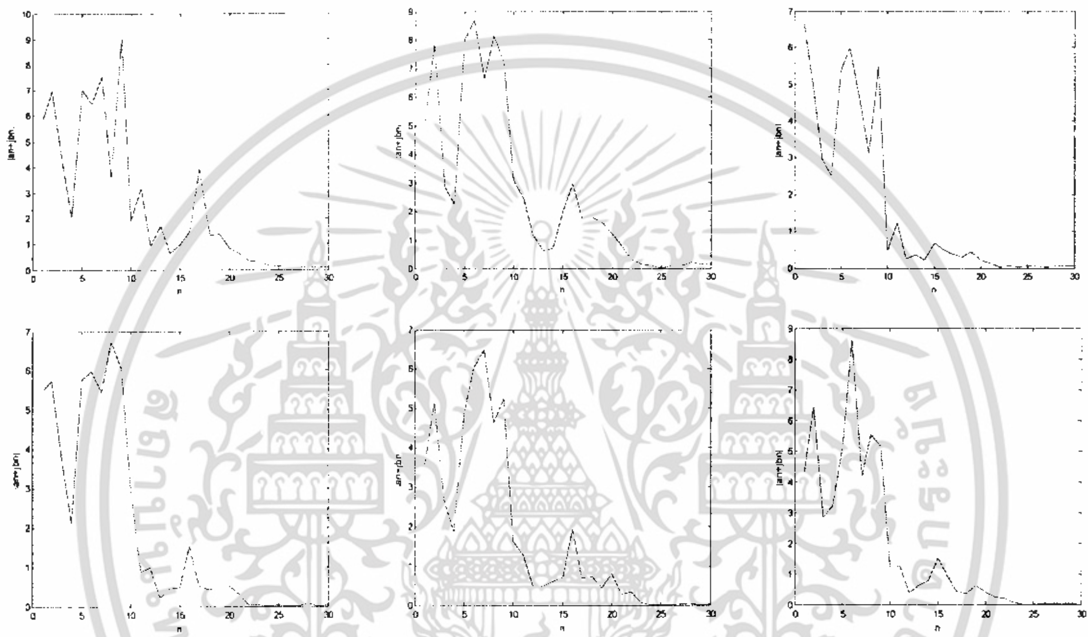
รูปที่ **4.19** แสดงตัวอย่างกราฟค่าสัมบูรณ์ของสัมประสิทธิ์ฟูเรียร์ของเสียงปิด



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

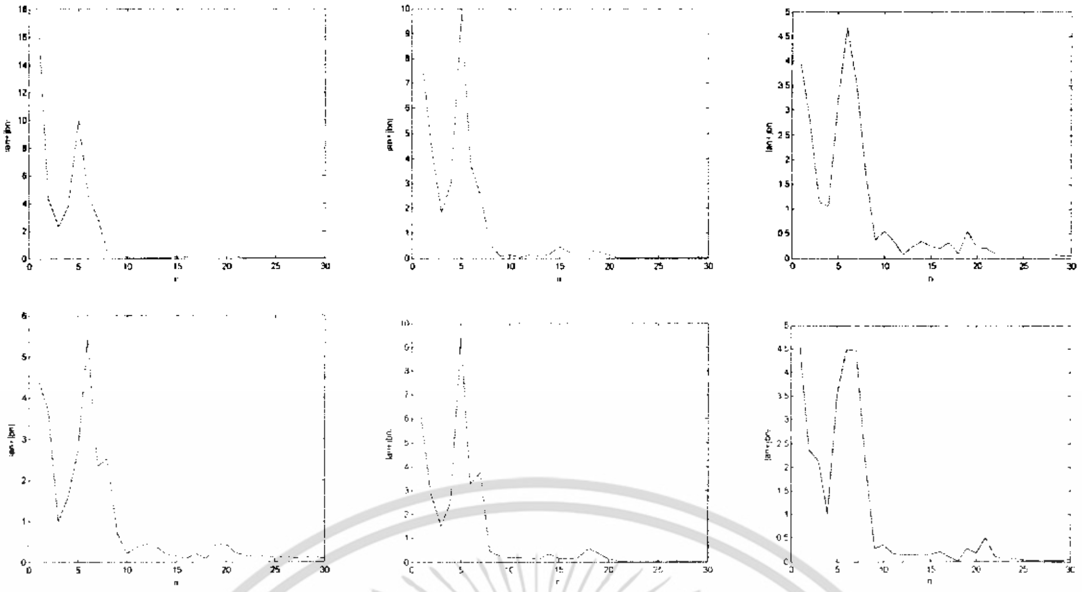


รูปที่ **4.20** แสดงตัวอย่างกราฟค่าสัมบูรณ์ของสัมประสิทธิ์ฟูเรียร์ของเสียงเปิด

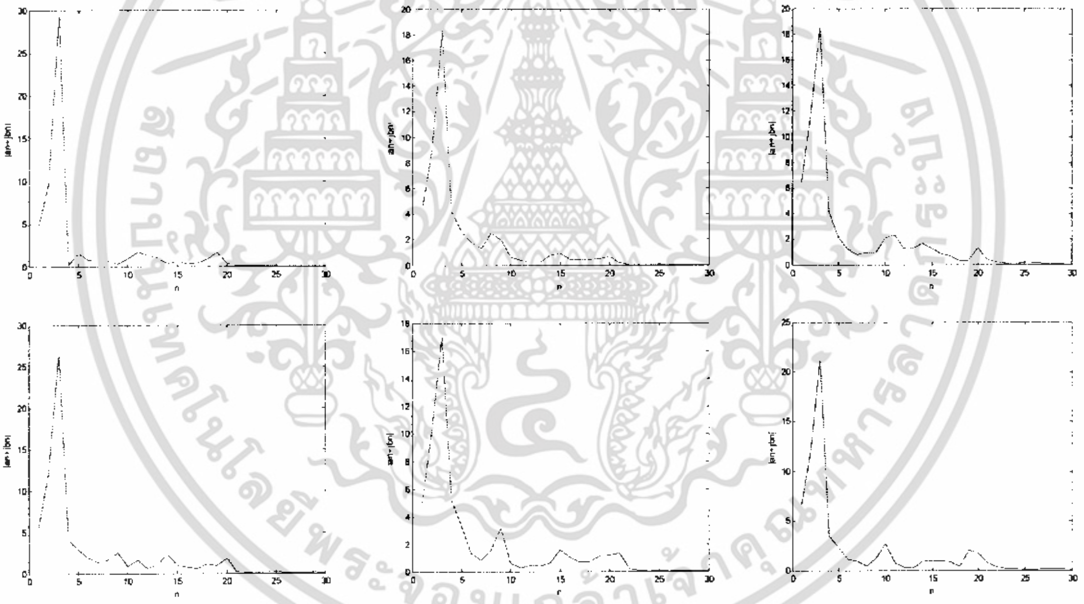


รูปที่ **4.21** แสดงตัวอย่างกราฟค่าสัมบูรณ์ของสัมประสิทธิ์ฟูเรียร์ของเสียงซ้าย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ **4.22** แสดงตัวอย่างกราฟค่าสัมบูรณ์ของสัมประสิทธิ์ฟูเรียร์ของเสียงขวา



รูปที่ **4.23** แสดงตัวอย่างกราฟค่าสัมบูรณ์ของสัมประสิทธิ์ฟูเรียร์ของเสียงหยุด

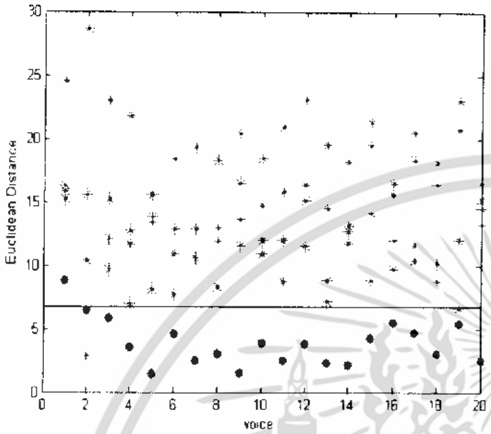
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.4 การรู้จำเสียง

ในการทดลองวิธีแรกจะใช้ค่าสัมบูรณ์ของสัมประสิทธิ์ฟูเรียร์ที่ได้จากรูปคลื่นสัญญาณเสียงเป็นลักษณะเด่นและหาระยะยูคลิดโดยใช้เสียงต่าง ๆ เป็นเสียงอ้างอิง

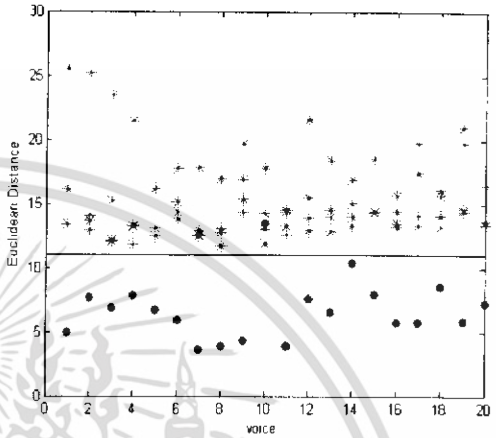
4.4.1 ใช้ค่าสัมบูรณ์ของสัมประสิทธิ์ฟูเรียร์ที่ได้จากรูปคลื่นสัญญาณเสียงเป็นลักษณะเด่น

ผู้พูดคนที่ 1



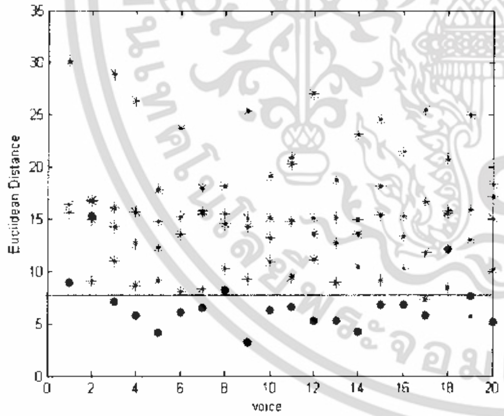
รูปที่ **4.24** ระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงปิด

ประสิทธิภาพ 97%



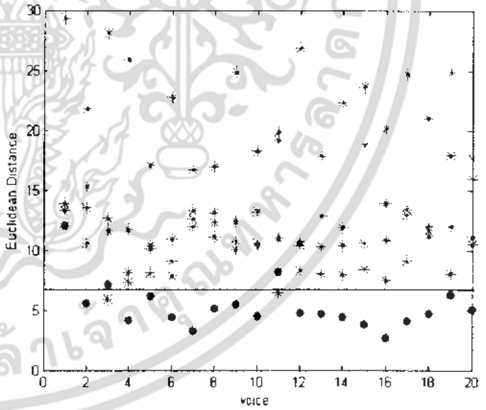
รูปที่ **4.25** ระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงเปิด

ประสิทธิภาพ 99%



รูปที่ **4.26** ระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงซ้าย

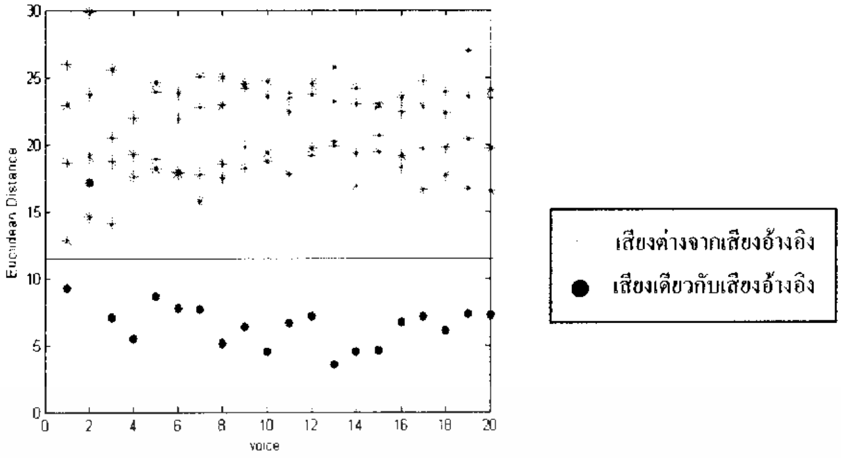
ประสิทธิภาพ 93%



รูปที่ **4.27** ระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงขวา

ประสิทธิภาพ 95%

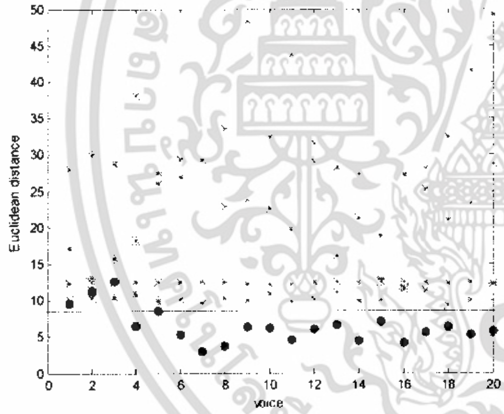
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

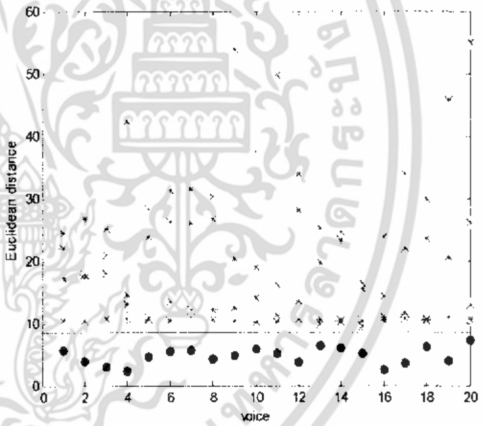


รูปที่ **4.28** ระยะยูคลิด โดยเสียงอ้างอิงเป็นเสียงหยุด

ประสิทธิภาพ 99%

ผู้พูดคนที่ 2





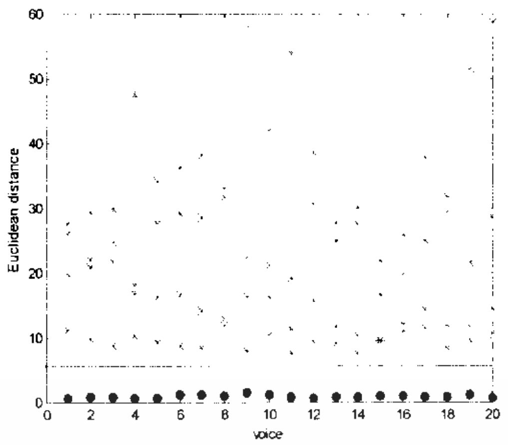
รูปที่ **4.29** ระยะยูคลิด โดยเสียงอ้างอิงเป็นเสียงปิด

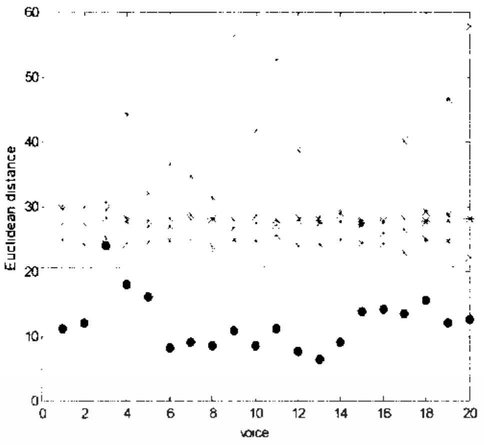
ประสิทธิภาพ 96%

รูปที่ **4.30** ระยะยูคลิด โดยเสียงอ้างอิงเป็นเสียงเปิด

ประสิทธิภาพ 100%

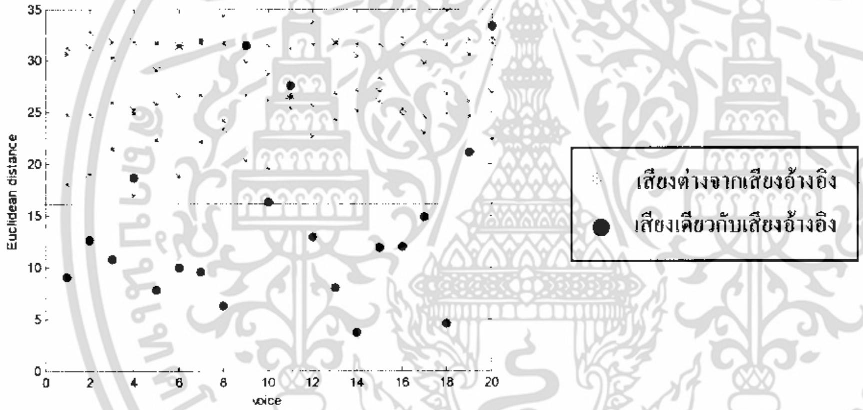
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้





รูปที่ **4.31** ระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงซ้าย รูปที่ **4.32** ระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงขวา

ประสิทธิภาพ 100% ประสิทธิภาพ 99%

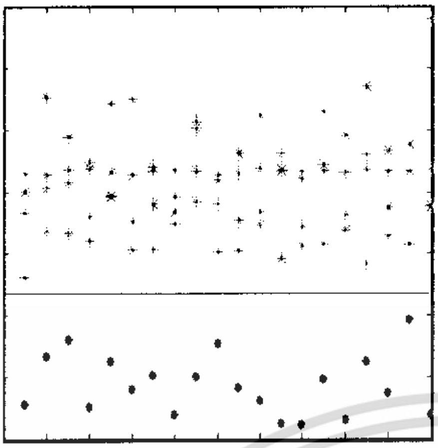


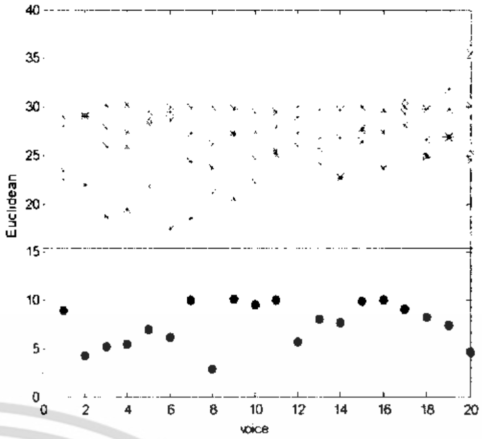
รูปที่ **4.33** ระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงหยุด

ประสิทธิภาพ 95%

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

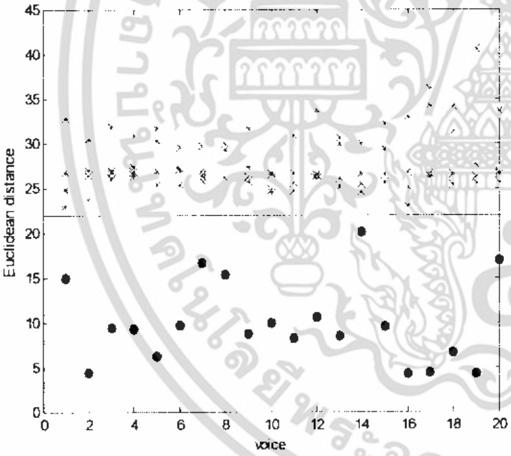
ผู้พูดคนที่ 3

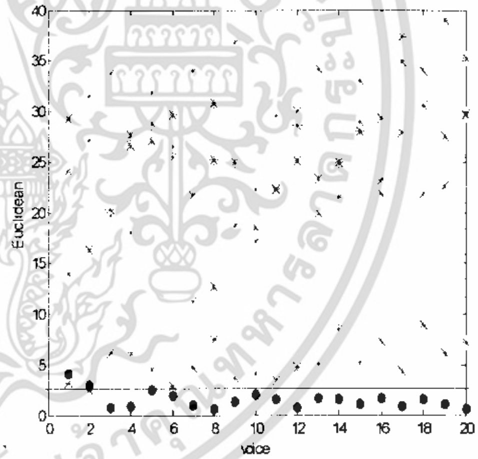




รูปที่ 4.34 ระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงปิด **รูปที่ 4.35** ระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงเปิด

ประสิทธิภาพ 100% ประสิทธิภาพ 100%

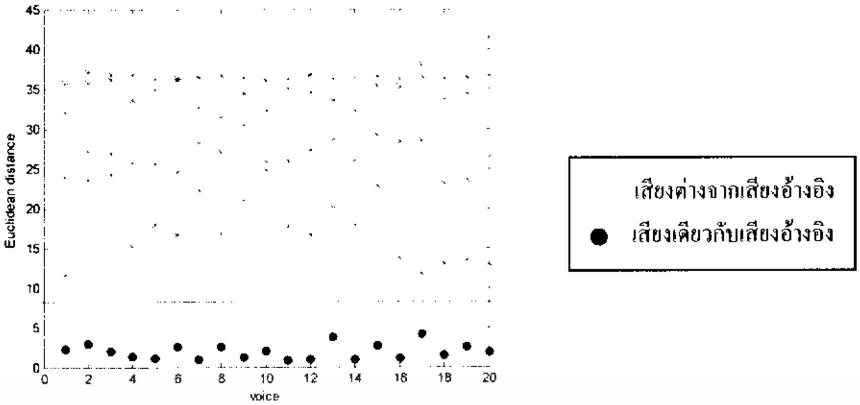




รูปที่ 4.36 ระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงซ้าย **รูปที่ 4.37** ระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงขวา

ประสิทธิภาพ 100% ประสิทธิภาพ 95%

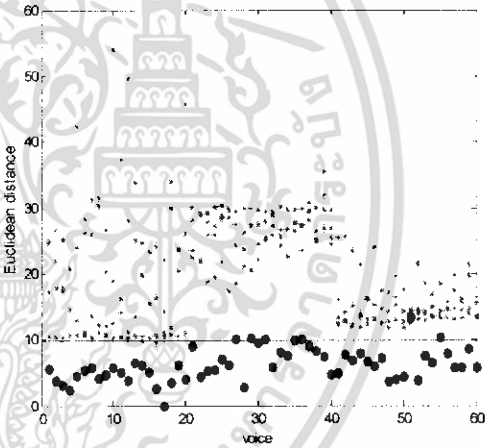
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ **4.38** ระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงหยุด

ประสิทธิภาพ 100%

ผู้พูดหลายคน



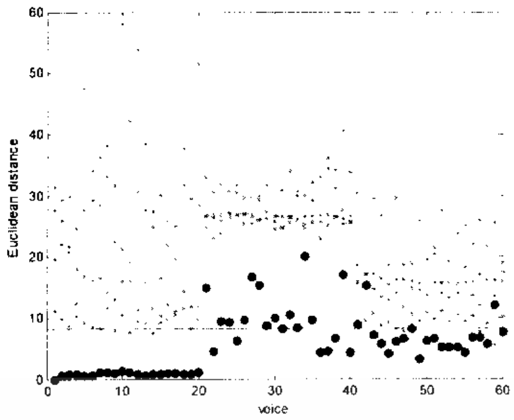
รูปที่ **4.39** ระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงปิด

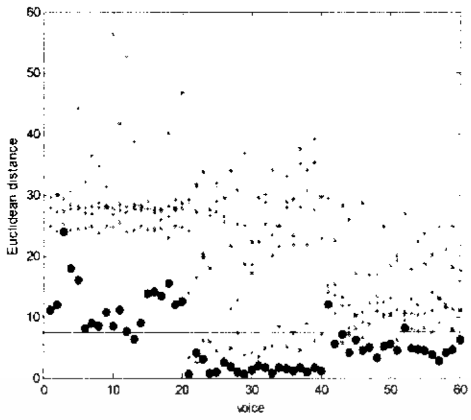
รูปที่ **4.40** ระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงเปิด

ประสิทธิภาพ = 88.67%

ประสิทธิภาพ = 97%

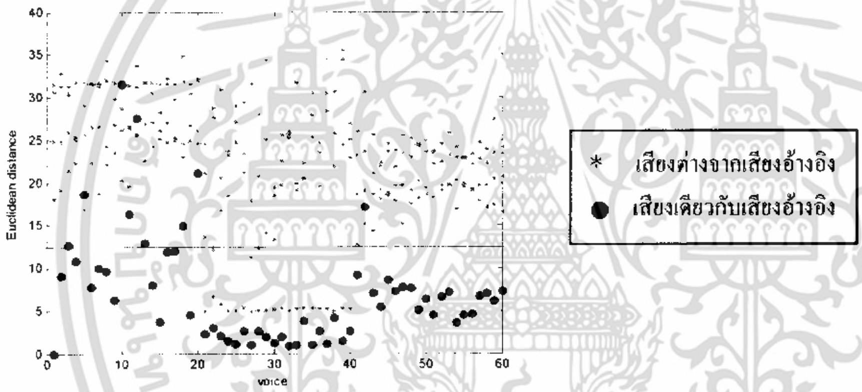
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้





รูปที่ **4.41** ระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงซ้าย รูปที่ **4.42** ระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงขวา

ประสิทธิภาพ = 91.33% ประสิทธิภาพ = 85%



รูปที่ **4.43** ระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงหยุด

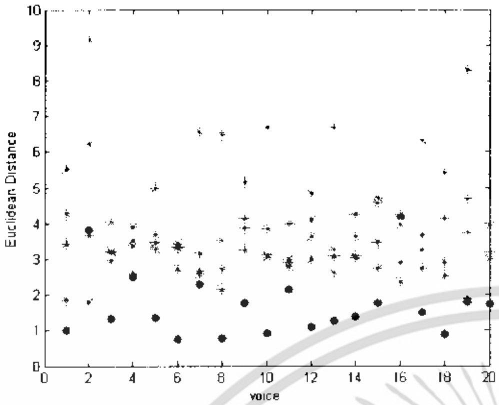
ประสิทธิภาพ = 89.33%

จากผลของการใช้ค่าสัมบูรณ์ของสัมประสิทธิ์ฟูเรียร์ที่ได้จากรูปคลื่นของสัญญาณเสียงเป็นลักษณะเด่นในการรู้จำ จะสังเกตเห็นได้จากกราฟระยะยูคลิดว่าสามารถแยกเสียงอ้างอิงออกจากเสียงทดสอบได้ดีในระดับที่น่าพึงพอใจ

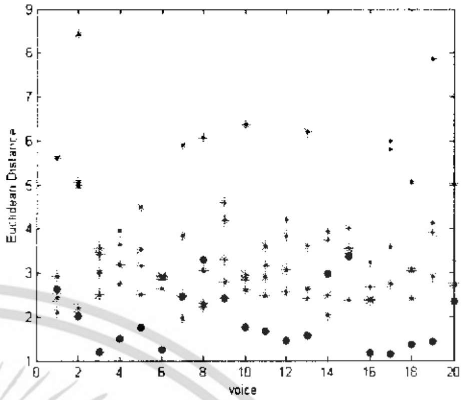
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.4.2 ใช้ค่าสัมบูรณ์ของสัมประสิทธิ์ฟูเรียร์ที่ได้จากอัตราการเปลี่ยนแปลงคลื่นเสียงตามเวลาเป็นลักษณะเด่น

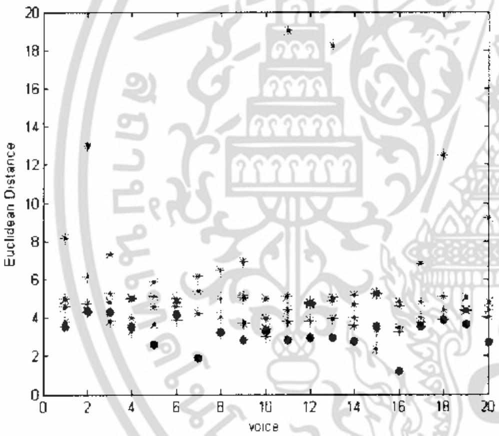
ผู้พูดคนที่ 1



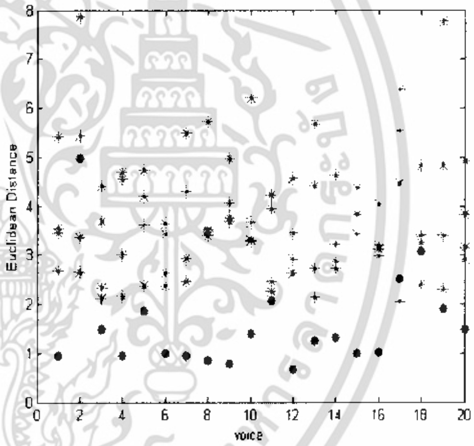
รูปที่ **4.44** ระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงปิด



รูปที่ **4.45** ระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงเปิด

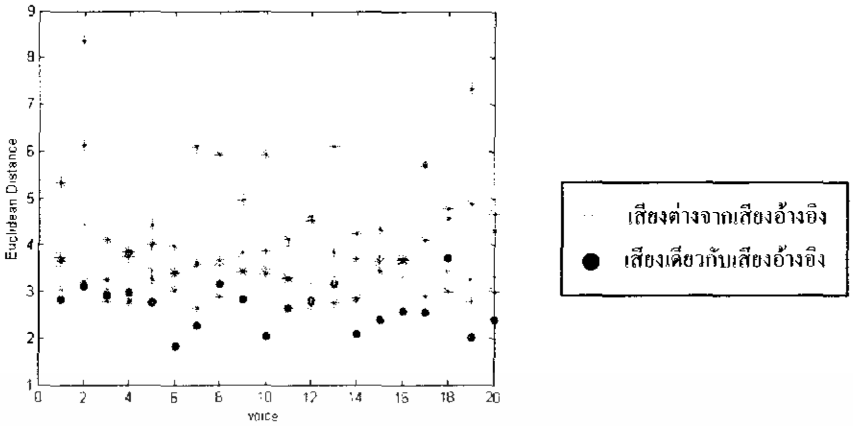


รูปที่ **4.46** ระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงซ้าย



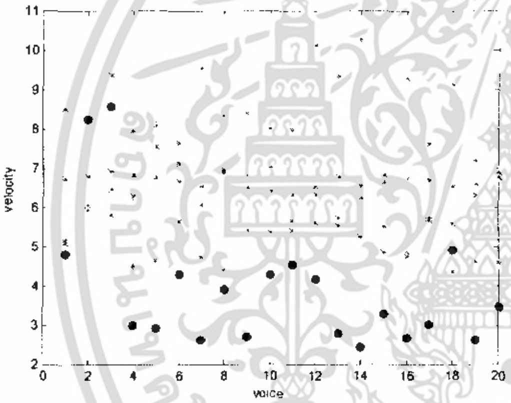
รูปที่ **4.47** ระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงขวา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

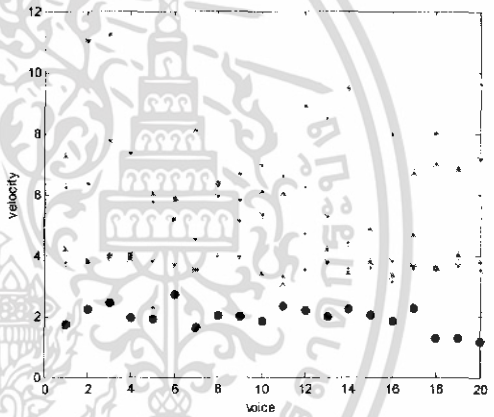


รูปที่ **4.48** ระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงหยุด

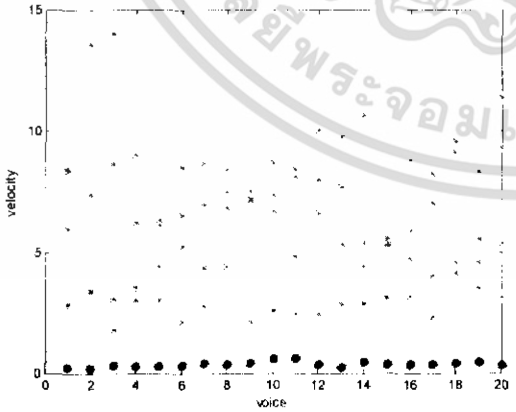
ผู้พูดคนที่ 2



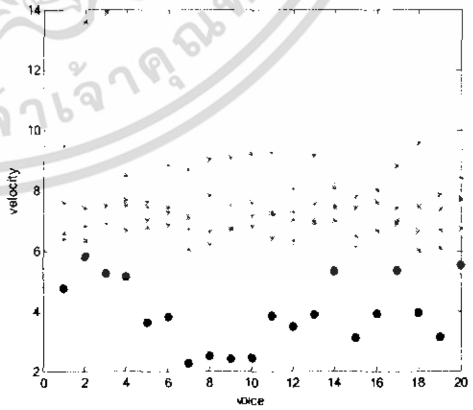
รูปที่ **4.49** ระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงปิด



รูปที่ **4.50** ระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงเปิด

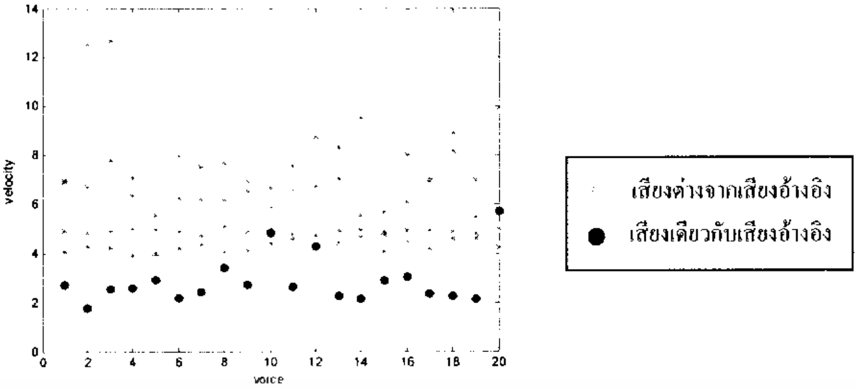


รูปที่ **4.51** ระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงซ้าย



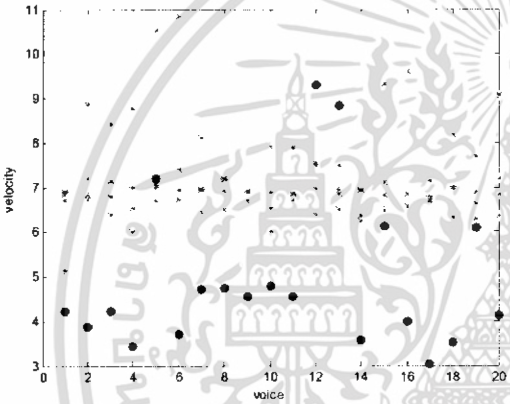
รูปที่ **4.52** ระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงขวา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

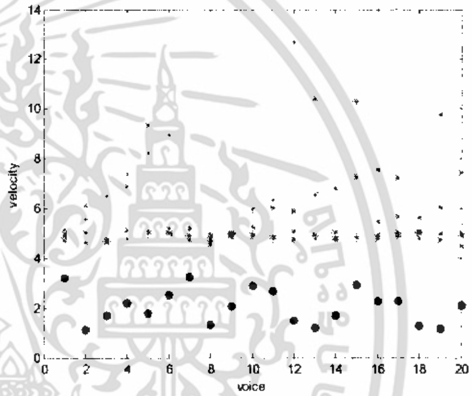


รูปที่ **4.53** ระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงหยุด

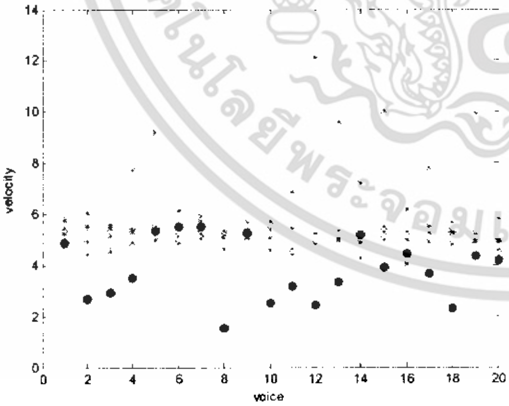
ผู้พูดคนที่ 3



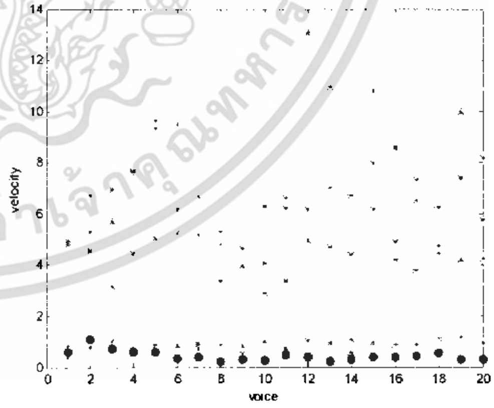
รูปที่ **4.54** ระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงปิด



รูปที่ **4.55** ระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงเปิด

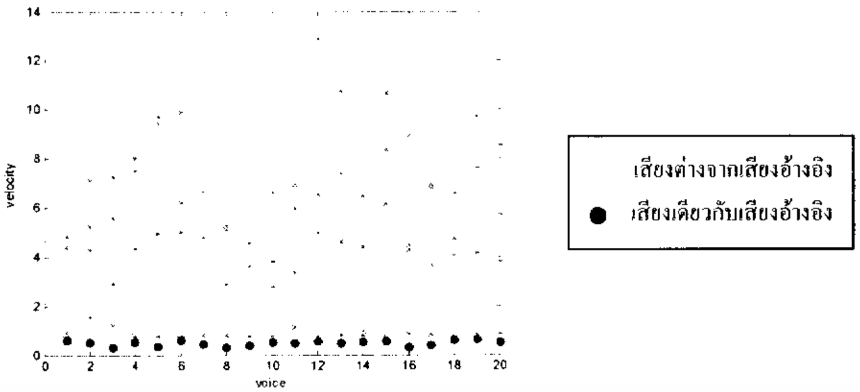


รูปที่ **4.56** ระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงซ้าย



รูปที่ **4.57** ระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงขวา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

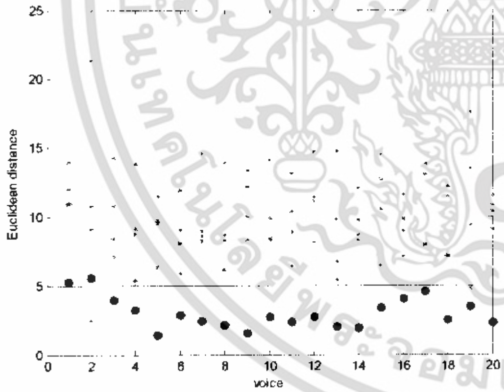


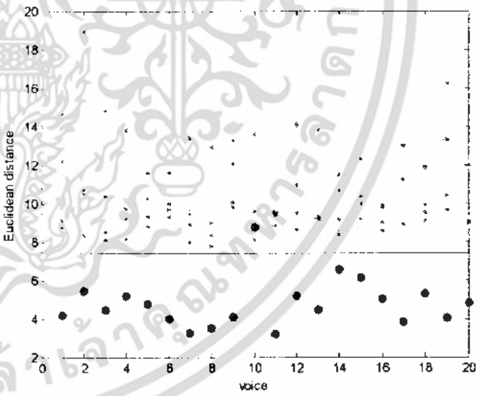
รูปที่ **4.58** ระยะยูคลิด โดยเสียงอ้างอิงเป็นเสียงหยุด

จากผลของการใช้ค่าสัมบูรณ์ของสัมประสิทธิ์ฟูเรียร์ที่ได้จากอัตราการเปลี่ยนแปลงคลื่นเสียงตามเวลาเป็นลักษณะเด่น จะเห็นได้ว่าสามารถแยกเสียงอ้างอิงออกจากเสียงทดสอบได้และเพื่อเพิ่มประสิทธิภาพในการแยกเสียงจึงนำอัตราการเปลี่ยนแปลงคลื่นเสียงตามเวลามาเป็นลักษณะเด่นร่วมด้วย

4.2.3 การใช้ค่าสัมบูรณ์ของสัมประสิทธิ์ฟูเรียร์ที่ได้จากรูปคลื่นของสัญญาณเสียงและอัตราการเปลี่ยนแปลงคลื่นเสียงตามเวลาเป็นลักษณะเด่น

ผู้พูดคนที่ 1

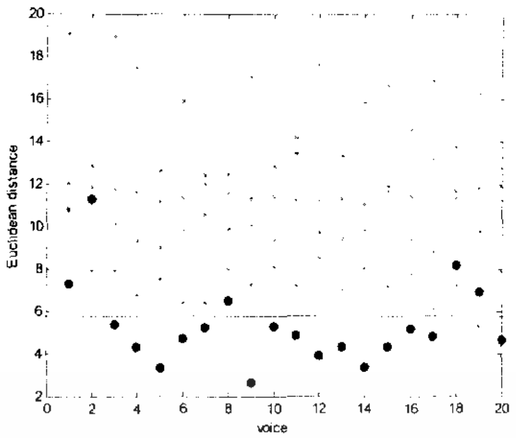




รูปที่ **4.59** ระยะยูคลิด โดยเสียงอ้างอิงเป็นเสียงปิด รูปที่ **4.60** ระยะยูคลิด โดยเสียงอ้างอิงเป็นเสียงเปิด

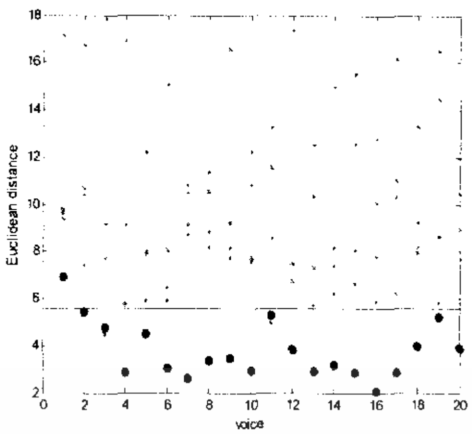
ประสิทธิภาพ = 98% ประสิทธิภาพ = 100%

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



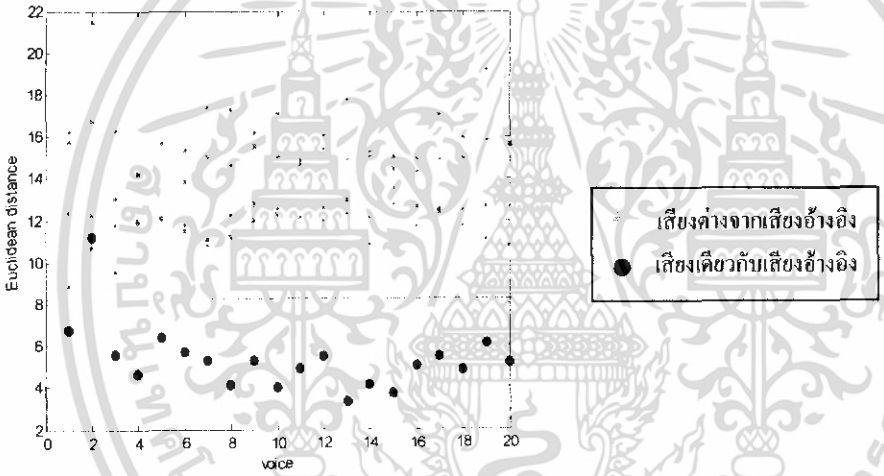
รูปที่ **4.61** ระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงซ้าย

ประสิทธิภาพ = 95%



รูปที่ **4.62** ระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงขวา

ประสิทธิภาพ = 98%

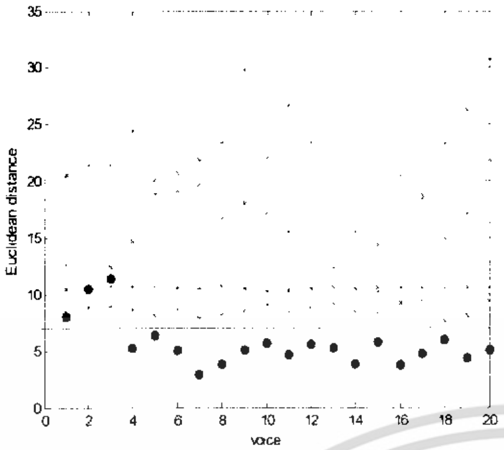


รูปที่ **4.63** ระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงหยุด

ประสิทธิภาพ = 99%

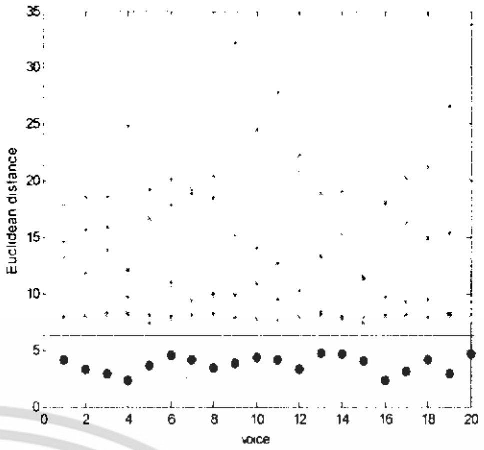
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ผู้พูดคนที่ 2



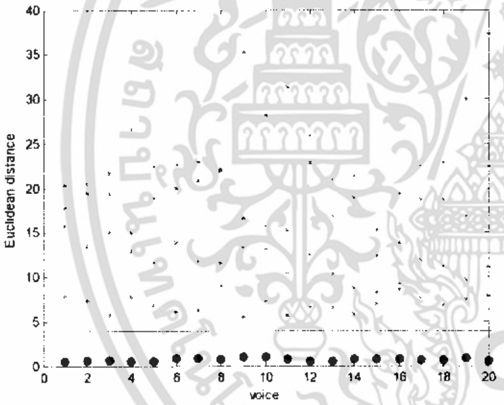
รูปที่ **4.64** ระยะยูคลิด โดยเสียงอ้างอิงเป็นเสียงปิด

ประสิทธิภาพ = 97%



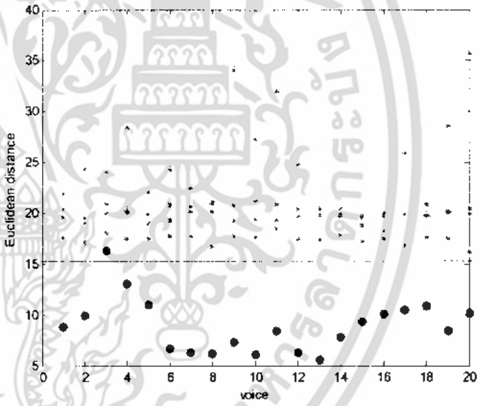
รูปที่ **4.65** ระยะยูคลิด โดยเสียงอ้างอิงเป็นเสียงเปิด

ประสิทธิภาพ = 100%



รูปที่ **4.66** ระยะยูคลิด โดยเสียงอ้างอิงเป็นเสียงซ้าย

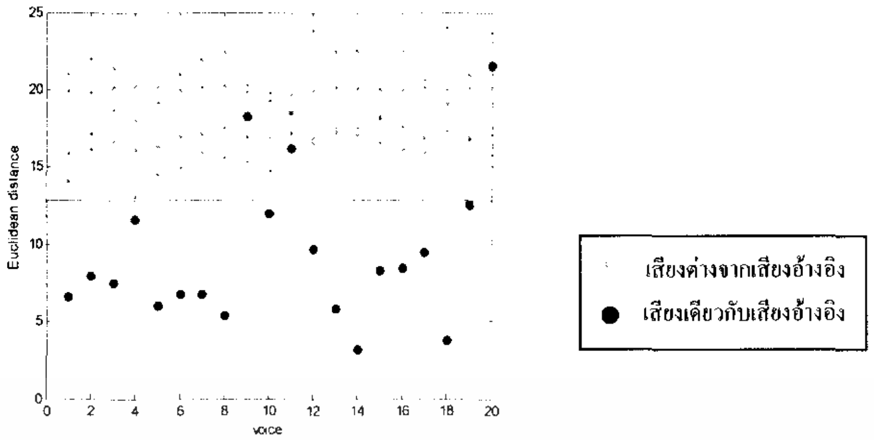
ประสิทธิภาพ = 100%



รูปที่ **4.67** ระยะยูคลิด โดยเสียงอ้างอิงเป็นเสียงขวา

ประสิทธิภาพ = 99%

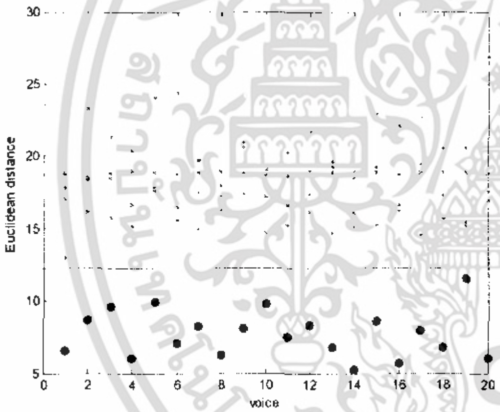
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

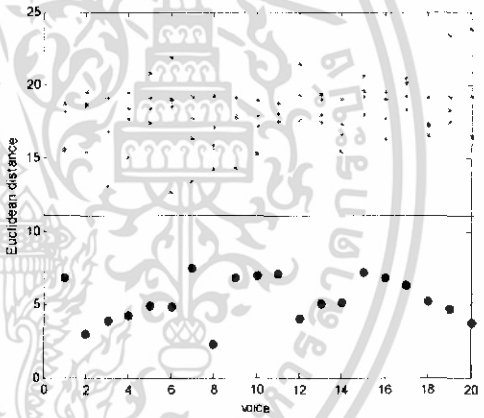


รูปที่ **4.68** ระยะยูคลิด โดยเสียงอ้างอิงเป็นเสียงหยุด

ประสิทธิภาพ = 97%

ผู้พูดคนที่ 3





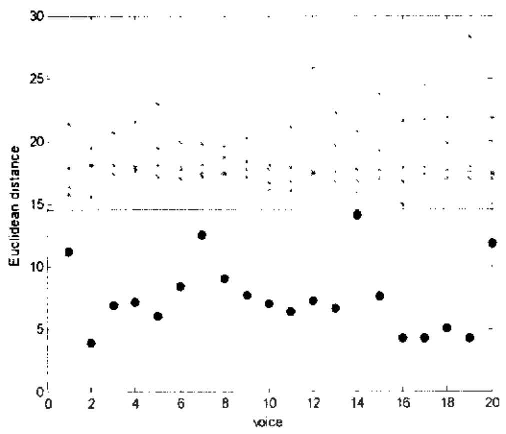
รูปที่ **4.69** ระยะยูคลิด โดยเสียงอ้างอิงเป็นเสียงปิด

ประสิทธิภาพ = 100%

รูปที่ **4.70** ระยะยูคลิด โดยเสียงอ้างอิงเป็นเสียงเปิด

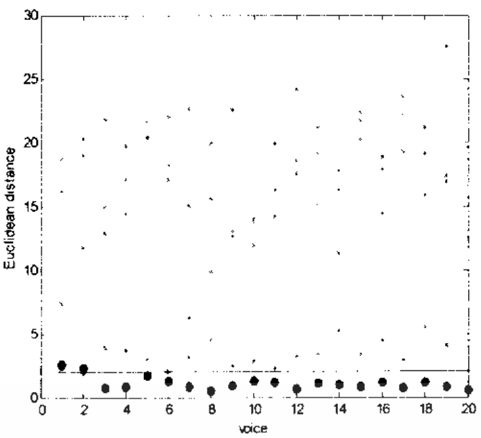
ประสิทธิภาพ = 100%

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



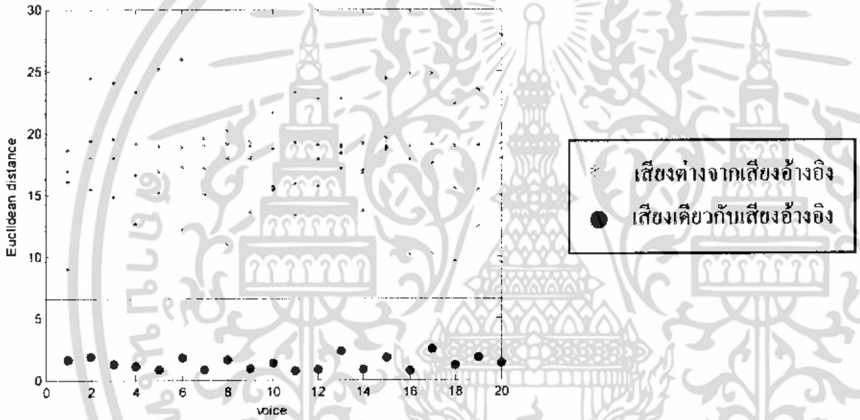
รูปที่ **4.71** ระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงซ้าย

ประสิทธิภาพ = 100%



รูปที่ **4.72** ระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงขวา

ประสิทธิภาพ = 96%

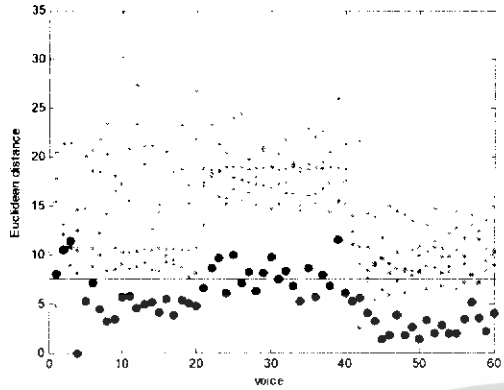


รูปที่ **4.73** ระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงหยุด

ประสิทธิภาพ = 100%

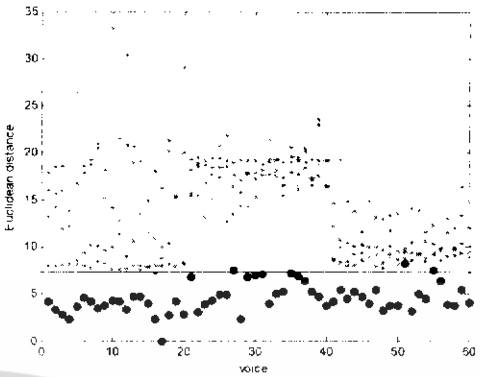
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ผู้พูดหลายคน



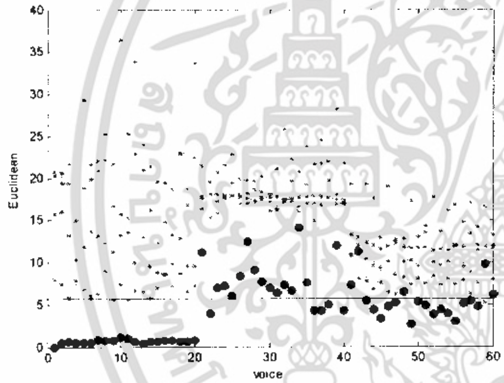
รูปที่ **4.74** ระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงปิด

ประสิทธิภาพ = 90.33%



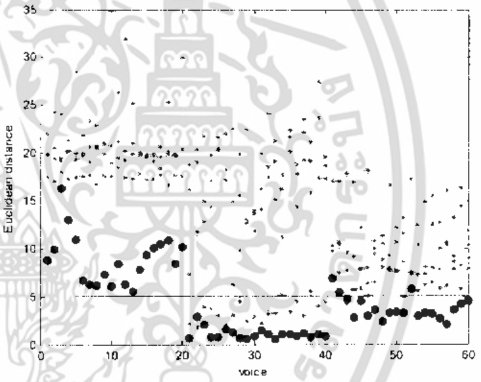
รูปที่ **4.75** ระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงเปิด

ประสิทธิภาพ = 98%



รูปที่ **4.76** ระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงซ้าย

ประสิทธิภาพ = 92.33%

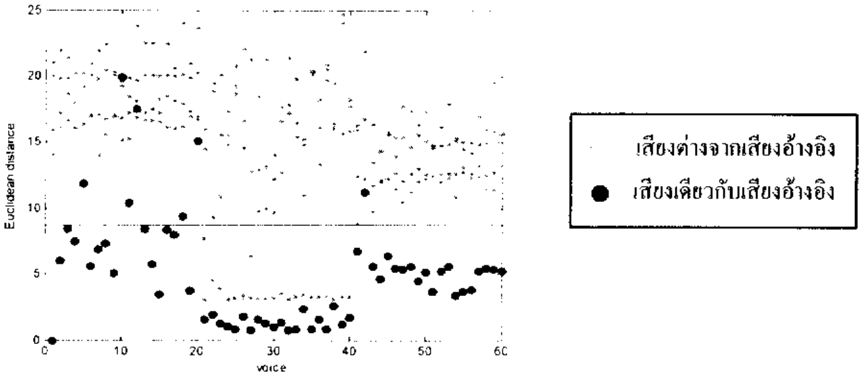


รูปที่ **4.77** ระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงขวา

ประสิทธิภาพ = 85%

จอมเกล้าเจ้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ **4.78** ระยะยูคลิดโดยเสียงอ้างอิงเป็นเสียงหยุด

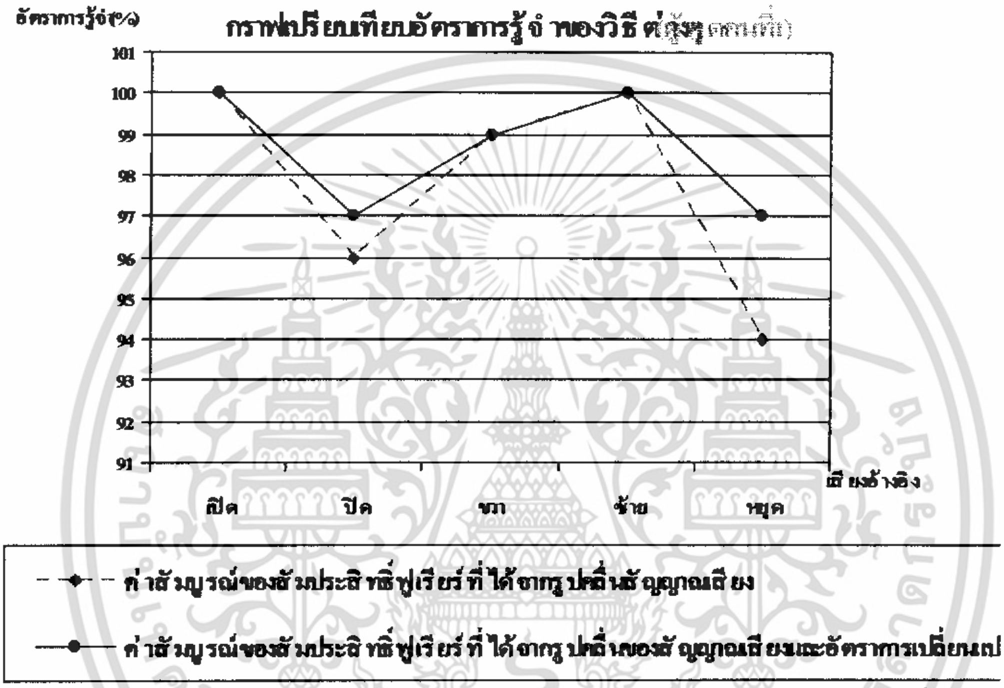
ประสิทธิภาพ = 90.33%

ตารางที่ **4.1** ตารางเปรียบเทียบอัตราการรู้จำของวิธีต่างๆ โดยใช้ผู้พูดคนเดียวกัน

อัตราการรู้จำ / เสียงอ้างอิง	จำนวนเสียงจริง	จำนวนเสียงอื่น	ค่าสัมบูรณ์ของสัมประสิทธิ์ฟูเรียร์ที่ได้จากรูปคลื่นสัญญาณเสียง	ค่าสัมบูรณ์ของสัมประสิทธิ์ฟูเรียร์ที่ได้จากรูปคลื่นสัญญาณเสียงและอัตราการเปลี่ยนแปลงคลื่นเสียงตามเวลา
ผู้พูดคนที่ 1				
เปิด	20	80	100%	100%
ปิด	20	80	96%	97%
ขวา	20	80	99%	99%
ซ้าย	20	80	100%	100%
หยุด	20	80	94%	97%
ผู้พูดคนที่ 2				
เปิด	20	80	99%	99%
ปิด	20	80	96%	96%
ขวา	20	80	95%	97%
ซ้าย	20	80	93%	94%
หยุด	20	80	99%	99%

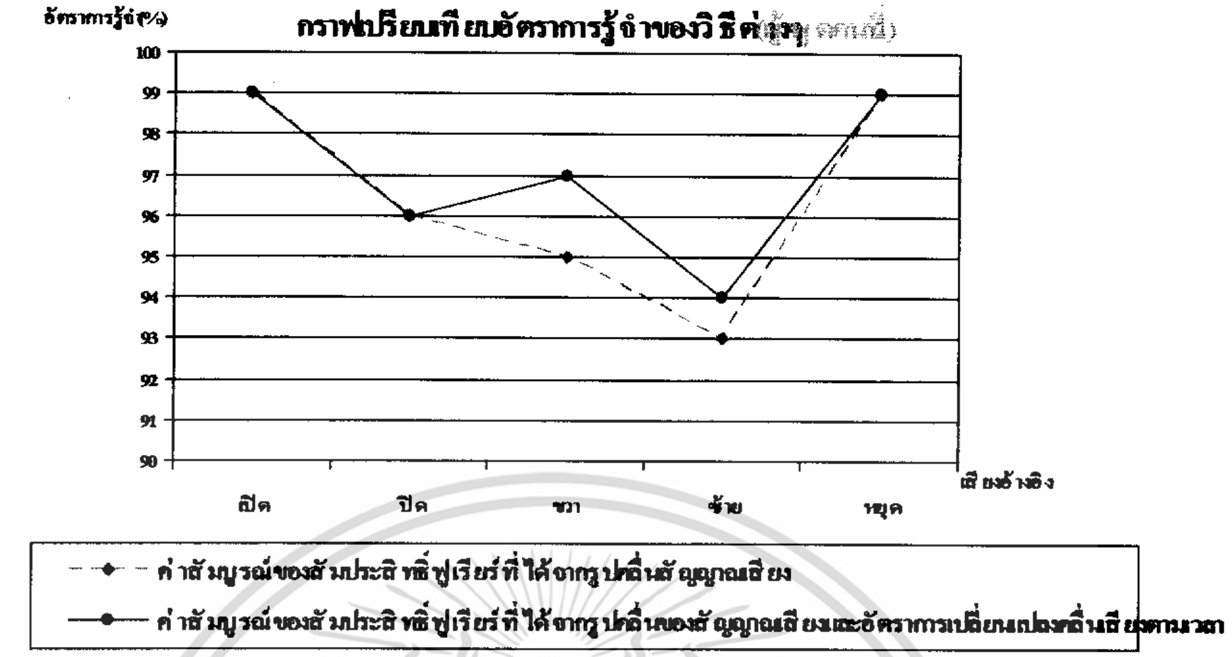
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ผู้พูดคนที่ 3				
เปิด	20	80	100%	100%
ปิด	20	80	100%	100%
ขวา	20	80	95%	96%
ซ้าย	20	80	100%	100%
หยุด	20	80	100%	100%

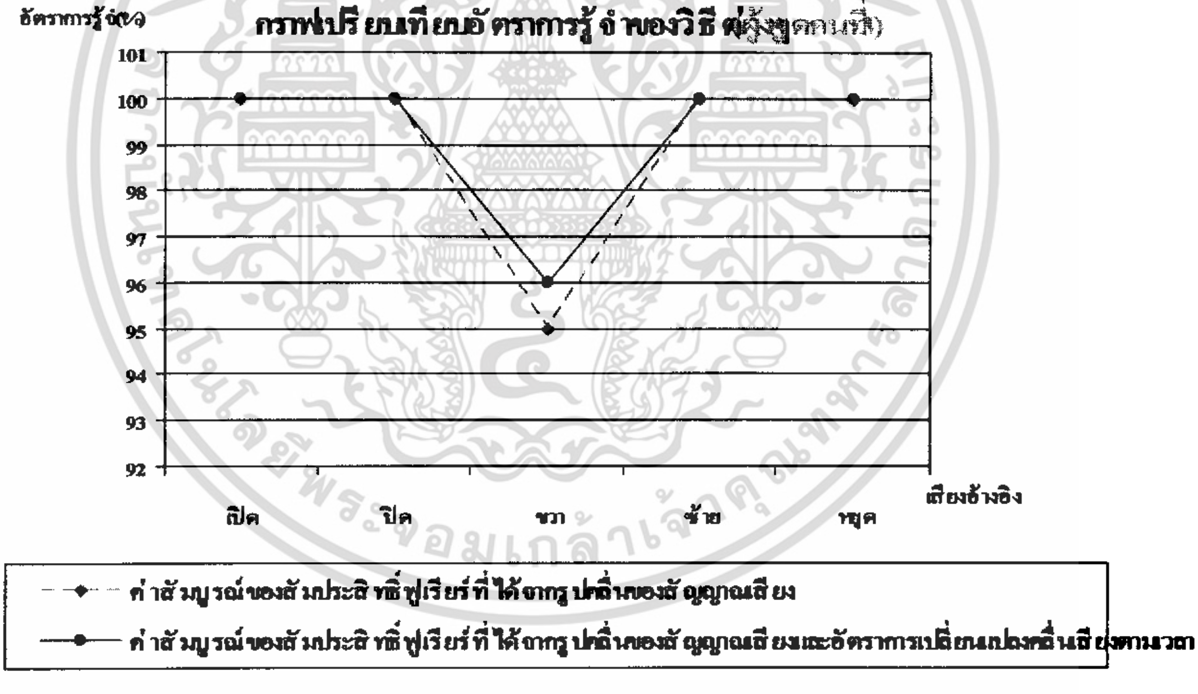


รูปที่ 4.79 กราฟเปรียบเทียบอัตราการรู้จำของวิธีต่าง ๆ ของผู้พูดคนที่ 1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ **4.80** กราฟเปรียบเทียบอัตราการรู้จำของวิธีต่าง ๆ ของผู้พูดคนที่ 2

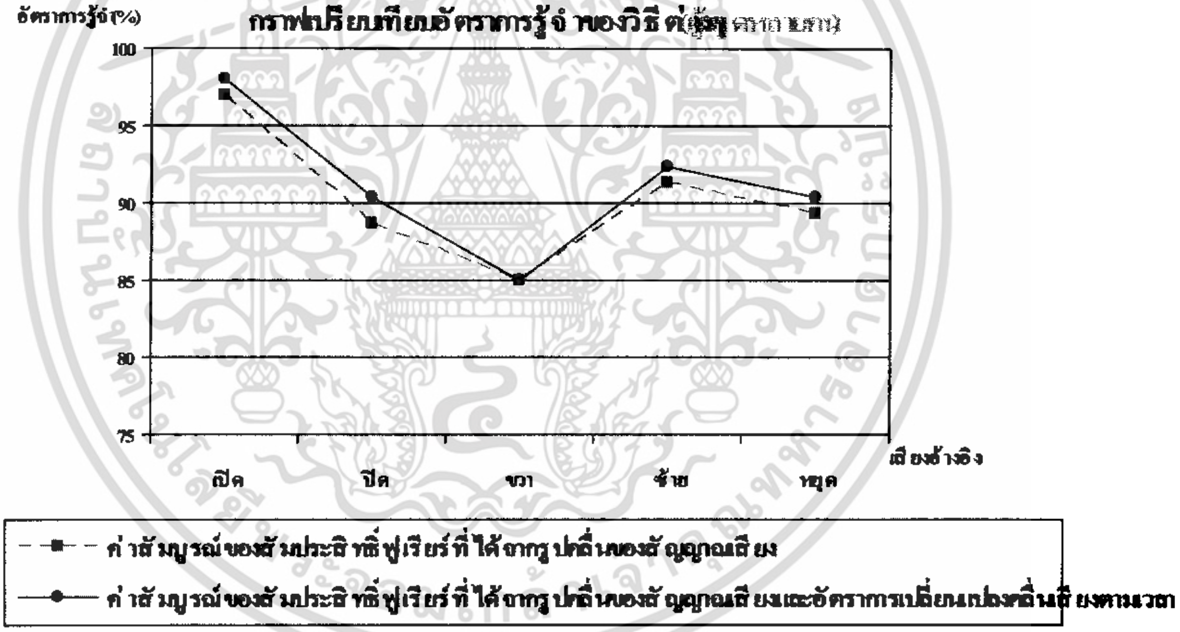


รูปที่ **4.81** กราฟเปรียบเทียบอัตราการรู้จำของวิธีต่าง ๆ ของผู้พูดคนที่ 3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ **4.2** ตารางเปรียบเทียบอัตราการรู้จำของวิธีต่างๆ โดยใช้ผู้พูดหลายคน

อัตราการรู้จำ / เสียงอ้างอิง	จำนวนเสียงจริง	จำนวนเสียงอื่น	ค่าสัมบูรณ์ของสัมประสิทธิ์ฟูเรียร์ที่ได้จากรูปคลื่นสัญญาณเสียง	ค่าสัมบูรณ์ของสัมประสิทธิ์ฟูเรียร์ที่ได้จากรูปคลื่นสัญญาณเสียงและอัตราการเปลี่ยนแปลงคลื่นเสียงตามเวลา
เปิด	60	240	97%	98%
ปิด	60	240	88.67%	90.33%
ขวา	60	240	85%	85%
ซ้าย	60	240	91.33%	92.33%
หยุด	60	240	89.33%	90.33%



รูปที่ **4.82** กราฟเปรียบเทียบอัตราการรู้จำของวิธีต่าง ๆ โดยผู้พูดหลายคน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5
สรุปผลการทดลอง

5.1 สรุปผลการทดลอง

ในปริญญานิพนธ์ฉบับนี้ได้นำเสนอวิธีการตรวจสอบและแยกสัญญาณเสียงโดยใช้ค่าสัมบูรณ์ของสัมประสิทธิ์ฟูเรียร์ที่ได้จากคุณลักษณะสำคัญ 2 อย่างคือ รูปคลื่นของสัญญาณเสียงและอัตราการเปลี่ยนแปลงคลื่นเสียงตามเวลา โดยถ้าใช้ทั้งรูปคลื่นของสัญญาณเสียงและอัตราการเปลี่ยนแปลงคลื่นเสียงโดยตรงจะทำให้ได้ประสิทธิภาพในการแยกเสียงไม่ดีเท่าที่ควร เนื่องจากมีความเบี่ยงเบน ซึ่งความเบี่ยงเบนนี้สามารถแก้ไขได้โดยการกระจายสัญญาณทั้งสองลงไปในอนุกรมฟูเรียร์ โดยใช้การประมาณค่าพร้อมทั้งค่าสัมประสิทธิ์ของอนุกรมฟูเรียร์ที่เหมาะสมซึ่งพิจารณาได้จากข้อมูลอ้างอิงของเสียงนั้น ๆ

และจากงานวิจัย [9] ซึ่งใช้วิธีการเปรียบเทียบโดยโปรแกรมแบบพลวัตมาใช้ในการรู้จำ จะเห็นได้ว่าวิธีการตรวจสอบสัญญาณเสียงโดยใช้ค่าสัมบูรณ์ของสัมประสิทธิ์ฟูเรียร์ที่ได้จากคลื่นของสัญญาณเสียงและอัตราการเปลี่ยนแปลงคลื่นเสียงตามเวลาให้ประสิทธิภาพในการรู้จำดีกว่ามาก

5.2 ปัญหาที่เกิดขึ้นระหว่างการทดลอง

1. เนื่องจากในขณะที่อัดเสียงอาจมีเสียงรบกวนอื่นรอบข้าง จึงทำให้เกิดค่าความเบี่ยงเบนในการแยกเสียงออกจากกัน

2. เสียงของผู้พูดคนเดียวกัน ในบางเสียงมีการออกเสียงที่แตกต่างกัน ทำให้รูปร่างของสัญญาณแตกต่างกันไปด้วย จึงทำให้ค่าความแตกต่างของเสียงมีค่ามาก

3. เสียงที่ต่างกันบางเสียงมีรูปร่างของสัญญาณคล้ายกัน ทำให้เมื่อเปรียบเทียบแล้วผลการทดลองแสดงให้เห็นว่าเป็นเสียงเดียวกัน ซึ่งทำให้เกิดความผิดพลาดขึ้น

5.3 แนวทางในการพัฒนาโครงงาน

1. เพิ่มเติมจำนวนคำที่ใช้ในการรู้จำให้มากขึ้น

2. เพิ่มเติมเสียงของผู้พูดเพศชายเข้าไปในระบบการรู้จำ

3. นำระบบการรู้จำเสียงไปใช้งานกับอุปกรณ์หรือโปรแกรมประยุกต์ต่าง ๆ ที่มีการสั่งงานด้วยเสียง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บรรณานุกรม

[1] ประเจียด ปฐมภาค. 2549. **เสียง**. สืบค้นเมื่อ 7 กรกฎาคม 2549, จาก http://www.pt.ac.th/ptweb/prajead/sound/sound1.htm

[2] มนัส สังวรศิลป์. 2543. **คู่มือการใช้งาน MATLAB ฉบับสมบูรณ์**. กรุงเทพฯ : อินโฟเพลส

[3] วิวัฒน์ กิรานนท์. 2544. **วิศวกรรมการสื่อสาร**. กรุงเทพฯ : สถาบันเทคโนโลยีพระจอมเกล้าเจาคุณทหารลาดกระบัง

[4] เอก ไชยสวัสดิ์. 2543. **สัญญาณและระบบ**. กรุงเทพฯ : วิศวกรรมสถานแห่งประเทศไทย

[5] Alan, V.O. and Alan S.W. **Signals & Systems**. Boston : Prentice-Hall PTR

[6] Lawrence, R. and Biing-Hwang, J. 1999. **Fundamentals of speech recognition**. China : Prentice Hall PTR.

[7] **Wave File Format**. 2006. Retrieved 21 August 2006, from http://www.sonicspot.com/guide/fileformatspecs.html

[8] Pitak Thumwarin and Takenobu Matsuura.**On-Line Writer Recognition for Thai Numerical Based on Barycenter Trajectory and Handwriting Velocity.**

[9] Hiroaki Sakoe and Seibi Chiba.**Dynamic Programming Algorithn Optimization for Spoken Word Recognition.**IEEE Transactions on Acoustics, speech and Signal Processing,Vol.ASSP-26,No.1.February 1978

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้