

การสังเคราะห์พยางค์ภาษาไทยด้วยวิธีการสังเคราะห์แบบวิเคราะห์

โดยใช้คู่เส้นสเปกตรัม



นายชัชวาลย์ หาญสกุลบรรเทิง

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต

สาขาวิชาวิศวกรรมไฟฟ้า ภาควิชาวิศวกรรมไฟฟ้า

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

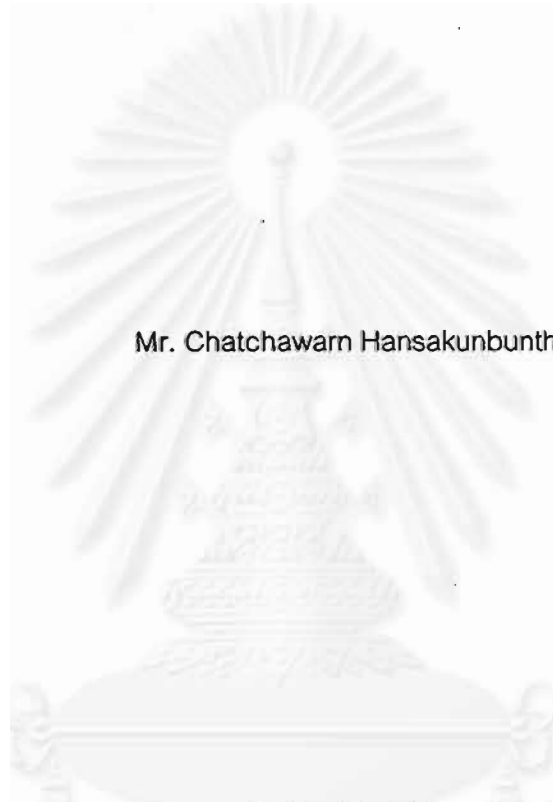
ปีการศึกษา 2542

ISBN 974-334-480-2

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

I 19278612

SYNTHESIS BY ANALYSIS OF THAI SYLLABLES  
BASED ON LINE SPECTRUM PAIRS



Mr. Chatchawarn Hansakunbuntheung

A Thesis Submitted in Partial Fulfillment of the Requirements  
for the Degree of Master of Engineering in Electrical Engineering

Department of Electrical Engineering

Faculty of Engineering

Chulalongkorn University

Academic Year 1999

ISBN 974-334-480-2

หัวข้อวิทยานิพนธ์      การสังเคราะห์พยางค์ภาษาไทยด้วยวิธีการสังเคราะห์แบบวิเคราะห์  
โดยใช้คู่เส้นสเปกตรัม  
โดย                      นาย ชัชวาลย์ หาญสกุลบรรเทียง  
ภาควิชา                      วิศวกรรมไฟฟ้า  
อาจารย์ที่ปรึกษา      รองศาสตราจารย์ ดร. เอกชัย ลีลารัมย์

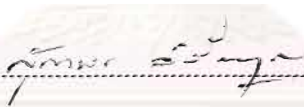
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วน  
หนึ่งของการศึกษาตามหลักสูตรปริญญาโท



คณบดีคณะวิศวกรรมศาสตร์

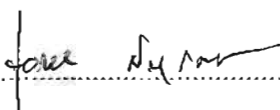
(ศาสตราจารย์ ดร. สมศักดิ์ บุญญาก้ว)

คณะกรรมการสอบวิทยานิพนธ์



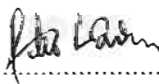
ประธานกรรมการ

(ผู้ช่วยศาสตราจารย์ ดร. สุดาพร ลักษณียนาวิน)



อาจารย์ที่ปรึกษา

(รองศาสตราจารย์ ดร. เอกชัย ลีลารัมย์)



กรรมการ

(อ. สุวิทย์ นาคพืระยเหตุ)

ชัชวาลย์ หาญสกุลบรรเทิง : การสังเคราะห์พยางค์ภาษาไทยด้วยวิธีการ  
สังเคราะห์แบบวิเคราะห์โดยใช้คู่เส้นสเปกตรัม (SYNTHESIS BY ANALYSIS  
OF THAI SYLLABLES BASED ON LINE SPECTRUM PAIRS) อ.ที่ปรึกษา :  
รศ. ดร.เอกชัย ลีลาวัศม์, 109 หน้า. ISBN 974-334-480-2.

คู่เส้นสเปกตรัม (Line Spectrum Pair: LSP) เป็นรูปแบบหนึ่งของสัมประสิทธิ์การเข้ารหัสการทำนายแบบเชิงเส้น (Linear Predictive Coding coefficients: LPC coefficients) ที่มีคุณสมบัติในการเข้ารหัสฟอร์แมนต์ของสัญญาณเสียงพูดและมีคุณสมบัติในการประมาณค่าพารามิเตอร์อย่างมีประสิทธิภาพ วิทยานิพนธ์นี้นำเสนอหน่วยเสียงและวิธีการสังเคราะห์เสียงพยางค์ภาษาไทยโดยใช้คุณสมบัติของคู่เส้นสเปกตรัมในการเข้ารหัสหน่วยเสียงและสังเคราะห์ฟอร์แมนต์ในบริเวณช่วงการเปลี่ยนแปลงของฟอร์แมนต์โดยวิธีการประมาณค่าแบบเชิงเส้น หน่วยเสียงเหล่านี้เก็บจุดอ้างอิงของฟอร์แมนต์ที่เป็นลักษณะเฉพาะของเสียงแต่ละเสียงในภาษาไทย นอกจากนี้ในการสังเคราะห์จะใช้ข้อมูลที่ได้จากการวิเคราะห์จากสัญญาณเสียงพูด อันได้แก่รูปแบบความถี่มูลฐานเพื่อใช้ในการสังเคราะห์เสียงวรรณยุกต์ กรอบแอมพลิจูดเพื่อใช้ควบคุมกรอบแอมพลิจูดของเสียงสังเคราะห์ และระยะเวลาเพื่อควบคุมระยะเวลาในการสังเคราะห์หน่วยเสียงแต่ละหน่วยเสียงของพยางค์แต่ละพยางค์ ในการสังเคราะห์เสียง วิทยานิพนธ์นี้นำเสนอกระบวนการสังเคราะห์โดยใช้หน่วยเสียงจากโครงสร้างพยางค์ที่มีลักษณะพิเศษ หน่วยเสียงเหล่านี้แบ่งประเภทตามลักษณะและหน้าที่ของเสียง จากนั้นนำหน่วยเสียงเหล่านี้ผ่านกระบวนการประมาณค่าเชิงเส้นบนคู่เส้นสเปกตรัมเพื่อสังเคราะห์เสียงพูด การสังเคราะห์เสียงวรรณยุกต์ใช้วิธีทีดี-โซลา(Time-Domain Pitch-Synchronous Overlap Add : TD-PSOLA) โดยกระบวนการที่นำเสนอนี้สามารถใช้สังเคราะห์พยางค์ในภาษาไทยได้ทุกพยางค์

การประเมินคุณภาพเสียงที่สังเคราะห์ตามวิธีในวิทยานิพนธ์นี้ทำโดยอาสาสมัครจำนวน 10 คน ผลการประเมินคุณภาพเสียงที่สังเคราะห์คือ ค่าเอ็มอาร์ที (Modified Rhyme Test : MRT) เท่ากับ 78 เปอร์เซนต์ และค่าเอ็มโอเอส (Most Opinion Score : MOS) เท่ากับ 3.98

ภาควิชา.....วิศวกรรมไฟฟ้า.....ลายมือชื่อนิสิต.....ชัชวาลย์ หาญสกุล  
สาขาวิชา.....วิศวกรรมไฟฟ้า.....ลายมือชื่ออาจารย์ที่ปรึกษา.....เอกชัย ลีลาวัศม์  
ปีการศึกษา.....2542.....

## 4170279521 : MAJOR ELECTRICAL ENGINEERING

KEY WORD: THAI SYLLABLES / SPEECH SYNTHESIS / PHONEMES /  
LINE SPECTRUM PAIR (LSP) / LINEAR INTERPOLATION /  
TIME-DOMAIN PITCH-SYNCHRONOUS OVERLAP ADD (TD-PSOLA)  
CHATCHAWARN HANSAKUNBUNTHEUNG : SYNTHESIS BY ANALYSIS  
OF THAI SYLLABLES0 BASED ON LINE SPECTRUM PAIRS. THESIS  
ADVISOR : ASSO. PROF. EKACHAI LEELARASMEE, Ph.D. 109 pp.  
ISBN 974-334-480-2.

Line spectrum pair (LSP) is one representation of linear predictive coding coefficients (LPC coefficients) which is used in formant coding of speech signal. It provides the stability on the interpolated parameters. This thesis proposes a Thai syllabic speech synthesizing method and its phonemes. This method uses the property of the line spectrum pair to encode phonemes and to generate formant transition between phonemes using linear interpolation. These units contain formant locus of Thai phonemes. In the synthesis method, the data are analyzed from speech signal and are used as a synthesizing database. These data consist of fundamental frequency patterns which are used in regenerating tone; amplitude envelopes which are used in controlling the amplitude envelope of synthesized speech; and time duration which is used in controlling the duration of the synthesized phoneme in each synthesized syllable. To synthesize Thai syllables, the thesis proposes a method of synthesizing speech from units of different syllabic structures. The units are classified into different types of segments by their functions and sounds, then the linear interpolation of line spectrum pairs is operated to generate speech from these units. To regenerate tone, the TD-PSOLA method is selected to implement this work. By this synthesis method, all Thai syllables can be synthesized.

The speech quality of this synthesis method was assessed by 10 volunteers. The results of speech quality assessments have MRT scores of 78 % and MOS of 3.98.

ภาควิชา.....วิศวกรรมไฟฟ้า.....      ลายมือชื่อนิสิต..... อังกรพร ทนทกุลทรัพย์.....  
สาขาวิชา.....วิศวกรรมไฟฟ้า.....      ลายมือชื่ออาจารย์ที่ปรึกษา..... วิชา.....  
ปีการศึกษา.....2542.....



## กิตติกรรมประกาศ

ในการทำวิทยานิพนธ์นี้ตั้งแต่เริ่มต้นจนกระทั่งสามารถสำเร็จได้นี้ ข้าพเจ้าขอกราบขอบพระคุณ รศ. ดร.เอกชัย ลีสารศรี อาจารย์ที่ปรึกษาของข้าพเจ้าที่ให้คำแนะนำต่างๆ อันเป็นประโยชน์ต่อการทำวิทยานิพนธ์และกรุณาให้ความช่วยเหลือในหลายๆ ด้านทำให้การทำวิทยานิพนธ์สามารถดำเนินไปได้ด้วยดี และขอกราบขอบพระคุณ ผศ. ดร.สุดาพร ลักขณียนาวิน ที่ให้ความรู้ทางด้านภาษาศาสตร์และกรุณาให้คำแนะนำอันเป็นประโยชน์ในการทำวิทยานิพนธ์

ขอขอบคุณ คุณประติษฐ์ มิตรราปิยานุรักษ์ ที่คอยให้คำแนะนำและความรู้ที่เป็นประโยชน์อย่างมากในการทำวิทยานิพนธ์และในการแก้ไขปัญหาต่างๆ ขอขอบคุณไว้ ณ ที่นี้ และขอขอบคุณ คุณปิยวรรณ มะธิปิไซ ที่ให้ความช่วยเหลือและช่วยในการประสานงานต่างๆ

ขอขอบคุณ คุณณพงศ์ ปณิธานธรรม ที่ช่วยเป็นผู้ให้เสียงสำหรับนำมาเป็นเสียงต้นแบบในการสังเคราะห์ และขอขอบคุณนิสิตภาควิชาวิศวกรรมไฟฟ้าที่เป็นอาสาสมัครในการทำประเมินคุณภาพเสียง นอกจากนี้ขอขอบคุณนิสิตในห้องปฏิบัติการวิจัยระบบเชิงเลขทุกคนที่ช่วยสร้างบรรยากาศที่ดีภายในห้องปฏิบัติการแห่งนี้

ในท้ายนี้ข้าพเจ้าขอกราบขอบพระคุณบิดาและมารดาของข้าพเจ้าที่เลี้ยงดู ให้กำลังใจ และความช่วยเหลือข้าพเจ้าตลอดมา และส่งเสริมด้านการศึกษาจนข้าพเจ้าสามารถดำเนินมาถึงทุกวันนี้ได้

## สารบัญ

บทคัดย่อภาษาไทย .....	ง
บทคัดย่อภาษาอังกฤษ .....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ .....	ช
สารบัญภาพ.....	ญ
สารบัญตาราง.....	ท
บทที่ 1 บทนำ.....	1
1.1 ที่มาของปัญหา.....	1
1.2 วัตถุประสงค์.....	2
1.3 ขอบเขตงานวิจัย .....	2
1.4 แนวความคิด .....	3
1.5 ขั้นตอนการวิจัย .....	4
1.6 ประโยชน์ที่คาดว่าจะได้รับจากงานวิจัย .....	4
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง.....	5
2.1 งานวิจัยที่เกี่ยวข้อง .....	5
2.2 ทฤษฎีทางด้านเสียง.....	6
2.2.1 การกำเนิดเสียงพูด .....	6
2.2.2 ลักษณะของเสียงพูด .....	7
2.2.3 เสียงพูดภาษาไทย.....	12
2.3 ทฤษฎีทางด้านการวิเคราะห์เสียงและสังเคราะห์เสียง.....	17
2.3.1 การทำหน้าต่าง .....	17
2.3.2 การแปลงฟูริเยร์แบบเต็มหน่วย.....	19
2.3.3 สเปกโตรแกรม .....	20
2.3.4 เอเอ็มดีเอฟ.....	22
2.3.5 การวิเคราะห์และการสังเคราะห์ด้วยการทำนายแบบเชิงเส้น .....	23
2.3.6 คู่เส้นสเปกตรัม.....	28
2.3.7 การประมาณค่าสัมประสิทธิ์.....	31

## สารบัญ (ต่อ)

2.3.8 ทีดี-โซลา.....	32
บทที่ 3 กระบวนการสังเคราะห์เสียง.....	33
3.1 ภาพรวมกระบวนการวิเคราะห์.....	34
3.2 หน่วยเสียง .....	35
3.3 การจัดเก็บหน่วยเสียง .....	43
3.4 รูปแบบความถี่มูลฐาน.....	44
3.5 รูปแบบกรอบแอมพลิจูด.....	45
3.6 ระยะเวลา.....	46
3.7 สรุปขั้นตอนการวิเคราะห์เสียง .....	51
บทที่ 4 กระบวนการสังเคราะห์เสียง.....	52
4.1 โครงสร้างพยางค์สังเคราะห์สำหรับภาษาไทย .....	52
4.2 ภาพรวมกระบวนการสังเคราะห์เสียง .....	54
4.3 การประมาณค่าพารามิเตอร์.....	55
4.4 การสังเคราะห์ด้วยคู่เส้นสเปกตรัม .....	56
4.5 การปรับกรอบแอมพลิจูด.....	57
4.6 การปรับความถี่มูลฐาน .....	59
4.7 สรุปขั้นตอนการสังเคราะห์เสียง .....	60
4.8 โปรแกรมสังเคราะห์เสียง .....	61
บทที่ 5 ผลการสังเคราะห์เสียง.....	62
5.1 ผลการสังเคราะห์ในเชิงเวลาและความถี่.....	62
5.2 ผลการสังเคราะห์เสียงวรรณยุกต์ .....	71
5.3 ผลการสังเคราะห์กรอบแอมพลิจูด .....	72
บทที่ 6 การประเมินผลการสังเคราะห์เสียง.....	74
6.1 เอ็มอาร์ที.....	74
6.2 เอ็มไอเอส.....	75
6.3 ข้อสมมุติฐานในการปรับปรุงหน่วยเสียง .....	75
6.4 การประเมินผล .....	80



## สารบัญ (ต่อ)

บทที่ 7 ผลการประเมินการสังเคราะห์เสียง .....	83
7.1 ผลการประเมินเสียงพยางค์ในระดับพยางค์ .....	83
7.2 ผลการประเมินเสียงสระในระดับพยางค์ .....	84
7.3 ผลการประเมินเสียงวรรณยุกต์ในระดับพยางค์ .....	85
7.4 ผลการประเมินในระดับพยางค์โดยรวม .....	85
7.5 ผลการประเมินระดับวลี .....	85
7.6 ผลการประเมินการปรับปรุงหน่วยเสียง .....	85
7.7 วิเคราะห์ผลการประเมิน .....	86
บทที่ 8 ฐานข้อมูลสำหรับการสังเคราะห์ .....	89
8.1 หน่วยเสียง .....	89
8.2 รูปแบบความถี่มูลฐาน .....	93
8.3 กรอบแอมพลิจูด .....	93
8.4 ระยะเวลา .....	94
8.5 สรุปรูปขนาดฐานข้อมูล .....	94
บทที่ 9 สรุปรงานวิจัยและข้อเสนอแนะ .....	95
9.1 สรุปรงานวิจัย .....	95
9.2 ข้อเสนอแนะ .....	96
รายการอ้างอิง .....	98
ภาคผนวก .....	101
ภาคผนวก ก. แบบประเมินคุณภาพ .....	102
ภาคผนวก ข. แผนผังแสดงขั้นตอนการแปลงสัมประสิทธิ์การทำนายแบบเชิงเส้น เป็นพารามิเตอร์คู่เส้นสเปกตรัม .....	107
ประวัติผู้เขียน .....	109

## สารบัญภาพ

รูปที่ 1 โครงสร้างระบบการออกเสียงของมนุษย์.....	7
รูปที่ 2 รูปคลื่นของลมที่เกิดที่ช่องระหว่างเส้นเสียง.....	8
รูปที่ 3 ฟังก์ชันถ่ายโอนของช่องทางเดินเสียง.....	8
รูปที่ 4 โครงสร้างสเปกตรัมของเสียงก้อง.....	9
รูปที่ 5 รูปคลื่นของเสียงก้อง "เออ".....	9
รูปที่ 6 สเปกโตรแกรมของเสียงก้อง "เออ".....	9
รูปที่ 7 รูปคลื่นของเสียงไม่ก้อง "ส".....	10
รูปที่ 8 สเปกโตรแกรมของเสียงไม่ก้อง "ส".....	10
รูปที่ 9 ตำแหน่งที่ออกเสียง.....	11
รูปที่ 10 การบิดเบนฟอร์แมนต์ของเสียงพยัญชนะ.....	12
รูปที่ 11 เสียงวรรณยุกต์และความถี่มูลฐาน.....	17
รูปที่ 12 ลักษณะหน้าตาที่ใช้โดยทั่วไป.....	18
รูปที่ 13 สเปกตรัมของหน้าตาที่เหลี่ยมเปรียบเทียบกับหน้าตาที่กลม.....	19
รูปที่ 14 สเปกโตรแกรมแถบกว้าง.....	21
รูปที่ 15 สเปกโตรแกรมแถบแคบ.....	21
รูปที่ 16 ผลจากวิธีเอเอ็มพีเอฟ.....	22
รูปที่ 17 ขั้นตอนการวิเคราะห์และสังเคราะห์การทำนายแบบเชิงเส้น.....	27
รูปที่ 18 การวางเรียงสลับของศูนย์ของพหุนามคู่เส้นสเปกตรัม $P(z)$ และ $Q(z)$ .....	29
รูปที่ 19 ความสัมพันธ์ระหว่างรากของ $A(z)$ กับรากของคู่เส้นสเปกตรัม $P(z)$ และ $Q(z)$ .....	30
รูปที่ 20 การทำทีดี-โซลา.....	32
รูปที่ 21 สเปกโตรแกรมต้นแบบ.....	33
รูปที่ 22 ภาพจำลองที่ใช้แทนสเปกโตรแกรม.....	34
รูปที่ 23 แผนผังกระบวนการวิเคราะห์.....	34
รูปที่ 24 ตัวอย่างสเปกโตรแกรมแสดงตำแหน่งการเก็บเสียงสระ "อา".....	36
รูปที่ 25 ตัวอย่างรูปคลื่นสัญญาณเสียงของหน่วยเสียงสระ "อา".....	36
รูปที่ 26 ตำแหน่งการเก็บหน่วยเสียงพยัญชนะ.....	37
รูปที่ 27 ลักษณะรูปคลื่นสัญญาณเสียงของหน่วยเสียงภายในกรอบเสียง.....	38

## สารบัญภาพ (ต่อ)

รูปที่ 28 ตัวอย่างสเปกโตรแกรมแสดงตำแหน่งหน่วยเสียงพยัญชนะต้นของเสียง "สี" .....	39
รูปที่ 29 ตัวอย่างรูปคลื่นสัญญาณเสียงของหน่วยเสียงพยัญชนะต้นของเสียง "สี" .....	39
รูปที่ 30 ตัวอย่างสเปกโตรแกรมแสดงตำแหน่งหน่วยเสียงพยัญชนะสะกดของเสียง "อาก" ..	40
รูปที่ 31 ตัวอย่างรูปคลื่นสัญญาณเสียงของหน่วยเสียงพยัญชนะสะกดของเสียง "อาก" .....	40
รูปที่ 32 ตัวอย่างสเปกโตรแกรมแสดงเสียงพยัญชนะสะกดอัมสระของเสียง "อาย" .....	41
รูปที่ 33 ตัวอย่างสเปกโตรแกรมแสดงตำแหน่ง หน่วยเสียงพยัญชนะสะกดอัมสระของเสียง "อาย" .....	41
รูปที่ 34 ตัวอย่างรูปคลื่นสัญญาณเสียงของ หน่วยเสียงพยัญชนะสะกดอัมสระของเสียง "อาย" .....	41
รูปที่ 35 สเปกโตรแกรมแสดงตำแหน่งเสียงพยัญชนะสะกดนาสิก "อาน" .....	42
รูปที่ 36 ตัวอย่างสเปกโตรแกรมแสดงตำแหน่ง หน่วยเสียงพยัญชนะสะกดนาสิกจากเสียง "อาน" .....	42
รูปที่ 37 ตัวอย่างรูปคลื่นสัญญาณเสียงของ หน่วยเสียงพยัญชนะสะกดนาสิกจากเสียง "อาน" .....	42
รูปที่ 38 การวิเคราะห์แบบหน้าต่างหน่วยเสียงพยัญชนะต้นและพยัญชนะควบกล้ำ ( $C_1T_{cl}$ ) .	43
รูปที่ 39 การวิเคราะห์แบบหน้าต่างหน่วยเสียงพยัญชนะสะกด และสระ ( $V, T_{cm}$ และ $C_m$ ) ..	44
รูปที่ 40 ขั้นตอนการหารูปแบบความถี่มูลฐาน .....	45
รูปที่ 41 ขั้นตอนการหากรอบแอมพลิจูด .....	46
รูปที่ 42 ตัวอย่างการวัดระยะเวลาเสียงพยัญชนะต้นของเสียง "สี" .....	47
รูปที่ 43 ตัวอย่างการวัดระยะเวลาเสียงพยัญชนะควบกล้ำของเสียง "พรา" .....	48
รูปที่ 44 ตัวอย่างการวัดระยะเวลาเสียงพยัญชนะสะกดของเสียง "อาก" .....	49
รูปที่ 45 ตัวอย่างการวัดระยะเวลาเสียงนาสิกของเสียงพยัญชนะสะกด "น" .....	50
รูปที่ 46 ขั้นตอนการวิเคราะห์เสียงโดยสรุป .....	51
รูปที่ 47 โครงสร้างพยางค์ที่ใช้ในการสังเคราะห์เสียง .....	53
รูปที่ 48 แผนผังกระบวนการสังเคราะห์เสียง .....	54
รูปที่ 49 กราฟค่าน้ำหนักสำหรับการประมาณเชิงเส้น .....	55

## สารบัญญภาพ (ต่อ)

รูปที่ 50 การสังเคราะห์เสียงด้วยใช้วงจรรองสังเคราะห์ การทำนายแบบเชิงเส้นผกผัน $H(z)$ .....	56
รูปที่ 51 การประมาณค่ากรอบแอมพลิจูด .....	57
รูปที่ 52 การปรับกรอบแอมพลิจูด.....	58
รูปที่ 53 การประมาณค่ารูปแบบความถี่มูลฐาน.....	59
รูปที่ 54 ขั้นตอนการสังเคราะห์เสียงโดยสรุป .....	60
รูปที่ 55 แผนผังโปรแกรมสังเคราะห์เสียง .....	61
รูปที่ 56 รูปคลื่นเสียง “กา” .....	63
รูปที่ 57 สเปกโตรแกรมของเสียง “กา” .....	63
รูปที่ 58 รูปคลื่นเสียง “ปึก”.....	64
รูปที่ 59 สเปกโตรแกรมของเสียง “ปึก” .....	64
รูปที่ 60 รูปคลื่นเสียง “สาย” .....	65
รูปที่ 61 สเปกโตรแกรมของเสียง “สาย” .....	65
รูปที่ 62 รูปคลื่นเสียง “สาม”.....	66
รูปที่ 63 สเปกโตรแกรมของเสียง “สาม” .....	66
รูปที่ 64 รูปคลื่นเสียง “เกีย”.....	67
รูปที่ 65 สเปกโตรแกรมของเสียง “เกีย” .....	67
รูปที่ 66 รูปคลื่นเสียง “เปี้ยก” .....	68
รูปที่ 67 สเปกโตรแกรมของเสียง “เปี้ยก”.....	68
รูปที่ 68 รูปคลื่นเสียง “เสียว”.....	69
รูปที่ 69 สเปกโตรแกรมของเสียง “เสียว” .....	69
รูปที่ 70 รูปคลื่นเสียง “เสียม” .....	70
รูปที่ 71 สเปกโตรแกรมของเสียง “เสียม”.....	70
รูปที่ 72 รูปแบบความถี่มูลฐานของเสียงวรรณยุกต์โทในฐานข้อมูล.....	71
รูปที่ 73 ผลการสังเคราะห์รูปแบบความถี่มูลฐานของเสียงวรรณยุกต์โท .....	72
รูปที่ 74 กรอบแอมพลิจูดของเสียงวรรณยุกต์โทในฐานข้อมูล .....	73
รูปที่ 75 ผลการสังเคราะห์กรอบแอมพลิจูดของเสียงวรรณยุกต์โท.....	73

## สารบัญตาราง

ตารางที่ 1 สระเดี่ยว.....	13
ตารางที่ 2 สระผสม .....	13
ตารางที่ 3 เสียงพยัญชนะภาษาไทย .....	14
ตารางที่ 4 เสียงพยัญชนะควบกล้ำ .....	15
ตารางที่ 5 เสียงพยัญชนะสะกด .....	16
ตารางที่ 6 ผลการประเมินเสียงพยัญชนะต้น .....	83
ตารางที่ 7 ผลการประเมินเสียงพยัญชนะควบกล้ำ.....	83
ตารางที่ 8 ผลการประเมินเสียงพยัญชนะสะกด .....	84
ตารางที่ 9 สรุปผลการประเมินเสียงพยัญชนะ .....	84
ตารางที่ 10 สรุปผลการประเมินเสียงสระ .....	84
ตารางที่ 11 สรุปผลการประเมินเสียงวรรณยุกต์.....	85
ตารางที่ 12 สรุปผลการประเมินเสียงสังเคราะห์ระดับพยางค์โดยรวม.....	85
ตารางที่ 13 ผลการประเมินการปรับปรุงหน่วยเสียงสระสำหรับเสียงพยัญชนะต้นนาสิก .....	85
ตารางที่ 14 ผลการประเมินการปรับปรุงหน่วยเสียงพยัญชนะสะกดกลุ่มอัมสระ .....	86
ตารางที่ 15 หน่วยเสียงกลุ่ม $C_iT_{Ci}$ .....	90
ตารางที่ 16 หน่วยเสียงกลุ่ม V.....	91
ตารางที่ 17 หน่วยเสียงกลุ่ม $T_{Cm}$ .....	92
ตารางที่ 18 หน่วยเสียงกลุ่ม $C_m$ .....	92

บทที่ 1  
บทนำ



1.1 ที่มาของปัญหา

ในปัจจุบันการสื่อสารเป็นสิ่งสำคัญในอันดับต้น อาจกล่าวได้ว่าปัจจุบันเป็นยุคของการสื่อสาร การสื่อสารเป็นสิ่งหนึ่งที่ช่วยลดช่องว่างระหว่างบุคคล รูปแบบหนึ่งของการสื่อสารที่สำคัญคือการสื่อสารด้วยภาษาพูดซึ่งเป็นรูปแบบที่สามารถถ่ายทอดได้รวดเร็วและได้ใจความ ดังนั้นจึงมีงานศึกษาและวิจัยทางด้านสื่อสารด้วยภาษาพูดเกิดขึ้นเป็นจำนวนมากซึ่งงานด้านหนึ่งที่สำคัญคือการสังเคราะห์เสียงพูด (Speech Synthesis) เป็นการสร้างขบวนการย้อนกลับในระบบการสื่อสารของคน งานทางด้านนี้นำไปใช้ในหลายๆด้านโดยเฉพาะในปัจจุบันซึ่งมีความก้าวหน้าทางด้านคอมพิวเตอร์และระบบเครือข่ายสื่อสาร ดังเช่นงานทางด้านสื่อสารโทรศัพท์ทางไกล (Telephony) ระบบแปลภาษาพูด (Machine Translator) ระบบช่วยเหลือสำหรับคนพิการโดยเฉพาะคนตาบอด และคนมีปัญญาในการพูด เป็นต้น

การวิจัยและพัฒนาทางด้านสังเคราะห์เสียงเริ่มต้นมานานสำหรับภาษาหลักต่างๆ เช่นภาษาอังกฤษ ภาษาญี่ปุ่น ภาษาจีน เป็นต้น การสังเคราะห์เสียงดังกล่าวถูกพัฒนาในหลากหลายวิธีเพื่อให้เสียงที่สังเคราะห์ขึ้นมีคุณภาพสูง สำหรับการสังเคราะห์เสียงสำหรับภาษาไทยมีการวิจัยมานานแต่มีผู้ทำการศึกษาและวิจัยทางด้านนี้น้อย ทำให้ผลงานวิจัยทางด้านนี้สำหรับภาษาไทยมีน้อย วิธีส่วนใหญ่ที่ใช้สังเคราะห์เสียงได้แก่ การสังเคราะห์เสียงด้วยวิธีการเชื่อมต่อ (Concatenation) และวิธีการสังเคราะห์โดยการวิเคราะห์ เป็นต้น สำหรับงานวิจัยในด้านการสังเคราะห์เสียงภาษาไทยมีงานวิจัยดังต่อไปนี้ ชัชวลิต สรวารี (Saravari, 1983) สุดาพร ลักษณะนิยานาวิน (Luksaneeyanawin, 1989, 1993, 1995) กฤษดา เรยเส (2530) อาทร นันทิยากุล (2533) สุดาพร ลักษณะนิยานาวิน (2534, 2535) สมนึก ศิริโตและคณะ (2538) และ เอกพล อนุสุเรนทร์ (2541)

การสังเคราะห์เสียงด้วยวิธีการเชื่อมต่อ เป็นวิธีการเก็บส่วนของเสียงที่มีลักษณะร่วมกันเพื่อนำมาเชื่อมต่อกันให้ได้เสียงคำพูดที่ต้องการ วิธีการนี้มีคุณภาพเสียงสูงเนื่องจากเป็นเสียงที่ได้จากการตัดมาจากเสียงพูดจริง แต่วิธีการดังกล่าวมีปัญหาทางด้านคุณภาพเสียงบริเวณรอยต่อและการปรับทำนองและจังหวะของเสียงเนื่องจากส่วนของเสียงที่เก็บไว้มีลักษณะคงตัว เพื่อให้การสังเคราะห์ได้ครอบคลุมเสียงที่เกิดขึ้นได้ทั้งหมด จำเป็นต้องเก็บส่วนของเสียงไว้ตามจำนวน

ลักษณะเสียงทั้งหมด จำเป็นต้องใช้หน่วยความจำเป็นจำนวนมากและไม่สะดวกในการนำไปใช้งานจริง

ในวิทยานิพนธ์นี้นำเสนอวิธีการสังเคราะห์เสียงที่ช่วยลดขนาดข้อมูลและจำนวนหน่วยเสียง โดยที่สามารถปรับลักษณะทำนองของเสียงที่สังเคราะห์ได้ เพื่อให้สามารถนำไปประยุกต์ปรับเสียงสังเคราะห์ให้ใกล้เคียงธรรมชาติ การสังเคราะห์เสียงที่กล่าวในวิทยานิพนธ์ฉบับนี้ใช้หน่วยเสียงที่เก็บลักษณะเฉพาะของเสียงพยัญชนะและสระต่างๆในรูปของพารามิเตอร์คู่เส้นสเปกตรัมเพื่อให้มีความยืดหยุ่นในการสังเคราะห์ การสังเคราะห์ใช้การประมาณค่าพารามิเตอร์ในการสร้างเสียงกลับคืนให้สมบูรณ์เพื่อให้เสียงที่ได้มีความต่อเนื่องในเชิงความถี่และใช้วิธีพิทซชิงโครนัส โอเวอร์แลปแอดด์สร้างเสียงวรรณยุกต์และช่วยให้เสียงสังเคราะห์มีความต่อเนื่องในเชิงเวลา

## 1.2 วัตถุประสงค์

1. เพื่อศึกษาและหาหน่วยเสียงที่จำเป็นสำหรับการสังเคราะห์เสียงภาษาไทย
2. เพื่อพัฒนาวิธีการสังเคราะห์เสียงโดยใช้คู่สเปกตรัมแบบเส้น
3. เพื่อพัฒนาโปรแกรมสังเคราะห์เสียงภาษาไทยแบบพยางค์เดียว

## 1.3 ขอบเขตงานวิจัย

1. ใช้หน่วยเสียงสำหรับสังเคราะห์เสียงน้อยกว่า 400 หน่วยเสียง
2. ทำการสังเคราะห์เสียงที่มีคุณสมบัติดังนี้
  - 2.1 สังเคราะห์เสียงพยางค์ภาษาไทยได้ทั้งพยางค์เปิดและพยางค์ปิด
  - 2.2 สังเคราะห์โดยใช้คู่สเปกตรัมแบบเส้น
  - 2.3 สังเคราะห์เสียงวรรณยุกต์โดยใช้ PSOLA เพื่อสังเคราะห์เสียงวรรณยุกต์ดังต่อไปนี้ เสียงสามัญ เสียงเอกเสียงโท เสียงตรี และเสียงจัตวา
  - 2.4 ปรับระยะเวลาของเสียงได้
3. พัฒนาโปรแกรมสำหรับสังเคราะห์เสียงบนโปรแกรม MatLab
4. หน่วยเสียงทำการบันทึกที่ 16 บิตที่อัตราสุ่ม 22,050 เฮิรตซ์

#### 1.4 แนวความคิด

ในขั้นต้นพิจารณาลักษณะของเสียงพูดพบว่า ลักษณะเสียงพูดส่วนใหญ่มีลักษณะเป็นเสียงก้องซึ่งเสียงดังกล่าวเกิดจากเสียงสระซึ่งมีลักษณะเป็นลูกคลื่นที่มีลักษณะซ้ำหรือใกล้เคียงกัน เรียงต่อกันในเชิงเวลาและเมื่อพิจารณาในเชิงความถี่พบว่าลักษณะของเสียงสระดังกล่าวมีลักษณะของฟอร์แมนต์ซึ่งเป็นลักษณะเฉพาะของเสียงในเชิงความถี่ที่ซ้ำกัน ในกรณีเสียงสระผสมเกิดจากการเชื่อมต่อฟอร์แมนต์ระหว่างเสียงสระสองตัว ดังนั้นในการเก็บส่วนของเสียงเพื่อใช้เป็นหน่วยเสียงสำหรับการสังเคราะห์ จึงเกิดความคิดที่จะเก็บลักษณะของเสียงสระหรือลักษณะฟอร์แมนต์แทนการเก็บเสียงในส่วนนี้ทั้งหมดแล้วนำหน่วยเสียงดังกล่าวมาเชื่อมต่อกันเพื่อสังเคราะห์เสียงเดิมกลับ

ในส่วนของเสียงพยัญชนะ โดยอาศัยผลการศึกษาทางด้านโสตสัทศาสตร์พบว่าเกิดจากการบิดเบนของฟอร์แมนต์บริเวณปลายของเสียงสระในเชิงความถี่ (Lieberman, 1988) ทั้งนี้มีตำแหน่งการบิดเบนแตกต่างกันสำหรับเสียงพยัญชนะแต่ละเสียงและมีความแตกต่างกันสำหรับแต่ละสระและในเสียงพูดมีพยางค์ในภาษาไทยเป็นจำนวนมากที่มีเสียงพยัญชนะหรือเสียงสระบางส่วนซ้ำกัน ดังนั้นการเก็บหน่วยเสียงสำหรับการสังเคราะห์เสียงพยัญชนะจึงไม่จำเป็นต้องทำการเก็บส่วนบิดเบนหรือส่วนที่มีการเปลี่ยนแปลงของฟอร์แมนต์แต่เก็บเฉพาะตำแหน่ง (Locus) ของฟอร์แมนต์บริเวณปลายของการบิดเบนของเสียงสระเพื่อใช้อ้างอิงตำแหน่งในการสังเคราะห์ส่วนบิดเบนกลับคืน ด้วยวิธีนี้จะช่วยลดการเก็บข้อมูลเสียงในส่วนที่มีการเปลี่ยนแปลงของฟอร์แมนต์และช่วยให้เสียงที่สังเคราะห์มีความต่อเนื่องในบริเวณรอยต่อของเสียงพยัญชนะและสระ

สำหรับเสียงวรรณยุกต์เกิดจากการเปลี่ยนแปลงระยะห่างระหว่างลูกคลื่นหรือระยะพิทช์ (Pitch) ของเสียงก้อง ซึ่งแต่ละเสียงวรรณยุกต์มีรูปแบบการเปลี่ยนแปลงระยะห่างดังกล่าวแตกต่างกัน ดังนั้นการเก็บเสียงของแต่ละพยางค์จึงไม่จำเป็นต้องเก็บทุกเสียงวรรณยุกต์ แต่อาศัยการปรับระยะห่างดังกล่าวตามเสียงวรรณยุกต์ที่ต้องการ วิธีการนี้จะช่วยลดจำนวนหน่วยเสียงและลดขนาดข้อมูลที่ต้องเก็บ

จากหลักการในการเก็บหน่วยเสียงข้างต้น ต้องหาวิธีการและตัวแทนในการเก็บหน่วยเสียงที่เหมาะสมสำหรับการสังเคราะห์ด้วยหน่วยเสียงดังกล่าว จากการศึกษาพบว่าพารามิเตอร์ของคู่เส้นสเปกตรัม (Line Spectrum Pair) ซึ่งเป็นพารามิเตอร์รูปแบบหนึ่งของการเข้ารหัสการทำนายแบบเชิงเส้น (Linear Prediction Coding) มีคุณสมบัติในการเข้ารหัสตำแหน่งฟอร์แมนต์ของเสียงในเชิงความถี่ของเสียงพูดและมีคุณสมบัติในการประมาณค่า (Interpolation) ฟอร์แมนต์ จึงได้



เกิดแนวคิดที่จะสังเคราะห์เสียงโดยอาศัยการเก็บหน่วยเสียงที่เก็บลักษณะเฉพาะของเสียงพยัญชนะ และสระแต่ละเสียงแทนการเก็บเสียงทั้งหมด โดยให้คู่เส้นสเปกตรัมในการเก็บและสังเคราะห์เสียง

#### 1.5 ขั้นตอนการวิจัย

1. ศึกษาและค้นคว้าข้อมูลทางการวิเคราะห์และสังเคราะห์เสียง
2. ศึกษาและค้นคว้าข้อมูลทางด้านภาษาศาสตร์
3. ออกแบบวิธีในการวิเคราะห์และสังเคราะห์เสียง
4. บันทึกเสียง คัดเลือกและจัดเก็บเป็นหน่วยเสียง
5. เขียนโปรแกรมในการสังเคราะห์เสียง
6. ทดลองสังเคราะห์เสียงและปรับปรุง
7. ทำการทดสอบการฟัง
8. ประเมินผล สรุปผลการทดสอบ และเสนอแนะแนวทาง
9. จัดทำวิทยานิพนธ์

#### 1.6 ประโยชน์ที่คาดว่าจะได้รับจากงานวิจัย

1. เป็นแนวทางพัฒนาทางด้านสังเคราะห์เสียงในประเทศไทย
2. สามารถนำไปพัฒนาเป็นระบบสื่อสารเพื่อติดต่อกับมนุษย์ โดยเฉพาะทางด้านระบบช่วยเหลือผู้พิการ
3. สามารถนำไปประยุกต์เพื่อสร้างเป็นเครื่องสังเคราะห์เสียงภาษาไทย

## บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในบทนี้กล่าวถึงงานวิจัยและงานที่เกี่ยวข้องทางการสังเคราะห์เสียงจากแหล่งต่างๆ เพื่อนำเสนอความก้าวหน้าทางการสังเคราะห์เสียง ในส่วนถัดมากล่าวถึงทฤษฎีที่เกี่ยวข้องและได้อ้างอิงในงานวิทยานิพนธ์ฉบับนี้เพื่อให้สามารถศึกษาและทำความเข้าใจเนื้อหาของวิทยานิพนธ์ได้เป็นอย่างดี

### 2.1 งานวิจัยที่เกี่ยวข้อง

งานวิจัยในอดีตที่เกี่ยวกับการสังเคราะห์เสียงโดยใช้การเข้ารหัสการทำนายแบบเชิงเส้น ดังเช่นในงานของ กฤษดา เรยเส (2530) และ อาทร นันทิยะกุล (2533) เป็นการสังเคราะห์เสียงภาษาไทยในเชิงวิเคราะห์ที่ใช้การเข้ารหัสการทำนายแบบเชิงเส้นในการเข้ารหัสเสียงพูดในระดับคำ โดยนำไปวิเคราะห์เป็นสัมประสิทธิ์การเข้ารหัสการทำนายแบบเชิงเส้น แล้วสังเคราะห์กลับโดยใช้สัมประสิทธิ์การเข้ารหัสการทำนายแบบเชิงเส้นจากการวิเคราะห์นำไปผ่านวงจรกรองการทำนายแบบเชิงเส้น (LP Filter) ซึ่งจะได้เสียงสังเคราะห์กลับคืนมา

สำหรับงานวิจัยทางการสังเคราะห์เสียงโดยใช้คู่เส้นสเปกตรัม (Line Spectrum Pair) มีงานวิจัยของ เอเวเรตต์ (Everett, 1988) เป็นการสังเคราะห์เสียงภาษาอังกฤษโดยการวิเคราะห์คู่เส้นสเปกตรัมและสังเคราะห์โดยการเชื่อมต่อ (Concatenation) หน่วยเสียง โดยหน่วยเสียงที่ใช้มีหลายระดับตั้งแต่ระดับหน่วยเสียงย่อย (Subphonemes) จนถึงระดับพยางค์ (Syllables) โดยเก็บหน่วยเสียงในรูปของพารามิเตอร์คู่เส้นสเปกตรัม ซึ่งผลการสังเคราะห์ที่ได้มีคุณภาพสูง

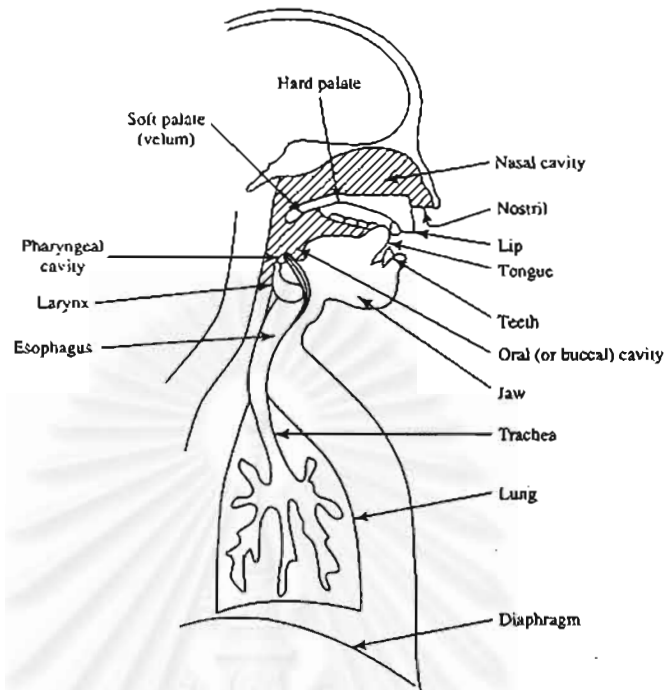
สำหรับงานวิจัยทางการสังเคราะห์เสียงวรรณยุกต์โดยใช้โวลลา (Pitch-Synchronous Overlap-Add : PSOLA) ดังในงานของ คาร์เพนเทีย และสเตลลา (Charpentier and Stella, 1986) และ คาร์เพนเทีย และมูลินส์ (Charpentier and Moulines, 1989) สำหรับการสังเคราะห์เสียงวรรณยุกต์ในภาษาไทยมีผลงานการวิจัยของ สุดาพร ลักษณะียนาวิน (Luksaneeyanawin, 1993, 1995) ซึ่งผลการสังเคราะห์ที่ได้มีคุณภาพสูง

## 2.2 ทฤษฎีทางด้านเสียง

ในขั้นแรกก่อนที่จะศึกษาทางด้านการสังเคราะห์เสียงต้องศึกษาเกี่ยวกับลักษณะและคุณสมบัติของเสียงพูดก่อนเพื่อศึกษาหาลักษณะหรือคุณสมบัติของเสียงที่เหมาะสมและสามารถนำมาใช้ในการสังเคราะห์เสียงได้ นอกจากนี้เสียงพูดในแต่ละภาษามีรูปแบบการเกิดของเสียงที่แตกต่างกัน เนื่องจากในวิทยานิพนธ์นี้เป็นการสังเคราะห์เสียงพูดภาษาไทยจึงได้นำเสนอทฤษฎีเกี่ยวกับเสียงพูดภาษาไทยประกอบไว้ ณ ที่นี้

### 2.2.1 การกำเนิดเสียงพูด

เสียงพูดของมนุษย์เกิดจากระบบการออกเสียงของมนุษย์ดังในรูปที่ 1 การกำเนิดเสียงเริ่มจากความดันลมจากปอด (Lung) ถูกดันผ่านเส้นเสียง (Vocal Cords) ซึ่งอยู่ภายในกล่องเสียง (Larynx) เส้นเสียงดังกล่าวเป็นตัวควบคุมจังหวะการปล่อยของลมที่เข้าสู่ช่องทางเดินเสียง (Vocal Tract) ซึ่งเริ่มตั้งแต่ทางออกของกล่องเสียงผ่านคอหอย (Pharynx) แล้วแยกออกไปสองช่องทางคือ ทางช่องปาก (Oral Cavity) ไปจนถึงริมฝีปากและทางช่องจมูก (Nasal Cavity) ไปจนถึงรูจมูก (Nostrils) (Ince, 1992) ส่วนนี้มีลักษณะเป็นท่อที่มีขนาดพื้นที่หน้าตัดต่างๆ กัน ซึ่งเป็นส่วนที่ทำให้เกิดการกำทอน (Resonance) ของเสียงที่มีความถี่ต่างๆ หรือเรียกว่าฟอร์มแมนต์ (Formants) ขึ้นกับการเปลี่ยนแปลงขนาดของท่อดังกล่าวและอาจเกิดการต่อต้านการกำทอน (Anti-resonance) ขึ้นด้วย (Donovan, 1996) ซึ่งเสียงที่เกิดจากช่องทางปากนี้ครอบคลุมเสียงสระ (Vowel Sounds) ทั้งหมดและเสียงพยัญชนะบางส่วน ส่วนเสียงที่เกิดจากช่องจมูกทำให้เกิดเสียงนาสิก (Nasal Sounds) ซึ่งเกิดการต่อต้านการกำทอนทำให้เสียงมีสเปกตรัมที่มีความเข้มต่ำในช่วง 500 ถึง 2000 เฮิรตซ์หรือบริเวณฟอร์มแมนต์ที่สอง เสียงที่เกิดขึ้นจากระบบการนี้ทำให้ได้ยินเป็นเสียงพูดของมนุษย์



รูปที่ 1 โครงสร้างระบบการออกเสียงของมนุษย์

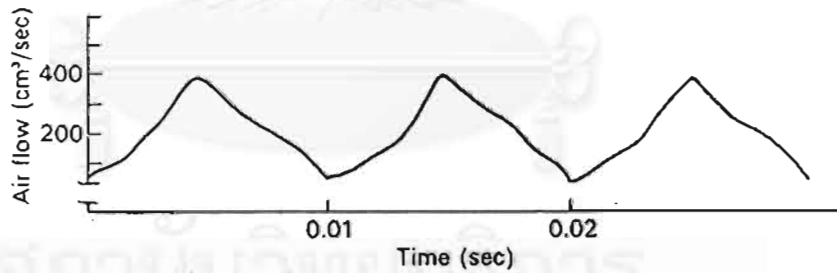
เมื่อมองทางด้านวิศวกรรม ปอดและเส้นเสียงเปรียบเสมือนกับแหล่งกำเนิดสัญญาณ (Source) และลมที่ออกจากปอดจนผ่านกล่องเสียงและเส้นเสียงเปรียบได้กับสัญญาณกระตุ้น ช่องทางเดินเสียงซึ่งทำหน้าที่ดัดแปลงคุณสมบัติของลมเปรียบได้กับวงจรกรองเสียง (Acoustic Filters) ที่มีฟังก์ชันถ่ายโอน (Transfer Function) ที่แปรเปลี่ยนตามเวลาได้ จากมุมมองดังกล่าว ได้มีการนำมาสร้างเป็นแบบจำลองทางคณิตศาสตร์เพื่อจำลองระบบการออกเสียงของมนุษย์ดัง เช่น แบบจำลองการทำนายแบบเชิงเส้น

## 2.2.2 ลักษณะของเสียงพูด

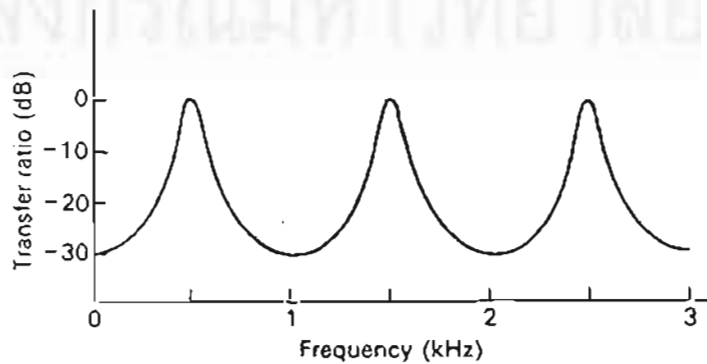
เสียงพูดของมนุษย์เกิดได้หลายลักษณะขึ้นกับลักษณะการกระตุ้น ตำแหน่งการออกเสียง และลักษณะการออกเสียง ซึ่งเกิดจากการเปลี่ยนแปลงลักษณะช่องทางเดินเสียงและการควบคุม เส้นเสียงดังนี้

1. ลักษณะการกระตุ้น คือ ลักษณะของลมที่ผ่านเส้นเสียงเมื่อเส้นเสียงเกิดการเปิดหรือปิด ในกรณีที่เส้นเสียงปิดจะทำให้เกิดการดัดแปลงลมลมและเกิดความดันจนกระทั่งเส้นเสียงถูกกันให้เปิด การปลดปล่อยลมที่เกิดขึ้นมีลักษณะเป็นกึ่งรายคาบ (Quasi-periodic) (Rabiner and Schafer, 1978) ดังในรูปที่ 2 โดยมีคาบการเปิดและปิดขึ้นกับความตึงของเส้นเสียง ลมดังกล่าว

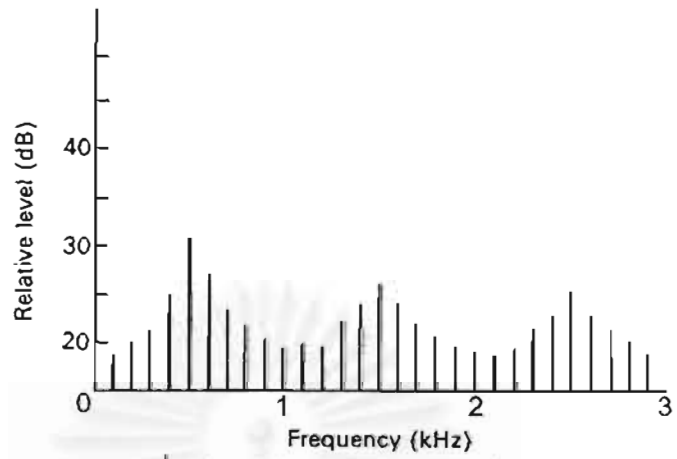
จะผ่านเข้าสู่ช่องทางเดินเสียงทำให้เกิดการกำทอนและเกิดฟอร์แมนต์ที่มีลักษณะเฉพาะสำหรับแต่ละรูปร่างของช่องทางเดินเสียงดังตัวอย่างในรูป 3 แสดงตัวอย่างลักษณะเฉพาะเชิงความถี่ของช่องทางเดินเสียง คาบเวลาการปลดปล่อยลมเมื่อนำมาคำนวณเป็นความถี่ ความถี่ที่ได้เรียกว่าความถี่มูลฐาน (Fundamental Frequency : F0) ซึ่งเป็นตัวกำหนดเสียงวรรณยุกต์ (Tone) ในภาษาพูด และลักษณะสเปกตรัมของเสียงก้องมีลักษณะโครงสร้างฮาร์โมนิคของความถี่มูลฐานอยู่ภายใต้กรอบลักษณะเฉพาะเชิงความถี่ของช่องทางเดินเสียงดังในรูปที่ 4 ซึ่งแสดงโครงสร้างฮาร์โมนิคและฟอร์แมนต์ เสียงที่เกิดขึ้นนี้เรียกว่าเสียงก้อง (Voiced Sound) ดังตัวอย่างรูปคลื่นในรูปที่ 5 ซึ่งแสดงให้เห็นลักษณะที่เป็นรายคาบของคลื่นของเสียง "เออ" และในรูปที่ 6 แสดงให้เห็นฟอร์แมนต์ของส่วนที่เป็นเสียงก้องของเสียงว่า "เออ" ซึ่งแสดงด้วยบริเวณที่เป็นแถบดำขนานกับแกนนอน ในกรณีเส้นเสียงเปิดตลอดจึงไม่เกิดการดัดแปลงลมและลมที่เกิดขึ้นจึงมีลักษณะไม่เป็นรายคาบ (Aperiodic) มีลักษณะคล้ายสัญญาณรบกวนสเปกตรัมกว้าง (Broad-Spectrum Noise) (Rabiner and Schafer, 1978) และไม่เกิดฟอร์แมนต์ในช่วงเวลาดังกล่าว เสียงที่เกิดขึ้นเรียกว่าเสียงไม่ก้อง (Unvoiced Sounds or Voiceless Sounds ) ดังในรูปที่ 7 แสดงตัวอย่างรูปคลื่นของเสียงไม่ก้องที่มีลักษณะไม่เป็นรายคาบของเสียง "ส" ในคำว่า "สี" และในรูปที่ 8 แสดงกราฟในเชิงความถี่เทียบกับเวลาซึ่งแสดงให้เห็นว่ามีลักษณะเป็นสัญญาณรบกวน



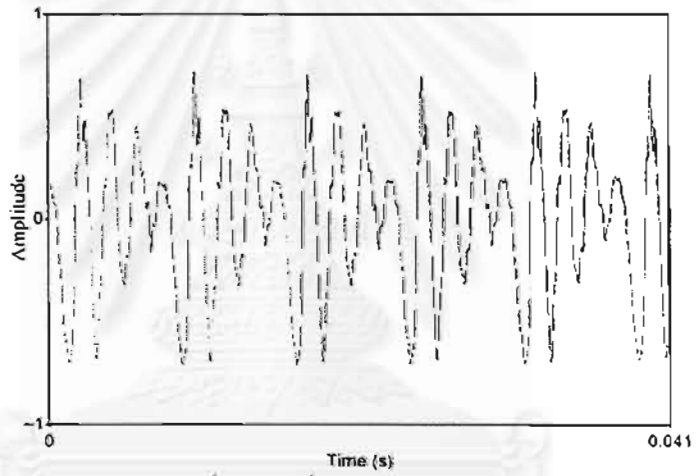
รูปที่ 2 รูปคลื่นของลมที่เกิดที่ช่องระหว่างเส้นเสียง



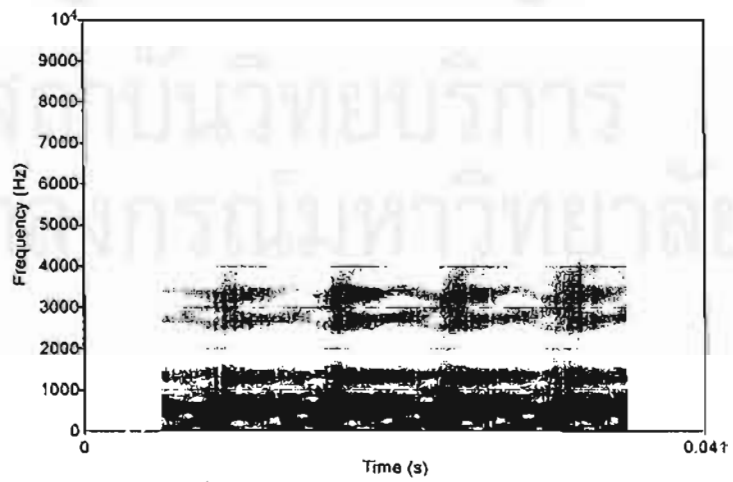
รูปที่ 3 ฟังก์ชันถ่ายโอนของช่องทางเดินเสียง



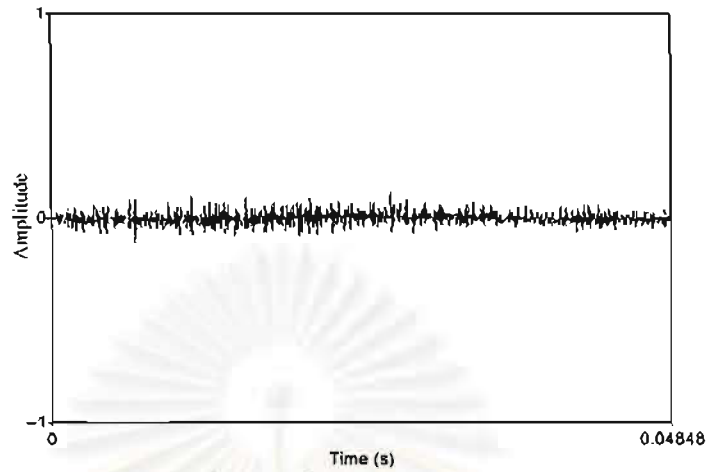
รูปที่ 4 โครงสร้างสเปกตรัมของเสียงก้อง



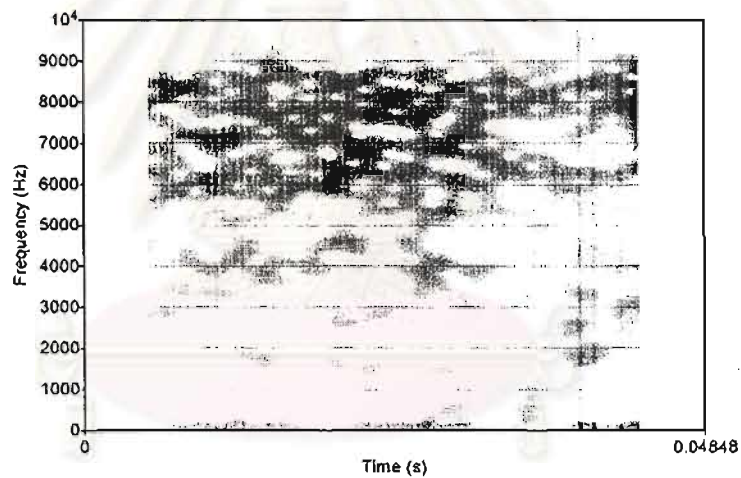
รูปที่ 5 รูปคลื่นของเสียงก้อง "เออ"



รูปที่ 6 สเปกโตรแกรมของเสียงก้อง "เออ"



รูปที่ 7 รูปคลื่นของเสียงไม่ก้อง "ส"



รูปที่ 8 สเปกโตรแกรมของเสียงไม่ก้อง "ส"

2. ตำแหน่งการออกเสียง (Place of Articulation) คือตำแหน่งฐานที่ทำให้เกิดเสียง ซึ่งเกิดจากการจำกัดช่องทางเดิน ณ ตำแหน่งต่างๆ ภายในปาก (Dew and Jensen, 1977) เช่น ริมฝีปาก ฟัน เหงือก เพดานอ่อน เป็นต้น ดังรูปที่ 9 โดยตำแหน่งที่ออกเสียงดังกล่าวมีผลต่อการเปลี่ยนแปลงตำแหน่งของฟอร์แมนต์ (Dew and Jensen, 1977; Deller, Proakis and Hansen, 1993)

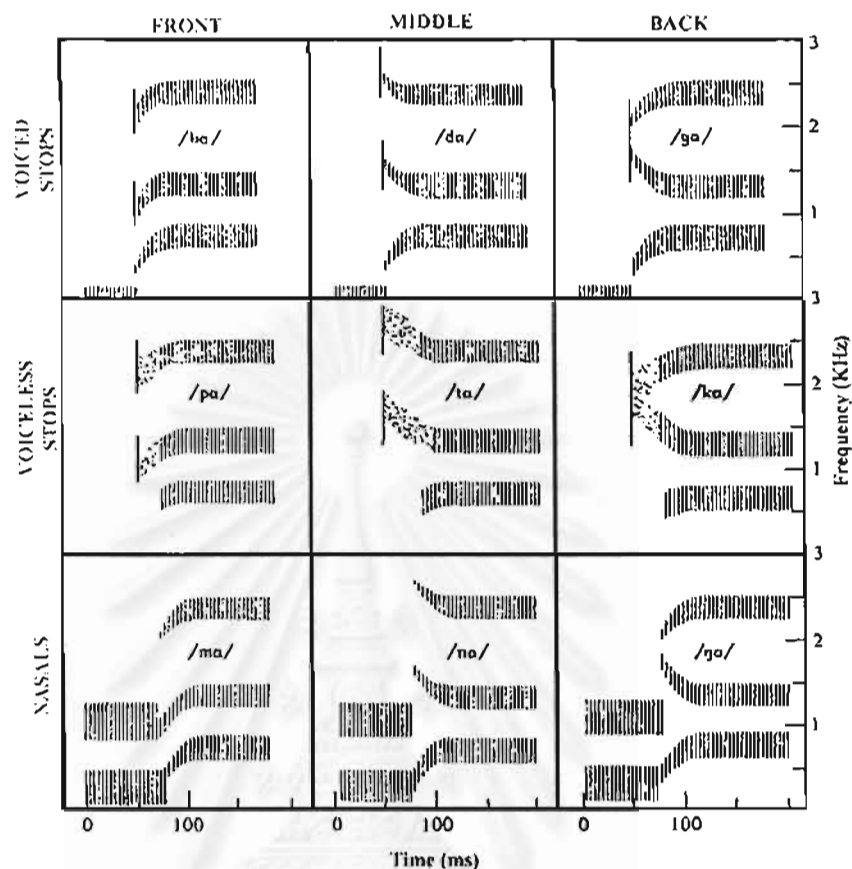


รูปที่ 9 ตำแหน่งที่ออกเสียง

3. ลักษณะการออกเสียง (Manner of Articulation) คือระดับการดัดแปลงลม ณ ตำแหน่งการออกเสียงต่างๆ ในช่องทางเดินเสียง โดยแสดงถึงขนาดของการดัดแปลงลมหรือการเปลี่ยนแปลงการดัดแปลงลมจากตำแหน่งการออกเสียงหนึ่งไปสู่ตำแหน่งการออกเสียงอีกตำแหน่งหนึ่ง (Dew and Jensen, 1977) การเปลี่ยนแปลงดังกล่าวเนื่องจากการเปลี่ยนแปลงตำแหน่งการออกเสียงจึงเป็นผลให้เกิดการเปลี่ยนแปลงฟอร์แมนต์หรือความถี่ก้ำทอนจากความถี่หนึ่งไปสู่อีกความถี่หนึ่งโดยมีอัตราการเปลี่ยนแปลงขึ้นกับอัตราการเปลี่ยนแปลงตำแหน่งการออกเสียงดังกล่าว

จากผลของตำแหน่งการออกเสียงและลักษณะการออกเสียงที่มีผลให้ฟอร์แมนต์หรือความถี่ก้ำทอนเกิดการบิดเบนตำแหน่งไป เมื่อพิจารณาเรื่องการรับรู้เสียงจากการทดสอบเสียงทางโสตศาสตร์พบว่า การบิดเบนดังกล่าวทำให้เกิดการรับรู้และแยกแยะเสียงพยัญชนะต่างๆ (Lieberman, 1988; Pickett, 1980) ดังตัวอย่างในรูปที่ 10 แสดงตัวอย่างเสียงพยัญชนะต่างๆ เปรียบเทียบกับการบิดเบนฟอร์แมนต์หรือความถี่ก้ำทอน ในเสียงพยัญชนะภาษาอังกฤษ และระยะเวลาในการบิดเบนก็มีผลต่อการรับรู้และแยกแยะดังกล่าวด้วย





รูปที่ 10 การบิดเบนฟอร์แมนต์ของเสียงพยัญชนะ

### 2.2.3 เสียงพูดภาษาไทย

ในภาษาไทยโดยทั่วไปมีโครงสร้างของเสียงพูดแต่ละพยางค์เป็น CCVC คือมีเสียงพยัญชนะต้น เสียงพยัญชนะตาม เสียงสระ หรือเสียงสระผสม และเสียงพยัญชนะสะกด และในภาษาไทยมีคุณสมบัติเด่นคือ คือเสียงวรรณยุกต์มีผลต่อความหมายของคำ สำหรับเสียงแต่ละเสียงในภาษาไทยมีรายละเอียดดังต่อไปนี้

1. เสียงสระ (Vowel Sounds) เกิดจากการวางตำแหน่งและความสูงของลิ้น และลักษณะของริมฝีปาก (Ince, 1992) ซึ่งในเสียงสระในภาษาไทยประกอบด้วยสระเดี่ยว และสระผสมดังนี้

- 1.1 สระเดี่ยว (Monophthongs) มีทั้งหมด 18 ตัว แบ่งเป็นสองกลุ่มคือสระเสียงสั้น 9 ตัวคือ อิ อี อุ เอะ เอะ โอะ แอะ อะ เอาะ และสระเสียงยาว 9 ตัว คือ อี อื อู เออ เออ ออ โดยสระเดี่ยวทั้งสองกลุ่มจะจับคู่กันได้ 9 คู่ ซึ่งแต่

ละคูมีตำแหน่งลิ้นและลักษณะฟอร์แมนต์เหมือนกันเป็นคู่ แต่ต่างกันในเรื่องระยะเวลาการออกเสียง (Luksaneeyanawin, 1993) ดังในตารางที่ 1

ความสูงลิ้น (Tongue Height)	ตำแหน่ง (Tongue Advancement)		
	หน้า (Front)	กลาง (Central)	หลัง (Back)
สูง (High)	i, ii อิ, อี	v, vv อี, อี	u, uu อุ, อู
กลาง (Mid)	e, ee เอะ, เอ	q, qq เออะ, เออ	o, oo โอะ, โอ
ต่ำ (Low)	x, xx แอะ, แอ	a, aa อะ, อา	@, @@ เออะ, ออ

ตารางที่ 1 สระเดี่ยว

1.2 สระผสม (Diphthongs) คือสระที่เกิดจากการเรียงต่อกันเป็นลำดับของสระเดี่ยวสองตัวซึ่งในภาษาไทยมีสระผสมทั้งหมด 6 ตัวโดยมีลักษณะเป็นคู่เสียงสั้นและยาวเช่นเดียวกับสระเดี่ยวดังตารางที่ 2

	ตำแหน่งลิ้นของเสียงสระตัวหน้า		
	หน้า (Front)	กลาง (Central)	หลัง (Back)
สระผสม (Diphthongs)	ia, iia เอียะ, เอีย	va, vva เอือะ, เอือ	ua, uua อัวะ, อิว

ตารางที่ 2 สระผสม

2. เสียงพยัญชนะ (Consonant Sounds) แบ่งได้เป็น 2 กลุ่มคือ

1. เสียงพยัญชนะต้น มีทั้งหมด 21 เสียง โดยแต่ละเสียงจะมีลักษณะเชิงความถี่และฟอร์แมนต์แตกต่างกันโดยแบ่งตามตำแหน่งการออกเสียงและลักษณะการออกเสียงได้ดังตารางที่ 3

ลักษณะการออกเสียง (Manner of Articulation)		ตำแหน่งการออกเสียง (Place of Articulation)				
		ริมฝีปาก (Labial)	ปุ่มเหงือก (Alveolar)	เพดานบน (Palatal)	เพดาน อ่อน (Velar)	เส้นเสียง (Glottal)
เสียงกัก (Stop)	ไม่ก้อง ไม่พ่นลม (Voiceless Unaspirated)	p ป	t ต, ฏ	c จ	k ก	? อ
	ไม่ก้อง พ่นลม (Voiceless Aspirated)	ph พ, ภ, ผ	th ท, ฑ, ฒ, ท, ถ, ฐ	ch ช, ฌ, ฉ	kh ค	
	ก้อง (Voiced)	b บ	d ด, ฎ, ฑ			
เสียงไม่กัก (Non-Stop)	นาสิก (Nasal)	m ม	n น, ณ		ng ง	
	เสียดแทรก (Fricative)	f ฟ, ฝ	s ซ, ศ, ษ, ส			h ฮ, ห
	รัว (Trill)		r ร, ฤ			
	ข้างลิ้น (Lateral)		l ล, ฤ			
	อัมสวะ (Approximant)	w ว		j ย, ญ		

ตารางที่ 3 เสียงพยัญชนะภาษาไทย

2. เสียงพยัญชนะควบกล้ำ (Double Consonant Sounds) คือเสียงพยัญชนะต้นลงเสียงออกเสียงต่อกัน โดยมีทั้งหมด 12 เสียงดังในตารางที่ 4

กลุ่มกักไม่พ่นลม (Unaspirated Stop Set)	pr ปร, ปฤ	tr ตร, ตฤ	kr กร, กฤ
	pl ปล		kl กล
			kw กว
กลุ่มกักไม่พ่นลม (Aspirated Stop Set)	phr พร, พฤ, ภร, ภฤ	thr ทร, ทฤ	chr คร, ชร
	phi พล, ผล		khl คล, ขล
			khw คว, ขว

ตารางที่ 4 เสียงพยัญชนะควบกล้ำ

สำหรับเสียงพยัญชนะในภาษาไทยที่สามารถออกเสียงในตำแหน่งท้ายพยางค์เพื่อทำหน้าที่เป็นเสียงตัวสะกดมีทั้งหมด 8 เสียงดังในตารางที่ 5 ในกลุ่มอัมสระเมื่อพิจารณาตามเสียงจริงที่เกิดขึ้นพบว่าเสียง "ว" ("w") เกิดจากเสียงสระตามด้วยเสียงสระอุ ("u") และเสียง "ย" ("j") เกิดจากเสียงสระตามด้วยเสียงสระอิ ("i") (Luksaneeyanawin, 1993)

ลักษณะการออกเสียง (Manner of Articulation)	ตำแหน่งการออกเสียง (Place of Articulation)			
	ริมฝีปาก (Labial)	ปุ่มเหงือก (Alveolar)	เพดานบน (Palatal)	เพดานอ่อน (Velar)
กัก (Stops)	p บ, ป, พ, ภ, ฟ	t ด, ถ, ต, ฏ, ท, ฐ, ฒ, ฑ, ฒ, ฐ, ฎ, ฏ, ช, ฅ, ฆ, ฌ		k ก, ต, ฏ, ข
นาสิก (Nasals)	m ม	n น, ณ, ร, ล, ฬ, ญ		ng ง
อัมสระ (Approximants)	w ว		j ย	

ตารางที่ 5 เสียงพยัญชนะสะกด

3. เสียงวรรณยุกต์ (Tones) (Luksaneeyanawin, 1993) ในภาษาไทยมี 5 เสียง แต่เนื่องจากในบางคำมีรูปเขียนและเสียงวรรณยุกต์ที่เกิดขึ้นจริงไม่ตรงกันเช่น ในกลุ่มอักษรสูง และอักษรต่ำ เป็นต้น ดังนั้นการอ้างอิงชื่อของเสียงวรรณยุกต์จึงอ้างอิงตามลักษณะของความถี่มูลฐานที่เกิดขึ้นจริงซึ่งอาจไม่สัมพันธ์กับรูปเขียน ดังตัวอย่างในรูปที่ 11 และแบ่งตามลักษณะการเปลี่ยนแปลงได้ดังนี้

1. วรรณยุกต์คงระดับ ประกอบด้วย 3 เสียงคือ

1.1 ระดับสูง (High) หรือเสียงวรรณยุกต์ตรี

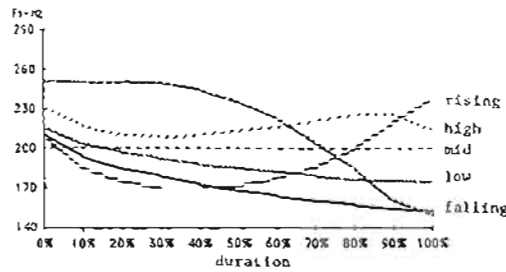
1.2 ระดับกลาง (Mid) หรือเสียงวรรณยุกต์สามัญ

1.3 ระดับต่ำ (Low) หรือเสียงวรรณยุกต์เอก

2. วรรณยุกต์เปลี่ยนระดับ ประกอบด้วย 2 เสียงคือ

2.1 เปลี่ยนขึ้น (Rising) หรือเสียงวรรณยุกต์จัตวา

2.2 เปลี่ยนตก (Falling) หรือเสียงวรรณยุกต์โท



รูปที่ 11 เสียงวรรณยุกต์และความถี่มูลฐาน

### 2.3 ทฤษฎีทางด้านการวิเคราะห์เสียงและการสังเคราะห์เสียง

ในการวิเคราะห์สัญญาณเสียง โดยพื้นฐานสัญญาณเสียงมีคุณสมบัติแปรเปลี่ยนไปตามเวลา (Non Stationary) ดังนั้นการวิเคราะห์หาคุณสมบัติต่างๆ ของเสียงจึงต้องทำในช่วงเวลาดั้ง (Short Time) ซึ่งคุณสมบัติชั่วคราวของเสียงในระยะเวลาดั้งสามารถอ้างได้ว่าคงที่ในช่วงระยะเวลา 10 ถึง 30 มิลลิวินาที (Rabiner and Schafer, 1978) จากคุณสมบัติของเสียงดังกล่าวนำไปสู่วิธีการวิเคราะห์ต่างๆ ดังต่อไปนี้

#### 2.3.1 การวิเคราะห์แบบหน้าต่าง (Windowing)

ในการวิเคราะห์ในช่วงเวลาดั้งของสัญญาณเสียงที่มีลักษณะแปรเปลี่ยนไปตามเวลา จึงต้องมีการเลือกช่วงหนึ่งของสัญญาณมาทำการวิเคราะห์ ซึ่งสามารถอ้างได้ว่าไม่แปรเปลี่ยนกับเวลา ช่วงเวลาที่เลือกเพื่อนำมาวิเคราะห์ดังกล่าวเรียกว่า หน้าต่าง (Window) (Deller et al., 1993) ในเชิงเวลา ซึ่งการวิเคราะห์แบบหน้าต่างดังกล่าวมีผลต่อการบิดเบี้ยวของสัญญาณเนื่องมาจากสัญญาณบริเวณของหน้าต่างเกิดจากการตัดปลายจากสัญญาณเสียงทำให้เกิดความไม่ต่อเนื่องและขาดหายไป ดังนั้นจึงต้องเลือกหน้าต่างที่เหมาะสมกับข้อมูลที่ต้องการวิเคราะห์เพื่อให้ข้อมูลที่ได้มีความผิดพลาดต่ำที่สุด

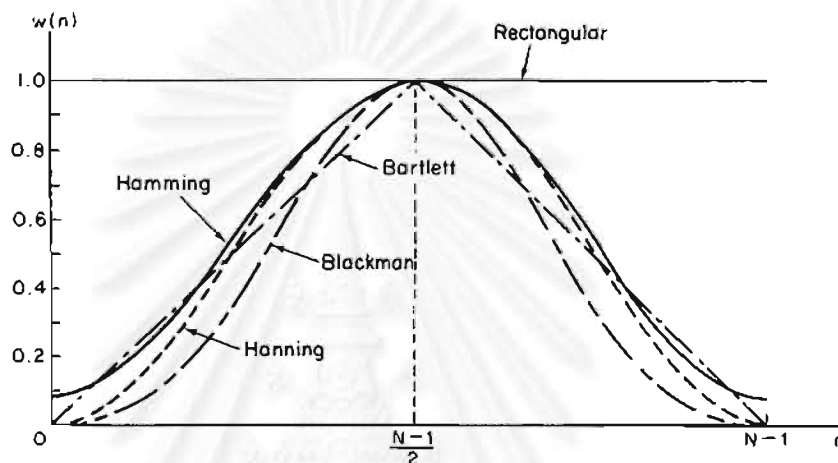
รูปแบบของหน้าต่างที่ใช้ในการวิเคราะห์มีหลายรูปแบบดังตัวอย่างในรูปที่ 12 สำหรับตัวอย่างของหน้าต่างที่ใช้กันทั่วไปมีดังนี้

หน้าต่างสี่เหลี่ยม (Rectangular Window) มีลักษณะตามสมการ 2.1

$$w(n) = \begin{cases} 1 & , 0 < n < N - 1 \\ 0 & , n \text{ มีค่าอื่นๆ} \end{cases} \quad (2.1)$$

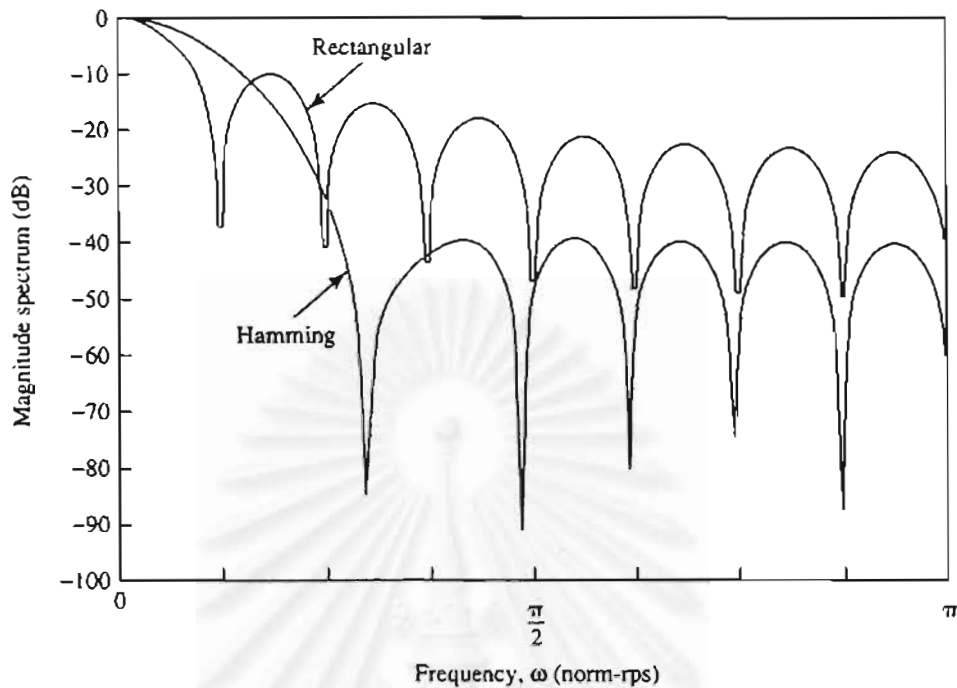
หน้าต่างแฮมมิง (Hamming Window) มีลักษณะตามสมการ 2.2

$$w(n) = 0.54 - 0.5 \cos\left(\frac{2\pi n}{N-1}\right) \quad \text{เมื่อ } 0 < n < N-1 \quad (2.2)$$



รูปที่ 12 ลักษณะหน้าต่างที่ใช้โดยทั่วไป

โดยหน้าต่างรูปแบบต่างๆ ที่ใช้กันทั่วไปมีลักษณะสเปกตรัมแบบผ่านต่ำ (Lowpass) ที่ด้านความถี่ต่ำและค่อยๆ ลดทอนที่ความถี่สูงขึ้น (Deller et al., 1993) ดังในรูปที่ 13 แสดงการเปรียบเทียบลักษณะสเปกตรัมของหน้าต่างสี่เหลี่ยมและหน้าต่างแฮมมิง เมื่อพิจารณาหน้าต่างสี่เหลี่ยมพบว่ามีความแบนด์วิดท์ (Bandwidth) แคบกว่าหน้าต่างแฮมมิงทำให้หน้าต่างสี่เหลี่ยมแสดงรายละเอียดเชิงความถี่ชัดกว่า แต่ทางด้านสัญญาณรบกวนพบว่าหน้าต่างแฮมมิงลดสัญญาณรบกวนได้ดีกว่าเนื่องจากการลดความไม่ต่อเนื่องบริเวณขอบหน้าต่าง ดังนั้นในการเลือกหน้าต่างต้องพิจารณาจุดแลกเปลี่ยนระหว่างความชัดเจนของรายละเอียดกับสัญญาณรบกวนเนื่องจากบริเวณขอบ (Deller et al., 1993) ซึ่งโดยการวิเคราะห์เสียงโดยทั่วไปนิยมใช้หน้าต่างแฮมมิงในการวิเคราะห์



รูปที่ 13 สเปกตรัมของหน้าต่างสี่เหลี่ยมเปรียบเทียบกับหน้าต่างแฮมมิง

### 2.3.2 การแปลงฟูรีเยร์แบบไม่ต่อเนื่อง (Discrete Fourier Transform)

การแปลงฟูรีเยร์แบบไม่ต่อเนื่อง (Oppenheim and Schaffer, 1975) ใช้ในการวิเคราะห์หาสัญญาณในเชิงความถี่หรือสเปกตรัมของสัญญาณที่ไม่ต่อเนื่องในเชิงเวลาและมีช่วงเวลาจำกัด การแปลงฟูรีเยร์แบบเต็มหน่วยของสัญญาณที่ไม่ต่อเนื่องในเชิงเวลาซึ่งเป็นสัญญาณที่เกิดจากการสุ่มและจำกัดจำนวนการสุ่มจะให้ผลการแปลงเป็นลำดับสัญญาณที่ไม่ต่อเนื่องในเชิงความถี่ และมีจำนวนตัวอย่างเท่ากับในเชิงเวลา สำหรับการแปลงฟูรีเยร์แบบเต็มหน่วยของสัญญาณสุ่มที่มี  $N$  ตัวอย่างจากเชิงเวลาเป็นเชิงความถี่เป็นตาม สมการ 2.3

$$\tilde{X}(k) = \sum_{n=0}^{N-1} \tilde{x}(n) e^{-j2\pi nk/N} \quad (2.3)$$

สำหรับการแปลงกลับจากสัญญาณเชิงความถี่เป็นเชิงเวลาเป็นไปตามสมการ 2.4

$$\tilde{x}(n) = \frac{1}{N} \sum_{k=0}^{N-1} \tilde{X}(k) e^{j2\pi nk/N} \quad (2.4)$$

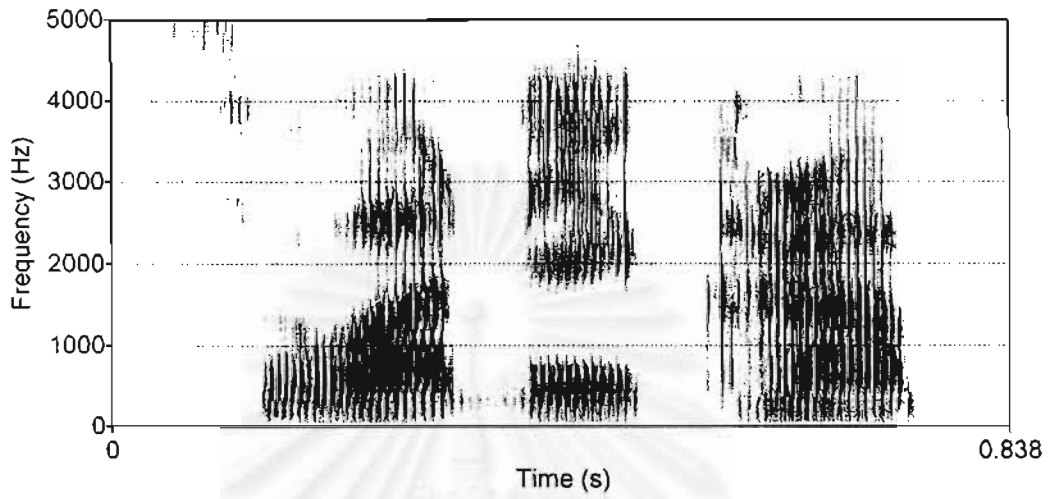


### 2.3.3 สเปกโตรแกรม (Spectrogram)

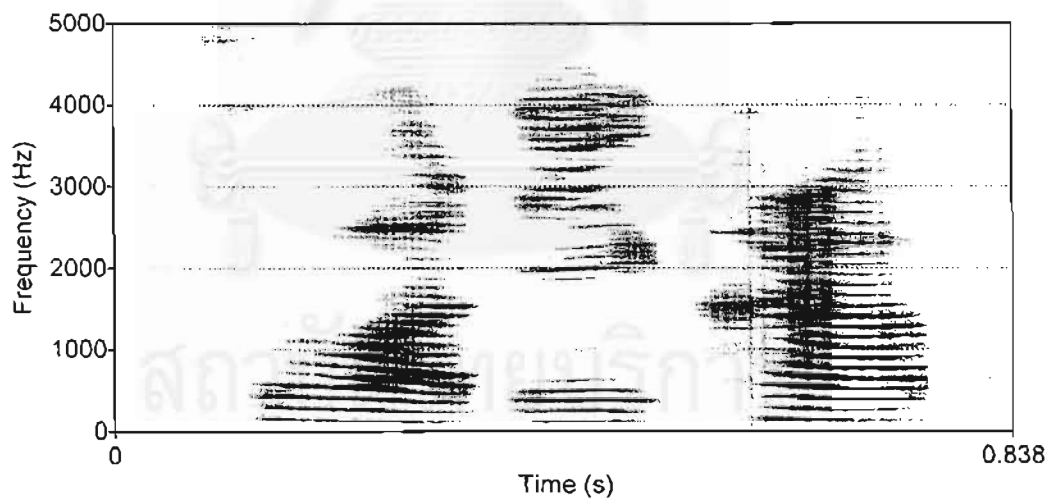
สเปกโตรแกรมเป็นรูปแบบหนึ่งของการแสดงสัญญาณเสียงเพื่อนำไปใช้ในการวิเคราะห์ที่นิยมใช้และมีความสำคัญสำหรับงานด้านนี้เป็นอย่างมาก การแสดงผลชนิดนี้ได้จากการคำนวณการแปลงฟูริเยร์ที่ขึ้นกับเวลา ผลที่แสดงคือกราฟสองมิติที่แสดงความสัมพันธ์ระหว่างสเปกตรัมที่ขึ้นกับเวลาในช่วงเวลานั้น เทียบกับเวลา (Rabiner and Schafer, 1978) โดยในแกนตั้งแสดงถึงสเปกตรัมและแกนนอนแสดงเวลา สำหรับขนาดสเปกตรัมแสดงด้วยสีที่มีระดับความเข้มแตกต่างกันซึ่งสามารถนำไปใช้ในการวิเคราะห์หาลักษณะที่สำคัญของเสียงและลักษณะเด่นสำหรับการรับรู้เสียงเช่น ความก้องของเสียง โครงสร้างฟอร์แมนต์ การเน้นเสียงและพิทช์ของเสียงเป็นต้น (Flanagan, 1972) ในการวิเคราะห์โดยใช้ สเปกโตรแกรมโดยทั่วไปแบ่งเป็น 2 แบบคือ

สเปกโตรแกรมแถบกว้าง (Wideband Spectrogram) เป็นการวิเคราะห์ที่แสดงรายละเอียดชั่วขณะในเชิงเวลาได้ดี แต่มีความละเอียดทางแกนความถี่ต่ำ (Robinson, 1998) ดังแสดงในรูปที่ 14 กราฟชนิดนี้แสดงถึงการเปลี่ยนแปลงของฟอร์แมนต์และแสดงคาบพิทช์โดยดูจากลายเส้นในแนวตั้งได้ (Rabiner and Schafer, 1978)

สเปกโตรแกรมแถบแคบ (Narrowband Spectrogram) เป็นการวิเคราะห์ที่แสดงรายละเอียดทางด้านฮาร์โมนิคได้ดี แต่แสดงรายละเอียดชั่วขณะในเชิงเวลาไม่ชัดเจน (Robinson, 1998) ดังแสดงในรูปที่ 15 กราฟชนิดนี้สามารถนำไปหาความถี่มูลฐานได้โดยการวิเคราะห์จากฮาร์โมนิคของความถี่มูลฐานและใช้ในการแยกช่วงที่เป็นเสียงก้องออกจากเสียงไม่ก้องได้ (Rabiner and Schafer, 1978)



รูปที่ 14 สเปกโตรแกรมแถบกว้าง



รูปที่ 15 สเปกโตรแกรมแถบแคบ

### 2.3.4 เอเอ็มดีเอฟ (Average Magnitude Difference Function : AMDF)

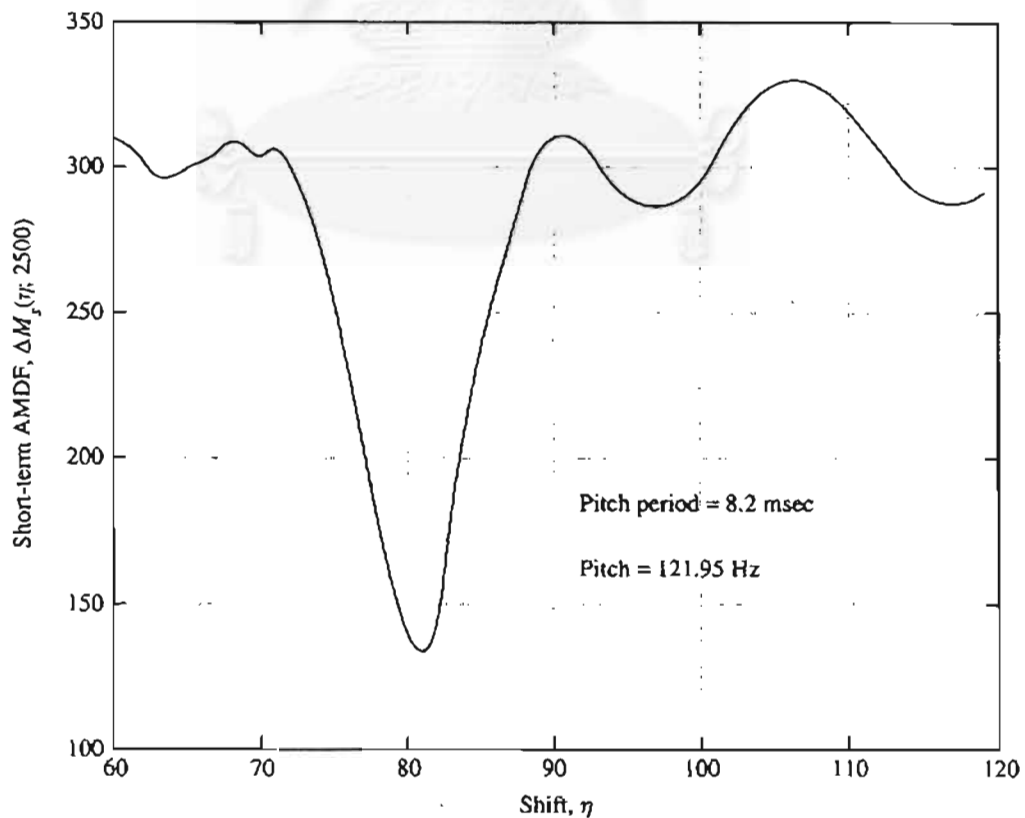
เอเอ็มดีเอฟเป็นวิธีที่ใช้ประมาณหาคาบพิทช์ของสัญญาณโดยอาศัยแนวความคิดที่ว่า สำหรับสัญญาณคาบที่มีคาบเท่ากับ  $P$  จะทำให้ความสัมพันธ์ตามสมการ 2.5 มีค่าเท่ากับศูนย์เมื่อ  $\eta = 0, \pm P, \pm 2P, \dots$

$$\Delta M_s(\eta) = |s(n) - s(n - \eta)| \quad (2.5)$$

สำหรับเสียงพูดที่เป็นเสียงก้อง ค่า  $\Delta M_s(\eta)$  ควรจะมีค่าน้อยทุกๆตำแหน่งที่เป็นจำนวนเท่าของคาบพิทช์ นั่นคือค่า  $\Delta M_s(\eta)$  ควรจะน้อยเมื่อ  $\eta$  มีค่าใกล้ค่าคาบพิทช์ (Rabiner and Schafer, 1978) จากหลักการดังกล่าวเขียนเป็นสมการของเอเอ็มดีเอฟได้ตามสมการ 2.6 โดยที่  $w(m-n)$  คือหน้าต่างที่ตำแหน่ง  $m-n$

$$\Delta M_s(\eta : m) = \frac{1}{N} \sum_{n=m-N+1}^m |s(n) - s(n - \eta)| w(m-n) \quad (2.6)$$

ผลที่ได้จากการทำด้วยวิธีนี้แสดงดังในรูปที่ 16 โดยค่า  $\eta$  ที่ได้มาสามารถนำไปคำนวณหาค่าความถี่มูลฐานได้ต่อไป



รูปที่ 16 ผลจากวิธีเอเอ็มดีเอฟ

### 2.3.5 การวิเคราะห์และการสังเคราะห์ด้วยการทำนายแบบเชิงเส้น

(Linear Predictive Analysis and Synthesis)

การทำนายแบบเชิงเส้นเป็นเทคนิคที่สำคัญทางด้านการวิเคราะห์เสียงและการสังเคราะห์เสียงเนื่องจากมีความแม่นยำสูงในการประมาณค่าพารามิเตอร์ของเสียงพูดเมื่อเทียบกับความเร็วในการประมวลผล หลักการพื้นฐานของการการทำนายแบบเชิงเส้นอาศัยแนวคิดที่ว่าตัวอย่างสัญญาณเสียงพูดสามารถประมาณค่าได้จากผลรวมของตัวอย่างสัญญาณเสียงพูดจากอดีต (Rabiner and Schafer, 1978) การวิเคราะห์หาพารามิเตอร์เพื่อใช้ในการทำนายโดยทั่วไปเรียกว่า การเข้ารหัสการทำนายแบบเชิงเส้นหรือแอลพีซี (Linear Predictive Coding : LPC) ในทางด้านการประมวลผลสัญญาณเสียง การเข้ารหัสการทำนายแบบเชิงเส้นนี้ถูกนำไปใช้ในสองแนวทาง (Kleijn and Paliwal, 1995) คือ

1. การเข้ารหัสสัญญาณเสียง โดยที่การเข้ารหัสการทำนายแบบเชิงเส้นถูกนำไปใช้เป็นวงจรกรองวิเคราะห์การทำนายแบบเชิงเส้น (LP Analysis Filter) เพื่อแยกส่วนซ้ำซ้อน (Redundancy) ของสัญญาณเสียงออก ส่วนที่เหลือเรียกว่าสัญญาณตกค้าง (Residual Signal)

2. การสังเคราะห์สัญญาณเสียงพูด โดยการเข้ารหัสการทำนายแบบเชิงเส้นถูกนำไปใช้เป็นวงจรกรองการทำนายแบบเชิงเส้นผกผัน (Inverse LP Filter) หรือวงจรกรองสังเคราะห์การทำนายแบบเชิงเส้น (LP Synthesis Filter) โดยที่ฟังก์ชันถ่ายโอนของ วงจรกรองดังกล่าวแสดงกรอบสเปกตรัมของสัญญาณเสียงพูด วงจรกรองสังเคราะห์การทำนายแบบเชิงเส้นถูกใช้แสดงแทนช่องทางเดินเสียงของมนุษย์และใช้หาสัญญาณกระตุ้นที่เหมาะสมสำหรับวงจรกรองดังกล่าว

ในการวิเคราะห์การเข้ารหัสการทำนายแบบเชิงเส้นเริ่มต้นจากพิจารณากรอบสัญญาณเสียงที่มี  $N$  ตัวอย่าง คือ  $s_1, s_2, \dots, s_N$  โดยอ้างว่าตัวอย่างสัญญาณปัจจุบันสามารถทำนายได้จากผลรวมของตัวอย่างสัญญาณในอดีต  $p$  ตัวอย่าง ดังสมการ 2.7

$$\tilde{s}_n = -a_1 s_{n-1} - a_2 s_{n-2} - a_3 s_{n-3} - \dots - a_p s_{n-p} = -\sum_{k=1}^p a_k s_{n-k} \quad (2.7)$$

เมื่อ  $p$  คืออันดับของการวิเคราะห์การทำนายแบบเชิงเส้นและ  $a_1, a_2, \dots, a_p$  คือสัมประสิทธิ์การเข้ารหัสการทำนายแบบเชิงเส้น กำหนด  $e_n$  แทนค่าผิดพลาดระหว่างค่าจริงและค่าที่ทำนายได้จะได้ตามสมการ 2.8

$$e_n = s_n - \tilde{s}_n \quad (2.8)$$

$$e_n = s_n + \sum_{k=1}^p a_k s_{n-k} \quad (2.9)$$

สัญญาณ  $e_n$  เรียกว่าสัญญาณตกค้าง เนื่องจากสัญญาณ  $e_n$  ได้จากผลการลบสัญญาณ  $s_n$  ด้วย  $\tilde{s}_n$  เนื่องจากค่าสหสัมพันธ์ช่วงสั้น (Short-term Correlation) ระหว่างตัวอย่างของสัญญาณตกค้างมีค่าต่ำ ดังนั้นประมาณได้ว่ากรอบสเปกตรัมกำลังของสัญญาณตกค้างมีลักษณะเรียบ (flat) เมื่อทำการแปลงแซด (Z Transform) ของสมการ 2.9 ได้ดังสมการ 2.10 และ

$$E(z) = A(z)S(z) \quad (2.10)$$

โดยที่  $S(z)$  และ  $E(z)$  เป็นผลการแปลงแซดของสัญญาณเสียงและสัญญาณตกค้างตามลำดับและ

$$A(z) = 1 + \sum_{k=1}^p a_k z^{-k} \quad (2.11)$$

โดยที่  $A(z)$  หรือวงจรกรองไวเทนนิง (Whitening Filter) มีหน้าที่แยกความสหสัมพันธ์ช่วงสั้นที่ปรากฏในสัญญาณเสียงพูดซึ่งก็คือการทำให้สเปกตรัมเรียบ เนื่องจาก  $E(z)$  ประมาณได้ว่ามีสเปกตรัมเรียบดังนั้นสามารถออกแบบกรอบสเปกตรัมของสัญญาณช่วงสั้นได้จากการวิเคราะห์การทำนายแบบเชิงเส้นรูปแบบออลโพล (All-pole Model) หรือ รูปแบบออตรีเกรสซีฟ (Autoregressive Model) ดังสมการ 2.12

$$H(z) = \frac{1}{A(z)} \quad (2.12)$$

วงจรกรอง  $A(z)$  เรียกได้อีกชื่อหนึ่งว่าวงจรกรองผกผัน (Inverse Filter) เนื่องจากเป็นส่วนผกผันของรูปแบบออลโพล  $H(z)$  ของสัญญาณเสียงพูด และรากของ  $A(z)$  ทำให้เกิดโพลใน  $H(z)$  นั่นคือตำแหน่งฟอร์แมนต์ของเสียงที่ได้จากช่องทางเสียงที่มีฟังก์ชันถ่ายโอน  $H(z)$

การหากรอบสเปกตรัมกำลังช่วงสั้นของเสียงพูดด้วยวิธีการวิเคราะห์การทำนายแบบเชิงเส้นคำนวณได้จาก  $H(z)$  บนวงกลมหนึ่งหน่วย โดยในขั้นแรกต้องหาสัมประสิทธิ์การทำนายแบบเชิงเส้นของสัญญาณเสียงพูดก่อน โดยปกติหาได้จากการทำให้ค่าผิดพลาดการทำนายแบบเชิงเส้นทั้งหมดยกกำลังสองดังสมการ 2.13 มีค่าต่ำที่สุด

$$E = \sum_{n=n_1}^{n_2} e_n^2 \quad (2.13)$$

โดยที่ผลรวมของช่วง  $n_1$  ถึง  $n_2$  ที่คำนวณได้ขึ้นอยู่กับวิธีการที่ใช้ซึ่งมีอยู่สองวิธี (Kleijn and Paliwal, 1995) คือ

1. วิธีอัตสหสัมพันธ์ (Autocorrelation Method) ในการวิเคราะห์การทำนายแบบเชิงเส้นช่วงสั้นหาได้โดยการวิเคราะห์แบบหน้าต่างสัญญาณเสียงพูดและอ้างว่าตัวอย่างสัญญาณภายนอกหน้าต่างนี้มีค่าเท่ากับศูนย์ แล้วจึงทำให้ค่าผิดพลาดตามสมการ 2.13 มีค่าต่ำสุดซึ่งได้ตามสมการ 2.14

$$\sum_{k=1}^p r_{|i-k|} a_k = -r_i \quad \text{เมื่อ } 1 \leq i \leq p \quad (2.14)$$

โดยที่  $r_k$  คือสัมประสิทธิ์อัตสหสัมพันธ์อันดับที่  $k$  ของหน้าต่างสัญญาณเสียงโดยที่

$$r_k = \frac{1}{N} \sum_{n=k}^N w_n s_n w_{n-k} s_{n-k} \quad (2.15)$$

เมื่อ  $\{w_n\}$  คือฟังก์ชันหน้าต่างที่มีระยะเวลา  $m$  ตัวอย่าง

การหาค่าสัมประสิทธิ์การเข้ารหัสการทำนายแบบเชิงเส้นสามารถหาได้จากการแก้สมการ 2.14 ซึ่งมีจำนวน  $p$  สมการ สมการดังกล่าวเรียกว่า สมการยูล-วอล์กเกอร์ (Yule-Walker) สมการทั้งหมดสามารถเขียนรูปเมตริกได้ดังนี้

$$\mathbf{R}\mathbf{a} = -\mathbf{r} \quad (2.16)$$

โดยที่

$$\mathbf{R} = \begin{bmatrix} r_0 & r_1 & r_2 & \cdots & r_{p-1} \\ r_1 & r_0 & r_1 & \cdots & r_{p-2} \\ r_2 & r_1 & r_0 & \cdots & r_{p-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{p-1} & r_{p-2} & r_{p-3} & \cdots & r_0 \end{bmatrix} \quad (2.17)$$

$$\mathbf{a} = [a_1, a_2, \dots, a_p]^T \quad (2.18)$$

$$\mathbf{r} = [r_1, r_2, \dots, r_p]^T \quad (2.19)$$

เมตริก R เรียกว่า เมตริกอัตโนมัติสัมพันธ์ (Autocorrelation Matrix) ซึ่งมีโครงสร้างแบบโทพลิตซ์ (Toeplitz) โครงสร้างนี้รับรองว่าโพลของวงจรรองส่งเคราะห์การทำนายแบบเชิงเส้น  $H(z)$  จะอยู่ภายในวงกลมหนึ่งหน่วย ดังนั้นจึงรับรองได้ว่าวงจรรองส่งเคราะห์  $H(z)$  ที่ได้จากวิธีอัตโนมัติสัมพันธ์จะเสถียรเสมอ สำหรับวิธีที่ช่วยหาผลเฉลยสมการยูล-วอล์กเกอร์ในสมการ 2.14 และหาสัมประสิทธิ์การทำนายแบบเชิงเส้นในสมการ 2.16 มีหลายวิธีดังเช่น วิธีเลวินสัน-เดอบิน (Levinson-Durbin Algorithm) และวิธีเชอ (Schur Algorithm) เป็นต้น

2. วิธีโคแวนเรียนซ์ (Covariance Method) ในการวิเคราะห์การทำนายแบบเชิงเส้น ช่วงของการรวมอยู่ในช่วง  $(p+1, N)$  ดังนั้นจึงไม่จำเป็นต้องใช้หน้าต่าง การทำให้ค่าผิดพลาดทั้งหมดยกกำลังสองมีค่าต่ำที่สุดหาได้จากสมการ  $p$  สมการดังต่อไปนี้

$$\sum_{k=1}^p c_{ik} a_k = -c_{i0} \quad \text{เมื่อ } 1 \leq i \leq p \quad (2.20)$$

โดยที่

$$c_{ik} = \sum_{n=p+1}^N s_{n-i} s_{n-k} \quad (2.21)$$

สมการจำนวน  $p$  สมการสามารถเขียนในรูปเมตริกได้ดังนี้

$$\mathbf{C}\mathbf{a} = -\mathbf{c} \quad (2.22)$$

$$\mathbf{C} = \begin{bmatrix} c_{11} & c_{12} & c_{13} & \cdots & c_{1p} \\ c_{21} & c_{22} & c_{23} & \cdots & c_{2p} \\ c_{31} & c_{32} & c_{33} & \cdots & c_{3p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{p1} & c_{p2} & c_{p3} & \cdots & c_{pp} \end{bmatrix} \quad (2.23)$$

$$\mathbf{c} = [c_{10}, c_{20}, \dots, c_{p0}]^T \quad (2.24)$$

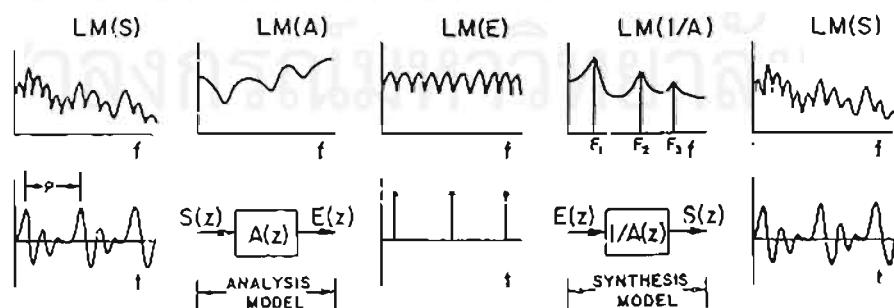
เมตริก  $c$  เรียกว่า เมตริกโคแวนเรียนซ์ (Covariance Matrix) และเป็นเมตริกสมมาตรนั่นคือ  $c_{n,m} = c_{m,n}$  แต่เมตริกไม่ได้มีโครงสร้างเป็นแบบโทพลิตซ์ ดังนั้นการหาสัมประสิทธิ์การเข้ารหัสการทำนายแบบเชิงเส้นด้วยวิธีนี้จึงมีประสิทธิภาพต่ำกว่าวิธีอัลทสสัมพันธ์ และสัมประสิทธิ์การเข้ารหัสการทำนายแบบเชิงเส้นที่ได้จากวิธีนี้จึงรับรองไม่ได้ว่าจะให้วงจรรองส่งเคราะห์ที่มีความเสถียรเสมอ นอกจากนี้โครงสร้างที่สมมาตรทำให้การคำนวณบางส่วนสามารถใช้เทคนิคการคำนวณให้เร็วขึ้นได้แต่ยังไม่เร็วเท่าวิธีเลวินสัน-เดอบิน

ในการสังเคราะห์สัญญาณเสียงด้วยการทำนายแบบเชิงเส้น โดยอาศัยสมการ 2.10 และ 2.12 จะได้วงจรรองส่งเคราะห์การทำนายแบบเชิงเส้นดังสมการต่อไปนี้

$$S(n) = \frac{1}{A(z)} \bullet E(z) \quad (2.25)$$

$$S(n) = H(z)E(s) \quad (2.26)$$

ดังสมการ 2.26 การสังเคราะห์สัญญาณเสียงทำได้โดยการผ่านสัญญาณตกค้างซึ่งทำหน้าที่เสมือนสัญญาณกระตุ้นในกระบวนการออกเสียงของมนุษย์เข้าสู่วงจรรองส่งเคราะห์การทำนายแบบเชิงเส้น  $H(z)$  ซึ่งจำลองรูปแบบช่องทางเดินเสียงและมีฟังก์ชันถ่ายโอนที่แสดงถึงกรอบสเปกตรัมของสัญญาณเสียงพูด สัญญาณเสียงสังเคราะห์ที่ได้จากการผ่านวงจรรองส่งนี้จะมีลักษณะสเปกตรัมตามกรอบสเปกตรัมของวงจรรองส่งเคราะห์ โดยสรุปขั้นตอนการวิเคราะห์และสังเคราะห์ด้วยการทำนายแบบเชิงเส้นแสดงได้ดังในรูปที่ 17 (Markel and Gray, 1976)



รูปที่ 17 ขั้นตอนการวิเคราะห์และสังเคราะห์การทำนายแบบเชิงเส้น



### 2.3.6 คู่เส้นสเปกตรัม (Line Spectrum Pair : LSP)

คู่เส้นสเปกตรัม หรือ ความถี่เส้นสเปกตรัม (Line Spectrum Frequency : LSF) เป็นพารามิเตอร์รูปแบบหนึ่งที่พัฒนามาจากพารามิเตอร์การทำนายแบบเชิงเส้น เนื่องจากพารามิเตอร์การทำนายแบบเชิงเส้นในขั้นตอนการประมาณค่าพารามิเตอร์อาจทำให้เกิดความไม่เสถียรของสัญญาณได้ซึ่งส่งผลต่อคุณภาพของเสียง ในขณะที่พารามิเตอร์คู่เส้นสเปกตรัมมีคุณสมบัติที่เด่นคือ ค่าพารามิเตอร์อยู่ภายในขอบเขตที่จำกัด มีการเรียงลำดับของค่าพารามิเตอร์และสามารถตรวจสอบเสถียรภาพของวงจรกรองได้ง่าย นอกจากนี้คู่เส้นสเปกตรัมยังแสดงในรูปเชิงความถี่จึงสามารถนำไปใช้ในการหาคุณสมบัติที่แน่นอนในระบบการรับรู้ของคนได้ (Kleijn and Paliwal, 1995)

ในการคำนวณหาคู่เส้นสเปกตรัมเริ่มต้นจากพหุนามอันดับ  $M$  ของวงจรกรองผกผันในเชิงแตรดิงสมการ 2.11 โดยทำการแยกส่วนสมการดังกล่าวเป็นพหุนามอันดับ  $M+1$  เป็นจำนวนสองพหุนามดังสมการ 2.27 และ 2.28

$$P(z) = A(z) + z^{-(M+1)} A(z^{-1}) \quad (2.27)$$

$$Q(z) = A(z) - z^{-(M+1)} A(z^{-1}) \quad (2.28)$$

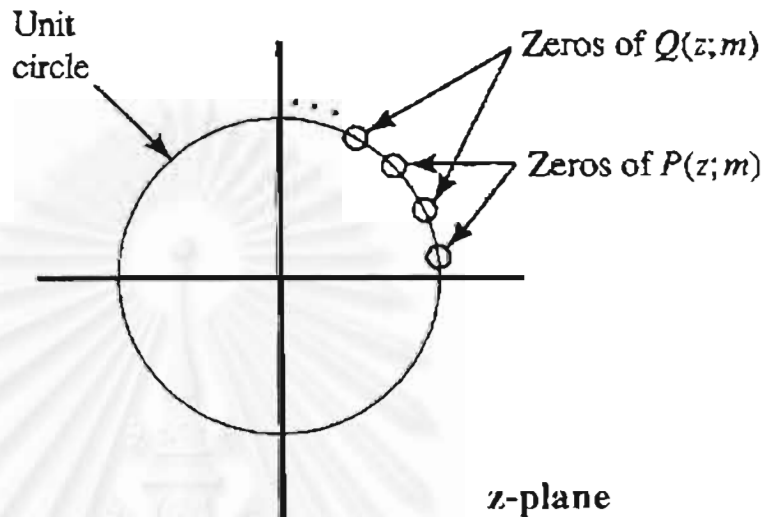
โดยพหุนาม  $P(z)$  และ  $Q(z)$  มีความสัมพันธ์กับ  $A(z)$  ตามสมการ 2.29

$$A(z) = \frac{P(z) + Q(z)}{2} \quad (2.29)$$

พหุนาม  $P(z)$  และ  $Q(z)$  สอดคล้องกับรูปแบบช่องทางเดินเสียงที่ไร้การสูญเสียขณะที่ช่องระหว่างเส้นเสียง (Glottis) ปิดและเปิดตามลำดับ (Deller et al., 1993) และรากของพหุนาม  $P(z)$  และ  $Q(z)$  เรียกว่าความถี่เส้นสเปกตรัม โดยพหุนามทั้งสองนี้มีคุณสมบัติดังต่อไปนี้ (Kleijn and Paliwal, 1995)

1. ศูนย์ (Zeros) หรือ รากทั้งหมดของพหุนาม  $P(z)$  และ  $Q(z)$  จะตั้งอยู่บนวงกลมหนึ่งหน่วยเสมอ

2. ศูนย์ของ  $P(z)$  และ  $Q(z)$  วางเรียงสลับซึ่งกันและกันตามลำดับจากน้อยไปมากดังแสดงในรูปที่ 18



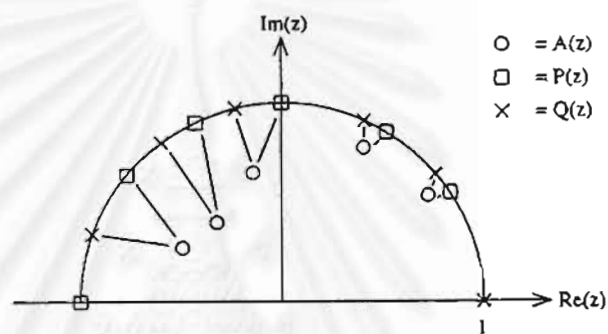
รูปที่ 18 การวางเรียงสลับของศูนย์ของพหุนามคู่เส้นสเปกตรัม  $P(z)$  และ  $Q(z)$

โดยสามารถแสดงได้ว่าวงจรรองผกผัน  $A(z)$  จะมีเฟสต่ำสุดถ้าคู่เส้นสเปกตรัมของ  $A(z)$  สอดคล้องกับเงื่อนไขทั้งสองนี้ ดังนั้นวงจрсังเคราะห์การเข้ารหัสการทำนายแบบเชิงเส้นสามารถยืนยันได้ว่ามีเสถียรภาพโดยการทำการประมวลผลพารามิเตอร์การเข้ารหัสการทำนายแบบเชิงเส้นในรูปแบบคู่เส้นสเปกตรัม (Kleijn and Paliwal, 1995)

เมื่อพิจารณารากของพหุนามทั้งสองพบว่าพหุนาม  $P(z)$  และ  $Q(z)$  มีรากจริงอยู่ที่  $-1$  และ  $1$  ตามลำดับ สำหรับรากอื่นๆ อยู่บนวงกลมหนึ่งหน่วยโดยวางเรียงสลับกันตามคุณสมบัติของพหุนามทั้งสอง และรากของทั้งสองพหุนามมีลักษณะเป็นคู่เชิงซ้อนสังยุค ดังนั้นในการเก็บรากของพหุนามเพื่อใช้เป็นพารามิเตอร์จึงเก็บเพียง  $M$  ค่า

เนื่องจากรากของพหุนาม  $A(z)$  แสดงตำแหน่งฟอร์แมนต์ของเสียงพูดและพหุนาม  $P(z)$  และ  $Q(z)$  สัมพันธ์กับ  $A(z)$  ตามสมการ 2.27 และ 2.28 ดังนั้นรากของพหุนามทั้งสองจึงสัมพันธ์กับฟอร์แมนต์ด้วย โดยรากของพหุนาม  $A(z)$  แต่ละอันจับคู่กับรากของพหุนาม  $P(z)$  และ  $Q(z)$  อย่างละหนึ่งราก (Deller et al., 1993) ดังในรูปที่ 19 การเกาะกลุ่มของรากหรือความถี่เส้นสเปกตรัมจำนวน 2 หรือ 3 ราก จะกำหนดลักษณะความถี่ฟอร์แมนต์และขนาดแบนด์วิดท์ของฟอร์แมนต์โดยขึ้นกับความชิดของกลุ่มรากดังกล่าว (Kleijn and Paliwal, 1995) ถ้า

รากชิดกันมากจะทำให้เกิดฟอร์แมนต์และมีแบนด์วิดท์แคบ ในทางตรงกันข้าม กลุ่มรากจะแสดงถึงสเปกตรัมที่มีแบนด์วิดท์กว้างคือไม่เกิดฟอร์แมนต์ (Deller et al., 1993) นอกจากนี้คุณสมบัติอีกอย่างหนึ่งของคู่เส้นสเปกตรัมคือความไวทางสเปกตรัมของคู่เส้นสเปกตรัมมีลักษณะเฉพาะที่คือเมื่อมีการเปลี่ยนแปลงคู่เส้นสเปกตรัมใดๆ จะทำให้เกิดการเปลี่ยนแปลงของสเปกตรัมกำลังของการเข้ารหัสการทำนายแบบเชิงเส้นเฉพาะ บริเวณรอบๆ เท่านั้น ทำให้สามารถทำการประเมินค่า (Quantization) ได้อย่างอิสระโดยไม่มีผลกระทบจากการลดทอนเนื่องจากการประเมินค่า จากสเปกตรัมหนึ่งไปสู่สเปกตรัมอื่นๆ (Kleijn and Paliwal, 1995)



รูปที่ 19 ความสัมพันธ์ระหว่างรากของ  $A(z)$  กับรากของคู่เส้นสเปกตรัม  $P(z)$  และ  $Q(z)$

### 2.3.7 การประมาณค่าสัมประสิทธิ์ (Interpolation of LP coefficients)

เนื่องจากค่าสัมประสิทธิ์ของวงจรรองส่งเคราะห์การทำนายแบบเชิงเส้นถูกปรับตลอดทุกๆ กรอบ การเปลี่ยนแปลงฉับพลันของค่าสัมประสิทธิ์นี้สามารถทำให้เกิดผลกระทบชั่วขณะซึ่งอาจมีผลต่อคุณภาพของเสียงที่ส่งเคราะห์ขึ้น การทำให้การเปลี่ยนแปลงค่าสัมประสิทธิ์วงจรรองการทำนายแบบเชิงเส้นมีความราบรื่น ณ เวลาต่างๆ สามารถทำได้โดยใช้การประมาณค่า (Interpolation) ของสัมประสิทธิ์เหล่านี้ โดยทำที่บริเวณขอบกรอบสัญญาณเพื่อให้ได้สัมประสิทธิ์ที่มีความราบรื่นสูงสุด (Kleijn and Paliwal, 1995)

ในทางปฏิบัติ ค่าสัมประสิทธิ์ในแต่ละกรอบย่อยสามารถหาได้จากสมการ 2.30

$$ST(i) = w_i \cdot ST_{future} + (1 - w_i) \cdot ST_{past} \quad (2.30)$$

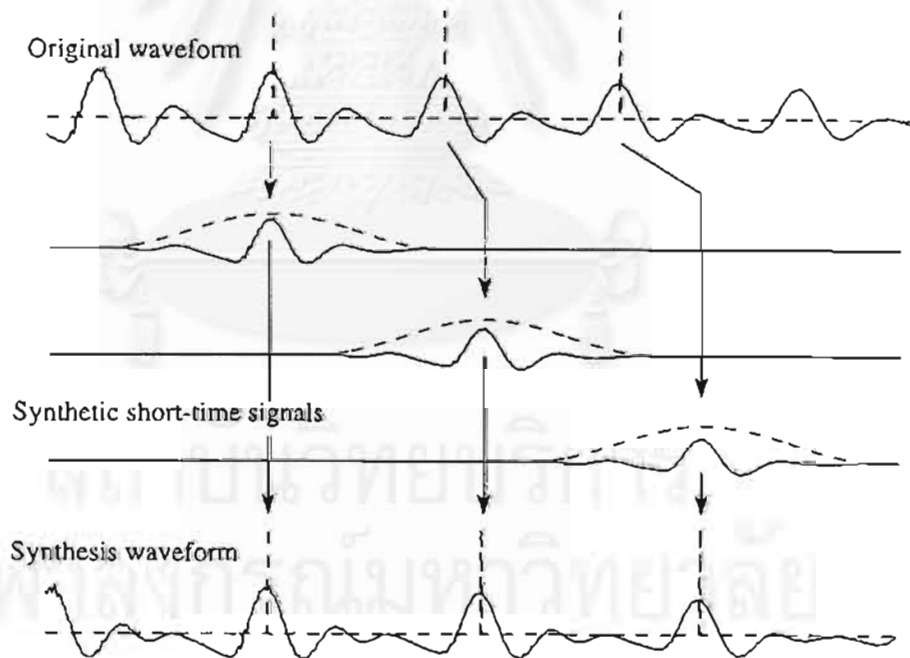
โดยที่  $w_i$  คือค่าน้ำหนักการประมาณค่า  $ST_{past}$  และ  $ST_{future}$  คือค่าสัมประสิทธิ์โดยประมาณของกรอบสัญญาณช่วงสั้นที่บริเวณขอบ

โดยทั่วไปการประมาณค่าสัมประสิทธิ์การทำนายแบบเชิงเส้นไม่กระทำบนสัมประสิทธิ์โดยตรง แต่จะทำบนเชิงอื่นๆ ที่ได้หลังจากการแปลงสัมประสิทธิ์ เนื่องจากสาเหตุแรกคือ มีรูปลักษณะอื่นๆ ของสัมประสิทธิ์ที่สามารถรับรองในเรื่องเสถียรภาพเมื่อทำการประมาณค่าได้ ในขณะที่การใช้สัมประสิทธิ์โดยตรงต้องมีการตรวจสอบเสถียรภาพในภายหลัง และอีกสาเหตุหนึ่งคือ พารามิเตอร์บางประเภทมีคุณสมบัติในการประมาณค่าได้ดีกว่าและพบว่าพารามิเตอร์เป็นตัวเลือกที่ดีคือ คู่เส้นสเปกตรัม และ พารามิเตอร์อัตราสัมพัทธ์ (Kleijn and Paliwal, 1995)

### 2.3.8 ทิดี-โซลา (Time-Domain Pitch-Synchronous Overlap Add : TD-PSOLA)

ทิดี-โซลา (Charpentier and Stella, 1986; Charpentier and Moulines, 1989) เป็นวิธีการที่ใช้ในการปรับทำนองของเสียงโดยการปรับตำแหน่งพิทช์ของเสียงเพื่อให้เสียงมีระยะเวลาและเสียงวรรณยุกต์ตามต้องการ

วิธีการนี้เริ่มจากการแยกเสียงพูดเป็นสัญญาณช่วงสั้นจำนวนมากโดยใช้การวิเคราะห์แบบหน้าต่างที่ตำแหน่งพิทช์ตลอดช่วงเสียงก้องหรือที่ตำแหน่งที่มีคาบเวลาคงที่สำหรับบริเวณที่เป็นเสียงไม่ก้อง แล้วนำสัญญาณช่วงสั้นดังกล่าวมารวมกันใหม่เพื่อสังเคราะห์เสียงขึ้นมาใหม่ การปรับพิทช์ทำได้โดยการปรับระยะระหว่างพิทช์ของสัญญาณช่วงสั้นระหว่างการสังเคราะห์ให้มีค่าเพิ่มหรือลดตามต้องการ สำหรับการปรับระยะเวลาทำในขณะเดียวกันโดยการทำซ้ำหรือลบสัญญาณช่วงสั้นออกจากเสียงสังเคราะห์ ในการรวมสัญญาณช่วงสั้นเข้าด้วยกันนั้นใช้หลักการบวกร้อน (Overlap-Add) สัญญาณเข้าด้วยกันโดยหลักการนี้เป็นการชดเชยแอมพลิจูดจากการวิเคราะห์แบบหน้าต่าง (Kleijn and Paliwal, 1995) วิธีการโดยสรุปแสดงดังในรูปที่ 20



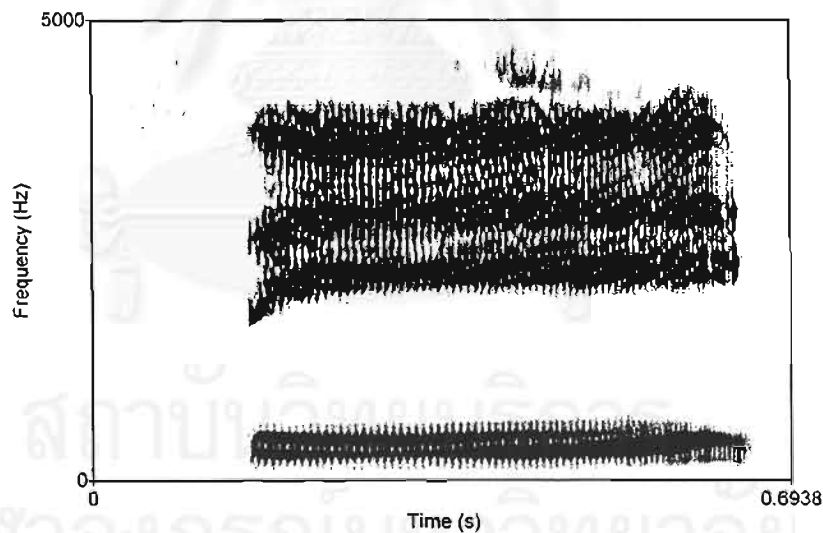
รูปที่ 20 การทำทิดี-โซลา

### บทที่ 3

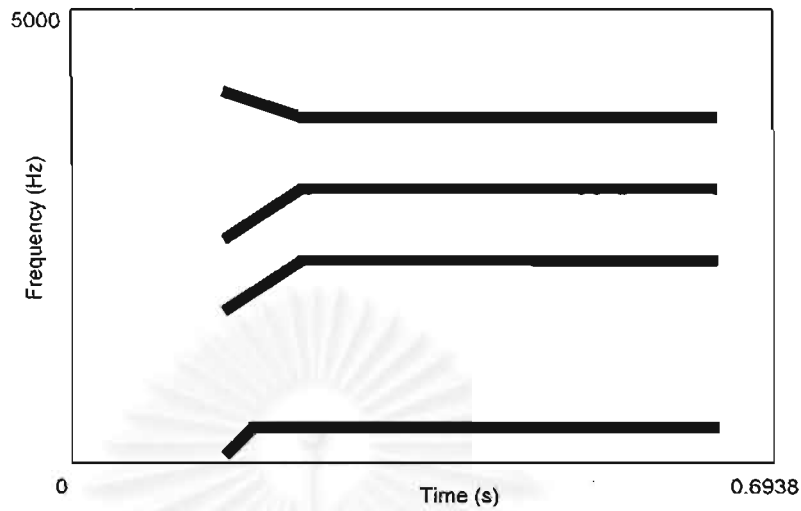
#### กระบวนการวิเคราะห์เสียง

กระบวนการแรกในการสังเคราะห์เสียงคือ การเตรียมฐานข้อมูลสำหรับการสังเคราะห์เสียง ฐานข้อมูลดังกล่าวประกอบไปด้วย หน่วยเสียง รูปแบบเสียงวรรณยุกต์หรือความถี่มูลฐาน รูปแบบกรอบแอมพลิจูด และระยะเวลา โดยลำดับและขั้นตอนต่างๆ ที่ใช้ในการวิเคราะห์และสร้างฐานข้อมูลมีรายละเอียดดังในเนื้อหาของบทนี้

เนื่องจากการแสดงผลด้วยสเปกโตรแกรมในบางภาพอาจไม่ชัดเจนต่อการนำเสนอ เพื่อให้แสดงภาพสัญญาณเสียงได้ชัดเจน การแสดงสัญญาณเสียงด้วยสเปกโตรแกรมในวิทยานิพนธ์นี้ ในบางส่วนจะแสดงด้วยภาพจำลองของสเปกโตรแกรม โดยแทนบริเวณที่เป็นฟอร์แมนต์ด้วยเส้นหนาที่บสีดำ ในส่วนบริเวณเสียงไม่ก้องแสดงด้วยพื้นที่แรเงา ดังแสดงตัวอย่างเปรียบเทียบโดยในรูปที่ 21 แสดงสเปกโตรแกรมและภาพที่ 22 แสดงภาพจำลองของสเปกโตรแกรมของสัญญาณเสียง



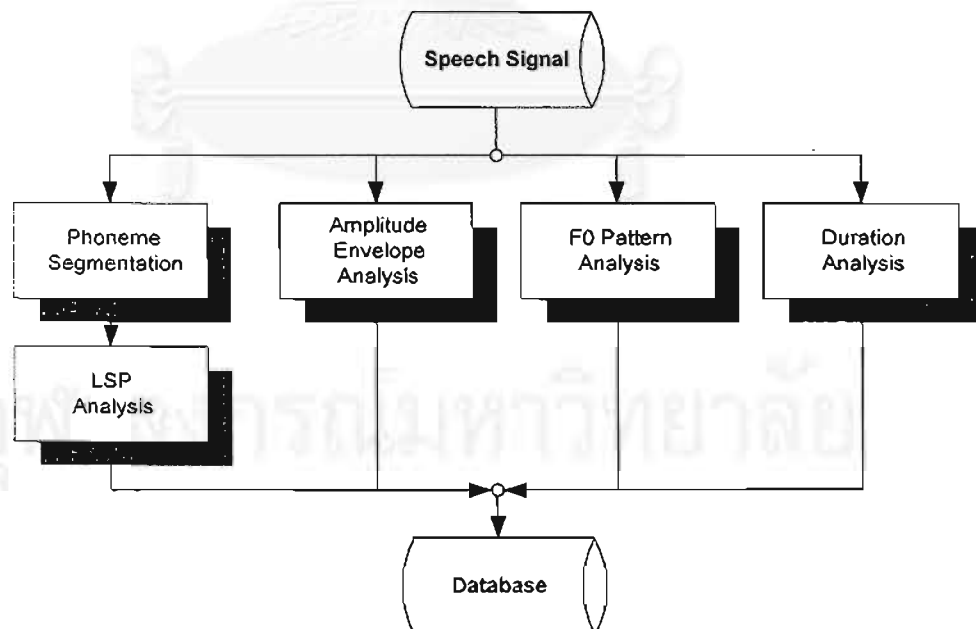
รูปที่ 21 สเปกโตรแกรมต้นแบบ



รูปที่ 22 ภาพจำลองที่ใช้แทนสเปกโตรแกรม

### 3.1 ภาพรวมกระบวนการวิเคราะห์

กระบวนการวิเคราะห์เพื่อสร้างฐานข้อมูลสำหรับการสังเคราะห์เสียงแสดงดังรูปที่ 23



รูปที่ 23 แผนผังกระบวนการวิเคราะห์

กระบวนการวิเคราะห์เริ่มจากการนำสัญญาณเสียงพูดจริงของมนุษย์มาทำการวิเคราะห์หาส่วนประกอบสำคัญซึ่งประกอบด้วยหน่วยเสียง รูปแบบความถี่มูลฐาน รูปแบบกรอบแอมพลิจูด และระยะเวลาการออกเสียง การหาหน่วยเสียงจะใช้การวิเคราะห์ในเชิงความถี่เพื่อแยกหน่วยเสียงออกมา หน่วยเสียงที่ได้จะนำไปวิเคราะห์การทำงานแบบเชิงเส้นโดยเก็บในรูปแบบพารามิเตอร์คู่เส้นสเปกตรัมและสัญญาณตกค้าง ในส่วนของการวิเคราะห์หารูปแบบความถี่มูลฐานจะนำเสียงที่ใช้เป็นต้นแบบมาทำการหารูปแบบวรรณยุกต์ต่างๆ เพื่อเก็บเป็นต้นแบบในการสังเคราะห์เสียงวรรณยุกต์ สำหรับการวิเคราะห์หารูปแบบกรอบแอมพลิจูดมีลักษณะเช่นเดียวกับการหารูปแบบความถี่มูลฐาน การวิเคราะห์อีกอย่างหนึ่งคือการหาระยะเวลาการออกเสียงเพื่อใช้อ้างอิงในกระบวนการสังเคราะห์ว่าการออกเสียงของแต่ละหน่วยเสียงควรมีระยะเวลาเท่าใด

### 3.2 หน่วยเสียง (Phonemes)

ในการสร้างหน่วยเสียงใช้การตัดมาจากเสียงพูดจริงโดยมีรายละเอียดของการบันทึกเสียงดังต่อไปนี้

1. คนออกเสียงต้นแบบเป็นเพศชาย
2. คนออกเสียงใช้ภาษาไทยกลางเป็นภาษาหลัก
3. ใช้คนออกเสียงต้นแบบจำนวน 1 คนในการสร้างหน่วยเสียง
4. ทำการออกเสียงทีละพยางค์โดยรักษาระดับแอมพลิจูดให้ใกล้เคียงกัน
5. บันทึกเสียงลงเครื่องคอมพิวเตอร์โดยตรง
6. เสียงแต่ละพยางค์ถูกปรับระดับแอมพลิจูดให้เท่ากับ 40 % ของค่าเต็มพิกัด

อุปกรณ์และซอฟต์แวร์ที่ใช้ในการบันทึกเสียงมีรายละเอียดดังนี้

1. การ์ดเสียง Sound Blaster AWE 64 Value ของบริษัท Creative Technology
2. ไมโครโฟน NPE รุ่น Pro-18 PROLOGUE
3. โปรแกรมบันทึกและแต่งเสียง Goldwave ของ Chris Craig

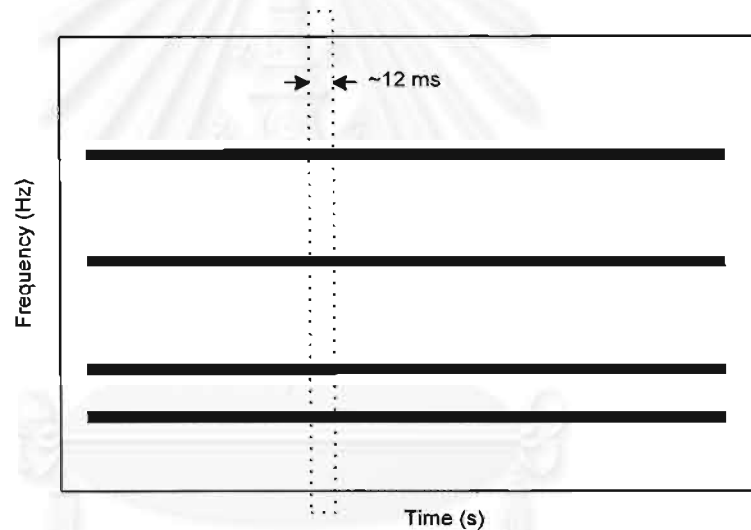
เสียงที่บันทึกมีรายละเอียดดังนี้

1. อัตราสุ่ม 22,050 เฮิรตซ์
2. ความละเอียดสัญญาณ 16 บิต
3. คำที่บันทึกเป็นต้นแบบได้จากการออกเสียงพยางค์คู่กับเสียงสระที่เกิดขึ้นได้ในภาษาไทยตามบทที่ 2 เสียงสระและเสียงวรรณยุกต์

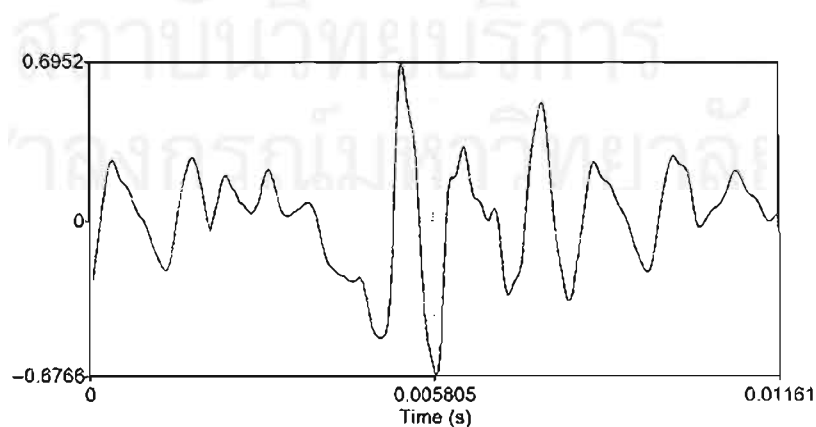


หน่วยเสียงที่ใช้เป็นหน่วยเสียงที่ออกแบบเฉพาะสำหรับการสังเคราะห์ด้วยวิธีการที่จะนำเสนอดังต่อไปนี้ในวิทยานิพนธ์นี้ หน่วยเสียงนี้ได้จากการวิเคราะห์หาส่วนของเสียงที่เก็บลักษณะเฉพาะของเสียงแต่ละเสียงโดยอาศัยทฤษฎีการรับรู้เสียงพูดในหัวข้อ 2.1 มาช่วยในการวิเคราะห์เพื่อให้ได้หน่วยเสียงที่จำเป็นสำหรับการรับรู้โดยที่จำนวนหน่วยเสียงน้อยและมีขนาดเล็ก ในการวิเคราะห์หาหน่วยเสียงนี้แยกออกเป็น 2 ส่วนหลักตามประเภทของเสียงคือ

1. หน่วยเสียงสระ เสียงสระเดี่ยวมีลักษณะฟอร์แมนต์ที่ชัดเจนและแตกต่างกันสำหรับเสียงสระแต่ละเสียง การเก็บหน่วยเสียงสระ (V) จึงเก็บเพียงหนึ่งกรอบเสียงสำหรับสระเดี่ยวแต่ละเสียง ดังในรูปที่ 24 แสดงตัวอย่างสเปกโตรแกรมของและตำแหน่งการเก็บหน่วยเสียงของเสียงสระ "อา" และรูปที่ 25 แสดงรูปคลื่นของหน่วยเสียงสระเดี่ยวกัน สำหรับสระผสมซึ่งเกิดจากการเชื่อมต่อของเสียงสระเดี่ยวสองเสียงอย่างต่อเนื่อง ดังนั้นจึงไม่เก็บหน่วยเสียงสำหรับสระผสม

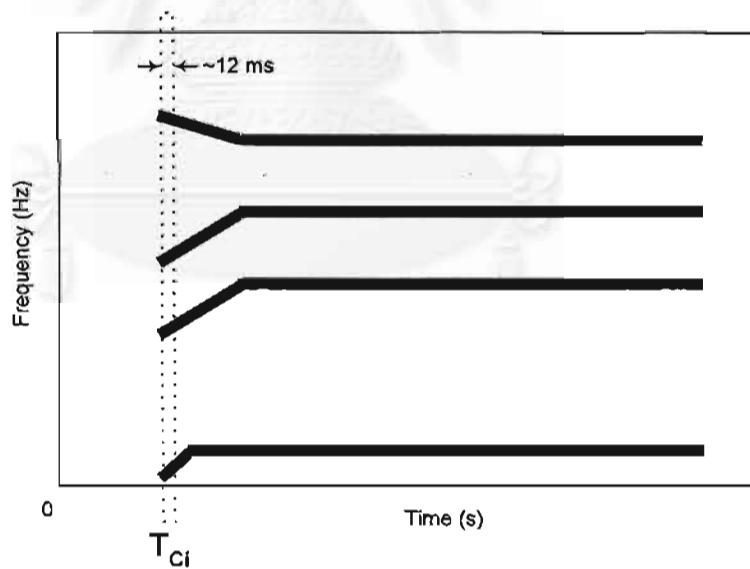


รูปที่ 24 ตัวอย่างสเปกโตรแกรมแสดงตำแหน่งการเก็บเสียงสระ "อา"

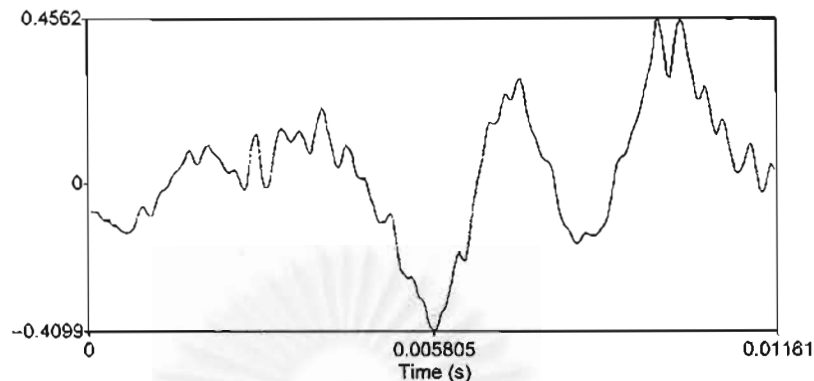


รูป 25 ตัวอย่างรูปคลื่นสัญญาณเสียงของหน่วยเสียงสระ "อา"

2. ส่วนหน่วยเสียงพยัญชนะ ได้แก่เสียงพยัญชนะต้น เสียงพยัญชนะควบกล้ำ และเสียงพยัญชนะสะกด จากทฤษฎีการรับรู้เสียงพยัญชนะ (Lieberman, 1988) คือ การบิดเบนของฟอร์แมนต์หรือการบิดเบนของความถี่กำหนดของเสียงสระแต่ละสระในลักษณะต่างๆ ทำให้เกิดการรับรู้เป็นเสียงพยัญชนะต่างๆกัน โดยบริเวณที่เกิดการบิดเบนของความถี่กำหนดดังกล่าวอาจเป็นเสียงก้องหรือเสียงผสมระหว่างก้องและไม่ก้อง ดังนั้นจึงเลือกเก็บกรอบเสียงเฉพาะบริเวณปลายของความถี่กำหนดที่เกิดการบิดเบนเพื่อใช้เป็นตำแหน่งอ้างอิงของการบิดเบนดังในรูปที่ 26 โดยเสียงสระแต่ละเสียงจะมีลักษณะการบิดเบนแตกต่างกันสำหรับพยัญชนะตัวเดียวกัน ดังนั้นจึงต้องเก็บเสียงในลักษณะคู่ของพยัญชนะและเสียงสระแต่ละตัว สำหรับความกว้างของกรอบเสียงที่ใช้คือประมาณ 12 มิลลิวินาที เนื่องจากความถี่มูลฐานต่ำสุดของเสียงพูดของผู้ชายโดยทั่วไปคือประมาณ 80 เฮิรท์ ดังนั้นหน่วยเสียงภายในกรอบเสียงที่เก็บที่ได้ยืนยันได้ว่าครอบคลุมสัญญาณอย่างน้อยหนึ่งพิทช์ และตำแหน่งการเก็บหน่วยเสียงต้องกำหนดตำแหน่งพิทช์ให้อยู่ตรงบริเวณกึ่งกลางของกรอบสัญญาณเพื่อใช้อ้างอิงตำแหน่งพิทช์ในการประมวลผลในภายหลังได้ดังรูปที่ 27 ตัวอย่างรูปคลื่นสัญญาณเสียงเชิงเวลาที่อยู่ภายในกรอบเสียง



รูปที่ 26 ตำแหน่งการเก็บหน่วยเสียงพยัญชนะ



รูปที่ 27 ลักษณะรูปคลื่นสัญญาณเสียงของหน่วยเสียงภายในกรอบเสียง

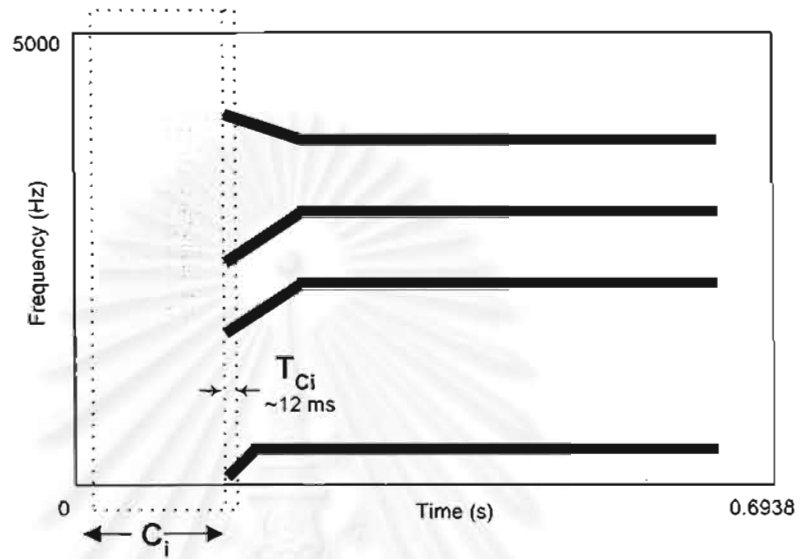
สำหรับเสียงพยัญชนะต้นและพยัญชนะควบกล้ำซึ่งมีเสียงที่มีลักษณะเชิงความถี่ซับซ้อนหรือเป็นส่วนเสียงไม่ก้องให้เกิดเสียงในส่วนนี้ (บริเวณ  $C_1$ ) พร้อมกับสัญญาณกรอบแรกที่ตำแหน่งปลายของฟอร์แมนต์ที่เกิดการบิดเบน (บริเวณ  $T_{C1}$ ) เพื่อใช้เป็นจุดอ้างอิง ดังตัวอย่างในรูปที่ 28 แสดงภาพจำลองสเปกโตรแกรมของตำแหน่งที่เก็บหน่วยเสียง  $C_1, T_{C1}$  ของเสียงพยัญชนะ "ส" จากเสียง "สี" และรูปที่ 29 แสดงรูปคลื่นสัญญาณเสียงของหน่วยเสียงพยัญชนะเดียวกัน

สำหรับการเก็บหน่วยเสียงพยัญชนะสะกดยกเว้นพยัญชนะสะกดอักษสระและพยัญชนะสะกดนาสิก โดยเก็บกรอบเสียงบริเวณปลายของการบิดเบนจำนวนหนึ่งกรอบ (บริเวณ  $T_{Cn}$ ) เป็นจุดอ้างอิง ดังตัวอย่างในรูปที่ 30 สเปกโตรแกรมของตำแหน่งที่เก็บหน่วยเสียงของเสียงพยัญชนะสะกด "ก" จากเสียง "อาก" และในรูปที่ 31 แสดงรูปคลื่นสัญญาณเสียงของหน่วยเสียงพยัญชนะสะกดเดียวกัน

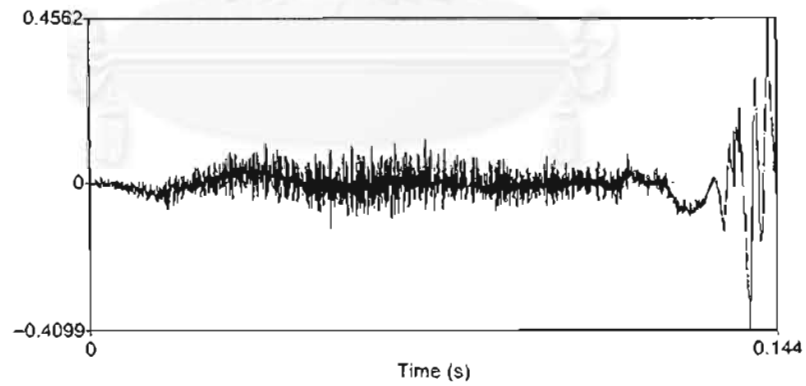
สำหรับกรณีหน่วยเสียงพยัญชนะสะกดอักษสระซึ่งมีลักษณะสเปกโตรแกรมเช่นเดียวกับสระผสมดังแสดงสเปกโตรแกรมในรูปที่ 32 ดังนั้นในการเก็บหน่วยเสียงจึงเก็บเพียงตำแหน่งฟอร์แมนต์อ้างอิง (บริเวณ  $C_n$ ) ในบริเวณที่ฟอร์แมนต์คงที่เพียงหนึ่งกรอบเช่นเดียวกับการเก็บหน่วยเสียงสระดังแสดงตัวอย่างตำแหน่งการเก็บหน่วยเสียงในรูปที่ 33 และแสดงรูปคลื่นดังในรูปที่ 34

สำหรับกรณีหน่วยเสียงพยัญชนะสะกดนาสิกดังแสดงในรูปที่ 35 แสดงสเปกโตรแกรมของเสียงพยัญชนะสะกด "น" หน่วยเสียงของพยัญชนะสะกดกลุ่มนี้มีลักษณะเฉพาะแตกต่างไปจากพยัญชนะสะกดกลุ่มอื่นๆ คือ การบิดเบนของฟอร์แมนต์แบ่งเป็น 2 ส่วน ดังนั้นการเก็บตำแหน่งฟอร์แมนต์อ้างอิงจึงเก็บเป็น 2 หน่วยเสียงย่อย (บริเวณ  $T_{Cm}$  และ  $C_n$ ) ดังแสดงในรูปที่ 36 สำหรับหน่วยเสียง  $C_n$  แสดงลักษณะของเสียงนาสิกคือมีลักษณะการต่อต้านการกำทอนโดยเกิดต่อเนื่องจากตำแหน่งปลายการบิดเบนของฟอร์แมนต์ ( $T_{Cm}$ ) จากสเปกโตรแกรมแสดงลักษณะของ

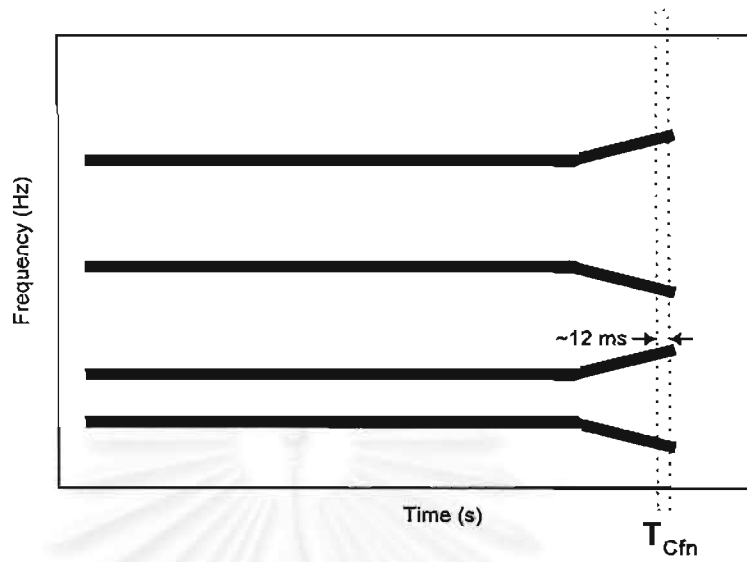
เสียงนาสิก  $C_{in}$  ที่มีลักษณะฟอร์แมนต์คงที่เช่นเดียวกับเสียงสระ การเก็บหน่วยเสียง  $C_{in}$  จึงเก็บเพียงหนึ่งกรอบเสียงและในรูปที่ 37 แสดงตัวอย่างรูปคลื่นของหน่วยเสียง  $C_{in}$



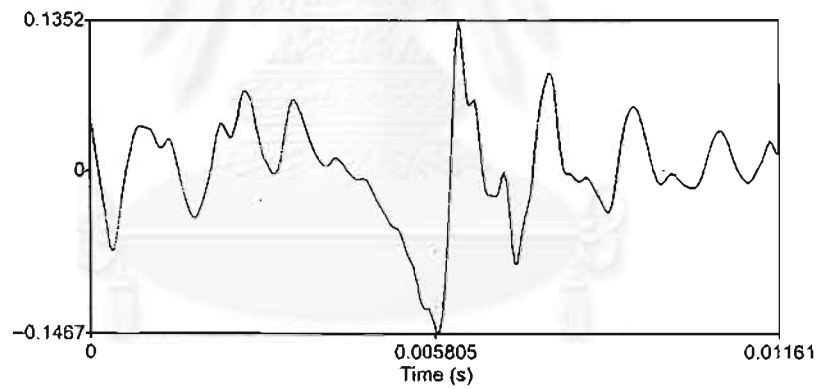
รูปที่ 28 ตัวอย่างสเปกโตรแกรมแสดงตำแหน่งหน่วยเสียงพยัญชนะต้นของเสียง "สิ"



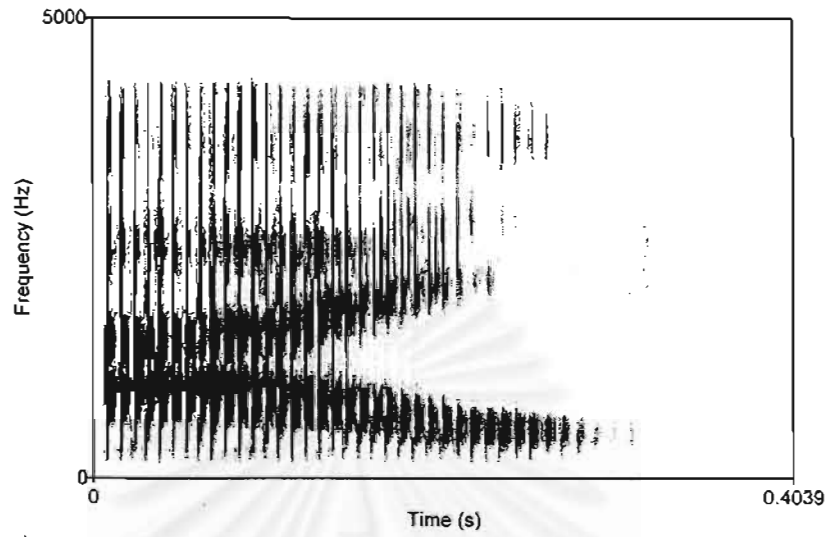
รูปที่ 29 ตัวอย่างรูปคลื่นสัญญาณเสียงของหน่วยเสียงพยัญชนะต้นของเสียง "สิ"



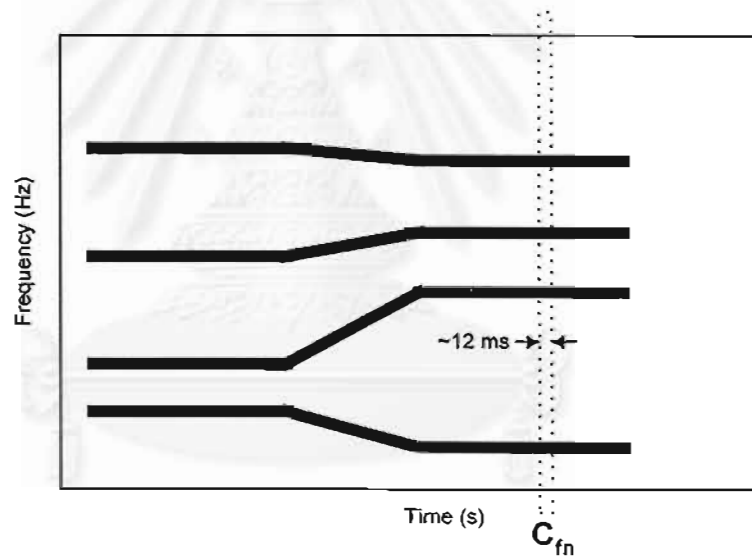
รูปที่ 30 ตัวอย่างสเปกโตรแกรมแสดงตำแหน่งหน่วยเสียงพยัญชนะสะกดของเสียง "าก"



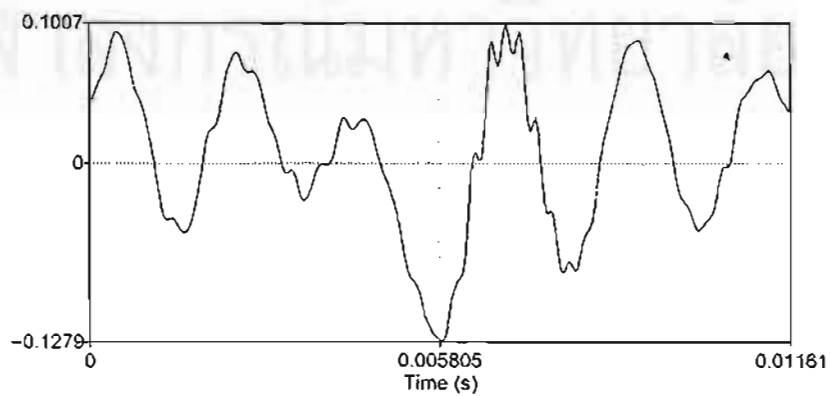
รูปที่ 31 ตัวอย่างรูปคลื่นสัญญาณเสียงของหน่วยเสียงพยัญชนะสะกดของเสียง "าก"



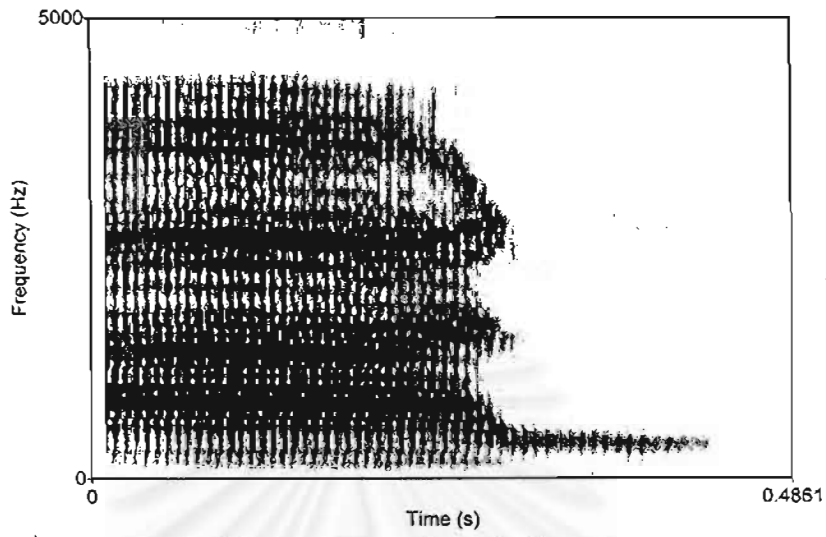
รูปที่ 32 ตัวอย่างสเปกโตรแกรมแสดงเสียงพยัญชนะสะกดอัฒสระของเสียง "อาย"



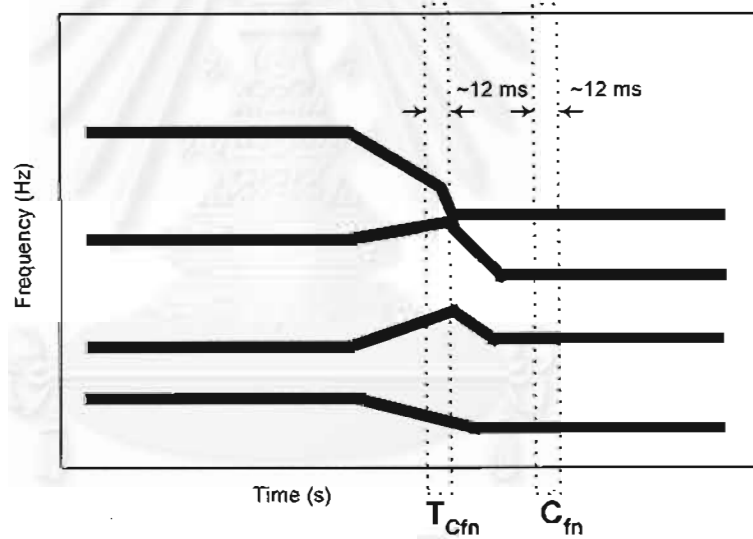
รูปที่ 33 ตัวอย่างสเปกโตรแกรมแสดงตำแหน่งหน่วยเสียงพยัญชนะสะกดอัฒสระของเสียง "อาย"



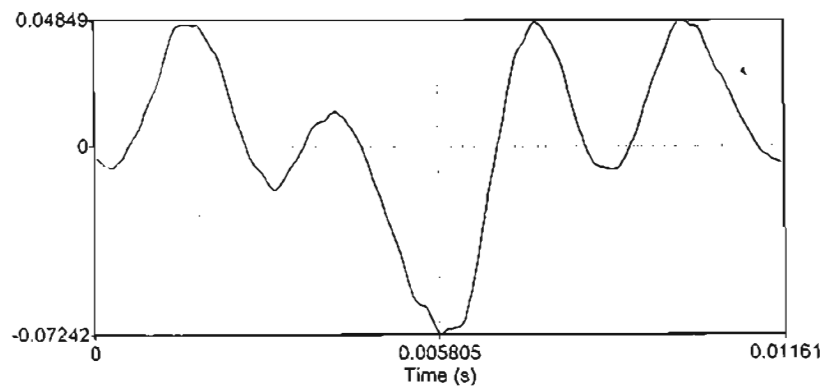
รูปที่ 34 ตัวอย่างรูปคลื่นสัญญาณเสียงของหน่วยเสียงพยัญชนะสะกดอัฒสระของเสียง "อาย"



รูปที่ 35 ตัวอย่างสเปกโตรแกรมแสดงตำแหน่งเสียงพยัญชนะสะกดนาสิก "อาน"



รูปที่ 36 ตัวอย่างสเปกโตรแกรมแสดงตำแหน่งหน่วยเสียงพยัญชนะสะกดนาสิกจากเสียง "อาน"

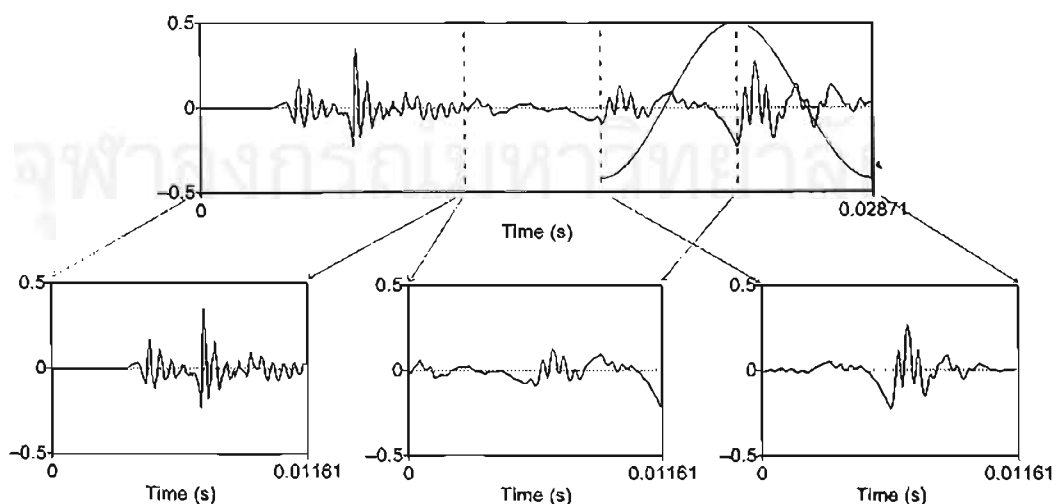


รูปที่ 37 ตัวอย่างรูปคลื่นสัญญาณเสียงของหน่วยเสียงพยัญชนะสะกดนาสิกจากเสียง "อาน"

### 3.3 การจัดเก็บหน่วยเสียง

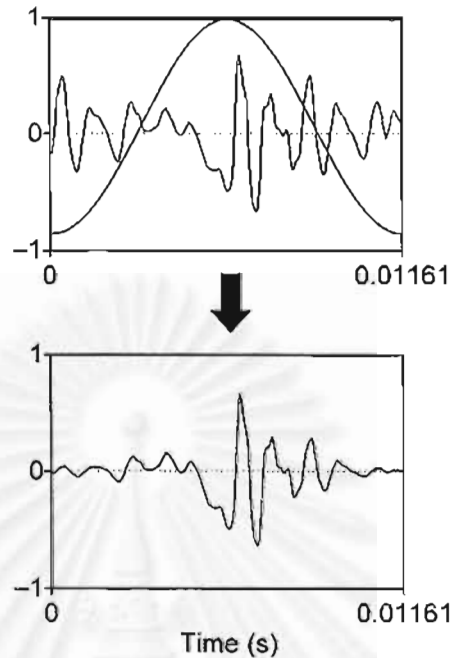
เมื่อทำการตัดหน่วยเสียงจากเสียงพูด หน่วยเสียงดังกล่าวจะถูกนำมาวิเคราะห์แบบหน้าต่างต่างเพื่อหาข้อมูลเชิงความถี่ของหน่วยเสียงโดยการวิเคราะห์การเข้ารหัสการทำนายแบบเชิงเส้น กรอบเชิงความถี่ดังกล่าวจะนำไปใช้ในการทำรอยต่อระหว่างหน่วยเสียงให้ราบรื่นเพื่อสร้างเสียงพัญชนะต่างๆกลับคืน ในวิทยานิพนธ์นี้เลือกใช้การเข้ารหัสการทำนายแบบเชิงเส้นที่เก็บสัมประสิทธิ์ในรูปคู่เส้นสเปกตรัมเนื่องจากคุณสมบัติในการเข้ารหัสความถี่ฟอร์แมนต์และความสามารถในการประมาณค่าพารามิเตอร์ดังที่กล่าวในบทที่ 2 ซึ่งจะช่วยให้การเชื่อมต่อหน่วยเสียงมีความราบรื่น

ในขั้นตอนนี้รูปคลื่นของหน่วยเสียงจะถูกทำเป็นกรอบเสียงย่อยและทำหน้าที่ต่างแต่หน้าต่างต่างกว้างประมาณ 12 มิลลิวินาทีซึ่งมีความกว้างเท่ากับคาบพิทช์ที่ความถี่มูลฐานต่ำสุดสำหรับเสียงผู้ชายคือ 80 เฮิรท์ โดยสัญญาณเสียงบริเวณปลายการบิดเบน  $T_{c1}$  และ  $T_{c2}$  จะใช้การวิเคราะห์แบบหน้าต่างชนิดแฮมมิง (Deller et al., 1993) เพื่อนำไปวิเคราะห์หาพารามิเตอร์สำหรับการเชื่อมต่อ ในขณะที่สัญญาณส่วนต้นของเสียงพัญชนะต้น ( $C_1$ ) จะถูกทำหน้าที่ต่างสี่เหลี่ยม ดังแสดงการแบ่งกรอบเสียงและทำหน้าที่ต่างของเสียงพัญชนะต้นและพัญชนะควบกล้ำในรูปที่ 38 และในรูปที่ 39 แสดงการวิเคราะห์แบบหน้าต่างหน่วยเสียงพัญชนะสะกดและสระ นำกรอบเสียงที่ทำหน้าที่ต่างแล้วไปทำการเข้ารหัสการทำนายแบบเชิงเส้นและหาสัญญาณตกค้าง นำสัมประสิทธิ์การเข้ารหัสการทำนายแบบเชิงเส้นดังกล่าวไปแปลงเป็นพารามิเตอร์คู่เส้นสเปกตรัม แล้วเก็บชุดพารามิเตอร์การทำนายแบบเชิงเส้นและสัญญาณตกค้างไว้เป็นฐานข้อมูลสำหรับการสังเคราะห์ต่อไป



รูปที่ 38 การวิเคราะห์แบบหน้าต่างหน่วยเสียงพัญชนะต้นและพัญชนะควบกล้ำ ( $C_1, T_{c1}$ )



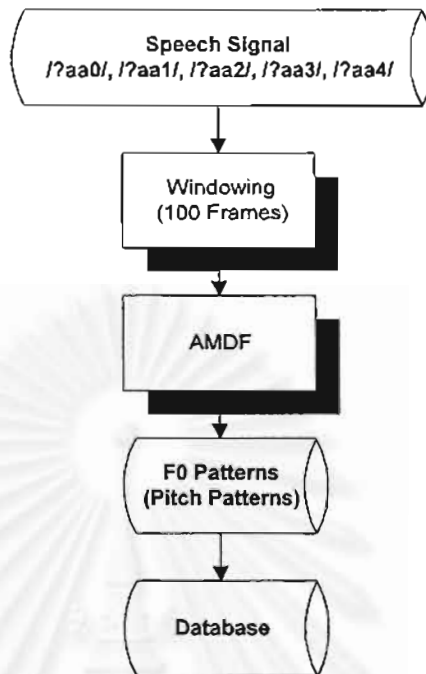


รูปที่ 39 การวิเคราะห์แบบหน้าต่างหน่วยเสียงพยัญชนะสะกดและสระ ( $V$ ,  $T_{Cn}$  และ  $C_n$ )

### 3.4 รูปแบบความถี่มูลฐาน

การเก็บรูปแบบความถี่มูลฐานใช้วิธีเอเอ็มดีเอฟเก็บรูปแบบความถี่มูลฐานของเสียงวรรณยุกต์แต่ละเสียง โดยการนำเสียงสระมาผ่านเสียงวรรณยุกต์ทั้งห้าเสียงประกอบด้วย เสียงสามัญ เสียงเอก เสียงโท เสียงตรี และเสียงจัตวา นำเสียงการผันวรรณยุกต์ดังกล่าวมาแบ่งกรอบเสียงเป็นจำนวน 100 กรอบแล้วผ่านวิธีเอเอ็มดีเอฟซึ่งจะได้ชุดของค่าพิทช์ของแต่ละเสียงวรรณยุกต์ เก็บชุดของค่าพิทช์ของวรรณยุกต์ทั้ง 5 วรรณยุกต์ดังกล่าวไว้เป็นต้นแบบในการสังเคราะห์ ดังแสดงขั้นตอนในรูปที่ 40

ในขั้นตอนนี้เลือกใช้โปรแกรม WinSAL ในการหารูปแบบความถี่มูลฐานโดยกำหนดวิธีการหาความถี่มูลฐานเป็นวิธีเอเอ็มพีเอฟ

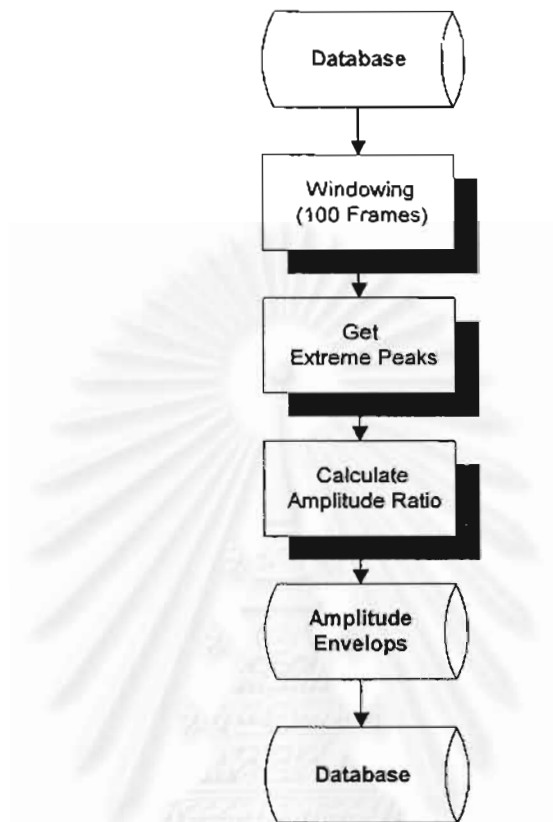


รูปที่ 40 ขั้นตอนการหารูปแบบความถี่มูลฐาน

### 3.5 รูปแบบกรอบแอมพลิจูด

การเก็บรูปแบบกรอบแอมพลิจูดใช้วิธีแบ่งสัญญาณเสียงเป็นกรอบเสียงย่อยจำนวน 100 กรอบแล้วทำการเก็บค่าแอมพลิจูด ณ ตำแหน่งที่มีแอมพลิจูดค่าสุด (Extreme Amplitude) ของแต่ละพิทช์ซึ่งก็คือตำแหน่งยอดพิทช์ของแต่ละกรอบ แล้วคำนวณอัตราส่วนของค่าแอมพลิจูดค่าสุดของแต่ละกรอบเทียบกับค่าแอมพลิจูดค่าสุดของกรอบแรกดังสมการที่ 3.1 เพื่อให้การปรับแอมพลิจูดมีความต่อเนื่อง ณ ตำแหน่งพิทช์แรกที่เริ่มปรับขนาดกรอบแอมพลิจูด แล้วทำการเก็บค่าอัตราส่วนแอมพลิจูดค่าสุดดังกล่าวไว้ ดังแสดงขั้นตอนการเก็บกรอบแอมพลิจูดในรูปที่ 41 โดยแยกเก็บกรอบแอมพลิจูดสำหรับเสียงวรรณยุกต์แต่ละเสียงเนื่องจากเสียงวรรณยุกต์แต่ละเสียงมีลักษณะกรอบแอมพลิจูดแตกต่างกัน ดังนั้นต้องเก็บกรอบแอมพลิจูดทั้งหมด 5 กรอบ โดยเสียงที่ใช้ในการวิเคราะห์หากรอบแอมพลิจูดเป็นชุดเดียวกับที่ใช้หารูปแบบความถี่มูลฐานในหัวข้อ 3.4 เพื่อรักษาความสัมพันธ์ทางด้านตำแหน่งของรูปแบบทั้งสอง

$$\text{อัตราส่วนของแอมพลิจูดค่าสุด} = \frac{A_{\text{Extreme},n}}{A_{\text{Extreme},1}} \quad (3.1)$$



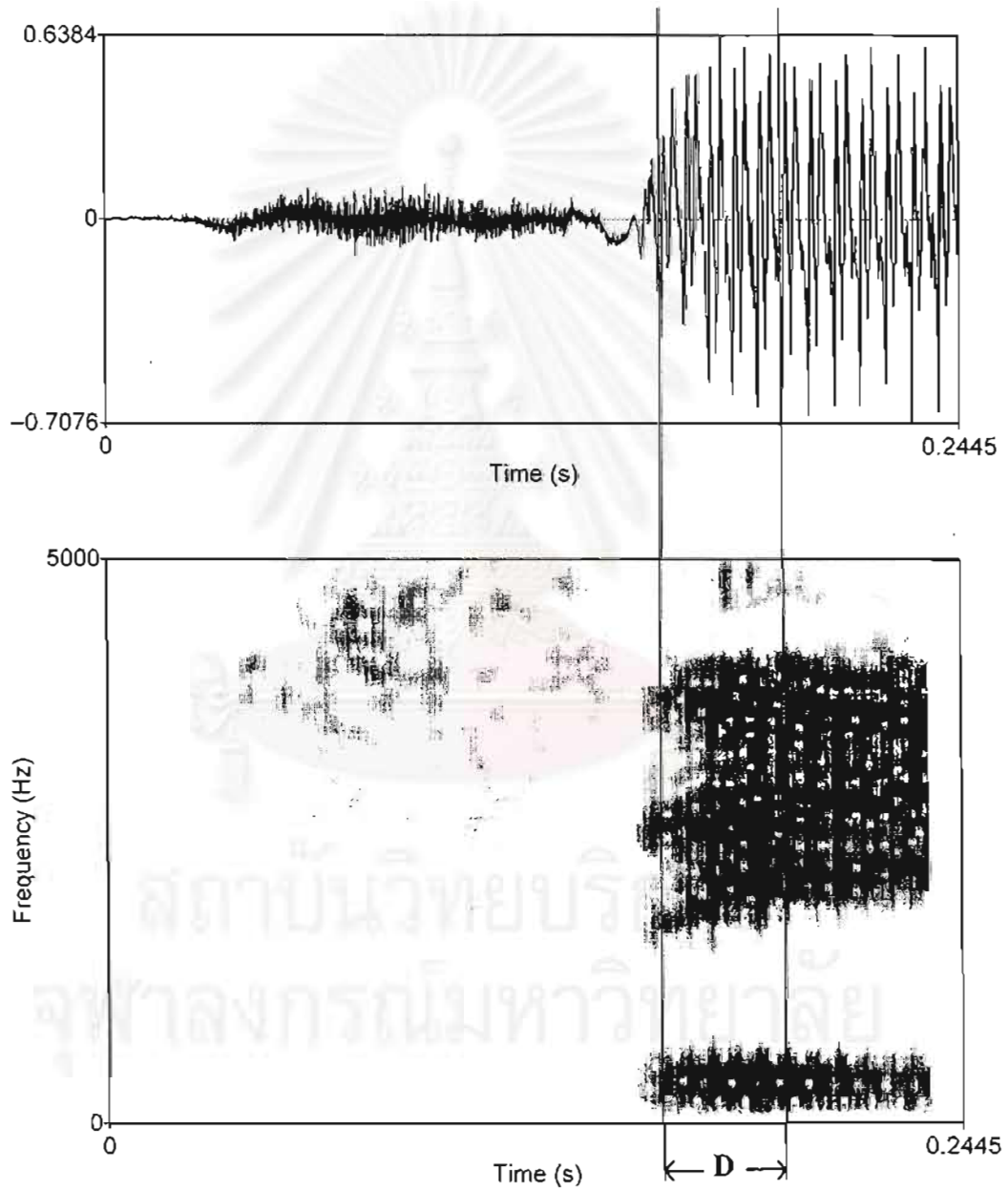
รูปที่ 41 ขั้นตอนการหากรอบแอมพลิจูด

ในขั้นตอนนี้เลือกใช้โปรแกรม WinSAL ในการหากรอบแอมพลิจูดของแต่ละวรรณยุกต์ โดยเลือกวิธีการคำนวณเป็นแอมพลิจูด

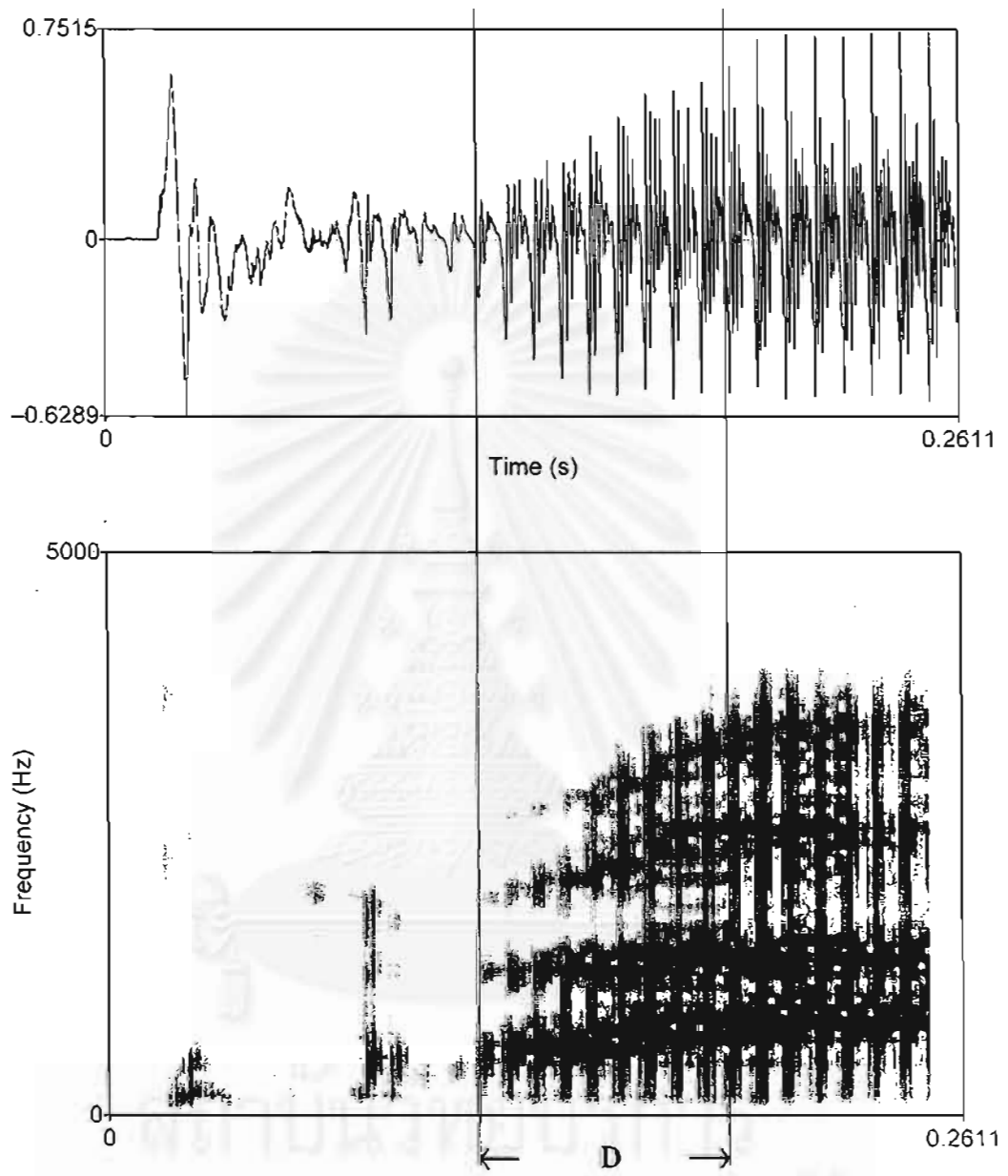
### 3.6 ระยะเวลา

การเก็บค่าระยะเวลาเพื่อใช้กำหนดระยะเวลาในการสังเคราะห์ของระหว่างแต่ละหน่วยเสียงได้จากการวัดระยะเวลาจากรูปคลื่นของเสียงพูดที่ใช้เป็นต้นแบบในการทำหน่วยเสียงเพื่อให้ได้ระยะเวลาที่เหมาะสมสำหรับเสียงต้นแบบ การวัดระยะเวลาเริ่มวัดระยะเวลาจากยอดพิทช์เริ่มต้นไปสิ้นสุดที่ยอดพิทช์ปลายทาง ในกรณีที่เป็นเสียงพยัญชนะต้น เสียงพยัญชนะควบกล้ำและเสียงพยัญชนะสะกด ตำแหน่งที่ใช้วัดระยะเวลาดังแสดงในรูปที่ 42 แสดงเสียง "สี" รูปที่ 43 แสดงเสียง "พรา" และรูปที่ 44 แสดงเสียง "อาท" โดยเริ่มวัดจากยอดพิทช์ที่เริ่มการบิดเบนของฟอร์แมนต์และไปสิ้นสุดที่ยอดพิทช์ที่ฟอร์แมนต์คงที่ ในกรณีของเสียงนาสิกของเสียงพยัญชนะสะกดเริ่มวัดที่ยอด

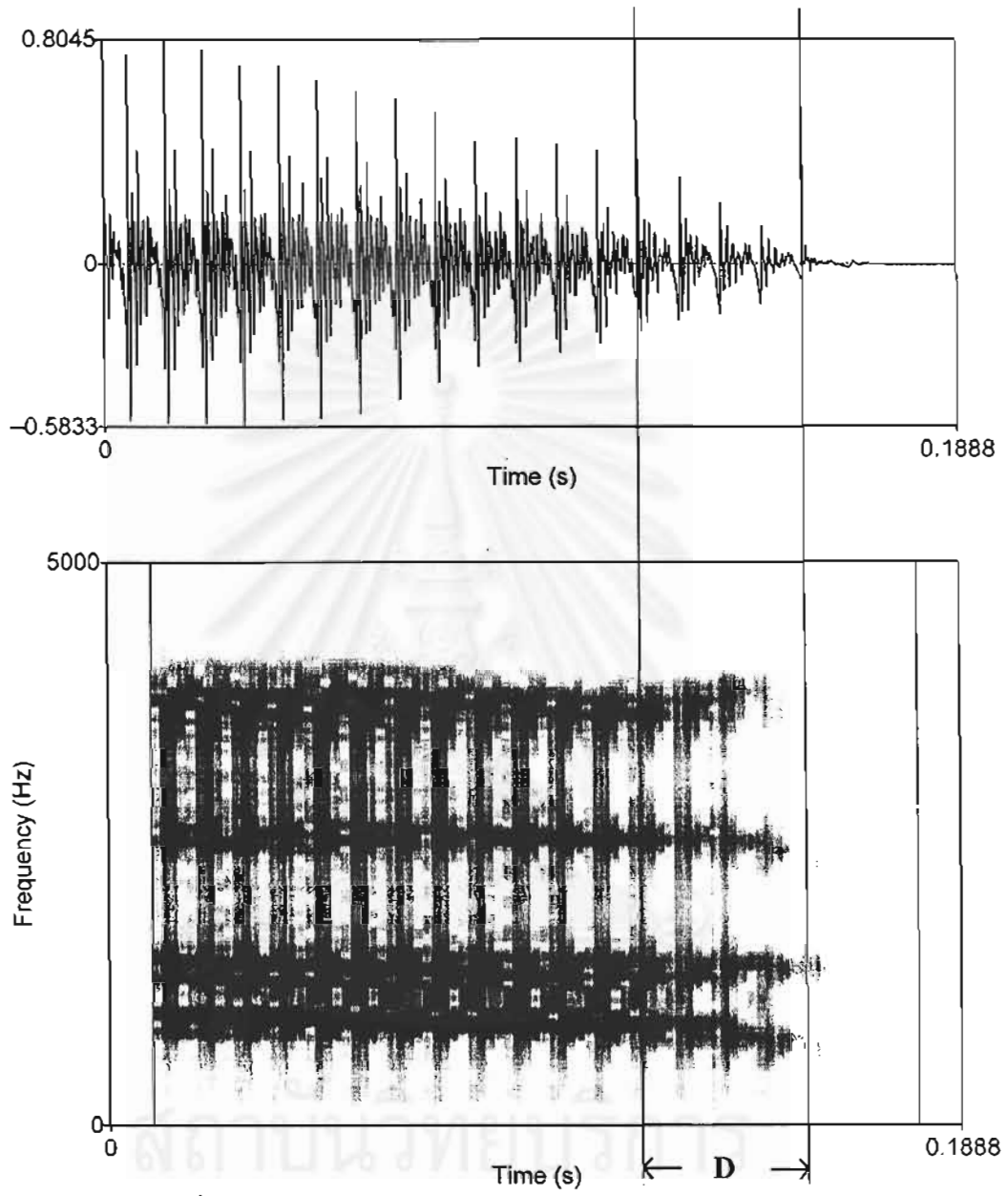
พิทช์สุดท้ายของการบิดเบือนของฟอร์แมนต์ของเสียงพยัญชนะสระกอดและไปสิ้นสุดที่ยอดพิทช์สุดท้ายของพยางค์ดังแสดงในรูปที่ 45 สำหรับระยะเวลาของส่วนสระมีระยะเวลาที่แปรเปลี่ยนได้ ดังนั้นระยะเวลาในส่วนนี้ใช้การทดลองปรับระยะเวลาในการหาระยะเวลาดังกล่าว



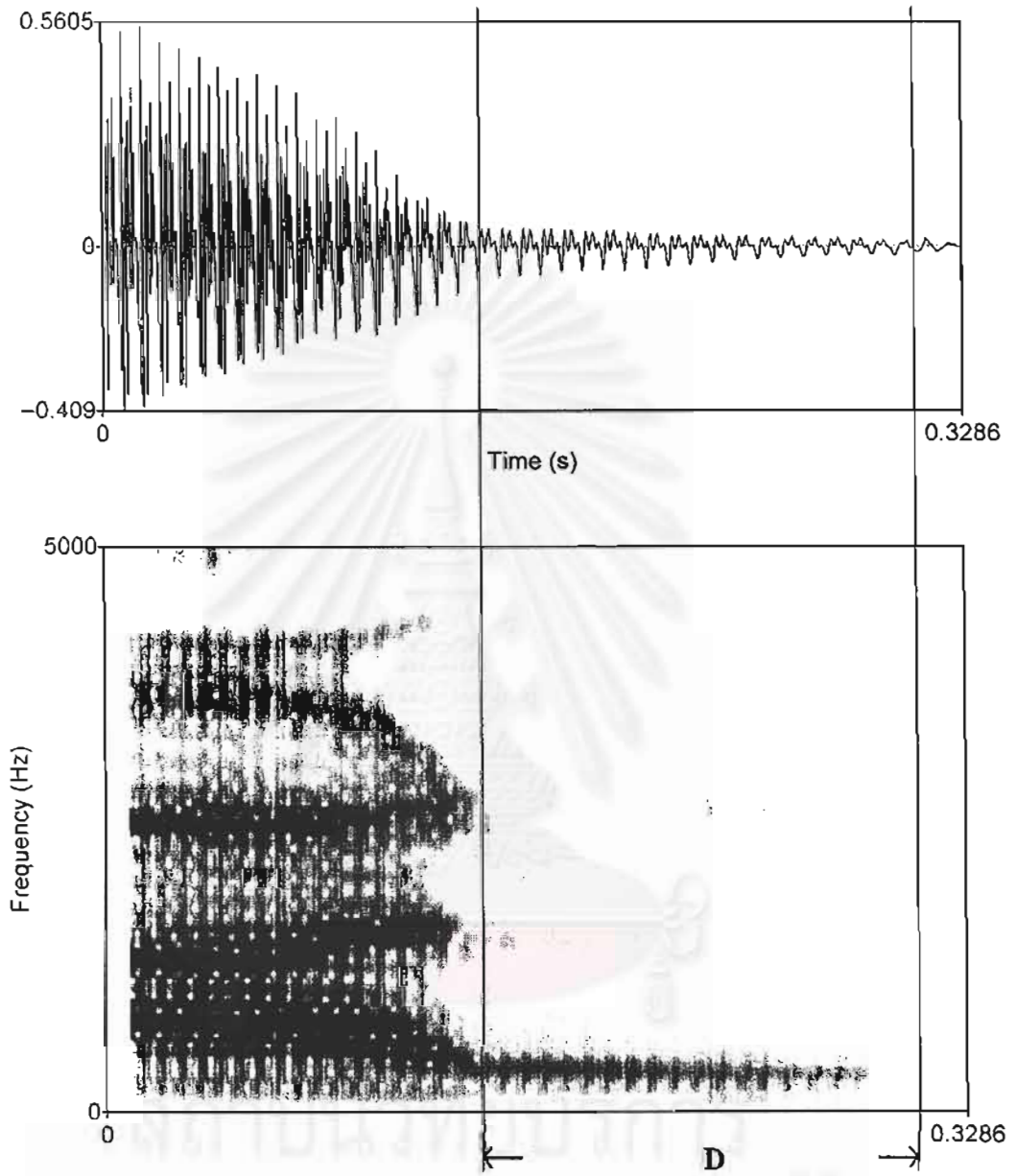
รูปที่ 42 ตัวอย่างการวัดระยะเวลาเสียงพยัญชนะต้นของเสียง "สี"



รูปที่ 43 ตัวอย่างการวัดระยะเวลาเสียงพยัญชนะควบกล้ำของเสียง "พรา"



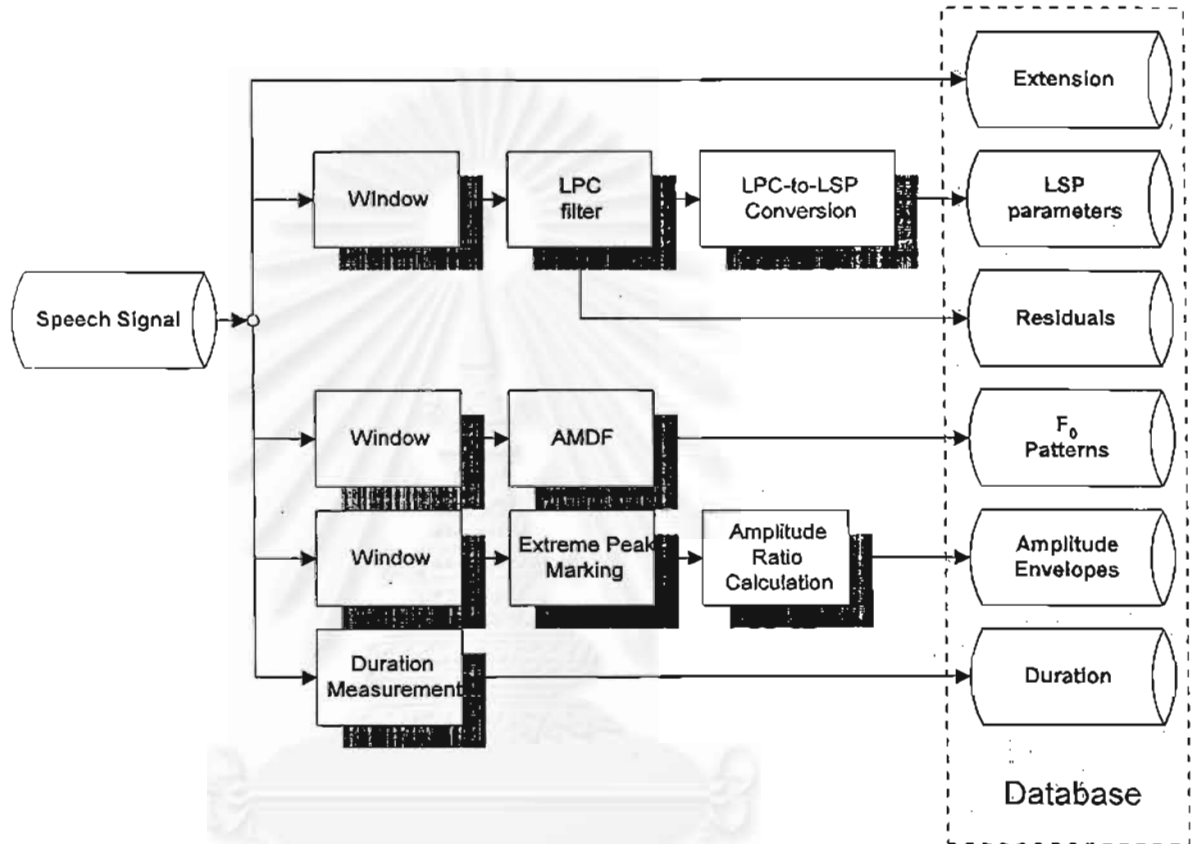
รูปที่ 44 ตัวอย่างการวัดระยะเวลาเสียงพยัญชนะสระคดของเสียง "อาก"



รูปที่ 45 ตัวอย่างการวัดระยะเวลาเสียงนาสิกของเสียงพยัญชนะสระกต "น"

### 3.7 สรุปขั้นตอนการวิเคราะห์เสียง

ในกระบวนการวิเคราะห์เสียงสรุปเป็นขั้นตอนการวิเคราะห์ได้ดังรูปที่ 46



รูปที่ 46 ขั้นตอนการวิเคราะห์เสียงโดยสรุป



## บทที่ 4

### กระบวนการสังเคราะห์เสียง

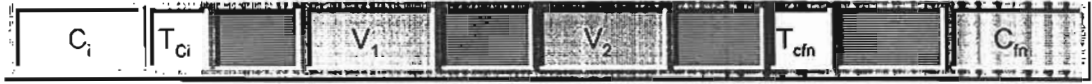
เมื่อสร้างฐานข้อมูลสำหรับการสังเคราะห์เสียงแล้ว ในบทนี้จะนำเสนอลำดับและขั้นตอนการสังเคราะห์เสียงโดยใช้คู่เส้นสเปกตรัม โดยอาศัยหน่วยเสียงและข้อมูลอื่น ๆ จากฐานข้อมูลที่ได้สร้างขึ้น นอกจากนี้การสังเคราะห์เสียงภาษาไทยที่จะนำเสนอคืออาศัยโครงสร้างพยางค์สังเคราะห์สำหรับหน่วยเสียงที่นำเสนอในวิทยานิพนธ์ฉบับนี้ และในตอนท้ายได้นำเสนอเกี่ยวโปรแกรมสังเคราะห์เสียง

#### 4.1 โครงสร้างพยางค์สังเคราะห์สำหรับภาษาไทย

หลักการที่ใช้ในการสังเคราะห์เสียงคือ บริเวณที่ฟอร์แมนต์หรือความถี่กำหนดเกิดการบิดเบนจะถูกสังเคราะห์ขึ้นจากการการประมาณค่าระหว่างหน่วยเสียง ณ บริเวณขอบหน่วยเสียง ซึ่งหน่วยเสียงดังกล่าวเก็บค่าตำแหน่งฟอร์แมนต์หรือความถี่กำหนดอ้างอิงไว้ ดังที่นำเสนอในบทที่ 3 และในบริเวณที่ฟอร์แมนต์หรือความถี่กำหนดมีลักษณะคงที่จะถูกสังเคราะห์ขึ้นด้วยวิธีการทำซ้ำ (Duplication) ของหน่วยเสียง

การสังเคราะห์เสียงที่จะนำเสนอในวิทยานิพนธ์นี้อาศัยโครงสร้างพยางค์สำหรับการสังเคราะห์ดังรูปที่ 47 โดยใช้หน่วยเสียงที่ได้นำเสนอในบทที่ 3 ภายในโครงสร้างนี้ประกอบด้วยหน่วยเสียง 3 ประเภทคือ

1. หน่วยเสียงพยัญชนะ (C) ประกอบด้วย  $C_i$  และ  $C_{in}$
2. หน่วยเสียงสระ (V) ประกอบด้วย  $V_1$  และ  $V_2$
3. หน่วยเสียงจุดอ้างอิง (T) ประกอบด้วย  $T_{ci}$  และ  $T_{cin}$



รูปที่ 47 โครงสร้างพยางค์ที่ใช้ในการสังเคราะห์เสียง

โดยที่

$C_i$	คือ หน่วยเสียงของเสียงพยัญชนะต้นพยางค์หรือพยัญชนะควบกล้ำ
$T_{Ci}$	คือ หน่วยเสียงจุดอ้างอิงของเสียงพยัญชนะต้นหรือพยัญชนะควบกล้ำ
$V_1$	คือ หน่วยเสียงของสระนำ
$V_2$	คือ หน่วยเสียงของสระตาม
$T_{cfn}$	คือ หน่วยเสียงจุดอ้างอิงของเสียงพยัญชนะสะกด
$C_{fn}$	คือ หน่วยเสียงท้ายพยางค์ที่ไม่ใช่เสียงกัก (Non-stop Final Consonant)

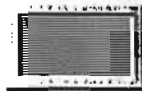
ประกอบด้วยเสียงนาสิกท้ายพยางค์และเสียงพยัญชนะสะกดกลุ่มอัมสระ



คือ บริเวณที่ไม่มีการประมวลผลใดๆ กับหน่วยเสียง



คือ บริเวณที่เกิดจากการทำซ้ำ (Duplication) ของหน่วยเสียง



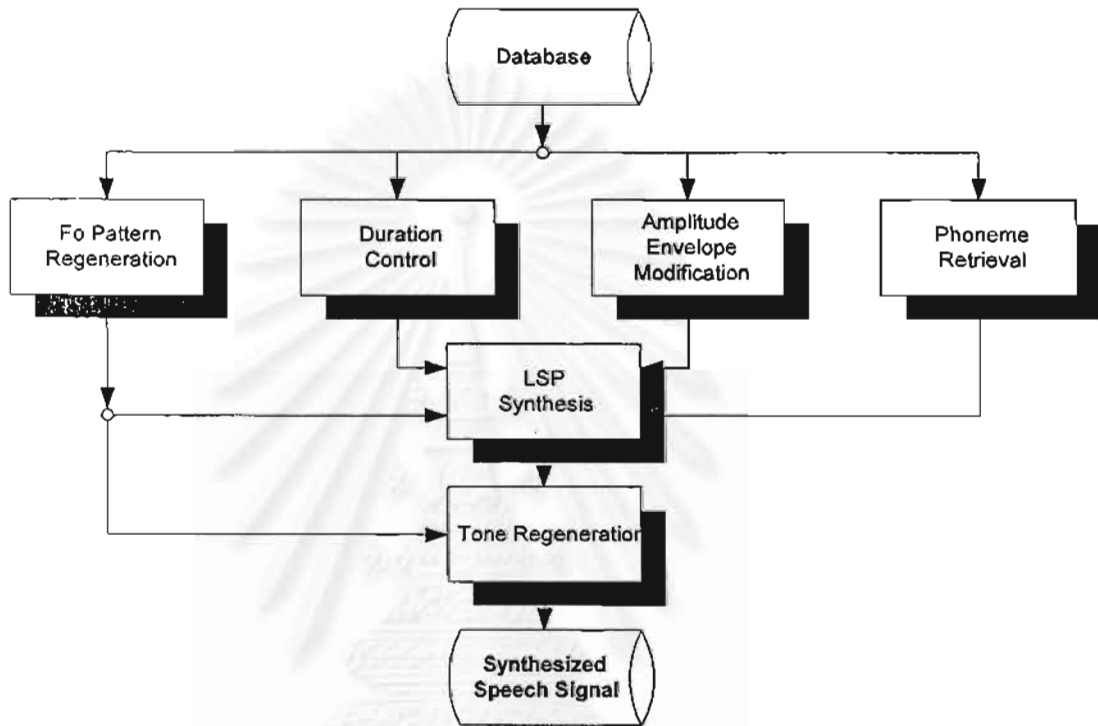
คือ บริเวณที่เกิดจากการประมาณค่าเพื่อเชื่อมต่อระหว่างหน่วยเสียง

โครงสร้างพยางค์สังเคราะห์นี้ครอบคลุมรูปแบบพยางค์ในภาษาไทยทั้งหมด โดยรูปแบบที่สามารถเกิดขึ้นได้มีดังนี้

1.  $C_i T_{Ci} V_1$  เช่น กะ ปลา คำ เป็นต้น
2.  $C_i T_{Ci} V_1 T_{cfn}$  เช่น กาก อาบ กั๊บ เป็นต้น
3.  $C_i T_{Ci} V_1 C_{fn}$  เช่น กาว ไก่ ลัย เป็นต้น
4.  $C_i T_{Ci} V_1 T_{cfn} C_{fn}$  เช่น ปาน ก้าม ร้าง เป็นต้น
5.  $C_i T_{Ci} V_1 V_2$  เช่น เพื้อ เพียะ ตัว เป็นต้น
6.  $C_i T_{Ci} V_1 V_2 T_{cfn}$  เช่น เปลือก เรียก ปวก เป็นต้น
7.  $C_i T_{Ci} V_1 V_2 C_{fn}$  เช่น กล้วย เกรียว เสือ เป็นต้น
8.  $C_i T_{Ci} V_1 V_2 T_{cfn} C_{fn}$  เช่น เปลี่ยน เพื่อน เชื่อม เป็นต้น

## 4.2 ภาพรวมกระบวนการสังเคราะห์เสียง

กระบวนการสังเคราะห์เสียงจากฐานข้อมูลที่เก็บไว้แสดงดังรูปที่ 48



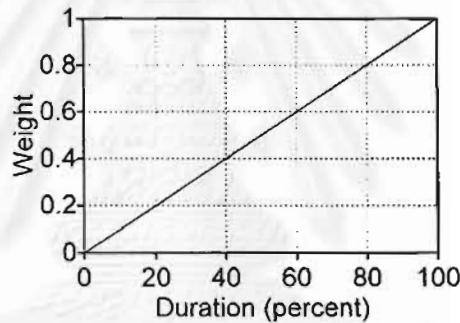
รูปที่ 48 แผนผังกระบวนการสังเคราะห์เสียง

การสังเคราะห์เริ่มจากการนำหน่วยเสียงที่เป็นส่วนประกอบของพยางค์ที่ต้องการสังเคราะห์ซึ่งถูกเก็บในรูปของคู่เส้นสเปกตรัมและสัญญาณตกค้าง มาเรียงลำดับตามโครงสร้างพยางค์สังเคราะห์ แล้วทำการประมาณค่าพารามิเตอร์คู่เส้นสเปกตรัมและสัญญาณตกค้างบริเวณรอยต่อระหว่างหน่วยเสียงเพื่อสร้างเสียงพยัญชนะและสระผสมต่างๆ กลับคืนหรือในกรณีของ  $V_1$ ,  $V_2$  และ  $C$ , อาศัยการทำซ้ำของพารามิเตอร์คู่เส้นสเปกตรัมและสัญญาณตกค้าง ในระหว่างการประมาณค่าหรือการทำซ้ำดังกล่าวจะทำการคำนวณค่าความถี่มูลฐานใหม่โดยมีระยะเวลาตามข้อมูลระยะเวลาที่เก็บไว้ในฐานข้อมูล ความถี่มูลฐานดังกล่าวจะนำไปใช้คำนวณตำแหน่งพิทช์อ้างอิงระหว่างการประมาณค่าพารามิเตอร์คู่เส้นสเปกตรัมและสัญญาณตกค้างข้างต้น รวมทั้งการคำนวณกรอบแอมพลิจูดกรอบใหม่ เมื่อได้ชุดพารามิเตอร์คู่เส้นสเปกตรัมและสัญญาณตกค้างครบทุกกรอบของพยางค์ที่ต้องการสังเคราะห์แล้ว นำค่าพารามิเตอร์คู่เส้นสเปกตรัมแปลงกลับเป็น

ค่าสัมประสิทธิ์การเข้ารหัสการทำนายแบบเชิงเส้น แล้วส่งผ่านวงจรกรองสัณเคราะห์การทำนายแบบเชิงเส้นผกผันหรือร่วมกับสัญญาณตกค้างเพื่อสังเคราะห์กรอบเสียงในเชิงเวลากลับคืน กรอบสัญญาณเสียงที่ได้จะถูกปรับกรอบแอมพลิจูดแล้วนำไปทำการสังเคราะห์เสียงวรรณยุกต์กลับด้วยโซลา ซึ่งจะได้เสียงสังเคราะห์ของพยางค์ที่ต้องการ

#### 4.3 การประมาณค่าพารามิเตอร์

จากทฤษฎีการประมาณค่าสัมประสิทธิ์ในบทที่ 2 นำมาใช้กับค่าพารามิเตอร์คู่เส้นสเปกตรัมและสัญญาณตกค้าง โดยใช้สมการ (2.30) โดยใช้การประมาณค่าเป็นแบบเชิงเส้นและมีค่าน้ำหนักเป็นเส้นตรงเมื่อเทียบกับระยะเวลาตามกราฟในรูปที่ 49



รูปที่ 49 กราฟค่าน้ำหนักสำหรับการประมาณเชิงเส้น

ซึ่งจะได้สมการสำหรับประมาณค่าพารามิเตอร์คู่เส้นสเปกตรัมดังสมการ 4.1

$$\tilde{p}_k = w_i \cdot p_{k, \text{future}} + (1 - w_i) \cdot p_{k, \text{past}} \quad (4.1)$$

โดยที่

- $\tilde{p}_k$  คือ ค่าพารามิเตอร์ของคู่สเปกตรัมแบบเส้นตัวที่  $k$  ที่ได้จากการประมาณค่า ณ เวลาที่ทำการประมาณค่า
- $p_{k, \text{past}}$  คือ ค่าพารามิเตอร์ของคู่สเปกตรัมแบบเส้นตัวที่  $k$  ของสัญญาณชุดเริ่มต้น
- $p_{k, \text{future}}$  คือ ค่าพารามิเตอร์ของคู่สเปกตรัมแบบเส้นตัวที่  $k$  ของสัญญาณชุดปลายทาง
- $w$  คือ ค่าน้ำหนักของการประมาณค่าแบบเชิงเส้น

สำหรับในส่วนสัญญาณตกค้างจะทำการประมาณค่าโดยการประมาณเชิงเส้นดังสมการที่ 4.2

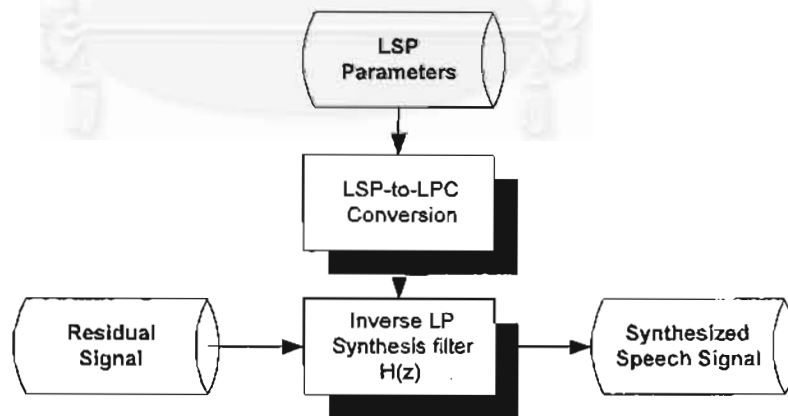
$$\tilde{r}_{ST} = w_i \cdot r_{ST, future} + (1 - w_i) \cdot r_{ST, past} \quad (4.2)$$

โดยที่

- $\tilde{r}_{ST}$  คือ สัญญาณตกค้างช่วงสั้นที่ได้จากการประมาณค่า
- $r_{ST, past}$  คือ สัญญาณตกค้างช่วงสั้นชุดเริ่มต้น
- $r_{ST, future}$  คือ สัญญาณตกค้างช่วงสั้นชุดปลายทาง
- $w_i$  คือ ค่าน้ำหนักของการประมาณค่าแบบเชิงเส้น

#### 4.4 การสังเคราะห์ด้วยคู่เส้นสเปกตรัม

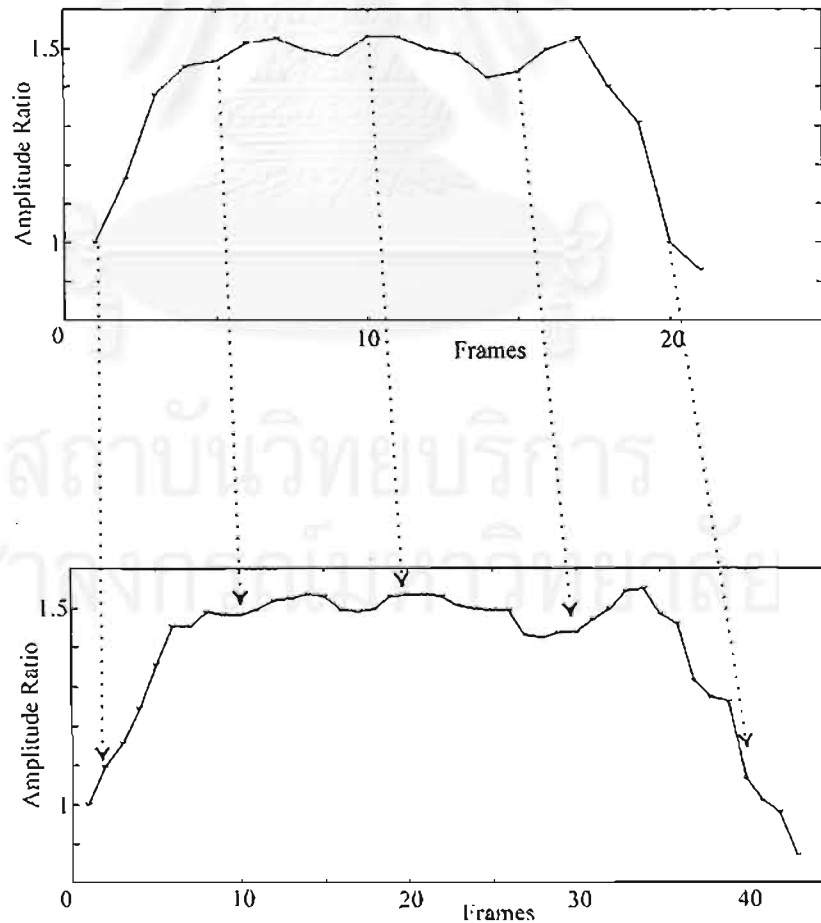
การสังเคราะห์รูปคลื่นของเสียงกลับคืนเริ่มจากการนำค่าพารามิเตอร์คู่เส้นสเปกตรัมของหน่วยเสียงและที่ได้จากการประมาณค่ามาแปลงกลับเป็นสัมประสิทธิ์การเข้ารหัสการทำนายแบบเชิงเส้น แล้วนำค่าสัมประสิทธิ์การเข้ารหัสการทำนายแบบเชิงเส้นที่คำนวณได้และสัญญาณตกค้างของแต่ละกรอบเสียงมาแปลงกลับเป็นสัญญาณเสียงเชิงเวลาโดยใช้วงจรกรองสังเคราะห์การทำนายแบบเชิงเส้นผกผัน  $H(z)$  ในสมการ 2.12 ดังแสดงขั้นตอนในรูปที่ 50



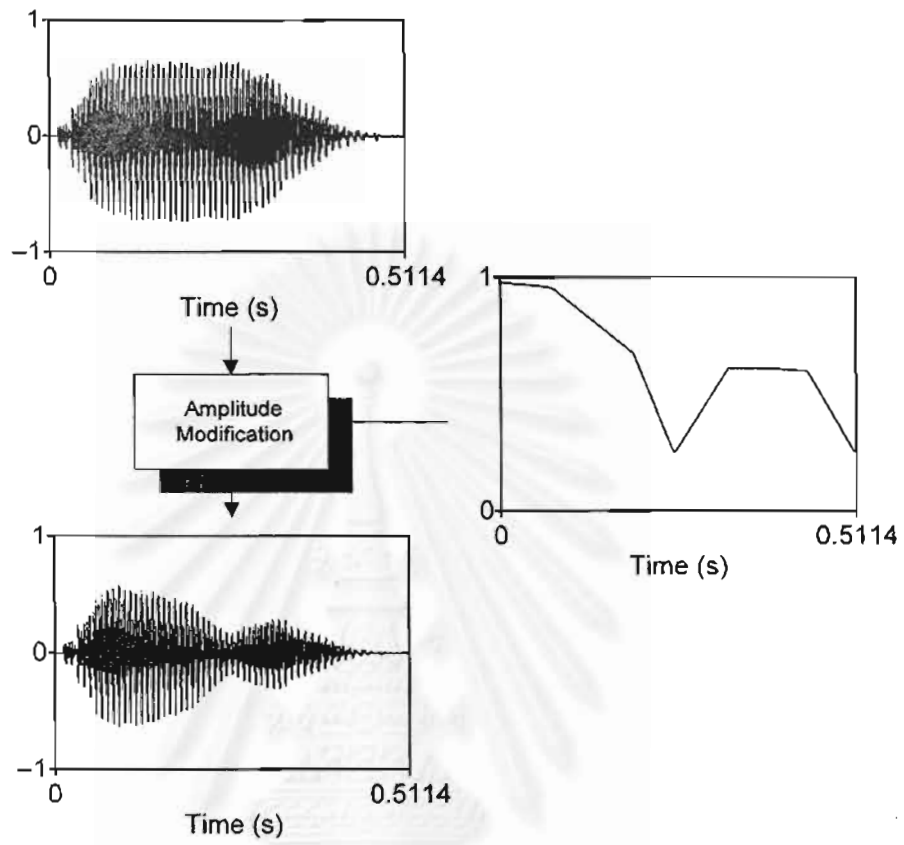
รูปที่ 50 การสังเคราะห์เสียงด้วยใช้วงจรกรองสังเคราะห์การทำนายแบบเชิงเส้นผกผัน  $H(z)$

#### 4.5 การปรับกรอบแอมพลิจูด

การปรับกรอบแอมพลิจูดจะกระทำก่อนทำการปรับความถี่มูลฐาน ในขั้นแรกค่าอัตราส่วนแอมพลิจูดที่เก็บไว้ในฐานข้อมูลจำนวน 100 ค่าสำหรับเสียงวรรณยุกต์แต่ละเสียงจะนำไปทำการประมาณค่าเชิงเส้นเพื่อให้มีระยะเวลาของกรอบแอมพลิจูดเท่ากับระยะเวลาของเสียงที่จะสังเคราะห์ดังในรูปที่ 51 การปรับกรอบแอมพลิจูดเริ่มตั้งแต่กรอบเสียงสุดท้ายของเสียง C<sub>1</sub> โดยการนำค่าอัตราส่วนแอมพลิจูดคูณกับค่าแอมพลิจูดของสัญญาณเสียงแต่ละกรอบที่ได้จากการการสังเคราะห์ด้วยตัวคูณเส้นสเปกตรัม เนื่องจากค่าอัตราส่วนแอมพลิจูดค่าแรกมีค่าเป็นหนึ่ง ดังนั้นเมื่อคูณค่าอัตราส่วนดังกล่าวกับค่าแอมพลิจูดของสัญญาณเสียงเดิม กรอบแอมพลิจูดของเสียงที่สังเคราะห์จะมีความต่อเนื่องกับสัญญาณเสียงเดิมดังในรูปที่ 52 แสดงการปรับกรอบแอมพลิจูดและผลที่ได้จากการปรับกรอบแอมพลิจูดแล้ว การปรับกรอบแอมพลิจูดด้วยอัตราส่วนแอมพลิจูดดังกล่าวทำให้เสียงที่สังเคราะห์ได้ยังคงลักษณะธรรมชาติของกรอบเสียงก่อนการปรับและมีลักษณะกรอบแอมพลิจูดตามที่กำหนดในขณะเดียวกัน



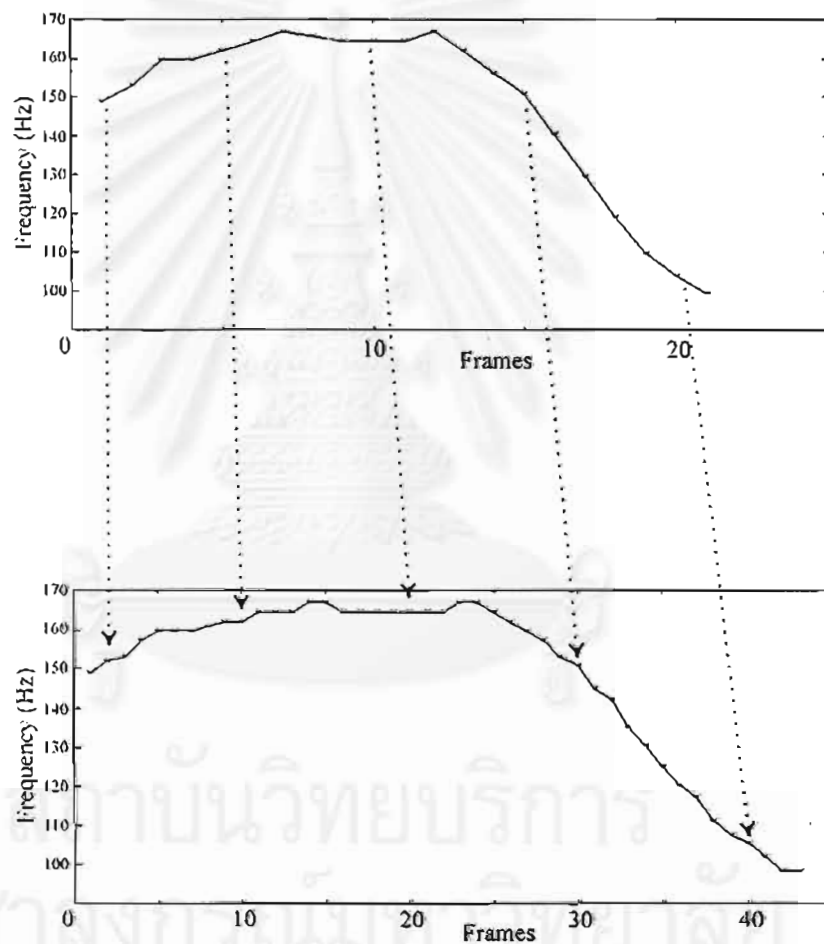
รูปที่ 51 การประมาณค่ากรอบแอมพลิจูด



รูปที่ 52 การปรับกรอบแอมพลิจูด

#### 4.6 การปรับความถี่มูลฐาน

รูปแบบความถี่มูลฐานที่เก็บไว้ในฐานข้อมูลจำนวน 100 ค่าสำหรับเสียงวรรณยุกต์แต่ละเสียงจะถูกนำมาประมาณค่าเชิงเส้นเพื่อหาค่าพิทช์ที่มีระยะเวลาเท่ากับระยะเวลาและจำนวนพิทช์ของเสียงที่จะสังเคราะห์ดังรูปที่ 53 แล้วนำค่าพิทช์ดังกล่าวที่ได้เป็นตำแหน่งอ้างอิงในการปรับความถี่มูลฐานด้วยวิธีโซลา

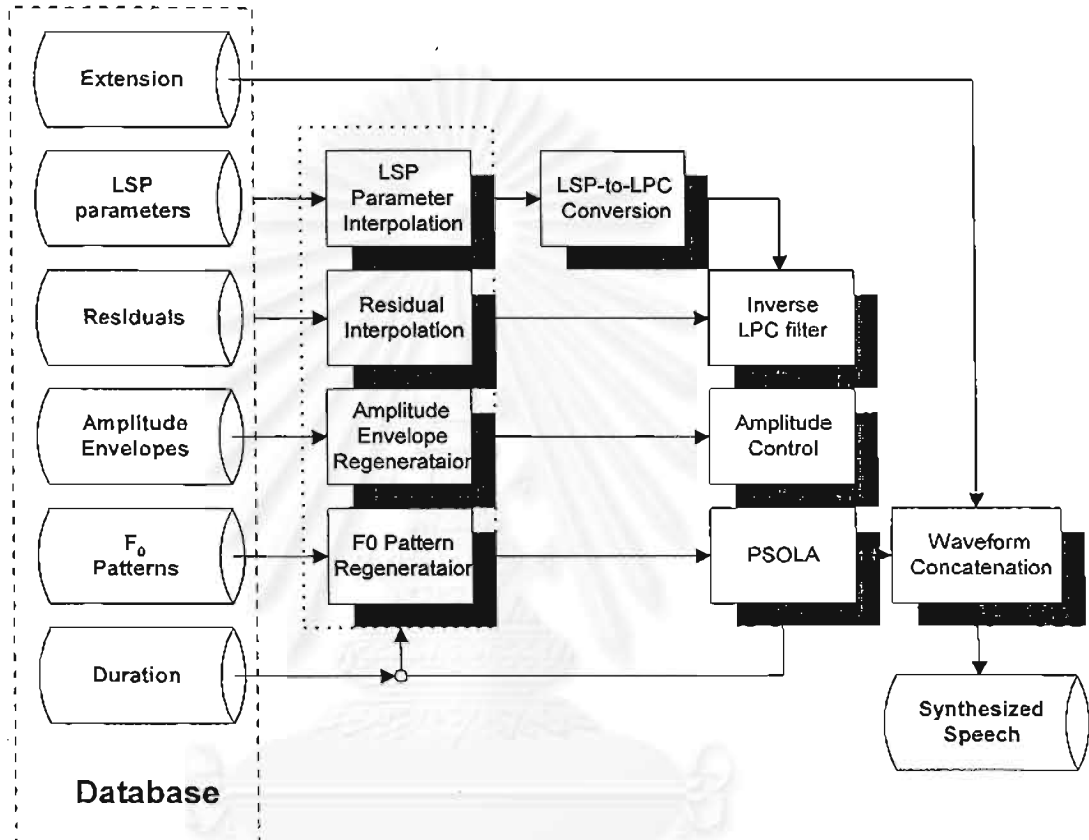


รูปที่ 53 การประมาณค่ารูปแบบความถี่มูลฐาน



#### 4.7 สรุปขั้นตอนการสังเคราะห์เสียง

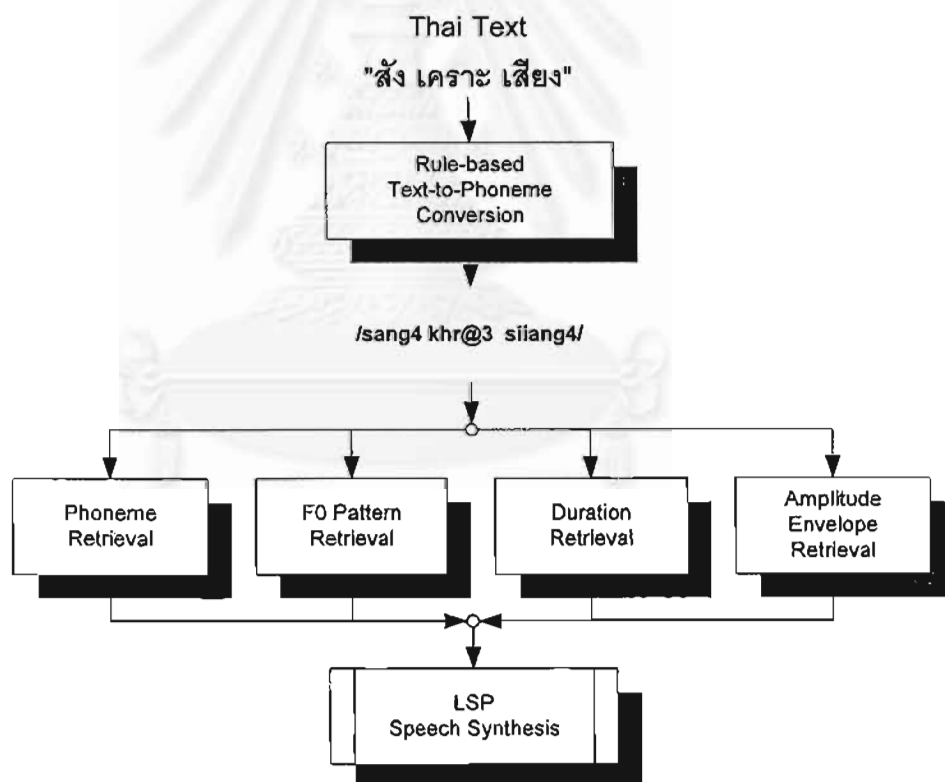
การสังเคราะห์เสียงสรุปขั้นตอนได้ดังในรูปที่ 54



รูปที่ 54 ขั้นตอนการสังเคราะห์เสียงโดยสรุป

#### 4.8 โปรแกรมสังเคราะห์เสียง

โปรแกรมสังเคราะห์เสียงที่นำเสนอนี้มีลักษณะเป็นโปรแกรมสังเคราะห์เสียงจากข้อความ (Text-to-Speech Synthesizer) โดยรับข้อมูลเป็นคำอ่านแยกแต่ละพยางค์ด้วยเว้นวรรคของประโยคที่ต้องการสังเคราะห์ คำอ่านของแต่ละพยางค์จะนำมาวิเคราะห์หาสัญลักษณ์แทนหน่วยเสียงและวรรณยุกต์โดยผ่านกระบวนการแปลงข้อความเป็นหน่วยเสียงโดยใช้กฎ หลังจากกระบวนการดังกล่าว สัญลักษณ์แทนหน่วยเสียงและวรรณยุกต์ดังกล่าวจะนำไปวิเคราะห์หาหน่วยเสียงจริง รูปแบบวรรณยุกต์ กรอบแอมพลิจูดและระยะเวลา ซึ่งหน่วยเสียงและค่าคุณสมบัติต่างๆ ที่ได้นี้จะส่งให้กระบวนการสังเคราะห์เสียงเพื่อสังเคราะห์เสียงต่อไป โดยสรุปเขียนเป็นแผนผังได้ดังรูปที่ 55



รูปที่ 55 แผนผังโปรแกรมสังเคราะห์เสียง

## บทที่ 5 ผลการสังเคราะห์เสียง

จากวิธีการสังเคราะห์โดยใช้คู่เส้นสเปกตรัมตามที่นำเสนอไป เมื่อนำมาทำการสังเคราะห์พยางค์ต่างๆ ในภาษาไทยได้ผลการสังเคราะห์ดังที่จะแสดงในบทนี้

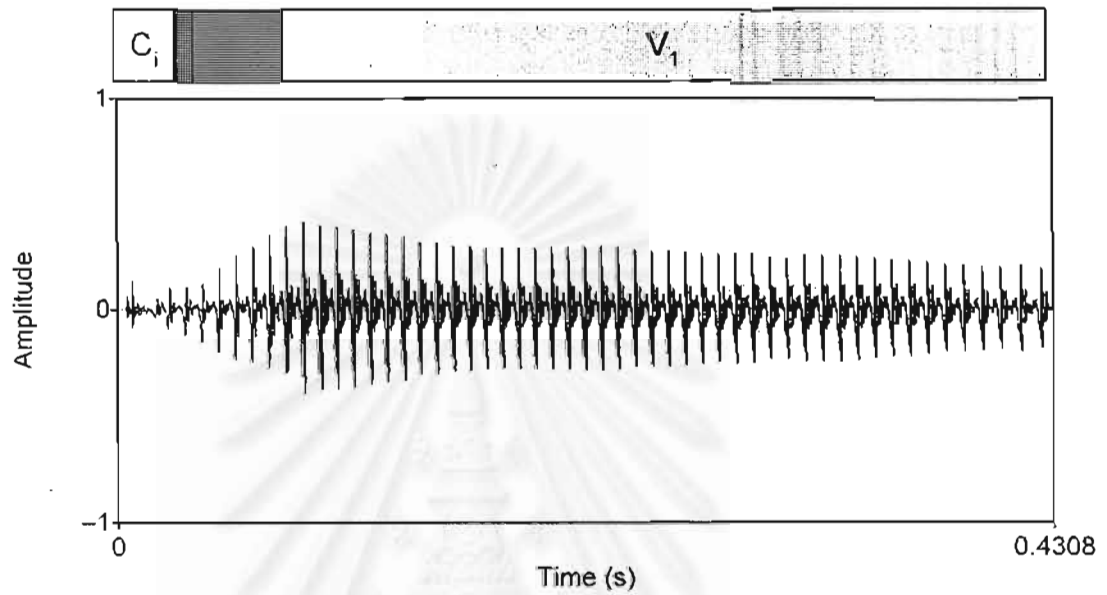
### 5.1 ผลการสังเคราะห์ในเชิงเวลาและเชิงความถี่

เมื่อทำการทดลองสังเคราะห์เสียงพยางค์ในภาษาไทยตามโครงสร้างพยางค์สังเคราะห์ โดยในการแสดงผลการสังเคราะห์ใช้สัญลักษณ์ในการบรรยายดังต่อไปนี้

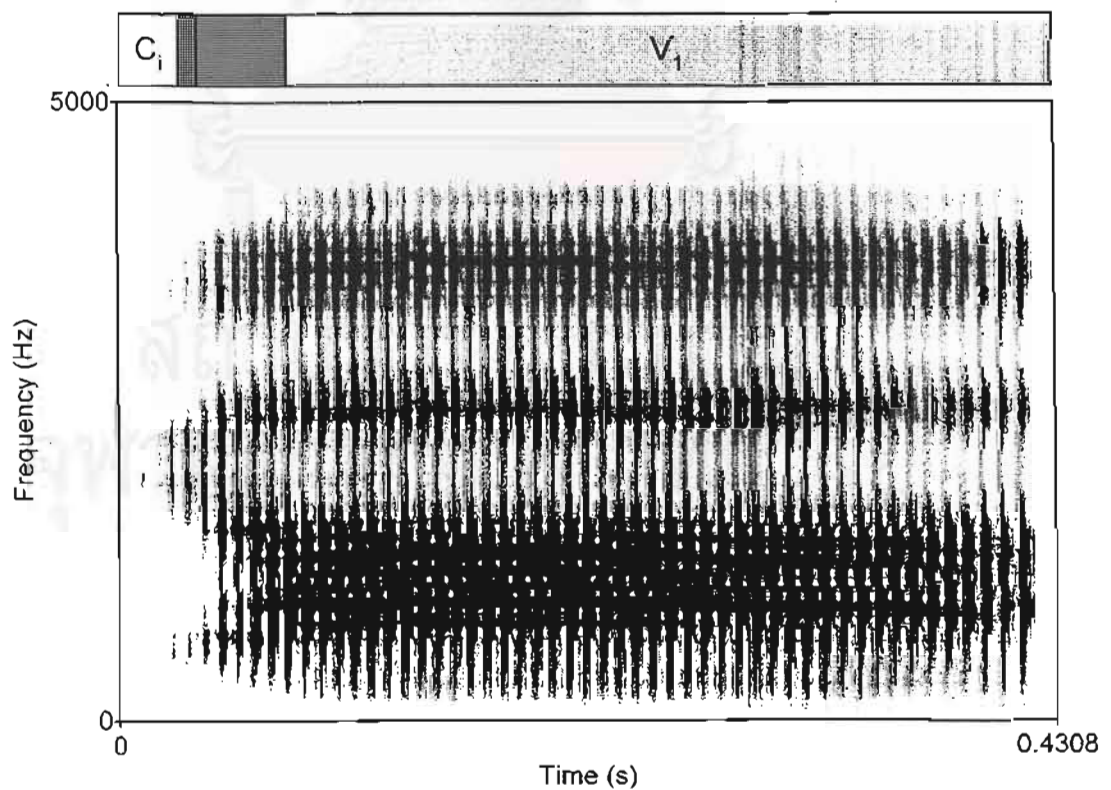
$C_i$	แสดงเสียงบริเวณที่เกิดจากหน่วยเสียง $C_i$
$T_{Ci}$	แสดงเสียงบริเวณที่เป็นจุดอ้างอิงที่ของหน่วยเสียง $T_{Ci}$
$V_1$	แสดงเสียงบริเวณที่เกิดการทำซ้ำของหน่วยเสียง $V_1$
$V_2$	แสดงเสียงบริเวณที่เกิดการทำซ้ำของหน่วยเสียง $V_2$
$T_{Cfn}$	แสดงเสียงบริเวณที่เป็นจุดอ้างอิงที่ของหน่วยเสียง $T_{Cfn}$
$C_{fn}$	แสดงเสียงบริเวณที่เกิดการทำซ้ำของหน่วยเสียง $C_{fn}$
	แสดงเสียงบริเวณที่เกิดจากการประมาณค่าระหว่างหน่วยเสียง

ตัวอย่างผลการสังเคราะห์เสียงพยางค์แบ่งตามลักษณะโครงสร้างพยางค์ที่เกิดขึ้นได้ดังนี้

1. โครงสร้าง  $C_i T_c V_1$  โดยทดลองสังเคราะห์เสียงพยางค์ว่า "กา" ได้ผลการสังเคราะห์ที่มีรูปคลื่นและสเปกโตรแกรมดังแสดงในรูปที่ 56 และ 57 ตามลำดับ

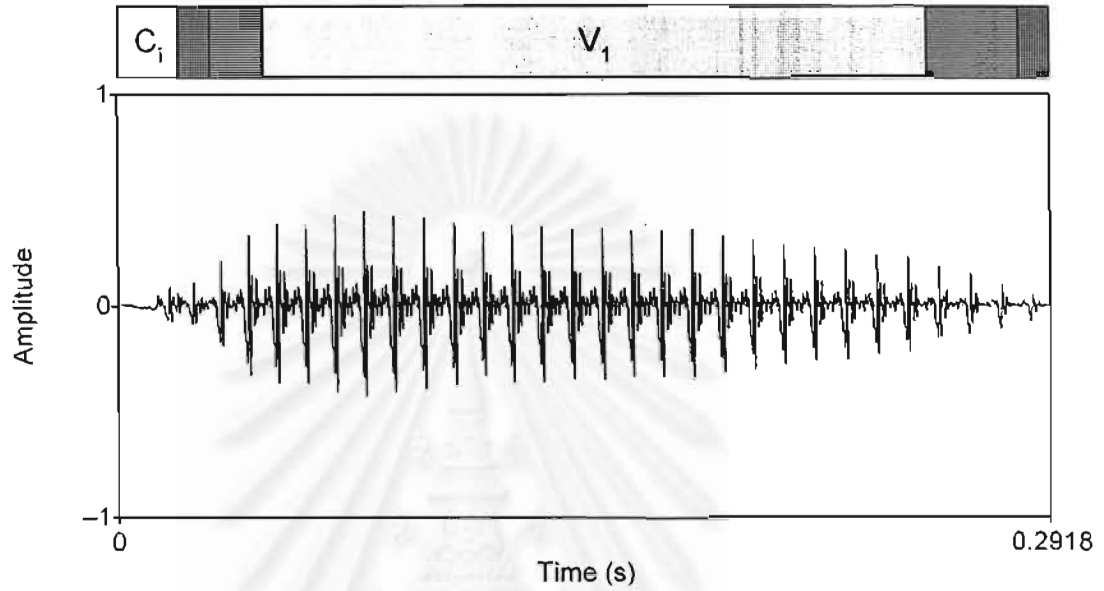


รูปที่ 56 รูปคลื่นเสียง "กา"

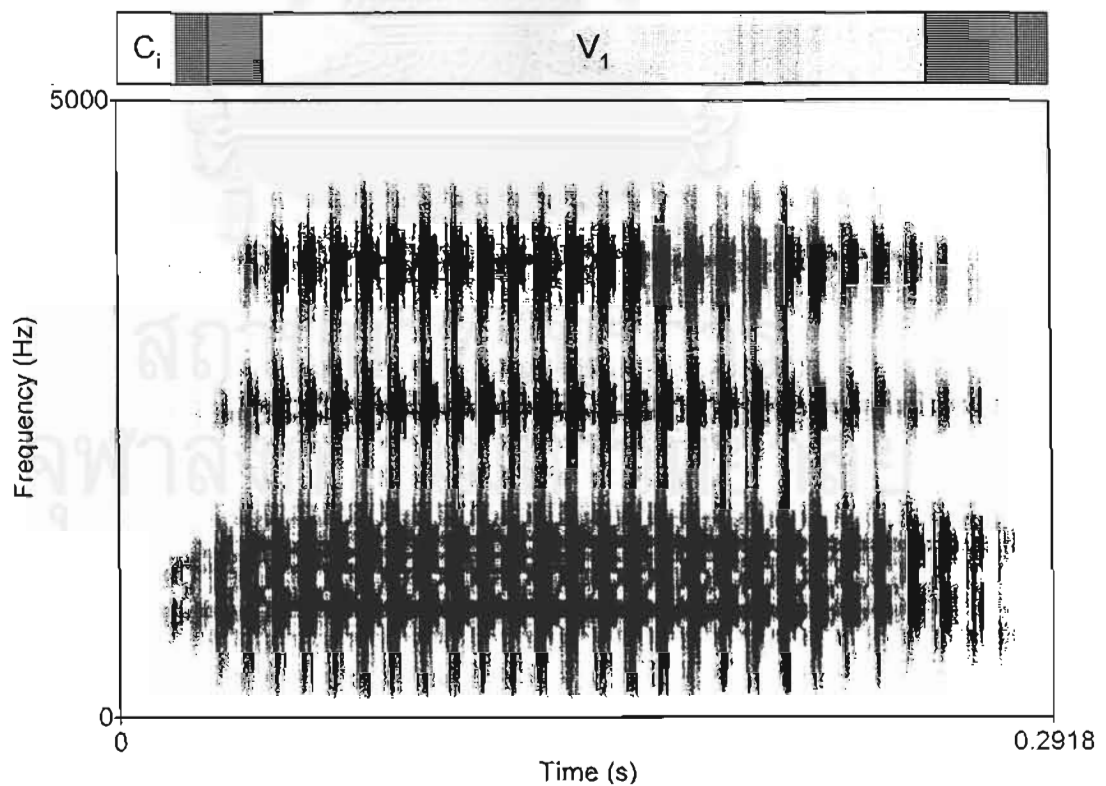


รูปที่ 57 สเปกโตรแกรมของเสียง "กา"

2. โครงสร้าง  $C_i, T_{ci}, V_1, T_{cm}$  โดยทดลองสังเคราะห์เสียงพยางค์ว่า "ปึก" ได้ผลการสังเคราะห์ที่มีรูปคลื่นและสเปกโตรแกรมดังแสดงในรูปที่ 58 และ 59 ตามลำดับ

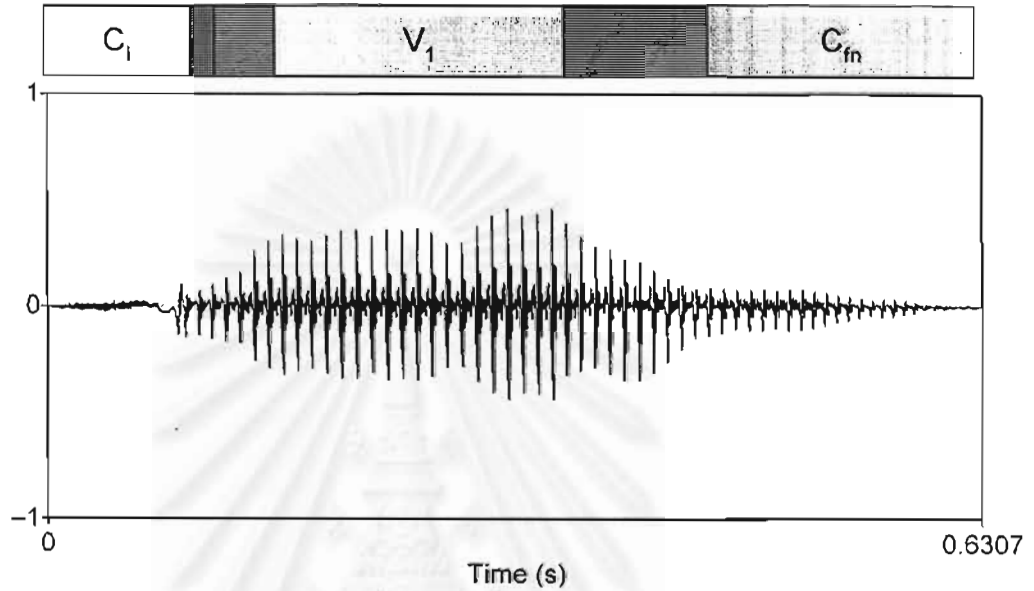


รูปที่ 58 รูปคลื่นเสียง "ปึก"

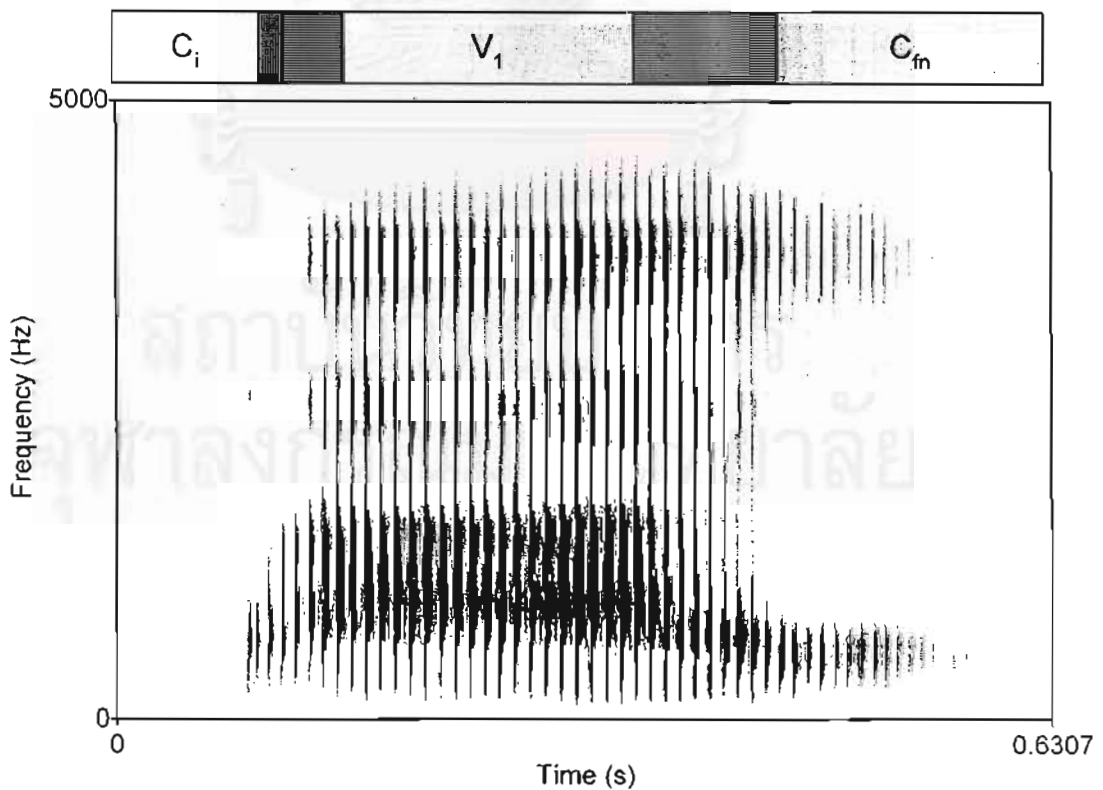


รูปที่ 59 สเปกโตรแกรมของเสียง "ปึก"

3. โครงสร้าง  $C_i T_{C_i} V_1 C_m$  โดยทดลองสังเคราะห์เสียงพยางค์ว่า "สาย" ได้ผลการสังเคราะห์ที่มีรูปคลื่นและสเปกโตรแกรมดังแสดงในรูปที่ 60 และ 61 ตามลำดับ

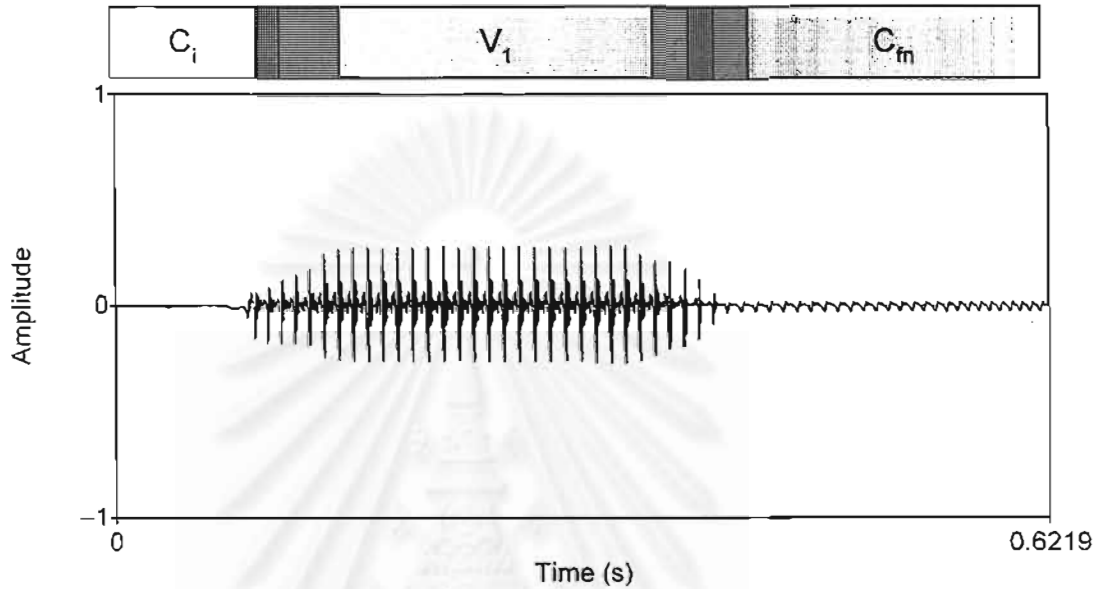


รูปที่ 60 รูปคลื่นเสียง "สาย"

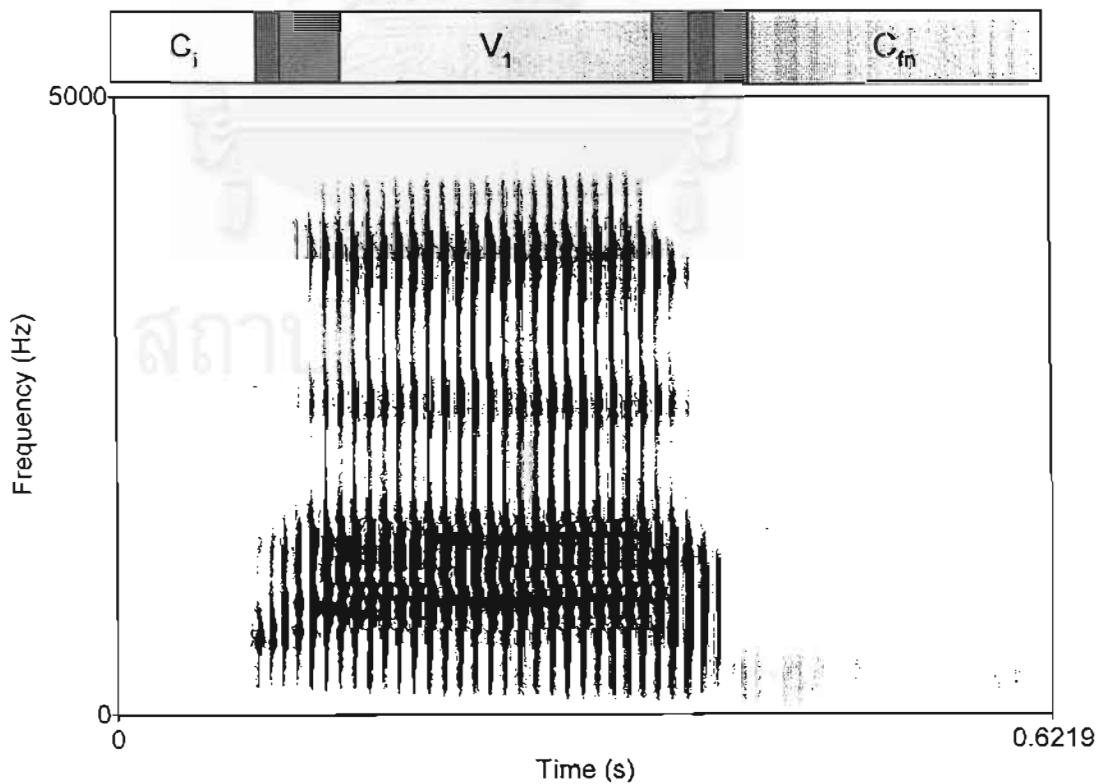


รูปที่ 61 สเปกโตรแกรมของเสียง "สาย"

4. โครงสร้าง  $C_i T_{cl} V_1 T_{cln} C_m$  โดยทดลองสังเคราะห์เสียงพยางค์ว่า "สาม" ได้ผลการสังเคราะห์ที่มีรูปคลื่นและสเปกโตรแกรมดังแสดงในรูปที่ 62 และ 63 ตามลำดับ

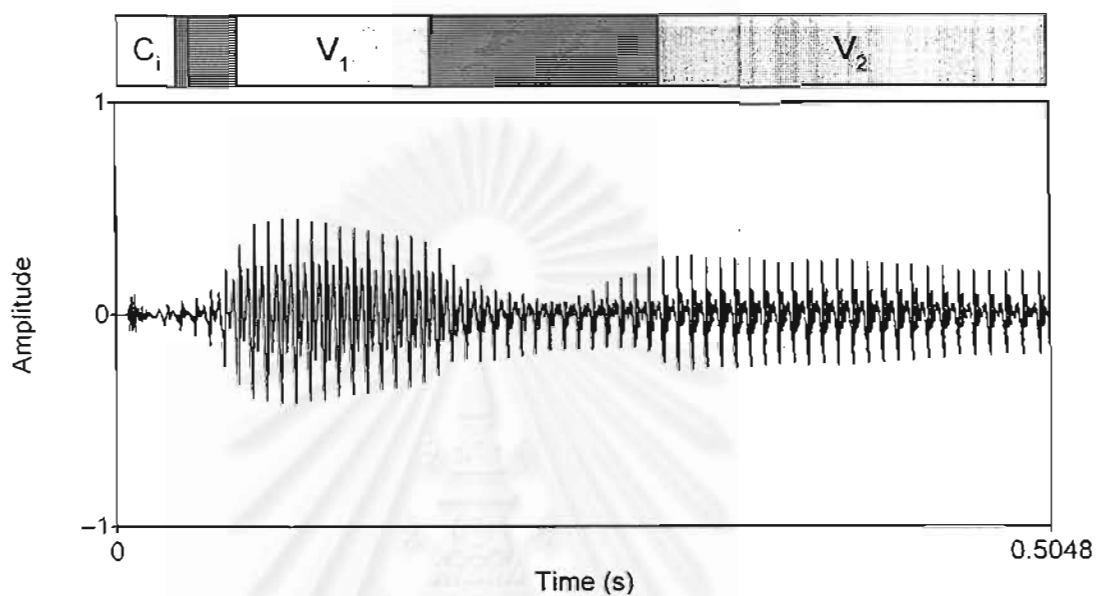


รูปที่ 62 รูปคลื่นเสียง "สาม"

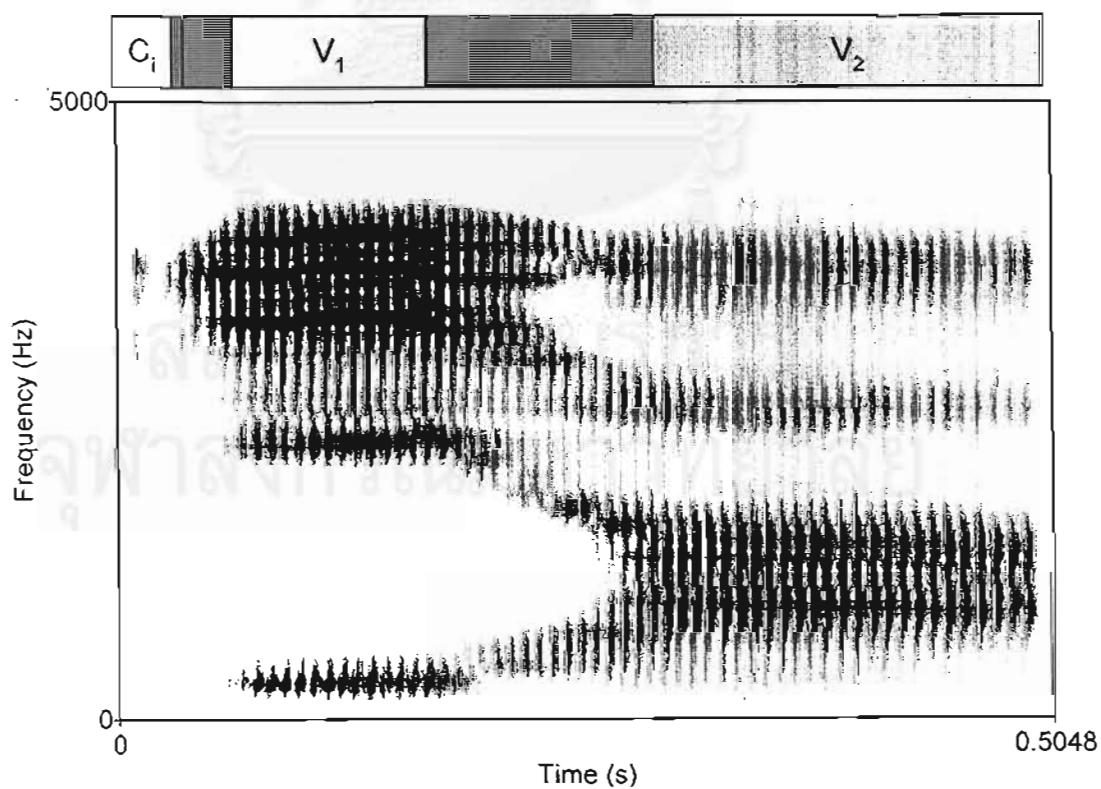


รูปที่ 63 สเปกโตรแกรมของเสียง "สาม"

5. โครงสร้าง  $C_i, T_{ci}, V_1, V_2$  โดยทดลองสังเคราะห์เสียงพยางค์ว่า "เกีย" ได้ผลการสังเคราะห์ที่มีรูปคลื่นและสเปกโตรแกรมดังแสดงในรูปที่ 64 และ 65 ตามลำดับ



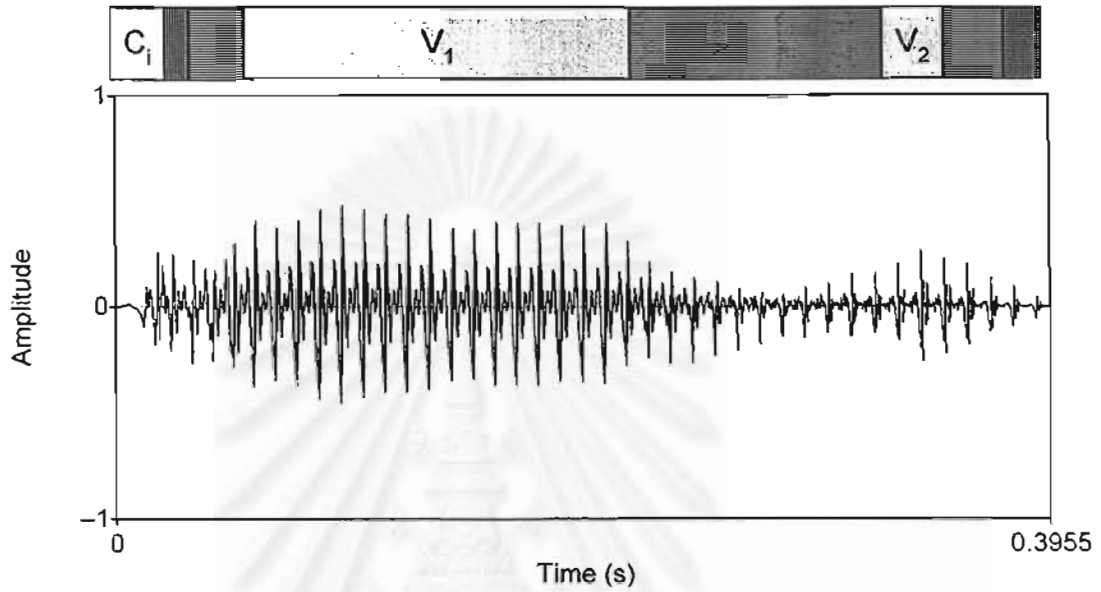
รูปที่ 64 รูปคลื่นเสียง "เกีย"



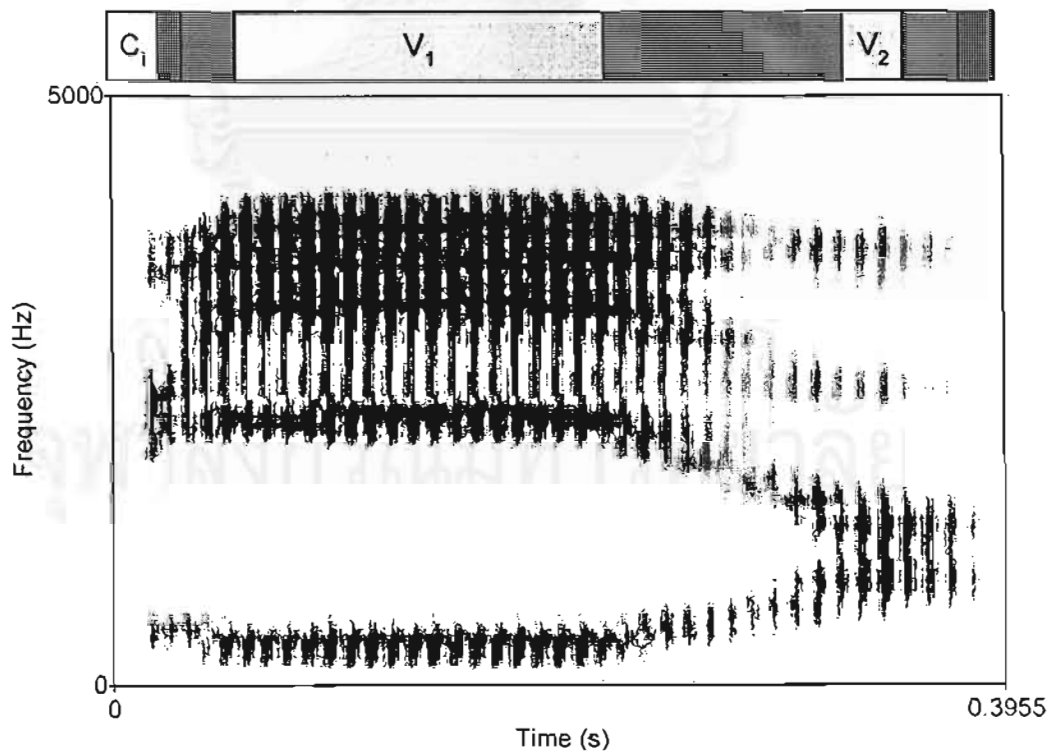
รูปที่ 65 สเปกโตรแกรมของเสียง "เกีย"



6. โครงสร้าง  $C_i T_{C_i} V_1 V_2 T_{C_m}$  โดยทดลองสังเคราะห์เสียงพยางค์ว่า "เป็ยก" ได้ผลการสังเคราะห์ที่มีรูปคลื่นและสเปกโตรแกรมดังแสดงในรูปที่ 66 และ 67 ตามลำดับ

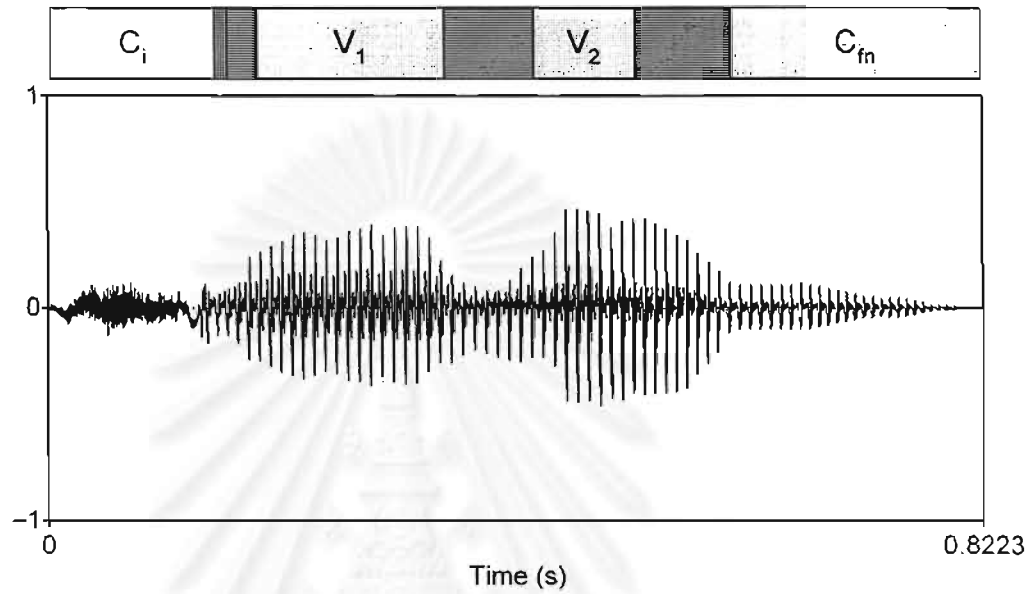


รูปที่ 66 รูปคลื่นเสียง "เป็ยก"

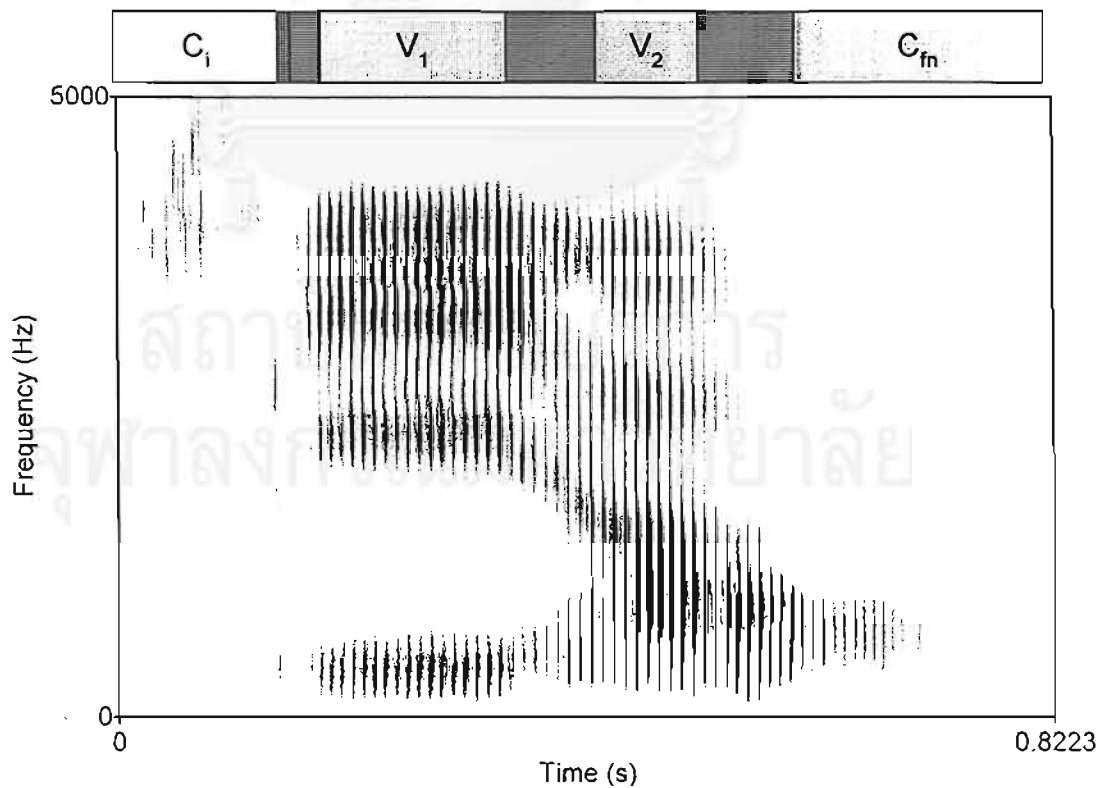


รูปที่ 67 สเปกโตรแกรมของเสียง "เป็ยก"

7. โครงสร้าง  $C_i T_{c_i} V_1 V_2 C_{f_n}$  โดยทดลองสังเคราะห์เสียงพยางค์ว่า "เสียว" ได้ผลการสังเคราะห์ที่มีรูปคลื่นและสเปกโตรแกรมดังแสดงในรูปที่ 68 และ 69 ตามลำดับ

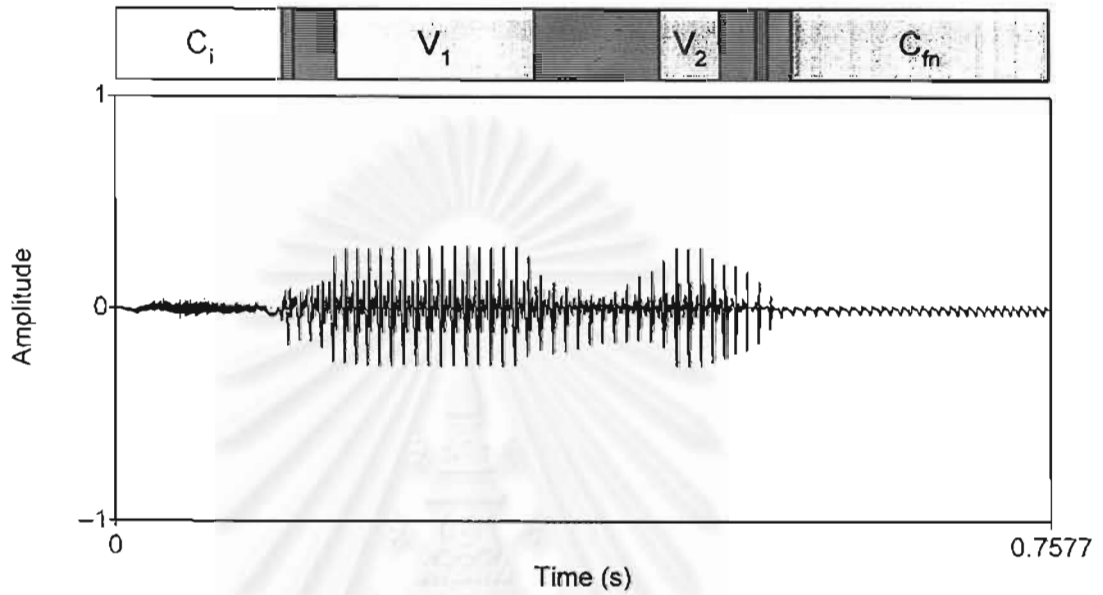


รูปที่ 68 รูปคลื่นเสียง "เสียว"

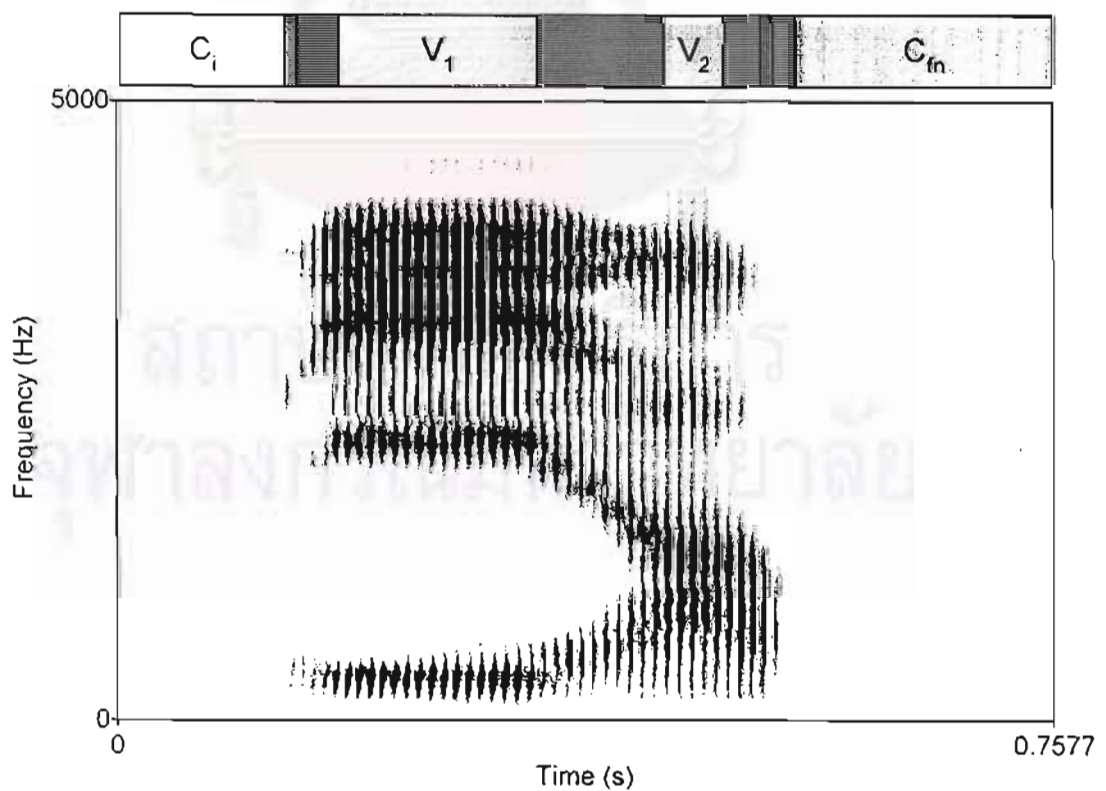


รูปที่ 69 สเปกโตรแกรมของเสียง "เสียว"

8. โครงสร้าง  $C_i T_{ci} V_1 V_2 T_{cm} C_m$  โดยทดลองสังเคราะห์เสียงพยางค์ว่า "เสียม" ได้ผลการสังเคราะห์ที่มีรูปคลื่นและสเปกโตรแกรมดังแสดงในรูปที่ 70 และ 71 ตามลำดับ



รูปที่ 70 รูปคลื่นเสียง "เสียม"

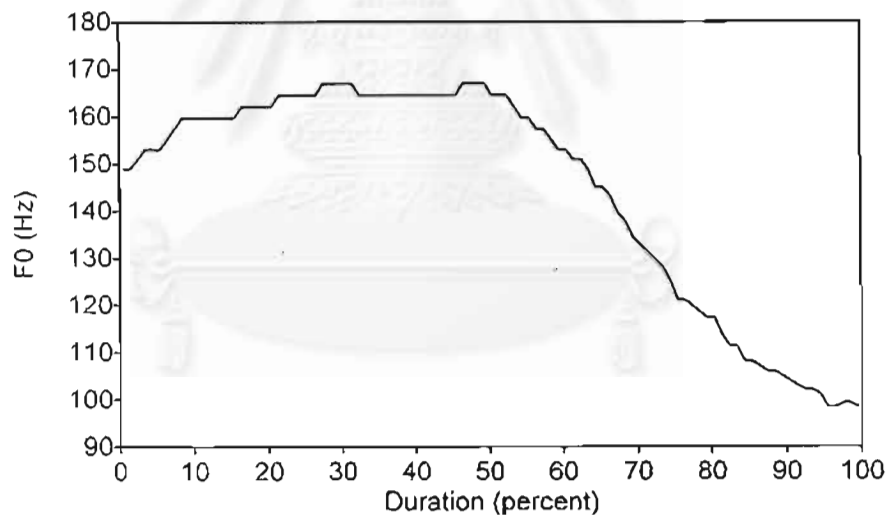


รูปที่ 71 สเปกโตรแกรมของเสียง "เสียม"

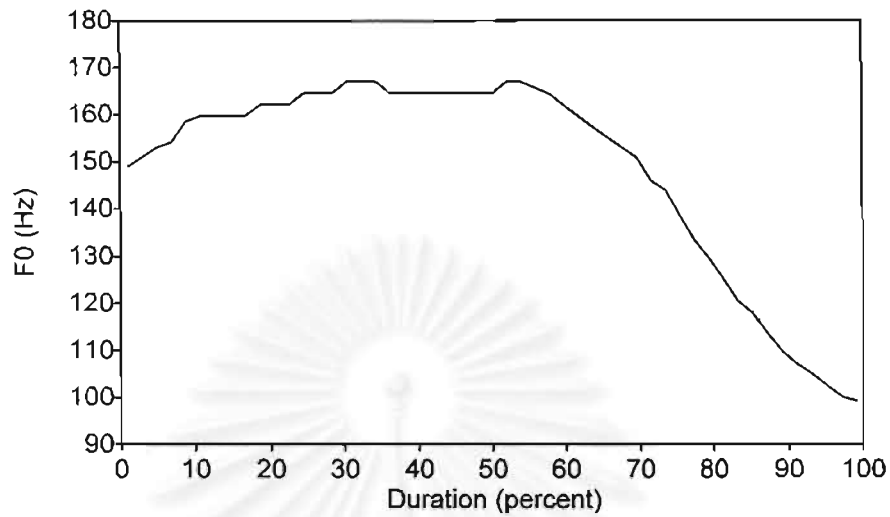
จากผลการสังเคราะห์พบว่าสามารถสังเคราะห์เสียงพยางค์ภาษาไทยได้ทุกพยางค์โดยใช้โครงสร้างพยางค์สังเคราะห์และเสียงที่สังเคราะห์ส่วนใหญ่ขึ้นมามีฟอร์แมนต์บริเวณรอยต่อระหว่างหน่วยเสียงที่ต่อเนื่องซึ่งการสังเคราะห์บริเวณที่ฟอร์แมนต์มีการบิดเบนดังกล่าวเป็นการสังเคราะห์เสียงพยัญชนะและสระผสมกลับคืน ในบางเสียงที่ฟอร์แมนต์เกิดการขาดช่วงอาจเนื่องมาจากคุณภาพของหน่วยเสียง เพราะในเสียงที่สังเคราะห์ส่วนใหญ่ฟอร์แมนต์สามารถเชื่อมต่อกันอย่างต่อเนื่องโดยใช้คุณสมบัติการประมาณค่าพารามิเตอร์คู่เส้นสเปกตรัม

## 5.2 ผลการสังเคราะห์เสียงวรรณยุกต์

การสังเคราะห์โดยใช้วิธีทีเอสโอแอลเอและการประมาณค่าเชิงเส้นเพื่อสังเคราะห์เสียงวรรณยุกต์แสดงตัวอย่างผลการสังเคราะห์เสียงวรรณยุกต์โทของเสียง "อ้อ" โดยใช้รูปแบบความถี่มูลฐานในฐานะข้อมูลดังแสดงในรูปที่ 72 ซึ่งได้ผลการสังเคราะห์ดังในรูปที่ 73



รูปที่ 72 รูปแบบความถี่มูลฐานของเสียงวรรณยุกต์โทในฐานะข้อมูล

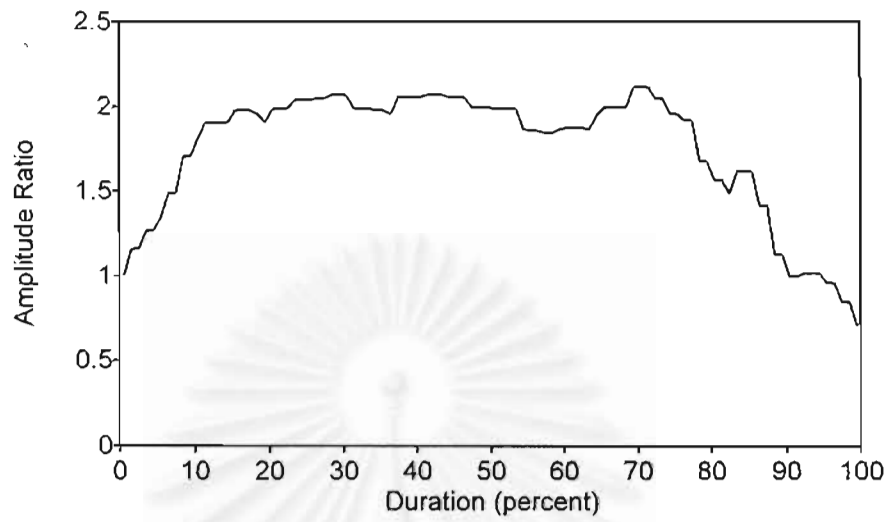


รูปที่ 73 ผลการสังเคราะห์รูปแบบความถี่มูลฐานของเสียงวรรณยุกต์โท

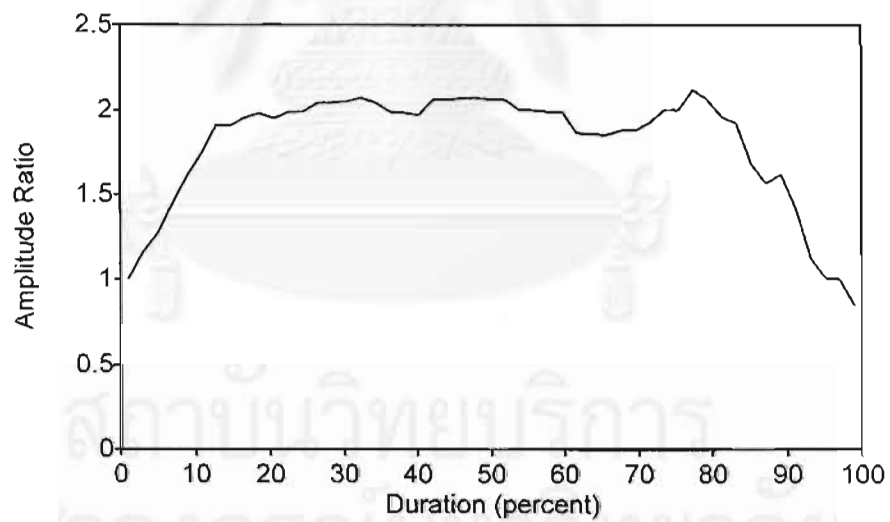
จากผลการสังเคราะห์เสียงวรรณยุกต์แสดงให้เห็นว่าการใช้วิธีพีเอสโอแอลเอและการประมาณค่าเชิงเส้นสามารถสังเคราะห์เสียงวรรณยุกต์ที่มีรูปแบบความถี่มูลฐานเช่นเดียวกับรูปแบบความถี่มูลฐานที่เป็นต้นถึงแม้ว่าระยะเวลาและจำนวนพิทช์ของเสียงสังเคราะห์ไม่เท่ากับรูปแบบที่เก็บในฐานข้อมูล

### 5.3 ผลการสังเคราะห์กรอบแอมพลิจูด

ตัวอย่างผลการสังเคราะห์กรอบแอมพลิจูดโดยใช้วิธีคุณค่าอัตราส่วนแอมพลิจูดของกรอบแอมพลิจูดของเสียงวรรณยุกต์โทที่เก็บในฐานข้อมูลดังรูปที่ 74 เมื่อทำการสังเคราะห์เสียงพยางค์ "พาก" จะได้เสียงสังเคราะห์ที่มีกรอบแอมพลิจูดดังแสดงในรูป 75



รูปที่ 74 กรอบแอมพลิจูดของเสียงวรรณยุกต์โทในฐานข้อมูล



รูปที่ 75 ผลการสังเคราะห์กรอบแอมพลิจูดของเสียงวรรณยุกต์โท

เมื่อทำการเปรียบเทียบลักษณะกรอบแอมพลิจูดที่สังเคราะห์ขึ้นพบว่า มีลักษณะใกล้เคียงกับกรอบแอมพลิจูดต้นแบบที่อยู่ในฐานข้อมูล

## บทที่ 6

### การประเมินผลการสังเคราะห์เสียง

การประเมินผลการสังเคราะห์เสียงทำเพื่อวัดคุณภาพเสียงที่สังเคราะห์ขึ้นโดยการประเมินที่จะทำต่อไปนี้ทำเพื่อวัดความสามารถในการรับรู้ (Intelligibility) และความเป็นธรรมชาติ (Naturalness) ของเสียงที่สังเคราะห์ นอกจากนี้ทำเพื่อใช้วิเคราะห์และปรับปรุงหน่วยเสียงที่จะนำมาใช้สังเคราะห์ ในการประเมินผลนี้มุ่งไปที่คุณภาพของระดับพยางค์เนื่องจากวิธีการที่นำเสนอนี้ตั้งอยู่บนฐานเสียงสังเคราะห์ระดับพยางค์โดยวิธีการวิเคราะห์ที่ใช้แบ่งเป็น 2 แบบคือ เอ็มอาร์ที และ เอ็มโอเอส โดยมีรายละเอียดดังต่อไปนี้

#### 6.1 เอ็มอาร์ที (Modified Rhyme Test : MRT)

เอ็มอาร์ที (Ince, 1992; Kleijn et al., 1995) เป็นวิธีที่ใช้วัดความสามารถในการรับรู้และแยกแยะเสียงที่สังเคราะห์ขึ้น กลุ่มคำหรือพยางค์ที่ใช้ในการทดสอบเป็นคำหรือพยางค์ที่มีเสียงคล้ายกัน (Rhyme) เสียงที่ใช้ทดสอบอยู่บนฐานเสียงระดับพยางค์หรือคำที่เนื่องจากการทดสอบระดับพยางค์หรือคำนี้มีความไวต่อความผิดพลาดที่เกิดขึ้นในเสียงสังเคราะห์ได้ง่ายทำให้สามารถนำไปใช้วิเคราะห์หาข้อผิดพลาดในการสังเคราะห์ได้ง่ายและเสียงที่ใช้ไม่จำเป็นต้องมีความหมายซึ่งจะทำให้ไวต่อข้อผิดพลาดได้มากขึ้นเนื่องจากไม่สามารถใช้การคาดเดาได้ ข้อดีอีกประการหนึ่งคือวิธีการนี้อาสาสมัครที่ทำแบบทดสอบไม่จำเป็นต้องได้รับการฝึกมาก่อน สำหรับวิธีการทดสอบใช้อาสาสมัครฟังชุดเสียงสังเคราะห์ที่ละพยางค์โดยไม่ทราบว่าเป็นเสียงที่ได้ยินคือเสียงอะไรแล้วให้เลือกตัวเลือกว่าเสียงที่ได้ยินเป็นเสียงใด โดยทั่วไปมี 6 ตัวเลือก (Ince, 1992) ตัวเลือกแต่ละตัวจะมีเสียงที่แตกต่างกันเพียงจุดเดียวและครอบคลุมเสียงที่จะเกิดขึ้นได้ทั้งหมด การทดสอบนี้โดยทั่วไปมุ่งเน้นไปที่เสียงพยัญชนะเนื่องจากเสียงพยัญชนะเป็นเสียงที่สังเคราะห์ได้ยากกว่าเสียงสระ ตัวอย่างพยางค์ที่ใช้ทดสอบดังเช่น กาก (kaak1) ปาก (paak1) ตาก (taak1) จาก (caak1) เป็นต้น ซึ่งจะเห็นได้ว่าเสียงแต่ละเสียงที่ใช้ทดสอบต่างกันเพียงแค่เสียงพยัญชนะต้นเพียงเสียงเดียว

การทดสอบนี้เหมาะสมที่จะใช้ทดสอบการสังเคราะห์ที่นำเสนอเนื่องจากการเป็นการสังเคราะห์ฟอร์แมนต์และการบิดเบนฟอร์แมนต์ซึ่งเป็นส่วนที่มีผลต่อการรับรู้เสียงพยัญชนะและสระผสมโดยตรง

## 6.2 เอ็มโอเอส (Mean Opinion Score : MOS)

วิธีนี้ใช้วัดคุณภาพเสียงโดยรวมเพื่อวัดความเป็นธรรมชาติของเสียง การทดสอบโดยให้อาสาสมัครฟังเสียงทีละพยางค์ซึ่งแต่ละพยางค์อาสาสมัครจะทราบล่วงหน้าว่าเป็นเสียงอะไร เมื่อฟังเสียงแต่ละพยางค์แล้วให้อาสาสมัครระบุความประทับใจต่อเสียงที่ได้ยิน โดยแบ่งระดับความประทับใจเป็น 5 ระดับคือ ดีมาก (Excellent) ดี (Good) ปานกลาง (Fair) แย่ (Poor) และแย่มาก (Bad) โดยแต่ละระดับมีน้ำหนักคะแนนเป็น 5 4 3 2 และ 1 ตามลำดับแล้วนำคะแนนที่ได้มาเฉลี่ยซึ่งคะแนนดังกล่าวเรียกว่า MOS

## 6.3 ข้อสมมุติฐานในการปรับปรุงหน่วยเสียง

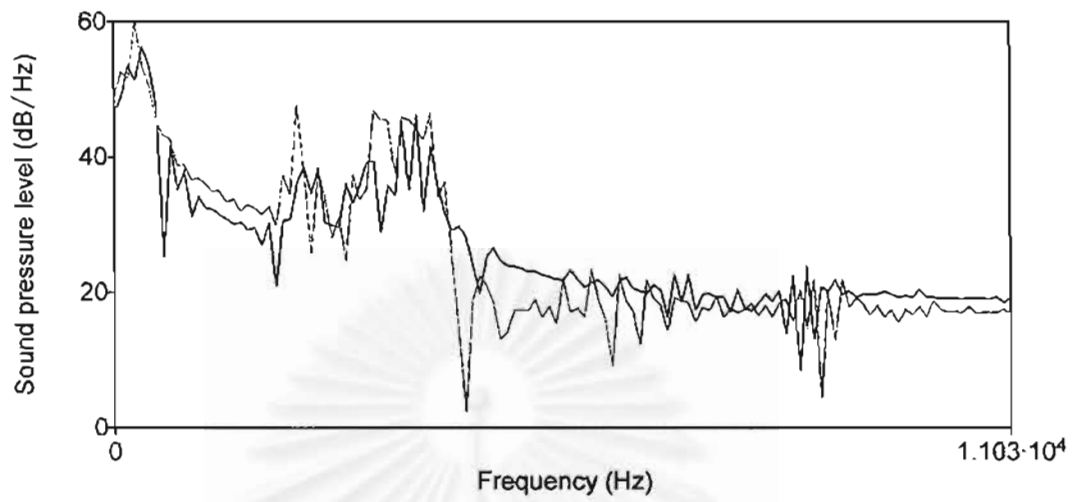
หน่วยเสียงที่จะทำการปรับปรุงมี 2 กลุ่มคือ

1. เสียงสระ เนื่องจากมีข้อสมมุติฐานว่าเสียงสระสำหรับพยางค์ที่มีพยัญชนะนาสิกเป็นพยัญชนะต้น เสียงสระที่ใช้ควรมีลักษณะเสียงนาสิกด้วย นอกจากนี้เมื่อพิจารณาสเปกตรัมพบว่าฟอร์แมนต์ของเสียงสระในพยางค์ที่มีพยัญชนะต้นเป็นเสียงนาสิกจะมีแบนด์วิดท์กว้างกว่าดังแสดงการเปรียบเทียบสเปกตรัมของสระทั้งสองกลุ่มในรูปที่ 76 ถึง 84 เนื่องจากพลังงานของเสียงนาสิกจะถูกหน่วงภายในช่องจมูก โดยผลการเปรียบเทียบแสดงเสียงสระกรณีที่มีเสียงนาสิกและไม่มีเสียงนาสิกด้วยเส้นกราฟสีแดงข้างล่าง

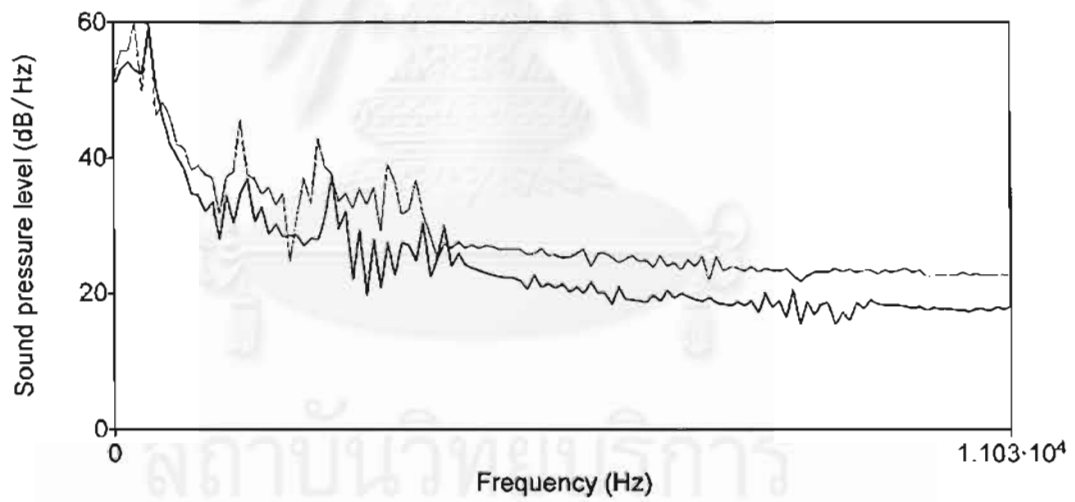
แสดงเสียงสระที่ไม่มีเสียงนาสิก

— แสดงเสียงสระที่มีเสียงนาสิก

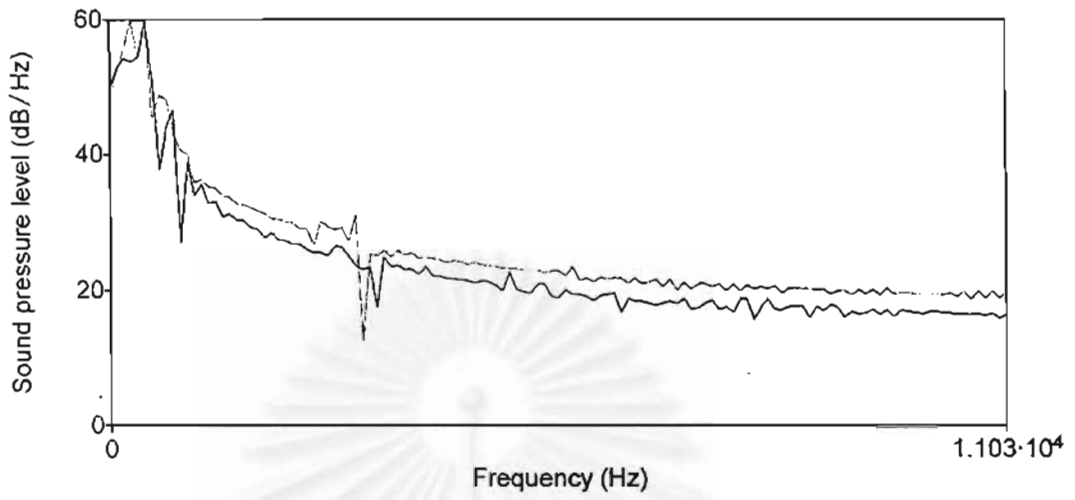




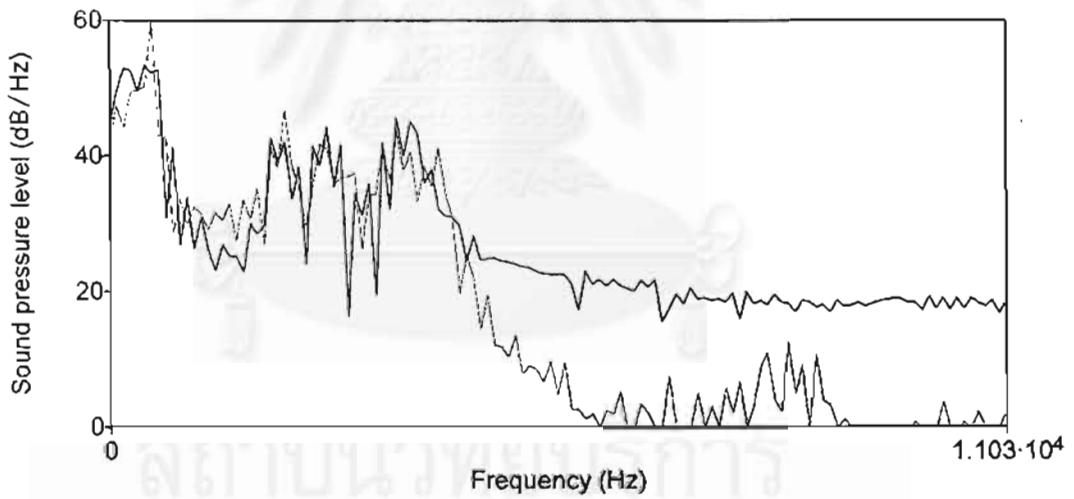
รูปที่ 76 การเปรียบเทียบสเปกตรัมของสระ "อ" ในกรณีที่มีเสียงนาสิกและไม่มีเสียงนาสิก



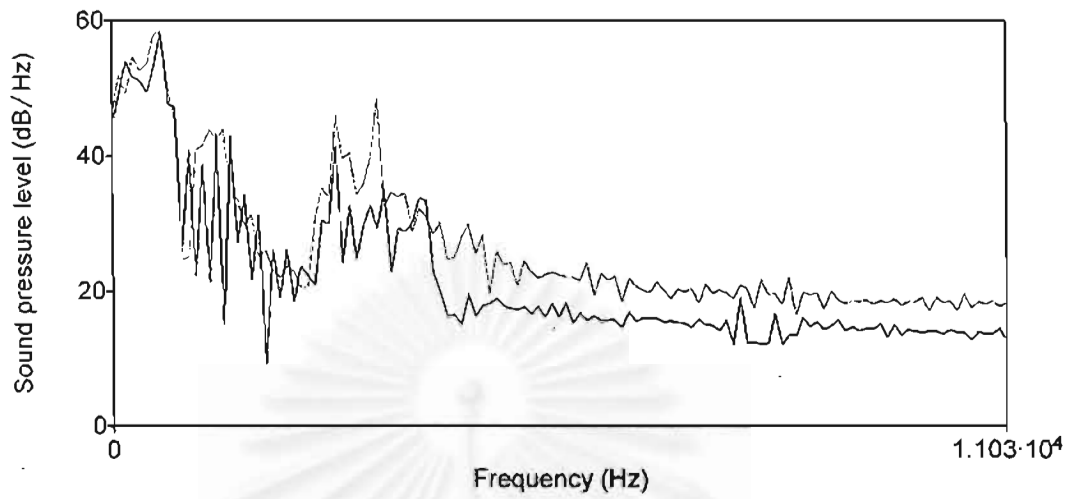
รูปที่ 77 การเปรียบเทียบสเปกตรัมของสระ "อ้อ" ในกรณีที่มีเสียงนาสิกและไม่มีเสียงนาสิก



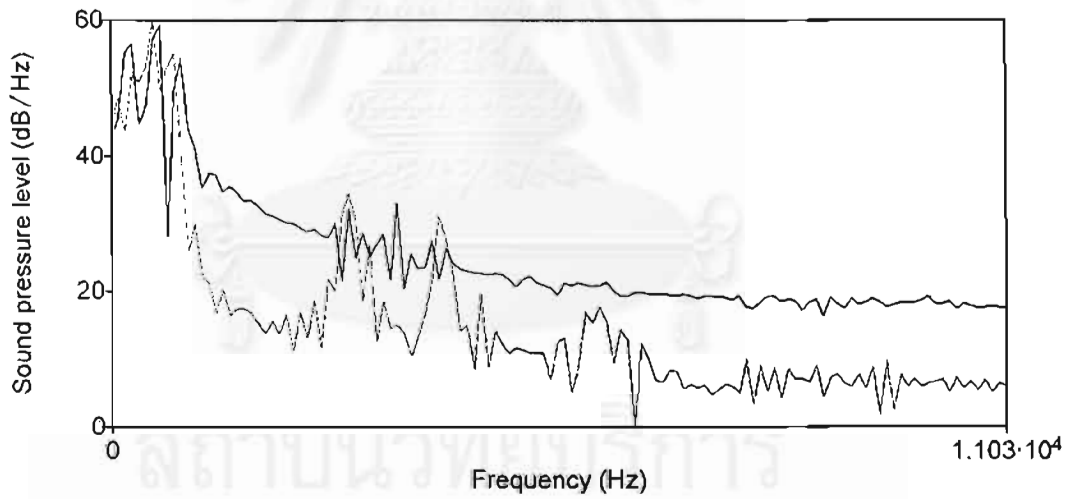
รูปที่ 78 การเปรียบเทียบสเปกตรัมของสระ "อู" ในกรณีที่มีเสียงนาสิกและไม่มีเสียงนาสิก



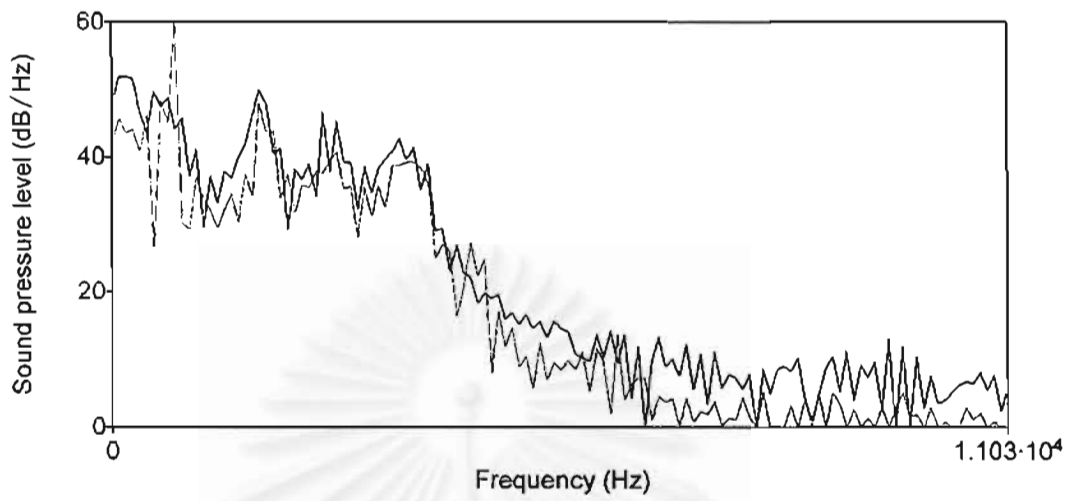
รูปที่ 79 การเปรียบเทียบสเปกตรัมของสระ "เอ" ในกรณีที่มีเสียงนาสิกและไม่มีเสียงนาสิก



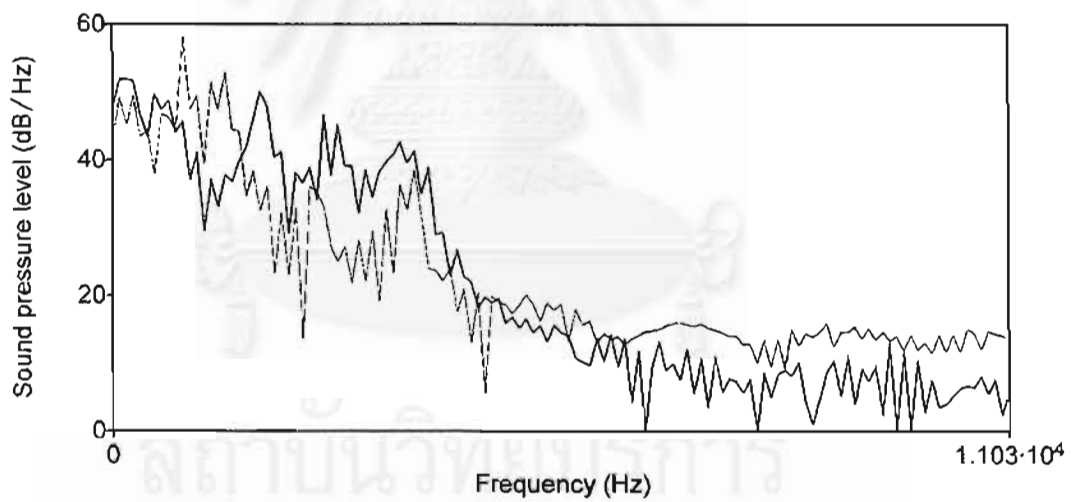
รูปที่ 80 การเปรียบเทียบสเปกตรัมของสระ "อ" ในกรณีที่มีเสียงนาสิกและไม่มีเสียงนาสิก



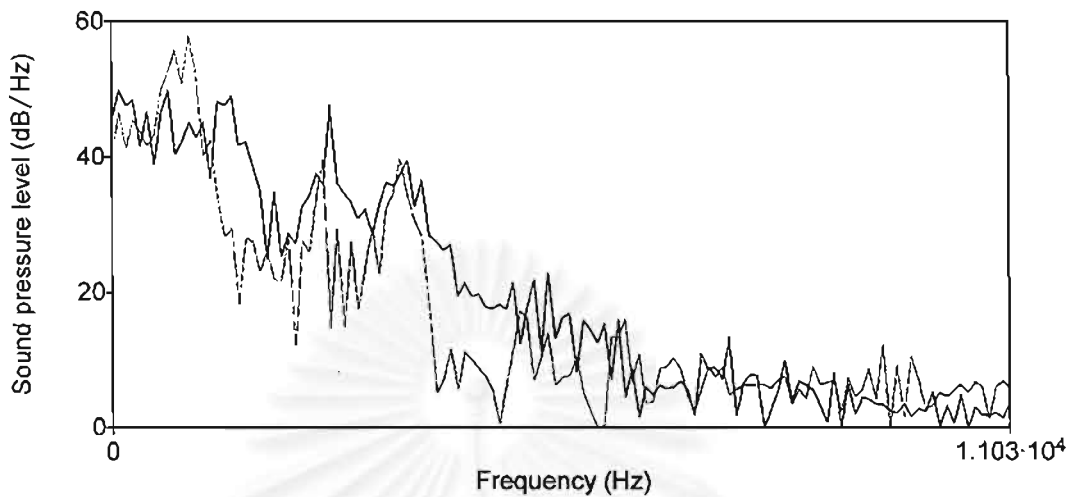
รูปที่ 81 การเปรียบเทียบสเปกตรัมของสระ "อ" ในกรณีที่มีเสียงนาสิกและไม่มีเสียงนาสิก



รูปที่ 82 การเปรียบเทียบสเปกตรัมของสระ "เอ" ในกรณีที่มีเสียงนาสิกและไม่มีเสียงนาสิก



รูปที่ 83 การเปรียบเทียบสเปกตรัมของสระ "อา" ในกรณีที่มีเสียงนาสิกและไม่มีเสียงนาสิก



รูปที่ 84 การเปรียบเทียบสเปกตรัมของสระ "อ" ในกรณีที่มีเสียงนาสิกและไม่มีเสียงนาสิก

2. เสียงพยัญชนะสะกดกลุ่มอัมสระ ในทางทฤษฎีของเสียงพยัญชนะสะกดที่กล่าวในบทที่ 2 ว่าเสียงพยัญชนะสะกดอัมสระ "ย" เกิดจากการเสียงสระตามด้วยเสียงสระ "อิ" และเสียงพยัญชนะสะกดอัมสระ "ว" เกิดจากการเสียงสระตามด้วยเสียงสระ "อุ" จึงตั้งสมมุติฐานว่าถ้านำหน่วยเสียงพยัญชนะสะกดกลุ่มอัมสระที่ได้จากเสียงจริงแทนการใช้เสียงสระ "อิ" และ "อุ" ควรจะให้ผลการสังเคราะห์ที่มีคุณภาพสูงกว่า

#### 6.4 การประเมินผล

การประเมินใช้อาสาสมัครในการทำแบบประเมินจำนวน 10 คน โดยใช้แบบประเมินในภาคผนวก ก. การเลือกพยางค์ที่นำมาทดสอบมีกฎดังนี้

1. เสียงพยัญชนะ ได้จากการสุ่มพยัญชนะในแต่ละตำแหน่งการออกเสียงเนื่องจากแต่ละตำแหน่งการออกเสียงมีการบิดเบนฟอร์แมนต์ที่แตกต่างกัน โดยสุ่มตัวอย่างพยางค์ตำแหน่งการออกเสียงละ 5 พยางค์
2. เสียงสระ เนื่องจากเสียงสระสั้นและเสียงสระยาวมีหน่วยเสียงเดียวกัน จึงเลือกใช้สังเคราะห์เฉพาะเสียงสระยาวเท่านั้น
3. เสียงวรรณยุกต์ ทดสอบด้วยการนำเสียงสระเท่านั้นมาทำการผันเสียงวรรณยุกต์ทั้ง 5 เสียง

สำหรับตัวเลือกของการทดสอบเอ็มอาร์ที การคัดเลือกตัวเลือกในแต่ละข้อมีกฎในการเลือกดังนี้

1. ตัวเลือกทุกตัวจะแตกต่างจากเสียงที่ถูกทดสอบเฉพาะคุณสมบัติที่ต้องการทดสอบ
2. ในกรณีทดสอบเสียงพยัญชนะ ตัวเลือกจะได้จากเสียงพยัญชนะที่มีคุณสมบัติต่างจากเสียงที่ถูกทดสอบเพียงคุณสมบัติเดียวคือต้องมีตำแหน่งการออกเสียงเหมือนกันหรือมีลักษณะการออกเสียงเหมือนกันเนื่องจากเสียงที่มีคุณสมบัติแตกต่างกันเพียงหนึ่งคุณสมบัติจะมีเสียงที่ใกล้เคียงกันซึ่งเป็นตัวเลือกที่ดีสำหรับการประเมินผล ยกเว้นในกรณีเสียงพยัญชนะที่มีคุณสมบัติแตกต่างกันหนึ่งคุณสมบัติถูกเลือกหมดแล้ว จึงสามารถสุ่มเลือกเสียงพยัญชนะอื่นๆ
3. ในกรณีทดสอบเสียงสระ ตัวเลือกมีหลักในการเลือกเช่นเดียวกับเสียงพยัญชนะคือเลือกเสียงสระที่มีตำแหน่งลิ้นที่ใกล้เคียงกับเสียงที่ต้องการทดสอบ เนื่องจากลิ้นซึ่งควบคุมเสียงสระที่ออกมาดังนั้นเสียงที่มีตำแหน่งลิ้นใกล้เคียงกันก็มีลักษณะเสียงที่ใกล้เคียงกัน ซึ่งเป็นตัวเลือกที่ดีสำหรับการทดสอบโดยเฉพาะการทดสอบการแยกแยะเสียงที่ได้ยิน

ในแบบประเมินผลดังกล่าวแบ่งเป็นการประเมินระดับพยางค์ 102 พยางค์โดยประเมินทั้งเอ็มอาร์ทีและเอ็มไอเอส และในระดับวลีจำนวน 5 วลีโดยประเมินเฉพาะเอ็มไอเอส การประเมินเริ่มจากการประเมินเอ็มอาร์ทีเป็นการประเมินแรก แล้วทำการประเมินเอ็มไอเอสเป็นอันดับต่อมา ทั้งนี้เพื่อให้ได้ค่าความสามารถในการรับรู้และแยกแยะเสียงที่ถูกต้อง เนื่องจากถ้าทำการประเมินค่าเอ็มไอเอสก่อน อาสาสมัครจะทราบว่าเป็นเสียงแต่เสียงคืออะไรและที่คุณภาพระดับใดเมื่อทำการประเมินค่าเอ็มอาร์ที ผลที่ได้อาจเกิดการโน้มเอียงเนื่องจากเคยเรียนรู้เสียงนั้นมาแล้วล่วงหน้า

ในระดับพยางค์ใช้หน่วยเสียงที่ได้ตามสมมุติฐานในหัวข้อ 6.3 คือ ใช้หน่วยเสียงสระนาสิกสำหรับพยางค์ที่ขึ้นต้นด้วยเสียงพยัญชนะนาสิกและใช้หน่วยเสียงพยัญชนะสะกดกลุ่มอัมผัสสระที่ได้จากเสียงจริง แล้วทำการประเมินผลดังต่อไปนี้

1. ประเมินผลการสังเคราะห์เสียงพยัญชนะต่างๆ
2. ประเมินผลการสังเคราะห์เสียงสระ
3. ประเมินผลการสังเคราะห์เสียงวรรณยุกต์

นอกจากนี้ได้ทำการประเมินและเปรียบเทียบข้อสมมุติฐานในหัวข้อ6.4 เพื่อทดสอบสมมุติฐานดังกล่าวได้ทำการประเมินดังต่อไปนี้

1. เสียงสระสำหรับเสียงพยัญชนะต้นกลุ่มนาสิก โดยเปรียบเทียบผลการสังเคราะห์ระหว่างการนำเสียงพยางค์ที่ใช้สระที่ไม่มีเสียงนาสิกและเสียงสระที่มีลักษณะเสียงนาสิก
2. เสียงพยัญชนะสะกดกลุ่มอัมสระ โดยเปรียบเทียบผลการสังเคราะห์ระหว่างการนำหน่วยเสียงสระอีและสระอุมาสังเคราะห์เสียงพยัญชนะสะกดกลุ่มอัมสระกับการเก็บหน่วยเสียงที่ได้จากการเก็บจากเสียงพยัญชนะสะกดจริง

ในระดับวลีทำการประเมินว่าการสังเคราะห์เสียงระดับพยางค์ด้วยวิธีการที่น่าเสนอนี้ เมื่อนำมาต่อเป็นวลีจะมีคุณภาพเป็นอย่างไร



## บทที่ 7

### ผลการประเมินการสังเคราะห์เสียง

ในบทนี้นำเสนอผลการประเมินที่ได้จากการประเมินในบทที่ 6 ผลการประเมินที่ได้แสดงถึงคุณภาพของเสียงทั้งทางด้านความสามารถในการรับรูปลักษณะและเป็นธรรมชาติของเสียงของเสียงที่สังเคราะห์ขึ้น นอกจากนี้ยังแสดงผลการคัดเลือกหน่วยเสียงเพิ่มเติมเพื่อปรับปรุงเสียงให้มีคุณภาพดีขึ้น โดยผลการประเมินมีดังต่อไปนี้

#### 7.1 ผลการประเมินเสียงพยัญชนะในระดับพยางค์

ผลการประเมินเสียงพยัญชนะต้น (ค่าเฉลี่ย)		
ตำแหน่งการออกเสียง	เอ็มอาร์ที (%)	เอ็มไอเอส
ริมฝีปาก	68	3.8
ปุ่มเหงือก	78	3.4
เพดานบน	84	3.5
เพดานอ่อน	92	3.3
เส้นเสียง	78	3.4
รวม	80	3.5

#### ตารางที่ 6 ผลการประเมินเสียงพยัญชนะต้น

ผลการประเมินเสียงพยัญชนะควบกล้ำ (ค่าเฉลี่ย)		
ตำแหน่งการออกเสียง	เอ็มอาร์ที (%)	เอ็มไอเอส
ริมฝีปาก	74	3.9
ปุ่มเหงือก	46	3.9
เพดานอ่อน	66	3.9
รวม	62	3.9

#### ตารางที่ 7 ผลการประเมินเสียงพยัญชนะควบกล้ำ



ผลการประเมินเสียงพยานะสะกด (ค่าเฉลี่ย)		
ตำแหน่งการออกเสียง	เอ็มอาร์ที (%)	เอ็มไอเอส
ริมฝีปาก	92	4.6
ปุ่มเหงือก	42	3.6
เพดานบน	74	3.4
เพดานอ่อน	56	3.9
รวม	66	3.9

ตารางที่ 8 ผลการประเมินเสียงพยานะสะกด

สรุปผลการประเมินเสียงพยานะ (ค่าเฉลี่ย)		
ประเภท	เอ็มอาร์ที (%)	เอ็มไอเอส
เสียงพยานะต้น	80	3.5
เสียงพยานะควบกล้ำ	62	3.9
เสียงพยานะสะกด	66	3.9
รวม	71	3.7

ตารางที่ 9 สรุปผลการประเมินเสียงพยานะ

## 7.2 ผลการประเมินเสียงสระในระดับพยางค์

สรุปผลการประเมินเสียงสระ (ค่าเฉลี่ย)		
ประเภทสระ	เอ็มอาร์ที (%)	เอ็มไอเอส
สระเดี่ยว	100	4.7
สระผสม	100	3.8
รวม	100	4.5

ตารางที่ 10 สรุปผลการประเมินเสียงสระ

### 7.3 ผลการประเมินเสียงวรรณยุกต์ในระดับพยางค์

สรุปผลการประเมินเสียงวรรณยุกต์						
ประเภทสระ	เอ็มอาร์ที (%)			เอ็มไอเอส		
	Mean	Max	Min	Mean	Max	Min
เสียงสระ (ปกติ)	98	100	90	5	5	4.9
เสียงสระ (นาสิก)	96	100	90	4.7	4.9	4.5
รวม	97	100	90	4.8	5	4.5

ตารางที่ 11 สรุปผลการประเมินเสียงวรรณยุกต์

### 7.4 ผลการประเมินในระดับพยางค์โดยรวม

ผลการประเมินในระดับพยางค์โดยรวม (เฉลี่ย)	
เอ็มอาร์ที (%)	เอ็มไอเอส
78	3.98

ตารางที่ 12 สรุปผลการประเมินเสียงสังเคราะห์ระดับพยางค์โดยรวม

### 7.5 ผลการประเมินระดับวลี

ค่าเอ็มไอเอสของการประเมินระดับวลีเท่ากับ 2.4

### 7.6 ผลการประเมินการปรับปรุงหน่วยเสียง

ผลการประเมินการปรับปรุงหน่วยเสียงสระสำหรับเสียงพยัญชนะต้นนาสิก (ค่าเฉลี่ย)		
ประเภท	เอ็มอาร์ที (%)	เอ็มไอเอส
เสียงสระ (ปกติ)	76	3.36
เสียงสระ (นาสิก)	80	3.98

ตารางที่ 13 ผลการประเมินการปรับปรุงหน่วยเสียงสระสำหรับเสียงพยัญชนะต้นนาสิก

ผลการประเมินการปรับปรุงหน่วยเสียงพยัญชนะสะกดกลุ่มอัมสระ (ค่าเฉลี่ย)		
ประเภท	เอ็มอาร์ที (%)	เอ็มไอเอส
ใช้เสียงสระ "อิ" และ "อุ"	92	2.94
ใช้เสียงจริง	68	3.68

ตารางที่ 14 ผลการประเมินการปรับปรุงหน่วยเสียงพยัญชนะสะกดกลุ่มอัมสระ

## 7.7 วิเคราะห์ผลการประเมิน

1. เสียงพยัญชนะ ผลการประเมินเสียงพยัญชนะที่สังเคราะห์ขึ้นโดยรวมมีความสามารถในการรับรู้และแยกแยะเสียงพยัญชนะค่อนข้างดีคือได้ค่าเอ็มอาร์ทีเท่ากับ 71 % โดยเฉพาะเสียงพยัญชนะต้น ในขณะที่เสียงพยัญชนะควบกล้ำได้ค่าเอ็มอาร์ทีต่ำที่สุดทั้งนี้อาจเนื่องมาจากผลของเสียงส่วนต้นของหน่วยเสียงกลุ่มนี้มีความซับซ้อนของสเปกตรัมมากกว่าการพิจารณาตำแหน่งการตัดหน่วยเสียงที่เหมาะสมจึงยากกว่าการตัดหน่วยเสียงในกลุ่มอื่นและระยะเวลาในช่วงนี้โดยทั่วไปยาวกว่าเสียงพยัญชนะในกลุ่มอื่นๆ ทำให้คุณภาพของเสียงส่วนต้นของหน่วยเสียงกลุ่มนี้มีผลมากต่อคุณภาพเสียงโดยรวม เมื่อพิจารณาความเป็นธรรมชาติของเสียงจากค่าเอ็มไอเอสพบว่ามีความในระดับค่อนข้างดีคือที่ค่าโดยเฉลี่ยเท่ากับ 3.7 ดังนั้นการสังเคราะห์ตามวิทยานิพนธ์นี้สามารถสังเคราะห์เสียงพยัญชนะได้ค่อนข้างดี ทั้งนี้ขึ้นอยู่กับการพิจารณาตำแหน่งตัดหน่วยเสียงที่เหมาะสมด้วย

2. เสียงสระ ผลการประเมินเสียงสระแบ่งเป็น 2 ส่วนคือ ส่วนเสียงสระเดี่ยว ในส่วนนี้ผลการทดสอบจะชี้ถึงคุณภาพของหน่วยเสียงที่นำมาใช้ในการสร้างเสียงสระทั้งหมดคือ สระเสียงสั้น สระเสียงยาว และสระผสม โดยผลการประเมินแสดงถึงคุณภาพของหน่วยเสียงที่สูง เนื่องจากได้ค่าเอ็มอาร์ทีเท่ากับ 100 % แสดงว่าความสามารถในการรับรู้และแยกแยะเสียงสระแต่ละเสียงมีค่าสูง และค่าเอ็มไอเอสมีค่าอยู่ในระดับเกือบดีมาก แสดงว่ามีความเป็นธรรมชาติของเสียงสูง ในอีกส่วนหนึ่งคือส่วนเสียงสระผสม ซึ่งเกิดจากการเชื่อมต่อของเสียงสระเดี่ยว ผลทางด้านกรรับรู้และแยกแยะเสียงมีค่าสูงเช่นเดียวกับส่วนเสียงสระเดี่ยวคือมีค่าเอ็มอาร์ทีเท่ากับ 100 % สำหรับค่าเอ็มไอเอสมีค่าเท่ากับ 3.8 เนื่องจากผลประเมินของเสียงสระเดี่ยวมีคุณภาพสูง แต่ความเป็นธรรมชาติของเสียงสระผสมมีค่าลดลงทั้งนี้อาจเนื่องมาจากผลของการเชื่อมต่อฟอร์แมนต์และระยะเวลาการสังเคราะห์เสียงในแต่ละส่วนของสระผสม แต่ผลความเป็นธรรมชาติที่ได้อยู่ใน

ระดับที่ค่อนข้างดี โดยสรุปจากผลประเมินโดยเฉลี่ยสรุปได้ว่าการสังเคราะห์เสียงตามวิทยานิพนธ์นี้สามารถสังเคราะห์เสียงสระที่มีคุณภาพในระดับสูงได้

3. เสียงวรรณยุกต์ ผลการสังเคราะห์เสียงวรรณยุกต์โดยใช้หน่วยเสียงที่นำเสนอในวิทยานิพนธ์และใช้วิธีพีเอสโอแอลเอ ผลประเมินทั้งในกรณีที่ใช้หน่วยเสียงสระปกติที่ไม่มีเสียงนาสิกสำหรับการสังเคราะห์เสียงพยางค์ทั่วไป และหน่วยเสียงสระที่มีลักษณะเสียงนาสิกสำหรับการสังเคราะห์เสียงพยางค์ที่ขึ้นต้นด้วยเสียงพยัญชนะนาสิก แสดงผลการสังเคราะห์ที่มีคุณภาพสูง เนื่องจากได้ค่าเอ็มอาร์ทีเฉลี่ยสูงและค่าค่าเอ็มโอเอสสูงถึง 97 % และ 4.8 ตามลำดับ

4. ระดับวลี ผลประเมินด้านนี้จะนำไปสู่การสังเคราะห์เสียงในระดับประโยคหรือวลี ซึ่งผลประเมินคุณภาพที่ได้อยู่ในระดับต่ำ คือมีค่าเอ็มโอเอสเท่ากับ 2.4 ทั้งนี้เนื่องจากคุณสมบัติของเสียงพยางค์ใดๆ จะแปรเปลี่ยนตามเสียงรอบๆ ถึงแม้ว่าเป็นการออกเสียงพยางค์เดียวกัน ดังนั้นเสียงพยางค์เดียวกันเมื่อฟังในระดับพยางค์ ผลประเมินคุณภาพเสียงอาจจะสูง แต่เมื่อนำมาสังเคราะห์ระดับพยางค์ ผลประเมินคุณภาพเสียงที่ได้จะลดลง โดยตัวแปรที่มีผลต่อเสียงในเสียงพูดต่อเนื่องคือทำนองการออกเสียง (Prosody) ประกอบด้วยตัวแปรด้านระยะเวลา รูปแบบความถี่มูลฐาน พลังงาน และคุณภาพเชิงความถี่ของเสียง ดังนั้นการสังเคราะห์เสียงพูดต่อเนื่องจึงต้องทำการปรับตัวแปรกลุ่มนี้เพื่อให้ได้เสียงสังเคราะห์ที่มีคุณภาพสูง

5. การเปรียบเทียบผลการปรับปรุงหน่วยเสียงสระสำหรับพยางค์ที่ขึ้นต้นด้วยเสียงพยัญชนะนาสิก สังเกตได้ว่าการใช้หน่วยเสียงสระที่มีลักษณะนาสิกมาสังเคราะห์เสียงพยางค์ที่ขึ้นต้นด้วยเสียงพยัญชนะต้นนาสิก ให้ผลการประเมินที่สูงกว่าทั้งทางด้านกรับรู้และแยกแยะเสียงและความเป็นธรรมชาติของเสียง ดังนั้นสรุปได้ว่าสมมติฐานที่เลือกใช้หน่วยเสียงที่มีลักษณะนาสิกช่วยทำให้คุณภาพการสังเคราะห์เสียงกลุ่มนี้สูงขึ้น

6. การเปรียบเทียบผลการปรับปรุงหน่วยเสียงพยัญชนะสะกดกลุ่มอัมสระ ผลที่ได้มีลักษณะที่ขัดกันคือการใช้หน่วยเสียงสระในการสังเคราะห์ให้ค่าเอ็มอาร์ทีสูงกว่าการใช้หน่วยเสียงจากเสียงจริง ในขณะที่ค่าเอ็มโอเอสของวิธีที่ใช้หน่วยเสียงจากเสียงจริงให้ค่าที่สูงกว่าการใช้หน่วยเสียงสระ ซึ่งหน่วยเสียงที่ได้เสียงจริงควรมีคุณภาพโดยรวมสูงกว่าการใช้หน่วยเสียงสระ ทั้งนี้เนื่องจากการใช้หน่วยเสียงสระทำให้เสียงสังเคราะห์มีความชัดเจนสูงทำให้คะแนนเอ็มอาร์ทีมีค่าสูงแต่ความชัดเจนสูงของเสียงสังเคราะห์นั้นทำให้เกิดความไม่เป็นธรรมชาติของเสียง ผลการประเมินเอ็มโอเอสจึงต่ำกว่าการใช้หน่วยเสียงจากเสียงอัมสระจริง

สรุปผลประเมินการสังเคราะห์ระดับพยางค์ตามวิธีในวิทยานิพนธ์นี้ให้ผลการสังเคราะห์โดยรวมในระดับค่อนข้างดีคือมีค่าเอ็มอาร์ทีเฉลี่ยเท่ากับ 78 % และค่าเอ็มโอเอสเฉลี่ยเท่ากับ 3.98

แต่ยังมีคุณภาพของเสียงสังเคราะห์บางกลุ่มที่ต้องทำการปรับปรุงเพื่อให้เสียงพยางค์สังเคราะห์โดยรวมมีคุณภาพที่ดีขึ้น สำหรับการสังเคราะห์ระดับประโยคหรือวลีซึ่งเป็นประโยคต่อเนื่องเมื่อสังเคราะห์โดยมีตัวแปรทางด้านทำนองเสียงคงที่โดยไม่ขึ้นกับเสียงรอบๆ เสียงสังเคราะห์ที่ได้จะมีคุณภาพต่ำดังนั้นต้องอาศัยการปรับตัวแปรของทำนองเสียงสำหรับการสังเคราะห์เสียงพูดต่อเนื่อง



บทที่ 8 (หน้า 89 – 94 ) ไม่มีในต้นฉบับที่นำมาสแกน

สถาบันวิทยบริการ  
จุฬาลงกรณ์มหาวิทยาลัย

## บทที่ 9 สรุปงานวิจัยและข้อเสนอแนะ

### 9.1 สรุปงานวิจัย

งานวิจัยทางการสังเคราะห์เสียงพูดของมนุษย์ที่นำเสนอในวิทยานิพนธ์ฉบับนี้เป็นวิธีการสังเคราะห์เสียงภาษาไทยในระดับพยางค์วิธีใหม่อยู่บนพื้นฐานการสังเคราะห์แบบวิเคราะห์ (Synthesis by Analysis) โดยใช้หน่วยเสียงที่มีจำนวนน้อยและขนาดเล็กแต่สามารถสังเคราะห์เสียงพยางค์ได้ทุกพยางค์ในภาษาไทยโดยอาศัยโครงสร้างพยางค์สังเคราะห์ที่แบ่งพยางค์ออกเป็นหน่วยย่อยตามลักษณะและหน้าที่ของเสียง โดยแต่ละหน่วยเสียงที่ประกอบเป็นพยางค์สังเคราะห์นี้จะเก็บจุดอ้างอิงเชิงความถี่ของเสียงแต่ละเสียงในภาษาไทย โดยเก็บในรูปแบบของคู่เส้นสเปกตรัมและใช้คุณสมบัติของคู่เส้นสเปกตรัมนี้ในการสร้างฟอร์แมนต์ของสัญญาณเสียงพูด ซึ่งเสียงสังเคราะห์ที่ได้มีความต่อเนื่องทั้งในเชิงเวลาและมีความต่อเนื่องของฟอร์แมนต์ในเชิงความถี่

การสังเคราะห์เสียงใช้คุณสมบัติในการประมาณค่าพารามิเตอร์ของคู่เส้นสเปกตรัมในการสร้างฟอร์แมนต์และเชื่อมต่อจุดอ้างอิงระหว่างหน่วยเสียงแล้วทำการแปลงกลับเป็นสัญญาณเสียงพูดช่วงสั้นโดยใช้วงจรกรองสังเคราะห์การทำนายแบบเชิงเส้น สัญญาณเสียงพูดช่วงสั้นดังกล่าวจะถูกปรับขนาดแอมพลิจูดแล้วปรับความถี่มูลฐานด้วยวิธีโซลาเพื่อให้ได้สัญญาณเสียงที่มีเสียงวรรณยุกต์ตามที่ต้องการออกมา

การสังเคราะห์นี้ใช้ฐานข้อมูลซึ่งประกอบไปด้วย หน่วยเสียงที่เก็บในรูปแบบคู่เส้นสเปกตรัมจำนวน 376 หน่วยเสียง รูปแบบความถี่มูลฐานจำนวน 5 รูปแบบ กรอบแอมพลิจูดจำนวน 5 รูปแบบ และระยะเวลาสำหรับการสังเคราะห์ ซึ่งขนาดฐานข้อมูลทั้งหมดมีขนาดประมาณเพียง 1.5 เมกกะไบต์สำหรับการสังเคราะห์เสียงผู้ชายที่มีคุณภาพเสียงที่อัตราสุ่ม 22,050 เฮอร์ทและความละเอียดสัญญาณ 16 บิตต่อหนึ่งตัวอย่าง โดยข้อมูลในฐานข้อมูลเหล่านี้ได้จากการวิเคราะห์สัญญาณเสียงพูดต้นแบบ

เสียงที่ได้จากการสังเคราะห์วิธีนี้ส่วนใหญ่มีความต่อเนื่องของฟอร์แมนต์เนื่องจากคุณสมบัติการประมาณค่าของคู่เส้นสเปกตรัมแต่ผลการสังเคราะห์บางเสียงอาจพบความไม่ต่อเนื่องเนื่องจากคุณภาพของหน่วยเสียงเนื่องจากการตัดหน่วยเสียงและคุณภาพของเสียงต้นแบบ สำหรับรูปแบบความถี่มูลฐานของเสียงวรรณยุกต์และกรอบแอมพลิจูดที่สังเคราะห์ขึ้นมีลักษณะเช่นเดียวกับต้นแบบในฐานข้อมูล

ในทำงงานวิจัยได้ประเมินคุณภาพเสียงที่สังเคราะห์ขึ้นได้ค่าเอ็มอาร์ทีเฉลี่ยเท่ากับ 78 % และได้ค่าเอ็มไอเอสเฉลี่ยเท่ากับ 3.98 โดยภาพรวมผลการสังเคราะห์ที่ได้อยู่ในระดับค่อนข้างดี

## 9.2 ข้อเสนอแนะ

1. เนื่องจากเสียงพูดในเชิงความถี่มีลักษณะการเปลี่ยนแปลงพลังงานลดลงเมื่อความถี่สูงขึ้น ดังนั้นฟอร์แมนต์ของเสียงที่มีความสูงกว่าจามีพลังงานต่ำกว่าฟอร์แมนต์ที่ความถี่ต่ำกว่า นอกจากนี้เสียงที่บันทึกต้องผ่านกระบวนการแปลงอนาลอกเป็นดิจิทัลซึ่งสัญญาณเสียงอนาลอกดังกล่าวต้องถูกนำไปผ่านวงจรกรองผ่านต่ำ (Lowpass Filter) ก่อนที่จะถูกสุ่มตัวอย่างเพื่อป้องกันการกลมกลืนเชิงสเปกตรัม (Spectral Aliasing) แต่ในขณะที่เดียวกันก็เป็นการลดพลังงานของฟอร์แมนต์ที่ความถี่สูง ซึ่งเป็นผลไม่ดีต่อการวิเคราะห์การทำนายเชิงเส้นเพราะอาจจะเป็นเหตุให้เมตริกอัตราสัมพันธ์มีความผิดพลาดได้และส่งผลกระทบต่อความถูกต้องในการคำนวณค่าสัมประสิทธิ์การทำนายแบบเชิงเส้นด้วย ในการแก้ปัญหาเรื่องนี้จึงต้องเสริมพลังงานที่ความถี่สูงของสัญญาณเสียงก่อนที่นำไปคำนวณสัมประสิทธิ์การทำนายแบบเชิงเส้น ซึ่งทำได้โดยนำสัญญาณเสียงผ่านวงจรกรองพรีเอมฟาสิส (Pre-Emphasis) (Choy, 1998) ซึ่งมีลักษณะตามสมการ (eq34)

$$H(z) = 1 - \alpha z^{-1} \quad (\text{eq34})$$

โดยที่  $\alpha$  เรียกว่า ตัวประกอบพรีเอมฟาสิส (Pre-emphasis Factor) เป็นตัวกำหนดความถี่คัท-ออฟ (Cut-off Frequency) ของวงจรกรองเดียวศูนย์โดยทั่วไปมี  $\alpha$  ค่าเท่ากับ 0.1 ซึ่งวิธีนี้จะทำให้พลังงานที่ความถี่สูงถูกยกกระดืบ และเมื่อต้องการทำสัญญาณกลับคืนสามารถทำได้โดยการทำดีเอมฟาสิส (De-emphasis) ซึ่งเป็นส่วนผกผันกับ  $H(z)$

2. เสียงที่สังเคราะห์ขึ้นพบว่า ในกรณีที่ความถี่มูลฐานมีค่าต่ำมากทำให้ระยะพิทช์สำหรับการทำไชลามีระยะห่างมาก พบว่าเสียงสระบางตัวมีลักษณะบ๊ซซี่ (Buzzy) เนื่องจากหน้าต่างที่มีความกว้างแคบใกล้เคียงหรือน้อยกว่าระยะพิทช์ทำให้สัญญาณบริเวณขอบมีลักษณะสปี เมื่อทำส่งผลให้บริเวณที่ทำการทับซ้อนมีแอมพลิจูดต่ำหรือเกิดสัญญาณขาดช่วงเนื่องจากมีค่าใกล้เคียงศูนย์เพื่อปรับปรุงคุณภาพเสียงดังกล่าวสามารถแก้ได้โดยการทำหน้าต่างที่มีความกว้างที่กว้างเท่ากับสองเท่าของคาบพิทช์ของสัญญาณ



3. การปรับปรุงคุณภาพเสียงในขั้นตอนโซลาอีกวิธีหนึ่งคือ แอลพี-โซลา (Linear Predictive PSOLA : LP-PSOLA) (Charpentier and Moulines, 1989) เป็นการทำให้โซลาบนสัญญาณตกค้างแทนการทำบนสัญญาณในเชิงเวลา วิธีนี้จะดีกว่าทีดี-โซลา คือการลดทอนสเปกตรัมต่ำกว่าเนื่องจากสัญญาณตกค้างมีลักษณะสเปกตรัมแบบราบ และสามารถแก้ปัญหาในกรณีที่คาบพิทช์กว้างเกินไปเนื่องจากเป็นการปรับพิทช์บนสัญญาณตกค้างก่อนผ่านวงจรกรองสักราะห้การทำนายแบบเชิงเส้น ดังนั้นในบริเวณรอยต่อของหน้าต่างจะถูกประมาณค่าโดยวงจรกรองสักราะห้การทำนายแบบเชิงเส้นทำให้สัญญาณเสียงที่ได้มีความต่อเนื่องโดยไม่เกิดปัญหาสัญญาณขาดช่วงดังเช่นวิธีทีดี-โซลา



## รายการอ้างอิง

### ภาษาไทย

- กฤษฎดา เรยเสธ. 2530. ต้นแบบเครื่องสังเคราะห์เสียงพูดด้วยวิธีเข้ารหัสแบบลิเนียร์พรีดิคทีฟ. วิทยานิพนธ์ปริญญาโทมหาบัณฑิต ภาควิชาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.
- ยี่น ภู่วรรณ. 2527. การสังเคราะห์เสียงพูดภาษาไทย. วารสารวิศวกรรมศาสตร์. คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเกษตรศาสตร์.
- สุดาพร ลักษณะนิยานาวิน. 2534. ปัญหาของระบบการสังเคราะห์เสียงจากข้อความภาษาไทย. การประชุมทางวิชาการโครงการวิจัยและพัฒนาอิเล็กทรอนิกส์และคอมพิวเตอร์ (เล่มที่ 1). ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ กระทรวงวิทยาศาสตร์เทคโนโลยี และการพลังงาน.
- สุดาพร ลักษณะนิยานาวิน. 2535. คอมพิวเตอร์อ่านออกเสียงไทย. การประชุมทางวิชาการครั้งที่ 4. ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ กระทรวงวิทยาศาสตร์เทคโนโลยี และการพลังงาน. หน้า 65-78.
- อาทร นันทิยะกุล. 2533. การสังเคราะห์เสียงพูดจากข้อความภาษาไทย. วิทยานิพนธ์ปริญญาโทมหาบัณฑิต ภาควิชาวิศวกรรมคอมพิวเตอร์ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.
- เอกพล อนุสุเรนทร์. 2541. การสังเคราะห์เสียงวรรณยุกต์และสระในพยางค์เปิดภาษาไทยโดยใช้หน่วยเสียงอนุภาค. วิทยานิพนธ์ปริญญาโทมหาบัณฑิต ภาควิชาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

## ภาษาอังกฤษ

- . (n.d.). WinSAL. [Computer Program]. Germany: Media Enterprise (Distributor).
- Boersma, P. and Weenink, D. 1992. Praat [Computer Program]. Cambridge: Institute of Phonetic Sciences.
- Charpentier, F. and Moulines, E. 1989. Pitch-Synchronous Waveform Processing Techniques for Text-to-Speech Synthesis using Diphones. European Conference on Speech Communication and Technology , pp. 013-019.
- Charpentier, F. and Stella, M.G. 1986. Diphone Synthesis Using an Overlap-Add Technique for Speech Waveforms Concatenation. Proceedings of ICASSP'86 International Conference on Acoustic, Speeh and Signal Processing. pp. 2015-2018.
- Choy, E. L.T. 1998. Waveform Interpolaiion Speech Coder at 4 kb/s. Master's Thesis. Faculty of Graduate Studies and Research. Department of Electrical and Computer Engineering, McGill University.
- Deller, J.R., Jr., Proakis, J.G. and Hansen, J.H. 1993. Discrete-Time Processing of Speech Signals. New Jersey: Prentice-Hall.
- Dew, D. and Jensen, P.J. 1977. Phonetic processing : The dynamics of speech. Ohio: Charles E. Merrill Publishing and A Bell & Howell.
- Donovan, R.E. 1996. Trainable Speech Synthesis. Doctoral dissertation, Department of Engineering, Cambridge University.
- Everett, S.S. Word Synthesis based on Line Spectrum Pairs. International Conference on Acoustic, Speeh and Signal Processing. Volume 1. pp. 675-678.
- Flanagan, J.L. 1972. Speech analysis synthesis and perception. Berlin: Springer-Verlag.
- Ince, A.N. 1992. Digital speech processing speech coding, synthesis and recognition. Kluwer Academic Publishers.
- Kiat-arpakul, R., Fakcharoenphol, J. and Keretho, S. 1995. A Combined Phoneme-based and Demisyllable-based Approach for Thai Speech Synthesis. Proceedings of SNLP' 95 The 2<sup>nd</sup> Symposium on Natural Language Processing. pp. 361-369.

- Kleijn, W.B. and Paliwal, K.K. 1995. Speech coding and synthesis. Amsterdam: Elsevier Science B.V.
- Lieberman, P. 1988. Speech physiology, speech perception, and acoustic phonetics. Cambridge: Cambridge University Press.
- Luksaneeyanawin, Sudaporn. 1989. A Thai text to speech system. Proceedings of The Regional Workshop on Computer Processing of Asian Language (CPAL). Asian Institute of Technology. pp. 305-15.
- Luksaneeyanawin, Sudaporn and Chokie, Alistair. 1993. Synthesis and Tone Transformation : Resynthesizing Thai tones using PSOLA technique. Ninth Australian International Conference on Language and Speech.
- Luksaneeyanawin, Sudaporn. 1993. Speech Computing and Speech Technology in Thailand. Proceedings of the Symposium on Natural Language Processing in Thailand. pp. 276-321.
- Luksaneeyanawin, Sudaporn. 1995. Tone Transformation. Proceedings of the Second Symposium on Natural Language Processing. pp. 345-353.
- Markel, J.D. and Gray, A.H., Jr. 1976. Linear prediction of speech. Berlin: Springer-Verlag.
- Pickett, J.M. 1980. The sound of speech communication. Baltimore: University Park Press.
- Rabiner, L.R. and Schafer, R.W. 1978. Digital processing of speech signals. New Jersey: Prentice-Hall.
- Robinson, T., Ph.D. 1998. Speech analysis [Machine readable data file]. University of Cambridge, Department of Engineering: Robinson, T. (Producer).
- Saravari, Chatchavalit and Satoshi Imai. 1983. A demisyllable approach to speech synthesis of Thai – A tone language. Journal of the Acoustic Society of Japan. 97-106.



ภาคผนวก



ภาคผนวก ก.

แบบประเมินคุณภาพ

หมายเหตุ : คำบางคำอาจไม่มีความหมาย และ กรุณาทำทุกข้อ

1	แป้น แมน	แพน แซน	แฮน แฟน	12	โชน โกน	โจน โปน	โยน โดน	23	ช้อด คอด	ฟอด ฮอด	วอด ชอด	34	ตริก คริก	ทริก พริก	ปริก คลิก
2	พาม ตาม	ปาม กาม	บาม ฟาม	13	ยีน ป็น	ซีน กีน	วิน จัน	24	เหิม เอ้ม	เพิ่ม เจิม	เลิม เฝิม	35	ทริก ตริก	พริก ปริก	คริก พลิก
3	ซุง ปุง	คุง พุง	ฟุง ทุง	14	ยอด จ้อด	วอด ชอด	ทอด ด้อด	25	เอิน เฟิน	เหิน เพิน	เชิน เซิน	36	เพริม เคลิม	เปลิม เปริม	เพลิม เคริม
4	พีด ปี้ด	ยี้ด วี้ด	ฟีด บี้ด	15	วิก ฟิก	ยิก ชิก	จิก ทิก	26	กั้ม คั้ม	ปรั้ม พลั้ม	พรั้ม ปลั้ม	37	ตริง คลึง	กวี้ง ปริง	กรึง กลึง
5	แบ็ก เนก	เมก เป็ก	เพก งก	16	เมน เนน	เงน เคน	เกน เทน	27	กลาย คลาย	ปลาย พวย	พลาย ปราย	38	ควอน ครอน	กรอน คลอน	กลอน กวอน
6	ป้าก ด้าก	ทาก พาก	ฟาก ชาก	17	คน ปน	งน ตน	กน จน	28	ตรีน พลีน	กรีน พริน	ปลีน ปรีน	39	เคริก เกริก	เทริก เควิก	เพริก เกลิก
7	ตอน ทอน	ซอน ซอน	ลอน คอน	18	จุ่น คุ่น	กุ่น นุ่น	มุ่น หุ่น	29	เปรม เพลม	เดรม เพรม	เปลม เกรม	40	เทรม ควม	เดรม เกลม	เพรม เกรม
8	นืด ดืด	ตืด ทืด	ปืด สืด	19	กิด งด	มด คด	นด พด	30	ทราก พลาก	คราก ปรัาก	พราก ปล้าก	41	ฟิม ฝิก	ฟิว ฟีน	ฝิด ฟิง
9	ชัน ตึน	ตึน ลึน	รึน ทึน	20	แกง แพง	แดง แนง	แงง แมง	31	ปราก ขลาก	กราก ผลาก	ถราก ตราด	42	รึก รึบ	รึด รึว	รึม รึน
10	เนิน เทิน	เดิน เงิน	เมิน เดิน	21	อึ้ง ปึ้ง	อึ้ง ตึ้ง	กึ้ง จึ้ง	32	เปริน เทริน	เกริน เคลิน	เดริน เพลิน	43	ฉาก ชาง	ชาว ชาน	ชาม ฉาบ
11	จั้ง กั้ง	ชั้ง ตั้ง	ปั้ง ยั้ง	22	อั้ง ฟั้ง	จั้ง หั้ง	สั้ง ชั้ง	33	ทริง ตริง	ปริง พริง	กรึง ครึง	44	เสง เสว	เสป เสม	เสน เสย

หมายเหตุ : คำบางคำอาจไม่มีความหมาย และ กรุณาทำทุกข้อ

45	แก้ว แลก	แก้ง แลบ	แล้ว แลด	56	แกง แกก	แกน แกด	แกว แกม	67	เอ ออ	อี แอ	อา เออ	78	มอ ม้อ	ม่อ หมอ	หม่อ (ไม่มี)
46	วีก วุ้น	วืบ วืม	วิด วืง	57	คอด คอบ	คอน คอม	คอง คอก	68	เออ อีอ	โอ อา	แอ ออ	79	หมอ ม้อ	มอ ม้อ	หม่อ (ไม่มี)
47	กิม กิก	กิว กิน	กิด กิง	58	ลาก ล่าง	ลาด ล่าน	ล้าม ลาบ	69	โอ แอ	ออ อู	อา เออ	80	หม่อ หมอ	มอ ม้อ	หม่อ (ไม่มี)
48	กาม กาย	กาก กาง	กาน กาว	59	พุม พุง	พุก พุย	พูน พูป	70	เอี้ย อัว	อี อา	เอื้อ อีอ	81	มอ หม่อ	ม้อ ม้อ	หมอ (ไม่มี)
49	แกน แกด	แกว แกก	แกม แกบ	60	ตบ ตั่ง	ตอก ตั่น	ตัม ตด	71	เอี้ย อา	อัว อีอ	เอื้อ เออ	82	ม้อ มอ	หมอ ม้อ	หม่อ (ไม่มี)
50	เกิน เกิบ	เกิม เกิง	เกิก เกิว	61	อีอ อู	อี แอ	เอ เออ	72	อู เอื้อ	อา อัว	อี เอี้ย				
51	คุม คูป	คูน คุย	คูก คุง	62	อี อา	อู อีอ	เออ โอ	73	อ้อ อีอ	ออ อ้อ	อ้อ (ไม่มี)				
52	เลม เลก	เลน เลย	เลง เลด	63	อี โอ	อู ออ	อีอ เออ	74	ออ อ้อ	อีอ อ้อ	อ้อ (ไม่มี)				
53	จอก จอย	จอน จอม	จอง จอด	64	อี แอ	เอ เออ	อู โอ	75	ออ อ้อ	อ้อ อีอ	อ้อ (ไม่มี)				
54	โลก โลน	โลม โลง	โลย โลบ	65	โอ เออ	เอ อา	อีอ ออ	76	อ้อ อ้อ	อีอ ออ	อ้อ (ไม่มี)				
55	ยาว ยาง	ยาน ยาม	ยาด ยาย	66	อู โอ	เออ ออ	เอ แอ	77	อ้อ อีอ	ออ อ้อ	อ้อ (ไม่มี)				



หมายเหตุ : คำบางคำอาจไม่มีความหมาย และ กรุณาทำทุกข้อ

83	เมิน	เงิน	เนิน	94	เวบ	เวม	เวน
	เกิน	เคิน	เวิน		เวก	เวย	เวว
84	เวิน	เมิน	เคิน	95	ยาย	ยาว	ยาม
	เนิน	เงิน	เกิน		ยาง	ยาก	ยาด
85	โมน	โดน	โงน	96	ยาว	ยาด	ยาม
	โนน	โทน	โดน		ยาง	ยาย	ยาก
86	โงน	โทน	โนน	97	ย้อม	ย้อย	ย้อก
	โดน	โมน	โดน		ย่อง	ย้อบ	ย้อน
87	เตน	เทน	เงน	98	ย้อม	ย่อง	ย้อย
	เนน	เมน	เดน		ย้อบ	ย้อน	ย้อก
88	งน	เตน	เนน	99	แวง	แวม	แวก
	เมน	เดน	เทน		แวน	แวน	แวง
89	นึ่ง	จึ่ง	มึ่ง	100	แวม	แวน	แวก
	ทึ่ง	ตึ่ง	ตึ่ง		แวง	แวน	แวก
90	มึ่ง	ตึ่ง	ตึ่ง	101	เกิม	เกิบ	เกย
	ทึ่ง	จึ่ง	นึ่ง		เกิน	เกิว	เกิด
91	กาม	มาม	คาม	102	เกิบ	เกิว	เกิด
	จาม	นาม	งาม		เกย	เกิม	เกิน
92	นาม	จาม	กาม				
	คาม	มาม	งาม				
93	เวย	เวก	เวว				
	เวน	เวม	เวบ				

หมายเหตุ : คำบางคำอาจไม่มีความหมาย และ กรุณาทำทุกข้อ

		ดีมาก	ดี	ปานกลาง	แย่มาก
1	แฟน				
2	ปาม				
3	พุง				
4	วีต				
5	แมก				
6	ทาก				
7	ทอน				
8	ดีด				
9	รีน				
10	เนน				
11	จิ่ง				
12	โจน				
13	ยีน				
14	ยอด				
15	ยิก				
16	เงน				
17	กน				
18	งุ่น				
19	งด				
20	แงง				
21	อั้ง				
22	หึ่ง				
23	ฮอด				
24	เหิม				
25	เอิน				

		ดีมาก	ดี	ปานกลาง	แย่มาก
26	ปลืม				
27	ปลาย				
28	ปรีน				
29	เปรม				
30	พราก				
31	ตราด				
32	เตริน				
33	ตริง				
34	ทริก				
35	ทริก				
36	เคลิม				
37	กริ่ง				
38	ควอน				
39	เคริก				
40	คววม				
41	ฟิว				
42	รีบ				
43	ชาม				
44	เสม				
45	แล้ว				
46	วีต				
47	กิน				
48	กาน				
49	แกน				
50	เกิน				
51	คุย				

		ดีมาก	ดี	ปานกลาง	แย่มาก
52	เลย				
53	จอย				
54	โดย				
55	ยาย				
56	แกง				
57	คอก				
58	ลาก				
59	พุง				
60	แตก				
61	อื				
62	อือ				
63	อู				
64	เอ				
65	เออ				
66	โอ				
67	แอ				
68	อา				
69	อช				
70	เอี้ย				
71	เอือ				
72	อัว				
73	อช				
74	อ้อ				
75	อ้อ				
76	อ้อ				

		ดีมาก	ดี	ปานกลาง	แย่มาก
77	อ้อ				
78	มอ				
79	หม่อ				
80	ม่อ				
81	ม้อ				
82	หมอ				
83	เงิน				
84	เงิน				
85	โนน				
86	โนน				
87	เงน				
88	เงน				
89	นึ่ง				
90	นึ่ง				
91	งาม				
92	งาม				
93	เวว				
94	เวว				
95	ยาย				
96	ยาย				
97	ย้อย				
98	ย้อย				
99	แวว				
100	แวว				
101	เกิว				
102	เกิว				

		ดีมาก	ดี	ปานกลาง	แย่มาก
103	วลี 1				
104	วลี 2				
105	วลี 3				
106	วลี 4				
107	วลี 5				

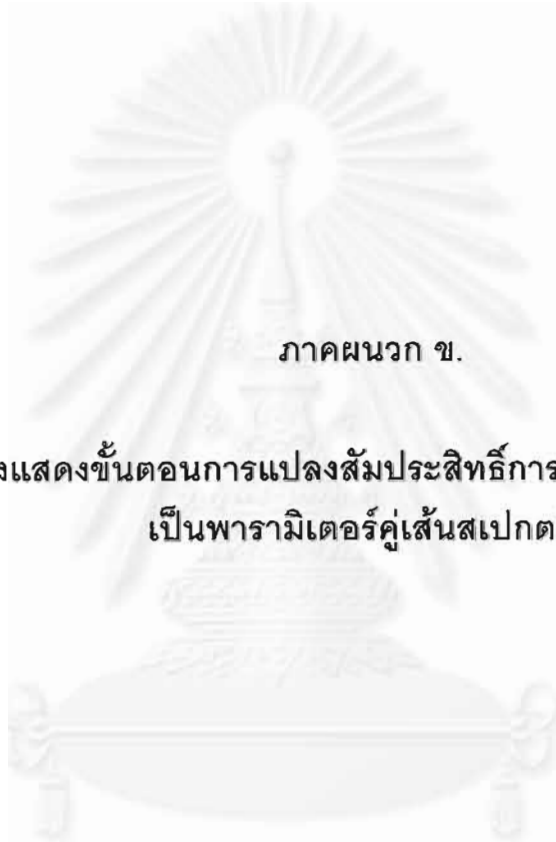
"สะ นวัด ดี ครับ ทุก ทุก ท่าน"

"การ สั้ง เคระะ เสียง ภา ชา โท"

"พาก วิ ชา วิ สะ วะ กำ ไฟ ฟ้า"

"พุด ทะ สัก กะ หลาด สอง พัน หน้า ร้อย สี่ สิบ สาม"

"น้ำ ใจ น่อง พี สี่ ชม พู"



ภาคผนวก ข.

แผนผังแสดงขั้นตอนการแปลงสัมประสิทธิ์การทำนายแบบเชิงเส้น  
เป็นพารามิเตอร์คู่เส้นสเปกตรัม

$$A_m(z) = 1 + \sum_{m=1}^M a_m z^{-m}$$



$$P(z) = A_M(z) + z^{-(M+1)} A(z^{-1}) = 1 + \sum_{m=1}^M (a_m + a_{M-m+1}) z^{-m} - z^{-(M+1)}$$

$$Q(z) = A_M(z) - z^{-(M+1)} A(z^{-1}) = 1 + \sum_{m=1}^M (a_m - a_{M-m+1}) z^{-m} + z^{-(M+1)}$$



หารากของ  $P(z)$  และ  $Q(z)$

$$P_{root} = Root(P(z)) = \sum_{m=1}^{M+1} (1 - z_{m,P} z^{-1})$$

$$Q_{root} = Root(Q(z)) = \sum_{m=1}^{M+1} (1 - z_{m,Q} z^{-1})$$



หามุมของค่าราก (เก็บ M ค่า)

P และ Q มี Zero ที่ -1 และ 1 ตามลำดับเสมอ  
(ค่าพารามิเตอร์ไม่ต้องเก็บค่ามุมของทั้งสองจุดนี้)

$$P_{root,angle} = \frac{1}{2\pi} \text{Angle}(P_{root}) = \frac{1}{2\pi} \text{arcTan}\left(\frac{\text{img}(z_{m,P})}{\text{real}(z_{m,P})}\right) \quad (M/2 + 1 \text{ ค่า})$$

$$Q_{root,angle} = \frac{1}{2\pi} \text{Angle}(Q_{root}) = \frac{1}{2\pi} \text{arcTan}\left(\frac{\text{img}(z_{m,Q})}{\text{real}(z_{m,Q})}\right) \quad (M/2 + 1 \text{ ค่า})$$



จัดเรียงค่ามุมจากน้อยไปมาก

$$P_{root,angle,sort} = \text{sort}(P_{root,angle})$$

$$Q_{root,angle,sort} = \text{sort}(Q_{root,angle})$$

ค่าพารามิเตอร์ LSP ได้จากการเรียง  $P_{root,angle,sort}$  ,  $Q_{root,angle,sort}$  สลับกัน

$$LSP - parameters = [p_1, q_1, p_2, q_2, \dots, p_{M/2}, q_{M/2}]$$

## ประวัติผู้เขียน

นายชัชวาลย์ หาญสกุลบรรเทิง เกิดวันที่ 28 พฤศจิกายน พ.ศ. 2520 ที่อำเภอเมือง จังหวัดยโสธร สำเร็จการศึกษาปริญญาตรีวิศวกรรมศาสตร์ ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2540 และเข้าศึกษาต่อในหลักสูตรวิศวกรรมศาสตรมหาบัณฑิต สาขาวิศวกรรมไฟฟ้า (แขนงวิชาการระบบเชิงเลข) ที่คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย เมื่อ พ.ศ. 2541 เคยได้รับการตีพิมพ์บทความทางวิชาการในการประชุมวิชาการทางวิศวกรรมไฟฟ้า ครั้งที่ 21 และครั้งที่ 22

