การรู้จำเสียงพูดภาษาไทยอย่างคงทนโดยใช้สัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอล ของค่าอัตสหสัมพันธ์ของเสียงพูดที่มีสัญญาณรบกวน

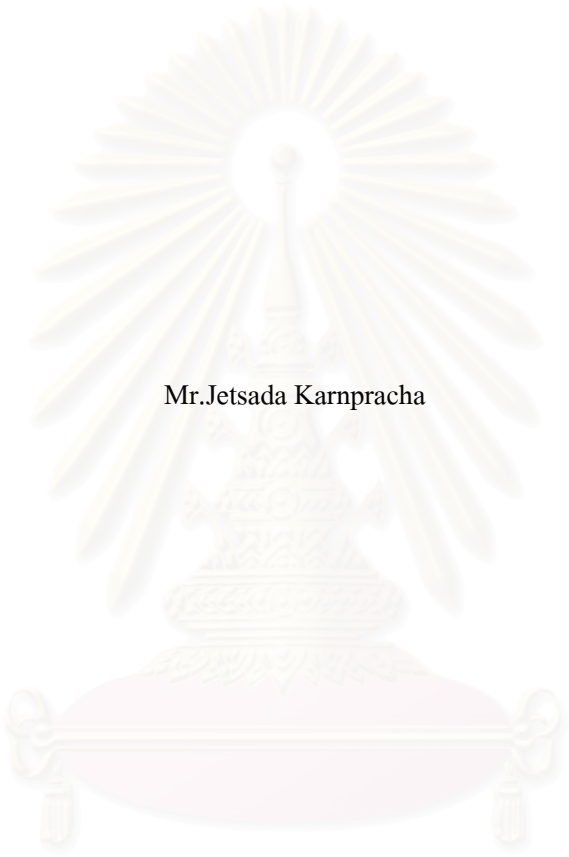
นาย เจษฎา กานต์ประชา

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมไฟฟ้า ภาควิชาวิศวกรรมไฟฟ้า
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2545
ISBN 974-17-2127-7
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

ROBUST THAI SPEECH RECOGNITION USING MFCC OF NOISY SPEECH AUTOCORRELATION

Mr.Jetsada Karnpracha



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

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 2002
ISBN 974-17-2127-7

หัวข้อวิทยานิพนธ์	การรู้จำเสียงพูดภาษาไทยอย่างคงทนโดยใช้สัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอลของค่าอัตสหสัมพันธ์ของเสียงพูดที่มีสัญญาณรบกวน
โดย	นายเจษฎา กานต์ประชา
ภาควิชา	วิศวกรรมไฟฟ้า
อาจารย์ที่ปรึกษา	รองศาสตราจารย์ ดร.สมชาย จิตะพันธ์กุล

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

..... คณบดีคณะวิศวกรรมศาสตร์
(ศาสตราจารย์ ดร.สมศักดิ์ ปัญญาแก้ว)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(ศาสตราจารย์ ดร.ประสิทธิ์ ประพิณมงคลการ)

..... อาจารย์ที่ปรึกษา
(รองศาสตราจารย์ ดร.สมชาย จิตะพันธ์กุล)

..... กรรมการ
(อาจารย์ สุวิทย์ นาคพีระยุทธ)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

เจษฎา กานต์ประชา : การรู้จำเสียงพูดภาษาไทยอย่างคงทนโดยใช้สัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอลของค่าอัตสหสัมพันธ์ของเสียงพูดที่มีสัญญาณรบกวน. (Robust Thai Speech Recognition Using MFCC of Noisy Speech Autocorrelation) อาจารย์ที่ปรึกษา : รองศาสตราจารย์ ดร.สมชาย จิตะพันธ์กุล, 54 หน้า. ISBN 974-17-2127-7.

วิทยานิพนธ์ฉบับนี้มีวัตถุประสงค์เพื่อนำเสนอการรู้จำเสียงพูดภาษาไทยอย่างคงทน การพัฒนาจะมุ่งความสนใจไปที่การหาค่าลักษณะสำคัญของสัญญาณเสียงพูดที่คงทนต่อสัญญาณรบกวนเกาส์เซียนขาวค่าเฉลี่ยศูนย์เพิ่มเข้ามา โดยใช้สัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอลของค่าอัตสหสัมพันธ์ของเสียงพูดซึ่งมีสัญญาณรบกวนเป็นสัญญาณเข้าในการหาค่าลักษณะสำคัญแทนสัญญาณเสียงพูด เนื่องจากมีความคงทนต่อสัญญาณรบกวนดังกล่าวมากกว่าตัวสัญญาณเสียงพูด กำหนดให้ระบบรู้จำเสียงพูดที่ใช้เป็นแบบไม่ขึ้นกับผู้พูด ใช้ระบบรู้จำเป็นแบบจำลองฮิดเดนมาร์คอฟ กลุ่มเสียงพูดคำโดดภาษาไทยที่ใช้ในการฝึกฝนและทดสอบจะแบ่งเป็น 5 กลุ่มตามเสียงวรรณยุกต์คือ สามัญ เอก โท ตรี และจัตวา การเปรียบเทียบอัตรารู้จำเฉลี่ยที่ได้จะเปรียบเทียบกับ อัตราการรู้จำเฉลี่ยที่ใช้สัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอลของเสียงพูดที่มีจำนวนอันดับเท่ากัน

ผลการทดสอบปรากฏว่า ค่าสัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอลของค่าอัตสหสัมพันธ์ของเสียงพูด อันดับ 32 เมื่อใช้กับเสียงพูดกลุ่มเสียงวรรณยุกต์สามัญ เอก โท และตรี ที่มีค่ากำลังสัญญาณต่อสัญญาณรบกวนน้อยกว่า 25 เดซิเบล ลงไป ให้ผลอัตราการรู้จำเฉลี่ยที่ดีกว่าสัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอลของเสียงพูดที่มีจำนวนอันดับเท่ากัน เฉลี่ยคิดเป็นร้อยละ 12.60 และให้ผลอัตราการรู้จำเฉลี่ยที่ดีกว่าในเสียงพูดกลุ่มเสียงวรรณยุกต์จัตวา เมื่อเลือกใช้จำนวนอันดับเป็น 16 ที่มีค่ากำลังสัญญาณต่อสัญญาณรบกวนในช่วงตั้งแต่ 5 เดซิเบล ไปจนถึง 20 เดซิเบล เฉลี่ยคิดเป็นร้อยละ 3.99

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาควิชา วิศวกรรมไฟฟ้า ลายมือชื่อนิสิต.....

สาขาวิชา วิศวกรรมไฟฟ้า ลายมือชื่ออาจารย์ที่ปรึกษา.....

ปีการศึกษา 2545

4270258421 : MAJOR ELECTRICAL ENGINEERING

KEY WORD: ROBUST SPEECH FEATURE / SPEECH RECOGNITION / MEL FREQUENCY CEPSTRAL COEFFICIENT / NOISY SPEECH AUTOCORRELATION

JETSADA KARNPRACHA : ROBUST THAI SPEECH RECOGNITION USING MFCC OF NOISY SPEECH AUTOCORRELATION. THESIS ADVISOR : ASSOC. PROF. SOMCHAI JITAPANKUL, Dr.Ing., 54 pp. ISBN 974-17-2127-7.

This thesis has the objective to develop a robust Thai speech recognition using robust speech feature with Hidden Markov Model. The system is an isolated word speaker independent system. Mel frequency cepstral coefficient of noisy speech autocorrelation is proposed to improved the robustness of the system to additive zero mean white gaussian noise. Thai isolated words are categorized to 5 classes according to their tones, mid tone, low tone, falling tone, high tone, and rising tone. The average recognition rate using MFCC of noisy speech autocorrelation is compared with MFCC of noisy speech in the same order.

The experiment results show that the average recognition rate of MFCC of noisy speech autocorrelation order 32 is 12.60 percent better than MFCC of noisy speech at SNR < 25 dB for the mid tone, low tone, falling tone, and high tone. In the rising tone, the average recognition rate of MFCC of noisy speech autocorrelation order 16 is 3.99 percent better than MFCC of noisy speech between 5 dB ข SNR ข 20 dB.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Department Electrical Engineering Student's signature .....

Field of study Electrical Engineering Advisor's signature .....

Academic year 2002

กิตติกรรมประกาศ

ในการดำเนินการวิจัยและจัดทำวิทยานิพนธ์ฉบับนี้ ผู้วิจัยขอกราบขอบพระคุณ รองศาสตราจารย์ ดร.สมชาย จิตะพันธ์กุล อาจารย์ที่ปรึกษา ที่ได้กรุณาให้คำแนะนำในการดำเนินการวิจัย แนวทางวิจัย ตลอดจนให้คำปรึกษารวมทั้งจัดหาอุปกรณ์และทุนวิจัยเพื่อดำเนินการวิจัยอย่างครบถ้วน

ขอขอบคุณ คุณวิศรุต อาขุบุตร คุณเอกฤทธิ์ มณีน้อย ที่ได้ให้คำแนะนำ และความช่วยเหลือต่างๆ ทำให้งานวิจัยลุล่วงได้ด้วยดี

ขอขอบคุณ คุณณัฐพล อุ่นศรี และเพื่อนนิสิตปริญญาโททุกท่านที่ได้ให้ความช่วยเหลือและเป็นกำลังใจให้เสมอมา

สุดท้ายนี้ ขอกราบขอบพระคุณ บิดา มารดา ที่ได้ให้กำลังใจตลอดเวลาที่ได้ศึกษาวิจัยจนกระทั่งวิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปด้วย

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง
บทคัดย่อภาษาอังกฤษ ..... จ
กิตติกรรมประกาศ ..... ฉ
สารบัญตาราง ..... ฌ
สารบัญภาพ ..... ญ
สารบัญคำศัพท์ ..... ฎ
บทที่ 1 บทนำ 1
ที่มาของปัญหา 1
วัตถุประสงค์ 4
ขั้นตอนการดำเนินงาน 4
เป้าหมายและขอบเขตงานวิจัย 4
ประโยชน์ที่คาดว่าจะได้รับ 4
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง 5
ส่วนรับสัญญาณเสียงพูด 5
การประมวลผลสัญญาณเบื้องต้น 5
การสกัดค่าลักษณะสำคัญ 7
การวิเคราะห์ค่าลักษณะสำคัญเมื่อมีสัญญาณรบกวน 19
บทที่ 3 ขั้นตอนวิธีในการดำเนินงานวิจัย 21
การกำหนดชุดคำศัพท์ 21
การเตรียมตัวอย่างข้อมูลเสียงพูด 22
การประมวลผลสัญญาณเบื้องต้น 24
การหาค่าลักษณะสำคัญ 25
วิธีการฝึกฝนและการทดสอบอัตราการรู้จำค่าลักษณะสำคัญ 26
บทที่ 4 ผลการทดลองและวิเคราะห์ผลการทดลอง 28
ขั้นตอนวิธีการรู้จำคำพูดภาษาไทย 28
ผลและการวิเคราะห์ผลการรู้จำเสียงพูดภาษาไทย 29
บทที่ 5 สรุปผลและข้อเสนอแนะ 49
สรุปผลการวิจัย 49
ข้อเสนอแนะสำหรับงานวิจัยในอนาคต 50

สารบัญ (ต่อ)

หน้า

รายการอ้างอิง..... 52
ประวัติผู้เขียนวิทยานิพนธ์..... 54

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญตาราง

หน้า

ตารางที่ 3.1 ชุดคำศัพท์ทั้งหมด.....21
ตารางที่ 3.2 การจัดแบ่งกลุ่มของผู้พูดสำหรับการฝึกฝนและทดสอบ.....27
ตารางที่ 4.1 ผลอัตราการรู้จำเฉลี่ยของเสียงพูดกลุ่มเสียงวรรณยุกต์สามัญ..... 29
ตารางที่ 4.2 ผลอัตราการรู้จำเฉลี่ยของเสียงพูดกลุ่มเสียงวรรณยุกต์เอก..... 31
ตารางที่ 4.3 ผลอัตราการรู้จำเฉลี่ยของเสียงพูดกลุ่มเสียงวรรณยุกต์โท..... 33
ตารางที่ 4.4 ผลอัตราการรู้จำเฉลี่ยของเสียงพูดกลุ่มเสียงวรรณยุกต์ตรี.....35
ตารางที่ 4.5 ผลอัตราการรู้จำเฉลี่ยของเสียงพูดกลุ่มเสียงวรรณยุกต์จัตวา.....37
ตารางที่ 4.6 ผลอัตราการรู้จำเฉลี่ยของเสียงพูดกลุ่มเสียงวรรณยุกต์สามัญ..... 39
ตารางที่ 4.7 ผลอัตราการรู้จำเฉลี่ยของเสียงพูดกลุ่มเสียงวรรณยุกต์เอก..... 41
ตารางที่ 4.8 ผลอัตราการรู้จำเฉลี่ยของเสียงพูดกลุ่มเสียงวรรณยุกต์โท..... 43
ตารางที่ 4.9 ผลอัตราการรู้จำเฉลี่ยของเสียงพูดกลุ่มเสียงวรรณยุกต์ตรี.....45
ตารางที่ 4.10 ผลอัตราการรู้จำเฉลี่ยของเสียงพูดกลุ่มเสียงวรรณยุกต์จัตวา.....47

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญภาพ

หน้า

รูปที่ 2.1 ฟังก์ชันกรอบชนิด Hamming Window.....7
รูปที่ 2.2 ฟังก์ชันกรอบตัวยก.....15
รูปที่ 2.3 Mel Scale ของความถี่สัญญาณเสียง.....16
รูปที่ 2.4 วงจรกรองแบบผ่านแถบความถี่ที่ให้ค่าสัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอล.....18
รูปที่ 3.1 แผนภูมิเส้นระดับพลังงานของเสียงพูด “มือซ้าย”.....23
รูปที่ 4.1 อัตราการรู้จำเฉลี่ยของเสียงพูดกลุ่มเสียงวรรณยุกต์สามัญ.....30
รูปที่ 4.2 อัตราการรู้จำเฉลี่ยของเสียงพูดกลุ่มเสียงวรรณยุกต์เอก.....32
รูปที่ 4.3 อัตราการรู้จำเฉลี่ยของเสียงพูดกลุ่มเสียงวรรณยุกต์โท.....34
รูปที่ 4.4 อัตราการรู้จำเฉลี่ยของเสียงพูดกลุ่มเสียงวรรณยุกต์ตรี.....36
รูปที่ 4.5 อัตราการรู้จำเฉลี่ยของเสียงพูดกลุ่มเสียงวรรณยุกต์จัตวา.....38
รูปที่ 4.6 อัตราการรู้จำเฉลี่ยของเสียงพูดกลุ่มเสียงวรรณยุกต์สามัญ.....40
รูปที่ 4.7 อัตราการรู้จำเฉลี่ยของเสียงพูดกลุ่มเสียงวรรณยุกต์เอก.....42
รูปที่ 4.8 อัตราการรู้จำเฉลี่ยของเสียงพูดกลุ่มเสียงวรรณยุกต์โท.....44
รูปที่ 4.9 อัตราการรู้จำเฉลี่ยของเสียงพูดกลุ่มเสียงวรรณยุกต์ตรี.....46
รูปที่ 4.10 อัตราการรู้จำเฉลี่ยของเสียงพูดกลุ่มเสียงวรรณยุกต์จัตวา.....48

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ความถี่ของสัญญาณเสียง	acoustic frequency
วิธีอัตสหสัมพันธ์	autocorrelation method
วงจรกรองแบบผ่านแถบความถี่	band pass filter
สัมประสิทธิ์เซปสตรอล	cepstral coefficient
เสียงพูดแบบคำต่อเนื่อง	connected word
เสียงพูดแบบต่อเนื่อง	continuous word
ค่าความแปรปรวนร่วม	covariance
ชุดวงจรกรองแบบดิจิตอล	digital filter bank
ช่วงพิสัยพลวัต	dynamic range
ค่าเจาะจง	eigenvalue
จุดเปลี่ยนระดับพลังงาน	energy level threshold
การสกัดค่าลักษณะสำคัญ	feature extraction
ตัวกรองเชิงเลขอันดับหนึ่ง	first order digital filter
การแปลงฟูริเยร์	fourier transformation
การแปลงฟูริเยร์ผกผัน	inverse fourier transform
เสียงพูดแบบคำโดด	isolated word
แบบจำลองซ้าย-ขวา	left-right model
ฟังก์ชันกรอบตัวยก	lifter
ความถี่ช่วงที่เป็นเชิงเส้น	linear frequency scale
สัมประสิทธิ์การประมาณพันธะเชิงเส้น	linear prediction coefficient
แบบจำลองการประมาณพันธะเชิงเส้น	linear prediction model
สัมประสิทธิ์แอลพีแอลอี	linear prediction with linear extrapolation coefficient
สัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอล	mel frequency cepstral coefficient
ค่าผิดพลาดกำลังสองเฉลี่ยน้อยสุด	minimum mean square error
ความถี่ที่ไม่เชิงเส้น	nonlinear frequency scale
การวิเคราะห์โดยไม่ใช้พารามิเตอร์	non parametric analysis
ความถี่ของการได้ยิน	perceptual frequency
สเปกตรัมกำลัง	power spectral
กรรมวิธีเน้นล่วงหน้า	preemphasis

ค่าผิดพลาดการรู้จำ	recognition error
อัตราการรู้จำ	recognition rate
แบบจำลองอนุกรม	serial model
การประมวลผลสัญญาณเบื้องต้น	signal preprocessing
อัตราส่วนสัญญาณต่อสัญญาณรบกวน	signal to noise ratio
กรรมวิธีวางกรอบขนาดสัญญาณ	smoothing window
ขึ้นกับผู้พูด	speaker dependent
ไม่ขึ้นกับผู้พูด	speaker independent
กรอบเสียงพูด	speech frame
ส่วนรับสัญญาณเสียงพูด	speech input
ไม่แปรเปลี่ยนตามเวลา	time invariant
แปรเปลี่ยนตามเวลา	time variant
ไม่เสถียร	unstable
เส้นเสียง	vocal cord
ชุดรูปร่องต้นแบบอ้างอิง	word reference templates
ฟังก์ชันกรอบ	window function

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 1

บทนำ

1.1 ที่มาของปัญหา

ปัจจุบันมีการพัฒนาเทคโนโลยีขึ้นอย่างรวดเร็ว โดยเฉพาะเทคโนโลยีด้านคอมพิวเตอร์ ความฝันหนึ่งของมนุษย์นับแต่อดีตคือ การสร้างเครื่องคอมพิวเตอร์ที่มีศักยภาพในการทำงานเทียบเท่ากับมนุษย์ คิดเองได้ ทำเองได้ มีความรู้สึก มีอารมณ์ และไม่รู้จักคำว่าผิดพลาด ดังปรากฏในนิยายวิทยาศาสตร์เรื่องยิ่งใหญ่และทรงคุณค่ามากที่สุด “2001: A Space Odyssey” เขียนโดย Arthur C. Clarke และถูกสร้างเป็นภาพยนตร์ในปี พ.ศ. 2511 โดยผู้กำกับชื่อดัง Stanley Kubrick

นิยายวิทยาศาสตร์เรื่องนี้จะมีคอมพิวเตอร์ที่ดีที่สุดในโลก “HAL” เป็นเครื่องคอมพิวเตอร์ที่ถูกสร้างให้เป็นไปตามความฝันของมนุษย์ทุกอย่าง ทำหน้าที่ควบคุมระบบต่างๆของยานอวกาศตลอดการเดินทางไปสู่ดาวพฤหัส แต่แล้วก็เกิดความผิดพลาดขึ้น ทำให้นักบินอวกาศเสียชีวิตไป 3 คน ด้วยความที่ HAL มีความรู้สึกนึกคิด จึงทำให้เกิดความกลัว กลัวว่าจะถูกทำลาย เพราะตัวเองได้ทำระบบผิดพลาด HAL จึงต้องพยายามกำจัดนักบินอวกาศอีก 2 คน ที่เหลือในยานด้วย โดยพยายามรวบรวมข้อมูลทุกอย่างที่ได้ ทั้งพฤติกรรมของนักบินอวกาศ รวมทั้งบทสนทนา

แม้ว่าจินตนาการเหล่านี้จะยังไม่เป็นจริง แต่ก็แสดงให้เห็นถึงแนวทางการพัฒนาเทคโนโลยีด้านนี้ต่อไปในอนาคต เทคโนโลยีด้านคอมพิวเตอร์ที่สำคัญต่อจินตนาการเหล่านี้คือ ปัญญาประดิษฐ์ (Artificial Intelligence) ที่สอนให้คอมพิวเตอร์คิดเป็น แต่กว่าจะพัฒนาให้เป็นปัญญาประดิษฐ์ที่สมบูรณ์ตามจินตนาการได้ จำเป็นต้องทำให้คอมพิวเตอร์สามารถจดจำและเข้าใจความหมายของสิ่งที่รับรู้ได้จากระบบสัมผัสทั้งห้า หนึ่งในระบบสัมผัสทั้งห้าที่เราสนใจขณะนี้คือ ระบบการฟัง

ระบบการฟังที่เราให้ความสำคัญเป็นอันดับแรกคือ ฟังภาษาพูดของมนุษย์เข้าใจ หรือที่เรียกว่า ระบบรู้จำเสียงพูด ระบบรู้จำเสียงพูดยังแยกย่อยออกเป็น การรู้จำแบบขึ้นกับผู้พูด (Speaker Dependent) และการรู้จำแบบไม่ขึ้นกับผู้พูด (Speaker Independent) นอกจากนี้ยังแยกย่อยได้อีกตามลักษณะเสียงพูดเป็น เสียงพูดแบบคำโดด (Isolated Word) เสียงพูดแบบคำต่อเนื่อง (Connected Word) และเสียงพูดแบบต่อเนื่อง (Continuous Speech)

ระบบรู้จำเสียงพูดที่มีอัตราการรู้จำเสียงพูดสูงคือ การรู้จำแบบขึ้นกับผู้พูดที่มีลักษณะเสียงพูดเป็นคำโดด ระบบการรู้จำแบบนี้ จะใช้ผู้ฝึกฝนและใช้งานเป็นบุคคลคนเดียวกัน มีการนำระบบรู้จำแบบนี้ไปใช้งานในชีวิตประจำวันแล้ว เช่น ในโทรศัพท์เคลื่อนที่รุ่นใหม่ จะใช้ระบบรู้จำแบบขึ้น

กับผู้พูด ทำหน้าที่หมุนโทรศัพท์ตามเสียงพูดสั่งงานของผู้ใช้ให้โทรไปยังผู้รับปลายทางที่เป็นเสียงพูดคำโดดที่ได้กำหนดไว้ล่วงหน้า หรือแม้แต่ในการพิมพ์งานโดยใช้คอมพิวเตอร์ มีโปรแกรมช่วยพิมพ์งานตามเสียงพูดของผู้ใช้งาน ในอนาคตข้างหน้าจะได้เห็นอุปกรณ์อิเล็กทรอนิกส์ใหม่ๆ ที่ไม่มีแผงควบคุมซับซ้อนเช่นปัจจุบัน ทุกอย่างจะถูกสั่งงานด้วยเสียงพูดของเจ้าของ ทำให้ดูกลมกลืนกับวิถีการดำเนินชีวิตของมนุษย์ยิ่งขึ้น

อย่างไรก็ตาม ระบบรู้จำเสียงพูดภาษาไทยที่พัฒนาขึ้นใช้ในปัจจุบัน ไม่ว่าจะเป็นเสียงพูดคำโดด (ระพีพัฒน์ เพ็ญศิริ, 2538), (ธีระ ภัทราพรนันท์, 2538), (เสาวลักษณ์ อารีย์พงศา, 2538), (วุฒิพงษ์ พรสุขจันทรา, 2538) และ คำต่อเนื่อง (วิศรุต อาขุบุตร, 2539), (ชัย วุฒิวิวัฒน์ชัย, 2540) ยังมีปัญหาต้องแก้ไขในการนำมาใช้งานจริงหลายจุด โดยเฉพาะปัญหาหนึ่งที่สำคัญมากคือ ระบบการสั่งงานด้วยเสียงพูดจะถูกรบกวนด้วยเสียงรบกวนจากสภาพแวดล้อมได้ง่าย ทำให้ระบบรู้จำเสียงพูดที่ถูกพัฒนาขึ้นในสภาพแวดล้อมที่ไม่มีเสียงรบกวน มีอัตราการรู้จำลดลงมากเมื่อนำไปใช้ในสภาพแวดล้อมทั่วไปที่มีเสียงรบกวน และเนื่องจากเป็นไปไม่ได้ที่จะหลีกเลี่ยงการใช้งานจริงในสภาพแวดล้อมที่มีเสียงรบกวนได้เสมอไป จึงจำเป็นต้องหาวิธีปรับปรุงระบบรู้จำเสียงพูดภาษาไทยให้ใช้งานได้ดีขึ้นในสภาพแวดล้อมที่มีเสียงรบกวน

แนวทางปรับปรุงให้เป็นระบบรู้จำเสียงพูดแบบคงทนที่มีการพัฒนาขึ้นสำหรับภาษาต่างประเทศ ประกอบด้วย 3 แนวทางคือ

1. กรองสัญญาณเสียงพูดที่มีสัญญาณรบกวนก่อนเข้าสู่ขั้นตอนการหาค่าลักษณะสำคัญ
2. หาค่าลักษณะสำคัญของสัญญาณเสียงพูดที่คงทนต่อสัญญาณรบกวน
3. ปรับเปลี่ยนพารามิเตอร์และแบบจำลองของการสร้างรูปแบบอ้างอิงและแยกความคล้ายของรูปแบบเสียงพูดให้คงทนต่อสัญญาณเสียงรบกวน

แนวทางแรกคือการออกแบบวงจรกรองที่สามารถลดสัญญาณรบกวนได้ตามค่าลักษณะของสัญญาณรบกวนที่มีในสัญญาณเสียงพูด เช่น เครื่องกรอง Kalman ซึ่งจะได้ผลดีเมื่อนำมาใช้กับสัญญาณเสียงพูด แต่ข้อด้อยคือใช้เวลานานในการคำนวณ เพราะมีค่าต่างๆในการคำนวณเป็นจำนวนมาก ทำให้การใช้งานจริงไม่สะดวก

แนวทางที่สอง เป็นแนวทางที่ใช้ในวิทยานิพนธ์นี้ คือแนวทางในการหาค่าลักษณะสำคัญของเสียงพูดที่คงทนต่อสัญญาณรบกวน ซึ่งเป็นวิธีการปรับเปลี่ยนค่าลักษณะสำคัญให้สามารถคงทนหรือลดผลกระทบของสัญญาณรบกวนก่อนที่จะนำไปใช้ในการฝึกฝนต่อไป ซึ่งจะยังผลให้ค่าผิดพลาดการรู้จำ (Recognition error) ลดลง จะเห็นว่าเป็นการแก้ไขกับตัวสัญญาณโดยตรง จึงทำได้ง่ายและใช้เวลาในการคำนวณน้อย

แนวทางสุดท้าย เป็นแนวทางที่ทำการวิจัยอย่างกว้างขวาง ซึ่งเป็นวิธีการปรับเปลี่ยนค่าพารามิเตอร์ของระบบการตัดสินใจเมื่อข้อมูลมีสัญญาณรบกวน เช่นวิธี Parallel Model Combination (PMC) ซึ่งหาแบบจำลองของข้อมูลสองชุดคือ สัญญาณเสียงพูดที่ไม่มีสัญญาณรบกวนและสัญญาณ

เสียงพูดที่มีสัญญาณรบกวนมาแยกกันฝึกฝนแล้วจึงรวมค่าพารามิเตอร์เพื่อฝึกฝนอีกครั้ง เพื่อเป็นการรวมแบบมาใช้ในกระบวนการตัดสินใจ ข้อด้อยของวิธีนี้คือใช้ได้ดีในระบบที่มีคำศัพท์ไม่มากนัก

ชนิดของสัญญาณรบกวนที่ให้ความสนใจคือ สัญญาณรบกวนเกาส์เซียนขาวค่าเฉลี่ยศูนย์เพิ่มเข้ามา (Additive Zero Mean White Gaussian Noise) สัญญาณรบกวนชนิดนี้จะทำให้เวกเตอร์ค่าลักษณะสำคัญของสัญญาณเสียงพูดทั้งที่เป็นสัมประสิทธิ์การประมาณพันธะเชิงเส้น และสัมประสิทธิ์เซปสตรอล มีการเปลี่ยนรูปแบบ Affine (R.J. Mammone, X. Zhang and R.P. Ramachandran, 1996) ทำให้ขนาดเวกเตอร์ค่าลักษณะสำคัญมีขนาดเล็กลงตามกำลังของสัญญาณรบกวน

งานวิจัยในต่างประเทศส่วนใหญ่จะเป็นการหาค่าลักษณะสำคัญของสัญญาณเสียงพูดที่คงทนต่อสัญญาณรบกวน เช่น การประมาณและกำจัดความแปรปรวนที่มารบกวนการหาค่าสัมประสิทธิ์การประมาณพันธะเชิงเส้น (K.F. Wong, S.H. Leung, and H.C. Ng., 1994) การออกแบบวงจรกรองสำหรับข้อเท็จจริงที่ว่ามนุษย์ให้ความสำคัญต่อสัญญาณที่มีการเปลี่ยนแปลงความถี่อย่างช้าๆ เพียงเล็กน้อย จึงไม่เกิดปัญหาเรื่องสัญญาณรบกวนในการติดต่อสื่อสาร (H. Hermansky and N. Morgan, 1994)

อย่างไรก็ตามมีงานวิจัยที่ใช้ประโยชน์จากค่าอัตสหสัมพันธ์ (Autocorrelation) ของสัญญาณเสียงพูด เนื่องจากมีความคงทนต่อสัญญาณรบกวนเกาส์เซียนขาวค่าเฉลี่ยศูนย์เพิ่มเข้ามามากกว่าตัวสัญญาณเสียงพูด เช่น การนำค่าอัตสหสัมพันธ์ของสัญญาณเสียงพูดมาใช้เพื่อปรับปรุงการหาค่าสัมประสิทธิ์การประมาณพันธะเชิงเส้นของสัญญาณเสียงพูดให้ดีขึ้น (D. Mansour and B.H. Juang, 1989) และถูกนำมาใช้ได้ดีกับสภาพแวดล้อมในรถยนต์ที่มีสัญญาณรบกวนตลอดเวลา (I. Lecomte, M. Lever, J. Boudy and A. Tassy, 1989)

สำหรับวิทยานิพนธ์นี้จะมุ่งความสนใจไปที่การหาค่าลักษณะสำคัญของสัญญาณเสียงพูดที่คงทนต่อสัญญาณรบกวน โดยเป็นการหาจากค่าอัตสหสัมพันธ์ของสัญญาณเสียงพูด เนื่องจากมีความคงทนต่อสัญญาณรบกวนเกาส์เซียนขาวค่าเฉลี่ยศูนย์เพิ่มเข้ามามากกว่าตัวสัญญาณเสียงพูด อัตราการรู้จำที่ได้จึงควรจะมากกว่า ระบบรู้จำเสียงพูดภาษาไทยแบบคงทนที่ใช้ในวิทยานิพนธ์นี้เป็นแบบไม่ขึ้นกับผู้พูด และใช้ลักษณะเสียงพูดแบบคำโดด โดยมีการนำแบบจำลองฮิดเดนมาร์คอฟมาใช้ในการรู้จำเสียงพูด

1.2 วัตถุประสงค์

1. เพื่อพัฒนาวิธีการหาค่าลักษณะสำคัญของสัญญาณเสียงพูดภาษาไทยจากค่าอัตสหสัมพันธ์ของสัญญาณเสียงพูด ที่มีความคงทนต่อสภาพแวดล้อมที่มีสัญญาณรบกวนเกาส์เซียนขาวค่าเฉลี่ยศูนย์เพิ่มเข้ามา ได้เป็นอย่างดี
2. เพื่อพัฒนาระบบรู้จำเสียงพูดภาษาไทยให้มีความคงทนต่อสัญญาณรบกวนมากขึ้น

1.3 ขั้นตอนการดำเนินการ

1. ศึกษา ค้นคว้า และเก็บรวบรวมข้อมูลงานวิจัยเกี่ยวกับ ค่าลักษณะสำคัญของสัญญาณเสียงพูดภาษาไทยที่มีความคงทนต่อสัญญาณรบกวนเกาส์เซียนขาวค่าเฉลี่ยศูนย์เพิ่มเข้ามา
2. วิเคราะห์ค่าลักษณะสำคัญต่างๆ เพื่อหาค่าลักษณะสำคัญของสัญญาณเสียงพูดภาษาไทยที่มีความคงทนต่อสัญญาณรบกวนได้เป็นอย่างดี
3. นำข้อดีของค่าลักษณะสำคัญแบบต่างๆมาใช้ด้วยกันเพื่อเพิ่มความคงทนยิ่งขึ้น
4. ทดสอบค่าลักษณะสำคัญที่เสนอขึ้นและปรับปรุงค่าลักษณะสำคัญ
5. สรุป วิจารณ์และจัดทำเอกสารวิทยานิพนธ์

1.4 เป้าหมายและขอบเขตของงานวิจัย

หาค่าลักษณะสำคัญของสัญญาณเสียงพูดภาษาไทยจากค่าอัตสหสัมพันธ์ ที่มีความคงทนต่อสภาพแวดล้อมที่มีสัญญาณรบกวนเกาส์เซียนขาวค่าเฉลี่ยศูนย์เพิ่มเข้ามา ที่ค่า SNR = 15 dB โดยเมื่อนำแบบจำลองฮิดเดนมาร์คอฟมาใช้ในการรู้จำเสียงพูดภาษาไทยในกลุ่มเสียงวรรณยุกต์ต่างๆ แล้วมีอัตราการรู้จำอย่างน้อยร้อยละ 60

1.5 ประโยชน์ที่คาดว่าจะได้รับ

เมื่อได้ระบบรู้จำเสียงพูดภาษาไทยที่ใช้งานได้เป็นอย่างดีในสภาพแวดล้อมที่มีสัญญาณรบกวนเกาส์เซียนขาวค่าเฉลี่ยศูนย์เพิ่มเข้ามา จะทำให้มีการนำระบบรู้จำเสียงพูดภาษาไทยมาใช้งานอย่างกว้างขวางยิ่งขึ้น ไม่จำกัดเพียงในสถานที่มีเสียงสัญญาณรบกวนน้อยเท่านั้น และเป็นอีกก้าวหนึ่งที่สำคัญของการทำความฝันข้างต้นให้เป็นจริง

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในบทนี้จะได้กล่าวถึงทฤษฎีต่างๆที่นำมาใช้ในงานวิทยานิพนธ์ โดยเริ่มจาก ส่วนรับสัญญาณเสียงพูด การประมวลผลสัญญาณเบื้องต้น การสกัดค่าลักษณะสำคัญ และการวิเคราะห์ค่าลักษณะสำคัญเมื่อมีสัญญาณรบกวนเกาส์เซียนขาวค่าเฉลี่ยศูนย์เพิ่มเข้ามา

2.1 ส่วนรับสัญญาณเสียงพูด (Speech Input)

เป็นขั้นตอนการรับสัญญาณเสียงพูดจากภายนอก ผ่านการแปลงสัญญาณอนาล็อกเป็นสัญญาณดิสครีต แล้วนำมาผ่านกระบวนการควอนไทเซชันเพื่อทำให้เป็นสัญญาณดิจิตอล

2.2 การประมวลผลสัญญาณเบื้องต้น (Signal Preprocessing)

การประมวลผลสัญญาณเบื้องต้นเป็นขั้นตอนกรรมวิธีในการจัดเตรียมข้อมูล จากข้อมูลดิบของเสียงพูดที่ได้จากการบันทึกเสียงนำมาผ่านกรรมวิธีประมวลผลสัญญาณเชิงเลข เพื่อใช้ในการประมวลผลในขั้นตอนต่อไป เนื่องจากสัญญาณเสียงพูดโดยรวมจะแปรเปลี่ยนตามเวลา (Time Variant) และไม่เสถียร (Unstable) ดังนั้นในการประยุกต์ใช้งานกรรมวิธีประมวลผลสัญญาณเชิงเลขกับสัญญาณเสียงพูด จะแบ่งสัญญาณเสียงพูดออกเป็นช่วงย่อยๆ (Rabiner and Levinson 1981; Furui, 1989) เรียกว่า “กรอบเสียงพูด” (Speech Frame) โดยแต่ละกรอบเสียงพูดจะมีความยาวประมาณ 10-40 มิลลิวินาที ซึ่งถือได้ว่าสัญญาณเสียงพูดในแต่ละกรอบเสียงพูดมีความเสถียรและไม่แปรเปลี่ยนตามเวลา (Time Invariant) จากนั้นจึงสามารถทำการประมวลผลสัญญาณเชิงเลขกับสัญญาณในแต่ละกรอบเสียงพูดได้

2.2.1 กรรมวิธีเน้นล่วงหน้า (Preemphasis)

ขั้นตอนการเน้นล่วงหน้านี้ เป็นการบีบอัดช่วงพิสัยพลวัต (Dynamic Range) ของสัญญาณเสียงพูด โดยการให้ความลาดเอียงในเชิงความถี่แบบราบลง ซึ่งจะส่งผลให้อัตราส่วนสัญญาณต่อสัญญาณรบกวนมีค่าสูงขึ้น ในทางปฏิบัติแล้วจะนำสัญญาณผ่านตัวกรองเชิงเลขอันดับหนึ่ง (First Order Digital Filter) ที่มีฟังก์ชันถ่ายโอนดังแสดงในสมการที่ (2.1) และ (2.2) (Furui, 1989)

$$H(z) = 1 - az^{-1} \tag{2.1}$$

$$\tilde{s}(n) = s(n) - as(n-1) \tag{2.2}$$

เมื่อ a เป็นสัมประสิทธิ์ตัวกรอง

$\tilde{s}(n)$ เป็นค่าของสัญญาณเสียงพูดขาออกที่ผ่านกรรมวิธีเน้นล่วงหน้า

$s(n)$ เป็นค่าของสัญญาณเสียงพูดขาเข้า

โดยกำหนดให้ค่าสัมประสิทธิ์ของตัวกรอง a มีค่าเข้าใกล้ 1 เมื่อใช้ร่วมกับการวิเคราะห์หาค่าสัมประสิทธิ์การประมาณพันธะเชิงเส้นจะกำหนดให้ค่า $a = 0.95$ เนื่องจากเป็นค่าที่ให้ผลดีที่สุด (วิศรุต อาขุบุตร, 2538) (เสาวลักษณ์ อารีพงศา, 2538) (เอกฤทธิ์ มณีน้อย, 2541)

2.2.2 กรรมวิธีวางกรอบขนาดสัญญาณ (Smoothing Window)

ขั้นตอนกรรมวิธีการวางกรอบขนาดสัญญาณจัดเป็นขั้นตอนในการเตรียมข้อมูลในแต่ละกรอบข้อมูลเสียงพูดเพื่อการวิเคราะห์อัตสหสัมพันธ์ โดยการคูณแต่ละค่าของสัญญาณในกรอบข้อมูลเสียงพูดด้วยค่าฟังก์ชันกรอบ (Window Function) ซึ่งมีหลายประเภทได้แก่ Rectangular Window, Hamming Window, Hanning Window, Blackman Window, Kaiser Window เป็นต้น (Oppenheim and Schafer, 1989) ผลของการวางกรอบขนาดสัญญาณมี 2 ประการ ประการแรก เป็นการลดทอนแอมปลิจูดอย่างช้าๆที่บริเวณปลายแต่ละข้างของกรอบข้อมูลเสียงพูดเพื่อป้องกันการเปลี่ยนแปลงที่ไม่ต่อเนื่องอย่างกะทันหันที่จุดปลายของกรอบสัญญาณ ประการที่สอง เป็นการสร้างค่าการประสานสำหรับผลการแปลงฟูริเยร์ (Fourier Transformation) ของฟังก์ชันกรอบและแถบสเปกตรัมของเสียงพูด ในงานวิจัยนี้ได้เลือกใช้ฟังก์ชันกรอบชนิด Hamming Window ดังแสดงในรูปที่ 2.1 สัญญาณเสียงพูดที่ผ่านขั้นตอนนี้จะได้เป็นข้อมูลเสียงพูดเพื่อนำไปใช้ในกรรมวิธีสัญญาณดิจิตอลต่อไป ซึ่งกรรมวิธีการวางกรอบสัญญาณจะเป็นไปตามสมการที่ (2.3) และ (2.4) คือ

$$\tilde{x}_l(n) = x_l(n) \cdot w(n) \tag{2.3}$$

$$w(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right) \tag{2.4}$$

เมื่อ $l = 0, 1, \ldots, L-1$ และ $n = 0, 1, \ldots, N-1$

กำหนดให้ $x_l(n)$ คือ ค่าสัญญาณเสียงพูดของข้อมูลที่ n

$\tilde{x}_l(n)$ คือ ค่าสัญญาณเสียงพูดที่ผ่านกรรมวิธีการวางกรอบ

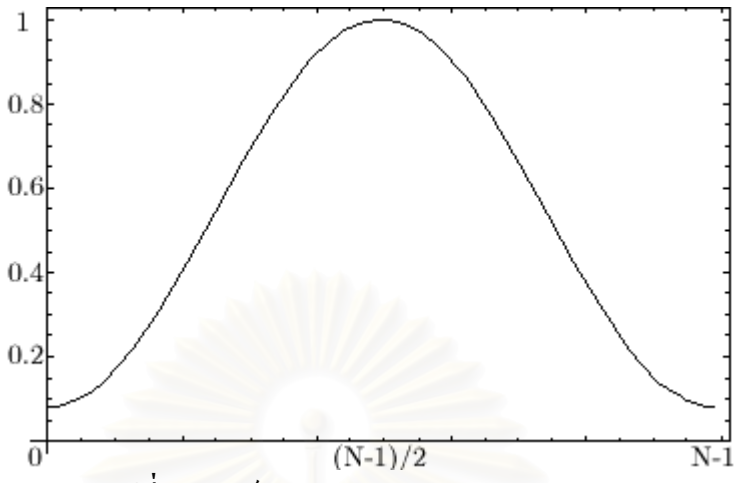
$w(n)$ คือ ฟังก์ชันกรอบชนิด Hamming Window

N คือ จำนวนของข้อมูลในแต่ละกรอบสัญญาณเสียงพูด

n คือ ลำดับข้อมูลในกรอบสัญญาณเสียงพูดที่ l

L คือ จำนวนของกรอบสัญญาณเสียงพูด

l คือ ลำดับของกรอบสัญญาณเสียงพูด



รูปที่ 2.1 ฟังก์ชันกรอบชนิด Hamming Window

2.3 การสกัดค่าลักษณะสำคัญ (Feature Extraction)

การวิเคราะห์และวัดค่าลักษณะสำคัญเป็นการวิเคราะห์สัญญาณเสียงพูด เพื่อเก็บรวบรวมลักษณะสำคัญของเสียงพูดแต่ละเสียง สำหรับการฝึกฝนระบบให้รับรู้ถึงความแตกต่างของเสียงพูดแต่ละเสียงและเพื่อใช้ในการเปรียบเทียบแบ่งแยกความแตกต่างของเสียงพูดแต่ละเสียงออกจากกัน ค่าลักษณะสำคัญที่เกี่ยวข้องในวิทยานิพนธ์มีด้วยกัน 7 แบบ ดังนี้

2.3.1 สัมประสิทธิ์การประมาณพันธะเชิงเส้น (Linear Prediction Coefficient, LPC)

คำว่า “การประมาณพันธะเชิงเส้น” หรือ Linear Prediction ถูกนำเสนอเป็นครั้งแรกโดย N.Weiner ในปี ค.ศ. 1966 โดยเทคนิคนี้ถูกนำมาใช้เป็นครั้งแรกกับการวิเคราะห์และการสังเคราะห์เสียงโดย Itakura ร่วมกับ Saito และ Atal ร่วมกับ Schoeder ในปี ค.ศ. 1968 (Furui, 1989) ความสำคัญของเทคนิคการประมาณพันธะเชิงเส้นนี้ก็คือการที่รูปและสมบัติทางความถี่ของเสียงพูดสามารถแสดงด้วยค่าพารามิเตอร์เพียงไม่กี่ค่าได้อย่างแม่นยำและมีประสิทธิภาพ นอกจากนี้ค่าพารามิเตอร์ดังกล่าวยังสามารถคำนวณได้ง่ายอีกด้วย

หลักการพื้นฐานของแบบจำลองการประมาณพันธะเชิงเส้น (Linear Prediction Model) คือ การคำนวณหาค่าอนาคตของสัญญาณจากการประมาณการผสมเชิงเส้น (Linear Combination) ของค่าสัญญาณก่อนหน้านั้น โดยแหล่งกำเนิดสัญญาณคือลมที่พ่นผ่านช่องว่างระหว่างเส้นเสียง (Vocal Cord) ออกมาเป็นรายคาบหรือสัญญาณรบกวนที่เกิดจากช่องแคบของการหดตัวของช่องทางเดิน

เสียงและที่มีผลกระทบกับช่องทางเดินเสียงส่วนบน การจำลองสัญญาณเสียงที่เปล่งออกมาสามารถจำลองได้ดังสมการที่ (2.5)

$$s(n) = \sum_{k=1}^{p} a_k s(n-k) + G u(n) \tag{2.5}$$

โดย $s(n)$ เป็นสัญญาณเสียงพูดเวลาปัจจุบัน

$s(n-k)$ เป็นสัญญาณเสียงพูดที่เป็นอดีตตัวที่ k

a_k เป็นสัมประสิทธิ์การประมาณพันธะเชิงเส้นตัวที่ k

p เป็นอันดับของการประมาณพันธะเชิงเส้น

$u(n)$ เป็นแหล่งกำเนิดสัญญาณกระตุ้น

G เป็นอัตราขยาย

โดยทั่วไป เราจะไม่สนใจคิดค่า $u(n)$ และถือว่าละเลยได้ จึงเหลือเพียงการประมาณสัญญาณเสียงพูดเวลาปัจจุบันจากสัญญาณเสียงพูดในอดีตเท่านั้น ดังสมการที่ (2.6)

$$\tilde{s}(n) = \sum_{k=1}^{p} a_k s(n-k) \tag{2.6}$$

การหาค่าสัมประสิทธิ์การประมาณพันธะเชิงเส้น a_k ที่ต้องการ จะใช้การพิจารณาจากค่าสัมประสิทธิ์ที่ทำให้เกิดความคลาดเคลื่อนของการประมาณสัญญาณเสียงพูด $\tilde{s}(n)$ จาก $s(n)$ น้อยที่สุด โดยเลือกใช้การวัดค่าความคลาดเคลื่อนแบบค่าผิดพลาดกำลังสองเฉลี่ยน้อยสุด (Minimum Mean Square Error)

กำหนดให้ค่าความผิดพลาด $e(n)$ เป็นไปตามสมการ (2.7)

$$e(n) = s(n) - \tilde{s}(n) = s(n) - \sum_{k=1}^{p} a_k s(n-k) \tag{2.7}$$

จะได้ค่าความคลาดเคลื่อนแบบค่าผิดพลาดกำลังสอง E คือ

$$E = \sum_{n} e^2(n) = \sum_{n} \left[s(n) - \sum_{k=1}^{p} a_k s(n-k) \right]^2 \tag{2.8}$$

ค่าสัมประสิทธิ์การประมาณพันธะเชิงเส้น a_l ที่ทำให้ได้ค่า E น้อยสุด คือค่าสัมประสิทธิ์ที่ทำให้

$$\frac{\partial E}{\partial a_l} = 0 = -2s(n-l) \left[s(n) - \sum_{k=1}^{p} a_k s(n-k) \right] \tag{2.9}$$

จัดรูปแบบใหม่เป็น

$$\sum_{n} s(n)s(n-l) = \sum_{k=1}^{p} a_k \left(\sum_{n} s(n-k)s(n-l) \right) \tag{2.10}$$

เนื่องจากค่าความแปรปรวนร่วม (Covariance) $f(a,b)$ ของ $s(n)$ คือ

$$f(a,b) = \sum_{n} s(n-a)s(n-b) \tag{2.11}$$

จึงเขียนสมการ (2.10) ได้ใหม่เป็น

$$f(l,0) = \sum_{k=1}^{p} a_k f(k,l) \tag{2.12}$$

วิธีแก้สมการหาค่าสัมประสิทธิ์ a_k จะใช้วิธีอัตสหสัมพันธ์ (Autocorrelation Method) โดยสมมติให้สัญญาณเสียงพูดในแต่ละกรอบขนาด N ตัวอย่าง มีค่าเป็นศูนย์ในช่วง $n < 0$ และ $n \geq N$ ในกรณีนี้ $f(a,b)$ จะลดรูปเป็นค่าอัตสหสัมพันธ์ $r(|a-b|)$

โดย

$$r(k) = \sum_{n=k}^{N-1} s(n)s(n-k) \tag{2.13}$$

จึงได้สมการที่ใช้หาค่าสัมประสิทธิ์ a_k คือ

$$r(l) = \sum_{k=1}^{p} a_k r(|k-l|) \tag{2.14}$$

หรือเขียนในรูปเมตริกซ์ $r = Ra$ คือ

$$\begin{bmatrix} r(1) \\ r(2) \\ M \\ r(p) \end{bmatrix} = \begin{bmatrix} r(0) & r(1) & L & r(p-1) \\ r(1) & r(0) & L & r(p-2) \\ M & M & O & M \\ r(p-1) & r(p-2) & L & r(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ M \\ a_p \end{bmatrix} \tag{2.15}$$

จะสังเกตได้ว่าเมตริกซ์ R เป็นเมตริกซ์ชนิด Toeplitz คือเป็นเมตริกซ์ที่มีความสมมาตรในแนวทแยงมุม การคำนวณหาค่าเมตริกซ์ a จึงสามารถใช้วิธีการวนซ้ำของ Levinson-Durbin (Hayes Manson H, 1996) ซึ่งสะดวกและรวดเร็วกว่าการหาโดยใช้อินเวอร์สเมตริกซ์แบบตรงๆมากได้

2.3.1.1 ขั้นตอนวิธีการวนซ้ำของ Levinson-Durbin

จากสมการ (2.15) สามารถเขียนใหม่เป็น

$$\begin{bmatrix} r(1) & r(0) & r(1) & L & r(p-1) \\ r(2) & r(1) & r(0) & L & r(p-2) \\ M & M & M & O & M \\ r(p) & r(p-1) & r(p-2) & L & r(0) \end{bmatrix} \begin{bmatrix} -1 \\ a_1 \\ a_2 \\ M \\ a_p \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ M \\ 0 \end{bmatrix} \tag{2.16}$$

หรือเขียนในอีกรูปแบบหนึ่งเป็น

$$\begin{bmatrix} r(0) & r(1) & r(2) & \cdots & r(p) \\ r(1) & r(0) & r(1) & \cdots & r(p-1) \\ r(2) & r(1) & r(0) & \cdots & r(p-1) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r(p) & r(p-1) & r(p-2) & \cdots & r(0) \end{bmatrix} \begin{bmatrix} -1 \\ a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} e_p \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{2.17}$$

โดย

$$e_p = -r(0) + \sum_{k=1}^{p} a_k r(k) \tag{2.18}$$

วิธีการวนซ้ำของ Levinson-Durbin จะเป็นการหาค่าเวกเตอร์สัมประสิทธิ์ a_{j+1} จากค่าเวกเตอร์ a_j ที่รู้มาก่อน สมมติว่ารู้ค่าเวกเตอร์สัมประสิทธิ์ a_j ที่ทำให้สมการนี้เป็นจริง

$$\begin{bmatrix} r(0) & r(1) & r(2) & \cdots & r(j) \\ r(1) & r(0) & r(1) & \cdots & r(j-1) \\ r(2) & r(1) & r(0) & \cdots & r(j-2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r(j) & r(j-1) & r(j-2) & \cdots & r(0) \end{bmatrix} \begin{bmatrix} -1 \\ a_j(1) \\ a_j(2) \\ \vdots \\ a_j(j) \end{bmatrix} = \begin{bmatrix} e_j \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{2.19}$$

และต้องการแก้สมการหาค่าเวกเตอร์ a_{j+1} ที่ทำให้สมการนี้เป็นจริง

$$\begin{bmatrix} r(0) & r(1) & r(2) & \cdots & r(j) & r(j+1) \\ r(1) & r(0) & r(1) & \cdots & r(j-1) & r(j) \\ r(2) & r(1) & r(0) & \cdots & r(j-2) & r(j-1) \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ r(j) & r(j-1) & r(j-2) & \cdots & r(0) & r(1) \\ r(j+1) & r(j) & r(j-1) & \cdots & r(1) & r(0) \end{bmatrix} \begin{bmatrix} -1 \\ a_{j+1}(1) \\ a_{j+1}(2) \\ \vdots \\ a_{j+1}(j) \\ a_{j+1}(j+1) \end{bmatrix} = \begin{bmatrix} e_{j+1} \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} \tag{2.20}$$

ในตอนเริ่มต้นหากกำหนดค่าค่าเริ่มต้นของ $a_{j+1}(k) = a_j(k)$ โดย k เริ่มจาก 1 ถึง j และ $a_{j+1}(j+1) = 0$ แทนค่าลงไป (2.20) จะได้

$$\begin{bmatrix} r(0) & r(1) & r(2) & \cdots & r(j) & r(j+1) \\ r(1) & r(0) & r(1) & \cdots & r(j-1) & r(j) \\ r(2) & r(1) & r(0) & \cdots & r(j-2) & r(j-1) \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ r(j) & r(j-1) & r(j-2) & \cdots & r(0) & r(1) \\ r(j+1) & r(j) & r(j-1) & \cdots & r(1) & r(0) \end{bmatrix} \begin{bmatrix} -1 \\ a_j(1) \\ a_j(2) \\ \vdots \\ a_j(j) \\ 0 \end{bmatrix} = \begin{bmatrix} e_j \\ 0 \\ 0 \\ \vdots \\ 0 \\ g_j \end{bmatrix} \tag{2.21}$$

หากค่า $g_j = 0$ จะได้ $a_{j+1} = [a_j(1), a_j(2), \ldots, a_j(j), 0]^T$ เป็นค่าเวกเตอร์สัมประสิทธิ์ที่ต้องการ แต่ส่วนใหญ่แล้ว $g_j \neq 0$

สมการ (2.21) เขียนในอีกรูปหนึ่งได้เป็น

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

$$\begin{bmatrix} r(0) & r(1) & r(2) & \cdots & r(j) & r(j+1) \\ r(1) & r(0) & r(1) & \cdots & r(j-1) & r(j) \\ r(2) & r(1) & r(0) & \cdots & r(j-2) & r(j-1) \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ r(j) & r(j-1) & r(j-2) & \cdots & r(0) & r(1) \\ r(j+1) & r(j) & r(j-1) & \cdots & r(1) & r(0) \end{bmatrix} \begin{bmatrix} 0 \\ a_j(j) \\ a_j(j-1) \\ \vdots \\ a_j(1) \\ -1 \end{bmatrix} = \begin{bmatrix} g_j \\ 0 \\ 0 \\ \vdots \\ 0 \\ e_j \end{bmatrix} \tag{2.22}$$

รวมสมการ (2.21) และ (2.22) ที่คูณด้วยค่าคงที่ G_{j+1} เข้าด้วยกันจะได้

$$R_{j+1} \begin{bmatrix} -1 \\ a_j(1) \\ a_j(2) \\ \vdots \\ a_j(j) \\ 0 \end{bmatrix} + G_{j+1} \begin{bmatrix} 0 \\ a_j(j) \\ a_j(j-1) \\ \vdots \\ a_j(1) \\ -1 \end{bmatrix} = \begin{bmatrix} e_j \\ 0 \\ 0 \\ \vdots \\ 0 \\ g_j \end{bmatrix} + G_{j+1} \begin{bmatrix} g_j \\ 0 \\ 0 \\ \vdots \\ 0 \\ e_j \end{bmatrix} \tag{2.23}$$

ผลลัพธ์ที่ได้จะมีลักษณะเดียวกับสมการ (2.20) เมื่อ $g_j + G_{j+1} e_j = 0$ จึงกำหนดให้

$$G_{j+1} = -\frac{g_j}{e_j} \tag{2.24}$$

จะได้ค่าเวกเตอร์สัมประสิทธิ์ a_{j+1} ที่ต้องการคือ

$$a_{j+1} = \begin{bmatrix} a_j(1) \\ a_j(2) \\ \vdots \\ a_j(j) \\ 0 \end{bmatrix} + G_{j+1} \begin{bmatrix} a_j(j) \\ a_j(j-1) \\ \vdots \\ a_j(1) \\ -1 \end{bmatrix} \tag{2.25}$$

และ $$e_{j+1} = e_j + G_{j+1} g_j = e_j (1 - G_{j+1}^2) \tag{2.26}$$

เมื่อนำวิธีแก้ปัญหาดังกล่าวมาเขียนเป็นขั้นตอนจะเป็นดังนี้

ขั้นตอนที่ 1 กำหนดเงื่อนไขเริ่มต้น ดังนี้

$$e_0 = -r(0) \tag{2.27}$$

$$a_1(1) = 0 \tag{2.28}$$

ขั้นตอนที่ 2 คำนวณค่าสัมประสิทธิ์การสะท้อน G_{j+1} โดย j เริ่มจาก $0, 1, \ldots, p-1$

$$g_j = -r(j+1) + \prod_{k=1}^{j} a_j(k) r(j-k+1) \tag{2.29}$$

$$G_{j+1} = -\frac{g_j}{e_j} \tag{2.30}$$

โดย $r(j+1)$ และ $r(j-k+1)$ คำนวณได้จากสมการ (2.13)

ขั้นตอนที่ 3 คำนวณค่าเวกเตอร์สัมประสิทธิ์ a_{j+1} โดย j เริ่มจาก $1, 2, \mathrm{K}, p-1$

$$a_{j+1}(i) = a_j(i) + G_{j+1}a_j(j-i+1) \tag{2.31}$$

$$a_{j+1}(j+1) = -G_{j+1} \tag{2.32}$$

ขั้นตอนที่ 4 คำนวณค่าผิดพลาด e_{j+1} ใหม่

$$e_{j+1} = e_j(1-G_{j+1}^2) \tag{2.33}$$

เมื่อคำนวณจนครบทั้ง 4 ขั้นตอน ที่ $j = p-1$ จะได้ค่าเวกเตอร์ a_p ที่ต้องการ

2.3.2 สัมประสิทธิ์แอลพีแอลอี (Linear Prediction with Linear Extrapolation Coefficient, LPLEC)

สัมประสิทธิ์แอลพีแอลอีใช้หลักการพื้นฐานเดียวกับสัมประสิทธิ์การประมาณพันธะเชิงเส้น คือ คำนวนหาค่าอนาคตของสัญญาณจากการประมาณการผสมของค่าสัญญาณก่อนหน้านั้น ความแตกต่างคือ สัมประสิทธิ์การประมาณพันธะเชิงเส้น จำนวน p อันดับ ใช้คำนวนหาค่าอนาคตของสัญญาณจากการประมาณการผสมของค่าสัญญาณก่อนหน้านั้น p ค่า หากต้องการให้สัมประสิทธิ์การประมาณพันธะเชิงเส้นใช้เป็นตัวแทนสัญญาณได้ดีขึ้น ทำได้โดยคำนวนหาค่าอนาคตของสัญญาณจากการประมาณการผสมของค่าสัญญาณก่อนหน้านั้นเป็นจำนวนค่าที่มากกว่าเดิม เช่น เป็น $2p$ ค่า จำนวนค่าสัญญาณที่เพิ่มขึ้นทำให้ต้องใช้จำนวนสัมประสิทธิ์การประมาณพันธะเชิงเส้นเพิ่มขึ้นเป็น $2p$ อันดับเช่นกัน แต่หากเลือกใช้สัมประสิทธิ์แอลพีแอลอีจะใช้จำนวนสัมประสิทธิ์เป็น p อันดับ เท่านั้น (S. Varho and P. Alku, 1999)

หลักการที่ใช้ในการลดจำนวนสัมประสิทธิ์คือ การจัดกลุ่มค่าสัญญาณที่ติดกันเป็นคู่ จำนวน p คู่ แล้วประมาณค่าสัญญาณใหม่แทนคู่ดังกล่าว จึงได้ค่าสัญญาณใหม่จำนวน p ค่า จากนั้นจึงนำมาใช้ในการหาสัมประสิทธิ์การประมาณพันธะเชิงเส้นต่อไป สัมประสิทธิ์ที่ได้เรียกว่าสัมประสิทธิ์แอลพีแอลอี

สำหรับสัญญาณคู่ที่ k คือสัญญาณของ $s(n-2k)$และ $s(n-2k+1)$ จะนำมาประมาณค่าสัญญาณใหม่คือ $\hat{s}(n-k)$ ดังสมการที่ (2.34)

$$\hat{s}(n-k) = s(n-2k) + [s(n-2k+1) - s(n-2k)]\Delta k \tag{2.34}$$

ค่าสัญญาณใหม่ที่ได้จะนำมาใช้หาสัมประสิทธิ์การประมาณพันธะเชิงเส้น a_k ต่อไป จึงได้การประมาณสัญญาณเวลาปัจจุบันจากสัญญาณในอดีต ดังสมการ (2.35)

$$\%(n) = \prod_{k=1}^{p} a_k \{s(n-2k) + [s(n-2k+1) - s(n-2k)]\Delta k\} \tag{2.35}$$

การหาค่าสัมประสิทธิ์ a_k ที่ต้องการ จะใช้การพิจารณาจากค่าสัมประสิทธิ์ที่ทำให้เกิดความคลาดเคลื่อนของการประมาณสัญญาณ $\%(n)$ จาก $s(n)$ น้อยที่สุด โดยเลือกใช้การวัดค่าความคลาดเคลื่อนแบบค่าผิดพลาดกำลังสองเฉลี่ยน้อยสุด (Minimum Mean Square Error) และใช้วิธีอัตสหสัมพันธ์เช่นเดียวกับการหาค่าสัมประสิทธิ์การประมาณพันธะเชิงเส้น จะได้สมการสำหรับใช้หาค่าสัมประสิทธิ์แอลพีแอลอี a_k ดังสมการ (2.36)

$$\begin{bmatrix} r(2q-1) - r(2q) \\ r(2q) - r(2q-1) \end{bmatrix} = \prod_{k=1}^{p} a_k \begin{bmatrix} r(8kq - 2k - 2q + 1) + r(|2q - 2k|) \\ r(2k - 4kq) + r(|2q - 2k + 1|) \\ r(2q - 4kq) + r(|2q - 2k - 1|) \end{bmatrix} \quad (2.36)$$

จากสมการพบว่ารูปแบบของเมตริกซ์ทางขวาที่คูณกับค่าสัมประสิทธิ์แอลพีแอลอีไม่ได้อยู่ในรูปแบบ Toeplitz ทำให้ไม่สามารถใช้วิธีการวนซ้ำของ Levinson-Durbin เข้ามาช่วยแก้สมการได้ สัมประสิทธิ์แอลพีแอลอีจึงมีความซับซ้อนในการหามากกว่าสัมประสิทธิ์การประมาณพันธะเชิงเส้น

2.3.3 ค่าอัตสหสัมพันธ์ที่มีความคงทน

สัญญาณรบกวนเกาส์เซียนขาวค่าเฉลี่ยศูนย์เพิ่มเข้ามา เป็นสัญญาณรบกวนที่เป็นอิสระทางค่าสถิติ มีค่ากำลังเท่ากันตลอดทุกความถี่ และมีค่าอัตสหสัมพันธ์ดังสมการ

$$r_{nn}(k) = \begin{cases} h & , k = 0 \\ 0 & , other\ k \end{cases} \quad (2.37)$$

โดย h เป็นค่าสเปกตรัมกำลัง (Power Spectral) ของสัญญาณรบกวน

หากมีสัญญาณรบกวน $g(n)$ ดังกล่าวเพิ่มเข้ามา สัญญาณเสียงพูดที่มีเสียงรบกวน $s_{ข}(n)$ คือ

$$s_{ข}(n) = s(n) + g(n) \quad (2.38)$$

ค่าอัตสหสัมพันธ์ของสัญญาณเสียงพูดที่มีเสียงรบกวน $r_{s_{ข}}(n)$ คือ

$$r_{s_{ข}}(n) = \begin{cases} r_{ss}(0) + h & , n = 0 \\ r_{ss}(n) & , other\ n \end{cases} \quad (2.39)$$

จากสมการ (2.39) จะพบว่าค่าอัตสหสัมพันธ์เมื่อมีสัญญาณรบกวนเกาส์เซียนขาวค่าเฉลี่ยศูนย์เพิ่มเข้ามา จะมีความคงทนต่อสัญญาณรบกวนมาก เนื่องจากมีเพียงค่า $r_{s_{ข}}(0)$ เท่านั้นที่เปลี่ยนแปลงไป นอกจากนี้ค่าอัตสหสัมพันธ์ยังมีความสัมพันธ์แบบเชิงเส้นกับค่าอัตสหสัมพันธ์ในอดีตด้วย โดยพิจารณาจากสัญญาณเสียงพูดที่มีความสัมพันธ์แบบเชิงเส้นกับค่าในอดีต ดังสมการ

$$\%(n) = \prod_{k=1}^{p} a_k s(n - k) \quad (2.40)$$

มีฟังก์ชันการถ่ายโอนคือ $$H(z) = \frac{1}{1 - \prod_{k=1}^{p} a_k z^{-k}} \quad (2.41)$$

เพราะว่า $$r(n) = s(n) * s^{*}(-n) \quad (2.42)$$

จึงได้

$$R(z) = H(z)H^{*}(1/z^{*}) = H(z)H(z^{-1})$$

$$= \frac{G z^{-p}}{1 - \prod_{k=1}^{2p} d_k z^{-k}} \quad (2.43)$$

ดังนั้นจึงใช้สัมประสิทธิ์การประมาณพันธะเชิงเส้นกับค่าอัตสหสัมพันธ์ได้เช่นกัน และจะใช้ค่าอัตสหสัมพันธ์เริ่มจาก $r_{s'v'}(1)$, $r_{s'v'}(2)$, $r_{s'v'}(3)$, K เป็นต้นไปเป็นค่าอัตสหสัมพันธ์ที่มีความคงทน การหาค่าอัตสหสัมพันธ์ที่มีความคงทนจะใช้ตามสมการ (2.44) (D. Mansour and B.H. Juang , 1989)

$$r_{s'v'}(k) = \prod_{n=0}^{N-1} s_v(n) s_v(n-k) \quad (2.44)$$

โดย $s_v(n)$ เป็นสัญญาณเสียงพูดที่มีเสียงรบกวนที่ยังไม่ได้ผ่านฟังก์ชันกรอบจำนวน $2N$ ค่า

2.3.4 สัมประสิทธิ์แอลพีแอลอีของค่าอัตสหสัมพันธ์ที่มีความคงทน (Autocorrelation Linear Prediction with Linear Extrapolation Coefficient, ALPLEC)

จากสมการ (2.43) จะพบว่า หากพิจารณาใช้จำนวนอันดับของสัมประสิทธิ์การประมาณพันธะเชิงเส้นของสัญญาณเสียงพูดเป็น p สัญญาณเสียงดังกล่าวเมื่อนำมาหาค่าอัตสหสัมพันธ์ที่มีความคงทน และนำค่าอัตสหสัมพันธ์ที่มีความคงทนมาใช้เป็นสัญญาณเข้าแทนสัญญาณเสียงพูด ในการหาค่าสัมประสิทธิ์การประมาณพันธะเชิงเส้นของค่าอัตสหสัมพันธ์ที่มีความคงทน จะต้องใช้จำนวนอันดับเป็นจำนวน $2p$ การใช้ค่าสัมประสิทธิ์การประมาณพันธะเชิงเส้นจึงไม่เหมาะสม

วิทยานิพนธ์นี้จะใช้สัมประสิทธิ์แอลพีแอลอีของค่าอัตสหสัมพันธ์ที่มีความคงทนแทนการใช้สัมประสิทธิ์การประมาณพันธะเชิงเส้นของค่าอัตสหสัมพันธ์ที่มีความคงทน เพราะใช้จำนวนอันดับเป็น p แต่สามารถที่ทำหน้าที่แทนสัมประสิทธิ์การประมาณพันธะเชิงเส้นที่อันดับ $2p$ และเนื่องจากค่าอัตสหสัมพันธ์ที่มีความคงทนไม่เปลี่ยนแปลงเมื่อมีสัญญาณรบกวนเกาส์เซียนขาวค่าเฉลี่ยศูนย์เพิ่มเข้ามา ทำให้สัมประสิทธิ์แอลพีแอลอีของค่าอัตสหสัมพันธ์ที่มีความคงทน ไม่เปลี่ยนแปลงเช่นกัน ดังนั้นเมื่อนำมาใช้ในระบบรู้จำเสียงพูดจะมีความคงทนต่อสัญญาณรบกวนมากขึ้น

2.3.5 สัมประสิทธิ์เซปสตรอล (Cepstral Coefficient, CEP)

สัมประสิทธิ์เซปสตรอลกำหนดจากการแปลงฟูริเยร์ (Fourier Transform) ของสัญญาณ ดังสมการ (2.45)

$$c\langle n\rangle = F^{-1}\{\log|F\langle s\langle n\rangle\rangle|\} \tag{2.45}$$

นอกจากหาด้วยวิธีการแปลงฟูริเยร์แล้ว สัมประสิทธิ์เซปสตรอลสามารถคำนวณได้จากสัมประสิทธิ์การประมาณพันธะเชิงเส้น ซึ่งง่ายต่อการคำนวณและใช้เวลาน้อยกว่า วิธีคำนวณหาสัมประสิทธิ์เซปสตรอล แสดงในสมการที่ (2.46) และ (2.47) (Rabiner and Juang, 1993)

$$c_m = a_m + \sum_{k=1}^{m-1}\left(\frac{k}{m}\right)c_k a_{m-k} \quad \text{เมื่อ } 1 \le m \le p \tag{2.46}$$

$$c_m = \sum_{k=1}^{m-1}\left(\frac{k}{m}\right)c_k a_{m-k} \quad \text{เมื่อ } m > p \tag{2.47}$$

โดย a_m คือค่าสัมประสิทธิ์การประมาณพันธะเชิงเส้น

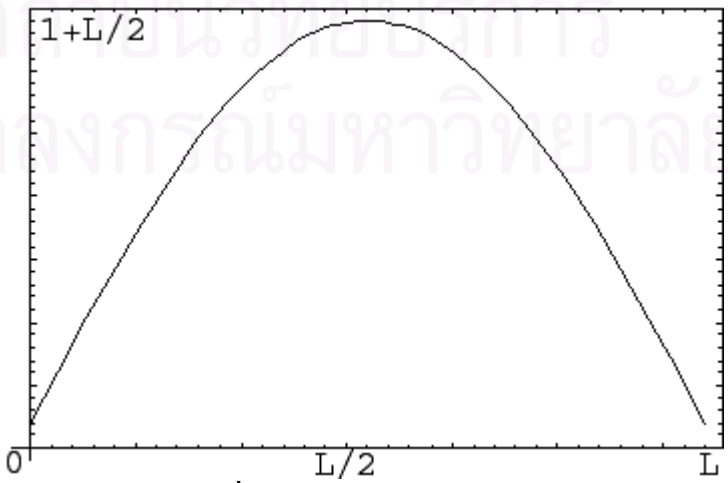
c_m คือค่าสัมประสิทธิ์เซปสตรอล

p คือ อันดับของค่าสัมประสิทธิ์การประมาณพันธะเชิงเส้น

สัมประสิทธิ์เซปสตรอลที่ได้จะถูกคูณกับฟังก์ชันกรอบตัวยก (Lifter) ก่อนนำไปใช้เพื่อตัดส่วนประกอบทางความถี่ของแหล่งกำเนิดสัญญาณกระตุ้นออกไป ฟังก์ชันกรอบตัวยกที่ใช้ในวิทยานิพนธ์เป็นดังรูปที่ 2.2 มีสมการคือ (Deller, Jr. John R., Hansen John H.L., and Proakis John G., 1993)

$$l\langle n\rangle = \begin{cases} 1 + \frac{L}{2}\sin\left(\frac{\pi n}{L}\right), & n = 0, 1, \ldots, L \\ 0, & \text{other } n \end{cases} \tag{2.48}$$

โดย L เป็นอันดับของค่าสัมประสิทธิ์เซปสตรอล



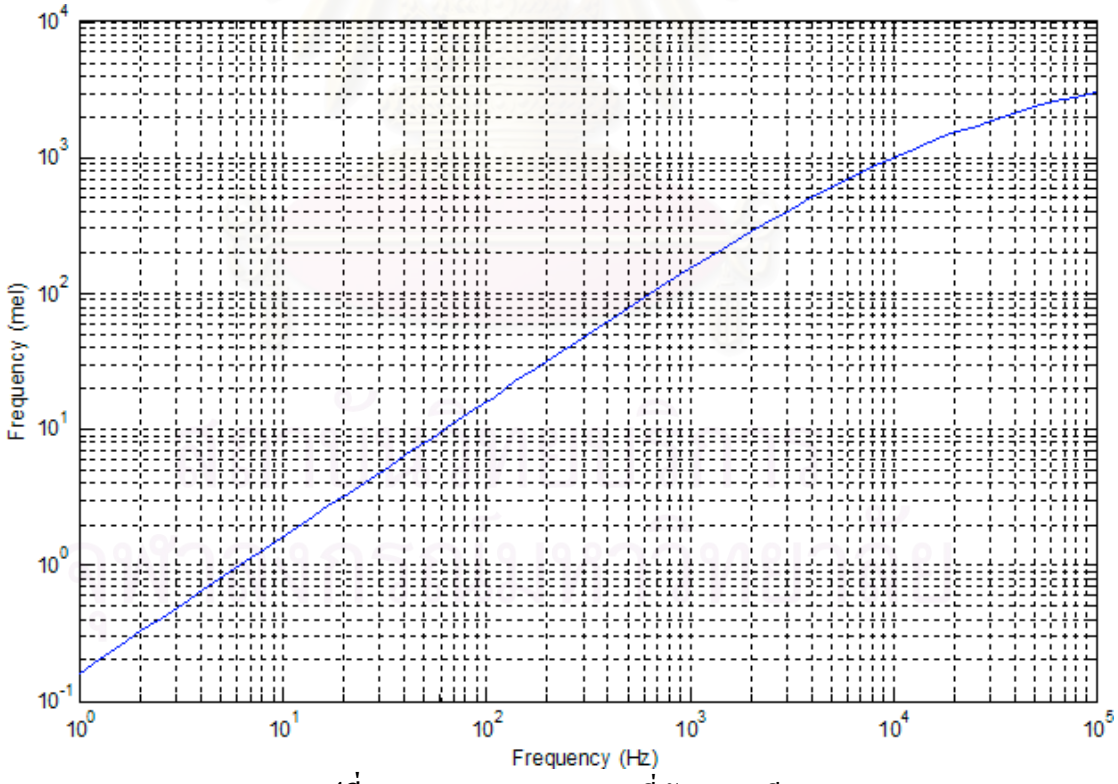
รูปที่ 2.2 ฟังก์ชันกรอบตัวยก

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

2.3.6 สัมประสิทธิ์เมลฟรีเควันซีเซปสตรอล (Mel Frequency Cepstral Coefficient, MFCC)

สัมประสิทธิ์เมลฟรีเควันซีเซปสตรอล คำนวณได้จากการวิเคราะห์โดยไม่ใช้พารามิเตอร์ (Non-Parametric Analysis) คือ ไม่มีการสร้างแบบจำลองสัญญาณเพื่อปรับค่าพารามิเตอร์ในแบบจำลองให้ได้สัญญาณที่เราต้องการ แต่เป็นการนำสัญญาณผ่านวงจรกรองแบบผ่านแถบความถี่ (Band Pass Filter) หลายวงจร โดยแต่ละวงจรกรองมีช่วงความถี่ที่ผ่านได้แตกต่างกันดังแสดงในรูปที่ 2.3 และวงจรกรองที่ใช้เป็นชุดวงจรกรองแบบดิจิตอล (Digital Filter Bank) ที่ใช้วิเคราะห์สเปกตรัมของสัญญาณเสียงพูด โดยเลียนแบบตามการได้ยินของมนุษย์ สามารถทำได้หลายวิธีขึ้นอยู่กับกรรมวิธีที่นำมาใช้กับแกนความถี่ได้แก่ (Tolba and O'Shaughnessy, 1998) (Tuzun, Demirekler and Nakiboglu, 1994)

- Uniform Spacing (Fourier Transform)
- Exponential Spacing (Wavelet Transform)
- Perceptually-Derived Spacing (Mel Scale หรือ Bark Scale)



รูปที่ 2.3 Mel Scale ของความถี่สัญญาณเสียง

ในวิทยานิพนธ์นี้ได้เลือกใช้กรรมวิธี Perceptually-derived Spacing หรือเรียกอีกชื่อหนึ่งว่า Mel Scale เป็นการเปลี่ยนแกนความถี่ของสัญญาณเสียง (Acoustic Frequency) ไปเป็นแกนความถี่ของการได้ยิน (Perceptual Frequency) จึงจะกล่าวถึงแต่กรรมวิธีนี้เท่านั้น Mel Scale ใช้หลักการ Mapping จากการประมาณความถี่ช่วงที่เป็นเชิงเส้น (Linear Frequency Scale) เป็นความถี่ที่ไม่เชิงเส้น (Nonlinear Frequency Scale) ตามลักษณะการได้ยินของมนุษย์ดังแสดงในสมการที่ (2.49) (Tolba and O'Shaughnessy, 1998)

$$mel(f) = 1127 \ln\left(1 + \frac{f}{7000}\right) \tag{2.49}$$

เมื่อ f คือช่วงความถี่ที่ยังเป็นเชิงเส้น จากรูปที่ 2.3 เห็นได้ว่าลักษณะการได้ยินของมนุษย์ในช่วงที่เป็นเชิงเส้น (Linear Scale) คือประมาณช่วงความถี่ตั้งแต่ 0 ถึง 1000 Hz และช่วงที่ไม่เป็นเชิงเส้น (Logarithm Scale) คือช่วงความถี่ตั้งแต่ 1000 Hz ขึ้นไป แบนด์วิดท์ของวงจรกรองสามารถคำนวณได้จากสมการที่ (2.50) ดังนี้

$$BW_{critical} = 25 + 75\left(1 + 1.4\left(\frac{f}{1000}\right)^2\right)^{0.69} \tag{2.50}$$

วิธีการคำนวณจากชุดวงจรกรอง (Filter Bank Model) ที่ง่ายที่สุดและให้ผลที่มีประสิทธิภาพดีคือการแปลงฟูริเยร์ (Fourier Transform) ของสัญญาณและหาข้อมูลออกของชุดวงจรกรองของแต่ละวงจร (X_n) จากความถี่ที่ใส่เข้าไป การคำนวณหาข้อมูลออกของชุดวงจรกรองสามารถคำนวณได้ตามสมการที่ (2.51)

$$X_N = \frac{1}{N_{FB}} \prod_{k=1}^{N_{FB}} w_{FB}(k) X(f + df(f,k)) \tag{2.51}$$

เมื่อ N_{FB} คือจำนวนของตัวอย่างที่ใช้ในการหาค่าเฉลี่ย

w_{FB} คือฟังก์ชันการถ่วงน้ำหนัก

$df(f,k)$ คือฟังก์ชันที่อธิบายถึงความถี่ที่อยู่ใกล้เคียงกับความถี่ f เพื่อนำมาใช้คำนวณหาค่าเฉลี่ย

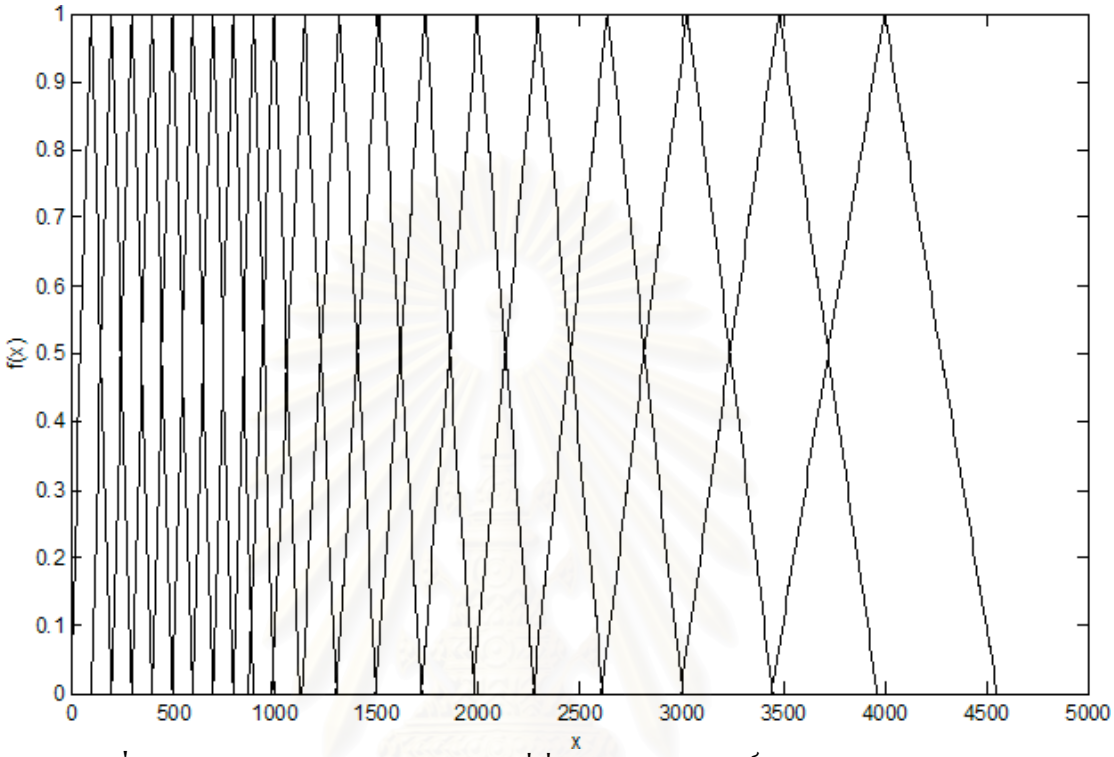
ในการคำนวณหาค่าสัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอลใช้ค่าข้อมูลออกที่ได้จากวงจรกรองแถบผ่าน N วงจรเมื่อใส่ค่าลอการิทึมของขนาดสเปกตรัมของสัญญาณเสียงเข้าไป เพื่อประมาณความถี่ตอบสนองของ Basilar Membrane ใน Cochlea ที่อยู่ในหูชั้นใน วงจรกรองที่นำมาใช้นี้เป็นวงจรกรองแถบผ่านแบบสามเหลี่ยมดังแสดงในรูปที่ 2.4 เพราะฉะนั้นการคำนวณหาค่าสัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอล (C_n) สามารถหาได้จากสมการดังต่อไปนี้ (Tolba and O'Shaughnessy, 1988)

$$C_n = \prod_{k=1}^{N} X_k \cos\left(\frac{pn}{N}\left(k - \frac{1}{2}\right)\right) \text{ เมื่อ } n = 1, 2, \mathrm{K}, M \tag{2.52}$$

เมื่อ M คือจำนวนของสัมประสิทธิ์เซปสตรอล

N คืออันดับในการวิเคราะห์

X_k คือค่าลอการิทึมของพลังงานที่ค่าข้อมูลออกของวงจรกรอง k วงจร เมื่อ $k = 1, 2, \mathrm{K}, N$



รูปที่ 2.4 วงจรกรองแบบผ่านแถบความถี่ที่ให้ค่าสัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอล

สัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอลที่ได้จะถูกคูณกับฟังก์ชันกรอบตัวยก (Lifter) ก่อนนำไปใช้ โดยใช้ฟังก์ชันตัวยกเป็นสมการ (2.48) เช่นเดียวกับสัมประสิทธิ์เซปสตรอล

2.3.7 สัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอลของค่าอัตสหสัมพันธ์ที่มีความคงทน (Autocorrelation Mel Frequency Cepstral Coefficient, AMFCC)

วิทยานิพนธ์นี้จะใช้สัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอลของค่าอัตสหสัมพันธ์ที่มีความคงทนแทนการใช้สัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอลของสัญญาณเสียงพูด เนื่องจากค่าอัตสหสัมพันธ์ที่มีความคงทนไม่เปลี่ยนแปลงเมื่อมีสัญญาณรบกวนเกาส์เซียนขาวค่าเฉลี่ยศูนย์เพิ่มเข้ามาทำให้สัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอลของค่าอัตสหสัมพันธ์ที่มีความคงทน ไม่เปลี่ยนแปลงเช่นกัน ดังนั้นเมื่อนำมาใช้ในระบบรู้จำเสียงพูดจะมีความคงทนต่อสัญญาณรบกวนมากขึ้น

2.4 การวิเคราะห์ค่าลักษณะสำคัญเมื่อมีสัญญาณรบกวน

2.4.1 สัมประสิทธิ์การประมาณพันธะเชิงเส้นและสัมประสิทธิ์แอลพีแอลอี

เมื่อนำค่าอัตสหสัมพันธ์มาใช้หาสัมประสิทธิ์การประมาณพันธะเชิงเส้น โดยแทนค่าลงในสมการ (2.15) จะได้

$$\begin{bmatrix} r_{ss}(1) \\ r_{ss}(2) \\ \vdots \\ r_{ss}(p) \end{bmatrix} = \begin{bmatrix} r_{ss}(0)+h & r_{ss}(1) & \cdots & r_{ss}(p-1) \\ r_{ss}(1) & r_{ss}(0)+h & \cdots & r_{ss}(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ r_{ss}(p-1) & r_{ss}(p-2) & \cdots & r_{ss}(0)+h \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} \tag{2.53}$$

ดังนั้น $$a = R_{ss}^{-1} r_{ss} = (R_{ss} + hI)^{-1} R_{ss} a \tag{2.54}$$

เนื่องจาก R_{ss} เป็นเมตริกซ์สมมาตรแบบ Toeplitz จึงสามารถเขียน R_{ss} ได้ในรูปของ

$$R_{ss} = P D P^{-1} \tag{2.55}$$

โดย D เป็นเมตริกซ์ทแยงมุม ที่มีค่าเจาะจง (Eigenvalue) l_i เป็นสมาชิกในแนวทแยงมุม

ทำให้ $$R_{ss} + hI = P(D + hI)P^{-1} \tag{2.56}$$

ดังนั้น $$a = R_{ss}^{-1} R_{ss} a = P \begin{bmatrix} \frac{l_1}{l_1 + h} & & \\ & \ddots & \\ & & \frac{l_k}{l_k + h} \end{bmatrix} P^{-1} a \tag{2.57}$$

จาก (2.57) เห็นได้ชัดว่าค่าสัมประสิทธิ์การประมาณพันธะเชิงเส้นมีขนาดเล็กลงเมื่อมีสัญญาณรบกวนเข้ามามากขึ้น เป็นเหตุให้สัมประสิทธิ์การประมาณพันธะเชิงเส้นไม่มีความคงทนต่อสัญญาณรบกวน

สำหรับสัมประสิทธิ์แอลพีแอลอี แม้เมตริกซ์อัตสหสัมพันธ์ที่ใช้คำนวณหาค่าสัมประสิทธิ์จะไม่เป็นเมตริกซ์สมมาตร แต่สมาชิกในแนวทแยงมุมมีเทอมของ $r_{ss}(0)$ รวมอยู่ จึงมีเทอมของ h รวมอยู่ด้วย ส่งผลให้ค่าสัมประสิทธิ์แอลพีแอลอีเปลี่ยนแปลงมากขึ้นเมื่อมีกำลังของสัญญาณรบกวนมากขึ้น

2.4.2 สัมประสิทธิ์เซปสตรอลและสัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอล

เนื่องจากสัมประสิทธิ์เซปสตรอลหาจากค่าสัมประสิทธิ์การประมาณพันธะเชิงเส้นในสมการ (2.46) และ (2.47) ดังนั้นจึงมีค่าสัมประสิทธิ์เล็กลงเมื่อมีสัญญาณรบกวนเข้ามามากขึ้น เป็นเหตุให้

สัมประสิทธิ์เซปสตรอลไม่มีความคงทนต่อสัญญาณรบกวนเช่นกัน สำหรับสัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอลก็เช่นกัน เนื่องจากการเพิ่มเข้ามาของกำลังของสัญญาณรบกวนตลอดช่วงความถี่ทำให้การหาค่าสัมประสิทธิ์ในขั้นตอนการแปลงฟูริเยร์ผกผัน (Inverse Fourier Transform) ของสมการ (2.52) มีค่าเปลี่ยนไปตามกำลังของสัญญาณรบกวน ทำให้สัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอลไม่มีความคงทนต่อสัญญาณรบกวน

2.4.3 สัมประสิทธิ์แอลพีแอลอีของค่าอัตสหสัมพันธ์ที่มีความคงทนและสัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอลของค่าอัตสหสัมพันธ์ที่มีความคงทน

เนื่องจากค่าลักษณะสำคัญทั้งสองเป็นการนำค่าอัตสหสัมพันธ์ที่มีความคงทนเป็นสัญญาณเข้าแทนสัญญาณเสียงพูด ซึ่งค่าอัตสหสัมพันธ์ที่มีความคงทนไม่เปลี่ยนแปลงเมื่อมีสัญญาณรบกวนเกาส์เซียนขาวค่าเฉลี่ยศูนย์เพิ่มเข้ามา ดังนั้นค่าลักษณะสำคัญทั้งสองจึงมีความคงทนต่อสัญญาณรบกวนเช่นกัน

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 3

ขั้นตอนวิธีในการดำเนินงานวิจัย

ในบทนี้จะได้กล่าวถึงรายละเอียดขั้นตอนต่างๆที่ใช้ในการดำเนินงานวิจัย ประกอบด้วย การกำหนดชุดคำศัพท์ การเตรียมตัวอย่างข้อมูลเสียงพูด การประมวลผลสัญญาณเบื้องต้น การหาค่าลักษณะสำคัญ วิธีการฝึกฝนและการทดสอบอัตราการรู้จำค่าลักษณะสำคัญในงานวิจัย

3.1 การกำหนดชุดคำศัพท์

เนื่องจากเสียงพูดภาษาไทย มีเสียงวรรณยุกต์ถึง 5 ระดับ และเพื่อให้แน่ใจว่าค่าลักษณะสำคัญที่เสนอในวิทยานิพนธ์ใช้กับภาษาไทยได้ดี จึงเลือกชุดคำศัพท์พยางค์เดียวให้ครอบคลุมกลุ่มเสียงวรรณยุกต์ทั้งหมด

ชุดคำศัพท์ที่เลือกมาใช้ จะจัดแบ่งเป็น 5 กลุ่ม ตามกลุ่มเสียงวรรณยุกต์ทั้ง 5 คือ สามัญ เอก โท ตรี และจัตวา ในแต่ละกลุ่มเสียงจะมีชุดคำศัพท์พยางค์เดียว 8 คำ รวมชุดคำศัพท์ทั้งหมดเป็น 40 คำ ดังตารางที่ 3.1

ตารางที่ 3.1 ชุดคำศัพท์ทั้งหมด

กลุ่มเสียง	คำศัพท์							
สามัญ	กิน (108,2)	จันทร์ (108,2)	เดิน (108,2)	ตา (107,2)	เตียง (108,2)	เทียน (108,2)	นอน (109,2)	มือ (108,2)
เอก	ไก่ (106,2)	เจ็ด (108,2)	ปาก (108,2)	เป็ด (108,2)	แปด (108,2)	สี่ (108,2)	หก (108,2)	หนึ่ง (108,2)
โท	กล้วย (108,2)	เก้า (108,2)	แก้ว (108,2)	นั่ง (108,2)	วิ่ง (108,2)	ส้ม (108,2)	หน้า (108,2)	ห้า (108,2)
ตรี	ซ้าย (104,6)	โต๊ะ (108,2)	ทุก (90,2)	นก (108,2)	น้ำ (108,2)	พุธ (108,2)	ไม้ (117,2)	รถ (110,0)
จัตวา	ขวา (108,2)	ศูนย์ (104,2)	สอง (108,2)	สาม (108,2)	เสาร์ (107,2)	เสือ (108,2)	หัว (108,2)	หู (108,2)

3.2 การเตรียมตัวอย่างข้อมูลเสียงพูด

ตัวอย่างข้อมูลเสียงพูดที่นำมาใช้ จะเป็นข้อมูลเสียงพูดของห้องปฏิบัติการวิจัยประมวลสัญญาณดิจิตอล ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ซึ่งได้รับการควบคุมสภาพแวดล้อมขณะทำการบันทึกเสียง ให้คล้ายคลึงกับสภาพแวดล้อมของสถานที่ทำงานทั่วไปและมีเสียงรบกวนน้อยที่สุด โดยเสียงพูดที่บันทึกไว้จะจัดเก็บด้วยตัวอย่างขนาด 16 บิต และมีอัตราการชักตัวอย่าง 11 kHz เนื่องจากเสียงพูดของมนุษย์จะอยู่ในช่วงไม่เกิน 5 kHz จึงต้องใช้อัตราการชักตัวอย่างที่สูงกว่าสองเท่าของความถี่ 5 kHz สำหรับเสียงที่บันทึกจะเป็นเสียงของผู้ที่ใช้ภาษาไทยสำเนียงกรุงเทพเป็นภาษาพูดทั่วไป มีอายุระหว่าง 18 – 25 ปี และเป็นผู้ที่มีการออกเสียงเป็นปกติ ตรงตามหลักการออกเสียงพูดภาษาไทย

หลังจากเลือกชุดคำศัพท์ทั้งหมดได้แล้ว จึงเลือกเสียงพูดในฐานข้อมูลที่ตรงกับชุดคำศัพท์มาคำศัพท์ละประมาณ 110 เสียง เป็นเสียงคละกันทั้งชายและหญิง จำนวนเสียงพูดของผู้ชายและจำนวนเสียงพูดของผู้หญิงสำหรับแต่ละคำศัพท์ แสดงในตารางที่ 3.1 โดยเป็นกลุ่มตัวเลขที่ใต้คำศัพท์แต่ละคำ ตัวเลขคู่ลำดับแรกแสดงจำนวนเสียงพูดของผู้ชาย ตัวเลขคู่ลำดับที่สองแสดงจำนวนเสียงพูดของผู้หญิง

3.2.1 การตัดพยางค์

เนื่องจากตัวอย่างเสียงพูดในฐานข้อมูลเสียงพูดที่เป็นพยางค์เดียว มีไม่ครบตามชุดคำศัพท์ที่ได้เลือกไว้ จึงนำตัวอย่างเสียงพูดที่เป็นสองพยางค์มาทำการตัดพยางค์เพื่อให้ได้ตัวอย่างเสียงพูดครบตามชุดคำศัพท์ทั้งหมด

หลักการตัดพยางค์ จะใช้วิธีการพิจารณาจากค่าระดับพลังงานของเสียงพูด $E\langle m\rangle$ ของสัญญาณเสียงพูดที่แบ่งเป็นกรอบเสียงพูด (Speech Frame) กรอบละ 100 ตัวอย่าง (9 มิลลิวินาที) และเลื่อนกรอบครั้งละ 1 คือ

$$E\langle m\rangle = \prod_{k=m}^{m+99} s^2\langle k\rangle \qquad (3.1)$$

กำหนดจุดเปลี่ยนระดับพลังงาน (Energy Level Threshold) ที่ระดับพลังงาน ร้อยละ 20 ของค่าเฉลี่ยของระดับพลังงานของสัญญาณเสียงพูดทั้งคำเพื่อให้เหมาะสมกับระดับพลังงานของแต่ละเสียงพูด และกำหนดจำนวนกรอบขั้นต่ำที่ใช้ยืนยันการเปลี่ยนค่าระดับพลังงานที่ 200 กรอบ

ขั้นตอนที่ 1

คือการหาจุดเริ่มต้นของพยางค์ที่หนึ่ง จึงเริ่มการค้นหาตั้งแต่กรอบที่หนึ่งจนกระทั่งพบกรอบแรกซึ่ง มีระดับพลังงานของกรอบสูงกว่าจุดเปลี่ยนระดับพลังงานเป็นช่วงเวลา 200 กรอบ

ขั้นตอนที่ 2

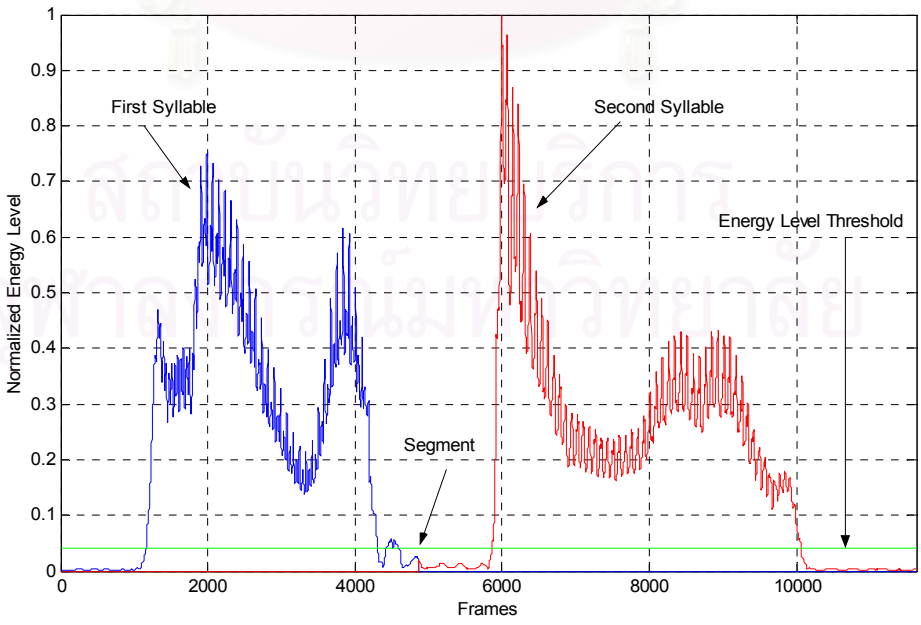
คือการหาจุดสิ้นสุดของพยางค์ที่หนึ่ง เป็นการหากรอบถัดมากรอบแรกซึ่ง มีระดับพลังงานต่ำกว่าจุดเปลี่ยนระดับพลังงานเป็นช่วงเวลา 200 กรอบ

ขั้นตอนที่ 3

เป็นการทำซ้ำขั้นตอนที่ 1 และ 2 จนครบทุกกรอบ เพื่อหาจุดเริ่มต้นและจุดสิ้นสุดของพยางค์ทั้งหมดในเสียงพูด

ขั้นตอนที่ 4

พิจารณาเลือกจุดเริ่มต้นและสิ้นสุดของพยางค์ที่แท้จริง เนื่องจากหลายๆครั้งพบว่า มีจำนวนจุดเริ่มต้นและจุดสิ้นสุดของพยางค์เป็นจำนวนมากกว่าที่ควรจะเป็น (มากกว่า 2 พยางค์) วิธีการเลือกจะพิจารณาจาก ค่าเฉลี่ยของระดับพลังงานในกรอบที่อยู่ระหว่างสองพยางค์ใดๆในขั้นตอนก่อนหน้านี้ ว่าช่วงใดมีค่าเฉลี่ยระดับพลังงานต่ำสุด ช่วงดังกล่าวควรจะเป็นช่วงรอยต่อระหว่างสองพยางค์ที่แท้จริง เมื่อหาช่วงดังกล่าวพบแล้ว จะแก้ไขให้จุดสิ้นสุดของพยางค์ที่หนึ่งและจุดเริ่มต้นของพยางค์ที่สองเป็นจุดเดียวกัน โดยใช้กรอบที่แบ่งครึ่งค่าเฉลี่ยของระดับพลังงานที่อยู่ในช่วงรอยต่อระหว่างสองพยางค์ให้เป็นจุดแบ่งพยางค์ ตัวอย่างของค่าระดับพลังงานและการตัดพยางค์แสดงในรูปที่ 3.1



รูปที่ 3.1 แผนภูมิเส้นระดับพลังงานของเสียงพูด “มือซ้าย”

ขั้นตอนที่ 5

เป็นขั้นตอนตรวจสอบเสียงพูดที่ได้จากการตัดพยางค์ในขั้นตอนที่ผ่าน โดยใช้การเปิดฟังเสียงพูดที่ได้ทุกเสียง หากพบว่าเสียงพูดใดมีการตัดพยางค์ไม่ถูกต้อง จะแก้ไขเสียงพูดแต่ละเสียงด้วยตนเอง โดยใช้โปรแกรมบันทึกเสียง GoldWave เวอร์ชัน 4.26 บนระบบปฏิบัติการ Windows XP

3.2.2 การเพิ่มเสียงสัญญาณรบกวน

ตัวอย่างเสียงพูดที่เตรียมไว้ในขั้นตอนที่ผ่านมา จัดว่าเป็นสัญญาณเสียงพูดที่ไม่มีสัญญาณรบกวน ยังขาดสัญญาณเสียงพูดที่มีสัญญาณรบกวนเกาส์เซียนขาวค่าเฉลี่ยศูนย์เพิ่มเข้ามาสำหรับใช้ในการทดสอบความคงทนของค่าลักษณะสำคัญ ในวิทยานิพนธ์นี้จะทดสอบความคงทนของค่าลักษณะสำคัญที่ค่าอัตราส่วนสัญญาณต่อสัญญาณรบกวน (Signal to Noise Ratio) 8 ค่า คือ 0 dB , 5 dB , 10 dB , 15 dB , 20 dB , 25 dB , 30 dB และไม่มีสัญญาณรบกวน

เพื่อให้สามารถเปรียบเทียบความคงทนของ ค่าลักษณะสำคัญแบบต่างๆที่ใช้จำนวนอันดับต่างกันได้ จึงต้องเตรียมตัวอย่างเสียงพูดของแต่ละชุดคำศัพท์ในทุกๆอัตราส่วนสัญญาณต่อสัญญาณรบกวนที่ใช้ในการทดสอบ บันทึกเก็บไว้บนฮาร์ดดิสก์ การเพิ่มสัญญาณรบกวนเกาส์เซียนขาวค่าเฉลี่ยศูนย์เพิ่มเข้ามาและบันทึกเก็บลงฮาร์ดดิสก์ จะทำผ่านโปรแกรม MATLAB เวอร์ชัน 6.1 บนระบบปฏิบัติการ Windows XP

3.3 การประมวลผลสัญญาณเบื้องต้น

ขั้นตอนนี้เป็นการเตรียมสัญญาณเสียงพูดที่ได้เตรียมไว้ในขั้นตอนก่อนหน้านี้ ก่อนจะนำไปหาค่าลักษณะสำคัญต่อไป ประกอบด้วย 2 ขั้นตอน คือ

3.3.1 กรรมวิธีเน้นล่วงหน้า

สัญญาณเสียงพูดที่เตรียมไว้จะนำมาผ่านกรรมวิธีเน้นล่วงหน้า เพื่อยกระดับสัญญาณ เสียงพูดต่อสัญญาณรบกวนให้สูงขึ้น โดยกำหนดให้ค่า $a = 0.95$

3.3.2 กรรมวิธีวางกรอบขนาดสัญญาณ

เนื่องจากสัญญาณเสียงพูดมีการเปลี่ยนแปลงตามเวลา (Time Variant) และไม่เสถียร (Unstable) ดังนั้นในการประยุกต์ใช้งานกรรมวิธีประมวลผลสัญญาณเชิงเลขกับสัญญาณเสียงพูด

จะแบ่งสัญญาณเสียงพูดออกเป็นช่วงย่อยๆ (Rabiner and Levinson 1981; Furui, 1989) เรียกว่า “กรอบเสียงพูด” (Speech Frame) โดยแต่ละกรอบเสียงพูดจะมีความยาวประมาณ 10-40 มิลลิวินาที ซึ่งถือได้ว่าสัญญาณเสียงพูดในแต่ละกรอบเสียงพูดมีความเสถียรและไม่แปรเปลี่ยนตามเวลา (Time Invariant) จากนั้นจึงสามารถทำการประมวลผลสัญญาณเชิงเลขกับสัญญาณในแต่ละกรอบเสียงพูดได้ ฟังก์ชันกรอบที่ใช้ในวิทยานิพนธ์นี้เลือกใช้ฟังก์ชันกรอบชนิด Hamming Window ที่มีจำนวนของข้อมูลในกรอบเสียงพูดเท่ากับ 512 ช่วงประสิทธิผลของฟังก์ชันกรอบมีค่าเท่ากับ 256 และเลื่อนฟังก์ชันกรอบครั้งละ 64 (เอกฤทธิ์ มณีน้อย, 2541) จึงทำให้แต่ละกรอบมีข้อมูลของสัญญาณเสียงพูดเท่ากับ 256 ค่า หรือคิดเป็นระยะเวลาจริง 23 มิลลิวินาที

3.4 การหาค่าลักษณะสำคัญ

ค่าลักษณะสำคัญของสัญญาณเสียงพูด ที่ใช้ในวิทยานิพนธ์นี้มี 4 แบบ คือ สัมประสิทธิ์การประมาณพันธะเชิงเส้น (LPC) สัมประสิทธิ์แอลพีแอลอีของค่าอัตสหสัมพันธ์ที่มีความคงทน (ALPLEC) สัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอล (MFCC) และ สัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอลของค่าอัตสหสัมพันธ์ที่มีความคงทน (AMFCC) ซึ่งมีรายละเอียดของการคำนวณหาดังนี้

3.4.1 สัมประสิทธิ์การประมาณพันธะเชิงเส้น (LPC)

การคำนวณหาค่าสัมประสิทธิ์ จะใช้วิธีอัตสหสัมพันธ์ และใช้วิธีการวนซ้ำของ Levinson-Durbin ในการแก้สมการ โดยหาค่าสัมประสิทธิ์ที่มีจำนวนอันดับเป็น 16 และ 32 ตามลำดับ

3.4.2 สัมประสิทธิ์แอลพีแอลอีของค่าอัตสหสัมพันธ์ที่มีความคงทน (ALPLEC)

การคำนวณหาค่าสัมประสิทธิ์ จะใช้ค่าอัตสหสัมพันธ์ของสัญญาณเสียงพูดที่มีความคงทนจากสมการ (2.44) เป็นสัญญาณเข้าแทนสัญญาณเสียงพูด และใช้วิธีอัตสหสัมพันธ์ แต่เนื่องจากเมตริกซ์อัตสหสัมพันธ์ไม่อยู่ในรูปแบบ Toeplitz จึงใช้วิธีการวนซ้ำของ Levinson-Durbin ในการแก้สมการไม่ได้ ต้องใช้วิธีแก้สมการเชิงเส้นทั่วไป และหาค่าสัมประสิทธิ์ที่มีจำนวนอันดับเป็น 16 และ 32 ตามลำดับ ค่าลักษณะสำคัญแบบนี้จะนำมาเปรียบเทียบความคงทนกับสัมประสิทธิ์การประมาณพันธะเชิงเส้นที่มีจำนวนอันดับเท่ากัน

3.4.3 สัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอล (MFCC)

การคำนวณหาค่าสัมประสิทธิ์ จะใช้การแปลงฟูริเยร์ (Fourier Transform) ของสัญญาณเสียงพูดแล้วนำไปเข้าชุดวงจรกรองรูปสามเหลี่ยมดังรูปที่ 2.4 กำหนดจำนวนสัมประสิทธิ์เป็น 16 และ 32 จึงใช้จำนวนวงจรกรองรูปสามเหลี่ยมเป็นจำนวน 16 และ 32 วงจร แล้วคำนวณหาค่าลอการิทึมของสัญญาณขาออกของแต่ละวงจร ได้เป็น X_k เมื่อ $k = 1, 2, \mathrm{K}, N$ ตามลำดับ และนำ X_k ไปคำนวณหาค่าสัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอลตามสมการ (2.52)

3.4.4 สัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอลของค่าอัตสหสัมพันธ์ที่มีความคงทน (AMFCC)

การคำนวณหาค่าสัมประสิทธิ์จะใช้การแปลงฟูริเยร์ (Fourier Transform) ของค่าอัตสหสัมพันธ์ของสัญญาณเสียงพูดที่มีความคงทนจากสมการ (2.44) แล้วนำไปเข้าชุดวงจรกรองรูปสามเหลี่ยมดังรูปที่ 2.4 กำหนดจำนวนสัมประสิทธิ์เป็น 16 และ 32 จึงใช้จำนวนวงจรกรองรูปสามเหลี่ยมเป็นจำนวน 16 และ 32 วงจร แล้วคำนวณหาค่าลอการิทึมของสัญญาณขาออกของแต่ละวงจร ได้เป็น X_k เมื่อ $k = 1, 2, \mathrm{K}, N$ ตามลำดับ และนำ X_k ไปคำนวณหาค่าสัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอลตามสมการ (2.52) ค่าลักษณะสำคัญแบบนี้จะนำมาเปรียบเทียบความคงทนกับสัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอลที่มีจำนวนอันดับเท่ากัน

3.5 วิธีการฝึกฝนและการทดสอบอัตราการรู้จำค่าลักษณะสำคัญ

การฝึกฝนและทดสอบค่าลักษณะสำคัญสำหรับใช้ในการทดสอบการรู้จำ ใช้ระบบรู้จำแบบจำลองฮิดเดนมาร์คอฟ ประเภทแบบจำลองซ้าย-ขวา (Left-Right Model) หรือแบบจำลองอนุกรม (Serial Model) แบบต่อเนื่อง ที่มี 7 สถานะและจำนวน Gaussian Mixture เป็น 2 แบบจำลองดังกล่าวถูกพัฒนาขึ้นโดยคุณเอกฤทธิ์ มณีน้อย (เอกฤทธิ์ มณีน้อย, 2541)

3.5.1 การฝึกฝนระบบ

การฝึกฝนเสียงพูดคำศัพท์แต่ละคำ จะนำเสียงพูดคำศัพท์คำเดียวกันที่ไม่มีสัญญาณรบกวนของผู้พูดในกลุ่มที่ได้กำหนดไว้ล่วงหน้าแล้วในตารางที่ 3.2 มาฝึกฝนกับแบบจำลองฮิดเดนมาร์คอฟเพื่อสร้างชุดรูปร่องต้นแบบอ้างอิง (Word Reference Templates) ของเสียงพูดคำศัพท์ของผู้พูดในกลุ่มดังกล่าว

การจัดแบ่งกลุ่มของผู้พูดในเสียงพูดคำศัพท์แต่ละคำสำหรับการฝึกฝนออกเป็น 3 กลุ่ม แต่ละกลุ่มมีจำนวนเสียงพูดเท่ากัน ดังนั้นหลังจากเสร็จสิ้นขั้นตอนการฝึกฝนระบบ จะมีชุดรูปร่องต้นแบบอ้างอิงของเสียงพูดคำศัพท์แต่ละคำ เป็นจำนวน 3 ชุด

3.5.2 การทดสอบอัตราการรู้จำค่าลักษณะสำคัญ

การทดสอบอัตราการรู้จำค่าลักษณะสำคัญในแต่ละแบบที่มีจำนวนอันดับสัมประสิทธิ์ต่างกันของแต่ละคำศัพท์ จะนำค่าลักษณะสำคัญเดียวกันของเสียงพูดคำศัพท์คำเดียวกันที่มีค่าอัตราส่วนสัญญาณต่อสัญญาณรบกวนเท่ากันของผู้พูดในกลุ่มที่ได้กำหนดไว้ล่วงหน้าแล้วในตารางที่ 3.2 มาทดสอบกับแบบจำลองฮิดเดนมาร์คอฟ

เนื่องจากมีชุดรูปร่องต้นแบบอ้างอิงของเสียงพูดคำศัพท์แต่ละคำ เป็นจำนวน 3 ชุด จึงแบ่งกลุ่มของผู้พูดในเสียงพูดคำศัพท์แต่ละคำสำหรับการทดสอบออกเป็น 3 กลุ่ม แต่ละกลุ่มมีจำนวนเสียงพูดเท่ากัน ดังนั้นหลังจากเสร็จสิ้นขั้นตอนการทดสอบระบบ จะมีผลการรู้จำเสียงพูดคำศัพท์แต่ละคำที่ค่าอัตราส่วนสัญญาณต่อสัญญาณรบกวนเท่ากันของแต่ละค่าลักษณะสำคัญ เป็นจำนวน 3 ชุด

3.5.3 การจัดแบ่งกลุ่มของผู้พูดสำหรับการฝึกฝนและทดสอบ

เพื่อให้ได้ผลการทดสอบรู้จำของค่าลักษณะสำคัญที่ไม่ขึ้นกับผู้พูด จึงจัดแบ่งกลุ่มของผู้พูดออกเป็น 3 กลุ่ม โดยชุดรูปร่องต้นแบบอ้างอิงของเสียงพูดจะไม่ถูกนำมาใช้ทดสอบกับกลุ่มผู้พูดที่ใช้สร้างชุดรูปร่องต้นแบบนั้น การจัดแบ่งกลุ่มของผู้พูดสำหรับการฝึกฝนและทดสอบแสดงในตารางที่ 3.2

ตารางที่ 3.2 การจัดแบ่งกลุ่มของผู้พูดสำหรับการฝึกฝนและทดสอบ

กลุ่มฝึกฝน	กลุ่มทดสอบ
กลุ่ม 1	กลุ่ม 2 และ กลุ่ม 3
กลุ่ม 2	กลุ่ม 1 และ กลุ่ม 3
กลุ่ม 3	กลุ่ม 1 และ กลุ่ม 2

บทที่ 4

ผลการทดลองและการวิเคราะห์ผลการทดลอง

ในบทนี้ จะกล่าวถึงรายละเอียดของผลการรู้จำเสียงพูดภาษาไทย ซึ่งแสดงผลการวิจัยที่ได้ในแต่ละกรณี และวิเคราะห์ผลที่ได้จากการทดสอบ โดยทดสอบกับกลุ่มเสียงวรรณยุกต์ภาษาไทยทั้งห้ากลุ่มคือกลุ่มเสียงสามัญ เอก โท ตรี และจัตวา ในกลุ่มคำศัพท์ที่ได้เตรียมไว้ เพื่อทดสอบความสามารถในการรู้จำเสียงพูดภาษาไทยซึ่งมีวรรณยุกต์เป็นส่วนสำคัญของภาษาไทย

4.1 ขั้นตอนวิธีการรู้จำคำพูดภาษาไทย

ขั้นตอนวิธีการรู้จำเสียงพูดภาษาไทยสำหรับงานวิจัยนี้ ใช้แบบจำลองฮิดเดนมาร์คอฟแบบต่อเนื่อง จะนำชุดพารามิเตอร์ของแบบจำลองฮิดเดนมาร์คอฟสำหรับคำศัพท์แต่ละคำที่ได้จากการฝึกฝนมาใช้เป็นชุดรูปร่องต้นแบบอ้างอิงในการรู้จำ โดยแบ่งชุดพารามิเตอร์ของคำศัพท์แต่ละคำดังนี้ คือ ชุดพารามิเตอร์ของคำศัพท์จำแนกตามวรรณยุกต์ ซึ่งประกอบด้วย ชุดพารามิเตอร์ของคำศัพท์จำแนกตามวรรณยุกต์เสียงสามัญ ชุดพารามิเตอร์ของคำศัพท์จำแนกตามวรรณยุกต์เสียงเอก ชุดพารามิเตอร์ของคำศัพท์จำแนกตามวรรณยุกต์เสียงโท ชุดพารามิเตอร์ของคำศัพท์จำแนกตามวรรณยุกต์เสียงตรี และชุดพารามิเตอร์ของคำศัพท์จำแนกตามวรรณยุกต์เสียงจัตวา ตามลำดับ

ในระหว่างกระบวนการรู้จำเสียงพูดภาษาไทย จะทำการเปรียบเทียบเสียงพูดทดสอบกับชุดรูปร่องต้นแบบอ้างอิง โดยใช้ชุดพารามิเตอร์ของคำศัพท์วรรณยุกต์เดียวกันตรงกับคำศัพท์ของวรรณยุกต์ที่นำมาทดสอบ การเปรียบเทียบจะกระทำกับชุดรูปร่องต้นแบบอ้างอิงของคำศัพท์ทุกคำในชุดพารามิเตอร์ ผลการรู้จำจะได้ความน่าจะเป็นของแต่ละคำศัพท์ที่ได้รับการเปรียบเทียบ เมื่อทำการเปรียบเทียบทั้งชุดแล้ว จะเรียงลำดับความน่าจะเป็นจากสูงที่สุดไปหาต่ำที่สุด และคำศัพท์ที่รู้จำได้จะเป็นคำศัพท์ที่มีค่าความน่าจะเป็นสูงที่สุด

อัตราการรู้จำ (Recognition Rate) หมายถึงอัตราความถูกต้องในการรู้จำคำศัพท์แต่ละคำโดยคิดเป็นร้อยละของจำนวนเสียงพูดทั้งหมดที่นำมาทดสอบ การคำนวณค่าอัตราการรู้จำที่ใช้ในงานวิจัยนี้จะนำชุดเสียงพูดทั้งหมดมาแบ่งเป็นสามส่วนเท่ากัน โดยทำการฝึกฝนกับส่วนที่หนึ่งและทดสอบกับอีกสองส่วนที่เหลือ ทำการฝึกฝนและทดสอบแบบนี้สามครั้ง พร้อมทั้งบันทึกจำนวนครั้งที่ระบบรู้จำถูกต้องและไม่ถูกต้อง อัตราการรู้จำเฉลี่ยของเสียงพูดในกลุ่มเสียงวรรณยุกต์สามารถคำนวณได้จาก อัตราส่วนระหว่างจำนวนเสียงที่ระบบรู้จำถูกต้องทั้งสามชุดรวมกันต่อจำนวนเสียงพูดทั้งหมดสามชุดรวมกัน

4.2 ผลและการวิเคราะห์ผลการรู้จำเสียงพูดภาษาไทย

ผลอัตราการรู้จำเฉลี่ยของค่าลักษณะสำคัญจะแบ่งเป็น 2 กลุ่ม ตามลักษณะของค่าลักษณะสำคัญ เพื่อจะได้เปรียบเทียบความคงทนได้เหมาะสม

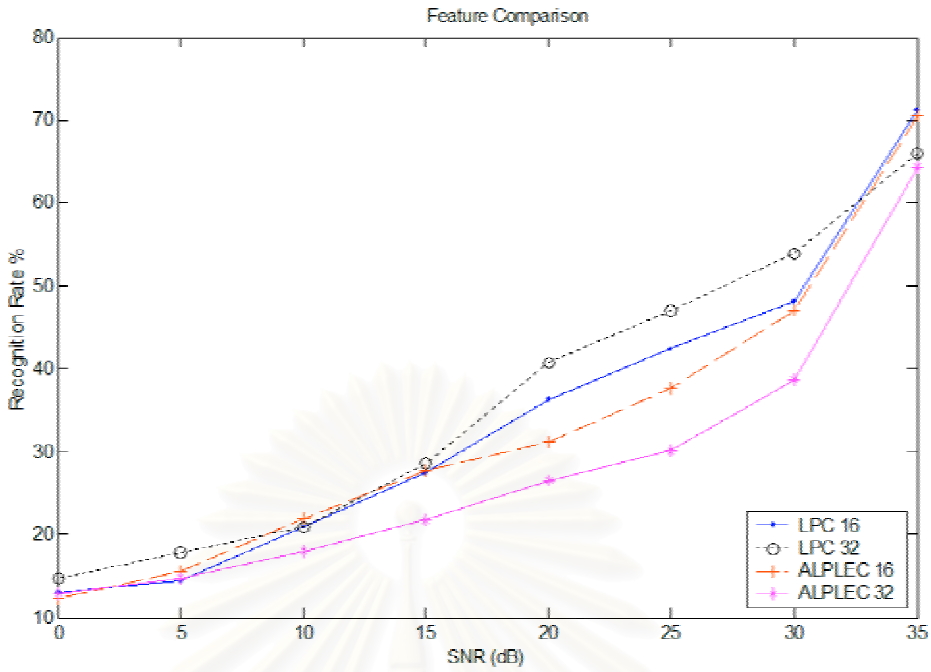
4.2.1 กลุ่มค่าลักษณะสำคัญแบบสัมประสิทธิ์การประมาณพันธะเชิงเส้น

4.2.1.1 อัตราการรู้จำเฉลี่ยของเสียงพูดกลุ่มเสียงวรรณยุกต์สามัญ

ผลอัตราการรู้จำเฉลี่ยของเสียงพูดกลุ่มเสียงวรรณยุกต์สามัญ ของค่าลักษณะสำคัญ คือ สัมประสิทธิ์การประมาณพันธะเชิงเส้น อันดับ 16 (LPC16) สัมประสิทธิ์การประมาณพันธะเชิงเส้น อันดับ 32 (LPC32) สัมประสิทธิ์แอลพีแอลอีของค่าอัตสหสัมพันธ์ที่มีความคงทน อันดับ 16 (ALPLEC16) และ สัมประสิทธิ์แอลพีแอลอีของค่าอัตสหสัมพันธ์ที่มีความคงทน อันดับ 32 (ALPLEC32) แสดงในตารางที่ 4.1 และนำมาเขียนเป็นกราฟในรูปที่ 4.1

ตารางที่ 4.1 ผลอัตราการรู้จำเฉลี่ยของ เสียงพูดกลุ่มเสียงวรรณยุกต์สามัญ

SNR	ค่าลักษณะสำคัญ			
	LPC16	ALPLEC16	LPC32	ALPLEC32
Clean	71.14	70.51	66.08	64.38
30 dB	48.07	46.99	53.86	38.75
25 dB	42.44	37.56	47.05	30.17
20 dB	36.25	31.08	40.68	26.53
15 dB	27.33	27.67	28.52	21.65
10 dB	20.80	21.88	20.91	17.90
5 dB	14.38	15.63	17.78	14.66
0 dB	13.07	12.33	14.72	12.84



รูปที่ 4.1 อัตราการรู้จำเฉลี่ยของ เสียงพูดกลุ่มเสียงวรรณยุกต์สามัญ

จากรูปที่ 4.1 จะเห็นว่าสัมประสิทธิ์แอลพีแอลอีของค่าอัตสหสัมพันธ์ที่มีความคงทน มีอัตราการรู้จำเฉลี่ยของเสียงพูดกลุ่มเสียงวรรณยุกต์สามัญน้อยกว่าสัมประสิทธิ์การประมาณพันธะเชิงเส้นในเกือบทุกค่าอัตราส่วนสัญญาณต่อสัญญาณรบกวน ส่วนในกรณีที่ให้ค่าอัตราการรู้จำมากกว่าก็คิดเป็นเพียงไม่เกินร้อยละ 2 เท่านั้น

นอกจากนี้จำนวนอันดับที่เพิ่มขึ้นของสัมประสิทธิ์แอลพีแอลอีของค่าอัตสหสัมพันธ์ที่มีความคงทนกลับทำให้อัตราการรู้จำเฉลี่ยลดลง เฉลี่ยร้อยละ 4.60 ทั้งนี้เพราะสัมประสิทธิ์แอลพีแอลอีที่จำนวนอันดับมากขึ้นให้ค่าความผิดพลาดทางสเปกตรัมมากกว่า ซึ่งตรงข้ามกับจำนวนอันดับที่เพิ่มขึ้นของสัมประสิทธิ์การประมาณพันธะเชิงเส้นที่การเพิ่มจำนวนอันดับทำให้อัตราการรู้จำเฉลี่ยเพิ่มขึ้น เฉลี่ยร้อยละ 2.02

เป็นที่น่าสังเกตว่าในกรณีที่ไม่มีสัญญาณรบกวน จำนวนอันดับของสัมประสิทธิ์ที่น้อยกว่ากลับให้ค่าอัตราการรู้จำเฉลี่ยมากกว่า ทั้งนี้เพราะสัมประสิทธิ์จำนวนมากขึ้นจะเป็นตัวแทนของข้อมูลในส่วนของสัญญาณรบกวนมากกว่าตัวสัญญาณจริงที่ต้องการ และในกรณีที่กำลังของสัญญาณรบกวนมีค่าเท่ากับกำลังของสัญญาณ จำนวนอันดับของสัมประสิทธิ์ที่มากขึ้นช่วยให้อัตราการรู้จำเพิ่มขึ้นเพียงไม่เกินร้อยละ 2

ดังนั้นสัมประสิทธิ์แอลพีแอลอีของค่าอัตสหสัมพันธ์ที่มีความคงทนเมื่อนำมาใช้กับเสียงพูดกลุ่มเสียงวรรณยุกต์สามัญ ไม่ช่วยให้อัตราการรู้จำเฉลี่ยดีขึ้นเมื่อมีสัญญาณรบกวนมากขึ้น

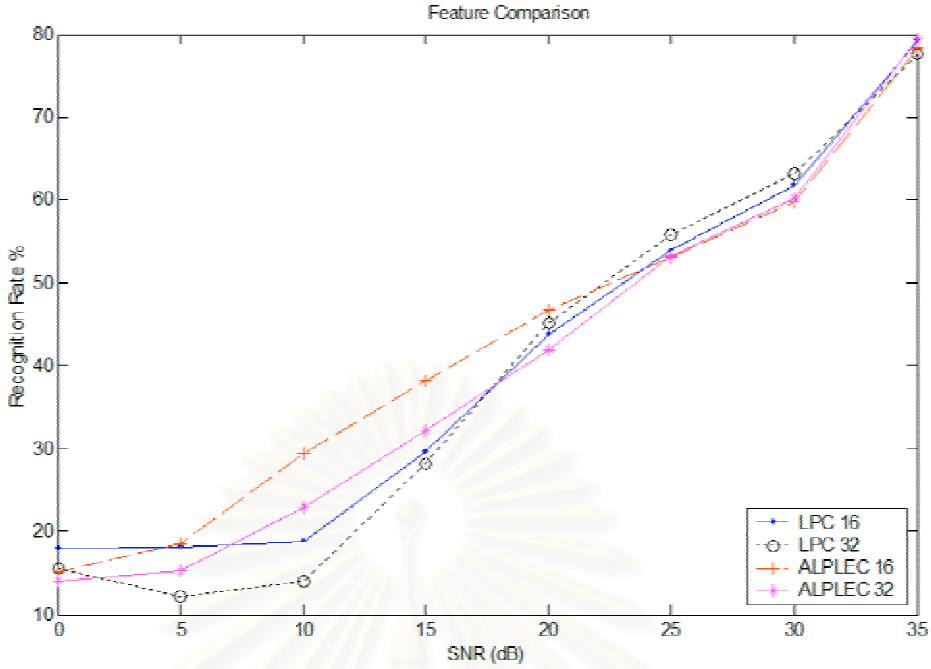
4.2.1.2 อัตราการรู้จำเฉลี่ยของเสียงพูดกลุ่มเสียงวรรณยุกต์เอก

ผลอัตราการรู้จำเฉลี่ยของเสียงพูดกลุ่มเสียงวรรณยุกต์เอก ของค่าลักษณะสำคัญ คือ สัมประสิทธิ์การประมาณพันธะเชิงเส้น อันดับ 16 (LPC16) สัมประสิทธิ์การประมาณพันธะเชิงเส้น อันดับ 32 (LPC32) สัมประสิทธิ์แอลพีแอลอีของค่าอัตสหสัมพันธ์ที่มีความคงทน อันดับ 16 (ALPLEC16) และ สัมประสิทธิ์แอลพีแอลอีของค่าอัตสหสัมพันธ์ที่มีความคงทน อันดับ 32 (ALPLEC32) แสดงในตารางที่ 4.2 และนำมาเขียนเป็นกราฟในรูปที่ 4.2

ตารางที่ 4.2 อัตราการรู้จำเฉลี่ยของ เสียงพูดกลุ่มเสียงวรรณยุกต์เอก

SNR	ค่าลักษณะสำคัญ			
	LPC16	ALPLEC16	LPC32	ALPLEC32
Clean	79.27	78.30	77.68	79.44
30 dB	61.73	59.74	63.10	60.31
25 dB	53.82	53.02	55.75	53.13
20 dB	43.74	46.64	45.16	41.86
15 dB	29.67	38.10	28.19	32.12
10 dB	18.85	29.38	13.95	22.84
5 dB	18.11	18.51	12.07	15.21
0 dB	17.88	14.98	15.49	13.95

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 4.2 อัตราการรู้จำเฉลี่ยของ เสียงพูดกลุ่มเสียงวรรณยุกต์เอก

จากรูปที่ 4.2 จะเห็นว่าในกรณีที่มีสัญญาณรบกวนน้อย (SNR > 20 dB) สัมประสิทธิ์แอลพีแอลอีของค่าอัตสหสัมพันธ์ที่มีความคงทน มีอัตราการรู้จำเฉลี่ยของเสียงพูดกลุ่มเสียงวรรณยุกต์เอกน้อยกว่าสัมประสิทธิ์การประมาณพันธะเชิงเส้นเพียงเล็กน้อย เฉลี่ยร้อยละ 2.47 แต่เมื่อมีสัญญาณรบกวนเพิ่มขึ้น (SNR ≤ 20 dB) จะให้ค่าอัตราการรู้จำเฉลี่ยที่มากกว่า เฉลี่ยร้อยละ 6.10

จำนวนอันดับที่เพิ่มขึ้นของสัมประสิทธิ์แอลพีแอลอีของค่าอัตสหสัมพันธ์ที่มีความคงทนกลับทำให้อัตราการรู้จำเฉลี่ยลดลง เฉลี่ยร้อยละ 2.48 ทั้งนี้เพราะสัมประสิทธิ์แอลพีแอลอีที่จำนวนอันดับมากขึ้นให้ค่าความผิดพลาดทางสเปกตรัมมากกว่า และจำนวนอันดับที่เพิ่มขึ้นของสัมประสิทธิ์การประมาณพันธะเชิงเส้นทำให้อัตราการรู้จำเฉลี่ยลดลง เฉลี่ยร้อยละ 1.46

เป็นที่น่าสังเกตว่าในกรณีที่ไม่มีสัญญาณรบกวน ทั้งค่าสัมประสิทธิ์การประมาณพันธะเชิงเส้นและสัมประสิทธิ์แอลพีแอลอีของค่าอัตสหสัมพันธ์ที่มีความคงทน ไม่ว่าจะใช้จำนวนอันดับเป็นเท่าใด จะให้อัตราการรู้จำเฉลี่ยใกล้เคียงกัน โดยมีความแตกต่างกันไม่เกินร้อยละ 1.76 และในกรณีที่กำลังของสัญญาณรบกวนมีค่าเท่ากับกำลังของสัญญาณ สัมประสิทธิ์การประมาณพันธะเชิงเส้น อันดับ 16 ให้อัตราการรู้จำเฉลี่ยดีที่สุดเป็นร้อยละ 17.88

ดังนั้นสัมประสิทธิ์แอลพีแอลอีของค่าอัตสหสัมพันธ์ที่มีความคงทน อันดับ 16 เมื่อนำมาใช้กับเสียงพูดกลุ่มเสียงวรรณยุกต์เอก ช่วยให้อัตราการรู้จำเฉลี่ยดีกว่าสัมประสิทธิ์การประมาณพันธะเชิงเส้นเมื่อมีค่า SNR < 25 dB

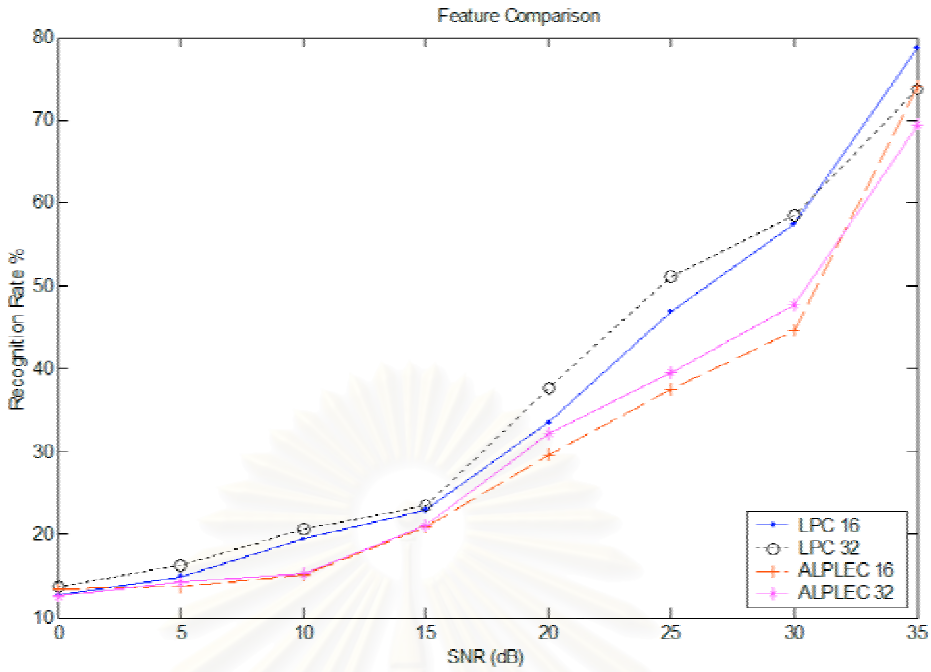
4.2.1.3 อัตราการรู้จำเฉลี่ยของเสียงพูดกลุ่มเสียงวรรณยุกต์โท

ผลอัตราการรู้จำเฉลี่ยของเสียงพูดกลุ่มเสียงวรรณยุกต์โท ของค่าลักษณะสำคัญ คือ สัมประสิทธิ์การประมาณพันธะเชิงเส้น อันดับ 16 (LPC16) สัมประสิทธิ์การประมาณพันธะเชิงเส้น อันดับ 32 (LPC32) สัมประสิทธิ์แอลพีแอลอีของค่าอัตสหสัมพันธ์ที่มีความคงทน อันดับ 16 (ALPLEC16) และ สัมประสิทธิ์แอลพีแอลอีของค่าอัตสหสัมพันธ์ที่มีความคงทน อันดับ 32 (ALPLEC32) แสดงในตารางที่ 4.3 และนำมาเขียนเป็นกราฟในรูปที่ 4.3

ตารางที่ 4.3 อัตราการรู้จำเฉลี่ยของ เสียงพูดกลุ่มเสียงวรรณยุกต์โท

SNR	ค่าลักษณะสำคัญ			
	LPC16	ALPLEC16	LPC32	ALPLEC32
Clean	78.64	74.15	73.81	69.49
30 dB	57.50	44.60	58.58	47.73
25 dB	46.82	37.44	51.08	39.60
20 dB	33.52	29.55	37.67	32.22
15 dB	22.90	20.91	23.35	21.02
10 dB	19.55	15.11	20.74	15.23
5 dB	14.94	13.69	16.31	14.20
0 dB	12.73	13.30	13.69	12.50

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 4.3 อัตราการรู้จำเฉลี่ยของ เสียงพูดกลุ่มเสียงวรรณยุกต์โท

จากรูปที่ 4.3 จะเห็นว่าสัมประสิทธิ์แอลพีแอลอีของค่าอัตสหสัมพันธ์ที่มีความคงทน มีอัตราการรู้จำเฉลี่ยของเสียงพูดกลุ่มเสียงวรรณยุกต์โทน้อยกว่าสัมประสิทธิ์การประมาณพันธะเชิงเส้นทุกค่าอัตราส่วนสัญญาณต่อสัญญาณรบกวน

นอกจากนี้จำนวนอันดับที่เพิ่มขึ้นของสัมประสิทธิ์แอลพีแอลอีของค่าอัตสหสัมพันธ์ที่มีความคงทนกลับทำให้อัตราการรู้จำเฉลี่ยเพิ่มขึ้นเล็กน้อย เฉลี่ยร้อยละ 0.40 ผลที่ได้ใกล้เคียงกับจำนวนอันดับที่เพิ่มขึ้นของสัมประสิทธิ์การประมาณพันธะเชิงเส้นที่การเพิ่มจำนวนอันดับทำให้อัตราการรู้จำเฉลี่ยเพิ่มขึ้น เฉลี่ยร้อยละ 1.08

เป็นที่น่าสังเกตว่าในกรณีที่ไม่มีสัญญาณรบกวน จำนวนอันดับของสัมประสิทธิ์ที่น้อยกว่ากลับให้ค่าอัตราการรู้จำเฉลี่ยมากกว่า ทั้งนี้เพราะสัมประสิทธิ์จำนวนมากขึ้นจะเป็นตัวแทนของข้อมูลในส่วนของสัญญาณรบกวนมากกว่าตัวสัญญาณจริงที่ต้องการ และในกรณีที่กำลังของสัญญาณรบกวนมีค่าเท่ากับกำลังของสัญญาณ ทั้งค่าสัมประสิทธิ์การประมาณพันธะเชิงเส้นและสัมประสิทธิ์แอลพีแอลอีของค่าอัตสหสัมพันธ์ที่มีความคงทน ไม่ว่าจะใช้จำนวนอันดับเป็นเท่าใดจะให้อัตราการรู้จำเฉลี่ยใกล้เคียงกัน โดยมีความแตกต่างกันไม่เกินร้อยละ 1.19

ดังนั้นสัมประสิทธิ์แอลพีแอลอีของค่าอัตสหสัมพันธ์ที่มีความคงทนเมื่อนำมาใช้กับเสียงพูดกลุ่มเสียงวรรณยุกต์โท ไม่ช่วยให้อัตราการรู้จำเฉลี่ยดีขึ้นเมื่อมีสัญญาณรบกวนมากขึ้น

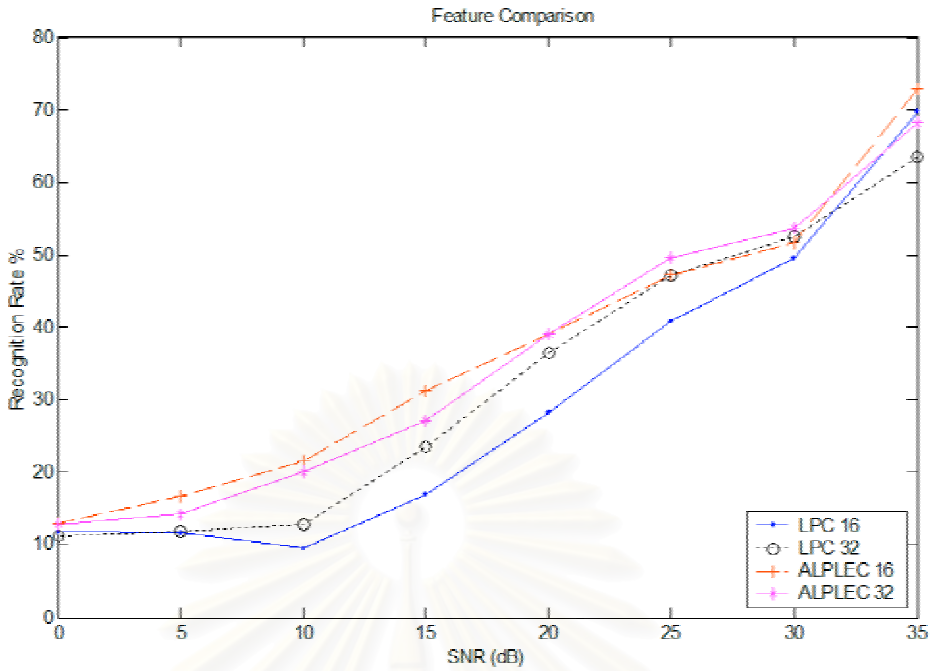
4.2.1.4 อัตราการรู้จำเฉลี่ยของเสียงพูดกลุ่มเสียงวรรณยุกต์ตรี

ผลอัตราการรู้จำเฉลี่ยของเสียงพูดกลุ่มเสียงวรรณยุกต์ตรี ของค่าลักษณะสำคัญ คือ สัมประสิทธิ์การประมาณพันธะเชิงเส้น อันดับ 16 (LPC16) สัมประสิทธิ์การประมาณพันธะเชิงเส้น อันดับ 32 (LPC32) สัมประสิทธิ์แอลพีแอลอีของค่าอัตสหสัมพันธ์ที่มีความคงทน อันดับ 16 (ALPLEC16) และ สัมประสิทธิ์แอลพีแอลอีของค่าอัตสหสัมพันธ์ที่มีความคงทน อันดับ 32 (ALPLEC32) แสดงในตารางที่ 4.4 และนำมาเขียนเป็นกราฟในรูปที่ 4.4

ตารางที่ 4.4 อัตราการรู้จำเฉลี่ยของ เสียงพูดกลุ่มเสียงวรรณยุกต์ตรี

SNR	ค่าลักษณะสำคัญ			
	LPC16	ALPLEC16	LPC32	ALPLEC32
Clean	69.75	72.79	63.38	68.14
30 dB	49.54	51.72	52.47	53.62
25 dB	40.70	47.19	47.24	49.60
20 dB	28.36	39.04	36.51	39.09
15 dB	16.93	31.11	23.54	27.10
10 dB	9.47	21.70	12.86	20.03
5 dB	11.65	16.65	11.88	14.18
0 dB	11.77	12.97	11.19	12.86

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 4.4 อัตราการรู้จำเฉลี่ยของ เสียงพูดกลุ่มเสียงวรรณยุกต์ตรี

จากรูปที่ 4.4 จะเห็นว่าไม่ว่าจะมีสัญญาณรบกวนมากน้อยแค่ไหน สัมประสิทธิ์แอลพีแอลอีของค่าอัตสหสัมพันธ์ที่มีความคงทน มีอัตราการรู้จำเฉลี่ยของเสียงพูดกลุ่มเสียงวรรณยุกต์ตรีมากกว่าสัมประสิทธิ์การประมาณพันธะเชิงเส้น โดยที่ค่า SNR ≥ 25 dB สัมประสิทธิ์แอลพีแอลอีของค่าอัตสหสัมพันธ์ที่มีความคงทน จำนวนอันดับ 16 ให้ค่าอัตราการรู้จำเฉลี่ยใกล้เคียงกับสัมประสิทธิ์การประมาณพันธะเชิงเส้นที่ใช้จำนวนอันดับ 32 และที่ค่า SNR < 25 dB ให้อัตราการรู้จำเฉลี่ยมากกว่าสัมประสิทธิ์การประมาณพันธะเชิงเส้นทุกอันดับ โดยให้อัตราการรู้จำเฉลี่ยมากกว่าสัมประสิทธิ์การประมาณพันธะเชิงเส้น อันดับ 16 เฉลี่ยร้อยละ 6.88 และ มากกว่าสัมประสิทธิ์การประมาณพันธะเชิงเส้น อันดับ 32 เฉลี่ยร้อยละ 4.26

นอกจากนี้จำนวนอันดับที่เพิ่มขึ้นของสัมประสิทธิ์แอลพีแอลอีของค่าอัตสหสัมพันธ์ที่มีความคงทนทำให้อัตราการรู้จำเฉลี่ยลดลง เฉลี่ยร้อยละ 1.07 ทั้งนี้เพราะสัมประสิทธิ์แอลพีแอลอีที่จำนวนอันดับมากขึ้นให้ค่าความผิดพลาดทางสเปกตรัมมากกว่า ตรงข้ามกับจำนวนอันดับที่เพิ่มขึ้นของสัมประสิทธิ์การประมาณพันธะเชิงเส้นที่การเพิ่มจำนวนอันดับทำให้มีอัตราการรู้จำเฉลี่ยเพิ่มขึ้น เฉลี่ยร้อยละ 2.61

เป็นที่น่าสังเกตว่าในกรณีที่ไม่มีสัญญาณรบกวน จำนวนอันดับของสัมประสิทธิ์ที่น้อยกว่ากลับให้ค่าอัตราการรู้จำเฉลี่ยมากกว่า ทั้งนี้เพราะสัมประสิทธิ์จำนวนมากขึ้นจะเป็นตัวแทนของข้อมูลในส่วนของสัญญาณรบกวนมากกว่าตัวสัญญาณจริงที่ต้องการ และในกรณีที่กำลังของสัญญาณรบกวนมีค่าเท่ากับกำลังของสัญญาณ ทั้งค่าสัมประสิทธิ์การประมาณพันธะเชิงเส้นและ

สัมประสิทธิ์แอลพีแอลอีของค่าอัตสหสัมพันธ์ที่มีความคงทน ไม่ว่าจะใช้จำนวนอันดับเป็นเท่าใดจะให้อัตราการรู้จำเฉลี่ยใกล้เคียงกัน โดยมีความแตกต่างกันไม่เกินร้อยละ 1.78

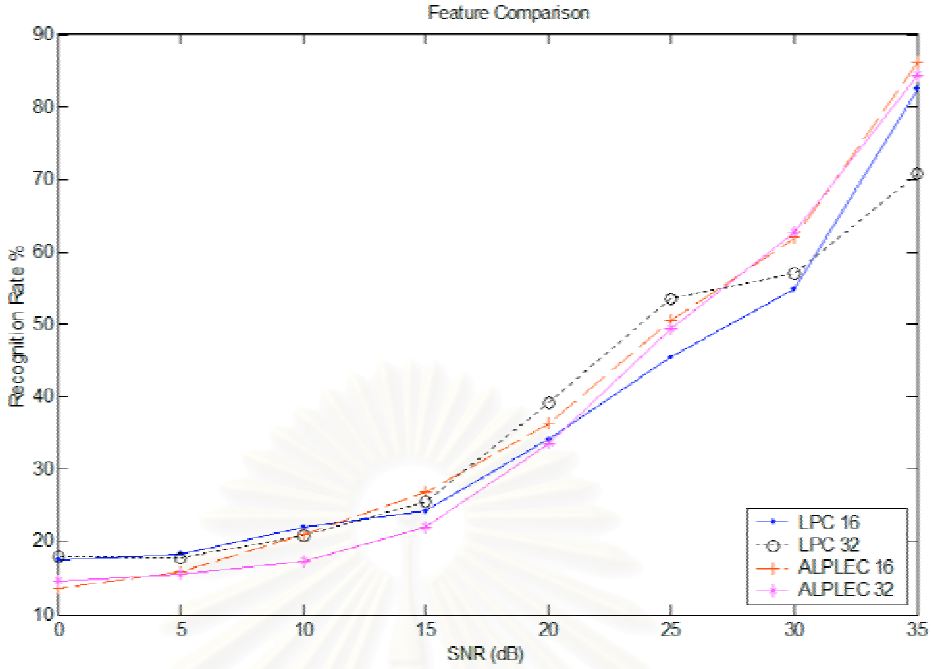
ดังนั้นสัมประสิทธิ์แอลพีแอลอีของค่าอัตสหสัมพันธ์ที่มีความคงทน อันดับ 16 เมื่อนำมาใช้กับเสียงพูดกลุ่มเสียงวรรณยุกต์ตรี ช่วยให้อัตราการรู้จำเฉลี่ยดีกว่าสัมประสิทธิ์การประมาณพันธะเชิงเส้นทุกค่า SNR

4.2.1.5 อัตราการรู้จำเฉลี่ยของเสียงพูดกลุ่มเสียงวรรณยุกต์จัตวา

ผลอัตราการรู้จำเฉลี่ยของเสียงพูดกลุ่มเสียงวรรณยุกต์จัตวา ของค่าลักษณะสำคัญ คือ สัมประสิทธิ์การประมาณพันธะเชิงเส้น อันดับ 16 (LPC16) สัมประสิทธิ์การประมาณพันธะเชิงเส้น อันดับ 32 (LPC32) สัมประสิทธิ์แอลพีแอลอีของค่าอัตสหสัมพันธ์ที่มีความคงทน อันดับ 16 (ALPLEC16) และ สัมประสิทธิ์แอลพีแอลอีของค่าอัตสหสัมพันธ์ที่มีความคงทน อันดับ 32 (ALPLEC32) แสดงในตารางที่ 4.5 และนำมาเขียนเป็นกราฟในรูปที่ 4.5

ตารางที่ 4.5 อัตราการรู้จำเฉลี่ยของ เสียงพูดกลุ่มเสียงวรรณยุกต์จัตวา

SNR	ค่าลักษณะสำคัญ			
	LPC16	ALPLEC16	LPC32	ALPLEC32
Clean	82.46	86.23	70.74	84.34
30 dB	54.97	61.94	56.97	62.63
25 dB	45.54	50.63	53.43	49.37
20 dB	34.23	36.23	39.31	33.66
15 dB	24.11	26.86	25.49	22.06
10 dB	21.94	21.03	20.91	17.26
5 dB	18.23	15.83	17.66	15.49
0 dB	17.60	13.66	18.00	14.51



รูปที่ 4.5 อัตราการรู้จำเฉลี่ยของ เสียงพูดกลุ่มเสียงวรรณยุกต์จัตวา

จากรูปที่ 4.5 จะเห็นว่าจำนวนอันดับที่เพิ่มขึ้นของสัมประสิทธิ์แอลพีแอลอีของค่าอัตสหสัมพันธ์ที่มีความคงทนไม่ช่วยให้อัตราการรู้จำเฉลี่ยมากขึ้น โดยมีค่าลดลงไม่เกินร้อยละ 4.80 ในช่วง 5 dB ๗ SNR ๗ 30 dB ทั้งนี้เพราะจำนวนสัมประสิทธิ์ที่มากขึ้นจะได้ข้อมูลของสัญญาณรบกวนมากกว่าตัวสัญญาณจริงที่ต้องการ ตรงข้ามกับจำนวนอันดับที่เพิ่มขึ้นของสัมประสิทธิ์การประมาณพันธะเชิงเส้นที่ช่วยให้อัตราการรู้จำเฉลี่ยเพิ่มขึ้นถึงร้อยละ 7.89 ที่ค่า SNR = 25 dB

แต่หากเราสนใจใช้จำนวนอันดับของสัมประสิทธิ์เป็น 16 จะพบว่าสัมประสิทธิ์แอลพีแอลอีของค่าอัตสหสัมพันธ์ที่มีความคงทน ให้อัตราการรู้จำเฉลี่ยมากกว่าสัมประสิทธิ์การประมาณพันธะเชิงเส้นที่ค่า SNR > 10 dB

นอกจากนี้หากเปรียบเทียบอัตราการรู้จำของทุกค่าลักษณะสำคัญในทุกกลุ่มเสียงวรรณยุกต์จะพบว่า กลุ่มเสียงวรรณยุกต์จัตวาให้ค่าอัตราการรู้จำเฉลี่ยมากที่สุดเมื่อใช้กับค่าลักษณะสำคัญที่เป็นสัมประสิทธิ์การประมาณพันธะเชิงเส้นและสัมประสิทธิ์แอลพีแอลอีของค่าอัตสหสัมพันธ์ที่มีความคงทน

สัมประสิทธิ์แอลพีแอลอีของค่าอัตสหสัมพันธ์ที่มีความคงทน เมื่อนำมาใช้กับเสียงพูดกลุ่มเสียงวรรณยุกต์จัตวา จะให้อัตราการรู้จำเฉลี่ยใกล้เคียงกับสัมประสิทธิ์การประมาณพันธะเชิงเส้น และจะให้อัตราการรู้จำเฉลี่ยดีกว่าเมื่อพิจารณาใช้จำนวนอันดับเป็น 16

4.2.2 กลุ่มค่าลักษณะสำคัญแบบสัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอล

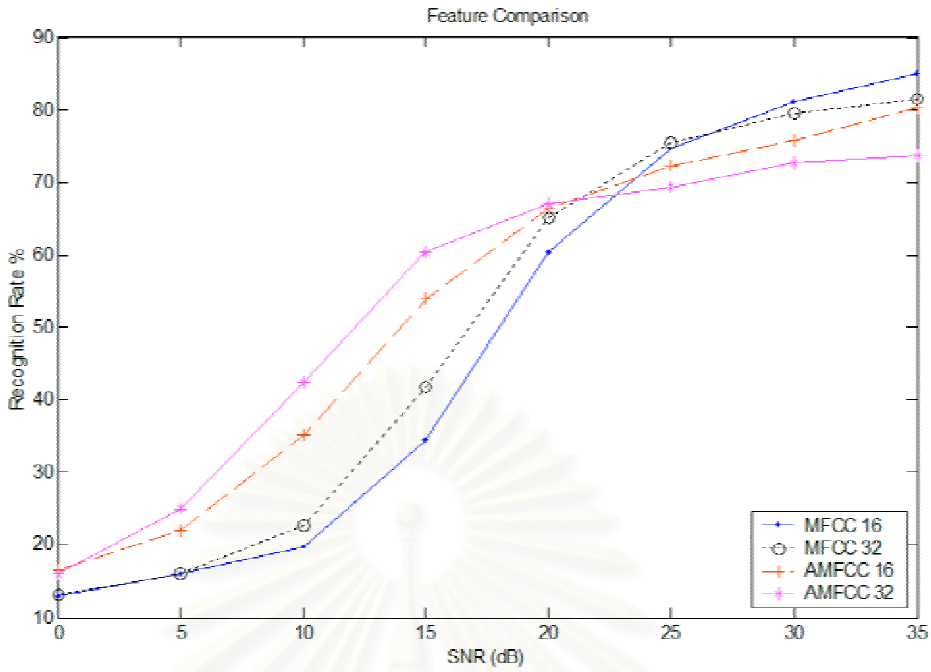
4.2.2.1 อัตราการรู้จำเฉลี่ยของเสียงพูดกลุ่มเสียงวรรณยุกต์สามัญ

ผลอัตราการรู้จำเฉลี่ยของเสียงพูดกลุ่มเสียงวรรณยุกต์สามัญ ของค่าลักษณะสำคัญ คือ สัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอล อันดับ 16 (MFCC16) สัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอล อันดับ 32 (MFCC32) สัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอลของค่าอัตสหสัมพันธ์ที่มีความคงทน อันดับ 16 (AMFCC16) และ สัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอลของค่าอัตสหสัมพันธ์ที่มีความคงทน อันดับ 32 (AMFCC32) แสดงในตารางที่ 4.6 และนำมาเขียนเป็นกราฟในรูปที่ 4.6

ตารางที่ 4.6 อัตราการรู้จำเฉลี่ยของ เสียงพูดกลุ่มเสียงวรรณยุกต์สามัญ

SNR	ค่าลักษณะสำคัญ			
	MFCC16	AMFCC16	MFCC32	AMFCC32
Clean	85.11	80.23	81.48	73.58
30 dB	81.02	75.80	79.49	72.61
25 dB	74.60	72.39	75.51	69.32
20 dB	60.40	66.42	65.11	66.99
15 dB	34.38	53.98	41.76	60.34
10 dB	19.77	35.11	22.67	42.44
5 dB	15.97	21.99	15.97	25.00
0 dB	12.90	16.48	13.01	16.02

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 4.6 อัตราการรู้จำเฉลี่ยของ เสียงพูดกลุ่มเสียงวรรณยุกต์สามัญ

จากรูปที่ 4.6 จะเห็นว่าสัมประสิทธิ์เมลฟรีเควันซีเซปสตรอลของค่าอัตสหสัมพันธ์ที่มีความคงทนให้อัตราการรู้จำเฉลี่ยน้อยกว่าสัมประสิทธิ์เมลฟรีเควันซีเซปสตรอลเมื่อค่า SNR $\geq$ 25 dB ทั้งนี้เพราะที่ค่า SNR มากๆข้อมูลที่สำคัญที่สุดของค่าอัตสหสัมพันธ์จะอยู่ที่เวลาศูนย์ ซึ่งเป็นค่าที่ไม่ถูกนำมาใช้ในการหาค่าอัตสหสัมพันธ์ที่มีความคงทน แต่เมื่อมีสัญญาณรบกวนมากขึ้น ที่ค่า SNR $\leq$ 20 dB จะให้อัตราการรู้จำเฉลี่ยมากกว่า โดย ที่จำนวนอันดับ 16 ให้อัตราการรู้จำเฉลี่ยมากกว่า เฉลี่ยร้อยละ 10.11 และ ที่จำนวนอันดับ 32 ให้อัตราการรู้จำเฉลี่ยมากกว่า เฉลี่ยร้อยละ 10.45

การเพิ่มขึ้นของจำนวนอันดับของสัมประสิทธิ์เมลฟรีเควันซีเซปสตรอล และสัมประสิทธิ์เมลฟรีเควันซีเซปสตรอลของค่าอัตสหสัมพันธ์ที่มีความคงทน ช่วยให้อัตราการรู้จำเฉลี่ยเพิ่มขึ้น ร้อยละ 3.02 และ 3.36 ตามลำดับ ที่ค่า SNR < 25 dB แต่เมื่อใช้กับเสียงที่มีสัญญาณรบกวนน้อย (SNR $\geq$ 25 dB) จะทำให้อัตราการรู้จำเฉลี่ยลดลงร้อยละ 1.42 และ 4.30 ตามลำดับ

เป็นที่น่าสังเกตว่า ในกรณีที่กำลังของสัญญาณรบกวนมีค่าเท่ากับกำลังของสัญญาณ จำนวนอันดับที่เพิ่มขึ้นแทบจะไม่ช่วยให้อัตราการรู้จำเฉลี่ยเพิ่มขึ้นหรือลดลงเลย

ดังนั้นสัมประสิทธิ์เมลฟรีเควันซีเซปสตรอลของค่าอัตสหสัมพันธ์ที่มีความคงทนอันดับ 32 เมื่อนำมาใช้กับเสียงพูดกลุ่มเสียงวรรณยุกต์สามัญ ช่วยให้อัตราการรู้จำเฉลี่ยดีกว่าสัมประสิทธิ์เมลฟรีเควันซีเซปสตรอล เมื่อมีค่า SNR < 25 dB

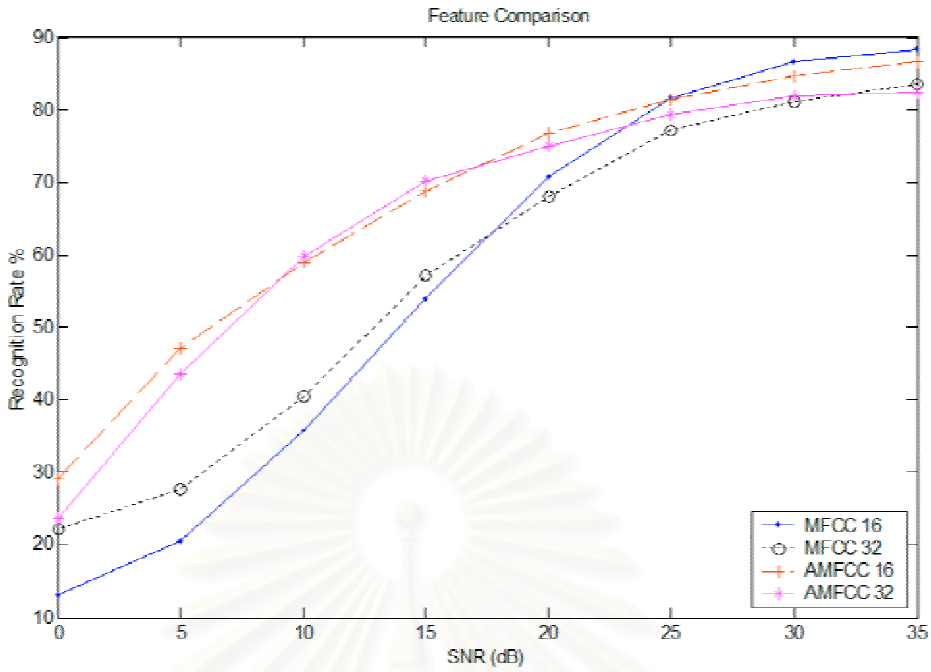
4.2.2.2 อัตราการรู้จำเฉลี่ยของเสียงพูดกลุ่มเสียงวรรณยุกต์เอก

ผลอัตราการรู้จำเฉลี่ยของเสียงพูดกลุ่มเสียงวรรณยุกต์เอก ของค่าลักษณะสำคัญ คือ สัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอล อันดับ 16 (MFCC16) สัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอล อันดับ 32 (MFCC32) สัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอลของค่าอัตสหสัมพันธ์ที่มีความคงทน อันดับ 16 (AMFCC16) และ สัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอลของค่าอัตสหสัมพันธ์ที่มีความคงทน อันดับ 32 (AMFCC32) แสดงในตารางที่ 4.7 และนำมาเขียนเป็นกราฟในรูปที่ 4.7

ตารางที่ 4.7 อัตราการรู้จำเฉลี่ยของ เสียงพูดกลุ่มเสียงวรรณยุกต์เอก

SNR	ค่าลักษณะสำคัญ			
	MFCC16	AMFCC16	MFCC32	AMFCC32
Clean	88.33	86.50	83.71	82.52
30 dB	86.56	84.68	81.15	81.95
25 dB	81.61	81.49	77.22	79.27
20 dB	70.79	76.77	68.05	75.00
15 dB	53.99	68.85	57.29	70.22
10 dB	35.65	58.94	40.43	59.74
5 dB	20.50	47.10	27.62	43.51
0 dB	12.98	29.10	22.21	23.63

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 4.7 อัตราการรู้จำเฉลี่ยของ เสียงพูดกลุ่มเสียงวรรณยุกต์เอก

จากรูปที่ 4.7 จะเห็นว่าสัมประสิทธิ์เมลฟรีเควันซีเซปสตรอลของค่าอัตสหสัมพันธ์ที่มีความคงทนให้อัตราการรู้จำเฉลี่ยน้อยกว่าสัมประสิทธิ์เมลฟรีเควันซีเซปสตรอลเมื่อค่า SNR > 25 dB ทั้งนี้เพราะที่ค่า SNR มากๆข้อมูลที่สำคัญที่สุดของค่าอัตสหสัมพันธ์จะอยู่ที่เวลาศูนย์ ซึ่งเป็นค่าที่ไม่ถูกนำมาใช้ในการหาค่าอัตสหสัมพันธ์ที่มีความคงทน แต่เมื่อมีสัญญาณรบกวนมากขึ้น ที่ค่า SNR ≤ 25 dB จะให้อัตราการรู้จำเฉลี่ยมากกว่า โดย ที่จำนวนอันดับ 16 ให้อัตราการรู้จำเฉลี่ยมากกว่า เฉลี่ยร้อยละ 14.46 และ ที่จำนวนอันดับ 32 ให้อัตราการรู้จำเฉลี่ยมากกว่า เฉลี่ยร้อยละ 9.76

จำนวนอันดับที่เพิ่มขึ้นของสัมประสิทธิ์เมลฟรีเควันซีเซปสตรอลของค่าอัตสหสัมพันธ์ที่มีความคงทน ไม่ช่วยให้อัตราการรู้จำเฉลี่ยเพิ่มขึ้นเพราะจะมีการเพิ่มขึ้นและลดลงในบางช่วง โดยมีค่าเฉลี่ยลดลงร้อยละ 2.00 ซึ่งแตกต่างกับ จำนวนอันดับที่เพิ่มขึ้นของสัมประสิทธิ์เมลฟรีเควันซีเซปสตรอล ที่ช่วยให้อัตราการรู้จำเฉลี่ยเพิ่มขึ้นที่ค่า SNR < 20 dB เฉลี่ยร้อยละ 6.11 แต่ก็ทำให้อัตราการรู้จำเฉลี่ยลดลงที่ค่า SNR ≥ 20 dB เฉลี่ยร้อยละ 4.29 เช่นกัน

เป็นที่น่าสังเกตว่า ในกรณีที่กำลังของสัญญาณรบกวนมีค่าเท่ากับกำลังของสัญญาณ จำนวนอันดับที่เพิ่มขึ้นของสัมประสิทธิ์เมลฟรีเควันซีเซปสตรอลของค่าอัตสหสัมพันธ์ที่มีความคงทน ทำให้อัตราการรู้จำเฉลี่ยลดลงร้อยละ 5.47 ตรงข้ามกับจำนวนอันดับที่เพิ่มขึ้นของสัมประสิทธิ์เมลฟรีเควันซีเซปสตรอล ทำให้อัตราการรู้จำเฉลี่ยเพิ่มขึ้นร้อยละ 9.23

ดังนั้นสัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอลของค่าอัตสหสัมพันธ์ที่มีความคงทนอันดับ 16 เมื่อนำมาใช้กับเสียงพูดกลุ่มเสียงวรรณยุกต์เอก ช่วยให้อัตราการรู้จำเฉลี่ยดีกว่าสัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอล เมื่อมีค่า SNR ≤ 25 dB

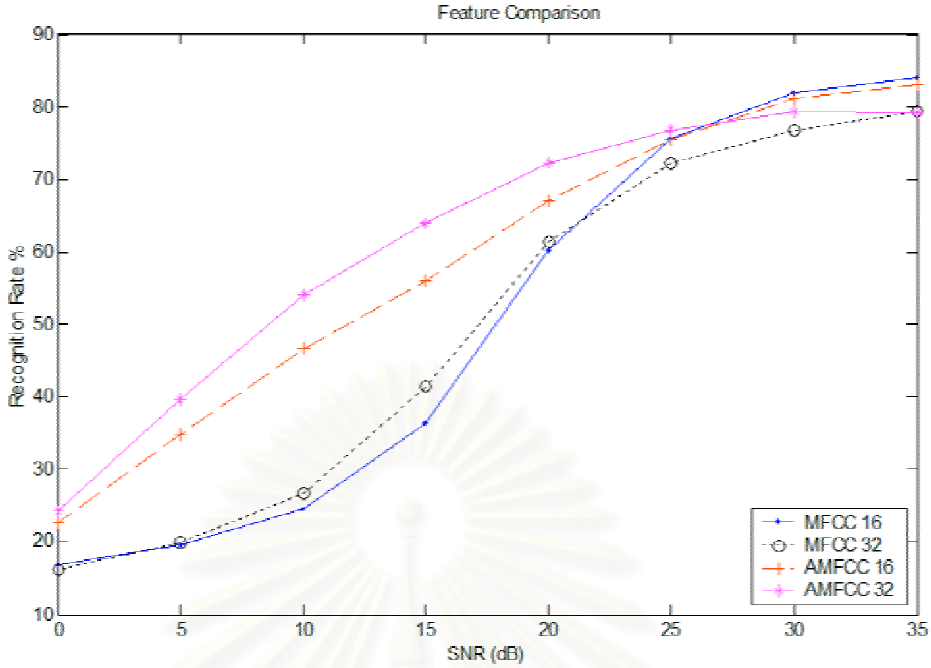
4.2.2.3 อัตราการรู้จำเฉลี่ยของเสียงพูดกลุ่มเสียงวรรณยุกต์โท

ผลอัตราการรู้จำเฉลี่ยของเสียงพูดกลุ่มเสียงวรรณยุกต์โท ของค่าลักษณะสำคัญ คือ สัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอล อันดับ 16 (MFCC16) สัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอล อันดับ 32 (MFCC32) สัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอลของค่าอัตสหสัมพันธ์ที่มีความคงทนอันดับ 16 (AMFCC16) และ สัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอลของค่าอัตสหสัมพันธ์ที่มีความคงทน อันดับ 32 (AMFCC32) แสดงในตารางที่ 4.8 และนำมาเขียนเป็นกราฟในรูปที่ 4.8

ตารางที่ 4.8 อัตราการรู้จำเฉลี่ยของ เสียงพูดกลุ่มเสียงวรรณยุกต์โท

SNR	ค่าลักษณะสำคัญ			
	MFCC16	AMFCC16	MFCC32	AMFCC32
Clean	84.09	83.07	79.32	79.15
30 dB	81.93	81.08	76.76	79.32
25 dB	75.68	75.45	72.33	76.82
20 dB	60.23	67.05	61.36	72.33
15 dB	36.25	56.14	41.42	63.86
10 dB	24.49	46.70	26.65	54.20
5 dB	19.43	34.77	19.89	39.55
0 dB	16.82	22.56	16.25	24.20

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 4.8 อัตราการรู้จำเฉลี่ยของ เสียงพูดกลุ่มเสียงวรรณยุกต์โท

จากรูปที่ 4.8 จะเห็นว่าสัมประสิทธิ์เมลฟรีเควันซีเซปสตรอลของค่าอัตสหสัมพันธ์ที่มีความคงทนให้อัตราการรู้จำเฉลี่ยน้อยกว่าสัมประสิทธิ์เมลฟรีเควันซีเซปสตรอลเมื่อค่า SNR > 25 dB ทั้งนี้เพราะที่ค่า SNR มากๆข้อมูลที่สำคัญที่สุดของค่าอัตสหสัมพันธ์จะอยู่ที่เวลาศูนย์ ซึ่งเป็นค่าที่ไม่ถูกนำมาใช้ในการหาค่าอัตสหสัมพันธ์ที่มีความคงทน แต่เมื่อมีสัญญาณรบกวนมากขึ้น ที่ค่า SNR ≤ 25 dB จะให้อัตราการรู้จำเฉลี่ยมากกว่า โดย ที่จำนวนอันดับ 16 ให้อัตราการรู้จำเฉลี่ยมากกว่า เฉลี่ยร้อยละ 11.63 และ ที่จำนวนอันดับ 32 ให้อัตราการรู้จำเฉลี่ยมากกว่า เฉลี่ยร้อยละ 15.51

จำนวนอันดับที่เพิ่มขึ้นของสัมประสิทธิ์เมลฟรีเควันซีเซปสตรอลของค่าอัตสหสัมพันธ์ที่มีความคงทน ช่วยให้อัตราการรู้จำเฉลี่ยเพิ่มขึ้น เฉลี่ยร้อยละ 5.38 ที่ค่า SNR < 25 dB แต่เมื่อใช้กับเสียงที่มีสัญญาณรบกวนน้อย (SNR ≥ 25 dB) จะให้อัตราการรู้จำเฉลี่ยลดลง เฉลี่ยร้อยละ 1.44 จำนวนอันดับที่เพิ่มขึ้นของสัมประสิทธิ์เมลฟรีเควันซีเซปสตรอลช่วยให้อัตราการรู้จำเฉลี่ยเพิ่มขึ้นที่ค่า SNR ≤ 20 dB เฉลี่ยร้อยละ 1.67 แต่ก็ทำให้อัตราการรู้จำเฉลี่ยลดลงที่ค่า SNR > 20 dB เฉลี่ยร้อยละ 4.43

เป็นที่น่าสังเกตว่า ในกรณีที่กำลังของสัญญาณรบกวนมีค่าเท่ากับกำลังของสัญญาณ จำนวนอันดับที่เพิ่มขึ้นแทบจะไม่ช่วยให้อัตราการรู้จำเฉลี่ยเพิ่มขึ้นหรือลดลงเลย

ดังนั้นสัมประสิทธิ์เมลฟรีเควันซีเซปสตรอลของค่าอัตสหสัมพันธ์ที่มีความคงทนอันดับ 32 เมื่อนำมาใช้กับเสียงพูดกลุ่มเสียงวรรณยุกต์โท ช่วยให้อัตราการรู้จำเฉลี่ยดีกว่าสัมประสิทธิ์เมลฟรีเควันซีเซปสตรอล เมื่อมีค่า SNR < 30 dB

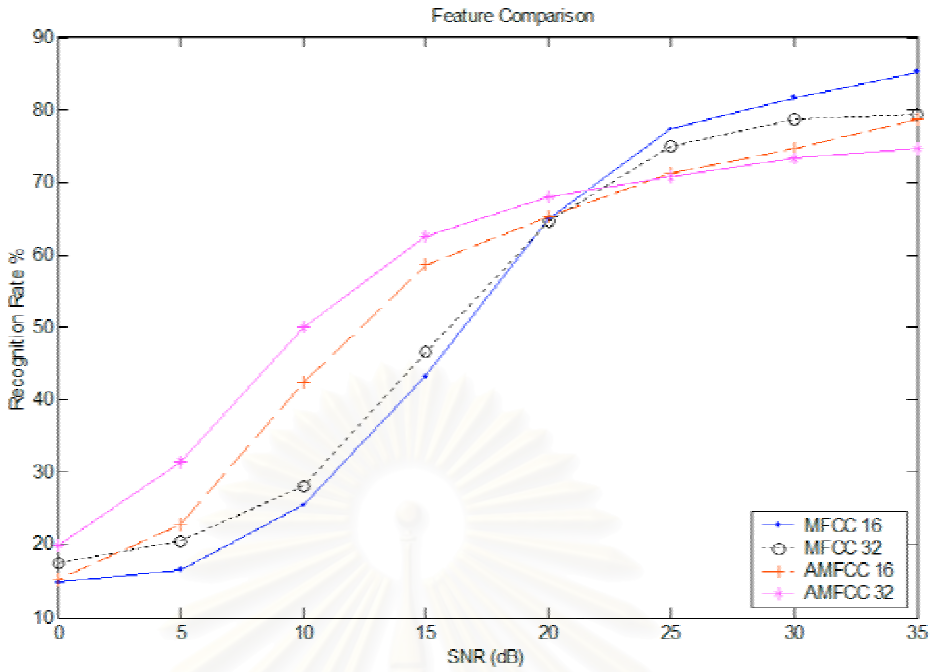
4.2.2.4 อัตราการรู้จำเฉลี่ยของเสียงพูดกลุ่มเสียงวรรณยุกต์ตรี

ผลอัตราการรู้จำเฉลี่ยของเสียงพูดกลุ่มเสียงวรรณยุกต์ตรี ของค่าลักษณะสำคัญ คือ สัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอล อันดับ 16 (MFCC16) สัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอล อันดับ 32 (MFCC32) สัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอลของค่าอัตสหสัมพันธ์ที่มีความคงทน อันดับ 16 (AMFCC16) และ สัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอลของค่าอัตสหสัมพันธ์ที่มีความคงทน อันดับ 32 (AMFCC32) แสดงในตารางที่ 4.9 และนำมาเขียนเป็นกราฟในรูปที่ 4.9

ตารางที่ 4.9 อัตราการรู้จำเฉลี่ยของ เสียงพูดกลุ่มเสียงวรรณยุกต์ตรี

SNR	ค่าลักษณะสำคัญ			
	MFCC16	AMFCC16	MFCC32	AMFCC32
Clean	85.19	78.70	79.39	74.63
30 dB	81.63	74.74	78.70	73.25
25 dB	77.38	71.30	74.97	70.67
20 dB	64.93	65.21	64.64	67.97
15 dB	43.11	58.55	46.61	62.51
10 dB	25.49	42.37	28.01	50.06
5 dB	16.65	22.90	20.38	31.40
0 dB	14.81	15.27	17.51	19.80

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 4.9 อัตราการรู้จำเฉลี่ยของ เสียงพูดกลุ่มเสียงวรรณยุกต์ตรี

จากรูปที่ 4.9 จะเห็นว่าสัมประสิทธิ์เมลฟรีเควันซีเซปสตรอลของค่าอัตสหสัมพันธ์ที่มีความคงทนให้อัตราการรู้จำเฉลี่ยน้อยกว่าสัมประสิทธิ์เมลฟรีเควันซีเซปสตรอลเมื่อค่า SNR ≥ 25 dB ทั้งนี้เพราะที่ค่า SNR มากๆข้อมูลที่สำคัญที่สุดของค่าอัตสหสัมพันธ์จะอยู่ที่เวลาศูนย์ ซึ่งเป็นค่าที่ไม่ถูกนำมาใช้ในการหาค่าอัตสหสัมพันธ์ที่มีความคงทน แต่เมื่อมีสัญญาณรบกวนมากขึ้น ที่ค่า SNR ≤ 20 dB จะให้อัตราการรู้จำเฉลี่ยมากกว่า โดย ที่จำนวนอันดับ 16 ให้อัตราการรู้จำเฉลี่ยมากกว่า เฉลี่ยร้อยละ 7.86 และ ที่จำนวนอันดับ 32 ให้อัตราการรู้จำเฉลี่ยมากกว่า เฉลี่ยร้อยละ 10.92

จำนวนอันดับที่เพิ่มขึ้นของสัมประสิทธิ์เมลฟรีเควันซีเซปสตรอลของค่าอัตสหสัมพันธ์ที่มีความคงทน ช่วยให้อัตราการรู้จำเฉลี่ยเพิ่มขึ้นที่ค่า SNR < 25 dB เฉลี่ยร้อยละ 5.49 และทำให้อัตราการรู้จำลดลงที่ค่า SNR ≥ 25 dB เฉลี่ยร้อยละ 2.06 จำนวนอันดับที่เพิ่มขึ้นของสัมประสิทธิ์เมลฟรีเควันซีเซปสตรอลช่วยให้อัตราการรู้จำเฉลี่ยลดลงที่ค่า SNR ≥ 20 dB ร้อยละ 2.86 และทำให้อัตราการรู้จำเฉลี่ยเพิ่มขึ้นที่ค่า SNR < 20 dB เฉลี่ยร้อยละ 3.11

ดังนั้นสัมประสิทธิ์เมลฟรีเควันซีเซปสตรอลของค่าอัตสหสัมพันธ์ที่มีความคงทนอันดับ 32 เมื่อนำมาใช้กับเสียงพูดกลุ่มเสียงวรรณยุกต์ตรี ช่วยให้อัตราการรู้จำเฉลี่ยดีกว่าสัมประสิทธิ์เมลฟรีเควันซีเซปสตรอล เมื่อมีค่า SNR < 25 dB

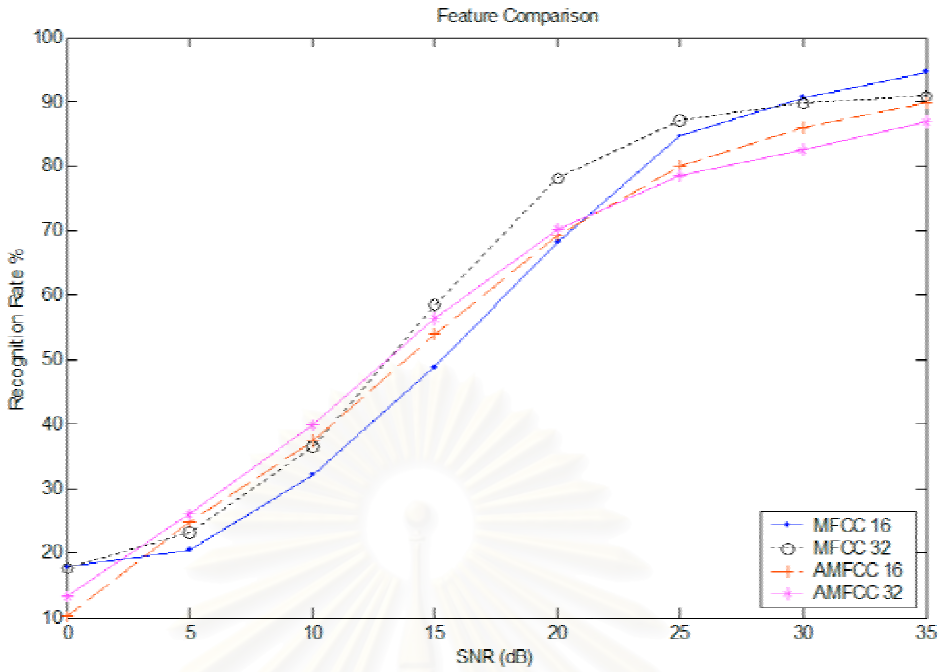
4.2.2.5 อัตราการรู้จำเฉลี่ยของเสียงพูดกลุ่มเสียงวรรณยุกต์จัตวา

ผลอัตราการรู้จำเฉลี่ยของเสียงพูดกลุ่มเสียงวรรณยุกต์จัตวา ของค่าลักษณะสำคัญ คือ สัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอล อันดับ 16 (MFCC16) สัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอล อันดับ 32 (MFCC32) สัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอลของค่าอัตสหสัมพันธ์ที่มีความคงทน อันดับ 16 (AMFCC16) และ สัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอลของค่าอัตสหสัมพันธ์ที่มีความคงทน อันดับ 32 (AMFCC32) แสดงในตารางที่ 4.10 และนำมาเขียนเป็นกราฟในรูปที่ 4.10

ตารางที่ 4.10 อัตราการรู้จำเฉลี่ยของ เสียงพูดกลุ่มเสียงวรรณยุกต์จัตวา

SNR	ค่าลักษณะสำคัญ			
	MFCC16	AMFCC16	MFCC32	AMFCC32
Clean	94.63	89.71	90.91	86.80
30 dB	90.63	85.94	89.83	82.40
25 dB	84.63	80.00	87.09	78.51
20 dB	68.29	69.31	78.17	70.23
15 dB	48.80	54.00	58.57	56.40
10 dB	32.11	37.43	36.46	39.71
5 dB	20.40	24.80	23.09	25.94
0 dB	17.71	10.11	17.49	13.14

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 4.10 อัตราการรู้จำเฉลี่ยของ เสียงพูดกลุ่มเสียงวรรณยุกต์จัตวา

จากรูปที่ 4.10 จะเห็นว่าจำนวนอันดับที่เพิ่มขึ้นของ สัมประสิทธิ์เมลฟรีเควันซีเซปสตรอลของค่าอัตสหสัมพันธ์ที่มีความคงทนช่วยให้อัตราการรู้จำเฉลี่ยเพิ่มขึ้นเล็กน้อย เฉลี่ยร้อยละ 1.95 เมื่อ SNR < 25 dB และมีค่าลดลง เฉลี่ยร้อยละ 2.65 เมื่อ SNR ≥ 25 dB ซึ่งแตกต่างจากจำนวนอันดับที่เพิ่มขึ้นของสัมประสิทธิ์เมลฟรีเควันซีเซปสตรอลที่ช่วยให้อัตราการรู้จำเฉลี่ยเพิ่มขึ้น เฉลี่ยร้อยละ 4.82 ที่ค่า SNR < 30 dB

แต่หากเราสนใจใช้จำนวนอันดับของสัมประสิทธิ์เป็น 16 จะพบว่า สัมประสิทธิ์เมลฟรีเควันซีเซปสตรอลของค่าอัตสหสัมพันธ์ที่มีความคงทน ให้อัตราการรู้จำเฉลี่ยมากกว่าสัมประสิทธิ์เมลฟรีเควันซีเซปสตรอลที่ค่า 5 dB ๗ SNR ๗ 20 dB

นอกจากนี้หากเราเปรียบเทียบอัตราการรู้จำของทุกค่าลักษณะสำคัญในทุกกลุ่มเสียงวรรณยุกต์จะพบว่า กลุ่มเสียงวรรณยุกต์จัตวาให้ค่าอัตราการรู้จำเฉลี่ยมากที่สุดเมื่อใช้กับค่าลักษณะสำคัญที่เป็นสัมประสิทธิ์เมลฟรีเควันซีเซปสตรอลและสัมประสิทธิ์เมลฟรีเควันซีเซปสตรอลของค่าอัตสหสัมพันธ์ที่มีความคงทน

โดยสรุปแล้วสัมประสิทธิ์เมลฟรีเควันซีเซปสตรอลของค่าอัตสหสัมพันธ์ที่มีความคงทน เมื่อนำมาใช้กับเสียงพูดกลุ่มเสียงวรรณยุกต์จัตวา ให้อัตราการรู้จำเฉลี่ยใกล้เคียงกับสัมประสิทธิ์เมลฟรีเควันซีเซปสตรอล และให้อัตราการรู้จำเฉลี่ยดีกว่าเมื่อพิจารณาใช้จำนวนอันดับเป็น 16

บทที่ 5

สรุปผลและข้อเสนอแนะ

5.1 สรุปผลการวิจัย

งานวิจัยนี้เป็นการพัฒนาระบบรู้จำเสียงพูดภาษาไทยให้มีความคงทนต่อสัญญาณรบกวนมากขึ้น การพัฒนาจะมุ่งความสนใจไปที่การหาค่าลักษณะสำคัญของสัญญาณเสียงพูดที่คงทนต่อสัญญาณรบกวนเกาส์เซียนขาวค่าเฉลี่ยศูนย์เพิ่มเข้ามา โดยใช้ค่าอัตสหสัมพันธ์ของสัญญาณเสียงพูดเป็นสัญญาณเข้าในการหาค่าลักษณะสำคัญแทนค่าสัญญาณเสียงพูด เนื่องจากมีความคงทนต่อสัญญาณรบกวนดังกล่าวมากกว่าตัวสัญญาณเสียงพูด จึงมีความคาดหวังว่าควรจะให้อัตราการรู้จำที่เพิ่มขึ้น

เพื่อให้แน่ใจได้ว่าค่าลักษณะสำคัญที่ได้มีความคงทนต่อสัญญาณรบกวนและให้อัตราการรู้จำที่เพิ่มขึ้นกับเสียงพูดภาษาไทยที่มีเสียงวรรณยุกต์ถึง 5 ระดับ ชุดคำศัพท์ที่เลือกมาใช้ในการฝึกฝนและทดสอบ จึงแบ่งเป็น 5 กลุ่ม คือ สามัญ เอก โท ตรี และจัตวา การฝึกฝนและทดสอบจึงแยกตามกลุ่มเสียง ไม่เกี่ยวข้องกัน โดยในแต่ละกลุ่มเสียงเลือกคำศัพท์มา 8 คำ แต่ละคำใช้จำนวนเสียงพูดประมาณ 110 เสียง

เนื่องจากตัวอย่างเสียงพูดที่บันทึกไว้ อยู่ในสภาพแวดล้อมที่ไม่มีสัญญาณรบกวนเท่านั้น ยังขาดสัญญาณเสียงพูดที่มีสัญญาณรบกวนเกาส์เซียนขาวค่าเฉลี่ยศูนย์เพิ่มเข้ามาสำหรับใช้ในการทดสอบความคงทนของค่าลักษณะสำคัญ และเพื่อให้สามารถเปรียบเทียบความคงทนของ ค่าลักษณะสำคัญแบบต่างๆที่ใช้จำนวนอันดับต่างกันได้ จึงต้องเตรียมตัวอย่างเสียงพูดของแต่ละชุดคำศัพท์ในทุกๆอัตราส่วนสัญญาณต่อสัญญาณรบกวนที่ใช้ในการทดสอบ บันทึกเก็บไว้บนฮาร์ดดิสก์

ค่าลักษณะสำคัญที่เลือกมาใช้คือ สัมประสิทธิ์การประมาณพันธะเชิงเส้นและสัมประสิทธิ์เมลฟรีเควินซีเซปสตรอล ที่มีจำนวนอันดับเป็น 16 และ 32 ค่าลักษณะสำคัญแต่ละแบบที่นำมาใช้จะนำสัญญาณเสียงพูดมาเป็นสัญญาณเข้าอันหนึ่ง และนำค่าค่าอัตสหสัมพันธ์ของสัญญาณเสียงพูดที่มีความคงทนเป็นสัญญาณเข้าอีกอันหนึ่ง แล้วจึงนำมาค่าลักษณะสำคัญทั้งสองในสภาพแวดล้อมที่ไม่มีสัญญาณรบกวนมาทำการฝึกฝน จากนั้นจึงนำค่าลักษณะสำคัญทั้งสองในสภาพแวดล้อมที่มีสัญญาณรบกวนต่างกันมาทำการทดสอบอัตราการรู้จำ และเปรียบเทียบผลที่ได้ว่า ค่าลักษณะสำคัญที่ใช้ค่าอัตสหสัมพันธ์ของสัญญาณเสียงพูดที่มีความคงทนเป็นสัญญาณเข้าจะช่วยให้ได้อัตราการรู้จำที่ดีขึ้นหรือไม่เมื่อเปรียบเทียบกับค่าลักษณะสำคัญที่ใช้สัญญาณเสียงพูดเป็นสัญญาณขาเข้าที่จำนวนอันดับเท่ากัน (สำหรับสัมประสิทธิ์การประมาณพันธะเชิงเส้น จะเปลี่ยนไปใช้เป็น

สัมประสิทธิ์แอลพีแอลอี เมื่อใช้ค่าอัตสหสัมพันธ์ของสัญญาณเสียงพูดที่มีความคงทนเป็นสัญญาณเข้า)

การฝึกฝนและทดสอบจะใช้ระบบรู้จำแบบจำลองฮิดเดนมาร์คอฟประเภทแบบจำลองซ้าย-ขวาแบบต่อเนื่องที่มี 7 สถานะและจำนวน Gaussian Mixture เป็น 2 นอกจากนี้เพื่อให้ได้ผลการทดสอบรู้จำของค่าลักษณะสำคัญที่ไม่ขึ้นกับผู้พูด จึงจัดแบ่งกลุ่มของผู้พูดออกเป็น 3 กลุ่ม โดยชุดรูปร่องต้นแบบอ้างอิงของเสียงพูดจะไม่ถูกนำมาใช้ทดสอบกับกลุ่มผู้พูดที่ใช้สร้างชุดรูปร่องต้นแบบนั้น

ผลการทดสอบค่าลักษณะสำคัญพบว่า ค่าสัมประสิทธิ์แอลพีแอลอีของค่าอัตสหสัมพันธ์ที่มีความคงทน ให้อัตราการรู้จำเฉลี่ยกับกลุ่มเสียงวรรณยุกต์เอก และตรี ที่ดีกว่าสัมประสิทธิ์การประมาณพันธะเชิงเส้น ให้อัตราการรู้จำเฉลี่ยกับกลุ่มเสียงวรรณยุกต์สามัญ และโท ที่น้อยกว่าสัมประสิทธิ์การประมาณพันธะเชิงเส้น และให้อัตราการรู้จำเฉลี่ยกลุ่มเสียงวรรณยุกต์จัตวา ใกล้เคียงกับสัมประสิทธิ์การประมาณพันธะเชิงเส้น

อย่างไรก็ตาม สัมประสิทธิ์แอลพีแอลอีของค่าอัตสหสัมพันธ์ที่มีความคงทน หากนำมาเปรียบเทียบกับสัมประสิทธิ์การประมาณพันธะเชิงเส้นที่จำนวนอันดับเท่ากันเป็น 16 จะให้อัตราการรู้จำเฉลี่ยทุกกลุ่มเสียงพูดและและทุกช่วงของค่า SNR ที่ดีกว่า เฉลี่ยร้อยละ 0.91 แต่เนื่องจากสัมประสิทธิ์แอลพีแอลอีที่จำนวนอันดับมากขึ้นให้ค่าความผิดพลาดทางสเปกตรัมมากกว่า ทำให้เมื่อนำมาเปรียบเทียบที่จำนวนอันดับเท่ากันเป็น 32 จะให้อัตราการรู้จำเฉลี่ยที่น้อยกว่า

ดังนั้นสัมประสิทธิ์แอลพีแอลอีของค่าอัตสหสัมพันธ์ที่มีความคงทนจะให้ผลอัตราการรู้จำเฉลี่ยที่ดีกว่าสัมประสิทธิ์การประมาณพันธะเชิงเส้นเมื่อเลือกใช้จำนวนสัมประสิทธิ์ที่ค่าน้อยๆ แต่หากเลือกใช้จำนวนสัมประสิทธิ์น้อยเกินไปจะทำให้ได้ผลอัตราการรู้จำเฉลี่ยน้อยลงเช่นกัน หากจะเลือกค่าสัมประสิทธิ์แอลพีแอลอีของค่าอัตสหสัมพันธ์ที่มีความคงทนมาใช้เป็นค่าลักษณะสำคัญ จึงต้องหาจำนวนสัมประสิทธิ์ที่เหมาะสม

สำหรับค่าสัมประสิทธิ์เมลฟรีเควนซีเซปสตรอลของค่าอัตสหสัมพันธ์ที่มีความคงทน พบว่าให้อัตราการรู้จำเฉลี่ยกับกลุ่มเสียงวรรณยุกต์สามัญ เอก โท และตรี ที่ดีกว่าสัมประสิทธิ์เซปสตรอลบนความถี่เมล ส่วนกลุ่มเสียงวรรณยุกต์จัตวา หากพิจารณาเลือกใช้จำนวนอันดับเป็น 16 จะให้อัตราการรู้จำเฉลี่ยที่ดีกว่า

อัตราการรู้จำเฉลี่ยทุกกลุ่มเสียงพูด และทุกช่วงของค่า SNR ของแต่ละค่าลักษณะสำคัญเรียงจากมากไปน้อยเป็นดังนี้ AMFCC 32 : 58.37 , AMFCC 16 : 57.57 , MFCC 32 : 53.46 , MFCC 16 : 52.50 , LPC 32 : 36.45 , ALPLEC 16 : 36.42 , LPC 16 : 35.51 และ ALPLEC 32 : 34.54

ดังนั้นค่าสัมประสิทธิ์เมลฟรีเควนซีเซปสตรอลของของค่าอัตสหสัมพันธ์ที่มีความคงทน (AMFCC) อันดับ 32 ให้อัตราการรู้จำเฉลี่ยดีที่สุด และให้อัตราการรู้จำเฉลี่ยทุกกลุ่มเสียงพูด อย่างน้อยร้อยละ 62.67 เมื่อนำไปใช้ที่ค่า SNR ณ 15 dB

สำหรับการนำไปใช้ที่ค่า SNR = 15 dB ค่าสัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอลของของค่าอัตสหสัมพันธ์ที่มีความคงทน อันดับ 32 ให้อัตราการรู้จำเสียงพูดเฉลี่ยในกลุ่มเสียงวรรณยุกต์สามัญ เอก โท ตรี และจัตวา คิดเป็นร้อยละตามลำดับคือ 60.34 , 70.22 , 63.86 , 62.51 และ 56.40

อย่างไรก็ตาม ที่ค่า SNR มากๆ (SNR ≥ 25 dB) ข้อมูลสำคัญที่สุดของค่าอัตสหสัมพันธ์จะอยู่ที่เวลาศูนย์ และไม่นำมาใช้ในการหาค่าอัตสหสัมพันธ์ที่มีความคงทน ค่าลักษณะสำคัญที่หาจากค่าอัตสหสัมพันธ์ที่มีความคงทน เช่น สัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอลของค่าอัตสหสัมพันธ์ที่มีความคงทน จึงให้อัตราการรู้จำเฉลี่ยต่ำกว่าสัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอล

5.2 ข้อเสนอแนะสำหรับงานวิจัยในอนาคต

1. ค่าลักษณะสำคัญที่เป็นสัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอลของของค่าอัตสหสัมพันธ์ที่มีความคงทน ให้อัตราการรู้จำเฉลี่ยน้อยกว่าสัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอล ในสภาพแวดล้อมที่มีสัญญาณรบกวนน้อย หากหาแนวทางแก้ไขได้จะช่วยให้ค่าลักษณะสำคัญนี้เหมาะแก่การนำไปใช้งานยิ่งขึ้น
2. เทคนิคการปรับปรุงความคงทนของ ค่าสัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอล ของเสียงพูด เช่น การหาอนุพันธ์ การหาผลต่าง และการปรับค่าเฉลี่ย อาจช่วยให้สัมประสิทธิ์เมลฟรีเคว็นซีเซปสตรอลของของค่าอัตสหสัมพันธ์ที่มีความคงทน มีอัตราการรู้จำเฉลี่ยมากขึ้น
3. ควรมีการพัฒนาโปรแกรมสำหรับช่วยในการฝึกฝนและทดสอบ เสียงพูดที่มีการแบ่งกลุ่มเสียงพูดจำนวนมาก จะช่วยให้การทดสอบค่าลักษณะสำคัญอื่นๆที่พัฒนาขึ้นใหม่ให้มีความคงทนต่อสัญญาณรบกวน เป็นไปอย่างราบรื่น

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

รายการอ้างอิง

ภาษาไทย

ชัย วุฒิวิวัฒน์ชัย. การรู้จำเสียงคำไทยหลายพยางค์แบบไม่ขึ้นกับผู้พูดโดยใช้เทคนิคแบบฟัซซีและนิวรอลเน็ตเวิร์ก. วิทยานิพนธ์ปริญญามหาบัณฑิต สาขาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2540.

ธีระ ภัทราพรนันท์. การรู้จำเสียงพูดสระภาษาไทยโดดๆไม่ขึ้นกับผู้พูดโดยการวัดสเปกตรัมดิสแตนซ์และใช้ไดนามิกไทม์วาร์ปปิง. วิทยานิพนธ์ปริญญามหาบัณฑิต สาขาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2538.

ระพีพัฒน์ เพ็ญศิริ. การรู้จำเสียงพูดตัวเลขไทยโดยไม่ขึ้นกับผู้พูดโดยการใช้ไดนามิกไทม์วาร์ปปิง. วิทยานิพนธ์ปริญญามหาบัณฑิต สาขาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2538.

วิศรุต อาขุบุตร. ระบบรู้จำคำไทยหลายพยางค์แบบไม่ขึ้นกับผู้พูดโดยใช้แบบจำลองฮิดเดนมาร์คอฟ. วิทยานิพนธ์ปริญญามหาบัณฑิต สาขาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2539.

วุฒิพงษ์ พรสุขจันทรา. การรู้จำเสียงตัวเลขภาษาไทยแบบไม่ขึ้นกับผู้พูดโดยใช้แอลพีซีและนิวรอลเน็ตเวิร์กแบบแบ็กพรอพาเกชัน. วิทยานิพนธ์ปริญญามหาบัณฑิต สาขาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2539.

เสาวลักษณ์ อารีย์พงศา. การรู้จำเสียงพูดตัวเลขเป็นภาษาไทยแบบไม่ขึ้นกับผู้พูดโดยวิธี ฮิดเดน มาร์คอฟ โมเดลและเวกเตอร์ควอนไตซ์เซชัน. วิทยานิพนธ์ปริญญามหาบัณฑิต สาขาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2538.

เอกฤทธิ์ มณีน้อย. การรู้จำหน่วยเสียงสระภาษาไทยโดยใช้โครงข่ายประสาทเทียม. วิทยานิพนธ์ปริญญามหาบัณฑิต สาขาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2541.

ภาษาอังกฤษ

D. Mansour and B.H. Juang. "The short-time modified coherence representation and noisy speech recognition". IEEE Transaction on Acoustics, Speech and Signal Processing Vol. 37, No. 6. (June 1989): 795-804.

Deller, Jr. John R., Hansen John H.L., and Proakis John G. Discrete-Time Processing of Speech Signals. New York: Macmillan, 1993.

Furui, S. Digital Speech Processing, Synthesis and Recognition. New York and Basel: Marcel Dekker, 1989.

H. Hermansky and N. Morgan. "RASTA processing of speech". IEEE Transactions on Speech and Audio Processing Vol. 24 (October 1994): 578-589.

Hayes Manson H. Statistical Digital Signal Processing and Modeling. John Wiley & Sons, 1996.

I. Lecomte, M. Lever, J. Boudy and A. Tassy. "Car Noise Processing for Speech Input", IEEE International Conference on ICASSP Vol. 1.1. (1989): 512-515.

J. Hernando and C. Nadeu. "Speech Recognition in Noisy Car Environment Based on OSALPC Representation and Robust Similarity Measuring Techniques". IEEE International Conference on ICASSP Vol ii (1994): II/69-II/72.

K.F. Wong, S.H. Leung, and H.C. Ng. "An Eigendecomposition Based Two Sided Linear Prediction Model for Robust Speech Recognition". 1994 International Symposium on Speech, Image Processing and Neural Networks Vol. 1.1 (April 1994): 249-252.

Oppenhiem, A. V., Schafer, R. W. Discrete-Time Signal Processing. Prentice Hall 1989.

R.J. Mammone, X. Zhang and R.P. Ramachandran. "Robust Speaker recognition: A Feature-Based Approach". IEEE Signal Processing Magazine Vol. 135 (September 1996): 58-71.

S. Varho and P. Alku. "A New Predictive Method for All-Pole Modelling of Speech Spectra with A Compressed Set of Parameters". Proceedings of the 1999 IEEE International Symposium (ISCAS 99) Vol. 3 (1999): 126-129.

Tolba, H. and O'Shaughnessay, D. "Automatic speech recognition based on cepstral coefficient and mel-based discrete energy operator". Proceedings of the 1998 IEEE International Conference on Acoustic, Speech and Signal Processing Vol. 2 (1998): 973-976.

Tuzan, O.B. Demirekle, M. and Nakiboglu, K.B. "Comparison of parametric and non-parametric representations of speech for recognition". Proceedings of the 7th Mediterranean Electrotechnical Conference Vol. 1 (1994): 65-68.

ประวัติผู้เขียนวิทยานิพนธ์

นายเจษฎา กานต์ประชา เกิดเมื่อวันที่ 4 กรกฎาคม พ.ศ. 2520 ที่อำเภอหาดใหญ่ จังหวัดสงขลา สำเร็จการศึกษาปริญญาวิศวกรรมศาสตร์บัณฑิต ภาควิชาวิศวกรรมไฟฟ้า จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2541 และเข้าศึกษาต่อในหลักสูตรวิศวกรรมศาสตร์มหาบัณฑิต แขนงวิชากรรมวิธีสัญญาณดิจิตอล ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2542

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย