ค่าลักษณะสำคัญแบบคงทนต่อสัญญาณรบกวนของการรู้จำเสียงพูดภาษาไทย
โดยใช้เทคนิคลดสัญญาณรบกวนแบบปรับตัวได้

นาย ณัฐพล อุ่นศรี

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

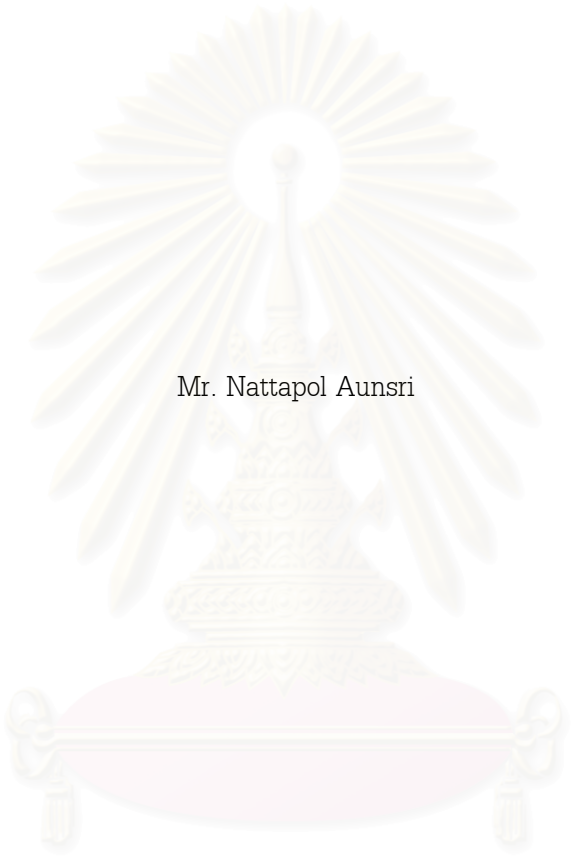
วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมไฟฟ้า ภาควิชาวิศวกรรมไฟฟ้า
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2545

ISBN 974-17-2126-9

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

NOISE ROBUST FEATURE OF THAI SPEECH RECOGNITION USING ADAPTIVE NOISE REDUCTION TECHNIQUE

Mr. Nattapol Aunsri



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 974-17-2126-9

หัวข้อวิทยานิพนธ์	ค่าลักษณะสำคัญแบบคงทนต่อสัญญาณรบกวนของการรู้จำเสียงพูดภาษาไทย โดยใช้เทคนิคลดสัญญาณรบกวนแบบปรับตัวได้
โดย	นาย ณัฐพล อุ่นศรี
สาขาวิชา	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

ณัฐพล อุ่นศรี : ค่าลักษณะสำคัญแบบคงทนต่อสัญญาณรบกวนของการรู้จำเสียงพูดภาษาไทยโดยใช้เทคนิคลดสัญญาณรบกวนแบบปรับตัวได้. (Noise Robust Feature of Thai Speech Recognition Using Adaptive Noise Reduction Technique) อ. ที่ปรึกษา : รศ. ดร.สมชาย จิตะพันธ์กุล, 77 หน้า. ISBN 974-17-2126-9.

วิทยานิพนธ์ฉบับนี้มีวัตถุประสงค์เพื่อนำเสนอค่าลักษณะสำคัญแบบคงทนต่อสัญญาณรบกวนเกาส์เซียนขาวค่าเฉลี่ยศูนย์แบบบวกโดยใช้เทคนิคลดสัญญาณรบกวนแบบปรับตัวได้โดยใช้ค่าเจาะจงต่ำสุดของเมตริกซ์อัตสหสัมพันธ์ของสัญญาณเสียงพูดเป็นค่าพารามิเตอร์ในการปรับลดสัญญาณรบกวนเพื่อให้ระบบรู้จำเสียงพูดภาษาไทยมีความคงทนเมื่อใช้งานในสภาพแวดล้อมที่มีสัญญาณรบกวนโดยใช้แบบจำลองฮิดเดนมาร์คอฟชนิดต่อเนื่องสำหรับการฝึกฝนและทดสอบ ค่าลักษณะสำคัญพื้นฐานที่ใช้เป็นตัวเปรียบเทียบประสิทธิภาพของค่าลักษณะสำคัญแบบคงทนคือ สัมประสิทธิ์การประมาณพันธะเชิงเส้น และสัมประสิทธิ์เซปตรอล ชุดคำศัพท์ประกอบไปด้วยคำศัพท์จำนวน 50 คำแบ่งเป็น 6 ชุด ได้แก่ชุดคำศัพท์ตัวเลขศูนย์ถึงเก้าจำนวน 10 คำ ชุดคำศัพท์เสียงวรรณยุกต์ไทยเสียงสามัญ จำนวน 8 คำ ชุดคำศัพท์เสียงวรรณยุกต์ไทยเสียงเอก จำนวน 8 คำ ชุดคำศัพท์เสียงวรรณยุกต์ไทยเสียงโท จำนวน 8 คำ ชุดคำศัพท์เสียงวรรณยุกต์ไทยเสียงตรี จำนวน 8 คำ และชุดคำศัพท์เสียงวรรณยุกต์ไทยเสียงจัตวา จำนวน 8 คำ ในการวิจัยนี้ได้ทดลองปรับเปลี่ยนจำนวนสถานะและจำนวน Gaussian mixture ของระบบเพื่อหาระบบที่เหมาะสมสำหรับการรู้จำเสียงพูด อีกทั้งยังทำการทดสอบเพื่อวิเคราะห์ผลของการทำเน้นล่วงหน้า ผลการทดสอบอัตราการรู้จำแบบไม่ขึ้นกับผู้พูดของชุดคำศัพท์ตัวเลขศูนย์ถึงเก้าเมื่อค่าอัตราส่วนสัญญาณต่อสัญญาณรบกวนเป็น 0 เดซิเบลโดยใช้ค่าลักษณะสำคัญแบบคงทนต่อสัญญาณรบกวนมีค่าร้อยละ 38.00 เมื่ออัตราการรู้จำก่อนลดสัญญาณรบกวนมีค่าร้อยละ 10.42 อัตราการรู้จำของชุดคำศัพท์เสียงวรรณยุกต์ไทยเสียงสามัญ เอก โท ตรี และจัตวาเมื่อค่าอัตราส่วนสัญญาณต่อสัญญาณรบกวนเป็น 0 เดซิเบลโดยใช้ค่าลักษณะสำคัญแบบคงทนต่อสัญญาณรบกวนมีค่าร้อยละ 31.65 33.64 38.69 34.26 และ 32.44 เมื่ออัตราการรู้จำก่อนลดสัญญาณรบกวนมีค่าร้อยละ 12.27 14.49 16.25 13.98 และ 13.75 ตามลำดับ

อัตราการรู้จำเฉลี่ยสูงสุดของเสียงตัวเลขภาษาไทย โดยคำนวณจากการนำอัตราการรู้จำของแต่ละค่าลักษณะสำคัญในแต่ละอันดับสำหรับทุกค่า SNR มาทำการเฉลี่ยเป็นร้อยละ 62.95 ขณะที่อัตราการรู้จำเฉลี่ยสูงสุดก่อนปรับลดสัญญาณรบกวนเป็นร้อยละ 36.65 และอัตราการรู้จำเฉลี่ยสูงสุดของเสียงวรรณยุกต์ไทยเสียงสามัญ เอก โท ตรี และจัตวา โดยคำนวณจากการนำอัตราการรู้จำของแต่ละค่าลักษณะสำคัญในแต่ละอันดับสำหรับทุกค่า SNR มาทำการเฉลี่ยเป็นร้อยละ 59.85 62.67 57.96 50.12 และ 55.33 ขณะที่อัตราการรู้จำเฉลี่ยสูงสุดก่อนปรับลดสัญญาณรบกวนเป็นร้อยละ 28.55 35.69 32.56 29.82 และ 34.26 ตามลำดับ

อัตราการรู้จำเฉลี่ยรวมสูงสุดของระบบเมื่อใช้ค่าลักษณะสำคัญแบบคงทนต่อสัญญาณรบกวนเป็นร้อยละ 57.68 ขณะที่อัตราการรู้จำเฉลี่ยรวมสูงสุดของระบบก่อนปรับลดสัญญาณรบกวนเป็นร้อยละ 32.28 ซึ่งเพิ่มขึ้นร้อยละ 78.69

ภาควิชา วิศวกรรมไฟฟ้า ลายมือชื่อนิสิต

สาขาวิชา วิศวกรรมไฟฟ้า ลายมือชื่ออาจารย์ที่ปรึกษา

ปีการศึกษา 2545

4270321221 : MAJOR ELECTRICAL ENGINEERING

KEYWORD : ADAPTIVE NOISE REDUCTION / NOISY SPEECH / NOISE ROBUST FEATURE / THAI SPEECH RECOGNITION

NATTAPOL AUNSRI : NOISE ROBUST FEATURE OF THAI SPEECH RECOGNITION USING ADAPTIVE NOISE REDUCTION TECHNIQUE.

THESIS ADVISOR : ASSOC.PROF. DR. SOMCHAI JITAPANKUL, Dr.Ing.,

77 pp. ISBN 974-17-2126-9.

The objective of this thesis is to propose robust feature to additive zero mean white Gaussian noise using adaptive noise reduction technique in order to obtain robust Thai speech recognition system in noisy environment. Continuous Hidden Markov Model is used for training and testing. Linear prediction coefficients and cepstrum coefficients are used to compare the performance of the robust features. The 50-word vocabularies are subdivided into six sets comprising middle, low, high, falling and rising tones for five Thai tonal levels, 8 words in each set, and last set consisting of 10-Thai numeric words, zero to nine. In this research, the number of states and Gaussian mixtures in continuous hidden Markov model are varied to obtain the optimal system for speech recognition, and also, analysis of preemphasis effect. The recognition rate of this speaker-independent recognition system of numeric words when signal to noise ratio is 0 decibel is 38.00 percent and 31.65, 33.64, 38.69, 34.26 and 32.44 percent respectively for the five Thai tonal levels, namely, middle, low, high, falling and rising tones when using the robust features and 12.27 percent for numeric word, and 12.27, 14.49, 16.25, 13.98 and 13.75 percent for the five Thai tonal levels respectively when using non-noise reduction features.

The maximum of average recognition rate that is calculated from all SNR for each order feature. The maximum of average recognition rate is 69.25 percent for numeric words and for five Thai tonal levels middle, low, high, falling and rising tones, are 59.85, 62.67, 57.96, 50.12 and 55.33 percent respectively when using the robust features and 36.65 for numeric words, and 28.55, 35.69, 32.56, 29.82 and 34.26 percent for five Thai tonal levels, respectively when using non-noise reduction features.

The maximum recognition rate of system when using noise robust features is 57.68 percent and 32.28 percent when using non-noise reduction features. Noise robust features improve the recognition rate of Thai speech recognition 78.69 percent.

Department Electrical Engineering ..... Student s signature

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

Academic year 2002

กิตติกรรมประกาศ

การดำเนินงานวิจัยนี้และการจัดทำวิทยานิพนธ์ฉบับนี้ ผู้วิจัยของกราบขอบพระคุณ รองศาสตราจารย์ ดร.สมชาย จิตะพันธ์กุล อาจารย์ที่ปรึกษา ที่ได้กรุณาให้คำแนะนำในการดำเนินงานวิจัย ทั้งในแง่ของการศึกษาและสังคม ตลอดจนแรงกระตุ้นผู้วิจัยเสมอมา

ขอขอบคุณ คุณเจษฎา กานต์ประชา และบุคลากรห้องปฏิบัติการกรรมวิธีเชิงเลขที่ได้ช่วยกันพัฒนาเครื่องมือในการวิจัยมาโดยตลอดและต่อเนื่องอันเป็นประโยชน์ยิ่งกับผู้วิจัย

สุดท้ายนี้ขอกราบขอบพระคุณ อากู๋ คุณป้า ที่ได้มอบความรัก ความปรารถนาดี ตลอดจนพลังในการดำเนินงานวิจัยจวบจนกระทั่งวิทยานิพนธ์สำเร็จลุล่วงด้วยดี

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญ

หน้า

บทคัดย่อภาษาไทย.....ง
บทคัดย่อภาษาอังกฤษ.....จ
กิตติกรรมประกาศ.....ฉ
สารบัญ.....ช
สารบัญตาราง.....ญ
สารบัญรูป.....ฏ
คำอธิบายศัพท์.....ฒ
บทที่ 1 บทนำ.....1
1.1 แนวเหตุผล.....1
1.2 ปัญหาของค่าลักษณะสำคัญสำหรับระบบรู้จำเสียงพูดเมื่อมีสัญญาณรบกวน และวิธีแก้ปัญหาที่มีผู้เสนอขึ้น.....4
1.3 วัตถุประสงค์.....5
1.4 เป้าหมายและขอบเขตของวิทยานิพนธ์.....5
1.5 ขั้นตอนและวิธีการดำเนินการวิจัย.....5
1.6 ประโยชน์ที่คาดว่าจะได้รับ.....5
1.7 ภาพรวมของวิทยานิพนธ์.....6
บทที่ 2 หลักการ ทฤษฎี ขั้นตอนวิธีการและงานวิจัยที่เกี่ยวข้อง.....7
2.1 การประมวลผลสัญญาณเบื้องต้น.....7
2.2 การสกัดค่าลักษณะสำคัญ.....8
2.2.1 แบบจำลองการประมาณพันธะเชิงเส้น.....9
2.2.2 สัมประสิทธิ์เซปตรอล.....14
2.3 การวิเคราะห์ค่าลักษณะสำคัญเมื่อมีสัญญาณรบกวนและการหาค่าลักษณะสำคัญที่คงทนต่อสัญญาณรบกวน.....14
2.3.1 สัญญาณรบกวนแบบขาว.....14
2.3.2 การวิเคราะห์ผลของสัญญาณรบกวนที่มีต่อค่าลักษณะสำคัญ.....14
2.3.3 ค่าลักษณะสำคัญที่คงทนต่อสัญญาณรบกวนบนพื้นฐานของการประมาณพันธะเชิงเส้น.....16
2.4 วิธีการเชิงตัวเลขที่เกี่ยวข้อง.....18
2.4.1 วิธีการแบ่งส่วน.....18
2.4.2 ค่าเจาะจง.....20

สารบัญ (ต่อ)

หน้า

2.4.3 วิธีการประมาณค่าเจาะจงด้วยวิธีกำลัง....21
2.5 งานวิจัยที่เกี่ยวข้อง....22
บทที่ 3 ขั้นตอนวิธีในการดำเนินการวิจัย....24
3.1 รายละเอียดการเตรียมตัวอย่างข้อมูลเสียงพูด....24
3.1.1 ชุดคำศัพท์....24
3.1.2 ขั้นตอนการตัดพยางค์....25
3.1.3 ขั้นตอนการเพิ่มเสียงสัญญาณรบกวน....26
3.2 รายละเอียดขั้นตอนการวิเคราะห์และสกัดค่าลักษณะสำคัญ....27
3.2.1 การประมวลผลสัญญาณเบื้องต้น....27
3.2.1.1 กรรมวิธีเน้นล่วงหน้า....27
3.2.1.2 กรรมวิธีวางกรอบขนาดสัญญาณ....27
3.2.2 การสกัดค่าลักษณะสำคัญ....28
3.2.2.1 สัมประสิทธิ์การประมาณพันธะเชิงเส้น....28
3.2.2.2 สัมประสิทธิ์เซปตรอล....28
3.2.2.3 สัมประสิทธิ์การประมาณพันธะเชิงเส้นจากเทคนิคลดสัญญาณรบกวนแบบปรับตัวได้แบบปรับตัวได้....29
3.2.2.4 สัมประสิทธิ์เซปสตรอลจากเทคนิคลดสัญญาณรบกวนแบบปรับตัวได้....29
3.3 รายละเอียดขั้นตอนการปรับลดสัญญาณรบกวนแบบปรับตัวได้....29
3.4 รายละเอียดขั้นตอนการฝึกฝนและการทดสอบ....31
บทที่ 4 ผลการวิจัยและการวิเคราะห์ผล....32
4.1 ขั้นตอนวิธีการรู้จำเสียงพูดภาษาไทย....32
4.2 การทดลองเพื่อหาระบบที่เหมาะสมสำหรับการรู้จำ....33
4.3 การทดลองเพื่อวิเคราะห์ผลของการทำเน้นล่วงหน้า....34
4.4 ผลการรู้จำเสียงพูดภาษาไทย....35
4.5 วิเคราะห์ผลการวิจัย....57
บทที่ 5 บทสรุปและข้อเสนอแนะ....62
5.1 สรุปผลการวิจัย....62
5.2 ข้อเสนอแนะสำหรับงานวิจัยในอนาคต....65
รายการอ้างอิง....66

สารบัญ (ต่อ)

หน้า

ภาคผนวก ก รายละเอียดอัตราการรู้จำเสียงพูดภาษาไทยเนื่องจากผลของการแปรค่าสถานะ และ Gaussian Mixture....69
ภาคผนวก ข รายละเอียดเปรียบเทียบอัตราการรู้จำสำหรับคุณลักษณะสำคัญที่ผ่านและไม่ผ่านกรรมวิธีเน้นล่วงหน้า....70
ภาคผนวก ค รายละเอียดผลการรู้จำเสียงพูดภาษาไทย....71
ประวัติผู้เขียนวิทยานิพนธ์....77

จุฬาลงกรณ์มหาวิทยาลัย

สารบัญตาราง

หน้า

ตารางที่ 3.1 ชุดคำศัพท์ที่เลือกมาทดสอบแบ่งตามกลุ่มเสียงวรรณยุกต์....25
ตารางที่ 4.1 ผลของระบบต่ออัตราการรู้จำเฉลี่ยรวม....34
ตารางที่ 5.1 อัตราการรู้จำเฉลี่ยสูงสุด เมื่อค่า SNR เป็น 0 dB....63
ตารางที่ 5.2 อัตราการรู้จำเฉลี่ยสูงสุดของแต่ละชุดคำศัพท์....64
ตารางที่ ก.1 รายละเอียดอัตราการรู้จำเสียงพูดภาษาไทยเนื่องจากผลผลของการแปรค่าสถานะ และ Gaussian Mixture....69
ตารางที่ ข.1 เปรียบเทียบอัตราการรู้จำสำหรับคุณลักษณะสำคัญที่ผ่านและไม่ผ่านกรรมวิธีเน้นล่วงหน้า....70
ตารางที่ ค.1 ผลการรู้จำเสียงตัวเลขภาษาไทยสำหรับค่าลักษณะสำคัญ LPC LPC-CEP อันดับ 10 12 14 และ 16....71
ตารางที่ ค.2 ผลการรู้จำเสียงตัวเลขภาษาไทยสำหรับค่าลักษณะสำคัญ ANR และ ANR-CEP อันดับ 10 12 14 และ 16....71
ตารางที่ ค.3 ผลการรู้จำเสียงวรรณยุกต์ไทยเสียงสามัญสำหรับค่าลักษณะสำคัญ LPC และ LPC-CEP อันดับ 10 12 14 และ 16....72
ตารางที่ ค.4 ผลการรู้จำเสียงวรรณยุกต์ไทยเสียงสามัญสำหรับค่าลักษณะสำคัญ ANR-LPC และ ANR-CEP อันดับ 10 12 14 และ 16....72
ตารางที่ ค.5 ผลการรู้จำเสียงวรรณยุกต์ไทยเสียงเอกสำหรับค่าลักษณะสำคัญ LPC และ LPC-CEP อันดับ 10 12 14 และ 16....73
ตารางที่ ค.6 ผลการรู้จำเสียงวรรณยุกต์ไทยเสียงเอกสำหรับค่าลักษณะสำคัญ ANR-LPC และ ANR-CEP อันดับ 10 12 14 และ 16....73
ตารางที่ ค.7 ผลการรู้จำเสียงวรรณยุกต์ไทยเสียงโทสำหรับค่าลักษณะสำคัญ LPC และ LPC-CEP อันดับ 10 12 14 และ 16....74
ตารางที่ ค.8 ผลการรู้จำเสียงวรรณยุกต์ไทยเสียงโทสำหรับค่าลักษณะสำคัญ ANR-LPC และ ANR-CEP อันดับ 10 12 14 และ 16....74
ตารางที่ ค.9 ผลการรู้จำเสียงวรรณยุกต์ไทยเสียงตรีสำหรับค่าลักษณะสำคัญ LPC และ LPC-CEP อันดับ 10 12 14 และ 16....75
ตารางที่ ค.10 ผลการรู้จำเสียงวรรณยุกต์ไทยเสียงตรีสำหรับค่าลักษณะสำคัญ ANR-LPC และ ANR-CEP อันดับ 10 12 14 และ 16....75
ตารางที่ ค.11 ผลการรู้จำเสียงวรรณยุกต์ไทยเสียงจัตวาสำหรับค่าลักษณะสำคัญ LPC และ LPC-CEP อันดับ 10 12 14 และ 16....76

สารบัญตาราง (ต่อ)

หน้า

ตารางที่ ค.12 ผลการรู้จำเสียงวรรณยุกต์ไทยเสียงจัตวาสำหรับค่าลักษณะสำคัญ ANR-LPC และ ANR-CEP อันดับ 10 12 14 และ 16.....76

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญรูป

หน้า

รูปที่ 1.1 อัตราการรู้จำเสียงตัวเลขภาษาไทยเมื่อมีสัญญาณรบกวน
โดยใช้ค่าลักษณะสำคัญสัมประสิทธิ์การประมาณพันธะเชิงเส้น....2
รูปที่ 1.2 รายละเอียดขั้นตอนการรู้จำคำพูดภาษาไทย....3
รูปที่ 2.1 ฟังก์ชันกรอบชนิด Hamming Window....8
รูปที่ 2.2 แบบจำลองการประมาณพันธะเชิงเส้นสำหรับเสียงพูด....10
รูปที่ 2.3 แบบจำลองสังเคราะห์เสียงพูดจากแบบจำลองการประมาณพันธะเชิงเส้น....10
รูปที่ 2.4 แสดงขั้นตอนวิธีการแบ่งส่วน....19
รูปที่ 2.5 แสดงการประสมเชิงเส้นแบบให้น้ำหนักของการแจกแจงแบบเกาส์....22
รูปที่ 3.1 แผนภูมิเส้นระดับพลังงานของเสียงพูด "มือซ้าย"....26
รูปที่ 4.1 อัตราการรู้จำเสียงพูดภาษาไทยเนื่องจากผลผลของการแปรค่าสถานะและ Gaussian Mixture....33
รูปที่ 4.2 ผลการทำเน้นล่วงหน้าต่ออัตราการรู้จำ....35
รูปที่ 4.3 อัตราการรู้จำเสียงตัวเลขภาษาไทยด้วยลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC และ ANR-CEP อันดับ 10....37
รูปที่ 4.4 อัตราการรู้จำเสียงตัวเลขภาษาไทยด้วยลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC และ ANR-CEP อันดับ 12....37
รูปที่ 4.5 อัตราการรู้จำเสียงตัวเลขภาษาไทยด้วยลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC และ ANR-CEP อันดับ 14....38
รูปที่ 4.6 อัตราการรู้จำเสียงตัวเลขภาษาไทยด้วยลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC และ ANR-CEP อันดับ 16....38
รูปที่ 4.7 อัตราการรู้จำเสียงวรรณยุกต์ไทยเสียงสามัญด้วยลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC และ ANR-CEP อันดับ 10....39
รูปที่ 4.8 อัตราการรู้จำเสียงวรรณยุกต์ไทยเสียงสามัญด้วยลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC และ ANR-CEP อันดับ 12....39
รูปที่ 4.9 อัตราการรู้จำเสียงวรรณยุกต์ไทยเสียงสามัญด้วยลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC และ ANR-CEP อันดับ 14....40
รูปที่ 4.10 อัตราการรู้จำเสียงวรรณยุกต์ไทยเสียงสามัญด้วยลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC และ ANR-CEP อันดับ 16....40
รูปที่ 4.11 อัตราการรู้จำเสียงวรรณยุกต์ไทยเสียงเอกด้วยลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC และ ANR-CEP อันดับ 10....41

สารบัญรูป (ต่อ)

หน้า

รูปที่ 4.12 อัตราการรู้จำเสียงวรรณยุกต์ไทยเสียงเอกด้วยลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC และ ANR-CEP อันดับ 12 41
รูปที่ 4.13 อัตราการรู้จำเสียงวรรณยุกต์ไทยเสียงเอกด้วยลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC และ ANR-CEP อันดับ 14 42
รูปที่ 4.14 อัตราการรู้จำเสียงวรรณยุกต์ไทยเสียงเอกด้วยลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC และ ANR-CEP อันดับ 16 42
รูปที่ 4.15 อัตราการรู้จำเสียงวรรณยุกต์ไทยเสียงโทด้วยลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC และ ANR-CEP อันดับ 10 43
รูปที่ 4.16 อัตราการรู้จำเสียงวรรณยุกต์ไทยเสียงโทด้วยลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC และ ANR-CEP อันดับ 12 43
รูปที่ 4.17 อัตราการรู้จำเสียงวรรณยุกต์ไทยเสียงโทด้วยลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC และ ANR-CEP อันดับ 14 44
รูปที่ 4.18 อัตราการรู้จำเสียงวรรณยุกต์ไทยเสียงโทด้วยลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC และ ANR-CEP อันดับ 16 44
รูปที่ 4.19 อัตราการรู้จำเสียงวรรณยุกต์ไทยเสียงตรีด้วยลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC และ ANR-CEP อันดับ 10 45
รูปที่ 4.20 อัตราการรู้จำเสียงวรรณยุกต์ไทยเสียงตรีด้วยลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC และ ANR-CEP อันดับ 12 45
รูปที่ 4.21 อัตราการรู้จำเสียงวรรณยุกต์ไทยเสียงตรีด้วยลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC และ ANR-CEP อันดับ 14 46
รูปที่ 4.22 อัตราการรู้จำเสียงวรรณยุกต์ไทยเสียงตรีด้วยลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC และ ANR-CEP อันดับ 16 46
รูปที่ 4.23 อัตราการรู้จำเสียงวรรณยุกต์ไทยเสียงจัตวาด้วยลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC และ ANR-CEP อันดับ 10 47
รูปที่ 4.24 อัตราการรู้จำเสียงวรรณยุกต์ไทยเสียงจัตวาด้วยลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC และ ANR-CEP อันดับ 12 47
รูปที่ 4.25 อัตราการรู้จำเสียงวรรณยุกต์ไทยเสียงจัตวาด้วยลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC และ ANR-CEP อันดับ 14 48
รูปที่ 4.26 อัตราการรู้จำเสียงวรรณยุกต์ไทยเสียงจัตวาด้วยลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC และ ANR-CEP อันดับ 16 48

สารบัญรูป (ต่อ)

หน้า

รูปที่ 4.27 เปรียบเทียบผลการรู้จำเสียงตัวเลขภาษาไทยโดยเฉลี่ยทุกค่า SNR สำหรับค่าลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC, ANR-CEP อันดับ 10 12 14 และ 16.....49
รูปที่ 4.28 เปรียบเทียบผลการรู้จำเสียงวรรณยุกต์ไทย เสียงสามัญโดยเฉลี่ยทุกค่า SNR สำหรับค่าลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC, ANR-CEP อันดับ 10 12 14 และ 16.....49
รูปที่ 4.29 เปรียบเทียบผลการรู้จำเสียงวรรณยุกต์ไทย เสียงเอกโดยเฉลี่ยทุกค่า SNR สำหรับค่าลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC, ANR-CEP อันดับ 10 12 14 และ 16.....50
รูปที่ 4.30 เปรียบเทียบผลการรู้จำเสียงวรรณยุกต์ไทย เสียงโทโดยเฉลี่ยทุกค่า SNR สำหรับค่าลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC, ANR-CEP อันดับ 10 12 14 และ 16.....50
รูปที่ 4.31 เปรียบเทียบผลการรู้จำเสียงวรรณยุกต์ไทย เสียงตรีโดยเฉลี่ยทุกค่า SNR สำหรับค่าลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC, ANR-CEP อันดับ 10 12 14 และ 16.....51
รูปที่ 4.32 เปรียบเทียบผลการรู้จำเสียงวรรณยุกต์ไทย เสียงจัตวาโดยเฉลี่ยทุกค่า SNR สำหรับค่าลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC, ANR-CEP อันดับ 10 12 14 และ 16.....51
รูปที่ 4.33 แสดงอัตราการรู้จำเฉลี่ยรวมทุกชุดคำศัพท์สำหรับค่า Clean speech.....52
รูปที่ 4.34 แสดงอัตราการรู้จำเฉลี่ยรวมทุกชุดคำศัพท์สำหรับค่า SNR 30 dB.....52
รูปที่ 4.35 แสดงอัตราการรู้จำเฉลี่ยรวมทุกชุดคำศัพท์สำหรับค่า SNR 25 dB.....53
รูปที่ 4.36 แสดงอัตราการรู้จำเฉลี่ยรวมทุกชุดคำศัพท์สำหรับค่า SNR 20 dB.....53
รูปที่ 4.37 แสดงอัตราการรู้จำเฉลี่ยรวมทุกชุดคำศัพท์สำหรับค่า SNR 15 dB.....54
รูปที่ 4.38 แสดงอัตราการรู้จำเฉลี่ยรวมทุกชุดคำศัพท์สำหรับค่า SNR 10 dB.....54
รูปที่ 4.39 แสดงอัตราการรู้จำเฉลี่ยรวมทุกชุดคำศัพท์สำหรับค่า SNR 5 dB.....55
รูปที่ 4.40 แสดงอัตราการรู้จำเฉลี่ยรวมทุกชุดคำศัพท์สำหรับค่า SNR 0 dB.....55
รูปที่ 4.41 แสดงอัตราการรู้จำเฉลี่ยรวมของระบบ.....56

จุฬาลงกรณ์มหาวิทยาลัย

สารบัญคำศัพท์

ปรับตัวได้	adaptive
ขั้นตอนวิธีการ	algorithm
ระบบที่มีแต่โพล	all pole system
อัตสหสัมพันธ์	autocorrelation
แบบจำลองออโตรีเกรสซีฟ	autoregressive model (AR model)
การมอดูเลตแบบสเปกตรัมแผ่	spread spectrum modulation
สัมประสิทธิ์เซปตรอล	cepstral coefficient
สัญญาณเสียงพูดที่ไม่มีสัญญาณรบกวน	clean speech
เสียงพูดแบบต่อเนื่อง	continuous speech
ความแปรปรวนร่วม	covariance
ลักษณะสำคัญ	feature
การสกัดลักษณะสำคัญ	feature extraction
การวัดค่าลักษณะสำคัญ	feature measurement
วงจรกรองดิจิตอลอันดับหนึ่ง	first-order digital filter
ส่วนย่อย, กรอบ	frame
แบบจำลองฮิดเดนมาร์คอฟ	hidden markov model
วนซ้ำ	iterative
การประมาณพันธะเชิงเส้น	linear prediction
สัมประสิทธิ์การประมาณพันธะเชิงเส้น	linear prediction coefficient
สัญญาณรบกวน	noise
สัญญาณเสียงพูดที่มีสัญญาณรบกวน	noisy speech
อันดับ	order
แบบจำลองที่ใช้พารามิเตอร์	parametric model
รูปแบบ	pattern
การจำแนกรูปแบบ	pattern classification
การเน้นล่วงหน้า	preemphasis
กรรมวิธีประมวลผลเบื้องต้น	preprocessing
การรู้จำ	recognition
อัตราการรู้จำ	recognition rate
การวางกรอบขนาดสัญญาณ	smoothing window
เสียงพูด	speech

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สถานะ	state
ไม่เปลี่ยนแปลงตามเวลา	stationary
ข้อมูลชุดทดสอบ	testing data
จุดเริ่มเปลี่ยน	threshold
การฝึกฝน	training
ข้อมูลชุดฝึกฝน	training data
ฟังก์ชันถ่ายโอน	transfer function
เสียงอโฆษะ	unvoice
เสียงโฆษะ	voice
สัญญาณรบกวนแบบขาว	white noise
ฟังก์ชันกรอบ	window function

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 1

บทนำ

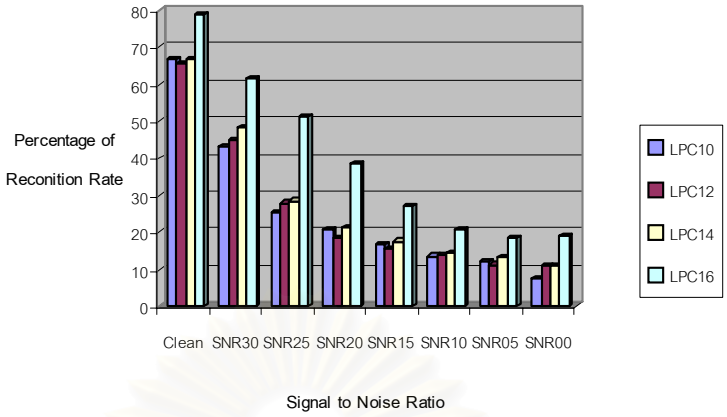
1.1 แนวเหตุผล

ระบบรู้จำเสียงพูด (Speech recognition system) มีความสำคัญมากในปัจจุบัน เพราะการพูดเป็นการติดต่อสื่อสารพื้นฐานของมนุษย์ ในปัจจุบัน คอมพิวเตอร์ได้เข้ามามีบทบาทมากในชีวิตประจำวัน และประสิทธิภาพของคอมพิวเตอร์ก็สูงขึ้น ดังนั้นจึงได้สร้างระบบรู้จำเสียงพูดที่ใช้งานได้บนคอมพิวเตอร์ และมีการพัฒนาระบบให้มีอัตราการรู้จำ (Recognition rate) ให้สูงขึ้นตามลำดับ

แท้จริงแล้ว ระบบรู้จำเสียงพูดแบบอัตโนมัติ (Automatic Speech Recognition) ด้วยเครื่องจักรเป็นเป้าหมายสูงสุดสำหรับงานวิจัยด้านการรู้จำเสียงพูดมากว่า 40 ปีแล้ว เนื่องจากได้รับแรงบันดาลใจอันสำคัญยิ่งจากนิยายวิทยาศาสตร์อันเลื่องชื่อของ Stanley Kubrick เรื่อง 2001-Space Odysey หรือไม่ว่าจะเป็นภาพยนต์ระดับมหากาพย์ของ George Lucus เรื่อง Star Wars ทุกตอน เมื่อหุ่นยนต์ R2D2 สามารถตอบสนองกับมนุษย์ได้ด้วยการสนทนา อย่างไรก็ดี เครื่องจักรกลที่ฉลาด (Intelligent Machine) ที่สามารถรู้จำและเข้าใจความหมายที่แท้จริงและยังสามารถสนทนากับมนุษย์ได้ทุกคนและทุกสถานการณ์สำหรับทุกสภาพแวดล้อมนั้นยังนับว่าห่างไกลความจริงยิ่งนัก

อัตราการรู้จำของระบบรู้จำเสียงพูดขึ้นอยู่กับ จำนวนของผู้พูด ซึ่งระบบที่ขึ้นกับผู้พูด (Speaker-dependent systems) มีประสิทธิภาพในการรู้จำดีกว่าระบบที่ไม่ขึ้นกับผู้พูด (Speaker-independent systems) ธรรมชาติของการออกเสียงของผู้พูด ความยุ่งยากและขนาดของคำศัพท์ ความยุ่งยากของภาษาและเงื่อนไขของสิ่งแวดล้อม [7] ในส่วนเงื่อนไขของสิ่งแวดล้อมนั้น ถ้าสิ่งแวดล้อมของชุดเสียงพูดที่ใช้ในการฝึกฝน (Training set) และชุดเสียงพูดที่ใช้ในการทดสอบ (Testing set) ไม่เหมือนกัน จะทำให้คุณลักษณะสำคัญ (Feature) ของชุดข้อมูลทั้งสองต่างกัน ยังผลให้อัตราการรู้จำของระบบเสียงพูดลดลง โดยปัญหาของสิ่งแวดล้อมที่ไม่เข้าคู่ (Mismatch environment) ของชุดฝึกฝนและชุดทดสอบอาจแบ่งเป็น ปัญหาความแตกต่างของไมโครโฟนของทั้งสองชุด และปัญหาของสัญญาณรบกวน (Noise)

จากงานวิจัยด้านการรู้จำเสียงพูดภาษาไทยที่ผ่านมานั้น ได้มีการพัฒนาตั้งแต่ระบบรู้จำแบบคำโดด (Isolated word) คำต่อเนื่อง (Connected word) ไม่ว่าจะเป็นการรู้จำตัวเลขไทย [1,2] การรู้จำเสียงสระภาษาไทย [3,4] การรู้จำคำไทยหลายพยางค์ [5] ล้วนแล้วแต่เป็นระบบที่มีสิ่งแวดล้อมในการฝึกฝนและการทดสอบเดียวกันด้วยสัญญาณเสียงพูดที่ไม่มีสัญญาณรบกวน (Clean speech) ซึ่งเมื่อนำไปใช้ในทางปฏิบัติอัตราการรู้จำจะลดลงอย่างมากเมื่อสิ่งแวดล้อมเปลี่ยนไป เพราะเราไม่สามารถควบคุมสิ่งแวดล้อมได้อย่างดีเยี่ยมเพื่อให้การทดสอบและการฝึกฝนมีสิ่งแวดล้อมเดียวกันได้ อัตราการรู้จำของระบบที่ลดลงไม่มากนักภาย



รูปที่ 1.1 อัตราการรู้จำเสียงตัวเลขภาษาไทยเมื่อมีสัญญาณรบกวนโดยใช้ค่าลักษณะสำคัญ คือสัมประสิทธิ์การประมาณพันธะเชิงเส้น

ใต้เงื่อนไขของสิ่งแวดล้อมที่เปลี่ยนไปในกรณีที่มีสัญญาณรบกวนเข้ามา ซึ่งเรียกว่า ระบบนั้นคงทน ต่อสัญญาณรบกวน (Noise-robust) ย่อมมีความสำคัญยิ่งสำหรับระบบรู้จำเสียงพูดเมื่อถูกนำไปใช้งานจริง ดังนั้น วิทยานิพนธ์นี้สนใจปัญหาของสิ่งแวดล้อมที่ไม่เข้ากันในปัญหาของสัญญาณรบกวนในสำนักงาน (Office noise)

รูปที่ 1.1 แสดงอัตราการรู้จำเสียงตัวเลขภาษาไทยเมื่อมีสัญญาณรบกวน ซึ่งจะพบว่าอัตราการรู้จำลดลงอย่างมากเมื่อมีสัญญาณรบกวน ซึ่งไม่ถึงร้อยละ 20 ที่ค่าอัตราส่วนสัญญาณเสียงพูดต่อสัญญาณรบกวน (Signal to Noise Ratio, SNR) 0 dB ซึ่งนับว่าเป็นปัญหาใหญ่มากเมื่อนำไปใช้งานจริง

ระบบรู้จำเสียงพูดแบบคงทนต่อสัญญาณรบกวน (Noise-robust speech recognition) อาจได้จากสามแนวทางใหญ่ๆ ได้ คือ

1 แนวทางการกรองสัญญาณเสียงพูดที่มีสัญญาณรบกวนก่อนเข้าสู่ระบบการตัดสินใจ

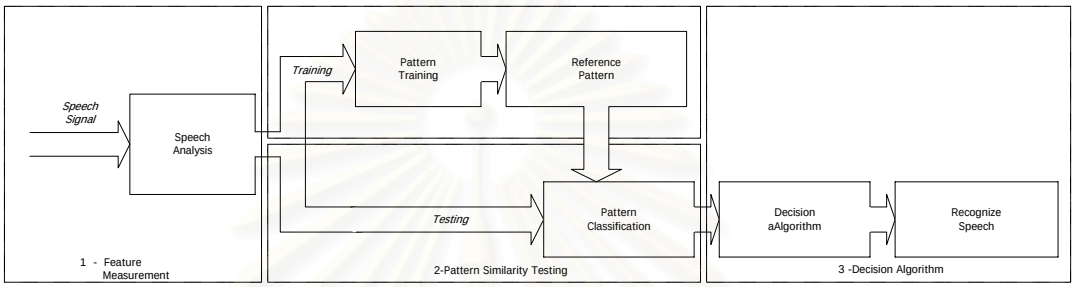
2 แนวทางการปรับแบบจำลองในการรู้จำตามผลกระทบของสัญญาณรบกวน

3 แนวทางการหาค่าลักษณะสำคัญของเสียงพูดที่คงทนต่อสัญญาณรบกวน

ในแนวทางแรกนั้นคือกระบวนการออกแบบวงจรกรองที่สามารถลดสัญญาณรบกวนได้ตามคุณลักษณะของสัญญาณรบกวนที่มีในสัญญาณเสียงพูด เช่นเครื่องกรอง Kalman เป็นต้น ซึ่งได้ผลดีเมื่อนำมาใช้กับสัญญาณเสียงพูด แต่ข้อด้อยคือใช้เวลานานในการคำนวณ เพราะมีค่าต่างๆในการคำนวณเป็นจำนวนมาก ทำให้การใช้งานจริงไม่สะดวก

แนวทางที่สอง เป็นแนวทางที่ทำการวิจัยอย่างกว้างขวาง ซึ่งเป็นวิธีการปรับเปลี่ยนค่าพารามิเตอร์ของระบบการตัดสินใจเมื่อข้อมูลมีสัญญาณรบกวน เช่นวิธี Parallel Model Combination (PMC) ซึ่งหาแบบจำลองของข้อมูลสองชุดคือ สัญญาณเสียงพูดที่ไม่มีสัญญาณรบกวนและสัญญาณเสียงพูดที่มีสัญญาณรบกวนมาแยกกันฝึกฝนแล้วจึงรวมค่าพารามิเตอร์เพื่อฝึกฝนอีกครั้ง เพื่อเป็นการรวมแบบมาใช้ในกระบวนการตัดสินใจ ข้อด้อยของวิธีนี้คือใช้ได้ดีในระบบที่มีคำศัพท์ไม่มากนัก [8]

แนวทางสุดท้าย เป็นแนวทางที่ใช้ในวิทยานิพนธ์นี้ คือแนวทางในการหาค่าลักษณะสำคัญของเสียงพูดที่คงทนต่อสัญญาณรบกวน ซึ่งเป็นวิธีการปรับเปลี่ยนค่าลักษณะสำคัญให้สามารถคงทนหรือลดผลกระทบของสัญญาณรบกวนก่อนที่จะนำไปใช้ในการฝึกฝนต่อไป ซึ่งจะยังผลให้ค่าผิดพลาดการรู้จำ (Recognition error) ลดลง แนวทางนี้เป็นกระบวนการที่ปรับปรุงค่าลักษณะสำคัญ ซึ่งเป็นค่าที่แทนสัญญาณเสียงพูดที่ใช้ในการฝึกฝน จะเห็นว่าเป็นการแก้ไขกับตัวสัญญาณโดยตรง จึงทำได้ง่ายและใช้เวลาในการคำนวณน้อย



รูปที่ 1.2 รายละเอียดขั้นตอนการรู้จำคำพูดภาษาไทย

ในรูปที่ 1.2 แสดงรายละเอียดขั้นตอนการรู้จำคำพูดภาษาไทย ซึ่งมีรายละเอียดดังนี้

ก. *การวัดค่าลักษณะสำคัญ (Feature measurement)* ประกอบด้วย การประมวลผลสัญญาณเบื้องต้น (Signal Preprocessing) และการหาค่าคุณลักษณะสำคัญของเสียงพูด เพื่อใช้แทนสัญญาณเสียงพูด ซึ่งวิทยานิพนธ์จะเกี่ยวข้องกับส่วนนี้ คือหาค่าลักษณะสำคัญที่คงทนต่อสัญญาณรบกวน

ข. *การแยกความคล้ายของรูปแบบ (Pattern similarity testing)* เป็นการจำแนกรูปแบบของค่าคุณลักษณะสำคัญ โดยมีรูปแบบอ้างอิง (Reference pattern) เป็นรูปแบบที่ใช้เปรียบเทียบ ในวิทยานิพนธ์นี้ใช้แบบจำลองฮิดเดนมาร์คอฟแบบต่อเนื่อง (Continous Hidden Markov Model, C-HMM)

ค. *การตัดสินใจ (Decision algorithm)* เป็นการตัดสินเลือกรูปแบบที่มีความคล้ายคลึงกันมากที่สุดระหว่างคำที่ไม่ทราบรูปแบบกับรูปแบบที่ได้เก็บเอาไว้แล้ว

1.2 ปัญหาของค่าลักษณะสำคัญสำหรับระบบรู้จำเสียงพูดเมื่อมีสัญญาณรบกวน และวิธีแก้ปัญหาที่มีผู้เสนอขึ้น

1. การแก้ไขสัมประสิทธิ์การประมาณพันธะเชิงเส้น โดยการหาอนุพันธ์อันดับสูงของสัมประสิทธิ์การประมาณพันธะเชิงเส้นเทียบกับอัตราส่วนสัญญาณรบกวนต่อสัญญาณข้อมูล (Noise to Signal Ratio, NSR) ซึ่งอนุพันธ์เหล่านี้มีประสิทธิภาพในการใช้ลดสัญญาณรบกวน และใช้การกระจายอนุกรมเทเลอร์เพื่อประมาณสัมประสิทธิ์การประมาณพันธะเชิงเส้นที่ใช้เป็นค่าลักษณะสำคัญแทน แต่ข้อจำกัดของขั้นตอนวิธีนี้คือสัญญาณรบกวนต้องมีระดับไม่สูงนัก ซึ่งโดยทั่วไปที่ใช้มีค่าประมาณ 6 dB [12]

2. การปรับปรุงสัมประสิทธิ์การประมาณพันธะเชิงเส้น โดยการลบค่ากำลังของสัญญาณรบกวนออกจากฟังก์ชันอัตสหสัมพันธ์ของเสียงพูดที่มีสัญญาณรบกวน ซึ่งลดค่า Cepstral distance ระหว่างค่า Cepstral ของสัญญาณเสียงพูดที่มีสัญญาณรบกวนกับสัญญาณเสียงพูดที่ไม่มีสัญญาณรบกวน และของสัญญาณพูดที่ปรับปรุงค่าสัมประสิทธิ์การประมาณพันธะเชิงเส้นแล้วกับสัญญาณเสียงพูดที่ไม่มีสัญญาณรบกวน แต่ผู้นำเสนอยังไม่นำไปใช้ในระบบรู้จำเสียงพูด [13]

3. การทำให้เป็นบรรทัดฐานเดียวกันของสัมประสิทธิ์เซปตรอล (Cepstral normalization) ในงานวิจัยที่ตรวจพบนั้นคือ การทำให้เป็นบรรทัดฐานเดียวกันของค่าเฉลี่ยของสัมประสิทธิ์เซปตรอล (Cepstral mean normalization, CMN) การทำให้เป็นบรรทัดฐานเดียวกันของค่าเฉลี่ยและความแปรปรวนของสัมประสิทธิ์เซปตรอล (Cepstral mean-variance normalization, CVN) และ การทำให้เป็นบรรทัดฐานเดียวกันของค่าโมเมนต์อันดับสามของสัมประสิทธิ์เซปตรอล (Cepstral third-order moment normalization, CTN) ซึ่งผู้เสนอได้แสดงไว้ว่าวิธีการนี้ให้ค่าอัตราการรู้จำสูงขึ้นมากสำหรับสัญญาณรบกวนในรถยนต์ แต่ยังต่ำสำหรับสัญญาณรบกวนเกาส์เซียนขาว [14]

4. การดัดแปลงค่าลักษณะสำคัญโดยใช้สัมประสิทธิ์เซปตรอล แทนค่าลักษณะสำคัญของเสียงพูดที่ไม่มีสัญญาณรบกวนรวมกับนิพจน์อนุพันธ์ของมันเมื่อเทียบกับสัญญาณรบกวนซึ่งได้จากการแปลงจากสัมประสิทธิ์ AR มาเป็นค่าลักษณะสำคัญในระบบรู้จำเสียงพูดที่มีสัญญาณรบกวน ซึ่งได้ค่าผิดพลาดการรู้จำโดยเฉลี่ยประมาณร้อยละห้าสิบ สำหรับระบบรู้จำเสียงพูดโดยใช้แบบจำลองฮิดเดนมาร์คอฟ (Hidden Markov Model, HMM) [15]

5. การแปลงค่าอัตสหสัมพันธ์ของสัญญาณเสียงพูดที่มีสัญญาณรบกวนไปสู่ค่าที่ไม่มีสัญญาณรบกวนโดยใช้ความเป็นรายคาบ (Periodicity) ของสัญญาณเสียงพูดมาใช้ ซึ่งผู้นำเสนอได้แสดงให้เห็นว่าวิธีการนี้สามารถปรับปรุงค่ายอดสเปกตรัล (Spectral peaks) ของสัญญาณเสียงพูดที่มีสัญญาณรบกวนให้ดีขึ้นได้อีกทั้งปรับปรุงค่า LPC Cepstral distance ด้วย แต่วิธีการนี้ไม่รับรองถึงเสถียรภาพของเครื่องกรองที่มีแต่โพล (All-pole filter) [16]

1.3 วัตถุประสงค์

1. เพื่อศึกษาระบบรู้จำเสียงพูดในสิ่งแวดล้อมที่มีสัญญาณรบกวน
2. เพื่อพัฒนากรรมวิธีในการลดสัญญาณรบกวนในระบบเสียงพูดภาษาไทย
3. พัฒนาระบบรู้จำเสียงพูดภาษาไทยให้มีอัตราการรู้จำมากขึ้น เมื่อสิ่งแวดล้อมในการฝึกฝนและการทดสอบต่างกันด้วยสัญญาณรบกวน

1.4 เป้าหมายและขอบเขตของวิทยานิพนธ์

สามารถพัฒนาค่าลักษณะสำคัญให้มีความคงทนต่อสัญญาณรบกวน ซึ่งทำให้ระบบรู้จำเสียงพูดภาษาไทยให้มีอัตราการรู้จำเพิ่มขึ้นจากเดิมไม่น้อยกว่าร้อยละ 60 ที่ SNR 0 dB เมื่อสัญญาณเสียงพูดชุดทดสอบเป็นสัญญาณเสียงพูดที่มีสัญญาณรบกวน โดยสัญญาณรบกวนกำหนดให้เป็นสัญญาณรบกวนเกาส์เซียนขาวค่าเฉลี่ยศูนย์แบบบวก (Additive zero-mean white Gaussian noise) และเป็นอิสระทางสถิติ (Statistical independent) กับสัญญาณเสียงพูด ซึ่งเดิมอัตราการรู้จำเสียงพูดเมื่อมีสัญญาณรบกวนที่ไม่ได้ปรับค่าคุณลักษณะสำคัญมีค่าประมาณร้อยละ 15 ที่ SNR 0 dB

1.5 ขั้นตอนและวิธีการดำเนินการวิจัย

1. ศึกษาค่าคุณลักษณะสำคัญที่คงทนต่อสัญญาณรบกวน โดยตั้งอยู่บนพื้นฐานของการประมาณเชิงเส้น พร้อมทั้งเปรียบเทียบประสิทธิภาพของขั้นตอนวิธีการที่มีผู้เสนอขึ้น
2. ศึกษาระบบรู้จำเสียงพูดภาษาไทย และระบบรู้จำเสียงพูดแบบคงทนต่อสัญญาณรบกวน
3. ออกแบบและปรับปรุงแก้ไขค่าคุณลักษณะสำคัญที่คงทนต่อสัญญาณรบกวนที่ได้เสนอขึ้น
4. จัดเก็บตัวอย่างและเตรียมข้อมูลเสียงพูด
5. เขียนโปรแกรมทดสอบผล
6. ประเมินและสรุปผลการทดลอง
7. เขียนวิทยานิพนธ์

1.6 ประโยชน์ที่คาดว่าจะได้รับ

ทราบถึงค่าลักษณะสำคัญของเสียงพูด และปัญหาของสัญญาณรบกวนต่อค่าลักษณะสำคัญ และออกแบบแก้ไขปรับปรุงค่าลักษณะสำคัญให้คงทนต่อสัญญาณรบกวนมากขึ้นโดยใช้การเทคนิคการลด

สัญญาณรบกวนแบบปรับตัวได้โดยตั้งอยู่บนพื้นฐานของการประมาณพันธะเชิงเส้น ซึ่งทำให้มีอัตราการรู้จำของระบบรู้จำเสียงพูดภาษาไทยเพิ่มขึ้น

1.7 ภาพรวมของวิทยานิพนธ์

วิทยานิพนธ์เล่มนี้ประกอบด้วยเนื้อหาส่วนต่างๆโดยแยกตามบทดังต่อไปนี้

บทที่ 1 บทนำ มีเนื้อหาที่กล่าวถึงความสำคัญของระบบรู้จำเสียงพูด ความเป็นมาของแนวเหตุผลสำหรับงานวิจัย ปัญหาของสัญญาณรบกวนและวิธีการแก้ปัญหาที่มีผู้เสนอขึ้น วัตถุประสงค์ เป้าหมายและขอบเขตของวิทยานิพนธ์ ขั้นตอนและวิธีการดำเนินการวิจัย และประโยชน์ที่คาดว่าจะได้รับ

บทที่ 2 หลักการ ทฤษฎี ขั้นตอนวิธีการ และงานวิจัยที่เกี่ยวข้อง ในบทนี้ จะกล่าวถึง หลักการทฤษฎีและขั้นตอนวิธีการที่เกี่ยวข้องกับงานวิจัยนี้ โดยแบ่งออกเป็นหัวข้อดังนี้คือ การประมวลผลสัญญาณเบื้องต้น การสกัดลักษณะสำคัญ การวิเคราะห์ค่าลักษณะสำคัญเมื่อมีสัญญาณรบกวนเกาส์เซียนขาวค่าเฉลี่ยศูนย์แบบบวก (Additive Zero-mean White Guassian Noise) และการหาค่าลักษณะสำคัญที่คงทนต่อสัญญาณรบกวนดังกล่าวด้วย และในส่วนของขั้นตอนวิธีการทางคณิตศาสตร์เชิงตัวเลขที่สำคัญสำหรับงานวิจัยคือ ขั้นตอนวิธีการแบ่งส่วน และขั้นตอนวิธีการหาค่าเจาะจงต่ำสุดตามลำดับ และส่วนสุดท้ายจะกล่าวถึงเอกสารงานวิจัยที่เกี่ยวข้องกับงานวิจัยนี้

บทที่ 3 ขั้นตอนวิธีในการดำเนินการวิจัย รายละเอียดในบทนี้จะกล่าวถึงรายละเอียดของขั้นตอนวิธีการทั้งหมดในการดำเนินงานวิจัย ซึ่งประกอบไปด้วยรายละเอียดเกี่ยวกับการเตรียมตัวอย่างข้อมูลเสียงพูดทั้งหมด รายละเอียดเกี่ยวกับการหาค่าลักษณะสำคัญ รายละเอียดขั้นตอนการปรับลดสัญญาณรบกวนแบบปรับตัวได้ และสุดท้านคือวิธีการฝึกฝนและการทดสอบอัตราการรู้จำของค่าลักษณะสำคัญในงานวิจัย

บทที่ 4 ผลการวิจัยและการวิเคราะห์ผล ในบทนี้ เบื้องต้นจะกล่าวถึงขั้นตอนวิธีการรู้จำเสียงพูดภาษาไทย จากนั้นจะแสดงผลการทดลองเพื่อหาระบบที่เหมาะสมสำหรับการรู้จำ ตามด้วยผลการทดลองเพื่อวิเคราะห์ผลของการทำเน้นล่วงหน้า และสุดท้ายซึ่งสำคัญที่สุดจะกล่าวถึงรายละเอียดของผลการรู้จำเสียงพูดภาษาไทย ซึ่งแสดงผลการวิจัยที่ได้ในแต่ละกรณี และวิเคราะห์ผลที่ได้จากการทดสอบ โดยทำการทดสอบกับเสียงพูดตัวเลขภาษาไทย และเสียงวรรณยุกต์ภาษาไทยทั้งห้าเสียงคือเสียงสามัญ เอก โท ตรี และจัตวา เพื่อทดสอบความสามารถในการรู้จำเสียงพูดภาษาไทยซึ่งมีวรรณยุกต์เป็นส่วนสำคัญของภาษาไทย จากนั้นจะแสดงรายละเอียดการวิเคราะห์ประกอบผลการวิจัยทั้งหมด

ในบทสุดท้าย *บทที่ 5 บทสรุปและข้อเสนอแนะ* เสนอผลสรุปทั้งหมดของงานวิจัย และข้อเสนอแนะสำหรับการวิจัยที่ต่อเนื่องกับงานวิจัยนี้ในอนาคตต่อไป

ในภาคผนวก จะแสดงผลการรู้จำทั้งหมดในงานวิจัยนี้ โดยแบ่งเป็น 5 ส่วน ภาคผนวก ก แสดงผลการรู้จำเสียงตัวเลขภาษาไทย และภาคผนวก ข ถึงภาคผนวก ฉ แสดงผลการรู้จำเสียงวรรณยุกต์ไทยเสียงสามัญ เอก โท ตรี และจัตวาตามลำดับ

บทที่ 2

หลักการ ทฤษฎี ขั้นตอนวิธีการและงานวิจัยที่เกี่ยวข้อง

ในบทนี้ จะกล่าวถึง หลักการ ทฤษฎีและขั้นตอนวิธีการที่เกี่ยวข้องกับงานวิจัยนี้ โดยแบ่งออกเป็นหัวข้อดังนี้คือ การประมวลผลสัญญาณเบื้องต้น การสกัดลักษณะสำคัญ การวิเคราะห์ค่าลักษณะสำคัญเมื่อมีสัญญาณรบกวนเกาส์เซียนขาวค่าเฉลี่ยศูนย์แบบบวก (Additive Zero-mean White Guassian Noise) และการหาค่าลักษณะสำคัญที่คงทนต่อสัญญาณรบกวนดังกล่าวด้วย และในส่วนของขั้นตอนวิธีการทางคณิตศาสตร์เชิงตัวเลขที่สำคัญสำหรับงานวิจัยคือ ขั้นตอนวิธีการแบ่งส่วน และขั้นตอนวิธีการหาค่าเจาะจงต่ำสุดตามลำดับ และส่วนสุดท้ายจะกล่าวถึงเอกสารงานวิจัยที่เกี่ยวข้องกับงานวิจัยนี้

2.1 การประมวลผลสัญญาณเบื้องต้น (Signal Preprocessing)

การประมวลผลสัญญาณเบื้องต้นเป็นขั้นตอนในการจัดเตรียมสัญญาณเสียงพูด (Speech Signal) โดยเริ่มตั้งแต่การแปลงสัญญาณเสียงพูดที่ได้จากการบันทึกเสียงมาเป็นสัญญาณเชิงเลขซึ่งได้เป็นข้อมูลเสียงพูด (Speech Data) เพื่อที่จะนำไปใช้ในการประมวลผลในขั้นตอนต่อไป

เนื่องจากสัญญาณเสียงพูดนั้นมีค่าทางสถิติที่เปลี่ยนแปลงตามเวลา (Non-stationary) ทำให้ไม่สามารถจำลองสัญญาณเสียงพูดเป็นค่าทางสถิติได้ ด้วยเหตุนี้ ในการประยุกต์ใช้งานเสียงพูดกับกรรมวิธีสัญญาณดิจิตอล (Digital Signal Processing) จึงจำเป็นยิ่งที่ต้องแบ่งสัญญาณเสียงพูดออกเป็นส่วนย่อยสั้นๆ เรียกว่า *กรอบเสียงพูด* (Speech Frame) โดยในแต่ละส่วนย่อยนั้นจะมีความยาวประมาณ 10-40 มิลลิวินาที ซึ่งทำให้กรอบเสียงพูดแต่ละกรอบมีค่าทางสถิติเปลี่ยนแปลงตามเวลาน้อยมาก จนถือได้ว่า *ในแต่ละกรอบเสียงพูดมีค่าทางสถิติไม่เปลี่ยนแปลงตามเวลา* (Stationary) จึงสามารถทำการประมวลผลโดยใช้ค่าทางสถิติกับสัญญาณเสียงพูดในแต่ละกรอบได้

ขั้นตอนกรรมวิธีการวางกรอบขนาดสัญญาณ (Smoothing Window)

ขั้นตอนกรรมวิธีการวางกรอบขนาดสัญญาณนี้เป็นการแบ่งสัญญาณเสียงพูดออกเป็นส่วนย่อยๆเพื่อใช้ในการหาค่าสหสัมพันธ์ (Autocorrelation) โดยการคูณแต่ละค่าของสัญญาณในกรอบข้อมูลเสียงพูดด้วยค่าฟังก์ชันกรอบ (Window Function) ซึ่งมีหลายชนิด เช่น Rectangular Window, Hamming Window, Hanning Window, Blackman Window, Kaiser Window เป็นต้น [9] โดยในงานวิจัยนี้ใช้ฟังก์ชันกรอบชนิด Hamming Window ดังแสดงในรูปที่ 2.1 ซึ่งการวางกรอบขนาดสัญญาณนั้น ยังผลให้เกิดผล 2 ประการ ซึ่งประการแรกคือเป็นการลดทอนแอมปลิจูดอย่างช้าๆที่บริเวณปลายแต่ละข้างของกรอบข้อมูลเสียงพูดเพื่อป้องกันการเปลี่ยนแปลงที่ไม่ต่อเนื่องอย่างกระทันหันที่ส่วนปลายของกรอบ

สัญญาณ และอีกประการหนึ่ง เป็นการสร้างค่าการประสานสำหรับผลการแปลงฟูริเยร์ (Fourier Transformation) ของฟังก์ชันกรอบและแถบสเปกตรัมของเสียงพูด เสียงพูดที่ผ่านขั้นตอนนี้จะได้เป็นข้อมูลเสียงพูดเพื่อนำไปใช้ในกรรมวิธีสัญญาณดิจิตอลต่อไป ซึ่งกรรมวิธีการวางกรอบสัญญาณเป็นไปตามสมการที่ (2.1) และ (2.2)

$$x_l(n) = x_l(n) \cdot w(n) \tag{2.1}$$

$$w(n) = 0.54 - 0.46\cos\left(2\pi n / N-1\right) \tag{2.2}$$

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

โดย $x_l(n)$ คือค่าสัญญาณเสียงพูดของข้อมูลชุดที่ n

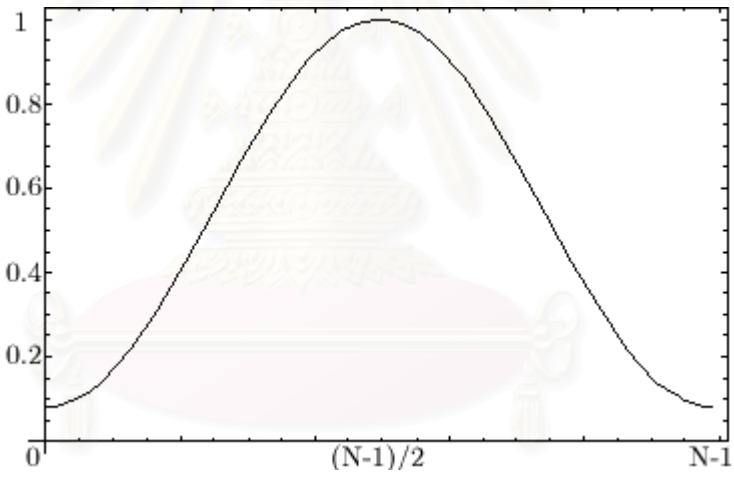
$w(n)$ คือฟังก์ชันกรอบชนิด Hamming Window

N คือจำนวนข้อมูลในแต่ละกรอบสัญญาณเสียงพูด

n คือลำดับข้อมูลในกรอบสัญญาณเสียงพูดที่ l

L คือจำนวนของกรอบสัญญาณเสียงพูด

l คือลำดับของกรอบสัญญาณเสียงพูด



รูปที่ 2.1 ฟังก์ชันกรอบชนิด Hamming Window

2.2 การสกัดค่าลักษณะสำคัญ (Feature Extraction)

การสกัดค่าลักษณะสำคัญเป็นการวิเคราะห์สัญญาณเสียงพูดเพื่อหาค่าที่จะนำมาแทนข้อมูลเสียงพูดซึ่งเก็บรวบรวมลักษณะสำคัญของเสียงพูดแต่ละเสียง โดยจะนำค่าเหล่านี้ไปทำการฝึกฝนระบบให้รับรู้ถึงความแตกต่างของเสียงพูดแต่ละเสียงและเพื่อใช้สำหรับการเปรียบเทียบเพื่อแบ่งแยกความแตกต่างของเสียงพูดแต่ละเสียงออกจากกัน

2.2.1 แบบจำลองการประมาณพันธะเชิงเส้น (Linear Prediction Model)

การประมาณพันธะเชิงเส้น (Linear Prediction) เป็นเทคนิคที่นิยมนำมาใช้ในการวิเคราะห์คุณสมบัติทางกายภาพของสัญญาณ เช่น Spectral Magnitude ของสัญญาณ เทคนิคนี้พิจารณาได้รวดเร็ว แม่นยำ และมีประสิทธิภาพ ซึ่งก็ใช้ได้ดีสำหรับสัญญาณเสียงพูดและนำมาประยุกต์ใช้กับระบบรู้จำเสียงพูดได้เป็นอย่างดี แบบจำลองการประมาณพันธะเชิงเส้นเป็นแบบจำลองแบบใช้ค่าพารามิเตอร์ (Parametric Model) ซึ่งจำลองให้ค่าพารามิเตอร์ที่ได้นี้เกิดจากช่องทางเดินเสียงเพื่อจำลองเป็นสัญญาณเสียงพูด

ให้ $s[n]$ แทนสัญญาณเสียงพูดซึ่งสามารถประมาณได้ด้วยผลรวมเชิงเส้นของสัญญาณในอดีต p ตัว คือ

$$s[n] \approx \hat{s}[n] = a_1 s[n-1] + a_2 s[n-2] + \ldots + a_p s[n-p] = \sum_{k=1}^{p} a_k s[n-k] \quad (2.3)$$

โดยที่ $a_1, a_2, \ldots, a_p$ เป็นสัมประสิทธิ์ที่มีค่าคงที่ตลอดกรอบข้อมูลเสียงพูดที่วิเคราะห์

เปลี่ยนสมการที่ (2.3) โดยให้มีพจน์การกระตุ้น (Excitation Term) $Gu[n]$ จะได้

$$s[n] = \sum_{i=1}^{p} a_i s[n-i] + Gu[n] \quad (2.4)$$

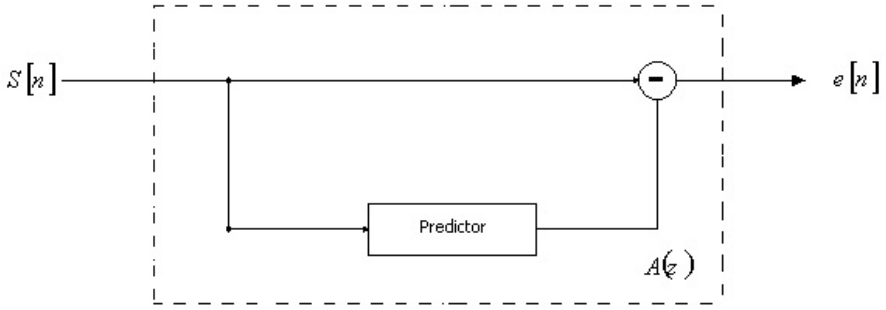
เมื่อ $u[n]$ เป็นการกระตุ้น และ G เป็นอัตราขยายของการกระตุ้น แบบจำลองตามสมการที่ (2.4) เรียกว่า Autoregressive Model (AR Model) ซึ่งเป็นสมการเชิงเส้น (Linear Equation) ถ้าวิเคราะห์สมการที่ (2.4) ในโดเมน z จะได้ความสัมพันธ์คือ

$$S(z) = \sum_{i=1}^{p} a_i z^{-i} S(z) + GU(z) \quad (2.5)$$

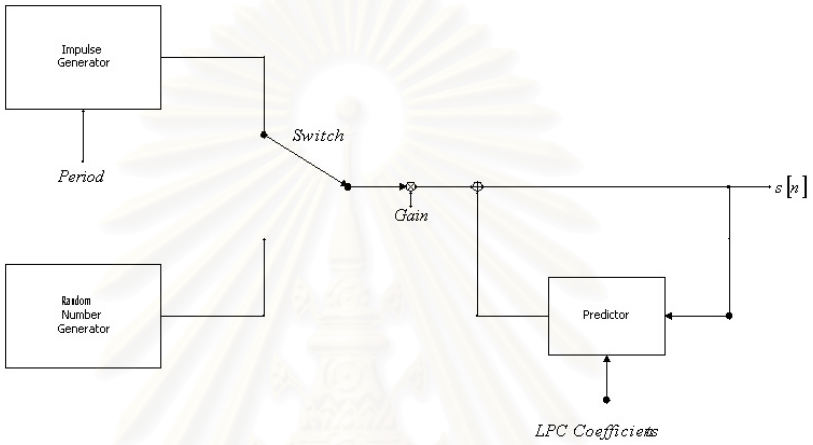
และฟังก์ชันการถ่ายโอน

$$H(z) = \frac{S(z)}{GU(z)} = \frac{1}{1 - \sum_{i=1}^{p} a_i z^{-i}} = \frac{1}{A(z)} \quad (2.6)$$

รูปที่ 2.2 แสดงความหมายของสมการที่ (2.6) โดยที่ $u[n]$ ถูกคูณด้วย G เป็นสัญญาณขาเข้าของระบบที่มีแต่โพล (All-pole System) $H(z) = \frac{1}{A(z)}$ เกิดเป็นสัญญาณเสียงพูด $s[n]$ ซึ่งฟังก์ชันการกระตุ้นของเสียงพูดเป็นขบวนคลื่นกึ่งรายคาบ (Quasi-periodic) สำหรับเสียงโฆษะ (Voiced Speech Sound) และฟังก์ชันการกระตุ้นของเสียงพูดเป็นแหล่งสัญญาณรบกวนแบบสุ่มสำหรับเสียงอโฆษะ (Unvoiced Sound) แบบจำลองที่เหมาะสมสำหรับการสังเคราะห์เสียงพูดที่มีความสอดคล้องกับการประมาณพันธะเชิงเส้นแสดงในรูปที่ 2.3 ซึ่งอธิบายได้ว่า สวิตช์สำหรับเลือกเสียงโฆษะหรืออโฆษะนั้นจะเลือกแหล่งกำเนิดการกระตุ้นโดยมีคุณลักษณะของเสียงพูดว่าเป็นเสียงโฆษะหรืออโฆษะเป็นตัวควบคุม โดยจะเลือกเป็นขบวนคลื่นกึ่งรายคาบเป็นตัวกระตุ้นสำหรับเสียงโฆษะ หรือจะเลือกเป็นแหล่งสัญญาณรบกวนแบบสุ่มสำหรับเสียงอโฆษะ



รูปที่ 2.2 แบบจำลองการประมาณพันธะเชิงเส้นสำหรับเสียงพูด [10]



รูปที่ 2.3 แบบจำลองสังเคราะห์เสียงพูดจากแบบจำลองการประมาณพันธะเชิงเส้น [11]

การวิเคราะห์การประมาณพันธะเชิงเส้น

จากรูปที่ 2.2 ความสัมพันธ์ระหว่าง $s[n]$ และ $u[n]$ เป็นดังนี้

$$s[n] = \sum_{i=1}^{p} a_i s[n-i] + Gu[n] \tag{2.7}$$

พิจารณาผลรวมเชิงเส้นของค่าตัวอย่างสัญญาณเสียงพูดในอดีตเพื่อที่จะประมาณค่า $\tilde{s}[n]$ โดย

$$\tilde{s}[n] = \sum_{k=1}^{p} a_k s[n-k] \tag{2.8}$$

พิจารณาค่าผิดพลาดจากการประมาณ $e[n]$

$$e[n] = s[n] - \tilde{s}[n] = s[n] - \sum_{k=1}^{p} a_k s[n-k] \tag{2.9}$$

ฟังก์ชันการถ่ายโอนของค่าผิดพลาดจากการประมาณ

$$A(z) = \frac{E(z)}{S(z)} = 1 - \sum_{k=1}^{p} a_k z^{-k} \tag{2.10}$$

เนื่องจากสัญญาณเสียงพูด $s[n]$ เกิดจากระบบเชิงเส้นดังแสดงในรูปที่ 2.2 ดังนั้นจากสมการที่ (2.7) (2.8) และ (2.9) จะได้

$$e[n] = s[n] - \tilde{s}[n] = s[n] - \sum_{k=1}^{p} a_k s[n-k] = Gu[n] \qquad (2.11)$$

ปัญหาพื้นฐานของการวิเคราะห์การประมาณพันธะเชิงเส้นคือการคำนวณหาค่าชุดของสัมประสิทธิ์การทำนาย $\{a_k\}_{K=1}^{p}$ โดยตรงจากสัญญาณเสียงพูดที่ทำให้ค่าคุณสมบัติของสเปกตรัมของวงจรเชิงเลขที่แสดงในรูปที่ 2.3 สอดคล้องกับรูปคลื่นสัญญาณเสียงภายในกรอบข้อมูลเสียงพูดที่วิเคราะห์ เนื่องจากว่าคุณลักษณะทางสเปกตรัมของสัญญาณเสียงพูดแปรเปลี่ยนตามเวลา ดังนั้น ค่าสัมประสิทธิ์การทำนายที่เวลา n ต้องเกิดจากการประมาณส่วนย่อยของสัญญาณเสียงพูดในช่วงเวลาสั้นๆ รอบเวลา n

ดังนั้น ปัญหาดังกล่าวข้างต้นจึงเป็นการหาค่าชุดสัมประสิทธิ์การทำนายที่ทำให้ค่าผิดพลาดกำลังสองเฉลี่ย (Mean Square Error) ของตัวทำนายบนรูปคลื่นสัญญาณเสียงภายในกรอบข้อมูลเสียงพูดมีค่าน้อยที่สุด

กำหนดสัญญาณเสียงพูดช่วงสั้นๆ (Short-term Speech) และค่าความผิดพลาดในส่วนย่อย (Error Segment) ที่เวลา n ตามลำดับเป็น

$$s_n[m] = s[n+m] \qquad (2.12)$$

$$e_n[m] = e[n+m] \qquad (2.13)$$

ต้องการให้ค่าผิดพลาดกำลังสองเฉลี่ยของค่าสัญญาณที่เวลา n มีค่าต่ำสุด

$$E_n = \sum_m e_n^2[m] \qquad (2.14)$$

$$= \sum_m \left[s_n[m] - \sum_{k=1}^{p} a_k s_n[m-k] \right]^2 \qquad (2.15)$$

$$= \sum_m (s_n[m])^2 - \sum_m \left(2 s_n[m] \sum_{k=1}^{p} a_k s_n[m-k] \right) + \sum_m \left(\sum_{k=1}^{p} a_k s_n[m-k] \right)^2 \qquad (2.16)$$

$$= \sum_m (s_n[m])^2 - 2\sum_{k=1}^{p} a_k \sum_m s_n[m] s_n[m-k] + \sum_m \left(\sum_{k=1}^{p} a_k s_n[m-k] \right)^2 \qquad (2.17)$$

หาอนุพันธ์ของ E_n เทียบกับค่าสัมประสิทธิ์ a_k และให้เท่ากับศูนย์

$$\frac{\partial E_n}{\partial a_k} = 0, \qquad k = 1,2,..., p \qquad (2.18)$$

จากสมการที่ (2.17) จะได้สมการที่ (2.18) เป็น

$$\frac{\partial E_n}{\partial a_k} = 0 = -2\sum_m s_n[m] s_n[m-k] + 2\sum_m \left(\sum_{k=1}^{p} a_k s_n[m-k] \right) s_n[m-i] \qquad (2.19)$$

จะได้

$$\sum_m s_n[m-i]s_n[m] = \sum_{k=1}^{p} a_k \sum_m s_n[m-i]s_n[m-k] \qquad (2.20)$$

ด้านขวามือของสมการ พจน์ $\sum_m s_n[m-i]s_n[m-k]$ คือความแปรปรวนร่วมของ $s_n[m]$ นั่นคือ

$$\phi_n(i,k) = \sum_m s_n[m-i]s_n[m-k] \qquad (2.21)$$

สมการที่ (2.21) เขียนได้ใหม่คือ

$$f(i,0) = \sum_{k=1}^{p} a_k f_n(i,k) \qquad (2.22)$$

ซึ่งเป็นชุดของ p สมการ p ตัวแปร ดังนั้น

$$E_n = \sum_m s_n^2[m] - \sum_{k=1}^{p} a_k \sum_m s_n[m]s_n[m-k] \qquad (2.23)$$

$$= f_n(0,0) - \sum_{k=1}^{p} a_k f_n(0,k) \qquad (2.24)$$

ค่าสัมประสิทธิ์การทำนายที่เหมาะสมที่สุดจะต้องคำนวณ $\phi_n(i,k)$ สำหรับ $1 \le i \le p$ และ $1 \le k \le p$ และแก้สมการ p สมการโดยมีสองแนวทางคือ วิธีอัตสหสัมพันธ์ (Autocorrelation method) และวิธีความแปรปรวนร่วม (Covariance method) สำหรับวิทยานิพนธ์นี้ใช้วิธีอัตสหสัมพันธ์

สมมติให้ส่วนย่อยของสัญญาณเสียง $s_n[m]$ มีค่าเป็นศูนย์นอกช่วง $0 \le m \le N-1$ โดยที่ N เป็นความกว้างของฟังก์ชันกรอบ ซึ่งสมมูลกับการคูณสัญญาณเสียงพูด $s[m+n]$ ด้วยฟังก์ชันกรอบขนาดจำกัด $w[m]$ ซึ่งมีค่าศูนย์นอกช่วง $0 \le m \le N-1$ ดังนั้น สัญญาณเสียงที่นำมาวิเคราะห์จะเขียนในรูปแบบสมการ คือ

$$s_n[m] = \begin{cases} s[m+n]w[n], & 0 \le m \le N-1 \\ 0, & otherwise \end{cases} \qquad (2.25)$$

จะได้

$$E_n = \sum_{m=0}^{N-1+p} e_n^2[m] \qquad (2.26)$$

และ

$$\phi_n(i,k) = \sum_{m=0}^{N-1+p} s_n[m-i]s_n[m-k] \qquad 1 \le i \le p, 0 \le k \le p \quad (2.27)$$ หรือ

$$\phi_n(i,k) = \sum_{m=0}^{N-1-(i-k)} s_n[m]s_n[m+i-k] \qquad 1 \le i \le p, 0 \le k \le p \quad (2.28)$$

สมการที่ (2.28) เป็นฟังก์ชันของ $i-k$ ดังนั้น ฟังก์ชันความแปรปรวนร่วม $\phi_n(i,k)$ จะลดรูปเป็นฟังก์ชันอัตสหสัมพันธ์อย่างง่าย

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

$$\phi_n(i,k) = R_n(i-k) = \sum_{m=0}^{N-1-(i-k)} s_n[m]s_n[m+i-k] \qquad (2.29)$$

เนื่องจากฟังก์ชันอัตสหสัมพันธ์มีความสมมาตรคือ $R_n(-k) = R_n(k)$ สมการประมาณพันธะเชิงเส้นเขียนได้เป็น

$$\sum_{k=1}^{p} R_n(|i-k|)a_k \approx R_n(i), \quad 1 \le i \le p \qquad (2.30)$$

และเขียนเป็นรูปแบบเมตริกซ์คือ

$$\begin{bmatrix} R_n(0) & R_n(1) & R_n(2) & \cdots & R_n(p-1) \\ R_n(1) & R_n(0) & R_n(1) & \cdots & R_n(p-2) \\ R_n(2) & R_n(1) & R_n(0) & \cdots & R_n(p-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ R_n(p-1) & R_n(p-2) & R_n(p-3) & \cdots & R_n(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} R_n(1) \\ R_n(2) \\ R_n(3) \\ \vdots \\ R_n(p) \end{bmatrix} \qquad (2.31)$$

หรือ

$$\mathbf{Ra} = \mathbf{r} \qquad (2.32)$$

จากสมการที่ (2.32) จะเห็นว่าเมตริกซ์ของค่าอัตสหสัมพันธ์ **R** อยู่ในรูป Toeplitz ซึ่งมีลักษณะสมมาตรและทุกๆสมาชิกในแนวทแยงมุมมีค่าเท่ากัน ทำให้การแก้สมการเพื่อหาค่าสัมประสิทธิ์ a_k โดยวิธีอัตสหสัมพันธ์ง่ายกว่าวิธีความแปรปรวนร่วม วิธีที่เหมาะสมในการแก้สมการที่อยู่ในรูปเมตริกซ์ Toeplitz คือขั้นตอนวิธีการ Levinson-Derbin ดังนี้

<u>ขั้นตอนที่ 1</u> เงื่อนไขเริ่มต้น

$$\begin{aligned} E_0 &= R(0) \\ a_0 &= 0 \end{aligned} \qquad (2.33)$$

<u>ขั้นตอนที่ 2</u> สัมประสิทธิ์การสะท้อน

$$k_m = \frac{R(m) - \sum_{i=1}^{m-1} a_{m-1}(i)R(m-i)}{E_{m-1}}, \quad |k_m| \le 1 \qquad (2.34)$$

<u>ขั้นตอนที่ 3</u> สัมประสิทธิ์ของการประมาณพันธะเชิงเส้นในแต่ละรอบของการคำนวณ

$$\begin{aligned} a_m(m) &= k_m \\ a_m(i) &= a_{m-1}(i) - k_m a_{m-1}(m-i), \quad (1 \le i \le m) \\ E_m &= (1-k_m^2)E_{m-1} \end{aligned} \qquad (2.35)$$

<u>ขั้นตอนที่ 4</u> สมการอัตสหสัมพันธ์

$$R(i) = \sum_{n=i}^{N-1} s[n]s[n-i], \quad i = 1,2,\ldots, p \qquad (2.36)$$

<u>ขั้นตอนที่ 5</u> สัมประสิทธิ์ของการประมาณพันธะเชิงเส้นเมื่อสิ้นสุดการคำนวณ

$$a_m = a_m(p), \quad 1 \le m \le p \qquad (2.37)$$

2.2.2 สัมประสิทธิ์เซปตรอล (Cepstral Coefficient)

สัมประสิทธิ์เซปตรอล สามารถคำนวณได้ 2 วิธี คือ วิธีการคำนวณจากการแปลงฟูริเยร์แบบเร็ว (Fast Fourier Transform, FFT) และวิธีการคำนวณจากแบบจำลองการประมาณพันธะเชิงเส้น ซึ่งมีข้อได้เปรียบคือใช้เวลาในการคำนวณน้อยกว่าโดยคำนวณจากสมการที่ (2.38) [10]

$$c_m = a_m + \sum_{k=1}^{n-1}\left(\frac{k}{m}\right)c_k a_{m-k}, \quad m = 1,2,..., p$$
$$c_m = \sum_{k=1}^{m-1}\left(\frac{k}{m}\right)c_k a_{m-k}, \qquad m > p \tag{2.38}$$

โดย a_m คือค่าสัมประสิทธิ์การประมาณพันธะเชิงเส้น

c_m คือค่าสัมประสิทธิ์เซปตรอล

p คืออันดับของค่าสัมประสิทธิ์การประมาณพันธะเชิงเส้น

2.3 การวิเคราะห์ค่าลักษณะสำคัญเมื่อมีสัญญาณรบกวนและการหาค่าลักษณะสำคัญที่คงทนต่อสัญญาณรบกวน

2.3.1 สัญญาณรบกวนขาว (White Noise)

สัญญาณรบกวนขาว เป็นสัญญาณรบกวนที่เป็นอิสระทางค่าสถิติ นั่นคือไม่ขึ้นกับสัญญาณอื่นใดเลย แม้แต่สัญญาณรบกวนขาวตัวอื่น ซึ่งสัญญาณรบกวนชนิดนี้มีค่ากำลังเท่ากันตลอดทุกความถี่ ดังนั้นค่าอัตสหสัมพันธ์ของมันจึงมีค่าที่ $R(0)$ เท่านั้น นั่นคือ

$$R_{nn}(k) = \begin{cases} \eta, & k = 0 \\ 0, & k \neq 0 \end{cases} \tag{2.39}$$

โดยที่ η เป็นค่าสเปกตรัมกำลัง (Power Spectral) ของสัญญาณรบกวน

สัญญาณรบกวนขาวที่สำคัญและใช้มากในทางวิศวกรรมไฟฟ้าสื่อสารคือ สัญญาณรบกวนเกาส์เซียนขาว (White Gaussian Noise, WGN) ซึ่งเกิดจากกรรมวิธีเชิงสุ่มจากค่าลำดับของตัวแปรเชิงสุ่มจริงแบบเกาส์เซียนที่ไม่มีความสัมพันธ์กัน ซึ่งเรียกว่า กรรมวิธีสัญญาณรบกวนแบบขาว (White Gaussian Noise Process)

2.3.2 การวิเคราะห์ผลของสัญญาณรบกวนต่อค่าลักษณะสำคัญ

สมมติมีสัญญาณรบกวนเข้ามากวนสัญญาณเสียงพูด เริ่มพิจารณาสมการที่ (2.3) จะเป็น

$$x[n] = s[n] + n[n] \tag{2.40}$$

โดยที่ $x[n]$ แทนสัญญาณเสียงพูดที่มีสัญญาณรบกวน (Noisy Speech) $s[n]$ แทนสัญญาณเสียงพูดที่ไม่มีสัญญาณรบกวน (Clean Speech) และ $n[n]$ แทนสัญญาณรบกวนเกาส์เซียนขาวค่าเฉลี่ยศูนย์ (Zero-mean White Gaussian Noise) และมีค่าความแปรปรวนหรือค่ากำลังเป็น η และเป็นอิสระทางสถิติ (Statistical Independent) กับ $s[n]$ ค่าอัตสหสัมพันธ์ของ $x[n]$ คือ

$$\phi_{n,xx}(i,k) = R_{n,xx}(i-k) = \sum_{m=0}^{N-1-(i-k)} x_n[m]x_n[m+i-k] \tag{2.41}$$

ซึ่ง $\phi_{n,xx}(i,k)$ สามารถลดรูปได้เป็น

$$\phi_{n,xx}(i,k) = \phi_{n,ss}(i,k) + \phi_{n,sn}(i,k) + \phi_{n,ns}(i,k) + \phi_{n,nn}(i,k) \tag{2.42}$$

จากคุณสมบัติของสัญญาณรบกวนที่ตั้งสมมติฐานไว้ทำให้

$$\phi_{n,sn}(i,k) = \phi_{n,ns}(i,k) = 0 \tag{2.43}$$

และ

$$\phi_{n,nn}(i,k) = \begin{cases} \eta, & if\ i = k \\ 0, & otherwise \end{cases} \tag{2.44}$$

ดังนั้น

$$\phi_{n,xx}(i,k) = \begin{cases} \phi_{n,ss}(i,i) + \eta, & if\ i = k \\ \phi_{n,ss}(i,k), & otherwise \end{cases} \tag{2.45}$$

หรือ

$$R_{xx}(i) = \begin{cases} R_{ss}(i) + \eta, & if\ i = 0 \\ R_{ss}(i), & otherwise \end{cases} \tag{2.46}$$

เพราะฉะนั้น เมตริกซ์อัตสหสัมพันธ์ของสัญญาณเสียงพูดที่มีสัญญาณรบกวนคือ

$$\mathbf{R}_{xx}(\eta) = \begin{bmatrix} R_{ss}(0)+\eta & R_{ss}(1) & R_{ss}(2) & . & . & R_{ss}(p-1) \\ R_{ss}(1) & R_{ss}(0)+\eta & R_{ss}(1) & . & . & R_{ss}(p-2) \\ R_{ss}(2) & R_{ss}(1) & R_{ss}(0)+\eta & . & . & R_{ss}(p-3) \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ R_{ss}(p-1) & R_{ss}(p-2) & R_{ss}(p-3) & . & . & R_{ss}(0)+\eta \end{bmatrix} \tag{2.47}$$

เวกเตอร์สัมประสิทธิ์การประมาณพันธะเชิงเส้นของสัญญาณเสียงพูดที่มีสัญญาณรบกวนคือ

$$\mathbf{a}(\eta) = \mathbf{R}_{xx}(\eta)^{-1}\mathbf{r} = (\mathbf{R}_{ss} + \eta\mathbf{I})^{-1}\mathbf{r} \tag{2.48}$$

หรือ

$$\mathbf{a}(\eta) = \begin{bmatrix} a_1(\eta) & a_2(\eta) & \ldots & a_p(\eta) \end{bmatrix}^T \tag{2.49}$$

เมื่อ $\mathbf{I}$ เป็นเมตริกซ์เอกลักษณ์ (Identity Matrix)

2.3.3 ค่าลักษณะสำคัญที่คงทนต่อสัญญาณรบกวนบนพื้นฐานของการประมาณพันธะเชิงเส้น

พิจารณาค่าผิดพลาดจากการประมาณพันธะเชิงเส้น จากสมการที่ (2.9)

$$e[n] = s[n] - \tilde{s}[n] = s[n] - \sum_{k=1}^{p} a_k s[n-k] \tag{2.50}$$

หรือเขียนใหม่ได้เป็น

$$e[n] = -\sum_{i=0}^{p} a_i x[n-i] \tag{2.51}$$

เมื่อ $a_0 = -1$ ดังนั้น ค่าผิดพลาดของการประมาณ (Prediction error) หรือค่าผิดพลาดเศษเหลือ (Residual error) ของสัญญาณที่เวลา n จึงเป็น

$$\begin{aligned} E^{(p)} &= \sum_{n=0}^{N-1+p} (e[n])^2 = \sum_{n=0}^{N-1+p} \left(\left[-\sum_{i=0}^{p} a_i x[n-i] \right] \left[-\sum_{j=0}^{p} a_j x[n-j] \right] \right) \\ &= \sum_{i=0}^{p} a_i \sum_{j=0}^{p} a_j \sum_{n=0}^{N-1+p} x[n-i]x[n-j] \end{aligned} \tag{2.52}$$

แต่

$$\sum_{n=0}^{N-1+p} x[n-i]x[n-j] = \sum_{n=0}^{N-1+p} x[n]x[n-j+i] = R(|i-j|) \tag{2.53}$$

ดังนั้น

$$E^{p} = \sum_{i=0}^{p} a_i \sum_{j=0}^{p} a_j R(|i-j|) = \mathbf{a}_p^{\mathrm{T}} \mathbf{R}_{xx} \mathbf{a}_p \tag{2.54}$$

เมื่อ $\mathbf{a} = [a_0 \quad a_1 \quad a_2 \quad \mathrm{K} \quad a_p]^{\mathrm{T}}$ โดยที่ $a_0 = -1$ และ $\mathbf{R}_{xx}$ คือ เมตริกซ์อัตสหสัมพันธ์ขนาด $(p+1)\times(p+1)$ นั่นคือ

$$\mathbf{R}_{xx}(\eta) = \begin{bmatrix} R_{xx}(0) & R_{xx}(1) & R_{xx}(2) & . & . & R_{xx}(p) \\ R_{xx}(1) & R_{xx}(0) & R_{xx}(1) & . & . & R_{xx}(p-1) \\ R_{xx}(2) & R_{xx}(1) & R_{xx}(0) & . & . & R_{xx}(p-2) \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ R_{xx}(p) & R_{xx}(p-1) & R_{xx}(p-2) & . & . & R_{xx}(0) \end{bmatrix} \tag{2.55}$$

จากสมการที่ (2.47) จะเห็นว่าเมตริกซ์อัตสหสัมพันธ์ $\mathbf{R}_{xx}$ มีค่าเป็น $\mathbf{R}_{ss} + \eta \mathbf{I}$ หรือ $\mathbf{R}_{xx} + \mathbf{R}_{nn}$ ซึ่ง η เป็นค่าแปกตรัมกำลังของสัญญาณรบกวนที่สามารถหาได้จากค่าอัตสหสัมพันธ์ที่ศูนย์ โดย $\mathbf{R}_{nn}$ แทนเมตริกซ์อัตสหสัมพันธ์ของสัญญาณรบกวนคือ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

$$\mathbf{R}_{nn}(\eta) = \begin{bmatrix} \eta & 0 & 0 & . & . & 0 \\ 0 & \eta & 0 & . & . & 0 \\ 0 & 0 & \eta & . & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & 0 & . & . & \eta \end{bmatrix} = diag(\eta) \qquad (2.56)$$

จากสมการที่ (2.54) จะได้

$$\begin{aligned} E^p &= \mathbf{a}_p{}^{\mathsf{T}} \mathbf{R}_{xx} \mathbf{a}_p \\ &= \mathbf{a}_p{}^{\mathsf{T}} (\mathbf{R}_{ss} + \mathbf{R}_{nn}) \mathbf{a}_p \\ &= \mathbf{a}_p{}^{\mathsf{T}} \mathbf{R}_{ss} \mathbf{a}_p + \mathbf{a}_p{}^{\mathsf{T}} \mathbf{R}_{nn} \mathbf{a}_p \end{aligned} \qquad (2.57)$$

ถ้าค่า SNR มีค่าค่อนข้างต่ำมากถึงปานกลาง สามารถพิจารณาพจน์แรกทางขวามือของสมการที่ (2.57) ในบรรทัดสุดท้ายว่ามีค่าน้อยมากเมื่อเทียบกับพจน์ที่สอง นั่นคือ $\mathbf{a}^{\mathsf{T}}\mathbf{R}_{ss}\mathbf{a} << \mathbf{a}^{\mathsf{T}}\mathbf{R}_{nn}\mathbf{a}$ ดังนั้น

$$\mathbf{a}^{\mathsf{T}}\mathbf{R}_{xx}\mathbf{a} \cong \mathbf{a}^{\mathsf{T}}\mathbf{R}_{nn}\mathbf{a} = \eta \sum_{k=0}^{p} a_k^2 = R_{nn}(0) \sum_{k=0}^{p} a_k^2 \qquad (2.58)$$

จากสมการที่ (2.58) กำหนดสมการการปรับลดสัญญาณรบกวนเป็น

$$\hat{\mathbf{a}}^{\mathsf{T}} (\mathbf{R}_{xx} - \beta \mathbf{I}) \hat{\mathbf{a}} \cong \hat{\eta} \sum_{k=0}^{p} \hat{a}_k^2 \qquad (2.59)$$

เมื่อ β แทนสัญญาณรบกวนส่วนที่หักออกจาก $R_{nn}(0)$ และ $\hat{\eta}$ แทนส่วนที่เหลืออยู่ของความแปรปรวนหรือกำลังของสัญญาณรบกวน และ $\hat{\mathbf{a}}$ แทนเวกเตอร์สัมประสิทธิ์การประมาณพันธะเชิงเส้นที่ได้จากลำดับอัตสหสัมพันธ์ที่ผ่านการปรับลดสัญญาณรบกวนแล้ว คือ $\{R_{xx}(0) - \beta, R_{xx}(1), \text{K}, R_{xx}(p)\}$

ในกระบวนการลดสัญญาณรบกวนที่นำเสนอนั้น เพื่อลดค่า $\hat{\eta}$ ลงในระดับที่ยอมรับได้ ซึ่งเป็นค่าที่ยังคงไว้ซึ่งคุณสมบัติของสัญญาณที่ยังไม่มีสัญญาณรบกวนเข้ามา ซึ่งในที่นี้เลือกค่า 15 dB [17] ระหว่างค่ากำลังของสัญญาณระหว่างสัญญาณที่ผ่านการปรับลดแล้วกับสัญญาณรบกวนที่เหลือ นั่นคือ

$$\hat{\eta} = \frac{R_{xx}(0) - \beta}{10^{1.5}} \qquad (2.60)$$

การหาค่า β ที่เหมาะสมคือนำสมการที่ (2.60) นำไปแทนในสมการที่ (2.59) แล้วทำให้

$$f(x) = \hat{\mathbf{a}}^{T} (\mathbf{R}_{xx} - x\mathbf{I}) \hat{\mathbf{a}} - \frac{R_{xx}(0) - x}{10^{1.5}} \left(\sum_{k=0}^{p} \hat{a}_k^2 \right) = 0 \qquad (2.61)$$

โดย x เป็นค่า β ที่ทำให้สมการที่ (2.61) เป็นจริง หรือกล่าวอีกนัยหนึ่งคือเป็นค่าสัญญาณรบกวนที่หักออกที่ดีที่สุด ซึ่งคำนวณด้วยวิธีการเชิงตัวเลข (Numerical Method) โดยในวิทยานิพนธ์นี้ใช้ขั้นตอนวิธีการ Bisection ในการหาคำตอบ (รายละเอียดของขั้นตอนวิธีการอยู่ในหัวข้อถัดไป) โดยกำหนดให้ค่าเริ่มต้นเป็น 0 และ $\lambda_{\min}$ ซึ่งเป็นค่าเจาะจงต่ำสุด (Minimum eigenvalue) ซึ่งคำนวณหาด้วยวิธีกำลังสมมาตร (Symmetric Power Method) ของ $\mathbf{R}_{xx}$ ซึ่งเป็นค่าที่มากที่สุดที่ยังทำให้เครื่องกรองที่มีแต่โพล (All-pole filter) ยังคงรักษาเสถียรภาพอยู่ได้

จากเทคนิคที่กล่าวข้างต้น เมื่อพิจารณาค่าเริ่มต้นในการวนซ้ำเพื่อคำนวณหาค่าการปรับลดสัญญาณรบกวนที่เหมาะสมนั้นจะได้ดังนี้

พิจารณา $f(x)|_{x=0}$ พบว่า

$$f(0) = \mathbf{a}^T \mathbf{R}_{xx} \mathbf{a} - \frac{R_{xx}(0)}{10^{1.5}} \left(\sum_{k=0}^{p} a_k^2 \right) > 0 \qquad (2.61)$$

และ $f(x)|_{x=\lambda_{\min}}$ พบว่า

$$f(\lambda_{\min}) = -\frac{R_{xx}(0) - x}{10^{1.5}} \left(\sum_{k=0}^{p} a_k^2 \right) < 0 \qquad (2.62)$$

ดังนั้นจึงมี $0 < x < \lambda_{\min}$ ที่ทำให้ $f(x) = 0$ จากคุณสมบัติดังกล่าว ทำให้เมตริกซ์อัตสหสัมพันธ์ $\mathbf{R}_{xx} - x\mathbf{R}_{nn}$ เป็น Positive-definite ดังนั้นรับรองเสถียรภาพของระบบอย่างแน่นอน

เมื่อได้ค่า β ที่เหมาะสมแล้ว แสดงว่าการลดสัญญาณรบกวนตามขั้นตอนวิธีการนี้สัมฤทธิ์ผล ค่าเวกเตอร์สัมประสิทธิ์การประมาณพันธะเชิงเส้นในรอบสุดท้ายในการคำนวณ จะกำหนดให้เป็นค่าลักษณะสำคัญสำหรับเสียงพูดเพื่อใช้ในการรู้จำต่อไป

จากข้างต้น แสดงให้เห็นถึงการปรับลดสัญญาณรบกวนแบบปรับตัวได้ (Adaptive Noise Reduction, ANR) โดยค่าพารามิเตอร์ในการปรับลดคือค่าเจาะจงต่ำสุดซึ่งเปลี่ยนไปในแต่ละกรอบสัญญาณเสียงพูดที่เข้ามา ดังนั้นเทคนิคนี้จึงมีการปรับลดสัญญาณรบกวนโดยขึ้นอยู่กับค่าสัญญาณรบกวนที่เข้ามาซึ่งได้ส่งผ่านคุณลักษณะของตัวสัญญาณรบกวนเองมาทางค่าเจาะจงนั่นเอง

2.4 วิธีการเชิงตัวเลขที่เกี่ยวข้อง (Numerical Method)

2.4.1 วิธีการแบ่งส่วน (Bisection Method) [18]

วิธีการแบ่งส่วน เป็นวิธีการหาราก หรือคำตอบของสมการที่อยู่ในรูปของ $f(x) = 0$ ซึ่ง

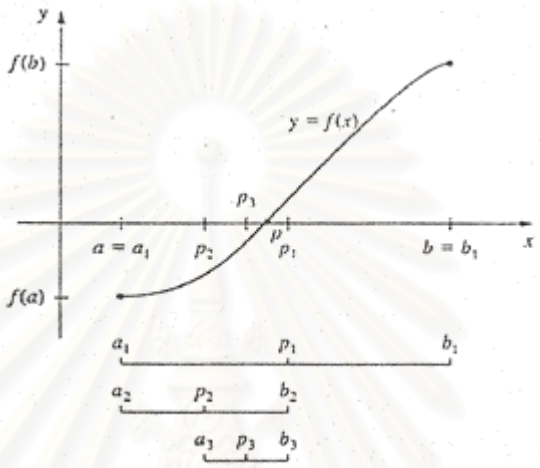
เป็นที่รู้จักกันดีและใช้กันอย่างกว้างขวางในปัญหาของการประมาณ (Approximation problems) ซึ่งเทคนิคนี้ตั้งอยู่บนพื้นฐานของทฤษฎีค่าระหว่างกลาง (Intermediate Value Theorem)

สมมุติว่า f เป็นฟังก์ชันต่อเนื่องบนช่วง $[a,b]$ โดยที่ $f(a)$ และ $f(b)$ มีเครื่องหมายตรงข้ามกัน จากทฤษฎีค่าระหว่างกลาง จะได้ว่าต้องมีค่า p ระหว่างช่วง (a,b) ที่ทำให้ $f(p) = 0$ นั่นคือจะต้องมีคำตอบของสมการ $f(x) = 0$ ในช่วงดังกล่าวอย่างแน่นอน โดยทำการแบ่งช่วงย่อยออกทีละครึ่งของช่วง $[a,b]$ และในแต่ละครั้งนั้นจะยังมีค่าคำตอบ p อยู่ในช่วงที่เหลืออยู่เสมอ

เริ่มต้น กำหนดให้ $a_1 = a$ และ $b_1 = b$ และ p_1 เป็นค่ากลางของช่วง $[a,b]$ นั่นคือ

$$p_1 = \frac{1}{2}(a_1 + b_1) \qquad (2.63)$$

ถ้า $f(p_1)=0$ ดังนั้น $p=p_1$ แต่ถ้าไม่ใช่ $f(p_1)$ จะต้องมีเครื่องหมายเดียวกับค่า $f(a)$ หรือ $f(b)$ อันใดอันหนึ่ง ดังนั้น ถ้า $f(p_1)$ และ $f(a)$ มีเครื่องหมายเดียวกัน ย่อมแสดงว่า $p\in(p_1,b_1)$ ดังนั้นจึงกำหนดให้ $a_2=a_1$ และ $b_2=b_1$ แทน แต่ถ้า $f(p_1)$ และ $f(a)$ มีเครื่องหมายตรงกันข้าม ย่อมแสดงว่า $p\in(a_1,p_1)$ จากนั้นก็ทำแบบเดียวกันกับช่วงใหม่ $[a_2,b_2]$ ซึ่งอธิบายขั้นตอนวิธีการได้ดังรูปที่ 2.4



รูปที่ 2.4 แสดงขั้นตอนวิธีการแบ่งส่วน [18]

<u>ขั้นตอนวิธีการแบ่งส่วน (Bisection Algorithm)</u>

ในการหาคำตอบหรือรากของสมการ $f(x)=0$ ต้องกำหนดฟังก์ชันต่อเนื่อง f บนช่วง $[a,b]$ เมื่อ $f(a)$ และ $f(b)$ มีเครื่องหมายตรงกันข้าม

<u>ค่าพารามิเตอร์อินพุท</u> คือ

ค่าสิ้นสุดช่วง a และ b

ค่ายอมทนหรือค่าผิดพลาดที่ยอมรับได้ (Tolerance) TOL

ค่าจำนวนการวนรอบที่มากที่สุด (Maximum number of iterations) N_0

<u>ค่าเอาท์พุท</u> คือ

แสดงค่าประมาณของคำตอบ p หรือข้อความฟ้องกระบวนการไม่สำเร็จอันใดอันหนึ่ง

<u>ขั้นตอนที่ 1</u> กำหนด $i=1$

<u>ขั้นตอนที่ 2</u> ขณะที่ $i\le N_0$ ทำขั้นตอนที่ 3 ถึง 6

<u>ขั้นตอนที่ 3</u> กำหนดให้ $p = a + \frac{b-a}{2}$ (คำนวณค่า p_i)

<u>ขั้นตอนที่ 4</u> ถ้า $f(p) = 0$ หรือ $\frac{b-a}{2} \le TOL$ ดังนั้น

<u>ค่าเอาท์พุท คือ p</u> (กระบวนการเสร็จสิ้นสมบูรณ์)

นั่นคือ <u>สิ้นสุด หรือหยุดกระบวนการ</u>

<u>ขั้นตอนที่ 5</u> กำหนด $i = i + 1$

<u>ขั้นตอนที่ 6</u> ถ้า $f(a)f(b) > 0$ ดังนั้น กำหนดให้ $a = p$ (คำนวณหา a_i และ b_i)

หรือถ้า $f(a)f(b) < 0$ จะกำหนดให้ $b = p$

<u>ขั้นตอนที่ 7</u> ถ้า $i > N_0$ นั่นคือกระบวนการหาคำตอบไม่สำเร็จ

นั่นคือ <u>สิ้นสุด หรือหยุดกระบวนการ</u>

2.4.2 ค่าเจาะจง (Eigenvalue) [19,20]

ค่าเจาะจง (Eigenvalue) และเวกเตอร์เจาะจง (Eigenvector) เป็นค่าที่แสดงถึงคุณสมบัติประจำตัวของเมตริกซ์ เช่นค่าเจาะจงสามารถบอกถึงคุณสมบัติความเป็นกำหนดบวก (Positive definite) รวมถึงสามารถคำนวณว่าเมตริกซ์นั้นมีเมตริกซ์ผกผัน (Inverse matrix) หรือไม่

กำหนดให้ $\mathbf{A}$ เป็นเมตริกซ์จัตุรัสขนาด $n \times n$ มี λ เป็นค่าคงที่ และ $\mathbf{v}$ เป็นเวกเตอร์ที่สอดคล้องกับสมการ

$$\mathbf{Av} = \lambda\mathbf{v} \tag{2.64}$$

สมการที่ (2.64) สามารถเขียนให้อยู่ในรูปของชุดสมการเอกพันธุ์เชิงเส้นคือ

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = \mathbf{0} \tag{2.65}$$

คำตอบของสมการที่ (2.65) จะไม่เป็นเวกเตอร์ศูนย์ก็ต่อเมื่อ เมตริกซ์ $\mathbf{A} - \lambda\mathbf{I}$ ต้องเป็นเมตริกซ์ (Singular matrix) ซึ่งหาค่าเมตริกซ์ผกผันได้ กล่าวคือ ตัวกำหนด (Determinant) ของเมตริกซ์ $\mathbf{A} - \lambda\mathbf{I}$ ต้องมีค่าเป็นศูนย์ นั่นคือ

$$\det(\mathbf{A} - \lambda\mathbf{I}) = 0 \tag{2.66}$$

เมื่อ $\mathbf{I}$ เป็นเมตริกซ์เอกลักษณ์ขนาด $n \times n$ สมการที่ (2.66) นำไปสู่สมการพหุนาม (Polynomial Equation) อันดับที่ n ดังสมการที่ (2.67)

$$\det(\mathbf{A} - \lambda\mathbf{I}) = \lambda^n + a_{n-1}\lambda^{n-1} + \text{L} \ + a_1\lambda + a_0 = 0 \tag{2.67}$$

ซึ่งจะเรียก $\det(\mathbf{A} - \lambda\mathbf{I})$ ว่า พหุนามคุณลักษณะ (Characteristic Polynomial) ของเมตริกซ์ $\mathbf{A}$ และเรียกสมการที่ (2.67) ว่า สมการคุณลักษณะ (Characteristic Equation) ดังนั้น สืบเนื่องจากทฤษฎีมูลฐานของพีชคณิต สมการที่ (2.67) จะต้องมีรากหรือคำตอบของสมการอยู่ n ค่า (อาจจะซ้ำได้) ซึ่งค่าเหล่านี้เอง เรียกว่า *ค่าเจาะจง* ของเมตริกซ์ $\mathbf{A}$ และแต่ละค่าเจาะจงนั้น เมตริกซ์ $\mathbf{A} - \lambda_i\mathbf{I}$ จะเป็นเมตริกซ์เอกฐาน (Singular Matrix) และจะมีเวกเตอร์ $\mathbf{v}_i$ ที่ไม่เวกเตอร์ศูนย์ที่ทำให้สมการ $\mathbf{Av}_i = \lambda_i\mathbf{v}_i$ เป็นจริง ซึ่ง

ค่า $\mathbf{v}_i$ เรียกว่า *เวกเตอร์เจาะจง* ของเมตริกซ์ $\mathbf{A}$ ดังนั้นจากทฤษฎีดังกล่าวจะได้ว่า เมตริกซ์ $\mathbf{A}$ ชนาด $n \times n$ จึงมีค่าเจาะจง n ค่า และสอดคล้องกับเวกเตอร์เจาะจง n ค่าเช่นเดียวกัน

ค่าเจาะจงและเวกเตอร์เจาะจงของเมตริกซ์อัตสหสัมพันธ์

ให้ $\mathbf{R}$ แทนเมตริกซ์อัตสหสัมพันธ์ ขนาด $n \times n$ ดันนั้นจากหัวข้อข้างต้นจึงกล่าวได้ว่า เมตริกซ์ $\mathbf{R}$ จึงมีค่าเจาะจงและเวกเตอร์เจาะจงอยู่ n ค่า โดยเมตริกซ์อัตสหสัมพันธ์นั้นเป็น Toeplitz และเป็นกำหนดบวกที่ซึ่งประกอบด้วยค่าเจาะจง $\{\lambda_i\}_{i=1}^{p}$ และเวกเตอร์เจาะจง $\{\mathbf{v}_i\}_{i=1}^{p}$ มีคุณสมบัติพิเศษ คือ ค่าเจาะจง $\{\lambda_i\}_{i=1}^{p}$ จะเป็นจำนวนจริงบวกเสมอ [21]

2.4.3 วิธีประมาณค่าเจาะจงด้วยวิธีกำลัง (Power Method) [18,22]

วิธีประมาณค่าเจาะจงด้วยวิธีกำลังนั้นใช้สำหรับค่าค่าเจาะจงที่มีขนาดมากที่สุด แต่ในวิทยานิพนธ์นี้ต้องการประมาณค่าเจาะจงต่ำสุด ดังนั้นเราจึงนำวิธีนี้มาประยุกต์โดยการนำเมตริกซ์ที่ต้องการหาค่าเจาะจงต่ำสุดมาหาค่าเมตริกซ์ผกผัน จากนั้นหาค่าเจาะจงด้วยวิธีกำลังของเมตริกซ์ผกผันแทน ซึ่งค่าเจาะจงสูงสุดที่ได้จากเมตริกซ์ผกผันนี้เมื่อนำมาหาค่าส่วนกลับของมัน จะให้ค่าเจาะจงต่ำสุดของเมตริกซ์เดิม [21] แต่เนื่องจากเมตริกซ์ที่เราจะหาค่าเจาะจง คือเมตริกซ์อัตสหสัมพันธ์เป็นเมตริกซ์สมมาตร การหาค่าเจาะจงของเมตริกซ์ชนิดนี้ยิ่งง่ายและลู่เข้าค่าที่ต้องการเร็วมากขึ้นด้วยวิธีกำลังสมมาตร (Symmetric Power Method) ซึ่งมีขั้นตอนวิธีการดังนี้คือ

ค่าพารามิเตอร์อินพุท คือ มิติ n เมตริกซ์ $\mathbf{A}$ เวกเตอร์ $\mathbf{x}$ ที่ไม่ใช่เวกเตอร์ศูนย์ ค่ายอมทนหรือค่าผิดพลาดที่ยอมรับได้ (Tolerance) TOL และค่าจำนวนการวนรอบที่มากที่สุด N_0

ค่าพารามิเตอร์อินพุท คือ ค่าการประมาณค่าเจาะจง μ และค่าเวกเตอร์ $\mathbf{x}$ ที่มีขนาดเป็นหนึ่ง ($\|\mathbf{x}\|_2 = 1$) หรือข้อความฟ้องกระบวนการไม่สำเร็จอย่างใดอย่างหนึ่ง

ขั้นตอนที่ 1 กำหนด $k = 1$ และให้ $\mathbf{x} = \mathbf{x} / \|\mathbf{x}\|_2$

ขั้นตอนที่ 2 ขณะที่ $k \le N_0$ ทำขั้นตอนที่ 3 ถึง 8

ขั้นตอนที่ 3 กำหนด $\mathbf{y} = \mathbf{A}\mathbf{x}$

ขั้นตอนที่ 4 กำหนด $\mu = \mathbf{x}^{\mathsf{T}}\mathbf{y}$

ขั้นตอนที่ 5 ถ้า $\|\mathbf{y}\|_2 = 0$ ดังนั้น

ค่าเอาท์พุท คือ $\mathbf{x}$ เป็นค่าเจาะจงเป็นศูนย์และกำหนดเมตริกซ์ $\mathbf{x}$ ใหม่

นั่นคือ *สิ้นสุด หรือหยุดกระบวนการ*

ขั้นตอนที่ 6 กำหนด ค่าผิดพลาด $ERR = \left\| \mathbf{x} - \frac{\mathbf{y}}{\|\mathbf{y}\|_2} \right\|_2$ และ $\mathbf{x} = \frac{\mathbf{y}}{\|\mathbf{y}\|_2}$

ขั้นตอนที่ 7 ถ้า $ERR < TOL$ ดังนั้น

ค่าเอาท์พุทคือ μ และ $\mathbf{x}$ เป็นค่าเจาะจงและเวกเตอร์เจาะจง

นั่นคือ <u>*สิ้นสุด หรือหยุดกระบวนการ*</u>

<u>*ขั้นตอนที่ 8*</u> กำหนด $k = k + 1$

<u>*ขั้นตอนที่ 9*</u> ถ้า $k > N_0$ นั่นคือกระบวนการหาคำตอบไม่สำเร็จ

นั่นคือ <u>*สิ้นสุด หรือหยุดกระบวนการ*</u>

2.5 งานวิจัยที่เกี่ยวข้อง

ในระยะเริ่มแรกนั้น การรู้จำเสียงพูดภาษาไทยใช้กรรมวิธีการเปรียบเทียบทางเวลาแบบพลวัต (Dynamic Time Warping, DTW) ซึ่งมีงานวิจัยดังต่อไปนี้

งานวิจัยของระพีพัฒน์ เพ็ญศิริ ทำการวิจัยการรู้จำตัวเลขไทย (0-9) โดยไม่ขึ้นกับผู้พูด โดยใช้กรรมวิธีเปรียบเทียบทางเวลาแบบพลวัต ได้อัตราการรู้จำร้อยละ 79.25 [1]

งานวิจัยของธีระ ภัทราพรนันท์ ทำการวิจัยการรู้จำเสียงสระโดยมีการแปลงดิสครีตฟูริเยร์ และค่าลอการิทึมของพลังงานเป็นค่าลักษณะสำคัญ โดยใช้วิธีวัดสเปกตรัมดิสแตนซ์และกรรมวิธีเปรียบเทียบทางเวลาแบบพลวัต ได้อัตราการรู้จำร้อยละ 86.17 สำหรับเสียงสระจำนวน 24 เสียง และร้อยละ 81.00 สำหรับเสียงวรรณยุกต์ 15 เสียง [3]

ในระยะต่อมา ได้นำแบบจำลองฮิดเดนมาร์คอฟมาใช้ ซึ่งมีงานวิจัยดังต่อไปนี้

งานวิจัยของเสาวลักษณ์ อารีพงศา ทำการวิจัยการรู้จำเสียงพูดตัวเลขเป็นภาษาไทยโดยไม่ขึ้นกับผู้พูด โดยมีสัมประสิทธิ์การทำนายพันธะเชิงเส้นอันดับ 10 กับเวกเตอร์ชุดรหัสขนาด 64 ชุดรหัสคำเป็นค่าลักษณะสำคัญ ได้อัตราการรู้จำร้อยละ 82 [2]

งานวิจัยของวิศรุต อาขุบุตร ทำการวิจัยการรู้จำคำไทยหลายพยางค์ซึ่งประกอบด้วยคำพยางค์เดียว สองพยางค์ และสามพยางค์ จำนวน 70 คำ โดยมีสัมประสิทธิ์การประมาณพันธะเชิงเส้นอันดับ 10 และเวกเตอร์ชุดรหัสขนาด 128 และ 256 เป็นค่าลักษณะสำคัญ ได้อัตราการรู้จำร้อยละ 89.91 [5]

งานวิจัยของ Tungthangthum ทำการวิจัยการรู้จำเสียงสระภาษาไทย โดยใช้ค่าความถี่มูลฐานร่วมกับความถี่ฟอร์แมนท์เป็นค่าลักษณะสำคัญ ได้อัตราการรู้จำร้อยละ 89.91 [23]

งานวิจัยของอุมาวสี ทาทอง ทำการวิจัยระบบรู้จำคำเรียกพยัญชนะภาษาไทย โดยสัมประสิทธิ์เซปตรอลบนความถี่เชิงเส้น สัมประสิทธิ์การประมาณพันธะเชิงเส้น สัมประสิทธิ์เซปตรอลที่คำนวณจากสัมประสิทธิ์การประมาณพันธะเชิงเส้น สัมประสิทธิ์เซปตรอลที่คำนวณจากการแปลงฟูริเยร์ และสัมประสิทธิ์เซปตรอลบนความถี่เมล เป็นค่าลักษณะสำคัญ ได้อัตราการรู้จำร้อยละ 83.75 [6]

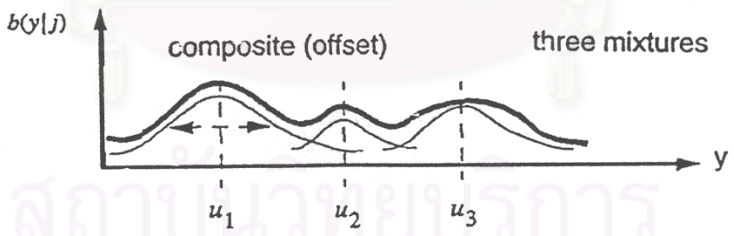
นอกจากนั้นยังมีการนำโครงข่ายประสาทเทียม (Artificial Neural Networks, ANNs) และเทคนิคฟัซซีมาใช้ดังต่อไปนี้

งานวิจัยของวุฒิพงษ์ พรสุขจันทรา ทำการวิจัยการรู้จำเสียงตัวเลขไทย โดยใช้สัมประสิทธิ์การประมาณพันธะเชิงเส้นอันดับ 10 เป็นค่าลักษณะสำคัญ ได้อัตราการรู้จำร้อยละ 89.40 [6]

งานวิจัยของชัย วุฒิวิวัฒน์ชัย ทำการวิจัยการรู้จำคำพยางค์เดียว สองพยางค์และสามพยางค์ซึ่งเป็นชุดคำเดียวกับงานของวิศรุต อาขุบุตร โดยใช้สัมประสิทธิ์การประมาณพันธะเชิงเส้นอันดับ 10 เป็นค่าลักษณะสำคัญ ได้อัตราการรู้จำร้อยละ 91.90 [6]

งานวิจัยของเอกฤทธิ์ มณีน้อย ทำการวิจัยการรู้จำหน่วยเสียงสระภาษไทยจำนวน 9 หน่วยเสียง โดยมีสัมประสิทธิ์การประมาณพันธะเชิงเส้น สัมประสิทธิ์เซปตรอล ความถี่ฟอร์แมนท์และความเข้มของสเปกตรัมเป็นค่าลักษณะสำคัญ ได้อัตราการรู้จำร้อยละ 90.34 [4]

ในงานวิจัยนี้เลือกใช้แบบจำลองฮิดเดนมาร์คอฟแบบต่อเนื่อง เนื่องจากถ้านำมาใช้กับแบบจำลองฮิดเดนมาร์คอฟแบบไม่ต่อเนื่องจะต้องทำการปรับขนาดข้อมูลด้วยการควอนไทซ์แบบเวกเตอร์ (Vector Quantization,VQ) ต่างจากแบบจำลองฮิดเดนมาร์คอฟแบบต่อเนื่องซึ่งใช้การประสมเชิงเส้นแบบให้น้ำหนัก Weighted Linear Combination) ของการแจกแจงแบบเกาส์ (Gaussians Distribution) หรือที่เรียกว่า Gaussians Mixture ดังแสดงในรูปที่ 2.5 ซึ่งในการเพิ่มจำนวน Gaussians Mixture เป็นการเพิ่มความซับซ้อนให้กับการแทนข้อมูลเสียง เพื่อให้ได้ข้อมูลเสียงที่ยังคงลักษณะสำคัญของข้อมูลเสียงพูดได้ครบถ้วนมากขึ้น



รูปที่ 2.5 แสดงการประสมเชิงเส้นแบบให้น้ำหนักของการแจกแจงแบบเกาส์

บทที่ 3

ขั้นตอนวิธีในการดำเนินการวิจัย

รายละเอียดในบทที่ 3 นี้ จะกล่าวถึงรายละเอียดของขั้นตอนวิธีการทั้งหมดในการดำเนินงานวิจัย ซึ่งประกอบไปด้วยรายละเอียดเกี่ยวกับการเตรียมตัวอย่างข้อมูลเสียงพูดทั้งหมด รายละเอียดเกี่ยวกับการหาค่าลักษณะสำคัญ รายละเอียดการปรับลดสัญญาณรบกวนแบบปรับตัวได้ และสุดท้ายคือวิธีการฝึกฝนและการทดสอบอัตราการรู้จำของค่าลักษณะสำคัญในงานวิจัย

3.1 รายละเอียดการเตรียมตัวอย่างข้อมูลเสียงพูด

3.1.1 รายละเอียดชุดคำศัพท์

ในการทดสอบคุณลักษณะสำคัญของสัญญาณเสียงพูดที่เสนอขึ้น จะใช้ฐานข้อมูลเสียงพูดของห้องปฏิบัติการวิจัยประมวลสัญญาณดิจิตอล ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ซึ่งได้รับการควบคุมสภาพแวดล้อมขณะทำการบันทึกเสียงให้คล้ายคลึงกับสภาพแวดล้อมของสถานที่ทำงานทั่วไปและมีเสียงรบกวนน้อยที่สุด โดยเสียงพูดที่บันทึกไว้จะจัดเก็บด้วยตัวอย่างขนาด 16 บิต และมีอัตราการชักตัวอย่าง 11 kHz เนื่องจากเสียงพูดของมนุษย์จะอยู่ในช่วงไม่เกิน 5 kHz จึงต้องใช้อัตราการชักตัวอย่างที่สูงกว่าสองเท่าของความถี่ 5 kHz สำหรับเสียงที่บันทึกจะเป็นเสียงของผู้ที่ใช้ภาษาไทยกรุงเทพเป็นภาษาพูดทั่วไป มีอายุระหว่าง 18 ถึง 25 ปี และเป็นผู้ที่มีการออกเสียงเป็นปกติและตรงตามหลักการออกเสียงพูดภาษาไทย

ชุดคำศัพท์ที่เลือกมาทดสอบนั้น จะจัดแบ่งเป็น 6 กลุ่ม คือกลุ่มเสียงตัวเลขภาษาไทยศูนย์ถึงเก้า จำนวน 10 คำ และกลุ่มตามกลุ่มเสียงวรรณยุกต์ทั้ง 5 คือ สามัญ เอก โท ตรี และจัตวา ในแต่ละกลุ่มเสียงจะมีชุดคำศัพท์พยางค์เดียว 8 คำ รวมชุดคำศัพท์ทั้งหมดเป็น 50 คำ โดยตารางที่ 3.1 แสดงชุดคำศัพท์ที่เลือกมาทดสอบแบ่งตามกลุ่มเสียงวรรณยุกต์ โดยตัวเลขในวงเล็บลำดับแรกแสดงจำนวนเสียงพูดของผู้ชายลำดับที่สองแสดงจำนวนเสียงพูดของผู้หญิง

ชุดคำศัพท์ตัวเลขศูนย์ถึงเก้านั้น แต่ละคำศัพท์ประกอบด้วยเสียง 120 เสียง เป็นเสียงผู้ชายและเสียงผู้หญิงอย่างละ 30 คน โดยแต่ละคนผู้ 2 ครั้ง และสำหรับชุดคำศัพท์เสียงวรรณยุกต์นั้น หลังจากเลือกชุดคำศัพท์ทั้งหมดได้แล้ว จึงเลือกเสียงในฐานข้อมูลที่ตรงกับคำศัพท์มา คำศัพท์ละประมาณ 110 เสียง เป็นเสียงพูดเสียงผู้ชาย 54 คนและเสียงพูดผู้หญิง 1 คน โดยแต่ละคนพูด 2 ครั้ง

ตารางที่ 3.1 ชุดคำศัพท์ที่เลือกมาทดสอบแบ่งตามกลุ่มเสียงวรรณยุกต์

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)

3.1.2 รายละเอียดขั้นตอนการตัดพยางค์

เนื่องจากฐานเสียงที่มีอยู่มีทั้งฐานเสียงที่เป็นพยางค์เดียว และสองพยางค์ โดยฐานเสียงพยางค์เดียวมีไม่ครบกลุ่มเสียงวรรณยุกต์ทั้ง 5 จึงต้องมีการตัดพยางค์จากฐานเสียงสองพยางค์เพื่อให้ได้เสียงครบตามชุดคำศัพท์ทั้งหมด

หลักการตัดพยางค์จะใช้วิธีการพิจารณาจากค่าระดับพลังงานของเสียงพูด $E[m]$ ของสัญญาณเสียงพูดที่แบ่งเป็นเฟรมๆละ 100 ตัวอย่าง คือ

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

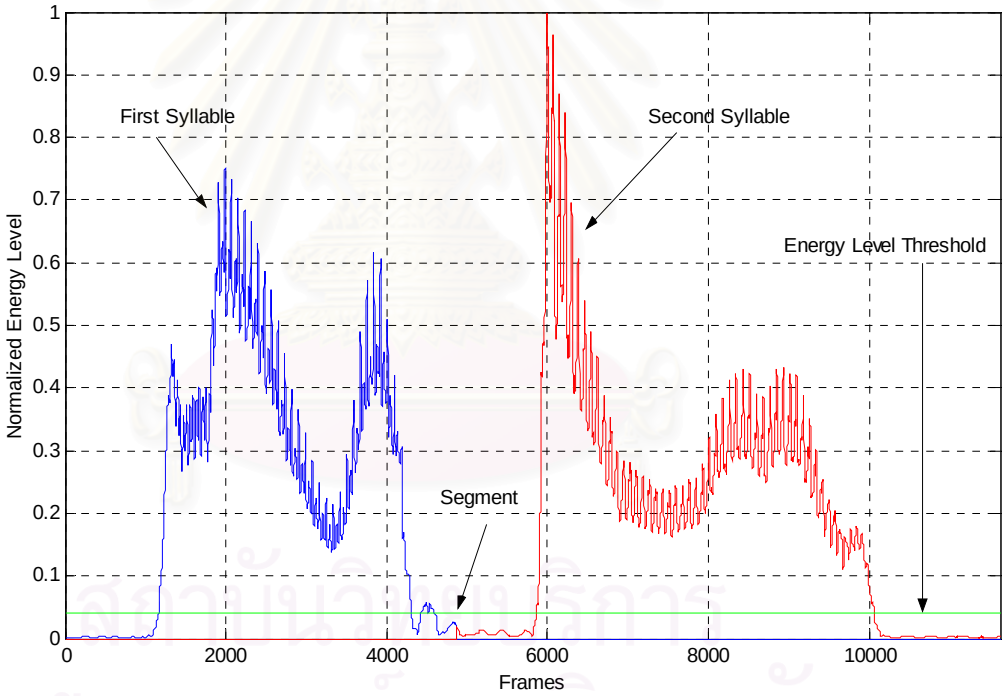
ในงานวิจัยนี้จะกำหนดจุดเปลี่ยนระดับพลังงาน (Energy Level Threshold) ที่ระดับพลังงาน ร้อยละ 20 ของค่าเฉลี่ยของระดับพลังงานเพื่อให้เหมาะสมกับระดับพลังงานของแต่ละเสียงพูด และกำหนดระยะเวลาขั้นต่ำของการเปลี่ยนระดับพลังงานที่ 200 เฟรม

ขั้นตอนแรกของการตัดพยางค์คือการหาจุดเริ่มต้นของพยางค์แรก จะเริ่มหาเฟรมแรกที่ระดับพลังงานสูงกว่าจุดเปลี่ยนระดับพลังงานเป็นระยะเวลา 200 เฟรม ต่อมาจึงหาจุดสิ้นสุดพยางค์แรก โดยหาเฟรมถัดมาที่ระดับพลังงานต่ำกว่าจุดเปลี่ยนระดับพลังงานเป็นระยะเวลา 200 เฟรม ซึ่งจะทำซ้ำกระบวนการดังกล่าวไปเรื่อยๆสำหรับพยางค์ถัดไปจนกระทั่งพิจารณาครบทุกเฟรม

ในหลายๆครั้งจะพบว่ามีจำนวนจุดเริ่มต้นและจุดสิ้นสุดของพยางค์เป็นจำนวนมากกว่าที่ควรจะเป็น (มากกว่า 2 พยางค์) จึงต้องหาจุดเริ่มต้นและสิ้นสุดของพยางค์ที่แท้จริงดังนี้ หาค่าเฉลี่ยของระดับพลังงานที่อยู่ระหว่างสองพยางค์ใดๆในขั้นตอนก่อนหน้านี้ แล้วหาว่าช่วงใดมีค่าเฉลี่ยระดับพลังงานต่ำสุด ช่วงดังกล่าวควรจะเป็นช่วงรอยต่อระหว่างสองพยางค์

เมื่อหาช่วงดังกล่าวพบแล้ว จะทำการแก้ไขให้จุดสิ้นสุดของพยางค์แรกและจุดเริ่มต้นพยางค์ถัดไปเป็นจุดเดียวกัน โดยใช้เฟรมที่แบ่งครึ่งค่าเฉลี่ยของระดับพลังงานที่อยู่ในช่วงรอยต่อระหว่างสองพยางค์ให้เป็นจุดแบ่งพยางค์

หลังจากขั้นตอนตัดพยางค์ข้างต้น จะทำการตรวจสอบเสียงพูดที่ตัดพยางค์แล้วด้วยการเปิดฟังเสียงพูดที่ได้ทุกเสียง หากพบว่าเสียงพูดใดมีการตัดพยางค์ไม่ถูกต้อง จะทำการแก้ไขด้วยตนเองโดยใช้โปรแกรมบันทึกเสียง Goldwave เวอร์ชัน 4.26 บนระบบปฏิบัติการ WindowXP



รูปที่ 3.1 แผนภูมิเส้นระดับพลังงานของเสียงพูด มือซ้าย

3.1.3 ขั้นตอนการเพิ่มเสียงสัญญาณรบกวน

ตัวอย่างเสียงพูดที่จัดเตรียมไว้ในขั้นตอนที่ผ่านมานั้นเป็นสัญญาณเสียงพูดที่ไม่มีสัญญาณรบกวน ดังนั้น ในขั้นตอนนี้จึงทำการเพิ่มสัญญาณรบกวนเข้าไปในสัญญาณเสียงพูดดังกล่าว โดยสัญญาณรบกวนเป็นเกาส์เซียนขาวค่าเฉลี่ยศูนย์แบบบวก โดยกำหนดค่าอัตราส่วนสัญญาณต่อสัญญาณรบกวน (Signal to

Noise Ratio, SNR) แตกต่างกันทั้งหมด 8 ค่า คือ Clean, 30 dB, 25 dB, 20 dB, 15 dB, 10 dB, 5 dB และ 0 dB ทั้งนี้เพื่อศึกษาความคงทนต่อสัญญาณรบกวนของค่าลักษณะสำคัญเมื่ออัตราส่วนสัญญาณต่อสัญญาณรบกวนแปรค่าไป ซึ่งการเพิ่มสัญญาณรบกวนจะกระทำการเพิ่มกับตัวอย่างเสียงพูดทั้งหมดในแต่ละชุดคำศัพท์กับทุกๆ SNR แล้วบันทึกไว้บนฮาร์ดดิสก์ โดยขั้นตอนการเพิ่มสัญญาณรบกวนจะกระทำผ่านโปรแกรม MATLAB เวอร์ชัน 6.1 บนระบบปฏิบัติการ WindowXP

3.2 รายละเอียดขั้นตอนการวิเคราะห์และการสกัดค่าลักษณะสำคัญ (Feature Analysis and Feature Extraction)

3.2.1 การประมวลผลสัญญาณเบื้องต้น (Signal Preprocessing)

ขั้นตอนการประมวลผลสัญญาณเบื้องต้นเป็นขั้นตอนแรก ซึ่งเป็นการจัดเตรียมข้อมูลจากข้อมูลเสียงที่ได้บันทึกไว้เพื่อนำไปสกัดค่าลักษณะสำคัญต่อไป ซึ่งขั้นตอนการประมวลผลสัญญาณเบื้องต้นในงานวิจัยนี้ประกอบไปด้วย 2 กรรมวิธีย่อยได้แก่ กรรมวิธีเน้นล่วงหน้า (Premphasis) และกรรมวิธีวางกรอบขนาดสัญญาณ (Smoothing Window)

3.2.1.1 กรรมวิธีเน้นล่วงหน้า (Premphasis)

กรรมวิธีเน้นล่วงหน้า เป็นขั้นตอนแรกที่กระทำกับสัญญาณเสียงพูด โดยนำสัญญาณเสียงพูดผ่านวงจรกรองผลตอบสนองอิมพัลส์จำกัด (Finite Impulse Response, FIR) อันดับหนึ่ง ดังสมการที่ (3.1) และ (3.2)

$$\tilde{s}[n] = s[n] - \alpha s[n-1] \tag{3.2}$$

$$H(z) = 1 - \alpha z^{-1} \tag{3.3}$$

เมื่อ α เป็นค่าสัมประสิทธิ์ของวงจรกรองซึ่งในงานวิจัยนี้กำหนดให้ $\alpha = 0.95$ [4,5,9] สำหรับงานวิจัยนี้การทำเน้นล่วงหน้าจะกระทำเพื่อทดสอบผลของการทำเน้นล่วงหน้ากับค่าลักษณะสำคัญว่ามีผลต่ออัตราการรู้จำอย่างไร

3.2.1.2 กรรมวิธีวางกรอบขนาดสัญญาณ (Smoothing Window)

กรรมวิธีวางกรอบขนาดสัญญาณเป็นกรรมวิธีต่อเนื่องจากกรรมวิธีเน้นล่วงหน้า กล่าวคือเนื่องมาจากสัญญาณเสียงพูดมีลักษณะเป็นสัญญาณที่มีการเปลี่ยนแปลงค่าสถิติตามเวลา (Non-stationary) และไม่เป็นสัญญาณเฟ้นสุ่ม (Non-Stochastic) ดังนั้นการประยุกต์ใช้สัญญาณเสียงพูดจึงจำเป็นอย่างยิ่งที่ต้องแบ่งสัญญาณเสียงพูดออกเป็นส่วนย่อยๆ ด้วยฟังก์ชันกรอบ (Window Function) ที่มีความยาวประมาณ 10 ถึง 30 มิลลิวินาที [10] ในงานวิจัยนี้ได้แบ่งสัญญาณเสียงพูดด้วยการคูณสัญญาณเสียงพูดกับ

ฟังก์ชันกรอบชนิด Hamming ดังสมการที่ (2.1) และ (2.2) เพื่อเป็นการลดทอนแอมปลิจูดที่บริเวณปลายแต่ละข้างของกรอบข้อมูลเสียงพูดอย่างช้าๆ และหลีกเลี่ยงความไม่ต่อเนื่อง

3.2.2 การสกัดค่าลักษณะสำคัญ (Feature Extraction)

การสกัดค่าลักษณะสำคัญคือขั้นตอนในการหาค่าลักษณะสำคัญของสัญญาณเสียงพูดจากข้อมูลเสียงพูด ค่าลักษณะสำคัญที่ใช้ในงานวิจัยนี้ประกอบไปด้วย ค่าลักษณะสำคัญ 4 ตัว คือ สัมประสิทธิ์การประมาณพันธะเชิงเส้นจากเทคนิคลดสัญญาณรบกวนแบบปรับตัวได้ (ANR-LPC) สัมประสิทธิ์เซปสตรอลจากเทคนิคลดสัญญาณรบกวนแบบปรับตัวได้ (ANR-CEP) ที่นำเสนอในงานวิจัย และสัมประสิทธิ์การประมาณพันธะเชิงเส้น (LPC) และสัมประสิทธิ์เซปสตรอล (CEP) เพื่อเป็นตัวเปรียบเทียบ โดยมีรายละเอียดการคำนวณหาค่าดังต่อไปนี้

3.2.2.1 สัมประสิทธิ์การประมาณพันธะเชิงเส้น (Linear Prediction Coefficient, LPC)

การคำนวณค่าสัมประสิทธิ์การประมาณพันธะเชิงเส้นจะใช้วิธีอัตสหสัมพันธ์ ซึ่งเมตริกซ์อัตสหสัมพันธ์อยู่ในรูปเมตริกซ์ Toeplitz ซึ่งมีลักษณะสมมาตร และทุกสมาชิกในแนวทแยงมุมมีค่าเท่ากัน จึงแก้สมการเพื่อหาค่าสัมประสิทธิ์ด้วยขั้นตอนวิธีการ Levinson-Derbin ในแต่ละกรอบข้อมูลเสียงพูด ดังมีรายละเอียดแสดงไว้ในหัวข้อที่ 2.2.1

ในการหาค่าสัมประสิทธิ์การประมาณพันธะเชิงเส้นสำหรับเสียงพูดนั้น จะหาค่าสัมประสิทธิ์เฉพาะแต่ละกรอบข้อมูลโดยการแทนที่แต่ละกรอบข้อมูลด้วยเวกเตอร์ของสัมประสิทธิ์ของการประมาณพันธะเชิงเส้น สำหรับงานวิจัยนี้จะแปรค่าลำดับของสัมประสิทธิ์ที่ 10 12 14 และ 16 เพื่อทดสอบศักยะภาพของค่าลักษณะสำคัญ

3.2.2.2 สัมประสิทธิ์เซปสตรอล (Cepstral Coefficient, CEP)

การคำนวณค่าสัมประสิทธิ์เซปสตรอล สามารถคำนวณได้ 2 วิธีคือคำนวณจากการแปลงฟูริเยร์อย่างเร็ว (Fast Fourier Transform, FFT) และคำนวณจากสัมประสิทธิ์แบบจำลองการประมาณพันธะเชิงเส้น โดยในงานวิจัยนี้เลือกวิธีคำนวณค่าสัมประสิทธิ์เซปสตรอลจากสัมประสิทธิ์แบบจำลองการประมาณพันธะเชิงเส้นซึ่งมีข้อดีคือใช้เวลาในการคำนวณน้อยกว่าและเพื่อต้องการเปรียบเทียบผลที่ได้กับสัมประสิทธิ์การประมาณพันธะเชิงเส้นด้วย ดังมีรายละเอียดแสดงไว้ในหัวข้อที่ 2.2.2 สำหรับงานวิจัยนี้จะแปรค่าลำดับของสัมประสิทธิ์ที่ 10 12 14 และ 16 เพื่อทดสอบศักยะภาพของค่าลักษณะสำคัญ

3.2.2.3 สัมประสิทธิ์การประมาณพันธะเชิงเส้นจากเทคนิคลดสัญญาณรบกวนแบบปรับตัวได้ (Adaptive Noise Reduction Technique Approach to Linear Prediction Coefficient, ANR-LPC)

การคำนวณค่าสัมประสิทธิ์การประมาณเชิงเส้นจากเทคนิคลดสัญญาณรบกวนแบบปรับตัวได้ คือการคำนวณค่าสัมประสิทธิ์การประมาณพันธะเชิงเส้นนั่นเอง แต่ได้ทำการปรับลดสัญญาณรบกวนตามขั้นตอนวิธีที่แสดงรายละเอียดในหัวข้อที่ 2.3.3 เพื่อให้ได้สัมประสิทธิ์การประมาณพันธะเชิงเส้นที่คงทนต่อสัญญาณรบกวน เพื่อใช้เป็นค่าลักษณะที่คงทนต่อสัญญาณรบกวนสำหรับงานวิจัยนี้ สำหรับงานวิจัยนี้จะแปรค่าลำดับของสัมประสิทธิ์ที่ 10 12 14 และ 16 เพื่อทดสอบศักยะภาพของค่าลักษณะสำคัญ

3.2.2.4 สัมประสิทธิ์เซปสตรอลจากเทคนิคลดสัญญาณรบกวนแบบปรับตัวได้ (Adaptive Noise Reduction Technique Approach to Cepstral Coefficient, ANR-CEP)

การคำนวณค่าสัมประสิทธิ์เซปตรอลจากเทคนิคลดสัญญาณรบกวนแบบปรับตัวได้ คือการคำนวณค่าสัมประสิทธิ์เซปตรอลนั่นเอง แต่ได้ทำการปรับลดสัญญาณรบกวนตามขั้นตอนวิธีที่แสดงรายละเอียดในหัวข้อที่ 2.3.3 เพื่อให้ได้สัมประสิทธิ์เซปตรอลที่คงทนต่อสัญญาณรบกวน เพื่อใช้เป็นค่าลักษณะที่คงทนต่อสัญญาณรบกวนสำหรับงานวิจัยนี้ สำหรับงานวิจัยนี้จะแปรค่าลำดับของสัมประสิทธิ์ที่ 10 12 14 และ 16 เพื่อทดสอบศักยะภาพของค่าลักษณะสำคัญ

3.3 รายละเอียดขั้นตอนการปรับลดสัญญาณรบกวนแบบปรับตัวได้

ในหัวข้อนี้ จะกล่าวถึงขั้นตอนการสกัดค่าลักษณะสำคัญแบบคงทนต่อสัญญาณรบกวนโดยใช้เทคนิคลดสัญญาณแบบปรับตัวได้ กล่าวคือ ค่าอินพุทคือสัญญาณเสียงพูดแต่ละคำโดยกระทำกับทุกกรอบสัญญาณเสียงพูด (Frame) ซึ่งมีรายละเอียดขั้นตอนดังนี้

1. ทำการแบ่งสัญญาณเสียงพูดเป็นส่วนย่อยโดยผ่านกรรมวิธีการวางกรอบขนาดสัญญาณ (Smoothing Window) เพื่อใช้ในการหาค่าอัตสหสัมพันธ์โดยการคูณค่าของสัญญาณในกรอบข้อมูลเสียงพูดด้วยค่าฟังก์ชันกรอบ (Window Function) ชนิด Hamming Window
2. หาค่าอัตสหสัมพันธ์ของสัญญาณเสียงพูด $R(i)|_{i=0}^{p}$ โดย p คืออันดับของค่าลักษณะสำคัญ
3. สร้างเมตริกซ์อัตสหสัมพันธ์ $\mathbf{R_{xx}}$ ขนาด $(p+1)\times(p+1)$ คือ

$$\mathbf{R}_{xx}(\eta)=\begin{bmatrix} R_{xx}(0) & R_{xx}(1) & R_{xx}(2) & . & . & R_{xx}(p) \\ R_{xx}(1) & R_{xx}(0) & R_{xx}(1) & . & . & R_{xx}(p-1) \\ R_{xx}(2) & R_{xx}(1) & R_{xx}(0) & . & . & R_{xx}(p-2) \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ R_{xx}(p) & R_{xx}(p-1) & R_{xx}(p-2) & . & . & R_{xx}(0) \end{bmatrix} \quad (3.3)$$

4. หาค่าเจาะจงต่ำสุด $\lambda_{\min}$ ของเมตริกซ์อัตสหสัมพันธ์ $\mathbf{R_{xx}}$ ด้วยวิธีกำลังสมมาตร (Power Method)

5. จากสมการการปรับลดสัญญาณรบกวน

$$\hat{\mathbf{a}}^{\mathsf{T}}(\mathbf{R_{xx}}-\beta\mathbf{I})\hat{\mathbf{a}} \cong \breve{\eta}\sum_{k=0}^{p}\hat{a}_k^2 \quad (3.4)$$

เมื่อ β แทนสัญญาณรบกวนส่วนที่หักออกจาก $R_{nn}(0)$ และ $\breve{\eta}$ แทนส่วนที่เหลืออยู่ของความแปรปรวนของสัญญาณรบกวน และ $\hat{\mathbf{a}}$ แทนเวกเตอร์สัมประสิทธิ์การประมาณพันธะเชิงเส้นที่ได้จากลำดับอัตสหสัมพันธ์ที่ผ่านการปรับลดสัญญาณรบกวนแล้ว คือ $\{R_{xx}(0)-\beta, R_{xx}(1),\ldots,R_{xx}(p)\}$

หลักการคือลดค่า $\breve{\eta}$ ลงในระดับที่ยอมรับได้ ซึ่งเป็นค่าที่ยังคงไว้ซึ่งคุณสมบัติของสัญญาณที่ยังไม่มีสัญญาณรบกวนเข้ามา ซึ่งในที่นี้เลือกค่า 15 dB [17] ระหว่างค่ากำลังของสัญญาณระหว่างสัญญาณที่ผ่านการปรับลดแล้วกับสัญญาณรบกวนที่เหลือ นั่นคือ

$$\breve{\eta}=\frac{R_{xx}(0)-\beta}{10^{1.5}} \quad (3.5)$$

โดยหาค่า β ที่ทำให้

$$f(x)=\hat{\mathbf{a}}^{T}(\mathbf{R_{xx}}-x\mathbf{I})\hat{\mathbf{a}}-\frac{R_{xx}(0)-x}{10^{1.5}}\left(\sum_{k=0}^{p}\hat{a}_k^2\right)=0 \quad (3.6)$$

โดย x เป็นค่า β ที่ทำให้สมการที่ (3.6) เป็นจริง

ในขั้นตอนนี้จะหาคำตอบด้วยขั้นตอนวิธีการแบ่งส่วน (Bisection Algorithm) โดยมีค่าเริ่มต้นเป็น 0 และ $\lambda_{\min}$ (คำนวณหาด้วยวิธีกำลังสมมาตรดังรายละเอียดในหัวข้อ 2.4.3

6. เมื่อได้ค่า x ซึ่งเป็นค่า β ที่ทำให้สมการที่ (3.6) เป็นจริงแล้ว ทำการหาค่าสัมประสิทธิ์การประมาณพันธะเชิงเส้นด้วยค่าอัตสหสัมพันธ์ $\{R_{xx}(0)-x, R_{xx}(1),\ldots,R_{xx}(p)\}$ ซึ่งเป็นค่าลักษณะสำคัญแบบคงทนต่อสัญญาณรบกวนชนิด ANR-LPC

7. ค่าลักษณะสำคัญ ANR-CEP หาได้จากค่าลักษณะสำคัญ ANR-LPC

จากขั้นตอนทั้งหมดแสดงให้เห็นถึงการปรับลดสัญญาณรบกวนแบบปรับตัวได้ (Adaptive Noise Reduction, ANR) โดยค่าพารามิเตอร์ในการปรับลดคือค่าเจาะจงต่ำสุดซึ่งเปลี่ยนไปในแต่ละกรอบสัญญาณเสียงพูดที่เข้ามา ดังนั้นเทคนิคนี้จึงมีการปรับลดสัญญาณรบกวนโดยขึ้นอยู่กับค่าสัญญาณรบกวนที่เข้ามาซึ่งได้ส่งผ่านคุณลักษณะของตัวสัญญาณรบกวนเองมาทางค่าเจาะจงนั่นเอง

3.4 รายละเอียดขั้นตอนการฝึกฝนและการทดสอบ

เมื่อสัญญาณเสียงพูดผ่านขั้นตอนการวิเคราะห์และสกัดค่าลักษณะสำคัญเป็นที่เรียบร้อยแล้ว ในขั้นนี้เรามีข้อมูลในรูปแบบของเวกเตอร์ของแต่ละกรอบสัญญาณ ค่าลักษณะสำคัญที่ได้นี้จะเป็นข้อมูลฝึกฝนในขั้นตอนการฝึกฝนเสียงพูดภาษาไทย

ในขั้นตอนการฝึกฝนชุดพารามิเตอร์ของแบบจำลองฮิดเดนมาร์คอฟ (Hidden Markov Model Parameter Training) จะใช้เสียงพูดแต่ละคำของผู้พูดตามชุดข้อมูลที่จัดเตรียมไว้มาฝึกฝนร่วมกันในแต่ละครั้ง หลังจากการฝึกฝนแล้วจะได้ชุดรูปร่องต้นแบบอ้างอิงของเสียงพูดแต่ละคำซึ่งเก็บค่าเฉลี่ย (Mean) และค่าความแปรปรวน (Variance) สำหรับแต่ละ Gaussian Mixture ของทุกสถานะสำหรับแต่ละเสียงพูดเอาไว้ และยังได้เก็บค่าเมตริกซ์ของการเปลี่ยนสถานะไว้ด้วย

ในขั้นตอนการทดสอบการรู้จำเสียงพูดโดยใช้แบบจำลองฮิดเดนมาร์คอฟแบบต่อเนื่องประกอบไปด้วย 3 ขั้นตอน ได้แก่ขั้นตอนการประมวลผลสัญญาณเบื้องต้น (Signal Preprocessing) ขั้นตอนการวิเคราะห์และวัดค่าลักษณะสำคัญ (Feature Measurement) และขั้นตอนการจำแนกรูปแบบร่วมกับขั้นตอนวิธีการตัดสินใจ (Pattern Classification and Decision Making Algorithm) ตามลำดับ ซึ่งในขั้นตอนการจำแนกรูปร่องต้นแบบอ้างอิงของเสียงพูดแต่ละคำโดยการเปรียบเทียบเสียงพูดที่นำมาทดสอบกับชุดรูปร่องต้นแบบอ้างอิงของเสียงพูดเพื่อใช้ในขั้นตอนวิธีการตัดสินใจ สำหรับในงานวิจัยนี้ ขั้นตอนวิธีการตัดสินใจจะอาศัยแบบจำลองฮิดเดนมาร์คอฟด้วยขั้นตอนวิธีการ Viterbi (Viterbi Algorithm) ผลลัพธ์ที่ได้ในขั้นตอนนี้จะเป็นชุดของเสียงพูดที่รู้จำได้ซึ่งให้ค่าความน่าจะเป็นสูงที่สุด

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 4

ผลการวิจัยและการวิเคราะห์ผล

ในบทนี้ เบื้องต้นจะกล่าวถึงขั้นตอนวิธีการรู้จำเสียงพูดภาษาไทย จากนั้นจะแสดงผลการทดลองเพื่อหาระบบที่เหมาะสมสำหรับการรู้จำ ตามด้วยผลการทดลองเพื่อวิเคราะห์ผลของการทำเน้นล่วงหน้า และสุดท้ายซึ่งสำคัญที่สุดจะกล่าวถึงรายละเอียดของผลการรู้จำเสียงพูดภาษาไทย ซึ่งแสดงผลการวิจัยที่ได้ในแต่ละกรณี และวิเคราะห์ผลที่ได้จากการทดสอบ โดยทำการทดสอบกับเสียงพูดตัวเลขภาษาไทย และเสียงวรรณยุกต์ภาษาไทยทั้งห้าเสียงคือเสียงสามัญ เอก โท ตรี และจัตวา เพื่อทดสอบความสามารถในการรู้จำเสียงพูดภาษาไทยซึ่งมีวรรณยุกต์เป็นส่วนสำคัญของภาษาไทย จากนั้นจะแสดงรายละเอียดการวิเคราะห์ประกอบผลการวิจัยทั้งหมด

4.1 ขั้นตอนวิธีการรู้จำเสียงพูดภาษาไทย

ขั้นตอนวิธีการรู้จำเสียงพูดภาษาไทยสำหรับงานวิจัยนี้ใช้แบบจำลองฮิดเดนมาร์คอฟแบบต่อเนื่อง จะนำชุดพารามิเตอร์ของแบบจำลองฮิดเดนมาร์คอฟสำหรับคำศัพท์แต่ละคำที่ได้จากการฝึกฝนมาใช้เป็นชุดรูปร่องต้นแบบอ้างอิงในการรู้จำ โดยแบ่งชุดพารามิเตอร์ของคำศัพท์แต่ละคำดังนี้คือ ชุดพารามิเตอร์ของคำศัพท์ตัวเลขภาษาไทยศูนย์ถึงเก้า ชุดพารามิเตอร์ของคำศัพท์จำแนกตามวรรณยุกต์ ซึ่งประกอบด้วย ชุดพารามิเตอร์ของคำศัพท์จำแนกตามวรรณยุกต์เสียงสามัญ ชุดพารามิเตอร์ของคำศัพท์จำแนกตามวรรณยุกต์เสียงเอก ชุดพารามิเตอร์ของคำศัพท์จำแนกตามวรรณยุกต์เสียงโท ชุดพารามิเตอร์ของคำศัพท์จำแนกตามวรรณยุกต์เสียงตรี และชุดพารามิเตอร์ของคำศัพท์จำแนกตามวรรณยุกต์เสียงจัตวา ตามลำดับ

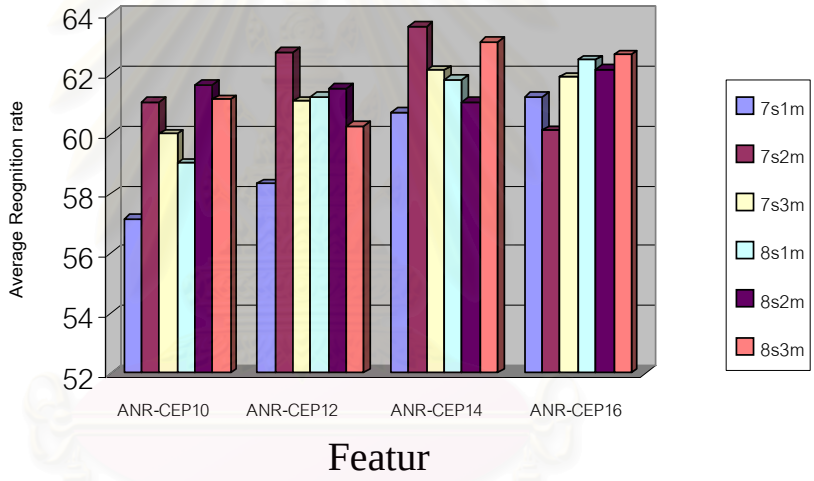
ในระหว่างกระบวนการรู้จำเสียงพูดภาษาไทย จะทำการเปรียบเทียบเสียงพูดทดสอบกับชุดรูปร่องต้นแบบอ้างอิง (Word Reference Templates) โดยใช้ชุดพารามิเตอร์ของคำศัพท์ตัวเลขภาษาไทยศูนย์ถึงเก้ามาทดสอบด้วยคำศัพท์ตัวเลขภาษาไทยจากศูนย์ถึงเก้าเช่นเดียวกัน และใช้ชุดพารามิเตอร์ของคำศัพท์จำแนกวรรณยุกต์เดียวกันตรงกับคำศัพท์จำแนกตามวรรณยุกต์ที่นำมาทดสอบ การเปรียบเทียบจะกระทำกับชุดรูปร่องต้นแบบอ้างอิงของคำศัพท์ทุกคำในชุดพารามิเตอร์ ผลการรู้จำจะได้ความน่าจะเป็นของแต่ละคำศัพท์ที่ได้รับการเปรียบเทียบ เมื่อทำการเปรียบเทียบทั้งชุดแล้ว จะเรียงลำดับความน่าจะเป็นจากสูงที่สุดไปหาต่ำที่สุด และคำศัพท์ที่รู้จำได้จะเป็นคำศัพท์ที่มีค่าความน่าจะเป็นสูงที่สุด

อัตราการรู้จำ (Recognition Rate) หมายถึงอัตราความถูกต้องในการรู้จำคำศัพท์แต่ละคำโดยคิดเป็นร้อยละของจำนวนเสียงพูดทั้งหมดที่นำมาทดสอบ การคำนวณค่าอัตราการรู้จำที่ใช้ในงานวิจัยนี้จะนำชุดเสียงพูดทั้งหมดมาแบ่งเป็นสามส่วนเท่ากัน โดยทำการฝึกฝนกับส่วนที่หนึ่งและทดสอบกับอีกสองส่วนที่เหลือ โดยทำการฝึกฝนและทดสอบแบบนี้สามครั้ง พร้อมทั้งบันทึกจำนวนครั้งที่ระบบรู้จำถูกต้องและไม่ถูกต้องและ

อัตราความแม่นยำในการรู้จำของคำศัพท์แต่ละคำสามารถคำนวณได้จากอัตราส่วนระหว่างจำนวนเสียงที่ระบบรู้จำถูกต้องกับจำนวนเสียงพูดทั้งหมดที่นำมาทดสอบรวมทั้งสามครั้ง อัตราความแม่นยำเฉลี่ยจะเป็นค่าเฉลี่ยของอัตราความแม่นยำในการรู้จำของคำศัพท์ทั้งหมดในชุดคำศัพท์

4.2 การทดลองเพื่อหาระบบที่เหมาะสมสำหรับการรู้จำ

การทดลองนี้เป็นการหาระบบการรู้จำที่เหมาะสมกับค่าลักษณะสำคัญแบบคงทนต่อสัญญาณรบกวนที่ใช้ โดยใช้ชุดคำศัพท์เสียงตัวเลขภาษาไทยจากศูนย์ถึงเก้ามาทำการทดสอบระบบ โดยการแปรค่าสถานะของแบบจำลองฮิดเดนมาร์คอฟ และจำนวน Gaussians Mixture เพื่อพิจารณาค่าอัตราการรู้จำเฉลี่ยสูงที่สุด ซึ่งได้ผลดังแสดงไว้ในตารางที่ ก.1



รูปที่ 4.1 อัตราการรู้จำเสียงพูดภาษาไทยเนื่องจากผลผลของการแปรค่าสถานะและ Gaussian Mixture

เมื่อพิจารณารูปที่ 4.1 จะเห็นว่า อัตราการรู้จำของระบบที่มีสถานะและค่า Gaussians Mixture ค่าต่างๆนั้นให้อัตราอัตราการรู้จำไม่แตกต่างกันนัก ซึ่งในบางครั้งการเพิ่มจำนวนสถานะและจำนวน Gaussians Mixture ก็ไม่ได้ส่งผลดีโดยรวมกับทุกค่าสัญญาณต่อสัญญาณรบกวน (SNR) เช่นถ้าพิจารณาอัตราการรู้จำจากระบบที่มีจำนวนสถานะและจำนวน Gaussian Mixture เท่ากับ 7 สถานะและ 2 Mixture (7s2m) ตามลำดับ กับระบบที่มีจำนวนสถานะและจำนวน Gaussian Mixture เท่ากับ 8 สถานะและ 3 Mixture (8s3m)ตามลำดับนั้น จะเห็นว่า ที่ค่า SNR Clean สำหรับค่าลักษณะสำคัญ ANR-CEP10 ระบบ 7s2m ให้อัตราการรู้จำร้อยละ 69.96 ส่วนระบบ 8s3m ให้อัตราการรู้จำสูงขึ้นเป็นร้อยละ 74.29 และสำหรับค่าลักษณะสำคัญ ANR-CEP14 ANR-CEP16 ก็ให้อัตราการรู้จำสูงขึ้นเช่นกัน แต่สำหรับค่าลักษณะสำคัญ ANR-CEP12 แทบไม่มีความแตกต่างของอัตราการรู้จำ เมื่อพิจารณาค่าอัตราการรู้จำโดยรวมของระบบแล้วจะพบว่าจำนวนสถานะและจำนวน Gaussians Mixture มีผลทำให้ระบบมีแนวโน้มของอัตราการรู้จำสูงขึ้นสำหรับค่า

SNR สูงๆคือประมาณ 25 dB ขึ้นไป และส่วนใหญ่จะให้ค่าอัตราการรู้จำลดลงเมื่อจำนวนสถานะและจำนวน Gaussian Mixture เพิ่มขึ้นสำหรับค่า SNR ที่ลดลง

ในสดมภ์สุดท้ายของตารางที่ ก.1 ได้แสดงค่าเฉลี่ยของอัตราการรู้จำของแต่ละระบบสำหรับทุกอันดับของค่าลักษณะสำคัญเอาไว้ จะเห็นว่าไม่มีระบบใดที่ให้อัตราการรู้จำสูงที่สุดอย่างเอกฉันท์ ดังนั้นเพื่อพิจารณาหาระบบที่เหมาะสมกับทุก SNR และทุกอันดับของค่าลักษณะสำคัญ จึงคำนวณหาค่าเฉลี่ยของอัตราการรู้จำของระบบแต่ละระบบด้วยการนำค่าอัตราการรู้จำแต่ละ SNR ของระบบนั้นๆมาหาค่าเฉลี่ยเพื่อหาอัตราการรู้จำเฉลี่ยรวมสูงที่สุด ซึ่งแสดงไว้ในตารางที่ 4.1

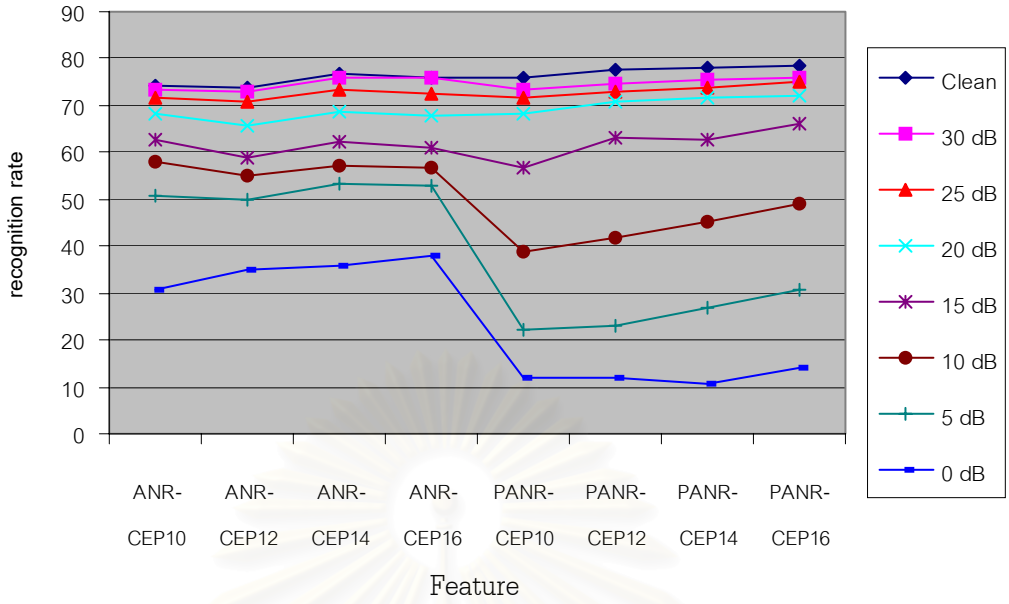
ตารางที่ 4.1 ผลของระบบต่ออัตราการรู้จำเฉลี่ยรวม

จำนวนสถานะ/ Guassian Mixture	**อัตราการรู้จำ (ร้อยละ)**
7s1m	59.3750
7s2m	**61.8925**
7s3m	61.3000
8s1m	61.1500
8s2m	61.6200
8s3m	61.7975

จากตารางที่ 4.1 จำนวนสถานะในแบบจำลองฮิดเดนมาร์คอฟเป็น 7 และจำนวน Gaussian Mixture เป็น 2 ให้อัตราการรู้จำเฉลี่ยรวมสูงที่สุด ดังนั้นวิทยานิพนธ์นี้จึงเลือกระบบฮิดเดนมาร์คอฟที่มีจำนวนสถานะเป็น 7 และจำนวน Gaussians Mixture เป็น 2 เป็นระบบสำหรับการรู้จำ

4.3 การทดลองเพื่อวิเคราะห์ผลของการทำเน้นล่วงหน้า

การทดลองนี้เป็นการศึกษาถึงผลของการทำเน้นล่วงหน้า (Preemphasis) กับข้อมูลเสียงพูดก่อนการหาค่าลักษณะสำคัญ ซึ่งได้ค่าอัตราการรู้จำดังแสดงในตารางที่ ข.1



รูปที่ 4.2 ผลการทำเน้นล่วงหน้าต่ออัตราการรู้จำ

จากรูปที่ 4.2 ANR-CEP แทนค่าลักษณะสำคัญชนิดค่าสัมประสิทธิ์เซปตรอลจากเทคนิคลดสัญญาณรบกวนแบบปรับตัวได้ที่ไม่ทำเน้นล่วงหน้า และ PANR-CEP แทนค่าลักษณะสำคัญชนิดค่าสัมประสิทธิ์เซปตรอลจากเทคนิคลดสัญญาณรบกวนแบบปรับตัวได้ที่ทำเน้นล่วงหน้า เมื่อพิจารณาผลที่ได้จะเห็นว่า การทำเน้นล่วงหน้ามีผลดีสำหรับการรู้จำเสียงพูดในบริเวณที่ไม่มีสัญญาณรบกวน ส่วนที่ค่า SNR 30 dB ลงมาจนถึงประมาณ 15 dB การทำเน้นล่วงหน้าแทบไม่มีผลต่ออัตราการรู้จำเลย หรือมีก็น้อยมาก เพราะการทำเน้นล่วงหน้าจะทำให้ค่าอัตราการรู้จำสูงขึ้นสำหรับเสียง Clean และ SNR สูงๆ เท่านั้น แต่เมื่อค่า SNR ต่ำลง จะเห็นว่าค่าอัตราการรู้จำลดต่ำลงอย่างเห็นได้ชัดเจนตั้งแต่ค่า SNR ต่ำกว่า 10 dB ดังนั้น เนื่องจากงานวิจัยนี้มุ่งเน้นการวิเคราะห์การรู้จำเสียงพูดเมื่อมีสัญญาณรบกวน จึงพิจารณาให้ค่าลักษณะสำคัญที่ใช้ไม่ทำเน้นล่วงหน้า

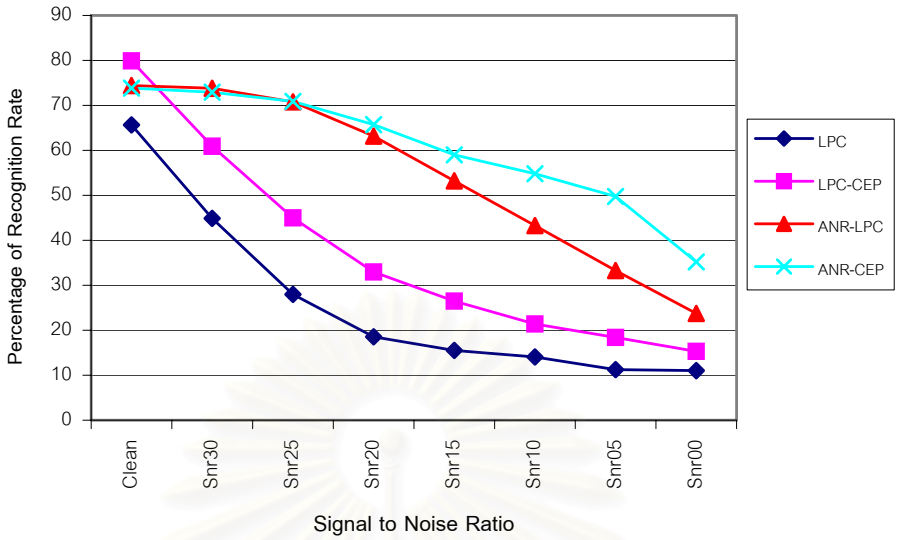
4.4 ผลการรู้จำเสียงพูดภาษาไทย

ผลการรู้จำเสียงพูดภาษาไทยของชุดคำศัพท์ แบ่งออกเป็น ชุดคำศัพท์ตัวเลขภาษาไทย ชุดคำศัพท์ชุดคำศัพท์เสียงวรรณยุกต์สามัญ ชุดคำศัพท์เสียงวรรณยุกต์เอก ชุดคำศัพท์เสียงวรรณยุกต์โท ชุดคำศัพท์เสียงวรรณยุกต์ตรี และชุดคำศัพท์เสียงวรรณยุกต์จัตวา โดยชุดเสียงพูดทั้งหมดมาแบ่งเป็นสามส่วนที่เท่ากัน โดยทำการฝึกฝนกับส่วนที่หนึ่งและทดสอบกับอีกสองส่วนที่เหลือ โดยทำการฝึกฝนและทดสอบแบบนี้สามครั้ง

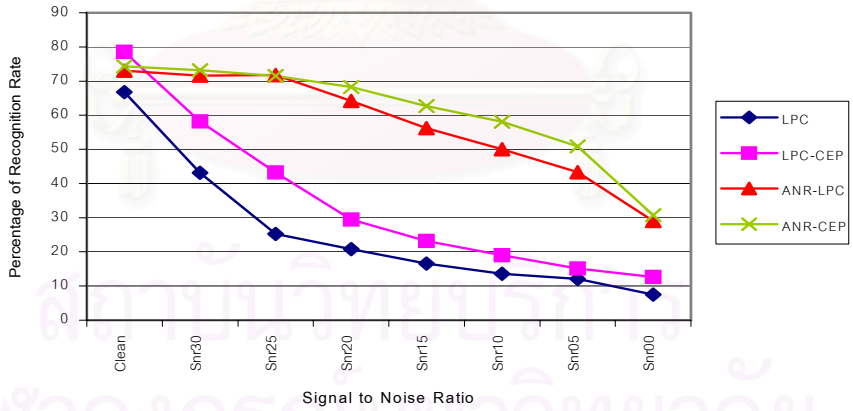
ผลการรู้จำเสียงพูดภาษาไทยของชุดทดสอบ แบ่งออกเป็นผลการรู้จำเสียงพูดโดยใช้ค่าลักษณะสำคัญคือ สัมประสิทธิ์การประมาณพันธะเชิงเส้น (LPC) สัมประสิทธิ์เซปตรอล (CEP) สัมประสิทธิ์การประมาณพันธะเชิงเส้นจากเทคนิคลดสัญญาณรบกวนแบบปรับตัวได้ (ANR-LPC) สัมประสิทธิ์เซปสตรอลจากเทคนิคลดสัญญาณรบกวนแบบปรับตัวได้ (ANR-CEP) โดยค่าลักษณะสำคัญแต่ละตัวจะแปรค่า SNR ต่างๆกัน 8 ค่า คือ ค่า SNR 00 dB, SNR 05 dB, SNR 10 dB, SNR 15 dB, SNR 20 dB, SNR 25 dB SNR 30 dB และ Clean speech ซึ่งผลการรู้จำทั้งหมดแสดงดังตารางในภาคผนวก ค

ในรูปที่ 4.3 ถึงรูปที่ 4.26 ได้แสดงอัตราการรู้ จำเสียงพูดภาษาไทยทุกชุดคำศัพท์สำหรับแต่ละอันดับของค่าลักษณะสำคัญ และรูปที่ 4.27 ถึงรูปที่ 4.32 แสดงผลการเปรียบเทียบของอัตราการรู้จำเสียงตัวเลขภาษาไทยและของชุดคำศัพท์เสียงวรรณยุกต์ต่างๆ โดยเฉลี่ยทุกค่า SNR สำหรับค่าลักษณะสำคัญอันนั้นๆ เอาไว้ และรูปที่ 4.33 ถึงรูปที่ 4.40 ได้แสดงอัตราการรู้จำเฉลี่ยรวมของทุกชุดคำศัพท์ในแต่ละค่า SNR เอาไว้ และสุดท้าย ในรูปที่ 4.41 ได้แสดงอัตราการรู้จำเฉลี่ยรวมของระบบ

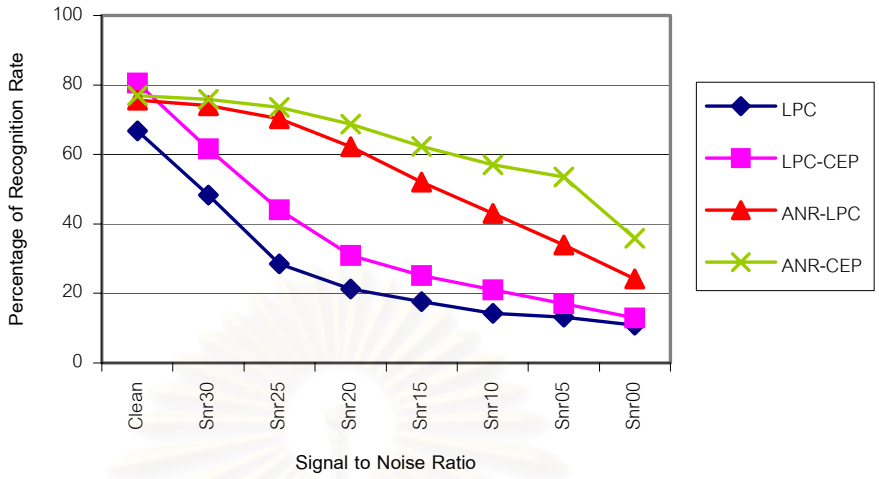
สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



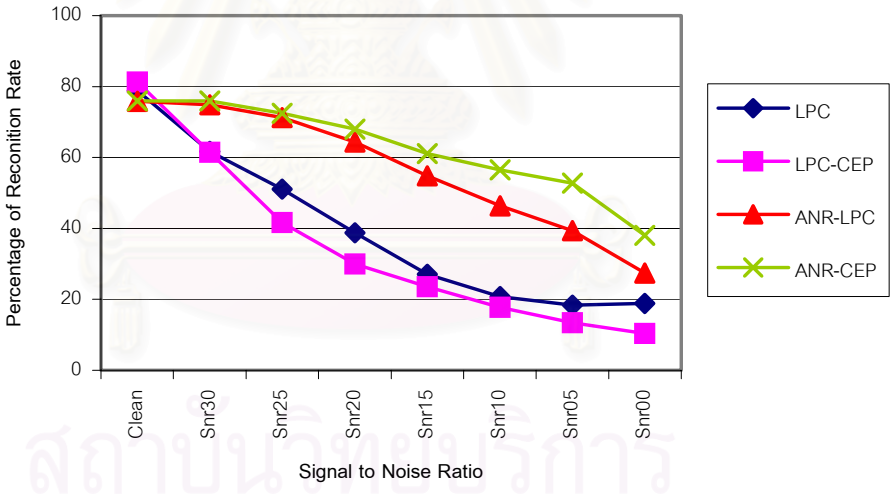
รูปที่ 4.3 อัตราการรู้จำเสียงตัวเลขภาษาไทยด้วยลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC และ ANR-CEP อันดับ 10



รูปที่ 4.4 อัตราการรู้จำเสียงตัวเลขภาษาไทยด้วยลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC และ ANR-CEP อันดับ 12

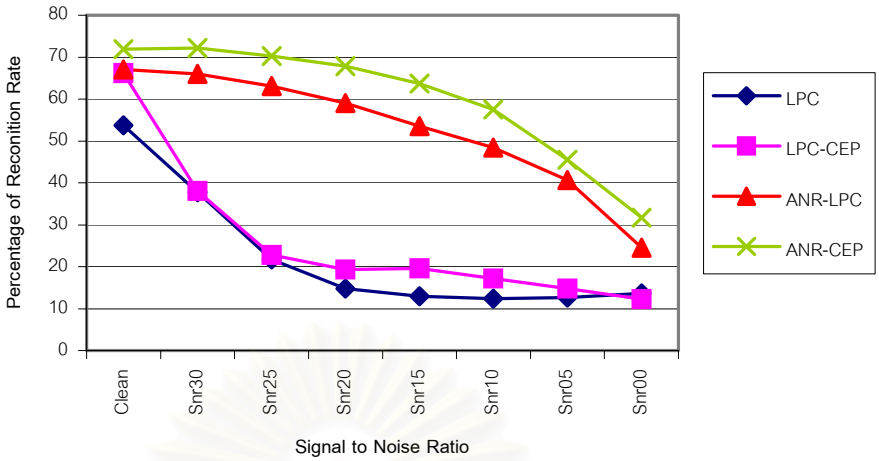


รูปที่ 4.5 อัตราการรู้จำเสียงตัวเลขภาษาไทยด้วยลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC และ ANR-CEP อันดับ 14

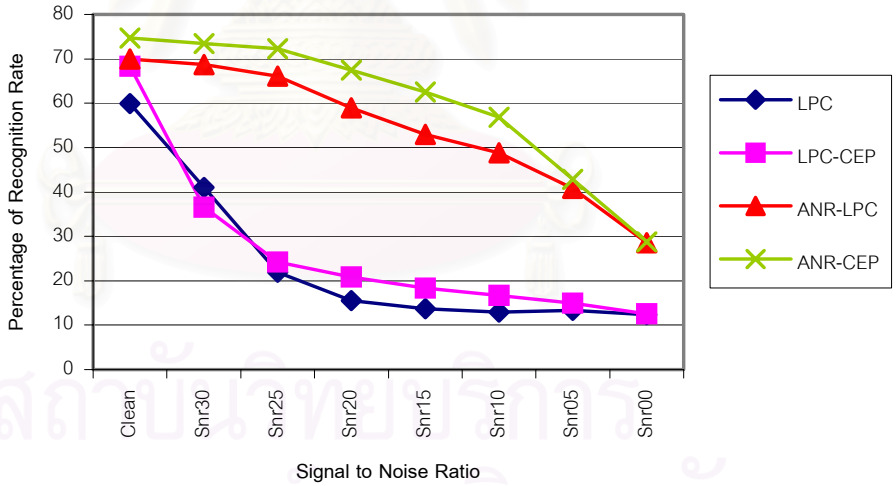


รูปที่ 4.6 อัตราการรู้จำเสียงตัวเลขภาษาไทยด้วยลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC และ ANR-CEP อันดับ 16

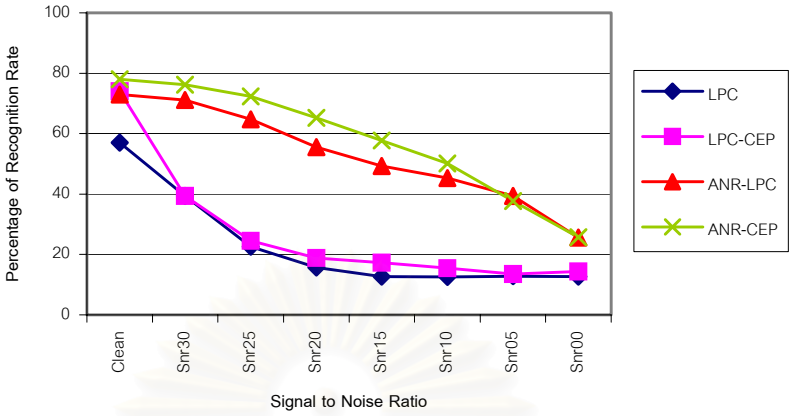
สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



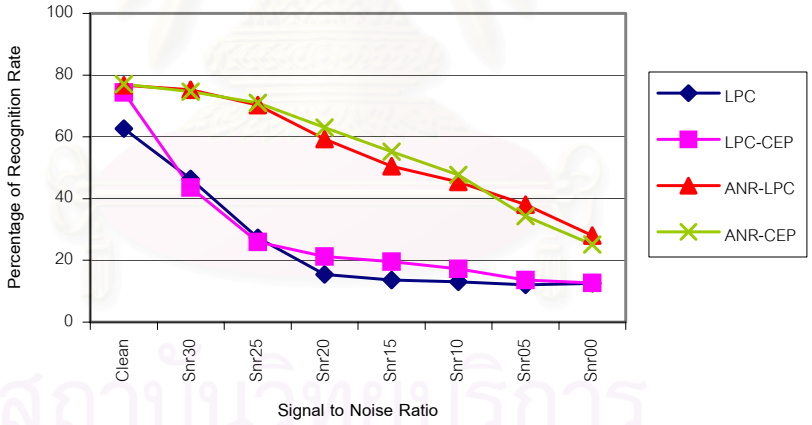
รูปที่ 4.7 อัตราการรู้จำเสียงวรรณยุกต์ไทยเสียงสามัญด้วยลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC และ ANR-CEP อันดับ 10



รูปที่ 4.8 อัตราการรู้จำเสียงวรรณยุกต์ไทยเสียงสามัญด้วยลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC และ ANR-CEP อันดับ 12

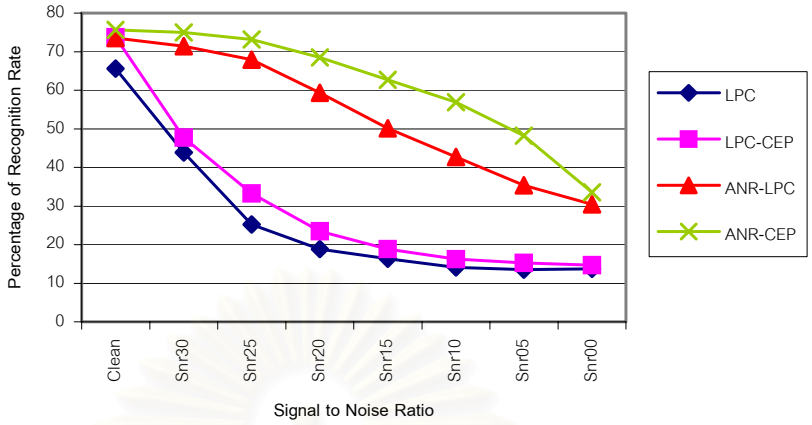


รูปที่ 4.9 อัตราการรู้จำเสียงวรรณยุกต์ไทยเสียงสามัญด้วยลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC และ ANR-CEP อันดับ 14

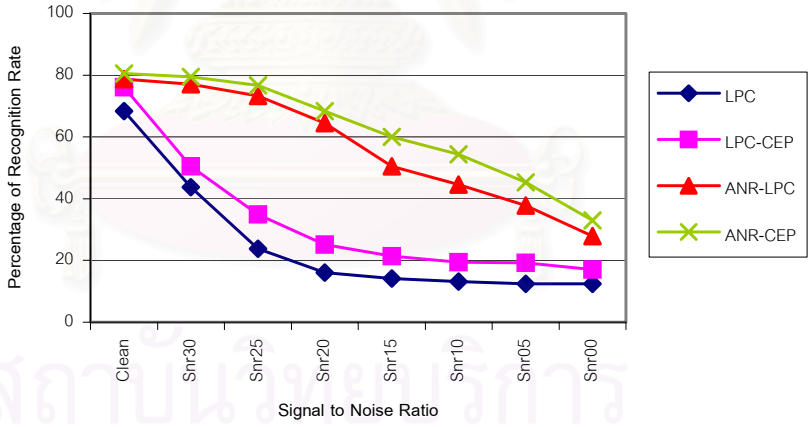


รูปที่ 4.10 อัตราการรู้จำเสียงวรรณยุกต์ไทยเสียงสามัญด้วยลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC และ ANR-CEP อันดับ 16

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

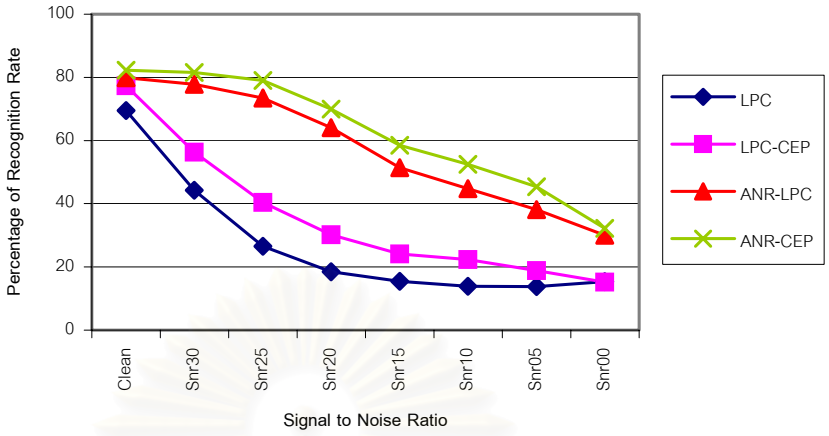


รูปที่ 4.11 อัตราการรู้จำเสียงวรรณยุกต์ไทยเสียงเอกด้วยลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC และ ANR-CEP อันดับ 10

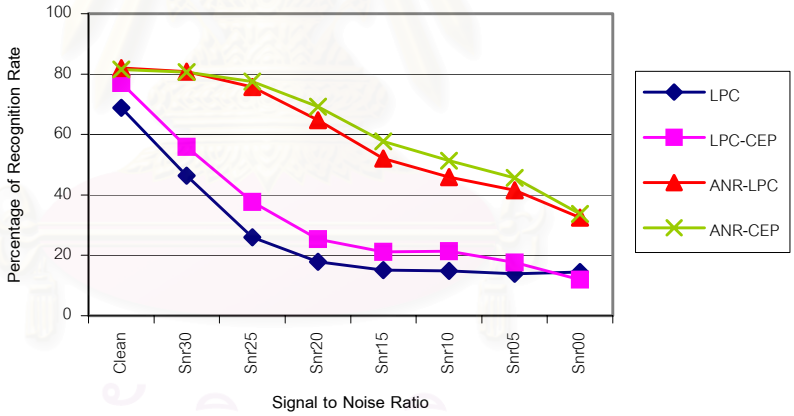


รูปที่ 4.12 อัตราการรู้จำเสียงวรรณยุกต์ไทยเสียงเอกด้วยลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC และ ANR-CEP อันดับ 12

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

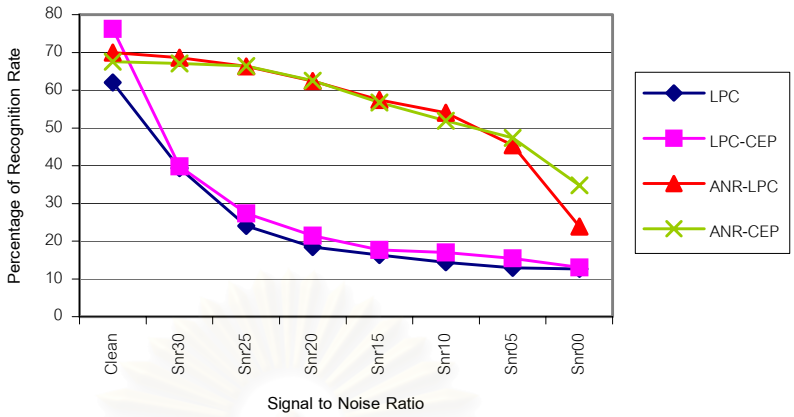


รูปที่ 4.13 อัตราการรู้จำเสียงวรรณยุกต์ไทยเสียงเอกด้วยลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC และ ANR-CEP อันดับ 14

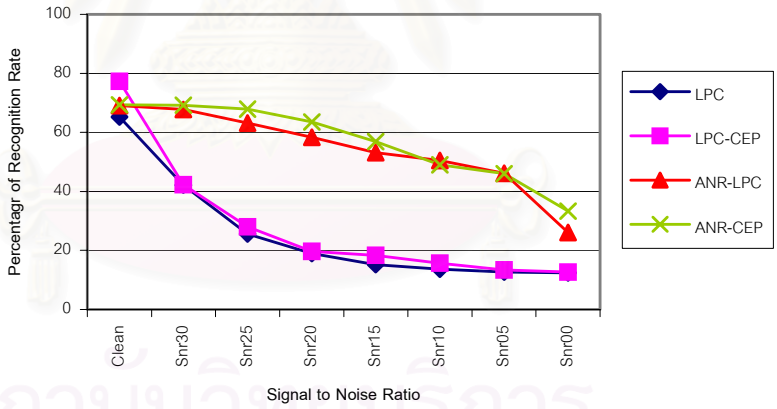


รูปที่ 4.14 อัตราการรู้จำเสียงวรรณยุกต์ไทยเสียงเอกด้วยลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC และ ANR-CEP อันดับ 16

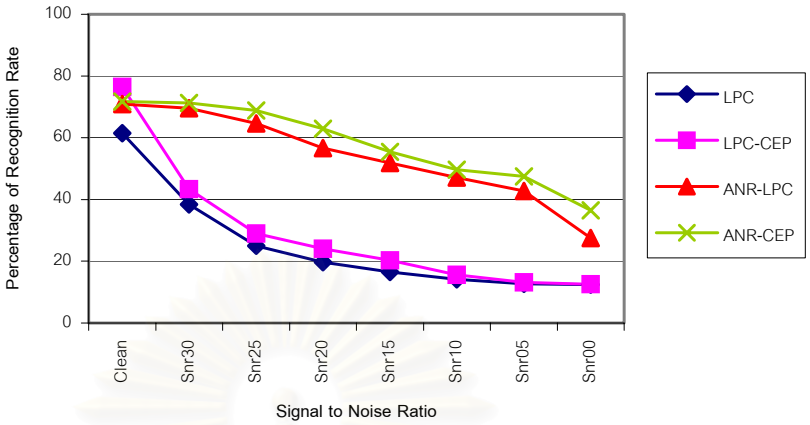
สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



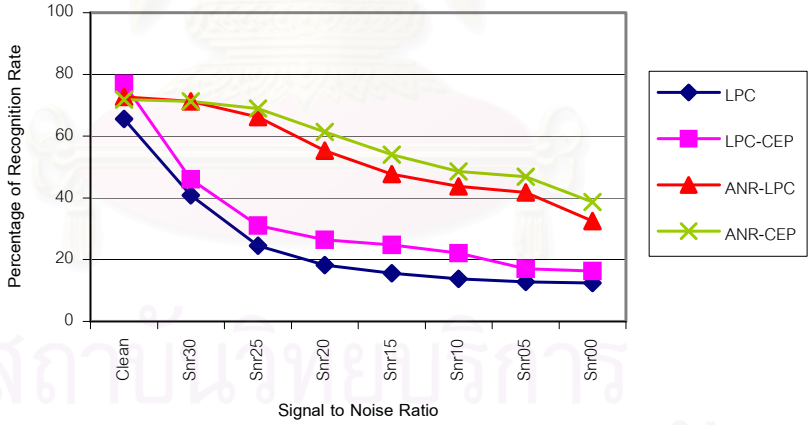
รูปที่ 4.15 อัตราการรู้จำเสียงวรรณยุกต์ไทยเสียงโทด้วยลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC และ ANR-CEP อันดับ 10



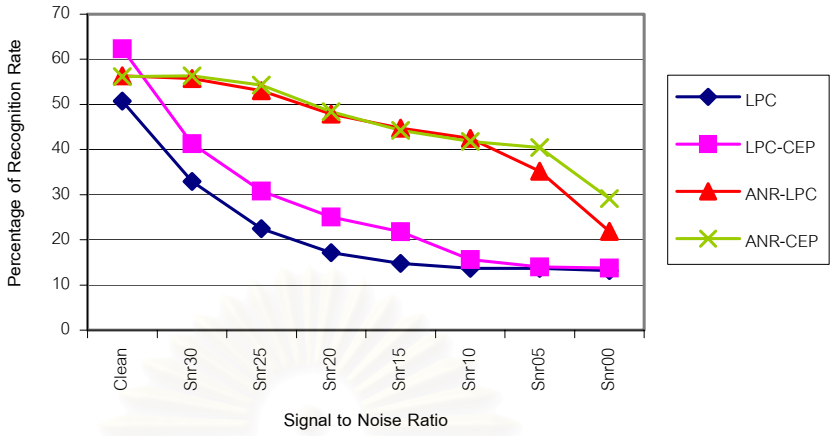
รูปที่ 4.16 อัตราการรู้จำเสียงวรรณยุกต์ไทยเสียงโทด้วยลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC และ ANR-CEP อันดับ 12



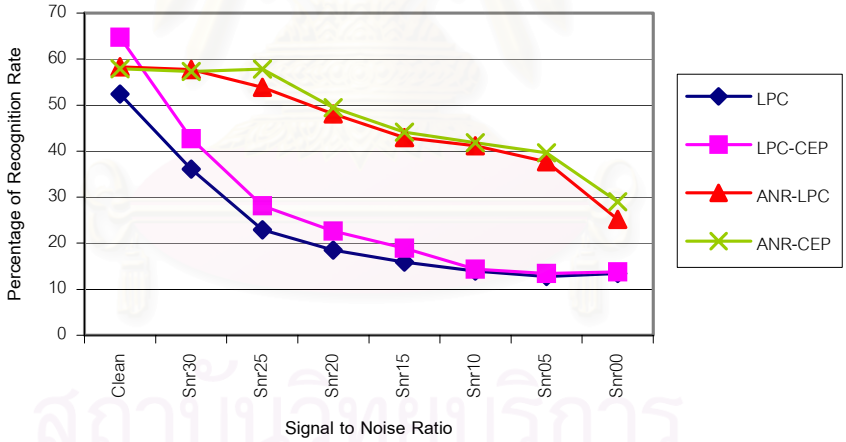
รูปที่ 4.17 อัตราการรู้จำเสียงวรรณยุกต์ไทยเสียงโทด้วยลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC และ ANR-CEP อันดับ 14



รูปที่ 4.18 อัตราการรู้จำเสียงวรรณยุกต์ไทยเสียงโทด้วยลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC และ ANR-CEP อันดับ 16

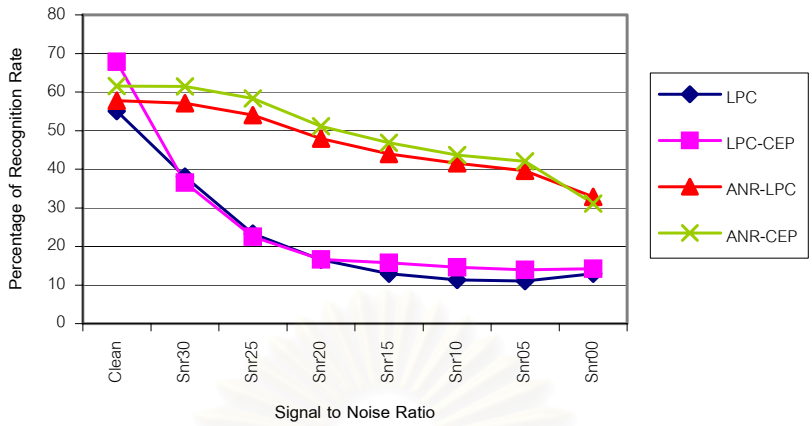


รูปที่ 4.19 อัตราการรู้จำเสียงวรรณยุกต์ไทยเสียงตรีด้วยลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC และ ANR-CEP อันดับ 10

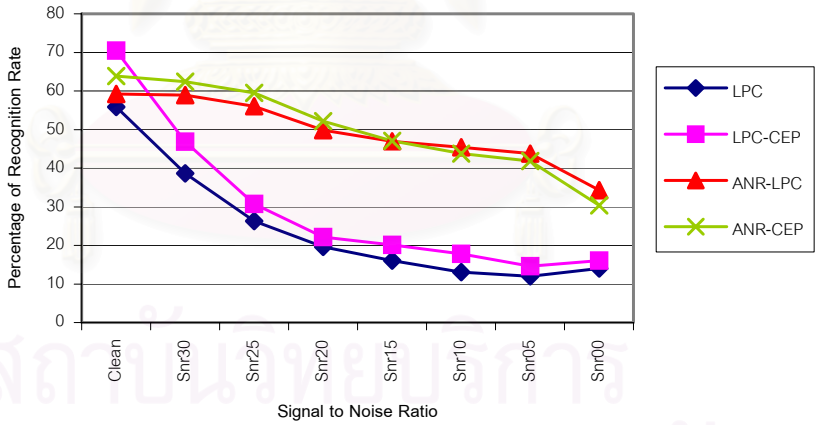


รูปที่ 4.20 อัตราการรู้จำเสียงวรรณยุกต์ไทยเสียงตรีด้วยลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC และ ANR-CEP อันดับ 12

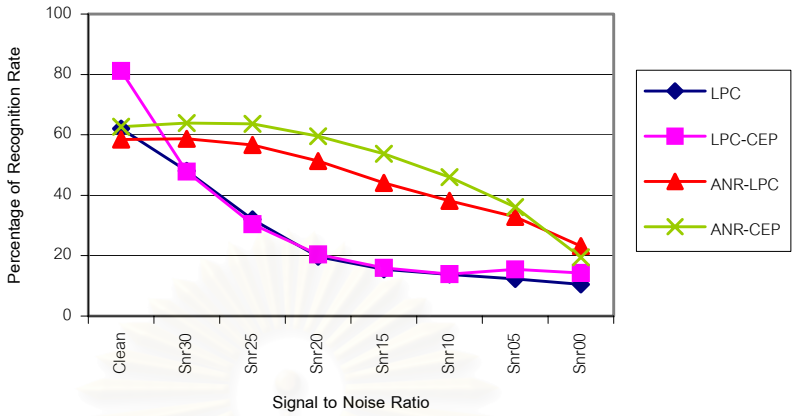
สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



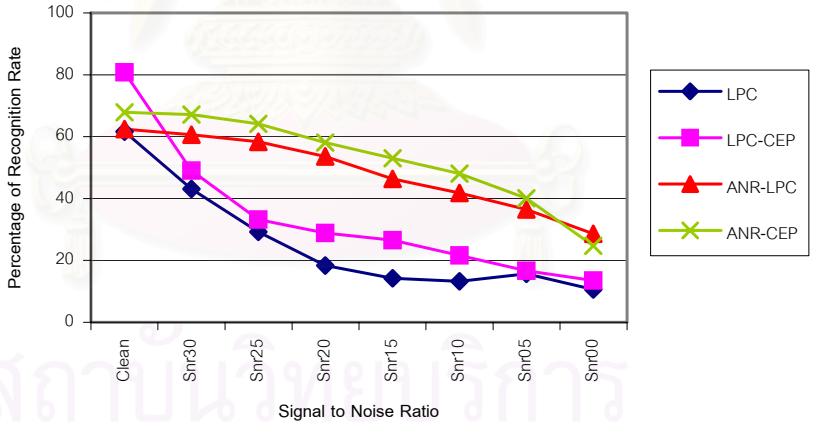
รูปที่ 4.21 อัตราการรู้จำเสียงวรรณยุกต์ไทยเสียงตรีด้วยลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC และ ANR-CEP อันดับ 14



รูปที่ 4.22 อัตราการรู้จำเสียงวรรณยุกต์ไทยเสียงตรีด้วยลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC และ ANR-CEP อันดับ 16

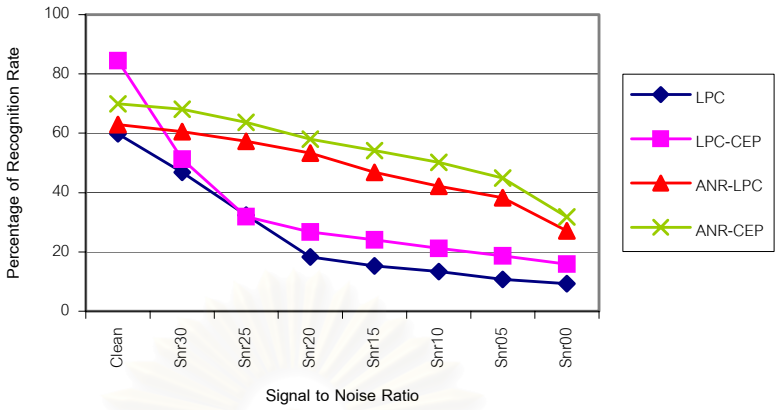


รูปที่ 4.23 อัตราการรู้จำเสียงวรรณยุกต์ไทยเสียงจัตวาด้วยลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC และ ANR-CEP อันดับ 10

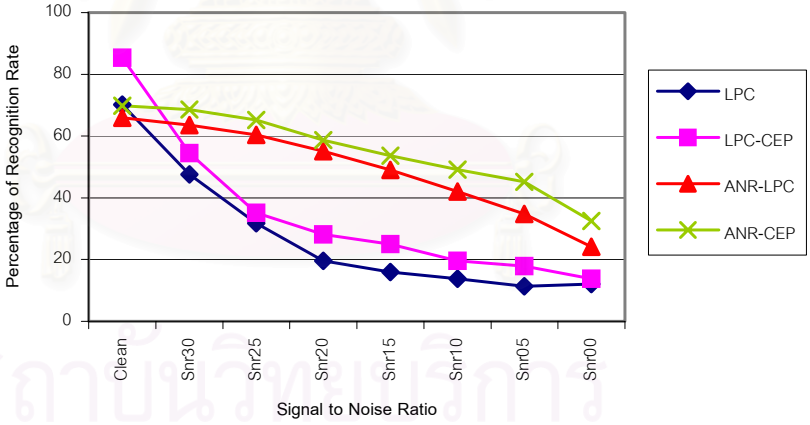


รูปที่ 4.24 อัตราการรู้จำเสียงวรรณยุกต์ไทยเสียงจัตวาด้วยลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC และ ANR-CEP อันดับ 12

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

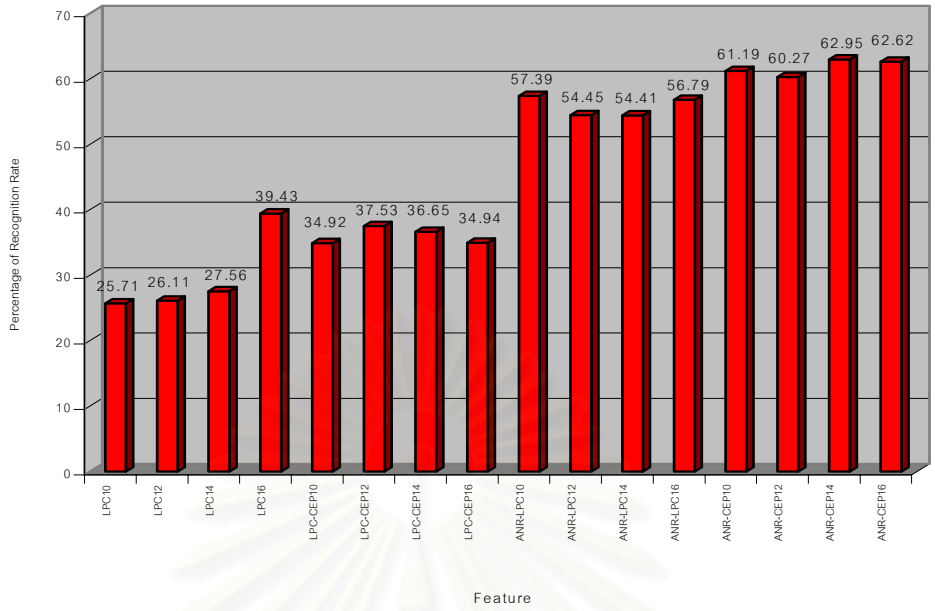


รูปที่ 4.25 อัตราการรู้จำเสียงวรรณยุกต์ไทยเสียงจัตวาด้วยลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC และ ANR-CEP อันดับ 14

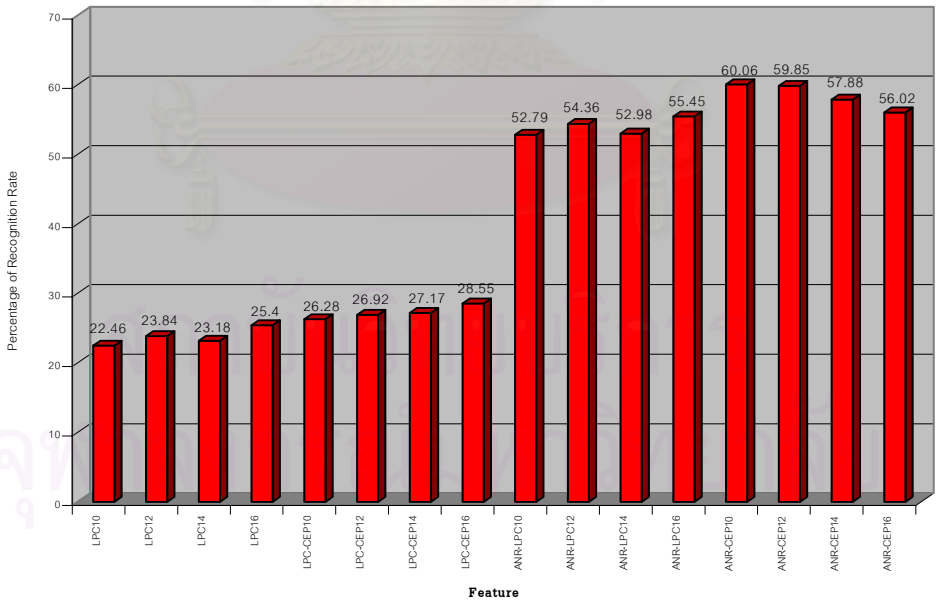


รูปที่ 4.26 อัตราการรู้จำเสียงวรรณยุกต์ไทยเสียงจัตวาด้วยลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC และ ANR-CEP อันดับ 16

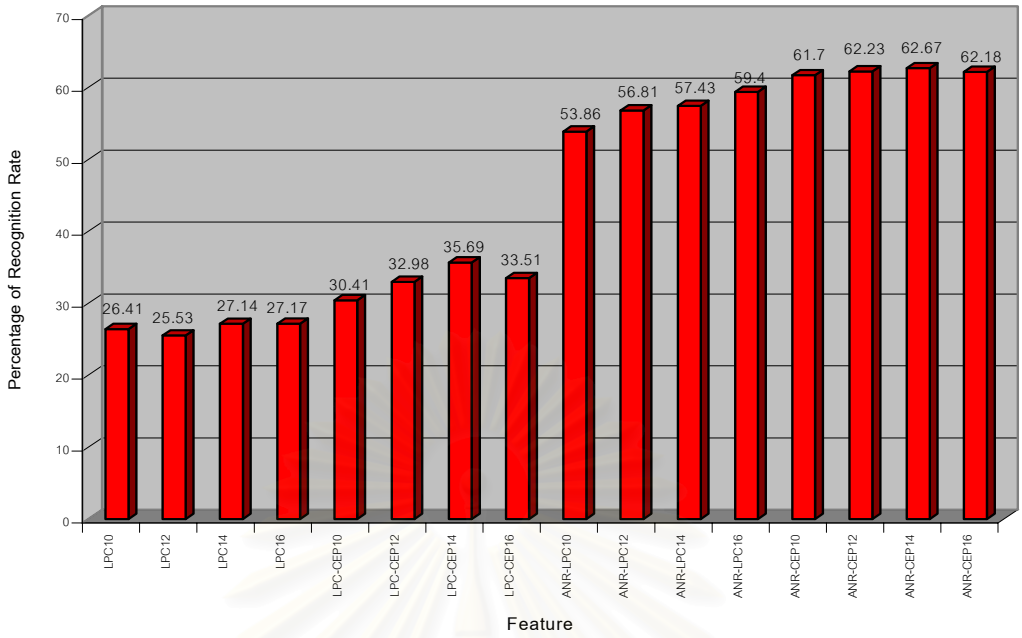
สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



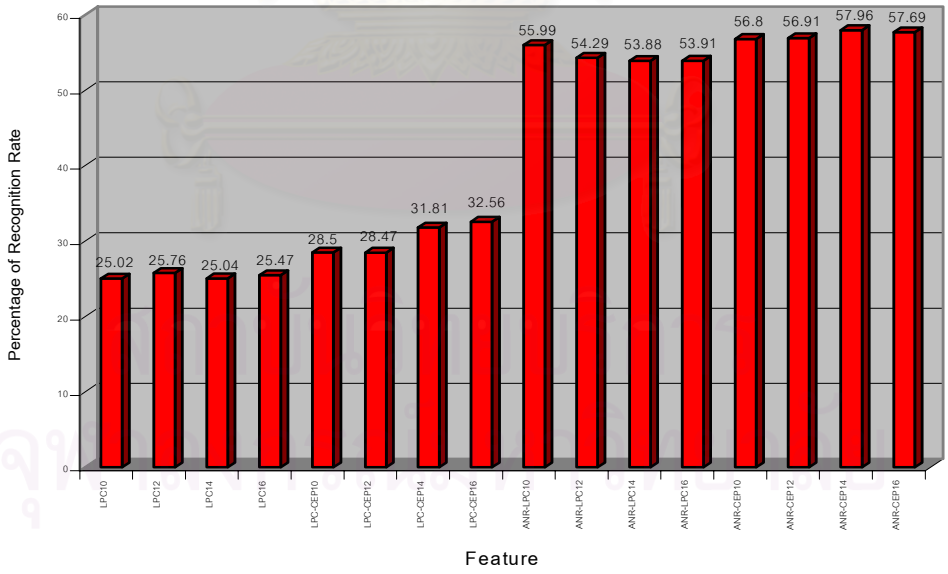
รูปที่ 4.27 เปรียบเทียบผลการรู้จำเสียงตัวเลขภาษาไทยโดยเฉลี่ยทุกค่า SNR สำหรับค่าลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC, ANR-CEP อันดับ 10 12 14 และ 16



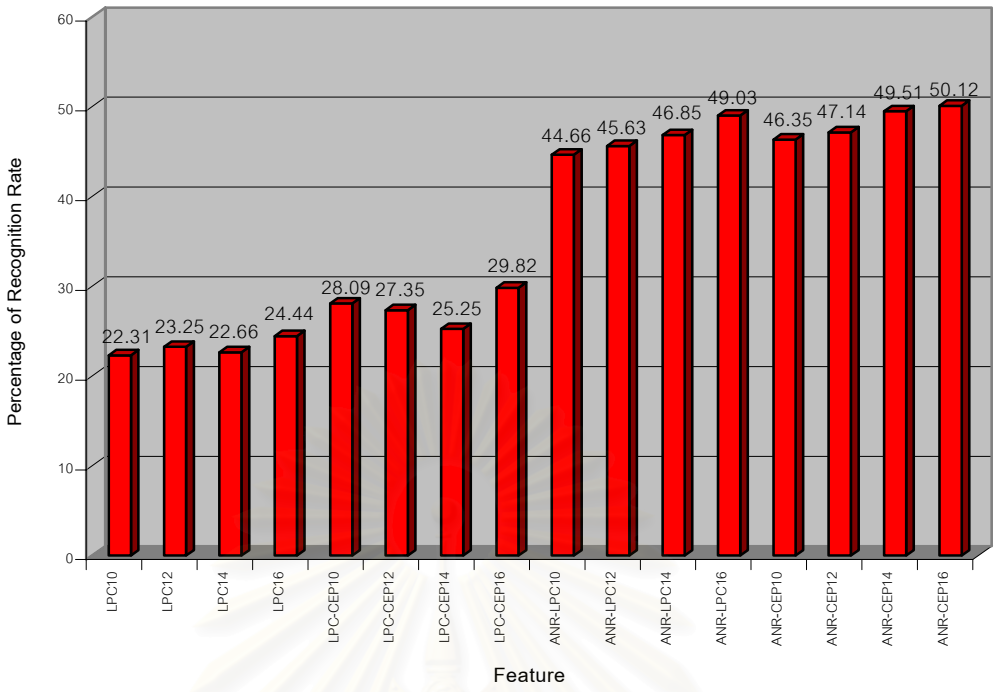
รูปที่ 4.28 เปรียบเทียบผลการรู้จำเสียงวรรณยุกต์ไทย เสียงสามัญโดยเฉลี่ยทุกค่า SNR สำหรับค่าลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC, ANR-CEP อันดับ 10 12 14 และ 16



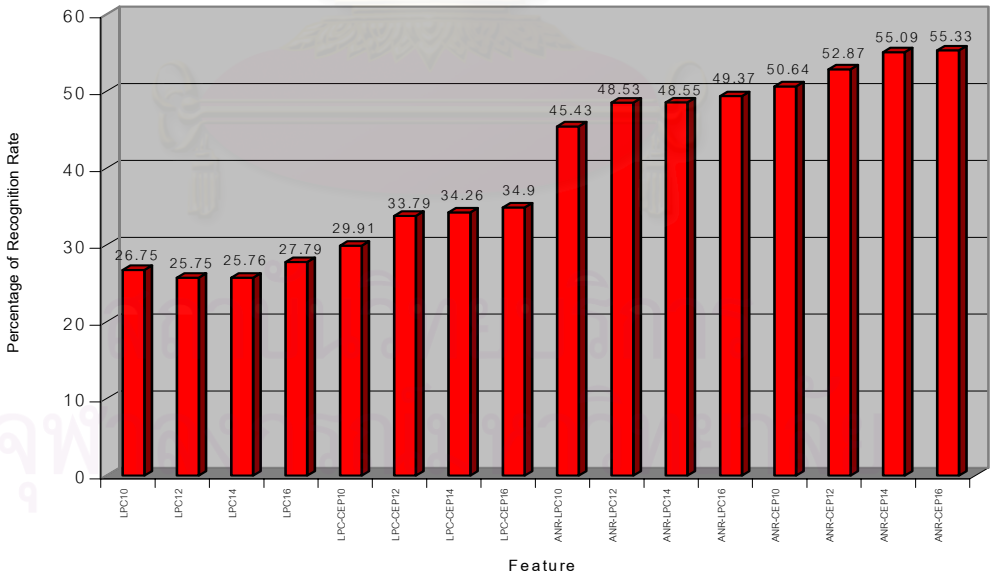
รูปที่ 4.29 เปรียบเทียบผลการรู้จำเสียงวรรณยุกต์ไทย เสียงเอกโดยเฉลี่ยทุกค่า SNR สำหรับค่าลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC, ANR-CEP อันดับ 10 12 14 และ 16



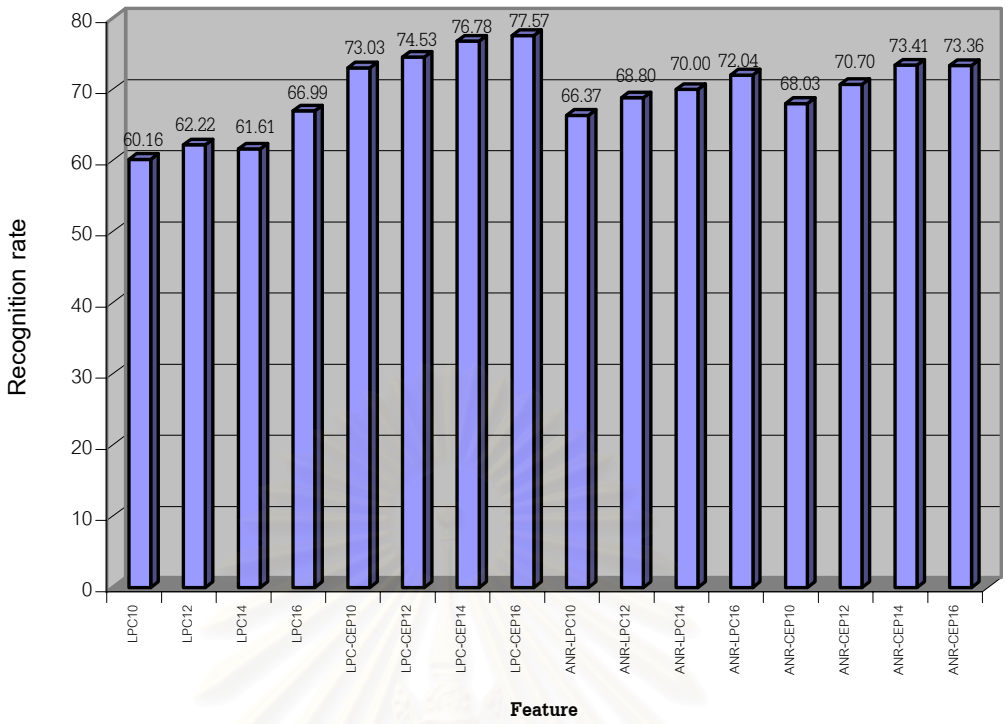
รูปที่ 4.30 เปรียบเทียบผลการรู้จำเสียงวรรณยุกต์ไทย เสียงโทโดยเฉลี่ยทุกค่า SNR สำหรับค่าลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC, ANR-CEP อันดับ 10 12 14 และ 16



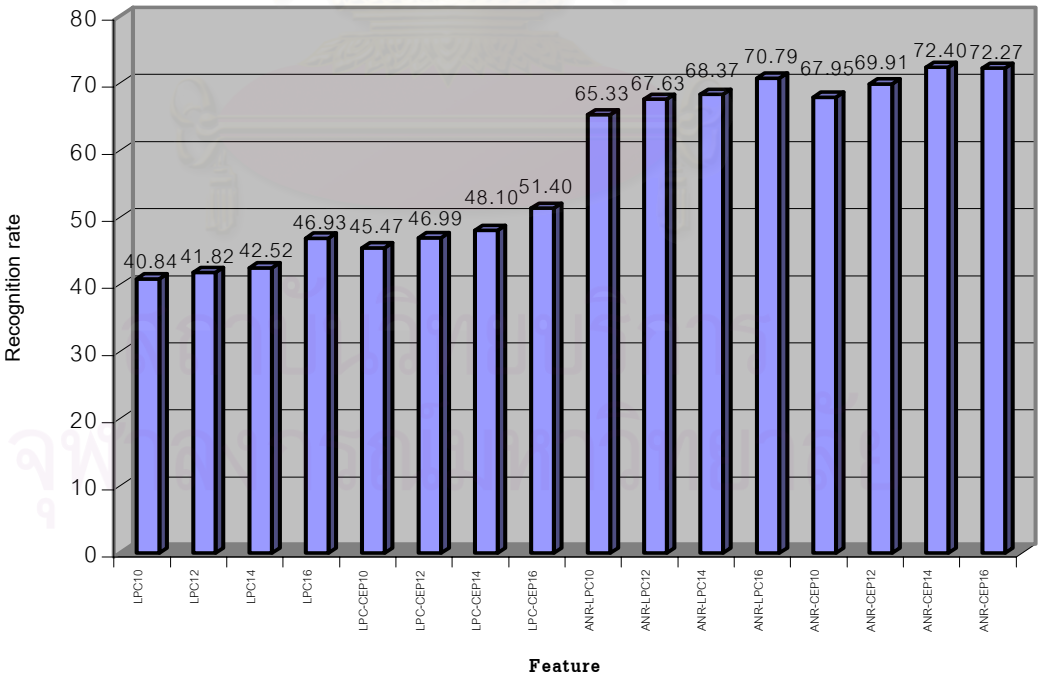
รูปที่ 4.31 เปรียบเทียบผลการรู้จำเสียงวรรณยุกต์ไทย เสียงตรีโดยเฉลี่ยทุกค่า SNR สำหรับค่าลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC, ANR-CEP อันดับ 10 12 14 และ 16



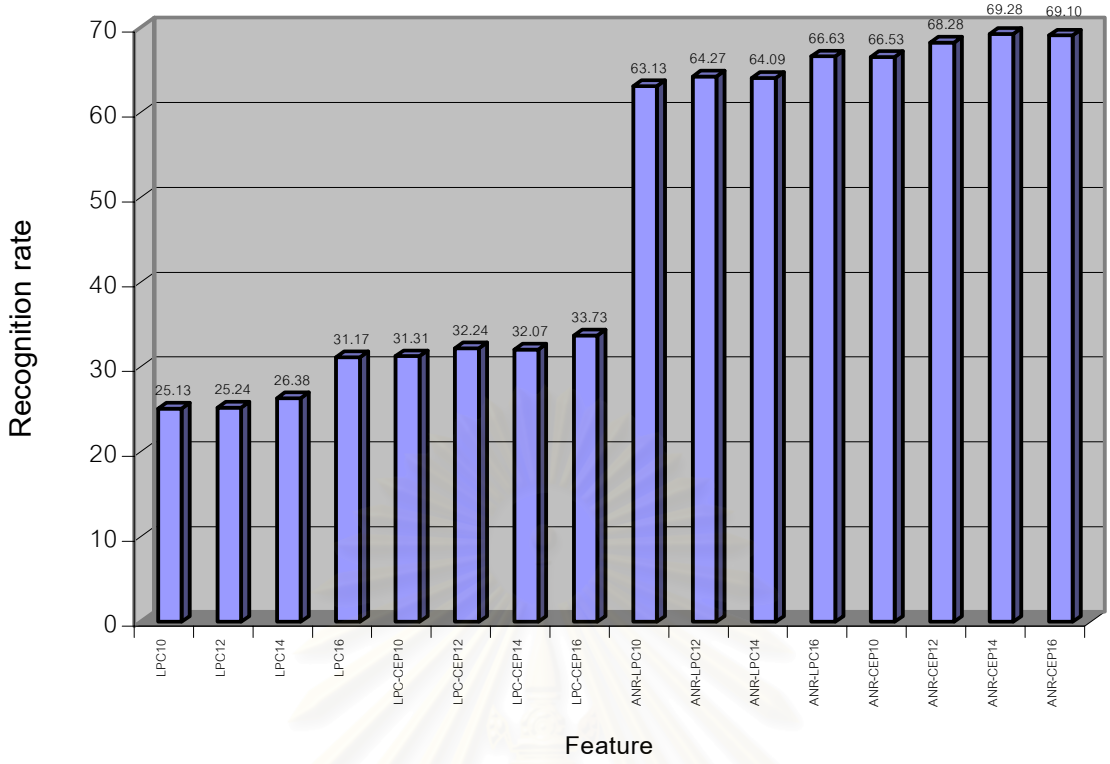
รูปที่ 4.32 เปรียบเทียบผลการรู้จำเสียงวรรณยุกต์ไทย เสียงจัตวาโดยเฉลี่ยทุกค่า SNR สำหรับค่าลักษณะสำคัญ LPC, LPC-CEP, ANR-LPC, ANR-CEP อันดับ 10 12 14 และ 16



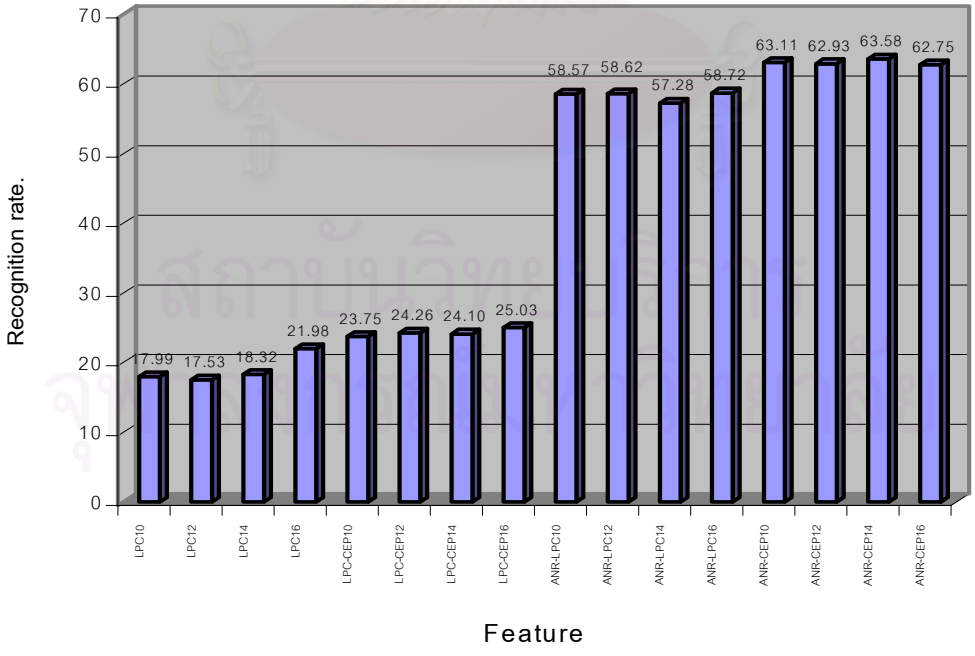
รูปที่ 4.33 แสดงอัตราการรู้จำเฉลี่ยรวมทุกชุดคำศัพท์สำหรับค่า Clean speech



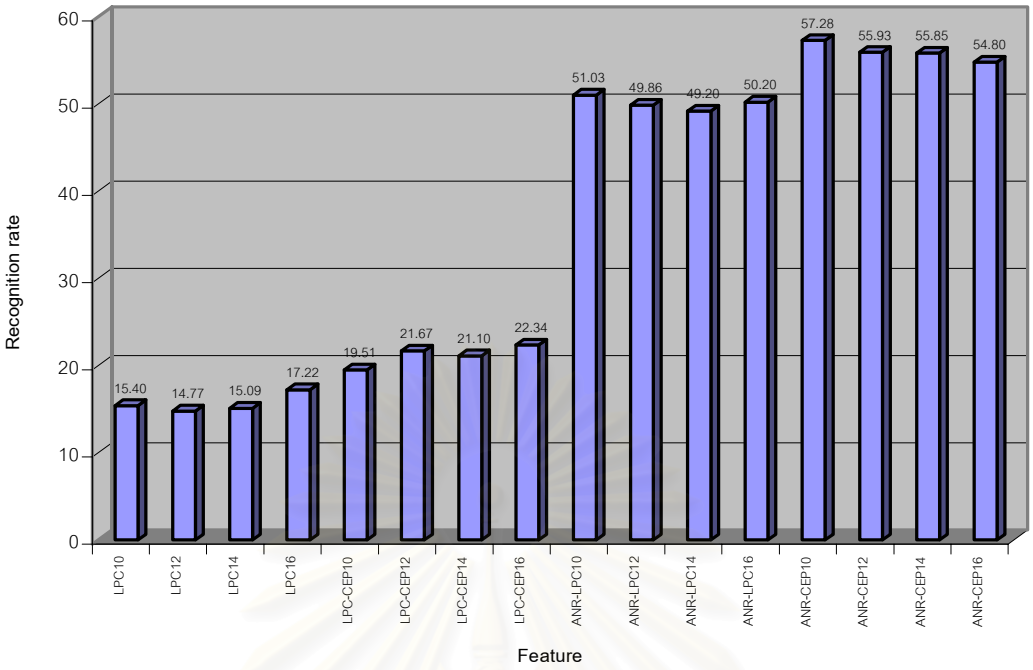
รูปที่ 4.34 แสดงอัตราการรู้จำเฉลี่ยรวมทุกชุดคำศัพท์สำหรับค่า SNR 30 dB



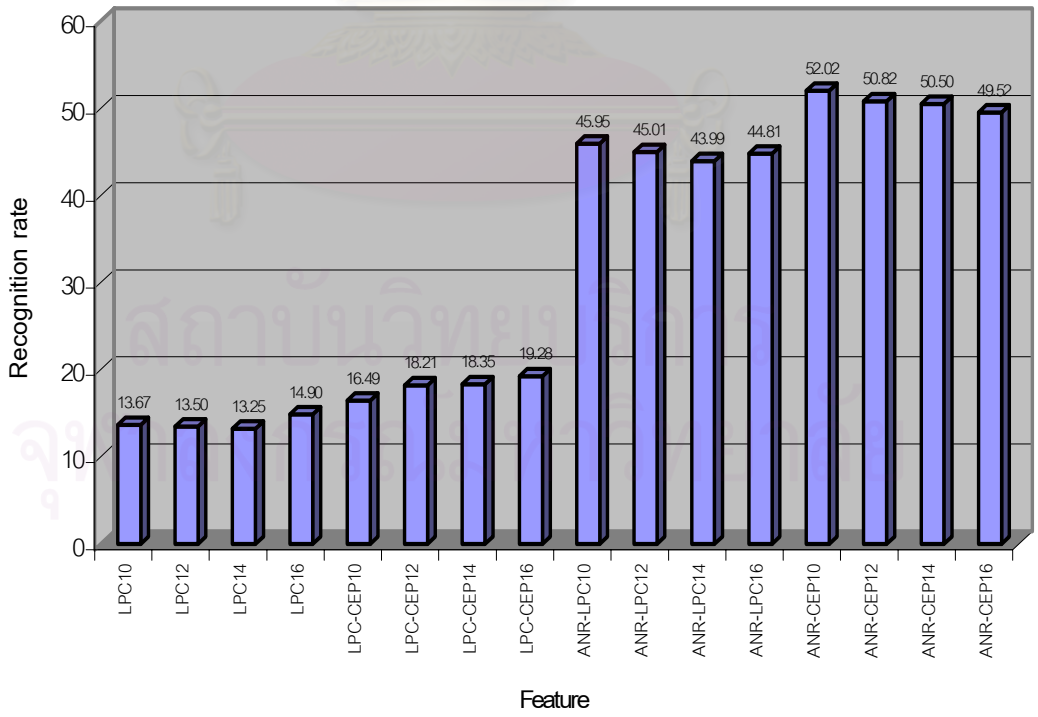
รูปที่ 4.35 แสดงอัตราการรู้จำเฉลี่ยรวมทุกชุดคำศัพท์สำหรับค่า SNR 25 dB



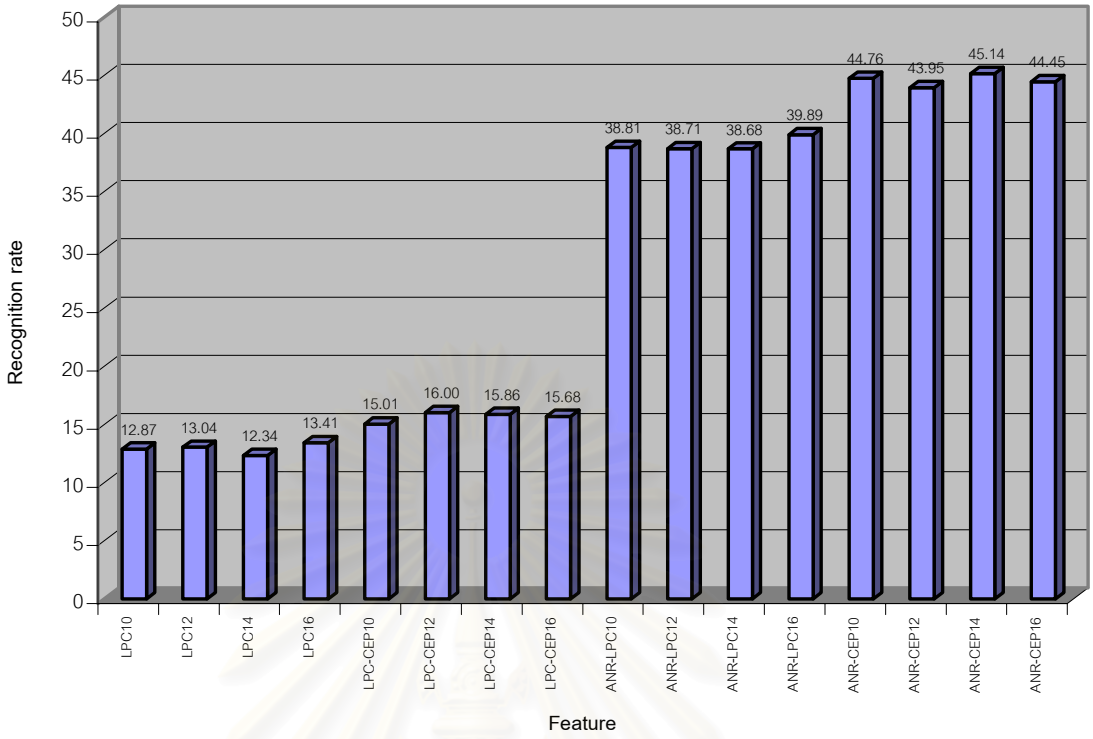
รูปที่ 4.36 แสดงอัตราการรู้จำเฉลี่ยรวมทุกชุดคำศัพท์สำหรับค่า SNR 20 dB



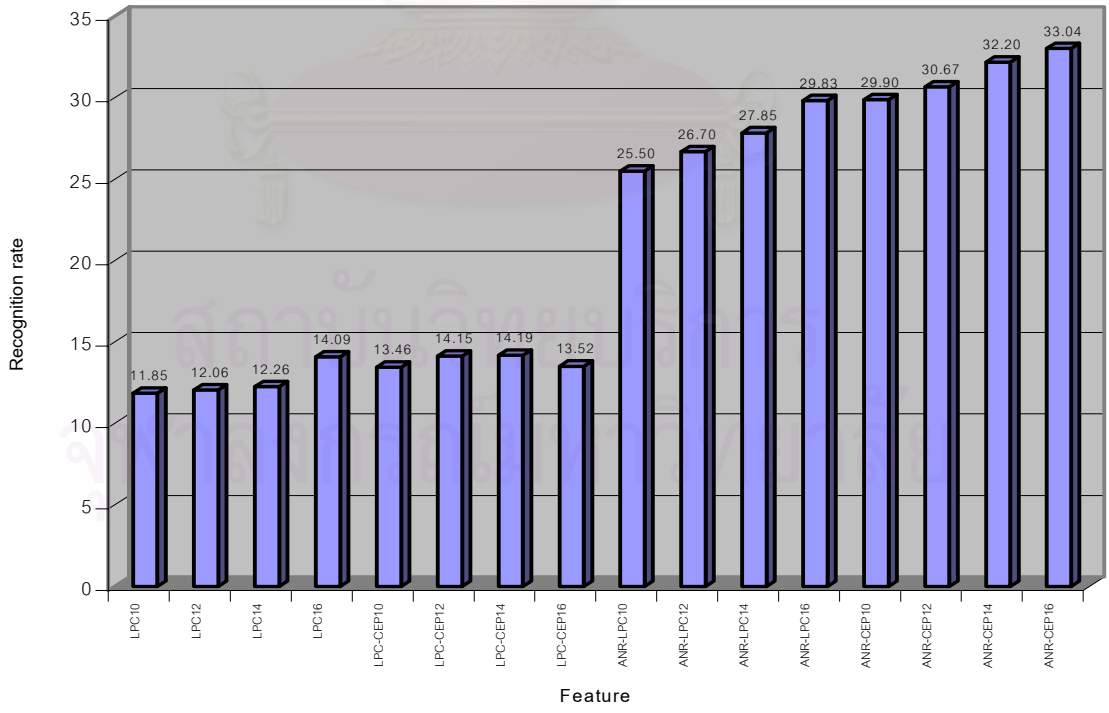
รูปที่ 4.37 แสดงอัตราการรู้จำเฉลี่ยรวมทุกชุดคำศัพท์สำหรับค่า SNR 15 dB



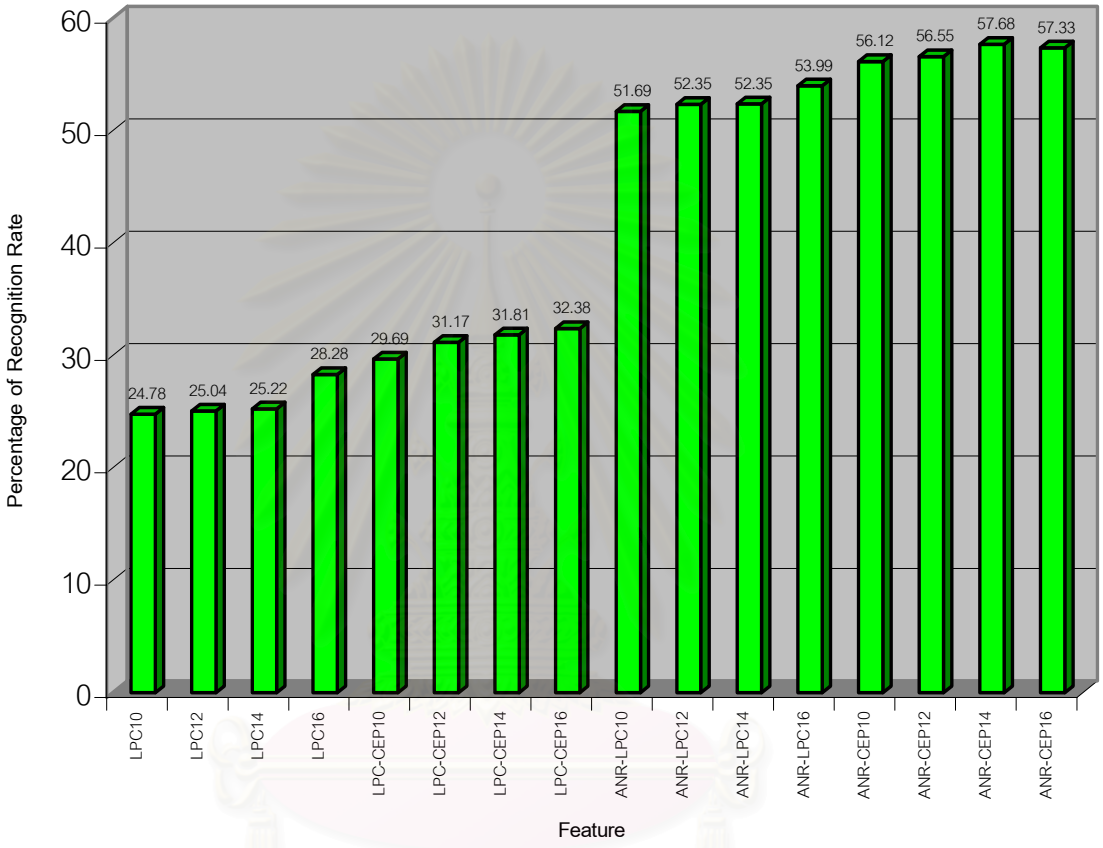
รูปที่ 4.38 แสดงอัตราการรู้จำเฉลี่ยรวมทุกชุดคำศัพท์สำหรับค่า SNR 10 dB



รูปที่ 4.39 แสดงอัตราการรู้จำเฉลี่ยรวมทุกชุดคำศัพท์สำหรับค่า SNR 5 dB



รูปที่ 4.40 แสดงอัตราการรู้จำเฉลี่ยรวมทุกชุดคำศัพท์สำหรับค่า SNR 0 dB



รูปที่ 4.41 แสดงอัตราการรู้จำเฉลี่ยรวมของระบบ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

4.5 วิเคราะห์ผลการวิจัย

1. ระบบที่เหมาะสมสำหรับการรู้จำเสียงพูดภาษาไทยที่มีสัญญาณรบกวนโดยใช้ค่าลักษณะสำคัญที่คงทนต่อสัญญาณรบกวน โดยการแปรค่าสถานะของแบบจำลองฮิดเดนมาร์คอฟ และจำนวน Gaussians Mixture เพื่อพิจารณาค่าอัตราการรู้จำเฉลี่ยสูงที่สุด ซึ่งได้ผลดังรูปที่ 4.1 และจากตารางที่ 4.1 จำนวนสถานะในแบบจำลองฮิดเดนมาร์คอฟเป็น 7 และจำนวน Gaussian Mixture เป็น 2 ให้อัตราการรู้จำเฉลี่ยรวมสูงที่สุด
2. จากรูปที่ 4.2 การทำเน้นล่วงหน้ามีผลดีสำหรับการรู้จำเสียงพูดในบริเวณที่ไม่มีสัญญาณรบกวน ส่วนที่ค่า SNR 30 dB ลงมาจนถึงประมาณ 15 dB การทำเน้นล่วงหน้าแทบไม่มีผลต่ออัตราการรู้จำเลย หรือมีก็น้อยมาก เพราะการทำเน้นล่วงหน้าจะทำให้ค่าอัตราการรู้จำสูงขึ้นสำหรับเสียง Clean และ SNR สูงๆ เท่านั้น แต่เมื่อค่า SNR ต่ำลง จะเห็นว่าค่าอัตราการรู้จำลดต่ำลงอย่างเห็นได้ชัดเจนตั้งแต่ค่า SNR ต่ำกว่า 10 dB ที่เป็นเช่นนี้เนื่องจากการทำเน้นล่วงหน้าเป็นการลดระดับความต่างของช่วงความถี่สูงและความถี่ต่ำให้ลดลง กล่าวคือเป็นการยกระดับความถี่สูงให้สูงขึ้น ซึ่งในกรณีที่ไม่มีสัญญาณรบกวนนั้น เป็นการช่วยให้คุณลักษณะสำคัญของสัญญาณเสียงพูดในช่วงความถี่สูงเด่นชัดขึ้น จึงทำให้อัตราการรู้จำสูงขึ้นในกรณีที่ไม่มีสัญญาณรบกวน หรือมีสัญญาณรบกวนน้อย แต่กรณีที่มีสัญญาณรบกวนมากแล้วนั้น เนื่องจากสัญญาณรบกวนเป็นสัญญาณรบกวนเกาส์เซียนขาวซึ่งมีค่าทุกความถี่ การทำเน้นล่วงหน้าเปรียบเสมือนการยกระดับขนาดสัญญาณรบกวนให้สูงขึ้น ผลกระทบของสัญญาณรบกวนจึงรุนแรงมากขึ้น ดังนั้นจึงยังผลให้อัตราการรู้จำลดลงเมื่อทำเน้นล่วงหน้ากับสัญญาณเสียงพูดที่มีสัญญาณรบกวนนั่นเอง
3. อัตราการรู้จำเสียงตัวเลขภาษาไทยเมื่อมีสัญญาณรบกวนด้วยค่าลักษณะสำคัญ ANR-LPC และ ANR-CEP นั้น สามารถเพิ่มอัตราการรู้จำของสัญญาณเสียงพูดที่มีสัญญาณรบกวนได้ทุกกรณี เว้นเสียแต่กรณีที่สัญญาณเสียงพูดที่ไม่มีสัญญาณรบกวนเท่านั้นที่ค่าลักษณะสำคัญ LPC-CEP ให้ค่าอัตราการรู้จำสูงกว่าค่าลักษณะสำคัญ ANR-CEP โดยปรับปรุงอัตราการรู้จำสูงสุดเป็นร้อยละ 38.00 ที่ SNR 0 dB เมื่ออัตราการรู้จำก่อนปรับลดสัญญาณรบกวนเป็นร้อยละ 10.42
4. อัตราการรู้จำเสียงวรรณยุกต์ไทยเสียงสามัญเมื่อมีสัญญาณรบกวนด้วยค่าลักษณะสำคัญ ANR-LPC และ ANR-CEP นั้น สามารถเพิ่มอัตราการรู้จำของสัญญาณเสียงพูดที่มีสัญญาณรบกวนได้ทุกกรณี เว้นเสียแต่กรณีที่สัญญาณเสียงพูดที่ไม่มีสัญญาณรบกวนเท่านั้นที่ค่าลักษณะสำคัญ LPC-CEP ให้ค่าอัตราการรู้จำพอๆค่าลักษณะสำคัญ ANR-CEP โดยปรับปรุงอัตราการรู้จำสูงสุดเป็นร้อยละ 31.65 ที่ SNR 0 dB เมื่ออัตราการรู้จำก่อนปรับลดสัญญาณรบกวนเป็นร้อยละ 12.27

5. อัตราการรู้จำเสียงวรรณยุกต์ไทยเสียงเอกเมื่อมีสัญญาณรบกวนด้วยค่าลักษณะสำคัญ ANR-LPC และ ANR-CEP นั้น สามารถเพิ่มอัตราการรู้จำของสัญญาณเสียงพูดที่มีสัญญาณรบกวนได้ทุกกรณี โดยปรับปรุงอัตราการรู้จำสูงสุดเป็นร้อยละ 33.64 ที่ SNR 0 dB เมื่ออัตราการรู้จำก่อนปรับลดสัญญาณรบกวนเป็นร้อยละ 11.93
6. อัตราการรู้จำเสียงวรรณยุกต์ไทยเสียงโทเมื่อมีสัญญาณรบกวนด้วยค่าลักษณะสำคัญ ANR-LPC และ ANR-CEP นั้น สามารถเพิ่มอัตราการรู้จำของสัญญาณเสียงพูดที่มีสัญญาณรบกวนได้ทุกกรณี เว้นเสียแต่กรณีที่สัญญาณเสียงพูดที่ไม่มีสัญญาณรบกวนเท่านั้นที่ค่าลักษณะสำคัญ LPC-CEP ให้ค่าอัตราการรู้จำสูงกว่าค่าลักษณะสำคัญ ANR-CEP โดยปรับปรุงอัตราการรู้จำสูงสุดเป็นร้อยละ 38.69 ที่ SNR 0 dB เมื่ออัตราการรู้จำก่อนปรับลดสัญญาณรบกวนเป็นร้อยละ 16.25
7. อัตราการรู้จำเสียงวรรณยุกต์ไทยเสียงตรีเมื่อมีสัญญาณรบกวนด้วยค่าลักษณะสำคัญ ANR-LPC และ ANR-CEP นั้น สามารถเพิ่มอัตราการรู้จำของสัญญาณเสียงพูดที่มีสัญญาณรบกวนได้ทุกกรณี เว้นเสียแต่กรณีที่สัญญาณเสียงพูดที่ไม่มีสัญญาณรบกวนเท่านั้นที่ค่าลักษณะสำคัญ LPC-CEP ให้ค่าอัตราการรู้จำสูงกว่าค่าลักษณะสำคัญ ANR-CEP โดยปรับปรุงอัตราการรู้จำสูงสุดเป็นร้อยละ 34.26 ที่ SNR 0 dB เมื่ออัตราการรู้จำก่อนปรับลดสัญญาณรบกวนเป็นร้อยละ 13.98
8. อัตราการรู้จำเสียงวรรณยุกต์ไทยเสียงจัตวาเมื่อมีสัญญาณรบกวนด้วยค่าลักษณะสำคัญ ANR-LPC และ ANR-CEP นั้น สามารถเพิ่มอัตราการรู้จำของสัญญาณเสียงพูดที่มีสัญญาณรบกวนได้ทุกกรณี เว้นเสียแต่กรณีที่สัญญาณเสียงพูดที่ไม่มีสัญญาณรบกวนเท่านั้นที่ค่าลักษณะสำคัญ LPC-CEP ให้ค่าอัตราการรู้จำพอๆค่าลักษณะสำคัญ ANR-CEP และ ค่าลักษณะสำคัญ LPC ให้ค่าอัตราการรู้จำสูงกว่าหรือสูงกว่าค่าลักษณะสำคัญ ANR-LPC ขึ้นอยู่กับอันดับของค่าลักษณะสำคัญ โดยปรับปรุงอัตราการรู้จำสูงสุดเป็นร้อยละ 32.44 ที่ SNR 0 dB เมื่ออัตราการรู้จำก่อนปรับลดสัญญาณรบกวนเป็นร้อยละ 13.75
9. อัตราการรู้จำเสียงตัวเลขภาษาไทยเฉลี่ยสูงสุด โดยคำนวณจากการนำอัตราการรู้จำของแต่ละค่าลักษณะสำคัญในแต่ละอันดับสำหรับทุกค่า SNR มาทำการเฉลี่ยเมื่อใช้ค่าลักษณะสำคัญแบบคงทนต่อสัญญาณรบกวน ANR-CEP อันดับ 14 เป็นร้อยละ 62.95 ขณะที่อัตราการรู้จำเฉลี่ยสูงสุดก่อนปรับลดสัญญาณรบกวนเป็นร้อยละ 36.65 โดยมีค่าลักษณะสำคัญเป็น LPC อันดับ 16
10. อัตราการรู้จำเสียงวรรณยุกต์ไทยเสียงสามัญเฉลี่ยสูงสุด โดยคำนวณจากการนำอัตราการรู้จำของแต่ละค่าลักษณะสำคัญในแต่ละอันดับสำหรับทุกค่า SNR มาทำการเฉลี่ยเมื่อใช้ค่าลักษณะสำคัญแบบคงทนต่อสัญญาณรบกวน ANR-CEP อันดับ 12 เป็นร้อยละ 59.85 ขณะที่อัตราการรู้จำเฉลี่ยสูงสุดก่อนปรับลดสัญญาณรบกวนเป็นร้อยละ 28.55 โดยมีค่าลักษณะสำคัญเป็น LPC-CEP อันดับ 16

11. อัตราการรู้จำเสียงวรรณยุกต์ไทยเสียงเอกเฉลี่ยสูงสุด โดยคำนวณจากการนำอัตราการรู้จำของแต่ละค่าลักษณะสำคัญในแต่ละอันดับสำหรับทุกค่า SNR มาทำการเฉลี่ยเมื่อใช้ค่าลักษณะสำคัญแบบคงทนต่อสัญญาณรบกวน ANR-CEP อันดับ 14 เป็นร้อยละ 62.67 ขณะที่อัตราการรู้จำเฉลี่ยสูงสุดก่อนปรับลดสัญญาณรบกวนเป็นร้อยละ 35.69 โดยมีค่าลักษณะสำคัญเป็น LPC-CEP อันดับ 14
12. อัตราการรู้จำเสียงวรรณยุกต์ไทยเสียงโทเฉลี่ยสูงสุด โดยคำนวณจากการนำอัตราการรู้จำของแต่ละค่าลักษณะสำคัญในแต่ละอันดับสำหรับทุกค่า SNR มาทำการเฉลี่ยเมื่อใช้ค่าลักษณะสำคัญแบบคงทนต่อสัญญาณรบกวน ANR-CEP อันดับ 14 เป็นร้อยละ 57.96 ขณะที่อัตราการรู้จำเฉลี่ยสูงสุดก่อนปรับลดสัญญาณรบกวนเป็นร้อยละ 32.56 โดยมีค่าลักษณะสำคัญเป็น LPC-CEP อันดับ 16
13. อัตราการรู้จำเสียงวรรณยุกต์ไทยเสียงตรีเฉลี่ยสูงสุด โดยคำนวณจากการนำอัตราการรู้จำของแต่ละค่าลักษณะสำคัญในแต่ละอันดับสำหรับทุกค่า SNR มาทำการเฉลี่ยเมื่อใช้ค่าลักษณะสำคัญแบบคงทนต่อสัญญาณรบกวน ANR-CEP อันดับ 16 เป็นร้อยละ 50.12 ขณะที่อัตราการรู้จำเฉลี่ยสูงสุดก่อนปรับลดสัญญาณรบกวนเป็นร้อยละ 29.82 โดยมีค่าลักษณะสำคัญเป็น LPC-CEP อันดับ 16
14. อัตราการรู้จำเสียงวรรณยุกต์ไทยเสียงจัตวาเฉลี่ยสูงสุด โดยคำนวณจากการนำอัตราการรู้จำของแต่ละค่าลักษณะสำคัญในแต่ละอันดับสำหรับทุกค่า SNR มาทำการเฉลี่ยเมื่อใช้ค่าลักษณะสำคัญแบบคงทนต่อสัญญาณรบกวน ANR-CEP อันดับ 16 เป็นร้อยละ 55.33 ขณะที่อัตราการรู้จำเฉลี่ยสูงสุดก่อนปรับลดสัญญาณรบกวนเป็นร้อยละ 34.26 โดยมีค่าลักษณะสำคัญเป็น LPC-CEP อันดับ 14
15. อัตราการรู้จำเฉลี่ยรวมของทุกชุดคำศัพท์สูงสุด ในกรณีไม่มีสัญญาณรบกวน (Clean speech) โดยคำนวณจากอัตราการรู้จำของทุกชุดคำศัพท์ของแต่ละอันดับของค่าลักษณะสำคัญมาทำการเฉลี่ย มีอัตราการรู้จำเป็นร้อยละ 73.41 ขณะที่อัตราการรู้จำเฉลี่ยรวมของทุกชุดคำศัพท์สูงสุดก่อนปรับลดสัญญาณรบกวนเป็นร้อยละ 77.67 เนื่องจากการใช้งานค่าลักษณะสำคัญแบบคงทนต่อสัญญาณรบกวนในกรณีที่ไม่มีสัญญาณรบกวนนั้น เปรียบเสมือนหนึ่งว่าต้องการลดสัญญาณรบกวนลงไปโดยความเป็นจริงนั้นไม่มีสัญญาณรบกวน ดังนั้นการลดสัญญาณรบกวนในกรณีนี้จึงทำให้เกิดการสูญเสียคุณลักษณะที่สำคัญบางส่วนของสัญญาณเสียงพูดไป จึงทำให้อัตราการรู้จำในกรณีนี้ลดลงเมื่อใช้ค่าลักษณะสำคัญแบบคงทนต่อสัญญาณรบกวน โดยเปรียบเทียบกับค่าลักษณะสำคัญที่ไม่ผ่านการปรับลดสัญญาณรบกวน
16. อัตราการรู้จำเฉลี่ยรวมของทุกชุดคำศัพท์สูงสุดในกรณีที่มีสัญญาณรบกวน โดยคำนวณจากอัตราการรู้จำของทุกชุดคำศัพท์ของแต่ละอันดับของค่าลักษณะสำคัญมาทำการเฉลี่ย ดังนี้
    - กรณี SNR 30 dB มีอัตราการรู้จำเฉลี่ยรวมของทุกชุดคำศัพท์สูงสุดเป็นร้อยละ 72.40 เมื่อค่าลักษณะสำคัญเป็น ANR-CEP 14 ขณะที่อัตราการรู้จำเฉลี่ยรวมของทุก

ชุดคำศัพท์สูงสุดก่อนปรับลดสัญญาณรบกวนเป็นร้อยละ 51.40 เมื่อค่าลักษณะสำคัญเป็น LPC-CEP 16

- กรณี SNR 25 dB มีอัตราการรู้จำเฉลี่ยรวมของทุกชุดคำศัพท์สูงสุดเป็นร้อยละ 69.28 เมื่อค่าลักษณะสำคัญเป็น ANR-CEP 14 ขณะที่อัตราการรู้จำเฉลี่ยรวมของทุกชุดคำศัพท์สูงสุดก่อนปรับลดสัญญาณรบกวนเป็นร้อยละ 33.73 เมื่อค่าลักษณะสำคัญเป็น LPC-CEP 16
- กรณี SNR 20 dB มีอัตราการรู้จำเฉลี่ยรวมของทุกชุดคำศัพท์สูงสุดเป็นร้อยละ 63.58 เมื่อค่าลักษณะสำคัญเป็น ANR-CEP 14 ขณะที่อัตราการรู้จำเฉลี่ยรวมของทุกชุดคำศัพท์สูงสุดก่อนปรับลดสัญญาณรบกวนเป็นร้อยละ 25.03 เมื่อค่าลักษณะสำคัญเป็น LPC-CEP 16
- กรณี SNR 15 dB มีอัตราการรู้จำเฉลี่ยรวมของทุกชุดคำศัพท์สูงสุดเป็นร้อยละ 57.28 เมื่อค่าลักษณะสำคัญเป็น ANR-CEP 12 ขณะที่อัตราการรู้จำเฉลี่ยรวมของทุกชุดคำศัพท์สูงสุดก่อนปรับลดสัญญาณรบกวนเป็นร้อยละ 22.34 เมื่อค่าลักษณะสำคัญเป็น LPC-CEP 16
- กรณี SNR 10 dB มีอัตราการรู้จำเฉลี่ยรวมของทุกชุดคำศัพท์สูงสุดเป็นร้อยละ 52.02 เมื่อค่าลักษณะสำคัญเป็น ANR-CEP 12 ขณะที่อัตราการรู้จำเฉลี่ยรวมของทุกชุดคำศัพท์สูงสุดก่อนปรับลดสัญญาณรบกวนเป็นร้อยละ 19.28 เมื่อค่าลักษณะสำคัญเป็น LPC-CEP 16
- กรณี SNR 5 dB มีอัตราการรู้จำเฉลี่ยรวมของทุกชุดคำศัพท์สูงสุดเป็นร้อยละ 45.14 เมื่อค่าลักษณะสำคัญเป็น ANR-CEP 14 ขณะที่อัตราการรู้จำเฉลี่ยรวมของทุกชุดคำศัพท์สูงสุดก่อนปรับลดสัญญาณรบกวนเป็นร้อยละ 16.00 เมื่อค่าลักษณะสำคัญเป็น LPC-CEP 12
- กรณี SNR 0 dB มีอัตราการรู้จำเฉลี่ยรวมของทุกชุดคำศัพท์สูงสุดเป็นร้อยละ 33.04 เมื่อค่าลักษณะสำคัญเป็น ANR-CEP 16 ขณะที่อัตราการรู้จำเฉลี่ยรวมของทุกชุดคำศัพท์สูงสุดก่อนปรับลดสัญญาณรบกวนเป็นร้อยละ 14.19 เมื่อค่าลักษณะสำคัญเป็น LPC-CEP 14

17. อัตราการรู้จำเฉลี่ยรวมสูงสุดของทั้งระบบเมื่อใช้ค่าลักษณะสำคัญแบบคงทนต่อสัญญาณรบกวนเป็นร้อยละ 57.68 เมื่อค่าลักษณะสำคัญเป็น ANR-CEP 14 ขณะที่อัตราการรู้จำเฉลี่ยรวมสูงสุดของระบบก่อนปรับลดสัญญาณรบกวนเป็นร้อยละ 32.28 ซึ่งเพิ่มขึ้นร้อยละ 78.69 เมื่อค่าลักษณะสำคัญเป็น LPC-CEP 16
18. ค่าลักษณะสำคัญที่มีอันดับสูงกว่าโดยส่วนใหญ่แทบทุกกรณีให้อัตราการรู้จำสูงกว่าค่าลักษณะสำคัญที่มีอันดับต่ำกว่า

19. จากผลการวิจัยทั้งหมดเป็นที่ประจักษ์ว่า ค่าลักษณะสำคัญแบบคงทนต่อสัญญาณรบกวนสามารถเพิ่มอัตราการรู้จำเมื่อนำไปใช้ในบริเวณที่มีสัญญาณรบกวนได้ดี ซึ่งจากการวิเคราะห์อัตราการรู้จำโดยเฉลี่ยในข้อที่ 9 ถึงข้อที่ 18 แสดงถึงประสิทธิภาพที่เพิ่มขึ้นของค่าลักษณะสำคัญดังกล่าวนี้เมื่อนำไปใช้งานในบริเวณที่มีสัญญาณรบกวนต่างๆ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 5

บทสรุปและข้อเสนอแนะ

5.1 สรุปผลการวิจัย

งายวิจัยนี้เป็นการพัฒนาระบบรู้จำเสียงพูดภาษาไทยเพื่อให้ค่าลักษณะสำคัญที่แทนข้อมูลเสียงพูดมีความคงทนต่อสัญญาณรบกวน (Noise Robust Feature) โดยพิจารณาให้สัญญาณรบกวนเป็นสัญญาณรบกวนเกาส์เซียนขาวค่าเฉลี่ยศูนย์แบบบวก (Additive zero mean White Gaussian Noise) ซึ่งค่าลักษณะสำคัญนี้จะตั้งอยู่บนพื้นฐานของการประมาณพันธะเชิงเส้น โดยใช้เทคนิคลดสัญญาณรบกวนแบบปรับตัวได้โดยใช้ค่าเจาะจงต่ำสุดของเมตริกซ์อัตสหสัมพันธ์ของสัญญาณเสียงพูดเป็นค่าพารามิเตอร์ในการปรับลดสัญญาณรบกวน ซึ่งขึ้นอยู่กับตัวสัญญาณรบกวนเอง

ชุดคำศัพท์ในการทดสอบคุณลักษณะสำคัญของสัญญาณเสียงพูดที่เสนอขึ้นนั้น ประกอบด้วย ชุดคำศัพท์ตัวเลขศูนย์ถึงเก้าจำนวน 10 คำ และชุดคำศัพท์แยกตามวรรณยุกต์ไทยโดยจัดแบ่งเป็น 5 กลุ่ม ตามกลุ่มเสียงวรรณยุกต์ทั้ง 5 คือ สามัญ เอก โท ตรี และจัตวา ในแต่ละกลุ่มเสียงจะมีชุดคำศัพท์พยางค์เดียว 8 คำ รวมชุดคำศัพท์ทั้งหมดเป็น 40 คำ

หลักการตัดพยางค์จะใช้วิธีการพิจารณาจากค่าระดับพลังงานของเสียงพูดที่แบ่งเป็นเฟรมๆละ 100 ตัวอย่าง โดยจะกำหนดจุดเปลี่ยนระดับพลังงาน (Energy Level Threshold) ที่ระดับพลังงาน ร้อยละ 20 ของค่าเฉลี่ยของระดับพลังงานเพื่อให้เหมาะสมกับระดับพลังงานของแต่ละเสียงพูด และกำหนดระยะเวลาขั้นต่ำของการเปลี่ยนระดับพลังงานที่ 200 เฟรม ซึ่งพบว่ามีจำนวนจุดเริ่มต้นและจุดสิ้นสุดของพยางค์เป็นจำนวนมากกว่าที่ควรจะเป็น (มากกว่า 2 พยางค์) เป็นจำนวนมาก จึงต้องหาจุดเริ่มต้นและสิ้นสุดของพยางค์ที่แท้จริงโดยหาค่าเฉลี่ยของระดับพลังงานที่อยู่ระหว่างสองพยางค์ใดๆในขั้นตอนก่อนหน้านี้ แล้วหาว่าช่วงใดมีค่าเฉลี่ยระดับพลังงานต่ำสุด ช่วงดังกล่าวควรจะเป็นช่วงรอยต่อระหว่างสองพยางค์ เมื่อหาช่วงดังกล่าวพบแล้ว จะทำการแก้ไขให้จุดสิ้นสุดของพยางค์แรกและจุดเริ่มต้นพยางค์ถัดไปเป็นจุดเดียวกัน โดยใช้เฟรมที่แบ่งครึ่งค่าเฉลี่ยของระดับพลังงานที่อยู่ในช่วงรอยต่อระหว่างสองพยางค์ให้เป็นจุดแบ่งพยางค์

การสกัดค่าลักษณะสำคัญจะใช้ค่าลักษณะสำคัญได้แก่ สัมประสิทธิ์การประมาณพันธะเชิงเส้นจากเทคนิคลดสัญญาณรบกวนแบบปรับตัวได้ และสัมประสิทธิ์เซปสตรอลจากเทคนิคลดสัญญาณรบกวนแบบปรับตัวได้ ที่นำเสนอในงานวิจัยโดยใช้ค่าเจาะจงต่ำสุดของเมตริกซ์อัตสหสัมพันธ์ของสัญญาณเสียงพูดเป็นค่าพารามิเตอร์ในการปรับลดสัญญาณรบกวนซึ่งคำนวณหาด้วยวิธีกำลังสมมาตรเพื่อใช้เป็นเงื่อนไขเริ่มต้นในการหาค่าการลดทอนสัญญาณรบกวนที่เหมาะสมที่ยังคงทำให้ระบบวงจรกรองที่มีแต่โพลมีเสถียรภาพ โดยใช้วิธีการแบ่งส่วนในกระบวนการหาค่าคำตอบของสมการ ซึ่งได้นำเสนอในงานวิจัย และสัมประสิทธิ์การประมาณ

พันธะเชิงเส้น และสัมประสิทธิ์เซปสตรอล เป็นตัวเปรียบเทียบ โดยใช้แบบจำลองฮิดเดนมาร์คอฟแบบต่อเนื่องเป็นระบบรู้จำเสียงพูดเพื่อฝึกฝนและทดสอบ

อัตราการรู้จำของระบบที่มีสถานะและค่า Gaussians Mixture ค่าต่างๆนั้นให้อัตราอัตราการรู้จำไม่แตกต่างกันมากนัก และการเพิ่มจำนวนสถานะและจำนวน Gaussians Mixture ก็ไม่ได้ส่งผลดีโดยรวมกับทุกค่าสัญญาณต่อสัญญาณรบกวน (SNR) โดยจำนวนสถานะและจำนวน Gaussians Mixture มีผลทำให้ระบบมีแนวโน้มของอัตราการรู้จำสูงขึ้นสำหรับค่า SNR สูงๆคือประมาณ 25 dB ขึ้นไป และส่วนใหญ่จะให้ค่าอัตราการรู้จำลดลงเมื่อจำนวนสถานะและจำนวน Gaussian Mixture เพิ่มขึ้นสำหรับค่า SNR ที่ลดลง และไม่มีระบบใดที่ให้อัตราการรู้จำสูงที่สุดอย่างเอกฉันท์ เมื่อคำนวณค่าเฉลี่ยของอัตราการรู้จำของระบบแต่ละระบบด้วยการนำค่าอัตราการรู้จำแต่ละ SNR ของระบบนั้นๆมาหาค่าเฉลี่ยเพื่อหาอัตราการรู้จำเฉลี่ยรวมสูงที่สุด จำนวนสถานะในแบบจำลองฮิดเดนมาร์คอฟเป็น 7 และจำนวน Gaussian Mixture เป็น 2 ให้อัตราการรู้จำเฉลี่ยรวมสูงที่สุด

การทำเน้นล่วงหน้ามีผลดีสำหรับการรู้จำเสียงพูดในบริเวณที่ไม่มีสัญญาณรบกวน ส่วนที่ค่า SNR 30 dB ลงมาจนถึงประมาณ 15 dB การทำเน้นล่วงหน้าแทบไม่มีผลต่ออัตราการรู้จำเลย หรือมีก็น้อยมาก เพราะการทำเน้นล่วงหน้าจะทำให้ค่าอัตราการรู้จำสูงขึ้นสำหรับเสียง Clean และ SNR สูงๆ เท่านั้น แต่เมื่อค่า SNR ต่ำลง ค่าอัตราการรู้จำลดต่ำลงอย่างเห็นได้ชัดเจนตั้งแต่ค่า SNR ต่ำกว่า 10 dB ดังนั้น เนื่องจากงานวิจัยนี้มุ่งเน้นการวิเคราะห์การรู้จำเสียงพูดเมื่อมีสัญญาณรบกวน จึงพิจารณาให้ค่าลักษณะสำคัญที่ใช้ไม่ทำเน้นล่วงหน้า

ผลการทดสอบอัตราการรู้จำเสียงตัวเลขภาษาไทยจากศูนย์ถึงเก้า และผลการทดสอบอัตราการรู้จำเสียงวรรณยุกต์ไทยทั้ง 5 คือเสียงสามัญ เอก โท ตรี และจัตวาที่ค่า SNR 0 dB โดยใช้ค่าลักษณะสำคัญแบบคงทนต่อสัญญาณรบกวน และก่อนลดสัญญาณรบกวนแสดงไว้ดังตารางที่ 5.1 พบว่าสามารถเพิ่มอัตราการรู้จำให้สูงขึ้นมากกว่าร้อยละหนึ่งร้อย ซึ่งสูงกว่าเป้าหมายที่ตั้งเอาไว้

ตารางที่ 5.1 อัตราการรู้จำเฉลี่ยสูงสุด เมื่อค่า SNR เป็น 0 dB

ชุดคำศัพท์	อัตราการรู้จำ ก่อนลดสัญญาณรบกวน (ร้อยละ)	อัตราการรู้จำ หลังลดสัญญาณรบกวน (ร้อยละ)
เสียงพูดตัวเลขภาษาไทย	10.42	38.00
เสียงวรรณยุกต์ไทย เสียงสามัญ	12.27	31.65
เสียงวรรณยุกต์ไทย เสียงเอก	14.49	33.64
เสียงวรรณยุกต์ไทย เสียงโท	16.25	38.69
เสียงวรรณยุกต์ไทย เสียงตรี	13.98	34.26
เสียงวรรณยุกต์ไทย เสียงจัตวา	13.75	32.44

อัตราการรู้จำเฉลี่ยสูงสุดของเสียงตัวเลขภาษาไทยศูนย์ถึงเก้า และอัตราการรู้จำเฉลี่ยสูงสุดของเสียงวรรณยุกต์ไทยทั้ง 5 คือเสียงสามัญ เอก โท ตรี และจัตวาโดยคำนวณจากการนำอัตราการรู้จำของแต่ละค่าลักษณะสำคัญในแต่ละอันดับสำหรับทุกค่า SNR มาทำการเฉลี่ยโดยใช้ค่าลักษณะสำคัญแบบคงทนต่อสัญญาณรบกวน และก่อนลดสัญญาณรบกวนแสดงไว้ดังตารางที่ 5.2

ตารางที่ 5.2 อัตราการรู้จำเฉลี่ยสูงสุดของแต่ละชุดคำศัพท์

ชุดคำศัพท์	อัตราการรู้จำ ก่อนลดสัญญาณรบกวน (ร้อยละ)	อัตราการรู้จำ หลังลดสัญญาณรบกวน (ร้อยละ)
เสียงพูดตัวเลขภาษาไทย	**36.65**	**62.95**
เสียงวรรณยุกต์ไทย เสียงสามัญ	**28.55**	**59.85**
เสียงวรรณยุกต์ไทย เสียงเอก	**35.69**	**62.67**
เสียงวรรณยุกต์ไทย เสียงโท	**32.56**	**57.96**
เสียงวรรณยุกต์ไทย เสียงตรี	**29.82**	**50.12**
เสียงวรรณยุกต์ไทย เสียงจัตวา	**34.26**	**55.33**

อัตราการรู้จำเฉลี่ยรวมสูงสุดของทุกชุดคำศัพท์ โดยคำนวณจากอัตราการรู้จำของทุกชุดคำศัพท์ของแต่ละอันดับของค่าลักษณะสำคัญมาทำการเฉลี่ย โดยกรณี Clean speech มีอัตราการรู้จำเป็นร้อยละ 73.41 ขณะที่อัตราการรู้จำเฉลี่ยรวมของทุกชุดคำศัพท์สูงสุดก่อนปรับลดสัญญาณรบกวนเป็นร้อยละ 77.67 และที่ค่า SNR 30 dB, 25 dB 20 dB, 15 dB, 10 dB, 5 dB และ 0 dB เป็นร้อยละ 72.40 69.28 63.58 57.28 52.02 45.14 และ 33.04 ขณะที่อัตราการรู้จำเฉลี่ยรวมของทุกชุดคำศัพท์สูงสุดก่อนปรับลดสัญญาณรบกวนเป็นร้อยละ 51.40 33.73 25.03 22.34 19.28 16.00 และ 14.19 ตามลำดับ

ค่าลักษณะสำคัญที่มีอันดับสูงกว่าโดยส่วนใหญ่แทบทุกกรณีให้อัตราการรู้จำสูงกว่าค่าลักษณะสำคัญที่มีอันดับต่ำกว่า และจากผลการวิจัยทั้งหมดเป็นที่ประจักษ์ว่า ค่าลักษณะสำคัญแบบคงทนต่อสัญญาณรบกวนสามารถเพิ่มอัตราการรู้จำเมื่อนำไปใช้ในบริเวณที่มีสัญญาณรบกวนได้ ซึ่งประสิทธิภาพที่เพิ่มขึ้นของค่าลักษณะสำคัญดังกล่าวเมื่อนำไปใช้งานในบริเวณที่มีสัญญาณรบกวนต่างๆกันทำให้มีอัตราการรู้จำเฉลี่ยเพิ่มขึ้นร้อยละ 78.69 โดยอัตราการรู้จำรวมสูงสุดของทั้งระบบหลังปรับลดสัญญาณรบกวนเป็นร้อยละ 57.68 ขณะที่โดยอัตราการรู้จำรวมสูงสุดของทั้งระบบก่อนปรับลดสัญญาณรบกวนเป็นร้อยละ 32.28

ดังนั้น ค่าลักษณะสำคัญแบบคงทนต่อสัญญาณรบกวนโดยใช้เทคนิคลดสัญญาณรบกวนแบบปรับตัวได้นี้ เหมาะสำหรับใช้เป็นค่าลักษณะสำคัญเมื่อใช้งานระบบรู้จำเสียงพูดภาษาไทยในบริเวณที่มีสัญญาณรบกวนได้เป็นอย่างดีเมื่อเทียบกับค่าลักษณะสำคัญเดิมที่ไม่ปรับลดสัญญาณรบกวน

5.2 ข้อเสนอแนะสำหรับงานวิจัยในอนาคต

1. ควรพิจารณาความสัมพันธ์ของค่าการปรับลดสัญญาณรบกวนในแต่ละกรอบสัญญาณเสียงพูด ว่าค่าการปรับลดสัญญาณรบกวนของกรอบสัญญาณเสียงพูดก่อนหน้าสามารถปรับตัว (Adapt) หรือสัมพันธ์กับค่าการปรับลดสัญญาณรบกวนของกรอบสัญญาณเสียงพูดปัจจุบันอย่างไร เพื่อจะได้มาตรฐานในการคำนวณค่าการปรับลดสัญญาณรบกวนในกรอบสัญญาณเสียงพูดกรอบต่อๆไป
2. ควรพิจารณาค่าการปรับลดสัญญาณรบกวนของทุกกรอบสัญญาณเสียงพูดในแต่ละคำในการปรับลดสัญญาณรบกวนโดยรวมของคำศัพท์ เช่นหาค่าเฉลี่ยของทุกกรอบสัญญาณเสียงพูดแล้วหักออกจากค่าอัตสหสัมพันธ์ของเมตริกซ์อัตสหสัมพันธ์
3. ควรทดลองนำไปประยุกต์ใช้หรือพัฒนากับฐานข้อมูลที่มีเสียงพูดผ่านช่องสัญญาณโทรศัพท์ (Telephone channel) เพื่อศึกษาประสิทธิภาพของค่าลักษณะสำคัญ
4. ควรทดลองใช้ค่าลักษณะสำคัญและค่าอนุพันธ์ร่วมกันเช่น สัมประสิทธิ์เซปสตรอลจากเทคนิคลดสัญญาณรบกวนแบบปรับตัวได้และค่าอนุพันธ์ของสัมประสิทธิ์เซปตรอล (Cepstral Derivative) จากเทคนิคลดสัญญาณรบกวนแบบปรับตัวได้อาจทำให้อัตราการรู้จำสูงขึ้น

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

รายการอ้างอิง

1. ระพีพัฒน์ เพ็ญศิริ. การรู้จำเสียงพูดตัวเลขไทยโดยไม่ขึ้นกับผู้พูดโดยใช้ไดนามิกไทม์วาร์ปปิง วิทยานิพนธ์ปริญญาวิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า จุฬาลงกรณ์มหาวิทยาลัย, 2538.
2. เสาวลักษณ์ อารีย์พงศา. การรู้จำเสียงพูดตัวเลขเป็นภาษาไทยแบบไม่ขึ้นกับผู้พูดโดยใช้วิธีฮิดเดนมาร์คอฟ โมเดล และเวกเตอร์ควอนไตซ์เซชั่น. วิทยานิพนธ์ปริญญาวิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า จุฬาลงกรณ์มหาวิทยาลัย, 2538.
3. ธีระ ภัทราพรนันท์. การรู้จำเสียงพูดภาษาไทยโดดๆ ไม่ขึ้นกับผู้พูดโดยการวัดสเปกตรัมดิสแตนส์และใช้ไดนามิกไทม์วาร์ปปิง. วิทยานิพนธ์ปริญญาวิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า จุฬาลงกรณ์มหาวิทยาลัย, 2538.
4. เอกฤทธิ์ มณีน้อย. ระบบรู้จำหน่วยเสียงสระภาษาไทยโดยใช้โครงข่ายประสาทเทียม. วิทยานิพนธ์ปริญญาวิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า จุฬาลงกรณ์มหาวิทยาลัย, 2541.
5. วิศรุต อาขุบุตร. ระบบการรู้จำคำไทยหลายพยางค์แบบไม่ขึ้นกับผู้พูดโดยใช้แบบจำลองฮิดเดนมาร์คอฟ. วิทยานิพนธ์ปริญญาวิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า จุฬาลงกรณ์มหาวิทยาลัย, 2539.
6. อุมาวสี ทาทอง. ระบบรู้จำคำเรียกพยัญชนะไทย : การศึกษาการวัดลักษณะสำคัญแบบต่างๆ. วิทยานิพนธ์ปริญญาวิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า จุฬาลงกรณ์มหาวิทยาลัย, 2544.
7. Junqua, J. C.,Haton, J. P., Robustness in Automatic Speech Recognition: Fundamentals and Applications. Kluwer Academic Publishers 1996.
8. Gales, M. J. F., Young, S. J., Robust Continuous Speech Recognition Using Parallel Model Combination. IEEE Transactions on Speech and Audio Processing 45 (Sept 1996): 352-359.
9. Oppenheim, A. V., Schafer, R.W., Discrete-Time Signal Processing. 3rd edition Prentice Hall 1999.
10. Rabiner, L. R., Juang, B. H., Fundamentals of Speech Recognition. Prentice Hall 1993.
11. O Shanughnessy, D., Linear Predictive Coding. IEEE Potentials. Volume: 7. 1, February 1988, Page(s): 29-32.

12. Cuntai GUAN, Yongbin CHEN, Boxiu WU, Direct Modification on LPC Coefficients with Application to Speech Enhancement and Improving The Performance of Speech Recognition in Noise.1993 IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP) (1993): 107-110.
13. Zhao Q., Shimamura T., SuZuki J, Linear Predictive Analysis of Noisy Speech. 1997 IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP) (1997): 585-588.
14. Suk, Y. H.,Choi, S. H., Lee, H. S., Cepstrum third-order Normalization Method for Noisy Speech Recognition. Electronics Letters 35 No.7 (1st April 1999): 527-528.
15. Lee L.-M., Chen J.-K., Wang H.-C., Nonlinear Cepstral Equalization method for Noisy Speech Recognition. IEE Proceedings 141 (December 1994): 397-402.
16. Shimamura, T., Kunieda, N., SuZuki, J., A Robust Linear Prediction Method for Noisy Speech. 1998 IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP) (1998): IV257-260.
17. Hu, H.T., Linear Prediction Analysis of Speech Signals in The Presence of White Gaussian Noise with Unknown Variance. IEE Proceedings 145 (August 1998): 303-308.
18. Burden, R.L., Faires, J.D., Numerical Analysis. 5th edition. PWS-KENT Publishing Company. 1993.
19. Hayes, M.H., Statistical Digital Signal Processing and Modeling. John Wiley & Sons, INC. 1996.
20. Chong, E.K.P., Zak, S.H., An Introduction to Optimization. 2nd edition. A Wiley-Interscience Publication. 2001.
21. Manolakis, D.G., Ingle, V.K., Kogon, S.M., Statistical and Adaptive Signal Processing; Spectral Estimation, Signal Modeling, Adaptive Filtering and Arra Processing, McGraw Hill. 2000.
22. Hamming, R.W., Numerical Methods for Scientist and Engineers, Prentice Hall. 1962
23. Tungthangthum A., Tone for Thai Speech Recognition. Proceedings of 1998 IEEE Asia-Pacific Conference on Circuits and Systems. Chiang Mai Thailand (November 1998): 157-160

ภาคผนวก

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ก

รายละเอียดอัตราการรู้จำเสียงพูดภาษาไทยเนื่องจากผลของการแปรค่าสถานะและ Gaussian Mixture

ตารางที่ ก.1 รายละเอียดอัตราการรู้จำเสียงพูดภาษาไทยเนื่องจากผลผลของการแปรค่าสถานะและ Gaussian Mixture

อันดับ	สถานะ/ G M	อัตราการรู้จำ (ร้อยละ)								
		Clean	**Snr30**	**Snr25**	**Snr20**	**Snr15**	**Snr10**	**Snr05**	**Snr00**	**ค่าเฉลี่ย**
ANR-CEP10	**7s1m**	65.54	65.08	65.67	64.33	61.08	55.25	47.54	32.71	57.15
	7s2m	69.96	69.96	70.04	68.67	65.79	60.88	52.08	31.33	61.09
	7s3m	72.50	72.08	69.92	66.33	62.42	57.79	49.67	29.46	60.02
	8s1m	66.71	66.71	66.67	65.58	62.67	58.42	49.79	35.75	59.04
	8s2m	71.21	71.13	71.21	67.92	65.04	60.67	51.71	34.33	**61.65**
	8s3m	74.29	73.17	71.50	68.29	62.67	58.04	50.83	30.75	61.17
ANR-CEP12	**7s1m**	68.46	68.13	66.88	66.08	61.75	55.96	47.83	31.67	58.35
	7s2m	73.58	72.83	71.58	68.04	64.25	60.00	53.75	37.88	**62.74**
	7s3m	72.92	72.21	70.88	67.04	62.50	57.38	50.63	35.42	61.12
	8s1m	70.54	70.50	69.71	67.42	63.50	58.71	51.79	37.83	61.25
	8s2m	73.04	72.63	70.50	66.75	62.58	58.08	52.38	36.38	61.54
	8s3m	73.83	72.92	70.92	65.71	59.00	54.83	49.85	35.17	60.28
ANR-CEP14	**7s1m**	71.42	70.79	69.38	65.79	62.71	57.96	51.79	36.17	60.75
	7s2m	74.00	73.29	71.67	68.71	63.42	59.92	56.54	41.21	**63.60**
	7s3m	74.88	74.17	72.42	66.96	60.96	57.46	53.67	36.71	62.15
	8s1m	71.92	71.25	69.92	66.67	62.67	59.17	52.29	40.67	61.82
	8s2m	73.38	73.00	70.96	66.25	60.79	56.13	51.38	36.88	61.10
	8s3m	76.96	75.83	73.54	68.71	62.29	56.96	53.46	35.88	63.08
ANR-CEP16	**7s1m**	71.67	71.00	69.92	66.92	62.42	57.42	52.67	37.96	61.25
	7s2m	74.42	73.96	71.38	66.63	60.08	56.08	51.58	37.00	60.14
	7s3m	75.46	74.83	71.83	66.96	60.58	56.04	53.08	36.50	61.91
	8s1m	73.17	72.75	71.29	67.92	62.63	58.75	52.54	40.83	62.49
	8s2m	75.38	74.58	73.00	67.88	60.63	56.29	52.83	36.92	62.19
	8s3m	76.00	75.96	72.46	67.96	61.13	56.58	52.83	38.00	**62.66**

ภาคผนวก ข

รายละเอียดเปรียบเทียบอัตราการรู้จำสำหรับคุณลักษณะสำคัญที่ผ่านและไม่ผ่านกรรมวิธีเน้นล่วงหน้า

ตารางที่ ข.1 เปรียบเทียบอัตราการรู้จำสำหรับคุณลักษณะสำคัญที่ผ่านและไม่ผ่านกรรมวิธีเน้นล่วงหน้า

	อัตราการรู้จำ (ร้อยละ)								
FEATURE	อันดับ/SNR	*Clean*	*SNR30*	*SNR25*	*SNR20*	*SNR15*	*SNR10*	*SNR05*	*SNR00*
	ANR-CEP10	74.29	73.17	71.50	68.29	**62.67**	**58.04**	50.83	30.75
	ANR-CEP12	73.83	72.92	70.92	65.71	59.00	54.83	49.75	35.17
	ANR-CEP14	**76.96**	75.83	**73.54**	**68.71**	62.29	56.96	**53.46**	35.88
	ANR-CEP16	76.00	**75.96**	72.46	67.96	61.13	56.58	52.83	**38.00**
	PANR-CEP10	75.79	73.50	71.63	68.42	56.92	38.92	22.21	11.83
	PANR-CEP12	77.63	74.83	73.04	70.63	62.96	42.00	23.21	12.08
	PANR-CEP14	77.92	75.71	73.83	71.54	62.54	45.13	26.92	10.63
	PANR-CEP16	**78.42**	**75.83**	**74.96**	**72.25**	**66.29**	**49.21**	**30.83**	**14.21**

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ค

รายละเอียดผลการรู้จำเสียงพูดภาษาไทย

ตารางที่ ค.1 ผลการรู้จำเสียงตัวเลขภาษาไทยสำหรับค่าลักษณะสำคัญ LPC และ LPC-CEP อันดับ 10 12 14 และ 16

เสียง	อัตราการรู้จำ (ร้อยละ)								
	อันดับ/SNR	Clean	SNR30	SNR25	SNR20	SNR15	SNR10	SNR05	SNR00	ค่าเฉลี่ย
ค่าลักษณะสำคัญ	LPC10	66.75	43.13	25.25	20.81	16.58	13.58	12.08	7.46	25.71
	LPC12	65.67	44.88	27.96	18.54	15.50	14.04	11.25	11.00	26.11
	LPC14	66.75	48.25	28.46	21.25	17.58	14.21	13.13	10.83	27.56
	LPC16	**78.88**	**61.67**	**51.04**	**38.75**	**27.04**	**20.79**	**18.38**	**18.88**	***39.43***
	LPC-CEP10	78.58	58.17	43.21	29.42	23.17	19.04	15.08	12.67	34.92
	LPC-CEP12	79.88	60.92	**45.00**	**32.96**	**26.46**	**21.38**	18.38	**15.29**	***37.53***
	LPC-CEP14	80.58	**61.63**	44.13	30.88	25.08	20.96	**17.04**	12.92	36.65
	LPC-CEP16	**81.29**	61.50	41.71	30.00	23.50	17.75	13.38	10.42	34.94

ตารางที่ ค.2 ผลการรู้จำเสียงตัวเลขภาษาไทยสำหรับค่าลักษณะสำคัญ ANR-LPC และ ANR-CEP อันดับ 10 12 14 และ 16

เสียง	อัตราการรู้จำ (ร้อยละ)								
	อันดับ/SNR	Clean	SNR30	SNR25	SNR20	SNR15	SNR10	SNR05	SNR00	ค่าเฉลี่ย
ค่าลักษณะสำคัญ	ANR-LPC10	73.00	71.58	71.71	64.21	**56.21**	**50.00**	**43.33**	**29.04**	***57.39***
	ANR-LPC12	74.42	73.79	**70.79**	63.17	53.21	43.25	33.25	23.75	54.45
	ANR-LPC14	75.58	74.04	70.25	62.25	52.08	43.00	33.92	24.13	54.41
	ANR-LPC16	**75.79**	**74.88**	71.21	**64.38**	54.88	46.38	39.33	27.46	56.79
	ANR-CEP10	74.29	73.17	71.50	68.29	**62.67**	58.04	50.83	30.75	61.19
	ANR-CEP12	73.83	72.92	70.92	65.71	59.00	54.83	49.75	35.17	60.27
	ANR-CEP14	**76.96**	75.83	**73.54**	**68.71**	62.29	**56.96**	**53.46**	35.88	***62.95***
	ANR-CEP16	76.00	**75.96**	72.46	67.96	61.13	56.58	52.83	**38.00**	62.62

ตารางที่ ค.3 ผลการรู้จำเสียงวรรณยุกต์ไทยเสียงสามัญสำหรับค่าลักษณะสำคัญ LPC และ LPC-CEP อันดับ 10 12 14 และ 16

เสียง	อัตราการรู้จำ (ร้อยละ)								
	อันดับ/SNR	Clean	SNR30	SNR25	SNR20	SNR15	SNR10	SNR05	SNR00	ค่าเฉลี่ย
สามัญ	LPC10	53.69	37.78	21.76	14.77	12.95	12.39	12.67	**13.64**	22.46
	LPC12	59.89	40.97	21.93	15.57	**13.69**	12.95	**13.30**	12.39	23.84
	LPC14	57.05	39.32	22.61	**15.74**	12.67	12.56	12.78	12.67	23.18
	LPC16	**62.67**	**46.36**	**27.27**	15.51	**13.69**	**13.07**	12.10	12.56	**25.40**
	LPC-CEP10	66.19	38.07	22.84	19.32	**19.60**	17.16	14.77	12.27	26.28
	LPC-CEP12	68.35	36.53	24.20	20.80	18.35	16.70	**14.94**	12.50	26.92
	LPC-CEP14	73.98	39.43	24.49	18.81	17.27	15.45	13.58	**14.32**	27.17
	LPC-CEP16	**74.38**	**43.64**	**25.97**	**21.31**	**19.60**	**17.22**	13.64	12.67	**28.55**

ตารางที่ ค.4 ผลการรู้จำเสียงวรรณยุกต์ไทยเสียงสามัญสำหรับค่าลักษณะสำคัญ ANR-LPC และ ANR-CEP อันดับ 10 12 14 และ 16

เสียง	อัตราการรู้จำ (ร้อยละ)								
	อันดับ/SNR	Clean	SNR30	SNR25	SNR20	SNR15	SNR10	SNR05	SNR00	ค่าเฉลี่ย
สามัญ	ANR-LPC10	67.05	65.97	63.07	59.03	**53.52**	48.41	40.68	24.55	52.79
	ANR-LPC12	69.94	68.75	66.14	58.92	53.01	**48.81**	**40.80**	**28.52**	54.36
	ANR-LPC14	72.95	71.14	64.72	55.51	49.26	45.34	39.32	25.57	52.98
	ANR-LPC16	**76.65**	**75.28**	**70.34**	**59.26**	50.45	45.45	38.07	28.07	**55.45**
	ANR-CEP10	71.88	72.22	70.23	**67.84**	**63.64**	**57.50**	**45.51**	**31.65**	**60.06**
	ANR-CEP12	74.66	73.41	72.27	67.50	62.56	56.82	42.84	28.75	59.85
	ANR-CEP14	**78.07**	**76.19**	**72.33**	65.23	57.78	50.11	37.67	25.68	57.88
	ANR-CEP16	77.05	74.66	70.97	63.07	55.23	47.73	34.32	25.11	56.02

ตารางที่ ค.5 ผลการรู้จำเสียงวรรณยุกต์ไทยเสียงเอกสำหรับค่าลักษณะสำคัญ LPC และ LPC-CEP อันดับ 10 12 14 และ 16

เสียง	อัตราการรู้จำ (ร้อยละ)								
	อันดับ/SNR	Clean	SNR30	SNR25	SNR20	SNR15	SNR10	SNR05	SNR00	ค่าเฉลี่ย
เอก	LPC10	65.63	43.86	25.23	**18.81**	**16.36**	14.15	13.52	13.69	26.41
	LPC12	68.41	43.75	23.75	16.02	14.15	13.13	12.50	12.50	25.53
	LPC14	**69.49**	44.26	**26.59**	18.41	15.45	13.92	13.75	**15.28**	27.14
	LPC16	68.81	**46.42**	26.02	17.84	15.06	**14.83**	**13.86**	14.49	**27.17**
	LPC-CEP10	73.75	47.73	33.24	23.47	18.86	16.19	15.28	14.72	30.41
	LPC-CEP12	76.08	50.45	34.94	25.17	21.42	19.49	**19.20**	**17.05**	32.98
	LPC-CEP14	**77.33**	**56.31**	**40.45**	**30.17**	**24.15**	**22.33**	18.86	15.23	**35.69**
	LPC-CEP16	77.05	55.91	37.73	25.34	21.14	21.36	17.61	11.93	33.51

ตารางที่ ค.6 ผลการรู้จำเสียงวรรณยุกต์ไทยเสียงเอกสำหรับค่าลักษณะสำคัญ ANR-LPC และ ANR-CEP อันดับ 10 12 14 และ 16

เสียง	อัตราการรู้จำ (ร้อยละ)								
	อันดับ/SNR	Clean	SNR30	SNR25	SNR20	SNR15	SNR10	SNR05	SNR00	Average
เอก	ANR-LPC10	73.52	71.42	67.95	59.32	50.17	42.73	35.34	30.45	53.86
	ANR-LPC12	78.69	77.10	73.30	64.55	50.51	44.55	37.84	27.95	56.81
	ANR-LPC14	79.83	77.84	73.52	64.03	51.36	44.72	38.13	30.00	57.43
	ANR-LPC16	**81.99**	**80.74**	**75.68**	**64.77**	**52.05**	**45.91**	**41.59**	**32.50**	**59.40**
	ANR-CEP10	75.68	74.94	73.13	68.52	**62.67**	**56.88**	**48.24**	33.52	61.70
	ANR-CEP12	80.57	79.49	76.65	68.41	60.06	54.32	45.34	33.01	62.23
	ANR-CEP14	**82.27**	**81.53**	**78.98**	**70.00**	58.53	52.44	45.40	32.22	**62.67**
	ANR-CEP16	81.53	80.68	77.56	69.26	57.73	51.36	45.68	**33.64**	62.18

ตารางที่ ค.7 ผลการรู้จำเสียงวรรณยุกต์ไทยเสียงโทสำหรับค่าลักษณะสำคัญ LPC และ LPC-CEP อันดับ 10 12 14 และ 16

เสียง	อัตราการรู้จำ (ร้อยละ)								
	อันดับ/SNR	Clean	SNR30	SNR25	SNR20	SNR15	SNR10	SNR05	SNR00	ค่าเฉลี่ย
โท	LPC10	62.05	39.32	24.03	18.47	16.31	**14.43**	**12.95**	**12.61**	25.02
	LPC12	65.23	**42.16**	**25.63**	18.98	15.17	13.69	12.73	12.50	**25.76**
	LPC14	61.48	38.41	24.94	**19.66**	**16.53**	14.09	12.67	12.50	25.04
	LPC16	**65.57**	40.80	24.55	18.18	15.57	13.81	12.78	12.50	25.47
	LPC-CEP10	76.19	39.77	27.33	21.48	17.67	17.05	15.45	13.07	28.50
	LPC-CEP12	**77.33**	42.27	28.01	19.72	18.35	15.68	13.35	12.67	28.47
	LPC-CEP14	76.48	43.41	28.98	23.98	20.34	15.63	13.13	12.56	31.81
	LPC-CEP16	76.93	**46.02**	**31.02**	**26.42**	**24.77**	**22.05**	**17.05**	**16.25**	**32.56**

ตารางที่ ค.8 ผลการรู้จำเสียงวรรณยุกต์ไทยเสียงโทสำหรับค่าลักษณะสำคัญ ANR-LPC และ ANR-CEP อันดับ 10 12 14 และ 16

เสียง	อัตราการรู้จำ (ร้อยละ)								
	อันดับ/SNR	Clean	SNR30	SNR25	SNR20	SNR15	SNR10	SNR05	SNR00	ค่าเฉลี่ย
โท	ANR-LPC10	69.94	68.64	**66.31**	**62.39**	**57.39**	**53.98**	45.40	23.86	**55.99**
	ANR-LPC12	69.03	67.78	63.20	58.35	53.13	50.51	**46.14**	26.14	54.29
	ANR-LPC14	70.91	**69.60**	64.66	56.70	51.76	47.10	42.78	27.50	53.88
	ANR-LPC16	**72.67**	71.31	66.19	55.28	47.73	43.75	41.82	**32.50**	53.91
	ANR-CEP10	67.50	67.10	66.42	62.56	56.70	**51.88**	47.39	34.83	56.80
	ANR-CEP12	69.38	69.20	67.90	**63.58**	**56.88**	49.03	46.02	33.30	56.91
	ANR-CEP14	71.70	71.25	68.81	62.90	55.45	49.60	**47.44**	36.53	**57.96**
	ANR-CEP16	**71.88**	**71.31**	**68.92**	61.31	54.03	48.52	46.88	**38.69**	57.69

ตารางที่ ค.9 ผลการรู้จำเสียงวรรณยุกต์ไทยเสียงตรีสำหรับค่าลักษณะสำคัญ LPC และ LPC-CEP อันดับ 10 12 14 และ 16

เสียง	อัตราการรู้จำ (ร้อยละ)								
	อันดับ/SNR	Clean	SNR30	SNR25	SNR20	SNR15	SNR10	SNR05	SNR00	ค่าเฉลี่ย
ตรี	LPC10	50.74	32.90	22.44	17.10	14.77	13.69	**13.69**	13.18	22.31
	LPC12	52.44	36.08	22.95	18.52	15.91	**13.92**	12.73	13.41	23.25
	LPC14	55.11	38.07	23.30	16.53	12.95	11.31	11.02	12.95	22.66
	LPC16	**55.85**	**38.69**	**26.31**	**19.60**	**16.02**	13.07	11.99	**13.98**	**24.44**
	LPC-CEP10	62.33	41.31	**30.80**	**25.06**	**21.82**	15.63	13.98	13.75	28.09
	LPC-CEP12	64.77	42.67	28.13	22.67	18.92	14.38	13.47	13.81	27.35
	LPC-CEP14	67.84	36.53	22.50	16.65	15.74	14.60	13.92	14.20	25.25
	LPC-CEP16	**70.45**	**46.82**	30.74	22.10	20.11	**17.73**	**14.55**	**16.08**	**29.82**

ตารางที่ ค.10 ผลการรู้จำเสียงวรรณยุกต์ไทยเสียงตรีสำหรับค่าลักษณะสำคัญ ANR-LPC และ ANR-CEP อันดับ 10 12 14 และ 16

เสียง	อัตราการรู้จำ (ร้อยละ)								
	อันดับ/SNR	Clean	SNR30	SNR25	SNR20	SNR15	SNR10	SNR05	SNR00	ค่าเฉลี่ย
ตรี	ANR-LPC10	56.31	55.74	53.01	47.90	44.77	42.44	35.23	21.88	44.66
	ANR-LPC12	58.30	57.78	53.86	48.13	42.95	41.14	37.73	25.17	45.63
	ANR-LPC14	57.78	57.10	54.03	47.90	43.92	41.59	39.66	32.84	46.85
	ANR-LPC16	**59.20**	**58.92**	**56.02**	**49.89**	**46.99**	**45.40**	**43.75**	**34.26**	**49.03**
	ANR-CEP10	56.14	56.36	54.26	48.35	44.26	41.82	40.45	29.15	46.35
	ANR-CEP12	57.90	57.33	57.83	49.43	44.09	41.88	39.66	29.03	47.14
	ANR-CEP14	61.53	61.48	58.35	51.08	46.82	43.69	**42.05**	**31.08**	49.51
	ANR-CEP16	**63.86**	**62.44**	**59.49**	**52.16**	**47.10**	**43.75**	41.82	30.34	**50.12**

ตารางที่ ค.11 ผลการรู้จำเสียงวรรณยุกต์ไทยเสียงจัตวาสำหรับค่าลักษณะสำคัญ LPC และ LPC-CEP อันดับ 10 12 14 และ 16

เสียง	อัตราการรู้จำ (ร้อยละ)								
	อันดับ/SNR	Clean	SNR30	SNR25	SNR20	SNR15	SNR10	SNR05	SNR00	ค่าเฉลี่ย
จัตวา	LPC10	62.10	48.07	32.05	19.72	15.45	13.75	12.33	10.51	26.75
	LPC12	61.65	43.07	29.20	18.41	14.20	13.24	**15.74**	10.57	25.75
	LPC14	59.77	46.82	**32.37**	18.35	15.34	13.41	10.68	9.32	25.76
	LPC16	**70.17**	**47.61**	31.82	**19.55**	**15.91**	**13.81**	11.36	**12.10**	**27.79**
	LPC-CEP10	81.14	47.78	30.45	20.40	15.91	13.86	15.51	**14.26**	29.91
	LPC-CEP12	80.74	49.09	33.18	**28.92**	**26.53**	**21.65**	16.65	13.58	33.79
	LPC-CEP14	84.49	51.31	31.88	26.70	24.03	21.14	**18.64**	15.91	34.26
	LPC-CEP16	**85.34**	**54.49**	**35.19**	28.13	24.94	19.55	17.84	13.75	**34.90**

ตารางที่ ค.12 ผลการรู้จำเสียงวรรณยุกต์ไทยเสียงจัตวาสำหรับค่าลักษณะสำคัญ ANR-LPC และ ANR-CEP อันดับ 10 12 14 และ 16

เสียง	อัตราการรู้จำ (ร้อยละ)								
	อันดับ/SNR	Clean	SNR30	SNR25	SNR20	SNR15	SNR10	SNR05	SNR00	ค่าเฉลี่ย
จัตวา	ANR-LPC10	58.41	58.64	56.70	51.36	44.09	38.13	32.90	23.24	45.43
	ANR-LPC12	62.44	60.57	58.30	53.64	46.36	41.82	36.48	**28.64**	48.53
	ANR-LPC14	62.95	60.51	57.33	53.35	46.82	**42.16**	**38.24**	27.05	48.55
	ANR-LPC16	**65.91**	**63.58**	**60.34**	**55.11**	**49.09**	41.99	34.77	24.20	**49.37**
	ANR-CEP10	62.67	63.92	63.64	**59.49**	53.75	46.02	36.14	19.49	50.64
	ANR-CEP12	67.84	67.10	64.09	58.07	53.01	48.01	40.06	24.77	52.87
	ANR-CEP14	**69.94**	68.13	63.69	57.95	**54.20**	**50.17**	44.83	31.82	55.09
	ANR-CEP16	69.83	**68.58**	**65.17**	58.69	53.58	49.20	**45.17**	**32.44**	**55.33**

ประวัติผู้เขียนวิทยานิพนธ์

นายณัฐพล อุ่นศรี เกิดเมื่อวันที่ 31 สิงหาคม พ.ศ. 2520 ที่ จ.อุบลราชธานี สำเร็จการศึกษาปริญญาตรีวิศวกรรมศาสตรบัณฑิต สาขาวิศวกรรมไฟฟ้า ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ มหาวิทยาลัยขอนแก่น ในปีการศึกษา 2541 และเข้าศึกษาต่อในหลักสูตรวิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย โดยได้รับทุนโครงการพัฒนาอาจารย์วิทยาเขตสารสนเทศ มหาวิทยาลัยเกษตรศาสตร์ ปัจจุบันดำรงตำแหน่งเป็นอาจารย์ประจำภาควิชาวิศวกรรมไฟฟ้าและคอมพิวเตอร์ มหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตเฉลิมพระเกียรติจังหวัดสกลนคร งานวิจัยที่สนใจได้แก่ หัวข้อทั่วไปเกี่ยวกับกรรมวิธีสัญญาณดิจิตอล (Digital Signal Processing) การรู้จำและการจำแนกรูปแบบ (Pattern Recognition and Classification) โดยเฉพาะอย่างยิ่งเกี่ยวกับ ระบบรู้จำเสียงพูดเป็นภาษาไทย (Thai Speech Recognition) และการลดสัญญาณรบกวน (Noise Reduction)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย