การลดสัญญาณรบกวนเติมแต่งแบบคงที่ก่อนการรู้จำเสียงพูดโดยการลบสเปกตรัมแบบปรับค่าถ่วงน้ำหนักได้

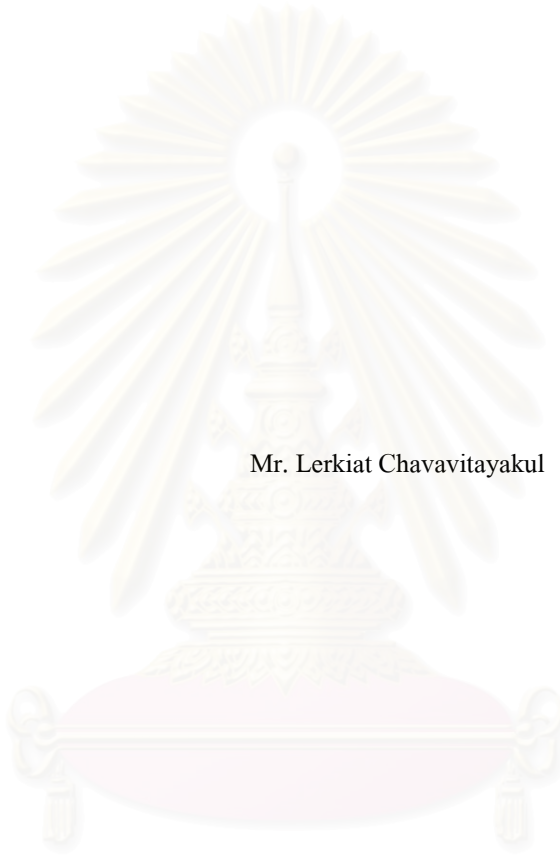
นายเลอเกียรติ ชววิทยากุล

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2549
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

ADDITIVE STATIONARY NOISE SUBTRACTION FOR SPEECH RECOGNITION USING ADJUSTABLE WEIGHT SPECTRAL SUBTRACTION

Mr. Lerkiat Chavavitayakul



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science Program in Computer Science
Department of Computer Engineering
Faculty of Engineering
Chulalongkorn University
Academic Year 2006

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การลดสัญญาณรบกวนเติมแต่งแบบคงที่ก่อนการรู้จำเสียงพูดโดยการลบสเปกตรัมแบบปรับค่าถ่วงน้ำหนักได้
โดย	นาย เลอเกียรติ ชววิทยากุล
สาขาวิชา	วิทยาศาสตร์คอมพิวเตอร์
อาจารย์ที่ปรึกษา	อาจารย์ ดร.อดิวงศ์ สุชาโต
อาจารย์ที่ปรึกษาร่วม	อาจารย์ ดร. โปรดปราน บุณยพุกกณะ

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

..... คณบดีคณะวิศวกรรมศาสตร์
(ศาสตราจารย์ ดร.ดิเรก ลาวัณย์ศิริ)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(ผู้ช่วยศาสตราจารย์ นงลักษณ์ โควาวิสารัช)

..... อาจารย์ที่ปรึกษา
(อาจารย์ ดร.อดิวงศ์ สุชาโต)

..... อาจารย์ที่ปรึกษาร่วม
(อาจารย์ ดร. โปรดปราน บุณยพุกกณะ)

..... กรรมการ
(อาจารย์ ดร.อรรถสิทธิ์ สุรฤกษ์)

..... กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร. นิศาชล ตั้งเสงี่ยมวิสัย)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

เลอเกียรติ ชววิทยากุล : การลดสัญญาณรบกวนเติมแต่งแบบคงที่ก่อนการรู้จำเสียงพูดโดยการลบสเปกตรัมแบบปรับค่าถ่วงน้ำหนักได้. (ADDITIVE STATIONARY NOISE SUBTRACTION FOR SPEECH RECOGNITION USING ADAPTIVE WEIGHT SPECTRAL SUBTRACTION) อ. ที่ปรึกษา : อ.ดร.อดิวงศ์ สุชาโต, อ.ที่ปรึกษาร่วม : อ.ดร. โปรดปราน บุณยพุกกณะ, 42 หน้า.

ในระบบรู้จำเสียงพูดนั้น หากมีสัญญาณรบกวนรวมอยู่ในสัญญาณเสียงจะมีผลทำให้ความแม่นยำและประสิทธิภาพในการรู้จำลดลง ทั้งนี้ได้มีงานวิจัยที่ทำการศึกษาขั้นตอนวิธีต่างๆอย่างกว้างขวาง โดยมีวัตถุประสงค์เพื่อลดสัญญาณรบกวนให้มากที่สุดก่อนผ่านเข้าสู่กระบวนการรู้จำต่อไป ในวิทยานิพนธ์นี้ได้นำเสนอขั้นตอนวิธีการลบสเปกตรัมแบบปรับค่าถ่วงน้ำหนักได้ โดยใช้การประมาณขนาดสัญญาณรบกวนจากสัญญาณเสียงพูดในช่วง 0 – 100 มิลลิวินาทีแรกซึ่งถือว่าเป็นช่วงสัญญาณเงียบเพื่อใช้เป็นตัวแปรเสริมในการเลือกค่าถ่วงน้ำหนักที่เหมาะสม งานวิจัยนี้นำเสนอการ ใช้ฟังก์ชันความสัมพันธ์ระหว่างค่าถ่วงน้ำหนักและค่าเฉลี่ยพลังงานสัญญาณรบกวนแบบเชิงเส้น ซึ่งสามารถเพิ่มความแม่นยำในการรู้จำเสียงพูดได้ดีในกรณีค่าอัตราส่วนสัญญาณต่อสัญญาณรบกวนมีค่าต่ำ ผลการทดลองแสดงให้เห็นว่าค่าความแม่นยำในการรู้จำมีค่ามากขึ้นเมื่อใช้กับขั้นตอนวิธีที่นำเสนอนี้ โดยเมื่อเปรียบเทียบขั้นตอนวิธีที่นำเสนอและขั้นตอนวิธีการลบสเปกตรัมแบบการประมาณแม็กซิมัมไลค์ลิฮูดพบว่า ค่าความแม่นยำโดยรวมเมื่อใช้ขั้นตอนวิธีที่นำเสนอดีขึ้น 17.79% เมื่อใช้ชุดข้อมูลเสียงพูดตัวเลข

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาควิชา.....วิศวกรรมคอมพิวเตอร์.....	ลายมือชื่อนิสิต.....เลอเกียรติ ชววิทยากุล.....
สาขาวิชา.....วิทยาศาสตร์คอมพิวเตอร์.....	ลายมือชื่ออาจารย์ที่ปรึกษา.....อดิวงศ์.....
ปีการศึกษา.....2549.....	ลายมือชื่ออาจารย์ที่ปรึกษาร่วม.....[illegible].....

487 14422 21 : MAJOR COMPUTER SCIENCE

KEY WORD: ADJUSTABLE WEIGHT / SPECTAL SUBTRACTION / SPEECH RECOGNITION

LERKIAT CHAVAVITAYAKUL : ADDITIVE STATIONARY NOISE SUBTRACTION FOR SPEECH RECOGNITION USING ADJUSTABLE WEIGHT SPECTRAL SUBTRACTION. THESIS ADVISOR : ATIWONG SUCHATO, Ph.D., THESIS COADVISOR : PROADPRAN PUNYABUKKANA, Ph.D., 42 pp.

Impact from Noisy signal degrades accuracy in speech recognition system. Noise reduction techniques have been developed continuously in order to reduce noise signal from desired speech before it is passed to the recognizer. In this thesis, an adaptive weight spectral subtraction is proposed. The method estimates noise from the first 100 milliseconds of the original speech and threats it as background noise. Then a linear function of weight and the average noise energy is calculated. This method is efficient with speech that SNR is low, Experiments show that the proposed method increases recognition accuracy by 17.79% when compared with maximum likelihood spectral subtraction and tested in single digit spoken speech data

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Department ..... Computer Engineering	Student's signature .....
Field of study .... Computer Science	Advisor's signature .....
Academic year ..... 2006.....	Co-advisor's signature .....

กิตติกรรมประกาศ

ขอขอบคุณ อ. ดร. อติวงศ์ สุชาโต อ.ดร.โปรดปราน บุณยพุกกณะ ที่คอยให้คำปรึกษาและรับฟังความคิดเห็นในการทำวิทยานิพนธ์ อีกทั้งยังคอยช่วยให้กำลังใจและวางแผนในการทำงานให้อย่างมีประสิทธิภาพ

ขอขอบคุณคณะกรรมการสอบวิทยานิพนธ์ ผศ.นงลักษณ์ โควาวิสารัช อ.ดร.อรรถสิทธิ์ สุรฤกษ์ และ ผศ.ดร.นิศาชล ตั้งเสงี่ยมวิสัย ที่ช่วยให้คำแนะนำที่เป็นประโยชน์ต่องานวิจัยและให้ความเป็นกันเองกับนิสิตอย่างมากในการสอบวิทยานิพนธ์

ขอขอบคุณบิดา มารดา ที่ช่วยเลี้ยงดูและอบรมให้ข้าพเจ้าเป็นคนดีและมีความตั้งใจในการศึกษาเล่าเรียน

ขอขอบคุณเพื่อนๆร่วมชั้นปี เพื่อนอาม เพื่อนบี เพื่อนโน๊ต พี่เอ๋ที่กำลังศึกษาในระดับปริญญาเอก กวาง เปิ้ลที่คอยช่วยเหลือผมอย่างมาก นุชที่ช่วยทำให้เกิดแนวคิดของวิจัย น้องนาถและน้องโรจน์จากห้องแลปที่คอยให้ความช่วยเหลือในด้านต่างๆ

ขอขอบคุณ นาย ชัชวาล ชิดชัยมงคล (พี่ชัช) ที่อนุญาตและส่งเสริมการเรียนปริญญาโทของข้าพเจ้าอย่างดีมาตลอด

ขอขอบคุณอาจารย์ นิศาชล ตั้งเสงี่ยมวิสัยที่ช่วยปลูกฝังความรู้และแนวคิดในการทำงานต่างๆตั้งแต่ครั้งที่ข้าพเจ้าเรียนอยู่ในระดับปริญญาตรีจนถึงปัจจุบัน

สุดท้ายนี้ขอขอบคุณ นส.พิรดี ตติยกวี ที่เป็นแรงผลักดันให้ข้าพเจ้าตั้งใจเรียนและทำงานจนถึงทุกวันนี้

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญ

หน้า

กิตติกรรมประกาศ ..... ฉ
สารบัญ ..... ช
สารบัญภาพ ..... ฌ
สารบัญตาราง ..... ฎ
บทที่ 1 บทนำ 1
ความเป็นมาและความสำคัญของปัญหา 1
วัตถุประสงค์ของการวิจัย 3
ขอบเขตของวิทยานิพนธ์ 3
วิธีดำเนินการวิจัย 3
ประโยชน์ที่คาดว่าจะได้รับ 4
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง 5
ทฤษฎีที่เกี่ยวข้อง 5
1. หลักการและทฤษฎีที่เกี่ยวข้องกับขั้นตอนวิธีในการลดสัญญาณรบกวน 6
2. หลักการและทฤษฎีทั่วไปที่เกี่ยวข้อง 11
งานวิจัยที่เกี่ยวข้อง 13
บทที่ 3 แนวคิดการวิจัย 16
การลดสัญญาณรบกวนด้วยขั้นตอนวิธีต่างๆ 16
1. การวิเคราะห์ผลค่าถ่วงน้ำหนักที่มีผลต่อความแม่นยำ 17
2. การประมาณขนาดของสัญญาณรบกวน 18
3. การหาฟังก์ชันการกำหนดค่าถ่วงน้ำหนักจากขนาดของสัญญาณรบกวน 19
4. การประเมินผล 19
บทที่ 4 การทดลองและวิเคราะห์ผล 20
วิธีการทดลอง 20
1. การเตรียมข้อมูลเสียงพูด 20
2. การสอนระบบรู้จำเสียงพูด 21
3. การลดสัญญาณรบกวนด้วยขั้นตอนวิธีต่างๆ 25
4. การวัดความแม่นยำในการรู้จำเสียงพูด 25
ผลการทดลอง 26
บทที่ 5 สรุปผลการวิจัย 38

หน้า

สรุปการวิจัย 38
ข้อจำกัดของงานวิจัย 39
ข้อเสนอแนะ 39
คุณค่าทางวิชาการของงานวิจัย 39
รายการอ้างอิง 41
ประวัติผู้เขียนวิทยานิพนธ์ 42

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญภาพ

หน้า

รูปที่ 1.1 สัญญาณเสียงพูดสะอาด 2
รูปที่ 1.2 สัญญาณเสียงพูดที่มีสัญญาณรบกวน 2
รูปที่ 2.1 ขั้นตอนในกระบวนการเรียนรู้เสียงพูด 5
รูปที่ 2.2 ขั้นตอนในกระบวนการรู้จำเสียงพูด 6
รูปที่ 2.3 (ก) สัญญาณเสียงพูดสะอาด (ข) สัญญาณเสียงพูดที่ตรวจหาได้ และ (ค) สัญญาณเสียงพูดที่ถูกลดสัญญาณรบกวนด้วยขั้นตอนวิธีการลบสเปกตรัมแบบการประมาณแม็กซิมัมไลค์ลิฮูด 9
รูปที่ 2.4 (ก) สัญญาณเสียงพูดสะอาด (ข) สัญญาณเสียงพูดที่เราทราบได้และ (ค) สัญญาณเสียงพูดที่ถูกลดสัญญาณรบกวนด้วยขั้นตอนวิธีการกรองวีเนอร์ 11
รูปที่ 2.5 การเปรียบเทียบค่าอัตราส่วนสัญญาณต่อสัญญาณรบกวนระหว่างขั้นตอนวิธีการลบสเปกตรัมและวิธีที่นำเสนอใน[4] 14
รูปที่ 3.1 กราฟแสดงผลการเปรียบเทียบความแม่นยำในการรู้จำระหว่างขั้นตอนวิธีการลบสเปกตรัมแบบตัวกรองลดสัญญาณรบกวน ขั้นตอนวิธีการลบสเปกตรัมแบบการประมาณแม็กซิมัมไลค์ลิฮูด ขั้นตอนการกรองวีเนอร์และกรณีไม่ใช้ขั้นตอนวิธีใดเลย 16
รูปที่ 4.1 ตัวอย่างเนื้อความแฟ้มฉลากของสัญญาณเสียงพูดตัวเลขภาษาไทย 21
รูปที่ 4.2 ตัวอย่างเนื้อความแฟ้มฉลากของสัญญาณเสียงพูดประโยคภาษาไทยทั่วไป 22
รูปที่ 4.3 กราฟแสดงผลการทดลองเปลี่ยนค่าถ่วงน้ำหนักต่อความแม่นยำในการรู้จำเสียงพูด เมื่อใช้ข้อมูลชุดเสียงพูดตัวเลข 27
รูปที่ 4.4 ฟังก์ชันความสัมพันธ์ระหว่างค่าถ่วงน้ำหนักและค่าเฉลี่ยพลังงานสัญญาณรบกวนแบบขั้นบันได 28
รูปที่ 4.5 กราฟแสดงผลการทดลองเปลี่ยนค่าขีดแบ่งค่าเฉลี่ยพลังงานของสัญญาณรบกวนต่อความแม่นยำในการรู้จำเสียงพูดเมื่อใช้ฟังก์ชันขั้นบันได 30
รูปที่ 4.6 ฟังก์ชันความสัมพันธ์ระหว่างค่าถ่วงน้ำหนักและค่าเฉลี่ยพลังงานสัญญาณรบกวนแบบเชิงเส้น 31
รูปที่ 4.7 กราฟแสดงผลการทดลองเปลี่ยนค่าขีดแบ่งค่าเฉลี่ยพลังงานของสัญญาณรบกวนต่อความแม่นยำในการรู้จำเสียงพูดเมื่อใช้ฟังก์ชันเชิงเส้น 33
รูปที่ 4.8 กราฟแสดงค่าความแม่นยำในการรู้จำเสียงพูดตัวเลขระหว่างขั้นตอนวิธีการลบสเปกตรัมแบบปรับค่าถ่วงน้ำหนักได้ ขั้นตอนวิธีการลบสเปกตรัมแบบการประมาณแมกซิมัมไลค์ลิฮูดและกรณีไม่ใช้ขั้นตอนวิธีใดเลย เมื่อใช้ชุดข้อมูลทดสอบประโยคตัวเลขภาษาไทย 34

รูปที่ 4.9 กราฟแสดงค่าความแม่นยำในการรู้จำเสียงพูดระหว่างขั้นตอนวิธีการลบสเปกตรัมแบบปรับค่าถ่วงน้ำหนักได้ ขั้นตอนวิธีการลบสเปกตรัมแบบการประมาณแมกซิมัมไลค์ลิฮูดและกรณีไม่ใช้ขั้นตอนวิธีใดเลย เมื่อใช้ชุดข้อมูลทดสอบประโยคภาษาไทยทั่วไป...35

รูปที่ 4.10 กราฟแสดงผลการทดลองเปลี่ยนค่าถ่วงน้ำหนักต่อความแม่นยำในการรู้จำเสียงพูดกรณีสัญญาณเสียงพูดประโยคทั่วไป.....37

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญตาราง

หน้า

ตารางที่ 4.1 รายชื่อหน่วยเสียงจับคู่กับพยัญชนะต้นในภาษาไทย [10] 23
ตารางที่ 4.2 รายชื่อหน่วยเสียงจับคู่กับสระในภาษาไทย [10] 24
ตารางที่ 4.3 รายชื่อหน่วยเสียงจับคู่กับตัวสะกดในภาษาไทย [10] 25
ตารางที่ 4.4 ความแม่นยำในการรู้จำเสียงพูดตัวเลขที่ค่าอัตราส่วนสัญญาณต่อสัญญาณรบกวนที่ค่าถ่วงน้ำหนักต่างๆ 26
ตารางที่ 4.5 ค่าความแม่นยำในการรู้จำเสียงพูดตัวเลขที่ค่าอัตราส่วนสัญญาณต่อสัญญาณรบกวนที่ค่าขีดแบ่งค่าเฉลี่ยพลังงานสัญญาณรบกวนค่าต่างๆเมื่อใช้ฟังก์ชันขั้นบันได 29
ตารางที่ 4.6 ค่าความแม่นยำในการรู้จำเสียงพูดตัวเลขที่ค่าอัตราส่วนสัญญาณต่อสัญญาณรบกวนที่ค่าขีดแบ่งค่าเฉลี่ยพลังงานสัญญาณรบกวนค่าต่างๆเมื่อใช้ฟังก์ชันเชิงเส้น 32
ตารางที่ 4.7 ค่าความแม่นยำในการรู้จำเสียงพูดระหว่างขั้นตอนวิธีการลบสเปกตรัมแบบปรับค่าถ่วงน้ำหนักได้ ขั้นตอนวิธีการลบสเปกตรัมแบบการประมาณแมกซิมัมไลค์ลิฮูดและกรณีไม่ใช้ขั้นตอนวิธีใดเลย เมื่อใช้ชุดข้อมูลทดสอบตัวเลขภาษาไทย 34
ตารางที่ 4.8 ค่าความแม่นยำในการรู้จำเสียงพูดระหว่างขั้นตอนวิธีการลบสเปกตรัมแบบปรับค่าถ่วงน้ำหนักได้ ขั้นตอนวิธีการลบสเปกตรัมแบบการประมาณแมกซิมัมไลค์ลิฮูดและกรณีไม่ใช้ขั้นตอนวิธีใดเลย เมื่อใช้ชุดข้อมูลทดสอบประโยคภาษาไทยทั่วไป 35
ตารางที่ 4.9 ความแม่นยำในการรู้จำเสียงพูดตัวเลขที่ค่าอัตราส่วนสัญญาณต่อสัญญาณรบกวนที่ค่าถ่วงน้ำหนักต่างๆสำหรับชุดข้อมูลประโยคภาษาไทยทั่วไป 36

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 1

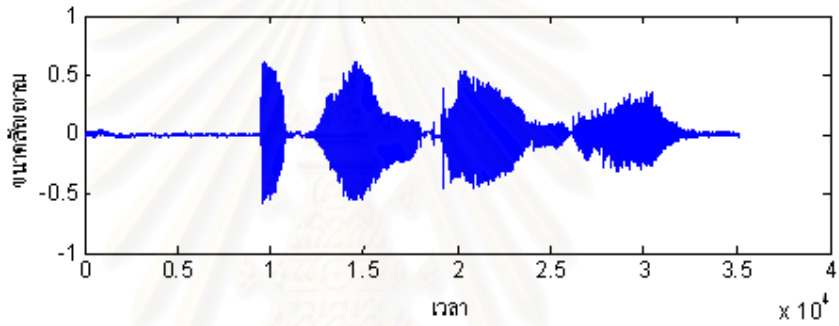
บทนำ

ความเป็นมาและความสำคัญของปัญหา

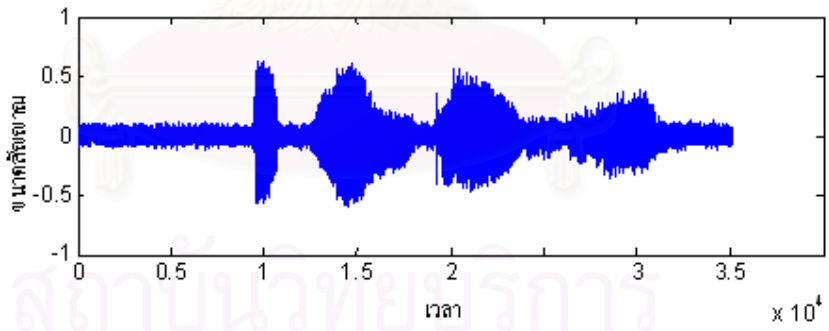
ปัจจุบันเทคโนโลยีทางด้านระบบรู้จำเสียงพูด (Speech Recognition System) ได้เริ่มถูกประยุกต์เพื่อนำมาใช้งานในชีวิตประจำวัน เช่น การเปิดปิดไฟในบ้านผ่านเสียงพูด การเปิดโปรแกรมในโทรศัพท์ผ่านเสียงพูด การรู้จำชื่อเพลงผ่านการอัดเสียงเพลงด้วยโปรแกรมรู้จำ เป็นต้น ดังนั้นการศึกษาวิจัยทางด้านการรู้จำเสียงพูดจึงได้เกิดขึ้นอย่างกว้างขวาง นักวิจัยได้ศึกษาค้นคว้าหาขั้นตอนวิธีต่างๆที่มีประสิทธิภาพอย่างต่อเนื่อง เพื่อนำมาพัฒนาปรับปรุงประสิทธิภาพในการรู้จำเสียงพูดให้มีความถูกต้องแม่นยำมากขึ้น ขั้นตอนวิธีหรือเทคนิคต่างๆสามารถเพิ่มประสิทธิภาพในการรู้จำได้หลายรูปแบบ เช่น เพื่อเพิ่มประสิทธิภาพให้กับระบบรู้จำเสียงพูดให้มีการเรียนรู้ที่ดียิ่งขึ้น หรือเพื่อให้ระบบรู้จำสามารถรู้จำเสียงพูดได้ดียิ่งขึ้น โดยกำจัดสัญญาณรบกวนในสัญญาณเสียงพูดให้ได้มากที่สุดก่อนผ่านกระบวนในระบบรู้จำเสียงพูด

จากการศึกษางานวิจัยที่ผ่านมาพบว่า งานวิจัยหลักๆสามารถแบ่งได้เป็นสองประเภท ได้แก่ งานวิจัยในส่วนของระบบรู้จำ (Recognizer) และงานวิจัยในส่วนของกระบวนการก่อนหน้า (Preprocessing) โดยงานวิจัยในประเภทที่สองจะมุ่งเน้นในการลดสัญญาณรบกวนออกจากสัญญาณเสียงพูด เพื่อเพิ่มประสิทธิภาพในการรู้จำเสียงพูดให้ดียิ่งขึ้น ในระบบรู้จำเสียงพูดนั้นเมื่อนำสัญญาณเสียงพูดสะอาด(แสดงในรูป 1.1) มาทำการรู้จำผ่านระบบรู้จำเสียงพูด ระบบจะทำการรู้จำเสียงพูดได้ดีกว่า กรณีนำการรู้จำด้วยสัญญาณเสียงพูดที่มีสัญญาณรบกวน(แสดงในรูปที่ 1.2) โดยทั่วไปแล้วสัญญาณรบกวนหมายถึงสัญญาณเสียงไม่พึงประสงค์ในระบบต่างๆ ที่เกิดขึ้นในสัญญาณเสียงที่นำมาทำการรู้จำ ปกติแล้วสัญญาณรบกวนนั้นสามารถเกิดขึ้นได้ทั่วไป เช่น เสียงรถยนต์ตามท้องถนน เสียงนักเรียนในห้องเรียนคุยกัน เสียงเพลงจากวิทยุ เป็นต้น ปัจจุบันขั้นตอนวิธีหรือเทคนิคที่ใช้ในการลดสัญญาณรบกวนได้ถูกศึกษาและพัฒนาอย่างมาก อาทิเช่น การลบสเปกตรัม (Spectral Subtraction) [1][2] การกรองวีเนอร์ (Wiener Filtering) [1] ซึ่งเป็นระเบียบวิธีที่มุ่งเน้นในการลดสัญญาณรบกวนเติมแต่ง (Additive Noise) ซึ่งเป็นสัญญาณรบกวนที่เกิดขึ้นจากแหล่งกำเนิดเสียงที่ไม่ต้องการ ขณะเดียวกันนักวิจัยยังได้ทำการศึกษาสัญญาณรบกวนอีกประเภทหนึ่งนั่นคือสัญญาณรบกวนสังวัตนาการ (Convolution Noise) ซึ่งเกิดขึ้นจากช่องสัญญาณที่ทำให้สัญญาณเสียงบิดเบือน ขั้นตอนวิธีที่สามารถลดทอนสัญญาณดังกล่าวได้แก่ การลบค่าเฉลี่ยเคปสตรัม (Cepstral Mean Subtraction) [1] เป็นต้น

กระบวนการรู้จำเสียงพูดนั้นประกอบไปด้วยขั้นตอนต่างๆหลายขั้นตอน เช่น การตัดแบ่งข้อมูลเป็นหน่วยเสียง (Phoneme Segmentation) คือ การแบ่งหน่วยเสียงพูดให้เป็นพยัญชนะต้น ตัวสะกดและสระ การหาลักษณะสำคัญของเสียง (Feature Extraction) คือการหาค่าสัมประสิทธิ์ในโดเมนความถี่ในสัญญาณเสียงพูดที่ถูกตัดแบ่งแล้ว โดยค่าดังกล่าวจะถูกนำมาใช้เป็นตัวแทนสัญญาณเสียง(Speech Representation) การตรวจหากิจกรรมเสียงพูด (Voice Activity Detection) คือ การแยกแยะสัญญาณเสียงพูดระหว่างส่วนที่เป็นสัญญาณเสียงพูดและส่วนที่ไม่ใช่สัญญาณเสียงพูด การลดสัญญาณรบกวน (Noise Reduction) และระบบในการรู้จำเสียงพูด ซึ่งในงานวิจัยนี้ใช้ระบบการรู้จำเสียงที่ใช้แบบจำลองฮิดเดนมาร์คอฟ (Hidden Markov Model, HMM)



รูปที่ 1.1 สัญญาณเสียงพูดสะอาด



รูปที่ 1.2 สัญญาณเสียงพูดที่มีสัญญาณรบกวน

วิทยานิพนธ์นี้นำเสนอการพัฒนาขั้นตอนวิธี ที่ใช้ในการลดสัญญาณรบกวนจากสัญญาณเสียงพูดก่อนที่สัญญาณจะถูกนำไปวิเคราะห์เพื่อหาค่าความแม่นยำ โดยเปรียบเทียบและวิเคราะห์ความแม่นยำในการรู้จำเสียงพูดระหว่างขั้นตอนวิธีที่ถูกนำเสนอและขั้นตอนวิธีเดิม ได้แก่ การลบสเปกตรัมแบบการประมาณแม็กซิมัมไลค์ลิฮูด การลบสเปกตรัมแบบตัวกรองลดสัญญาณ การกรองวีเนอร์ ทั้งนี้เพื่อเป็นแนวทางในการพัฒนาเครื่องมือเพื่อเพิ่มประสิทธิภาพในการรู้จำเสียงพูดต่อไป

วัตถุประสงค์ของการวิจัย

งานวิจัยนี้มีจุดประสงค์เพื่อนำเสนอการพัฒนาขั้นตอนวิธี ที่ใช้ในการลดสัญญาณรบกวนเพื่อเพิ่มความแม่นยำในการรู้จำเสียงพูดให้มีประสิทธิภาพดียิ่งขึ้น งานวิจัยนี้นำเสนอรายละเอียดของขั้นตอนวิธีและทฤษฎีต่างๆที่เกี่ยวข้อง และยังนำเสนอการเปรียบเทียบประสิทธิภาพของขั้นตอนวิธีต่างๆอีกด้วย จากนั้นทำการวิเคราะห์ขั้นตอนวิธีที่พิจารณาและนำเสนอแนวคิดในการพัฒนาขั้นตอนวิธีดังกล่าว เพื่อเพิ่มประสิทธิภาพในการรู้จำเสียงพูดให้มีความแม่นยำมากขึ้น โดยงานวิจัยนี้ ใช้ข้อมูลชุดตัวเลขภาษาไทยและชุดข้อมูลประโยคภาษาไทยทั่วไปในการสอนระบบและทำการทดสอบค่าความแม่นยำโดยใช้ระบบการรู้จำเสียงพูดที่ใช้แบบจำลองฮิดเดนมาร์คอฟ อีกทั้งนำเสนอการศึกษาแนวโน้มการเปลี่ยนแปลงของค่าความแม่นยำในการรู้จำเสียงพูด เมื่อทำการเปลี่ยนค่าถ่วงน้ำหนักที่ค่าต่างๆ ซึ่งเป็นแนวคิดเริ่มต้นของขั้นตอนวิธีการลบสเปกตรัมแบบปรับค่าถ่วงน้ำหนักได้

ขอบเขตของวิทยานิพนธ์

1. ศึกษาและออกแบบขั้นตอนการดำเนินการวิจัย เพื่อลดทอนสัญญาณรบกวนเติมแต่งในสัญญาณเสียงพูดโดยมีจุดประสงค์เพื่อเพิ่มความแม่นยำในการรู้จำเสียงพูดขั้นตอนการดำเนินการวิจัย
2. จัดเตรียมข้อมูลเสียงพูดตัวเลขภาษาไทยจากผู้ชายจำนวน 13 คนและผู้หญิง 13 คน จำนวนมากกว่า 800 ประโยค เพื่อใช้ในการสอนระบบและทดสอบความแม่นยำในการรู้จำเสียงพูด
3. จัดเตรียมข้อมูลเสียงพูดประโยคภาษาไทยทั่วไป จำนวนมากกว่า 1,000 ประโยคในการสอนระบบให้ทำการรู้จำเสียงพูดและทดสอบความแม่นยำในการรู้จำเสียงพูด
4. พัฒนาขั้นตอนวิธีที่สามารถทำการลดสัญญาณรบกวนเพื่อเพิ่มประสิทธิภาพในการรู้จำเสียงพูดได้
5. เปรียบเทียบความแม่นยำในการรู้จำเสียงพูดระหว่างขั้นตอนวิธีใหม่ที่ถูกพัฒนากับขั้นตอนวิธีเดิมโดยใช้ระบบรู้จำเสียงพูดที่ใช้แบบจำลองฮิดเดนมาร์คอฟ

วิธีดำเนินการวิจัย

1. ศึกษารูปแบบและทฤษฎีที่เกี่ยวข้องกับการรู้จำเสียงพูด เพื่อให้ทราบถึงองค์ประกอบที่สำคัญในการเพิ่มประสิทธิภาพในการรู้จำเสียงพูด โดยในที่นี้ทำการศึกษาขั้นตอนวิธีที่สามารถลดทอนสัญญาณรบกวนในสัญญาณเสียงพูดได้

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

2. ออกแบบวิธีการทดลอง
3. จัดเตรียมข้อมูลสัญญาณเสียงพูดเพื่อนำมาใช้ในการทดลอง
4. ทำการทดลองโดยการเปรียบเทียบประสิทธิภาพของขั้นตอนวิธีที่นำมาใช้ในการลดสัญญาณรบกวนในสัญญาณเสียงพูด
5. วิเคราะห์ผลการทดลองและสรุปผลการทดลองที่ได้
6. สรุปผลการวิจัยและจัดทำรายงานวิทยานิพนธ์

ประโยชน์ที่คาดว่าจะได้รับ

สามารถพัฒนาขั้นตอนวิธีใหม่ ซึ่งสามารถลดสัญญาณรบกวนและเพิ่มความแม่นยำในการรู้จำเสียงพูดได้โดยมีประสิทธิภาพในการลดสัญญาณรบกวนดีกว่าขั้นตอนวิธีเดิม ทั้งนี้เพื่อให้ได้เครื่องมือที่สามารถนำมาใช้ในการลดทอนสัญญาณรบกวนในสัญญาณเสียงพูดได้ อีกทั้งยังเป็นแนวทางในการพัฒนาเครื่องมือเพื่อลดทอนสัญญาณรบกวน และเพิ่มความแม่นยำในการรู้จำสัญญาณเสียงพูดต่อไป พร้อมทั้งสามารถทราบข้อดีข้อเสียของขั้นตอนวิธีที่ใช้ในการลดสัญญาณรบกวนจากการเปรียบเทียบความแม่นยำในการรู้จำเสียงพูด

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

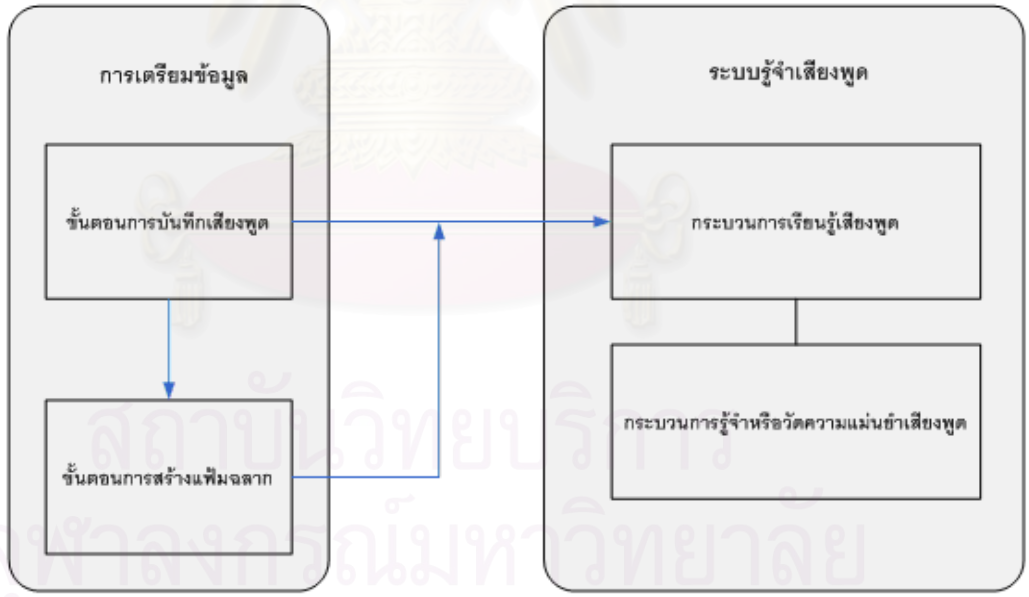
บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ทฤษฎีที่เกี่ยวข้อง

ในกระบวนการรู้จำเสียงพูดนั้นประกอบไปด้วยขั้นตอนต่างๆหลายขั้นตอน ได้แก่ ขั้นตอนการเตรียมข้อมูล ขั้นตอนกระบวนการก่อนหน้าเพื่อลดสัญญาณรบกวนและขั้นตอนของระบบรู้จำเสียงพูด

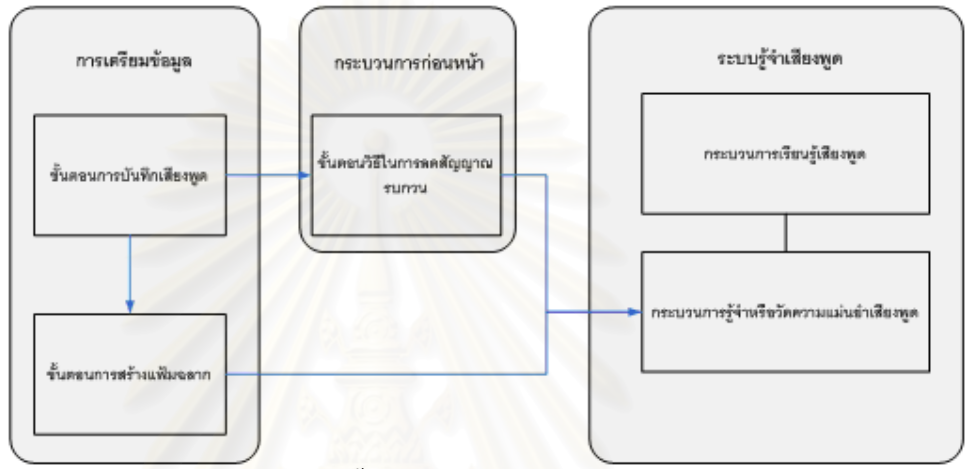
การลดผลกระทบจากสัญญาณรบกวนนั้นเริ่มต้นจากการเตรียมข้อมูลเสียงพูดเพื่อใช้ในการสอนระบบดังแสดงในรูปที่ 2.1 โดยข้อมูลเสียงพูดที่ใช้ในการสอนระบบต้องเป็นเสียงพูดที่สะอาดเพื่อให้ระบบสามารถทำการรู้จำได้อย่างถูกต้องมากที่สุด นอกจากนี้ต้องทำการเตรียมแฟ้มฉลาก (Label File) ซึ่งระบุคำพูดที่ถูกต้องของสัญญาณเสียงพูดเพื่อให้ระบบใช้ในการรู้จำอีกด้วย



รูปที่ 2.1 ขั้นตอนในกระบวนการเรียนรู้เสียงพูด

ในกรณีการรู้จำหรือวัดความแม่นยำเสียงพูดดังแสดงในรูปที่ 2.2 นั้น สัญญาณเสียงพูดดังกล่าว จะถูกประมวลผลด้วยกระบวนการก่อนหน้าเพื่อลดสัญญาณรบกวนด้วยขั้นตอนวิธีต่างๆ โดยมีจุดประสงค์ เพื่อให้สัญญาณเสียงพูดที่บันทึกนั้นสะอาดมากที่สุดเพื่อเพิ่มความแม่นยำในการรู้จำ

เสียงพูด ก่อนเข้าสู่ในกระบวนการถัดไป หลังจากขั้นตอนข้างต้นสัญญาณเสียงพูดที่ได้จึงผ่านเข้าสู่ระบบรู้จำเสียงพูดพร้อมกับแฟ้มฉลากที่เตรียมไว้ในขั้นตอนแรก ในระบบรู้จำเสียงพูดนั้นสัญญาณเสียงพูดจะถูกประมวลผลเพื่อใช้ในการเรียนรู้หรือทดสอบความความแม่นยำโดยใช้ข้อมูลจากแฟ้มฉลากเป็นสิ่งอ้างอิง ในงานวิจัยนี้จะใช้ข้อมูลเสียงพูดสองประเภทคือ ข้อมูลเสียงพูดตัวเลขและข้อมูลเสียงพูดประโยคทั่วไป



รูปที่ 2.2 ขั้นตอนในกระบวนการรู้จำเสียงพูด

งานวิจัยนี้มุ่งเน้นการพัฒนาขั้นตอนวิธีที่ใช้ในการลดสัญญาณรบกวนในสัญญาณเสียงพูดเพื่อเพิ่มประสิทธิภาพในการรู้จำให้ดียิ่งขึ้น หลักการและทฤษฎีที่เกี่ยวข้องได้แก่

1. หลักการและทฤษฎีที่เกี่ยวข้องกับขั้นตอนวิธีในการลดสัญญาณรบกวน

โดยทั่วไปสัญญาณเสียงพูด($y(n)$) จะประกอบด้วยสัญญาณเสียงพูดสะอาด ($x(n)$) และสัญญาณรบกวนพื้นหลัง (Background Noise) ($b(n)$) ที่เราไม่ต้องการอยู่ในสัญญาณดังกล่าว โดยสามารถเขียนเป็นสมการได้ดังสมการที่ 2.1

$$y(n) = x(n) + b(n) \tag{2.1}$$

จากสมการข้างต้น เมื่อพิจารณาสัญญาณให้อยู่ในแกนความถี่โดยใช้การแปลงฟูเรียร์แบบช่วงเวลาสั้น (Short-Time Fourier Transform, STFT) เราสามารถเขียนสมการใหม่ได้ดังนี้

$$y_{pL}[n] = w(pL-n)[x(n) + b(n)] \tag{2.2}$$

$$Y(pL,\omega) \quad = \quad X(pL,\omega) + B(pL,\omega) \tag{2.3}$$

โดยที่ $Y(pL,\omega)$ คือ การแปลงฟูเรียร์แบบช่วงเวลาสั้นของสัญญาณเสียงพูด

$X(pL,\omega)$ คือ การแปลงฟูเรียร์แบบช่วงเวลาสั้นของสัญญาณเสียงพูดสะอาด

$B(pL,\omega)$ คือ การแปลงฟูเรียร์แบบช่วงเวลาสั้นของสัญญาณรบกวนพื้นหลัง

p คือ ลำดับของกรอบ

L คือ ความยาวกรอบ

โดยทั่วไปเราสามารถตั้งสมมติฐานได้ว่าสำหรับสัญญาณเสียงที่เราทราบได้นั้น ในช่วงเวลา 0 – 100 มิลลิวินาทีแรกของสัญญาณเสียงนั้นเป็นช่วงที่ไม่มีสัญญาณเสียงพูดอยู่ ดังนั้นเราสามารถทำการประมาณสัญญาณรบกวนพื้นหลังได้จากสัญญาณเสียงที่เรามีในช่วงดังกล่าว

1.1 การลบสเปกตรัม

จากข้อมูลข้างต้นทำให้เราทราบได้ว่า สัญญาณเสียงพูดที่เราทราบนั้นจะเป็นสัญญาณเสียงที่ประกอบด้วยสัญญาณเสียงพูดสะอาดและสัญญาณรบกวน ดังนั้นภายใต้สมมติฐานที่ว่าสัญญาณรบกวนพื้นหลังและสัญญาณเสียงสะอาดไม่มีสหสัมพันธ์ร่วมกัน เราจะสามารถประมาณสัญญาณเสียงพูดสะอาดได้ดังนี้ [1]

$$|\hat{X}(pL,\omega)| \begin{cases} = |Y(pL,\omega)|^2 - \hat{S}_b(\omega) & if \ |Y(pL,\omega)|^2 \geq 0 \\ = 0 & otherwise \end{cases} \tag{2.4}$$

$$\hat{X}(pL,\omega) = |\hat{X}(pL,\omega)| e^{j\angle Y(pL,\omega)} \tag{2.5}$$

โดยที่ $\hat{X}(pL,\omega)$ คือ การแปลงฟูเรียร์แบบช่วงเวลาสั้นของสัญญาณเสียงพูดที่ต้องการทราบค่า

$\hat{S}_b(\omega)$ คือ ค่าเฉลี่ยสเปกตรัมกำลังของสัญญาณรบกวนที่ประมาณขึ้น

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

นอกจากนี้การลบสเปกตรัมยังสามารถนำมาประยุกต์ใช้เพื่อเป็นตัวกรองลดสัญญาณรบกวน (Suppression Filter) ได้ดังนี้

$$
\begin{aligned}
|\hat{X}(pL,\omega)|^2 &= |Y(pL,\omega)|^2 - \hat{S}_b(\omega) \\
&= |Y(pL,\omega)|^2 \left[1 - \frac{\hat{S}_b(\omega)}{|Y(pL,\omega)|^2}\right] \\
&\approx |Y(pL,\omega)|^2 \left[1 + \frac{1}{R(pL,\omega)}\right]^{-1}
\end{aligned}
$$

เมื่อกำหนดให้

$$R(pL,\omega) = \frac{|X(pL,\omega)|^2}{\hat{S}_b(\omega)} \tag{2.6}$$

เมื่อใช้การประมาณที่ $|Y(pL,\omega)|^2 \approx |X(pL,\omega)|^2 + \hat{S}_b(\omega)$ ดังนั้นเราสามารถประมาณสัญญาณเสียงพูดสะอาดได้โดยการใช้ตัวกรองลดทอนสัญญาณรบกวนได้ดังนี้

$$\hat{X}(pL,\omega) = Y(pL,\omega)H_s(pL,\omega) \tag{2.7}$$

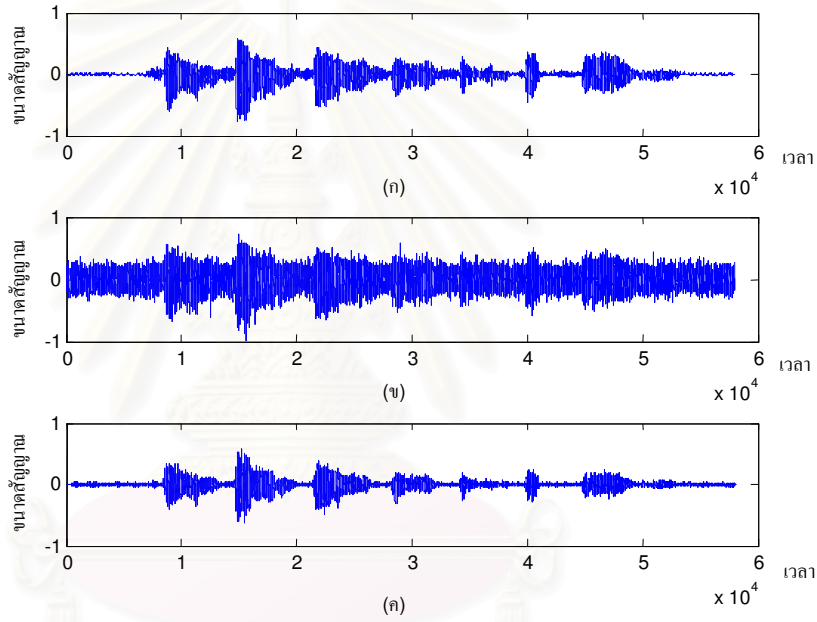
โดยที่

$$H_s(pL,\omega) = \left[1 + \frac{1}{R(pL,\omega)}\right]^{-1/2} \tag{2.8}$$

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

นอกจากการประยุกต์การลบสเปกตรัมในรูปแบบของตัวกรองแล้ว ปัจจุบันการประมาณการแปลงฟูเรียร์แบบช่วงเวลาสั้นยังคงถูกศึกษาและพัฒนาอย่างต่อเนื่อง ซึ่งวิธีดังกล่าวสามารถนำมาประยุกต์ใช้กับการลบสเปกตรัมได้ วิธีที่นำมาใช้ได้แก่ การประมาณแบบแม็กซิมัมไลค์ลิฮูด (Maximum Likelihood Estimation) จากวิธีดังกล่าวทำให้เราสามารถประมาณสัญญาณเสียงพูดสะอาดได้จากสัญญาณเสียงพูดที่เราทราบได้จากสมการที่ 2.9 [1] [2]

$$|\hat{X}(pL,\omega)| = \frac{1}{2}|Y(pL,\omega)| + \frac{1}{2}[|Y(pL,\omega)|^2 - \hat{S}_b(\omega)]^{1/2} \qquad (2.9)$$



รูปที่ 2.3 (ก) สัญญาณเสียงพูดสะอาด (ข) สัญญาณเสียงพูดที่ตรวจหาได้ และ (ค) สัญญาณเสียงพูดที่ถูกลดสัญญาณรบกวนด้วยขั้นตอนวิธีการลบสเปกตรัมแบบการประมาณแม็กซิมัมไลค์ลิฮูด

จากรูปที่2.3 แสดงถึงการผลการประมาณสัญญาณเสียงพูดสะอาดที่ได้จากการลดสัญญาณรบกวนออกจากสัญญาณเสียงพูดที่ตรวจหาได้ สามารถสังเกตได้ว่าสัญญาณรบกวนถูกลดทั้งบริเวณที่เป็นสัญญาณเสียงพูดและไม่ใช่สัญญาณเสียงพูด ซึ่งโดยทั่วไปแล้วการลดสัญญาณรบกวนที่มากขึ้นจะทำให้สัญญาณเสียงพูดที่พิจารณาเกิดการบิดเบือนมากขึ้นตามไปด้วย ซึ่งทำให้ระบบสามารถรู้จำเสียงพูดได้แย่ลง [2]

1.2 การกรองวีเนอร์

ขั้นตอนวิธีนี้นำเสนอแนวคิดการใช้ตัวกรองเชิงเส้น $h(n)$ เมื่อ $\hat{x}(n) = y(n) * h(n)$ โดยใช้ตัวกรองในการลดสัญญาณรบกวนจากสัญญาณเสียงพูดที่ตรวจจับได้ ภายใต้เงื่อนไขสัญญาณเสียงสะอาด $x(n)$ และสัญญาณรบกวน $b(n)$ ที่ไม่มีสหสัมพันธ์ร่วมกันและสัญญาณรบกวนที่พิจารณาเป็นสัญญาณรบกวนแบบคงที่ (Stationary Noise) ดังนั้นเมื่อเราพิจารณาสัญญาณให้อยู่ในแกนความถี่ เราสามารถเขียนสมการได้เป็น [1]

$$H_s(pL,\omega) = \frac{\hat{S}_x(pL,\omega)}{\hat{S}_x(pL,\omega) + \hat{S}_b(\omega)} \tag{2.10}$$

โดยที่ $\hat{S}_x(pL,\omega)$ คือ ค่าพลังงานสเปกตรัมของสัญญาณเสียงพูดที่เราทราบค่า ดังนั้นเราสามารถเขียนสมการข้างต้นได้ใหม่เป็น

$$H_s(pL,\omega) = \left[1 + \frac{1}{R(pL,\omega)}\right]^{-1} \tag{2.11}$$

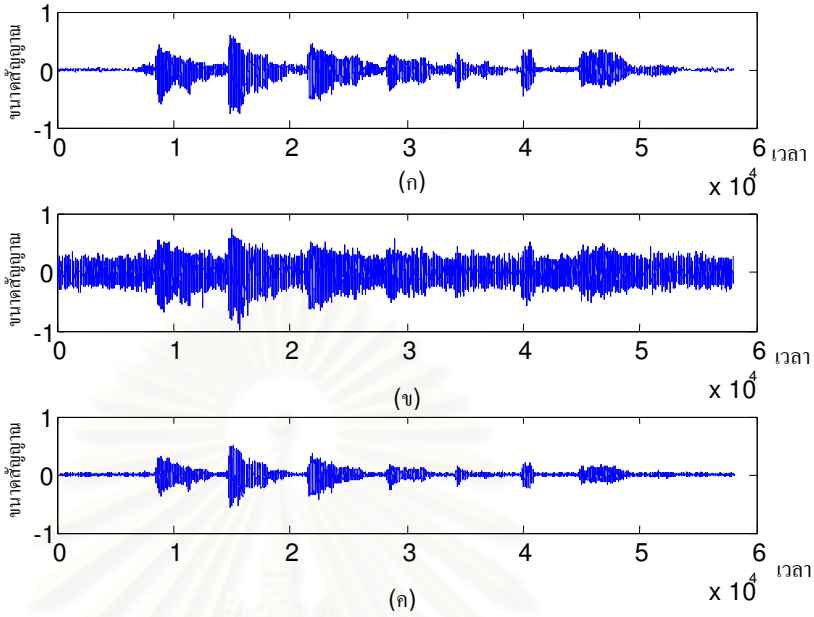
เมื่อ

$$R(pL,\omega) = \frac{\hat{S}_x(pL,\omega)}{\hat{S}_b(\omega)} \tag{2.12}$$

ดังนั้นเราสามารถหาสัญญาณเสียงพูดที่ต้องการทราบได้จาก สมการดังต่อไปนี้

$$\hat{X}(pL,\omega) = Y(pL,\omega)H_s(pL,\omega)$$

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 2.4 (ก) สัญญาณเสียงพูดสะอาด (ข) สัญญาณเสียงพูดที่เราทราบได้และ (ค) สัญญาณเสียงพูดที่ถูกลดสัญญาณรบกวนด้วยขั้นตอนวิธีการกรองวีเนอร์

2. หลักการและทฤษฎีทั่วไปที่เกี่ยวข้อง

การลดสัญญาณรบกวนในสัญญาณเสียงพูดที่พิจารณานั้น งานวิจัยโดยทั่วไปจะทำการวัดค่าพลังงานสัญญาณรบกวนเพื่อใช้เป็นค่าบ่งชี้ขนาดหรือปริมาณสัญญาณรบกวน และใช้ค่าดังกล่าวเป็นตัวแปรเสริมในกระบวนการลดสัญญาณรบกวน และทำการวัดค่าความแม่นยำในการรู้จำเสียงพูดของสัญญาณเสียงที่ผ่านกระบวนการก่อนหน้าด้วยขั้นตอนวิธีต่างๆ ที่ค่าอัตราส่วนต่อสัญญาณรบกวนค่าต่างๆกัน เพื่อศึกษาถึงประสิทธิภาพในการลดสัญญาณรบกวนที่ขนาดสัญญาณรบกวนต่างๆ โดยในการวัดค่าความแม่นยำในการรู้จำเสียงพูดนั้น ระบบจะทำการวัด

จากที่กล่าวมาเบื้องต้นหลักการและทฤษฎีที่เกี่ยวข้องได้แก่

2.1 ค่าพลังงาน (Energy Value)

โดยทั่วไปพลังงานของสัญญาณเสียง จะถูกพิจารณาจากการแบ่งสัญญาณเสียงออกเป็นช่วงเล็กๆตามแกนเวลาหรือเรียกว่าแบ่งออกเป็นกรอบ เช่น กรอบละประมาณ 10 - 50 มิลลิวินาที เราสามารถหาพลังงานของเสียงแต่ละกรอบได้เป็น

$$E(m) = \sum_{n=0}^{N-1} [w(m)s(m-n)]^2 \qquad (2.13)$$

เมื่อ $E(m)$ คือพลังงานของสัญญาณเสียง ณ กรอบที่พิจารณา

โดยที่ $w(m)$ หมายถึงวินโดว์พังก์ชันที่ใช้ในการกำหนดกรอบที่พิจารณาสัญญาณเสียงพูด $s(n)$ และ N หมายถึงจำนวนตัวอย่างของสัญญาณเสียงในหนึ่งกรอบ

2.2 อัตราส่วนสัญญาณต่อสัญญาณรบกวน (Signal-to-Noise Ratio, SNR)

ค่าอัตราส่วนสัญญาณต่อสัญญาณรบกวนเป็นตัวแปรเสริม ที่ใช้ในการวัดปริมาณสัญญาณรบกวนที่มีอยู่ในสัญญาณ ในกรณีที่สัญญาณมีค่าอัตราส่วนสัญญาณต่อสัญญาณรบกวนสูงนั่นหมายถึงสัญญาณที่มีปริมาณสัญญาณรบกวนอยู่ น้อยกว่ากรณีที่สัญญาณมีค่าอัตราส่วนสัญญาณต่อสัญญาณรบกวนต่ำ

ค่าอัตราส่วนสัญญาณต่อสัญญาณรบกวนสามารถคำนวณหาได้จากสมการดังนี้

$$SNR(db) = 10\log_{10}(\frac{S}{N}) \tag{2.14}$$

เมื่อ S คือ พลังงานของสัญญาณเสียงสะอาด

N คือ พลังงานของสัญญาณรบกวน

2.3 การวัดความแม่นยำในการรู้จำเสียงพูด

ในการวัดค่าความแม่นยำของการรู้จำนั้น ผู้ทดสอบจะต้องทำการระบุแฟ้มฉลากซึ่งระบุรายชื่อคำในสัญญาณเสียงพูดนั้นๆ ทั้งนี้เพื่อใช้เป็นสิ่งอ้างอิงและเปรียบเทียบกับคำที่ระบบรู้จำได้จากสัญญาณเสียงพูด เราสามารถคำนวณค่าความแม่นยำได้จากสมการดังนี้

$$Accuracy\% = \frac{H - I}{N} \times 100\% \tag{2.15}$$

เมื่อ H คือ จำนวนคำที่ถูกรู้จำถูกต้องตรงกับรายชื่อคำในไฟล์ฉลาก

I คือ จำนวนคำที่ถูกรู้จำผิดหรือไม่มีอยู่ในรายชื่อคำในไฟล์ฉลาก

N คือ จำนวนคำทั้งหมดที่ถูกระบุอยู่ในไฟล์ฉลาก

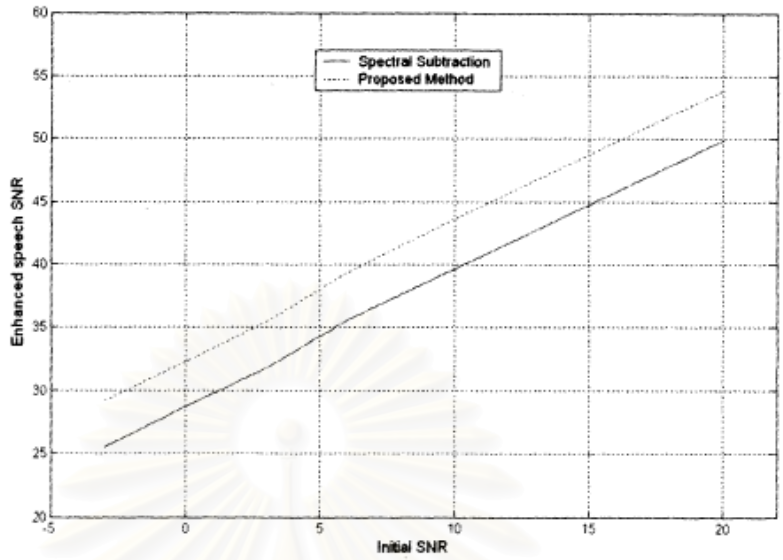
โดยทั่วไปแล้วในการวัดค่าความแม่นยำ ความผิดพลาด (Error) เกิดขึ้นได้จากกรณีอื่นๆอีก ได้แก่ การแทนที่ (Substitution) หมายถึงการที่ระบบทำการรู้จำคำที่ไม่ตรงกับแฟ้มฉลากในตำแหน่งต่างๆ และการหลุดหาย (Deletion) หมายถึงการที่ระบบไม่สามารถรู้จำคำที่ต้องได้

งานวิจัยที่เกี่ยวข้อง

จากการศึกษางานวิจัยที่ผ่านมาพบว่า นักวิจัยจำนวนมากได้ทำการศึกษาและพัฒนาขั้นตอนวิธีต่างๆในการลดสัญญาณรบกวนออกจากสัญญาณเสียงพูดทั้งกรณีสัญญาณรบกวนคงที่และกรณีสัญญาณรบกวนไม่คงที่ โดยขั้นตอนในส่วนการประเมินผลจะใช้วิธีการเปรียบเทียบผลความแม่นยำในการรู้จำกับวิธีพื้นฐาน (Baseline) ซึ่งหมายถึงขั้นตอนวิธีที่นักวิจัยใช้ในการเปรียบเทียบกับขั้นตอนวิธีของตนเองที่นำเสนอ โดยงานวิจัยนี้ใช้ขั้นตอนวิธีการประมาณแบบแม็กซิมัมไลค์ลิฮูดเป็นวิธีพื้นฐาน และทำการเปรียบเทียบที่สัญญาณเสียงที่มีค่าอัตราส่วนสัญญาณต่อสัญญาณรบกวนต่างๆกัน โดยทั่วไปแล้วจะใช้ค่าอัตราส่วนสัญญาณต่อสัญญาณรบกวนที่มีค่าอยู่ระหว่าง 30 ถึง -10 เดซิเบล [3]

โดยตัวอย่างขั้นตอนวิธีแบบสัญญาณรบกวนแบบคงที่ ได้แก่ Krishnamoorthy และ Prasanna [4] เสนอการประยุกต์ใช้วิธีการตรวจหากิจกรรมเสียงพูดและขั้นตอนวิธีการลบสเปกตรัม โดยในส่วนของการตรวจหากิจกรรมเสียงพูดนั้น นักวิจัยได้นำเสนอการใช้วิธีภาวะยอดมน (Kurtosis) และวิธีค่าพลังงาน (Energy Value) ในการตรวจหากิจกรรมเสียงพูดซึ่งวิธีภาวะยอดมนเป็นวิธีที่มีประสิทธิภาพและเหมาะสม ในกรณีสัญญาณเสียงพูดนั้นมีค่าอัตราส่วนสัญญาณต่อสัญญาณรบกวนต่ำ ส่วนวิธีค่าพลังงานเป็นวิธีที่เหมาะสมในกรณีสัญญาณมีค่าอัตราส่วนสัญญาณต่อสัญญาณรบกวนสูง ดังนั้นการรวมข้อดีของทั้งสองวิธีมาใช้ร่วมโดยพิจารณากันจึงได้ผลที่ดีกว่า จากนั้นนำค่าที่ได้จากวิธีทั้งสองมาทำการหาค่าถ่วงน้ำหนัก เพื่อใช้ถ่วงน้ำหนักสัญญาณเสียงพูดก่อนที่จะผ่านขั้นตอนวิธีการลบสเปกตรัมต่อไป และจากผลการทดลองที่ได้พบว่าการใช้วิธีตรวจหากิจกรรมเสียงเพื่อหาค่าถ่วงน้ำหนักสามารถช่วยลดสัญญาณรบกวนได้ดีขึ้นดังแสดงในรูปที่ 2.5

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 2.5 การเปรียบเทียบค่าอัตราส่วนสัญญาณต่อสัญญาณรบกวนระหว่างขั้นตอนวิธีการลบสเปกตรัมและวิธีที่นำเสนอใน[4]

จะเห็นได้ว่างานวิจัยนี้ นำวิธีการตรวจหากิจกรรมเสียงมาช่วยในการเพิ่มประสิทธิภาพของขั้นตอนวิธีการลบสเปกตรัม ซึ่งวิธีตรวจหากิจกรรมเสียงยังคงได้รับการพัฒนาอย่างต่อเนื่อง Harsha [5] ได้ประยุกต์ใช้วิธีค่าพลังงานและอัตราการข้ามผ่านศูนย์ (Zero-Crossing Rate) เพื่อเพิ่มประสิทธิภาพในการตรวจหากิจกรรมเสียง โดยในงานวิจัยดังกล่าวประเมินผลการทดลองด้วยการวัดอัตราส่วนค่าความถูกต้องในการตรวจหาส่วนที่เป็นเสียงพูดและไม่ใช่เสียงพูด

ส่วนในกรณีขั้นตอนวิธีที่เกี่ยวข้องกับสัญญาณรบกวนไม่คงที่ เช่น Denda, Nishiura, Kawahara และ Irino [6] เนื่องด้วยขั้นตอนวิธีการลบสเปกตรัมซึ่งไม่เหมาะสมในการลดสัญญาณรบกวนแบบไม่คงที่ งานวิจัยดังกล่าวจึงได้นำเสนอขั้นตอนวิธีที่สามารถประยุกต์ใช้กับสัญญาณรบกวนแบบไม่คงที่ได้ คือ การลบสเปกตรัมแบบเวฟเล็ต (Wavelet Spectral Subtraction) โดยผลการทดลองแสดงการเปรียบเทียบค่าความแม่นยำในการรู้จำ ระหว่างขั้นตอนวิธีที่นำเสนอ ขั้นตอนวิธีพื้นฐานและกรณีไม่ใช้ขั้นตอนวิธีใดเลยในการลดสัญญาณรบกวน ซึ่งผลที่ได้แสดงให้เห็นว่าวิธีที่นำเสนอมีประสิทธิภาพที่ดีกว่าขั้นตอนวิธีพื้นฐาน เมื่อพิจารณาสำหรับกรณีสัญญาณรบกวนแบบไม่คงที่ ขณะที่ Lin, Holms และ Ambikairajah [7] ได้นำเสนอวิธีการพัฒนาขั้นตอนวิธีการกรองวีเนอร์สำหรับกรณีสัญญาณรบกวนแบบไม่คงที่ และนำเสนอการใช้วิธีการติดตามประมาณสัญญาณรบกวนไม่คงที่ในสัญญาณเสียงพูดที่คิดค้นขึ้น [8] ซึ่งการประมาณสัญญาณรบกวนแบบไม่คงที่ก็ได้ถูกศึกษาและพัฒนาอย่างต่อเนื่องเช่นกัน โดยมีการใช้ขั้นตอนวิธีในการประมาณสัญญาณรบกวนที่เปลี่ยนแปลงตามเวลา [9]

จากการศึกษาวิเคราะห์งานวิจัยที่ผ่านมาพบว่า นักวิจัยยังคงมุ่งเน้นการนำเสนอขั้นตอนวิธีต่างๆโดยเริ่มต้นจากการประมาณสัญญาณรบกวนที่เกิดขึ้นในสัญญาณเสียงพูด โดยสัญญาณรบกวนที่พิจารณาอาจเป็นสัญญาณรบกวนแบบคงที่หรือหรือแบบไม่คงที่ ซึ่งขั้นตอนวิธีที่ใช้จะแตกต่างกันไป นอกจากนี้ยังมีการนำวิธีการตรวจหากิจกรรมเสียงพูดมาช่วยในการประมาณสัญญาณรบกวนให้ดียิ่งขึ้นได้อีกด้วย จากนั้นจึงนำเสนอหรือเลือกขั้นตอนวิธีในการลดสัญญาณรบกวนที่เหมาะสมในกระบวนการรู้จำเสียงพูดต่อไป ซึ่งโดยส่วนมากการเปรียบเทียบผลการทดลองจะพิจารณาจากการวัดค่าความแม่นยำในการรู้จำเสียงพูดของแต่ละขั้นตอนวิธี ที่อัตราส่วนสัญญาณต่อสัญญาณรบกวนต่างๆกัน

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

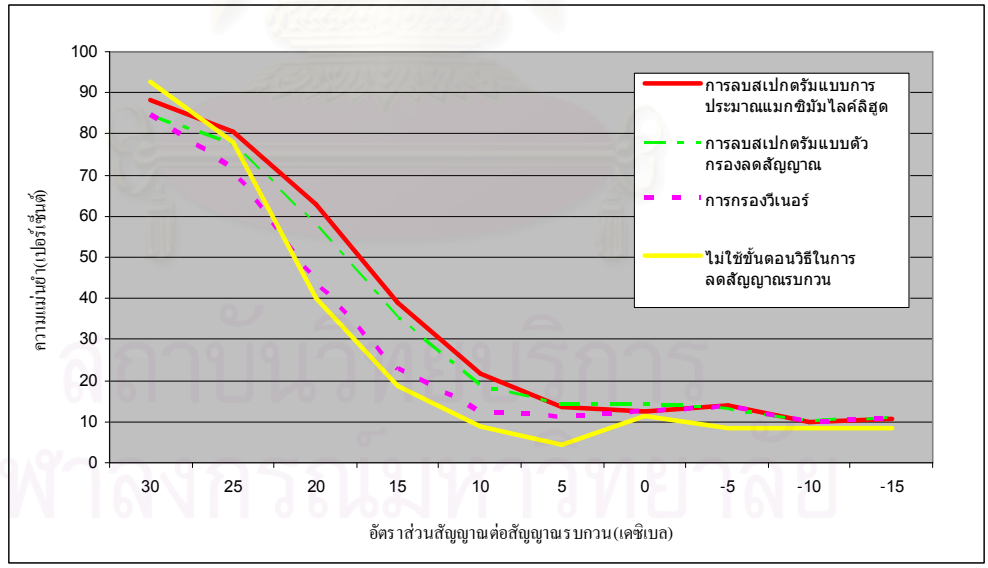
บทที่ 3

แนวคิดการวิจัย

การลดสัญญาณรบกวนด้วยขั้นตอนวิธีต่างๆ

จากการศึกษาและทดลองเบื้องต้นพบว่า ปัจจุบันสัญญาณรบกวนในสัญญาณเสียงพูดสำหรับระบบรู้จำเสียงพูดนั้นยังคงเป็นปัญหาที่เกิดขึ้นในปัจจุบัน โดยเฉพาะอย่างยิ่งในกรณีสัญญาณรบกวนแบบคงที่ซึ่งเกิดขึ้นบ่อยครั้งในชีวิตประจำวันและยังคงเป็นปัญหาที่มีอยู่ ทำให้นักวิจัยจำนวนมาก ยังคงศึกษาและพัฒนาขั้นตอนวิธีต่างๆในการลดสัญญาณรบกวนแบบคงที่ ดังกล่าวให้มีประสิทธิภาพดียิ่งขึ้น

จากการศึกษาทฤษฎีที่เกี่ยวข้องและเปรียบเทียบประสิทธิภาพในการรู้จำเสียงพูด ระหว่างขั้นตอนวิธี การลบสเปกตรัมแบบตัวกรองลดสัญญาณรบกวน การลบสเปกตรัมแบบการประมาณแม็กซิมัมไลค์ลิฮูด และการกรองวีเนอร์ดังแสดงในรูปที่ 3.1 และใช้ข้อมูลเสียงพูดที่จัดเตรียมในการทดสอบความแม่นยำ จากผลการทดลองเบื้องต้นเราสามารถวิเคราะห์ผลการทดลองเบื้องต้นได้ดังนี้



รูปที่ 3.1 กราฟแสดงผลการเปรียบเทียบความแม่นยำในการรู้จำระหว่างขั้นตอนวิธีการลบสเปกตรัมแบบตัวกรองลดสัญญาณรบกวน ขั้นตอนวิธีการลบสเปกตรัมแบบการประมาณแม็กซิมัมไลค์ลิฮูด ขั้นตอนการกรองวีเนอร์และกรณีไม่ใช้ขั้นตอนวิธีใดเลย

1. สัญญาณรบกวนมีผลทำให้ความแม่นยำในการรู้จำมีค่าลดลง ทั้งนี้เนื่องจากสัญญาณรบกวนทำให้ลักษณะสำคัญของเสียงพูดมีค่าเปลี่ยนแปลง ส่งผลให้ระบบรู้จำเสียงพูดไม่สามารถทำการรู้จำได้อย่างถูกต้องแม่นยำเนื่องจากลักษณะสำคัญของเสียงพูดไม่ใกล้เคียงกับข้อมูลที่ทำการสอน
2. ขั้นตอนวิธีในการลดสัญญาณรบกวนสามารถช่วยลดทอนสัญญาณรบกวนและเพิ่มประสิทธิภาพในการรู้จำได้ โดยกรณีสัญญาณเสียงมีสัญญาณรบกวนอยู่มากสัญญาณรบกวนดังกล่าวจะถูกลดทอนมากกว่าในกรณีสัญญาณเสียงมีสัญญาณรบกวนอยู่น้อย
3. ในกรณีสัญญาณเสียงพูดที่มีค่าอัตราส่วนสัญญาณต่อสัญญาณรบกวนสูงเมื่อผ่านขั้นตอนการลดสัญญาณรบกวน ค่าความแม่นยำในการรู้จำมีค่าลดลงเมื่อเทียบกับกรณีสัญญาณไม่ได้ถูกลดทอนด้วยขั้นตอนวิธีใดๆ
4. ในกรณีสัญญาณเสียงพูดที่มีค่าอัตราส่วนสัญญาณต่อสัญญาณรบกวนต่ำเมื่อผ่านขั้นตอนการลดสัญญาณรบกวน ค่าความแม่นยำในการรู้จำมีค่าสูงขึ้นเมื่อเทียบกับกรณีสัญญาณไม่ได้ถูกลดทอนด้วยขั้นตอนวิธีใดๆ

จากการศึกษางานวิจัยที่เกี่ยวข้อง ในการเพิ่มประสิทธิภาพให้กับขั้นตอนการลดสัญญาณรบกวน ทำให้เกิดแนวคิดในการประยุกต์และพัฒนาขั้นตอนวิธีที่ใช้ในปัจจุบันเพื่อเพิ่มความแม่นยำในการรู้จำให้ดียิ่งขึ้น ทั้งนี้กระบวนการในการลดสัญญาณรบกวนสามารถนำมาประยุกต์ได้เป็น 4 ขั้นตอนดังนี้

1. การวิเคราะห์ผลค่าถ่วงน้ำหนักที่มีผลต่อความแม่นยำ

จากการศึกษางานวิจัยที่เกี่ยวข้อง[4] ในการใช้ค่าถ่วงน้ำหนักเพื่อปรับขนาดสัญญาณเสียงพูดโดยให้ความสำคัญในส่วนของสัญญาณที่เป็นเสียงพูดและส่วนที่ไม่ใช่เสียงพูด ให้มีค่าแตกต่างกัน ดังนั้นในงานวิจัยนี้จึงเสนอการใช้ค่าถ่วงน้ำหนักกับขั้นตอนวิธีการลบสเปกตรัมแบบการประมาณแม็กซิมัมไลค์ลิฮูด โดยค่าถ่วงน้ำหนักจะมีความสัมพันธ์กับสัญญาณรบกวนที่มีอยู่ในสัญญาณเสียงพูด โดยสมการการลบสเปกตรัมมีสมการดังที่กล่าวไปแล้วคือ

$$|\hat{X}(pL,\omega)| = \frac{1}{2}|Y(pL,\omega)| + \frac{1}{2}[|Y(pL,\omega)|^2 - \hat{S}_b(\omega)]^{1/2}$$

จากการวิเคราะห์สมการข้างต้นพบว่า ในทุกๆกรณีไม่ว่าสัญญาณเสียงพูดที่ระบบทำการรู้จำหรือวัดความแม่นยำนั้นจะประกอบด้วยสัญญาณรบกวนมากน้อยเพียงใด การให้ความสำคัญต่อ

การลดสัญญาณรบกวนจากสัญญาณเสียงพูดที่ทราบได้ $\frac{1}{2}\left[|Y(pL,\omega)|^2 - \hat{S}_b(\omega)\right]^{1/2}$ จะคงที่เสมอ จากจุดนี้จึงเกิดแนวคิดในงานวิจัยโดยใช้หลักการพิจารณาที่ว่า ในกรณีสัญญาณเสียงพูดที่มีสัญญาณรบกวนแตกต่างกัน การให้ความสำคัญกับสัญญาณเสียงพูดที่เราทราบจึงควรแตกต่างกันด้วย ดังนั้นเมื่อพิจารณาสมการดังกล่าวอีกรูปแบบหนึ่ง เราสามารถพิจารณาสมการใหม่ได้ดังนี้

$$|\hat{X}(pL,\omega)| = \beta|Y(pL,\omega)| + (1-\beta)[|Y(pL,\omega)|^2 - \hat{S}_b(\omega)]^{1/2} \qquad (3.1)$$

เมื่อ β แสดงถึงค่าถ่วงน้ำหนักในการให้ความสำคัญของสัญญาณเสียงพูดที่ตรวจหาได้ โดยค่า β มีค่าที่เป็นไปได้อยู่ระหว่างศูนย์ถึงหนึ่ง ทั้งนี้สามารถสังเกตได้ว่ากรณีการลบสเปกตรัมแบบการประมาณแบบแม็กซิมัมไลค์ลิฮูดคือการใช้ค่า β เท่ากับ 0.5 นั่นเอง

โดยในงานวิจัยนี้จะทำการทดลองวัดค่าความแม่นยำในการรู้จำเสียงพูดโดยทดลองเปลี่ยน β ที่ค่าต่างๆ ได้แก่ 0.1, 0.2, 0.3, 0.4, 0.6, 0.7, 0.8, 0.9 ทั้งนี้เพื่อศึกษาแนวโน้มและผลความแม่นยำในการรู้ที่ได้จากการปรับค่าดังกล่าว

2. การประมาณขนาดของสัญญาณรบกวน

จากสมมติฐานเบื้องต้น ค่า β ในขั้นตอนที่หนึ่งควรมีความสัมพันธ์กับขนาดและปริมาณสัญญาณรบกวนหรืออัตราส่วนสัญญาณต่อสัญญาณรบกวนที่อยู่ในสัญญาณเสียงพูด ดังนั้นหากสามารถทราบได้ว่าสัญญาณรบกวนในสัญญาณเสียงพูดนั้นมีค่ามากน้อยเพียงใด เราสามารถเลือกค่าถ่วงน้ำหนักที่เหมาะสมเพื่อเพิ่มความแม่นยำในการรู้จำเสียงพูดต่อไปได้

ในงานวิจัยจำนวนมากซึ่งพิจารณาสัญญาณรบกวนในแบบคงที่ เราสามารถทำการประมาณขนาดของสัญญาณรบกวนได้จากสัญญาณเสียงในช่วง 0 – 100 มิลลิวินาทีแรกเพื่อนำมาใช้เป็นตัวแทนสัญญาณรบกวนที่เกิดขึ้นในสัญญาณเสียงพูดกรอบอื่นๆ ทั้งนี้เนื่องจากในช่วงเวลาดังกล่าวถือว่าเป็นช่วงที่ไม่มีสัญญาณเสียงพูดอยู่ ดังนั้นเราสามารถใช้สัญญาณเสียงในช่วงดังกล่าวเพื่อวิเคราะห์และประมาณปริมาณของสัญญาณรบกวนได้

ขั้นตอนนี้นำเสนอการใช้ค่าพลังงานของสัญญาณเสียงพูดที่พิจารณา ในการประมาณขนาดและปริมาณสัญญาณรบกวนของสัญญาณเสียงที่พิจารณา ทั้งนี้เนื่องจากสัญญาณเสียงพูดที่พิจารณาอาจมีขนาดสัญญาณเสียงพูดแตกต่างกันมาก ดังนั้นในขั้นตอนนี้จะปรับสัญญาณเสียงพูดให้เป็นบรรทัดฐานเดียวกันด้วยขนาดสูงสุดของสัญญาณเสียงพูดนั้นๆ จากนั้นจึงทำการวัดปริมาณค่าพลังงานสัญญาณเสียงพูด

3. การหาฟังก์ชันการกำหนดค่าถ่วงน้ำหนักจากขนาดของสัญญาณรบกวน

หลังจากได้ผลในขั้นตอนที่หนึ่งและขั้นตอนที่สอง งานวิจัยนี้เสนอการใช้ฟังก์ชันซึ่งมีความสัมพันธ์ระหว่างค่าถ่วงน้ำหนักและค่าพลังงานของสัญญาณรบกวน เพื่อใช้หาค่าถ่วงน้ำหนักที่เหมาะสมที่สุดสำหรับสัญญาณเสียงพูดที่มีสัญญาณรบกวนขนาดต่างๆกัน เพื่อใช้ในขั้นตอนวิธีการลดสัญญาณรบกวนด้วยขั้นตอนวิธีที่นำเสนอ

4. การประเมินผล

งานวิจัยนี้ทำการวิเคราะห์และเปรียบเทียบค่าความแม่นยำจากระบบรู้จำเสียง ระหว่างขั้นตอนวิธีที่นำเสนอ ขั้นตอนวิธีการลบสเปกตรัมแบบการประมาณแม็กซิมัมไลค์ลิฮูดและกรณีไม่ใช้ขั้นตอนวิธีใดเลย โดยใช้ข้อมูลเสียงพูดสำหรับทดสอบทั้งในกรณีเสียงพูดตัวเลขภาษาไทยและเสียงพูดที่เป็นประโยคภาษาไทยทั่วไป โดยทำการเปรียบเทียบค่าความแม่นยำในการรู้จำเสียงพูดของสัญญาณเสียงพูดที่ถูกลดสัญญาณรบกวน ด้วยขั้นตอนวิธีที่นำเสนอและขั้นตอนวิธีอื่นๆในปัจจุบันด้วยการใช้ระบบรู้จำเสียงพูดที่ใช้แบบจำลองฮิดเดนมาร์คอฟ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 4

การทดลองและวิเคราะห์ผล

วิธีการทดลอง

การทดลองเพื่อหาความแม่นยำในการรู้จำเสียงพูดสามารถแบ่งเป็นขั้นตอนต่างๆได้ดังนี้

1. การเตรียมข้อมูลเสียงพูด

ข้อมูลเสียงพูดที่ใช้ในงานวิจัยมี 2 ประเภทซึ่งมีลักษณะรูปแบบภาษา (Language Model) แตกต่างกัน ได้แก่

1. สัญญาณเสียงพูดตัวเลขภาษาไทย

ผู้วิจัยทำการจัดเตรียมตัวเลขสุ่มเจ็ดหลักจำนวน 1,040 ชุด โดยจัดแบ่งข้อมูลเป็นกลุ่มๆละ 40 ชุด ดังนั้นจะได้กลุ่มข้อมูลทั้งสิ้น 26 กลุ่ม จากนั้นทำการบันทึกเสียงพูดตัวเลขโดยพูดเป็นภาษาไทย เช่น 0213237 ให้อ่านว่า “ศูนย์ สอง หนึ่ง สาม สอง สาม เจ็ด” โดยให้ผู้ชายจำนวน 13 คนและผู้หญิงจำนวน 13 คนทำการบันทึกเสียงพูดกลุ่มคำตัวเลขภาษาไทยคนละหนึ่งกลุ่ม เมื่อได้ข้อมูลสัญญาณเสียงพูดเป็นที่เรียบร้อยทำการแบ่งข้อมูลทั้งหมดจากผู้ชายจำนวน 7 คนและผู้หญิง 7 คน เพื่อใช้เป็นข้อมูลในการสอนระบบเพื่อทำการเรียนรู้เสียงพูด(Training Set) ใช้ข้อมูลจากผู้ชาย 3 คนและผู้หญิง 3 คนเพื่อใช้เป็นข้อมูลในการพัฒนา(Development Set) โดยชุดข้อมูลในการพัฒนานี้จะใช้ในการทดลองเพื่อปรับเปลี่ยนค่าตัวแปรต่างๆให้ได้ผลการทดลองที่ดีที่สุด ก่อนที่จะใช้ค่าตัวแปรชุดดังกล่าวในชุดข้อมูลทดสอบต่อไป และใช้ชุดข้อมูลจากผู้ชาย 3 คนและผู้หญิง 3 คนในการทดสอบความแม่นยำในการรู้จำเสียงพูดของระบบ(Test Set)

2. สัญญาณเสียงพูดประโยคภาษาไทยทั่วไป

งานวิจัยนี้ใช้ชุดข้อมูลสำหรับสัญญาณเสียงพูดประโยคภาษาไทยทั่วไป จากฐานข้อมูล LOTUS [10] โดยใช้สัญญาณเสียงพูดประโยคภาษาไทยทั่วไปจำนวน 1,295 ประโยคในการสอนระบบเพื่อทำการเรียนรู้เสียงพูดและใช้สัญญาณเสียงพูดประโยคภาษาไทยทั่วไปจำนวน 105 ประโยคเพื่อเป็นชุดข้อมูลในการพัฒนา และใช้สัญญาณเสียงพูดประโยคภาษาไทยทั่วไปจำนวน 280 ประโยคในการทดสอบความแม่นยำในการรู้จำเสียงพูดของระบบ รวมข้อมูลที่ใช้ทั้งสิ้น 1,680 ประโยค

2. การสอนระบบรู้จำเสียงพูด

ในงานวิจัยนี้ใช้เครื่องมือในการรู้จำเสียงพูดด้วยชุดเครื่องมือฮิดเดนมาร์คอฟ (Hidden Markov Toolkit) ในการสอนระบบให้ทำการรู้จำชุดข้อมูลสัญญาณเสียงพูดสะอาดที่จัดเตรียมไว้สำหรับทำการสอนระบบ นอกจากนี้ยังต้องทำการจัดเตรียมข้อมูลแฟ้มฉลากของสัญญาณเสียงพูดแต่ละประโยคด้วย

ทั้งนี้ข้อมูลในแฟ้มฉลากนั้นจะประกอบไปด้วยส่วนสำคัญ2 ส่วนคือ ชื่อของแฟ้มฉลากซึ่งจะตรงกับชื่อแฟ้มของสัญญาณเสียงพูดและอีกส่วนหนึ่งคือส่วนของเนื้อประโยคซึ่งเขียนให้อยู่ในรูปของหน่วยเสียงย่อยๆต่อกัน (phoneme) ดังแสดงในตัวอย่างดังนี้

ตัวอย่างข้อมูลแฟ้มฉลากของสัญญาณเสียงพูดตัวเลขภาษาไทย

แฟ้มสัญญาณเสียงพูดมีชื่อว่า “0071262.wav” และเนื้อความในประโยคที่บันทึกเสียงพูดคือ“ศูนย์ ศูนย์ เจ็ด หนึ่ง สอง หก สอง” แฟ้มฉลากของสัญญาณเสียงนี้มีเนื้อความดังแสดงในรูปที่ 4.1

```
#!MLF!#
"*/0071262.lab"
sil
s
uu
n^
s
uu
n^
c
e
t^
n
v
ng^
s
@@
ng^
h
o
k^
s
@@
ng^
sil
```

รูปที่ 4.1 ตัวอย่างเนื้อความแฟ้มฉลากของสัญญาณเสียงพูดตัวเลขภาษาไทย

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตัวอย่างข้อมูลแฟ้มฉลากของสัญญาณเสียงพูดประโยคภาษาไทยทั่วไป

แฟ้มสัญญาณเสียงพูดมีชื่อว่า "UOF001_Pa001_001.wav" และเนื้อความในประโยคที่บันทึกเสียงพูดคือ "นายสง่า สรรพศรี"

```
#!MLF!#
"*/UOF001_Pa001_001.lab"
sil
n
aa
j^
s
a
ng
aa
s
a
p^
sp
ph
a
s
ii
sil
```

รูปที่ 4.2 ตัวอย่างเนื้อความแฟ้มฉลากของสัญญาณเสียงพูดประโยคภาษาไทยทั่วไป

จำนวนหน่วยเสียงทั้งหมด ที่ใช้ในเป็นตัวแทนตัวอักษรสำหรับเสียงพูดภาษาไทยในกระบวนการรู้จำเสียงพูดจะมีทั้งสิ้น 74 ตัว โดยสามารถแบ่งประเภทของหน่วยเสียงได้เป็น 3 ประเภทได้แก่ หน่วยเสียงประเภทพยัญชนะต้น หน่วยเสียงประเภทสระและหน่วยเสียงประเภทตัวสะกด ทั้งนี้ในแต่ละประเภทจะมีทั้งหน่วยเสียงเดี่ยวและหน่วยเสียงผสม ดังแสดงในตาราง 4.1,4.2 และ 4.3 ตามลำดับ นอกจากหน่วยเสียงย่อยที่จับคู่กับตัวอักษรภาษาไทยทั้ง 74 ตัวแล้ว ยังมีหน่วยเสียงย่อยที่ใช้อยู่อีก 2 ตัวได้แก่ "sil" และ "sp" โดย sil ถูกใช้ในการแทนช่วงสัญญาณเสียงเงียบที่หัวท้ายของประโยคในสัญญาณเสียงพูด ขณะที่ sp ใช้ในกรณีช่วงสัญญาณเสียงเงียบระหว่างหน่วยเสียงมากกว่า 20 มิลลิวินาที [10]

ตารางที่ 4.1 รายชื่อหน่วยเสียงจับคู่กับพยัญชนะต้นในภาษาไทย [10]

พยัญชนะต้น			
เดี่ยว	**ตัวอย่าง**	**ผสม**	**ตัวอย่าง**
p	ปาก	pr	ประสาน
t	เต้น, กุฏิ	phr	พราน
c	จะ	tr	เตรียม
k	ก่อน	kr	กราบ
z	อาน	khr	คร่า
ph	พบ, ภัย, ผ่าน	pl	ปลา
th	ทิ้ง, ธง, เฒ่า,	phl	พลาด
	ฐาน, มณโฑ		
ch	ชอบ, เฌอ	thr	จันทรา
kh	คน, เขิน, ฆ่า	kl	เกลอ
b	บอก	khl	เคลื่อน
d	ด้าน, ชฎา	kw	กวาง
m	ไม่	khw	ขวา
n	นาน, เณร	**เสียงทับศัพท์**	
ng	เงิน	br	เบรน
l	เล่น, กีฬา	bl	บลู
r	รอ, ฤทัย	fr	ฟราย
f	ฝน, ฟัน	fl	เฟลม
s	สาย, ศิลา,	dr	ดราก้อน
	รักษา, ซ่อน		
h	โหน, เฮฮา	17 หน่วย	
w	ว่า		
j	ย้อน, หญิง		
21 หน่วย			

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.2 รายชื่อหน่วยเสียงจับคู่กับสระในภาษาไทย [10]

สระ			
เดี่ยว	**ตัวอย่าง**	**ผสม**	**ตัวอย่าง**
a	อะ	Ia	เอียะ
aa	อา	Iia	เอีย
i	อิ	Va	เอือะ
ii	อี	vva	เอือ
v	อึ	Ua	อัวะ
vv	อื	uua	อัว
u	อุ	6 หน่วย	
uu	อู		
e	เอะ		
ee	เอ		
x	แอะ		
xx	แอ		
o	โอะ		
oo	โอ		
@	เอาะ		
@@	ออ		
q	เออะ		
qq	เออ		
18 หน่วย			

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.3 รายชื่อหน่วยเสียงจับคู่กับตัวสะกดในภาษาไทย [10]

ตัวสะกด	
เดี่ยว	**ตัวอย่าง**
p^	พบ
t^	เกร็ด
k^	ปาก
n^	หาร
m^	ลม
ng^	ฟาง
j^	ยาย
w^	กาว
เสียงทับศัพท์	
f^	กราฟ
l^	แอล
s^	เอส
ch^	คลัช
12 หน่วย	

3. การลดสัญญาณรบกวนด้วยขั้นตอนวิธีต่างๆ

การทดลองในงานวิจัยนี้ใช้เครื่องมือในการประมวลผลสัญญาณด้วยโปรแกรมแมทแลปซึ่งใช้ในการนำเข้าสัญญาณเสียงพูดและประมวลผลสัญญาณด้วยฟังก์ชันและขั้นตอนวิธีต่างๆรวมถึงขั้นตอนวิธีที่นำเสนอ จากนั้นทำการส่งออกสัญญาณเสียงพูดที่ผ่านกระบวนการลดสัญญาณรบกวนเพื่อนำเข้าสู่กระบวนการทดสอบความแม่นยำในการรู้จำเป็นขั้นตอนถัดไป

4. การวัดความแม่นยำในการรู้จำเสียงพูด

เช่นเดียวกับขั้นตอนการสอนระบบรู้จำเสียงพูด การทดลองในขั้นตอนนี้ต้องทำการสร้างแฟ้มฉลากของสัญญาณเสียงพูดที่จะใช้ในการทดสอบความแม่นยำ โดยข้อมูลในแฟ้มฉลากระบุข้อมูลประโยคที่ถูกต้อง เพื่อใช้เป็นคำตอบในการเปรียบเทียบผลลัพธ์ของการรู้จำสัญญาณเสียงพูดด้วยระบบรู้จำเสียงพูด

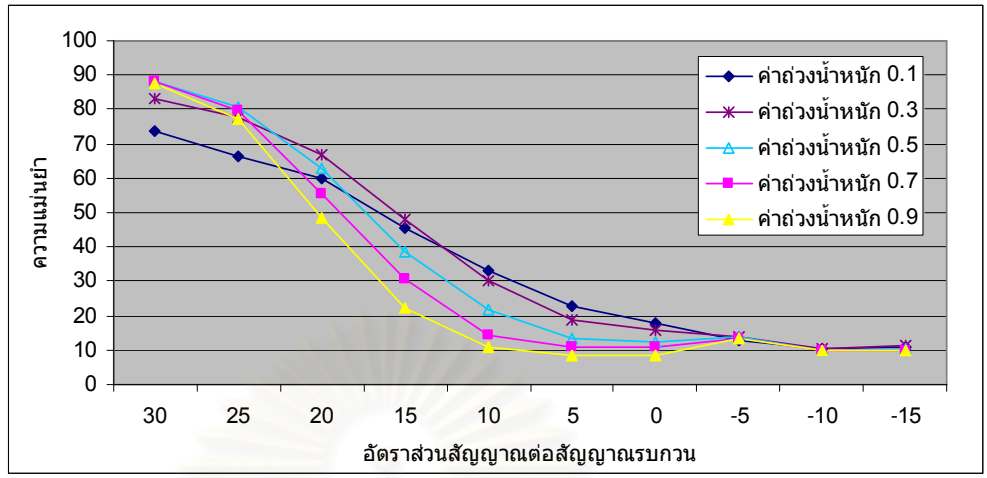
ผลการทดลอง

จากแนวคิดในการปรับค่าถ่วงน้ำหนักในสมการที่ 3.1 เพื่อดูผลของความแม่นยำในการรู้จำเสียงพูดที่ได้จากการปรับค่าถ่วงน้ำหนัก เมื่อใช้ชุดข้อมูลในการพัฒนาและทำการปรับค่าถ่วงน้ำหนัก β ให้มีค่าเท่ากับ 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8 และ 0.9 ได้ผลการทดลองดังตารางที่ 4.4

ตารางที่ 4.4 ความแม่นยำในการรู้จำเสียงพูดตัวเลขที่ค่าอัตราส่วนสัญญาณต่อสัญญาณรบกวนที่ค่าถ่วงน้ำหนักต่างๆ

ค่าถ่วงน้ำหนัก (β)	**อัตราส่วนสัญญาณต่อสัญญาณรบกวน(เดซิเบล)**									
	30	**25**	**20**	**15**	**10**	**5**	**0**	**-5**	**-10**	**-15**
0.1	73.57	66.49	59.88	45.30	33.21	22.74	17.74	12.91	10.22	10.99
0.2	79.35	73.15	63.39	48.45	32.80	20.95	17.80	13.63	9.98	10.25
0.3	83.27	77.86	66.73	47.98	30.00	18.93	16.07	13.99	10.28	11.37
0.4	87.02	79.88	65.48	44.00	25.42	15.71	14.11	13.51	10.04	11.08
0.5	88.15	80.48	62.98	38.81	21.67	13.45	12.38	13.87	9.91	10.59
0.6	88.33	80.77	59.35	33.39	17.62	11.61	11.96	13.63	10.16	10.11
0.7	88.27	79.88	55.48	30.60	14.46	10.89	10.89	13.45	10.10	10.14
0.8	88.15	78.45	51.61	26.67	12.62	9.70	10.12	13.45	9.98	9.94
0.9	87.86	77.32	48.27	22.44	10.89	8.51	8.45	13.57	10.10	9.85

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

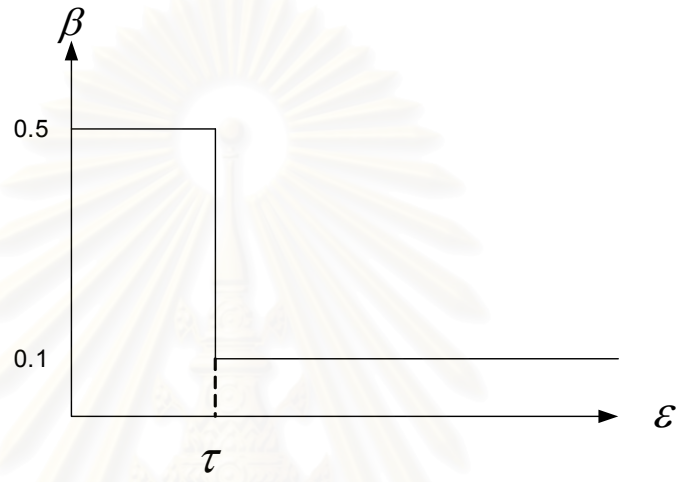


รูปที่ 4.3 กราฟแสดงผลการทดลองเปลี่ยนค่าถ่วงน้ำหนักต่อความแม่นยำในการรู้จำเสียงพูด เมื่อใช้ข้อมูลชุดเสียงพูดตัวเลข

จากผลการทดลองดังแสดงในรูปที่ 4.3 พบว่ากรณีที่สัญญาณเสียงพูดมีสัญญาณรบกวนน้อยหรือค่าอัตราส่วนต่อสัญญาณรบกวนมีค่าสูง ค่าความแม่นยำในการรู้จำเสียงพูดกรณีที่ค่าถ่วงน้ำหนักน้อยจะมีค่าน้อยกว่ากรณีเมื่อค่าถ่วงน้ำหนักมากขึ้น ทั้งนี้เนื่องจากค่าถ่วงน้ำหนักที่เพิ่มขึ้นจะหมายถึงการให้ความสำคัญสัญญาณเสียงพูดที่ตรวจหาได้ นั่นคือสัญญาณเสียงพูดดังกล่าวจะถูกลดทอนสัญญาณรบกวนน้อยลง ทำให้สัญญาณเสียงพูดที่พิจารณายังคงรักษาคุณลักษณะเดิมของสัญญาณเสียงไว้ซึ่งส่งผลให้ความแม่นยำมีค่าสูง ส่วนในกรณีที่สัญญาณเสียงพูดที่มีสัญญาณรบกวนอยู่มากหรือมีค่าอัตราส่วนสัญญาณต่อสัญญาณรบกวนต่ำ เราสามารถสังเกตได้ว่าค่าถ่วงน้ำหนักที่ลดลงสามารถเพิ่มประสิทธิภาพในการรู้จำสัญญาณเสียงพูดได้อย่างมีประสิทธิภาพ

จะเห็นได้ว่าที่สัญญาณเสียงพูดมีสัญญาณรบกวนน้อยๆ ระบบควรเลือกใช้ค่าถ่วงน้ำหนักที่ 0.5 ซึ่งให้ผลในการรู้จำเสียงพูดได้ดีกว่ากรณีค่าถ่วงน้ำหนักที่น้อยกว่าและมีแนวโน้มการลดต่ำของความแม่นยำน้อยกว่ากรณีที่ใช้ค่าถ่วงน้ำหนักที่มีค่ามากกว่า ขณะเดียวกันกรณีที่สัญญาณเสียงพูดมีสัญญาณรบกวนอยู่มากระบบควรเลือกใช้ค่าถ่วงน้ำหนักที่น้อยที่สุด โดยในงานวิจัยนี้เลือกใช้ที่ค่าถ่วงน้ำหนักเท่ากับ 0.1 ซึ่งให้ประสิทธิภาพในการลดสัญญาณรบกวนและมีความแม่นยำในการรู้จำเสียงพูดสูงสุด ดังนั้นหากเราสามารถทราบได้ว่าสัญญาณเสียงพูดที่พิจารณานั้นมีค่าสัญญาณรบกวนมากหรือน้อยเพียงใด เราก็สามารถทำการเลือกค่าถ่วงน้ำหนักที่เหมาะสมเพื่อเพิ่มประสิทธิภาพในการรู้จำเสียงพูดได้

งานวิจัยนี้นำเสนอการใช้สมการความสัมพันธ์ระหว่างค่าถ่วงน้ำหนักและค่าเฉลี่ยพลังงานของสัญญาณรบกวนซึ่งใช้ค่าเฉลี่ยพลังงานของสัญญาณเสียงพูดที่พิจารณาในช่วง 0-100 มิลลิวินาทีแรกซึ่งถือว่าเป็นช่วงสัญญาณเงียบ โดยเริ่มต้นพิจารณาด้วยการใช้ฟังก์ชันขั้นบันไดซึ่งมีลักษณะดังรูปที่ 4.4



รูปที่ 4.4 ฟังก์ชันความสัมพันธ์ระหว่างค่าถ่วงน้ำหนักและค่าเฉลี่ยพลังงานสัญญาณรบกวนแบบขั้นบันได

เมื่อ τ หมายถึงค่าขีดแบ่งค่าเฉลี่ยพลังงานสัญญาณรบกวนในการกำหนดค่าถ่วงน้ำหนัก และ ε หมายถึงค่าเฉลี่ยพลังงานสัญญาณรบกวน

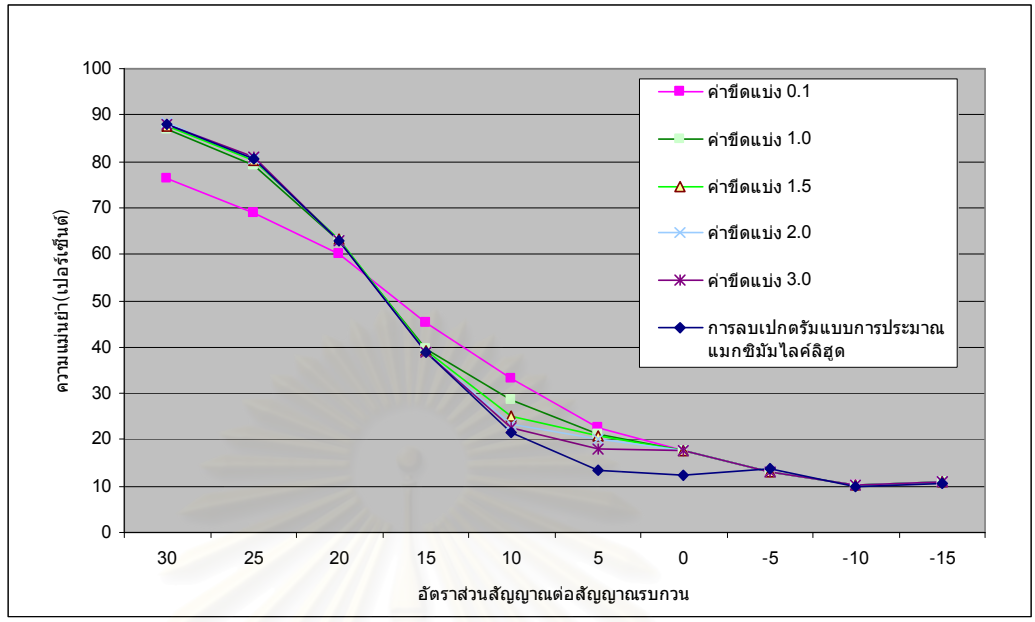
เนื่องจากเราไม่สามารถทราบได้ว่า สัญญาณเสียงพูดที่มีค่าสัญญาณต่อสัญญาณรบกวนค่าหนึ่งจะมีค่าเฉลี่ยพลังงานสัญญาณรบกวนเท่ากับเท่าใด เนื่องจากสัญญาณเสียงพูดแต่ละประโยคนั้นจะมีลักษณะเฉพาะที่แตกต่างกัน กล่าวคือ สัญญาณเสียงพูดสองสัญญาณที่มีค่าอัตราส่วนสัญญาณต่อสัญญาณรบกวนเท่ากันอาจมีค่าเฉลี่ยพลังงานสัญญาณรบกวนไม่เท่ากัน

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ดังนั้นงานวิจัยนี้จึงเริ่มต้นด้วยการทดลอง การเปลี่ยนค่าขีดแบ่งค่าเฉลี่ยพลังงานสัญญาณรบกวน τ ที่ค่าต่างๆเพื่อศึกษาแนวโน้มของค่าความแม่นยำที่เปลี่ยนไป โดยได้ผลการทดลองดังตารางที่ 4.5

ตารางที่ 4.5 ค่าความแม่นยำในการรู้จำเสียงพูดตัวเลขที่ค่าอัตราส่วนสัญญาณต่อสัญญาณรบกวนที่ค่าขีดแบ่งค่าเฉลี่ยพลังงานสัญญาณรบกวนค่าต่างๆเมื่อใช้ฟังก์ชันขั้นบันได

ค่าขีดแบ่งค่าเฉลี่ยพลังงานสัญญาณรบกวน (τ)	**อัตราส่วนสัญญาณต่อสัญญาณรบกวน(เดซิเบล)**									
	30	**25**	**20**	**15**	**10**	**5**	**0**	**-5**	**-10**	**-15**
0.1	76.49	68.81	60.06	45.12	33.21	22.74	17.74	12.91	10.22	10.99
0.2	78.99	71.19	60.71	44.64	33.27	22.74	17.74	12.91	10.22	10.99
0.3	81.31	73.75	61.43	43.57	33.15	22.74	17.74	12.91	10.22	10.99
0.4	83.04	75.83	61.79	41.96	31.61	22.50	17.74	12.91	10.22	10.99
0.5	83.99	76.55	62.38	41.61	31.43	22.26	17.74	12.91	10.22	10.99
0.6	84.94	77.02	62.44	41.25	31.01	22.02	17.74	12.91	10.22	10.99
0.7	85.83	78.21	62.92	41.07	30.65	22.14	17.74	12.91	10.22	10.99
0.8	86.31	78.75	62.92	40.48	30.06	21.55	17.74	12.91	10.22	10.99
0.9	86.55	79.17	62.98	40.12	28.99	21.55	17.74	12.91	10.22	10.99
1.0	86.90	79.23	62.86	39.46	28.57	21.25	17.74	12.91	10.22	10.99
1.5	87.68	80.36	63.10	39.17	24.94	20.71	17.62	12.91	10.22	10.99
2.0	88.40	80.60	63.00	38.80	23.40	20.00	17.62	12.91	10.22	10.99
2.5	88.15	80.77	63.04	38.87	23.15	18.99	17.62	12.91	10.22	10.99
3.0	88.15	80.77	63.04	38.93	22.74	18.10	17.62	12.91	10.22	10.99



รูปที่ 4.5 กราฟแสดงผลการทดลองเปลี่ยนค่าขีดแบ่งค่าเฉลี่ยพลังงานของสัญญาณรบกวนต่อความแม่นยำในการรู้จำเสียงพูดเมื่อใช้ฟังก์ชันขั้นบันได

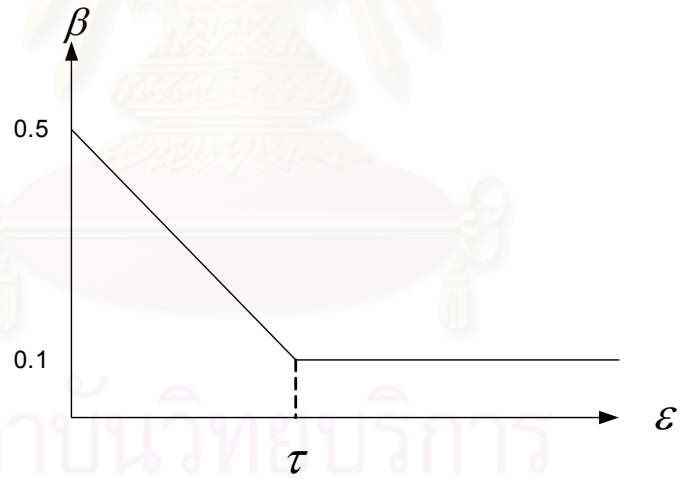
จากผลการทดลองพบว่าที่ค่าขีดแบ่งค่าเฉลี่ยพลังงานที่เพิ่มขึ้นสามารถช่วยเพิ่มความแม่นยำในการรู้จำเสียงพูดได้ดี สำหรับกรณีที่สัญญาณเสียงพูดมีสัญญาณรบกวนน้อยหรือมีค่าสัญญาณต่อสัญญาณรบกวนสูง แต่ในกรณีที่สัญญาณเสียงพูดมีสัญญาณรบกวนมากหรือค่าสัญญาณต่อสัญญาณรบกวนสูงค่าความแม่นยำจะมีค่าลดลงแทน ทั้งนี้เนื่องจากค่าขีดกั้นค่าเฉลี่ยพลังงานที่เพิ่มขึ้นหมายถึงการยังคงใช้ค่าถ่วงน้ำหนักที่ 0.5 ซึ่งเป็นการให้ความสำคัญในการลดสัญญาณรบกวนที่น้อยที่สุด แม้ว่าสัญญาณรบกวนจะมีขนาดสูงขึ้นจนถึงค่าขีดกั้นค่าเฉลี่ยพลังงาน ขณะเดียวกันที่ค่าขีดกั้นค่าเฉลี่ยพลังงานน้อยๆ ค่าความแม่นยำในการรู้จำเสียงพูดกรณีที่สัญญาณเสียงพูดมีสัญญาณรบกวนมากหรือค่าสัญญาณต่อสัญญาณรบกวนต่ำจะมีค่าสูง เนื่องจากสัญญาณเสียงพูดดังกล่าวถูกลดสัญญาณรบกวนด้วยค่าถ่วงน้ำหนักที่ 0.1 ซึ่งให้ความสำคัญในการลดสัญญาณรบกวนสูงสุด โดยเมื่อใช้ฟังก์ชันความสัมพันธ์ระหว่างค่าถ่วงน้ำหนักและค่าเฉลี่ยพลังงานสัญญาณรบกวนแบบขั้นบันได ระบบสามารถทำการรู้จำสัญญาณได้ดีขึ้นกว่าขั้นตอนวิธีการลบสเปกตรัมแบบการประมาณแมกซิมัมไลค์ลิฮูด ซึ่งเมื่อรวมค่าผลต่างความแม่นยำในการรู้จำเสียงพูด ระหว่างขั้นวิธีการลบสเปกตรัมแบบการประมาณแมกซิมัมไลค์ลิฮูดและขั้นตอนวิธีที่นำเสนอพบว่า ค่าความแม่นยำโดยรวมดีขึ้นในทุกกรณีและดีที่สุดที่ค่าความแม่นยำโดยรวมเท่ากับ 20.27% (ค่าดังกล่าวคำนวณจากผลรวมของผลต่างความแม่นยำระหว่างขั้นตอนวิธีการประมาณแบบแมกซิมัมไลค์ลิฮูดและขั้น

ตอนวิธีที่นำเสนอในทุกค่าอัตราส่วนสัญญาณต่อสัญญาณรบกวน) เมื่อใช้ค่าขีดแบ่งค่าเฉลี่ยพลังงานสัญญาณรบกวนเท่ากับ 0.7

อย่างไรก็ตามการใช้ฟังก์ชันขั้นบันไดยังมีข้อดีข้อเสีย กล่าวคือในกรณีที่สัญญาณเสียงพูดมีค่าสัญญาณรบกวนเพิ่มขึ้นแต่ยังไม่เกินค่าขีดแบ่งค่าเฉลี่ยพลังงานสัญญาณรบกวน ระบบจะยังคงใช้ค่าถ่วงน้ำหนักที่ 0.5 ซึ่งนั่นหมายถึงการให้ความสำคัญในการลดสัญญาณรบกวนคงที่ตลอด ซึ่งที่ค่าถ่วงน้ำหนักดังกล่าวจะมีผลดีเฉพาะที่ค่าเฉลี่ยพลังงานสัญญาณรบกวนน้อยๆเท่านั้น แต่ขณะเดียวกันระบบควรจะสามารถลดสัญญาณรบกวนได้มากขึ้นที่ค่าถ่วงน้ำหนักที่ลดลง

ด้วยเหตุผลข้างต้นดังกล่าว งานวิจัยนี้จึงเลือกใช้ข้อดีของทั้งสองกรณีไว้ด้วยกันด้วยการนำเสนอการใช้ฟังก์ชันเชิงเส้นที่มีลักษณะดังรูปที่ 4.6 และสามารถเขียนเป็นสมการได้ดังสมการที่ 4.1

$$\begin{aligned} \beta &= 0.5 - \left(\frac{0.4 \times \varepsilon}{\tau}\right) && if \quad \varepsilon < \tau \\ &= 0.1 && if \quad \varepsilon \geq \tau \end{aligned} \tag{4.1}$$



รูปที่ 4.6 ฟังก์ชันความสัมพันธ์ระหว่างค่าถ่วงน้ำหนักและค่าเฉลี่ยพลังงานสัญญาณรบกวนแบบเชิงเส้น

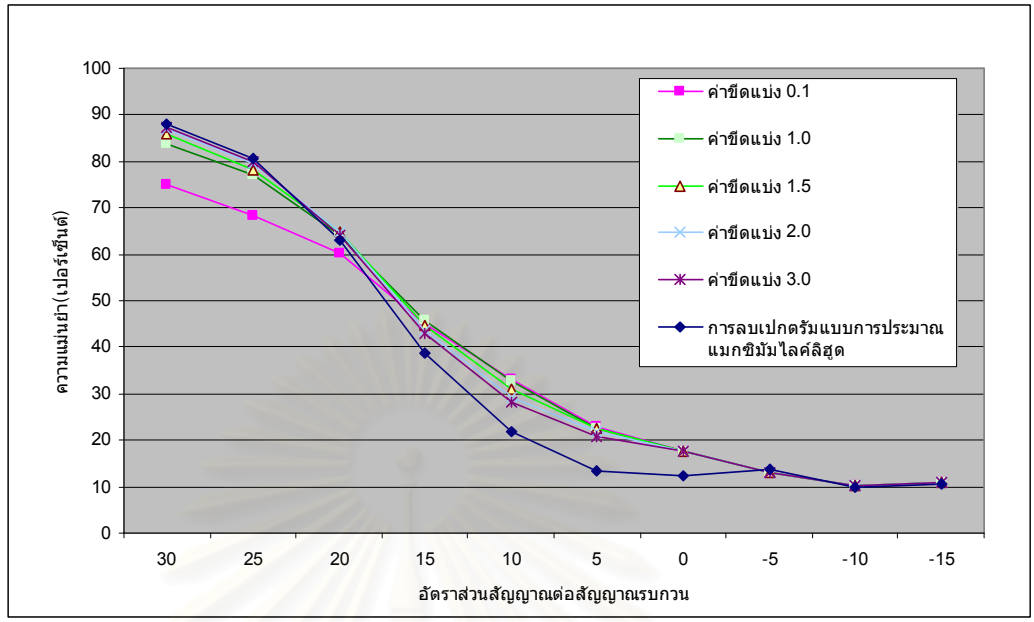
เมื่อพิจารณาฟังก์ชันเชิงเส้นในรูปที่ 4.6 จะพบว่าในกรณีที่สัญญาณเสียงพูดที่พิจารณามีสัญญาณรบกวนอยู่น้อย ระบบจะเลือกใช้ค่าถ่วงน้ำหนักที่มีค่าใกล้เคียงกับ 0.5 โดยค่าถ่วงน้ำหนักจะค่อยๆลดลงเมื่อค่าเฉลี่ยพลังงานสัญญาณรบกวนมีค่าเพิ่มขึ้น จนกระทั่งค่าเฉลี่ยพลังงานสัญญาณรบกวนมีค่ามากกว่า τ ค่าถ่วงน้ำหนักจะเท่ากับ 0.1

ดังนั้นงานวิจัยนี้จึงทำการทดลองในขั้นตอนต่อไป เพื่อหาแนวโน้มผลของค่าความแม่นยำในการรู้จำเสียงพูดที่ค่าสัญญาณต่อสัญญาณรบกวนต่างๆเมื่อทำการเปลี่ยนค่า τ โดยได้ผลการทดลองดังตารางที่ 4.6

ตารางที่ 4.6 ค่าความแม่นยำในการรู้จำเสียงพูดตัวเลขที่ค่าอัตราส่วนสัญญาณต่อสัญญาณรบกวนที่ค่าขีดแบ่งค่าเฉลี่ยพลังงานสัญญาณรบกวนค่าต่างๆเมื่อใช้ฟังก์ชันเชิงเส้น

ค่าขีดแบ่งค่าเฉลี่ยพลังงานสัญญาณรบกวน (τ)	**อัตราส่วนสัญญาณต่อสัญญาณรบกวน(เดซิเบล)**									
	30	**25**	**20**	**15**	**10**	**5**	**0**	**-5**	**-10**	**-15**
0.1	74.94	68.27	60.24	45.24	33.21	22.74	17.74	12.91	10.22	10.99
0.5	80.95	75.00	63.57	46.37	33.15	22.74	17.74	12.91	10.22	10.99
0.7	82.82	76.10	64.05	46.07	32.74	22.44	17.74	12.91	10.22	10.99
1.0	83.63	77.14	63.99	45.60	32.74	22.68	17.74	12.91	10.22	10.99
1.5	85.77	78.33	64.82	44.64	31.01	22.50	17.74	12.91	10.22	10.99
2.0	86.85	79.52	64.64	43.81	29.35	21.73	17.68	12.91	10.22	10.99
2.5	87.08	79.70	64.05	43.75	28.93	21.43	17.62	12.91	10.22	10.99
3.0	87.20	79.90	64.20	43.00	28.15	20.77	17.62	12.91	10.22	10.99

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 4.7 กราฟแสดงผลการทดลองเปลี่ยนค่าขีดแบ่งค่าเฉลี่ยพลังงานของสัญญาณรบกวนต่อความแม่นยำในการรู้จำเสียงพูดเมื่อใช้ฟังก์ชันเชิงเส้น

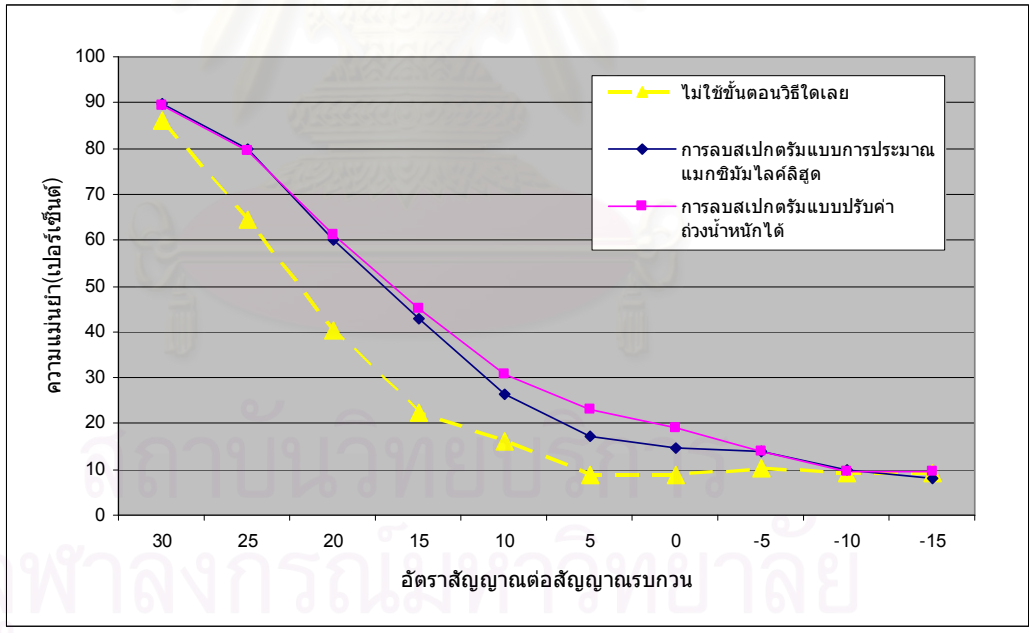
จากผลการทดลองพบว่าขั้นตอนวิธีที่นำเสนอสามารถช่วยลดสัญญาณรบกวนได้ โดยมีค่าความแม่นยำในการรู้จำเสียงพูดต่ำกว่าขั้นตอนวิธีการประมาณแบบแมกซิมัมไลค์ลิฮูดเล็กน้อย เมื่อใช้ค่าขีดแบ่งค่าเฉลี่ยพลังงานของสัญญาณรบกวนมากกว่า 1.0 แต่ขณะเดียวกันจะมีค่าความแม่นยำในการรู้จำที่ดีกว่าเมื่อสัญญาณเสียงพูดมีสัญญาณรบกวนมากขึ้น โดยค่าความแม่นยำโดยรวมนั้นดีกว่าขั้นตอนวิธีการประมาณแบบแมกซิมัมไลค์ลิฮูดในทุกกรณี และดีที่สุดที่ค่าความแม่นยำโดยรวมมากกว่าขั้นตอนวิธีการประมาณแบบแม็กซิมัมไลค์ลิฮูดถึง 26.64% เมื่อใช้ค่าขีดแบ่งค่าเฉลี่ยพลังงานสัญญาณรบกวนเท่ากับ 1.5 ซึ่งมีค่าความแม่นยำในการรู้จำเสียงที่ดีกว่าการใช้ฟังก์ชันขั้นบันได นอกจากนี้ยังสังเกตได้ว่าที่ค่าขีดแบ่งค่าเฉลี่ยพลังงานสัญญาณรบกวนที่ 0.7 การใช้ฟังก์ชันเชิงเส้นให้ค่าความแม่นยำมากกว่าการใช้ฟังก์ชันขั้นบันได

จากผลการทดลองทั้งหมดเราสามารถสรุปได้ว่า ขั้นตอนวิธีการลบสเปกตรัมแบบปรับค่าถ่วงน้ำหนักได้เมื่อใช้ฟังก์ชันเชิงเส้นสามารถช่วยลดสัญญาณรบกวนได้มีประสิทธิภาพดีที่สุด เมื่อเปรียบเทียบกับกรณีการใช้ขั้นตอนวิธีการลบสเปกตรัมแบบปรับค่าถ่วงน้ำหนักได้และขั้นตอนการประมาณแบบแมกซิมัมไลค์ลิฮูด โดยในชุดข้อมูลตัวเลขเราควรเลือกใช้ค่าขีดแบ่งค่าเฉลี่ยพลังงานให้อยู่ในช่วง 1.0 ถึง 2.0 โดยในการทดลองขั้นตอนต่อไป งานวิจัยนี้เลือกใช้ค่าขีดแบ่งค่าเฉลี่ยพลังงานที่ 1.5 ในการทดลอง

เมื่อทดลองใช้ขั้นตอนวิธีที่นำเสนอกับชุดข้อมูลทดสอบเสียงพูดตัวเลขภาษาไทยและประโยคภาษาไทยทั่วไปได้ผลการทดลองดังนี้

ตารางที่ 4.7 ค่าความแม่นยำในการรู้จำเสียงพูดระหว่างขั้นตอนวิธีการลบสเปกตรัมแบบปรับค่าถ่วงน้ำหนักได้ ขั้นตอนวิธีการลบสเปกตรัมแบบการประมาณแมกซิมัมไลค์ลิฮูดและกรณีไม่ใช้ขั้นตอนวิธีใดเลย เมื่อใช้ชุดข้อมูลทดสอบตัวเลขภาษาไทย

ขั้นตอนวิธี	อัตราส่วนสัญญาณต่อสัญญาณรบกวน(เดซิเบล)									
	30	**25**	**20**	**15**	**10**	**5**	**0**	**-5**	**-10**	**-15**
การลบสเปกตรัมแบบปรับค่าถ่วงน้ำหนักได้	89.40	79.46	61.19	45.00	30.60	22.98	19.05	13.81	9.67	9.49
การลบสเปกตรัมแบบการประมาณแมกซิมัมไลค์ลิฮูด	89.76	79.70	60.00	42.98	26.25	17.38	14.82	13.81	9.93	8.23
ไม่ใช้ขั้นตอนวิธีใดเลย	86.07	64.64	40.36	22.44	15.95	8.93	8.69	10.12	9.05	9.05

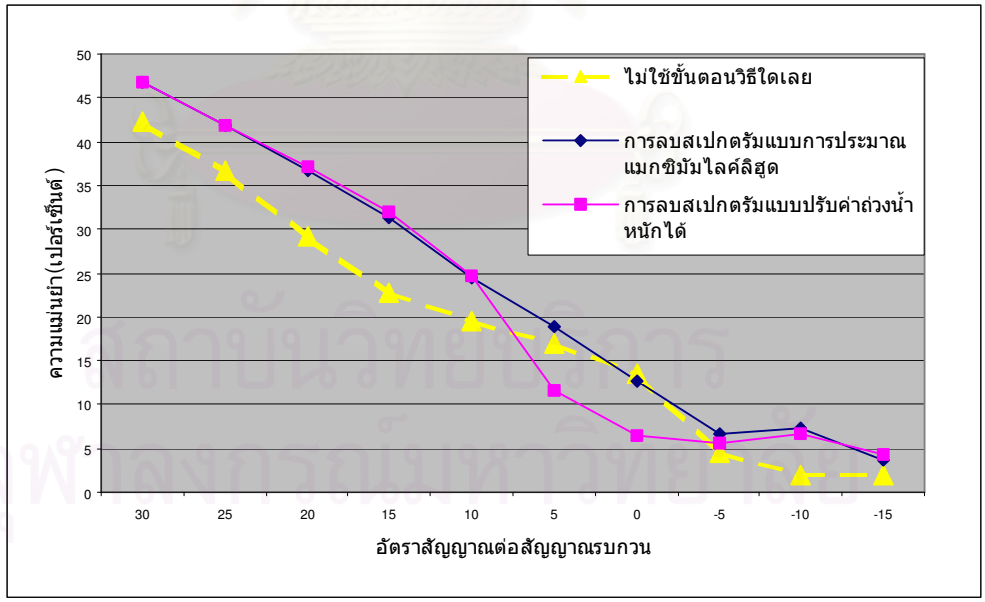


รูปที่ 4.8 กราฟแสดงค่าความแม่นยำในการรู้จำเสียงพูดตัวเลขระหว่างขั้นตอนวิธีการลบสเปกตรัมแบบปรับค่าถ่วงน้ำหนักได้ ขั้นตอนวิธีการลบสเปกตรัมแบบการประมาณแมกซิมัมไลค์ลิฮูดและกรณีไม่ใช้ขั้นตอนวิธีใดเลย เมื่อใช้ชุดข้อมูลทดสอบประโยคตัวเลขภาษาไทย

จากการทดลองกับชุดข้อมูลทดสอบพบว่าขั้นตอนวิธีที่นำเสนอ สามารถเพิ่มความแม่นยำโดยรวมในการรู้จำเสียงพูดได้ดีกว่าขั้นตอนวิธีการลบสเปกตรัมแบบการประมาณแม็กซิมัมไลค์ลิฮูด ถึง 17.79%

ตารางที่ 4.8 ค่าความแม่นยำในการรู้จำเสียงพูดระหว่างขั้นตอนวิธีการลบสเปกตรัมแบบปรับค่าถ่วงน้ำหนักได้ ขั้นตอนวิธีการลบสเปกตรัมแบบการประมาณแมกซิมัมไลค์ลิฮูดและกรณีไม่ใช้ขั้นตอนวิธีใดเลย เมื่อใช้ชุดข้อมูลทดสอบประโยคภาษาไทยทั่วไป

ขั้นตอนวิธี	อัตราส่วนสัญญาณต่อสัญญาณรบกวน(เดซิเบล)									
	30	25	20	15	10	5	0	-5	-10	-15
การลบสเปกตรัมแบบปรับค่าถ่วงน้ำหนักได้	46.79	41.78	37.02	32.01	24.69	11.51	6.48	5.61	6.60	4.30
การลบสเปกตรัมแบบการประมาณแมกซิมัมไลค์ลิฮูด	46.78	41.75	36.73	31.39	24.43	18.85	12.69	6.62	7.29	3.58
ไม่ใช้ขั้นตอนวิธีใดเลย	42.19	36.63	29.15	22.69	19.43	16.89	13.42	4.50	1.96	1.96



รูปที่ 4.9 กราฟแสดงค่าความแม่นยำในการรู้จำเสียงพูดระหว่างขั้นตอนวิธีการลบสเปกตรัมแบบปรับค่าถ่วงน้ำหนักได้ ขั้นตอนวิธีการลบสเปกตรัมแบบการประมาณแมกซิมัมไลค์ลิฮูดและกรณีไม่ใช้ขั้นตอนวิธีใดเลย เมื่อใช้ชุดข้อมูลทดสอบประโยคภาษาไทยทั่วไป

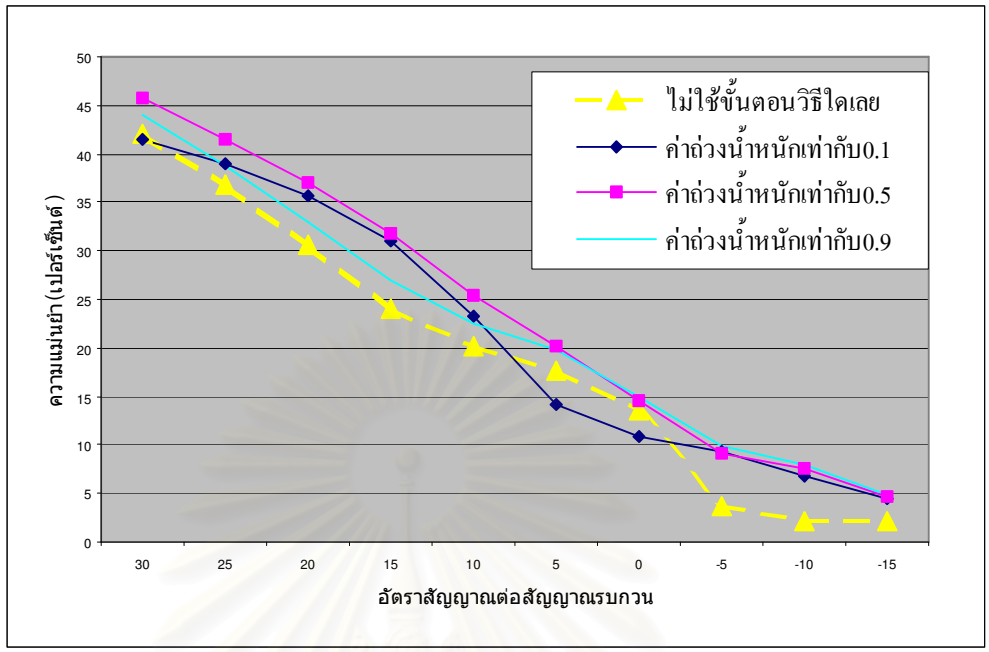
จากผลการทดลองด้วยชุดข้อมูลทดสอบประโยคภาษาไทยทั่วไปพบว่า ในกรณีสัญญาณเสียงพูดที่พิจารณามีค่าสัญญาณต่อสัญญาณรบกวนสูงหรือมีสัญญาณรบกวนต่ำ ค่าความแม่นยำในการรู้จำเสียงพูดที่ใช้ขั้นตอนวิธีการลบสเปกตรัมแบบปรับค่าถ่วงน้ำหนักได้นั้น มีค่าใกล้เคียงกับค่าความแม่นยำในการรู้จำเสียงพูด ที่ใช้ขั้นตอนวิธีการลบสเปกตรัมแบบการประมาณแมกซิมัมไลค์ลิฮูดอย่างมาก แต่ในกรณีที่สัญญาณเสียงพูดที่พิจารณามีค่าสัญญาณต่อสัญญาณรบกวนต่ำหรือมีสัญญาณรบกวนสูง ค่าความแม่นยำในการรู้จำเสียงพูดที่ใช้ขั้นตอนวิธีการลบสเปกตรัมแบบปรับค่าถ่วงน้ำหนักได้ มีค่าน้อยกว่ากรณีที่ใช้ขั้นตอนวิธีการลบสเปกตรัมแบบการประมาณแมกซิมัมไลค์ลิฮูด

ผลการทดลองที่ได้จากชุดข้อมูลประโยคภาษาไทยทั่วไปอาจเกิดจากสาเหตุดังนี้

1. การเพิ่มความสำคัญในการลดสัญญาณรบกวนทำให้สัญญาณเสียงพูดบิดเบือน ทำให้ไม่สามารถทำการรู้จำได้ดีขึ้น
2. แบบจำลองภาษาของประโยคทั่วไปมีความซับซ้อนมากกว่าแบบจำลองภาษาของตัวเลข โดยกรณีแบบจำลองภาษาของตัวเลขนั้น ถึงแม้สัญญาณเสียงพูดจะถูกบิดเบือนไป แต่ระบบยังคงสามารถรู้จำตัวเลขดังกล่าว
3. การประมาณค่าเฉลี่ยพลังงานสัญญาณรบกวนจากสัญญาณเสียงพูดในช่วง 100 มิลลิวินาทีแรกของสัญญาณเสียงพูดอาจไม่เหมาะสมกับข้อมูลทุกชุด

ตารางที่ 4.9 ความแม่นยำในการรู้จำเสียงพูดตัวเลขที่ค่าอัตราส่วนสัญญาณต่อสัญญาณรบกวนที่ค่าถ่วงน้ำหนักต่างๆสำหรับชุดข้อมูลประโยคภาษาไทยทั่วไป

ค่าถ่วงน้ำหนัก (β)	**อัตราส่วนสัญญาณต่อสัญญาณรบกวน(เดซิเบล)**									
	30	**25**	**20**	**15**	**10**	**5**	**0**	**-5**	**-10**	**-15**
0.1	41.53	38.91	35.63	30.94	23.29	14.22	10.77	9.26	6.86	4.52
0.3	43.40	40.30	38.28	32.74	27.19	19.62	13.50	9.22	7.34	4.58
0.5	45.81	41.51	37.04	31.86	25.31	20.21	14.60	9.14	7.56	4.72
0.7	45.42	40.73	35.10	29.13	24.12	19.21	14.62	9.95	7.79	4.97
0.9	43.97	38.71	33.00	27.00	22.54	19.80	14.99	9.87	7.87	4.86
ไม่ใช้ขั้นตอนวิธีใดเลย	41.97	36.85	30.70	23.94	20.25	17.68	13.56	3.75	2.20	2.20



รูปที่ 4.10 กราฟแสดงผลการทดลองเปลี่ยนค่าถ่วงน้ำหนักต่อความแม่นยำในการรู้จำเสียงพูดกรณีสัญญาณเสียงพูดประโยคทั่วไป

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 5

สรุปผลการวิจัย

สรุปการวิจัย

ในระบบรู้จำเสียงพูดนั้น สัญญาณรบกวนมีผลกระทบทำให้ประสิทธิภาพหรือความแม่นยำในการรู้จำเสียงพูดนั้นมีค่าน้อยลง อย่างไรก็ตามนักวิจัยยังคงศึกษาและพัฒนาขั้นตอนวิธีต่างๆในการลดสัญญาณรบกวนจากสัญญาณเสียงพูดที่ตรวจวัดได้ โดยมีจุดประสงค์เพื่อใช้ขั้นตอนวิธีดังกล่าวในการทำการลดสัญญาณรบกวนที่อยู่ในสัญญาณเสียงพูดให้สะอาดมากที่สุด ก่อนที่สัญญาณเสียงพูดดังกล่าวจะผ่านเข้ากระบวนการรู้จำเสียงพูดต่อไป ขั้นตอนวิธีที่ได้รับความนิยมและถูกศึกษาอย่างกว้างขวางได้แก่ ขั้นตอนวิธีการลบสเปกตรัม การกรองวีเนอร์ เป็นต้น โดยในขั้นตอนวิธีการลบสเปกตรัมนั้นได้ถูกพัฒนาและประยุกต์ในการใช้งานอย่างกว้างขวาง ได้แก่ ขั้นตอนวิธีการลบสเปกตรัมแบบการประมาณแม็กซิมัมไลค์ลิฮูด ขั้นตอนวิธีการลบสเปกตรัมแบบตัวกรองลดสัญญาณรบกวน โดยในงานวิจัยนี้ได้ทำการเปรียบเทียบค่าความแม่นยำในการรู้จำเสียงพูดในขั้นตอนวิธีต่างๆ พบว่า ขั้นตอนวิธีการลบสเปกตรัมแบบการประมาณแม็กซิมัมไลค์ลิฮูดให้ค่าความแม่นยำสูงสุด อย่างไรก็ตาม เมื่อพิจารณาจากสมการของขั้นตอนวิธีการลบสเปกตรัมแบบการประมาณแม็กซิมัมไลค์ลิฮูดพบว่า การให้ความสำคัญในส่วนของการลดสัญญาณรบกวนจะคงที่เสมอ ไม่ว่าสัญญาณเสียงพูดที่พิจารณาจะมีขนาดของสัญญาณรบกวนมากเพียงใดก็ตาม งานวิจัยนี้จึงเสนอขั้นตอนวิธีการลบสเปกตรัมแบบปรับค่าถ่วงน้ำหนักได้ โดยค่าถ่วงน้ำหนักหมายถึงการให้ความสำคัญสัญญาณเสียงพูดที่ตรวจหาได้ งานวิจัยนี้เริ่มต้นด้วยการทดลองเปลี่ยนค่าถ่วงน้ำหนักเพื่อศึกษาถึงแนวโน้มการเปลี่ยนแปลงของค่าความแม่นยำในการรู้จำเสียงพูด โดยพบว่าที่สัญญาณรบกวนมากๆระบบควรเลือกใช้ค่าถ่วงน้ำหนักให้มีค่าน้อยที่สุด ซึ่งงานวิจัยนี้เลือกใช้ค่าถ่วงน้ำหนักเท่ากับ 0.1 และเลือกใช้ค่าถ่วงน้ำหนักเท่ากับ 0.5 ซึ่งให้ค่าความแม่นยำสูงและมีแนวโน้มการลดลงของค่าความแม่นยำ ในกรณีสัญญาณรบกวนมีค่าสูงขึ้นดีที่สุด ความสัมพันธ์ระหว่างค่าถ่วงน้ำหนักและค่าเฉลี่ยพลังงานสัญญาณรบกวนถูกนำเสนอเพื่อใช้เป็นตัวกำหนดค่าถ่วงน้ำหนัก โดยฟังก์ชันความสัมพันธ์ระหว่างค่าถ่วงน้ำหนักและค่าเฉลี่ยพลังงานสัญญาณรบกวน ควรเลือกใช้ฟังก์ชันเชิงเส้นซึ่งมีประสิทธิภาพในการลดสัญญาณรบกวนดีกว่าการใช้ฟังก์ชันแบบขั้นบันได เนื่องจากการใช้ฟังก์ชันขั้นบันได ทำให้ประสิทธิภาพในการลดสัญญาณรบกวนแย่ลงเมื่อปริมาณสัญญาณรบกวนเพิ่มขึ้นและยังไม่เกินค่าขีดแบ่งค่าเฉลี่ยพลังงานสัญญาณรบกวน เนื่องจากช่วงดังกล่าวระบบทำการใช้ค่าถ่วงน้ำหนักคงที่ ที่ 0.5 ซึ่งจะได้ค่าความแม่นยำเทียบเท่ากับการใช้ขั้นตอนวิธีการลบ

สเปกตรัมแบบการประมาณแม็กซิมัมไลค์ลิฮูด โดยในขั้นตอนวิธีที่นำเสนอนั้นระบบควรเลือกใช้ค่าขีดแบ่งค่าเฉลี่ยพลังงานสัญญาณรบกวนอยู่ในช่วงระหว่าง 1.0 ถึง 2.0 จึงจะให้ผลในการรู้จำดีที่สุด โดยงานวิจัยนี้เลือกใช้ค่าขีดแบ่งค่าเฉลี่ยพลังงานเท่ากับ 1.5 ซึ่งผลในการรู้จำดีที่สุดสำหรับชุดข้อมูลในการพัฒนา โดยเมื่อทดลองใช้ขั้นตอนวิธีดังกล่าวกับชุดข้อมูลทดสอบประโยคตัวเลขภาษาไทย พบว่าค่าความแม่นยำในการรู้จำเสียงพูดโดยรวมเพิ่มขึ้น 17.79% แต่เมื่อใช้ขั้นตอนวิธีดังกล่าวกับชุดข้อมูลประโยคภาษาไทยทั่วไป พบว่าขั้นตอนวิธีดังกล่าวไม่สามารถเพิ่มความแม่นยำในการรู้จำเสียงได้ดีเทียบเท่ากรณีใช้ชุดข้อมูลตัวเลขโดด โดยอาจเกิดขึ้นเนื่องจากความซับซ้อนของรูปแบบภาษาของระบบรู้จำเสียงพูดที่ใช้ การบิดเบือนของสัญญาณเสียงพูดหรือการประมาณสัญญาณรบกวนที่ผิดพลาด

ข้อจำกัดของงานวิจัย

ขั้นตอนวิธีการลบสเปกตรัมแบบปรับค่าถ่วงน้ำหนักได้ ถูกพัฒนาขึ้นเพื่อใช้ในการลดสัญญาณรบกวนเติมแต่งแบบคงที่เท่านั้น โดยสัญญาณรบกวนและสัญญาณเสียงพูดจะต้องไม่มีอัตสหสัมพันธ์กัน

ข้อเสนอแนะ

1. ขั้นตอนวิธีในการลดสัญญาณรบกวนนั้น อาจไม่ส่งผลใดๆต่อความแม่นยำในการรู้จำเสียงพูด หากสัญญาณเสียงพูดดังกล่าวมีปริมาณสัญญาณรบกวนอยู่มากเกินไป
2. น่าจะมีการทดลองการประยุกต์ใช้ขั้นตอนวิธีการลบสเปกตรัมแบบปรับค่าถ่วงน้ำหนักได้ ในกรณีสัญญาณรบกวนเติมแต่งแบบไม่คงที่ ซึ่งน่าสามารถช่วยในการลดสัญญาณรบกวนได้ดีขึ้น
3. ควรมีการทดลองใช้ขั้นตอนวิธีนี้ร่วมกับขั้นตอนวิธีอื่นๆ เช่น ขั้นตอนวิธีการตรวจหากิจกรรมเสียงพูด ขั้นตอนการลดสัญญาณเสียงสะท้อน เป็นต้น เพื่อศึกษาว่าระบบจะสามารถเพิ่มความแม่นยำในการรู้จำเสียงพูดได้หรือไม่
4. ควรมีการทดลองใช้ฟังก์ชันความสัมพันธ์ระหว่างสัญญาณรบกวนและค่าถ่วงน้ำหนักแบบเชิงเส้นในรูปแบบอื่นเพื่อศึกษาถึงข้อดีเสียของฟังก์ชันแต่ละแบบ

คุณค่าทางวิชาการของงานวิจัย

1. ได้ขั้นตอนวิธีใหม่ที่สามารถเพิ่มความแม่นยำในการรู้จำเสียงพูด โดยการลดปริมาณสัญญาณรบกวนในสัญญาณเสียงพูดที่พิจารณา

2. การใช้แนวคิดในการปรับค่าถ่วงน้ำหนัก โดยพิจารณาจากการตรวจหาสัญญาณรบกวนขนาดต่างๆกัน

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

รายการอ้างอิง

[1] Quatieri, T.F. Discrete-Time Speech Signal Processing: Principle and Practice. New Jersy: Prentice Hall, 2002, pp 665-708.

[2] McAulay, R.J., and Malpass, M.L. Speech Enhancement Using a Soft-Decision Maximum Likelihood Noise Suppression Filter. IEEE Transaction on Acoustics, Speech, and Signal Processing. (1990): 137-145.

[3] Kermorvant, C. A Comparison of Noise Reduction Techniques for Robust Speech Recognition. IDIAP Research Report, RR-99-10. 1999.

[4] Krishnamoorthy, P., and Prasanna, S.R.M. Modified Spectral Subtraction Method for Enhancement of Noisy Speech, Third International Conference on Intelligent Sensing and Information Processing. (2005): 146-150.

[5] Harsha, B.V. A Noise Robust Speech Activity Detection Algorithm. In proceeding of International Symposium on Intelligent Multimedia, Video, and Speech Processing. (2004): 322-325.

[6] Denda, Y., Nishiura, T., Kawahara, H., and Irino, T. Speech Recognition with Wavelet Spectral Subtraction in Real Noisy Environment. In Proceeding of International Conference in Signal Processing. (2004): 638-641.

[7] Lin, L., Holms, W.H., and Ambikairajah, E. Subband Noise Estimation for Speech Enhancement Using A Perceptual Wiener Filter. In Proceeding of Acoustics, Speech, and Signal Processing. (2003): 80-83.

[8] Lin, L., Holms, W.H., and Ambikairajah, E. Adaptive Noise Estimation for Speech Enhancement. Electronics Letters. 39 (May 2003): 754-755

[9] Rangachari, S., Loizou, P.C., and Hu, Y. A Noise Estimation Algorithm with Rapid Adaptation for Highly Non-Stationary Environments. International Conference on Acoustics, Speech, and Signal Processing. (2004): 305-308.

[10] Sawit, K., Virach, S., Patcharika, C., Supphanat, K., and Nattanun, T. Thai Speech Corpus for Thai Speech Recognition. The Oriental COCOSDA. (2003): 54-61

ประวัติผู้เขียนวิทยานิพนธ์

นายเลอเกียรติ ชววิทยากุล เกิดเมื่อวันที่ 13 พฤษภาคม 2524 จบการศึกษาระดับปริญญาตรีจากคณะวิศวกรรมศาสตร์ ภาควิชาวิศวกรรมไฟฟ้า จุฬาลงกรณ์มหาวิทยาลัยในปีการศึกษา 2545 โดยขณะทำการศึกษาระดับปริญญาตรีได้เข้าร่วมชมรมจุฬาฯสู่ชุมชน และได้รับตำแหน่งประธานชมรมจุฬาฯสู่ชุมชนขณะศึกษาอยู่ชั้นปีที่ 3 หลังจากสอบวิทยานิพนธ์ผ่านแล้วได้ทำการเล่นดนตรีแสดงเป็นการขอบคุณคณะกรรมการสอบวิทยานิพนธ์

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย