

เทคนิคการลดสัญญาณรบกวนแบบไมโครโฟนสองตัวสำหรับการโทรศัพท์แบบไม่ใช้มือ

นางสาวหทัยชนก ชรรณเชิดชูพงษ์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต

สาขาวิชาวิศวกรรมไฟฟ้า ภาควิชาวิศวกรรมไฟฟ้า

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2555

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

บทคัดย่อและแฟ้มข้อมูลฉบับเต็มของวิทยานิพนธ์ตั้งแต่ปีการศึกษา 2554 ที่ให้บริการในคลังปัญญาจุฬาฯ (CUIR)

เป็นแฟ้มข้อมูลของนิสิตเจ้าของวิทยานิพนธ์ที่ส่งผ่านทางบัณฑิตวิทยาลัย

The abstract and full text of theses from the academic year 2011 in Chulalongkorn University Intellectual Repository (CUIR)

are the thesis authors' files submitted through the Graduate School.

TWO-MICROPHONE NOISE REDUCTION TECHNIQUES FOR HANDS-FREE
TELEPHONY

Miss Hathaichanok Thumchirdchupong

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

Department of Electrical Engineering

Faculty of Engineering

Chulalongkorn University

Academic Year 2012

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์

เทคนิคการลดสัญญาณรบกวนแบบไมโครโฟนสอง
ตัวสำหรับการโทรศัพท์แบบไม่ใช้มือ

โดย

นางสาวหทัยชนก ธรรมเชิดชูพงษ์

สาขาวิชา

วิศวกรรมไฟฟ้า

อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

รองศาสตราจารย์ ดร.นิสาชล ตั้งเสงี่ยมวิสัย

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้บัณฑิตวิทยาลัยเป็นส่วน
หนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรบัณฑิต

.....คณบดีคณะวิศวกรรมศาสตร์
(ศาสตราจารย์ ดร.บุญสม เลิศธีรชูวงศ์)

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.สุภาวดี อร่ามวิทย์)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(รองศาสตราจารย์ ดร.นิสาชล ตั้งเสงี่ยมวิสัย)

.....กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.ชาญชัย ปลื้มปิติวิริยะเวช)

.....กรรมการภายนอกมหาวิทยาลัย
(ผู้ช่วยศาสตราจารย์ ดร.เผ่าถัก ศิริสุข)

หทัยชนก ธรรมเชิดชูพงษ์ : เทคนิคการลดสัญญาณรบกวนแบบไมโครโฟนสองตัว
สำหรับการโทรศัพท์แบบไม่ใช้มือ. (TWO-MICROPHONE NOISE REDUCTION
TECHNIQUES FOR HANDS-FREE TELEPHONY) อ.ที่ปรึกษาวิทยานิพนธ์หลัก : รศ.
ดร.นิสาชล ตั้งเสงี่ยมวิสัย, 84 หน้า.

ในสภาพแวดล้อมรถยนต์ สมรรถนะของการโทรศัพท์แบบไม่ใช้มือถูกรบกวนด้วย
สัญญาณรบกวนต่างๆ ได้แก่ เสียงเครื่องยนต์ เสียงลม และเสียงพูดคุยจ่อแจ เป็นต้น เทคนิคการลด
สัญญาณรบกวนจึงมีความจำเป็นต่ออุปกรณ์ของระบบการโทรศัพท์แบบไม่ใช้มือ เพื่อปรับปรุง
คุณภาพของสัญญาณเสียงพูดและการสื่อสารทางเสียงที่มีความถูกต้อง

วิทยานิพนธ์นี้เสนอการใช้งานโครงสร้างแถบความถี่ย่อยกับเทคนิคการลดสัญญาณรบกวน
ที่ใช้ไมโครโฟน 2 ตัวที่มีอยู่แล้ว คือ เทคนิคการลบสเปกตรัมข้ามตัดแปร (Modified Cross Spectral
Subtraction, MCSS) และระบบการตัดออกสัญญาณรบกวนแบบปรับตัว (Adaptive Noise
Cancellation, ANC) ที่ใช้งานขั้นตอนวิธีดับเบิลแอฟฟายโปรเจกชัน (Double Affine Projection,
DAP) เทคนิค MCSS ใช้ลดสัญญาณรบกวนในช่วงแถบความถี่ต่ำ (ต่ำกว่า 1.5 kHz) เพื่อสมรรถนะ
การลดทอนสัญญาณรบกวนที่ดีที่สุด และเลือกใช้ขั้นตอนวิธี DAP ในช่วงแถบความถี่สูง (สูงกว่า
1.5 kHz) เพื่อมีให้ข้อมูลความถี่ของสัญญาณเสียงพูดที่ปรับปรุงแล้วถูกรบกวน

ถัดมาขั้นตอนวิธี DAP ที่มีอันดับการโปรเจกชันสูงและต่ำ ถูกนำเสนอโดยใช้โครงสร้าง
แถบความถี่ย่อย อันดับการโปรเจกชันสูงถูกเลือกใช้งานในช่วงแถบความถี่ต่ำ ซึ่งมีองค์ประกอบ
ทางความถี่ของข้อมูลของเสียงพูดอยู่มากเพื่อการลดทอนสัญญาณรบกวนที่ดี ในช่วงความถี่สูงซึ่งมี
องค์ประกอบทางความถี่ของข้อมูลของสัญญาณเสียงพูดอยู่น้อยกว่า อันดับการโปรเจกชันต่ำถูก
เลือกใช้งานเพื่อลดความซับซ้อนในการคำนวณโดยรวมลง เมื่อเปรียบเทียบกับการทำงานของ
ขั้นตอนวิธี DAP โดยลำพังที่เลือกอันดับการโปรเจกชันสูง

นอกจากนี้ เทคนิคการสร้างทดแทนฮาร์โมนิกของสัญญาณเสียงพูดถูกนำเสนอใน
วิทยานิพนธ์ ฟังก์ชันไม่เชิงเส้นถูกนำเสนอเพื่อสร้างทดแทนฮาร์โมนิกของเสียงพูดที่ถูกปรับปรุง
ด้วยเทคนิคการลดสัญญาณรบกวนแล้ว การจำลองระบบแสดงให้เห็นถึงสมรรถนะที่ดีขึ้นของ
แนวคิดต่างๆที่นำเสนอในด้านการลดทอนสัญญาณรบกวน และความสามารถในการรักษาไว้ซึ่ง
องค์ประกอบทางความถี่ของข้อมูลสัญญาณเสียงพูด พร้อมกันนี้ผลทดสอบฟังก์ชันแสดงให้เห็นถึง
คุณภาพของสัญญาณเสียงพูดที่ถูกรักษาไว้ด้วย

ภาควิชา วิศวกรรมไฟฟ้าลายมือชื่อ.....

สาขาวิชา วิศวกรรมไฟฟ้าลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก.....

ปีการศึกษา 2555

5470437921 : MAJOR ELECTRICAL ENGINEERING

KEYWORDS : NOISE REDUCTION / HANDS-FREE TELEPHONY / SUBBAND / SPEECH HARMONICS / ADAPTIVE NOISE CANCELLATION

HATHAICHANOK THUMCHIRDCHUPONG : TWO-MICROPHONE NOISE REDUCTION TECHNIQUES FOR HANDS-FREE TELEPHONY. ADVISOR : ASSOC. PROF. NISACHON TANGSANGIUMVISAI, Ph.D., 84 pp.

In a car environment, the performance of hands-free telephony is normally disturbed by engine noise, wind noise, and babble noise, etc. Noise Reduction (NR) techniques are therefore necessary for hands-free car kits in order to improve speech quality and speech intelligibility of the voice communication systems. This thesis proposes to utilise two existing two-microphone NR techniques; namely the Modified Cross Spectral Subtraction (MCSS) method and the Adaptive Noise Cancellation (ANC) employing the Double Affine Projection (DAP) algorithm, based upon a subband structure. The MCSS technique is employed in the low-frequency band (below 1.5 kHz) for superior noise attenuation performance. As for frequencies beyond 1.5 kHz, the DAP algorithm is suggested so that high frequency components of the enhanced speech spectrum are not disturbed. Next, a subband structure between the DAP algorithm with low and high projection order is also proposed. The DAP algorithm with high projection-order is chosen to be operated in the low-frequency band, where dominant cues of speech spectral components are located, to obtain excellent noise attenuation performance. As for the high-frequency band, where less-dominant frequency information of speech spectrum is located, a low projection-order of the DAP algorithm will be operated. The overall computational complexity can be reduced, as compared to the stand-alone DAP algorithm with high projection-order. In addition, a speech Harmonic Regeneration (HR) technique is proposed. A non-linear function is proposed to regenerate the harmonic components of the enhanced speech signal in order to improve the speech quality. Simulation results suggest improved performance of the proposed NR schemes with and without use of the proposed speech HR technique in terms of noise attenuation and preservation of speech spectral components of the enhanced speech signals. Furthermore, a subjective listening test is given to ensure the speech quality preservation of proposed techniques.

Department :Electrical Engineering..... Student's Signature.....

Field of Study :Electrical Engineering..... Advisor's Signature.....

Academic Year :2012.....

กิตติกรรมประกาศ

งานวิจัยฉบับนี้สำเร็จลุล่วงไปด้วยดีเนื่องจากผู้วิจัยได้รับความช่วยเหลือ ดูแลเอาใจใส่เป็นอย่างดีจากหลายๆฝ่าย โดยเฉพาะอาจารย์ที่ปรึกษาวิทยานิพนธ์ รศ. ดร. นิสาชล ตั้งเสงี่ยมวิสัย ที่ได้ให้คำปรึกษาแนะนำ ติดตามความก้าวหน้าในการดำเนินการวิจัย พร้อมทั้ง ให้กำลังใจและแรงบันดาลใจ ในการทำวิทยานิพนธ์เสมอมา ผู้วิจัยขอขอบพระคุณเป็นอย่างสูงไว้ ณ โอกาสนี้

ขอขอบพระคุณอาจารย์กรรมการสอบวิทยานิพนธ์ที่ได้ให้ข้อเสนอแนะ และแนวคิดต่างๆ ที่เป็นประโยชน์ และขอขอบพระคุณอาจารย์ทุกท่าน ที่ได้สั่งสอน และ ให้ความรู้เสมอมา

สุดท้ายนี้ ขอขอบคุณบิดา มารดา ครอบครัว ญาติ และ เพื่อนๆ พี่ น้อง ทุกคนที่เป็นกำลังใจและให้การสนับสนุนด้วยดีเสมอ

สารบัญ

| | หน้า |
|---|------|
| บทคัดย่อภาษาไทย..... | ง |
| บทคัดย่อภาษาอังกฤษ..... | จ |
| กิตติกรรมประกาศ..... | ฉ |
| สารบัญ..... | ช |
| สารบัญตาราง..... | ญ |
| สารบัญรูป..... | ฎ |
| บทที่ | |
| 1. บทนำ..... | 1 |
| 1.1 ความเป็นมาและความสำคัญของปัญหา..... | 1 |
| 1.2 วัตถุประสงค์ของการวิจัย..... | 3 |
| 1.3 ขอบเขตของการวิจัย..... | 3 |
| 1.4 ประโยชน์ที่คาดว่าจะได้รับ..... | 4 |
| 1.5 วิธีดำเนินการวิจัย..... | 4 |
| 2. งานวิจัยที่เกี่ยวข้อง..... | 5 |
| 2.1 เทคนิคการลดเสียงรบกวนที่ประมวลผลในโดเมนความถี่..... | 5 |
| 2.1.1 เทคนิคการลบสเปกตรัม (Spectrum Subtraction, SS) | 6 |
| 2.1.2 การจำลองสัญญาณสำหรับระบบที่มีไมโครโฟน 2 ตัว..... | 14 |
| 2.1.3 เทคนิคการลบสเปกตรัมข้าม (Cross Spectrum Subtraction, CSS) | 15 |
| 2.1.4 เทคนิคการลบสเปกตรัมข้ามดัดแปร (Modification Cross Spectrum Subtraction, MCSS) | 15 |
| 2.2 เทคนิคการลดเสียงรบกวนที่ประมวลผลในโดเมนเวลา..... | 19 |
| 2.2.1 ขั้นตอนวิธีกำลังสองเฉลี่ยน้อยที่สุด (Least Mean Square, LMS) | 20 |

| | | |
|-------|--|----|
| 2.2.2 | ขั้นตอนวิธีกำลังสองเฉลี่ยน้อยที่สุดแบบนอร์มาไลซ์ (Normalize Least Mean Square, NLMS) | 21 |
| 2.2.3 | ขั้นตอนวิธีกำลังสองน้อยที่สุดแบบวนซ้ำ (Recursive Least Square, RLS) | 21 |
| 2.2.4 | ขั้นตอนวิธีแอฟฟายโปรเจกชัน (Affine Projection, AP) | 22 |
| 2.2.5 | ขั้นตอนวิธีดับเบิลแอฟฟายโปรเจกชัน (Double Affine Projection, DAP) | 23 |
| 2.3 | การสร้างทดแทนฮาร์โมนิกของเสียงพูด (Harmonic Regeneration, HR, Technique) | 26 |
| 3. | เทคนิคการลดสัญญาณรบกวนที่นำเสนอในวิทยานิพนธ์..... | 29 |
| 3.1 | สัญญาณรบกวนประเภทต่างๆ..... | 30 |
| 3.2 | เทคนิคการสร้างทดแทนฮาร์โมนิกของสัญญาณเสียงที่นำเสนอ..... | 32 |
| 3.3 | ขั้นตอนวิธี DAP แบบหลายอันดับ..... | 34 |
| 3.4 | เทคนิคร่วมระหว่างขั้นตอนวิธี DAP และ MCSS..... | 36 |
| 4. | ผลการวิเคราะห์ข้อมูล..... | 38 |
| 4.1 | การจำลองระบบ..... | 38 |
| 4.2 | ตัวชี้วัดประสิทธิภาพ..... | 41 |
| 4.3 | การทดลองที่ 1 ประสิทธิภาพของวิธี MCSS, วิธี HR[10] และ วิธีนำเสนอ..... | 44 |
| 4.4 | การทดลองที่ 2 ผลกระทบของอันดับการโปรเจกชันของวิธี DAP..... | 53 |
| 4.5 | การทดลองที่ 3 ประสิทธิภาพของขั้นตอนวิธี DAP แบบหลายอันดับ..... | 54 |
| 4.6 | การทดลองที่ 4 ประสิทธิภาพของเทคนิคร่วมระหว่างขั้นตอนวิธี DAP และ MCSS..... | 61 |
| 5. | สรุปผลการทดลอง..... | 72 |
| 5.1 | สรุปผลการทดลอง..... | 72 |
| 5.2 | ข้อเสนอแนะ..... | 74 |
| | รายการอ้างอิง..... | 75 |
| | ภาคผนวก..... | 77 |

| | หน้า |
|---------------------------------|------|
| ภาคผนวก ก..... | 78 |
| ภาคผนวก ข..... | 82 |
| ประวัติผู้เขียนวิทยานิพนธ์..... | 84 |

สารบัญตาราง

| ตารางที่ | | หน้า |
|----------|--|------|
| 2.1 | ความซับซ้อนในการคำนวณ (Computation complexity)..... | 25 |
| 2.2 | ความซับซ้อนในการคำนวณเนื่องจากการคูณ | 26 |
| 4.1 | ระดับคะแนนความคิดเห็นเฉลี่ย (Mean Opinion Score, MOS)..... | 43 |
| 4.2 | SegSNR ของวิธี MCSS, วิธี HR[10] และ วิธี HR ที่นำเสนอ..... | 51 |
| 4.3 | Log Spectral Distance ของวิธี MCSS, วิธี HR[10] และ วิธี HR ที่เสนอ..... | 52 |
| 4.4 | Segmental Signal-To-Noise Ratio (SegSNR) | 68 |
| 4.5 | Log Spectral Distance (LSD) | 69 |
| 4.6 | Mean opines score test (MOS), Female speech..... | 70 |
| 4.7 | Mean opines score test (MOS), Male speech..... | 71 |

สารบัญรูป

| รูปที่ | | หน้า |
|--------|--|------|
| 2.1 | ขั้นตอนวิธีการลบสเปกตรัม..... | 10 |
| 2.2 | การจำลองสัญญาณสำหรับระบบที่มีไมโครโฟน 2 ตัว..... | 15 |
| 2.3 | วิธีลดสัญญาณรบกวนที่ใช้ไมโครโฟน 2 ตัว..... | 16 |
| 2.4 | การหาแบบจำลองของระบบ..... | 19 |
| 2.5 | ขั้นตอนวิธีดับเบิ้ลแอฟฟายโปรเจกชัน (DAP Algorithm) | 24 |
| 2.6 | วิธีการสร้างฮาร์โมนิกตามงานวิจัย [10] | 26 |
| 2.7 | สัญญาณเสียงพูดที่ถูกสร้างทดแทนฮาร์โมนิกตามงานวิจัย [10] (1 เฟรม) $\hat{s}(n) \geq 0$ (เส้นทึบ) และ $p(\hat{s}(n))$ (เส้นประ) | 27 |
| 2.8 | สัญญาณเสียงพูดที่ถูกสร้างทดแทนฮาร์โมนิกตามงานวิจัย [10] (1 เฟรม) $\hat{s}(n) \geq 0$ (เส้นทึบ) และ $p(\hat{s}(n))$ (เส้นประ)..... | 28 |
| 3.1 | สเปกโตรแกรมของ ก) สัญญาณรบกวนขาวแบบเกาส์ (White Gaussian Noise, WGN), ข) สัญญาณรบกวนแบบเสียงพูดคุย (Babble Noise) และ ค) สัญญาณรบกวนในรถยนต์ (Car Noise)..... | 31 |
| 3.2 | เทคนิคการสร้างทดแทนฮาร์โมนิกของสัญญาณเสียงที่นำเสนอ..... | 32 |
| 3.3 | สัญญาณเสียงพูดที่ถูกสร้างทดแทนฮาร์โมนิกด้วยวิธีที่นำเสนอ (1 เฟรม) $s'_{harmonic}(n)$ (เส้นทึบ) และ $p(\hat{s}(n))$ (เส้นประ) เมื่อ $\zeta = 10$ และ $\eta = 0.2$ | 33 |
| 3.4 | ขั้นตอนวิธี DAP แบบหลายอันดับที่นำเสนอ..... | 34 |
| 3.5 | เทคนิคร่วมระหว่างขั้นตอนวิธี DAP และ MCSS ที่นำเสนอ..... | 36 |
| 4.1 | สเปกโตรแกรมของ ก) สัญญาณเสียงพูดเพศหญิงพูดว่า 'A bicycle has two wheels.', ข) สัญญาณเสียงพูดที่ปนด้วยสัญญาณรบกวนขาวแบบเกาส์ (White Gaussian Noise, WGN), ค) สัญญาณเสียงพูดที่ปนด้วยสัญญาณรบกวนแบบเสียงพูดคุย (Babble Noise) และ ง) สัญญาณเสียงพูดที่ปนด้วยเสียงรบกวนในรถยนต์ (Car Noise) SNR ขนาด 10 dB..... | 39 |

4.2 สเปกโตรแกรมของ ก) สัญญาณเสียงพูดเพศชายพูดว่า ‘A bicycle has two wheels.’, ข) สัญญาณเสียงพูดที่ปนด้วยสัญญาณรบกวนขาวแบบเกาส์ (White Gaussian Noise, WGN), ค) สัญญาณเสียงพูดที่ปนด้วยสัญญาณรบกวนแบบเสียงพุดคุย (Babble Noise) และ ง) สัญญาณเสียงพูดที่ปนด้วยเสียงรบกวนในรถยนต์ (Car Noise) SNR ขนาด 10 dB..... 40

4.3 สเปกโตรแกรมของ ก) สัญญาณเสียงพูดต้นฉบับเสียงพูดเพศหญิง ‘A bicycle has two wheels.’, ข) สัญญาณเสียงพูดที่ถูกรบกวนด้วยสัญญาณรบกวนขาวแบบเกาส์ (WGN) 10 dB, ค) การลดเสียงรบกวนด้วยวิธี MCSS, ง) การเพิ่มฮาร์มอนิกของเสียงพูดด้วยวิธี MCSS + HR[10], จ) การเพิ่มฮาร์มอนิกของเสียงพูดด้วยวิธี MCSS + HR ที่เสนอ $\beta = 15$ และ $\gamma = 0.1$ 45

4.4 สเปกโตรแกรมของ ก) สัญญาณเสียงพูดต้นฉบับเสียงพูดเพศหญิง ‘A bicycle has two wheels.’, ข) สัญญาณเสียงพูดที่ถูกรบกวนด้วยเสียงพุดคุย (Babble) 10 dB, ค) การลดเสียงรบกวนด้วยวิธี MCSS, ง) การเพิ่มฮาร์มอนิกของเสียงพูดด้วยวิธี MCSS + HR[10], จ) การเพิ่มฮาร์มอนิกของเสียงพูดด้วยวิธี MCSS + HR ที่เสนอ $\beta = 15$ และ $\gamma = 0.1$ 46

4.5 สเปกโตรแกรมของ ก) สัญญาณเสียงพูดต้นฉบับเสียงพูดเพศหญิง ‘A bicycle has two wheels.’, ข) สัญญาณเสียงพูดที่ถูกรบกวนด้วยเสียงรบกวนในรถยนต์ (Car noise) 10 dB, ค) การลดเสียงรบกวนด้วยวิธี MCSS, ง) การเพิ่มฮาร์มอนิกของเสียงพูดด้วยวิธี MCSS + HR[10], จ) การเพิ่มฮาร์มอนิกของเสียงพูดด้วยวิธี MCSS + HR ที่เสนอ $\beta = 15$ และ $\gamma = 0.1$ 47

4.6 สเปกโตรแกรมของ ก) สัญญาณเสียงพูดต้นฉบับเสียงพูดเพศชาย ‘A bicycle has two wheels.’, ข) สัญญาณเสียงพูดที่ถูกรบกวนด้วยสัญญาณรบกวนขาวแบบเกาส์ (WGN) 10 dB, ค) การลดเสียงรบกวนด้วยวิธี MCSS, ง) การเพิ่มฮาร์มอนิกของเสียงพูดด้วยวิธี MCSS + HR[10], จ) การเพิ่มฮาร์มอนิกของเสียงพูดด้วยวิธี MCSS + HR ที่เสนอ $\beta = 15$ และ $\gamma = 0.1$ 48

- 4.7 สเปกโตรแกรมของ ก) สัญญาณเสียงพูดต้นฉบับเสียงพูดเพศชาย ‘A bicycle has two wheels.’, ข) สัญญาณเสียงพูดที่ถูกรบกวนด้วยเสียงพูดคุย (Babble) 10 dB, ค) การลดเสียงรบกวนด้วยวิธี MCSS, ง) การเพิ่มฮาร์โมนิกของเสียงพูดด้วยวิธี MCSS + HR[10], จ) การเพิ่มฮาร์โมนิกของเสียงพูดด้วยวิธี MCSS + HR ที่ เสนอ $\beta = 15$ และ $\gamma = 0.1$ 49
- 4.8 4.8 สเปกโตรแกรมของ ก) สัญญาณเสียงพูดต้นฉบับเสียงพูดเพศชาย ‘A bicycle has two wheels.’, ข) สัญญาณเสียงพูดที่ถูกรบกวนด้วยเสียงรบกวนในรถยนต์ (Car noise) 10 dB, ค) การลดเสียงรบกวนด้วยวิธี MCSS, ง) การเพิ่มฮาร์โมนิกของเสียงพูดด้วยวิธี MCSS + HR[10], จ) การเพิ่มฮาร์โมนิกของเสียงพูดด้วยวิธี MCSS + HR ที่ เสนอ $\beta = 15$ และ $\gamma = 0.1$ 50
- 4.9 สเปกโตรแกรมของ ก) สัญญาณเสียงพูดต้นฉบับเสียงพูดเพศหญิง ‘A bicycle has two wheels.’, ข) สัญญาณเสียงพูดที่ถูกรบกวนด้วยสัญญาณรบกวนขาวแบบเกาส์ (WGN) 10 dB, ค) การลดเสียงรบกวนด้วยวิธี DAP ที่ $p = 4$, ง) การลดเสียงรบกวนด้วยวิธี DAP ที่ $p = 20$, จ) ขั้นตอนวิธี DAP แบบหลายอันดับ.. 55
- 4.10 สเปกโตรแกรมของ ก) สัญญาณเสียงพูดต้นฉบับเสียงพูดเพศหญิง ‘A bicycle has two wheels.’, ข) สัญญาณเสียงพูดที่ถูกรบกวนด้วยเสียงพูดคุย (Babble Noise) 10 dB, ค) การลดเสียงรบกวนด้วยวิธี DAP ที่ $p = 4$, ง) การลดเสียงรบกวนด้วยวิธี DAP ที่ $p = 20$, จ) ขั้นตอนวิธี DAP แบบหลายอันดับ..... 56
- 4.11 สเปกโตรแกรมของ ก) สัญญาณเสียงพูดต้นฉบับเสียงพูดเพศหญิง ‘A bicycle has two wheels.’, ข) สัญญาณเสียงพูดที่ถูกรบกวนด้วยเสียงรบกวนในรถยนต์ (Car Noise) 10 dB, ค) การลดเสียงรบกวนด้วยวิธี DAP ที่ $p = 4$, ง) การลดเสียงรบกวนด้วยวิธี DAP ที่ $p = 20$, จ) ขั้นตอนวิธี DAP แบบหลายอันดับ.... 57
- 4.12 สเปกโตรแกรมของ ก) สัญญาณเสียงพูดต้นฉบับเสียงพูดเพศชาย ‘A bicycle has two wheels.’, ข) สัญญาณเสียงพูดที่ถูกรบกวนด้วยสัญญาณรบกวนขาวแบบเกาส์ (WGN) 10 dB, ค) การลดเสียงรบกวนด้วยวิธี DAP ที่ $p = 4$, ง) การลดเสียงรบกวนด้วยวิธี DAP ที่ $p = 20$, จ) ขั้นตอนวิธี DAP แบบหลายอันดับ..... 58

| | | |
|------|---|----|
| 4.13 | สเปกโตรแกรมของ ก) สัญญาณเสียงพูดต้นฉบับเสียงพูดเพศชาย ‘A bicycle has two wheels.’, ข) สัญญาณเสียงพูดที่ถูกรบกวนด้วยเสียงพูดคุย (Babble Noise) 10 dB, ค) การลดเสียงรบกวนด้วยวิธี DAP ที่ $p = 4$, ง) การลดเสียงรบกวนด้วยวิธี DAP ที่ $p = 20$, จ) ขั้นตอนวิธี DAP แบบหลายอันดับ..... | 59 |
| 4.14 | สเปกโตรแกรมของ ก) สัญญาณเสียงพูดต้นฉบับเสียงพูดเพศชาย ‘A bicycle has two wheels.’, ข) สัญญาณเสียงพูดที่ถูกรบกวนด้วยเสียงรบกวนในรถยนต์ (Car Noise) 10 dB, ค) การลดเสียงรบกวนด้วยวิธี DAP ที่ $p = 4$, ง) การลดเสียงรบกวนด้วยวิธี DAP ที่ $p = 20$, จ) ขั้นตอนวิธี DAP แบบหลายอันดับ..... | 60 |
| 4.15 | สเปกโตรแกรมของ ก) สัญญาณเสียงพูดต้นฉบับเสียงพูดเพศหญิง ‘A bicycle has two wheels.’, ข) สัญญาณเสียงพูดที่ถูกรบกวนด้วยสัญญาณรบกวนขาวแบบเกาส์ (WGN) 10 dB, ค) เทคนิคร่วมระหว่างขั้นตอนวิธี DAP และ MCSS. | 62 |
| 4.16 | สเปกโตรแกรมของ ก) สัญญาณเสียงพูดต้นฉบับเสียงพูดเพศหญิง ‘A bicycle has two wheels.’, ข) สัญญาณเสียงพูดที่ถูกรบกวนด้วยเสียงพูดคุย (Babble) 10 dB, ค) เทคนิคร่วมระหว่างขั้นตอนวิธี DAP และ MCSS..... | 63 |
| 4.17 | สเปกโตรแกรมของ ก) สัญญาณเสียงพูดต้นฉบับเสียงพูดเพศหญิง ‘A bicycle has two wheels.’, ข) สัญญาณเสียงพูดที่ถูกรบกวนด้วยเสียงรบกวนในรถยนต์ (Car Noise) 10 dB, ค) เทคนิคร่วมระหว่างขั้นตอนวิธี DAP และ MCSS | 64 |
| 4.18 | สเปกโตรแกรมของ ก) สัญญาณเสียงพูดต้นฉบับเสียงพูดเพศชาย ‘A bicycle has two wheels.’, ข) สัญญาณเสียงพูดที่ถูกรบกวนด้วยสัญญาณรบกวนขาวแบบเกาส์ (WGN) 10 dB, ค) เทคนิคร่วมระหว่างขั้นตอนวิธี DAP และ MCSS..... | 65 |
| 4.19 | สเปกโตรแกรมของ ก) สัญญาณเสียงพูดต้นฉบับเสียงพูดเพศชาย ‘A bicycle has two wheels.’, ข) สัญญาณเสียงพูดที่ถูกรบกวนด้วยเสียงพูดคุย (Babble) 10 dB, ค) เทคนิคร่วมระหว่างขั้นตอนวิธี DAP และ MCSS..... | 66 |
| 4.20 | สเปกโตรแกรมของ ก) สัญญาณเสียงพูดต้นฉบับเสียงพูดเพศชาย ‘A bicycle has two wheels.’, ข) สัญญาณเสียงพูดที่ถูกรบกวนด้วยเสียงรบกวนในรถยนต์ (Car Noise) 10 dB, ค) เทคนิคร่วมระหว่างขั้นตอนวิธี DAP และ MCSS | 67 |

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

คุณภาพของสัญญาณเสียงพูดที่ชัดเจนดีและสามารถเข้าใจได้ดี (Intelligibility) นั้นเป็นปัจจัยสำคัญของการสื่อสารทางเสียงพูด (Speech Communication) ซึ่งในสภาพแวดล้อมในรถยนต์นั้น สัญญาณเสียงพูดจะมีผลกระทบจากสัญญาณรบกวนพื้นหลัง (Background Noise) สัญญาณเสียงสะท้อน (Echo) สัญญาณรบกวนที่ปรากฏในรถยนต์ เช่น เสียงเครื่องยนต์, เสียงที่เกิดจากการกระทบระหว่างล้อกับถนน, เสียงจากลม, เสียงพูดคุย (Babble noise) และ เสียงจากเครื่องเสียงสเตอริโอ เป็นต้น สัญญาณรบกวนเหล่านี้ทำให้ประสิทธิภาพของการสื่อสารเป็นไปไม่ดีเท่าที่ควรและอาจทำให้เกิดการสื่อสารข้อมูลที่ผิดพลาดได้ ดังนั้นเทคนิคการลดสัญญาณรบกวนจะถูกนำมาใช้เพื่อเพิ่มประสิทธิภาพของการสื่อสารทางเสียงพูด โดยการลดปริมาณสัญญาณรบกวนที่ปนอยู่ออกไป อย่างไรก็ตามความพยายามในการลดสัญญาณรบกวนอาจทำให้สัญญาณเสียงพูดถูกลดทอนไปด้วย ดังนั้นจึงต้องคำนึงถึงสัญญาณเสียงพูดที่ได้จะต้องมีความบิดเบี้ยวต่ำและสามารถลดระดับของสัญญาณรบกวนได้ดี ตัวอย่างการใช้งานการลดสัญญาณรบกวน ได้แก่ เครื่องช่วยฟัง (Hearing Aids), ระบบการสัมมนาทางไกล (Teleconference System) และ การใช้โทรศัพท์เคลื่อนที่ด้วยอุปกรณ์แฮนด์ฟรี เป็นต้น

ในวิทยานิพนธ์นี้ พิจารณาที่การใช้งานโทรศัพท์เคลื่อนที่ด้วยอุปกรณ์แฮนด์ฟรีในรถยนต์ ตามพระราชบัญญัติจราจรทางบก พ.ศ. 2551 (ฉบับที่ 8) ซึ่งบังคับไม่ให้มีการใช้โทรศัพท์เคลื่อนที่ขณะขับรถยนต์ เว้นแต่ในกรณีการใช้โทรศัพท์เคลื่อนที่โดยใช้อุปกรณ์เสริมสำหรับการสนทนา โดยผู้ขับขี่ไม่ต้องถือหรือจับโทรศัพท์เคลื่อนที่นั้น[1] ทำให้การใช้งานโทรศัพท์เคลื่อนที่ด้วยอุปกรณ์แฮนด์ฟรีในรถยนต์แพร่หลาย เพื่อเพิ่มความปลอดภัยด้วยเช่นกัน

สำหรับการพัฒนาระบบการสื่อสารทางเสียงเพื่อลดสัญญาณรบกวนมีมากมายหลายวิธี สำหรับวิธีหนึ่งที่เป็นที่นิยมคือ วิธีการลบทางสเปกตรัม (Spectral Subtraction, SS Method) ซึ่งเป็นวิธีการลบค่าประมาณของสเปกตรัมของสัญญาณรบกวนออกจากสเปกตรัมของสัญญาณเสียงพูดที่ถูกรบกวน โดยมีข้อดีคือวิธี SS มีความซับซ้อนในการคำนวณต่ำ แต่อย่างไรก็ตามปัญหาที่เกิดขึ้นกับวิธีนี้คือมี 'Musical Noise' ซึ่งเป็นสัญญาณรบกวนที่เกิดขึ้นที่ความถี่ต่างๆ

แบบสุ่มปนเบื่อนอยู่กับสัญญาณที่ต้องการ[2] ได้ยินคล้ายเสียงดนตรีเมทัลลิกปนอยู่กับสัญญาณเสียงพูดที่ต้องการ ดังนั้นเทคนิคการลดสัญญาณรบกวนจึงต้องแก้ไขการเกิดปัญหา Musical Noise ด้วย เพื่อไม่ให้เกิดความเพี้ยนต่อสัญญาณเสียงพูด ทั้งนี้เทคนิควิธีการลดสัญญาณรบกวนได้รับการคิดค้นและพัฒนาอย่างต่อเนื่อง สำหรับแต่ละสภาวะแวดล้อม ก็มีเทคนิคการลดสัญญาณรบกวนที่เหมาะสมแตกต่างกันไป รายละเอียดของวิธี SS จะถูกกล่าวถึงอย่างละเอียดในบทที่ 2

เมื่อพิจารณาการใช้โทรศัพท์เคลื่อนที่ในรถยนต์ด้วยอุปกรณ์แฮนด์ฟรี (Hands-Free Telephony) พบว่าสัญญาณเสียงพูดจะถูกปนด้วยสัญญาณรบกวนที่เรียกว่า สัญญาณรบกวนพื้นหลังที่ปรากฏในรถยนต์ตัวอย่าง เช่น เสียงเครื่องยนต์, เสียงที่เกิดจากการกระทบระหว่างล้อกับถนน, เสียงจากลม, เสียงพูดคุย และ เสียงจากเครื่องเสียงสเตอริโอ เป็นต้น ทั้งนี้สัญญาณรบกวน ถูกแบ่งออกเป็น 2 ประเภท คือ สัญญาณรบกวนแบบคงที่ (Stationary noise) และสัญญาณรบกวนแบบไม่คงที่ (Non-stationary noise) สำหรับสัญญาณรบกวนแบบคงที่นั้นสามารถอธิบายในทางสถิติได้ คือ เป็นสัญญาณที่มีค่าเฉลี่ยและค่าความแปรปรวนคงที่และติดต่อกันยาวนาน[3] ตัวอย่างของสัญญาณรบกวนแบบนี้ได้แก่ เสียงเครื่องยนต์, เสียงที่เกิดจากการกระทบระหว่างล้อกับถนน และเสียงจากลม เป็นต้น สำหรับตัวอย่างของสัญญาณรบกวนแบบไม่คงที่ที่ปรากฏในรถยนต์ยกตัวอย่างคือ เสียงพูดคุย และ เสียงจากเครื่องเสียงสเตอริโอ เป็นต้น โดยสัญญาณแบบนี้จะมีคุณลักษณะทางสถิติไม่คงที่[3] ทั้งนี้เทคนิคการลดสัญญาณรบกวนก็จะมีวิธีและแนวคิดที่แตกต่างกันไปขึ้นกับชนิดของสัญญาณรบกวนที่พิจารณา

เทคนิคการลดสัญญาณรบกวนสามารถแบ่งได้เป็น 2 กลุ่มวิธีตามจำนวนไมโครโฟนได้แก่ เทคนิคที่ใช้ไมโครโฟนหนึ่งตัว (Single-Microphone Technique) และเทคนิคที่ใช้ไมโครโฟนหลายตัว (Multi-Microphone Technique) จากการศึกษางานวิจัยพบว่าเทคนิคการลดเสียงรบกวนที่ใช้ไมโครโฟนหลายตัวจะสามารถประมาณสัญญาณรบกวนได้ดีกว่า เนื่องจากข้อมูลเชิงพื้นที่ (Spatial information) ถูกนำมาใช้เพิ่มความแม่นยำในการประมาณสัญญาณรบกวน นอกจากนี้เทคนิคสามารถแบ่งย่อยตามการประมวลผลคือ เทคนิคที่ประมวลผลในโดเมนความถี่ และเทคนิคที่ประมวลผลในโดเมนเวลา โดยเทคนิคที่ประมวลผลในโดเมนเวลา โดยที่การลดเสียงรบกวนที่ใช้ในโดเมนความถี่นั้นมีข้อดีคือมีความซับซ้อนต่ำแต่ก็อาจเกิดความผิดเพี้ยนของสัญญาณเสียงได้ สำหรับการลดเสียงรบกวนที่ประมวลผลในโดเมนเวลา เรียกว่าการกำจัดสัญญาณรบกวนโดยใช้ตัวกรองปรับตัว (Adaptive Noise Cancelation, ANC) ไม่เกิดความผิดเพี้ยนของสัญญาณเสียงแต่มีความซับซ้อนสูงกว่า รายละเอียดจะถูกกล่าวถึงอย่างละเอียดในหัวข้อย่อยที่ 2.1

และ 2.2 การลดเสียงรบกวนที่ใช้ในโดเมนความถี่ และการลดเสียงรบกวนที่ใช้ในโดเมนเวลาตามลำดับ

วิทยานิพนธ์นี้มุ่งเน้นไปที่การพัฒนาเทคนิคการลดสัญญาณรบกวนที่ใช้ไมโครโฟน 2 ตัว สำหรับการใช้งานโทรศัพท์แบบไม่ใช้มือด้วยอุปกรณ์แฮนด์ฟรีในรถยนต์ เพื่อให้เทคนิคการลดเสียงรบกวนมีประสิทธิภาพการทำงานดีขึ้น กล่าวคือยังคงไว้ซึ่งประสิทธิภาพด้านการลดเสียงรบกวน และคำนึงถึงการรักษาคุณภาพของสัญญาณเสียงพูดขาออกที่มีความเพี้ยนต่ำ

1.2 วัตถุประสงค์ของการวิจัย

1. เพื่อพัฒนาเทคนิคการลดสัญญาณรบกวนที่ใช้ไมโครโฟนสองตัวสำหรับการใช้โทรศัพท์แบบไม่ใช้มือ
2. เพื่อลดความเพี้ยนของสัญญาณเสียงพูดที่ถูกสัญญาณรบกวนแล้วและรักษาคุณภาพของสัญญาณเสียงพูดดังกล่าวให้ใกล้เคียงกับต้นฉบับมากที่สุด

1.3 ขอบเขตของการวิจัย

1. จำลองเทคนิคการลดสัญญาณรบกวนต่างๆ ด้วยโปรแกรม MATLAB โดยสัญญาณขาเข้าของระบบเป็นสัญญาณเสียงพูด โดยใช้ฐานข้อมูลสัญญาณเสียงจาก TIMIT[23]
2. พิจารณาสัญญาณรบกวนเป็นประเภท สัญญาณรบกวนขาวแบบเกาส์, สัญญาณรบกวนในรถยนต์ และสัญญาณรบกวนเสียงพูดคุย
3. พิจารณาและศึกษาเทคนิคการลดสัญญาณรบกวนในทางโดเมนความถี่ด้วยวิธีการลบสเปกตรัม SS และวิธี MCSS รวมถึงเทคนิคที่ทำงานในทางโดเมนเวลา วิธีดับเบิลแอฟฟายโปรเจกชัน (Double Affine Projection, DAP)
4. พิจารณาความเป็นไปได้ของการใช้งานร่วมกันระหว่างทั้ง 2 เทคนิค และ เพื่อหาลำดับการทำงานที่เหมาะสม เพื่อให้ได้มาซึ่งข้อดีของทั้ง 2 เทคนิค
5. เปรียบเทียบประสิทธิภาพการลดสัญญาณรบกวนผ่านตัวชี้วัดแบบ objective 3 ตัว คือ Segmental Signal-to-Noise ratio (SegSNR), Log spectral Distance (LSD) และ สเปกโตรแกรม

6. ทำการทดสอบฟัง MOS (Subjective test) กับกลุ่มตัวอย่างผู้ฟังประมาณ 20 คน

1.4 ประโยชน์ที่คาดว่าจะได้รับ

เทคนิคการลดสัญญาณรบกวนที่มีประสิทธิภาพสำหรับการใช้งานการลดสัญญาณรบกวนของอุปกรณ์แฮนด์ฟรีในรถยนต์ โดยสามารถลดระดับของสัญญาณรบกวนได้ดี โดยการลดปริมาณสัญญาณรบกวนที่ปนอยู่ออกไป โดยก่อให้เกิดผลกระทบต่อคุณภาพสัญญาณพูดน้อยที่สุดคือก่อให้เกิดความเพี้ยนต่ำ

1.5 วิธีดำเนินการวิจัย

1. ศึกษางานวิจัยที่เกี่ยวข้องกับเทคนิคการลดสัญญาณรบกวน สำหรับการใช้งานในอุปกรณ์แฮนด์ฟรีในรถยนต์ และการใช้งานอื่นๆที่เกี่ยวข้อง
2. พิจารณาศึกษาหลักการทำงานของเทคนิคการลดสัญญาณรบกวนต่างๆ โดยเปรียบเทียบข้อดีข้อเสีย อันได้แก่ วิธีการลบสัญญาณรบกวนโดยใช้ไมโครโฟน 1 ตัว, วิธีการลดสัญญาณรบกวนโดยใช้ไมโครโฟน 2 ตัว และวิธีการเพิ่มฮาร์โมนิกของสัญญาณเสียงพูด การแก้ปัญหา Musical Noise ตัวตรวจจับกิจกรรมของสัญญาณเสียงพูด เป็นต้น
3. จำลองผลเทคนิคการลดสัญญาณรบกวนที่น่าสนใจเพื่อให้ทราบถึงประสิทธิภาพและปัญหาที่เกิดขึ้น
4. นำเสนอแนวคิดการลดสัญญาณรบกวนเพื่อแก้ปัญหาต่างๆที่เกิดขึ้นในวิธีการลดสัญญาณรบกวนแบบดั้งเดิม
5. จำลองผลเทคนิคการลดสัญญาณรบกวนที่เสนอและเปรียบเทียบผลการทดลองกับเทคนิคอื่นๆ ที่พิจารณาในวิทยานิพนธ์นี้ โดยพิจารณาประสิทธิภาพของเทคนิคการลดสัญญาณรบกวนต่างๆ ทั้งในด้านประสิทธิภาพการลดทอนสัญญาณรบกวน และในด้านคุณภาพของเสียงหรือวัดความเพี้ยนของสัญญาณเสียงพูด ผ่านตัวชี้วัดต่างๆ ทั้งแบบ Objective และแบบ Subjective
6. เขียนบทความทางวิชาการและนำเสนอผลงาน
7. สรุปผลการวิจัย และจัดทำวิทยานิพนธ์ฉบับสมบูรณ์

บทที่ 2

งานวิจัยที่เกี่ยวข้อง

เทคนิคการลดสัญญาณรบกวน ได้รับการพัฒนามาอย่างต่อเนื่อง จึงทำให้เกิดขั้นตอนวิธีต่าง ๆ ขึ้นเป็นจำนวนมาก จากการศึกษางานวิจัยที่เกี่ยวข้องพบว่า เทคนิคการลดสัญญาณรบกวนสามารถแบ่งย่อยตามการประมวลผลคือ การประมวลผลในโดเมนความถี่[4]-[9] และการประมวลผลในโดเมนเวลา รายละเอียดจะถูกกล่าวถึงอย่างละเอียดในหัวข้อย่อยที่ 2.1 และ 2.2 การลดสัญญาณรบกวนที่ใช้ในโดเมนความถี่ และการลดสัญญาณรบกวนที่ใช้ในโดเมนเวลา ตามลำดับ นอกจากนี้ ปัจจัยสำคัญที่จำเป็นในการลดสัญญาณรบกวนคือ ตัวตรวจจับสัญญาณเสียง (Voice Activity Detector, VAD) ซึ่งเป็นการแยกช่วงที่เป็นสัญญาณเสียงพูด (Speech-Activity Period) ออกจากช่วงที่ไม่ใช่สัญญาณเสียงพูด (Non-Speech Activity Period) คือมีแต่สัญญาณรบกวนเท่านั้น จะอธิบายไว้ในหัวข้อ 2.3

นอกจากนี้เมื่อทำการลดสัญญาณรบกวนแล้วจะพบว่าสัญญาณเสียงพูดที่ได้อาจมีความเพี้ยนไป เช่น การสูญเสียองค์ประกอบทางความถี่บางส่วน โดยเฉพาะฮาร์โมนิกของสัญญาณเสียงพูด ดังนั้นจึงต้องทำการสร้างสัญญาณที่หายไปกลับคืนมาโดยจะใช้วิธีการสร้างหรือก่อกำเนิดฮาร์โมนิกของสัญญาณเสียงพูดขึ้นใหม่ (Harmonic Regeneration)[10]-[12] ซึ่งจะอธิบายในหัวข้อ 2.4

2.1 เทคนิคการลดเสียงรบกวนที่ใช้ในโดเมนความถี่

สำหรับการลดสัญญาณรบกวนโดยอาศัยไมโครโฟน 1 ตัวที่ทำงานในโดเมนความถี่นั้น เทคนิคที่เป็นที่นิยมใช้กันมากคือเทคนิคการลบทางสเปกตรัม (Spectral subtraction technique, SS) [4]-[7] วิธีนี้มีหลักการคือจะทำการประมาณสเปกตรัมของสัญญาณรบกวน (Noise spectrum) แล้วนำมาลบออกจากสเปกตรัมของสัญญาณเสียงพูดที่ปนด้วยสัญญาณรบกวน (Noisy Spectrum) ซึ่งสเปกตรัมของสัญญาณรบกวนนี้จะถูกประมาณจากช่วงที่ไม่มีกิจกรรมของเสียงพูด ซึ่งวิธี SS สามารถอธิบายได้ดังนี้

2.1.1 เทคนิคการลบสเปกตรัม (Spectral Subtraction Technique, SS)

สำหรับการลดสัญญาณรบกวนเพื่อเพิ่มคุณภาพของสัญญาณเสียงพูดในระบบสนทนาหนึ่งๆ จะสามารถทำได้ โดยทำการประมาณสเปกตรัมของสัญญาณรบกวน แล้วลบออกจากสเปกตรัมของสัญญาณเสียงพูดที่มีสัญญาณรบกวนปนเปื้อนอยู่ ทั้งนี้วิธีการ SS นี้จะขึ้นกับความแม่นยำของการประมาณค่าสเปกตรัมของสัญญาณรบกวน โดยถ้าสเปกตรัมของสัญญาณรบกวนที่ประมาณขึ้นมานั้นมีค่าใกล้เคียงกับสเปกตรัมของสัญญาณรบกวนจริง จะทำให้ระบบสามารถลบสัญญาณรบกวนออกได้อย่างมีประสิทธิภาพ แต่ถ้าสเปกตรัมของสัญญาณรบกวนที่ประมาณขึ้นมามีความผิดพลาด เช่นมีค่ามากเกินไป โดยเฉพาะอย่างยิ่งบางองค์ประกอบความถี่ จะมีผลกระทบกับสัญญาณเสียงพูดต้นฉบับ (Original signal) ณ ความถี่นั้นๆ กล่าวคือจะทำให้องค์ประกอบสัญญาณเสียงพูดบางความถี่ขาดหายไป แต่ในทางกลับกันถ้าสเปกตรัมของสัญญาณรบกวนที่ประมาณขึ้นมานั้น มีค่าน้อยเกินไปจะทำให้ระบบไม่สามารถลบสัญญาณรบกวนออกไปได้อย่างมีประสิทธิภาพ ดังนั้นการประมาณสเปกตรัมของสัญญาณรบกวนให้เหมาะสมจึงมีความสำคัญมากเพื่อให้ได้สัญญาณเสียงพูดที่มีความเพี้ยนไปจากสัญญาณต้นฉบับน้อยที่สุด

เมื่อพิจารณาสัญญาณเสียงพูดที่ถูกรบกวน ดังแสดงในสมการ (2.1) ในโดเมนเวลา เมื่อ $d(n)$, $s(n)$ และ $b(n)$ คือ สัญญาณที่แปรตามเวลา n

$$d(n) = s(n) + b(n) \quad (2.1)$$

เมื่อ $s(n)$ คือ สัญญาณเสียงพูด

$b(n)$ คือ สัญญาณรบกวน

$d(n)$ คือ สัญญาณเสียงพูดที่ปนด้วยสัญญาณรบกวน

โดยมีสมมติฐานว่าสัญญาณเสียงพูด $s(n)$ และสัญญาณรบกวน $b(n)$ ไม่มีความสัมพันธ์กัน (Uncorrelated) เมื่อทำการแปลงฟูริเยร์แบบไม่ต่อเนื่องในช่วงเวลาสั้น (Short-Time Discrete Fourier Transform, STFT) ต่อสมการที่ (2.1) จะได้ความสัมพันธ์ในโดเมนความถี่ดังต่อไปนี้[2]

$$D(k) = S(k) + B(k) \quad (2.2)$$

เมื่อ $D(k)$ คือ สเปกตรัมของสัญญาณเสียงพูดที่ปนด้วยสัญญาณรบกวน

$S(k)$ คือ สเปกตรัมของสัญญาณเสียงพูด

$B(k)$ คือ สเปกตรัมของสัญญาณรบกวน

เมื่อ k คือ องค์กรประกอบทางความถี่ (Frequency-bin index) $k = 0, 1, \dots, N-1$
เมื่อทำการวิเคราะห์สเปกตรัม ด้วย STFT จำนวน N จุด

เนื่องจากสัญญาณเสียงพูดมีภาวะไม่นิ่ง (Non-stationary) หรือแปรตามเวลา (Time-Varying) ในการวิเคราะห์สัญญาณเสียงพูด จึงต้องพิจารณาเป็นเฟรมสั้นๆ ช่วงละ 15-30 มิลลิวินาที[13] ซึ่งการแบ่งสัญญาณเสียงพูดที่มีสัญญาณรบกวนปนเปื้อนอยู่ออกเป็นเฟรมจะทำได้ โดยกรองด้วยหน้าต่าง (Window) เช่น หน้าต่างแฮมมิง (Hamming window), หน้าต่างแฮนนิง (Hanning window) เป็นต้น ตามสมการที่ (2.3)

$$\begin{aligned} d_w(n) &= w(n) \cdot d(n) \\ &= w(n) \cdot [s(n) + b(n)] \\ &= s_w(n) + b_w(n) \end{aligned} \quad (2.3)$$

เมื่อ $w(n)$ คือ ฟังก์ชันหน้าต่าง

$d_w(n)$ คือ สัญญาณเสียงพูดที่ปนด้วยสัญญาณรบกวนที่ผ่านการกรองด้วยหน้าต่างแล้ว

$s_w(n)$ คือ สัญญาณเสียงพูดที่ผ่านการกรองด้วยหน้าต่างแล้ว

$b_w(n)$ คือ สัญญาณรบกวนที่ผ่านการกรองด้วยหน้าต่างแล้ว

ในทำนองเดียวกัน เมื่อทำการแปลงฟูริเยร์แบบไม่ต่อเนื่องในช่วงเวลาสั้นต่อสัญญาณในสมการที่ (2.3) จะได้ความสัมพันธ์ในโดเมนความถี่ดังนี้

$$\begin{aligned} D_w(k) &= W(k) * D(k) \\ &= S_w(k) + B_w(k) \end{aligned} \quad (2.4)$$

- เมื่อ $W(k)$ คือ สเปกตรัมของฟังก์ชันหน้าต่าง
- $D_w(k)$ คือ สเปกตรัมของสัญญาณเสียงพูดที่ปนด้วยสัญญาณรบกวนที่กรองด้วยหน้าต่างแล้ว
- $S_w(k)$ คือ สเปกตรัมของสัญญาณเสียงพูดที่ผ่านการกรองด้วยหน้าต่างแล้ว
- $B_w(k)$ คือ สเปกตรัมของสัญญาณรบกวนที่ผ่านการกรองด้วยหน้าต่างแล้ว

เมื่อตัวดำเนินการ $\langle * \rangle$ คือ การคอนโวลูชัน (Convolution) กำหนดให้สัญญาณที่กล่าวต่อจากนี้เป็นสัญญาณที่ผ่านการกรองด้วยหน้าต่างแล้วดังนั้นจะยกเว้นตัวห้อย w ในสมการ โดยจะแสดงเป็นดัชนีบ่งบอกเฟรม (Frame index) $l = 1, 2, \dots$ และองค์ประกอบทางความถี่ $k = 0, 1, \dots, N-1$ หรือเรียกว่าการแปลงฟูริเยร์แบบไม่ต่อเนื่องในช่วงเวลาสั้น

จากนั้นค่าประมาณสเปกตรัมของสัญญาณเสียงพูดจะหาได้จากการนำขนาดของค่าประมาณสเปกตรัมของสัญญาณรบกวนลบออกจากสเปกตรัมของสัญญาณพูดที่มีสัญญาณรบกวนปนเปื้อนอยู่ ซึ่งวิธี SS จะพิจารณาเฉพาะขนาดของสเปกตรัมโดยไม่นำเฟสมาพิจารณา เนื่องจากมนุษย์มีความไวต่อขนาดมากกว่าเฟส[3] ดังนั้นจากสมการที่ 2.2 กลายเป็น

$$D(k, l) = S(k, l) + B(k, l) \quad (2.5)$$

ทำให้

$$|\hat{S}(k, l)|^\sigma = |D(k, l)|^\sigma - \alpha \cdot \overline{|B(k, l)|^\sigma} \quad (2.6)$$

- เมื่อ $|S(k, l)|$ คือ ขนาดของค่าประมาณสเปกตรัมของสัญญาณเสียงพูด
- $|D(k, l)|$ คือ ขนาดสเปกตรัมของสัญญาณเสียงพูดที่ปนด้วยสัญญาณรบกวน
- $\overline{|B(k, l)|}$ คือ ขนาดของค่าประมาณสเปกตรัมของสัญญาณรบกวน

ณ องค์ประกอบทางความถี่ k และ เฟรมที่ l เมื่อ $\sigma = 1$ จะเป็นการพิจารณาขนาดของสเปกตรัมและ $\sigma = 2$ จะเป็นการพิจารณาสเปกตรัมกำลัง สมการ (2.6) เป็นการปรับปรุงจากสมการ (2.2) โดยเพิ่มตัวแปรการลบสเปกตรัม (Subtraction Parameter) α เพื่อควบคุมการลบสัญญาณรบกวน โดยเราสามารถพิจารณาได้ 2 เงื่อนไขคือ $\alpha=1$ การลบสัญญาณรบกวนแบบเต็ม (Full Noise Subtraction) และ $\alpha > 1$ การลบสัญญาณรบกวนเกิน (Over-Subtraction)[2]

โดยส่วนมากค่าประมาณสเปกตรัมของสัญญาณรบกวนจะหาได้จากช่วงที่ไม่มีกิจกรรมเสียงพูด (Non-Speech Activity period) ซึ่งโดยส่วนมากอาศัย VAD ในการระบุเฟรมหนึ่งๆว่าเป็นช่วงดังกล่าวหรือไม่

$$\overline{|B(k)|^\sigma} = \frac{1}{K} \cdot \sum_{i=0}^{K-1} |B_i(k)|^\sigma \quad (2.7)$$

เมื่อ $|B_i(k)|$ คือ ขนาดสเปกตรัมของสัญญาณรบกวน ณ เฟรมที่ i เมื่อ $i = 0, 1, \dots, K-1$ และ K คือ จำนวนเฟรมที่ไม่มีกิจกรรมของเสียงพูดทั้งหมด

สเปกตรัมของสัญญาณรบกวนนี้สามารถพิจารณาจากตัวกรองแถบความถี่ต่ำแบบดิจิทัลอันดับหนึ่ง (First Order Digital Low-pass Filter) เมื่อ ρ คือ สัมประสิทธิ์ตัวกรองแถบความถี่ต่ำ มีขอบเขตที่เหมาะสมอยู่ระหว่าง 0.85 ถึง 0.99[2]

$$\overline{|B(k)|^\sigma} = \rho \cdot \overline{|B_{i-1}(k)|^\sigma} + (1 - \rho) \cdot |B_i(k)|^\sigma \quad (2.8)$$

สำหรับการหาค่าประมาณสัญญาณเสียงพูดจากค่าประมาณสเปกตรัมที่ได้จากวิธี SS ทำได้โดยรวมเฟสของสเปกตรัมของสัญญาณเสียงพูดที่มีสัญญาณรบกวนกับขนาดของค่าประมาณสัญญาณเสียงพูด จากนั้นใช้วิธีการแปลงกลับของฟูรีเยร์แบบไม่ต่อเนื่องผกผันในช่วงเวลาสั้น และวิธีการ Overlap add[2] เพื่อแปลงให้อยู่ในโดเมนความเวลาต่อไป ดังนี้

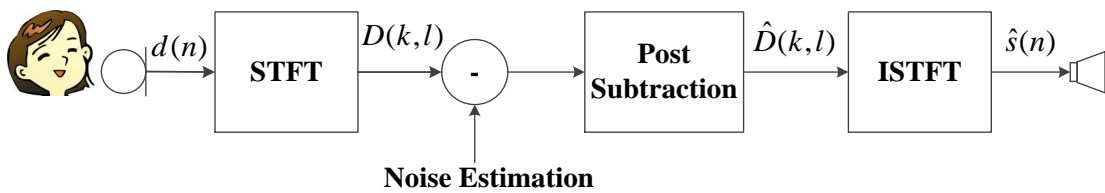
$$\hat{s}(n) = \sum_l \sum_{k=0}^{N-1} \hat{S}(k, l) \cdot \tilde{w}(n - l \cdot M) \cdot e^{j \cdot \frac{2\pi}{T} \cdot k(n - l \cdot M)} \quad (2.9)$$

เมื่อ $\hat{s}(n)$ คือ ค่าประมาณสัญญาณเสียงพูด

M คือ ค่าช่วงก้าวระหว่างเฟรม

$w(t)$ คือ หน้าต่างสังเคราะห์ (Synthesis window)

ณ องค์ประกอบทางความถี่ k และ เฟรมที่ l



รูปที่ 2.1 ขั้นตอนวิธีการลบสเปกตรัม[2]

รูปที่ 2.1 อธิบายภาพรวมของวิธีการลบสเปกตรัม โดยเริ่มจากสัญญาณขาเข้าในที่นี้คือสัญญาณเสียงพูดที่มีสัญญาณรบกวนปนเปื้อนอยู่ ผ่านวิธีการแปลงกลับของฟูริเยร์แบบไม่ต่อเนื่องผกผันในช่วงเวลาสั้นๆ จะได้สเปกตรัมสัญญาณเสียงพูดที่ปนเปื้อนด้วยสัญญาณรบกวน จากนั้นหาค่าประมาณสเปกตรัมของสัญญาณรบกวนและนำไปหักออกจากสเปกตรัมของสัญญาณเสียงพูดขาเข้า สำหรับค่าประมาณสเปกตรัมของสัญญาณเสียงพูดที่ได้จากสมการที่ (2.6) ในกรณีที่ถ้าสเปกตรัมของสัญญาณรบกวนที่ประมาณขึ้น ณ ความถี่ใดๆ มีค่ามากกว่าสเปกตรัมของสัญญาณเสียงที่ปนด้วยสัญญาณรบกวน เมื่อนำไปผ่านการลบทางสเปกตรัมตามสมการที่ (2.6) อาจมีผลลัพธ์เป็นค่าลบได้ ซึ่งจะไม่เกิดขึ้นกับขนาดหรือสเปกตรัมกำลัง เพื่อแก้ปัญหานี้จึงต้องใช้ฟังก์ชันต่อไปนี้ ดังแสดงในรูป 2.1 บล็อกหลังการลบ (Post Subtraction) เพื่อป้องกันการประมาณสเปกตรัมแล้วได้ขนาดเป็นลบ ดังสมการ (2.10)

$$T[|\hat{S}(k,l)|] = \begin{cases} |\hat{S}(k,l)| & |\hat{S}(k,l)| \geq \beta |D(k,l)| \\ fn[|D(k,l)|] & \text{Otherwise} \end{cases} \quad (2.10)$$

โดยส่วนมากมักจะเลือกตัวแปรพื้นของสเปกตรัม (Spectral floor parameter, β) ขอบเขตของ β ที่เหมาะสมคือ $0 < \beta \leq 1$ เพื่อลดระดับของสัญญาณรบกวนไว้[2]

จากนั้น หาค่าประมาณของสัญญาณเสียงพูดที่ถูกหักลบสเปกตรัมของสัญญาณรบกวนออกแล้วโดยใช้การแปลงกลับของฟูรีเยร์แบบไม่ต่อเนื่องผกผันในช่วงเวลาสั้น (Inverse Short-Time Discrete Fourier Transform) ดังสมการที่ (2.9)

จากการศึกษางานวิจัยในปี ค.ศ. 1979 ของ S. F. Boll[4] เป็นการลดสัญญาณรบกวนด้วยเทคนิคการลบทางสเปกตรัมทางขนาด (Magnitude Spectral Subtraction, MSS) นอกจากนี้ M. Berouti, R. Schwartz และ J. Makhoul[5], และในปี ค.ศ. 1980 R. J. McAulay และ M. L. Malpass[6] รวมทั้งในปี ค.ศ. 1984 Y. Ephraim และ D. Malah[7] เป็นการลดสัญญาณรบกวนด้วยเทคนิคการลบสเปกตรัมทางกำลัง (Power Spectral Subtraction, PSS) ข้อดีของวิธี PSS นี้คือ มีการทำงานที่ง่ายและราคาถูก แต่มีข้อจำกัดคือ ไม่เหมาะสมกับการลดสัญญาณรบกวนแบบไม่คงที่[3] เช่นเสียงพูดคุย อย่างไรก็ตามปัญหาที่เรียกว่า ‘Musical Noise’ เกิดขึ้นเนื่องจากการลบทางสเปกตรัม

2.1.1.1 การลบสเปกตรัมทางขนาด (Magnitude Spectral Subtraction-MSS)

การลบสเปกตรัมทางขนาด เมื่อ $\alpha = 1$ สามารถอธิบายได้ตามสมการนี้

$$|\hat{S}(k, l)| = D|Y(k, l)| - \overline{|B(k, l)|} \quad (2.11)$$

เมื่อพิจารณาค่าคาดหวัง ทั้งสองข้างของสมการจะได้

$$\begin{aligned} E[|\hat{S}(k, l)|] &= E[|D(k, l)|] - E[\overline{|B(k, l)|}] \\ &= E[|S(k, l) + B(k, l)|] - E[\overline{|B(k, l)|}] \\ &\approx E[|S(k, l)|] \end{aligned} \quad (2.12)$$

พบว่าค่าคาดหวังของค่าประมาณสเปกตรัมทางขนาดจะได้ค่าใกล้เคียงกับค่าคาดหวังสเปกตรัมสัญญาณเสียงพูด[2]

2.1.1.2 การลบสเปกตรัมทางกำลัง (Power Spectral Subtraction-PSS)

การลบสเปกตรัมทางกำลัง หรือ กำลังสองของการลบสเปกตรัมทางขนาด เมื่อ $\alpha = 2$ สามารถอธิบายตามสมการดังนี้

$$|\hat{S}(k, l)|^2 = |D(k, l)|^2 - \overline{|B(k, l)|^2} \quad (2.13)$$

จากนั้นหาค่าคาดหวังของสมการที่ (2.13) โดยมีสมมติฐานว่า สัญญาณเสียงพูด และสัญญาณเสียงรบกวนไม่สัมพันธ์กัน (uncorrelated) และมีคุณสมบัติเออร์годิกส์ (Ergodic) จะให้ความสัมพันธ์ดังต่อไปนี้

$$E\left[|\hat{S}(k, l)|^2\right] = E\left[|S(k, l)|^2\right] \quad (2.14)$$

ซึ่งหมายความว่าค่าคาดหวังของค่าประมาณสเปกตรัมกำลังของสัญญาณเสียงพูดมีค่าใกล้เคียงกับค่าคาดหวังของสเปกตรัมสัญญาณเสียงพูดยกกำลัง

2.1.1.3 การตรวจจับกิจกรรมของเสียงพูด (Voice Activity Detector, VAD)

การตรวจจับกิจกรรมของสัญญาณเสียงพูดเพื่อแยกช่วงของสัญญาณออกเป็น 2 ช่วง คือ ช่วงที่มีกิจกรรมของเสียงพูด และ ช่วงที่ไม่มีกิจกรรมของเสียงพูด สามารถทำได้มากมายหลายวิธีเช่น วิธีการตรวจจับจุดเริ่มเปลี่ยนแบบพิจารณาที่พลังงาน (Energy Threshold Method), อัตราการผ่านค่าศูนย์ (Zero Crossings Rate Method, ZCR) และวิธีประมาณคาบกำลังสองที่น้อยที่สุด (Least-Square Periodicity Estimator Method, LSPE)[14] เป็นต้น สำหรับวิธีการตรวจจับจุดเริ่มเปลี่ยนแบบพิจารณาที่พลังงานที่ถูกเลือกใช้ในวิทยานิพนธ์นี้ เป็นขั้นตอนวิธีการตรวจจับกิจกรรมของเสียงสัญญาณที่ทำได้ง่าย โดยจะทำการเปรียบเทียบพลังงานของสัญญาณเสียงพูด ณ เวลาหนึ่งๆ กับระดับจุดเริ่มเปลี่ยนที่กำหนดไว้ โดยมีสมมติฐานว่า ระดับพลังงานของสัญญาณเสียงพูดจะมีค่ามากกว่าพลังงานของสัญญาณรบกวนในกรณีของอัตราส่วนสัญญาณเสียงพูดต่อสัญญาณรบกวน (Signal-to-Noise Ratio, SNR) สูง ดังนั้นที่ระดับพลังงานที่ต่างกันนี้ เราจะสามารถตรวจจับกิจกรรมของเสียงพูดได้ การกำหนดระดับของจุดเริ่มเปลี่ยนจะขึ้นอยู่กับลักษณะสัญญาณเสียงพูดที่กำลังพิจารณา

$$\hat{s}_l = \begin{cases} 1, & E[y^2(n,l)] \geq \theta \\ 0, & E[y^2(n,l)] < \theta \end{cases} \quad (2.15)$$

เมื่อ \hat{s}_l คือ ค่าประมาณพลังงานของสัญญาณ ณ เฟรมที่ l
 $E[y^2(n,l)]$ คือ ค่าเฉลี่ยกำลังสองของสัญญาณขาเข้า ณ เฟรมที่ l
 θ คือ จุดเริ่มเปลี่ยน (Threshold)

2.1.1.4 ปัญหาที่เกิดขึ้นจากการใช้งานของวิธีการลบทางสเปกตรัม

ปัญหาหลักที่เกิดขึ้นจากการใช้งานวิธีการลบทางสเปกตรัม (SS) คือมี ‘Musical Noise’ ซึ่งเป็นสัญญาณรบกวนที่เกิดขึ้นที่ความถี่ต่างๆแบบสุ่มปนเปื้อนอยู่กับสัญญาณที่ต้องการ[2] ได้ ยืนคล้ายเสียงดนตรีเมทัลลิกปนอยู่กับสัญญาณเสียงพูดที่ต้องการ ดังนั้นเทคนิคการลดสัญญาณรบกวนจึงต้องแก้ไขการเกิดปัญหา Musical Noise ด้วย เพื่อไม่ให้เกิดความเพี้ยนต่อสัญญาณเสียงพูด

นอกจากนี้ เพื่อเป็นการเพิ่มประสิทธิภาพของเทคนิคการลบทางสเปกตรัม จากการศึกษางานวิจัยพบว่าเทคนิคการลดเสียงรบกวนที่ใช้ไมโครโฟนหลายตัวจะสามารถประมาณเสียงรบกวนได้ดีกว่าเทคนิคการลดสัญญาณรบกวนที่ใช้ไมโครโฟน 1 ตัว เนื่องจากข้อมูลเชิงพื้นที่ (Spatial information) ถูกนำมาใช้เพิ่มความแม่นยำในการประมาณเสียงรบกวน

- ปี ค.ศ. 1982 L. J. Griffiths และ C. Jim[15] เสนองานวิจัยการลดสัญญาณรบกวนโดยใช้ไมโครโฟนหลายตัวต่อกันแบบอาร์เรย์ วิธีนี้จะสามารถปรับค่าสัมประสิทธิ์แบบเวลาจริง (Real-Time) ได้ วิธีนี้เรียกว่าเทคนิคไมโครโฟนอาร์เรย์แบบบีม (Beam forming Microphone Array Technique)
- ปี ค.ศ. 1987 J. Meyer และ K. U. Simmer[16] ได้เสนอการลดสัญญาณรบกวนโดยใช้ไมโครโฟนมากกว่า 1 ตัว และใช้วิธีการกรองแบบวีเนอร์ (Wiener Filter) กับ การลบทางสเปกตรัมแบบเรียบ (Smooth Spectral Subtraction, SSS) ร่วมกัน โดยมีหลักการคือ แบ่งสัญญาณเสียงพูดที่ถูกรบกวนออกเป็นช่วงความถี่ต่างๆ ใช้วิธีการกรองแบบวีเนอร์ที่ช่วงความถี่สูงกว่าความถี่ที่กำหนดและใช้วิธี SSS กับช่วงความถี่ที่ต่ำกว่าที่กำหนด ซึ่งการใช้สองวิธีร่วมกัน

จะได้ผลดีกว่าการใช้งานวิธีใดวิธีหนึ่งเพียงอย่างเดียว เทคนิคนี้ทำให้สัญญาณรบกวนลดลงและมีความเพี้ยนของสัญญาณเสียงพูดที่ต่ำ อย่างไรก็ตามก็ข้อเสียคือการใช้ไมโครโฟนจำนวนมาก

ดังนั้นเทคนิคการลดสัญญาณรบกวนที่ใช้ไมโครโฟน 2 ตัว จึงถูกพิจารณาเนื่องจากมีค่าใช้จ่ายที่เพิ่มขึ้นไม่สูงมากนักเมื่อเปรียบเทียบกับเทคนิคการลดสัญญาณรบกวนที่ใช้ไมโครโฟนหลายตัว โดยมีประสิทธิภาพการลดสัญญาณรบกวนที่ดีขึ้นกว่าเทคนิคการลดสัญญาณรบกวนที่ใช้ไมโครโฟน 1 ตัว โดยวิทยานิพนธ์นี้ศึกษาวิธีการลบสเปกตรัมข้าม (Cross-Spectral Subtraction Method, CSS) และ วิธีการลบสเปกตรัมข้ามดัดแปร (Modified Cross-Spectral Subtraction Method, MCSS) ดังจะได้อธิบายต่อไปในหัวข้อย่อย 2.1.3 และ 2.1.4 ตามลำดับ

2.1.2 การจำลองสัญญาณสำหรับระบบที่มีไมโครโฟน 2 ตัว

สัญญาณเสียงพูดที่ปนด้วยเสียงรบกวนสำหรับระบบไมโครโฟน 2 ตัว สามารถพิจารณาได้จากสมการที่ (2.16) และ (2.17) ดังนี้

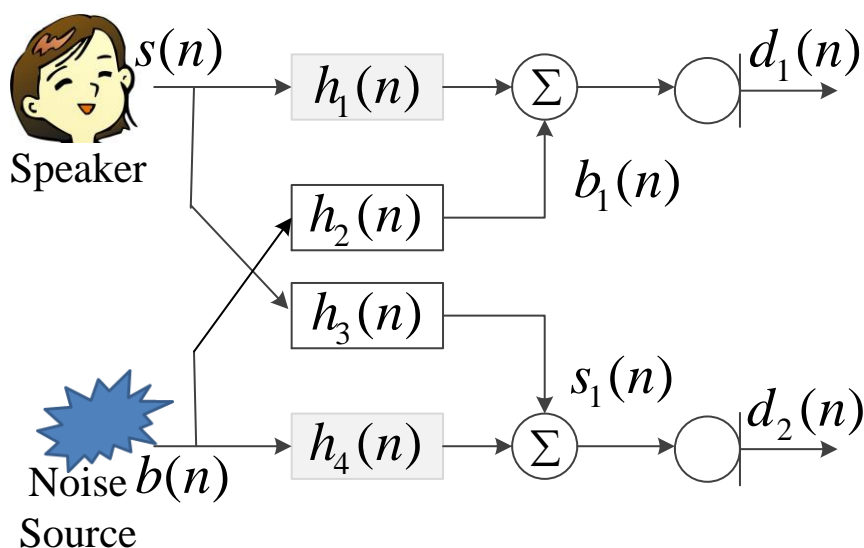
$$d_1(n) = s(n) + h_2(n) * b(n) = s(n) + b_1(n) \quad (2.16)$$

$$d_2(n) = b(n) + h_3(n) * s(n) = b(n) + s_1(n) \quad (2.17)$$

เมื่อ $d_1(n)$ คือ สัญญาณเสียงพูดที่ปนด้วยสัญญาณรบกวนที่เข้าไมโครโฟนตัวที่ 1

$d_2(n)$ คือ สัญญาณเสียงพูดที่ปนด้วยสัญญาณรบกวนที่เข้าไมโครโฟนตัวที่ 2

โดยที่ ผลตอบสนองอิมพัลส์ $h_2(n)$ คือ ผลกระทบที่เกิดจากทางเดินของสัญญาณเสียงระหว่างสัญญาณสัญญาณรบกวนกับไมโครโฟนตัวที่ 1 และ ผลตอบสนองอิมพัลส์ $h_3(n)$ คือผลกระทบที่เกิดจากทางเดินของสัญญาณเสียงระหว่างสัญญาณเสียงพูดกับไมโครโฟนตัวที่ 2 เมื่อ $b_1(n)$ เกิดจากการคอนโวลูชันระหว่าง $h_2(n)$ และ สัญญาณรบกวน $b(n)$ และ $s_1(n)$ เกิดจากการคอนโวลูชันระหว่าง $h_3(n)$ และสัญญาณเสียงพูด $s(n)$ แสดงดังรูปที่ 2.2



รูปที่ 2.2 การจำลองสัญญาณสำหรับระบบที่มีไมโครโฟน 2 ตัว

โดยสมมติให้ไมโครโฟนตัวที่ 1 ใกล้เคียงกับเสียงพูด และ ไมโครโฟนตัวที่ 2 ใกล้เคียงกับแหล่งกำเนิดสัญญาณรบกวน ดังนั้นจะไม่มีผลตอบสนองอิมพัลส์ $h_1(n)$ และ $h_4(n)$ และกำหนดให้เสียงพูด และเสียงรบกวนพื้นหลังไม่มีความสัมพันธ์กัน

2.1.3 เทคนิคการลบสเปกตรัมข้าม (Cross-Spectral Subtraction Method, CSS)

ปี ค.ศ. 1997 R. L. Bouquin-Jeannes, A. A. Azirani, และ G. Faucon[8] ได้เสนอการปรับปรุงสัญญาณเสียงพูดโดยมีหลักการคือ ใช้ไมโครโฟน 2 ตัว เพื่อรับสัญญาณมาวิเคราะห์ฟังก์ชันสหพันธ์ (Coherence function) ระหว่างสัญญาณรบกวนและสัญญาณเสียงพูด ฟังก์ชันสหพันธ์คือฟังก์ชันที่แสดงความสัมพันธ์ของความหนาแน่นกำลังระหว่างสัญญาณเสียงพูดและสัญญาณรบกวน ซึ่งการลดสัญญาณรบกวนในงานวิจัยนี้ ได้นำเสนอการใช้ 2 วิธีร่วมกันคือ ใช้การกรองแบบวีเนอร์ และ การลบทางสเปกตรัมแบบข้าม (Cross Spectral Subtraction) ข้อจำกัดของวิธีนี้คือไม่เหมาะสมที่จะใช้งานกับสัญญาณรบกวนที่มีคุณสมบัติไม่คงที่

2.1.4 เทคนิคการลบสเปกตรัมข้ามดัดแปร (Modified Cross-Spectral Subtraction Method, MCSS)

ปี ค.ศ. 2003 A Guerin, R. Le Bouquin-Jeannes และ G. Faucon[9] ได้เสนอวิธีลดสัญญาณรบกวนโดยใช้ไมโครโฟน 2 ตัวสำหรับอุปกรณ์แฮนด์ฟรีที่ใช้ในรถยนต์ โดยจะใช้ข้อมูล

เสียงที่รับมาจากไมโครโฟนทั้ง 2 ตัว เพื่อสร้างข้อมูลในเชิงปริภูมิ (Spatial Signal Characteristic) ข้อมูลนี้สามารถพิจารณาได้จากฟังก์ชันสหพันธ์ ระหว่างสัญญาณ d_1 และ d_2 ถูกนิยามดังสมการ

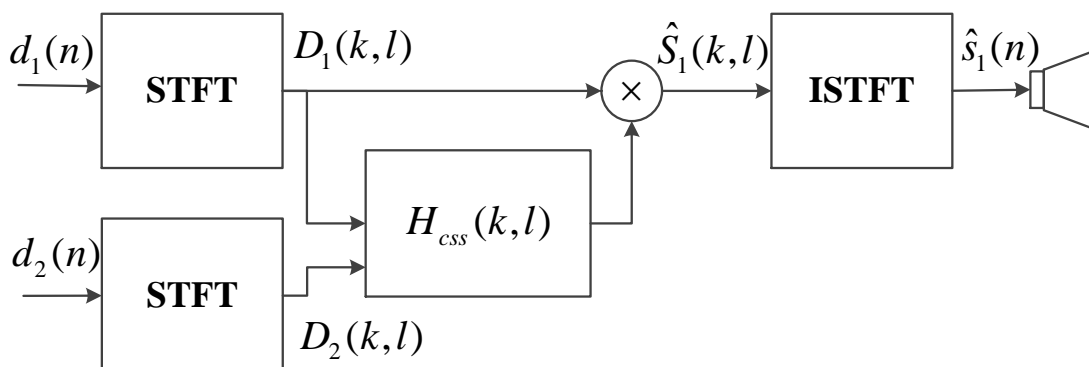
$$\rho(k) = \frac{\gamma_{d_1 d_2}(k)}{\sqrt{\gamma_{d_1}(k) \cdot \gamma_{d_2}(k)}} \quad (2.18)$$

$$MSC(k) = |\rho(k)|^2 \quad (2.19)$$

เมื่อ $\gamma_{d_i}(k)$ คือ ความหนาแน่นสเปกตรัมกำลัง (Power Spectral Density, PSD) ของสัญญาณเสียงพูดที่ปนด้วยสัญญาณรบกวน ไมโครโฟนตัวที่ i เมื่อ i คือ ไมโครโฟนตัวที่ 1 และ 2

$\gamma_{d_1 d_2}(k)$ คือ ความหนาแน่นสเปกตรัมกำลังแบบข้าม (Cross Power Spectral Density, Cross-PSD) ของสัญญาณเสียงพูดที่ปนด้วยสัญญาณรบกวน

เทคนิค MCSS นี้พิจารณาการลบทางสเปกตรัมในรูปการคูณด้วยตัวกรอง ที่เรียกว่า การลบสเปกตรัมแบบข้าม (Cross-spectral subtraction) หรือ $H_{css}(k, l)$ ซึ่งจะเป็นการกรองสัญญาณส่วนที่ไม่สัมพันธ์กันออกไปคือสัญญาณรบกวนนั่นเอง เทคนิคนี้จะแสดงได้ด้วยบล็อกไดอะแกรมดังรูปที่ 2.3



รูปที่ 2.3 วิธีลดสัญญาณรบกวนที่ใช้ไมโครโฟน 2 ตัว[9]

จากรูปที่ 2.3 ระบบจะรับสัญญาณเสียงพูดที่ปนเปื้อนด้วยเสียงรบกวนจากไมโครโฟนสองตัว $d_i(n)$ (เมื่อ i คือ ไมโครโฟนตัวที่ 1 และ 2) วิธี MCSS จะใช้ฟังก์ชันสหพันธ์ (Coherence function) ระหว่าง $d_1(n)$ และ $d_2(n)$ ทำหน้าที่เหมือนตัวกรองเพื่อลบสเปกตรัมของเสียงรบกวนออกไป ตัวกรองนี้จะอาศัยการลบกับระหว่างความหนาแน่นสเปกตรัมกำลังแบบข้ามของสัญญาณไมโครโฟน $\gamma_{d_1 d_2}(k, l)$ และความหนาแน่นสเปกตรัมกำลังแบบข้ามของเสียงรบกวนทั้งสอง $\gamma_{b_1 b_2}(k, l)$ ดังสมการต่อไปนี้

$$H_{css}(k, l) = \frac{|\gamma_{d_1 d_2}(k, l)| - |\gamma_{b_1 b_2}(k, l)|}{\sqrt{\gamma_{d_1}(k, l) \cdot \gamma_{d_2}(k, l)}} \quad (2.20)$$

เมื่อความหนาแน่นสเปกตรัมกำลังของสัญญาณไมโครโฟน $\gamma_{d_i}(k)$ และความหนาแน่นสเปกตรัมกำลังแบบข้ามของสัญญาณไมโครโฟน $\gamma_{d_1 d_2}(k)$ หาได้จากความสัมพันธ์ต่อไปนี้ [9]

$$\gamma_{d_i}(k, l) = \lambda(k, l)\gamma_{d_i}(k, l-1) + (1 - \lambda(k, l))D_i(k, l)D_i^*(k, l) \quad (2.21)$$

$$\gamma_{d_1 d_2}(k, l) = \lambda(k, l)\gamma_{d_1 d_2}(k, l-1) + (1 - \lambda(k, l))D_1(k, l)D_2^*(k, l) \quad (2.22)$$

เมื่อ $\langle \cdot \rangle^*$ แทนการคอนจูเกตเวกเตอร์และ $0 < \lambda(k, l) \leq 1$ เป็นสัมประสิทธิ์การลืม (Forgetting factor) โดยเลือกให้ $\lambda(k, l)$ มีค่าน้อยสำหรับเฟรมที่เป็นสัญญาณเสียงพูดหรือมีอัตราเสียงพูดต่อเสียงรบกวน (Signal-to-Noise Ratio, SNR) สูง และเลือกให้ $\lambda(k, l)$ มีค่าสูงสำหรับเฟรมที่ไม่มีกิจกรรมของสัญญาณเสียงพูดหรือมี SNR ต่ำ นอกจากนี้ เพื่อลดผลกระทบของสเปกตรัมของเสียงรบกวนที่เหลืออยู่ หรือที่เรียกว่า Musical Noise ซึ่งเป็นองค์ประกอบทางความถี่ของสัญญาณรบกวนที่เกิดขึ้นแบบสุ่ม [2] ได้ยีนคล้ายเสียงดนตรีเมทัลลิกปนอยู่กับเสียงพูด งานวิจัย [9] จึงเสนอ $\lambda(k, l)$ ดังนี้

$$\lambda(k, l) = 0.98 - 0.3 \cdot H_{css}(k, l-1) \quad (2.23)$$

ทั้งนี้ความหนาแน่นสเปกตรัมกำลังแบบข้ามของเสียงรบกวน $\gamma_{b_1 b_2}(k, l)$ ต้องมีการประมาณขึ้น โดย [9] เสนอให้หาจาก

$$\sqrt{\gamma_{b_1}(k,l)\gamma_{b_2}(k,l)} = \alpha(SNR_{post}(k,l)) \cdot \sqrt{\gamma_{b_1}(k,l-1) \cdot \gamma_{b_2}(k,l-1)} \quad (2.24)$$

เมื่อ γ_{b_i} คือ ความหนาแน่นสเปกตรัมกำลังของเสียงรบกวนในไมโครโฟนตัวที่ i โดยที่ $i = 1$ และ ฟังก์ชัน $\alpha(SNR_{post}(k,l))$ หาได้จาก

$$\alpha(SNR_{post}(k,l)) = L + \left(\frac{(1-L)}{1 + \frac{1}{g \cdot SNR_{post}(k,l)}} \right) \cdot \left(1 + \frac{1}{1 + g \cdot h \cdot SNR_{post}(k,l)} \right) \quad (2.25)$$

เมื่อตัวแปร h , g และ L เป็นค่าคงที่บวก โดยที่ $g = \frac{1}{1-h}$ และ $SNR_{post}(k,l)$ หาจาก

$$SNR_{post}(k,l) = \frac{|D_1(k,l) \cdot D_2(k,l)|}{\sqrt{\gamma_{b_1}(k,l-1) \cdot \gamma_{b_2}(k,l-1)}} \quad (2.26)$$

ซึ่งการประมาณ Cross-PSD ของเสียงรบกวนโดยเทคนิค MCSS นี้จะไม่อาศัยตัวตรวจจับกิจกรรมของเสียงพูด แต่จะพิจารณาจากลักษณะของ Cross-PSD ของเสียงรบกวนที่คงที่ต่อเนื่องกันตลอดทั้งช่วงที่มีกิจกรรมและไม่มีกิจกรรมของเสียงพูด จากนั้นค่าประมาณสเปกตรัมของเสียงพูดในไมโครโฟนตัวที่ i หาได้จาก

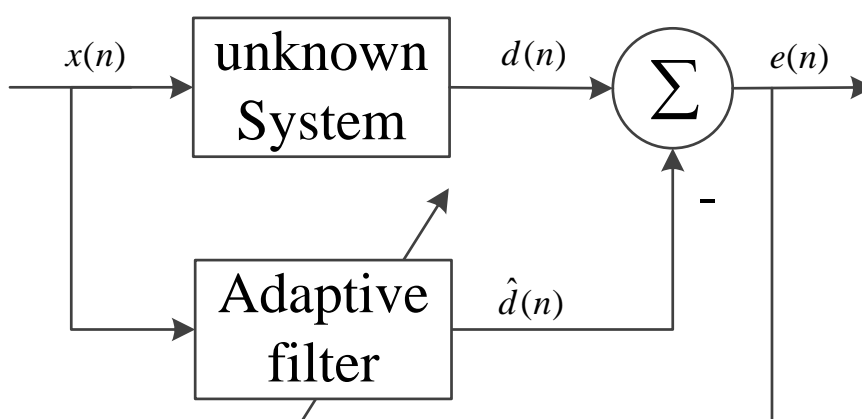
$$\hat{S}_i(k,l) = H_{css}(k,l) \cdot D_i(k,l) \quad (2.27)$$

เมื่อทำการแปลงผกผันฟูริเยร์แบบไม่ต่อเนื่องในช่วงเวลาสั้น ต่อค่าประมาณสเปกตรัมของเสียงพูด จะได้สัญญาณเสียงพูดที่ปรับปรุงแล้ว $\hat{s}_i(n)$ โดยสามารถประยุกต์ที่ช่องสัญญาณ เมื่อ $i = 1$ และ 2

งานวิจัยนี้ได้เสนอการประมาณสัญญาณรบกวนเกิน (Noise overestimation) โดยการคูณค่าคงที่ $\alpha = 8$ ใช้ในเทอมที่ 2 ในสมการที่ (2.19) เพื่อลดผลกระทบของ Musical Noise ซึ่งพบว่าสามารถลด Musical Noise ได้ดี แต่ทำให้สัญญาณเสียงพูดผิดเพี้ยนไปด้วย โดยผลการทดลองของเทคนิค MCSS นี้จะถูกนำเสนอต่อไปในบทที่ 4

2.2 เทคนิคการลดเสียงรบกวนที่ประมวลผลในโดเมนเวลา

การลดสัญญาณรบกวนยังสามารถประมวลผลในโดเมนเวลาได้ โดยใช้เทคนิคการกำจัดสัญญาณรบกวนแบบปรับตัว (Adaptive Filter)[17] ซึ่งมีหลักการทำงานคือ การนำตัวกรองที่สามารถปรับค่าสัมประสิทธิ์ของตัวเองได้ มาใช้งานในการประมาณระบบหรือช่องสัญญาณที่ไม่ทราบค่าเพื่อกำจัดสัญญาณรบกวน โดยขั้นตอนวิธีที่ใช้ในการควบคุมการปรับตัวได้แก่ ขั้นตอนวิธีกำลังสองเฉลี่ยน้อยที่สุด (Least Mean Square, LMS)[17], ขั้นตอนวิธีกำลังสองเฉลี่ยน้อยที่สุดแบบนอร์มาไลซ์ (Normalize Least Mean Square, NLMS)[17], ขั้นตอนวิธีกำลังสองน้อยที่สุดแบบวนซ้ำ (Recursive Least Square, RLS)[17] และขั้นตอนวิธีแอฟฟายโปรเจกชัน (Affine Projection, AP) [17] เป็นต้น ดังจะได้อธิบายในหัวข้อย่อย 2.2.1 – 2.2.4 ตามลำดับ



รูปที่ 2.4 แบบจำลองของระบบ

หลักการทำงานของตัวกรองประเภทนี้ จะพิจารณาผลทุกขณะเวลาคือ จะใช้ข้อมูลเสียงพูดขณะช่วงเวลานั้นๆ มาประมาณค่าสัญญาณที่ต้องการ และคำนวณหาค่าความคลาดเคลื่อนในการประมาณ $e(n)$

$$e(n) = d(n) - \hat{d}(n) \quad (2.28)$$

เมื่อ $d(n)$ คือ สัญญาณที่ต้องการของระบบ และสัญญาณที่ถูกประมาณเป็นสัญญาณขาออกของตัวกรองดังนี้

$$\hat{d}(n) = \mathbf{w}^T(n)\mathbf{x}(n) \quad (2.29)$$

เมื่อ $\mathbf{x}(n)$ คือ เวกเตอร์ของสัญญาณเข้า
 $\mathbf{w}(n)$ คือ เวกเตอร์ของสัมประสิทธิ์ตัวกรอง
 L คือ ความยาวของสัมประสิทธิ์ตัวกรอง

โดย $\mathbf{x}(n) = [x(n-1) x(n-2) \dots x(n-L+1)]^T$ และ $\mathbf{w}(n) = [w_1(n) w_2(n) \dots w_L(n)]^T$

ทั้งนี้การปรับตัวของตัวกรองจะแตกต่างกันไปสำหรับแต่ละขั้นตอนวิธี โดยมีรายละเอียดในหัวข้อย่อต่อไปนี้

2.2.1 ขั้นตอนวิธีกำลังสองเฉลี่ยน้อยที่สุด (Least Mean Square, LMS)

ขั้นตอนวิธี LMS[17] จะทำการประมาณสัมประสิทธิ์ของตัวกรองเพื่อให้ได้ค่าผิดพลาดกำลังสองเฉลี่ย (Mean-Squared error) ที่มีค่าต่ำสุด โดยจะใช้วิธีการวนซ้ำ สมการการปรับค่าพารามิเตอร์ของตัวกรองดังนี้

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu \cdot e(n) \cdot \mathbf{x}(n) \quad (2.30)$$

สมการ (2.27) จะใช้ข้อมูลขณะช่วงเวลานั้นๆของ $x(n)$, $d(n)$ และ $w(n)$ ในการปรับค่าพารามิเตอร์ของวงจรเพื่อประมาณค่าสัญญาณที่ต้องการ คำนวณหาค่าความคลาดเคลื่อน $e(n)$ โดยที่ μ (Step-Size Parameter) คือพารามิเตอร์ค่าช่วงก้าว เพื่อควบคุมขนาดของการปรับค่าพารามิเตอร์[17]

$$0 < \mu < \frac{2}{\sum_{m=0}^{L-1} x(n-m)^2} \quad (2.31)$$

เมื่อค่า μ มีค่ามากตัวกรองจะมีการปรับตัวอย่างรวดเร็วแต่ก็ทำให้การประมาณค่ามีความแม่นยำลดลง เมื่อ μ มีค่าน้อยจะมีความละเอียดของการประมาณค่าเพิ่มขึ้นแต่ก็ทำให้ระบบช้าลง ดังนั้นวิธี LMS มีความซับซ้อนต่ำ แต่อาจมีอัตราการลู่เข้า (Convergent Rate) ช้า

2.2.2 ขั้นตอนวิธีกำลังสองเฉลี่ยน้อยที่สุดแบบนอร์มาไลซ์ (Normalize Least Mean Square, NLMS)

ขั้นตอนวิธี NLMS[17] ปรับปรุงมาจากขั้นตอนวิธี LMS เพื่อป้องกันปัญหาสัญญาณขาเข้าขนาดใหญ่ โดยมีการนอร์มาไลซ์ด้วยขนาดของสัญญาณขาเข้า ดังสมการที่ (2.30)

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \frac{\mu \cdot e(n) \cdot \mathbf{x}(n)}{\varepsilon + \|\mathbf{x}(n)\|_2^2} \quad (2.31)$$

เมื่อ $\|\mathbf{x}(n)\|_2^2 = \mathbf{x}^T(n) \cdot \mathbf{x}(n) = x(n) + x^2(n-1) \cdot x^2(n-2) \cdot \dots \cdot x^2(n-L+1) = \sum_{m=0}^{L-1} x(n-m)^2$

และ ε คือค่าคงที่บวกขนาดเล็กๆ เพื่อป้องกันส่วนเป็นค่าศูนย์ โดย μ มีค่าอยู่ระหว่าง 0 และ 2

2.2.3 ขั้นตอนวิธีกำลังสองน้อยที่สุดแบบวนซ้ำ (Recursive Least Square, RLS)

ขั้นตอนวิธี RLS สร้างเวกเตอร์ $\mathbf{X}(n+1)$ จากสัญญาณขาเข้า ณ เวลา $n+1$

$$\mathbf{X}(n+1) = \begin{bmatrix} \mathbf{x}^T(n+1) \\ \dots \\ \mathbf{X}(n) \end{bmatrix} \quad (2.32)$$

เมื่อ $\mathbf{x}^T(n+1)$ คือ เวกเตอร์ของข้อมูลใหม่ และ $\mathbf{X}(n)$ คือ เมตริกซ์ของข้อมูล ณ เวลา n โดยกำหนดค่าเริ่มต้นของเวกเตอร์สัมประสิทธิ์ของตัวกรอง $\mathbf{w}(0)$ เป็นดังนี้

$$\mathbf{w}(0) = 0$$

ขั้นตอนวิธี RLS มีขั้นตอนการทำงานดังต่อไปนี้

$$\gamma = \mathbf{x}^T(n+1) \cdot \Phi^{-1}(n) \cdot \mathbf{x}(n+1) \quad (2.33)$$

$$\mathbf{k}(n+1) = \frac{\Phi^{-1}(n) \cdot \mathbf{x}(n+1)}{\gamma + 1} \quad (2.34)$$

$$\xi(n+1) = d(n+1) - \mathbf{w}^T(n) \cdot \mathbf{x}(n+1) \quad (2.35)$$

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mathbf{k}(n+1) \cdot \xi^*(n+1) \quad (2.36)$$

$$\Phi^{-1}(n+1) = \Phi^{-1}(n) - \mathbf{k}(n+1) \cdot \mathbf{x}^T(n+1) \cdot \Phi^{-1}(n) \quad (2.37)$$

เมื่อ $\mathbf{k}(n)$ คือ เวกเตอร์ของอัตราขยายกาลมาน (Kalman gain vector), γ คือ ตัวแปรไลคิลูด (Likelihood) และค่า $\delta^{-1}I$ ที่เพิ่มในสมการเพื่อป้องกันการหา $\Phi^{-1}(n)$ ให้มีค่าเสมอ โดยที่ $\Phi(n) = \delta^{-1}I$ ซึ่งวิธี RLS มีอัตราการลู่เข้าเร็วกว่าเมื่อเทียบกับขั้นตอนวิธี LMS และ NLMS ทั้งนี้กำหนดให้ $\Phi(0) = \delta^{-1}I$ [17]

2.2.4 ขั้นตอนวิธีแอฟฟายโปรเจกชัน (Affine Projection, AP)

ขั้นตอนวิธี AP[17] เป็นการปรับปรุงจากขั้นตอนวิธี NLMS เพื่อให้มีประสิทธิภาพใกล้เคียงขั้นตอนวิธี RLS โดยพิจารณาพารามิเตอร์ดังนี้ L คือ ความยาวของสัมประสิทธิ์ตัวกรอง, p คืออันดับโปรเจกชัน โดยที่ $1 \leq p \leq L$ และ $p=1$ คือขั้นตอนวิธี NLMS โดยเวกเตอร์ข้อมูลเข้ามีขนาด $(L \times p)$ ดังนี้

$$\begin{aligned} \mathbf{X}(n) &= [\mathbf{x}(n-1) \ \mathbf{x}(n-2) \ \dots \ \mathbf{x}(n-p+1)] \\ &= \begin{bmatrix} x(n) & x(n-1) & \dots & x(n-p+1) \\ x(n-1) & x(n-2) & \dots & x(n-p) \\ \vdots & \vdots & \ddots & \vdots \\ x(n-L+1) & x(n-L) & \dots & x(n-p-L+2) \end{bmatrix} \end{aligned} \quad (2.38)$$

และเมตริกซ์ข้อมูลเข้าขนาด $(L \times 1)$ คือ $\mathbf{x}(n) = [x(n-1) \ x(n-2) \ \dots \ x(n-L+1)]^T$ สำหรับการคำนวณสำหรับวิธี AP แสดงดังนี้ เมื่อ $n = 1, 2, \dots$

$$\mathbf{d}(n) = [d(n-1) \ d(n-2) \ \dots \ d(n-p+1)]^T \quad (2.39)$$

$$\mathbf{e}(n) = \mathbf{d}(n) - \mathbf{X}^T(n) \cdot \mathbf{w}(n) \quad (2.40)$$

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu \cdot \mathbf{X}^T(n) \cdot (\mathbf{X}(n) \cdot \mathbf{X}^T(n))^{-1} \cdot \mathbf{e}(n) \quad (2.41)$$

เมื่อ $\mathbf{d}(n)$ คือ เวกเตอร์สัญญาณออกจากระบบขนาด $(p \times 1)$

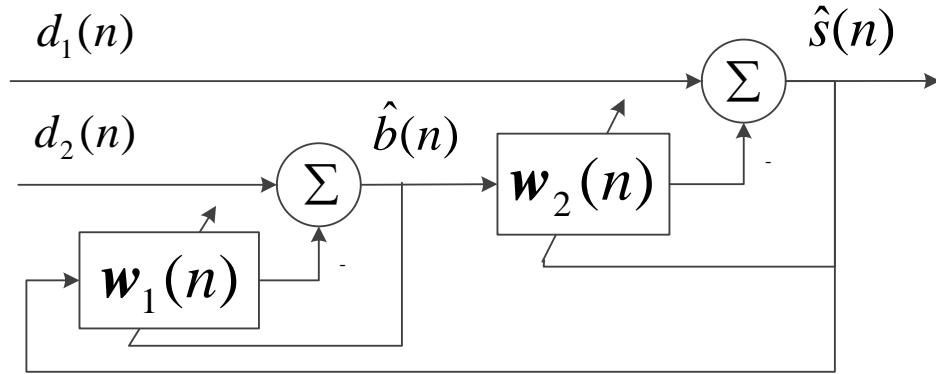
$\mathbf{e}(n)$ คือ เวกเตอร์ค่าความผิดพลาดขนาด $(p \times 1)$

$\mathbf{w}(n+1)$ คือ เวกเตอร์ของสัมประสิทธิ์ตัวกรองขนาด $(L \times 1)$

2.2.5 ขั้นตอนวิธีดับเบิลแอฟฟายโปรเจกชัน (Double Affine Projection, DAP)

ปี ค.ศ. 2003 M. Gabrea[18] ได้เสนอการกำจัดสัญญาณรบกวนแบบใช้ไมโครโฟน 2 สองตัว ที่เรียกว่า ดับเบิลแอฟฟายโปรเจกชัน หรือ DAP ซึ่งเป็นการนำตัวกรองปรับตัวที่ใช้กับวิธีแอฟฟายโปรเจกชันมาต่อแบบป้อนกลับเพื่อประมาณสัญญาณรบกวน วิธีนี้จะทำให้ตัวกรองมีประสิทธิภาพเพิ่มขึ้นเมื่อเทียบกับวิธีการลดสัญญาณรบกวนที่ใช้ตัวกรองแบบปรับตัววิธีอื่นๆ

ขั้นตอนวิธี DAP รับเสียงพูดที่ปนด้วยเสียงรบกวนจากไมโครโฟน 2 ตัว $d_1(n)$ และ $d_2(n)$ ไปยังตัวกรองปรับค่าได้ $w_1(n)$ และ $w_2(n)$ โดยที่ $w_1(n)$ มีจุดประสงค์เพื่อประมาณผลตอบสนองอิมพัลส์ (Impulse responses) $h_2(n)$ ซึ่งเป็นผลกระทบที่เกิดจากทางเดินของเสียงระหว่างเสียงรบกวนกับไมโครโฟนตัวที่ 1 โดยมีสัญญาณออกที่ได้คือ $\hat{s}_1(n)$ หรือค่าประมาณของสัญญาณ $s_1(n)$ จากนั้นนำสัญญาณ $\hat{s}_1(n)$ ลบกับสัญญาณจากไมโครโฟนตัวที่ 2 เพื่อให้ได้ค่าประมาณสัญญาณรบกวน $\hat{b}(n)$ และใช้เป็นสัญญาณเข้าของตัวกรองปรับค่าได้ $w_2(n)$ สำหรับ $w_2(n)$ นั้นมีจุดประสงค์เพื่อประมาณผลตอบสนองอิมพัลส์ $h_3(n)$ ซึ่งเป็นผลกระทบที่เกิดจากทางเดินของเสียงระหว่างเสียงพูดกับไมโครโฟนตัวที่ 2 โดยที่สัญญาณออกที่ได้คือ $\hat{b}_1(n)$ เพื่อนำไปลบออกจากสัญญาณจากไมโครโฟนตัวที่ 1 สัญญาณเสียงที่ได้คือค่าประมาณเสียงพูดที่ต้องการ $\hat{s}(n)$ สามารถแสดงดังรูปที่ 2.5 และอธิบายตามสมการที่ (2.42) และ (2.43) ดังนี้



รูปที่ 2.5 ขั้นตอนวิธีดับเบิลแอฟฟายโปรเจกชัน (DAP Algorithm)

$$\hat{\mathbf{b}}(n) = \mathbf{d}_2(n) - \mathbf{S}^T(n) \cdot \mathbf{w}_1(n-1) \quad (2.42)$$

$$\hat{\mathbf{s}}(n) = \mathbf{d}_1(n) - \hat{\mathbf{B}}^T(n) \cdot \mathbf{w}_2(n-1) \quad (2.43)$$

เมื่อ

$$\mathbf{d}_1(n) = [d_1(n), \dots, d_1(n-p+1)]^T$$

$$\mathbf{d}_2(n) = [d_2(n), \dots, d_2(n-p+1)]^T$$

$$\hat{\mathbf{s}}(n) = [\hat{s}(n), \dots, \hat{s}(n-p+1)]^T$$

$$\hat{\mathbf{b}}(n) = [\hat{b}(n), \dots, \hat{b}(n-p+1)]^T$$

$$\hat{\mathbf{S}} = [\hat{\mathbf{s}}(n), \dots, \hat{\mathbf{s}}(n-L+1)]^T$$

$$\hat{\mathbf{B}} = [\hat{\mathbf{b}}(n), \dots, \hat{\mathbf{b}}(n-L+1)]^T$$

$$\mathbf{w}_1(n) = [w_{1,0}(n), \dots, w_{1,L-1}(n)]^T$$

$$\mathbf{w}_2(n) = [w_{2,0}(n), \dots, w_{2,L-1}(n)]^T$$

เมื่อ L คือ ความยาวของตัวกรอง และ p คือ อันดับโปรเจกชัน โดยที่ $1 \leq p \leq L$ เมื่อค่า $p = 1$ ตัวกรองทำงานแบบขั้นตอนวิธี NLMS และเมื่อมีการปรับค่า p มากกว่า 1 เป็นการเพิ่ม

ประสิทธิภาพของตัวกรอง และ ตัวกรองปรับค่าได้ $w_1(n)$ และ $w_2(n)$ แสดงดังสมการที่ (2.44) และ (2.45) ตามลำดับ

$$\mathbf{w}_1(n) = \mathbf{w}_1(n-1) + \mu_1 \cdot \hat{S}(n) \cdot [\hat{S}^T(n) \cdot \hat{S}(n) + \delta_1 \cdot I]^{-1} \cdot \hat{\mathbf{b}}(n) \quad (2.44)$$

$$\mathbf{w}_2(n) = \mathbf{w}_2(n-1) + \mu_2 \cdot \hat{B}(n) \cdot [\hat{B}^T(n) \cdot \hat{B}(n) + \delta_2 \cdot I]^{-1} \cdot \hat{\mathbf{s}}(n) \quad (2.45)$$

เมื่อ $\delta_1 I$ และ $\delta_2 I$ คือเมตริกซ์เรกูลาไรเซชัน (Regularization matrix) เพื่อป้องกันเมตริกซ์ $\hat{S}^T(n) \cdot \hat{S}(n)$ และ $\hat{B}^T(n) \hat{B}(n)$ เป็นค่าลบ โดยที่ δ_1 และ δ_2 เป็นค่าบวกที่น้อยมากๆ และ μ_i มีค่าระหว่าง $0 < \mu_i < 2$ สำหรับ $i = 1$ และ 2

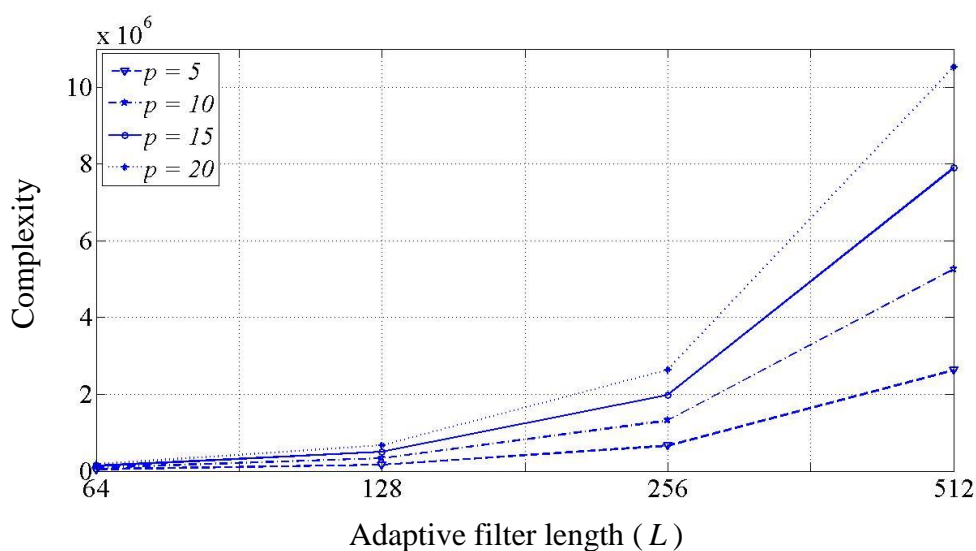
ตารางที่ 2.1 ความซับซ้อนในการคำนวณ (Computation complexity)

| Algorithms | Multiplication | Addition |
|--|---------------------------|------------------|
| $\hat{\mathbf{b}}(n) = \mathbf{d}_2(n) - S^T(n) \cdot \mathbf{w}_1(n-1)$ | Lp | Lp |
| $\hat{\mathbf{s}}(n) = \mathbf{d}_1(n) - \hat{B}^T(n) \cdot \mathbf{w}_2(n-1)$ | Lp | Lp |
| $\mathfrak{R}_1 = \hat{S}^T(n) \cdot \hat{S}(n) + \delta_1 \cdot I$ | $pL^2 + p$ | p^2 |
| $\mathfrak{R}_2 = \hat{B}^T(n) \cdot \hat{B}(n) + \delta_2 \cdot I$ | $pL^2 + p$ | p^2 |
| $\mathbf{w}_1(n) = \mathbf{w}_1(n-1) + \mu_1 \cdot \hat{S}(n) \cdot [\mathfrak{R}_1]^{-1} \cdot \hat{\mathbf{b}}(n)$ | $p^2 + p + Lp$ | $(p-1)(p+L)$ |
| $\mathbf{w}_2(n) = \mathbf{w}_2(n-1) + \mu_2 \cdot \hat{B}(n) \cdot [\mathfrak{R}_2]^{-1} \cdot \hat{\mathbf{s}}(n)$ | $p^2 + p + Lp$ | $(p-1)(p+L)$ |
| Total | $2pL^2 + 2p^2 + 4Lp + 3p$ | $4p^2 - 2p - 2L$ |

ความซับซ้อนในการคำนวณของขั้นตอนวิธี DAP สามารถแสดงดังตารางที่ 2.1 เมื่อการคูณ (Multiplication) และการบวก (Addition) ในการประมวลผลการทำงานของขั้นตอนวิธี DAP ถูกนำมาพิจารณา ซึ่งการคูณคือการคูณจำนวนจริงต่อตัวอย่างขาเข้า 1 ตัวอย่าง (Real Multiplication per Input Sample, RMPs) ขึ้นอยู่กับความยาวกำลังสองของสัมประสิทธิ์ตัวกรอง L^2 เป็นหลัก ดังแสดงในตารางที่ 2.2 และ รูปที่ 2.6

ตารางที่ 2.2 ความซับซ้อนในการคำนวณเนื่องจากการคูณ

| Computation Complexity ($2pL^2 + 2p^2 + 4Lp + 3p$) | | | | |
|--|--------|--------|---------|----------|
| $p \backslash L$ | 64 | 128 | 256 | 512 |
| 5 | 42305 | 166465 | 660545 | 2631745 |
| 10 | 84710 | 333030 | 1321190 | 5263590 |
| 15 | 127215 | 499695 | 1981935 | 7895535 |
| 20 | 169820 | 666460 | 2642780 | 10527580 |



รูปที่ 2.6 ความซับซ้อนในการคำนวณจากการคูณ เมื่อ

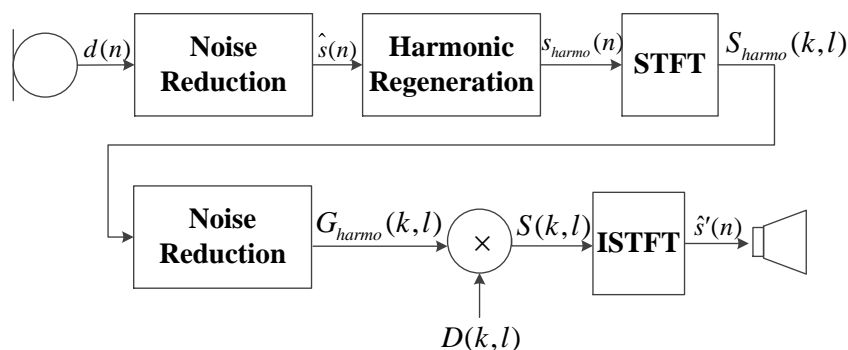
(---- $p = 5$, - - - $p = 10$, — $p = 15$, และ $p = 20$)

2.3 การสร้างทดแทนฮาร์โมนิกของเสียงพูด (Harmonic Regeneration, HR, Technique)

การเพิ่มฮาร์โมนิกของเสียงพูด ยกตัวอย่างงานวิจัย[10]-[12] โดยงานวิจัยเหล่านี้จะทำการสร้างฮาร์โมนิกของสัญญาณเสียงพูดขึ้นมาใหม่เพื่อปรับปรุง และ เพิ่มคุณภาพของเสียง ดังนี้

- ปี ค.ศ. 2005 C. Plapous, C. Marro และ P. Scalart[10] นำเสนอการปรับปรุงสัญญาณเสียงพูดที่ใช้ไมโครโฟน 1 ตัว เนื่องจากว่าหลังจากทำการการลบสัญญาณรบกวนแล้วฮาร์โมนิกของสัญญาณเสียงพูดจะหายไปบางส่วน ดังนั้นจึงเสนอวิธีการสร้างฮาร์โมนิกของสัญญาณเสียงพูดขึ้นมาใหม่ โดยใช้วิธีความไม่เป็นเชิงเส้น (Non-Linearity, NL) เรียกวิธีนี้ว่า Harmonic

Regeneration Noise Reduction (HRNR) มีจุดประสงค์เพื่อเพิ่มฮาร์โมนิก และปรับปรุงคุณภาพของสัญญาณเสียงพูด โดยมีหลักการทำงานดังนี้



รูปที่ 2.7 วิธีการสร้างฮาร์โมนิกตามงานวิจัย [10]

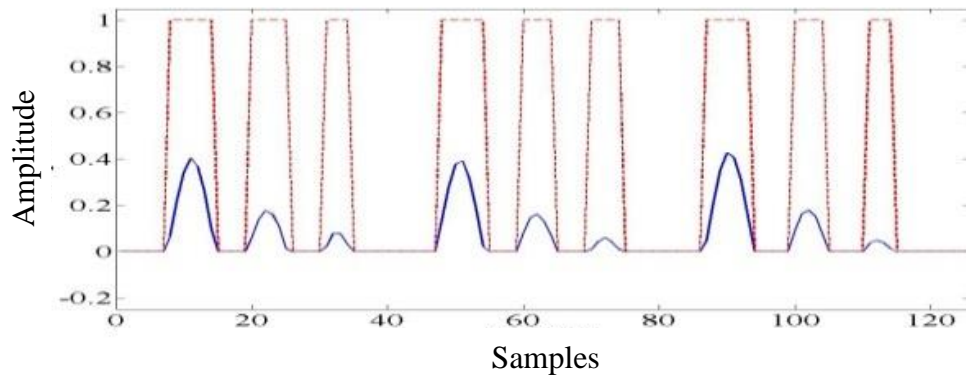
จากรูปที่ 2.7 หลังจากผ่านเทคนิคการลดสัญญาณรบกวนแล้วจะทำการสร้างฮาร์โมนิกของสัญญาณเสียงพูดขึ้นมาใหม่ตามงานวิจัย [10] เมื่อได้สัญญาณเสียงพูดที่ผ่านวิธีการสร้างฮาร์โมนิกแบบไม่เป็นเชิงเส้นแล้ว งานวิจัยนี้จะทำการคำนวณอัตราการกคสัญญาณรบกวนขึ้นมาใหม่ ($G_{harmo}(k, l)$) ซึ่งอัตราการกคสัญญาณรบกวนนี้จะสามารถรักษาสัญญาณฮาร์โมนิกของสัญญาณเสียงพูดไว้ได้

การสร้างทดแทนฮาร์โมนิกของเสียงพูด (HR) เพื่อเพิ่มคุณภาพของเสียงพูดในงานวิจัย[10] จะใช้ฟังก์ชัน ไม่เชิงเส้น ดังนี้

$$s_{harmo}(n) = \hat{s}(n) \cdot p(\hat{s}(n)) \quad (2.46)$$

$$p(\hat{s}(n)) = \begin{cases} 1 & \text{if } \hat{s}(n) \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (2.47)$$

เมื่อ $s_{harmo}(n)$ คือเสียงพูดที่ได้จากการสร้างทดแทนฮาร์โมนิก, $p(\hat{s}(n))$ แสดงตามสมการที่ (2.48) และ $\hat{s}(n)$ คือค่าประมาณเสียงพูด



รูปที่ 2.8 สัญญาณเสียงพูดที่ถูกสร้างทดแทนฮาร์โมนิกตามงานวิจัย [10] (1 เฟรม)

$$\hat{s}(n) \geq 0 \text{ (เส้นทึบ) และ } p(\hat{s}(n)) \text{ (เส้นประ)}$$

จากรูปที่ 2.8 เสียงพูดช่วง $\hat{s}(n) \geq 0$ แสดงในเส้นทึบ และ $p(\hat{s}(n))$ คือการสร้างรูปคลื่นแบบซำ (เส้นประ) เมื่อทำการแปลงฟูริเยร์ต่อ $s_{harmonic}(n)$ จะได้

$$F(s_{harmonic}(n)) = F(\hat{s}(n)) * \frac{e^{-j \cdot \theta}}{T} \cdot \sum_{m=-\infty}^{\infty} R\left(\frac{m}{T}\right) \cdot \delta\left(f - \frac{m}{T}\right) \quad (2.48)$$

เมื่อ T คือคาบของเสียง, $R(m/T)$ คือองค์ประกอบของรูปในโดเมนความถี่ ทำให้ $F(s_{harmonic}(n))$ ประกอบด้วยลำดับของอิมพัลส์ในโดเมนความถี่เป็นระยะห่างกัน m/T ทำให้สเปกตรัมของเสียงพูดมีฮาร์โมนิกลำดับต่างๆที่ถูกสร้างมีระยะห่างกัน m/T และขนาดของอิมพัลส์จะถูกกำหนดโดยขอบของ $F(s_{harmonic}(n))$

- ปี ค.ศ. 2010 T. Shen, D. P. Lun และ T. Hsung[11] ข้อดีของงานวิจัยนี้คือสามารถลดสัญญาณรบกวนที่มี SNR ค่าต่ำได้ โดยสามารถรักษาคุณภาพของสัญญาณเสียงพูดไว้ได้ หลักการของวิธีนี้ คือ จะใช้การประมาณค่า a priori SNR และการสร้างฮาร์โมนิกของสัญญาณเสียงพูดเช่นเดียวกันในงานวิจัยก่อนหน้านี้ [10] เพื่อลดสัญญาณรบกวนที่ยังเหลือ วิธีนี้สามารถลดสัญญาณรบกวนและรักษาคุณภาพของสัญญาณเสียงพูดได้
- ปี ค.ศ. 2011 H. Xuchu และ Z. Xiaojing[12] งานวิจัยนี้แก้ปัญหา ฮาร์โมนิกของสัญญาณเสียงพูดที่หายไปหลังจากการลดสัญญาณรบกวนแล้ว โดยใช้วิธี HRNR[10] ร่วมกับการแบ่งช่วงของสัญญาณเสียงพูดออกเป็น 3 ช่วงความถี่และทำการปรับค่าสัมประสิทธิ์ของสัญญาณฮาร์โมนิกไปเรื่อยๆ วิธีนี้สามารถสร้างฮาร์โมนิกของสัญญาณเสียงพูดขึ้นมาใหม่ได้ด้วย

บทที่ 3

เทคนิคการลดสัญญาณรบกวนที่นำเสนอในวิทยานิพนธ์

ในบทนี้การลดสัญญาณรบกวนที่พิจารณาสำหรับการใช้งานในระบบแฮนด์ฟรี ถูกนำเสนอ 3 วิธี ได้แก่ เทคนิคการสร้างทดแทนฮาร์โมนิกของสัญญาณเสียงพูดที่นำเสนอ, เทคนิคการแบ่งช่วงความถี่ของขั้นตอนวิธี DAP ร่วมกับวิธี MCSS และเทคนิคการแบ่งช่วงความถี่ย่อยของขั้นตอนวิธี DAP

แนวคิดที่นำเสนอคือการแบ่งช่วงความถี่ย่อย ที่ใช้วิธีที่ประมวลผลในโดเมนความถี่ และ วิธีที่ประมวลผลในโดเมนเวลารวมกัน โดยที่วิธีที่ประมวลผลในโดเมนความถี่มีข้อดีคือมีความซับซ้อนต่ำ แต่ข้อเสียคือเกิดความเพี้ยนของสัญญาณเสียง จากงานวิจัย [9] วิธี MCSS สามารถลดสัญญาณรบกวนได้ดีแต่มีผลกระทบกับองค์ประกอบของเสียงในช่วงความถี่สูง สำหรับวิธีที่ประมวลผลในโดเมนเวลานั้น มีข้อดีคือไม่ก่อให้เกิดความเพี้ยนของสัญญาณเสียงแต่ มีความซับซ้อนสูง[17] โดยวิธีที่เสนอจะคงไว้ซึ่งข้อดีของทั้งสองเทคนิค

แนวคิดแรก จะเรียกว่า เทคนิคร่วมระหว่างขั้นตอนวิธี DAP และ MCSS เกิดจากการพิจารณาลักษณะทางความถี่ของสัญญาณเสียงพูดที่ต่างกัน ในช่วงความถี่ต่ำมีองค์ประกอบทางความถี่ของข้อมูลสัญญาณเสียงพูดอยู่มากกว่าช่วงความถี่สูง[21] จึงเลือกใช้เทคนิคการลดสัญญาณรบกวนที่ทำงานในโดเมนความถี่ คือวิธี MCSS เพื่อลดระดับของสัญญาณรบกวน ในขณะที่ช่วงความถี่สูงนั้นองค์ประกอบของข้อมูลสัญญาณเสียงพูด ที่มีผลกระทบต่อคุณภาพ และความคมชัดของเสียงพูด[21] จึงเลือกใช้เทคนิคการลดสัญญาณรบกวนที่ทำงานในโดเมนเวลา คือวิธี DAP เพื่อรักษาองค์ประกอบทางความถี่สูงข้อมูลสัญญาณเสียงพูด[18]

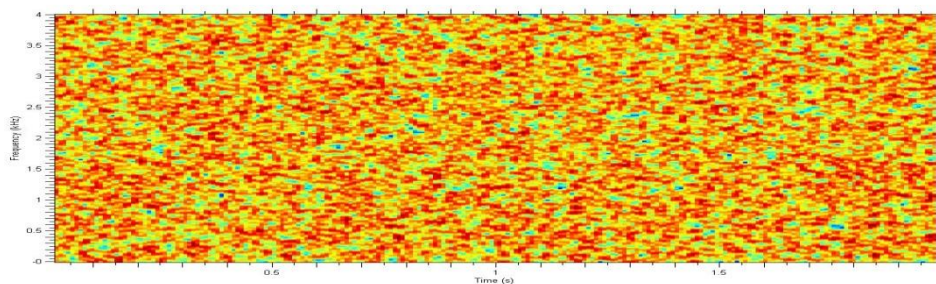
แนวคิดที่สองจะเรียกว่า ขั้นตอนวิธี DAP แบบหลายอันดับ สำหรับการแบ่งช่วงความถี่จะสามารถลดความซับซ้อนทั้งหมดของระบบลงได้ โดยเลือกใช้ขั้นตอนวิธี DAP เนื่องจากข้อดีของขั้นตอนวิธี DAP คือสามารถลดสัญญาณรบกวนโดยไม่ก่อให้เกิดความเพี้ยนของสัญญาณเสียง[18] จากงานวิจัย [18] อันดับโปรเจกชัน p ในขั้นตอนวิธี DAP นั้น มีผลกับประสิทธิภาพและความซับซ้อนของตัวกรอง ดังนั้นวิทยานิพนธ์นี้สนใจที่การปรับอันดับโปรเจกชันหรือค่า p ที่เหมาะสมกับช่วงความถี่เพื่อลดความซับซ้อนของตัวกรองและยังรักษาประสิทธิภาพการลดสัญญาณรบกวนของตัวกรองไว้ด้วย

นอกจากนี้ไม่ว่าจะใช้เทคนิคการลดสัญญาณรบกวนแบบใด หากเกิดปัญหาที่ข้อมูลความถี่สูง (Harmonic) ของสัญญาณเสียงพูดหายไปหลังจากทำการลดสัญญาณรบกวนแล้ว ให้สามารถสร้างทดแทนฮาร์โมนิก ดังนั้นเพื่อไม่ทำให้เกิดผลกระทบต่อคุณภาพของเสียงพูดที่ถูกลดเสียงรบกวน เทคนิคการสร้างทดแทนฮาร์โมนิกของเสียงพูดใน [10] จะถูกนำมาศึกษาเพื่อพัฒนาวิธีทางเลือกสำหรับการสร้างฮาร์โมนิกของเสียงพูด เพื่อให้เทคนิคการลดเสียงรบกวนมีประสิทธิภาพการทำงานดีขึ้น กล่าวคือยังคงไว้ซึ่งประสิทธิภาพด้านการลดเสียงรบกวน และก่อให้เกิดความเพี้ยนที่ต่ำต่อเสียงพูด

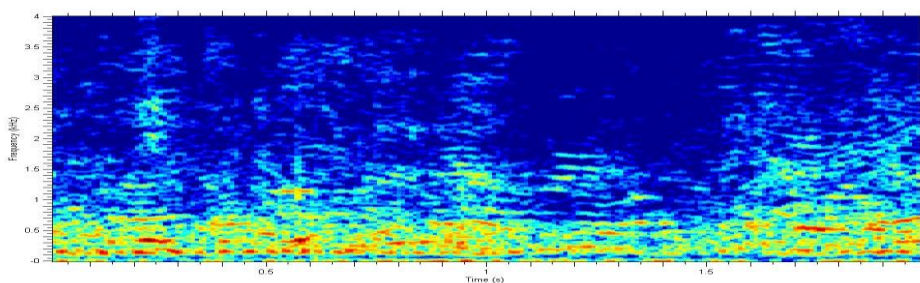
เทคนิคที่เสนอในงานวิจัยครั้งนี้ เทคนิคการสร้างทดแทนฮาร์โมนิกของสัญญาณเสียงพูดที่น่าเสนอ, เทคนิคการแบ่งช่วงความถี่ของขั้นตอนวิธี DAP และ เทคนิคการแบ่งช่วงความถี่ของขั้นตอนวิธี DAP ร่วมกับวิธี MCSS อธิบายในหัวข้อย่อยที่ 3.2, 3.3 และ 3.4 ตามลำดับ

3.1 สัญญาณรบกวนประเภทต่างๆ

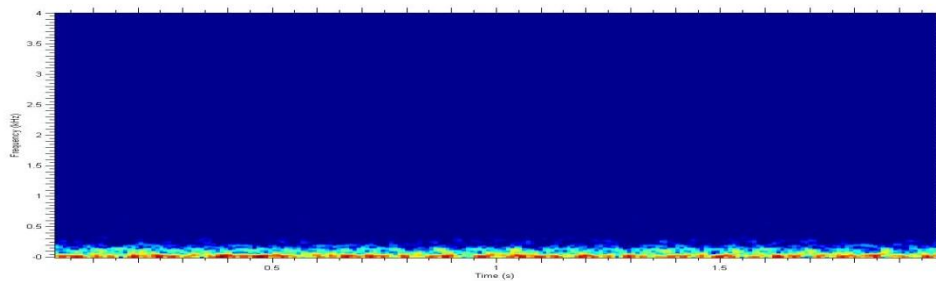
ในการสื่อสารในรถยนต์ด้วยระบบแฮนด์ฟรี พบสัญญาณรบกวนที่ปรากฏ เช่น เสียงเครื่องยนต์, เสียงที่เกิดจากการกระทบระหว่างล้อกับถนน, เสียงจากลม, เสียงพูดคุย และ เสียงจากเครื่องเสียงสเตอริโอ เป็นต้น ในวิทยานิพนธ์นี้พิจารณาสัญญาณรบกวน 3 แบบ เมื่อความถี่ในการสุ่มตัวอย่าง แสดงสเปกโตรแกรมดังนี้



ก)



ข)



ค)

รูปที่ 3.1 สเปกโตรแกรมของ

ก) สัญญาณรบกวนขาวแบบเกาส์ (White Gaussian Noise, WGN)

ข) สัญญาณรบกวนแบบเสียงพูดคึก (Babble Noise)

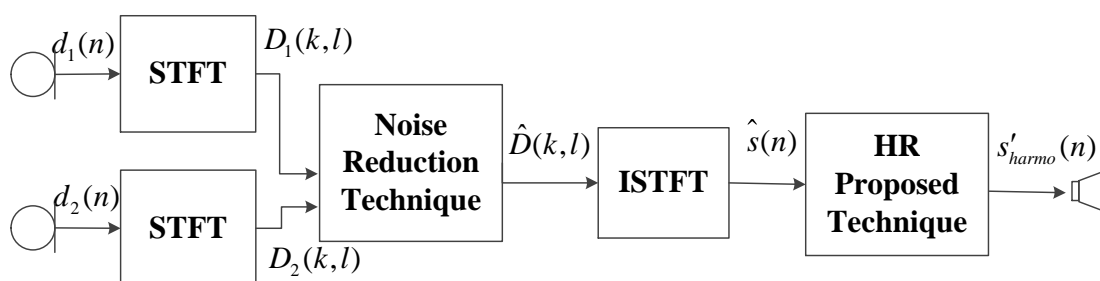
ค) สัญญาณรบกวนในรถยนต์ (Car Noise)

เมื่อพิจารณาสเปกโตรแกรมรูปที่ 3.1 ก) สเปกโตรแกรมของสัญญาณรบกวนขาวแบบเกาส์ เป็นสัญญาณรบกวนที่มีการกระจายทางความถี่แบบคงที่เท่ากับทุกความถี่ รูปที่ 3.1 ข) สเปกโตรแกรมของสัญญาณรบกวนแบบเสียงพูดคูลย เป็นสัญญาณรบกวนแบบไม่คงที่ หรือ มีคุณลักษณะทางสถิติไม่คงที่ คือมีการแปรตามเวลา มีการกระจายตัวอย่างหนาแน่นอยู่ในช่วงความถี่ต่ำกว่า 1.5 kHz และรูปที่ 3.1 ค) สเปกโตรแกรมของสัญญาณรบกวนแบบเสียงรถยนต์ เป็นสัญญาณรบกวนแบบคงที่ หรือ มีคุณลักษณะทางสถิติคงที่ คือไม่แปรตามเวลา มีการกระจายตัวอย่างหนาแน่นอยู่ในช่วงความถี่ต่ำกว่า 0.5 kHz ซึ่งพบว่าสัญญาณรบกวนแบบเสียงพูดคูลยและเสียงรถยนต์ มีผลกระทบต่อสัญญาณเสียงพูดในช่วงความถี่ต่ำ จึงเป็นที่มาของแนวคิดการแบ่งช่วงความถี่ย่อยสำหรับการลดสัญญาณรบกวนที่นำเสนอ โดยสัญญาณรบกวนขาวแบบเกาส์เป็นกรณีเส้นหลักล่าง (Base line) ที่ศึกษาไว้ด้วย

จากคุณลักษณะของสัญญาณรบกวนที่ต่างกันนี้ จึงมีแนวคิดการแบ่งช่วงความถี่เพื่อใช้งานกับประเภทของสัญญาณรบกวนที่ต่างกัน นอกจากนี้การแบ่งช่วงความถี่อาจไม่ได้ใช้ความกว้างของความถี่ย่อย (Subband frequency) ที่เหมือนกัน กล่าวคือ สัญญาณรบกวนแบบเสียงพูดคูลย เลือกความกว้างของความถี่ย่อย 0 - 1.5 kHz และสัญญาณรบกวนในรถยนต์เลือกความกว้างของความถี่ย่อย 0 - 1 kHz โดยเทคนิคการทำงานจะถูกอธิบายต่อไป

3.2 เทคนิคการสร้างทดแทนฮาร์โมนิกของสัญญาณเสียงที่นำเสนอ

เพื่อพัฒนาคุณภาพของสัญญาณเสียงพูดที่ถูกลดสัญญาณรบกวนแล้ว โดยเทคนิคนี้แตกต่างจากเทคนิคที่มีอยู่แล้วใน [10] ดังแสดงในบล็อกไดอะแกรมรูปที่ 2.7 วิทยานิพนธ์นี้เสนอวิธีการสร้างทดแทนฮาร์โมนิกของสัญญาณเสียงที่แตกต่างไปดังแสดงในรูป 3.2



รูปที่ 3.2 เทคนิคการสร้างทดแทนฮาร์โมนิกของสัญญาณเสียงที่นำเสนอ

การสร้างทดแทนฮาร์โมนิกของสัญญาณเสียงพูดมีแนวคิดในการสร้างคีนองค์ประกอบทางความถี่ที่เป็นฮาร์โมนิกของเสียงพูด และคงไว้ซึ่งองค์ประกอบเดิมของสัญญาณเสียงพูดด้วย ดังนี้

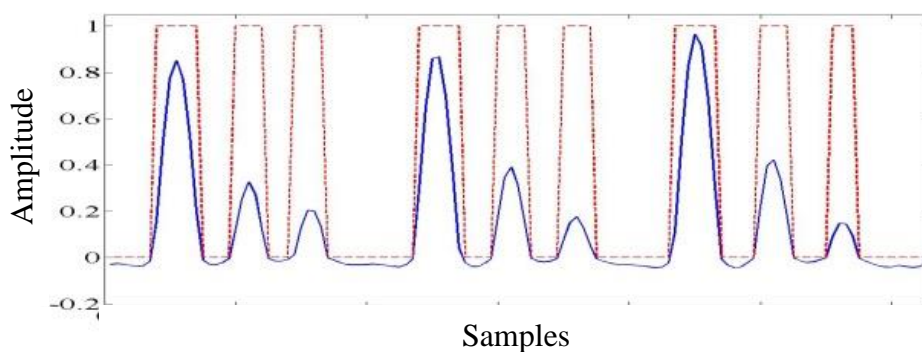
$$s'_{harmonic}(n) = \eta \cdot (\hat{s}(n) + \zeta \cdot \hat{s}(n) \cdot p(\hat{s}(n))) \quad (3.1)$$

เมื่อ $s'_{harmonic}(n)$ คือ เสียงพูดที่ผ่านการสร้างทดแทนฮาร์โมนิกที่นำเสนอ

ζ คือ ค่าคงที่ที่ใช้ในการปรับขนาดของสัญญาณเสียงพูด

η คือ ตัวแปรเพื่อปรับขนาดของ $s'_{harmonic}(n)$

ซึ่งความแตกต่างระหว่างวิธีที่นำเสนอกับวิธีเดิมใน [10] คือ เสียงพูดที่ผ่านการสร้างทดแทนฮาร์โมนิกที่นำเสนอ $s'_{harmonic}(n)$ ในสมการที่ (3.1) มีการคำนึงถึงข้อมูลเสียงเมื่อ $\hat{s}(n) < 0$ เพื่อรักษาค่าขององค์ประกอบของรูปคลื่นของเสียงพูด และ ζ คือค่าคงที่ที่ใช้ในการปรับขนาดของสัญญาณเสียงพูด โดยที่ $p(\hat{s}(n))$ เป็นดังสมการที่ (2.48) และ $\eta < 1$ เป็นการปรับขนาดของ $s'_{harmonic}(n)$ ที่ได้รับผลกระทบจากการคูณด้วย ζ เพื่อให้ใกล้เคียงกับสัญญาณต้นฉบับมากที่สุด ตัวอย่างสัญญาณเสียงพูดที่สร้างทดแทนฮาร์โมนิกด้วยวิธีที่นำเสนอเป็นดังรูปที่ 3.3

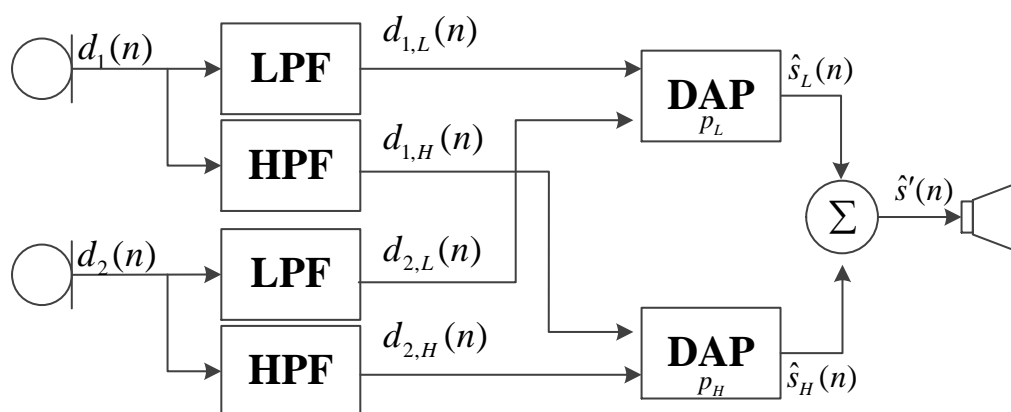


รูปที่ 3.3 สัญญาณเสียงพูดที่ถูกสร้างทดแทนฮาร์โมนิกด้วยวิธีที่นำเสนอ (1 เฟรม)

$s'_{harmonic}(n)$ (เส้นทึบ) และ $p(\hat{s}(t))$ (เส้นประ) เมื่อ $\zeta = 10$ และ $\eta = 0.2$

3.3 ขั้นตอนวิธี DAP แบบหลายอันดับ

ขั้นตอนวิธี DAP ที่ใช้ไมโครโฟน 2 ตัวที่ทำงานในโดเมนเวลา โดยใช้ตัวกรองปรับค่า 2 ตัว ($w_1(n)$ และ $w_2(n)$) ดังบล็อกไดอะแกรมรูปที่ 2.6 นั้น สามารถลดเสียงรบกวนพื้นหลังได้ดีโดยที่ไม่เกิดความเพี้ยนต่อสัญญาณเสียงพูด และสามารถรักษาคุณภาพของเสียงได้เป็นอย่างดี[18] แต่อย่างไรก็ตามวิธี DAP มีความซับซ้อนสูงเมื่อเทียบกับวิธีที่ทำงานในโดเมนความถี่ อันดับโปรเจกชันที่ปรากฏในวิธี DAP มีผลต่อความซับซ้อนของและอัตราการสุ่มเข้าของระบบซึ่งจะแปรผกผันกันกล่าวคือ เมื่อมีอันดับโปรเจกชันที่สูงระบบมีความซับซ้อนสูงแต่มีอัตราการสุ่มเข้าที่เร็วกว่า สำหรับอันดับโปรเจกชันที่ต่ำระบบมีความซับซ้อนต่ำแต่มีอัตราการสุ่มเข้าที่ช้ากว่า[18]



รูปที่ 3.4 ขั้นตอนวิธี DAP แบบหลายอันดับที่นำเสนอ

จากรูปที่ 3.4 LPF และ HPF แทนตัวกรองความถี่ต่ำ (Low-pass filter) และตัวกรองความถี่สูง (High-pass filter) ตามลำดับ ที่มีความถี่ตัด 0 - 1.5 kHz

เมื่อ $d_{i,L}(n)$ คือ เสียงพูดที่ปนด้วยเสียงรบกวนที่ความถี่ต่ำ

$d_{i,H}(n)$ คือ เสียงพูดที่ปนด้วยเสียงรบกวนที่ความถี่สูง

p_L คือ อันดับการโปรเจกชันในช่วงความถี่ต่ำ

p_H คือ อันดับการโปรเจกชันในช่วงความถี่สูง

$\hat{s}_L(n)$ คือ ค่าประมาณเสียงพูดในช่วงความถี่ต่ำที่ใช้วิธี DAP

$\hat{s}_H(n)$ คือ ค่าประมาณเสียงพูดในช่วงความถี่สูงที่ใช้วิธี DAP

$\hat{s}'(n)$ คือ ค่าประมาณเสียงพูดของการแบ่งช่วงความถี่ด้วยวิธี DAP ที่เสนอ

เมื่อ i คือ ไมโครโฟนตัวที่ 1 และ 2

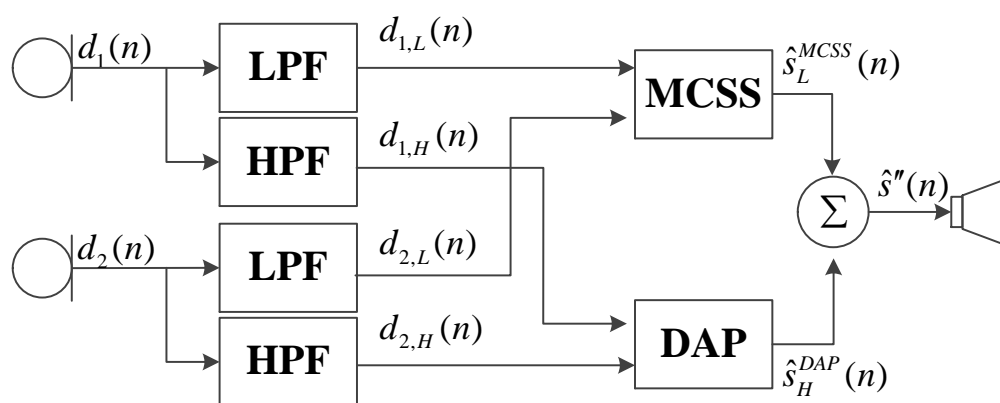
การทำงานร่วมกันระหว่างวิธี DAP ที่มีอันดับโปรเจกชันที่สูงและต่ำโดยใช้การแบ่งแถบความถี่ย่อยที่เสนอในวิทยานิพนธ์นี้ มุ่งไปที่การลดสัญญาณรบกวนแบบใช้ไมโครโฟน 2 ตัวที่ใช้งานกับอุปกรณ์แฮนด์ฟรีในรถยนต์ การแบ่งช่วงความถี่แบ่งออกเป็น 2 ช่วงความถี่ ขึ้นอยู่กับลักษณะทางความถี่ของสัญญาณรบกวนกล่าวคือ ในกรณีสัญญาณรบกวนแบบเสียงพูดคุย จะใช้ความถี่ตัด (Cutoff frequency) เป็น 1.5 kHz สำหรับสัญญาณรบกวนรถยนต์ จะใช้ความถี่ตัดเป็น 1 kHz

เนื่องจากในช่วงความถี่ต่ำมีปริมาณข้อมูลของเสียงพูดอย่างความหนาแน่น[21] ในขณะที่ปริมาณข้อมูลของเสียงพูดจะน้อยลงในช่วงความถี่ที่สูงขึ้น ดังนั้นเพื่อรักษาองค์ประกอบทางความถี่ของเสียงพูด วิทยานิพนธ์นี้จึงเสนอให้เลือกอันดับโปรเจกชันสูงเพื่อทำงานในช่วงความถี่ต่ำ ในทางกลับกันเสนอให้เลือกขั้นตอน DAP ที่มีอันดับโปรเจกชันที่ต่ำกับช่วงความถี่สูง ทั้งนี้ยังสามารถลดความซับซ้อนในการคำนวณโดยรวมลง เมื่อเปรียบเทียบกับการทำงานของขั้นตอนวิธี DAP โดยลำพังที่เลือกอันดับการโปรเจกชันสูง

3.4 เทคนิคร่วมระหว่างขั้นตอนวิธี DAP และ MCSS

เทคนิคการลดเสียงรบกวนด้วยไมโครโฟน 2 ตัวด้วยวิธี MCSS ที่ทำงานในโดเมนความถี่ สามารถประมาณเสียงพูดและยังสามารถลดเสียงรบกวนได้อย่างไรก็ตามปัญหาที่เกิดขึ้นกับวิธีนี้คือ ฮาร์โมนิกของเสียงพูดที่ความถี่สูงก็ถูกลบออกไปด้วย[9] ซึ่งทำให้คุณภาพของเสียงลดลงในทางกลับกัน วิธี DAP ที่ใช้ไมโครโฟน 2 ตัวที่ทำงานในโดเมนเวลา ดังบล็อกไดอะแกรมรูปที่ 2.6 นั้น สามารถลดเสียงรบกวนพื้นหลังได้ดีโดยที่ไม่เกิดความเพี้ยนของเสียงพูด แต่อย่างไรก็ตามวิธี DAP มีความซับซ้อนสูงเมื่อเทียบกับวิธีที่ทำงานในโดเมนความถี่ ดังนั้นเพื่อแก้ปัญหาฮาร์โมนิกของเสียงพูดที่ความถี่สูงที่หายไปของวิธี MCSS จึงใช้ขั้นตอนวิธี DAP เพื่อมาเสริมองค์ประกอบของเสียงพูดที่ถูกลบออกไปที่ความถี่สูง

การทำงานร่วมกันระหว่างวิธีที่ทำงานในโดเมนเวลา (วิธี DAP) และ วิธีที่ทำงานในโดเมนความถี่ (วิธี MCSS) ที่เสนอในวิทยานิพนธ์นี้ มุ่งไปที่การลดสัญญาณรบกวน โดยสามารถรักษาองค์ประกอบความถี่ของเสียงพูด และยังคงมีความซับซ้อนต่ำ โดยทำการแบ่งเป็น 2 ช่วงความถี่คือ ความถี่ที่สูงกว่า 1.5 kHz และความถี่ที่ต่ำกว่า 1.5 kHz ในกรณีสัญญาณรบกวนแบบเสียงพูดคุย และ ความถี่ที่สูงกว่า 1 kHz และความถี่ที่ต่ำกว่า 1 kHz ในกรณีสัญญาณรบกวนแบบเสียงพูดคุย



รูปที่ 3.5 เทคนิคร่วมระหว่างขั้นตอนวิธี DAP และ MCSS ที่นำเสนอ

เมื่อ $\hat{S}_L^{MCSS}(n)$ คือ ค่าประมาณเสียงพูดที่ความถี่ต่ำที่ใช้วิธี MCSS

$\hat{S}_H^{DAP}(n)$ คือ ค่าประมาณเสียงพูดที่ความถี่สูงที่ใช้วิธี DAP

$\hat{s}''(n)$ คือ ค่าประมาณเสียงพูดของเทคนิคที่ใช้วิธี MCSS และ DAP ที่เสนอ

ผลรวมของค่าประมาณเสียงพูดที่ความถี่ต่ำ ($\hat{s}_L^{MCSS}(n)$) และความถี่สูง ($\hat{s}_H^{DAP}(n)$) ที่ได้คือ ค่าประมาณเสียงของพูด $\hat{s}''(n)$

สำหรับขั้นตอนวิธี DAP ที่ทำงานในความถี่สูง เลือกใช้อัตราการโปรเจกชันที่ต่ำ เนื่องจากในช่วงความถี่สูงมีปริมาณความหนาแน่นของข้อมูลน้อยกว่า อัตราการรู้ที่ช้ากว่าสามารถเป็นที่ยอมรับได้และโปรเจกชันที่ต่ำยังมีข้อดีคือ ระบบมีความซับซ้อนต่ำลงโดยสามารถดูได้จากตารางที่ 2.1

สำหรับวิธี MCSS ที่ทำงานในความถี่ต่ำ เพื่อเป็นการลดความซับซ้อนของระบบสามารถเลือกใช้ค่า α หรือ ตัวแปรควบคุมการลบสเปกตรัมที่น้อยกว่าค่า α ที่เสนอในงานวิจัย [9] ได้

$$H_{css}(k,l) = \frac{|\gamma_{d_1 d_2}(k,l) - \alpha_L \cdot |\gamma_{b_1 b_2}(k,l)|}{\sqrt{\gamma_{d_1}(k,l) \cdot \gamma_{d_2}(k,l)}} \quad 3.2$$

เมื่อ α_L คือ ตัวแปรควบคุมการลบสเปกตรัมที่ทำงานในความถี่ต่ำ

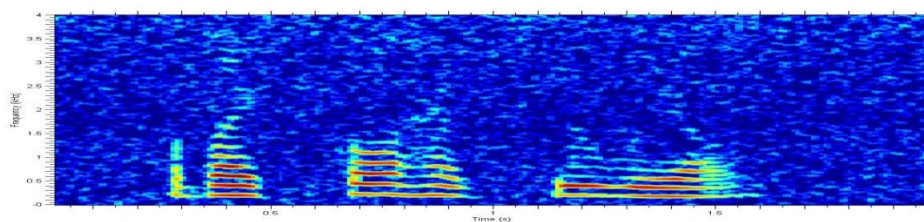
บทที่ 4

การจำลองระบบบนคอมพิวเตอร์และการวิเคราะห์

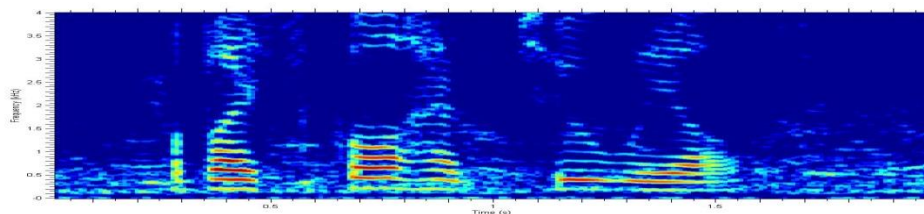
ในบทนี้จะนำเสนอผลการวิเคราะห์ข้อมูลของเทคนิคการลดสัญญาณรบกวนต่างๆ โดยแบ่งการทดลองเป็น 4 การทดลอง ได้แก่ การศึกษาประสิทธิภาพของวิธี MCSS, วิธี HR[10] และ วิธี HR ที่นำเสนอ, การศึกษาประสิทธิภาพของอันดับการโปรเจกชันของวิธี DAP, การศึกษาประสิทธิภาพของเทคนิคการแบ่งช่วงความถี่ย่อยของขั้นตอนวิธี DAP ที่นำเสนอ และ ประสิทธิภาพของเทคนิคการแบ่งช่วงความถี่ของขั้นตอนวิธี DAP ร่วมกับวิธี MCSS ที่นำเสนอ พร้อมนำเสนอตัวชี้วัด

4.1 การจำลองระบบ

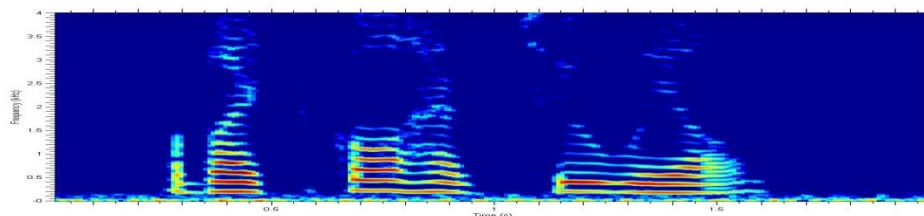
ในการจำลองระบบด้วยโดยใช้โปรแกรม MATLAB ใช้สัญญาณเสียงพูดจากฐานข้อมูลของ TIMIT เป็นสัญญาณเสียงพูดของระบบที่พิจารณา ‘A bicycle has two wheels.’ ทั้งเพศหญิงและชาย ความถี่ในการซัดตัวอย่าง 8 kHz พิจารณาเสียงรบกวน 3 แบบคือ สัญญาณรบกวนขาวแบบเกาส์ (White Gaussian Noise, WGN), สัญญาณรบกวนแบบเสียงพูดคุย (Babble Noise) และเสียงรบกวนในรถยนต์ (Car Noise) สเปกโตรแกรมของสัญญาณเสียงพูดและสัญญาณเสียงพูดที่ปนด้วยสัญญาณรบกวนทั้ง 3 แบบถูกแสดงดังรูปที่ 4.1 และ 4.2 ทั้งเพศหญิงและชาย ตามลำดับ



ก)



ข)

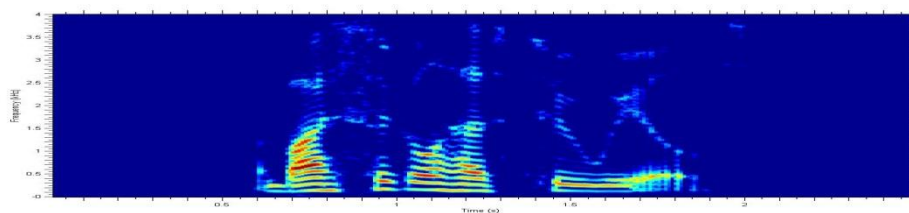


ค)

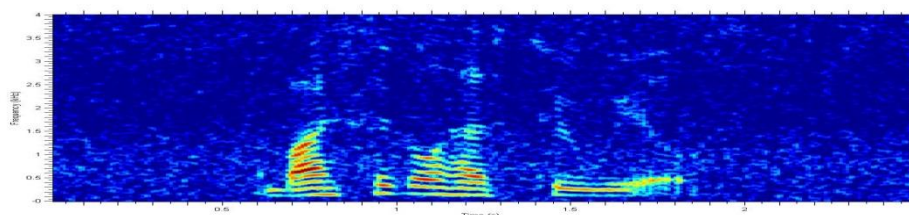
รูปที่ 4.1 สเปกโตรแกรมของ

- ก) สัญญาณเสียงพูดเพศหญิงพูดว่า 'A bicycle has two wheels.'
- ข) สัญญาณเสียงพูดที่ปนด้วยสัญญาณรบกวนขาวแบบเกาส์ (White Gaussian Noise, WGN),
- ค) สัญญาณเสียงพูดที่ปนด้วยสัญญาณรบกวนแบบเสียงพูดคึก (Babble Noise) และ
- ง) สัญญาณเสียงพูดที่ปนด้วยเสียงรบกวนในรถยนต์ (Car Noise)

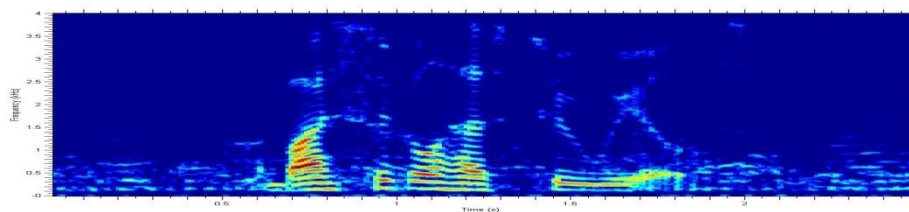
SNR ขนาด 10 dB



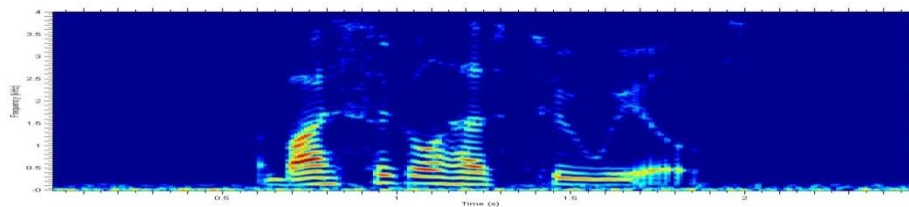
ก)



ข)



ค)



ง)

รูปที่ 4.2 สเปกโตรแกรมของ

ก) สัญญาณเสียงพูดเพศชายพูดว่า 'A bicycle has two wheels.'

ข) สัญญาณเสียงพูดที่ปนด้วยสัญญาณรบกวนขาวแบบเกาส์ (White Gaussian Noise, WGN)

ค) สัญญาณเสียงพูดที่ปนด้วยสัญญาณรบกวนแบบเสียงพูดคูย (Babble Noise) และ

ง) สัญญาณเสียงพูดที่ปนด้วยเสียงรบกวนในรถยนต์ (Car Noise)

SNR ขนาด 10 dB

จากรูปที่ 4.1 ก) แสดงสเปกโตรแกรมของสัญญาณเสียงพูดเพศหญิงพูดว่า ‘A bicycle has two wheels.’ และรูปที่ 4.1 ข) แสดงสเปกโตรแกรมของสัญญาณแบบเกาส์พบว่ามีการกระจายตลอดทั้งความถี่ รูปที่ 4.1 ค) สัญญาณรบกวนแบบเสียงพูดคุย พบว่า มีการกระจายตัวที่ความถี่ต่ำกว่า 1.5 kHz และเป็นสัญญาณรบกวนแบบไม่คงที่ และรูปที่ 4.1 ง) เสียงรบกวนในรถยนต์ พบว่ามีปริมาณเสียงรบกวนอยู่อย่างหนาแน่นที่ความถี่ความถี่ต่ำกว่า 1 kHz และเป็นสัญญาณรบกวนแบบคงที่ ออกแบบการทดลองโดยใช้ไมโครโฟน 2 ตัวเพื่อรับสัญญาณขาเข้าตามรูปที่ 2.4 โดยมี SNR ขาเข้าเป็น 5 - 20 dB

4.2 ตัวชี้วัดประสิทธิภาพ

ตัวชี้วัดแบบ Objective ที่พิจารณาคือ ค่าระยะสเปกตรัมลอการิทึม (Log-Spectral Distance, LSD) ใช้เพื่อประเมินความเพี้ยน (Distortion) ของเสียงพูด และค่าอัตราส่วนระหว่างสัญญาณต่อสัญญาณรบกวนความผิดเพี้ยนแยกส่วน (Segmental SNR, SEGSNR) ใช้เพื่อประเมินการลดสัญญาณรบกวน โดยอธิบายในหัวข้อ 4.2.1 และ 4.2.2 ตามลำดับ (ITU. T P. 835[19]) นอกจากนี้ยังมีตัวชี้วัดแบบ Subjective ตาม ITU P. 800[20] คือการทดสอบคุณภาพของเสียงโดยใช้การฟังแล้วให้คะแนนที่เรียกว่า Mean Opinion Score (MOS) โดยประเมินผู้ฟังจำนวน 20 คน โดยอธิบายในหัวข้อ 4.2.3 และ สเปกโตรแกรม ในหัวข้อ 4.2.4

4.2.1 ค่าอัตราส่วนระหว่างสัญญาณต่อสัญญาณรบกวนความผิดเพี้ยนแยกส่วน (Segmental SNR, SegSNR)[3]

SegSNR ใช้เป็นตัวชี้วัดประสิทธิภาพในเรื่องความสามารถในการลดทอนสัญญาณรบกวนของเทคนิคที่พิจารณา กล่าวคือถ้าระบบสามารถลดสัญญาณรบกวนลงได้มาก ค่า SegSNR จะสูงในขณะที่ถ้าระบบยังคงมีสัญญาณรบกวนคงเหลืออยู่มากค่า SegSNR จะต่ำ โดย SegSNR สามารถหาได้ดังนี้[3]

$$SegSNR(dB) = \frac{1}{L_s} \sum_l F_s \left\{ 10 \times \log_{10} \left(\frac{\sum_{t=0}^{T-1} s^2 \left(n + \frac{lT}{2} \right)}{\sum_{t=0}^{T-1} \left(s \left(n + \frac{lT}{2} \right) - \hat{s} \left(n + \frac{lT}{2} \right) \right)^2} \right) \right\} \quad (4.1)$$

เมื่อ $\hat{s}(n)$ คือ ค่าของสัญญาณเสียงพูด

$F_s \{\bullet\}$ คือ ฟังก์ชันที่เลือกเฉพาะเฟรมที่มีเสียงพูด

L_s คือ จำนวนเฟรมที่มีเสียงพูด

ในทางปฏิบัติจะใช้ $\Delta SegSNR$ ในการแสดงผล[3] หาได้จาก

$$\Delta SegSNR = SegSNR_s - SegSNR_d \quad (4.2)$$

เมื่อ $SegSNR_s$ หาได้จากสมการที่ (2.50) โดยแทน $\hat{s}(n)$ ด้วยสัญญาณเสียงพูดที่ปนด้วยสัญญาณรบกวนออก $d(n)$ ถ้าค่า $\Delta SegSNR$ มากแสดงว่าลดสัญญาณรบกวนได้ดี และถ้าค่า $\Delta SegSNR$ น้อยแสดงว่ายังมีสัญญาณรบกวนเหลืออยู่มากคือลดสัญญาณรบกวนได้ไม่ดี

4.2.2 ระยะสเปกตรัมลอการิทึม (Log-Spectral Distance, LSD)

LSD คือตัวชี้วัดเพื่อประเมินความผิดพลาดของสัญญาณเสียงพูด ซึ่งจะพิจารณาในโดเมนความถี่ สามารถหาได้จาก

$$LSD(dB) = \frac{1}{L_s} \sum_l F_s \left\{ \frac{1}{\frac{K}{2} + 1} \sum_{k=0}^{K/2} \left[20 \times \log_{10} \left(\frac{|S(k,l)|}{|\hat{S}(k,l)|} \right) \right]^2 \right\} \quad (4.3)$$

เมื่อ K คือ จำนวนข้อมูลใน 1 เฟรม, l คือดัชนีบ่งบอกเฟรม และ k คือองค์ประกอบทางความถี่

4.2.3 คะแนนความคิดเห็นเฉลี่ย (Mean Opinion Score, MOS)

นอกจากตัวชี้วัดที่กล่าวไปข้างต้นซึ่งเป็นการวัดแบบ Objective การทดสอบฟัง (Subjective Test) ก็เป็นอีกวิธีหนึ่งที่น่ามาบ่งบอกถึงคุณภาพของสัญญาณเสียงโดยโครงร่างวิทยานิพนธ์นี้เลือกใช้ค่าคะแนนความคิดเห็นเฉลี่ย (MOS) ซึ่งเป็นการวัดคุณภาพของสัญญาณเสียงโดยใช้การฟังแล้วให้คะแนน ตามมาตรฐานของ ITU-T P.800[19] ได้กำหนดหลักการให้คะแนนไว้คือ

ตารางที่ 4.1 ระดับคะแนนความคิดเห็นเฉลี่ย (Mean Opinion Score, MOS)

| คะแนน | คุณภาพของสัญญาณเสียง | ระดับความบิตเพี้ยนของสัญญาณเสียง |
|-------|----------------------|------------------------------------|
| 5 | Excellent | Imperceptible |
| 4 | Good | Just Perceptible, but not annoying |
| 3 | Fair | Perceptible and slightly annoying |
| 2 | Poor | Annoying but not objectionable |
| 1 | Bad | Very annoying and objectionable |

วิธีการวัดคุณภาพสัญญาณเสียงแบบ MOS นั้นหลังจากที่ผู้ฟังฟังสัญญาณเสียงที่กำหนดให้แล้ว (โดยไม่เรียงตามคุณภาพ) จะให้ผู้ฟังให้คะแนน ซึ่งจะมีอยู่ 5 ระดับ เสียงที่แย่ที่สุดจะให้คะแนนเป็น 1 และดีที่สุดเป็น 5 แล้วนำคะแนนจากผู้ทดสอบฟังทั้งหมดมาหาค่าเฉลี่ย (Mean) เพื่อเป็นคะแนนสำหรับสัญญาณเสียงแต่ละเสียง โดยส่วนมากในการทดลองจะให้ผู้ฟัง ได้ทำการทดสอบฟังกับสัญญาณเสียงพูดก่อนฉบับและสัญญาณเสียงพูดที่ปนเปื้อนที่มีค่า SNR ที่ต่าง ๆ กัน ด้วย รวมถึงสัญญาณเสียงพูดที่ถูกสัญญาณรบกวนแล้วด้วยวิธีต่างๆ

4.2.4 สเปกโตรแกรม (Spectrogram)

สเปกโตรแกรม แสดงองค์ประกอบทางความถี่ของสัญญาณเสียงพูดก่อนและหลังเทคนิคการลดสัญญาณรบกวน และผลกระทบต่อองค์ประกอบฮาร์มอนิก ของสัญญาณเสียงพูดเป็นตัวชี้วัดแบบ Subjective ใช้แสดงค่าขนาดของสเปกตรัมสำหรับแต่ละองค์ประกอบทางความถี่ ณ เวลาต่างๆ กล่าวคือ ทำการแสดงสเปกตรัม $|X(k,l)|$ ที่ความถี่ k และเฟรม l ใดๆ โดยใช้สีเป็นตัวแสดงผลกล่าวคือ สีเข้มแสดงถึงขนาดของสเปกตรัมที่สูงกว่าสีอ่อน

4.3 การทดลองที่ 1 ศึกษาประสิทธิภาพของวิธี MCSS, วิธี HR[10] และ วิธี HR ที่เสนอ

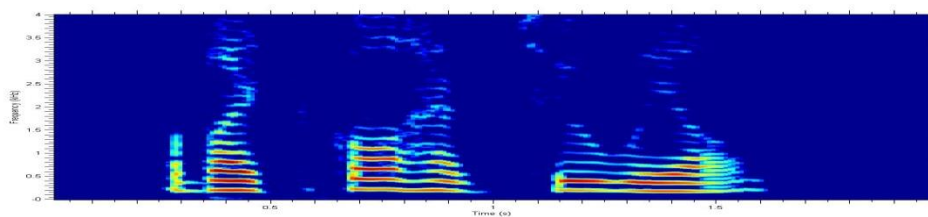
หัวข้อนี้แสดงผลเปรียบเทียบวิธี MCSS, วิธี HR[10] และ วิธี HR ที่เสนอ ในด้านการเพิ่มฮาร์มอนิกของเสียงพูด ตัวชี้วัดประสิทธิภาพที่ใช้ได้แก่ LSD และ สเปกโตรแกรม ทดสอบโดยใช้เสียงพูดที่ปนด้วยเสียงรบกวน 3 กรณีตามรูปที่ 4.3, 4.4 และ 4.5 คือกรณีสัญญาณรบกวนขาแบบเกาส์, สัญญาณรบกวนแบบเสียงพูดคุย และ เสียงในรถยนต์ ตามลำดับ

เมื่อพิจารณา กรณีเสียงพูดที่ปนด้วยเสียงรบกวนสัญญาณรบกวนขาแบบเกาส์ จากผลทดลองยกตัวอย่างในกรณีของเพศหญิง รูปที่ 4.4 ค) พบว่า วิธี MCSS สามารถลดเสียงรบกวนได้ดี แต่มีผลกระทบทำให้องค์ประกอบทางความถี่ของเสียงพูดโดยเฉพาะในช่วงความถี่สูงกว่า 1.5 kHz ซึ่งเป็นฮาร์มอนิกของเสียงพูดสูญหายไป เมื่อเปรียบเทียบกับสเปกโตรแกรมของเทคนิค MCSS+HR[10] (รูปที่ 4.3 ง) กับ สเปกโตรแกรมของวิธี MCSS+HR ที่เสนอ (รูปที่ 4.3 จ) พบว่าวิธี MCSS+HR ที่เสนอสามารถชดเชยฮาร์มอนิกที่บางความถี่ของสัญญาณเสียงพูดกลับคืนมาได้

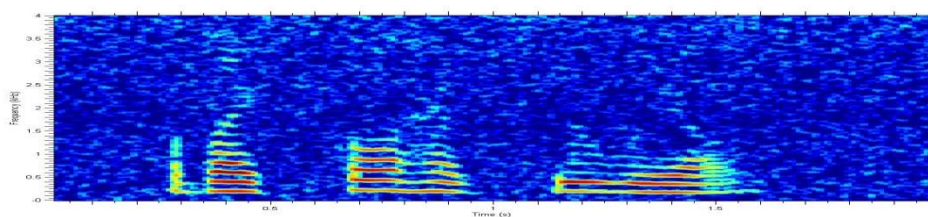
การใช้งานร่วมกันระหว่าง MCSS และ HR ที่เสนอมีประสิทธิภาพไม่ด้อยไปกว่าวิธี HR ใน[10] ดูได้จากตารางที่ 4.4 ค่า LSD ซึ่งสนับสนุนความถี่ที่ต่ำของวิธี HR ที่เสนอ ความถี่ที่ต่ำของวิธี HR ที่เสนอนั้น จะเป็นผลมาจากการปรับขนาดด้วยสัมประสิทธิ์ $\gamma \geq 1$ และ $\beta < 1$ เพื่อให้ใกล้เคียงกับเสียงพูดต้นฉบับมากที่สุด จากการจำลองผลพบว่าค่า γ มีช่วงการใช้งานอยู่ที่ $1 < \gamma \leq 15$ โดยที่ $\gamma < 1$ จะไม่มีการสร้างฮาร์มอนิก กลับคืนมา และการสร้างฮาร์มอนิกจะมีประสิทธิภาพใกล้เคียงกันตลอดเมื่อ $\gamma > 15$ สำหรับวิธี MCSS+HR[10] นั้นไม่มีการรักษาองค์ประกอบของรูปคลื่นในช่วง $\hat{s}(n) < 0$

สำหรับเสียงกรณีของเสียงพูดที่ปนด้วยเสียงรบกวนในรถยนต์ (รูปที่ 4.5) พบว่าการลดเสียงรบกวนด้วยวิธี MCSS อย่างเดียวมีค่าความถี่ต่ำสุด โดยดูจากตารางที่ 4.3 มีค่า LSD น้อยที่สุด เนื่องจากเสียงรบกวนในรถยนต์มีผลกระทบที่ความถี่ต่ำกว่า 1.5 kHz ดังนั้นจะไม่ก่อให้เกิดการสูญหายขององค์ประกอบความถี่สูง แต่อย่างไรก็ดี เมื่อเปรียบเทียบเทคนิคของการเพิ่มฮาร์มอนิก พบว่าวิธี MCSS+HR ที่เสนอ มีค่า LSD ที่ดีกว่าวิธี MCSS+HR[10]

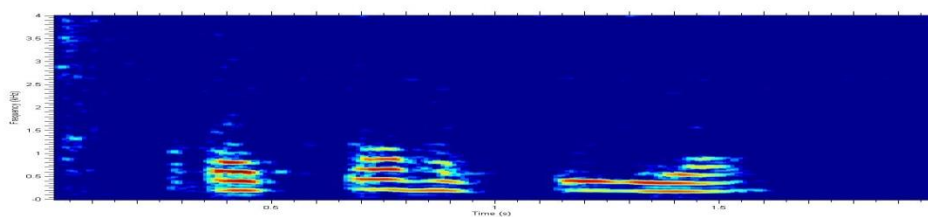
จากผลทดลองทั้งในกรณีเสียงพูดของเพศหญิงและเพศชาย เสียงพูดที่ปนด้วยเสียงรบกวนทั้ง WGN, เสียงพูดคุย และ เสียงรบกวนในรถยนต์ มีแนวโน้มของผลการทดลองไปในทางเดียวกัน ซึ่งการลดสัญญาณรบกวนแสดงค่า SegSNR ดังตารางที่ 4.2



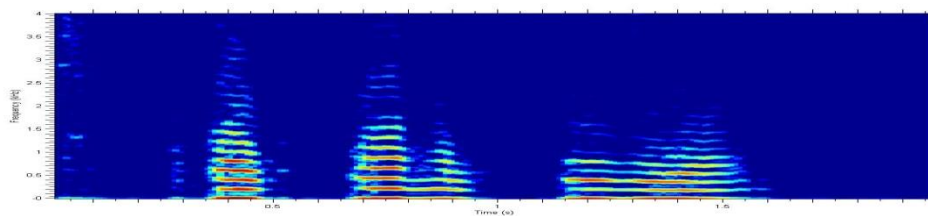
ก)



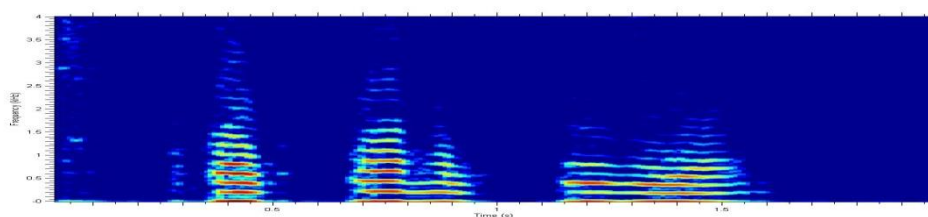
ข)



ค)



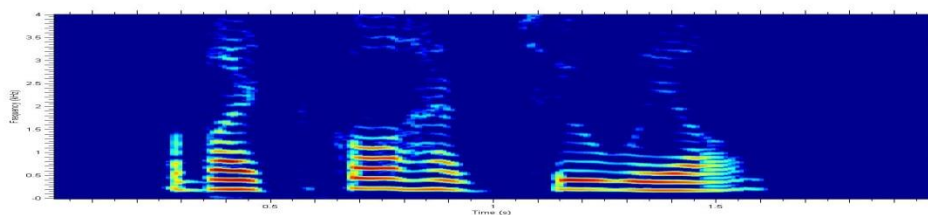
ง)



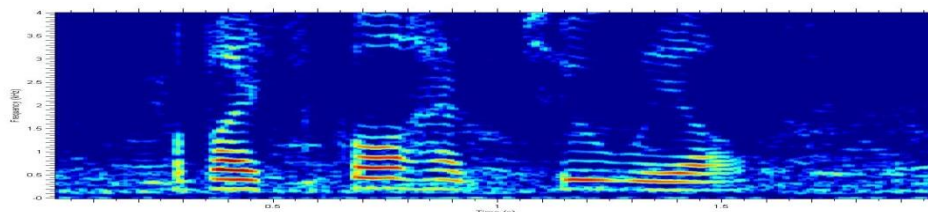
จ)

รูปที่ 4.3 สเปกโตรแกรมของ

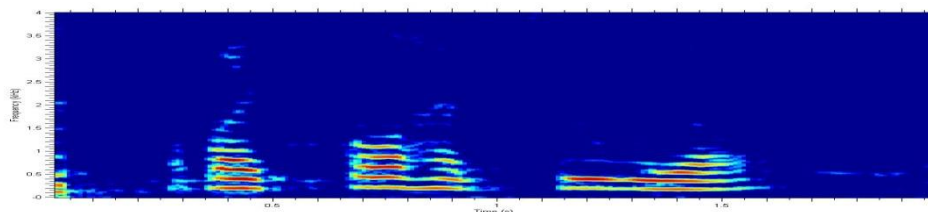
- ก) สัญญาณเสียงพูดต้นฉบับเสียงพูดเทศาหญิง ‘A bicycle has two wheels.’
 ข) สัญญาณเสียงพูดที่ถูกรบกวนด้วยสัญญาณรบกวนขาวแบบเกาส์ (WGN) 10 dB
 ค) การลดเสียงรบกวนด้วยวิธี MCSS
 ง) การเพิ่มฮาร์โมนิกของเสียงพูดด้วยวิธี MCSS + HR[10]
 จ) การเพิ่มฮาร์โมนิกของเสียงพูดด้วยวิธี MCSS + HR ที่เสนอ $\beta = 15$ และ $\gamma = 0.1$



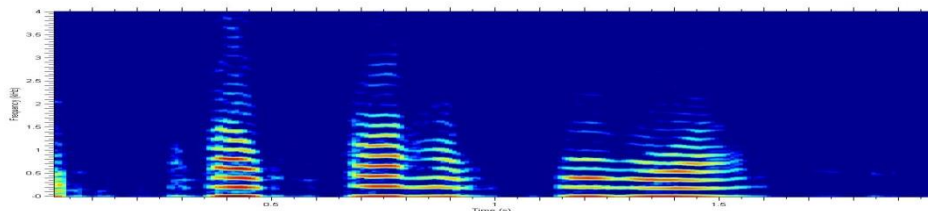
ก)



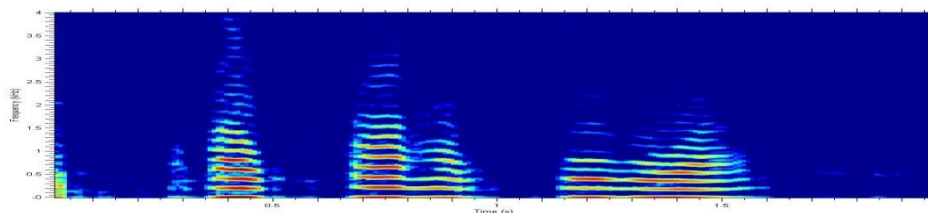
ข)



ค)



ง)



จ)

รูปที่ 4.4 สเปกโตรแกรมของ

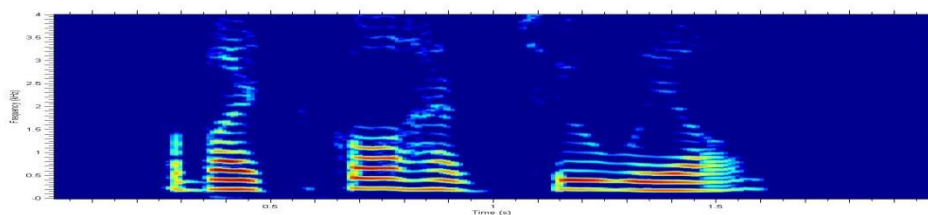
ก) สัญญาณเสียงพูดต้นฉบับเสียงพูดเพศหญิง ‘A bicycle has two wheels.’

ข) สัญญาณเสียงพูดที่ถูกรบกวนด้วยเสียงพูดคุษ (Babble) 10 dB

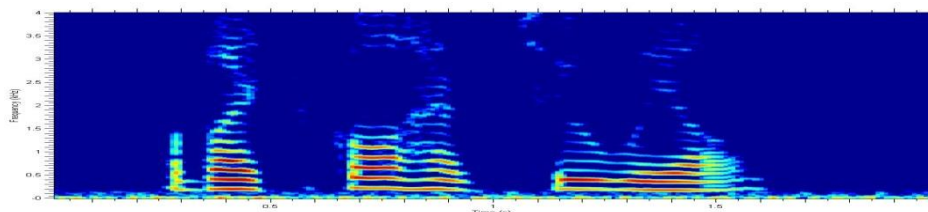
ค) การลดเสียงรบกวนด้วยวิธี MCSS

ง) การเพิ่มฮาร์โมนิกของเสียงพูดด้วยวิธี MCSS + HR[10]

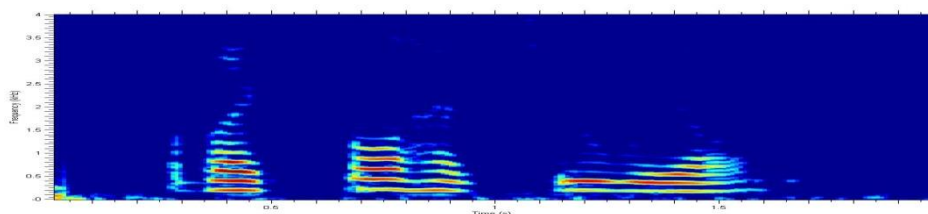
จ) การเพิ่มฮาร์โมนิกของเสียงพูดด้วยวิธี MCSS + HR ที่เสนอ $\beta = 15$ และ $\gamma = 0.1$



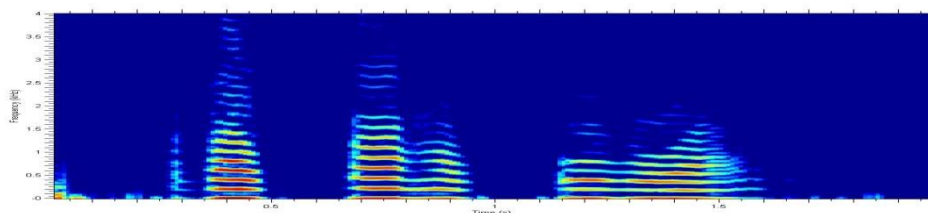
ก)



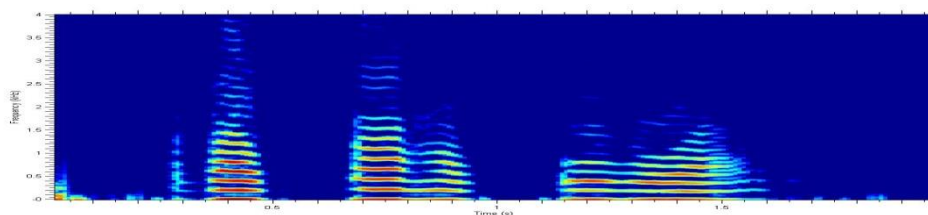
ข)



ค)



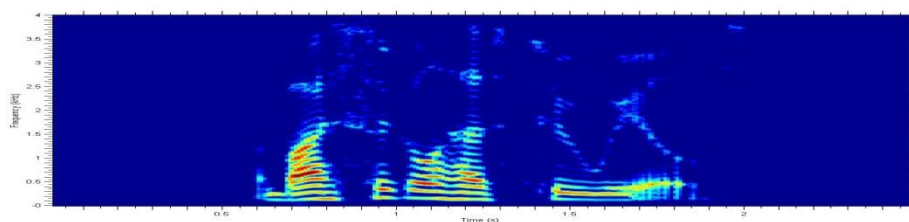
ง)



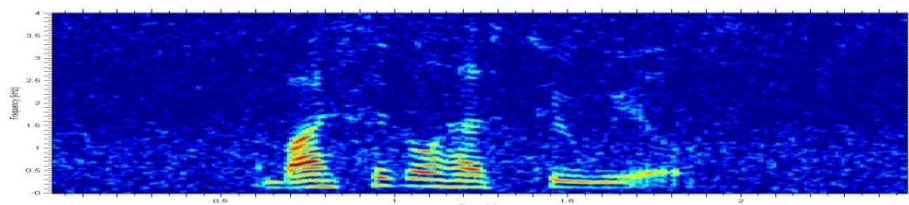
จ)

รูปที่ 4.5 สเปกโตรแกรมของ

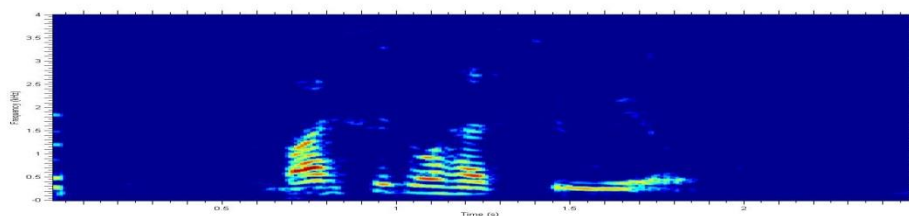
- ก) สัญญาณเสียงพูดต้นฉบับเสียงพูดเทศาภิบาล 'A bicycle has two wheels.'
- ข) สัญญาณเสียงพูดที่ถูกรบกวนด้วยเสียงรบกวนในรถยนต์ (Car noise) 10 dB
- ค) การลดเสียงรบกวนด้วยวิธี MCSS
- ง) การเพิ่มฮาร์โมนิกของเสียงพูดด้วยวิธี MCSS + HR[10]
- จ) การเพิ่มฮาร์โมนิกของเสียงพูดด้วยวิธี MCSS + HR ที่เสนอ $\beta = 15$ และ $\gamma = 0.1$



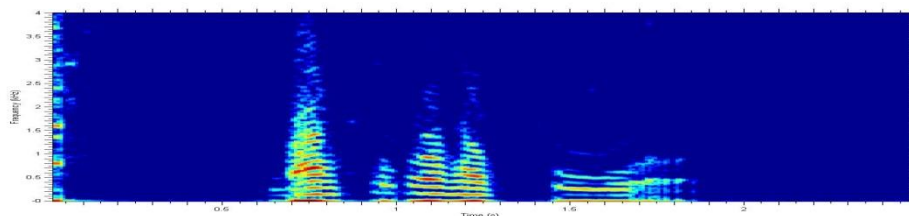
ก)



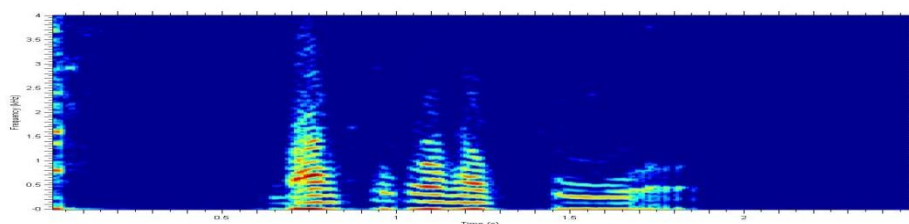
ข)



ค)



ง)



จ)

รูปที่ 4.6 สเปกโตรแกรมของ

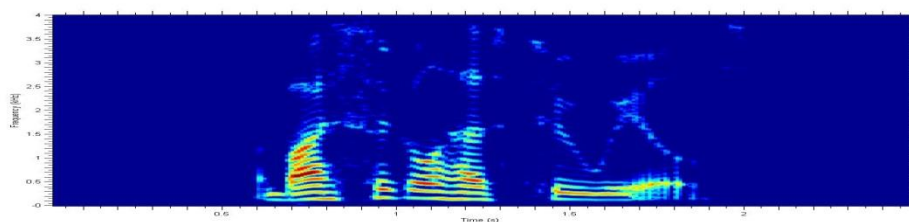
ก) สัญญาณเสียงพูดต้นฉบับเสียงพูดเพศชาย 'A bicycle has two wheels.'

ข) สัญญาณเสียงพูดที่ถูกรบกวนด้วยสัญญาณรบกวนขาวแบบเกาส์ (WGN) 10 dB

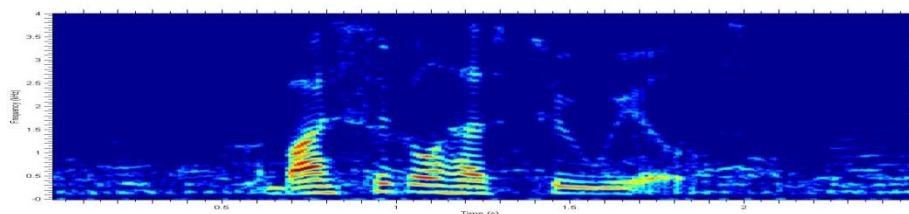
ค) การลดเสียงรบกวนด้วยวิธี MCSS

ง) การเพิ่มฮาร์โมนิกของเสียงพูดด้วยวิธี MCSS + HR[10]

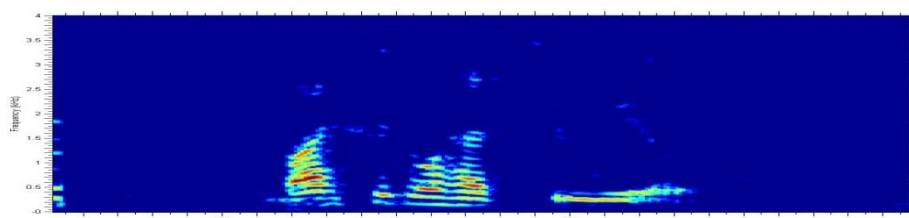
จ) การเพิ่มฮาร์โมนิกของเสียงพูดด้วยวิธี MCSS + HR ที่เสนอ $\beta = 15$ และ $\gamma = 0.1$



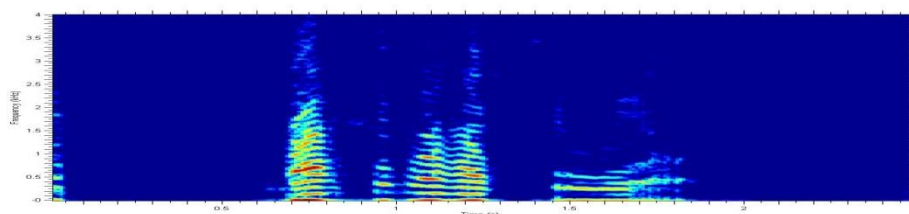
ก)



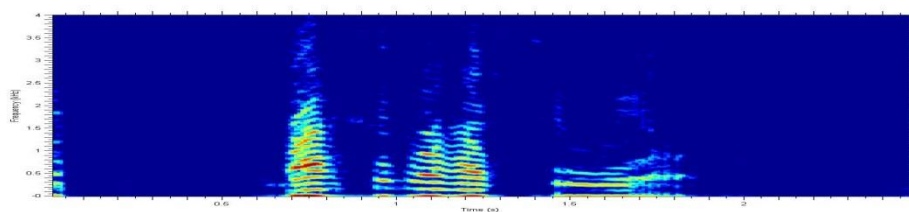
ข)



ค)



ง)



จ)

รูปที่ 4.7 สเปกโตรแกรมของ

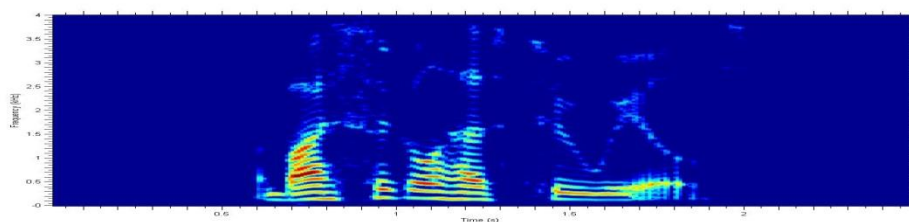
ก) สัญญาณเสียงพูดต้นฉบับเสียงพูดเพศชาย 'A bicycle has two wheels.'

ข) สัญญาณเสียงพูดที่ถูกบกรบกวนด้วยเสียงพูดคุย (Babble) 10 dB

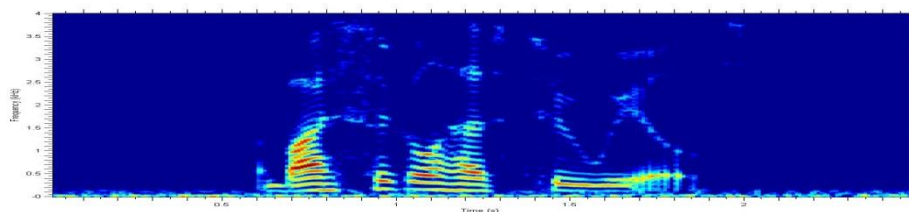
ค) การลดเสียงรบกวนด้วยวิธี MCSS

ง) การเพิ่มฮาร์โมนิกของเสียงพูดด้วยวิธี MCSS + HR[10]

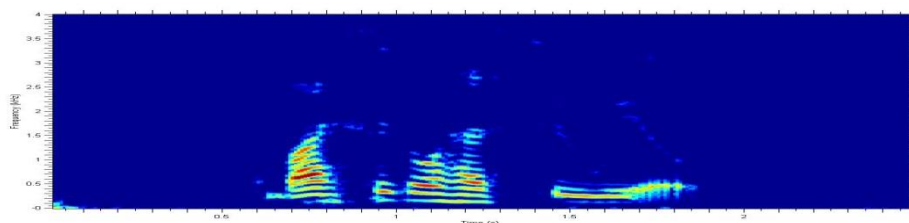
จ) การเพิ่มฮาร์โมนิกของเสียงพูดด้วยวิธี MCSS + HR ที่เสนอ $\beta = 15$ และ $\gamma = 0.1$



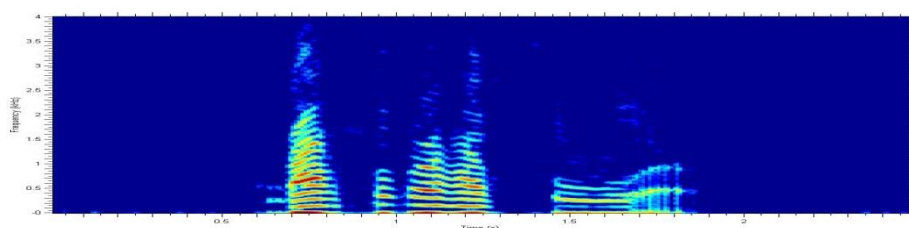
ก)



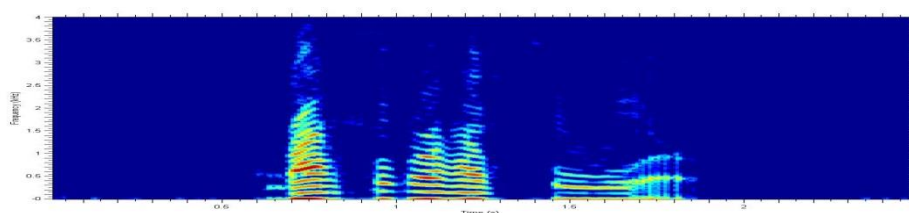
ข)



ค)



ง)



จ)

รูปที่ 4.8 สเปกโตรแกรมของ

- ก) สัญญาณเสียงพูดต้นฉบับเสียงพูดเพศชาย 'A bicycle has two wheels.'
 ข) สัญญาณเสียงพูดที่ถูกรบกวนด้วยเสียงรบกวนในรถยนต์ (Car noise) 10 dB
 ค) การลดเสียงรบกวนด้วยวิธี MCSS
 ง) การเพิ่มฮาร์โมนิกของเสียงพูดด้วยวิธี MCSS + HR[10]
 จ) การเพิ่มฮาร์โมนิกของเสียงพูดด้วยวิธี MCSS + HR ที่เสนอ $\beta = 15$ และ $\gamma = 0.1$

ตารางที่ 4.2 SegSNR ของวิธี MCSS, วิธี HR[10] และ วิธี HR ที่นำเสนอ

| Speech Type | Noise Type | Input SNR | SegSNR | | |
|-------------|--------------|-----------|--------|---------------|--------------------|
| | | | MCSS | MCSS + HR[10] | MCSS + HR Proposed |
| Female | WGN | 5 | 8.61 | 6.45 | 10.54 |
| | | 10 | 12.47 | 10.21 | 14.31 |
| | | 15 | 17.81 | 15.78 | 19.87 |
| | | 20 | 23.40 | 21.35 | 25.43 |
| | Babble Noise | 5 | 7.79 | 5.59 | 9.68 |
| | | 10 | 12.86 | 10.77 | 14.87 |
| | | 15 | 17.91 | 15.91 | 20.01 |
| | | 20 | 23.27 | 21.28 | 25.37 |
| | Car Noise | 5 | 8.45 | 6.41 | 10.50 |
| | | 10 | 13.27 | 11.30 | 15.39 |
| | | 15 | 18.15 | 16.16 | 20.25 |
| | | 20 | 23.29 | 21.32 | 25.42 |
| Male | WGN | 5 | 7.21 | 4.31 | 8.03 |
| | | 10 | 11.02 | 8.65 | 12.75 |
| | | 15 | 16.98 | 13.12 | 17.20 |
| | | 20 | 22.10 | 18.25 | 23.45 |
| | Babble Noise | 5 | 6.09 | 3.91 | 7.05 |
| | | 10 | 11.34 | 8.80 | 12.89 |
| | | 15 | 16.10 | 14.29 | 17.75 |
| | | 20 | 22.13 | 20.10 | 22.29 |
| | Car Noise | 5 | 7.49 | 5.20 | 9.29 |
| | | 10 | 12.61 | 10.23 | 14.33 |
| | | 15 | 17.21 | 15.38 | 19.02 |
| | | 20 | 22.12 | 20.27 | 24.41 |

ตารางที่ 4.3 Log Spectral Distance ของวิธี MCSS, วิธี HR[10] และ วิธี HR ที่เสนอ

| Speech Type | Noise Type | Input SNR | LSD | | |
|-------------|--------------|-----------|------|---------------|--------------------|
| | | | MCSS | MCSS + HR[10] | MCSS + HR Proposed |
| Female | WGN | 5 | 1.99 | 1.75 | 1.67 |
| | | 10 | 1.87 | 1.53 | 1.43 |
| | | 15 | 1.06 | 1.01 | 0.83 |
| | | 20 | 0.85 | 0.68 | 0.65 |
| | Babble Noise | 5 | 2.89 | 2.72 | 2.66 |
| | | 10 | 2.07 | 1.82 | 1.59 |
| | | 15 | 1.98 | 1.61 | 1.56 |
| | | 20 | 0.47 | 0.35 | 0.30 |
| | Car Noise | 5 | 0.98 | 1.21 | 1.08 |
| | | 10 | 0.86 | 0.93 | 0.87 |
| | | 15 | 0.28 | 0.42 | 0.39 |
| | | 20 | 0.20 | 0.31 | 0.23 |
| Male | WGN | 5 | 1.97 | 1.88 | 1.72 |
| | | 10 | 1.93 | 1.62 | 1.50 |
| | | 15 | 1.12 | 1.05 | 0.89 |
| | | 20 | 0.81 | 0.74 | 0.72 |
| | Babble Noise | 5 | 2.39 | 3.01 | 2.11 |
| | | 10 | 2.20 | 2.11 | 1.81 |
| | | 15 | 2.01 | 1.71 | 1.59 |
| | | 20 | 1.51 | 1.35 | 0.29 |
| | Car Noise | 5 | 2.33 | 2.52 | 3.57 |
| | | 10 | 2.29 | 2.35 | 2.31 |
| | | 15 | 1.81 | 2.01 | 1.94 |
| | | 20 | 1.71 | 1.95 | 1.82 |

4.4 การทดลองที่ 2 ผลกระทบของอันดับการโปรเจกชันของวิธี DAP

การทดลองเปรียบเทียบประสิทธิภาพของอันดับการโปรเจกชันของวิธี DAP โดยการทดลองเปรียบเทียบ $p_H = 4$ และ $p_L = 20$ ในด้านความซับซ้อนและอัตราการรู้เข้าของระบบ ทดสอบโดยใช้เสียงพูดที่ปนด้วยเสียงรบกวน 3 แบบเมื่อเป็นสัญญาณรบกวนขาวแบบเกาส์, สัญญาณรบกวนพุดคุย และ เสียงในรถยนต์

สำหรับในกรณีสัญญาณรบกวนขาวแบบเกาส์ ยกตัวอย่างเสียงพูดเพศหญิง แสดงสเปกโตรแกรมรูปที่ 4.9 ค) และ 4.9 ง), การลดเสียงรบกวนด้วยวิธี DAP ที่ $p_H = 4$ และการลดเสียงรบกวนด้วยวิธี DAP ที่ $p_L = 20$ ตามลำดับ และตารางที่ 4.4 ค่า SegSNR พบว่า การลดเสียงรบกวนด้วยวิธี DAP ที่อันดับการโปรเจกชันสูง ($p_L = 20$) สามารถลดสัญญาณรบกวนได้ดีกว่าเมื่ออันดับการโปรเจกชันต่ำ ($p_H = 4$) แต่อย่างไรก็ตามวิธี DAP ที่อันดับการโปรเจกชันสูงมีความซับซ้อนสูงกว่า แสดงดังตาราง 3.1

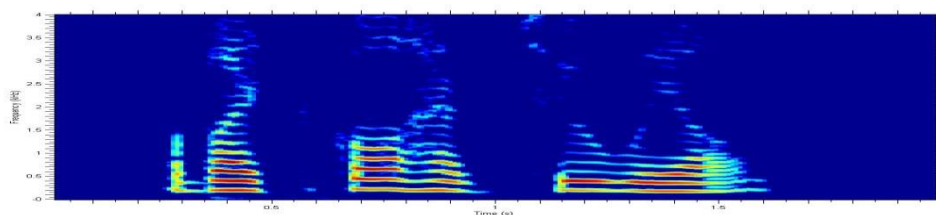
จากผลทดลองทั้งในกรณีเสียงพูดของเพศหญิงและเพศชาย เสียงพูดที่ปนด้วยเสียงรบกวนทั้ง WGN, เสียงพุดคุย และ เสียงรบกวนในรถยนต์ มีแนวโน้มของผลการทดลองไปในทางเดียวกัน และค่าคุณภาพของเสียงพูดที่ถูกลดสัญญาณรบกวนแล้วแสดงค่า LSD ดังตารางที่ 4.5

4.5 การทดลองที่ 3 ประสิทธิภาพของขั้นตอนวิธี DAP แบบหลายอันดับ

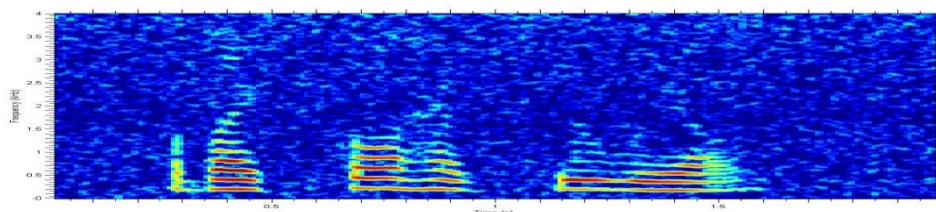
การทดลองเปรียบเทียบประสิทธิภาพของอันดับการโปรเจกชันของวิธี DAP โดยการทดลองเปรียบเทียบ $p_H = 4$ และ $p_L = 20$ กับ ขั้นตอนวิธี DAP ที่ $p_H = 4$ ใช้งานร่วมกับ $p_L = 20$ โดยการแบ่งช่วงความถี่ย่อย ในด้านความซับซ้อน การลดสัญญาณรบกวนและคุณภาพของสัญญาณเสียงพูดที่ถูกลดสัญญาณรบกวนแล้ว ทดสอบโดยใช้เสียงพูดที่ปนด้วยเสียงรบกวน 3 แบบ คือเป็นสัญญาณรบกวนขาวแบบเกาส์, สัญญาณรบกวนพูดคุย และ เสียงในรถยนต์

การทดลองทำการแบ่งสัญญาณเสียงพูดที่ปนด้วยเสียงรบกวนออกเป็น 2 ช่วงความถี่ โดยใช้ตัวกรองความถี่ต่ำ (LPF) และตัวกรองความถี่สูง (HPF) ซึ่งมีความถี่ตัดอยู่ที่ 0 – 1.5 kHz ในช่วงความถี่สูงใช้การลดสัญญาณรบกวนด้วยวิธี DAP ที่ $p_H = 4$ และในช่วงความถี่ต่ำใช้การลดสัญญาณรบกวนด้วยวิธี DAP ที่ $p_L = 20$

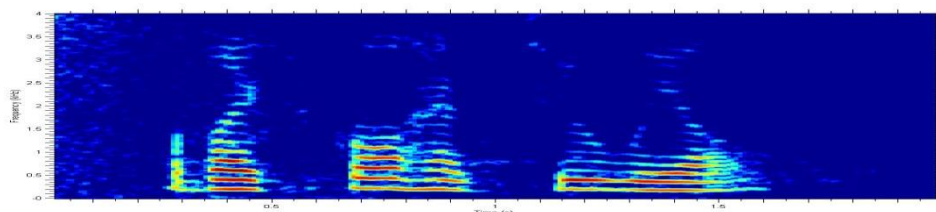
สำหรับในกรณีสัญญาณรบกวนขาวแบบเกาส์ ยกตัวอย่างเสียงพูดเพศหญิง แสดงสเปกโตรแกรมรูปที่ 4.9 จ) พบว่าวิธีที่เสนอสามารถลดสัญญาณรบกวนในช่วงความถี่ต่ำและสามารถรักษารูปประกอบทางความถี่ที่ความถี่สูงกว่า 1.5 kHz ประสิทธิภาพของการลดสัญญาณรบกวนและความเพี้ยนของสัญญาณแสดงในค่า SegSNR (ตารางที่ 4.4) และ LSD (ตารางที่ 4.5) ตามลำดับ เป็นเช่นเดียวกันกับในกรณีที่เป็นสัญญาณรบกวนแบบเสียงพูดคุย และ ในรถยนต์ และ เทคนิควิธี DAP แบบหลายอันดับที่เสนอยังมีความซับซ้อนต่ำกว่าเมื่อใช้ DAP ที่ $p_L = 20$ อย่างเดียว ความซับซ้อนในการคำนวณแสดงดังตารางที่ 3.1



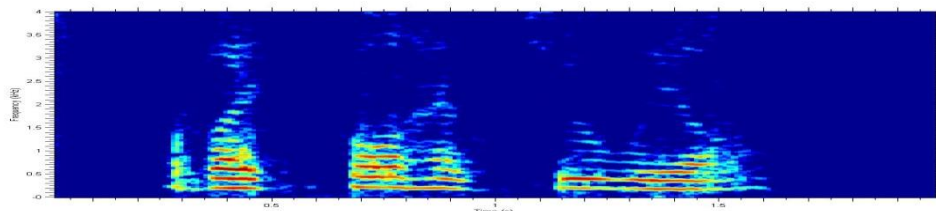
ก)



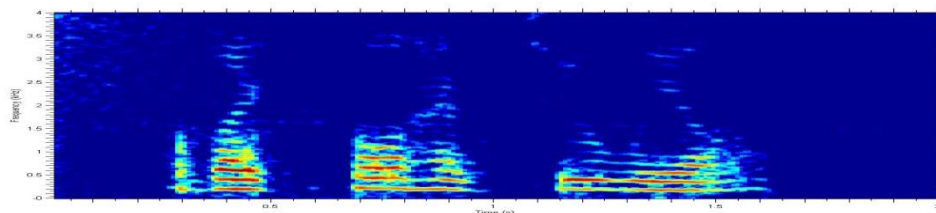
ข)



ค)



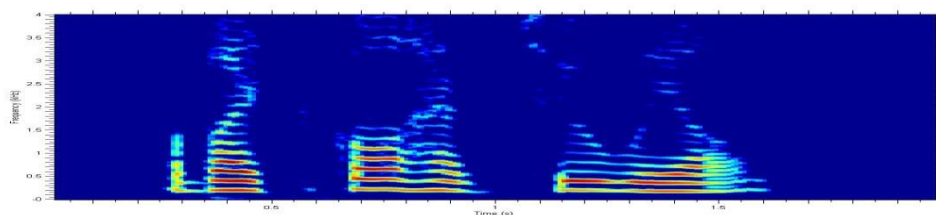
ง)



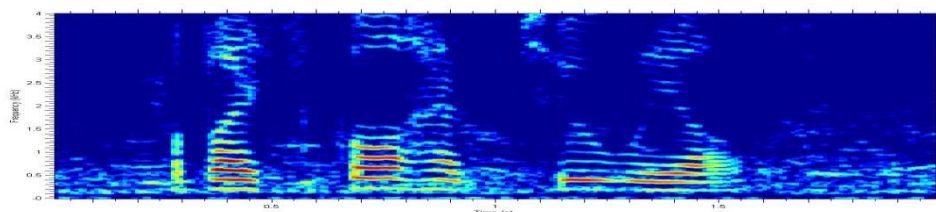
จ)

รูปที่ 4.9 สเปกโตรแกรมของ

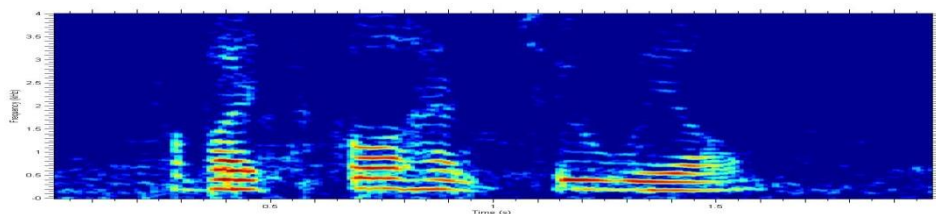
- ก) สัญญาณเสียงพูดต้นฉบับเสียงพูดเพชหึง ‘A bicycle has two wheels.’
 ข) สัญญาณเสียงพูดที่ถูกรบกวนด้วยสัญญาณรบกวนขาวแบบเกาส์ (WGN) 10 dB
 ค) การลดเสียงรบกวนด้วยวิธี DAP ที่ $p = 4$
 ง) การลดเสียงรบกวนด้วยวิธี DAP ที่ $p = 20$
 จ) ขั้นตอนวิธี DAP แบบหลายอันดับ



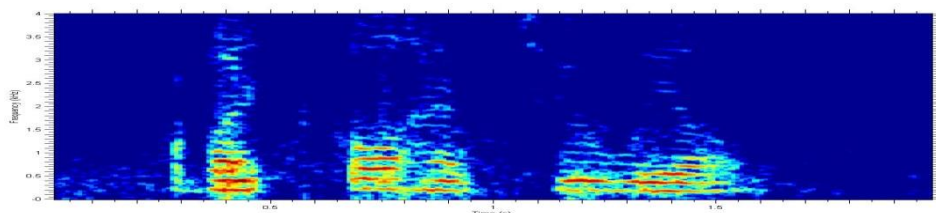
ก)



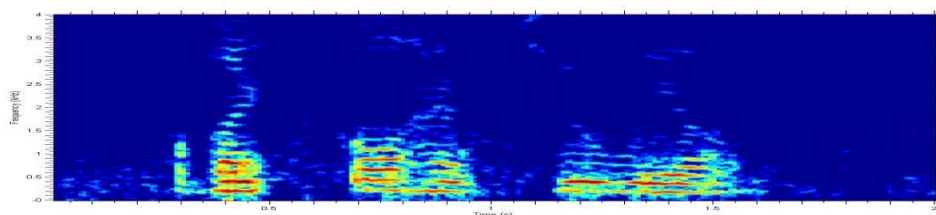
ข)



ค)



ง)



จ)

รูปที่ 4.10 สเปกโตรแกรมของ

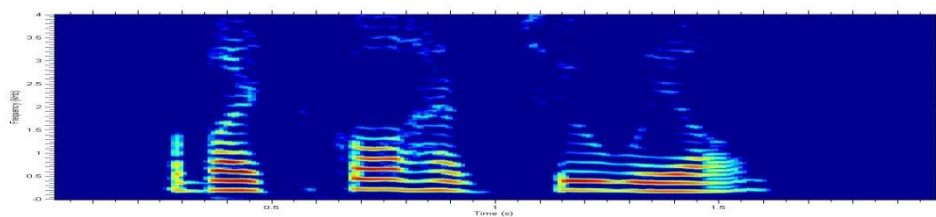
ก) สัญญาณเสียงพูดต้นฉบับเสียงพูดเทศหึง ‘A bicycle has two wheels.’

ข) สัญญาณเสียงพูดที่ถูกรบกวนด้วยเสียงพูดคุย (Babble Noise) 10 dB

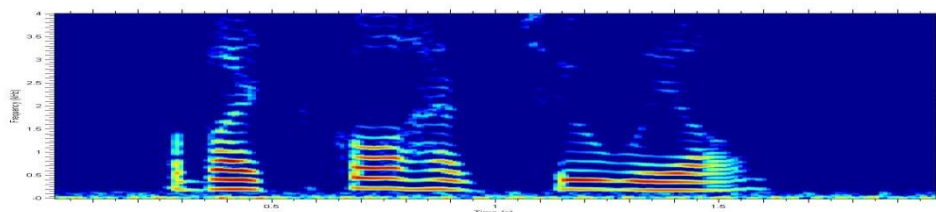
ค) การลดเสียงรบกวนด้วยวิธี DAP ที่ $p = 4$

ง) การลดเสียงรบกวนด้วยวิธี DAP ที่ $p = 20$

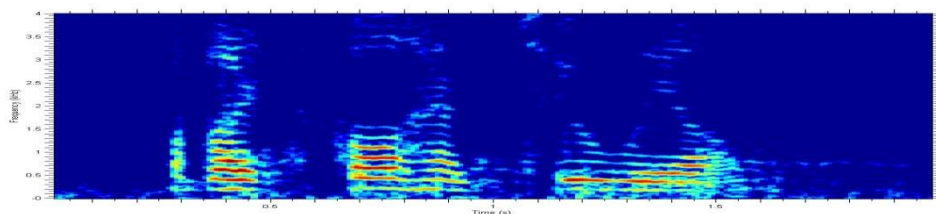
จ) ขั้นตอนวิธี DAP แบบหลายอันดับ



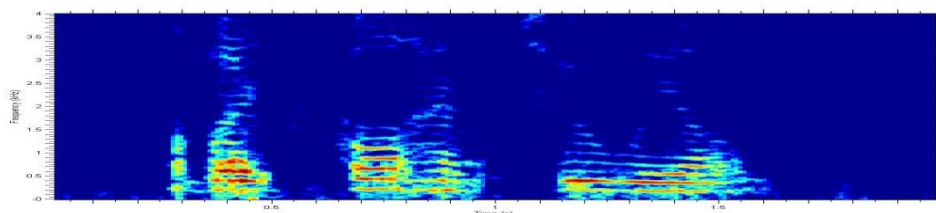
ก)



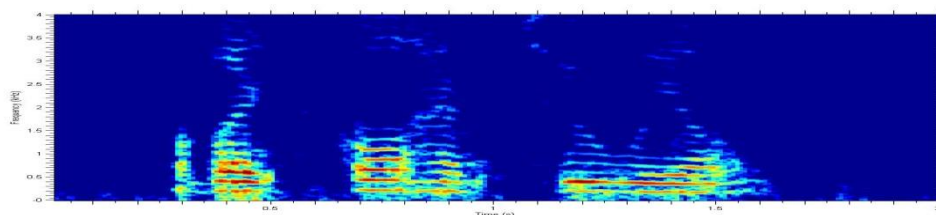
ข)



ค)



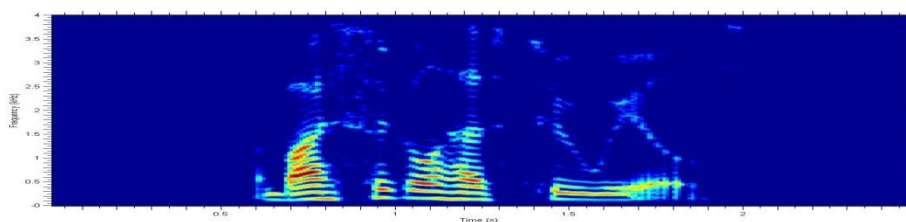
ง)



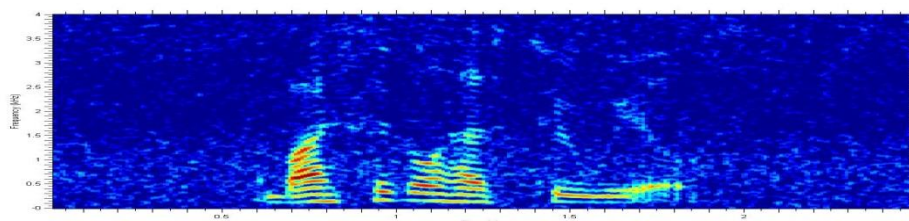
จ)

รูปที่ 4.11 สเปกโตรแกรมของ

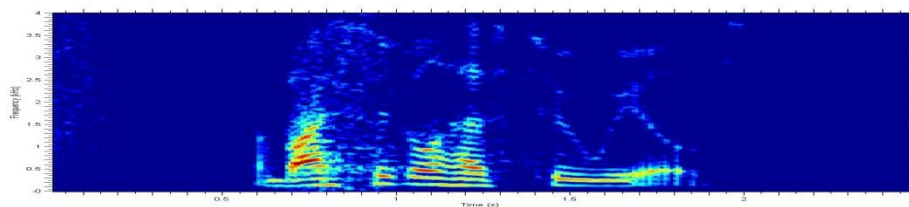
- ก) สัญญาณเสียงพูดต้นฉบับเสียงพูดเทศาภิบาล 'A bicycle has two wheels.'
 ข) สัญญาณเสียงพูดที่ถูกรบกวนด้วยเสียงรบกวนในรถยนต์ (Car Noise) 10 dB
 ค) การลดเสียงรบกวนด้วยวิธี DAP ที่ $p = 4$
 ง) การลดเสียงรบกวนด้วยวิธี DAP ที่ $p = 20$
 จ) ขั้นตอนวิธี DAP แบบหลายอันดับ



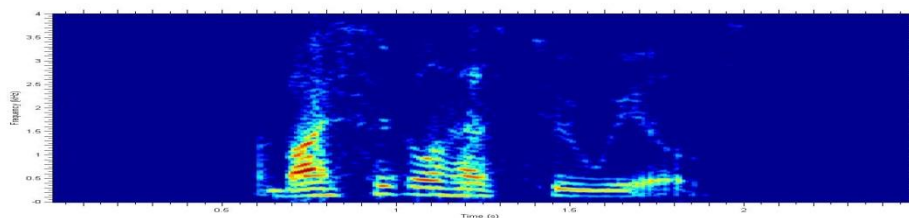
ก)



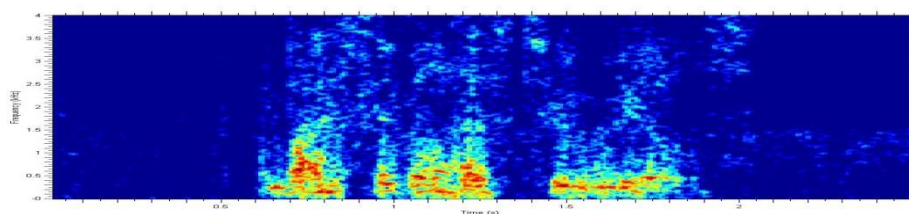
ข)



ค)



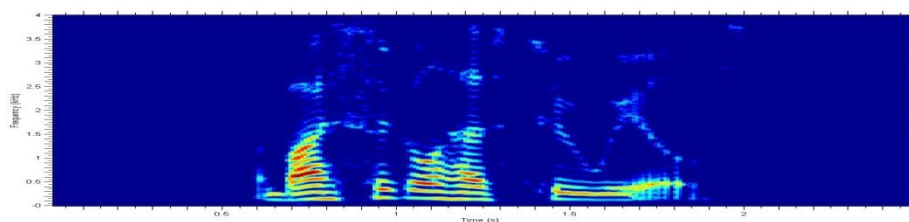
ง)



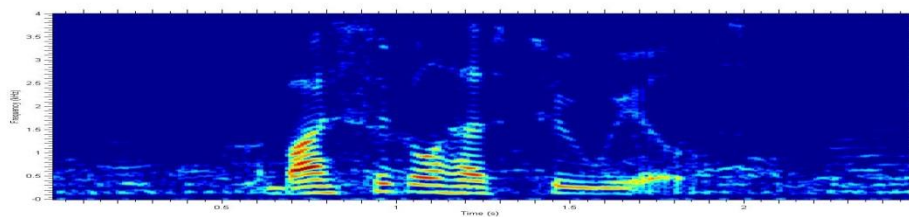
จ)

รูปที่ 4.12 สเปกโตรแกรมของ

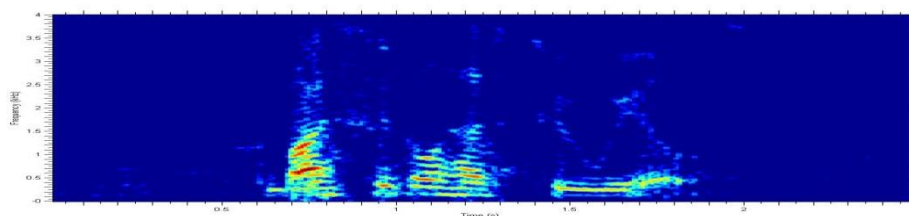
- ก) สัญญาณเสียงพูดต้นฉบับเสียงพูดเพศชาย 'A bicycle has two wheels.'
 ข) สัญญาณเสียงพูดที่ถูกรบกวนด้วยสัญญาณรบกวนขาวแบบเกาส์ (WGN) 10 dB
 ค) การลดเสียงรบกวนด้วยวิธี DAP ที่ $p = 4$
 ง) การลดเสียงรบกวนด้วยวิธี DAP ที่ $p = 20$
 จ) ขั้นตอนวิธี DAP แบบหลายอันดับ



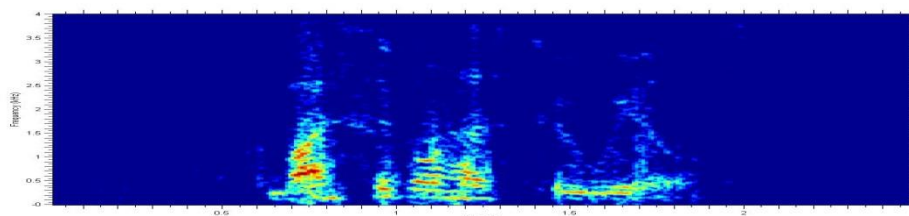
ก)



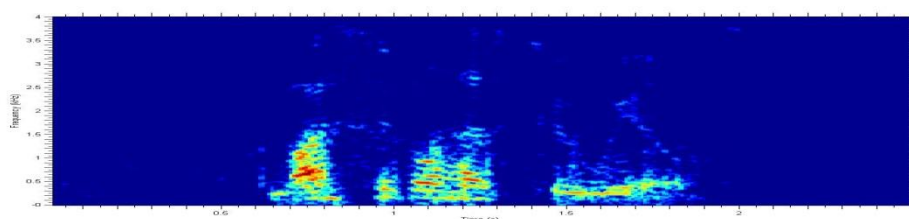
ข)



ค)



ง)



จ)

รูปที่ 4.13 สเปกโตรแกรมของ

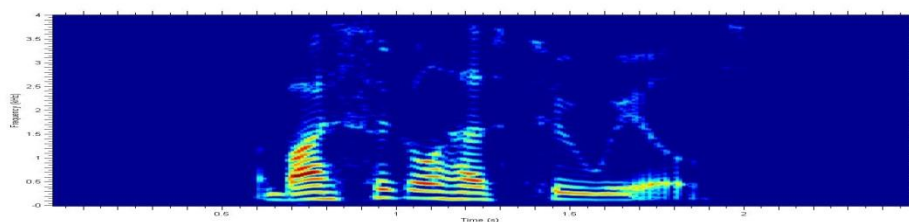
ก) สัญญาณเสียงพูดต้นฉบับเสียงพูดเพศชาย 'A bicycle has two wheels.'

ข) สัญญาณเสียงพูดที่ถูกรบกวนด้วยเสียงพูดคุย (Babble Noise) 10 dB

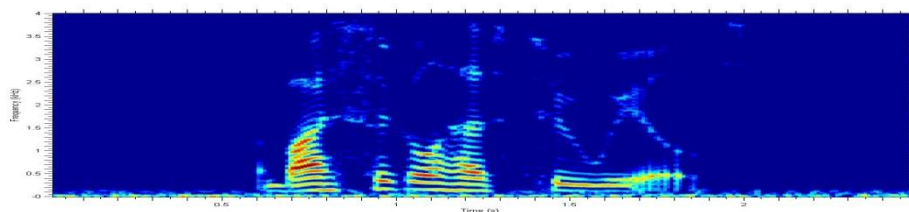
ค) การลดเสียงรบกวนด้วยวิธี DAP ที่ $p = 4$

ง) การลดเสียงรบกวนด้วยวิธี DAP ที่ $p = 20$

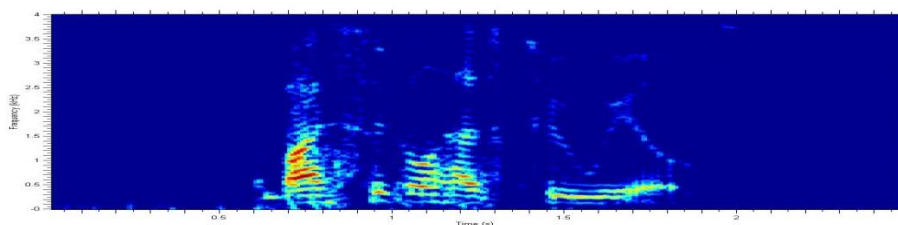
จ) ขั้นตอนวิธี DAP แบบหลายอันดับ



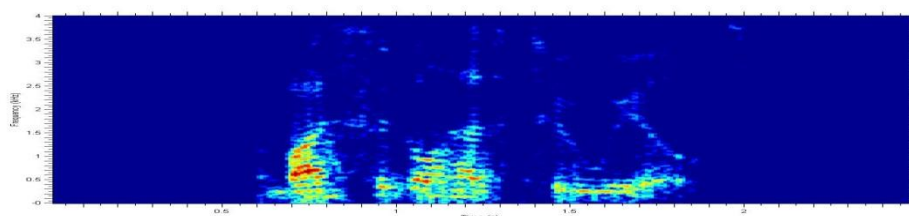
ก)



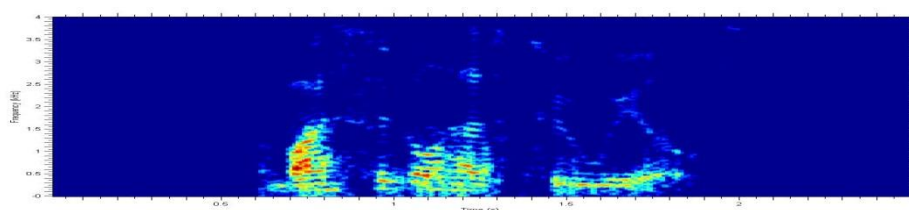
ข)



ค)



ง)



จ)

รูปที่ 4.14 สเปกโตรแกรมของ

- ก) สัญญาณเสียงพูดต้นฉบับเสียงพูดเพศชาย 'A bicycle has two wheels.'
- ข) สัญญาณเสียงพูดที่ถูกรบกวนด้วยเสียงรบกวนในรถยนต์ (Car Noise) 10 dB
- ค) การลดเสียงรบกวนด้วยวิธี DAP ที่ $p = 4$
- ง) การลดเสียงรบกวนด้วยวิธี DAP ที่ $p = 20$
- จ) ขั้นตอนวิธี DAP แบบหลายอันดับ

4.6 การทดลองที่ 4 ประสิทธิภาพของเทคนิคร่วมระหว่างขั้นตอนวิธี DAP และ MCSS

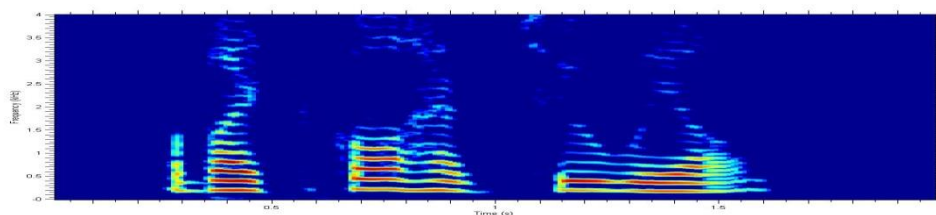
การทดลองเพื่อดูประสิทธิภาพของเทคนิคการแบ่งแถบความถี่ย่อยที่ใช้กับวิธี MCSS ร่วมกับวิธี DAP ที่เสนอในด้านความซับซ้อน การลดสัญญาณรบกวนและความเพี้ยนของสัญญาณ ทดสอบโดยใช้เสียงพูดที่ปนด้วยเสียงรบกวน 3 แบบเมื่อเป็นสัญญาณรบกวนขาวแบบเกาส์, สัญญาณรบกวนพุดคุย (Babble) และ เสียงในรถยนต์

การทดลองทำการแบ่งสัญญาณเสียงพูดที่ปนด้วยเสียงรบกวนออกเป็น 2 แถบความถี่ย่อย โดยใช้ตัวกรองความถี่ต่ำ (LPF) และตัวกรองความถี่สูง (HPF) ซึ่งมีความถี่ตัดอยู่ที่ 0 - 1 kHz ในแถบความถี่สูงใช้การลดสัญญาณรบกวนด้วยวิธี DAP ที่ $p = 4$ และที่แถบความถี่ต่ำใช้การลดสัญญาณรบกวนด้วยวิธี MCSS

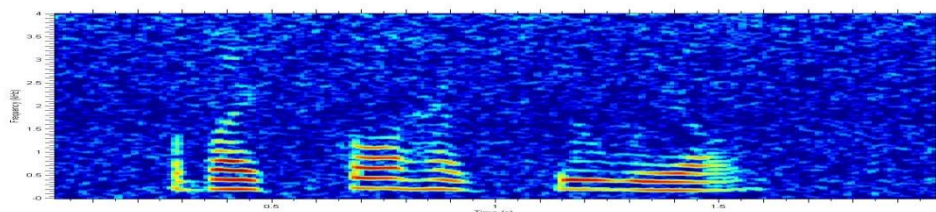
จากรูปที่ 4.15 ค) แสดงสเปกโตรแกรมในกรณีที่เป็นสัญญาณรบกวนขาวแบบเกาส์ พบว่าเทคนิคร่วมระหว่างขั้นตอนวิธี DAP และ MCSS ที่นำเสนอสามารถลดสัญญาณรบกวนที่แถบความถี่ต่ำและสามารถรักษารักษาองค์ประกอบทางความถี่ที่ความถี่สูงกว่า 1.5 kHz ประสิทธิภาพของการลดสัญญาณรบกวนและคุณภาพของสัญญาณเสียงพูดที่ถูกลดสัญญาณรบกวนแล้ว แสดงในค่า SegSNR (ตารางที่ 4.4) และ LSD (ตารางที่ 4.5) ตามลำดับ เป็นเช่นเดียวกันกับในกรณีที่เป็นสัญญาณรบกวนในรถยนต์

ในกรณีที่เป็นสัญญาณรบกวนแบบเสียงพุดคุย มีประสิทธิภาพอยู่ระหว่างวิธี DAP ที่ใช้ $p = 20$ และ เทคนิคร่วมระหว่างขั้นตอนวิธี DAP และ MCSS เนื่องจาก สัญญาณรบกวนแบบเสียงพุดคุยมีคุณสมบัติทางสถิติไม่คงที่ การลดสัญญาณรบกวนด้วยวิธี DAP ที่ความถี่สูงมีการลู่เข้าช้า

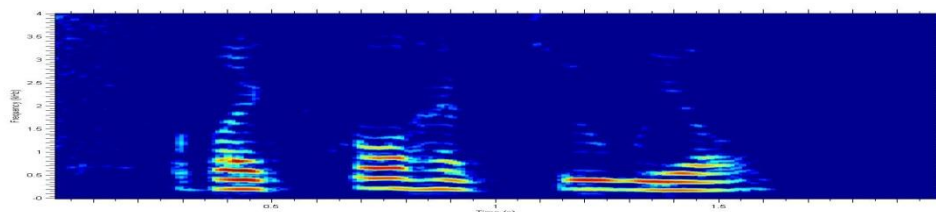
เทคนิคร่วมระหว่างขั้นตอนวิธี DAP และ MCSS ที่เสนอยังมีความซับซ้อนต่ำกว่าเมื่อใช้ DAP ที่ $p_L = 20$ อย่างเดียว ความซับซ้อนในการคำนวณ แสดงดังตารางที่ 3.1 สำหรับการ ใช้วิธี MCSS อย่างเดียว วิธีที่เสนอใน [9] ใช้ตัวแปรควบคุมการลบสเปกตรัม $\alpha = 8$ ซึ่งแตกต่างจากเทคนิคร่วมระหว่างขั้นตอนวิธี DAP และ MCSS ที่นำเสนอ คือใช้ $\alpha = 4$



ก)



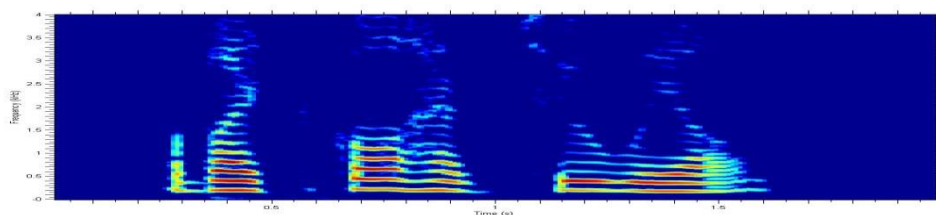
ข)



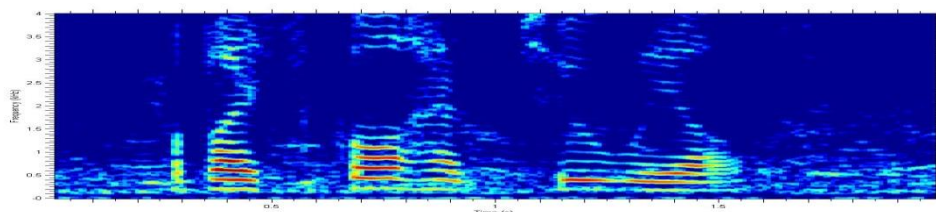
ค)

รูปที่ 4.15 สเปกโตรแกรมของ

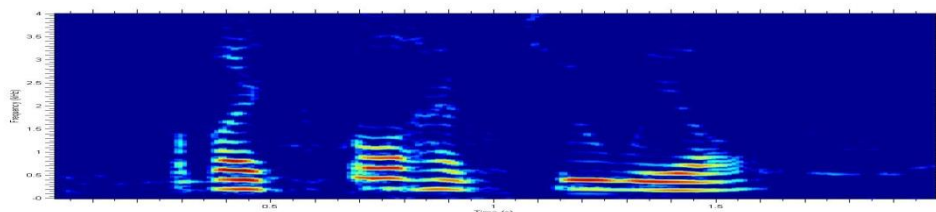
- ก) สัญญาณเสียงพูดต้นฉบับเสียงพูดเทศาหญิง 'A bicycle has two wheels.'
- ข) สัญญาณเสียงพูดที่ถูกรบกวนด้วยสัญญาณรบกวนขาวแบบเกาส์ (WGN) 10 dB
- ค) เทคนิคร่วมระหว่างขั้นตอนวิธี DAP และ MCSS



ก)



ข)



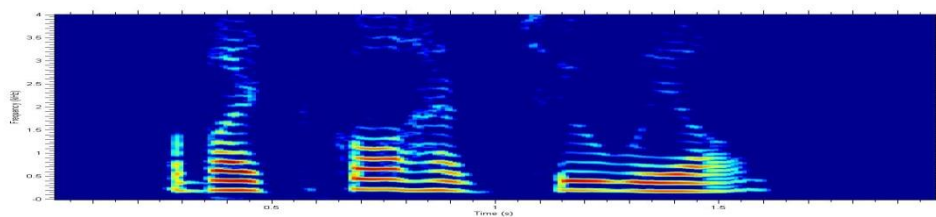
ค)

รูปที่ 4.16 สเปกโตรแกรมของ

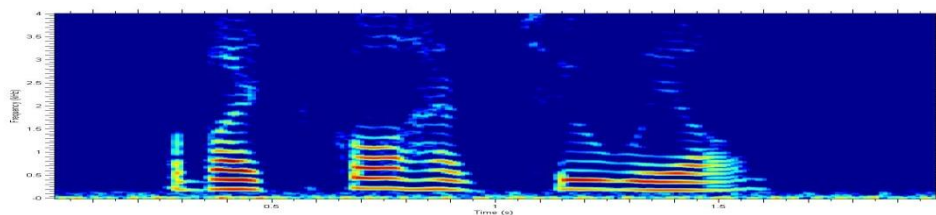
ก) สัญญาณเสียงพูดต้นฉบับเสียงพูดเทศาหญิง 'A bicycle has two wheels.'

ข) สัญญาณเสียงพูดที่ถูกรบกวนด้วยเสียงพูดคุย (Babble) 10 dB

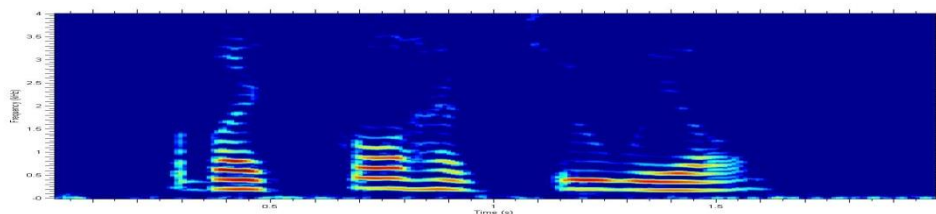
ค) เทคนิคร่วมระหว่างขั้นตอนวิธี DAP และ MCSS



ก)



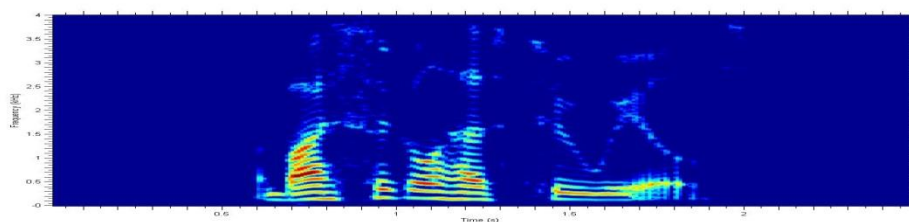
ข)



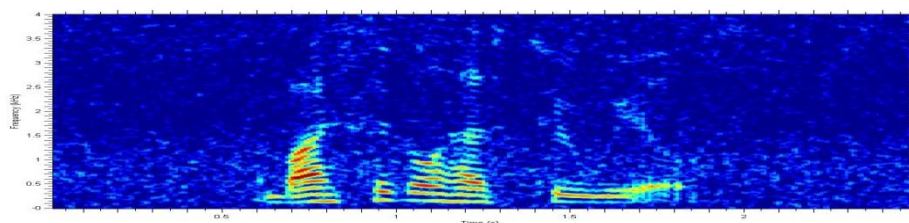
ค)

รูปที่ 4.17 สเปกโตรแกรมของ

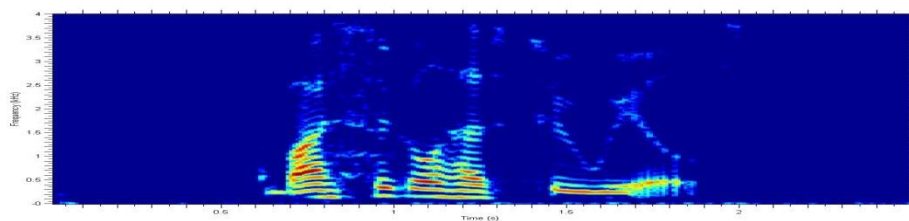
- ก) สัญญาณเสียงพูดต้นฉบับเสียงพูดเทศาหญิง 'A bicycle has two wheels.'
- ข) สัญญาณเสียงพูดที่ถูกรบกวนด้วยเสียงรบกวนในรถยนต์ (Car Noise) 10 dB
- ค) เทคนิคร่วมระหว่างขั้นตอนวิธี DAP และ MCSS



ก)



ข)



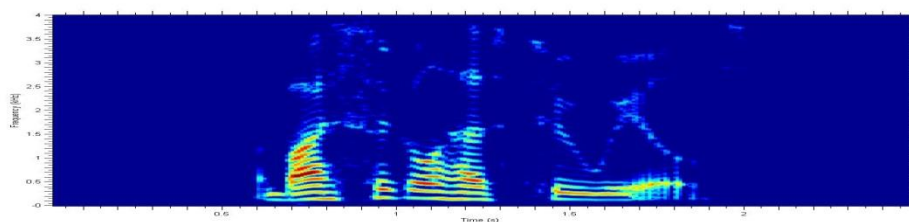
ค)

รูปที่ 4.18 สเปกโตรแกรมของ

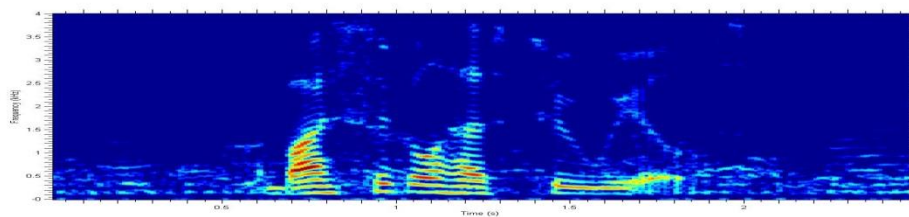
ก) สัญญาณเสียงพูดต้นฉบับเสียงพูดเพศชาย 'A bicycle has two wheels.'

ข) สัญญาณเสียงพูดที่ถูกรบกวนด้วยสัญญาณรบกวนขาวแบบเกาส์ (WGN) 10 dB

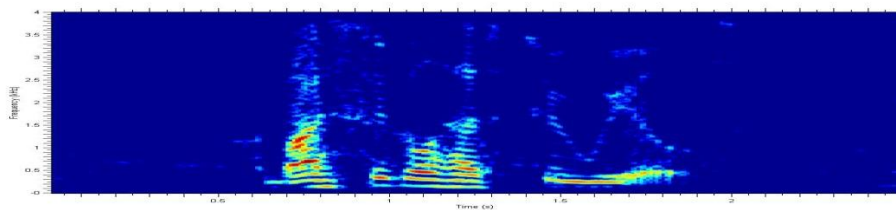
ค) เทคนิคร่วมระหว่างขั้นตอนวิธี DAP และ MCSS



ก)



ข)



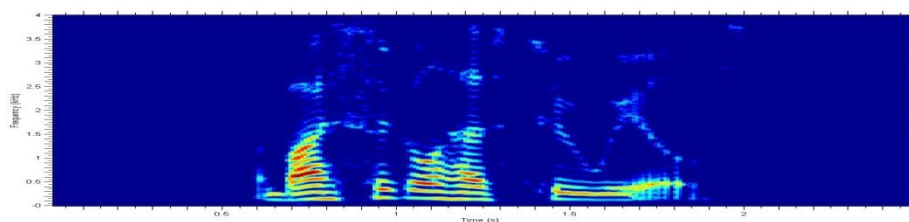
ค)

รูปที่ 4.19 สเปกโตรแกรมของ

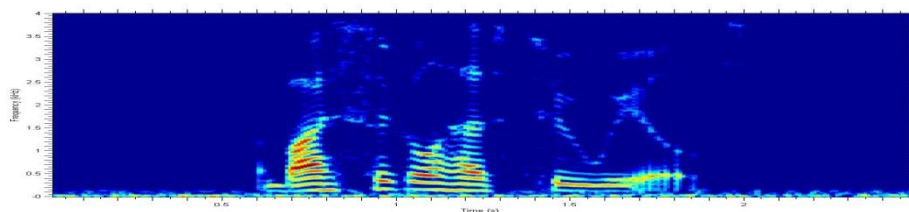
ก) สัญญาณเสียงพูดต้นฉบับเสียงพูดเพศชาย 'A bicycle has two wheels.'

ข) สัญญาณเสียงพูดที่ถูกรบกวนด้วยเสียงพูดคุย (Babble) 10 dB

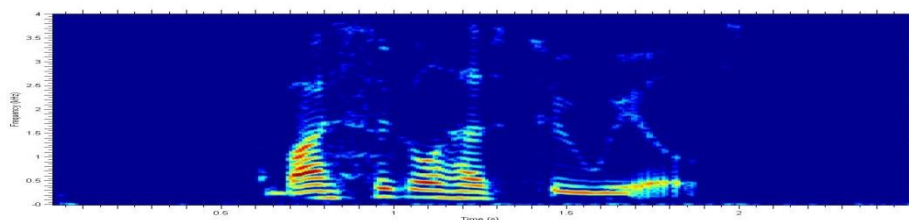
ค) เทคนิคร่วมระหว่างขั้นตอนวิธี DAP และ MCSS



ก)



ข)



ค)

รูปที่ 4.20 สเปกโตรแกรมของ

- ก) สัญญาณเสียงพูดต้นฉบับเสียงพูดเพศชาย 'A bicycle has two wheels.'
 ข) สัญญาณเสียงพูดที่ถูกรบกวนด้วยเสียงรบกวนในรถยนต์ (Car Noise) 10 dB
 ค) เทคนิคร่วมระหว่างขั้นตอนวิธี DAP และ MCSS

ตารางที่ 4.4 Segmental Signal-To-Noise Ratio (SegSNR)

| Speech Type | Noise Type | Input SNR | SegSNR | | | | |
|-------------|--------------|-----------|--------|-----------------|------------------|----------------------|---------------------------|
| | | | MCSS | DAP ($p = 4$) | DAP ($p = 20$) | DAP Subband Proposed | DAP+MCSS Subband Proposed |
| Female | WGN | 5 | 8.24 | 5.91 | 8.03 | 3.03 | 7.96 |
| | | 10 | 11.48 | 10.32 | 11.49 | 11.48 | 12.93 |
| | | 15 | 18.04 | 15.13 | 15.59 | 15.30 | 18.27 |
| | | 20 | 22.74 | 20.03 | 20.83 | 18.51 | 23.08 |
| | Babble Noise | 5 | 7.79 | 7.43 | 7.87 | 6.03 | 8.11 |
| | | 10 | 11.62 | 10.91 | 12.82 | 12.52 | 12.79 |
| | | 15 | 18.17 | 15.39 | 16.81 | 17.14 | 18.24 |
| | | 20 | 22.92 | 20.23 | 21.71 | 21.86 | 23.16 |
| | Car Noise | 5 | 8.08 | 4.53 | 7.56 | 7.38 | 8.14 |
| | | 10 | 11.78 | 10.26 | 12.08 | 11.98 | 13.38 |
| | | 15 | 18.13 | 15.88 | 16.95 | 16.64 | 18.32 |
| | | 20 | 23.34 | 20.65 | 21.48 | 21.37 | 23.26 |
| Male | WGN | 5 | 7.68 | 5.45 | 7.27 | 2.27 | 7.1 |
| | | 10 | 11.02 | 9.86 | 10.83 | 10.72 | 12.17 |
| | | 15 | 17.38 | 14.57 | 14.93 | 14.54 | 17.41 |
| | | 20 | 22.08 | 19.37 | 20.17 | 17.55 | 22.22 |
| | Babble Noise | 5 | 7.41 | 7.15 | 7.29 | 5.45 | 7.43 |
| | | 10 | 11.34 | 10.63 | 12.34 | 11.94 | 12.21 |
| | | 15 | 17.69 | 15.01 | 16.33 | 16.56 | 17.56 |
| | | 20 | 22.44 | 19.75 | 21.23 | 21.08 | 22.48 |
| | Car Noise | 5 | 8.81 | 5.36 | 8.09 | 7.91 | 8.57 |
| | | 10 | 12.61 | 11.09 | 12.71 | 12.51 | 13.91 |
| | | 15 | 18.76 | 16.61 | 17.58 | 17.17 | 18.75 |
| | | 20 | 23.97 | 21.28 | 22.11 | 21.7 | 23.69 |

ตารางที่ 4.5 Log Spectral Distance (LSD)

| Speech Type | Noise Type | Input SNR | LSD | | | | |
|-------------|-----------------|-----------|------|--------------------|---------------------|----------------------------|---------------------------------|
| | | | MCSS | DAP ($p = 4$) | DAP ($p = 20$) | DAP Subband Proposed | DAP+MCSS Subband Proposed |
| Female | WGN | 5 | 1.99 | 1.42 | 1.16 | 1.16 | 1.14 |
| | | 10 | 1.87 | 1.36 | 1.14 | 1.15 | 1.11 |
| | | 15 | 1.06 | 1.30 | 1.12 | 1.14 | 1.08 |
| | | 20 | 0.85 | 1.24 | 1.1 | 1.13 | 1.05 |
| | Babble Noise | 5 | 2.89 | 1.52 | 1.13 | 1.17 | 1.2 |
| | | 10 | 2.07 | 1.35 | 1.12 | 1.16 | 1.16 |
| | | 15 | 1.98 | 1.18 | 1.11 | 1.15 | 1.12 |
| | | 20 | 0.47 | 1.00 | 1.09 | 1.13 | 1.07 |
| | Car Noise | 5 | 0.98 | 1.3 | 1.11 | 1.15 | 1.12 |
| | | 10 | 0.86 | 1.11 | 0.87 | 0.93 | 0.94 |
| | | 15 | 0.28 | 1.04 | 0.63 | 0.71 | 0.76 |
| | | 20 | 0.30 | 0.76 | 0.18 | 0.28 | 0.37 |
| Male | WGN | 5 | 2.05 | 2.11 | 2.13 | 2.14 | 2.17 |
| | | 10 | 1.93 | 1.99 | 1.88 | 1.89 | 1.92 |
| | | 15 | 1.72 | 1.78 | 1.80 | 1.81 | 1.71 |
| | | 20 | 0.87 | 1.78 | 1.08 | 1.21 | 1.26 |
| | Babble Noise | 5 | 2.32 | 2.38 | 2.4 | 2.41 | 2.44 |
| | | 10 | 2.2 | 2.26 | 2.15 | 2.16 | 2.19 |
| | | 15 | 1.99 | 2.05 | 2.07 | 2.08 | 1.98 |
| | | 20 | 1.14 | 0.81 | 0.97 | 0.95 | 0.93 |
| | Car Noise | 5 | 2.41 | 2.34 | 2.36 | 2.37 | 2.4 |
| | | 10 | 2.29 | 2.35 | 2.24 | 2.25 | 2.28 |
| | | 15 | 2.08 | 2.01 | 2.03 | 2.04 | 1.94 |
| | | 20 | 1.1 | 0.77 | 0.93 | 0.91 | 0.89 |

ตารางที่ 4.6 Mean opines score test (MOS), Female speech

| Signals | | MOS values |
|--|---------------------------|------------|
| Clean speech signal | | 4.67 |
| Noisy speech signal (WGN, SNR=10dB) | | 1.65 |
| WGN, SNR=10dB | MCSS | 2.2 |
| | MCSS + HR | 2.35 |
| | DAP ($p = 4$) | 3.5 |
| | DAP ($p = 20$) | 3.5 |
| | DAP Subband Proposed | 3.65 |
| | MCSS+DAP Subband Proposed | 3.85 |
| Noisy speech signal (Babble noise, SNR=10dB) | | 3.00 |
| Babble noise, SNR=10dB | MCSS | 2.35 |
| | MCSS + HR | 2.9 |
| | DAP ($p = 4$) | 3.4 |
| | DAP ($p = 20$) | 3.00 |
| | DAP Subband Proposed | 2.5 |
| | MCSS+DAP Subband Proposed | 3.5 |
| Noisy speech signal (Car noise, SNR=10dB) | | 3.40 |
| Car noise, SNR=10dB | MCSS | 3.50 |
| | MCSS + HR | 4.20 |
| | DAP ($p = 4$) | 2.80 |
| | DAP ($p = 20$) | 2.60 |
| | DAP Subband Proposed | 3.50 |
| | MCSS+DAP Subband Proposed | 3.55 |

ตารางที่ 4.7 Mean opines score test (MOS), Male speech

| Signals | | MOS values |
|--|---------------------------|------------|
| Clean speech signal | | 4.05 |
| Noisy speech signal (WGN, SNR=10dB) | | 1.60 |
| WGN, SNR=10dB | MCSS | 2.25 |
| | MCSS + HR | 2.45 |
| | DAP ($p = 4$) | 3.45 |
| | DAP ($p = 20$) | 3.45 |
| | DAP Subband Proposed | 3.55 |
| | MCSS+DAP Subband Proposed | 3.50 |
| Noisy speech signal (Babble noise, SNR=10dB) | | 3.05 |
| Babble noise, SNR=10dB | MCSS | 2.30 |
| | MCSS + HR | 2.85 |
| | DAP ($p = 4$) | 2.45 |
| | DAP ($p = 20$) | 2.90 |
| | DAP Subband Proposed | 2.70 |
| | MCSS+DAP Subband Proposed | 3.00 |
| Noisy speech signal (Car noise, SNR=10dB) | | 3.65 |
| Car noise, SNR=10dB | MCSS | 3.40 |
| | MCSS + HR | 3.60 |
| | DAP ($p = 4$) | 2.70 |
| | DAP ($p = 20$) | 2.85 |
| | DAP Subband Proposed | 3.05 |
| | MCSS+DAP Subband Proposed | 3.40 |

บทที่ 5

สรุปผลการวิจัย และข้อเสนอแนะ

5.1 สรุปผลการวิจัย

วิทยานิพนธ์นี้เสนอเทคนิคการลดสัญญาณรบกวนแบบใช้ไมโครโฟน 2 ตัวที่ใช้ในอุปกรณ์แฮนด์ฟรี โดยพิจารณาที่เสียงรบกวน 3 แบบคือ เสียงรบกวนขาวแบบเกาส์ (WGN), เสียงพูดคุย (Babble Noise) และ เสียงรบกวนในรถยนต์ (Car Noise) ซึ่งเป็นสัญญาณรบกวนที่เกิดขึ้นกับการสื่อสารระบบแฮนด์ฟรี หลักการลดสัญญาณรบกวนแบ่งตามการทำงานได้ 2 แบบคือ การลดสัญญาณรบกวนที่ทำงานในโดเมนความถี่ และ การลดสัญญาณรบกวนที่ทำงานในโดเมนเวลา ซึ่งการลดสัญญาณรบกวนที่ทำงานในโดเมนความถี่นั้นมีข้อดีคือมีความซับซ้อนต่ำแต่สัญญาณเสียงพูดมีความเพี้ยน เนื่องจากองค์ประกอบทางความถี่ของสัญญาณที่ได้ในช่วงความถี่สูงหายไป สำหรับการลดสัญญาณรบกวนที่ทำงานในโดเมนเวลานั้น มีข้อดีคือไม่เกิดความผิดเพี้ยนขึ้นกับค่าประมาณสัญญาณเสียงพูด สามารถรักษารายละเอียดของเสียงพูดไว้ได้ แต่มีความซับซ้อนสูงกว่า

วิทยานิพนธ์นี้เสนอการลดสัญญาณรบกวนโดยทำการวิเคราะห์ทั้งในโดเมนความถี่และโดเมนเวลา ได้แก่ เทคนิคการเพิ่มฮาร์โมนิกของเสียงพูด, ขั้นตอนวิธี DAP แบบหลายอันดับ และเทคนิคร่วมระหว่างขั้นตอนวิธี DAP และ MCSS ที่เสนอ

การลดสัญญาณรบกวนที่ทำงานในโดเมนความถี่ที่เป็นที่นิยมใช้เรียกว่า เทคนิคการลบสเปกตรัม (SS) วิธีนี้มีความซับซ้อนต่ำ สามารถลดสัญญาณรบกวนได้ แต่ปัญหาที่เกิดขึ้นคือเกิดสัญญาณรบกวนอื่นขึ้นเรียกว่า Musical Noise ซึ่งเกิดจากการประมาณสเปกตรัมของเสียงรบกวนที่ไม่แม่นยำ สำหรับงานวิจัยที่ศึกษาในงานวิทยานิพนธ์นี้เรียกว่า วิธีการลบสเปกตรัมข้ามตัดแปร (Modified Cross-Spectral Subtraction Method, MCSS) วิธีนี้รับสัญญาณด้วยไมโครโฟน 2 ตัว และนำสัญญาณทั้งสองมาหาความสัมพันธ์ด้วยฟังก์ชันโคไซน์ พบว่าตรงช่วงความถี่ต่ำมีสัญญาณเสียงพูดอยู่อย่างหนาแน่นและในขณะที่ช่วงความถี่สูงกว่า 210Hz[9] จะมีสัญญาณเสียงรบกวนอยู่อย่างหนาแน่น ดังนั้น จึงใช้คุณลักษณะนี้เพื่อแยกช่วงที่เป็นสัญญาณเสียงพูดออกจากสัญญาณรบกวนโดยไม่ต้องใช้วิธี VAD วิธีนี้สามารถแก้ปัญหของ Musical Noise โดยการทำการประมาณสเปกตรัมของเสียงรบกวนเกิน (Overestimate) โดยใช้ $\alpha = 8$ แต่ปัญหาที่เกิดขึ้นกับวิธีนี้

คือองค์ประกอบทางความถี่ของเสียงพูดที่ความถี่สูงกว่า 1.5 kHz หายไป ทำให้เกิดทำความผิดเพี้ยนของสัญญาณ

เนื่องจากการลดเสียงรบกวนจะมีผลกระทบต่อคุณภาพของเสียงพูด ดังนั้นฟังก์ชันไม่เชิงเส้นจึงถูกเสนอเพื่อสร้างทดฮาร์โมนิกของเสียงพูด โดยพบว่าการใช้งานร่วมกันระหว่าง MCSS และ HR ที่เสนอมีประสิทธิภาพไม่ด้อยไปกว่าวิธี HR ใน [10] โดยผลการจำลองระบบผ่านทางค่า LSD, สเปกโตรแกรม และค่า MOS แสดงให้เห็นว่าวิธีการสร้างทดแทนฮาร์โมนิกที่นำเสนอจะให้คุณภาพเสียงที่ดีกว่าวิธีเดิมใน [10] ดังนั้นวิธีที่นำเสนอสามารถสร้างทดแทนฮาร์โมนิกของสัญญาณเสียงพูดที่มีความใกล้เคียงกับสัญญาณเสียงพูดต้นฉบับ

การลดสัญญาณรบกวนที่ทำงานในโดเมนเวลาแบบใช้ไมโครโฟน 2 ตัวที่ศึกษาในวิทยานิพนธ์นี้เรียกว่า วิธีดับเบิลแอฟฟายโปรเจกชัน (DAP) วิธีนี้เป็นการเพิ่มประสิทธิภาพของวิธี NLMS เพื่อให้มีประสิทธิภาพใกล้เคียงกับวิธี RLS โดยวิธี DAP มีตัวแปร p หรือ อันดับโปรเจกชัน โดยที่ $1 \leq p \leq L$ เมื่อ $p=1$ คือวงจรกรองแบบ NLMS และเมื่อมีการปรับค่า p มากกว่า 1 เป็นการเพิ่มประสิทธิภาพของตัวกรอง โดยที่อันดับโปรเจกชันสูงสามารถลดสัญญาณรบกวนได้ดีกว่าเมื่อเทียบกับอันดับโปรเจกชันต่ำ แต่อย่างไรก็ตามวิธี DAP ที่มีที่อันดับโปรเจกชันสูงนั้นทำให้ระบบมีความซับซ้อนสูง

วิธีที่เสนอในวิทยานิพนธ์จะนำข้อดีของวิธี DAP ที่ขึ้นกับอันดับการโปรเจกชันคือ วิธีที่เสนอสามารถลดสัญญาณรบกวนได้และมีความซับซ้อนที่ต่ำ โดยจะใช้ร่วมกับวิธีการแบ่งแถบความถี่ โดยผลการจำลองระบบผ่านทางตัวชี้วัด โดยที่การลดสัญญาณรบกวนดูจากค่า SegSNR และความเพี้ยนของสัญญาณเสียงพูดดูจากค่า LSD นอกจากนี้ยังมีตัวชี้วัดที่พิจารณาจากผู้ฟังที่เรียกว่าค่า MOS แสดงให้เห็นว่าวิธีที่เสนอสามารถลดสัญญาณรบกวนในขณะเดียวกันยังสามารถรักษาองค์ประกอบความถี่ของเสียงพูดที่ความถี่สูงกว่า 1.5 kHz ไว้ได้

วิธีที่เสนอในวิทยานิพนธ์อีกวิธีหนึ่งคือ เทคนิคร่วมระหว่างขั้นตอนวิธี DAP และ MCSS ที่เสนอ กล่าวคือในช่วงความถี่ต่ำใช้วิธี MCSS และ ในช่วงความถี่สูงใช้วิธี DAP ที่มีอันดับโปรเจกชัน $p=4$ และวิธีที่เสนอใช้ตัวแปรควบคุมการลบสเปกตรัม $\alpha = 4$ ซึ่งแตกต่างจากที่ใช้วิธี MCSS อย่างเดียวคือ $\alpha = 8$ โดยวิธีที่เสนอจะนำข้อดีของทั้ง 2 วิธีมารวมกัน ซึ่งดูได้จากตัวชี้วัดประสิทธิภาพ SegSNR, LSD และ MOS แสดงให้เห็นว่า วิธีที่เสนอสามารถลดสัญญาณรบกวนในขณะเดียวกันยังสามารถรักษาองค์ประกอบความถี่ของเสียงพูดที่ความถี่สูงกว่า 1.5 kHz ไว้ได้ อีกทั้งมีความซับซ้อนที่ต่ำ

5.2 ข้อเสนอแนะ

แนวทางการพัฒนาที่สามารถทำได้แบ่งออกเป็น 2 แบบคือ

1. จำนวนไมโครโฟน

ในวิทยานิพนธ์นี้มุ่งเน้นไปที่การใช้งานด้วยไมโครโฟน 2 ตัว เนื่องจากต้นทุนในการเพิ่มไมโครโฟนในอุปกรณ์นั้นมีค่าต่ำสุดเมื่อเปรียบเทียบกับการใช้งานไมโครโฟนหลายตัว และการเพิ่มประสิทธิภาพการลดสัญญาณรบกวนที่สามารถยอมรับได้ ในส่วนของการใช้งานด้วยไมโครโฟนหลายตัวนั้นสามารถประมาณเสียงรบกวนได้ดีกว่า เนื่องจากข้อมูลเชิงพื้นที่ (Spatial information) ถูกนำมาใช้เพิ่มความแม่นยำในการประมาณเสียงรบกวน[9] จึงเป็นเทคนิคที่น่าสนใจในการพัฒนาเทคนิคการลดสัญญาณรบกวนต่อไป

2. การแบ่งช่วงความถี่ย่อย

ในวิทยานิพนธ์นี้มุ่งเน้นไปที่การแบ่งช่วงความถี่ออกเป็น 2 ช่วงความถี่ย่อย ทั้งนี้เทคนิคการแบ่งช่วงความถี่ย่อยยังสามารถพัฒนา โดยสามารถแบ่งได้ออกเป็นหลายช่วงความถี่ ตามลักษณะการใช้งาน ซึ่งเทคนิคการทำงานจะมีความเหมาะสมในแต่ละช่วงความถี่แตกต่างกันไป ในด้านของประสิทธิภาพการทำงาน การลดความซับซ้อน เป็นต้น

รายการอ้างอิง

- [1] An Act of Legislation Traffic. Contents Land Act B.E. 2552(Online). 2010.
<http://www.thailandlawyercenter.com>[March. 29 2012]
- [2] Vaseghi, S. V. Spectral Subtraction, Advanced Digital Signal Processing and Noise Reduction.
Second Edition. Chapter 11. 333 - 341.
- [3] Thoonsaengngam, R. Acoustic Echo and Noise Reduction Techniques for Hand-Free
Communication System. Degree of Master of Engineering in Electrical Engineering,
Chulalongkorn University, 2007.
- [4] Boll, S. F. Suppression of Acoustic Noise in Speech Using Spectral Subtraction. IEEE
Transactions on Acoustics, Speech, and Signal Processing 2 (1979) : 113 - 120.
- [5] Berouti, R. Schwartz and Makhoul, J. Enhancement of Speech Corrupted by Acoustic Noise.
IEEE International Conference on ICASSP '79, Acoustics, Speech, and Signal Processing.
4 (1979) : 208 - 211.
- [6] McAulay, R. J. & Malpass, M. L. Speech Enhancement Using a Soft-Decision Noise
Suppression Filter. IEEE Transactions on Acoustics, Speech, and Signal Processing. 2
(1980) : 137-145.
- [7] Ephraim, Y. and Malah, D. Speech Enhancement Using a Minimum Mean-Square Error Short-
Time Spectral Amplitude Estimator. IEEE Transactions on Acoustics, Speech, and Signal
Processing. 6 (1984) : 1109 - 1121.
- [8] Bouquin-Jeannes, R. L., Azirani, A. A. and Faucon, G. Enhancement of speech degraded by
coherent and incoherence noise using a cross-spectral estimator. IEEE Transactions on
Acoustics, Speech, and Signal Processing. 5 (1997) : 484 - 487.
- [9] Guerin, A. Le Bouquin-Jeannes, R. and Faucon, G. A Two-Sensor Noise Reduction System:
Applications for Hands-Free Car Kit. EURASIP Journal on Applied Signal Processing.
(2003) : 1125 - 1134.
- [10] Plapous, C. Marro, C. and Scalart, P. Speech Enhancement Using Harmonic Regeneration.
IEEE International Conference. 1 (2005) : 157 - 160.

- [11] Shen, T. Lun, D. P. and Hsung, T. Speech Enhancement Using Harmonic Regeneration with Improved Wevelet Based A-Priori Signal to Noise Ratio Estimator. International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS). (2010) : 1 - 4.
- [12] Xuchu, H. and Xiaojing, Z. Speech Enhancement Using Harmonic Regeneration. IEEE International Conference. 1 (2011) : 150 - 152.
- [13] Benesty, J. Makino, S. and Chen, J. Speech Enhancement. Netherlands: Springer Berlin Heidelberg. (2005).
- [14] Tanyer, S. G. and Ozer, H. Voice Activity Detection in Nonstationary Noise. IEEE transactions on Speech, audio processing. (2000) : 478 - 482.
- [15] Griffiths, L. J. and Jim, C. An Alternative Approach to Linearly Constrained Adaptive Beamforming. IEEE Transactions on Acoustics, Speech, and Signal Processing. 1 (1982) : 27 - 34.
- [16] Meyer, J. and Simmer, K. U. Multi-channel speech enhancement in car environment using wiener filter and spectral subtraction. IEEE International Conference. 2 (1987) : 1167 - 1170.
- [17] Haykin, S. Adaptive Filter Theory. Fourth Edition. United States of America : Prentice-Hall, 2002.
- [18] Gabrea, M. Two Microphone Speech Enhancement System Based on Double Affine Projection Algorithm. IEEE International Conference on Acoustics, Speech, and Signal Processing. 2 (2003) : 544 - 547.
- [19] International Telecommunication Union. Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm (Online). ITU-T Recommendation P.835. (2003).
- [20] International Telecommunication Union. Method for Subjective determination of Transmission Quality (Online). ITU-T Recommendation P.800. (1996).
- [21] Fastl, H. and Zwicker, E. Psychoacoustics : Facts and models, Third Edition. Springer-Verlag. (2007).

ภาคผนวก

ภาคผนวก ก

ภาคผนวก ก

ดัชนีคำศัพท์

| คำศัพท์ | ความหมาย |
|-------------------------------------|---|
| Adaptive Noise Cancellation | การกำจัดสัญญาณรบกวนโดยใช้ตัวกรองปรับตัว |
| Affine Projection | ขั้นตอนวิธีแอฟฟายโปรเจกชัน |
| Babble Noise | สัญญาณเสียงพูดคุย |
| Background Noise | สัญญาณรบกวนพื้นหลัง |
| Car Noise | สัญญาณรบกวนในรถยนต์ |
| Coherence Function | ฟังก์ชันสหพันธ์ |
| Convergent Rate | อัตราการลู่เข้า |
| Convolution | การคอนโวลูชัน |
| Cross Power Spectral Density | ความหนาแน่นสเปกตรัมกำลังแบบข้าม |
| Cross-Spectral Subtraction | การลบสเปกตรัมแบบข้าม |
| Cross-Spectral Subtraction Method | วิธีการลบสเปกตรัมข้าม |
| Distortion | ความเพี้ยน |
| Double Affine Projection Algorithm | ขั้นตอนวิธีดับเบิลแอฟฟายโปรเจกชัน |
| Echo | สัญญาณเสียงสะท้อน |
| First Order Digital Low-pass Filter | ตัวกรองแถบความถี่ต่ำแบบดิจิทัลอันดับหนึ่ง |
| Frame Index | ดัชนีบ่งบอกเฟรม |
| Frequency-Bin Index | องค์ประกอบทางความถี่ |
| Full Noise Subtraction | การลบสัญญาณรบกวนแบบเต็ม |
| Hamming Window | หน้าต่างแฮมมิง |
| Hands-Free Telephony | การใช้โทรศัพท์เคลื่อนที่ในรถยนต์ด้วยอุปกรณ์แฮนด์ฟรี |
| Hanning Window | หน้าต่างแฮนนิ่ง |
| Harmonic Regeneration | วิธีการสร้างทดแทนฮาร์โมนิกของสัญญาณเสียง |

| คำศัพท์ | ความหมาย |
|---|--|
| Hearing Aids | เครื่องช่วยฟัง |
| High-Pass Filter | ตัวกรองความถี่สูง |
| Impulse Responses | ผลตอบสนองอิมพัลส์ |
| Inverse Short-Time Discrete Fourier Transform | การแปลงกลับของฟูรีเยร์แบบไม่ต่อเนื่องผกผันในช่วงเวลาสั้น |
| Kalman Gain Vector | เวกเตอร์อัตราขยายกาลมาน |
| Least Mean Square | ขั้นตอนวิธีกำลังสองเฉลี่ยน้อยที่สุด |
| Log-Spectral Distance | ค่าระยะสเปกตรัมลอการิทึม |
| Low-Pass Filter | ตัวกรองความถี่ต่ำ |
| Magnitude Spectral Subtraction | การลบสเปกตรัมทางขนาด |
| Mean Opinion Score | คะแนนความคิดเห็นเฉลี่ย |
| Modified Cross-Spectral Subtraction Method | วิธีการลบสเปกตรัมข้ามตัดแปร |
| Multi-Microphone Technique | เทคนิคที่ใช้ไมโครโฟนหลายตัว |
| Noise Overestimation | การประมาณสัญญาณรบกวนเกิน |
| Noise Spectrum | สเปกตรัมของสัญญาณรบกวน |
| Noisy | สัญญาณเสียงพูดที่ปนด้วยสัญญาณเสียงรบกวน |
| Noisy Spectrum | สเปกตรัมของสัญญาณเสียงพูดที่ปนด้วยสัญญาณรบกวน |
| Non-Linearity | วิธีความไม่เป็นเชิงเส้น |
| Non-Speech Activity Period | ช่วงที่ไม่ใช่สัญญาณเสียงพูด |
| Non-Stationary Noise | สัญญาณรบกวนแบบไม่คงที่ |
| Normalize Least Mean Square | ขั้นตอนวิธีกำลังสองเฉลี่ยน้อยที่สุดแบบนอร์มาไลซ์ |
| Original Signal | สัญญาณเสียงพูดต้นฉบับ |
| Over-Subtraction | การลบสัญญาณรบกวนเกิน |
| Post Subtraction | หลังการลบ |
| Power Spectral Density | ความหนาแน่นสเปกตรัมกำลัง |

คำศัพท์

Power Spectral Subtraction
 Recursive Least Square
 Regularization Matrix
 Segmental Signal-to-Noise Ratio
 Short-Time Discrete Fourier Transform
 Signal-to-Noise Ratio
 Single-Microphone
 Spatial Information
 Spectral Floor Parameter
 Spectral Subtraction Method
 Spectrogram
 Speech
 Speech Communication
 Speech-Activity Period
 Stationary Noise
 Step-Size Parameter
 Subtraction Parameter
 Synthesis window
 Teleconference System
 Voice Activity Detector
 White Gaussian Noise
 Wiener Filter
 Window

ความหมาย

เทคนิคการลบสเปกตรัมทางกำลัง
 ขั้นตอนวิธีกำลังสองน้อยที่สุดแบบวนซ้ำ
 เมตริกซ์เรกูลาไรเซชัน
 ค่าอัตราส่วนระหว่างสัญญาณต่อสัญญาณรบกวน
 ความผิดพลาดแยกส่วน
 การแปลงฟูริเยร์แบบไม่ต่อเนื่องในช่วงเวลาสั้น
 อัตราส่วนสัญญาณเสียงพูดต่อสัญญาณรบกวน
 เทคนิคที่ใช้ไมโครโฟนหนึ่งตัว
 ข้อมูลเชิงพื้นที่
 ตัวแปรพื้นของสเปกตรัม
 วิธีการลบทางสเปกตรัม
 สเปกโตรแกรม
 สัญญาณเสียงพูด
 การสื่อสารทางเสียงพูด
 ช่วงที่เป็นสัญญาณเสียงพูด
 สัญญาณรบกวนแบบคงที่
 พารามิเตอร์ค่าช่วงก้าว
 ตัวแปรการลบสเปกตรัม
 หน้าต่างสังเคราะห์
 ระบบการสัมมนาทางไกล
 ตัวตรวจจับสัญญาณเสียง
 สัญญาณรบกวนขาวแบบเกาส์
 วิธีการกรองแบบวีเนอร์
 หน้าต่าง

ภาคผนวก ข

ภาคผนวก ข

บทความ และ การนำเสนอผลงาน

1. H. Thumchirdchupong and N. Tangsangiumvisai, A noise reduction technique with speech harmonic regeneration for hands-free telephony. Accepted for National Conference on Electrical Engineering (EECON-35). (December 2012) : 849 - 852.
2. H. Thumchirdchupong and N. Tangsangiumvisai, A two-microphone noise reduction scheme for hands-free telephony in a car environment. Summited to Electrical Engineering / Electronics, Computer, Telecommunications and Information Technology (ECTI) Association. (January 2013).

ประวัติผู้เขียนวิทยานิพนธ์

นางสาวหทัยชนก ธรรมเชิดชูพงษ์ เกิดวันที่ 6 มีนาคม พ.ศ. 2532 ที่จังหวัด กรุงเทพมหานคร เข้าศึกษาในหลักสูตรวิศวกรรมศาสตรบัณฑิต คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ ในปีการศึกษา 2550 สำเร็จการศึกษา วิศวกรรมศาสตรบัณฑิต สาขาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ ในปีการศึกษา 2553 เข้าศึกษา ต่อในหลักสูตรวิศวกรรมศาสตรมหาบัณฑิต สาขาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2554