

วิธีการดีคอรี่ไลซ์ของสัญญาณสเตริโอสำหรับการกำจัดสัญญาณเสียงสะท้อนแบบสเตริโอ  
โดยการเพิ่มสัญญาณรบกวนแบบปรับตัวได้



นางสาว พิมพกา สุรินทร์

สถาบันวิทยบริการ

จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต

สาขาวิชาวิศวกรรมไฟฟ้า ภาควิชาวิศวกรรมไฟฟ้า


คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2547

ISBN 947-17-6422-7

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

A SIGNAL DECORRELATION TECHNIQUE FOR STEREOPHONIC ACOUSTIC ECHO  
CANCELLATION USING ADAPTIVE NOISE ADDITION



Miss Pimphaka Surin

สถาบันวิทยบริการ

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

Department of Electrical Engineering

Faculty of Engineering

Chulalongkorn University

Academic Year 2004

ISBN 974-17-6422-7



พิมพ์กา สุรินทร์ : วิธีการตัดคอร์เลชันของสัญญาณสเตอริโอสำหรับการกำจัดสัญญาณเสียงสะท้อนแบบสเตอริโอโดยการเพิ่มสัญญาณรบกวนแบบปรับตัวได้ (A SIGNAL DECORRELATION TECHNIQUE FOR STEREOPHONIC ACOUSTIC ECHO CANCELLATION USING ADAPTIVE NOISE ADDITION) อ.ที่ปรึกษา : อ.ดร.นิศาชล ตั้งเสียมวิสัย, อ.ที่ปรึกษาร่วม : ผศ.ดร.สุภาวดี อร่ามวิทย์, 91 หน้า. ISBN 974-17-6422-7

ในวิทยานิพนธ์นี้ได้นำเสนอวิธีการตัดคอร์เลชันสัญญาณสเตอริโอโดยการเพิ่มสัญญาณรบกวนแบบปรับตัวได้ ซึ่งมีการปรับเปลี่ยนค่าไปตามลักษณะของสัญญาณเสียงพูดขาเข้า เพื่อให้สามารถลดสหสัมพันธ์ของสัญญาณเสียงสเตอริโอลงได้ โดยไม่ส่งผลกระทบต่อคุณภาพสัญญาณเสียงของสัญญาณที่ถูกตัดคอร์เลชัน ผลจากการตัดคอร์เลชันสัญญาณสเตอริโอทำให้ระบบ SAEC สามารถประมาณค่าวิถีสะท้อนทางเสียงได้ใกล้เคียงมากขึ้น ดังผลการทดสอบของระบบ SAEC ทั้งในด้านของผลที่เป็นตัวเลขชี้วัดและผลการทดสอบฟังคุณภาพเสียง ทั้งนี้วิธีตัดคอร์เลชันที่นำเสนอมีความซับซ้อนในการคำนวณต่ำ นอกจากนี้วิทยานิพนธ์นี้ยังศึกษาถึงความเป็นไปได้ในการนำระบบ SAEC และวิธีตัดคอร์เลชันสัญญาณสเตอริโอที่นำเสนอ ไปประยุกต์ใช้กับระบบการสื่อสารที่มีการเข้ารหัสและถอดรหัสสัญญาณเสียงด้วยมาตรฐาน G723.1 ผลการจำลองระบบดังกล่าวบนคอมพิวเตอร์พบว่า ระบบ SAEC สามารถทำงานได้ดีเช่นเดียวกับเมื่อไม่มีการเข้ารหัสและถอดรหัสสัญญาณเสียง

สถาบันวิทยบริการ  
จุฬาลงกรณ์มหาวิทยาลัย

ภาควิชา.....วิศวกรรมไฟฟ้า..... ลายมือชื่อนิสิต.....  
สาขาวิชา.....วิศวกรรมไฟฟ้า..... ลายมือชื่ออาจารย์ที่ปรึกษา.....  
ปีการศึกษา.....2547..... ลายมือชื่ออาจารย์ที่ปรึกษาร่วม .....

## 4570730521 : MAJOR ELECTRICAL ENGINEERING

KEY WORD : ADAPTIVE FILTER/STEREOPHONIC ACOUSTIC ECHO CANCELLATION/

DECORRELATION/NON -UNIQUENESS PROMBLEM/MIS ALIGNMENT PROBLEM

PIMPHAKA SURIN : A SIGNAL DECORRELATION TECHNIQUE FOR STEREOPHONIC ACOUSTIC ECHO CANCELLATION USING ADAPTIVE NOISE ADDITION. THESIS ADVISOR : NISACHON TANGSANGIUMVISAI, Ph.D., THESIS CO-ADVISOR : ASSIST. PROF. SUPAVADEE ARAMVIT, Ph.D., 91 pp. ISBN 974-17-6422-7

This thesis presents a new signal decorrelation technique, based upon adaptive noise addition. The added noise level is adjusted according to the characteristics of the speech inputs so that the quality of the decorrelated stereo speech signals is kept undisturbed. Therefore, the performance of an SAEC system can be improved by employing the adaptive noise addition technique. Test results on recorded speech signals indicate improved performance in both objective and subjective manners. In addition, the computational complexity for the proposed technique is insignificant as compared to the whole system. Furthermore, this SAEC system with the proposed decorrelation technique is applied to a communication system with G723.1 speech codec. Simulation results demonstrate well-behaved performance of the SAEC system as well as in the case without speech codec.

สถาบันวิทยบริการ  
จุฬาลงกรณ์มหาวิทยาลัย

Department Electrical Engineering

Field of study Electrical Engineering

Academic year 2004

Student's signature.....

Advisor's signature.....

Co Advisor's signature.....

## กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปด้วยดี ดิฉันใคร่ขอกราบขอบพระคุณอย่างสูงสำหรับความช่วยเหลืออย่างดียิ่งของ อาจารย์ ดร.นิศาชล ตั้งเสงี่ยมวิสัย อาจารย์ที่ปรึกษาวิทยานิพนธ์ และ ผู้ช่วยศาสตราจารย์ ดร.สภาวดี อร่ามวิทย์ อาจารย์ที่ปรึกษาร่วม ซึ่งท่านได้ให้คำแนะนำและข้อคิดเห็นต่าง ๆ พร้อมทั้งแรงกระตุ้นและแรงบันดาลใจในการทำวิจัยมาด้วยดีตลอดมา

ขอขอบคุณอาจารย์ ทุก ๆ ท่านที่ได้ให้ความรู้ในการศึกษาในการนำไปใช้เกี่ยวกับงานวิจัยนี้

ขอขอบคุณ รุ่นพี่ รุ่นน้อง เพื่อน ๆ และคนรอบตัวผู้วิจัยทุก ๆ คน ในห้องปฏิบัติการวิจัย การประมวลผลสัญญาณดิจิทัลทุก ๆ คน สำหรับความช่วยเหลือและกำลังใจในการทำวิจัยตลอดมา ขอขอบคุณที่ทำให้วันและเวลาของผู้วิจัยผ่านไปอย่างมีความหมาย

สุดท้ายขอกราบขอบพระคุณบิดามารดาและครอบครัว ตลอดจนญาติ ๆ ทุกคนที่เป็นกำลังใจและให้การสนับสนุนแก่ผู้วิจัยมาโดยตลอดจนสำเร็จการศึกษา



สถาบันวิทยบริการ  
จุฬาลงกรณ์มหาวิทยาลัย

# สารบัญ

หน้า

บทคัดย่อภาษาไทย .....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ .....	ช
สารบัญตาราง .....	ญ
สารบัญภาพ.....	ฎ
บัญชีคำศัพท์.....	ฏ
บทที่	
1. บทนำ.....	1
1.1 แนวคิดและเหตุผล.....	1
1.2 วัตถุประสงค์.....	4
1.3 เป้าหมายและขอบเขตของงานวิจัย .....	4
1.4 ขั้นตอนและวิธีการดำเนินการ .....	5
1.5 ประโยชน์ที่คาดว่าจะได้รับ .....	5
1.6 ภาพรวมของวิทยานิพนธ์.....	5
2. หลักการและขั้นตอนวิธีที่เกี่ยวข้อง.....	7
2.1 ระบบการกำจัดสัญญาณเสียงสะท้อน .....	7
2.1.1 ปัญหาสัญญาณเสียงสะท้อนในระบบการสื่อสารทางเสียง .....	7
2.1.2 การแสดงเอกลักษณ์ของระบบ (System Identification) ของวงจรกรองแบบ ปรับตัวสำหรับระบบ AEC.....	8
2.2 ขั้นตอนวิธีที่ใช้กับวงจรกรองแบบปรับตัว .....	12
2.2.1 ขั้นตอนวิธี Least Mean Square (LMS) .....	12
2.2.2 ขั้นตอนวิธี Normalized Least Mean Square (NLMS) .....	16
2.3 การแสดงเอกลักษณ์ของระบบ ของวงจรกรองแบบปรับตัว สำหรับระบบ SAEC .....	19
2.4 ปัญหา “Non-uniqueness Problem” ในระบบ SAEC.....	21
2.5 ลักษณะพื้นฐานของสัญญาณเสียงพูด .....	23

3. การเพิ่มประสิทธิภาพในระบบการกำจัดสัญญาณเสียงสะท้อนแบบสเตอริโอ.....	27
3.1 วิธีการดีคอดรีเลชันสัญญาณเสียงสเตอริโอ. ....	27
3.1.1 การดีคอดรีเลชันสัญญาณเสียงสเตอริโอโดยการเพิ่มสัญญาณรบกวนสุ่ม.....	28
3.1.2 การดีคอดรีเลชันสัญญาณเสียงสเตอริโอโดยการเพิ่มภาวะไม่เชิงเส้น.....	29
3.1.3 การดีคอดรีเลชันสัญญาณเสียงสเตอริโอโดยการเพิ่มสัญญาณรบกวนแบบปรับ ตัวได้.....	31
3.1.3.1 วิธีการดีคอดรีเลชันสัญญาณโดยใช้สัญญาณรบกวน แบบปรับตัวได้.....	31
3.1.3.2 ฟังก์ชันสัญญาณรบกวนแบบปรับตัวได้ .....	33
3.2 การวัดประสิทธิภาพการทำงานของระบบ .....	35
3.2.1 สมรรถนะของวงจรรองแบบปรับตัวในโดเมนเวลา.....	35
3.2.2 ฟังก์ชันร่วมนัย (Coherence Function).....	36
3.2.3 การทดสอบคุณภาพเสียงทางกายภาพ (Subjective Listening Test).....	37
3.3 การเข้ารหัสสัญญาณเสียงพูด.....	38
3.3.1 วิธีการเข้ารหัสสัญญาณเสียงพูด.....	39
3.3.1.1 Waveform Coder.....	39
3.3.1.2 Source Coder (Vocoder).....	40
3.3.1.3 Hybrid Coder.....	41
3.3.2 การเข้ารหัสเสียงพูดตามมาตรฐาน G.723.1.....	44
4. ผลการทดลองและการวิเคราะห์ผล.....	47
4.1 การจำลองการทำงานของระบบ SAEC.....	47
4.2 การจำลองการทำงานของระบบ SAEC เมื่อมีการดีคอดรีเลชันสัญญาณเสียงสเตอริโอ โดยเพิ่มสัญญาณรบกวนสุ่ม .....	50
4.3 การจำลองการทำงานของระบบ SAEC เมื่อดีคอดรีเลชันสัญญาณเสียงสเตอริโอ โดยเพิ่มภาวะไม่เชิงเส้น.....	54
4.4 การจำลองการทำงานของระบบ SAEC เมื่อมีการดีคอดรีเลชันสัญญาณเสียงสเตอริโอ โดย เพิ่มสัญญาณรบกวนแบบปรับตัวได้ ซึ่งเป็นวิธีที่น่าเสนอในวิทยานิพนธ์นี้.....	57



4.5 การศึกษาผลกระทบต่อคุณภาพสัญญาณเสียงสเตอริโอ ด้วยการทดสอบ Subjective Listening Test.....	68
4.6 ผลการจำลองการทำงานของระบบ SAEC ที่มีการเข้ารหัสสัญญาณเสียง ตามมาตรฐาน G.723.1 .....	71
4.6.1 ผลการจำลองการทำงานของระบบ SAEC ที่ระบบมีการเข้ารหัสสัญญาณเสียง ตามมาตรฐาน G.723.1 เมื่อไม่ได้มีการดีคอดิเลชันสัญญาณเข้าแบบสเตอริโอ....	71
4.6.2 ผลการจำลองการทำงานของระบบ SAEC ที่มีการเข้ารหัสสัญญาณเสียง ตามมาตรฐาน G.723.1 เมื่อมีการดีคอดิเลชันสัญญาณเข้าแบบสเตอริโอ.....	73
5. สรุปผลการวิจัยและข้อเสนอแนะในอนาคต .....	76
5.1 สรุปผลการวิจัย .....	76
5.2 ข้อเสนอแนะสำหรับงานวิจัยในอนาคต .....	77
รายการอ้างอิง.....	78
ภาคผนวก.....	81
บทความที่ได้รับการเผยแพร่ .....	82
ประวัติผู้เขียนวิทยานิพนธ์ .....	91

## สารบัญตาราง

ตาราง	หน้า
ตารางที่ 2.1 ความซับซ้อนในการคำนวณของขั้นตอนวิธี LMS .....	17
ตารางที่ 2.2 ความซับซ้อนในการคำนวณของขั้นตอนวิธี NLMS.....	18
ตารางที่ 3.1 คุณภาพของสัญญาณเสียงสเตอริโอตามคะแนนต่างๆ ที่ให้กับการทดสอบ MOS Testing .....	37
ตารางที่ 4.1 เปรียบเทียบคุณภาพของสัญญาณเสียงสเตอริโอในแต่ละวิธีการดีคอดิเลชัน สัญญาณตามค่า MOS .....	69



สถาบันวิทยบริการ  
จุฬาลงกรณ์มหาวิทยาลัย

## สารบัญภาพ

ภาพประกอบ	หน้า
รูปที่ 1.1 แบบจำลองการกำจัดสัญญาณเสียงสะท้อนแบบสเตอริโอ.....	2
รูปที่ 2.1 การเกิดสัญญาณเสียงสะท้อนในระบบการสื่อสารทางเสียง.....	8
รูปที่ 2.2 แบบจำลองระบบ AEC โดยใช้วงจรรองแบบปรับตัวจำลอง AEP .....	9
รูปที่ 2.3 การแสดงเอกลักษณ์ของระบบ (System Identification) ของวงจรรองแบบปรับตัวสำหรับระบบ AEC.....	10
รูปที่ 2.4 แบบจำลองระบบ SAEC โดยใช้วงจรรองแบบปรับตัวจำลอง AEP .....	19
รูปที่ 2.5 บล็อกไดอะแกรมของระบบ SAEC.....	21
รูปที่ 2.6 ลักษณะของเสียงและพลังงานของเสียงแบบ Voiced Sound .....	24
รูปที่ 2.7 ลักษณะของเสียงและพลังงานของเสียงแบบ Unvoiced Sound .....	25
รูปที่ 3.1 ระบบ SAEC ที่มีการดีคอดรีเลชันสัญญาณเสียงสเตอริโอโดยใช้ส่วนประมวลผลก่อน.....	27
รูปที่ 3.2 แบบจำลองการดีคอดรีเลชันสัญญาณเสียงสเตอริโอในบล็อก PP ด้วยการเพิ่มสัญญาณรบกวนสุ่ม.....	28
รูปที่ 3.3 แบบจำลองการดีคอดรีเลชันสัญญาณเสียงสเตอริโอในบล็อก PP ด้วยการเพิ่มภาวะไม่เชิงเส้น .....	29
รูปที่ 3.4 การดีคอดรีเลชันสัญญาณเสียงสเตอริโอโดยการเพิ่มสัญญาณรบกวนแบบปรับตัวได้.....	31
รูปที่ 3.5 การตรวจสอบช่วงเวลาของคาบที่มีเสียงพูดและคาบเงียบ .....	32
รูปที่ 3.6 ไดอะแกรมตรวจสอบคาบและเพิ่มสัญญาณรบกวนตามฟังก์ชันที่กำหนด.....	33
รูปที่ 3.7 สัญญาณเสียงสเตอริโอที่ใช้ในการทดสอบคุณภาพ .....	37
รูปที่ 3.8 แบบจำลองระบบ SAEC ที่มีการเข้ารหัสสัญญาณเสียงพูด .....	39
รูปที่ 3.9 แบบจำลองการเข้ารหัสแบบ Source Coder .....	41
รูปที่ 3.10 แบบจำลองส่วนของการเข้ารหัสและถอดรหัสแบบ Hybrid Coder.....	42
รูปที่ 3.11 เปรียบเทียบคุณภาพและจำนวนบิตข้อมูลในการเข้ารหัสเสียงพูดแต่ละแบบ .....	44
รูปที่ 3.12 บล็อกไดอะแกรมมาตรฐานการเข้ารหัส G.723.1 .....	45
รูปที่ 3.13 บล็อกไดอะแกรมการถอดรหัสสัญญาณเสียงตามมาตรฐาน G.723.1.....	46
รูปที่ 4.1 สัญญาณเสียงพูดแบบสเตอริโอ ที่ใช้ในการจำลองการทำงานของระบบ SAEC .....	48
รูปที่ 4.2 วิธีสะท้อนทางเสียงในห้องรับที่จำลองขึ้น มีขนาดความยาว $L=256$ .....	49
รูปที่ 4.3 ค่า WEVN ในระบบ SAEC ในขณะที่ยังไม่ได้มีการดีคอดรีเลชันสัญญาณ.....	50
รูปที่ 4.4 ตัวอย่างสัญญาณที่ถูกดีคอดรีเลชันโดยเพิ่มสัญญาณรบกวนสุ่มเมื่อ $SNR=20$ dB .....	51

ภาพประกอบ	หน้า
รูปที่ 4.5 (a) ค่า WEVN ของระบบ SAEC เมื่อมีการเพิ่มสัญญาณรบกวนสุ่ม.....	52
รูปที่ 4.5 (b) ค่า Coherence Function ของสัญญาณเสียงสเตอริโอ เมื่อมีการเพิ่มสัญญาณรบกวนสุ่ม.....	52
รูปที่ 4.6 เปรียบเทียบความหนาแน่นกำลังเชิงสเปกตรัมสัญญาณรบกวนสุ่ม กับความหนาแน่นกำลังเชิงสเปกตรัมสัญญาณเสียงพูด.....	53
รูปที่ 4.7 ตัวอย่างสัญญาณที่ถูกดีคอดรีเลชันเมื่อเพิ่มภาวะไม่เชิงเส้น เมื่อ $\alpha=0.8$ .....	54
รูปที่ 4.8 (a) ค่า WEVN ของระบบ SAEC เมื่อเพิ่มภาวะไม่เชิงเส้น .....	55
รูปที่ 4.8 (b) ค่า Coherence Function ของสัญญาณเสียงสเตอริโอเมื่อเพิ่มภาวะไม่เชิงเส้น .....	55
รูปที่ 4.9 เปรียบเทียบความหนาแน่นกำลังเชิงสเปกตรัมสัญญาณภาวะไม่เชิงเส้น กับความหนาแน่นกำลังเชิงสเปกตรัมสัญญาณเสียงพูด.....	57
รูปที่ 4.10 ตัวอย่างพลังงานเสียงสเตอริโอในช่องสัญญาณหนึ่ง เมื่อเพิ่มสัญญาณรบกวนแบบปรับตัวได้.....	58
รูปที่ 4.11 (a) ค่า WEVN ของระบบ SAECเมื่อเพิ่มสัญญาณรบกวนแบบปรับตัวได้.....	60
รูปที่ 4.11 (b) ค่า Coherence Function ของสัญญาณเสียงสเตอริโอเมื่อเพิ่มสัญญาณรบกวน แบบปรับตัวได้.....	60
รูปที่ 4.12 พลังงานเฉลี่ยของสัญญาณที่เปลี่ยนแปลงตามสมการที่ (3.12) .....	61
รูปที่ 4.13 (a) ค่า WEVN ของระบบ SAEC เมื่อทำการดีคอดรีเลชันสัญญาณเสียงสเตอริโอ ด้วยภาวะไม่เชิงเส้นที่นำเสนอ ที่ปรับค่าตามพลังงานเฉลี่ยของสัญญาณเสียงพูด .....	62
รูปที่ 4.13 (b) ค่า Coherence Function ของสัญญาณเสียงสเตอริโอ เมื่อทำการดีคอดรีเลชัน สัญญาณเสียงสเตอริโอด้วยภาวะไม่เชิงเส้นที่นำเสนอ ที่ปรับค่าตามพลังงานเฉลี่ย ของสัญญาณเสียงพูด.....	62
รูปที่ 4.14 เปรียบเทียบความหนาแน่นกำลังเชิงสเปกตรัมสัญญาณภาวะไม่เชิงเส้นที่นำเสนอ กับความหนาแน่นกำลังเชิงสเปกตรัมสัญญาณเสียงพูด.....	63
รูปที่ 4.15 (a) ค่า WEVN ของระบบ SAEC เมื่อดีคอดรีเลชันสัญญาณเสียงสเตอริโอด้วยสัญญาณ รบกวนแบบปรับตัวได้ที่นำเสนอ ( $\beta = 0.6$ ).....	65
รูปที่ 4.15 (b) ค่า Coherence Function ของสัญญาณเสียงสเตอริโอเมื่อดีคอดรีเลชันสัญญาณ เสียงสเตอริโอด้วยสัญญาณรบกวนแบบปรับตัวได้ที่นำเสนอ ( $\beta = 0.6$ ).....	65

รูปที่ 4.16 เปรียบเทียบความหนาแน่นกำลังเชิงสเปกตรัมสัญญาณรบกวนแบบปรับตัวได้  
ที่นำเสนอกับความหนาแน่นกำลังเชิงสเปกตรัมสัญญาณเสียงพูด.....66

รูปที่ 4.17 (a) ค่า WEVN ของระบบ SAEC เมื่อดีคอดรีเลชันสัญญาณเสียงสเตอริโอด้วยวิธีต่างๆ... 67

รูปที่ 4.17 (b) ค่า Coherence Function ของสัญญาณเสียงสเตอริโอเมื่อดีคอดรีเลชัน  
สัญญาณเสียงสเตอริโอด้วยวิธีต่างๆ .....67

รูปที่ 4.18 ค่า MOS ของสัญญาณเสียงสเตอริโอจากการดีคอดรีเลชันสัญญาณโดยการเพิ่ม  
สัญญาณรบกวนสุ่ม เปรียบเทียบการเพิ่มสัญญาณรบกวนแบบปรับตัวได้.....70

รูปที่ 4.19 (a) การเปรียบเทียบ WEVN ของระบบ SAEC กรณีที่มีการเข้ารหัสและถอดรหัส  
สัญญาณเสียงพูดตามมาตรฐาน G.723.1 กับระบบที่ใช้สัญญาณเสียงเดิม  
ที่ยังไม่ได้เข้ารหัสและถอดรหัส.....72

รูปที่ 4.19 (b) การเปรียบเทียบค่า Coherence Function ของสัญญาณเสียงสเตอริโอ  
กรณีที่มีการเข้ารหัสและถอดรหัสสัญญาณเสียงพูดตามมาตรฐาน G723.1  
กับระบบที่ใช้สัญญาณเสียงเดิมที่ยังไม่ได้เข้ารหัสและถอดรหัส .....72

รูปที่ 4.20 (a) ค่า WEVN ของระบบ SAEC ที่ใช้การเข้ารหัสและถอดรหัสตามมาตรฐาน  
G.723.1 เมื่อใช้วิธีดีคอดรีเลชันสัญญาณเสียงสเตอริโอดังกล่าวด้วยวิธีที่นำเสนอ.....74

รูปที่ 4.20 (b) ค่า Coherence Function ของสัญญาณเสียงสเตอริโอที่ใช้การเข้ารหัสและถอดรหัส  
ตามมาตรฐาน G.723.1 เมื่อใช้วิธีดีคอดรีเลชันสัญญาณเสียงสเตอริโอดังกล่าว  
ด้วยวิธีที่นำเสนอ.....74

## บัญชีคำศัพท์

การกำจัดสัญญาณเสียงสะท้อน	Acoustic Echo Cancellation
การกำจัดสัญญาณเสียงสะท้อนแบบสเตอริโอ	Stereophonic Acoustic Echo Cancellation
การควอนไทซ์	Quantization
การซัดตัวอย่าง	Sampling
การเชื่อมต่อ	Coupling
การดีคอรีเลชัน	Decorrelation
การติดตาม	Tracking
การทำนาย	Prediction
การนอร์มัลไลซ์	Normalization
การเบี่ยงเบนของคำตอบ	Misadjustment
การประมวลผลก่อน	Pre-processing
การประมาณกำลังเชิงสเปกตรัม	Power Spectral Estimation
การประมาณขณะหนึ่ง	Instantaneous Estimate
การป้อนกลับ	Feedback
การเพิ่มสัญญาณรบกวนแบบปรับตัวได้	Adaptive Noise Addition
การเพิ่มสัญญาณรบกวนสุ่ม	Random Noise Addition
การวนซ้ำ	Iteration
การสังเคราะห์	Synthesis
การแสดงเอกลักษณ์ของระบบ	System Identification
ขั้นตอนวิธี	Algorithm
ความซับซ้อนในการคำนวณ	Computational Complexity
ความถี่ในการซัดตัวอย่าง	Sampling Frequency
ความเพี้ยน	Distortion
ความเพี้ยนจากการควอนไทซ์	Quantization Distortion
คอนโวลูชัน	Convolution
คอหอย/โพรงคอ	Pharynx
ค่าจริง	Real Value

ค่าเจาะจง	Eigenvalue
ค่าเฉลี่ยทั้งหมดเชิงสถิติ	Ensemble Average
ค่าช่วงก้าว	Step-size
คาบ	Period
คาบเงียบ	Silence Period
คาบที่มีเสียงพูด	Speech Active Period
ค่าสัมบูรณ์	Absolute Value
จุดเริ่มเปลี่ยน	Threshold
ช่องสัญญาณ	Channel
ตัวเข้ารหัสและถอดรหัสเสียงพูด	Speech Codec
ตัวเข้ารหัสเสียงพูด	Speech Encoder
ตัวถอดรหัสเสียงพูด	Speech Decoder
ตัวเรียงกระแสแบบครึ่งคลื่น	Half-wave Rectifier
ต่ำสุด	Minimum
เทคนิคการดีคอรีเลชัน	Decorrelation Technique
ปอด	Lung
ผลตอบสนองต่ออิมพัลส์	Impulse Response
ผลตอบสนองต่ออิมพัลส์จำกัด	Finite Impulse Response
พิตช์	Pitch
โพรงจมูก	Velum
ฟังก์ชันคอนเวกซ์	Convex Function
ฟังก์ชันต้นทุน	Cost Function
ฟังก์ชันร่วมนัย	Coherence Function
เฟรม	Frame
ภาวะไม่เชิงเส้น	Non-linearity
เมตริกซ์ผกผัน	Inverse Matrix
เมตริกซ์เอกลักษณ์	Identity Matrix
ระบบการประชุมทางไกล	Teleconferencing System

ระบบการประชุมทางวิดีโอ	Video Conferencing System
ระบบช่องสัญญาณเดียว	Mono Channel System
ระบบเชิงเส้นไม่เปลี่ยนแปลงตามเวลา	Linear Time Invariant System
ระบบโทรศัพท์เคลื่อนที่	Mobile Phone System
ระบบเปลี่ยนแปลงตามเวลา	Time-varying System
ระบบไม่เปลี่ยนแปลงตามเวลา	Time Invariant System
ระบบสเตอริโอ	Stereo System
ลักษณะเฉพาะ	Characteristic
วงจรรองแบบปรับตัว	Adaptive filter
วงจขยาย	Amplifier
วิถีสะท้อนทางเสียง	Acoustic Echo Path
สถานะคงตัว	Steady State
สมการปรับให้ทันกาล	Update Equation
สมรรถนะที่ทนทาน	Robust Performance
สหสัมพันธ์	Correlation
สหสัมพันธ์ข้าม	Cross-correlation
สัญญาณรบกวนขาวแบบเกาส์	White Gaussian Noise
สัญญาณสะท้อน	Echo
สัญญาณเสียงสะท้อน	Acoustic Echo
เส้นเสียง/สายเสียง	Local Cord
หลอดลม	Larynx
ห้องรับ	Receiving Room
ห้องส่ง	Transmitting Room
เหมาะสมที่สุด	Optimum
อัตราการลู่เข้า	Convergence Rate
อัตราบิต	Bit Rate
อัตราส่วนสัญญาณต่อสัญญาณรบกวน	Signal-to-Noise Ratio
อัตสหสัมพันธ์	Autocorrelation



อุดมคติ

Ideal

แอมพลิจูด

Amplitude



สถาบันวิทยบริการ  
จุฬาลงกรณ์มหาวิทยาลัย

# บทที่ 1

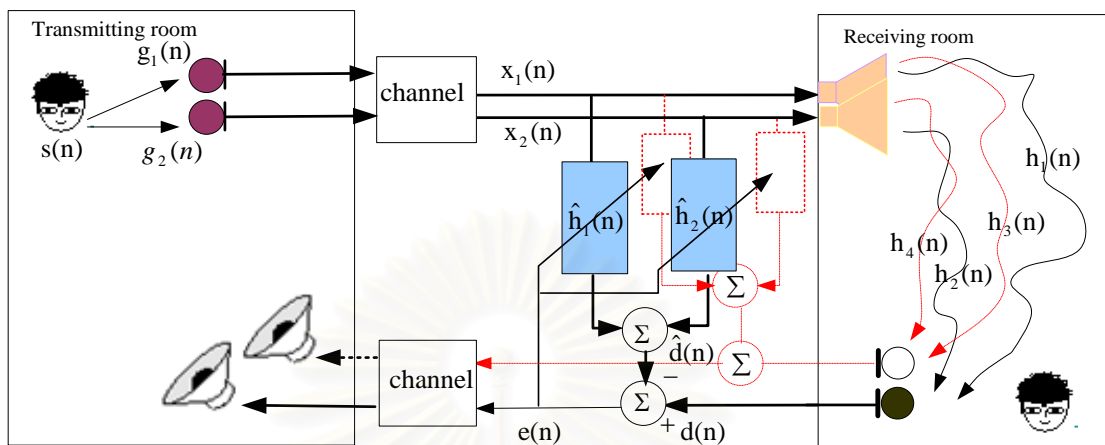
## บทนำ

### 1.1 แนวคิดและเหตุผล

ระบบการสื่อสารทางเสียงมีความสำคัญอย่างมากในการสื่อสารยุคปัจจุบัน หลายครั้งที่การส่งสัญญาณทางเสียงมักได้รับการพัฒนาควบคู่กับการส่งสัญญาณทางภาพเพื่อให้สามารถแสดงผลได้อย่างมีประสิทธิภาพมากที่สุด ทั้งนี้ เทคโนโลยีทางด้านการสื่อสารทางเสียงได้รับการพัฒนาไปสู่การใช้ระบบเสียงแบบหลายช่องสัญญาณ เริ่มตั้งแต่ระบบสเตอริโอ (Stereo System) ขึ้นไป ถือว่ามีความสำคัญและได้รับความนิยมอย่างมากต่อการใช้งานในปัจจุบัน เนื่องจากสามารถให้คุณภาพเสียงที่ดีกว่าและสมจริงกว่าการสื่อสารทางเสียงในระบบช่องสัญญาณเดียว (Mono Channel System) เพราะมิติทางเสียงสามารถช่วยให้ผู้ฟังสามารถแยกแยะทิศทางเสียงของผู้พูดได้ [1-4][7] โดยเฉพาะอย่างยิ่งในระบบการสื่อสารทางเสียงที่ผู้พูดมีการเคลื่อนที่ได้เป็นอย่างดี เช่น ระบบการประชุมทางไกล (Teleconferencing System) หรือระบบโทรศัพท์เคลื่อนที่ (Mobile Phone System) เป็นต้น

อย่างไรก็ตามทุกครั้งที่มีการสื่อสารข้อมูลทางเสียงไม่ว่าจะเป็นด้วยระบบใด มักจะพบปัญหาที่เกิดขึ้นร่วมกันเสมอคือ การเกิดสัญญาณเสียงสะท้อน (Acoustic Echo) ซึ่งถือว่าเป็นอุปสรรคสำคัญต่อการสื่อสารทางเสียง ไม่ว่าจะเป็นการใช้เทคโนโลยีการสื่อสารทางโทรศัพท์ วิทยุ โทรทัศน์ ดาวเทียม เคเบิลหรือแม้กระทั่งเครือข่ายไร้สาย ทั้งนี้สัญญาณเสียงสะท้อนเกิดจากการที่สัญญาณเสียงของผู้พูดทางด้านห้องส่ง (Transmitting Room) ถูกส่งผ่านไปในช่วงสัญญาณ (Channel) ไปยังห้องรับ (Receiving Room) และสัญญาณเสียงที่ออกจากลำโพงทางห้องรับเกิดการสะท้อนกลับเข้าสู่ไมโครโฟนทางห้องรับ จึงทำให้มีพลังงานเสียงบางส่วนถูกส่งผ่านช่องสัญญาณกลับไปยังห้องส่ง ทำให้ผู้พูดทางด้านห้องส่งได้ยินสัญญาณเสียงสะท้อนของตนเอง ส่งผลให้เกิดการรบกวนการสนทนา ดังนั้นการกำจัดสัญญาณเสียงสะท้อน (Acoustic Echo Cancellation, AEC) ในระบบการสื่อสารทางเสียงจึงมีความจำเป็น เพื่อเพิ่มประสิทธิภาพของระบบการสื่อสารให้มากขึ้น วิธีหนึ่งในการกำจัดสัญญาณเสียงสะท้อนดังกล่าว ทำได้โดยการนำวงจรกรองแบบปรับตัว (Adaptive Filter) มาใช้ประมาณสัญญาณเสียงสะท้อนที่เกิดขึ้นโดยการจำลองวิถีสะท้อนทางเสียง (Acoustic Echo Path, AEP) ทางด้านห้องรับ แล้วนำไปลบออกจากสัญญาณเสียงสะท้อนที่เกิดขึ้นจริงก็ส่งผลให้สามารถกำจัดสัญญาณเสียงสะท้อนที่เกิดขึ้นได้

เมื่อพิจารณาการกำจัดสัญญาณเสียงสะท้อนแบบสเตอริโอ (Stereophonic Acoustic Echo Cancellation, SAEC) ดังรูปที่ 1.1



รูปที่ 1.1 แบบจำลองการกำจัดสัญญาณเสียงสะท้อนแบบสเตอริโอ

ไมโครโฟนทั้งสองตัวในห้องส่งทำหน้าที่เปลี่ยนสัญญาณเสียงของผู้พูด  $s(n)$  ในห้องส่งให้เป็นสัญญาณทางไฟฟ้า  $x_1(n)$  และ  $x_2(n)$  เพื่อส่งต่อไปยังห้องรับ โดยทางเดินของคลื่นเสียง  $s(n)$  ไปยังไมโครโฟนทั้งสองตัวในห้องส่ง สามารถจำลองได้ด้วยผลตอบสนองต่ออิมพัลส์ (Impulse Response)  $g_1(n)$  และ  $g_2(n)$  เมื่อสัญญาณสเตอริโอ  $x_1(n)$  และ  $x_2(n)$  มาถึงยังห้องรับจะถูกลำโพงทางฝั่งห้องรับเปลี่ยนเป็นสัญญาณเสียงให้ผู้ฟังได้ยิน ทั้งนี้สัญญาณเสียงที่ออกจากลำโพง  $x_1(n)$  และ  $x_2(n)$  อาจเกิดการสะท้อนกลับเข้าสู่ไมโครโฟนในห้องรับ โดยวิธีสะท้อนทางเสียง (Acoustic Echo Path, AEP) จากลำโพงทั้งสองตัวมายังไมโครโฟนตัวที่ A ในห้องรับถูกจำลองได้โดยผลตอบสนองต่ออิมพัลส์  $h_1(n)$  และ  $h_2(n)$  ส่งผลให้เกิดสัญญาณสะท้อน  $d(n)$  ถูกส่งไปยังห้องส่ง ทำให้ผู้พูดในห้องส่งได้ยินสัญญาณเสียงสะท้อนของตนเอง ถ้าอัตราขยาย (Gain) ของระบบมีค่าสูง สัญญาณสะท้อนดังกล่าวจะวนเป็นวงป้อนกลับ (Feedback Loop) เกิดเป็นเสียงหอน (Howling) ได้ สำหรับการกำจัดสัญญาณเสียงสะท้อนซึ่งเกิดจากผลของวิธีสะท้อนทางเสียงจากลำโพงทั้งสองตัวในห้องรับมายังไมโครโฟนตัวที่ B ที่ถูกจำลองด้วยผลตอบสนองต่ออิมพัลส์  $h_3(n)$  และ  $h_4(n)$  จะเป็นไปในทำนองเดียวกันกับการกำจัดสัญญาณเสียงสะท้อนที่ไมโครโฟนตัวที่ A จึงไม่ถูกพิจารณาในวิทยานิพนธ์ฉบับนี้ ดังนั้นจึงเห็นได้ว่าการกำจัดสัญญาณเสียงสะท้อนแบบสเตอริโอเป็นสิ่งที่จำเป็นในระบบการสื่อสารทางเสียงแบบสเตอริโอ เพื่อเพิ่มประสิทธิภาพของการสื่อสารให้ดีขึ้น

การกำจัดสัญญาณเสียงสะท้อนแบบสเตอริโอหรือ SAEC โดยใช้วงจรกรองแบบปรับตัวสามารถทำได้ 2 แบบ แบบที่ 1 ทำได้โดยการใช้อัตรากรองแบบปรับตัวเพียงตัวเดียวจำลองวิถีสะท้อนทางเสียง  $h_1(n)$  และ  $h_2(n)$  [1] แบบที่ 2 ทำได้โดยการใช้อัตรากรองแบบปรับตัวสองตัวจำลองวิถีสะท้อนทางเสียง  $h_1(n)$  และ  $h_2(n)$  อย่างไรก็ตามพบว่า วิธีใช้อัตรากรองแบบปรับตัวเพียงตัวเดียวเพื่อจำลองวิถีสะท้อนทางเสียงทั้งสองตัวนั้น จะมีความยุ่งยากสำหรับการหาขั้นตอนวิธี (Algorithm) และฟังก์ชันในการประมาณค่าสัญญาณเสียงในช่องสัญญาณหนึ่งด้วยสัญญาณเสียงอีกช่องสัญญาณหนึ่ง ที่สำคัญฟังก์ชันที่นำมาใช้ในการประมาณสัญญาณเสียงนั้น อาจจะทำให้เกิดการหน่วงเวลา จนกระทั่งทำให้ระบบไม่มีเสถียรภาพได้เช่น ฟังก์ชันอิมพัลส์ [1][11]

โดยทั่วไป แบบจำลองระบบ SAEC จะใช้แบบที่ 2 ดังนั้นในวิทยานิพนธ์ฉบับนี้ได้เลือกใช้แบบจำลองการกำจัดสัญญาณเสียงสะท้อนที่ใช้อัตรากรองแบบปรับตัว 2 ตัว ที่ถูกจำลองด้วยผลตอบสนองต่ออิมพัลส์  $\hat{h}_1(n)$  และ  $\hat{h}_2(n)$  สัญญาณเสียงสะท้อนที่ประมาณได้จากการจำลองของ  $\hat{h}_1(n)$  และ  $\hat{h}_2(n)$  คือ  $\hat{d}(n)$  จะถูกนำไปหักออกจากสัญญาณเสียงสะท้อน  $d(n)$  ทำให้ได้ค่าผิดพลาดเป็น  $e(n)$  ค่าผิดพลาดดังกล่าวจะถูกนำไปควบคุมการทำงานของวงจรกรองแบบปรับตัวเพื่อประมาณวิถีสะท้อนทางเสียงในรอบถัดไป จนกระทั่งสามารถปรับปรุงสัมประสิทธิ์ของวงจรกรองให้มีความใกล้เคียงกับวิถีสะท้อนจริงมากที่สุด เมื่อค่าผิดพลาดมีค่าเข้าใกล้ศูนย์ ซึ่งแสดงว่าสัญญาณเสียงสะท้อนถูกกำจัดไปได้

เนื่องจากเสียงแบบสเตอริโอจะมีผลต่อมิติกการได้ยินของผู้ฟัง [2-3] ดังนั้นเมื่อมีสัญญาณเสียงสะท้อนเกิดขึ้น จึงเป็นการยากอย่างยิ่งในการแยกแยะผู้พูดและทิศทางที่เกิดขึ้น โดยเฉพาะอย่างยิ่งในระบบการประชุมแบบวีดิทัศน์ (Video Conferencing System) ที่ต้องระวังในเรื่องการประมาณวิถีสะท้อนอย่างแม่นยำ เพื่อสามารถแสดงภาพและเสียงได้อย่างสัมพันธ์กัน [3] เนื่องจากสัญญาณเสียงสเตอริโอ  $x_1(n)$  และ  $x_2(n)$  ในระบบ SAEC มีสหสัมพันธ์ (Correlation) กันอยู่ [1-5] จึงส่งผลต่อการปรับค่าสัมประสิทธิ์วงจรกรองแบบปรับตัว  $\hat{h}_1(n)$  และ  $\hat{h}_2(n)$  ให้มีลักษณะใกล้เคียงกับวิถีสะท้อนทางเสียง  $h_1(n)$  และ  $h_2(n)$  ตามลำดับ กล่าวคือเมื่อทำการปรับค่าสัมประสิทธิ์วงจรกรองแบบปรับตัว  $\hat{h}_1(n)$  จะเกิดผลกระทบต่อค่าสัมประสิทธิ์วงจรกรองแบบปรับตัว  $\hat{h}_2(n)$  ด้วย โดยเรียกปัญหาที่เกิดขึ้นนี้ว่า Non-uniqueness Problem [1-3] ซึ่งจะเป็นประเด็นหลักที่ถูกพิจารณาในวิทยานิพนธ์ฉบับนี้ คือการหาวิธีการแก้ไขปัญหา Non-uniqueness Problem เพื่อให้ระบบ SAEC สามารถกำจัดสัญญาณเสียงสะท้อนได้อย่างมีประสิทธิภาพมากขึ้นและเป็นหลักในการนำไปประยุกต์ใช้ในระบบการทำงานจริงต่อไป

วิธีหนึ่งในการแก้ปัญหา Non-uniqueness Problem คือ การลดสหสัมพันธ์ของสัญญาณเสียงสเตอริโอ  $x_1(n)$  และ  $x_2(n)$  หรือเรียกว่า การดีคอรีเลชัน (Decorrelation) ของ

สัญญาณสเตอริโอ [2-3] เพื่อให้การปรับค่าสัมประสิทธิ์วงจรรองแบบปรับตัวทั้งสอง คือ  $\hat{h}_1(n)$  และ  $\hat{h}_2(n)$  เป็นอิสระต่อกันมากที่สุด ส่งผลให้สามารถประมาณค่าวิถีสะท้อนทางเสียง  $h_1(n)$  และ  $h_2(n)$  ได้ใกล้เคียงมากยิ่งขึ้น นอกจากนี้ควรมีการคำนึงถึง การรักษาไว้ซึ่งคุณภาพของเสียงสเตอริโอที่ผ่านวิธีการดีคอดิเลชันแล้ว ให้มีคุณภาพคล้ายเสียงสเตอริโอเดิมก่อนการดีคอดิเลชันมากที่สุด หรือมีผลกระทบต่อคุณภาพเสียงสเตอริโอ (Stereo Perception) น้อยที่สุด เนื่องจากเป็นเสียงเดียวกับที่ผู้ฟังทางด้านห้องรับจะต้องได้ยิน

ทั้งนี้วิธีการดีคอดิเลชันสัญญาณเสียงสเตอริโอนั้นมีอยู่หลายวิธี [1-4] หรือแม้แต่วิธีการเข้ารหัสสัญญาณเสียงพูดก่อนส่งไปในช่องสัญญาณก็สามารถช่วยให้สหสัมพันธ์สัญญาณเสียงสเตอริโอนั้นลดลงได้เช่นกัน [8]

ดังนั้นวิทยานิพนธ์ฉบับนี้จึงทำการศึกษาแนวคิดในการเพิ่มประสิทธิภาพระบบการกำจัดสัญญาณเสียงสะท้อนแบบสเตอริโอ ด้วยการนำเสนอวิธีหนึ่งในการดีคอดิเลชันสัญญาณเสียงสเตอริโอให้มีสหสัมพันธ์กันน้อยลง คือ การเพิ่มสัญญาณรบกวนแบบปรับตัวได้ (Adaptive Noise Addition) โดยการเปรียบเทียบกับวิธีการดีคอดิเลชันแบบอื่นๆ คือ การเพิ่มสัญญาณรบกวนสุ่ม (Random Noise Addition) ในแต่ละช่องสัญญาณ [1] และการเพิ่มภาวะไม่เชิงเส้น (Non-linearity) ในแต่ละช่องสัญญาณ [2-3][5] ทั้งในด้านของประสิทธิภาพการดีคอดิเลชัน และการรักษาไว้ซึ่งคุณภาพเสียงเชิงสเตอริโอ ของวิธีทั้งสองนี้และวิธีที่นำเสนอ

## 1.2 วัตถุประสงค์

ศึกษา วิเคราะห์ เพื่อพัฒนาประสิทธิภาพการกำจัดสัญญาณเสียงสะท้อนแบบสเตอริโอโดยใช้วิธีการดีคอดิเลชันสัญญาณเสียงสเตอริโอที่อาศัยการเพิ่มสัญญาณรบกวนแบบปรับตัวได้

## 1.3 เป้าหมายและขอบเขตของงานวิจัย

พัฒนาวิธีการดีคอดิเลชันสัญญาณเสียงสเตอริโอ โดยการเพิ่มสัญญาณรบกวนแบบปรับตัวได้ ในระบบการกำจัดสัญญาณเสียงสะท้อนแบบสเตอริโอ และทำการเปรียบเทียบกับวิธีการดีคอดิเลชันสัญญาณเสียงสเตอริโอที่ใช้การเพิ่มสัญญาณรบกวนสุ่ม และการเพิ่มภาวะไม่เชิงเส้น

#### 1.4 ขั้นตอนและวิธีการดำเนินการ

1. ศึกษาถึงที่มาของปัญหาการเกิดสัญญาณเสียงสะท้อนแบบสเตริโอ รวมทั้งวิธีการกำจัดสัญญาณเสียงสะท้อนแบบสเตริโอ หรือระบบ SAEC
2. ศึกษาวิธีการดีคอดิเลชันสัญญาณเสียงสเตริโอแบบต่างๆ โดยคำนึงถึงผลกระทบต่อคุณภาพเสียงสเตริโอ พร้อมกับประสิทธิภาพการทำงานของระบบ SAEC
3. ศึกษาและนำเสนอวิธีการเพิ่มสัญญาณรบกวนแบบปรับตัวได้ในการดีคอดิเลชันสัญญาณเสียงสเตริโอเพื่อปรับปรุงประสิทธิภาพการทำงานของระบบ SAEC ให้ดีขึ้น
4. ทำการทดสอบเขียนโปรแกรมจำลองการทำงานของระบบ SAEC และทำการทดสอบคุณภาพเสียงสเตริโอที่ถูกดีคอดิเลชันด้วยวิธีต่างๆ
5. วิเคราะห์และประเมินผลการทำงานของระบบ SAEC โดยวิธีการดีคอดิเลชันสัญญาณเสียงสเตริโอที่นำเสนอไว้ แล้วเปรียบเทียบผลที่เกิดขึ้นจริง ทั้งในด้านประสิทธิภาพของระบบ SAEC และคุณภาพเสียงสเตริโอที่ได้
6. สรุปผลงานวิจัยและจัดทำรูปเล่มวิทยานิพนธ์ฉบับสมบูรณ์

#### 1.5 ประโยชน์ที่คาดว่าจะได้รับ

1. การศึกษาถึงวิธีการต่าง ๆ ที่ใช้ในการดีคอดิเลชันสัญญาณเสียงสเตริโอและผลที่เกิดขึ้นในระบบ SAEC และสามารถเปรียบเทียบข้อดีและข้อจำกัดของแต่ละวิธีที่มีผลต่อเสียงสเตริโอได้
2. ได้วิธีการดีคอดิเลชันสัญญาณเสียงสเตริโอแบบใหม่สำหรับระบบ SAEC

#### 1.6 ภาพรวมของวิทยานิพนธ์

สำหรับเนื้อหาของวิทยานิพนธ์ฉบับนี้แบ่งเป็น 5 บทรายละเอียดต่อไปนี้

บทที่ 1 บทนำ มีเนื้อหาเกี่ยวกับแนวเหตุผลในการศึกษาถึงระบบ SAEC แนวทางของวิทยานิพนธ์วัตถุประสงค์ ของเขตวิทยานิพนธ์ และภาพรวมของวิทยานิพนธ์

บทที่ 2 นำเสนอทฤษฎีที่เกี่ยวข้องในเรื่องการเกิดสัญญาณเสียงสะท้อนในระบบช่องสัญญาณเดี่ยว การเกิดสัญญาณเสียงสะท้อนแบบสเตริโอ การจำลองการทำงานของระบบ SAEC รวมถึงปัญหาที่เกิดขึ้นในระบบ SAEC นอกจากนี้ก็ได้กล่าวถึงลักษณะสัญญาณเสียงพูดเพื่อเป็นพื้นฐานในการทำความเข้าใจในวิธีที่จะนำเสนอในบทถัดไป

บทที่ 3 อธิบายวิธีการดีคอล์เลชันสัญญาณเสียงสเตอริโอเพื่อให้มีสหสัมพันธ์กันน้อยลง โดยวิธีการเพิ่มสัญญาณรบกวนสุ่ม การเพิ่มภาวะไม่เชิงเส้น รวมทั้งได้นำเสนอวิธีการดีคอล์เลชันสัญญาณเสียงสเตอริโอโดยการเพิ่มสัญญาณรบกวนแบบปรับตัวได้ รวมถึงเกณฑ์ในการเปรียบเทียบประสิทธิภาพการดีคอล์เลชัน นอกจากนี้ได้กล่าวถึงการเข้ารหัสสัญญาณเสียงพูดสำหรับใช้ในการอธิบายผลการทดลองในบทถัดไป

บทที่ 4 มีเนื้อหาเกี่ยวกับผลการทดลอง โดยจะเป็นการแสดงผลการเปรียบเทียบของประสิทธิภาพการทำงานของระบบ SAEC เมื่อมีการดีคอล์เลชันสัญญาณเสียงสเตอริโอด้วยวิธีต่างๆ คือ การเพิ่มสัญญาณรบกวนสุ่ม การเพิ่มภาวะไม่เชิงเส้นและวิธีที่นำเสนอในบทที่ 3 รวมทั้งการแสดงผลประสิทธิภาพการทำงานของระบบ SAEC ที่ใช้วิธีการดีคอล์เลชันที่นำเสนอ เมื่อมีการนำไปใช้ร่วมด้วยกับระบบที่มีการเข้ารหัสและถอดรหัสเสียงพูด

บทที่ 5 บทสรุปและข้อเสนอแนะ จะสรุปเกี่ยวกับเนื้อหาที่สำคัญในวิทยานิพนธ์รวมทั้งปัญหาที่เกิดขึ้นและข้อเสนอแนะสำหรับงานวิจัยในอนาคต



สถาบันวิทยบริการ  
จุฬาลงกรณ์มหาวิทยาลัย

## บทที่ 2

### หลักการและขั้นตอนวิธีที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงการใช้งานของวงจรกรองแบบปรับตัวเพื่อการสื่อสารทางเสียง ในระบบการกำจัดสัญญาณเสียงสะท้อน (Acoustic Echo Cancellation, AEC) และระบบการกำจัดสัญญาณเสียงสะท้อนแบบสเตอริโอ (Stereophonic Acoustic Echo Cancellation, SAEC) โดยจะกล่าวถึงการทำงานของวงจรกรองแบบปรับตัวในด้านการแสดงเอกลักษณ์ของระบบ (System Identification) และสรุปถึงขั้นตอนวิธีที่นำมาใช้กับวงจรกรองแบบปรับตัว รวมถึงปัญหาที่เกิดขึ้นเฉพาะในระบบ SAEC นอกจากนี้จะได้กล่าวถึงพื้นฐานของสัญญาณเสียงพูด เพื่อเป็นพื้นฐานในการทำความเข้าใจถึงวิธีที่จะนำเสนอในบทต่อไป

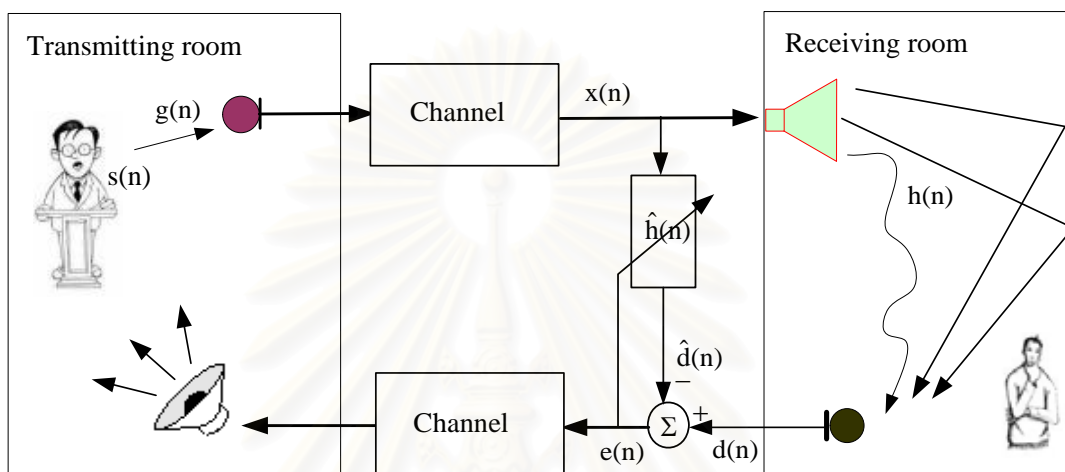
#### 2.1 ระบบการกำจัดสัญญาณเสียงสะท้อน

##### 2.1.1 ปัญหาสัญญาณเสียงสะท้อนในระบบการสื่อสารทางเสียง

การศึกษาเกี่ยวกับปัญหาสัญญาณเสียงสะท้อนที่เกิดขึ้นในระบบการสื่อสารทางเสียงในระบบช่องสัญญาณเดียว ถือว่ามีบทบาทที่สำคัญยิ่งต่อการนำไปสู่การพิจารณาแก้ปัญหาสัญญาณเสียงสะท้อนที่เกิดขึ้นในระบบที่มีการสื่อสารทางเสียงแบบหลายช่องสัญญาณ เช่นระบบสเตอริโอ เป็นต้น เมื่อพิจารณาปัญหาการเกิดสัญญาณเสียงสะท้อนในระบบการสื่อสารทางเสียงแบบช่องสัญญาณเดียว ระหว่าง 2 ห้อง ในแต่ละห้อง มีไมโครโฟนและลำโพงอย่างละ 1 ชุด ดังรูปที่ 2.1 จะเห็นว่า เมื่อผู้พูดในห้องส่ง (Transmitting Room) กำลังพูดอยู่ ทางเดินของคลื่นเสียง  $s(n)$  ของผู้พูดสามารถจำลองด้วยผลตอบสนองต่ออิมพัลส์  $g(n)$  สัญญาณทางไฟฟ้า  $x(n)$  จะถูกส่งผ่านช่องสัญญาณซึ่งสมมติให้เป็นช่องสัญญาณอุดมคติ (Ideal) ไปยังห้องรับ (Receiving Room) แล้วถูกเปลี่ยนเป็นเสียงพูดด้วยลำโพงที่อยู่ทางด้านห้องรับ สัญญาณเสียงที่ออกจากลำโพงในห้องรับ  $x(n)$  อาจเกิดการสะท้อนกับบริเวณโดยรอบกลับเข้าสู่ไมโครโฟนในห้องรับ ทำให้มีพลังงานเสียงบางส่วนถูกส่งผ่านช่องสัญญาณกลับไปยังห้องส่ง ผู้พูดในห้องส่งจึงได้ยินสัญญาณเสียงสะท้อนของตนเอง ทำให้เกิดการรบกวนการสนทนา ดังนั้นเพื่อแก้ไขปัญหานี้ สัญญาณเสียงสะท้อนที่เกิดขึ้น ทางห้องรับจึงจำเป็นต้องมีระบบที่ประมาณวิธีสะท้อนทางเสียง  $h(n)$  ที่เกิดขึ้นและกำจัดสัญญาณเสียงสะท้อนดังกล่าว ก่อนที่ผู้พูดทางห้องรับจะพูดสื่อสาร



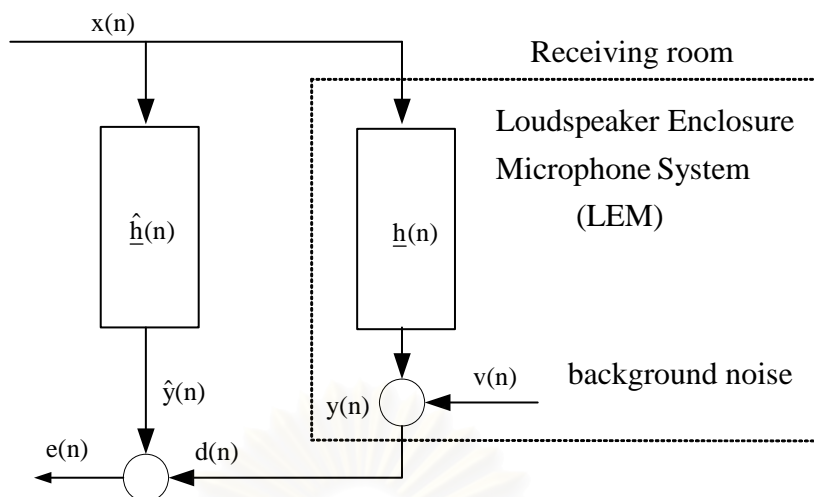
กลับไปยังห้องส่ง โดยใช้การกำจัดสัญญาณเสียงสะท้อน หรือเรียกว่าระบบ AEC ทางด้านภาครับนั่นเอง ในทำนองเดียวกัน ทางห้องส่งก็จำเป็นต้องมีระบบ AEC ด้วย เพื่อป้องกันไม่ให้ผู้พูดทางห้องรับได้ยินสัญญาณเสียงสะท้อนของตนเอง สำหรับในวิทยานิพนธ์นี้จะพิจารณาถึงระบบ AEC ทางห้องรับเท่านั้น ทั้งนี้หากระบบสามารถทำการกำจัดสัญญาณเสียงสะท้อนได้อย่างมีประสิทธิภาพ ก็สามารถนำไปพิจารณาใช้กับทางด้านห้องส่งได้เช่นเดียวกัน



รูปที่ 2.1 การเกิดสัญญาณเสียงสะท้อนในระบบการสื่อสารทางเสียง

### 2.1.2 การแสดงเอกลักษณ์ของระบบ (System Identification) ของวงจรกรองแบบปรับตัว สำหรับระบบ AEC

การทำงานของระบบการกำจัดสัญญาณเสียงสะท้อน คือ การหาแบบจำลองของการเชื่อมต่อ (Coupling) ระหว่างไมโครโฟนและลำโพงในสถานที่ปิด (Loudspeaker–Enclosure–Microphone, LEM) [12] หรือที่เรียกว่าการจำลองวิถีสะท้อนทางเสียง (Acoustic Echo Path, AEP) ระหว่างไมโครโฟนและลำโพง ซึ่งถูกจำลองโดยผลตอบสนองต่ออิมพัลส์  $\hat{h}(n)$  ดังแสดงในรูปที่ 2.2 ดังนั้นในการกำจัดสัญญาณเสียงสะท้อน จึงสามารถทำได้โดยใช้วงจรกรองแบบปรับตัว  $\hat{h}(n)$  ต่อขนานกับ LEM ดังรูปที่ 2.2



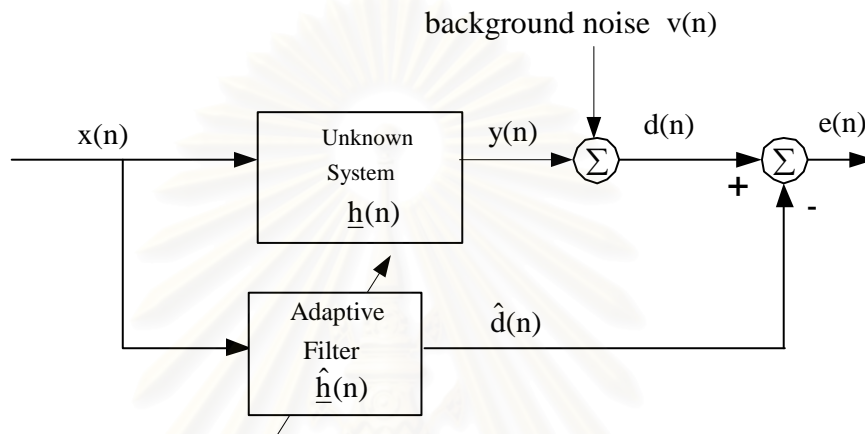
รูปที่ 2.2 แบบจำลองระบบ AEC โดยใช้วงจรรองแบบปรับตัวจำลอง AEP

พิจารณารูปที่ 2.2 ทางห้องรับ เมื่อสัญญาณเข้าของระบบ  $x(n)$  ถูกส่งมายังห้องรับ ซึ่งจำลองการเชื่อมต่อระหว่างไมโครโฟนและลำโพงในสถานที่ปิดในห้องรับด้วยวิธีสะท้อนทางเสียงที่เกิดขึ้น ส่งผลให้เกิดสัญญาณสะท้อน  $y(n)$  สะท้อนกลับไปยังผู้พูดในห้องส่งผ่านทางไมโครโฟนตัวหนึ่ง เมื่อ  $d(n)$  เกิดจากผลรวมของสัญญาณสะท้อน  $y(n)$  กับสัญญาณรบกวนพื้นหลัง  $v(n)$

เมื่อนำวงจรรองแบบปรับตัว  $\hat{h}(n)$  มาต่อขนานกับระบบ LEM ดังรูปที่ 2.2 สัญญาณออกของวงจรรองแบบปรับตัว  $\hat{y}(n)$  จะเป็นสัญญาณค่าประมาณของสัญญาณเสียงสะท้อน ผลต่างของสัญญาณเสียงสะท้อนจากไมโครโฟนกับสัญญาณออกของวงจรรองแบบปรับตัวจะเรียกว่า สัญญาณผิดพลาดในการประมาณ  $e(n)$  โดยสัญญาณผิดพลาดนี้จะถูกส่งกลับไปยังผู้พูดในห้องส่งผ่านทางสัญญาณไมโครโฟน ถ้าสัญญาณผิดพลาดมีค่าเข้าใกล้ศูนย์แสดงว่าระบบสามารถทำการกำจัดสัญญาณเสียงสะท้อนได้ ทำให้ผู้ฟังทางด้านห้องส่งไม่ได้ยินสัญญาณเสียงสะท้อนของตนเองอีกต่อไป โดยให้วงจรรองแบบปรับตัว  $\hat{h}(n)$  มีโครงสร้างแบบผลตอบสนองต่ออิมพัลส์จำนวนจำกัด (Finite Impulse Response, FIR) และมีจำนวนสัมประสิทธิ์หรือความยาวเท่ากับ  $L$  ทั้งนี้การเลือกความยาวของวงจรรองแบบปรับตัวให้เหมาะสมกับระบบหนึ่งๆ โดยความยาวของวงจรรองแบบปรับตัวจะสอดคล้องกับค่าประมาณของเวลาสะท้อนกลับ (Reverberation Time) ของ AEP ในห้องภาครับว่ามีค่าประมาณของเวลาสะท้อนเป็นอย่างไร เช่น ในสภาพแวดล้อมสำนักงานที่มีเวลาสะท้อนกลับประมาณ 300-500 ms ถ้าความถี่ซักรั่วอย่างของสัญญาณในระบบเท่ากับ 8 kHz ความยาว  $L$  ควรมีค่าประมาณ 2,400-4,000 ในขณะที่

สภาพแวดล้อมภายในรถยนต์ที่มีเวลาสะท้อนกลับประมาณ 50-100 ms ถ้าความถี่ซีกตัวอย่างของสัญญาณในระบบเท่ากับ 8 kHz ความยาว  $L$  ควรมีค่าประมาณ 400-800 [21]

รูปแบบการนำวงจรกรองแบบปรับตัวไปใช้งานมีหลายแบบ [13][15] สำหรับในระบบ AEC จะใช้วงจรกรองแบบปรับตัวที่มีการใช้งานแบบการแสดงเอกลักษณ์ของระบบ ดังแสดงในรูปที่ 2.3



รูปที่ 2.3 การแสดงเอกลักษณ์ของระบบ (System Identification)  
ของวงจรกรองแบบปรับตัวสำหรับระบบ AEC

การแสดงเอกลักษณ์ของระบบเป็นการสร้างแบบจำลองของระบบที่ไม่ทราบค่าระบบหนึ่ง โดยอาศัยวงจรกรองแบบปรับตัว ซึ่งมีขั้นตอนต่าง ๆ ในการจำลองตามแบบเอกลักษณ์ดังนี้คือ

- (1) การออกแบบวางแผนการทดลอง
- (2) การเลือกโครงสร้างของแบบจำลอง
- (3) การประมาณค่าสัมประสิทธิ์
- (4) การพิสูจน์แบบจำลอง

ขั้นตอนและวิธีการต่าง ๆ เหล่านี้จะกระทำวนซ้ำไปเรื่อย ๆ จนกระทั่งได้แบบจำลองที่น่าพอใจ [15]

ในที่นี้จะอธิบายการใช่วงจรกรองแบบปรับตัวในการประมาณค่าสัมประสิทธิ์ของระบบที่ไม่ทราบค่า ระบบที่ไม่ทราบค่าในวิทยานิพนธ์ฉบับนี้คือ วิธีสะท้อนทางเสียง AEP ที่จำลองโดยผลตอบสนองต่ออิมพัลส์  $\underline{h}(n)$  ที่มีจำนวนจำกัดความยาว  $L$  และกำหนดให้เป็นระบบเชิงเส้นไม่เปลี่ยนแปลงตามเวลา (Linear Time Invariant System) คือ

$$\underline{h}(n)=[h_0(n), h_1(n), \dots, h_{L-1}(n)]^T$$

โดยที่  $[\cdot]^T$  คือ เมตริกซ์สลับเปลี่ยน (Transposition)

และขนาดวงจรกรองแบบปรับตัวถูกกำหนดให้มีความยาวเท่ากับจำนวนสัมประสิทธิ์ของ  $\underline{h}(n)$  โดยเวกเตอร์สัมประสิทธิ์ของวงจรกรองแบบปรับตัวหรือเรียกว่าสัมประสิทธิ์นำหน้าของวงจรกรองแบบปรับตัวคือ

$$\underline{\hat{h}}(n)=[\hat{h}_0(n), \hat{h}_1(n), \dots, \hat{h}_{L-1}(n)]^T$$

และเวกเตอร์สัญญาณเข้าของระบบที่เวลา  $n$  ใดๆ คือ

$$\underline{x}(n)=[x(n), x(n-1), \dots, x(n-L+1)]^T$$

สัญญาณไมโครโฟนในห้องรับ  $d(n)$  หาได้จาก

$$\begin{aligned} d(n) &= y(n) + v(n) \\ &= \sum_{k=0}^{L-1} \underline{h}_k(n) \underline{x}(n-k) + v(n) \end{aligned} \quad (2.1)$$

เมื่อ  $\sum_{k=0}^{L-1} \underline{h}_k(n) \underline{x}(n-k)$  เป็นผลรวมคอนโวลูชัน (Convolution) ระหว่าง  $\underline{h}(n)$  และ  $\underline{x}(n)$

นอกจากนี้ สัญญาณออกของวงจรกรองแบบปรับตัว  $\hat{y}(n)$  คำนวณได้จากผลรวมคอนโวลูชันระหว่างสัมประสิทธิ์นำหน้าของวงจรกรองแบบปรับตัว  $\hat{\underline{h}}_k(n)$  กับสัญญาณเข้า  $\underline{x}(n)$  ดังสมการที่ (2.2)

$$\hat{y}(n) = \sum_{k=0}^{L-1} \hat{\underline{h}}_k(n) \underline{x}(n-k) \quad (2.2)$$

ดังนั้นสัญญาณผิดพลาดของการประมาณ  $e(n)$  ซึ่งหาได้จากความแตกต่างระหว่าง สัญญาณออกของวงจรกรองแบบปรับตัว  $\hat{y}(n)$  กับสัญญาณไมโครโฟน  $d(n)$  ดังสมการ

$$\begin{aligned} e(n) &= y(n) + v(n) - \hat{y}(n) \\ &= \sum_{k=0}^{L-1} [\underline{h}_k(n) - \hat{\underline{h}}_k(n)] \underline{x}(n-k) + v(n) \end{aligned} \quad (2.3)$$

เมื่อพิจารณาสมการที่ (2.3) จะพบว่าที่เวลา  $n$  ใดๆ ถ้าวงจรกรองแบบปรับตัวสามารถจำลองระบบ  $\underline{h}(n)$  ได้ กล่าวคือ  $\hat{y}(n)$  มีค่าใกล้เคียงกับ  $y(n)$  ค่าสัญญาณผิดพลาด  $e(n)$  จะมีขนาดเข้าใกล้ค่าสัญญาณรบกวนพื้นหลัง  $v(n)$  ดังนั้นค่าสัญญาณผิดพลาดจึงถูกนำไปใช้

เป็นตัวบอกการเบี่ยงเบนของแบบจำลองวิธีสะท้อนทางเสียงเมื่อเปรียบเทียบกับ AEP ในห้องรับ สำหรับขั้นตอนวิธี (Algorithm) ต่าง ๆ ที่มีการเลือกใช้งาน ทั้งนี้ค่าสัญญาณผิดพลาด  $e(n)$  จะถูกนำไปใช้ในการควบคุมการเปลี่ยนแปลงสัมประสิทธิ์น้ำหนักของวงจรรองแบบปรับตัวที่คำนวณได้ในแต่ละรอบการวนซ้ำ (Iteration) เพื่อใช้หาค่าสัมประสิทธิ์น้ำหนักของวงจรรองแบบปรับตัวในการคำนวณรอบถัดไป เมื่อการปรับปรุงค่าสัมประสิทธิ์น้ำหนักของวงจรรองแบบปรับตัวกระทำซ้ำไปเรื่อยๆ แบบจำลองวิธีสะท้อนทางเสียงที่ใกล้เคียงมากที่สุดกับวิธีสะท้อนทางเสียงจริงในห้องรับ AEP หาได้จากสัมประสิทธิ์  $\hat{h}(n)$  เมื่อขนาดของสัญญาณความผิดพลาด  $e(n)$  มีค่าน้อยจนเป็นที่ยอมรับได้ [13]

สำหรับในกรณีที่ AEP มีการเปลี่ยนแปลงตามเวลา (Time-varying) สัญญาณออกของระบบจะเป็นแบบเปลี่ยนแปลงตามเวลาด้วย ดังนั้นขั้นตอนวิธีที่ใช้ควบคุมการทำงานของวงจรรองแบบปรับตัวนอกจากจะต้องมีความสามารถในการทำให้ขนาดของสัญญาณความผิดพลาดของระบบ มีค่าน้อย ๆ แล้ว ยังควรที่จะต้องมีความสามารถในการติดตาม (Tracking) การเปลี่ยนแปลงของ AEP อย่างต่อเนื่องตามเวลาอีกด้วย [17] ซึ่งจะอยู่นอกขอบเขตการศึกษาของวิทยานิพนธ์นี้

## 2.2 ขั้นตอนวิธีที่ใช้กับวงจรรองแบบปรับตัว

### 2.2.1 ขั้นตอนวิธี Least Mean Square (LMS)

ขั้นตอนวิธี LMS เป็นขั้นตอนวิธีที่นิยมและมีการใช้งานกันอย่างกว้างขวางสำหรับควบคุมการทำงานของวงจรรองแบบปรับตัว เนื่องจากเป็นขั้นตอนวิธีที่มีสมรรถนะที่ทนทาน (Robust Performance) และมีความซับซ้อนในการคำนวณต่ำ (Low Computational Complexity) ถึงแม้จะมีอัตราการลู่เข้า (Convergence Rate) ที่ค่อนข้างช้า [12][15-16] เมื่อเปรียบเทียบกับขั้นตอนวิธีอื่นๆ ขั้นตอนวิธีนี้พัฒนามาจากขั้นตอนวิธี Steepest Descent โดยพยายามลดค่าเฉลี่ยกำลังสองของค่าสัญญาณผิดพลาด  $e(n)$  ให้มีค่าน้อยที่สุด [15-16] ขั้นตอนวิธี LMS มีการทำงานที่เริ่มจากการกำหนดค่าเริ่มต้นให้แก่สัมประสิทธิ์ของวงจรรองแบบปรับตัวหรือ เวกเตอร์สัมประสิทธิ์น้ำหนัก (Tap Weight Vector) ของวงจรรองแบบปรับตัว โดยค่าสัมประสิทธิ์น้ำหนักของวงจรรองแบบปรับตัวจะถูกปรับปรุงไปพร้อมกับจำนวนรอบหรือจำนวนการวนซ้ำของการคำนวณ ค่าของเวกเตอร์สัมประสิทธิ์น้ำหนัก ที่ได้จากการวนซ้ำในรอบสุดท้ายจะลู่เข้าสู่ค่าตอบของระบบเรียกว่า Wiener Solution [15] การทำงานของขั้นตอนวิธี LMS นี้จะมีลักษณะคล้ายกับระบบการควบคุมแบบป้อนกลับ โดยอาศัยกระบวนการพื้นฐาน 2 กระบวนการคือ

1. Adaptive Process จะทำการปรับเปลี่ยนเวกเตอร์สัมประสิทธิ์น้ำหนักในแต่ละรอบการวนซ้ำ

2. Filtering Process จะเป็นการทำผลคูณภายในระหว่างชุดข้อมูลสัญญาณเข้าและชุดเวกเตอร์สัมประสิทธิ์น้ำหนักของวงจรรองแบบปรับตัวที่ผ่านกระบวนการปรับปรุงค่าสัมประสิทธิ์น้ำหนักแล้วจากในขั้นตอน Adaptive Process

ฟังก์ชันต้นทุน (Cost Function) ที่ใช้ในขั้นตอนวิธี LMS มีค่าเป็น ค่าเฉลี่ยกำลังสองของค่าสัญญาณผิดพลาดดังสมการ

$$J(n)=E\{e^2(n)\} \quad (2.4)$$

เมื่อ  $E\{\cdot\}$  คือ Expectation ของค่าสัญญาณ

พิจารณารูปที่ 2.3 สมมติให้  $\underline{x}(n)$  และ  $d(n)$  เป็นสัญญาณที่มีคุณลักษณะแบบ Jointly Stationary และ  $J(n)$  เป็นฟังก์ชันคองเวกซ์ (Convex Function) ของ  $\underline{\hat{h}}(n)$

ถ้าสมมติให้สัญญาณรบกวนพื้นหลังมีค่าน้อยมากจนเข้าใกล้ศูนย์ จากสมการที่ (2.3) ค่าสัญญาณผิดพลาดสามารถเขียนได้เป็น

$$e(n)=\left[\underline{h}(n)-\underline{\hat{h}}(n)\right]^T \underline{x}(n) \quad (2.5)$$

แทนค่า  $e(n)$  ลงในฟังก์ชันต้นทุนในสมการที่ (2.4) จะได้

$$J(n)=E\left\{\left[\left[\underline{h}(n)-\underline{\hat{h}}(n)\right]^T \underline{x}(n)\right]^2\right\} \quad (2.6)$$

ในการหาค่าสัมประสิทธิ์น้ำหนัก  $\underline{\hat{h}}(n)$  ค่าที่เหมาะสมที่สุด (Optimum) เมื่อ  $J(n)$  มีค่าต่ำสุด (Minimum) สามารถทำได้โดยการหาค่าเกรเดียนท์เวกเตอร์ของ  $J(n)$  ซึ่งจะมีค่าเป็น

$$\nabla(J(n))=-2\underline{p}(n)-2\mathbf{R}(n)\underline{\hat{h}}(n) \quad (2.7)$$

โดยที่  $\underline{p}(n)$  คือ เวกเตอร์สหสัมพันธ์ข้าม (Cross-correlation) ระหว่าง  $\underline{x}(n)$  และ  $d(n)$  สามารถหาได้จาก

$$\underline{p}(n)=E\{\underline{x}(n)d^*(n)\} \quad (2.8)$$

เมื่อ  $\mathbf{H}$  คือ Hermitian Transpose และ  $\mathbf{R}$  คือ เมตริกซ์อัตโนมัติสหสัมพันธ์ (Autocorrelation) ของ  $\underline{\mathbf{x}}(n)$  สามารถหาได้จาก

$$\mathbf{R}(n) = E\{\underline{\mathbf{x}}(n)\underline{\mathbf{x}}^H(n)\} \quad (2.9)$$

สมการปรับปรุงค่าสัมประสิทธิ์น้ำหนักของวงจรรองแบบปรับตัว  $\hat{\mathbf{h}}(n)$  หรือสมการปรับให้ทันกาล (Update Equation) ของระบบ สามารถหาได้จากการคำนวณแบบวนซ้ำๆ ดังต่อไปนี้

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \frac{1}{2}\mu[-\nabla(J(n))] \quad (2.10)$$

เมื่อ  $\mu$  คือ ค่าคงที่จำนวนจริงบวกที่ส่งผลต่ออัตราการปรับปรุงค่าสัมประสิทธิ์ของวงจรรองแบบปรับตัว และจะเรียกว่า ค่าช่วงก้าว (Step-size)

อย่างไรก็ตามค่าของ  $\underline{\mathbf{p}}(n)$  และ  $\mathbf{R}(n)$  ดังสมการที่ (2.8) และ (2.9) ไม่สามารถหาได้จากสัญญาณเข้าเพียงชุดเดียว ดังนั้น ในขั้นตอนวิธี LMS จึงปรับปรุงจากขั้นตอนวิธี Steepest Descent โดยการใช้ค่าประมาณ (Estimate) ของ  $\underline{\mathbf{p}}(n)$  และ  $\mathbf{R}(n)$  ขณะเวลาหนึ่ง (Instantaneous) ในการหาค่าเกรเดียนต์เวกเตอร์ในสมการที่ (2.7) โดยค่าประมาณขณะเวลาหนึ่งของ  $\mathbf{R}$  คือ

$$\hat{\mathbf{R}}(n) = \underline{\mathbf{x}}(n)\underline{\mathbf{x}}^T(n) \quad (2.11)$$

เนื่องจากสัญญาณที่พิจารณาเป็นสัญญาณค่าจริง (Real Value) จึงใช้  $\langle \cdot \rangle^T$  แทน  $\langle \cdot \rangle^H$  และค่าประมาณขณะเวลาหนึ่งของ  $\underline{\mathbf{p}}$  คือ

$$\hat{\underline{\mathbf{p}}}(n) = \underline{\mathbf{x}}(n)d(n) \quad (2.12)$$

เมื่อนำสมการที่ (2.11) และ (2.12) แทนในสมการที่ (2.7) และ (2.5) สมการปรับให้ทันกาลของ  $\hat{\mathbf{h}}(n)$  ในสมการที่ (2.10) จะกลายเป็น

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \mu \underline{\mathbf{x}}(n) [d(n) - \hat{\mathbf{h}}^T(n)\underline{\mathbf{x}}(n)] \quad (2.13)$$

ดังนั้นเมื่อนำขั้นตอนวิธี LMS มาใช้ในการควบคุมการทำงานของวงจรรองแบบปรับตัวตามรูปที่ 2.3 จะได้ความสัมพันธ์ต่าง ๆ ที่จำเป็นคือ

$$\hat{y}(n) = \hat{\mathbf{h}}^T(n) \mathbf{x}(n) \quad (2.14)$$

$$\mathbf{e}(n) = d(n) - \hat{y}(n) \quad (2.15)$$

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \mu \mathbf{x}(n) \mathbf{e}(n) \quad (2.16)$$

ค่าช่วงก้าว  $\mu$  จะมีผลโดยตรงต่ออัตราการลู่เข้าสู่สถานะคงตัว (Steady State) ของระบบ เนื่องจากการป้อนกลับ (Feedback) ในขั้นตอนวิธี LMS โดยอาจมีผลทำให้ระบบไม่เสถียรได้ ทั้งนี้ค่า  $\mu$  ควรเลือกให้อยู่ในช่วงของ [12][16]

$$0 < \mu < \frac{2}{\lambda_{\max}} \quad (2.17)$$

เพื่อรักษาความเสถียรของระบบ โดย  $\lambda_{\max}$  คือ ค่าเจาะจง (Eigenvalue) ของ  $\mathbf{R}$  ที่มีค่ามากที่สุด โดยปกติ  $\hat{\mathbf{h}}(n)$  จะไม่ลู่เข้าสู่จุดที่น้อยที่สุดของ  $J(n)$  แต่จะแกว่งรอบ ๆ จุดที่น้อยที่สุดนี้ เนื่องจากผลของการมี Gradient Noise [15] กล่าวคือ ค่าผิดพลาดในการประมาณค่าเกรเดียนท์เวกเตอร์ผ่าน  $\mathbf{R}(n)$  และ  $\hat{\mathbf{p}}(n)$  เป็นผลให้ฟังก์ชันต้นทุนในสภาวะอยู่ตัว หรือ  $J(\alpha)$  มีค่ามากกว่าค่าเฉลี่ยกำลังสองของสัญญาณความผิดพลาดที่มีค่าน้อยสุด ( $J_{\min}$ ) ความแตกต่างระหว่าง  $J(\alpha)$  และ  $J_{\min}$  จะเรียกว่า Excess Mean Square Error หรือ  $J_{\text{ex}}(\alpha)$  กล่าวคือ

$$J_{\text{ex}}(\alpha) = J(\alpha) - J_{\min} \quad (2.18)$$

การวัดผลการเบี่ยงเบนของคำตอบสามารถหาได้จากอัตราส่วนระหว่าง  $J_{\text{ex}}(\alpha)$  และ  $J_{\min}$  โดยเรียกว่าค่า Misadjustment

$$\text{Misadjustment (M)} = \frac{J_{\text{ex}}(\alpha)}{J_{\min}} \quad (2.19)$$

สำหรับอัตราการลู่เข้าของ  $J(n)$  สูตรคำตอบของ Wiener Solution (เมื่อ  $\hat{\mathbf{h}}_{\text{opt}} = \mathbf{R}^{-1}(n) \mathbf{p}(n)$ ) จะขึ้นอยู่กับการกระจายของค่าเจาะจง (Eigenvalue Spread) ของ  $\mathbf{R}$  ซึ่งหาจาก  $\lambda_{\max} / \lambda_{\min}$  ถ้า



การกระจายของค่าเจาะจงของ  $\mathbf{R}$  มีค่าสูง ( $>1$ ) จะทำให้อัตราการลู่เข้าช้าลง ในกรณีสัญญาณเข้าเป็นสัญญาณรบกวนขาวจะมีค่าเจาะจงของ  $\mathbf{R}$  เท่ากับ 1 หากแต่ถ้าในระบบ AEC สัญญาณเข้าส่วนมากมักจะเป็นสัญญาณเสียงพูดซึ่งมีการกระจายของค่าเจาะจงของ  $\mathbf{R}$  สูงทำให้อัตราการลู่เข้าช้ากว่าเมื่อเทียบกับกรณีที่สัญญาณเข้าของระบบเป็นสัญญาณรบกวนขาว ดังนั้นในการเลือกขั้นตอนวิธีที่จะนำมาใช้ในการหาคำตอบของระบบ AEC เพื่อให้มีอัตราการลู่เข้าที่ดี จึงสามารถปรับผลของสัญญาณเข้าให้น้อยลงได้จากการเลือกใช้ขั้นตอนวิธี Normalized Least Mean Square (NLMS) ดังจะกล่าวในหัวข้อถัดไป

## 2.2.2 ขั้นตอนวิธี Normalized Least Mean Square (NLMS)

ดังได้กล่าวไว้ในหัวข้อที่ 2.2.1 ถึงขั้นตอนวิธี LMS แล้วนั้น ปัจจัยที่มีความสัมพันธ์กับเวกเตอร์สัมประสิทธิ์ของวงจรรองแบบปรับตัว  $\hat{\mathbf{h}}(n)$  ถูกคำนวณจากค่าประมาณของเกเรเดียนท์  $\mathbf{x}(n)\mathbf{e}^*(n)$  เนื่องจากปริมาณนี้เป็นสัดส่วนโดยตรงกับเวกเตอร์สัญญาณเข้า  $\mathbf{x}(n)$  ค่าประมาณของเกเรเดียนท์จะมากขึ้นเมื่อสัญญาณเข้า  $\mathbf{x}(n)$  มีขนาดใหญ่ขึ้น ซึ่งในระบบ AEC จะพบปัญหาดังกล่าว เนื่องจากสัญญาณเข้าของระบบเป็นสัญญาณเสียงพูดที่ถือว่าการกระจายของค่าเจาะจงของ  $\mathbf{R}$  มีค่าสูง แต่ปัญหานี้สามารถหลีกเลี่ยงได้โดยการทำนอร์มัลไลซ์ (Normalize) ค่าประมาณเกเรเดียนท์ด้วยขนาดของเวกเตอร์สัญญาณเข้า  $\|\mathbf{x}(n)\|^2$  เมื่อ  $\|\cdot\|$  คือ Euclidean Norm [15] การปรับปรุงค่าสัมประสิทธิ์นำหนักของวงจรรองแบบปรับตัว  $\hat{\mathbf{h}}(n)$  สามารถกระทำได้โดยขั้นตอนดังนี้

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \frac{\tilde{\mu}\mathbf{x}(n)\mathbf{e}(n)}{\varepsilon + \|\mathbf{x}(n)\|^2} \quad (2.20)$$

เมื่อให้  $\varepsilon$  คือ ค่าคงที่เล็ก ๆ ที่บวกเพิ่มในสมการ เพื่อป้องกันการลู่ออกของวงจรรองแบบปรับตัว ในกรณีที่สัญญาณเข้ามีขนาดน้อย ๆ และค่าช่วงก้าว  $\tilde{\mu} \in (0,2)$  จะอยู่ภายใต้เงื่อนไขที่จำเป็นและพอเพียงสำหรับการลู่เข้าสู่คำตอบของขั้นตอนวิธี NLMS เมื่อทำการนอร์มัลไลซ์ค่าประมาณเกเรเดียนท์ด้วยขนาดของสัญญาณเข้าแล้ว ค่าช่วงก้าว  $\tilde{\mu}$  จะเป็นอิสระต่อลักษณะเฉพาะ (Characteristic) ของสัญญาณเข้า [15]

สำหรับขั้นตอนวิธี LMS ถ้าเราเลือกใช้ค่า  $\mu$  น้อยๆ จะทำให้การเบี่ยงเบนของคำตอบน้อย หรือความผิดพลาดของระบบน้อย แต่อัตราการลู่เข้าสู่คำตอบจะช้า ในทางตรงกันข้ามถ้าเราใช้ค่า  $\mu$  มากจะทำให้อัตราการลู่เข้าสู่คำตอบของระบบเร็วขึ้นแต่การเบี่ยงเบนของคำตอบก็จะ

มากขึ้นด้วย สำหรับค่า  $\mu$  ที่คงที่ค่าหนึ่งๆ จะมีความเหมาะสมสำหรับสัญญาณเข้าชุดหนึ่ง แต่ อาจจะไม่ใช่ว่าค่าที่เหมาะสมกับสัญญาณเข้าอีกชุดหนึ่ง กล่าวคือ อาจส่งผลให้ระบบเกิดการลู่ออกได้ เมื่อใช้ขั้นตอนวิธี NLMS พบว่าค่าช่วงก้าว  $\mu(n)$  จะมีการเปลี่ยนแปลงตามเวลาหรือ เปลี่ยนไปตามสัญญาณเข้า ดังสมการที่ (2.21)

$$\mu(n) = \frac{\tilde{\mu}}{\varepsilon + \|\underline{x}(n)\|^2} \quad (2.21)$$

ซึ่งแตกต่างจากการใช้ค่า  $\mu$  ที่เป็นค่าคงที่ในขั้นตอนวิธี LMS ค่า  $\mu(n)$  ที่เปลี่ยนแปลงตามเวลา จะมีความยืดหยุ่นเกี่ยวกับอัตราการลู่เข้าและการเบี่ยงเบนของคำตอบ ดังนั้นการกำจัดสัญญาณเสียงสะท้อนโดยใช้ขั้นตอนวิธี NLMS จะมีอัตราการลู่เข้าที่เร็วกว่าขั้นตอนวิธี LMS เมื่อสัญญาณเข้าเป็นสัญญาณเสียงพูด หรืออาจจะกล่าวได้ว่าขั้นตอนวิธี NLMS มีความเหมาะสมในระบบ AEC เมื่อสัญญาณเข้าเป็นสัญญาณเสียงพูดมากกว่าการใช้ขั้นตอนวิธี LMS

สำหรับการเปรียบเทียบความซับซ้อนในการคำนวณของขั้นตอนวิธี LMS และขั้นตอนวิธี NLMS สามารถคำนวณจากตัวดำเนินการ ในแต่ละรอบของการวนซ้ำของขั้นตอนวิธีดังตารางข้างล่างนี้ โดยจะกำหนดให้  $L$  คือความยาวของวงจรรองแบบปรับตัว

LMS algorithm	+ / -	x
$\hat{y}(n) = \hat{h}^T(n)x(n)$	$L - 1$	$L$
$e(n) = d(n) - \hat{y}(n)$	$1$	$-$
$\hat{h}(n+1) = \hat{h}(n) + \mu x(n)e(n)$	$L$	$L + 1$
Total computation complexity	$2L$	$2L + 1$

ตารางที่ 2.1 ความซับซ้อนในการคำนวณของขั้นตอนวิธี LMS

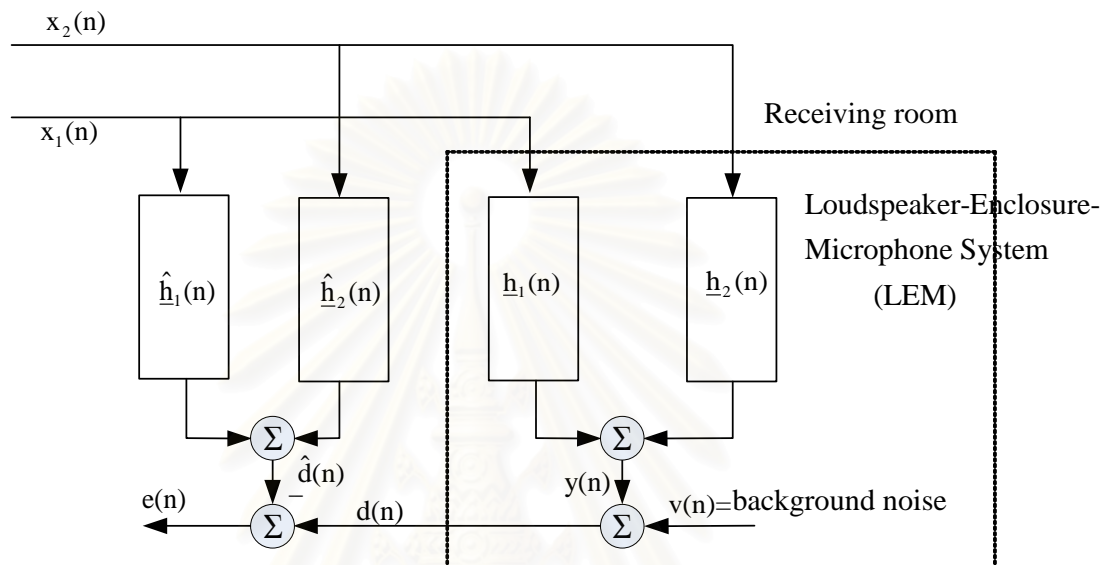
NLMS algorithm	+ / -	x
$e(n) = d(n) - \hat{h}^T(n)\underline{x}(n)$	L	L
$\hat{h}(n+1) = \hat{h}(n) + \frac{\tilde{\mu}\underline{x}(n)e(n)}{\varepsilon + \ \underline{x}(n)\ ^2}$	L + 1	L + 3
Total computation complexity	2 L + 1	2 L + 3

### ตารางที่ 2.2 ความซับซ้อนในการคำนวณของขั้นตอนวิธี NLMS

การเปรียบเทียบความซับซ้อนในการคำนวณของขั้นตอนวิธี LMS และ NLMS พบว่า ความซับซ้อนของขั้นตอนวิธีทั้งตัวดำเนินการบวก และตัวดำเนินการคูณนั้นใกล้เคียงกัน หรือพูดอีกอย่างหนึ่งว่า ความซับซ้อนในการคำนวณของขั้นตอนวิธี NLMS ที่เพิ่มขึ้นจากขั้นตอนวิธี LMS มีขนาดน้อยมาก ดังนั้นในการเลือกใช้ขั้นตอนวิธีในระบบการกำจัดสัญญาณเสียงสะท้อนนั้นก็ สามารถเลือกทั้งสองขั้นตอนวิธีในการใช้งาน แต่อย่างไรก็ตามในการพิจารณานั้นต้องอาศัยการ พิจารณาลักษณะเฉพาะของสัญญาณเข้าของระบบเป็นส่วนประกอบสำคัญ

### 2.3 การแสดงเอกลักษณ์ของระบบ ของวงจรรองแบบปรับตัว สำหรับระบบ SAEC

ขั้นตอนวิธีต่างๆ สำหรับวงจรรองแบบปรับตัวที่ใช้ในระบบ AEC สามารถนำมาประยุกต์ใช้ในระบบการกำจัดสัญญาณเสียงสะท้อนแบบสเตอริโอ (SAEC) โดยมีการขยายระบบ AEC ให้เป็นระบบ SAEC ด้วยการเพิ่มวงจรรองแบบปรับตัวเป็น 2 ตัว ดังรูปที่ 2.4



รูปที่ 2.4 แบบจำลองระบบ SAEC โดยใช้วงจรรองแบบปรับตัวจำลอง AEP

ในระบบ SAEC นี้ สัญญาณเสียงสะท้อนทางด้านห้องรับหาได้จาก

$$y(n) = \underline{h}_1^T(n) \underline{x}_1(n) + \underline{h}_2^T(n) \underline{x}_2(n) \quad (2.22)$$

เมื่อ

$$\underline{x}_i(n) = [x_i(n) \quad x_i(n-1) \quad \dots \quad x_i(n-L+1)]^T$$

$$\underline{h}_i(n) = [h_{i,0}(n) \quad h_{i,1}(n) \quad \dots \quad h_{i,L-1}(n)]^T$$

โดยที่  $i=1, 2$

สัญญาณที่ได้จากการจำลองวิธีการสะท้อนทางเสียง  $\hat{h}_i(n)$  โดยใช้วงจรรองแบบปรับตัวทั้ง 2 ตัวมีค่าเป็น  $\hat{y}(n)$

$$\hat{y}(n) = \hat{\mathbf{h}}_1^T(n) \mathbf{x}_1(n) + \hat{\mathbf{h}}_2^T(n) \mathbf{x}_2(n) \quad (2.23)$$

เมื่อ  $\hat{\mathbf{h}}_i(n) = [h_{i,0}(n) \ h_{i,1}(n) \ \dots \ h_{i,L-1}(n)]^T$

ค่าผิดพลาดที่เกิดขึ้นจากการใช้วงจรรองแบบปรับตัวในการประมาณค่าสัญญาณเสียงสะท้อนหาได้จาก

$$\mathbf{e}(n) = y(n) - [\hat{\mathbf{h}}_1^T(n) \mathbf{x}_2(n) + \hat{\mathbf{h}}_2^T(n) \mathbf{x}_2(n)] \quad (2.24)$$

เมื่อใช้ขั้นตอนวิธี NLMS ในการปรับปรุงการทำงานของวงจรรองแบบปรับตัวในระบบ SAEC สมการปรับให้ทันกาลของระบบเป็นดังนี้

$$\hat{\mathbf{h}}_i(n+1) = \hat{\mathbf{h}}_i(n) + \frac{\mu}{\varepsilon + \|\mathbf{x}_i(n)\|^2} \mathbf{x}_i(n) \mathbf{e}(n) \quad (2.25)$$

เมื่อให้  $\varepsilon$  คือ ค่าคงที่น้อย ๆ เพื่อป้องกันการสูญเสียเสถียรภาพของระบบในกรณีที่เวกเตอร์สัญญาณมีขนาดน้อย ๆ

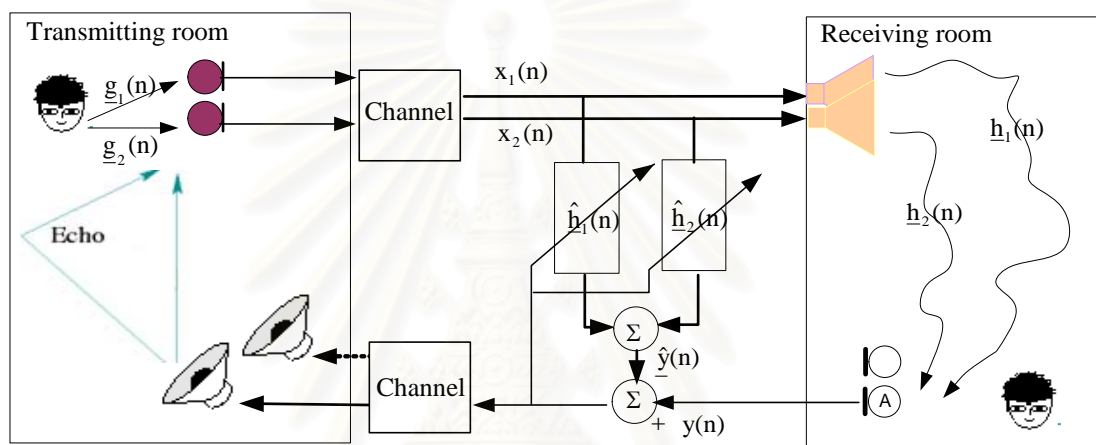
จากที่ได้กล่าวในบทที่ 1 แล้วว่า การกำจัดสัญญาณเสียงสะท้อนโดยการใช้วงจรรองแบบปรับตัวในระบบ SAEC นั้นค่อนข้างทำได้ยากกว่าในระบบ AEC เนื่องจากสัญญาณเข้าของระบบสเตอริโอ  $\mathbf{x}_1(n)$  และ  $\mathbf{x}_2(n)$  ในแบบจำลองดังกล่าวจะมีสหสัมพันธ์กันอยู่ [1-4] ซึ่งจะส่งผลต่อการปรับค่าสัมประสิทธิ์น้ำหนักวงจรรองแบบปรับตัวทั้งสองให้มีลักษณะใกล้เคียงกับค่าวิถีสะท้อนทางเสียงแต่ละชุดของช่องสัญญาณดังจะได้กล่าวถึงในหัวข้อถัดไป

สถาบันวิทยบริการ  
จุฬาลงกรณ์มหาวิทยาลัย

## 2.4 ปัญหา “Non-uniqueness Problem” ในระบบ SAEC

จากการจำลองระบบการกำจัดสัญญาณเสียงสะท้อนแบบสเตอริโอในรูปที่ 2.5 เมื่อสัญญาณเข้าของระบบเป็นสัญญาณเชิงเส้นไม่เปลี่ยนแปลงตามเวลา (Linear Time Invariant, LTI) จะได้ความสัมพันธ์ของผลตอบสนองต่ออิมพัลส์ของสัญญาณเข้าในห้องรับเป็น [2]

$$\underline{x}_1^T(n)\underline{g}_2 = \underline{x}_2^T(n)\underline{g}_1 \quad (2.26)$$



รูปที่ 2.5 บล็อกไดอะแกรมของระบบ SAEC

เมื่อให้ผลตอบสนองต่ออิมพัลส์ของสัญญาณเข้าในห้องส่งมีขนาดความยาว  $M$  และเป็นระบบเชิงเส้นไม่เปลี่ยนแปลงตามเวลาดังนั้น

$$\underline{g}_i = [g_{i1} \ g_{i2} \ \dots \ g_{iM}]^T$$

เมื่อ  $i=1, 2$  และสัญญาณเข้าที่ผ่านไมโครโฟนในห้องส่งมีค่าเป็น

จากสมการที่ (2.6) เมื่อเราต้องการวิเคราะห์ห้ระบบเพื่อให้ระบบมีค่าผิดพลาดต่ำสุด โดยใช้สมการของ Wiener Solution จะพบว่า

$$J(n) = \sum_{p=1}^n \lambda^{n-p} [y(p) - \hat{h}_1^T(n)\underline{x}_1(p) - \hat{h}_2^T(n)\underline{x}_2(p)]^2 \quad (2.27)$$

เมื่อให้  $0 < \lambda \leq 1$  เป็นค่า Exponential Forgetting Factor ส่งผลให้ได้สมการความสัมพันธ์ของ ฟังก์ชันอัตโนมัติสหสัมพันธ์ (Autocorrelation) และฟังก์ชันสหสัมพันธ์ข้าม (Cross-correlation) ของ สัญญาณเข้าของระบบเป็นดังนี้

$$\mathbf{R}(n) \begin{bmatrix} \hat{\mathbf{h}}_1(n) \\ \hat{\mathbf{h}}_2(n) \end{bmatrix} = \mathbf{r}(n) \rightarrow \begin{bmatrix} \hat{\mathbf{h}}_1(n) \\ \hat{\mathbf{h}}_2(n) \end{bmatrix} = \mathbf{R}^{-1}(n) \mathbf{r}(n) \quad (2.28)$$

เมื่อ  $\mathbf{R}(n)$  คือ เมตริกซ์ Covariance หรือเมตริกซ์อัตโนมัติสหสัมพันธ์ของสัญญาณเข้า  
 $\mathbf{r}(n)$  คือ ฟังก์ชันสหสัมพันธ์ข้ามระหว่างสัญญาณเข้าและสัญญาณออกจากระบบ  
 และ

$$\mathbf{R}^{-1}(n) = \sum_{p=1}^n \lambda^{n-p} \begin{bmatrix} \mathbf{x}_1(p) \\ \mathbf{x}_2(p) \end{bmatrix} \begin{bmatrix} \mathbf{x}_1^T(p) & \mathbf{x}_2^T(p) \end{bmatrix} = \begin{bmatrix} \mathbf{R}_{x_1x_1}(n) & \mathbf{R}_{x_1x_2}(n) \\ \mathbf{R}_{x_2x_1}(n) & \mathbf{R}_{x_2x_2}(n) \end{bmatrix} \quad (2.29)$$

$$\mathbf{r}(n) = \sum_{p=1}^n \lambda^{n-p} y(p) \begin{bmatrix} \mathbf{x}_1(p) \\ \mathbf{x}_2(p) \end{bmatrix} \quad (2.30)$$

ดังนั้น เพื่อให้ได้คำตอบของวงจรรองแบบปรับตัวให้มีความใกล้เคียงกับจุดต่ำสุดของ  $J(n)$  โดยให้

$$\mathbf{u} = \begin{bmatrix} \mathbf{g}_2 \\ \mathbf{g}_1 \end{bmatrix} \quad (2.31)$$

และเมื่อพิจารณาจากสมการที่ (2.28) จะได้ว่า

$$\mathbf{R}(n) \mathbf{u} = \mathbf{0} \quad (2.32)$$

จากสมการที่ (2.28) พบว่าการหาค่าเวกเตอร์สัมประสิทธิ์วงจรรองแบบปรับตัว ขึ้นอยู่กับเมตริกซ์ผกผันของ  $\mathbf{R}(n)$  ซึ่งถ้าเมตริกซ์  $\mathbf{R}(n)$  ไม่มีเมตริกซ์ผกผัน (Invertible) คำตอบสมการที่ได้จะไม่มีผลเฉลยเดียว อย่างไรก็ตามในระบบจริงสามารถหาเมตริกซ์ผกผันของ  $\mathbf{R}(n)$  ได้ ดังนั้นจึงสามารถหาผลเฉลยของสมการได้เช่นกัน [2] เมื่อพิจารณาสมการที่ (2.29) จะพบความซ้ำซ้อนในการทำงานสูงในการหาเมตริกซ์ผกผันของ  $\mathbf{R}(n)$  เนื่องจากผลของสหสัมพันธ์ของสัญญาณเสียงสเตอริโอในระบบ ที่ทำให้เมตริกซ์  $\mathbf{R}^{-1}(n)$  มีค่าเมตริกซ์ย่อย  $\mathbf{R}_{x_1x_2}(n)$  และ

$R_{x2x1}(n)$  อยู่ด้วย จึงส่งผลให้ระบบหาค่าเวกเตอร์สัมประสิทธิ์วงจรรองปรับตัวได้ค่อนข้างยาก เราเรียกเมตริกซ์ที่มีคุณลักษณะดังกล่าวว่า “illed Condition” [2] ดังนั้นจึงกล่าวได้ว่า คำตอบของสมการที่ (2.28) ขึ้นอยู่กับการกระจายของค่าเจาะจงของเมตริกซ์  $R(n)$  ซึ่งถ้าเป็นระบบ AEC การกระจายของค่าเจาะจงของ  $R(n)$  จะขึ้นอยู่กับสัญญาณเข้า  $x(n)$  เพียงช่องเดียว ในขณะที่ถ้าเป็นระบบ SAEC การกระจายของค่าเจาะจงของ  $R(n)$  ดังกล่าวจะขึ้นอยู่กับสัญญาณเข้าสเตอริโอทั้ง 2 ช่องสัญญาณที่มีสหสัมพันธ์กัน ดังนั้นจึงส่งผลให้คำตอบของสมการที่ (2.29) หาค่าได้ยากนั่นเอง [2-3] จากเหตุผลดังกล่าวได้ส่งผลไปถึงค่าผิดพลาดที่น้อยที่สุดของระบบด้วย กล่าวคือระบบอาจจะมีค่าผิดพลาดที่น้อยที่สุดในขณะที่วงจรรองแบบปรับตัวยังไม่สามารถจะปรับค่าเวกเตอร์สัมประสิทธิ์นำหน้าวงจรรองแบบปรับตัวให้ใกล้เคียงกับวิธีสะท้อนทางเสียงจริงได้ เรียกปัญหาดังกล่าวว่า “Misalignment Problem” [1-4] หาได้จาก

$$\epsilon = \frac{\|\underline{h} - \hat{\underline{h}}\|}{\|\underline{h}\|} \quad (2.33)$$

เมื่อ  $\underline{h} = [\underline{h}_{1,L}^T \quad \underline{h}_{2,L}^T]^T$  และ  $\hat{\underline{h}} = [\hat{\underline{h}}_{1,L}^T \quad \hat{\underline{h}}_{2,L}^T]^T$

ปัญหาดังกล่าวจะไม่เกิดขึ้นในระบบ AEC ดังนั้น เพื่อกำจัดสัญญาณเสียงสะท้อนที่เกิดขึ้นในระบบสเตอริโอให้ได้ผลดีที่สุด การดีคอดิเลชันสัญญาณเข้าแบบสเตอริโอให้ลดน้อยลงจนถึงน้อยที่สุด จึงเป็นวิธีหนึ่งที่มีการเสนอขึ้นใช้กันแพร่หลาย [2][5] เพื่อให้การปรับค่าเวกเตอร์สัมประสิทธิ์นำหน้าวงจรรองแบบปรับตัวทั้ง 2 ชุดนั้น เป็นอิสระต่อกันให้มากที่สุด จะได้ส่งผลให้สัญญาณเสียงสะท้อนที่ประมาณได้มีความใกล้เคียงกับสภาพความเป็นจริงมากที่สุดนั่นเอง

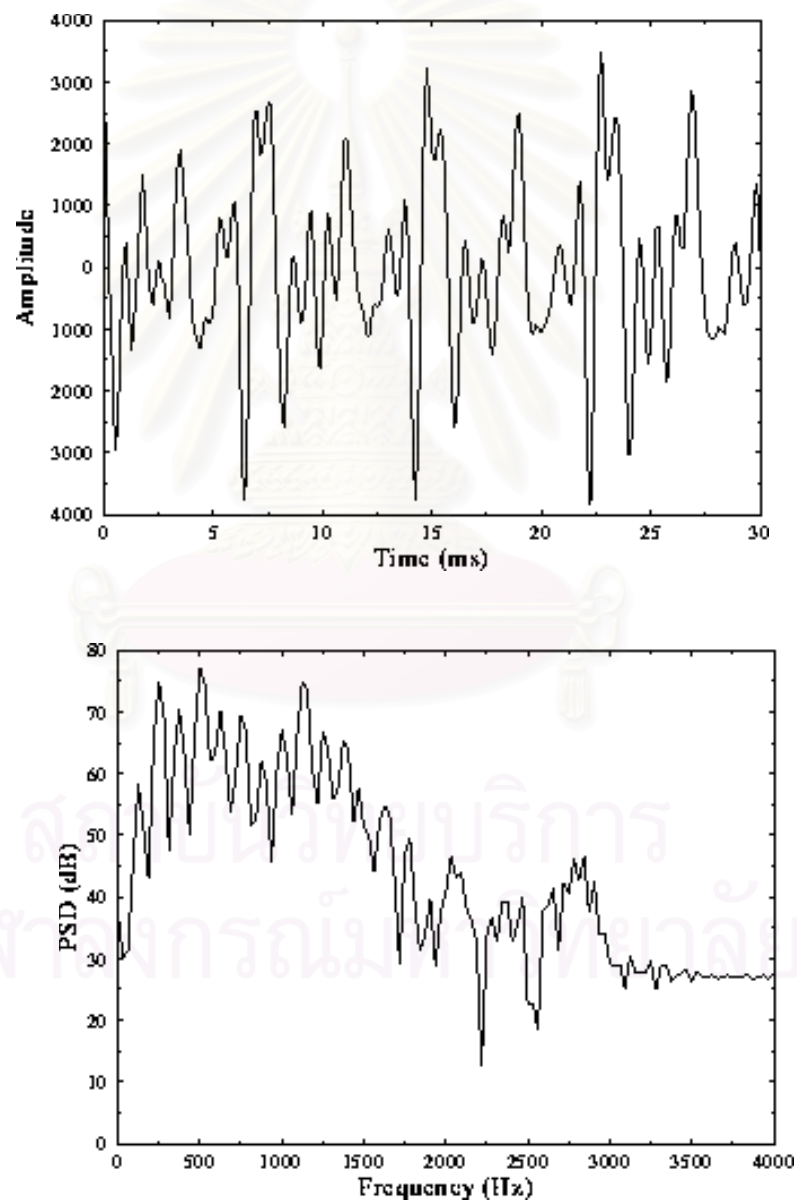
## 2.5 ลักษณะพื้นฐานของสัญญาณเสียงพูด

เสียงพูดเกิดขึ้นจากการที่ลมถูกขับออกจากปอด (Lung) ผ่านช่องว่างระหว่างเส้นเสียง (Vocal Cord/Vocal Fold) ไปยังหลอดลม (Larynx) ส่งผ่านไปยังโพรงคอ (Pharynx) และปล่อยออกสู่ช่องปากหรือจมูกตามการกำหนดของช่องกั้นระหว่างปากกับโพรงจมูก (Velum) ทั้งนี้ความยาวของช่องทางเดินเสียงตั้งแต่ปอดจนถึงช่องปากหรือจมูกคิดเป็นประมาณ 17 เซนติเมตรในเพศชายวัยผู้ใหญ่ [20-21] สัญญาณเสียงพูดที่เปล่งออกมาสามารถแบ่งได้เป็นหลายลักษณะ ขึ้นอยู่กับว่าจะใช้เกณฑ์ใดในการแบ่งประเภทและลักษณะเสียงที่เปล่งออกมามีดังต่อไปนี้ สำหรับเกณฑ์ที่



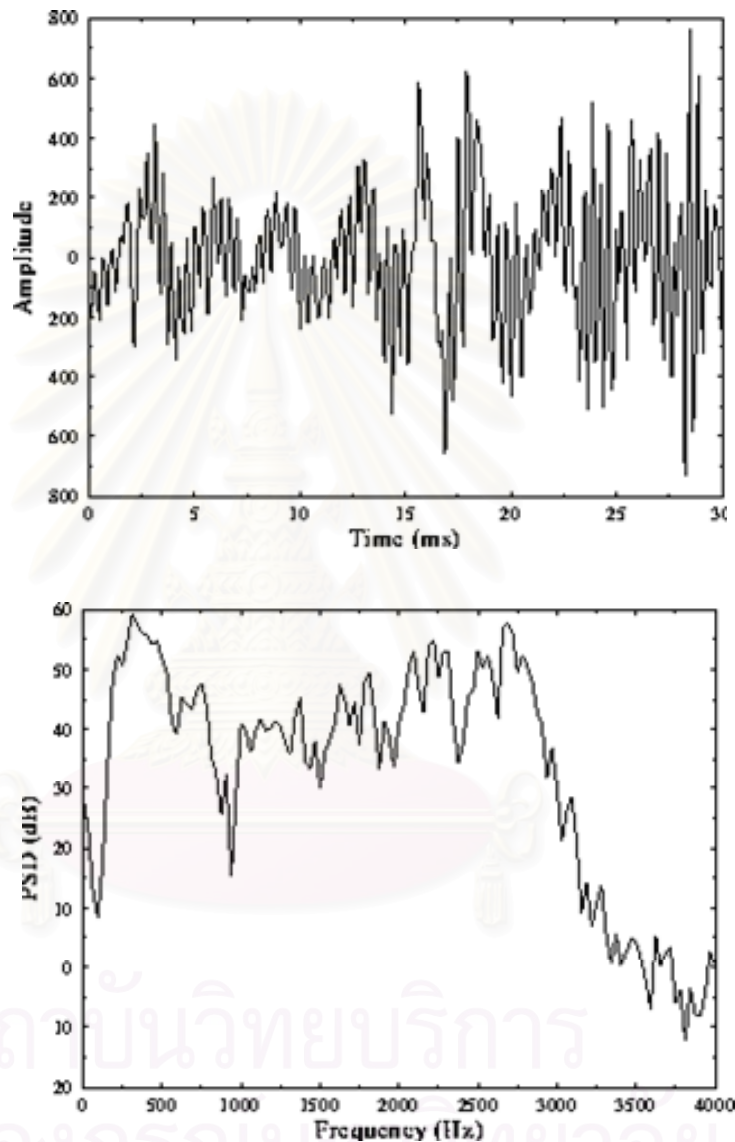
ใช้กันมากในการแบ่งลักษณะสัญญาณเสียงพูดคือแบ่งตามการกำหนดของเส้นเสียง โดยแบ่งออกเป็น 2 ประเภท

1. **เสียงก้อง (Voiced Sound)** เกิดจากการบังคับเส้นเสียงให้ปิดในขณะที่ลมจากปอดถูกปล่อยผ่านออกมาเป็นช่วงๆ ช่วงละประมาณ 2 ถึง 20 มิลลิวินาที เช่น เสียงพยัญชนะในภาษาอังกฤษตัว A, M , O หรือ I เป็นต้น โดยมีข้อสังเกตอย่างง่ายคือเมื่อใช้มือจับบริเวณหลอดลมซึ่งมีเส้นเสียงอยู่ในขณะพูดจะเกิดการสั่นขึ้นจนรู้สึกได้ ลักษณะตัวอย่างของเสียงแบบ Voiced Sound สามารถแสดงได้ดังรูปที่ 2.6



รูปที่ 2.6 ลักษณะของเสียงและพลังงานของเสียงแบบ Voiced Sound [20]

2. **เสียงไม่ก้อง (Unvoiced Sound)** คือเสียงเกิดขึ้นจากการปล่อยลมออกจากปอดโดยไม่มี การปิดของเส้นเสียง เช่น เสียงพยัญชนะในภาษาอังกฤษตัว S, C หรือ P เมื่อใช้มือจับบริเวณหลอดลมขณะพูดเสียงประเภทนี้จะไม่เกิดการสั่นจนรู้สึกได้ในบริเวณดังกล่าว ลักษณะตัวอย่างเสียงแบบ Unvoiced Sound แสดงได้ดังรูปที่ 27.



รูปที่ 2. 7 ลักษณะของเสียงและพลังงานของเสียงแบบ Unvoiced Sound [20]

การจัดประเภทของเสียงทั้งสองลักษณะนี้ ขึ้นอยู่กับการสั่นสะเทือนของเส้นเสียงในการออกเสียงแต่ละครั้ง ทั้งนี้การทราบถึงลักษณะและประเภทของเสียงจะมีประโยชน์อย่างมากต่อการวิเคราะห์และประมวลสัญญาณเสียงในระบบการสื่อสารสัญญาณทางเสียง เช่น การตรวจสอบรูปคลื่นพลังงานของช่วงสัญญาณเสียงที่แตกต่างกันก็สามารถทราบถึงลักษณะของสัญญาณเสียงที่

แตกต่างกัน นอกจากนี้ การเข้ารหัสเสียงก็ต้องอาศัยความรู้เกี่ยวกับลักษณะของสัญญาณเสียง เพื่อทำการเข้ารหัสของเสียงเพื่อให้ได้คุณภาพที่ดีด้วย



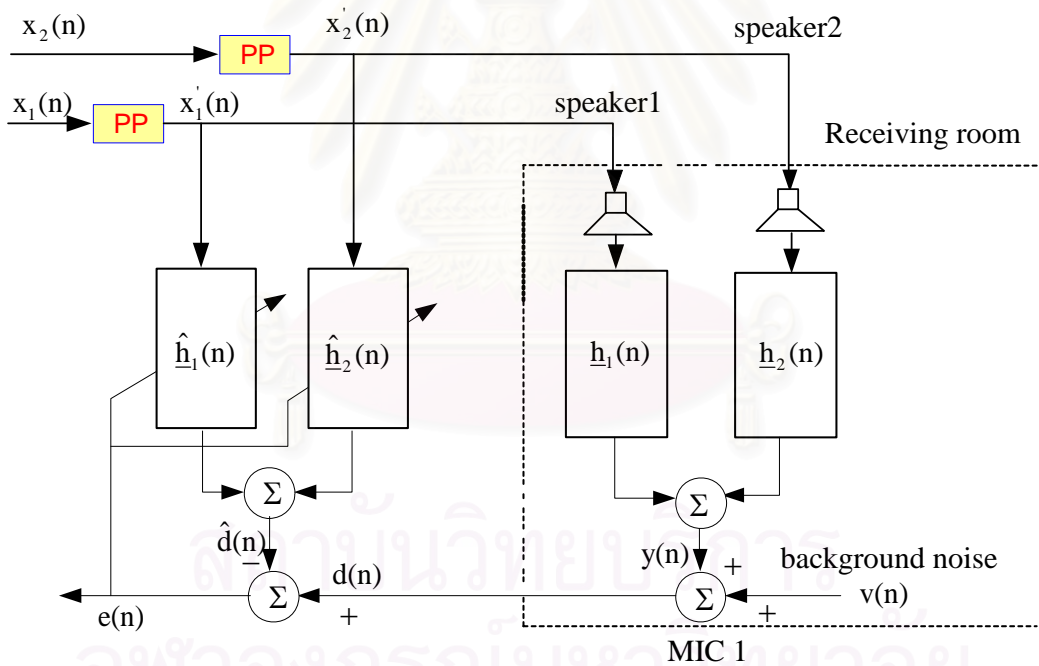
สถาบันวิทยบริการ  
จุฬาลงกรณ์มหาวิทยาลัย

### บทที่ 3

## การเพิ่มประสิทธิภาพในระบบการกำจัดสัญญาณเสียงสะท้อนแบบสเตอริโอ

### 3.1 วิธีการดีคორิเลชันสัญญาณเสียงสเตอริโอ

วิธีหนึ่งในการเพิ่มประสิทธิภาพการทำงานของระบบกำจัดสัญญาณเสียงสะท้อนแบบสเตอริโอ (SAEC) คือ การดีคอริเลชันสัญญาณเสียงสเตอริโอของระบบ SAEC ให้มีสหสัมพันธ์กันน้อยลง โดยจะแทนวิธีการดีคอริเลชันสัญญาณเสียงสเตอริโอด้วยส่วนประมวลผลก่อน (Pre-processing, PP) จากนั้นจึงทำการป้อนสัญญาณเสียงสเตอริโอที่ผ่านการประมวลผลก่อนเข้าสู่ระบบ SAEC ดังบล็อกไดอะแกรมของระบบ SAEC ที่ปรับปรุงใหม่ ดังแสดงในรูปที่ 3.1



รูปที่ 3.1 ระบบ SAEC ที่มีการดีคอริเลชันสัญญาณเสียงสเตอริโอโดยใช้ส่วนประมวลผลก่อน

ทั้งนี้วิธีการดีคอริเลชันสัญญาณเสียงสเตอริโอสามารถทำได้หลายวิธี [1] แต่ละวิธีมีข้อดีและข้อเสียที่แตกต่างกัน ในวิทยานิพนธ์นี้ได้เลือกศึกษาสามวิธี ได้แก่ วิธีที่หนึ่งการดีคอริเลชันสัญญาณเสียงสเตอริโอด้วยการเพิ่มสัญญาณรบกวนสุ่มทุกช่วงเวลาในแต่ละช่องสัญญาณ [1] เพื่อลดสหสัมพันธ์ของสัญญาณทั้งสอง และทำการเปรียบเทียบผลการกำจัดสัญญาณเสียงสะท้อนที่เกิด

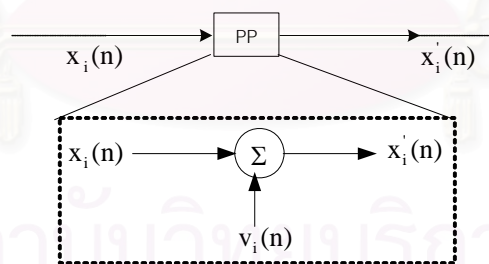
ขึ้นกับกรณีปกติของระบบ SAEC ที่ไม่ได้มีการดีคอดิเลชันสัญญาณเสียงสเตอริโอ วิธีที่สองการดีคอดิเลชันสัญญาณเสียงสเตอริโอโดยการเพิ่มภาวะไม่เชิงเส้น [2-5] และวิธีที่สามเป็นวิธีที่น่าเสนอขึ้นในวิทยานิพนธ์ฉบับนี้คือ การดีคอดิเลชันสัญญาณเสียงสเตอริโอโดยเพิ่มสัญญาณรบกวนแบบปรับตัวได้ ทั้งนี้การศึกษาเปรียบเทียบวิธีต่างๆ ในการดีคอดิเลชันสัญญาณเสียงสเตอริโอจะพิจารณาทั้งในแง่ของประสิทธิภาพการกำจัดสัญญาณเสียงสะท้อนของระบบ SAEC และในแง่ของผลกระทบของวิธีดีคอดิเลชันต่อคุณภาพเสียงสเตอริโอของสัญญาณที่ถูกดีคอดิเลชัน  $x_i'(n)$  ซึ่งออกสู่ลำโพงทางห้องรับ

### 3.1.1 การดีคอดิเลชันสัญญาณเสียงสเตอริโอโดยการเพิ่มสัญญาณรบกวนสุ่ม

เมื่อทำการเพิ่มสัญญาณรบกวนสุ่มทุกช่วงเวลาในแต่ละช่องสัญญาณของเสียงสเตอริโอ เพื่อลดสหสัมพันธ์ของสัญญาณด้วยสมการ

$$x_i'(n) = x_i(n) + v_i(n), \quad i=1,2 \quad (3.1)$$

หรือสามารถแสดงการเพิ่มสัญญาณรบกวนสุ่มตามแบบจำลองในรูปที่ 3.2 สัญญาณเสียงสเตอริโอ  $x_i(n)$  เมื่อ  $i=1, 2$  ในแต่ละช่องสัญญาณ จะถูกดีคอดิเลชันเป็น  $x_i'(n)$  และส่งต่อไปยังระบบ SAEC



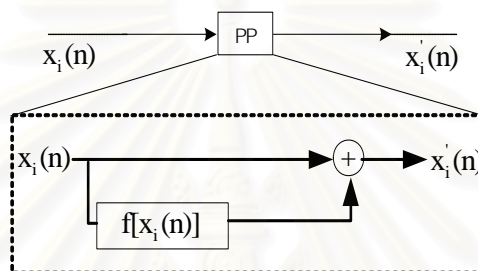
รูปที่ 3.2 แบบจำลองการดีคอดิเลชันสัญญาณเสียงสเตอริโอในบล็อก PP ด้วยการเพิ่มสัญญาณรบกวนสุ่ม

ขนาดของสัญญาณ  $v_i(n)$  สามารถกำหนดได้ตามต้องการ โดยทฤษฎีแล้วสัญญาณรบกวนสุ่มเป็นสัญญาณที่มีสหสัมพันธ์กันน้อยมากเนื่องจากเป็นสัญญาณที่ได้จากการสุ่มในแต่ละช่วงเวลา ดังนั้นเป็นไปได้ว่าสัญญาณ  $x_i'(n)$  ที่ได้จากการเพิ่มสัญญาณรบกวนสุ่มยังมีสหสัมพันธ์กันน้อยลงเมื่อใส่สัญญาณรบกวนสุ่มที่มากขึ้น อย่างไรก็ตาม การใส่สัญญาณรบกวนสุ่มทุกช่วงเวลาจะทำได้

ขอบเขตหนึ่งเท่านั้น อันเนื่องจากผลกระทบที่มีต่อคุณภาพสัญญาณเสียงในระบบ SAEC คือ อาจส่งผลกระทบต่อคุณภาพของสัญญาณเสียงต้นฉบับจนไม่สามารถยอมรับได้

### 3.1.2 การดีคอดิเลชันสัญญาณเสียงสเตอริโอโดยการเพิ่มภาวะไม่เชิงเส้น

การดีคอดิเลชันสัญญาณเสียงสเตอริโอโดยการเพิ่มภาวะไม่เชิงเส้น (Non-linearity) ทำได้ด้วยการเพิ่มภาวะไม่เชิงเส้นในส่วนประมวลผลก่อนดัดบล็อกไดอะแกรมรูปที่ 3.3 เพื่อลดสหสัมพันธ์ของสัญญาณเสียงสเตอริโอ



รูปที่ 3.3 แบบจำลองการดีคอดิเลชันสัญญาณเสียงสเตอริโอในบล็อก PP ด้วยการเพิ่มภาวะไม่เชิงเส้น

สัญญาณขาออกของวงจรไม่เป็นเชิงเส้นในแต่ละช่องสัญญาณมีค่าเป็น

$$x'_i(n) = x_i(n) + \alpha [f(x_i(n))], \quad i=1,2 \quad (3.2)$$

เมื่อ  $\alpha$  คือ สัมประสิทธิ์การแปลงของภาวะไม่เชิงเส้น ซึ่งสามารถปรับขนาดได้ ค่า  $\alpha$  ที่เหมาะสมควรจะอยู่ในช่วง  $0 \leq \alpha \leq 1.5$  [3] ทั้งนี้เพื่อลดสหสัมพันธ์สัญญาณเสียงสเตอริโอทั้ง 2 ช่องสัญญาณ โดยที่ไม่ทำให้คุณภาพเสียงสเตอริโอของสัญญาณที่ถูกดีคอดิเลชันลดลงจากเดิมมากนัก

ภาวะไม่เชิงเส้นที่สามารถนำมาใช้ลดสหสัมพันธ์สัญญาณเสียงสเตอริโอ นั้นมีหลายแบบ [5] ซึ่งสามารถสรุปได้ดังต่อไปนี้

1. Half-wave Rectifier เป็นภาวะไม่เชิงเส้นที่มีความซับซ้อนในการดำเนินการต่ำ จะมีการพิจารณาให้สัญญาณออกของวงจรนี้เป็นค่าบวกเท่านั้น คือเมื่อสัญญาณเข้าที่มี

ขนาดสัญญาณ (Amplitude) เป็นค่าบวก ก็สามารถผ่านไปในวงจรนี้ได้ ส่วนกรณีที่สัญญาณเข้ามีขนาดสัญญาณเป็นค่าลบก็จะถูกตัดทิ้งไป ดังสมการ

$$x'_i(n) = \begin{cases} x_i(n), & \text{if } x \geq 0 \\ 0, & \text{otherwise} \end{cases} \quad (3.3)$$

2. Full-wave Rectifier สัญญาณออกของวงจร จะเป็นค่าสัมบูรณ์ (Absolute) ของสัญญาณเสียงสเตอริโอในแต่ละช่องสัญญาณดังสมการ

$$x'_i(n) = |x_i(n)| \quad (3.4)$$

3. Hard Limiter เป็นภาวะไม่เชิงเส้นที่ให้สัญญาณออกของวงจรเป็นค่า Signum ดังสมการต่อไปนี้

$$x'_i(n) = \text{sgn}(x_i(n)) \quad (3.5)$$

4. Square-law เป็นภาวะไม่เชิงเส้นที่ให้สัญญาณออกของวงจรเป็นค่ากำลังสองของสัญญาณเข้า ดังนั้น สัญญาณออกที่ได้มีค่าดังสมการ

$$x'_i(n) = x_i^2(n) \quad (3.6)$$

5. Square-sign เป็นภาวะไม่เชิงเส้นที่ให้สัญญาณออกของวงจรเป็นค่ากำลังสองของสัญญาณเข้าคูณกันกับค่า Signum ของสัญญาณเข้า สัญญาณที่ได้มีค่าดังสมการ

$$x'_i(n) = x_i^2(n) \text{sgn}(x_i(n)) \quad (3.7)$$

6. Cubic เป็นภาวะไม่เชิงเส้นที่ให้สัญญาณออกของวงจรเป็นค่ากำลังสามของสัญญาณเข้าของระบบ สัญญาณออกที่ได้ก็จะมีทั้งค่าที่เป็นบวกและที่เป็นลบดังสมการ

$$x'_i(n) = x_i^3(n) \quad (3.8)$$

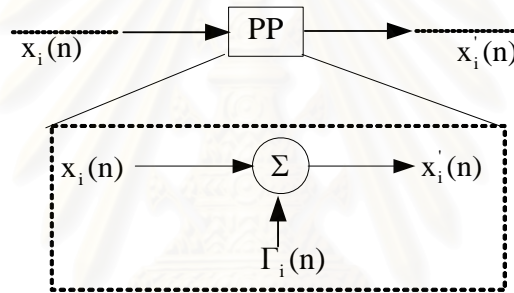
สำหรับในวิทยานิพนธ์ฉบับนี้ได้เลือกศึกษาภาวะไม่เชิงเส้นแบบตัวเรียงกระแสครึ่งคลื่น (Half-wave Rectifier) ที่เป็นดังสมการที่ (3.3) ในการดีคอดิเคชันสัญญาณเสียงสเตอริโอในระบบ

SAEC เนื่องจากเป็นวิธีที่มีผลกระทบต่อคุณภาพของสัญญาณเสียงดั้งเดิมค่อนข้างน้อย ในขณะที่ยังสามารถลดสหสัมพันธ์ของสัญญาณเสียงสเตอริโอได้ [3][5] เพื่อทดสอบผลการกำจัดสัญญาณเสียงสะท้อนเปรียบเทียบกับการใช้สัญญาณรบกวนสุ่มและวิธีที่จะนำเสนอต่อไป

### 3.1.3 การดีคอริเลชันสัญญาณเสียงสเตอริโอโดยการเพิ่มสัญญาณรบกวนแบบปรับตัวได้

#### 3.1.3.1 วิธีการดีคอริเลชันสัญญาณเสียงสเตอริโอโดยใช้สัญญาณรบกวนแบบปรับตัวได้

การดีคอริเลชันสัญญาณเสียงสเตอริโอด้วยวิธีนี้ เสียงสเตอริโอที่ถูกลดสหสัมพันธ์หลังจากผ่านบล็อก PP จะมีค่าเป็นไปตามรูปที่ 3.4



รูปที่ 3.4 การดีคอริเลชันสัญญาณเสียงสเตอริโอโดยการเพิ่มสัญญาณรบกวนแบบปรับตัวได้

สัญญาณสเตอริโอที่ได้เมื่อทำการใส่สัญญาณรบกวนแบบปรับตัวได้มีค่าเป็น

$$x'_i(n) = x_i(n) + \Gamma_i(n), \quad i=1,2 \quad (3.9)$$

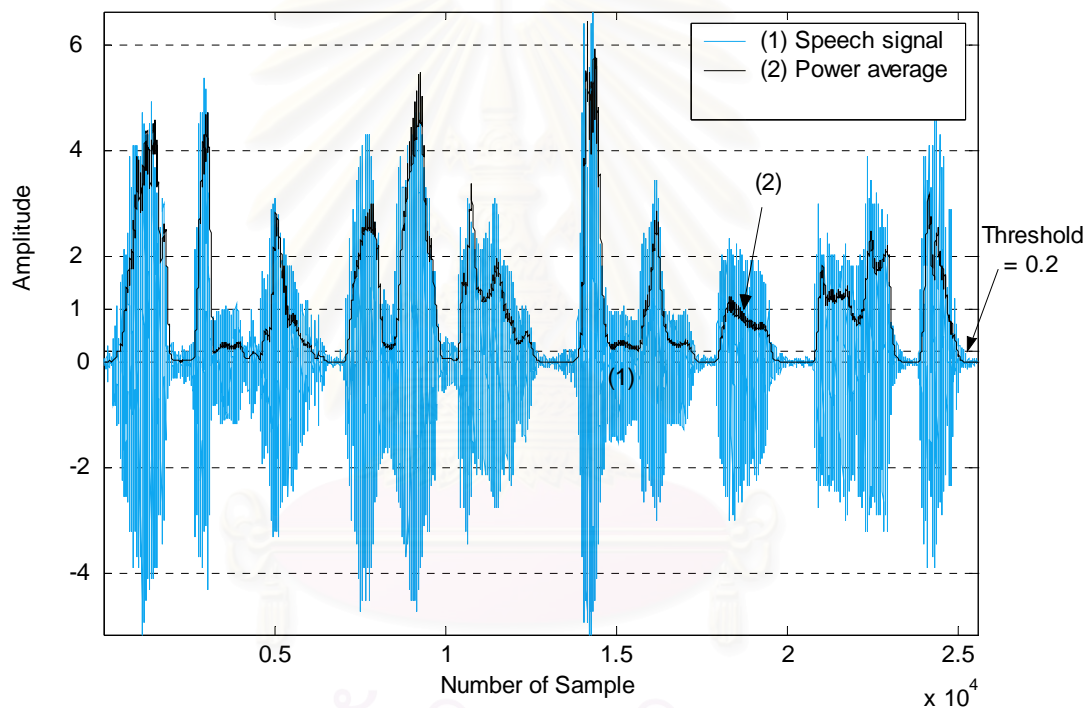
เมื่อ  $\Gamma_i(n)$  เป็นฟังก์ชันสัญญาณรบกวนแบบปรับตัวได้ ที่เพิ่มเข้าไปในสัญญาณเสียงสเตอริโอในแต่ละช่องสัญญาณเพื่อการดีคอริเลชัน ซึ่งจะกล่าวถึงในหัวข้อถัดไป

จากลักษณะพื้นฐานของเสียงดังหัวข้อที่ 2.5 เนื่องจากความต้องการที่จะให้ระบบการกำจัดสัญญาณเสียงสะท้อนแบบ SAEC มีประสิทธิภาพการทำงานที่ดีขึ้น ในขณะเดียวกันก็ยังคงไว้ซึ่งคุณภาพของสัญญาณเสียงสเตอริโอ ดังนั้น วิทยานิพนธ์ฉบับนี้จึงได้นำเสนอวิธีดีคอริเลชันของสัญญาณเสียงสเตอริโอด้วยการเพิ่มสัญญาณรบกวนแบบปรับตัวได้ (Adaptive Noise Addition) ตามขนาดที่ต่างกันของสัญญาณเสียงพูดในเฉพาะคาบที่มีเสียงพูด (Speech Active Period)



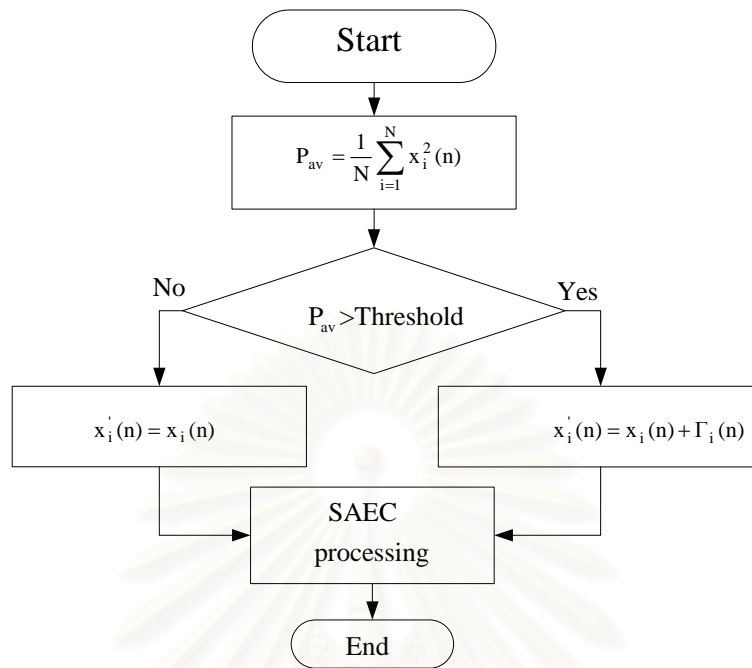
เท่านั้น กล่าวคือจะเลือกใส่สัญญาณรบกวนแบบปรับตัวได้ตามฟังก์ชัน  $\Gamma_i(\mathbf{n})$  ดังรูปที่ 3.4 เฉพาะช่วงที่ถูกตัดสินว่าเป็นคาบที่มีเสียงพูด ส่วนช่วงที่ถูกตัดสินว่าเป็นคาบเงียบ (Silence Period) ก็จะไม่ใส่สัญญาณรบกวนดังกล่าว ทั้งนี้เพื่อไม่ให้เกิดผลกระทบต่อคุณภาพเสียงสเตอริโอมากจนสามารถรู้สึกได้ ในขณะที่เดียวกันก็ส่งผลให้ระบบ SAEC มีการทำงานที่ดีขึ้น

สำหรับการตรวจสอบเพื่อตัดสินคาบที่มีเสียงพูดหรือคาบเงียบของสัญญาณเสียงพูดสามารถแสดงได้ดังรูปที่ 3.5 โดยเปรียบเทียบค่าพลังงานเฉลี่ยของสัญญาณเสียงพูดกับจุดเริ่มเปลี่ยน (Threshold) ที่กำหนดขึ้น



รูปที่ 3.5 การตรวจสอบช่วงเวลาของคาบที่มีเสียงพูดและคาบเงียบ

ผลการตรวจสอบคาบที่มีสัญญาณเสียงพูด จะตรวจสอบจากค่าพลังงานเฉลี่ยของสัญญาณเสียงพูดแต่ละช่วงเวลาตามกราฟเส้นที่ 2 ในรูปที่ 3.5 เปรียบเทียบกับจุดเริ่มเปลี่ยนคงที่ค่าหนึ่ง ซึ่งในที่นี้ถูกเลือกเป็นค่า 0.2 เมื่อพลังงานเฉลี่ยของสัญญาณเสียงพูดช่วงใดมากกว่าค่าจุดเริ่มเปลี่ยน (Threshold) จะถูกตัดสินว่าเป็นคาบที่มีสัญญาณเสียงพูด ส่วนช่วงใดที่พลังงานเฉลี่ยน้อยกว่าค่าจุดเริ่มเปลี่ยน จะถูกตัดสินว่าเป็นคาบเงียบ หลังจากนั้น ทำการเพิ่มสัญญาณรบกวนแบบปรับตัวได้ตามขนาดของสัญญาณเสียงพูดในคาบที่มีเสียงพูด ซึ่งบล็อกไดอะแกรมการทำงานของวิธีที่นำเสนอแสดงดังรูปที่ 3.6



รูปที่ 3.6 ไตอะแกรมตรวจสอบคานและเพิ่มสัญญาณรบกวนตามฟังก์ชันที่กำหนด

### 3.1.3.2 ฟังก์ชันสัญญาณรบกวนแบบปรับตัวได้

ในการพิจารณาฟังก์ชันสัญญาณรบกวนแบบปรับตัวได้ ( $\Gamma_i(n)$ ) ที่จะใส่ในบล็อก PP เพื่อให้สัญญาณเสียงสเตอริโอมีสหสัมพันธ์กันน้อยลงในขณะที่ยังคงไว้ซึ่งคุณภาพเสียงของสัญญาณสเตอริโอนั้น ได้ใช้ลักษณะพื้นฐานของสัญญาณเสียงดังหัวข้อที่ 2.5 กล่าวคือ เลือกใส่สัญญาณรบกวนในช่วงที่ถูกตัดสินว่าเป็นคานที่มีสัญญาณเสียงพูดดังได้กล่าวในหัวข้อ 3.1.3.1 ซึ่งการกำหนดขนาดของฟังก์ชันสัญญาณรบกวนแบบปรับขนาดดังกล่าว อธิบายได้ตามลำดับแนวคิดดังต่อไปนี้

ลำดับแนวคิดที่ 1 ขนาดของสัญญาณรบกวนแบบปรับตัวได้ ถูกกำหนดให้เป็นฟังก์ชันที่ขึ้นอยู่กับขนาดของสัญญาณเสียงพูด ด้วยอัตราส่วนขนาดสัญญาณเสียงพูดขณะนั้นต่อสัญญาณรบกวนขาวแบบเกาส์ (Signal-to-Noise Ratio, SNR) สัญญาณที่ถูกดีคอดรีเลชันมีค่าดังสมการ

$$x'_i(n) = x_i(n) + \underbrace{\gamma_i(n) \times v_i(n)}_{\Gamma_i(n)} \quad (3.10)$$

เมื่อ  $i=1, 2$  โดย  $\gamma_i(\mathbf{n})$  ถูกเลือกให้เป็นสัดส่วนของการเปลี่ยนแปลงของสัญญาณเสียงพูดเปรียบเทียบกับสัญญาณรบกวนขาวแบบเกาส์ (White Gaussian Noise) มีค่าเป็น

$$\gamma_i(\mathbf{n}) = \sqrt{\frac{x_i^2(\mathbf{n})}{v_i^2(\mathbf{n})} \times 10^{-\left(\frac{\text{SNR}}{10}\right)}} \quad (3.11)$$

เมื่อ  $x_i^2(\mathbf{n})$  เป็นค่าพลังงานของสัญญาณเสียงสเตริโอในแต่ละช่องสัญญาณ และ  $v_i^2(\mathbf{n})$  เป็นค่าพลังงานของสัญญาณรบกวนขาวแบบเกาส์ ดังนั้นเมื่อช่วงไหนที่สัญญาณเสียงสเตริโอในแต่ละช่องสัญญาณมีขนาดของพลังงานสูง ก็จะสามารถใส่สัญญาณรบกวนขาวแบบเกาส์ที่มีขนาดมากลงไปในช่วงสัญญาณดังกล่าวได้ โดยกำหนดให้ค่า SNR ในสมการที่ (3.11) มีค่าน้อยสุดเป็น 20 dB ซึ่งหากค่า SNR มีค่าต่ำกว่านี้ อาจส่งผลให้สัญญาณเสียงสเตริโอที่ถูกดีคอดิเรตมีคุณภาพเสียงที่ด้อยลงจนไม่สามารถยอมรับได้

ลำดับแนวคิดที่ 2 เนื่องด้วยฟังก์ชัน  $\Gamma_i(\mathbf{n})$  ในแนวคิดที่หนึ่งสามารถทำให้สัญญาณเสียงสเตริโอมีสหสัมพันธ์กันลดลงโดยอาศัยคุณสมบัติของสัญญาณรบกวนขาวแบบเกาส์ อย่างไรก็ตาม อาจส่งผลกระทบต่อคุณภาพสัญญาณเสียงสเตริโอให้ด้อยลง เมื่อค่า SNR ในสมการที่ (3.11) มีค่าต่ำกว่า 20 dB ดังนั้นเพื่อเป็นการรักษาคุณภาพของสัญญาณเสียงสเตริโอให้คล้ายสัญญาณเดิมให้มากที่สุด แนวคิดที่สองนี้จึงเลือกให้ฟังก์ชัน  $\Gamma_i(\mathbf{n})$  เป็นสัดส่วนของพลังงานเฉลี่ยสัญญาณเสียงพูดเดิมด้วยค่า  $\beta$  ส่งผลให้ฟังก์ชัน  $\Gamma_i(\mathbf{n})$  มีคุณสมบัติการดีคอดิเรตคล้ายกับภาวะไม่เชิงเส้นที่กล่าวถึงในหัวข้อที่ 3.1.2 สัญญาณเสียงสเตริโอที่ถูกดีคอดิเรตในแต่ละช่องสัญญาณจึงมีค่าเป็นดังสมการ

$$\hat{x}_i'(\mathbf{n}) = x_i(\mathbf{n}) + \beta \times P_{x_i}(\mathbf{n}) \quad (3.12)$$

เมื่อ  $\beta$  คือ ค่าคงที่ เพื่อใช้กำหนดสัดส่วนของพลังงานเฉลี่ยสัญญาณเสียงพูดที่จะเพิ่มในสัญญาณเสียงสเตริโอเดิมและค่าพลังงานเฉลี่ยของสัญญาณเสียงสเตริโอในแต่ละช่องสัญญาณสามารถหาได้จาก

$$P_{x_i}(\mathbf{n}) = \frac{1}{K} \sum_{j=0}^{K-1} |x_i(\mathbf{n}-j)|^2 \quad (3.13)$$

โดยที่  $K$  เป็นขนาดของข้อมูลที่ต้องการหาพลังงานเฉลี่ย ทั้งนี้ เพื่อลดความซับซ้อนในการคำนวณของสมการที่ (3.13) จึงสามารถหาค่าพลังงานเฉลี่ยของสัญญาณได้จาก

$$P_{x_i}(n) = P_{x_i}(n-1) + x_i^2(n) - x_i^2(n-K+1) \quad (3.14)$$

ดังนั้นจากแนวคิดทั้งสองดังกล่าวจึงส่งผลให้วิทยานิพนธ์ฉบับนี้ได้นำเสนอวิธีการเพิ่มสัญญาณรบกวนแบบปรับตัวได้ที่มีฟังก์ชัน  $\Gamma_i(n)$  ดังสมการ

$$\hat{x}_i'(n) = x_i(n) + \underbrace{\beta \times P_{x_i}(n) \times \gamma_i(n) \times v_i(n)}_{\Gamma_i(n)} \quad (3.15)$$

สำหรับการเปรียบเทียบประสิทธิภาพในการดีคอดิเลชันสัญญาณเสียงสเตอริโอ และความสามารถในการรักษาคุณภาพเสียงของสัญญาณเสียงสเตอริโอจะกล่าวในบทต่อไป

## 3.2 การวัดประสิทธิภาพการทำงานของระบบ

### 3.2.1 สมรรถนะของวงจรกรองแบบปรับตัวในโดเมนเวลา

ในการศึกษาและเปรียบเทียบลักษณะการกำจัดสัญญาณเสียงสะท้อนแบบช่องสัญญาณเดี่ยวและแบบสเตอริโอ นั้นจะเปรียบเทียบจากลักษณะการลู่เข้าสู่ค่าตอบของวงจรกรองแบบปรับตัวว่าสามารถปรับค่าสัมประสิทธิ์ของวงจรกรองให้เข้าใกล้ค่าวิถีการสะท้อน (AEP) ได้ใกล้เคียงกันเพียงใด ยิ่งมีความใกล้เคียงมากก็แสดงให้เห็นถึงค่าความผิดพลาดของระบบที่น้อยลง ซึ่งหมายถึงสามารถกำจัดสัญญาณเสียงสะท้อนที่เกิดขึ้นในระบบได้นั่นเอง

สำหรับค่าที่ใช้วัดเปรียบเทียบลักษณะการกำจัดสัญญาณเสียงสะท้อนว่า วงจรกรองสามารถช่วยกำจัดสัญญาณเสียงสะท้อนได้ดีเพียงใดสามารถสังเกตและเปรียบเทียบได้จากค่าต่างๆ ดังนี้คือ

1. ค่าของความผิดพลาดเฉลี่ย (Mean Square Error, MSE) หาได้จากสมการที่

$$MSE = |e(n)|^2 \quad (3.16)$$

2. คำนวณน้ำหนักผลต่างยกกำลังสองของสัมประสิทธิ์วงจรรองกับวิธีการสะท้อน (Weight Error Vector Norm, WEVN) หาได้จากสมการที่ (3.17) สำหรับระบบ AEC และจากสมการที่ (3.18) สำหรับระบบ SAEC

$$\text{WEVN}(n) = 10 \times \log \left( \frac{\|\mathbf{h} - \hat{\mathbf{h}}\|^2}{\|\mathbf{h}\|^2} \right) \quad (3.17)$$

$$\text{WEVN}(n) = 10 \times \log \left( \frac{\|\mathbf{h}_1(n) - \hat{\mathbf{h}}_1(n)\|^2 + \|\mathbf{h}_2(n) - \hat{\mathbf{h}}_2(n)\|^2}{\|\mathbf{h}_1(n)\|^2 + \|\mathbf{h}_2(n)\|^2} \right) \quad (3.18)$$

โดยจะนำไปใช้ตลอดการทดลองเพื่อทดสอบสมรรถนะหรือประสิทธิภาพการทำงานของระบบ SAEC และเปรียบเทียบผลที่จะเกิดขึ้นในการจำลองการทำงานของระบบต่อไป

### 3.2.2 ฟังก์ชันร่วมนัย (Coherence Function)

โดยปกติแล้วสหสัมพันธ์หรือการคอรีเลชันกันระหว่างสัญญาณเสียงสเตอริโอทั้ง 2 ช่องสัญญาณนี้ สามารถวัดได้ด้วยฟังก์ชันร่วมนัย (Coherence Function) ซึ่งสามารถนิยามในโดเมนของความถี่ [3][5] ได้ดังสมการนี้

$$\gamma(f) = \frac{S_{x_1 x_2}(f)}{\sqrt{S_{x_1 x_1}(f) S_{x_2 x_2}(f)}} \quad (3.19)$$

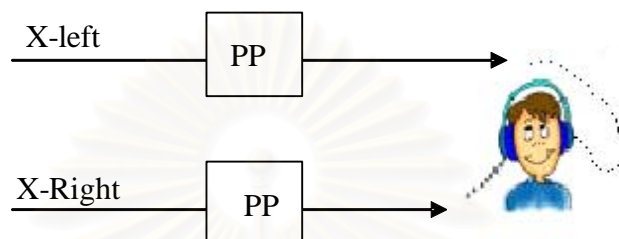
เมื่อ  $S_{x_k(x_l)}$  คือ ค่าประมาณกำลังเชิงสเปกตรัม (Power Spectral Estimation) ของสัญญาณ หาได้จากสมการ

$$S_{x_k x_l}(f) = \sum_{\tau=-\infty}^{+\infty} E \{ x_k(n) x_l(n-\tau) \} e^{-i2\pi f \tau} = \sum_{\tau=-\infty}^{+\infty} R_{x_k x_l}(\tau) e^{-i2\pi f \tau}, \quad k, l=1, 2 \quad (3.20)$$

เมื่อสัญญาณมีสหสัมพันธ์หรือคอรีเลชันกันมาก ค่า  $\gamma(f)$  จะใกล้เคียงกับ 1 มาก โดยเฉพาะถ้าสัญญาณมีสหสัมพันธ์กันเป็นเชิงเส้นอย่างสมบูรณ์ (Completely Linearly Related) ค่า  $\gamma(f)$  จะเท่ากับ 1 แต่โดยความเป็นจริงแล้ว ถึงแม้เป็นสัญญาณจากแหล่งกำเนิดเดียวกัน ก็ยังมีสัญญาณรบกวนอยู่ด้วย ฟังก์ชันร่วมนัยที่ได้จึงไม่เป็นสหสัมพันธ์เชิงเส้นแบบสมบูรณ์ในทางปฏิบัติจริง

### 3.2.3 การทดสอบคุณภาพเสียงทางกายภาพ (Subjective Listening Test)

การทดสอบคุณภาพทางกายภาพในวิทยานิพนธ์ฉบับนี้ได้เลือกใช้แบบการคิดคะแนนเฉลี่ย (Mean Opinion Score, MOS) ตามมาตรฐานของ ITU-R [14] โดยมีขั้นตอนการทำการทดสอบคือให้ผู้ทำการทดสอบรับฟังสัญญาณเสียงสเตอริโอแต่ละชุดที่ต้องการทดสอบดังรูปที่ 3.7



รูปที่ 3.7 สัญญาณเสียงสเตอริโอที่ใช้ในการทดสอบคุณภาพ

หลังจากนั้นจึงให้คะแนนจาก 1 ถึง 5 แก่คุณภาพสัญญาณเสียงสเตอริโอในแต่ละชุด โดยเป็นการให้คะแนนอย่างอิสระต่อกันในการฟังสัญญาณเสียงแต่ละชุด โดยที่ความสัมพันธ์ของคุณภาพเสียงกับคะแนนถูกอธิบายดังตารางที่ 3.1

Rating	Speech Quality	Level of Distortion
5	Excellent	Imperceptible
4	Good	Just perceptible, but not annoying
3	Fair	Perceptible and slightly annoying
2	Poor	Annoying, but not objectionable
1	Unsatisfactory (Bad)	Very annoying and objectionable

ตารางที่ 3.1 คุณภาพของสัญญาณเสียงตามคะแนนต่าง ๆ ที่ให้กับการทดสอบ MOS Testing [14]

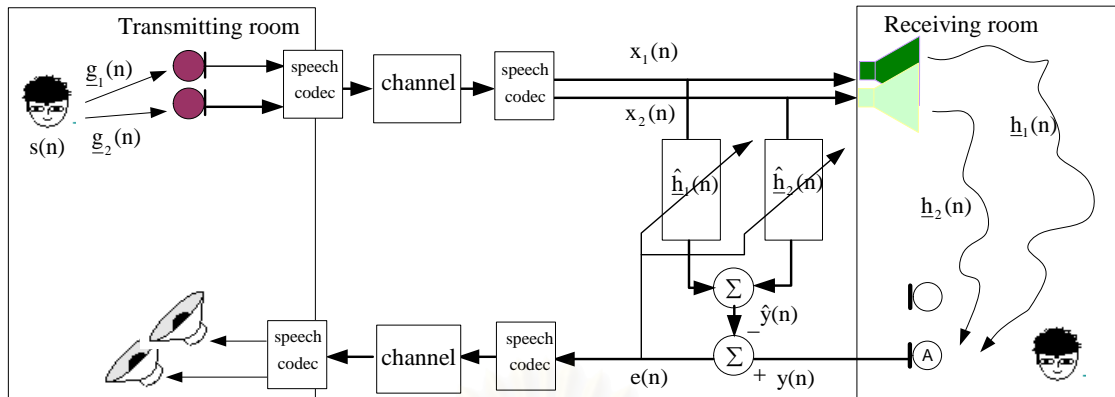
ในการทดสอบจะให้ผู้ทดสอบได้ฟังสัญญาณเสียงสัญญาณต้นแบบที่คุณภาพต่างกัน 3 ระดับ เพื่อให้ผู้ทำการทดสอบได้มีบรรทัดฐานในการตัดสินใจคุณภาพของสัญญาณเสียงที่จะทำการทดสอบต่อไป ซึ่งความสัมพันธ์ของสัญญาณเสียงต้นแบบทั้ง 3 ระดับเป็นดังสมการต่อไปนี้

$$s_{\text{new}}(n) = \frac{s(n) + \beta v(n)}{(1 + \beta^2)} \quad (3.21)$$

โดยที่  $s(n)$  คือสัญญาณเสียงต้นแบบเดิม และ  $v(n)$  คือ สัญญาณรบกวนขาวแบบเกาส์ ซึ่งมีขนาด SNR เท่ากับ 30 dB ค่า  $\beta$  จะเป็นค่าปรับระดับของสัญญาณรบกวนเพื่อให้ได้สัญญาณที่คุณภาพต่าง ๆ และ  $s_{\text{new}}(n)$  คือ สัญญาณเสียงต้นแบบทั้ง 3 ระดับ หลังจากผู้ทำการทดสอบได้ฟังสัญญาณเสียงต้นแบบที่คุณภาพต่าง ๆ กัน 3 ระดับแล้ว ต่อจากนั้นจะให้ผู้ทำการทดสอบได้ฟังสัญญาณเสียงสเตอริโอที่ผ่านการดีคอดริเคชันแบบต่างๆ แล้วทำการให้คะแนนแต่ละสัญญาณ แล้วนำผลการให้คะแนนดังกล่าวมาหาค่าเฉลี่ยเพื่อนำไปเปรียบเทียบคุณภาพสัญญาณต่อไป

### 3.3 การเข้ารหัสสัญญาณเสียงพูด

การเข้ารหัสสัญญาณเสียงพูดเป็นอีกวิธีการหนึ่งซึ่งส่งผลให้เกิดการเปลี่ยนแปลงในทางที่ดีขึ้นต่อการทำงานของระบบ SAEC [8] จากรูปที่ 3.8 เป็นระบบ SAEC ที่มีการเข้ารหัสและถอดรหัสสัญญาณเสียงพูด เมื่อจำลองทางเดินของคลื่นเสียงของผู้พูดในห้องส่ง  $s(n)$  ด้วยผลตอบสนองต่ออิมพัลส์  $g_1(n)$  และ  $g_2(n)$  และเสียงพูดจะถูกเปลี่ยนเป็นสัญญาณทางไฟฟ้าด้วยไมโครโฟนในห้องส่งทั้งสองตัว สัญญาณสเตอริโอดังกล่าวจะอยู่ในรูปของดิจิทัลแล้ว เมื่อนำการเข้ารหัสและถอดรหัสสัญญาณเสียงพูดมาใช้งานร่วมกับระบบ SAEC สัญญาณที่ออกจากไมโครโฟนในห้องส่งจะถูกบีบอัดด้วยกระบวนการเข้ารหัส และถูกส่งผ่านไปในห้องสัญญาณสื่อสาร หลังจากนั้นทางห้องรับจะทำการแปลงสัญญาณที่ได้รับด้วยกระบวนการถอดรหัสให้เป็นสัญญาณสเตอริโอ  $x_1(n)$  และ  $x_2(n)$  ซึ่งวิธีการเข้ารหัสและถอดรหัสสัญญาณเสียงพูดจะกล่าวถึงในหัวข้อถัดไป โดยสัญญาณเสียงสเตอริโอ  $x_1(n)$  และ  $x_2(n)$  จะมีสหสัมพันธ์กันลดลง เนื่องจากความเพี้ยน (Distortion) ที่เกิดขึ้นในขณะที่ทำการเข้ารหัสและถอดรหัสสัญญาณเสียงดังกล่าว ส่งผลให้ระบบ SAEC มีการทำงานที่ดีขึ้น และจะแสดงผลการทำงานของระบบดังกล่าวนี้ในบทที่ 4 ต่อไป



รูปที่ 3.8 แบบจำลองระบบ SAEC ที่มีการเข้ารหัสสัญญาณเสียงพูด

### 3.3.1 วิธีการเข้ารหัสสัญญาณเสียงพูด

ระบบการสื่อสารทางเสียงที่มีประสิทธิภาพย่อมส่งผลให้ข้อมูลในการสื่อสารมีความถูกต้องแม่นยำ ปกติแล้วเสียงพูดจะมีลักษณะสัญญาณเป็นแอนะล็อกดังกล่าวกว่าในหัวข้อที่ 2.5 แต่เนื่องจากปริมาณข้อมูลส่งในช่องสัญญาณสื่อสารมีมาก แต่แบนด์วิดท์ (Bandwidth) ของช่องสัญญาณมีจำกัด ดังนั้นวิธีการเข้ารหัสหรือบีบอัดเสียงพูดจึงเข้ามามีบทบาทต่องานประยุกต์ประเภทการส่งสัญญาณเสียงพูดผ่านช่องสัญญาณ ซึ่งจะต้องมีการเปลี่ยนเสียงพูดจากอะนาลอกเป็นดิจิทัลแล้วจึงทำการเข้ารหัสต่อไป การเข้ารหัสเสียงพูดที่นิยมใช้กันอยู่ในปัจจุบันนี้สามารถแยกได้เป็น 3 ประเภท [19] คือ Waveform Coder, Source Coder (Vocoder) และ Hybrid Coder ซึ่งมีคุณลักษณะของการแปลงข้อมูล และคุณภาพสัญญาณเสียงแตกต่างกันดังต่อไปนี้

#### 3.3.1.1 Waveform Coder

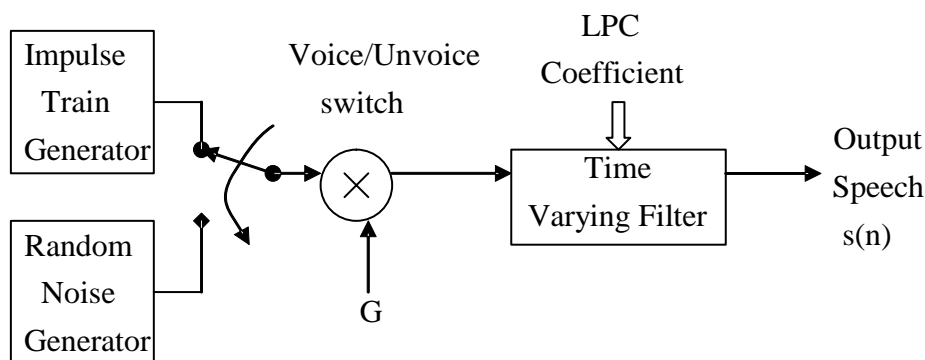
การเข้ารหัสสัญญาณเสียงแบบนี้จะให้คุณภาพสัญญาณเสียงที่ดี หากแต่ว่าอัตราบิตในการเข้ารหัสที่ใช้ค่อนข้างสูงมากเมื่อเปรียบเทียบกับวิธีการเข้ารหัสสัญญาณเสียงประเภทอื่น กระบวนการที่ใช้ในการเข้ารหัสแบบนี้ไม่มีความซับซ้อนมาก เริ่มจากกระบวนการ Pulse Code Modulation (PCM) ที่ใช้หลักการผ่านสัญญาณเสียงที่เป็นสัญญาณแอนะล็อกไปทำการชักตัวอย่าง (Sampling) หลังจากนั้นแทนค่าสัญญาณที่ชักตัวอย่างได้ด้วยค่าทางดิจิทัลเรียกว่า การควอนไทซ์ (Quantization) ทั้งนี้จำนวนบิตข้อมูลที่ใช้แทนค่า ขึ้นอยู่กับความละเอียดที่ต้องการ ตัวอย่างเช่นสัญญาณเสียงพูดของคนปกติที่ช่วงสัญญาณแคบ (Narrow-band) จะมีความถี่สูงสุดของข้อมูลอยู่ที่ประมาณ 4 kHz ดังนั้น ความถี่ในการชักตัวอย่างสัญญาณ (Sampling Frequency) จึงควรมีค่าอย่างน้อยเป็น 2 เท่าของความถี่สูงสุดคือประมาณ 8 kHz เมื่อเลือกใช้จำนวนบิตเป็น 8 บิต ในการ



แทนค่าข้อมูลของสัญญาณ อัตราบิตข้อมูล (Bit Rate) ที่ต้องใช้ในการเข้ารหัสเพื่อแทนค่าของสัญญาณเสียงพูดนี้จึงเป็น 64 kbps ซึ่งพบว่าคุณภาพเสียงที่ได้ใกล้เคียงกับสัญญาณเสียงดั้งเดิมมาก [19][22] อย่างไรก็ตามด้วยอัตราบิตข้อมูลที่ใช้นี้ค่อนข้างสูง จึงมีการพยายามหาวิธีลดอัตราบิตข้อมูลนี้ลง ด้วยการใช้อัลกอริทึม (Algorithm) ในการแทนค่าสัญญาณเพื่อให้จำนวนบิตข้อมูลที่ใช้แทนค่าสัญญาณลดลง เรียกว่า Logarithmic Quantizer ได้แก่ A-law และ  $\mu$ -law [19] ที่เป็นมาตรฐานใช้งานในปัจจุบัน นอกจากนี้ยังมีวิธีอื่นอีกเพื่อลดอัตราบิตข้อมูลในการเข้ารหัสสัญญาณลง เช่น การใช้กระบวนการ Difference Pulse Code Modulation (DPCM) และ Adaptive Difference Pulse Code Modulation (ADPCM) ซึ่งเป็นวิธีการที่อาศัยลักษณะทางกายภาพของสัญญาณเสียงที่มีความสัมพันธ์กันในการส่งสัญญาณแต่ละครั้ง ดังนั้นการเข้ารหัสสัญญาณเสียงด้วยวิธีนี้ จึงใช้เฉพาะความแตกต่างของสัญญาณเสียงพูดที่เวลาติดกันมาทำการเข้ารหัสในกระบวนการ PCM ผลที่เกิดขึ้นคือ สามารถลดอัตราบิตข้อมูลที่ใช้ลงได้ นอกจากนี้ยังพัฒนาต่อไปเป็นการใช้การทำนายสัญญาณเสียงพูด (Prediction) โดยนำไปเปรียบเทียบกับสัญญาณจริงเพื่อหาความแตกต่างของข้อมูลเสียง แล้วใช้ความแตกต่างนั้นปรับสัมประสิทธิ์วงจรรองแบบปรับตัวของการทำนายสัญญาณเสียงพูด เพื่อนำไปเปรียบเทียบกับเสียงรอบต่อไป จนค่าความแตกต่างนั้นมีค่าลดลงน้อยลงเรื่อยๆ อัตราบิตข้อมูลที่ใช้ก็จะมีการลดลงไปอีก วิธีการนี้เรียกว่า ADPCM ทั้งนี้การพัฒนาการเข้ารหัสแบบ Waveform Coder โดยการแบ่งสัญญาณเสียงออกเป็นช่วงความถี่แคบๆ ในโดเมนของความถี่ (Subband Frequency) เพื่อแยกการเข้ารหัสในแต่ละช่วงความถี่ด้วยอัตราบิตต่างกัน ก็สามารถลดอัตราบิตข้อมูลที่ได้อีกเรียกว่า Subband Coder วิธีนี้มีข้อดีคือสัญญาณรบกวน (Noise) ที่เข้ามากระทบต่อการเข้ารหัสสัญญาณจะมีปริมาณน้อยมากเพราะระบบทำงานในโดเมนของความถี่ อย่างไรก็ตามเนื่องจากการแบ่งสัญญาณออกเป็นช่วงๆ จึงทำให้การเข้ารหัสแบบนี้มีผลของการประวิงเวลา (Delay) เข้ามาเกี่ยวข้องด้วย

### 3.3.1.2 Source Coder (Vocoder)

การเข้ารหัสสัญญาณเสียงพูดแบบนี้สามารถลดอัตราบิตข้อมูลลงได้ถึงประมาณ 2.4 kbps โดยอาศัยหลักการของความแตกต่างในการกำเนิดเสียงจากเส้นเสียงระหว่างเสียงพูดแบบ Voiced Sound และแบบ Unvoiced Sound ในการสร้างแบบจำลองเสียงพูด ตามหัวข้อที่ 2.5 โดยสามารถแสดงการจำลองการเข้ารหัสแบบนี้ดังรูปที่ 3.9

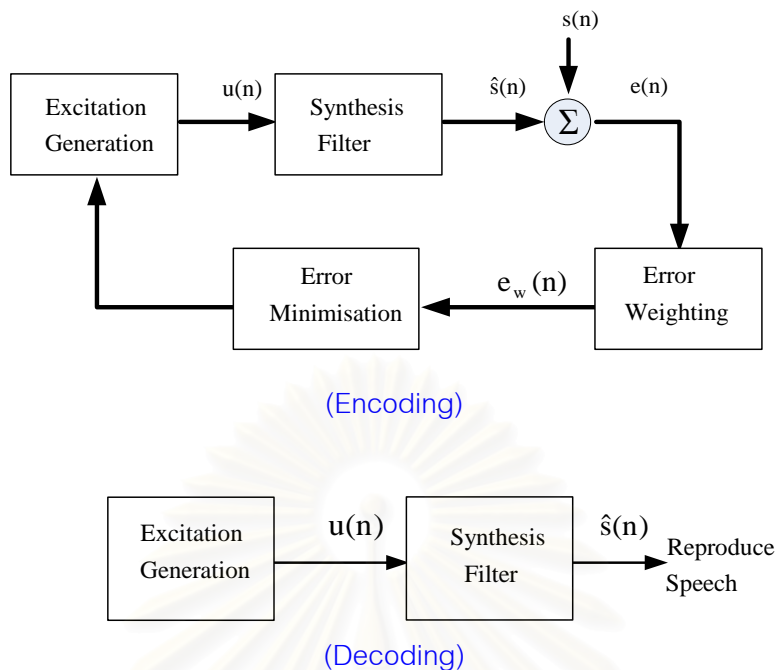


รูปที่ 3.9 แบบจำลองการเข้ารหัสแบบ Source Coder

แบบจำลองในรูปที่ 3.9 จะทำการเปรียบเทียบลักษณะเสียงพูด  $G$  ที่เข้าสู่ระบบกับสัญญาณที่ระบบสร้างขึ้น แล้วจำลองลักษณะเส้นเสียงด้วยผลตอบสนองต่ออิมพัลส์ที่ระบบสร้างขึ้นในช่วงที่มีเสียงพูดและสัญญาณรบกวนในช่วงที่ไม่มีเสียงพูด เมื่อหาค่าสัมประสิทธิ์ของวงจรกรองแบบปรับตัวในรูปที่ 3.9 เพื่อใช้ในการจำลองเส้นเสียงได้ ก็ทำการส่งสัมประสิทธิ์เหล่านี้ไปยังภาครับ เพื่อใช้ในการสังเคราะห์ (Synthesis) สัญญาณเสียงพูดแบบ Voiced Sound หรือแบบ Unvoiced Sound แต่วิธีนี้อาจทำให้มีความผิดพลาดในการจำลองสัญญาณเสียงที่ภาครับเกิดขึ้น ซึ่งเรียกการเข้ารหัสโดยวิธีนี้อีกอย่างว่า Linear Predictive Coding (LPC) ข้อเสียของการเข้ารหัสแบบนี้คือ เสียงที่ได้ อาจจะมีลักษณะคล้ายเสียงหุ่นยนต์มากกว่าเสียงพูด อย่างไรก็ตามการเข้ารหัสแบบนี้ก็มีประโยชน์สำหรับการสื่อสารที่ต้องการอัตราบิตข้อมูลน้อยๆ ความปลอดภัยสูง โดยยังสามารถฟังเสียงนั้นได้รู้เรื่องไม่จำเป็นต้องชัดเจนมากนัก เช่น การสื่อสารที่ใช้ในการทหาร เป็นต้น [19]

### 3.3.1.3 Hybrid Coder

การเข้ารหัสประเภท Hybrid Coder เป็นการรวมข้อดีของการเข้ารหัสทั้งสองแบบที่กล่าวมาข้างต้น [19] คือ อัตราบิตข้อมูลที่ต่ำและคุณภาพสัญญาณเสียงพูดที่ได้จะอยู่ระหว่างการเข้ารหัสแบบ Waveform Coder และ Source Coder รูปแบบของการเข้ารหัสประเภทนี้ที่นิยมและถือว่ามีประสิทธิภาพสำเร็จในเรื่องของคุณภาพเสียงคือ วิธี Time Domain Analysis by Synthesis (Abs) ซึ่งมีแบบจำลองการเข้ารหัสดังรูปที่ 3.10



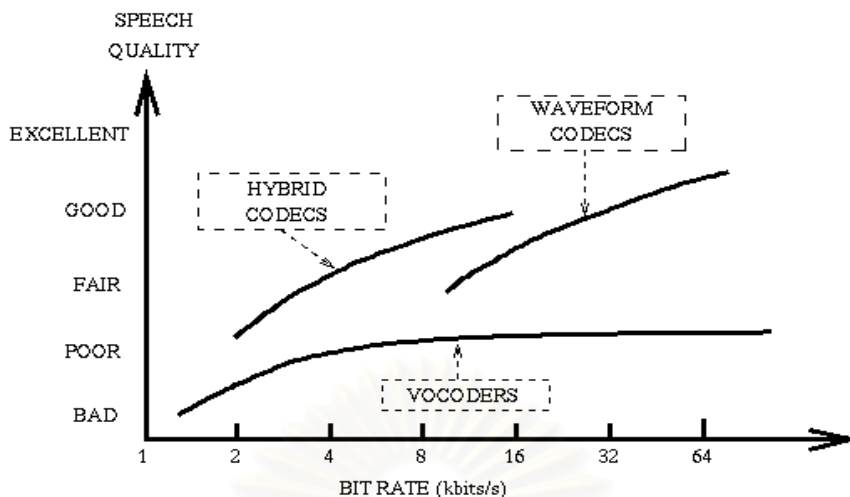
รูปที่ 3.10 แบบจำลองส่วนของการเข้ารหัสและถอดรหัสแบบ Hybrid Coder

แบบจำลองการเข้ารหัสแบบนี้มีหลักการเช่นเดียวกับวิธี LPC ใน Source Coder ดังรูปที่ 3.10 ระบบจะมีส่วนของการจำลองเส้นเสียงในบล็อกของ Excitation Generation แต่วิธีนี้นอกจากจะมีการสร้างแบบจำลองของเสียงทั้ง 2 ลักษณะคือ Voiced Sound และ Unvoiced Sound ยังทำการเพิ่มจำนวนพัลส์กระตุ้น (Exciting Pulse) ให้เป็นแบบ Multi-pulse เพื่อให้สามารถสังเคราะห์สัญญาณเสียงใกล้เคียงกับเสียงต้นฉบับมากขึ้น ในขณะเดียวกันก็นำเอาสัญญาณจริงเปรียบเทียบกับสัญญาณเสียงที่จำลองขึ้น เพื่อหาสัญญาณผิดพลาดที่เกิดขึ้น แล้วนำสัญญาณผิดพลาดไปใช้ในการปรับค่าสัมประสิทธิ์ของวงจรรองแบบปรับตัว จนกระทั่งสามารถจำลองเสียงให้เหมือนสัญญาณต้นฉบับมากที่สุด สำหรับคาบเวลาในการสังเคราะห์สัญญาณเสียงแต่ละเฟรม (Frame) นั้น อยู่ที่ประมาณ 20 ms กรณีที่เป็นเสียงผู้ชายที่มีความถี่ต่ำ จนถึง 2 ms กรณีที่เป็นเสียงความถี่สูง เช่นเสียงผู้หญิงหรือเสียงเด็ก [18] สำหรับการถอดรหัส (Decode) สัญญาณเสียงพูดนี้ ทำได้ด้วยการสังเคราะห์สัญญาณเสียงขึ้นโดยผ่านวงจรรองที่มีสัมประสิทธิ์เช่นเดียวกับสัมประสิทธิ์ของวงจรรองที่หาได้จาก LPC ทั้งนี้ ตัวอย่างการเข้ารหัสแบบนี้ได้แก่วิธี Multi Pulse Excited (MPE), Regular Pulse Excited (RPE) และ Code Excited Linear Predictive (CELP) เป็นต้น โดยวิธี MPE และ RPE จะทำการเข้ารหัสเช่นเดียวกับในวิธี Abs เพียงแต่แตกต่างกันในส่วนของลักษณะพัลส์ของสัญญาณที่ใช้กระตุ้นในการสร้างแบบจำลอง Synthesis Filter ส่วนความแตกต่างระหว่างวิธี MPE และวิธี RPE อยู่ที่จำนวนพัลส์ที่ใช้ในการสังเคราะห์สัญญาณเสียงพูด (Excitation Pulse)

คือ กรณียของวิธี RPE แต่ละพัลส์ที่ใช้ในการสังเคราะห์สัญญาณเสียงในภาคส่งมีขนาดแอมพลิจูด (Amplitude) เท่ากัน และมีตำแหน่งของพัลส์ (Position) ที่มีระยะห่างเท่ากันในแต่ละเฟรม ส่วนกรณียของวิธี MPE จะมีตำแหน่งและขนาดของพัลส์ที่ไม่เท่ากันตลอดทั้งเฟรม ดังนั้นในการส่งข้อมูลไปยังฝั่งภาครับ ทางภาคส่งจึงจำเป็นต้องทำการส่งตำแหน่งของพัลส์ และขนาดของพัลส์ไปด้วยในกรณียของวิธี MPE ส่วนกรณียของวิธี RPE นั้น เพียงแค่ส่งตำแหน่งของพัลส์และจำนวนของพัลส์ไปยังภาครับ โดยไม่จำเป็นต้องส่งขนาดของพัลส์ไปด้วย ถ้าพิจารณาที่อัตราบิตของมุลที่เท่ากัน วิธี RPE จึงให้คุณภาพที่ดีกว่าวิธี MPE

สำหรับวิธี CELP เป็นวิธีการเข้ารหัสที่สร้างขึ้นเพื่อต้องการลดอัตราบิตข้อมูลในการแทนค่าสัญญาณที่เข้ารหัสให้น้อยลงต่ำกว่าวิธี MPE และวิธี RPE ขณะที่พยายามทำให้สัญญาณเกิดความผิดพลาดน้อยสุด โดยมีพื้นฐานการเข้ารหัสเช่นเดียวกับวิธี MPE และ RPE เพียงแต่ลักษณะสัญญาณพัลส์ที่ใช้ในการกระตุ้นการสร้างแบบจำลอง Synthesis Filter ได้เลือกใช้ข้อมูลจากเวกเตอร์สัญญาณที่มีการจัดเตรียมไว้แล้วเรียกว่า Code Book โดยทุกๆเวกเตอร์ที่เลือกมาใช้งานจะเป็นอัตราส่วนสัญญาณจริงที่เหมาะสมที่สุดในแต่ละขนาดสัญญาณพัลส์ ซึ่งจะมีดัชนีในการบอกถึงค่าเวกเตอร์นั้นๆ ดังนั้นขณะเข้ารหัสสัญญาณจึงสามารถสร้าง Synthesis Filter ด้วยการระบุเพียงแต่ค่าดัชนีที่มีอยู่ใน Code Book ก็สามารถที่จะทำให้ฝั่งรับสังเคราะห์หรือถอดรหัสสัญญาณดังกล่าวได้ ส่งผลให้อัตราบิตข้อมูลที่ใช้้น้อยลง ทั้งนี้การใช้วิธี CELP สามารถลดอัตราบิตข้อมูลในการเข้ารหัสสัญญาณลงได้ต่ำกว่า 10 kbps โดยทุกวันนี้ยังคงมีความพยายามในการปรับปรุงวิธีการเข้ารหัสแบบ CELP นี้เพื่อให้มีอัตราบิตข้อมูลลดลงได้อีกในขณะที่คุณภาพสัญญาณเสียงที่สังเคราะห์ขึ้นที่ภาครับยังดีอยู่ โดยปัจจุบันที่ใช้งานอยู่ในกระทรวงกลาโหมสหรัฐอเมริกา (American Department of Defense) อยู่ที่ 4.8 kbps

จากความแตกต่างของการเข้ารหัสเสียงพูดทั้ง 3 ประเภท ดังกล่าว สามารถแสดงให้เห็นความแตกต่างทางด้านคุณภาพ และอัตราบิตข้อมูลที่ใช้ในการเข้ารหัสแต่ละประเภทได้ดังนี้คือ การเข้ารหัสสัญญาณเสียงพูดแบบ Waveform Coder จะให้คุณภาพเสียงที่ดีที่สุดในขณะเดียวกันจำนวนบิตข้อมูลที่ใช้ก็มากที่สุดด้วย การเข้ารหัสสัญญาณเสียงพูดแบบ Source Coder จะใช้บิตข้อมูลในการแทนค่าสัญญาณน้อยสุดแต่ในขณะเดียวกันคุณภาพเสียงพูดที่ได้ก็มีคุณภาพต่ำสุด ส่วนการเข้ารหัสสัญญาณเสียงแบบ Hybrid Coder จะให้คุณภาพเสียงที่ดีว่าการเข้ารหัสแบบ Source Coder เกือบจะใกล้เคียงกับการเข้ารหัสแบบ Waveform Coder ในขณะที่จำนวนบิตข้อมูลที่ใช้ในการแทนค่าสัญญาณเสียงจะใช้น้อยกว่าแบบ Waveform Coder ดังรูปที่ 3.11

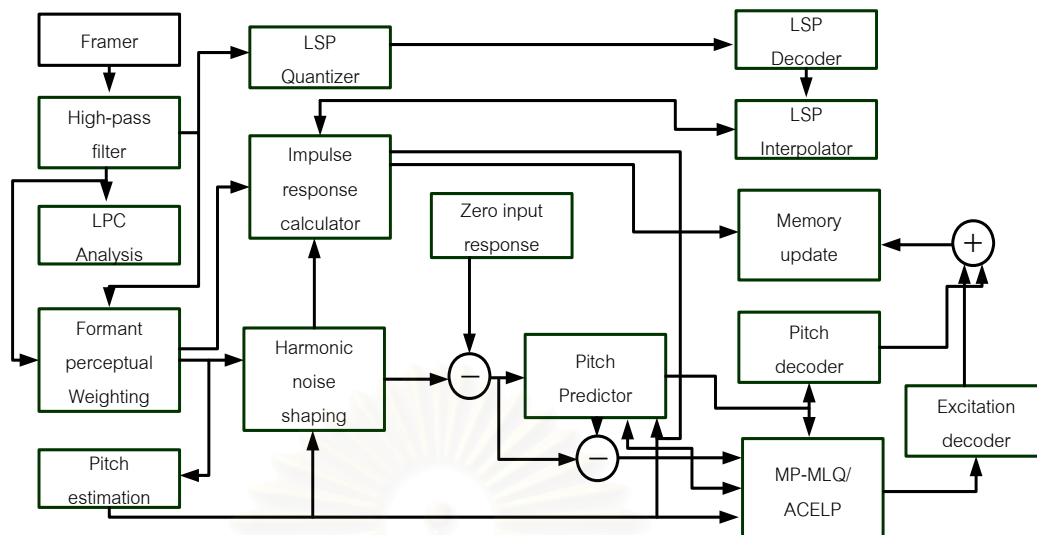


รูปที่ 3.11 เปรียบคุณภาพและจำนวนบิตข้อมูลในการเข้ารหัสเสียงพูดแต่ละแบบ [22]

### 3.3.2 การเข้ารหัสเสียงพูดตามมาตรฐาน G.723.1

เป็นมาตรฐานการเข้ารหัสเสียงพูดที่ถูกออกแบบมาใช้ในระบบการสื่อสารทางเสียงผ่านเครือข่ายอินเทอร์เน็ต [18] เนื่องจากอัตราบิตข้อมูลต่ำ ในขณะที่คุณภาพสัญญาณเสียงยังอยู่ในระดับที่ยอมรับได้ G.723.1 เป็นมาตรฐานการเข้ารหัสเสียงพูดจากสถาบัน ITU-T ที่ใช้วิธีการเข้ารหัสเสียงพูดแบบ Hybrid Coder แล้วใช้รหัส Code Book ในการกระตุ้นการจำลองเสียงให้กับระบบการสร้าง Synthesis Filter เพื่อส่งไปทางภาครับ โดยมาตรฐาน G.723.1 มีอัตราบิตข้อมูลที่ 5.3 kbps โดยใช้ Algebraic Code Excited Linear Prediction (ACELP) กระตุ้นสำหรับสร้าง Synthesis Filter และอัตราบิตข้อมูล 6.3 kbps ใช้วิธี Multi Pulse Maximum Likelihood Quantization ในการสร้าง Synthesis Filter มาตรฐานการเข้ารหัสเสียงพูดแบบ G.723.1 นี้ มักถูกนำไปใช้เป็นตัวเข้ารหัสสัญญาณเสียงพูดร่วมกับมาตรฐาน H.32X [18] ยกตัวอย่างเช่น มาตรฐาน H.324 ที่ใช้ในระบบโทรศัพท์ภาพ (Videophone) และ H.323 ที่ใช้ในระบบการประชุมทางไกล (Teleconferencing System) ผ่านเครือข่ายอินเทอร์เน็ต เป็นต้น

การเข้ารหัสเสียงพูดตามมาตรฐาน G.723.1 จะมีการแบ่งช่วงเข้ารหัสเป็นลักษณะเฟรม (Frame) โดย 1 เฟรมจะมีความยาวข้อมูลจำนวน 240 ตัวอย่าง (Sample) ซึ่งสัญญาณเข้าตามมาตรฐาน G.723.1 ต้องเป็นแบบดิจิตอลที่ผ่านกระบวนการซิกตัวอย่างที่มีความถี่ 8 kHz และแทนค่าข้อมูลด้วยจำนวนบิตข้อมูล 16 บิต ดังนั้นความยาวแต่ละเฟรมในการเข้ารหัสจะอยู่ที่ 30 ms บล็อกไดอะแกรมการเข้ารหัสแสดงดังรูปที่ 3.12



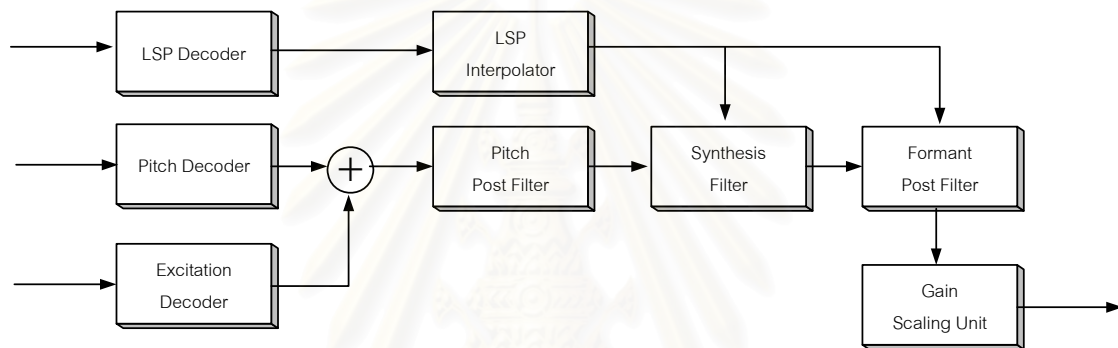
รูปที่ 3.12 บล็อกไดอะแกรมมาตรฐานการเข้ารหัส G.723.1

จากรูปที่ 3.12 สามารถอธิบายลักษณะการเข้ารหัสตามมาตรฐาน G.723.1 ได้ดังต่อไปนี้

1. ข้อมูลจะถูกแบ่งเป็นเฟรมๆ 1 เฟรม มีความยาวข้อมูล 240 ตัวอย่าง แต่ละเฟรมจะถูกแบ่งออกเป็น 4 เฟรมย่อย (Sub Frame) แต่ละเฟรมย่อยมีขนาดข้อมูล 60 ตัวอย่าง
2. ข้อมูลสัญญาณเข้าจะถูกนำเอาส่วนประกอบดีซี (DC-component) ออกก่อนที่จะนำข้อมูลไปผ่านการประมวลผล
3. แต่ละเฟรมย่อยที่มีขนาดข้อมูล 60 ตัวอย่างจะถูกส่งไปคำนวณสัมประสิทธิ์วงจรรอง LPC (LPC filter) ซึ่งจะมีขนาดอันดับ (Order) เป็น 10 ส่วนเฟรมย่อยใดที่ยังไม่ได้มีการคำนวณสัมประสิทธิ์วงจรรอง LPC ก็จะถูกส่งไปหาค่าสัมประสิทธิ์น้ำหนักวงจรรองแบบย่อ (Short Term Perceptual Weighting Filter) เพื่อจะทำการแทนค่าสัญญาณเสียงที่จำลองขึ้น
4. จากนั้นแต่ละ 2 เฟรมย่อยที่มีข้อมูลจำนวน 120 ตัวอย่าง จะถูกส่งไปเพื่อคำนวณช่วงคาบระดับสัญญาณเสียงจากข้อมูลที่ได้ในข้อ 3
5. ค่าช่วงคาบที่คำนวณได้จะถูกนำไปคำนวณขนาดวงจรรอง Harmonic Noise
6. นำค่าต่างๆที่คำนวณได้ ส่งไปยังส่วนของการคำนวณผลตอบลัพท์อิมพัลส์ (Impulse Response) ซึ่งจะนำไปใช้ในการคำนวณในส่วนอื่นๆต่อไป

7. สัญญาณที่ได้จากการประมาณในส่วนการทำนายพิตช์ (Pitch Predictor) จะถูกนำไปเปรียบเทียบกับผลต่างกับข้อมูลเสียงที่ได้จากข้อ 6 เพื่อส่งไปทำการแทนค่าสัญญาณแบบ Multi Pulse Maximum Likelihood Quantization (MP-MLQ) สำหรับอัตราข้อมูลบิตที่ 6.3 kbps และ Algebraic Code Excited Linear Prediction (ACELP) สำหรับอัตราข้อมูลบิตที่ 5.3 kbps

ข้อมูลที่ถูกส่งออกไปจากส่วนของการเข้ารหัสจะเป็นลักษณะข้อมูลที่เป็นกระแสบิต (Bit Stream) ส่งผ่านไปในช่วงสัญญาณสื่อสาร ทางด้านผู้รับเมื่อได้รับข้อมูลที่เป็นกระแสบิตแล้ว ต้องนำไปถอดรหัสสัญญาณข้อมูล ซึ่งส่วนของการถอดรหัสข้อมูลเสียงพูดตามมาตรฐาน G.723.1 ก็อธิบายการทำงานตามบล็อกไดอะแกรมดังรูปที่ 3.13 นี้



รูปที่ 3.13 บล็อกไดอะแกรมการถอดรหัสสัญญาณเสียงตามมาตรฐาน G.723.1

ส่วนของการถอดรหัสเสียงพูดตามมาตรฐาน G.723.1 ทำได้ด้วยการถอดรหัสข้อมูลของสัมประสิทธิ์ฟังก์ชันวงจรรอง ทำให้ได้ค่าของ Linear Spectral Pair Related Function (LSP) ข้อมูลช่วงคาบ (Period) ของระดับสัญญาณพิตช์ (Pitch) และรหัส Code Book เพื่อส่งต่อไปยังส่วนของวงจรรองปรับตัวในการประมาณช่วงคาบสัญญาณเสียงพูดและการกำหนดขนาดอัตราส่วน (Gain) กับสัญญาณจริง เพื่อส่งให้ผู้ฟังเป็นลักษณะสัญญาณแบบดิจิทัลเช่นเดียวกับสัญญาณเข้าต่อไป

จุฬาลงกรณ์มหาวิทยาลัย

## บทที่ 4

### ผลการทดลองและการวิเคราะห์ผล

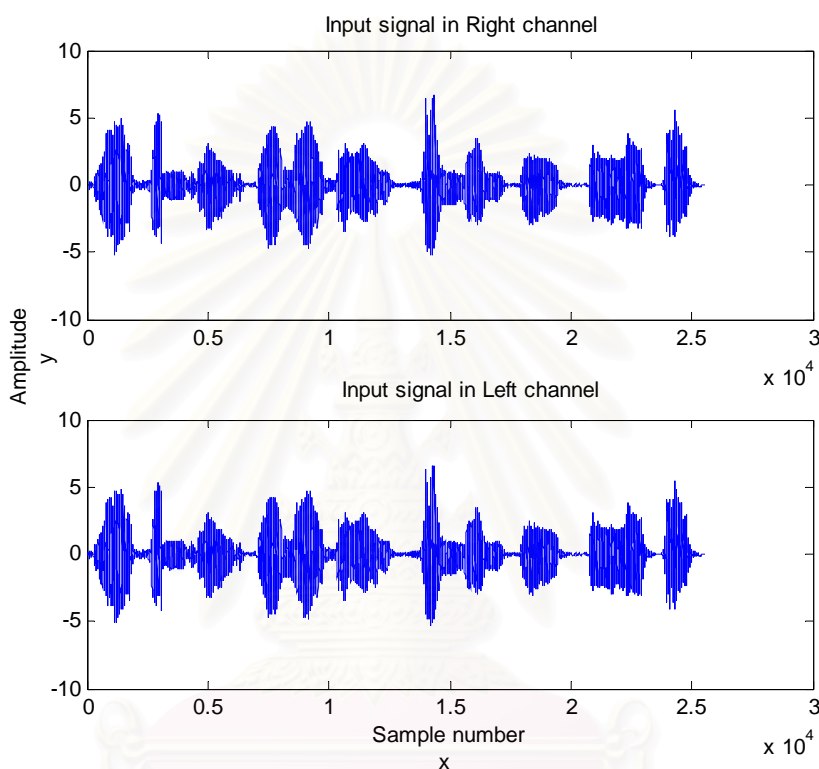
ในบทนี้ได้ทำการทดสอบผลการจำลองการทำงานของระบบ SAEC ในกรณีต่างๆ ที่กล่าวไว้ในบทที่ 3 ซึ่งในการทดสอบและวัดประสิทธิภาพการทำงานของระบบจะใช้ค่าต่างๆ ที่ได้กล่าวไว้แล้วในหัวข้อที่ 3.2 โดยแบ่งการจำลองการทำงานของระบบเป็นกรณีต่างๆ คือ ลักษณะสัญญาณเสียงสเตอริโอที่นำมาใช้ในการทดลอง วิธีสะท้อนทางเสียงที่จำลองขึ้น การเลือกใช้อัตราการสุ่มเข้าของขั้นตอนวิธีที่นำมาใช้ พารามิเตอร์ต่างๆ และผลการจำลองการทำงานของระบบ SAEC เมื่อยังไม่ได้มีการดีคอดิเคชันสัญญาณเสียงสเตอริโอ จะกล่าวถึงในหัวข้อที่ 4.1 หลังจากนั้นทำการดีคอดิเคชันสัญญาณเสียงสเตอริโอด้วยวิธีการเพิ่มสัญญาณรบกวนสุ่ม ซึ่งจะกล่าวในหัวข้อที่ 4.2 และวิธีดีคอดิเคชันสัญญาณเสียงสเตอริโอโดยเพิ่มภาวะไม่เชิงเส้นจะกล่าวถึงในหัวข้อที่ 4.3 สำหรับหัวข้อที่ 4.4 จะกล่าวถึงผลการจำลองการทำงานของระบบ SAEC ตามแนวคิดที่นำเสนอคือการเพิ่มสัญญาณรบกวนแบบปรับตัวได้ที่เปลี่ยนตามลักษณะสัญญาณเข้าในแต่ละช่องสัญญาณตามหัวข้อที่ 3.1.3 จนสามารถนำมาประยุกต์ใช้ในการเพิ่มประสิทธิภาพการทำงานของระบบ SAEC นำผลที่ได้เปรียบเทียบกับในแง่ของประสิทธิภาพในการกำจัดสัญญาณเสียงสะท้อนเมื่อสัญญาณเข้าแบบสเตอริโอถูกลดสหสัมพันธ์ลงไปจากเดิมกับวิธีในข้อที่ 4.2 และข้อที่ 4.3 นอกจากนั้นจะได้ศึกษาผลกระทบต่อคุณภาพสัญญาณเสียงสเตอริโอที่ผ่านการดีคอดิเคชันโดยวิธีต่างๆ ผ่านทาง subjective listening test ในหัวข้อที่ 4.5 และลำดับสุดท้ายที่ในบทนี้จะกล่าวถึงการทดสอบการทำงานของระบบ SAEC กรณีที่ต้องนำไปใช้งานร่วมกับมาตรฐานการเข้ารหัสเสียงพูด G.723.1 ในหัวข้อที่ 4.6 ทั้งนี้ได้ทำการเปรียบเทียบผลที่เกิดขึ้นกับประสิทธิภาพการทำงานของระบบ SAEC เมื่อมีการเข้ารหัสสัญญาณเสียงตามมาตรฐานข้างต้น เปรียบเทียบกับระบบ SAEC ที่ไม่ได้มีการเข้ารหัสสัญญาณเสียงก่อนที่จะป้อนให้ระบบ และประสิทธิภาพการทำงานของระบบ SAEC เมื่อนำเอาวิธีการดีคอดิเคชันสัญญาณที่นำเสนอไปใช้ร่วมด้วยกับการเข้ารหัสสัญญาณเสียงตามมาตรฐานดังกล่าว

#### 4.1 การจำลองการทำงานของระบบ SAEC

ในหัวข้อนี้ได้แสดงผลการจำลองการทำงานของระบบการกำจัดสัญญาณเสียงสะท้อนแบบสเตอริโอดังแสดงในรูปที่ 3.1 ในขณะที่ยังไม่ได้มีการดีคอดิเคชันสัญญาณในบล็อก PP โดยในการจำลองระบบนี้ยังไม่คำนึงถึงการส่งสัญญาณผ่านช่องสัญญาณสื่อสาร จึงสมมุติให้ไม่มีความ

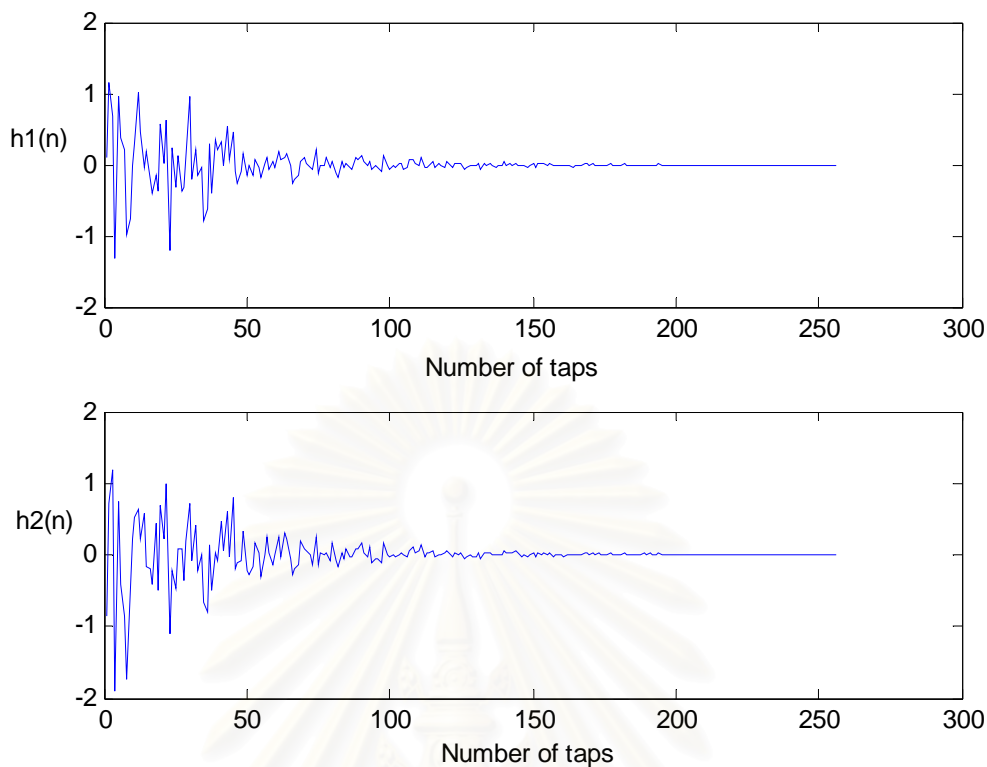


ผิดพลาดเกิดขึ้นระหว่างการส่งสัญญาณไปในช่องสัญญาณสื่อสาร เริ่มจากการนำสัญญาณเสียงสเตอริโอที่ส่งมาจากห้องส่ง  $x_1(n)$  และ  $x_2(n)$  ผ่านเข้าไปในระบบ SAEC แล้วจำลองการกำจัดสัญญาณเสียงสะท้อน ทดสอบประสิทธิภาพการทำงานของระบบจากค่า WEVN เมื่อให้สัญญาณเสียงพูดแบบสเตอริโอที่ใช้ในการทดลองตลอดทุกการทดลองสามารถแสดงได้ดังรูปที่ 4.1 คือ “ราชดำริ ศาลาแดง ชื่องนนทรี สุรศักดิ์”



รูปที่ 4.1 สัญญาณเสียงพูดแบบสเตอริโอ ที่ใช้ในการจำลองการทำงานของระบบ SAEC

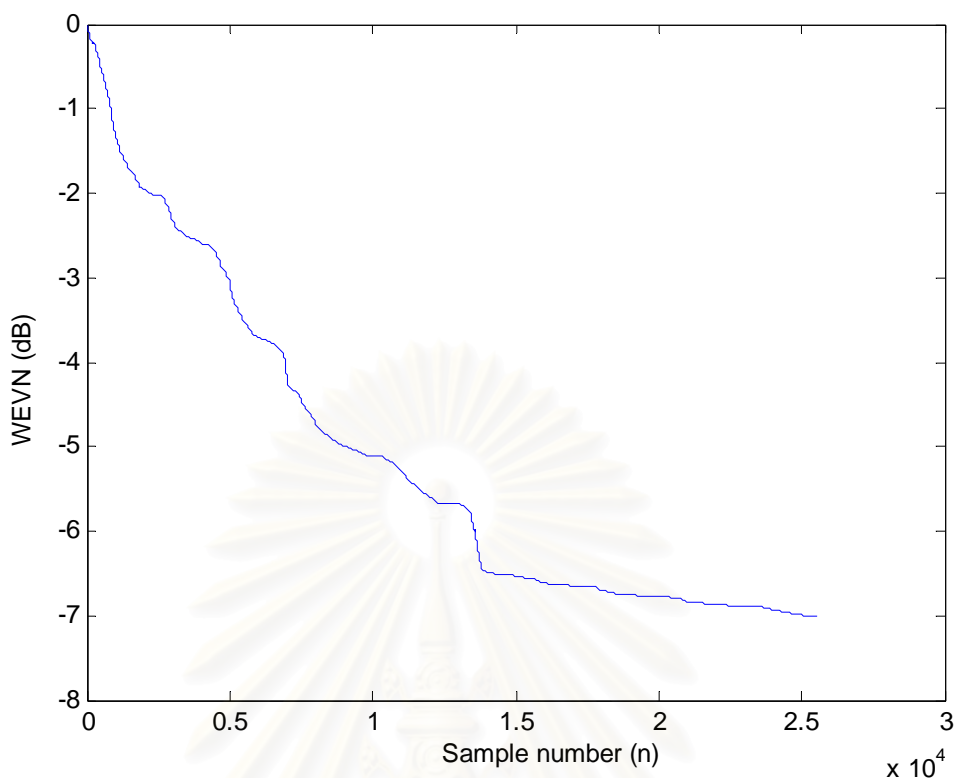
ในรูปที่ 4.1 สัญญาณเสียงที่ใช้ในการจำลองระบบเป็นสัญญาณที่มีความถี่ชักตัวอย่างเท่ากับ 8,000 Hz แทนค่าสัญญาณข้อมูล 16 บิตต่อ 1 ตัวอย่าง จำนวนสัญญาณที่สุ่มมีขนาด 25,537 ตัวอย่าง นอกจากนี้ในการจำลองระบบได้สมมติให้วิถีสะท้อนทางเสียงในห้องรับมีความยาว  $L$  เท่ากับ 256 และเป็นแบบเชิงเส้นไม่เปลี่ยนแปลงตามเวลา แสดงให้เห็นดังรูปที่ 4.2 เมื่อ  $h_1(n)$  เป็นวิถีสะท้อนทางเสียงที่เกิดระหว่างไมโครโฟนและลำโพงคู่ที่ 1 และ  $h_2(n)$  เป็นวิถีสะท้อนทางเสียงระหว่างไมโครโฟนและลำโพงคู่ที่ 2 โดยจะถูกนำไปใช้เพื่อศึกษาการทำงานของระบบ SAEC ตลอดทุกการทดลอง



รูปที่ 4.2 วิธีสะท้อนทางเสียงในห้องรับที่จำลองขึ้น มีขนาดความยาว  $L=256$  ไม่เปลี่ยนแปลงตามเวลา เมื่อ  $h_1(n)$  เป็นวิธีสะท้อนทางเสียงที่จำลองขึ้นระหว่างไมโครโฟนและลำโพงคู่ที่ 1 และ  $h_2(n)$  เป็นวิธีสะท้อนทางเสียงที่จำลองขึ้นระหว่างไมโครโฟนและลำโพงคู่ที่ 2

การจำลองการทำงานของระบบ SAEC ได้เลือกใช้ขั้นตอนวิธี NLMS ในการควบคุมการทำงานของวงจรรองแบบปรับตัว ที่ใช้สมการปรับปรุงการทำงานตามสมการที่ (2.25) ค่าที่ใช้ในการควบคุมอัตราการลู่เข้าของขั้นตอนวิธีนี้คือค่า  $\tilde{\mu}$  ซึ่งในการจำลองระบบนี้ได้เลือกใช้ค่า  $\tilde{\mu}=0.8$  ซึ่งเป็นค่าที่ทำให้ระบบ SAEC มีการทำงานได้เหมาะสม คือ มีอัตราการลู่เข้าที่ดีในขณะที่ยังสามารถรักษาเสถียรภาพของระบบได้ดีนั่นเอง และได้เลือกค่า  $\epsilon$  ให้เป็นค่าคงที่น้อยๆ เพื่อป้องกันการสูญเสียเสถียรภาพของระบบ หากว่าเมื่อสัญญาณเข้ามีขนาดน้อยๆ โดยเลือกให้  $\epsilon$  เท่ากับ 0.1

จากการจำลองการทำงานของระบบ SAEC ในกรณีที่ยังไม่ได้ทำการดีคอรรีเลชันสัญญาณเสียงสเตอริโอ  $x_1(n)$  และ  $x_2(n)$  ตามรูปที่ 3.1 พบว่ากราฟค่า WEVN ของระบบ SAEC เป็นดังรูปที่ 4.3 โดยมีค่า WEVN ในสถานะคงตัว (Steady State) อยู่ที่ประมาณ -7 dB

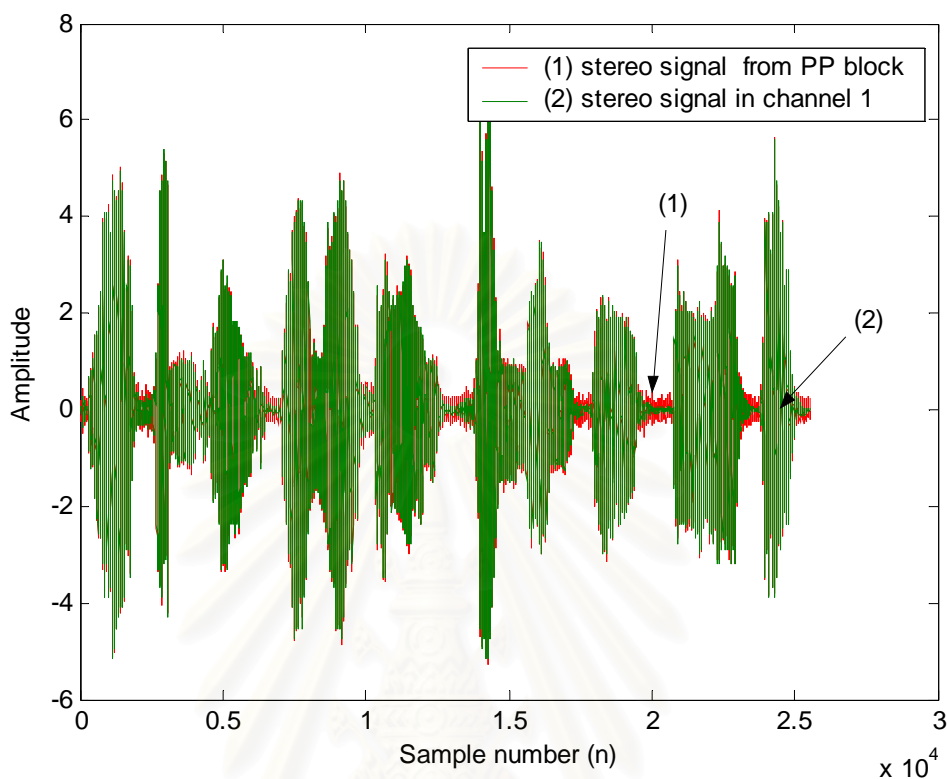


รูปที่ 4.3 ค่า WEVN ในระบบ SAEC ในขณะที่ยังไม่ได้มีการดีคอดรีเลชันสัญญาณ

#### 4.2 การจำลองการทำงานของระบบ SAEC เมื่อมีการดีคอดรีเลชันสัญญาณเสียงสเตอริโอ โดยเพิ่มสัญญาณรบกวนสุ่ม

การศึกษาค่าผลของการดีคอดรีเลชันของสัญญาณเสียงสเตอริโอโดยการเพิ่มสัญญาณรบกวนสุ่ม ทำได้ด้วยการเพิ่มสัญญาณรบกวนตามหัวข้อ 3.1.1 เข้าไปในระบบตามรูปที่ 3.2 เพื่อลดสหสัมพันธ์ของสัญญาณเข้าแบบสเตอริโอ โดยเริ่มจากการเพิ่มสัญญาณรบกวนสุ่มเพียงช่องสัญญาณเดียวก่อน แล้วตรวจสอบสมรรถนะของระบบ SAEC หลังจากนั้นค่อยๆเพิ่มสัญญาณรบกวนที่อีกช่องสัญญาณหนึ่ง เปรียบเทียบประสิทธิภาพการทำงานของระบบ SAEC และตรวจสอบคุณภาพเสียงสเตอริโอที่ผ่านการดีคอดรีเลชันแล้ว พบว่าการเพิ่มสัญญาณรบกวนสุ่มที่ช่องสัญญาณเดียวทำให้ระบบ SAEC มีประสิทธิภาพการทำงานของระบบได้ด้อยกว่าเช่นเดียวกันคุณภาพเสียงสเตอริโอที่ได้หลังจากดีคอดรีเลชันแล้วก็มีคุณภาพด้อยกว่าการเพิ่มสัญญาณรบกวนสุ่มทั้ง 2 ช่องสัญญาณ ดังนั้นวิทยานิพนธ์นี้จึงเลือกทำการเพิ่มสัญญาณรบกวนสุ่มทั้ง 2 ช่องสัญญาณเพื่อดีคอดรีเลชันสัญญาณเสียงสเตอริโอ สามารถแสดงผลการเปลี่ยนแปลงของสัญญาณใหม่ดังรูปที่ 4.4 ซึ่งเป็นตัวอย่างของลักษณะสัญญาณเสียงแบบสเตอริโอที่

ช่องสัญญาณหนึ่งที่เปลี่ยนแปลงไปเมื่อทำการเพิ่มสัญญาณรบกวนสุ่มที่เลือกค่า SNR เท่ากับ 20 dB

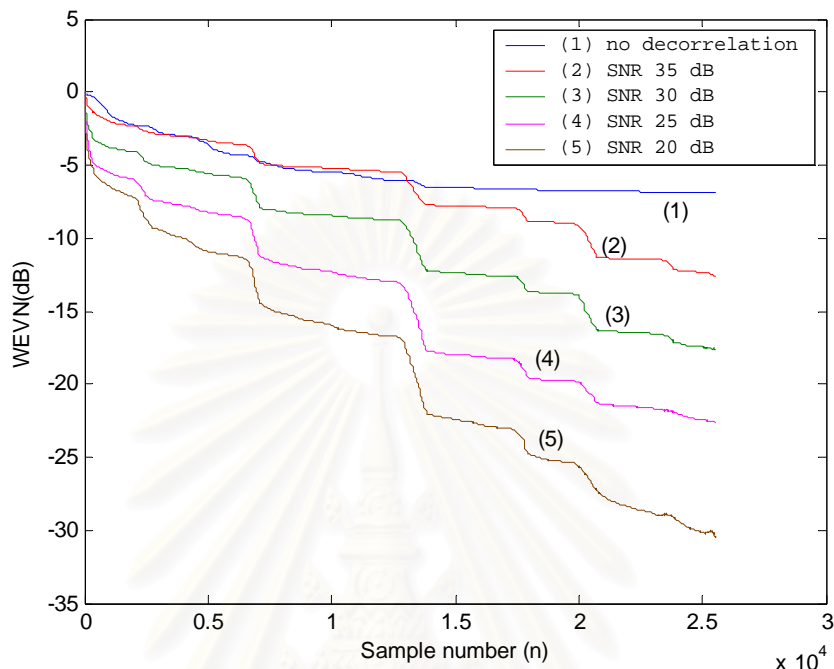


รูปที่ 4.4 ตัวอย่างสัญญาณที่ถูกดีคอดริเลชันโดยเพิ่มสัญญาณรบกวนสุ่มเมื่อ SNR=20 dB

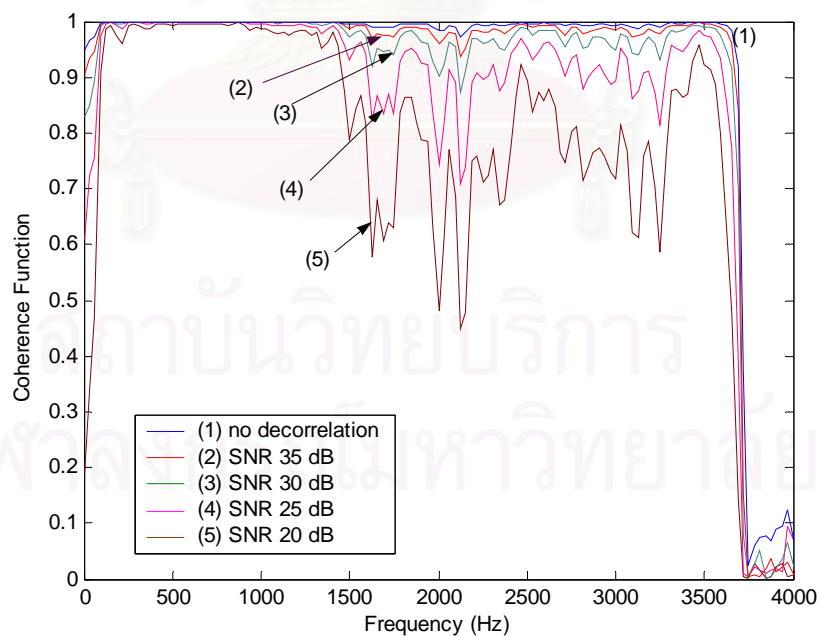
จากการทดลองเมื่อทำการเปลี่ยนแปลงค่า SNR ของสัญญาณรบกวนสุ่มที่เพิ่มเข้าไปในระบบ พบว่าวิธีนี้สามารถดีคอดริเลชันสัญญาณเสียงสเตอริโอได้ค่อนข้างดี จึงทำให้ระบบ SAEC มีประสิทธิภาพการทำงานที่ดี โดยสังเกตจากค่า WEVN ที่ลดลงได้ค่อนข้างมาก ดังรูปที่ 4.5 (a) นอกจากนี้ ค่าฟังก์ชันร่วมนัย (Coherence Function) ที่แสดงถึงสหสัมพันธ์ของสัญญาณเสียงสเตอริโอก็มีค่าลดลงจากเดิมดังรูปที่ 4.5 (b) เมื่อค่า SNR ลดลง

ในการทดลองได้ทำการปรับระดับสัญญาณรบกวนสุ่มที่ใส่ให้แก่ระบบตามค่า SNR ของแต่ละช่องสัญญาณเป็น 35, 30, 25 และ 20 dB โดยพบว่าค่า WEVN และค่า coherence function มีค่าลดลงเมื่อทำการเพิ่มสัญญาณรบกวนสุ่ม ตามกราฟเส้นที่ (2) – (5) และมีแนวโน้มว่าจะลดลงได้อีกถ้าเพิ่มระดับสัญญาณรบกวนสุ่มให้มากขึ้นอีก หรือเมื่อค่า SNR มีค่าลดลงเรื่อยๆ ซึ่งหมายถึงสหสัมพันธ์ของสัญญาณเสียงสเตอริโอก็จะยิ่งลดลงอีกเรื่อยๆ อย่างไรก็ตาม การเพิ่มระดับสัญญาณรบกวนสุ่มก็มีข้อจำกัดในเรื่องของคุณภาพเสียงของสัญญาณเสียงสเตอริโอ โดยผู้ฟังทางห้องรับจะได้ยินเสียงรบกวนเมื่อสัญญาณรบกวนสุ่มที่เพิ่มเข้าไปในระบบมีระดับมากขึ้น ดังนั้นการดีคอดริเลชันสัญญาณเสียงสเตอริโอโดยใช้สัญญาณรบกวนสุ่มนี้ จึงถูกจำกัดระดับการเพิ่ม

สัญญาณรบกวนไว้ที่ประมาณ 20 dB เพื่อให้วงจรระบบปรับตัวมีการลู่เข้าสู่ค่าตอบได้ดี โดยการเพิ่มสัญญาณรบกวนสุ่มที่ค่า SNR 20 dB ก็เริ่มเกิดผลกระทบต่อคุณภาพเสียงของสัญญาณที่ถูกดีโคอริเลชันแล้ว



รูปที่ 4.5 (a)

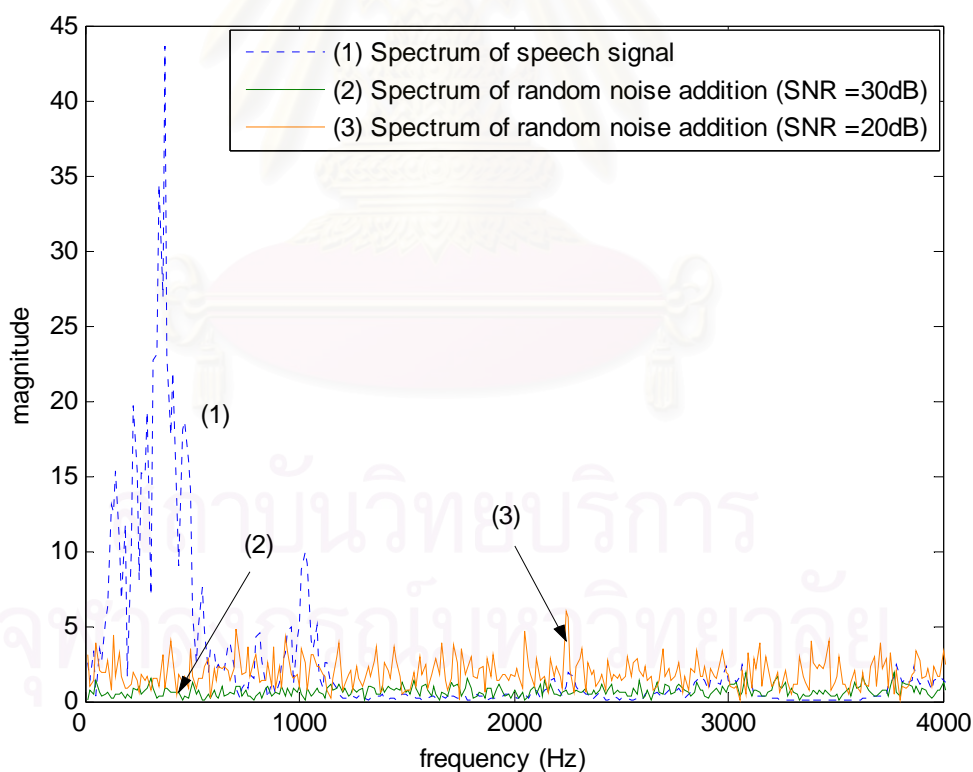


รูปที่ 4.5 (b)

รูปที่ 4.5 (a) ค่า WEVN ของระบบ SAEC และ (b) Coherence Function

ของสัญญาณเสียงสเตอริโอ เมื่อมีการเพิ่มสัญญาณรบกวนสุ่ม

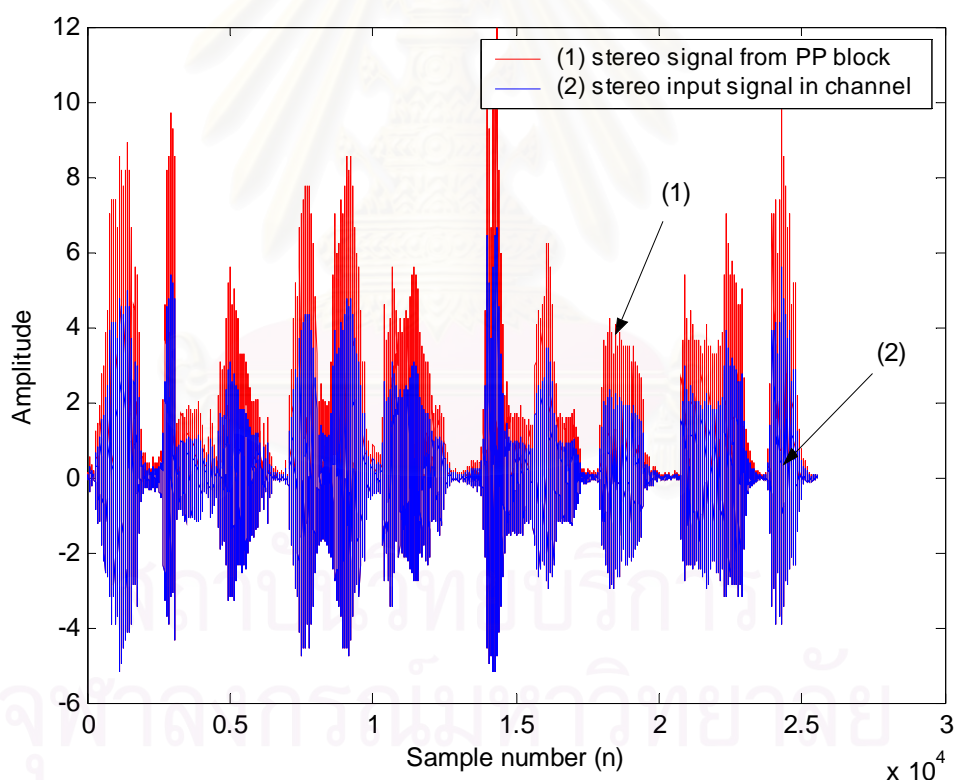
การเพิ่มสัญญาณรบกวนสุ่มในสัญญาณเสียงสเตอริโอส่งผลให้ระบบ SAEC ทำงานได้ค่อนข้างดี แต่ขณะเดียวกันก็ถูกจำกัดขนาดสัญญาณรบกวนสุ่ม ที่ค่า SNR ไม่ควรต่ำกว่า 20 dB ทั้งนี้หากเราพิจารณาความหนาแน่นกำลังเชิงสเปกตรัม (Power Spectral Density) ของสัญญาณรบกวนสุ่ม เปรียบเทียบกับสัญญาณเสียงพูดในโดเมนความถี่ดังรูปที่ 4.6 ความหนาแน่นกำลังเชิงสเปกตรัมของสัญญาณรบกวนสุ่มจะมีค่าโดยเฉลี่ยเท่าๆกันในทุกๆความถี่ เมื่อเพิ่มขนาดสัญญาณรบกวนสุ่มจากค่า SNR เท่ากับ 30 dB เป็น 20 dB จะเห็นได้ว่าขนาดความหนาแน่นกำลังเชิงสเปกตรัมสัญญาณรบกวนสุ่มจะเพิ่มขึ้นจากกราฟเส้นที่ (2) เป็นกราฟเส้นที่ (3) อย่างเท่าๆกันในทุกๆความถี่ ซึ่งแสดงให้เห็นว่าการดีโคอริเลชันสัญญาณเสียงสเตอริโอด้วยวิธีนี้ถึงแม้จะส่งผลให้ระบบ SAEC ทำงานได้ดี แต่ก็ต้องถูกจำกัดระดับสัญญาณรบกวนสุ่มไว้ที่ระดับหนึ่งเท่านั้นเพราะผลกระทบของสัญญาณรบกวนที่มากกว่า 20 dB จะส่งผลให้มีคุณภาพเสียงที่ด้อยลง โดยเฉพาะช่วงความถี่สูงที่ขนาดความหนาแน่นกำลังเชิงสเปกตรัมของสัญญาณรบกวนมีค่ามากกว่าความหนาแน่นกำลังเชิงสเปกตรัมของสัญญาณเสียงพูดนั่นเอง



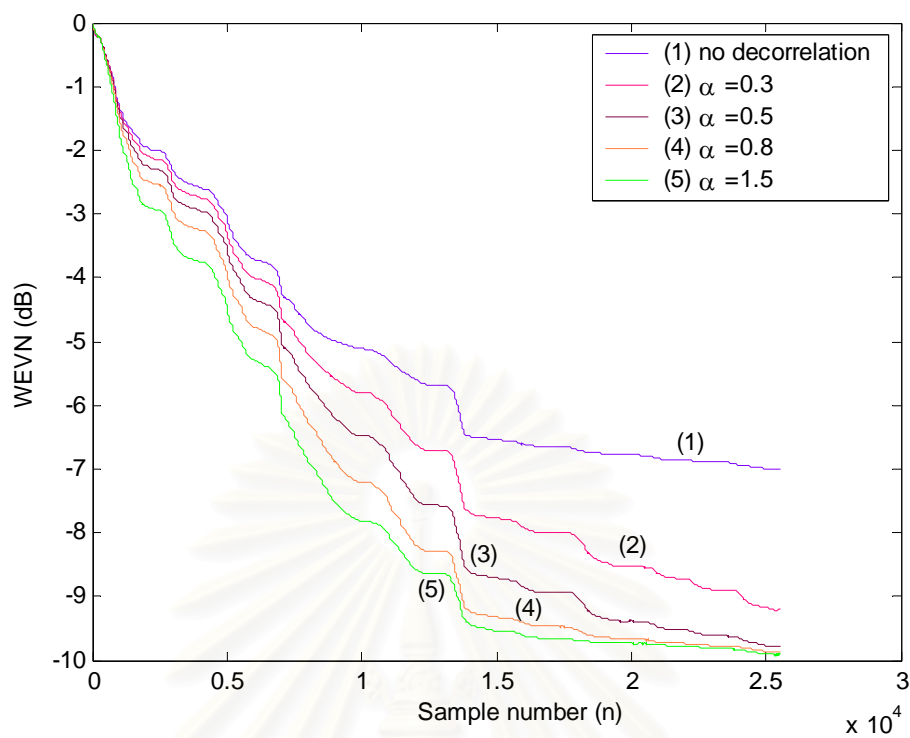
รูปที่ 4.6 เปรียบเทียบความหนาแน่นกำลังเชิงสเปกตรัมสัญญาณรบกวนสุ่ม  
กับความหนาแน่นกำลังเชิงสเปกตรัมสัญญาณเสียงพูด

### 4.3 การจำลองการทำงานของระบบ SAEC เมื่อดีคอรিলักษณ์สัญญาณเสียงสเตอริโอ โดยเพิ่มภาวะไม่เชิงเส้น

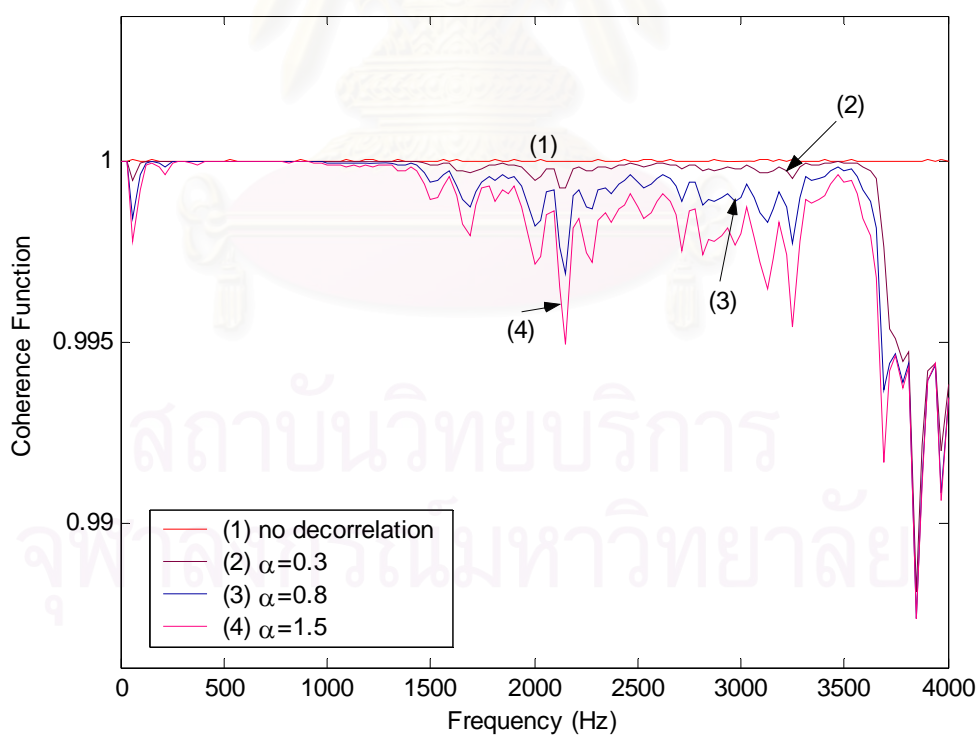
ในการทดลองศึกษาผลของการดีคอรিলักษณ์ของสัญญาณเสียงสเตอริโอ ด้วยการภาวะไม่เชิงเส้นตามหัวข้อที่ 3.1.2 ค่าการเปลี่ยนแปลงภาวะไม่เชิงเส้นของสัญญาณเข้าแบบสเตอริโอจะถูกกำหนดด้วยค่า  $\alpha$  ตามสมการที่ (3.2) เมื่อเปลี่ยนแปลงค่า  $\alpha$  ก็จะส่งผลให้ภาวะไม่เชิงเส้นของสัญญาณเสียงสเตอริโอมีค่าเปลี่ยนแปลงไป ส่งผลให้การดีคอรিলักษณ์สัญญาณเสียงสเตอริโอมีการเปลี่ยนแปลงด้วย ดังรูปที่ 4.7 มีการเลือกค่า  $\alpha = 0.8$  สัญญาณเสียงสเตอริโอที่ผ่านบล็อก PP จะมีลักษณะแตกต่างไปจากเดิม สำหรับผลของการเปลี่ยนแปลงค่า  $\alpha$  ที่มีต่อการสมรรถนะการทำงานของระบบ SAEC แสดงผ่านทางกราฟ WEVN ดังรูปที่ 4.8 (a) และจากค่า Coherence Function ดังรูปที่ 4.8 (b)



รูปที่ 4.7 ตัวอย่างสัญญาณที่ถูกดีคอริลักษณ์เมื่อเพิ่มภาวะไม่เชิงเส้น เมื่อ  $\alpha=0.8$



รูปที่ 4.8 (a)



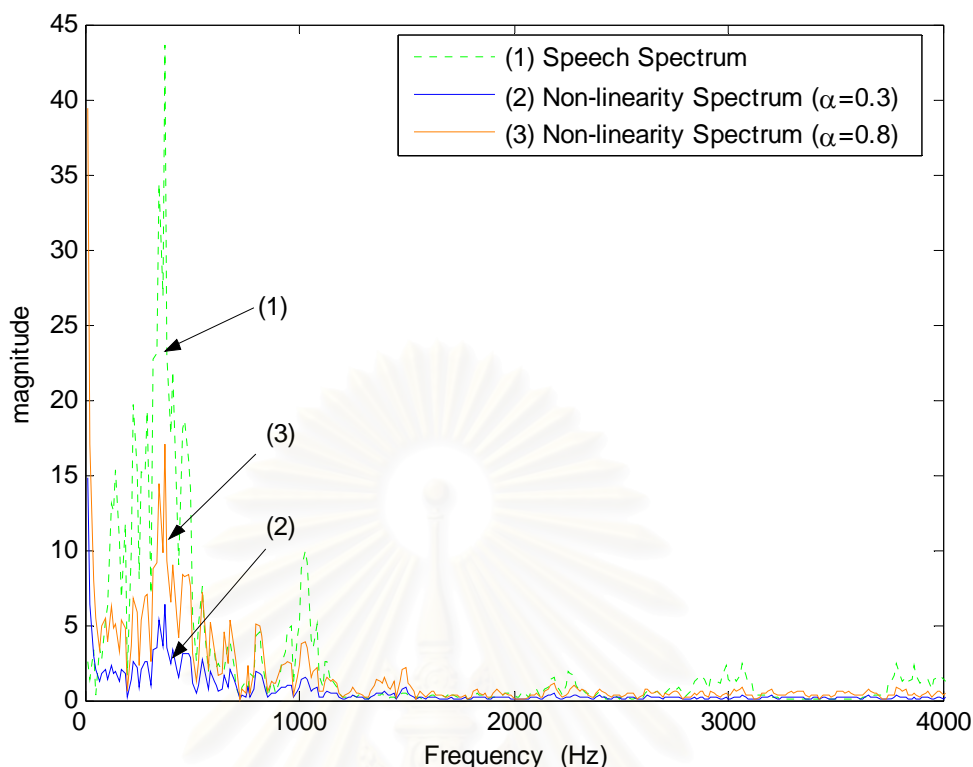
รูปที่ 4.8 (b)

รูปที่ 4.8 (a) ค่า WEVN ของระบบ SAEC และ (b) Coherence Function  
ของสัญญาณเสียงสเตอริโอ เมื่อเพิ่มภาวะไม่เชิงเส้น



จากกราฟ WEVN ในรูปที่ 4.8 (a) เมื่อเราเพิ่มภาวะไม่เชิงเส้นผ่านทางการปรับค่าพารามิเตอร์  $\alpha$  ให้แก่สัญญาณเสียงสเตอริโอปริมาณน้อยๆ สหสัมพันธ์ของสัญญาณก็จะเริ่มเปลี่ยนไป เมื่อให้  $\alpha$  เท่ากับ 0.3, 0.5, 0.8 และ 1.5 ตามลำดับ กราฟ WEVN จะเป็นดังเส้นที่ (2), (3), (4) และ (5) ทั้งนี้ค่า  $\alpha$  เท่ากับ 1.5 ระบบ SAEC สามารถจำลองการทำงานของวงจรกรองแบบปรับตัวเพื่อกำจัดสัญญาณเสียงสะท้อนแบบสเตอริโอได้ดีที่สุดในขณะที่ยังไม่ก่อให้เกิดการรบกวนต่อคุณภาพเสียงสเตอริโอ อย่างไรก็ตามพบว่า ค่า WEVN ในการดีคอดรีเลชันของสัญญาณเสียงสเตอริโอโดยการเพิ่มภาวะไม่เชิงเส้นมีค่าลดลงได้ประมาณ  $-9$  dB เท่านั้น สำหรับ สหสัมพันธ์ของสัญญาณเสียงสเตอริโอ สังเกตได้จากกราฟ Coherence Function ดังรูปที่ 4.8 (b) เมื่อค่า  $\alpha$  เพิ่มขึ้นเรื่อยๆ สหสัมพันธ์ของสัญญาณเสียงสเตอริโอก็จะลดลงเรื่อยๆ ตามกราฟเส้นที่ (2) – (4) เมื่อ  $\alpha$  เป็น 0.3, 0.8 และ 1.5

การดีคอดรีเลชันโดยการเพิ่มภาวะไม่เชิงเส้นช่วยให้ประสิทธิภาพการทำงานของระบบ SAEC ดีขึ้นเพียงเล็กน้อยเมื่อเปรียบกับวิธีในข้อ 4.2 แต่เมื่อพิจารณาถึงคุณภาพเสียงสเตอริโอที่ผ่านการดีคอดรีเลชันแล้ว กลับพบว่าวิธีนี้สามารถรักษาคุณภาพเสียงได้ใกล้เคียงกับเสียงสเตอริโอเดิมค่อนข้างมาก ทั้งนี้หากพิจารณาความหนาแน่นกำลังเชิงสเปกตรัมของสัญญาณภาวะไม่เชิงเส้นที่เพิ่มกับเสียงสเตอริโอเดิมในโดเมนความถี่ดังแสดงในรูปที่ 4.9 จะมีลักษณะคล้ายกับความหนาแน่นกำลังเชิงสเปกตรัมของสัญญาณเสียงพูด เมื่อเพิ่มค่า  $\alpha$  ให้มากขึ้น จาก 0.3 เป็น 0.8 พบว่าขนาดความหนาแน่นกำลังเชิงสเปกตรัมของสัญญาณภาวะไม่เชิงเส้นจะเพิ่มขึ้นในช่วงความถี่ที่มีความหนาแน่นกำลังเชิงสเปกตรัมของสัญญาณเสียงพูดเท่านั้น แสดงให้เห็นว่าคุณภาพเสียงสเตอริโอเดิมจะยังไม่ถูกรบกวนจากการเพิ่มภาวะไม่เชิงเส้นมากนัก ซึ่งมีลักษณะคล้ายหลักของการบังความถี่ (Frequency Masking) นั่นเอง

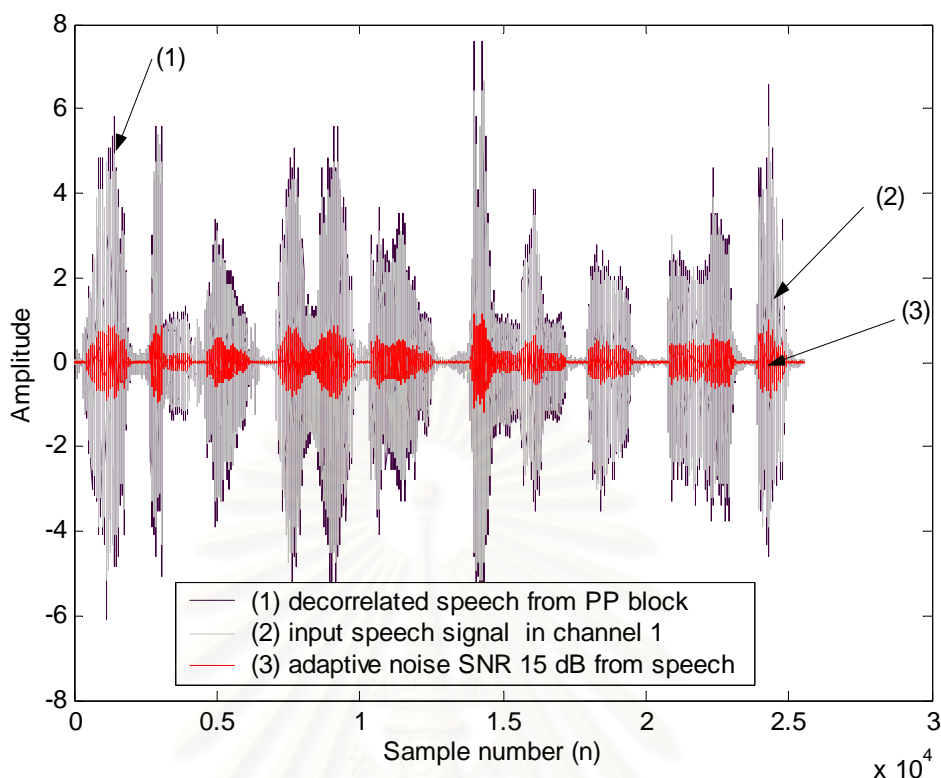


รูปที่ 4.9 เปรียบเทียบความหนาแน่นกำลังเชิงสเปกตรัมสัญญาณภาวะไม่เชิงเส้น  
กับความหนาแน่นกำลังเชิงสเปกตรัมสัญญาณเสียงพูด

#### 4.4 การจำลองการทำงานของระบบ SAEC เมื่อมีการดีคอรিলักษณ์สัญญาณเสียงสเตอริโอ โดยเพิ่มสัญญาณรบกวนแบบปรับตัวได้ ซึ่งเป็นวิธีที่นำเสนอในวิทยานิพนธ์นี้

ในการจำลองการทำงานของระบบ SAEC ที่มีการดีคอรিলักษณ์สัญญาณเสียงสเตอริโอด้วยการเพิ่มสัญญาณรบกวนแบบปรับตัวได้ โดยให้ค่าจุดเริ่มเปลี่ยน (Threshold) ในการทดลองเป็น 0.2 เนื่องจากจุดเริ่มเปลี่ยนดังกล่าวเป็นค่าที่มีความเหมาะสมที่สามารถทำการแยกแยะสัญญาณเสียงพูดได้ว่า ช่วงใดเป็นคาบที่สัญญาณเสียงพูด และช่วงใดเป็นคาบเงียบได้ชัดเจน ดังรูปที่ 3.5 โดยขนาดของสัญญาณรบกวนแบบปรับตัวได้ที่เพิ่มในแต่ละช่องสัญญาณตามสมการที่ (3.9) ( $\Gamma_i(n)$ ) จะมีการเปลี่ยนไปตามลักษณะของสัญญาณเสียงพูดที่เปลี่ยนไปตามเวลา โดยมีลำดับการจำลองการทำงานของระบบ SAEC เป็นไปตามแนวคิดที่ได้เสนอในหัวข้อที่ 3.1.3.2

เมื่อทำการดีคอรিলักษณ์สัญญาณเสียงสเตอริโอด้วยสัญญาณรบกวนที่ปรับตัวได้ตามอัตราส่วนสัญญาณเสียงพูดต่อสัญญาณรบกวนขาวแบบเกาส์ ดังสมการที่ (3.10) และ (3.11) สัญญาณรบกวนแบบปรับตัวได้จะมีสัดส่วนที่เปลี่ยนแปลงไปตามสัญญาณเสียงพูดดังแสดงในกราฟรูปที่ 4.10



รูปที่ 4.10 ตัวอย่างพลังงานเสียงสเตอริโอในช่องสัญญาณหนึ่งเมื่อเพิ่มสัญญาณรบกวนแบบปรับตัวได้ เมื่อ (1) เป็นสัญญาณรวมของสัญญาณเสียงสเตอริโอช่องที่ 1 ตามกราฟเส้นที่ (2) และสัญญาณรบกวนที่ปรับตัวได้ตามกราฟเส้นที่ (3) (2) เป็นสัญญาณเสียงสเตอริโอในช่องที่ 1 และ (3) เป็นสัญญาณรบกวนแบบปรับตัวได้ตาม SNR ของสัญญาณเสียงในข้อ 3.1.3.2

จากกราฟต่างๆในรูปที่ 4.10 สังเกตได้ว่า ในคาบเงียบ สัญญาณรบกวนแบบปรับตัวได้จะไม่ถูกเพิ่มเข้าไปในสัญญาณเสียงพูดในช่วงนั้นๆ ซึ่งเป็นข้อดีที่ไม่ทำให้เกิดผลกระทบต่อคุณภาพสัญญาณเสียงให้ด้อยลงมากนัก

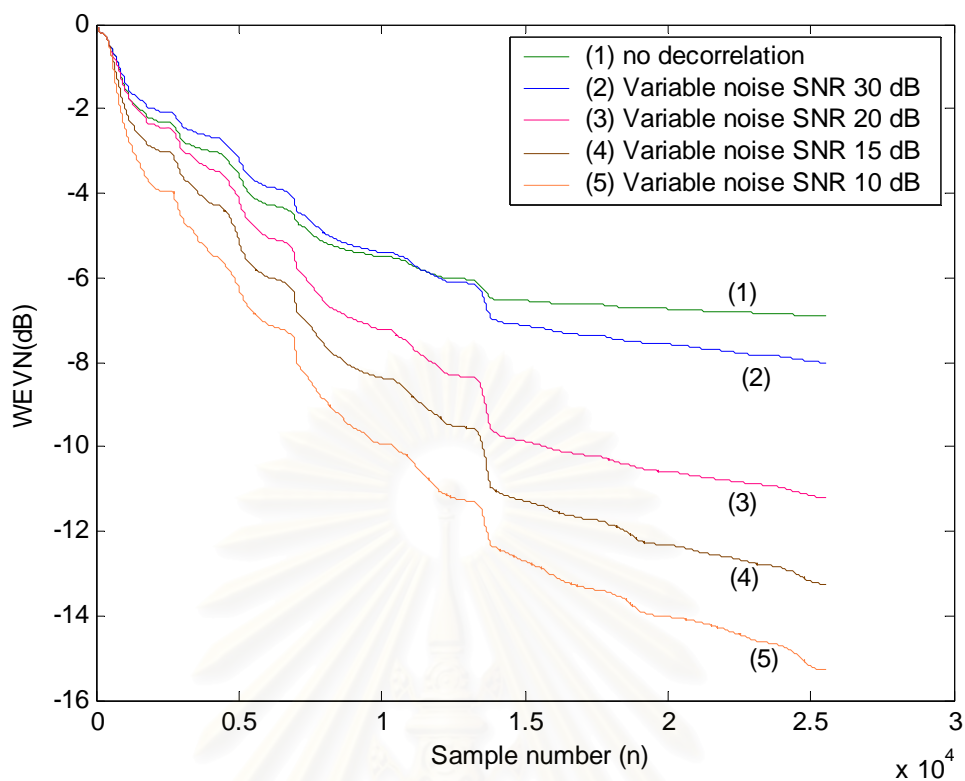
หลังจากผ่านสัญญาณที่ถูกดีคอดรีเลชันแล้วเข้าไปในระบบ SAEC พบว่า ประสิทธิภาพการทำงานของระบบ SAEC ดีขึ้นตามขนาดของสัญญาณรบกวนที่มี SNR ลดลง ดังรูปที่ 4.11 (a) ทั้งนี้สหสัมพันธ์ของสัญญาณเสียงสเตอริโอก็ลดลงตามขนาดของสัญญาณรบกวนที่มี SNR ลดลงเช่นกัน ดังรูปที่ 4.11 (b)

เมื่อทำการปรับขนาดสัญญาณรบกวนแบบปรับตัวได้ในแต่ละช่องสัญญาณให้มีค่าขึ้นอยู่กับอัตราส่วน SNR ระหว่างสัญญาณเสียงพูดกับสัญญาณรบกวน เป็น 30, 20, 15 และ 10 dB ตามรูปที่ 4.11 (a) พบว่า ส่งผลให้ประสิทธิภาพการทำงานของระบบ SAEC ดีขึ้นตามลำดับ ตามค่า WEVN ที่ลดลงดังกราฟเส้นที่ (2) – (5) และลดลงได้ถึงประมาณ -15 dB เมื่อค่า SNR มีค่า 10 dB และเมื่อพิจารณาค่า Coherence Function ดังรูปที่ 4.11 (b) พบว่าการดีคอดรีเลชัน

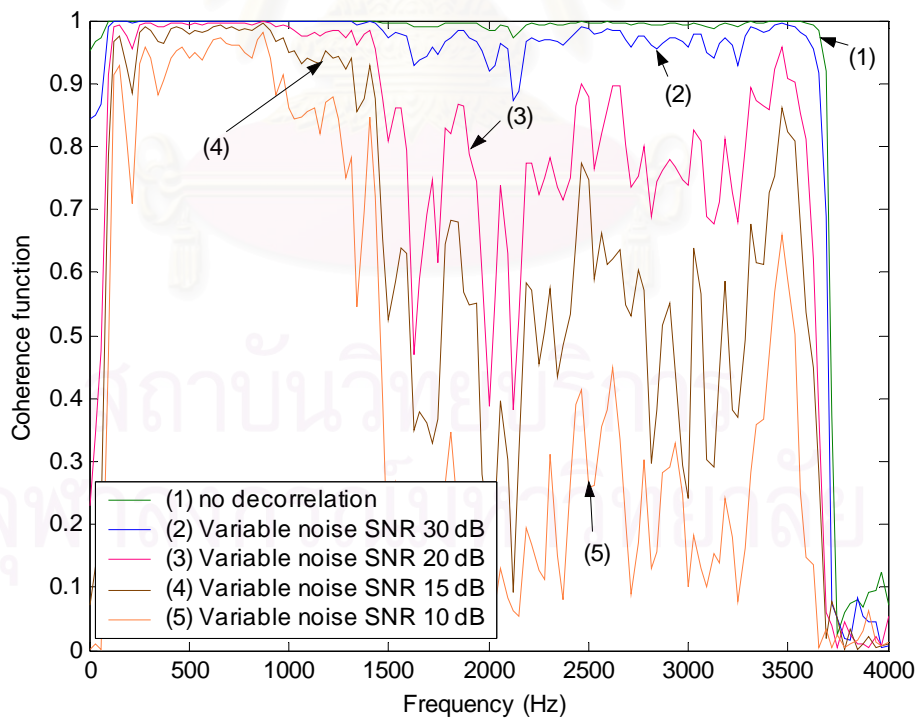
สัญญาณเสียงสเตอริโอ สามารถลดสหสัมพันธ์ของสัญญาณเสียงสเตอริโอทั้งช่องสัญญาณลงได้ค่อนข้างมาก โดยเฉพาะในช่วงความถี่สูง อย่างไรก็ตามการดีคอดรีเลชันสัญญาณเสียงสเตอริโอด้วยการเพิ่มสัญญาณรบกวนแบบปรับตัวได้ยังคงมีผลกระทบต่อคุณภาพสัญญาณเสียงสเตอริโอที่ถูกดีคอดรีเลชันคล้ายกับการเพิ่มสัญญาณรบกวนขาวสุ่ม กล่าวคือ เมื่อ SNR ของสัญญาณรบกวนลดลง ผลกระทบต่อคุณภาพสัญญาณเสียงจะมีมากขึ้น



สถาบันวิทยบริการ  
จุฬาลงกรณ์มหาวิทยาลัย



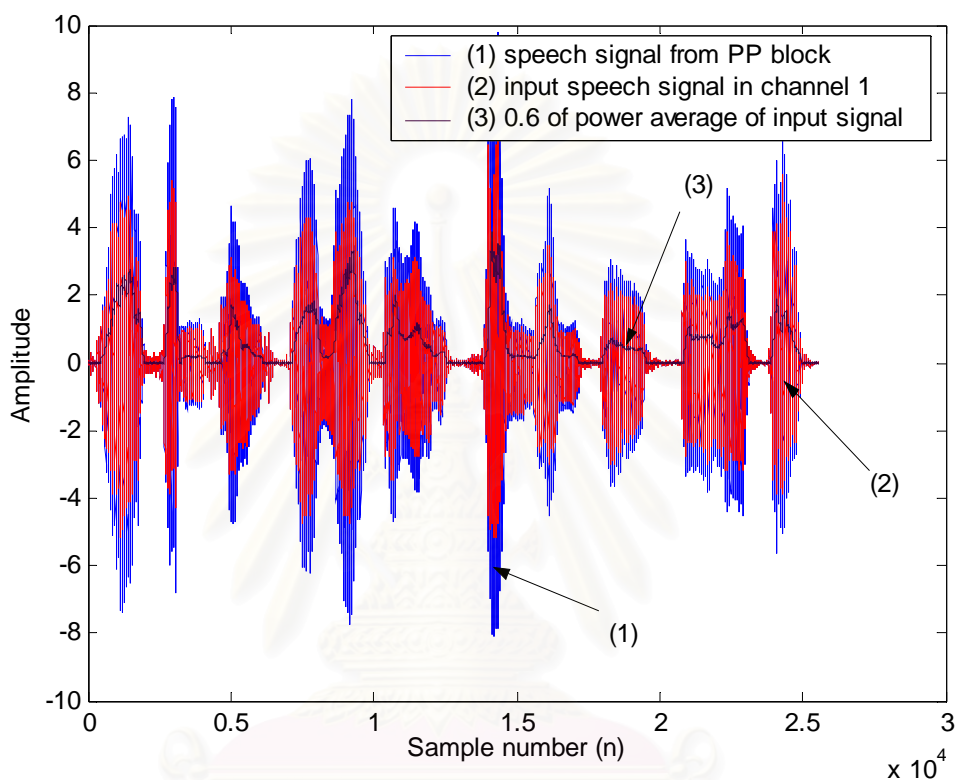
รูปที่ 4.11 (a)



รูปที่ 4.11 (b)

รูปที่ 4.11 (a) ค่า WEVN ของระบบ SAEC และ (b) Coherence Function ของสัญญาณเสียงสเตอริโอ เมื่อเพิ่มสัญญาณรบกวนแบบปรับตัวได้

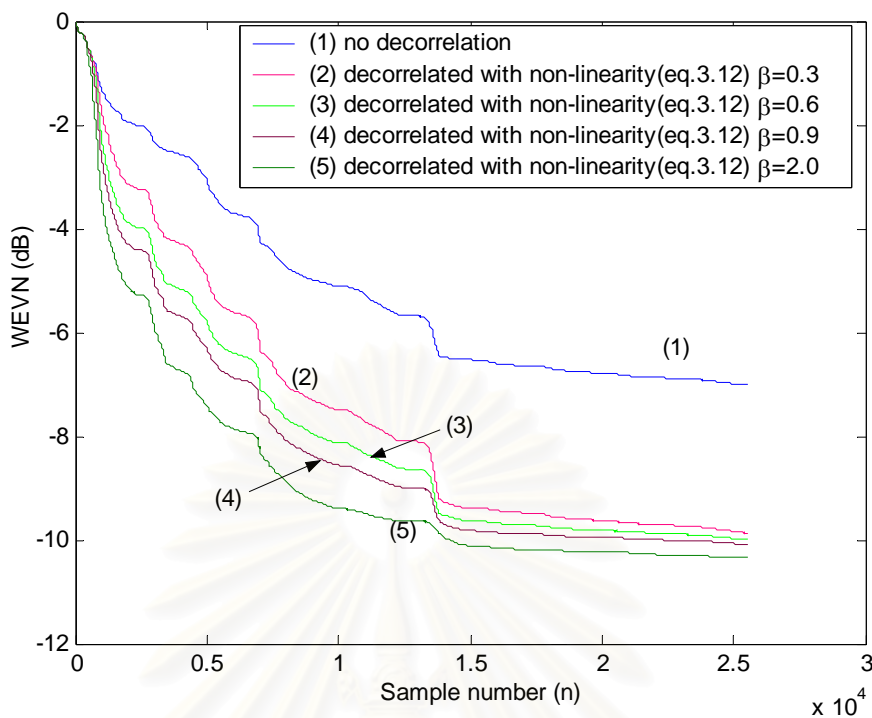
ดังนั้นเพื่อป้องกันการลดคุณภาพของสัญญาณเสียงสเตอริโอดังกล่าว จึงมีแนวคิดตามหัวข้อที่ 3.1.3.2 ในการเพิ่มภาวะไม่เชิงเส้นให้แก่ระบบ โดยฟังก์ชัน  $\Gamma_i(n)$  ในช่องสัญญาณ  $i$  มีค่าปรับเปลี่ยนตามค่าพลังงานเฉลี่ยของสัญญาณเสียงพูดในช่วงนั้นๆ ตามแฟกเตอร์  $\beta$  สัญญาณที่ถูกดีคอดิเรชันเป็นไปตามสมการที่ (3.12) ซึ่งลักษณะสัญญาณที่ถูกดีคอดิเรชันที่ได้จากบล็อก PP เป็นดังรูปที่ 4.12



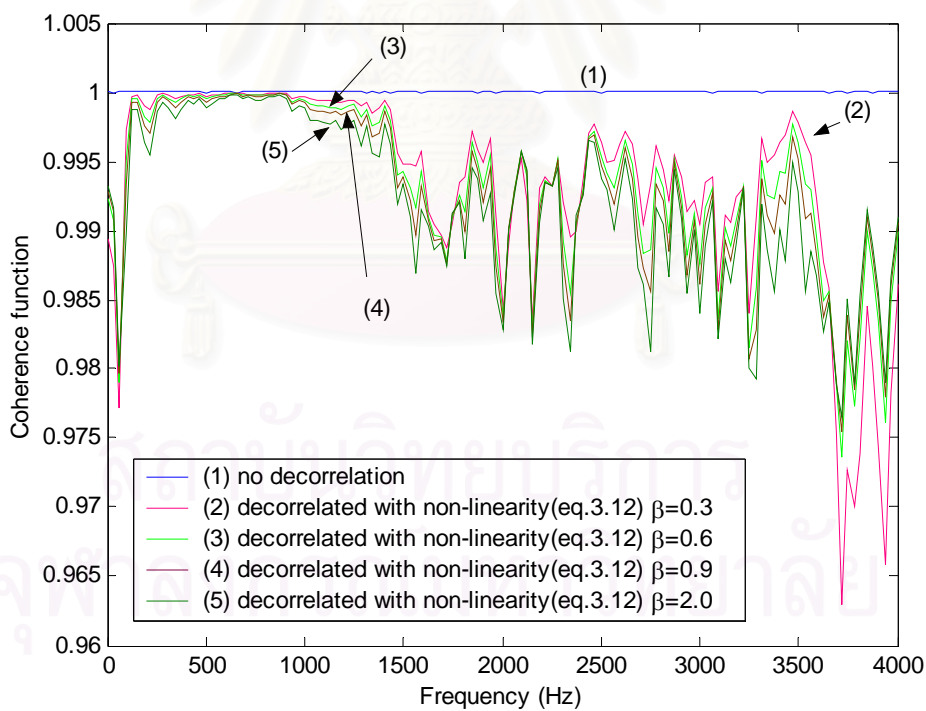
รูปที่ 4.12 พลังงานเฉลี่ยของสัญญาณที่เปลี่ยนแปลงตามสมการที่ (3.12)

(1) เป็นสัญญาณรวมของสัญญาณเสียงสเตอริโอช่องที่ 1 และสัญญาณรบกวนในอัตราส่วน 0.6 ของสัญญาณเสียงสเตอริโอในช่องที่ 1 (2) เป็นสัญญาณเสียงสเตอริโอในช่องที่ 1 และ (3) เป็นสัญญาณรบกวนแบบปรับตัวได้ตามอัตราส่วนของพลังงานเฉลี่ยของสัญญาณเสียงในข้อ 3.1.3.2

สำหรับการหาค่า  $\beta$  ตามสมการที่ (3.12) โดยพิจารณากราฟรูปที่ 4.12 พบว่าสามารถเลือกใช้ได้หลายค่า ซึ่งค่า  $\beta$  ต่างๆ จะมีผลต่อประสิทธิภาพการทำงานของระบบ SAEC ที่ต่างกัน ด้วย โดยการเลือกใช้ค่า  $\beta$  ควรอยู่ในช่วง 0.3 - 2.0 ซึ่งค่า WEVN และค่า Coherence Function จะถูกแสดงในรูปที่ 4.13 (a) และ รูปที่ 4.13 (b) ตามลำดับ



รูปที่ 4.13 (a)

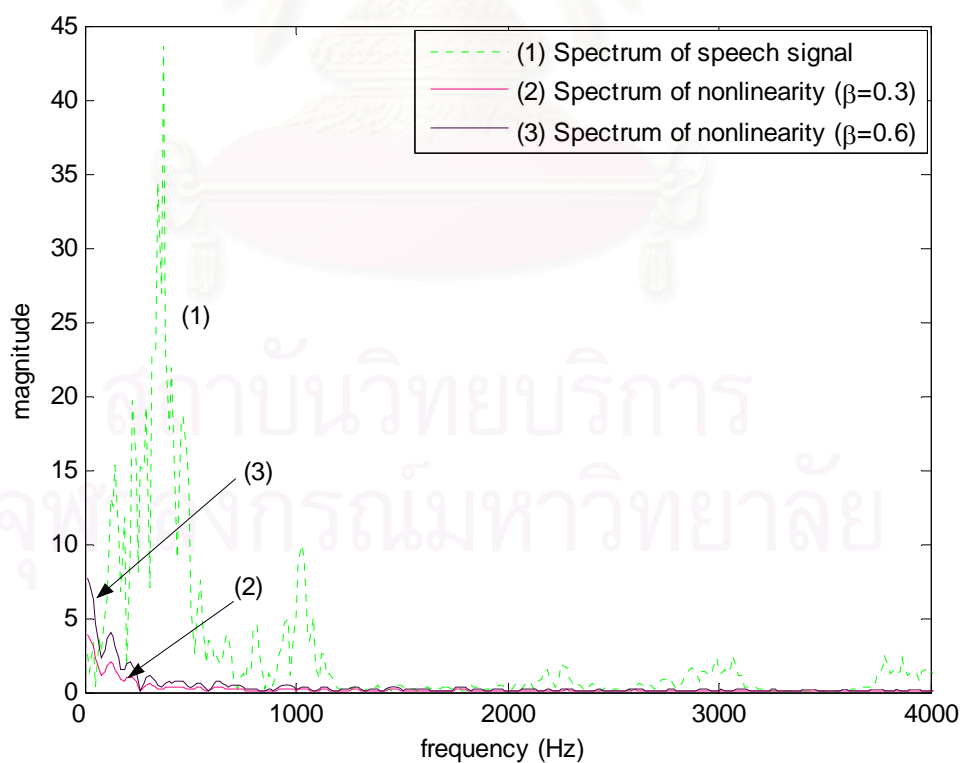


รูปที่ 4.13 (b)

รูปที่ 4.13 (a) ค่า WEVN ของระบบ SAEC และ (b) Coherence Function ของสัญญาณเสียงสเตอริโอเมื่อทำการดีคორเรเลชันสัญญาณเสียงสเตอริโอ ด้วยภาวะไม่เชิงเส้นที่นำเสนอที่ปรับค่าตามพลังงานเฉลี่ยของสัญญาณเสียงพูด

ค่า WEVN จะมีค่าลดลงตามค่า  $\beta$  ที่เพิ่มขึ้นดังกราฟเส้นที่ (2) – (5) รูปที่ 4.13 (a) ซึ่งในความเป็นจริงสามารถเพิ่มค่า  $\beta$  ให้มากกว่านี้ได้อีก แต่ก็พบว่าประสิทธิภาพการทำงานของระบบ SAEC ไม่ได้เพิ่มขึ้นเลย ในขณะที่คุณภาพสัญญาณเสียงเริ่มถูกรบกวนมากขึ้นจนสังเกตได้ชัดเจน (เสียงดัง)

ในรูปที่ 4.13 (a) เมื่อทำการเปลี่ยนแปลงค่า  $\beta$  ตั้งแต่ 0.3 - 2.0 พบว่า สามารถเพิ่มประสิทธิภาพการทำงานของระบบ SAEC ให้ดีขึ้น ค่า WEVN ลดลงได้ประมาณ -10 dB แต่เมื่อสังเกตค่า Coherence Function ในรูป (b) พบว่า แนวคิดในการดีคอดรีเลชันโดยการเพิ่มภาวะไม่เชิงเส้นที่ปรับเปลี่ยนค่าตามพลังงานเฉลี่ยของสัญญาณเสียงพูดตามที่น่าเสนอนี้ ไม่สามารถดีคอดรีเลชันสัญญาณเสียงสเตอริโอได้มากนัก แต่สามารถรักษาไว้ซึ่งคุณภาพสัญญาณเสียงให้ใกล้เคียงคุณภาพสัญญาณเสียงเดิมเช่นเดียวกับวิธีการเพิ่มภาวะไม่เชิงเส้นในหัวข้อที่ 3.1.2 เมื่อพิจารณาความหนาแน่นกำลังเชิงสเปกตรัมของสัญญาณภาวะไม่เชิงเส้นที่น่าเสนอเปรียบเทียบกับสัญญาณเสียงสเตอริโอเดิมในโดเมนความถี่ แสดงดังรูปที่ 4.14 จะเห็นว่า ผลกระทบต่อคุณภาพสัญญาณเสียงจากการเพิ่มภาวะไม่เชิงเส้น ( $\beta$ ) มีไม่มากนัก เนื่องจากในช่วงความถี่สูงที่ความหนาแน่นกำลังเชิงสเปกตรัมสัญญาณเสียงพูดมีขนาดน้อย ความหนาแน่นกำลังเชิงสเปกตรัมของสัญญาณภาวะไม่เชิงเส้นนั้นมีขนาดน้อยด้วย



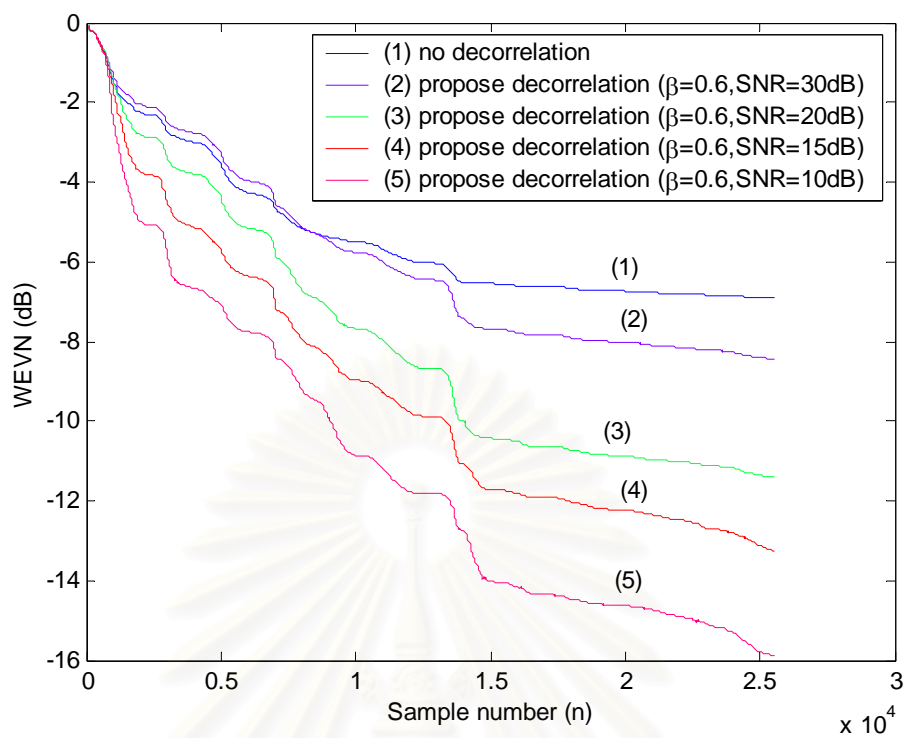
รูปที่ 4.14 เปรียบเทียบความหนาแน่นกำลังเชิงสเปกตรัมสัญญาณภาวะไม่เชิงเส้นที่น่าเสนอ กับความหนาแน่นกำลังเชิงสเปกตรัมสัญญาณเสียงพูด



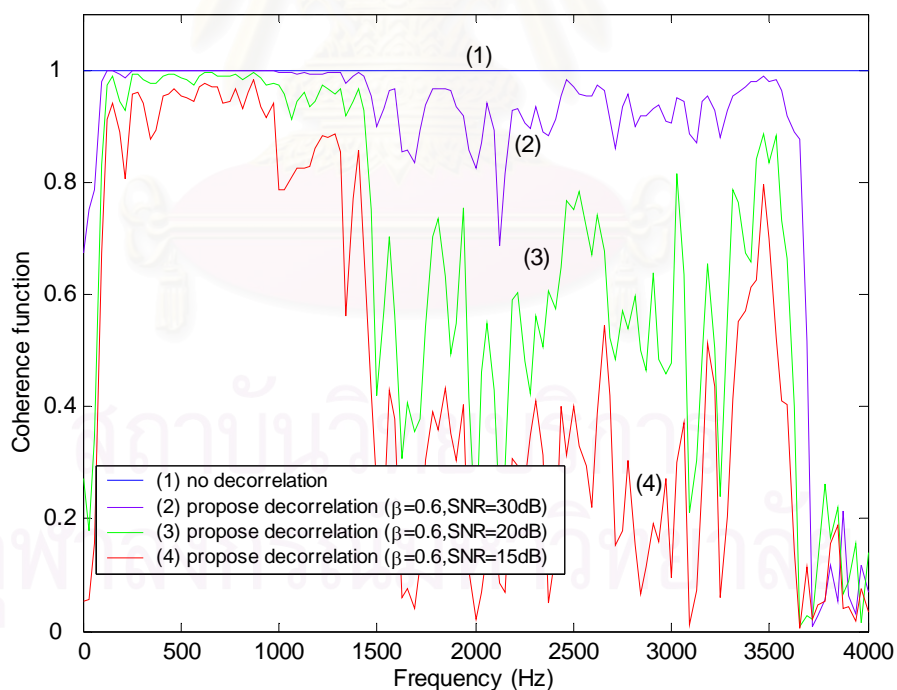
ดังนั้นจากแนวคิดทั้งสองข้างต้น จึงศึกษาผลของการเพิ่มสัญญาณรบกวนแบบปรับตัวได้ตามสมการที่ (3.15) เพื่อลดสหสัมพันธ์ของสัญญาณเสียงสเตอริโอ โดยศึกษาถึงค่า SNR ของสัญญาณรบกวนที่ปรับค่าได้โดยที่คุณภาพสัญญาณเสียงสเตอริโอที่ถูกดีคอดรีเลชันไม่ถูกรบกวนมากนัก ผลการจำลองการทำงานของระบบ SAEC ภายหลังจากที่มีการดีคอดรีเลชันตามสมการที่ (3.15) แล้ว แสดงค่า WEVN ได้ดังรูปที่ 4.15 (a) และค่า Coherence Function ในรูปที่ 4.15 (b) ตามลำดับ



สถาบันวิทยบริการ  
จุฬาลงกรณ์มหาวิทยาลัย



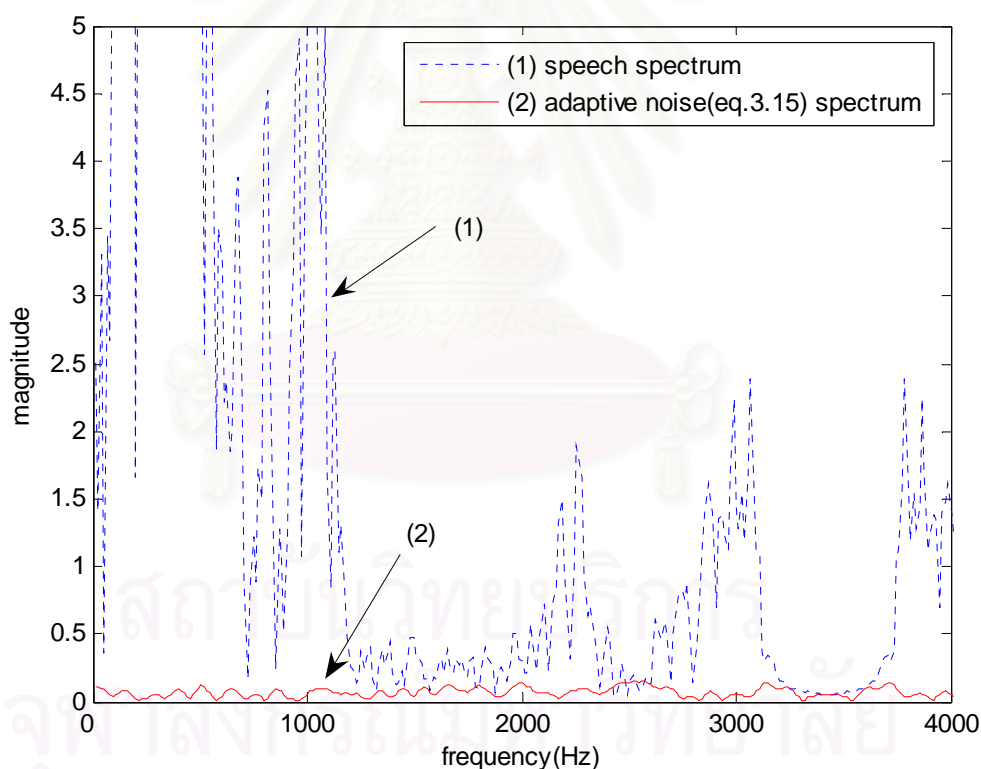
รูปที่ 4.15 (a)



รูปที่ 4.15 (b)

รูปที่ 4.15 (a) ค่า WEVN ของระบบ SAEC (b) Coherence Function  
ของสัญญาณเสียงสเตอริโอ เมื่อดีคอรเรลชันสัญญาณเสียงสเตอริโอ  
ด้วยสัญญาณรบกวนแบบปรับตัวได้ที่น่าเสนอ ( $\beta=0.6$ )

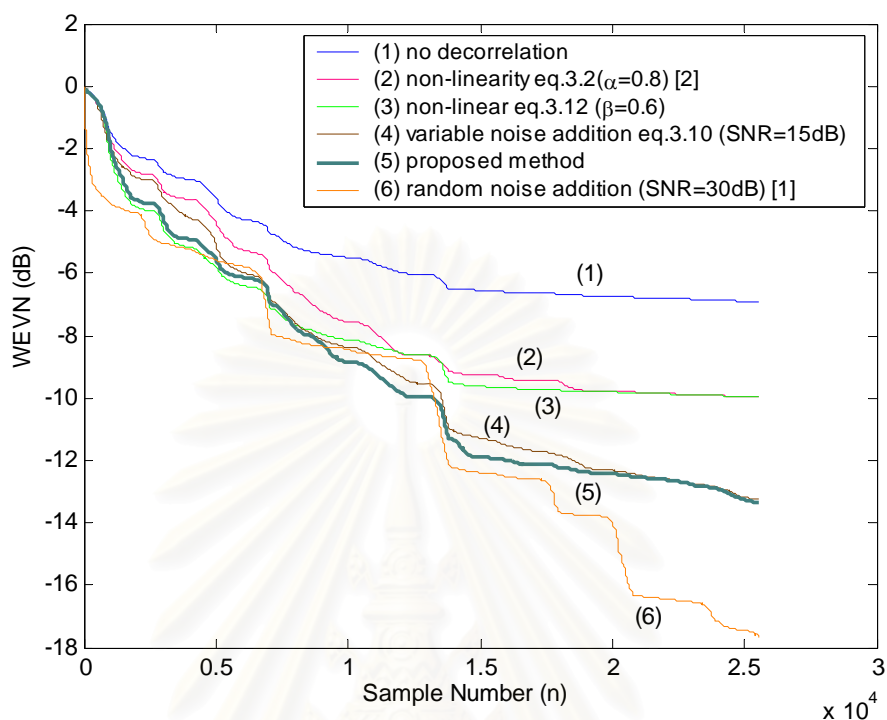
จากค่า WEVN ในรูปที่ 4.15 (a) พบว่า วิธีที่นำเสนอสมาสามารถช่วยเพิ่มประสิทธิภาพการทำงานของระบบ SAEC ให้ดีขึ้นได้ โดยสังเกตจากค่า WEVN ที่ค่อยๆ ลดลงตามกราฟเส้นที่ (2) ถึง (5) ตามค่า SNR ของสัญญาณรบกวนแบบปรับตัวได้ที่ลดลง โดยสามารถลดลงได้ถึงประมาณ -16 dB เมื่อค่า SNR ของสัญญาณรบกวนแบบปรับตัวได้มีค่า 10 dB และเมื่อพิจารณาจากกราฟค่า Coherence Function ในรูปที่ 4.15 (b) พบว่าสัญญาณเสียงสเตอริโอทั้งสองช่องสัญญาณถูกดีคორิเลชันลงค่อนข้างมากจากสัญญาณเสียงสเตอริโอเดิม โดยสหสัมพันธ์ของสัญญาณถูกลดลงอย่างมาก โดยเฉพาะในช่วงความถี่สูง ทั้งนี้เมื่อพิจารณาถึงค่าความหนาแน่นกำลังเชิงสเปกตรัมของสัญญาณรบกวนแบบปรับตัวได้ตามพารามิเตอร์ที่ทำให้ระบบ SAEC มีประสิทธิภาพการทำงานที่ดี ในขณะที่สัญญาณรบกวนยังไม่มีผลกระทบต่อคุณภาพสัญญาณเสียงสเตอริโอเดิมคือ ค่า  $\beta=0.6$  และค่า SNR ที่ 15 dB ดังรูปที่ 4.16 กล่าวคือ ขนาดความหนาแน่นกำลังเชิงสเปกตรัมมีค่าน้อยกว่าภาวะไม่เชิงเส้นเมื่อเลือกค่า  $\beta$  ที่เท่ากัน



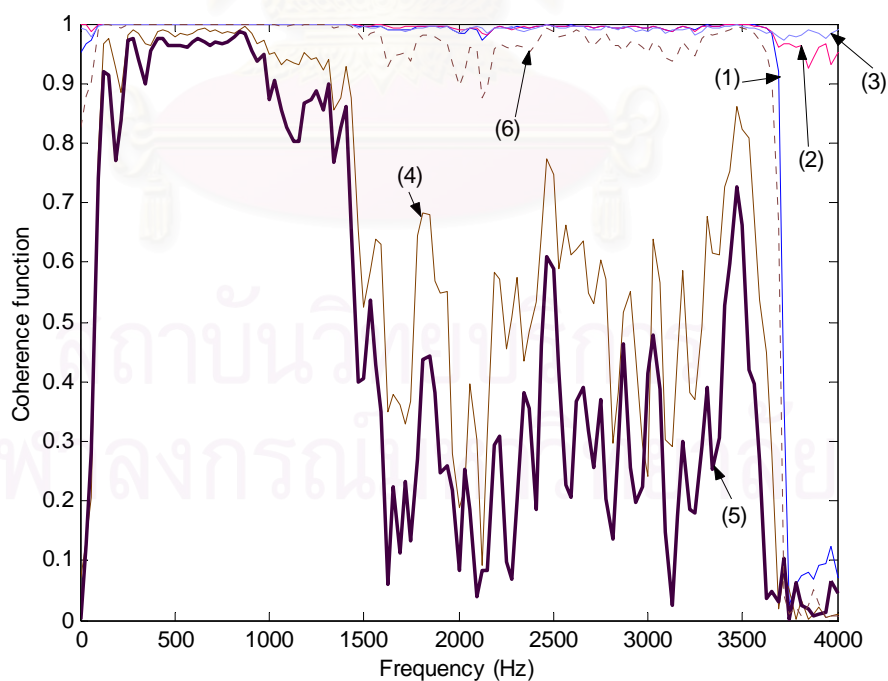
รูปที่ 4.16 เปรียบเทียบความหนาแน่นกำลังเชิงสเปกตรัมสัญญาณรบกวนแบบปรับตัวได้ที่นำเสนอกับความหนาแน่นกำลังเชิงสเปกตรัมสัญญาณเสียงพูด

เพื่อแสดงให้เห็นผลของการทำงานของระบบ SAEC ที่แตกต่างกันในการเลือกวิธีดีคอริเลชันแต่ละวิธี จึงนำผลของค่า WEVN และค่า Coherence Function ในแต่ละวิธีมาแสดงในกราฟเดียวกัน ดังรูปที่ 4.17 (a) และ 4.17 (b) โดยได้ทำการเลือกตัวแปรที่เหมาะสมที่สุดในแต่ละวิธี

เพื่อจะทำให้ประสิทธิภาพของระบบ SAEC ทำงานได้ดี ในขณะที่ยังรักษาคุณภาพสัญญาณเสียงสเตอริโอได้ดีเช่นกัน



รูปที่ 4.17 (a)



รูปที่ 4.17 (b)

รูปที่ 4.17 (a) ค่า WEVN ของระบบ SAEC และ (b) ค่า Coherence Function ของสัญญาณเสียงสเตอริโอ เมื่อใส่คอรืเลขันสัญญาณเสียงสเตอริโอด้วยวิธีต่างๆ

ในการเลือกค่าตัวแปรที่เหมาะสมที่สุดสำหรับแต่ละวิธีในการดีคอดิเลชันสัญญาณตามกราฟรูปที่ 4.17 (a) คือ กรณีดีคอดิเลชันสัญญาณโดยวิธีเพิ่มภาวะไม่เชิงเส้นตามกราฟเส้นที่ (2) ได้เลือกค่า  $\alpha=0.8$  ซึ่งเป็นค่าที่เหมาะสมที่สุดที่ยังคงให้คุณภาพสัญญาณเสียงสเตอริโอไม่ถูกรบกวนมากนัก พบว่าค่า WEVN ของระบบ SAEC มีค่าประมาณ -10 dB ค่า Coherence Function ในรูปที่ 4.17 (b) ของสัญญาณเสียงสเตอริโอลดลงค่อนข้างน้อย แสดงว่า สหสัมพันธ์ของสัญญาณเสียงสเตอริโอถูกลดลงไปได้เล็กน้อย สำหรับวิธีดีคอดิเลชันโดยเพิ่มสัญญาณรบกวนสุ่ม ตามกราฟเส้นที่ (6) ได้เลือกใช้ตัวแปรของสัญญาณรบกวนสุ่มดังกล่าวที่ SNR เท่ากับ 30 dB โดยคุณภาพสัญญาณเสียงสเตอริโอยังสามารถยอมรับได้ พบว่าค่า WEVN ของระบบ SAEC ลดลงได้ถึงประมาณ -17 dB ซึ่งถือว่ามีประสิทธิภาพดีที่สุดจากการเปรียบเทียบกับกราฟเส้นอื่นๆ ในกราฟรูปที่ 4.17 (a) สำหรับวิธีดีคอดิเลชันสัญญาณเสียงสเตอริโอด้วยวิธีที่น่าเสนอคือ การเพิ่มสัญญาณรบกวนแบบปรับตัวได้เฉพาะคาบที่มีเสียงพูดดังกราฟเส้นที่ (5) ได้เลือกใช้อัตราส่วน  $\beta=0.6$  และอัตราส่วนสัญญาณ (Variable Gain,  $\gamma$ ) เท่ากับ 15 dB ซึ่งยังคงรักษาไว้ซึ่งคุณภาพสัญญาณเสียงสเตอริโอไว้ได้ใกล้เคียงกับสัญญาณเสียงสเตอริโอเดิม ค่า WEVN ของระบบ SAEC สามารถลดลงได้ถึงประมาณ -14 dB ซึ่งดีกว่าวิธีการเพิ่มภาวะไม่เชิงเส้นในกราฟเส้นที่ (2) แต่ดีกว่าวิธีการใช้สัญญาณรบกวน หากแต่ถ้าเมื่อพิจารณาที่กราฟ Coherence Function รูปที่ 4.17 (b) พบว่า สหสัมพันธ์ของสัญญาณเสียงสเตอริโอถูกทำให้ลดลงได้ค่อนข้างมากเมื่อเปรียบเทียบกับวิธีดีคอดิเลชันทั้ง 2 วิธีดังกล่าวข้างต้น ดังแสดงในกราฟเส้นที่ (5) จะสังเกตเห็นได้ว่าในช่วงความถี่สูง กล่าวคือ  $> 1,500$  Hz สัญญาณเสียงสเตอริโอถูกลดสหสัมพันธ์ลงได้มากกว่าในช่วงความถี่ต่ำ นอกจากนี้วิธีที่น่าเสนอนี้ก็สามารถที่จะรักษาคุณภาพสัญญาณเสียงให้ได้ใกล้เคียงกับสัญญาณเสียงสเตอริโอเดิมด้วย สำหรับการทดสอบคุณภาพของสัญญาณเสียงโดยการทดสอบ Subjective Listening Test จะแสดงให้เห็นในหัวข้อถัดไป

#### 4.5 การศึกษาผลกระทบต่อคุณภาพสัญญาณเสียงสเตอริโอ ด้วยการทดสอบ Subjective Listening Test

ในหัวข้อนี้จะใช้การทดสอบ MOS Test (ดังหัวข้อที่ 3.3.3) เพื่อศึกษาคุณภาพเสียงของสัญญาณเสียงสเตอริโอชุดต่างๆ โดยได้เลือกบุคคลเข้าทำการทดสอบจำนวน 18 คน แบ่งเป็นเพศชายอายุประมาณ 20-25 ปีจำนวน 10 คน และเพศหญิงอายุประมาณ 20-25 ปีจำนวน 8 คน แต่ละคนจะได้รับฟังสัญญาณเสียงสเตอริโอที่มีความแตกต่างกันระหว่างสัญญาณลำโพงซ้าย กับ ลำโพงทางขวา ตามรูปที่ 3.7 สำหรับสัญญาณที่ใช้ในการทดสอบ ได้แก่

1. สัญญาณเสียงสเตอริโอเดิมในขณะที่ยังไม่ได้มีการดีคอดิเรชันสัญญาณ
2. สัญญาณเสียงสเตอริโอที่ทำการดีคอดิเรชันสัญญาณด้วยการเพิ่มภาวะไม่เชิงเส้นตามสมการที่ (3.2) และ (3.3) โดยเลือกค่า  $\alpha=0.8$  ซึ่งเป็นค่าที่เหมาะสมที่สุดที่คุณภาพสัญญาณเสียงสเตอริโอไม่ถูกรบกวนมากนัก
3. สัญญาณเสียงสเตอริโอ จากการดีคอดิเรชันโดยเพิ่มสัญญาณรบกวนแบบปรับตัวได้ตามสมการที่ (3.10) โดยเลือกค่า SNR ( $\gamma$ ) ที่ 15 dB
4. สัญญาณเสียงสเตอริโอ จากการดีคอดิเรชันโดยการเพิ่มขนาดสัญญาณตามสมการที่ (3.12) โดยเลือกค่า  $\beta$  เท่ากับ 0.6
5. สัญญาณเสียงสเตอริโอ จากการดีคอดิเรชันสัญญาณโดยใช้สัญญาณรบกวนแบบปรับตัวได้ตามแนวคิดที่นำเสนอในสมการที่ (3.15) เมื่อเลือก  $\beta$  เท่ากับ 0.6 และ  $\gamma$  เท่ากับ 15 dB
6. สัญญาณเสียงสเตอริโอ จากการดีคอดิเรชันโดยการใช้สัญญาณรบกวนสุ่ม ตามสมการที่ (3.1) โดยเลือกใช้ค่า SNR ที่ 30 dB

เมื่อทำการหาค่าเฉลี่ยของระดับคะแนนตามตารางที่ 3.1 สามารถสรุปผลการให้คะแนนที่บอกถึงระดับของคุณภาพของสัญญาณเสียงสเตอริโอได้ดังตารางที่ 4.1

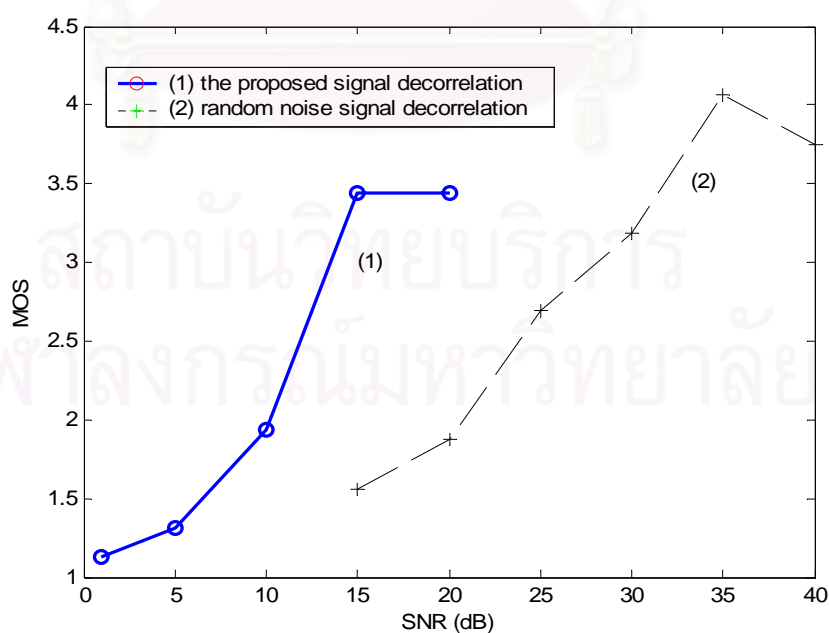
Signals	MOS values
1. Original Speech Signal	4.61
2. $S_{new}$ (eq. 3.21) [ $\beta=0.5$ ]	3.33
3. $S_{new}$ (eq. 3.21) [ $\beta=1.0$ ]	2.89
4. Random Noise Addition in [1]	2.58
5. Non-linearity [3.1.2]	4.14
6. Adaptive Noise Addition [proposed]	4.04

ตารางที่ 4.1 เปรียบเทียบคุณภาพของสัญญาณเสียงสเตอริโอ  
ในแต่ละวิธีการดีคอดิเรชันสัญญาณ ตามค่า MOS

จากการเปรียบเทียบคุณภาพของสัญญาณเสียงสเตอริโอตามตารางที่ 4.1 พบว่า คุณภาพสัญญาณเสียงสเตอริโอหลังจากที่มีการดีคอดิเรชันสัญญาณโดยใช้สัญญาณรบกวนสุ่มจะให้คุณภาพเสียงที่ค่อนข้างด้อยที่สุดเมื่อเปรียบเทียบกับคุณภาพเสียงสเตอริโอที่ได้จากวิธีการดีคอดิเร

ชั้นแบบอื่นๆ โดยมีค่า MOS เท่ากับ 2.58 ถึงแม้ว่าวิธีการดีคอรিলेशनแบบใช้สัญญาณรบกวนสุ่มนี้จะส่งผลให้ระบบ SAEC มีประสิทธิภาพการทำงานดีที่สุดดังกราฟค่า WEVN ในรูปที่ 4.17 (a) ก็ตาม และจากตารางเดียวกันนี้พบว่า วิธีดีคอริลेशनสัญญาณโดยการเพิ่มสัญญาณรบกวนแบบปรับตัวได้ตามที่นำเสนอมีค่า MOS เท่ากับ 4.04 ในขณะที่ การดีคอริลेशनสัญญาณโดยการเพิ่มภาวะไม่เชิงเส้นมีค่า MOS เท่ากับ 4.14 แสดงว่าคุณภาพสัญญาณเสียงสเตอริโอที่ผ่านการดีคอริลेशनวิธีที่นำเสนอสามารถรักษาคุณภาพเสียงสเตอริโอไว้ได้ใกล้เคียงกับวิธีการเพิ่มภาวะไม่เชิงเส้น แต่เมื่อพิจารณาด้านประสิทธิภาพการทำงานของระบบ SAEC โดยสังเกตจากค่า WEVN ในกราฟรูปที่ 4.17 (a) หรือค่า Coherence Function ที่สามารถลดลงได้มากกว่าตามรูปที่ 4.17 (b) กลับพบว่า วิธีดีคอริลेशनสัญญาณโดยการเพิ่มสัญญาณรบกวนแบบปรับตัวได้ตามที่นำเสนอสามารถที่จะให้ประสิทธิภาพที่ดีกว่าวิธีดีคอริลेशनที่เพิ่มภาวะไม่เชิงเส้น

เนื่องจากวิธีดีคอริลेशनสัญญาณเสียงสเตอริโอที่นำเสนอในวิทยานิพนธ์นี้ยังคงอาศัยคุณสมบัติการดีคอริลेशनสัญญาณจากสัญญาณรบกวนขาวแบบเกาส์ ซึ่งให้คุณสมบัติการดีคอริลेशनคล้ายกับสัญญาณรบกวนสุ่ม เพียงแต่เป็นสัญญาณรบกวนที่ปรับขนาดได้ตามขนาดสัญญาณเสียงขาเข้าของระบบ SAEC ดังนั้นเพื่อเป็นการเปรียบเทียบถึงคุณภาพเสียงสเตอริโอที่ถูกดีคอริลेशनสัญญาณแล้ว ในกระบวนการดีคอริลेशनที่เพิ่มสัญญาณรบกวนสุ่มเปรียบเทียบกับวิธีเพิ่มสัญญาณรบกวนแบบปรับตัวได้ตามที่นำเสนอ จึงทำการทดลองปรับขนาดสัญญาณรบกวนทั้ง 2 ประเภทตามค่า SNR ต่างๆ และทำการทดสอบ MOS Test ดังกราฟในรูปที่ 4.18



รูปที่ 4.18 ค่า MOS ของสัญญาณเสียงสเตอริโอจากการดีคอริลेशनสัญญาณ โดยการเพิ่มสัญญาณรบกวนสุ่ม เปรียบเทียบการเพิ่มสัญญาณรบกวนแบบปรับตัวได้

จากค่า MOS ในกราฟดังรูปที่ 4.18 แสดงให้เห็นว่า การเพิ่มสัญญาณรบกวนแบบปรับตัวได้ สามารถเพิ่มขนาดสัญญาณรบกวนได้ในปริมาณที่มากกว่าการเพิ่มสัญญาณรบกวนสุ่มสังเกต จากค่า SNR ที่ลดลงได้ต่ำกว่า นอกจากนี้ยังสามารถรักษาคุณภาพเสียงของสัญญาณเสียงสเตอริโอได้ดีกว่าอีกด้วย ซึ่งสังเกตได้ว่าเมื่อให้ระดับการเพิ่มสัญญาณรบกวนที่มีค่า SNR เท่ากัน สัญญาณเสียงสเตอริโอจากการเพิ่มสัญญาณรบกวนแบบปรับตัวได้จะให้ค่า MOS ที่สูงกว่าหรือถ้า ต้องการคุณภาพเสียงที่มีค่า MOS ระดับเดียวกันเช่นที่ค่า MOS ประมาณ 2.5 การดีคอดิเลชันสัญญาณเสียงสเตอริโอโดยการเพิ่มสัญญาณรบกวนแบบปรับตัวได้สามารถเพิ่มระดับสัญญาณรบกวนได้ในปริมาณที่มากกว่าการใช้สัญญาณรบกวนสุ่มโดยดูจากค่า SNR ที่ต่ำกว่า

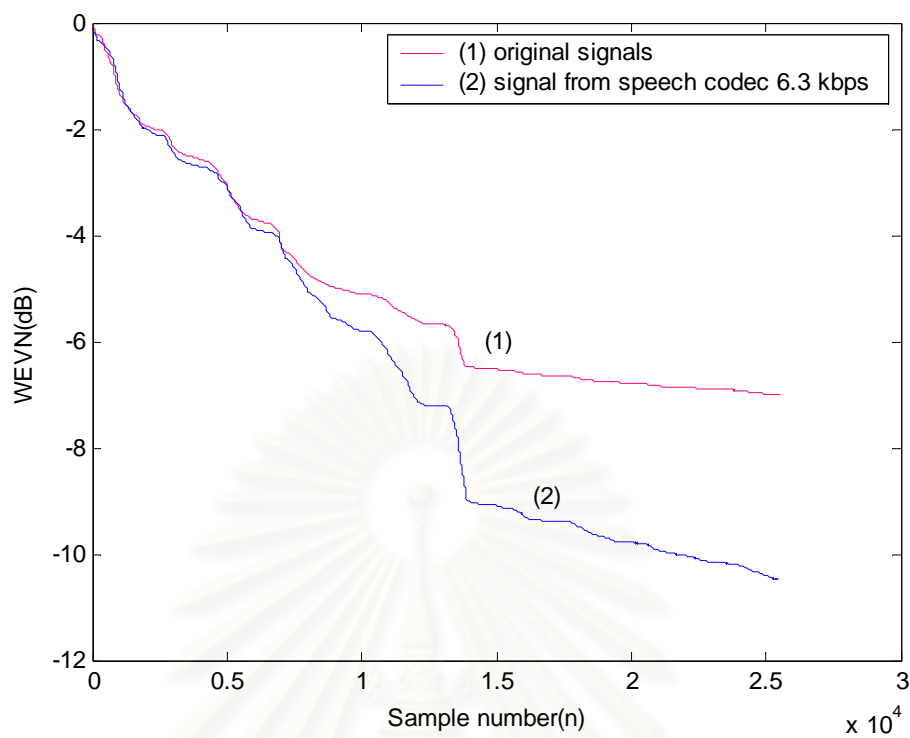
#### 4.6 ผลการจำลองการทำงานของระบบ SAEC ที่มีการเข้ารหัสสัญญาณเสียงตามมาตรฐาน G.723.1

##### 4.6.1 ผลการจำลองการทำงานของระบบ SAEC ที่ระบบมีการเข้ารหัสสัญญาณเสียงตามมาตรฐาน G.723.1 เมื่อไม่ได้มีการดีคอดิเลชันสัญญาณเข้าแบบสเตอริโอ

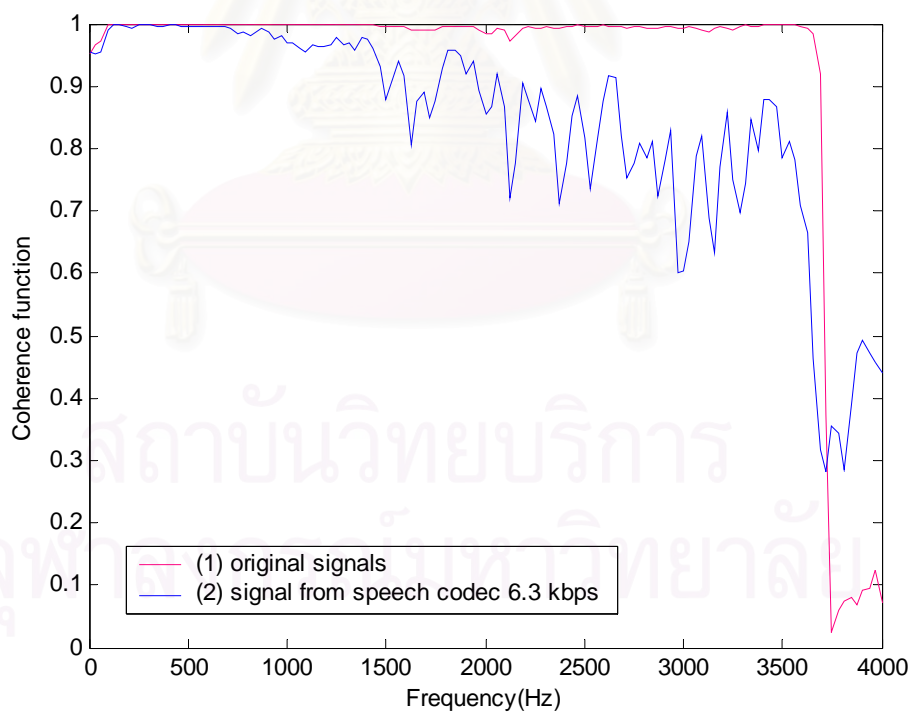
จากหัวข้อที่ 3.3 ได้กล่าวถึงการเข้ารหัสสัญญาณเสียงพูดที่เป็นวิธีการหนึ่งซึ่งส่งผลให้เกิดการเปลี่ยนแปลงในทางที่ดีขึ้นต่อการทำงานของระบบ SAEC ได้ เมื่อกรณีที่มีการสื่อสารทางเสียงโดยผ่านการเข้ารหัสสัญญาณเสียงพูดตามรูปที่ 3.8 ซึ่งในวิทยานิพนธ์ฉบับนี้ได้เลือกศึกษาผลของการเข้ารหัสเสียงพูดตามมาตรฐาน G.723.1 ต่อประสิทธิภาพการทำงานของระบบ SAEC เนื่องจากเป็นมาตรฐานที่ถูกออกแบบมาใช้ในระบบการสื่อสารทางเสียงผ่านเครือข่ายอินเทอร์เน็ต [18] และอัตราบิตข้อมูลต่ำ ในขณะที่คุณภาพสัญญาณเสียงยังคงอยู่ในระดับที่ยอมรับได้ ซึ่งมีรายละเอียดวิธีการเข้ารหัสดังกล่าวในหัวข้อที่ 3.3.2 สำหรับการจำลองการทำงานของระบบ SAEC ในวิทยานิพนธ์นี้ได้สมมติให้ช่องสัญญาณสื่อสารตามรูปที่ 3.8 เป็นแบบจุดมคติ หรือไม่มีการสูญเสียเกิดขึ้นในระหว่างการส่งข้อมูลผ่านช่องสัญญาณสื่อสาร สำหรับอัตราการเข้ารหัสได้เลือกใช้ที่อัตราบิตข้อมูล 6.3 kbps ซึ่งถือว่ามีคุณภาพเสียงที่ดีกว่าที่อัตราบิตข้อมูล 5.3 kbps [18]

จากการจำลองการทำงานของระบบได้ใช้สัญญาณเสียงพูดตามรูปที่ 4.1 แล้วทำการเข้ารหัสสัญญาณเสียงพูดตามมาตรฐานข้างต้น พบว่า การเข้ารหัสและถอดรหัสสัญญาณเสียงพูดมีส่วนทำให้สหสัมพันธ์ของสัญญาณเสียงสเตอริโอทั้ง 2 ช่องสัญญาณลดน้อยลง ส่งผลให้ระบบ SAEC มีประสิทธิภาพการทำงานที่ดีขึ้นตามผลค่า WEVN และค่า Coherence Function ดังรูปที่ 4.19 (a) และ (b)





รูปที่ 4.19 (a)



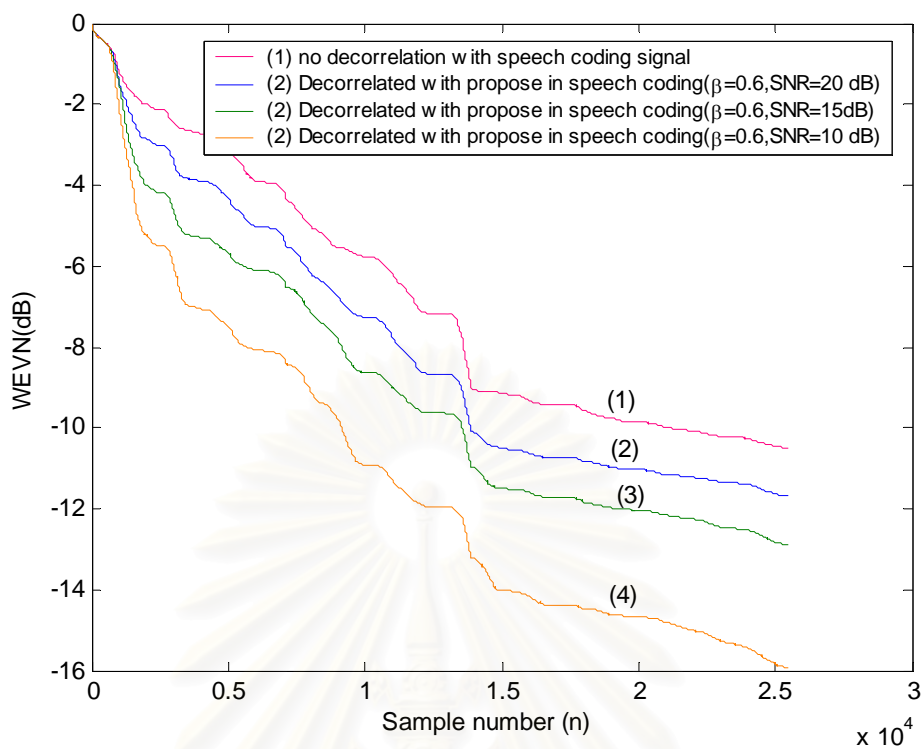
รูปที่ 4.19 (b)

รูปที่ 4.19 (a) การเปรียบเทียบ WEVN ของระบบ SAEC และ (b) การเปรียบเทียบค่า Coherence Function ของสัญญาณเสียงสเตอริโอ กรณีที่มีการเข้ารหัสและถอดรหัสสัญญาณเสียงพูด ตามมาตรฐาน G.723.1 กับระบบที่ใช้สัญญาณเสียงเดิมที่ยังไม่ได้เข้ารหัสและถอดรหัส

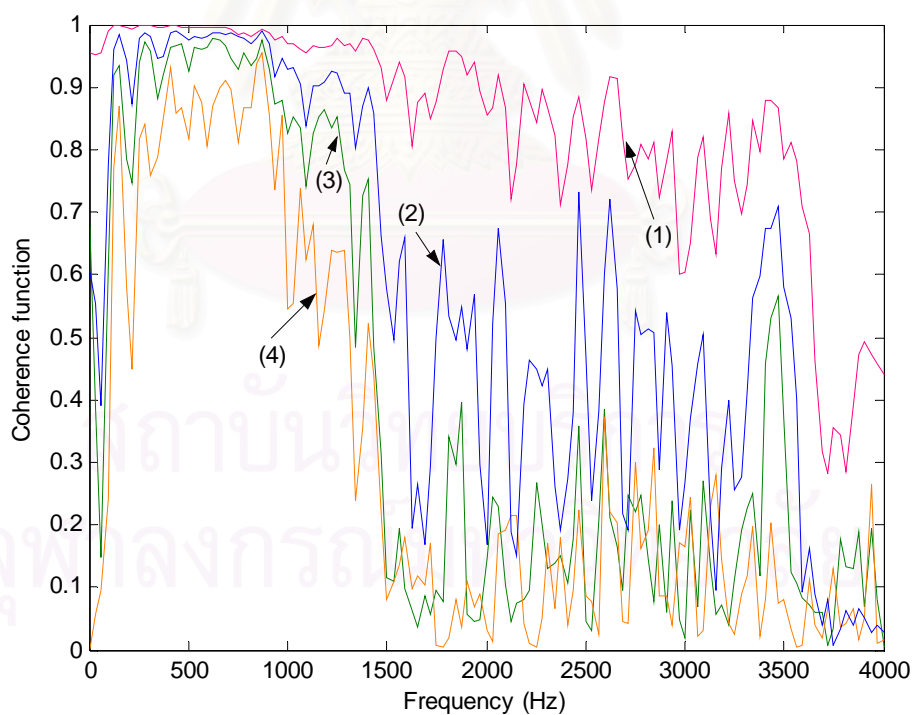
กรณีที่จำลองระบบ SAEC จากสัญญาณเสียงดิจิทัลที่มีความถี่ซ้ำตัวอย่าง 8 kHz โดยที่ยังไม่ได้เข้ารหัสตามมาตรฐาน G.723.1 ค่า WEVN เป็นดังกราฟเส้นที่ (1) ในรูปที่ 4.19 (a) มีค่าประมาณ -7 dB เมื่อเปรียบเทียบกับค่า WEVN ของระบบ SAEC เมื่อทำการเข้ารหัสและถอดรหัสสัญญาณเสียงสเตอริโอก่อนผ่านสัญญาณเข้ารหัสระบบ SAEC พบว่า ผลของความเพี้ยน (Distortion) ที่เกิดจากการเข้ารหัสและถอดรหัสสัญญาณเสียงพูดมีผลทำให้สหสัมพันธ์ของสัญญาณเสียงสเตอริโอลดน้อยลงได้ สังเกตจากค่า WEVN ที่สามารถลดลงได้น้อยกว่า คือลดลงได้ถึงประมาณ -10 dB หรือถ้าสังเกตจากค่า Coherence Function ในรูปที่ 4.19 (b) ความสัมพันธ์ของสัญญาณในกราฟเส้นที่ (2) จะถูกทำให้ลดลงได้มากกว่ากราฟเส้นที่ (1) ในช่วงความถี่ที่มีสัญญาณเสียงพูดอยู่ สำหรับกรณีที่ใช้อัตราบิตการเข้ารหัสข้อมูล 5.3 kbps กราฟค่า WEVN และ Coherence Function ที่ได้ คงยังมีความใกล้เคียงกันกับรูปที่ 4.19 (a) และ 4.19 (b) ดังนั้นจึงไม่ได้แสดงให้เห็นเนื่องจากค่าที่ได้มีความใกล้เคียงกันมาก ทั้งนี้เพื่อเป็นการทดสอบถึงวิธีการดีคอดิเลชันสัญญาณที่นำเสนอว่าสามารถที่จะนำมาใช้ร่วมด้วยกับระบบที่มีการเข้ารหัสสัญญาณเสียงตามมาตรฐานดังกล่าวนี้ได้หรือไม่ ซึ่งจะแสดงผลการจำลองการทำงานของระบบในหัวข้อถัดไป

#### 4.6.2 ผลการจำลองการทำงานของระบบ SAEC ที่มีการเข้ารหัสสัญญาณเสียงตามมาตรฐาน G.723.1 เมื่อมีการดีคอดิเลชันสัญญาณเข้าแบบสเตอริโอ

เมื่อจำลองการทำงานของระบบ SAEC ที่มีการเข้ารหัสและถอดรหัสสัญญาณเสียงพูดและประยุกต์นำวิธีดีคอดิเลชันสัญญาณเสียงสเตอริโอที่ได้นำเสนอไปในหัวข้อที่ 3.1.3.2 มาใช้พร้อมกัน โดยพารามิเตอร์ที่ใช้คือ สัญญาณเสียงสเตอริโอเป็นสัญญาณเดียวกับที่ใช้ในข้อที่ 4.6.1 ค่าพารามิเตอร์ของสัญญาณรบกวนแบบปรับตัวได้เป็นไปตามหัวข้อที่ 4.4 เนื่องจากเป็นค่าที่ทำให้ระบบ SAEC มีประสิทธิภาพการทำงานที่ดีขณะที่ยังสามารถรักษาคุณภาพเสียงไว้ได้ แล้วสังเกตผลการทำงานของระบบ SAEC จากค่า WEVN และ Coherence Function ของสัญญาณเสียงสเตอริโอ ดังแสดงในรูปที่ 4.20



รูปที่ 4.20 (a)



รูปที่ 4.20 (b)

รูปที่ 4.20 (a) ค่า WEVN ของระบบ SAEC และ (b) ค่า Coherence Function ของสัญญาณเสียง

สเตอริโอ ที่ใช้การเข้ารหัสและถอดรหัสตามมาตรฐาน G.723.1

เมื่อใช้วิธีดีคორเรลชันสัญญาณเสียงสเตอริโอดังกล่าวด้วยวิธีที่นำเสนอ

ผลการจำลองการทำงานของระบบ SAEC ดังกล่าว พบว่ากรณีที่มีการประยุกต์นำเอาวิธีดีคอรี่เดชันที่ได้นำเสนอไว้ในหัวข้อที่ 4.4 มาใช้กับระบบ SAEC ที่มีการเข้ารหัสสัญญาณเสียง พบว่าสามารถส่งผลให้ระบบ SAEC ทำงานได้ดีขึ้นกัน เมื่อเราเพิ่มค่าสัญญาณรบกวนแบบปรับตัวได้ให้เพิ่มขึ้น ค่า WEVN มีค่าค่อยๆลดลง ตามกราฟเส้นที่ (1) ถึง (4) ในรูปที่ 4.20 (a) ส่วนค่า Coherence Function ในรูปที่ 4.20 (b) ก็ค่อยๆ ลดลง ซึ่งหมายถึงสหสัมพันธ์ของสัญญาณเสียงสเตอริโอที่ลดลงนั่นเอง ทั้งนี้แสดงให้เห็นว่ากราฟที่ได้มีค่าใกล้เคียงกับกราฟในรูปที่ 4.15 กล่าวได้ว่าวิธีการดีคอรี่เดชันสัญญาณเสียงสเตอริโอตามที่น่าเสนอ สามารถนำไปใช้ร่วมกับระบบ SAEC ที่ระบบมีการเข้ารหัสและถอดรหัสสัญญาณเสียงพูดตามมาตรฐาน G.723.1 ได้



สถาบันวิทยบริการ  
จุฬาลงกรณ์มหาวิทยาลัย

## บทที่ 5

### สรุปผลการวิจัยและข้อเสนอแนะในอนาคต

#### 5.1 สรุปผลการวิจัย

วิทยานิพนธ์ฉบับนี้นำเสนอวิธีการดีคอรืเลขันสัญญาณเสียงสเตอริโอโดยการเพิ่มสัญญาณรบกวนแบบปรับตัวได้ เพื่อพัฒนาให้ประสิทธิภาพการทำงานของระบบ SAEC ทำงานได้ดีขึ้น เนื่องจากปกติประสิทธิภาพการทำงานของระบบ SAEC ค่อนข้างทำได้ยุ่งยากกว่าในระบบ AEC อันเป็นผลจากการที่สัญญาณเข้าแบบสเตอริโอมีสหสัมพันธ์กันค่อนข้างสูง ดังนั้นวิธีหนึ่งในการเพิ่มประสิทธิภาพการทำงานของระบบ SAEC คือการดีคอรืเลขันสัญญาณเสียงสเตอริโอดังกล่าวเพื่อให้สัญญาณเสียงสเตอริโอมีสหสัมพันธ์กันน้อยลงก่อนที่จะผ่านสัญญาณเข้าสู่ระบบ SAEC ตามปกติ ส่งผลให้ตัวกรองแบบปรับตัวในระบบ SAEC สามารถประมาณวิถีสะท้อนทางเสียงได้ใกล้เคียงกับวิถีสะท้อนทางเสียงจริงมากที่สุด

วิธีดีคอรืเลขันสัญญาณเสียงสเตอริโอวิธีหนึ่งที่ถูกนำเสนอมาในอดีตและได้รับความนิยมเป็นอย่างมาก คือการเพิ่มภาวะไม่เชิงเส้น เนื่องจากเป็นวิธีที่สามารถรักษาคุณภาพสัญญาณเสียงสเตอริโอที่ผ่านการดีคอรืเลขันแล้วได้ใกล้เคียงกับคุณภาพสัญญาณเสียงสเตอริโอเดิมค่อนข้างมาก อย่างไรก็ตามวิธีดังกล่าวก็มีข้อจำกัดที่ไม่สามารถเพิ่มประสิทธิภาพการทำงานของระบบ SAEC ได้ดีมากเท่าที่ควร นอกจากนี้วิธีดีคอรืเลขันสัญญาณเสียงสเตอริโออีกวิธีหนึ่งที่มีความสนใจคือ การเพิ่มสัญญาณรบกวนสุ่ม เนื่องจากอาศัยคุณลักษณะเฉพาะของสัญญาณรบกวนที่ไม่มีคอรืเลขันกัน เมื่อเพิ่มสัญญาณรบกวนสุ่มเพื่อการดีคอรืเลขันสัญญาณเสียงสเตอริโอ จึงส่งผลให้สัญญาณเสียงสเตอริโอมีสหสัมพันธ์กันลดน้อยลงได้ แต่ข้อจำกัดของวิธีนี้คือคุณภาพเสียงสเตอริโอที่ผ่านการดีคอรืเลขันแล้วมีคุณภาพด้อยที่สุด โดยเฉพาะอย่างยิ่ง เมื่อระดับของสัญญาณรบกวนสุ่มที่เพิ่มมีค่าสูง ซึ่งผู้ฟังจะสังเกตเห็นได้ชัดเจนว่ามีสัญญาณรบกวนปนอยู่กับสัญญาณเสียงสเตอริโอด้วย

ดังนั้นวิธีการดีคอรืเลขันที่นำเสนอในวิทยานิพนธ์นี้คือ การดีคอรืเลขันสัญญาณเสียงสเตอริโอโดยการเพิ่มสัญญาณรบกวนแบบปรับตัวได้ จึงได้นำข้อดีของแต่ละวิธีดังกล่าวมาใช้คือ การใช้คุณสมบัติเด่นของการดีคอรืเลขันสัญญาณจากการเพิ่มสัญญาณรบกวนสุ่ม แต่ปรับเปลี่ยนให้สัญญาณรบกวนดังกล่าวมีขนาดเปลี่ยนไปตามขนาดของสัญญาณเสียงสเตอริโอเพื่อไม่ให้มีผลกระทบต่อคุณภาพสัญญาณเสียงที่ถูกดีคอรืเลขันแล้ว ดังเช่นในวิธีที่ใช้สัญญาณรบกวนสุ่ม โดยใช้สัดส่วน

SNR ของสัญญาณเสียงสเตอริโอต่อสัญญาณรบกวนขาวแบบเกาส์ที่เปลี่ยนไปตามเวลา นอกจากนี้ การเพิ่มภาวะไม่เป็นเชิงเส้นเพื่อปรับเปลี่ยนขนาดสัญญาณรบกวนที่เพิ่มให้ระบบตามค่าสัดส่วนของ พลังงานเฉลี่ยของสัญญาณเสียงพูด ก็สามารถลดสหสัมพันธ์ของสัญญาณเสียงสเตอริโอลงได้ โดยยัง รักษาไว้ซึ่งคุณภาพเสียงสเตอริโอให้ใกล้เคียงกับสัญญาณเสียงเดิมมากที่สุด ดังนั้นระบบ SAEC จึง สามารถทำงานได้อย่างมีประสิทธิภาพเพิ่มมากขึ้น

นอกจากนี้ ผลของการเพิ่มประสิทธิภาพการกำจัดสัญญาณเสียงสะท้อนในระบบ SAEC โดยการเพิ่มสัญญาณรบกวนแบบปรับตัวได้ก็ยังสามารถนำไปประยุกต์ใช้ร่วมกับระบบ SAEC ที่มีการ เข้ารหัสและถอดรหัสสัญญาณเสียงตามมาตรฐาน G.723.1 ได้ โดยระบบ SAEC ยังมีประสิทธิภาพ การทำงานที่ดีขึ้น คุณภาพเสียงสเตอริโอที่ได้ก็ไม่มีการเปลี่ยนแปลงมากนัก เมื่อเปรียบเทียบกับผลที่ได้ จากระบบที่ไม่มีการเข้าและถอดรหัสสัญญาณเสียงพูด นอกจากนี้ทำให้พบว่าการเข้าและถอดรหัส สัญญาณเสียงพูดมีผลต่อสหสัมพันธ์ของสัญญาณเสียงสเตอริโอด้วย กล่าวคือ ผลของความเพี้ยนใน การเข้ารหัสและถอดรหัสสัญญาณเสียงพูดสามารถดีคอดรีเลชันสัญญาณเสียงแบบสเตอริโอได้อีกส่วน หนึ่ง แสดงให้เห็นว่า ถ้านำวิธีดีคอดรีเลชันสัญญาณเสียงสเตอริโอที่นำเสนอไปใช้ในระบบการสื่อสารจริง ที่ระบบมีการเข้ารหัสและถอดรหัสสัญญาณเสียงด้วยนั้น ประสิทธิภาพการทำงานของระบบ SAEC จะยังคงเป็นดังเช่นผลที่ได้จากการจำลองระบบบนคอมพิวเตอร์ ดังได้เสนอไว้ในวิทยานิพนธ์นี้

## 5.2 ข้อเสนอแนะสำหรับงานวิจัยในอนาคต

งานวิจัยนี้สามารถนำไปประยุกต์ใช้กับระบบการสื่อสารทางเสียงสเตอริโอได้ รวมถึงระบบที่มี การเข้ารหัสและถอดรหัสสัญญาณเสียงพูดแบบสเตอริโอตามมาตรฐานที่มีอยู่เพื่อเพิ่มประสิทธิภาพการ ทำงานของระบบการกำจัดสัญญาณเสียงสะท้อน โดยควรคำนึงถึงคุณภาพสัญญาณเสียงสเตอริโอที่ถูก ดีคอดรีเลชันแล้วเป็นอย่างยิ่งว่าจะจะเป็นไปตามเกณฑ์ที่ยอมรับได้หรือไม่ สำหรับการปรับค่าหรือขนาด สัญญาณรบกวนที่เปลี่ยนไปตามเวลาเพื่อเพิ่มเข้าไปเพื่อดีคอดรีเลชันสัญญาณเสียงสเตอริโอ นั้น อาจจะ ปรับปรุงให้มีขนาดเปลี่ยนไปจากวิทยานิพนธ์ฉบับนี้ได้ การศึกษาถึงค่าจุดเปลี่ยนหรือระบบการตัดสินใจ คอบที่มีเสียงพูดและคอบเงียบที่สามารถปรับเปลี่ยนได้ตามสัญญาณเข้าในสถานการณ์ต่างๆ เช่น ระบบที่สัญญาณเข้ามีการเปลี่ยนแปลงฉับพลันหรือมีสัญญาณรบกวนพื้นหลังเกิดขึ้นบางช่วง และ การศึกษาถึงผลของวิธีสะท้อนทางเสียงที่เป็นระบบเปลี่ยนตามเวลา ก็เป็นหัวข้อที่สามารถนำไป พัฒนาต่อไปได้

## รายการอ้างอิง

1. M. M. Sondhi, D. R. Morgan, and J. L. Hall. "Stereophonic Acoustic Echo Cancellation - An Overview of the Fundamental Problem". IEEE Signal Processing Letter Vol. 2, No. 8 (August 1995) : 148 - 151.
2. J. Benesty, D. R. Morgan, and M. M. Sondhi. "A Better Understanding and An Improved Solution to the Specific Problems of Stereophonic Acoustic Echo Cancellation". IEEE Transactions on Speech and Audio Processing Vol. 6, No. 2 (March 1998) : 156 - 165.
3. J. Benesty, D. R. Morgan, and M. M. Sondhi. "A Better Understanding and An Improved Solution to the Problem of Stereophonic Acoustic Echo Cancellation". Proceedings of IEEE ICASSP Vol. 1 (April 1997) : 303 - 306.
4. J. Benesty, F. Amand, A. Gilloire, and Y. Grenier. "Adaptive Filtering Algorithms for Stereophonic Acoustic Echo Cancellation". Proceedings of IEEE International Conference on Acoustics Speech and Signal Processing Vol. 5 (May 1995) : 3099 - 3101.
5. D. R. Morgan. "Investigation of Several Types of Nonlinearities for Use in Stereo Acoustic Echo Cancellation". IEEE Transactions on Speech and Audio processing Vol. 9, No.6 (September 2001) : 686-695.
6. T. N. Yensen, R. A. Goubran, and I. Lambadaris. "Synthetic Stereo Acoustic Echo Cancellation Structure for Multiple Participant VOIP Conferences". IEEE Transactions on Speech and Audio Processing Vol. 9, No.2 (February 2001) : 168 - 174.
7. M. M. Sondhi, and D. R. Morgan. "Echo Cancellation for Stereophonic Teleconferencing". Proceedings of IEEE Acoustics Speech and Signal Processing Workshop on Applications of Signal Processing to Audio and Acoustic Vol. 64 (May 1991) : 2093-2096.

8. T. Gansler, and P. Eneroth. "Influence of Audio Coding on Stereophonic Acoustic Echo Cancellation". Proceedings of IEEE International Conference on Acoustics Speech and Signal Processing Vol. 6 (May 1998) : 3649 – 3652.
9. A. Gilloire, and V. Turbin. "Using Auditory Properties to Improve the Behaviour of Stereophonic Acoustic Echo Cancellers". Proceedings of IEEE International Conference on Acoustics Speech and Signal Processing Vol. 6 (May 1998) : 3681 – 3684.
10. Y. Jung, J. H. Lee, Y. C. Park, and D. H. Youn. "A new adaptive algorithm for stereophonic acoustic echo canceller". Proceedings of IEEE International Conference on Acoustics Speech and Signal Processing Vol. 2 (June 2000) : II801 - II804.
11. S. Minami. "A stereophonic echo canceler using single adaptive filter". Proceedings of IEEE International Conference on Acoustics Speech and Signal Processing Vol. 5 (May 1995) : 3027 - 3030.
12. N. Tangsangiumvisai. 2002. "Introduction to Adaptive Filters for Acoustic Echo Cancellation in Hands-Free Mobile Communication" (Laboratory sheet). Communication Engineering Laboratory : Chulalongkorn University, Thailand.
13. H. Hayes. 1991. "Adaptive Filter Theory". United States of America : Prentice Hall.
14. S. R. Quackenbush, T. P. Barnwell, and M. A. Clements. 1988. "Objective Measures of Speech Quality". United States of America : Prentice Hall.
15. H. Hayes. 1996. "Statistical digital Signal Processing and Modeling". United States of America : John Wiley & Sons.
16. N. Tangsangiumvisai. 2002. "Statistical Digital Signal Processing" (Handout). Communication Engineering Laboratory : Chulalongkorn University, Thailand.
17. S. L. Gay, and J. Benesty. 2000. "Acoustic Signal Processing for Telecommunication". 1<sup>st</sup> edition. Boston : Kluwer Academic Publisher.
18. ITU Recommendation. February 2002. "G.723.1 Recommendation".



19. A. M. Kondoz. January 2000. "Digital Speech". United States of America : John Wiley & Sons.
20. L. Rabiner, and B. H. Juang. 1993. "Fundamentals of speech recognition". United States of America : Prentice Hall.
21. G. U. Schmidt. "Acoustic Echo and Noise Control for Low-Cost Processor". Published at DSP World Spring Design Conference 2000 (April 2000).
22. "<http://svr-www.eng.cam.ac.uk/comp.speech/>" : Speech Applications Group, Sun Microsystems Laboratories.



สถาบันวิทยบริการ  
จุฬาลงกรณ์มหาวิทยาลัย



ภาคผนวก

สถาบันวิทยบริการ  
จุฬาลงกรณ์มหาวิทยาลัย

## บทความที่ได้รับการเผยแพร่

1. P. Surin, N. Tangsangiumvisai and S. Aramvith. "A New Decorrelation Technique for Stereophonic Acoustic Echo Cancellation". Proceedings of Thailand Electrical Engineering Conference (EECON 27<sup>th</sup>) Vol. 2 (November 2004) : 213 – 216.
2. P. Surin, N. Tangsangiumvisai and S. Aramvith. "An Adaptive Noise Decorrelation Technique for Stereophonic Acoustic Echo Cancellation". Proceedings of IEEE international Conference (TENCON2004) on Analog And Digital Technique for Electrical Engineering Vol. A, No. 1 (November 2004) : 112 - 115.



สถาบันวิทยบริการ  
จุฬาลงกรณ์มหาวิทยาลัย

# A NEW SIGNAL DECORRELATION TECHNIQUE FOR STEREOPHONIC ACOUSTIC ECHO CANCELLATION

P. Surin, N. Tangsangiumvisai, and S. Aramvith

Digital Signal Processing Research Laboratory (DSPRL),  
Department of Electrical Engineering, Faculty of Engineering, Chulalongkorn University,  
Bangkok, 10330, Thailand, *e-mail*: Nisachon.T@eng.chula.ac.th

## ABSTRACT

In this paper, a new signal decorrelation technique, based upon adaptive noise addition is presented for Stereophonic Acoustic Echo Cancellation (SAEC). The added noise level is adjusted according to the characteristics of the speech inputs, hence, the quality of the decorrelated stereo speech signals are kept undisturbed. In addition, the computational complexity for the proposed technique is insignificant. Test results on recorded speech signals indicate improved performance in both objective and subjective manners.

## 1. INTRODUCTION

Voice communication systems nowadays have shown the necessity of multiple channels, such as within a stereo teleconferencing system, so that the audience is able to indicate the presence of the talkers in a more realistic manner. Within such an audio system, echoes normally/naturally occur due to acoustic coupling between loudspeakers and microphones. This phenomenon, however, disturbs the conversation. Stereophonic Acoustic Echo Cancellation (SAEC) is therefore required to mitigate the effect of such coupling. A schematic diagram of an SAEC system in a stereo teleconferencing environment is illustrated in Fig. 1, where the transmitting and receiving rooms can be referred to as far-end and near-end rooms, respectively. The mathematical explanation of the SAEC situation will be given in the next section.

Since the two loudspeaker signals in the near-end room of a stereo teleconferencing system come from a common source in the far-end room, this results in a strong correlation between the transmitted stereo signal and thus prevents the two adaptive filters of the SAEC system from identifying the impulse responses of both unknown acoustic echo paths (AEPs) in the near-end room uniquely and correctly. One possible way to tackle this so-called non-uniqueness problem is to decorrelate artificially the stereo speech signals so that the two input signals of the adaptive filters become less correlated to each other. As a result, the adaptive filters converge more correctly and the non-uniqueness problem can be alleviated. However, the quality of the decorrelated stereo signals that travel to the near-end room has to be minimally affected.

Several signal decorrelation techniques have been suggested for SAEC, including random noise addition to loudspeaker signals [1], non-linear transformation of the

loudspeaker signals [2], and the employment of first-order time-varying allpass filters (TV-APFs) [3]. These techniques show improved performance in terms of misalignment of the adaptive filters, but the stereo perception of the signals are somewhat degraded. Other signal decorrelation techniques based on human psychoacoustics have also been proposed in [4], [5]. In [4], spectral-shaped random noise, which is designed to be auxiliary signals based on human auditory properties are added to the stereo signals in order to make the added noise inaudible. Although this approach claims to yield more efficient performance than the non-linear transformation method in [3], the operation of the frequency masking results in the undesirable processing delay. To reduce the computational complexity of this process, it is suggested in [5] that the masking patterns can be approximated by employing the Autoregressive (AR) analysis for the source of the transmitting speech signals. The AR analysis is used to generate speech-like noise in each channel so that the quality of stereo speech signals is not affected after the addition of such noise. Moreover, a time-varying gain is utilized in order to keep the added noise level under the masking patterns. However, the computational complexity of the AR analysis part within the decorrelation technique in [5] is rather high. In addition, the employment of the GS-AP algorithm, as suggested in [5] is unnecessary. Once the stereo input signals of the SAEC system are decorrelated adequately by any signal decorrelation technique, the Normalised Least Mean Square (NLMS) algorithm is sufficient to eliminate the acoustic echo signal at low implementation cost.

In this paper, a signal decorrelation technique based on the criterion of random noise addition to stereo input signals, using adaptive noise addition is proposed. The added noise level is controlled by stereo input energy. Improvement of the SAEC system in terms of misalignment is demonstrated in this paper, together with minimum degradation of stereo signal perception based upon on the subjective listening test. Furthermore, the computational complexity of the proposed technique is insignificant, as compared to that in [5].

This paper is organized as follows. Section II described the structure of the SAEC system. The proposed adaptive noise addition for signal decorrelation technique is introduced in Section III, followed by the experimental results based on stereo speech signals in Section IV. The Conclusions are given in Section V.

## 2. AN SAEC SYSTEM

This section describes the structure of an SAEC system in brief. Two microphones in the far-end room pick up speech signals from a source  $s(n)$  through two AEPs, which are characterized by impulse responses  $\mathbf{g}_i(n)$  for  $i = 1, 2$ . These two stereo signals,  $x_i(n)$ , are transmitted to the near-end room and travel to one of the microphones via another set of AEPs, which are denoted by  $L$ -tap impulse responses  $\mathbf{h}_i(n)$ . Note that only half of the system is considered in Fig. 1, i.e. one of the two microphones in the near-end room is neglected, due to similar comments for its independent AEPs  $\mathbf{h}_3(n)$  and  $\mathbf{h}_4(n)$  between another pair of loudspeakers and microphones.

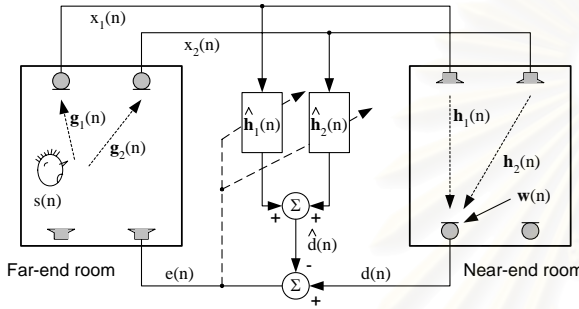


Figure 1: A diagram for Stereophonic Acoustic Echo Cancellation.

Two Finite Impulse Response (FIR) adaptive filters are needed for the identification of the two unknown AEPs,  $\mathbf{h}_i(n)$ , in the near-end room. In this paper, the length of the adaptive filters is assumed to be the same as those of the unknown AEPs in the near-end room. The adaptive filters  $\hat{\mathbf{h}}_i(n)$  try to identify  $\mathbf{h}_i(n)$ , yielding the output sum of

$$\hat{d}(n) = \sum_{i=1}^2 \hat{\mathbf{h}}_i^T(n) \mathbf{x}_i(n) \quad (1)$$

where  $\mathbf{x}_i(n) = [x_i(n), x_i(n-1), \dots, x_i(n-L+1)]^T$  is the input signal vector and the adaptive filters are given by  $\hat{\mathbf{h}}_i(n) = [\hat{h}_{i,1}(n), \dots, \hat{h}_{i,L}(n)]^T$ . The microphone signal in the near-end room is given by

$$d(n) = \sum_{i=1}^2 \mathbf{h}_i^T(n) \mathbf{x}_i(n) + w(n) \quad (2)$$

where  $w(n)$  is the background noise that is assumed to be of zero-mean and uncorrelated with the input signals. The acoustic echo signal can be eliminated when the error signal, given by

$$e(n) = \sum_{i=1}^2 [\mathbf{h}_i(n) - \hat{\mathbf{h}}_i(n)]^T \mathbf{x}_i(n) + w(n) \quad (3)$$

is as close to zero as possible.

Due to strong correlation of the stereo input signals, the non-uniqueness problem exists [1], and results in difficulty of the identification of the two unknown AEPs in the near-end room. This is best shown in the frequency-domain representation. By assuming that the echo signal has been completely eliminated, it

can be observed that the error signal approaches zero, i.e.  $\lim_{n \rightarrow \infty} e(n) = 0$ . Thus, the error signal can be represented in the frequency-domain as

$$\sum_{i=1}^2 [H_i(\omega) - \hat{H}_i(\omega)]^T G_i(\omega) = 0 \quad (4)$$

where  $H_i(\omega)$  and  $\hat{H}_i(\omega)$  are the frequency-domain representation of the near-end room impulse responses and the converged adaptive filters, respectively,  $\omega = 2\pi k/L$  and  $k = 0, 1, \dots, L-1$ . Given that the impulse responses in the far-end room are not equal to zero;  $G_i(\omega) \neq 0$ , it does not imply that there is a unique solution to (4). In fact, there can be many solutions obtained by the adaptive filters. Thus, the unknown AEPs cannot be identified uniquely. In addition, the variation of the impulse responses  $G_i(\omega)$  has strong impact on the identification of the unknown systems [1]. Hence, it is necessary to identify the unknown systems correctly and uniquely.

## 3. A SIGNAL DECORRELATION TECHNIQUE USING ADAPTIVE NOISE ADDITION

The performance of the SAEC system can be improved by employing a pre-processing block to decorrelate the transmitted stereo signals, while keeping the information of the transmitted signals unchanged. In addition, the quality of the decorrelated signals, which are the loudspeaker signals in the near-end room, should be minimally affected. In this section, a proposed signal decorrelation technique based upon adaptive noise addition is present in order to decorrelate the stereo signals sufficiently while guaranteeing the imperceptible noise level in the decorrelated stereo signals.

The proposed technique involves three steps of investigation as follows. First, the level of random noise addition to each stereo signal can be varied according to signal-to-noise ratio (SNR) between the stereo speech signal and the added random noise signal in that channel. This variable gain can be obtained from

$$\gamma_i(n) = \sqrt{\frac{x_i^2(n)}{v_i^2(n)} \times 10^{-\left(\frac{SNR}{10}\right)}} \quad (5)$$

where  $v_i(n)$  is a white Gaussian noise sequence, for  $i = 1, 2$ . The decorrelated stereo signals obtained using this variable gain  $\gamma_i(n)$  are given by

$$x_i'(n) = x_i(n) + \gamma_i(n) \cdot v_i(n) \quad (6)$$

This approach is expected to give approximately the same capability for signal decorrelation as the random noise addition in [1]. The advantage is due to the variable SNR of the noise level added to the stereo signals, as compared to the fixed noise level in [1]. Hence, the quality of the decorrelated stereo signals employing this approach should be less degraded.

Second, another type of signal decorrelation based on non-linear transformation is observed. The decorrelated signals using this technique are given by

$$x'_i(n) = x_i(n) + \beta \cdot P_{x_i}(n) \quad (7)$$

The parameter  $\beta$  is used to control the level of non-linearity added to the stereo speech signals in each channel. The energy of the speech input signal in each channel is found from

$$P_{x_i}(n) = \frac{1}{K} \sum_{j=0}^{K-1} |x_i(n-j)|^2 \quad (8)$$

and  $K$  is the averaged window length. In fact,  $P_{x_i}(n)$  can be found recursively from

$$P_{x_i}(n) = P_{x_i}(n-1) + x_i^2(n) - x_i^2(n-K+1) \quad (9)$$

As compared to the non-linear transformation using half-wave rectifier in [2], this technique should give approximate capability for signal decorrelation.

In order to maximize the signal decorrelation level, the adaptive noise addition technique is proposed. The variable gain random noise addition and the non-linear transformation methods are combined. The decorrelated stereo signals become

$$x'_i(n) = x_i(n) + \beta \cdot P_{x_i}(n) \cdot \gamma_i(n) \cdot v_i(n) \quad (10)$$

By controlling the level of noise signal  $v_i(n)$  according to the energy of the speech input signal  $P_{x_i}(n)$ , the added noise is imperceptible. In terms of the computational complexity, the energy of speech input signal in each channel will already be calculated in the adaptation part of the adaptive filters, provided that the NLMS algorithm are employed. Thus, by choosing the average window length to be  $K = L$ , the additional computational complexity required for calculating the adaptive noise level is insignificant.

## 4. EXPERIMENTAL RESULTS

### 4.1. Simulation Results

The experiments were carried out under a statistically stationary assumption for the near-end room, i.e. the coefficients of the AEPs were fixed and modeled as given in Fig. 2 of length  $L = 256$ . Stereo speech signals, sampled at 8kHz, were used as the input signals of the system. The background noise at the microphone in the near-end room was 50dB SNR. The adaptive filters employed in this system were also of length  $L = 256$  and controlled by the two-channel NLMS algorithm with the step-size parameter  $\mu = 0.8$ .

For comparison, the Weight Error Vector Norm (WEVN) performance was evaluated. The WEVN performance of the SAEC system employing the signal decorrelation techniques using random noise addition in [1], non-linear transformation in [2], and those three steps of the proposed technique are illustrated in Fig. 3. As compared to the SAEC system without any decorrelation methods, the WEVN performance of system employing the random noise addition method in [1] with SNR of 30dB was improved by 9dB. When employing the random noise addition with variable gain

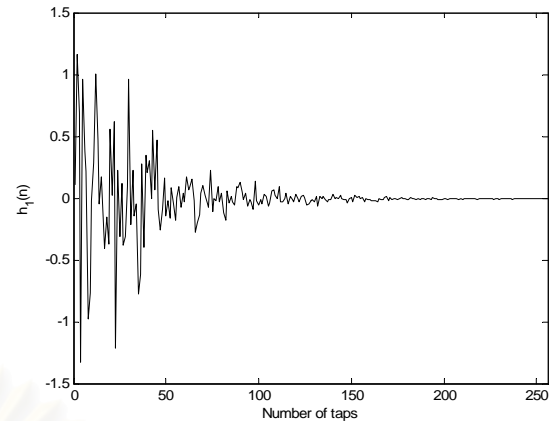


Figure 2: Acoustic Echo Path  $h_1(n)$  in the near-end room.

$\gamma_i(n)$ , the improvement of WEVN of 8dB was obtained. For the non-linear transformation method in [2] when the non-linear factor was set to be  $\alpha = 0.8$  in this paper to maintain the speech quality, the WEVN performance was improved by 3dB. Similarly to the method in [2], the SAEC system with non-linear transformation as given in (7) and (8) reached about the same level of WEVN when the parameter  $\beta = 0.6$  was chosen. For the proposed adaptive noise addition, with  $\beta = 0.6$ , the WEVN performance can be improved by 7dB.

In addition, the magnitude squared coherence, defined as

$$\gamma_{12}(f) = \frac{|P_{x_1 x_2}(f)|^2}{P_{x_1 x_1}(f) P_{x_2 x_2}(f)} \quad (11)$$

was used to indicate the correlation between the stereo signals in each channel at each frequency, i.e.  $\gamma_{12}(f) \in [0, 1]$ , where  $P_{x_1 x_1}(f)$  and  $P_{x_2 x_2}(f)$  were the power spectra of the speech input signals  $x_1(n)$  and  $x_2(n)$ , and the cross-spectral density between  $x_1(n)$  and  $x_2(n)$  was  $P_{x_1 x_2}(f)$ . From Fig. 4, the coherence function of the stereo speech signals with no decorrelation methods is about unity at all frequency. The technique that gives the best decorrelation level of the stereo signals among these methods was the proposed method employing adaptive noise addition.

### 4.2. Subjective Listening Test

Based on the ITU recommendation, the first one subjective listening tests were undertaken with 20 listeners. The original stereo signals were compared with the decorrelated stereo signals employing the random noise addition technique in [1], the non-linear transformation in [2], and those from three steps of the proposed method employing adaptive noise addition. The Mean Opinion Score (MOS) value of the original stereo signals was 4.38, whereas those decorrelated stereo signals were graded as given in Table 1.

The test shows that the signal decorrelation approaches employing the proposed adaptive noise

addition and the non-linear transformation preserve much better stereo perception of the signals than the other investigated methods, whereas the technique using random noise addition in [1] gives the lowest MOS value. It can be seen that the signal decorrelation performance and the ability to maintain the quality of the processed stereo signals have to be compromised.

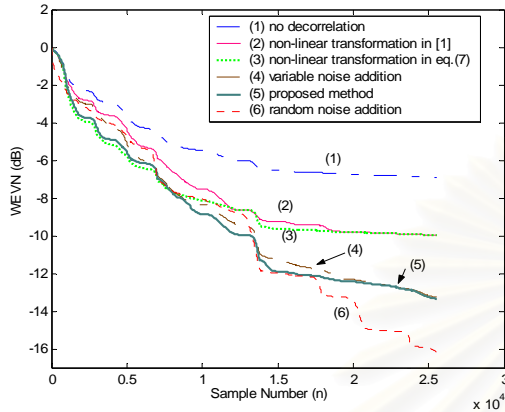


Figure 3: Comparison of WEVN performance.

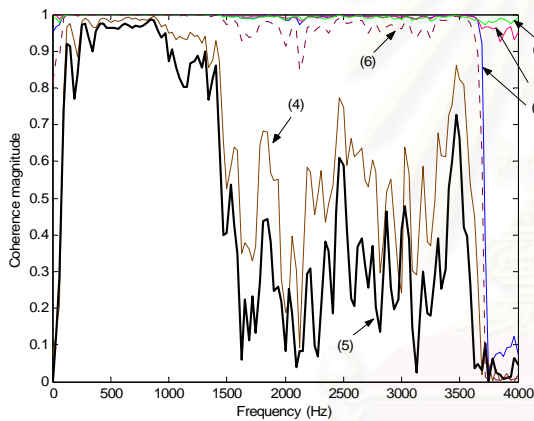


Figure 4: Coherence function of decorrelated stereo signals employing the proposed adaptive noise addition.

Table 1: Mean Opinion Score value of the stereo signals employing different decorrelation methods.

Signals	MOS values
No signal decorrelation method	4.38
Non-linear transformation in [2]	3.56
Non-linear transformation in (7)	2.22
Variable noise addition in (6)	2.88
Adaptive noise addition	3.50
Random noise addition in [1]	2.00

In addition, the effect of SNR level of the added noise signals to the SAEC system is shown. By carrying out another subjective listening test, the MOS values of the proposed adaptive noise addition and the random noise addition techniques were investigated. At the same

level of added noise, the proposed technique offers higher MOS values than the method in [1]. Moreover, the proposed technique yields more signal decorrelation level as compared to the method in [1] while preserving better quality of the decorrelated signals than the one in [1].

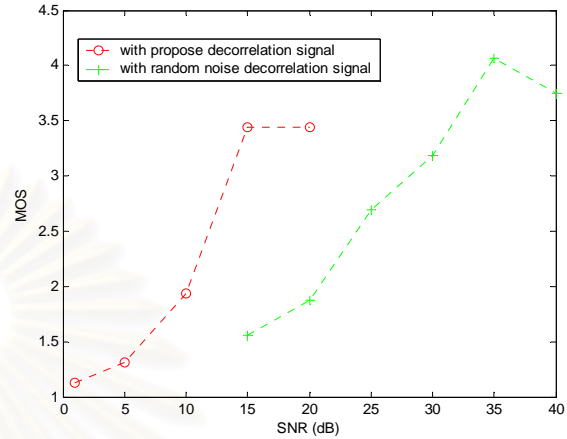


Figure 5: Effect of SNR level on MOS score for the signal decorrelation in [1] and the proposed technique.

## 5. CONCLUSIONS

It has been shown in terms of WEVN that stereo signals can be decorrelated efficiently by employing the proposed low-complexity adaptive noise addition technique. Furthermore, the proposed technique results in minimum degradation of the quality of the stereo signals, as compared to the other techniques investigated in this paper.

## 6. REFERENCES

- [1] M. M. Sondhi, D. R. Morgan, and J. L. Hall, "Stereophonic Acoustic Echo Cancellation – An Overview of the Fundamental Problem," *IEEE Signal Processing Letters*, pp. 148-151, August 1995.
- [2] J. Benesty, D. R. Morgan, and M. M. Sondhi, "A better understanding and an improved solution to the problems of stereophonic acoustic echo cancellation", *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'97, Munich, Germany)*, vol. 1, pp. 303-306, April 1997.
- [3] M. Ali, "Stereophonic acoustic echo cancellation system using time-varying all-pass filtering for signal decorrelation", *Proceedings of ICASSP'98 (Seattle, U.S.A.)*, vol. 6, pp. 3689-3692, May 1998.
- [4] A. Gilloire and V. Turbin, "Using auditory properties to improve the behaviour of stereophonic acoustic echo cancellers", *Proceedings of ICASSP'98 (Seattle, U.S.A.)*, vol. 6, pp. 3681-3684, May 1998.
- [5] Y. W. Jung, J. H. Lee, Y. C. Park and D. H. Youn, "A new adaptive algorithm for stereophonic acoustic echo canceller", *Proceedings of ICASSP'2000 (Istanbul, Turkey)*, vol. 2, pp. 801-804, June 2000.
- [6] E. Zwicker and H. Fastl, "Psychoacoustics : Facts and Models", *Springer-Verlag*, 1990.

# AN ADAPTIVE NOISE DECORRELATION TECHNIQUE FOR STEREOPHONIC ACOUSTIC ECHO CANCELLATION

*P. Surin, N. Tangsangiumvisai, and S. Aramvith*

Digital Signal Processing Research Laboratory (DSPRL),  
Department of Electrical Engineering, Faculty of Engineering, Chulalongkorn University,  
Bangkok, 10330, Thailand, *e-mail*: Nisachon.T@eng.chula.ac.th

## ABSTRACT

In this paper, a new signal decorrelation technique, based upon adaptive noise addition is presented for Stereophonic Acoustic Echo Cancellation (SAEC), with insignificant additional computational complexity. The added noise level is adjusted according to the characteristics of the speech inputs, hence, the quality of the decorrelated stereo speech signals are kept undisturbed. Test results on recorded speech signals indicate improved performance in both objective and subjective manners.

## 1. INTRODUCTION

Voice communication systems nowadays have shown the necessity of multiple channels, such as within a stereo teleconferencing system, so that the audience is able to indicate the presence of the talkers in a more realistic manner. Within such an audio system, echoes normally/naturally occur due to acoustic coupling between loudspeakers and microphones. This phenomenon, however, disturbs the conversation. Stereophonic Acoustic Echo Cancellation (SAEC) is therefore required to mitigate the effect of such coupling. A schematic diagram of an SAEC system in a stereo teleconferencing environment is illustrated in Fig. 1, where the transmitting and receiving rooms can be referred to as far-end and near-end rooms, respectively. The mathematical explanation of the SAEC situation will be given in the next section.

Since the two loudspeaker signals in the near-end room of a stereo teleconferencing system come from a common source in the far-end room, this results in a strong correlation between the transmitted stereo signal and thus prevents the two adaptive filters of the SAEC system from identifying the impulse responses of both unknown acoustic echo paths (AEPs) in the near-end room uniquely and correctly. One possible way to tackle this so-called non-uniqueness problem is to decorrelate artificially the stereo speech signals so that the two input signals of the adaptive filters become less correlated to each other. As a result, the adaptive filters converge more correctly and the non-uniqueness problem can be alleviated. There is another important factor that has to be taken care of when decorrelating the stereo signals. The quality of the decorrelated stereo signals that travel to the near-end room is required to be minimally affected.

Several decorrelation techniques have been suggested, including random noise addition to loudspeaker signals [1], [2], non-linear transformation of the loudspeaker signals [3], [4], and the employment of first-order time-varying allpass filters (TV-APFs) [5]. These techniques show improved performance in terms of misalignment of the adaptive filters, but the stereo perception of the signals are somewhat degraded. Other signal decorrelation techniques based on human psychoacoustics have also been proposed in [6], [7]. In [6], spectral-shaped random noise, which is designed to be auxiliary signals based on human auditory properties are added to the stereo signals in order to make the added noise inaudible. Although this approach claims to yield more efficient performance than the non-linear transformation method in [3], [4], the operation of the frequency masking results in the undesirable processing delay.

The signal decorrelation technique in [7] suggests to approximate the masking patterns by employing the Autoregressive (AR) analysis for the source of the transmitting speech signals in order to reduce the computational complexity, as compared to [6]. The AR analysis is used to generate the speech-like noise in each channel so that the quality of the stereo speech signals is not affected after the addition of such noise. Moreover, a time-varying gain is utilized in order to keep the added noise level under the masking patterns. The computational complexity of the AR analysis part within this technique is still rather high. In addition, the employment of the GS-AP algorithm as suggested in [7] is unnecessary. By applying any signal decorrelation technique within an SAEC system, the Normalised Least Mean Square (NLMS) algorithm is sufficient to eliminate the acoustic echo signal with low implementation cost.

In this paper, a signal decorrelation technique based on the criterion of random noise addition to stereo input signals, using adaptive noise addition is proposed. The added noise level is controlled by the energy of the stereo input signals. Improvement of the SAEC system in terms of the misalignment is demonstrated in this paper, together with minimum degradation of stereo signal perception based on the subjective listening test. Furthermore, the computational complexity of the proposed technique is insignificant as compared to that in [7].

This paper is organized as follows. Section II described the structure of the SAEC system. The



proposed adaptive noise addition for signal decorrelation technique is introduced in Section III, followed by the experimental results based on stereo speech signals in Section IV. The Conclusions are given in Section V.

## 2. THE SAEC SYSTEM

This section describes the structure of an SAEC system in brief. Two microphones in the far-end room pick up speech signals from a source  $s(n)$  through two AEPs, which are characterized by impulse responses  $\mathbf{g}_i(n)$  for  $i = 1, 2$ . These two stereo signals,  $x_i(n)$ , are transmitted to the near-end room and travel to one of the microphones via another set of AEPs, which are denoted by  $L$ -tap impulse responses  $\mathbf{h}_i(n)$ . Note that only half of the system is considered in Fig. 1, i.e. one of the two microphones in the near-end room is neglected, due to similar comments for its independent AEPs  $\mathbf{h}_3(n)$  and  $\mathbf{h}_4(n)$  between another pair of loudspeakers and microphones.

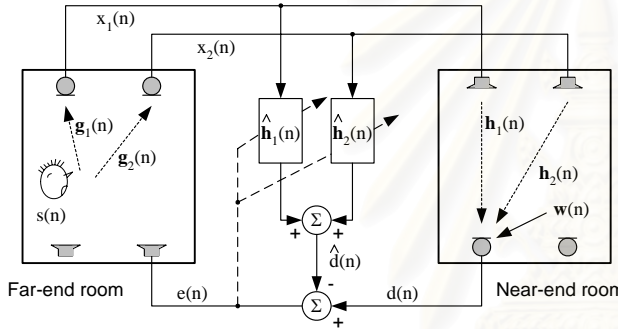


Figure 1: A diagram for Stereophonic Acoustic Echo Cancellation.

Two Finite Impulse Response (FIR) adaptive filters are needed for the identification of the two unknown AEPs,  $\mathbf{h}_i(n)$ , in the near-end room. In this paper, the length of the adaptive filters is assumed to be the same as those of the unknown AEPs in the near-end room. The adaptive filters  $\hat{\mathbf{h}}_i(n)$  try to identify  $\mathbf{h}_i(n)$ , yielding the output sum of

$$\hat{d}(n) = \sum_{i=1}^2 \hat{\mathbf{h}}_i^T(n) \mathbf{x}_i(n) \quad (1)$$

where  $\mathbf{x}_i(n) = [x_i(n), x_i(n-1), \dots, x_i(n-L+1)]^T$  is the input signal vector and the adaptive filters are given by  $\hat{\mathbf{h}}_i(n) = [\hat{h}_{i,1}(n), \dots, \hat{h}_{i,L}(n)]^T$ . The microphone signal in the near-end room is given by

$$d(n) = \sum_{i=1}^2 \mathbf{h}_i^T(n) \mathbf{x}_i(n) + w(n) \quad (2)$$

where  $w(n)$  is the background noise that is assumed to be of zero-mean and uncorrelated with the input signals. The acoustic echo signal can be eliminated when the error signal, given by

$$e(n) = \sum_{i=1}^2 [\mathbf{h}_i(n) - \hat{\mathbf{h}}_i(n)]^T \mathbf{x}_i(n) + w(n) \quad (3)$$

is as close to zero as possible.

Due to strong correlation of the stereo input signals, the non-uniqueness problem exists [1], [2], and results in difficulty of the identification of the two unknown AEPs in the near-end room. This is best shown in the frequency-domain representation. By assuming that the echo signal has been completely eliminated, it can be observed that the error signal approaches zero, i.e.  $\lim_{n \rightarrow \infty} e(n) = 0$ . Thus, the error signal can be represented in the frequency-domain as

$$\sum_{i=1}^2 [H_i(\omega) - \hat{H}_i(\omega)]^T G_i(\omega) = 0 \quad (4)$$

where  $H_i(\omega)$  and  $\hat{H}_i(\omega)$  are the frequency-domain representation of the near-end room impulse responses and the converged adaptive filters, respectively,  $\omega = 2\pi k/L$  and  $k = 0, 1, \dots, L-1$ . Given

that the impulse responses in the far-end room are not equal to zero;  $G_i(\omega) \neq 0$ , it does not imply that there is a unique solution to (4). In fact, there can be many solutions obtained by the adaptive filters. Thus, the unknown AEPs cannot be identified uniquely by the adaptive filters. In addition, the variation of the impulse responses  $G_i(\omega)$  has strong impact on the identification of the unknown systems [2]. Hence, it is therefore necessary to identify the unknown systems correctly and uniquely.

## 3. A SIGNAL DECORRELATION TECHNIQUE USING ADAPTIVE NOISE ADDITION

The performance of the SAEC system can be improved by employing a pre-processing block to decorrelate the transmitted stereo signals, while keeping the information of the transmitted signals unchanged. In addition, the quality of the decorrelated signals, which are the loudspeaker signals in the near-end room, should be minimally affected. In this section, a proposed signal decorrelation technique based upon adaptive noise addition is present in order to decorrelate the stereo signals sufficiently while guaranteeing the imperceptible noise level in the decorrelated stereo signals.

The proposed technique involves three steps of investigation as follows. First, the level of random noise addition to each stereo signal can be varied according to signal-to-noise ratio (SNR) between the stereo speech signal and the added random noise signal in that channel. This variable gain can be obtained from

$$\gamma_i(n) = \sqrt{\frac{x_i^2(n)}{v_i^2(n)} \times 10^{-\left(\frac{SNR}{10}\right)}} \quad (5)$$

where  $v_i(n)$  is a white Gaussian noise sequence, for  $i = 1, 2$ . The decorrelated stereo signals obtained using this variable gain  $\gamma_i(n)$  are given by

$$\hat{x}_i(n) = x_i(n) + \gamma_i(n) \cdot v_i(n) \quad (6)$$

This approach is expected to give approximately the same capability for signal decorrelation as the random

noise addition in [1]. The advantage is due to the variable SNR of the noise level added to the stereo signals, as compared to the fixed noise level in [1]. Hence, the quality of the decorrelated stereo signals employing this approach should be less degraded.

Second, another type of signal decorrelation based on non-linear transformation is observed. The decorrelated signals using this technique are given by

$$\hat{x}_i(n) = x_i(n) + \beta \cdot P_{x_i}(n) \quad (7)$$

The parameter  $\beta$  is used to control the level of non-linearity added to the stereo speech signals in each channel. The energy of the speech input signal in each channel is found from

$$P_{x_i}(n) = \frac{1}{K} \sum_{j=0}^{K-1} |x_i(n-j)|^2 \quad (8)$$

and  $K$  is the averaged window length. In fact,  $P_{x_i}(n)$  can be found recursively from

$$P_{x_i}(n) = P_{x_i}(n-1) + x_i^2(n) - x_i^2(n-K+1) \quad (9)$$

As compared to the non-linear transformation using half-wave rectifier in [3], this technique should give approximate capability for signal decorrelation.

In order to maximize the signal decorrelation level, the adaptive noise addition technique is proposed. The variable gain random noise addition and the non-linear transformation methods are combined. The decorrelated stereo signals become

$$\hat{x}_i(n) = x_i(n) + \beta \cdot P_{x_i}(n) \cdot \gamma_i(n) \cdot v_i(n) \quad (10)$$

By controlling the level of noise signal  $v_i(n)$  according to the energy of the speech input signals  $P_{x_i}(n)$ , the added noise is imperceptible. In terms of the computational complexity, the energy of speech input signal in each channel will already be calculated in the adaptation part of the adaptive filters, provided that the Normalised Least Mean Square (NLMS) algorithm are employed. Thus, by choosing the average window length to be  $K = L$ , the additional computational complexity required for calculating the adaptive noise level is insignificant.

## 4. EXPERIMENTAL RESULTS

### 4.1. Simulation Results

The experiments were carried out under a statistically stationary assumption for the near-end room, i.e. the coefficients of the AEPs were fixed and modeled as given in Fig. 2 of length  $L = 256$ . Stereo speech signals, sampled at  $8\text{kHz}$ , were used as the input signals of the system. The background noise at the microphone in the near-end room was  $50\text{dB}$  SNR. The adaptive filters employed in this system were also of length  $L = 256$  and controlled by the two-channel NLMS algorithm with the step-size parameter  $\mu = 0.8$ .

For comparison, the Weight Error Vector Norm (WEVN) performance, as described in (11) was evaluated.

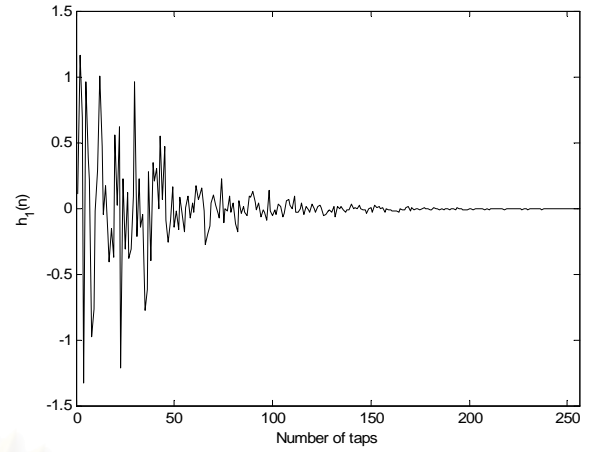


Figure 2: Acoustic Echo Path  $h_i(n)$  in the near-end room.

$$\text{WEVN}(n) = 10 \times \log_{10} \frac{\sum_{i=1}^2 \|\mathbf{h}_i - \hat{\mathbf{h}}_i(n)\|^2}{\sum_{i=1}^2 \|\mathbf{h}_i\|^2} \quad (11)$$

The WEVN performance of the SAEC system employing the signal decorrelation techniques using random noise addition in [1], non-linear transformation in [3], and those three steps of the proposed technique are illustrated in Fig. 3. As compared to the SAEC system without any decorrelation methods, the WEVN performance of system employing the random noise addition method in [1] with SNR of  $30\text{dB}$  was improved by  $9\text{dB}$ . By considering at the performance when employing the random noise addition with variable gain  $\gamma_i(n)$ , the improvement of  $8\text{dB}$  was obtained. For the non-linear transformation method in [3] when the non-linear factor was set to be  $\alpha = 0.8$  in this paper to maintain the speech quality, the WEVN performance was improved by  $3\text{dB}$ . Similarly to the method in [3], the SAEC system with the non-linear transformation as given in (7) and (8) reached about the same level of WEVN when the parameter  $\beta = 0.6$  was chosen. For the proposed adaptive noise addition, with  $\beta = 0.6$ , the WEVN performance can be improved by  $7\text{dB}$ .

In addition, the magnitude squared coherence, defined as

$$\gamma_{12}(f) = \frac{|P_{x_1 x_2}(f)|^2}{P_{x_1 x_1}(f) P_{x_2 x_2}(f)} \quad (12)$$

was used to indicate the correlation between the stereo signals in each channel at each frequency, i.e.  $\gamma_{12}(f) \in [0, 1]$ , where  $P_{x_1 x_1}(f)$  and  $P_{x_2 x_2}(f)$  were the power spectra of the speech input signals  $x_1(n)$  and  $x_2(n)$ , and the cross-spectral density between  $x_1(n)$  and  $x_2(n)$  was  $P_{x_1 x_2}(f)$ . From Fig. 4, the coherence

function of the stereo speech signals with no decorrelation methods is about unity at all frequency. The technique that gives the best decorrelation level of the stereo signals among these methods was the proposed method employing adaptive noise addition.

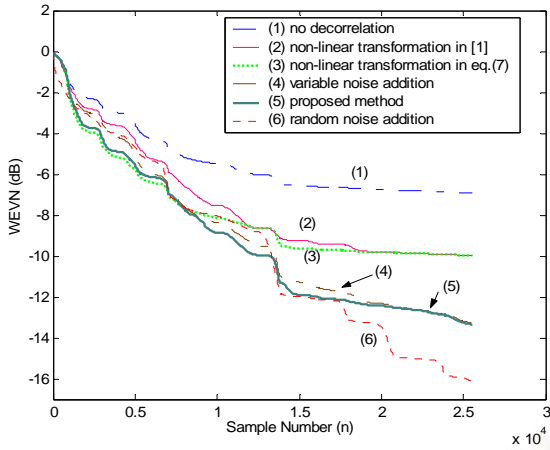


Figure 3: Weight Error Vector Norm Comparison.

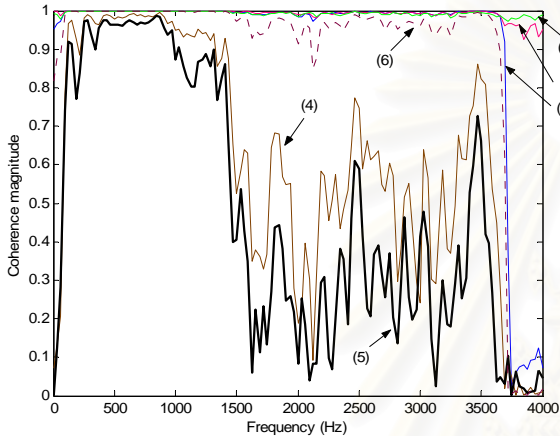


Figure 4: Coherence function of decorrelated stereo signals employing the proposed adaptive noise addition.

4.2. Subjective Listening Test

Based on the ITU recommendation [9], the subjective listening tests were undertaken with 20 listeners. The original stereo signals were compared with the decorrelated stereo signals employing the random noise addition technique in [1], the non-linear transformation in [3], and those from three steps of the proposed method employing adaptive noise addition. The Mean Opinion Score (MOS) value of the original signal was 4.38, whereas those decorrelated stereo signals were graded as given in Table 1.

Table 1: Mean Opinion Score value of the stereo signals employing different decorrelation methods.

Signals	MOS values
No signal decorrelation method	4.38
Non-linear transformation in [3]	3.56
Non-linear transformation in (7)	2.22
Variable noise addition in (6)	2.88
Adaptive noise addition	3.50
Random noise addition in [1]	2.00

The test shows that the signal decorrelation approaches employing the proposed adaptive noise addition and the non-linear transformation approach in [3] preserve much better stereo perception of the signals than the other investigated methods, whereas the technique using random noise addition in [1] gives the lowest MOS value. It can be seen that the signal decorrelation performance and the ability to maintain the quality of the processed stereo signals have to be compromised.

5. CONCLUSIONS

It has been shown in terms of WEVN that stereo signals can be decorrelated efficiently by employing the proposed low-complexity adaptive noise addition technique. Furthermore, the proposed technique results in minimum degradation of the quality of the stereo signals, as compared to the other techniques investigated in this paper.

6. ACKNOWLEDGEMENTS

This work has been supported by Ratchadaphisek Somphot Endowment, Chulalongkorn University, Thailand.

7. REFERENCES

- [1] M. M. Sondhi, D. R. Morgan, and J. L. Hall, "Stereophonic Acoustic Echo Cancellation – An Overview of the Fundamental Problem," *IEEE Signal Processing Letters*, pp. 148-151, August 1995.
- [2] M. M. Sondhi and D. R. Morgan, "Acoustic echo cancellation for stereophonic teleconferencing", *IEEE Workshop on Applications of Signal Processing to Audio Acoustics (ASSP'91, New Paltz, NY)*, pp. 141-142, October 1991.
- [3] J. Benesty, D. R. Morgan, and M. M. Sondhi, "A better understanding and an improved solution to the problems of stereophonic acoustic echo cancellation", *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'97, Munich, Germany)*, vol. 1, pp. 303-306, April 1997.
- [4] J. Benesty, D. R. Morgan and M. M. Sondhi, "A Better Understanding and an Improved Solution to the Specific Problems of Stereophonic Acoustic Echo Cancellation", *IEEE Trans. on Speech and Audio Processing*, vol. 6, no. 2, pp. 156-165, March 1998.
- [5] M. Ali, "Stereophonic acoustic echo cancellation system using time-varying all-pass filtering for signal decorrelation", *Proceedings of ICASSP'98 (Seattle, U.S.A.)*, vol. 6, pp. 3689-3692, May 1998.
- [6] A. Gilloire and V. Turbin, "Using auditory properties to improve the behaviour of stereophonic acoustic echo cancellers", *Proceedings of ICASSP'98 (Seattle, U.S.A.)*, vol. 6, pp. 3681-3684, May 1998.
- [7] Y. W. Jung, J. H. Lee, Y. C. Park and D. H. Youn, "A new adaptive algorithm for stereophonic acoustic echo canceller", *Proceedings of ICASSP'2000 (Istanbul, Turkey)*, vol. 2, pp. 801-804, June 2000.
- [8] E. Zwicker and H. Fastl, "Psychoacoustics : Facts and Models", *Springer-Verlag*, 1990.
- [9] CCITT Recommendation E.432, International Telecommunication Union (ITU), 1992.

## ประวัติผู้เขียนวิทยานิพนธ์

นางสาว พิมพกา สุรินทร์ เกิดวันที่ 1 พฤษภาคม พ.ศ. 2520 ที่จังหวัดน่าน ได้เข้าศึกษาในหลักสูตรวิศวกรรมศาสตรบัณฑิต คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเชียงใหม่ในปีการศึกษา 2538 ซึ่งสำเร็จการศึกษาวิศวกรรมศาสตรบัณฑิต สาขาวิศวกรรมไฟฟ้า ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเชียงใหม่ในปีการศึกษา 2542 และเข้าทำงานที่การไฟฟ้าส่วนภูมิภาค หลังจากนั้นได้เข้าศึกษาต่อในหลักสูตรวิศวกรรมศาสตรมหาบัณฑิต คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัยในปีการศึกษา 2545



สถาบันวิทยบริการ  
จุฬาลงกรณ์มหาวิทยาลัย