

## โครงการวิจัยย่อยลำดับที่ 8

### เรื่อง การปรับปรุงคุณภาพการส่งสัญญาณเสียงบนเครือข่าย X เฟสที่ 3

Achieving high subjective voice quality in Voice over “X”-networks (Phase III)

ผู้รับผิดชอบโครงการ ผู้ช่วยศาสตราจารย์ ดร.นิศาชล ตั้งเสงี่ยมวิสัย

#### บทที่ 1

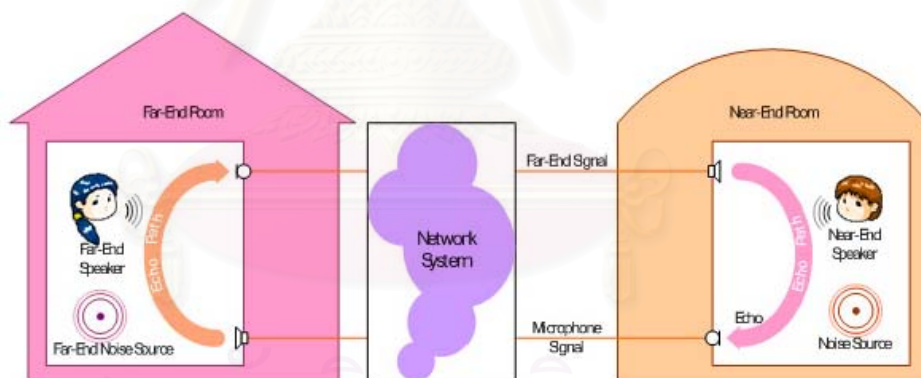
#### บทนำ

##### 1.1 ความสำคัญและที่มาของโครงการวิจัย

คุณภาพของการส่งสัญญาณเสียงในระบบการสื่อสารบนเครือข่ายใด ๆ ขึ้นอยู่กับความชัดเจนของสัญญาณเสียง และความหน่วงเวลา (Delay) เป็นปัจจัยหลัก โดยปัจจัยเรื่องความหน่วงเวลาทำให้เกิดปัญหาสัญญาณเสียงสะท้อน (Echo) ขึ้น ซึ่งแนวทางการตัดหรือลดสัญญาณเสียงสะท้อนประเภท Acoustic Echo (Acoustic Echo Cancellation, AEC) [1] ได้ทำการศึกษาในสองเฟสแรกของโครงการวิจัย “การตัดสัญญาณสะท้อนทางเสียงเมื่อมีสถานการณ์ Double-Talk ในระบบการสนทนาแบบสเตอริโอ (เฟสที่หนึ่งและสอง)” (1 มิถุนายน 2545 – 30 พฤษภาคม 2547) โดยได้ทำการศึกษากการตัดสัญญาณสะท้อนทางเสียงในระบบการสนทนาแบบโมโน (ช่องสัญญาณเดียว) และแบบสเตอริโอ (สองช่องสัญญาณ) (Stereophonic Acoustic Echo Cancellation, SAEC) [1, 2, 8] สำหรับปัจจัยเรื่องความชัดเจนของสัญญาณเสียงนั้น ในเฟสที่สามของโครงการวิจัย “การปรับปรุงคุณภาพการส่งสัญญาณเสียงบนเครือข่าย X (เฟสที่หนึ่ง)” (1 มิถุนายน 2547 – 30 พฤษภาคม 2548) ได้ศึกษาการนำระบบ AEC และ SAEC มาใช้ในการตัดสัญญาณเสียงสะท้อนออกจากระบบการสื่อสารบนเครือข่าย “X” อาทิ เช่น VoIP VoFR VoATM เป็นต้น เพื่อรักษาไว้ซึ่งคุณภาพของสัญญาณเสียงในระบบ และพิจารณาการเข้ารหัสสัญญาณเสียงที่ทำการตัดเสียงสะท้อนแล้ว ด้วยตัวเข้ารหัสสัญญาณเสียง (Speech Coder) เช่น G.729 Codec เป็นต้น หรือตัวเข้ารหัสอื่นที่มีความเหมาะสมต่อคุณสมบัติของเครือข่าย “X” นั้น ๆ และในเฟสที่สี่ของโครงการวิจัย “การปรับปรุงคุณภาพการส่งสัญญาณเสียงบนเครือข่าย X (เฟสที่สอง)” (1 มิถุนายน 2548 – 30 พฤษภาคม 2549) เน้นถึงการศึกษาด้านเทคนิคต่างๆ ที่ใช้ในการตัดสัญญาณรบกวน และนำเทคนิคการตัดสัญญาณรบกวนที่เหมาะสม มาประยุกต์ใช้งานร่วมกับระบบตัดสัญญาณเสียง

สะท้อน โดยเน้นไปที่ระบบ AEC ก่อน เพื่อพัฒนาประสิทธิภาพการตัดสัญญาณเสียงสะท้อนให้เพิ่มขึ้น และทำให้คุณภาพของสัญญาณเสียงที่ส่งออกไปในระบบมีความชัดเจนมากขึ้น ปรากฏจากเสียงสะท้อนที่หลงเหลืออยู่ (Residual Echo) หรือสัญญาณเสียงรบกวนใดๆ ในโครงการวิจัยนี้ “การปรับปรุงคุณภาพการส่งสัญญาณเสียงบนเครือข่าย X (เฟสที่สาม)” (1 มิถุนายน 2549 – 30 พฤษภาคม 2550) ศึกษาเทคนิคการตัดหรือลดผลกระทบของสัญญาณรบกวนและสัญญาณเสียงสะท้อนพร้อมๆ กัน เพื่อเพิ่มคุณภาพสัญญาณเสียงที่จะส่งไปบนเครือข่ายใดๆ ให้ความคมชัดของสัญญาณเสียงพูด

เมื่อพิจารณาระบบการสนทนาแบบแฮนด์ฟรี (Hands-Free Communication Systems) ดังรูปที่ 1 เช่น ระบบการสนทนาทางไกล ระบบโทรศัพท์แฮนด์ฟรีในรถยนต์ และระบบสนทนาผ่านทางอินเทอร์เน็ตแบบแฮนด์ฟรี เป็นต้น ไมโครโฟนและลำโพงจะถูกติดตั้งไว้ ณ จุดหนึ่งในบริเวณที่ทำกรสนทนา เพื่อรับเสียงพูดของผู้พูดทางห้องใกล้ (Near-end Room) หรือทางฝั่งห้องรับ (Receiving Room) และกระจายเสียงพูดของคู่สนทนาหรือผู้พูดทางห้องไกล (Far-end Speech) หรือทางฝั่งห้องส่ง (Transmitting Room) ตามลำดับ จึงทำให้การสนทนาเป็นไปในรูปแบบที่มีความเป็นอิสระ



รูปที่ 1 ระบบการสนทนาแบบแฮนด์ฟรี

ปัญหาสัญญาณเสียงสะท้อน (Acoustic Echo) เกิดขึ้นจากเสียงของผู้พูดทางห้องไกลถูกส่งย้อนกลับไปหาตัวผู้พูดทางห้องไกลเอง ทำให้ผู้พูดทางฝั่งห้องไกลได้ยินเสียงสะท้อนของตนเอง เนื่องมาจากการวางตัวของไมโครโฟนและลำโพงในลักษณะดังรูป นอกจากนี้ สัญญาณเสียงรบกวน (Acoustic Noise) จากภายในบริเวณการสนทนาจะถูกส่งไปยังผู้พูดทางห้องไกลโดยไม่อาจหลีกเลี่ยงได้ ซึ่งปัญหาทั้งสองนี้มีผลต่อคุณภาพของเสียงที่ถูกส่งให้กับผู้พูดทางห้องไกล

กล่าวคือไม่เฉพาะเสียงรบกวนที่ถูกปะปนมาเท่านั้น ยังมีผลจากความสับสนและไม่เป็นธรรมชาติที่เกิดจากการได้ยินเสียงสะท้อนของตนเองอีกด้วย ดังนั้นการเพิ่มคุณภาพของเสียงก่อนถูกส่งไปยังคู่สนทนาทางห้องไกลจึงมีความจำเป็นในระบบสนทนาแบบแฮนด์ฟรี

ในอดีตที่ผ่านมา ปัญหาเสียงรบกวนและเสียงสะท้อนได้รับการแก้ไขอย่างอิสระจากกัน ดังเห็นได้จากตัวอย่างบทความเกี่ยวกับการลดสัญญาณเสียงรบกวน (Noise Suppression, NS) [1]-[20] และตัวอย่างบทความการตัดสัญญาณเสียงสะท้อน (AEC) [21]-[26] อย่างไรก็ตาม เมื่อไม่นานมานี้ ได้มีความพยายามที่จะค้นหาและพัฒนาระเบียบวิธีรวมในแก้ไขปัญหาสัญญาณรบกวน (Disturbance Signal) ทั้งสอง [27]-[34] เพื่อเพิ่มประสิทธิภาพการทำงานที่เหมาะสมให้กับระบบการสนทนาแบบแฮนด์ฟรี

## 1.2 วัตถุประสงค์ของโครงการวิจัย

- 1.2.1 ศึกษา วิเคราะห์ และพัฒนาเทคนิคการตัดหรือลดผลกระทบของสัญญาณรบกวน ภายหลังจาก/หรือรวมกันกับการตัดสัญญาณเสียงสะท้อน เพื่อเพิ่มคุณภาพสัญญาณเสียงที่จะส่งไปบนเครือข่ายใดๆ โดยเน้นทางด้านความคมชัดของสัญญาณเสียงพูด
- 1.2.2 ศึกษาความเป็นไปได้ในการทดสอบระบบที่พัฒนาขึ้นในโครงการนี้บนสถานการณ์จริง (Real-Time) โดยอาศัย FPGA Board ที่ได้พัฒนาแล้วบางส่วนจากเฟสก่อนหน้า

## 1.3 ขอบเขตของโครงการวิจัย

การปรับปรุงคุณภาพของสัญญาณเสียงที่ส่งออกไปบนเครือข่ายการสื่อสารทางเสียงที่มีหนึ่งช่องสัญญาณ โดยมุ่งเน้นไปที่การตัดสัญญาณเสียงสะท้อนด้วยระบบ (Acoustic Echo Cancellation – AEC) และการตัดสัญญาณรบกวนที่ปนเปื้อนอยู่กับสัญญาณเสียงพูดในช่องสัญญาณที่ทำการศึกษา เพื่อเพิ่มคุณภาพสัญญาณเสียงก่อนส่งออกไปในเครือข่าย

## บทที่ 2

### วิธีการดำเนินการวิจัย

แผนงานวิจัยของโครงการนี้ถูกจัดไว้เป็นระยะเวลา 12 เดือน โดยจะถูกดำเนินการดังแสดงในตารางที่ 2-1 และมีการจ้างผู้ช่วยวิจัยจำนวน 4 คนตลอดโครงการวิจัยนี้ ดังรายละเอียดต่อไปนี้

1. นายรัฐพล ทูลแสงงาม (ตลอด 12 เดือน ของโครงการนี้)
2. นายสมภพ จรรยาวิลาส (ในช่วง 3 เดือนแรก ของโครงการนี้)
3. นายธันย์ชนก ตั้งวัชรพงศ์ (ตลอด 12 เดือน ของโครงการนี้)
4. นายปพิชญา ชัยสกุล (ในช่วง 6 เดือนหลังของโครงการนี้)

สถาบันวิทยบริการ  
จุฬาลงกรณ์มหาวิทยาลัย



ตารางที่ 2-1 แผนการดำเนินงานของโครงการนี้ตลอดระยะเวลา 12 เดือน

กิจกรรม	ระยะเวลาปฏิบัติการ					
	เดือน 1-2	เดือน 3-4	เดือน 5-6	เดือน 7-8	เดือน 9-10	เดือน 11-12
2.1 ศึกษาถึงปัญหาของระบบการตัดเสียงสะท้อน (AEC)						
2.2 ศึกษาเทคนิคการตัดเสียงรบกวนแบบต่างๆ เพื่อนำเทคนิคที่เหมาะสมมาใช้ในการตัดเสียงสะท้อน						
2.3 ศึกษาความเป็นไปได้ในการทดสอบระบบที่พัฒนาขึ้นในโครงการนี้บนสถานการณ์จริง (Real-Time) โดยอาศัย FPGA Board ที่ได้พัฒนาแล้วบางส่วนจากเฟสก่อนหน้า						
2.4 จำลองระบบบนคอมพิวเตอร์การตัดเสียงสะท้อนและเสียงรบกวน						
2.5 สรุปผลการทดสอบและเขียนรายงานฉบับสมบูรณ์						

## 2.1 ศึกษาถึงปัญหาของระบบการตัดเสียงสะท้อน (AEC)

ในช่วง 4 เดือนแรกของโครงการ จะทำการศึกษา วิเคราะห์ ถึงปัญหาที่เกิดขึ้นในระบบการตัดเสียงสะท้อน (AEC) เมื่อมีสัญญาณรบกวนรวมอยู่ด้วย โดยสัญญาณรบกวนที่พิจารณาเป็นสัญญาณรบกวนพื้นหลัง (Background Noise) ที่รวมอยู่ในสัญญาณไมโครโฟนของระบบการสนทนาแบบแฮนด์ฟรี เมื่อทำการตัดสัญญาณเสียงสะท้อนทางห้องใกล้ โดยใช้ฐานข้อมูลที่มีอยู่ทั่วไป เช่น เอกสารทางวิชาการ วารสารทางวิชาการ หรือรายงานจาก ห้องสมุด และอินเทอร์เน็ต เป็นต้น

## 2.2 ศึกษาเทคนิคการตัดเสียงรบกวนแบบต่างๆ เพื่อนำเทคนิคที่เหมาะสมมาใช้เพื่อการตัดเสียงสะท้อน

ศึกษาเทคนิคการตัดเสียงรบกวนต่างๆ เพื่อพิจารณาหาเทคนิคที่สามารถนำไปใช้งานเพื่อการตัดเสียงรบกวนพร้อมๆ กับการตัดเสียงสะท้อนในระบบการสนทนาแบบแฮนด์ฟรี โดยพิจารณาถึงความเป็นไปได้ในการใช้งานร่วมกันของระบบ AEC และเทคนิค ANS ซึ่งเป็นเทคนิคหนึ่งที่ใช้ตัดเสียงรบกวน โดยมุ่งเน้นในด้านประสิทธิภาพการตัดสัญญาณรบกวน และการรักษาความบิดเบี้ยวต่อคุณภาพสัญญาณเสียงพูดที่ถูกตัดสัญญาณรบกวนให้ต่ำที่สุด ทั้งนี้เทคนิคหนึ่งทำงานในลักษณะดังกล่าว ที่เรียกว่า Acoustic Echo Suppression (AES) ก็ถูกนำมาวิเคราะห์ และศึกษาประสิทธิภาพการทำงาน เพื่อเปรียบเทียบกับระบบ AEC และระบบร่วมของ AEC และ AES ด้วย

## 2.3 ศึกษาความเป็นไปได้ในการทดสอบระบบที่พัฒนาขึ้นในโครงการนี้บนสถานการณ์จริง (Real-Time) โดยอาศัยบอร์ด FPGA ที่ได้พัฒนาแล้วบางส่วนจากเฟสก่อนหน้า

ทดสอบการใช้งานของระบบ AEC ที่ทำงานร่วมกับเทคนิค ANS ด้วยบอร์ด FPGA เพื่อประเมินผลประสิทธิภาพการตัดเสียงสะท้อนในเวลาจริง

## 2.4 จำลองระบบบนคอมพิวเตอร์การตัดเสียงสะท้อนและเสียงรบกวน

พิจารณาคัดเลือกเทคนิคการตัดสัญญาณเสียงรบกวนที่เหมาะสม ทำการจำลองระบบการตัดสัญญาณเสียงรบกวนเพียงลำพัง กับการจำลองระบบร่วมของการตัดสัญญาณเสียงสะท้อนและเสียงรบกวนบนคอมพิวเตอร์ ผ่านทางโปรแกรม MATLAB

สำหรับสัญญาณเสียงที่ใช้ในการจำลองระบบบนคอมพิวเตอร์ เป็นสัญญาณเสียงพูดที่ทำการบันทึกเสียงเอาไว้ในฐานข้อมูลในเฟสก่อนหน้านี โดยทำการบันทึกเสียงในช่องที่ไม่มีสัญญาณเสียงรบกวน ที่ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ส่วนการเติมสัญญาณเสียงรบกวนให้กับสัญญาณเสียงเข้าของระบบ เพื่อจำลองสภาวะแวดล้อมต่างๆ ในบริเวณโดยรอบของผู้ใช้ให้สมจริงนั้น ใช้การเพิ่มสัญญาณรบกวนขาว (Additive White Gaussian Noise) ในการจำลองระบบบนคอมพิวเตอร์ นอกจากนี้ อาจมีการบันทึกเสียงเพิ่มเติมสำหรับสัญญาณเสียงพูดในสภาวะแวดล้อมที่จัดให้มีสัญญาณเสียงรบกวนด้วย

## 2.5 สรุปผลการทดสอบและเขียนรายงานฉบับสมบูรณ์

สรุปแนวทางในการปรับปรุงระบบที่น่าเสนอในโครงการวิจัยนี้ เพื่อเพิ่มประสิทธิภาพการทำงาน และเขียนรายงานฉบับสมบูรณ์ของโครงการวิจัย

สถาบันวิทยบริการ  
จุฬาลงกรณ์มหาวิทยาลัย

## บทที่ 3

### ผลการดำเนินการ

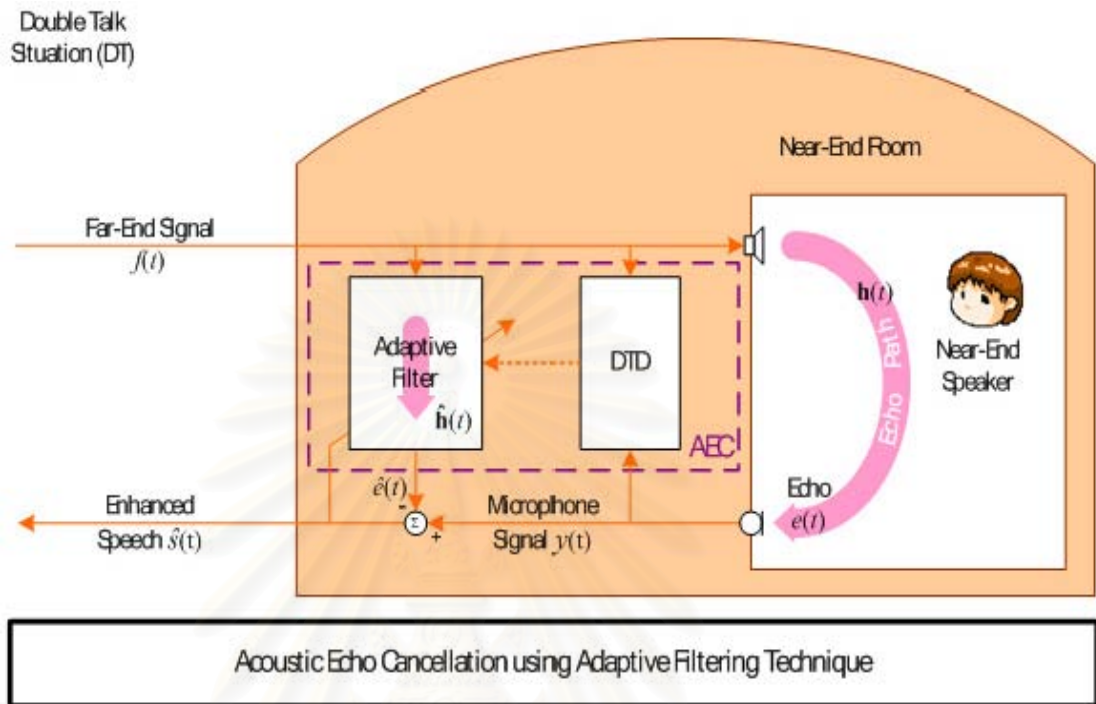
การศึกษาประสิทธิภาพการลดเสียงสะท้อนและเสียงรบกวนในโครงการนี้แบ่งออกเป็น 4 ส่วน คือ ส่วนแรกจะกล่าวถึงการลดเสียงสะท้อน (Acoustic Echo Reduction, AER) ซึ่งมี 2 แนวทางหลัก คือ การตัดเสียงสะท้อน (Acoustic Echo Cancellation, AEC) และการลดเสียงสะท้อนโดยการกดเสียงสะท้อน (Acoustic Echo Suppression, AES)

จากนั้น ส่วนที่สองจะนำเสนอเทคนิคการลดเสียงรบกวน (Noise Reduction Techniques, NR) เช่น การตัดสัญญาณรบกวนแบบปรับตัว (Adaptive Noise Cancellation, ANC) การลบออกของสเปกตรัม (Spectral Subtraction, SS) และการกดสัญญาณรบกวน (Noise Suppression, NS) เป็นต้น เพื่อนำไปใช้ในส่วนที่สามคือ การลดเสียงสะท้อนและเสียงรบกวนพร้อมๆ กัน (Acoustic Echo and Noise Reduction, AENR) ตามด้วยการพัฒนาการกดทั้งเสียงสะท้อนและเสียงรบกวน (AENS) ในส่วนสุดท้ายเป็นการแสดงผลการจำลองระบบบนคอมพิวเตอร์

### 3.1 การลดเสียงสะท้อน (AER)

แนวทางหลัก 2 แนวทางของการลดเสียงสะท้อน คือ การตัดเสียงสะท้อน (AEC) ที่ใช้วงจรกรองแบบปรับตัว (Adaptive Filter) ในการประมาณสัญญาณเสียงสะท้อน และนำไปลบออกจากสัญญาณไมโครโฟนของระบบการสนทนาแบบแฮนด์ฟรี วิธี AEC นี้ไม่ทำให้เกิดความผิดเพี้ยนของสัญญาณเสียงพูดที่ถูกรับปรุง อีกแนวทางหนึ่ง คือ การกดเสียงสะท้อน (AES) ซึ่งมีความซับซ้อนในการคำนวณของระบบที่ต่ำกว่าระบบ AEC อยู่มาก แต่เนื่องจากระบบ AES ถูกพัฒนามาจากหลักการของการลดเสียงรบกวนแบบการลดทอนสเปกตรัมในช่วงเวลาสั้นๆ (ดังจะได้กล่าวต่อไป) ทำให้สัญญาณเสียงพูดที่ถูกรับปรุงด้วยระบบนี้ เกิดความผิดเพี้ยนของสัญญาณเสียงพูด

### 3.1.1 การตัดเสียงสะท้อน (AEC)



รูปที่ 2 การตัดเสียงสะท้อน (AEC) ในห้องใกล้

เมื่อพิจารณาทางฝั่งห้องใกล้ สัญญาณเสียงสะท้อนแบบ Acoustic Echo เกิดขึ้นจากการเชื่อมต่อทางเสียง (Acoustic Coupling) ระหว่างลำโพงกับไมโครโฟนภายในห้องใกล้ ดังแสดงในรูปที่ 2 ทั้งนี้ การตัดสัญญาณเสียงสะท้อนที่เกิดขึ้นภายในห้องใกล้นี้ ระบบ AEC ใช้วงจรกรองแบบปรับตัวในการจำลองวิถีสะท้อนทางเสียง (Acoustic Echo Path) และหักออกจากสัญญาณไมโครโฟน ก่อนที่จะถูกส่งกลับไปยังคู่สนทนาในห้องใกล้ สำหรับการตัดสัญญาณเสียงสะท้อนที่เกิดขึ้นภายในห้องใกล้ ก็สามารถใช้ระบบ AEC อีกชุดหนึ่งทางห้องใกล้ เพื่อทำการจำลองวิถีสะท้อนทางเสียงทางฝั่งห้องใกล้ได้เช่นกัน

เมื่อสัญญาณขาเข้าของระบบ AEC เป็นสัญญาณเสียงพูดจากห้องไกล  $f(t)$  สัญญาณไมโครโฟนที่ห้องใกล้  $y(t)$  ประกอบด้วยเสียงสะท้อนที่เกิดจากวิถีสะท้อนทางเสียงในห้องใกล้  $h(t)$  ซึ่งถูกจำลองขึ้นจากการคอนโวลูชันสัญญาณขาเข้า  $f(t)$  กับวิถีสะท้อนทางเสียง  $h(t)$  ดังสมการ

$$y(t) = \mathbf{h}^T(t)\mathbf{f}(t) \quad (1)$$

โดยที่  $\mathbf{h}(t) = [h_1(t), h_2(t), \dots, h_L(t)]^T$  เป็นเวกเตอร์สัมประสิทธิ์ของวิถีสะท้อนทางเสียงในห้องโถงที่มีขนาด  $L$  และ  $\mathbf{f}(t) = [f(t), f(t-1), \dots, f(t-L+1)]^T$  เป็นเวกเตอร์สัญญาณเสียงขาเข้าที่มีขนาด  $L$  เช่นกัน โดยทั่วไปแล้ว จะกำหนดให้สัญญาณไมโครโฟน  $y(t)$  มีสัญญาณรบกวนพื้นหลังภายในห้องโถงรวมอยู่ด้วย เพื่อความสมจริงมากขึ้น สำหรับสัญญาณรบกวนพื้นหลังนี้ สามารถจำลองได้โดยใช้สัญญาณรบกวนขาว

สำหรับสัญญาณผิดพลาด  $e_y(t)$  จากการประมาณสัญญาณเสียงสะท้อนของระบบ AEC หาได้จากผลต่างของสัญญาณเสียงสะท้อนในห้องโถงและค่าประมาณของสัญญาณเสียงสะท้อนที่เป็นสัญญาณออกของวงจรรองแบบปรับตัว  $\hat{\mathbf{h}}(t)$  ที่ทำการจำลองวิถีสะท้อนทางเสียงในห้องโถงและมีจำนวนสัมประสิทธิ์เท่ากับ  $L$  เช่นกัน

$$e_y(t) = y(t) - \hat{\mathbf{h}}^T(t)\mathbf{f}(t) \quad (2)$$

โดยสัญญาณผิดพลาด  $e_y(t)$  เป็นสัญญาณขาออกของระบบ AEC และจะถูกส่งต่อไปยังห้องโถง ดังนั้นเมื่อวงจรรองแบบปรับตัว  $\hat{\mathbf{h}}(t)$  สามารถจำลองวิถีสะท้อนทางเสียงได้อย่างใกล้เคียง เสียงสะท้อนก็สามารถถูกตัดไปได้ในที่สุด

การปรับปรุงค่าสัมประสิทธิ์ของวงจรรองแบบปรับตัวในโครงการวิจัยนี้จะใช้ระเบียบวิธี Normalised Least Mean Square (NLMS) โดยมีสมการปรับให้ทันกาล (Update Equation) ดังสมการที่ (3)

$$\hat{\mathbf{h}}(t+1) = \hat{\mathbf{h}}(t) + \mu \frac{\mathbf{f}(t)e_y(t)}{\varepsilon + \|\mathbf{f}(t)\|^2} \quad (3)$$



เมื่อ  $\square$  คือค่าช่วงก้าว (Step-size) ของระเบียบวิธี NLMS ที่ใช้ในการปรับการทำงานของวงจรรองแบบปรับตัว และ  $\epsilon$  เป็นค่าคงที่เล็กๆ เพื่อป้องกันการลู่ออกของวงจรรองแบบปรับตัว เมื่อขนาดของสัญญาณเสียงเข้ามีขนาดเล็กมากในช่วงหนึ่งๆ

สำหรับปัจจัยที่ส่งผลต่อการทำงานของวงจรรองแบบปรับตัวในระบบ AEC ปัจจัยแรกคือจำนวนสัมประสิทธิ์ของวงจรรองแบบปรับตัว โดยหากเลือกใช้จำนวนสัมประสิทธิ์น้อยเกินไป จะทำให้ไม่สามารถประมาณวิถีสะท้อนทางเสียงได้อย่างแม่นยำ แต่หากเลือกจำนวนสัมประสิทธิ์ของวงจรรองยาวมากเกินไป ก็จะทำให้มีผลของส่วนหางของวงจรรองแบบปรับตัว (Tail Effect) ทำให้สัญญาณเสียงสะท้อนที่ประมาณได้มีค่าผิดเพี้ยนไปจากที่ควรจะเป็น ในทางปฏิบัตินั้นการเลือกใช้จำนวนสัมประสิทธิ์ของวงจรรองแบบปรับตัวจะเลือกให้น้อยที่สุดเท่าที่จะเป็นไปได้ เพราะจำนวนสัมประสิทธิ์ของวงจรรองแบบปรับตัวแปรผันตรงกับความซับซ้อนในการคำนวณของระบบ AEC

ปัจจัยที่สองที่มีผลต่อการทำงานของวงจรรองแบบปรับตัวในระบบ AEC โดยทำให้ระบบ AEC ทำงานได้ไม่เต็มประสิทธิภาพคือ การที่สัญญาณไมโครโฟน  $y(t)$  มีสัญญาณอื่นๆ นอกเหนือจากสัญญาณเสียงสะท้อน  $e(t)$  ปะปนอยู่ด้วย เช่น กรณีที่คู่สนทนาทั้งสองพูดพร้อมกัน หรือที่เรียกว่า สถานการณ์ดับเบิลทอล์ก (Double-Talk Situation, DTS) โดยแตกต่างจากสถานการณ์ปกติที่พิจารณาในระบบ AEC คือ คู่สนทนาพูดทีละฝ่าย ซึ่งเรียกว่าสถานการณ์ซิงเกิ้ลทอล์ก (Single-Talk Situation, STS) ทั้งนี้เนื่องจากสัญญาณเสียงดับเบิลทอล์กหรือสัญญาณเสียงของผู้พูดในห้องใกล้มีพลังงานที่สูงมากเมื่อเทียบกับสัญญาณเสียงสะท้อนที่ไมโครโฟนทางห้องใกล้ จนอาจทำให้วงจรรองแบบปรับตัวเกิดลู่ออกจากสถานะอยู่ตัว (Steady State) ได้ ดังนั้นการแก้ไขปัญหา DTS จึงทำได้โดยการสั่งให้วงจรรองแบบปรับตัวทำการคงค่าวิถีสะท้อนทางเสียงเดิมที่ประมาณเอาไว้เมื่อก่อนเกิดสถานการณ์ดับเบิลทอล์กขึ้น เนื่องจากสถานการณ์ดับเบิลทอล์กจะเกิดขึ้นช่วงสั้นๆ โดยอาศัยการทำงานของตัวตรวจหาสถานการณ์ดับเบิลทอล์ก (Double-Talk Detector, DTD) เพื่อระบุว่าช่วงเวลาใดเกิดสถานการณ์ดับเบิลทอล์ก หลังจากนั้นจึงสามารถทำการปรับค่าสัมประสิทธิ์ของวงจรรองแบบปรับตัวเสียใหม่เพื่อติดตามการเปลี่ยนแปลงของวิถีสะท้อนทางเสียงที่อาจเกิดขึ้น

นอกจากนี้แล้ว สัญญาณอื่นๆ ที่อาจปะปนมากับสัญญาณไมโครโฟน  $y(t)$  ได้แก่ สัญญาณเสียงรบกวนภายในห้องใกล้นั่นเอง โดยสัญญาณเสียงรบกวนนี้มีคุณลักษณะทางสถิติที่แตกต่างจากสัญญาณดับเบิลทอล์ก คือจะเกิดขึ้นตลอดเวลา ทำให้ไม่สามารถหาช่วงที่ไม่มี

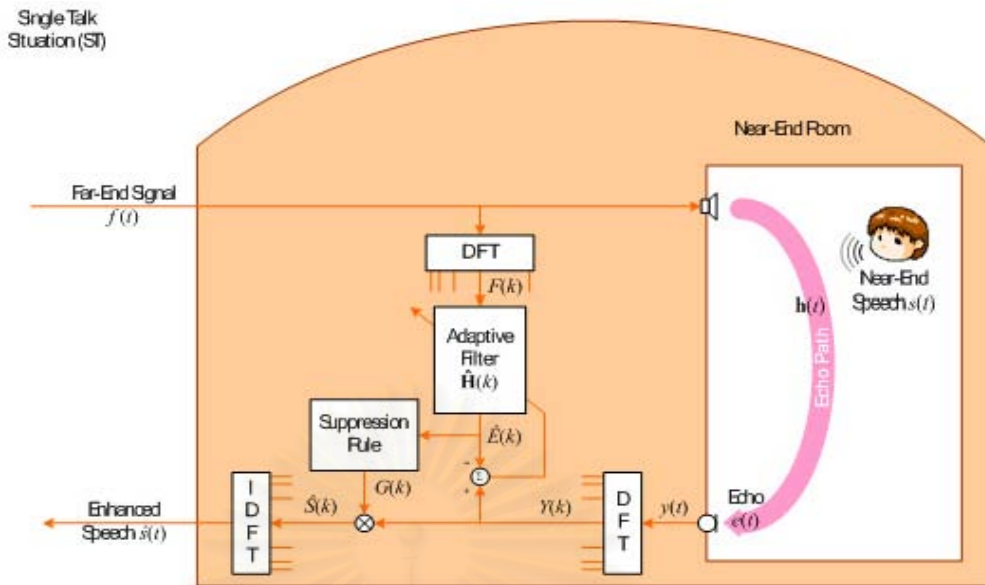
สัญญาณเสียงรบกวนได้ อย่างไรก็ตามพลังงานของสัญญาณเสียงรบกวนจะมีค่าต่ำกว่าพลังงานของสัญญาณดับเบิลทอล์ก ทั้งนี้ระเบียบวิธีที่ใช้ในการปรับตัวของวงจรกรองในระบบ AEC ที่ทนทานต่อผลของสัญญาณเสียงรบกวน จึงเป็นสิ่งที่มีความสำคัญต่อการใช้งานในสถานการณ์จริงเป็นอย่างยิ่ง นอกจากนี้ปัญหาความสัมพันธ์แบบไม่เป็นเชิงเส้นของสัญญาณเสียงสะท้อน  $e(t)$  กับสัญญาณทางฝั่งห้องไกล  $f(t)$  จะทำให้วงจรกรองแบบปรับตัวไม่สามารถตัดสัญญาณเสียงสะท้อนส่วนนี้ได้ และทำให้เกิดสัญญาณเสียงสะท้อนตกค้างขึ้นด้วยเช่นกัน

อย่างไรก็ดี จากการศึกษาพบว่า ประสิทธิภาพการตัดเสียงสะท้อนของระบบ AEC จะลดลงมากเมื่ออยู่ในสภาพแวดล้อมที่มีสัญญาณรบกวน จึงมีความพยายามที่จะทำให้เกิดการใช้งานร่วมกันของระบบ AEC กับเทคนิคที่ใช้ในการตัดสัญญาณเสียงรบกวน ดังรายละเอียดต่อไป

### 3.1.2 การกีดเสียงสะท้อน (AES)

การกีดเสียงสะท้อน ดังรูปที่ 3 ถูกพัฒนาขึ้นโดยใช้แนวคิดเดียวกับการลดทอนสเปกตรัม (Spectral Subtraction, SS) ในช่วงเวลาสั้นๆ ที่ใช้ในการลดเสียงรบกวน (โปรดดูรายละเอียดในหัวข้อ 3.2 การลดเสียงรบกวน) กล่าวคือ ในขณะที่ระบบ AEC พยายามตัดหรือหักลบสัญญาณเสียงสะท้อนในระดับของรูปคลื่น (Waveform) ระบบ AES กลับทำการกีดสัญญาณเสียงสะท้อนในโดเมน STDFT โดยใช้กฎการลดทอน (Suppression Rule) แทน และด้วยรูปแบบการทำงานในโดเมน STDFT นี้เอง ทำให้ AES สามารถลดสัญญาณเสียงสะท้อนลงได้โดยไม่จำเป็นต้องทำการประมาณวิถีสะท้อนทางเสียงอย่างแม่นยำ เป็นเหตุให้ระบบ AES มีความทนทานต่อการเปลี่ยนแปลงวิถีสะท้อนทางเสียง (Echo Path Change) แต่ทั้งนี้แลกมาด้วยความผิดเพี้ยนของสัญญาณเสียงพูดทางฝั่งห้องใกล้ที่ถูกปรับปรุง

สถาบันวิทยบริการ  
จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 3 การกดเสียงสะท้อน, AES

ระบบ AES ทำการประมาณสัญญาณเสียงพูดในลักษณะเดียวกับการกดเสียงรบกวน (NS) (โปรดดูรายละเอียดของวิธี NS ในหัวข้อ 3.2 การกดเสียงรบกวน) กล่าวคือ ในแต่ละองค์ประกอบทางความถี่ สัญญาณเสียงพูดที่ถูกปรับปรุงจะเกิดจากการคูณสัญญาณไมโครโฟน  $Y(k)$  ด้วยกฎการลดทอน  $G(k)$  ก่อนที่จะนำเฟสของสัญญาณเสียงพูดที่ถูกรบกวนเดิมมารวมเข้ากลายเป็นสัญญาณเสียงพูดที่ถูกปรับปรุง ทั้งนี้สัญญาณเสียงรบกวนในสถานการณ์ของระบบ AES ได้แก่ สัญญาณเสียงสะท้อน

จะเห็นว่ากฎการลดทอนแบบหักลบทางความถี่ที่ต้องการค่ากำลังของสัญญาณเสียงสะท้อนขณะนั้น การประมาณค่ากำลังของสัญญาณเสียงสะท้อนขณะนั้นใน [35] ใช้วงจรกรองแบบปรับตัวที่มีสัมประสิทธิ์เป็นจำนวนเชิงซ้อน โดยทำการประมาณสเปกตรัมของสัญญาณเสียงสะท้อน  $\hat{E}(k)$  ดังนี้

$$\hat{E}(k, n) = F(k, n) * \hat{H}_c(k, n) \tag{4}$$

$$\varepsilon(k, n) \approx |\hat{E}(k, n)|^2 \tag{5}$$

เมื่อ  $\hat{E}(k, n)$  คือ ค่าเชิงซ้อนของสัญญาณเสียงสะท้อนขณะนั้น ณ องค์ประกอบทางความถี่ที่  $k$

$\langle * \rangle$  คือ สัญลักษณ์แทนการคอนโวลูชัน

$\hat{H}_c(k, n)$ <sup>1</sup> คือ วงจรกรองแบบปรับตัวที่มีสัมประสิทธิ์เป็นจำนวนเชิงซ้อน ณ องค์ประกอบทางความถี่ที่  $k$

การประมาณค่าเชิงซ้อนของสัญญาณเสียงสะท้อนขณะนั้น ทำให้ AES ตาม [35] ถูกเรียกว่า Short-Time Spectral-based Echo Suppression (STSES) และเนื่องจากจำนวนของสัมประสิทธิ์ของวงจรกรองแบบปรับตัวในแต่ละองค์ประกอบทางความถี่มีจำนวนน้อยลง ดังนั้นจึงเลือกใช้ระเบียบวิธี (Recursive Least Square, RLS) ในการปรับตัวของวงจรกรองเพื่อให้ระบบมีอัตราการลู่เข้าที่รวดเร็ว

อย่างไรก็ตามเนื่องจากกฎการลดทอนแบบหักลบทางความถี่มีการแกว่งตัวสูงทำให้ STSES ได้รับผลกระทบจากเสียงรบกวนตกค้างแบบเสียงดนตรี (Musical Noise) ดังนั้นการกีดเสียงสะท้อนตามการรับรู้ (Perceptual Acoustic Echo Suppression, PAES) [36] จึงถูกเสนอขึ้น PAES ทำการกีดสัญญาณเสียงสะท้อนบนโดเมนของสเปกตรัม (Spectral Envelop) ซึ่งโดเมนของสเปกตรัมดังกล่าวคือโดเมนที่พยายามเลียนแบบความละเอียดทางความถี่ (Frequency Resolution) ของการรับรู้ของมนุษย์ ดังนั้นจึงอาจเรียกว่า โดเมนของสเปกตรัมทางโสต (Auditory Spectral Envelop) เมื่อย้ายมาพิจารณาในโดเมนดังกล่าว ค่ากำลังของสัญญาณเสียงสะท้อนขณะนั้น ณ ของสเปกตรัมที่  $\tilde{k}$  จะถูกประมาณขึ้นจากวงจรกรองแบบปรับตัวที่มีสัมประสิทธิ์เป็นจำนวนจริง ดังนี้

$$\tilde{\varepsilon}(\tilde{k}, n) = \tilde{\varphi}(\tilde{k}, n) * \hat{H}_r(\tilde{k}, n) \quad (6)$$

เมื่อ  $\tilde{\varphi}(\tilde{k}, n)$  คือ ค่ากำลังของสัญญาณจากทางห้องไกล ณ ของสเปกตรัมที่

$\tilde{k}$

<sup>1</sup> ความสัมพันธ์ระหว่างวงจรกรองแบบปรับตัวที่มีสัมประสิทธิ์เป็นจำนวนเชิงซ้อน และวิธีสะท้อนทางเสียงจริงถูกบรรยายไว้ใน [35]

<sup>2</sup>  $\tilde{k}$  ถูกใช้แทนองค์ประกอบของโดเมนของสเปกตรัม ในขณะที่  $k$  ใช้แทนองค์ประกอบของโดเมนความถี่

$\hat{H}_R(\tilde{k}, n)$  คือ ค่าจริงของสัมประสิทธิ์วงจรรองแบบปรับตัว ณ ของสเปกตรัมที่  $\tilde{k}$

PAES ได้รับผลกระทบจากเสียงรบกวนตกค้างแบบเสียงดนตรีน้อยกว่า STSES ทั้งนี้เนื่องจากบนโดเมนของสเปกตรัม กฎการลดทอนมีการแกว่งตัวที่ค่อนข้างเรียบ [36] และพร้อมกันกับที่ PAES ยังคงรักษาข้อดีของ STSES ไว้ คือความทนต่อการเปลี่ยนแปลงวิธีสะท้อนทางเสียง ความซับซ้อนในการคำนวณ (โดยพิจารณาจากการคูณจำนวนจริงต่อหนึ่งรอบการชักตัวอย่างสัญญาณเข้าที่เรียกว่า Real Multiplication per Input Sample, RMPs) ของ PAES ก็ยังลดลงเป็นอย่างมากเมื่อเปรียบเทียบกับระบบ AEC ดังแสดงในตารางที่ 1 ทำให้ระบบ AES เหมาะสำหรับงานประยุกต์ที่ไม่สามารถใช้ความซับซ้อนที่สูงได้ เช่น ในเครื่องคอมพิวเตอร์ส่วนบุคคล เป็นต้น แต่อย่างไรก็ตาม ความผิดเพี้ยนของสัญญาณเสียงพูดที่ถูกปรับปรุงซึ่งสังเกตได้ในช่วงสถานการณ์ดับเบิ้ลทอล์กก็มีมากด้วยเช่นกัน

ตารางที่ 1 ความซับซ้อนในการคำนวณของแต่ละระเบียบวิธีของ AES

ระเบียบวิธี	ความซับซ้อนในการคำนวณ (RMPs)	ความซับซ้อนในการคำนวณเมื่อสัมประสิทธิ์ $L$ เป็น 1024 ตัว (RMPs)
NLMSbased STSES	STFT + 4WP/N	120
FRLSbased STSES	STFT + 64WP/N	1080
NLMSbased PAES	STFT + 2IP/N + 4W/N	57
STFT	$3W(1 + \log_2 W)/N$	48

W : SIFT Window Size      N : SIFT Window hop size      I : #Spectral Envelop

### 3.2 การลดเสียงรบกวน (NR)

ปัญหาเสียงรบกวนสามารถแก้ไขได้หลายแนวทาง ซึ่งสามารถแบ่งออกได้เป็น 2 แนวทางหลักได้แก่

1. การลดเสียงรบกวนโดยใช้ไมโครโฟนหนึ่งตัว [1]-[17] เทคนิคที่ใช้ไมโครโฟนหนึ่งตัวได้แก่ การลดทอนสเปกตรัมในช่วงเวลาสั้นๆ (Short-Time Spectral Attenuation) [1]-[14], การ



ลดในปริภูมิย่อย (Signal Subspace Method) [15] และ การอาศัยวงจรรองแบบปรับตัวในการ  
ทำนายสัญญาณเสียงพูดและสัญญาณเสียงรบกวน (Adaptive Line Enhancement) [16]-[17]  
เป็นต้น โดยการลดทอนสเปกตรัมในช่วงเวลาสั้นๆ และการลดในปริภูมิย่อย เป็นการกีดเสียง  
รบกวน (Noise Suppression, NS) และ การอาศัยวงจรรองแบบปรับตัวในการทำนาย  
สัญญาณเสียงพูดและสัญญาณเสียงรบกวน เป็นการตัดเสียงรบกวนแบบปรับตัวได้  
(Adaptive Noise Cancellation, ANC)

2. การลดเสียงรบกวนโดยใช้ไมโครโฟนหลายตัว [1], [9], [18]-[20] ได้แก่ การลดทอน  
สเปกตรัมในช่วงเวลาสั้นๆ [9], การประมาณสัญญาณรบกวนจากสัญญาณรบกวนอ้างอิง ซึ่งก็คือ  
การตัดเสียงรบกวนแบบปรับตัวที่ใช้ไมโครโฟนอย่างน้อย 2 ตัว โดยสัญญาณจากไมโครโฟนที่  
เพิ่มเข้ามาจะทำหน้าที่เป็นสัญญาณรบกวนอ้างอิง [18], การเลือกทิศทาง (Beamforming) [19]  
และการแยกแยะแหล่งกำเนิดโดยปราศจากวิสัย (Blind Source Separation) [20] เป็นต้น แนว  
ทางการใช้ไมโครโฟนมากกว่าหนึ่งตัวนี้ทำให้สามารถรับข่าวสารที่สำคัญมากขึ้นทั้งในโดเมนปริภูมิ  
(Spatial Domain) และ โดเมนเวลา (Time Domain) ส่งผลให้คุณภาพของสัญญาณเสียงที่ได้  
สูงขึ้น

อย่างไรก็ตาม อุปกรณ์สื่อสารทางเสียงโดยทั่วไปประกอบด้วยไมโครโฟนเพียงหนึ่งตัว จึง  
ทำให้แนวทางการลดเสียงรบกวนโดยใช้ไมโครโฟนหนึ่งตัวยังคงเป็นแนวทางที่ได้รับความสนใจ  
จากนักวิจัยอยู่เช่นกัน [32], [37] โครงการวิจัยนี้จึงพิจารณาแนวทางการลดเสียงรบกวนโดยใช้  
ไมโครโฟนหนึ่งตัว เนื่องจากมีความซับซ้อนต่ำและมีความเหมาะสมสำหรับสภาพแวดล้อมที่ค่า  
อัตราส่วนสัญญาณต่อสัญญาณรบกวน (Signal-to-Noise Ratio, SNR) ที่ไม่ต่ำมาก คือประมาณ  
30 dB ถึง -5 dB

หลักการทํางานของวิธี NR คือการหาค่าสัมประสิทธิ์ของวงจรรองที่เหมาะสมที่สุดในการ  
กรองสัญญาณเสียงรบกวนออกจากสัญญาณเสียงพูดที่ถูกรบกวน (Noisy Speech Signal) ทำให้  
สัญญาณเสียงออกของวงจรรองมีปริมาณสัญญาณเสียงรบกวนที่ต่ำลง อย่างไรก็ตาม การลด  
สัญญาณเสียงรบกวนเหล่านี้ จะทำให้เกิดความผิดเพี้ยนของสัญญาณเสียงพูดขึ้น (Speech  
Distortion) [1] ดังนั้น วิธี NR จำนวนมากจึงพยายามที่จะลดสัญญาณเสียงรบกวนลงให้ได้มาก  
ที่สุด และควบคุมความผิดเพี้ยนของสัญญาณเสียงพูดออกให้น้อยที่สุด [4]-[5], [15] โดยวิธี NR ที่  
ได้รับความนิยมมากเนื่องจากสอดคล้องกับข้อถ่วงดุล (Tradeoff) ทั้งด้านประสิทธิภาพการลด  
สัญญาณเสียงรบกวนและความผิดเพี้ยนของสัญญาณเสียงพูด รวมทั้งมีความซับซ้อนในการ  
คำนวณ (Computational Complexity) ที่ต่ำ ได้แก่



### 3.2.1 การลดเสียงรบกวนด้วยการลดทอนสเปกตรัมในช่วงเวลาสั้น ๆ (Short-Time Spectral Amplitude, STSA) หรือ การกดเสียงรบกวน (Noise Suppression, NS)

ขนาดของสเปกตรัมของสัญญาณเสียงพูดที่ไร้เสียงรบกวน (Clean Speech Signal) จะถูกประมาณขึ้นจากขนาดของสเปกตรัมของสัญญาณเสียงพูดที่ถูกรบกวน (Noisy Speech Signal) เมื่อทำการการแปลงฟูรีเยร์แบบไม่ต่อเนื่องในช่วงเวลาสั้น ๆ (Short-Time Discrete Fourier Transform, STDFT) ทั้งนี้เมื่อนำมารวมกับเฟสในช่วงสั้น ๆ (Short-Time Phase) ของสัญญาณเสียงพูดที่ถูกรบกวน จะสามารถประกอบคืน (Reconstruct) เป็นสัญญาณเสียงที่ถูกรบกวน (Enhanced Signal) ด้วยเหตุผลที่ว่า หูมนุษย์รับรู้ผลของการเปลี่ยนแปลงข้อมูลทางเฟสของสัญญาณเสียงได้ยากกว่าการเปลี่ยนแปลงข้อมูลทางขนาด [7]

การที่วิธีนี้ทำการประมวลผลสัญญาณในโดเมนความถี่ แทนการประมวลผลสัญญาณในโดเมนเวลา เนื่องจากความต้องการพิจารณาและประมวลผลในแต่ละองค์ประกอบ (Component หรือ Frequency Bin ในกรณีของ STDFT) ได้อย่างอิสระจากกัน (Independent) หรืออย่างไม่มีสหสัมพันธ์กัน (Uncorrelated) กล่าวคือสามารถปรับปรุงคุณภาพของสัญญาณเสียงในแต่ละองค์ประกอบได้โดยที่พิจารณาเพียงแค่องค์ประกอบนั้นๆ และเมื่อปรับปรุงแล้วก็จะมีผลต่อองค์ประกอบนั้นๆ เพียงอย่างเดียวไม่มีผลกระทบต่อองค์ประกอบอื่น ๆ

การแยกองค์ประกอบ (Decomposition) ที่ไม่มีสหสัมพันธ์กัน (Uncorrelated) สามารถทำได้โดยใช้การแปลงคาร์ฮูเนน-เลอว์ฟ (Karhunen-Loève Transform, KLT) [15] หากแต่ว่าการแปลง KLT สามารถใช้การแปลง STDFT แทนได้โดยมีข้อแม้ว่าขนาดหน้าต่าง (Window) ที่ใช้ในการแปลง STDFT จะต้องมีความเป็นอนันต์ [15], [9] แต่ในทางปฏิบัติ เนื่องจากสัญญาณเสียงพูดเป็นสัญญาณที่มีความเป็นจุดนิ่ง (Stationary) เพียงช่วงสั้นๆ ประมาณ 20 – 40 มิลลิวินาที หรือเทียบเท่ากับ 160 – 320 ตัวอย่าง ที่ความถี่ซีกตัวอย่าง (Sampling Frequency) 8 kHz [15] ดังนั้นการจะใช้หน้าต่างขนาดเป็นอนันต์หรือขนาดใหญ่มากจึงเป็นเรื่องที่เป็นไปไม่ได้ การแปลง STDFT จึงเลือกใช้ขนาดหน้าต่างที่ยังคงรักษาความเป็นจุดนิ่งของสัญญาณเสียงพูดเอาไว้คือประมาณ 20-30 มิลลิวินาที การรักษาความเป็นจุดนิ่งของสัญญาณเสียงพูดนั้นจะทำให้สามารถหาค่าสัมประสิทธิ์ของวงจรรองที่เหมาะสมที่สุดได้

การประมวลผลสัญญาณเสียงในโดเมน STDFT ซึ่งถูกมองว่าสัญญาณเสียงแต่ละองค์ประกอบทางความถี่มีความเป็นอิสระจากกัน ขนาดของสัญญาณเสียงพูดก็จะถูกประมาณขึ้นสำหรับแต่ละองค์ประกอบทางความถี่ แบบจำลองทางสถิติ (Statistical Model) ของสัญญาณเสียงพูดในโดเมน STDFT และวิธีการประมาณขนาดของสัญญาณเสียงพูดที่ต่างกัน

นำมาซึ่งกฎการลดทอน (Suppression Rule) หรือ ค่าอัตราขยาย (Gain) ที่แตกต่างกันเป็นจำนวนมาก เช่น

### 1. กฎการลดทอนแบบวีเนอร์ (Wiener Suppression Rule)

ใช้แบบจำลองทางสถิติของสัญญาณเสียงพูดในโดเมน STDFT เป็นแบบเกาส์เซียน (Gaussian Distribution) และทำการประมาณขนาดของสัญญาณเสียงพูดโดยใช้ตัวประมาณแบบค่าเฉลี่ยกำลังสองน้อยสุด (Minimum Mean Square Error, MMSE) ในการประมาณทั้งองค์ประกอบของสัญญาณ [3]

### 2. กฎการลดทอนแบบการประมาณแอมพลิจูดของสเปกตรัมในช่วงเวลาสั้น ๆ (STSA Suppression Rule)

ใช้แบบจำลองทางสถิติของสัญญาณเสียงพูดในโดเมน STDFT เป็นแบบเกาส์เซียน และทำการประมาณขนาดของสัญญาณเสียงพูดโดยใช้ตัวประมาณแบบ MMSE ในการประมาณเฉพาะขนาดขององค์ประกอบของสัญญาณ (STSA) [4]

### 3. กฎการลดทอนแบบการประมาณลอการิทึมของขนาดของสเปกตรัม (Log-Spectral Amplitude, LSA)

ใช้แบบจำลองทางสถิติของสัญญาณเสียงพูดในโดเมน STDFT เป็นแบบเกาส์เซียน และทำการประมาณลอการิทึมของขนาดของสเปกตรัมของสัญญาณเสียงพูดโดยใช้ตัวประมาณแบบ MMSE [5]

จากการศึกษาพบว่า กฎการลดทอนแบบ STSA และ LSA เป็นที่ยอมรับในด้านคุณภาพของสัญญาณเสียงที่ถูกปรับปรุง กล่าวคือ มีประสิทธิภาพการลดสัญญาณรบกวนสูง และควบคุมความผิดเพี้ยนของสัญญาณเสียงที่ถูกเพิ่มคุณภาพให้ต่ำที่สุด ทั้งนี้เนื่องจากกฎการลดทอนแบบ STSA มีความซับซ้อนค่อนข้างสูง จึงมีความพยายามที่จะลดความซับซ้อนในการคำนวณลง โดยเปลี่ยนวิธีการประมาณจากการใช้ตัวประมาณแบบ MMSE เป็นการหาค่าสูงสุดหลังประสม (Maximum a Posterior, MAP) [10] รวมถึงการเปลี่ยนแบบจำลองทางสถิติของสัญญาณเสียงพูดให้เหมาะสมขึ้น เช่น จากเกาส์เซียนเป็นแกมมา (Gamma) หรือ ลาปลาเซียน (Laplacian) [8]-[9] เป็นต้น

กฎการลดทอนหรือค่าอัตราขยายที่หาได้โดยวิธีการต่างๆ นี้จะอยู่ในรูปฟังก์ชันของค่าอัตราส่วนสัญญาณต่อสัญญาณรบกวนก่อนประสพ (A Priori SNR) และค่าอัตราส่วนสัญญาณต่อสัญญาณรบกวนหลังประสพ (A Posterior SNR) เช่น กฎการลดทอนแบบ LSA ณ องค์ประกอบทางความถี่ที่  $k$  [5]

$$G_{LSA}(k) = \frac{\zeta(k)}{\zeta(k)+1} \exp\left\{\frac{1}{2} \int_{v(k)}^{\infty} \frac{e^{-t}}{t} dt\right\}$$

$$v(k) = \frac{\zeta(k)}{\zeta(k)+1} \gamma(k)$$
(7)

เมื่อ  $\zeta(k)$  คือ ค่า A Priori SNR ณ องค์ประกอบที่  $k$

$\gamma(k)$  คือ ค่า A Posterior SNR ณ องค์ประกอบที่  $k$

การประมาณค่า SNR ทั้งสอง ต่างต้องอาศัยค่าสเปกตรัมกำลังของสัญญาณเสียงรบกวน (Noise Power Spectrum) ในการประมาณทั้งสิ้น ดังนั้นการลดเสียงรบกวนแบบการลดทอนสเปกตรัมในช่วงเวลาสั้นๆ สามารถแบ่งออกได้เป็น 3 กระบวนการหลักดังแสดงในรูปที่ 4 ได้แก่

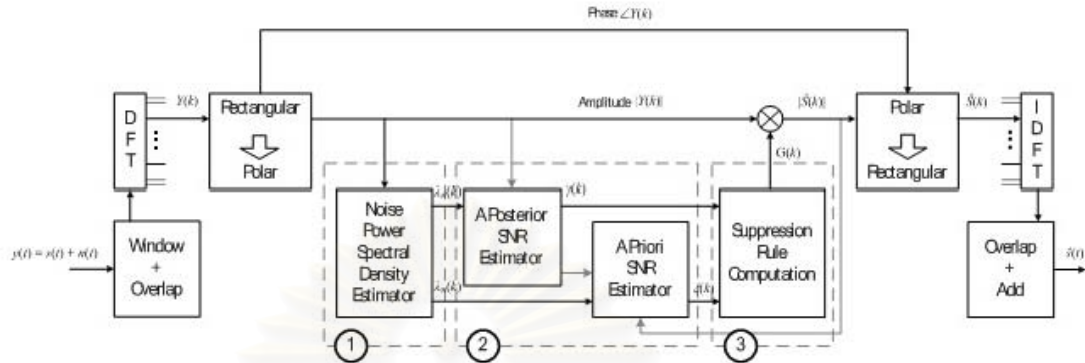
- กระบวนการที่ 1. การประมาณค่าสเปกตรัมกำลังของสัญญาณเสียงรบกวน
- กระบวนการที่ 2. การประมาณค่า A Priori SNR และค่า A Posterior SNR
- กระบวนการที่ 3. การคำนวณค่าอัตราขยายจากกฎการลดทอนที่เลือกใช้ โดยค่าประมาณของขนาดของสเปกตรัมของสัญญาณเสียงพูดที่ถูกปรับปรุงสามารถหา

ได้จากการนำค่าอัตราขยายที่หาได้คูณกับขนาดของสเปกตรัมของสัญญาณ

เสียงพูดที่ถูกรบกวน จากนั้นจึงนำมารวมกับเฟสของสัญญาณเสียงพูดที่

ถูกรบกวน ทำให้ได้เป็นสัญญาณเสียงพูดที่ถูกเพิ่มคุณภาพในโดเมน

STDFFT ก่อนที่จะทำการแปลงไปสู่โดเมนเวลาต่อไป



รูปที่ 4 การลดเสียงรบกวนด้วยการลดทอนสเปกตรัมในช่วงเวลาสั้นๆ

- เมื่อ  $(t)$  คือดัชนีเวลา (Time Index)
- $(k)$  คือดัชนีบิน หรือดัชนีองค์ประกอบทางความถี่ (Bin Index)
- $(n)$  คือดัชนีเฟรม (Frame Index)
- $y, Y$  คือสัญญาณเสียงพูดที่ถูกรบกวน หรือ สัญญาณไมโครโฟน
- $s, S$  คือสัญญาณเสียงพูดที่ปราศจากเสียงรบกวน
- $n, N$  คือสัญญาณเสียงรบกวน
- $\hat{s}, \hat{S}$  คือสัญญาณเสียงพูดที่ถูกปรับปรุง หรือถูกเพิ่มคุณภาพ

\*หมายเหตุ อักษรตัวเล็กถูกใช้กับสัญญาณในโดเมนเวลา และอักษรตัวใหญ่ถูกใช้กับสัญญาณในโดเมนความถี่

การประมาณค่าสเปกตรัมกำลังของสัญญาณเสียงรบกวนในกระบวนการที่ 1 จะกระทำในช่วงที่ไม่มีสัญญาณเสียงพูด (Non-Speech Period) โดยอาศัยสมมุติฐานที่ว่าสัญญาณเสียงรบกวนมีความเป็นจุดนิ่งยาวนาน ดังนั้นสเปกตรัมกำลังของสัญญาณเสียงรบกวนที่ประมาณได้ในช่วงที่ไม่มีสัญญาณเสียงพูดสามารถนำไปใช้เป็นค่าสเปกตรัมกำลังของสัญญาณเสียงรบกวนในทุกๆ ช่วงได้ [2] อย่างไรก็ตามการประมาณค่าดังกล่าวจำเป็นต้องอาศัยตัวตรวจหาช่วงเวลาที่มืเสียงพูด (Voice Activity Detector, VAD) เพื่อบ่งชี้ช่วงที่มีและไม่มีสัญญาณเสียงพูด [3] ดังนั้นในการแก้ไขปัญหาเสียงรบกวนและเสียงสะท้อนควบคู่กัน VAD ของวิธี NR จะ

มีหน้าที่คล้ายคลึงกับตัวตรวจหาสถานการณ์ดับเบิ้ลทอล์ก (DTD) ของระบบ AEC ทั้งนี้การประมาณค่าสเปกตรัมกำลังของสัญญาณเสียงรบกวนโดยอาศัย VAD มีข้อเสียคือ หากสัญญาณรบกวนเกิดการเปลี่ยนแปลงไม่เป็นจุดนิ่งในช่วงที่มีสัญญาณเสียงพูด จะทำให้ระบบไม่สามารถติดตามการเปลี่ยนแปลงดังกล่าวได้ ส่งผลให้ไม่สามารถลดสัญญาณเสียงรบกวนในช่วงดังกล่าวลงได้

ค่าประมาณสเปกตรัมกำลังของสัญญาณเสียงรบกวน มีบทบาทสำคัญต่อประสิทธิภาพของวิธี STSA เป็นอย่างมาก กล่าวคือ ยิ่งสามารถทำการประมาณสเปกตรัมกำลังของสัญญาณเสียงรบกวนได้อย่างแม่นยำมากเท่าไร คุณภาพของสัญญาณเสียงพูดที่ถูกปรับปรุงด้วยวิธี STSA นี้ จะอยู่ในเกณฑ์ที่ดี กล่าวคือมีความบิดเพี้ยนของสัญญาณเสียงที่ต่ำและสามารถลดปริมาณของสัญญาณเสียงรบกวนลงได้อย่างถูกต้อง แต่หากค่าประมาณสเปกตรัมกำลังของสัญญาณเสียงรบกวน มีความผิดพลาดมาก จะทำให้กฎการลดทอนทำงานได้แย่ง ส่งผลให้คุณภาพเสียงของสัญญาณออกของระบบมีความบิดเพี้ยนมาก และไม่สามารถลดปริมาณของสัญญาณเสียงรบกวนลงได้อย่างถูกต้อง ดังนั้นจึงมีการพัฒนาวิธีการประมาณค่าสเปกตรัมกำลังของสัญญาณเสียงรบกวนขึ้นมาหลาย เช่น การประมาณค่าสเปกตรัมกำลังของสัญญาณเสียงรบกวน ตลอดทั้งช่วงที่มีและช่วงที่ไม่มีเสียงพูด [13]-[14] โดยไม่ต้องอาศัย VAD ซึ่งวิธีดังกล่าวนี้ทำให้ระบบสามารถติดตามการเปลี่ยนแปลงของสัญญาณรบกวนได้ดียิ่งขึ้นเมื่อสัญญาณรบกวนมีลักษณะทางสถิติที่ไม่มีความเป็นจุดนิ่ง (Non-Stationary)

หลังจากทำการประมาณสเปกตรัมกำลังของสัญญาณเสียงรบกวนแล้ว กระบวนการต่อไปคือการประมาณค่า A Priori SNR และค่า A Posterior SNR โดยค่า A Priori SNR  $\zeta(k)$  ถูกนิยามดังต่อไปนี้ [4]

$$\zeta(k) = \frac{\lambda_s(k)}{\lambda_n(k)} \quad (8)$$

เมื่อ

$$\lambda_s(k) = E\{|S(k)|^2\} \quad (9)$$

คือ ค่าความแปรปรวนของสัญญาณเสียงพูด ณ องค์ประกอบทางความถี่ที่  $k$ ,  $E\{\cdot\}$  แทนค่าคาดหวัง (Expectation) ของตัวแปรภายในวงเล็บ  $|S(k)|^2$  คือขนาดกำลังสองของ



สัญญาณเสียงพูด ณ องค์ประกอบทางความถี่ที่  $k$  และ  $S(k)$  เป็นสัญญาณเสียงพูดในโดเมน STDFT ที่มีค่าเฉลี่ยทั้งในส่วนจริงและส่วนจินตภาพเป็นศูนย์ และ

$$\lambda_N(k) = E\{|N(k)|^2\} \quad (10)$$

คือค่าความแปรปรวนของสัญญาณเสียงรบกวน ณ องค์ประกอบทางความถี่ที่  $k$  โดย  $|N(k)|^2$  คือขนาดกำลังสองของสัญญาณเสียงรบกวน ณ องค์ประกอบทางความถี่ที่  $k$  โดยที่  $N(k)$  เป็นสัญญาณเสียงรบกวนในโดเมน STDFT ที่มีค่าเฉลี่ยทั้งในส่วนจริงและส่วนจินตภาพเป็นศูนย์ ทั้งนี้ค่าสเปกตรัมกำลังของสัญญาณเสียงรบกวนที่ทำการประมาณมาได้จากกระบวนการที่ 1 จะถูกใช้แทนค่าความแปรปรวนของสัญญาณเสียงรบกวน ณ องค์ประกอบทางความถี่นั้นๆ

สำหรับค่า A Posterior SNR  $\gamma(k)$  ถูกนิยามดังต่อไปนี้ [4]

$$\gamma(k) = \frac{|X(k)|^2}{\lambda_N(k)} \quad (11)$$

เมื่อ  $|X(k)|^2$  คือขนาดกำลังสองของสัญญาณเสียงพูดที่ถูกรบกวนที่องค์ประกอบทางความถี่ที่  $k$

จากนิยามของค่า A Posterior SNR ในสมการที่ (11) นั้น จะเห็นว่าขึ้นอยู่กับค่าสเปกตรัมกำลังของสัญญาณเสียงรบกวนที่ทำการประมาณไว้ และค่าขนาดกำลังสองของสัญญาณเสียงพูดที่ถูกรบกวน ณ องค์ประกอบทางความถี่ที่  $k$  แต่ค่า A Priori SNR ในสมการที่ (8) นั้นขึ้นอยู่กับค่าความแปรปรวนของสัญญาณเสียงพูด ซึ่งไม่ทราบค่า จึงต้องทำการประมาณค่า A Priori SNR โดยมี 2 วิธีที่นิยมใช้กันดังนี้

วิธีที่ 1. การถ่วงน้ำหนัก หรือ วิธีความควรจะเป็นสูงสุด (เรียกตามที่มาของวิธีการ) [4]



$$\bar{\gamma}(k, n) = \alpha \bar{\gamma}(k, n-1) + \frac{(1-\alpha)}{\beta} \gamma(k, n) - 1 \quad (12)$$

$$\zeta(k, n) = \begin{cases} \bar{\gamma}(k, n) - 1 & \bar{\gamma}(k, n) - 1 \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (13)$$

เมื่อ  $\alpha$  คือค่าความจำ (Forgetting Factor) และ  $\beta$  เป็นค่าคงที่ในการถ่วงน้ำหนัก (Weighting Factor) การประมาณนี้ถูกใช้ในช่วงระยะแรกๆ ของการพัฒนาวิธีการลดเสียงรบกวน หากแต่ถ้าข้อเสียของวิธีการประมาณนี้คือ ค่า A Priori SNR ที่ประมาณได้จะมีการแกว่งตัว (Deviation) ค่อนข้างมาก ซึ่งจะทำให้ค่าอัตราขยายที่ได้มีการแกว่งตัวมากตามไปด้วย ส่งผลให้เกิดสัญญาณเสียงรบกวนตกค้างที่เรียกว่า สัญญาณเสียงรบกวนแบบเสียงดนตรี (Musical Noise) เหตุที่เรียกว่าสัญญาณเสียงรบกวนแบบเสียงดนตรีนั้นเนื่องมาจากสัญญาณเสียงรบกวนตกค้างนี้เกิดขึ้นอย่างสม่ำเสมอในทางองค์ประกอบทางความถี่และในทางเวลา โดยเป็นผลจากการที่ค่าอัตราขยายในแต่ละองค์ประกอบทางความถี่แกว่งตัวตามเวลาอย่างสม่ำเสมอ ทำให้สัญญาณเสียงรบกวนตกค้างที่ได้ยินมีความคล้ายคลึงกับเสียงดนตรี ที่โน้ตต่างๆ กันตามความถี่ต่างกันเหล่านั้น

วิธีที่ 2. การประมาณค่า A Priori SNR ด้วยการตัดสินใจโดยตรง (Decision-Directed, DD) [4]

$$\zeta(k, n) = \alpha \frac{|\hat{S}(k, n-1)|^2}{\lambda_N(k, n)} + (1-\alpha)[\gamma(k, n) - 1] \quad (14)$$

โดยที่  $|\hat{S}(k, n-1)|^2$  คือขนาดกำลังสองของสัญญาณเสียงพูดที่ถูกปรับปรุงแล้ว ณ องค์ประกอบทางความถี่ที่  $k$  ของเฟรมก่อนหน้า ( $n-1$ ) และ  $\alpha$  เป็นค่าคงที่ระหว่าง 0 และ 1 วิธีการประมาณนี้ทำให้ค่า A Priori SNR มีค่าค่อนข้างคงที่ในช่วงที่ค่า A Posterior SNR มีค่าต่ำ หรือ ช่วงที่มีแต่สัญญาณเสียงรบกวนเพียงอย่างเดียว และมีค่าได้ตามค่า A Posterior SNR ในช่วงที่ค่า A Posterior SNR มีค่าสูง หรือ ช่วงที่มีทั้งสัญญาณเสียงพูดและสัญญาณเสียงรบกวน วิธีการประมาณนี้จึงให้ค่า A Priori SNR  $\zeta(k, n)$  ที่มีการแกว่งตัวน้อยกว่าวิธีการประมาณโดยการถ่วงน้ำหนัก ทำให้สามารถลดสัญญาณเสียงรบกวนตกค้างแบบเสียงดนตรีลงได้ พฤติกรรมของการประมาณค่า A Priori SNR ด้วยการตัดสินใจโดยตรง ถูกอธิบายอย่างละเอียดใน [6] นอกจากนี้ ประสิทธิภาพในการ

ประมาณค่า A Priori SNR ที่แม่นยำและความซับซ้อนในการคำนวณต่ำของการตัดสินใจ โดยตรงทำให้วิธีการนี้เป็นที่นิยมใช้กันอย่างแพร่หลาย

หลังจากได้กฎการลดทอน  $G(k)$  ที่เหมาะสมแล้ว ขนาดของสัญญาณเสียงพูดที่ถูกรบกวน  $|Y(k)|$  จะถูกนำมาคูณเพื่อประมาณขนาดของสัญญาณเสียงพูดที่ถูกรับปรุง  $|\hat{S}(k)|$

$$|\hat{S}(k)| = G(k) |Y(k)| \quad (15)$$

จากนั้นขนาดของสัญญาณเสียงพูดที่ถูกรับปรุงจะถูกรวมเข้ากับเฟสของสัญญาณเสียงพูดที่ถูกรบกวน เพื่อประกอบเป็นสัญญาณเสียงพูดที่ถูกรับปรุง

$$\hat{S}(k) = |\hat{S}(k)| \exp\{j\angle Y(k)\} \quad (16)$$

ก่อนถูกแปลงกลับเป็นสัญญาณเสียงพูดที่ถูกรับปรุงในโดเมนเวลา  $\hat{s}(t)$  ต่อไป

ความซับซ้อนในการคำนวณ ของการลดเสียงรบกวน จะพิจารณาที่การคูณจำนวนจริงต่อหนึ่งรอบการซีกตัวอย่าง (RMPs) สำหรับวิธี STSA จะมีความซับซ้อนในการคำนวณประมาณ 60-80 RMPs นอกจากนี้หากนำวิธีนี้ ไปใช้ในระบบสื่อสารแล้ว จะทำให้เกิดการประวิงเวลาในการประมวลผลประมาณ 32 มิลลิวินาที ซึ่งเป็นการประวิงเวลาที่ยอมรับได้ ทั้งนี้ ในส่วนที่เหลือต่อไปในรายงานจะขอแทนคำว่า “การลดเสียงรบกวนแบบการลดทอนสเปกตรัมในช่วงเวลาสั้นๆ” ด้วย “การกดเสียงรบกวน” (Noise Suppression, NS) เพื่อความกระชับ

### 3.3 การลดเสียงสะท้อนและเสียงรบกวนพร้อมๆ กัน (AENR)

สถานการณ์หนึ่งของระบบสื่อสารแบบแชนด์ฟรีดังแสดงในรูปที่ 1 โดยสัญญาณไมโครโฟนในห้องใกล้  $y(t)$  ประกอบไปด้วยสัญญาณเสียงจากผู้พูดทางห้องใกล้  $s(t)$

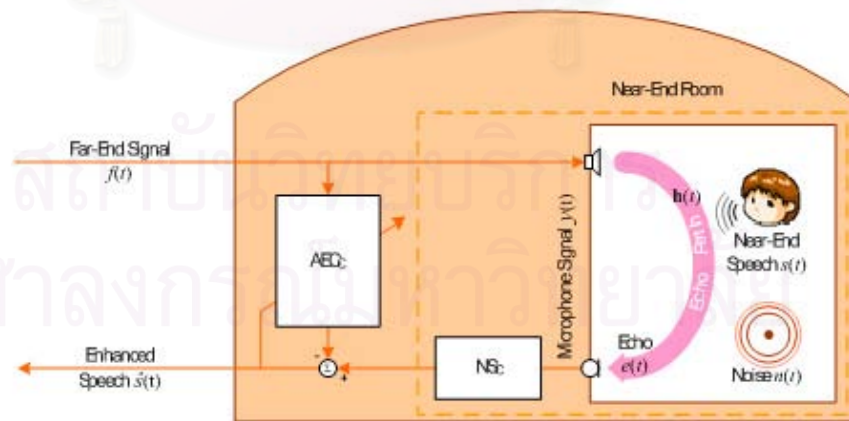
สัญญาณเสียงรบกวน  $n(t)$  และสัญญาณเสียงสะท้อน  $e(t)$  ซึ่งเกิดจากสัญญาณเสียงจากผู้พูดทางฝั่งห้องไกล  $f(t)$  ผ่านวิถีสะท้อนทางเสียงในห้องใกล้  $h(t)$

แรงจูงใจของการพิจารณาการลดเสียงสะท้อนและเสียงรบกวนพร้อมๆ กันหรือระบบ AENR นั้นอยู่ที่ข้อสังเกตที่ว่า **คู่สนทนาทางห้องไกลจะไม่ได้ยินสัญญาณเสียงสะท้อนตกค้าง หากสัญญาณรบกวนที่เหลืออยู่มีกำลังมากกว่า** ทั้งนี้เป็นเพราะคุณสมบัติจิตวิทยาทางเสียงของมนุษย์ (Psychoacoustic) ที่เรียกว่าการพรางทางความถี่ (Frequency Masking) ซึ่งหมายความว่า ระบบ AENR สามารถลดความซับซ้อนในการคำนวณของระบบ AER ลงได้ งานวิจัยของระบบ AENR เพื่อแก้ไขปัญหาดังกล่าวทั้งสัญญาณเสียงสะท้อนและสัญญาณเสียงรบกวนพร้อมๆ กันในห้องใกล้ในระบบการสื่อสารแบบแฮนด์ฟรี สามารถแบ่งออกได้เป็น 4 แนวทาง ได้แก่

**แนวทางที่ 1: ระบบ AENR แบบวิธี NS<sub>C</sub><sup>3</sup> ต่อด้วยระบบ AEC<sub>C</sub><sup>4</sup>**

จากข้อสังเกตที่ว่าระบบ AEC<sub>C</sub> มีประสิทธิภาพการทำงานแยกลงเมื่อสัญญาณเสียงรบกวนที่รวมอยู่ในห้องใกล้มีระดับสูง (หรือมี SNR ต่ำ) ดังนั้นจึงควรลดเสียงรบกวนก่อนการตัดเสียงสะท้อนดังบล็อกไดอะแกรมในรูปที่ 5

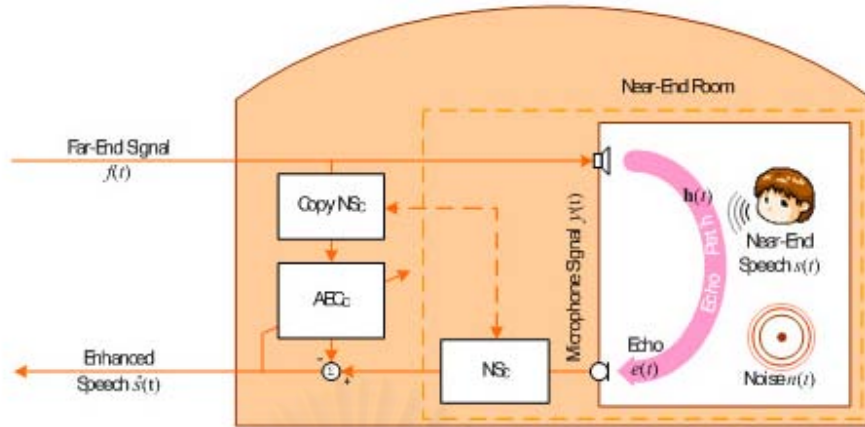
ปัญหาที่เกิดขึ้นเมื่อใช้แนวทางที่ 1 นี้ คือ วิถีสะท้อนทางเสียงที่ระบบ AEC<sub>C</sub> จะต้องทำการประมาณได้รับผลกระทบจากวิธี NS<sub>C</sub> ซึ่งมีความไม่เป็นเชิงเส้น (Non-linear) จึงมีการเสนอให้นำวงจรกรองที่เหมือนกับวงจรกรองของระบบ NS<sub>C</sub> ไปต่อเรียงไว้ก่อนนำวงจรกรองแบบปรับตัวในระบบ AEC<sub>C</sub> ดังรูปที่ 6 เพื่อลดผลของปัญหาความไม่เป็นเชิงเส้นดังกล่าว [30]



รูปที่ 5 บล็อกไดอะแกรมของระบบ AENR ตามแนวทางที่ 1

<sup>3</sup> NS<sub>C</sub> หมายถึง การลดเสียงรบกวนดั้งเดิม (Conventional) ไม่ได้รับการปรับปรุง

<sup>4</sup> AEC<sub>C</sub> หมายถึง การตัดเสียงสะท้อนดั้งเดิม (Conventional) ไม่ได้รับการปรับปรุง



รูปที่ 6 การแก้ไขปัญหาคำไม่เป็นเชิงเส้นในระบบ AENR ตามแนวทางที่ 1

อย่างไรก็ตามเนื่องจากแนวทางที่ 1 เป็นการนำระบบ  $NS_c$  มาต่อเรียงกับระบบ  $AEC_c$  ดังนั้นความซับซ้อนในการคำนวณของระบบรวมจะเพิ่มขึ้น และความซับซ้อนในการคำนวณของระบบยังขึ้นอยู่กับจำนวนสัมประสิทธิ์ของวงจรรองแบบปรับตัวของระบบ  $AEC_c$  ที่ต้องมากเพียงพอสำหรับการประมาณเสียงสะท้อนทางเสียง

ทั้งนี้ ลำดับก่อนหลังของระบบ AEC และระบบ NS เป็นสิ่งสำคัญ [27] ในการพัฒนาระบบ AENR จึงมีการหาคำตอบที่เหมาะสม (Optimal Solution) ในเชิงค่าเฉลี่ยของค่าผิดพลาดกำลังสองน้อยสุด (MMSE) บนสมมติฐานที่ว่าสัญญาณที่พิจารณาในระบบทุกสัญญาณมีความเป็นจุดนิ่ง และสัญญาณเสียงพูดทางฝั่งห้องใกล้  $s(t)$  ถูกประมาณได้จากการคอนโวลูชันของวงจรรองเชิงเส้นกับสัญญาณไมโครโฟน  $y(t)$  และสัญญาณเสียงทางฝั่งห้องไกล  $f(t)$  ดังนี้ [32]

$$\hat{s}(t) = y(t) * w_1(t) + f(t) * w_2(t) \quad (17)$$

โดยที่  $w_1(t)$  และ  $w_2(t)$  เป็นวงจรรองเชิงเส้นที่มีจำนวนสัมประสิทธิ์เป็นอนันต์ (Infinite Impulse Response, IIR) หรือ มีจำนวนสัมประสิทธิ์จำกัด (Finite Impulse Response, FIR) ก็ได้

คำตอบที่ได้หลังจากการหาคำตอบที่เหมาะสมของค่าเฉลี่ยของค่าผิดพลาดกำลังสองน้อยสุด  $E\{(s(t) - \hat{s}(t))^2\}$  มีความสอดคล้องกับระบบ AENR ที่มีโครงสร้าง AEC ต่อ ด้วย NS ดังนี้

$$\hat{s}(t) = \left( y(t) + f(t) * \mathfrak{F}^{-1} \left\{ \begin{matrix} R_{fy}(k) \\ R_{ff}(k) \end{matrix} \right\} \right) * \mathfrak{F}^{-1} \left\{ \frac{R_{ss}(k)}{R_{yy}(k) - R_{yf}(k)R_{ff}(k)^{-1}R_{fy}(k)} \right\} \quad (18)$$

เมื่อ  $\mathfrak{F}^{-1}\{\cdot\}$  แทนการแปลงฟูริเยร์ผกผันแบบไม่ต่อเนื่อง (Inverse Discrete Fourier Transform)

$R_{ff}(k)$  เป็นความหนาแน่นสเปกตรัมกำลังข้าม (Cross Power Spectral Density) ระหว่างสัญญาณเสียงทางฝั่งห้องไกล  $f(t)$  และสัญญาณเสียงทางฝั่งห้องไกล  $f(t)$

$R_{fy}(k)$  เป็นความหนาแน่นสเปกตรัมกำลังข้ามระหว่างสัญญาณเสียงทางฝั่งห้องไกล  $f(t)$  และสัญญาณไมโครโฟน  $y(t)$

$R_{yf}(k)$  เป็นความหนาแน่นสเปกตรัมกำลังข้ามระหว่างสัญญาณไมโครโฟน  $y(t)$  และสัญญาณเสียงทางฝั่งห้องไกล  $f(t)$

$R_{yy}(k)$  เป็นความหนาแน่นสเปกตรัมกำลังข้ามระหว่างสัญญาณไมโครโฟน  $y(t)$  และสัญญาณไมโครโฟน  $y(t)$

$R_{ss}(k)$  เป็นความหนาแน่นสเปกตรัมกำลังข้ามระหว่างสัญญาณสัญญาณเสียงพูดทางฝั่งห้องใกล้  $s(t)$  และสัญญาณเสียงพูดทางฝั่งห้องใกล้  $s(t)$

จะเห็นได้อย่างชัดเจนว่าคำตอบที่เหมาะสมในสมการ (18) สามารถแบ่งการทำงานออกได้เป็น 2 ส่วนได้แก่ ส่วนแรกคือส่วนตัดสัญญาณเสียงสะท้อน โดยมีผลตอบเชิงความถี่ของระบบเป็น

$$C(k) = R_{fy}(k)R_{ff}(k)^{-1} \quad (19)$$

และส่วนที่สองคือส่วนที่ทำหน้าที่กีดสัญญาณเสียงสะท้อนตกค้างและสัญญาณเสียงรบกวน โดยมีผลตอบเชิงความถี่เป็น

$$H(k) = W_1(k) = R_{ss}(k)[R_{yy}(k) - R_{yf}(k)R_{ff}(k)^{-1}R_{fy}(k)]^{-1} \quad (20)$$



ความหมายของสมการที่ (18) นำมาซึ่งแนวทางการพัฒนาระบบ AENR อีก 3 แนวทาง [29] ดังจะได้อธิบายต่อไป

## แนวทางที่ 2: ระบบ AENR แบบระบบ AEC<sub>R</sub><sup>5</sup> ต่อด้วยวิธี NS<sub>C</sub>

หากมองว่าระบบ AEC ซึ่งถูกวางไว้ก่อนหน้าระบบ NS สามารถทำงานได้อย่างมีประสิทธิภาพ กล่าวคือ สัญญาณสะท้อนถูกตัดไปได้ ส่งผลให้สัญญาณไมโครโฟนและสัญญาณลำโพงที่มาจากห้องใกล้เคียงไม่มีสหสัมพันธ์กันแล้ว  $H(k)$  ในสมการที่ (20) จะลดรูปลงเหลือเพียง

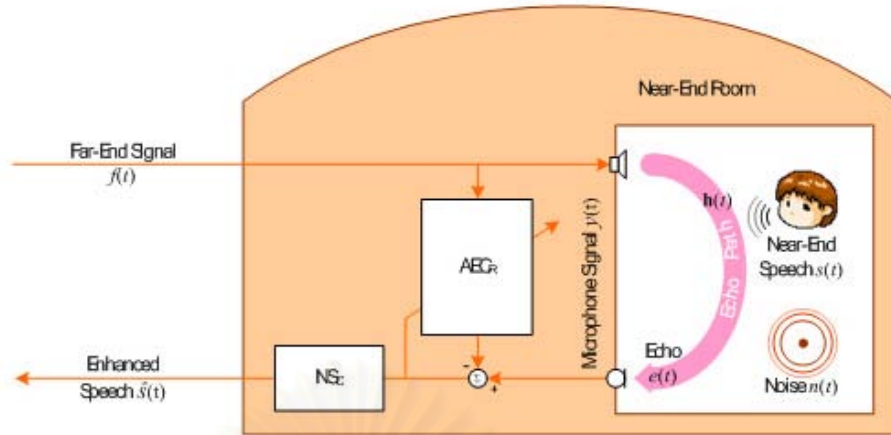
$$\begin{aligned} H(k) &= W_1(k) = R_{ss}(k)R_{yy}(k)^{-1} \\ &= R_{ss}(k)[R_{ss}(k) + R_{mm}(k)]^{-1} \end{aligned} \quad (21)$$

ซึ่งตรงกับผลเฉลยวีเนอร์สำหรับระบบ NS แบบดั้งเดิม กล่าวคือหากระบบ AEC สามารถทำงานได้อย่างมีประสิทธิภาพแล้ว ส่วนที่ทำหน้าที่กดสัญญาณเสียงสะท้อนตกค้างและสัญญาณเสียงรบกวนตามสมการ (21) จะสามารถถูกแทนที่ด้วยระบบ NS แบบดั้งเดิมได้ ดังรูปที่ 7 เนื่องจากข้อสังเกตนี้เองทำให้การพัฒนาของระบบ AENR ในแนวทางที่ 2 มุ่งเน้นไปที่การเพิ่มประสิทธิภาพการทำงานของระบบ AEC ให้สามารถทำงานภายใต้สภาพแวดล้อมที่มีสัญญาณรบกวนได้ โดยในที่นี้จะเรียกว่าระบบ AEC<sub>R</sub> [29]-[31]

สถาบันวิทยบริการ  
จุฬาลงกรณ์มหาวิทยาลัย

<sup>5</sup> AEC<sub>R</sub> หมายถึง ระบบ AEC ที่ได้รับการปรับปรุงให้มีความทนทานต่อสัญญาณเสียงรบกวน





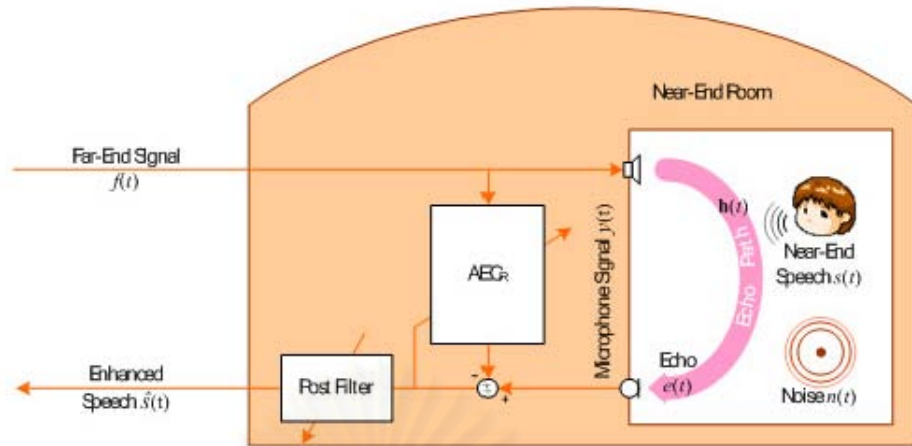
รูปที่ 7 บล็อกไดอะแกรมของระบบ AENR ตามแนวทางที่ 2

อย่างไรก็ดี ระบบ AENR ตามแนวทางที่ 2 ยังคงมีความซับซ้อนในการคำนวณที่ขึ้นอยู่กับจำนวนสัมประสิทธิ์ของวงจรรองของระบบ  $AEC_R$  ที่สูงอยู่

**แนวทางที่ 3:** ระบบ AENR แบบ  $AEC_R$  ต่อเรียงกับ วงจรรองภายหลัง (Combined Echo and Noise Reduction with Post-Filter)

จากสมการที่ (18) จะเห็นได้ว่าประสิทธิภาพสูงสุดของระบบ  $AEC_R$  ในการตัดสัญญาณเสียงสะท้อนภายใต้สภาวะแวดล้อมที่มีสัญญาณรบกวน จำเป็นต้องอาศัยวงจรรองอีกตัวหนึ่งเพื่อทำการกรองทั้งสัญญาณเสียงสะท้อนตกค้าง และสัญญาณเสียงรบกวน ออกจากสัญญาณออกของระบบ  $AEC_R$  โดยวงจรรองตัวที่เพิ่มขึ้นมานี้จะถูกพัฒนามาจากระบบ NS ซึ่งถูกเรียกว่า วงจรรองภายหลัง (Post Filter) บล็อกไดอะแกรมของระบบ AENR ในแนวทางนี้ถูกแสดงไว้ในรูปที่ 8

สถาบันวิทยบริการ  
จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 8 บล็อกไดอะแกรมของระบบ AENR ตามแนวทางที่ 3

เนื่องจากวงจรกรองภายหลังสามารถกำจัดได้ทั้งสัญญาณเสียงสะท้อนตกค้าง และสัญญาณเสียงรบกวน ทำให้สามารถลดจำนวนสัมประสิทธิ์ของวงจรกรองแบบปรับตัวในส่วน ของระบบ  $AEC_R$  ลงได้ จึงส่งผลให้อัตราการสุ่มเข้าของระบบ  $AEC_R$  ด้วยแนวทางที่ 3 นี้เร็วขึ้น และมีความทนทานต่อสัญญาณรบกวนที่สูงขึ้น [32] งานวิจัยของระบบ AENR ที่ใช้แนวทางที่ 3 จึงมุ่งเน้นไปที่การพัฒนาวงจรกรองภายหลังนี้เพื่อให้สามารถกำจัดได้ทั้งสัญญาณเสียงรบกวนและสัญญาณเสียงสะท้อนตกค้างอย่างมีประสิทธิภาพ ตัวอย่างของแนวทางนี้ได้แก่ [32]-[33] สำหรับความซับซ้อนในการคำนวณของระบบ AENR ในแนวทางนี้ จะมีแนวโน้มลดลงจาก 2 แนวทางแรก

#### แนวทางที่ 4: ระบบ AENR แบบครอบคลุมทั้งระบบ (Global System)

แนวทางนี้พิจารณาปัญหาเสียงสะท้อนและเสียงรบกวนแบบครอบคลุมทั้งระบบ (Global System) โดยนำสมการที่ (18) มาเรียบเรียงใหม่ [34] เป็น

$$\hat{s}(t) = (y(t) * w_1 + f(t) * w_1) * w_3 \quad (22)$$

โดย

$$w_1 = \mathfrak{F}^{-1}\{W_1\} \quad (23)$$

และ

$$w_3 = \mathfrak{F}^{-1} \left\{ \frac{R_{(y(t)*w_1)(x(t)*w_1)}(k)}{R_{(x(t)*w_1)(x(t)*w_1)}(k)} \right\} \quad (24)$$

ความพยายามที่จะหาระบบที่เหมาะสมนี้จะพิจารณาคำตอบที่เหมาะสมที่สุดของระบบรวมระบบหนึ่งที่สามารถกำจัดได้ทั้งสัญญาณเสียงสะท้อนและสัญญาณเสียงรบกวนพร้อมๆ กัน ดังนั้นสัญญาณเสียงพูดทางฝั่งห้องใกล้ หรือสัญญาณดับเบิ้ลทอล์กจะถูกประมาณขึ้นเป็นสัญญาณเสียงพูดที่ถูกปรับปรุง  $\hat{s}(t)$  จากสัญญาณไมโครโฟน  $y(t)$  เพียงสัญญาณเดียวและใช้วงจรรอง  $g(t)$  เพียงตัวเดียว

$$\hat{s}(t) = y(t) * g(t) \quad (25)$$

เมื่อ  $g(t)$  เป็นวงจรรองเชิงเส้นที่มีจำนวนสัมประสิทธิ์เป็นอนันต์ (IIR) หรือ มีจำนวนสัมประสิทธิ์จำกัด (FIR) ก็ได้ และสามารถแปลงให้อยู่ในโดเมนความถี่ได้เป็น

$$\hat{S}(k) = Y(k)G(k) \quad (26)$$

หากทำการหาค่าประมาณสัญญาณเสียงพูดที่ถูกปรับปรุง โดยอาศัยค่าเฉลี่ยของค่าผิดพลาดกำลังสองน้อยสุดระหว่างสัญญาณเสียงพูดที่ถูกปรับปรุง  $\hat{S}(k)$  กับสัญญาณเสียงพูดทางฝั่งห้องใกล้  $S(k)$  จะได้ว่า

$$G(k) = \frac{SDR(k)}{1 + SDR(k)} \quad (27)$$

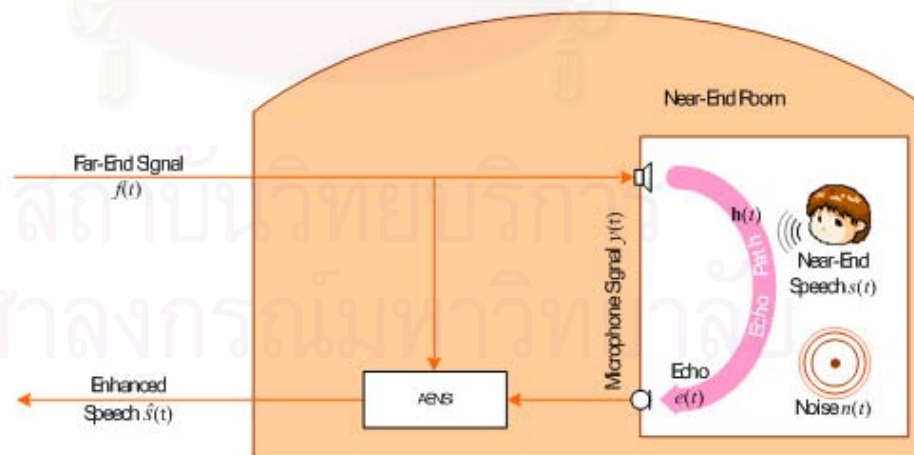
โดย  $SDR(k)$  คือค่าอัตราสัญญาณต่อสัญญาณก่อกวน (Signal-to-Disturbance Ratio) ณ องค์ประกอบทางความถี่ที่  $k$  ซึ่งคือ

$$\text{SDR}(k) = \frac{\lambda_s(k)}{\lambda_D(k)} \quad (28)$$

และ  $\lambda_D(k)$  คือค่าความแปรปรวนของสัญญาณก่อนกรอง องค์ประกอบทางความถี่ที่  $k$

จะเห็นว่าระบบ AENR แบบครอบคลุมทั้งระบบเปรียบเสมือนกับระบบ NS โดยพิจารณา สัญญาณเสียงสะท้อนเป็นสัญญาณเสียงรบกวนอีกชนิดหนึ่ง นอกจากนี้จะเห็นได้ว่าสมการที่ (27) คือ กฎการลดทอนแบบวีเนอร์นั่นเอง

ขั้นตอนหนึ่งซึ่งหลีกเลี่ยงไม่ได้ในระบบ AENR แบบครอบคลุมทั้งระบบนี้ ได้แก่ การประมาณค่าสเปกตรัมกำลังของสัญญาณก่อนกรอง ซึ่งก็คือค่าสเปกตรัมกำลังของทั้งสัญญาณเสียงสะท้อนและสัญญาณเสียงรบกวน การประมาณค่าสเปกตรัมกำลังของสัญญาณเสียงรบกวนสามารถหาได้จากสัญญาณ  $y(t)$  ในช่วงที่ไม่มีสัญญาณเสียงพูด แต่การประมาณค่าสเปกตรัมกำลังของสัญญาณเสียงสะท้อนนั้นต้องทำการประมาณโดยวิธีการที่ต่างออกไป ทั้งนี้เพราะสัญญาณเสียงสะท้อนเป็นสัญญาณที่มีความเป็นจุดนิ่งเพียงช่วงสั้นๆ เช่นเดียวกับสัญญาณเสียงพูด การประมาณค่า SDR สามารถใช้วิธีการตัดสินใจโดยตรงดังเช่นในระบบ NS ได้



รูปที่ 9 บล็อกไดอะแกรมของระบบ AENR ตามแนวทางที่ 4

ระบบ AENR แบบครอบคลุมทั้งระบบนี้มีความซับซ้อนในการคำนวณที่ต่ำกว่าระบบ AENR ในแนวทางที่ 2 และ แนวทางที่ 3 อยู่มาก โดยความซับซ้อนในการคำนวณของระบบ AENR แบบครอบคลุมทั้งระบบ เทียบเท่าได้กับวิธี NS<sub>C</sub> แต่อย่างไรก็ตามแนวทางที่ 4 นี้ทำให้เกิดความผิดเพี้ยนต่อสัญญาณออกของระบบ AENR ในช่วงสถานการณ์ดับเบิ้ลทอล์กอย่างมาก

การแก้ไขปัญหาสัญญาณเสียงสะท้อนและสัญญาณเสียงรบกวนในแนวทางที่ 4 นำไปสู่มุมมองใหม่ที่แตกต่างไปจากการแก้ไขปัญหาสัญญาณเสียงสะท้อนด้วยระบบ AEC แบบเดิม โดยพิจารณาสัญญาณเสียงสะท้อนเป็นสัญญาณเสียงรบกวนประเภทหนึ่งและทำการกดสัญญาณเสียงสะท้อนในโดเมนความถี่เช่นเดียวกับระเบียบวิธีที่ใช้ในระบบ NS<sub>C</sub> และนี่เป็นที่มาของระบบ AES ที่กล่าวถึง

ระบบ AENR แบบครอบคลุมทั้งระบบอาจพิจารณาได้ว่าเป็นการกดทั้งเสียงสะท้อนและเสียงรบกวน ดังนั้นจึงสามารถเรียกระบบ AENR แบบครอบคลุมทั้งระบบตามแนวทางที่ 4 ว่าการกดทั้งเสียงสะท้อนและเสียงรบกวน (Acoustic Echo and Noise Suppression, AENS) นอกจากนี้แล้ว ระบบ AENR แบบครอบคลุมทั้งระบบนี้อาจนำไปใช้เป็นวงจรกรองภายหลัง ตามแนวทางที่ 3 ได้อีกด้วย โดยแทนที่การประมาณค่าสเปกตรัมกำลังของสัญญาณเสียงสะท้อนด้วยการประมาณค่าสเปกตรัมกำลังของสัญญาณเสียงสะท้อนตกค้าง

### 3.3.1 การกดทั้งเสียงสะท้อนและเสียงรบกวน (AENS)

จากการศึกษาแนวทางการพัฒนาระบบ AENR ทั้ง 4 แนวทาง สามารถสรุปความแตกต่างในด้านการลดลงของเสียงรบกวน (Noise Attenuation, NA), การลดลงของเสียงสะท้อน (Echo Attenuation, EA), ความผิดเพี้ยนของเสียงพูด (Speech Distortion, SD) และความซับซ้อนในการคำนวณ (Computational Complexity) ดังแสดงในตารางที่ 2

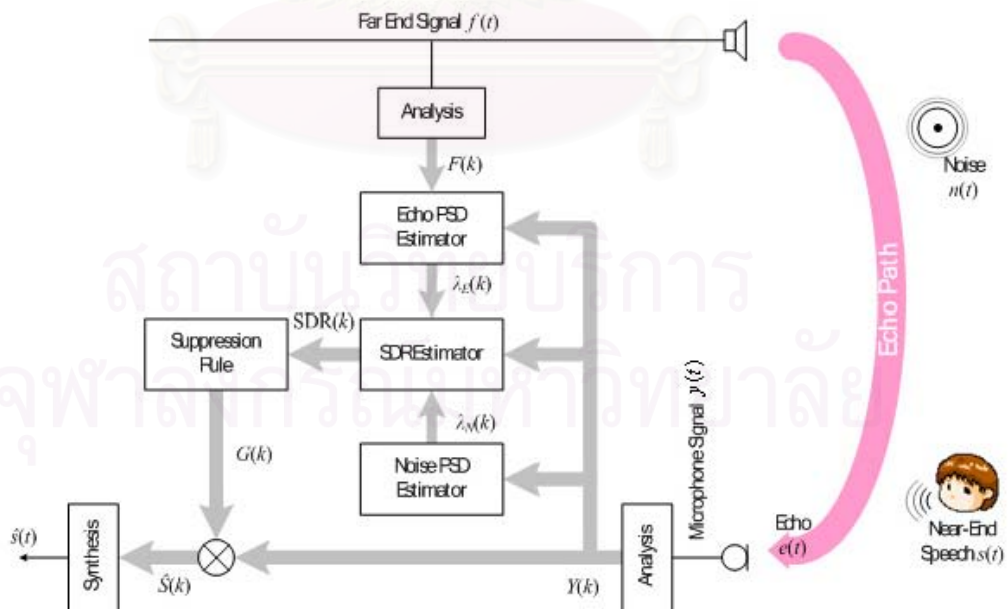


ตารางที่ 2 การเปรียบเทียบการทำงานโดยสังเขปของระบบ AENR ทั้ง 4 แนวทาง

	แนวทางที่ 1		แนวทางที่ 2		แนวทางที่ 3		แนวทางที่ 4	
	ST	DT	ST	DT	ST	DT	ST	DT
การลดลงเสียงรบกวน	ดี	ดี	ดี	ดี	ดี	ดี	ดี	ดี
การลดลงของเสียงสะท้อน	ดี	ดี	ดี	ดี	ดี	ดี	ปานกลาง	ปานกลาง
ความผิดเพี้ยนของสัญญาณเสียงพูด	น้อย	น้อย	น้อย	น้อย	น้อย	ปานกลาง	น้อย	สูง
ความซับซ้อนในการคำนวณ	สูง		สูง		ปานกลาง		ต่ำ	

หมายเหตุ เมื่อ ST คือ สถานการณ์ซึ่งเกิดทอล์ก และ DT คือ สถานการณ์ดับเบิ้ลทอล์ก

โครงการวิจัยนี้ต้องการเพิ่มคุณภาพของสัญญาณเสียงในระบบการสื่อสารแบบแฮนด์ฟรีโดยเลือกพิจารณาวิธีการแก้ปัญหาสัญญาณเสียงสะท้อนและสัญญาณเสียงรบกวน ด้วยระบบ AENR แบบครอบคลุมทั้งระบบตามแนวทางที่ 4 หรือที่เรียกว่า ระบบ AENS เนื่องจากเป็นระบบที่มีความซับซ้อนในการคำนวณต่ำ เหมาะสำหรับการใช้งานที่ไม่สามารถรองรับความซับซ้อนที่สูงมากได้ เช่น การสนทนาทางเสียงบนโครงข่ายอินเทอร์เน็ตที่มีเครื่องคอมพิวเตอร์ส่วนบุคคลเป็นเครื่องรับส่งปลายทาง และการสื่อสารผ่านทางโทรศัพท์มือถือแบบแฮนด์ฟรี เป็นต้น



รูปที่ 10 บล็อกไดอะแกรมของระบบ AENR ตามแนวทางที่ 4 หรือ ระบบ AENS ใน [34]

ระบบ AENS ใน [34] ถูกแสดงดังรูปที่ 10 โดยเสนอให้ทำการประมาณค่าสเปกตรัมกำลัง (Echo Power Spectrum) จากการหาค่าอัตราส่วนระหว่างความหนาแน่นสเปกตรัมกำลังข้าม (Cross Power Spectral Density) ของสัญญาณเสียงพูดทางห้องไกล  $f(t)$  และสัญญาณไมโครโฟน  $y(t)$  กับความหนาแน่นสเปกตรัมกำลังของสัญญาณทางฝั่งห้องไกล  $f(t)$  ดังนี้

$$\lambda_E(k) = \frac{|\lambda_{fy}(k)|^2}{\lambda_{ff}(k)} \quad (29)$$

เมื่อ  $\lambda_{fy}(k)$  คือความหนาแน่นสเปกตรัมกำลังข้ามของ  $f(t)$  และ  $y(t)$

$\lambda_{ff}(k)$  คือความหนาแน่นสเปกตรัมกำลังของ  $f(t)$

โดยค่า  $\lambda_{fy}(k)$  และ  $\lambda_{ff}(k)$  ถูกประมาณได้โดยใช้วงจรรองแบบ IIR อันดับหนึ่ง และทำการประมาณสเปกตรัมกำลังของสัญญาณรบกวนในช่วงที่ไม่มีเสียงพูดทั้งทางห้องไกลและห้องใกล้ จากนั้นจึงคำนวณหาค่าอัตราส่วนสัญญาณต่อสัญญาณก่อนรบกวนก่อนประลบ

$$\text{SDR}(k) = \frac{1}{\text{SNR}(k)^{-1} + \text{SER}(k)^{-1}} \quad (30)$$

เมื่อ  $\text{SNR}(k)$  คือค่าอัตราส่วนสัญญาณต่อสัญญาณรบกวนก่อนประลบ

$\text{SER}(k)$  คือค่าอัตราส่วนสัญญาณต่อสัญญาณเสียงสะท้อนก่อนประลบ

ค่าอัตราส่วนสัญญาณต่อสัญญาณก่อนรบกวนก่อนประลบ  $\text{SDR}(k)$  จะถูกประมาณโดยการตัดสินใจโดยตรงแยกจากกันดังสมการที่ (31)-(32)

$$\text{SNR}(k) = \alpha_N \frac{|S(k, n-1)|^2}{\lambda_N(k, n)} + (1 - \alpha_N) \left[ \frac{|Y(k, n)|^2}{\lambda_N(k, n)} - 1 \right] \quad (31)$$

และ

$$\text{SER}(k) = \alpha_E \frac{|S(k, n-1)|^2}{\lambda_E(k, n)} + (1 - \alpha_E) \left[ \frac{|Y(k, n)|^2}{\lambda_E(k, n)} - 1 \right] \quad (32)$$

เมื่อ  $\alpha_E, \alpha_N$  เป็นค่าคงที่ระหว่าง 0 และ 1

และใช้คู่กับกฎการลดทอนแบบวีเนอร์  $G(k)$  ที่หาได้จากสมการที่ (27)

หน้าที่การทำงานต่างๆ ของระบบ AENS ใน [34] ถูกสรุปไว้ในตารางที่ 3

ตารางที่ 3 ระเบียบวิธีที่ถูกใช้ในระบบ AENS ใน [34]

Function	Technique
Echo Power Spectral Density Estimator	$ \lambda_{fy}(k) ^2 /  \lambda_{ff}(k) ^2$
Noise Power Spectral Density Estimator	Estimate in Non-Speech Period
SDRstimator	Decision Directed Method
Suppression Rule	Wener Suppression Rule

เนื่องจากระบบ AENS ใช้การลดทอนสเปกตรัมในช่วงเวลาสั้นๆ ที่พิจารณาสัญญาณเสียงสะท้อนเป็นสัญญาณเสียงรบกวนอีกสัญญาณหนึ่ง ดังนั้นการวิเคราะห์ระบบ AENS สามารถกระทำโดยอิงคุณสมบัติของการลดทอนสเปกตรัมในช่วงเวลาสั้นๆ ได้ และจากการวิเคราะห์พบว่าระบบ AENS ใน [34] ประสบกับปัญหา 2 ประการดังนี้

## 1. ปัญหาความผิดเพี้ยนของสัญญาณเสียงพูดที่ถูกปรับปรุง

ระบบ AENS ประสบกับปัญหาความผิดเพี้ยนของสัญญาณเสียงพูดที่ถูกปรับปรุงเป็นอย่างมากในช่วงสถานการณ์ดับเบิลทอล์ก เนื่องจากสัญญาณก่อกวนในช่วงดังกล่าวคือสัญญาณเสียงรบกวนและสัญญาณเสียงสะท้อน ทั้งนี้สัญญาณเสียงสะท้อนมีกำลังใกล้เคียงกับสัญญาณพูดของผู้พูดทางห้องไกล หากทำการเปรียบกับการลดเสียงรบกวนแบบการลดทอนสเปกตรัมในช่วงเวลานั้นๆ ก็คือในกรณีที่สัญญาณเสียงพูดที่ถูกรบกวนมีค่า SNR ที่ต่ำมากนั่นเอง

## 2. ปัญหาสัญญาณก่อกวนตกค้าง

นอกจากความผิดเพี้ยนของสัญญาณเสียงพูดที่ถูกปรับปรุงแล้ว ระบบ AENS ยังประสบกับปัญหาสัญญาณเสียงก่อกวนตกค้างอีกด้วย หากระบบการสื่อสารที่พิจารณามีเพียงสัญญาณเสียงรบกวนพื้นหลัง เป็นสัญญาณที่ไม่ต้องการของระบบ การใช้วิธี STSA จะไม่ส่งผลต่อคุณภาพเสียงของสัญญาณที่ถูกปรับปรุง ทั้งนี้ เนื่องจากสัญญาณเสียงรบกวนมีความเป็นจุดนิ่งยาวนาน เช่น เสียงเครื่องปรับอากาศ และเสียงเครื่องรถยนต์ เป็นต้น ดังนั้นคู่สนทนาทางห้องไกลจะรับรู้ถึงสัญญาณเสียงรบกวนตกค้างนี้ได้อย่างเป็นธรรมชาติ

แต่สำหรับกรณีของระบบการสื่อสารแบบแฮนด์ฟรีที่พิจารณาในโครงการนี้ ต้องการแก้ปัญหาสัญญาณเสียงสะท้อนและสัญญาณเสียงรบกวนพร้อมๆ กัน เมื่อใช้วิธี AENS สัญญาณเสียงก่อกวนตกค้างจะประกอบด้วย สัญญาณเสียงรบกวนตกค้างและสัญญาณเสียงสะท้อนตกค้าง จากคุณสมบัติความไม่เป็นจุดนิ่งของสัญญาณเสียงสะท้อนตกค้าง คู่สนทนาทางห้องไกลจึงสามารถรับรู้สัญญาณเสียงสะท้อนตกค้างนี้ และก่อให้เกิดความไม่เป็นธรรมชาติมากกว่าผลของสัญญาณเสียงรบกวนตกค้าง

นอกจากนี้ การประมาณค่าสเปกตรัมกำลังของสัญญาณเสียงสะท้อนที่ถูกนำเสนอใน [34] ยังไม่มีประสิทธิภาพการประมาณที่ดีพอ เนื่องจากถูกออกแบบมาเพื่อใช้กับงานประยุกต์ประเภทโทรศัพท์ภายในรถยนต์ จึงไม่สามารถรองรับการใช้งานในกรณีระบบการสื่อสารแบบแฮนด์ฟรีที่เกิดขึ้นในห้องที่มีระยะเวลาที่ก่อกวนยาวนานได้

ดังนั้น โครงการวิจัยนี้จึงนำแนวคิดการประมาณค่าสเปกตรัมกำลังของ สัญญาณเสียงสะท้อนของระบบ PAES ใน [36] มาใช้เพื่อทำการประมาณค่าสเปกตรัมกำลังของ สัญญาณเสียงสะท้อนในระบบ AENS ดังนั้นค่าสเปกตรัมกำลังของสัญญาณเสียงสะท้อนที่จะ นำเสนอในโครงการวิจัยนี้ จะหาจาก

$$\varepsilon(k, n) = \hat{\mathbf{H}}_R^T(k, n) \boldsymbol{\varphi}(k, n) \quad (33)$$

เมื่อ

$$\boldsymbol{\varphi}(k, n) = [\varphi(k, n) \quad \varphi(k, n-1) \quad \dots \quad \varphi(k, n-P+1)]^T \quad (34)$$

คือ ค่ากำลังของสัญญาณจากทางห้องไกล ณ องค์ประกอบทางความถี่ที่  $k$  และ

$$\hat{\mathbf{H}}_R(k, n) = [\hat{H}_R(k, n) \quad \hat{H}_R(k, n-1) \quad \dots \quad \hat{H}_R(k, n-P+1)]^T \quad (35)$$

คือ สัมประสิทธิ์เป็นจำนวนจริงของวงจรรองแบบปรับตัว ณ องค์ประกอบทางความถี่ที่  $k$  โดยจำนวนสัมประสิทธิ์ของวงจรรองแบบปรับตัว  $P$  ต้องได้รับการเลือกที่เหมาะสม

ข้อสังเกตประการหนึ่งคือ สัญญาณเข้าของวงจรรองแบบปรับตัวในกรณีนี้ เป็น สัญญาณที่มีค่าบวกเสมอ เนื่องจากเป็นค่ากำลังของสัญญาณเสียงพูดทางฝั่งห้องไกล

การตัดสินใจโดยตรงถูกใช้ควบคู่กับการประมาณค่าสเปกตรัมกำลังของสัญญาณเสียง สะท้อน เพื่อป้องกันการแกว่งของกฎการลดทอน ซึ่งจะเป็นผลให้เกิดสัญญาณเสียงรบกวนตกค้าง แบบเสียงดนตรี ดังนี้



$$\text{SER}(k) = \alpha_E \frac{|S(k, n-1)|^2}{\varepsilon(k, n)} + (1 - \alpha_E) \left[ \frac{|Y(k, n)|^2}{\varepsilon(k, n)} - 1 \right] \quad (36)$$

โดยทำการประมาณค่าสเปกตรัมกำลังของสัญญาณเสียงรบกวนจากค่าเฉลี่ยกำลังของสัญญาณเสียงรบกวนในช่วงที่ไม่ทั้งสัญญาณเสียงพูดทางฝั่งห้องใกล้และห้องไกล การหาค่าอัตราส่วนสัญญาณต่อสัญญาณรบกวนก่อนประสบ  $\text{SNR}(k)$  ใช้ตามสมการที่ (31) และเลือกใช้กฎการลดทอนแบบวีเนอร์ตามสมการที่ (27)

สมการที่ (36) แสดงให้เห็นว่าวิธีที่นำเสนอ ใช้ค่ากำลังของสัญญาณเสียงสะท้อนขณะนั้น  $\varepsilon(k, n)$  แทนค่าสเปกตรัมกำลังของสัญญาณเสียงสะท้อน  $\lambda_E(k, n)$  ของวิธีใน [34] นั้นเอง หน้าที่การทำงานต่างๆ ของระบบ AENS ที่นำเสนอขึ้นในโครงการวิจัยนี้เป็นดังแสดงในตารางที่ 4

ตารางที่ 4 แสดงขั้นตอนต่างๆ ที่ใช้ในระบบ AENS ที่นำเสนอ

Function	Technique
Echo Power Spectral Density Estimator	Adaptive Filter in Power Spectral Domain
Noise Power Spectral Density Estimator	Estimate in Non-Speech Period
SDR estimator	Decision Directed Method
Suppression Rule	Wiener Suppression Rule

ถึงแม้ว่าประสิทธิภาพการลดลงของเสียงสะท้อน (EA) ของระบบ AENS ที่นำเสนอจะสูงและไม่มีผลกระทบจากสัญญาณเสียงรบกวนตกค้างแบบเสียงดนตรี แต่ความผิดเพี้ยนของสัญญาณเสียงพูดยังคงเกิดขึ้นในช่วงสถานการณ์ดับเบิลทอล์ก ดังนั้นเพื่อลดความผิดเพี้ยนของสัญญาณเสียงพูด โดยพิจารณาผู้พูดทางห้องใกล้ กฎการลดทอนแบบ LSA จึงถูกนำมาทดลองใช้

ในระบบ AENS ที่นำเสนอด้วย นอกจากนี้ หากพิจารณาเฉพาะส่วนของ AES ในระบบ AENS ที่นำเสนอ พบว่าความซับซ้อนในการคำนวณของระบบอยู่ที่ 124 RMPs ซึ่งน้อยกว่าความซับซ้อนในการคำนวณของระบบ AEC อยู่มาก



สถาบันวิทยบริการ  
จุฬาลงกรณ์มหาวิทยาลัย

### 3.4 ผลการจำลองระบบบนคอมพิวเตอร์

ในการจำลองระบบบนคอมพิวเตอร์ด้วยโปรแกรม MATLAB จะกล่าวถึงตัวชี้วัดที่ใช้ในการทดสอบระบบ การทดสอบระบบ การทดสอบวิธีการประมาณสเปกตรัมกำลังที่นำเสนอในโครงการวิจัยนี้ ตามด้วยการทดสอบประสิทธิภาพของระบบ AENS ที่นำเสนอ

#### 3.4.1 ตัวชี้วัดที่ใช้ในการทดสอบระบบ

1. การลดลงของเสียงสะท้อน (Echo Attenuation, EA) หากจากอัตราส่วนสัญญาณเสียงสะท้อนต่อสัญญาณเสียงสะท้อนตกค้างแบบเป็นช่วงๆ (Segmental Error Return Loss Enhancement) โดยดัดแปลงจาก [32]

$$EA \text{ (dB)} = \frac{1}{L_E} \sum_{l \in L_E} F_{EA} \left\{ 10 \times \log_{10} \left( \frac{\sum_{n=0}^{T-1} e^2(t + lT/2)}{\sum_{n=0}^{T-1} \tilde{e}^2(t + lT/2)} \right) \right\} \quad (37)$$

เมื่อ  $\tilde{e}(t)$  คือ สัญญาณเสียงสะท้อนตกค้างจากระบบ

$e(t)$  คือ สัญญาณเสียงสะท้อนก่อนถูกประมวลผล

$F_{EA} \{ \}$  คือ ฟังก์ชันที่เลือกเอาเฉพาะค่าที่อยู่ระหว่าง -50 dB และ 80 dB

$L_E$  คือ จำนวนเฟรมที่มีค่าอัตราส่วนสัญญาณต่อสัญญาณผิดเพี้ยนในช่วงระหว่าง 80 dB และ -50 dB

ดังนั้น ถ้าระบบสามารถลดสัญญาณเสียงสะท้อนได้มาก EA จะมีค่ามาก

2. การลดลงของเสียงรบกวน (Noise Attenuation, NA) หากจากอัตราส่วนสัญญาณเสียงรบกวนตกค้างต่อสัญญาณเสียงรบกวนแบบเป็นช่วงๆ โดยดัดแปลงจาก [32]

$$NA \text{ (dB)} = \frac{1}{L_N} \sum_{l \in L_N} F_{NA} \left\{ 10 \times \log_{10} \left( \frac{\sum_{n=0}^{T-1} \tilde{n}^2(t + lT/2)}{\sum_{n=0}^{T-1} n^2(t + lT/2)} \right) \right\} \quad (38)$$

เมื่อ  $\tilde{n}(t)$  คือ สัญญาณเสียงรบกวนตกค้างจากระบบ

$n(t)$  คือ สัญญาณเสียงรบกวนก่อนถูกประมวลผล

$F_{NA} \{ \}$  คือ ฟังก์ชันที่เลือกเอาเฉพาะค่าที่มีค่าอยู่ระหว่าง -80 dB และ 50 dB

$L_N$  คือ จำนวนเฟรมที่มีค่าอัตราส่วนสัญญาณต่อสัญญาณผิดเพี้ยนในช่วงระหว่าง 50 dB และ -80 dB

ดังนั้น ถ้าระบบสามารถลดสัญญาณเสียงรบกวนลงได้มาก NA จะมีค่าน้อย

3. ความผิดเพี้ยนของสัญญาณเสียงพูด (Speech Distortion, SD) ถูกบ่งชี้ด้วยอัตราส่วนสัญญาณต่อสัญญาณผิดเพี้ยนแบบเป็นช่วงๆ (Segmental Signal to Distortion Ratio, SEGSDR) โดยดัดแปลงจาก [11]

$$SEGSDR \text{ (dB)} = \frac{1}{L_S} \sum_{l \in L_S} F_{SD} \left\{ 10 \times \log_{10} \left( \frac{\sum_{n=0}^{T-1} s^2(t + lT/2)}{\sum_{n=0}^{T-1} (s(t + lT/2) - \hat{s}(t + lT/2))^2} \right) \right\} \quad (39)$$

เมื่อ  $\hat{s}(t)$  คือ สัญญาณเสียงพูดทางห้องใกล้ที่ได้จากการประมวลผล

$s(t)$  คือ สัญญาณเสียงพูดทางห้องใกล้ที่ปราศจากสิ่งก่อกวนใดๆ

$T$  คือ จำนวนตัวอย่างในแต่ละเฟรมที่  $l$

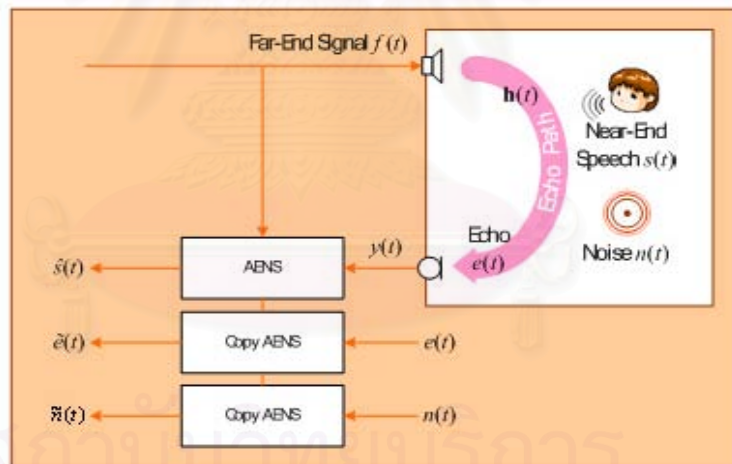
$F_{SD} \{ \}$  คือ ฟังก์ชันซึ่งจะเลือกเอาเฉพาะค่าที่อยู่ระหว่าง 35 dB และ -10 dB เพื่อป้องกันค่าอัตราส่วนในบางเฟรมที่สูงและต่ำจนผิดปกติ

$L_S$  คือ จำนวนเฟรมที่มีค่าอัตราส่วนสัญญาณต่อสัญญาณผิดเพี้ยนในช่วงระหว่าง 35 dB และ -10 dB

ดังนั้น หากสัญญาณเสียงพูดที่ปรับปรุงมีความผิดเพี้ยนน้อย SD จะมีค่าสูง

### 3.4.2 การทดสอบระบบ

ระบบที่จำลองขึ้นเพื่อการทดสอบระบบเป็นดังรูปที่ 11 เนื่องจากทราบค่าสัญญาณเสียงสะท้อนและสัญญาณเสียงรบกวนจึงสามารถหาสัญญาณเสียงสะท้อนตกค้างและสัญญาณเสียงรบกวนตกค้างของระบบซึ่งจะนำไปใช้ในการคำนวณ EA และ NA ได้ นอกจากนี้จะทราบค่าสัญญาณเสียงพูดทางห้องใกล้ที่ปราศจากสัญญาณเสียงก้องรบกวนด้วย จึงสามารถวัดค่า SD ได้



รูปที่ 11 ชุดการทดสอบระบบ

การประมาณค่าสเปกตรัมกำลังของสัญญาณเสียงรบกวนของระบบ AENS ที่นำเสนอใช้การประมาณค่ากำลังของสัญญาณเสียงรบกวนในช่วงที่ไม่มีทั้งสัญญาณเสียงพูดทางฝั่งห้องใกล้และทางฝั่งห้องไกล [30] ดังนี้



$$\lambda_N(k, n+1) = \alpha_N \lambda_N(k, n) + (1 - \alpha_N) |Y(k)|^2 \quad (40)$$

เมื่อ  $\alpha_N$  คือค่าความจำ (Forgetting Factor) มีค่าอยู่ระหว่าง 0 ถึง 1

ซึ่งในช่วงเวลาที่ไม่มีการมีทั้งสัญญาณเสียงพูดทางฝั่งห้องใกล้และทางฝั่งห้องไกล  $|Y(k)|^2 = |N(k)|^2$  โดยในการทดสอบสามารถระบุช่วงเวลาที่ไม่มีทั้งสัญญาณเสียงพูดทางฝั่งห้องใกล้และทางฝั่งห้องไกล ได้โดยการตรวจสอบโดยตรงจากไฟล์ข้อมูลของสัญญาณเสียงพูดทั้งสอง

### 3.4.3 การทดสอบระเบียบวิธีการประมาณสเปกตรัมกำลังที่นำเสนอ

สัญญาณเสียงพูดทางฝั่งห้องไกลเป็นเสียงผู้ชาย ที่ความถี่ซีกตัวอย่าง (Sampling Frequency,  $f_s$ ) เท่ากับ 8 kHz โดยทำการทดสอบระบบเฉพาะสถานการณ์ซึ่งเกิดทอล์กเท่านั้น ค่าอัตราส่วนสัญญาณต่อสัญญาณรบกวนขาเข้า (Input SNR) เป็น 20 dB ตลอดทุกการทดสอบ

ในการทดสอบ ทำการเปลี่ยนค่าจำนวนสัมประสิทธิ์ ( $L$ ) ของผลตอบสนองอิมพัลส์ของห้องใกล้ (Room Impulse Response) เป็น 128, 256, 512, 1024 และ 2048 โดยในวิธีที่นำเสนอ นั้นจำนวนสัมประสิทธิ์ของวงจรรองแบบปรับตัวในแต่ละองค์ประกอบทางความถี่ จะถูกเลือกใช้ตามสมการที่ (41)

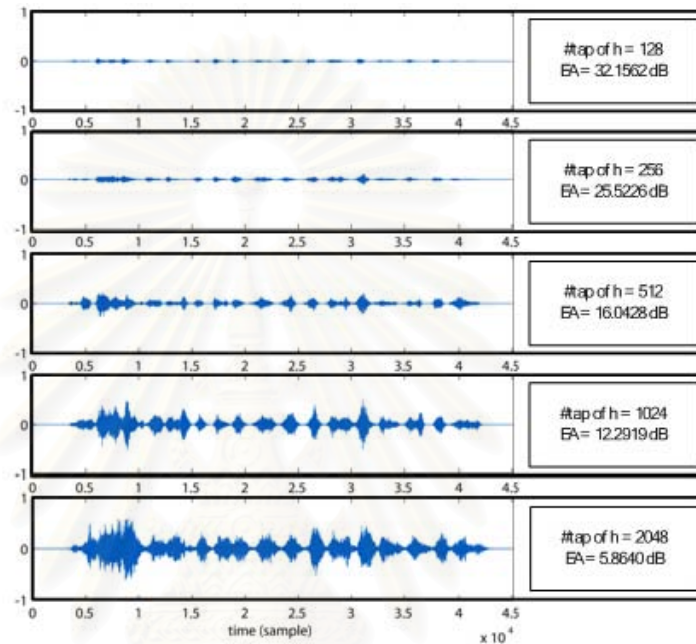
(41)

$$\begin{aligned} & \text{จำนวนสัมประสิทธิ์ของวงจรรองแบบปรับตัวในแต่ละองค์ประกอบทางความถี่} \\ & = \text{จำนวนสัมประสิทธิ์ของผลตอบสนองอิมพัลส์ของห้องใกล้} / 128 \end{aligned}$$

ทั้งนี้ 128 คือค่าที่ใช้ในการเลื่อนหน้าต่างการแปลง FFT

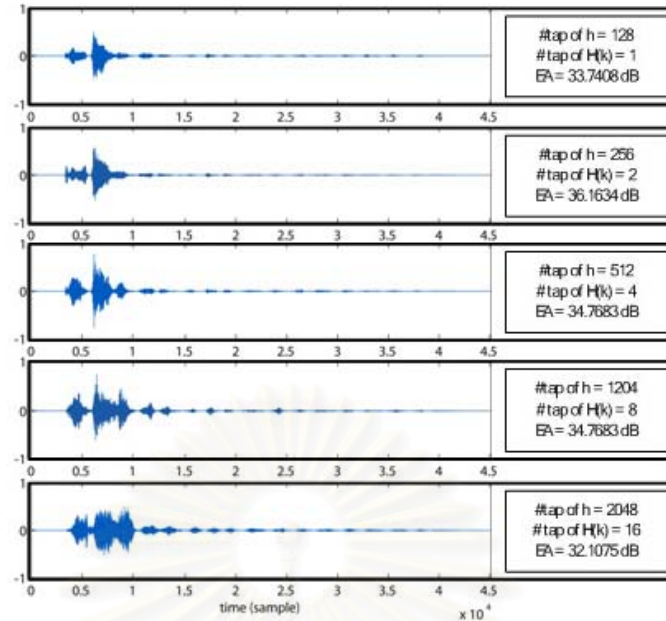
จากรูปที่ 12 และรูปที่ 13 จะเห็นว่า วิธีที่นำเสนอสามารถประมาณสเปกตรัมกำลังของสัญญาณเสียงสะท้อนได้ ดีกว่าวิธีใน [34] ดังจะเห็นจากสัญญาณเสียงสะท้อนตกค้าง (Residual Echo) ที่มีขนาดลดลงเมื่อเปรียบเทียบกับสัญญาณเสียงสะท้อนก่อนการผ่าน

ระบบ AENS ที่ใช้ วิธีในการประมาณสเปกตรัมกำลังที่นำเสนอ และจากการคำนวณค่า EA พบว่ามีค่าค่อนข้างสูง คือ ประมาณ 33 dB เมื่อจำนวนสัมประสิทธิ์ของผลตอบสนองอิมพัลส์ของห้องโถง  $L$  เป็น 128 โดยความสามารถในการลดสัญญาณเสียงสะท้อนจะลดลงเมื่อจำนวนสัมประสิทธิ์ของผลตอบสนองอิมพัลส์เพิ่มขึ้น



รูปที่ 12 รูปคลื่นของสัญญาณเสียงสะท้อนตกค้างที่จำนวนสัมประสิทธิ์ของผลตอบสนองอิมพัลส์ของห้องโถง ตามที่ระบุไว้ทางขวา และ ค่า EA ตามสมการที่ (37) ของวิธีใน [34]

สถาบันวิทยบริการ  
จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 13 รูปคลื่นของสัญญาณเสียงสะท้อนตกค้างที่จำนวนสัมประสิทธิ์ของผลตอบสนองอิมพัลส์ของห้องโถง ตามที่ระบุไว้ทางขวา และ ค่า EA ตามสมการที่ (37) ของวิธีที่นำเสนอ

### 3.4.4 การทดสอบประสิทธิภาพของระบบ AENS ที่นำเสนอ

สัญญาณเสียงพูดทางฝั่งห้องโถงและทางฝั่งห้องโถงเป็นเสียงผู้ชาย ที่ความถี่ชักตัวอย่าง 8 kHz โดยบทสนทนาจะแบ่งออกเป็น 3 ช่วง ได้แก่

- ช่วงที่ 1. มีเพียงเสียงพูดทางห้องโถงเท่านั้นสถานการณ์จึงเกิดทอล์ก
- ช่วงที่ 2. มีเพียงเสียงพูดทางห้องโถง
- ช่วงที่ 3. เสียงพูดทางห้องโถงและห้องโถงเกิดพร้อมกันแทนสถานการณ์ดับเบิ้ลทอล์ก

จำนวนสัมประสิทธิ์ของผลตอบสนองอิมพัลส์ของห้องโถง ( $L$ ) เป็น 1024 สัญญาณเสียงรบกวนที่ใช้มีการกระจายตัวแบบเกาส์เซียนที่ค่าอัตราส่วนสัญญาณต่อสัญญาณรบกวนขาเข้า (Input SNR) ณ ไมโครโฟนทางห้องโถง 4 กรณี คือ -5, 0, 5, 10 dB โดยทำการทดสอบ 3 ครั้งด้วยสัญญาณเสียงพูดทางห้องโถงที่แตกต่างกันแล้วหาค่าเฉลี่ยของ ค่า EA, NA และ SD ทั้งสาม

จากตารางที่ 5 พบว่าระบบ AENS ที่ใช้การลดทอนแบบวินเนอร์ให้ผลการลดลงของสัญญาณเสียงสะท้อน (EA) ที่น้อยกว่าระบบ AENS ที่ใช้การลดทอนแบบ LSA โดยหากคิดตามมาตรฐาน G.167 [38] ที่ว่าสัญญาณเสียงสะท้อนควรลดลงมากกว่า 20 dB จะเห็นได้ว่า

ระบบ AENS ซึ่งใช้กฎการลดทอนแบบวินเนอร์ ตอบรับกับมาตรฐานดังกล่าวได้ดีกว่ากฎการลดทอนแบบ LSA

แต่อย่างไรก็ตาม ระบบ AENS ซึ่งใช้กฎการลดทอนแบบวินเนอร์ให้ความผิดเพี้ยนของเสียงพูด (SD) ที่มากกว่า ดังแสดงในตารางที่ 7 โดยจากการทดลองฟังอย่างไม่เป็นทางการ พบว่าระบบ AENS ที่ใช้กฎการลดทอนแบบ LSA ให้ความผิดเพี้ยนของสัญญาณเสียงพูดที่น้อยกว่าแบบการใช้กฎการลดทอนวินเนอร์ ทั้งนี้ เมื่อพิจารณาสเปกโตรแกรมของสัญญาณเสียงพูดที่ถูกปรับปรุงโดยกฎการลดทอนแบบวินเนอร์ และกฎการลดทอนแบบ LSA ในรูปที่ 14 ถึงรูปที่ 17 ก็สนับสนุนข้อสรุปดังกล่าว หากแต่ถ้าเมื่อพิจารณาในช่วงสถานการณ์ดับเบิ้ลทอล์ก สัญญาณเสียงพูดที่ถูกปรับปรุงโดยกฎการลดทอนแบบวินเนอร์ ไม่ค่อยมีความเป็นธรรมชาติ (Naturalness) มากนัก ดังนั้นระบบ AENS ที่นำเสนอขึ้นนี้จึงควรได้รับการพัฒนาต่อไปทั้งในด้านประสิทธิภาพการลดทอนของเสียงรบกวนและในด้านความผิดเพี้ยนของสัญญาณเสียงพูดที่ถูกปรับปรุง

สำหรับการลดทอนของเสียงรบกวน (NA) นั้น ระบบ AENS ที่อาศัยกฎการลดทอนแบบวินเนอร์ให้ประสิทธิภาพที่ดีกว่า ระบบ AENS ที่อาศัยกฎการลดทอนแบบ LSA ดังแสดงในตารางที่ 6 อย่างไรก็ตาม สัญญาณเสียงรบกวนตกค้างของทั้งสองกรณีอยู่ในเกณฑ์ต่ำและไม่ก่อให้เกิดความรำคาญ

ตารางที่ 5 ผลการทดสอบ EA (dB) ของระบบ AENS ที่นำเสนอ

	Wiener	LSA
Input SNR (dB)	Suppression Rule	Suppression Rule
-5	40.4902	14.7915
0	35.4511	15.4356
5	30.6755	15.2475
10	27.1163	14.3565

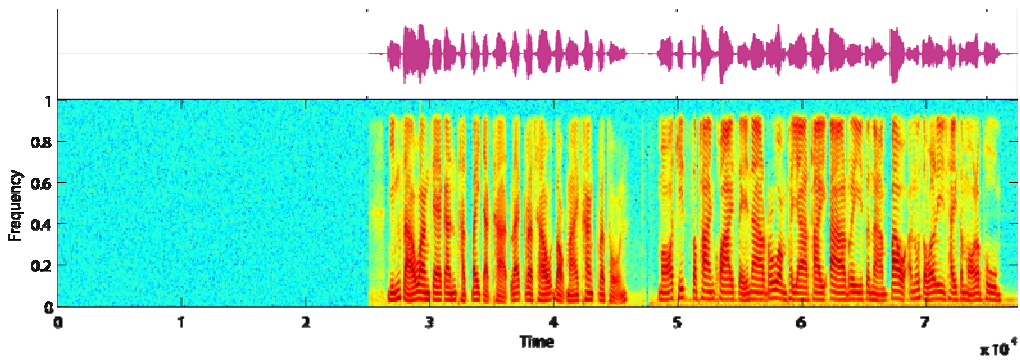
ตารางที่ 6 ผลการทดสอบ NA (dB) ของระบบ AENS ที่นำเสนอ

Input SNR (dB)	Wiener	LSA
	Suppression Rule	Suppression Rule
-5	-27.1023	-16.5590
0	-24.2990	-14.9749
5	-21.8161	-13.4302
10	-19.6719	-11.6431

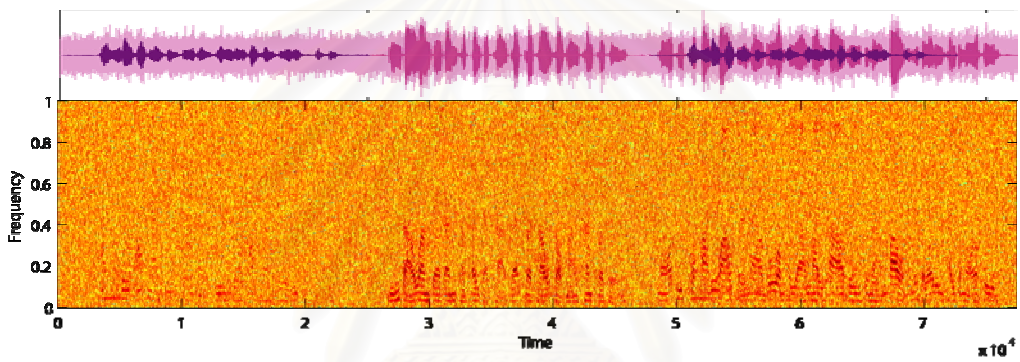
ตารางที่ 7 ผลการทดสอบ SD (dB) ของระบบ AENS ที่นำเสนอ

Input SNR (dB)	Microphone Signal	Wiener	LSA
		Suppression Rule	Suppression Rule
-5	1.7472	2.7660	3.2716
0	3.8787	4.6280	5.2777
5	6.6405	6.8332	7.6337
10	9.7980	8.8847	9.8349

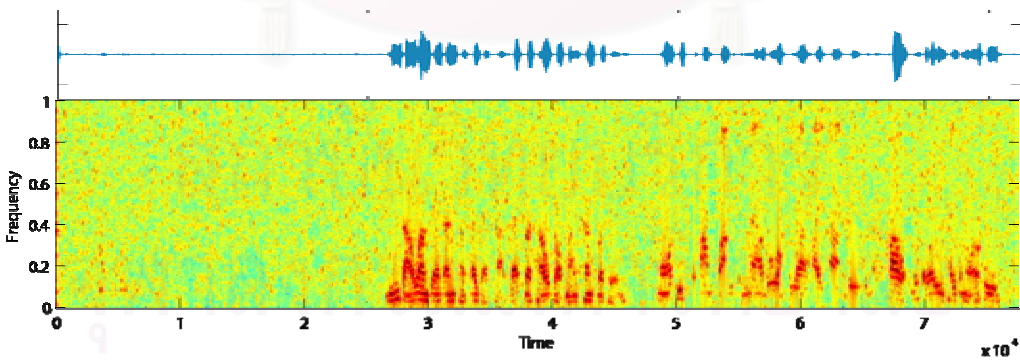




รูปที่ 14 รูปคลื่นและสเปกโตรแกรมของสัญญาณเสียงพูดทางฝั่งห้องใกล้ที่ปราศจากเสียงก่อกวน

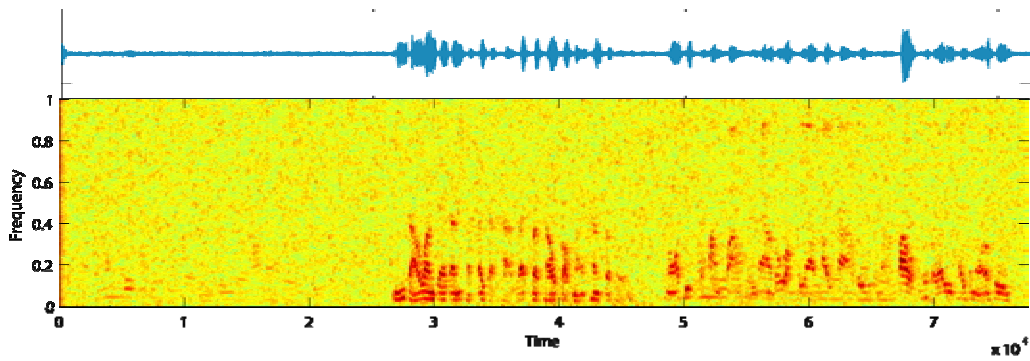


รูปที่ 15 รูปคลื่นและสเปกโตรแกรมของสัญญาณเสียงพูดที่ถูกก่อกวน(SNR = 0 dB)



รูปที่ 16 รูปคลื่นและสเปกโตรแกรมของสัญญาณเสียงพูดที่ถูกประมวลผลด้วยระบบ AENS

ที่ใช้กฎการลดทอนแบบวีเนอร์



รูปที่ 17 รูปคลื่นและสเปกโตรแกรมของสัญญาณเสียงพูดที่ถูกประมวลผลด้วยระบบ AENS

ที่ใช้กฎการลดทอนแบบ LSA

### 3.5 ผลการทดสอบระบบ AENR ที่พัฒนาขึ้นด้วยการสร้างจริงบนบอร์ด FPGA

สำหรับการทดสอบระบบ AENR ที่พัฒนาขึ้นในโครงการนี้บนสถานการณ์จริง ด้วยบอร์ด FPGA ยังติดปัญหาเรื่องเวลาของโครงการ จึงทำการสร้างจริง (Implementation) เฉพาะระบบ AEC เท่านั้น ด้วยบอร์ด FPGA ที่ได้พัฒนาแล้วบางส่วนจากเฟสก่อนหน้านี้ รุ่น DS-KIT-2VP7FF672-EURO โดยใช้โครงสร้างแบบ DA ร่วมกับ MAC เป็นวงจรมิติขนาด 32 แท็บ

ผลการสังเคราะห์วงจรแสดงดังในตารางที่ 8 โดยการสังเคราะห์วงจรได้ใช้วงจรรวม FPGA ของบริษัท Xilinx ในตระกูล Virtex-II Pro™ เบอร์ 2VP7FF672-6 จากผลการสังเคราะห์วงจรพบว่าได้ใช้ Slices ไป 2497 หรือคิดเป็น 50% ของจำนวน Slices ทั้งหมดที่มีอยู่ในวงจรรวม FPGA เบอร์นี้ และนอกจากนี้ยังได้ใช้ทรัพยากรอื่นๆ ภายในชิปซึ่งแสดงดังในตารางที่ 8 เมื่อคิดเป็นจำนวนเกตเสมือน (Equivalent gate) จะได้เท่ากับ 69,363 เกต

ในด้านของประสิทธิภาพวงจรในแง่ของเวลานั้นแสดงดังในตารางที่ 9 ซึ่งพบว่าโครงสร้างนี้สามารถทำงานได้ที่ความถี่สูงสุดเท่ากับ 62.636 MHz เนื่องจากวงจรมี Gate delay ต่ำสุด (Minimum period) เท่ากับ 15.965 nSec ส่วนค่า Minimum input arrival time before clock, Maximum output required time after clock และ Maximum combinational path delay เป็นค่าเวลา (ดีเลย์) มากสุดระหว่างอินพุตไปยังวงจรชนิดซีควนเชียล (Sequential Circuit) เวลามากที่สุดระหว่างวงจรชนิดซีควนเชียลไปยังเอาต์พุต และสุดท้ายเป็นเวลามากสุดระหว่างอินพุตไปยังเอาต์พุต ตามลำดับ ซึ่งโดยปกติแล้วเราจะสนใจที่ค่า Minimum period มากกว่าเนื่องจากเป็นค่า

ดีเลย์มากที่สุด (เมื่อเทียบกับค่าดีเลย์อื่นๆ) ที่เกิดขึ้นระหว่างสัญญาณนาฬิกาภายในวงจรชนิดที่  
ควมเร็ว

ตารางที่ 8 ผลการสังเคราะห์วงจรโดยใช้โครงสร้างแบบ DA ร่วมกับ MAC

Device utilization summary		
Number of Slices	2497 out of 4928	50%
Number of Slice Flip Flops	2998 out of 9856	30%
Number of 4 input LUTs	2974 out of 9856	30%
Number of bonded IOBs	49 out of 396	12%
Number of GCLKs	1 out of 16	6%
Total equivalent gate count for design : 69,363		

ตารางที่ 9 ประสิทธิภาพของวงจรในด้านเวลาของโครงสร้างแบบ DA ร่วมกับ MAC

Timing Summary
Minimum period: 15.965 ns (Maximum Frequency: 62.636 MHz)
Minimum input arrival time before clock: No path found
Maximum output required time after clock: 8.469 ns
Maximum combinational path delay: No path found

จากตารางที่ 10 และ 11 พบว่าวงจรที่สร้างใช้ทรัพยากรเกินกว่าที่มีอยู่ในชิป ได้แก่  
จำนวนของ Slice ซึ่งในชิปมีเพียง 4928 slice เท่านั้นในขณะที่วงจรใช้มากถึง 7427 slice ใน  
ทำนองเดียวกัน จำนวน Slice Flip Flop ก็ใช้มากถึง 10794 slice ซึ่งมากกว่าที่มีอยู่ในชิป ดังนั้น  
จำเป็นจะต้องปรับปรุงวงจรเพื่อให้ใช้ทรัพยากรให้น้อยลงโดยการลดชุดการประมวลผลลงเพื่อให้  
ใช้ทรัพยากรน้อยลง และทำการเพิ่มรอบในการประมวลผลให้มากขึ้นเพื่อให้วงจรยังคงมีขนาด  
128 แท้ปดั้งเดิม

ตารางที่ 10 ผลการสังเคราะห์วงจรในเชิงทรัพยากร

Device utilization summary : 2vp7ff672-6			
Number of Slices:	7427	out of 4928	150% (*)
Number of Slice Flip Flops:	10794	out of 9856	109% (*)
Number of 4 input LUTs:	8183	out of 9856	83%
Number of bonded IOBs:	37	out of 396	9%
Number of BRAMs:	3	out of 44	6%
Number of GCLKs:	3	out of 16	18%
WARNING:Xst:1336 - (*) More than 100% of Device resources are used			

ตารางที่ 11 ผลการสังเคราะห์วงจรในเชิงเวลา

Timing Summary: Speed Grade: -6
Minimum period: 21.142ns (Maximum Frequency: 47.300MHz)
Minimum input arrival time before clock: 5.244ns
Maximum output required time after clock: 10.089ns
Maximum combinational path delay: No path found

ทั้งนี้ วงจรที่สร้างใช้ทรัพยากรมากเกินไป ทำให้ไม่สามารถโปรแกรมลงบนชิปได้ ดังนั้นจึงได้ปรับปรุงวงจรใหม่โดยการลดจำนวนชุดของการประมวลผลลงเป็น 16 ชุด และใช้จำนวนรอบในการประมวลผลเป็น 8 รอบ เพื่อให้ได้วงจรขนาด 128 แท็บ (16X8 = 128) ซึ่งจะทำให้

ให้วงจรใช้ทรัพยากรน้อยลง หากแต่วงจรดังกล่าวยังไม่เสร็จเนื่องจากในการเปลี่ยนจำนวนรอบของการประมวลผลจำเป็นต้องสร้างวงจรควบคุมการทำงานของวงจรทั้งหมดใหม่ โดยวงจรควบคุมการทำงานนี้ได้สร้างในรูปแบบของ สเตตแมชชีน (State Machine) ซึ่งในการเปลี่ยนรอบของการประมวลผลจะทำให้สเตต (State) ของวงจรเปลี่ยนไป ดังนั้นจะต้องทำการออกแบบสร้างวงจรควบคุมให้ดีขึ้น



## สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย



## บทที่ 4

### สรุปและวิจารณ์ผลการวิจัย

#### 4.1 สรุปผลการวิจัย

ในโครงการวิจัยนี้ การปรับปรุงคุณภาพของสัญญาณเสียงพูดที่ส่งออกไปบนเครือข่ายการสื่อสารทางเสียงที่มีหนึ่งช่องสัญญาณ โดยมุ่งเน้นที่การตัดสัญญาณเสียงสะท้อนและสัญญาณรบกวน สามารถใช้ระบบ AENS ในการตัดสัญญาณก่อกวนทั้งสองนี้ได้อย่างพร้อมๆ กัน โดยมีประสิทธิภาพใกล้เคียงกับการใช้งานของระบบ AEC และระบบ NR 2 ระบบต่อกัน

เมื่อเปรียบเทียบประสิทธิภาพการทำงานของระบบ AENS ด้วยวิธีการประมาณสเปกตรัมกำลังที่นำเสนอแล้ว พบว่าให้ผลการลด / กำจัดสัญญาณเสียงสะท้อนได้อย่างใกล้เคียงกับมาตรฐาน ITU Standard G.165 ที่กำหนดไว้ นอกจากนี้ คุณภาพของสัญญาณเสียงที่ส่งออกไปในระบบมีความชัดเจนมากขึ้น ดังเห็นได้จากผลการจำลองและทดสอบระบบ

#### 4.2 ผลิตผลและหรือความสัมฤทธิ์ผลของงานที่ได้ดำเนินการไปแล้ว

4.2.1 โปรแกรมภาษา MATLAB ในการจำลองการทำงานของระบบ AENS

4.2.2 โปรแกรมภาษา VHDL ในการควบคุมการทำงานของบอร์ด FPGA สำหรับจำลองระบบ AEC

4.2.3 บทความทางวิชาการ จำนวน – เรื่อง

(ยังอยู่ในระหว่างการร่างบทความทางวิชาการ ยังไม่ได้ส่งตีพิมพ์ใดๆ)

#### 4.3 งบประมาณเฉพาะหมวดค่าตอบแทนและหมวดวัสดุและค่าใช้จ่าย

ได้รับจัดสรรงบประมาณในหมวดนี้ เป็นเงินจำนวน 700,000 บาท ดังรายละเอียดต่อไปนี้

- หมวดครุภัณฑ์	-	บาท
- หมวดวัสดุ / หมวดค่าใช้สอย	365,000	บาท
- ผู้ช่วยวิจัย 4 คน	<u>335,000</u>	บาท
<b>รวมทั้งสิ้น</b>	<b>700,000</b>	<b>บาท</b>



#### 4.4 ข้อเสนอแนะและงานที่จะดำเนินการต่อไปในอนาคต

- หากใช้ความรู้จากจิตวิทยาทางเสียง (Psychoacoustic) ในด้านการบังทางความถี่ (Frequency Masking) เป็นต้น มาร่วมกับระบบ AENS ในการลดผลกระทบของสัญญาณเสียงสะท้อนและสัญญาณรบกวน คุณภาพของเสียงพูดที่ถูกปรับปรุงควรจะเป็นที่ดีขึ้นอีก กล่าวคือ ลดความผิดเพี้ยนของสัญญาณเสียงพูดลงได้กว่าที่เป็นอยู่
- เนื่องจากเวลาที่จำกัด ในโครงการนี้จึงยังไม่สามารถทำการทดสอบคุณภาพสัญญาณเสียงที่ถูกปรับปรุงโดยวิธีที่พัฒนาขึ้นด้วย Subjective Listening Test เพื่อยืนยันผล Objective เหล่านี้



สถาบันวิทยบริการ  
จุฬาลงกรณ์มหาวิทยาลัย

## เอกสารอ้างอิง

- [1] J. Benesty, S. Makino and J. Chen, *Speech Enhancement*. Springer, 2005.
- [2] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-26, pp. 113-120, Apr. 1979.
- [3] R. J. McAulay and M. L. Malpass, "Speech enhancement using a soft-decision noise suppression filter," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, pp. 137-145, Apr. 1980.
- [4] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-32, pp. 1109-1121, Dec. 1984.
- [5] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, no. 2, pp. 443-445, Apr. 1985.
- [6] O. Cappe, "Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor," *IEEE Trans. Speech and Audio Processing*, vol. 2, no. 2, pp. 345-349, Apr. 1994.
- [7] D. Wang and J. Lim, "The unimportance of phase in speech enhancement," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 30, no. 4, pp.679-681, Aug. 1982.
- [8] R. Martin, "Speech enhancement based on minimum mean-square error estimation and supergaussian priors," *IEEE Trans. Speech and Audio Processing*, vol. 13, no. 5, part 2, pp. 845-856, Sept. 2005.
- [9] T. Lotter and P. Vary, "Speech enhancement by MAP spectral amplitude estimation using a super-gaussian speech model," *EURASIP Journal on Applied Signal Processing*, vol. 2005, no. 7, pp. 1110-1126, 2005.
- [10] J. Patrick, J. Godsill Wolfe and Simon, "Efficient alternatives to the Ephraim and Malah suppression rule for audio signal enhancement," *EURASIP Journal on Applied Signal Processing*, vol. 2003, no. 10, pp. 1043-1051, 2003.

- [11] I. Cohen and B. Berdugo, "Speech enhancement using super-Gaussian models and noncausal a priori SNR estimation," *Signal Processing*, vol. 47, pp. 336-350, Feb. 2005.
- [12] I. Cohen, "Modeling speech signals in the time-frequency domain using GARCH," *Signal Processing*, vol. 84, pp. 2453-2459, Dec. 2004.
- [13] R. Martin, "Noise power spectral density estimation based on optimal and minimum statistics," *IEEE Trans. Speech and Audio Processing*, vol. 9, no. 5, pp. 504-512, Jul. 2001
- [14] M. Kato, M. Serizawa, N. Toki, U. Mori, Y. Morishita and K. Hayashi, "Noise suppression with high speech quality based on weighted noise estimation for 3G handsets," *NEC Res. & Develop.*, vol. 44, pp. 66-73, Oct. 2003.
- [15] Y. Ephraim and H.L. Van Trees, "A signal subspace approach for speech enhancement," *IEEE Trans. Speech and Audio Processing*, vol. 3, no. 4, pp. 251-266, Jul. 1995.
- [16] B. Widrow, U.S. Patent No 2005075866, filed 28 Nov. 2004, 7 Apr. 2005.
- [17] N. Sasaoka, K. Sumi, Y. Itoh and K. Fujji, "A study on noise reduction system based on ALE and noise reconstruction filter," in *Proc., ISPACS 2004, Intelligent Signal Processing and Communication Systems*, pp. 305-308, Nov. 2004.
- [18] B. Widrow et al., "Adaptive noise cancelling: Principles and applications," in *Proc. IEEE*, vol.63, pp. 1692 - 1716, Dec. 1975.
- [19] I. A. McCowan, "Robust Speech Recognition using Microphone Arrays," PhD Thesis, Queensland University of Technology, Australia, Apr. 2001.
- [20] S. Haykin, *Unsupervised Adaptive Filtering*. John Wiley & Sons, 2000.
- [21] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, 3<sup>rd</sup> edition, 1996.
- [22] E. Hansler, "Adaptive echo compensation applied to the hands-free telephone problem," in *Proc IEEE int. Circuits and Systems*, vol. 1, pp. 279-282, May 1990.
- [23] K. Ochiai, T. Araseki and T. Ogihara, "Echo canceller with two echo path models," *IEEE Trans. Commun.*, vol. 25, no. 6, pp. 589-595, Jun. 1977.

- [24] G.-O. Glentis, K. Berberidis and S. Theodoridis, "Efficient least squares adaptive algorithms for FIR transversal filtering," *IEEE Signal Processing Magazine*, vol. 16, no. 4, pp. 13-41, Jul. 1999.
- [25] J. J. Shynk, "Frequency-domain and multirate adaptive filtering," *IEEE Signal Processing Magazine*, vol. 9, no. 1, pp. 14-37, Jan. 1992.
- [26] S. L. Gay, "A fast converging, low complexity adaptive filtering algorithm," *IEEE Workshop, Application of Signal Processing to Audio and Acoustics*, pp. 4-7, Oct. 1993.
- [27] R. Martin and P. Vary, "Combined Acoustic Echo Control and noise reduction for hands-free telephony state of the art and perspectives," in *Proc. EUSIPCO '96*, pp. 1107-1110, 1996.
- [28] F. Capman, J. Boudy, P. Lockwood, "Acoustic Echo Cancellation and Noise Reduction in the Frequency-Domain: A Global Optimisation," in *Proc. EUSIPCO '96*, Sept. 1996.
- [29] R. L. B. Jeannes, P. Scalart, G. Faucon and C. Beaugeant, "Combined noise and echo reduction in hands-free systems: A survey," *IEEE Trans. Speech and Audio Processing*, vol. 9, no. 8, Nov. 2001.
- [30] Y. Guelou, A. Benamar and P. Scalart, "Analysis of two structures for combined acoustic echo cancellation and noise reduction," in *Proceedings, ICASSP-96, Acoust., Speech, and Signal Processing*, pp. 637-340, May 1996.
- [31] B. Ayad, G. Faucon and R. Bouquin-Jeannes, "Optimization of noise reduction preprocessing in an acoustic echo and noise controller," in *Proc. ICASSP-96, Acoust., Speech, and Signal Processing*, pp. 953-956, May 1996.
- [32] S. Gustafsson, R. Martin and P. Vary, "Combined acoustic echo control and noise reduction for hands-free telephony," *Signal Processing*, vol. 64, pp. 21-32, Jan. 1998.
- [33] S. Gustafsson, R. Martin, P. Jax and P. Vary, "A psychoacoustic approach to combined acoustic echo cancellation and noise reduction," *IEEE Trans. Speech and Audio Processing*, vol. 10, no. 5, pp. 245-256, Jul. 2002.

- [34] C. Beaugeant, V. Turbin, P. Scalart and A. Gilloire, "New optimal filtering approaches for hands-free telecommunication terminals," *Signal Processing*, vol. 64, pp. 33-47, Jan. 1998.
- [35] C. Avendano, "Acoustic echo suppression in the STFT domain," *IEEE Workshop, Application of Signal Processing to Audio and Acoustics*, pp 175-178, Oct. 2001.
- [36] C. Faller and J. Chen, "Suppression acoustic echo in a spectral envelope space," *IEEE Trans. Speech and Audio Processing*, vol. 13, no. 5, pp. 1048-1062, Sept. 2005.
- [37] A. Gilloire, P. Scalart and C. Lamblin, C. Mokbel and S. Proust, "Innovative speech processing for mobile terminals: an annotated bibliography," *Signal Processing*, vol. 80, pp. 1149-1166, Jul. 2000.
- [38] ITU-T Recommendation G.167



## ภาคผนวก ก

### รายชื่อผู้ทรงคุณวุฒิที่สามารถประเมินโครงการวิจัยนี้

- ผศ. ดร. พีระพล ยุวภูษิตานนท์

ภาควิศวกรรมไฟฟ้าอิเล็กทรอนิกส์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีมหานคร

- ผศ. ดร. เผ่าศักดิ์ ศิริสุข

ภาควิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีมหานคร



สถาบันวิทยบริการ  
จุฬาลงกรณ์มหาวิทยาลัย