

การปรับปรุงคุณภาพเสียงโดยใช้วิธีการหักล้างทางสเปกตรัม

SPEECH ENHANCEMENT USING SPECTRAL SUBTRACTION



สุเมธี นพากาศ
SUMATHEE NAPAKAS

เลขหม.....
เลขทะเบียน **45651**
วัน, เดือน, ปี **1 2 ก.พ. 2546**

b.....
i.....

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมไฟฟ้า
บัณฑิตวิทยาลัย
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
พ.ศ. 2546

SPEECH ENHANCEMENT USING SPECTRAL SUBTRACTION



**A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENT FOR THE DEGREE OF
MASTER OF ENGINEERING IN ELECTRICAL ENGINEERING
SCHOOL OF GRADUATE STUDIES
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG**

2003

ISBN 974-324-181-7

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



COPYRIGHT 2003

SCHOOL OF GRADUATE STUDIES

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ไว้เพื่อการวิจัยเท่านั้น เมื่ออนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อวิทยานิพนธ์	การปรับปรุงคุณภาพเสียง โดยใช้วิธีการหักล้างทางสเปกตรัม
นักศึกษา	นางสาวสุเมธี นภากาศ
รหัสประจำตัว	42061152
ปริญญา	วิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชา	วิศวกรรมไฟฟ้า
พ.ศ.	2546
อาจารย์ผู้ควบคุมวิทยานิพนธ์	รศ.ดร.ยุทธพงษ์ รังสรรค์เสรี

บทคัดย่อ

วิทยานิพนธ์นี้เป็นการศึกษาเทคนิคการปรับแต่งคุณภาพเสียง โดยใช้วิธีการหักล้างทางสเปกตรัม ซึ่งอัลกอริทึมของการปรับแต่งคุณภาพเสียงได้นำเสนอ การลดสัญญาณเสียงรบกวนในเสียงพูด โดยการใช้เทคนิคการหักล้างทางสเปกตรัมในการจัดสัญญาณรบกวนออกจากเสียงพูด เพื่อให้ได้ค่าอัตราส่วนสัญญาณต่อสัญญาณรบกวน (Signal to noise ratio) ของสัญญาณเสียงสูงขึ้น การกำจัดสัญญาณรบกวนออกจากเสียงพูดโดยใช้วิธีการหักล้างทางสเปกตรัมนี้ สัญญาณรบกวนจะประมาณได้จากช่วง non-speech ของสัญญาณ โดยที่เฟสของสัญญาณที่ผิดเพี้ยนจะถูกเก็บไว้ก่อน ในอัลกอริทึมนี้จะอธิบายถึงตัวอย่างของสัญญาณอินพุตที่ถูกซัดตัวอย่าง 8 kHz และมีโมเดลของสัญญาณรบกวนแบบ White Gaussian เสียงพูดที่ผ่านตัวกรองความถี่ต่ำและการแปลงสัญญาณจากอนาลอกเป็นดิจิตอลจะถูกวิเคราะห์ด้วยการวินโดว์ข้อมูล จากบัพเฟอร์ข้อมูลอินพุตแบบ Half-Overlap ส่วนขนาดของข้อมูลที่ถูกวินโดว์แล้ว จะถูกคำนวณโดยใช้ DFT และสเปกตรัมของสัญญาณรบกวนจะถูกคำนวณระหว่างช่วง non-speech การกำจัดสัญญาณรบกวน จะคำนวณจากการหักล้างสเปกตรัมของสัญญาณรบกวน ออกจากสเปกตรัมของสัญญาณที่ผิดเพี้ยน ผลลัพธ์ที่ได้ ถ้าเป็นค่าลบจะถูกปรับค่าให้เป็นศูนย์ และทำการคำนวณ IDFT เป็นลำดับสุดท้าย ผลลัพธ์ของสัญญาณเสียงแสดงผล โดยการพล็อตรูปคลื่นและสเปกโตรแกรม จากการวัดประสิทธิภาพและการฟังเสียง แสดงให้เห็นว่าวิธีการปรับแต่งคุณภาพเสียง โดยใช้วิธีการหักล้างทางสเปกตรัม ให้ผลลัพธ์เป็นที่น่าพอใจ

Thesis Title	Speech enhancement using spectral subtraction
Student	Miss.Sumathee Napakas
Student ID.	42061152
Degree	Master of Engineering
Programme	Electrical Engineering
Year	2003
Thesis Advisor	Assoc.Prof.Dr.Yuttapong Rangsanseri

ABSTRACT

This thesis presents a study of speech enhancement using spectral subtraction technique. An enhancement of speech signal algorithm is presented for reducing the spectral effects of acoustically added noise in speech. The spectral subtraction technique is used to remove noise components from noisy speech in order to improve the signal to noise ratio of the signal. Suppression stationary noise from speech by subtracting the noise spectrum is obtained from the signal measured during non-speech activity while the phase of noisy speech was kept. In this algorithm, the implementation described is designed for an input sample rate of 8 kHz and the background disturbance are modeled as additive white Gaussian noise. Speech, suitably low pass filtered and digitized, is analyzed by windowing data from half-overlapped input data buffers. The magnitude spectra of the windowed data are calculated using discrete Fourier transform and the spectral noise bias is calculated during non-speech activity. Suppression noise by subtracting the spectral noise from the spectral noisy. Resulting negative amplitudes are then zeroed out. At last compute the inverse discrete Fourier transform. Performance assessment based on the time waveform plots, objective measures and informal subjective listening tests all indicate that our method gives consistently good results.

กิตติกรรมประกาศ

ขอขอบพระคุณ บิดามารดา ที่ให้ความสนับสนุนและให้โอกาสในการศึกษา ให้ผู้เขียนได้ทำในสิ่งที่ตั้งใจไว้ จนถึงวันนี้

ขอขอบพระคุณ รศ.ดร.บุทรพงษ์ รั้งสรรค์เสรี เป็นอย่างสูงที่ให้คำแนะนำ ให้ความรู้ และให้คำปรึกษา ตั้งแต่เริ่มเข้ามาศึกษา จนกระทั่งสำเร็จการศึกษา

ขอขอบคุณทุนสนับสนุนการทำวิทยานิพนธ์ จากบัณฑิตวิทยาลัย สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

และขอขอบคุณทุกท่านที่ให้การช่วยเหลือจนทุกอย่างสำเร็จลุล่วงไปได้ด้วยดี แม้ไม่ได้กล่าวถึง ณ ที่นี้

ขอขอบพระคุณ

สุเมธ นภาภาส



สารบัญ

หน้า

บทคัดย่อภาษาไทย.....	I
บทคัดย่อภาษาอังกฤษ.....	II
กิตติกรรมประกาศ.....	III
สารบัญ.....	IV
สารบัญตาราง.....	VI
สารบัญรูป.....	VII
บทที่ 1 บทนำ.....	1
1.1 วัตถุประสงค์ของวิทยานิพนธ์.....	1
1.2 ขอบเขตของการวิจัย.....	1
บทที่ 2 เสียงและสัญญาณรบกวน.....	3
2.1 กายวิภาคของระบบการแปลงเสียงของมนุษย์.....	3
2.2 คุณสมบัติพื้นฐานของเสียงพูด.....	8
2.3 การแสดงสัญญาณรบกวนทางคณิตศาสตร์.....	17
2.4 การนิยามหน่วยเดซิเบล.....	28
2.5 การวัดคุณภาพของเสียง.....	30
บทที่ 3 ทฤษฎีที่เกี่ยวข้อง.....	31
3.1 การแปลงสัญญาณอนาล็อกเป็นดิจิทัล.....	31
3.2 การแปลงสัญญาณแบบ DFT และ FFT.....	36
3.3 การวิเคราะห์สเปกตรัมเวลาสั้นๆ.....	46
3.4 การหากล้างทางสเปกตรัม.....	51
บทที่ 4 วิธีที่นำเสนอในการปรับแต่งคุณภาพเสียง.....	65
4.1 การอธิบายทางคณิตศาสตร์.....	65
4.2 ขั้นตอนการทำงานด้วยโปรแกรมคอมพิวเตอร์.....	68

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ(ต่อ)

	หน้า
บทที่ 5 ผลการทดลอง.....	71
5.1 ผลการทดลองกับสัญญาณเสียงที่จำลองขึ้น.....	71
5.2 ผลการทดลองกับสัญญาณเสียงที่ใช้งานจริง.....	72
บทที่ 6 บทสรุปผลการวิจัย.....	89
บรรณานุกรม.....	90
ภาคผนวก.....	91
ประวัติผู้เขียน.....	96



สารบัญตาราง

ตารางที่	หน้า
2.1 ความถี่ฟอร์แมนท์ของเสียงสระชาวญี่ปุ่น 5 เสียง.....	14
2.2 การแบ่งแยกเสียงพยัญชนะของชาวญี่ปุ่น	15
2.3 ความสัมพันธ์ที่แตกต่างของคุณลักษณะ หน่วยพื้นฐานของเสียงภาษาอังกฤษ.....	16
2.4 อัตราเวลาของเสียงสนทนาสำหรับระดับความเร็วของการพูด.....	16
3.1 การแปลงแบบต่างๆ กับคุณลักษณะของสัญญาณที่เกี่ยวข้อง.....	37
3.2 โครงสร้างของการแปลงแบบต่างๆ	38
5.1 ค่า SNR ของสัญญาณเสียงพูดประโยค “หนึ่ง สอง สาม สี่ ห้า” ก่อนและหลังผ่าน กระบวนการปรับแต่งคุณภาพเสียงด้วยวิธีการหักล้างทางสเปกตรัม.....	87
5.2 ค่า SNR ของสัญญาณเสียงพูดประโยค “เอ็ม ไอ ที” ก่อนและหลังผ่านกระบวนการ ปรับแต่งคุณภาพเสียงด้วยวิธีการหักล้างทางสเปกตรัม	88

สารบัญรูป

รูปที่	หน้า
2.1 ภาพตัดขวางแสดงอวัยวะในระบบการพูดของมนุษย์	3
2.2 รูปกล่องเสียงขณะ (ก) หายใจปกติ (ข) หายใจเข้าลึกๆ (ค) กำลังส่งเสียง (ง) ส่งเสียงกระซิบ หรือเสียงแผ่ว.....	4
2.3 แผนภาพระบบเสียงพูดของมนุษย์.....	6
2.4 (ก) ตัวอย่างรูปคลื่นของสัญญาณเสียงก้อง /a/	7
2.4 (ข) ตัวอย่างรูปคลื่นของสัญญาณเสียงไม่ก้อง /sh/.....	7
2.5 แผนภาพกรอบจำลองระบบกำเนิดเสียงเริ่มต้น.....	8
2.6 รูปคลื่นของเสียงพูด	8
2.7 การกระจายตัวที่เพิ่มขึ้นของแอมพลิจูดสัญญาณเสียงพูด.....	9
2.8 ความถี่สเปกตรัมของเสียงพูด.....	11
2.9 ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของความถี่ค่าพิทช์ของเสียง.....	11
2.10 (ก) ความถี่ฟอร์แมนท์ที่หนึ่งและที่สองของเสียงชาวญี่ปุ่น 5 เสียง.....	13
2.10 (ข) ความถี่ฟอร์แมนท์ที่หนึ่งและที่สองของเสียงชาวอเมริกัน 10 เสียง.....	13
2.11 ความเหมือนระหว่างการพูดออกเสียงพยัญชนะและสเปกตรัม	15
2.12 ไดโอด.....	17
2.13 รูปแบบของรีซีสเตอร์ที่มีสัญญาณรบกวน.....	19
2.14 คุณสมบัติของไวท์นอยส์.....	20
2.15 (ก) รูปร่างของแชนเนลสัญญาณรบกวน.....	21
2.15 (ข) รูปร่างสัญญาณมีคาบเวลาที่ได้จาก (ก) ในช่วง $-T/2$ ถึง $+T/2$	21
2.16 สเปกตรัมกำลังของ $(n_i^{(s)}(t))$	23
2.17 (ก) เกาส์เซียนนอยส์ $n_i(t)$	25
2.17 (ข) พัลเซอร์จากเกาส์เซียนนอยส์ $n_i(t)$	25
2.18 การคิดกำลังสัญญาณในรูปเดซิเบล.....	29
3.1 กระบวนการแชนเปลลิง	32
3.2 การแชนเปลลิงที่คิดเพี้ยน.....	33
3.3 กระบวนการควอนไทซ์	35
3.4 แผนภาพของกระบวนการประมวลผลสัญญาณดิจิทัล.....	36
3.5 ความเกี่ยวข้องกันระหว่างการแปลง DFT กับการแปลงฟูเรียร์.....	40

สารบัญรูป(ต่อ)

รูปที่	หน้า
3.6 ผลของการเติมศูนย์กับการแปลง DFT	41
3.7 การวิเคราะห์สเปกตรัมในช่วงสั้นๆ ของวินโดว์.....	47
3.8 ฟังก์ชันของหน้าต่างเวลา.....	49
3.9 ค่าที่ลดลงด้วยการวินโดว์ และการวิเคราะห์แบ่งออกเป็น 2 ส่วน.....	50
3.10 การวิเคราะห์การเลื่อมอย่างต่อเนื่อง	50
3.11 การวิเคราะห์เสียงสระ 5 เสียง /a, i, u, e, o/.....	51
3.12 สัญญาณรบกวนในโดเมนเวลาและความถี่.....	53
3.13 บล็อกโคออร์ดิเนตของกระบวนการหักล้างทางสเปกตรัม.....	53
3.14 ความผิดเพี้ยนที่อาจจะเกิดขึ้น ได้จากกระบวนการหักล้างทางสเปกตรัม.....	58
3.15 ความผิดเพี้ยนที่อาจจะเกิดขึ้น ได้จากกระบวนการหักล้างทางสเปกตรัม ในรูปขนาด สเปกตรัมของสัญญาณ.....	59
3.16 วิธีการบ่งชี้ตัวกรองสำหรับ musical noise.....	61
3.17 ผลกระทบของการหักล้างทางสเปกตรัมแบบไม่เป็นเชิงเส้น.....	64
4.1 กระบวนการกำจัดสัญญาณรบกวนโดยวิธีการหักล้างทางสเปกตรัม.....	69
5.1 รูปคลื่นของสัญญาณเสียงพูด “หนึ่ง สอง สาม สี่ ห้า” SNR 10 dB	73
5.2 สเปกโตรแกรมของสัญญาณเสียงพูด “หนึ่ง สอง สาม สี่ ห้า” SNR 10 dB.....	74
5.3 รูปคลื่นของสัญญาณเสียงพูด “หนึ่ง สอง สาม สี่ ห้า” SNR 5 dB	75
5.4 สเปกโตรแกรมของสัญญาณเสียงพูด “หนึ่ง สอง สาม สี่ ห้า” SNR 5 dB.....	76
5.5 รูปคลื่นของสัญญาณเสียงพูด “เอ็ม ไอ ที” SNR 10 dB.....	77
5.6 สเปกโตรแกรมของสัญญาณเสียงพูด “เอ็ม ไอ ที” SNR 10 dB	78
5.7 รูปคลื่นของสัญญาณเสียงพูด “เอ็ม ไอ ที” SNR 5 dB.....	79
5.8 สเปกโตรแกรมของสัญญาณเสียงพูด “เอ็ม ไอ ที” SNR 5 dB	80
5.9 รูปคลื่นของสัญญาณเสียงพูดจากวิทยุ FM “นั่นคือกีฬาประเภทไหน ระหว่าง”... ..	81
5.10 สเปกโตรแกรมของสัญญาณเสียงพูดจากวิทยุ FM “นั่นคือกีฬาประเภทไหน ระหว่าง”.....	82
5.11 รูปคลื่นของสัญญาณเสียงพูดจากวิทยุ FM “อยากตัดคำว่าวัดออก”.....	83
5.12 สเปกโตรแกรมของสัญญาณเสียงพูดจากวิทยุ FM “อยากตัดคำว่าวัดออก”.....	84
5.13 รูปคลื่นของสัญญาณเสียงพูดจากวิทยุ FM “ส่วนใหญ่ก็เป็นชาวพุทธที่เรียนอยู่”.....	85
5.14 สเปกโตรแกรมของสัญญาณเสียงพูดจากวิทยุ FM “ส่วนใหญ่ก็เป็นชาวพุทธที่เรียนอยู่”.....	86

สารบัญรูป(ต่อ)

รูปที่	หน้า
5.15 กราฟค่า SNR ของสัญญาณเสียงพูดประโยค “หนึ่ง สอง สาม สี่ ห้า”.....	87
5.16 กราฟค่า SNR ของสัญญาณเสียงพูดประโยค “เอ็ม ไอ ที”.....	88



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1

บทนำ

1.1 วัตถุประสงค์ของวิทยานิพนธ์

เสียงพูดที่มีสัญญาณรบกวนนั้น เป็นสิ่งที่ไม่ต้องการ ดังนั้นการกำจัดสัญญาณรบกวนออกจากเสียงพูด เป็นสิ่งที่มีความสำคัญมาก เพราะจะทำให้เสียงนั้นมีคุณภาพดีขึ้น เหมาะสมสำหรับการนำไปใช้งานต่อไป ซึ่งเทคนิคการกำจัดสัญญาณรบกวนในเสียงพูดนั้น ได้มีการศึกษากันอย่างกว้างขวางถึงวิธีการต่างๆ และสามารถนำไปประยุกต์ใช้ในระบบการสื่อสารได้อย่างมากมาย เช่น การกำจัดสัญญาณรบกวนจากเสียงพูดที่ออกจากไมโครโฟน

ลักษณะการเกิดสัญญาณรบกวนในรอบการสื่อสารนั้น สามารถเกิดขึ้นได้หลายจุด ตั้งแต่ตอนสร้างสัญญาณ และขยายสัญญาณที่ต้นทาง หรือระหว่างส่งมาตามสายนำสัญญาณ และแม้แต่ที่ภาครับสัญญาณของระบบ เมื่อผลกระทบของสัญญาณรบกวนมากขึ้น ทำให้ขนาดของสัญญาณจริงเล็กลงตามลำดับ

ดังนั้นวิทยานิพนธ์ฉบับนี้ จึงได้ทำการวิจัยเพื่อหาวิธีการกำจัดสัญญาณรบกวนแบบเกาส์เซียน ที่เกิดขึ้นในเสียงพูด โดยวิธีการที่ใช้ในการปรับแต่งคุณภาพเสียงครั้งนี้ใช้เทคนิควิธีการหักล้างทางสเปกตรัม ซึ่งคาดว่าเทคนิควิธีการที่ใช้ในงานวิจัยนี้ จะเป็นประโยชน์ต่องานวิจัย และการพัฒนาของงานในสาขาต่างๆ ที่เกี่ยวข้องกัน

1.2 ขอบเขตการวิจัย

งานวิจัยนี้จะนำเสนอถึงเทคนิควิธีการปรับแต่งคุณภาพเสียงโดยใช้วิธีหักล้างทางสเปกตรัม ซึ่งเทคนิควิธีการดังกล่าว จะช่วยแก้ปัญหาการกำจัดสัญญาณรบกวนของสัญญาณรบกวนแบบเกาส์เซียนที่เกิดขึ้นในเสียงพูดที่ให้ผลของคุณภาพของเสียงพูดที่ดีขึ้น โดยใช้อัตราส่วนสัญญาณต่อสัญญาณรบกวน (signal to noise ratio) เป็นดัชนีชี้วัดคุณภาพของเสียงพูด รายละเอียดของการวิจัยของวิทยานิพนธ์ฉบับนี้ จะแบ่งย่อยออกเป็น 6 บท โดยที่ในแต่ละบทจะมีหัวข้อและเนื้อหาที่จะกล่าวต่อไปนี้

- บทที่ 1 เป็นบทนำ กล่าวถึงวัตถุประสงค์ และขอบเขตของวิทยานิพนธ์
- บทที่ 2 เป็นการกล่าวถึง สัญญาณเสียง และสัญญาณรบกวน ซึ่งจะอธิบายถึงกายวิภาคของระบบการแปลงเสียงของมนุษย์ คุณลักษณะพื้นฐานของเสียงพูด ส่วนสัญญาณรบกวนจะเสนอรายละเอียดของสัญญาณรบกวนทางคณิตศาสตร์ และอธิบายถึงการวัดคุณภาพเสียง
- บทที่ 3 เป็นการกล่าวถึง ทฤษฎีที่เกี่ยวข้อง คือ การแปลงสัญญาณอนาลอกเป็นดิจิทัล การแปลงสัญญาณแบบ DFT และ FFT การวิเคราะห์สเปกตรัมเวลาสั้นๆ การหักล้างทางสเปกตรัม ซึ่งอธิบายถึงโมเดลของสัญญาณเสียงที่ผิดเพี้ยน กำลังของการหักล้างทางสเปกตรัม ขนาดของการหักล้างทางสเปกตรัม ตัวกรองการหักล้างทางสเปกตรัม รวมถึงผลกระทบที่จะเกิดจากการหักล้างทางสเปกตรัม และการหักล้างทางสเปกตรัมแบบไม่เป็นเชิงเส้น
- บทที่ 4 เป็นการกล่าวถึง การนำทฤษฎีของกระบวนการหักล้างทางสเปกตรัมมาประยุกต์ใช้ในการกำจัดสัญญาณรบกวนในเสียงพูด และจำลองกระบวนการทำงานด้วยโปรแกรมคอมพิวเตอร์ โดยกล่าวถึงกระบวนการต่างๆ ที่เกี่ยวข้องซึ่งประกอบด้วย โมเดลการปรับสัญญาณรบกวน การวินโดว์ข้อมูล การประมาณค่าสัญญาณรบกวน และการทำ Half-wave rectification และกล่าวถึง การทำงานและองค์ประกอบของแบบจำลองการทำงานของการปรับแต่งคุณภาพเสียงด้วยโปรแกรมคอมพิวเตอร์ และแสดงขั้นตอนของโปรแกรมการกำจัดสัญญาณรบกวนในเสียงพูด ด้วยวิธีการหักล้างทางสเปกตรัม
- บทที่ 5 เป็นผลการทดลองที่แสดงให้เห็นถึงประสิทธิภาพของกระบวนการกำจัดสัญญาณรบกวน ด้วยเทคนิควิธีการหักล้างทางสเปกตรัม ซึ่งแบ่งออกเป็น 2 ส่วน คือ ผลการทดลองกับสัญญาณเสียงที่จำลองขึ้น และ ผลการทดลองกับสัญญาณเสียงที่ใช้งานจริง
- บทที่ 6 เป็นบทสรุป และการวิจารณ์ผลการทดลองที่ได้จากบทที่ 5

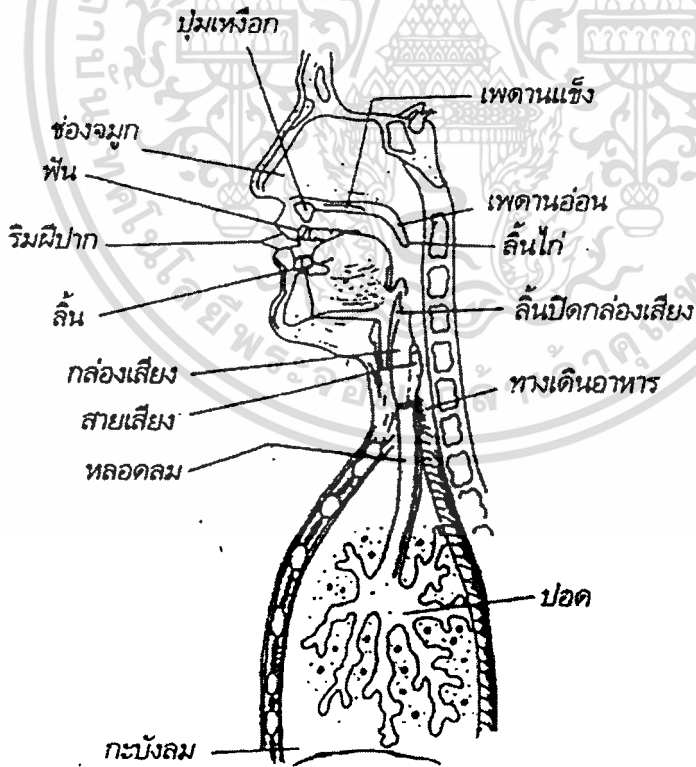
บทที่ 2

เสียงและสัทวิทยาบทกวี

2.1 กายวิภาคของระบบการเปล่งเสียงของมนุษย์

จากการศึกษาด้าน กายวิภาคศาสตร์ของมนุษย์ (Human anatomy) วิชาที่ว่าด้วยเสียงของภาษา (Phonetics) และศาสตร์ทางด้านเสียง (acoustics) ช่วยให้เข้าใจขั้นตอนการทำงานร่วมกันของอวัยวะต่างๆ ในการเปล่งเสียงพูดตลอดจนลักษณะทางกายภาพของเสียงพูดเพื่อนำวิเคราะห์เสียงพูดของมนุษย์ [1]

การพูดของมนุษย์มีใซ่อาการที่เกิดเฉพาะที่ปากเท่านั้น หากเริ่มจากลมหายใจเข้าของมนุษย์เองที่นำลมเข้าสู่ปอด จากนั้นจะใช้ลมจากปอดซึ่งก็คือลมหายใจออก มาทำให้เกิดเสียงพูด โดยลมจะถูกบังคับให้ผ่านอวัยวะต่างๆ ที่สำคัญ คือ เส้นเสียงซึ่งอยู่ในช่องของหลอดลม หรือบริเวณที่เรียกว่าลูกกระเดือก แล้วลมนั้นจะผ่านต่อออกมาทางช่องปากหรือช่องจมูก อวัยวะในช่องปากก็จะดัดแปลงให้เป็นเสียงต่างๆ ตามที่ผู้พูดต้องการ



รูปที่ 2.1 ภาพตัดขวางแสดงอวัยวะในระบบการพูดของมนุษย์

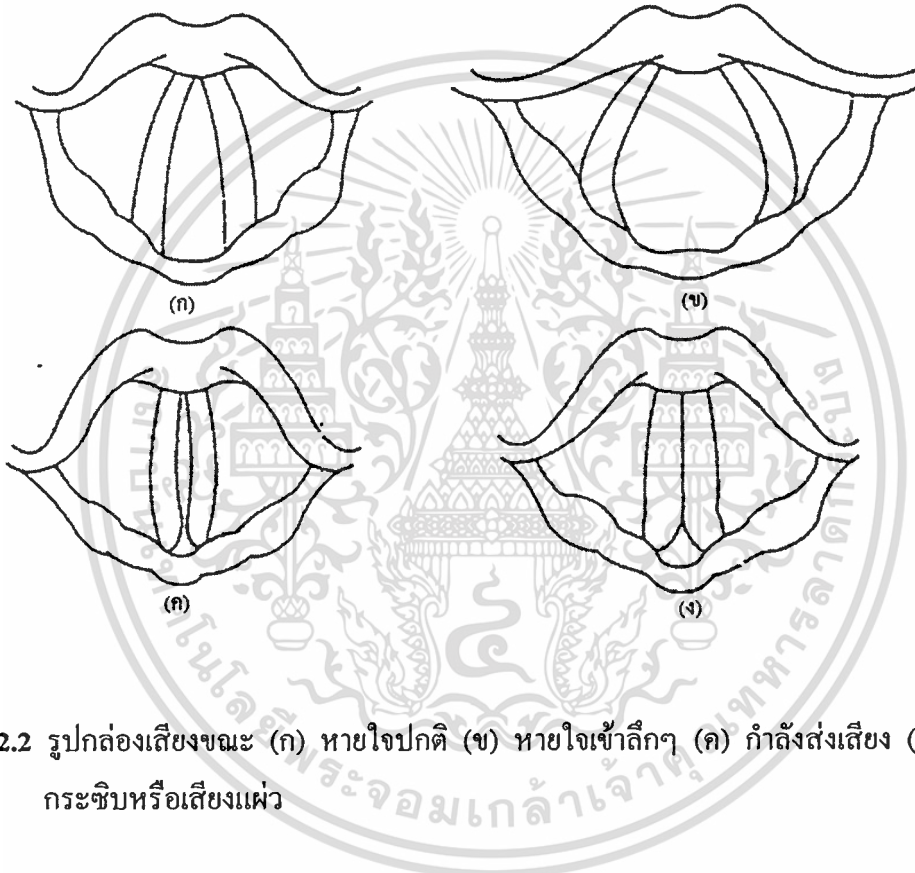
เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ของมหาวิทยาลัยราชภัฏวชิราวุฒวิทยาลัย ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.1 อวัยวะที่ใช้ในการออกเสียงพูด

ปอดและกระบังลม ทำหน้าที่สำคัญในการหายใจ และเป็นต้นกำเนิดการไหลของอากาศ ในกระบวนการผลิตเสียง

หลอดลม ทำหน้าที่นำอากาศจากปอดผ่านกล่องเสียง และเป็นอวัยวะที่อยู่ด้านหน้าของหลอดอาหาร

กล่องเสียง เป็นอวัยวะพิเศษที่ทำหน้าที่เป็นทางเดินอากาศเวลาหายใจ และเป็นตัวผลิตพัลส์ (Pulse) ของอากาศขณะเปล่งเสียง ซึ่งประกอบด้วยเส้นเสียง (Vocal cords) และช่องสายเสียง (Glottis) รูปร่างของกล่องเสียงและเสียงเสียงในลักษณะต่างๆ แสดงในรูปที่ ๒.๒



รูปที่ 2.2 รูปกล่องเสียงขณะ (ก) หายใจปกติ (ข) หายใจเข้าลึกๆ (ค) กำถึงส่งเสียง (ง) ส่งเสียง กระซิบหรือเสียงแผ่ว

ช่องปากและส่วนของหลอดอาหารตอนต้น อวัยวะกลุ่มนี้อยู่ต่อจากกล่องเสียงอาจเรียกว่า อวัยวะทำทอนเสียง (Vocal tract) ทำหน้าที่กำทอนเสียง โดยให้กำทอนทั้งเสียงขึ้นอยู่กับตำแหน่งของลิ้น ริมฝีปาก ขากรรไกร และเพดานอ่อน และเปลี่ยนแปลงไปตามการออกเสียง

โพรงจมูก เริ่มจากเพดานอ่อนจนถึงรูจมูกทั้งสอง ทำหน้าที่กำทอนเสียงร่วมกับช่องปากเมื่อมีการเปล่งเสียงที่ออกทางจมูก (nasal sounds) เช่น เสียง /ม./น./ และ /ง/ เป็นต้น

ริมฝีปาก เป็นอวัยวะส่วนที่เคลื่อนไหวได้มาก และทำให้เสียงแตกต่างกันได้มาก เราอาจจะบังคับริมฝีปากให้อยู่ชิดกัน ห่างกัน ขึ้นออก หรือห่อกลม ฯลฯ ลักษณะริมฝีปากต่างๆ นี้ล้วนแต่มีอิทธิพลต่อการออกเสียง และการทำให้เสียงแตกต่างกันไปทั้งสิ้น อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ฟัน เป็นอวัยวะที่เกิดของเสียงหลายชนิด เช่น เมื่อฟันบนกดลงบนริมฝีปากล่าง หรือฟันล่าง ลมที่ผ่านออกมาโดยแรงจะลอคช่องที่พอฟันได้ออกมา ทำให้เกิดเสียงเป็นเสียงชนิดที่เรียกว่า เสียงเสียดแทรกเป็นต้น

ปุ่มเหงือก เป็นส่วนนูนออกมาอยู่หลังฟันด้านบน ถ้าเอาลิ้นแตะจะรู้สึกว่ามีลักษณะเป็นคลื่น

เพดานแข็งหรือเพดานปาก คือ ส่วนเฉพาะที่โค้งเป็นกระดูกแข็ง

เพดานอ่อน คือ ส่วนของเพดานที่อยู่ต่อจากเพดานแข็งไปข้างในมีลักษณะเป็นกระดูกอ่อนที่ขยับขึ้น-ลงได้ เวลาหายใจเพดานอ่อนและลิ้นไก่ซึ่งอยู่ตอนปลายจะลกระดักลงมาเปิดช่องให้ลมออกไปทางจมูก นอกจากเวลาออกเสียงนาสิกเท่านั้นที่เพดานอ่อนจะลกระดักลงมา เพื่อให้ลมออกทางช่องจมูก

ลิ้นไก่ เป็นก้อนเนื้อเล็กๆ อยู่ต่อปลายเพดานตรงกลางปาก อวัยวะส่วนนี้สั้นร้วได้

ลิ้น เป็นส่วนที่เคลื่อนไหวมากที่สุดในการออกเสียงพูด จึงต้องแบ่งออกเป็น 3 ส่วน ตามหน้าที่ในการออกเสียง คือ

ปลายลิ้น คือ ส่วนปลายลิ้นซึ่งสามารถยกขึ้นไปแตะอวัยวะส่วนต่างๆ ในปากตอนบนได้โดยง่าย

หน้าลิ้น คือ ลิ้นที่อยู่ตรงข้ามกับเพดานแข็ง

หลังลิ้น คือ ส่วนของลิ้นที่อยู่ตรงข้ามกับเพดานอ่อน

แผ่นเนื้อปากหลดกลม เป็นก้อนเนื้อเล็กๆ คล้ายลิ้นไก่ อยู่ต่อโคนลิ้นลงไปในคอ มีหน้าที่ปิดช่องลมเมื่อรับประทานอาหาร และเปิดช่องลมเมื่อพูด

กรวยคอ หมายถึง โพรงคอที่อยู่ถัดจากปากลงไปจนถึงเส้นเสียง

ช่องจมูก หมายถึง โพรงในช่องจมูก ซึ่งอยู่เหนือลิ้นไก่ขึ้นไป เป็นช่องที่ลมซึ่งผ่านเส้นเสียงขึ้นจะผ่านออกไปทางจมูกได้เมื่อเวลาหายใจและเวลาออกเสียงนาสิก ในเวลาที่พูดเสียงอื่นลิ้นไก่จะถูกยกขึ้นไปปิดช่องจมูก เพื่อให้ลมออกมาทางปาก

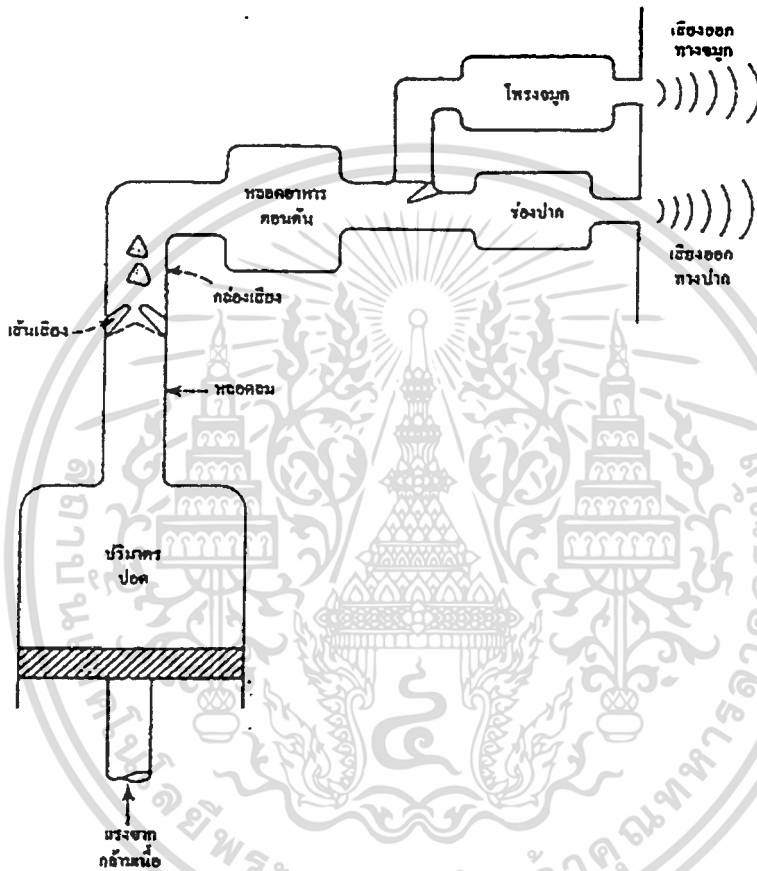
เส้นเสียงปลอม เป็นอวัยวะที่มีลักษณะเหมือนเส้นเสียงแต่อยู่เหนือเส้นเสียงขึ้นไป เส้นเสียงปลอมนี้เข้าใจกันว่าจะดึงเข้าหากันเมื่อเวลาพูดเสียงกระซิบ

2.1.2 กระบวนการผลิตเสียงพูด

จากระบบเสียงพูด สามารถแสดงเป็นแผนภาพของระบบกำเนิดเสียง ดังรูปที่ 2.3 ซึ่งสามารถจำแนกกลไกการสร้างเสียงพูดของมนุษย์ได้ 3 แบบ ดังนี้

1. อากาศไหลจากปอดจะถูกมอดูเลท (modulate) โดยการสั่นของเส้นเสียงทำให้เกิดคลื่นเสียงลักษณะคล้ายพัลส์ที่มีคาบเวลาแบบควอไซ (quasi-periodic pulse - excitation)

2. อากาศที่ไหลจากปอดถูกทำให้ปั่นป่วน ด้วยการบังคับให้ไหลผ่านช่องแคบอันเกิดจากการบีบตัวของอวัยวะในช่องปากทำให้เกิดเสียงลักษณะคล้ายเสียงรบกวน (noise-like excitation)
3. อากาศที่ไหลถูกกัก และเกิดแรงดันอยู่ภายในส่วนของช่องปากที่ปิด จากนั้นจึงปล่อยให้ อากาศที่มีแรงดัน พุ่งออกไปอย่างรวดเร็ว ทำให้เกิดการกระตุ้นเป็นเสียงในช่วงเริ่มต้น (Transient excitation)



รูปที่ 2.3 แผนภาพระบบเสียงพูดของมนุษย์

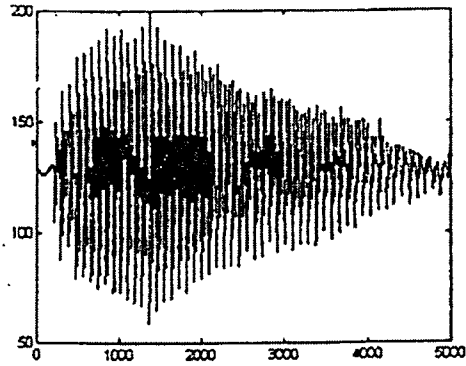
2.1.3 ลักษณะร่วมของเสียงพูด

เสียงที่ใช้ในภาษาพูดนั้นจะมีลักษณะที่สำคัญบางประการร่วมกัน ซึ่งเรียกได้ว่าเป็นลักษณะร่วมของเสียงพูด ลักษณะที่กล่าวถึงนี้มีอยู่หลายประการ คือ ความก้อง หรือ ไม่ก้องของเสียง

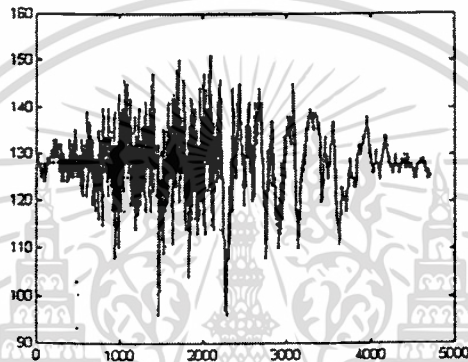
2.1.3.1 เสียงก้องหรือเสียงโฆษะ (voiced) เกิดจากการบังคับอากาศให้ผ่านช่องเสียงทำให้เกิดการเปลี่ยนแปลงความตึงของหย่อนเส้นเสียง โดยเส้นเสียงจะสั่นและเกิดเป็นพัลส์ (pulse) ของอากาศไปกระตุ้นอวัยวะกำทอนเกิดเป็นเสียงก้อง ตัวอย่างเสียงก้องได้แก่ เสียงสระ เสียงพยัญชนะที่

ต้องออกเสียงจากลำคอ (Voiced consonants) เช่น เสียง /a/ ดังรูปที่ 2.4 (ก) ให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.4 (ก) ตัวอย่างรูปคลื่นของสัญญาณเสียง /s/



รูปที่ 2.4 (ข) ตัวอย่างรูปคลื่นของสัญญาณเสียงไม่ก้อง /sh/

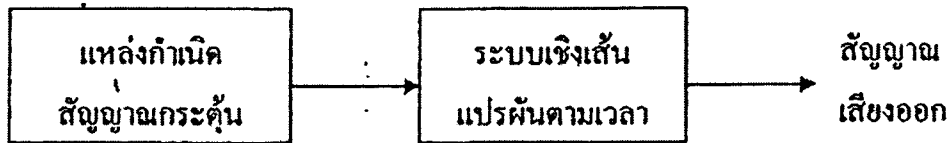
2.1.3.2 เสียงไม่ก้องหรือโหน้ะ (unvoiced หรือ voiceless) เป็นเสียงที่เกิดในช่องปากหรือโพรงจมูก โดยอวัยวะภายในช่องปาก ริมฝีปาก ขวางการไหลของอากาศให้ผ่านได้เป็นช่องเล็กๆ อากาศจึงไหลผ่านอย่างรวดเร็วและปั่นป่วนจนกระทั่งสร้างเป็นเสียงรบกวนช่วงความถี่กว้าง (broad-spectrum noise) ตัวอย่างเสียงไม่ก้องได้แก่ เสียงพยัญชนะที่ไม่ได้เกิดจากลำคอ (Voiceless consonants) เช่น เสียง /sh/ ดังรูปที่ 2.4 (ข)

2.1.4 แบบจำลองระบบกำเนิดเสียงพูด

จากรูปที่ 2.3 สามารถแสดงแผนภาพกรอบจำลองระบบกำเนิดเสียงเริ่มต้น ดังรูปที่ 2.5 จากรูปมีการแยกภาคแหล่งกำเนิดสัญญาณกระตุ้นออกจากส่วนกำหนดเสียง ซึ่งแทนด้วยระบบเชิงเส้นแปรผันตามเวลา แหล่งกำเนิดสัญญาณกระตุ้นทำหน้าที่แทนการทำงานของปอด และกล่องเสียง ส่วนนี้จะผลิตขบวนการพัลส์ที่มีคาบเวลาพิชชณะเปล่งเสียงว้อยซ์ และทำให้กำเนิดเสียงซึ่งคล้ายเสียงรบกวนขณะเปล่งเสียงโหน้ะ (Unvoiced) ส่วนที่สองเป็นท่อกำหนดเสียง จะแทน

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การทำงานของช่องปากและโพรงจมูกทำหน้าที่เสมือนตัวกรองสัญญาณ (Filter) ที่ยอมให้ความถี่ฟอร์แมนท์ผ่านได้ ซึ่งสามารถแทนด้วยระบบเชิงเส้นแปรผันตามเวลา (Time-vary linear system)



รูปที่ 2.5 แผนภาพกรอบจำลองระบบกำเนิดเสียงเริ่มต้น

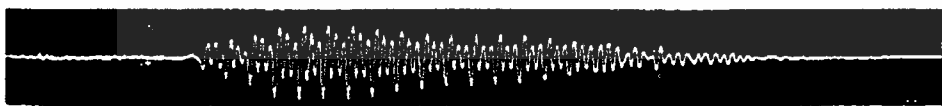
2.2 คุณลักษณะพื้นฐานของเสียงพูด

2.2.1 รูปลักษณ์ของเสียงพูด

รูปลักษณ์ของเสียงพูดที่ออกมาจากริมฝีปาก แสดงในรูปที่ 2.6 แอมพลิจูดแปรตามเวลาในแนวทางที่ซับซ้อน และอยู่เหนือช่วงกว้าง จากระดับที่เป็นนาฬิกาเป็นระดับที่หนาแน่น การกระจายที่เพิ่มขึ้นทางสถิติเมื่อค่าเฉลี่ยเกิน 10 นาฬิกาของเสียงที่พูดออกมา

แนวนอนในรูปแทนแอมพลิจูดของรูปลักษณ์เสียงพูด และวัดในหน่วยเดซิเบลกล่าวค่าที่เป็นผลของค่าเฉลี่ยระยะเวลา แนวตั้งแทนการกระจายที่เพิ่มขึ้นของแอมพลิจูด ในเป็นเปอร์เซ็นต์ระดับที่เกินถูกระบุโดยแนวนอน ข้อมูลของชาวญี่ปุ่นและอเมริกัน ดังที่แสดงมีช่วงที่เปลี่ยนแปลงแต่ละช่วงอยู่ที่ 55 dB

ความแตกต่างระหว่างระดับที่เพิ่มขึ้น 1% ของรูปลักษณ์เสียงพูด และระดับที่เป็นผลของค่าเฉลี่ยระยะยาวเรียกว่า จุดยอดปัจจัย (peak factor) และอยู่ที่ 12 dB ซึ่งสามารถเห็นได้ในรูปที่ 2.7 [5] จุดยอดปัจจัยวัดได้จากความแปรปรวนของรูปลักษณ์ ความแปรปรวนคือ σ^2 ดังนั้นรูปลักษณ์ของเสียงพูด

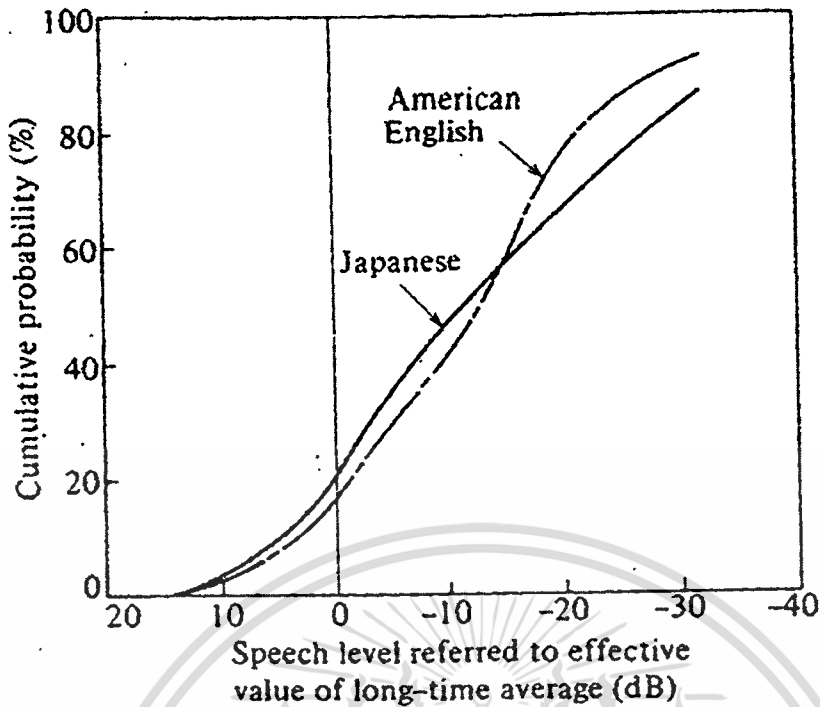


(ก)



(ข)

เอกสารที่ 2.6 รูปลักษณ์ของเสียงพูด (ก) เสียงผู้ชาย /pu/ (ข) เสียงผู้หญิง /tu/ าดให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.7 การกระจายตัวที่เพิ่มขึ้นของแอมพลิจูดสัญญาณเสียงพูด

ฟังก์ชันความหนาแน่นการกระจายสามารถแจกแจงจากความแตกต่างของการกระจายที่เพิ่มขึ้น ดังแสดงในรูปที่ 2.7 ฟังก์ชันความหนาแน่นนี้ใกล้เคียงกับเอ็กซ์โปเนนเชียล (exponential) ในช่วงแอมพลิจูดสูงและเกาส์เซียน ในช่วงแอมพลิจูดต่ำ ซึ่งฟังก์ชันความน่าจะเป็นความหนาแน่นของรูปคลื่นเสียงพูดสามารถประมาณค่าได้จากการรวมของฟังก์ชันเอ็กซ์โปเนนเชียลและเกาส์เซียน [6]

2.2.2 กำลังของเสียงพูด

ระดับที่เป็นผลของค่าเฉลี่ยระยะยาว ของรูปคลื่นเสียงพูด แปรผันเนื่องจากการเปลี่ยนตัวผู้พูด รวมทั้งการเปลี่ยนแปลงในสภาวะของเสียงที่เปล่งออกมา ในสภาพแวดล้อมที่เงียบ ค่าเฉลี่ยระยะยาวที่เป็นผลของเสียงพูดวัดได้จากกริมฟี ปากของผู้พูดจาก 1 เมตรข้างหน้าของผู้พูดอยู่ในช่วง 58.2 dB นี่คือการรวมค่าเฉลี่ยของผู้พูดที่เป็นผลของค่าเฉลี่ยระยะยาว สำหรับเสียงของผู้ชายและผู้หญิงเป็นการประมาณค่าเกาส์เซียน กับความเบี่ยงเบนมาตรฐานของการประมาณค่าอยู่ที่ 3.8 dB

ปัจจัยที่มีอิทธิพลอย่างมากระหว่างสภาวะเสียงที่เปล่งออกมา คือ สัญญาณรบกวนในห้อง โดยธรรมชาติเราจะพูดดังขึ้น ถ้าอยู่ในสภาวะที่เสียงรบกวนดัง ในกรณีของการคุยทางโทรศัพท์ โดยธรรมชาติจะสะท้อนกลับเป็นการลดความดัง เมื่อการพูดกลุ่มอื่นพูดดัง หรือเมื่อระดับน้ำเสียงผ่านวงจรโทรศัพท์เป็นค่าสูง ดังนั้นค่าเฉลี่ยระยะยาวของระดับเสียงพูดสำหรับ สำหรับการสนทนา

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปะเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ทางโทรศัพท์อยู่ที่ 63.1 dB ถ้าสมมติว่าค่าเฉลี่ยระยะยาวของระดับเสียงพูดที่เป็นผลเท่ากันในทุกทิศทางรอบศีรษะของผู้พูด กำลังงานที่เปล่งออกมาจากครึ่งวงกลม ของรัศมี 1 เมตร คือ 4π ครั้ง ที่ 63.1 dB และนี่จะเป็น $51.3 \mu W$ จริงๆ แล้วความหนาแน่นข้างหลังผู้พูดอยู่ที่ 4 dB ต่ำกว่า ด้านหน้าของผู้พูด ส่วนกำลังงานรวมที่เปล่งออกมาจากริมฝี ปากอยู่ที่ 20–30 μW

2.2.3 สเปกตรัมความถี่ของเสียงพูด

องค์ประกอบความถี่ของเสียงพูดเปลี่ยนแปลงอย่างต่อเนื่อง เป็นความหนาแน่นของเสียงพูด ค่าเฉลี่ยระยะยาวของสเปกตรัมความถี่แจกแจงได้จากการวัดระดับพลังงานในแต่ละแบนด์แคบของวงจรกรองความถี่ผ่าน (BPF) ครอบคลุมกับช่วงความถี่ทั้งหมดของสัญญาณเสียงพูด สเปกตรัมของค่าเฉลี่ยระยะยาวของผู้พูดชายและหญิงชาวญี่ปุ่น แสดงในรูปที่ 2.6 ความแตกต่างระหว่างเสียงของผู้ชายและผู้หญิงปรากฏในแบนด์ความถี่ต่ำที่สุดที่ค่าประมาณ 0.13 kHz โดยที่พลังงานเสียงพูดผู้หญิงเกือบจะไม่มีการวัดค่าที่คล้ายกันของเสียงผู้ชายและผู้หญิง แสดงว่าความแตกต่างเนื่องจากภาษา แสดงออกก่อนข้างยาก

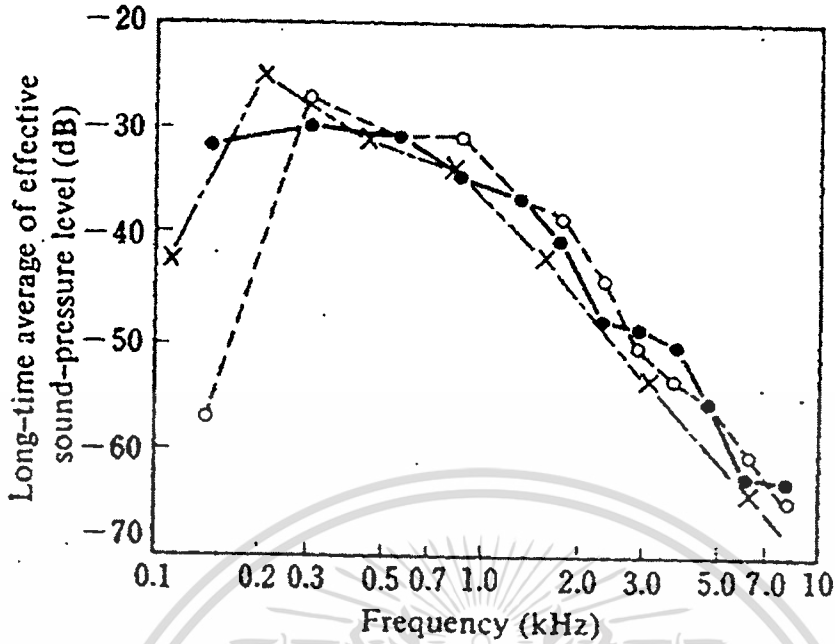
ในรูปที่ 2.8 มีองค์ประกอบความถี่ต่ำกว่า 0.8 kHz เกือบ 80% ของพลังงานทั้งหมด ดังนั้น สเปกตรัมเสียงพูดเป็นการ ใช้การประมาณค่าสเปกตรัมที่เป็นในแนวราบที่ 0.8 kHz และความชันที่ระดับของ -10 dB /วินาที เกินกว่า 0.8 kHz

2.2.4 พื้นฐานความถี่ของเสียงพูด

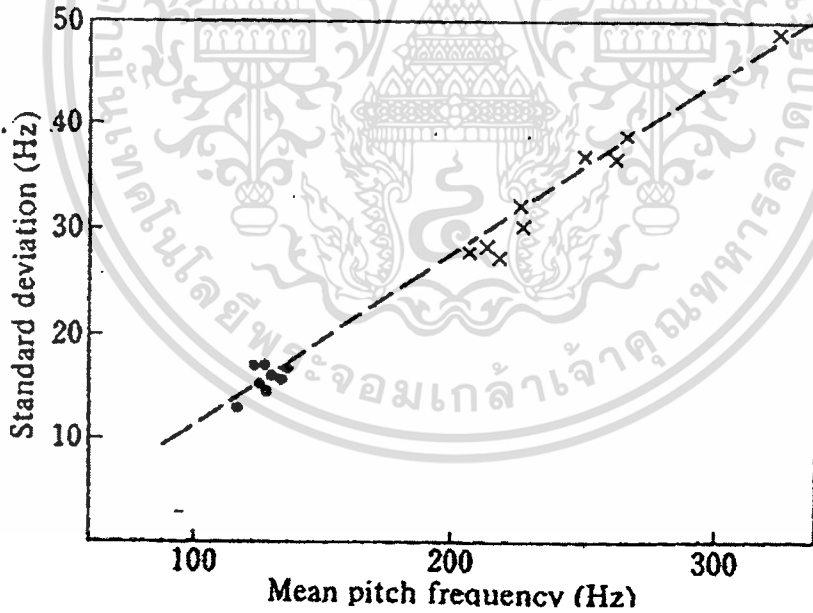
รูปคลื่นเสียงพูดประกอบขึ้นจาก 2 ส่วน คือ : ส่วนความเหมือนสัญญาณรบกวนที่แอมพลิฟิเคชันแปรผันอย่างสุ่ม และส่วน Periodic ซึ่งทำซ้ำเกือบจะเหมือนกันของการหมุนรอบรูปคลื่น ช่วงเวลาการวนซ้ำเรียกว่าช่วงเวลาพื้นฐาน และสิ่งที่เป็นไปได้ทั้งสองของช่วงเวลาเรียกว่า ความถี่พื้นฐาน รูปคลื่นที่ต่อเนื่องจะสร้างองค์ประกอบความถี่ในทุกๆ ความสัมพันธ์ฮาร์โมนิก ความถี่ต่ำที่สุด คือ ความถี่พื้นฐาน ที่สอดคล้องกับการสั่นของเส้นเสียง

ความถี่พื้นฐานสำหรับรูปคลื่นเสียงพูดที่มีการสนทนาแปรผันอย่างต่อเนื่องและซ้ำในเวลา ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพื้นฐานความถี่สำหรับผู้พูดคนเดียวแสดงในรูปที่ 2.9 การเบี่ยงเบนสำหรับเสียงผู้หญิงเป็น 2 ครั้ง จะมากเท่ากับเสียงผู้ชาย [7]

รูปแบบการแปรผันของเวลาของความถี่พื้นฐาน สำหรับเสียงพูดของการสนทนา แสดงการเลื่อนลงค่าประมาณค่าที่ 50% ของเวลาและเลื่อนขึ้นบนประมาณค่าได้ 18% ของเวลา การวิเคราะห์สเปกตรัมของการแปรผันตามเวลาของความถี่พื้นฐานระบุว่าอัตราการแปรผันน้อยกว่า 10 Hz หลังจากการเชื่อมต่อเส้น ไค้งเรียบในที่ไม่มีเสียง และช่วงหยุด



รูปที่ 2.8 ความถี่สเปกตรัมของเสียงพูด : (x---) คือ เสียงพูดผู้ชายชาวอเมริกัน, (●-) คือ เสียงพูดผู้ชายชาวญี่ปุ่น, (O---) คือ เสียงพูดผู้หญิงชาวญี่ปุ่น



รูปที่ 2.9 ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของความถี่ค่าพิทช์ของเสียง (●) ผู้ชาย, (x) ผู้หญิง

2.2.5 คุณลักษณะของสัญญาณเสียงพูด

เสียงพูดเป็นสัญญาณเสียงอนาล็อก ดังแสดงในรูป 2.6 ที่เวลาเดียวกันจะแทนรายละเอียดภาษาอธิบายโดยสัญลักษณ์ของการศึกษาหน่วยพื้นฐานของเสียงที่ไม่ต่อเนื่อง จำนวนของสัญลักษณ์ของหน่วยพื้นฐานเสียงในการใช้งานปกติขึ้นอยู่กับภาษา แต่โดยทั่วไปมีอย่างน้อย 20 หน่วยพื้นฐานเสียงที่แตกต่างกัน อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

20–30 ชนิดของภาษาที่ระบุ การแปรผันอย่างกว้างของคุณสมบัติการควบคุมเสียง เรียกว่า หน่วยพื้นฐานของเสียงและจัดระดับเป็นเสียงสระหรือ เสียงพยัญชนะ

2.2.5.1 เสียงสระ

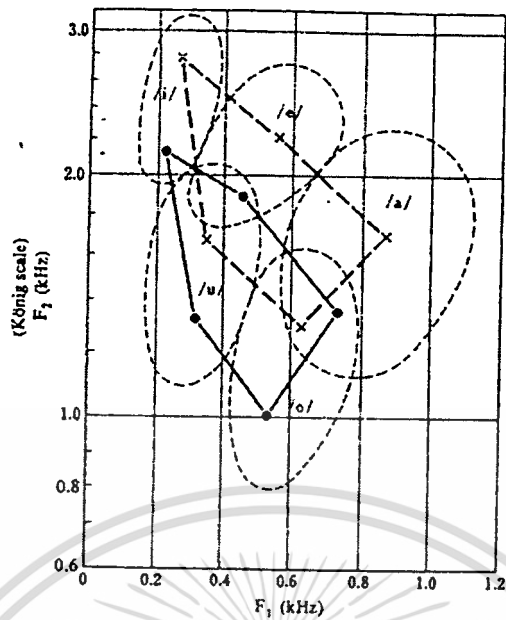
เสียงสระสร้างขึ้นมากดเสียงสะท้อนของบริเวณเสียงที่เปล่งออกมา โพรงปากและจมูกเป็นรูปร่างนำไปสู่แบบที่ถูกตัว และถูกกระตุ้นโดยการไหลของอากาศ ผลลัพธ์จากการเปิดเป็นช่วงๆ และการปิดของเส้นเสียง สระแต่ละตัวเป็นการจัดคุณลักษณะจากาการระบุเสียงสะท้อน บางภาษามีมากกว่า 10 เสียง

ฟอร์แมนท์ (formant) เป็นองค์ประกอบความถี่ที่เด่นกว่าของคุณสมบัติแต่ละเสียงสระ มีการจัดลำดับ ไปเป็นระดับความถี่ที่หนึ่ง ระดับความถี่ที่สอง และระดับความถี่ที่สาม และอื่นๆ จากที่สูงกว่าต่ำที่สุดในความถี่เสียงสะท้อน

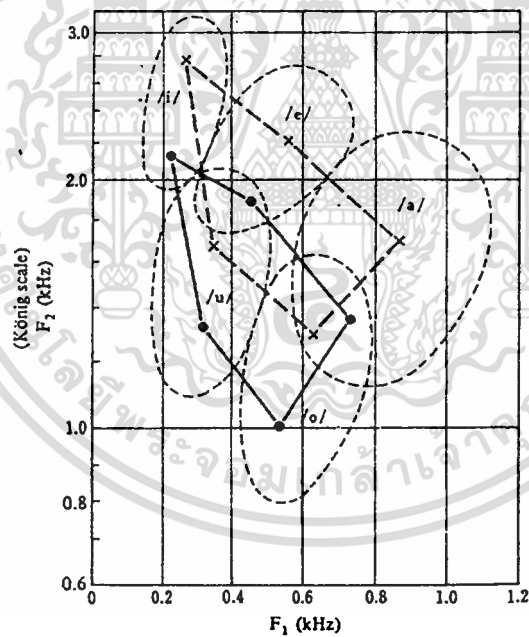
ความถี่ฟอร์แมนท์จะแปรผันเหนือช่วงกว้าง ขึ้นอยู่กับเพศของผู้พูด และคุณลักษณะอื่นๆ รวมทั้งการแปรผันที่ขึ้นกับส่วนก่อนหน้าและหน่วยพื้นฐานเสียงตามมา การแปรผันเหล่านั้นเหมือนกับว่าเป็นธรรมชาติ พิจารณาความสัมพันธ์ที่ใกล้ชิดระหว่างระดับความถี่ และรูปร่าง กับมิติของบริเวณเสียงที่เปล่งออกมา คุณลักษณะของเสียงสระขึ้นกับระดับในช่วงความถี่ต่ำ ซึ่งเป็นระดับความถี่ที่หนึ่งและที่สอง การกระจายของเสียงสระของชาวญี่ปุ่น 5 คน ในระดับความถี่ที่หนึ่ง และที่สอง ดังแสดงในรูปที่ 2.10 (ก) การนำเสนอที่คล้ายกันของระดับตำแหน่งของฟอร์แมนท์ที่หนึ่งและที่สอง สำหรับเสียงสระภาษาอังกฤษของชาวอเมริกันดังแสดงในรูปที่ 2.10(ข) [8] แกนแนวนอนและแนวตั้งแทนระดับความถี่ที่หนึ่งและที่สอง ตามลำดับ ตัวอย่างของแต่ละเสียงสระระบุด้วยสัญลักษณ์เสียงสระของผู้ชายและผู้หญิง และโดเมนการประมาณค่าของการแปรผันเนื่องจากการเปลี่ยนผู้พูดหรือสภาวะเสียงที่เปล่งออกมาจะถูกระบุโดยเส้นประที่ปิดเส้นโค้งสำหรับแต่ละเสียงสระในรูปที่ 2.10 (ก)

การพิจารณานี้คาบเกี่ยวระหว่างโดเมนความถี่ฟอร์แมนท์ แต่การแปรผันตามรูปแบบทั่วไปของการเลื่อนแนวขนาน การเก็บความสัมพันธ์ของตำแหน่งของแต่ละเสียงสระไม่เปลี่ยนตัวอย่าง เช่น ความสัมพันธ์ระหว่างเสียงพูดของผู้ชายและผู้หญิง การรวมแกน (F_3) ในแนวตั้งกับระนาบ (F_1) และ (F_2) จะแยกเสียงสระ 5 ตัวได้อย่างชัดเจนมากกว่า

ระดับความถี่ที่หนึ่ง ระดับความถี่ที่สองและที่สามของเสียงสระชาวญี่ปุ่น แสดงในตารางที่ 2.1 ระดับความถี่ที่ลำดับสูงกว่าแสดงช่วงเล็กๆ อย่างสัมพันธ์กันกับการแปรผันระหว่างเสียงสระซึ่งขึ้นกับผู้พูดเกือบทั้งหมด



(ก)



(ข)

รูปที่ 2.10 (ก) ความถี่ฟอร์แมนที่หนึ่งและที่สองของเสียงชาวญี่ปุ่น 5 เสียง : (●—) ผู้ชาย, (×—) ผู้หญิง (ข) ความถี่ฟอร์แมนที่หนึ่งและที่สองของเสียงชาวอเมริกัน 10 เสียง : * : /i/ in heed, ■ : /ɪ/ in hid, ○ : /ɛ/ in head, ▲ : /æ/ in had, ○ : /a/ in hod, □ : /ɔ/ in hawed, △ : /ʊ/ in hood, ● : /u/ in who'd, × : /ʌ/ in hud, + : /ə/ in heard ซึ่งการออกเสียงนี้จากผู้พูด 10 คน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 2.1 ความถี่ฟอร์แมนต์ของเสียงสระชาวญี่ปุ่น 5 เสียง

Vowel	F_1 (kHz)		F_2 (kHz)		F_3 (kHz)	
	Male	Female	Male	Female	Male	Female
/i/	0.16–0.3	0.2–0.35	1.9–2.4	2.5–3.1	2.5–3.2	3.3–4.0
/e/	0.3–0.6	0.4–0.75	1.7–2.2	1.9–2.5	2.3–3.1	2.6–3.5
/a/	0.6–0.9	0.7–1.2	1.0–1.6	1.2–1.9	2.2–3.0	2.5–3.3
/o/	0.4–0.65	0.5–0.8	0.65–1.3	0.9–1.6	2.2–3.0	2.4–3.3
/u/	0.2–0.5	0.2–0.5	1.1–1.8	1.2–2.1	2.0–2.8	2.4–3.1

2.2.5.2 เสียงพยัญชนะ

เสียงพยัญชนะสร้างจากการไหลทะลักของอากาศที่ไหลสร้างขึ้นใกล้ๆ กับจุดที่มีการบีบตัวของบริเวณที่มีการเปล่งเสียงออกมาอันเนื่องมาจากการตอกัน อากาศที่ไหลที่จุดปิดของโพรงปากสร้างการหยุด เสียงเสียดสีถูกสร้างขึ้นจากสถานที่ของการบีบตัวสูงสุด และเสียงกระทบของลิ้นกับไรฟันก็ถูกสร้างในระหว่างนี้ด้วย เสียงพยัญชนะเป็นเสียงหรือไม่มีเสียง ขึ้นกับว่าไม่มีการสั่นของเส้นเสียง

เสียงจากจมูก เปล่งออกมาช่องจมูกผ่านโพรงจมูกกับโพรงปากปิดที่ริมฝีปาก หรือที่อื่นข้างในโพรงปาก เป็นการเชื่อมต่อกับโพรงจมูกในแนวนอน นี่เป็นสาเหตุการป้องกันเสียงสะท้อน (ศูนย์) จากด้านผลกระทบของโพรงปากในสเปกตรัมความถี่

นอกจากนั้นมีเสียงที่เรียกว่าเสียงลิ้น และเสียงที่ออกคล้ายเสียงกระพือ เสียงมีการส่งผ่านในทางตรงข้ามกับเสียงสระที่คงที่ พยัญชนะทุกตัวจัดลำดับถูกต้องมากกว่าสถานที่ของการตอกัน (บริเวณที่มีการบีบตัวมากที่สุด) ดังแสดงในตารางที่ 2.2 คุณลักษณะของแต่ละหน่วยพื้นฐานเสียงขึ้นอยู่กับความต่อเนื่องของรูปคลื่น ความถี่สเปกตรัม ระยะเวลา และการส่งผ่าน

ความถี่ระหว่าง ความถี่สเปกตรัมหรือระดับการส่งผ่านและการให้สัญญาณการตอกันแสดงในพยางค์เปิด พยัญชนะสระ (CV) ในรูปที่ 2.11 [9] รูปนี้แสดงว่าความแตกต่างในท่าทางของการผลิตสะท้อนบนการเคลื่อนที่ในระดับความถี่ที่หนึ่ง และแตกต่างในสถานที่ตอกันสะท้อนบนการส่งผ่านในระดับความถี่ที่สอง คุณลักษณะของพยัญชนะเหล่านี้เป็นหัวข้อหลักของการเปลี่ยนขึ้นอยู่กับสระที่ขึ้นต้นก่อนและตามมา

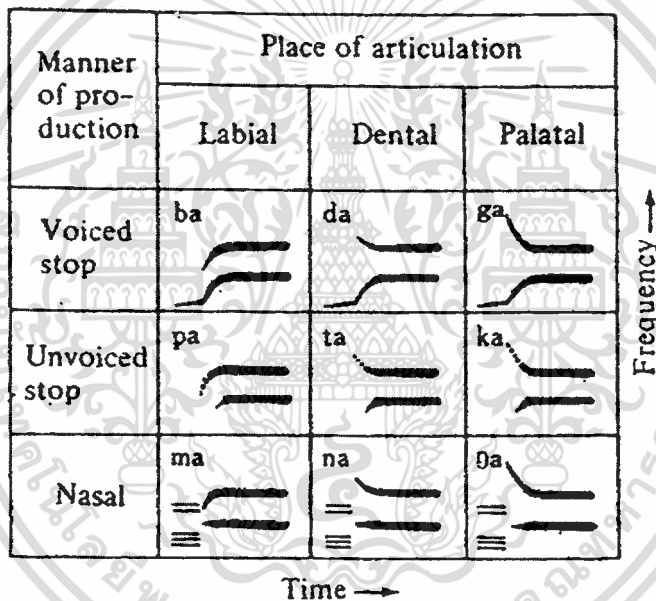
คุณสมบัติการควบคุมเสียงแปรผันอย่างต่อเนื่องในระหว่างแต่ละหน่วยพื้นฐานเสียงระหว่างเสียงพูดของการสนทนาปกติ เพราะว่าการเคลื่อนที่ที่ยุ่งเหยิง จากอวัยวะในการส่งผ่านระหว่างแต่ละหน่วยพื้นฐาน เรียกว่า การต่อร่วมกัน

ตารางที่ 2.2 การแบ่งแยกเสียงพยัญชนะของชาวญี่ปุ่น

Manner of articulations	Place of articulation								
	Bilabial		Dental		Alveolar		Palatal		Glottal
	<U ^a >	<V ^b >	<U>	<V>	<U>	<V>	<U>	<V>	<U>
Stop	p	b			t	d	k	g	
Affricate			ts	dz	tʃ	dʒ			
Fricative	f		s	z	ʃ	ʒ			h
Nasal		m				n		ɲ	
Liquid		w						j	
Flap					r				

^a Voiced.

^b Unvoiced.



รูปที่ 2.11 ความเหมือนระหว่างการพูดออกเสียงพยัญชนะและสเปคตรัม

คุณสมบัติการควบคุมเสียงแปรผันกล่าวคือการเปลี่ยนแปลงในหน่วยพื้นฐานที่ติดกัน เนื่องจากการต่อรวมกัน ซึ่งเป็นงานที่ซับซ้อนที่จะระบุ ความสัมพันธ์หนึ่งต่อหนึ่ง ระหว่างหน่วยพื้นฐานกับความสัมพันธ์กับการควบคุมเสียงที่ออกมาแทนคุณลักษณะไบนารีหลายๆ ค่า โดยเลือกจากการระบุแต่ละหน่วยพื้นฐานของสเปคตรัม อย่างเช่น กระชับ / กระจ่ายได้ ซึ่งเรียกว่าคุณลักษณะเด่น [10] ตัวอย่างการจัดลำดับหน่วยพื้นฐานภาษาอังกฤษจากคุณลักษณะเด่น แสดงในตารางที่ 2.3 ภายในคุณสมบัติ 9 คู่ : เสียงพูด (vocalic) / ไม่มีเสียงพูด (Nonvocalic) พยัญชนะ (Consonantal) / ไม่มีพยัญชนะ (nonconsonantal) กระชับ (compact) / กระจ่าย (diffuse) เสียงเข้ม (grave) / เสียงแหลม (acute) เสียงแบน (flat) / เสียงเรียบ (plane) จมูก (nasal) / ปาก (oral) เครียด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การวิเคราะห์ของระยะเวลาของหน่วยพื้นฐาน แสดงว่าระยะเวลาของเสียงสระจะไม่เปลี่ยนแปลง ระยะเวลาเฉลี่ยประมาณ 70 msec อย่างไรก็ตามระยะเวลาของพยัญชนะแปรผันจาก 5 ถึง 130 msec ขึ้นอยู่กับระดับชั้นนั้น ซึ่งระยะเวลาพยางค์แปรผันจาก 75 ถึง 200 msec หรือ ค่าเฉลี่ยอยู่ที่ 130 msec

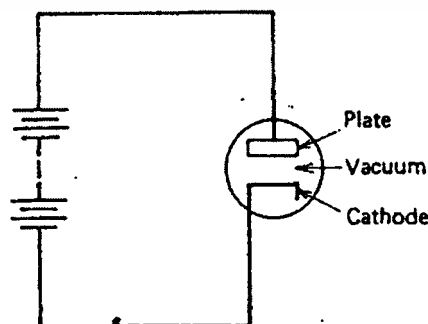
2.2 การแสดงสัญญาณรบกวนทางคณิตศาสตร์

2.3.1 สัญญาณรบกวน

สัญญาณรบกวน นิยมใช้ระบุคลื่นไฟฟ้า ซึ่งพยายามจะเข้าไปรบกวนการส่งสัญญาณและการประมวลสัญญาณ (signal processing) ในระบบสื่อสาร ซึ่งจะไม่สามารถกำจัดมันออกมาได้อย่างสมบูรณ์ ในทางปฏิบัตินั้นจะพบว่าแหล่งกำเนิดสัญญาณรบกวนมากมาย ในระบบสื่อสารนั้น สัญญาณรบกวนอาจเป็นประเภทที่เกิดขึ้นจากภายนอกระบบ และประเภทที่เกิดขึ้นภายในระบบเอง สัญญาณรบกวนที่เกิดจากภายนอกนั้นได้แก่ สัญญาณรบกวนจากบรรยากาศ (atmospheric noise), galactic noise และสัญญาณรบกวนที่มนุษย์สร้างขึ้น (man-made noise) เป็นต้น นอกจากนี้ยังมีสัญญาณรบกวนที่สำคัญซึ่งเกิดจากการเปลี่ยนแปลงโดยธรรมชาติของกระแสหรือโวลเตจในวงจรไฟฟ้า สัญญาณรบกวนเหล่านี้คือ ช็อตนอยส์ (shot noise) และเทอร์มัลนอยส์ (thermal noise) ซึ่งจะกล่าวพอสังเขปดังต่อไปนี้

2.3.1.1 ช็อตนอยส์

เป็นสัญญาณรบกวนที่เกิดขึ้นในอุปกรณ์อิเล็กทรอนิกส์ เนื่องมาจากธรรมชาติของกระแสที่ไหลในอุปกรณ์ ซึ่งสัญญาณรบกวนนี้จะเป็นแบบไม่ต่อเนื่อง เช่นการไหลของกระแสในไดโอด (diode) ตามรูปที่ 2.12 จะประกอบด้วยขั้วอิเล็กโทรด 2 ตัว ที่อยู่ในสุญญากาศ กล่าวคือ มีแคโทด (cathode) จะถูกเผาจนกระทั่งทำให้อิเล็กตรอนแผ่กระจายออกมา และมีแอนโนด (anode) หรือเพลท (plate) ซึ่งมีศักดาไฟฟ้าเป็นบวก



รูปที่ 2.12 ไดโอด

เมื่อเทียบกับแคโทดเพื่อที่จะรวบรวมอิเล็กตรอน ในขณะที่สมมติว่าความต่างศักย์ระหว่างแคโทด และเพลทสูงมาก เพื่อที่จะให้อิเล็กตรอนที่วิ่งจาก cathode ไปยัง plate มีความเร็วสูง ดังนั้นจึงไม่ต้องคำนึงถึง space charge effect และกระแสที่เพลทจะถูกกำหนดโดยอัตราการเคลื่อนที่ของอิเล็กตรอนจาก cathode โดยอาศัยการพิจารณาว่ากระแสที่เพลทเกิดจากผลรวมของกระแสที่เป็นแบบพัลส์ ซึ่งเนื่องจากการเคลื่อนที่ของอิเล็กตรอนผ่านช่องว่างระหว่าง cathode และ plate เป็นที่ทราบกันว่าค่า mean-square ของกระแสที่เพิ่มขึ้นและลดลงตามการสุ่ม (random) ถูกกำหนดให้เป็นดังสมการ 2.1

$$\overline{I_{SN}^2} = 2eI_0\Delta f \quad (2.1)$$

ซึ่ง e เป็นอิเล็กตรอนชาร์จมีค่าเท่ากับ 1.59×10^{-19} คูโลมปี I_0 เป็นค่า mean ของกระแสในหน่วยแอมแปร์ และ Δf เป็นแบนด์วิดธ์ของเครื่องมือวัดในหน่วย Hertz สมการ 2.1 นี้เป็นสูตรของ Scotty โดยทั่วไปการเคลื่อนที่ของอิเล็กตรอนจาก cathode ไปยัง plate จะใช้เวลาประมาณอันดับของ 10^{-9} วินาที ทั้งนี้ไม่คำนึงถึง transit time effects สำหรับคุณสมบัติที่สำคัญของซ็อนทอยส์ก็คือ มันจะเป็นการแจกแจงแบบเกาส์เซียน (Gaussian distributed) ซึ่งมีค่า mean เป็นศูนย์

2.3.1.2 เทอร์มัลนอยส์

เทอร์มัลนอยส์เป็นชื่อของสัญญาณรบกวนทางไฟฟ้า ซึ่งเกิดจากการเคลื่อนที่ตามการสุ่มของอิเล็กตรอนในตัวนำ ในทางปฏิบัติทุกกรณี ค่า mean square ของเทอร์มัลนอยส์ที่ปรากฏระหว่างขั้วทั้งสองของรีซิสเตอร์เมื่อวัดในช่องแบนด์วิดธ์ Δf ถูกกำหนดให้เป็นดังสมการ 2.2

$$\overline{V_{TN}^2} = 4kTR\Delta f \quad (2.2)$$

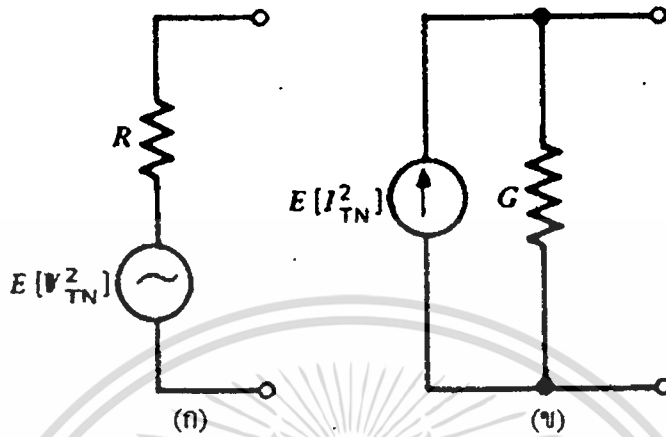
ซึ่ง k เป็น Boltzmann's constant มีค่าเท่ากับ $1.38 \times 10^{-23} J / ^\circ K$

T เป็นองศาสัมบูรณ์ในหน่วยของศาเคลวิน

R เป็นรีซิสเตนซ์ในหน่วยโอห์ม

รีซิสเตอร์ซึ่งมีสัญญาณรบกวนอยู่นั้น อาจแสดงด้วยวงจรสมมูลเทเวนิน (Thevenin) ซึ่งประกอบด้วย แหล่งกำเนิดสัญญาณรบกวน $E = \overline{V_{TN}^2}$ ต่ออนุกรมอยู่กับรีซิสเตอร์ที่ปราศจากสัญญาณรบกวนตามรูปที่ 2.13 (ก) หรือในทางกลับกันอาจแสดงด้วยวงจรสมมูลนอร์ตัน (Norton)

ซึ่งประกอบด้วยแหล่งกำเนิดกระแสซึ่งต่อขนานอยู่กับความนำที่ปราศจากสัญญาณรบกวน ตามรูปที่ 2.13 (ข)



รูปที่ 2.13 รูปแบบของรีซิสเตอร์ที่มีสัญญาณรบกวน (ก) วงจรสมมูลย์เทวินิน (ข) วงจรสมมูลย์นอร์ตัน

ค่า mean square ของแหล่งกำเนิดกระแสของสัญญาณรบกวน คือ

$$\begin{aligned} \overline{I_{TN}^2} &= \frac{1}{R^2} \overline{V_{TN}^2} \\ &= 4kTG\Delta f \end{aligned} \tag{2.3}$$

ในที่นี้ $G = 1/R$ เป็นความนำ และพึงเข้าใจว่าเทอร์มัลนอยส์นี้เป็นการแจกแจงแบบเกาส์เซียนซึ่งมีค่า mean เป็นศูนย์

ในการคำนวณหาสัญญาณรบกวนจะเกี่ยวข้องกับทฤษฎีของการขนถ่ายกำลังเพื่อให้ได้ค่าสูงสุด (maximum power transfer) ซึ่งในที่นี้สมมติว่ามีโหลด R_L มาต่อ ถ้า $R_L = R$ แล้ว ภายใต้เงื่อนไขที่เรียกว่าแมชชิง (matching) นี้ โหลดจะได้รับกำลังเพียงครึ่งหนึ่งของกำลังที่จ่ายมาทั้งหมด และกำลังที่จ่ายให้โหลดในขณะนี้ถูกเรียกว่า อะไวเลเบิลนอยส์พาวเวอร์ (available noise power) ซึ่งจะทราบว่า อะไวเลเบิลพาวเวอร์ P_a จากรูปที่ 2.13 (ก) คือ

$$P_a = \frac{V_{TN}^2}{4R} \tag{2.4}$$

ดังนั้นจะได้อะไวเลเบิลเทอร์มัลนอยส์พาวเวอร์ของรีซิสเตอร์ R ในช่วงแบนด์วิดท์ Δf เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 คิงนี่
 ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$P_a = \frac{4kTR\Delta f}{kT\Delta f} \tag{2.5}$$

2.3.1.3 ไวท์นอยส์ (White Noise)

การวิเคราะห์เรื่องของสัญญาณรบกวนในระบบสื่อสารนั้น โดยทั่วไปจะพิจารณาสัญญาณรบกวนในรูปของอุดมคติ ซึ่งเรียกว่าไวท์นอยส์ ซึ่งจะมีความหนาแน่นสเปกตรัมเป็นอิสระกับความถี่ “white” ถูกพิจารณาในทฤษฎีที่ว่ามันเป็นแสงสีขาวซึ่งมีองค์ประกอบของทุกความถี่ในช่วงการแผ่กระจายของแม่เหล็กไฟฟ้าที่เห็นได้ โดยทั่วไปจะกำหนดให้ความหนาแน่นของสเปกตรัมกำลังของไวท์นอยส์เป็นดังนี้

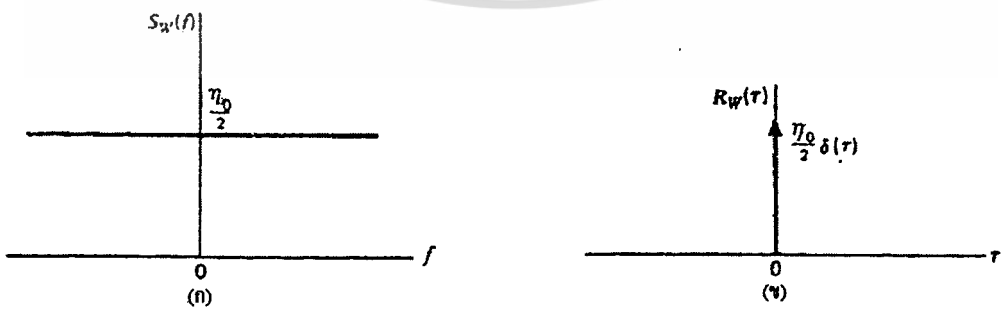
$$S_w(f) = \frac{\eta_0}{2} \tag{2.6}$$

แฟกเตอร์ 1/2 แสดงโดยสเปกตรัม 2 ด้าน คือทั้งทางบวกและทางลบดังแสดงไว้ในรูปที่ 2.14 (ก)

η จะมีหน่วยเป็น *watts / Hz* และตามปกติ η นี้จะเกี่ยวข้องกับอยู่ที่อินพุทของระบบสื่อสารซึ่งสามารถคำนวณได้จาก

$$\eta_0 = kT_e \tag{2.7}$$

- ซึ่ง k เป็น Boltmann’s constant
- T_e เป็น equivalent noise temperature ของเครื่องรับ



รูปที่ 2.14 คุณสมบัติของไวท์นอยส์ (ก) ความหนาแน่นของสเปกตรัมกำลัง (ข) ออกได้คอร์เรชันฟังก์ชัน

เนื่องจากออโต้คอร์เรเลชันฟังก์ชันเป็นอินเวอร์สฟูรีเยร์ทรานส์ฟอร์มของความหนาแน่นของสเปกตรัมกำลัง ดังนั้นสำหรับกรณีที่เป็นไวท์นอยส์ เมื่อคำนวณออโต้คอร์เรเลชันฟังก์ชัน $R_{yy}(\tau)$ แล้วจะได้

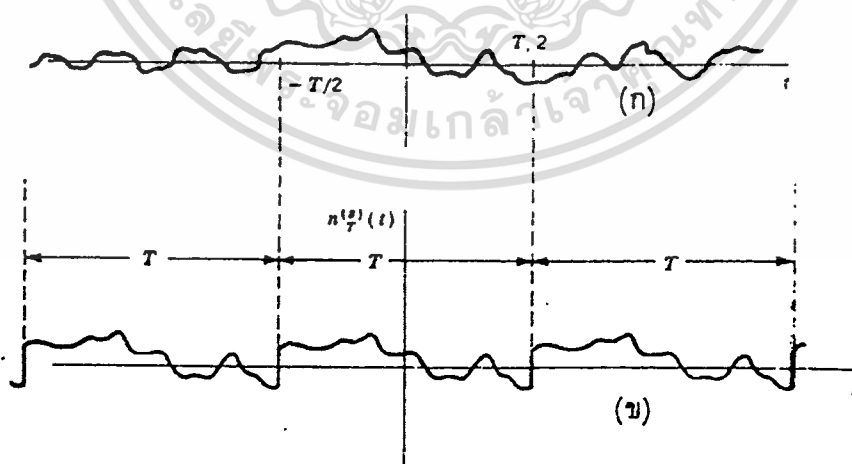
$$R_{yy}(\tau) = \frac{\eta_0}{2} \delta(\tau) \quad (2.8)$$

นั่นคือ ออโต้คอร์เรเลชันของไวท์นอยส์ก็คือ เดลต้าฟังก์ชันที่ถ่วงคูณโดย $\eta/2$ และเกิดขึ้นที่ $\tau=0$ ตามรูปที่ 2.14 (ข) ซึ่งจะสังเกตเห็นได้ว่า $R_{yy}(\tau)=0$ เมื่อ $\tau \neq 0$ ฉะนั้น แชนเนลเปิดของไวท์นอยส์ที่ต่างกัน 2 ค่าใดๆ แม้ว่าแชนเนลทั้งสองจะถูกเก็บในเวลาทีใกล้เคียงกันก็ตาม ก็จะไม่มีความคล้ายคลึง (uncorrelated) กันเลย

2.3.2 การแสดงสัญญาณรบกวนในโดเมนความถี่

ในระบบสื่อสารนั้น สัญญาณรบกวนที่ปนเข้ามาับสัญญาณจะผ่านไปยังฟิลเตอร์ ซึ่งฟิลเตอร์ที่กล่าวนี้โดยทั่วไปจะแสดงคุณสมบัติของมันในโดเมนความถี่ ดังนั้นในการคำนวณหาค่าต่างๆ ที่มีอิทธิพลกับฟิลเตอร์เหล่านี้จะสะดวกมากถ้าใช้คุณสมบัติของสัญญาณรบกวนในโดเมนความถี่ กล่าวคือจะใช้ความหนาแน่นของสเปกตรัมกำลังของสัญญาณรบกวนมาพิจารณา ซึ่งก็จะมีคุณสมบัติกับการคำนวณที่ใช้กับรูปร่างสัญญาณ (waveform) ต่างๆ ตามที่ได้กล่าวมาแล้วข้างต้น

จากรูปที่ 2.15 (ก) ขึ้นแรกลองเลือกแชนเนลเปิดของสัญญาณรบกวนฟังก์ชันในช่วงคาบเวลา T คือเลือกเวลาระหว่าง $t = -T/2$ ถึง $+T/2$ และสมมติว่าฟังก์ชันนี้คือ $n^{(s)}(t)$



รูปที่ 2.15 (ก) รูปร่างของแชนเนลเปิดสัญญาณรบกวน (ข) รูปร่างสัญญาณมีคาบเวลาที่ได้จาก (ก) ในช่วง $-T/2$ ถึง $+T/2$

ต่อไปกำหนดให้ฟังก์ชันของเวฟฟอร์มมีคาบเวลา T เป็น $n_T^{(s)}(t)$ ตามรูปที่ 2.15 (ข) ซึ่งเวฟฟอร์มนี้สามารถแสดงในรูปของอนุกรมฟูรีเยร์ในช่วงจาก $-T/2$ ถึง $+T/2$ ได้ ทั้งกำหนดให้ $\Delta f = 1/T$ เป็นความถี่เริ่มต้น และสมมติว่าเวฟฟอร์มนี้ไม่มีองค์ประกอบของกระแสตรง ดังนั้นจึงเขียนเป็นอนุกรมฟูรีเยร์ได้ดังนี้

$$n_T^{(s)}(t) = \sum_{k=1}^{\infty} \{a_k \cos(2\pi k \Delta f t) + b_k \sin(2\pi k \Delta f t)\} \quad (2.9)$$

หรือแสดงในเทอมของโคไซน์เพียงอย่างเดียวได้

$$n_T^{(s)}(t) = \sum_{k=1}^{\infty} c_k \cos(2\pi k \Delta f t - \theta_k) \quad (2.10)$$

ในที่นี้ a_k , b_k และ c_k เป็นสัมประสิทธิ์อนุกรมฟูรีเยร์ θ_k เป็นมุมของเฟส (phase angle) ซึ่งมีความสัมพันธ์ระหว่างกันดังนี้

$$c_k^2 = a_k^2 + b_k^2 \quad (2.11)$$

และในที่นี้

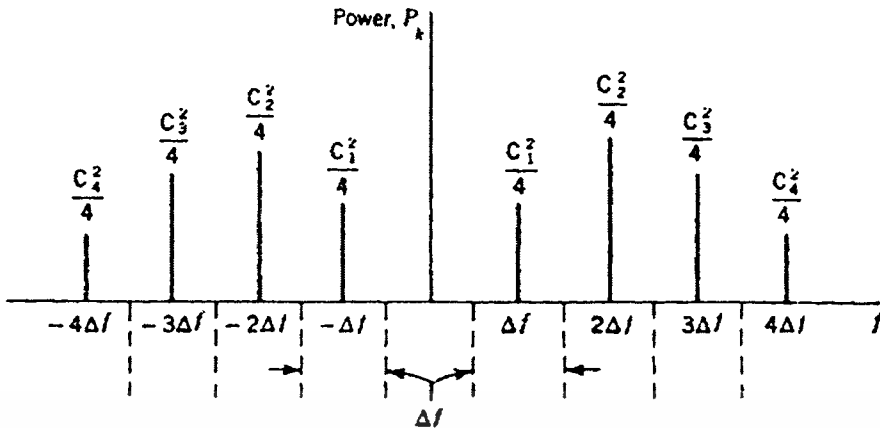
$$\theta_k = \tan^{-1} \frac{b_k}{a_k} \quad (2.12)$$

การเลือกพิเศษสำหรับสัมประสิทธิ์เหล่านี้จะได้จากแอมพลิจูดฟังก์ชันเฉพาะของสัญญาณรบกวนซึ่งมีคาบเวลาการเลือกที่แตกต่างกันไปจะได้แอมพลิจูดฟังก์ชันที่ต่างกัน

เนื่องจากกำลังเฉลี่ยของทุกองค์ประกอบของสเปกตรัม c_k คือ $c_k^2/2$ เมื่อแสดงเป็นสเปกตรัมสองด้านจะมีขนาดเป็น $c_k^2/4$ ตามที่แสดงไว้ในรูปที่ 2.16

ขณะนี้จะกำหนดให้ความหนาแน่นของสเปกตรัมกำลังที่ความถี่ $k\Delta f$ เป็นปริมาณ ดังนี้

$$\begin{aligned} S_n(k\Delta f) &\equiv S_n(-k\Delta f) \equiv \frac{c_k^2}{4\Delta f} \\ &= \frac{a_k^2 + b_k^2}{4\Delta f} \end{aligned} \quad (2.13)$$



รูปที่ 2.16 สเปกตรัมกำลังของ $n_t^{(s)}(t)$

ดังนั้น กำลังทั้งหมด P_k ที่เกี่ยวข้องกับในช่วงความถี่ Δf ณ ที่ความถี่ $k\Delta f$ คือ

$$P_k = 2S_n(k\Delta f)\Delta f \quad (2.14)$$

เนื่องจาก P_k เกี่ยวข้องกับสเปกตรัมที่ความถี่ $k\Delta f$ ทั้งสองด้าน ดังนั้น $S_n(k\Delta f) \equiv S_n(-k\Delta f)$ จะเป็นกำลังทางด้านบวกหรือลบเมื่อถูกแบ่งโดยขนาดของ Δf ฉะนั้น $S_n(k\Delta f)$ จะเป็นความหนาแน่นของสเปกตรัมกำลังเฉลี่ย (สองด้าน) ของทุกช่วงความถี่

ถ้ากำหนดให้คาบเวลา $T \rightarrow \infty (\Delta f \rightarrow 0)$ ดังนั้นเวฟฟอร์มที่กำหนดให้มีคาบเวลาก็จะมีสภาพเหมือนแชนเนลสัญญาณรบกวนเวฟฟอร์มจริงๆ นั่นคือสามารถพิจารณาได้ว่า $n_t^{(s)}(t) = n(t)$ และสามารถเขียนสมการ 2.9 ใหม่ได้ดังนี้

$$n(t) = \lim_{\Delta f \rightarrow 0} \sum_{k=1}^{\infty} \{a_k \cos(2\pi k \Delta f t) + b_k \sin(2\pi k \Delta f t)\} \quad (2.15)$$

หรืออีกแบบหนึ่งคือ

$$n(t) = \lim_{\Delta f \rightarrow 0} \sum_{k=1}^{\infty} c_k \cos(2\pi k \Delta f t + \theta_k) \quad (2.16)$$

จากสมการ 2.13 แทนค่า c_k^2 ด้วย c_k^2 โดยพิจารณาว่าเป็นค่าเฉลี่ยของกำลังสองของตัวแปร แรนดอม c_k และนอกจากนั้น เมื่อพิจารณาว่า Δf เข้าใกล้ศูนย์แล้วจุด Δf ก็จะอยู่ชิดกันมาก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จนกลายเป็นตัวแปรความถี่ ซึ่งต่อเนื่อง f ดังนั้น ค่า $k\Delta f$ จากสมการ 2.13 จึงสามารถเขียนแทนด้วยตัวแปรต่อเนื่อง f ได้ และก็สามารถเขียนสมการ 2.11 ใหม่ได้ดังนี้

$$\overline{c_k^2} = \overline{a_k^2} + \overline{b_k^2} \quad (2.17)$$

ผลสุดท้ายจะเขียนสมการ 2.13 ใหม่ได้ดังนี้

$$S_n(f) = \lim_{\Delta f \rightarrow 0} \frac{\overline{c_k^2}}{4\Delta f} = \lim_{\Delta f \rightarrow 0} \frac{\overline{a_k^2} + \overline{b_k^2}}{4\Delta f} \quad (2.18)$$

และพึงสังเกตว่ากำลังในช่วงความถี่จาก f_1 ถึง f_2 คือ

$$\begin{aligned} P(f_1 \rightarrow f_2) &= \int_{-f_2}^{-f_1} S_n(f) df + \int_{f_2}^{f_1} S_n(f) df \\ &= 2 \int_{f_2}^{f_1} S_n(f) df \end{aligned} \quad (2.19)$$

ซึ่งในขณะที่กำลังทั้งหมด P_T จะมีค่าดังนี้

$$P_T = \int_{-\infty}^{\infty} S_n(f) df = 2 \int_0^{\infty} S_n(f) df \quad (2.20)$$

2.3.3 ผลของการกรองความถี่ต่อความน่าจะเป็นของความหนาแน่นของเกาส์เซียนนอยส์

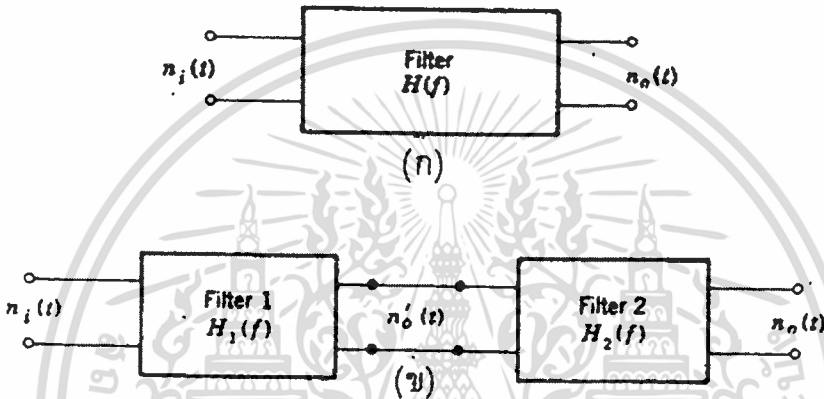
พิจารณาจากรูปที่ 2.17 (ก) ถ้าอินพุตเป็นเกาส์เซียนนอยส์ $n_i(t)$ สัญญาณรบกวนที่เอาท์พุท $n_o(t)$ ก็จะเป็นเกาส์เซียนด้วย สมมติว่าอิมพัลส์เรสปอนส์ของฟิลเตอร์เป็น $h(t)$ จะได้

$$n_o(t) = \int_{-\infty}^{\infty} n_i(\tau) h(t-\tau) d\tau = \int_{-\infty}^t n_i(\tau) h(t-\tau) d\tau \quad (2.21)$$

สมการ 2.21 นั้น โดยทั่วไปลิมิตด้านบนอาจต้องใช้ $\tau=t$ เนื่องจากตัวแปรในการอินทิเกรตคือ τ และ เมื่อ $\tau > t$ นั้น $h(t-\tau)=0$ สมการนี้แสดงการวางซ้อนของอิมพัลส์ $n_i(\tau) d\tau$ ที่ป้อนให้กับอินพุตของฟิลเตอร์ในประเด็นนี้จะเขียนสมการ 2.21 ใหม่ได้ในรูปของ

$$n_0(t) = \lim_{\Delta\tau \rightarrow 0} \sum_{k=-\infty}^{k=t/\Delta\tau} n_i(k\Delta\tau)h(t-k\Delta\tau)\Delta\tau \quad (2.22)$$

ซึ่ง k จะมีขอบเขตสูงกว่าค่าในอินทิกรัล ในทฤษฎีทางลิมิตนั้น $k\Delta\tau$ จะกลายเป็นตัวแปร τ และผลรวมก็จะกลายเป็นอินทิกรัล จากสมการ 2.22 $n_i(k\Delta\tau)$ จะเป็นตัวแปรแบบเรคตองมนั้นคือ สัญญาณรบกวนจะถูกแสดงโดยฟังก์ชันต่างๆไป และค่าของ $n_i(k\Delta\tau)$ ก็จะขึ้นอยู่กับแอมพลิจูดฟังก์ชันที่จะพิจารณา ในทางตรงกันข้ามปริมาณ $h(t-k\Delta\tau)$ จะเป็นค่าคงที่โดยมีจำนวนที่กำหนดแน่นอน และถูกกำหนดโดยตัวฟิลเตอร์



รูปที่ 2.17 (ก) เกาส์เซียนนอยส์ $n_i(t)$ ซึ่งป้อนให้กับบลีเนียร์ฟิลเตอร์แล้วเกิดเอาต์พุต ($n_o(t)$)
(ข) ฟิลเตอร์จาก (ก) เมื่อแยกเป็นสองส่วน

สมมติว่าฟิลเตอร์มีทรานส์เฟอร์ฟังก์ชัน $H(f)$ และแยกออกเป็น 2 ส่วนตามรูปที่ 2.17 (ข) โดยที่ $H(f) = H_1(f)H_2(f)$ ดังนั้น ถ้า $n_i(t)$ เป็นไวทน์นอยส์แบบเกาส์เซียนแล้วทั้ง $n'_o(t)$ และ $n_o(t)$ จะเป็นเกาส์เซียนด้วยโดยอาศัยการพิจารณาจากข้างต้น นอกจากนั้นเมื่อ $n'_o(t)$ และ $n_o(t)$ เป็นเกาส์เซียนแล้วผลที่ได้ก็คือ แม้ว่าอินพุตจะไม่ใช่ไวทน์นอยส์แบบเกาส์เซียนก็ตาม แต่เมื่อผ่านเข้าไปยังฟิลเตอร์แบบเชิงเส้น แล้วจะทำให้เกิดสัญญาณรบกวนที่เอาต์พุตเป็นแบบเกาส์เซียนอีก

สามารถจะใช้สมการ 2.22 ได้โดยตรงกับฟิลเตอร์ 2 ในรูปที่ 2.17 (ข) อย่างง่ายๆ โดยแทนค่าอิมพัลส์ $n_i(k\Delta\tau)$ ด้วยอิมพัลส์ $n'_i(k\Delta\tau)$

2.3.4 องค์ประกอบสเปกตรัมของสัญญาณรบกวน

องค์ประกอบสเปกตรัมที่เกี่ยวข้องกับช่วงความถี่ที่ k ซึ่งกำหนดให้เป็น $n_k(t)$ เมื่อ $\Delta f \rightarrow 0$ จะเขียนได้ดังนี้

$$n_k(t) = a_k \cos(2\pi k\Delta f t) + b_k \sin(2\pi k\Delta f t) \quad (2.23)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับกรใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หรือ

$$n_k(t) = c_k \cos(2\pi k \Delta f t - \theta_k) \quad (2.24)$$

องค์ประกอบสเปกตรัมในขณะนี้ประกอบด้วยสัญญาณรบกวนและสิ่งเหล่านี้จะเป็น random process ดังนั้น a_k , b_k และ θ_k ตามสมการ 2.23 และ 2.24 จะเป็น random variable สำหรับ $n_k(t)$ จะเป็น ensemble ของแอมพลิจูดฟังก์ชัน ซึ่งทุกฟังก์ชันเหล่านี้จะเป็นลักษณะคลื่นแบบ deterministic จากความจริงที่ว่าคลื่นลักษณะชาวยน์ที่แท้จริงแต่ละคลื่นจะมีแอมมพลิจูดและเฟสที่แตกต่างกันซึ่งขึ้นอยู่กับค่าของ c_k และ θ_k random process $n_k(t)$ จะเป็น stationary นั่นคือคุณสมบัติทางสถิติจะไม่เปลี่ยนแปลงไปกับเวลา แต่อย่างไรก็ตาม ค่าเฉลี่ยตามเวลาของแต่ละแอมพลิจูดฟังก์ชันของ ensemble จะแตกต่างกันไปจากแอมพลิจูดอื่น ๆ

ขณะนี้จะลองดูคุณสมบัติบางประการของ random variables a_k และ b_k นอร์มัลไลซ์กำลัง P_k ของ $n_k(t)$ ถูกกำหนดโดยค่าเฉลี่ยทั่ว ensemble ของ $[n_k(t)]^2$ ซึ่งได้จากสมการ 2.23 คือ

$$P_k = \overline{[n_k(t)]^2} = \overline{a_k^2 \cos^2 2\pi k \Delta f t + b_k^2 \sin^2 2\pi k \Delta f t + 2a_k b_k \sin 2\pi k \Delta f t \cos 2\pi k \Delta f t} \quad (2.25)$$

พึงเข้าใจว่า $n_k(t)$ เป็น stationary process ดังนั้น $[n_k(t)]^2$ จะไม่ขึ้นอยู่กับเวลา ดังนั้นในการคำนวณ P_k นั้น จะกำหนดให้ $t = t_1$ และ $\cos 2\pi k \Delta f t_1 = 1$ ซึ่งจะทำให้ $\sin 2\pi k \Delta f t_1 = 0$ ดังนั้น จะได้

$$P_k = \overline{a_k^2} \quad (2.26)$$

ในทำนองเดียวกันก็อาจแสดงได้ว่า $P_k = \overline{b_k^2}$ ดังนั้นจะได้

$$\overline{a_k^2} = \overline{b_k^2} \quad (2.27)$$

จากสมการ 2.14, 2.17, 2.26 และ 2.27 จะได้

$$P_k = 2S_n(k\Delta f)\Delta f = 2S_n(-k\Delta f)\Delta f = \overline{a_k^2} = \overline{b_k^2} = \frac{\overline{a_k^2}}{2} = \frac{\overline{b_k^2}}{2} = \frac{\overline{c_k^2}}{2} \quad (2.28)$$

เนื่องจาก $\overline{a_k^2} = \overline{b_k^2}$ จึงเขียนสมการ 2.25 ใหม่ได้เป็น

$$P_k = \overline{a_k^2} (\cos^2 2\pi k \Delta f t + \sin^2 2\pi k \Delta f t) + \overline{2a_k b_k \sin 2\pi k \Delta f t \cos 2\pi k \Delta f t} \quad (2.29)$$

หรือ

$$P_k = \overline{a_k^2} + \overline{2a_k b_k \sin 2\pi k \Delta f t \cos 2\pi k \Delta f t} \quad (2.30)$$

แต่อย่างไรก็ตาม จากสมการ 2.26 นั้น $P_k = \overline{a_k^2}$ จะไม่ขึ้นอยู่กับเวลา เพื่อที่จะทำให้สมการ 2.30 มีค่าเดียวกันกับค่าดังกล่าว จึงต้องทำให้

$$\overline{a_k b_k} = 0 \quad (2.31)$$

ดังนั้น สัมประสิทธิ์ a_k และ b_k จะไม่เกี่ยวข้องกัน (uncorrelate)

2.3.4 ผลของฟิลเตอร์ที่มีต่อความหนาแน่นของสเปกตรัมกำลังของสัญญาณรบกวน กำหนดให้องค์ประกอบสเปกตรัมของสัญญาณรบกวน $n_k(t)$ ที่กำหนดไว้ตามสมการ 2.23 เป็นอินพุทของฟิลเตอร์ ซึ่งมีทรานส์เฟอร์ฟังก์ชัน ณ ความถี่ $k\Delta f$ เป็นดังนี้

$$H(k\Delta f) = |H(k\Delta f)| e^{j\varphi_k} = |H(k\Delta f)| \angle \varphi_k \quad (2.32)$$

องค์ประกอบสเปกตรัมของสัญญาณรบกวน $n_{k_0}(t)$ ที่สอดคล้องกัน ณ ที่เอาต์พุทจะมีค่า ดังนี้

$$n_{k_0}(t) = |H(k\Delta f)| a_k \cos(2\pi k \Delta f t + \varphi_k) + |H(k\Delta f)| b_k \sin(2\pi k \Delta f t + \varphi_k) \quad (2.33)$$

จากสมการ 2.28 กำลังที่สอดคล้องกับ $n_{k_i}(t)$ คือ

$$P_{k_i} = \frac{\overline{a_k^2} + \overline{b_k^2}}{2} \quad (2.24)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เนื่องจาก $|H(k\Delta f)|$ เป็น deterministic function $\overline{[H(k\Delta f)a_k]^2} = |H(k\Delta f)|^2 \overline{a_k^2}$ และ $\overline{[H(k\Delta f)b_k]^2} = |H(k\Delta f)|^2 \overline{b_k^2}$ ฉะนั้น เมื่อเปรียบเทียบกับสมการ 2.32 กับสมการ 2.23 จะได้ $P_{k_0}(t)$ ที่สอดคล้องกับ $n_{k_0}(t)$ ดังนี้

$$P_{k_0} = |H(k\Delta f)|^2 \frac{\overline{a_k^2} + \overline{b_k^2}}{2} \quad (2.35)$$

ผลสุดท้าย จากสมการ 2.5 และ 2.6 และโดยการใช้สมการ 2.26 จะได้ความสัมพันธ์ของความหนาแน่นของสเปกตรัมกำลังที่อินพุตและเอาต์พุตดังนี้

$$S_{n_0}(k\Delta f) = |H(k\Delta f)|^2 S_n(k\Delta f) \quad (2.36)$$

ในทฤษฎีทางลิมิตนั้น ในขณะที่ $\Delta f \rightarrow 0$, $k\Delta f$ จะกลายเป็นตัวแปร f ซึ่งต่อเนื่อง ดังนั้นจึงเขียนสมการ 2.36 ใหม่ได้ดังนี้

$$S_{n_0}(f) = |H(f)|^2 S_n(f) \quad (2.37)$$

2.4 การนิยามหน่วยเดซิเบล

ในการส่งผ่านสัญญาณในระบบสื่อสารนั้น ถ้าระบบนั้นประกอบด้วยชิ้นส่วนแพสซีฟล้วนๆ เช่น วงจรฟิลเตอร์ หรือสายนำสัญญาณ เป็นต้น สัญญาณขาออกมักจะมีกำลังลดลงกว่าสัญญาณขาเข้าในกรณีที่ระบบนั้นเป็นวงจรขยาย ระดับสัญญาณออกจะใหญ่กว่าสัญญาณขาเข้า อัตราส่วนระหว่างกำลังของสัญญาณขาออกกับกำลังของสัญญาณขาเข้านี้ ถ้ามีค่าน้อยกว่า 1 ก็จะเป็นการสูญเสียกำลังสัญญาณ และถ้ามีค่ามากกว่า 1 ก็จะเป็นการขยายกำลังสัญญาณ ถ้าให้ P_o และ P_i เป็นกำลังสัญญาณขาเข้าและกำลังของสัญญาณขาออกตามลำดับ เราจะนิยามการสูญเสียหรือการขยายกำลังของสัญญาณได้ดังนี้

$$\begin{aligned} L &= \frac{P_o}{P_i} & P_o < P_i \\ G &= \frac{P_o}{P_i} & P_o > P_i \end{aligned} \quad (2.38)$$

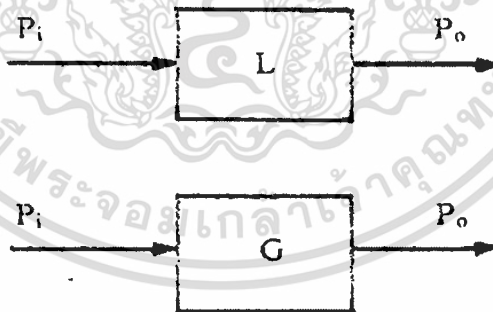
การสูญเสียกำลังสัญญาณ L และอัตรายขยาย G ที่นิยามตามสมการข้างบนนี้ จะมีหน่วยเป็นจำนวนเท่า หน่วยที่เป็นจำนวนเท่านี้จะไม่ให้ความสะดวกในการคำนวณมากนัก เพราะจะต้องทำการคูณหารตัวเลขที่ละมากๆ จึงได้มีการนิยามค่าดังกล่าวนี้และค่ากำลังสัญญาณในรูปของเดซิเบลเพื่อเปลี่ยนการคำนวณจากการคูณหารให้เป็นการบวกลบดังนี้

$$\begin{aligned} L &= 10 \log L \\ G &= 10 \log G \end{aligned} \quad (2.39)$$

$$\begin{aligned} P \text{ dB}_W &= 10 \log \frac{P}{1W} \\ P \text{ dB}_m &= 10 \log \frac{P}{1mW} \end{aligned} \quad (2.40)$$

เมื่อใช้นิยามของเดซิเบลตามสมการ (2.39) และสมการ (2.40) การคำนวณกำลังของสัญญาณในสมการ (2.38) จะเปลี่ยนจากการคูณหารเป็นการบวกลบดังนี้

$$\begin{aligned} P_o \text{ dB}_W &= P_i \text{ dB}_W + L \text{ dB} \\ P_o \text{ dB}_W &= P_i \text{ dB}_W + G \text{ dB} \end{aligned} \quad (2.41)$$



รูปที่ 2.18 การคิดกำลังสัญญาณในรูปเดซิเบล

ประโยชน์ของการใช้หน่วยเดซิเบลนี้ นอกจากจะเปลี่ยนเลขคูณหารเป็นเลขบวกลบแล้ว ยังทำให้ตัวเลขที่มีค่าใหญ่กลายเป็นตัวเลขที่มีขนาดเล็กลง อย่างเช่น 10,000 เท่า จะเท่ากับ 40 dB เป็นต้น ซึ่งคุณสมบัตินี้จะถูกนำไปใช้ในการออกแบบเครื่องวัดที่ต้องการให้สามารถวัดและแสดงค่าที่ต่างกันมากเป็นหมื่นหรือเป็นแสนเท่า การคิดเป็นหน่วยเดซิเบลก็ทำให้เราสามารถเปลี่ยนตัวเลข 9 หลัก ให้เป็น ตัวเลข 2 หลักเท่านั้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.5 การวัดคุณภาพของเสียง

ในการพิจารณาว่าคุณภาพของสัญญาณเสียงจะดีมากน้อยเท่าไรนั้นเราจะนิยามจากอัตราส่วนระหว่างกำลังของสัญญาณที่ต้องการกับกำลังของสัญญาณรบกวน (signal to noise ratio) และเพื่อความกะทัดรัดจะขอใช้ตัวย่อภาษาอังกฤษว่า SNR และสามารถเขียนเป็นสมการได้ดังนี้

$$\begin{aligned} SNR &= \frac{S}{N} \\ &= \frac{x^2}{n^2} S \end{aligned} \quad (2.41)$$

ค่า SNR นั้นโดยทั่วไปเราจะใช้หน่วยเป็นเดซิเบล (dB) ซึ่งหาได้โดยการทำอัตราส่วนที่มีหน่วยเป็นเท่าให้เป็นเดซิเบล หรือถ้าหน่วยของกำลังเป็นหน่วยที่คิดเป็นเดซิเบลอยู่แล้ว การหาค่า SNR จะหาได้จากผลต่างของสัญญาณและสัญญาณรบกวน ดังสมการ (2.42)

$$\begin{aligned} SNR &= 10 \times \log_{10} \left(\frac{|y|^2}{|n|^2} \right) \\ &= 10 \times \log_{10} \left(\frac{|y|^2}{|y-x|^2} \right) \end{aligned} \quad (2.42)$$

เมื่อ n คือ สัญญาณรบกวน
 x คือ สัญญาณเสียง
 y คือ สัญญาณเสียงที่ผิดเพี้ยน

สำหรับความต้องการค่า SNR ของสัญญาณแต่ละชนิด เช่น สัญญาณเสียง สัญญาณวิดีโอ และสัญญาณเสียงตามโทรศัพท์ เป็นต้น จะไม่เท่ากันยกตัวอย่างเช่น สัญญาณวิดีโอจะต้องมี SNR สูงกว่า 40 เดซิเบลขึ้นไป สัญญาณเสียงโดยเฉพาะเสียงดนตรีที่ฟังเสนาะหูจะต้องมีค่า SNR ที่สูงกว่า 50 เดซิเบลขึ้นไป สำหรับสัญญาณโทรศัพท์นั้นถ้ามีค่า SNR สูงกว่า 25 เดซิเบลขึ้นไปก็จัดว่าใช้ได้เป็นต้น

บทที่ 3

ทฤษฎีที่เกี่ยวข้อง

3.1 การแปลงสัญญาณอนาลอกเป็นดิจิทัล

รูปคลื่นเสียงพูด โดยปกติถูกแปลงเป็นรูปของดิจิทัล สำหรับการประมวลผล การแปลงจากสัญญาณอนาลอก เป็นรูปแบบดิจิทัล เกี่ยวข้องกับ 2 ขั้นตอน คือ 1. การแซมปลิง (Sampling) และ 2. การควอนไทซ์และการเข้ารหัส “การแซมปลิง” หมายถึง การเปลี่ยนรูปคลื่น ซึ่งต่อเนื่องกันในเวลานั้นเป็นสัญญาณดิจิทัล นั่นคือ ลำดับของตัวอย่างที่ไม่ต่อเนื่อง (discrete) ในเวลา “การควอนไทซ์” หมายถึง การประมาณค่าตัวอย่างจากตัวแปรโดยหนึ่งของค่าที่จำกัดประกอบอยู่ในกลุ่มของรหัส ทั้งสองขั้นตอนนี้ยอมให้รูปคลื่นที่ต่อเนื่องถูกแสดงโดยลำดับของค่าแต่ละค่าเป็นสมาชิกของกลุ่มของค่าที่จำกัด หลังจากการแซมปลิงและการควอนไทซ์ สัญญาณคือรหัสนำไปแสดงในรูปไบนารี นั่นคือ รหัสเลขฐานสอง (binary code)

3.1.1 การแซมปลิง

การแซมปลิงเป็นการแปลงจากสัญญาณอนาลอกที่ต่อเนื่องในช่วงเวลาเป็นลำดับของตัวอย่างที่ไม่ต่อเนื่องในช่วงเวลาแสดงได้เป็น $\{f_n\} = \{f(nT)\}$ ที่ $t_n = nT, n = 0, \dots, \infty$ ช่วงของการแซมปลิง คือ T วินาที และแลกเปลี่ยนกัน $S = 1/T$ Hz เป็นความถี่ของแซมปลิง กระบวนการการแซมปลิง แสดงในรูปที่ 3.1 ตามทฤษฎีการแซมปลิงของ Shannon-Someya [4] จำกัดความสัมพันธ์ระหว่างช่วงการแซมปลิง และคุณสมบัติของสัญญาณเป็นตัวอย่างขณะที่การแซมปลิงเป็นเหตุของการผิดเพี้ยนในสัญญาณตัวอย่าง

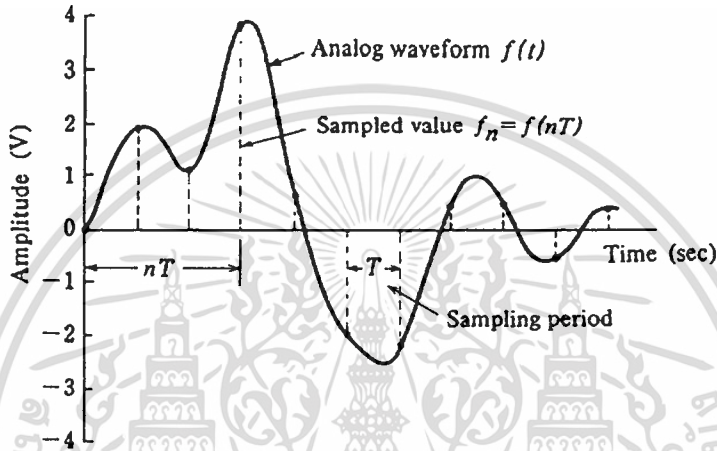
เมื่อสัญญาณอนาลอก $f(t)$ เป็นแบนด์ความถี่ที่จำกัดในช่วง 0 ถึง W Hz, $f(t)$ เป็นการสร้างใหม่โดยการใช้ค่าแซมปลิงที่ทุกๆ ช่วง $1/2W$ วินาที ดังสมการ 3.1

$$f(t) = \sum_{n=-\infty}^{\infty} f\left(\frac{n}{2W}\right) \left\{ \frac{\sin \pi(2Wt - n)}{\pi(2Wt - n)} \right\} \quad (3.1)$$

โดยที่ $f(n/2W)$ เป็นค่าแซมปลิงของ $f(t)$ ที่ $t = n/2W, n = -\infty, \dots, 0, \dots, \infty, T = 1/2W$ ช่วงการแซมปลิงที่ต้องการ และ $S = 2W$ ความถี่ตัวอย่างที่ต่ำที่สุดที่อนุญาตให้ได้ ในกรณีของรูปคลื่นของเสียงพูดส่งผ่านสายโทรศัพท์แบบเดิม, $W = 4$ kHz และช่วงการแซมปลิงเป็น $T = \frac{1}{8}$ มิลลิวินาที

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในหลายๆกรณีที่ W ไม่ได้กำหนดไว้อย่างชัดเจน แต่ถือว่า W คือ ความถี่ที่มีพลังงานของสัญญาณมากกว่า 30–40 dB ซึ่งต่ำกว่า (ลดทอนประมาณ 10^4) ระดับค่าเฉลี่ยของพลังงานในแบนด์ความถี่หลัก โดยปกติแล้ววงจรกรองความถี่ต่ำ (LPF) จะกำหนดค่าก่อนที่จะแซมปลิง เพื่อป้องกันการบิดเบี้ยวของสัญญาณ เมื่อ W ไม่ได้กำหนดอย่างชัดเจน แต่แม้ว่า W ควรจะเลือกช่วงความถี่อย่างระมัดระวังซึ่งก็คือ การลดทอนจาก 30 ถึง 40 dB และไม่ต่ำกว่า 3 dB ความถี่คัทออฟ ของ LPF ความสัมพันธ์เหล่านี้ ซึ่งแสดงในรูปที่ 3.2



รูปที่ 3.1 กระบวนการแซมปลิง

ถ้าการแซมปลิงไม่เป็นไปตามทฤษฎีการแซมปลิง แซมปลิงโดย $T > 1/2W$ สำหรับสัญญาณของการจำกัดแบนด์ W Hz องค์ประกอบความถี่สูงมีการบิดเบี้ยวแสดงในรูปที่ 3.2 (ข) ยิ่งไปกว่านั้นผลลัพธ์ของการแซมปลิงที่ $T > 1/W$ ไม่สามารถแสดงได้เด่นชัดจากการแซมปลิงสัญญาณ DC แสดงในรูปที่ 3.2 (ค)

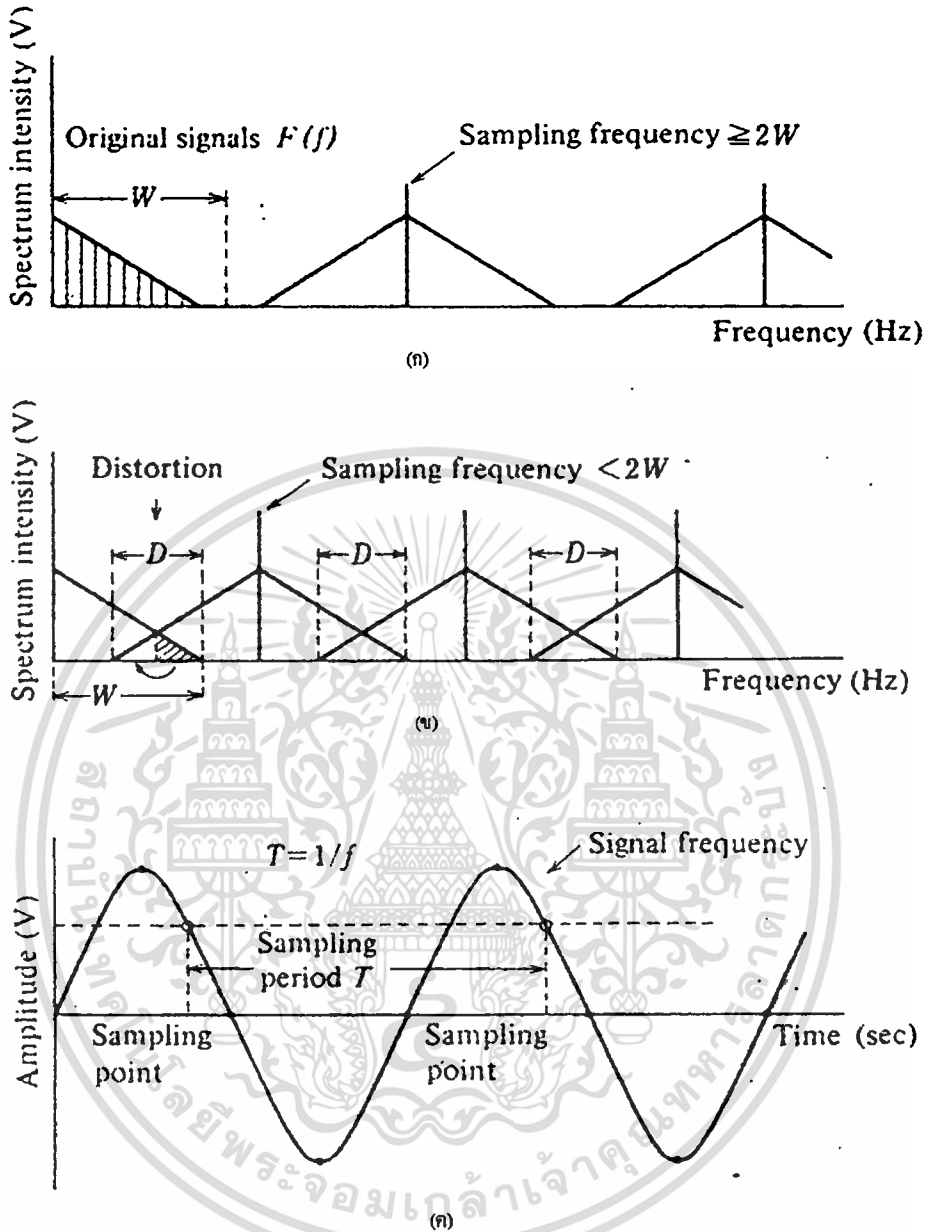
สัญญาณหลังจากการแซมปลิงเป็นแบบ ไม่ต่อเนื่องในเวลา แต่ยังคงมีค่าที่ต่อเนื่องอยู่ เรียกว่า สัญญาณไม่ต่อเนื่อง การประมวลผลของสัญญาณเหล่านี้ เรียกว่า การประมวลผลสัญญาณไม่ต่อเนื่อง และดำเนินการเป็นประจำโดยอุปกรณ์ประจุกุ (CCD) หรือ วงจร switch-capacitor

3.1.2 การควอนไทส์

การควอนไทส์ หมายถึง การประมาณค่าของค่าตัวอย่างของสัญญาณโดยการใช้หนึ่งค่าของการเลือกค่าจำกัดก่อน และเป็นอย่างไร้ระบบเทียบเท่ากับการตัดการคำนวณในตัวเลขที่แน่นอนหรือหุคที่จำนวนเต็มที่ดีที่สุด ข้อมูลบางอย่างเป็นการสูญหายซึ่งหลีกเลี่ยงไม่ได้ในการประมวลผลการควอนไทส์ การสูญเสียเรียกว่า การบิดเบี้ยวเนื่องจากการควอนไทส์ หรือสัญญาณรบกวนการควอนไทส์

เอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.2 การแซมปลิงที่ผิดเพี้ยน (ก) กรณีที่การแซมปลิงเป็นไปอย่างถูกต้อง (ข) กรณีที่การแซมปลิงเป็นไปอย่างไม่ถูกต้อง (ค) ตัวอย่างรูปคลื่นที่ไม่ถูกต้อง

ถ้าสัญญาณสมมติหยุดอยู่กับที่ (stationary) เป็นไปได้ว่าการแทนที่ เวลาเฉลี่ยของตัวอย่าง โดยค่าเฉลี่ยทั้งหมดเหล่านั้น และอัตราความผิดพลาดการควอนไทส์เป็นการกระจายแบบยูนิฟอร์ม ในช่วงจาก $-\frac{1}{2}\Delta L$ ถึง $\frac{1}{2}\Delta L$ (ΔL คือ ขนาดของขั้นการควอนไทส์ แสดงในรูปที่ 3.3 แล้ว) ค่าตั้งเฉลี่ยของสัญญาณรบกวนการควอนไทส์ \bar{Q} สามารถคำนวณได้เป็น

$$\bar{Q} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \varepsilon_k^2 \rightarrow \int_{-\Delta L/2}^{\Delta L/2} \varepsilon^2 p(\varepsilon) d\varepsilon = \int_{-\Delta L/2}^{\Delta L/2} \varepsilon^2 \frac{d\varepsilon}{\Delta L} = \frac{\Delta L^2}{12} \quad (3.2)$$

กำลังสัญญาณเฉลี่ย \bar{P}_L ระบุได้จากค่าการควอนไทล์ คือ

$$\begin{aligned} \bar{P}_L &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^{n-1} (\Delta L l_k)^2 \rightarrow \sum_{j=1}^m (\Delta L l_k)^2 p(l_j) \\ &= \Delta L^2 \sum_{j=1}^m \left(j - \frac{1}{2} \right)^2 \frac{1}{m} = \frac{\Delta L^2}{12} [(2m)^2 - 1] \end{aligned} \quad (3.3)$$

โดย $2m$ คือจำนวนรวมของค่าการควอนไทล์ที่เลือกก่อน ถ้าสัญญาณมีอยู่ในช่วงของ $\pm L$ ดังที่แสดงในรูปที่ 3.3 แล้วจะได้

$$m = \left[\frac{L}{\Delta L} \right]_{\text{int}} \quad (3.4)$$

อัตราสัญญาณต่อสัญญาณรบกวนหลังจากการควอนไทล์คือ

$$R_Q = \frac{(\Delta L^2/12) [(2m)^2 - 1]}{\Delta L^2/12} = (2m)^2 - 1 \approx (2m)^2, \quad m \gg 1 \quad (3.5)$$

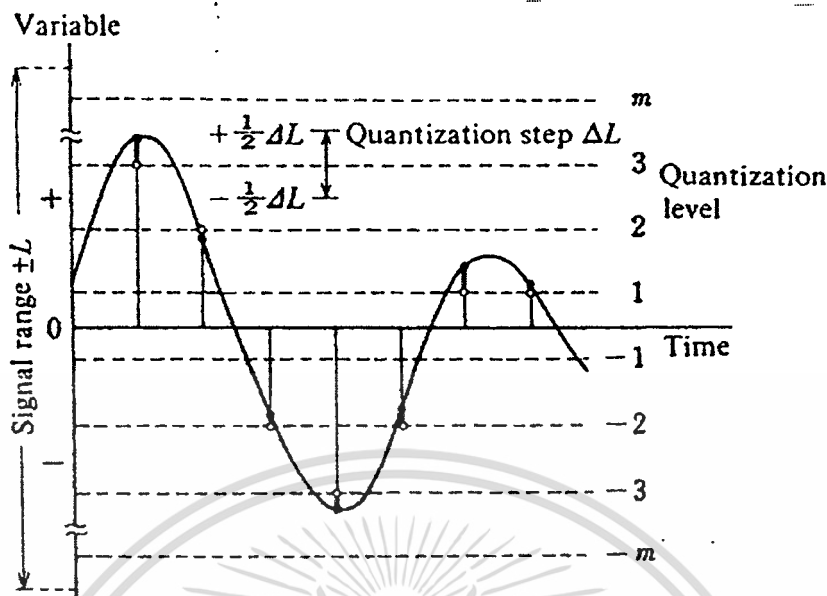
จากการเข้ารหัสโดยใช้ B เป็นบิตเอาฐานสอง $2m \leq 2^B$ เราจะได้

$$\begin{aligned} R_Q \Big|_{db} &= 10 \log_{10} R_Q = 20 \log_{10} (2m) \leq 20 \log_{10} (2^B) \\ &= B 20 \log_{10} 2 \approx 6B \text{ dB} \end{aligned} \quad (3.6)$$

ข้อแนะนำการควอนไทล์เป็นตามข้างล่างนี้

(1) ในกรณีของการประเมินค่าเสียงที่กระทบในหูของการประมวลสัญญาณดิจิทัล อัตราสัญญาณต่อสัญญาณรบกวน R_Q dB ระบุโดยการควอนไทล์ ต้องเป็นตามจำนวนอัตราสัญญาณต่อสัญญาณรบกวนที่อนุญาตให้มีได้ในการสัมผัสในปราสาทหู

(2) ในกรณีของการประเมินค่าจากเครื่องอัตราสัญญาณต่อสัญญาณรบกวนหลังจากการควอนไทล์ต้องเกินกว่าอัตราสัญญาณต่อสัญญาณรบกวนที่เริ่มต้นของสัญญาณอนาล็อก ซึ่งสามารถแนะนำโดยการเซตอัตราลำดับแรกสูงกว่าอัตราส่วนเริ่มต้น เพื่อหลีกเลี่ยงผลกระทบการควอนไทล์



รูปที่ 3.3 กระบวนการควอนไทส์ (●) ตัวอย่าง (○) หลังจากทำการควอนไทส์

(i) การควอนไทส์ที่ผิดพลาด

3.1.3 การแปลงอนาลอกเป็นดิจิทัล

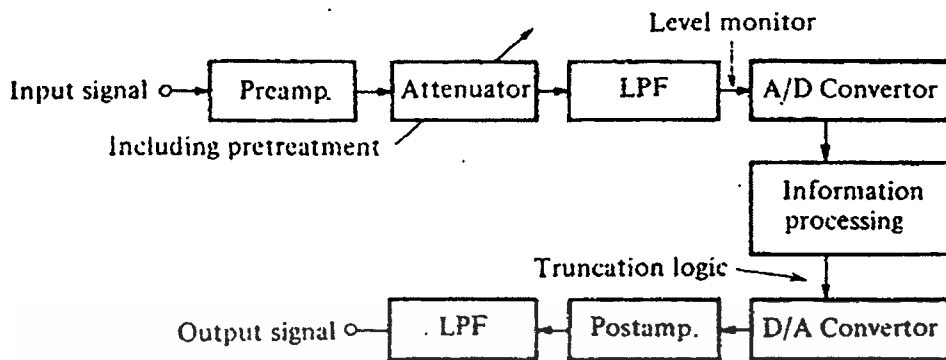
วิธีการประมวลผลสัญญาณโดยทั่วไป การแปลงจากอนาลอกเป็นดิจิทัล (A/D) และดิจิทัลเป็นอนาลอก (D/A) แสดงในรูปที่ 3.4

ในกรณีการแปลงดิจิทัลเป็นอนาลอก ช่วงบิตของการประมวลผลสัญญาณ โดยทั่วไปใหญ่กว่าการแปลงดิจิทัลเป็นอนาลอก แล้วช่วงที่เป็นผลสำหรับการแปลงดิจิทัลเป็นอนาลอก ควรจะได้รับเลือกอย่างเหมาะสม กระบวนการนี้เรียกว่า การตัดเชิงตรรกะ ความสัมพันธ์ระหว่างความถี่การแซมปลิง และคุณลักษณะของ LPF ซึ่งจะจัดความผิดเพี้ยนความถี่สูงเป็นสิ่งที่เหมือนกันทั้งสองของกระบวนการแปลงจากดิจิทัลเป็นอนาลอก และ อนาลอกเป็นดิจิทัล

ในการประมวลผลสัญญาณเสียงพูด มีสัญญาณ 2 ชนิด คือ สัญญาณเสียงพูดในแบนด์โทรศัพท์ และสัญญาณเสียงที่แม่นยำสูง ชนิดของพารามิเตอร์ของการแปลงอนาลอกเป็นดิจิทัล เป็นดังนี้

- (1) สำหรับเสียงพูดในแบนด์โทรศัพท์ : 3 dB LPF ที่ 3.4 kHz กับการลดทอนมากกว่า 24 dB / oct ความถี่การแซมปลิง 8–10 kHz และ 10–12 bit กับการแปลงเชิงเส้น
- (2) สำหรับเสียงที่มีความชัดสูง : 3 dB LPF ที่ 5-8 kHz กับการลดทอนมากกว่า 48 dB / oct ความถี่การแซมปลิง 16-20 kHz และ 12-14 bit กับการแปลงเชิงเส้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.4 แผนภาพของกระบวนการประมวลผลสัญญาณดิจิทัล

นอกจากนั้น 6 dB/oct จะถูกใช้เป็นการประจักษ์ก่อนที่จะแปลงอนาลอกเป็นดิจิทัล จะลดลงในช่วงที่เปลี่ยนแปลงตลอดของสัญญาณเสียงพูด เพื่อปรับปรุงอัตราส่วนของสัญญาณต่อสัญญาณรบกวนให้ดีขึ้น

วงจรที่ใช้อยู่จำเป็นอย่างยิ่งสำหรับการความเร็วสูงของการแปลงอนาลอกเป็นดิจิทัล เป็นสิ่งที่จำเป็นที่จะต้องกำหนดช่วงของระดับสัญญาณอินพุทภายในช่วงที่ยอมรับได้ของการแปลงอนาลอกเป็นดิจิทัลอย่างเหมาะสม กับตัวแปลงอนาลอกเป็นดิจิทัลที่ได้ จำนวนบิตของตัวแปลงอนาลอกเป็นดิจิทัล ควรจะใหญ่พอและตัวลดทอนควรจะใส่ในด้านหน้าของ LPF เพื่อปรับระดับอินพุท ให้ไม่มีการสั่นเกิดขึ้น

ช่วงของการแปลงอนาลอกเป็นดิจิทัล ของสัญญาณเสียงพูดต้องครอบคลุม 10 bit เพราะว่าช่วงไดนามิกของสัญญาณเสียงพูดเป็นการประมาณค่าที่ 55 dB ปัจจัยสูงสุดของสัญญาณเสียงพูดเป็น 12 dB ซึ่งอินพุทที่เป็นผลของระดับสัญญาณเสียงพูดควรจะเซตเป็น 12 dB ต่ำกว่าค่ามากที่สุดที่ยอมรับได้ของระดับอินพุทที่สูงสุด Automatic gain control (AGC) หรือ ความหมายอื่นของการบีบอัดของช่วงไดนามิก ของสัญญาณเสียงพูดอย่างไม่เป็นเชิงเส้น เป็นการไม่พิจารณาถึงผลกระทบหรือความต้องการสำหรับการประมวลผลสัญญาณเสียงพูดอย่างเป็นเชิงเส้น

3.2 การแปลงสัญญาณแบบ DFT และ FFT

การแปลงระหว่างสัญญาณในเชิงความถี่ และเชิงเวลามีอยู่หลายแบบ ซึ่งทุกแบบก็มีวัตถุประสงค์เดียวกัน คือ ต้องการแปลงระหว่างสัญญาณในเชิงเวลา กับสัญญาณในเชิงความถี่ (หรือสเปกตรัม) จุดที่แตกต่างกันของการแปลงแต่ละแบบก็คือ คุณลักษณะของสัญญาณที่จะแปลงเท่านั้น เช่น การแปลงฟูริเยร์จะใช้กับสัญญาณในเชิงเวลาที่มีความต่อเนื่อง และมีพลังงานจำกัด ไม่วาร์กณใดๆทั้งสิ้น อีกทั้งห้ามมีที่ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในขณะที่อนุกรมฟูรีเยร์ใช้กับในเชิงเวลาที่มีความต่อเนื่อง เป็นคาบ และมีพลังงานไม่จำกัด สรุปได้ดังตารางที่ 3.1 และ ตารางที่ 3.2

ตารางที่ 3.1 การแปลงแบบต่างๆ กับคุณลักษณะของสัญญาณที่เกี่ยวข้อง

การแปลงฟูรีเยร์ Fourier Transform (FT) เชิงเวลาต่อเนื่อง \longleftrightarrow เชิงความถี่ต่อเนื่อง	อนุกรมฟูรีเยร์ Fourier Series (FS) เชิงเวลาต่อเนื่อง \longleftrightarrow เชิงความถี่ไม่ต่อเนื่อง และเป็นคาบ
การแปลงฟูรีเยร์แบบเวลาไม่ต่อเนื่อง Discrete Time Fourier Transform (DTFT) เชิงเวลาไม่ต่อเนื่อง \longleftrightarrow เชิงความถี่ต่อเนื่อง และเป็นคาบ	การแปลงฟูรีเยร์แบบเวลาไม่ต่อเนื่อง Discrete Fourier Series (DFS) หรือ Discrete Fourier Transform (DFT) เชิงเวลาไม่ต่อเนื่อง \longleftrightarrow เชิงความถี่ไม่ต่อเนื่อง และเป็นคาบ

ถ้าไม่ระบุว่าเป็นคาบหมายถึงสัญญาณที่ไม่เป็นคาบ และมีพลังงานจำกัด (ที่อนันต์มีค่าเป็นศูนย์)

3.2.1 การแปลง DFT (Discrete Fourier Transform)

จากตารางที่ 3.1 จะเห็นได้ว่า มีเพียงการแปลง DFT เท่านั้นที่มีทั้งสัญญาณในเชิงเวลา และในเชิงความถี่เป็นแบบไม่ต่อเนื่อง จุดนี้เป็นจุดที่สำคัญมาก เพราะมันบ่งบอกว่า เราสามารถจะกระทำการแปลงนี้ได้โดยใช้การคำนวณ (การคูณและการบวก) ทางดิจิทัลได้ ซึ่งสามารถประยุกต์ได้สะดวกมากในคอมพิวเตอร์ หรือในฮาร์ดแวร์โดยตรงก็ได้ การแปลงแบบอื่นมีสัญญาณแบบต่อเนื่องเกี่ยวข้องด้วย ซึ่งทำให้การแปลงต้องใช้วิธีอินทิเกรตซึ่งยุ่งยากกว่ามาก

DFT มีความเหมือนกันกับ DFS มาก ทั้ง DFT กับ DFS มีสมการในการแปลงเหมือนกัน จุดที่ต่างกันก็คือที่มาและความหมายของทั้งสอง DFS คือ อนุกรมฟูรีเยร์แบบไม่ต่อเนื่อง (Fourier Series) ใช้ในกรณีที่มีสัญญาณเชิงเวลาเป็นแบบไม่ต่อเนื่อง และเป็นคาบ ซึ่งก็จะได้ว่าสัญญาณในเชิงความถี่จะเป็นแบบไม่ต่อเนื่อง และเป็นคาบเช่นเดียวกัน ส่วน DFT เป็นการนำเอาความจริงที่เกิดขึ้นจาก DFS มาใช้ นั่นคือ

1. สัญญาณทั้งในเชิงเวลา และความถี่เป็นแบบไม่ต่อเนื่อง
2. สัญญาณเป็นรายคาบทั้งในเชิงเวลาและความถี่ สามารถทำการแปลงโดยการใช้ค่าที่เกิดขึ้นใน 1 คาบเท่านั้น (สังเกตว่าในสูตรจะเป็นการหาผลรวมของสัญญาณในตำแหน่งที่ 0 ถึง $N-1$ เท่านั้น) ดังนั้น จำนวนค่าที่นำมาคำนวณในการแปลงไป และแปลงผกผันจึงมีความจำกัด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.2 สรุปสมการของการแปลงแบบต่างๆ

การแปลง	จากเชิงเวลาไปเป็นความถี่	จากเชิงความถี่ไปเป็นเวลา (การแปลงผกผัน)
การแปลงฟูรีเยร์ (FT)	$X(f) = \int_{-\infty}^{\infty} x(t)e^{-j\omega t} dt$	$x(n) = \int_{-\infty}^{\infty} x(f)e^{j\omega t} df$
การแปลงฟูรีเยร์แบบ เวลาต่อเนื่อง (DTFT) มีคาบในเชิงความถี่ = 2π	$X(e^{j\omega}) = \sum_{n=-\infty}^{\infty} x(n)e^{-j\omega n}$	$x(n) = \int_{-\pi}^{\pi} X(e^{j\omega'})e^{j\omega' n} \frac{d\omega'}{2\pi}$
อนุกรมฟูรีเยร์ (FS) มีคาบในเชิงเวลา = T	$c(k) = \frac{1}{T} \int_0^T x(t)e^{-jk\omega_0 t} dt$ โดยที่ $c(k)$ เป็นสัมประสิทธิ์ ของความถี่ $k\omega_0$	$x(t) = \sum_{k=-\infty}^{\infty} c(k)e^{jk\omega_0 t}$
อนุกรมฟูรีเยร์ แบบไม่ต่อเนื่อง (DFS) หรือ การแปลงฟูรีเยร์ แบบไม่ต่อเนื่อง (DFT) มี คาบในเชิงเวลา = คาบเชิงเวลา = N ตัวอย่าง	$X(k) = \sum_{n=0}^{N-1} x(n)e^{-j2\pi kn/N}$	$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k)e^{j2\pi kn/N}$

DFT ก็คือ DFS ที่เราสนใจเพียงคาบเดียวเท่านั้น ก็คือ เราสนใจว่าสัญญาณในเชิงเวลาเป็นสัญญาณไม่ต่อเนื่องมีความยาวจำกัดเท่ากับ N และเป็นสัญญาณที่มีรูปร่างใดๆก็ได้ เมื่อทำการแปลง DFT แล้ว ก็จะได้สัญญาณในเชิงความถี่เป็นสัญญาณไม่ต่อเนื่อง และมีความยาวจำกัดเท่ากับ N เท่ากัน

สมมติให้ $x(n)$ เป็นสัญญาณในเชิงเวลา และ $X(k)$ เป็นสัญญาณในเชิงความถี่ที่เกิดจาก DFT โดย k แทนตัวชี้ลำดับของสัญญาณทางด้านความถี่ ทั้งสองสัญญาณมีความยาวเท่ากัน คือ N เราจะเขียนสัญลักษณ์ได้ว่า

$$x(n) \xleftrightarrow{DFT, N} X(k)$$

จากสูตรในตารางที่ 3.2 เราจะได้ว่า $x(n)$ และ $X(k)$ มีความสัมพันธ์กันดังนี้

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-j2\pi kn/N} \quad (3.7)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับนักเรียนงานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เพื่อจัดรูปสมการให้ง่ายขึ้น ขอนิยามให้ $W_N = e^{-j2\pi/N}$ เป็นค่าที่ขึ้นกับ N เท่านั้น สำหรับในการแปลงครั้งหนึ่ง ๆ N จะมีค่าคงที่ ดังนั้น W_N จึงเสมือนเป็นค่าคงที่ เราสามารถเขียนการแปลง DFT ได้ใหม่เป็น

$$X(k) = \sum_{n=0}^{N-1} x(n)W_N^{kn} \quad (3.8)$$

หรือเขียนในรูปเมทริกซ์ได้เป็น

$$\begin{bmatrix} X(0) \\ X(1) \\ \vdots \\ X(N-1) \end{bmatrix} = \begin{bmatrix} W_N^0 & W_N^0 & W_N^0 & \dots & W_N^0 \\ W_N^0 & W_N^{1 \times 1} & W_N^{2 \times 1} & \dots & W_N^{1 \times (N-1)} \\ W_N^0 & W_N^{2 \times 1} & W_N^{2 \times 2} & \dots & W_N^{2 \times (N-1)} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ W_N^0 & W_N^{(N-1) \times 1} & W_N^{(N-1) \times 2} & \dots & W_N^{(N-1) \times (N-1)} \end{bmatrix} \begin{bmatrix} X(0) \\ X(1) \\ \vdots \\ X(N-1) \end{bmatrix} \quad (3.9)$$

3.2.1.1 ที่มาและความหมายของการแปลง DFT

จากรูปที่ 3.5 สามารถที่จะอธิบายได้ว่า

ก) ให้ $X_o(t)$ เป็นสัญญาณแบบต่อเนื่องใดๆ ที่มีความยาวจำกัด (พลังงานจำกัด) และมีสเปกตรัม (ซึ่งหามาได้จากการแปลงฟูเรียร์) คือ $X_o(f)$ สมมติว่าได้ $X_o(f)$ มีความถี่จำกัด โดยมีความถี่สูงสุดอยู่ที่ f_{\max}

ข) ถ้าเราร่วมสัญญาณ $X_o(f)$ ด้วยอัตรา f_s โดยที่ $f_s > 2f_{\max}$ จะได้สัญญาณไม่ต่อเนื่อง $X_b(n)$ ซึ่งมีสเปกตรัมคือ $X_b(e^{j\omega'})$ มีลักษณะเป็นรายคาบ โดยที่สัญญาณในช่วงความถี่ดีจิตอล $\omega' = -\pi$ ถึง π จะเหมือนกับสเปกตรัมของสัญญาณอนาลอกเดิม คือ $X_o(f)$ ในช่วง $-f/2$ ถึง $f/2$ ทุกประการ เพราะไม่เกิด aliasing ในการสุ่มครั้งนี้

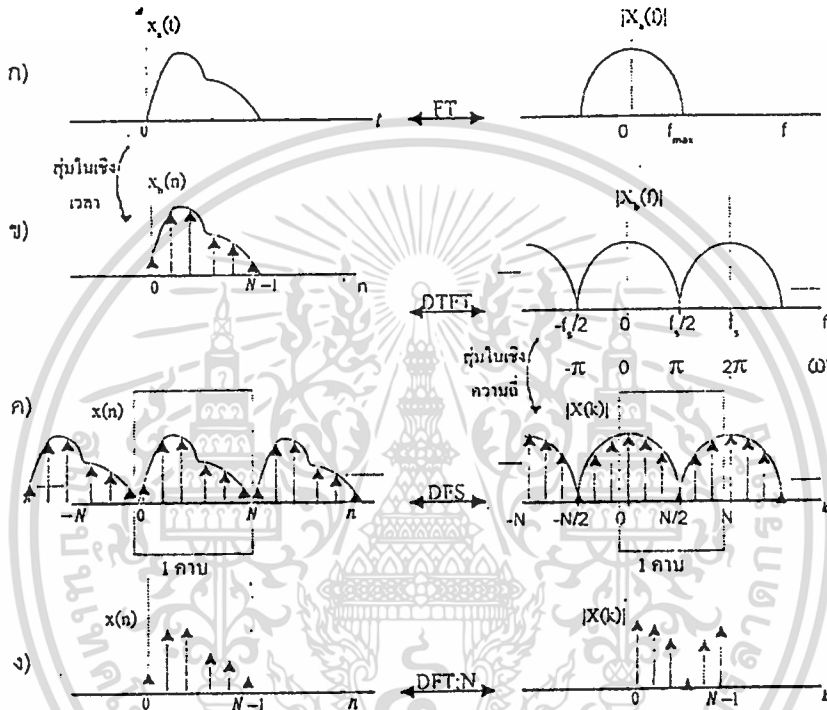
ค) ถ้าเราร่วมสัญญาณในเชิงความถี่ $X_b(e^{j\omega'})$ ด้วยความถี่ N ตัวอย่างต่อ 1 คาบ หรือมีคาบในการสุ่มเท่ากับ $2\pi/N$ เราจะได้สัญญาณในเชิงความถี่เป็นแบบไม่ต่อเนื่อง และเป็นรายคาบ ให้สัญญาณใหม่นี้เป็น $X(k)$ สิ่งที่เกิดขึ้นในเชิงเวลาก็คือ สัญญาณในเชิงเวลาจะเกิดเป็นรายคาบขึ้น (เช่นเดียวกับที่เราสุ่มสัญญาณในเชิงเวลา แล้วเกิดสำเนาของสัญญาณขึ้นในเชิงความถี่ การสุ่มสัญญาณในเชิงความถี่ ก็จะทำให้เกิดสำเนาของสัญญาณขึ้นในเชิงเวลา)

ง) กับเราดึงเอาเฉพาะสัญญาณในช่วง 1 คาบออกมาออกมาทั้งในเชิงเวลา และความถี่ คือในช่วง n และ k เท่ากับ 0 ถึง $N-1$ ส่วนนี้ก็คือ การแปลง DFS ในคาบเวลาเดียว หรือ การแปลง DFT นั่นเอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จะสังเกตได้ว่า คาบหนึ่งๆ ของสัญญาณ $x(t)$ คือ สัญญาณที่มาจากการสุ่มของสัญญาณ $x_a(t)$

นั่นก็คือ DFT สามารถใช้หาสัญญาณในเชิงความถี่แทนการแปลงฟูเรียร์ได้อย่างสมบูรณ์ โดยให้สเปกตรัมเป็นสัญญาณไม่ต่อเนื่องซึ่งจะมีรูปร่างเหมือนสเปกตรัมจริงของสัญญาณ โดยมีเงื่อนไขว่า สัญญาณที่ต้องการหาควมในเชิงความถี่ต้องมีพลังงานจำกัด และความถี่จำกัด คือ ไม่เกิน f_{\max} ซึ่งความถี่ที่ใช้ในการสุ่มเพื่อแปลงสัญญาณเป็นดิจิตอลต้องมีค่ามากกว่า $2f_{\max}$



รูปที่ 3.5 ความเกี่ยวข้องกันระหว่างการแปลง DFT กับการแปลงฟูเรียร์

ถ้าเงื่อนไขดังกล่าวเป็นจริง เราสามารถสรุปได้ว่า การแปลง DFT ให้ผลลัพธ์เป็นสัญญาณไม่ต่อเนื่องซึ่งมีค่าเท่ากับเป็นการสุ่มสเปกตรัมที่ได้จากการแปลง DTFT หรือ หมายถึงว่าค่าทุกค่าที่ได้จากการแปลง DFT จะอยู่บนเส้นของผลลัพธ์ที่ได้จากการแปลง DTFT เสมอ

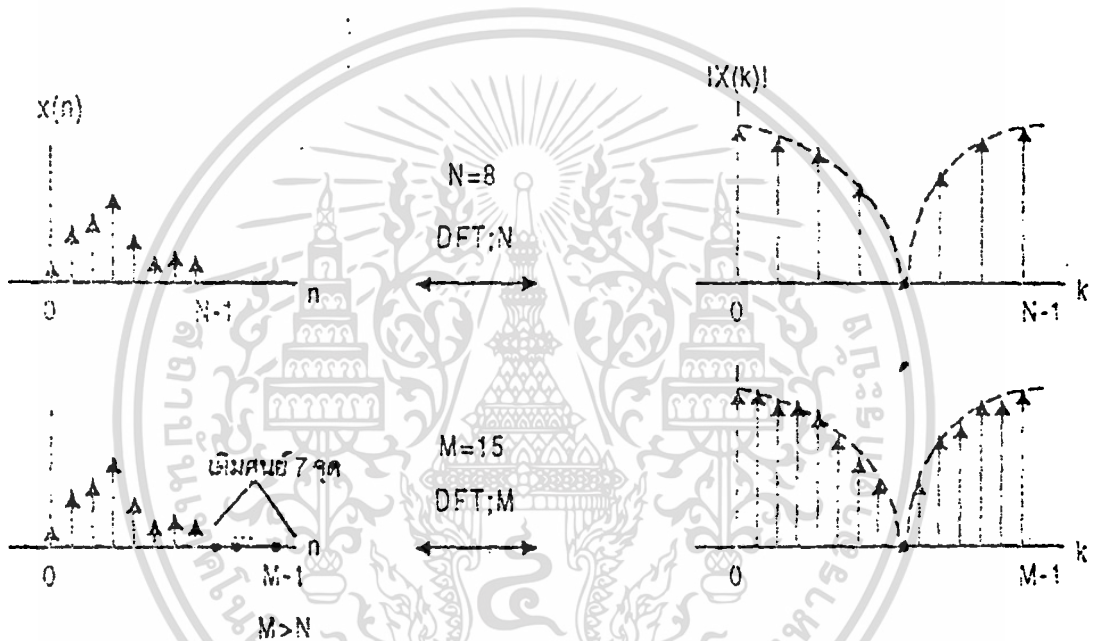
การค้นพบนี้เป็นสิ่งที่สำคัญมาก เพราะทำให้เราสามารถแปลงสัญญาณกลับไปกลับมา ระหว่างเชิงเวลา กับเชิงความถี่ได้ โดยกระทำกับสัญญาณแบบไม่ต่อเนื่องล้วนๆ ดังนั้น การแปลง DFT ก็เป็นการประมวลผลแบบดิจิตอลอย่างหนึ่งเพื่อหาสเปกตรัมของสัญญาณ

ข้อสังเกตอีกอันหนึ่งก็คือ ผลตอบสนองที่ได้จาก DFT ถ้าเทียบกับผลที่ได้จาก DTFT จะอยู่ในช่วง 0 ถึง 2π หรือคือความถี่จริงๆ ที่ 0 ถึง f_s ซึ่งผลตอบนี้เพื่อพบที่จุดกึ่งกลางจะสมมาตรกัน เราสนใจผลตอบในช่วงครึ่งแรกเท่านั้น ซึ่งคือความถี่จริงที่ 0 ถึง $f_s/2$ ซึ่งก็คือช่วง $k=0$ ถึง $N/2$ ในกรณีที่ N เป็นเลขคู่ และ $k=0$ ถึง $(N+1)/2$ ในกรณีที่ N เป็นเลขคี่

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2.1.2 การเติมศูนย์ (zero padding)

การเติมศูนย์ เป็นการเติมจุดที่มีค่าเป็นศูนย์ต่อท้ายเข้าไปในสัญญาณ $x(n)$ ก่อนที่จะทำการแปลง DFT ซึ่งจะส่งผลให้สเปกตรัมที่ได้มีจำนวนจุดมากขึ้น ซึ่งเสนอเป็นการซูมสเปกตรัมด้วยจำนวนจุดที่มากขึ้น การเติมศูนย์ช่วยให้มองเห็นรูปร่างได้ละเอียด และชัดเจนขึ้น แต่ในทางทฤษฎีแล้ว ไม่ได้เป็นการเพิ่มข้อมูลใดๆ ให้แก่สัญญาณเลย เส้นประที่แสดงในรูปที่ 3.6 คือ เส้นที่แสดงผลลัพธ์ที่เกิดจากการแปลง DFT ซึ่งไม่ว่าเราจะเติมศูนย์เข้าไปมากเท่าไรเส้นนี้ก็จะคงเดิม กล่าวคือ ผลลัพธ์ที่เราได้ดีที่สุดจากการเติมศูนย์มากขึ้นๆ ก็คือผลที่เข้าใกล้ผลของการแปลง DFT นั่นเอง



รูปที่ 3.6 ผลของการเติมศูนย์กับการแปลง DFT

สเปกตรัมของพลังงานและสเปกตรัมของกำลัง

ผลลัพธ์ที่เกิดจากการแปลง DTFT หรือ DFT ที่ได้กล่าวถึงมานี้ เรียกว่า สเปกตรัมทางขนาด (หรือสเปกตรัมของโวลท์เทจ) ถ้าเรามีสมมติฐานว่าสัญญาณที่นำมาหาสเปกตรัมนี้เกิดขึ้นและสิ้นสุดภายในช่วงที่นำมาคิดเท่านั้น หรือที่เวลาอื่นๆ สัญญาณมีค่า สัญญาณประเภทนี้เรียกว่า สัญญาณที่มีพลังงานจำกัด (finite-energy signal) สัญญาณที่มีพลังงานจำกัดมีสเปกตรัมที่เราสนใจคือ สเปกตรัมของพลังงาน (energy spectrum) ซึ่งสามารถหาได้จาก

$$S_x(k) = |X(k)|^2 \quad (3.10)$$

โดย $S_x(k)$ คือ สเปกตรัมของพลังงานของสัญญาณ $x(n)$ และ $X(k)$ คือ ผลการแปลง DFT ของสัญญาณ $x(n)$ สำหรับการแปลง DTFT ก็มีสูตรในลักษณะเดียวกัน ก็จะได้ $S_x(\omega) = |X(e^{j\omega})|^2$ ซึ่งสามารถพิสูจน์ได้ [2]

แต่ถ้าเรามีสมมติฐานว่า สัญญาณที่นำมาหาสเปกตรัมนี้เป็นเพียงส่วนย่อยหนึ่งของสัญญาณที่มีคุณลักษณะทางสถิติไม่แปรตามเวลา (stationary signal) สัญญาณหนึ่ง กล่าวคือ สัญญาณนี้ยาวไปจนถึงเวลาเป็นอนันต์ แต่เราตัดเอาเพียงส่วนหนึ่งมาดูเท่านั้น ซึ่งสัญญาณนี้อาจเป็นสัญญาณที่มีรูปร่างไม่แน่นอน (random signal) แต่ขอให้มันมีคุณลักษณะทางสถิติที่คงที่ เช่น มีค่าเฉลี่ยคงที่ และมีกำลังคงที่

สัญญาณประเภทนี้เรียกว่า สัญญาณที่มีพลังงานไม่จำกัด ซึ่งไม่สามารถหาค่าพลังงานได้ แต่สามารถหาค่ากำลังเฉลี่ยของสัญญาณได้ ดังนั้นสเปกตรัมที่เราสนใจของสัญญาณประเภทนี้ คือ สเปกตรัมของกำลัง (power spectrum) ตามสมการ 3.11

$$P_x(k) = \frac{1}{N} |X(k)|^2 \quad (3.11)$$

สเปกตรัมของกำลังเป็นตัวบอกว่า สัญญาณมีกำลัง (หรือพลังงาน) กระจายอยู่ในความถี่ต่างๆ อย่างไร ถ้าเราบวกค่าทุกค่าของ $S_x(k)$ เข้าด้วยกัน ก็จะได้กำลังรวมของสัญญาณในทุกๆ ความถี่ ซึ่งก็ควรจะเท่ากับกำลังเฉลี่ยของสัญญาณนั่นเอง ซึ่งความจริงข้อนี้มีกล่าวในทฤษฎีบทของ Parseval ที่ว่า กำลังเฉลี่ยในทางเวลา จะเท่ากับกำลังเฉลี่ยในทางความถี่ ซึ่งเขียนเป็นสูตรสำหรับสัญญาณไม่ต่อเนื่อง ได้ว่า

$$\sum_{n=0}^{N-1} |x(n)|^2 = \frac{1}{N} \sum_{k=0}^{N-1} |X(k)|^2 \quad (3.12)$$

มักนิยมแสดง สเปกตรัมของกำลังในหน่วย dB ซึ่งมีสูตรว่า

$$P_x(k) = 10 \log \left(\frac{1}{N} |X(k)|^2 \right) \quad (3.13)$$

สังเกตว่าสเปกตรัมของกำลังที่แสดงในหน่วย dB จะสามารถแสดงให้เห็นถึงในรายละเอียดของสเปกตรัมได้ดีกว่าโดยเฉพาะในย่านที่มีกำลังของสัญญาณต่ำๆ และโดยทั่วไปแล้ว สเปกตรัมของกำลังเป็นค่าที่นิยมมากกว่าสเปกตรัมของขนาด ถึงแม้สัญญาณจะไม่เป็นแบบ stationary ก็ตาม ดังจะได้กล่าวถึงตัวอย่างในส่วนหลังในเรื่องเครื่องวิเคราะห์สเปกตรัม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2.2 การแปลง FFT (Fast Fourier Transform)

เนื่องจาก การแปลง DFT มีประโยชน์ในการใช้งานมาก จึงมีได้มีความพยายามคิดค้นหาวิธีที่จะคำนวณ DFT ให้เร็วขึ้น และมีประสิทธิภาพขึ้นกว่าปกติ การแปลง FFT ก็คือวิธีการคำนวณ DFT อย่างรวดเร็ว กว่าความคิดปกติ เพราะฉะนั้น เมื่อกล่าวถึงการแปลง FFT โดยหลักการแล้วก็คือ การแปลง DFT นั้นเอง และการแปลง FFT ไม่ใช่การแปลงชนิดใหม่แต่อย่างใด

การคำนวณ DFT โดยตรงจากนิยาม ถ้าสัญญาณมีความยาวเท่ากับ N จะต้องใช้การคำนวณถึงประมาณ N^2 CMAC, (CMAC คือ Complex Multiplication and Accumulation เป็นหน่วยวัดการคำนวณ ซึ่ง 1 CMAC เท่ากับการกระทำทางคณิตศาสตร์ที่ประกอบด้วย การคูณเลขเชิงซ้อน 2 จำนวน เสร็จแล้วนำเอาผลลัพธ์ที่ได้ไปบวกสมทบเข้ากับเลขเชิงซ้อนอีกจำนวนหนึ่ง) ซึ่งมีค่าที่มาก โดยเฉพาะเมื่อ N มีค่าสูงๆ การใช้ FFT จะช่วยลดจำนวน CMAC ที่ต้องใช้ลงได้มาก

ในปัจจุบันได้มีผู้คิดค้นการคำนวณ DFT อย่างรวดเร็วได้หลายวิธี คำว่า FFT เป็นชื่อกลางๆ ที่ไม่ได้บ่งบอกว่าเป็นวิธีไหน วิธีทำ FFT วิธีพื้นฐานวิธีหนึ่ง คือ วิธี radix-2 แบบ decimation-in-time (แตกเป็นส่วนย่อยทางฝั่งเวลา) ย้อนกลับไปดูการแปลง DFT ในสมการ 3.8 คือ

$$X(k) = \sum_{n=0}^{N-1} x(n)W_N^{kn}$$

โดยที่ $W_N = e^{-j2\pi/N}$ และ $k = 0, 1, \dots, N-1$

ถ้าให้ N เป็นเลขคู่ เราสามารถกระจาย $X(k)$ ให้อยู่ในรูปของผลบวกของเทอมที่ n เป็นคู่และเทอมที่ n เป็นคู่ได้ ดังสมการ 3.14

$$X(k) = \underbrace{\sum_{n=0}^{\frac{N}{2}-1} x(2n)W_N^{2nk}}_{\text{เทอมคู่}} + \underbrace{\sum_{n=0}^{\frac{N}{2}-1} x(2n+1)W_N^{(2n+1)k}}_{\text{เทอมคี่}} \quad (3.14)$$

$$X(k) = \sum_{n=0}^{\frac{N}{2}-1} x(2n)W_N^{2nk} + \sum_{n=0}^{\frac{N}{2}-1} x(2n+1)W_N^{2nk}W_N^k \quad (3.15)$$

ถ้าพิจารณาเทอม W_N^{ab} ที่มี a และ b เป็นจำนวนใดๆ ที่ไม่เท่ากับ 0 จะพบว่าเราสามารถย้ายตัวยกกำลังของ W ไปเป็นตัวหารของ N ได้ดังสมการ 3.16

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$W_N^{ab} = e^{-j\frac{2\pi}{N}ab} = e^{-j\frac{2\pi}{N/b}ab} = W_{N/b}^a \quad (3.16)$$

เราใช้ความจริงข้อนี้ แทนค่าเทอม W_N^{2nk} ด้วย $W_{N/2}^{nk}$ ในสมการ 3.15 จะได้

$$X(k) = \underbrace{\sum_{n=0}^{N/2-1} x(2n)W_{N/2}^{nk}}_{\text{DFT } N/2 \text{ จุด}} + \underbrace{\sum_{n=0}^{N/2-1} x(2n+1)W_{N/2}^{nk}}_{\text{DFT } N/2 \text{ จุด}} \times \underbrace{W_N^k}_{\substack{\text{สัมประสิทธิ์บิด} \\ \text{ใช้คุณสมบัติ}}} \quad (3.17)$$

จะเห็นได้ว่า $X(k)$ ได้กลายเป็นผลบวกของสองเทอม แต่ละเทอมเป็นรูปแบบของการคำนวณ DFT $N/2$ จุด โดยเทอมแรกกระทำกับสัญญาณ $x(0), x(2), \dots, x(N-2)$ และเทอมที่สองกระทำกับสัญญาณ $x(1), x(3), \dots, x(N-1)$

ถ้าเราดูการแตกกระจายเป็นเทอมย่อยแต่เพียงเท่านี้ และคำนวณ DFT โดยใช้สมการ 3.17 จะได้ว่า เราต้องคำนวณ DFT $N/2$ จุด เป็นจำนวน 2 ชุด ซึ่งแต่ละชุดจะต้องใช้จำนวน $CMAC$ ในการคำนวณเท่ากับ $(N/2)^2$ ดังนั้น ต้องใช้จำนวน $CMAC$ ในการคำนวณทั้งสิ้นประมาณ $2\left(\frac{N}{2}\right)^2 = \frac{N^2}{2}$ (จริงๆ แล้ว ต้องใช้การบวกอีก N จุด เพื่อนำผลลัพธ์ของแต่ละชุดมาบวกกัน แต่เนื่องจาก ถ้า N ใหญ่พอประมาณ N จะมีค่าน้อยเมื่อเทียบกับ $\frac{N^2}{2}$ จึงประมาณว่าไม่ต้องคิดการบวก N ครั้งนี้ได้)

สรุปว่า การหา $W(k)$ ซึ่งเป็น DFT N จุด สามารถกระจายให้อยู่ในเทอมของ DFT $N/2$ จุด ซึ่งจะทำให้จำนวน $CMAC$ ที่ต้องใช้ลดลงประมาณครึ่งหนึ่ง เช่นเดียวกัน ถ้าเราทำการแตกเทอม DFT $N/2$ จุดที่อยู่ในสมการ 3.17 นี้ต่อไป แต่ละเทอมก็จะสามารถกระจายให้กลายเป็นผลบวกของ DFT $N/4$ จุดสองเทอม ซึ่งก็จะทำให้จำนวน $CMAC$ ที่ต้องใช้ลดลงอีกประมาณครึ่งหนึ่ง เราสามารถกระจายเช่นนี้ไปเรื่อยๆ จนกระทั่งทุกตัวอยู่ในรูปของ DFT 2 จุด ซึ่ง DFT 2 จุดสามารถคำนวณได้ง่ายๆ คือ สมมติ $x(n)$ 2 จุด จะได้

$$X(k) = \sum_{n=0}^1 x(n)W_2^{kn} \quad (3.18)$$

ซึ่ง $W_2^0 = 1$ และ $W_2^1 = e^{-j\pi} = -1$ จะได้ว่า

$$\begin{aligned} X(0) &= x(0) + x(1) \\ X(1) &= x(0) - x(1) \end{aligned} \quad (3.19)$$

ขั้นตอนที่ได้อธิบายทั้งหมดนี้รวมเรียกว่า การแปลง FFT เรามักเขียนวิธีคำนวณ FFT โดยใช้แผนภาพเรียกว่า แผนภาพผีเสื้อ (butterfly diagram)[3]

3.2.3 การแปลง DFT ผกผัน (IDFT, Inverse Discrete Fourier Transform)

พิจารณาสูตรของ IDFT เทียบกับ DFT จะพบว่ามีความคล้ายกันมาก ซึ่งก็พบว่า การหา IDFT สามารถหาได้โดยการใช้ DFT ดังสมการ 3.2

$$x = IDFT(X) = \frac{1}{N} (DFT(X^*))^* \quad (3.20)$$

เครื่องหมาย * หมายถึง conjugate ลองพิสูจนสูตรนี้ โดยใช้สมการของ DFT และแทนค่า X^* ลงไป จะได้

$$DFT(X^*) = \sum_{k=0}^{N-1} X(k)^* e^{-j2\pi kn/N}$$

ดังนั้นจะได้

$$\frac{1}{N} (DFT(X^*))^* = \frac{1}{N} \sum_{k=0}^{N-1} \{X(k)^* e^{-j2\pi kn/N}\}^* = \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{j2\pi kn/N}$$

ซึ่งมีค่าเท่ากับ $IDFT(X) = x(n)$ จริงตามสมการ 3.20

ในทำนองเดียวกัน IFFT ซึ่งมีความหมายเดียวกับ IDFT และสามารถหาได้จากการคำนวณ FFT ดังสมการต่อไปนี้

$$x = IFFT(X) = \frac{1}{N} (FFT(X^*))^* \quad (3.21)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คุณสมบัติของ DFT

สมมติให้ $X_1 = DFT(x_1)$ และ $X_2 = DFT(x_2)$ จะได้ว่า

1. คุณสมบัติความเป็นเชิงเส้น

$$a_1x_1(n) + a_2x_2(n) \xrightarrow{DFT:N} a_1X_1(k) + a_2X_2(k) \quad (3.22)$$

2. คุณสมบัติการเลื่อนทางเวลา

$$x_1(n-m) \xrightarrow{DFT:N} e^{-j2\pi m/N} X_1(k) \quad (3.23)$$

3.3 การวิเคราะห์สเปกตรัมเวลาสั้น ๆ

ในการวิเคราะห์เสียงพูด การวิเคราะห์สเปกตรัมเวลาสั้น ๆ หรือการรับค่าเฉลี่ยสเปกตรัม เป็นหลักการที่ใช้สำหรับ ; (1) สเปกตรัมควรจะถูกวิเคราะห์ในช่วงเวลา stationary (การวิเคราะห์ quasi-stationary) และ (2) การเปลี่ยนแปลงสเปกตรัมเกี่ยวกับเวลาอย่างช้า แสดงสมาชิกที่สำคัญมากของข้อมูลเสียงพูด (ความสำคัญของสเปกตรัมเปลี่ยนแปลงในเวลา)

ในทางทฤษฎีการวิเคราะห์การแปลงฟูเรียร์ การอินทิเกรตเป็นการดำเนินการในช่วงไม่จำกัด อย่างไรก็ตามการวิเคราะห์ต้องถูกดำเนินการจากจำนวนที่จำกัดของตัวอย่างและควรจะเป็นการวิเคราะห์สเปกตรัมเวลาสั้น ๆ ความสัมพันธ์ถูกแสดงในเชิงคณิตศาสตร์ดังสมการ 3.24

$$F_T(\omega) = \int_{-\infty}^{\infty} f(t)w(t)e^{-j\omega t} dt = \int_{-T}^T f(t)w(t)e^{-j\omega t} dt \quad (3.24)$$

โดยที่ $w(t)$ คือ ฟังก์ชันการถ่วงน้ำหนักที่กำหนดช่วงเวลาจำกัดของการอินทิเกรตจาก $-T$ จนถึง $+T$ บน $f(t)$ และเรียกว่าเป็นฟังก์ชันวินโดว์เวลา (time-window) และเรียก $F_T(\omega)$ ว่าเป็นสเปกตรัมเวลาสั้น ๆ

สมมติว่าสเปกตรัมเชิงทฤษฎีของ $f(t)$ เป็น $F(\omega)$ และของ $w(t)$ คือ $W(\omega)$ จะได้

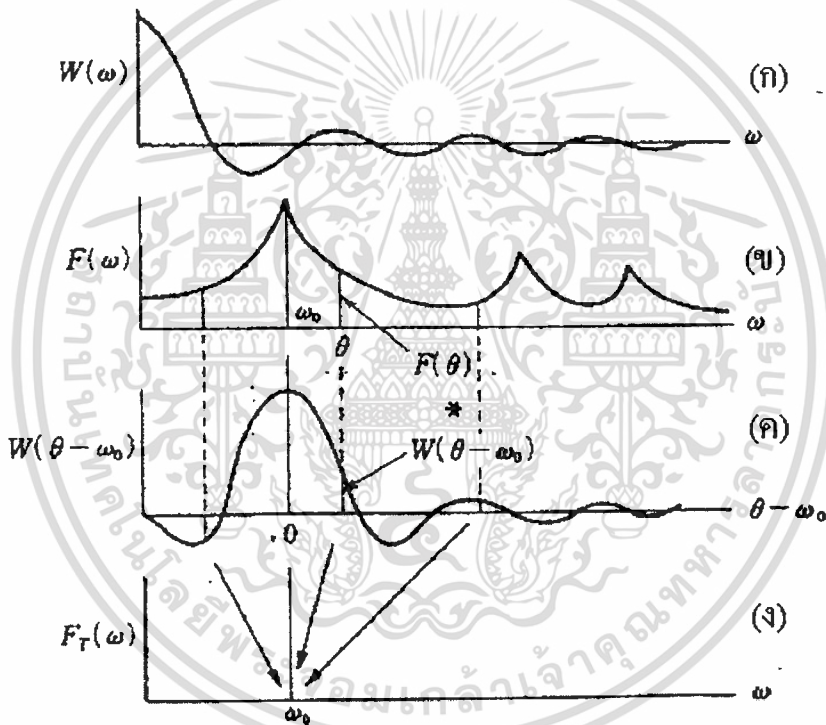
$$F_T(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\theta)W(\theta - \omega)d\theta \quad (3.25)$$

ซึ่งหมายความว่า $F_T(\omega)$ เป็นการคอนโวลูชันของ $F(\omega)$ และ $W(\omega)$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ความสัมพันธ์แสดงในรูปที่ 3.7 เทอมของ $F_T(\omega)$ ควรจะเป็น $F(\omega)$ ได้เอง ถ้าเป็นไปได้ แต่จริงๆแล้วเป็นผลรวมของ $F(\omega)$ รวมทั้งสมาชิกสเปกตรัมด้านนอกที่ไม่ต้องการ $F(\theta)$ ถ่วงน้ำหนักโดยสเปกตรัมของฟังก์ชันวินโดว์ $W(\theta - \omega)$

ดังนั้นฟังก์ชันวินโดว์ $w(t)$ ควรจะเป็นฟังก์ชัน ซึ่งสเปกตรัม $W(\omega)$ มีคุณลักษณะตามนี้ :
 (1) ความละเอียดความถี่สูง (โลบหลักต้องแคบและคม) และ (2) การบิดเบี้ยวมีสาเหตุมาจากการอินทิเกรตนอกโลบหลัก (การตอบสนองของ $W(\omega)$ ต้องเล็ก) ความต้องการสองข้อนั้นขัดแย้งและไม่สามารถเป็นที่พอใจในเวลาเดียวกัน ดังนั้น $w(t)$ ควรจะเลือกที่ความต้องการเหล่านั้น เข้ากันได้เหมาะสมเท่าที่เป็นไปได้ ความสัมพันธ์ที่เหมือนกันถืออยู่ใน DFT - FFT ของตัวอย่างสัญญาณไม่ต่อเนื่อง และ $W(n)$ ของ $0 \leq n \leq N-1$



รูปที่ 3.7 การวิเคราะห์สเปกตรัมในช่วงสั้นๆ ของวินโดว์ (ก) สเปกตรัมของฟังก์ชันของวินโดว์เวลา (ข) สเปกตรัมสัญญาณ (ค) และ (ง) คือผลลัพธ์จากสมการ

$$F_T(\omega_0) = (1/2\pi) \int_{-\infty}^{\infty} F(\theta)W(\theta - \omega_0)d\theta$$

รูปแบบของฟังก์ชันวินโดว์แบบต่างๆมีสมการดังต่อไปนี้

1. วินโดว์ Rectangular (มุมฉาก) :

$$W(n) = 1, \quad 0 \leq n \leq N-1$$

2. วินโดว์ Barlett (สามเหลี่ยม)

$$W(n) = \begin{cases} \frac{2n}{N-1}, & 0 \leq n \leq \frac{N-1}{2}, \\ \frac{2-2n}{N-1}, & \frac{N-1}{2} \leq n \leq N-1 \end{cases}$$

3. วินโดว์ Hanning

$$W(n) = \frac{1}{2} \left\{ 1 - \cos \left(2\pi \frac{n}{N-1} \right) \right\}, \quad 0 \leq n \leq N-1$$

4. วินโดว์ Hamming

$$W(n) = 0.54 - 0.46 \cos \left(2\pi \frac{n}{N-1} \right), \quad 0 \leq n \leq N-1$$

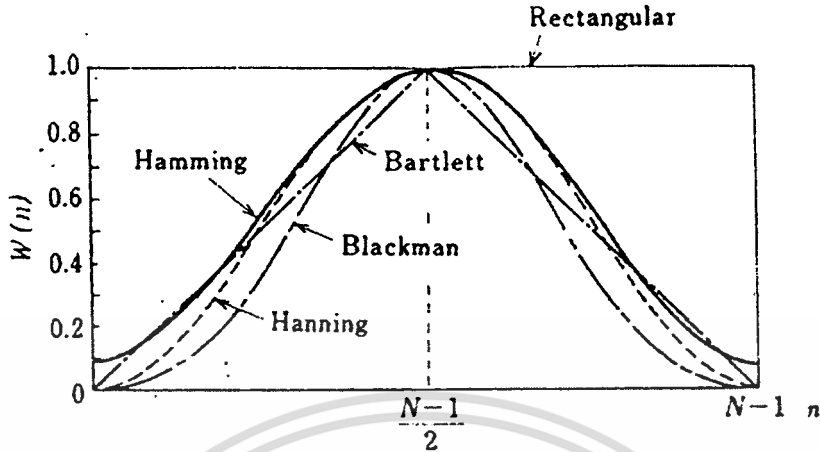
5. วินโดว์ Blackman

$$W(n) = 0.42 - 0.5 \cos \left(2\pi \frac{n}{N-1} \right) + 0.08 \cos \left(4\pi \frac{n}{N-1} \right), \quad 0 \leq n \leq N-1$$

ฟังก์ชันวินโดว์ $w(t)$ และสเปกตรัม $W(\omega)$ แสดงในรูปที่ 3.8 ซึ่งสามารถยกตัวอย่างให้เห็นถึงความแตกต่างของวินโดว์แบบต่างๆดังนี้

- (1) วินโดว์ Rectangula มีความละเอียดความถี่สูงแต่ด้านข้างโลบสูงสุด (ประมาณ -20 dB)
- (2) วินโดว์ Bartlett ไม่มีทั้งความละเอียดความถี่สูง และ side โลบต่ำ
- (3) วินโดว์ Hanning ต่ำในลำดับข้างโลบที่สูงกว่า แต่อันดับแรก และอันดับสองของด้านข้างโลบ ไม่ลดทอนต่ำกว่า 40 dB ของจุดยอดหลัก
- (4) วินโดว์ Hamming มีด้านข้างโลบต่ำกว่า -40 dB ในทุกลำดับ
- (5) วินโดว์ Blackman มีด้านข้างโลบเล็กมากน้อยกว่า -60 dB แต่ความสัมพันธ์ของความละเอียดความถี่ไม่ดีเท่าที่ควร

จากการวิเคราะห์ตามรายละเอียด การวินโดว์ที่เหมาะสม โดยปกติจะใช้วินโดว์ Hamming เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.8 ฟังก์ชันของหน้าต่างเวลา

การวิเคราะห์สเปกตรัมบนพื้นฐานช่วงเวลาจำกัดของรูปคลื่นสัญญาณ หรือลำดับของจำนวนจำกัดของค่าตัวอย่าง เรียกว่า การวิเคราะห์สเปกตรัมเวลาสั้นๆ ระหว่างปัจจัยของช่วงการแซมปลิง T วินาที ข้อมูลตัวอย่าง จำนวน N ตัว และความละเอียดความถี่ของสเปกตรัมผลลัพธ์ $\Delta f = 1/TN$ Hz สำหรับตัวอย่าง ถ้า $T = 0.1$ มิลลิวินาที (แซมปลิง 10 kHz) และ $N = 512$ แล้ว $\Delta f = 10 \times 10^3 / 512 \approx 20$ Hz

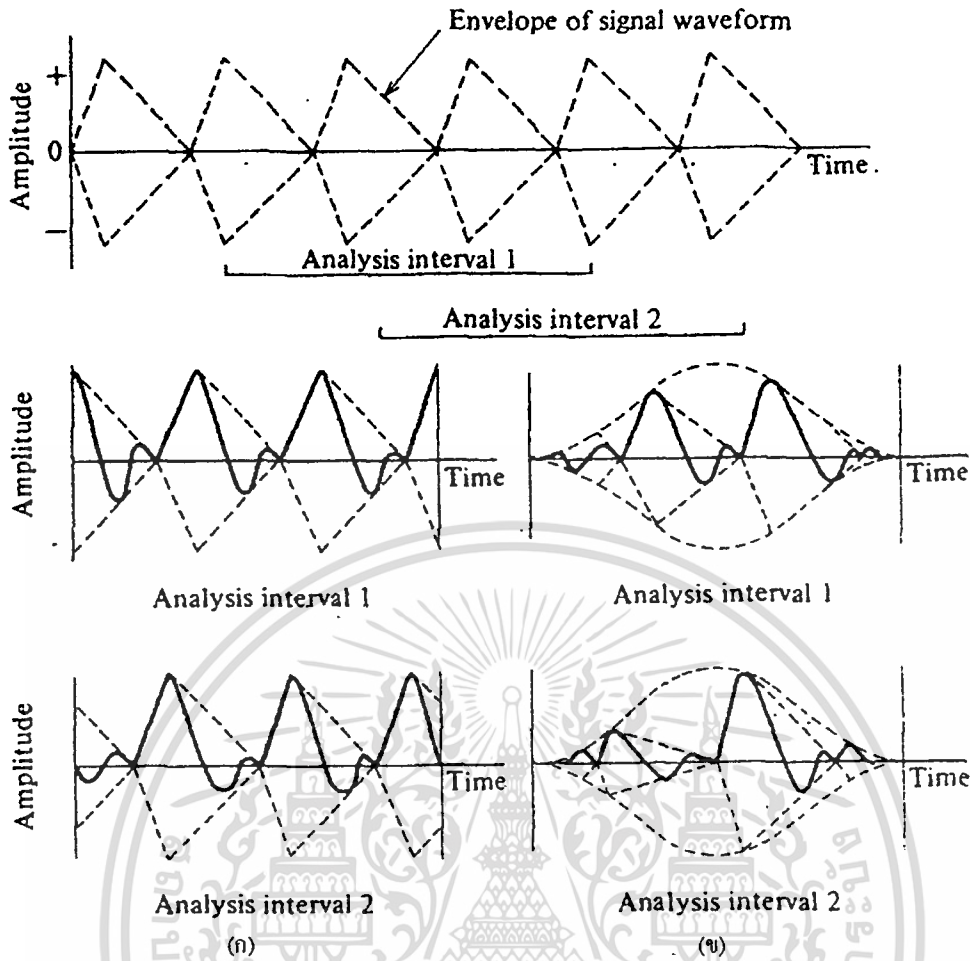
ในการวิเคราะห์สเปกตรัมของเสียงพูดโดยเฉพาะส่วนของเสียง ซึ่งเป็นช่วงอย่างสม่ำเสมอ (Periodicity) ของรูปคลื่น เนื่องจากโครงสร้างที่ตั้งชัดเจน ฟังก์ชันวินโดว์สามารถลดความผันแปรของการวิเคราะห์ผล ซึ่งเป็นการแจกแจงการเปลี่ยนในตำแหน่งของช่วงการวิเคราะห์เกี่ยวข้องกับระดับเสียงที่เปลี่ยนแปลงอย่างสม่ำเสมอ

ในวินโดว์มุลทิกแสดงในรูปที่ 3.9 (ก) ขอบที่มีผลกระทบเป็นขนาดใหญ่เนื่องจากระดับเสียงที่เปลี่ยนแปลงอย่างสม่ำเสมอ โดยเฉพาะอย่างยิ่งการวิเคราะห์เวลาสั้นๆ ของช่วงระดับเสียงน้อยกว่าสองหรือสามช่วง จากการประยุกต์วินโดว์ Hanning ค่าแสดงในรูปที่ 3.9 (ข) ถ้าวินโดว์เป็นศูนย์ที่จุดสิ้นสุดทั้งคู่ ซึ่งลดขอบผลกระทบ ยิ่งไปกว่านั้นในการติดตามก็ต่อเนื่องของการเคลื่อนที่สเปกตรัมที่ความเร็วสูงอย่างสัมพันธ์กันมีการคาบเกี่ยวกันของช่วงการวิเคราะห์ถูกใช้เพื่อรักษาความต่อเนื่องเป็นประจำ ดังแสดงในรูปที่ 3.10

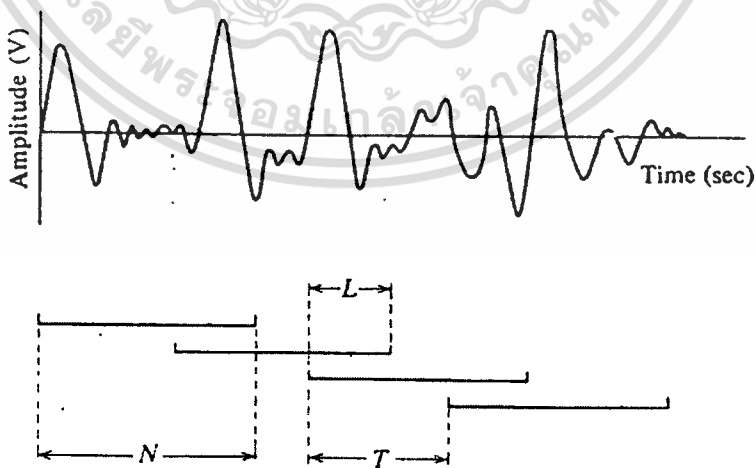
การวิเคราะห์เสียงพูด ที่ผ่านมา ความยาวเวลาวินโดว์อยู่ที่ 10 ถึง 30 มิลลิวินาที และการข้ามช่วงเวลาการวิเคราะห์ (อัตราเฟรม) อยู่ที่ 10 ถึง 20 มิลลิวินาที เมื่อใช้กับวินโดว์ที่เป็น long-time ความละเอียดความถี่สูงพอสำหรับสเปกตรัมเฉลี่ย แต่พอปรับใช้กับวินโดว์ที่เป็น short-time การวิเคราะห์ทางสเปกตรัมก็จะปรับเปลี่ยนความละเอียดของความถี่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่นอนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.9 ค่าที่ลดลงด้วยการวินโดว์ และการวิเคราะห์แบ่งออกเป็น 2 ส่วน (ก) การวินโดว์แบบ rectangular (ข) การวินโดว์แบบ Hanning

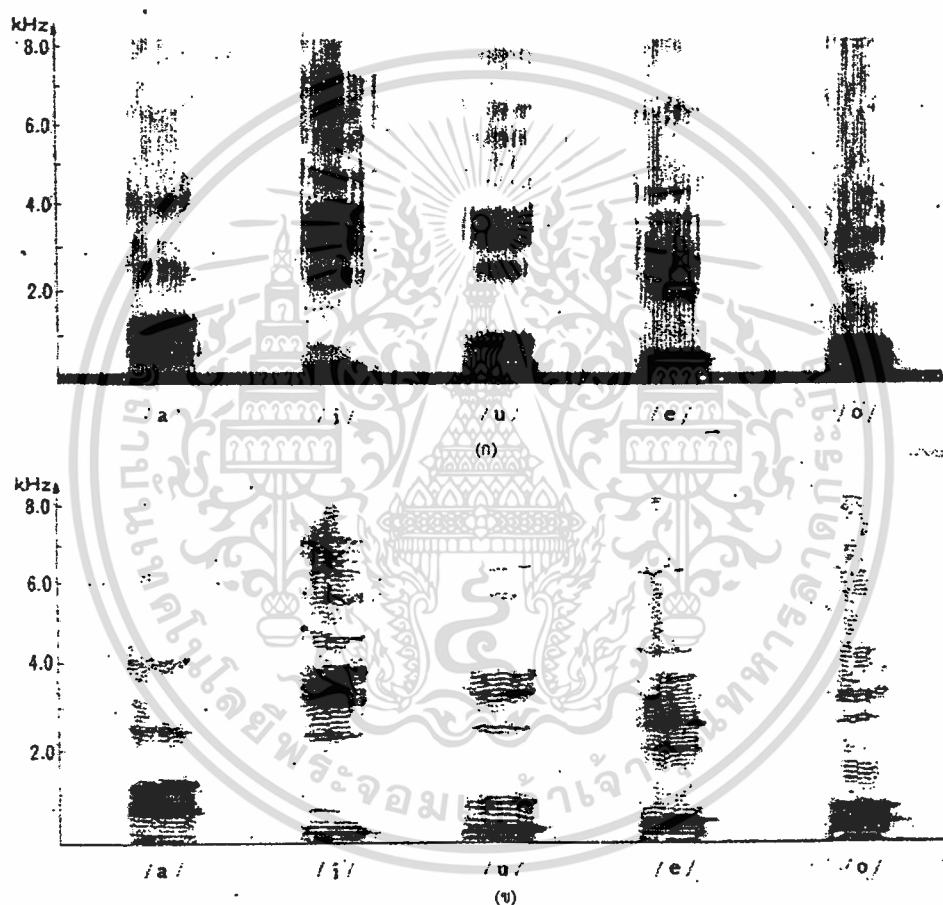


รูปที่ 3.10 การวิเคราะห์การเลื่อนอย่างต่อเนื่อง โดยที่ L คือช่วงเวลาในการเลื่อน, N คือ การวิเคราะห์ช่วงเวลาวินโดว์, T คือ การวิเคราะห์ช่วงเวลาการ shift และค่าเฟรมเรท คือ

$$F = 1/T$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ใน Sonagraph เป็นอุปกรณ์การวิเคราะห์สเปกตรัมอนุภาคที่เป็นที่รู้จักดี แต่ละชั้นใน 300 Hz วงจรกรองผ่านแบนด์กว้าง (BPF) หรือ 40 Hz แบนด์แคบ BPF เครื่องสร้างได้สอดคล้องกับ 3 มิลลิวินาที เวลาวินโดว์และส่วนที่สองเป็นวินโดว์ 22 มิลลิวินาที ตัวอย่างที่วิเคราะห์บางตัวดังแสดงในรูปที่ 3.11 ใช้การวิเคราะห์แบนด์กว้าง รูปแบบการเคลื่อนที่ความถี่เป็นที่ทราบว่าเป็นการติดตามอย่างไรก็ตามมีความละเอียดความถี่ค่อนข้างต่ำสำหรับการตรวจจับฮาร์โมนิกระดับเสียง การใช้การวิเคราะห์แบนด์แคบ ฮาร์โมนิกความถี่ระดับเสียงเป็นที่ทราบว่าจะสังเกตการรบกวนค่อนข้างต่ำของรูปการเคลื่อนที่



รูปที่ 3.11- การวิเคราะห์เสียงสระ 5 เสียง /a, i, u, e, o/ (ก) การวิเคราะห์สัญญาณแถบกว้าง (ข) การวิเคราะห์สัญญาณแถบแคบ

3.4 การหักล้างทางสเปกตรัม (Spectral Subtraction)

การหักล้างทางสเปกตรัม คือวิธีการสำหรับการหาค่าล้างของสเปกตรัมหรือขนาดของสเปกตรัมที่แท้จริง จากสัญญาณที่มีสัญญาณรบกวน รวมถึงการหักล้างสัญญาณรบกวนที่ประมาณได้ จากสเปกตรัมของสัญญาณที่ผิดเพี้ยน (noisy signal) วิธีดังกล่าวจะพบว่า สเปกตรัมไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ของสัญญาณรบกวนจะถูกประมาณค่าจากค่าของสัญญาณ เมื่อสัญญาณขาดหายและมีเพียงสัญญาณรบกวนที่ปรากฏ โดยสันนิษฐานว่าสัญญาณหยุดนิ่ง หรือเปลี่ยนไปอย่างช้ามาก และสเปกตรัมของสัญญาณรบกวนจะหยุดนิ่งหรือเปลี่ยนไปอย่างช้ามาก และสเปกตรัมของสัญญาณรบกวนจะไม่เปลี่ยนแปลงระหว่างคาบเวลาช่วง Update สำหรับการกู้คืนของสัญญาณในโดเมนของเวลา (time domain) และขนาดของสเปกตรัมที่ประมาณได้ช่วงสั้นๆ จะถูกนำรวมกับเฟสของสัญญาณที่ผิดเพี้ยน (noisy signal) และเมื่อทำการแปลงฟูร์มกลับด้วย inverse discrete fourier ในเทอมของค่าที่นำมาคำนวณจะเป็นกลุ่มคอมเพล็กซ์ โมเดลของสัญญาณที่ผิดเพี้ยนในโดเมนของเวลา สามารถแทนด้วยสมการ 3.26

$$y(m) = x(m) + n(m) \quad (3.26)$$

โดย $y(m)$ คือ สัญญาณที่ผิดเพี้ยน (Noisy signal)
 $x(m)$ คือ สัญญาณต้นแบบ (Original Signal)
 $n(m)$ คือ สัญญาณรบกวน

และในโดเมนของความถี่สัญญาณเสียงที่ผิดเพี้ยนสามารถแทนด้วยสมการ 3.27

$$Y(f) = X(f) + N(f) \quad (3.27)$$

โดย $Y(f)$ คือ ฟูเรียร์ทรานส์ฟอร์มของสัญญาณที่ผิดเพี้ยน
 $X(f)$ คือ ฟูเรียร์ทรานส์ฟอร์มของสัญญาณต้นแบบ
 $N(f)$ คือ ฟูเรียร์ทรานส์ฟอร์มของสัญญาณรบกวน

ในกระบวนการหักล้างทางสเปกตรัมสัญญาณ $x(m)$ คือสัญญาณที่เข้ามาจะถูกนำมาแบ่งเป็น N ตัวอย่าง โดยแต่ละกลุ่มจะถูกวินโดว์ด้วย วินโดว์ Hamming และถูกทรานส์ฟอร์มด้วย Discrete Fourier transform (DFT) ดังสมการ 3.28

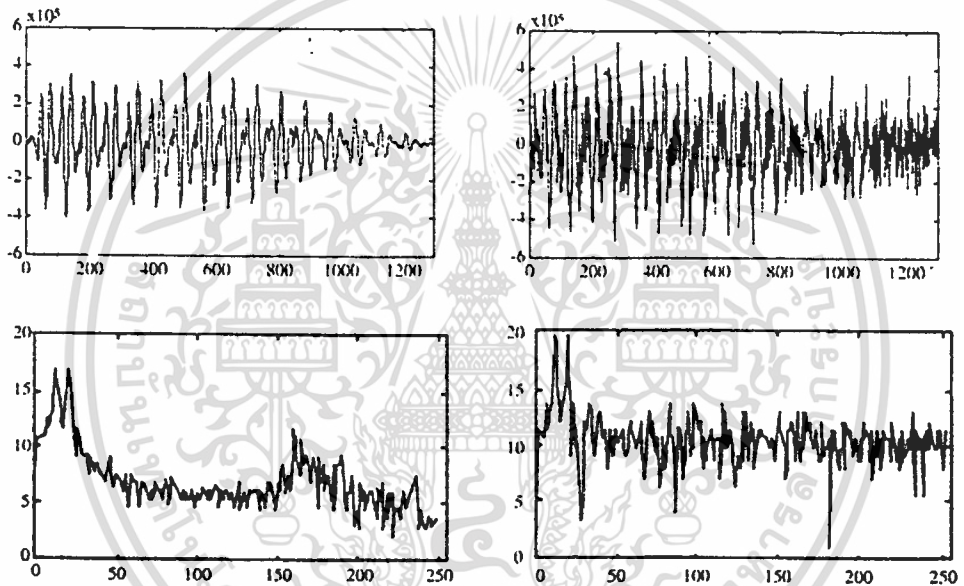
$$\begin{aligned} y_w(m) &= w(m)y(m) \\ &= w(m)[x(m) + n(m)] \\ &= x_w(m) + n_w(m) \end{aligned} \quad (3.28)$$

สามารถแทนค่าในโดเมนของความถี่ได้ดังสมการ 3.29

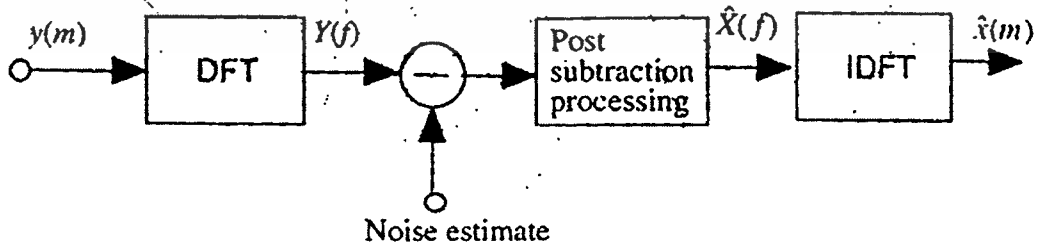
$$\begin{aligned}
 Y_w(f) &= W(f) * Y(f) \\
 &= X_w(f) + N_w(f)
 \end{aligned}
 \tag{3.29}$$

โดย * คือ การคอนโวลูชัน

W คือ สัญญาณลักษณะของสัญญาณที่ผ่านการวินโดว์แล้ว



รูปที่ 3.12. สัญญาณรบกวนใน โดเมนเวลาและความถี่



รูปที่ 3.13 บล็อกไดอะแกรมของกระบวนการหักล้างทางสเปกตรัม

ซึ่งรูปที่ 3.13 แสดงถึงโครงสร้างของวิธีการหักล้างทางสเปกตรัม โดยรายละเอียดที่กล่าวมา

แล้วสามารถอธิบายกระบวนการหักล้างทางสเปกตรัมได้อีก ดังสมการ 3.30

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ไว้เพื่อใช้ประโยชน์ในวงจำกัด ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$|\hat{X}(f)|^b = |Y(f)|^b - \alpha \overline{|N(f)|^b} \quad (3.30)$$

โดย $\frac{|\hat{X}(f)|^b}{\overline{|N(f)|^b}}$ คือสัญญาณเสียงต้นแบบของสัญญาณ $|X(f)|^b$
 คือ ค่าเฉลี่ยของคาบเวลาสเปกตรัมของสัญญาณรบกวน

ซึ่งเราสมมติให้สัญญาณรบกวนคือ สัญญาณแรนดอมในช่วงปฏิบัติการ โดยขนาดของการหักล้างทางสเปกตรัมกำหนดให้ $b=1$ และกำลังของการหักล้างทางสเปกตรัม กำหนดให้ $b=2$ ส่วนพารามิเตอร์ α จะเป็นตัวควบคุมการหักล้างสัญญาณรบกวนออกจากสัญญาณเสียงที่ผิดเพี้ยน ค่า α จะมีค่าเท่ากับ 1 ก็คือการหักล้างเป็นแบบ Full noise และค่า α จะมีค่ามากกว่า 1 ก็ต่อเมื่อการหักล้างเป็นแบบ over-subtraction และค่าเฉลี่ยสเปกตรัมของสัญญาณรบกวนหาได้จากคาบเวลาที่มีสัญญาณว่างและมีเพียงแต่สัญญาณรบกวน ดังสมการ 3.31

$$\overline{|N(f)|^b} = \frac{1}{K} \sum_{i=0}^{K-1} |N_i(f)|^b \quad (3.31)$$

จากสมการข้างบนค่าของ $\overline{|N(f)|^b}$ คือสเปกตรัมของสัญญาณรบกวนในลำดับเฟรมที่ i^{th} ซึ่งถูกสุ่มนิยฐานไว้ โดยมีจำนวนเฟรม K เฟรม ในคาบของสัญญาณรบกวน ค่าเฉลี่ยของสเปกตรัมของสัญญาณรบกวนหาได้จากลำดับที่หนึ่งของตัวกรอง Low-pass filter

$$\overline{|N_i(f)|^b} = \rho \overline{|N_{i-1}(f)|^b} + (1-\rho) |N_i(f)|^b \quad (3.32)$$

ซึ่งค่า ρ คือสัมประสิทธิ์ของ Low pass filter และถูกกำหนดค่าไว้ระหว่าง 0.85 และ 0.9 ส่วนการกู้กลับคืนของสัญญาณในโดเมนของเวลา ขนาดของสเปกตรัม $|\hat{X}(f)|$ จะถูกนำมาพร้อมกับเฟสของสัญญาณที่ผิดเพี้ยนและสามารถทรานส์ฟอร์มกลับเข้าไปในโดเมนของเวลา โดยผ่านการแปลงกลับของ Discrete Fourier transform ดังสมการ 3.33

$$\hat{x}(m) = \sum_{k=0}^{N-1} |\hat{X}(k)| e^{j\theta_r(k)} e^{-j\frac{2\pi}{N}km} \quad (3.33)$$

โดยที่ $\theta_r(k)$ คือเฟสของสัญญาณที่ผิดเพี้ยน $Y(k)$

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เนื่องจากสเปกตรัมของสัญญาณรบกวนมีความผันผวนทำให้ ผลของการหักล้างทางสเปกตรัมอาจให้ค่าของกำลังหรือขนาดของสเปกตรัมเป็นลบ ซึ่งผลที่ได้นี้ทำให้ค่า SNR ลดลง จึงมีการหลีกเลี่ยงที่จะมีให้ค่าของขนาดในวิธีการหักล้างทางสเปกตรัมเป็นค่าติดลบจะใช้กระบวนการ mapping Function $r[\cdot]$ ดังสมการ 3.34

$$T\left[|\hat{X}(f)|\right] = \begin{cases} |\hat{X}(f)| & \text{if } |\hat{X}(f)| > \beta|Y(f)| \\ fn[Y(f)] & \text{otherwise} \end{cases} \quad (3.34)$$

สำหรับตัวอย่างเราอาจจะเลือกกฎนี้ถ้าการประมาณค่าของ $|\hat{X}(f)| > 0.01|Y(f)|$ (ในขนาดของสเปกตรัม 0.01 เท่ากับ -40 dB) ดังนั้น $|\hat{X}(f)|$ ควรจะกำหนดให้เป็นฟังก์ชันของสัญญาณที่ผิดเพี้ยน (noisy signal) โดยค่าของ $fn[Y(f)]$ จะอยู่ในฟอร์มที่ง่ายที่สุดคือ $fn[Y(f)] =$ พื้นของสัญญาณรบกวน (noise floor) ซึ่งของพื้นของสัญญาณรบกวนนี้จะเป็นค่าคงที่ที่เป็นบวก อีกทางเลือกหนึ่งคือ $fn[Y(f)] = \beta|Y(f)|$ ในทางเลือกนี้จะได้สมการ 3.35

$$T\left[|\hat{X}(f)|\right] = \begin{cases} |\hat{X}(f)| & \text{if } |\hat{X}(f)| > \beta|Y(f)| \\ \beta|Y(f)| & \text{otherwise} \end{cases} \quad (3.35)$$

วิธีการหักล้างทางสเปกตรัมอาจจะถูกเพิ่มในโดเมนของกำลังหรือขนาดของสเปกตรัม ในสองวิธีที่เหมือนกันแม้ว่าตามหลักการแล้วผลลัพธ์ของมันอาจแตกต่างกัน

3.4.1 กำลังของการหักล้างทางสเปกตรัม

กำลังของกระบวนการหักล้างทางสเปกตรัม หรือขนาดของสเปกตรัม สามารถหาได้จากสมการ 3.36

$$|\hat{X}(f)|^2 = |Y(f)|^2 - \overline{|N(f)|^2} \quad (3.36)$$

จากสมการ 3.30 จะสมมติด้วยค่า α เราจะสังเกตได้ว่าการคำนวณหา กำลังของสเปกตรัมจะได้จากฟังก์ชันของ $E\left[|X(f)|^2\right]$ ส่วนกำลังของสเปกตรัมในช่วง Time-averaged หาได้จาก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$\overline{|X(f)|^2}$ และกำลังของสเปกตรัมในช่วงเวลาได้จาก $|X(f)|^2$ การขยายตัวของสเปกตรัมกำลังของสัญญาณเสียงที่ผิดเพี้ยนที่เกิดขึ้นในช่วงเวลาสั้นๆคือ $|Y(f)|^2$ และการจัดกลุ่มของเทอมต่างๆทำให้สมการ 3.36 อาจสามารถเขียนได้ดังสมการ 3.37

$$|\hat{X}(f)|^2 = |X(f)|^2 + \underbrace{\left(|N(f)|^2 - \overline{|N(f)|^2} \right)}_{\text{Noise variations}} + \underbrace{X^*(f)N(f) + X(f)N^*(f)}_{\text{Cross products}} \quad (3.37)$$

โดยที่ $\left(|N(f)|^2 - \overline{|N(f)|^2} \right)$ เป็นเทอมของการแปรผันของสัญญาณรบกวน (Noise variations)
 $X^*(f)N(f) + X(f)N^*(f)$ เป็นเทอมของการ cross products

$$E\left[|\hat{X}(f)|^2\right] = E\left[|X(f)|^2\right] \quad (3.38)$$

จากสมการ 3.38 จะเห็นได้ว่า ค่าเฉลี่ยของการประมาณค่าของกำลังของสัญญาณช่วงจะรวมตัวกับกำลังของสัญญาณ noise-free ซึ่งเป็นที่น่าสังเกตสำหรับสัญญาณ non-stationary เช่นเดียวกับสัญญาณเสียงพูด ในช่วงช่วงระยะเวลาหรือ short-time spectrum เพียงแค่นำมาเทียบเคียงกันระหว่างค่าเฉลี่ยก็สามารถนำมาประยุกต์ได้ โดยทั่วไปการเฉลี่ยจะมีการผิดเพี้ยนของสัญญาณและยากแก่การเข้าใจ ดังนั้นสามารถอธิบายได้ตามสมการ 3.38 ดังที่กล่าวมาแล้ว

3.4.2 ขนาดของการหักล้างทางสเปกตรัม

ขนาดของการหักล้างทางสเปกตรัม สามารถคำนวณได้จากสมการ (3.39)

$$|\hat{X}(f)| = |Y(f)| - \overline{|N(f)|} \quad (3.39)$$

โดยที่ $\overline{|N(f)|}$ คือ ขนาดของสเปกตรัมของสัญญาณรบกวนในช่วง Time-averaged

$$\begin{aligned} E\left[|\hat{X}(f)|\right] &= E\left[|Y(f)|\right] - E\left[\overline{|N(f)|}\right] \\ &= E\left[|X(f) + N(f)|\right] - E\left[\overline{|N(f)|}\right] \\ &= E\left[|X(f)|\right] \end{aligned} \quad (3.40)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การกู้กลับของสัญญาณทำได้โดยเอาขนาดที่ได้มารวมกับเฟสของสัญญาณที่ผิดเพี้ยนแล้วทำการ ทรานส์ฟอร์มเข้าไปใน โดเมนของเวลา

3.4.3 ตัวกรองการหักล้างทางสเปกตรัม

สมการของการหักล้างทางสเปกตรัม สามารถแสดงในรูปของผลคูณระหว่างสเปกตรัมของสัญญาณที่ผิดเพี้ยนกับความถี่ตอบสนองของตัวกรองการหักล้างทางสเปกตรัม ดังสมการ 3.41

$$\begin{aligned} |\hat{X}(f)|^2 &= |Y(f)|^2 - |N(f)|^2 \\ &= H(f)|Y(f)|^2 \end{aligned} \quad (3.41)$$

โดยที่ $H(f)$ คือ ความถี่ตอบสนองของตัวกรองหักล้างของสเปกตรัมซึ่งเกิดจากสมการ 3.42

$$\begin{aligned} H(f) &= 1 - \frac{|N(f)|^2}{|Y(f)|^2} \\ &= \frac{|Y(f)|^2 - |N(f)|^2}{|Y(f)|^2} \end{aligned} \quad (3.42)$$

ในการหักล้างทางสเปกตรัมตัวกรอง $H(f)$ คือ ตัวกรอง Zero phase และมีขนาดระหว่างช่วง $0 \geq H(f) \geq 1$

3.4.4 ความคิดพึ้นในการประมวลผล

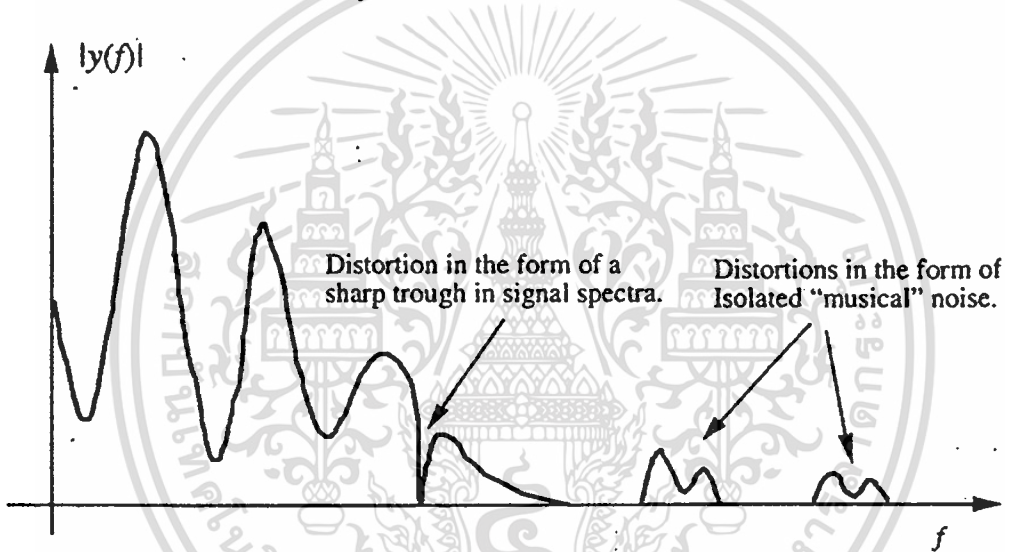
ปัญหาหลักที่เกิดในวิธีการหักล้างทางสเปกตรัม คือ การประมวลผลของสัญญาณรบกวนที่ไม่เป็นเชิงเส้น (non-linear) ซึ่งสาเหตุเกิดจากการเรนดอมของสเปกตรัมของสัญญาณรบกวนนั่นเอง จากสมการ 3.37 เราจะได้ว่าค่าของขนาดของสเปกตรัมจะต้องไม่เป็นค่าลบ ซึ่งเราสามารถหาค่าได้จาก การประมวลค่าสัญญาณรบกวน หรือกำลังของสัญญาณรบกวน คือ (ก) การแปรผันของสเปกตรัมกำลังของสัญญาณของรบกวนสามารถประมาณจากค่าเฉลี่ย, (ข) สัญญาณและสัญญาณรบกวนอยู่ในเทอมของ Cross Product , (ค) การ mapping อย่างไม่เป็นเชิงเส้นของการประมาณค่าสเปกตรัม จะคืออยู่ต่ำกว่าค่า threshold

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เช่นเดียวกับสัญญาณรบกวนในที่ทอมของสเปกตรัมขนาดและกำลังในวิธีการหักล้างทางสเปกตรัม จากข้อมูลของสัญญาณรบกวน 3 กลุ่ม ข้างบน ส่วนประกอบสำคัญของสัญญาณรบกวนที่เกิดขึ้นบ่อยๆ เนื่องจากค่าของการ mapping แบบไม่เป็นเชิงเส้นเป็นลบหรือมีค่าน้อยมากๆ ในการประมาณค่าทางสเปกตรัม สัญญาณรบกวนนี้ทำให้เกิด metallic sounding noise หรือที่เรียกกันว่า "musical tone noise" สิ่งเหล่านี้เกิดขึ้นเนื่องจากสเปกตรัมมีแถบแคบ และเสียง tin-like ผลสำเร็จของวิธีการหักล้างทางสเปกตรัม ขึ้นอยู่กับความสามารถในการกำจัดสัญญาณรบกวนของอัลกอริทึม และการจัดการประมวลผล สัญญาณรบกวน ในกรณีที่ยากและไม่ปกติ สัญญาณรบกวนที่ตกค้างจะมี 2 ลักษณะ คือ

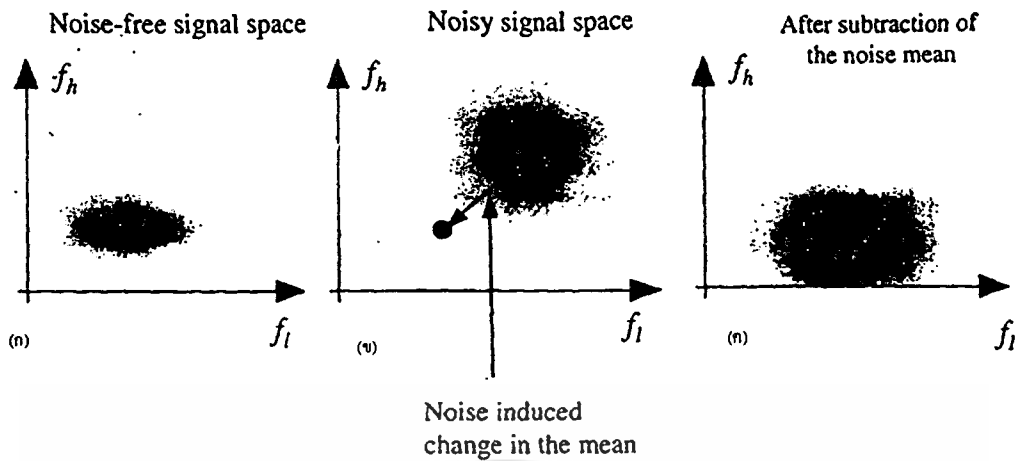
(ก) รูปร่างเป็นร่องหรือมีระดับสูงสุดในสัญญาณ spectra

(ข) แถบแคบของความถี่อยู่แยกจากกัน



รูปที่ 3.14 ความผิดเพี้ยนที่อาจเกิดขึ้นได้จากกระบวนการหักล้างทางสเปกตรัม

ในบริเวณรอบๆ ของสัญญาณความถี่ที่มีขนาดสูง จะมีสัญญาณรบกวนที่ทำให้เกิดร่องหรือจุดสูงสุดปกคลุมทั่วบริเวณ ทำให้ไม่สามารถได้ยินเสียงเนื่องจากสัญญาณมีพลังสูง สาเหตุหลักที่ทำให้การได้ยินลดลงหรือแย่งคือส่วนประกอบความถี่แยกจากกัน ที่เห็นได้คือ musical tone หรือ musical noise ที่อธิบายในรูปที่ 3.14 ลักษณะของ musical noise จะมีแถบความถี่แคบและสั้น หรือจะกล่าวได้ว่าส่วนประกอบความถี่อยู่ในระดับต่ำ ในการกู้กลับคืนของสัญญาณเสียงโดยวิธีหักล้างทางสเปกตรัม (spectral subtraction) ผลลัพธ์ที่ได้เป็นเครื่องหมายที่ใช้ในการตัดสินคุณภาพของสัญญาณ ในลักษณะนี้ผลลัพธ์จะมีค่า SNR ต่ำ ซึ่งผลกระทบในการใช้อัลกอริทึมของการหักล้างทางสเปกตรัมที่มีคุณภาพต่ำ ทำให้ผลลัพธ์ในการได้ยินเสียงและข้อมูลต่ำกว่าสัญญาณเสียงดั้งเดิม



รูปที่ 3.15 ความผิดเพี้ยนที่อาจจะเกิดขึ้นได้จากกระบวนการหักล้างทางสเปกตรัม ในรูปขนาดสเปกตรัมของสัญญาณ

3.4.5 ผลกระทบของการหักล้างของสเปกตรัมในการกระจายตัวของสัญญาณ

จากรูปที่ 3.15 เป็นการอธิบายถึง ผลกระทบสัญญาณรบกวนของการหักล้างทางสเปกตรัม ในการกระจายตัวของขนาดของสเปกตรัมของสัญญาณถูกแบ่งเป็นสองส่วน คือแถบความถี่ต่ำ f_l และแถบความถี่สูง f_h แต่ละจุดในรูปที่ 3.15 จะพรีออดอยู่ระหว่างสเปกตรัมความถี่สูง (high-frequency spectrum) กับสเปกตรัมความถี่ต่ำ (low-frequency spectrum) ในลักษณะ 2 มิติ โดยรูปรูปที่ 3.15 (ก) แสดงการกระจายตัวของตัวอย่างสเปกตรัมของสัญญาณในขนาดความถี่ 2 มิติ ซึ่งผลกระทบของการเรนดอมสัญญาณรบกวน แสดงในรูปที่ 3.15 (ข) จะมีการเพิ่มขึ้นในช่วง mean และช่วงแปรผัน (variance) ของสเปกตรัมโดยค่ารวมที่ได้จะขึ้นอยู่กับค่า mean และค่าแปรผันจะเป็นสัญญาณรบกวนที่ไม่สามารถขจัดได้ ส่วนการเพิ่มขึ้นในช่วง mean ของขนาดสเปกตรัม จะสามารถขจัดได้ด้วยวิธีการหักล้างทางสเปกตรัม รูปที่ 3.15 (ค) แสดงผลของสัญญาณรบกวนของวิธีการหักล้างทางสเปกตรัม ในการกระจายตัวของสเปกตรัมสัญญาณ จากรูปจะเห็นว่าสัญญาณรบกวนที่เพิ่มขึ้นในช่วงการแปรผันของสเปกตรัมสัญญาณ หลังจากผ่านการหักล้างค่า SNR จะมีค่าน้อยมากจนกลายเป็นลบ จากภาพแสดงให้เห็นถึงการกระจายตัวของค่า SNR ที่ต่ำของสัญญาณสเปกตรัม

3.4.6 การลดสัญญาณรบกวนแบบแปรผัน

สัญญาณรบกวนที่เป็นผลลัพธ์จากการหักล้างทางสเปกตรัม เนื่องจากค่าสัญญาณรบกวนแบบแปรผัน เราจะพิจารณาวิธีการลดการแปรผันของค่ากำลังของสเปกตรัม สัญญาณรบกวนแบบ

white noise ซึ่งมีค่าแปรผันเป็น เราสามารถแสดงค่าแปรผันของสเปกตรัมที่ผ่านกระบวนการ DFT สเปกตรัม โดยมี $N(f)$ เป็นสัญญาณ ได้ดังสมการ 3.43

$$VAR\left[|N(f)|^2\right] \approx P_{NN}^2(f) = \sigma_n^4 \quad (3.43)$$

ถ้าค่า K เป็นตัวแปรอิสระจะได้สมการ

$$VAR\left[\frac{1}{K} \sum_{i=0}^{K-1} |N_i(f)|^2\right] \approx \frac{1}{K} P_{NN}^2(f) \approx \frac{1}{K} \sigma_n^4 \quad (3.44)$$

ในวิธีการหักล้างทางสเปกตรัม (spectral subtraction) สัญญาณเสียงที่ผิดเพี้ยน (noisy signal) หรือ ในสมการคือ $y(m)$ จะถูกแบ่งเป็นกลุ่มโดยกลุ่มละ N ตัวอย่าง สัญญาณแต่ละกลุ่ม จะถูกทรานส์ฟอร์ม ด้วย DFT จะได้สัญญาณเป็น $Y(f)$ หลังจากนั้นสเปกตรัมที่ได้จะอยู่ในรูปของ เมตริกซ์ของความถี่และเวลาสองมิติ แทนด้วย $Y(f, t)$ ซึ่งตัวแปร t คือตัวชี้ของกลุ่มที่ถูกแบ่ง และ ใช้แทนมิติของเวลาด้วย สัญญาณ $Y(f, t)$ สามารถพิจารณาได้เช่นเดียวกับ band-pass ของช่อง f ซึ่งประกอบด้วยเวลาที่เปลี่ยนไปเรื่อยๆ (time-varying) ของสัญญาณ $X(f, t)$ บวกกับสัญญาณ รบกวน $N(f, t)$ ที่เกิดจากการแรนดอม ส่วนวิธีการหนึ่งที่ใช้สำหรับลดสัญญาณรบกวนแปรผัน ก็คือ การนำขนาดของแต่ละความถี่มาผ่าน low-pass filter ตัวอย่างที่ยกขึ้นมาสามารถเขียนได้ดัง สมการ 3.45

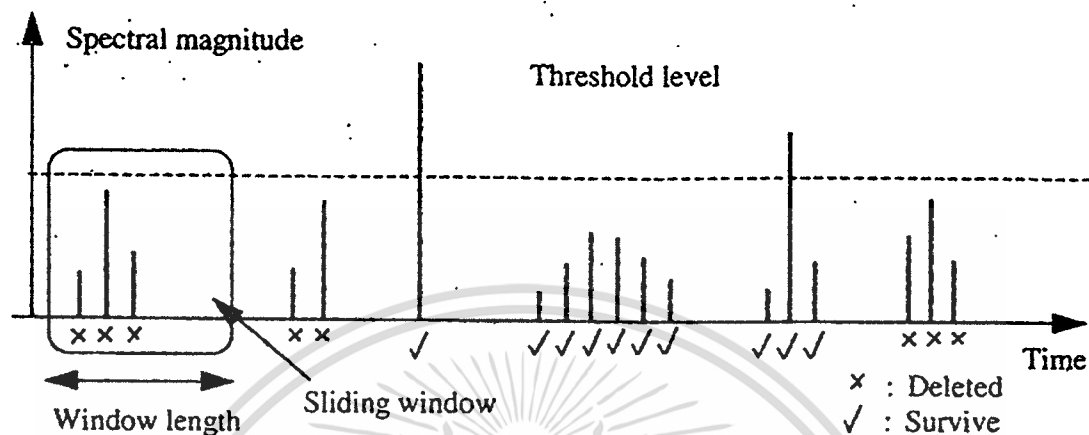
$$|Y_{LP}(f, t)| = \rho |Y_{LP}(f, t-1)| + (1-\rho) |Y(f, t)| \quad (3.45)$$

โดย LP แทนค่าผลลัพธ์ของสัญญาณที่ผ่าน low-pass filter และค่าสัมประสิทธิ์ ρ เป็นตัวควบคุม แถบกว้างโดยมีเวลาคงที่

3.4.7 การกรองสัญญาณที่ผิดเพี้ยน

สัญญาณเสียงพูดและเสียงดนตรี จะถูกปรับแต่งในช่วง non-stationary และจะมี lifetime เปลี่ยนไปเรื่อย ๆ จะต่างกันไปขึ้นอยู่กับความเข้มของสเปกตรัมที่ถูกปรับแต่ง โดยปกติ time varying ของสัญญาณเสียงเมื่อเปิดจะถ่ายทอดข้อมูล ความรู้สึก และคุณภาพของเสียงออกมา แต่สิ่ง ที่เราไม่ต้องการก็คือ musical tone noise ที่เป็นผลพวงของ spectral subtraction ในช่วง time-varying ด้วยนั่นเอง อย่างไรก็ตามจะมีเครื่องหมายแสดงถึงความแตกต่างระหว่าง ลักษณะของสัญญาณเสียง (audio signal) กับ musical noise ซึ่งมีลักษณะที่แตกต่างนี้อาจจะแสดงให้เห็นชัดเจน หรืออาจจะถูก

ขจัดออกไปบ้างในบางส่วน การแสดงตัวของ musical noise อาจทำงานในช่วงที่เปลี่ยนไปในเทอมของความถี่และเวลา ลักษณะหลักของ musical noise มักจะมีความสัมพันธ์กับสัญญาณ narrow band ขาวสั้นๆ ที่แยกกลุ่มออกมา โดยมีการแรนดอม และมีแอมพลิจูดขนาดเล็ก



รูปที่ 3.16 วิธีการบ่งชี้ตัวกรองสำหรับ musical noise

การใช้ DFT โดยขนาดของบล็อกคือ 128 ตัวอย่าง มีอัตราการแซมปลิง (sampling) ที่ 20 kHz ในการทดลองจะเห็นว่ามี musical noise ไม่น้อยกว่า 3 เฟรม ตัวอย่างในรูปที่ 3.16 แสดงถึงวิธีการที่จะบ่งชี้ว่าตัวใดเป็น musical noise โดยในแต่ละช่อง DFT จะมีการตรวจสอบช่องความถี่เป็น short-lived ถึงองค์ประกอบความถี่มีระยะเวลาสั้นกว่าช่วงเวลาที่เลือกครั้งก่อน และมีแอมพลิจูดที่เล็กกว่าค่า threshold และไม่บังสัญญาณที่อยู่ติดกัน จะถือว่าเป็นสัญญาณช่วงที่มีคุณสมบัติดังกล่าวเป็นสัญญาณรบกวน และถูกค้นพบ

3.4.8 การหักล้างทางสเปกตรัมแบบไม่เป็นเชิงเส้น

การใช้วิธีการหักล้างทางสเปกตรัมในรูปของฟอร์มพื้นฐาน ดังสมการ 3.30 อาจจะทำให้คุณภาพและข้อมูลของสัญญาณไม่ดีเท่าที่ควร เช่น ในรูปของการกู่กลับคือของสัญญาณ audio , การที่สัญญาณ musical noise ที่เป็นสาเหตุให้คุณภาพในภาครับลดต่ำลงและการรู้จักของเสียงพูด เมื่อใช้การหักล้างทาง สเปกตรัมแบบเหลือจะทำให้มีผลต่อความแม่นยำในการรู้จำ ในการวิจัยจะมีหลายวิธีในการคำนวณหาค่าเพื่อให้ได้มาซึ่งการประมาณค่าสเปกตรัมของสัญญาณรบกวน (noise spectrum) การหาค่าเฉลี่ยสัญญาณที่ผิดเพี้ยน (noise signal spectrum) และการกำจัดสัญญาณรบกวน สำหรับวิธีการหักล้างของสเปกตรัมแบบไม่เป็นเชิงเส้น ก็เป็นวิธีที่เป็นประโยชน์กับการประมาณค่า SNR เพื่อทำให้สัญญาณมีคุณภาพดีขึ้น โดยสมการจะอยู่ในรูปแบบ over-subtraction ดังสมการ 3.46

$$\begin{aligned}
 |\hat{X}(f)| &= |Y(f)| - \overline{|N(f)|} \\
 &= |X(f)| + |N(f)| - \overline{|N(f)|} \\
 &= |X(f)| + V_N(f)
 \end{aligned} \tag{3.46}$$

โดยค่า $V_N(f)$ คือค่า Zero-mean ที่เกิดจากการแรนดอมของสเปกตรัมของสัญญาณรบกวน (noise spectrum)

ด้วยหลักการพิจารณาจะพบว่าถ้าค่า $V_N(f)$ มีค่าติดลบ สัญญาณ $X(f)$ แล้วนั้นหมายถึง อาจจะพิจารณาได้ว่าการสูญเสียของสัญญาณรบกวน ในลักษณะนี้คือ over-subtraction ด้วยวิธีการประมาณค่าในเชิงลบแบบไม่เป็นเชิงเส้น ผลลัพธ์ที่ได้คือ มีสัญญาณรบกวนน้อยลง และการหักล้างทางสเปกตรัมแบบไม่เป็นเชิงเส้น อาจอธิบายได้ด้วยสมการ 3.47

$$|\hat{X}(f)| = |Y(f)| - \alpha(SNR(f)) \overline{|N(f)|}_{NL} \tag{3.47}$$

โดยที่ $\alpha(SNR(f))$ คือ ค่า SNR ที่ขึ้นอยู่กับแฟคเตอร์ของวิธีการหักล้าง
 $\overline{|N(f)|}_{NL}$ คือ ค่าประมาณของสเปกตรัมของสัญญาณรบกวนแบบไม่เป็นเชิงเส้น

กระบวนการของการประมาณค่าสเปกตรัมจะหลีกเลี่ยงเพื่อไม่ให้เกิดการประมาณค่าเป็นลบ ได้ด้วยสมการ 3.48

$$|\hat{X}(f)| = \begin{cases} |\hat{X}(f)| & \text{if } |\hat{X}(f)| > |\beta Y(f)| \\ |\beta Y(f)| & \text{otherwise} \end{cases} \tag{3.48}$$

ค่า SNR ที่ขึ้นอยู่กับแฟคเตอร์ของวิธีการหักล้าง จากสมการ 3.47 หาได้จากสมการ 3.49

$$\alpha(SNR(f)) = 1 + \frac{sd(|N(f)|)}{\overline{|N(f)|}} \tag{3.49}$$

จากสมการฟังก์ชัน $sd(|N(f)|)$ คือค่าเบี่ยงเบนมาตรฐานของสัญญาณรบกวนที่ความถี่ f สำหรับ white noise ค่า $sd(|N(f)|) = \sigma_n$ โดยที่ค่า σ_n^2 คือค่าแปรผันของสัญญาณรบกวน ส่วนการหักล้างของสมการ 3.49 ในรูปแบบตามสมการ 3.47 จะได้ดังสมการ 3.50

$$|\hat{X}(f)| = |Y(f)| - \left[1 + \frac{sd(|N(f)|)}{|N(f)|} \right] |N(f)| \quad (3.50)$$

ในสมการ 3.50 จะพบว่าแฟคเตอร์ของการหักล้างจะขึ้นอยู่กับค่า mean และค่าความแปรผันของสัญญาณรบกวน สังเกตได้ว่าการเพิ่มขึ้นของ over-subtracted ก็คือค่าเบี่ยงเบนมาตรฐานของสัญญาณรบกวน วิธีการที่ค้นคว้าได้นี้ น่าสนใจเป็นอย่างมากเนื่องจาก ค่าสูงสุดหนึ่งของค่าแปรผันของสัญญาณรบกวนที่กำหนดได้เป็นศูนย์ จะทำให้เกิดสัญญาณ Sine wave นั่นคือ $\alpha(SNR(f)) = 1$ และสำหรับค่าสูงสุดของสัญญาณรบกวนแบบ white noise $\alpha(SNR(f)) = 2$ ซึ่งสำหรับการประยุกต์ใช้กับการหักล้างทางสเปกตรัมในการรู้จำเสียงพูด จะพบว่าค่าแฟคเตอร์ที่ดีที่สุดที่มักจะใช้คือ 1 และ 2

ในวิธีการหักล้างทางสเปกตรัมแบบไม่เป็นเชิงเส้นของ Lockwood และ Boudy จะใช้ตัวกรองการหักล้างทางสเปกตรัมที่หาได้จาก

$$H(f) = \frac{\overline{|Y(f)|^2} - \overline{|N(f)|^2}_{NL}}{\overline{|Y(f)|^2}} \quad (3.51)$$

และฟังก์ชันการประมาณค่าของสเปกตรัมสัญญาณรบกวนนั้น Lockwood และ Boudy ได้นำเสนอดังสมการ 3.52

$$\overline{|N(f)|^2}_{NL} = \Phi \left(\max_{\text{over } M \text{ frames}} (|N(f)|^2), SNR(f), \overline{|N(f)|^2} \right) \quad (3.52)$$

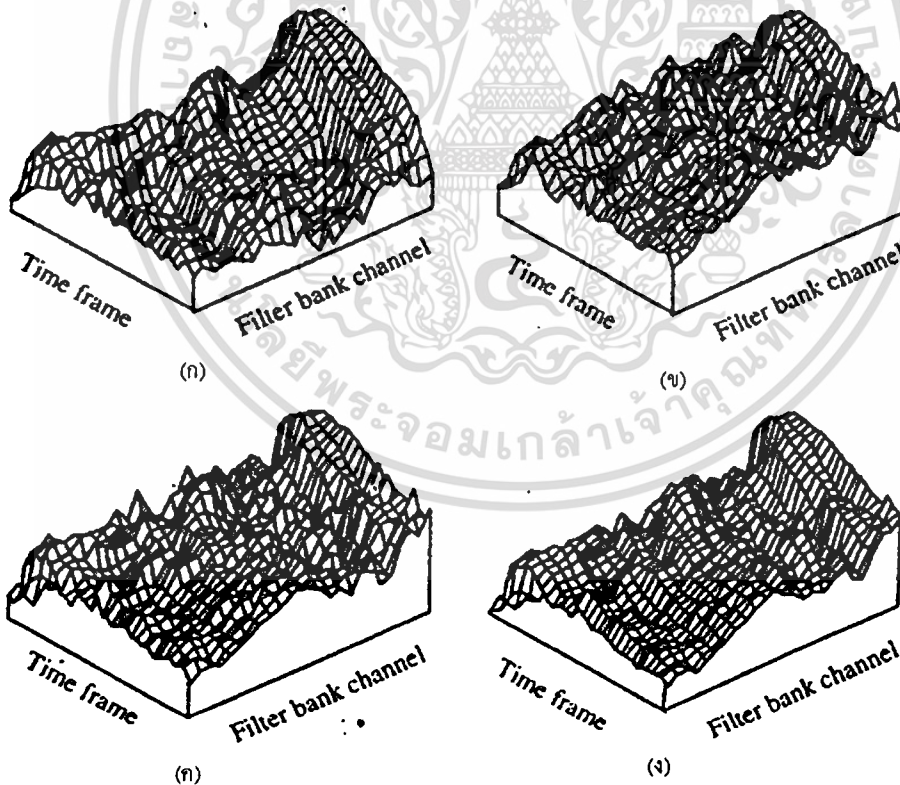
ค่าประมาณของสเปกตรัมของสัญญาณรบกวน คือ ฟังก์ชันของค่าสูงสุดของสเปกตรัมของสัญญาณรบกวนที่เกินกว่า M เฟรม และ SNR (Signal -to-noise ratio) ซึ่งสมการของ $\Phi(\cdot)$ แบบไม่เป็นเชิงเส้นเป็นไปตามสมการ 3.53

$$\Phi\left(\max_{\text{over } M \text{ frames}} (|N(f)|^2), SNR(f)\right) = \frac{\max_{\text{over } M \text{ frames}} (|N(f)|^2)}{1 + \gamma SNR(f)} \quad (3.53)$$

โดยค่า γ คือ ตัวแปรที่ออกแบบจากสมการ 3.53 พบว่าเมื่อค่า SNR ลดลงทำให้ผลลัพธ์ของตัวประมาณค่า $\Phi(\cdot)$ แบบไม่เป็นเชิงเส้นมีค่าเข้าใกล้ค่า $\max(|N(f)|^2)$ และเมื่อค่า SNR เพิ่มขึ้นผลลัพธ์จะเข้าใกล้ศูนย์สำหรับ over-subtraction ค่าสัญญาณรบกวนโดยประมาณจะถูกบังคับให้เป็น over-estimation โดยใช้ฟังก์ชันที่มีเขตจำกัดดังนี้

$$\overline{|N(f)|^2} \leq \Phi\left(\max_{\text{over } M \text{ frames}} (|N(f)|^2), SNR(f), \overline{|N(f)|^2}\right) \leq 3\overline{|N(f)|^2} \quad (3.54)$$

ตัวกรองของวิธีการหักล้างทางสเปกตรัมที่มากที่สุดถูกจำกัดอยู่ที่ $H(f) \geq \beta$ ซึ่งโดยปกติจะจำกัดค่าจะใช้ $\beta \geq 0.01$ โดยรูปที่ 3.17 แสดงให้เห็นถึงผลกระทบของการหักล้างทางสเปกตรัมแบบไม่เป็นเชิงเส้น และการกู้กลับคืนของสเปกตรัมของสัญญาณเสียงพูด



รูปที่ 3.17 ผลกระทบของการหักล้างทางสเปกตรัมแบบไม่เป็นเชิงเส้น (ก) สัญญาณเสียงพูดต้นแบบ (ข) สัญญาณเสียงพูดที่ผิดเพี้ยน 12 dB (ค) การหักล้างทางสเปกตรัมแบบไม่เป็นเชิงเส้น (ง) การหักล้างทางสเปกตรัมแบบไม่เป็นเชิงเส้นด้วย smoothing

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4

วิธีที่นำเสนอในการปรับแต่งคุณภาพเสียงพูด

4.1 การอธิบายทางคณิตศาสตร์

การกำจัดสัญญาณรบกวนออกจากสัญญาณเสียงที่ผิดเพี้ยน ด้วยวิธีการหักล้างทางสเปกตรัม มีรายละเอียดและวิธีการต่างๆ ที่ควรทราบถึงดังต่อไปนี้

4.1.1 โมเดลของการปรับสัญญาณรบกวน

โมเดลการปรับสัญญาณรบกวนนี้ จำลองขึ้นเพื่อให้เราทราบถึงค่าสัญญาณรบกวนเพื่อที่จะทำให้เราได้ทราบว่าอัลกอริทึมการกำจัดสัญญาณรบกวนของเรามีคุณภาพมากน้อยเพียงใด ส่วนของโมเดลการปรับสัญญาณรบกวนอยู่ในช่วงของการเตรียมข้อมูล ก่อนทำการกำจัดสัญญาณรบกวน ซึ่งภาพลักษณ์โดยรวมเริ่มจาก เราสมมุติให้สัญญาณหนึ่งเกิดจากผลรวมของสัญญาณสองสัญญาณที่ปรับเข้าด้วยกัน ตามสมการ 4.1

$$x(k) = s(k) + n(k) \quad (4.1)$$

โดยที่ $s(k)$ คือสัญญาณเสียงต้นแบบ
 $n(k)$ คือสัญญาณรบกวน
 $x(k)$ คือสัญญาณเสียงพูดที่ผิดเพี้ยน

สมการ 4.1 เมื่อผ่านการแปลงฟูเรียร์ทรานส์ฟอร์มจะได้สมการ 4.2

$$X(e^{j\omega}) = S(e^{j\omega}) + N(e^{j\omega}) \quad (4.2)$$

โดย

$$\begin{aligned} x(k) &\leftrightarrow X(e^{j\omega}) \\ X(e^{j\omega}) &= \sum_{k=0}^{L-1} x(k)e^{-j\omega k} \\ x(k) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{j\omega})e^{j\omega k} d\omega \end{aligned} \quad (4.3)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้.

4.1.2 การวินโดว์ข้อมูล

สัญญาณเสียงที่ผ่านการแปลงอนาลอกเป็นดิจิทัล จะถูกแบ่งเฟรมๆ ละ 256 จุด โดยสัญญาณจะมีการเลื่อน (over-lap) เท่ากับ 128 จุด ผลลัพธ์ที่ได้จะเก็บไว้ในบัฟเฟอร์ แล้วแต่ละบัฟเฟอร์จะถูกนำไปคูณกับ Hamming window ซึ่งการ Hamming window นี้ เป็นส่วนขยายของ Hanning window ซึ่งมีรูปแบบตัวสมการ 4.4

$$h(n) = \alpha + (1.0 - \alpha) \cos \left[\left(\frac{2\pi}{N} \right) n \right] \quad (4.4)$$

สมการ 4.4 จะสอดคล้องกับสเปกตรัมของสมการ 4.5

$$H(\theta) = \alpha D(\omega) + \frac{(1.0 - \alpha)}{2} \left[D\left(\omega - \frac{2\pi}{N}\right) + D\left(\omega + \frac{2\pi}{N}\right) \right] \quad (4.5)$$

โดยค่า α คือค่าประมาณ 0.54 ดังนั้นเราจะสามารถแทนค่าได้ ดังสมการ 4.6

$$H(\theta) = 0.54 + 0.46 \cos \left[\left(\frac{2\pi}{N} \right) n \right] \quad (4.6)$$

4.1.3 การประมาณค่าสัญญาณรบกวน

การกำจัดสัญญาณรบกวน โดยวิธีการหักล้างทางสเปกตรัม เราทำได้โดยการหักล้างสเปกตรัมของสัญญาณรบกวนออกจาก สเปกตรัมของสัญญาณเสียงที่ผิดเพี้ยน โดยเราสามารถประมาณค่าสัญญาณรบกวนได้ระหว่างช่วงที่ไม่มีเสียงพูด (non-speech) ดังสมการ 4.7

$$\mu(e^{j\omega}) = \frac{1}{N} \sum_{i=0}^{N-1} |X_i(e^{j\omega})| \quad (4.7)$$

โดยค่า	$\mu(e^{j\omega})$	คือ ค่าเฉลี่ยของขนาดของสัญญาณรบกวน
	$X_i(e^{j\omega})$	คือ สัญญาณเสียงพูดที่ผิดเพี้ยน
	N	คือ จำนวนข้อมูลในเฟรมแรก

ในสถานะที่หยุดอยู่กับที่ระหว่างช่วงที่ไม่มีเสียงพูดนี้ เราสามารถประมาณได้เท่ากับ 300 มิลลิวินาที และเราเลือกที่จะใช้ข้อมูลในเฟรมแรก คือ 256 จุด

4.1.4 การประมาณค่าการหักล้างทางสเปกตรัม

การหักล้างทางสเปกตรัม สามารถหาค่าตัวกรองการหักล้าง $H(e^{j\omega})$ ได้โดยการแทนค่าสเปกตรัมของสัญญาณรบกวนด้วย $N(e^{j\omega})$ ซึ่งสามารถหาได้จากการประมาณ [4] และขนาดของ $N(e^{j\omega})$ ก็คือ $|N(e^{j\omega})|$ จะถูกแทนค่าด้วยค่าประมาณของตัวเอง นั่นคือ ค่า $\mu(e^{j\omega})$ ซึ่งสามารถหาได้จากช่วงไม่มีเสียงพูด (non-speech activity) ส่วนเฟสของสัญญาณรบกวน $N(e^{j\omega})$ ก็คือ $\theta_N(e^{j\omega})$ จะถูกแทนค่าด้วยเฟสของ $X(e^{j\omega})$ ซึ่งก็คือ ค่า $\theta_X(e^{j\omega})$ นั่นเองในที่สุดเราจะได้ผลลัพธ์ของการหักล้างออกมา คือ $\hat{S}(e^{j\omega})$ สามารถแสดงได้โดยสมการ 4.8

$$\hat{S}(e^{j\omega}) = [|X(e^{j\omega})| - \mu(e^{j\omega})] e^{j\theta_X(e^{j\omega})} \quad (4.8)$$

$$\hat{S}(e^{j\omega}) = H(e^{j\omega})X(e^{j\omega})$$

โดย

$$H(e^{j\omega}) = 1 - \frac{\mu(e^{j\omega})}{|X(e^{j\omega})|} \quad (4.9)$$

$$\mu(e^{j\omega}) = E\{|N(e^{j\omega})|\}$$

4.1.5 การทำ Half-wave rectification

สำหรับแต่ละความถี่ ω ขนาดของสเปกตรัมของสัญญาณที่ผิดเพี้ยน $X(e^{j\omega})$ ที่มีค่าน้อยกว่าค่าเฉลี่ยของขนาดของสเปกตรัมของสัญญาณรบกวน $\mu(e^{j\omega})$ จะทำให้ผลลัพธ์ของสัญญาณถูกทำให้เป็นศูนย์ การปรับสัญญาณให้เป็นไปตามที่กำหนดทำได้โดยการหา half-wave rectification ซึ่งแทนด้วย $H(e^{j\omega})$ การประมาณค่าตัวกรองจะหาได้จากสมการ 4.10

$$\hat{S}(e^{j\omega}) = H_R(e^{j\omega})X(e^{j\omega}) \quad (4.10)$$

โดย

$$H_R(e^{j\omega}) = \frac{H(e^{j\omega}) + |H(e^{j\omega})|}{2} \quad (4.11)$$

ในกระบวนการดังกล่าวค่าประมาณ \hat{S} จะหาได้จากการหักล้าง ขนาดของสัญญาณรบกวน μ ออกจาก ขนาดของสัญญาณเสียงที่ผิดเพี้ยน $|X|$ ดังสมการ 4.12

$$|\hat{S}(k)| = |X(k)| - \mu \quad k = 0, 1, \dots, L-1 \quad (4.12)$$

หรืออาจเขียนแทนได้ดังสมการ 4.13

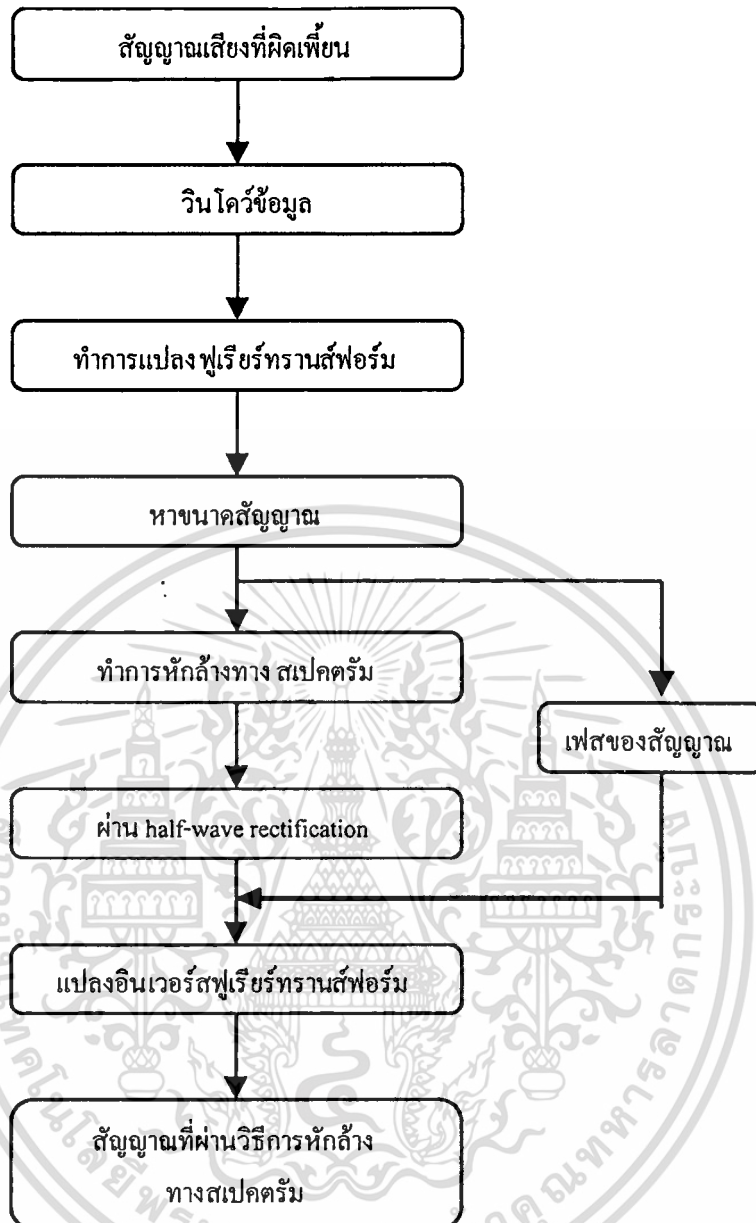
$$\begin{aligned} \hat{S}(k) &= H(k) \cdot X(k) \\ H(k) &= 1 - \frac{\mu(k)}{|X(k)|} \quad k = 0, 1, \dots, L-1 \end{aligned} \quad (4.13)$$

โดยค่า L คือ ความยาวของบัพเฟออร์ที่ผ่านการทำ DFT แล้ว ซึ่งหลังจากทำการหักล้างแล้ว ค่า \hat{S} ที่มีค่าเป็นลบ จะถูกเปลี่ยนให้เป็นศูนย์ทั้งหมด

4.2 ขั้นตอนการทำงานด้วยโปรแกรมคอมพิวเตอร์

เทคนิคการปรับแต่งคุณภาพเสียงด้วยวิธีการหักล้างทางสเปกตรัม มีกระบวนการทำงานที่ประกอบด้วยขั้นตอนและตัวกระทำทางคณิตศาสตร์มากมาย ซึ่งสามารถแทนกระบวนการเหล่านี้ได้โดย การใช้โปรแกรมภาษาคอมพิวเตอร์จำลองการทำงาน ซึ่งภาษาคอมพิวเตอร์สามารถเขียนโปรแกรมให้ทำงานเป็นขั้นตอนตามเทคนิควิธีการหักล้างทางสเปกตรัมได้สะดวก เพราะมีชุดคำสั่งและตัวกระทำทางคณิตศาสตร์มากมาย และยังสามารถแก้ไขโปรแกรมได้ง่าย มีความสะดวกต่อการทดลอง

แบบจำลองการปรับแต่งคุณภาพเสียง จะมีการทำงานโดยเริ่มจากการนำไฟล์เสียงป้อนเข้าสู่คอมพิวเตอร์ แล้วจะถูกแปลงสัญญาณเสียงให้เป็นสัญญาณดิจิทัล จากนั้นจำลองสัญญาณรบกวนแบบ White Gaussian ผสมเข้าไปในสัญญาณเสียงพูด เราจะได้สัญญาณเสียงพูดที่ผิดเพี้ยนที่เกิดจากสัญญาณรบกวน นำมาป้อนลงในโปรแกรมการปรับแต่งคุณภาพเสียง โปรแกรมจะทำการกำจัดสัญญาณรบกวนออกจากสัญญาณเสียงพูด โดยใช้เทคนิควิธีการหักล้างทางสเปกตรัม จากนั้นจึงนำผลลัพธ์ที่ได้มาทำการวัดคุณภาพเสียง และแสดงค่าวัดคุณภาพเสียงในรูปของกราฟ และสเปกโตรแกรม ของสัญญาณเสียงผลลัพธ์ ผลลัพธ์ดังกล่าวจะแสดงออกมาทางจอภาพ และบอกค่า SNR ของสัญญาณ



รูปที่ 4.1 กระบวนการกำจัดสัญญาณรบกวนโดยวิธีการหักล่างทางสเตครัม

4.2.1 ขั้นตอนการทำงานของ การปรับแต่งคุณภาพเสียงด้วยวิธีการหักล่างทางสเตครัม

สำหรับขั้นตอนการทำงานของกระบวนการปรับแต่งคุณภาพเสียงด้วยวิธีการหักล่างทางสเตครัม แสดงไว้ในรูปที่ 4.1 และสามารถอธิบายได้ถึงขั้นตอนต่างๆ ดังต่อไปนี้

1. นำสัญญาณเสียงที่คิดเพี้ยนด้วยสัญญาณรบกวนแบบ White Gaussian ด้วยค่า SNR 0 ถึง 20 เดซิเบล ป้อนเข้าไปใน โปรแกรมการปรับแต่งคุณภาพเสียงพูด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. ทำการวินโดว์สัญญาณเสียงที่ผิดเพี้ยน ด้วย Hamming window ออกเป็น เฟรมๆ ละ 256 จุด โดยให้สัญญาณมีการเลื่อม (overlap) เท่ากับ 128 จุด
3. ทำการแปลงฟูเรียร์ทรานส์ฟอร์ม กับสัญญาณเสียงแต่ละเฟรม ซึ่งหลังจากทำการแปลงฟูเรียร์ทรานส์ฟอร์มแล้ว จะได้เฟสของสัญญาณออกมาด้วย โดยที่เรานำเฟสของข้อมูลมาพักไว้ก่อน
4. หาขนาดของสัญญาณที่ได้ รวมทั้งหาขนาดของสัญญาณรบกวนด้วย
5. ทำการหักล้างทางสเปกตรัมระหว่างสัญญาณเสียงที่ผิดเพี้ยนกับสัญญาณเสียงรบกวน
6. นำผลที่ได้จากขั้นตอนที่ 5 มาผ่านกระบวนการ half-wave rectification แล้วนำค่าที่ได้มารวมกับเฟสของสัญญาณที่เราพักไว้ก่อนหน้า
7. ทำอินเวอร์สฟูเรียร์ทรานส์ฟอร์มกลับ แล้วทำการวัดคุณภาพด้วยค่า SNR
8. จะได้สัญญาณผลลัพธ์ที่ผ่านการปรับแต่งคุณภาพเสียงด้วยวิธีการหักล้างทางสเปกตรัม ซึ่งเราจะแสดงผลโดยการพล็อตรูปคลื่น และสเปกโตแกรม



บทที่ 5

ผลการทดลอง

การทดลองในบทนี้ จะแสดงให้เห็นถึงประสิทธิภาพในการปรับแต่งคุณภาพเสียงพูด โดยวิธีการหักล้างทางสเปกตรัม ซึ่งการทดลองแบ่งออกเป็น 2 ส่วน ดังนี้

5.1 ผลการทดลองกับสัญญาณเสียงที่จำลองขึ้น

การทดลองในส่วนนี้เกิดจากสัญญาณเสียงพูดต้นแบบมีสัญญาณรบกวนแบบเกาส์เซียนผสมอยู่ การทดลองนี้ได้นำสัญญาณเสียงพูดต้นแบบเป็นสัญญาณเสียงพูดประโยค “หนึ่ง สอง สาม สี่ ห้า” และประโยค “เอ็ม ไอ ที” โดยมีสัญญาณรบกวนผสมอยู่ด้วยค่า SNR ตั้งแต่ 0 ถึง 20 เดซิเบล โดยสัญญาณเสียงพูดดังกล่าว ถูกชักตัวอย่างด้วยความถี่ 8 kHz แล้วปฏิบัติตามขั้นตอนการทำงานของกระบวนการกำจัดสัญญาณรบกวนดังรูปที่ 4.1 โดยทำวินโดว์สัญญาณเสียง เฟรมละ 256 จุด และมีการเลื่อนของสัญญาณ 128 จุด จากนั้นจึงทำการแปลง DFT เราจะได้ผลออกมาเป็นจำนวนจินตภาพ ซึ่งเป็นข้อมูลของสัญญาณเสียงที่ผิดเพี้ยน และเฟสของข้อมูล นำสัญญาณเสียงที่ผิดเพี้ยนมาหาขนาด ส่วนเฟสของข้อมูลพักไว้ก่อน พร้อมทั้งหาขนาดของสัญญาณรบกวนในช่วงเวลาที่ไม่มีเสียงพูด (non-speech) จากนั้นนำสัญญาณเสียงมาผ่านกระบวนการหักล้างทางสเปกตรัม โดยนำสัญญาณรบกวน มาหักล้างออกจากสัญญาณเสียงที่ผิดเพี้ยน ผลลัพธ์จากกระบวนการหักล้างทางสเปกตรัม นำมาผ่านขั้นตอนของ Half-wave rectification เพื่อปรับแต่งสัญญาณให้ได้ค่าที่ดีขึ้น ขนาดของสัญญาณเสียงที่ได้ถูกนำไปรวมกับเฟสของข้อมูลที่พักไว้ จากนั้นจึงทำการแปลง-IDFT วัดประสิทธิภาพผลลัพธ์ที่ได้ด้วยค่า SNR

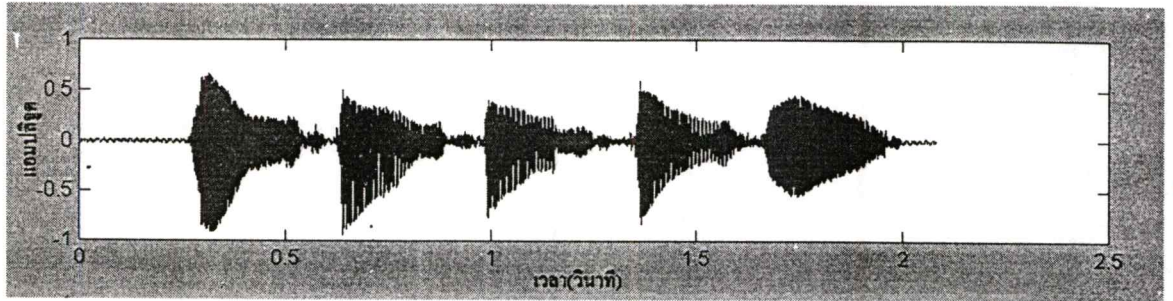
ผลการทดลองดังรูปที่ 5.1 ถึง รูปที่ 5.4 แสดงรูปคลื่นและสเปกโตรแกรมของสัญญาณเสียงพูดประโยค “หนึ่ง สอง สาม สี่ ห้า” โดยรูป (ก) คือสัญญาณเสียงพูดต้นแบบ รูป (ข) คือ สัญญาณเสียงที่ผิดเพี้ยน มีเสียงรบกวนด้วยค่า SNR 5 dB และ 10 dB รูป (ค) คือ สัญญาณเสียงหลังจากกระบวนการปรับแต่งคุณภาพเสียงแล้ว จะพบว่าสัญญาณในรูป (ค) มีลักษณะใกล้เคียงกับสัญญาณในรูป (ก) มาก เช่นเดียวกับผลการทดลองรูปที่ 5.5 ถึง รูปที่ 5.8 ซึ่งแสดงรูปคลื่นและสเปกโตรแกรมของสัญญาณเสียงพูดประโยค “เอ็ม ไอ ที” ก็พบว่าสัญญาณในรูป (ค) มีลักษณะใกล้เคียงกับสัญญาณในรูป (ก) มากเช่นกัน

5.1 ผลการทดลองกับสัญญาณเสียงที่ใช้งานจริง

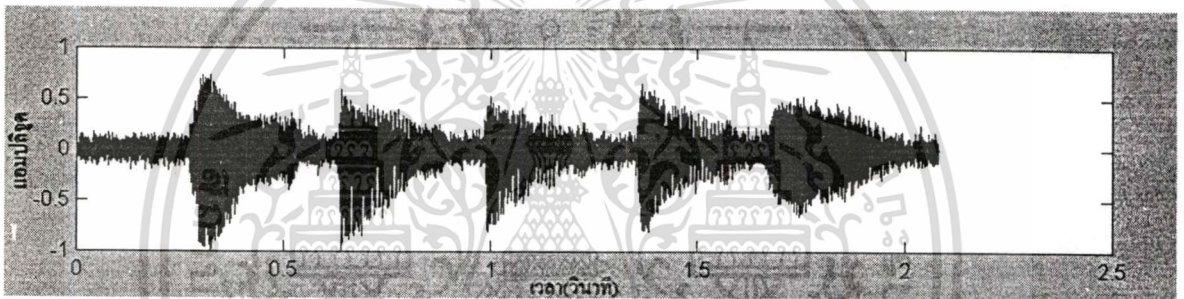
การทดลองในส่วนนี้ เราใช้เสียงที่ใช้งานจริงจากวิทยุ FM ในประโยค “นั่นคือกีฬาประเภทไหน ระหว่าง”, ประโยค “อยากตัดคำว่าวัดออก” และ ประโยค “ส่วนใหญ่ก็เป็นชาวพุทธที่เรียนอยู่” เป็นเสียงพูดที่ผิดเพี้ยน เนื่องจากสัญญาณรบกวนต่างๆ ที่อยู่ในสภาพแวดล้อม จากนั้นนำสัญญาณเสียงพูดที่ผิดเพี้ยนมากระทำการกำจัดสัญญาณรบกวน ตามกระบวนการจากรูปที่ 4.1 ซึ่งผลการทดลองสามารถวัดประสิทธิภาพของเสียงพูดได้ด้วยการฟัง

ผลการทดลองดังรูปที่ 5.9 ถึง รูปที่ 5.14 แสดงรูปคลื่นและสเปกโตรแกรมของสัญญาณเสียงพูดที่ใช้งานจริงจากวิทยุ FM ในประโยค “นั่นคือกีฬาประเภทไหน ระหว่าง”, ประโยค “อยากตัดคำว่าวัดออก” และ ประโยค “ส่วนใหญ่ก็เป็นชาวพุทธที่เรียนอยู่” ตามลำดับ โดยรูป (ก) สัญญาณเสียงที่ผิดเพี้ยนด้วยเสียงรบกวนจากสภาพแวดล้อมต่างๆ (ข) สัญญาณเสียงหลังผ่านกระบวนการปรับแต่งคุณภาพเสียงแล้ว จากการฟังเสียงพบว่าสัญญาณเสียงในรูป (ข) มีเสียงชัดเจนกว่าสัญญาณในรูป (ก) อย่างชัดเจน

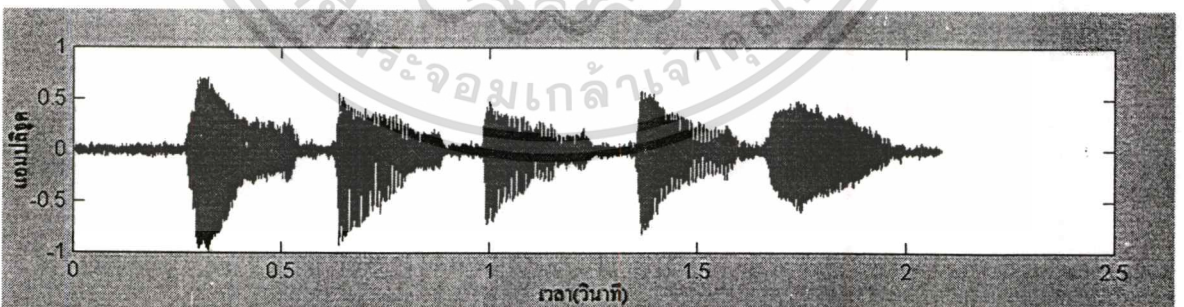




(ก)



(ข)

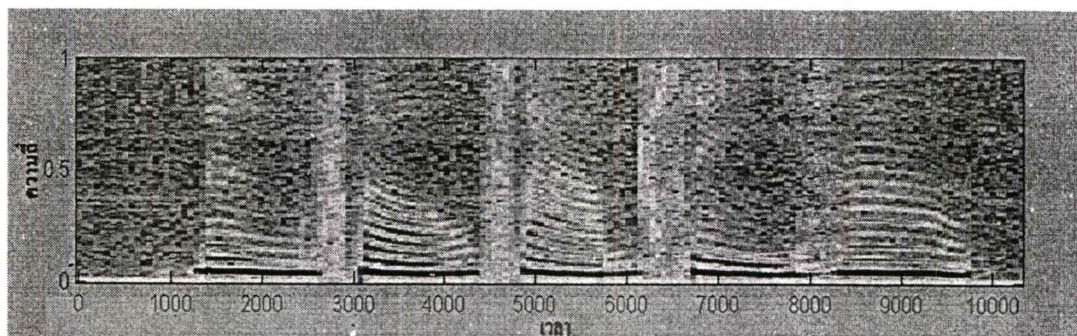


(ค)

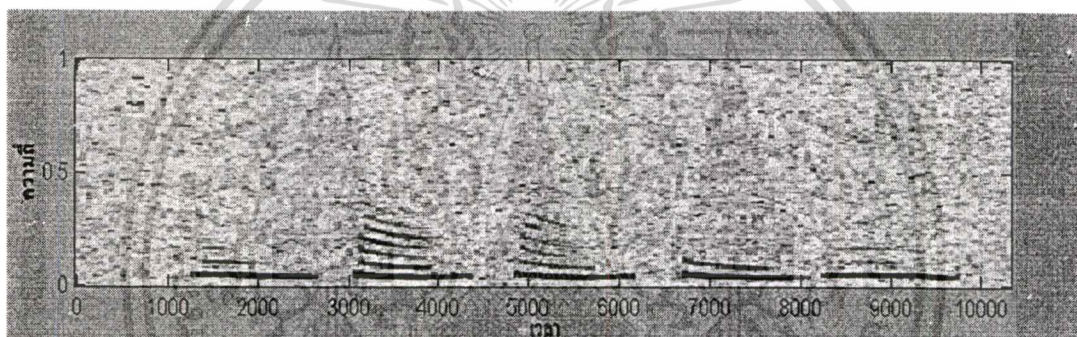
รูปที่ 5.1 รูปคลื่นของสัญญาณเสียงพูด “หนึ่ง สอง สาม สี่ ห้า” (ก) สัญญาณเสียงพูด
ต้นแบบ (ข) สัญญาณเสียงที่ผิดเพี้ยน มีเสียงรบกวนด้วยค่า SNR 10 dB (ค) สัญญาณ

เสียงหลังผ่านกระบวนการปรับแต่งคุณภาพเสียง

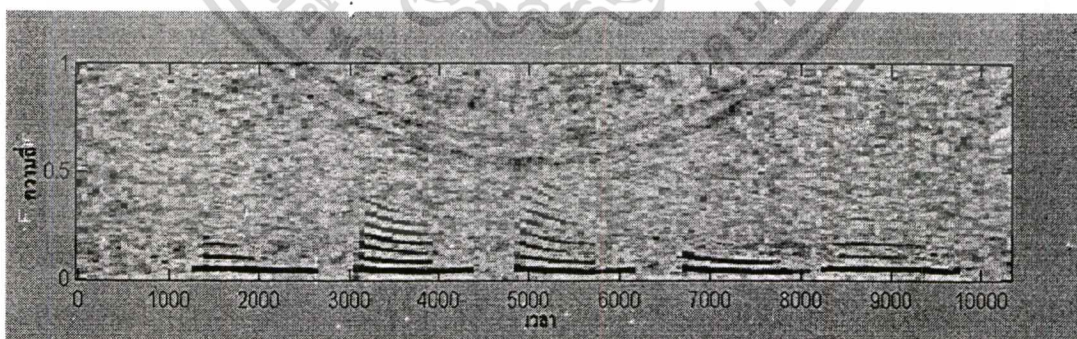
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



(ก)



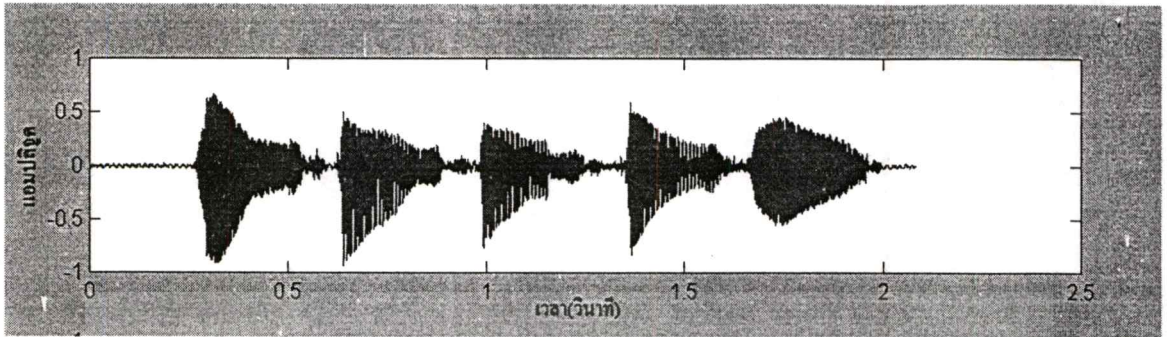
(ข)



(ค)

รูปที่ 5.2 สเปกโตรแกรมของสัญญาณเสียงพูด “หนึ่ง สอง สาม สี่ ห้า” (ก) สัญญาณเสียงพูด
ต้นแบบ (ข) สัญญาณเสียงที่ผิดเพี้ยน มีเสียงรบกวนด้วยค่า SNR 10 dB (ค) สัญญาณ

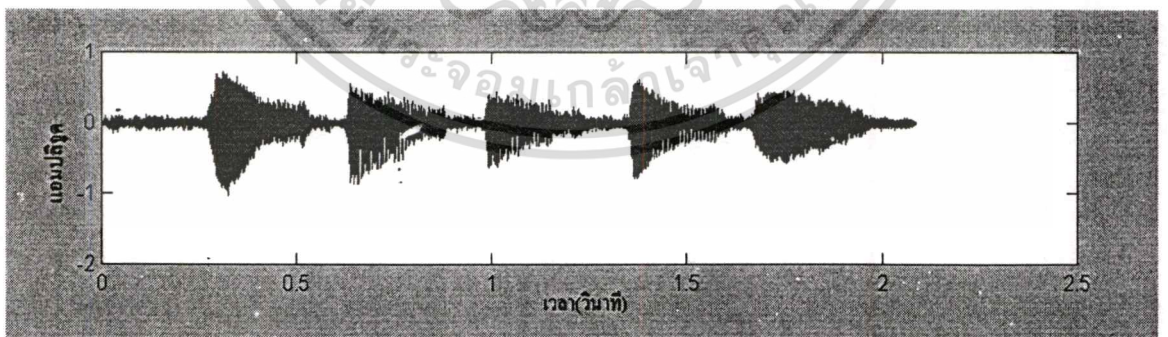
เสียงหลังผ่านกระบวนการปรับแต่งคุณภาพเสียง
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการเรียนเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



(ก)



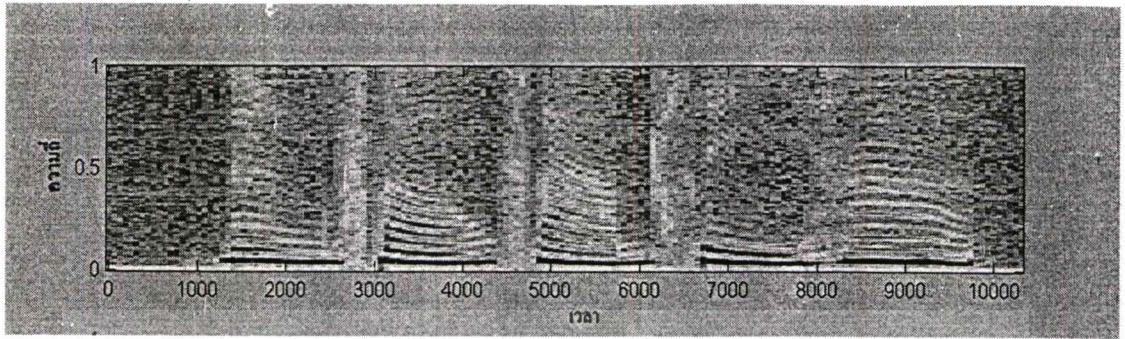
(ข)



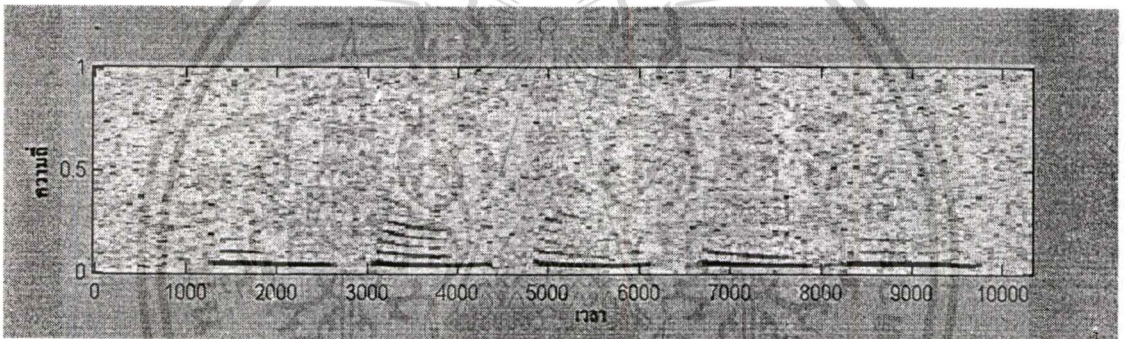
(ค)

รูปที่ 5.3 รูปคลื่นของสัญญาณเสียงพูด “หนึ่ง สอง สาม สี่ ห้า” (ก) สัญญาณเสียงพูดต้นแบบ (ข) สัญญาณเสียงที่ผิดเพี้ยน มีเสียงรบกวนด้วยค่า SNR 5 dB (ค) สัญญาณเสียงหลังผ่านกระบวนการปรับแต่งคุณภาพเสียง

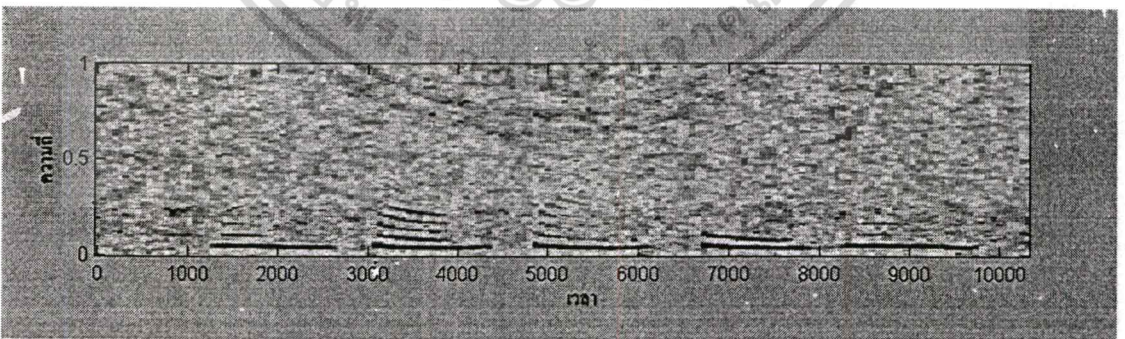
เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ไว้เพื่อใช้ในการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



(ก)



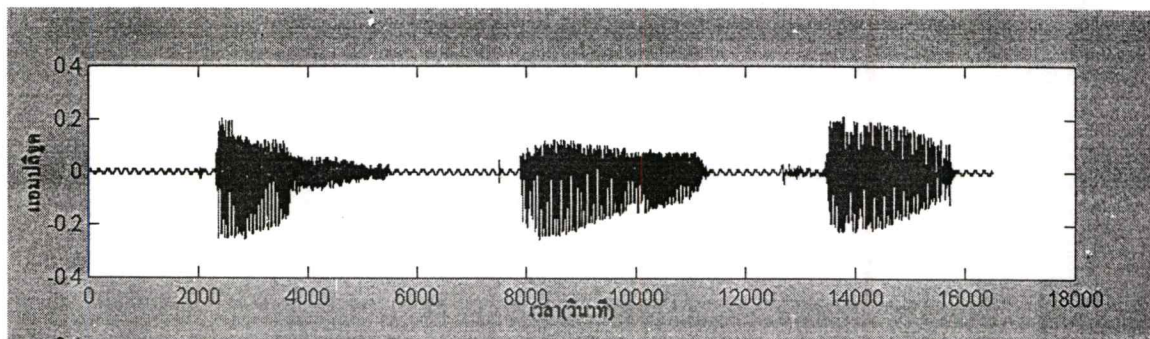
(ข)



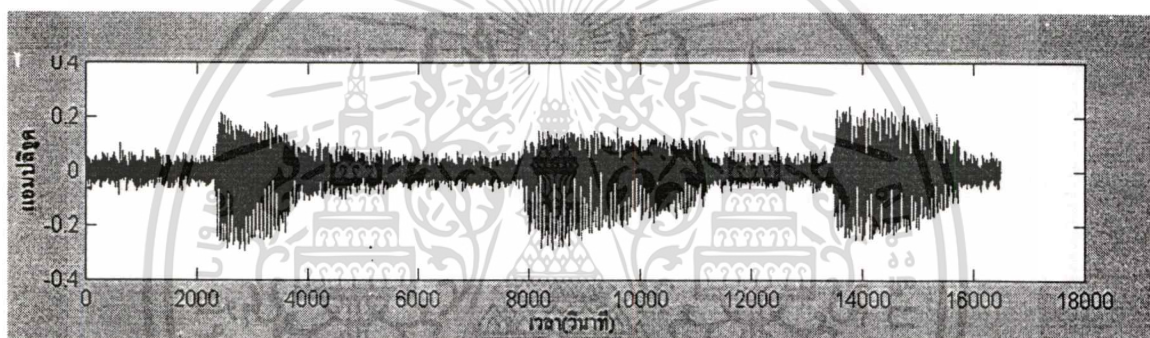
(ค)

รูปที่ 5.4 สเปกโตรแกรมของสัญญาณเสียงพูด “หนึ่ง สอง สาม สี่ ห้า” (ก) สัญญาณเสียงพูดต้นแบบ (ข) สัญญาณเสียงที่ผิดเพี้ยน มีเสียงรบกวนด้วยค่า SNR 5 dB (ค) สัญญาณเสียงหลังผ่านกระบวนการปรับแต่งคุณภาพเสียง

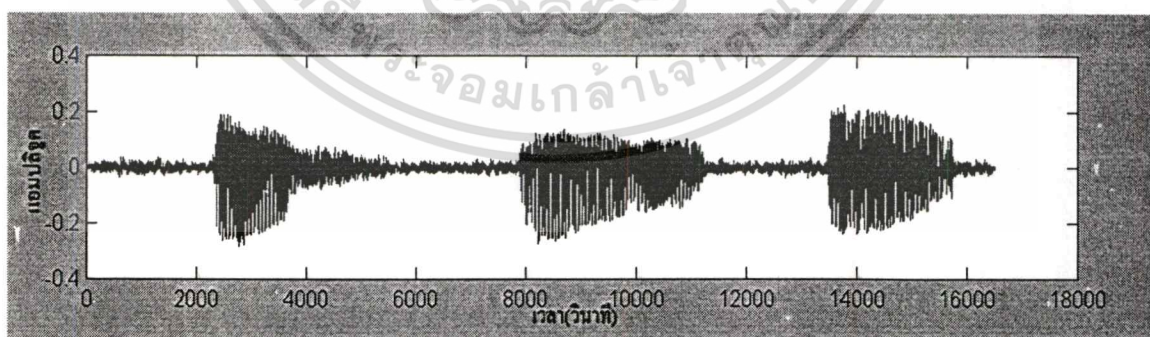
เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ไว้สำหรับการใช้ในแวดวงการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



(ก)



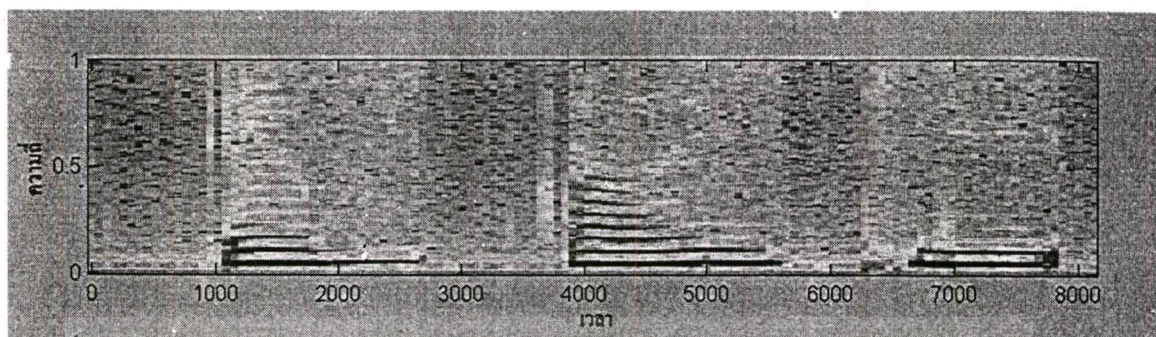
(ข)



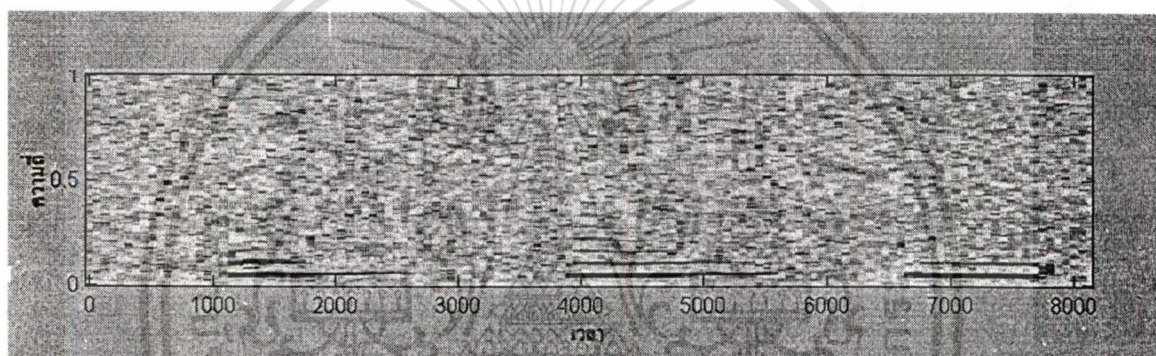
(ค)

รูปที่ 5.5 รูปคลื่นของสัญญาณเสียงพูด “เอ็ม ไอ ที” (ก) สัญญาณเสียงพูดต้นแบบ (ข) สัญญาณเสียงที่ผิดเพี้ยน มีเสียงรบกวนด้วยค่า SNR 10 dB (ค) สัญญาณเสียงหลังผ่านกระบวนการปรับแต่งคุณภาพเสียง

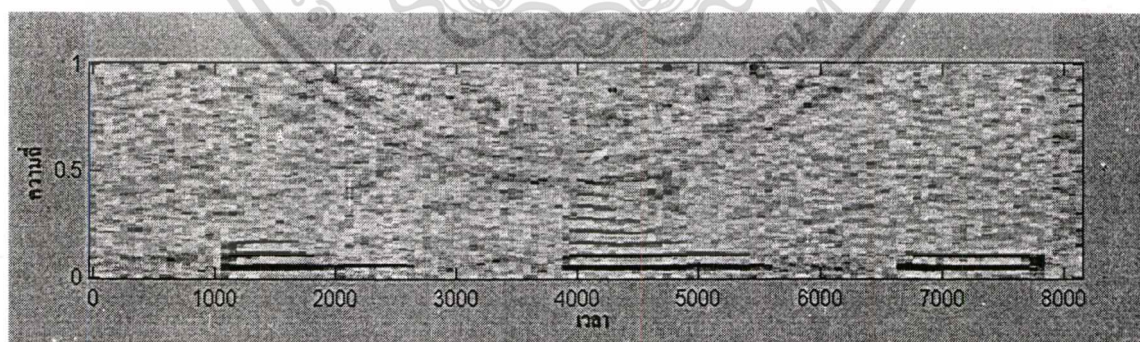
เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ไว้เพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



(ก)



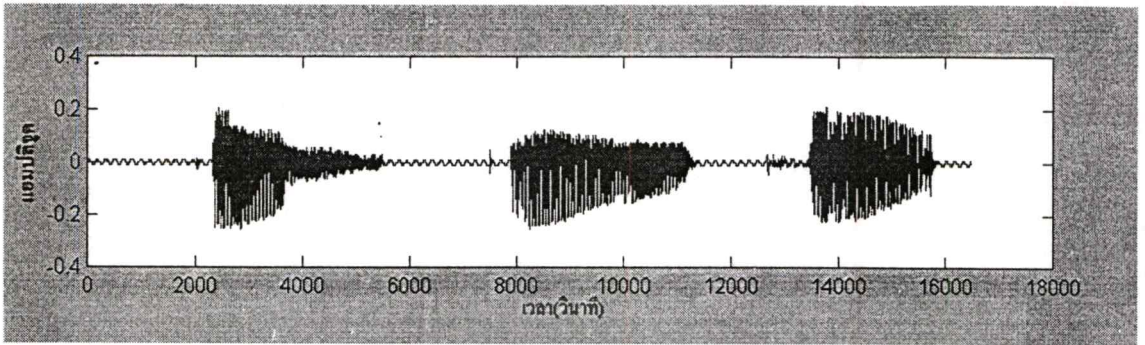
(ข)



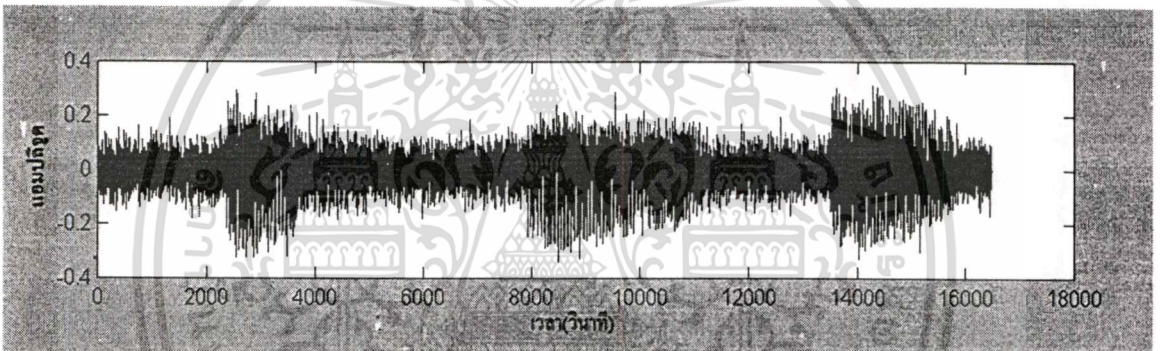
(ค)

รูปที่ 5.6 สเปกโตรแกรมของสัญญาณเสียงพูด “เอ็ม ไอ ที” (ก) สัญญาณเสียงพูดต้นแบบ
 (ข) สัญญาณเสียงที่ผิดเพี้ยน มีเสียงรบกวนด้วยค่า SNR 10 dB (ค) สัญญาณเสียง
 หลังผ่านกระบวนการปรับแต่งคุณภาพเสียง

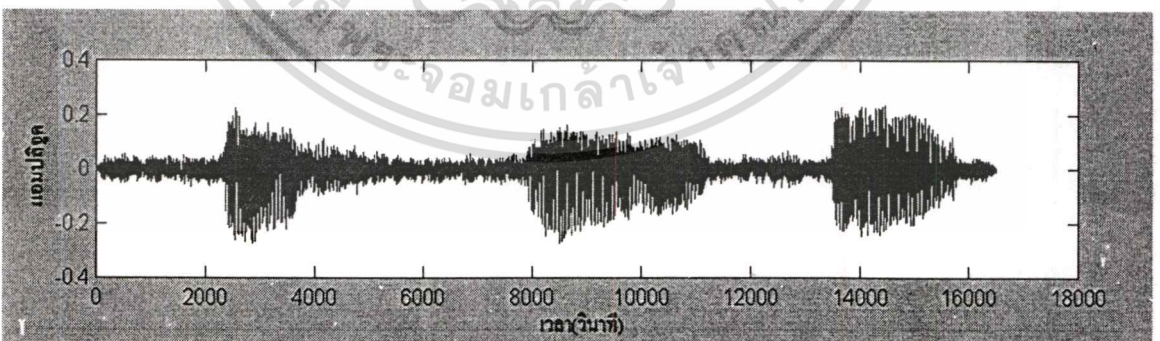
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



(ก)



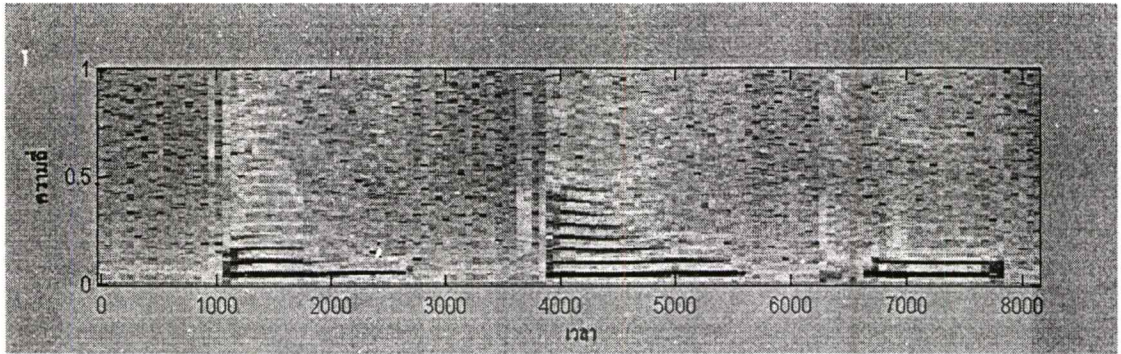
(ข)



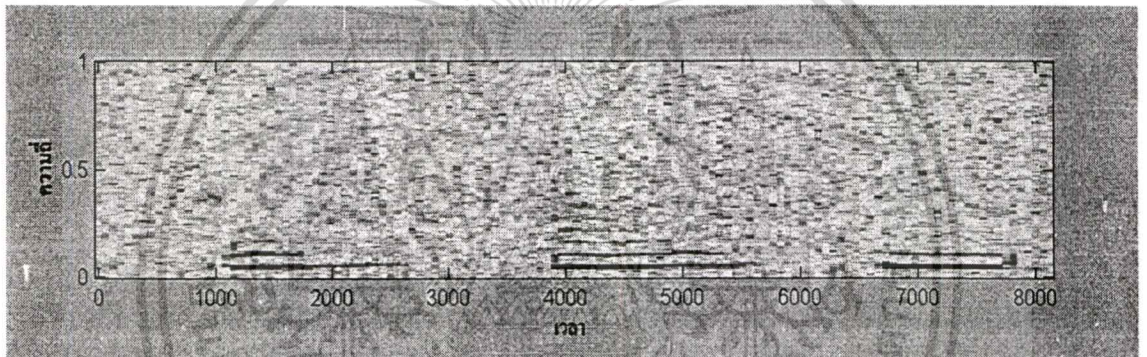
(ค)

รูปที่ 5.7 รูปคลื่นของสัญญาณเสียงพูด “เอ็ม ไอ ที” (ก) สัญญาณเสียงพูดต้นแบบ (ข) สัญญาณเสียงที่ผิดเพี้ยน มีเสียงรบกวนด้วยค่า SNR 5 dB (ค) สัญญาณเสียงหลังผ่านกระบวนการปรับแต่งคุณภาพเสียง

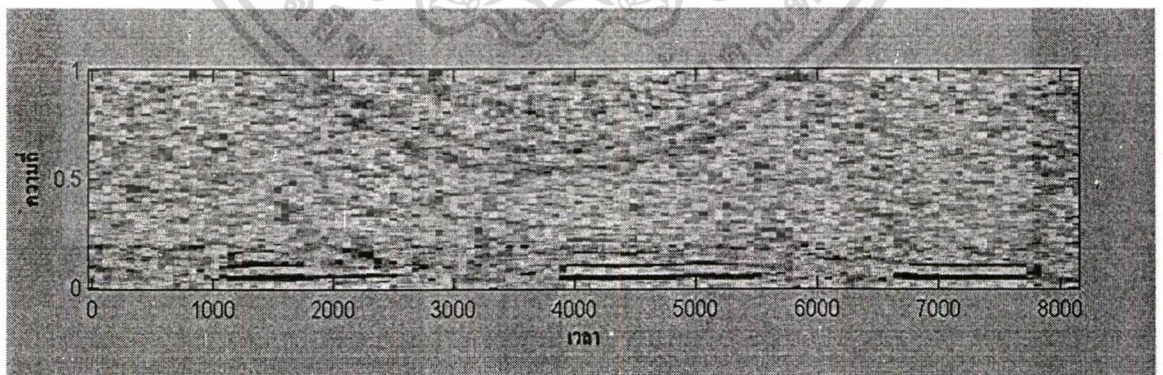
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



(ก)



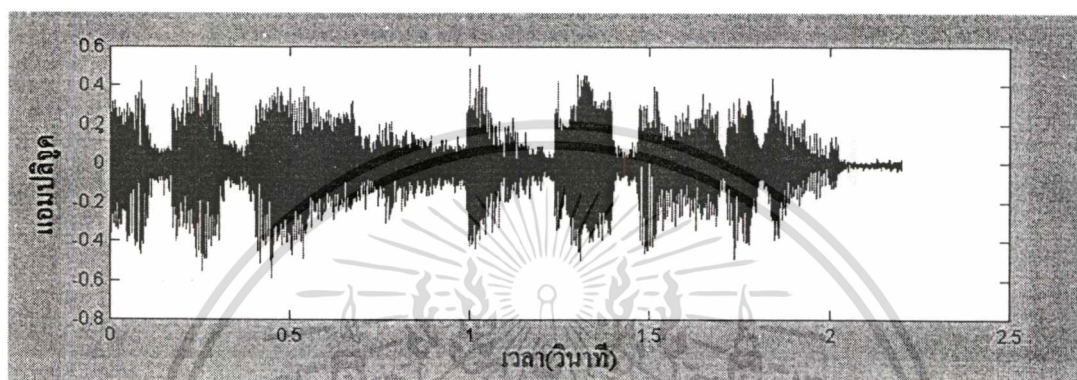
(ข)



(ค)

รูปที่ 5.8 สเปกโตรแกรมของสัญญาณเสียงพูด “เอ็ม ไอ ที” (ก) สัญญาณเสียงพูดต้นแบบ
 (ข) สัญญาณเสียงที่ผิดเพี้ยน มีเสียงรบกวนด้วยค่า SNR 5 dB (ค) สัญญาณเสียง
 หลังผ่านกระบวนการปรับแต่งคุณภาพเสียง

เอกสารนี้เป็นเอกสารลิขสิทธิ์สงวนสำหรับโรงเรียนเอกชนเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

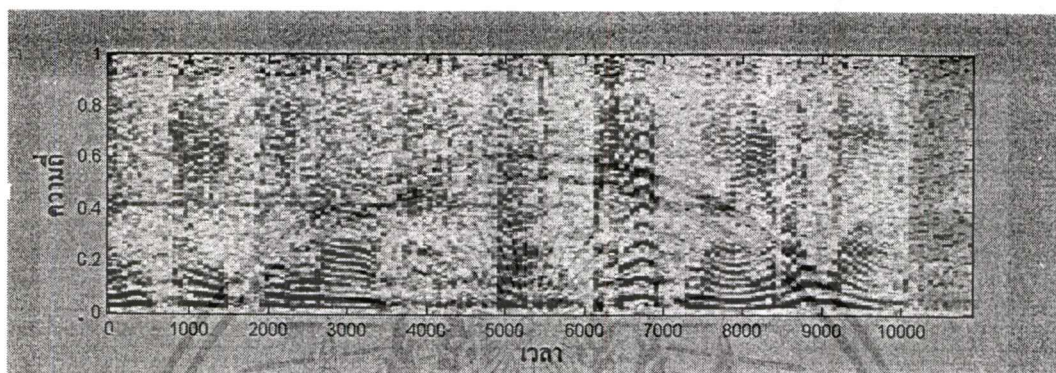


(ก)



(ข)

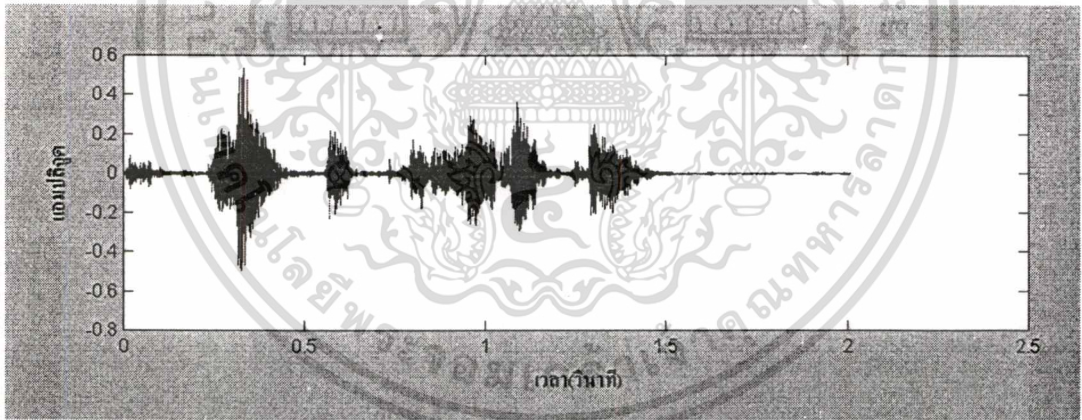
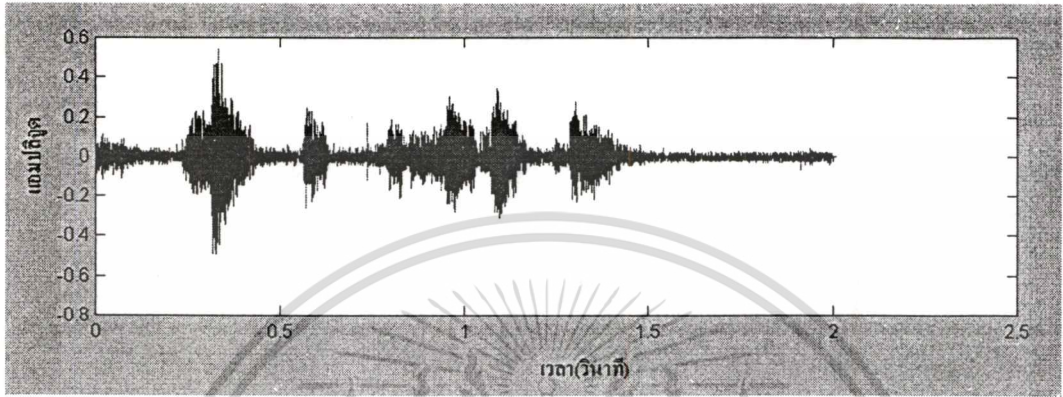
รูปที่ 5.9 รูปคลื่นของสัญญาณเสียงพูดจากวิทยุ FM “นั่นคือกีฬาประเภทไหน ระหว่าง”
 (ก) สัญญาณเสียงที่ผิดเพี้ยนด้วยเสียงรบกวนจากสภาพแวดล้อมต่างๆ (ข) สัญญาณเสียงหลังผ่านกระบวนการปรับแต่งคุณภาพเสียง



(ข)

รูปที่ 5.10 สเปกโตรแกรมของสัญญาณเสียงพูดจากวิทยุ FM “นั่นคือกีฬาประเภทไหน ระหว่าง”
 (ก) สัญญาณเสียงที่ผิดเพี้ยนด้วยเสียงรบกวนจากสภาพแวดล้อมต่างๆ (ข) สัญญาณเสียงหลังผ่านกระบวนการปรับแต่งคุณภาพเสียง

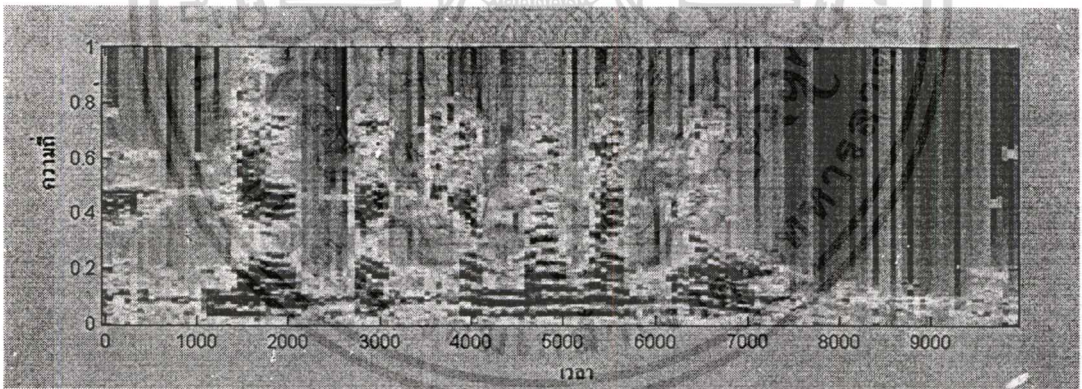
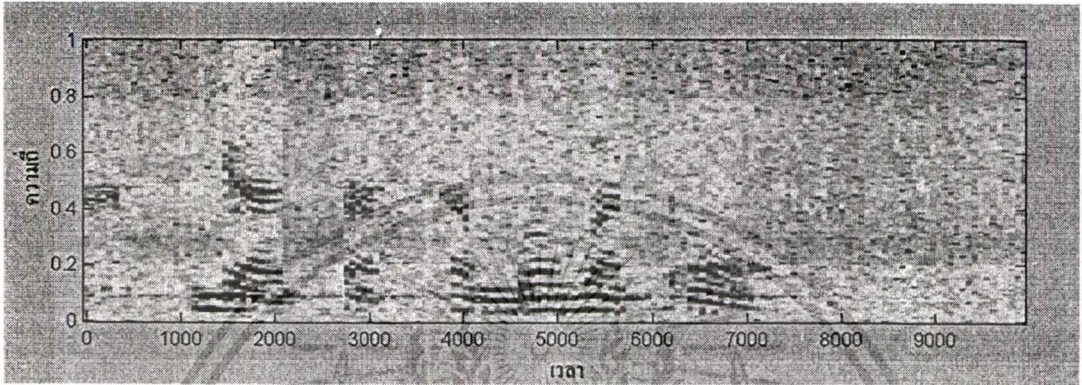
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



(ข)

รูปที่ 5.11 รูปคลื่นของสัญญาณเสียงพูดจากวิทยุ FM “อยากตัดคำว่าวัดออก” (ก) สัญญาณเสียงที่ผิดเพี้ยนด้วยเสียงรบกวนจากสภาพแวดล้อมต่างๆ (ข) สัญญาณเสียงหลังผ่านกระบวนการปรับแต่งคุณภาพเสียง

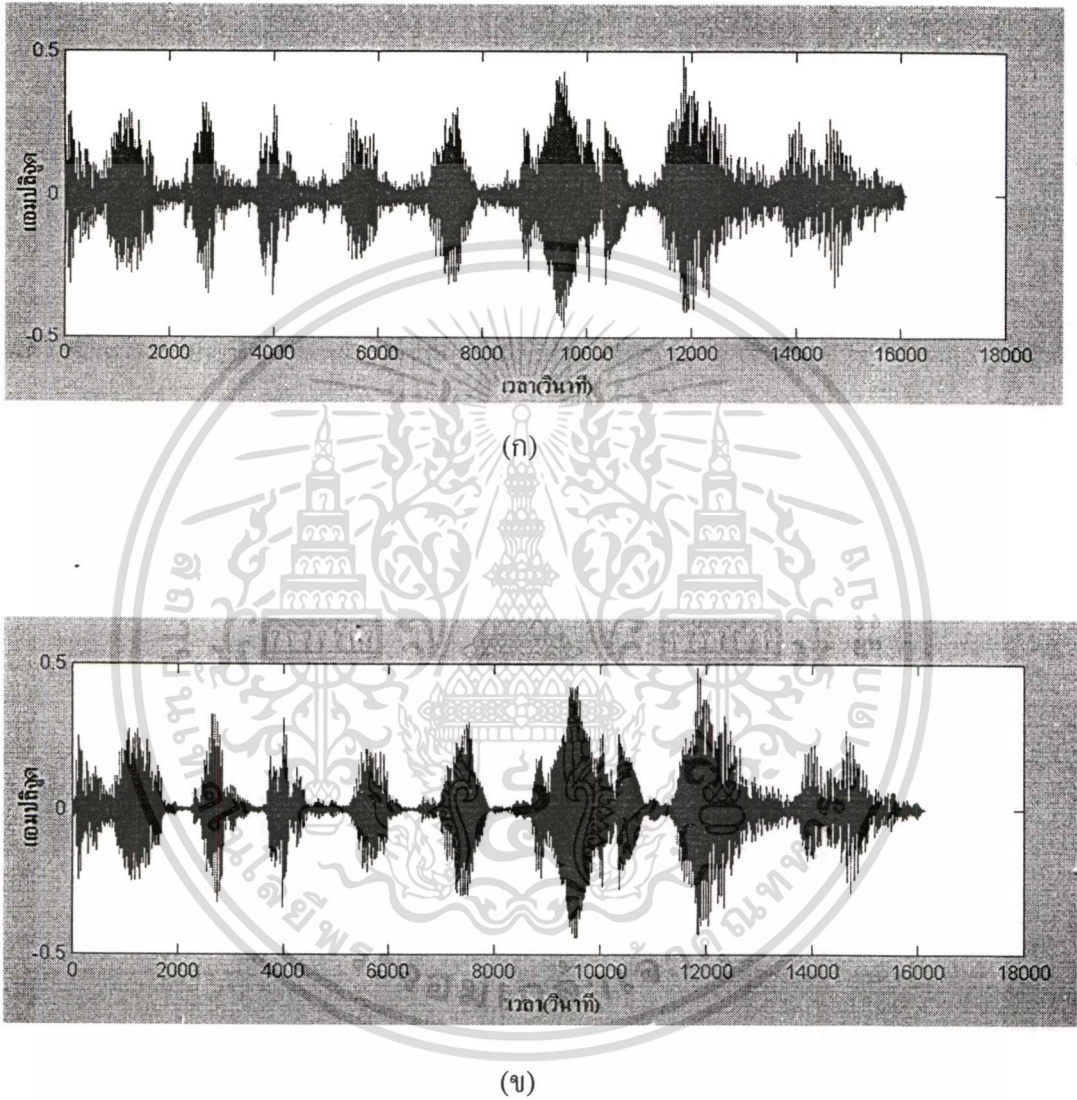
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



(จ)

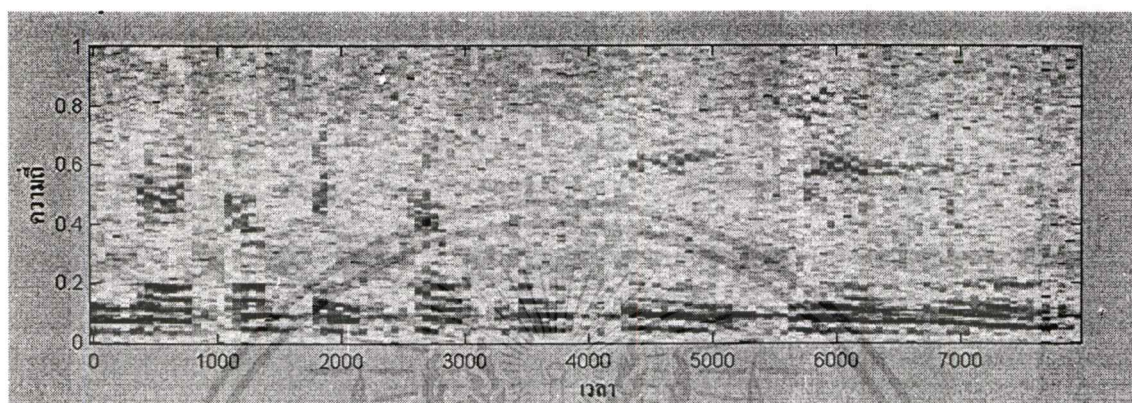
รูปที่ 5.12 สเปกโตรแกรมของสัญญาณเสียงพูดจากวิทยุ FM “อยากตัดคำว่าวัดออก” (ก) สัญญาณเสียงที่ผลิตขึ้นด้วยเสียงรบกวนจากสภาพแวดล้อมต่างๆ (ข) สัญญาณเสียงหลังผ่านกระบวนการปรับแต่งคุณภาพเสียง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

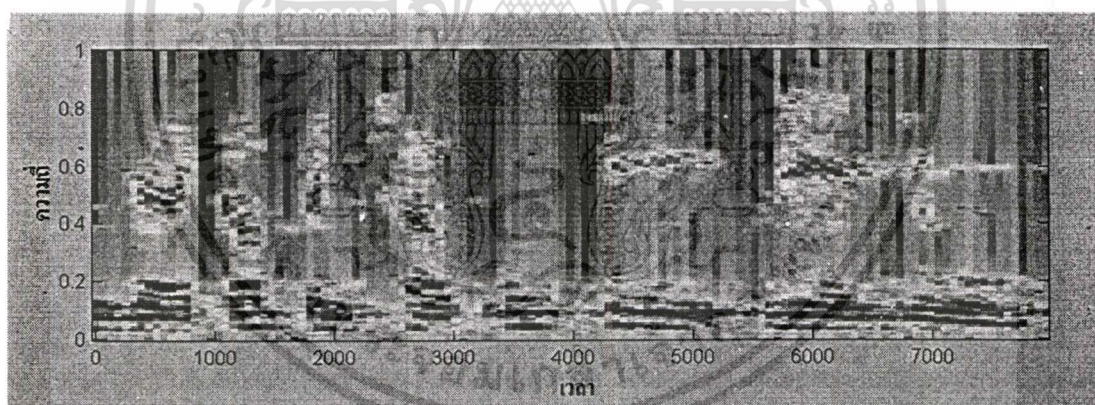


รูปที่ 5.13 รูปคลื่นของสัญญาณเสียงพูดจากวิทยุ FM “ส่วนใหญ่ก็เป็นชาวพุทธที่เรียนอยู่”
 (ก) สัญญาณเสียงที่ผิดเพี้ยนด้วยเสียงรบกวนจากสภาพแวดล้อมต่างๆ (ข) สัญญาณเสียงหลังผ่านกระบวนการปรับแต่งคุณภาพเสียง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



(ก)



(ข)

รูปที่ 5.14 สเปกโตรแกรมของสัญญาณเสียงพูดจากวิทยุ FM “ส่วนใหญ่ก็เป็นชาวพุทธที่เรียนอยู่”
 (ก) สัญญาณ-เสียงที่ผิดเพี้ยนด้วยเสียงรบกวนจากสภาพแวดล้อมต่างๆ (ข) สัญญาณ-เสียงหลังผ่านกระบวนการปรับแต่งคุณภาพเสียง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับทำใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่าจะกรณีใดๆทั้งสิ้น. อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.1 ค่า SNR ของสัญญาณเสียงพูดประโยค “หนึ่ง สอง สาม สี่ ห้า” ก่อนและหลังผ่านกระบวนการปรับแต่งคุณภาพเสียงด้วยวิธีการหักล้างทางสเปกตรัม

อัตราส่วนสัญญาณต่อสัญญาณรบกวน(dB)	
สัญญาณเสียงพูดที่ผิดเพี้ยน	สัญญาณเสียงพูดที่ปรับแต่งแล้ว
20.00	21.57
15.00	19.80
10.00	17.06
5.00	13.49
0.00	8.69

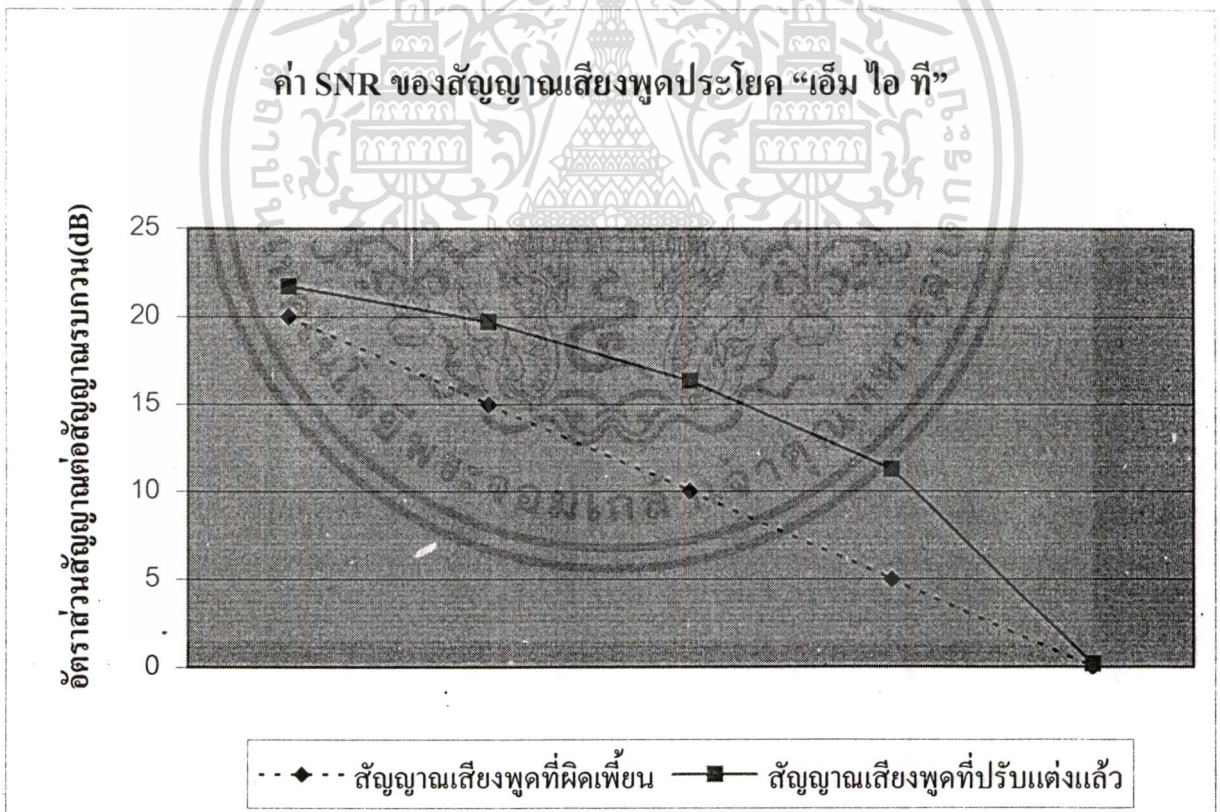


รูปที่ 5.15 กราฟค่า SNR ของสัญญาณเสียงพูดประโยค “หนึ่ง สอง สาม สี่ ห้า”

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 5.2 ค่า SNR ของสัญญาณเสียงพูดประโยค “เอ็ม ไอ ที” ก่อนและหลังผ่านกระบวนการปรับแต่งคุณภาพเสียงด้วยวิธีการหักล้างทางสเปกตรัม

อัตราส่วนสัญญาณต่อสัญญาณรบกวน(dB)	
สัญญาณเสียงพูดที่ผิดเพี้ยน	สัญญาณเสียงพูดที่ปรับแต่งแล้ว
20.00	21.74
15.00	19.69
10.00	16.32
5.00	11.26
0.00	0.20



รูปที่ 5.16 กราฟค่า SNR ของสัญญาณเสียงพูดประโยค “เอ็ม ไอ ที”

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 6

สรุปผลการวิจัย

การปรับแต่งคุณภาพเสียงมีด้วยกันหลายวิธีการ การกำจัดสัญญาณเสียงรบกวนออกจากสัญญาณเสียงที่ผิดเพี้ยน ก็เป็นวิธีหนึ่งที่ทำให้เสียงมีคุณภาพดีขึ้น ซึ่งวิธีการกำจัดสัญญาณรบกวนก็มีด้วยกันหลายวิธี กระบวนการหักล้างทางสเปกตรัมก็เป็นวิธีหนึ่งที่มีอัลกอริทึมที่มีประสิทธิภาพในการกำจัดสัญญาณรบกวน และยังมีโมเดลที่ง่ายต่อการศึกษาเพื่อจะนำไปประยุกต์ใช้อีกด้วย ในวิทยานิพนธ์ฉบับนี้ทำการปรับแต่งคุณภาพเสียงที่ผิดเพี้ยนจากสัญญาณรบกวนแบบเกาส์เซียนโดยใช้วิธีการหักล้างการสเปกตรัม เริ่มจากการประมาณค่าสัญญาณรบกวน ระหว่างช่วงเงียบ (non-speech) เพื่อนำมาหักล้างจากสัญญาณเสียงที่ผิดเพี้ยน ผลลัพธ์ที่ได้วัดประสิทธิภาพด้วยค่า SNR และนำมาเปรียบเทียบกับสัญญาณเสียงที่ผิดเพี้ยน พร้อมทั้งแสดงการเปรียบเทียบกับสัญญาณเสียงต้นแบบ ในรูปของการพล็อตรูปคลื่นและสเปกโตรแกรมของสัญญาณ

ในการทดลองเพื่อทดสอบประสิทธิภาพของอัลกอริทึมการปรับแต่งคุณภาพเสียงโดยวิธีการหักล้างทางสเปกตรัม ใช้สัญญาณเสียงที่จำลองขึ้น และสัญญาณเสียงที่งานจริง จากวิทยุ FM ผลการทดลองแสดงให้เห็นถึงประสิทธิภาพของอัลกอริทึม ซึ่งสามารถกำจัดสัญญาณรบกวนอย่างได้ผล เนื่องจากค่า SNR ของสัญญาณสูงขึ้นอย่างชัดเจน สำหรับการทดลองที่ใช้กับสัญญาณเสียงที่ใช้งานจริง พบว่า จากการฟังสัญญาณเสียงผลลัพธ์มีความชัดเจนขึ้น เนื่องจากเสียงของสัญญาณรบกวนขาดหายไป แม้ว่าแอมพลิจูดของสัญญาณเสียงจะต่ำลง ไปบ้าง

แนวทางในการพัฒนางานวิจัยต่อไปนั้น อัลกอริทึมของการปรับแต่งคุณภาพเสียง อาจจะต้องมีความสามารถในการกำจัดสัญญาณรบกวนหลายๆ แบบเพื่อปรับแต่งคุณภาพของเสียงพูดให้ดีขึ้น เนื่องจากสภาพแวดล้อมตามความเป็นจริงนั้น เราจะไม่สามารถทราบได้ว่า สัญญาณเสียงที่ผิดเพี้ยนไปนั้น เกิดจากสัญญาณรบกวนแบบใด ทั้งนี้เราอาจจะเพิ่มเติมกระบวนการกำจัดสัญญาณรบกวนที่ตกค้างในสัญญาณเสียงผลลัพธ์และเพิ่มส่วนที่จะทำให้แอมพลิจูดของสัญญาณเสียงผลลัพธ์มีค่าคงเดิม เพื่อที่จะได้สัญญาณเสียงที่ใกล้เคียงกับเสียงต้นแบบมากที่สุด งานวิจัยนี้เป็นเพียงการศึกษาอัลกอริทึม ด้วยการจำลองการทำงานด้วยโปรแกรมคอมพิวเตอร์ยังไม่สามารถประมวลผลให้เป็นแบบ real time ได้ หากต้องการให้อัลกอริทึมทำงานแบบ real time จำเป็นต้องมีตัวประมวลผลที่เป็นฮาร์ดแวร์มาเกี่ยวข้อง เพื่อที่จะสามารถช่วยในการประมวลผลให้เร็วขึ้น จึงขอเสนอให้เพื่อเป็นแนวทางในการพัฒนางานวิจัยเพื่อให้กระบวนการปรับแต่งคุณภาพเสียงเป็นไปอย่างสมบูรณ์ที่สุด

บรรณานุกรม

- [1] สุนทร อรอินทร์, อัฐ เครือฟัก, การประมวลเสียงพูดโดยการประมาณเชิงเส้น, ปรินซ์ตันนิพนธ์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง, พ.ศ. 2537.
- [2] J. G. Proakis, and D. G. Manolakis, "Digital Signal Processing, Principles, Algorithm, and Application", 3rd ed., Prentice Hall, NJ, 1996.
- [3] W. W. Smith, and J. M. Smith, "Handbook of Real-Time Fast Fourier Transforms", IEEE Press, Piscataway, NJ, 1995.
- [4] C. E. Shannon, and W. Weaver, "The Mathematical Theory of Communication", University of Illinois Press, Urbana, Illinois, 1949.
- [5] T. Miura, and T. Koshikawa, "A study of speech level distribution and spectra (J)", Electr. Commun. Lab. Tech. J. 4, pp. 45. 1955.
- [6] W. B. Davenport, "A study of speech probability distribution", MIT Technical Report No. 148, Massachusetts Institute of Technology, Cambridge, Massachusetts. 1950.
- [7] S. Saito, K. Kato, and N. Teranishi, "Statistical properties of the fundamental frequencies on Japanese speech voices (J)", JASJ 14, pp. 111, 1958.
- [8] R. K. Potter, and J. C. Steinberg, "Toward the specification of speech", J. Acoust. Soc. Am. 22(6), pp. 807, 1950.
- [9] A. M. Liberman, F. Ingeman, L. Lisker, p. Delattre, and F. S. Cooper, "Minimal rules for synthesizing speech", J. Acoust. Soc. Am. 31(11), pp. 1490, 1959.
- [10] R. Jakobson, G. Fant, and M. Halle, "Preliminaries to speech analysis", MIT Press, Cambridge, Massachusetts. 1963.

ภาคผนวก

ผลงานวิจัย

Sumathee Napakas and Yuttapong Rangsanseri "Speech Enhancement Using Spectral Subtraction Technique" Proceeding of ICT, Vol.1, pp 340-343, 2002.



Speech Enhancement Using Spectral Subtraction Technique

Sumathee Napakas and Yuttapong Rangsanseri

Department of Telecommunications Engineering, Faculty of Engineering
King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand
e-mail: rangsanseri@yahoo.com

ABSTRACT

An enhancement of speech signal algorithm is presented for reducing the spectral effects of acoustically added noise in speech. The spectral subtraction technique is used to remove noise components from noisy speech in order to improve the signal to noise ratio of the signal. Suppression stationary noise from speech by subtracting the noise spectrum is obtained from the signal measured during non-speech activity while the phase of noisy speech was kept. Performance assessment based on the time waveform and spectrogram plots, objective measures and informal subjective listening tests all indicate that our method gives consistently good results.

1. INTRODUCTION

Degradation of the quality of speech caused by the acoustic background noise is common in most of speech processing applications such as mobile communication. One way to reduce this problem is removing noise components from the noisy speech. Speech enhancement has been widely studying since the past because it is an important issue in the field of speech research [1]. The spectral subtraction technique [2] is used to remove noise components from noisy speech in order to improve the signal to noise ratio of the signal.

In this paper, we discuss some developments in speech enhancement. Our focus is on suppression noise from speech signal. We start by reviewing, in the first section, the spectral subtraction technique. This section includes noise model, windowing data, spectral subtraction estimator, noise estimator; half-wave rectification and system block diagram. Experimental results are provided in the second section. A conclusion is given in the last section.

2. SPECTRAL SUBTRACTION TECHNIQUE

The proposed speech enhancement is show in Figure 1, and is summarized as follows.

- Input data is the noisy speech signal.
- Data are segmented and windowed that if a sequence is separated into half-overlapped data buffers and each buffer is multiplied by Hamming window.
- Compute the discrete Fourier transform and keep the phase information of noisy speech. Since real data are being transformed, two data windows can be transformed using one FFT. The FFT size is set equal to the window size of 256. Augmentation with zero was not incorporated.

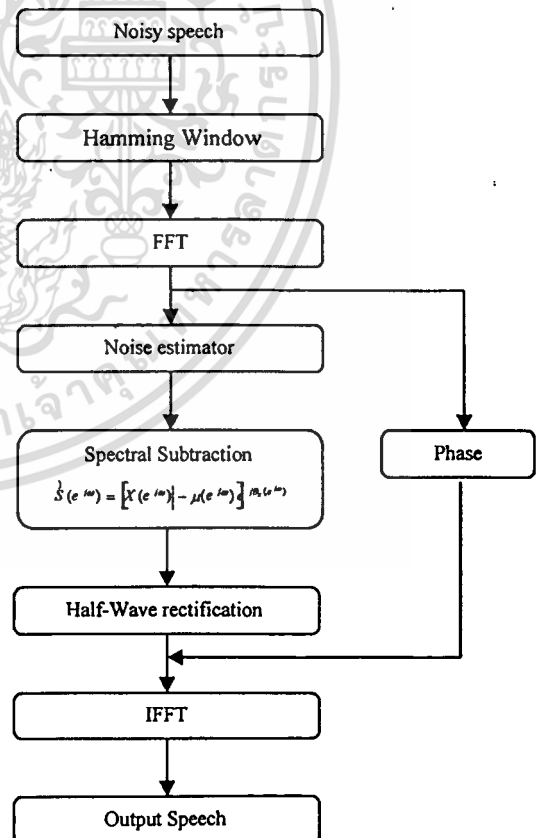


Figure 1 Block diagram of the spectral subtraction method.

- Compute the noisy speech magnitude, the averaging of noise magnitude and the noise estimator.
- Spectral subtraction analysis.
- Half-wave rectification analysis and additive by phase of noisy speech.
- Compute the inverse discrete Fourier transform.

This section describes the theory of enhancement system. The details include noise model, windowing data, spectral subtraction estimator, noise estimator and half-wave rectification.

2.1 Noise model

Additive noise model is an ingredient of an algorithm. Assume that a noise signal has been added to a speech signal, with their sum is the noisy speech:

$$x(k) = s(k) + n(k)$$

where $s(k)$ is an original clean speech signal, $n(k)$ is a noise signal, and $x(k)$ is a noisy speech.

Taking the Fourier transform gives

$$X(e^{j\omega}) = S(e^{j\omega}) + N(e^{j\omega})$$

where $x(k) \leftrightarrow X(e^{j\omega})$,

$$X(e^{j\omega}) = \sum_{k=0}^{L-1} x(k)e^{-j\omega k},$$

$$\text{and } x(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{j\omega}) e^{j\omega k} d\omega$$

2.2 Windowing data

The speech signal from the A-D converter is segmented and windowed such that in the absence of spectral modifications, if the speech segments are added together, the resulting overall system reduces to an identity. The data are segmented and windowed using the result that if a sequence is separated into half-overlapped data buffers and each buffer is multiplied by Hamming window.

The Hamming window is an extension of Hann window [3] in the sense that it is a raise cosine window of the form

$$h(n) = \alpha + (1.0 - \alpha) \cos\left[\left(\frac{2\pi}{N}\right)n\right]$$

with a corresponding spectrum of the form

$$H(\theta) = \alpha D(\omega) + \frac{(1.0 - \alpha)}{2} \left[D\left(\omega - \frac{2\pi}{N}\right) + D\left(\omega + \frac{2\pi}{N}\right) \right]$$

The parameter α permits the optimization of the destructive sidelobe cancellation mentioned in the description of the Hann window. In particular, when α is adjusted to 25/46 (0.543478261...), the first sidelobe is canceled. The common approximation to this value of α is 0.54, for which the window is called the Hamming window and is of the form

$$H(\theta) = 0.54 + 0.46 \cos\left[\left(\frac{2\pi}{N}\right)n\right]$$

2.3 Spectral subtraction estimator

The spectral subtraction filter $H(e^{j\omega})$ is calculated by replacing the noise spectrum $N(e^{j\omega})$ with spectra which can be readily measured [4]. The magnitude $|N(e^{j\omega})|$ of $N(e^{j\omega})$ is replaced by its average value $\mu(e^{j\omega})$ taken during non-speech activity, and the phase $\theta_N(e^{j\omega})$ of $N(e^{j\omega})$ is replaced by the phase $\theta_X(e^{j\omega})$ of $X(e^{j\omega})$. These substitutions result in the spectral subtraction estimator $\hat{S}(e^{j\omega})$:

$$\hat{S}(e^{j\omega}) = \left[|X(e^{j\omega})| - \mu(e^{j\omega}) \right] e^{j\theta_X(e^{j\omega})}$$

$$\hat{S}(e^{j\omega}) = H(e^{j\omega}) X(e^{j\omega})$$

$$\text{where } H(e^{j\omega}) = 1 - \frac{\mu(e^{j\omega})}{|X(e^{j\omega})|}$$

$$\text{and } \mu(e^{j\omega}) = E\left\{ |N(e^{j\omega})| \right\}$$

2.4 Noise estimator

The noise-suppressed estimator is obtained by subtracting an estimate of the noise spectrum from the noisy speech spectrum. The background noise environment remains locally stationary to the degree

that the noise spectrum is obtained from the signal measured during non-speech activity:

$$\mu(e^{j\omega}) = \frac{1}{N} \sum_{i=0}^{N-1} |X_i(e^{j\omega})|$$

where $\mu(e^{j\omega})$ is an averaging of noise magnitude, $X_i(e^{j\omega})$ is a noisy speech, and N is the number of data in the first frame.

In stationary state, there exists enough time (about 300 ms) to estimating a background noise spectral magnitude and we choose the first frame of data (256 points).

2.5 Half-wave Rectification

For each frequency ω where the noisy signal spectrum magnitude $X(e^{j\omega})$ is less than the average noise spectrum magnitude $\mu(e^{j\omega})$, the output is set to zero. This modification can be simply implemented by half-wave rectifying $H(e^{j\omega})$. The estimator then becomes:

$$\hat{S}(e^{j\omega}) = H_R(e^{j\omega})X(e^{j\omega})$$

where

$$H_R(e^{j\omega}) = \frac{H(e^{j\omega}) + |H(e^{j\omega})|}{2}$$

The spectral subtraction spectral estimate \hat{S} is obtained subtracting the noise magnitude μ from noisy speech magnitude $|X|$:

$$|\hat{S}(k)| = |X(k)| - \mu \quad k = 0, 1, \dots, L-1$$

or

$$\hat{S}(k) = H(k) \cdot X(k),$$

$$H(k) = 1 - \frac{\mu(k)}{|X(k)|} \quad k = 0, 1, \dots, L-1$$

where L is discrete Fourier transform buffer length. After subtracting, the difference values \hat{S} having negative magnitudes are set to zero.

3. EXPERIMENTAL RESULTS

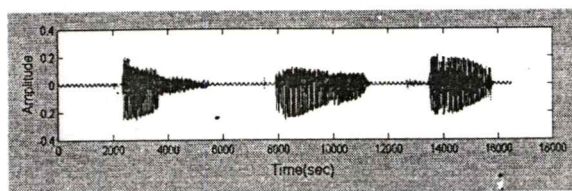
The proposed speech enhancement algorithm has been tested on the spoken English passages "M-I-T". The speech signals corrupted by additive white Gaussian noise with various global signal to noise ratio from 0 dB to 20 dB. Speech signal has the 8 kHz sampling rate. A window length of 256 points shifted in steps of 128 points was used. The discrete Fourier transform of each data window is taken while the phase of data was kept and the magnitude is computed. The suppression stationary noise from noisy speech by subtracting the noise spectrum is obtained from the signal measured during non-speech activity. The result from subtracting is less than the average noise spectrum magnitude, the output is set to zero (half-wave rectification).

Table 1 Comparison of signal to noise ratio.

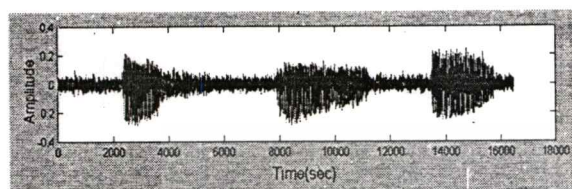
Signal to noise ratio (dB)	
Noisy speech	Enhanced
20.00	21.74
15.00	19.69
10.00	16.32
5.00	11.26
0.00	0.20

The phase of data is combined in this section. We compute the inverse discrete Fourier transform of each data window. Experimental result is measured by the signal to noise ratio of the signal and showed in Table 1.

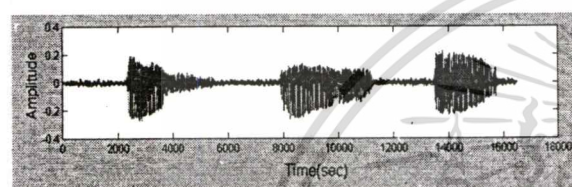
Figure 2 shows the waveform of clean speech "M-I-T". The noisy speech of 10 dB SNR and the enhanced speech signal. We can be seen that the enhance speech signal (c) becomes very close to the original clean speech (a). Likewise Figure 3 shows the spectrogram of speech in the same sentence. The most popular representation of this type is the sound spectrogram [5] in which a three-dimensional representation of the speech intensity, in different frequency bands, over time is portrayed. From this figure, the enhanced speech signal (c) is similar to the original clean speech (a).



(a)

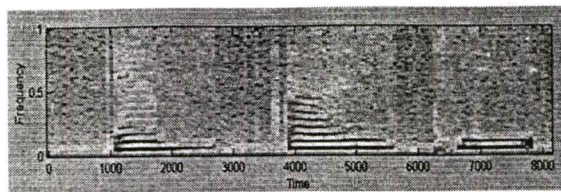


(b)

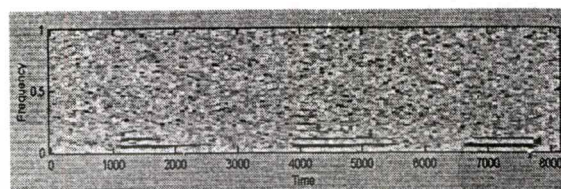


(c)

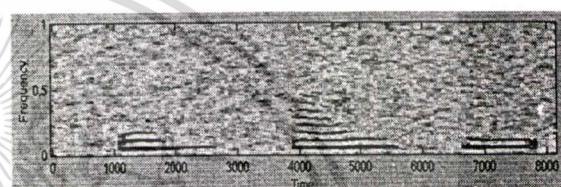
Figure 2 Time waveform of English speech "M-I-T".
(a) Original clean speech. (b) Noisy speech SNR 10 dB. (c) Enhanced speech



(a)



(b)



(c)

Figure 3 Spectrogram of English speech "M-I-T".
(a) Original clean speech. (b) Noisy speech SNR 10 dB. (c) Enhanced speech.

4. CONCLUSION

This paper proposed the speech enhancement algorithm using spectral subtraction technique. Results presented for additive stationary white Gaussian noise show the algorithm to be effective. The experimental results show that the suggested method gives the improvement in signal to noise ratio. The algorithm is potentially extendible to non-stationary noise and this will be the subject of future investigations. Considering the simplicity of the proposed algorithm, the results have proven to be very promising.

REFERENCES

- [1] Vaseghi, S.V., *Advanced Digital Signal Processing and Noise Reduction*, 2nd ed. John Wiley & Sons, 2000.
- [2] Boll, S.F., Suppression of Acoustic Noise in Speech using Spectral Subtraction. *IEEE Trans. Acoustic, Speech, and Signal Processing*. 27 (2):113-120, 1979.
- [3] Oppenheim, A.V., Shafer, R.W., and Buck, J.R. *Discrete-Time Signal Processing*, 2nd ed., Prentice-Hall, 1999.

- [4] Bodin, P., and Villemoes, L.F., Spectral Subtraction in Time-Frequency Domain using Wavelet Packets. *Proc. Speech Coding for Telecommunication*. pp 47-48, 1997.
- [5] Rabiner, L. and Juang, B.H. *Fundamentals of Speech Recognition*. pp 19-20, Prentice-Hall, 1993.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประวัติผู้เขียน

- ชื่อ นางสาวสุเมธิ นภาภาศ
 เกิดวันที่ 14 มกราคม 2520 ที่จังหวัดชัยภูมิ
 การศึกษา สำเร็จการศึกษาระดับปริญญาตรีวิศวกรรมศาสตรบัณฑิต (วศ.บ.) สาขาวิศวกรรมคอมพิวเตอร์มหาวิทยาลัยรังสิต ปีการศึกษา 2540
 การทำงาน สำนักวิจัยและบริการคอมพิวเตอร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ตำแหน่งนักวิชาการคอมพิวเตอร์ ระดับ 3 ในปี พ.ศ. 2542-2544



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้