

การประมาณพหุนามเชิงเส้นเสียงพูด
(LINEAR PREDICTION OF SPEECH)







สถาบันเทคโนโลยีพระจอมเกล้า เจ้าคุณทหาร-ลาดกระบัง
ปีการศึกษา ๒๕๒๕

แบบฟอร์มการให้คะแนนการสอบวิทยานิพนธ์
สำหรับนักศึกษาปริญญาโท

ชื่อนักศึกษา นายสุเชียร เกียรติสุนทร

ชื่อหัวข้อเรื่องวิทยานิพนธ์ "การประมาณค่าและเชิงเส้นleast square"

ชื่ออาจารย์ผู้สอบสัมภาษณ์	ลายเซ็น	ผลการสอบ
1. ผศ.วิวัฒน์ ปรีชาพานิช		ผ่าน
2. รศ.ดร.ไพรัช รัชพงษ์		ผ่าน
3. ดร.สมชาย จิตพันธ์กุล		ผ่าน
4. ดร.ประสิทธิ์ ประทีปมงคลการ		ผ่าน

วันที่, เดือน, ปี, ที่สอบ..... 10 มี.ย. ๕6 เวลา 11.00 น.

สารบัญ

	หน้า
กิตติกรรมประกาศ	iv
บทคัดย่อ	v
ABSTRACT	vi
บทที่ 1 บทนำ	1
บทที่ 2 ระบบการพูด	4
2.1 ระบบสรีระในการเปล่งเสียงพูด	4
2.2 แบบจำลองของระบบกำเนิดเสียงพูด	8
2.2.1 แหล่งกำเนิดสัญญาณกระตุ้น	8
2.2.2 ท่อกำธรเสียง	8
2.3 ระบบกำเนิดเสียงพูดแบบดิจิทัล	8
2.3.1 แหล่งกำเนิดสัญญาณกระตุ้น	8
2.3.2 วงจรกรองความถี่แบบดิจิทัล	8
2.3.3 พารามิเตอร์ของระบบกำเนิดเสียงพูด	9
2.4 สรุป	9
บทที่ 3 การประมาณพื่นอะเชิงเส้น	10
3.1 หลักการประมาณพื่นอะเชิงเส้น	10
3.2 วิธีในการประมาณพื่นอะเชิงเส้น	14
3.2.1 วิธีหาค่าความแปรปรวนร่วม	14
3.2.2 วิธีอัตสัมพันธ์	14
3.3 อัตราขยายของวงจรกรองความถี่	15
3.4 วิธีอัตสัมพันธ์ย่อย	16
3.5 ผลคุณภาพในของฟังก์ชันพหุคูณ	18
3.5.1 คุณสมบัติของผลคุณภาพใน	18
3.5.2 ความคลาดเคลื่อนกำลังสองรวม	20

3.6	ความสัมพันธ์เชิงพิกัดฉากของฟังก์ชันพหุนาม	20
3.7	สัมประสิทธิ์ PARCOR	21
3.7.1	ขั้นตอนการหาค่าสัมประสิทธิ์ PARCOR ด้วยวิธี หาค่าความแปรปรวนร่วม	25
3.7.2	ขั้นตอนการหาค่าสัมประสิทธิ์ PARCOR ด้วยวิธี อัตรสัมพันธ์	27
3.8	เสถียรภาพของวงจรรองความถี่แบบดิจิทัล	29
3.8.1	การตรวจสอบเสถียรภาพของวงจรรองความถี่	30
3.8.2	การปรับปรุงเสถียรภาพของวงจรรองความถี่	31
3.9	โครงสร้างของระบบกำเนิดเสียง	31
3.10	สรุป	34
บทที่ 4	ความถี่กำลังของเสียงพูด	35
4.1	การหาค่าความถี่กำลัง	35
4.1.1	การจัดเตรียมข้อมูล	35
4.1.2	การหาพารามิเตอร์ของวงจรรองความถี่	36
4.1.3	การหาข้อมูลดิบ	37
4.1.4	การตัดสินใจข้อมูล	42
4.2	สรุป	47
บทที่ 5	ช่วงกระตุ้นของเสียงพูด	48
5.1	การหาช่วงกระตุ้นของเสียงพูดด้วยวิธี SIFT	49
5.1.1	การสุ่มสัญญาณ	49
5.1.2	การหาสัญญาณกระตุ้นจากกล่องเสียง	52
5.1.3	การหาช่วงกระตุ้นของเสียงพูด	52
5.1.4	การตัดสินใจ	57
5.1.5	การปรับปรุงข้อมูล	57
5.2	สรุป	57

	หน้า
บทที่ 6 ระบบกำเนิดเสียงพูด	58
6.1 การวิเคราะห์สัญญาณเสียง	58
6.1.1 การหาพารามิเตอร์ของวงจรรองความถี่	59
6.1.2 การหาช่วงกระตุ้นของเสียงพูด	62
6.1.3 การเข้ารหัสพารามิเตอร์	62
6.2 การสังเคราะห์สัญญาณเสียงพูด	63
6.3 สรุป	66
บทที่ 7 ผลการทดลอง	67
7.1 ผลจากการวิเคราะห์สัญญาณเสียงพูด	67
7.1.1 พารามิเตอร์ของวงจรรองความถี่	73
7.1.2 การหาช่วงกระตุ้นของเสียงพูด	73
7.2 ผลจากการสังเคราะห์เสียงพูด	74
7.3 สรุป	75
บทที่ 8 บทสรุป	76
บทความและหนังสืออ้างอิง	79
ภาคผนวก	80

กิติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จขึ้นได้ด้วยคำแนะนำในการแก้ปัญหาอันมีค่าตลอดระยะเวลา
 ที่ผู้เขียนศึกษา ณ สถาบันแห่งนี้จาก อาจารย์ ดร.ไพรัช ธีชัยพงษ์ และอาจารย์วิพันธ์ ปรีชาพานิช
 นอกจากนี้ผู้เขียนยังได้รับความช่วยเหลือในการจัดหาอุปกรณ์ที่ใช้ประกอบการวิจัยจาก อาจารย์
 ดร.สุรเดช จันทรานุรักษ์ ซึ่งถ้าไม่ได้รับความช่วยเหลือดังกล่าวจากบุคคลใดบุคคลหนึ่งที่ได้กล่าว
 นามมาข้างต้นแล้ว วิทยานิพนธ์ฉบับนี้คงจะเสร็จสมบูรณ์ขึ้นไม่ได้ผู้เขียนจึงขอกราบขอบพระคุณบุคคล
 เหล่านี้มาไว้ ณ ที่นี้ด้วย



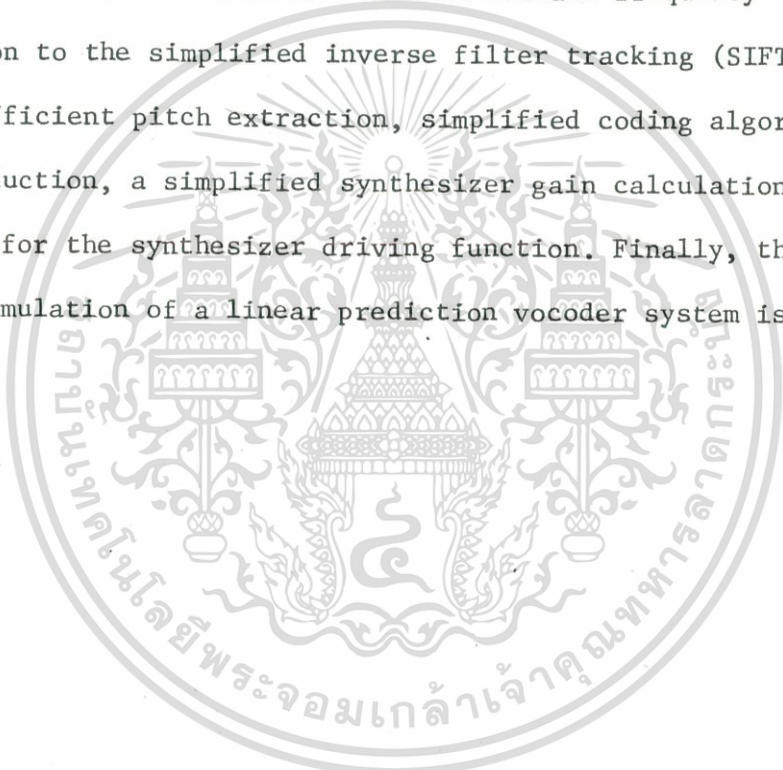
บทคัดย่อ

วิทยานิพนธ์ฉบับนี้จะกล่าวถึงการวิเคราะห์และสังเคราะห์เสียงพูดโดยการประมวล
 พันธะเชิงเส้นด้วยวิธีอัตโนมัติ ต่อไปจะกล่าวถึงการนำเทคนิคดังกล่าวมาใช้ในการหาค่าความ
 ที่กำกวมของเสียงพูด ช่วงกระตุ้นของเสียงพูดโดยใช้วิธี SIFT การเข้ารหัสพารามิเตอร์ของ
 เสียงพูด เพื่อลดจำนวนหน่วยความจำที่ใช้เก็บคุณลักษณะของสัญญาณดังกล่าว การหาอัตราขยาย
 ของระบบกำเนิดเสียงพูดและลักษณะของสัญญาณเข้าที่ใช้กระตุ้นระบบให้ทำงาน ในบทสุดท้ายของ
 วิทยานิพนธ์ฉบับนี้จะกล่าวถึงรายละเอียดการใช้ดิจิทัลคอมพิวเตอร์จำลองระบบกำเนิดเสียงพูด
 ของมนุษย์ ตั้งแต่การวิเคราะห์ถึงการสังเคราะห์เสียงพูดโดยใช้เทคนิคการประมวลพันธะเชิง
 เส้น เป็นรากฐานสำคัญ



ABSTRACT

This thesis describes a method of speed analysis and synthesis. The technique used for speech analysis is the autocorrelation method of linear prediction. The results include formant frequency estimation, modification to the simplified inverse filter tracking (SIFT) algorithm for more efficient pitch extraction, simplified coding algorithm for data storage reduction, a simplified synthesizer gain calculation and a bias correction for the synthesizer driving function. Finally, the digital computer simulation of a linear prediction vocoder system is presented in detail.



บทที่ 1

บทนำ

การประมาณพหุระเชิงเส้น (linear predictive) เป็นขบวนการทางคณิตศาสตร์ที่ใช้หาเอกลักษณ์ของระบบต่าง ๆ (systems identification) ซึ่งมีมานานแล้วแต่ยังไม่แพร่หลายเท่าที่ควร เนื่องจากในอดีตเหล่าวิศวกรและนักฟิสิกส์ยังขาดเครื่องคำนวณที่มีประสิทธิภาพเพียงพอ จนกระทั่งเทคโนโลยีด้านสารกึ่งตัวนำ (semiconductor) ได้รับการพัฒนาขึ้น จนสามารถสร้างเครื่องจักรที่ใช้ในการคำนวณที่มีประสิทธิภาพสูงที่สุดเท่าที่มนุษย์เคยรู้จักซึ่งเรียกกันว่า ดิจิตอลคอมพิวเตอร์ (digital computer) และจากการมีระบบอุปกรณ์ดังกล่าวก็ได้ทำให้มีวิชาการแขนงหนึ่งเกิดขึ้นเคียงคู่กันมาคือ การประมวลสัญญาณดิจิตอล (digital signal processing) ซึ่งมีการประมวลสัญญาณเสียงพูด (speech signal processing) เป็นส่วนหนึ่งของแขนงวิชาดังกล่าว การประมาณพหุระเชิงเส้น² เป็นวิธีการหนึ่งที่มีผู้นำมาใช้ในการทราพารามิเตอร์ (parameter) ของเสียงพูดตั้งแต่ปี ค.ศ. 1960 หรือเมื่อประมาณ 23 ปีมาแล้ว จนถึงปัจจุบันวิธีการดังกล่าวได้รับการพัฒนาขึ้นเรื่อย ๆ จนเหมาะสมกับการประมวลสัญญาณเสียงพูดในงานหลาย ๆ ด้าน นับตั้งแต่การวิเคราะห์เสียง (speech analysis) การสังเคราะห์เสียง (speech synthesis) การจำคำพูด (speech recognition) และการจำเสียงพูด (voice recognition)

วัตถุประสงค์ในการทำวิทยานิพนธ์ เรื่องการประมาณพหุระเชิงเส้นเสียงพูด มีสามประการ คือ

1. ทดลองหาแบบจำลองของระบบกำเนิดเสียงพูด (vocal system) พร้อมทั้งการวิเคราะห์หาค่าพารามิเตอร์ของระบบที่สร้างขึ้น โดยใช้เทคนิคการประมาณพหุระเชิงเส้น ในการวิเคราะห์หาค่าต่าง ๆ
2. การทดลองนำค่าพารามิเตอร์ต่าง ๆ ที่วิเคราะห์ได้ในขั้นตอนแรกกำเนิดเสียงพูด ขึ้นใหม่ เป็นการทดลองเลียนเสียงพูดของมนุษย์
3. ทดลองนำเสียงพูดที่เลียนแบบขึ้นในขั้นตอนที่สอง เปรียบเทียบกับสัญญาณเสียงพูดจริงจะทำให้สามารถตรวจสอบได้ว่า การวิเคราะห์หาค่าพารามิเตอร์ของเสียงพูด ด้วยการประมาณพหุระ

เชิงเส้น มีความน่าเชื่อถือมากน้อยเพียงไร สามารถนำไปใช้ประโยชน์สำหรับการค้นคว้าที่เกี่ยวข้องกันมากน้อยเพียงไร

เนื่องจากงานวิจัยที่ทำอยู่ในขณะนี้ เป็นขั้นตอนแรกในการค้นคว้าวิจัยที่เกี่ยวกับการประมวลสัญญาณเสียงคำพูดทำให้เนื้อหาส่วนใหญ่ในวิทยานิพนธ์ฉบับนี้ จะเป็นการทดลองนำทฤษฎีทางคณิตศาสตร์ในการประมาณพินธะเชิงเส้น เสียงคำพูดมาใช้ประโยชน์เกี่ยวกับการวิเคราะห์พารามิเตอร์ของระบบกำเนิดเสียงพูด และการทดสอบเกี่ยวกับคุณภาพของเสียงพูดเป็นส่วนใหญ่ โดยใช้ดิจิทัลคอมพิวเตอร์เป็นเครื่องมือประมวลผลต่าง ๆ สาเหตุวัตถุประสงค์และขั้นตอนการทำวิทยานิพนธ์ที่ได้กล่าวไปข้างต้น ทำให้สามารถแบ่งวิทยานิพนธ์ฉบับนี้ออกเป็น 8 บท คือ

บทที่ 2 จะกล่าวถึงระบบการพูดของมนุษย์^{5,8} ในด้านองค์ประกอบทางสรีระ และหน้าที่ของอวัยวะต่าง ๆ ของมนุษย์ในการเปล่งเสียงพูด เพื่อให้เกิดแนวความคิดอย่างกว้าง ๆ เกี่ยวกับแบบจำลองระบบกำเนิดเสียงพูด ซึ่งจะกล่าวถึงในตอนท้ายของบทนี้

บทที่ 3 จะกล่าวถึงทฤษฎีทางคณิตศาสตร์ที่ใช้ในการประมาณพินธะเชิงเส้น^{1,2,7} ที่เกี่ยวข้องในการประมวลเสียงพูดที่จะใช้ทำการค้นคว้าในบทต่อ ๆ ไป เพื่อให้ผู้ที่ทำการค้นคว้าวิจัยในงานแขนงนี้ต่อไปสามารถเข้าใจวิทยานิพนธ์ฉบับนี้อ้างอิงทางทฤษฎีอย่างคร่าว ๆ ได้ จึงทำให้รายละเอียดต่าง ๆ เกี่ยวกับวิทยานิพนธ์ฉบับนี้มีค่อนข้างมาก

บทที่ 4 บทนี้จะ เป็นบทแรกของงานวิจัยที่ทำโดยอาศัยทฤษฎีต่าง ๆ ที่กล่าวไว้ในบทที่ 2 และบทที่ 3 เป็นรากฐานโดยจะแสดงขั้นตอนการหาความถี่กำลังของเสียงพูด² ซึ่งใช้ในการทดสอบคุณภาพของเสียงพูด

บทที่ 5 บทนี้จะแสดงขั้นตอนการหาช่วงกระตุ้นของเสียงพูด^{2,4,6} ซึ่งเป็นพารามิเตอร์ที่สำคัญในการสังเคราะห์เสียงเพื่อเลียนเสียงพูดของมนุษย์

บทที่ 6 บทนี้เป็นส่วนสุดท้ายของงานวิจัยเป็นการแสดงขั้นตอนในการจำลองระบบกำเนิดเสียงพูด^{2,3,9} โดยใช้ดิจิทัลคอมพิวเตอร์ นับตั้งแต่การวิเคราะห์พารามิเตอร์ต่าง ๆ ของเสียงพูดจนถึงเลียนเสียงพูดเต็ม โดยใช้พารามิเตอร์ที่วิเคราะห์ได้โดยโปรแกรมต่าง ๆ ทำหน้าที่แทนทั้งหมด

บทที่ 7 บทนี้จะรวบรวมผลการทดลอง ขณะทำการวิจัยมาไว้ในบทเดียวกันทั้งหมด เพื่อให้สะดวกแก่ผู้ที่สนใจจะนำขั้นตอนการทดลองในบทที่ 4 ถึงบทที่ 6 ไปทดสอบในการตรวจสอบผล
ที่ได้รับนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่นอนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

บทที่ 8 บทนี้จะเป็นบทสรุปเกี่ยวกับผลการค้นคว้าและข้อเสนอแนะสำหรับผู้ทำการค้นคว้าวิจัยต่อไป

ในส่วนสุดท้ายของวิทยานิพนธ์ จะเป็นการรวบรวมโปรแกรม (program) ทั้งหมดที่เขียนขึ้นในขณะทำวิจัย เพื่อเป็นตัวอย่างสำหรับผู้ที่จะทำวิจัยต่อไป โดยแบ่งออกเป็นโปรแกรมน้อย (subroutine) ที่ใช้ในงานแต่ละส่วน โดยไม่ได้แสดงโปรแกรมหลัก (main program) เนื่องจากโปรแกรมหลักจะมีลักษณะอย่างไรขึ้นอยู่กับวัตถุประสงค์ของการวิจัย และลักษณะของเครื่องคอมพิวเตอร์ที่ใช้ และสาเหตุที่ไม่ได้นำโปรแกรมแสดงไว้ในบทต่าง ๆ นับตั้งแต่บทที่ 4 ถึงบทที่ 6 โดยตรง เนื่องจากต้องการให้บทดังกล่าวแสดงขั้นตอนการทดลองอย่างกว้างๆ พร้อมทั้งแทรกทฤษฎีที่ใช้ประกอบและไม่ได้กล่าวไว้ก่อน ในบทที่ 2 และบทที่ 3 เพื่อให้ผู้ที่จะทำการค้นคว้าต่อไป สามารถนำไปใช้ประโยชน์ให้เหมาะสมกับอุปกรณ์ที่ตนมีอยู่



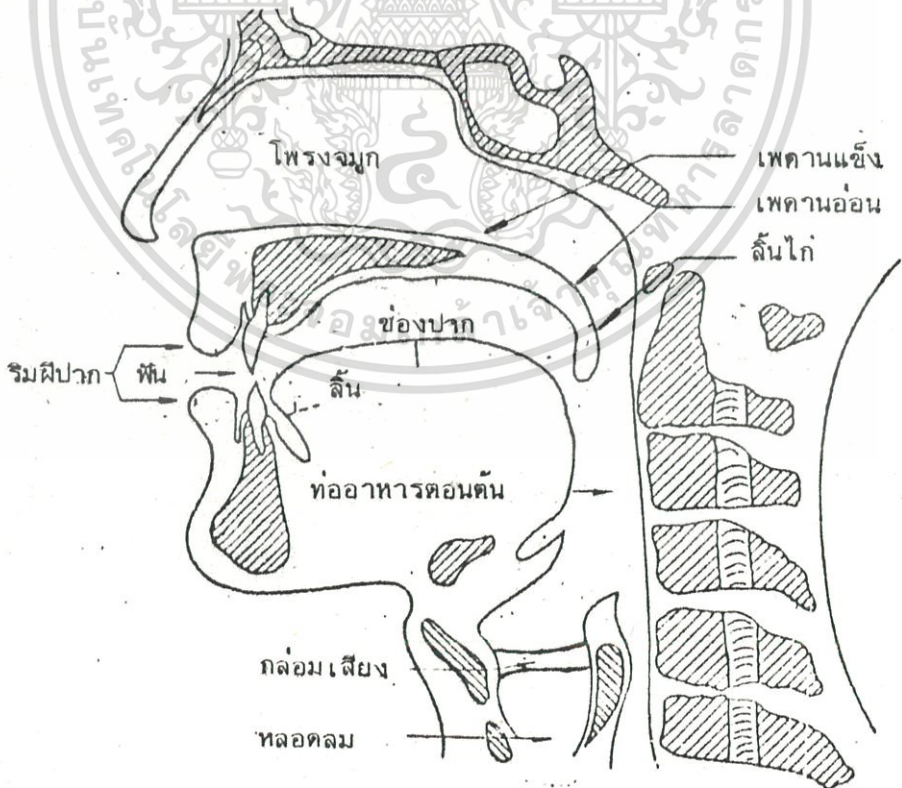
บทที่ 2

ระบบการพูด

ผลจากการพัฒนาวิชาการทางด้านเสียง (acoustic) และสรีระวิทยาทำให้ในปัจจุบันมนุษย์เราสามารถเข้าใจขั้นตอนต่าง ๆ ในการเปล่งเสียงพูดได้อย่างสมบูรณ์จนสามารถสร้างระบบกำเนิดเสียงได้สำเร็จ ในบทนี้จะกล่าวถึงระบบสรีระของมนุษย์ในการเปล่งเสียงพูด และโครงสร้างของแบบจำลองกำเนิดเสียงพูดที่มนุษย์สร้างขึ้น

2.1 ระบบสรีระในการเปล่งเสียงพูด

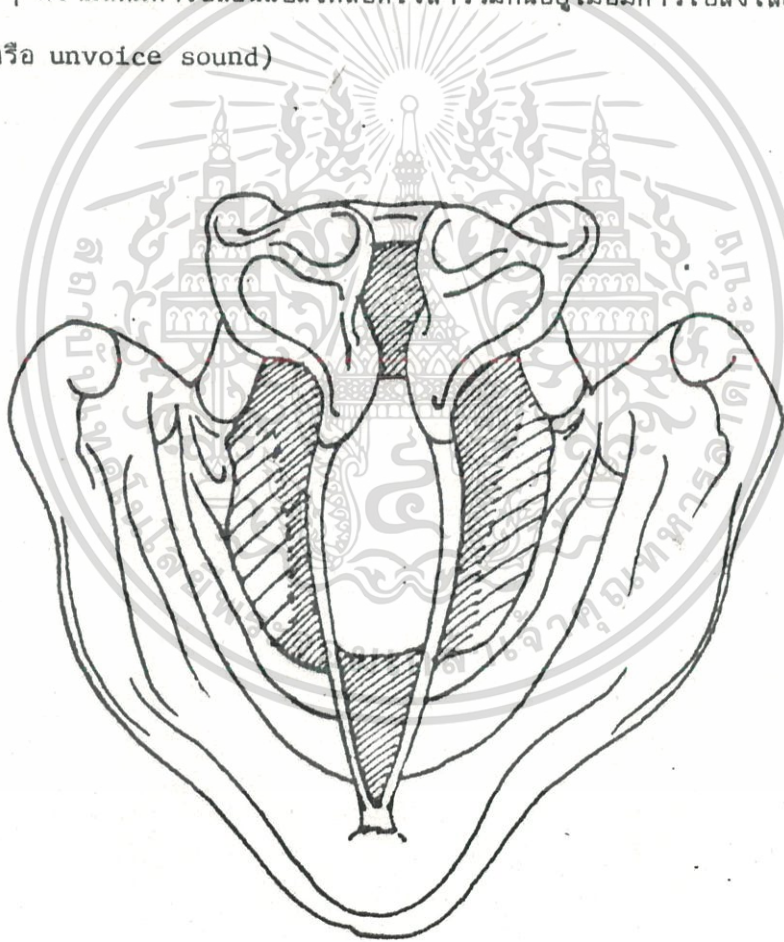
การเปล่งเสียงพูดของมนุษย์เรานั้น นอกจากปากและลิ้นตามที่พวกเราส่วนใหญ่เข้าใจกันแล้วยังมีอวัยวะอื่นอีกหลายอย่างที่ทำางานร่วมกันเพื่อให้การเปล่งเสียงพูดแต่ละครั้ง เป็นผลสำเร็จ รูปที่ 2.1 ได้แสดงระบบสรีระในการเปล่งเสียงพูดของมนุษย์ ซึ่งประกอบด้วย



รูปที่ 2.1 แสดงระบบการพูดของมนุษย์

1. ปอด เป็นอวัยวะที่ทำให้เกิดลวอากาศที่เป็นต้นกำเนิดของเสียง ขณะที่ปอดทำหน้าที่หายใจ

2. กล่องเสียง ทำหน้าที่บีบและกักอากาศจากปอดโดยการบังคับช่องว่างระหว่างเส้นเสียง (vocal cords) หรือช่องสายเสียง (glottis) ให้มีขนาดต่าง ๆ กัน ซึ่งจะทำให้กระแสอากาศที่ผ่านออกจากกล่องเสียงมีลักษณะเป็นสัญญาณพัลส์ (pulse) ที่มีความถี่ค่อนข้างแน่นอนขณะเปล่งเสียงโห่ (voice sound) และมีลักษณะเป็นสัญญาณนอยส์ (noise) ที่ประกอบด้วยสัญญาณความถี่สูงหลาย ๆ ความถี่ที่มีค่าเปลี่ยนแปลงตลอดเวลาารวมกันอยู่เมื่อมีการเปล่งเสียงอโห่ (voiceless หรือ unvoice sound)



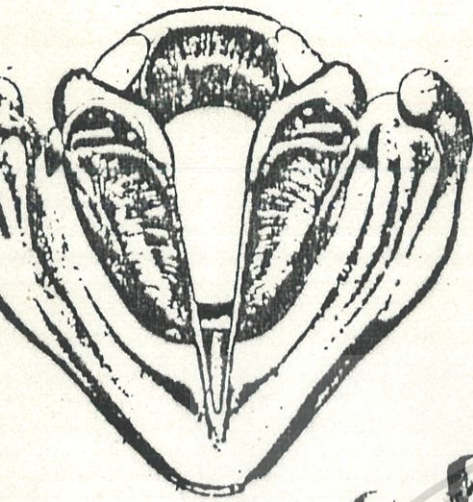
รูปที่ 2.2 แสดงลักษณะกล่องเสียง



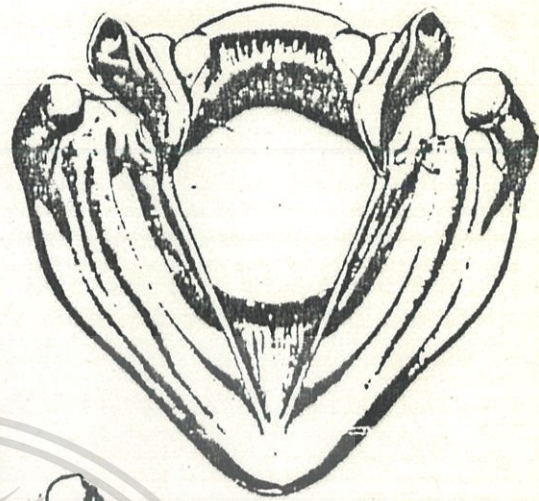
รูปที่ 2.3 แสดงที่ตั้งของกล่องเสียงภายในช่องปาก

3. ช่องปาก อวัยวะส่วนนี้เริ่มต้นจากท่ออาหารตอนบนบริเวณลิ้นไก่ไปสิ้นสุดที่บริเวณริมฝีปาก ทำหน้าที่เป็นท่อกำธรมเสียง (vocal tract) กับลักษณะความถี่ต่าง ๆ ที่รวมกันอยู่ภายในกระแสอากาศจากกล่องเสียง ซึ่งค่าความถี่กำธรม (resonance frequency หรือ formant frequency) ของท่อกำธรมเสียงจะขึ้นอยู่กัตำแหน่งของลิ้น ฟัน ขนาดและแรงตึงผิวของผนังช่องปาก

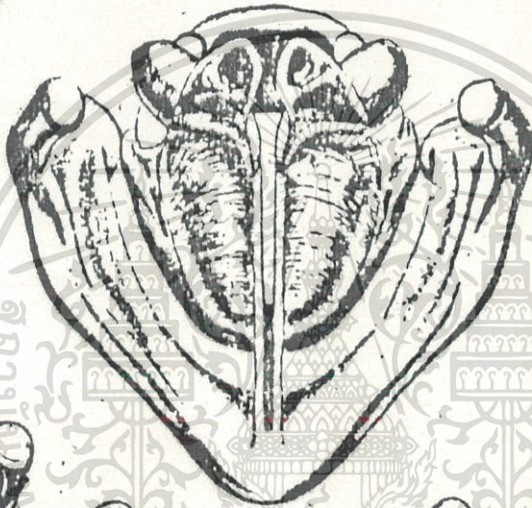
4. โทรงจมูก อวัยวะส่วนนี้เริ่มต้นจากเพดานอ่อน (velum) ไปสิ้นสุดที่รูจมูกทั้งสอง ทำหน้าที่เป็นส่วนหนึ่งของท่อกำธรมเสียงร่วมกับช่องปาก เมื่อมีการเปล่งเสียงนาสิก (nasal sound) ซึ่งเป็นเสียงอโฆษะประเภทหนึ่ง



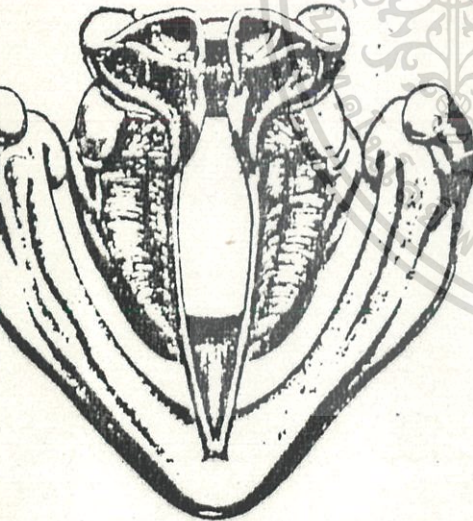
(ก)



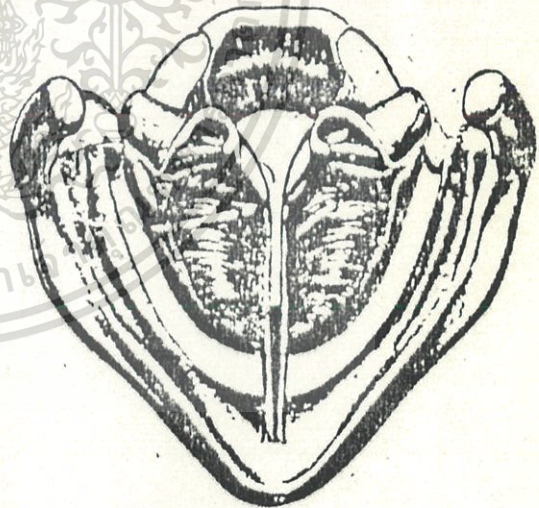
(ข)



(ค)



(ง)



(จ)

รูปที่ 2.4 แสดงลักษณะของกล่องเสียงขณะเปล่งเสียงชนิดต่าง ๆ

- ก) เสียงอโหหะ ข) เสียงตะโกน ค) เสียงโหหะ
ง) เสียงกระซิบ และ จ) เสียงนาสิก

2.2 แบบจำลองของระบบกำเนิดเสียงพูด

ในการประมวลสัญญาณเสียงพูดระบบกำเนิดเสียงพูดจะถูกแยกพิจารณาเป็น 2 ส่วนคือ

1. แหล่งกำเนิดสัญญาณกระตุ้น
2. ท่อกำธรเสียง

2.2.1 แหล่งกำเนิดสัญญาณกระตุ้น

แหล่งกำเนิดสัญญาณกระตุ้น หมายถึง เส้นเสียงและช่องสายเสียง ซึ่งสามารถแทนด้วยแหล่งกำเนิดสัญญาณพัลซซ์ที่มีช่วงกระตุ้น (pitch period) หรือคาบเวลาการเกิดสัญญาณพัลซซ์ ค่อนข้างแน่นอนเมื่อมีการเปล่งเสียงโห่หะ และทำหน้าที่สร้างสัญญาณนอยส์ เมื่อมีการเปล่งเสียงอโห่หะ

2.2.2 ท่อกำธรเสียง

ท่อกำธรเสียงใช้แทนการทำงานของช่องปาก และโพรงจมูก ซึ่งทำหน้าที่เป็นอุปกรณ์กรองความถี่ (filter) ที่ยอมให้ความถี่กำธรของสัญญาณเสียงผ่านได้เท่านั้น

2.3 ระบบกำเนิดเสียงพูดแบบดิจิทัล (digital vocoder)

การนำเทคนิคการประมวลสัญญาณดิจิทัลมาใช้กับเสียงพูดจำเป็นต้องกำหนดโครงสร้างของระบบดังกล่าวขึ้นก่อน เพื่อเป็นแนวทางในการจัดหาพารามิเตอร์ที่เหมาะสมให้กับแบบจำลองที่สร้างขึ้น รูปที่ 2.5 ได้แสดงโครงสร้างของระบบกำเนิดเสียงพูดแบบดิจิทัล ซึ่งประกอบด้วย

1. แหล่งกำเนิดสัญญาณกระตุ้น
2. วงจรกรองความถี่แบบดิจิทัล (digital filter)

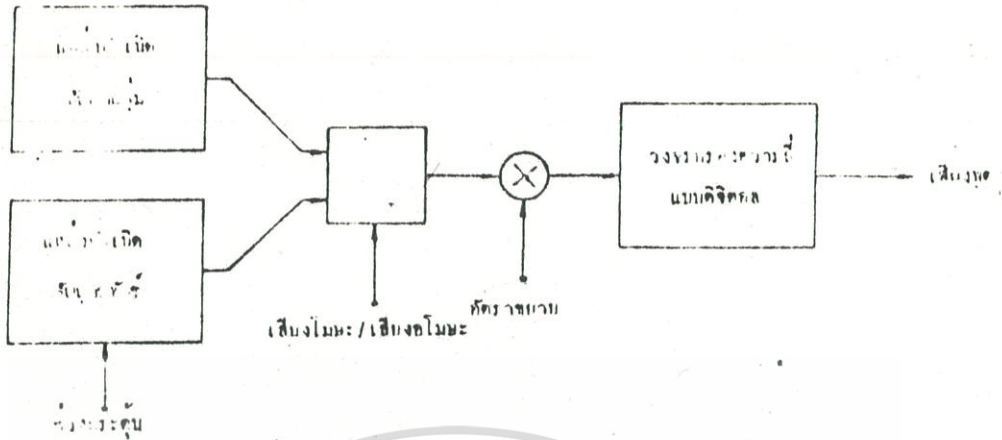
2.3.1 แหล่งกำเนิดสัญญาณกระตุ้น

ทำหน้าที่สร้างสัญญาณเพื่อกระตุ้นวงจรกรองความถี่แบบดิจิทัลให้ทำงาน ซึ่งสัญญาณที่สร้างขึ้นจะมีลักษณะ เป็นสัญญาณพัลซซ์ที่มีช่วงกระตุ้นคงที่ขณะกำเนิดเสียงโห่หะ และสร้างสัญญาณสุ่ม (random signals) ขณะกำเนิดเสียงอโห่หะ

2.3.2 วงจรกรองความถี่แบบดิจิทัล

ทำหน้าที่เป็นตัวกำธรตามค่าความถี่ธรรมชาติ (natural frequency) ของท่อกำธรเสียง การกำหนดคุณลักษณะ (characteristic) ของวงจรกรองความถี่อาจใช้ฟังก์ชันส่งผ่าน (transfer function) ที่มีค่าสัมประสิทธิ์เป็นพารามิเตอร์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า



รูปที่ 2.5 แสดงระบบกำเนิดเสียงพูดแบบดิจิทัล

2.3.3 พารามิเตอร์ของระบบกำเนิดเสียงพูด

การประมวลสัญญาณเสียงพูดแบบดิจิทัล จำเป็นต้องมีพารามิเตอร์ที่ใช้ในการกำหนดคุณลักษณะของเสียงพูด ซึ่งพารามิเตอร์เหล่านี้จะประกอบด้วย

1. ดัชนีบอกชนิดของเสียงพูดว่าเป็นโฆชะหรือเสียงอโฆชะ
2. ช่วงกระตุ้นสำหรับสัญญาณเสียงโฆชะ
3. อัตราขยาย (gain) ของวงจรกรองความถี่แบบดิจิทัลที่ใช้กำหนดความดังของเสียง
4. พารามิเตอร์ของวงจรกรองความถี่แบบดิจิทัล

2.4 สรุป

การประมวลสัญญาณเสียงพูดในการวิเคราะห์และสังเคราะห์เสียงพูด มักจะแบ่งระบบกำเนิดเสียงพูดออกเป็น 2 ส่วน เพื่อแยกกันพิจารณา คือแหล่งกำเนิดสัญญาณกระตุ้น ซึ่งใช้แทนการทำงานของปอด และกล่องเสียงส่วนหนึ่ง และท่อทางเสียงใช้แทนการทำงานของช่องปากและโพรงจมูกอีกส่วนหนึ่ง จากนั้นจึงทำการวิเคราะห์เพื่อหาค่าพารามิเตอร์ที่เหมาะสมให้กับโครงสร้างของระบบกำเนิดเสียง โดยแบ่งสัญญาณเสียงพูดที่จะวิเคราะห์ออกเป็นส่วน ๆ แต่ละส่วนใช้ระยะเวลาช่วงสั้น ๆ ประมาณ 15 ถึง 20 มิลลิวินาที (m-Sec) ซึ่งสัญญาณเสียงพูดจะมีการเปลี่ยนแปลงคุณลักษณะอย่างช้า ๆ จนอาจถือได้ว่าระบบกำเนิดเสียงมีคุณลักษณะไม่เปลี่ยนแปลง (stationary)

บทที่ 3

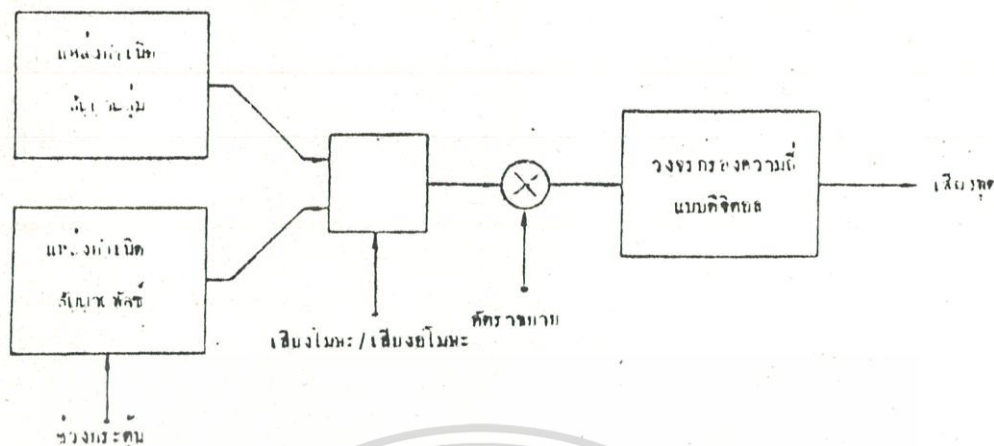
การประมาณพหุระเชิงเส้น

การประมาณพหุระเชิงเส้นเป็นขบวนการทางคณิตศาสตร์ที่ใช้ในการหาเอกลักษณ์ ของระบบโดยวิธีเลือกค่าพารามิเตอร์ที่เหมาะสม ซึ่งทำให้สามารถแสดงค่าสัญญาณในขณะเวลาใด ๆ ก็ตามที่ต้องการทราบ จากผลรวมเชิงเส้น (linear combination) ของสัญญาณที่ทราบค่าแล้ว โดยใช้วิธีกำลังสองน้อยที่สุด (method of least square) ในการเลือกค่าพารามิเตอร์ของระบบดังกล่าว ในบทนี้จะกล่าวถึงทฤษฎีทางคณิตศาสตร์ในการประมาณพหุระเชิงเส้นที่เกี่ยวข้องกับการประมวลสัญญาณเสียงพูด ซึ่งประกอบด้วย หลักการประมาณพหุระเชิงเส้นอย่างกว้าง ๆ เพื่อให้เกิดความเข้าใจพื้นฐานเกี่ยวกับเทคนิคดังกล่าวขึ้นก่อน จากนั้นจะกล่าวถึงวิธีการใหญ่ ๆ 2 วิธีที่ใช้ในการประมาณพหุระเชิงเส้นเสียงพูดคือ วิธีการหาค่าความแปรปรวนร่วม และวิธีอัลติมัทัมพร้อมทั้งเปรียบเทียบข้อดีข้อเสียของแต่ละวิธีการหาอัตราขยายของวงจรรองความถี่ที่ได้จากการประมาณพหุระเชิงเส้น เพื่อกำหนดความถี่ของเสียงพูด

หัวข้อต่อไปจะกล่าวถึงวิธีอัลติมัทัมพร้อมย่อ ซึ่งเป็นเทคนิคการหาค่าพารามิเตอร์ของวงจรรองความถี่ ที่จะใช้ต่อไปในภายหลัง พร้อมทั้งทฤษฎีทางคณิตศาสตร์ใช้ในเทคนิคดังกล่าวคือ คุณสมบัติของผลคูณภายใน และความสัมพัทธ์เชิงฟังก์ชันลากรางจ์ของฟังก์ชันพหุคูณ ต่อไปจะกล่าวถึงขั้นตอนการวิเคราะห์หาค่าสัมประสิทธิ์ PARCOR ซึ่งเป็นพารามิเตอร์ของวงจรรองความถี่โดยใช้วิธีอัลติมัทัมพร้อมย่อ และใน 2 หัวข้อสุดท้ายในบทนี้ จะกล่าวถึงการพิจารณาเสถียรภาพของระบบกำเนิดเสียงและการปรับปรุงเสถียรภาพดังกล่าว พร้อมทั้งโครงสร้างของระบบกำเนิดเสียงที่จะนำไปใช้ในบทที่ 6

3.1 หลักการประมาณพหุระเชิงเส้น

จากแบบจำลองของระบบกำเนิดเสียงที่ได้กล่าวไปแล้วในบทก่อนจะสามารถแสดง



รูปที่ 3.1 แสดงแบบจำลองของระบบกำเนิดเสียงพูดแบบดิจิทัล

ระบบได้ด้วยความสัมพันธ์

$$H(z) = \frac{S(z)}{U(z)} = \frac{G}{1 - \sum_{i=1}^m a_i z^{-i}} \quad (3.1)$$

หรือแสดงความสัมพันธ์ดังกล่าวในเชิงเวลา (time domain)

$$s(n) = \sum_{i=1}^m a_i s(n-i) + G u(n) \quad (3.2)$$

เมื่อ $\{a_i\}$ เป็นสัมประสิทธิ์ของวงจรกรองความถี่แบบดิจิทัลที่มีโพล (pole) เพียงอย่างเดียว G คืออัตราขยายของวงจรกรองความถี่ และ $u(n)$ เป็นสัญญาณกระตุ้นที่ถูกป้อนให้วงจรกรองความถี่ ซึ่งจะมีลักษณะเป็นพัลส์ที่มีคาบเวลาเท่ากับช่วงกระตุ้นของเสียงพูดในขณะกำเนิดเสียงโฆชะและเป็นสัญญาณสุ่มในขณะกำเนิดเสียงอโฆชะ

การหาเอกลักษณ์ของระบบจากสมการ (3.1) และ (3.2) ด้วยวิธีประมาณพันธะเชิงเส้นจะเริ่มด้วยการกำหนดระบบที่มีสัญญาณออก (output signal)

$$s(n) = \sum_{i=1}^m \alpha_i s(n-i) \quad (3.3)$$

เป็นค่าประมาณของสัญญาณจริง $s(n)$ ซึ่งทำให้เกิดความคลาดเคลื่อนในการประมาณค่า (prediction error)

$$e(n) = s(n) - \hat{s}(n) \quad (3.4)$$

$$= s(n) - \sum_{i=1}^m \alpha_i s(n-i) \quad (3.5)$$

จากสมการ (3.5) จะพบว่าความคลาดเคลื่อนในการกะประมาณค่าเป็นสัญญาณออก ของระบบที่มี ฟังก์ชันส่งผ่าน

$$A(z) = 1 - \sum_{i=1}^m \alpha_i z^{-i} \quad (3.6)$$

จากสมการ (3.2) และ (3.5) ถ้าการกะประมาณค่าเป็นจริงและ $\alpha_i = a_i$ แล้ว $e(n) = G u(n)$ และ

$$H(z) = \frac{G}{A(z)} \quad (3.7)$$

งานขั้นต่อไปในการประมาณพหุนามเชิงเส้น คือการเลือกเซต (set) ของสัมประสิทธิ์ $\{\alpha_i\}$ หรือ $\{a_i\}$ ที่จะทำให้ความคลาดเคลื่อนในการกะประมาณค่ามีค่าต่ำสุด จากสมการ (3.5)

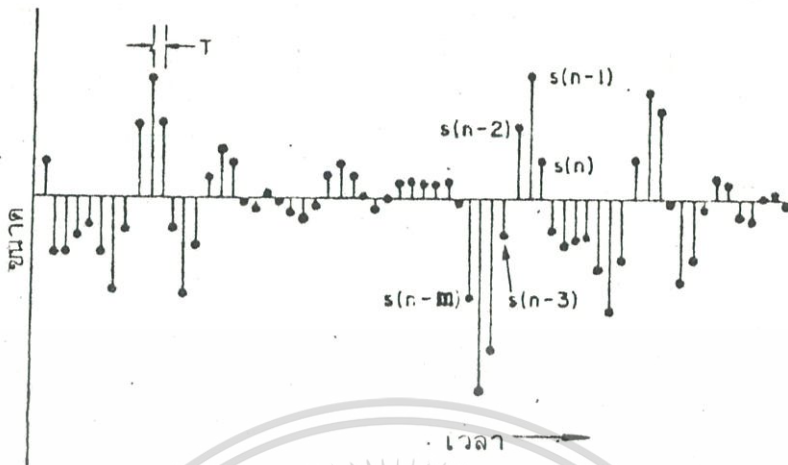
$$E = \sum_{n=n_0}^n e^2(n) \quad (3.8)$$

$$= \sum_{n=n_0}^n \left| \sum_{i=0}^m \alpha_i s(n-i) \right|^2 \text{ เมื่อ } \alpha_0 = 1 \quad (3.9)$$

โดยมีเงื่อนไขจำเป็น (necessary condition) และเงื่อนไขพอเพียง (sufficient condition) ที่ทำให้ความคลาดเคลื่อนรวม E มีค่าต่ำสุด คือ

$$\left. \begin{aligned} \frac{\partial E}{\partial \alpha_i} &= 0 \end{aligned} \right\} \quad (3.10)$$

$$\left. \begin{aligned} \frac{\partial^2 E}{\partial \alpha_i^2} &> 0 \end{aligned} \right\} \text{ เมื่อ } i = 1, 2, \dots, m \quad (3.11)$$



รูปที่ 3.2 แสดงการประมาณฟังก์ชันเชิงเส้น

จากสมการ (3.10)

$$\sum_{\substack{j=1 \\ j \neq i}}^m \alpha_j \sum_{n=n_0}^{n_1} s(n-1)s(n-j) = - \sum_{n=n_0}^{n_1} s(n)s(n-1) \quad (3.12)$$

ถ้ากำหนดให้

$$\phi(i, j) = \sum_{n=n_0}^{n_1} s(n-1)s(n-j) \quad (3.13)$$

จะทำให้สมการ (3.12) สามารถแสดงด้วยระบบสมการเชิงเส้น (set of linear equations)

$$\begin{aligned} \phi(1,1)\alpha_1 + \phi(1,2)\alpha_2 + \dots + \phi(1,m)\alpha_m &= -\phi(1,0) \\ \phi(2,1)\alpha_1 + \phi(2,2)\alpha_2 + \dots + \phi(2,m)\alpha_m &= -\phi(2,0) \\ \dots & \dots \dots \\ \phi(m,1)\alpha_1 + \phi(m,2)\alpha_2 + \dots + \phi(m,m)\alpha_m &= -\phi(m,0) \end{aligned} \quad (3.14)$$

สัมประสิทธิ์การกะประมาณค่า $\{\alpha_i\}$ สามารถหาได้จากคำตอบ (solution) ของสมการ (3.14)

3.2 ขอบเขตของสัญญาณที่ใช้ในการประมาณฟังก์ชันเชิงเส้น

จากสมการที่ (3.8) ถึง (3.9) และสมการ (3.12) ถึง (3.13) จะพบว่าความคลาดเคลื่อนในการประมาณค่าและสัมประสิทธิ์ของวงจรถองความถี่ขึ้นกับขอบเขตของสัญญาณ ภายใต้เครื่องหมายผลรวม (summation) โดยตรง ทำให้การประมาณฟังก์ชันเชิงเส้น ของสัญญาณสามารถแบ่งวิธีการที่ใช้ในการกำหนดขอบเขตสัญญาณที่ใช้ออกเป็น 2 กรณี คือ

1. วิธีหาค่าความแปรปรวนร่วม (covariance method)
2. วิธีอัตสัมพันธ์ (autocorrelation method)

3.2.1 วิธีหาค่าความแปรปรวนร่วม

การประมาณฟังก์ชันเชิงเส้นด้วยวิธีหาค่าความแปรปรวน จะกำหนดให้ $n_0 = m$ และ $n_1 = N-1$ เพื่อทำการลดความคลาดเคลื่อนในการประมาณค่าในช่วง $[m, N-1]$ ซึ่งทำให้สามารถแสดงขั้นตอนการคำนวณด้วยสมการต่อไปนี้

$$\sum_{i=1}^m a_i c_{ij} = -c_{oj} \quad \text{เมื่อ } j = 1, 2, \dots, m \quad (3.15)$$

$$c_{ij} = \sum_{n=m}^{N-1} s(n-i)s(n-j) \quad (3.16)$$

$$e(n) = \sum_{i=0}^m a_i s(n-i) \quad (a_0=1) \quad n=m, m+1, \dots, N-1 \quad (3.17)$$

3.2.2 วิธีอัตสัมพันธ์

การประมาณฟังก์ชันเชิงเส้นด้วยวิธีอัตสัมพันธ์จะกำหนดให้ $n_0 = -\infty$ และ $n_1 = +\infty$ โดย $s(n) = 0$ เมื่อ $n < 0$ และ $n > N$ เพื่อลดความคลาดเคลื่อนในการประมาณค่าในช่วง $[0, N+m-1]$ โดยมีขั้นตอนการคำนวณดังนี้

$$\sum_{i=1}^m a_i r(i-j) = -r(j) \quad \text{เมื่อ } j = 1, 2, \dots, m \quad (3.18)$$

$$r(1) = \sum_{n=0}^{N-1-1} s(n)s(n+1) \quad (1 \geq 0) \quad (3.19)$$

$$e(n) = \sum_{i=0}^m a_i s(n-i) \quad (a_0=1) \quad \text{เมื่อ } n=0,1,\dots,N+m-1 \quad (3.20)$$

3.3 อัตราขยายของวงจรรองความถี่

อัตราขยายของวงจรรองความถี่สามารถหาได้ โดยทำให้ระดับพลังงานโดยเฉลี่ยของข้อมูลดิบที่ใช้ในการประมาณค่ามีค่าเท่ากับผลที่ได้จากสมการ (3.2) และ (3.5)

$$G u(n) = s(n) - \sum_{i=1}^m a_i s(n-i) \quad (3.21)$$

$$e(n) = s(n) - \sum_{i=1}^m \alpha_i s(n-i) \quad (3.22)$$

ถ้าการประมาณค่าเป็นจริง $a_i = \alpha_i$ จะทำให้

$$e(n) = G u(n) \quad (3.23)$$

ถ้าทำให้ระดับพลังงานโดยเฉลี่ยทั้งสองด้านของสมการเท่ากัน จะได้

$$G^2 \sum_{n=n_0}^{n_1} u^2(n) = \sum_{n=n_0}^{n_1} e^2(n) \quad (3.24)$$

ซึ่ง $u(n) = \delta(n)$ ในกรณีที่ระบบกำเนิดเสียงอิสระและเป็นสัญญาณสุ่มที่มีค่ามัธยฐาน (mean) 0 และความแปรปรวน (variance) มีค่า 1 จากสมการ (3.8) จะได้

$$G^2 = \sum_{n=n_0}^{n_1} e^2(n) \quad (3.25)$$

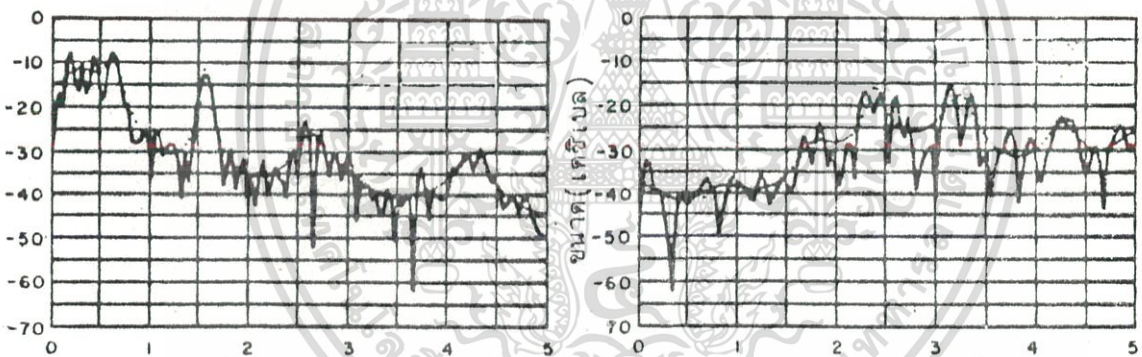
$$G^2 = E \quad (3.26)$$

$$G = \sqrt{E} \quad (3.27)$$

3.4 วิธีอัตโนมัติย่อย

วิธีอัตโนมัติย่อยหรือวิธี PARCOR (Partial Autocorrelation) เป็นวิธีการหาค่าพหุนามตัวหารในการประมาณค่าสัญญาณ ด้วยวิธีประมาณพหุคูณเชิงเส้นวิธีหนึ่ง โดยใช้คุณสมบัติของผลคูณภายใน (inner product) และความสัมพันธ์เชิงตั้งฉาก (orthogonal relationship) ของฟังก์ชันพหุคูณ (polynomial function) ในการเลือกค่าพหุนามตัวหาร ที่ทำให้เกิดความคลาดเคลื่อนต่ำสุด

กำหนดให้ $x_m^+(n)$ เป็นความคลาดเคลื่อนในการประมาณค่าแบบก้าวหน้า (forward prediction) ของสัญญาณ $\{x(n) | n_0 \leq n \leq n_1\}$ โดยใช้สัมประสิทธิ์ในการประมาณค่า m ตัว $\{a_{mi} | 1 \leq i \leq m\}$



ความถี่ (กิโลเฮิรตซ์)
(ก)

ความถี่ (กิโลเฮิรตซ์)
(ข)

รูปที่ 3.3 แสดงแถบพลังงาน (power spectrum) ของสัญญาณเสียง (เส้นทึบ) เปรียบเทียบกับแถบพลังงานของวงจรกรองความถี่ (เส้นประ) เมื่อ (ก) เสียงโฆษะ และ (ข) เสียงอโฆษะ

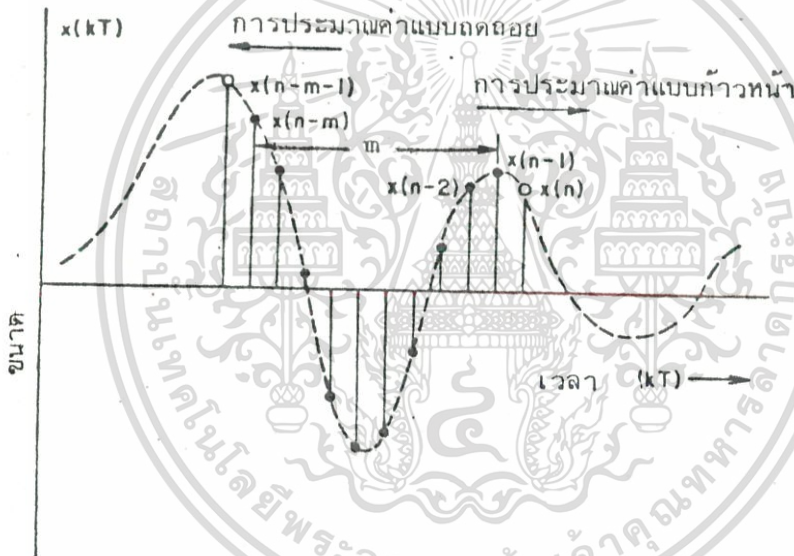
$$x_m^+(n) = x(n) - \left\{ - \sum_{i=1}^m a_{mi} x(n-i) \right\} \quad (3.28)$$

$$= \sum_{i=0}^m a_{mi} x(n-i) \quad \text{เมื่อ } a_{m0} = 1 \quad (3.29)$$

ในทำนองเดียวกันกับความคลาดเคลื่อนในการประมาณค่าแบบถดถอย (backward prediction) $x_m^-(n)$ ของสัญญาณ $\{x(n-m-1) | n_0 \leq n \leq n_1\}$ โดยใช้สัมประสิทธิ์ในการกะประมาณค่า m ตัว $\{b_{mi} | 1 \leq i \leq m\}$

$$x_m^-(n) = x(n-m-1) - \left\{ - \sum_{i=1}^m b_{mi} x(n-i) \right\} \quad (3.30)$$

$$= \sum_{i=1}^{m+1} b_{mi} x(n-i) \quad \text{เมื่อ} \quad b_{m, m+1} = 1 \quad (3.31)$$



รูปที่ 3.4

ค่าความคลาดเคลื่อนกำลังสองรวม (total square error) แบบก้าวหน้า α_m และถดถอย β_m เมื่อ $m = 1, 2, \dots, M$ ในช่วง $[n_0, n_1]$ มีค่า

$$\alpha_m = \sum_{n=n_0}^{n_1} \{x_m^+(n)\}^2 \quad (3.32)$$

$$\beta_m = \sum_{n=n_0}^{n_1} \{x_m^-(n)\}^2 \quad (3.33)$$

ขั้นตอนต่อไปในการประมาณพหุคูณเชิงเส้น คือการเลือกค่าสัมประสิทธิ์ในการประมาณค่า $\{a_{mi}\}$ และ $\{b_{mi}\}$ ที่ทำให้ค่า α_m และ β_m มีค่าต่ำสุด

3.5 ผลคูณภายในฟังก์ชันพหุคูณ

ถ้า $F(z)$ และ $G(z)$ เป็นฟังก์ชันพหุคูณที่มีเทอม (term) อนันต์

$$F(z) = \sum_{i=0}^{\infty} f_i z^{-i} \quad (3.34)$$

$$G(z) = \sum_{i=0}^{\infty} g_i z^{-i} \quad (3.35)$$

ถ้าฟังก์ชันพหุคูณทั้งสอง เป็นฟังก์ชันส่งผ่านของวงจรกรองความถี่แบบดิจิทัล... เมื่อป้อนสัญญาณเข้า (input signal) $\{x(n)\}$ จะได้สัญญาณออก $\{u(n)\}$ และ $\{v(n)\}$ แล้ว ผลคูณภายในของฟังก์ชันพหุคูณทั้งสอง จะมีค่า

$$\langle F(z), G(z) \rangle = \sum_{n=n_0}^{n_1} \{u(n)\} \{v(n)\} \quad (3.36)$$

$$= \sum_{n=n_0}^{n_1} \left\{ \sum_{i=0}^{\infty} f_i x(n-i) \right\} \left\{ \sum_{j=0}^{\infty} g_j x(n-j) \right\} \quad (3.37)$$

$$= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} f_i \left\{ \sum_{n=n_0}^{n_1} x(n-i)x(n-j) \right\} g_j \quad (3.38)$$

จากสมการ (3.38) จะพบว่าค่าผลคูณภายในของฟังก์ชันพหุคูณจะขึ้นกับสัมประสิทธิ์ และสัญญาณเข้าของวงจรกรองความถี่ทั้งสองเท่านั้น

3.5.1 คุณสมบัติของผลคูณภายใน

ผลคูณภายในของฟังก์ชันพหุคูณมีคุณสมบัติที่สำคัญ ดังนี้

1. คุณสมบัติเชิงสมมาตร (symmetry property) จาก

$$\langle F(z), G(z) \rangle = \sum_{i=0}^{\alpha} \sum_{j=0}^{\alpha} f_i \left\{ \sum_{n=n_0}^{n_1} x(n-i)x(n-j) \right\} g_j \quad (3.39)$$

$$= \sum_{j=0}^{\alpha} \sum_{i=0}^{\alpha} g_j \left\{ \sum_{n=n_0}^{n_1} x(n-j)x(n-i) \right\} f_i \quad (3.40)$$

$$\langle F(z), G(z) \rangle = \langle G(z), F(z) \rangle \quad (3.41)$$

2. คุณสมบัติเชิงเส้น (linear property) ถ้า $F(z)$, $G(z)$ และ $H(z)$ เป็นฟังก์ชันพหุคูณ และ c_1, c_2 เป็นค่าคงที่จากสมการ (3.37)

$$\begin{aligned} \langle F(z), c_1 G(z) + c_2 H(z) \rangle &= \sum_{n=n_0}^{n_1} \left\{ \sum_{i=0}^{\alpha} f_i x(n-i) \right\} \left\{ c_1 \sum_{j=0}^{\alpha} g_j x(n-j) \right. \\ &\quad \left. + c_2 \sum_{k=0}^{\alpha} h_k x(n-k) \right\} \end{aligned} \quad (3.42)$$

$$\begin{aligned} &= c_1 \sum_{n=n_0}^{n_1} \left\{ \sum_{i=0}^{\alpha} f_i x(n-i) \right\} \left\{ \sum_{j=0}^{\alpha} g_j x(n-j) \right\} \\ &\quad + c_2 \sum_{n=n_0}^{n_1} \left\{ \sum_{i=0}^{\alpha} f_i x(n-i) \right\} \left\{ \sum_{k=0}^{\alpha} h_k x(n-k) \right\} \end{aligned} \quad (3.43)$$

$$\langle F(z), c_1 G(z) + c_2 H(z) \rangle = c_1 \langle F(z), G(z) \rangle + c_2 \langle F(z), H(z) \rangle \quad (3.44)$$

3. ค่าสัมบูรณ์ของฟังก์ชันพหุคูณ (opolynomial norm) ถ้า $F(z)$ เป็นฟังก์ชันพหุคูณ ค่าสัมบูรณ์ของฟังก์ชันพหุคูณจะมีค่า

$$\|F(z)\| = \sqrt{\langle F(z), F(z) \rangle} \quad (3.45)$$

จากสมการ (3.37) และคุณสมบัติเชิงเส้นของผลคูณภายใน

$$\|F(z)\| \geq 0 \quad (3.46)$$

ถ้า $\|F(z)\| = 0$

$$\|F(z) + G(z)\| = \|G(z)\| \quad (3.47)$$

$$\begin{aligned} \text{ถ้า } \|F(z)\| = 0 \text{ หรือ } \|G(z)\| = 0 \\ \|F(z), G(z)\| = 0 \end{aligned} \quad (3.48)$$

3.5.2 ความคลาดเคลื่อนกำลังสองรวม

ค่าความคลาดเคลื่อนกำลังสองรวมสามารถแสดงในรูปค่าสัมบูรณ์ของฟังก์ชันพหุนาม

$$\alpha_m = \sum_{n=n_0}^{n_1} \{x_m^+(n)\}^2 \quad (3.49)$$

$$\alpha_m = \sum_{n=n_0}^{n_1} \left\{ \sum_{i=0}^m a_i x(n-i) \right\} \left\{ \sum_{j=0}^m a_j x(n-j) \right\} \quad (3.50)$$

$$= \langle A_m(z), A_m(z) \rangle \quad (3.51)$$

$$= \|A_m(z)\|^2 \quad (3.52)$$

เมื่อ

$$A_m(z) = \sum_{i=0}^m a_m z^{-i} \text{ และ } a_{m0} = 1 \quad (3.53)$$

ในทำนองเดียวกัน

$$\beta_m = \|B_m(z)\|^2 \quad (3.54)$$

และ

$$B_m(z) = \sum_{i=1}^{m+1} b_{mi} z^{-i} \text{ และ } b_{m, m+1} = 1 \quad (3.55)$$

3.6 ความสัมพันธ์เชิงพิกัดฉากของฟังก์ชันพหุนาม

เซตของฟังก์ชันพหุนาม $\{P_i(z) \mid 0 \leq i \leq m\}$ จะเป็นพหุนามพิกัดฉาก (orthogonal polynomial) ถ้า

$$\langle P_i(z), P_j(z) \rangle \begin{cases} \neq 0 & \text{เมื่อ } i = j \\ = 0 & \text{เมื่อ } i \neq j \end{cases} \quad (3.56)$$

ถ้าฟังก์ชันพหุนาม $A_m(z), m = 1, 2, \dots, M$ ทำให้เกิดความคลาดเคลื่อนกำลังสองรวมต่ำสุด

$$\|A_m(z) + cz^{-i}\|^2 \geq \|A_m(z)\|^2 \quad (3.57)$$

เมื่อ $i = 1, 2, \dots, m$ และ c เป็นค่าคงที่ใด ๆ ที่ไม่เป็นศูนย์

$$\|A_m(z)\|^2 + 2c\langle A_m(z), z^{-i} \rangle + c^2\langle z^{-i}, z^{-i} \rangle \geq \|A_m(z)\|^2 \quad (3.58)$$

$$2c\langle A_m(z), z^{-i} \rangle + c^2\langle z^{-i}, z^{-i} \rangle \geq 0 \quad (3.59)$$

เงื่อนไขที่สอดคล้องกับสมการ (3.59) คือ

$$\langle A_m(z), z^{-i} \rangle = 0 \quad (3.60)$$

ในทำนองเดียวกัน

$$\langle B_m(z), z^{-i} \rangle = 0 \quad (3.61)$$

จากสองสมการสุดท้ายจะพบว่าฟังก์ชันพหุนาม $A_m(z)$ และ $B_m(z)$ เมื่อ $m = 1, 2, \dots, M$ มีความสัมพันธ์เชิงพิกัดฉากกับ z^{-i} เมื่อ $i = 1, 2, \dots, m$

3.7 สมประสิทธิ์ PARCOR

การเลือกค่าพารามิเตอร์ของระบบกำเนิดเสียงโดยวิธี PARCOR คือการหาค่าของสัมประสิทธิ์ $\{k_m | 1 \leq m \leq M\}$ ที่ทำให้สามารถแสดงฟังก์ชันพหุนาม $A_m(z)$ ในรูปผลรวมเชิงเส้นของ $A_{m-1}(z)$ และ $B_{m-1}(z)$ โดยเริ่มจาก

$$A_0(z) = 1 \tag{3.62}$$

และ

$$B_0(z) = z^{-1} \tag{3.63}$$

และสำหรับค่า m ใด ๆ ที่มากกว่าศูนย์

$$A_m(z) = A_{m-1}(z) + k_m B_{m-1}(z) \tag{3.64}$$

$$\langle A_m(z), z^{-m} \rangle = \langle A_{m-1}(z), z^{-m} \rangle + k_m \langle B_{m-1}(z), z^{-m} \rangle \tag{3.65}$$

จากคุณสมบัติเชิงฟังก์ชันของฟังก์ชันพหุนาม

$$\langle A_m(z), z^{-m} \rangle = 0 \tag{3.66}$$

$$\langle A_{m-1}(z), z^{-m} \rangle = \langle A_{m-1}(z), B_{m-1}(z) \rangle \tag{3.67}$$

$$= \langle 1, B_{m-1}(z) \rangle \tag{3.68}$$

และ

$$\langle B_{m-1}(z), z^{-m} \rangle = \langle B_{m-1}(z), B_{m-1}(z) \rangle \tag{3.69}$$

$$= \|B_{m-1}(z)\|^2 \tag{3.70}$$

$$= \beta_{m-1} \tag{3.71}$$

แทนค่าลงในสมการ (3.65)

$$k_m = \frac{-1}{\beta_{m-1}} \langle A_{m-1}(z), B_{m-1}(z) \rangle \tag{3.72}$$

$$= \frac{-1}{\beta_{m-1}} \left\{ \sum_{n=n_0}^{n_1} \sum_{i=0}^{m-1} a_{mi} x^{(n-i)} \right\}$$

$$= \left\{ \sum_{n=n_0}^{n_1} \sum_{j=0}^m b_{mj} x^{(n-j)} \right\} \tag{3.73}$$

$$= \frac{-1}{\beta_{m-1}} \left\{ \sum_{n=n_0}^{n_1} x_m^+(n) \right\} \left\{ \sum_{n=n_0}^{n_1} x_m^-(n) \right\} \quad (3.74)$$

จาก

$$A_m(z) = \sum_{i=0}^m a_{mi} z^{-i} \quad (3.75)$$

$$z^{-(m+1)} A_m(z^{-1}) = z^{-(m+1)} \sum_{i=0}^m a_{mi} z^i \quad (3.76)$$

$$= \sum_{i=0}^m a_{mi} z^{i-m-1} \quad (3.77)$$

ให้ $j = -(i-m-1)$

$$z^{-(m+1)} A_m(z^{-1}) = \sum_{j=1}^{m+1} a_{m, m+1-j} z^{-j} \quad (3.78)$$

เมื่อ

$$b_{mj} = a_{m, m+1-j} \quad (3.79)$$

$$B_m(z) = z^{-(m+1)} A_m(z^{-1}) \quad (3.80)$$

และ

$$A_m(z) = z^{-(m+1)} B_m(z^{-1}) \quad (3.81)$$

จากสมการ (3.64)

$$z^{-(m+1)} A_m(z^{-1}) = z^{-1} \{ z^{-m} A_{m-1}(z^{-1}) + k_m z^{-m} B_{m-1}(z^{-1}) \} \quad (3.82)$$

$$B_m(z) = z^{-1} \{ k_m A_{m-1}(z) + B_{m-1}(z) \} \quad (3.83)$$

จากสมการ (3.29) และ (3.31)

$$X_m^+(z) = A_m(z) X(z) \quad (3.84)$$

$$X_m^-(z) = B_m(z) X(z) \quad (3.85)$$

แทนค่าทั้งสองลงในสมการ (3.64) และ (3.83)

$$X_m^+(z) = X_{m-1}^+(z) + k_m X_{m-1}^-(z) \quad (3.86)$$

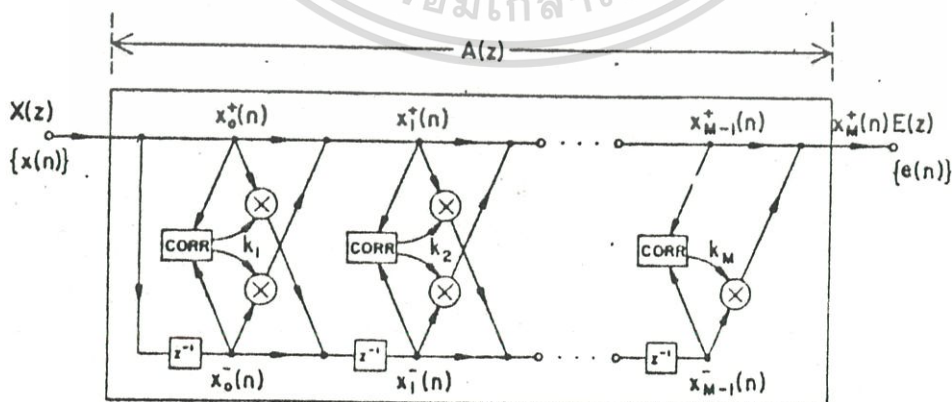
$$X_m^-(z) = z^{-1} \{k_m X_{m-1}^+(z) + X_{m-1}^-(z)\} \quad (3.87)$$

หรือแสดงความสัมพันธ์ในเชิงเวลา

$$x_m^+(z) = x_{m-1}^+(z) + k_m x_{m-1}^-(z) \quad (3.88)$$

$$x_m^-(z) = k_m x_{m-1}^+(z) + x_{m-1}^-(z) \quad (3.89)$$

จากความสัมพันธ์ทั้งหมดที่กล่าวมาทำให้สามารถหาค่าพารามิเตอร์ของการวิเคราะห์สัญญาณ ซึ่งต่อไปจะเรียกพารามิเตอร์ที่ได้ด้วยวิธีว่าสัมประสิทธิ์ PARCOR หรือสัมประสิทธิ์การสะท้อน (reflection coefficient)



รูปที่ 3.5 แสดงแผนผังการวิเคราะห์ที่เสียงพูดด้วยวิธีประมาณพหุคูณเชิงเส้น

จากสมการ (3.64) เมื่อ $A_0(z) = 1$

$$A_m(z) = 1 + \sum_{i=1}^m k_i B_{i-1}(z) \tag{3.90}$$

$$\|A_m(z)-1\|^2 = \sum_{i=1}^m k_i^2 \beta_{i-1} \tag{3.91}$$

และ

$$\|A_m(z)-1\|^2 = \|A_m(z)\|^2 - 2\langle A_m(z), 1 \rangle + \|1\|^2 \tag{3.92}$$

$$= \alpha_m - 2\alpha_m + \|1\|^2 \tag{3.93}$$

$$= \|1\|^2 - \alpha_m \tag{3.94}$$

แทนค่าดังกล่าวลงในสมการ (3.91)

$$\alpha_m = \|1\|^2 - \sum_{i=1}^m k_i^2 \beta_{i-1} \tag{3.95}$$

และ

$$\alpha_{m+1} = \|1\|^2 - \sum_{i=1}^{m+1} k_i^2 \beta_{i-1} \tag{3.96}$$

จากผลต่างของสมการทั้งสอง

$$\alpha_{m+1} = \alpha_m - k_{m+1}^2 \beta_m \tag{3.97}$$

3.7.1 ขั้นตอนการหาค่าสัมประสิทธิ์ PARCOR ด้วยวิธีหาค่าความแปรปรวนร่วม เมื่อ $m = 0$ จะกำหนดค่าเริ่มต้น

$$A_0(z) = 1 \text{ หรือ } a_{00} = 1 \tag{3.98}$$

$$B_0(z) = z^{-1} \text{ หรือ } b_{01} = 1 \tag{3.99}$$

$$\alpha_0 = \langle A_0(z), A_0(z) \rangle = c_{00} \quad (3.100)$$

$$\beta_0 = \langle B_0(z), B_0(z) \rangle = c_{11} \quad (3.101)$$

เมื่อ $m = 1$

$$k_1 = \frac{-1}{\beta_0} \langle A_0(z), B_0(z) \rangle = \frac{-c_{10}}{\beta_0} \quad (3.102)$$

$$A_1(z) = A_0(z) + k_1 B_0(z) = 1 + k_1 z^{-1} \quad (3.103)$$

$$B_1(z) = z^{-1} \{k_1 A_0(z) + B_0(z)\} = k_1 z^{-1} + z^{-2} \quad (3.104)$$

$$\alpha_1 = \alpha_0 + k_1^2 \beta_0 \quad (3.105)$$

$$\beta_1 = \langle B_1(z), z^{-2} \rangle = \sum_{i=1}^2 c_{2i} b_{1i} \quad (3.106)$$

เมื่อ $m = 2, 3, \dots, M$

$$k_m = \frac{-1}{\alpha_{m-1}} \sum_{i=0}^{m-1} c_{mi} a_{m-1 i} \quad (3.107)$$

$$A_m(z) = A_{m-1}(z) + k_m B_{m-1}(z) \quad (3.108)$$

$$a_{m0} = 1 \quad (3.109)$$

$$a_{mi} = a_{m-1 i} + k_m b_{m-1 i} \quad \text{เมื่อ } i = 1, 2, \dots, m-1 \quad (3.110)$$

$$a_{mm} = k_m \quad (3.111)$$

$$b_{mi} = a_{m m+1-i} \quad \text{เมื่อ } i = 1, 2, \dots, m+1 \quad (3.112)$$

$$\alpha_m = \alpha_{m-1} - k_m^2 \beta_{m-1} \quad (3.113)$$

$$\beta_m = \sum_{i=1}^{m+1} c_{m+1, i} b_{m, i} \quad (3.114)$$

3.7.2 ขั้นตอนการหาค่าสัมประสิทธิ์ PARCOR ด้วยวิธีอัลติมพันธ์ เมื่อ $m = 0$ จะกำหนดค่าเริ่มต้น

$$a_{00} = 1 \quad (3.115)$$

$$b_{01} = 1 \quad (3.116)$$

$$\alpha_0 = r_0 \quad (3.117)$$

$$\beta_0 = r_0 \quad (3.118)$$

เมื่อ $m = 1$

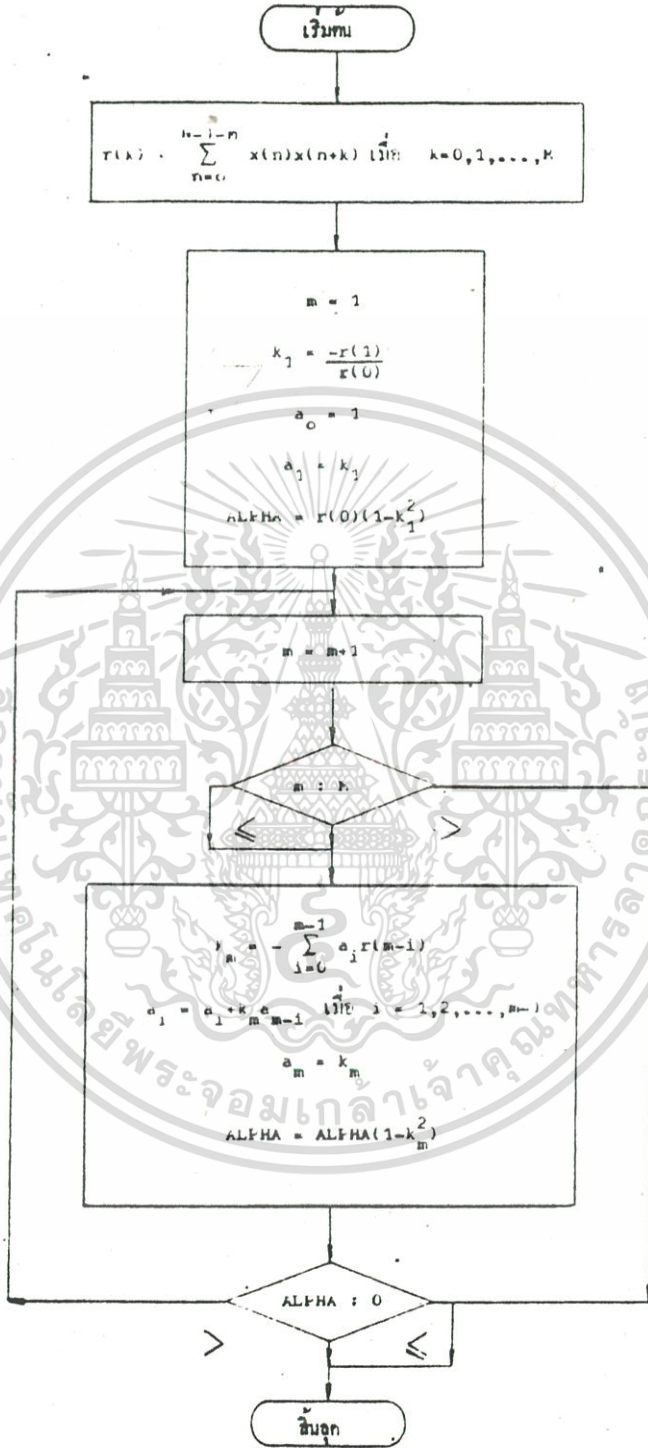
$$k_1 = \frac{-r(1)}{\beta_0} \quad (3.119)$$

$$A_1(z) = 1 + k_1 z^{-1} \quad (3.120)$$

$$B_1(z) = k_1 z^{-1} + z^{-2} \quad (3.121)$$

$$\alpha_1 = \alpha_0 + k_1^2 \beta_0 \quad (3.122)$$

$$\beta_1 = \alpha_1 \quad (3.123)$$



รูปที่ 3.6 แสดงขั้นตอนการประมาณฟังก์ชันเชิงเส้นด้วยวิธีฮัดสันสัน

เมื่อ $m = 2, 3, \dots, M$

$$k_m = \frac{-1}{\alpha_{m-1}} \sum_{i=0}^{m-1} r^{(m-i)} a_{m-1 i} \quad (3.124)$$

$$a_{m0} = 1 \quad (3.125)$$

$$a_{mi} = a_{m-1 i} + k_m b_{m-1 i} \quad \text{เมื่อ } i = 1, 2, \dots, m-1 \quad (3.126)$$

$$a_{mm} = k_m \quad (3.127)$$

$$b_{mi} = a_{m m+1-i} \quad \text{เมื่อ } i = 1, 2, \dots, m+1 \quad (3.128)$$

$$\alpha_m = \alpha_{m-1} (1+k_m^2) \quad (3.129)$$

$$\beta_m = \alpha_m \quad (3.130)$$

การคำนวณจะยุติเมื่อ $m = M$ หรือ α_m มีค่าเป็นลบ รูปที่ 3.6 ได้แสดงขั้นตอนการคำนวณเมื่อใช้วิธีตัดสัมพันธ์ส่วนการคำนวณโดยการหาค่าความแปรปรวนร่วม ซึ่งคล้ายคลึงกันจึงไม่ได้นำมาแสดงไว้

3.8 เสถียรภาพของวงจกรองความถี่แบบดิจิทัล

ฟังก์ชันส่งผ่านของวงจกรองความถี่ $1/A_m(z)$ จะมีเสถียรภาพ ถ้ารากของฟังก์ชันพหุนาม $A_m(z)$ ทุกตัวมีขนาดน้อยกว่า 1 หรือ

$$A_m(z) = \sum_{i=0}^m a_{mi} z^{-i} \quad (3.131)$$

$$A_m(z_{mi}) = 0 \quad \text{เมื่อ } i = 1, 2, \dots, m \quad (3.132)$$

$$A_m(z) = \prod_{i=1}^m (1 - z_{mi} z^{-1}) \tag{3.133}$$

วงจรรองความถี่จะมีเสถียรภาพ เมื่อ

$$|z_{mi}| < 1 \quad \text{เมื่อ} \quad i = 1, 2, \dots, m \tag{3.134}$$

การตรวจสอบเสถียรภาพของวงจรรองความถี่ด้วยวิธีดังกล่าวถึงแม้ว่าจะให้ผลที่แน่นอน แต่ต้องใช้เวลาอย่างมากในทางปฏิบัติ จึงนิยมใช้วิธีอื่นซึ่งสะดวกกว่า

3.8.1 การตรวจสอบเสถียรภาพของวงจรรองความถี่

การตรวจสอบเสถียรภาพของวงจรรองความถี่แบบง่าย ๆ สามารถทำได้โดยอาศัย

สมการ

$$A_m(z) = A_{m-1}(z) + k_m B_{m-1}(z) \tag{3.135}$$

$$A_m(z) = \sum_{i=0}^{m-1} a_{m-1-i} z^{-i} + k_m \sum_{i=0}^{m-1} a_{m-1-i} z^{-(m-i)} \tag{3.136}$$

จากสมการ (3.133)

$$k_m = \prod_{i=1}^m (-z_{mi}) \tag{3.137}$$

ถ้าฟังก์ชันพหุคูณ $A_m(z)$ ทำให้วงจรรองความถี่มีเสถียรภาพ หรือค่ารากทุกค่าของ $A_m(z)$ มีขนาดน้อยกว่า 1 จากสมการ (3.137) จะพบว่า

$$|k_m| < 1 \tag{3.138}$$

สมการ (3.138) เป็นเงื่อนไขจำเป็นสำหรับการรักษาเสถียรภาพของวงจรรองความถี่ $1/A_m(z)$ แต่ยังไม่เพียงพอที่จะใช้สมการดังกล่าวในการตรวจสอบเสถียรภาพของวงจรรองความถี่ ในสมการ (3.135) ถ้า m มีค่า 1

$$A_1(z) = 1 + k_1 z^{-1} \tag{3.139}$$

และเนื่องจาก $A_1(z)$ มีรากเพียงรากเดียว ซึ่งถ้า k_1 มีขนาดน้อยกว่า 1 แล้วหมายความว่า รากของ $A_1(z)$ จะมีขนาดน้อยกว่า 1 ด้วย และทำให้วงจรรองความถี่สามารถรักษาเสถียรไว้ได้ซึ่งจากผลดังกล่าวสามารถพิสูจน์ต่อไปได้ว่าสำหรับวงจรรองความถี่อันดับที่ M ใด ๆ ที่เกิดจากฟังก์ชันพหุนาม $A_M(z)$ ถ้า

$$|k_m| < 1 \quad \text{เมื่อ} \quad m = 1, 2, \dots, M \quad (3.140)$$

วงจรรองความถี่ดังกล่าวจะมีเสถียรภาพ

3.8.2 การปรับปรุงเสถียรภาพของวงจรรองความถี่

การวิเคราะห์สัญญาณที่ใต้วงจรรองความถี่ที่ขาดเสถียรภาพจะทำให้ค่าที่ได้หมดความหมายอย่างสิ้นเชิง ในทางปฏิบัติเมื่อพหุนามจรรองความถี่ที่ขาดเสถียรภาพจึงต้องมีการแก้ไขคุณลักษณะดังกล่าว ด้วยการแทนโพลของวงจรรองความถี่ที่เป็นต้นเหตุ ซึ่งมีขนาดเท่ากับ หรือมากกว่า 1 ด้วย ค่าคู่แฝด (image) ของโพลดังกล่าวที่มีคุณลักษณะเหมือนกัน จากสมการ (3.133)

$$A_m(z) = \prod_{i=1}^m (1 - z_{mi} z^{-1}) \quad (3.141)$$

ถ้ามีค่า j บางค่าในช่วง $[1, m]$ ที่มีขนาดมากกว่าหรือเท่ากับ 1

$$|z_{mj}| \geq 1 \quad (3.142)$$

จะมีการแทนค่าดังกล่าวด้วยคู่แฝด ซึ่งมีค่า

$$\tilde{z}_{mj} = \frac{z_{mj}}{|z_{mj}|} \quad (3.143)$$

3.9 โครงสร้างของระบบกำเนิดเสียง

โครงสร้างของระบบกำเนิดเสียงหรือวงจรรองความถี่ที่ได้จากการประมาณพี้นะเชิงเส้นในรูปของสัมประสิทธิ์การสะท้อน สามารถหาได้จากสมการ (3.64) และ (3.83)

$$\frac{A_m(z)E(z)}{A(z)} = \frac{A_{m-1}(z)E(z)}{A(z)} + \frac{k_m B_{m-1}(z)E(z)}{A(z)} \quad (3.144)$$

$$\frac{zB_m(z)E(z)}{A(z)} = \frac{k_m A_{m-1}(z)E(z)}{A(z)} + \frac{B_{m-1}(z)E(z)}{A(z)} \quad (3.145)$$

เมื่อ $m = 0$

$$X(z) = \frac{E(z)}{A(z)} \tag{3.146}$$

$$= \frac{A_0(z)E(z)}{A(z)} \tag{3.147}$$

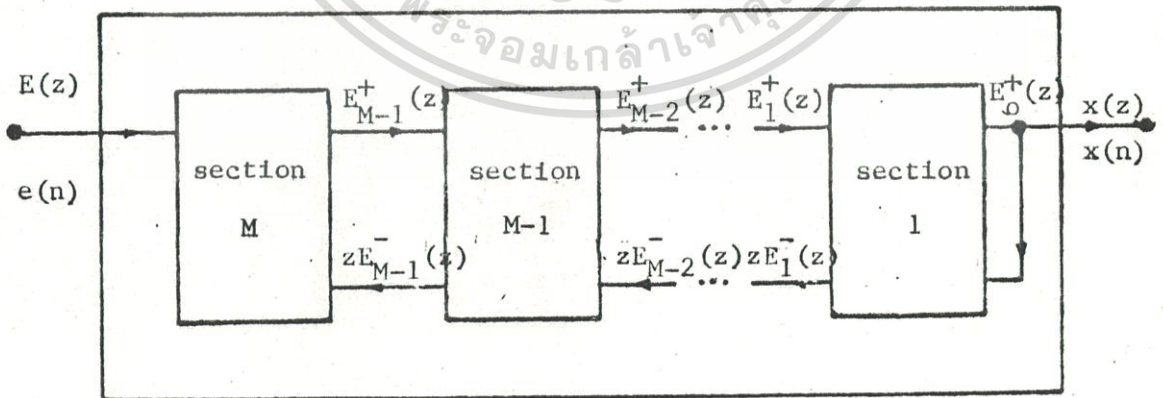
$$= \frac{z B_0(z)E(z)}{A(z)} \tag{3.148}$$

เมื่อ $m = M$

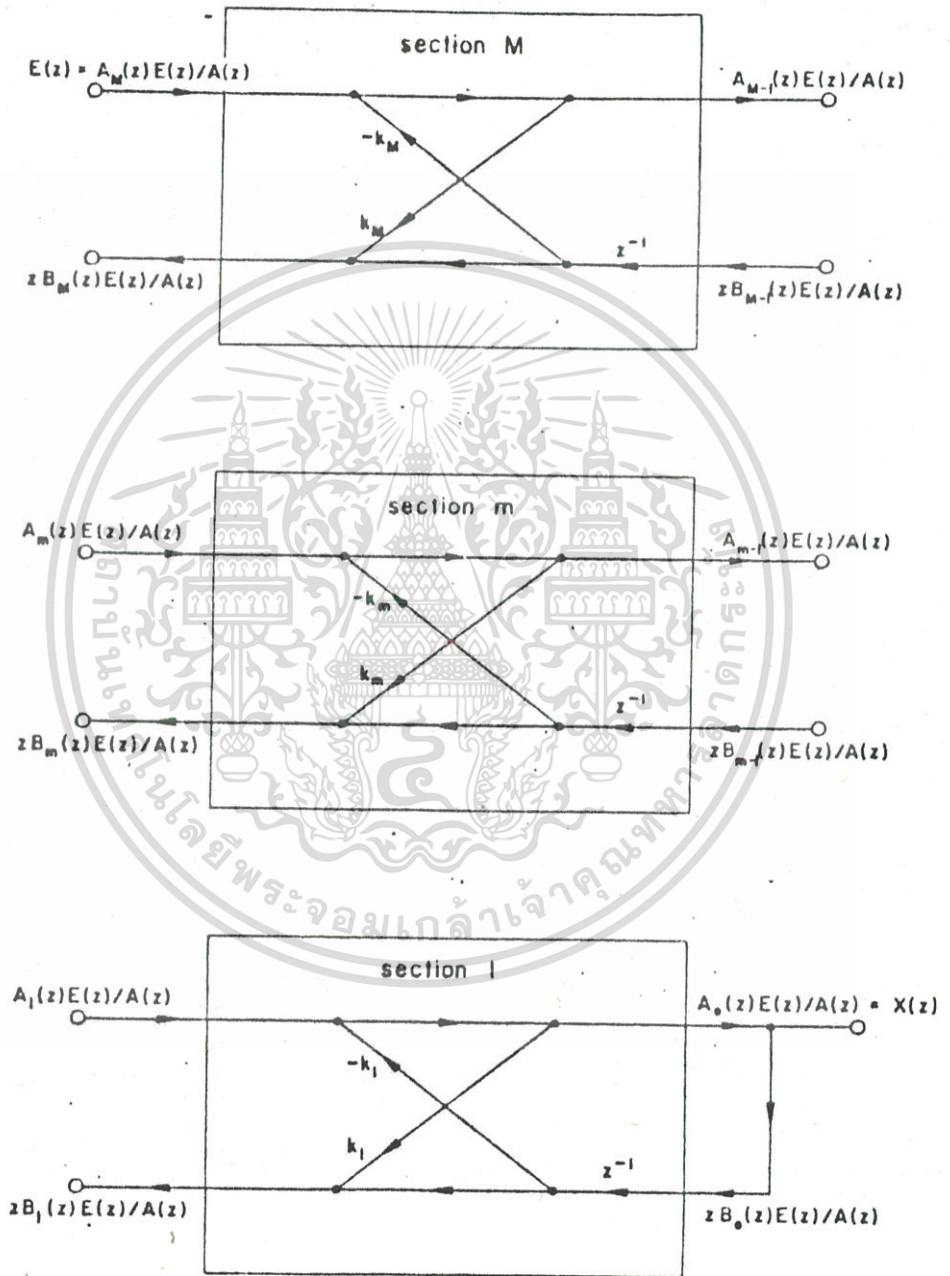
$$A(z) = A_M(z) \tag{3.149}$$

$$E(z) = \frac{A_M(z)E(z)}{A(z)} \tag{3.150}$$

จากสมการ (3.144) ถึง (3.150) ทำให้สามารถสร้างระบบกำเนิดเสียงพูด ที่ประกอบด้วยส่วน (section) ย่อย ๆ ได้ดังรูปที่ 3.7 และ 3.8



รูปที่ 3.7 แสดงโครงสร้างของระบบกำเนิดเสียงพูด



รูปที่ 3.8 แสดงรายละเอียดของระบบกำเนิดเสียงพูดส่วนย่อย ๆ แต่ละส่วน

3.10 สรุป

ในสองบทที่ผ่านมาจะกล่าวถึงการประมาณพันธะเชิงเส้น และเทคนิคที่เกี่ยวข้องในการประมวลสัญญาณเสียงพูดทางด้านทฤษฎี เป็นส่วนใหญ่ เพื่อสร้างพื้นฐานทางทฤษฎีที่จำเป็น ในงานวิจัยที่จะทำต่อไปในวิทยานิพนธ์ฉบับนี้ สิ่งที่ควรสนใจในบทนี้คือ หัวข้อ 3.7 ที่กล่าวถึงขั้นตอนการวิเคราะห์สัญญาณเสียงพูดทั้งสองวิธี และได้แสดงขั้นตอนการวิเคราะห์ในรูปที่ 3.6 เพื่อเป็นแนวทางให้ผู้สนใจนำไปใช้ และในภาคผนวกได้แสดงโปรแกรมย่อย AUTO ที่เขียนขึ้นเพื่อเป็นตัวอย่าง ซึ่งอาจจะต้องปรับปรุงบ้าง เพื่อให้เหมาะสมกับคอมพิวเตอร์แต่ละเครื่อง



บทที่ 4

ความถี่กำลังของเสียงพูด

จุดประสงค์ในการวิเคราะห์เสียงพูดคือ การหาค่าพารามิเตอร์ที่ใช้แสดงคุณลักษณะของสัญญาณเสียง ซึ่งประกอบด้วยความถี่กำลัง และช่วงกระตุ้น

ความถี่กำลังของเสียงพูดหมายถึง ความถี่ธรรมชาติของท่อนำเสียงซึ่งขึ้นกับตำแหน่งของลิ้น ฟัน และริมฝีปากตลอดจนอวัยวะอื่น ๆ ภายในช่องปาก ตามปกติความถี่กำลังจะมีค่าเฉลี่ยไม่เกิน 3 กิโลเฮิร์ตซ์ สำหรับ 3 ความถี่แรก และ 5 กิโลเฮิร์ตซ์ สำหรับ 5 ความถี่แรก บทนี้จะเป็นบทแรกของงานวิจัย ซึ่งเกี่ยวกับการหาค่าความถี่กำลังของเสียงพูด เพื่อตรวจสอบคุณภาพของเสียงพูดโดยอาศัยทฤษฎีจากบทที่ 2 และบทที่ 3 ประกอบกันในแต่ละขั้นตอนจะมีรูปประกอบ เพื่อให้ง่ายต่อการเข้าใจ และอ้างอิงถึงขณะใช้งาน ส่วนโปรแกรมที่ใช้งานจริงที่เขียนขึ้นขณะทำการวิจัยได้รวบรวมไว้ในภาคผนวก เนื่องจากโปรแกรมเหล่านี้จะต้องแก้ไขตัวถูกนำไปใช้กับคอมพิวเตอร์เครื่องอื่น ที่มีโครงสร้างและรายละเอียดต่างกัน โปรแกรมเหล่านี้จึงเป็นเพียงตัวอย่างในการใช้งานเท่านั้นจึงไม่ได้นำมาแสดงไว้ในแต่ละบทที่กล่าวถึงงานวิจัยที่เกี่ยวข้อง

4.1 การหาค่าความถี่กำลัง

วิธีหาค่าความถี่กำลังด้วยวิธีประมาณพหุระเชิงเส้นของสัญญาณเสียงพูด

$\{s(n) | 0 \leq n \leq N-1\}$ มีขั้นตอนต่าง ๆ ดังนี้

1. การจัดเตรียมข้อมูล
2. การหาพารามิเตอร์ของวงจรรองความถี่
3. การหาข้อมูลคิบ (row data) ที่ใช้ในการตัดสินใจ
4. การตัดสินใจข้อมูล

4.1.1 การเตรียมข้อมูล

การจัดเตรียมข้อมูลที่จะนำไปใช้ในการประมาณพหุระเชิงเส้น $\{x(n)\}$ จากสัญญาณ

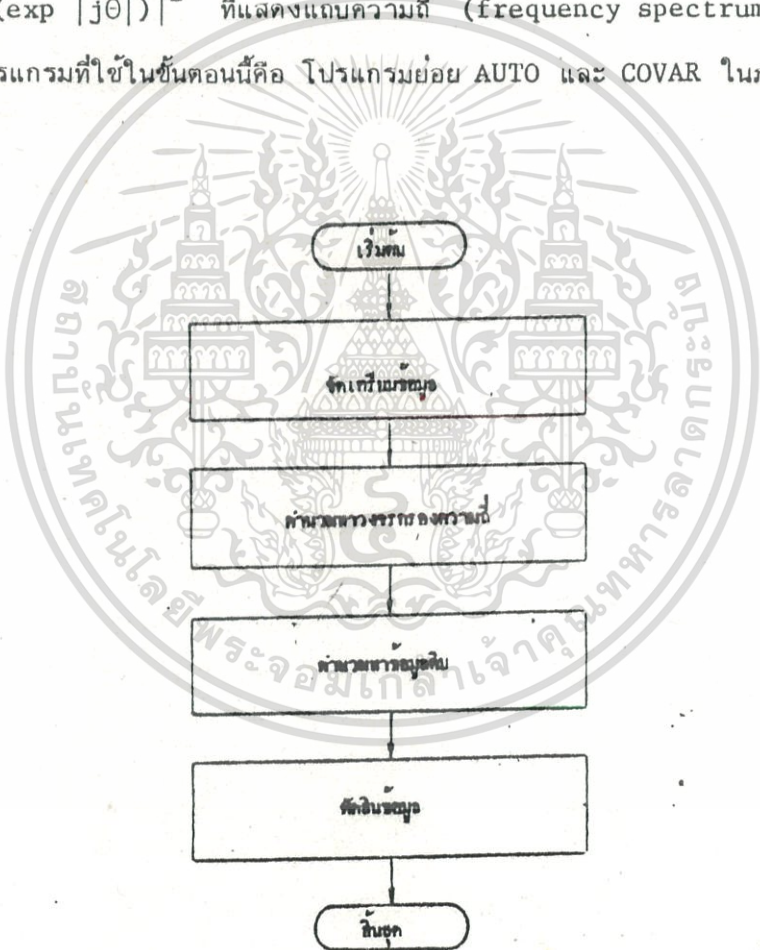
เสียง $\{s(n)\}$ ประกอบด้วยขั้นตอนย่อย ๆ สองขั้นตอน คือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

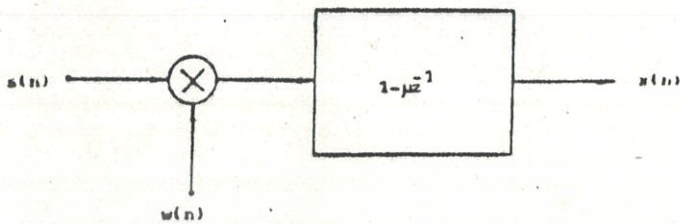
1. การนำสัญญาณผ่านวินโดว์ฟังก์ชัน (window function)
2. การพรีเอมฟาสิส (pre-emphasis)

4.1.2 การหาพารามิเตอร์ของวงจรรองความถี่

สัมประสิทธิ์ของวงจรรองความถี่ $A(z)$ ที่ได้กล่าวถึงวิธีการหาแล้วในบทก่อนจะถูกนำมาใช้ในบทนี้ เพื่อหาลักษณะการตอบสนองต่อความถี่ (frequency response) ของวงจรรองความถี่ $1/|A(\exp |j\omega|)|^2$ ที่แสดงแถบความถี่ (frequency spectrum) ของเสียงพูด ตัวอย่างของโปรแกรมที่ใช้ในขั้นตอนนี้คือ โปรแกรมย่อย AUTO และ COVAR ในภาคผนวก



รูปที่ 4.1 แสดงขั้นตอนการหาความถี่กำลังของเสียงพูด



รูปที่ 4.2 แสดงบล็อกไดอะแกรม (block diagram) การจัดเตรียมข้อมูล

4.1.3 การหาข้อมูลดิบ

ข้อมูลดิบของสัญญาณเสียงพูดหมายถึง ความถี่กำลังและช่วงกว้างของการตอบสนอง (band width) ของสัญญาณแต่ละความถี่ ซึ่งวิธีการหาค่าดังกล่าวทั้งสองจากวงจรรองความถี่ $1/A(z)$ มี 2 วิธี คือ

1. การหาค่าจากรากของฟังก์ชันพหุนาม
2. การหาค่าจากลักษณะการตอบสนองต่อความถี่ของวงจรรองความถี่

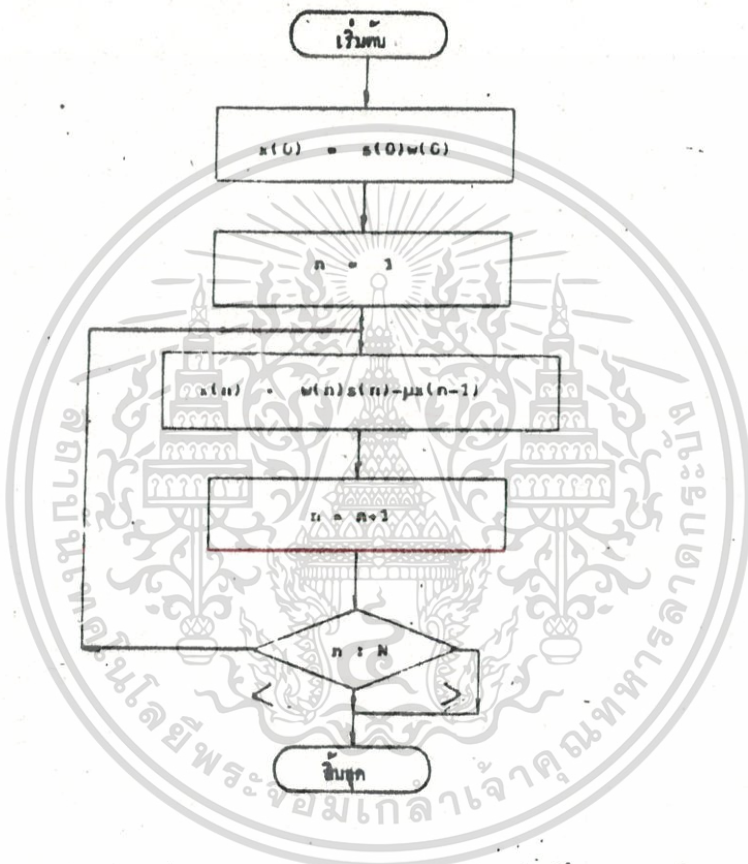
การหาค่าข้อมูลดิบจากรากของฟังก์ชันพหุนาม $A(z)$ ซึ่งมีค่า

$$z = \text{Re}(z) + j \text{Im}(z) \tag{4.1}$$

จากนิยามของตัวแปลง z (z-transform) ซึ่งมีค่า

$$z = \exp(sT) \tag{4.2}$$

$$s = -\pi \hat{B} \pm 2\pi \hat{F} j \tag{4.3}$$



$w(n) = 1$ เมื่อ $0 \leq n \leq N-1$ สำหรับวินโดวโค้งสี่เหลี่ยมขี้นก
(rectangular window)

$w(n) = (1.54 - 0.46 \cos \frac{2\pi n}{N-1})$ เมื่อ $0 \leq n \leq N-1$
สำหรับวินโดวโค้งสี่เหลี่ยมคางหมู
(Hamming window)

รูปที่ 4.3 แสดงขั้นตอนการจัดเตรียมข้อมูล

โดย \hat{B} และ \hat{F} เป็นข้อมูลดิบของช่วงกว้างการตอบสนอง และความถี่กำลังของเสียงพูด ซึ่งมีค่า

$$\hat{F} = \frac{f_s}{\pi} \tan^{-1} \left[\frac{I_m(z)}{R_e(z)} \right] \quad (4.4)$$

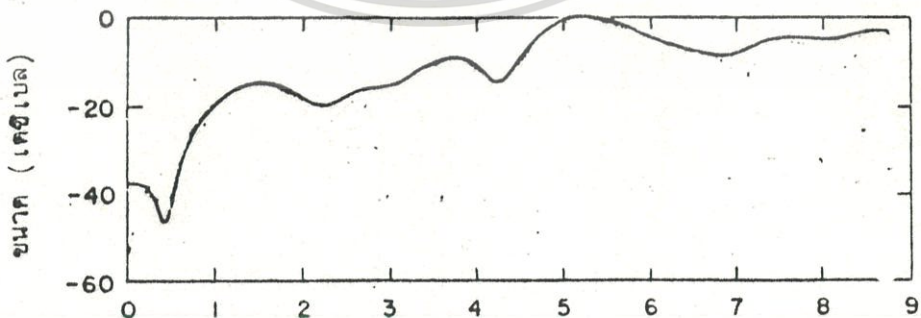
$$\hat{B} = \frac{-f_s}{\pi} \ln |z| \quad (4.5)$$

การหาค่าข้อมูลดิบอีกวิธีหนึ่งคือ การหาค่าจากลักษณะการตอบสนองต่อความถี่ของวงจรรองความถี่ $1/A(z)$ ซึ่งมีค่า

$$X(n) = \frac{1}{|A(\exp[j\theta_n])|^2} \quad (4.6)$$

$$\theta_n = \frac{\pi n}{N} \quad (4.7)$$

ที่ความถี่ $f_s n/2N$ ในรูปที่ 4.4 ได้แสดงค่า $\{X(n) | 0 \leq n \leq N-1\}$ ซึ่งจะพบความถี่กำลังอยู่ตามยอดคลื่น (peak) ของค่าดังกล่าว ส่วนค่าที่ได้จากการทดลองขณะทำวิจัยได้แสดงไว้ในรูปที่ 7.3 แต่เนื่องจากค่ายอดคลื่นที่พบ เกิดจากค่าที่ไม่ต่อเนื่อง ซึ่งอาจไม่ใช่ค่าสูงสุดที่แท้



ความถี่ (กิโลเฮิรตซ์)

รูปที่ 4.4 แสดงลักษณะการตอบสนองของวงจรรองความถี่ $1/A(z)$

จริง ถ้า $X(i)$ เป็นค่าสูงสุดดังกล่าว จะสามารถหาค่าของฟังก์ชันที่ต่อเนื่อง $X(\lambda)$ ได้ด้วยสมการ

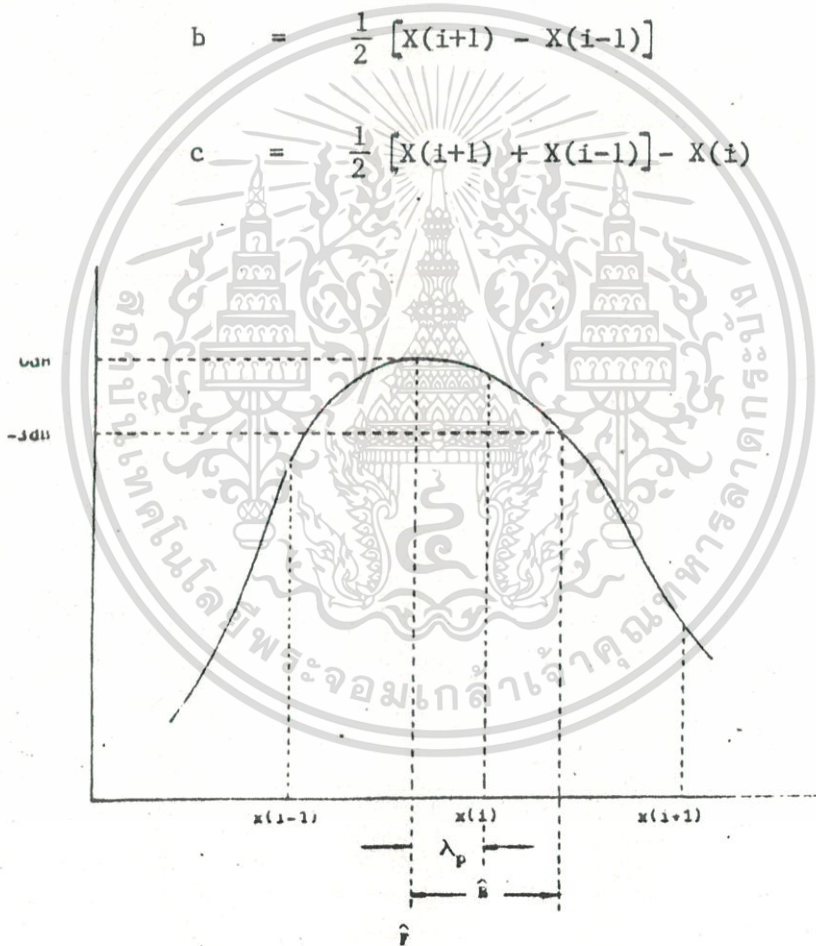
$$X(\lambda) = a + b\lambda + c\lambda^2 \quad (4.8)$$

เมื่อค่าสัมประสิทธิ์ต่าง ๆ มีค่า

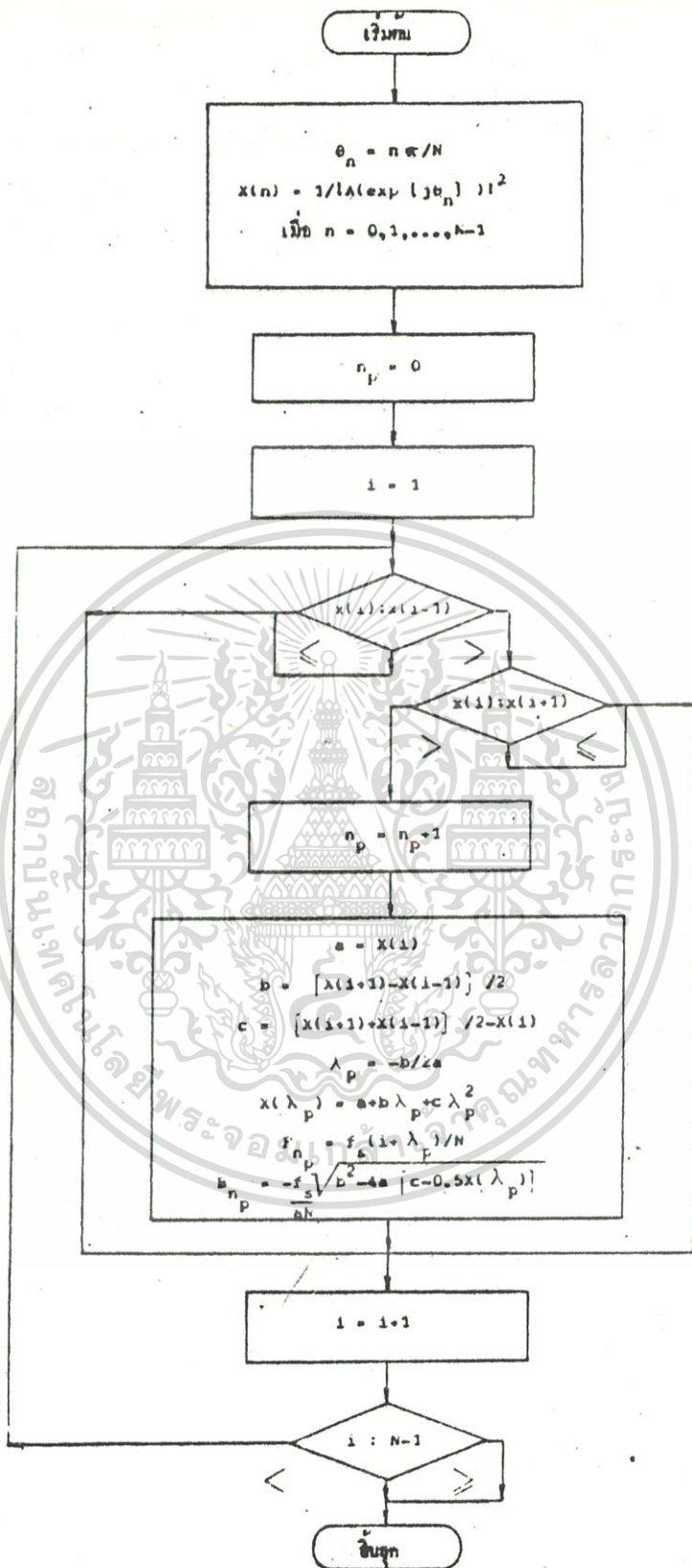
$$a = X(i-1) \quad (4.9)$$

$$b = \frac{1}{2} [X(i+1) - X(i-1)] \quad (4.10)$$

$$c = \frac{1}{2} [X(i+1) + X(i-1)] - X(i) \quad (4.11)$$



รูปที่ 4.5 แสดงลักษณะยอดคลื่น



รูปที่ 4.6 แสดงขั้นตอนการหาขั้วมูลติบ

ค่าสูงสุดของฟังก์ชัน $X(\lambda)$ สามารถหาได้จากเงื่อนไข

$$\frac{dX(\lambda)}{d\lambda} = 0 \tag{4.12}$$

ซึ่งจะได้ค่าสูงสุดของฟังก์ชัน $X(\lambda)$ ซึ่งมีค่า

$$\lambda_p = -\frac{b}{2a} \tag{4.13}$$

$$\hat{F} = \frac{f_s}{N} (i + \lambda_p) \tag{4.14}$$

$$\hat{B} = \frac{-f_s}{aN} \sqrt{b^2 - 4a[c - 0.5X(\lambda_p)]} \tag{4.15}$$

ตัวอย่างของโปรแกรมที่ใช้ในขั้นตอนนี้คือ โปรแกรมย่อย FFT และ LOWDAT

4.1.4 การตัดสินใจข้อมูล

การตัดสินใจข้อมูลเป็นขั้นสุดท้ายในการหาค่าความถี่กำลังของเสียงพูด เนื่องจากข้อมูลดิบของความถี่กำลังที่ได้ อาจไม่ใช่ค่าที่แท้จริง จึงจำเป็นต้องตัดสินใจ เลือกความถี่กำลังที่ถูกต้องโดยอาศัยข้อมูลดิบที่ได้ และข้อมูลที่ได้จากสัญญาณเสียงพูดก่อนช่วยในการตัดสินใจ โดยเงื่อนไขที่ใช้ในการตัดสินใจตั้งอยู่บนสมมติฐาน 2 ข้อ คือ

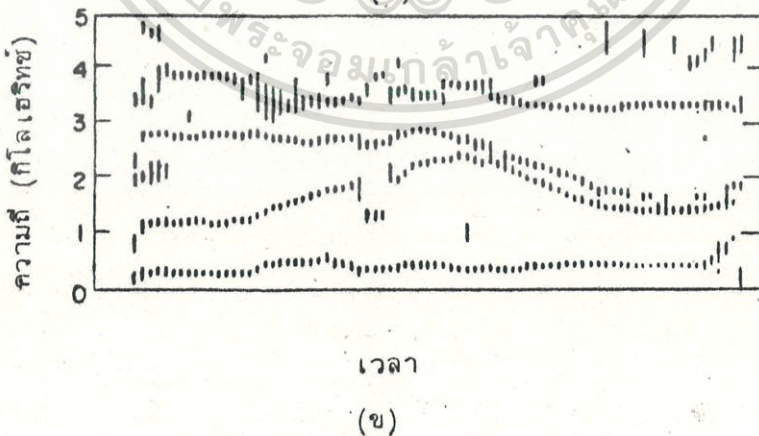
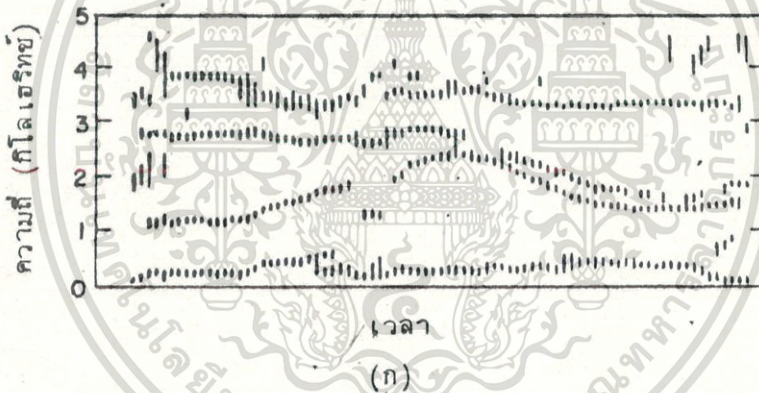
1. ความถี่กำลัง 3 ความถี่แรกจะมีค่าไม่เกิน 3 กิโลเฮิรตซ์ และแต่ละความถี่ต้องมีช่วงกว้างการตอบสนองไม่เกิน 500 เฮิรตซ์ ซึ่งจากผลการทดลองที่มีผู้รวบรวมไว้ ปรากฏว่าสมมติฐานดังกล่าวใช้ได้กับเสียงผู้ชายส่วนใหญ่ ซึ่งมีเสียงต่ำกว่าเสียงผู้หญิงและเด็ก

2. ข้อมูลดิบแต่ละชุด จะประกอบด้วยความถี่กำลังไม่เกิน 4 ความถี่จากข้อมูลที่มีผู้รวบรวมไว้ ปรากฏว่า 85-90 เปอร์เซ็นต์ จะมีความถี่กำลังในข้อมูลดิบแต่ละชุด 3 ความถี่ และ 10-15 เปอร์เซ็นต์ ที่พบความถี่กำลัง 2 หรือ 4 ความถี่ และมีน้อยกว่า 1 เปอร์เซ็นต์ ที่พบความถี่กำลังเพียงความถี่เดียว ส่วนข้อมูลดิบที่มีความถี่กำลังเกิน 4 ความถี่แทบจะไม่พบเลย ถ้าการวิเคราะห์สัญญาณถูกกระทำอย่างถูกต้องและเหมาะสม

จากสมมุติฐานที่กำหนดไว้ การตัดสินใจข้อมูลมีขั้นตอนต่าง ๆ ดังนี้

1. ถ้าข้อมูลดิบมีความถี่กำลัง 3 ความถี่ มีค่าไม่เกิน 3 กิโลเฮิรตซ์และแต่ละความถี่มีช่วงกว้างการตอบสนองไม่เกิน 500 เฮิรตซ์ จะตัดสินใจโดยเลือกความถี่ทั้ง 3 ความถี่ของข้อมูลดิบ เป็นความถี่กำลังที่ถูกต้องของสัญญาณยุคปัจจุบัน โดยไม่สนใจว่าความถี่กำลังของข้อมูลชุดก่อน ๆ จะมีค่าเท่าไร

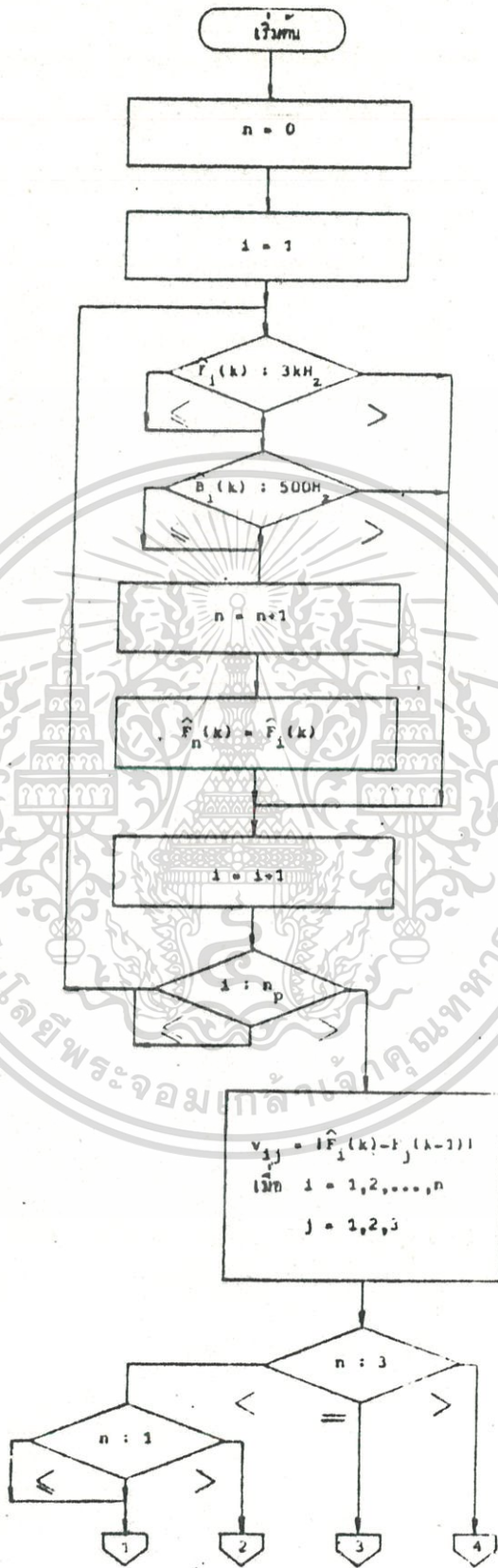
2. ถ้าข้อมูลดิบมีความถี่กำลังมากหรือน้อยกว่า 3 ความถี่ที่มีค่าอยู่ในช่วง 3 กิโลเฮิรตซ์ และมีช่วงกว้างการตอบสนองไม่เกิน 500 เฮิรตซ์ การเลือกความถี่หรือหาความถี่กำลังมาเพิ่มเติม จะต้องอาศัยข้อมูลจากสัญญาณเสียงชุดก่อน ๆ ที่ผ่านมาช่วยในการตัดสินใจโดยกำหนดตัวพารามิเตอร์ เพื่อช่วยช่วยในการตัดสินใจ



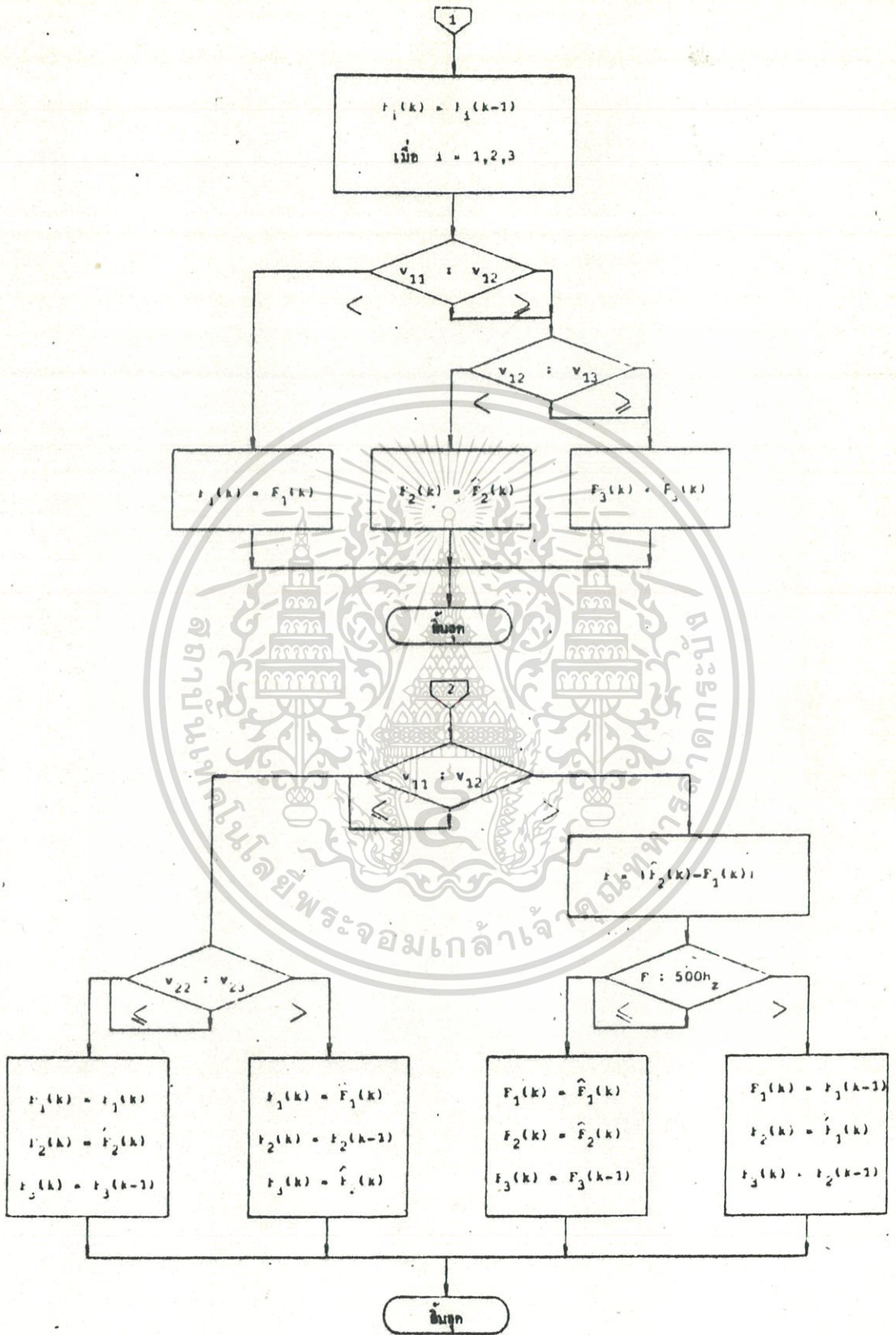
รูปที่ 4.7 ข้อมูลดิบของสัญญาณเสียงพูด

ก) เมื่อไม่มีการพรีเอมฟาซิส

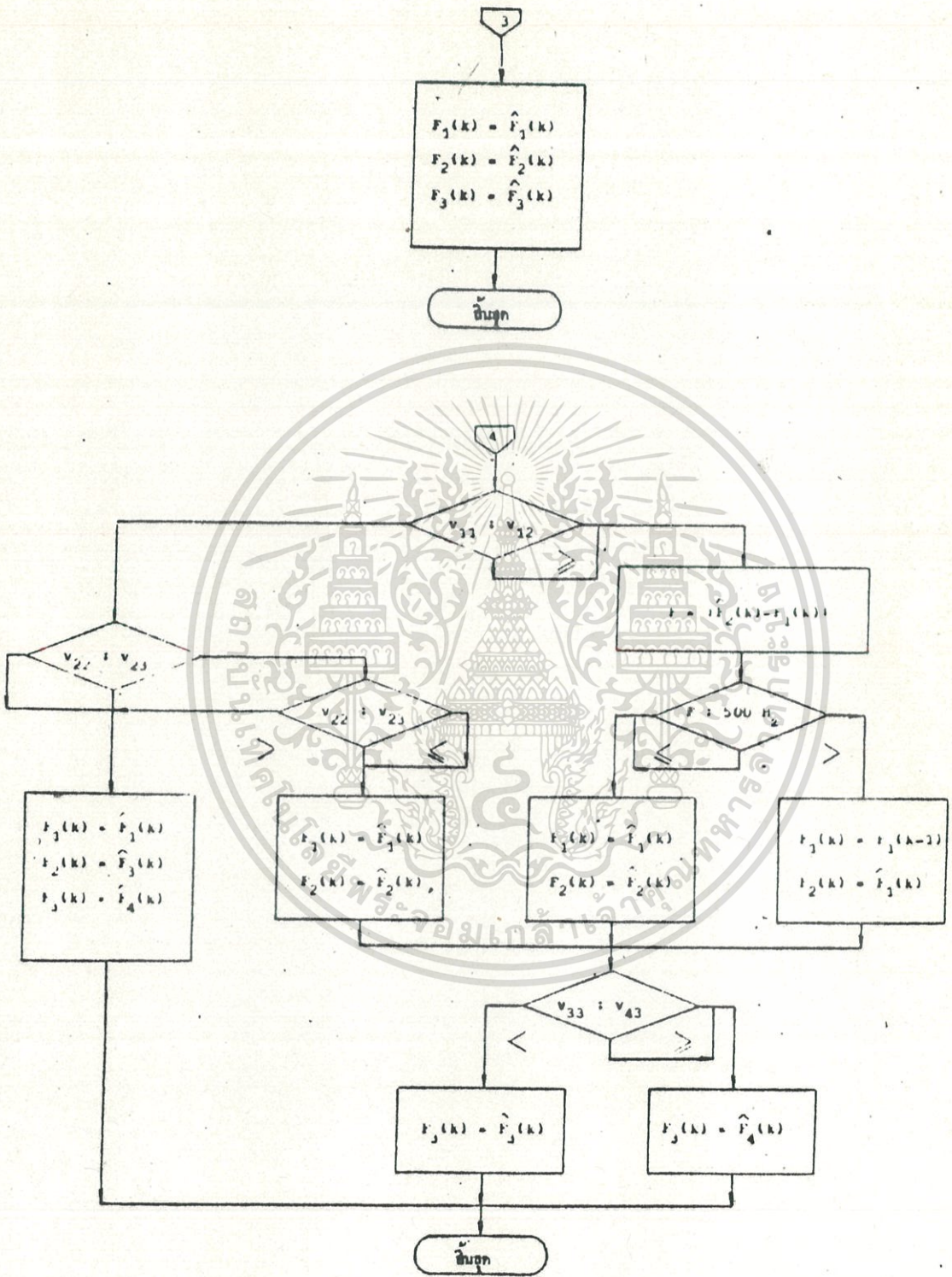
ข) เมื่อมีการพรีเอมฟาซิส



รูปที่ 4.8 แสดงขั้นตอนการตัดสินใจ



รูปที่ 4.8 (ต่อ)



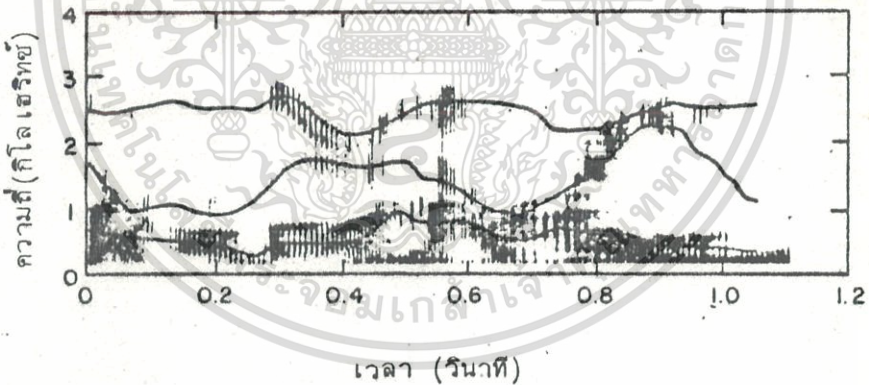
รูปที่ 4.8 (ต่อ)

$$V_{ij} = |\hat{F}_i(k) - F_j(k-1)| \quad (4.16)$$

เมื่อ $\hat{F}_i(k)$ เป็นข้อมูลดิบของความถี่ลำดับที่ i จากข้อมูลชุดที่ k และ $F_j(k-1)$ เป็นความถี่ลำดับที่ j จากข้อมูลชุดที่ $k-1$ โดยมีเงื่อนไขการตัดสินใจดังนี้

1. ถ้าข้อมูลดิบมีความถี่ค่าธรมไม่ครบ 3 ความถี่ จะนำความถี่ค่าธรมจากข้อมูลชุดก่อนมาใช้กับข้อมูลชุดปัจจุบันทุกความถี่ ยกเว้นบางความถี่ที่มีค่าใกล้เคียงกับความถี่ค่าธรมของข้อมูลดิบชุดปัจจุบัน จะใช้ค่าจากข้อมูลดิบแทน
2. ถ้าข้อมูลดิบมีความถี่ค่าธรมเกิน 3 ความถี่ ให้เลือกค่าเหล่านี้ขึ้นมา 3 ค่าที่ใกล้เคียงกับความถี่ค่าธรมของข้อมูลชุดก่อน

ตัวอย่างโปรแกรมที่ใช้ในขั้นตอนนี้คือ โปรแกรมย่อย FORMNT ในภาคผนวก



รูปที่ 4.9 แสดงความถี่ค่าธรมเทียบกับผลที่ได้จากสเปกโตรแกรม (spectrogram)

4.2 สรุป

ความถี่ค่าธรมของเสียงพูดถึงแม้ว่าจะไม่ได้ใช้ในการกำเนิดเสียงพูดโดยตรงในปัจจุบัน แต่ก็ตาม แต่ในการทดสอบคุณภาพของเสียงพูดตลอดจนการจำคำพูด และการจำเสียงพูดยังคงต้องใช้ความถี่ค่าธรม เป็นพารามิเตอร์ที่สำคัญซึ่งการใช้วิธีประมาณพินธระเชิงเส้นในการหาค่าดังกล่าว มีข้อดีที่เหนือกว่าวิธีอื่น คือการคำนวณทำได้ง่ายไม่ซับซ้อน ใช้เวลาน้อยและมีความแม่นยำสูง

บทที่ 5

ช่วงกระตุ้นของเสียงพูด

ช่วงกระตุ้นของเสียงพูด เป็นพารามิเตอร์ที่ใช้กำหนดคุณลักษณะหรือคุณภาพของเสียงพูด ร่วมกับเวลาที่กำหนดของเสียงพูด ช่วงกระตุ้นของเสียงพูด หมายถึงระยะเวลาที่เส้นเสียงสั่นปิดเปิด กล้องเสียงครบ 1 รอบ ซึ่งจากจังหวะการปิดเปิดกล้องเสียงนี้เองที่ทำให้เสียงพูดของคนเรา แต่ละคนมีน้ำเสียงและท่วงทำนองการออกเสียงแตกต่างกัน

ในบทนี้จะแสดงการทำช่วงกระตุ้นของเสียงพูด โดยใช้วิธี SIFT (Simplified Inverse Filter Tracking) ซึ่งใช้การประมาณพหุคูณเชิงเส้นที่กล่าวไปแล้วในบทที่ 3 เป็นพื้นฐาน สาเหตุที่เลือกวิธี SIFT ในการหาช่วงกระตุ้น เนื่องจาก

1. สัญญาณเสียงพูดไม่เป็นสัญญาณรายคาบ (periodic signal) ที่แท้จริงมีการเปลี่ยนแปลงทั้งลักษณะรายละเอียดของสัญญาณ และคาบเวลาการเกิดช่วงกระตุ้นตลอดเวลา ทำให้การหาช่วงกระตุ้นด้วยวิธีอื่นเกิดความผิดพลาดอยู่เสมอ แต่วิธี SIFT จะใช้ค่าฟังก์ชันอัตโนมัติ (autocorrelation) ในการตัดสินใจครั้งสุดท้าย ทำให้ความผิดพลาดจากสาเหตุดังกล่าวไม่เกิดขึ้น
2. สัญญาณเสียงพูดเป็นผลรวมของการตอบสนองต่อสัญญาณพัลซของท่อกำธรเสียง และสัญญาณกระตุ้นจากกล้องเสียงซึ่งเป็นการยากที่จะแยกสัญญาณทั้งสองออกจากกัน เมื่อใช้วิธีการอื่น ๆ ในการหาช่วงกระตุ้น จึงมักเกิดการตัดสินใจเลือกค่าที่ผิดพลาดเนื่องจากค่าความถี่กำธรที่มีค่าต่ำ แต่วิธี SIFT จะแยกความถี่กำธรของเสียงพูดออกขั้นหนึ่งก่อน แล้วจึงทำการเลือกค่าช่วงกระตุ้นจากสัญญาณกระตุ้นโดยตรง จึงได้เปรียบกว่าวิธีอื่น
3. วิธี SIFT จะทำการลดจำนวนข้อมูลดิบที่ใช้ในการประมวลผล ทำให้สิ้นเปลืองเวลาวิเคราะห์น้อยกว่าวิธีอื่น ๆ

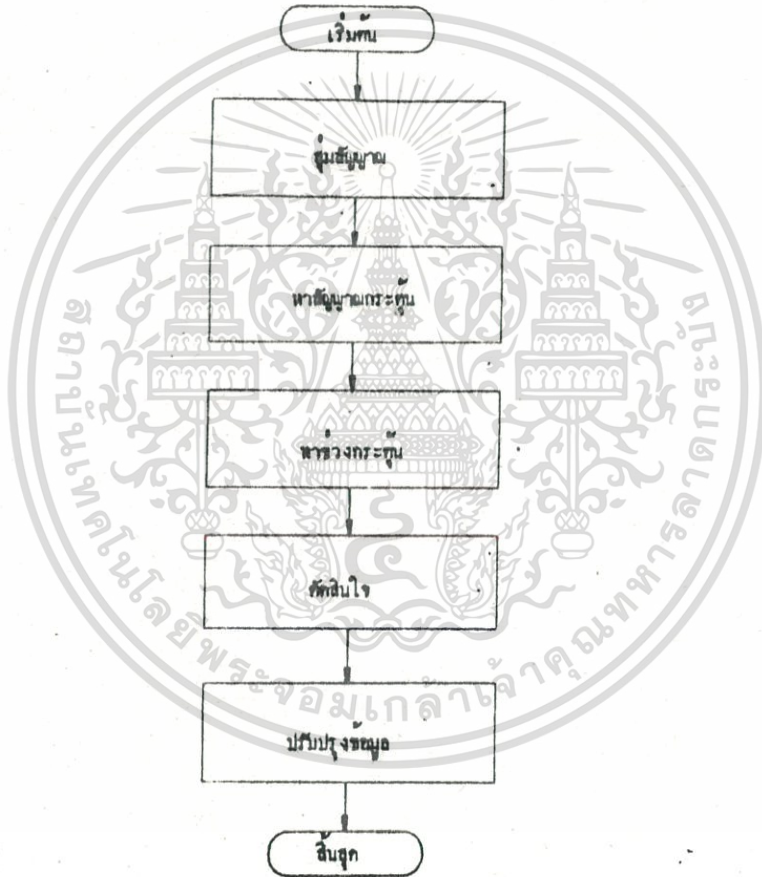
5.1 การหาช่วงกระตุ้นของเสียงพูดด้วยวิธี SIFT

SIFT เป็นวิธีการหนึ่งที่ใช้ในการหาช่วงกระตุ้นเฉลี่ยของสัญญาณเสียงช่วงสั้น ๆ ที่มีความถี่อยู่ระหว่าง 60 ถึง 300 เฮิรตซ์ ซึ่งมีขั้นตอนดังนี้

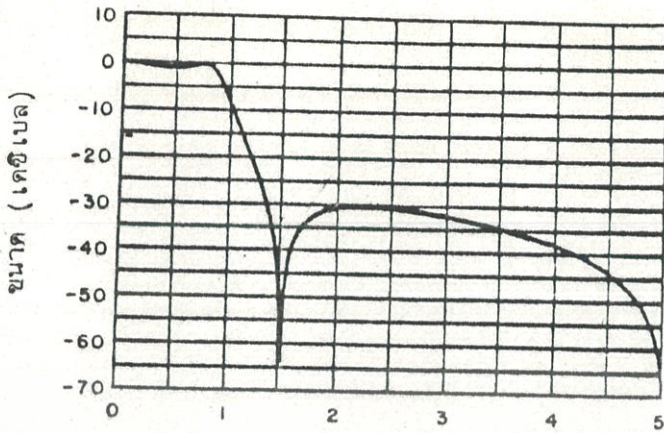
1. การสุ่มสัญญาณ (sampling)
2. การหาสัญญาณกระตุ้น
3. การหาช่วงกระตุ้น
4. การตัดสินใจ
5. การปรับปรุงข้อมูล

5.1.1 การสุ่มสัญญาณ

การสุ่มสัญญาณเป็นขั้นตอนแรกของการหาช่วงกระตุ้นโดยวิธี SIFT เพื่อลดจำนวนข้อมูลที่จะใช้ในการประมวลผลในขั้นต่อไปให้น้อยลง แต่ก่อนที่จะมีการสุ่มสัญญาณตามทฤษฎีการสุ่มสัญญาณ (sampling theorem) จำเป็นต้องทำให้สัญญาณที่จะทำการสุ่มมีความถี่สูงสุดเป็นครึ่งหนึ่ง หรือน้อยกว่าความถี่ ในการสุ่มสัญญาณโดยนำสัญญาณป้อนผ่านวงจรรองความถี่แบบความถี่ต่ำผ่าน (lowpass filter) รูปที่ 5.2 แสดงคุณลักษณะในเชิงความถี่ของวงจรรองความถี่ที่มีช่วงกว้างการตอบสนองประมาณ 1 กิโลเฮิรตซ์ เพื่อใช้จำกัดความถี่สูงสุดของสัญญาณเสียงพูดให้สามารถทำการสุ่ม ตัวอย่างสัญญาณโดยใช้ความถี่ 2 กิโลเฮิรตซ์ หรือมากกว่าการสุ่มสัญญาณโดยใช้วงจรรองความถี่แบบดีจิตอลที่มีคุณลักษณะตามรูปที่ 5.2 เป็นตัวจำกัดความถี่ของสัญญาณเสียงพูดที่ถูกสุ่มสัญญาณด้วยอัตรา 10 กิโลเฮิรตซ์ ให้มีความถี่สูงสุดไม่เกิน 1 กิโลเฮิรตซ์ เพื่อทำการสุ่มตัวอย่างใหม่ในอัตรา 2 กิโลเฮิรตซ์ ซึ่งจะทำให้สามารถลดข้อมูลที่ใช้ในการประมวลผลเพื่อหาช่วงกระตุ้นจากจำนวนเดิม 400 จุดต่อข้อมูล 1 ชุด ให้เหลือเพียง 80 จุดต่อข้อมูล 1 ชุด ซึ่งจะทำให้การประมวลผลในขั้นต่อไปสามารถ ลดเวลาลงได้มาก รูปที่ 7.4 จะแสดงสัญญาณเสียงที่ผ่านวงจรรองความถี่แบบความถี่ต่ำผ่าน และรูปที่ 7.5 จะแสดงสัญญาณที่ผ่านการสุ่มสัญญาณ



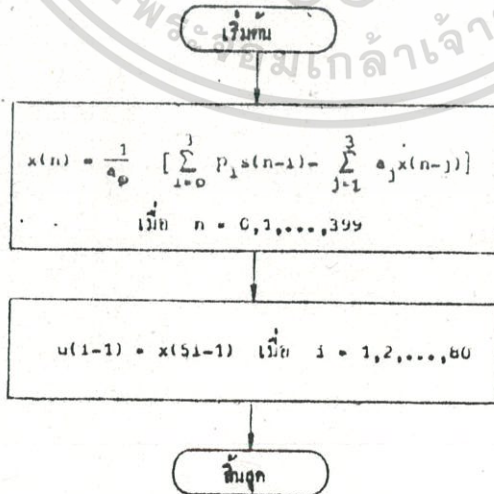
รูปที่ 5.1 แสดงขั้นตอนการหาช่วงกระดูกของเสียงพูด



ความถี่ (กิโลเฮิร์ตซ์)

i	a_i	p_i
0	1.000000	0.0357082
1	-2.340366	0.0069956
2	2.011900	0.0069956
3	-0.614109	0.0357082

รูปที่ 5.2 แสดงสัมประสิทธิ์และคุณลักษณะในเชิงความถี่ของวงจรกรองความถี่ที่ใช้ในการสุ่มสัญญาณในขั้นตอนแรก



รูปที่ 5.3 แสดงขั้นตอนการสุ่มสัญญาณ

5.1.2 การหาสัญญาณกระตุ้นจากกล่องเสียง

การประมาณพัลส์เชิงเส้นสัญญาณเสียงพูดที่กล่าวไว้ในบทก่อนพบว่า

$$\frac{S(z)}{E(z)} = \frac{1}{A(z)} \tag{5.1}$$

$$E(z) = S(z) A(z) \tag{5.2}$$

$$A(z) = \sum_{i=0}^m a_i z^{-i} \quad \text{เมื่อ} \quad a_0 = 1 \tag{5.3}$$

$$e(n) = \sum_{i=0}^m a_i s(n-i) \tag{5.4}$$

ขั้นตอนการหาค่าของสัญญาณกระตุ้นจากกล่องเสียง ประกอบด้วย

1. การพรีเอมฟาริสและการวินโดวสัญญาณ

2. การหาวงจรรองความถี่แบบดิจิทัลด้วยวิธีประมาณพัลส์เชิงเส้น ซึ่งต้องเลือกอันดับ (order) ของวงจรรองความถี่ให้เหมาะสม เพราะถ้าใช้ค่าน้อยเกินไปการหาค่าสัญญาณกระตุ้นด้วยการป้อนสัญญาณเสียงพูดผ่านวงจรรองความถี่จะไม่ได้ผล แต่ถ้าใช้มากเกินไปจะทำให้ใช้เวลาในการคำนวณมาก จากการทดลองพบว่าค่าที่เหมาะสมควรใช้วงจรรองความถี่ที่มีลำดับอยู่ระหว่างอันดับที่ 3 ถึง 5

3. การนำสัญญาณเสียงผ่านวงจรรองความถี่หาได้โดยใช้สมการ (5.4) จะได้สัญญาณกระตุ้นของเสียงพูด ซึ่งได้แสดงไว้ในรูปที่ 7.6

5.1.3 การหาช่วงกระตุ้นของเสียงพูด

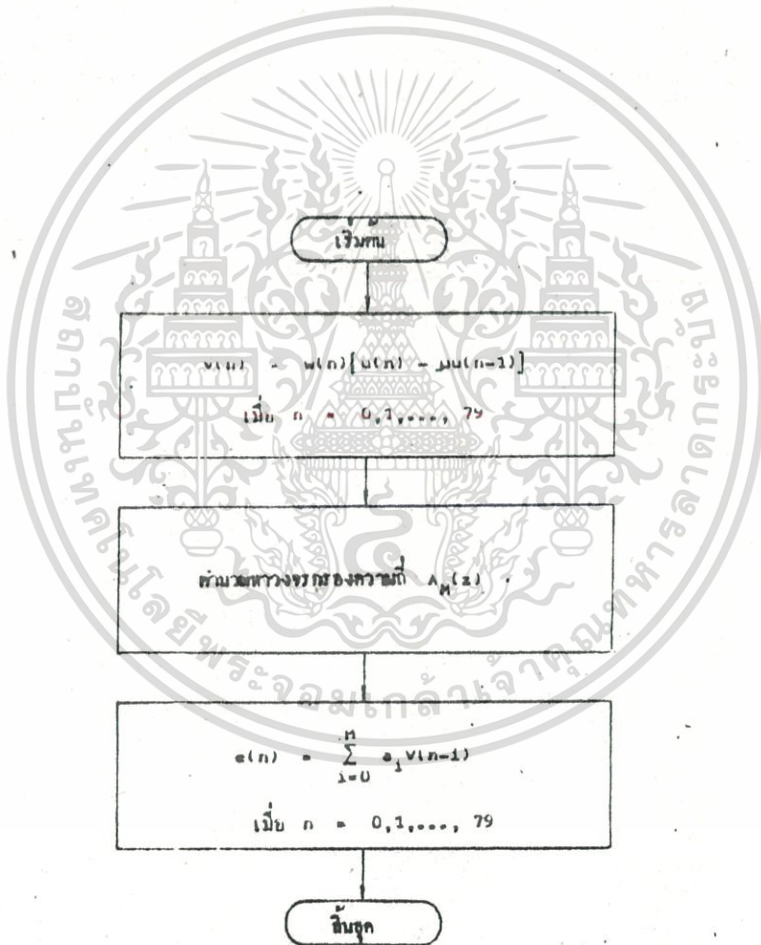
ช่วงกระตุ้นของเสียงพูดหรือคาบเวลา การเกิดสัญญาณกระตุ้นจากกล่องเสียงของสัญญาณ $\{e(n) \ 0 \leq n \leq N-1\}$ สามารถหาได้โดยใช้ฟังก์ชันอัตโนมัติ ซึ่งมีค่า

$$r(n) = \frac{N-1-n}{2} \sum_{i=0}^{N-1-n} e(i)e(i+n) \quad \text{เมื่อ} \quad n = 0, 1, \dots, \frac{N-1}{2} \tag{5.5}$$

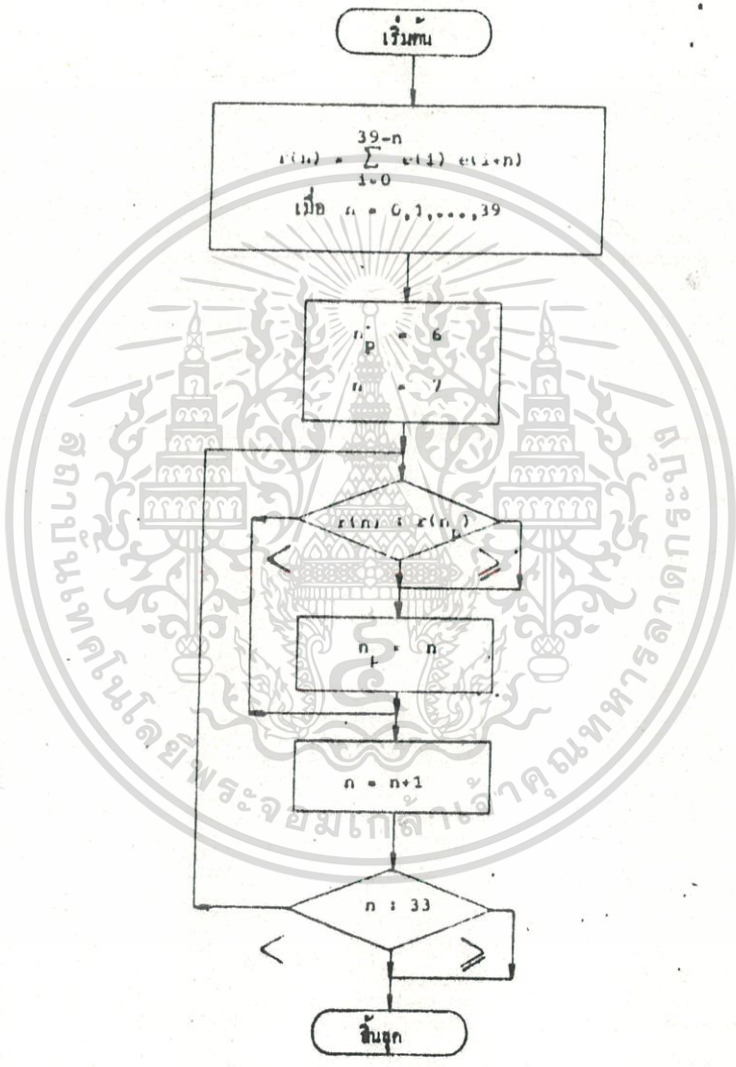
คาบเวลาการเกิดสัญญาณกระตุ้น ซึ่งจะเกิดขึ้นที่สัญญาณระดับที่ n_p ของเซต $\{e(n)\}$ มีค่า

$$r(n_p) = \frac{N-1}{2} \text{ MAX}_{n=1} \{r(n)\} \quad (5.6)$$

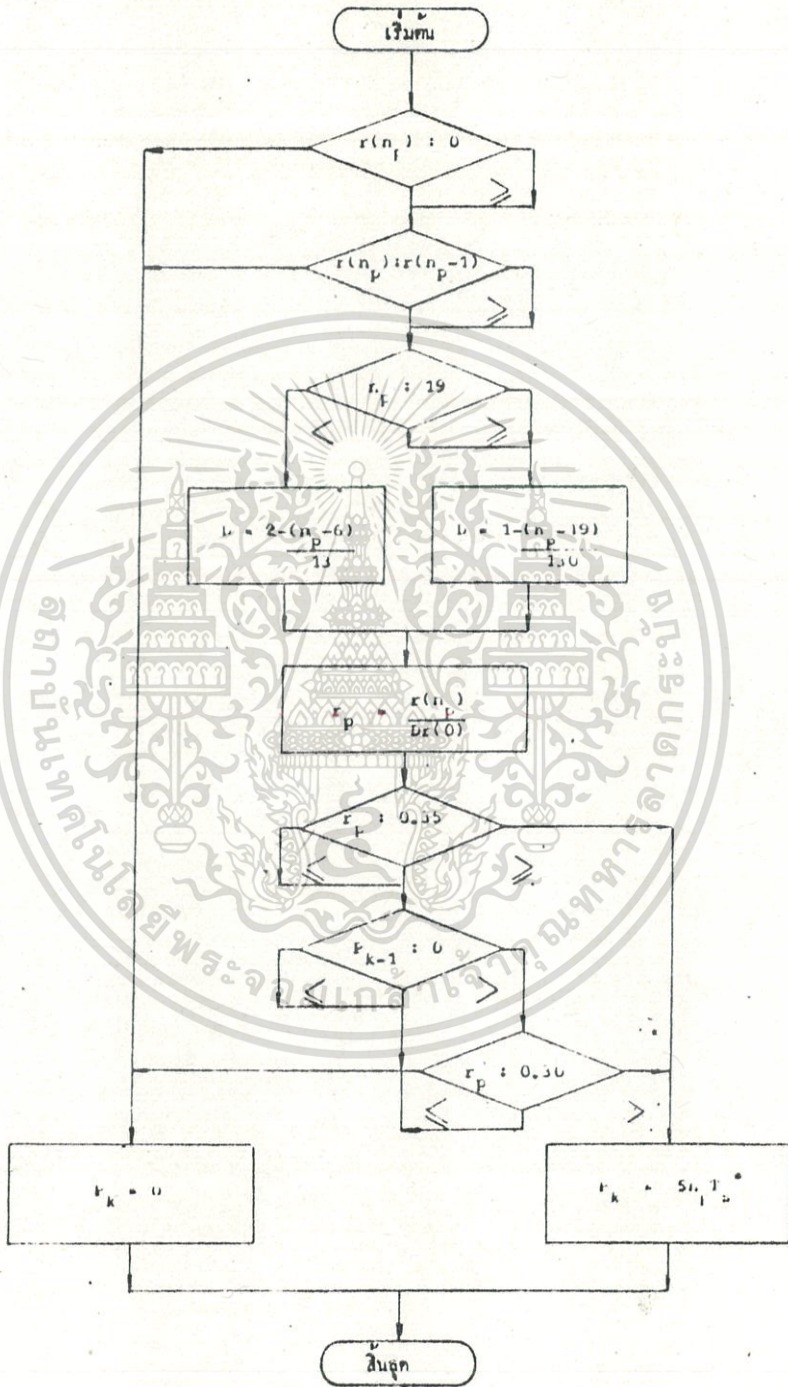
รูปที่ 7.7 ได้แสดงลักษณะของฟังก์ชันอัตโนมัติ ที่จะเห็นช่วงกระตุ้นของเสียงพูดอย่างชัดเจน



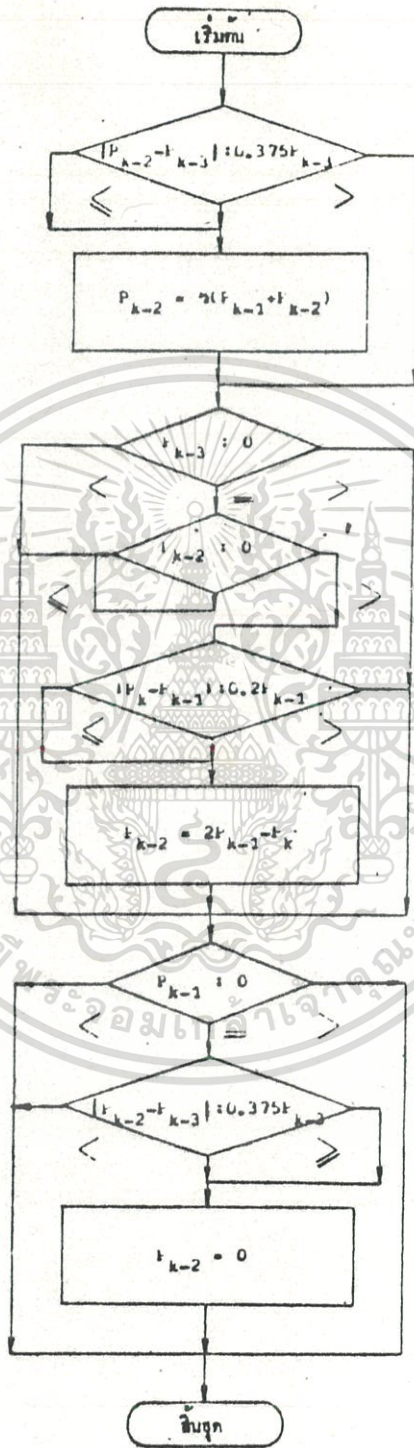
รูปที่ 5.4 แสดงขั้นตอนการหาสัญญาณกระตุ้นจากกล่องเสียง



รูปที่ 5.5 แสดงขั้นตอนการหาช่วงกระดุนของเสียงพูด



รูปที่ 5.6 แสดงขั้นตอนการตัดสินใจเลือกค่าช่วงกระตุ้นของเสียงพูด



รูปที่ 5.7 แสดงขั้นตอนการปรับปรุงข้อมูลของช่วงกระตุ้นในขั้นสุดท้าย

5.1.4 การตัดสินใจ

ช่วงกระตุ้นของเสียงพูดที่ได้จากค่าสูงสุดของฟังก์ชันอัตโนมัติ อาจจะไม่ใช่วงกระตุ้นที่แท้จริง จึงจำเป็นต้องใช้ค่าดังกล่าวเป็นข้อมูลดิบในการตัดสินใจร่วมกับค่าช่วงกระตุ้นของข้อมูลเสียงชุดก่อน ขั้นตอนการตัดสินใจได้แสดงไว้ในรูปที่ 5.6

5.1.5 การปรับปรุงข้อมูล

การปรับปรุงข้อมูลเป็นขั้นสุดท้ายในการหาช่วงกระตุ้นของสัญญาณเสียงโดยการลดความไม่ต่อเนื่องของค่าช่วงกระตุ้นของสัญญาณเสียง สำหรับข้อมูลชุดต่าง ๆ ที่อยู่ใกล้เคียงกัน รูปที่ 5.7 ได้แสดงขั้นตอนต่าง ๆ ในการปรับปรุงข้อมูล

5.2 สรุป

บทนี้ได้แสดงรายละเอียดต่าง ๆ อย่างชัดเจนเกี่ยวกับขั้นตอนการหาช่วงกระตุ้นของเสียงพูดด้วยวิธี SIFT ที่ใช้ในขณะทำวิจัย ทั้งการอธิบายถึงขั้นตอนและภาพแสดงขั้นตอนเช่นเดียวกับในบทที่ 4 ส่วนโปรแกรมใช้งานที่เขียนขึ้นได้แสดงไว้ในภาคผนวก คือโปรแกรมย่อย SIFT เพื่อเป็นตัวอย่างสำหรับผู้สนใจ ส่วนผลการทดลองตามขั้นตอนย่อย ๆ ต่าง ๆ ได้แสดงไว้ในบทที่ 7 เพื่อให้ผู้สนใจสามารถนำไปเปรียบเทียบได้

บทที่ 6

ระบบกำเนิดเสียง

การสร้างระบบกำเนิดเสียงพูดขึ้นเพื่อเลียนเสียงพูด เป็นจุดประสงค์ประการหนึ่ง ในการประมวลสัญญาณเสียงพูด ซึ่งในบทนี้จะกล่าวถึงขั้นตอนการใช้ดิจิทัลคอมพิวเตอร์จำลองระบบกำเนิดเสียงขึ้น โดยใช้วิธีอัตรสัมพันธ์ในการประมาณพันธะเชิงเส้น วิเคราะห์หาพารามิเตอร์ของวงจรกรองความถี่ โดยอาศัยทฤษฎีต่าง ๆ ในบทที่ 3 เป็นรากฐาน และใช้วิธี SIFT ที่ได้กล่าวไปในบทก่อนในการหาค่าช่วงกระตุ้นของเสียงพูด ในบทนี้ซึ่งเป็นบทสุดท้ายที่กล่าวถึงรายละเอียดในการค้นคว้าวิจัย สำหรับวิทยานิพนธ์ฉบับนี้มีจุดประสงค์ 2 ประการ คือ

1. ทดลองสร้างระบบกำเนิดเสียงพูดเพื่อเลียนเสียงพูดโดยใช้ดิจิทัลคอมพิวเตอร์
2. ทดลองเลียนเสียงพูดเพื่อทดสอบดูว่า พารามิเตอร์ที่วิเคราะห์ได้จากบทก่อน ๆ สามารถเชื่อถือได้มากน้อยเพียงไร

ระบบกำเนิดเสียงพูดที่จะกล่าวต่อไป จะทำการสุ่มสัญญาณด้วยความถี่ 10 กิโลเฮิรตซ์ และแบ่งสัญญาณเสียงพูดออกเป็นช่วง ๆ มีความยาวช่วงละ 40 มิลลิวินาที และเปลี่ยนพารามิเตอร์ทุกช่วงเวลา 20 มิลลิวินาที เนื่องจากการวิจัยในขั้นต่อไป ต้องการใช่วงจรรวม (Interated circuit) เบอร์ TMS 5220 ของบริษัทเท็กซัสอินสตรูเมนต์ (Texas Instruments) ช่วยในการเลียนเสียงพูด จึงต้องกำหนดรายละเอียดการวิเคราะห์ขั้นให้เหมาะสมกับวงจรรวมดังกล่าว

6.1 การวิเคราะห์สัญญาณเสียง

การวิเคราะห์สัญญาณเสียง หมายถึงการหาพารามิเตอร์ของระบบกำเนิดเสียงพูดซึ่งได้เคยกล่าวมาแล้วในบทที่ 3 โดยพารามิเตอร์ดังกล่าวประกอบด้วย

1. พารามิเตอร์ของวงจรกรองความถี่แบบดิจิทัล
2. ช่วงกระตุ้นของสัญญาณเสียงพูด

6.1.1 การหาพารามิเตอร์ของวงจรรองความถี่

ขั้นตอนการหาพารามิเตอร์ของวงจรรองความถี่ด้วยวิธีประมาณพหุระเชิงเส้น ได้เคยกล่าวมาแล้วในบทที่ 3 ซึ่งในบทนี้จะกล่าวเพิ่มเติมเกี่ยวกับรายละเอียดบางประการ ที่ใช้ในการทดลอง ซึ่งประกอบด้วย

1. การกรองสัญญาณเสียง เพื่อกำจัดสัญญาณความถี่สูงทิ้งไปตามทฤษฎีการสุ่มสัญญาณ ซึ่งในการทดลองได้ใช้วงจรรองความถี่แบบความถี่ต่ำผ่านที่มีช่วงกว้างการตอบสนองประมาณ 4 กิโล-เฮิร์ตซ์ เพื่อจัดเตรียมสัญญาณเสียงให้เหมาะสมกับการสุ่มสัญญาณที่ความถี่ 10 กิโลเฮิร์ตซ์
2. การสุ่มสัญญาณในการทดลองใช้อุปกรณ์สุ่มสัญญาณ (sampler) ซึ่งรวมอุปกรณ์ที่ใช้แปลงสัญญาณอนาล็อกเป็นสัญญาณดิจิทัล หรือ A/D (analog to digital converter) ไว้ด้วย โดยทำการสุ่มเก็บตัวอย่างสัญญาณด้วยอัตรา 10 กิโลเฮิร์ตซ์
3. การพรีเอมพาซิส คือการนำสัญญาณเสียงพูดผ่านวงจรรองความถี่ที่มีฟังก์ชันผ่าน

$$P(z) = \frac{1}{1 - \mu z^{-1}} \tag{6.1}$$

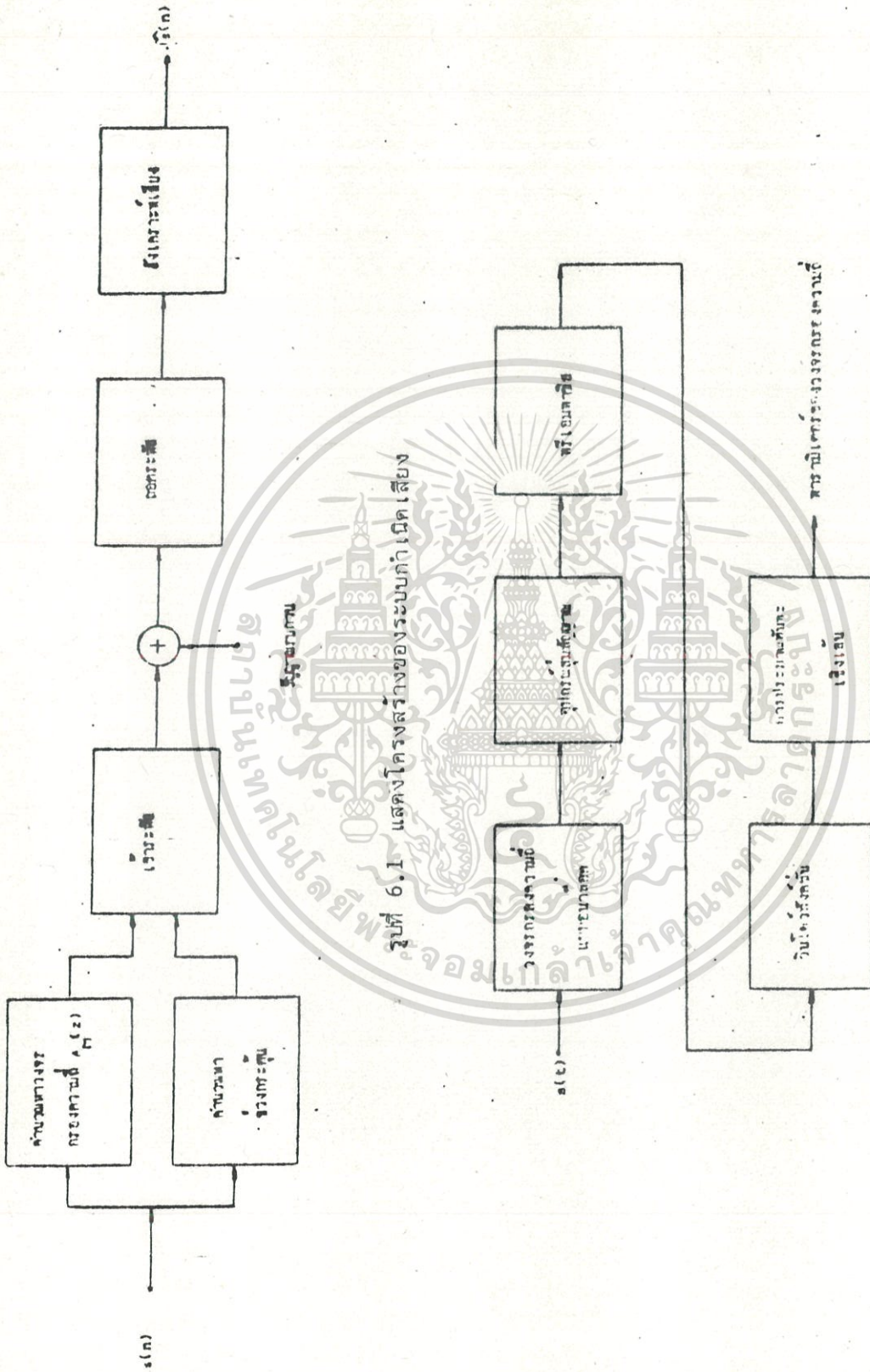
และมีแถบพลังงาน

$$P(\exp |j\omega T|)^2 = \frac{1}{1 + \mu^2 - 2\mu \cos\omega T} \tag{6.2}$$

ซึ่งทำให้สัญญาณเสียงที่ผ่านมีค่าแถบพลังงานสม่ำเสมอขึ้นทำให้การประมาณพหุระเชิงเส้นได้ผลดีขึ้น โดยค่า μ ที่เหมาะสมควร มีค่า

$$\mu = r(1)/r(0) \tag{6.3}$$

เมื่อ $r(.)$ คือฟังก์ชันอัตโนมัติสัมพันธ์ของสัญญาณเสียง แต่เนื่องจากสัญญาณเสียงพูดตามปกติ จะมีค่า μ ตามสมการ (6.3) ประมาณ 0.90 ถึง 0.95 ในการทดลองจึงใช้ค่าดังกล่าวคงที่ประมาณ 0.90 ทุกครั้งเพื่อลดขั้นตอนการคำนวณ



รูปที่ 6.2 แสดงบล็อกไดอะแกรมการหาพหามิติเตอร์ของวงจรถองความถี่

4. การวินโดว์สัญญาณ เป็นการจำกัดขอบเขตของสัญญาณที่ใช้ในการวิเคราะห์แต่ละชุดตามปกติวินโดว์ฟังก์ชันที่ใช้มี 2 ชนิด คือ

ก) วินโดว์แบบจตุรัส มีค่า

$$w(n) = 1 \quad n = 0, 1, \dots, N-1 \quad (6.4)$$

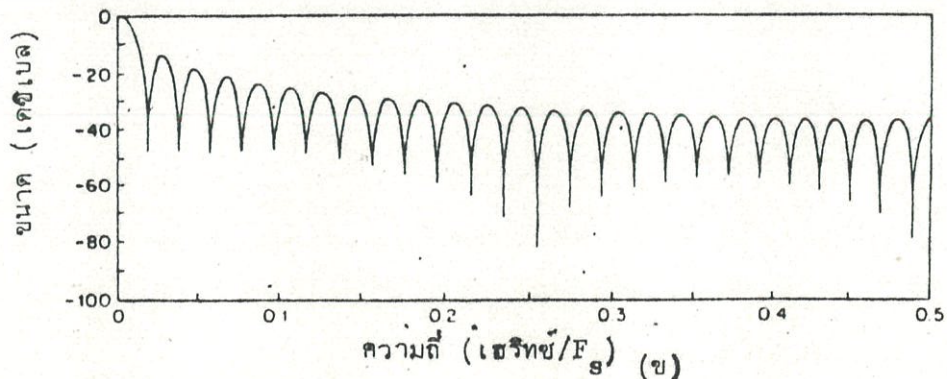
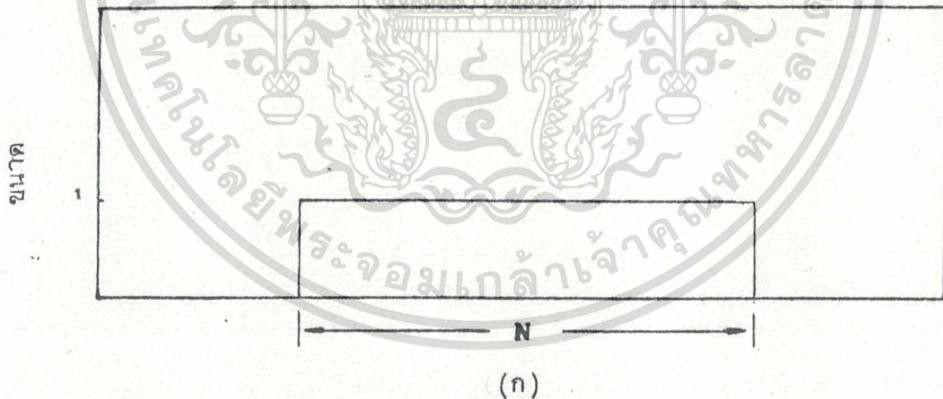
โดยวินโดว์มีความกว้าง N รูปที่ 6.3 ได้แสดงลักษณะของวินโดว์แบบจตุรัสพร้อมแถบพลังงาน

ข) วินโดว์แบบแฮมมิง มีค่า

$$w(n) = 0.54 - 0.46 \cos (2\pi n/N) \quad (6.5)$$

รูปที่ 6.4 ได้แสดงลักษณะของวินโดว์แบบแฮมมิงพร้อมแถบพลังงาน

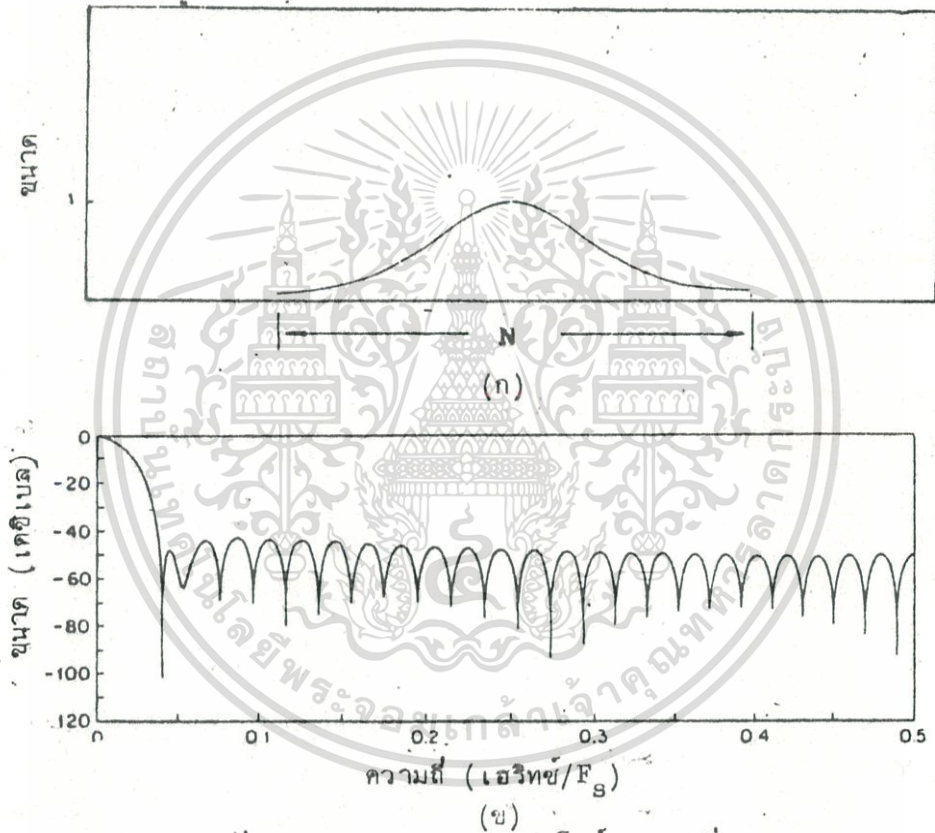
5. การประมาณพหุคูณเชิงเส้น เป็นขั้นสุดท้ายในการหาพารามิเตอร์ของระบบกำเนิดเสียงในส่วนของวงจรกรองความถี่ ซึ่งในการทดลองได้ใช้วิธีออสซิลโลสโคปในการประมาณพหุคูณเชิงเส้น โดยเลือกวงจรกรองความถี่ของระบบกำเนิดเสียงที่มีอันดับ 10 ซึ่งรายละเอียดต่างๆ ได้กล่าวไปแล้วในบทที่ 3



รูปที่ 6.3 ก) ลักษณะของวินโดว์แบบจตุรัส ข) แถบพลังงานของวินโดว์แบบจตุรัส

6.1.2 การหาช่วงกระตุ้นของเสียงพูด

การหาช่วงกระตุ้นของเสียงพูดในที่นี้จะหมายถึงการแยกชนิดของเสียงพูดว่า เป็นเสียงประเภทโฆษะหรือโฆษะด้วย ในการทดลองได้ใช้วิธี SIFT ในการหาช่วงกระตุ้น เนื่องจากวิธีการดังกล่าวใช้เวลาในการคำนวณน้อยและมีประสิทธิภาพค่อนข้างดีสามารถกำจัดต้นเหตุบางประการที่ทำให้การหาช่วงกระตุ้นด้วยวิธีอื่นผิดพลาด ซึ่งได้กล่าวไปแล้วในบทที่ 5



รูปที่ 6.4 ก) ลักษณะของวินโดว์แบบแอมมิง
ข) แอมพลิจูดของวินโดว์แบบแอมมิง

6.1.3 การเข้ารหัสพารามิเตอร์

การเข้ารหัสในที่นี้หมายถึง การเก็บพารามิเตอร์ของระบบกำเนิดเสียงที่วิเคราะห์ ได้ในรูปของรหัสเลขฐานสอง (binary number) ซึ่งพารามิเตอร์ดังกล่าวประกอบด้วย

1. พารามิเตอร์ที่แสดงคุณลักษณะของวงจรกรองความถี่จำนวน 10 ตัว โดยพารามิเตอร์ดังกล่าวอาจเก็บในรูปของสัมประสิทธิ์ของวงจรกรองความถี่หรือสัมประสิทธิ์ PARCOR ก็ได้ขึ้นอยู่กับ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ลักษณะโครงสร้างของระบบกำเนิดเสียงที่ใช้ในการทดลองพารามิเตอร์แต่ละตัว เป็นค่าจริง (real value) จะใช้เนื้อที่เก็บ 4 ไบท์ (byte) หรือ 32 บิต (bit) ต่อหนึ่งค่า

2. อัตราขยายของวงจรรองความถี่หรือความคลาดเคลื่อนกำลังสองรวมอย่างใดอย่างหนึ่ง เป็นค่าจริงเช่นเดียวกัน ใช้เนื้อที่ในหน่วยความจำ 4 ไบท์ หรือ 32 บิต

3. ช่วงกระตุ้นของเสียงพูดใช้เนื้อที่เก็บ 2 ไบท์ หรือ 16 บิต เนื่องจากค่าดังกล่าวเป็น เลขจำนวนเต็ม (integer)

รวมพารามิเตอร์ทั้ง 3 ส่วนแล้วข้อมูลเสียง 1 ชุดจะใช้เนื้อที่หน่วยความจำ 46 ไบท์

6.2 การสังเคราะห์สัญญาณเสียงพูด

การสังเคราะห์สัญญาณเสียงพูดมีขั้นตอนต่าง ๆ ดังนี้

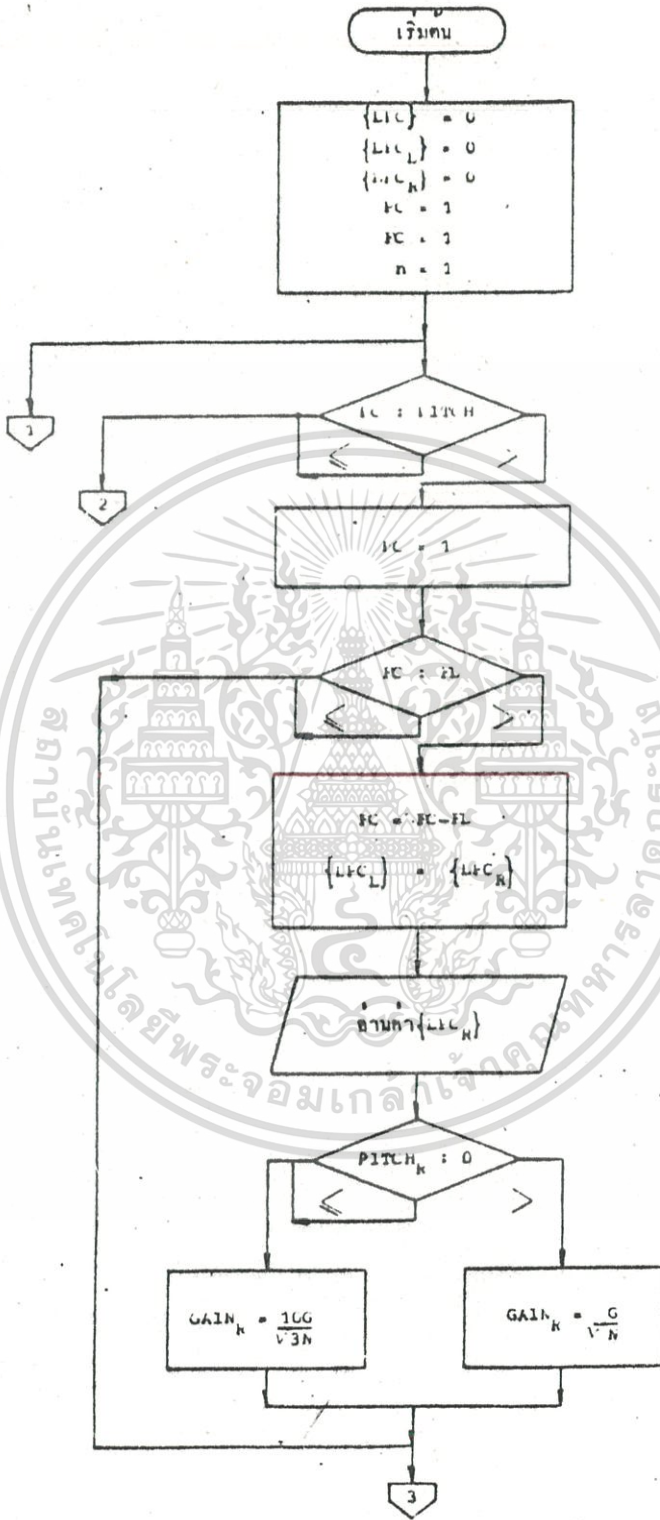
1. การกำหนดเริ่มต้นจะกำหนดให้พารามิเตอร์ของระบบกำเนิดเสียงชุดปัจจุบันคือ {LPC} (linear predictive codes) ซึ่งประกอบด้วยพารามิเตอร์ทั้งหมดที่เข้ารหัสไว้จากการวิเคราะห์ ด้วยการประมาณพันธะเชิงเส้นที่ประกอบด้วย พารามิเตอร์ของวงจรรองความถี่ซึ่งอาจเป็น

$\{k_i | i = 1, 2, \dots, M\}$ หรือ $\{a_i | i = 1, 2, \dots, M\}$ อัตราขยายหรือ G และช่วงกระตุ้น PITCH ตลอดจน $\{LPC_L\}$ และ $\{LPC_R\}$ ซึ่งเป็นพารามิเตอร์ของสัญญาณเสียงชุดข้างเคียงอีกสอง ชุดเป็น 0 หมด พร้อมกันนั้นตัวแปรที่ใช้นับจำนวนสัญญาณ ซึ่งประกอบด้วย PC (pitch counter) FC (frame counter) และ n จำนวนข้อมูลในหน่วยความจำชั่วคราวจะมีค่าเป็น 1 หมด

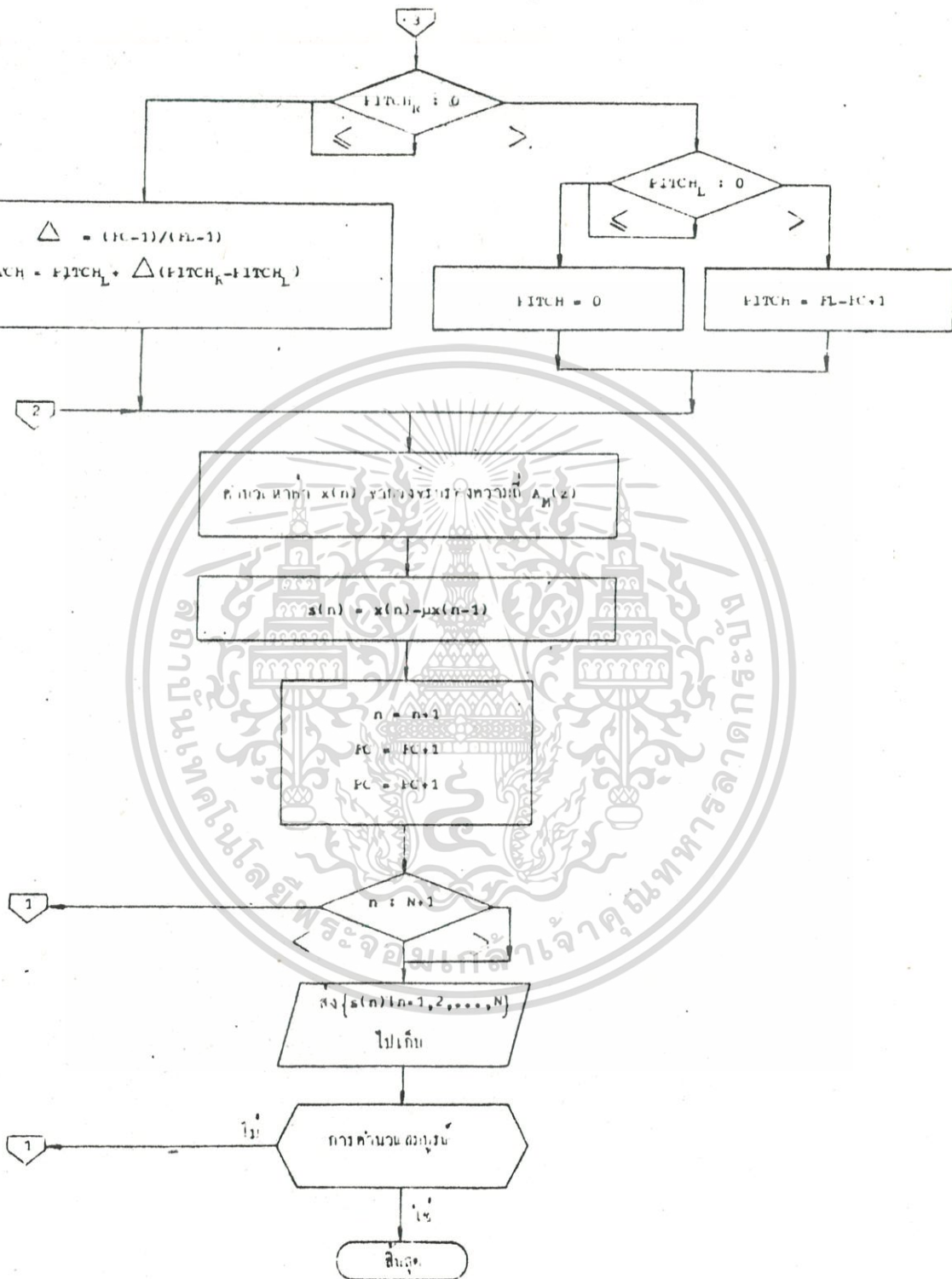
2. ระบบกำเนิดเสียงจะสร้างสัญญาณใหม่ขึ้นเรื่อย ๆ ทีละ 1 ค่า ส่วนวิธีการกำเนิดเสียง จะแตกต่างกันขึ้นกับลักษณะของพารามิเตอร์และโครงสร้างของวงจรรองความถี่ที่ใช้ ซึ่งได้กล่าวไปแล้วในบทที่ 3 ขณะที่ระบบกำเนิดเสียงสร้างสัญญาณเสียงขึ้นใหม่ 1 ค่า ตัวแปรที่ใช้นับจะมีค่าเพิ่มขึ้น 1 ทุกครั้ง

3. ตัวแปร PC จะถูกนำมาเปรียบเทียบกับ PITCH ถ้ามีค่าน้อยกว่าระบบกำเนิดเสียงจะย้อนไปทำงานตามเดิมในข้อ 2 แต่ถ้ามีค่ามากกว่าแสดงว่าการกำเนิดเสียงได้ทำครบช่วงกระตุ้นแล้ว ต้องเริ่มต้นช่วงกระตุ้นใหม่โดยทำตามขั้นตอนข้อ 4

4. PC จะเริ่มต้นค่า 1 ใหม่ ขณะเดียวกัน FC จะถูกนำมาเปรียบเทียบกับ FL (analysis frame length) ถ้ามีค่าน้อยกว่าแสดงว่าช่วงกระตุ้นที่จะสร้างสัญญาณขึ้นใหม่ จะใช้



รูปที่ 6.5 แสดงขั้นตอนการสังเคราะห์เสียงพูด



รูปที่ 6.5 (ต่อ)

พารามิเตอร์ชุดเก่าที่มีอยู่ โดยทำตามขั้นตอนที่ 5 แต่ถ้า FC มากกว่า FL แล้วแสดงว่าพารามิเตอร์ของระบบกำเนิดเสียงถึงเวลาต้องเปลี่ยนใหม่ โดยการนำค่าใน $\{LPC_R\}$ ทั้งหมดเข้าแทนที่ใน $\{LPC_L\}$ แล้วอ่านค่า $\{LPC_R\}$ เข้ามาใหม่ จากนั้นทำตามขั้นตอนที่ 5 ต่อไป

5. ข้อมูลของสัญญาณเสียงชุดข้างเคียงจะถูกทดสอบว่าเป็นเสียงโฆชะทั้งสองชุดหรือไม่ ถ้ามีข้อมูลชุดใหม่หรือทั้งสองชุดเป็นเสียงโฆชะค่าใน $\{LPC\}$ จะถูกกำหนดให้เหมือนกับ $\{LPC_L\}$ แต่ถ้าข้อมูลชุดข้างเคียงทั้งสองเป็นเสียงโฆชะหมดค่าใน $\{LPC\}$ จะกำหนดจากอัตราส่วนของ $\{LPC_L\}$ และ $\{LPC_R\}$ ส่วนอัตราขยายของวงจรรองความถี่ หรือ GAIN สามารถคำนวณหาจากระดับพลังงานเฉลี่ย ซึ่งมีค่า

$$GAIN = 10 G/\sqrt{3N} \tag{6.6}$$

สำหรับเสียงโฆชะ และ

$$GAIN = G/\sqrt{N} \tag{6.7}$$

สำหรับเสียงโฆชะ

6. ระบบกำเนิดเสียงจะทำการสร้างสัญญาณเสียงขึ้น 1 ค่า ตามขั้นตอนที่ 2 ในขณะเดียวกันอาจมีการถ่ายสัญญาณที่สร้างขึ้นในหน่วยความจำไปสู่ที่เก็บถาวรต่อไป เมื่อมีจำนวนมากพอสมควร

6.3 สรุป

ในบทนี้ได้กล่าวถึงรายละเอียดการจำลองระบบกำเนิดเสียงพูด โดยใช้ดิจิทัลคอมพิวเตอร์ ซึ่งรายละเอียดเกี่ยวกับขั้นตอนการวิเคราะห์ที่ได้อธิบายไปแล้วในบทที่ 3 และบทที่ 5 ซึ่งผลลัพธ์ที่ได้จากการวิเคราะห์ที่ได้แสดงไว้ในรูปที่ 7.2 ส่วนการกำเนิดเสียงพูดที่ได้กล่าวไว้ในบทนี้มีตัวอย่างโปรแกรมแสดงไว้ในโปรแกรม SYN ในภาคผนวก และผลที่ได้จากโปรแกรมดังกล่าว ได้แสดงไว้ในรูปที่ 7.8 วิธีการเข้ารหัสเพื่อลดจำนวนหน่วยความจำที่ใช้เก็บพารามิเตอร์ของระบบกำเนิดเสียงในบทนี้เป็นเพียงวิธีการง่าย ๆ ซึ่งช่วยให้ลดจำนวนหน่วยความจำประมาณ 5 ถึง 10 เท่า ส่วนการเข้ารหัสเพื่อให้สามารถลดจำนวนหน่วยความจำประมาณ 50 เท่า ถ้ามีผู้สนใจอาจศึกษาได้จากรายละเอียดของวงจรรวม TMS 5220 ต่อไป

บทที่ 7

ผลการทดลอง

ในบทนี้ได้รวบรวมผลต่าง ๆ ที่ได้จากการทดลองบางส่วนตามขั้นตอนต่าง ๆ ที่ได้กล่าวไว้ในบทที่ 4 ถึงบทที่ 6 โดยเริ่มจากสัญญาณเสียงพูด 1 ชุดที่ใช้เป็นตัวอย่าง มีจำนวนสัญญาณ 400 จุด ใช้เวลา 40 มิลลิวินาที ค่าพารามิเตอร์ต่าง ๆ ของระบบกำเนิดเสียงที่ได้จากการวิเคราะห์ พร้อมทั้งรูปภาพแสดงลักษณะของสัญญาณ เมื่อผ่านการประมวลผลในขั้นตอนต่าง ๆ และในที่สุดท้ายได้แสดงผลที่ได้จากระบบกำเนิดเสียงที่จำลองขึ้น โดยใช้ดิจิทัลคอมพิวเตอร์ ในบทที่ 6

7.1 ผลการทดลองการวิเคราะห์สัญญาณเสียงพูด

ผลจากการวิเคราะห์สัญญาณเสียงพูดสำหรับ 1 ชุดข้อมูลซึ่งมี 400 จุดที่มีลักษณะตามรูปที่ 7.1 ซึ่งประกอบด้วยขั้นตอน 2 ขั้นตอน

1. การหาพารามิเตอร์ของวงจรถ่วงความถี่
2. การหาช่วงกระตุ้นของเสียงพูด

LINEAR PREDICTION ANALYSIS

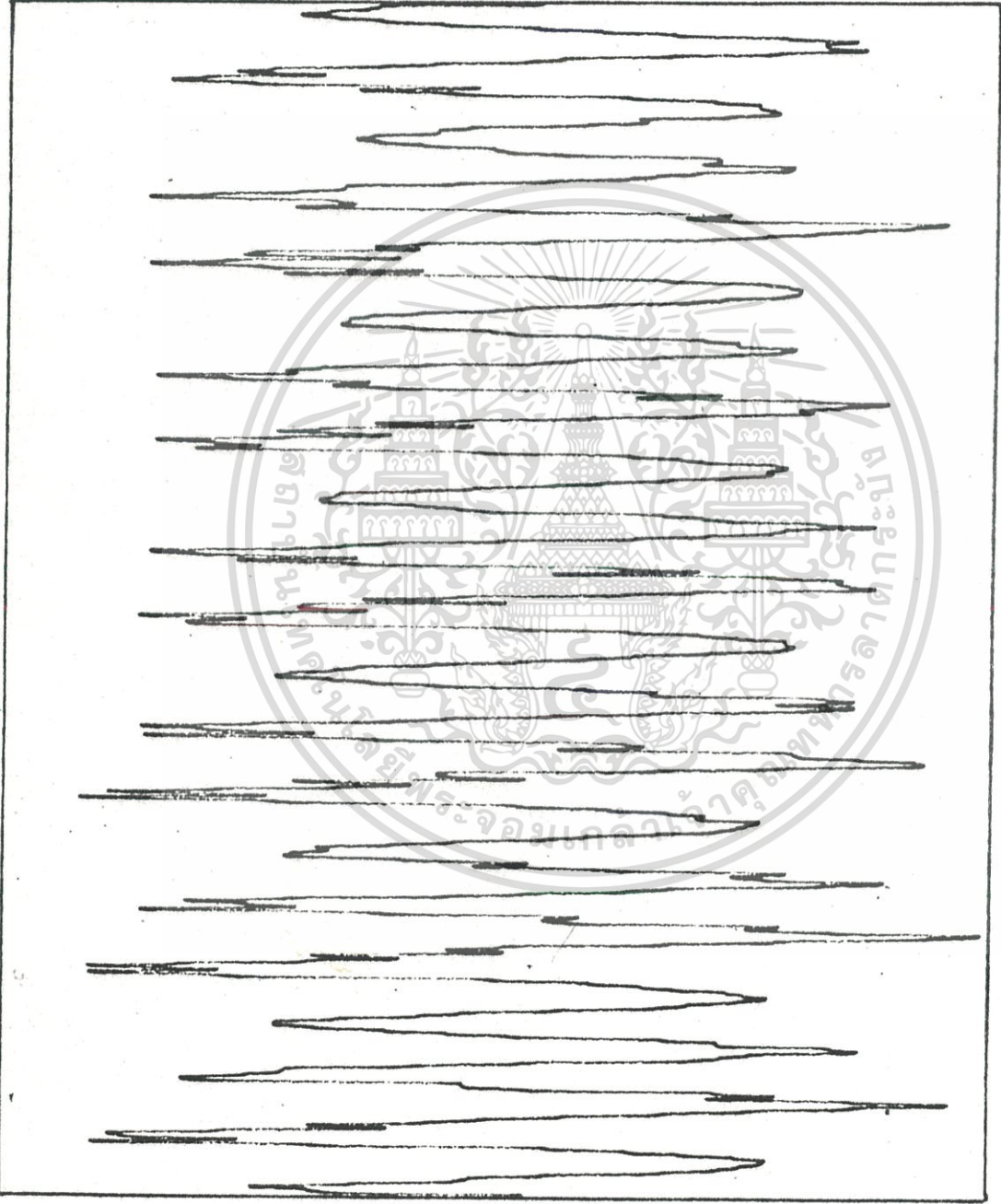
FRAME NUMBER 4

STAGE	PREDICTOR COEFFICIENT	PARCOR COEFFICIENT
0	.10000E+01	
1	.43669E+00	.31050E+00
2	.16619E+00	-.15310E+00
3	-.33822E+00	-.48326E+00
4	-.24817E+00	-.16401E+00
5	-.46495E+00	-.52846E+00
6	-.19547E-01	.64362E-01
7	-.22597E+00	-.27794E+00
8	.23777E+01	.56310E-01
9	-.78656E-01	-.39311E-01
10	-.90840E-01	-.90840-01

GAIN 95.55

PITCH 59

รูปที่ 7.2 แสดงผลจากการวิเคราะห์สัญญาณเสียงพูด



รูปที่ 7.1 แสดงสัญญาณเสียงพูดที่ใช้เป็นตัวอย่างในการวิเคราะห์

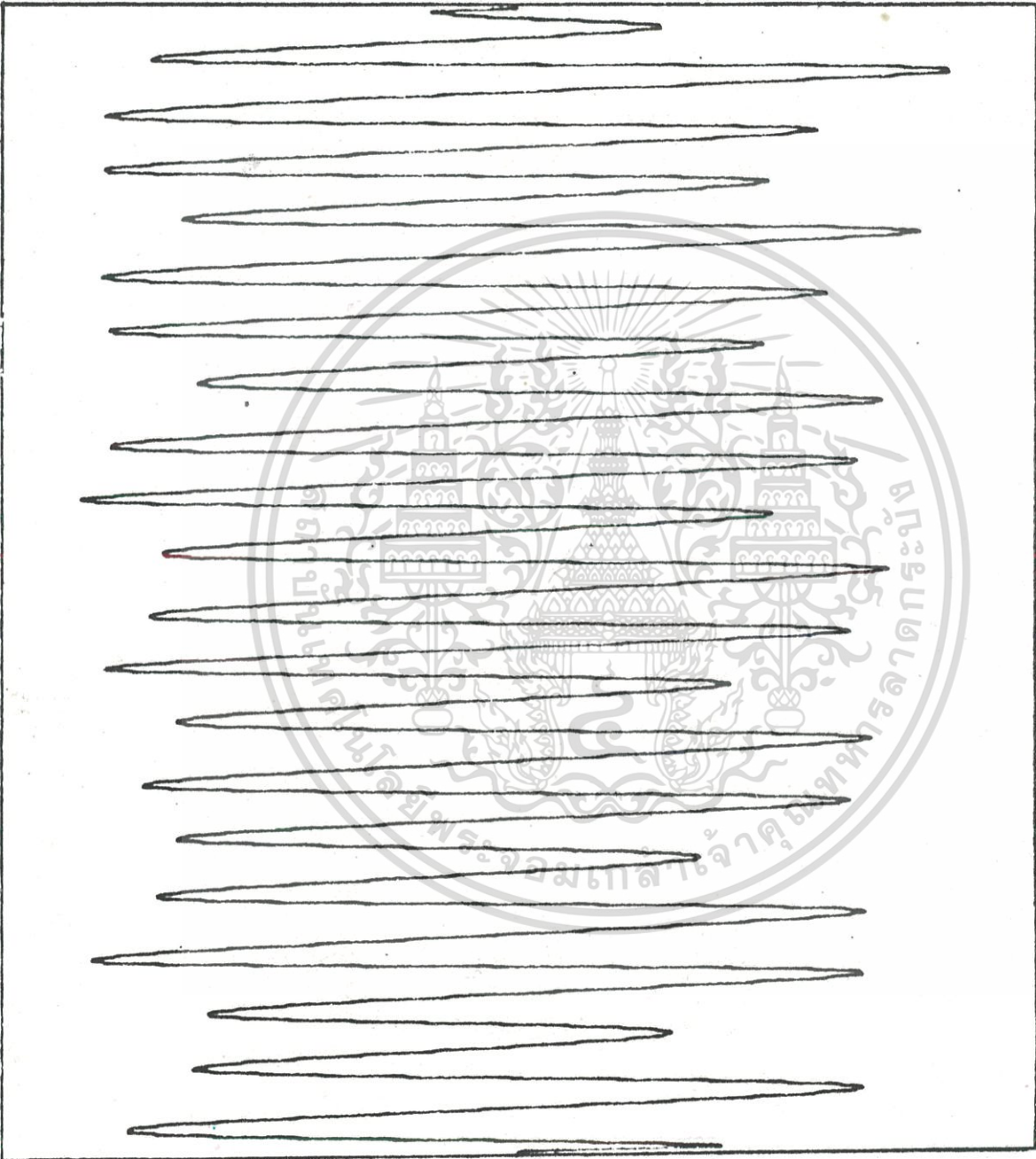


รูปที่ 7.3 (ก) แสดงแถบพลังงานของสัญญาณเสียงพูด

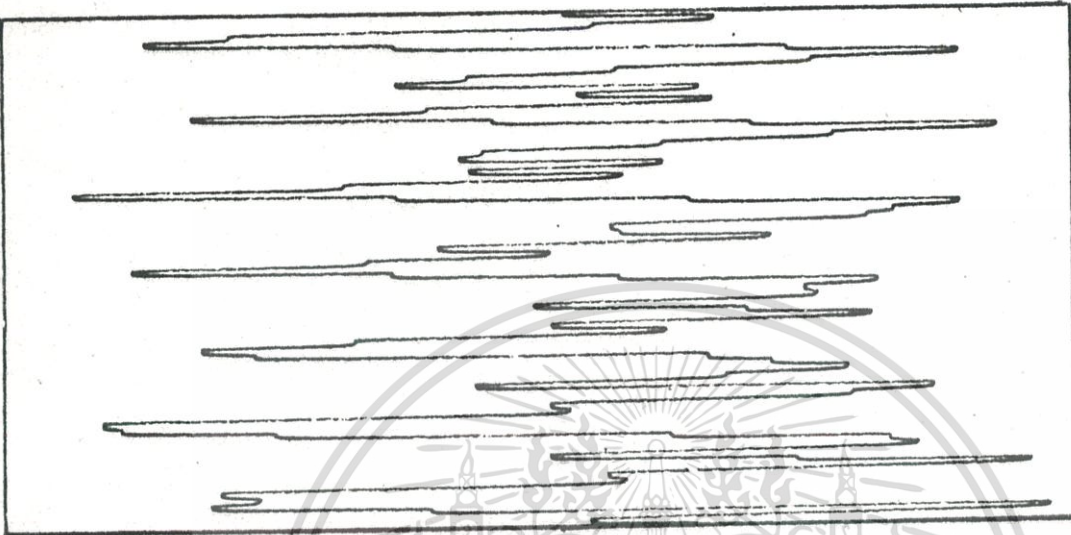
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า



รูปที่ 7.3 (ข) แสดงแถบพลังงานของวงจรงอมความถี่วีเคราะห้ได้



รูปที่ 7.4 แสดงสัญญาณที่ออกจากรวมการของความถี่ที่ใช้ในการส่งสัญญาณ



รูปที่ 7.6 แสดงสัญญาณกระตุกของเสียงพูด



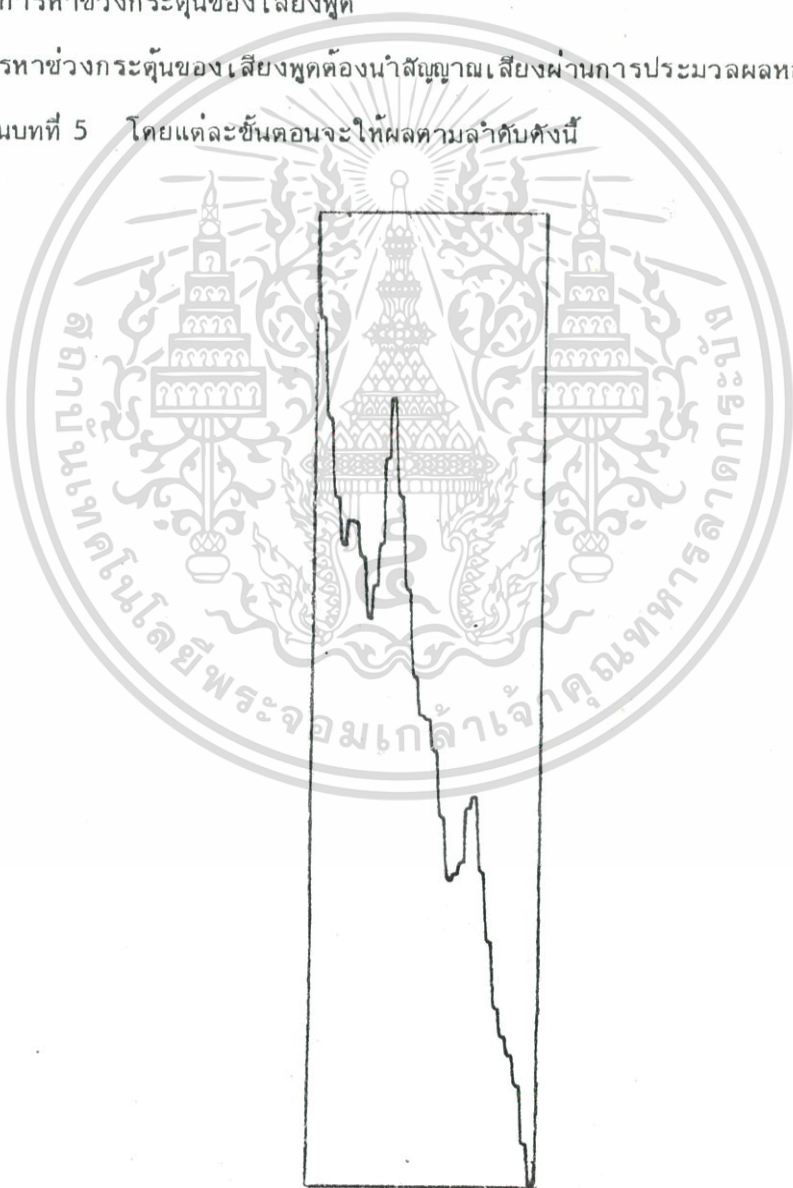
รูปที่ 7.5 แสดงสัญญาณที่ผ่านการสุ่มสัญญาณ

7.1.1 พารามิเตอร์ของวงจรรองความถี่

พารามิเตอร์ของวงจรรองความถี่และสัมประสิทธิ์ PARCOR พร้อมด้วยอัตราขยายของวงจรรองความถี่จากโปรแกรมย่อย AUTO ผ่านช่วงกระตุ้นของเสียงพูดหาได้จากโปรแกรมย่อย SIFT ในขณะเดียวกันก็มีการตรวจสอบคุณภาพของวงจรรองความถี่โดยใช้โปรแกรมย่อย FFT และ LOWDAT ที่กล่าวไว้ในบทที่ 4 เปรียบเทียบกับสัญญาณเสียงจริง จึงได้แสดงไว้ในรูปที่ 7.3

7.1.2 ผลจากการหาช่วงกระตุ้นของเสียงพูด

การหาช่วงกระตุ้นของเสียงพูดต้องนำสัญญาณเสียงผ่านการประมวลผลหลายขั้นตอน ซึ่งได้กล่าวไปแล้วในบทที่ 5 โดยแต่ละขั้นตอนจะให้ผลตามลำดับดังนี้

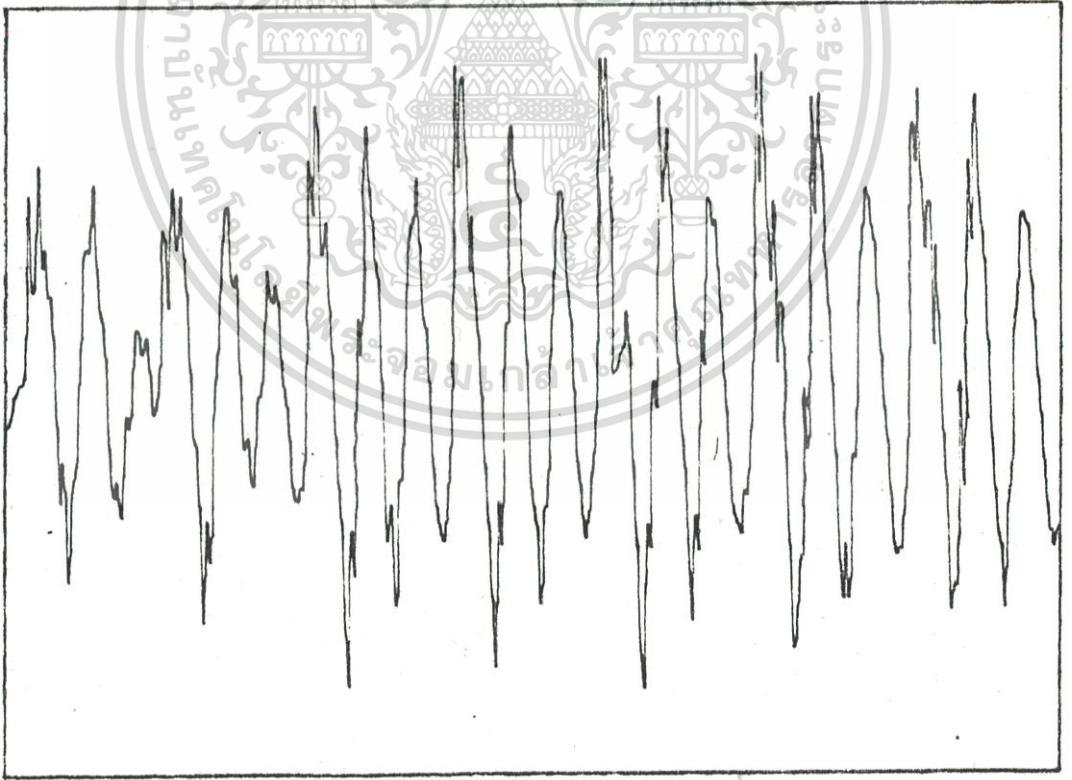


รูปที่ 7.7 แสดงฟังก์ชันอัตโนมัติสัมพันธ์ของสัญญาณกระตุ้น

1. สัญญาณออกจากวงจรรองความถี่แบบความถี่ต่ำผ่านได้แสดงไว้ในรูปที่ 7.4
2. สัญญาณเสียงที่ผ่านการลุ่มสัญญาณเพื่อลดจำนวนสัญญาณลง 5 เท่าให้เหลือสัญญาณ เพียง 80 จุด ได้แสดงไว้ในรูปที่ 7.5
3. สัญญาณกระตุ้นของเสียงพูดได้จากการนำสัญญาณที่ผ่านการลุ่มสัญญาณแล้วผ่านวงจรรองความถี่ที่ได้จากการประมาณพันธะเชิงเส้นได้แสดงไว้ในรูปที่ 7.6
4. พังก์ชันอัตโนมัติสัมพันธ์ของสัญญาณกระตุ้น ซึ่งจะสามารถตรวจพบช่วงกระตุ้นได้โดยง่าย ได้แสดงไว้ในรูปที่ 7.7

7.2 ผลจากการสังเคราะห์เสียงพูด

ผลจากการสังเคราะห์เสียงพูดหรือสัญญาณที่ได้จากระบบกำเนิดเสียงได้แสดงไว้ ในรูปที่ 7.8



รูปที่ 7.8 แสดงสัญญาณที่ได้จากระบบกำเนิดเสียง

7.3 สรุป

ในบทนี้ได้นำผลทั้งหมดที่ได้จากการทดลองมาแสดงโดยคัดเลือกตัวอย่างข้อมูล เพียงชุดเดียวจำนวน 400 จุด ซึ่งใช้ระยะเวลา 40 มิลลิวินาที จากข้อมูลทั้งหมด 1 ค่า ซึ่งมีจำนวนข้อมูลทั้งหมดประมาณ 5000 จุด และใช้เวลาประมาณ 0.5 วินาที เนื่องจากเครื่องบันทึกกราฟ (recorder) ที่ใช้รวบรวมผลการประมวลสัญญาณ ต้องใช้กระดาษบันทึกจำนวนมาก ในการแสดงผลดังกล่าว จึงไม่สะดวกที่จะนำผลที่ได้ทั้งหมดมาแสดงไว้



บทที่ 8

บทสรุป

วัตถุประสงค์ของวิทยานิพนธ์ฉบับนี้คือ การวิเคราะห์พารามิเตอร์ของเสียงพูด หรือคุณลักษณะของระบบกำเนิดเสียงพูดซึ่งเป็นรากฐานที่จำเป็นและสำคัญยิ่งสำหรับการประมวลสัญญาณเสียงพูดในระดับสูงขึ้นไป โดยพารามิเตอร์ที่สำคัญของเสียงพูดมี 2 ส่วนคือ ความถี่ก่าธรของเสียงพูดซึ่งเป็นตัวกำหนดคุณลักษณะของท่อนำเสียง และช่วงกระตุ่นของเสียงพูด ซึ่งเป็นตัวที่ทำให้เสียงมนุษย์แต่ละคนเกิดน้ำเสียงแตกต่างกัน ตลอดจนแสดงอารมณ์ของผู้พูด

งานวิจัยที่กล่าวไว้ในวิทยานิพนธ์ฉบับนี้เป็นขั้นตอนแรกของการวิจัยที่เกี่ยวกับการประมวลสัญญาณเสียงพูด ในคณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้า วิทยาเขตเจ้าคุณทหารลาดกระบัง จึงทำให้เนื้อหาส่วนใหญ่ของวิทยานิพนธ์ฉบับนี้เกี่ยวข้องกับการศึกษาทฤษฎีต่าง ๆ และการทดลองเพื่อตรวจสอบและสร้างความเข้าใจเกี่ยวกับทฤษฎีต่าง ๆ ที่น่าสนใจ เพื่อเป็นการวางรากฐานและสร้างแนวทางสำหรับผู้ที่ทำงานวิจัยขั้นต่อไป การทดลองตามขั้นตอนต่าง ๆ จะพบว่าอุปกรณ์หลักที่ใช้มีเพียงดิจิตอลคอมพิวเตอรฺ์ อุปกรณ์ส่งสัญญาณ และอุปกรณ์ที่ใช้แปลงสัญญาณอนาล็อก เป็นสัญญาณดิจิตอลเท่านั้น ส่วนเครื่องมืออื่น ๆ จะใช้โปรแกรมต่าง ๆ ทำหน้าที่แทนทั้งสิ้น เนื่องจากงานวิจัยขณะนี้ เป็นเพียงขั้นตอนแรก ซึ่งใช้ในการทดสอบทฤษฎีต่าง ๆ ยังไม่จำเป็นต้องจัดหาอุปกรณ์อื่น ๆ ที่มีราคาแพง และไม่จำเป็นในขณะนี้ อีกทั้งดิจิตอลคอมพิวเตอรฺ์สามารถนำมาใช้ทำงานแทนอุปกรณ์ใด ๆ ก็ได้ ถ้าการเขียนโปรแกรมมีประสิทธิภาพเพียงพอ จะแตกต่างกันบ้างก็เพียง เวลาที่ใช้ในการประมวลผลเท่านั้น

สาเหตุที่เลือกใช้การประมวลผลขั้นเชิงเส้น ในการวิเคราะห์คุณลักษณะของเสียงพูด เนื่องจากทฤษฎีดังกล่าว เข้าใจง่ายเมื่อเปรียบเทียบกับวิธีอื่นสำหรับผู้เริ่มศึกษาและผลที่ได้รับสามารถไปใช้กับการวิจัยอื่น ๆ ที่เกี่ยวข้อง เช่นการจำคำพูด และการจำเสียงพูดได้ดีกว่าแบบอื่นจากผลการทดลองขณะทำการวิจัยมีสิ่งที่น่าสนใจ ดังนี้

1. การตรวจสอบคุณภาพของเสียงในบทที่ 4 ส่วนใหญ่จะได้ผลดี สำหรับคำพยางค์เดียวที่ไม่มีเสียงควบกล้ำ และเป็นเสียงโห่หะแท้ ๆ คือไม่เป็นเสียงเสียดสีโห่หะหรือเสียงนาสิก สิ่งที่น่า

ทดลองต่อไปคือการนำผลที่ได้ในบทนี้ไปเปรียบเทียบกับผลที่ได้จากสเปคโตรแกรม

2. การหาช่วงกระตุ้นของเสียงพูดในบทที่ 5 ได้ผลคืออำนาจของสัญญาณใหญ่เพียงพอสำหรับเสียงโฆชะทุกประเภท และเสียงอโฆชะบางประเภท เช่น เสียงเสียดสือโฆชะ และเสียงนาสิก เนื่องจากสัญญาณที่มีขนาดต่ำมักจะมีสัญญาณรบกวนปะปนทำให้การคำนวณต่าง ๆ ผิดพลาด การแก้ไขที่ทำได้ในขณะนี้คือ ใช้วงจรรองความถี่แบบความถี่ต่ำผ่านกำจัดสัญญาณรบกวนจากภายนอกซึ่งสามารถแก้ไขจุดบกพร่องดังกล่าวได้บ้าง แต่ยังไม่สามารถกำจัดได้ทั้งหมด จุดบกพร่องอีกประการหนึ่ง ที่เกิดขึ้นคือ วิธี SIFT ที่ใช้ในการหาช่วงกระตุ้นของเสียงพูดไม่สามารถแยกประเภทของเสียงว่าเป็นเสียงโฆชะหรืออโฆชะได้ ถ้าชุดข้อมูลที่วิเคราะห์เป็นช่วงเวลาที่เกิดการเปลี่ยนแปลงประเภทของเสียงพอดี ซึ่งการแก้ไขอาจทำได้ 2 วิธี คือการลดจำนวนข้อมูลที่วิเคราะห์เพื่อหาช่วงกระตุ้นของเสียงพูดลง เพื่อให้การวิเคราะห์ละเอียดขึ้น แต่จะทำให้การประมวลผลต่าง ๆ ใช้เวลามากขึ้นหรือพยายามหาวิธีวิเคราะห์อย่างอื่นมาช่วยเมื่อเกิดกรณีดังกล่าวขึ้น

3. การเลียนเสียงพูดในบทที่ 6 ขณะนี้ได้ผลดีสำหรับคำที่มีการออกเสียงง่าย ๆ เป็นเสียงโตด ๆ และไม่เป็นคำควบกล้ำโดยเฉพาะ เสียงสระและการเก็บตัวอย่างสัญญาณจะต้องทำอย่างละเอียดรอบคอบ ไม่มีสัญญาณรบกวนแทรกเข้ามามากนัก ขนาดของสัญญาณต้องมีขนาดใหญ่พอเพื่อเปรียบเทียบกับสัญญาณรบกวนที่ปนอยู่ในขณะเลียนเสียงพูด สัญญาณเสียงก่อนจะส่งออกมาที่ลำโพง ควรผ่านวงจรรองความถี่แบบความถี่ต่ำผ่านก่อน เพื่อกำจัดสัญญาณความถี่สูงออกให้หมด

สำหรับงานวิจัยที่ควรทำต่อไป คือพยายามหาอุปกรณ์อื่นมาทดแทนขั้นตอนการประมวลผลต่าง ๆ ซึ่งในขณะนี้ใช้โปรแกรมต่าง ๆ ทำหน้าที่อยู่บางส่วนหรือทั้งหมด เพื่อลดเวลาต่าง ๆ ในการทำงานให้เหลือน้อยลง เนื่องจากในขณะทำการเลียนเสียงพูด 1 คำ ซึ่งเป็นระยะเวลาประมาณ 0.5 วินาที จะต้องใช้เวลาในการประมวลต่าง ๆ ตั้งแต่เริ่มต้นวิเคราะห์สัญญาณเสียง จนกำเนิดเสียงพูดขึ้นใหม่ประมาณ 10 นาที ซึ่งใช้เวลานานมาก ยังไม่สามารถนำไปใช้ประโยชน์ได้ ส่วนงานวิจัยในแขนงอื่นที่เกี่ยวข้องและควรทำต่อไปมี 2 แขนง คือ

1. การจำคำพูด เพื่อช่วยให้เครื่องคอมพิวเตอร์สามารถตอบสนองต่อคำพูดได้ ซึ่งเมื่อทำงานร่วมกับโปรแกรมแปลภาษา (compiler) จะทำให้คอมพิวเตอร์สามารถโต้ตอบกับมนุษย์ได้โดยตรง หรืออาจใช้ช่วยเหลือคนพิการให้ทำงานต่าง ๆ ได้

2. การจำเสียงพูด เพื่อช่วยในการตรวจสอบลักษณะเฉพาะของเสียงพูดแต่ละคน ซึ่งประโยชน์ส่วนใหญ่มักใช้ในระบบรักษาความปลอดภัย

อย่างไรก็ตาม งานที่ควรทำต่อไปควรจะเป็นการเพิ่มความเร็วในการประมวลผลเสียก่อนเป็นลำดับแรก เพราะถึงแม้ว่าระบบการทำงานอื่นจะได้ผลดีเพียงไรก็ตาม ถ้าการทำงานต่าง ๆ ยังคงต้องใช้เวลามากอยู่ ก็ไม่สามารถนำผลงานต่าง ๆ ที่ได้จากการค้นคว้าไปใช้ประโยชน์ได้



บทความและหนังสืออ้างอิง

1. B.S. Atal and Suzanne L. Hanauer, 'Speech Analysis and Synthesis by Linear Prediction of the Speech Wave', J. Acoust. Soc. Am., vol 50, pp. 637-655, Aug. 1971.
2. J.D. Markel and A.H. Gray. 'Linear Prediction of Speech', Springer-Verlag, 1976.
3. J.D. Markel and A.H. Gray, 'A Linear Prediction Vocoder Simulation Based upon the Autocorrelation Method', IEEE Trans. Acoust., Speech and Signal Process., vol ASSP-22, pp. 124-134, Apr, 1974.
4. J.D. Markel, 'The SIFT Algorithm for Fundamental Frequency Estimation', IEEE Trans. Audio Electroacoust., vol AU-20, pp. 367 - 377, Dec. 1972.
5. J.L. Flanagan, 'Voices of Men and Machines', J. Acoust. Soc. Am., vol 51, pp. 1375-1387, Mar. 1972.
6. L.R. Rabiner, M.J. Cheng, A.E. Rosenberg and C.A. McGonegal, 'A Comparative Performance Study of Several Pitch Detection Algorithms', IEEE Trans. Acoust., Speech and Signal Process., vol ASSP-24, pp. 399-417, Oct. 1976.
7. Nobuhiko Kitawaki, Kenzo Itoh and Fumitada Itakura, 'PARCOR Speech Analysis-Synthesis System', Review of the Electrical Communication Laboratories, vol 26, NOS 11-12, Nov-Dec, 1978.
8. R.W. Schafer and L.R. Rabiner, 'Digital Representation of Speech Signal', Proc. IEEE, vol 63, pp. 662-677, Apr. 1975.
9. The Engineering Staff of Texas Instruments Incorporated Semiconductor Group, 'TMS 5200 Voice Synthesis Processor Data Manual', Texas Instruments Incorporated, 1980.

ภาคผนวก

สำหรับภาคผนวก เป็นการแสดงโปรแกรมที่ใช้ในการประมวลผลต่าง ๆ ดังนี้

1. โปรแกรมย่อย AUTO
2. โปรแกรมย่อย COVAR
3. โปรแกรมย่อย FFT
4. โปรแกรมย่อย ROWDAT
5. โปรแกรมย่อย FORMNT
6. โปรแกรมย่อย SIFT
7. โปรแกรม SYN

โปรแกรมย่อย AUTO

โปรแกรมย่อย AUTO ใช้วิเคราะห์หาสัมประสิทธิ์ของวงจรรองความถี่แบบดีจิตอล และสัมประสิทธิ์ PARCOR โดยการประมาณฟังก์ชันเชิงเส้น ด้วยวิธีอัลกอริทึมปริมาตรและแยกดังนี้

1. พารามิเตอร์ที่โปรแกรมย่อยต้องการ

- N จำนวนสัญญาณที่ต้องการวิเคราะห์
- X(N) สัญญาณที่ต้องการวิเคราะห์
- M อันดับของวงจรรองความถี่ที่ต้องการ

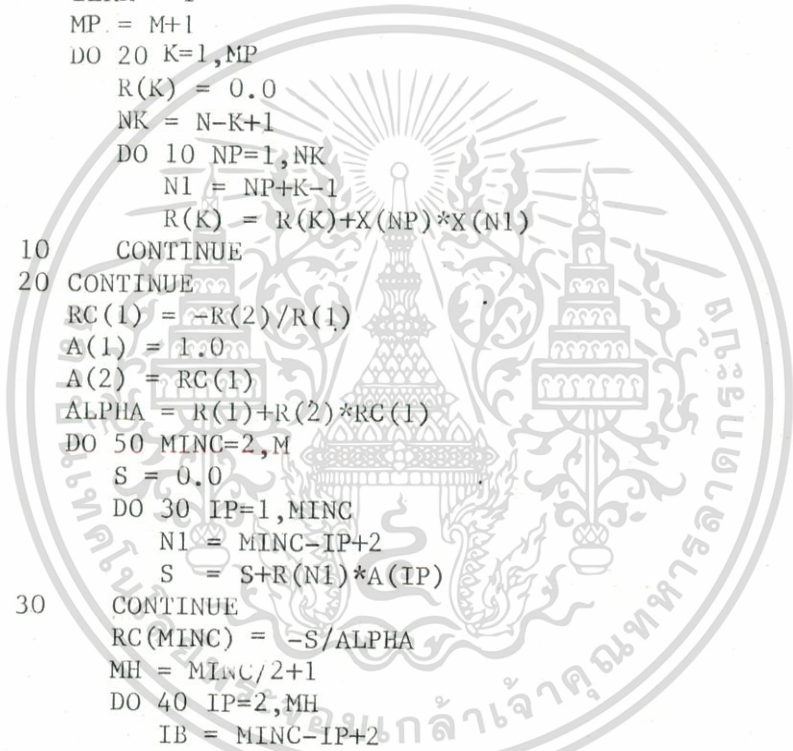
2. ผลจากโปรแกรมย่อย

- A(M+1) สัมประสิทธิ์ของวงจรรองความถี่แบบดีจิตอล
- ALPHA ความคลาดเคลื่อนกำลังสองรวมที่เกิดจากการประมาณฟังก์ชันเชิงเส้น
- RC(M) สัมประสิทธิ์ PARCOR ของระบบกำเนิดเสียง
- IERR เลขดัชนีบอกผลวิเคราะห์ ถ้า
 - IERR = 0 การวิเคราะห์เสร็จสมบูรณ์
 - IERR = 1 ข้อมูลที่ใช้ไม่สมบูรณ์

C
C-----
C
C
C
C-----
C

SUBROUTINE AUTO

```
SUBROUTINE AUTO(N,X,M,A,ALPHA,RC,IERR)
REAL X(1),A(1),RC(1)
REAL R(21)
IERR = 1
MP = M+1
DO 20 K=1,MP
  R(K) = 0.0
  NK = N-K+1
  DO 10 NP=1,NK
    N1 = NP+K-1
    R(K) = R(K)+X(NP)*X(N1)
10  CONTINUE
20  CONTINUE
  RC(1) = -R(2)/R(1)
  A(1) = 1.0
  A(2) = RC(1)
  ALPHA = R(1)+R(2)*RC(1)
  DO 50 MINC=2,M
    S = 0.0
    DO 30 IP=1,MINC
      N1 = MINC-IP+2
      S = S+R(N1)*A(IP)
30  CONTINUE
    RC(MINC) = -S/ALPHA
    MH = MINC/2+1
    DO 40 IP=2,MH
      IB = MINC-IP+2
      AT = A(IP)+RC(MINC)*A(IB)
      A(IB) = A(IB)+RC(MINC)*A(IP)
      A(IP) = AT
40  CONTINUE
    A(MINC+1) = RC(MINC)
    ALPHA = ALPHA+RC(MINC)*S
    IF(ALPHA.LE.0.0) RETURN
50  CONTINUE
  IERR = 0
  RETURN
END
```



โปรแกรมย่อย COVAR

โปรแกรมย่อย COVAR ใช้วิเคราะห์หาสัมประสิทธิ์ของวงจรรองความถี่แบบดิจิทัล และสัมประสิทธิ์ PARCOR โดยการประมาณพหุคูณเชิงเส้นด้วยการหาค่าความแปรปรวนร่วม มีรายละเอียดดังนี้

1. พารามิเตอร์ที่โปรแกรมย่อยต้องการ

- N จำนวนสัญญาณที่ต้องการวิเคราะห์
- X(N) สัญญาณที่ต้องการวิเคราะห์
- M อันดับของวงจรรองความถี่ที่ต้องการหา

2. ผลจากโปรแกรมย่อย

- A(M+1) สัมประสิทธิ์ของวงจรรองความถี่
- ALPHA ความคลาดเคลื่อนกำลังสองรวมที่เกิดจากการประมาณพหุคูณเชิงเส้น
- GRC(M) สัมประสิทธิ์ PARCOR ของระบบกำเนิดเสียง
- IERR เลขชี้ชี้บอกผลวิเคราะห์
 - IERR = 0 การวิเคราะห์สมบูรณ์
 - IERR = 1 สัญญาณที่ใช้วิเคราะห์ไม่สมบูรณ์

C
C-----
C
C SUBROUTINE COVAR
C-----
C

```
SUBROUTINE COVAR(N,X,M,A,ALPHA,GRC,IERR)
REAL X(1),A(1),GRC(1)
REAL B(190),BETA(20),CC(21)
IERR = 1
MP = M+1
MT = MP*M/2
DO 10 J=1,MT
    B(J) = 0.0
10 CONTINUE
ALPHA = 0.0
CC(1) = 0.0
CC(2) = 0.0
DO 20 NP=MP,N
    NP1 = NP-1
    ALPHA = ALPHA+C(NP)*X(NP)
    CC(1) = CC(1)+X(NP)*X(NP1)
    CC(2) = CC(2)+X(NP1)*X(NP1)
20 CONTINUE
B(1) = 1.0
BETA(1) = CC(2)
GRC(1) = -CC(1)/CC(2)
A(1) = 1.0
A(2) = GRC(1)
ALPHA = ALPHA+GRC(1)*CC(1)
MF = M
DO 130 MINC=2,MF
    DO 30 J=1,MINC
        JP = MINC+2-j
        N1 = MP+1-JP
        N2 = N+1-MINC
        N3 = N+2-JP
        N4 = MP-MINC
        CC(JP) = CC(JP-1)+X(N4)*X(N1)-X(N2)*X(N3)
30 CONTINUE
CC(1) = 0.0
DO 40 NP=MP,N
    N1 = NP-MINC
    CC(1) = CC(1)+X(N1)*X(NP)
40 CONTINUE
MSUB = (MINC*MINC-MINC)/2
MM1 = MINC-1
N1 = MSUB+MINC
B(N1) = 1.0
DO 80 IP=1,MM1
```



```
ISUB = (IP*IP-IP)/2
IF (BETA(ISUB).LE.0.0) RETURN
50  GAM = 0.0
    DO 60 J=1,IP
        N1 = ISUB+J
        GAM = GAM+CC(J+1)*B(N1)
60  CONTINUE
    GAM = GAM/BETA(IP)
    DO 70 J=1,IP
        N1 = MSUB+JP
        N2 = ISUB+JP
        B(N1) = B(N1)-GAM*B(N2)
70  CONTINUE
80  CONTINUE
    BETA(MINC) = 0.0
    DO 90 J=1,MINC
        N1 = MSUB+J
        BETA(MINE) = BETA(MINE)+CC(J+1)*B(N1)
90  CONTINUE
    IF (BETA(MINC).LE.0.0) RETURN
100 S = 0.0
    DO 110 IP=1,MINC
        S = S+CC(IP)*A(IP)
110 CONTINUE
    GRC(MINC) = -S/BETA(MINC)
    DO 120 IP=2,MINC
        M2 = MSUB+IP-1
        A(IP) = A(IP)+GRC(MINC)*BETA(MINC)
120 CONTINUE
    A(MINC+1) = GRC(MINC)
    S = GRC(MINC)*GRC(MINC)*BETA(MINC)
    ALPHA = ALPHA-S
    IF (ALPHA.LE.0.0) RETURN
130 CONTINUE
    IERR = 0
    RETURN
    END
```

โปรแกรมย่อย FFT

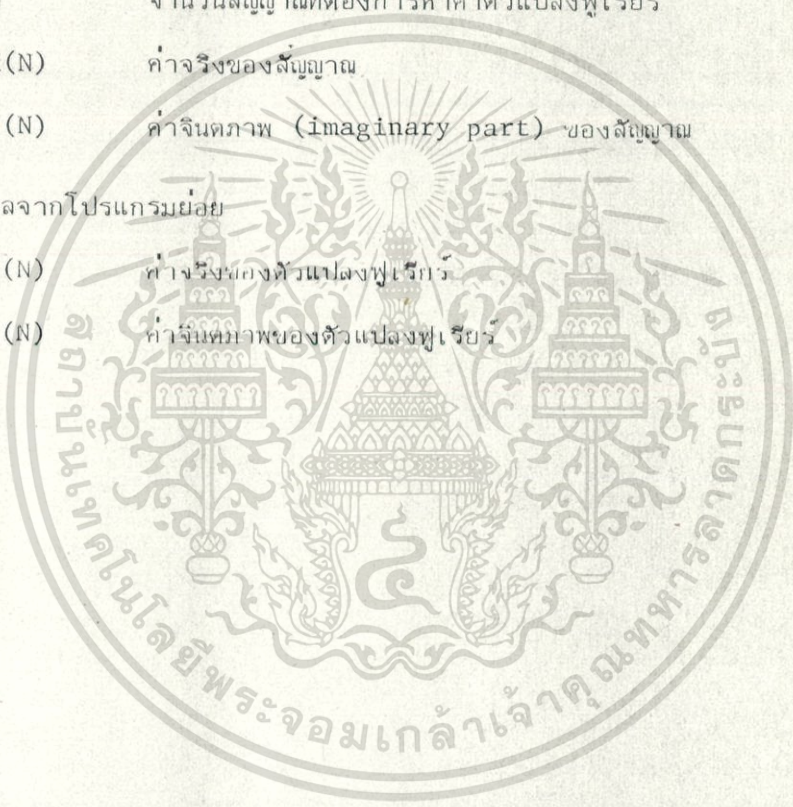
โปรแกรมย่อย FFT ใช้หาค่าตัวแปลงฟูเรียร์ (Fourier Transform) โดยใช้เทคนิค FFT (Fast Fourier Transform) มีรายละเอียดดังนี้

1. พารามิเตอร์ที่โปรแกรมย่อยต้องการ

- N จำนวนสัญญาณที่ต้องการหาค่าตัวแปลงฟูเรียร์
- X(N) ค่าจริงของสัญญาณ
- Y(N) ค่าจินตภาพ (imaginary part) ของสัญญาณ

2. ผลจากโปรแกรมย่อย

- X(N) ค่าจริงของตัวแปลงฟูเรียร์
- Y(N) ค่าจินตภาพของตัวแปลงฟูเรียร์



C
C-----
C
C RADIX-2 FAST FOURIER TRANSFORM SUBROUTINE
C-----
C

SUBROUTINE FFT (X,Y,N)
REAL X(1),Y(1)

C
C
C CALCULATE THE NUMBER OF ITERATIONS (LOG. N TO THE BASE 2)
C

```
L = 1
NP = 1
10 NP = NP*2
IF (NP.GE.N) GOTO 20
L = L+1
GOTO 10
20 NP1 = N+1
DO 25 I=NP1,NP
  X(I) = 0.0
  Y(I) = 0.0
25 CONTINUE
LMX = NP
SCL = 6.283185/FLOAT(NP)
DO 50 LO=1,L
  LIX = LMX
  LMX = LMX/2
  ARG = 0.0
  DO 40 LM=1,LMX
    C = COS(ARG)
    S = SIN(ARG)
    ARG = ARG+SCL
    DO 30 LI=LIX,NP,LIX
      J1 = LI-LIX+LM
      J2 = J1-LMX
      T1 = X(J1)-X(J2)
      T2 = Y(J1)-Y(J2)
      X(J1) = X(J1)+X(J2)
      Y(J1) = Y(J1)+Y(J2)
      X(J2) = C*T1+S*T2
      Y(J2) = C*T2-S*T1
    30 CONTINUE
  40 CONTINUE
  SCL = 2.0*SCL
50 CONTINUE
```

C
C
C
C
C

BIT REVERSAL

```
J = 1
NY2 = NP/2
NPM1 = NP1
DO 90 I=1,NPM1
  IF (I.GE..J) GOTO 60
  T1 = X(J)
  T2 = Y(J)
  X(J) = X(I)
  Y(J) = Y(I)
  X(I) = T1
  Y(I) = T2
60  K = NY2
  IF (K.GE..J) GOTO 80
  J = J-K
  K = K/2
  GOTO 70
80  J = J+K
90 CONTINUE
RETURN
END
```



โปรแกรมย่อย ROWDAT

โปรแกรมย่อย ROWDAT ใช้หาข้อมูลดิบของความถี่กำลัง มีรายละเอียดดังนี้

1. พารามิเตอร์ที่โปรแกรมย่อยต้องการ

NPTS จำนวนข้อมูลดิบ มีค่า N/2 จากโปรแกรมย่อย FFT
 ค่าแอมพลิจูดของสัญญาณเสียง จากโปรแกรมย่อย FFT มีค่า

$$F(i) = X^2(i) + Y^2(i)$$

FSK ความถี่ที่ใช้ในการส่งสัญญาณ

2. ผลจากโปรแกรมย่อย

NPEAK จำนวนยอดคลื่นที่พบในข้อมูลดิบ

IX(3*NPEAK) ข้อมูลดิบของสัญญาณ

$$IX(i) = \hat{F}_i$$

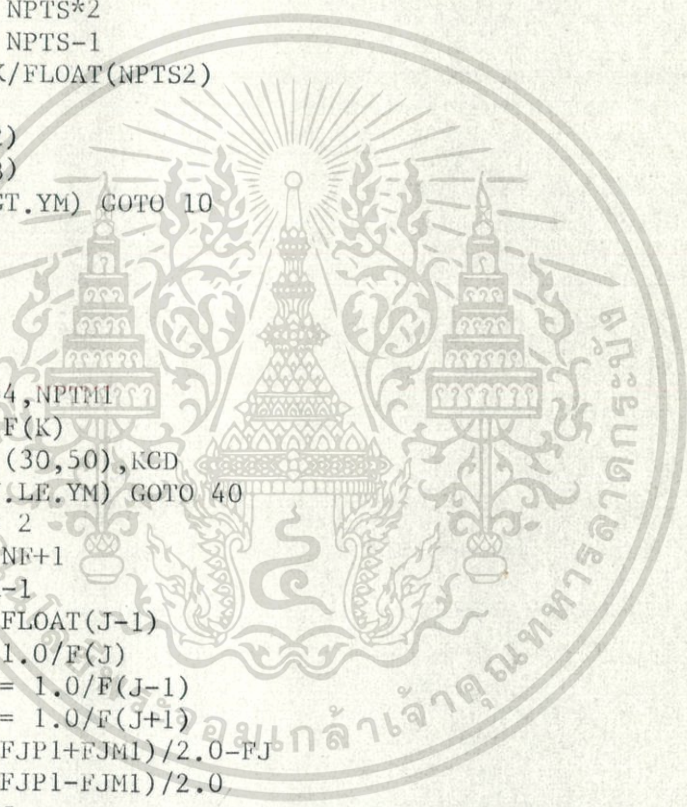
$$IX(i+1) = \hat{B}_i$$

$$IX(i+2) = \hat{P}_i$$

C
C
C
C
C
C
C
C
C

THIS SUBROUTINE IS USED TO PERFORM PEAK PICKING AND
INTERPOLATE THE FILTER POWER SPECTRUM TO FIND ROW DATA

```
SUBROUTINE ROWDAT(F,NPTS,FSK,NPEAK,IX)
REAL F(1)
INTEGER IX(1)
JOO = 0
NPEAK = 0
NPTS2 = NPTS*2
NPTM1 = NPTS-1
TS = FSK/FLOAT(NPTS2)
NF = 0
YM = F(2)
YP = F(3)
IF (YP.GT.YM) GOTO 10
KCD = 1
YM = YP
GOTO 20
10 KCD = 2
YM = YP
20 DO 60 K=4,NPTM1
  YP = F(K)
  GOTO (30,50),KCD
30 IF (Y.LE.YM) GOTO 40
  KCD = 2
  NF = NF+1
  J = K-1
  FN = FLOAT(J-1)
  FJ = 1.0/F(J)
  FJM1 = 1.0/F(J-1)
  FJP1 = 1.0/F(J+1)
  A = (FJP1+FJM1)/2.0-FJ
  B = (FJP1-FJM1)/2.0
  C = FJ
  D = -B/(2.0*A)
  E = D*(A*D+B)+C
  G = FN+D
  FI = TS*G
  BW = -SORT(B*B-4.0*A*(C-0.5*E))*TS/A
  PK = E
  IX(JOO+1) = IFIX(FI)
  IX(JOO+2) = IFIX(BW)
```



```
IX(J00+3) = IFIX(434.2945*ALOG(PX))
J00 = J00+3
NPEAK = NPEAK+1
40  YM = YP
    GOTO 60
50  IF (YP.LT.YM) KCD = 1
    YM = YP
60  CONTINUE
    RETURN
    END
```



โปรแกรมย่อย FORMNT

โปรแกรมย่อย FORMNT ใช้ตัดสิน เลือกค่าความถี่กำลังจากข้อมูลดิบที่ได้จากโปรแกรมย่อย ROWDAT มีรายละเอียดดังนี้

1. พารามิเตอร์ที่โปรแกรมย่อยต้องการ

NPEAK จำนวนยอดคลื่นที่ตรวจพบจากโปรแกรมย่อย ROWDAT

IX(3*NPEAK) ข้อมูลดิบจากโปรแกรมย่อย

IF1(3) ความถี่กำลังที่ได้จากข้อมูลชุดก่อน

IUV เลขดัชนีแยกประเภทของเสียงพูดชุดปัจจุบัน

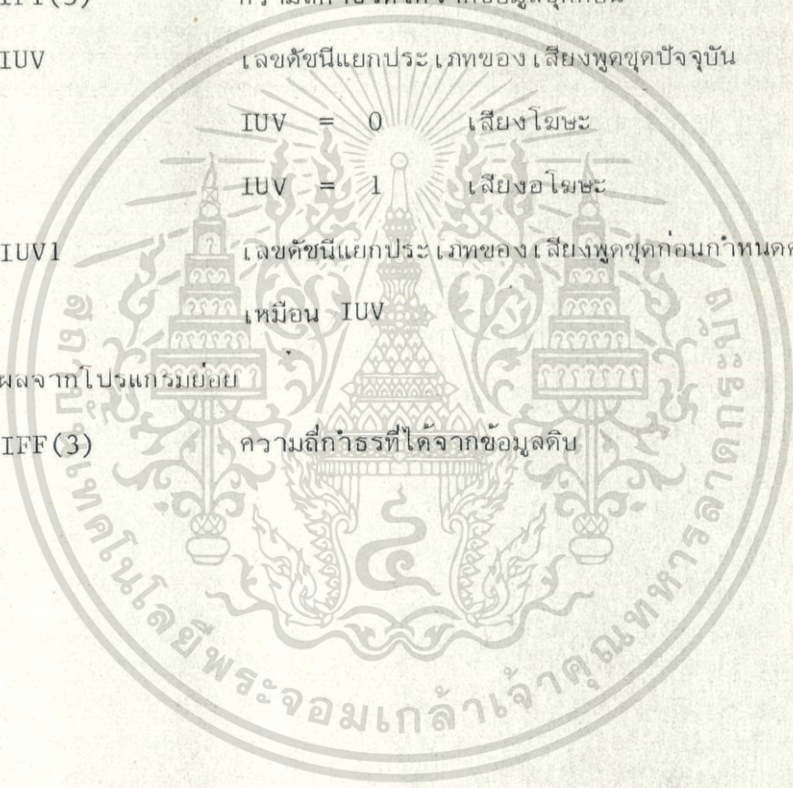
IUV = 0 เสียงโฆษะ

IUV = 1 เสียงอโฆษะ

IUV1 เลขดัชนีแยกประเภทของเสียงพูดชุดก่อนกำหนดค่าเหมือน IUV

2. ผลจากโปรแกรมย่อย

IFF(3) ความถี่กำลังที่ได้จากข้อมูลดิบ



C
C
C
C
C
C
C
C
C

THIS SUBROUTINE IS USED TO PERFORMING FORMANT
ESTIMATION ON ROW DATA

C
C
C
C
C

SUBROUTINE FORMNT(IX, IFF, IF1, IUV, IUV1, NPEAK, IFC)
INTEGER IP (8), IX(1), IFF(1), IF1(1), IV(10,10)
IF (NPEAK.EQ.0) GOTO 10
IHLF = 500

C
C
C
C
C

IS FRAME VOICED ?

IF (IUV.EQ.1) GOTO 20
10 IFF(1) = 0
 IFF(2) = 0
 IFF(3) = 0
 GOTO 210
20 L = 0

C
C
C
C
C

OBTAIN PEAK FOR THIS FRAME

DO 30 J=1, NPEAK
 I = 3*J-2
 IF (IX(I) .GT. 3000) GOTO 30
 IF (IX(I) .LE. 0) GOTO 30
 IF (IX(I+1) .GT. 500) GOTO 30
 L = L+1
 IP(L) = IX(I)
30 CONTINUE
 DO 35 I=1, L
 DO 35 J=1, L
 IV(I, J) = IABS(IP(I)-IF1(J))
35 CONTINUE

DETERMINE F1-F3 FOR PRESENT FRAME

IF (L.EQ.3) GOTO 40
IF (L.EQ.1) GOTO 60
GOTO 100

C
C
C
C
C

WHEN L = 3

```

40 DO 50 J=1,3
    IFF(J) = IP(J)
50 CONTINUE
    GOTO 210

```

C
C
C
C
C

WHEN L = 1

```

60 DO 70 J=1,3
    IFF(J) = IF1(J)
70 CONTINUE
    IF (IV(1,1).GT.IV(2,1)) GOTO 80
    IFF(1) = IP(1)
    GOTO 210
80 IF (IV(3,1).GT.IV(2,1)) GOTO 90
    IFF(2) = IP(1)
    GOTO 210
90 IFF(3) = IP(1)
    GOTO 210

```

C
C
C
C
C

WHEN L = 2 OR L > 3

```

100 IF (IV(1,1).LT.IV(1,2)) GOTO 120
    IF (IP(2)-IP(1).LE.IHLF) GOTO 110
    ICD = 2
    IFF(2) = IP(1)
    IFF(1) = IF1(1)
    GOTO 170
110 IFF(1) = IP(1)
    IFF(2) = IP(2)
    ICD = 3
    GOTO 170
120 IFF(1) = IP(1)
    IF (L.EQ.2) GOTO 150
    IF (IV(2,2).GT.IV(3,2)) GOTO 130
    IF (IV(2,2).LT.IV(2,3)) GOTO 140
130 IFF(2) = IP(3)
    IFF(3) = IP(4)
    GOTO 210
140 IFF(2) = IP(2)
    ICD = 3
    GOTO 180

```

150 IF(IV(2,2).LT IV(2,3)) GOTO 160
IFF(3) = IP(2)
IFF(2) = IF1(2)
GOTO 210
160 IFF(2) = IP(2)
IFF(3) = IF1(3)
GOTO 210
170 IF (ICD.GT.L) GOTO 200
IF (ICD.EQ.L) GOTO 190
180 IF (IV(ICD,3).LT.IV(ICD+1,3)) GOTO 190
IFF(3) = IP(ICD+1)
GOTO 210
190 IFF(3) = IP(ICD)
GOTO 210
200 IFF(3) = IF1(3)
210 NPEAK = L
RETURN
END



โปรแกรมย่อย SIFT

โปรแกรมย่อย SIFT ให้หาช่วงกระตุ้นของเสียงพูด มีรายละเอียดดังนี้

1. พารามิเตอร์ที่โปรแกรมย่อยต้องการ

SPCH(400) สัญญาณเสียงพูด

2. ผลจากโปรแกรมย่อย

PITCH(3) ช่วงกระตุ้นของเสียงพูดชุดปัจจุบันและ 2 ชุดก่อน

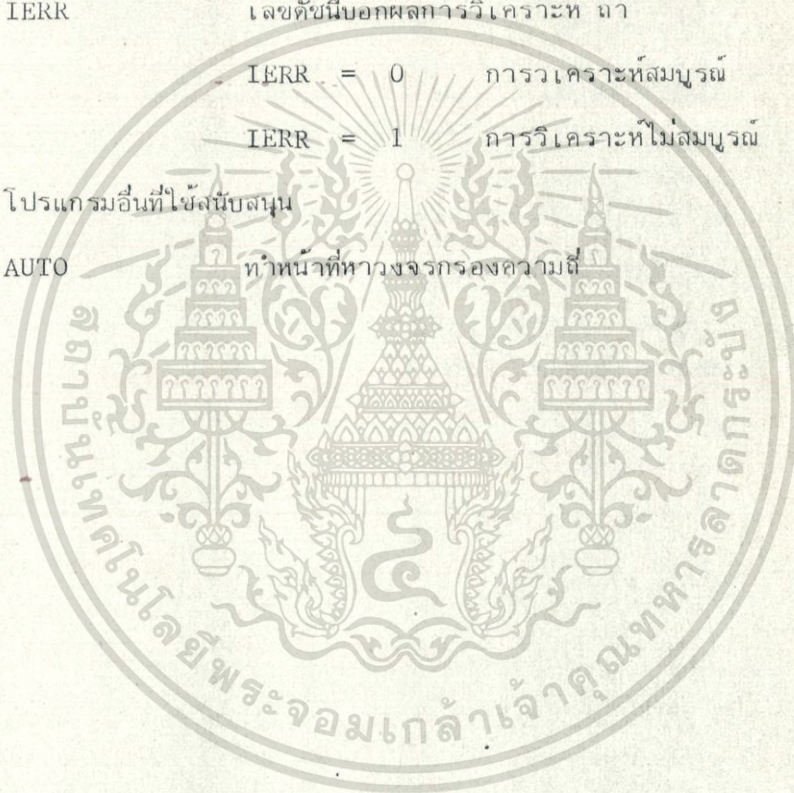
IERR เลขดัชนีบอกผลการวิเคราะห์ที่ น่า

IERR = 0 การวิเคราะห์สมบูรณ์

IERR = 1 การวิเคราะห์ไม่สมบูรณ์

3. โปรแกรมอื่นที่ใช้สนับสนุน

AUTO ทำหน้าที่หาวงจรรองความถี่



C
C
C
C
C
C
C
C
C

SIFT ALGORITHM PROCESSING FOR PERFORM PITCH DETECTOR.
NOTE: PARAMETERS FIXED FOR SAMPLING FREQUENCY 10 kHz.

C
C
C
C
C

SUBROUTINE SIFT (APCH,PITCH,IERR)
REAL SPCH(400),SOUT(400)
REAL A(21),RC(20),U(80),PBUF (80)
REAL AF(4),PF(4),PITCH(3),ABUF(33)
DATA AF/1.000000,-2.340366,2.011900,-0.614109/
DATA PF/0.035708,-0.006996,-0.006996,0.035708/

PRE-FILTER BY LOWPASS DIGITAL FILTER

SOUT(1) = PF(1)*SPCH(1)/AF(1)
DO 20 I=2,400
SOUT(I) = PF(1)*SPCH(1)
DO 10 J=2,4
K = I-J+1
IF (K.GE.1) SOUT(I) = SOUT(I)+PF(J)*SPCH(K)
-AF(J)*SOUT(K)
10¹ CONTINUE
SOUT(I) = SOUT(I)/AF(I)
20 CONTINUE

C
C
C
C
C

DOWN-SAMPLER,DIFFERENCER AND HAMMING WINDOWER

PBUF(1) = SOUT(5)
U(1) = 0.08*SOUT(5)
DO 30 I=2,80
ID = I*5
IDP = ID-5
PBUF(I) = SOUT(ID)
U(I) = (SOUT(ID)-SOUT(IDP))
*(0.54-0.46*COS(FLOAT(I-1)*6.28318/79.0))
30¹CONTINUE

C
C
C
C
C

COMPUTE INVERSE FILTER COEFFICIENTS
BY AUTOCORRELATION METHOD.

CALL AUTO(80,U,5,A,ALPHA,RC,IERR)
IF (IERR.NE.0) RETURN

C
C
C
C
C

PERFORM INVERSE FILTERING TO FIND ERROR SIGNALS.

```

U(1) = PBUF(1)
DO 50 I=2,80
  U(I) = PBUF(I)
  DO 40 J=2,6
    K = I-J+1
    IF (K.GE.1) U(I) = U(I)+A(J)*PBUF(K)
40  CONTINUE
50  CONTINUE

```

C
C
C
C
C

PERFORM HAMMING WINDOW

```

DO 60 I=5,80
  PBUF(I-4) = U(I)*(0.54-0.46
    *COS(FLOAT(I-5)*6.28318/75.0))
601 CONTINUE

```

C
C
C
C
C

PERFORM AUTOCORRELATION ON PITCH BUFFER "PBUF"

```

DO 80 J=1,33
  JM = J-1
  NMJ = 76-JM
  SUM = 0.0
  DO 70 I=1,NMJ
    IPJ = I+JM
    SUM = SUM+PBUF(I)*PBUF(IPJ)
70  CONTINUE
  ABUF(J) = SUM
80  CONTINUE

```

C
C
C
C
C

OBTAIN PITCH VALUES FROM LAST 3 FRAMES

```

P1 = PITCH(1)
P2 = PITCH(2)
P3 = PITCH(3)

```

C
C
C
C
C

GET PEAK WITHIN RANGE 6 to 32

```

L = 6
AMAX = ABUF(L)
DO 90 J=6,32
  IF (ABUF(J).LE.AMAX) GOTO 90
  AMAX = ABUF(J)
  L = J
90 CONTINUE

```

C
C
C
C
C

TEST FOR MAXIMUM PEAK = 0

IF (AMAX.EQ.0,0) GOTO 120

C
C
C
C
C

TEST FOR LEFT HAND EDGE

IF ABUF(L) IS NOT A PEAK, SET,UNVOICBD

IF (ABUF(L).LT.ABUF(L-1)) GOTO 120

C
C
C
C
C

PERFORM PARABOLIC INTERPOLATION ABOUT LOCATION L

AA = (ABUF(L+1)-2.0*ABUF(L)+ABUF(L-1))/2.0

BB = (ABUF(L+1)-ABUF(L-1))/4.0

AP = ABUF(L)-BB*BB/AA

AL = FLOAT(L)-BB/AA

V = AP/ABUF(L)

C
C
C
C
C

TEST WITH VARIABLE THRESHOLD

IF (L.GE.19) GOTO 100

D = 2.0-FLOAT(L-6)/13.0

GOTO 110

100 D = 1.0-FLOAT(L-19)/130.0

110 V = V/D

C
C
C
C
C

DICISIONS

IF (V.GE.0.35) GOTO 130

IF (P1.EQ.0.0) GOTO 120

IF (V.GE.0.30) GOTO 130

120 P0 = 0.0

GOTO 140

130 P0 = AL

โปรแกรม SYN

โปรแกรม SYN ใช้จำลองระบบกำเนิดด้วยวิธีประมาณพื้นธะเชิงเส้นมีรายละเอียด

ดังนี้

1. พารามิเตอร์ที่โปรแกรมต้องการ

- PE สัมประสิทธิ์การพรีเอมพาซิส
- N จำนวนสัญญาณ 1 ชุด ที่ใช้ในขณะวิเคราะห์
- M อันดับของวงจรกรองความถี่ที่ใช้

2. โปรแกรมอื่นที่ใช้สนับสนุน

- RAN ทำหน้าที่สร้างค่าสุ่ม (random number)



IF (IVLST.EQ.0).OR.(IVL.EQ.1)) GOTO 60

ZERO BUFFER DURING TRANSITION FROM VOICE TO UNVOICE

DO 50 I=1,M
RCBUF(I) = 0.0
50 CONTINUE

GET NEW PARAMETERS
RCR(1),RCR(2),...,RCR(M),SIGMA,IPTCHR
IF END OF FILE GOTO 70

60 NB = 1
70 NB = NB+1
IF (IPTCHR.GT.0).GOTO 80
GAINR = SIGMA*SQRT(3.0/FLOAT(N))*10.0
IVR = 0
GOTO 30
80 GAINR = SIGMA/SQRT(FLOAT(N))
IVR = 1

END OF THE CROSSING FRAME BOUNDARY WORK
PERFORM INTERPOLATION

90 IP ((IVR.EQ.1).AND.(IVL.EQ.1)) GOTO 100
IPITCH = IPTCHL
IF (IVL.EQ.0) IPITCH = IFLGTH-IFC+1
WLC = 0.0
GOTO 110
100 WLC = FLOAT(IFC-1)/FLOAT(IFLGTH-1)
IPITCH = IFIX(WLC*FLOAT(IPTCHR-IPTCHL))+IPTCHL

END OF SETUP FOR A NEW PITCH PERIOD

110 DRV = 1.0
DRVN = -1.0/FLOAT(IPITCH-1)
DO 120 I=1,M
RC(I) = WLC*(RCR(I)-RCL(I))+RCL(I)
120 CONTINUE
GAIN = WLC*(GAINR-GAINL)+GAINL
IF (IVL.EQ.1) GAIN = GAIN*SQRT(FLOAT(IPITCH))

```
130 IF (IVL.EQ.1) GOTO 140
    DRV = SIGN*RAN(1.0)
    SIGN = 0.01*RAN(1.0)
    IF (DRV.GT.0.0) SIGN = -SIGN
140 TAP(1) = GAIN
```

C
C
C
C
C

TWO MULTIPLIER LATTIC SYNTHESIS MODEL

```
YOUT = 0.0
DO 150 I=1,M
    II = M+1-I
    JJ = II+1
    DRV = DRV-RC(II)*RCBUF(II)
    RCBUF(II+1) = RCBUF(II)+RC(II)*DRV
    YOUT = YOUT+RCBUF(II+1)*TAP(II+1)
150 CONTINUE
RCBUF(1) = DRV
YOUT = YOUT+RCBUF(1)*TAP(1)
Y(NN) = YOUT+PE*YPREV
DRV = DRVN
YPREV = Y(NN)
IFC = IFC+1
NN = NN+1
IPC = IPC+1
IF (NN.LE.IFLGTH) GOTO 10
NN = 1
```

C
C
C
C
C
C
C
C
C

PUT OUT IFLGTH SYNTHESIZED SAMPLES

GOTO 10

END OF JOB

```
160 STOP
    END
```