



ใบรับรองวิทยานิพนธ์

บัณฑิตวิทยาลัย สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ

เรื่อง การเพิ่มประสิทธิภาพของกระบวนการตรวจจับค่าพิตช์ของเสียงพูดภาษาไทยเพื่อ
การจำแนกเพศบุคคล

โดย นางสาวนิชประภา วิรัชโกติ

ได้รับอนุมัติให้นับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ครุศาสตร์อุตสาหกรรมมหาบัณฑิต สาขาวิชาเทคโนโลยีคอมพิวเตอร์

คณบดีบัณฑิตวิทยาลัย

(รองศาสตราจารย์ ดร.วิบูลย์ ชื่นแขก)

28 กรกฎาคม 2549

คณะกรรมการสอบวิทยานิพนธ์

ประธานกรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.กัณฑ์พงษ์ วรรณปัญญา)

กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.สุพจน์ นิตย์สุวัฒน์)

กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.พยุ่ง มีสัจ)

การเพิ่มประสิทธิภาพของกระบวนการตรวจจับคำพิตช์ของเสียงพูดภาษาไทย
เพื่อการจำแนกเพศบุคคล

นางสาวนิชประภา วิรัชโกศล

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ครุศาสตรบัณฑิต สาขาวิชาเทคโนโลยีคอมพิวเตอร์ ภาควิชาคอมพิวเตอร์ศึกษา
บัณฑิตวิทยาลัย สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ
ปีการศึกษา 2549
ISBN 974-19-0782-6
ลิขสิทธิ์ของสถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ

ชื่อ : นางสาวนิชประภา วิรัชโกดี
ชื่อวิทยานิพนธ์ : การเพิ่มประสิทธิภาพของกระบวนการตรวจจับคำพิตช์ของเสียงพูด
ภาษาไทยเพื่อการจำแนกเพศบุคคล
สาขาวิชา : เทคโนโลยีคอมพิวเตอร์
สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ
ที่ปรึกษาวิทยานิพนธ์ : ผู้ช่วยศาสตราจารย์ ดร.กนต์พงษ์ วรรณปัญญา
ผู้ช่วยศาสตราจารย์ ดร.สุพจน์ นิตยสุวัฒน์
ปีการศึกษา : 2549

บทคัดย่อ

วิทยานิพนธ์ฉบับนี้นำเสนอการเพิ่มประสิทธิภาพการตรวจจับคำพิตช์ของเสียงพูดภาษาไทย สำหรับการจำแนกเพศบุคคล ในงานวิจัยนี้ได้มุ่งเน้นไปยังขั้นตอนวิธีการประมวลผลสัญญาณเบื้องต้นก่อนกระบวนการตรวจจับคำพิตช์ ทั้งนี้เนื่องจากประสิทธิภาพโดยรวมของการตรวจจับคำพิตช์ขึ้นอยู่กับคุณภาพของสัญญาณนำเข้า ด้วยเหตุผลดังกล่าวขั้นตอนวิธีการประมวลผลสัญญาณเบื้องต้นด้วยวิธีการลบดีซีไบอัส (Removed-DC-Bias) ร่วมกับตัวกรองบัตเตอร์เวิร์ธซึ่งเป็นวิธีการที่ง่ายและมีประสิทธิภาพจึงถูกประยุกต์ใช้กับระบบการจำแนกเพศบุคคล เพื่อพิสูจน์ความสามารถของขั้นตอนวิธีดังกล่าว วิธีออโตคอร์รีเลชัน (Autocorrelation) และวิธีเซปสตรัม (Cepstrum) ซึ่งเป็นสองวิธีการที่มีความน่าเชื่อถือและมีประสิทธิภาพในการประมวลผลบนขอบเขตทางเวลาและขอบเขตทางความถี่ได้ถูกเลือกมาเป็นวิธีในการตรวจจับคำพิตช์ จากนั้นได้นำตัวอย่างสัญญาณเสียงจำนวน 200 เสียงพูดโดยแบ่งออกเป็น 2 กลุ่ม คือ สัญญาณเสียงจำนวน 40 เสียงพูดสำหรับกระบวนการเรียนรู้ และสัญญาณเสียงพูด 160 เสียงพูดสำหรับกระบวนการทดสอบผลลัพธ์ จากการทดลองพิสูจน์ให้เห็นว่าประสิทธิภาพของขั้นตอนวิธีที่นำเสนอเมื่อประยุกต์ใช้งานร่วมกับวิธีออโตคอร์รีเลชันสามารถจำแนกเพศบุคคลให้มีความถูกต้องโดยรวมเพิ่มขึ้นจาก 92.94% เป็น 94.37% และความถูกต้องโดยรวมเพิ่มขึ้นจาก 93.17% เป็น 94.84% เมื่อประยุกต์ใช้งานร่วมกับวิธีเซปสตรัม จากผลการทดลองดังกล่าวพิสูจน์ได้ว่าวิธีการลบดีซีไบอัสร่วมกับตัวกรองบัตเตอร์เวิร์ธเป็นวิธีการที่เหมาะสมสำหรับการเตรียมสัญญาณที่มีคุณภาพให้กับกระบวนการการตรวจจับคำพิตช์ของเสียงพูดภาษาไทย

(วิทยานิพนธ์มีจำนวนทั้งสิ้น 65 หน้า)

คำสำคัญ : การจำแนกเพศบุคคล, การตรวจจับคำพิตช์ของเสียงพูดภาษาไทย, ออโตคอร์รีเลชัน, เซปสตรัม



อาจารย์ที่ปรึกษาวิทยานิพนธ์

Name : Miss Nidchpapha Wiratpokee
Thesis Title : Improvement of Thai Speech Pitch Detection for Gender Classification
Major Field : Computer Technology
King Mongkut's Institute of Technology North Bangkok
Thesis Advisors : Assistant Professor Dr. Kuntpong Woraratpanya
Assistant Professor Dr. Supot Nitsuwat
Academic Year : 2006

Abstract

In this thesis, an improvement of Thai speech pitch detection for gender classification is proposed. This research focuses on a preprocessing phase of the pitch detection, since the overall performance of a pitch detection procedure depends on quality of signals. For this reason, a simple and effective preprocessing algorithm, removed-DC-bias collaborating with butterworth filter, is applied to gender classification system. In order to prove capabilities of the intended algorithm, two reliable techniques of the pitch detection are chosen, i.e., autocorrelation and cepstrum for spatial- and frequency-domain processing, respectively. Then 200 speech samples are set up and divided into two groups: 40 samples for training and 160 samples for experimenting. The experimental results show that the proposed algorithm can increase the overall accuracy of the gender classification system from 92.94% to 94.37% for applying to autocorrelation technique and from 93.17% to 94.84% for utilizing cepstrum approach, respectively. This verifies that the removed-DC-bias collaborating with butterworth filter is suitable for providing good quality signals for the Thai-speech-pitch-detection procedure.

(Total 65 Pages)

Keyword : gender classification, Thai speech pitch detection, autocorrelation, cepstrum



Advisor

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยความช่วยเหลือจากหลายฝ่าย ขอขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร.กัมต์พงษ์ วรรัตน์ปัญญา และผู้ช่วยศาสตราจารย์ ดร.สุพจน์ นิตย์สุวัฒน์ เป็นอาจารย์ที่ปรึกษาที่กรุณาให้คำแนะนำที่เป็นประโยชน์ ข้อคิดเห็นและแก้ไขข้อบกพร่องต่างๆ ในการทำวิจัยมาโดยตลอด และขอขอบคุณเพื่อนร่วมงาน เพื่อนร่วมห้องเรียนที่ให้ความช่วยเหลือ และให้กำลังใจตลอดเวลา

ขอขอบพระคุณบิดา มารดา พี่และน้อง ที่ให้ความช่วยเหลือในด้านต่างๆ ให้สามารถทำงาน วิจัยนี้ให้สำเร็จลุล่วงด้วยดีไว้ ณ โอกาสนี้

นิชประภา วิรัชโกศล

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ข
บทคัดย่อภาษาอังกฤษ	ค
กิตติกรรมประกาศ	ง
สารบัญตาราง	ช
สารบัญภาพ	ซ
บทที่ 1 บทนำ	1
1.1 ความเป็นมาและความสำคัญของปัญหา	1
1.2 วัตถุประสงค์	2
1.3 สมมติฐานของการวิจัย	2
1.4 ขอบเขตของการวิจัย	2
1.5 ประโยชน์ที่ได้รับ	2
บทที่ 2 ทฤษฎีและเอกสารที่เกี่ยวข้อง	3
2.1 ทฤษฎีเบื้องต้นเกี่ยวกับเสียง	3
2.2 การแยกความถี่มูลฐาน	5
2.3 กระบวนการประมวลผลสัญญาณเบื้องต้น	12
2.4 งานวิจัยที่เกี่ยวข้อง	20
2.5 วิธีการที่นำเสนอในการวิจัย	21
บทที่ 3 วิธีการดำเนินการวิจัย	22
3.1 การเตรียมข้อมูลเสียงพูด	22
3.2 กระบวนการประมาณค่าความถี่มูลฐาน	25
3.3 กระบวนการจำแนกเพศบุคคล	31
บทที่ 4 ผลการวิจัย	33
4.1 ผลการเก็บตัวอย่างสัญญาณเสียงพูด	33
4.2 ผลการทดลองการจำแนกเพศบุคคลกรณีใช้ประโยคคำพูดเหมือนกัน	34
4.3 ผลการทดลองการจำแนกเพศบุคคลกรณีใช้ประโยคคำพูดต่างกัน	42
4.4 วิเคราะห์ผลการวิจัย	50
บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ	53
5.1 สรุปผลการวิจัย	53
5.2 อภิปรายผลการวิจัย	55
5.3 ข้อเสนอแนะ	56
เอกสารอ้างอิง	57

สารบัญ (ต่อ)

	หน้า
ภาคผนวก ก ผลการทดสอบประสิทธิภาพกรอบสัญญาณที่เหมาะสมกับแต่ละวิธีการ ประมาณค่าความถี่มูลฐาน	59
ประวัติผู้วิจัย	65

สารบัญตาราง

ตารางที่	หน้า
3-1 จำนวนเสียงพูดจากการเก็บข้อมูลกลุ่มตัวอย่าง	23
4-1 จำนวนตัวอย่างเสียงพูดประโยคคำว่า “กาลครั้งหนึ่งนานมาแล้ว”	33
4-2 จำนวนตัวอย่างเสียงพูดประโยคแนะนำตัว	33
4-3 การเปรียบเทียบร้อยละความถูกต้องของกระบวนการ Autocorrelation	50
4-4 การเปรียบเทียบร้อยละความถูกต้องของกระบวนการวิเคราะห์ Cepstrum	51
4-5 การเปรียบเทียบร้อยละความถูกต้องของกระบวนการ Autocorrelation	51
4-6 การเปรียบเทียบร้อยละความถูกต้องของกระบวนการวิเคราะห์ Cepstrum	52
5-1 การเปรียบเทียบประสิทธิภาพขั้นตอนวิธีการประมวลผลสัญญาณเบื้องต้น	55

สารบัญภาพ

ภาพที่	หน้า
2-1 รูปแบบคลื่นที่ไม่มีฮาร์โมนิคอื่นเจือปน	6
2-2 รูปแบบคลื่นที่มีฮาร์โมนิคพลังงานต่ำกว่าความถี่มูลฐานเจือปนอยู่	7
2-3 รูปแบบคลื่นที่มีฮาร์โมนิคพลังงานสูงกว่าความถี่มูลฐานเจือปนอยู่	7
2-4 รูปแบบคลื่นที่มีฮาร์โมนิคพลังงานสูงกว่าความถี่มูลฐานมากจนเกิดปรากฏการณ์ความถี่มูลฐานที่หายไป	7
2-5 ลักษณะการเปรียบเทียบความเหมือนกันวิธี Autocorrelation	9
2-6 ขั้นตอนในกระบวนการวิเคราะห์เซปตรัม	11
2-7 ลักษณะกรอบสัญญาณแบบต่าง ๆ	13
2-8 กรอบสัญญาณแบบสี่เหลี่ยม	14
2-9 กรอบสัญญาณแบบแฮมมิง	15
2-10 ลักษณะของวงจรกรองความถี่ผ่านต่ำในอุดมคติ	15
2-11 การเปรียบเทียบลักษณะของตัวกรองบัตเตอร์เวิร์ธกับตัวกรองชนิดอื่น	16
2-12 ขั้นตอนวิธีการ Clipping สัญญาณเสียง	16
2-13 กระบวนการ Clipping สัญญาณเสียง	18
2-14 ลักษณะสัญญาณเสียงในแต่ละขั้นตอนการ Clipping	19
3-1 ตัวอย่างสัญญาณเสียงพูดนำเข้าเพศชายประโยค “กาลครั้งหนึ่งนานมาแล้ว”	23
3-2 ตัวอย่างสัญญาณเสียงพูดนำเข้าเพศหญิงประโยค “กาลครั้งหนึ่งนานมาแล้ว”	23
3-3 ตัวอย่างสัญญาณเสียงพูดนำเข้าเพศชายสำหรับประโยคคำพูดที่ต่างกันคำว่า “ศุภวุฒิ สุขสวัสดิ์ เพศชาย อายุ 21 ปี วันเกิด 5 ธันวาคม 2526”	24
3-4 ตัวอย่างสัญญาณเสียงพูดนำเข้าเพศชายสำหรับประโยคคำพูดที่ต่างกันคำว่า “สุภาภรณ์ ไชยาพรพรรณ อายุ 54 ปี เกิดวันที่ 24 มิถุนายน 2494 ค่ะ”	24
3-5 ขั้นตอนวิธีที่ 1	25
3-6 ขั้นตอนวิธีที่ 2	26
3-7 ขั้นตอนวิธีที่ 3	26
3-8 ขั้นตอนวิธีที่ 4	27
3-9 ตัวอย่างสัญญาณเสียงก่อนและหลังลบ DC Bias	28
3-10 ขั้นตอนวิธีที่ 5	29
3-11 ขั้นตอนวิธีที่ 6	29
3-12 ขั้นตอนวิธีที่ 7	30
3-13 ขั้นตอนวิธีที่ 8	30

สารบัญภาพ (ต่อ)

ภาพที่	หน้า
3-14 ขั้นตอนวิธีการจำแนกเพศบุคคลโดยใช้ค่าพิตช์ของเสียงพูดภาษาไทย	32
4-1 ค่าพิตช์ของเสียงพูดทดสอบที่คำนวณจากวิธี Autocorrelation	34
4-2 ค่าพิตช์ของเสียงพูดทดสอบที่คำนวณจากวิธี Cepstrum	35
4-3 ค่าพิตช์ของเสียงพูดทดสอบที่คำนวณจากวิธี Autocorrelation โดยเพิ่มกระบวนการ Clipping สัญญาณเสียงพูดเข้าไปในกระบวนการประมวลผลสัญญาณเบื้องต้น	36
4-4 ค่าพิตช์ของเสียงพูดทดสอบที่คำนวณจากวิธี Cepstrum โดยเพิ่มกระบวนการ Clipping สัญญาณเสียงพูดเข้าไปในกระบวนการประมวลผลสัญญาณเบื้องต้น	37
4-5 ค่าพิตช์ของเสียงพูดทดสอบที่คำนวณจากวิธี Autocorrelation โดยเพิ่มการลบ DC Bias และ Butterworth Filter เข้าไปในกระบวนการประมวลผลสัญญาณเบื้องต้น	38
4-6 ค่าพิตช์ของเสียงพูดทดสอบที่คำนวณจากวิธี Cepstrum โดยเพิ่มการลบ DC Bias และ Butterworth Filter เข้าไปในกระบวนการประมวลผลสัญญาณเบื้องต้น	39
4-7 ค่าพิตช์ของเสียงพูดทดสอบที่คำนวณจากวิธี Autocorrelation ที่เพิ่มการลบ DC Bias, Butterworth Filter และการ Clipping สัญญาณเสียงพูดเข้าไปในกระบวนการประมวลผลสัญญาณเบื้องต้น	40
4-8 ค่าพิตช์ของเสียงพูดทดสอบที่คำนวณจากวิธี Cepstrum ที่เพิ่มการลบ DC Bias, Butterworth Filter และการ Clipping สัญญาณเสียงพูดเข้าไปในกระบวนการประมวลผลสัญญาณเบื้องต้น	41
4-9 ค่าพิตช์ของเสียงพูดทดสอบที่คำนวณจากวิธี Autocorrelation	42
4-10 ค่าพิตช์ของเสียงพูดทดสอบที่คำนวณจากวิธี Cepstrum	43
4-11 ค่าพิตช์ของเสียงพูดทดสอบที่คำนวณจากวิธี Autocorrelation โดยเพิ่มกระบวนการ Clipping สัญญาณเสียงพูดเข้าไปในกระบวนการประมวลผลสัญญาณเบื้องต้น	44
4-12 ค่าพิตช์ของเสียงพูดทดสอบที่คำนวณจากวิธี Cepstrum โดยเพิ่มกระบวนการ Clipping สัญญาณเสียงพูดเข้าไปในกระบวนการประมวลผลสัญญาณเบื้องต้น	45

สารบัญภาพ (ต่อ)

ภาพที่	หน้า
4-13 คำพิตช์ของเสียงพูดทดสอบที่คำนวณจากวิธี Autocorrelation โดยเพิ่มการลบ DC Bias และ Butterworth Filter เข้าไปในกระบวนการประมวลผลสัญญาณเบื้องต้น	46
4-14 คำพิตช์ของเสียงพูดทดสอบที่คำนวณจากวิธี Cepstrum โดยเพิ่มการลบ DC Bias และ Butterworth Filter เข้าไปในกระบวนการประมวลผลสัญญาณเบื้องต้น	47
4-15 คำพิตช์ของเสียงพูดทดสอบที่คำนวณจากวิธี Autocorrelation ที่เพิ่มการลบ DC Bias และ Butterworth Filter และการ Clipping สัญญาณเสียงพูดเข้าไปในกระบวนการประมวลผลสัญญาณเบื้องต้น	48
4-16 คำพิตช์ของเสียงพูดทดสอบที่คำนวณจากวิธี Cepstrum ที่เพิ่มการลบ DC Bias, Butterworth Filter และการ Clipping สัญญาณเสียงพูดเข้าไปในกระบวนการประมวลผลสัญญาณเบื้องต้น	49
ก-1 ประสิทธิภาพการจำแนกเพศบุคคลด้วยการประมาณค่าความถี่มูลฐานด้วยวิธี Autocorrelation โดยใช้กรอบสัญญาณแบบสี่เหลี่ยม	60
ก-2 ประสิทธิภาพการจำแนกเพศบุคคลด้วยการประมาณค่าความถี่มูลฐานด้วยวิธี Autocorrelation โดยใช้กรอบสัญญาณแบบสามมิ่ง	61
ก-3 ประสิทธิภาพการจำแนกเพศบุคคลด้วยการประมาณค่าความถี่มูลฐานด้วยวิธี Cepstrum โดยใช้กรอบสัญญาณแบบสี่เหลี่ยม	62
ก-4 ประสิทธิภาพการจำแนกเพศบุคคลด้วยการประมาณค่าความถี่มูลฐานด้วยวิธี Cepstrum โดยใช้กรอบสัญญาณแบบสามมิ่ง	63

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

การตรวจจับค่าพิตช์เป็นปัญหาในการประมวลผลสัญญาณเสียงพูด ปัญหานี้เป็นปัญหาสำคัญที่ยังต้องได้รับการแก้ไข เนื่องจากจะส่งผลกระทบต่อเนื่องไปยังการประมวลผลสัญญาณเสียงในรูปแบบอื่น เช่น การรู้จำเสียง การสังเคราะห์เสียง เป็นต้น การตรวจจับค่าพิตช์คือ ส่วนกลับของการประมาณค่าความถี่มูลฐาน (Fundamental Frequency estimation : f_0) สัญญาณเสียงที่เข้าสู่ระบบนั้น นอกจากสัญญาณเสียงที่ต้องการพิจารณาแล้ว ยังประกอบด้วย ความถี่อื่นที่ปะปนอยู่ในช่วงความกว้างของคลื่นสัญญาณเดียวกันและกระบวนการทางธรรมชาติอื่น ๆ ของผู้พูดที่เกิดขึ้นในขณะที่ทำการบันทึกเสียง เช่น รายละเอียดในการออกเสียงคำพูด อารมณ์ และ ฉันทลักษณ์ของผู้พูดแต่ละคน รวมถึงเสียงรบกวนที่เกิดจากสภาพแวดล้อมในการบันทึกเสียง ที่ทำให้ระยะของค่าพิตช์ของผู้พูดแต่ละคนเปลี่ยนแปลงไป ซึ่งมีผลกระทบทำให้ค่าพิตช์ของสัญญาณเสียงที่ต้องการพิจารณาลาดเคลื่อนไปจากความเป็นจริง จนกระทั่งปัจจุบันยังไม่มีทฤษฎีใดในการตรวจจับค่าพิตช์ที่มีความสามารถแยกเฉพาะค่าพิตช์ของสัญญาณเสียงที่สนใจออกจากฮาร์โมนิคอื่นได้อย่างถูกต้องแม่นยำครบถ้วนทุกรายละเอียดได้ตามต้องการ โปรแกรมการรู้จำเสียงพูดในปัจจุบันส่วนใหญ่จะละทิ้งข้อมูลพิตช์โดยถือว่าไม่เกี่ยวข้องกับงานรู้จำ ในขณะที่ในความเป็นจริงเสียงพยางค์ เสียงสระของแต่ละบุคคลมีความเป็นเอกลักษณ์สูง ถ้าไม่เอาใจใส่ในการรู้จำค่าพิตช์มาช่วยเป็นฟังก์ชันในการขับเคลื่อน หรือแม้แต่ไม่เอาใจใส่ค่าพิตช์ในเสียงกระซิบแล้ว จะทำให้ผลลัพธ์ที่ได้มีกำกวม

จากงานวิจัยของ David Gerhard [1] ได้รวบรวมเทคนิคที่มีประสิทธิภาพในการประมาณค่าความถี่มูลฐาน (Fundamental Frequency : f_0) พบว่าวิธีการที่มีประสิทธิภาพและได้รับความนิยมมากที่สุดในโดเมนเวลาคือวิธีออโตคอร์เรลชัน (Autocorrelation) และในโดเมนความถี่คือวิธีวิเคราะห์สเปกตรัม (Spectrum Analysis) ดังจะเห็นประสิทธิภาพได้จากโปรแกรม Colea ที่พัฒนาโดย Philip Loizou [2] ที่ใช้โปรแกรม MATLAB ในการประมวลผล และโปรแกรม Terez's [3,4] ซึ่งใช้โปรแกรม MATLAB ในการประมวลผล เช่นกัน โดยทั้งสองโปรแกรมได้เลือกใช้วิธีออโตคอร์เรลชัน และวิธีวิเคราะห์สเปกตรัมเป็นตัวประมาณค่าพิตช์ ซึ่งผลลัพธ์ที่ได้จากการทดลองมีประสิทธิภาพเป็นที่น่าสนใจ

งานวิจัยของ Tan และ มนตรี เรื่อง Pitch Detection Algorithm : Autocorrelation Method and AMDF [5] ได้เพิ่มประสิทธิภาพกระบวนการประมวลผลสัญญาณเบื้องต้นด้วยกระบวนการ Clipping สัญญาณเสียงก่อนการหาค่าพิตช์ด้วยวิธีออโตคอร์เรลชันซึ่งทำให้ประสิทธิภาพการหาค่าพิตช์ด้วยวิธีการออโตคอร์เรลชันเพิ่มขึ้น

จากการวิจัยข้างต้น พบว่าประสิทธิภาพโดยรวมของการตรวจจับคำพิตซ์ขึ้นอยู่กับคุณภาพของสัญญาณนำเข้า ผู้วิจัยจึงปรับปรุงประสิทธิภาพขั้นตอนวิธีการประมวลผลสัญญาณเบื้องต้นของสัญญาณเสียงนำเข้า เพื่อเตรียมความพร้อมให้กับสัญญาณเสียงพูดก่อนเข้าสู่การตรวจจับคำพิตซ์ โดยประยุกต์ใช้วิธีการลบดีซีไบอัส (Removed-DC-Bias) ร่วมกับตัวกรองบัตเตอร์เวิร์ธ (Butterworth Filter) ซึ่งเป็นวิธีการที่ง่าย และมีประสิทธิภาพในการกรองสัญญาณฮาร์โมนิคอื่นที่อยู่ล้อมรอบสัญญาณเสียงที่ต้องการพิจารณา จากนั้นจึงเข้าสู่กระบวนการหาค่าพิตซ์โดยเลือกใช่วิธีออโตคอร์เรเลชันในโดเมนเวลา และวิธีวิเคราะห์เซปตรัมในโดเมนความถี่แล้วจึงเข้าสู่กระบวนการจำแนกเพศบุคคลจากคำพิตซ์ของเสียงพูดภาษาไทยเพื่อพิสูจน์ความสามารถของขั้นตอนวิธีดังกล่าว

1.2 วัตถุประสงค์ของการวิจัย

เพื่อพัฒนาขั้นตอนวิธีการปรับปรุงประสิทธิภาพการประมวลผลสัญญาณเบื้องต้นสำหรับการจำแนกเพศบุคคลจากเสียงพูดภาษาไทยโดยใช้กระบวนการตรวจจับคำพิตซ์ของเสียงพูด

1.3 สมมติฐานของการวิจัย

ขั้นตอนวิธีเพื่อการจำแนกเพศของบุคคลที่พัฒนาขึ้นนี้สามารถระบุเพศของแต่ละบุคคลได้ถูกต้องอย่างน้อย 80 เปอร์เซ็นต์

1.4 ขอบเขตของการวิจัย

1.4.1 เสียงพูดที่นำเข้าไปในกระบวนการสอนและทดสอบกับโปรแกรมการจำแนกเพศของบุคคลมีลักษณะดังนี้ เป็นเสียงประเภท Mono, มีอัตราสุ่มตัวอย่าง (Sampling Rate) เท่ากับ 16,000, มีความละเอียด (Resolution) เป็น 16 บิต และเป็นไฟล์เสียงประเภท .WAV

1.4.2 สภาพแวดล้อมที่ติดมากับเสียงพูดนำเข้าจะเป็นสภาพแวดล้อมปกติ เช่น สำนักงาน เป็นต้น

1.4.3 การประมวลผลสัญญาณแต่ละครั้งจะต้องใช้วิธีการประมาณค่าความถี่มูลฐานวิธีเดียวกันทั้งกระบวนการเรียนรู้และกระบวนการทดสอบ

1.5 ประโยชน์ที่ได้รับ

1.5.1 ได้ขั้นตอนวิธีการของกระบวนการประมวลผลสัญญาณเบื้องต้นที่มีประสิทธิภาพในการตรวจจับคำพิตซ์ของเสียงพูดภาษาไทย

1.5.2 นำหลักการและขั้นตอนวิธีในงานวิจัยนี้ไปปรับปรุงประสิทธิภาพระบบการรู้จำเสียงวรรณยุกต์ภาษาไทย

บทที่ 2

ทฤษฎีและเอกสารที่เกี่ยวข้อง

2.1 ทฤษฎีเบื้องต้นเกี่ยวกับเสียง

เสียงเป็นศาสตร์สาขาหนึ่งทางวิทยาศาสตร์ เรียกว่า สวณศาสตร์ (อ่านว่า สะ-วะ-นะ-สาด) หมายถึง วิชาที่ว่าด้วยเรื่องเสียง เสียงเกิดจากการสั่นสะเทือนของวัตถุ การสั่นสะเทือนเกิดจากความดัน ถ้าความดันยิ่งสูง จะยิ่งมีเสียงดัง วัตถุใดที่สั่นสะเทือนได้ก็จะเกิดเสียงได้ เมื่อวัตถุสั่นสะเทือนจะทำให้เกิดการอัดตัวและการขยายตัวของคลื่นเสียง ซึ่งถูกส่งไปยังหูโดยผ่านชั้นบรรยากาศ ดังนั้นเสียงที่ไต่ย็นก็เป็นผลจากการที่คลื่นเสียงถูกส่งจากวัตถุที่สั่นสะเทือนไปยังหูนั่นเอง เสียงที่ไต่ย็นจะเป็นเสียงสูง เสียงต่ำ เสียงดัง เสียงเบา หรือเสียงมีคุณภาพเสียงในลักษณะต่าง ๆ จะขึ้นอยู่กับแหล่งกำเนิดเสียง และจำนวนรอบต่อวินาที (Cycle per Second) ของการสั่นสะเทือนที่ทำให้เกิดเสียงนั้น โดยมีหน่วยของความถี่ของเสียงเรียกว่า เฮิรทซ์ (Hz) และมีหน่วยวัดการสั่นสะเทือนจากความดันเป็น เดซิเบล (Decibel : dB) [6]

2.1.1 ความถี่ของเสียงกับการไต่ย็น

ความถี่ของเสียงที่มนุษย์เราสามารถรับฟังเสียงได้ต่ำสุดคือ 20 Hz และสูงสุดที่ 20,000 Hz บางคนอาจไต่ย็นเสียงสูงมากกว่านี้ถึง 40,000 Hz ส่วนใหญ่เด็กที่มีอายุน้อยจะไต่ย็นเสียงสูงดีมาก แต่ครั้นถึงวัยชราการไต่ย็นเสียงสูงจะลดลงและคนที่ประสาทหูเสียเสียงสูงจะเสียทั้งหมด อย่างไรก็ตาม ความถี่ที่คนเรารับฟังอยู่ทุกวันนี้ เป็นช่วงความถี่ของเสียงระหว่าง 125-8,000 Hz เท่านั้น ความถี่น้อยคือเสียงต่ำ ความถี่มากคือเสียงสูง สำหรับความถี่ที่ใช้วัดการไต่ย็นเพื่อวินิจฉัยโรคหรือตรวจสอบการไต่ย็นนั้น จะอยู่ระหว่าง 500-2,000 Hz ความถี่ช่วงนี้ใช้สำหรับ วัดหาความสามารถในการไต่ย็นเสียงคำพูดเพื่อเข้าใจคำพูด ช่วงความดังของเสียงที่คนเราไต่ย็น คำพูดธรรมดาอยู่ระหว่าง 20-40 เดซิเบล หากใครไต่ย็นต่ำกว่า 20 เดซิเบล จะไต่ย็นเป็นเสียงกระซิบ ฉะนั้นความดังของเสียงที่พูดกันธรรมดา ๆ ไต่ย็นชัดเจนนั้นอยู่ระหว่าง 30-40 เดซิเบล และไม่เกิน 65 เดซิเบล ซึ่งเป็นช่วงของการไต่ย็นเสียงคำพูดที่ชัดที่สุด

2.1.2 กระบวนการทำให้เกิดเสียง (Speech Production)

ระบบของเสียงพูดสามารถที่จะพิจารณาได้ว่า ประกอบด้วยลำดับของท่อ และ ช่องที่ต่อออกมาจากปอดไปยังปากและจมูก ท่อและช่องนี้จะมีความยาวโดยประมาณ 7 นิ้ว เส้นเสียงจะอยู่ในตำแหน่งปลายที่ตรงข้ามกับข้อปอด ทำหน้าที่ควบคุมการไหลของลมให้ผ่านปอดเข้าสู่ช่องทางเดินเสียง ภายใต้การควบคุมของกล้ามเนื้อส่วนประกอบของช่องทางเดินเสียงที่มีลักษณะเป็นท่อ จะสามารถเปลี่ยนแปลงรูปร่างได้ในอัตราถึง 10 ครั้งต่อวินาที ส่วนเส้นเสียงนั้นจะสามารถเปิด-ปิดด้วยอัตราเร็วประมาณ 100-300 ครั้งต่อวินาที

การเปลี่ยนแปลงรูปร่างของช่องทางเดินเสียง รูปร่างและตำแหน่งของสื่อกกลางที่ทำให้เกิดเสียง ดังกล่าวนี้นี้รวมเรียกว่า กระบวนการทำให้เกิดเสียง [7] ในขั้นตอนการเกิดของเสียงแบ่งได้เป็น 3 ขั้นตอนหลัก คือ ขั้นเริ่มต้น (Initiation) ขั้นตอนนี้เป็นขั้นตอนที่ลมเริ่มถูกขับออกจากปอด เพื่อเข้าสู่ขั้นตอนที่สองต่อไป ขั้นตอนที่สอง คือ ขั้นดัดแปลงลมที่เส้นเสียง (Phonation) เป็นขั้นตอนที่ลมจากปอดจะผ่านมายังหลอดลมและกล่องเสียง ซึ่งที่กล่องเสียงนี้ เส้นเสียงจะทำหน้าที่เป็นลิ้นเปิดปิด ทำให้เกิดเสียงได้ 2 ชนิด คือ เสียงก้อง (Voiced Sounds) และ เสียงไม่ก้อง (Unvoiced Sounds) อวัยวะที่ใช้ในขั้นตอนนี้ คือ ส่วนที่ต่อจากปอดขึ้นมาจนถึงกล่องเสียง และ ขั้นตอนที่สาม คือ ขั้นเปลี่ยนแปลงลักษณะเสียง (Articulation) ในขั้นตอนนี้ลมที่ผ่านออกมาจากกล่องเสียง จะถูกแปลงให้เกิดเสียงในลักษณะต่าง ๆ อวัยวะที่ใช้ในขั้นตอนนี้ คือ ส่วนที่ต่อจากกล่องเสียงจนถึงริมฝีปาก

2.1.3 คุณสมบัติของเสียงแต่ละเสียง [7] สามารถจำแนกได้ดังนี้

2.1.3.1 ระดับเสียง (Pitch) หมายถึง เสียงสูงเสียงต่ำ เราสามารถเปรียบเทียบเสียง 2 เสียงได้ว่าเสียงไหนมีระดับเสียงสูงกว่า และเสียงไหนมีระดับเสียงต่ำกว่าโดยการฟัง ถ้าระดับเสียงต่างกันมากก็จะฟังง่าย ถ้าต่างกันเพียงเล็กน้อยจะฟังยากกว่า โดยทั่วไประดับเสียงของผู้ชายจะต่ำกว่าระดับเสียงของผู้หญิง และในเพศเดียวกัน ระดับเสียงของเด็กจะสูงกว่าระดับเสียงของผู้ใหญ่ ระดับเสียงที่ต่ำกว่าจะเกิดจากแหล่งกำเนิดเสียงที่ใหญ่มากกว่า ระดับเสียงเกิดจากความถี่ของการสั่นสะเทือนของวัตถุหรืออาจกล่าวได้ว่า ความถี่ของการสั่นสะเทือนเป็นตัวกำหนดระดับเสียง วัตถุที่สั่นสะเทือนเร็วกว่าทำให้เกิดระดับเสียงสูงกว่า ในขณะที่วัตถุที่สั่นสะเทือนช้ากว่าทำให้เกิดระดับเสียงต่ำกว่า ความถี่ซึ่งเป็นตัวกำหนดระดับเสียงได้มาจากการวัดความถี่ของการสั่นสะเทือนของวัตถุ โดยมีหน่วยเป็นรอบต่อวินาที วัตถุที่สั่นสะเทือนมากกว่าจะมีความถี่มากกว่าทำให้เกิดระดับเสียงสูงกว่า

2.1.3.2 ความเข้มเสียง (Volume หรือ Intensity) หมายถึง เสียงเบาเสียงดัง ความเข้มเสียงเกิดจากแรงสั่นสะเทือนของวัตถุที่เป็นแหล่งกำเนิดเสียง ถ้าวัตถุสั่นสะเทือนมากเสียงจะดัง และในทางตรงกันข้ามถ้าวัตถุสั่นสะเทือนน้อย เสียงก็จะเบา ความเข้มเสียงขึ้นอยู่กับความแรงที่ส่งจากแหล่งกำเนิดเสียงไปยังหูเวลาที่เราระโกนเส้นเสียงจะสั่นสะเทือนมาก ทำให้เกิดเสียงดัง ความเข้มเสียงวัดได้จากความสูงของคลื่นเสียง (Amplitude) ถ้าคลื่นเสียงยิ่งสูงเสียงก็จะยิ่งดัง ในขณะที่เสียงเดินทางผ่านชั้นบรรยากาศนั้น ความเข้มเสียงจะน้อยลงตามลำดับ กล่าวคือ เสียงจะเบาลง ดังนั้น ระยะทางที่เสียงต้องเดินทางก็จะมีผลต่อความเข้มเสียงด้วย

2.1.3.3 สีสันทเสียง (Timbre หรือ Tone Color) หมายถึง เสียงที่มาจากแหล่งกำเนิดเสียงต่าง ๆ กัน เช่น เสียงเปียโน เสียงกีตาร์ เสียงพูด เสียงเคาะประตู เสียงกริ่ง เสียงลม เสียงร้องเพลง ฯลฯ คลื่นเสียงที่เกิดจากแหล่งที่มีสีสันทเสียงต่างกันจะมีรูปร่างลักษณะต่างกัน เช่น คลื่นเสียงของไวโอลินจะมีรูปร่างต่างจากคลื่นเสียงของทรัมเป็ต เป็นต้น สื่อดนตรี (Medium) ที่ต่างกันจะทำให้เกิดสีสันทเสียงต่างกัน สื่อดนตรี ก็คือ เครื่องดนตรี หรือเส้นเสียงใน

กรณีการขับร้องเครื่องดนตรีต่างชนิดกันจะทำให้เกิดสีนเสียงต่างกัน เพลงเดียวกันที่บรรเลงโดยเครื่องดนตรีต่างชนิดกันจะทำให้เกิดรสชาติต่างกันเพราะสีนเสียงต่างกัน

2.1.3.4 คุณภาพเสียง (Tone Quality) หมายถึง เสียงดีมากหรือดีน้อย เช่น เสียงโน้ตตัวเดียวกันที่เกิดจากเปียโน 2 ตัว จะมีคุณภาพเสียงไม่เหมือนกันทีเดียว แม้จะมีระดับเสียง ความเข้มเสียง และสีนเสียงเหมือนกันทุกประการ เวลาเลือกซื้อเครื่องดนตรีจึงต้องคำนึงถึงคุณภาพเสียง ซึ่งแต่ละบุคคลจะชอบคุณภาพเสียงไม่เหมือนกัน

2.1.3.5 ความยาวเสียง (Duration) เป็นพื้นฐานของดนตรีที่เกี่ยวข้องกับระยะเวลา เสียงแต่ละเสียงที่เกิดขึ้นต้องมีระยะเวลาซึ่งทำให้เกิดเสียงยาวเสียงสั้น เสียงไม่ว่าจะมีระดับเสียงที่แน่นอนหรือไม่ก็ต้องมีความยาวเข้ามาเกี่ยวข้อง ความยาวเสียงเป็นที่มาของจังหวะซึ่งมีส่วนสำคัญอย่างยิ่งในดนตรีของทุกชาติทุกภาษา ความยาวเสียงในที่นี้หมายรวมถึง ความเงียบของตัวหยุดด้วย เนื่องจากดนตรีเป็นผลของกระบวนการเกิดเสียง (Sound) สลับกับความเงียบ (Silence) ไม่ว่าจะเป็นเสียง (ซึ่งใช้สัญลักษณ์ตัวโน้ต) หรือความเงียบ (ซึ่งใช้สัญลักษณ์ตัวหยุด) ย่อมต้องมีความยาวหรือระยะเวลาทั้งสิ้น

มนุษย์มีความสามารถแยกแยะระดับเสียง หรือ พิตช์ได้จากการฟัง หากต้องการให้คอมพิวเตอร์สามารถแยกแยะระดับเสียงได้ จะต้องนำสัญญาณเสียงพูดที่สนใจเข้าสู่กระบวนการแยกความถี่มูลฐานของเสียงพูด (Fundamental Frequency) เพื่อทำการวิเคราะห์ระดับความถี่ของเสียงพูดนั้น ๆ

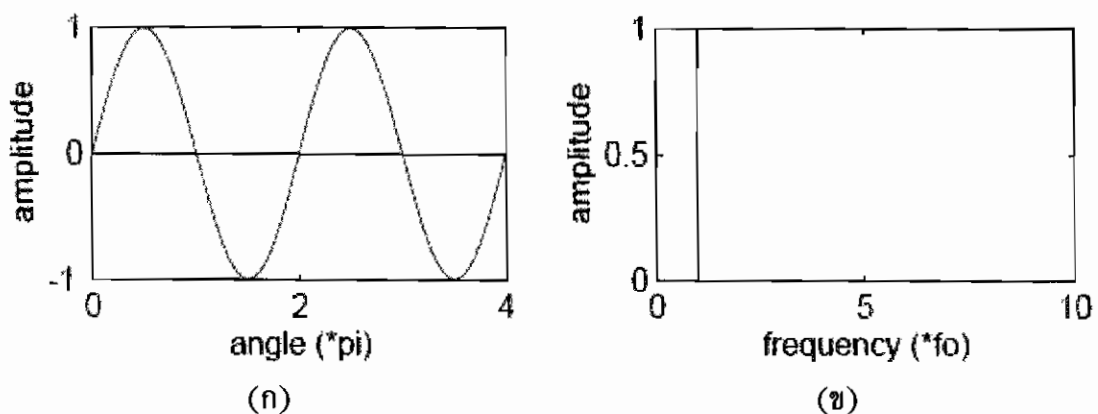
2.2 การแยกความถี่มูลฐาน (Fundamental Frequency Extraction)

กระบวนการตรวจจับค่าพิตช์ คือ กระบวนการที่ถูกออกแบบมาสำหรับประมาณค่าพิตช์หรือความถี่มูลฐานของเสียงพูดหรือเสียงโน้ตดนตรีหรือโทนเสียง โดยสามารถประมวลผลได้ทั้งในโดเมนเวลาและโดเมนความถี่ โดยการตรวจจับค่าพิตช์จะทำการประมาณค่าคาบเวลาของสัญญาณเสียงก่อน จากนั้นจึงทำการคำนวณความถี่มูลฐานในทางกลับกันของคาบเวลาซึ่งก็คือค่าพิตช์นั่นเอง

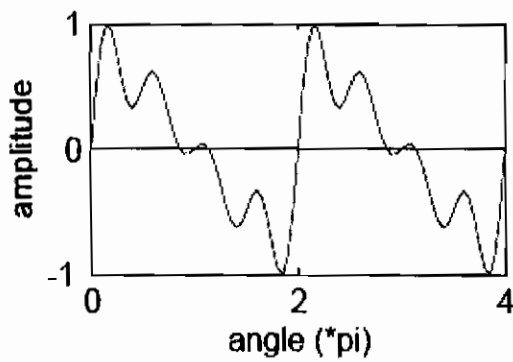
รูปแบบของการสั่นของเส้นเสียงทำให้เกิดเป็นเสียงแบบต่าง ๆ การสั่นของเส้นเสียงแบบธรรมดา (Normal Vibration) ทำให้เกิดเสียงก้อง (Voiced Sound) ส่วนเสียงไม่ก้อง (Voiceless Sound) จะไม่มีการสั่นสะเทือนของเส้นเสียง ต่อมาอากาศก็จะเดินทางเข้าสู่ช่องปาก ซึ่งประกอบด้วยอวัยวะแปรเสียงหรือฐานกรณ (Articulators) มากมาย เพื่อทำหน้าที่ในการกล่อมเกลาเสียงให้ออกมามีคุณลักษณะแตกต่างกัน สระซึ่งเป็นแกนกลางพยางค์มีลักษณะเป็นเสียงก้อง ซึ่งเสียงก้องจะมีความเป็นคาบเวลาของระดับเสียง (Pitch Period) และส่วนกลับของระยะเวลารายคาบของสัญญาณเสียงคือ ความถี่มูลฐาน (Fundamental Frequency : f_0) [8] ความยากในการค้นหา f_0 ของรูปแบบคลื่นขึ้นอยู่กับตัวของรูปแบบคลื่นเอง ถ้ารูปแบบคลื่นที่สนใจมีฮาร์โมนิคอื่นมาปะปนโดยฮาร์โมนิคนั้นมีพลังงานต่ำกว่า f_0 ก็สามารถตรวจจับได้ง่ายกว่า

คลื่นเสียง (Waveform) มีลักษณะเป็นช่วง ๆ ซ้ำ ๆ กัน หรือเรียกว่า สัญญาณแบบมีคาบ (Periodic) หากทำการพลอตกราฟตามแกนจินตภาพ ให้แกนนอนเป็นมุมคูณกับ π โดยที่ π เท่ากับ 180 องศา แกนตั้งเป็นค่า Amplitude ที่มีค่าอยู่ระหว่าง -1 ถึง 1 แล้ว พบว่าคลื่นเสียงจะมีลักษณะคล้ายรูปไซน์ (Sine) ดังภาพที่ 2-1 (ก) หนึ่งรูปคลื่นเสียงคือสัญญาณตั้งแต่ตำแหน่ง (x,y) ที่ (0,0) จนถึงตำแหน่ง (2,0) คือหนึ่งรูปคลื่นเสียงจะมีขนาดตามแกนจินตภาพเท่ากับ 360 องศาพอดี (2 คูณ π) ภาพที่ 2-1 (ข) แสดงความถี่มูลฐาน ณ ตำแหน่งแรกของกราฟ จะเห็นว่า ค่าความถี่มูลฐานแสดงค่าแอมพลิจูด (Amplitude) ที่ 1 จากกราฟสามารถแยกความถี่มูลฐานได้อย่างชัดเจนเนื่องจากเป็นสัญญาณเสียงเพียงเสียงเดียวไม่มีฮาร์โมนิคอื่นมารบกวนสัญญาณเลย ภาพที่ 2-2 (ก) แสดงสัญญาณเสียงที่สนใจปะปนอยู่กับฮาร์โมนิคอื่นที่มีค่าพลังงานต่ำกว่า เมื่อนำมาพลอตในกราฟดังภาพที่ 2-2 (ข) จะเห็นว่า มีฮาร์โมนิคปะปนกันอยู่ 5 คลื่นฮาร์โมนิค แต่สามารถมองเห็นความถี่มูลฐาน ณ ตำแหน่งแรกได้อย่างชัดเจน เนื่องจากฮาร์โมนิคอื่นอีก 4 ฮาร์โมนิคเป็นฮาร์โมนิคที่มีคลื่นพลังงานต่ำกว่าความถี่มูลฐานที่สนใจ ภาพที่ 2-3 แสดงฮาร์โมนิคที่สนใจปะปนอยู่กับฮาร์โมนิคอื่นที่มีค่าพลังงานสูงกว่า เมื่อทำการพลอต กราฟความถี่จะเห็นว่า มีฮาร์โมนิคปะปนกันอยู่ทั้งหมด 7 ฮาร์โมนิค ฮาร์โมนิคที่ 2,3,4,5,6 มีพลังงานมากกว่าความถี่มูลฐานมีผลทำให้ความถี่มูลฐานถูกกดให้ต่ำลงไป แต่ยังสามารถมองเห็นได้อยู่ หากฮาร์โมนิคอื่นมีพลังงานสูงกว่ามาก ๆ อาจส่งผลให้ความถี่มูลฐานถูกกดลงไปจนไม่สามารถมองเห็นได้ ปรากฏการณ์นี้เรียกว่า ปรากฏการณ์ของความถี่มูลฐานที่หายไป (Missing Fundamental Frequency) ซึ่งส่งผลให้การตรวจจับความถี่มูลฐานเลื่อนไปจับถูกตำแหน่งที่ 2 ซึ่งเป็นตำแหน่งที่ผิดไปจากความเป็นจริงแทน

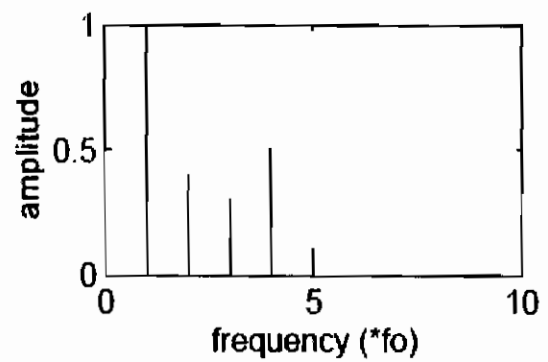
ในการประมวลผลสัญญาณเสียงพูด เสียงพูดที่นำเข้าสู่กระบวนการแยกความถี่มูลฐาน มักจะมีฮาร์โมนิคอื่นเจือปนติดอยู่เสมอ ไม่ว่าจะเป็นฮาร์โมนิคของเสียงลมพัด หรือ เสียงเครื่องปรับอากาศ เป็นต้น ทำให้จำเป็นต้องหาวิธีการในการแยกเอาเฉพาะความถี่มูลฐานที่สนใจนั้นออกจากฮาร์โมนิคอื่นที่เกาะติดอยู่ให้ได้อย่างถูกต้องแม่นยำ



ภาพที่ 2-1 รูปแบบคลื่นที่ไม่มีฮาร์โมนิคอื่นเจือปน

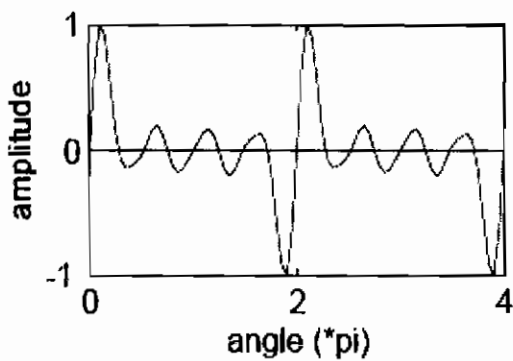


(ก)

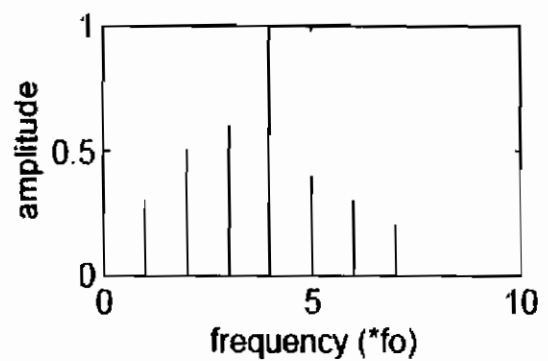


(ข)

ภาพที่ 2-2 รูปแบบคลื่นที่มีฮาร์โมนิคพลังงานต่ำกว่าความถี่มูลฐานเจือปนอยู่

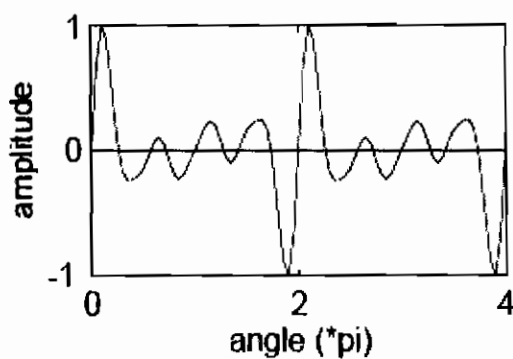


(ก)

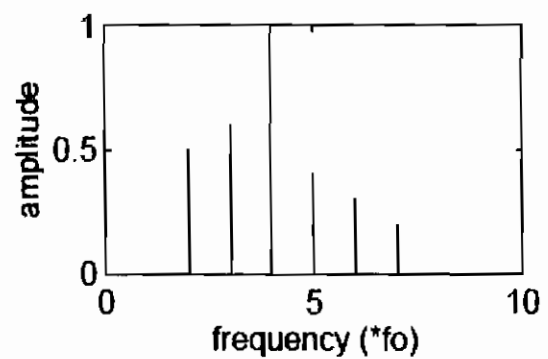


(ข)

ภาพที่ 2-3 รูปแบบคลื่นที่มีฮาร์โมนิคพลังงานสูงกว่าความถี่มูลฐานเจือปนอยู่



(ก)



(ข)

ภาพที่ 2-4 รูปแบบคลื่นที่มีฮาร์โมนิคพลังงานสูงกว่าความถี่มูลฐานมากจนเกิดปรากฏการณ์ความถี่มูลฐานหายไป

กระบวนการในการหาค่าพิตช์ หรือ ค่าความถี่มูลฐาน (f_0) มีหลายวิธีทั้งในโดเมนเวลา และในโดเมนความถี่ ในงานวิจัยนี้จะเลือกวิธีที่มีประสิทธิภาพและได้รับความนิยมในโดเมนเวลา และโดเมนความถี่มาโดเมนละ 1 วิธี คือ วิธี Autocorrelation ในโดเมนเวลา และวิธีวิเคราะห์แบบ Cepstrum ในโดเมนความถี่

2.2.1 วิธี Autocorrelation

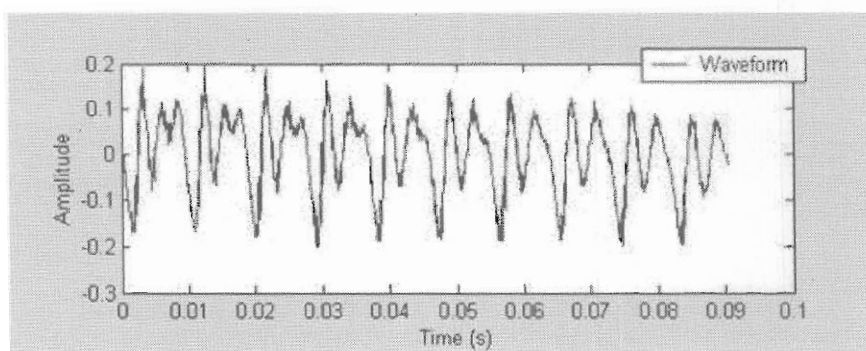
Autocorrelation เป็นเครื่องมือทางคณิตศาสตร์ที่ใช้ในการหาความเหมือนกันของสัญญาณ ด้วยตัวสัญญาณเองในแต่ละเฟรม (Frame) หรือกรอบสัญญาณของสัญญาณเสียงจะประกอบไปด้วยสัญญาณเสียงที่สนใจปะปนอยู่กับฮาร์โมนิคอื่น ๆ สิ่งหนึ่งที่คาดการณ์ได้ก็คือ ในแต่ละเฟรม จะต้องปรากฏความถี่มูลฐานอยู่เสมอแต่อาจมีฮาร์โมนิคอื่นปกคลุมไว้ แสดงว่าทุก ๆ เฟรมจะมีสิ่งที่เหมือนกันอยู่สิ่งหนึ่ง นั่นคือ ความถี่มูลฐาน การที่จะแยกค่าความถี่มูลฐานออกมาก็คือการหาความเหมือนกันของข้อมูลในแต่ละเฟรม ดังนั้นจึงนำเอา Autocorrelation มาใช้หาค่าความถี่มูลฐานของสัญญาณเสียง วิธีการนี้สามารถเขียนเป็นสมการทางคณิตศาสตร์ได้ดังนี้

$$R_x(m) = \frac{1}{N} \sum_{n=0}^{N'-1} [x(n+l)w(n)][x(n+l+m)w(n+m)] ; 0 \leq m \leq M_0 \quad (2-1)$$

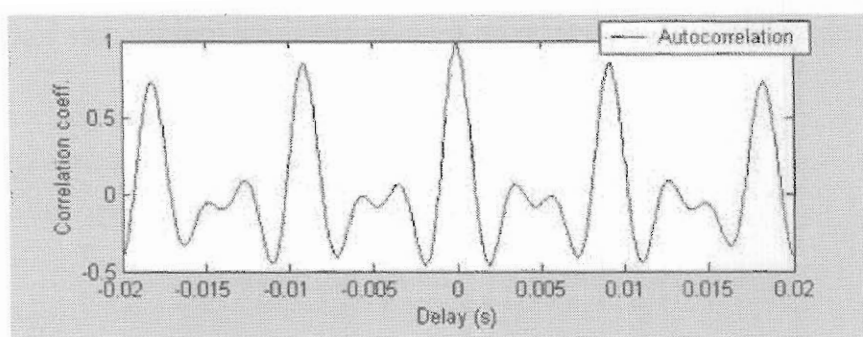
- เมื่อ $w(n)$ คือกรอบสัญญาณที่เหมาะสมสำหรับการวิเคราะห์
 N คือความยาวของแต่ละส่วนที่จะถูกวิเคราะห์
 N' คือจำนวนของตัวอย่างสัญญาณที่ใช้ในการคำนวณ $R(m)$
 M_0 คือจำนวนของจุดออโตคอร์เรลชันที่ถูกคำนวณ
 l คือตัวชี้ตัวอย่างที่เริ่มต้นของเฟรม

ภาพที่ 2-5 (ก) แสดงกราฟของสัญญาณเสียงพูดโดยที่แกนนอนเป็นเวลา และแกนตั้งเป็น Amplitude แสดงให้เห็นลักษณะของสัญญาณเสียงที่สนใจถูกปกคลุมไปด้วยฮาร์โมนิคอื่น ภาพที่ 2-5 (ข) แสดงความเหมือนกัน (Correlation) ของสัญญาณเสียงในเฟรมที่พิจารณากับเฟรมก่อนหน้า และเฟรมถัดไป จากกราฟ แกนนอนคือแกนของเวลาที่มีดีเลย์ (Delay) ต่างกัน คือ ณ เฟรมที่พิจารณาคือดีเลย์ที่ 0 เฟรมก่อนหน้าก็จะเป็นค่าลบ และเฟรมถัดไปเป็น ค่าบวก ในแกนตั้งแสดงค่าความเหมือนกัน พิจารณาเฟรมที่ดีเลย์ 0 คือเฟรมที่พิจารณาเทียบความเหมือนกันกับตัวเฟรมนั่นเอง จะมีความเหมือนกันสูงที่สุดคือ 1 เนื่องจากคือเฟรมเดียวกันมาเปรียบเทียบกัน จึงมีความเหมือนกันทุกประการ จากนั้นเทียบความเหมือนกันกับเฟรมก่อนหน้า และ เฟรมถัดไปก็จะได้ค่าเปลี่ยนแปลงไปตามค่า Autocorrelation ที่คำนวณได้

วิธี Autocorrelation จะทำงานได้ดีที่สุดเมื่อสัญญาณเสียงเป็นความถี่ต่ำและมีค่าพิตช์สม่ำเสมอ



(ก)



(ข)

ภาพที่ 2-5 ลักษณะการเปรียบเทียบความเหมือนกันวิธี Autocorrelation

2.2.2 วิธีวิเคราะห์แบบ Cepstrum

ปัญหาทั่วไปในการประมาณค่าในโดเมนความถี่คือการแบ่งสัญญาณและค้นหาความถี่ส่วนสำคัญที่เกิดขึ้นในลักษณะเป็นคาบที่ซ้ำ ๆ กัน หรือ เรียกว่าสัญญาณแบบมีคาบ (Periodic) ความยากเกิดขึ้นจากสาเหตุต่อไปนี้ 1) สัญญาณไม่ได้เป็นสัญญาณแบบมีคาบทั้งหมด 2) สัญญาณแบบมีคาบเหล่านั้นบางครั้งจะเปลี่ยนแปลงไปในความถี่มูลฐานที่เกินกว่าเวลาที่สนใจ 3) บางครั้งสัญญาณจะปะปนด้วยการรบกวนของสัญญาณแบบมีคาบของ ความถี่มูลฐานอื่น 4) สัญญาณที่เป็นแบบมีคาบด้วยเวลาระหว่าง T ยังคงเป็นสัญญาณแบบมีคาบที่เวลาระหว่าง $2T$ และ $3T$ ด้วย เป็นต้น ดังนั้นเราจำเป็นต้องค้นหาเวลาระหว่างสัญญาณแบบมีคาบที่เล็กที่สุด หรือ ความถี่มูลฐานที่สูงที่สุด และ 5) แม้แต่สัญญาณของความถี่มูลฐานที่สม่ำเสมออาจจะเปลี่ยนแปลงได้ในหลายทางที่เกินกว่าระยะเวลาที่สนใจ

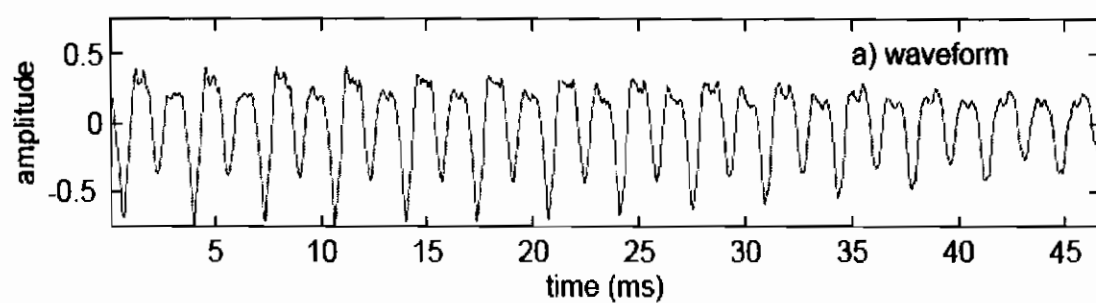
วิธีที่เชื่อถือได้ก็คือการประมาณค่าความถี่มูลฐาน โดยที่ความถี่มูลฐานที่สำคัญต้องยาวพอ และเกลี้ยงเกลาเพียงพอ ดังนั้นจึงจำเป็นต้องใช้การหาสัญญาณเสียงพูดคงที่โดยใช้การวิเคราะห์แบบ Cepstrum วิธีกันนี้เป็นวิธีการวิเคราะห์ Fourier ของ Logarithmic Amplitude Spectrum ของสัญญาณ ถ้า Logarithmic Amplitude Spectrum บรรจุหลาย ๆ ฮาร์โมนิกเป็นระยะทางสม่ำเสมอแล้ว การวิเคราะห์ Fourier ของ Spectrum จะแสดงค่าพีค (Peak) ที่สอดคล้องกับระยะทาง

ระหว่างฮาร์โมนิค คือ ความถี่มูลฐาน ประสิทธิภาพที่ได้จากการประมวลผลสเปกตรัมสัญญาณ เช่นเดียวกับสัญญาณอื่น แล้วค้นหาความซ้ำ ๆ กันในสเปกตรัมของตัวมันเอง การประมาณค่าความถี่มูลฐานด้วย Cepstrum เราจะเห็นค่าพีคในขอบเขต Quefrequency ที่สอดคล้องกับความถี่มูลฐานของเสียงพูดที่เป็นตัวอย่าง [1]

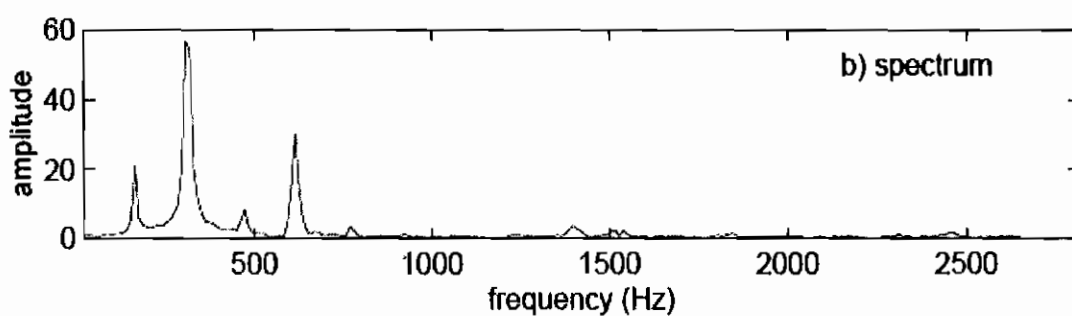
วิธี Cepstrum จะทำการวิเคราะห์ Fourier ของ Logarithmic Amplitude Spectrum ของสัญญาณที่ได้จากการประมวลผลสัญญาณขั้นต้นแล้ว นำมาแปลงเป็นโดเมนความถี่ด้วย Fast Fourier Transform : FFT ก็จะได้แถบสเปกตรัมของสัญญาณ จากนั้นก็ใส่ Logarithmic Amplitude Spectrum เข้าไปเพื่อให้เห็นพีคของฮาร์โมนิคต่าง ๆ แล้วก็ทำการ Inverse Fast Fourier Transform : IFFT อีกครั้งเพื่อให้กลับมาสู่โดเมนเวลา จะทำให้เห็นค่าพีคที่สูงที่สุดของสัญญาณชัดเจน นั่นก็คือ ค่าพีคซ์ของสัญญาณนั่นเอง ที่เรียกชื่อวิธีการนี้ว่า Cepstrum เนื่องมาจากการกลับ Spectrum อย่างทะลุปรุโปร่ง จึงทำการ ย้อนกลับตัวอักษร 4 ตัวแรกในคำว่า “Spectrum” เพื่อแสดงถึงการแก้ไขสเปกตรัม แกน x ของ Cepstrum จะมีหน่วยเป็น Quefrequency เป็นส่วนกลับของคำว่า “Frequency” เนื่องจากมันคือตัวแปรอิสระที่สัมพันธ์กับการเปลี่ยนรูป Cepstrum จึงเรียกว่า “Quefrequency” เพื่อแสดงว่า ตัวแปรมีความสัมพันธ์ใกล้ชิดมาก ๆ กับเวลา ซึ่งจะรับตัวแปรอ้างอิงเป็นหน่วยเวลา และพีคใน Cepstrum ที่สัมพันธ์กับการเกิดซ้ำ ๆ กันใน Spectrum จะถูกเรียกว่า Rahmonics ซึ่งแปลงมาจากคำว่า Harmonics เช่นกัน

วิธีการ Cepstrum ทำงานได้ดีที่สุดเมื่อความถี่มูลฐานไม่เปลี่ยนแปลงกะทันหันมากนัก ความถี่ไม่สูงเกินไปและเมื่อสัญญาณที่มีความสัมพันธ์กันไม่มีสัญญาณรบกวน ข้อเสียของวิธีการนี้ คือ การคำนวณที่ยุงยากของกระบวนการในโดเมนความถี่ จากภาพที่ 2-6 แสดงความคืบหน้าของอัลกอริธึม Cepstrum โดยภาพที่ 2-6 (ก) แสดงสัญญาณเสียงในโดเมนเวลาแกนนอนเป็นแกนเวลาและแกนตั้งเป็น Amplitude จากนั้นทำการแปลงเป็นโดเมนเวลาด้วยการใช้ Fast Fourier Transform จะได้ดังภาพที่ 2-6 (ข) ภาพที่ 2-6 (ค) แสดงการ Logarithmic Magnitude Spectrum ของสัญญาณเดียวกันเพื่อให้เห็นค่าพีคของฮาร์โมนิคต่าง ๆ จากนั้นทำการย้อนกลับมาในโดเมนเวลาอีกครั้งด้วยการทำ Inverse Fast Fourier Transform สังเกต Periodicity ของทั้งสอง Spectra แล้วแสดงลักษณะของ Logarithmic Magnitude Spectrum ขึ้นมาใหม่เพื่อให้เห็นค่าพีคซ์ของคลื่นเสียงในเฟรมที่สนใจชัดเจนขึ้น ผลลัพธ์ที่ได้แสดงในภาพที่ 2-6 (ง)

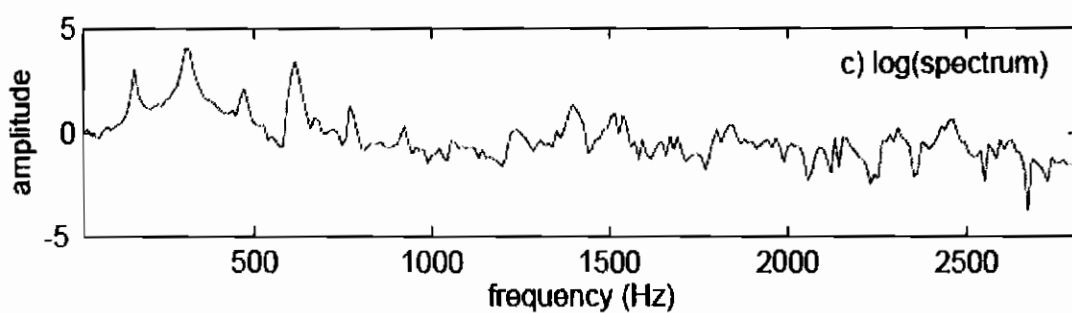
กระบวนการในการหาค่าความถี่มูลฐานทั้งสองวิธีข้างต้น เป็นวิธีการที่มีประสิทธิภาพและได้รับความนิยมอย่างมาก แต่ยังพบข้อผิดพลาดอยู่ เนื่องจากสัญญาณเสียงนำเข้าที่สนใจนั้นถูกฮาร์โมนิคอื่นปกคลุมอยู่ทั้งฮาร์โมนิคที่มีพลังงานสูงและต่ำกว่าทำให้ค่าที่ได้จากการประมาณค่าความถี่มูลฐานถูกบิดเบือนไป ดังนั้นหากต้องการเพิ่มประสิทธิภาพในการประมาณค่าความถี่มูลฐาน จึงจำเป็นต้องนำเอากระบวนการประมวลผลสัญญาณเบื้องต้น (Preprocessing) มาช่วยปรับปรุงสัญญาณนำเข้าให้พร้อมก่อนเข้าสู่กระบวนการประมาณค่าความถี่มูลฐานต่อไป



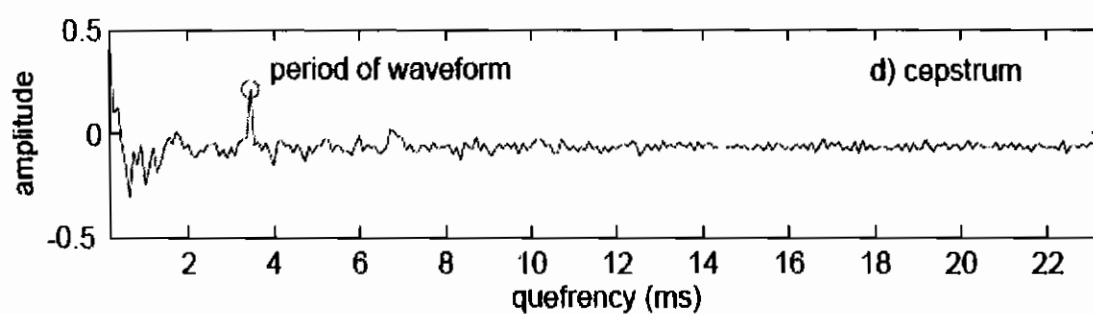
(ก)



(ข)



(ค)



(ง)

ภาพที่ 2-6 ขั้นตอนในกระบวนการวิเคราะห์เซปตรัม

2.3 กระบวนการประมวลผลสัญญาณเบื้องต้น (Preprocessing)

การประมวลผลสัญญาณเบื้องต้น เป็นขั้นตอนในการจัดเตรียมข้อมูลนำเข้าให้พร้อมก่อนนำไปสกัดคุณลักษณะเด่นของสัญญาณเสียง สำหรับการวิจัยครั้งนี้การประมวลผลสัญญาณเบื้องต้นเพื่อให้ได้สัญญาณเสียงนำเข้าที่พร้อมสำหรับนำไปหาค่าพิตช์ของสัญญาณนั้น จะเน้นไปที่การกรองฮาร์โมนิคในช่วงความถี่สูงออกไปเพื่อให้เหลือสัญญาณในช่วงความถี่ต่ำที่เป็นส่วนที่จะใช้สำหรับนำไปพิจารณาหาค่าความถี่มูลฐาน ซึ่งคือความถี่ต่ำสุดของสัญญาณนั่นเอง สัญญาณเสียงพูดประกอบด้วยฮาร์โมนิคจำนวนมากรวมกันอยู่ ค่า f_0 น้อยสุดจะอยู่ที่ 80 Hz และ มากสุดอยู่ที่ 500 Hz โดยส่วนใหญ่จะเป็นช่วงประมาณ 100–200 Hz ดังนั้นสัญญาณจึงปะปนกันอยู่ 30–40 ส่วนประกอบฮาร์โมนิค (Harmonic Component) และส่วนประกอบ f_0 จะไม่คงที่อยู่ที่ค่าเดียว เนื่องจากฟอร์แมนค่าแรก (First Formant) จะอยู่ระหว่าง 300–1,000 Hz ในตำแหน่งนั้นจะมี 2–8 ฮาร์โมนิคที่มีความเข้มข้นสูงกว่าค่า f_0 ฮาร์โมนิคที่เข้มข้นนี้จะทำให้การจับค่าพิตช์ซับซ้อนมากขึ้น โดยปกติจะมีความผิดพลาดของฮาร์โมนิค และความผิดพลาดของฮาร์โมนิคประกอบเกิดขึ้น การที่จะแก้ไขข้อผิดพลาดนี้จึงเป็นสิ่งจำเป็น เมื่อช่วงของ f_0 โดยทั่วไปอยู่ในช่วง 80–500 Hz แล้ว ส่วนประกอบของความถี่ที่อยู่เหนือกว่า 500 Hz จึงไม่ใช่สาระสำคัญในการแก้ไขประสิทธิภาพในการตรวจจับค่าพิตช์ การประมวลผลสัญญาณเบื้องต้นที่ใช้เพิ่มประสิทธิภาพในการประมาณค่าความถี่มูลฐานในการวิจัยครั้งนี้ มีดังนี้

2.3.1 การวางกรอบสัญญาณ (Smoothing Window)

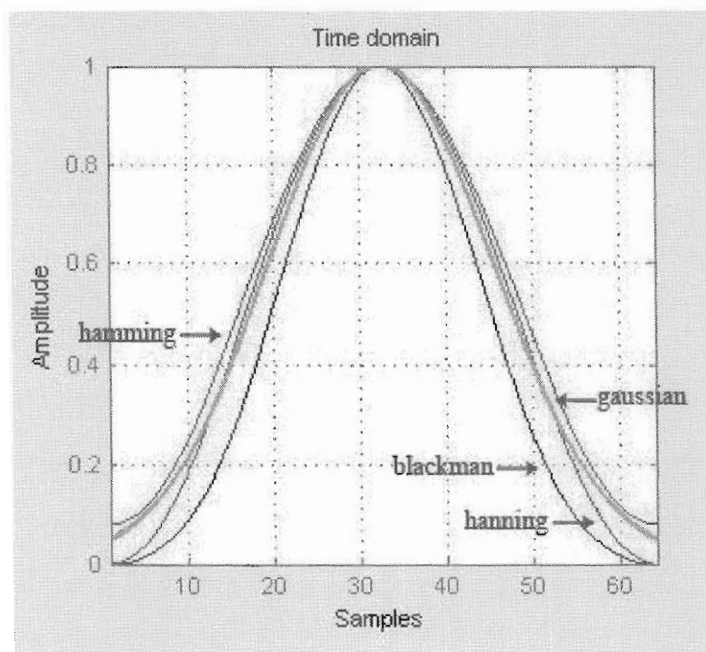
เนื่องจากสัญญาณเสียงพูดโดยรวมจะแปรเปลี่ยนตามเวลาและไม่เสถียร (Nonstationary) ทำให้ไม่สามารถจำลองสัญญาณเสียงพูดเป็นค่าทางสถิติได้ จึงจำเป็นต้องแบ่งสัญญาณเสียงพูดออกเป็นส่วนย่อย (Frame) [1] เรียกว่า “กรอบเสียงพูด” (Speech Frame) โดยแต่ละกรอบเสียงพูดจะมีความยาวประมาณ 10–40 มิลลิวินาที ทำให้สัญญาณเสียงพูดมีคุณสมบัติเปลี่ยนแปลงตามเวลาน้อยมากหรือไม่เปลี่ยนแปลงตามเวลา (Stationary) จึงสามารถนำส่วนย่อยนี้มาคำนวณหาค่าทางสถิติเพื่อใช้แทนสัญญาณเสียงพูดได้

การวางกรอบสัญญาณ เป็นการวิเคราะห์สัญญาณเสียงโดยกำหนดกรอบสัญญาณเป็นช่วงสั้น ๆ ที่มีขนาดเท่า ๆ กัน ซึ่งการวิเคราะห์ จะใช้วิธีการคูณสัญญาณเสียงด้วยฟังก์ชันกรอบสัญญาณ (Window Function) สัญญาณช่วงสั้น ๆ ที่ได้ สามารถนำไปใช้ในการวิเคราะห์หาค่าพารามิเตอร์ของสัญญาณเสียง ซึ่งฟังก์ชันกรอบสัญญาณมีหลายประเภท เช่น Rectangular Window, Hamming Window, Hanning Window, Blackman Window, Gaussian Window เป็นต้น [6] ในที่นี้จะกล่าวถึงเฉพาะกรอบสัญญาณที่ใช้ในการวิจัยนี้เท่านั้น

การกำหนดขนาดของกรอบสัญญาณที่ใช้ จะขึ้นอยู่กับองค์ประกอบต่อไปนี้

- 1) ขนาดของกรอบสัญญาณจะต้องสั้นพอ ที่จะทำให้คุณสมบัติของเสียงที่นำมาวิเคราะห์ ไม่มีการเปลี่ยนแปลงอย่างมีนัยสำคัญในกรอบสัญญาณนั้น
- 2) ขนาดของกรอบสัญญาณจะต้องยาวพอที่จะทำให้จำนวนตัวอย่างสำคัญในกรอบสัญญาณสามารถนำมาคำนวณหาค่าพารามิเตอร์ได้ตามต้องการ
- 3) ขนาดของกรอบสัญญาณจะต้องเหมาะสม คือ ไม่ควรสั้นเกินกว่าช่วงหนึ่ง

คาบของสัญญาณเสียงในช่วงที่ทำการวิเคราะห์ ถ้าสั้นกว่าจะส่งผลกระทบต่อค่าเฟรมเรต (Frame Rate) ตามปกติเฟรมเรตจะมีค่าประมาณ 2 เท่าของส่วนกลับของขนาดกรอบสัญญาณ นั่นคือ กรอบสัญญาณถัดไปจะมีการซ้อนทับกันประมาณ 50 เปอร์เซ็นต์



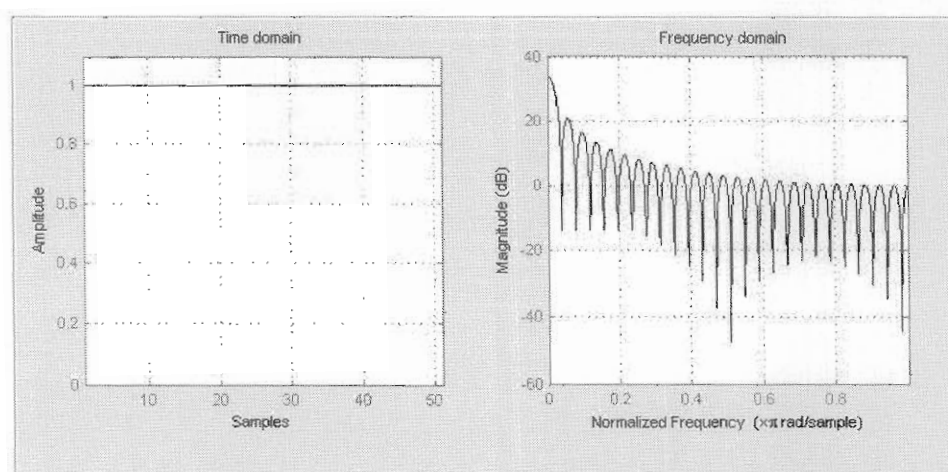
ภาพที่ 2-7 ลักษณะกรอบสัญญาณแบบต่าง ๆ

2.3.1.1 กรอบสัญญาณแบบสี่เหลี่ยม (Rectangular Window)

การนำฟังก์ชันกรอบสัญญาณที่มีช่วงขนาดจำกัด $w(n)$ มาคูณกับสัญญาณ $s(n)$ จะทำให้ได้กลุ่มตัวอย่างของเสียงพูดที่ถูกกำหนดน้ำหนักให้แปรไปตามรูปร่างของกรอบสัญญาณ ซึ่งเขียนเป็นสมการได้ดังนี้

$$w(n) = \begin{cases} 1 & 0 \leq n \leq N-1 \\ 0 & \text{otherwise} \end{cases} \quad (2-2)$$

ในสมการที่ (2-2) นี้ เป็นการกำหนดกรอบสัญญาณการวิเคราะห์ให้มีจำนวนตัวอย่าง N ตัวอย่าง



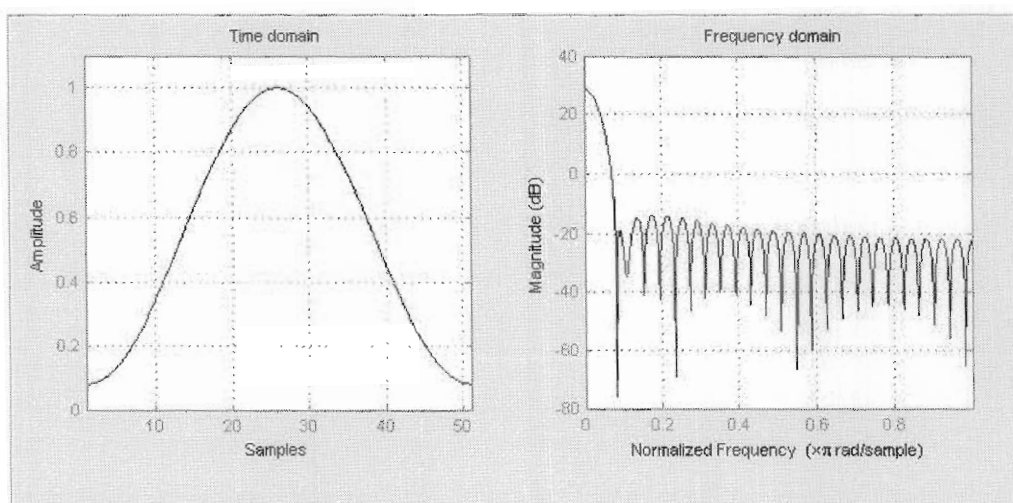
ภาพที่ 2-8 กรอบสัญญาณแบบสี่เหลี่ยม (Rectangular Window)

2.3.1.2 กรอบสัญญาณแบบแฮมมิง (Hamming Window)

การประยุกต์ใช้กรอบสัญญาณที่มีช่วงกว้างกว่าสัญญาณคงที่ และเน้นเฉพาะส่วนตรงกลางของกรอบสัญญาณให้เป็นส่วนของข้อมูลที่จะทำการวิเคราะห์ นั่นคือ การทำให้ขนาดของสัญญาณลดลงอย่างช้า ๆ ที่บริเวณปลายของกรอบกรอบสัญญาณจึงเป็นการช่วยป้องกันการเปลี่ยนแปลงอย่างกะทันหันที่บริเวณปลายของกรอบเสียงพูด เช่น ข้อมูลสัญญาณเสียงคงที่ในช่วงเวลา 10 มิลลิวินาที อาจใช้กรอบสัญญาณขนาด 20 มิลลิวินาที โดยส่วนตรงกลางของกรอบสัญญาณขนาด 10 มิลลิวินาที จะมีการเน้นเพิ่มน้ำหนักให้มากกว่าช่วง 5 วินาที ที่ส่วนปลายทั้งสองข้างของกรอบสัญญาณ โดยฟังก์ชันกรอบสัญญาณที่นิยมใช้ในการวิเคราะห์สัญญาณเสียง คือ ฟังก์ชัน Hamming Window ที่มีรูปร่างตามลักษณะของโคไซน์พัลส์ (Cosine Pulse) เขียนเป็นสมการได้ดังนี้

$$w(n) = \begin{cases} 0.54 - 0.46 \cos(2\pi n / (N-1)) & 0 \leq n \leq N-1 \\ 0 & \text{otherwise} \end{cases} \quad (2-3)$$

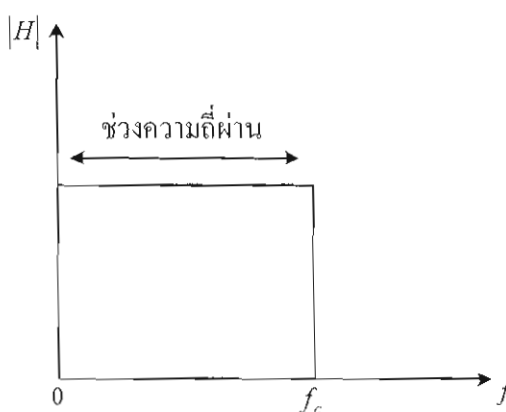
กรอบสัญญาณแบบแฮมมิงนี้ เหมาะสำหรับการประมวลผลสัญญาณเสียง เพราะสามารถเน้นสัญญาณเสียงในเฟรมที่กำลังพิจารณาให้มีความสำคัญสูงสุด โดยลดความสำคัญของสัญญาณเสียงที่อยู่ในเฟรมรอบข้าง แต่ยังคงความต่อเนื่องของสัญญาณเสียงให้มีความต่อเนื่องกัน ทำให้ได้เสียงที่ผ่านการวางกรอบสัญญาณที่ยังคงข้อมูลครบถ้วน ความต่อเนื่องของสัญญาณจะช่วยลดปัญหาในกรณีที่เสียงพูดมีการเปลี่ยนแปลงอย่างรวดเร็วได้ [11]



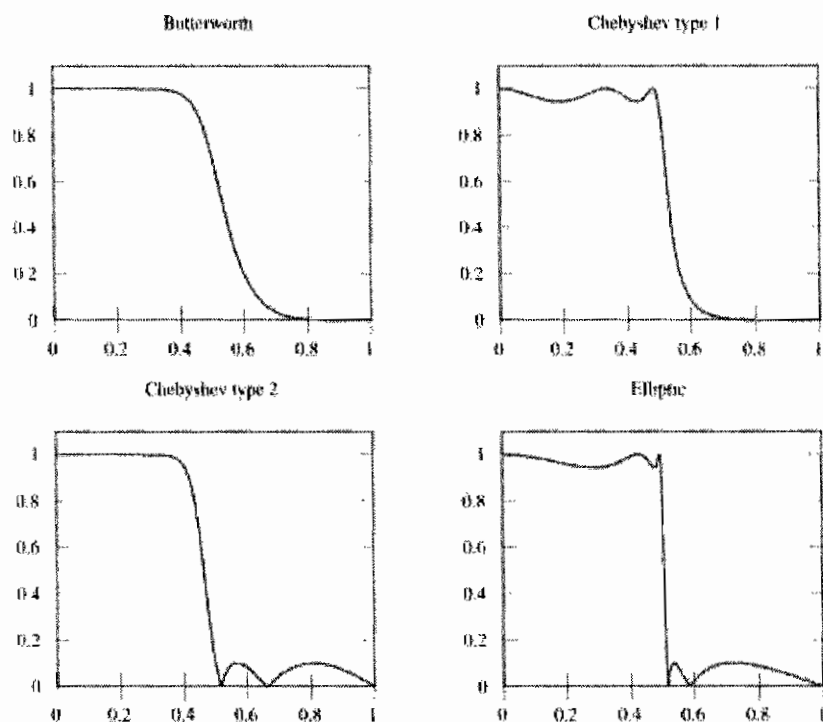
ภาพที่ 2-9 กรอบสัญญาณแบบแฮมมิง (Hamming Window)

2.3.2 ตัวกรองบัตเตอร์เวิร์ธ (Butterworth Filter)

บัตเตอร์เวิร์ธเป็นตัวกรองชนิดหนึ่งที่ถูกบรรยายเป็นครั้งแรกโดยวิศวกรชาวอังกฤษ ชื่อ Stephen Butterworth [12] ตัวกรองบัตเตอร์เวิร์ธเป็นตัวกรองสัญญาณที่มีลักษณะใกล้เคียงกับวงจรกรองความถี่ผ่านต่ำ (Low-Pass Filter : LPF) ซึ่งเป็นวงจรกรองความถี่ทางอุดมคติมากที่สุด ความถี่ตอบสนองของตัวกรองบัตเตอร์เวิร์ธจะมีลักษณะการเปลี่ยนสัญญาณได้ราบรื่นมากที่สุด คือไม่มีระลอกคลื่นเลย เมื่อเปรียบเทียบกับตัวกรองชนิดอื่น ลักษณะของตัวกรองบัตเตอร์เวิร์ธจะมีลักษณะราบเรียบเป็นเส้นตรงใน Passband และจะเริ่มค่อยๆ ลดต่ำลงไปทาง 0 ใน Stopband เมื่อถึงจุดที่เป็นความถี่ตัดออก (Cutoff Frequency)



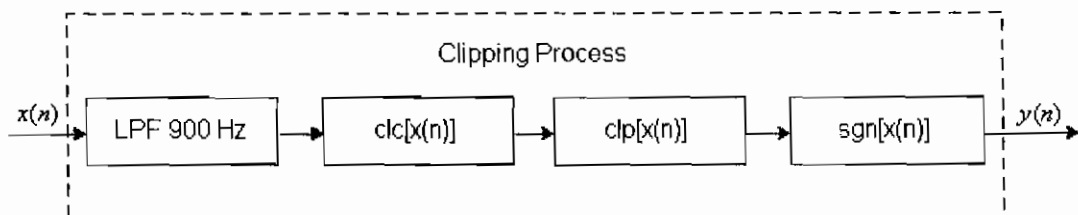
ภาพที่ 2-10 ลักษณะของวงจรกรองความถี่ผ่านต่ำในอุดมคติ



ภาพที่ 2-11 การเปรียบเทียบลักษณะของตัวกรองบัตเตอร์เวิร์ธกับตัวกรองชนิดอื่น

2.3.3 การ Clipping สัญญาณเสียง

กระบวนการ Clipping สัญญาณเสียง ถูกเสนอครั้งแรกโดย M. M. Sondhi [9] ขั้นแรกจะทำการกรองความถี่ผ่านต่ำ (Low-Pass Filter) ที่ความถี่ตัดออก (Cutoff Frequency) 900 Hz เมื่อสัญญาณผ่านการกรองความถี่ต่ำผ่านแล้ว จะเข้าสู่กระบวนการ Clipping สัญญาณโดยมีขั้นตอนวิธีดังนี้



ภาพที่ 2-12 ขั้นตอนวิธีการ Clipping สัญญาณเสียง

ขั้นตอนการทำงานของวิธีการ Clipping สัญญาณเสียงขั้นแรก คือ การทำ Center-Clipping ของเสียงพูด กระบวนการจะทำการบีบอัดสัญญาณนำเข้า $x(n)$ ด้วยฟังก์ชัน $clc[x]$ ผลลัพธ์จากการบีบอัดจะได้สัญญาณ $y(n)$ ซึ่งมีความสัมพันธ์ระหว่างสัญญาณนำเข้า $x(n)$ และ $y(n)$ เป็นดังนี้

$$y(n) = clc[x(n)] = \begin{cases} (x(n) - C_L), & x(n) \geq C_L \\ 0, & |x(n)| < C_L \\ (x(n) + C_L), & x(n) \leq -C_L \end{cases} \quad (2-4)$$

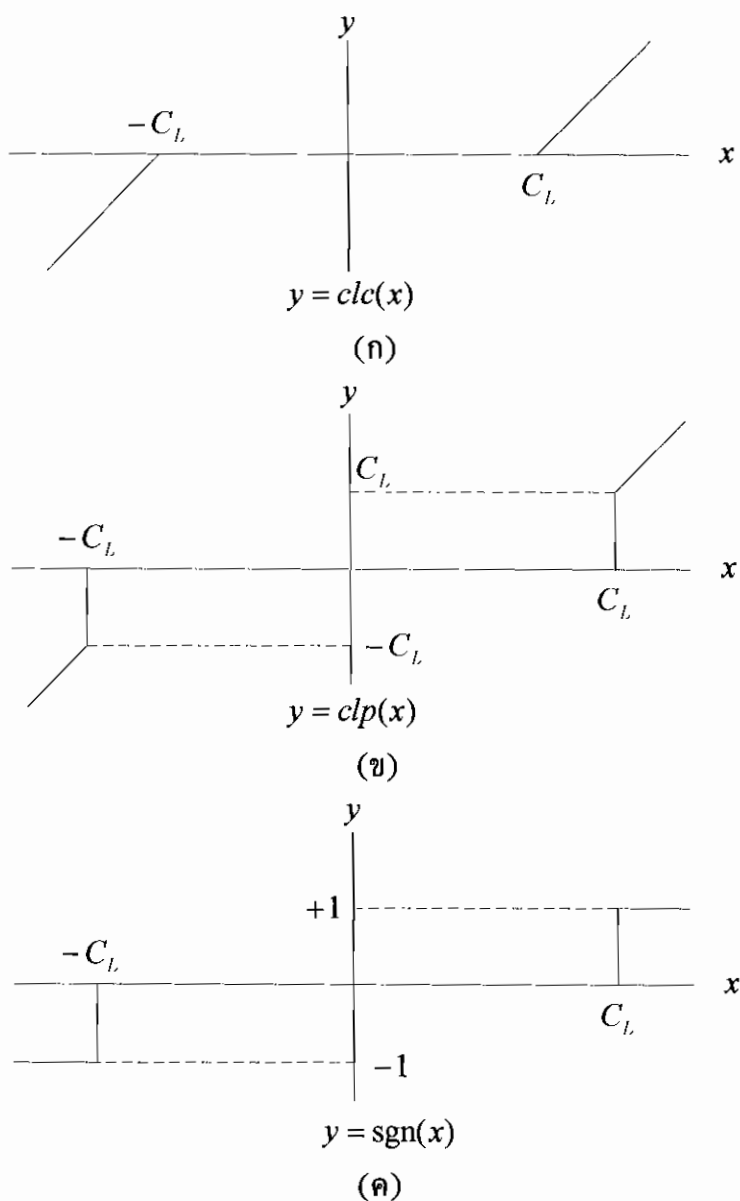
เมื่อ C_L เป็นขอบเขต Clipping ถ้าสัญญาณมากกว่า C_L การบีบอัดสัญญาณจะนำสัญญาณลบออกด้วย C_L ถ้าสัญญาณน้อยกว่า $-C_L$ จะไม่ทำการบีบอัดสัญญาณ ถ้าสัญญาณน้อยกว่าหรือเท่ากับ $-C_L$ จะบวก C_L เข้ากับสัญญาณนั้น ดังแสดงในภาพที่ 2-13 (ก) ลำดับที่สองคือการทำ Infinite-Peak-Clipping จะทำการเล็มสัญญาณที่เกินกว่าศูนย์กลาง หรือเกินกว่าระดับ C_L ตามสมการดังนี้

$$y(n) = clp[x(n)] = \begin{cases} x(n), & x(n) \geq C_L \\ 0, & |x(n)| < C_L \\ -x(n), & x(n) \leq -C_L \end{cases} \quad (2-5)$$

ผลลัพธ์ที่ได้ดังแสดงในภาพที่ 2-13 (ข) สุดท้ายลำดับที่สามเป็นการรวมกันของสองลำดับข้างต้นด้วยความสัมพันธ์ดังนี้

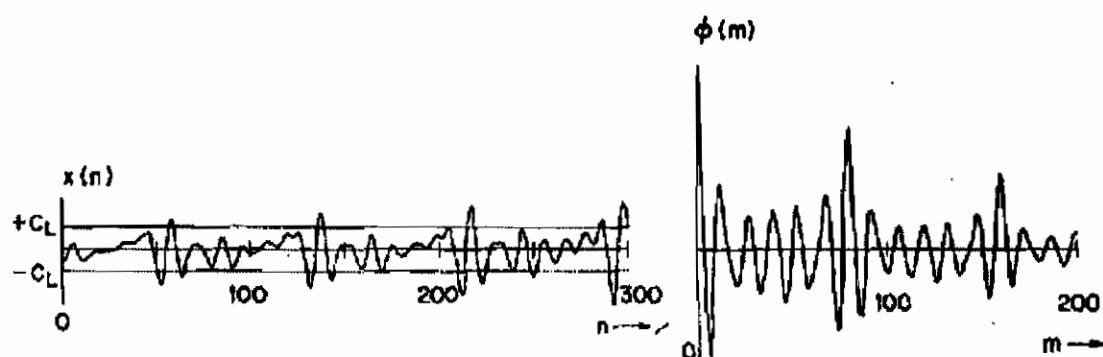
$$y(n) = \text{sgn}[x(n)] = \begin{cases} 1, & x(n) \geq C_L \\ 0, & |x(n)| < C_L \\ -1, & x(n) \leq -C_L \end{cases} \quad (2-6)$$

ผลลัพธ์ที่ได้ดังแสดงในภาพที่ 2-13 (ค) โดยทั่วไป C_L จะอยู่ที่ 30% ของแมกนิจูดสูงสุดของสัญญาณในโปรแกรม C_L ควรจะสูงเพียงพอที่จะเป็นไปได้ ถ้าให้ C_L สูง เราจะสามารถจับค่าพิคตำแหน่งแรกที่ 1/3 ของสัญญาณ และตำแหน่งสุดท้ายที่ 1/3 ของสัญญาณ และใช้จำนวนหนึ่งเป็นแมกนิจูดสูงสุด จากนั้นจะกำหนดค่าแมกนิจูดที่มากที่สุด 60-80% ให้เป็น C_L เสร็จแล้วก็จะได้ผลลัพธ์ของ Center-Clipping และ Infinite-Peak-Clipping

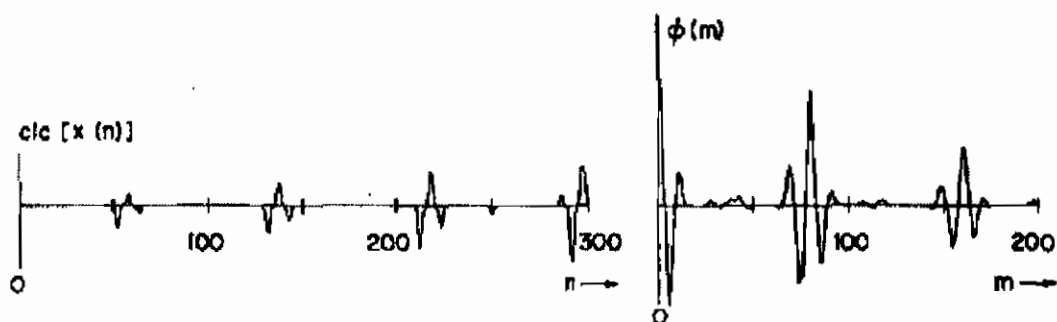


ภาพที่ 2-13 กระบวนการ Clipping สัญญาณเสียง

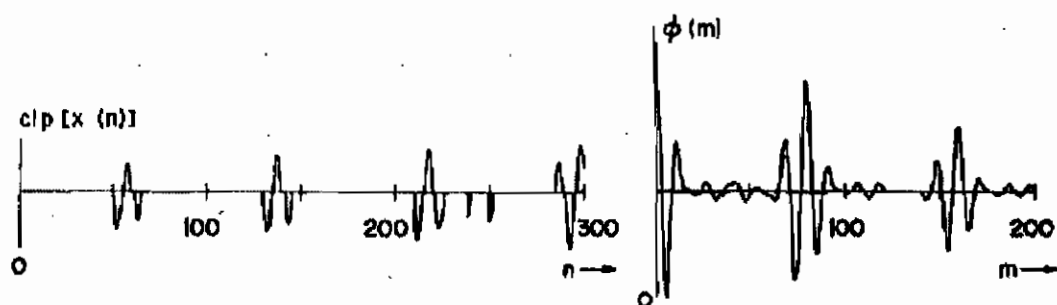
สัญญาณที่ผ่านกระบวนการ Clipping สัญญาณเสียงจะมีลักษณะดังภาพที่ 2-14 จะเห็นว่าในภาพที่ 2-14 (ก) แสดงลักษณะของสัญญาณเสียงนำเข้า $x(n)$ เมื่อสัญญาณนำเข้าถูกบีบอัดด้วยฟังก์ชัน $clc(x)$ สัญญาณจะมีลักษณะดังภาพที่ 2-14 (ข) จากนั้นสัญญาณจะถูกเล็มเฉพาะบริเวณที่บริเวณที่เกินกว่าระดับ C_L ดังในภาพที่ 2-14 (ค) สุดท้ายนำผลลัพธ์จากทั้งสองขั้นตอนรวมกันจะได้ดังภาพที่ 2-14 (ง)



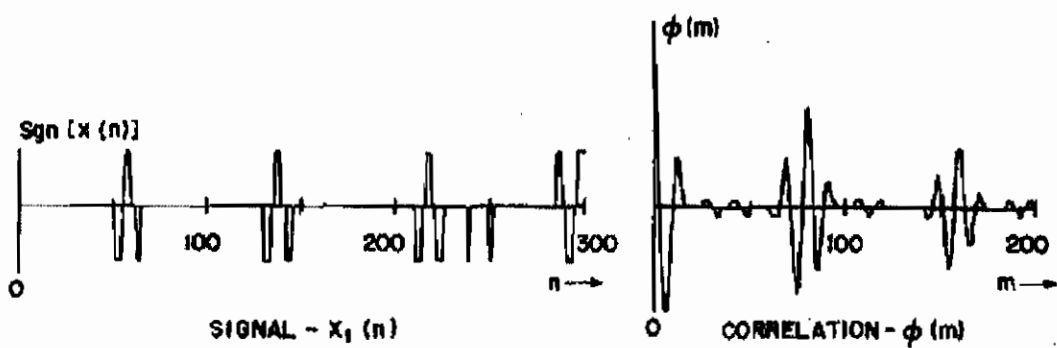
(ก)



(ข)



(ค)



(ง)

ภาพที่ 2-14 ลักษณะสัญญาณเสียงในแต่ละขั้นตอนการ Clipping

2.4 งานวิจัยที่เกี่ยวข้อง

ผู้วิจัยได้ศึกษาค้นคว้างานวิจัยเกี่ยวกับการประมวลผลสัญญาณเสียงพูด เพื่อใช้เป็นพื้นฐานในการวิจัย ซึ่งพอจะสรุปได้ดังนี้

งานวิจัยของ Gerhard [1] ได้รวบรวมเทคนิคที่มีประสิทธิภาพในการประมาณค่าความถี่มูลฐาน (Fundamental Frequency : f_0) และการแยกค่าพิตซ์ที่เหมาะสมสำหรับเสียงพูดของมนุษย์และเสียงจากเครื่องดนตรี กล่าวว่า ตัวประมาณค่าความถี่มูลฐานที่เหมาะสมกับทั้งเสียงพูดและเสียงจากเครื่องดนตรีนั้นยังไม่สามารถทำได้ เนื่องจากความแตกต่างกันในด้านโดเมนที่ใช้ในการออกแบบตัวประมาณค่า โปรแกรมที่ถูกออกแบบมาอย่างมีประสิทธิภาพในโดเมนหนึ่งอาจจะมีประสิทธิภาพลดลงเมื่อนำมาทดลองกับอีกโดเมนหนึ่งก็ได้ ซึ่งการวิจัยนี้ได้รวบรวมการประมาณค่าความถี่มูลฐานออกเป็น 3 วิธีมาตรฐาน คือ วิธีในโดเมนเวลา (Time Domain Method) ซึ่งมีการประมาณค่าความถี่มูลฐานประกอบด้วย วิธี Time-Event Rate Detection, วิธี Autocorrelation, วิธี Phase Space วิธีในโดเมนความถี่ (Frequency Domain) ประกอบด้วย วิธี Component Frequency Ratios, วิธี Filter-Based Methods, วิธี Cepstrum Analysis, วิธี Multi-Resolution Methods และใน Statistical Frequency Domain Methods ประกอบด้วย วิธี Neural Networks, วิธี Maximum Likelihood Estimators ด้วยวิธีการต่าง ๆ จากทั้งสามโดเมนที่กล่าวมาข้างต้น ในงานวิจัยนี้ได้สรุปว่าวิธีการที่มีประสิทธิภาพและได้รับความนิยมมากที่สุดในโดเมนเวลา คือ วิธี Autocorrelation และในโดเมนความถี่ คือ วิธี Cepstrum Analysis ดังจะเห็นประสิทธิภาพได้จากโปรแกรม Colea ที่พัฒนาโดย Philip Loizou [2] ที่ใช้โปรแกรม MATLAB ในการประมวลผล และ โปรแกรม Terez's [3,4] ซึ่งใช้โปรแกรม MATLAB ในการประมวลผลเช่นกัน โดยทั้งสองโปรแกรมได้เลือกใช้วิธี Autocorrelation และ วิธี Cepstrum เป็นตัวประมาณค่า ซึ่งผลลัพธ์ที่ได้จากการทดลองมีประสิทธิภาพเป็นที่น่าพอใจ

งานวิจัยของ Parris และ Carey [13] ทำการพัฒนาเทคนิคใหม่สำหรับระบุเพศบุคคลด้วยการวิเคราะห์เสียงพูดและค่าพิตซ์ของสัญญาณเสียง ด้วยการใช้โมเดลเสียงพูดเพศชาย และเพศหญิงด้วย Hidden Markov Model และจับคู่ความเหมือนของเสียงพูดโดยใช้ขั้นตอนวิธี Viterbi จากนั้นปรับปรุงกระบวนการประมาณค่าพิตซ์ด้วยตัวจำแนกแบบลิเนียร์เพื่อระบุเพศบุคคลจากเสียงพูด ระบบจะถูกทดสอบจากเสียงพูดภาษาอังกฤษจากฐานข้อมูลประสิทธิภาพการจำแนกเพศบุคคลจากเสียงพูดความยาว 2 วินาทีต่ำกว่ามีอัตราความผิดพลาดต่ำกว่า 1 % และนำไปทดสอบกับ 11 ภาษาพูดจากฐานข้อมูล OGI พบว่ามีอัตราความผิดพลาดต่ำกว่า 5.2 %

งานวิจัยของ Boersma [14] ทำการวิจัยเกี่ยวกับขั้นตอนวิธีที่มีประสิทธิภาพสำหรับการตรวจจับสัญญาณแบบมีคาบทำงานใน Lag Domain หรือ Autocorrelation เมื่อนำมาทดสอบกับสัญญาณแบบมีคาบ และสัญญาณที่เพิ่มเสียงรบกวนเข้าไป ขั้นตอนวิธีนี้ก็สามารถจับแมกนิจูดได้อย่างแม่นยำกว่าวิธีอื่นที่ใช้กับการวิเคราะห์เสียงพูด ขั้นตอนวิธีนี้สามารถวัดอัตรา Harmonics-to-Noise ใน Lag Domain ได้อย่างแม่นยำและเชื่อถือได้มากกว่าวิธีอื่นที่ใช้ในโดเมนความถี่

งานวิจัยของ Tan และ Kamjanadecha [5] นำเสนอเทคนิคการจับค่าพิตช์โดยใช้วิธี Autocorrelation และ วิธี AMDF (Average Magnitude Difference Function) โดยนำวิธีการ Clipping สัญญาณเสียงและเทคนิคการแยกรูปแบบพิตช์ (Pitch Pattern) มาเพิ่มประสิทธิภาพให้กับกระบวนการประมวลผลสัญญาณขั้นต้นก่อนนำไปประมาณค่าพิตช์ด้วยวิธี Autocorrelation และวิธี AMDF สำหรับเสียงพูดภาษาจีน จากการวิจัยพบว่ากระบวนการประมวลผลสัญญาณเบื้องต้นด้วยวิธีการ Clipping สัญญาณเสียงและเทคนิคการแยกรูปแบบพิตช์ (Pitch Pattern) สามารถนำมาใช้งานร่วมกับวิธี Autocorrelation และวิธี AMDF ได้อย่างมีประสิทธิภาพ โดยมีค่าความถูกต้องอยู่ที่ 79.02 %

2.5 วิธีการที่นำเสนอในการวิจัย

จากความรู้ที่ได้จากการศึกษางานวิจัยเกี่ยวกับเสียงพูด ผู้วิจัยทำการทดสอบประสิทธิภาพของกระบวนการหาค่าพิตช์ด้วยวิธี Autocorrelation และวิธีการวิเคราะห์แบบ Cepstrum โดยมีการประมวลผลสัญญาณเบื้องต้นด้วยการวางกรอบสัญญาณ เพื่อให้ทราบประสิทธิภาพของแต่ละวิธี จากนั้นนำการ Clipping สัญญาณเสียงเพิ่มเข้ามาในส่วนของการประมวลผลสัญญาณเบื้องต้นแล้วนำไปใช้งานร่วมกับวิธี Autocorrelation ตามงานวิจัยของ Tan และ Kamjanadecha แต่ในงานวิจัยครั้งนี้ได้นำเอากระบวนการ Clipping สัญญาณเสียงมาประยุกต์ใช้งานร่วมกับการวิเคราะห์แบบ Cepstrum ด้วยเพื่อเปรียบเทียบประสิทธิภาพ จากนั้นนำกระบวนการลบ DC Bias มาใช้งานร่วมกับตัวกรองบัตเตอร์เวิร์ธในการประมวลผลสัญญาณเบื้องต้นเป็นขั้นตอนวิธีสุดท้าย โดยทั่วไปขั้นตอนวิธีการประมวลผลสัญญาณเบื้องต้นจะเน้นการกำจัดสัญญาณฮาร์โมนิกที่มีความถี่สูงเกินกว่าระดับที่สนใจโดยไม่คำนึงถึงสัญญาณ Bias ที่แฝงอยู่ในสัญญาณทำให้ยังพบปัญหาความคลาดเคลื่อนของการประมาณค่าความถี่มูลฐาน เนื่องจากสัญญาณ Bias เป็นสัญญาณที่ทำให้ค่าความถี่ของสัญญาณเสียงลอยตัวขึ้นเกินกว่าค่าที่แท้จริง ดังนั้นการกำจัดสัญญาณ Bias จะทำให้การประมาณค่าความถี่มูลฐานมีความถูกต้องมากขึ้น จากแนวคิดดังกล่าวผู้วิจัยจึงเพิ่มประสิทธิภาพในการประมวลผลสัญญาณเบื้องต้นด้วยการนำกระบวนการลบ DC Bias มาใช้งานร่วมกับตัวกรองบัตเตอร์เวิร์ธและการ Clipping สัญญาณเสียงเพื่อให้สัญญาณเสียงที่ผ่านกระบวนการประมวลผลสัญญาณเบื้องต้นสามารถกำจัดสัญญาณ Bias และฮาร์โมนิกพลังงานสูงอื่น ๆ ที่ปะปนอยู่ออกไปได้มากที่สุดแล้วจึงนำเข้าสู่กระบวนการหาค่าพิตช์ของเสียงพูดด้วยสองวิธีข้างต้นเพื่อให้ได้ผลลัพธ์ที่มีประสิทธิภาพมากที่สุด

บทที่ 3

วิธีการดำเนินการวิจัย

จากปัญหาในการหาค่าความถี่มูลฐานในบทที่ 1 และทฤษฎีที่เกี่ยวข้องกับการหาค่าความถี่มูลฐานในบทที่ 2 ผู้วิจัยจึงได้ทำการปรับปรุงขั้นตอนวิธีในการประมวลผลสัญญาณเบื้องต้นของสัญญาณเสียงพูดเพื่อนำไปใช้กับการจำแนกเพศบุคคลด้วยการตรวจจับค่าพิทช์ของสัญญาณเสียงพูดภาษาไทยโดยมีขั้นตอนการทำงานดังนี้

3.1 การเตรียมข้อมูลเสียงพูด

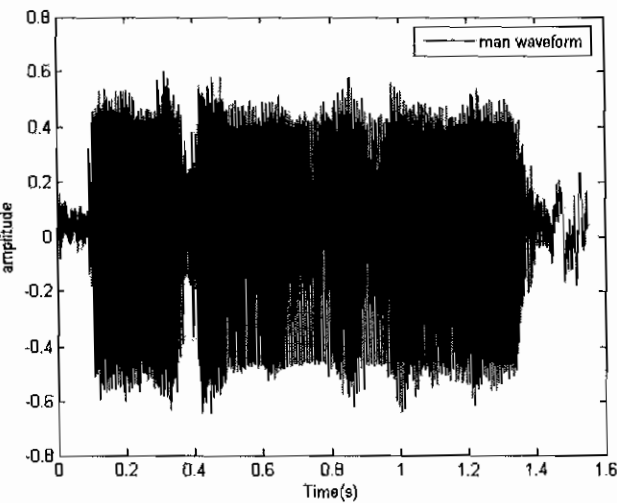
ในงานวิจัยครั้งนี้ ผู้วิจัยได้ทำการเก็บตัวอย่างเสียงพูดภาษาไทยโดยใช้การบันทึกเสียงผ่านเครื่องบันทึกเสียงดิจิทัล จากนั้นนำสัญญาณเสียงที่ได้แปลงเป็น WAV File ด้วยโปรแกรม Adobe Audition 1.5 กำหนดลักษณะของเสียงพูดให้เป็นเสียงประเภท Mono มีอัตราสุ่มตัวอย่าง (Sampling Rate) เท่ากับ 16,000 มีความละเอียด (Resolution) เป็น 16 บิต แบ่งสัญญาณเสียงจากการบันทึกตามการทดลองออกเป็น 2 ประเภท คือ

3.1.1 สัญญาณเสียงพูดประโยคเดียวกันด้วยประโยคคำว่า “กาลครั้งหนึ่งนานมาแล้ว” บันทึกจากกลุ่มตัวอย่างจำนวน 200 คน แบ่งเป็นเพศชายจำนวน 90 คน และเพศหญิงจำนวน 110 คน จากนั้นแบ่งออกเป็นสัญญาณเสียงพูดสำหรับกลุ่มเรียนรู้เพศชายจำนวน 20 เสียงพูด และเพศหญิงจำนวน 20 เสียงพูด สัญญาณเสียงพูดทดสอบเพศชายจำนวน 70 เสียงพูด และเพศหญิงจำนวน 90 เสียงพูด

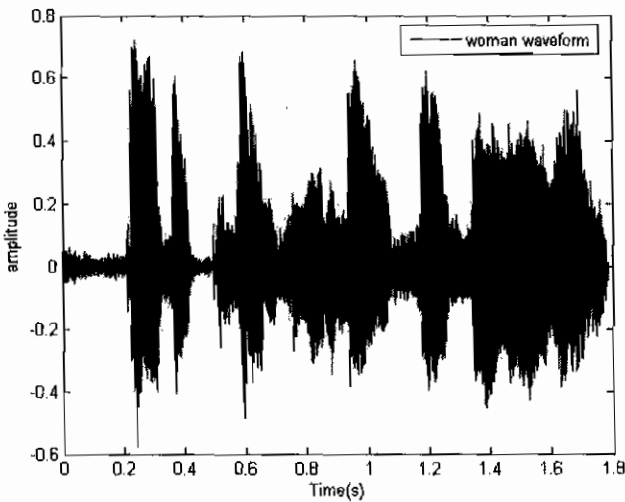
3.1.2 สัญญาณเสียงพูดต่างประโยคกันด้วยประโยคแนะนำตัวของกลุ่มตัวอย่างจำนวน 208 คนแบ่งเป็นเพศชายจำนวน 81 คน และเพศหญิงจำนวน 127 คน จากนั้นแบ่งออกเป็นสัญญาณเสียงพูดสำหรับกลุ่มเรียนรู้เพศชายจำนวน 20 เสียงพูด และเพศหญิงจำนวน 20 เสียงพูด สัญญาณเสียงพูดทดสอบเพศชายจำนวน 61 เสียงพูด และเพศหญิงจำนวน 107 เสียงพูด

ตารางที่ 3-1 จำนวนเสียงพูดจากการเก็บข้อมูลกลุ่มตัวอย่าง

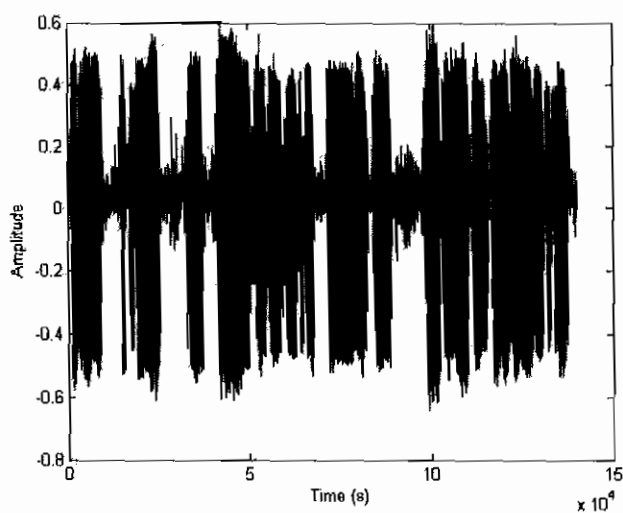
ประโยคคำพูด	จำนวนเสียงพูดเพศชาย		จำนวนเสียงพูดเพศหญิง	
	เรียนรู้	ทดสอบ	เรียนรู้	ทดสอบ
“กาลครั้งหนึ่งนานมาแล้ว”	20	70	20	90
“แนะนำตัว”	20	61	20	107
รวม (หน่วยเป็นเสียงพูด)	40	131	40	197



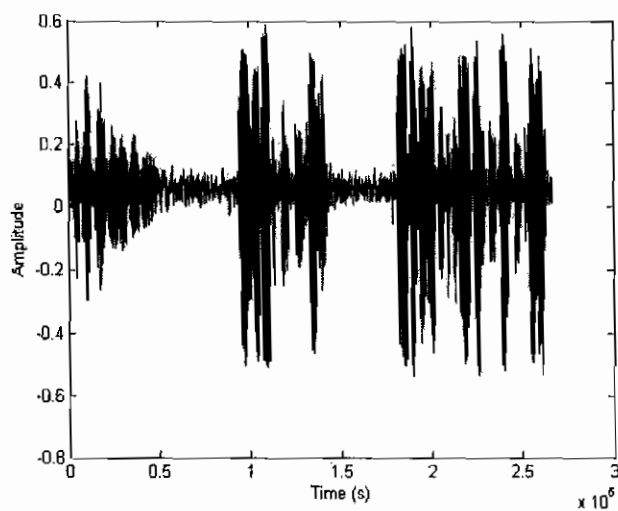
ภาพที่ 3-1 ตัวอย่างสัญญาณเสียงพูดนำเข้าเพศชายประโยค “กาลครั้งหนึ่งนานมาแล้ว”



ภาพที่ 3-2 ตัวอย่างสัญญาณเสียงพูดนำเข้าเพศหญิงประโยค “กาลครั้งหนึ่งนานมาแล้ว”



ภาพที่ 3-3 ตัวอย่างสัญญาณเสียงพูดนำเข้าเพศชายสำหรับประโยคคำพูดที่ต่างกันคำว่า
“ศุภวุฒิ สุขสวัสดิ์ เพศชาย อายุ 21 ปี วันเกิด 5 ธันวาคม 2526”



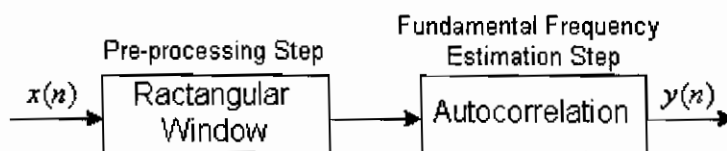
ภาพที่ 3-4 ตัวอย่างสัญญาณเสียงพูดนำเข้าเพศหญิงสำหรับประโยคคำพูดที่ต่างกันคำว่า
“สุภาภรณ์ ไชยาพรพรรณ อายุ 54 ปี เกิดวันที่ 24 มิถุนายน 2494 ค่ะ”

3.2 กระบวนการประมาณค่าความถี่มูลฐาน (Fundamental Frequency Estimation Process)

ความพยายามในการเพิ่มประสิทธิภาพขั้นตอนวิธีการประมวลผลสัญญาณเบื้องต้นสำหรับกระบวนการประมาณค่าความถี่มูลฐานที่นิยมกันมีหลายวิธี ในการวิจัยครั้งนี้จึงนำขั้นตอนวิธีดังกล่าวมาใช้งานร่วมกับกระบวนการประมาณค่าความถี่มูลฐานที่เลือกใช้ พร้อมทั้งนำวิธีการที่นำเสนอมาทดลองด้วยเช่นกัน เพื่อเปรียบเทียบประสิทธิภาพของแต่ละขั้นตอนวิธี โดยแบ่งการทดลองเป็น 8 ขั้นตอนวิธี ดังนี้

3.2.1 ขั้นตอนวิธีการหาค่าพิตซ์ของเสียงพูดด้วยวิธี Autocorrelation โดยมีการประมวลผลสัญญาณเบื้องต้นด้วยการวางฟังก์ชันกรอบสัญญาณแบบสี่เหลี่ยม

ขั้นตอนวิธีที่ 1 นี้สัญญาณเสียงนำเข้าจะถูกแบ่งออกเป็นเฟรมแต่ละเฟรมมีขนาด 30 ms ส่วนซ้อนทับกันมีขนาด 20 ms เพื่อทำให้สัญญาณเสียงในแต่ละเฟรมสามารถครอบคลุมคาบสัญญาณเสียงได้ไม่ต่ำกว่า 4 คาบสัญญาณเสียงต่อ 1 เฟรม จากนั้นเลือกใส่ฟังก์ชันกรอบสัญญาณแบบสี่เหลี่ยมเนื่องจากผลการทดสอบประสิทธิภาพการใช้กรอบสัญญาณแบบสี่เหลี่ยมเปรียบเทียบกับกรอบสัญญาณแบบแฮมมิงเมื่อนำมาประมาณค่าความถี่มูลฐานด้วยวิธี Autocorrelation แล้วประสิทธิภาพในการประมาณค่าความถี่มูลฐานเมื่อมีการประมวลผลสัญญาณเบื้องต้นด้วยกรอบสัญญาณแบบสี่เหลี่ยมมีประสิทธิภาพดีกว่ากรอบสัญญาณแบบแฮมมิง ผู้วิจัยจึงตัดสินใจเลือกใช้กรอบสัญญาณแบบสี่เหลี่ยมในการประมาณค่าความถี่มูลฐานกับทุกขั้นตอนวิธีที่มีการประมาณค่าความถี่มูลฐานด้วยวิธี Autocorrelation ผลการทดสอบประสิทธิภาพของกรอบสัญญาณแบบสี่เหลี่ยมกับกรอบสัญญาณแบบแฮมมิงแสดงในภาคผนวก ก จากนั้นเมื่อสัญญาณเสียงผ่านการวางกรอบสัญญาณแบบสี่เหลี่ยมแล้วจะถูกนำเข้าสู่กระบวนการคำนวณหาค่าความถี่มูลฐานด้วยวิธี Autocorrelation ตามที่ได้กล่าวไว้แล้วในบทที่ 2 เมื่อได้ค่าพิตซ์ของสัญญาณเสียงนั้นแล้วสัญญาณเสียงจะถูกส่งต่อไปยังกระบวนการจำแนกเพศบุคคลต่อไป

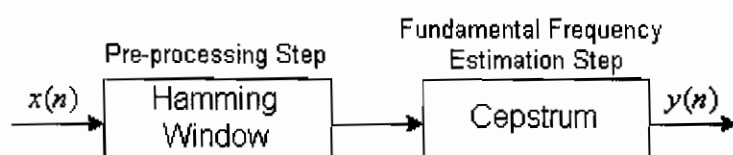


ภาพที่ 3-5 ขั้นตอนวิธีที่ 1

3.2.2 ขั้นตอนวิธีการหาค่าพิตซ์ของเสียงพูดด้วยวิธี Cepstrum โดยมีการประมวลผลสัญญาณเบื้องต้นด้วยการวางฟังก์ชันกรอบสัญญาณแบบแฮมมิง

ขั้นตอนวิธีที่ 2 นี้สัญญาณเสียงนำเข้าจะถูกแบ่งออกเป็นเฟรมแต่ละเฟรมมีขนาด 40 ms ส่วนซ้อนทับกันมีขนาด 20 ms เพื่อทำให้สัญญาณเสียงในแต่ละเฟรมมีความคงที่ ณ เวลาที่พิจารณามากที่สุด จากนั้นเลือกใส่ฟังก์ชันกรอบสัญญาณแบบแฮมมิงเนื่องจากผลการทดสอบประสิทธิภาพการใช้กรอบสัญญาณแบบสี่เหลี่ยมเปรียบเทียบกับกรอบสัญญาณแบบแฮมมิงเมื่อ

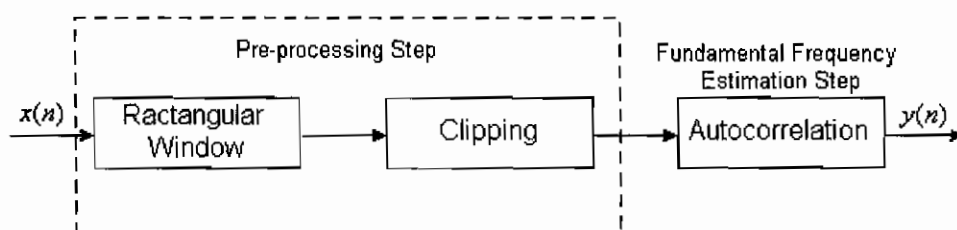
นำมาประมาณค่าความถี่มูลฐานด้วยวิธี Cepstrum แล้วประสิทธิภาพในการประมาณค่าความถี่มูลฐานเมื่อมีการประมวลผลสัญญาณเบื้องต้นด้วยฟังก์ชันกรอบสัญญาณแบบแฮมมิงมีประสิทธิภาพดีกว่าฟังก์ชันกรอบสัญญาณแบบสี่เหลี่ยม ผู้วิจัยจึงเลือกใช้ฟังก์ชันกรอบสัญญาณแบบแฮมมิงในการประมาณค่าความถี่มูลฐานกับทุกขั้นตอนวิธีที่มีการประมาณค่าความถี่มูลฐานด้วยวิธี Cepstrum ผลการทดสอบประสิทธิภาพของฟังก์ชันกรอบสัญญาณแบบสี่เหลี่ยมกับฟังก์ชันกรอบสัญญาณแบบแฮมมิงแสดงในภาคผนวก ก จากนั้นเมื่อสัญญาณเสียงผ่านฟังก์ชันกรอบสัญญาณแบบแฮมมิงแล้วจะถูกนำเข้าสู่กระบวนการคำนวณหาความถี่มูลฐานด้วยวิธี Cepstrum ที่ได้กล่าวไว้แล้วในบทที่ 2 เมื่อได้ค่าพิตซ์ของสัญญาณเสียงนั้นแล้วสัญญาณเสียงจะถูกส่งไปสู่กระบวนการจำแนกเพศบุคคลต่อไป



ภาพที่ 3-6 ขั้นตอนวิธีที่ 2

3.2.3 ขั้นตอนวิธีการหาค่าพิตซ์ของเสียงพูดด้วยวิธี Autocorrelation โดยมีการประมวลผลสัญญาณเบื้องต้นด้วยการวางฟังก์ชันกรอบสัญญาณและการ Clipping สัญญาณเสียง

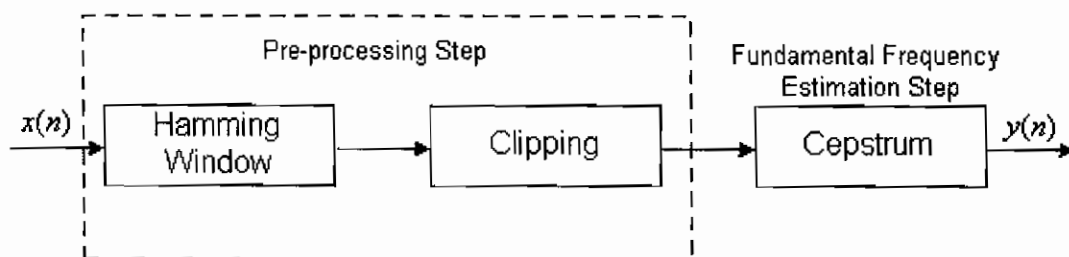
ขั้นตอนวิธีที่ 3 นี้ในขั้นตอนการประมวลผลสัญญาณเบื้องต้นสัญญาณเสียงจะถูกแบ่งเฟรมและวางฟังก์ชันกรอบสัญญาณแบบสี่เหลี่ยมจากนั้นสัญญาณเสียงจะเข้าสู่กระบวนการ Clipping สัญญาณเสียงโดยทำการกรองสัญญาณเสียงนำเข้าด้วยตัวกรองความถี่ผ่านต่ำ โดยกำหนดความถี่ตัดออก (Cutoff Frequency) ที่ 900 Hz เพื่อกรองฮาร์โมนิกที่มีความสูงเกินกว่าขอบเขตที่ต้องการพิจารณาออกไป จากนั้นทำการ Clipping สัญญาณเสียงตามทฤษฎีในบทที่ 2 แล้วจึงทำการประมาณค่าความถี่มูลฐานด้วยวิธี Autocorrelation เมื่อได้ค่าพิตซ์ของสัญญาณเสียงนั้นแล้วสัญญาณเสียงจะถูกส่งไปสู่กระบวนการจำแนกเพศบุคคลต่อไป



ภาพที่ 3-7 ขั้นตอนวิธีที่ 3

3.2.4 ขั้นตอนวิธีการหาค่าพิตช์ของเสียงพูดด้วยวิธี Cepstrum โดยมีการประมวลผลสัญญาณเบื้องต้นด้วยการวางฟังก์ชันกรอบสัญญาณและการ Clipping สัญญาณเสียง

ขั้นตอนวิธีที่ 4 นี้สัญญาณเสียงนำเข้าจะถูกแบ่งออกเป็นเฟรมและวางฟังก์ชันกรอบสัญญาณแบบแฮมมิง จากนั้นเมื่อสัญญาณเสียงผ่านการวางกรอบสัญญาณแบบแฮมมิงแล้ว จากนั้นสัญญาณเสียงจะเข้าสู่กระบวนการ Clipping สัญญาณเสียงตามทฤษฎีในบทที่ 2 จากนั้นจึงทำการประมาณค่าความถี่มูลฐานด้วยวิธี Cepstrum เมื่อได้ค่าพิตช์ของสัญญาณเสียงนั้นแล้ว สัญญาณเสียงจะถูกส่งไปสู่กระบวนการจำแนกเพศบุคคลต่อไป



ภาพที่ 3-8 ขั้นตอนวิธีที่ 4

3.2.5 ขั้นตอนวิธีการหาค่าพิตช์ของเสียงพูดด้วยวิธี Autocorrelation โดยมีการประมวลผลสัญญาณเบื้องต้นด้วยการลบ DC Bias และฟังก์ชันการวางกรอบสัญญาณ

ขั้นตอนวิธีที่ 5 นี้ในขั้นตอนการประมวลผลสัญญาณเบื้องต้นจะถูกเพิ่มประสิทธิภาพด้วย 2 กระบวนการดังนี้คือ

3.2.5.1 กระบวนการลบ DC Bias

สารโมโนคอื่นที่ติดมากับสัญญาณเสียงที่เข้าสู่อุปกรณ์ประมวลผลสัญญาณนั้น ส่วนหนึ่งเป็นสัญญาณ Bias ที่ทำให้ค่าความถี่ของสัญญาณเสียงลอยตัวขึ้นเกินกว่าค่าที่แท้จริง ซึ่งส่งผลกระทบต่อกระบวนการประมาณค่าความถี่มูลฐานและยังส่งผลกระทบต่อไปยังกระบวนการจำแนกเพศบุคคล เนื่องจากระดับสัญญาณเสียงพูดเพศชายที่มีระดับเสียงสูงกว่าปกติ ส่วนใหญ่เกิดในกรณีเสียงพูดเพศชายอายุน้อย หรือมีการดัดเสียง เมื่อมีสัญญาณ Bias ปะปนอยู่ด้วยจะทำให้สัญญาณเสียงมีระดับสูงยิ่งขึ้นทำให้การประมาณค่าความถี่มูลฐานได้ผลลัพธ์เป็นค่าระดับเสียงที่สูงเกินกว่าความเป็นจริง เมื่อค่าจากการประมาณค่าความถี่มูลฐานนั้นเข้าสู่กระบวนการจำแนกเพศบุคคลจึงเกิดความผิดพลาดจากการจำแนกขึ้น เนื่องจากระดับสัญญาณเสียงนั้นจะเข้าสู่เกณฑ์ระดับเสียงพูดเพศหญิง ดังนั้นหากกำจัดสัญญาณ Bias ออกจากสัญญาณเสียงแล้วจะส่งผลให้ค่าจากการประมาณค่าความถี่มูลฐานมีความใกล้เคียงกับระดับเสียงที่แท้จริงและผลลัพธ์จากการจำแนกเพศบุคคลมีความถูกต้องมากยิ่งขึ้น จากผลกระทบข้างต้นจึงจำเป็นต้องทำการกำจัดสัญญาณ Bias ออกจากสัญญาณเสียงที่พิจารณาในขั้นตอนการประมวลผลสัญญาณเบื้องต้นเสียก่อน กระบวนการลบ DC Bias จะทำการหาค่าเฉลี่ยของสัญญาณเสียง แล้วนำค่าเฉลี่ยของ

สัญญาณเสียงนั้นลบออกจากสัญญาณเสียงที่พิจารณา เพื่อเป็นการกำจัดสัญญาณ Bias ภายในสัญญาณเสียง ดังสมการต่อไปนี้

$$b = \frac{1}{N} \sum_{n=1}^N X_n \quad (3-1)$$

$$S = \sum_{n=1}^N (X_n - b) \quad (3-2)$$

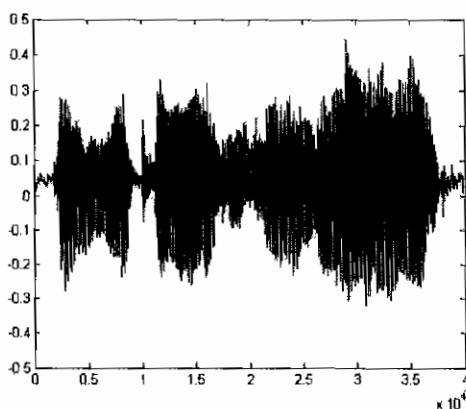
เมื่อ X_n คือ ค่าของสัญญาณเสียง ณ ตำแหน่งที่ n

N คือ จำนวนเฟรมของสัญญาณเสียง

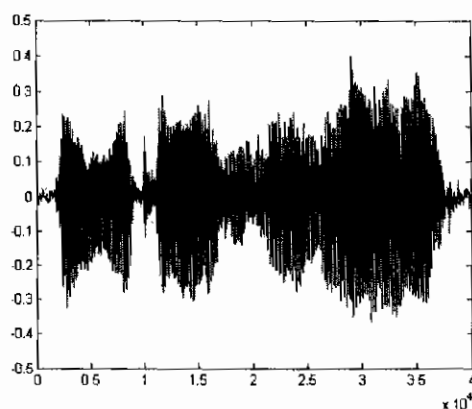
b คือ ผลรวมของตัวอย่างของสัญญาณเสียง

S คือ สัญญาณเสียงหลังการลบ DC Bias

จากภาพที่ 3-9 (ก) คือ สัญญาณเสียงนำเข้า และภาพ (ข) เป็นสัญญาณเสียงที่ผ่านการลบ DC Bias แล้ว จะเห็นว่าระดับสัญญาณเสียงจะลดต่ำลงเมื่อสัญญาณ Bias ถูกกำจัดออกไป แม้ความแตกต่างจะมองเห็นเพียงเล็กน้อย แต่ส่วนเล็กน้อยนั้นกลับส่งผลกระทบต่อกระบวนการประมาณค่าความถี่มูลฐานและยังส่งผลกระทบต่อประสิทธิภาพของกระบวนการจำแนกเพศบุคคล



(ก)



(ข)

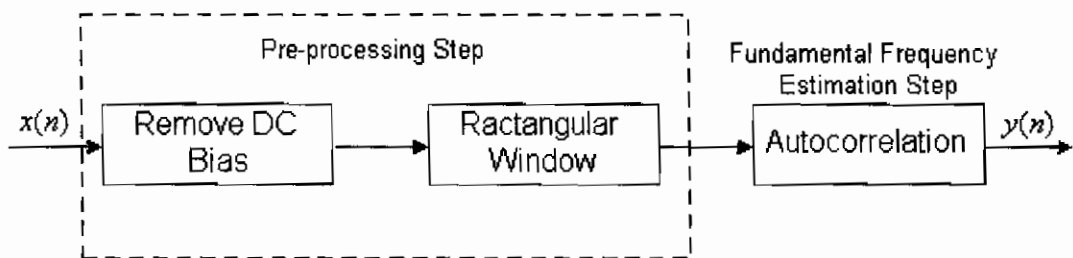
ภาพที่ 3-9 ตัวอย่างสัญญาณเสียงก่อนและหลังลบ DC bias

3.2.5.2 ตัวกรองบัตเตอร์เวิร์ธ

การนำตัวกรองบัตเตอร์เวิร์ธมาใช้งานร่วมกับการลบ DC Bias เพื่อช่วยกรองฮาร์โมนิคที่มีความถี่สูงเกินกว่า 900 Hz เนื่องจากความถี่มูลฐานของเสียงพูดส่วนใหญ่จะอยู่ในช่วงความถี่ระหว่าง 50-500 Hz [2] ผู้วิจัยจึงกำหนดระยะความถี่ตัดออก (Cutoff Frequency) ที่ 900 Hz

เพื่อกรองความถี่ที่สูงเกินกว่าขอบเขตที่สนใจออกไป พร้อมทั้งเพื่อระยะความถี่ช่วงระหว่าง 500–900 Hz ไว้สำหรับกรณีความถี่มูลฐานของเสียงพูดที่พิจารณาอาจจะสูงเกินกว่าระดับทั่วไปเพื่อป้องกันความผิดพลาดจากการกรองสัญญาณเสียงในส่วนที่เป็นสาระสำคัญออกไป

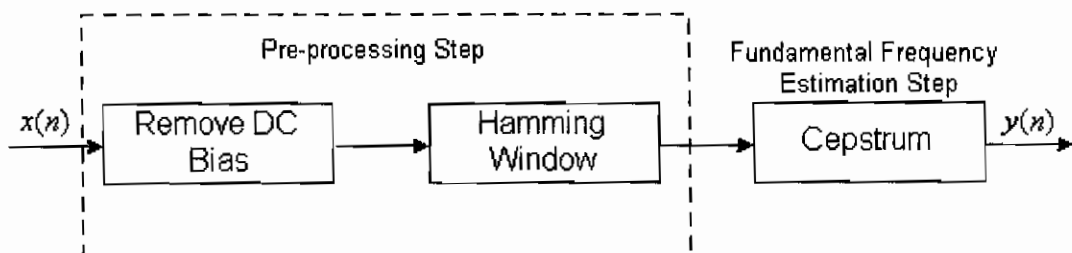
เมื่อสัญญาณเสียงนำเข้าผ่านกระบวนการลบ DC Bias และผ่านการกรองด้วยตัวกรองบัตเตอร์เวิร์ธแล้ว จะทำการแบ่งเฟรมและวางฟังก์ชันกรอบสัญญาณแบบสี่เหลี่ยม เมื่อผ่านกระบวนการประมวลผลสัญญาณเบื้องต้นแล้วจึงทำการประมาณค่าความถี่มูลฐานด้วยวิธี Autocorrelation เพื่อหาค่าพิตซ์ของสัญญาณเสียงนั้น จากนั้นจึงเข้าสู่กระบวนการจำแนกเพศบุคคล



ภาพที่ 3-10 ขั้นตอนวิธีที่ 5

3.2.6 ขั้นตอนวิธีการหาค่าพิตซ์ของเสียงพูดด้วยวิธี Cepstrum โดยมีการประมวลผลสัญญาณเบื้องต้นด้วยการลบ DC Bias และการวางกรอบสัญญาณ

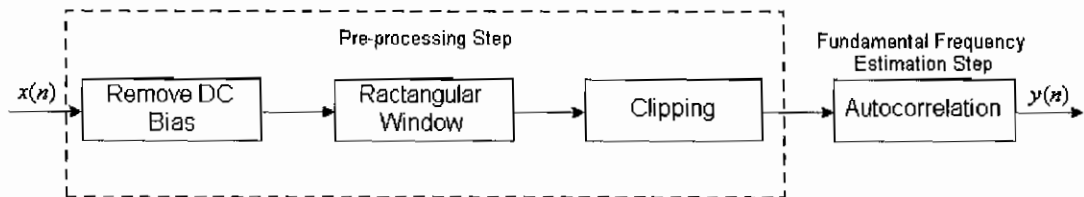
ขั้นตอนวิธีที่ 6 นี้สัญญาณเสียงนำเข้าจะเข้าสู่กระบวนการประมวลผลสัญญาณเบื้องต้นด้วยการลบ DC Bias และการกรองสัญญาณด้วยตัวกรองบัตเตอร์เวิร์ธ จากนั้นสัญญาณเสียงพูดจะถูกแบ่งเฟรมและวางฟังก์ชันกรอบสัญญาณแบบแฮมมิง แล้วจึงเข้าสู่กระบวนการประมาณค่าความถี่มูลฐานด้วยวิธี Cepstrum เมื่อได้ค่าพิตซ์ของสัญญาณเสียงนั้นแล้วสัญญาณเสียงจะถูกส่งไปสู่กระบวนการจำแนกเพศบุคคลต่อไป



ภาพที่ 3-11 ขั้นตอนวิธีที่ 6

3.2.7 ขั้นตอนวิธีการหาค่าพิตช์ของเสียงพูดด้วยวิธี Autocorrelation โดยมีการประมวลผลสัญญาณเบื้องต้นด้วยการลบ DC Bias ร่วมกับการวางฟังก์ชันกรอบสัญญาณ และการ Clipping สัญญาณเสียง

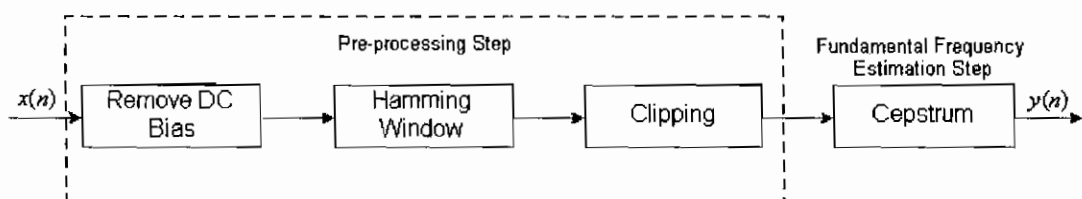
ขั้นตอนวิธีที่ 7 นี้ผู้วิจัยได้นำกระบวนการประมวลผลสัญญาณเบื้องต้นทุกวิธีข้างต้นมาใช้งานร่วมกันเพื่อเพิ่มประสิทธิภาพให้กับกระบวนการประมวลผลสัญญาณเบื้องต้น สัญญาณเสียงนำเข้าจะถูกลบ DC Bias และถูกกรองสัญญาณด้วยตัวกรองบัตเตอร์เวิร์ธ จากนั้นสัญญาณเสียงพูดจะถูกแบ่งเฟรมและวางฟังก์ชันกรอบสัญญาณแบบสี่เหลี่ยม จากนั้นสัญญาณเสียงพูดจะถูก Clipping สัญญาณเสียงก่อนแล้วจึงเข้าสู่กระบวนการประมาณค่าความถี่มูลฐานด้วยวิธี Autocorrelation เมื่อได้ค่าพิตช์ของสัญญาณเสียงนั้นแล้วสัญญาณเสียงจะถูกส่งไปสู่กระบวนการจำแนกเพศบุคคลต่อไป



ภาพที่ 3-12 ขั้นตอนวิธีที่ 7

3.2.8 ขั้นตอนวิธีการหาค่าพิตช์ของเสียงพูดด้วยวิธี Cepstrum โดยมีการประมวลผลสัญญาณเบื้องต้นด้วยการลบ DC Bias ร่วมกับการวางฟังก์ชันกรอบสัญญาณ และการ Clipping สัญญาณเสียง

ขั้นตอนวิธีที่ 8 นี้กระบวนการประมวลผลสัญญาณเบื้องต้นจะทำการลบ DC Bias และกรองสัญญาณด้วยตัวกรองบัตเตอร์เวิร์ธ จากนั้นสัญญาณเสียงพูดจะถูกแบ่งเฟรมและวางฟังก์ชันกรอบสัญญาณแบบแฮมมิง จากนั้นสัญญาณเสียงพูดจะถูก Clipping สัญญาณเสียงแล้วเข้าสู่กระบวนการประมาณค่าความถี่มูลฐานด้วยวิธี Cepstrum เพื่อให้ได้ค่าพิตช์ของสัญญาณเสียงนั้น แล้วจึงเข้าสู่กระบวนการจำแนกเพศบุคคลต่อไป



ภาพที่ 3-13 ขั้นตอนวิธีที่ 8

3.3 กระบวนการจำแนกเพศบุคคล (Gender Classification Process)

ในกระบวนการจำแนกเพศบุคคลนี้ สามารถแยกออกได้เป็น 2 กระบวนการ คือ

3.3.1 กระบวนการเรียนรู้ (Training Process)

ขั้นตอนวิธีในกระบวนการเรียนรู้ถูกแบ่งออกเป็นกระบวนการเรียนรู้เพื่อการจำแนกเพศบุคคลสำหรับเสียงพูดไทยด้วยคำพูดประโยคเดียวกัน และเสียงพูดไทยด้วยคำพูดต่างประโยคกัน โดยทั้งสองกรณี มีขั้นตอนวิธีเหมือนกันต่างกันเฉพาะคำพูดในประโยคเสียงพูดนำเข้าเท่านั้น

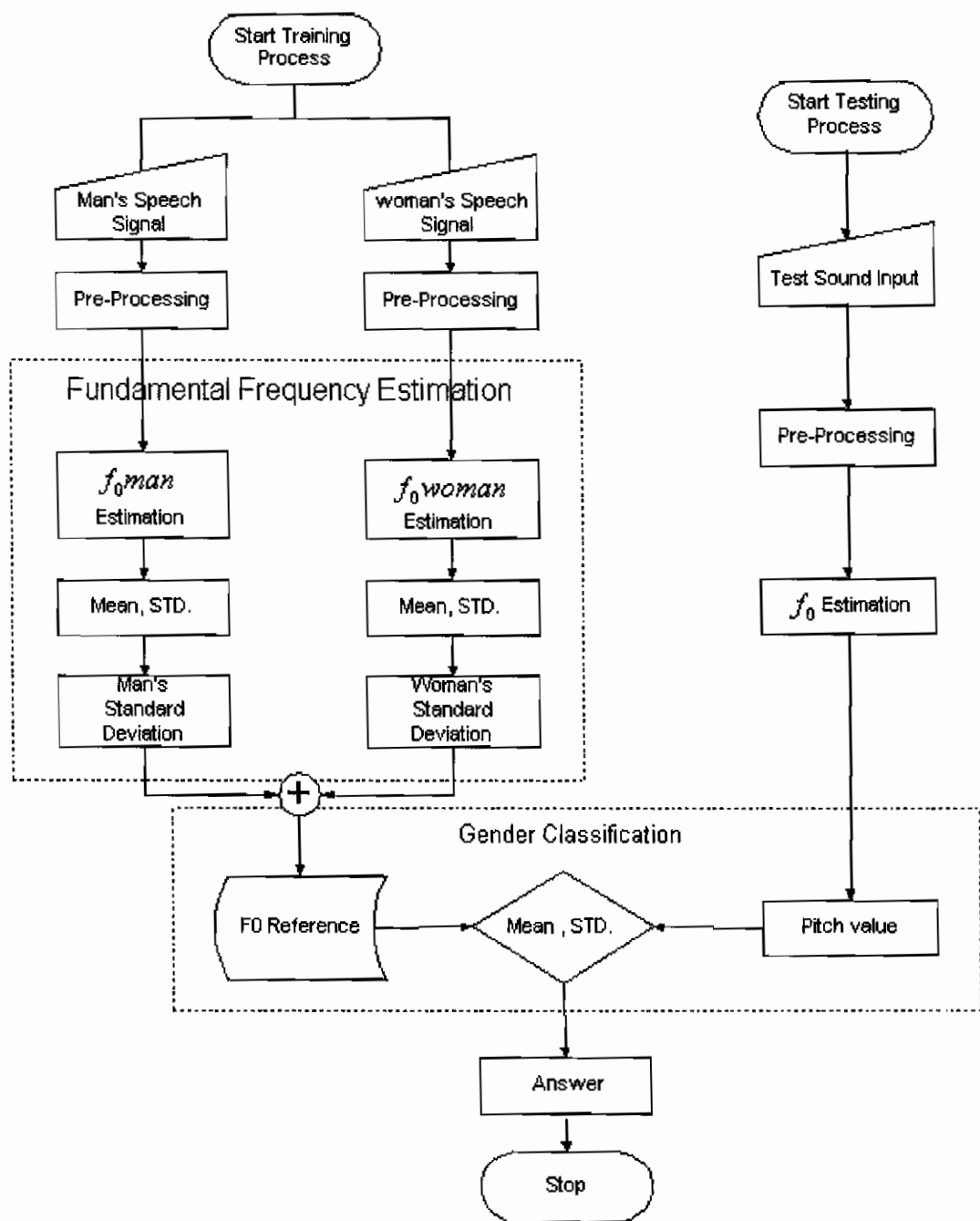
ขั้นตอนวิธีการเรียนรู้ดังแสดงในภาพที่ 3-14 เริ่มต้นด้วยการนำสัญญาณเสียงเรียนรู้เพศชายจำนวน 20 เสียงพูด และเสียงเรียนรู้เพศหญิงจำนวน 20 เสียงพูดเข้าสู่ขั้นตอนวิธีในกระบวนการเรียนรู้ทำการประมวลผลสัญญาณเบื้องต้นและประมาณค่าความถี่มูลฐานตามขั้นตอนวิธีที่เลือกใช้ในหัวข้อ 3.2 จะได้ค่าเฉลี่ย (Mean) และค่าเบี่ยงเบนมาตรฐาน (Standard Deviation) จากกลุ่มเรียนรู้เสียงพูดเพศชาย และกลุ่มเรียนรู้เสียงพูดเพศหญิงเพื่อนำไปใช้เป็นค่าอ้างอิงสำหรับเปรียบเทียบกับเสียงพูดทดสอบในกระบวนการจำแนกเพศบุคคลต่อไป

3.3.2 กระบวนการทดสอบ (Testing Process)

กระบวนการทดสอบจะใช้ขั้นตอนวิธีการประมวลผลสัญญาณเบื้องต้นและการประมาณค่าความถี่มูลฐานเช่นเดียวกันกับกระบวนการเรียนรู้แต่จะดำเนินการกับเสียงทดสอบครั้งละ 1 เสียงทดสอบเพื่อประมาณค่าความถี่มูลฐาน เมื่อได้ค่าพิตช์จากการประมาณค่าความถี่มูลฐานแล้วจะนำค่าพิตช์ที่ได้เข้าสู่ขั้นตอนการจำแนกเพศต่อไป

3.3.3 กระบวนการจำแนกเพศบุคคลด้วยค่าพิตช์จากการทดสอบ

เมื่อค่าพิตช์จากการทดสอบเข้าสู่การจำแนกเพศบุคคล ระบบจะนำค่าพิตช์ของเสียงพูดที่ทำการทดสอบนั้นไปรวมกลุ่มกับกลุ่มเรียนรู้เสียงพูดเพศชายเพื่อคำนวณหาค่าเฉลี่ย (Mean) และค่าเบี่ยงเบนมาตรฐาน (Standard Deviation) และนำค่าพิตช์ของเสียงพูดที่ทำการทดสอบนั้นไปรวมกลุ่มกับกลุ่มเรียนรู้เสียงพูดเพศหญิงเพื่อคำนวณหาค่าเฉลี่ย และค่าเบี่ยงเบนมาตรฐานด้วยเช่นกัน จากนั้นระบบจะพิจารณาความเปลี่ยนแปลงของค่าเฉลี่ย และค่าเบี่ยงเบนมาตรฐาน ที่เปลี่ยนแปลงไปเมื่อเปรียบเทียบกับค่าเฉลี่ย และค่าเบี่ยงเบนมาตรฐานอ้างอิงที่ได้จากขั้นตอนที่ 3.3.1 ว่ามีความน่าจะเป็นที่จะเป็นเสียงพูดเพศใด โดยใช้เกณฑ์ในการระบุเพศ คือ ค่าเฉลี่ย (Mean) และค่าเบี่ยงเบนมาตรฐาน (Standard Deviation) ของกลุ่มใดเปลี่ยนแปลงไปน้อยที่สุด ระบบก็จะแสดงผลลัพธ์ว่าเสียงพูดทดสอบคือเสียงพูดเพศนั้น



ภาพที่ 3-14 ขั้นตอนวิธีการจำแนกเพศบุคคลโดยใช้ค่าพิตช์ของเสียงพูดภาษาไทย

จากวิธีการดำเนินการวิจัยข้างต้น ผู้วิจัยได้ทำการทดลองกระบวนการจำแนกเพศบุคคลด้วยค่าพิตช์ของเสียงพูดภาษาไทยโดยแบ่งสัญญาณนำเข้าออกเป็นสัญญาณเสียงพูดในประโยคเดียวกัน และต่างประโยคกันเข้าสู่ขั้นตอนวิธีการประมาณค่าความถี่มูลฐานที่พัฒนาขึ้นทั้ง 8 ขั้นตอนวิธี แล้วนำผลลัพธ์ที่ได้เข้าสู่กระบวนการจำแนกเพศบุคคลเพื่อนำผลลัพธ์จากการจำแนกเพศบุคคลมาใช้เปรียบเทียบกับประสิทธิภาพของแต่ละขั้นตอนวิธี ดังผลการทดลองในบทที่ 4

บทที่ 4

ผลการวิจัย

4.1 ผลการเก็บตัวอย่างสัญญาณเสียงพูด

สัญญาณเสียงพูดที่ได้จากกลุ่มตัวอย่างแบ่งออกเป็น 2 กรณี

4.1.1 สัญญาณเสียงกรณีใช้ประโยคคำพูดเหมือนกัน

เก็บตัวอย่างเสียงพูดประโยค “กาลครั้งหนึ่งนานมาแล้ว” จำนวน 200 ตัวอย่างเสียงแบ่งออกเป็นเสียงพูดเพศชายจำนวน 90 สัญญาณเสียง และเสียงพูดเพศหญิงจำนวน 110 สัญญาณเสียง

ตารางที่ 4-1 จำนวนตัวอย่างเสียงพูดประโยคคำว่า “กาลครั้งหนึ่งนานมาแล้ว”

สัญญาณเสียงพูด	กลุ่มเรียนรู้	กลุ่มทดสอบ	รวม
เพศชาย	20	70	90
เพศหญิง	20	90	110
รวม	40	160	200

4.1.2 สัญญาณเสียงกรณีใช้ประโยคคำพูดต่างกัน

เก็บตัวอย่างเสียงพูดประโยคแนะนำตัว “ชื่อ นามสกุล เพศ อายุ วันเกิด” จำนวน 208 ตัวอย่างเสียงแบ่งออกเป็นเสียงพูดเพศชายจำนวน 81 สัญญาณเสียง และเสียงพูดเพศหญิงจำนวน 127 สัญญาณเสียง

ตารางที่ 4-2 จำนวนตัวอย่างเสียงพูดประโยคนำตัว

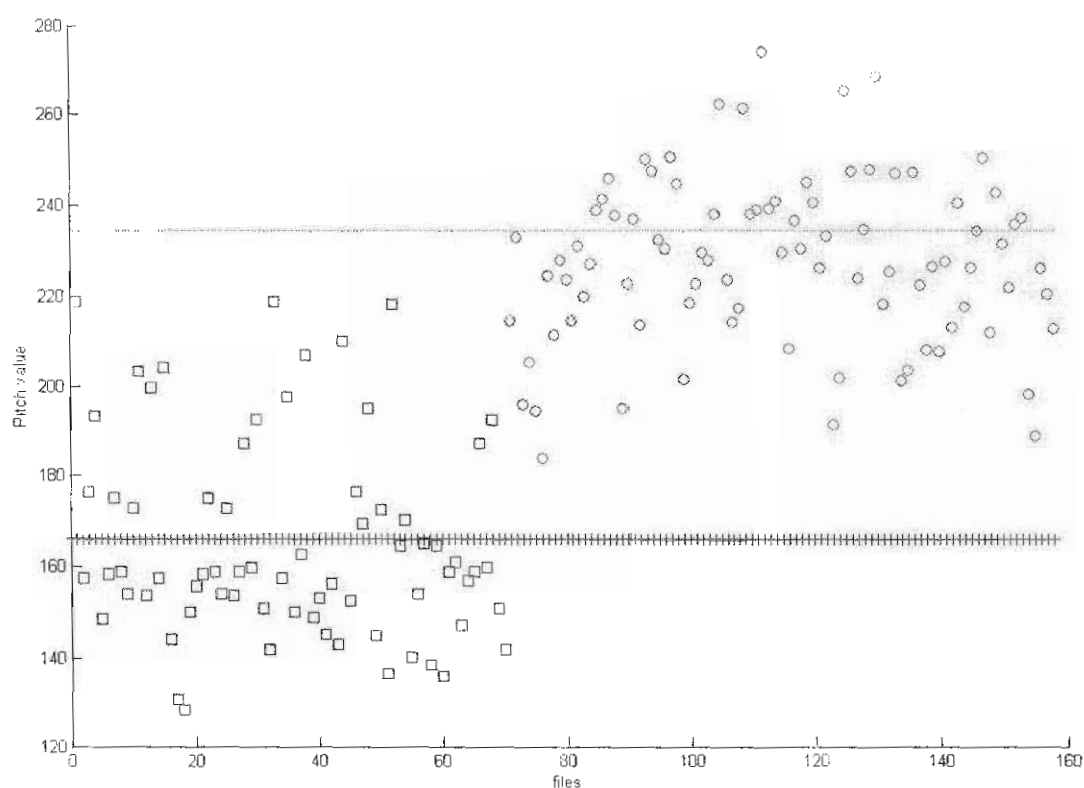
สัญญาณเสียงพูด	กลุ่มเรียนรู้	กลุ่มทดสอบ	รวม
เพศชาย	20	61	81
เพศหญิง	20	107	127
รวม	40	168	208

4.2 ผลการทดลองการจำแนกเพศบุคคลกรณีใช้ประโยคคำพูดเหมือนกัน

การทดลองการจำแนกเพศบุคคลด้วยคำพิตช์จากเสียงพูดภาษาไทยกรณีใช้ประโยคคำพูดเหมือนกันแบ่งการทดลองตามการเพิ่มประสิทธิภาพกระบวนการประมวลผลสัญญาณเบื้องต้นออกเป็น 8 กรณีดังในหัวข้อที่ 3.2 ในบทที่ 3 ซึ่งได้ผลการทดลองดังนี้

4.2.1 ผลการทดลองขั้นตอนวิธีการหาค่าพิตช์ของเสียงพูดด้วยวิธี Autocorrelation โดยมีการประมวลผลสัญญาณเบื้องต้นด้วยการวางฟังก์ชันกรอบสัญญาณแบบสี่เหลี่ยม

ค่าเฉลี่ย (Mean) และค่าเบี่ยงเบนมาตรฐาน (Standard Deviation) อ้างอิงของกลุ่มเรียนรู้เพศชาย คือ 153.85 และ 17.64 ตามลำดับ ค่าเฉลี่ย และค่าเบี่ยงเบนมาตรฐานของกลุ่มเรียนรู้เพศหญิง คือ 226.44 และ 16.65 ตามลำดับ จากนั้นจึงเข้าสู่กระบวนการทดสอบ จากการทดสอบพบว่าระบบมีร้อยละความถูกต้องจากการจำแนกเสียงพูดเพศชายอยู่ที่ 91.43 และร้อยละความถูกต้องของเสียงพูดเพศหญิงคือ 94.44 ดังภาพที่ 4-1

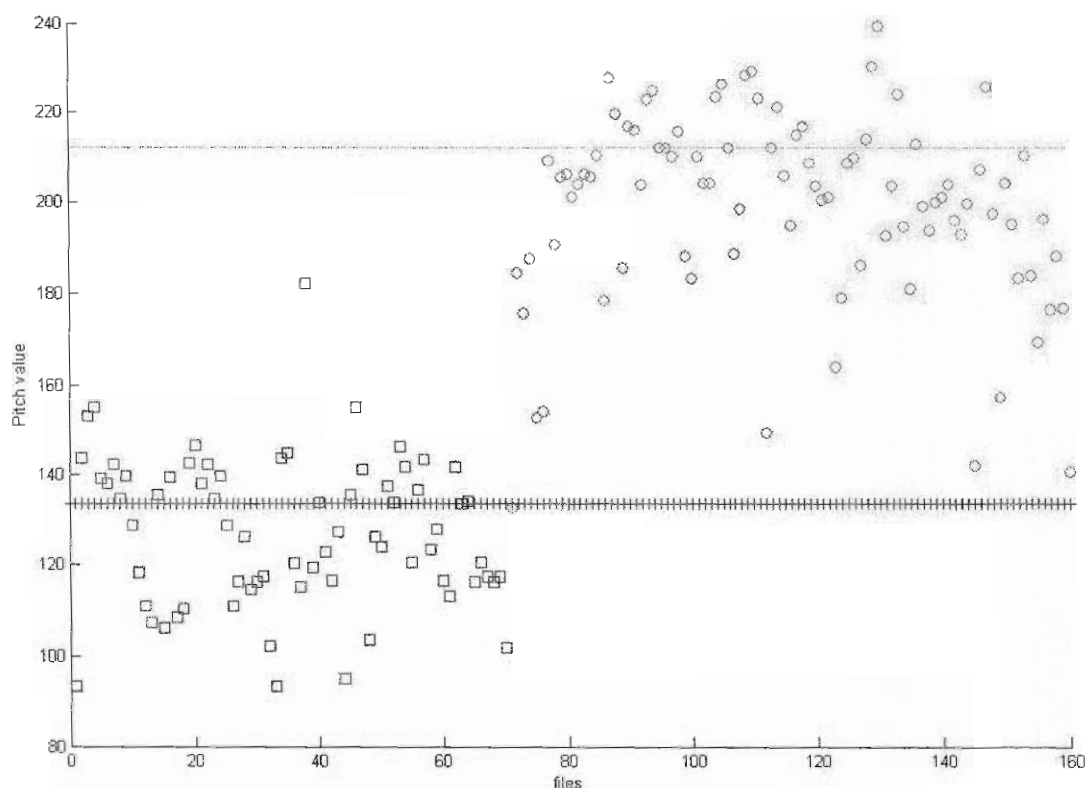


- เมื่อ +++ แทน ค่าพิตช์อ้างอิงเสียงพูดเพศชาย
 --- แทน ค่าพิตช์อ้างอิงเสียงพูดเพศหญิง
 □ แทน ค่าพิตช์ของสัญญาณเสียงพูดเพศชาย
 ○ แทน ค่าพิตช์ของสัญญาณเสียงพูดเพศหญิง

ภาพที่ 4-1 ค่าพิตช์ของเสียงพูดทดสอบที่คำนวณจากวิธี Autocorrelation

4.2.2 ผลการทดลองขั้นตอนวิธีการหาค่าพิตช์ของเสียงพูดด้วยวิธี Cepstrum โดยมีการประมวลผลสัญญาณเบื้องต้นด้วยการวางฟังก์ชันกรอบสัญญาณแบบแฮมมิง

ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานอ้างอิงของกลุ่มเรียนรู้เพศชาย คือ 133.60 และ 22.10 ตามลำดับ ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของกลุ่มเรียนรู้เพศหญิง คือ 212.17 และ 13.50 ตามลำดับ จากนั้นจึงเข้าสู่กระบวนการทดสอบ จากการทดสอบพบว่าระบบมีร้อยละความถูกต้องจากการจำแนกเสียงพูดเพศชายอยู่ที่ 98.57 และร้อยละความถูกต้องจากการจำแนกเสียงพูดเพศหญิงคือ 87.78 ดังภาพที่ 4-2

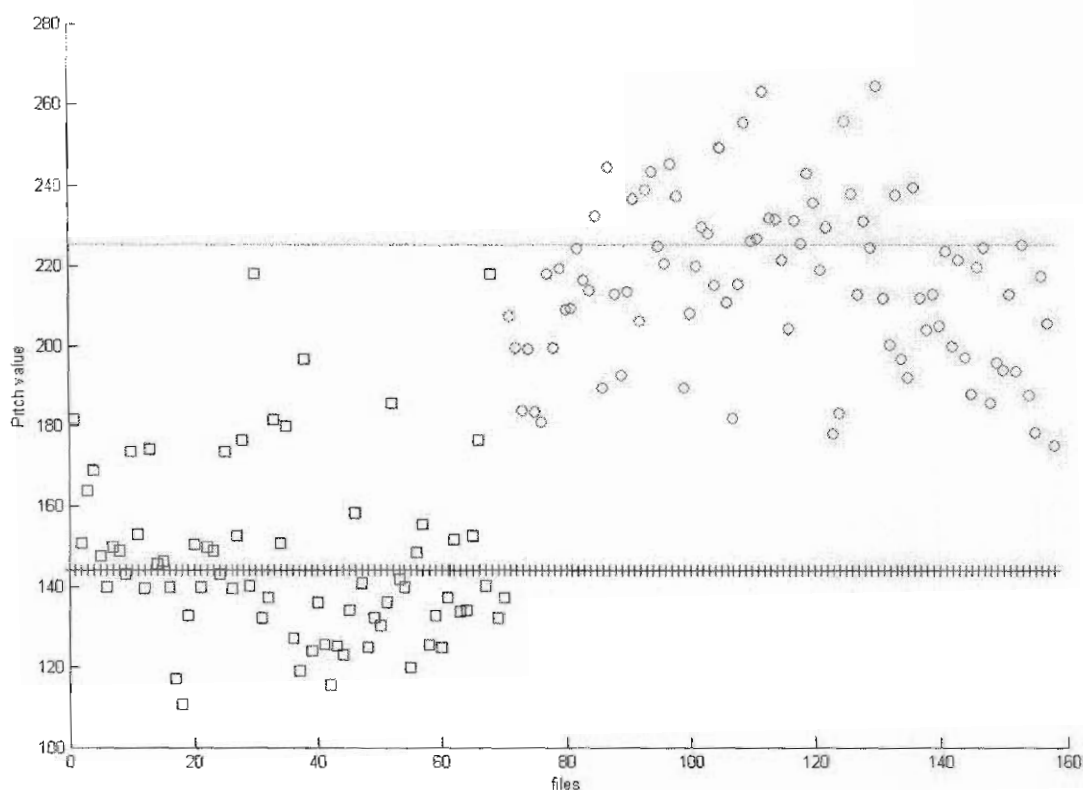


- เมื่อ +++ แทน ค่าพิตช์อ้างอิงเสียงพูดเพศชาย
 --- แทน ค่าพิตช์อ้างอิงเสียงพูดเพศหญิง
 □ แทน ค่าพิตช์ของสัญญาณเสียงพูดเพศชาย
 ○ แทน ค่าพิตช์ของสัญญาณเสียงพูดเพศหญิง

ภาพที่ 4-2 ค่าพิตช์ของเสียงพูดทดสอบที่คำนวณจากวิธีวิเคราะห์ Cepstrum

4.2.3 ผลการทดลองขั้นตอนวิธีการหาค่าพิตช์ของเสียงพูดด้วยวิธี Autocorrelation โดยมีการประมวลผลสัญญาณเบื้องต้นด้วยการวางฟังก์ชันกรอบสัญญาณและการ Clipping สัญญาณเสียง

ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานอ้างอิงของกลุ่มเรียนรู้เพศชาย คือ 144.00 และ 13.97 ตามลำดับ ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของกลุ่มเรียนรู้เพศหญิง คือ 225.56 และ 20.27 ตามลำดับ จากนั้นจึงเข้าสู่กระบวนการทดสอบ จากการทดสอบพบว่าระบบมีร้อยละความถูกต้องจากการจำแนกเสียงพูดเพศชายอยู่ที่ 92.86 และร้อยละความถูกต้องจากการจำแนกเสียงพูดเพศหญิงคือ 96.67 ดังภาพที่ 4-3

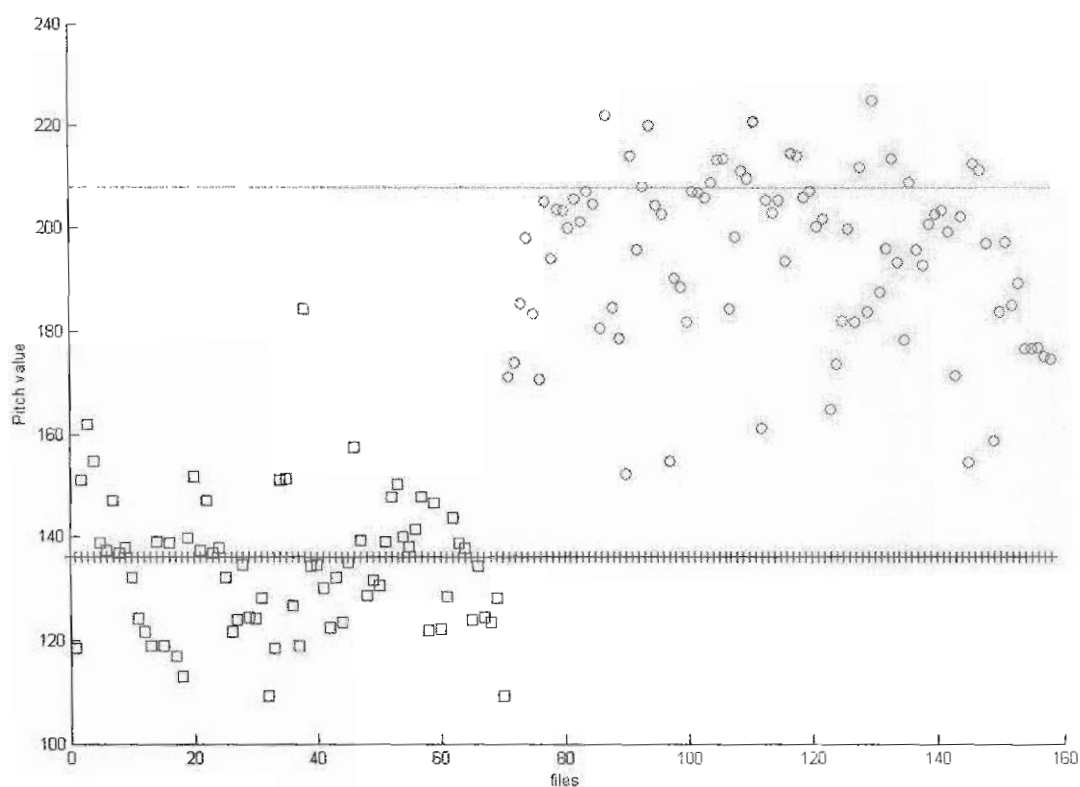


- เมื่อ +++ แทน ค่าพิตช์อ้างอิงเสียงพูดเพศชาย
 --- แทน ค่าพิตช์อ้างอิงเสียงพูดเพศหญิง
 □ แทน ค่าพิตช์ของสัญญาณเสียงพูดเพศชาย
 ○ แทน ค่าพิตช์ของสัญญาณเสียงพูดเพศหญิง

ภาพที่ 4-3 ค่าพิตช์ของเสียงพูดทดสอบที่คำนวณจากวิธี Autocorrelation โดยเพิ่มกระบวนการ Clipping สัญญาณเสียงพูดเข้าไปในกระบวนการประมวลผลสัญญาณเบื้องต้น

4.2.4 ผลการทดลองขั้นตอนวิธีการหาค่าพิตช์ของเสียงพูดด้วยวิธี Cepstrum โดยมีการประมวลผลสัญญาณเบื้องต้นด้วยการวางฟังก์ชันกรอบสัญญาณและการ Clipping สัญญาณเสียง

ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานอ้างอิงของกลุ่มเรียนรู้เพศชาย คือ 135.96 และ 14.05 ตามลำดับ ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของกลุ่มเรียนรู้เพศหญิง คือ 207.92 และ 11.25 ตามลำดับ จากนั้นจึงเข้าสู่กระบวนการทดสอบ จากการทดสอบพบว่าระบบมีร้อยละความถูกต้องจากการจำแนกเสียงพูดเพศชายอยู่ที่ 98.57 และร้อยละความถูกต้องจากการจำแนกเสียงพูดเพศหญิงคือ 90.00 ดังภาพที่ 4-4

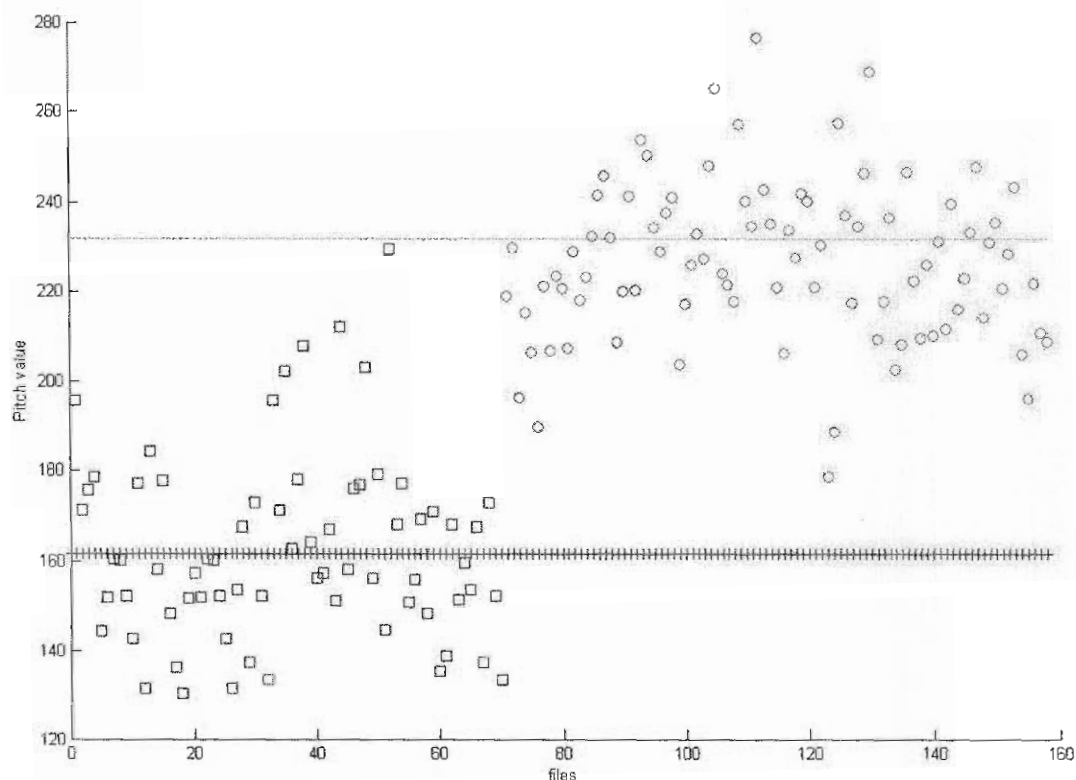


- เมื่อ +++ แทน ค่าพิตช์อ้างอิงเสียงพูดเพศชาย
 --- แทน ค่าพิตช์อ้างอิงเสียงพูดเพศหญิง
 □ แทน ค่าพิตช์ของสัญญาณเสียงพูดเพศชาย
 ○ แทน ค่าพิตช์ของสัญญาณเสียงพูดเพศหญิง

ภาพที่ 4-4 ค่าพิตช์ของเสียงพูดทดสอบที่คำนวณจากวิธีวิเคราะห์ Cepstrum โดยเพิ่มกระบวนการ Clipping สัญญาณเสียงพูดเข้าไปในกระบวนการประมวลผลสัญญาณเบื้องต้น

4.2.5 ผลการทดลองขั้นตอนวิธีการหาค่าพิตช์ของเสียงพูดด้วยวิธี Autocorrelation โดยมีการประมวลผลสัญญาณเบื้องต้นด้วยการลบ DC Bias และฟังก์ชันการวางกรอบสัญญาณ

ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานอ้างอิงของกลุ่มเรียนรู้เพศชาย คือ 150.77 และ 15.91 ตามลำดับ ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของกลุ่มเรียนรู้เพศหญิง คือ 223.36 และ 18.50 ตามลำดับ จากนั้นจึงเข้าสู่กระบวนการทดสอบ จากการทดสอบพบว่าระบบมีร้อยละความถูกต้องจากการจำแนกเสียงพูดเพศชายอยู่ที่ 94.29 และร้อยละความถูกต้องจากการจำแนกเสียงพูดเพศหญิงคือ 94.45 ดังภาพที่ 4-5

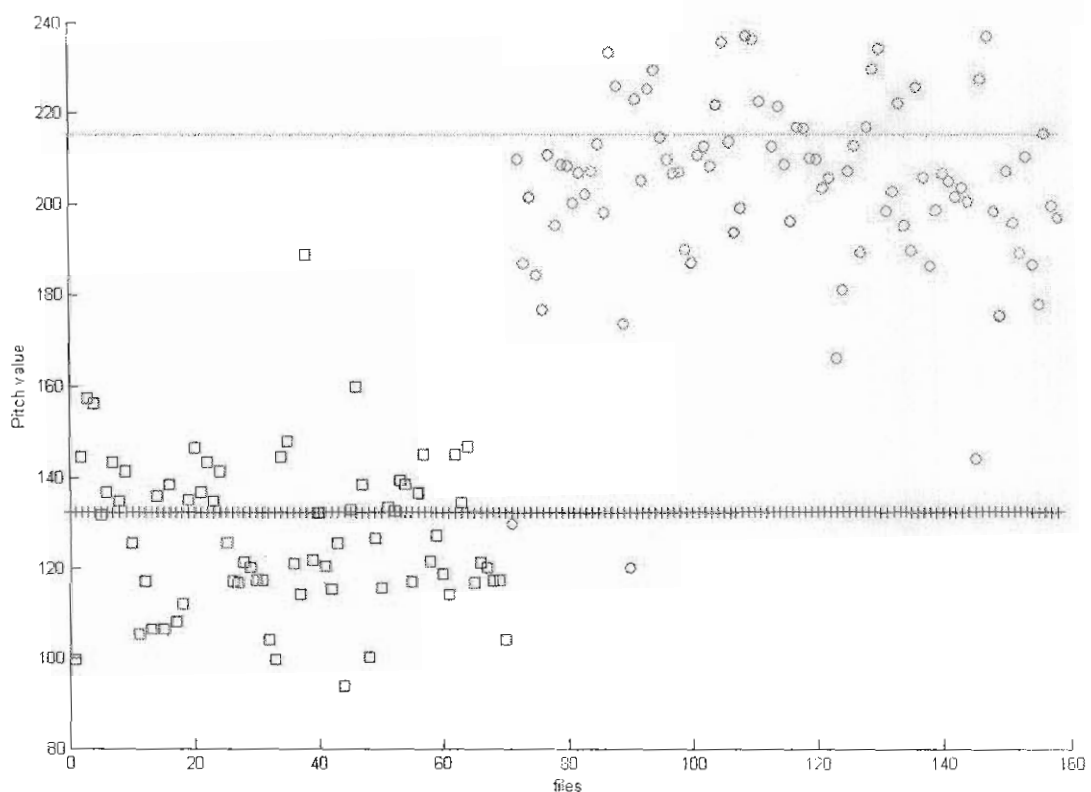


- เมื่อ +++ แทน ค่าพิตช์อ้างอิงเสียงพูดเพศชาย
 --- แทน ค่าพิตช์อ้างอิงเสียงพูดเพศหญิง
 □ แทน ค่าพิตช์ของสัญญาณเสียงพูดเพศชาย
 ○ แทน ค่าพิตช์ของสัญญาณเสียงพูดเพศหญิง

ภาพที่ 4-5 ค่าพิตช์ของเสียงพูดทดสอบที่คำนวณจากวิธี Autocorrelation โดยเพิ่มการลบ DC Bias และ Butterworth Filter เข้าไปในกระบวนการประมวลผลสัญญาณเบื้องต้น

4.2.6 ผลการทดลองขั้นตอนวิธีการหาค่าพิตช์ของเสียงพูดด้วยวิธี Cepstrum โดยมีการประมวลผลสัญญาณเบื้องต้นด้วยการลบ DC Bias และการวางกรอบสัญญาณ

ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานอ้างอิงของกลุ่มเรียนรู้เพศชาย คือ 132.49 และ 19.18 ตามลำดับ ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของกลุ่มเรียนรู้เพศหญิง คือ 215.20 และ 13.48 ตามลำดับ จากนั้นจึงเข้าสู่กระบวนการทดสอบ จากการทดสอบพบว่าระบบมีร้อยละความถูกต้องจากการจำแนกเสียงพูดเพศชายอยู่ที่ 98.57 และร้อยละความถูกต้องจากการจำแนกเสียงพูดเพศหญิงคือ 91.11 ดังภาพที่ 4-6

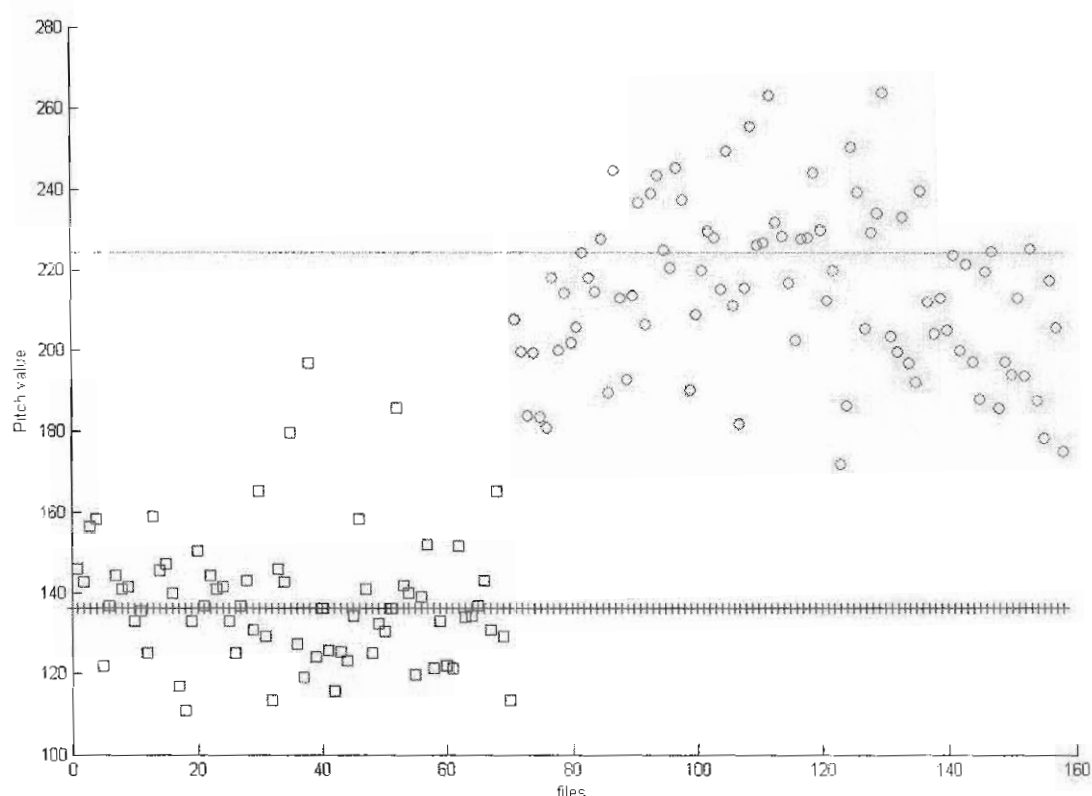


- เมื่อ +++ แทน ค่าพิตช์อ้างอิงเสียงพูดเพศชาย
 --- แทน ค่าพิตช์อ้างอิงเสียงพูดเพศหญิง
 □ แทน ค่าพิตช์ของสัญญาณเสียงพูดเพศชาย
 ○ แทน ค่าพิตช์ของสัญญาณเสียงพูดเพศหญิง

ภาพที่ 4-6 ค่าพิตช์ของเสียงพูดทดสอบที่คำนวณจากวิเคราะห์ Cepstrum โดยเพิ่มการลบ DC Bias และ Butterworth Filter เข้าไปในกระบวนการประมวลผลสัญญาณเบื้องต้น

4.2.7 ผลการทดลองขั้นตอนวิธีการหาค่าพิตช์ของเสียงพูดด้วยวิธี Autocorrelation โดยมีการประมวลผลสัญญาณเบื้องต้นด้วยการลบ DC Bias การวางฟังก์ชันกรอบสัญญาณ และการ Clipping สัญญาณเสียง

ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานอ้างอิงของกลุ่มเรียนรู้เพศชาย คือ 136.07 และ 16.43 ตามลำดับ ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของกลุ่มเรียนรู้เพศหญิง คือ 224.65 และ 20.77 ตามลำดับ จากนั้นจึงเข้าสู่กระบวนการทดสอบ จากการทดสอบพบว่าระบบมีร้อยละความถูกต้องจากการจำแนกเสียงพูดเพศชายอยู่ที่ 95.71 และร้อยละความถูกต้องจากการจำแนกเสียงพูดเพศหญิงคือ 97.78 ดังภาพที่ 4-7

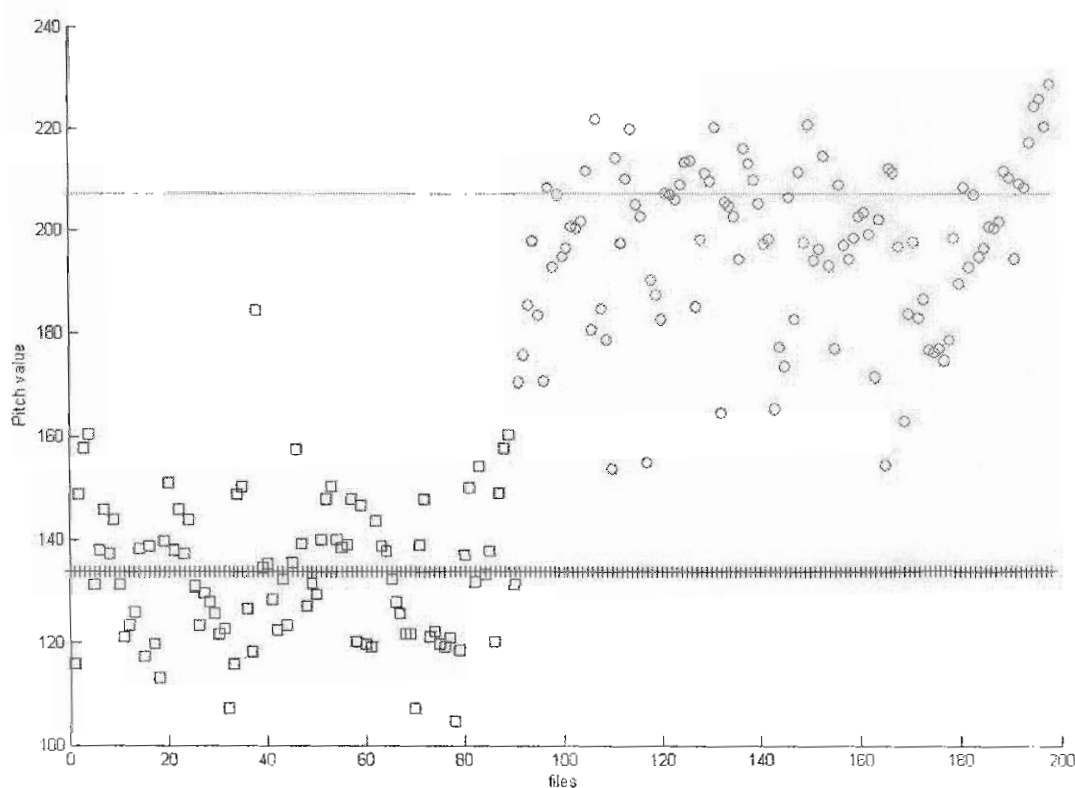


- เมื่อ +++ แทน ค่าพิตช์อ้างอิงเสียงพูดเพศชาย
 --- แทน ค่าพิตช์อ้างอิงเสียงพูดเพศหญิง
 □ แทน ค่าพิตช์ของสัญญาณเสียงพูดเพศชาย
 ○ แทน ค่าพิตช์ของสัญญาณเสียงพูดเพศหญิง

ภาพที่ 4-7 ค่าพิตช์ของเสียงพูดทดสอบที่คำนวณจากวิธี Autocorrelation ที่เพิ่มการลบ DC Bias, Butterworth Filter และการ Clipping สัญญาณเสียงพูดเข้าไปในกระบวนการประมวลผลสัญญาณเบื้องต้น

4.2.8 ผลการทดลองขั้นตอนวิธีการหาค่าพิตช์ของเสียงพูดด้วยวิธี Cepstrum โดยมีการประมวลผลสัญญาณเบื้องต้นด้วยการลบ DC Bias จากนั้นทำการวางฟังก์ชันกรอบสัญญาณ และการ Clipping สัญญาณเสียง

ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานอ้างอิงของกลุ่มเรียนรู้เพศชาย คือ 133.77 และ 15.59 ตามลำดับ ค่าเฉลี่ย และค่าเบี่ยงเบนมาตรฐานของกลุ่มเรียนรู้เพศหญิง คือ 207.13 และ 11.65 ตามลำดับ จากนั้นจึงเข้าสู่กระบวนการทดสอบ จากการทดสอบพบว่าระบบมีร้อยละความถูกต้องจากการจำแนกเสียงพูดเพศชายอยู่ที่ 98.57 และร้อยละความถูกต้องจากการจำแนกเสียงพูดเพศหญิงคือ 90.00 ดังภาพที่ 4-8



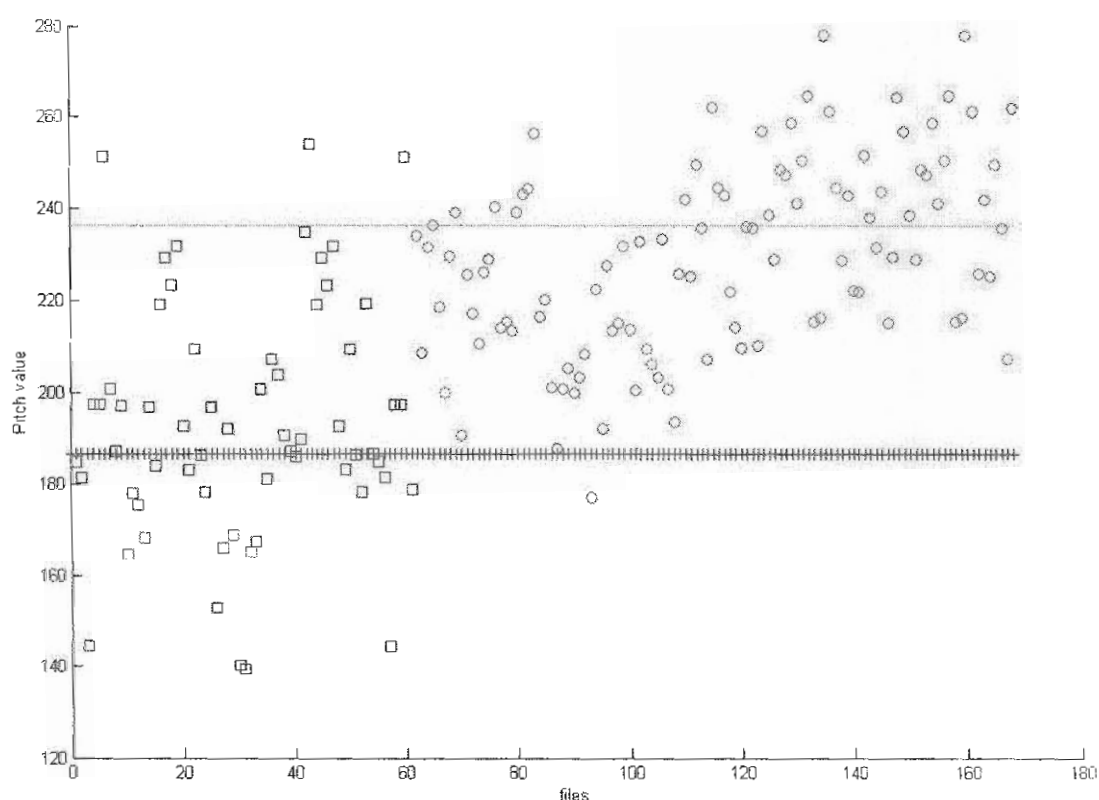
- เมื่อ +++ แทน ค่าพิตช์อ้างอิงเสียงพูดเพศชาย
 --- แทน ค่าพิตช์อ้างอิงเสียงพูดเพศหญิง
 □ แทน ค่าพิตช์ของสัญญาณเสียงพูดเพศชาย
 ○ แทน ค่าพิตช์ของสัญญาณเสียงพูดเพศหญิง

ภาพที่ 4-8 ค่าพิตช์ของเสียงพูดทดสอบที่คำนวณจากวิธีวิเคราะห์ Cepstrum ที่เพิ่มการลบ DC Bias, Butterworth Filter และการ Clipping สัญญาณเสียงพูดเข้าไปในกระบวนการประมวลผลสัญญาณเบื้องต้น

4.3 ผลการทดลองการจำแนกเพศบุคคลกรณีใช้ประโยคคำพูดต่างกัน

4.3.1 ผลการทดลองขั้นตอนวิธีการหาค่าพิตช์ของเสียงพูดด้วยวิธี Autocorrelation โดยมีการประมวลผลสัญญาณเบื้องต้นด้วยการวางฟังก์ชันกรอบสัญญาณแบบสี่เหลี่ยม

ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานอ้างอิงของกลุ่มเรียนรู้เพศชาย คือ 186.65 และ 27.05 ตามลำดับ ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของกลุ่มเรียนรู้เพศหญิง คือ 236.20 และ 13.31 ตามลำดับ จากนั้นจึงเข้าสู่กระบวนการทดสอบ จากการทดสอบพบว่าระบบมีร้อยละความถูกต้องจากการจำแนกเสียงพูดเพศชายอยู่ที่ 78.69 และร้อยละความถูกต้องจากการจำแนกเสียงพูดเพศหญิงคือ 64.49 ดังภาพที่ 4-9

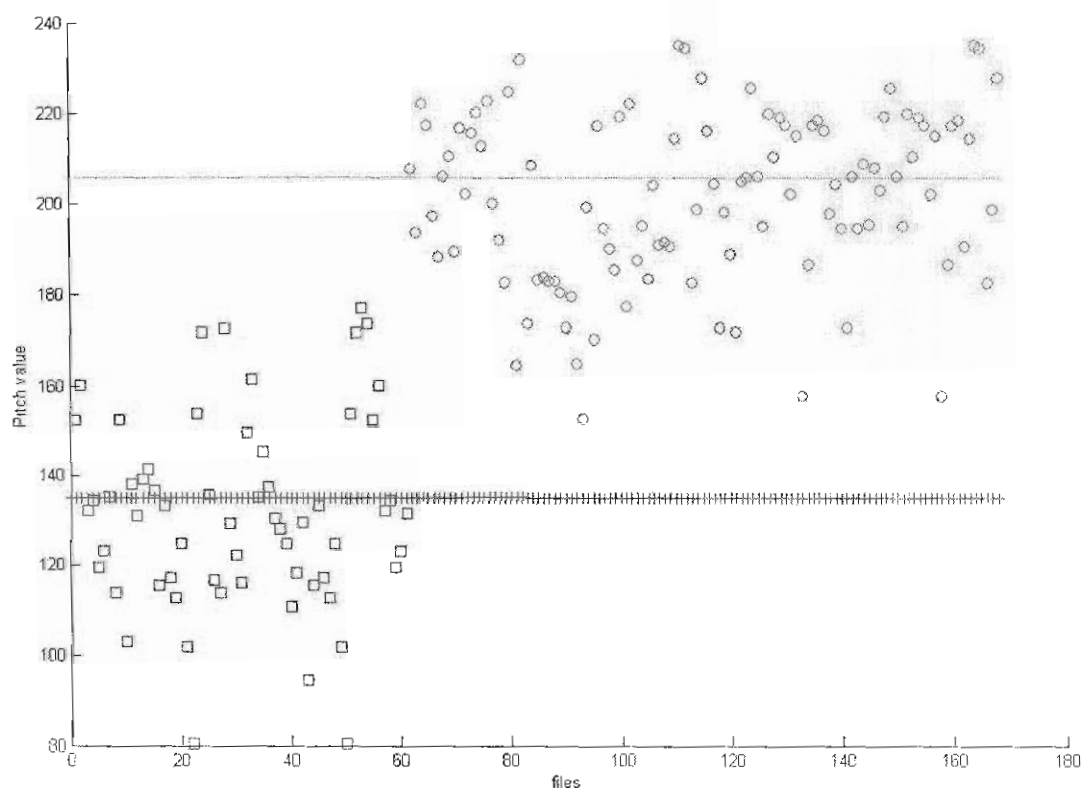


- เมื่อ +++ แทน ค่าพิตช์อ้างอิงเสียงพูดเพศชาย
 --- แทน ค่าพิตช์อ้างอิงเสียงพูดเพศหญิง
 □ แทน ค่าพิตช์ของสัญญาณเสียงพูดเพศชาย
 ○ แทน ค่าพิตช์ของสัญญาณเสียงพูดเพศหญิง

ภาพที่ 4-9 ค่าพิตช์ของเสียงพูดทดสอบที่คำนวณจากวิธี Autocorrelation

4.3.2 ผลการทดลองขั้นตอนวิธีการหาค่าพิตช์ของเสียงพูดด้วยวิธี Cepstrum โดยมีการประมวลผลสัญญาณเบื้องต้นด้วยการวางฟังก์ชันกรอบสัญญาณแบบแฮมมิง

ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานอ้างอิงของกลุ่มเรียนรู้เพศชาย คือ 135.08 และ 20.21 ตามลำดับ ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของกลุ่มเรียนรู้เพศหญิง คือ 205.80 และ 21.58 ตามลำดับ จากนั้นจึงเข้าสู่กระบวนการทดสอบ จากการทดสอบพบว่าระบบมีร้อยละความถูกต้องจากการจำแนกเสียงพูดเพศชายอยู่ที่ 91.80 และร้อยละความถูกต้องจากการจำแนกเสียงพูดเพศหญิงคือ 95.33 ดังภาพที่ 4-10

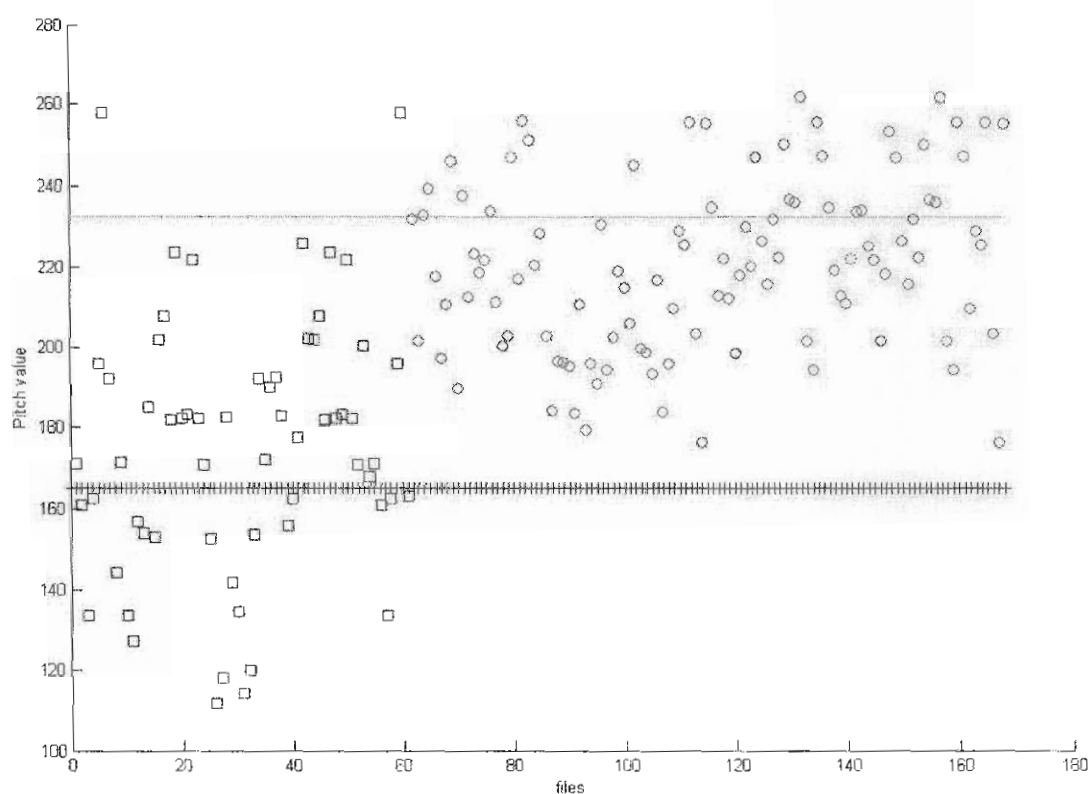


- เมื่อ +++ แทน ค่าพิตช์อ้างอิงเสียงพูดเพศชาย
 --- แทน ค่าพิตช์อ้างอิงเสียงพูดเพศหญิง
 □ แทน ค่าพิตช์ของสัญญาณเสียงพูดเพศชาย
 ○ แทน ค่าพิตช์ของสัญญาณเสียงพูดเพศหญิง

ภาพที่ 4-10 ค่าพิตช์ของเสียงพูดทดสอบที่คำนวณจากวิธีวิเคราะห์ Cepstrum

4.3.3 ผลการทดลองขั้นตอนวิธีการหาค่าพิตช์ของเสียงพูดด้วยวิธี Autocorrelation โดยมีการประมวลผลสัญญาณเบื้องต้นด้วยการวางฟังก์ชันกรอบสัญญาณและการ Clipping สัญญาณเสียง

ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานอ้างอิงของกลุ่มเรียนรู้เพศชาย คือ 165.09 และ 28.12 ตามลำดับ ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของกลุ่มเรียนรู้เพศหญิง คือ 232.75 และ 16.94 ตามลำดับ จากนั้นจึงเข้าสู่กระบวนการทดสอบ จากการทดสอบพบว่าระบบมีร้อยละความถูกต้องจากการจำแนกเสียงพูดเพศชายอยู่ที่ 86.89 และร้อยละความถูกต้องจากการจำแนกเสียงพูดเพศหญิงคือ 71.96 ดังภาพที่ 4-11

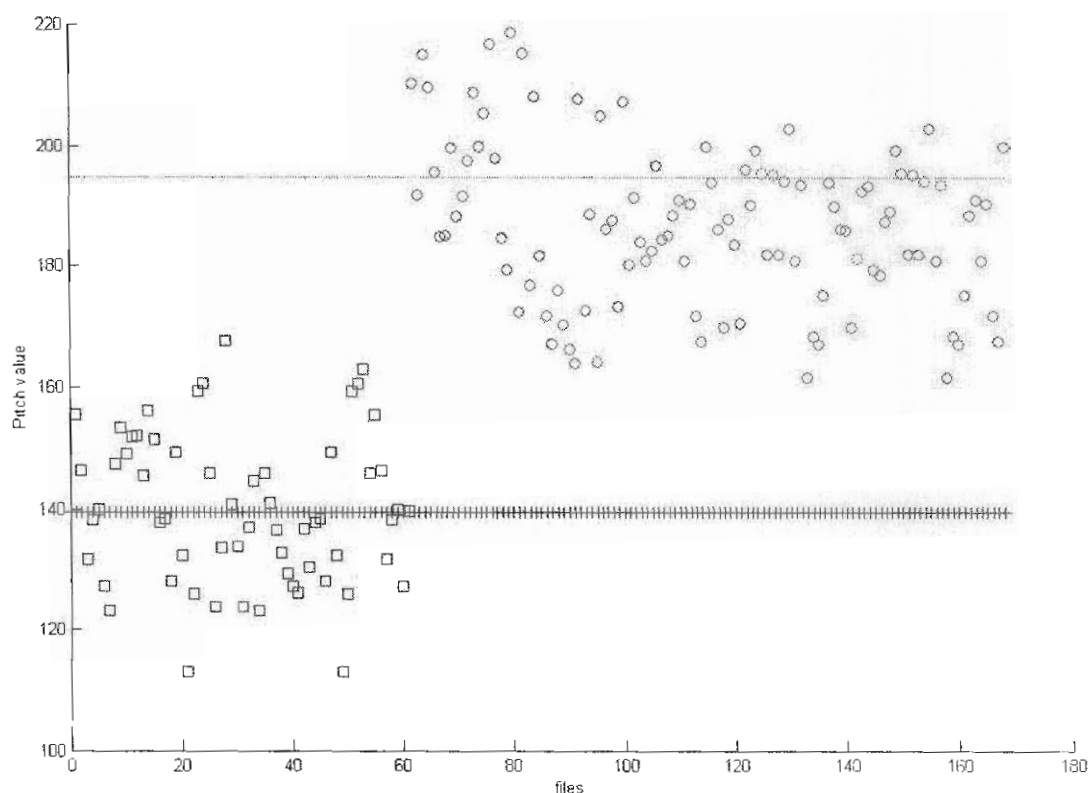


- เมื่อ +++ แทน ค่าพิตช์อ้างอิงเสียงพูดเพศชาย
 --- แทน ค่าพิตช์อ้างอิงเสียงพูดเพศหญิง
 □ แทน ค่าพิตช์ของสัญญาณเสียงพูดเพศชาย
 ○ แทน ค่าพิตช์ของสัญญาณเสียงพูดเพศหญิง

ภาพที่ 4-11 ค่าพิตช์ของเสียงพูดทดสอบที่คำนวณจากวิธี Autocorrelation โดยเพิ่มการ Clipping สัญญาณเสียงพูดเข้าไปในกระบวนการประมวลผลสัญญาณเบื้องต้น

4.3.4 ผลการทดลองขั้นตอนวิธีการหาค่าพิตช์ของเสียงพูดด้วยวิธี Cepstrum โดยมีการประมวลผลสัญญาณเบื้องต้นด้วยการวางฟังก์ชันกรอบสัญญาณและการ Clipping สัญญาณเสียง

ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานอ้างอิงของกลุ่มเรียนรู้เพศชาย คือ 139.50 และ 11.45 ตามลำดับ ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของกลุ่มเรียนรู้เพศหญิง คือ 194.91 และ 11.72 ตามลำดับ จากนั้นจึงเข้าสู่กระบวนการทดสอบ: จากการทดสอบพบว่าระบบมีร้อยละความถูกต้องจากการจำแนกเสียงพูดเพศชายอยู่ที่ 98.36 และร้อยละความถูกต้องจากการจำแนกเสียงพูดเพศหญิงคือ 95.33 ดังภาพที่ 4-12

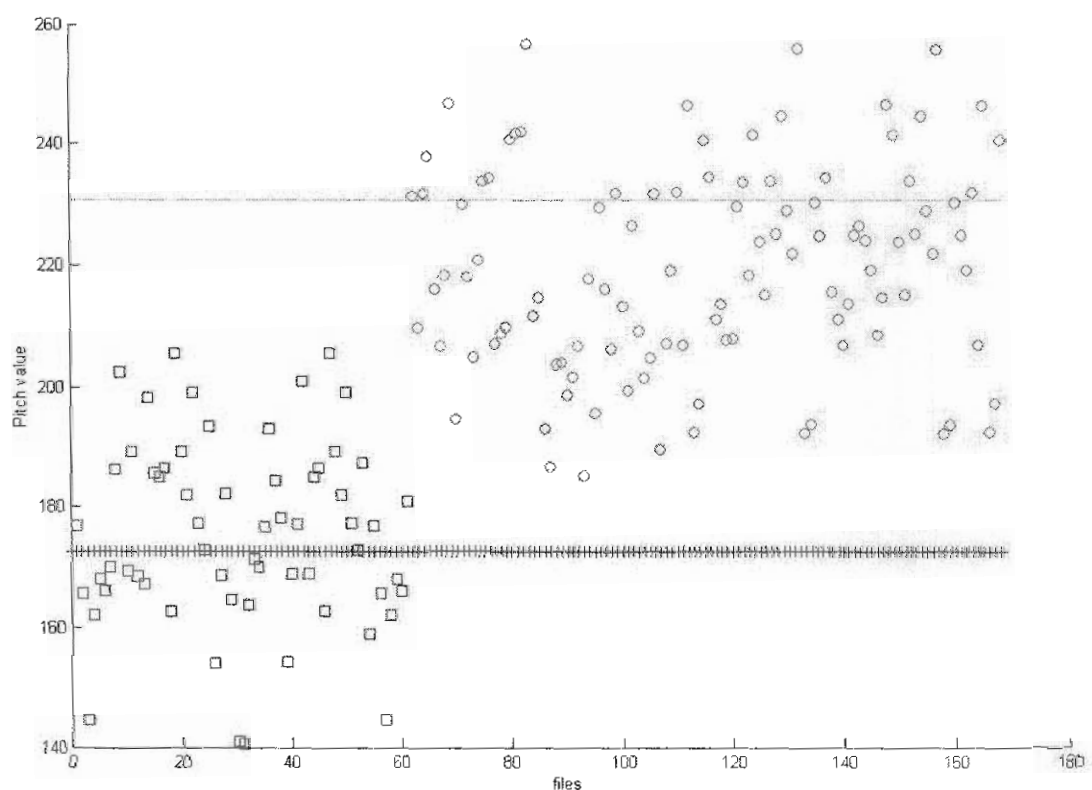


- เมื่อ +++ แทน ค่าพิตช์อ้างอิงเสียงพูดเพศชาย
 --- แทน ค่าพิตช์อ้างอิงเสียงพูดเพศหญิง
 □ แทน ค่าพิตช์ของสัญญาณเสียงพูดเพศชาย
 ○ แทน ค่าพิตช์ของสัญญาณเสียงพูดเพศหญิง

ภาพที่ 4-12 ค่าพิตช์ของเสียงพูดทดสอบที่คำนวณจากวิธีวิเคราะห์ Cepstrum โดยเพิ่มกระบวนการ Clipping สัญญาณเสียงพูดเข้าไปในกระบวนการประมวลผลสัญญาณเบื้องต้น

4.3.5 ผลการทดลองขั้นตอนวิธีการหาค่าพิตช์ของเสียงพูดด้วยวิธี Autocorrelation โดยมีการประมวลผลสัญญาณเบื้องต้นด้วยการลบ DC Bias และฟังก์ชันการวางกรอบสัญญาณ

ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานอ้างอิงของกลุ่มเรียนรู้เพศชาย คือ 172.50 และ 15.39 ตามลำดับ ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของกลุ่มเรียนรู้เพศหญิง คือ 230.58 และ 15.46 ตามลำดับ จากนั้นจึงเข้าสู่กระบวนการทดสอบ จากการทดสอบพบว่าระบบมีร้อยละความถูกต้องจากการจำแนกเสียงพูดเพศชายอยู่ที่ 92.72 และร้อยละความถูกต้องจากการจำแนกเสียงพูดเพศหญิงคือ 84.11 ดังภาพที่ 4-13

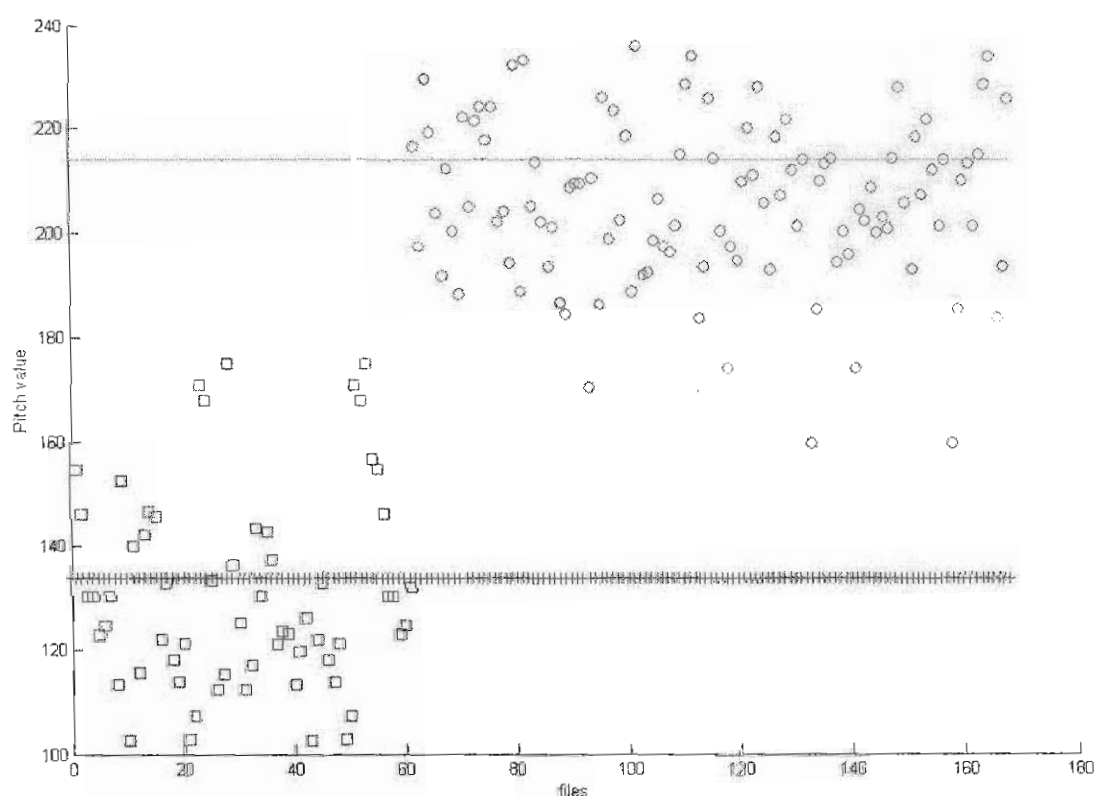


- เมื่อ +++ แทน ค่าพิตช์อ้างอิงเสียงพูดเพศชาย
 --- แทน ค่าพิตช์อ้างอิงเสียงพูดเพศหญิง
 □ แทน ค่าพิตช์ของสัญญาณเสียงพูดเพศชาย
 ○ แทน ค่าพิตช์ของสัญญาณเสียงพูดเพศหญิง

ภาพที่ 4-13 ค่าพิตช์ของเสียงพูดทดสอบที่คำนวณจากวิธี Autocorrelation โดยเพิ่มการลบ DC Bias และ Butterworth Filter เข้าไปในกระบวนการประมวลผลสัญญาณเบื้องต้น

4.3.6 ผลการทดลองขั้นตอนวิธีการหาค่าพิตช์ของเสียงพูดด้วยวิธี Cepstrum โดยมีการประมวลผลสัญญาณเบื้องต้นด้วยการลบ DC Bias และการวางกรอบสัญญาณ

ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานอ้างอิงของกลุ่มเรียนรู้เพศชาย คือ 133.68 และ 17.61 ตามลำดับ ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของกลุ่มเรียนรู้เพศหญิง คือ 214.04 และ 14.18 ตามลำดับ จากนั้นจึงเข้าสู่กระบวนการทดสอบ จากการทดสอบพบว่าระบบมีร้อยละความถูกต้องจากการจำแนกเสียงพูดเพศชายอยู่ที่ 100 และร้อยละความถูกต้องจากการจำแนกเสียงพูดเพศหญิงคือ 95.33 ดังภาพที่ 4-14

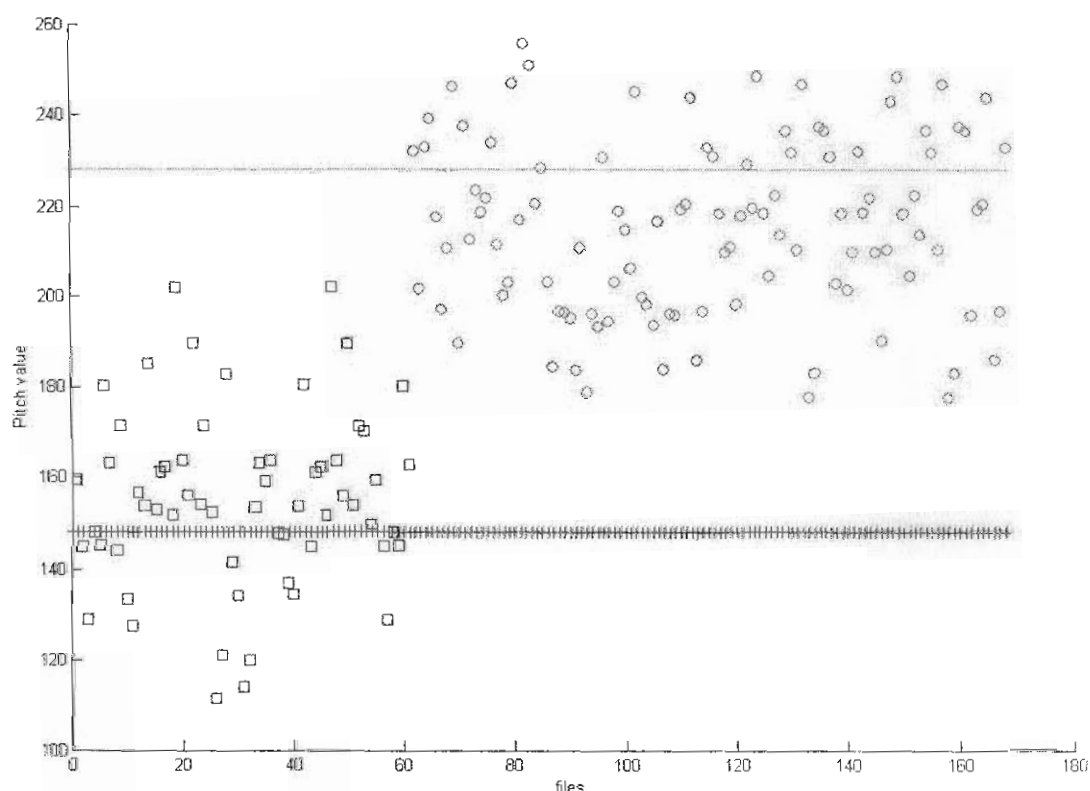


- เมื่อ ++ แทน ค่าพิตช์อ้างอิงเสียงพูดเพศชาย
 -- แทน ค่าพิตช์อ้างอิงเสียงพูดเพศหญิง
 □ แทน ค่าพิตช์ของสัญญาณเสียงพูดเพศชาย
 ○ แทน ค่าพิตช์ของสัญญาณเสียงพูดเพศหญิง

ภาพที่ 4-14 ค่าพิตช์ของเสียงพูดทดสอบที่คำนวณจากวิเคราะห์ Cepstrum โดยเพิ่มการลบ DC Bias และ Butterworth Filter เข้าไปในกระบวนการประมวลผลสัญญาณเบื้องต้น

4.3.7 ผลการทดลองขั้นตอนวิธีการหาค่าพิตช์ของเสียงพูดด้วยวิธี Autocorrelation โดยมีการประมวลผลสัญญาณเบื้องต้นด้วยการลบ DC Bias การวางฟังก์ชันกรอบสัญญาณ และการ Clipping สัญญาณเสียง

ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานอ้างอิงของกลุ่มเรียนรู้เพศชาย คือ 148.10 และ 15.02 ตามลำดับ ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของกลุ่มเรียนรู้เพศหญิง คือ 228.26 และ 20.27 ตามลำดับ จากนั้นจึงเข้าสู่กระบวนการทดสอบ จากการทดสอบพบว่าระบบมีร้อยละความถูกต้องจากการจำแนกเสียงพูดเพศชายอยู่ที่ 91.80 และร้อยละความถูกต้องจากการจำแนกเสียงพูดเพศหญิงคือ 92.52 ดังภาพที่ 4-15

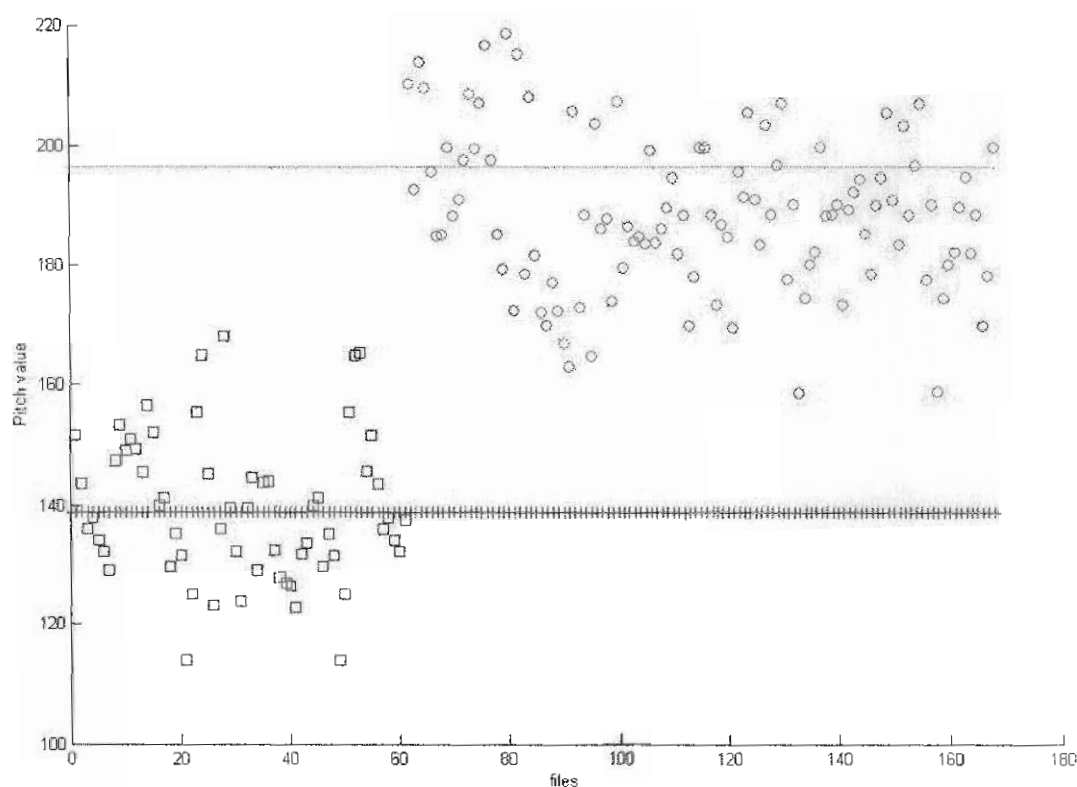


- เมื่อ +++ แทน ค่าพิตช์อ้างอิงเสียงพูดเพศชาย
 --- แทน ค่าพิตช์อ้างอิงเสียงพูดเพศหญิง
 □ แทน ค่าพิตช์ของสัญญาณเสียงพูดเพศชาย
 ○ แทน ค่าพิตช์ของสัญญาณเสียงพูดเพศหญิง

ภาพที่ 4-15 ค่าพิตช์ของเสียงพูดทดสอบที่คำนวณจากวิธี Autocorrelation ที่เพิ่มการลบ DC Bias, Butterworth Filter และการ Clipping สัญญาณเสียงพูดเข้าไปในกระบวนการประมวลผลสัญญาณเบื้องต้น

4.3.8 ผลการทดลองขั้นตอนวิธีการหาค่าพิตช์ของเสียงพูดด้วยวิธี Cepstrum โดยมีการประมวลผลสัญญาณเบื้องต้นด้วยการลบ DC Bias ร่วมกับการวางฟังก์ชันกรอบสัญญาณ และการ Clipping สัญญาณเสียง

ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานอ้างอิงของกลุ่มเรียนรู้เพศชาย คือ 138.83 และ 12.32 ตามลำดับ ค่าเฉลี่ย และค่าเบี่ยงเบนมาตรฐานของกลุ่มเรียนรู้เพศหญิง คือ 196.60 และ 11.21 ตามลำดับ จากนั้นจึงเข้าสู่กระบวนการทดสอบ จากการทดสอบพบว่าระบบมีร้อยละความถูกต้องจากการจำแนกเสียงพูดเพศชายอยู่ที่ 100 และร้อยละความถูกต้องจากการจำแนกเสียงพูดเพศหญิงคือ 95.33 ดังภาพที่ 4-16



- เมื่อ +++ แทน ค่าพิตช์อ้างอิงเสียงพูดเพศชาย
 --- แทน ค่าพิตช์อ้างอิงเสียงพูดเพศหญิง
 □ แทน ค่าพิตช์ของสัญญาณเสียงพูดเพศชาย
 ○ แทน ค่าพิตช์ของสัญญาณเสียงพูดเพศหญิง

ภาพที่ 4-16 ค่าพิตช์ของเสียงพูดทดสอบที่คำนวณจากวิธีวิเคราะห์ Cepstrum ที่เพิ่มการลบ DC Bias, Butterworth Filter และการ Clipping สัญญาณเสียงพูดเข้าไปในกระบวนการประมวลผลสัญญาณเบื้องต้น

4.4 วิเคราะห์ผลการวิจัย

จุดมุ่งหมายของการสร้างระบบจำแนกเพศของบุคคลในวิทยานิพนธ์นี้ เพื่อต้องการเพิ่มประสิทธิภาพขั้นตอนวิธีการประมวลผลสัญญาณเบื้องต้นที่เหมาะสมกับการประมาณค่าความถี่มูลฐานเพื่อนำไปใช้กับระบบการจำแนกเพศบุคคลด้วยคำพิตซ์ของเสียงพูดภาษาไทย จากกราฟแสดงประสิทธิภาพในการจำแนกเพศบุคคลในหัวข้อ 4.2 และ 4.3 สามารถสรุปผลแยกออกเป็น 2 กรณี ดังนี้

4.4.1 กรณีใช้ประโยคคำพูดเหมือนกัน พบว่าขั้นตอนวิธีการประมวลผลสัญญาณเบื้องต้นก่อนการประมาณค่าความถี่มูลฐานวิธี Autocorrelation ขั้นตอนวิธีที่ 1 สามารถจำแนกเสียงพูดที่มีความถี่สูงได้ดีกว่า ดังจะเห็นได้จากประสิทธิภาพในการจำแนกเสียงพูดเพศหญิงมีร้อยละความถูกต้องสูงกว่าเสียงพูดเพศชาย ขั้นตอนวิธีที่ 2 เมื่อนำการ Clipping สัญญาณเสียงมาช่วยในการประมวลผลสัญญาณเบื้องต้น ผลลัพธ์ที่ได้แสดงให้เห็นว่าขั้นตอนวิธีนี้สามารถเพิ่มประสิทธิภาพการจำแนกเสียงพูดเพศชายและเพศหญิงได้ดีขึ้นแต่ยังมีประสิทธิภาพในการจำแนกเสียงพูดเพศชายต่ำกว่าขั้นตอนวิธีที่ 3 ซึ่งมีร้อยละความถูกต้องในเสียงพูดเพศชายสูงกว่า แต่ขั้นตอนวิธีที่สามารถเพิ่มประสิทธิภาพการจำแนกเพศบุคคลได้ดีที่สุดคือ ขั้นตอนวิธีที่ 7 ซึ่งเป็นการนำทุกขั้นตอนวิธีมารวมกันทำให้ได้ร้อยละความถูกต้องของเสียงพูดเพศชายและเพศหญิงดีที่สุดที่ 95.71 และ 97.78 ตามลำดับ

ตารางที่ 4-3 การเปรียบเทียบร้อยละความถูกต้องของกระบวนการ Autocorrelation

หัวข้อ	กระบวนการ	ร้อยละความ ถูกต้องเสียงพูด เพศชาย	ร้อยละความ ถูกต้องเสียงพูด เพศหญิง
4.2.1	Autocorrelation	91.43	94.44
4.2.3	Clipping+ Autocorrelation	92.86	96.67
4.2.5	Remove DC+ Autocorrelation	94.29	94.44
4.2.7	Remove DC+ Clipping+ Autocorrelation	95.71	97.78

หมายเหตุ : “+” หมายถึง การใช้งานร่วมกัน

ในกระบวนการประมาณค่าความถี่มูลฐานด้วยวิธีการวิเคราะห์แบบ Cepstrum พบว่าการวิเคราะห์แบบ Cepstrum ซึ่งเป็นการวิเคราะห์ในโดเมนความถี่มีประสิทธิภาพในการจำแนกเพศบุคคลจากเสียงพูดเพศชายดีกว่าเสียงพูดเพศหญิง โดยไม่ว่าจะทำการเพิ่มประสิทธิภาพการประมวลผลสัญญาณเบื้องต้นด้วยวิธีใดก็มีความมีประสิทธิภาพเท่ากัน แต่สำหรับการจำแนกเพศบุคคลเพศหญิงการประมวลผลสัญญาณเบื้องต้นด้วยวิธีลบ DC Bias มีประสิทธิภาพสูงที่สุด

ตารางที่ 4-4 การเปรียบเทียบร้อยละความถูกต้องของกระบวนการวิเคราะห์ Cepstrum

หัวข้อ	กระบวนการ	ร้อยละความถูกต้อง เสียงพูดเพศชาย	ร้อยละความถูกต้อง เสียงพูดเพศหญิง
4.2.2	Cepstrum	98.57	87.78
4.2.4	Clipping+ Cepstrum	98.57	90
4.2.6	Remove DC+ Cepstrum	98.57	91.11
4.2.8	Remove DC+ Clipping+ Cepstrum	98.57	90

หมายเหตุ : “+” หมายถึง การใช้งานร่วมกัน

4.4.2 กรณีใช้ประโยคคำพูดต่างกัน พบว่าขั้นตอนวิธีการประมวลผลสัญญาณเบื้องต้นก่อนการประมาณค่าความถี่มูลฐานด้วยวิธี Autocorrelation นั้น ขั้นตอนวิธีการลบ DC Bias มีประสิทธิภาพสูงที่สุดเมื่อใช้ในการจำแนกเสียงพูดเพศชาย ส่วนการจำแนกเสียงพูดเพศหญิงขั้นตอนวิธีการลบ DC Bias ร่วมกับการ Clipping สัญญาณเสียงมีประสิทธิภาพสูงที่สุด

ตารางที่ 4-5 การเปรียบเทียบร้อยละความถูกต้องของกระบวนการ Autocorrelation

หัวข้อ	กระบวนการ	ร้อยละความ ถูกต้องเสียงพูด เพศชาย	ร้อยละความ ถูกต้องเสียงพูด เพศหญิง
4.3.1	Autocorrelation	78.69	64.49
4.3.3	Clipping+ Autocorrelation	86.89	71.96
4.3.5	Remove DC+ Autocorrelation	96.72	84.11
4.3.7	Remove DC+ Clipping+ Autocorrelation	91.80	92.52

หมายเหตุ : “+” หมายถึง การใช้งานร่วมกัน

ในกระบวนการวิเคราะห์ Cepstrum จากการทดลองพบว่าขั้นตอนวิธีการ Clipping สัญญาณเสียงช่วยเพิ่มประสิทธิภาพการจำแนกเสียงพูดเพศชายได้เพิ่มขึ้นจากเดิมร้อยละ 91.8032 เป็นร้อยละ 98.3606 แต่ขั้นตอนวิธีการลบ DC Bias สามารถเพิ่มประสิทธิภาพการจำแนกเพศบุคคลด้วยการประมาณค่าพิตช์ของเสียงพูดเพศชายด้วยวิธี Cepstrum ได้ดีที่สุด สำหรับกรณีประโยคเสียงพูดต่างกัน เนื่องจากสามารถจำแนกเสียงพูดเพศชายได้ถึงร้อยละ 100 คือไม่มีเสียงพูดเพศชายเสียงใดที่ระบบจำแนกผิดเลย ส่วนการจำแนกเสียงพูดเพศหญิงนั้นทุกขั้นตอนวิธีมีร้อยละความถูกต้องเท่ากันคือ 95.3271

ตารางที่ 4-6 การเปรียบเทียบร้อยละความถูกต้องของกระบวนการวิเคราะห์ Cepstrum

หัวข้อ	กระบวนการ	ร้อยละความถูกต้อง เสียงพูดเพศชาย	ร้อยละความถูกต้อง เสียงพูดเพศหญิง
4.3.2	Cepstrum	91.80	95.33
4.3.4	Clipping+ Cepstrum	98.36	95.33
4.3.6	Remove DC+ Cepstrum	100	95.33
4.3.8	Remove DC+ Clipping+ Cepstrum	100	95.33

หมายเหตุ : “+” หมายถึง การใช้งานร่วมกัน

บทที่ 5

สรุปผลการวิจัยและข้อเสนอแนะ

5.1 สรุปผลการวิจัย

จากผลการวิจัยการจำแนกเพศบุคคลโดยใช้ค่าพิตช์ของเสียงพูดภาษาไทย พบว่ากรณีเสียงพูดประโยคเดียวกันด้วยคำว่า “กาลครั้งหนึ่งนานมาแล้ว” เมื่อนำไปทดลองกับกระบวนการประมาณค่าความถี่มูลฐานด้วยวิธี Autocorrelation นั้น พบว่ากระบวนการประมวลผลสัญญาณเบื้องต้นด้วยฟังก์ชันกรอบสัญญาณแบบสี่เหลี่ยมมีร้อยละความถูกต้องเพศชาย 91.43 และเพศหญิง 94.44 จะเห็นว่าระบบสามารถจำแนกเสียงพูดเพศหญิงได้ดีกว่าเสียงพูดเพศชาย เมื่อนำกระบวนการ Clipping สัญญาณเสียงมาใช้พบว่าระบบมีประสิทธิภาพเพิ่มมากขึ้นเป็นร้อยละความถูกต้องเพศชาย 92.86 และเพศหญิง 96.67 เมื่อนำกระบวนการลบ DC Bias มาใช้พบว่าร้อยละความถูกต้องเพศชายอยู่ที่ 94.29 และเพศหญิง 94.44 จะเห็นว่าเมื่อเปรียบเทียบวิธีการที่ 4.2.3 กับวิธีการที่ 4.2.5 พบว่าวิธีการที่ 4.2.5 มีประสิทธิภาพสูงกว่าในกรณีที่เป็นการจำแนกเสียงพูดเพศชาย แต่จะมีประสิทธิภาพต่ำกว่าในกรณีที่เป็นการจำแนกเสียงพูดเพศหญิง และเมื่อนำทั้งกระบวนการลบ DC Bias และกระบวนการ Clipping สัญญาณเสียงมาใช้งานร่วมกันแล้วพบว่าประสิทธิภาพในการจำแนกเพศบุคคลด้วยการประมาณค่าพิตช์ด้วยวิธี Autocorrelation มีประสิทธิภาพสูงที่สุดคือ ร้อยละความถูกต้องเพศชาย 95.71 และเพศหญิง 97.78

กรณีเสียงพูดประโยคเดียวกันด้วยคำว่า “กาลครั้งหนึ่งนานมาแล้ว” เมื่อนำไปทดลองกับกระบวนการประมาณค่าความถี่มูลฐานด้วยวิธีวิเคราะห์แบบ Cepstrum พบว่ากระบวนการประมวลผลสัญญาณเบื้องต้นด้วยฟังก์ชันกรอบสัญญาณแบบแฮมมิงมีร้อยละความถูกต้องเพศชาย 98.57 และเพศหญิง 87.78 จะเห็นว่าระบบสามารถจำแนกเสียงพูดเพศชายได้ดีกว่าเสียงพูดเพศหญิง เมื่อนำกระบวนการ Clipping สัญญาณเสียงมาใช้พบว่าระบบมีประสิทธิภาพเพิ่มมากขึ้นในกรณีเสียงพูดเพศหญิงคือเพิ่มขึ้นเป็นร้อยละ 90.00 แต่ในเสียงพูดเพศชายมีประสิทธิภาพเท่าเดิม เนื่องจากเดิมระบบก็สามารถจำแนกได้อย่างมีประสิทธิภาพสูง คือจำแนกเสียงพูดเพศชายผิดเพียงสัญญาณเสียงเดียวซึ่งเป็นเสียงพูดเพศชายที่มีเสียงสูงกว่าเสียงพูดเพศชายทั่วไป เมื่อนำกระบวนการลบ DC Bias เข้ามาเพิ่มประสิทธิภาพพบว่าสามารถเพิ่มประสิทธิภาพได้เพิ่มขึ้นในเสียงพูดเพศหญิงคือมีร้อยละความถูกต้อง 91.11 แต่ไม่สามารถเพิ่มประสิทธิภาพการจำแนกเสียงพูดเพศชายได้เช่นกัน และเมื่อนำทั้งกระบวนการลบ DC Bias และกระบวนการ Clipping สัญญาณเสียงมาใช้งานร่วมกันแล้วพบว่าประสิทธิภาพในการจำแนกเพศบุคคลด้วยการประมาณค่าพิตช์ด้วยวิธี Cepstrum ไม่ได้มีประสิทธิภาพเพิ่มสูงขึ้น ดังนั้นในกรณีนี้

การประมวลผลสัญญาณเบื้องต้นด้วยขั้นตอนวิธีการลบ DC Bias มีประสิทธิภาพสูงที่สุดคือมีร้อยละความถูกต้องเพศชาย 98.57 และเพศหญิง 91.11

ในกรณีคำพูดในประโยคต่างกันด้วยประโยคแนะนำตัวของกลุ่มตัวอย่าง เมื่อนำไปทดลองกับกระบวนการประมาณค่าความถี่มูลฐานด้วยวิธี Autocorrelation นั้น พบว่ากระบวนการประมวลผลสัญญาณเบื้องต้นด้วยฟังก์ชันกรอบสัญญาณแบบสี่เหลี่ยมมีร้อยละความถูกต้องเพศชาย 78.69 และเพศหญิง 64.49 จะเห็นว่าระบบสามารถจำแนกเสียงพูดเพศชายได้ดีกว่าเสียงพูดเพศหญิงเล็กน้อย เมื่อนำกระบวนการ Clipping สัญญาณเสียงมาใช้พบว่าระบบมีประสิทธิภาพเพิ่มมากขึ้นเป็นร้อยละความถูกต้องเพศชาย 86.89 และเพศหญิง 71.96 เมื่อนำกระบวนการลบ DC Bias มาใช้พบว่าร้อยละความถูกต้องเพศชายอยู่ที่ 96.72 และ เพศหญิง 84.11 จะเห็นว่าการลบ DC Bias มีประสิทธิภาพในการจำแนกเพศบุคคลดีกว่าวิธีการ Clipping สัญญาณเสียงทั้งในสัญญาณเสียงเพศชายและเพศหญิง และเมื่อนำทั้งกระบวนการลบ DC Bias และกระบวนการ Clipping สัญญาณเสียงมาใช้งานร่วมกันแล้วพบว่าประสิทธิภาพในการจำแนกเพศบุคคลสูงขึ้นกว่าขั้นตอนวิธีที่ใช้ Clipping สัญญาณเสียงเพียงอย่างเดียวแต่มีประสิทธิภาพต่ำกว่าขั้นตอนวิธีที่ใช้การลบ DC Bias ดังนั้นขั้นตอนวิธีการจำแนกเพศบุคคลด้วยการประมาณค่าพิตช์ด้วยวิธี Autocorrelation ที่มีประสิทธิภาพสูงที่สุดในกรณีใช้คำพูดในประโยคต่างกันคือขั้นตอนวิธีการประมวลผลสัญญาณเบื้องต้นด้วยวิธีการลบ DC Bias

เมื่อนำไปทดลองกับกระบวนการประมาณค่าความถี่มูลฐานด้วยวิธีวิเคราะห์แบบ Cepstrum พบว่ากระบวนการประมวลผลสัญญาณเบื้องต้นด้วยฟังก์ชันกรอบสัญญาณแบบสามมิ่งมีร้อยละความถูกต้องเพศชาย 91.80 และเพศหญิง 95.33 จะเห็นว่าระบบสามารถจำแนกเสียงพูดเพศหญิงได้ดีกว่าเสียงพูดเพศชาย เมื่อนำกระบวนการ Clipping สัญญาณเสียงมาใช้พบว่าระบบมีประสิทธิภาพเพิ่มมากขึ้นในกรณีเสียงพูดเพศชายคือเพิ่มขึ้นเป็นร้อยละ 98.36 แต่ในเสียงพูดเพศหญิงมีประสิทธิภาพเท่าเดิม เมื่อนำกระบวนการลบ DC Bias เข้ามาเพิ่มประสิทธิภาพพบว่าสามารถเพิ่มประสิทธิภาพได้เพิ่มขึ้นร้อยละความถูกต้อง 100 คือไม่มีการจำแนกเสียงพูดเพศชายผิดเลย แต่ไม่สามารถเพิ่มประสิทธิภาพการจำแนกเสียงพูดเพศหญิงได้เช่นกัน และเมื่อนำทั้งกระบวนการลบ DC Bias และกระบวนการ Clipping สัญญาณเสียงมาใช้งานร่วมกันแล้วพบว่าประสิทธิภาพในการจำแนกเพศบุคคลด้วยการประมาณค่าพิตช์ด้วยวิธี Cepstrum สามารถจำแนกเสียงพูดเพศชายได้ถูกต้องถึงร้อยละ 100 เช่นเดียวกับวิธีการลบ DC Bias แต่ไม่สามารถเพิ่มประสิทธิภาพในเสียงพูดเพศหญิงได้เช่นกัน

ขั้นตอนวิธีการประมวลผลเสียงพูดเบื้องต้นที่เหมาะสมกับการประมาณค่าพิตช์ด้วยวิธี Autocorrelation คือขั้นตอนวิธีการลบ DC Bias ร่วมกับกระบวนการ Clipping สัญญาณเสียง และขั้นตอนวิธีการประมวลผลเสียงพูดเบื้องต้นที่เหมาะสมกับการประมาณค่าพิตช์ด้วยวิธีวิเคราะห์แบบ Cepstrum คือขั้นตอนวิธีการลบ DC Bias และจากการนำเอาวิธีการลบ DC Bias มาเพิ่มประสิทธิภาพการประมวลผลสัญญาณเบื้องต้นเพื่อเตรียมความพร้อมให้สัญญาณเสียงพูดก่อนนำไปประมาณค่าความถี่มูลฐานสำหรับการจำแนกเพศบุคคลด้วยค่าพิตช์ พบว่าวิธีการลบ

DC Bias สามารถช่วยเพิ่มประสิทธิภาพให้กับการจำแนกเพศบุคคลด้วยค่าพิตซ์ได้เป็นอย่างดี ทั้งการใช้งานร่วมกับวิธี Autocorrelation และ วิธีวิเคราะห์แบบ Cepstrum โดยสามารถเพิ่มประสิทธิภาพได้มากกว่าการใช้กระบวนการ Clipping สัญญาณเสียง ตารางที่ 5-1 แสดงการเปรียบเทียบประสิทธิภาพขั้นตอนวิธีการประมวลผลสัญญาณเบื้องต้น

ตารางที่ 5-1 การเปรียบเทียบประสิทธิภาพขั้นตอนวิธีการประมวลผลสัญญาณเบื้องต้น

วิธีการประมวลค่า ความถี่มูลฐาน	เสียงพูดเพศชาย			เสียงพูดเพศหญิง		
	การ Clipping	การลบ DC Bias	สองวิธี	การ Clipping	การลบ DC Bias	สองวิธี
กาลครั้งหนึ่งนานมาแล้ว :						
Autocorrelation	92.86	94.29	95.71	96.67	94.44	97.78
Cepstrum	98.57	98.57	98.57	90	91.11	90
แนะนำตัว :						
Autocorrelation	86.89	96.72	91.80	71.96	84.11	92.52
Cepstrum	98.36	100	100	95.33	95.33	95.33

5.2 อภิปรายผลการวิจัย

จากการสรุปผลการวิจัยข้างต้นนั้น จะเห็นได้ว่างานวิจัยนี้ได้ผลลัพธ์ตรงตามวัตถุประสงค์การวิจัย คือ สามารถเพิ่มประสิทธิภาพขั้นตอนวิธีการประมวลผลสัญญาณเบื้องต้นของสัญญาณเสียงเพื่อกรองสัญญาณเสียงพร้อมทั้งปรับเรียบสัญญาณเสียงก่อนนำไปประมวลค่าพิตซ์สำหรับการจำแนกเพศบุคคลด้วยเสียงพูดภาษาไทย โดยกระบวนการประมวลผลสัญญาณเบื้องต้นที่มีประสิทธิภาพได้จากการเพิ่มขั้นตอนวิธีการลบ DC Bias ของสัญญาณเสียงนั่นเอง การจำแนกเพศบุคคลสามารถจำแนกเพศบุคคลจากเสียงพูดได้อย่างมีประสิทธิภาพเป็นที่น่าพอใจคือ มีร้อยละความถูกต้องเกินกว่า 80 ในงานวิจัยครั้งนี้ยังพบข้อผิดพลาดอยู่บ้างในการจำแนกเพศบุคคลจากสัญญาณเสียงพูดเพศชายที่มีระดับเสียงสูงกว่าเสียงพูดเพศชายทั่วไป และเสียงพูดเพศหญิงที่มีลักษณะแหบ และพูดเสียงเบาว่าปกติ

ในการนำขั้นตอนวิธีดังกล่าวไปประยุกต์ใช้งานจริงกับระบบการจำแนกเพศบุคคล เสียงพูดที่เข้าสู่ระบบมักจะไม่ถูกกำหนดให้เป็นการพูดประโยคเดียวกันเพื่อประสิทธิภาพในด้านความยืดหยุ่นของระบบ ดังนั้นขั้นตอนวิธีการลบ DC Bias ของสัญญาณเสียงจึงเป็นวิธีการที่เหมาะสมที่จะนำมาใช้ในขั้นตอนวิธีการประมวลผลสัญญาณเบื้องต้นเพื่อนำไปใช้ร่วมกับการประมวลค่าความถี่มูลฐาน ดังผลการทดลองในตารางที่ 5-1 จะเห็นได้ว่าประโยคแนะนำตัวซึ่งเป็นประโยคคำพูดที่ไม่เฉพาะเจาะจงมีประสิทธิภาพในการจำแนกเสียงพูดเพศชายด้วยขั้นตอนวิธีการลบ DC

Bias ของสัญญาณเสียงมากที่สุด ทั้งการใช้งานร่วมกับการประมาณค่าความถี่มูลฐานวิธี Autocorrelation และวิธีวิเคราะห์แบบ Cepstrum

5.3 ข้อเสนอแนะ

จากการอภิปรายผลข้อสรุป จะเห็นว่างานวิจัยนี้ยังสามารถนำไปพัฒนาต่อเพิ่มเติมให้มีประสิทธิภาพเพิ่มขึ้นได้อีก โดยควรปรับปรุงในส่วนต่าง ๆ ดังต่อไปนี้

1. การปรับปรุงกรอบสัญญาณเสียงที่เหมาะสมสำหรับเสียงพูดเพศชายเสียงสูง และเสียงพูดเพศหญิงเสียงต่ำในขั้นตอนการประมวลผลสัญญาณเบื้องต้นก่อนการประมาณค่าความถี่มูลฐาน

2. การเพิ่มการหาคุณลักษณะเด่นอื่นของสัญญาณเสียงมาใช้ประกอบกับการประมาณค่าความถี่มูลฐานที่ได้จากงานวิจัยนี้เพื่อใช้ประกอบการตัดสินใจในกระบวนการจำแนกเพศบุคคลให้มีความถูกต้องมากที่สุด

เอกสารอ้างอิง

1. David Gerhard. "Pitch Extraction and Fundamental Frequency : History and Current Techniques." Technical Report TR-CS 2003-06. Department of Computer Science University of Regina, 2003.
2. Philip Loizou. Colea: A matlab software tool for speech analysis [Online] 2003. [cited 11 August 2005]. Available from : <http://www.utdallas.edu/~loizou/speech/colea.htm>.
3. Dmitry Terez. "Fundamental frequency estimation using signal embedding in state space." Journal of the Acoustical Society of America. 112(5) (November 2002) : 2279.
4. _____. "Robust pitch determination using nonlinear state-space embedding." International Conference on Acoustics, Speech and Signal Processing. volume I (2002) : 345-348.
5. Li Tan and Montri Karnjanadecha. Pitch Detection Algorithm : Autocorrelation method and AMDF. Thesis, Faculty of Engineering, Prince of Songkhla University, 2003.
6. สราญ บุญญานุสนธิ์. ความรู้ทั่วไปเกี่ยวกับเรื่องหู การได้ยินเสียงและการสอนพูดแก่เด็กหูตึง. กรุงเทพมหานคร : ศักติโสภาคการพิมพ์, 2532.
7. ณัฏชา โสคติยานุรักษ์. ทฤษฎีดนตรี. กรุงเทพมหานคร : สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย, 2543.
8. วิศรุต อาชุนบุตร. ระบบการรู้จำคำไทยหลายพยางค์แบบไม่ขึ้นกับผู้พูดโดยใช้แบบจำลองฮิดเดนมาร์คอฟ. วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต ภาควิชาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2539.
9. Lawrence R. Rabiner. "On the Use of Autocorrelation Analysis for Pitch Detection." IEEE Transactions on Acoustics, Speech, and Signal Processing. vol. 25 (February 1977) : 24-33.
10. ณัฏฐา จิตติวรังกุล. ขั้นตอนการหาขอบเขตพยางค์สำหรับคำพูดต่อเนื่องภาษาไทย. วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต ภาควิชาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2541.

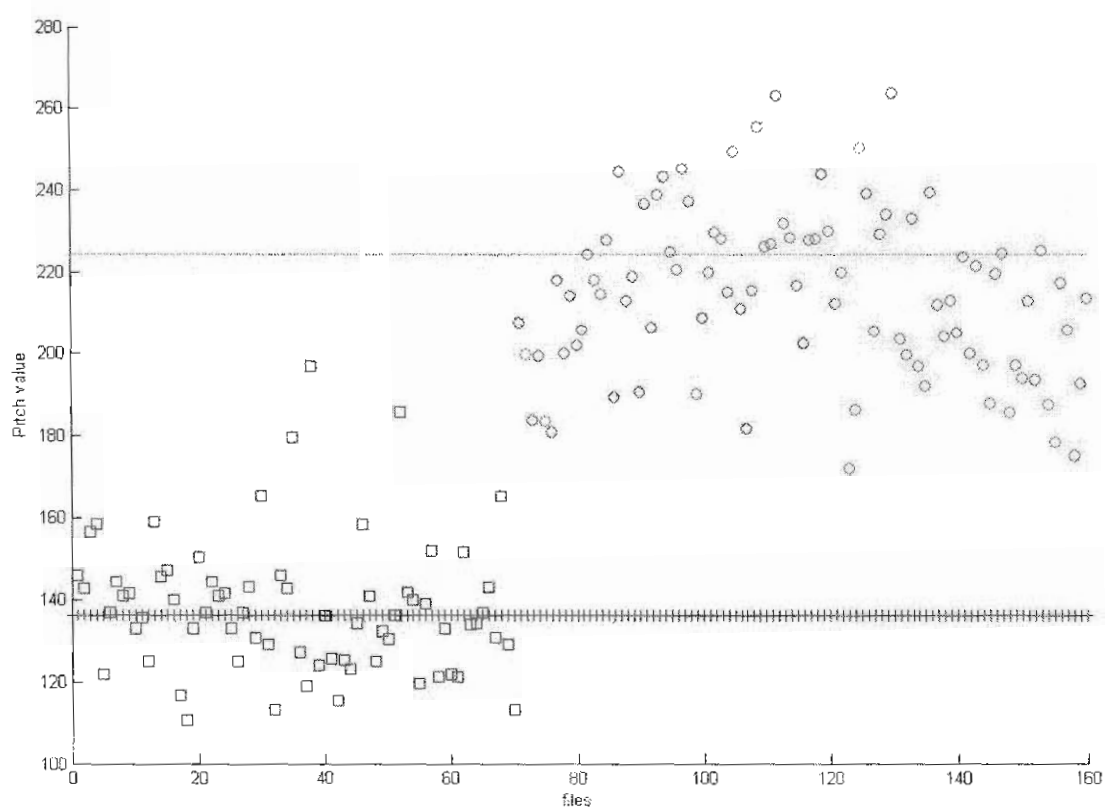
11. L. R. Rabiner, et al. "A comparative performance study of several pitch detection algorithms." IEEE Transactions on Audio, Signal, and Speech Processing. vol. 24 (1976) : 399-417.
12. Stephen Butterworth. "On the Theory of Filter Amplifiers." Experimental Wireless and the Radio Engineer. vol. 7 (1930) : 536-541.
13. Eluned S Parris and Michael J Carey. "Language Independent Gender Identification." IEEE International Conference on Acoustics, Speech, and Signal Processing. vol. 2 (1996) : 685-688.
14. Paul Boersma. Accurate short-term analysis of the Fundamental Frequency and The Harmonics-to-noise ratio of a sampled sound. Phonetic Sciences, University of Amsterdam, 1993.

ภาคผนวก ก

ผลการทดสอบประสิทธิภาพกรอบสัญญาณที่เหมาะสมกับแต่ละวิธีการประมาณค่าความถี่มูลฐาน

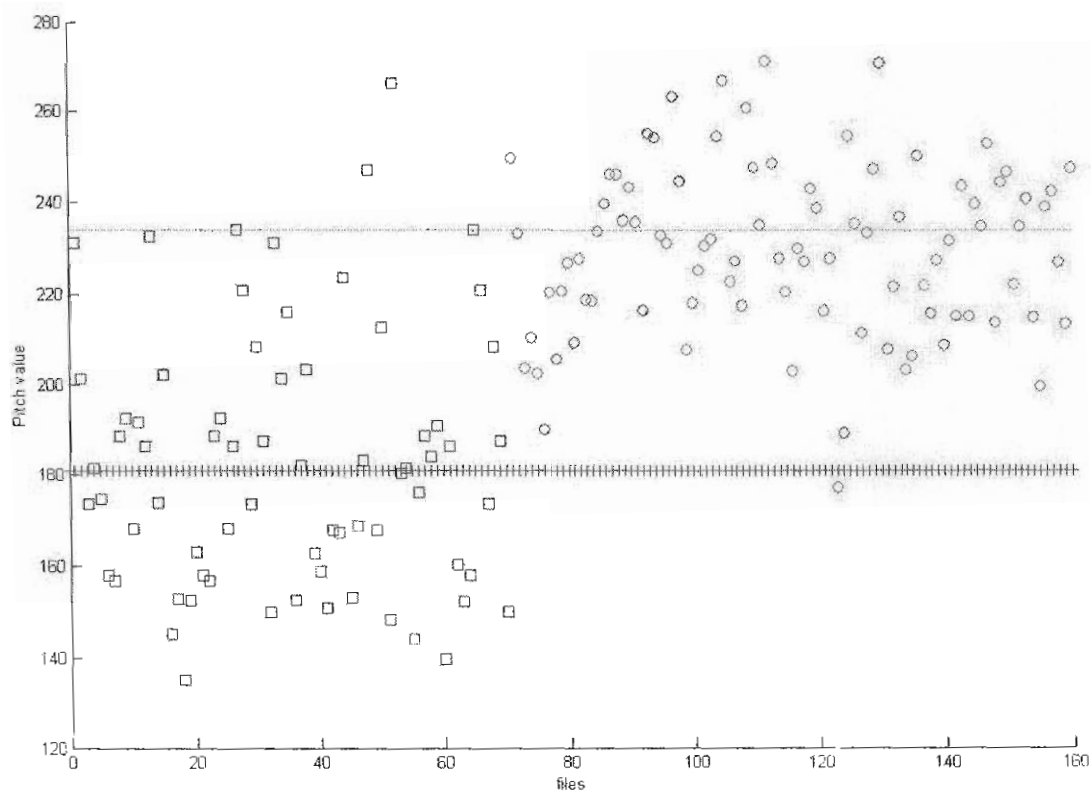
ผลการทดสอบประสิทธิภาพฟังก์ชันกรอบสัญญาณที่เหมาะสมกับแต่ละวิธีการประมาณค่าความถี่มูลฐาน

ผู้วิจัยได้ทำการทดสอบฟังก์ชันกรอบสัญญาณแบบสี่เหลี่ยม (Rectangular Window) และกรอบสัญญาณแบบแฮมมิง (Hamming Window) เพื่อให้ทราบถึงประสิทธิภาพในการใช้งาน ร่วมกับการประมาณค่าความถี่มูลฐานวิธี Autocorrelation และวิธีการวิเคราะห์แบบ Cepstrum ดังแสดงในภาพที่ ก-1, ก-2, ก-3 และ ก-4



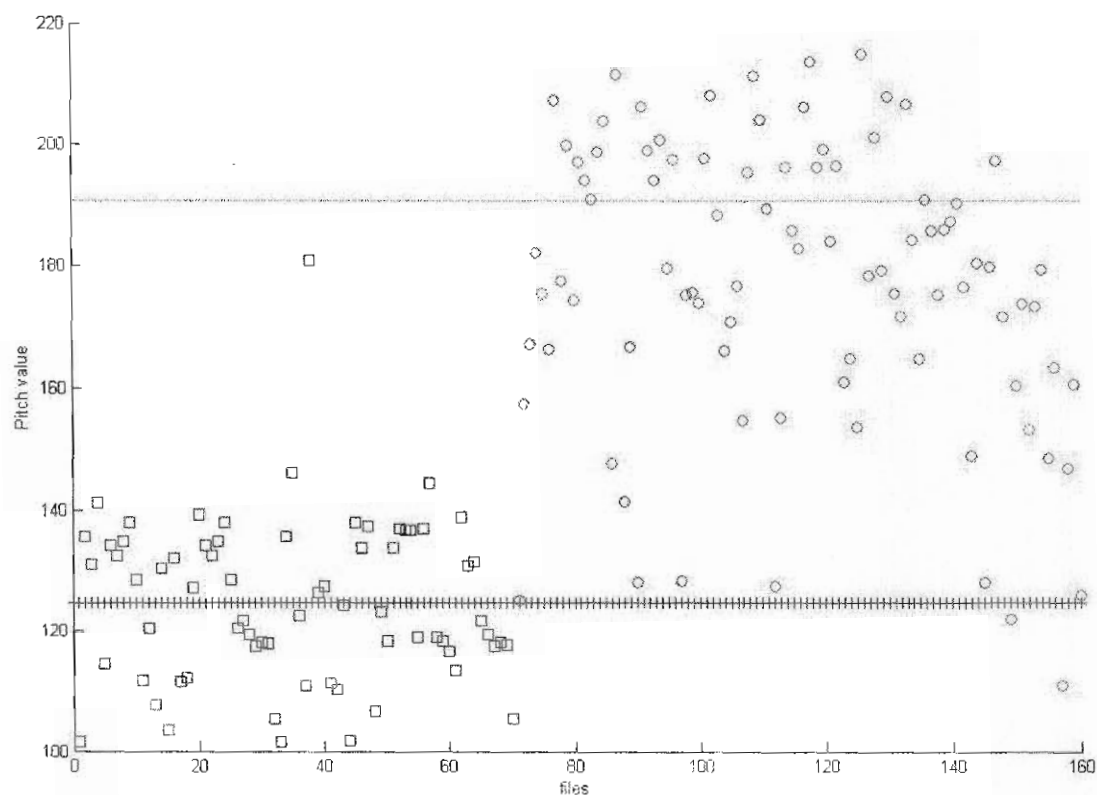
- เมื่อ +++ แทน ค่าพิตช์อ้างอิงเสียงพูดเพศชาย
 --- แทน ค่าพิตช์อ้างอิงเสียงพูดเพศหญิง
 □ แทน ค่าพิตช์ของสัญญาณเสียงพูดเพศชาย
 ○ แทน ค่าพิตช์ของสัญญาณเสียงพูดเพศหญิง

ภาพที่ ก-1 ประสิทธิภาพการจำแนกเพศบุคคลด้วยการประมาณค่าความถี่มูลฐานด้วยวิธี Autocorrelation โดยใช้กรอบสัญญาณแบบสี่เหลี่ยม



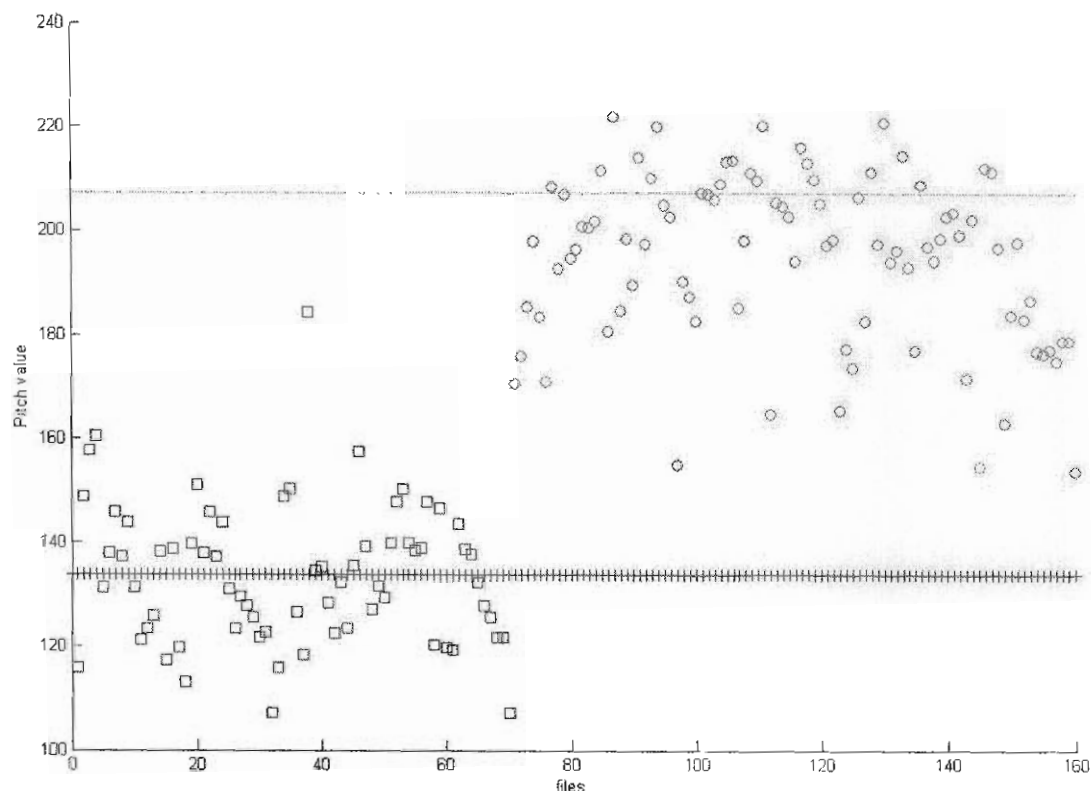
- เมื่อ +++ แทน ค่าพิตช์อ้างอิงเสียงพูดเพศชาย
 --- แทน ค่าพิตช์อ้างอิงเสียงพูดเพศหญิง
 □ แทน ค่าพิตช์ของสัญญาณเสียงพูดเพศชาย
 ○ แทน ค่าพิตช์ของสัญญาณเสียงพูดเพศหญิง

ภาพที่ ก-2 ประสิทธิภาพการจำแนกเพศบุคคลด้วยการประมาณค่าความมูลฐานด้วยวิธี Autocorrelation โดยใช้กรอบสัญญาณแบบแสมมิ่ง



- เมื่อ +++ แทน ค่าพิตช์อ้างอิงเสียงพูดเพศชาย
 --- แทน ค่าพิตช์อ้างอิงเสียงพูดเพศหญิง
 □ แทน ค่าพิตช์ของสัญญาณเสียงพูดเพศชาย
 ○ แทน ค่าพิตช์ของสัญญาณเสียงพูดเพศหญิง

ภาพที่ ก-3 ประสิทธิภาพการจำแนกเพศบุคคลด้วยการประมาณค่าความมูลฐานด้วยวิธี Cepstrum โดยใช้กรอบสัญญาณแบบสี่เหลี่ยม



- เมื่อ +++ แทน ค่าพิตช์อ้างอิงเสียงพูดเพศชาย
 --- แทน ค่าพิตช์อ้างอิงเสียงพูดเพศหญิง
 □ แทน ค่าพิตช์ของสัญญาณเสียงพูดเพศชาย
 ○ แทน ค่าพิตช์ของสัญญาณเสียงพูดเพศหญิง

ภาพที่ ก-4 ประสิทธิภาพการจำแนกเพศบุคคลด้วยการประมาณค่าความมูลฐานด้วยวิธี Cepstrum โดยใช้กรอบสัญญาณแบบแฮมมิง

จากภาพที่ ก-1 และ ก-2 พบว่าฟังก์ชันกรอบสัญญาณแบบสี่เหลี่ยมเหมาะสมกับการประมาณค่าความถี่มูลฐานด้วยวิธี Autocorrelation มากกว่าฟังก์ชันกรอบสัญญาณแบบแฮมมิงดังจะเห็นได้จากประสิทธิภาพในการจำแนกเพศบุคคลในภาพที่ ก-1 ระบบสามารถจำแนกเพศบุคคลได้อย่างชัดเจนแม้จะมีบางส่วนของสัญญาณเสียงพูดที่จำแนกผิดพลาดบ้าง แต่เมื่อนำฟังก์ชันกรอบสัญญาณแบบแฮมมิงมาใช้ในการประมาณค่าความถี่มูลฐานด้วยวิธี Autocorrelation แล้วจากภาพที่ ก-2 จะเห็นว่าการจำแนกเพศบุคคลมีประสิทธิภาพลดลงอย่างเห็นได้ชัด สัญญาณเสียงเพศชายปะปนอยู่ในบริเวณที่ควรจะเป็นสัญญาณเสียงเพศหญิงเป็นจำนวนมาก และสัญญาณเสียงเพศหญิงก็ปะปนอยู่ในบริเวณที่ควรจะเป็นสัญญาณเสียงเพศชายเช่นกัน

จากภาพที่ ก-3 และ ก-4 พบว่าฟังก์ชันกรอบสัญญาณแบบแฮมมิงเหมาะสมกับการประมาณค่าความถี่มูลฐานด้วยวิธี Cepstrum มากกว่าฟังก์ชันกรอบสัญญาณแบบสี่เหลี่ยมดังจะ

เห็นได้จากประสิทธิภาพในการจำแนกเพศบุคคลในภาพที่ ก-4 ระบบสามารถจำแนกเพศบุคคลได้อย่างชัดเจนแม้จะมีบางส่วนของสัญญาณเสียงพูดที่จำแนกผิดพลาดบ้าง แต่เมื่อนำฟังก์ชันกรอบสัญญาณแบบสี่เหลี่ยมมาใช้ในการประมาณค่าความถี่มูลฐานด้วยวิธี Cepstrum แล้วจากภาพที่ ก-3 จะเห็นว่าการจำแนกเพศบุคคลในสัญญาณเสียงพูดเพศหญิงมีประสิทธิภาพลดลงอย่างเห็นได้ชัด

ประวัติผู้วิจัย

ชื่อ : นางสาวนิชประภา วิรัชโกศล
ชื่อวิทยานิพนธ์ : การเพิ่มประสิทธิภาพของกระบวนการตรวจจับค่าพิตช์ของเสียงพูดภาษาไทย
เพื่อการจำแนกเพศบุคคล
สาขาวิชา : เทคโนโลยีคอมพิวเตอร์

ประวัติ

เกิดวันที่ 13 กันยายน 2520 ที่กรุงเทพมหานคร สำเร็จการศึกษาระดับปริญญาตรี สาขาวิชา
บัญชีต้นทุน คณะบัญชี จากมหาวิทยาลัยหอการค้าไทย