การจำลองพยัญชนะภาษาไทยด้วยเสียง
Thai Consonant Voice Model

ปัณณธร พัฒนชลาลัย
Pannatorn Pattanachalalai
วรปรัชญ์ พันทวีสุข
Voraprat Puntaveesuk

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต
สาขาวิชาวิศวกรรมอิเล็กทรอนิกส์
คณะวิศวกรรมศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
พ.ศ.2560

การจำลองพยัญชนะภาษาไทยด้วยเสียง

Thai Consonant Voice Model

โดย

ปัณณธร	พัฒนชลาลัย	รหัส	57010766
วรปรัชญ์	พันทวีสุข	รหัส	57011100

อาจารย์ที่ปรึกษา

ผศ.ดร.ยุทธนา คิดใจเดียว

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต

สาขาวิชาวิศวกรรมอิเล็กทรอนิกส์

คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ.2560

ใบรับรองวิชาวิชาการประยุกต์วงจรอิเล็กทรอนิกส์ 2

รายงาน วิชา Project2 ปีการศึกษา 2560

สาขา วิชาวิศวกรรมอิเล็กทรอนิกส์

คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เรื่อง การจำลองพยัญชนะภาษาไทยด้วยเสียง

ผู้จัดทำ

1.ปัณณธร พัฒนชลาลัย รหัส 57010766

2.วรปรัชญ์ พันทวีสุข รหัส 57011100

รายงานนี้ผ่านการตรวจสอบโดยอาจารย์ที่ปรึกษาแล้ว

.....

อาจารย์ที่ปรึกษา

(ผศ.ดร.ยุทธนา คิดใจเดียว)

18 / 05 / 2561

การจำลองพยัญชนะภาษาไทยด้วยเสียง

นายปัณณธร พัฒนชลาลัย, นายวรปรัชญ์ พันทวีสุข

อาจารย์ที่ปรึกษา: ผศ.ดร.ยุทธนา คิดใจเดียว

ภาควิชาวิศวกรรมอิเล็คทรอนิคส์

คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ภาษาพูดถือได้ว่าเป็นการสื่อสารของมนุษย์ที่เป็นธรรมชาติที่สุดวิธีการหนึ่ง การติดต่อระหว่างมนุษย์กับเครื่องจักรโดยอาศัยเสียง หรือภาษาพูดจึงนับเป็นก้าวที่สำคัญที่นำไปสู่ยุคใหม่ ระหว่างมนุษย์กับเครื่องจักร Speech Recognition หรือการรู้จำเสียงพูดเป็นการเปิดโอกาสให้คอมพิวเตอร์สามารถเข้าใจคำพูดของมนุษย์ได้ งานวิจัยนี้เสนอการจำลองพยัญชนะภาษาไทยด้วยเสียงโดยการนำเอาโครงข่ายประสาทเทียมซึ่งเป็นระบบจำลองการเรียนรู้ของสมองมนุษย์มาประยุกต์ใช้ในการศึกษารูปแบบของคุณลักษณะเด่นของคำ โดยใช้ Mel Frequency Cepstral Coefficient: MFCC ในการดึงลักษณะเด่นของสัญญาณเสียงด้วยวิธีการหาค่าสัมประสิทธิ์เซปสตรัมบนเมลสเกลเป็นวิธีการสกัดคุณลักษณะที่ใช้มากที่สุดในการรู้จำเสียงพูดอัตโนมัติ ผลที่ได้คือ เสียงของผู้ชายนั้น การใช้ค่า MFCC + ΔMFCC ให้ Accuracy Rate สูงที่สุด มีค่า 78.38% แต่ในส่วนของเสียงผู้หญิงนั้น การใช้ค่า MFCC + ΔMFCC + ΔΔMFCC +ΔE+ ΔΔE ให้ Accuracy Rate สูงที่สุด มีค่า 78.47%

คำสำคัญ – การจำลอง ; พยัญชนะภาษาไทย ; เสียง ; พูด ; คอมพิวเตอร์

Thai Consonant Voice Model

Mr.Pannatorn Pattanachalalai, Mr.Voraprat Puntaveesuk
Advisor: Asst.Prof.Dr.Yuttana Kitjaidure
Department of Electronics Engineering
Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang
Bangkok, Thailand

Spoken language is one of the most intuitive communication methods for humans. Interfacing between human and machine by using speech or spoken language is an important step towards the new era of human-computer interfacing. Speech recognition gives opportunity to the ability to understand human speech. This research presents Thai Consonant Voice Model by using neural network, which models a human brain for learning pattern of speech by Mel Frequency Cepstral Coefficient: MFCC to extract the characteristic of the voice signal by finding the cepstral coefficient on the Mel scale. The Mel scale is the method of extracting the most commonly used features of speech recognition automatically. The result is the male voice. The MFCC + ΔMFCC gives the highest accuracy of 78.38%. In terms of female voice, the MFCC + ΔMFCC + ΔΔMFCC + ΔE + ΔΔE give the highest accuracy of 78.47%

Keywords – Model; Thai Consonant ; Sound ; Speech ; Computer

กิตติกรรมประกาศ

โครงงานเรื่อง การจำลองพยัญชนะภาษาไทยด้วยเสียง “Thai Consonant Voice Model” จะสำเร็จลุล่วงไม่ได้ถ้าไม่ได้รับการช่วยเหลือจากอาจารย์ที่ปรึกษา ผศ.ดร.ยุทธนา คิดใจเดียว ภาควิชาวิศวกรรมอิเล็คทรอนิคส์ คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ที่ช่วยให้คำปรึกษา ช่วยแก้ไขปัญหาต่าง ๆ เกี่ยวกับโครงงานนี้

สุดท้ายนี้คณะผู้จัดทำหวังว่าโครงงานนี้จะเป็นประโยชน์สำหรับผู้ที่สนใจ และสามารถนำความรู้จากโครงงานนี้ไปใช้ในอนาคตได้

นายปัณณธร พัฒนชลาลัย

นายวรปรัชญ์ พันทวีสุข

คณะผู้จัดทำ

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	I
บทคัดย่อภาษาอังกฤษ	II
กิตติกรรมประกาศ	III
สารบัญ	IV
สารบัญรูป	VIII
สารบัญตาราง	IX
บทที่ 1 บทนำ	1
1.1 ที่มาและความสำคัญ	1
1.2 ความมุ่งหมายและวัตถุประสงค์	1
1.2.1สร้างระบบรู้จำหน่วยเสียงพยัญชนะภาษาไทยโดยใช้โครงข่ายประสาทเทียม	1
1.2.2เพื่อศึกษาข้อมูลเกี่ยวกับโปรแกรมที่ใช้ประมวลผลด้านเสียงโดยโปรแกรม MATLAB	1
1.2.3 เพื่อนำไปเป็นตัวอย่างหรือต่อยอดในโครงงานอื่นต่อไป	1
1.3 สมมติฐานของการศึกษา	
1.3.1 เข้าใจขั้นตอนและวิธีการจำลองพยัญชนะภาษาไทยด้วยเสียงผ่านคอมพิวเตอร์	1
1.3.2 สามารถใช้โปรแกรม MATLAB ในการวิเคราะห์ข้อมูลเสียง	1
1.4 ขอบเขตของโครงงาน	2
1.4.1 โปรแกรม MATLAB ที่ออกแบบสามารถวิเคราะห์เสียงพูดพยัญชนะภาษาไทยได้อย่างถูกต้อง	2
1.4.2 เข้าใจถึงขั้นตอนและกระบวนการที่นำมาใช้วิเคราะห์	2
1.5 ประโยชน์ที่คาดว่าจะได้รับ	2
1.5.1 สามารถใช้โปรแกรม MATLAB ในการวิเคราะห์เสียงได้	2
1.5.2 เข้าใจถึงกระบวนการวิเคราะห์เสียงพูดโดยใช้โปรแกรม MATLAB	2
1.5.3 ทราบถึงแนวทางที่เหมาะสมในการสร้างระบบรู้จำหน่วยเสียงพยัญชนะภาษาไทย	2
1.5.4 สามารถนำความรู้ที่ได้ไปต่อยอดและพัฒนาต่อไป	2
บทที่ 2 ทฤษฎี	3
2.1 กล่าวนำ	3
2.2 ทฤษฎีเสียง	3

หน้า

2.2.1 อวัยวะที่ใชในการเปลงเสียง 4

2.2.1.1 ปอดและกระบังลม 4

2.2.1.2 หลอดลม (Larynx) 4

2.2.1.3 กลองเสียงและเสนเสียง (Vocal Card) 4

2.2.14 ชองปากและสวนของหลอดอาหารตอนตน 4

2.2.1.5 โพรงจมูก เริ่มจากเพดานออนจนถึงรูจมูกทั้งสอง 4

2.3 เสียงพยัญชนะในภาษาไทย 5

2.3.1 ลักษณะของเสียงพยัญชนะ 5

2.3.2 ประเภทของเสียงพยัญชนะ 5

2.4 ไมโครโฟน (microphone) 6

2.5 การประมวลผลสัญญาณเสียง (audio signal processing) 7

2.5.1 สัญญาณต่อเนื่อง (continuous-time signal) 7

2.5.2 สัญญาณไม่ต่อเนื่อง (discrete-time signal) 8

2.6 การสกัดลักษณะเด่น (feature extraction) 8

2.6.1 นอร์มัลไลเซชัน (Normalization) 8

2.6.2 การลดทอนสัญญาณรบกวน (noise reduction) 9

2.6.3 พรีเอมฟาซิส (Pre-emphasis) 9

2.6.4 การแบ่งเฟรม (Framing) 9

2.6.5 แฮมมิ่งวินโดว์ (Hamming window) 9

2.6.6 การวิเคราะห์สัญญาณเชิงเวลา-ความถี่ (time-frequency analysis) 10

2.6.7 เมลฟิลเตอร์แบงค์ (Mel filter bank) 11

2.6.8 การแปลงโคซายน์แบบไม่ต่อเนื่อง (Discrete Cosine Transform: DCT) 13

2.6.9 การหาสัมประสิทธิ์พลังงานและเดลต้าเซปสตรัม 13

2.7 โครงข่ายประสาทเทียม 15

2.7.1 เซลล์ประสาท (neuron) 15

2.7.2 โครงข่ายประสาทเทียมกับเซลล์ประสาทของมนุษย์ 16

2.8 เซลล์ประสาทเทียม 17

2.8.1 ลักษณะของโครงขายประสาทเทียม 18

2.8.1.1 โครงขายแบบชั้นเดียว 18

2.8.1.2 โครงขายแบบหลายชั้น 19

หน้า

2.9 ประเภทของการเรียนรูของโครงขายประสาทเทียม 19

2.9.1 การเรียนรูแบบมีผูสอน (supervised learning) 19

2.9.2 การเรียนรูแบบไมมีผูสอน (unsupervised learning) 20

บทที่ 3 การออกแบบและการทดลอง 21

3.1 กล่าวนำ 21

3.2 การเก็บตัวอยางสัญญาณเสียงพูด 23

3.3 ขั้นตอนดำเนินการ 24

3.3.1การเตรียมสัญญาณสำหรับใช้ในอัลกอริทึม (MFCC) 25

3.3.2 การสกัดคุณลักษณะของสัญญาณเสียงพูดโดยใช้ในอัลกอริทึม (MFCC) 25

3.3.2.1 พรีเอมฟาซีสของสัญญาณเสียงพูด 25

3.3.2.2 การแบ่งเฟรมเสียงพูด 25

3.3.2.3 การทำแฮมมิ่งวินโดว์ 25

3.3.2.4 การแปลงฟูเรียร์แบบไม่ต่อเนื่อง 26

3.3.2.5 การทำเมลฟิลเตอร์แบงค์ 26

3.3.2.6 การแปลงโคซายน์แบบไม่ต่อเนื่อง 28

3.3.2.7 การหาสัมประสิทธิ์พลังงานและเดลต้าเซปสตรัม 28

บทที่ 4 ผลการทดลอง 29

4.1 การสกัดคุณลักษณะของสัญญาณเสียงพูดโดยใช้ในอัลกอริทึม MFCC 29

4.1.1 การเตรียมสัญญาณเสียง 29

4.1.2 พรีเอมฟาซีสของสัญญาณเสียงพูด 30

4.1.3 การแบ่งเฟรมเสียงพูด 31

4.1.4 การแบ่งเฟรมเสียงพูด 31

4.1.5 การแปลงฟูเรียร์แบบไม่ต่อเนื่อง 32

4.1.6 การทำเมลฟิลเตอร์แบงค์ 33

4.1.7 การแปลงโคซายน์แบบไม่ต่อเนื่อง 34

4.2.8 การหาสัมประสิทธิ์พลังงานและเดลต้าเซปสตรัม 34

4.3 การทดลองการรู้จำเสียง 36

บทที่ 5 สรุปและวิเคราะห์ผลการทดลอง 46

5.1 สรุปผลการทดลอง 46

5.2 ปัญหาและอุปสรรค 46

หน้า

5.3 ข้อเสนอแนะ 47

เอกสารอ้างอิง 48

สารบัญรูป

	หน้า
รูปที่ 2.1 อวัยวะที่ช่วยในการออกเสียง	3
รูปที่ 2.2 ลักษณะของกลองเสียงขณะออกเสียงและไมออกเสียง	4
รูปที่ 2.3 ตัวอย่างสัญญาณต่อเนื่องและสัญญาณไม่ต่อเนื่อง	7
รูปที่ 2.4 แฮมมิ่งวินโดว์	10
รูปที่ 2.5 ชุดตัวกรองแบบฟิลเตอร์แบงค์	12
รูปที่ 2.6 ส่วนต่างๆ ของระบบเซลล์ประสาท	16
รูปที่ 2.7 โครงข่ายประสาทเทียมเทียบกับเซลล์ประสาทมนุษย์	16
รูปที่ 2.8 โครงสร้างการทำงานของโครงข่ายประสาทเทียม	17
รูปที่ 2.9 โครงสร้างโครงข่ายประสาทเทียมแบบชั้นเดียว	18
รูปที่ 2.10 โครงข่ายประสาทเทียมแบบหลายชั้น	19
รูปที่ 3.1 บล็อกไดอะแกรมของระบบ	24
รูปที่ 3.2 ตัวกรองสามเหลี่ยมในความถี่อนาล็อก	27
รูปที่ 3.2 ตัวกรองสามเหลี่ยมในความถี่ดิจิตอล	27
รูปที่ 4.1 การตัดส่วนของสัญญาณที่ไม่ใช่เสียงพูด	29
รูปที่ 4.2 กราฟแสดงก่อนและหลังทำกระบวนการพรีเอมฟาซีส	30
รูปที่ 4.3 เฟรมย่อยของสัญญาณเสียง	31
รูปที่ 4. 4 สัญญาณที่ผ่านการทำแฮมมิ่งวินโดว์	32
รูปที่ 4.5 เฟรมย่อยของสัญญาณเสียงในโดเมนเวลา(ซ้าย) และโดเมนความถี่(ขวา)	32
รูปที่ 4.6 Validation and test data	33
รูปที่ 4.7 เมลฟิลเตอร์แบงค์และสเปคตรัมในแต่ละตัวกรองทั้งหมด	33
รูปที่ 4.8 ค่าสัมประสิทธิ์เซปสตรัม	34
รูปที่ 4.9 26 MFCC	35

สารบัญตาราง

หน้า

ตารางที่ 3.1 เสียงคำพูดพยัญชนะภาษาไทย 3

ตารางที่ 4.1 ค่าลักษณะสำคัญที่ใช้ในการทดลอง 4.3.1 36

ตารางที่ 4.2 อัตราการรู้จำเสียง โดยใช้ค่าสัมประสิทธิ์ MFCC จำนวน 12 ตัว 36

ตารางที่ 4.3 ค่าลักษณะสำคัญที่ใช้ในการทดลอง 4.3.2 38

ตารางที่ 4.4 อัตราการรู้จำเสียง โดยใช้ค่าสัมประสิทธิ์ MFCC และ delta MFCC 38

ตารางที่ 4.5 ค่าลักษณะสำคัญที่ใช้ในการทดลอง 4.3.3 40

ตารางที่ 4.6 อัตราการรู้จำเสียง โดยใช้ค่าสัมประสิทธิ์ MFCC , delta MFCC และ energy 40

ตารางที่ 4.7 ค่าลักษณะสำคัญที่ใช้ในการทดลอง 4.3.4 42

ตารางที่ 4.8 อัตราการรู้จำเสียง โดยใช้ค่าสัมประสิทธิ์ MFCC,delta MFCC,energy และ delta-delta MFCC 42

ตารางที่ 4.9 ตารางแสดงผลการทดลองการวิเคราะห์เสียงผู้ชายแบบหลายบุคคล 45

ตารางที่ 4-10 ตารางแสดงผลการทดลองการวิเคราะห์เสียงผู้หญิงแบบหลายบุคคล 45

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญ

ปัจจุบันนี้การทำให้คอมพิวเตอร์เข้าใจเสียงพูดของมนุษย์มีประโยชน์หลายประการ รวมทั้งช่วยให้ผู้พิการทางสายตาและผู้พิการทางร่างกายสั่งงานได้ง่ายและสะดวกขึ้น ถ้ามีระบบรู้จำเสียงก็จะทำให้คนพิการสั่งงานด้วยเสียงแทนได้ ทำให้เขาเหล่านั้นมีความสามารถทัดเทียมคนปกติ เป็นการเพิ่มโอกาสในการมีอาชีพทำให้มีคุณภาพชีวิตที่ดีขึ้น

ระบบการจำลองพยัญชนะภาษาไทยด้วยเสียง เป็นโครงงานที่สร้างขึ้นโดยใช้เทคโนโลยีในปัจจุบันเข้ามาช่วยในการรู้จำเสียงพูดของมนุษย์เพื่อนำไปอำนวยความสะดวกแก่มนุษย์ในอนาคต และยังมีประโยชน์อย่างมากต่อผู้พิการทางร่างกายอีกด้วย โดยยังสามารถนำไปประยุกต์ใช้และพัฒนาได้ในอนาคต

1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา

1. สร้างระบบรู้จำหน่วยเสียงพยัญชนะภาษาไทย โดยใช้โครงข่ายประสาทเทียม
2. เพื่อศึกษาข้อมูลเกี่ยวกับโปรแกรมที่ใช้ประมวลผลด้านเสียงโดยโปรแกรม MATLAB
3. เพื่อนำไปเป็นตัวอย่างหรือต่อยอดในโครงงานอื่นต่อไป

1.3 สมมุติฐานของการศึกษา

1. เข้าใจขั้นตอนและวิธีการจำลองพยัญชนะภาษาไทยด้วยเสียงผ่านคอมพิวเตอร์
2. สามารถใช้โปรแกรม MATLAB ในการวิเคราะห์ข้อมูลเสียง

1.4 ขอบเขตการวิจัย

1. โปรแกรม MATLAB ที่ออกแบบสามารถวิเคราะห์เสียงพูดพยัญชนะภาษาไทยได้อย่างถูกต้อง

2. เข้าใจถึงขั้นตอนและกระบวนการที่นำมาใช้วิเคราะห์

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1. สามารถใช้โปรแกรม MATLAB ในการวิเคราะห์เสียงได้
2. เข้าใจถึงกระบวนการวิเคราะห์เสียงพูดโดยใช้โปรแกรม MATLAB
3. ทราบถึงแนวทางที่เหมาะสมในการสร้างระบบรู้จำหน่วยเสียงพยัญชนะภาษาไทย
4. สามารถนำความรู้ที่ได้ไปต่อยอดและพัฒนาต่อไป

บทที่ 2

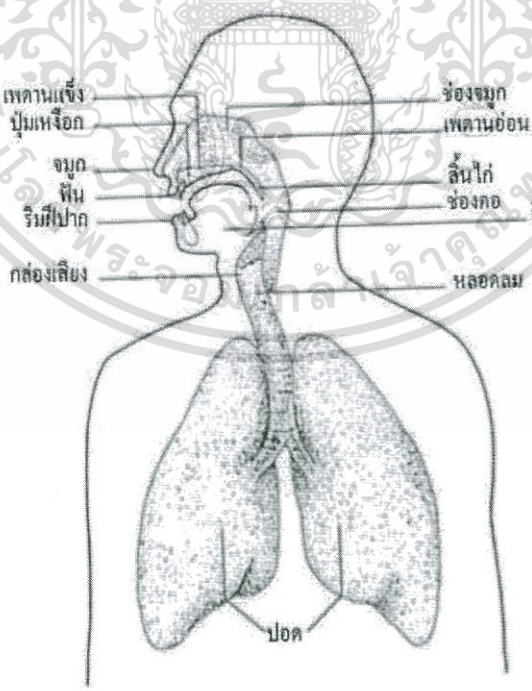
ทฤษฎีและหลักการที่เกี่ยวข้อง

2.1 กล่าวนำ

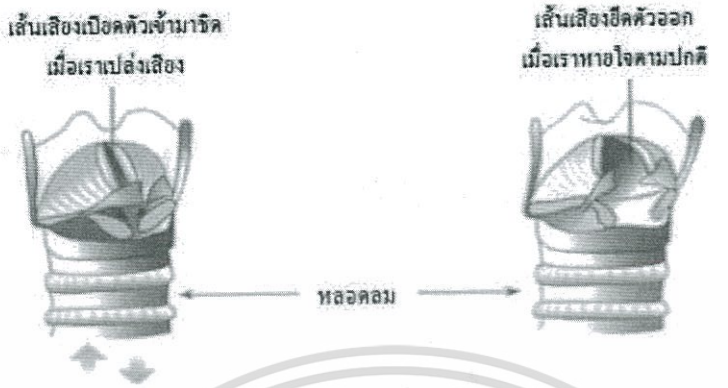
ในบทที่ 2 นี้ อธิบายถึงทฤษฎีที่เกี่ยวข้องกับการใช้อุปกรณ์บันทึกข้อมูล การเตรียมข้อมูล สู่กระบวนการประมวลผลสัญญาณเสียง ซึ่งเป็นวิธีการที่วิทยานิพนธ์นี้ ใช้ในการจำลองพยัญชนะด้วยเสียงและได้นำเสนอผลงานวิจัยที่เกี่ยวข้องกับระบบ ในการดำเนินงานการศึกษาวิจัยทำผู้วิจัยได้แบ่งหลักการและทฤษฎีต่างๆ ที่เกี่ยวข้องกับวิทยานิพนธ์ออกเป็นกลุ่มๆ ดังต่อไปนี้

2.2 ทฤษฎีเสียง

เสียงพูดเกิดจากการที่อวัยวะหลายส่วนในร่างกายของเราทำงาน ประสานกันอวัยวะเหล่านี้จะเคลื่อนไหวตามหน้าที่ และทำให้มนุษย์สามารถเปล่งเสียงออกมาเป็นภาษาได้[5] อวัยวะที่ช่วยในการออกเสียงดังแสดงในรูปที่ 2.1 และแสดงลักษณะของกลองเสียงขณะออกเสียงและไม่ออกเสียงดังรูปที่ 2.2



รูปที่ 2.1 อวัยวะที่ช่วยในการออกเสียง



รูปที่ 2.2 ลักษณะของกลองเสียงขณะออกเสียงและไมออกเสียง

2.2.1 อวัยวะที่ใช้ในการเปล่งเสียง

การเปล่งเสียงของมนุษย์ต้องอาศัยการทำงานของอวัยวะเหล่านี้ [6] คือ

2.2.1.1 ปอดและกระบังลม ทำหน้าที่สำคัญในการหายใจ และเป็นต้นกำเนิดการไหลของอากาศในกระบวนการผลิตเสียง

2.2.1.2 หลอดลม (Larynx) ทำหน้าที่นำอากาศจากปอดผ่านกล่องเสียง และเป็นอวัยวะที่อยู่ด้านหน้าของหลอดอาหาร

2.2.1.3 กล่องเสียงและเส้นเสียง (Vocal Card) มีหน้าที่หลักในการปิดกั้นไม่ให้อาหารพลัดลงไปในหลอดลม ในการเปล่งเสียง เส้นเสียงมีหน้าที่เปลี่ยนลมจากปอดให้เป็นคลื่นเสียง เส้นเสียงทำให้เกิดข้อแตกต่างระหว่างเสียงประเภทต่างๆ

2.2.14 ช่องปากและส่วนของหลอดอาหารตอนต้น อวัยวะกลุ่มนี้อยู่ต่อจากกล่องเสียง อาจเรียกว่าอวัยวะกำทอนเสียง (Vocal Tract) ทำหน้าที่กำทอนเสียง โดยทำให้กำทอนทั้งเสียงที่เกิดจากกล่องเสียงและเสียงที่เกิดภายในช่องปาก

2.2.1.5 โพรงจมูก เริ่มจากเพดานอ่อนจนถึงรูจมูกทั้งสอง ทำหน้าที่กำทอนเสียงร่วมกับช่องปากเมื่อมีการเปล่งเสียงที่ออกทางจมูก (Nasal Sounds) เช่นเสียง /ม/, /น/, และ /ง/ เป็นต้น

2.3 เสียงพยัญชนะในภาษาไทย

เสียงพยัญชนะหรือเสียงเป็นเสียงที่เปล่งออกมาจากลำคอแล้วถูกสกัดกั้นจากฐานต่างๆ จนทำให้เกิดเสียงก้องและไม่ก้อง [5]

2.3.1 ลักษณะของเสียงพยัญชนะ

เสียงพยัญชนะเกิดจากลมหายใจที่ส่งมาจากปอดผ่านมาตามหลอดลม กระทบเส้นเสียงในหลอดลมแล้วผ่านมาถึงลำคอ ลมที่ออกมานี้จะถูกกักกันไว้ในส่วนต่างๆ ซึ่งมากระทบอวัยวะบางส่วนในช่องปากหรือทั้งหมด แล้วจึงปล่อยลมนั้นออกมาทางปากหรือขึ้นจมูกก็ได้ ทำให้รู้สึกว่าการออกเสียง พยัญชนะไม่สะดวกเท่ากับการออกเสียงสระ จุดที่ลมถูกกักกั้นแล้วปล่อยให้ลมออกมานั้นเป็นที่เกิดของเสียงพยัญชนะ ซึ่งมีที่เกิดหลายแห่งดังนี้

1. เกิดจากกักลมแล้วปล่อยออกมาจากลำคอ เช่น เสียง ก ค ง
2. เกิดจากการที่เอาลิ้นไปแตะที่เพดานปากแล้วปล่อยลมออกมา เช่น เสียง จ ฉ ย
3. เกิดจากการเอาลิ้นไปแตะที่ปุ่มเหงือแล้วปล่อยลมออกมา เช่น เสียง ฏ ฑ ฒ ณ
4. เกิดจากการเอาลิ้นไปแตะที่ฟันแล้วปล่อยลมออกมา เช่น เสียง ต ถ ท น
5. เกิดจากการกักลมที่ริมฝีปาก แล้วปล่อยเสียงออกมา เช่น เสียง บ ป พ ฟ ม
6. เกิดในที่ต่างๆ เช่น เสียง ร ล เกิดที่ลิ้น เสียง ว เกิดจาการห่อริมฝีปาก เสียง อ ห เกิดจาก การปล่อยลมออกมาจากลำคอโดยตรงโดยไม่กักลมไว

2.3.2 ประเภทของเสียงพยัญชนะ

2.2.2.1 การออกเสียงพยัญชนะแบ่งเป็น 4 ลักษณะดังนี้

2.3.2.1.1 พยัญชนะเสียงก้องหรือโฆษะ คือ เสียงที่เกิดจากการที่ลมถูกดันออกมากระทบกับเส้นเสียงอย่างแรง ทำให้เส้นสียงสั่นและเกิดเป็นพัลส์(Pulse)ของอากาศไปกระตุ้นอวัยวะกำทอนเกิดเป็นเสียงก้อง เช่น เสียง บ ด อ เป็นต้น

2.3.2.1.2 พยัญชนะเสียงไม่ก้องหรืออโฆษะ คือ เสียงที่เกิดจากการที่ลมถูกดันออกมาขณะที่เส่นเสียงอยู่ในลักษณะเปิดลมพุ่งออกมาโดยสะดวกไม่สั่นสะเทือนแรงมากนักจะมีลักษณะเสียงไม่ก้อง เช่น เสียง ก ค ป เปนตน

2.3.2.1.3 พยัญชนะเสียงหนัก (ธนิต) คือ พยัญชนะเสียงไม่ก้อง ที่ขณะออกเสียงมีลมจำนวนหนึ่งพุ่งออกมาด้วย เช่น เสียง พ ท เป็นต้น

2.3.2.1.4 พยัญชนะเสียงเบา (สิถิล) คือ พยัญชนะที่ขณะออกเสียงไม่มีกลุ่มลมพุ่งตามมา เช่น เสียง ป ต เป็นต้น

2.3.2.2 เสียงพยัญชนะ สามารถแยกได้เป็น 6 ประเภท ดังนี้

2.3.2.2.1 พยัญชนะเสียงระเบิด คือ พยัญชนะที่เกิดจากลมถูกกักไว้ในช่องปาก แล้วให้ลมพุ่งออกมาอย่างรวดเร็ว เช่น เสียง ก ค จ เป็นต้น

2.3.2.2.2 พยัญชนะเสียงเสียดแทรกคือ พยัญชนะที่เกิดจากลมที่พุ่งออกมาแล้ว ถูกบีบตัวให้เสียดแทรกออก เชน เสียง ฟ ซ ฮ เป็นต้น

2.3.2.2.3 พยัญชนะเสียงนาสิก คือ พยัญชนะที่เกิดจากลมที่ถูกดันออกมาทาง จมูก เช่น เสียง ง ย ม เป็นต้น

2.3.2.2.4 พยัญชนะเสียงลิ้นรัว คือ เสียงพยัญชนะที่เกิดจากเสียงลมที่ถูกลิ้นส่วน หน้ากระดกขึ้นไปแตะเพดานให้กักไว้แล้วปล่อยลมให้ลิ้นสะบัดรัวเร็วเรียกว่า พยัญชนะลิ้นรัว เช่น เสียง ร

2.3.2.2.5 พยัญชนะเสียงข้างลิ้น คือ เสียงพยัญชนะที่เกิดจากเสียงลมที่ลิ้นกัก เอาไว้แล้วยกลิ้นขึ้นไปแตะเพดาน ปล่อยให้ลมออกมาทางขางลิ้น เรียกว่าพยัญชนะข้างลิ้น เช่น เสียง ล

2.3.2.2.6 พยัญชนะไอน้ำ คือ เสียงพยัญชนะที่เกิดขึ้นจากลำคอและมีไอน้ำ ประสมออกมาด้วยรียกว่า พยัญชนะไอน้ำ (อุสุม) เช่น เสียง อ ห

2.4 ไมโครโฟน (microphone)

ไมโครโฟนเป็นอุปกรณ์ช่วยในการรับเสียงและเปลี่ยนพลังงานเสียงมาเป็นพลังงาน ไฟฟ้า ด้วยแผ่นรับเสียงที่เรียกว่า ไดอะแฟรม (diaphragm) ซึ่งจะรับและถ่ายทอดแรงสั่นสะเทือน ที่มาจาก เสียงเปลี่ยนเป็นสัญญาณไฟฟ้าอ่อนๆ แล้วถึงจะส่งต่อไปยังไมโครโฟน ปรีแอมป์ (mic pre-amp) เพื่อ ขยายสัญญาณให้มีขนาดใหญ่พอที่จะส่งต่อไปยังเครื่องขยายเสียง ซึ่งเป็น หลักการเบื้องต้นของการ ทำงานของไมโครโฟน

ไมโครโฟนมีมากมายหลายชนิดด้วยกันในปัจจุบัน ทั้งแบบไม่ใช้ไฟฟ้า รวมถึงขอบข่ายความ กว้างในการรับสัญญาณเสียงของไมโครโฟน อิมพีแดนซ์ (impedance) มีส่วนในการแยกลักษณะ ของไมโครโฟน ซึ่งจะใช้ตัว z เป็นสัญลักษณ์ ไมโครโฟนจะมีอิมพีแดนซ์อยู่สองลักษณะได้แก่

1) ไมโครโฟนลักษณะที่มีอิมพีแดนซ์สูงหรือ high impedance (unbalance) หมายถึง ไมโครโฟนที่มีเอาท์พุต (output) ที่มีค่าความต้านทานสูงเกินกว่า 10 กิโลโอห์ม (ohms) และมีความ ไวต่อการรับเสียงแต่ในการผลิตนั้นจะมีต้นทุนในการผลิตที่ค่อนข้างสูง ซึ่งในการใช้ งานไม่ควรใช้สาย ยาวเกิน 5-6 เมตร เพราะจะทำ ให้มีผลต่อคุณภาพของเสียง

2) ไมโครโฟนลักษณะที่มีอิมพีแดนซ์ต่ำ หรือ low impedance (balance) หมายถึง ไมโครโฟนที่มีเอาท์พุตที่มีค่าความต้านทานต่ำกว่า 10 กิโลโอห์ม (อยู่ในช่วงประมาณ 200-300 โอห์ม) สามารถใช้สายยาวมากกว่า 5 เมตร โดยไม่มีปัญหาของการลดทอนของสัญญาณ ด้วยเหตุนี้ ระบบเครื่องเสียงส่วนใหญ่ จึงออกแบบมาใช้กับไมโครโฟนลักษณะ low impedance

2.5 การประมวลผลสัญญาณเสียง (audio signal processing)

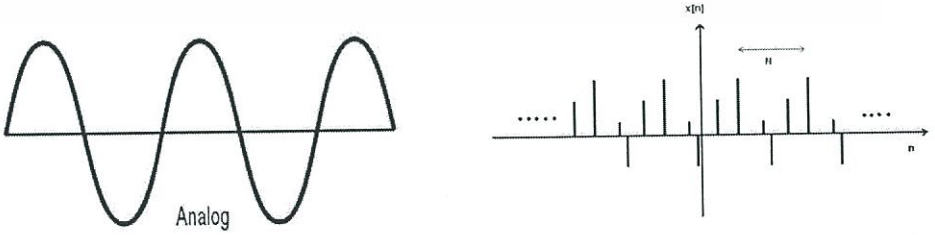
การประมวลผลสัญญาณเป็นกระบวนการปรับแต่งสัญญาณก่อนนำไปหาลักษณะเด่น เนื่องจากสัญญาณเสียงอินพุต (input) ที่บันทึกได้จากกลุ่มตัวอย่าง มีความแตกต่างในด้านระดับความดังของเสียงความถี่ของเสียง และช่วงเวลากำเนิดเสียง จึงต้องทำ การปรับแต่งสัญญาณเสียง เพื่อให้ง่ายต่อการวิเคราะห์ข้อมูล

2.5.1 สัญญาณต่อเนื่อง (continuous-time signal)

เป็นสัญญาณที่มีค่าต่อเนื่องในทางเวลา สัญญาณจะแกว่งขึ้นแกว่งลงอย่างต่อเนื่อง และราบเรียบตลอดเวลา ไม่มีการเปลี่ยนแปลงแบบทันทีทันใด เช่น สัญญาณเสียง สัญญาณภาพ คลื่นวิทยุ สัญญาณไฟบ้าน 50 Hz และอื่นๆ ถ้าแทนสัญญาณด้วยสัญลักษณ์ x และแทนเวลาด้วย สัญลักษณ์ t เราจะกล่าวว่า x เป็นฟังก์ชันของ t หรือ x มีค่าที่เวลา t ใดๆ เขียนแทนสัญญาณนี้ได้ ว่า $x(t)$ ซึ่งเป็นฟังก์ชันที่ต่อเนื่อง สัญญาณต่อเนื่องนี้เรียกอีกอย่างหนึ่งว่า สัญญาณแอนะล็อก (analog signal)

2.5.2 สัญญาณไม่ต่อเนื่อง (discrete-time signal)

เป็นสัญญาณที่มีค่าเพียงบางจุดของเวลา โดยทั่วไปเกิดจากการสุ่มสัญญาณ ต่อเนื่องด้วยคาบเวลาของการสุ่มคงที่จะใช้สัญลักษณ์ n แทนเวลาแบบไม่ต่อเนื่อง โดย n เป็นตัว แปรที่มีค่าเป็นจำนวนเต็มเท่านั้น คือ n = ..., -2, -1, 0, 1, 2, 3, ... และสัญญาณไม่ต่อเนื่องจะเป็น ฟังก์ชันของ n ดังนั้น จะเขียนแทนสัญญาณนี้ได้ว่า $x(n)$



รูปที่ 2.3 ตัวอย่างสัญญาณต่อเนื่องและสัญญาณไม่ต่อเนื่อง

2.6 การสกัดลักษณะเด่น (feature extraction)

การสกัดลักษณะเด่น เป็นการดึงลักษณะเฉพาะของหน่วยเสียงแต่ละหน่วยเสียง ที่แตกต่างกันออกมา แล้วให้ระบบทำการรู้จำลักษณะเด่นของหน่วยเสียงแต่ละหน่วยเสียงไว้ เมื่อสัญญาณที่เข้ามาภายหลัง มีลักษณะเด่นที่เหมือนหรือใกล้เคียงกับลักษณะเด่นของหน่วยเสียงใด ระบบรู้จำจะสามารถบอกได้ว่าเป็นหน่วยเสียงกลุ่มใด หรือใกล้เคียงกับหน่วยเสียงกลุ่มใดมากที่สุด และสามารถลดจำนวนข้อมูล โดยที่ข้อมูลจำนวนมากจะถูกแปลงเป็นชุดข้อมูลที่มีจำนวนน้อยลง และยังคงคุณสมบัติสำคัญของข้อมูลเดิมไว้ได้อย่างถูกต้อง

2.6.1 นอร์มัลไลเซชัน (Normalization)

เป็นทฤษฎีที่ผู้ออกแบบฐานข้อมูลจะต้องนำมาใช้ในการแปลงข้อมูลที่อยู่ในรูปแบบซับซ้อนให้อยู่ในรูปแบบที่ง่ายต่อการนำใช้งานและก่อให้เกิดปัญหาน้อยที่สุด ซึ่งเราได้ทำการการปรับความกว้างของสัญญาณคือการเปลี่ยนความกว้างให้เป็นไปตามเกณฑ์ที่กำหนด ลักษณะหนึ่งของการนอร์มัลไลเซชันคือการเปลี่ยนแอมปลิจูดเพื่อให้ขนาดสูงสุดของสัญญาณเท่ากับระดับที่กำหนดโดยกระบวนการทาง Matlab ความกว้างของสัญญาณเสียงจะมีช่วงระหว่าง -1 ถึง +1 ดังนั้นขนาดสูงสุด (ความแตกต่างจาก 0) สัญญาณสามารถทำได้คือ 1 ค่าสูงสุดนี้สามารถใช้เป็นระดับอ้างอิง เรียกว่า full scale (FS) เป็นระดับอ้างอิงจะมีค่าเดซิเบล

การสกัดคุณลักษณะของสัญญาณเสียงด้วยวิธีการหาค่าสัมประสิทธิ์เซปสตรัมบน เมลสเกล (Mel Frequency Cepstral Coefficient: MFCC)

การดึงลักษณะเด่นของสัญญาณเสียงด้วยวิธีการหาค่าสัมประสิทธิ์เซปสตรัมบนเมลสเกล เป็นวิธีการสกัดคุณลักษณะที่ใช้มากที่สุดในการรู้จำเสียงพูดอัตโนมัติ (Automatic Speech Recognition : ASR) เพื่อดึงเวกเตอร์คุณลักษณะที่มีข้อมูลทั้งหมดที่เกี่ยวกับข้อความทางภาษา MFCC จำลองการรับรู้แบบลอการิทึมของเสียงดังและระดับเสียงของระบบการฟังของมนุษย์และพยายามจะกำจัดคุณลักษณะที่เป็นผลมาจากผู้พูดยกเว้นความถี่มูลฐานและฮาร์มอนิค เพื่อแสดงลักษณะธรรมชาติของคำพูดแล้ว MFCC ยังรวมถึงการเปลี่ยนแปลงเมื่อเวลาผ่านไปของคุณลักษณะ ของเสียงอีกด้วย โดยขั้นตอนของการทำ MFCC นั้นประกอบด้วยหลักการต่างๆ ดังนี้

2.6.2 การลดทอนสัญญาณรบกวน (noise reduction)

การลดทอนสัญญาณรบกวนคือกระบวนการการปรับแต่งความถี่ของสัญญาณ ให้มีลักษณะตามที่ต้องการโดยต้องการให้มีเฉพาะ ความถี่ต่ำ ความถี่สูง ช่วงความถี่บางช่วง หรือต้องการให้บางช่วงความถี่ไม่สามารถผ่านไปแสดงที่เอาท์พุตได้ ซึ่งสามารถทำได้ดังนี้

2.6.3 พรีเอมฟาซิส (Pre-emphasis)

พรีเอมฟาซิสคือการเพิ่มปริมาณของพลังงานที่ความถี่สูงของสัญญาณเสียง ซึ่ง โดยทั่วไปสัญญาณเสียงมีปริมาณพลังงานที่ความถี่ต่ำมากกว่าความถี่สูง จึงใช้ตัวกรองความถี่สูงผ่าน อันดับที่หนึ่ง แสดงความสัมพันธ์ระหว่างข้อมูลอินพุต X[n] และข้อมูลขาออก Y[n] โดยพิจารณาค่า 0.95 นั้นว่าที่ตัวอย่างใดๆ มาจากตัวอย่างก่อนหน้าเก้าสิบห้าเปอร์เซ็นต์ [1] ได้ดังสมการที่ 2.1

$$y[n] = x[n] - 0.95x[n-1] \tag{2.1}$$

2.6.4 การแบ่งเฟรม (Framing)

เนื่องจากสัญญาณเสียงที่ได้ในแต่ละครั้งจากการบันทึก จะมีค่าตรงแกนกลางที่สูง หรือต่ำกว่าศูนย์ ทำให้การวิเคราะห์ข้อมูลเป็นไปได้ยาก เพื่อให้ง่ายต่อการวิเคราะห์และประมวลผล สัญญาณ จึงต้องทำการปรับสัญญาณที่นอกแกนศูนย์กลับเข้าสู่แกนศูนย์

การแบ่งเฟรมเป็นการเตรียมสัญญาณก่อนทำการแฮมมิ่งวินโดว์โดยแบ่งสัญญาณเสียง ให้มีขนาดสั้นลง สัญญาณเสียงจะถูกแบ่งออกเป็นเฟรมขนาด N จุดข้อมูล โดยเฟรมถัดไปจะมีข้อมูล เหลื่อมทับกันจำนวน M จุดข้อมูลกับเฟรมก่อนหน้า [2]

2.6.5 แฮมมิ่งวินโดว์ (Hamming window)

แฮมมิ่งวินโดว์ถูกนำไปคูณกับสัญญาณเสียงผ่านการแบ่งเฟรมแล้ว เพื่อให้แต่ละเฟรม ย่อยนั้นมีลักษณะเป็นสัญญาณที่ต่อเนื่องในจุดเริ่มต้นของเฟรมและจุดปลายของเฟรม โดยเมื่อนำ เฟรมสัญญาณเสียงนี้ไปแปลงฟูเรียร์จะทำให้สเปคตรัมความถี่มูลฐานและความถี่ฮาร์มอนิกของ สัญญาณเสียงแสดงออกมาในลักษณะสัญญาณยอดแหลมจำนวนมากที่มีระยะห่างใกล้เคียงกัน ซึ่ง จะเป็นผลตอบสนองความถี่แบบที่ต้องการ สมการของแฮมมิ่งวินโดว์แสดงได้ดังสมการที่ 2.2 และ สัญญาณที่ได้จากการทำแฮมมิ่งวินโดว์แสดงได้ดังสมการที่ 2.3

$$w(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right) \tag{2.2}$$

$$y(n) = x(n)^{*}w(n) \tag{2.3}$$

โดยที่ $w(n)$ คือ ค่าสัมประสิทธิ์ของแฮมมิ่งวินโดว์

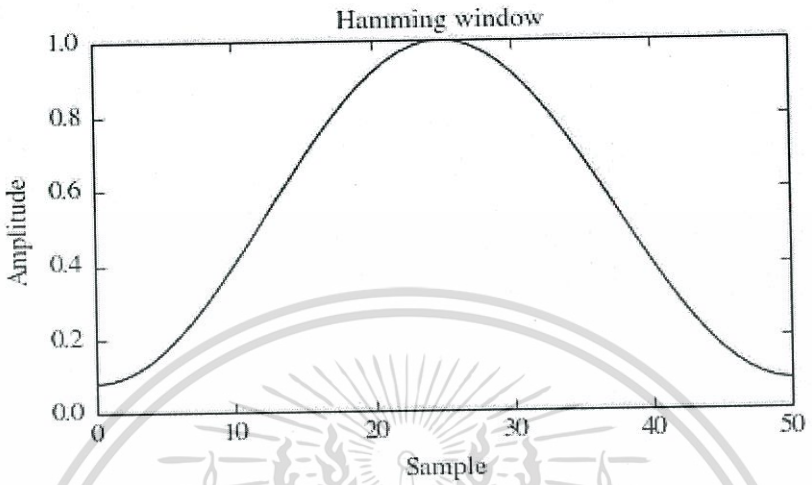
$x(n)$ คือ สัญญาณอินพุทก่อนการทำวินโดว์

$y(n)$ คือ สัญญาณเอาท์พุทที่ได้จากการทำวินโดว์

N คือ ขนาดของเฟรมสัญญาณเสียงและ

และ n มีค่าตั้งแต่ 0 ถึง N-1

โดยตัวอย่างของแฮมมิ่งวินโดว์สามารถแสดงได้ดังรูปที่ 2.4



รูปที่ 2.4 แฮมมิ่งวินโดว์

2.6.6 การวิเคราะห์สัญญาณเชิงเวลา-ความถี่ (time-frequency analysis)

ตั้งแต่ช่วงต้นศตวรรษที่ 19 เป็นต้นมาการแปลงฟูเรียร์ (fourier transform) กลายเป็นเครื่องมือ ในการวิเคราะห์สัญญาณโดยใช้กันอย่างแพร่หลายในงานด้านวิทยาศาสตร์ และวิศวกรรมศาสตร์ แนวคิดพื้นฐานของการแปลงฟูเรียร์มาจากสมมติฐานที่ว่าสัญญาณใด ๆ ก็ตาม โดยปกติแล้วจะสามารถแยกองค์ประกอบออกเป็นกลุ่มของสัญญาณรูปคลื่นไซน์หลาย ๆ ความถี่ ซึ่งเกิดจากกระบวนการโปรเจคชันสัญญาณบนกลุ่มของฟังก์ชันพื้นฐาน โดยในแต่ละ ฟังก์ชันพื้นฐานจะสร้างมาจากสัญญาณรูปคลื่นไซน์ที่มีความถี่เดียว ค่าที่ได้จากการโปรเจคชัน ที่ความถี่หนึ่ง ๆ จะเป็นตัวบ่งชี้ถึงความใกล้เคียงของสัญญาณกับฟังก์ชันพื้นฐานรูปคลื่นไซน์ ที่ความถี่นั้น แล้วนำมาจัดเรียงให้อยู่ในรูปสเปคตรัมความถี่ ดังนั้นผลจากการแปลงฟูเรียร์ ของสัญญาณใด ๆ จะแสดงถึงองค์ประกอบความถี่ทั้งหมดของสัญญาณนั้น ๆ

การแปลงฟูเรียร์ เป็นการวิเคราะห์องค์ประกอบเชิงความถี่ของสัญญาณที่นำมาใช้ประโยชน์เป็นอย่างมากสำหรับสัญญาณคงที่ (stationary signal) ในขณะที่สัญญาณส่วนใหญ่ ที่พบในโลกของความเป็นจริงนั้น ค่อนข้างซับซ้อนและมีองค์ประกอบเชิงความถี่ที่มีการเปลี่ยนแปลงตลอดเวลา ในกรณีนี้การใช้กราฟรูปคลื่นไซน์อย่างง่ายมาแทนเป็นฟังก์ชันพื้นฐานของ สัญญาณ อาจจะไม่ง่ายนัก ขณะเดียวกันการอธิบายคุณลักษณะของสัญญาณด้วยสเปกตรัมความถี่ เพียงอย่างเดียวอาจจะไม่เพียงพอ ดังนั้นการแปลงสัญญาณทั้งในเชิงเวลาและความถี่

พร้อม ๆกัน จึงถูกพัฒนาขึ้นเพื่อใช้ในการอธิบายคุณลักษณะของสัญญาณที่มีองค์ประกอบเชิงความถี่ ที่เปลี่ยนแปลงตามเวลา และเรียกกราฟที่แสดงองค์ประกอบของสัญญาณลักษณะดังกล่าวว่า สเปกโตรแกรม (spectrogram) โดยพื้นฐานแล้วการสร้างสเปกโตรแกรมจะอาศัยหลักการหาองค์ประกอบเชิงความถี่ของสัญญาณในแต่ละวินโดว์ของเวลา (time window) ที่มีการเคลื่อนที่ดังนั้นสเปกโตรแกรมจะประกอบด้วยข้อมูลขององค์ประกอบเชิงความถี่ของสัญญาณที่เวลาขณะใดขณะหนึ่งที่แตกต่างกัน สรุปคือการแปลงสัญญาณโดยใช้สมการฟูเรียร์นั้นเพื่อเปลี่ยน สัญญาณเสียงจากโดเมนของเวลา $x(n)$ ให้อยู่ในโดเมนของความถี่ ตามสมการ (2.4)

$$X(m) = \sum_{n=0}^{N-1} x(n) e^{\frac{j2\pi mn}{N}} \quad ; m = 0,1,2,3,...,N-1 \qquad (2.4)$$

โดยที่ x(n) คือสัญญาอินพุท

X(m) คือสัญญาณสเปคตรัมของสัญญาณ x(n)

N คือขนาดของเฟรมเสียงที่ผ่านการทำวินโดว์

และ m มีค่าตั้งแต่ 0 ถึง N-1

ซึ่งอาจเรียก X(m) ว่าเป็น DFT ขนาด N จุด (N-point DFT) ของ x(n) เพราะจำนวนจุดที่ใช้ในการคำนวณ DFT นั้นมีผลกับการใช้งาน DFT แต่ปัญหาของการใช้การแปลงฟูเรียร์แบบไม่ต่อเนื่อง จาก สมการที่ 2.4 จะเห็นได้ว่าต้องคูณจำนวนเชิงซ้อนเป็นสัดส่วนโดยตรงกับ N^2 ซึ่งถ้าจำนวนเฟรม ข้อมูล N มีค่ามากจะต้องใช้เวลาในการคำนวณมากขึ้นไปด้วย ดังนั้นในทางปฏิบัติจึงใช้การแปลงฟูเรียร์แบบเร็ว (Fast Fourier Transform : FFT) ซึ่งจะช่วยทำให้การแปลงฟูเรียร์มีประสิทธิภาพ และรวดเร็วมากยิ่งขึ้น

2.6.7 เมลฟิลเตอร์แบงค์ (Mel filter bank)

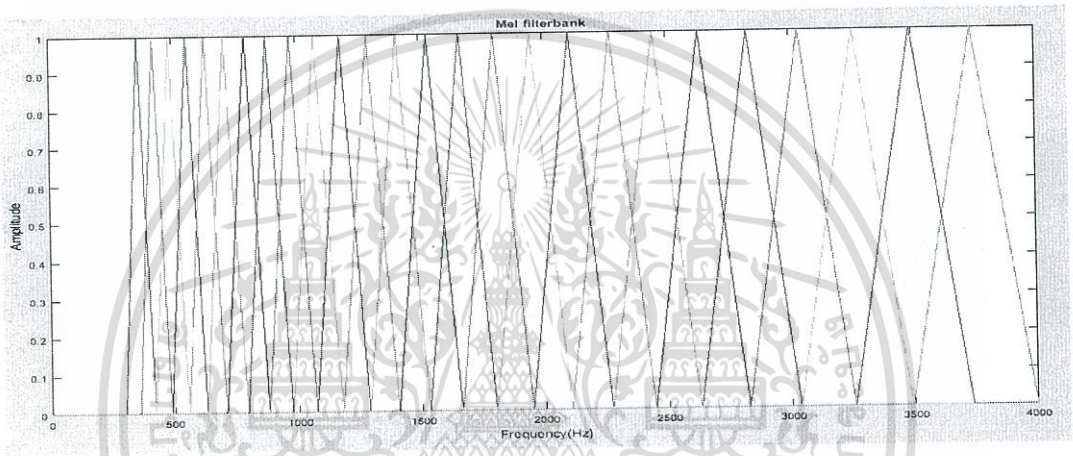
การได้ยินของมนุษย์มีความไวในการได้ยินที่ไม่เท่ากันในแต่ละช่วงความถี่ โดยมีความไวในการได้ยินต่ำในช่วงความถี่สูงที่เหนือกว่า 1000 เฮิรตซ์ ซึ่งในการสร้างแบบจำลองการได้ยินของมนุษย์ระหว่างการสกัดคุณลักษณะนั้นช่วยเพิ่มประสิทธิภาพในการรู้จำเสียง [3]

ในที่นี้ได้อาศัยหลักการของ MFCC โดยนำความถี่ที่ได้จากการทำ DFT เปลี่ยนให้เป็นความถี่บนเมลสเกล การทำเมลสเกลจะเป็นหน่วยการวัดที่เหมาะสมกับระดับการได้ยินของเสียง การแปลงระหว่างความถี่ในหน่วยเฮิรตซ์เป็นความถี่บนเมลสเกลจะให้ค่าความเป็นเชิงเส้นที่ความถี่ต่ำกว่า 1000 เฮิรตซ์และเป็นลอการึทึ่มที่ความถี่สูงกว่า 1000 เฮิรตซ์ขึ้นไป ความถี่เมลสามารถคำนวณได้จากสมการที่ 2.5 และสามารถคำนวณความถี่ในหน่วยเฮิรตซ์จากความถี่บนเมลสเกลได้จากสมการที่ 2.6

$$\text{Mel (f)} = 1125\ln(1 + \frac{f}{700}) \quad (2.5)$$

$$\text{f(m)} = M^{-1}\text{ (m)} = 700(\exp(\text{m}/1125) - 1 \quad (2.6)$$

โดยที่ M(f) คือ ค่าความถี่บนเมลสเกล
f คือ ความถี่ในหน่วยเฮิรตซ์
f(m) คือ ความถี่ในหน่วยเฮิรตซ์
m คือ ค่าความถี่บนเมลสเกล



รูปที่ 2.5 ชุดตัวกรองแบบฟิลเตอร์แบงค์

สำหรับการคำนวณ MFCC นั้น เมลฟิลเตอร์แบงค์จะถูกสร้างจากค่าความถี่เมลที่คำนวณได้ (การคำนวณฟิลเตอร์แบงค์จะกล่าวถึงในบทที่ 3) ความกว้างของฟิลเตอร์แบงค์มีความ เป็นเชิงเส้นที่ความถี่ต่ำกว่า 1000 เฮิรตซ์และเป็นลอการิทึ่มที่ความถี่สูงกว่า 1000 เฮิรตซ์ซึ่งสมการในการคำนวณฟิลเตอร์แบงค์แสดงได้ดังสมการที่ 2.7

$$H_m(k) = \begin{cases} 0 & k < f(m-1) \\ \dfrac{k - f(m-1)}{f(m) - f(m-1)} & f(m-1) \le k \le f(m) \\ \dfrac{f(m+1) - k}{f(m+1) - f(m)} & f(m) \le k \le f(m+1) \\ 0 & k > f(m+1) \end{cases} \quad (2.7)$$

2.6.8 การแปลงโคซายน์แบบไม่ต่อเนื่อง (Discrete Cosine Transform: DCT)

การแปลงโคซายน์แบบไม่ต่อเนื่องใช้สำหรับการแปลงจากค่าพลังงานลอการิทึ่มเมลสเปคตรัมที่อยู่ในโดเมนความถี่มาเป็นโดเมนเวลา [1,2] และเนื่องจากฟิลเตอร์แบงค์ที่มีการซ้อนทับ กันค่าพลังงานที่ได้นั้นก็จะมีความสัมพันธ์กันอยู่ด้วย จึงใช้การแปลงโคซายน์แบบไม่ต่อเนื่องในการ ลดความสัมพันธ์และความซับซ้อนของพลังงานลงเพื่อการจำลองรูปแบบของคุณลักษณะได้ [3]

ซึ่ง ค่าสัมประสิทธิที่ได้คือค่าสัมประสิทธิ์เซปสตรัม การแปลงโคซายน์แบบไม่ต่อเนื่องแสดงได้ดังสมการที่ 2.8

$$C(k) = w(k)\sum_{n=1}^{N} e(n)\cos\left(\frac{\pi}{2N}(2n-1)(k-1)\right), \quad k = 1,2,...,N$$

$$w(k) = \begin{cases} \frac{1}{\sqrt{2}}, & k = 1 \\ \sqrt{\frac{2}{N}}, & 2 \le k \le N \end{cases} \qquad (2.8)$$

โดยที่ C(k) คือ ค่าสัมประสิทธิ์เซปสตรัม

e(n) คือ ค่าพลังงานลอการิทึ่มเมลสเปคตรัม

N คือ จำนวนของค่าพลังงาน

2.6.9 การหาสัมประสิทธิ์พลังงานและเดลต้าเซปสตรัม

เนื่องจากพลังงานมีความสัมพันธ์กับอัตลักษณ์ของเสียงและเพื่อเป็นประโยชน์ในการตรวจสอบอัตลักษณ์ พลังงานในแต่ละเฟรมซึ่งคือผลรวมของค่ากำลังงานของจุดข้อมูลของสัญญาณเสียงในแต่ละเฟรมที่ผ่านกระบวนการท าวินโดว์จะถูกนำมาใช้เป็นหนึ่งในลักษณะเด่นของ เสียง โดยพลังงานในแต่ละเฟรมสามารถหาได้ดังสมการที่ 2.9

$$Energy = \sum_{n=n_1}^{n_2} x^2[n] \qquad (2.9)$$

โดยที่ x(n) คือ สัญญาณเสียงที่ผ่านการทำวินโดว์

การทำเดลต้าเซปสตรัมเป็นการหาอัตราการเปลี่ยนแปลงของค่าสัมประสิทธิ์เซปสตรัมสำหรับค่าการเปลี่ยนแปลงระหว่างเฟรม ซึ่งค่าสัมประสิทธิ์เซปตรัมของแต่ละเฟรมเป็นค่าของสัมประสิทธิ์หนึ่งชุด โดยการคำนวณหาค่าสัมประสิทธิ์เดลต้าเซปสตรัม สามารถคำนวณได้โดยใช้สมการ ดังแสดงในสมการที่ 2.10

$$d_t = \frac{\sum_{n=1}^{N} n(c_{t+n} - c_{t-n})}{2\sum_{n=1}^{N} n^2} \tag{2.10}$$

โดยที่ d_t คือ สัมประสิทธิ์เดลต้าเซปสตรัมของเฟรมที่ t

C_t คือ สัมประสิทธิ์เซปสตรัมของเฟรมที่ t

N คือ ค่าที่กำหนดให้เท่ากับ 2

และการคำนวณหาค่าเดลต้าเดลต้าเซปสตรัมก็สามารถคำนวณได้จากสมการที่ 2.10 เพียงเปลี่ยนค่าจาก C_t คือ สัมประสิทธิ์เซปสตรัมของเฟรมที่ t เป็น d_t คือ สัมประสิทธิ์เดลต้าเซปสตรัมของเฟรมที่ t แทน

ส่วนการหาค่าพลังงานของเดลต้าเซปสตรัมและเดลต้าเดลต้าเซปสตรัมสามารถคำนวณได้จากสมการที่ 2.11

$$Energy = \sum_{n=n_1}^{n_2} x^2[n] \tag{2.11}$$

โดยที่ x(n) คือ ค่าของ เดลต้าเซปสตรัม และ เดลต้าเดลต้าเซปสตรัม

2.7 โครงข่ายประสาทเทียม

โครงข่ายประสาทเทียม (Artificial Neural Network: ANN) เป็นโมเดลทางคณิตศาสตร์สำหรับประมวลผลสารสนเทศด้วยการคำนวณแบบคอนเนคชันนิสต์ (connectionist) เพื่อจำลองการทำงานของเครือข่ายประสาทในสมองมนุษย์ โครงข่ายประสาทเทียมมีวัตถุประสงค์ที่จะสร้างเครื่องมือที่มีความสามารถในการเรียนรู้การจดจำแบบรูป และการทำนายอนาคตเช่นเดียวกับความสามารถที่มีในสมองมนุษย์ [11]

2.7.1 เซลล์ประสาท (neuron)

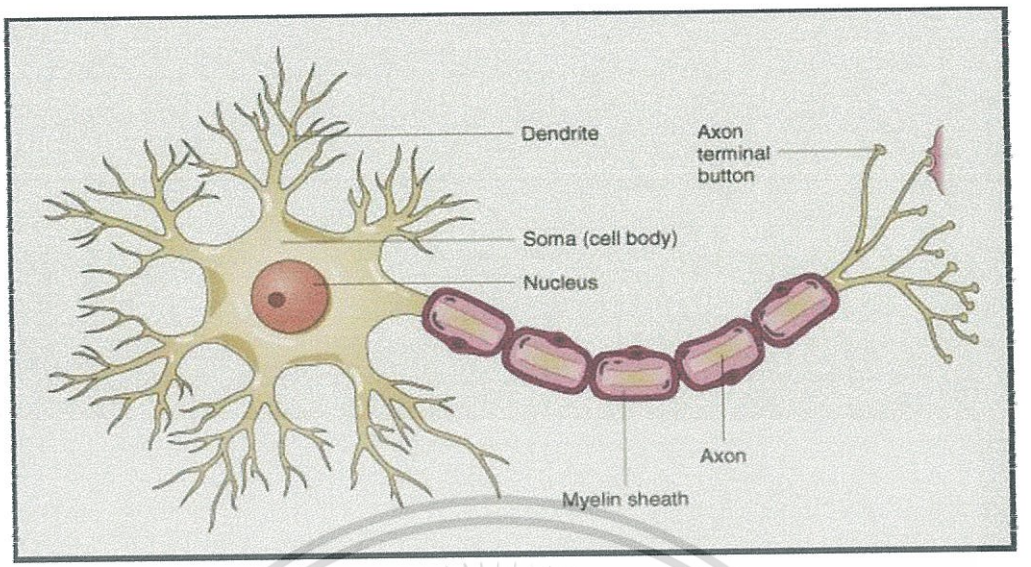
เซลล์ประสาทเป็นส่วนที่เล็กที่สุดของระบบประสาท เซลล์ประสาทหนึ่งเซลล์มีส่วนประกอบที่สำคัญอยู่ 4 ส่วน ดังแสดงในรูปที่ 2.4 [10] ดังนี้

2.7.1.1 ตัวเซลล์ (cell body) เป็นจุดศูนย์กลางของเซลล์ประสาท ประกอบด้วยนิวเคลียส (nucleus) อยู่ตรงกลางเซลล์ ล้อมรอบด้วยของเหลวที่เรียกว่าไซโตพลาส (cytoplast) มีผนังเซลล์ทำหน้าที่เป็นผนังห่อหุ้มเซลล์

2.7.1.2 เดนไดรท์ (dendrite) เป็นเส้นใยที่ยื่นออกจากตัวเซลล์มีหน้าที่รับความรู้สึกมีกิ่งก้านสาขาเป็นแขนงสั้น ๆ มีลักษณะคล้ายรากแขนงของต้นไม้

2.7.1.3 แอกซอน (axon) เป็นเส้นใยเดี่ยวๆที่ยื่นออกจากตัวเซลล์ ทำหน้าที่ส่งความรู้สึกของเซลล์ไปยังเซลล์ประสาทตัวอื่นๆ แอกซอนมีเปลือกหุ้มเรียกว่า ไมอิลินชีท (myelin sheath) ปลายสุดของแอกซอนเป็นพุ่มต่อกับอวัยวะเรียกเอนด์บลาส (end brush) ใยแอกซอนจะมีความยาวมากเป็นพิเศษ แต่ละเซลล์จะมีเพียงเส้นเดียวเท่านั้น ปลายแขนงย่อยของแอกซอน ทุกแขนงจะมีตุ่มเล็กๆ เรียกว่าตุ่มปลายประสาท (terminal buttons) การทำงานของแอกซอนจะเกิดขึ้น เมื่อตัวเซลล์ได้รับกระแสประสาทความรู้สึกจากเดนไดรท์จากนั้นจะส่งกระแสความรู้สึกนั้นไปยังแอกซอน แล้วแอกซอนจะส่งกระแสประสาทความรู้สึกนั้น ต่อไปยังเซลล์ประสาทตัวอื่นๆ หรือส่งไปยังอวัยวะต่างๆ ที่ต้องการให้เกิดความรู้สึก

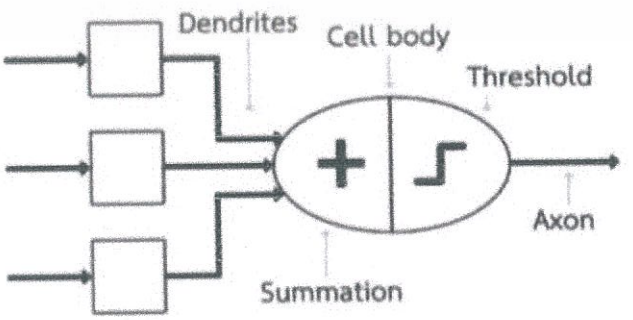
2.7.1.4 ซิแนปส์ (synapse) เป็นจุดต่อระหว่างใยแอกซอนของเซลล์ประสาทตัดหนึ่งกับเดนไดรท์ของเซลล์ประสาทอีกตัวหนึ่ง โดยที่เมื่อเซลล์ประสาทตัวหนึ่งส่งกระแสประสาทความรู้สึกเข้าสู่แอกซอนจนถึงปลายตุ่มประสาทแล้ว กระแสความรู้สึกนั้นจะถูกส่งเข้าสู่บริเวณซิแนปส์ จากนั้นซิแนปส์จะรับกระแสประสาทและส่งต่อไปยังเดนไดรท์เพื่อเข้าสู่เซลล์ประสาทอีกตัวหนึ่งทันที ซิแนปส์จึงทำหน้าที่เป็นตัวเชื่อมสัญญาณกระแสประสาทระหว่างเซลล์ประสาทตัวหนึ่งกับเซลล์ประสาทอีกตัวหนึ่ง



รูปที่ 2.6 ส่วนต่างๆ ของระบบเซลล์ประสาท

2.7.2 โครงข่ายประสาทเทียมกับเซลล์ประสาทของมนุษย์

จากการทำงานของเซลล์ประสาทของมนุษย์ ทำให้เกิดโครงข่ายประสาทเทียมที่จำลองการทำงานและหน้าที่ในส่วนต่างๆของเซลล์จนเกิดเป็นโครงข่าย ซึ่งประกอบด้วยปลายการรับกระแสประสาท เดนไดรท์ เปรียบเสมือนเป็นอินพุทของระบบและปลายการส่งกระแสประสาทเรียกว่า แอคซอน ซึ่งเปรียบเสมือนเป็นเอาท์พุทของระบบ เซลล์เหล่านี้ทำงานด้วยปฏิกิริยาไฟฟ้าเคมีเมื่อมีการกระตุ้นด้วยสิ่งเร้า กระแสประสาทจะวิ่งผ่านเดนไดรท์เข้าสู่นิวเคลียสซึ่งจะเป็นตัวตัดสินว่าต้องกระตุ้นเซลล์อื่นๆ ต่อหรือไม่ถ้ากระแสประสาทแรงพอ นิวเคลียสจะกระตุ้นเซลล์อื่นๆต่อไปผ่านทางแอคซอน ตามโมเดลนี้ข่ายงานประสาทเกิดจากการเชื่อมต่อระหว่างเซลล์ประสาทจนเป็นเครือข่ายที่ทำงานร่วมกัน ซึ่งรูปที่ 2.5 จะเป็นการอธิบายโครงข่ายประสาทเทียมเทียบกับเซลล์ประสาทของมนุษย์

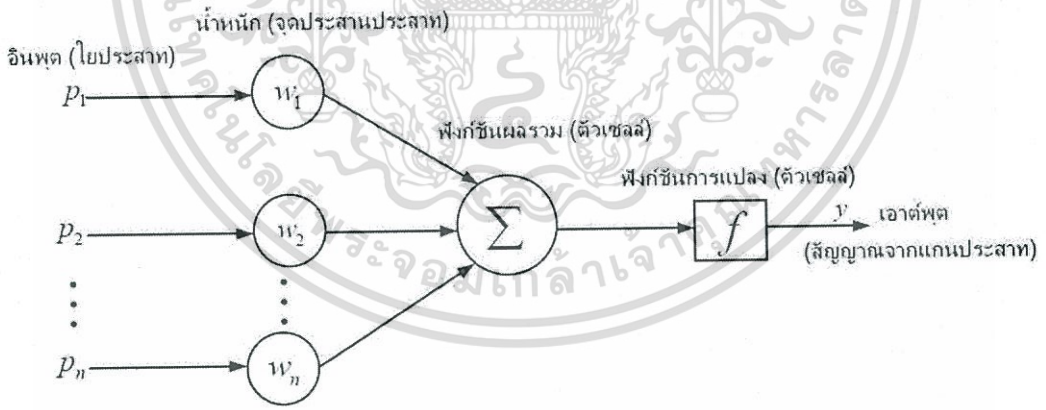


รูปที่ 2.7 โครงข่ายประสาทเทียมเทียบกับเซลล์ประสาทมนุษย์

2.8 เซลล์ประสาทเทียม

แบบจำลองของเซลล์ประสาททางชีวภาพหรือที่เรียกว่า เซลล์ประสาทเทียม (Artificial Neuron) แสดงได้ดังรูปที่ 2.6 โดยประกอบด้วยโครงสร้างพื้นฐานที่สำคัญคือ หน่วย (Node) หรือ ยูนิต (Unit), ตัวแปรด้านเข้า (Input), ตัวแปรด้านออก (Output) และค่าถ่วงน้ำหนัก (Weight) ซึ่ง สามารถสรุปความสัมพันธ์ระหว่างเซลล์ประสาททางชีววิทยาและเซลล์ประสาทเทียมดังนี้

- การประมวลผล จะเกิดขึ้นในหน่วยประมวลผลย่อยคือ หน่วย (Node) หรือ ยูนิต (Unit) ซึ่งจำลองมาจากลักษณะการทำงานของตัวเซลล์
- การส่งสัญญาณระหว่างหน่วยด้วยส่วนที่เชื่อมติดกันจำลองมาจากการเชื่อมต่อของเดนไดรต์และแอกซอน
- แต่ละการเชื่อมต่อประกอบด้วยค่าน้ำหนักที่แตกต่างกัน โดยขึ้นอยู่กับอิทธิพลที่หน่วยจะได้รับจากหน่วยอื่นๆ ซึ่งจำลองมาจากไซแนปส์ โดยค่าน้ำหนักที่ได้รับจะทำหน้าที่เสมือนความรู้ ที่ถูกรวบรวมไว้ใช้แก้ปัญหาเฉพาะอย่างของมนุษย์
- ภายในหน่วยมีฟังก์ชั่นที่ใช้ในการกำหนดสัญญาณด้านออกที่เรียกว่า ฟังก์ชั่นถ่ายโอน (Transfer Function) หรือ ฟังก์ชั่นกระตุ้น (Activation Function)



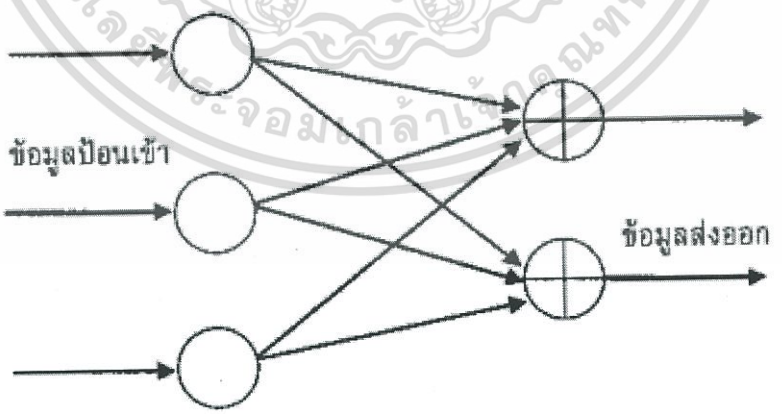
รูปที่ 2.8 โครงสร้างการทำงานของโครงข่ายประสาทเทียม

2.8.1 ลักษณะของโครงข่ายประสาทเทียม

โครงข่ายประสาทเทียมประกอบด้วยเซลล์ประสาทเทียมหรือโหนดจำนวนมากเชื่อมต่อกัน ซึ่งการเชื่อมต่อแบ่งออกเป็นกลุ่มย่อย เรียกว่า ชั้น (layer) ชั้นแรกเป็นชั้นนำข้อมูลเข้า เรียกว่า ชั้นรับข้อมูลป้อนเข้า (input layer) ส่วนชั้นสุดท้าย เรียกว่า ชั้นส่งข้อมูลออก (output layer) และชั้นที่อยู่ระหว่างชั้นรับข้อมูลป้อนเข้าและชั้นส่งข้อมูลออก เรียกว่า ชั้นแอบแฝง (hidden layer) ซึ่งโดยทั่วไปชั้นแอบแฝงอาจมีมากกว่า 1 ชั้น ก็ได้ด้วยเหตุผลนี้จึงสามารถแบ่งประเภทของโครงข่ายประสาทเทียมตามจำนวนชั้นของโครงข่ายแบบกว้างๆได้ 2 แบบ ได้แก่ โครงข่ายแบบชั้นเดียว (single layer) และโครงข่ายแบบหลายชั้น (multi-layer)

2.8.1.1 โครงข่ายแบบชั้นเดียว

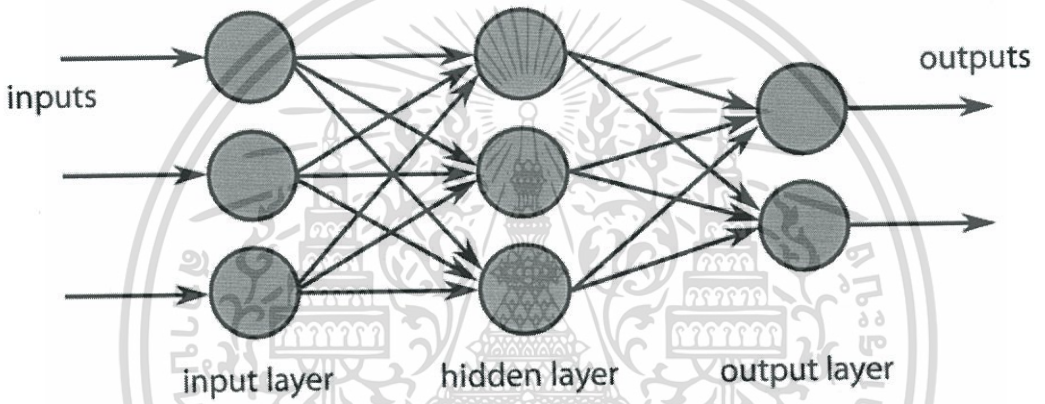
โครงข่ายแบบชั้นเดียวเป็นโครงประสาทเทียมอย่างง่ายที่มีชั้นรับข้อมูลป้อนเข้าและชั้นส่งข้อมูลออกเท่านั้น โหนดในชั้นรับข้อมูลป้อนเข้าทำหน้าที่ รับข้อมูลเข้า (input value) แล้วส่งข้อมูลผ่านเส้นเชื่อมโยงต่างๆไปให้โหนดในชั้นส่งข้อมูลออก ความเข้มของสัญญาณหรือปริมาณข้อมูลที่นำเข้าสู่โหนดในชั้นส่งข้อมูลออกจะขึ้นอยู่กับค่าน้ำหนักที่อยู่บนเส้นเชื่อมโยงโหนดในชั้นส่งข้อมูลออกจะนำข้อมูลที่ได้นับมาคำนวณโดยใช้ฟังก์ชันทางคณิตศาสตร์ที่ เรียกว่า ฟังก์ชันการแปลง (transfer function) ที่เหมาะสมกับปัญหา แล้วส่งผลลัพธ์ที่ได้ออกมา เป็นข้อมูลส่งออก เช่น โครงข่ายแบบชั้นเดียว แบบเพอเซบตรอนอย่างง่าย (simple perceptron) และโครงข่ายโฮบฟีลด์ (Hopfield networks) ลักณษะโครงข่ายแบบชั้นเดียว แสดงดังรูปที่ 2.7



รูปที่ 2.9 โครงสร้างโครงข่ายประสาทเทียมแบบชั้นเดียว

2.8.1.2 โครงข่ายแบบหลายชั้น

โครงข่ายแบบหลายชั้นเป็นโครงข่ายที่มีชั้นแอบแฝงตั้งแต่ 1 ชั้นขึ้นไป โครงข่ายแบบหลายชั้นจะใช้ในกรณีที่ปัญหามีความซับซ้อน ซึ่งโครงข่ายแบบชั้นเดียวไม่สามารถแก้ปัญหาได้จึงเพิ่มจำนวนโหนดที่มีการคำนวณหรือชั้นแอบแฝงให้กับโครงข่ายตัวอย่างของโครงข่ายแบบหลายชั้น เช่น กการแพร่ย้อนกลับ (back propagation) เซลฟ์ออร์แกนไนซิ่งแมปซ์ (self organizing maps) และเคาน์เตอร์พรอพะเกชัน (counter propagation) เป็นต้น ลักษณะโครงสร้างโครงข่ายแบบหลายชั้นแสดง ดังรูปที่ 2.7



รูปที่ 2.10 โครงข่ายประสาทเทียมแบบหลายชั้น

2.9 ประเภทของการเรียนรู้องโครงข่ายประสาทเทียม

2.9.1 การเรียนรู้แบบมีผู้สอน (supervised learning) ข้อมูลจะประกอบด้วยตัวอย่างข้อมูลที่ต้องการสอนและผลลัพธ์ที่ต้องการให้โครงข่ายสร้าง เมื่อมีการนำข้อมูลในลักษณะเดียวกันมาเป็นข้อมูลป้อนเข้า โครงข่ายจะกำหนดค่าผลลัพธ์ที่เป็นเป้าหมายให้กับข้อูลป้อนเข้าแต่ละตัว โครงข่ายจะนำค่าผิดพลาดระหว่างค่าเป้าหมายกับ ค่าผลลัพธ์ที่ได้มาใช้ในการปรับค่าน้ำหนักเพื่อ ให้ค่าผลลัพธ์ใกล้เคียงกับเป้าหมายมากที่สุดถ้าหากเปรียบเทียบกับมนุษย์จะเหมือนกับการสอนนักเรียน โดยมีครูผู้สอนคอยให้คำแนะนำตัวอย่างแบบจำลองนี้ได้แก่ การแพร่ย้อนกลับและเพอเซบตรอน (perceptron) เป็นต้น

2.9.2 การเรียนรู้แบบไม่มีผู้สอน (unsupervised learning) การเรียนรู้แบบนี้จะสอนโครงข่ายโดยการนำข้อมูลป้อนเข้าอย่างต่อเนื่องเพียงอย่างเดียว ไม่มีการส่งค่าผลลัพธ์เป้าหมาย ให้กับข้อมูลป้อนเข้าแต่ละตัว การปรับน้ำหนักจะใช้ข้อมูลที่นำมาสอนเป็นตัวปรับค่า โดยค่าน้ำหนักจะปรับตามกลุ่มที่ข้อมูลป้อนเข้าที่มีรูปแบบคล้ายคลึงกัน ถ้าหากเปรียบเทียบกับมนุษย์จะเหมือนกับการที่เราสามารถแยกแยะพันธ์สัตว์ตามลักษณะรูปร่างของมันได้ด้วยตนเอง ตัวอย่างแบบจำลองนี้ ได้แก่ เคาน์เตอร์พรอพะเกชัน (counter propagation : CPN) แบบจำลองอะแดปทีฟรีโซแนนซ์เทียรี่ (Adaptive Resonance Theory neural networks : ART) เป็นต้น

บทที่ 3

การออกแบบและการทดลอง

3.1 กล่าวนำ

ในบทที่ 3 นี้ได้กล่าวถึงวิธีดำเนินการศึกษางานวิจัย การออกแบบและการทดลองข้อมูลจากสัญญาณเสียง เพื่อการเรียนรู้ โดยอาศัยขั้นตอน และวิธีต่าง ๆ ในการจำลองเสียงพยัญชนะ ซึ่งได้แบ่งเนื้อหาออกมาตามหัวข้อต่างๆ เพื่อให้บรรลุวัตถุประสงค์ที่กำหนดไว้ ผู้วิจัยได้ดำเนินการวิจัยตามขั้นตอนดังนี้

ตารางที่ 3.1 เสียงคำพูดพยัญชนะภาษาไทย

ลำดับที่	พยัญชนะ	การออกเสียง
1	ก	กอ-ไก่
2	ข	ขอ-ไข่
3	ฃ	ฃอ-ขวด
4	ค	คอ-ควาย
5	ฅ	คอ-ฅน
6	ฆ	คอ-ระ-คัง
7	ง	งอ-งู
8	จ	จอ-จาน
9	ฉ	ฉอ-ฉิ่ง
10	ช	ชอ-ช้าง
11	ซ	ซอ-โซ่
12	ฌ	ชอ-เชอ
13	ญ	ยอ-หญิง
14	ฎ	ชอ-ชะ-ดา
15	ฏ	ตอ-ปะ-ตัก
16	ฐ	ถอ-ถาน
17	ฑ	ทอ-มน-โท
18	ฒ	ทอ-ผู้-เท่า

19	ณ	นอ-เนน
20	ด	ดอ-เด็ก
21	ต	ตอ-เต่า
22	ถ	ถอ-ถุง
23	ท	ทอ-ทะ-หาน
24	ธ	ทอ-ทง
25	น	นอ-หนู
26	บ	บอ-ใบ-ไม้
27	ป	ปอ-ปลา
28	ผ	ผอ-ผึ้ง
29	ฝ	ฝอ-ฝา
30	พ	พอ-พาน
31	ฟ	ฟอ-ฟัน
32	ภ	พอ-สำ-เพา
33	ม	มอ-ม้า
34	ย	ยอ-ยักษ์
35	ร	รอ-เรือ
36	ล	ลอ-ลิง
37	ว	วอ-แหวน
38	ศ	สอ-สา-ลา
39	ษ	สอ-รือ-สี
40	ส	สอ-เสือ
41	ห	หอ-หีบ
42	ฬ	รอ-จุ-ลา
43	อ	ออ-อ่าง
44	ฮ	ฮอ-นก-ฮูก

3.2 การเก็บตัวอย่างสัญญาณเสียงพูด

ในการเก็บตัวอย่างสัญญาณเสียงพูดในเบื้องต้นผู้วิจัย ได้ทำการเก็บตัวอย่างเสียงพูดตัวอย่าง โดยรายละเอียดการเก็บเสียงพูดตัวอย่างมีดังนี้

3.2.1 กลุ่มต้นแบบ หมายถึง กลุ่มต้นแบบสำหรับระบบรู้จำ ทำการบันทึกเสียงพูดเสียง กอ-ไก่ ถึง ฮอ-นก-ฮูก

3.2.1.1. สำหรับวิเคราะห์เสียงแบบบุคคลเดี่ยว

- เพศชายจำนวน 2 คน และ เพศหญิงจำนวน 2 คน คนละ 10 ชุดโดยแต่ละชุดประกอบด้วยคำพูดทั้งหมด 44 คำ

- เสียงคำพูดชุดที่ 1 ใช้สำหรับฝึกฝนระบบให้รู้จำ (Training Set)

- เสียงคำพูดชุดที่ 2 ใช้สำหรับการทดสอบแบบขึ้นกับผู้พูด (Testing Set) ใช้เสียงเพศชายจำนวน 2 คน และ เพศหญิงจำนวน 2 คน

3.2.1.2 สำหรับวิเคราะห์เสียงแบบหลายบุคคล

- เพศชายจำนวน 10 คน คนละ 2 ชุดโดยแต่ละชุดประกอบด้วยคำพูดทั้งหมด 44คำ

- เสียงคำพูดชุดที่ 1 ใช้สำหรับฝึกฝนระบบให้รู้จำ (Training Set)

- เสียงคำพูดชุดที่ 2 ใช้สำหรับการทดสอบแบบขึ้นกับผู้พูด (Testing Set) ใช้เสียงเพศชาย 10 คน

3.2.2 กลุ่มทดสอบ หมายถึง กลุ่มของเสียงที่ไม่ได้นำไปให้ระบบรู้จำ

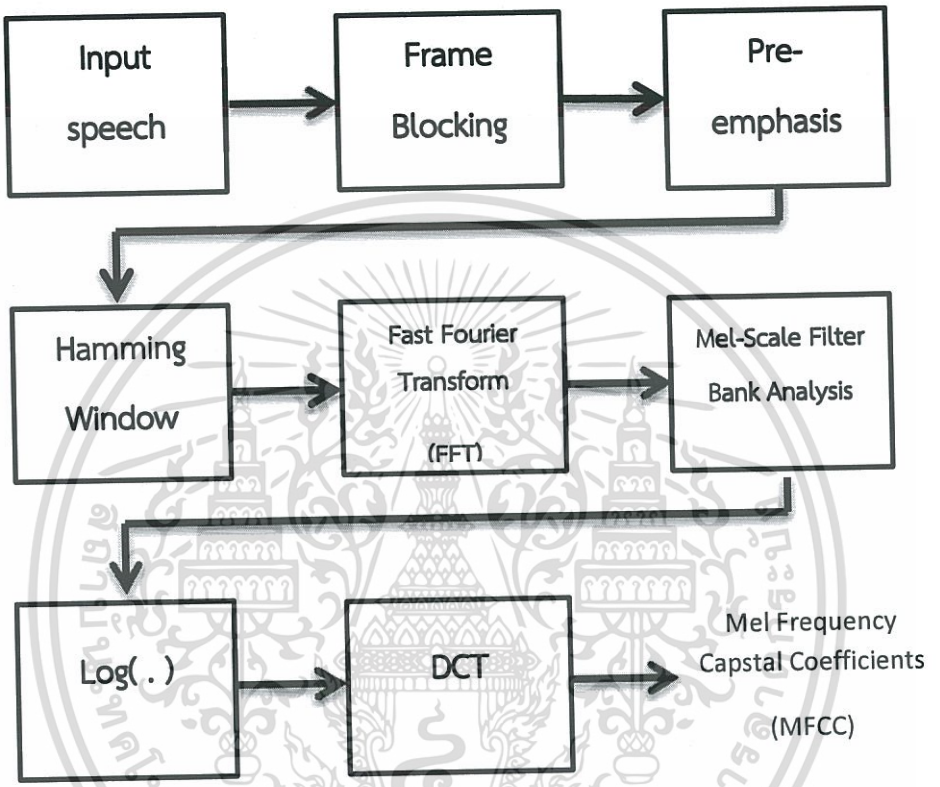
3.2.2.1 สำหรับวิเคราะห์แบบบุคคลเดี่ยว ทำการบันทึกเสียงพูด เสียง กอ-ไก่ ถึง ฮอ-นก-ฮูก คนละ 5 ชุดโดยแต่ละชุดประกอบด้วยคำพูดทั้งหมด 44 คำ เป็นเพศชายจำนวน 2 คน และ เพศหญิงจำนวน 2 คน

3.2.2.2 สำหรับวิเคราะห์แบบหลายบุคคลทำการบันทึกเสียงพูด เสียง กอ-ไก่ ถึง ฮอ-นก-ฮูก คนละ 1 ชุดโดยแต่ละชุดประกอบด้วยคำพูดทั้งหมด 44 คำ เป็นเพศชายจำนวน 10 คน

3.2.3 กลุ่มต้นแบบและกลุ่มทดสอบ อายุระหว่าง 15-23 ปี

3.3 ขั้นตอนดำเนินการ

ในงานวิจัยนี้ได้ทำการวิเคราะห์แนวทางในการออกแบบการวิจัย ซึ่งมีขั้นตอนการดำเนินงานและวิธีการศึกษาข้อมูลในกระบวนการต่างๆ ดังที่ได้แสดงในรูปที่ 3.1



รูปที่ 3.1 บล็อกไดอะแกรมของระบบ

การทำงานของระบบในส่วนแรกจะเริ่มจากการนำเสียงมาผ่านระบบ MFCC เพื่อสกัด หาคุณลักษณะเด่นออกมาเป็นค่าสัมประสิทธิ์

3.3.1 การเตรียมสัญญาณสำหรับใช้ในอัลกอริทึม Mel Frequency Cepstral Coefficient (MFCC)

การบันทึกเสียงพูดเป็นการบันทึกเสียงในโปรแกรม Matlab โดยพูดผ่าน ไมโครโฟนที่ต่อเข้าวงจรขยายไมโครโฟน (Microphone preamplifier) และทำการบันทึกเสียงพูด ยาว 2 วินาที ความถี่แซมปลิ้ง 8000 เฮิรตซ์ในโปรแกรม Matlab ซึ่งในความยาว 2 วินาทีนี้ประกอบ ไปด้วยส่วนที่เป็นเสียงพูดและส่วนที่ไม่ใช่เสียงพูด ดังนั้นจึงทำการตัดส่วนที่ไม่ใช่เสียงพูดออกด้วย วิธีการกำหนดขอบเขตที่รับได้ (Thresholding)

3.3.2 การสกัดคุณลักษณะของสัญญาณเสียงพูดโดยใช้ในอัลกอริทึม Mel Frequency Cepstral Coefficient (MFCC)

3.3.2.1 พรีเอมฟาซีสของสัญญาณเสียงพูด

เนื่องจากโดยทั่วไปสัญญาณเสียงมีปริมาณพลังงานที่ความถี่ต่ำมากกว่า ความถี่สูง เพื่อเพิ่มความชัดเจนของเสียงพูดสามารถเพิ่มระดับพลังงานที่ความถี่สูงโดยนำเสียงพูด ผ่านการทำพรีเอมฟาซีส ด้วยการใช้ตัวกรองความถี่สูงผ่านอันดับที่หนึ่ง แสดงความสัมพันธ์ระหว่าง ข้อมูลอินพุต x[n] และข้อมูลขาออก y[n] ดังแสดงในสมการที่ 2.1 [1]

3.3.2.2 การแบ่งเฟรมเสียงพูด

ขั้นตอนการแบ่งเฟรมจะนำเสียงที่ผ่านการทำฟรีเอมฟาซีสมาแบ่งเป็นเฟรม ย่อย โดยเสียงพูดที่ถูกตัดมานั้นจะมีความยาวที่ไม่เท่ากัน ซึ่งก่อนทำการแบ่งเฟรมสัญญาณเสียงจะ ผ่านการทำ zero-pad เพื่อปรับให้สัญญาณเสียงแต่ละสัญญาณมีความยาวเท่ากันที่ความยาว 8000 จุดข้อมูลหรือ 1 วินาที เนื่องจากการสัญญาณเสียงแต่ละสัญญาณที่ถูกตัดมานั้นจะมีความยาวที่ไม่ เกิน 1 วินาที เมื่อทำการ zero-pad สัญญาณเสียงเรียบร้อยแล้ว เสียงที่ได้จะถูกนำมาแบ่งเฟรม ให้เป็นเฟรมย่อยที่มีขนาด 256 จุดข้อมูลและข้อมูลของเฟรมต่อไปเลื่อนจากเฟรมก่อนหน้า 100 จุด ข้อมูลซึ่งสัญญาณเสียงพูดหนึ่งสัญญาณจะถูกแบ่งเป็นเฟรมย่อยได้ 78 เฟรม โดยขั้นตอนนี้ทำการ แบ่งเฟรมที่ 256 จุดข้อมูลเพื่อให้เหมาะสมกับขั้นตอนการแปลงฟูเรียร์แบบไม่ต่อเนื่อง [1,2]

3.3.2.3 การทำแฮมมิ่งวินโดว์

ขั้นตอนการทำแฮมมิ่งวินโดว์เป็นการนำแต่ละเฟรมย่อยของสัญญาณ เสียงพูดมาคูณกับสัมประสิทธิ์แฮมมิ่งวินโดว์ เพื่อให้แต่ละเฟรมย่อยนั้นมีลักษณะเป็นสัญญาณที่ต่อเนื่องในจุดเริ่มต้นของเฟรมและจุดปลายของเฟรม โดยแฮมมิ่งวินโดว์แสดงได้ดังสมการที่ 2.2 และสัญญาณที่ได้จากการทำแฮมมิ่งวินโดว์แสดงได้ดังสมการที่ 2.3

3.3.2.4 การแปลงฟูเรียร์แบบไม่ต่อเนื่อง

ขั้นตอนการแปลงฟูเรียร์แบบไม่ต่อเนื่องจะนำเฟรมสัญญาณที่ผ่านการทำวินโดว์มาแปลงจากโดเมนเวลาเป็นโดเมนความถี่เพื่อทำการประมาณหาค่าสเปคตรัมของสัญญาณเสียงแต่ละเฟรมเพื่อนำไปเข้ากระบวนการขั้นถัดไปคือการทำเมลฟิลเตอร์แบงค์ซึ่งเป็นการคำนวณสัญญาณได้แกนความถี่โดยการแปลงฟูเรียร์แบบไม่ต่อเนื่องแสดงได้ดังสมการที่ 2.4

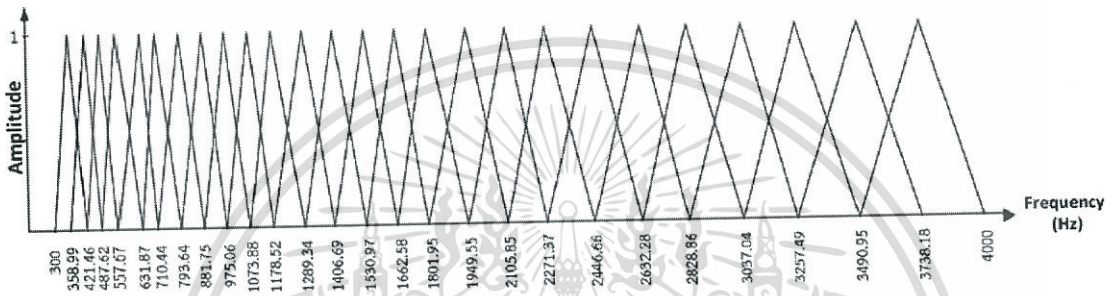
3.3.2.5 การทำเมลฟิลเตอร์แบงค์

การทำเมลฟิลเตอร์แบงค์จะเริ่มจากการออกแบบตัวกรองรูปสามเหลี่ยมซึ่งเป็นตัวกรองแบบความถี่ผ่านในที่นี้ใช้ตัวกรองสามเหลี่ยม 26 ตัวกรอง ซึ่งในการออกแบบความ กว้างย่านความถี่ของตัวกรองแต่ละตัวกรองนั้นจะถูกคำนวณในโดเมนความถี่บนเมลสเกล โดยการ คำนวณเมลฟิลเตอร์แบงค์สามารถทำได้ตามขั้นตอนต่อไปนี้ [3]

1) เลือกความถี่ล่างและความถี่บนซึ่งโดยปกติค่าความถี่ล่างจะใช้ ค่าความถี่ 300 เฮิรตซ์และจากการบันทึกเสียงโดยใช้ความถี่แซมปลิ้ง 8000 เฮิรตซ์ ทำให้ความถี่บน ถูกจำกัดที่ความถี่ 4000 เฮิรตซ์ เมื่อได้ค่าความถี่ล่างและความถี่บนแล้วก็ทำการแปลงความถี่นี้เป็น ความถี่บนเมลสเกล โดยการแปลงความถี่ในหน่วยเฮิรตซ์เป็นความถี่บนเมลสเกลทำได้ดังสมการที่ 2.5 ได้เป็นค่าความถี่ล่างและความถี่บนบนเมลสเกลคือ 401.25 และ 2142.26 ตามลำดับ ซึ่งในโครงงานนี้ใช้ตัวกรอง 26 ตัวกรอง จึงต้องคำนวณให้ได้ค่าความถี่รวมทั้งหมด 28 จุด ในช่วง ความถี่ 401.25 , 465.74 , 530.22 , 594.70 , 659.18 , 723.66 , 788.14 , 852.14 , 917.11 , 981.59 , 1046.07 , 1110.55 , 1175.04 , 1239.52 , 1304.00 , 1368.48 , 1432.96 , 1497.44 , 1561.93 , 1626.41 , 1690.89 , 1755.37 , 1819.85 , 1884.34 , 1948.82 , 2013.30 , 2077.78 และ 2141.26

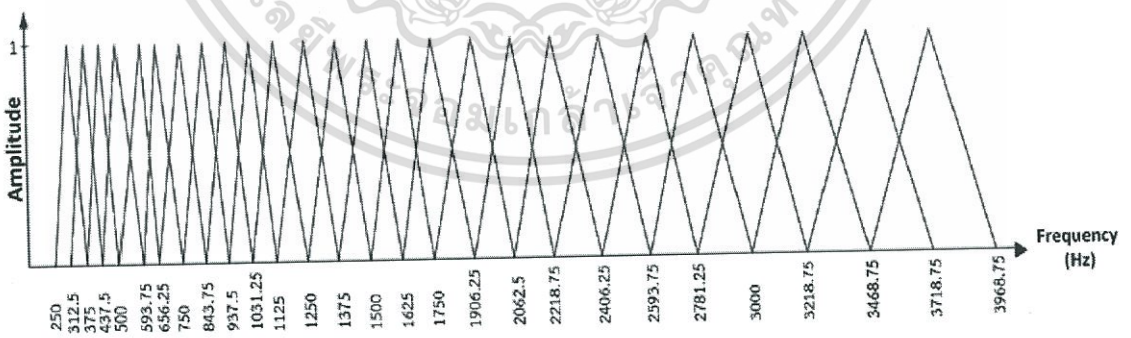
2) แปลงค่าความถี่บนเมลสเกลทั้ง 28 ค่ากลับไปเป็นความถี่ใน หน่วยเฮิรตซ์โดยสามารถคำนวณได้จากสมการที่ 2.6 ได้เป็นความถี่ในหน่วยเฮิรตซ์ คือ 300 , 358.99 , 421.46 , 487.62 , 557.67 , 631.87 , 710.44 , 793.64 , 881.75 , 975.06 , 1073.88 , 1178.52 , 1289.34 , 1406.69 , 1530.97 , 1662.58 , 1801.95 , 1949.55 , 2105.85 , 2271.37 , 2446.66 , 2632.28 , 2828.86 , 3037.04 , 3257.49 , 3490.95 , 3738.18 และ 4000 เฮิรตซ์ตามลำดับ

3) ขั้นตอนของการสร้างตัวกรองสามเหลี่ยมจากค่าความถี่ที่คำนวณได้จาก ข้อที่ 2) จะสร้างตัวกรองสามเหลี่ยมตัวแรกโดยเริ่มจากจุดค่าความถี่ที่หนึ่งโดยมีค่า เพิ่มจนมียอดสูงสุดที่จุดความถี่ที่สองและลดลงมาเป็นศูนย์ที่จุดความถี่ที่สาม ตัวกรองตัวที่สองจะ เริ่มจากจุดค่าความถี่ที่สองมีค่าขึ้นสูงสุดที่ความถี่จุดที่สามและมีค่าเป็นศูนย์ที่จุดความถี่ที่สี่ และตัว กรองลำดับถัดไปจะมีรูปแบบความสัมพันธ์แบบนี้ไปเรื่อยๆ สามารถแสดงการคำนวณนี้ได้ดังสมการ ที่ 2.7 โดยตัวกรองสามเหลี่ยมทั้งหมด 26 ตัวกรอง แสดงได้ดังรูปที่ 3.2



รูปที่ 3.2 ตัวกรองสามเหลี่ยมในความถี่อนาล็อก

แต่เนื่องจากนำมาประมวลผลบนคอมพิวเตอร์โดยใช้โปรแกรม Matlab ซึ่งความถี่ที่ใช้เป็นความถี่ดิจิตอลความถี่ที่คำนวณได้จากสมการ 2.7 จะถูกนำไปประมาณค่าตำแหน่งที่มีค่าความถี่ใกล้เคียงกับความถี่ที่คำนวณได้ และนำค่าความถี่นั้นมาสร้างฟิลเตอร์แบงค์แสดงได้ดังรูปที่ 3.3



รูปที่ 3.3 ตัวกรองสามเหลี่ยมในความถี่ดิจิตอล

ซึ่งชุดของตัวกรองสามเหลี่ยมที่ได้นี้คือฟิลตอร์แบงค์ ฟิลเตอร์แบงค์นี้จะถูกนำไปคูณกับค่าสเปคตรัมของแต่ละเฟรมเสียงที่ได้จากขั้นตอนการแปลงฟูเรียร์แบบไม่ต่อเนื่องภายในช่วงความถี่ของแต่ละ ตัวกรองนั้นจะได้ค่าสเปคตรัมที่ความถี่ต่างๆ ซึ่งสเปคตรัมความถี่นั้นๆจะถูกรวมเป็นสเปคตรัมความถี่เพียงค่าเดียวและทำการหาค่าพลังงานที่มีอยู่ในแต่ละตัวกรอง จากขั้นตอนนี้จะทำให้ได้ค่าพลังงานออกมาทั้งหมด 26 ค่า ค่าพลังงานนี้จะถูกทำลอกการิทึ่มและนำไปใช้ในขั้นตอนการแปลงโคซายน์แบบไม่ต่อเนื่องต่อไป

3.3.2.6 การแปลงโคซายน์แบบไม่ต่อเนื่อง

ค่าการแปลงโคซายน์แบบไม่ต่อเนื่องจะถูกนำมาใช้กับค่าพลังงานที่ได้จาก ขั้นตอนการใช้เมลฟิลเตอร์แบงค์ สำหรับการแปลงจากค่าที่อยู่ของลอกการิทึ่มที่อยู่ในโดเมนความถี่ มาเป็นโดเมนเวลา เนื่องจากฟิลเตอร์แบงค์ที่มีการซ้อนทับกันค่าพลังงานที่ได้นั้นก็จะมีความสัมพันธ์ กันอยู่ด้วย จึงใช้การแปลงโคซายน์แบบไม่ต่อเนื่องในการลดความสัมพันธ์และความซับซ้อนของ พลังงานลง ซึ่งค่าสัมประสิทธิที่ได้จากการแปลงโคซายน์แบบไม่ต่อเนื่องนั้นก็คือค่าสัมประสิทธิ์เซปสตรัมนั่นเอง โดยค่าสัมประสิทธิ์เซปสตรัมที่ได้มีทั้งหมด 26 ค่า แต่ค่าสัมประสิทธิ์ลำดับที่ 2-13 เท่านั้นที่ถูกนำมาใช้ จากขั้นตอนนี้ทำให้ได้ค่าสัมประสิทธิ์ทั้งหมด 12 ค่า

3.3.2.7 การหาสัมประสิทธิ์พลังงานและเดลต้าเซปสตรัม

การหาสัมประสิทธิ์พลังงานและเดลต้าเซปสตรัมเป็นการคำนวณเพื่อเพิ่มค่าสัมประสิทธิ์ โดยค่าสัมประสิทธิ์พลังงานนั้นเป็นการหาค่าพลังงานจากเฟรมเสียงย่อยที่ผ่านกระบวนการแฮมมิ่งวินโดว์ สามารถคำนวณได้จากสมการที่ 2.9 ได้เป็นค่าสัมประสิทธิ์อีกหนึ่งค่าออกมา ส่วน การหาสัมประสิทธิ์เดลต้าเซปสตรัมเป็นการหาอัตราการเปลี่ยนแปลงของค่าสัมประสิทธิ์เซปสตรัมระหว่างเฟรม ซึ่งคำนวณดังสมการที่ 2.10 จากการคำนวณนี้จะได้ค่าสัมประสิทธิ์อีก 12 ค่า และค่าสัมประสิทธิ์ตัวที่ 26 สามารถหาได้จากการคำนวณอัตราการเปลี่ยนแปลงของค่าสัมประสิทธิ์พลังงานระหว่างเฟรมซึ่งเรียกว่าเดลต้าพลังงาน คำนวณได้ดังสมการที่ 2.11

บทที่ 4

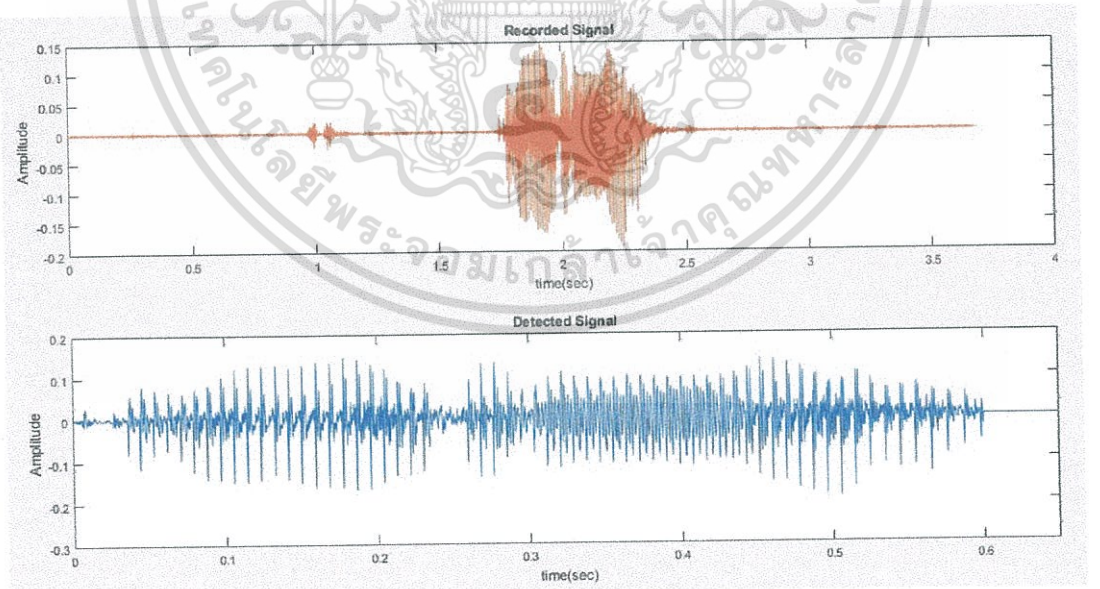
ผลการทดลอง

4.1 การสกัดคุณลักษณะของสัญญาณเสียงพูดโดยใช้ในอัลกอริทึ่ม MFCC

การสกัดคุณลักษณะของสัญญาณเสียงพูดซึ่งเลือกใช้อัลกอริทึ่ม MFCC มีขั้นตอนของการเตรียมสัญญาณเสียงก่อนการทำการสกัดคุณลักษณะเด่นของเสียงทั้งหมด 7 ขั้นตอนซึ่งประกอบด้วย 1) การเตรียมสัญญาณเสียง 2) การทำพรีเอมฟาซีสกับสัญญาณเสียงพูด 3) การแบ่งเฟรมเสียงพูด 4) การทำแฮมมิ่งวินโดว์ 5) การแปลงฟูเรียร์แบบไม่ต่อเนื่อง 6) การทำเมลฟิลเตอร์แบงค์ 7) การแปลงโคซายน์แบบไม่ต่อเนื่อง และ 8)การหาสัมประสิทธิ์พลังงานและเดลต้าเซปสตรัม โดยแสดงผลการทดลองในแต่ละขั้นตอนตามลำดับดังนี้

4.1.1 การเตรียมสัญญาณเสียง

สัญญาณเสียงที่ถูกบันทึกขนาด 2 วินาทีนั้นประกอบไปด้วยส่วนที่เป็นเสียงพูดและส่วนที่ไม่ใช่เสียงพูด ดังนั้นจึงทำการตัดส่วนที่ไม่ใช่เสียงพูดออกด้วยวิธีการกำหนดขอบเขตที่รับได้ (Thresholding) แสดงได้ในรูปที่ 4.1

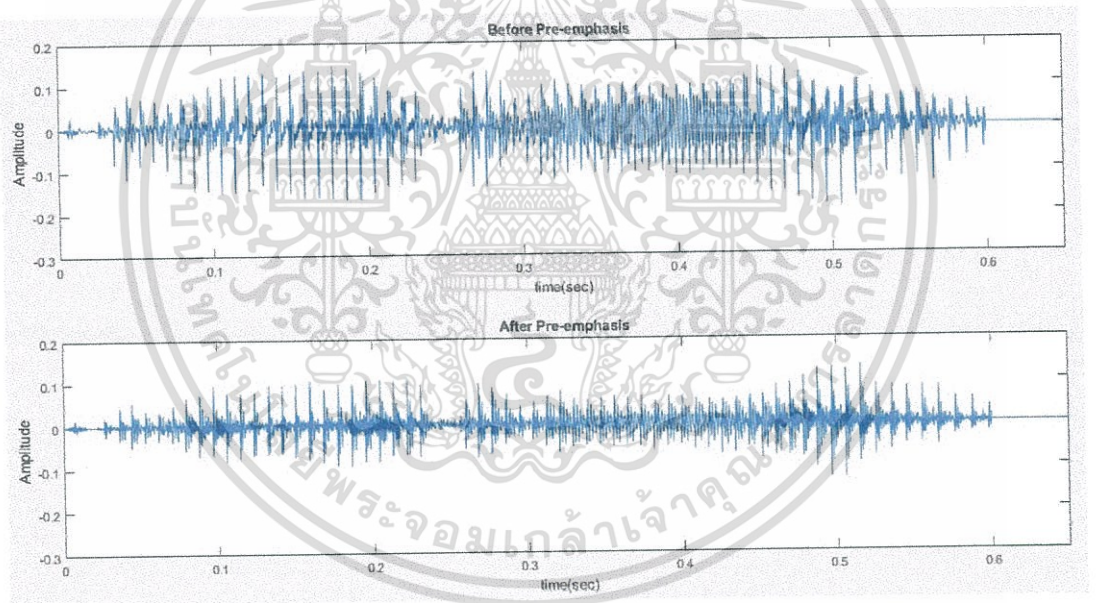


รูปที่ 4.1 การตัดส่วนของสัญญาณที่ไม่ใช่เสียงพูด

จากรูปที่ 4.1 แถบรูปด้านบนแสดงสัญญาณเสียงที่ถูกบันทึก และแถบรูปด้านล่างแสดงสัญญาณเสียงที่ถูกตัดส่วนสัญญาณที่ไม่ใช่เสียงพูดออก

4.1.2 พรีเอมฟาซีสของสัญญาณเสียงพูด

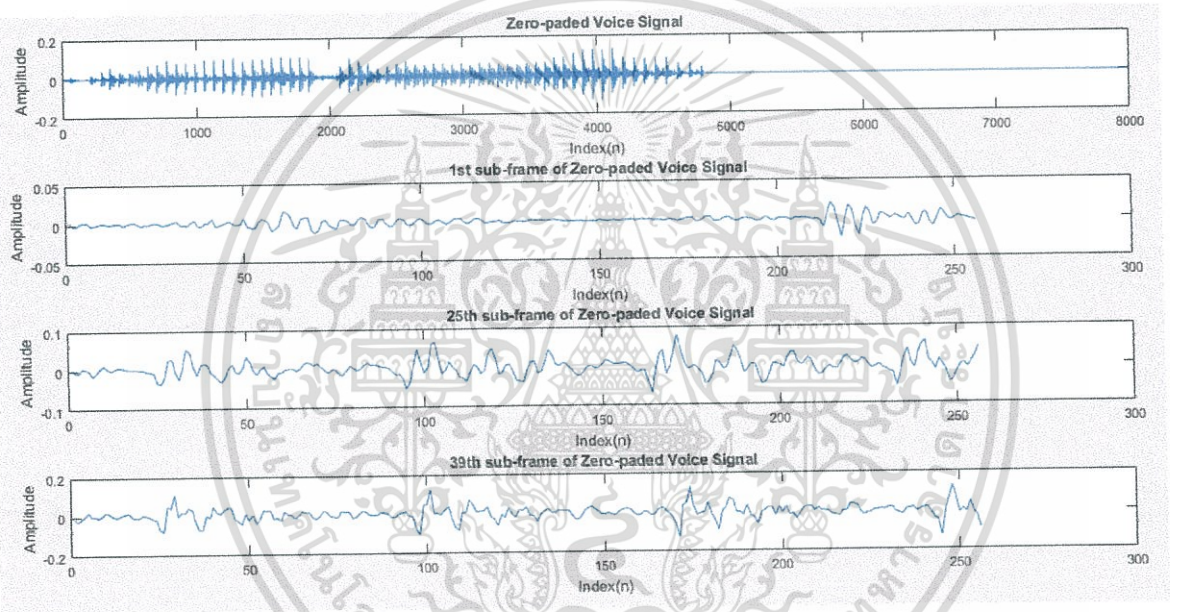
พรีเอมฟาซีสของสัญญาณเสียงพูดเพื่อเพิ่มปริมาณของพลังงานที่ความถี่สูงของสัญญาณเสียง ซึ่งโดยทั่วไปสัญญาณเสียงมีปริมาณพลังงานที่ความถี่ต่ำมากกว่าความถี่สูง ผลการพรีเอมฟาซีสของสัญญาณเสียงพูดแสดงได้ดังรูปที่ 4.2 โดยรูปแถบที่หนึ่งเป็นสัญญาณเสียงที่ถูกตัดเสียงที่ไม่ใช่เสียงพูดออก แล้วนำสัญญาณเสียงนั้นมาผ่านกระบวนการพรีเอมฟาซีสแสดงผลได้ดังแถบรูปที่สอง แต่ผลในรูปแถบที่สองเป็นการแสดงผลในโดเมนเวลาทำให้ไม่สามารถเห็นผลการทำพรีเอมฟาซีสได้ชัดเจน จึงแสดงผลของเสียงก่อนและหลังทำพรีเอมฟาซีสเป็นโดเมนความถี่ในรูปแถบที่สาม และสี่ตามลาดับ ในรูปแถบที่สี่จะแสดงให้เห็นว่าช่วงความถี่สูงถูกขยายขึ้น



รูปที่ 4.2 กราฟแสดงก่อนและหลังทำกระบวนการพรีเอมฟาซีส

4.1.3 การแบ่งเฟรมเสียงพูด

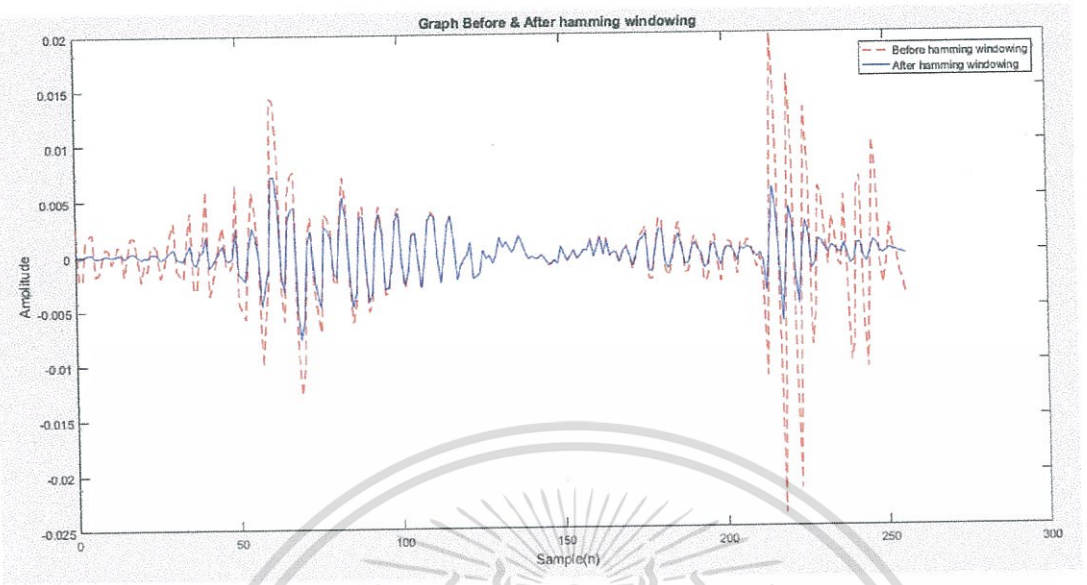
การแบ่งเฟรมจะนำเสียงที่ผ่านการทำฟรีเอมฟาซีสมาแบ่งเป็นเฟรมย่อย โดยเสียงพูดที่ถูกตัดมานั้นจะมีความยาวที่ไม่เท่ากัน ซึ่งก่อนทำการแบ่งเฟรมสัญญาณเสียงจะผ่านการทำ zero-pad เพื่อปรับให้สัญญาณเสียงแต่ละสัญญาณมีความยาวเท่ากัน เมื่อทำการzero-pad สัญญาณเสียงเรียบร้อยแล้ว เสียงที่ได้จะถูกนำมาแบ่งเฟรมให้เป็นเฟรมย่อย แสดงได้ดังรูปที่ 4.3 โดยแถบรูปแรกแสดงสัญญาณเสียงพูดที่ผ่านการทำ Zero-pad ก่อนทำการแบ่งเฟรม ส่วนแถบรูปแถบที่สอง สาม และสี่แสดงเฟรมย่อยของสัญญาณเสียงที่ถูกแบ่งเป็นเฟรมย่อยขนาด 256 จุดข้อมูล



รูปที่ 4.3 เฟรมย่อยของสัญญาณเสียง

4.1.4 การแบ่งเฟรมเสียงพูด

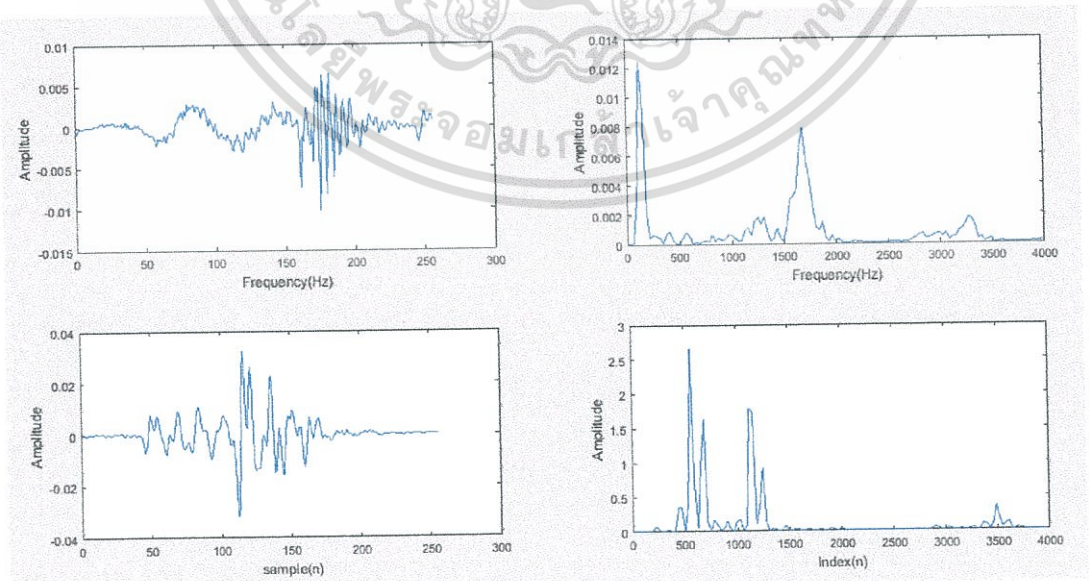
การทำแฮมมิ่งวินโดว์เป็นการนำแต่ละเฟรมย่อยของสัญญาณเสียงพูดมาคูณกับแฮมมิ่งวินโดว์ เพื่อให้แต่ละเฟรมย่อยนั้นมีลักษณะเป็นสัญญาณที่ต่อเนื่องในจุดเริ่มต้นของเฟรมและจุดปลายของเฟรม โดยผลของการทำแฮมมิ่งวินโดว์แสดงได้ดังรูปที่ 4.4 โดยสัญญาณเส้นประแสดงเฟรมย่อยของสัญญาณเสียงก่อนทำแฮมมิ่งวินโดว์ และสัญญาณเส้นทึบแสดงเฟรมย่อยของสัญญาณเสียงหลังจากผ่านการทำแฮมมิ่งวินโดว์แล้ว



รูปที่4.4 สัญญาณที่ผ่านการทำแฮมมิ่งวินโดว์

4.1.5 การแปลงฟูเรียร์แบบไม่ต่อเนื่อง

ขั้นตอนการแปลงฟูเรียร์แบบไม่ต่อเนื่องจะนำเฟรมสัญญาณที่ผ่านการทำวินโดว์มาแปลงจากโดเมนเวลาเป็นโดเมนความถี่เพื่อทำการประมาณหาค่าสเปคตรัมของสัญญาณเสียงแต่ละเฟรม แสดงได้ดังรูปที่ 4.5 โดยแถบรูปด้านซ้ายแสดงเฟรมย่อยของสัญญาณเสียงพูดที่ผ่านการทำแฮมมิ่งวินโดว์ และแถบรูปด้านขวาแสดงผลลัพธ์จากการแปลงฟูเรียร์แบบไม่ต่อเนื่องของสัญญาณที่ผ่านการทำวินโดว์

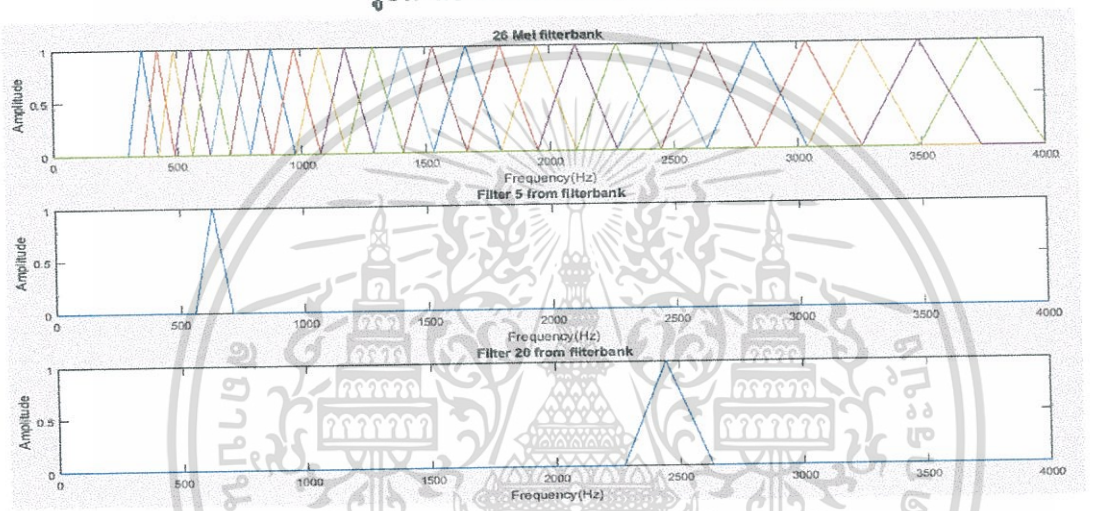


รูปที่ 4.5 เฟรมย่อยของสัญญาณเสียงในโดเมนเวลา(ซ้าย) และโดเมนความถี่(ขวา)

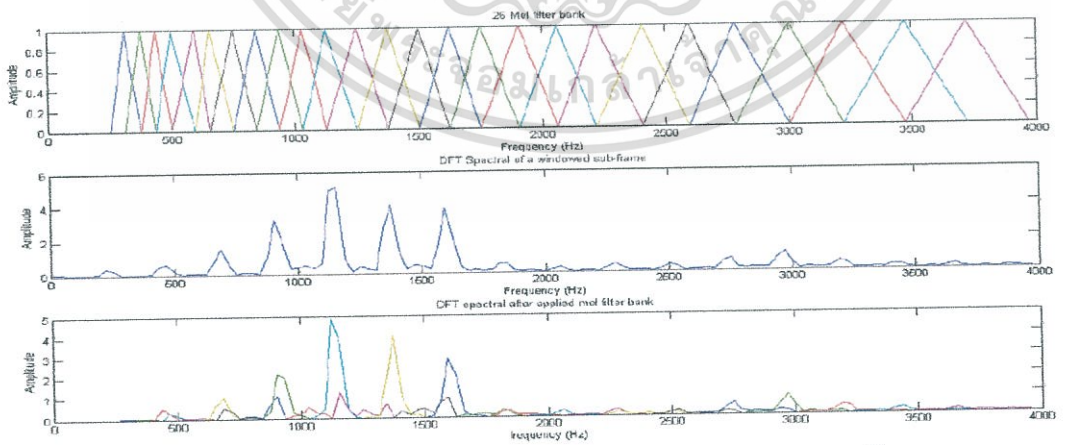
4.1.6 การทำเมลฟิลเตอร์แบงค์

การทำเมลฟิลเตอร์แบงค์เป็นการใช้ชุดของตัวกรองสามเหลี่ยม 26 ตัวคูณกับค่าสเปคตรัมของเฟรมย่อยของสัญญาณเสียง ค่าที่ได้หลังจากการใช้ตัวกรองสามเหลี่ยมแต่ละตัวเป็นค่าสเปคตรัมที่ความถี่ต่างๆ แสดงได้ดังรูปที่ 4.6 และค่าสเปคตรัมที่ได้จากการใช้ตัวกรองสามเหลี่ยมทั้งหมด 26 ตัว แสดงได้ดังรูปที่ 4.7 โดยค่าเสปคตรัมในแต่ละตัวกรองจะถูกรวมกันแล้วทำการหาค่าพลังงาน และถูกทำลอการิทึมเพื่อใช้สำหรับขั้นตอนการแปลงโคซายน์แบบไม่ต่อเนื่องต่อไป

รูปที่ 4.6 เมลฟิลเตอร์แบงค์ แบบเต็มและย่อย



จากรูปที่ 4.6 แถบรูปด้านบนแสดงเมลฟิลเตอร์ที่เป็นชุดของตัวกรองสามเหลี่ยม 26 ตัวกรองแถบรูปแถบที่สองและสามแสดงตัวกรองสามเหลี่ยมในช่วงความถี่ที่ต่างกัน

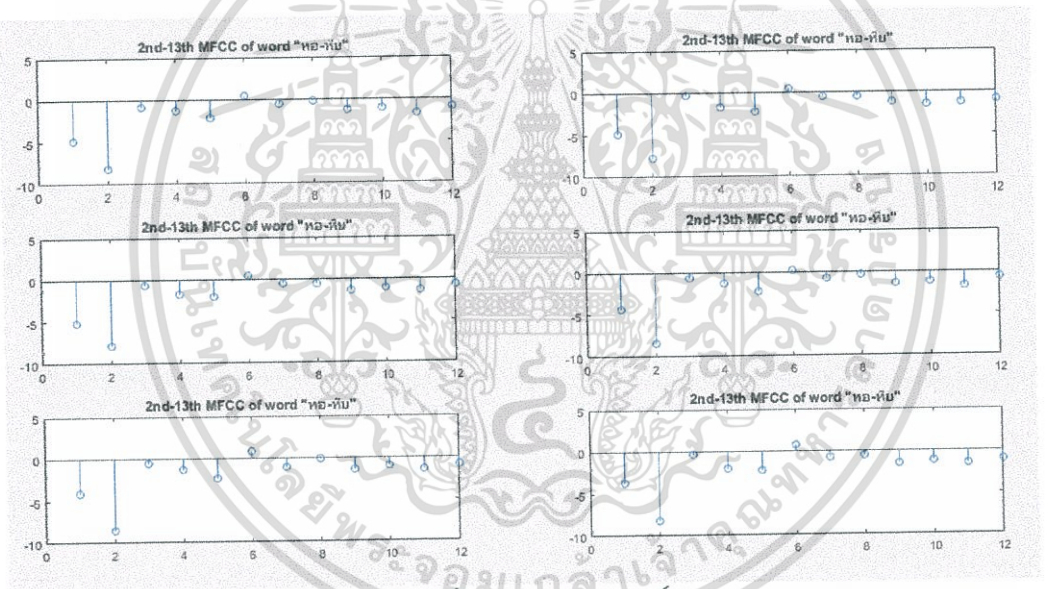


รูปที่ 4.7 เมลฟิลเตอร์แบงค์และสเปคตรัมในแต่ละตัวกรองทั้งหมด

จากรูปที่ 4.7 แถบรูปแรกแสดงเมลฟิลเตอร์ที่เป็นชุดของตัวกรองสามเหลี่ยม 26 ตัวกรอง แถบรูปแถบที่สองแสดงแถบสเปคตรัมของเฟรมย่อยของสัญญาณเสียงพูด และแถบรูปแถบสุดท้ายแสดงผลจากการคูณฟิลเตอร์แบงค์กับค่าสเปคตรัมของสัญญาณได้เป็นค่าสเปคตรัมที่ความถี่ต่างๆ ที่มีอยู่ในแต่ละตัวกรอง

4.1.7 การแปลงโคซายน์แบบไม่ต่อเนื่อง

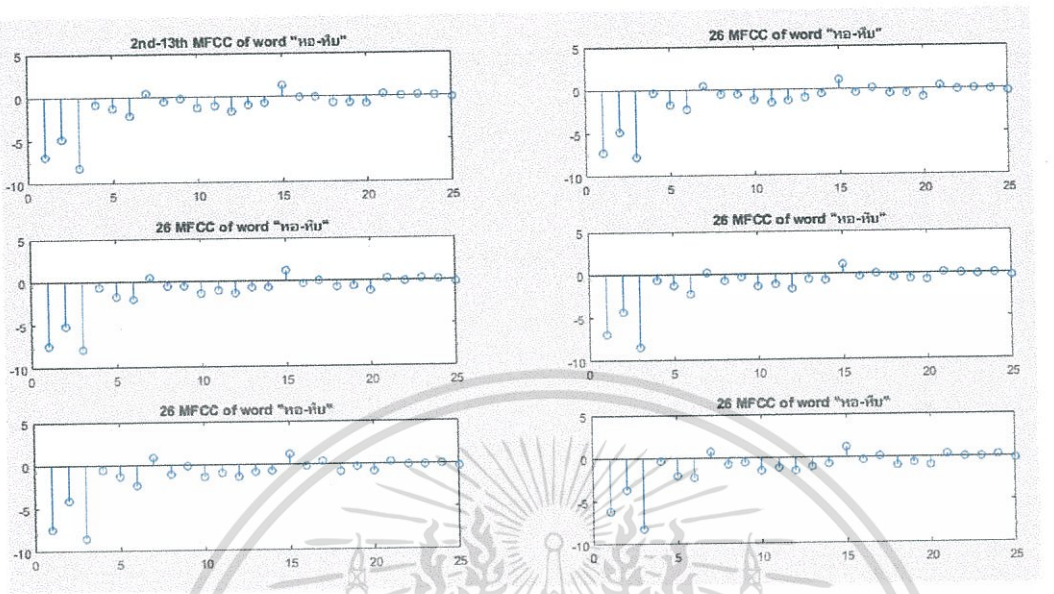
การแปลงโคซายน์แบบไม่ต่อเนื่องใช้สำหรับการแปลงจากค่าพลังงานลอการิทึ่มเมลสเปคตรัมที่อยู่ในโดเมนความถี่มาเป็นโดเมนเวลา โดยค่าสัมประสิทธิ์ที่ได้จากการแปลงโคซายน์แบบไม่ต่อเนื่องคือค่าสัมประสิทธิ์เซปสตรัม โดยได้ค่าสัมประสิทธิ์ 12 ค่าของแต่ละเฟรมย่อยของสัญญาณเสียงหนึ่งเสียงแสดงดังรูปที่ 4.8



รูปที่ 4.8 ค่าสัมประสิทธิ์เซปสตรัม

4.1.8 การหาสัมประสิทธิ์พลังงานและเดลต้าเซปสตรัม

การหาสัมประสิทธิ์พลังงานและเดลต้าเซปสตรัมเป็นการคำนวณเพื่อเพิ่มค่าสัมประสิทธิ์สำหรับเฟรมย่อย โดยค่าสัมประสิทธิ์พลังงานนั้นเป็นการหาค่าพลังงานจากเฟรมเสียงย่อยที่ผ่านกระบวนการแฮมมิ่งวินโดว์ ได้เป็นค่าสัมประสิทธิ์หนึ่งค่าออกมา ส่วนการหาสัมประสิทธิ์เดลต้าเซปสตรัมเป็นการหาอัตราการเปลี่ยนแปลงของค่าสัมประสิทธิ์เซปสตรัมระหว่างเฟรม จากการคานวณนี้จะได้ค่าสัมประสิทธิ์อีก 12 ค่า และค่าสัมประสิทธิ์ตัวที่ 26 สามารถหาได้จากการคำนวณอัตราการเปลี่ยนแปลงของค่าสัมประสิทธิ์พลังงานระหว่างเฟรมซึ่งเรียกว่า เดลต้าพลังงาน



รูปที่ 4.9 26 MFCC

4.2 การทดลองการรู้จำเสียง

ในส่วนต้นนี้เป็นการทดลองเพื่อวิเคราะห์เสียงแบบบุคคลเดี่ยว ตามหัวข้อ 3.2.1.1 และ 3.2.2.1 เพื่อที่จะทำการเลือก ค่าลักษณะสำคัญที่ดีที่สุดไปใช้ในการทดลองต่อไปส่วนต่อไป

4.3.1 การทดลองการรู้จำหน่วยเสียงพยัญชนะต้นภาษาไทยโดยใช้ค่าสัมประสิทธิ์ MFCC จำนวน 12 ตัว (สัมประสิทธิ์ตัวที่ 2-13) และ delta MFCC รวมสัมประสิทธิ์ทั้งหมด 24 ค่า

ตาราง 4-1 ค่าลักษณะสำคัญที่ใช้ในการทดลอง 4.3.1

ค่าลักษณะสำคัญ	จำนวน
MFCC	12
delta MFCC	12

นำค่าลักษณะสำคัญที่ได้เข้าสู่โครงข่ายประสาทเทียมเพื่อรู้จำและทดสอบการรู้จำ ผลอัตราการรู้จำดังตารางที่ 4-2

ตารางที่ 4-2 อัตราการรู้จำเสียง โดยใช้ค่าสัมประสิทธิ์ MFCC และ delta MFCC

พยัญชนะ	อัตราการรู้จำ			
	เพศชาย		เพศหญิง	
	คนที่1	คนที่2	คนที่1	คนที่2
ก	100	100	50	70
ข	100	100	50	90
ฃ	100	100	80	90
ค	80	60	90	80
ฅ	80	80	80	80
ฆ	100	90	100	90
ง	90	100	100	90
จ	80	90	80	70
ฉ	100	100	50	90
ช	50	50	80	90
ซ	100	90	30	10

ฌ	100	90	80	90
ญ	80	80	80	80
ฏ	80	90	70	90
ฏ	40	40	20	70
ฐ	100	60	70	50
ฑ	80	100	80	90
ฒ	90	90	70	50
ณ	80	80	70	90
ด	80	80	40	20
ต	90	60	80	80
ถ	60	100	90	90
ท	90	80	90	70
ธ	100	100	40	40
น	100	60	90	90
บ	80	70	90	100
ป	90	40	60	20
ผ	100	70	90	100
ฝ	60	80	60	70
พ	60	70	30	50
ฟ	80	70	40	70
ภ	80	80	60	60
ม	100	90	100	60
ย	50	80	80	60
ร	90	90	60	90
ล	60	90	30	70
ว	90	90	90	70
ศ	60	80	30	70
ษ	100	100	100	60
ส	50	90	20	10
ห	80	80	100	60

ฬ	90	100	80	40
อ	90	70	70	40
ฮ	100	100	80	100
เฉลี่ย	83.18	82.05	68.41	69.28

4.3.2 การทดลองการรู้จำหน่วยเสียงพยัญชนะต้นภาษาไทยโดยใช้ค่าสัมประสิทธิ์ MFCC จำนวน 12 ตัว (สัมประสิทธิ์ตัวที่ 2-13) , delta MFCC และ delta delta MFCC รวมสัมประสิทธิ์ทั้งหมด 36 ค่า

ตาราง 4-3 ค่าลักษณะสำคัญที่ใช้ในการทดลอง 4.3.2

ค่าลักษณะสำคัญ	จำนวน
MFCC	12
delta MFCC	12
delta delta MFCC	12

นำค่าลักษณะสำคัญที่ได้เข้าสู่โครงข่ายประสาทเทียมเพื่อรู้จำและทดสอบการรู้จำ ผลอัตราการรู้จำดังตารางที่ 4-4

ตารางที่ 4-4 อัตราการรู้จำเสียง โดยใช้ค่าสัมประสิทธิ์ MFCC , delta MFCC และ delta deltaMFCC

พยัญชนะ	อัตราการรู้จำ			
	เพศชาย		เพศหญิง	
	คนที่1	คนที่2	คนที่1	คนที่2
ก	100	100	80	70
ข	70	50	60	60
ฃ	90	90	100	100
ค	80	90	90	90
ฅ	90	70	80	90
ฆ	100	80	100	100
ง	100	80	80	100

จ	90	100	80	90
ฉ	100	100	50	80
ช	90	80	90	70
ซ	90	70	20	10
ฌ	100	60	80	100
ญ	70	100	80	60
ฎ	70	100	70	60
ฏ	50	100	30	40
ฐ	100	70	80	60
ฑ	70	60	100	90
ฒ	100	90	80	60
ณ	80	80	100	90
ด	80	90	60	60
ต	70	70	80	90
ถ	60	70	80	40
ท	90	90	100	60
ธ	90	100	20	30
น	100	90	100	100
บ	60	50	90	100
ป	70	70	30	30
ผ	50	80	70	100
ฝ	50	70	60	70
พ	60	60	30	20
ฟ	80	70	20	80
ภ	90	80	60	30
ม	70	60	100	40
ย	70	60	30	40
ร	100	50	70	90
ล	90	70	07	40
ว	100	100	80	90

ศ	70	100	10	50
ษ	100	100	80	70
ส	100	100	80	20
ห	60	100	80	70
ฬ	80	70	30	70
อ	80	100	50	60
ฮ	100	100	100	100
เฉลี่ย	82.05	81.14	68.41	67.50

4.3.3 การทดลองการรู้จำหน่วยเสียงพยัญชนะต้นภาษาไทยโดยใช้ค่าสัมประสิทธิ์ MFCC จำนวน 12 ตัว (สัมประสิทธิ์ตัวที่ 2-13) , delta MFCC + Energy และ delta delta MFCCรวม สัมประสิทธิ์ทั้งหมด 37 ค่า

ตาราง 4-5 ค่าลักษณะสำคัญที่ใช้ในการทดลอง 4.3.3

ค่าลักษณะสำคัญ	จำนวน
MFCC	12
delta MFCC + Energy	13
delta delta MFCC	12

นำค่าลักษณะสำคัญที่ได้เข้าสู่โครงข่ายประสาทเทียมเพื่อรู้จำและทดสอบการรู้จำ ผลอัตราการรู้จำดังตารางที่ 4-6

ตารางที่ 4-6 อัตราการรู้จำเสียง โดยใช้ค่าสัมประสิทธิ์ MFCC , delta MFCC + Energy และ delta delta MFCC

พยัญชนะ	อัตราการรู้จำ			
	เพศชาย		เพศหญิง	
	คนที่1	คนที่2	คนที่1	คนที่2
ก	100	90	90	70
ข	100	80	80	80
ฃ	100	100	100	100

ค	70	80	90	80
ศ	80	80	80	90
ฆ	80	80	100	100
ง	80	100	100	100
จ	60	50	100	80
ฉ	80	90	90	90
ช	50	70	100	70
ซ	100	100	20	20
ฌ	80	80	80	100
ญ	70	90	90	80
ฎ	40	70	70	100
ฏ	90	50	30	70
ฐ	90	90	70	20
ฑ	70	100	100	90
ฒ	100	90	90	70
ณ	80	90	70	90
ด	60	60	80	50
ต	90	60	80	70
ถ	90	90	80	70
ท	100	80	90	100
ธ	100	100	70	40
น	90	60	100	100
บ	100	80	90	100
ป	70	60	10	50
ผ	70	60	100	100
ฝ	70	70	70	20
พ	60	70	40	30
ฟ	60	50	70	50
ภ	80	80	50	40
ม	60	90	60	70

ย	90	80	20	80
ร	90	100	60	100
ล	90	60	50	30
ว	100	80	50	80
ศ	70	60	50	30
ษ	90	100	80	90
ส	80	80	60	30
ห	80	80	80	30
ฬ	80	90	50	60
อ	80	90	50	70
ฮ	100	70	70	100
เฉลี่ย	81.14	79.09	71.36	70.23

4.3.4 การทดลองการรู้จำหน่วยเสียงพยัญชนะต้นภาษาไทยโดยใช้ค่าสัมประสิทธิ์ MFCC จำนวน 12 ตัว (สัมประสิทธิ์ตัวที่ 2-13) , delta MFCC + energy และ delta delta MFCCรวม สัมประสิทธิ์ทั้งหมด 38 ค่า

ตาราง 4-7 ค่าลักษณะสำคัญที่ใช้ในการทดลอง 4.3.4

ค่าลักษณะสำคัญ	จำนวน
MFCC	12
delta MFCC + Energy	13
delta delta MFCC + Energy	13

นำค่าลักษณะสำคัญที่ได้เข้าสู่โครงข่ายประสาทเทียมเพื่อรู้จำและทดสอบการรู้จำ ผลอัตราการรู้จำดังตารางที่ 4-8

ตารางที่ 4-8 อัตราการรู้จำเสียง โดยใช้ค่าสัมประสิทธิ์ MFCC,delta MFCC + energy และ delta-delta MFCC + Energy

พยัญชนะ	อัตราการรู้จำ			
	เพศชาย		เพศหญิง	
	คนที่1	คนที่2	คนที่1	คนที่2
ก	80	100	70	80
ข	90	80	70	70
ฃ	100	50	100	90
ค	80	100	100	80
ฅ	100	80	80	90
ฆ	90	70	90	100
ง	70	100	80	70
จ	80	100	80	60
ฉ	100	80	70	100
ช	60	90	100	90
ซ	90	40	30	30
ฌ	90	90	50	70
ญ	90	100	80	70
ฎ	60	100	70	70
ฏ	90	100	20	60
ฐ	90	100	100	80
ฑ	50	50	100	70
ฒ	100	40	90	90
ณ	100	100	70	80
ด	70	100	50	90
ต	70	100	50	70
ถ	80	50	90	100
ท	90	100	100	100
ธ	100	90	60	70
น	80	40	100	80

บ	90	30	80	60
ป	50	100	80	50
ผ	80	80	100	60
ฝ	80	50	50	90
พ	70	90	60	30
ฟ	100	30	07	50
ภ	80	70	40	80
ม	70	100	70	90
ย	60	60	90	80
ร	60	50	70	80
ล	60	70	30	60
ว	80	100	70	90
ศ	80	100	80	40
ษ	100	100	100	50
ส	80	90	80	60
ห	80	100	100	70
ฬ	70	80	60	90
อ	80	100	80	70
ฮ	100	90	70	100
เฉลี่ย	81.14	80.45	74.55	74.09

จากผลการทดลองโดยใช้ข้อมูลจากการวิเคราะห์เสียงแบบบุคคลเดี่ยว ในการฝึกฝนระบบรู้จำเสียง

ผลการทดลองในตาราง 4-2 , 4-4 , 4-6 และ 4-8 พบว่าอัตราการรู้จำหน่วยเสียงพยัญชนะภาษาไทยเฉลี่ยของเพศชายใน 4 ตารางมีค่าเฉลี่ยที่ใกล้เคียงกัน แต่จากค่าเฉลี่ยตารางที่ 4-2 มีค่าสูงที่สุด ซึ่งใช้ค่าสัมประสิทธิ์ MFCC และ delta MFCC ให้ผลการรู้จำเสียงพยัญชนะภาษาไทยเฉลี่ยสูงที่สุด แต่ในส่วนของเพศหญิงตารางที่ 4-8 มีค่าสูงที่สุด ซึ่งใช้ค่าสัมประสิทธิ์ MFCC , delta MFCC + energy และ delta-delta MFCC + Energy

หลังจากที่ได้วิธีการที่ดีที่สุดจากการวิเคราะห์เสียงแบบบุคคลเดียวแล้วก็นำวิธีการนั้นมาทดลองวิเคราะห์เสียงแบบหลายบุคคลต่อไปและมีผลการทดลองดังนี้

ตารางที่ 4-9 ตารางแสดงผลการทดลองการวิเคราะห์เสียงผู้ชายแบบหลายบุคคล

ค่าลักษณะสำคัญ	จำนวน Features	อัตราการรู้จำ	
		Train	Test
MFCC , delta MFCC	24	93.24	78.38
MFCC, delta MFCC , delta-delta MFCC	36	94.05	78.12
MFCC, delta MFCC+ energy , delta-delta MFCC	37	94.13	77.79
MFCC, delta MFCC+ energy , delta-delta MFCC+ energy	38	95.02	78.04

ตารางที่ 4-10 ตารางแสดงผลการทดลองการวิเคราะห์เสียงผู้หญิงแบบหลายบุคคล

ค่าลักษณะสำคัญ	จำนวน Features	อัตราการรู้จำ	
		Train	Test
MFCC , delta MFCC	24	88.05	74.02
MFCC, delta MFCC , delta-delta MFCC	36	88.62	74.68
MFCC, delta MFCC+ energy , delta-delta MFCC	37	88.74	75.88
MFCC, delta MFCC+ energy , delta-delta MFCC+ energy	38	89.23	78.47

จากผลการทดลองอัตราการรู้จำเสียงของเพศชายแบบหลายบุคคลในตารางที่ 4-9 ปรากฏว่าเมื่อใช้ค่า MFCC , delta MFCC ให้อัตราการรู้จำสูงสุดคือ 78.38 และสำหรับเพศหญิงแบบหลายบุคคลในตารางที่ 4-10 มีค่าสูงสุดคือ 78.47 เมื่อใช้ค่า MFCC, delta MFCC+ energy , delta-delta MFCC+ energy

บทที่ 5

สรุปและวิเคราะห์ผลการทดลอง

5.1 สรุปผลการทดลอง

จากการทดลองในโครงงานเรื่อง การจำลองพยัญชนะภาษาไทยด้วยเสียง "Thai Consonant Voice Model" โดยใช้โครงข่ายประสาทเทียม โดยใช้เสียงพูดพยัญชนะภาษาไทย จำนวน 44 เสียง ของผู้ทดลองผู้จัดทำได้นำวิธีการและขั้นตอนต่างๆในการรู้จำเสียงพูด เข้ามาช่วยในการวิเคราะห์เพื่อให้สามารถรู้จำเสียงพยัญชนะไทย โดยเริ่มจากการนำสัญญาณผ่านขั้นตอนการประมวลผลสัญญาณเบื้องต้น ขั้นตอนการสกัดลักษณะสำคัญ ด้วยเทคนิคสัมประสิทธิ์เซปสตรัมบนเมลสเกล ค่าพลังงานเสียง และใช้ค่า MFCC มาคำนวณค่า delta MFCC และ delta delta MFCC ส่วนในขั้นตอนการทดสอบความคล้ายคลึงกันของรูปแบบและกฎเกณฑ์การตัดสินใจใช้โครงข่ายประสาทเทียมในการฝึกฝน รู้จำเสียงพยัญชนะไทยและตัดสินใจเปรียบเทียบผลการออกเสียง

จากการทดลองพบว่าในขั้นตอนการรู้จำเสียงโดยใช้ค่าสัมประสิทธิ์เซปสตรัมบนเมลสเกล จำนวน 12 ค่า (สัมประสิทธิ์ตัวที่ 2-13) ค่า delta MFCC , delta delta MFCC และค่าพลังงานมาใช้วิเคราะห์ข้อมูลเสียงแบบบุคคลเดี่ยวในเพศหญิงและเพศชายให้มีผลการทดสอบถูกต้องมากขึ้น โดยให้ผลอัตราการรู้จำคิดเป็นร้อยละ 83.18 , 82.05 ในเพศชายและผลอัตราการรู้จำคิดเป็นร้อยละ 74.55 และ 74.09 สำหรับเพศหญิง และจากการทดลองในการวิเคราะห์ข้อมูลเสียงแบบหลายบุคคล ได้ผลอัตราการรู้จำสูงสุดคิดเป็นร้อยละ 78.38 ในเพศชายและผลอัตราการรู้จำสูงสุดคิดเป็นร้อยละ 78.47 สำหรับเพศหญิง

5.2 ปัญหาและอุปสรรค

1. การเก็บตัวอย่างเสียงกลุ่มต้นแบบ มีการออกเสียงในบางเสียงที่ไม่ชัดเจน เช่นเสียง รอ-เรือ และ เสียง ลอ-ลิง

2. ในขั้นตอนการเก็บตัวอย่างมีเสียงภายนอกแทรกเข้ามาบ้างทำให้มีผลกระทบในการวิเคราะห์ผล

5.3 ข้อเสนอแนะ

1. ในขั้นตอนการเก็บตัวอย่างเสียงเรื่อง การจำลองพยัญชนะภาษาไทยด้วยเสียงต้องมีการควบคุมในเรื่องเครื่องมือ อุปกรณ์ และสิ่งแวดล้อมเนื่องจากส่งผลต่อการรู้จำ ในการเลือกคำพูดเพื่อให้สำหรับการรู้จำเสียงพูดซึ่งอาจจะเป็นไปได้ค่อยข้างยาก

2. ในการรู้จำเสียงพูดโดยใช้โครงข่ายประสาทเทียม ข้อมูลที่ส่งให้โครงข่ายประสาทเทียมเรียนรู้หรือทดสอบต้องมีขนาดเวกเตอร์เท่ากันทั้งหมด ดังนั้นสามารถทดลองปรับเปลี่ยนวิธีการรู้จำเสียงด้วยวิธีอื่นๆ ที่สามารถเรียนรู้และเปรียบเทียบคุณลักษณะสำคัญของเสียงโดยข้อมูลเวกเตอร์ไม่จำเป็นต้องเท่ากัน เช่น วิธีไดนามิกไทม์วาร์ปปิง (DTW) วิธีฮิดเดนมาร์คอฟ โมเดล (HMM) เป็นต้น

3. โปรเจคนี้เป็นการรู้จำที่เน้นในส่วนของเสียงพยัญชนะภาษาไทย ซึ่งเป็นพยัญชนะ 44 เสียงเท่านั้น ดังนั้นสามารถพัฒนาให้สามารถรู้จำในส่วนอื่นๆ ของการออกเสียงได้ เช่น ส่วนของการออกเสียงสระ เป็นต้น

เอกสารอ้างอิง

[1] L. Muda, M. Begam, and I. Elamvazuthi. 2010. "Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques," Journal of Computing. 2(3) : 138 – 143

[2] P. Pal Singh and P. Rani. 2014. "An Approach to Extract Feature using MFCC," IOSR Journal of Engineering (IOSRJEN). 04(08)

[3] James Lyons. "Mel Frequency Cepstral Coefficient (MFCC) tutorial" http://practicalcryptography.com/miscellaneous/machine-learning/guide-melfrequency-cepstral-coefficients-mfccs/#eqn2

[4] http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/

[5] เยาวลักษณ ชาติสุขศิริเดช. การใช้เสียงในภาษาไทย. กรุงเทพฯ : สำนักพิมพ์อักษรเจริญทัศน, 2548.

[6] ปรีดาวรรณ เกษเมธีการุณ. การพัฒนาการรู้จำเสียงสำหรับพยัญชนะต้นของอัฒพยางค. วิทยานิพนธ์ครุศาสตรอุตสาหกรรมมหาบัณฑิต สาขาวิชาเทคโนโลยีคอมพิวเตอร์ ภาควิชาคอมพิวเตอร์ศึกษา บัณฑิตวิทยาลัย สถาบันเทคโนโลยีพระจอมเกล้าพระนคร เหนือ, 2548.

[7] Mohaddeseh Nosratighods. "*SPEAKER VERIFICATION USING A NOVEL SET OF DYNAMIC FEATURES*"

[8] *Amarin Deemagarn, Asanee Kawtrakul* "Thai Connected Digit Speech Recognition Using Hidden Markov Models"

[9] http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/

[10] ถนัด ศรีบุญเรือง และคณะ. "ขนาดและรูปร่างของเซลล์."
http://www.trueplookpanya.com/learning/detail/2169-002640.

[11] Wikipedia. "Artificial neural network."
https://en.wikipedia.org/wiki/Artificial_neural_network.