ระบบรู้จำเสียงพูดชนิดแยกคำขึ้นกับผู้พูดสำหรับผู้ป่วยโรคหลอดเลือดสมอง

Speaker Dependent Isolated Word Recognition for Stroke Patient

โอฬาร ดาวเวียง

Olarn Daowieng

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญา

วิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า

มหาวิทยาลัยสงขลานครินทร์

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Engineering in Electrical Engineering

Prince of Songkla University

2553

ชื่อวิทยานิพนธ์ ระบบรู้จำเสียงพูดชนิดแยกคำขึ้นกับผู้พูดสำหรับผู้ป่วยโรคหลอดเลือดสมอง
ผู้เขียน นายโอฬาร ดาวเวียง
สาขาวิชา วิศวกรรมไฟฟ้า

อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก **คณะกรรมการสอบ**

…….. ..……..……ประธานกรรมการ
(รองศาสตราจารย์บุญเจริญ วงศ์กิตติศึกษา) (รองศาสตราจารย์ ดร. มนตรี กาญจนะเดชะ)

..……...กรรมการ
(รองศาสตราจารย์บุญเจริญ วงศ์กิตติศึกษา)

...……..กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.พรชัย พฤกษ์ภัทรานนต์)

..…….……..กรรมการ
(ดร.ฉัตรชัย ศุภพิทักษ์สกุล)

บัณฑิตวิทยาลัย มหาวิทยาลัยสงขลานครินทร์ อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษา ตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า

…...……...…….
(รองศาสตราจารย์ ดร.เกริกชัย ทองหนู)
คณบดีบัณฑิตวิทยาลัย

ชื่อวิทยานิพนธ์ ระบบรู้จำเสียงพูดชนิดแยกคำขึ้นกับผู้พูดสำหรับผู้ป่วยโรคหลอดเลือดสมอง

ผู้เขียน นายโอฬาร ดาวเวียง

สาขาวิชา วิศวกรรมไฟฟ้า

ปีการศึกษา 2552

บทคัดย่อ

โรคหลอดเลือดสมองเป็นโรคที่เกิดกับระบบประสาท ส่งผลต่อความสามารถและการควบคุมกล้ามเนื้อในการพูด ทำให้ผู้ป่วยโรคนี้เปล่งเสียงได้ด้อยกว่าคนปกติ ระบบสั่งการโดยใช้เสียงทั่วไป จึงไม่เหมาะที่จะใช้กับกลุ่มคนเหล่านี้ ดังนั้นงานวิจัยนี้จึงได้ศึกษาการออกแบบระบบสั่งการโดยใช้เสียงพูดชนิดขึ้นกับผู้พูดสำหรับผู้ป่วยโรคหลอดเลือดสมองที่มีภาวะกลไกการออกเสียงบกพร่อง (Dysarthria) โดยใช้ระบบฟัซซี่ลอจิกในการรู้จำ, วิธีการแปลงเวฟเลทในการแยกช่วงความถี่ก่อนที่จะทำการสกัดคุณลักษณะของเสียงซึ่งประกอบด้วยความถี่ฟอร์แมนต์ที่หนึ่งความถี่ฟอร์แมนต์ที่สองและช่วงระยะเวลาในการออกเสียง คำสั่งที่ใช้เป็นเสียงสระแท้ในภาษาไทยจำนวน 6 คำ ประกอบด้วยสระแท้เสียงสั้น (อึ, เออะ, อะ) และสระแท้เสียงยาว (อือ, เออ, อา) จากผลการทดสอบกับผู้ป่วยอัมพาตเพศชาย 2 คนและคนปกติเพศชาย 1 คน พูดคนละ 6 คำ จำนวน 5 รอบ พบว่าวิธีการที่นำเสนอสามารถแยกเสียงของผู้ป่วยอัมพาตได้ดีกว่าคนปกติ โดยให้อัตราการรู้จำสูงถึง 97.2 %, 87.0%, 86.4% สำหรับกลุ่มผู้ทดสอบผู้ป่วยอัมพาตระดับ 6 (C6), ระดับ 4 (C4) และคนปกติตามลำดับ

คำสำคัญ รู้จำเสียงพูด ภาวะกลไกการออกเสียงบกพร่อง

Thesis Title Speaker Dependent Isolated Word Recognition for Stroke Patient

Author Mr. Olarn Daowieng

Major Program Electrical Engineering

Academic Year 2009

ABSTRACT

Stroke is associated with nervous system and affected to the ability of speech. As a result, conventional speech recognition is not suitable for stroke patient. This thesis studies the command system design of speaker dependence for stroke patient with dysarthric speech. The six vowels were used in recognition system consisting of 3 short unmixed vowels (ฺ, ั, a) and 3 long unmixed vowels (ฺฺ, ัั, aa). Wavelet transform is used to separate the frequency band before feature extraction, which includes the first formant frequency, second formant frequency, and length portion of vowel tone. Subsequently, fuzzy logic was used in speech recognition. Our approach was tested with two male stroke patients and one normal male. All volunteers pronounced 6 vowels for 5 rounds. Results show that the proposed methods can be appropriately used for the stroke patient. The accuracy of recognition rate is 97.2%, 87.0%, 86.4% for two stroke patients (C6 & C4) and normal people, respectively.

Keywords: Speech recognition, Dysarthria

สารบัญ

สารบัญ (ต่อ)

สารบัญ (ต่อ)

รายการตาราง

รายการภาพประกอบ

รายการภาพประกอบ (ต่อ)

รายการภาพประกอบ (ต่อ)

รายการภาพประกอบ (ต่อ)

บทที่ 1

บทนำ

1.1 ความสำคัญและที่มาของการวิจัย

การสื่อสารด้วยคำพูดเป็นหนึ่งในความสามารถขั้นพื้นฐานของมนุษย์ที่มีความจำเป็นและสำคัญอย่างมาก การสื่อสารด้วยคำพูดนับว่าเป็นวิธีการสื่อสารที่ใช้สนองความต้องการของมนุษย์ที่ปฏิบัติได้ง่าย ชัดเจน และสะดวกที่สุด เพราะใช้เพียง การเปล่งเสียงคำพูดออกมาเท่านั้น จากข้อดีเหล่านี้จึงทำให้มีผู้นำเอาการสื่อสารด้วยคำพูดเข้ามาประยุกต์ใช้กับเทคโนโลยีต่างๆ ในโลกปัจจุบันที่กำลังพัฒนาไปอย่างรวดเร็ว โดยเฉพาะทางด้านคอมพิวเตอร์ ระบบควบคุม และระบบโทรศัพท์ เป็นต้น การประยุกต์ใช้งานอย่างหนึ่งที่ได้รับความสนใจอย่างแพร่หลายและมีงานวิจัยอย่างต่อเนื่อง ก็คือ การสั่งงาน เครื่องมือเครื่องใช้ และอุปกรณ์ต่างๆด้วยเสียงพูด เช่น ระบบสั่งงานหุ่นยนต์ด้วยเสียงพูด ระบบหมุนหมายเลขโทรศัพท์ ด้วยเสียงพูด และ ระบบป้อนข้อมูลตัวเลขที่มีความยาวมากๆด้วยเสียงพูด เป็นต้น และการที่จะสามารถสั่งงานอุปกรณ์ต่างๆได้ด้วยเสียงพูดนี้ จำเป็นต้องมีระบบการรู้จำเสียงพูดประกอบอยู่ด้วย

ระบบสั่งการโดยใช้เสียงพูดส่วนใหญ่นั้นได้ถูกออกแบบมาใช้สำหรับคนทั่วไปที่มีการพูดแบบปกติ ในขณะเดียวกันผู้ที่บกพร่องทางการพูดอันเนื่องมาจากโรคหลอดเลือดสมอง (Dysarthria) ยังไม่สามารถใช้งานได้ดีเท่าที่ควรกับระบบสั่งการเหล่านี้ จากการสำรวจเอกสารวิจัยพบว่าส่วนใหญ่ในการออกแบบระบบสั่งการไม่ได้พิจารณาถึงความบกพร่องนี้และความบกพร่องในแต่ละคนจะแตกต่างกันเนื่องจากความรุนแรงของโรคหลอดเลือดสมอง ดังนั้นงานวิจัยนี้จึงได้ทำการศึกษาการออกแบบระบบสั่งการแบบใช้เสียงพูดเพื่อใช้สำหรับผู้ที่มีความบกพร่องทางการพูดชนิด Dysarthria และขึ้นกับผู้พูด โดยใช้ฟัซซี่ลอจิกสำหรับการรู้จำและตัดสินใจคำสั่งของระบบ

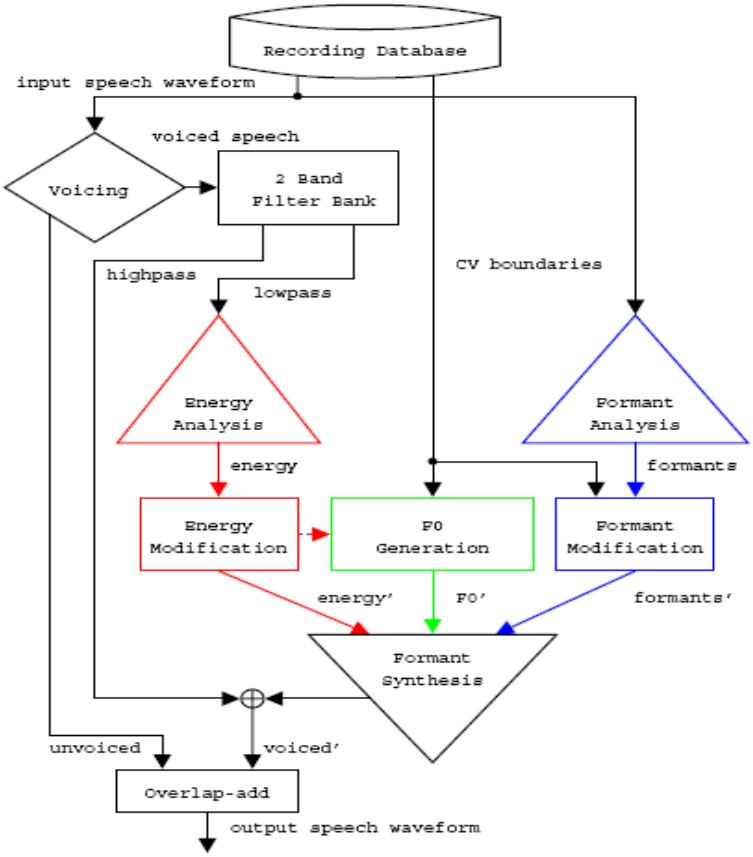
1.2 การตรวจเอกสาร บทความ และงานวิจัยที่เกี่ยวข้องกับประเด็นวิจัย

1.2.1 Improving the Intelligibility of Dysarthric Speech [1]

งานวิจัยนี้ได้นำเสนอการปรับปรุงคุณภาพสัญญาณเสียงพูดของผู้พูดประเภท Dysarthric ให้มีคุณภาพเทียบเท่ากับเสียงพูดปกติมากขึ้น โดยเปรียบเทียบกับกลุ่มทดสอบผู้พูดปกติ ซึ่งมีรายละเอียดดังนี้

- ขอบเขตของคำที่ใช้รู้จำ คือ เสียงสระทั้ง 8 คำ (i, I, E, @, u, U, ^, A)

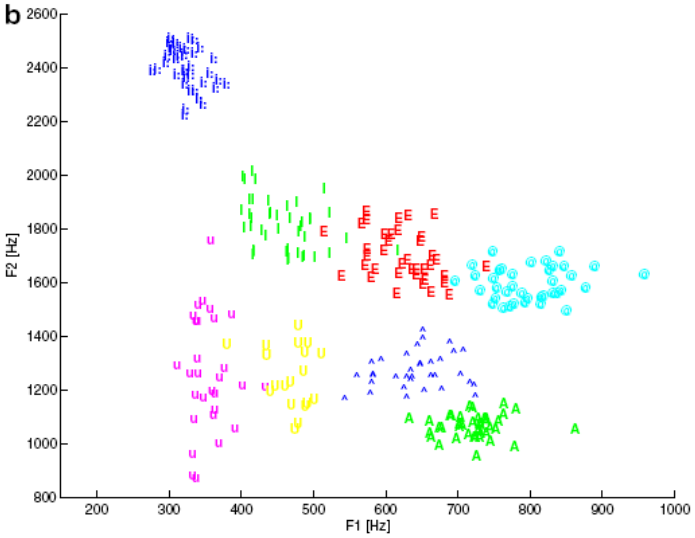
- พารามิเตอร์ที่ใช้ คือ ความถี่ฟอร์แมนต์ (Formant Frequency)
- กลุ่มผู้ทดสอบ ผู้ป่วยโรคหลอดเลือดสมองเพศชายมีความผิดปกติทางการพูดชนิด Dysarthria จำนวน 1 คน
- ผลลัพธ์ที่ได้คือ อัตราการรู้จำเพิ่มขึ้นจากร้อยละ 48 เป็นร้อยละ 54 โดยใช้ฐานข้อมูลจากเสียงพูดปกติเป็นเกณฑ์



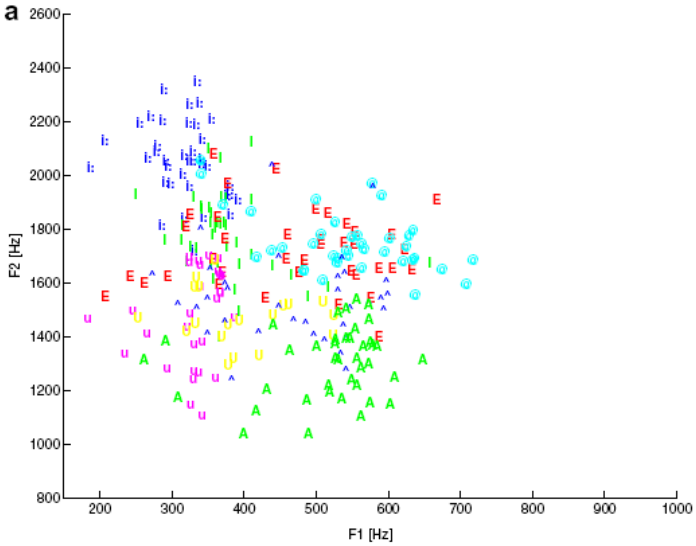
ภาพประกอบ 1-1 แผนผังการทำงานของระบบ [1]

ขั้นตอนแรกในการอัดเสียงโดยใช้ไมโครโฟนชนิด AKG HSC 200 ค่าความถี่สุ่มตัวอย่างเท่ากับ 16 kHz 16 bit จากนั้นสัญญาณถูกส่งต่อมาที่ Recording database ซึ่งทำหน้าที่เก็บสัญญาณเสียงในโดเมนเวลาของผู้ป่วยและสัญญาณเสียงที่ถูกต้องของผู้พูดปกติ (Data base) ผ่านมาทางขั้นตอน Voicing จะทำหน้าที่แยกสัญญาณเสียงพูด (Voiced speech) และไม่ใช่สัญญาณเสียงพูด (Unvoiced) ออกจากกัน ถัดไปที่ขั้นตอน 2 Band filter bank จะทำหน้าที่เป็นตัวกรองสัญญาณในรูปแบบความถี่สูงผ่าน (High pass) และความถี่ต่ำผ่าน (Low pass) ส่วนที่เป็น Low pass จะถูกส่งมายังขั้นตอน Energy analysis เพื่อทำการวิเคราะห์อัตราการแกว่งของพลังงานเสียง ถ้าพลังงานเสียงมีการแกว่งมากก็จะถูกปรับแต่งให้เรียบขึ้นโดยขั้นตอน Energy Modification ถัดไป

สัญญาณเสียงของผู้ป่วยจะผ่านมาทางขั้นตอน Formant Analysis ทำหน้าที่เป็นตัวตรวจเช็คความถี่ฟอร์แมนต์เบื้องต้นว่าผิดปกติจากความเป็นจริงมากเท่าใด หลังจากนั้นจะถูกส่งต่อไปยังขั้นตอน Formant Modification เพื่อทำการปรับแต่งความถี่ฟอร์แมนต์ให้มีความใกล้เคียงกับเสียงปกติ โดยมีข้อมูลจาก F0 Generation เป็นฐานข้อมูลจากผู้พูดปกติ สุดท้ายพารามิเตอร์ที่ถูกปรับแต่งทั้งหมดจะถูกส่งมายังขั้นตอน Formant Synthesis เพื่อทำการสังเคราะห์ความถี่ฟอร์แมนต์ใหม่ทำให้ได้รูปแบบเสียงพูดที่ถูกต้องมากยิ่งขึ้น



ภาพประกอบ 1-2 กราฟความถี่ฟอร์แมนต์ที่หนึ่งและสองของผู้พูดปกติจำแนกตามกลุ่มเสียง [1]



ภาพประกอบ 1-3 กราฟความถี่ฟอร์แมนต์ที่หนึ่งและสองของผู้พูดชนิด Dysarthria จำแนกตามกลุ่มเสียง [1]

ตารางที่ 1-1 แสดงความถูกต้องของสัญญาณเสียงหลังจากการปรับแต่งสัญญาณ มีหน่วยเป็นร้อยละ

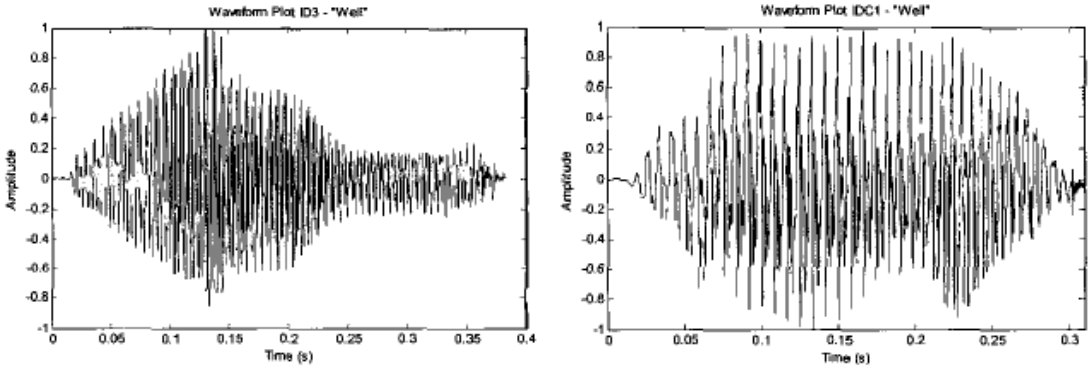
Vowels / Type	i	I	E	@	u	U	^	A	Average
Dysarthric	73	63	40	10	92	73	27	6	48
Dysarthric (improved)	67	42	54	83	46	52	19	73	54
Normal	100	98	100	100	98	92	100	100	93

1.2.2 Signal Processing for use in the Assessment of Dysarthric Speech [2]

เป็นงานวิจัยสำหรับใช้ในทางการแพทย์เพื่อประเมินความผิดปกติและความรุนแรงของอาการ Dysarthria ในผู้ป่วย โดยมีรายละเอียดดังนี้

- แพทย์ผู้เชี่ยวชาญจะประเมินผู้ป่วยจากกราฟสัญญาณเสียง, ความดังและอัตราเร็วในการพูด 1 ประโยค โดยอ้างอิงจากฐานข้อมูลที่ถูกต้อง
- ประโยคที่ใช้ทดสอบคือ “You wish to know all about my grandfather. Well, he is nearly 90 years old. Yet he still thinks as swiftly as ever.”
- พารามิเตอร์ที่ใช้ทดสอบ คือ ค่าความถี่ฟอร์แมนต์

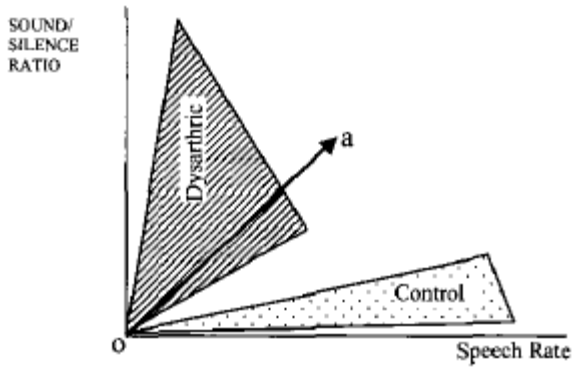
ขั้นตอนแรกแพทย์ผู้เชี่ยวชาญจะประเมินจากกราฟสัญญาณเสียงของผู้ป่วยเพื่อเปรียบเทียบกับเสียงพูดปกติ สังเกตบริเวณยอดของสัญญาณเสียงผู้ป่วยจะลดลงอย่างรวดเร็วเป็นขั้นบันไดและไม่มีความต่อเนื่องเมื่อเทียบกับเสียงพูดปกติ



(ก) ผู้พูดชนิด Dysarthria (ข) ผู้พูดปกติ

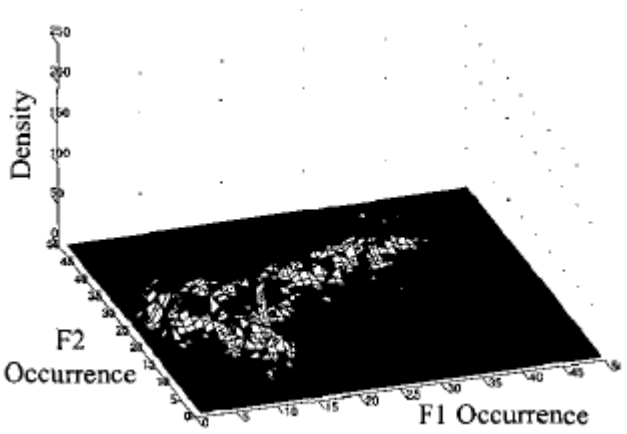
ภาพประกอบ 1-4 กราฟสัญญาณเสียงของคำพูด “well” [2]

ขั้นตอนต่อไปจะเป็นการประเมินจากอัตราความเร็วในการพูด 1 ประโยค (แกน x) และความดังของน้ำเสียง (แกน y) จากภาพประกอบที่ 1-5 จุด a คือ จุดที่ประเมินได้จากผู้ป่วย เมื่อลากเส้นจากจุดกำเนิด (origin) มายังจุด a จะได้เวกเตอร์ซึ่งลากผ่านโซน Dysarthric จะเข้าข่ายผู้ป่วย Dysarthria ทันที

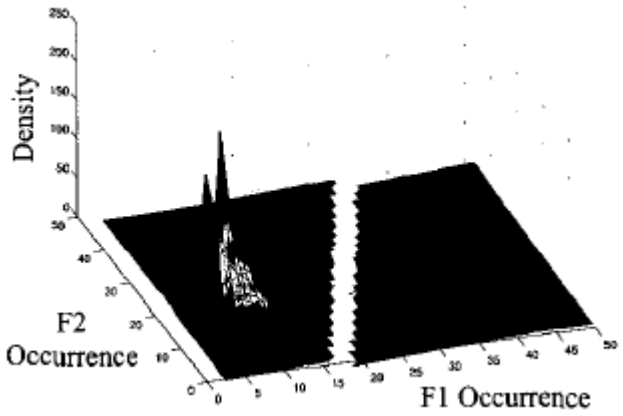


ภาพประกอบ 1-5 กราฟที่ใช้ประเมินผู้ป่วย [2]

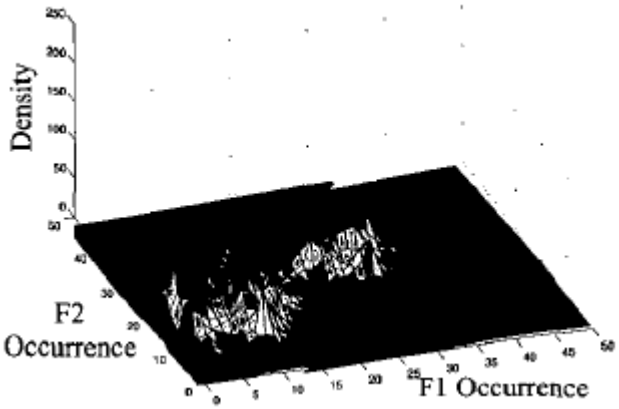
เมื่อเข้าข่ายผู้ป่วย Dysarthria แล้ว ก็จะนำสัญญาณเสียงของผู้ป่วยมาเช็คความถี่ฟอร์แมนต์โดยอ้างอิงจากฐานข้อมูลเสียงพูดปกติ จากภาพประกอบที่ 1-6 ถึง 1-8 เป็นการแสดงกราฟใน 3 มิติ โดยแกน x เป็นความถี่ฟอร์แมนต์ที่หนึ่ง แกน y เป็นความถี่ฟอร์แมนต์ที่สองและแกน z เป็นความเข้มของความถี่ (Density) โดยแพทย์ผู้เชี่ยวชาญจะทำหน้าที่สังเกตค่าความถี่ที่ปรากฏขึ้นของผู้ป่วยว่าความถี่ที่ปรากฏนั้นมีการซ้อนทับกับฐานข้อมูลหรือไม่ ถ้าย่านความถี่ที่ปรากฏหลุดออกนอกขอบเขตของฐานข้อมูลแสดงว่าบุคคลนั้นมีความผิดปกติทางการพูดชนิด Dysarthria



ภาพประกอบ 1-6 ขอบเขตความถี่ฟอร์แมนต์ของฐานข้อมูล (ผู้พูดปกติ) [2]



ภาพประกอบ 1-7 แสดงผลการทดสอบความถี่ฟอร์แมนต์จำนวน 1 คำ ของผู้ป่วย [2]



ภาพประกอบ 1-8 แสดงผลการทดสอบความถี่ฟอร์แมนต์ทั้งประโยคของผู้ป่วย [2]

1.2.3 Experiments with Fast Fourier Transform, Linear Predictive and Cepstral Coefficients in Dysarthric Speech Recognition Algorithms Using Hidden Markov Model [3]

งานวิจัยนี้เป็นการนำเสนอวิธีการออกแบบการรู้จำเสียงพูดสำหรับผู้พูดชนิด Dysarthric speech เปรียบเทียบระหว่างเทคนิคการหาค่าพารามิเตอร์ทั้ง 3 รูปแบบ มีรายละเอียดดังนี้

เทคนิคที่ใช้ประกอบด้วย

- Fast Fourier Transform (FFT)
- Linear Predictive Coefficients (LPC)
- Mel Frequency Cepstral Coefficients (MFCC)

ระบบการรู้จำเป็นแบบชนิด ขึ้นกับผู้พูด (Speaker Dependent), กลุ่มผู้ทดสอบเป็นผู้ป่วยโรคหลอดเลือดสมองเพศชาย 3 คนที่มีความผิดปกติทางการพูดชนิด Dysarthria

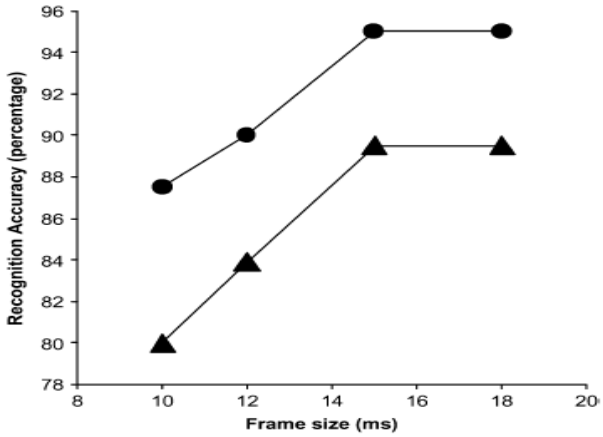
ในการทดสอบการอัดเสียงจะประกอบด้วยคำสั่ง 10 คำสั่ง, ตัวเลข 10 ตัว (1-10) ความถี่ปฏิบัติงานของเสียงจะอยู่ที่ 100-5500 Hz ค่าความถี่สุ่มตัวอย่าง อยู่ที่ 11.025 kHz 24 bit กรอบหน้าต่างเป็นแบบชนิด Hamming ใช้ร่วมกับโปรแกรม Matlab โดยใช้วิธีการฮิดเดนมาคอฟ (Hidden Markov Model : HMM) ชนิดไวเทอบี (Viterbi algorithm) เป็นตัวตัดสินคำสั่ง

ในแต่ละกรอบสัญญาณเสียงจะประกอบด้วยการคำณวนแบบ FFT 256 จุด สำหรับวิธีแรก ส่วนวิธีที่สองและสามจะเป็นแบบ LPC 15-coefficients และ MFCC 15-coefficients ตามลำดับ

ตารางที่ 1-2 คำสั่งที่ใช้ในการทดลอง

Utterance	**Digits**	**Words**
1	One	Back
2	Two	Front
3	Three	Go
4	Four	Good
5	Five	Is
6	Six	Left
7	Seven	My
8	Eight	Name
9	Nine	No
10	Ten	Ok
11		Right
12		Start
13		Stop
14		Very
15		Yes

ขั้นตอนถัดมาจะเป็นการทดลองหาขนาดของกรอบหน้าต่างที่เหมาะสมเพื่อทำให้อัตราการรู้จำเสียงมีประสิทธิภาพสูงที่สุด ในการทดลองนี้เลือกใช้วิธี MFCC ร่วมกับการปรับเปลี่ยนขนาดของกรอบหน้าต่างชนิด Hamming ตั้งแต่ 10 มิลลิวินาที ถึง 18 มิลลิวินาที ได้ผลลัพธ์สรุปออกมาดังภาพประกอบที่ 1-9



ภาพประกอบ 1-9 อัตราการรู้จำเสียงเฉลี่ยต่อขนาดความกว้างของหน้าต่างเป็นมิลลิวินาที กำหนดให้ รูปสามเหลี่ยม คือ เสียงคำพูดและรูปวงกลม คือ เสียงตัวเลข [3]

หลังจากได้ขนาดความกว้างของหน้าต่างที่เหมาะสมแล้ว จะนำเทคนิคทั้งสามมาทำการทดสอบการรู้จำร่วมกับขนาดของความกว้างของหน้าต่างเท่ากับ 15 มิลลิวินาที จะได้ผลสรุปแสดงดังภาพประกอบที่ 1-10

Dysarthric data	*LPC model*	*FFT model*	*MFCC model*
D1D	85%	90%	92.5%
D2D	90%	95%	97.5%
D3D	87.5%	95%	95%
Average DXD Recognition Accuracy	*87.5%*	*93%*	*95%*
D1W	66.7%	80%	86.7%
D2W	75%	86.7%	91.7%
D3W	73.3%	85%	90%
Average DXW Recognition Accuracy	*71.6%*	*83.9%*	*89.4%*
Overall Recognition Accuracy (for 25 utterances)	***79.5%***	***89%***	***92%***

ภาพประกอบ 1-10 ผลสรุปจากการทดลองโดยใช้เทคนิคทั้ง 3 วิธี (LPC, FFT & MFCC Model) กำหนดให้ DXD คือ เสียงตัวเลขและ DXW คือ เสียงคำพูด [3]

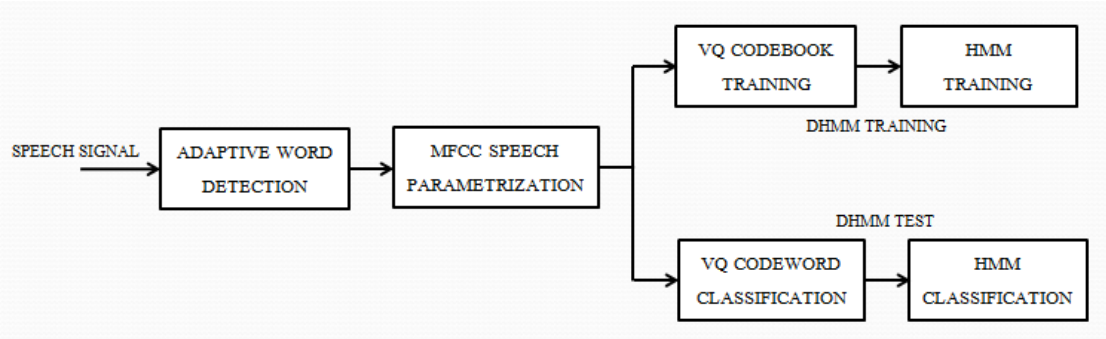
จากภาพประกอบที่ 1- 10 แสดงให้เห็นว่าการใช้ความกว้างของหน้าต่างขนาด 15 มิลลิวินาที ร่วมกับเทคนิค MFCC จะให้อัตราการรู้จำดีที่สุด คือร้อยละ 92 ถัดมา คือ เทคนิค FFT ร้อยละ 89 และเทคนิค LPC ร้อยละ 79 ตามลำดับ

1.2.4 Optimization of Dysarthric Speech Recognition [4]

งานวิจัยนี้ได้นำเสนอวิธีการสร้างตัวรู้จำเสียงพูด ASR (Automatic Speech Recognition) สำหรับผู้พูดชนิด Dysarthric Speech ซึ่งมีรายละเอียดดังนี้

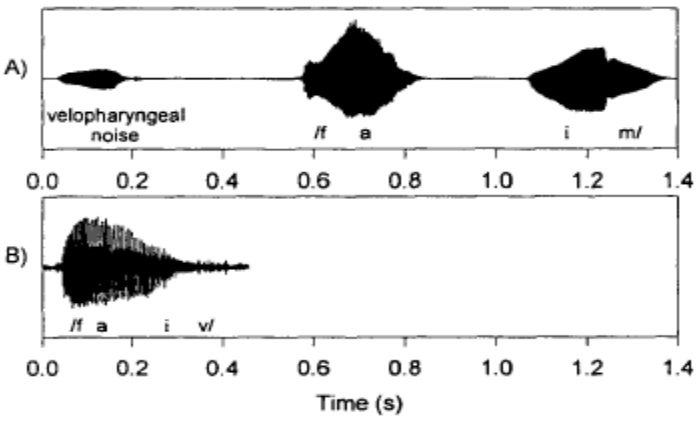
- การรู้จำเป็นชนิดขึ้นกับผู้พูด (Dependent System)
- พารามิเตอร์ที่ใช้ คือ Mel Frequency Cepstral Coefficients (MFCC)
- แบบจำลองที่ใช้ในการรู้จำ คือ Hidden Markov Model (Viterbi algorithm)
- กลุ่มผู้ทดสอบเป็นผู้ป่วยโรคหลอดเลือดสมองที่มีความผิดปกติทางการพูดชนิด Dysarthria เพศหญิงจำนวน 1 คนอายุ 39 ปี
- ขอบเขตของคำสั่งเป็นตัวเลขในภาษาอังกฤษ 0-9

หลักการทำงาน

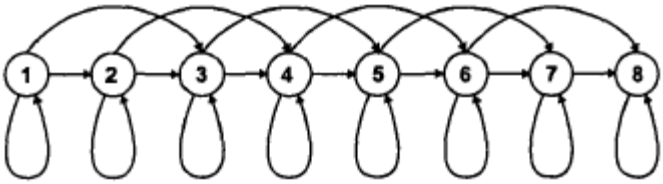


ภาพประกอบ 1-11 Automatic Speech Recognition Systems. [4]

เริ่มต้นที่ Adaptive Word Detection จะทำหน้าที่รับสัญญาณเสียงและทำการตัดสัญญาณเฉพาะช่วงที่เป็นสัญญาณเสียงพูดมาวิเคราะห์หลังจากนั้นสัญญาณจะถูกส่งต่อมายัง MFCC Parameterization เพื่อทำการสกัดพารามิเตอร์ของสัญญาณเสียง จากนั้นการทำงานจะแบ่งออกเป็น 2 โหมด คือ โหมดสอน (Training mode) ร่วมกับฮิดเดนมาคอฟและโหมดทดสอบ (Test mode) ร่วมกับฮิดเดนมาคอฟ โดยทั้ง 2 โหมด จะทำการรู้จำในขั้นตอนสุดท้าย จากภาพประกอบที่ 1-12 แสดงตัวอย่างสัญญาณเสียงพูดคำว่า “five” ของผู้ป่วย และภาพประกอบที่ 1-13 แสดงรูปแบบของแบบจำลองฮิดเดนมาคอฟที่ใช้ในงานวิจัยนี้



ภาพประกอบ 1-12 กราฟสัญญาณเสียงพูดคำว่า “five” (a) ของผู้ป่วย (b) ของผู้พูดปกติ [4]



ภาพประกอบ 1-13 left to right DHMM. [4]

การทดลองจะให้ผู้ป่วยเปล่งเสียงตัวเลขภาษาอังกฤษ (one - ten) ทั้งหมด 4 รอบ (A, B, C, D) ค่าความถี่ในการแซมปลิ้งจะอยู่ที่ 10kHz 16bit โดยใช้ไมโครโฟนที่ตอบสนองในย่านความถี่ 100 – 5000 Hz ส่วนในการทดสอบจะใช้ข้อมูลเดิมทั้ง 4 ชุด (A, B, C, D) มาทำการสอนระบบและทดสอบโดยใช้จำนวนการสอนจากน้อยไปหามาก คือ 1 ชุด, 2 ชุด และ 3 ชุดตามลำดับ โดยผลที่ได้คือความแม่นยำจะเพิ่มขึ้นเรื่อยๆตามจำนวนครั้งในการสอนระบบดังตารางที่ 1-3

ตารางที่ 1-3 ผลการทดสอบระบบ

Train sessions	Test Sessions	Recognition Rate
A	B, C, D	83%
B	A, C, D	83%
C	A, B, D	73%
D	A, B, C	84%
	Average	**81%**
A, B	C, D	88%
A, C	B, D	87%
A, D	B, C	85%
B, C	A, D	93%
B, D	A, C	91%
C, D	A, B	88%
	Average	**89%**

A, B, C	D	94%
A, B, D	C	94%
B ,C, D	A	88%
A, C, D	B	90%
	Average	**92%**

จากงานวิจัยนี้แสดงให้เห็นว่าระบบสามารถใช้กับผู้ป่วยชนิด Dysarthria ได้ ซึ่งจะให้ความแม่นยำในการรู้จำค่อนข้างสูงคือร้อยละ 92 แต่ระบบต้องเป็นชนิดขึ้นกับผู้พูดและจะต้องมีการสอนระบบทุกครั้งที่ใช้ ถ้าเราสามารถสอนระบบให้มากยิ่งขึ้น ระบบจะเพิ่มความแม่นยำและเหมาะสมกับผู้ป่วยคนนั้นมากขึ้น

1.2.5 การศึกษาการรู้จำเสียงพูดตัวเลขภาษาไทยแบบแยกคำชนิดไม่ขึ้นกับผู้พูดโดยใช้โครงข่ายประสาทที่มีการเรียนรู้แบบแพร่กลับ [5]

งานวิจัยนี้ได้นำเสนอการศึกษาการรู้จำสัญญาณเสียงพูดตัวเลข 0-9 แบบแยกคำชนิดไม่ขึ้นกับผู้พูดโดยใช้โครงข่ายประสาทที่มีการเรียนรู้แบบแพร่กลับ โดยเริ่มทำการศึกษาการรู้จำสัญญาณเสียงพูดตัวเลข 0-9 แบบแยกคำชนิดขึ้นกับผู้พูดก่อน เพื่อหาพารามิเตอร์ที่เป็นตัวแทนสัญญาณเสียงพูดที่เหมาะสมสำหรับการศึกษาในกรณีไม่ขึ้นกับผู้พูดต่อไป ทั้งนี้พารามิเตอร์ที่ใช้ในการศึกษาประกอบด้วย ค่าสัมประสิทธิ์พหุนามอันดับสองของคาบเวลาพิตช์ ค่าความหนาแน่นกลุ่มความถี่ฟอร์แมนต์ ค่าสัมประสิทธิ์การประมาณพันธะเชิงเส้น ค่าระยะเวลาในการออกเสียงและค่าผลรวมของความหนาแน่นของสเปกตรัมเชิงกำลังในแต่ละกลุ่มความถี่ จากการทดลองการศึกษาการรู้จำแบบขึ้นกับผู้พูดกับข้อมูลกลุ่มตัวอย่างและกลุ่มทดสอบกลุ่มละ 50 ตัวอย่างพบว่าพารามิเตอร์ที่เหมาะสมในการรู้จำ คือ ค่าสัมประสิทธิ์การประมาณพันธะเชิงเส้น ร่วมกับ ค่าผลรวมของความหนาแน่นของสเปกตรัมเชิงกำลังในแต่ละกลุ่มความถี่ โดยให้อัตราการรู้จำถูกต้องสูงถึงร้อยละ 98.0 และ 94.0 สำหรับข้อมูลกลุ่มตัวอย่างและกลุ่มทดสอบ ตามลำดับ ส่วนในการศึกษาการรู้จำแบบไม่ขึ้นกับผู้พูดของผู้ชายและผู้หญิงกับข้อมูลกลุ่มตัวอย่างและกลุ่มทดสอบเพศละ 300 และ 300 ตัวอย่าง ตามลำดับ พบว่าพารามิเตอร์ที่เหมาะสมในการรู้จำ คือ ค่าระยะเวลาในการออกเสียงร่วมกับ ค่าสัมประสิทธิ์พหุนามอันดับสองของคาบเวลาพิตช์ และค่าผลรวมของความหนาแน่นของสเปกตรัมเชิงกำลังในแต่ละกลุ่มความถี่ ซึ่งเมื่อใช้โครงข่ายประสาททำการรู้จำด้วยพารามิเตอร์เหล่านี้ร่วมกับการพิจารณาแยกสัญญาณเสียงพูดตัวเลข 5 และ 9 แล้วพบว่า สำหรับข้อมูลกลุ่มตัวอย่างและกลุ่มทดสอบของผู้ชายมีความถูกต้องร้อยละ 99.0 และ 85.3 ตามลำดับ และมีความถูกต้องร้อยละ 99.7 และ 82.0 สำหรับข้อมูลกลุ่มตัวอย่างและกลุ่มทดสอบของผู้หญิง ตามลำดับ

1.2.6 การแยกแยะคำสั่งควบคุมรถเข็นคนพิการในระบบรู้จำเสียงพูด [6]

งานวิจัยนี้ได้นำเสนอกระบวนการแยกแยะคำสั่งในระบบรู้จำเสียงพูดสำหรับควบคุมรถเข็นคนพิการ ในการเคลื่อนที่ไปในทิศทางต่างๆ โดยสามารถแยกแยะคำพูดที่มีความหมายเป็นคำสั่งจากกลุ่มคำพูด หรือประโยคที่เป็นคำสั่งหรือไม่ใช่ประโยคคำสั่ง แล้วแปลความหมายของคำสั่งนั้น ยกตัวอย่างเช่น เดินหน้า, เลี้ยวซ้าย, เลี้ยวขวา เป็นต้น ซึ่งในงานวิจัยนี้ได้แบ่งกระบวนการแยกแยะคำสั่งออกเป็นสองส่วน คือส่วนการแยกแยะประโยคคำสั่งและประโยคที่ไม่ใช่ประโยคคำสั่ง โดยใช้การตรวจสอบองค์ประกอบของประโยค และการตรวจสอบไวยากรณ์ เป็นการแยกแยะประโยคคำสั่ง ซึ่งในการค้นหาข้อมูลของคำและไวยากรณ์ในฐานข้อมูลคำที่เกี่ยวข้องกับคำสั่ง ได้ใช้วิธีการค้นหาแบบลิเนียร์ ส่วนที่สองคือ ส่วนของการแปลความหมายประโยคคำสั่งที่ถูกแยกแยะในส่วนที่หนึ่งให้เป็นความหมายของคำสั่งโดยใช้หลักความน่าจะเป็น เพื่อทำการหาค่าน้ำหนักของคำแต่ละคำที่มีความสัมพันธ์กับคำสั่งต่างๆโดยใช้ทฤษฎีของเบย์มาช่วยในการหาค่าน้ำหนักของคำ ในการทดลองระบบตามวิธีที่เสนอใช้ตัวอย่างทิศทางที่นำมาทดลองจำนวน 9 ทิศทาง โดยใช้คำพูด หรือประโยคคำสั่งจำนวน 548 ประโยค จากบุคคลจำนวน 42 คนในการแยกแยะความหมายของคำสั่งได้ผลความถูกต้องเฉลี่ยเป็นร้อยละ 88.06

1.2.7 การเปรียบเทียบระหว่างวิธีการ Pitch Extraction สองวิธีสำหรับการรู้จำเสียงวรรณยุกต์ภาษาไทยในสภาวะที่มีสัญญาณรบกวนขาว [7]

งานวิจัยนี้ได้เสนอวิธีการ Pitch Extraction ในการรู้จำหน่วยเสียงวรรณยุกต์สำหรับภาษาไทยในสภาวะแวดล้อมที่มีเสียงรบกวน โดยใช้ออโตคอร์รีเลชั่นที่มีขั้นตอนการคลิปยอดของสัญญาณ (Autocorrelation Method using Center Clipping: AUTOC) ในการคำนวณหาค่าคาบเวลาพิทช์ ซึ่งจะทำการคลิปยอดของสัญญาณในระดับที่ต่างกัน ค่าคาบเวลาพิตซ์ที่ได้จากการคลิปยอดของสัญญาณในระดับที่ต่างกันนี้ จะถูกแปลงเป็นค่าความถี่มูลฐานซึ่งเป็นตัวบ่งชี้ระดับสูงต่ำของเสียง ซึ่งลำดับของความถี่มูลฐานที่ได้จะถูกปรับปรุงให้มีความต่อเนื่องของข้อมูลโดยใช้มีเดียนฟิลเตอร์ จากนั้นทำการหาค่าการเปลี่ยนแปลงของความถี่มูลฐานของพิตซ์นั้นๆเทียบกับเวลาโดยทำการควอนไตซ์การเบี่ยงเบนออกเป็น 3 ระดับตามทิศทางการเพิ่มขึ้นหรือลดลงของความถี่มูลฐาน ซึ่งค่าที่ได้จากการควอนไตซ์นี้จะถูกนำไปใช้เป็นข้อมูลฝึกสอนให้กับการสร้างแบบจำลองแบบรู้จำของหน่วยเสียงวรรณยุกต์ทั้ง 5 ระดับด้วยวิธี Hidden Markov Model

งานวิจัยนี้ได้ทำการสอดแทรกสัญญาณรบกวนที่เป็นสัญญาณรบกวนเกาส์เซียนขาวเข้าไปในเสียงพูดวรรณยุกต์ในภาษาไทยพยางค์เดียวทั้ง 5 ระดับ เพื่อทดสอบความทนทานในการรู้จำเสียงวรรณยุกต์ในภาษาไทยต่อเสียงรบกวนและหาค่าการคลิปยอดของสัญญาณที่เหมาะสม

สำหรับการหาค่าคาบเวลาพิตซ์ในสภาวะที่มีเสียงรบกวน โดยใช้ข้อมูลเสียงที่ได้จากเพศชาย 5 คน และเพศหญิง 5 คน ผลที่ได้คือการคลิปยอดของสัญญาณที่ระดับ 30 เปอร์เซ็นต์ให้ผลการรู้จำเสียงวรรณยุกต์ภาษาไทยแม่นยำถึงร้อยละ 90.00 ในระดับสัญญาณต่อสัญญาณรบกวนที่ 10 dB

1.2.8 การรู้จำเสียงวรรณยุกต์ต่อเนื่องภาษาไทย ด้วยแบบจำลองฮิดเดน มาร์คอฟ 3 ส่วน [8]

งานวิจัยนี้ได้นำเสนอระบบการรู้จำเสียงวรรณยุกต์แบบต่อเนื่อง โดยแบ่งการทำงานออกมาเป็น 3 ส่วนใหญ่ๆ ได้แก่ ส่วนการสร้างเส้นโครงร่างระดับเสียง ส่วนการสร้างแบบจำลองเสียงวรรณยุกต์ และส่วนการตัดสินผลลัพธ์การรู้จำเมื่อพูดต่อเนื่อง การทำงานในแต่ละส่วนมีรายละเอียดดังนี้ ส่วนการสร้างเส้นโครงร่างระดับเสียง ในงานวิจัยนี้ เลือกใช้วิธีการแยกความถี่พื้นฐานตามแบบของพอลเบอร์สมา (Paul Boersma) คำนวณจากสัญญาณเสียงพูดที่ตัดแบ่งออกเป็นช่วงๆ ใช้การแปลงฟูเรียร์แปลงสัญญาณให้อยู่ในอาณาจักรของความถี่ เน้นค่าความถี่ที่มีพลังงานสูงให้เพิ่มมากขึ้น แปลงสัญญาณให้กลับมาอยู่ในอาณาจักรเวลา คำนวณหาค่าความถี่พื้นฐานจากสัญญาณที่ผ่านการเน้นความถี่ จะได้ลำดับการเปลี่ยนแปลงความถี่พื้นฐานของสัญญาณเสียงพูดและลดผลกระทบจากการแปรปรวนของผู้พูด ด้วยการคำนวณค่าคะแนนแซดและตัวกรองมัธยฐาน ส่วนการสร้างแบบจำลองเสียงวรรณยุกต์ เริ่มจากตัดแบ่งเส้นโครงร่างความถี่พื้นฐาน เป็น 3 ส่วน ได้แก่ ครึ่งพยางค์หน้า ครึ่งพยางค์หลัง และส่วนครึ่งพยางค์เชื่อมต่อ นำลำดับการเปลี่ยนแปลงความถี่พื้นฐานแต่ละส่วนไปฝึกฝนกับแบบจำลองฮิดเดน มาร์คอฟประเภทกึ่งต่อเนื่อง คุณลักษณะสำคัญที่ใช้ คือ ค่าความถี่พื้นฐาน ผลต่างของค่าความถี่พื้นฐานระหว่าง 2 จุด และค่าพลังงานของเสียง การฝึกฝนใช้วิธีการหยุดฝึกฝนก่อนที่จะจดจำ เพื่อเพิ่มประสิทธิภาพการรู้จำให้ดียิ่งขึ้น ส่วนตัดสินผลลัพธ์การรู้จำเมื่อพูดต่อเนื่อง ใช้ขั้นตอนวิธีแบบส่งผ่านโทเคน เพื่อหาลำดับของแบบจำลองที่ให้ค่าความน่าจะเป็นสูงสุด

ผลของการทดสอบประสิทธิภาพแบบจำลอง เมื่อใช้ข้อมูลฝึกฝนจากกลุ่มผู้พูดเป็นชาย 5 คนและหญิง 5 คน พูดคนละ 2 รอบ รอบละ 125 กลุ่มคำ กลุ่มละ 3 คำเรียงติดกัน ครอบคลุมทุกรูปแบบการเชื่อมต่อของเสียงวรรณยุกต์ทั้ง 5 รูปแบบ ข้อมูลเพื่อวัดประสิทธิภาพของแบบจำลองที่ผ่านการฝึกฝน จากกลุ่มผู้พูดชายและหญิงที่ไม่ซ้ำกับกลุ่มเดิม เป็นชาย 2 คนและหญิง 2 คน พูดคนละ 2 รอบ รอบละ 125 กลุ่มคำเช่นกัน แบบจำลองที่มีประสิทธิภาพดีที่สุด เป็นแบบจำลอง ฮิดเดน มาร์คอฟประเภทกึ่งต่อเนื่อง 4 สถานะ 5 กลุ่มเกาส์เซียน จำนวนทั้งหมด 35 แบบจำลอง แบ่งเป็นแบบจำลองส่วนครึ่งพยางค์หน้า 5 แบบจำลองแบบจำลองส่วนครึ่งพยางค์หลัง 5 แบบจำลอง และแบบจำลองส่วนครึ่งพยางค์เชื่อมต่อ 25 แบบจำลอง ให้ความถูกต้องร้อยละ 87.4

1.2.9 การรู้จำเสียงสระเสียงเดี่ยวในภาษาไทยโดยการใช้สเปกตรัมแอลพีซีบนสเกลบาร์คประสิทธิภาพ [9]

งานวิจัยนี้ได้นำเสนอวิธีในการรู้จำหน่วยเสียงสระเสียงเดี่ยวในภาษาไทยและหาวิธีการลดมิติของความเข้มแถบความถี่วิกฤติ (Critical Band Intensities: CBI) บนสเกลความเข้มแถบความถี่วิกฤติ เพื่อเพิ่มประสิทธิภาพในการรู้จำโดยลดความยุ่งยากซับซ้อน และเวลาในกระบวนการคำนวณลง หน่วยเสียงสระที่จะทำการรู้จำคือเสียง อิ, เอะ, แอะ, อึ, เออะ, อะ, อุ, โอะ และเอาะ แต่ละหน่วยเสียงจะถูกนำไปคำนวณสัมประสิทธิ์ของการประมาณเชิงเส้น แล้วทำการแปลงสเปกตรัมของการประมาณเชิงเส้นที่ได้ไปอยู่ในสเกลความเข้มแถบความถี่วิกฤติ ซึ่งจะได้เวกเตอร์คุณสมบัติของความเข้มแถบความถี่วิกฤติ (CBI) 18 มิติ นำมาหาความเข้มแถบความถี่วิกฤติที่มีประสิทธิภาพโดยวิธีการหาค่าสูงสุดของความถูกต้องในการรู้จำเพิ่มทีละมิติ ซึ่งเป็นวิธีการลดขนาดมิติของเวกเตอร์คุณสมบัติที่จะถูกนำไปใช้สร้างและทดสอบแบบอ้างอิงต่อไป การทดสอบแบบอ้างอิงจะอาศัยเทคนิค K-Nearest Neighbor (KNN) ในการตัดสินใจแยกแยะเสียงสระ

ในงานวิจัยนี้เราสามารถลดขนาดมิติที่ใช้ในการรู้จำในแบบอ้างอิงต่างๆ ได้ประมาณร้อยละ 52 โดยค่าร้อยละความถูกต้องของทุกแบบอ้างอิงจะสูงขึ้นเมื่อเทียบกับการรู้จำเสียงสระเดี่ยวในภาษาไทยแบบใช้ข้อมูลจากทุกมิติแบบเดิม

1.2.10 การเพิ่มประสิทธิภาพของกระบวนการตรวจจับค่าพิตช์ของเสียงพูดภาษาไทยเพื่อการจำแนกเพศบุคคล [10]

งานวิจัยนี้ได้นำเสนอการเพิ่มประสิทธิภาพการตรวจจับค่าพิตช์ของเสียงพูดภาษาไทยสำหรับการจำแนกเพศบุคคล ในงานวิจัยนี้ได้มุ่งเน้นไปยังขั้นตอนวิธีการประมวลผลสัญญาณเบื้องต้นก่อนกระบวนการตรวจจับค่าพิตช์ ทั้งนี้เนื่องจากประสิทธิภาพโดยรวมของการตรวจจับค่าพิตช์ขึ้นอยู่กับคุณภาพของสัญญาณนำเข้า ด้วยเหตุผลดังกล่าวขั้นตอนวิธีการประมวลผลสัญญาณเบื้องต้นด้วยวิธีการลบดีซีไบอัส (Removed-DC-Bias) ร่วมกับตัวกรองบัตเตอร์เวิร์ธซึ่งเป็นวิธีการที่ง่ายและมีประสิทธิภาพจึงถูกประยุกต์ใช้กับระบบการจำแนกเพศบุคคล เพื่อพิสูจน์ความสามารถของขั้นตอนวิธีดังกล่าว วิธีออโตคอรีเลชัน (Autocorrelation) และวิธีเซปสตรัม (Cepstrum) ซึ่งเป็นสองวิธีการที่มีความน่าเชื่อถือและมีประสิทธิภาพในการประมวลผลบนขอบเขตทางเวลาและขอบเขตทางความถี่ได้ถูกเลือกมาเป็นวิธีในการตรวจจับค่าพิตช์ จากนั้นได้นำตัวอย่างสัญญาณเสียงจำนวน 200 เสียงพูดโดยแบ่งออกเป็น 2 กลุ่ม คือ สัญญาณเสียงจำนวน 40 เสียงพูดสำหรับกระบวนการเรียนรู้ และสัญญาณเสียงพูด 160 เสียงพูดสำหรับกระบวนการทดสอบผลลัพธ์จากการทดลองพิสูจน์ให้เห็นว่าประสิทธิภาพของขั้นตอนวิธีที่นำเสนอเมื่อประยุกต์ใช้งานร่วมกับ

วิธีออโตคอรีเลชันสามารถจำแนกเพศบุคคลให้มีความถูกต้องโดยรวมเพิ่มขึ้นจากร้อยละ 92.94 เป็นร้อยละ 94.37 และความถูกต้องโดยรวมเพิ่มขึ้นจากร้อยละ 93.17 เป็นร้อยละ 94.84 เมื่อประยุกต์ใช้งานร่วมกับวิธีเซปตรัมได้แสดงให้เห็นว่าวิธีการลบดีซีไบอัสร่วมกับตัวกรองบัตเตอร์เวิร์ธเป็นวิธีการที่เหมาะสมสำหรับการเตรียมสัญญาณที่มีคุณภาพให้กับกระบวนการการตรวจจับค่าพิตช์ของเสียงพูดภาษาไทย

1.2.11 การรู้จำเสียงตัวเลขต่อเนื่องภาษาไทยแบบไม่ขึ้นกับบุคคลโดยใช้ ฮิดเดน มาร์คอฟ โมเดล [11]

งานวิจัยนี้ได้นำเสนอระบบรู้จำเสียงพูดตัวเลขต่อเนื่องภาษาไทยที่ไม่ขึ้นกับบุคคล ประกอบไปด้วย 5 ส่วน คือ 1) การประมวลผลสัญญาณเสียงพูดเบื้องต้นซึ่งทำหน้าที่ในการวางกรอบสัญญาณเพื่อเตรียมข้อมูลเสียงให้เหมาะสม 2) การดึงค่าลักษณะสำคัญของเสียง ซึ่งทำหน้าที่หาลักษณะสำคัญที่สามารถใช้เป็นตัวแทนของเสียงพูด 3) การฝึกฝนรูปแบบที่ทำหน้าที่ในการสอนระบบให้เรียนรู้ลักษณะสำคัญของเสียงพูดแต่ละตัวเลข 4) การจำแนกรูปแบบสำหรับวิเคราะห์เปรียบเทียบค่าคุณลักษณะสำคัญว่าเป็นตัวเลขใด และ 5) การหาคำตอบโดยหาลำดับที่ดีที่สุดเพื่อเป็นคำตอบ ในงานวิจัยนี้จะเน้น 3 ขั้นตอนคือ 2, 3 และ 4 ในส่วนของขั้นตอนที่ 2 การดึงค่าคุณลักษณะสำคัญของเสียงที่นำมาใช้ในระบบประกอบด้วย 6 คุณลักษณะ ได้แก่ ค่าสัมประสิทธิ์เซปตรัมบนสเกลเมล ผลต่างทางเวลาของสัมประสิทธิ์เซปตรัมบนสเกลเมลลำดับที่หนึ่ง ผลต่างทางเวลาของสัมประสิทธิ์เซปตรัมบนสเกลเมลลำดับที่สอง ค่าพลังงานเสียง ค่าผลต่างพลังงาน และ ค่าผลต่างพลังงานลำดับที่สอง ในขั้นตอนที่ 3 และ 4 การฝึกฝนรูปแบบและการจำแนกรูปแบบใช้ทฤษฎีแบบจำลอง ฮิดเดนมาร์คอฟแบบต่อเนื่อง ผลการทดลองประสิทธิภาพของระบบ ใช้ชุดข้อมูลฝึกฝนเสียงพูดจากผู้พูดจำนวน 100 คน แบ่งเป็น ชาย 50 คน และ หญิง 50 คน รวม 2000 เสียง แต่ละเสียงยาว 7 คำเรียงติดกัน ข้อมูลทดสอบประกอบด้วย 40 คน แบ่งเป็นชาย 20 คน และหญิง 20 คน พบว่าจำนวนคุณลักษณะที่ทำให้อัตรารู้จำสูงสุดในการทดลองนี้ คือใช้ 5 คุณลักษณะ (จำนวนรวม 41 ค่า) และแบบจำลองเสียงพูดที่ให้ประสิทธิภาพดีที่สุดคือ แบบจำลองที่มีจำนวน 9 สถานะ 8 กลุ่มเกาส์เซียน อัตราการรู้จำเสียงพูดเฉลี่ยเท่ากับร้อยละ 74.25

จากงานวิจัยรู้จำเสียงพูดในประเทศไทยทั้งหมดตั้งแต่หัวข้อที่ 1.2.5 ถึ หัวข้อที่ 1.2.11 สามารถสรุปรายละเอียดโดยเรียงลำดับตามจำนวนปี พ.ศ. ตั้งแต่ พ.ศ. 2544 ถึ งพ.ศ. 2550 แสดงดังตารางที่ 1-4

ตารางที่ 1-4 แสดงรายละเอียดงานวิจัยรู้จำเสียงพูดในประเทศไทยที่ผ่านมา

ปี พ.ศ.	หลักการ	กลุ่มผู้ทดสอบ	รูปแบบ	ความแม่นยำ
2544	สเปกตรัมเชิงความถี่และ โครงข่ายประสาทที่มีการเรียนรู้แบบแพร่กลับ	ชายและหญิง อย่างละ 300 เสียงพูด	ตัวเลขภาษาไทย 0-9 ไม่ขึ้นกับผู้พูด	85.30% และ 82.00%
2546	ทฤษฎีของเบย์	บุคคล 42 คน 548 เสียงพูด	ประโยคคำศัพท์ 9 คำ ขึ้นกับผู้พูด	88.06%
2547	Pitch Extraction และ ฮิดเดนมาร์คอฟโมเดล	ชาย 5 คน หญิง 5 คน	ประโยคคำศัพท์ 100 คำ ขึ้นกับผู้พูด	90.00%
2547	แยกความถี่แบบพอลเบอร์สมาและฮิดเดน มาร์คอฟประเภทกึ่งต่อเนื่อง (รู้จำแบบคำต่อเนื่อง)	ชาย 7 คน หญิง 7 คน	เสียงวรรณยุกต์ 125 กลุ่มคำ ไม่ขึ้นกับผู้พูด	87.40%
2548	สเปกตรัมแอลพีซีและ K-Nearest Neighbor (KNN)	ชาย 7 คน หญิง 7 คน	สระเดี่ยว 21 เสียง ไม่ขึ้นกับผู้พูด	90.00%
2550	ค่าสัมประสิทธิ์เซปตรัมบนสเกลเมลและฮิด เดนมาร์คอฟแบบต่อเนื่อง (รู้จำแบบคำต่อเนื่อง)	ชาย 20 คน หญิง 20 คน	ตัวเลขภาษาไทย 0-9 ไม่ขึ้นกับผู้พูด	74.25%

1.3 วัตถุประสงค์

1.3.1 เพื่อออกแบบระบบสั่งการสำหรับอำนวยความสะดวกแก่ผู้ป่วยโรคหลอดเลือดสมองโดยใช้เสียงพูด

1.3.2 เพื่อศึกษาวิธีการรู้จำเสียงพูดและวิธีการหาค่าพารามิเตอร์สำหรับใช้ในการจำแนกสัญญาณเสียงพูดแบบแยกคำและขึ้นกับผู้พูดที่มีความผิดปกติทางการพูดชนิด Dysarthria

1.4 ขอบเขตการวิจัย

1.4.1 ออกแบบคำสั่งโดยใช้เสียงสระในภาษาไทยแบ่งออกเป็นสระเดี่ยวเสียงสั้น (อึ เออะ อะ) และสระเดี่ยวเสียงยาว (อือ เออ อา)

1.4.2 ใช้กับกลุ่มผู้ป่วยโรคหลอดเลือดสมองที่มีความผิดปกติทางการพูดประเภท Dysarthria

1.4.3 ระบบสั่งการเป็นแบบชนิดขึ้นกับผู้พูด

1.5 ประโยชน์ที่ได้รับ

1.5.1 อำนวยความสะดวกให้แก่ผู้ป่วยโรคหลอดเลือดสมองในการสั่งงานอุปกรณ์โดยใช้เสียงพูด

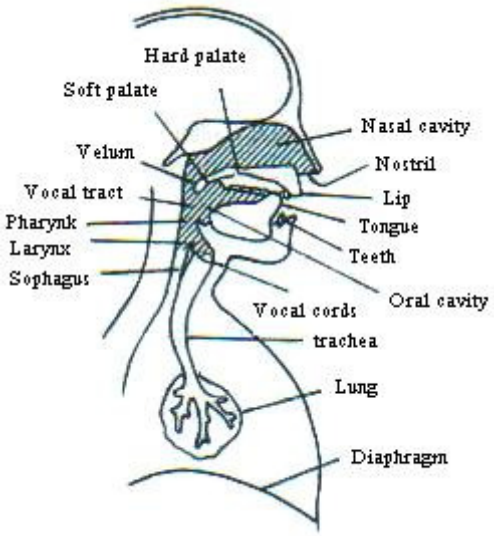
1.5.2 ประยุกต์ใช้กับสิ่งอำนวยความสะดวกอื่นๆสำหรับผู้ป่วยโรคหลอดเลือดสมองในการสั่งงานโดยใช้เสียงพูด

บทที่ 2

ทฤษฎีและหลักการ

ในการจัดทำวิทยานิพนธ์เรื่องการออกแบบระบบสั่งการแบบรู้จำเสียงพูดชนิดขึ้นกับผู้พูดสำหรับผู้ป่วยโรคหลอดเลือดสมองนั้น ผู้วิจัยได้แบ่งหลักการและทฤษฎีต่างๆที่เกี่ยวข้องกับงานวิจัยออกเป็นกลุ่มๆดังต่อไปนี้ 1) กระบวนการทำให้เกิดเสียงพูด 2) เสียงสระในภาษาไทย 3) การพูดผิดปกติเนื่องจากความผิดปกติของระบบประสาทที่ควบคุมการเคลื่อนไหว 4) การรู้จำเสียงคำพูด และ 5) การประมวลผลสัญญาณเบื้องต้น 6) การสกัดค่าลักษณะสำคัญ

2.1 กระบวนการทำให้เกิดเสียงพูด (Speech Production)



ภาพประกอบ 2-1 อวัยวะภายในของกระบวนการทำให้เกิดเสียงพูดของมนุษย์ [5]

กระบวนการทำให้เกิดเสียงพูดของมนุษย์เริ่มต้นจากลมที่ขับออกมาจากปอด (Lung) ผ่านทางหลอดลม (Trachea) และช่องเส้นเสียง (Glottis) ไปยังช่องทางเดินเสียง (Vocal Tract) ซึ่งประกอบด้วยช่องคอ (Larynx) ช่องปาก (Oral Cavity) และโพรงจมูก (Nasal Cavity) ช่องเส้นเสียงดังกล่าวข้างต้น ก็คือช่องว่างระหว่างเส้นเสียง (Vocal Cords) ซึ่งมีลักษณะเป็นกล้ามเนื้อ 2 แผ่นปิดขวางอยู่ที่ช่องหลอดลม ลมที่ผ่านช่องเส้นเสียงนี้จะทำให้เกิดการสั่นของเส้นเสียงขึ้น และจากการสั่นนี้จะทำให้เกิดคลื่นเสียงความถี่ต่างๆ ผ่านเข้าสู่ลำคอ เข้าสู่ช่องปากหรือโพรงจมูกต่อไป

การเปลี่ยนแปลงรูปร่างและตำแหน่งของกล้ามเนื้อและอวัยวะภายในช่องทางเดินเสียง เช่น ช่องปาก, ขากรรไกร (Jaw), ลิ้น (Tongue) , เพดานแข็ง (Hard Palate), เพดานอ่อน (Soft Palate) และริมฝีปาก (Lip) ทำให้มนุษย์สามารถเปล่งเสียงออกมาเป็นคำพูดในลักษณะต่างๆกันได้ อวัยวะภายในของกระบวนการทำให้เกิดเสียงพูดของมนุษย์แสดงดังภาพประกอบ 2-1 เสียงที่เกิดขึ้นจากกระบวนการดังกล่าว สามารถแบ่งออกได้เป็น 2 ประเภท คือ

1. เสียงก้อง (Voiced Sounds) คือ เสียงที่ขณะเปล่งออกมามีการสั่นของเส้นเสียงด้วย ได้แก่ เสียงของสระและพยัญชนะบางตัว เช่น ‘ด’, ‘น’, ‘ว’ และ ‘ร’ ฯลฯ
2. เสียงไม่ก้อง (Unvoiced Sounds) คือ เสียงที่ขณะเปล่งออกมาไม่มีการสั่นของเส้นเสียง เช่น คำว่า ‘ปะ’, ‘จะ’, ‘ชะ’ ‘พะ’, ‘ฟะ’ และ ‘สะ’ เป็นต้น

จากกระบวนการทำให้เกิดเสียงพูดข้างต้น อาจกล่าวได้ว่า ประกอบด้วยขั้นตอนใหญ่ 3 ขั้นตอนด้วยกัน คือ

1. การผลิตเสียงจากแหล่งกำเนิดเสียง (Sound Source Production)
2. การเปลี่ยนแปลงรูปร่างของช่องทางเดินเสียง (Articulation by Vocal Tract)
3. การแพร่เสียงออกจากปากหรือโพรงจมูก (Radiation from the Lip and/or Nostrils)

เสียงพูดของมนุษย์ที่เปล่งออกมาและมนุษย์ได้ยินจะมีคุณสมบัติดังต่อไปนี้

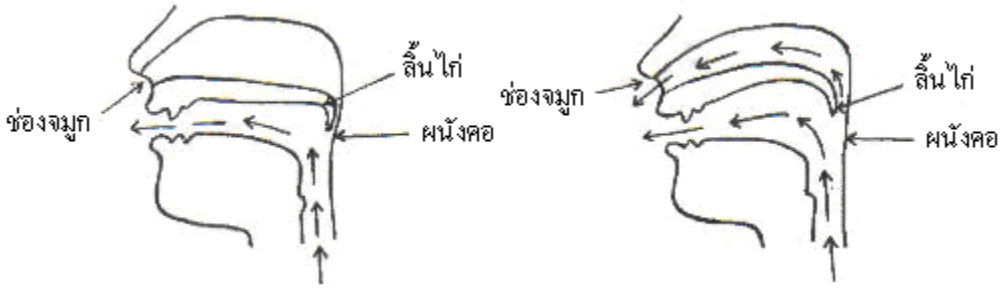
1. ความยาวของเสียง (Length) เป็นช่วงเวลาของการเกิดเสียงใดเสียงหนึ่งเปล่งออกมา เสียงพูดบางเสียงอาจจะเปล่งออกมาได้ติดต่อกันได้นาน เช่น เสียงสระเสียงพยัญชนะที่เป็นเสียงสอดแทรก เป็นต้น สำหรับในภาษาไทย เสียงพูดที่มีความยาว-สั้น ก็มีเพียงเสียงสระเท่านั้น เช่น อะ อิ อุ เป็นเสียงสั้น อา อี อู เป็นเสียงยาว เป็นต้น
2. ระดับเสียงสูง-ต่ำ (Pitch) เสียงพูดจะมีระดับเสียงสูงหรือต่ำ อยู่ที่ความถี่มูลฐานของเสียง (Fundamental frequency) ถ้าความถี่ต่ำเสียงจะต่ำ อวัยวะส่วนที่ทำให้เสียงมีระดับ สูง-ต่ำ คือ เส้นเสียง ดังนั้น ระดับเสียงสูง-ต่ำ ก็คือ อัตราการสั่นสะบัดของเส้นเสียงนั่นเอง ในการพูดเสียงที่มีระดับสูง-ต่ำ ได้คือเสียงก้องเท่านั้น เพราะมีการสั่นสะเทือนของเส้นเสียง ที่ทำให้เกิดมีความถี่ระดับต่างๆได้ ในภาษาไทยระดับเสียงสูง-ต่ำ ของคำ เราเรียกว่า วรรณยุกต์
3. ความดัง (Loudness) ความดังขึ้นอยู่กับปริมาณของลม ที่ผู้พูดเปล่งเสียงออกมาในช่วงเวลาหนึ่งๆ
4. การลงน้ำหนัก (Stress) เป็นการออกเสียงพยางค์ใดพยางค์หนึ่งให้ดังเน้นมาก

หรือน้อยกว่าพยางค์อื่นๆที่อยู่ข้างเคียง (แสดงอารมณ์อย่างใดอย่างหนึ่ง)

5. ช่วงต่อของเสียง (Juncture) ช่วงระยะที่ผู้พูดเปล่งเสียงหนึ่งแล้วต่อไปเปล่งเสียงอีกเสียง ซึ่งเรียงกันมาเป็นลำดับเสียงที่ประกอบกันเข้าเป็นพยางค์ จะมีช่วงต่อของเสียงแนบสนิทจนไม่เห็นร่องรอย (Close juncture) แต่ถ้าเสียงปรากฏอยู่คนละพยางค์ หรือคนละคำจะมีช่วงต่อหางจนสังเกตได้ชัด (Open Juncture) ดังนั้นช่วงต่อของเสียง โดยเฉพาะช่วงต่อหางจะเป็นลักษณะการแบ่งวรรคตอนของเสียงพูด

2.2 เสียงสระในภาษาไทย

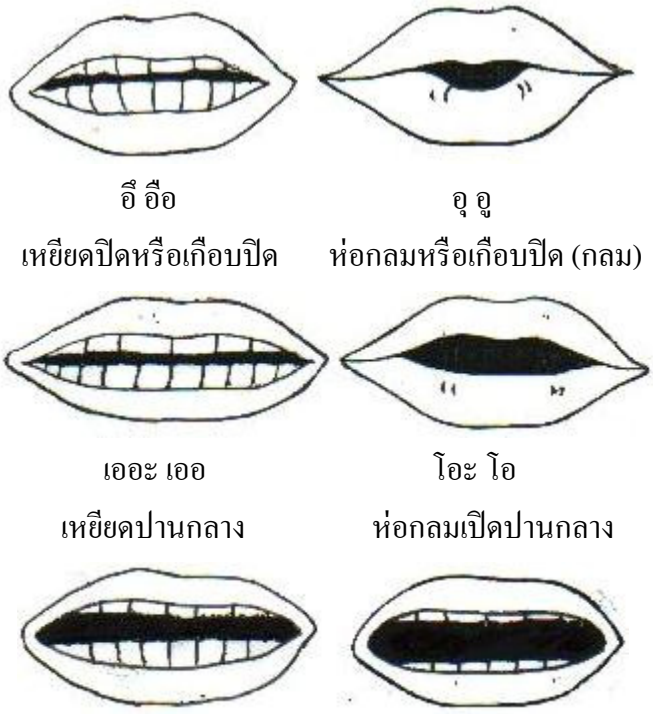
เสียงสระในภาษาไทยที่ใช้พูดกันมีอยู่ 3 อย่าง คือ เสียงแท้ เสียงแปรและเสียงดนตรี โดยเสียงแท้คือเสียงที่เปล่งออกมาจากลำคอโดยตรง ไม่ผ่านการดัดแปลงจากอวัยวะอื่นๆ ให้ผิดเพี้ยนไป เสียงแปรคือเสียงที่เปล่งออกมาจากลำคอแล้วกระดกลิ้นให้กระทบกับอวัยวะอื่นๆ ในช่องปาก ทำให้เสียงแท้เปลี่ยนเป็นเสียงต่างๆ และเ สียงดนตรีเป็นเสียงแท้หรือเสียงแปรที่ผู้เปล่งเสียงออกเสียงให้สูงต่ำเช่นเดียวกับเสียงดนตรีที่มีช่วงของเสียงสูงต่ำต่างกัน จากภาพ ประกอบที่ 2-2 เป็นการแสดงถึงการเดินทางของกระแสลมจากปอดที่ทำให้เกิดเสียงในลักษณะต่างๆ



ภาพประกอบ 2-2 การเดินทางของกระแสลมจากปอดที่ทำให้เกิดเสียง [15]

เสียงสระจัดเป็นเสียงแท้ กล่าวคือเป็นเสียงที่เปล่งออกจากลำคอเป็นเสียงเดียวไม่ถูกปิดกั้นทางลม (แต่อาจมีการกล่อมเกลาเสียงโดยอวัยวะอื่นๆ เช่น ริมฝีปากและลิ้น แต่ไม่มากนัก) ในขณะที่ลมผ่าน เส้นเสียงจะเกร็งตัวสั่ นสะบัดทำให้เกิดความกังวาน เป็นผลให้เสียงสระทุกเสียงจัดเป็นเสียงแบบโฆษะ (เสียงก้อง) ลิ้นและริมฝีปากทำให้เกิดเสียงสระที่แตกต่างกันหลายเสียง โดยที่ลิ้นมีส่วนเกี่ยวข้องในสองลักษณะคือส่วนของลิ้นที่มีผลต่อเสียงสระและระดับความสูงของลิ้นส่วนดังกล่าว ส่วนของลิ้นแบ่ งเป็นลิ้นส่วนหน้า (Blade) ลิ้นส่วนกลาง (Centre) และลิ้นส่วนหลัง (Dorsum) หากมีการกระดกลิ้นส่วนหน้าจะทำให้เกิดสระหน้า กระดกลิ้นส่วนกลางทำให้เกิดสระ

กลางและกระดกลิ้นส่วนหลังทำให้เกิดสระหลัง สำหรับระดับความสูงของลิ้นหากลิ้นส่วนหนึ่งๆ กระดกขึ้นในระดับสูงจะทำให้เกิดสระระดับสูง ทำนองเดียวกันกับการกระดกลิ้นในระดับกลางและต่ำจะทำให้เกิดสระระดับกลางและสระระดับต่ำ นอกจากนี้ เสียงสระที่แตกต่างกันนั้นมีผลมาจากริมฝีปากด้วย กล่าวคือการเปลี่ยนแปลงรูปริมฝีปากไปในลักษณะต่างๆ เช่น ห่อกลมเปิดกว้าง เหยียดกึ่งปิด จะส่งผลโดยตรงต่อ ส่วนของลิ้นและระดับความสูงของลิ้นด้วยแสดงดังภาพประกอบที่ 2-3



ภาพประกอบ 2-3 ลักษณะของริมฝีปากในการเปล่งเสียงสระ

2.2.1 ลักษณะที่สำคัญของเสียงสระ

1. เป็นเสียงที่ลมผ่านออกมาได้สะดวกโดยไม่ถูกอวัยวะในปากกักทางลม
2. ส่วนของลิ้นที่โค้งขึ้นและระดับความสูงของลิ้นส่วนนั้นๆทำให้เกิดเสียงสระที่ต่างกัน
3. เสียงสระออกเสียงได้ยาวนาน
4. เสียงสระทุกเสียงเป็นเสียงก้อง
5. เสียงสระมีทั้งเสียงสั้นและเสียงยาว

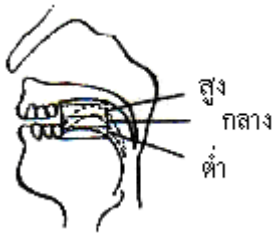
2.2.2 การจำแนกเสียงสระในภาษาไทย

รูปสระมี 21 รูป เสียงสระมี 24 เสียง โดยการจำแนกเสียงตามช่วงเวลาในการเปล่งเสียงจำแนกได้สองกลุ่ม คือกลุ่มที่ออกเสียงเป็นสระสั้นและกลุ่มที่ออกเสียงเป็นสระยาว นอกจากนี้ยังจำแนกเสียงตามลักษณะการเกิดเสียงได้ 3 กลุ่ม คือ

1. สระเดี่ยวหรือสระแท้ เป็นเสียงสระที่เปล่งออกมาจากลำคอแล้วถูกลิ้นหรือริมฝีปากเกลาเสียง มีทั้งเสียงสั้นและเสียงยาวเป็นคู่จำนวน 18 เสียง สระแท้ยังแบ่งออกเป็น 2 พวก คือ สระที่เปล่งออกมากระทบกับอวัยวะออกเสียงในช่องปากเพียงแห่งเดียวมี 8 เสียงประกอบด้วย อะ อา อิ อี อึ อือ อุ อู และสระแท้ทั้งสองฐาน คือ สระที่เปล่งเสียงออกมากระทบกับอวัยวะออกเสียงในช่องปากสองแห่งมี 10 เสียงประกอบด้วย เอะ เอ แอะ แอ โอะ โอ เอาะ ออ เออะ เออ
2. สระประสมหรือเรียกว่าสระเลื่อน เป็นสระที่ออกเสียงสระแท้ทั้งสองเสียงประสมกัน มี 6 เสียง คือ เอียะ เอีย เอือะ เอือ อัวะ และ อัว ทั้งนี้นักภาษาศาสตร์บางกลุ่มถือว่าสระประสมมีเพียง 3 เสียง คือ เอีย เอือ อัว เนื่องจากคำที่มีเสียงสระประสมเสียงสั้นมีใช้น้อย ส่วนมากยืมมาจากภาษาอื่นหรือเป็นคำเลียนเสียง
3. สระเกิน เป็นสระที่เกินความจำเป็นโดยทางภาษาศาสตร์ไม่ถือว่าเป็นสระ เนื่องจากมีเสียงพยัญชนะประสมอยู่กับสระด้วย ได้แก่ อำ ไอ ใอ เอา ฤ และ ฤา จากการจำแนกเสียงสระแท้ข้างต้นสามารถแบ่งตามการเปลี่ยนแปลงของลิ้นและริมฝีปากได้ดังตารางที่ 2-1

ตารางที่ 2-1 สระแท้ที่แบ่งตามการเปลี่ยนแปลงของลิ้นและริมฝีปาก

ลักษณะริมฝีปาก	รี		ปกติ		กลม	
ลักษณะเสียง	เสียงสั้น	เสียงยาว	เสียงสั้น	เสียงยาว	เสียงสั้น	เสียงยาว
ลิ้นอยู่ระดับสูง	อิ	อี	อึ	อือ	อุ	อู
ลิ้นอยู่ระดับกลาง	เอะ	เอ	เออะ	เออ	โอะ	โอ
ลิ้นอยู่ระดับต่ำ	แอะ	แอ	อะ	อา	เอาะ	ออ



ภาพประกอบ 2-4 ระดับความสูงของลิ้นที่มีความสัมพันธ์ต่อเสียงสระต่างๆ [15]

จากตารางที่ 2-1 ลักษณะริมฝีปากที่ใช้ในการออกเสียงสระแบ่งได้ 3 แบบ คือ รี ปกติ และ กลม ซึ่งการปรับริมฝีปากให้เป็นวงรีและกลมนั้น ผู้ป่วยจะทำไม่ค่อยได้ สาเหตุ อันเนื่องมาจากอาการอัมพาตบริเวณใบหน้า ดังนั้นเพื่อให้ผู้ป่วยสามารถสั่งการได้ง่ายขึ้น จึงพิจารณาเฉพาะคำไม่ต้องเปลี่ยนแปลงลักษณะของปากมากจนเกินไป งานวิจัยนี้ จึงใช้เสียงสระทั้ง 6 คำ ได้แก่ อึ อือ เออะ เออ อะ และ อา สำหรับออกแบบระบบสั่งการเพื่อให้เกิดความเหม าะสมกับผู้ป่วย

2.3 การพูดผิดปกติเนื่องจากความผิดปกติของระบบประสาทที่ควบคุมการเคลื่อนไหว (Motor Speech Disorders)

การพูดผิดปกติเนื่องจากความผิดปกติของระบบประสาทที่ควบคุมการเคลื่อนไหว แบ่งออกเป็น 2 ประเภทใหญ่ๆ คือ Dysarthria และ Apraxia of speech

2.3.1 Dysarthria

เป็นการพูดผิดปกติที่เกิดจากความผิดปกติของระบบประสาทที่ควบคุมการทำงานของอวัยวะที่ใช้ในการพูด ซึ่งอาจเป็นความผิดปกติที่ Central หรือ Peripheral Nervous System ก็ได้ ทำให้กล้ามเนื้อที่เกี่ยวข้องกับการพูดเป็นอัมพาต อ่อนแรง รับความรู้สึกได้ไม่ดี หรื อมีการทำงานประสานงานกันไม่ดี จึงทำให้ผู้ป่วยมีความผิดปกติของการหายใจ การออกเสียง (Phonation) การเปล่งเสียงพูด (Articulation) การกำทอนของเสียง (Resonation) และมีลีลาจังหวะการพูด (Prosody) ผิดปกติ นอกจากนั้นผู้ป่ วยอาจมีความผิดปกติของการเคลื่อนไหวแบบอัตโนมัติของกล้ามเนื้อที่เกี่ยวข้องกับการพูด ทำให้เกิดมีความผิดปกติของการเคี้ยว กลืน การเคลื่อนไหวของลิ้นและขากรรไกรด้วย

การจัดแบ่งประเภทของ Dysarthria นั้นมีหลายแบบ เช่น แบ่งตามสาเหตุของโรคที่ทำให้ระบบประสาททำงานบกพร่องไป บางคนแบ่งตามตำแหน่งที่เกิดพยาธิสภาพหรือแบ่งตาม

ลักษณะการพูดของผู้ป่วย โดยวิเคราะห์จากการฟังเสียงพูดของผู้ป่วย ซึ่งเป็นการแบ่งประเภทของ Dysarthria ตามลักษณะการพูดของผู้ป่วย และพบว่ามีความสัมพันธ์กับสาเหตุของโรคและตำแหน่งพยาธิสภาพด้วย ผู้ทำการศึกษาได้แบ่ง Dysarthria ออกเป็น 6 ชนิด คือ

1. Flaccid Dysarthria เป็นความผิดปกติของการพูด ที่เกิดจากมีการอ่อนแรงของกล้ามเนื้อ กล้ามเนื้อตึงตัวน้อยกว่าปกติ ผู้ป่วยมักจะพูดเสียงมีลมแทรก พูดเสียงขึ้นจมูก พูดเสียงสระ พยัญชนะไม่ชัดเจน ตำแหน่งของพยาธิสภาพอยู่ที่ Lower Motor Neuron พวกนี้พบในผู้ป่วยที่เป็น CVA (Cerebro Vascular Accident), Tumor, Poliomyelitis, Myasthenia gravis, Bulbar palsy, Facial Palsy เป็นต้น
2. Spastic Dysarthria เป็นความผิดปกติของการพูดที่เกิดจากมีการเกร็งของกล้ามเนื้อ กล้ามเนื้อเคลื่อนไหวได้ในขีดจำกัดและเคลื่อนไหวค่อนข้างช้า ผู้ป่วยมักจะพูดเสียงห้าว พูดเสียงต่ำ พูดเสียงขึ้นจมูก พูดไม่ชัด จังหวะการพูดค่อนข้างช้า ตำแหน่งของพยาธิสภาพอยู่ที่ Upper Motor Neuron พวกนี้พบในผู้ป่วยที่เป็น CVA, Tumor, Encephalitis, Cerebral palsy เป็นต้น
3. Ataxic Dysarthria เป็นความผิดปกติของการพูดที่เกิดจากอวัยวะที่ใช้ในการพูด ทำงานประสานกันไม่ดีเท่าที่ควร การตึงตัวของกล้ามเนื้อน้อยกว่าปกติ มีการสั่นของกล้ามเนื้อในขณะเคลื่อนไหว ควบคุมแรงและทิศทางการเคลื่อนไหวไม่ดี ผู้ป่วยมักจะพูดเสียงห้าว พูดไม่ชัด พูดเสียงระดับเดียวกันหมด (Mono pitch) พูดเสียงดังระดับเดียวกัน (Mono loudness) ไม่มีการเน้นเสียงหนักเบาและมีจังหวะการพูดค่อนข้างช้าด้วย ตำแหน่งของพยาธิสภาพอยู่ที่ Cerebellar System มักพบในผู้ป่วยที่เป็น CVA, Tumor, Infection, Cerebral Palsy เป็นต้น
4. Hypokinetic Dysarthria เป็นความผิดปกติของการพูดที่เกิดจากการที่กล้ามเนื้อที่เกี่ยวข้องกับการพูดเคลื่อน ไหวได้ในขีดจำกัด มีความสามารถในการเคลื่อนไหวน้อยกว่าปกติ มีความลำบากในการเริ่มต้นเคลื่อนไหว มีการสั่นเมื่ออยู่ในท่าพัก ปัญหาทางการพูด คือ พูดเสียงระดับเสียงเดียวกันหมด พูดเสียงดังระดับเดียว พูดไม่ชัด มีการหยุดพูดในจังหวะที่ไม่เหมาะสม ตำแหน่งของพยาธิสภาพอยู่ที่ Extrapyramidal System พวกนี้พบในผู้ป่วยที่เป็น Parkinson’s Disease, พวก Drug Induced เป็นต้น
5. Hyperkinetic Dysarthria เป็นความผิดปกติของการพูดที่เกิดมีการเคลื่อนไหว

ของกล้ามเนื้อที่อยู่นอกอำนาจจิตใจ (involuntary movement) มาขัดขวางการเคลื่อนไหวอย่างปกติในขณะพูด การเคลื่อนไหวแบบนอกอำนาจจิตใจนี้ บางชนิดเกิดขึ้นอย่างรวดเร็ว บางชนิดเกิดขึ้นอย่างช้าๆ หรือบางคนอาจเป็นทั้ง 2 แบบ ซึ่งเกิดไม่สม่ำเสมอ พวกนี้จะมีการพูดไม่ชัดทั้งเสียงสระและพยัญชนะ พูดระดับเสียงผิดปกติ ความดังของเสียงผิดปกติ และมักจะหยุดพูดในที่ไม่ควรหยุด ตำแหน่งของพยาธิสภาพอยู่ที่ Extrapyramidal System พวกนี้พบในผู้ป่วยที่เป็น Infection, Chorea, Athetosis เป็นต้น

6. Mixed Dysarthria เป็นการพูดผิดปกติที่เกิดจากการมีพยาธิสภาพที่ระบบประสาทมากกว่า 1 แห่งการพูดผิด ปกติของพวกนี้จึงมีลักษณะที่แตกต่างกันไปแล้วแต่ว่าจะเป็นแบบใด เช่น ในผู้ป่วย Amyotrophic Lateral Sclerosis จะมีพยาธิสภาพที่ Upper และ Lower Motor Neurons ผู้ป่วยจะพูดไม่ชัด พูดเสียงห้าว พูดเสียงขึ้นจมูก จังหวะการพูดช้า พูดเสียงระดับเดียวกันหมด พูดเป็นวลีสั้นๆ เป็นต้น ส่วนในผู้ป่วย Multiple Sclerosis จะมีพยาธิสภาพที่ upper motor neurons และ cerebellar systems ผู้ป่วยจะพูดไม่ชัด พูดเสียงห้าว ควบคุมความดังของเสียงและควบคุมระดับเสียงไม่ได้ พูดเสียงขึ้นจมูก พูดเสียงลมแทรก เป็นต้น

จากความผิดปกติทั้ง 6 ประเภท สามารถสรุปภาพรวมของความผิดปกติทางการพูดประเภท Dysarthria ได้ดังตารางที่ 2-2

ตารางที่ 2-2 รายละเอียดของความผิดปกติประเภท Dysarthria

คุณสมบัติ	**Dysarthria**
1. การเคลื่อนไหว ลิ้น ริมฝีปาก เพดานอ่อน	มีความผิดปกติของกล้ามเนื้อ มากกว่า 1 ตำแหน่ง เช่น กล้ามเนื้อที่ลิ้นขากรรไกร และอื่นๆ
2. ความผิดปกติทางร่างกาย	เมื่อเป็นร่วมกับอัมพาต สมองพิการ กล้ามเนื้ออ่อนแรง การเคลื่อนไหวร่างกายและบริเวณใบหน้ารวมถึงริมฝีปาก ลำบากหรือแทบไม่ได้
3. พยาธิสภาพ	ระบบประสาทส่วนกลาง หรือปลาย
4. พัฒนาการพูด	มักจะล่าช้า
5. การเปล่งเสียง	การทำงานของกล้ามเนื้ออาจไม่ประสานกัน จึงทำให้หายใจ

	ผิดปกติ เสียงผิดปกติ แหบ มีลมแทรก ความดังและโทนเสียงมีระดับเดียว
6. การฟังและพูดตาม	การฟังแยกเสียงปกติ เข้าใจภาษา แต่พูดได้เท่าที่กล้ามเนื้อจะทำได้
7. อัตราเร็วการออกเสียง	ไม่แน่นอน มักช้ากว่าปกติ
8. ลักษณะความผิดปกติ	มีความผิดปกติทั้งในการพูด ทั้งขณะพูดเองและพูดตาม
9. การแปรเสียง	ผิดปกติแน่นอน ลักษณะของเสียงหรือแทนด้วยเสียงอื่น
10. ฉันทลักษณ์	อาจช้าหรือเร็วขึ้นกับชนิดของความผิดปกติ ไม่มีการพูดเน้นเสียง

2.3.2 Apaxia

เป็นการพูดไม่ชัดและมีลีลาจังหวะการพูดผิดปกติ เนื่องจากอวัยวะที่ใช้ในการพูดไม่สามารถเคลื่อนไหวให้เป็นไปตามลำดับขั้นของการพูดได้ถูกต้อง ซึ่งเป็นผลจากการมีพยาธิสภาพที่สมอง ทั้งๆที่ผู้ป่วยสามารถเคลื่อนไหวอวัยวะเหล่านี้ได้โดยไม่มีความผิดปกติ ทั้งในด้านโครงสร้างและในการทำงานที่เกี่ยวข้องกับการเคลื่อนไหว ไม่มีอัมพาตของกล้ามเนื้อ ไม่มีความผิดปกติในด้านการประสานงานกันของกล้ามเนื้อเลย อวัยวะเหล่านี้จะทำงานได้เป็นปกติถ้าทำการเคลื่อนไหวอย่างอัตโนมัติ เช่น แลบลิ้นเลียปาก แต่ถ้าให้เคลื่อนไหวอวัยวะเหล่านี้เพื่อใช้ในการพูดแล้ว ผู้ป่วยไม่สามารถทำได้ดังที่ตั้งใจจะพูด ความผิดพลาดในการพูดนี้จะเกิดขึ้นในลักษณะของการพูดไม่ชัด แต่ไม่สม่ำเสมอ บางครั้งเมื่อผู้ป่วยออกเสียงได้ถูกต้องครั้งหนึ่งแล้ว เมื่อพูดออกเสียงนั้นใหม่ ผู้ป่วยอาจทำไม่ได้ดังต้องการ ความผิดพลาดในการพูดบางครั้ง เกิดจากผู้ป่วยคาดการณ์ล่วงหน้าว่าจะพูดผิดและพยายามแก้ไขตัวเอง เพื่อให้มีความผิดพลาดในการออกเสียงน้อยลง แต่ก็ไม่สามารถควบคุมอวัยวะที่ใช้ในการพูดให้ทำงานไปตามลำดับขั้นได้ดังต้องการ ปัญหาการพูดของผู้ป่วยมักจะเกิดมากขึ้น ถ้าพูดเสียงหรือคำที่ยากและซับซ้อนมากขึ้น หรือพูดพยางค์หรือคำที่ยาวขึ้น

เนื่องจากความผิดปกติประเภท Apaxia of speech ในการพูดแต่ละครั้งผู้ป่วยไม่สามารถพูดให้เหมือนเดิมและความผิดพลาดในการพูดของผู้ป่วยในแต่ละครั้งจะเพิ่มมากขึ้น อาจมีการใช้คำผิดและใช้คำไม่เหมือนเดิม ทำให้คาดการณ์การพูดของผู้ป่วยไม่ได้ ซึ่งต่างจากผู้ป่วยประเภท Dysarthria คือ สามารถพูดคำเดิมได้และความผิดพลาดจะเกิดขึ้นเฉพาะคุณสมบัติของเสียงที่พูดเท่านั้น ดังนั้นงานวิจัยนี้จึงได้พิจารณาเฉพาะผู้ป่วยประเภท Dysarthria เพียงอย่างเดียว

2.4 การรู้จำเสียงคำพูด (Speech Recognition)

การรู้จำเสียงคำพูด เป็นศาสตร์หนึ่งของการวิจัยที่มุ่งศึกษาให้คอมพิวเตอร์เข้าใจเสียงคำพูดของมนุษย์ เหตุส่วนหนึ่งอันเนื่องมาจากข้อได้เปรียบของการใช้เสียงพูดในงานต่างๆ เช่น การจัดทำเอกสาร ผู้จัดทำไม่จำเป็นต้องมีทักษะในการพิมพ์ดีดก็สามารถสั่งการโดยใช้การพูดของตนได้ หรือในแง่ของความสะดวกสบาย สามารถใช้เสียงพูดในการเปิด-ปิด เครื่องใช้ไฟฟ้าในบ้าน อีกทั้งการพูดเป็นการป้อนคำสั่งหรือข้อมูลที่ใช้เวลาในการทำงานที่น้อยกว่าการสั่งงานด้วยวิธีอื่น สามารถแบ่งประเภทของการรู้จำเสียงคำพูดได้ดังนี้

2.4.1 แบ่งตามลักษณะของผู้พูด

1. แบบขึ้นกับผู้พูด (Speaker Dependence)
2. แบบหลายผู้พูด (Multi-Speaker)
3. แบบไม่ขึ้นกับผู้พูด (Speaker Independence)

2.4.2 แบ่งตามลักษณะเสียงคำพูด

1. แบบเสียงคำพูดเดี่ยว (Isolated Speech)
2. แบบเสียงคำพูดต่อกัน (Connected Speech)
3. แบบเสียงคำพูดต่อเนื่อง (Continuous Speech)

2.4.3 แบ่งตามจำนวนคำที่ให้ระบบทำการรู้จำ

1. แบบจำนวนคำน้อย (Small Vocabulary) เป็นการฝึกฝนให้ระบบทำการรู้จำคำพูดไม่เกิน 100 คำ
2. แบบจำนวนคำปานกลาง (Moderate Vocabulary) เป็นการฝึกฝนให้ระบบทำการรู้จำคำพูดระหว่าง 100 ถึง 5000 คำ
3. แบบจำนวนคำมาก (Large Vocabulary) เป็นการฝึกฝนให้ระบบทำการรู้จำคำพูดตั้งแต่ 5000 คำขึ้นไป

2.5 การประมวลผลสัญญาณเบื้องต้น (Pre-processing)

โดยธรรมชาติของสัญญาณเสียงพูดจะไม่เสถียรและเปลี่ยนแปลงตามเวลา (Non - Stationary) ดังนั้น เมื่อต้องการนำสัญญาณเสียงพูดมาประมวลผลสัญญาณดิจิตอล (Digital Signal Processing) จึงจำเป็นต้องแบ่งสัญญาณพูดออกเป็นช่วงเวลาสั้นๆ (Short Time) เพื่อให้สัญญาณ

เสียงมีความเสถียรและไม่เปลี่ยนแปลงตามเวลา (Stationary) จากนั้นจึงจะสามารถนำสัญญาณเสียงไปประมวลผลต่อไปได้ กรอบเสียงพูด (Speech Frame) ความยาวประมาณ 10 - 40 มิลลิวินาที ทำให้สัญญาณเสียงพูดในแต่ละกรอบเสียงพูดเป็นสัญญาณที่มีความเสถียรและไม่เปลี่ยนแปลงตามเวลา การเหลื่อมกรอบเสียงพูด (Frame Overlap) จะทำให้รอยต่อของลักษณะสำคัญของเสียงพูดจากกรอบเสียงพูดหนึ่งไปยังอีกกรอบเสียงพูดหนึ่งเรียบ (Smooth) ขึ้น กรรมวิธีการวิเคราะห์สัญญาณเสียงพูดเบื้องต้นสามารถแบ่งออกเป็นหัวข้อย่อยๆได้ดังนี้

2.5.1 กรรมวิธีการปรับบรรทัดฐานแอมปลิจูด (Amplitude Normalization)

กรรมวิธีการปรับบรรทัดฐานแอมปลิจูดของสัญญาณเสียงพูดเป็นการเพิ่มหรือลดขนาดของสัญญาณเสียงพูด เพื่อให้ขนาดของสัญญาณเสียงพูดมีความเหมาะสม เนื่องจากสัญญาณเสียงพูดของแต่ละบุคคลมีขนาดไม่เท่ากัน จึงจำเป็นต้องปรับให้มีขนาดของสัญญาณเสียงพูดอยู่ในบรรทัดฐานเดียวกัน เพื่ อง่ายต่อการวัดคุณลักษณะและเปรียบเทียบสัญญาณเสียง การปรับบรรทั ดฐานแอมปลิจูด แสดงดังสมการที่ 2.1

$$s'(n) = \frac{S(n)}{S_{max}} \quad , n = 0, 1, 2, 3, \ldots, N-1 \qquad (2.1)$$

เมื่อ $s'(n)$ คือ สัญญาณเสียงที่ปรับบรรทัดฐานแอมปลิจูดแล้ว

$s(n)$ คือสัญญาณเสียงพูดดิจิตอล

S_{max} คือ ค่าแอมปลิจูดสูงสุดของสัญญาณ $s(n)$

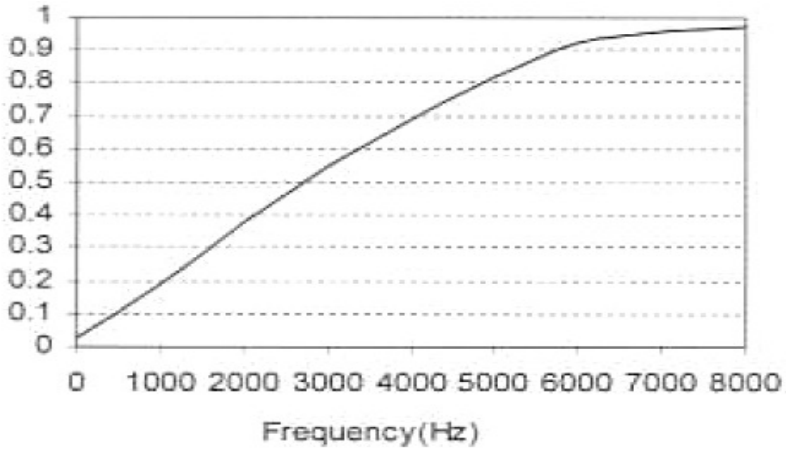
2.5.2 กรรมวิธีการเน้นล่วงหน้า (Pre-emphasis) [12]

เนื่องจากสัญญาณเสียงพูดของมนุษย์ จะมีองค์ประกอบส่วนใหญ่อยู่บริเวณความถี่ต่ำ เมื่อเทียบกับแถบความถี่ปฏิบัติงาน (Bandwidth) ไม่เกิน 5 กิโลเฮิรตซ์ ดังนั้น เพื่อให้อัตราส่วนสัญญาณเสียงต่อสัญญาณรบกวน (Signal-to-Noise Ratio: SNR) มีค่าค่อนข้างคงที่ตลอดช่วงความถี่ปฏิบัติงานนี้ เราจึงต้องมีการพรีเอมฟาซิส โดยเน้นความถี่สูงให้มีขนาดสูงขึ้นนั่นคือ การพรีเอมฟาซิส คือการกรองสัญญาณด้วยวงจรกรองความถี่สูงผ่าน (High pass filter) ซึ่งมักนิยมใช้วงจรกรองอันดับหนึ่ง มีฟังก์ชันถ่ายโอนเป็น

$$H(z) = 1 - az^{-1} \quad , 0.93 < a < 0.98 \qquad (2.2)$$

เมื่อเทียบกับภาพประกอบที่ 2-5 จะได้

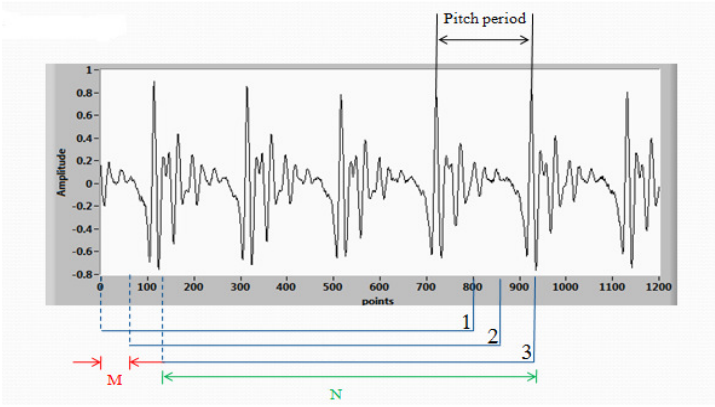
$$S(n) = s(n) - as(n-1) \qquad (2.3)$$



ภาพประกอบ 2-5 ขนาดสเปกตรัมของฟังก์ชันถ่ายโอนของการพรีเอมฟาซิส (α=0.94) [16]

2.5.3 การแบ่งช่วงสัญญาณ (Frame Blocking)

สัญญาณที่ผ่านการพรีเอมฟาซิสแล้ว *s(n)* จะถูกตัดแบ่งออกเป็นช่วงๆ หรือ เฟรม ช่วงละ *N* ตัวอย่างสัญญาณ การวิเคราะห์จะวิเคราะห์ทีละช่วงของแต่ละ *N* ตัวอย่างสัญญาณ ดังภาพประกอบที่ 2-6 โดยช่วงในการวิเคราะห์แต่ละช่วงจะถูกเลื่อนไปเป็นระยะ *M* ช่วงสัญญาณ จะเห็นได้ว่า ถ้าค่า *M* โตกว่าค่า *N* ในการเลื่อนของช่วงในการวิเคราะห์ จะทำให้บางสัญญาณไม่ถูกใช้ในการวิเคราะห์ จะเป็นการสูญเสียส่วนหนึ่งทำให้ผลที่ได้ไม่ถูกต้องเท่าที่ควร ถ้าค่ า *M* เล็กกว่า *N* จะทำให้ตัวอย่างสัญญาณ ทุกตัวถูกนำมาวิเคราะห์ ยิ่งค่า *M* เล็กเท่าใด ความแม่นยำในการวิเคราะห์ก็จะยิ่งสูงขึ้นเท่านั้น แต่จะทำให้การคำนวณช้าลง



ภาพประกอบที่ 2-6 แสดงการแบ่งช่วงของสัญญาณที่ใช้ในการวิเคราะห์

2.5.4 การวินโดว์ (Windowing)

โดยทั่วไปการวินโดว์สามารถกระทำได้หลายแบบตามความเหมาะสมของสัญญาณที่จะนำมาวิเคราะห์ ซึ่งจะยึดหลักตามรูปแบบทั่วไปของการวินโดว์ สามารถแบ่งรายละเอียดออกเป็นหัวข้อได้ดังนี้

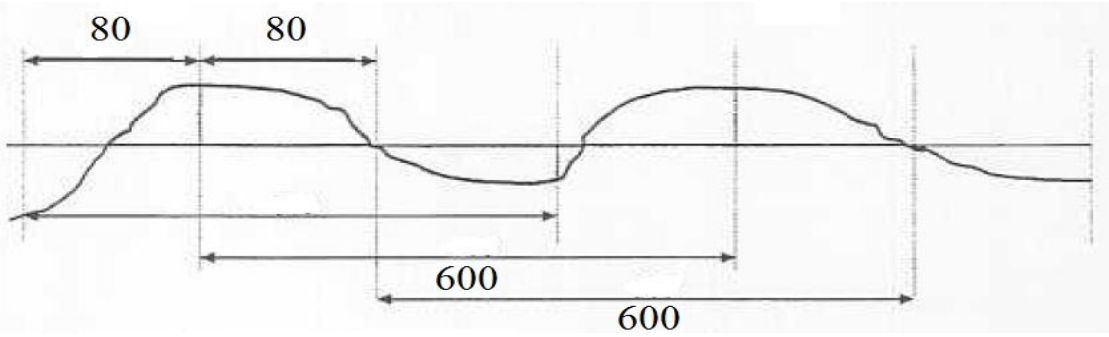
1. วินโดว์จะต้องสั้นพอที่คุณสมบัติของเสียงที่เราสนใจจะวิเคราะห์นั้นยังไม่เปลี่ยนแปลงในวินโดว์
2. วินโดว์จะต้องยาวพอที่จำนวนตัวอย่างสัญญาณในวินโดว์สามารถนำมาคำนวณหาคุณสมบัติที่ต้องการได้
3. วินโดว์ที่ติดกันไม่ควรจะสั้นจนกระโดดข้ามข้อมูลบางส่วนไป แต่ควรเลื่อนวินโดว์ให้น้อยกว่าขนาดของเฟรม

ในงานวิจัยนี้เลือกใช้ความถี่ในการสุ่มสัญญาณ 20 กิโลเฮิรตซ์ ค่า $N = 600$ และค่า $M = 80$ (ระยะในการเลื่อนเฟรม 4 มิลลิวินาที)

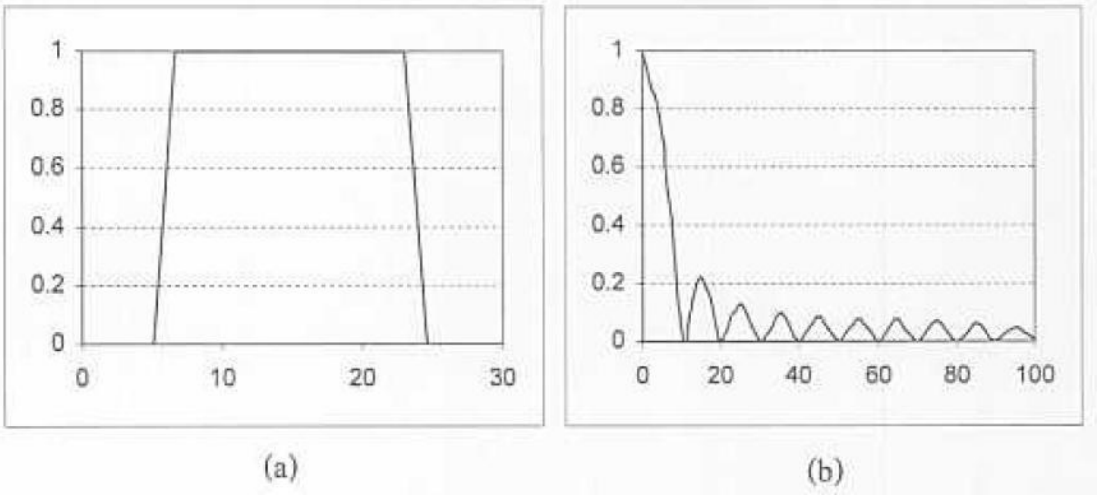
พิจารณาช่วงสัญญาณ N ตัวอย่างสัญญาณของช่วงใดๆ ที่ตัดมาวิเคราะห์ (ภาพประกอบที่ 2-7) จะเห็นว่าที่ขอบของเฟรมที่ตัดมานี้มีความไม่ต่อเนื่องของสัญญาณ ถ้ามองในโดเมนความถี่ จะมีความถี่สูงเกิดขึ้นดังนั้นเพื่อที่จะลดองค์ประกอบทางความถี่ที่สูงเหล่านี้ จะคูณด้วยฟังก์ชันวินโดว์เพื่อลดความไม่ต่อเนื่องของสัญญาณที่ขอบ และไม่ทำให้สเปกตรัมของสัญญาณในช่วงความถี่ต่ำเปลี่ยนไปมากนัก สำหรับงานวิจัยนี้จะเลือกใช้วินโดว์ชนิดแฮมมิ่ง (Hamming Window Function) ซึ่งเป็นวินโดว์ที่นิยมใช้กับสัญญาณเสียง นิยามโดยสมการดังนี้

$$w(n) = 0.54 - 0.46\left(\frac{2\pi n}{N-1}\right) \quad , n = 1, 2, \dots, N-1 \qquad (2.4)$$

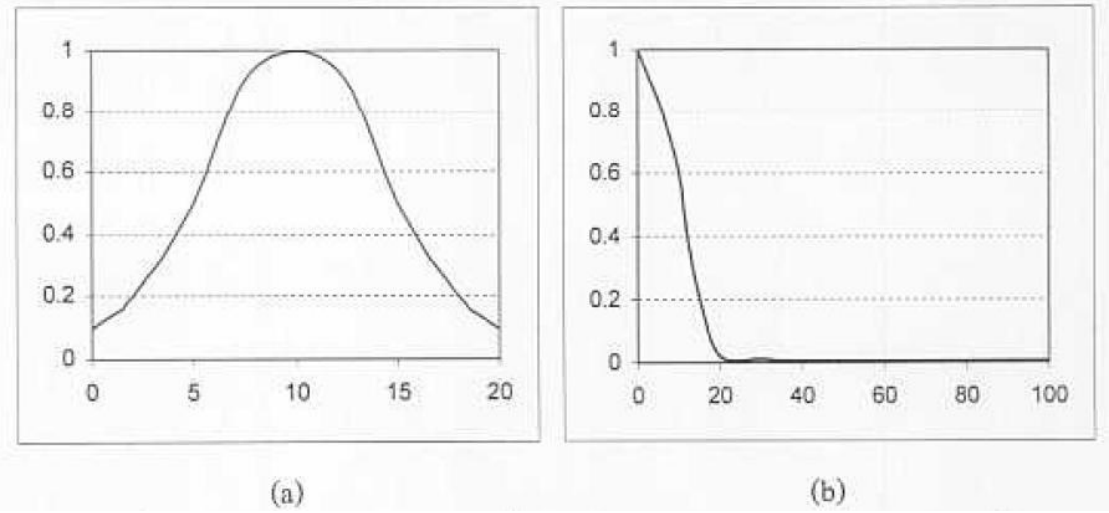
ภาพประกอบที่ 2-8 และ 2-9 แสดงองค์ประกอบทางเวลาและทางความถี่ของฟังก์ชันวินโดว์แบบสี่เหลี่ยม และแฮมมิ่ง ตามลำดับ จะเห็นว่าสเปกตรัมของวินโดว์แฮมมิ่งมีริพเพิล (Ripple) น้อยกว่าของวินโดว์สี่เหลี่ยม



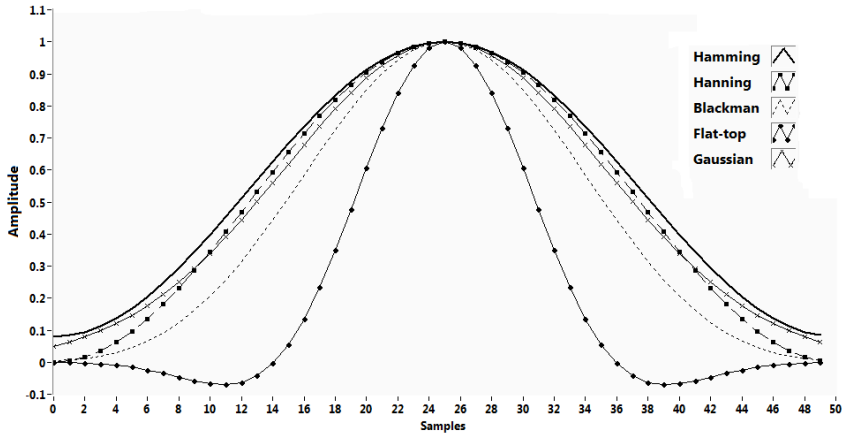
ภาพประกอบ 2-7 แสดงส่วนของสัญญาณที่ตัดมาวิเคราะห์ [16]



ภาพประกอบ 2-8 แสดงวินโดว์แบบสี่เหลี่ยม (a) ในโดเมนเวลา (b) ในโดเมนความถี่ [16]



ภาพประกอบ 2-9 แสดงวินโดว์แบบแฮมมิ่ง (a) ในโดเมนเวลา (b) ในโดเมนความถี่ [16]



ภาพประกอบ 2-10 ลักษณะของวินโดว์ชนิดต่างๆ

ดังนั้นจะได้ว่า

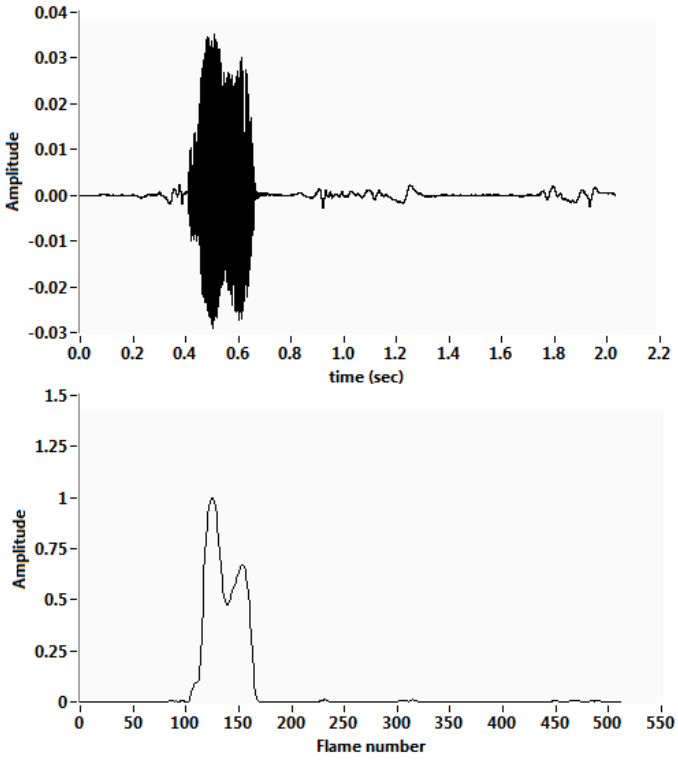
$$x'(n) = x(n) \cdot w(n) \qquad (2.5)$$

โดย $x'(n)$ คือ สัญญาณเสียงที่ผ่านการวินโดว์เรียบร้อยแล้ว, $x(n)$ คือ สัญญาณเสียงในโดเมนเวลา

2.5.5 พลังงานของสัญญาณเสียงพูด

พลังงานของสัญญาณเสียงพูด ถูกนำมาใช้ในการวิเคราะห์เสียงพูด และเป็นคุณลักษณะที่นิยมนำมาใช้กันอย่างแพร่หลาย เนื่องจากเป็นวิธีที่คำนวณง่ายและรวดเร็ว พลังงานของสัญญาณเสียงพูดเป็นคุณสมบัติที่แสดงให้เห็นว่ามีสัญญาณเสียง (รวมทั้งสัญญาณรบกวน) เกิดขึ้น ณ เวลานั้นหรือไม่ การคำนวณหาค่าพลังงาน จะทำทีละกรอบเสียงพูด โดย E_s คือ สัญญาณเสียงพูดในกรอบเสียงพูด m และในแต่ละกรอบเสียงพูดจะมีสัญญาณเสียงพูดจำนวน n ถึง N-1 ซึ่งวิธีการคำนวณหาค่าพลังงานของสัญญาณเสียงพูดในงานวิจัยนี้ เลือกใช้วิธีพลังงานกำลังสอง (Square Energy) ซึ่งเป็นการวัดค่าพลังงานจากสัญญาณเสียงพูดยกกำลังสอง ทำให้ค่าพลังงานมีความไวต่อสัญญาณที่มีขนาดใหญ่ การหาค่าพลังงานกำลังสองแสดงดังสมการที่ (2.6)

$$E_s(m) = \sum_{n=0}^{N-1} s^2(n) \qquad (2.6)$$



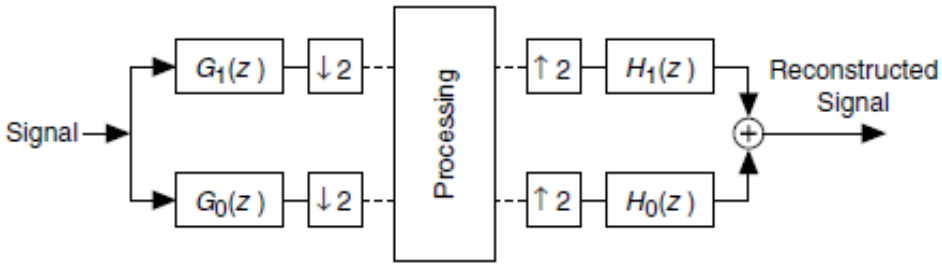
ภาพประกอบ 2-11 การเปลี่ยนระดับพลังงานของสัญญาณเสียง
(รูปบน คือ สัญญาณเสียง, รูปล่าง คือ สัญญาณพลังงานเสียง)

2.5.6 การคัดแยกความถี่ของสัญญาณโดยเทคนิค Wavelet Transform [13]

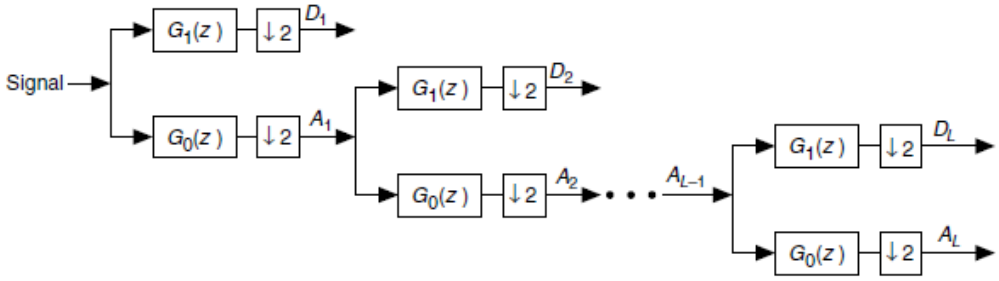
เนื่องจากสัญญาณเสียงพูดของมนุษย์ จะมีองค์ประกอบส่วนใหญ่อยู่บริเวณความถี่ต่ำ เมื่อเทียบกับแถบความถี่ปฏิบัติงาน (Bandwidth) ไม่เกิน 5 กิโลเฮิรตซ์ ซึ่งประกอบด้วยความถี่ฟอร์แมนต์ที่ 1 และ 2 ปนกันอยู่ ก่อนที่จะวิเคราะห์สัญญาณเสียงในโดเมนความถี่ ถ้าเราสามารถคัดแยกเสียงในโดเมนเวลาออกเป็นย่านความถี่แต่ละระดับได้แล้ว จะช่วยให้การวิเคราะห์สัญญาณเสียงในโดเมนความถี่แม่นยำและสะดวกมากยิ่งขึ้น งานวิจัยนี้จึงทำการคัดแยกสัญญาณเสียงในโดเมนเวลาโดยใช้หลักการแปลงเวฟเล็ตแบบฟิลเตอร์แบงค์ (Filter Bank) ซึ่งมีหลักการทำงานเบื้องต้นดังนี้

ฟิลเตอร์แบงค์แบบสองช่องสัญญาณ (Two Channel Filter Bank) และชุดตัวกรองสำหรับการสังเคราะห์ (Synthesis Filter Bank) ตัวกรองในการวิเคราะห์ทำหน้าที่ในการแยกสัญญาณอินพุทเป็นสองส่วนคือ ความถี่สูงด้วยตัวกรองความถี่สูง $G_1(z)$ และความถี่ต่ำด้วยตัวกรองความถี่ต่ำ $G_0(z)$ ตามด้วย Decimators หรือ Down Sampling สัญญาณ เมื่อนำฟิลเตอร์แบงค์แบบ

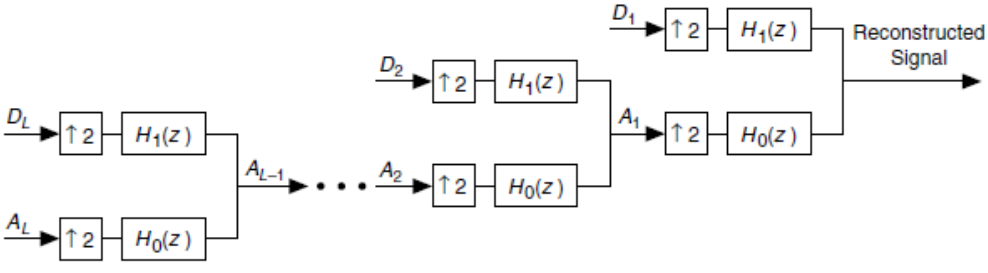
สองช่องสัญญาณมาเรียงต่อกันในลักษณะ Tree Structure จะได้ Octave Analysis Filter Banks ดังภาพประกอบที่ 2-13 ฟิลเตอร์แบงค์แบบสองช่องสัญญาณชุดแรกจะแยกสัญญาณ *Signal* เป็นสัญญาณความถี่สูงครึ่งบน D_1 และสัญญาณที่ความถี่ต่ำครึ่งล่าง A_1 ออกจากกัน สัญญาณเอาท์พุต A_1 เป็นส่วนที่ถูกแยกความถี่ต่อไป ดังนั้นส่วนของความถี่ต่ำที่ได้จากแต่ละขั้นตอนจะถูกทำการแยกแบนด์ต่อไปเรื่อยๆ ตามระดับที่ต้องการ



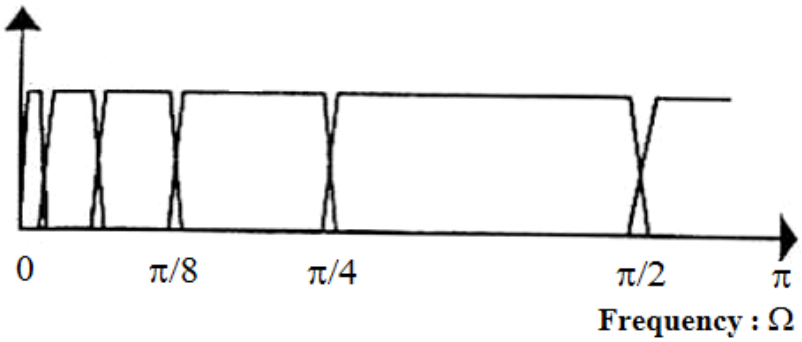
ภาพประกอบ 2-12 ฟิลเตอร์แบงค์แบบสองช่องสัญญาณ [13]



ภาพประกอบ 2-13 Octave Band Analysis filter banks. [13]



ภาพประกอบ 2-14 Octave Band Synthesis filter banks. [13]



ภาพประกอบ 2-15 ลักษณะการแยกแบนด์ความถี่ (Frequency Band) ของ Octave analysis filters banks. [13]

ฟิลเตอร์แบงค์ในการสังเคราะห์หรือสร้างกลับสัญญาณ (Synthesis Filter Banks) จะประกอบด้วย Expander หรือ Up-Sampling และตัวกรองในการสังเคราะห์ (Synthesis Filter) H0(z) และ H1(z) ลักษณะของ Octave Band Synthesis Filter Banks แสดงดังภาพประกอบที่ 2-14 การสร้างกลับสัญญาณ (Reconstruction) เป็นการหาค่าของ Reconstructed Signal ตัวกรองในการวิเคราะห์และตัวกรองในการสังเคราะห์มีความสัมพันธ์กันในลักษณะของวงจรควอดราเจอร์มิเรอร์

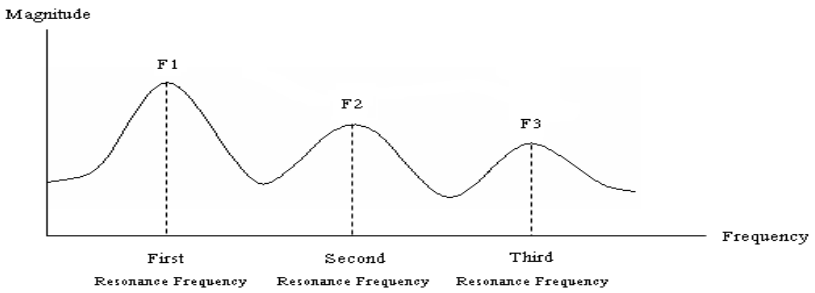
ในการแยกแบนด์ความถี่ของ Octave Filter Bank แบนด์ของความถี่ต่ำจะถูกแบ่งออกไปเรื่อยๆ สัมพันธ์กับอัตราสุ่มที่ลดลงครั้งละสองเท่า ดังนั้นเอาท์พุทในส่วนของความถี่ต่ำจะลดลงในอัตราสองเท่าในแต่ละสเตทและความถี่คัทออฟ (Cut Off Frequency) จะมีค่าลดลงครั้งละสองเท่า โดยลดลงไปทางด้านความถี่ที่ต่ำลง ทรานสเฟอร์ฟังก์ชัน A_L คือส่วนของความถี่ต่ำผ่านครั้งสุดท้ายในขณะที่ A_5, A_4, A_3, A_2 มีลักษณะเป็นแถบความถี่ผ่านและ A_1 เป็นส่วนของความถี่สูงผ่านแสดงดังรูปภาพประกอบที่ 2-15 ดังนั้น Octave Filter Bank จึงสามารถแยกแบนด์ความถี่ต่างๆออกได้ (Band – Separating Filter)

2.6 การสกัดค่าลักษณะสำคัญ (Feature Extraction)

2.6.1 ความถี่กำธรหรือความถี่ฟอร์แมนต์ (Formant Frequency)

ช่องทางเดินเสียงมีลักษณะเหมือนกับท่ออากาศ และทำหน้าที่เป็นตัวทำ ให้เกิดการกำธร (Resonance) ซึ่งการกำธรของช่องทางเดินเสียงนี้เรียกว่าฟอร์แมนต์ (Formant) ในช่องทางเดินเสียงจะมีความถี่ธรรมชาติอยู่ค่าหนึ่ง และจะมีการตอบสนองกับคลื่นเสียงที่มีความถี่ตรงกับความถี่

ช่องทางเดินเสียง ถ้าช่องทางเดินเสียงมีรูปร่างแตกต่างกันออกไป ค วามถี่ธรรมชาติของช่องทางเดินเสียงจะมีค่าเปลี่ยนไปด้วย ความถี่ที่เกิดจากก ารกำธรนี้เองเรียกว่าความถี่กำธรหรือความถี่ฟอร์แมนต์ โดยเกิดจากรูปร่างของช่องทางเดินเสียง ความถี่ที่เกิดจากการสั่นของเส้นเสียงและความถี่ที่เกิดจากช่องทางเดินเสียงจะมีความเป็นอิสระต่อกัน ความถี่ฟอร์แมนต์มีชื่อเรียกตามขนาดของความถี่ เริ่มจากฟอร์แมนต์ที่ 1 ซึ่งมีความถี่ต่ำสุดในกลุ่มของความถี่ฟอร์แมนต์ และเรียกความถี่ถัดไปว่าความถี่ฟอร์แมนต์ที่ 2 ตามลำดับ ทั้งนี้ค่าความถี่ฟอร์แมนต์ที่มีผลต่อการจำแนกเสียงสระจะมีไม่เกิน 3 ค่า คือ ความถี่ฟอร์แมนต์ที่1, 2 และ 3



ภาพประกอบที่ 2-16 ความถี่ฟอร์แมนต์บนสเปกตรัมของเสียงพูด [5]

ความถี่ฟอร์แมนต์สำหรับแต่ละเสียงสระค่อนข้างจะแตกต่างกันอย่างชัดเจน ตัวอย่างเช่น สำหรับผู้พูดที่มีความถี่มูลฐานของเสียงเท่ากับ 100 Hz และมีฮาร์โมนิคที่ 200, 300, 400, 500,…Hz เมื่อพูดคำที่มีเสียงสระอี เช่น “he” จะให้ค่าความถี่ฟอร์แมนต์ที่หนึ่งและสองเท่ากับ 300 และ 2100 Hz ตามลำดับ และเมื่อพูดคำที่มีเสียงสระอา เช่น “hard” จะให้ค่าความถี่ฟอร์แมนต์ที่หนึ่งที่สองเท่ากับ 700 และ 900 Hz ตามลำดับ เป็นต้น ค่าคว ามถี่มูลฐานของผู้พูดจะแปรเปลี่ยนไปไม่แน่นอนขึ้นอยู่กับการพูด (Speaking) อารมณ์ของผู้พูด (Mood) และการเน้นเสียงพูด (Emphasis) แต่ค่าของขนาดและความสัมพันธ์ของความถี่ฟอร์แมนต์บนสเปกตรัมของเสียงพูดค่อนข้างแน่นอน จึงทำให้การรู้จำเสียงพูดทำได้ง่ายขึ้นเมื่อพิจารณาจากค่าความถี่ฟอร์แมนต์

ค่าความถี่ฟอร์แมนต์ของผู้พูดแต่ละคนจะแตกต่างกันขึ้นอยู่กับความยาวและการเปลี่ยนแปลงรูปร่างของช่องทางเดินเสียงของผู้พูดแต่ละคน อย่างไรก็ตาม ยังคงสามารถพิจารณาฟอร์แมนต์อยู่ในช่วงความถี่หนึ่งๆได้ โดยฟอร์แมนต์ที่หนึ่งจะอยู่ในช่วงความถี่ประมาณ 200-1000

Hz ฟอร์แมนต์ที่สองจะอยู่ในช่วงความถี่ประมาณ 500 – 2500 Hz และ ฟอร์แมนต์ที่สามจะอยู่ในช่วงความถี่ประมาณ 1500 – 3500 Hz สำหรับผู้พูดปกติ เป็นต้น Abramson [5] ได้วัดค่าความถี่ฟอร์แมนต์ ของสระในภาษาไทยมาตรฐานโดยใช้เพศชาย 2 คน สรุปได้ดังตาราง 2-3

ตารางที่ 2-3 ค่าความถี่ฟอร์แมนต์ของสระในภาษาไทย (ผู้พูดปกติ)

เสียงสระ (คำ) / ประเภทความถี่ (Hz)	อึ	เออะ	อะ	อือ	เออ	อา
ฟอร์แมนต์ที่ 1 (Hz)	300	540	720	300	540	780
ฟอร์แมนต์ที่ 2 (Hz)	1380	1200	1380	1380	1260	1380

ค่าของความถี่ฟอร์แมนต์ที่ 1 จะลดลงถ้าบริเวณระหว่างลิ้นส่วนกลางรวมทั้งเพดานแข็งและบริเวณระหว่างลิ้นส่วนหลังกับเพดานอ่อนแคบลง ค่าของความถี่ฟอร์แมนต์ที่ 2 จะเพิ่มขึ้นถ้าช่องระหว่างลิ้นส่วนกลางกับเพดานแข็งแคบลง นอกจากนี้ ค่าของความถี่ฟอร์แมนต์ทุกตัวจะลดลงถ้ามีการห่อริมฝีปาก

สำหรับวิธีการที่ใช้ในการหาค่าความถี่ฟอร์แมนต์ คือการแปลงฟูเรียร์แบบเร็ว [14] ดังสมการที่ 2.7 เมื่อ $Y(k)$ คือชุดแถวของสัญญาณที่ได้จากการแปลงฟูเรียร์ โดยที่ N คือจำนวนข้อมูลในการแปลงฟูเรียร์และ k มีค่าตั้งแต่ 0 จนถึงจำนวน N-1

$$Y(k) = \sum_{n=0}^{N-1} x_n e^{-j\,2\pi kn\,/N} \qquad (2.7)$$

$$n=0,\ 1,\ 2,\ldots,\ N-1$$

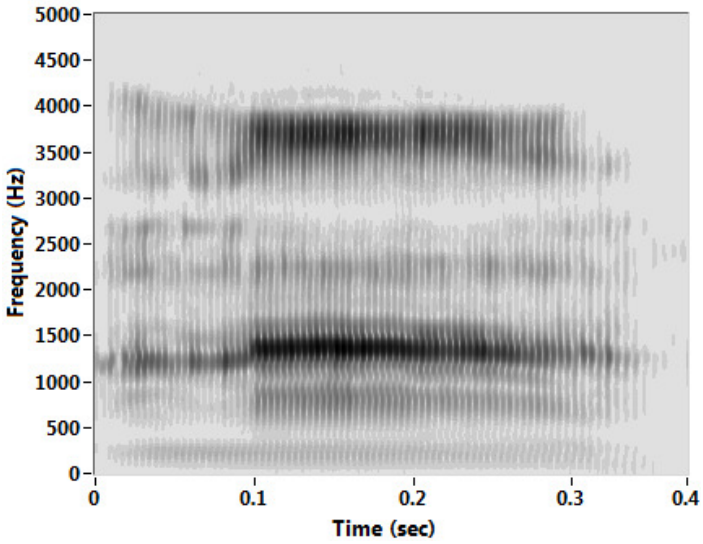
กำหนดให้ค่าที่อยู่ระหว่าง $Y(k)$ ทางด้านแกน x (frequency resolution) คือ $\Delta F = f_s/N$ โดย f_s คือ อัตราสุ่ม (Sampling Rate) ของสัญญาณ

เมื่อทำการคูณสัญญาณแต่ละช่วงกับฟังก์ชันกรอบแล้ว จะคำนวณการแปลงฟูเรียร์แบบเร็วจำนวน N จุด การแปลงฟูเรียร์นี้นอกจากจะนำไปวิเคราะห์ในโดเมนความถี่แล้วยังสามารถนำไปใช้ในการแสดงค่าสเปกตรัมของสัญญาณได้โดยแสดงค่าสเปกตรัมของแต่ละความถี่ตั้งแต่ 0 *Hz* จนถึงครึ่งหนึ่งของความถี่สุ่มตัวอย่าง (Nyquist Frequency) ผลของสเปกตรัมสามารถแสดงได้ในรูปแบบของสเปกโตรแกรม เป็นรูปแบบหนึ่งของการแสดงสัญญาณเสียงเพื่อนำไปใช้ในการ

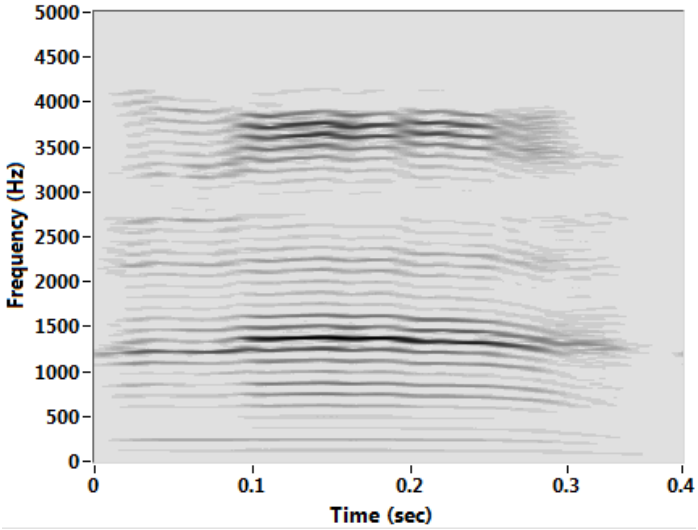
วิเคราะห์ที่นิยมใช้และมีความสำคัญสำหรับงานด้านนี้เป็นอย่างมาก การแสดงผลชนิดนี้ได้จากการคำนวณการแปลงฟูเรียร์ที่ขึ้นกับเวลา ผลที่แสดงคือกราฟสองมิติ ที่แสดงความสัมพันธ์ระหว่างสเปกตรัมที่ขึ้นกับเวลา สำหรับขนาดสเปกตรัมแสดงด้วยสีที่มีความเข้มแตกต่างกันซึ่งสามารถนำไปใช้ในการวิเคราะห์หาลักษณะที่สำคัญของเสียงและลักษณะเด่นสำหรับการรับรู้เสียง ความก้องของเสียง โครงสร้างฟอร์แมนต์ การเน้นเสียงและพิตช์ของเสียง เป็นต้น ในการวิเคราะห์โดยใช้สเปกโตรแกรมโดยทั่วไปแบ่งเป็น 2 แบบคือ

สเปกโตรแกรมแถบกว้าง (Wideband Spectrogram) เป็นการวิเคราะห์ที่แสดงรายละเอียดชั่วขณะในเชิงเวลาได้ดี แต่มีความละเอียดทางแกนความถี่ต่ำ ดัง ภาพประกอบที่ 2-17 กราฟชนิดนี้แสดงถึงการเปลี่ยนแปลงของฟอร์แมนต์และแสดงคาบพิตช์โดยดูจากลายเส้นในแนวตั้งได้

สเปกโตรแกรมแถบแคบ (Narrowband Spectrogram) เป็นการวิเคราะห์ที่แสดงรายละเอียดทางด้านฮาร์โมนิคได้ดี แต่แสดงรายละเอียดชั่วขณะในเชิงเวลาไม่ชัดเจน ดังภาพประกอบที่ 2-18 กราฟชนิดนี้สามารถนำไปหาความถี่มูลฐานได้โดยการวิเคราะห์จากฮาร์โมนิคของความถี่มูลฐานและใช้ในการแยกช่วงที่เป็นเสียงก้องออกจากเสียงไม่ก้องได้



ภาพประกอบที่ 2-17 สเปกโตรแกรมแถบกว้าง



ภาพประกอบที่ 2-18 สเปกโตรแกรมแถบแคบ

2.6.2 การแยกสระเสียงสั้น-เสียงยาว

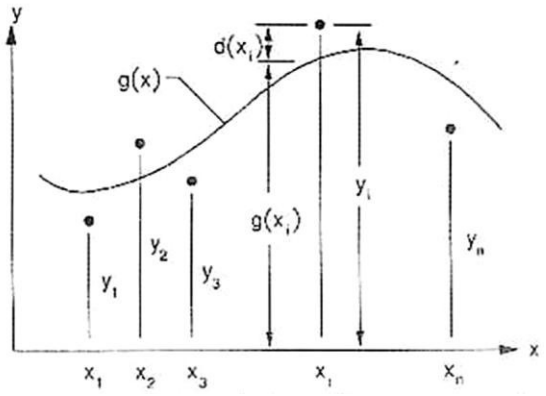
เสียงสระในภาษาไทย นอกจากจะจำแนกประเภทตามรูปเสียงแล้ว อีกทางหนึ่งยังสามารถจำแนกประเภทโดยใช้ช่วงระยะเวลาในการออกเสียงได้อีก 2 กลุ่ม คือ สระเสียงสั้นและเสียงยาว สำหรับงานวิจัยนี้ได้ทำการคัดแยกสระเสียงสั้น-เสียงยาวโดยใช้แนวทางในการคัดแยกทั้งหมด 2 วิธีเปรียบเทียบกันประกอบด้วย 1) การถดถอยของพหุนาม และ 2) การวิเคราะห์จากช่วงระยะเวลาในการออกเสียง ซึ่งมีรายละเอียดดังนี้

2.6.2.1 การถดถอยแบบพหุนาม

การถดถอยแบบพหุนาม (Polynomial regression) เป็นระเบียบวิธีที่ใช้ประดิษฐ์ฟังก์ชันพหุนามสำหรับข้อมูลที่มีการกระจายโดยทั่วไปที่ไม่อยู่ในรูปแบบของเชิงเส้นหรือสมการกำลัง ภาพประกอบที่ 2-19 แสดงการใช้ฟังก์ชันพหุนามในรูปแบบทั่วไปกับชุดของข้อมูลที่ไม่ได้อยู่ในรูปแบบเชิงเส้นชุดหนึ่งที่กำหนดมาให้

ชุดของข้อมูลในภาพประกอบที่ 2-19 นี้ประกอบด้วย x_i, y_i, $i=1.\ 2,\ldots,\ n$ กล่าวคือมีจำนวนข้อมูลทั้งสิ้น n ข้อมูล ในที่นี้เราจะประดิษฐ์ฟังก์ชันพหุนามอันดับ m สำหรับข้อมูลชุดนี้

$$g(x) = a_0 + a_1 x + a_2 x^2 + \cdots a_m x^m \qquad (2.8)$$



ภาพประกอบ 2-19 การถดถอยแบบพหุนามโดยการประดิษฐ์ฟังก์ชันพหุนามจากชุดของข้อมูล [16]

โดย $a_0, a_1, a_2, \ldots, a_m$ เป็นค่าคงตัวที่ไม่รู้ค่าซึ่งจะคำนวณหาจากเงื่อนไขที่ว่า สมการพหุนามที่จะประดิษฐ์ขึ้นมานี้จะก่อให้เกิดค่าความผิดพลาด โดยเฉลี่ยที่น้อยที่สุดจากข้อมูลทั้งหมดที่กำหนดมาให้ ขั้นตอนในการประดิษฐ์สมการพหุนามนี้ เริ่มจากการหาค่าความผิดพลาด (E) ทั้งหมดที่เกิดขึ้นจาก n ข้อมูลในรูปแบบดังนี้

$$E = \sum_{i=0}^{n}[d(x_i)]^2 \qquad (2.9)$$

ซึ่งสามารถเขียนให้ประกอบด้วยฟังก์ชันพหุนามดังแสดงในสมการ (2.8) ได้คือ

$$E = \sum_{i=1}^{n}[y_i - g(x_i)]^2 \qquad (2.10)$$

$$E = \sum_{i=1}^{n}[y_i - (a_0 + a_1x + a_2x^2 + \cdots + a_mx^m)]^2 \qquad (2.11)$$

ในการหาตัวไม่รู้ค่า $a_0, a_1, a_2, \ldots, a_m$ รวมทั้งสิ้น m+1 ค่านั้นเราจะใช้วิธีกำลังสองน้อยสุด (*least – squares*) ซึ่งทำจากการหาค่าต่ำที่สุด (*minimization*) ของค่าความผิดพลาด E โดยเกี่ยวข้องกับตัวไม่รู้ค่า ก่อให้เกิดระบบสมการที่ประกอบด้วย m+1 สมการย่อยนั่นคือ

$$\text{m+1สมการ} \begin{cases} \frac{\partial \text{E}}{\partial \text{a}_0} = 0 \\ \frac{\partial \text{E}}{\partial \text{a}_1} = 0 \\ \frac{\partial \text{E}}{\partial \text{a}_\text{m}} = 0 \end{cases} \qquad (2.12)$$

ดังตัวอย่าง เช่น สมการแรกในระบบสมการนี้สามารถประดิษฐ์ขึ้นได้ดังนี้

$$2\sum_{i=1}^{n}[y_i - (a_0 + a_1x_i + a_2x_i^2 + \cdots + a_mx_i^m)](-1) = 0$$

$$\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} a_0 - \sum_{i=1}^{n} x_i a_1 - \sum_{i=1}^{n} x_i^2 a_2 - \ldots - \sum_{i=1}^{n} x_i^m a_m = 0$$

$$na_0 + (\sum_{i=1}^{n} x_i)a_1 + (\sum_{i=1}^{n} x_i^2)a_2 + \cdots + (\sum_{i=1}^{n} x_i^m)a_m = \sum_{i=1}^{n} y_i$$

และเช่นเดียวกันกับสมการที่สอง ซึ่งคือ

$$2\sum_{i=1}^{n}[y_i - (a_0 + a_1x_i + a_2x_i^2 + \cdots + a_mx_i^m)](-x) = 0$$

$$\sum_{i=1}^{n} x_iy_i - \sum_{i=1}^{n} x_ia_0 - \sum_{i=1}^{n} x_i^2a_1 - \sum_{i=1}^{n} x_i^3a_2 - \ldots - \sum_{i=1}^{n} x_i^{m+1}a_m = 0$$

$$(\sum_{I=1}^{n} x_i)a_0 + (\sum_{i=1}^{n} x_i^2)a_1 + (\sum_{i=1}^{n} x_i^3)a_2 + \cdots + (\sum_{i=1}^{n} x_i^{m+1})a_m = \sum_{i=1}^{n} x_iy_i$$

และสมการอื่นๆที่เหลือก็สามารถทำได้ทำนองเดียวกัน สมการทั้งหมดที่ประดิษฐ์ขึ้นมาได้นี้สามารถเขียนให้อยู่ในรูปแบบของระบบสมการที่ประกอบด้วย m+1 สมการย่อยดังนี้

$$\begin{bmatrix} n & \sum x_i & \sum x_i^2 & \cdots & \sum x_i^m \\ \sum x_i & \sum x_i^2 & \sum x_i^3 & \cdots & \sum x_i^{m+1} \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 & \cdots & \sum x_i^{m+2} \\ \vdots & \vdots & \vdots & & \\ \sum x_i^m & \sum x_i^{m+1} & \sum x_i^{m+2} & \cdots & \sum x_i^{2m} \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_m \end{Bmatrix} = \begin{Bmatrix} \sum y_i \\ \sum x_iy_i \\ \sum x_i^2y_i \\ \vdots \\ \sum x_i^my_i \end{Bmatrix} \quad (2.13)$$

โดยเมตริกซ์จัตุรัสขนาด $(m+1) \times (m+1)$ ทางด้านซ้ายของระบบสมการนี้เป็นเมตริกซ์สมมาตรที่รู้ค่า และเวกเตอร์ขนาด $(m+1) \times 1$ ทางด้านขวาของระบบสมการก็รู้ค่าเช่นกัน ดังนั้นตัวไม่รู้ค่า $a_0, a_1, a_2, \ldots, a_m$ ทั้งหมด m+1 ค่า สามารถคำนวณหาได้จากระบบสมการนี้โดยใช้ระเบียบวิธีการแก้ระบบสมการวิธีใดวิธีหนึ่ง

2.6.2.2 การวิเคราะห์จากช่วงระยะเวลาในการออกเสียง

การวิเคราะห์ช่วงระยะเวลาในการออกเสียง สำหรับงานวิจัยนี้จะใช้วิธีการนับจำนวนกรอบหน้าต่างที่ใช้สำหรับการสร้างสัญญาณพลังงานเสียงพูด ซึ่งโปรแกรมจะทำหน้าที่

แสดงสัญญาณพลังงานและก็นับจำนวนกรอบหน้าต่างโดยสังเกตจากจุดเริ่มต้นและจุดสุดท้ายของสัญญาณพลังงานเสียง ก็จะทำให้ทราบถึงจำนวนกรอบหน้าต่างที่ใช้

2.6.3 ฟัซซี่ลอจิก (Fuzzy Logic)

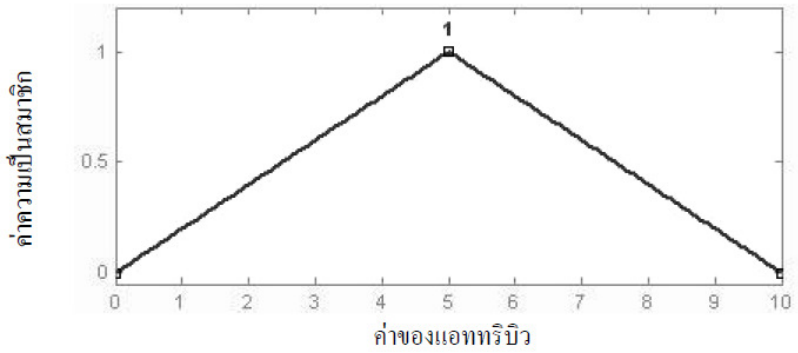
ฟัซซี่ลอจิกเป็นตรรกะที่อยู่บนพื้นฐานความเป็นจริงที่ว่า ทุกสิ่งบนโลกแห่งความเป็นจริงไม่ใช่มีเฉพาะสิ่งมีความแน่นอนเท่านั้น แต่มีหลายสิ่งหลายเหตุการณ์ที่เกิด ขึ้นอย่างไม่เที่ยงและไม่แน่น่อน (Uncertain) อาจเป็นสิ่งที่คลุมเครือ (Fuzzy) ไม่ใช่ชัดเจน (Exact) ยกตัวอย่างเช่น เซตของอายุคน อาจแบ่งเป็น วัยทารก วัยเด็ก วัยรุ่น วัยกลางคน และวัยชรา จะเห็นได้ว่าในแต่ละช่วงอายุคนไม่สามารถระบุได้แน่ชัดว่าวัยทารกกับวัย เด็กแยกจากกันแน่ชัดช่วงใด วัยทารกอาจถูกตีความว่าเป็นอายุระหว่าง 0 ถึง 1 ปี บางคนอาจตีความว่าวัยทารกอยู่ในช่วงอายุ 0 ถึง 2 ปี ในทำนองเดียวกัน วัยเด็กและวัยรุ่น ก็ ไม่สามารถระบุได้ชัดเจนว่าช่วงต่อของอายุควรจะอยู่ใน ช่วงใด อาจตีความว่าวัยเด็กมีอายุอยู่ในช่วง 1 ถึง 12 ปี หรืออาจเป็น 2 ถึง 10 ปี เป็นต้น สิ่งเหล่าเป็นตัวอย่างของความไม่แนน่อน ซึ่งเป็นลักษณะทางธรรมชาติที่เกิดขึ้นทั่วไป เซตของเหตุการณ์ที่ไม่แน่นอนเช่นนี้เรียกว่าฟัซซี่เซต (Fuzzy Set)

ในทฤษฎีฟัซซี่เซตนั้นสามารถแก้ปัญหาข้อจำกัดของเซตแบบดั้งเดิมที่มีค่าความเป็นสมาชิกเป็น 0 หรือ 1 ถูก หรือ ผิด เพียงอย่างเดียวเท่านั้น แต่ในฟัซซี่เซตยอมให้มีค่าความเป็นสมาชิก (Degree of Membership - μ) แสดงค่าตัวเลขตั้งแต่ 0 ถึง 1 หรือเขียนเป็นสัญลักษณ์ [0, 1] โดยที่เลข 0 หมายถึง ไม่เป็นสมาชิกในเซต และเลข 1 หมายถึง เป็นสมาชิกในเซต และค่าระหว่าง 0 กับ 1 เป็นสมาชิกบางส่วนในเซต

องค์ประกอบที่สำคัญในการสร้างฟัซซี่ลอจิกประกอบไปด้วย 3 ส่วนดังนี้

1) ฟังก์ชันความเป็นสมาชิก (Membership Function) คือ ฟังก์ชันที่ กำหนดระดับค่าความเป็นสมาชิกของแอททริบิว ซึ่งค่าความเป็นสมาชิกของแอททริบิวนั้นจะขึ้นอยู่กับรูปแบบของกราฟฟังก์ชันความเป็นสมาชิก ดังนั้นฟังก์ชันความเป็นสมาชิกจึงเป็นส่วนที่สำคัญต่อกระบวนการคิดและแก้ไขปัญหา โดยที่กราฟฟังก์ชั นความเป็นสมาชิกจะสมมาตรหรือไม่สมมาตรกันทุกประการก็ได้ แล้วแต่ชนิดของฟังก์ชันความเป็นสมาชิก ซึ่งฟังก์ชันความเป็นสมาชิกที่ใช้งานทั่วไปมีหลายชนิดดังนี้

ก) ฟังก์ชันความเป็นสมาชิกรูปสามเหลี่ยม (Triangular Membership Function) การสร้างฟังก์ชันความเป็นสมาชิกรูปสามเหลี่ยมสามารถทำได้โดยการกำหนดตำแหน่งจุดมุมทั้ง 3 มุม โดยจุดแรกเป็นจุดที่บอกตำแหน่งของจุดมุมที่ฐานด้านซ้ายของสามเหลี่ยม (a) จุดที่สองบอกถึงจุดยอดของสามเหลี่ยม (b) และจุดที่สามบอกถึงจุดมุมที่ฐานด้านขวาของสามเหลี่ยม (c) ดังตัวอย่างเมื่อกำหนดค่าพารามิเตอร์ $a = 0$, $b = 5$ และ $c = 10$ จะได้กราฟดังที่แสดงในภาพประกอบที่ 2-20



ภาพประกอบ 2-20 กราฟฟังก์ชันความเป็นสมาชิกรูปสามเหลี่ยม เมื่อ a=0, b=5 และ c=10 [15]

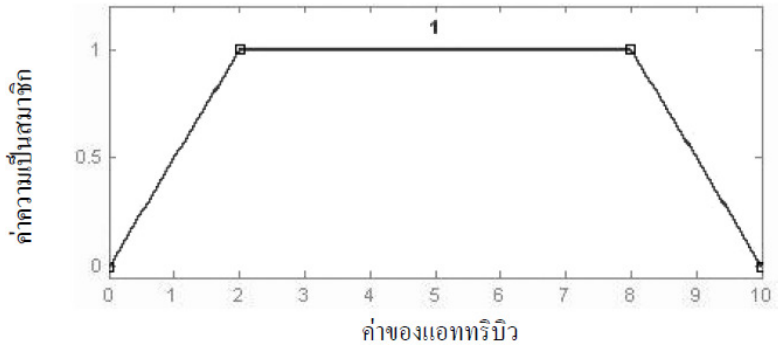
สมการคำนวณหาค่าความเป็นสมาชิกมีดังนี้

$$Triangular(x, a, b, c) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a < x \leq b \\ \frac{c-x}{c-b} & b < x \leq c \\ 0 & x > c \end{cases} \qquad (2.14)$$

โดยที่ x แทน ค่าของแอทริบิว

ข) ฟังก์ชันความเป็นสมาชิกรูปสี่เหลี่ยมคางหมู (Trapezoidal Membership Function) การสร้างฟังก์ชันความเป็นสมาชิกรูปสี่เหลี่ยมคางหมูสามารถทำได้โดยการกำหนดตำแหน่งจุดมุมทั้ง 4 มุม โดยจุดแรกเป็นจุดที่บอกตำแหน่งของจุดมุมล่างด้านซ้าย (a) จุดที่สองบอกถึงจุดมุมบนด้านซ้าย (b) จุดที่สามบอกถึงจุดมุมบนด้านขวา (c) และจุดที่สี่บอกถึงตำแหน่งของจุดมุม

ล่างด้านขวา (*d*) ดังตัวอย่างเมื่อกำหนดค่าพารามิเตอร์ $a = 0$, $b = 2$, $c = 8$ และ $d = 10$ จะได้กราฟดังที่แสดงในภาพประกอบที่ 2-21



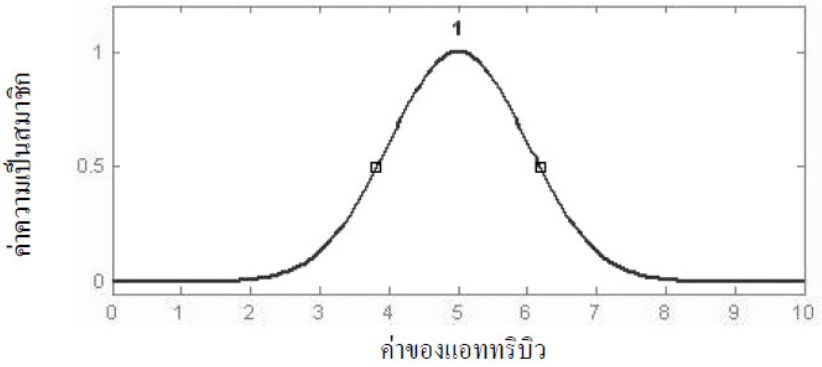
ภาพประกอบ 2-21 ฟังก์ชันความเป็นสมาชิกรูปสี่เหลี่ยมคางหมู [15]
เมื่อ $a = 0$, $b = 2$, $c = 8$ และ $d = 10$

สมการคำนวณหาค่าความเป็นสมาชิกมีดังนี้

$$Trapezoidal(x: a, b, c, d) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \le x < b \\ 1 & b \le x < c \\ \frac{c-x}{b-a} & c \le x < d \\ 0 & x \le d \end{cases} \qquad (2.15)$$

โดยที่ x แทน ค่าของแอทริบิว

ค) ฟังก์ชันความเป็นสมาชิกรูปแบบเกาส์เซียน (Gaussian Membership Function) เป็นฟังก์ชันความเป็นสมาชิกที่รับค่าพารามิเตอร์ทั้งหมด 2 ค่า คือ ค่า m หมายถึงค่าเฉลี่ย และ σ หมายถึง ค่าเบี่ยงเบนมาตรฐาน ดังตัวอย่างเมื่อกำหนดค่าพารามิเตอร์ให้ $m = 5$ และ $\sigma = 1$ จะได้กราฟดังแสดงในภาพประกอบที่ 2-22



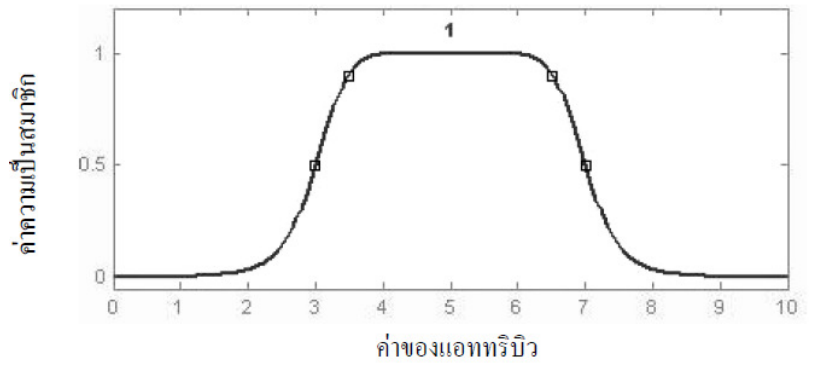
ภาพประกอบ 2-22 ฟังก์ชันความเป็นสมาชิกแบบเกาส์เซียน เมื่อ m=5 และ σ=1 [15]

สมการคำนวณหาค่าความเป็นสมาชิกมีดังนี้

$$Guassian(x: m, \sigma) = exp - \frac{(x-m)^2}{\sigma^2} \qquad (2.16)$$

โดยที่ x แทน ค่าของแอทริบิว

ง) ฟังก์ชันความเป็นสมาชิกรูประฆังคว่ำ (Bell-Shaped Membership Function) ฟังก์ชันความเป็นสมาชิกรูประฆังคว่ำที่รับค่าพารามิเตอร์ทั้งหมด 3 ค่าคือ (a, b, c) ดังตัวอย่างเมื่อกำหนดค่าพารามิเตอร์ a=2, b=4 และ c=5 จะได้กราฟดังแสดงในภาพประกอบที่ 2-23



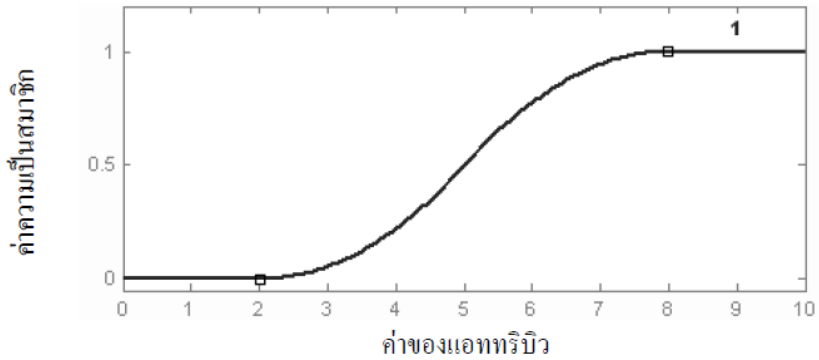
ภาพประกอบ 2-23 ฟังก์ชันความเป็นสมาชิกรูประฆังคว่ำ เมื่อ a=2, b=4 และ c=5 [15]

สมการคำนวณหาค่าความเป็นสมาชิกมีดังนี้

$$Bell - shaped(x: a, b, c) = \frac{1}{1+\left|\frac{x-c}{a}\right|^{2b}} \quad (2.17)$$

โดยที่ x แทน ค่าของแอทริบิว

จ) ฟังก์ชันความเป็นสมาชิกรูปตัวเอส (Smooth Membership Function) ฟังก์ชันความเป็นสมาชิกรูปตัวเอสที่รับพารามิเตอร์ทั้งหมด 2 ค่าคือ (a, b) ดังตัวอย่างเมื่อกำหนดค่าพารามิเตอร์ $a = 2$ และ $b = 8$ จะได้กราฟดังแสดงในภาพประกอบที่ 2-24



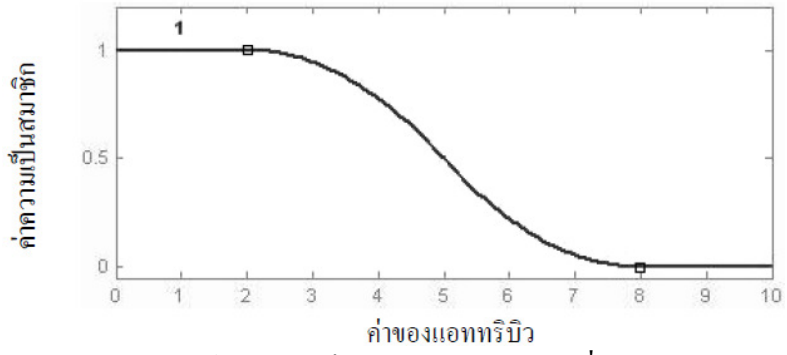
ภาพประกอบ 2-24 ฟังก์ชันความเป็นสมาชิกรูปตัวเอส เมื่อ a=2, b=8 [15]

สมการคำนวณหาค่าความเป็นสมาชิกมีดังนี้

$$Smooth(x: a, b) = \begin{cases} 0 & x < a \\ 2\left(\frac{x-b}{b-a}\right)^2 & a \le x \le \frac{a+b}{2} \\ 1 - 2\left(\frac{x-b}{b-a}\right)^2 & \frac{a+b}{2} \le x \le b \\ 0 & x \ge b \end{cases} \quad (2.18)$$

โดยที่ x แทน ค่าของแอทริบิว

ฉ) ฟังก์ชันความเป็นสมาชิกรูปตัวแซด (Z-Membership Function) ฟังก์ชันความเป็นสมาชิกรูปตัวเอสที่รับพารามิเตอร์ทั้งหมด 2 ค่าคือ (a, b) ดังตัวอย่างเมื่อกำหนดค่าพารามิเตอร์ $a = 2$ และ $b = 8$ จะได้กราฟดังแสดงในภาพประกอบที่ 2-25



ภาพประกอบ 2-25 ฟังก์ชันความเป็นสมาชิกรูปตัวแซด เมื่อ a=2 และ b=8 [15]

สมการคำนวณหาค่าความเป็นสมาชิกมีดังนี้

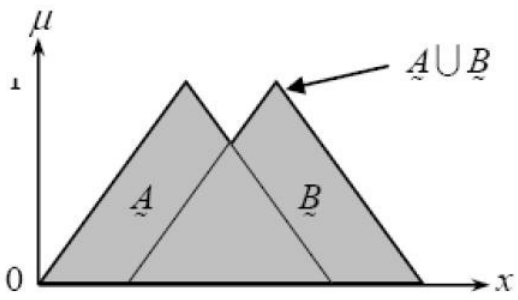
$$Z(x: a, b) = \begin{cases} 1 & x < a \\ 1 - 2\left(\frac{x-b}{b-a}\right)^2 & a \le x \le \frac{a+b}{2} \\ 2\left(\frac{x-b}{b-a}\right)^2 & \frac{a+b}{2} \le x \le b \\ 0 & x \ge b \end{cases} \qquad (2.19)$$

โดยที่ x แทน ค่าของแอทริบิว

2) ตัวดำเนินการของฟัซซี่ลอจิก ประกอบด้วย คำสั่งยูเนียน (Union) อินเตอร์เซกชัน (Intersection) และคำสั่งคอมพลีเมนต์ (Complement) ทั้ง 3 คำสั่งนี้มีคุณสมบัติที่เหมือนกับคำสั่งในเซตทั่วๆไป ซึ่งมีรายละเอียดดังนี้

ก) คำสั่งยูเนียน เป็นคำสั่งในการเชื่อมฟังก์ชันความเป็นสมาชิกของเซตทั้ง 2 เซต รวมเข้าด้วยกัน โดยเขียนให้อยู่ในรูปสมการของฟัซซี่ลอจิกได้ดังนี้

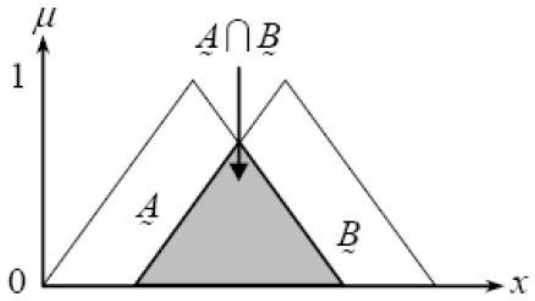
$$\mu_{A \cup B}(x) = \max[\mu_A(x), \mu_B(x)] \qquad (2.20)$$



ภาพประกอบ 2-26 การยูเนียนของฟังก์ชันความเป็นสมาชิกของ A และ B [15]

ข) คำสั่งอินเตอร์เซกชัน เป็นคำสั่งในการเลือกฟังก์ชันความเป็นสมาชิกของเซตทั้ง 2 เซตที่มีค่าเหมือนกันเท่านั้น เขียนให้อยู่ในรูปสมการของฟัซซี่ลอจิกได้ดังนี้

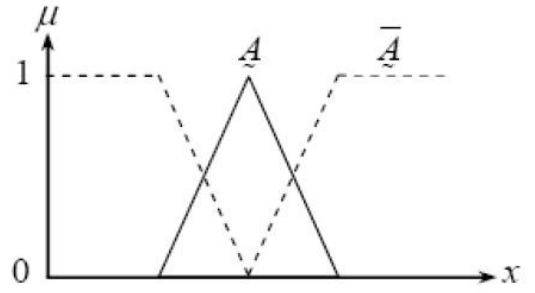
$$\mu_{A\cap B}(x) = \min[\mu_A(x), \mu_B(x)) \qquad (2.21)$$



ภาพประกอบ 2-27 การอินเตอร์เซกชันของฟังก์ชันความเป็นสมาชิกของ A และ B [15]

ค) คำสั่งคอมพลีเมนต์ เป็นคำสั่งที่เลือกฟังก์ชันความเป็นสมาชิกที่ไม่อยู่ในเซตนั้น โดยเขียนให้อยู่ในรูปสมการของฟัซซี่ลอจิกได้ดังนี้

$$\mu_{\bar{A}} = 1 - \mu_A(x) \qquad (2.22)$$



ภาพประกอบที่ 2-28 การคอมพลีเมนต์ของฟังก์ชันความเป็นสมาชิกของ A [15]

บทที่ 3

วิธีดำเนินการ

บทนี้จะกล่าวถึงวิธีดำเนินการในการทำวิทยานิพนธ์ โดยเริ่มจากการเก็บข้อมูลสัญญาณเสียงพูด ถัดไปเป็นการเตรียมข้อมูลสัญญาณเสียงพูด ซึ่งในขั้นตอนนี้จะประกอบไปด้วยขั้นตอนย่อยๆทั้งหมดที่ใช้ในกระบวนการวิเคราะห์สัญญาณเสียงพูด พร้อมทั้งได้แสดงรายละเอียดของการดำเนินการในแต่ละขั้นตอนย่อยไว้อย่างละเอียด สุดท้ายจะกล่าวถึงรายละเอียดของการทดลองและผลการทดลอง เพื่อแสดงให้เห็นถึงความแตกต่างระหว่างผู้พูดชนิด Dysarthria และผู้พูดปกติ ผลจากการศึกษาในบทนี้เป็นปัจจัยสำคัญที่ก่อให้เกิดแนวทางในการออกแบบระบบคำสั่งที่เหมาะสมกับผู้ป่วย ซึ่งประกอบด้วย 1) การคัดแยกคำสั่งเสียงสระโดยวิเคราะห์จากความถี่ฟอร์แมนต์และ 2) การคัดแยกสระเสียงสั้น-เสียงยาวโดยวิเคราะห์จากค่าพลังงานเสียง

3.1 การประมวลผลสัญญาณเบื้องต้น (Pre-processing)

3.1.1 การเก็บข้อมูลสัญญาณเสียงพูด

การบันทึกเสียงพูดสำหรับวิทยานิพนธ์นี้กระทำในห้องที่ไม่มีสัญญาณรบกวนจากภายนอกโดยใช้โปรแกรมการบันทึกซึ่งเขียนขึ้นด้วยโปรแกรม Labview 8.6 ร่วมกับไมโครโฟนยี่ห้อ Genius (Desktop Microphone) ซึ่งมีย่านความถี่ตอบสนองในช่วง 0 – 10 kHz ใช้เวลาในการบันทึกแต่ละคำนาน 2 วินาที เก็บบันทึกในไฟล์ข้อมูลรูปแบบ .lvm เสียงที่พูดออกจากไมโครโฟนจะผ่านโปรแกรมกรองความถี่สูงผ่าน (High-pass Filter) ที่มีความถี่คัตออฟ (Cutoff Frequency) เท่ากับ 200 Hz ทั้งนี้เพื่อให้ได้สัญญาณซึ่งได้ค่าความถี่ฟอร์แมนต์ที่ต้องการเด่นชัดขึ้นสำหรับใช้ในการวิเคราะห์จากสเปกตรัมของสัญญาณเสียงต่อไป ค่าความถี่แซมปลิ้ง (Sampling Frequency) ที่ใช้ในการเก็บข้อมูลมีค่าเท่ากับ 20 kHz ซึ่งเพียงพอสำหรับครอบคลุมแบนด์วิดท์ (Band Width) ของเสียงพูดโดยทั่วไปได้ จากค่าความถี่แซมปลิ้งที่ใช้นี้จะทำให้ได้ตำแหน่งของข้อมูลสัญญาณเสียงพูดแต่ละตำแหน่งห่างกันตำแหน่งละ 0.05 มิลลิวินาที ใช้คำสั่งเป็นเสียงสระในภาษาไทยจำนวน 6 คำ ประกอบด้วยสระเสียงสั้นจำนวน 3 คำ (อึ เออะ อะ) และสระเสียงยาวจำนวน 3 คำ (อือ เออ อา) รวมทั้งสิ้น 6 คำสั่ง ร่วมกับกลุ่มผู้ทดสอบเป็นเพศชาย จำนวน 3 คน คือ

1) ผู้พูดผิดปกติชนิด Dysarthria อายุ 57 ปี อัมพาตครึ่งซีก (Hemiplegia) ระดับ 6 (C6)

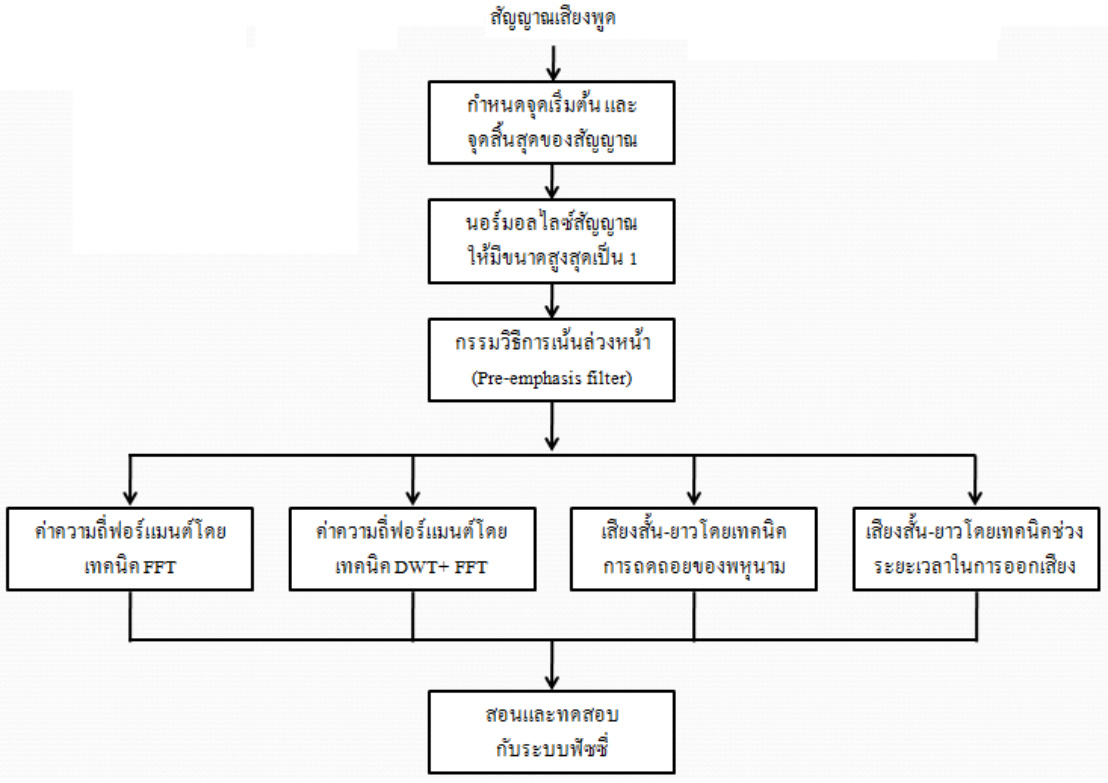
2) ผู้พูดผิดปกติชนิด Dysarthria อายุ 70 ปี อัมพาตครึ่งซีก (Hemiplegia) ระดับ 4 (C4)

3) ผู้พูดปกติ อายุ 25 ปี

โดยให้กลุ่มผู้ทดสอบเปล่งเสียงสระในแต่ละคำซ้ำกันคำละ 5 ครั้ง ข้อมูลสัญญาณเสียงพูดที่ได้ทั้งหมดนี้จะถูกนำมาใช้เป็นข้อมูลสัญญาณเสียงพูดกลุ่มตัวอย่างและข้อมูลสัญญาณเสียงพูดกลุ่มทดสอบ สำหรับใช้ในการสอนและทดสอบกับระบบฟัซซี่ลอจิกต่อไป

3.1.2 การเตรียมข้อมูลสัญญาณเสียงพูด

ข้อมูลสัญญาณเสียงพูดที่ได้จะถูกนำมาวิเคราะห์หาค่าความถี่ฟอร์แมนต์และค่าพลังงานเสียง กระบวนการวิเคราะห์ทั้งหมดถูกเขียนขึ้นด้วยโปรแกรม Labview 8.6 โดยมีขั้นตอนสำหรับการวิเคราะห์แสดงในภาพประกอบ3-1ซึ่งสามารถอธิบายแต่ละขั้นตอนการวิเคราะห์ได้ดังนี้

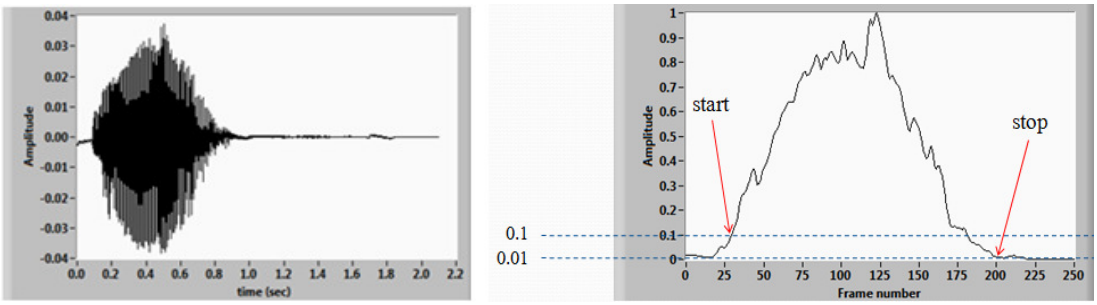


ภาพประกอบ 3-1 ขั้นตอนการวิเคราะห์ข้อมูลสัญญาณเสียงพูด

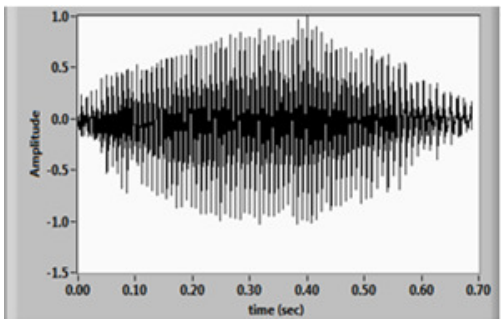
เสียงพูดที่ผ่านกระบวนการเก็บสัญญาณเสียงแล้ว จะได้เป็นข้อมูลสัญญาณเสียงพูดซึ่งจะมีช่วงข้อมูลที่เป็นสัญญาณเงียบทั้งก่อนเริ่มต้นและสุดท้ายของเสียงพูดเสมอ จึงต้องนำมาผ่านเข้าสู่กระบวนการกำหนดจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณก่อน เพื่อให้ได้เฉพาะข้อมูลในช่วงที่เป็นเสียงคำพูดเท่านั้น

3.1.3 การกำหนดจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณ

การกำหนดจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณที่นำมาวิเคราะห์สำหรับวิทยานิพนธ์นี้ จะเลือกใช้วิธีการพิจารณาจากค่าพลังงานกำลังสอง (Square Energy) ระหว่างสัญญาณเสียงพูดกับหน้าต่างชนิดแฮมมิ่ง (Hamming) ที่มีความกว้างเท่ากับ 30 มิลลิวินาที (600 ตำแหน่ง) และระยะเลื่อนเฟรมเท่ากับ 4 มิลลิวินาที (80 ตำแหน่ง) ซึ่งจะทำให้ขนาดของสัญญาณเรียบและมีค่ามากในช่วงที่เป็นเสียงคำพูดและมีค่าน้อยในช่วงที่ไม่ใช่เสียงคำพูด เมื่อได้สัญญาณพลังงานเสียงกำลังสองแล้วขั้นตอนแรกเป็นการนอร์มอลไลซ์ค่าพลังงานเสียงให้มีค่าสูงสุดเท่ากับ 1 จากนั้นจึงพิจารณาจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณเสียง โดยให้ตำแหน่งแรกนับจากต้นเสียงซึ่งให้ค่าพลังงานเสียงเท่ากับ 0.1 เป็นจุดเริ่มต้น และให้ตำแหน่งแรกนับจากปลายเสียงซึ่งให้ค่าพลังงานเสียงเท่ากับ 0.01 เป็นจุดสิ้นสุด ดังแสดงในภาพประกอบ 3-2 โดยภาพประกอบ 3-2ก แสดงสัญญาณเสียงพูดตลอดช่วงการเก็บสัญญาณเสียงสระอือ ภาพประกอบ 3-2ข แสดงค่าพลังงานของสัญญาณเสียงและตำแหน่งของเส้นตรงสองเส้นแสดงจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณที่นำมาวิเคราะห์ ซึ่งเมื่อพิจารณาเฉพาะสัญญาณในช่วงที่ทำการวิเคราะห์แล้ว จะได้สัญญาณดังแสดงในภาพประกอบ 3-2ค



(ก) (ข)



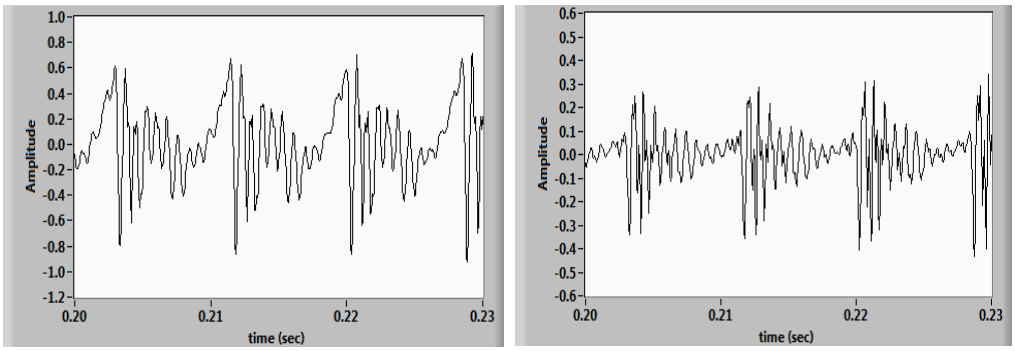
(ค)

ภาพประกอบ 3-2 ตัวอย่างการพิจารณาจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณที่นำมาวิเคราะห์
(ก) สัญญาณเสียงพูดตลอดช่วงการเก็บสัญญาณเสียงสระอือ
(ข) ค่าพลังงานของสัญญาณเสียงในภาพประกอบ 3-2ก
(ค) สัญญาณเสียงพูดในช่วงที่ทำการวิเคราะห์และทำการนอร์มอลไลซ์แล้ว

3.1.4 กรรมวิธีการเน้นล่วงหน้า (Pre-Emphasis)

ในการทำให้อัตราเสียงพูดต่อสัญญาณรบกวน (Signal to Noise Ratio : SNR) ให้มีค่าคงที่นั้น สามารถทำได้โดยการเน้นความถี่ให้มีขนาดสูงขึ้น นอกจากนี้ยังเป็นการลดสเปกตรัมที่เกิดจากแบบจำลองช่องเส้นเสียงและแบบจำลองการแพร่นั้นมีความชันรวมกันประมาณ -6 dB/ Octave[15] ดังนั้นเพื่อลดผลทางสเปกตรัมนี้ สัญญาณเสียงต้องผ่านกรรมวิธีการเน้นล่วงหน้า ซึ่งอาจใช้วิธีการกรองความถี่สูงผ่าน โดยใช้วงจรกรองอันดับหนึ่ง มีฟังก์ชันถ่ายโอนดังสมการ 3.1 ในงานวิจัยนี้กำหนดให้ค่า α = 0.94

$$H(z) = 1 - az^{-1} \quad , 0.93 < a < 0.98 \qquad (3.1)$$



ภาพประกอบ 3-3 ตัวอย่างสัญญาณเสียงขนาด 30 มิลลิวินาทีก่อนผ่านกรรมวิธีการเน้นล่วงหน้า (ซ้าย) และสัญญาณเสียงหลังผ่านกระบวนการเน้นล่วงหน้า(ขวา)

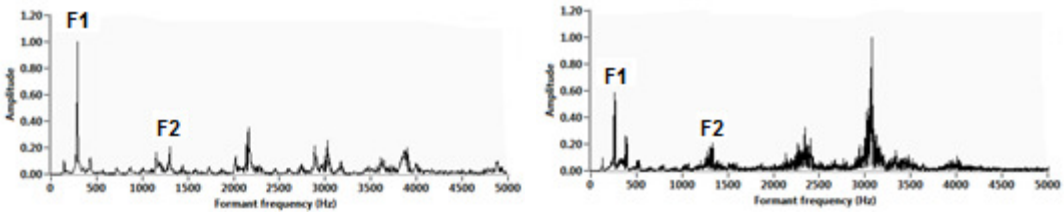
3.2 การคัดแยกคำสั่งเสียงสระโดยวิเคราะห์จากความถี่ฟอร์แมนต์

วัตถุประสงค์ของการศึกษาการคัดแยกคำสั่งเสียงสระโดยวิเคราะห์จากความถี่ฟอร์แมนต์เพื่อสามารถแสดงถึงความแตกต่างของความถี่ในแต่ละคำได้และทราบถึงผลกระทบจากความผิดปกติของผู้ป่วยส่งผลต่อความผิดเพี้ยนของความถี่ฟอร์แมนต์ทั้งสองโดยใช้ความถี่ฟอร์แมนต์ของผู้พูดปกติเป็นเกณฑ์ สำหรับแนวทางการศึกษาจะแบ่งออกเป็น 2 แนวทาง คือ 1) การคัดแยกเสียงสระโดยใช้เทคนิคฟาสฟูเรียร์ทรานฟอร์ม (FFT) และ 2) การคัดแยกเสียงสระโดยใช้เทคนิคเวฟเล็ทร่วมกับฟาสฟูเรียร์ทรานฟอร์ม (DWT + FFT)

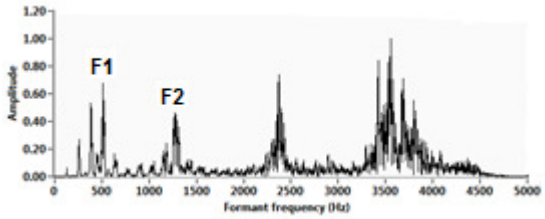
3.2.1 การคัดแยกเสียงโดยใช้เทคนิคฟาสฟูเรียร์ทรานฟอร์ม (FFT)

ความถี่ฟอร์แมนต์เป็นพารามิเตอร์ที่ใช้สำหรับจำแนกเสียงสระที่ต่างกันและเนื่องจากความถี่ฟอร์แมนต์ที่หนึ่งและสองของเสียงสระทั้ง 6 นั้นถ้าเป็นของผู้พูดปกติก็จะอยู่ในช่วงที่ไม่เกิน 2000 Hz แต่สำหรับผู้ป่วย ความถี่ฟอร์แมนต์ทั้งสองอาจจะสูงกว่า 2000 Hzได้ ดังนั้นการทดลองนี้จึงใช้ย่านความถี่ปฏิบัติงานอยู่ในช่วง 0-5000 Hz ในการตรวจสอบความคลาดเคลื่อนของความถี่ฟอร์แมนต์ทั้งสอง

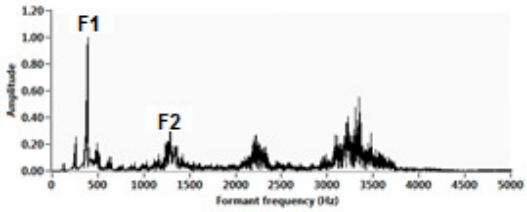
จากการพิจารณาสเปกตรัมของสัญญาณเสียงพูดสำหรับความถี่ฟอร์แมนต์ที่หนึ่งเสียงสระ อึ, อือ, เออะ, เออ นั้นจะมีค่าใกล้เคียงกับผู้พูดปกติ คืออยู่ในช่วง ไม่เกิน 1000 Hz แต่สำหรับสระ อะ, อา นั้นความถี่จะอยู่บริเวณ 1000 Hz ขึ้นไปซึ่งปกติอยู่ในช่วง 700-800 Hz ส่วนความถี่ฟอร์แมนต์ที่สองจะอยู่กระจัดกระจายในช่วง 1000 – 3000 Hz แสดงดังภาพประกอบที่ 3-4 ถึง 3-6 คือสัญญาณเสียงในโดเมนความถี่ของกลุ่มผู้ทดสอบทั้ง 3



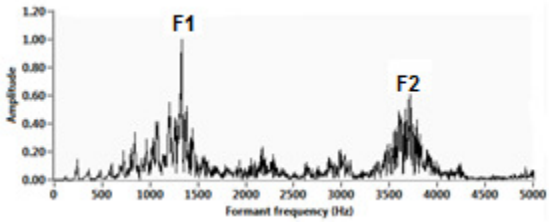
(ก) (ข)


F1
F2
Amplitude
Formant frequency (Hz)

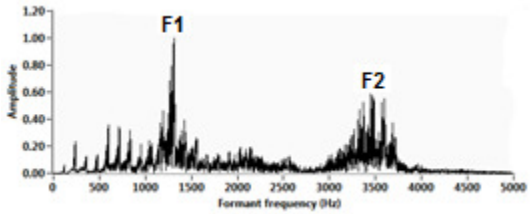
(ค)


F1
F2
Amplitude
Formant frequency (Hz)

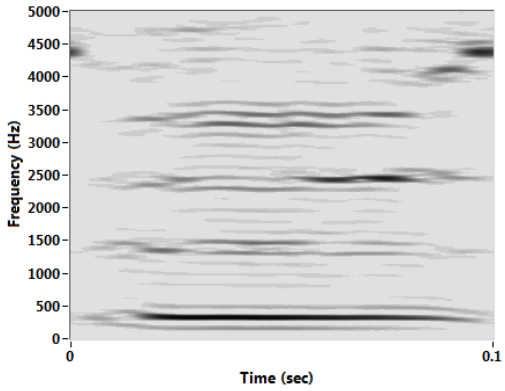
(ง)


F1
F2
Amplitude
Formant frequency (Hz)

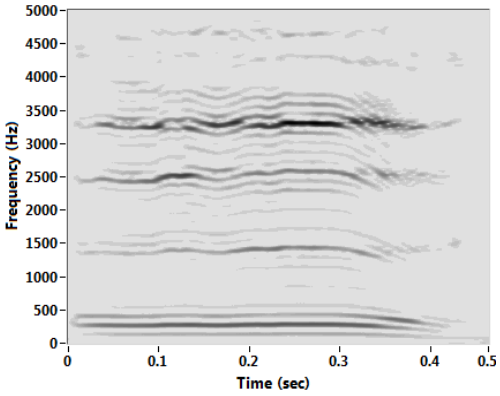
(จ)


F1
F2
Amplitude
Formant frequency (Hz)

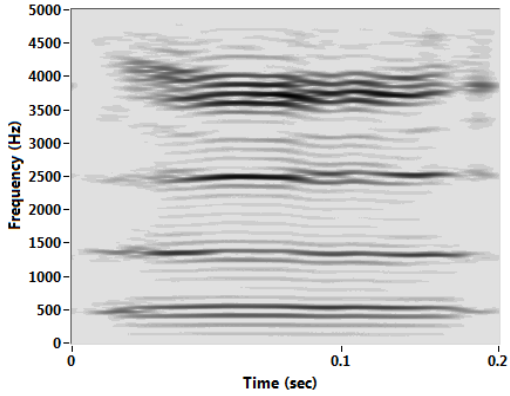
(ฉ)


Frequency (Hz)
Time (sec)

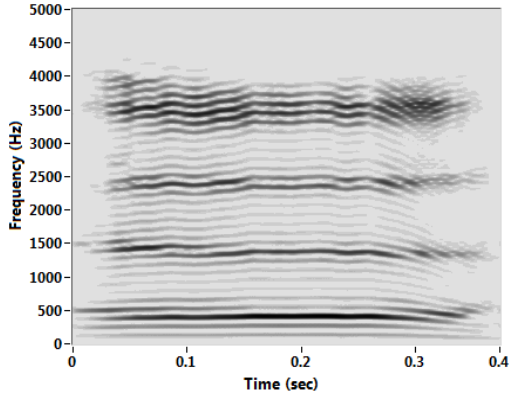
(ช)


Frequency (Hz)
Time (sec)

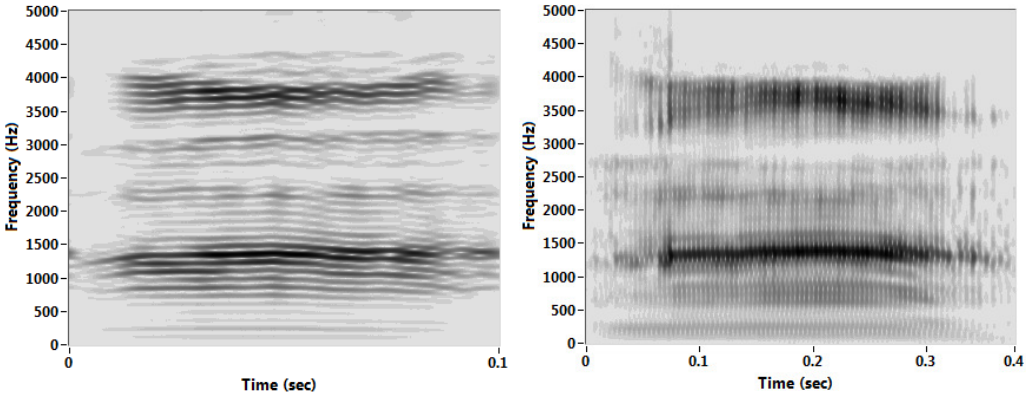
(ซ)


Frequency (Hz)
Time (sec)

(ฌ)


Frequency (Hz)
Time (sec)

(ญ)

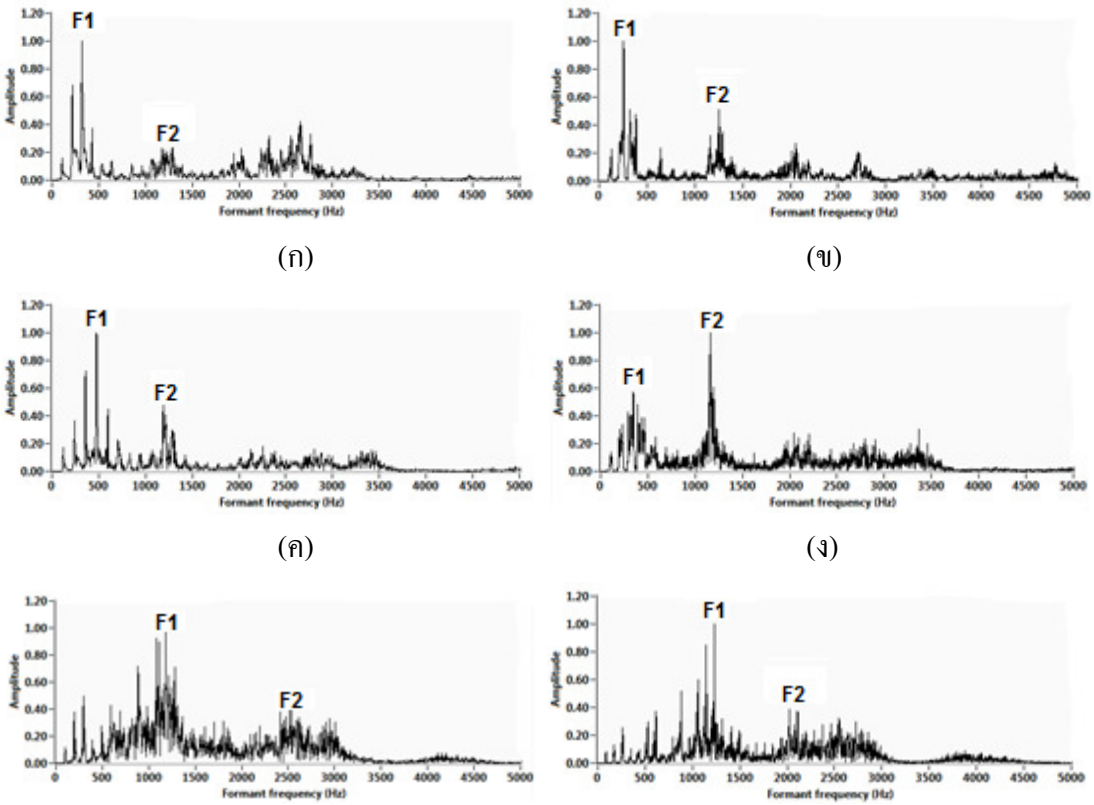


(ฎ) (ฏ)

ภาพประกอบ 3-4 ตัวอย่างความถี่ฟอร์แมนต์ที่หนึ่งและสองของสัญญาณเสียงสระทั้ง 6 คำโดยผู้ทดสอบคนที่หนึ่ง (Dysarthric Speech 1)

ก)-ฉ) ตัวอย่างสัญญาณเสียงพูดในโดเมนความถี่ของเสียงสระ อึ, อือ, เออะ, เออ, อะ, อา ตามลำดับ

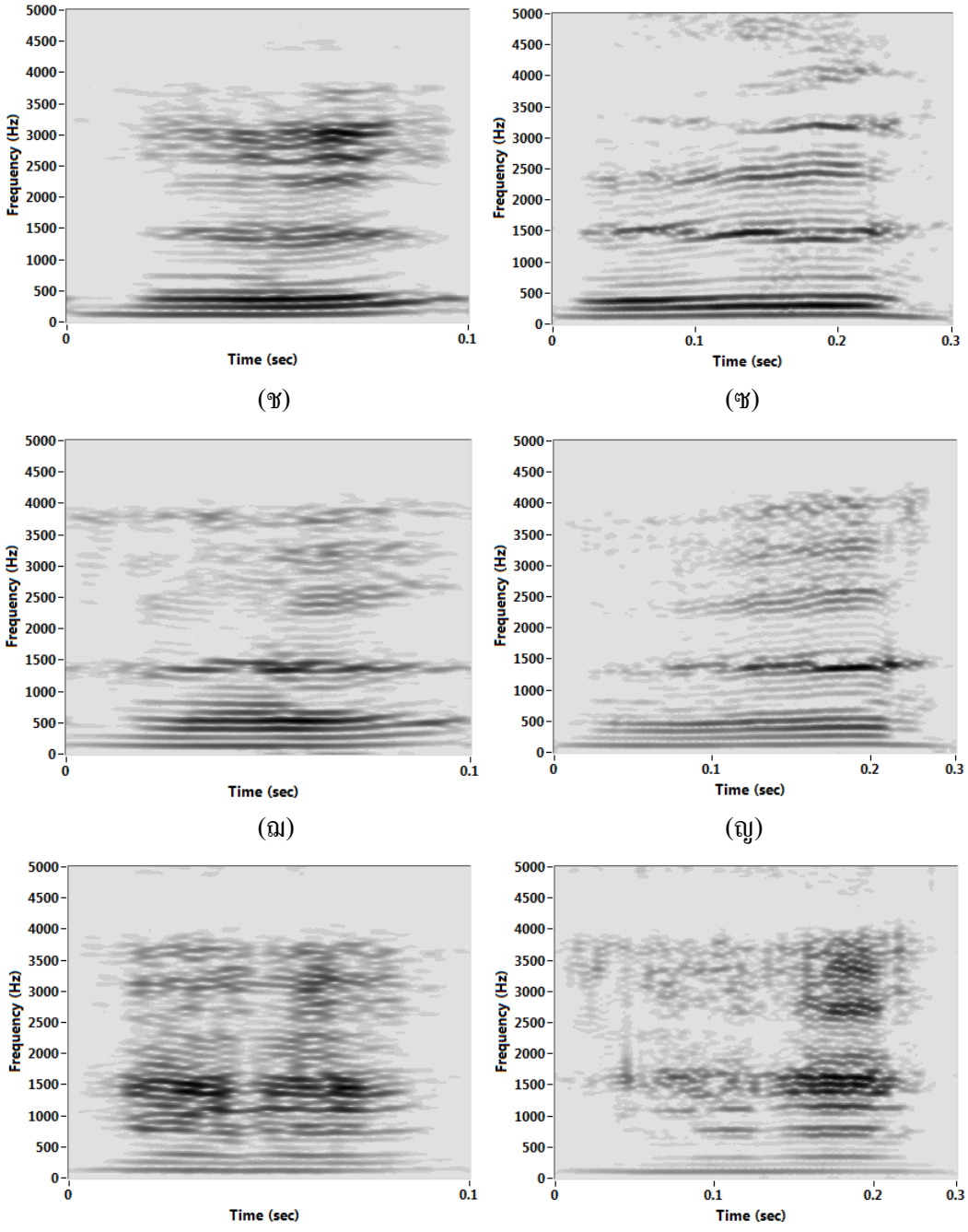
ช)-ฏ) ตัวอย่างสัญญาณเสียงพูดในรูปแบบสเปกโตรแกรมของเสียงสระ อึ, อือ, เออะ, เออ, อะ, อา ตามลำดับ



(ก) (ข)

(ค) (ง)

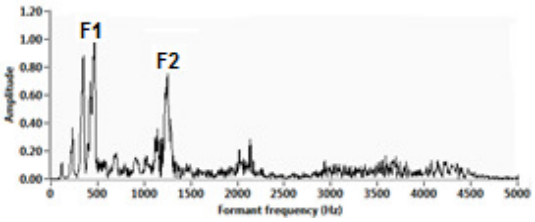
(จ) (ฉ)



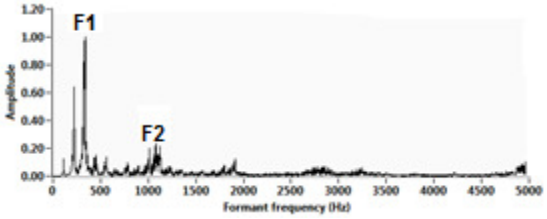
ภาพประกอบ 3-5 ตัวอย่างความถี่ฟอร์แมนต์ที่หนึ่งและสองของสัญญาณเสียงสระทั้ง 6 คำโดยผู้ทดสอบคนที่สอง (Dysarthric Speech 2)

ก)-ฉ) ตัวอย่างสัญญาณเสียงพูดในโดเมนความถี่ของเสียงสระ อึ, อือ, เออะ, เออ, อะ, อา ตามลำดับ

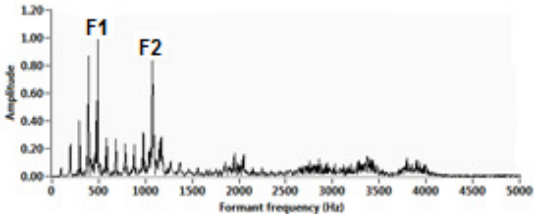
ช)-ฏ) ตัวอย่างสัญญาณเสียงพูดในรูปแบบสเปกโตรแกรมของเสียงสระ อึ, อือ, เออะ, เออ, อะ, อา ตามลำดับ


F1
F2
Amplitude
Formant frequency (Hz)

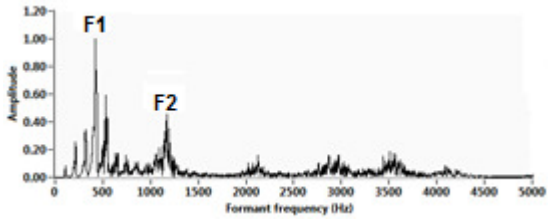
(ก)


F1
F2
Amplitude
Formant frequency (Hz)

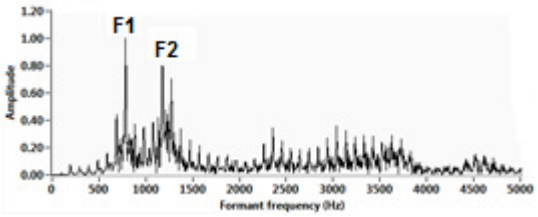
(ข)


F1
F2
Amplitude
Formant frequency (Hz)

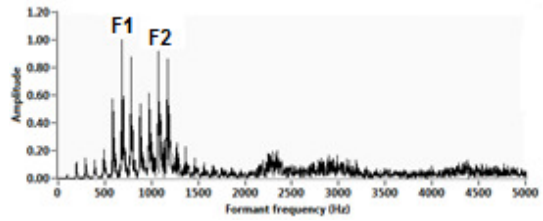
(ค)


F1
F2
Amplitude
Formant frequency (Hz)

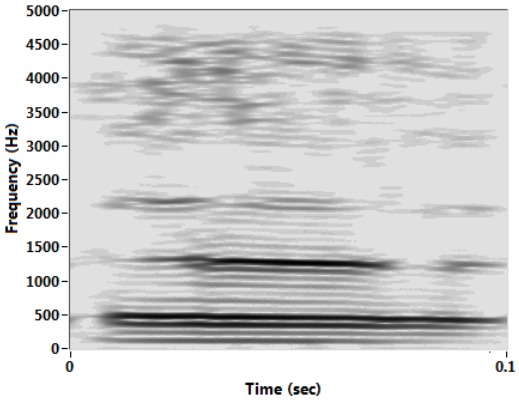
(ง)


F1
F2
Amplitude
Formant frequency (Hz)

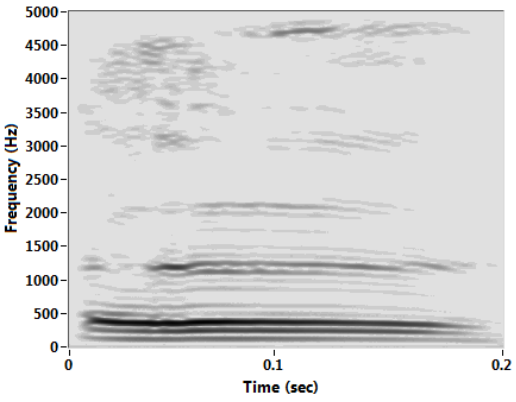
(จ)


F1
F2
Amplitude
Formant frequency (Hz)

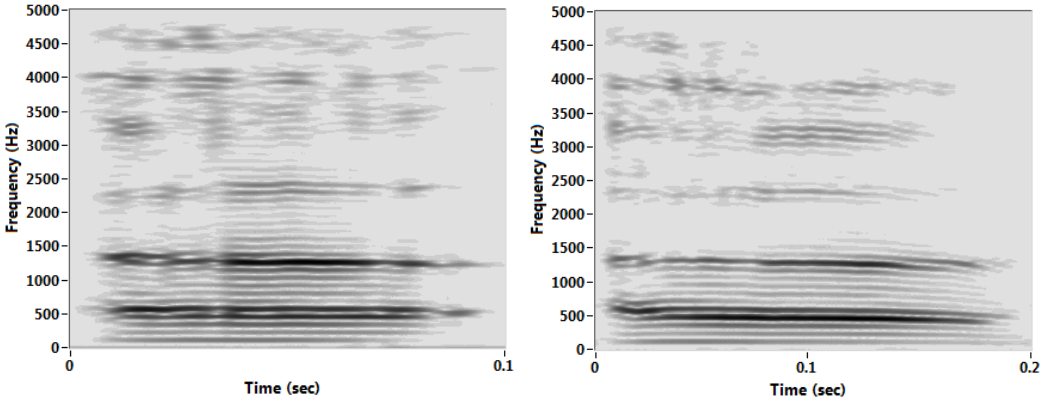
(ฉ)


Frequency (Hz)
Time (sec)

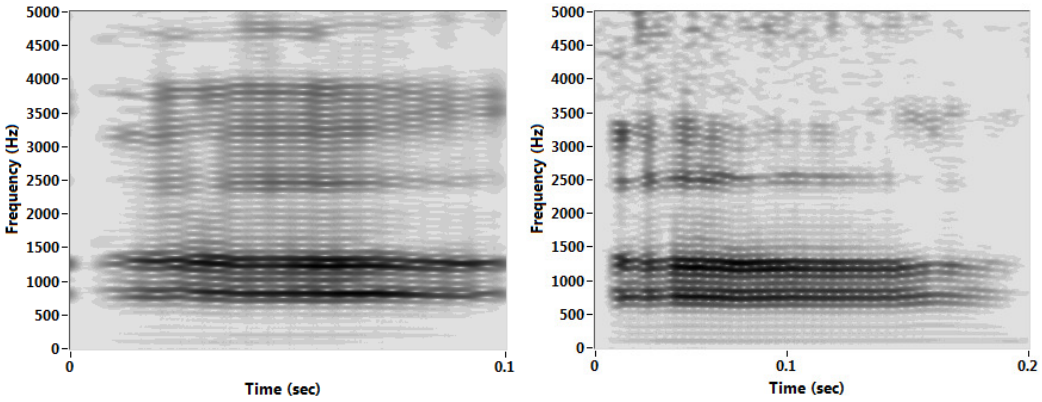
(ช)


Frequency (Hz)
Time (sec)

(ซ)



(ฌ) (ญ)



(ฎ) (ฏ)

ภาพประกอบ 3-6 ตัวอย่างความถี่ฟอร์แมนต์ที่หนึ่งและสองของสัญญาณเสียงสระทั้ง 6 คำโดยผู้ทดสอบคนที่สาม (Normal Speech)

ก)-ฉ) ตัวอย่างสัญญาณเสียงพูดในโดเมนความถี่ของเสียงสระ อึ, อือ, เออะ, เออ, อะ, อา ตามลำดับ

ช)-ฏ) ตัวอย่างสัญญาณเสียงพูดในรูปแบบสเปกโตรแกรมของเสียงสระ อึ, อือ, เออะ, เออ, อะ, อา ตามลำดับ

จากขั้นตอนทั้งหมด สามารถสรุปค่าความถี่ฟอร์แมนต์แยกตามรายบุคคลได้ดังตารางที่ 3-1 ถึง 3-3 (หมายเหตุ; เสียง อึ-อือ, เออะ-เออ และ อะ-อา ถือว่าเป็นเสียงเดียวกัน ต่างกันที่ช่วงเวลาในการออกเสียงเท่านั้น)

ตารางที่ 3-1 ความถี่ฟอร์แมนต์ที่ 1 และ 2 (F1, F2) จำแนกตามกลุ่มเสียงสระ ของผู้ป่วยคนที่หนึ่ง (Dysarthric Speech 1)

เสียงสระ	อึ-อือ		เออะ-เออ		อะ-อา	
ครั้งที่พูด	F1(Hz)	F2(Hz)	F1(Hz)	F2(Hz)	F1(Hz)	F2(Hz)
1	296	2380	505	1256	1329	3700
2	332	3310	397	1330	1298	3770
3	318	2392	526	1304	1348	3580
4	310	2352	407	2405	1350	3714
5	323	3240	415	2406	1333	3660
6	309	2417	404	1335	1426	3636
7	278	3226	398	1334	1340	3648
8	270	2448	403	1349	1316	3593
9	288	3315	407	1350	1348	3540
10	286	3530	422	2370	1409	3640

ตารางที่ 3-2 ความถี่ฟอร์แมนต์ที่ 1 และ 2 (F1, F2) จำแนกตามกลุ่มเสียงสระ ของผู้ป่วยคนที่สอง (Dysarthric Speech 2)

เสียงสระ	อึ-อือ		เออะ-เออ		อะ-อา	
ครั้งที่พูด	F1(Hz)	F2(Hz)	F1(Hz)	F2(Hz)	F1(Hz)	F2(Hz)
1	220	1397	393	1180	1555	3675
2	360	1410	529	1190	1530	3050
3	388	1357	520	1300	1310	3040
4	367	1345	420	1260	1593	3082
5	394	1452	542	1362	1600	3210
6	275	1360	433	1310	1495	3195
7	380	1390	474	1350	1582	2600
8	294	1440	414	1365	1548	2955
9	291	1325	392	1314	1555	3675
10	220	1397	373	1324	1530	3050

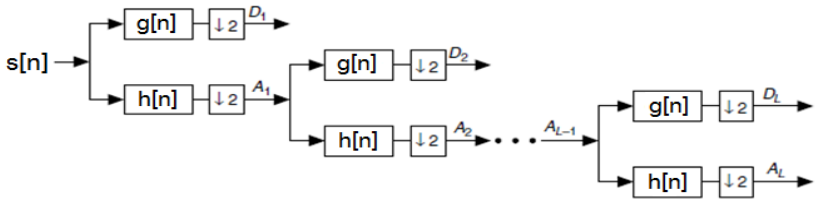
ตารางที่ 3-3 ความถี่ฟอร์แมนต์ที่ 1 และ 2 (F1, F2) จำแนกตามกลุ่มเสียงสระของผู้พูดปกติ (Normal Speech)

เสียงสระ	อึ-อือ		เออะ-เออ		อะ-อา	
ครั้งที่พูด	F1(Hz)	F2(Hz)	F1(Hz)	F2(Hz)	F1(Hz)	F2(Hz)
1	347	1158	528	1292	810	1314
2	398	1250	552	1214	780	1340
3	407	1230	504	1236	800	1298
4	359	1195	542	1200	800	1200
5	369	1228	553	1200	783	1273
6	362	1212	420	1265	780	1250
7	357	1295	539	1190	792	1205
8	360	1306	453	1243	766	1150
9	360	1426	449	1130	776	1162
10	347	1158	453	1250	810	1314

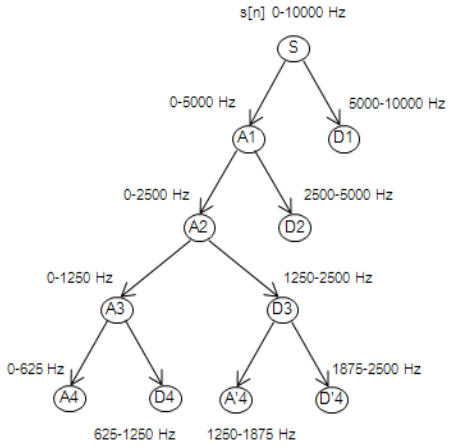
3.2.2 การคัดแยกเสียงโดยใช้เทคนิคเวฟเล็ทร่วมกับฟาสฟูเรียร์ทรานฟอร์ม (DWT + FFT)

จากการทดลองก่อนหน้านี้ ทำให้เราทราบถึงย่านความถี่ฟอร์แมนต์ที่หนึ่งและสองของผู้ป่วย การทดลองนี้จึงได้นำเทคนิคเวฟเล็ทมาช่วยในการแยกส่วนประกอบของความถี่ ซึ่งจะกระทำกับสัญญาณในโดเมนเวลา ผลที่ได้คือสามารถคัดแยกรายละเอียดของความถี่ในช่วงที่เราต้องการได้ดีขึ้นและขจัดสัญญาณในช่วงความถี่ที่ไม่เกี่ยวข้องออกไป ทำให้ความถี่ฟอร์แมนต์ที่ได้หลังจากการทำฟาสฟูเรียร์ทรานฟอร์มมีความแม่นยำและน่าเชื่อถือมากขึ้น สำหรับการทดลองนี้จะแสดงเฉพาะวิธีการและขั้นตอนของผู้ป่วยคนที่หนึ่งกับเสียงสระอา เท่านั้น ถัดไปจะเป็นผลสรุปความถี่ฟอร์แมนต์ทั้งหมดเพื่อเปรียบเทียบระหว่างการทดลองทั้งสอง ซึ่งมีราย ละเอียดและขั้นตอนการทดลองดังนี้

ที่ขั้นตอน Wavelet Decomposition งานวิจัยนี้ใช้เวฟเล็ทแม่แบบ Harr ในการแยกส่วนประกอบความถี่ของสัญญาณ ซึ่งสามารถแยกส่วนประกอบความถี่ออกมาเป็นสัญญาณในแต่ละระดับดังแสดงในภาพประกอบที่ 3-7 และ 3-8

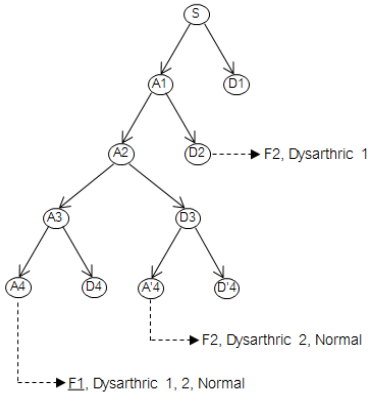


ภาพประกอบ 3-7 แผนภูมิการกระจายเวฟเล็ทสำหรับการแยกส่วนประกอบความถี่ [13]

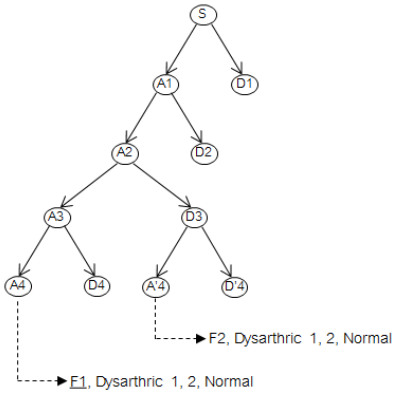


ภาพประกอบ 3-8 แผนภูมิต้นไม้สำหรับการแยกส่วนประกอบความถี่

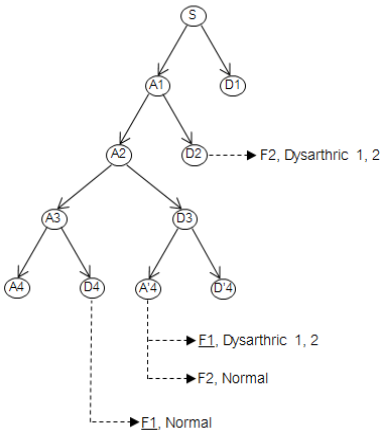
สัญญาณ s[n] เป็นสัญญาณที่ได้มาจากขั้นตอน Preprocessing เมื่อทำการแยกส่วนประกอบความถี่ด้วยวิธีเวฟเล็ทจะได้ส่วนประกอบสัญญาณความถี่สูง g[n] และส่วนประกอบสัญญาณความถี่ต่ำ h[n] โดยมีอัตราสุ่ม (Sampling Rate) ของสัญญาณลดลงเหลือครึ่งหนึ่งของอัตราสุ่มเดิม หลังจากนั้นจึงนำส่วนประกอบสัญญาณความถี่ต่ำไปแยกส่วนประกอบความถี่ของสัญญาณในระดับต่อไปจนถึงระดับที่ต้องการ งานวิจัยนี้ได้วิเคราะห์ระดับสัญญาณรายละเอียดอยู่ที่ระดับ 4 (Wavelet Decomposition level 4) หลังจากนั้นสัญญาณก็จะถูกส่งต่อไปยังขั้นตอน Formant analysis เฉพาะบางช่วงสัญญาณโดยพิจารณาจากคุณลักษณะของความถี่ฟอร์แมนต์ทั้งสอง จากตารางที่ 3-1 ถึง 3-3 ซึ่งสามารถสรุปย่านความถี่ปฏิบัติงานในแต่ละคำโดยใช้แผนภูมิต้นไม้ ดังรูปที่ 3-9 ถึง 3-11 (กำหนดให้ F1, F2 = ความถี่ฟอร์แมนต์ที่ 1, 2 Dysarthric 1, 2 คือ ผู้พูดชนิด Dysarthric Speech คนที่ 1, 2 และ Normal คือ ผู้พูดปกติ ตามลำดับ)



ภาพประกอบ 3-9 การเลือกช่วงความถี่ปฏิบัติงานโดยใช้แผนภูมิการกระจายเวฟเล็ท (A4, A′4, D2) เปรียบเทียบระหว่างผู้ทดสอบทั้ง 3 (เสียงสระ “อึ-อือ”)

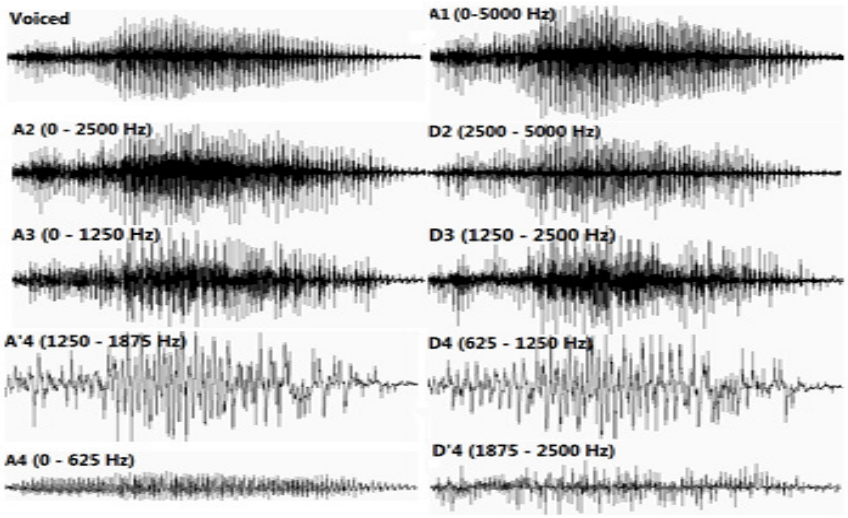


ภาพประกอบ 3-10 การเลือกช่วงความถี่ปฏิบัติงานโดยใช้แผนภูมิการกระจายเวฟเล็ท (A4, A′4) เปรียบเทียบระหว่างผู้ทดสอบทั้ง 3 (เสียงสระ “เออะ-เออ”)



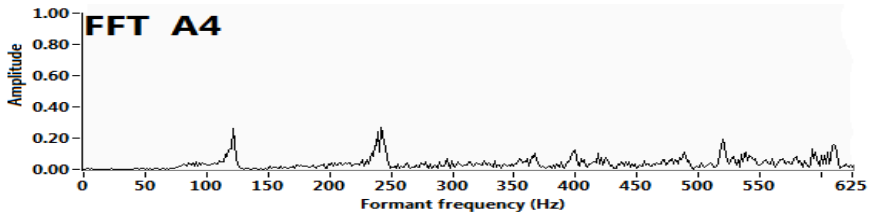
ภาพประกอบ 3-11 การเลือกช่วงความถี่ปฏิบัติงานโดยใช้แผนภูมิการกระจายเวฟเล็ท (D4, A′4, D2) เปรียบเทียบระหว่างผู้ทดสอบทั้ง 3 (เสียงสระ “อะ-อา”)

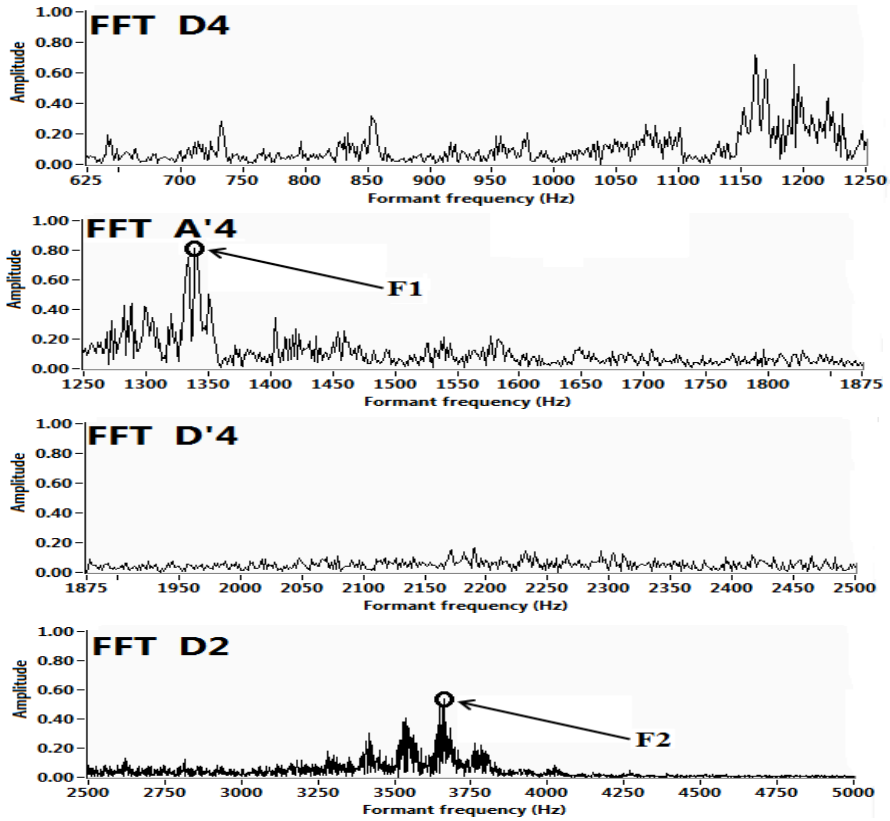
ยกตัวอย่างการเลือกช่วงปฏิบัติงาน จากภาพประกอบที่ 3-11 ในการวิเคราะห์เสียงสระ “อะ-อา” กรณีผู้ป่วยทั้ง สองจะเห็นว่าความถี่ฟอร์แมนต์ที่ 1 อยู่ในช่วง 1250 Hz ขึ้นไป และความถี่ฟอร์แมนต์ที่ 2 อยู่ในช่วง 2500 Hz ขึ้นไป ดังนั้นในการเลือกความถี่ปฏิบัติงาน จึงเลือกเฉพาะช่วง A′4 และ D2 ซึ่งแตกต่างกับผู้พูดปกติ จะเลือกเฉพาะช่วง D4 และ A′4 เมื่อได้ช่วงความถี่ที่ต้องการแล้ว ถัดไปเป็นขั้นตอนการแยกส่วนประกอบความถี่ในโดเมนเวลา แสดงดังภาพประกอบที่ 3-12 (ในการใช้งานจริงไม่ได้ใช้หมดทุกช่วง เพียงแต่แสดงให้เห็นถึงรายละเอียดหลังจากการแยกเท่านั้น)



ภาพประกอบ 3-12 สัญญาณรายละเอียดที่ได้จากการแปลงเวฟเล็ทของเสียงสระ “อา” (Dysarthric Speech 1)

หลังจากเลือกช่วงความถี่ปฏิบัติงานแล้วสัญญาณเสียงในแต่ละช่วงก็จะถูกนำมาแยกความถี่ฟอร์แมนต์โดยใช้เทคนิค FFT ดังภาพประกอบที่ 3-13 (แสดงเฉพาะช่วง D2, D4, D′4, A4 และ A′4)





ภาพประกอบ 3-13 สัญญาณเสียงสระ “อา” (Dysarthric Speech 1) ในโดเมนความถี่แต่ละช่วง

จากขั้นตอนทั้งหมด สามารถสรุปค่าความถี่ฟอร์แมนต์แยกตามรายบุคคลได้ดังตารางที่ 3-4 ถึง 3-5 (หมายเหตุ; เสียง อึ-อือ, เออะ-เออ และ อะ-อา ถือว่าเป็นเสียงเดียวกัน ต่างกันที่ช่วงเวลาในการออกเสียงเท่านั้น)

ตารางที่ 3-4 ความถี่ฟอร์แมนต์ที่ 1 และ 2 (F1, F2) จำแนกตามกลุ่มเสียงสระของผู้ป่วยคนที่หนึ่ง (Dysarthric Speech 1)

เสียงสระ	อึ-อือ		เออะ-เออ		อะ-อา	
ครั้งที่พูด	F1(Hz)	F2(Hz)	F1(Hz)	F2(Hz)	F1(Hz)	F2(Hz)
1	296	2360	509	1256	1333	3671
2	335	3323	400	1330	1303	3837
3	322	3325	528	1333	1355	3576
4	322	2397	407	2405	1368	3715
5	312	3286	405	1335	1345	3665
6	309	2420	398	1335	1426	3635
7	278	3226	405	1349	1342	3648

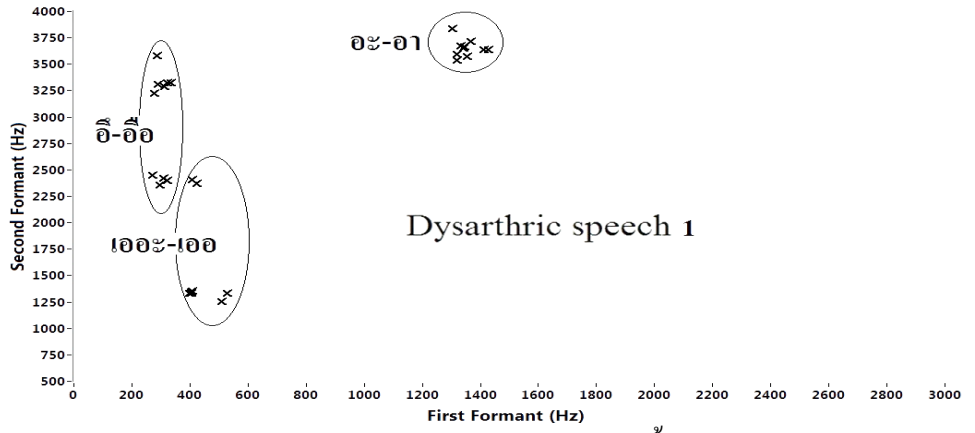
8	270	2448	408	1352	1318	3591
9	288	3308	422	2370	1319	3540
10	286	3580	400	1330	1410	3635

ตารางที่ 3-5 ความถี่ฟอร์แมนต์ที่ 1 และ 2 (F1, F2) จำแนกตามกลุ่มเสียงสระของผู้ป่วยคนที่สอง (Dysarthric Speech 2)

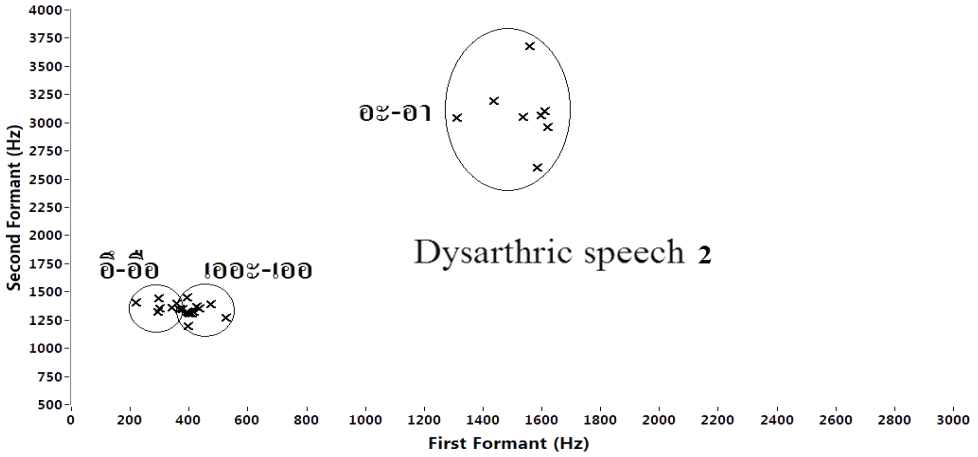
เสียงสระ	อึ-อือ		เออะ-เออ		อะ-อา	
ครั้งที่พูด	F1(Hz)	F2(Hz)	F1(Hz)	F2(Hz)	F1(Hz)	F2(Hz)
1	220	1405	395	1312	1558	3676
2	358	1400	400	1195	1535	3048
3	340	1359	526	1307	1310	3043
4	368	1347	424	1267	1595	3063
5	394	1450	409	1367	1608	3100
6	300	1355	433	1310	1435	3195
7	378	1342	474	1352	1585	2602
8	295	1445	414	1393	1620	2957
9	292	1325	392	1320	1595	3063
10	300	1355	395	1326	1535	3048

ตารางที่ 3-6 ความถี่ฟอร์แมนต์ที่ 1 และ 2 (F1, F2) จำแนกตามกลุ่มเสียงสระของผู้พูดปกติ (Normal Speech)

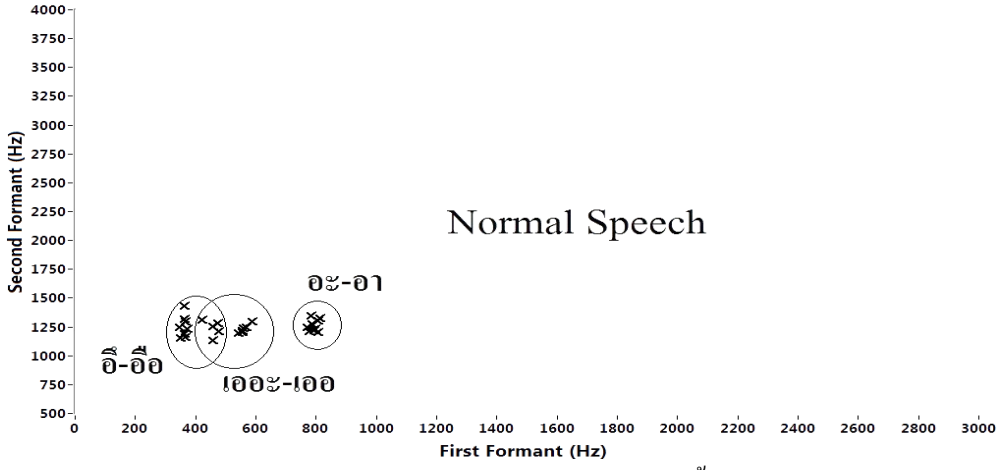
เสียงสระ	อึ-อือ		เออะ-เออ		อะ-อา	
ครั้งที่พูด	F1(Hz)	F2(Hz)	F1(Hz)	F2(Hz)	F1(Hz)	F2(Hz)
1	350	1158	588	1295	811	1322
2	346	1250	555	1222	781	1343
3	472	1280	566	1247	806	1305
4	362	1192	557	1241	805	1205
5	372	1230	553	1205	785	1277
6	365	1163	422	1310	795	1230
7	364	1295	541	1200	792	1220
8	362	1318	475	1210	768	1245
9	362	1429	455	1134	777	1208
10	362	1192	457	1252	806	1305



ภาพประกอบ 3-14 กราฟแจกแจงแบบกระจายของเสียงสระทั้ง 6 คำ (Dysarthric Speech 1)



ภาพประกอบ 3-15 กราฟแจกแจงแบบกระจายของเสียงสระทั้ง 6 คำ (Dysarthric Speech 2)



ภาพประกอบ 3-16 กราฟแจกแจงแบบกระจายของเสียงสระทั้ง 6 คำ (Normal Speech)

(กำหนดให้ แกนตั้ง คือ ความถี่ฟอร์แมนต์ที่ 2 และแกนนอน คือ ความถี่ฟอร์แมนต์ที่ 1)

3.3 การคัดแยกสระเสียงสั้น-เสียงยาวโดยวิเคราะห์จากค่าพลังงานเสียง

วัตถุประสงค์ของการศึกษาการคัดแยกสระเสียงสั้น-เสียงยาวโดยวิเคราะห์จากค่าพลังงานเสียงเพื่อคัดแยกสระเสียงสั้นและยาวอันเนื่องมาจากผลกระทบจากความผิดปกติของผู้ป่วย ส่งผลให้ช่วงระยะเวลาในการเปล่งเสียงสั้นและเสียงยาวมีความแตกต่างกันมากกว่าคนปกติกล่าวคือผู้ป่วยไม่สามารถเปล่งเสียงครึ่งๆกลางๆเหมือนกับผู้พูดปกติได้ ดังนั้นงานวิจัยนี้จึงทำการทดลองโดยแบ่งออกเป็น 2 วิธี เปรียบเทียบกันคือ 1) วิธีการถดถอยของพหุนาม และ 2) วิธีการวิเคราะห์จากช่วงระยะเวลาในการออกเสียง ซึ่งมีรายละเอียดดังนี้

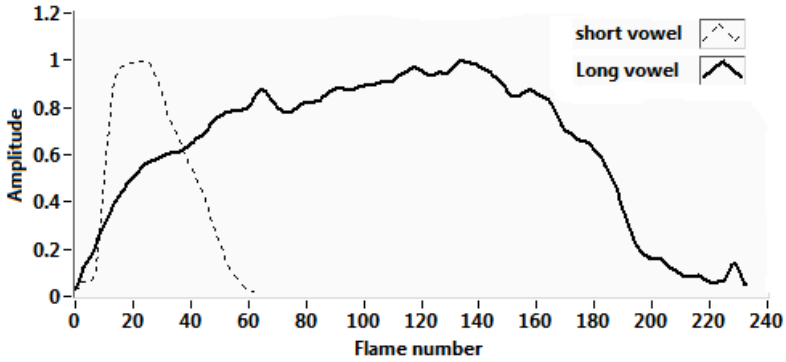
3.3.1 วิธีการถดถอยของพหุนาม (Polynomial regression)

การทดลองนี้ เลือกใช้การถดถอยพหุนามอันดับสอง เพราะลักษณะการกระจายของข้อมูลพลังงานเสียง ไม่เป็นเชิงเส้นที่มีลักษณะคล้ายรูปพาราโบล่าร์คว่ำ ที่มีการเปลี่ยนแปลงระดับพลังงานเสียง ระหว่างช่วงที่ออกเสียงเมื่อนำฟังก์ชันพหุนามอันดับสองมาประมาณค่า จะมีค่าใกล้เคียงกับ ฟังก์ชันพลังงานเสียงและสามารถนำฟังก์ชันประมาณค่าที่ได้ ไปวิเคราะห์เพื่อแยกเสียงสระเสียงสั้นและเสียงสระเสียงยาว

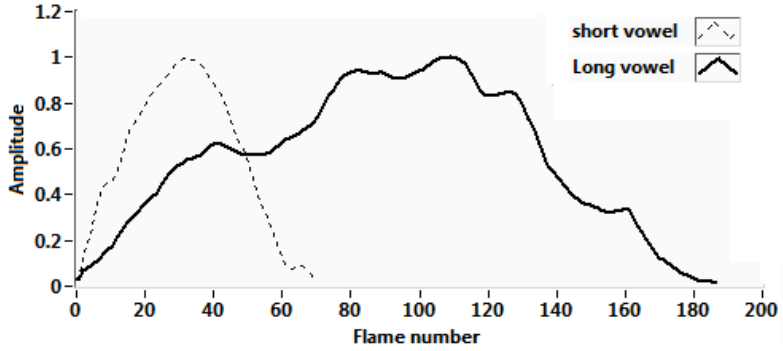
โดยการถดถอยพหุนามอันดับสอง จะทำให้ได้ค่าสัมประสิทธิ์เป็น a_0, a_1, และ a_2 แต่เนื่องจาก a_0 เป็นส่วนตัดแกน y ซึ่งคือจุดเริ่มต้นของพลังงานเสียง ไม่เป็นส่วนสำคัญของการบ่งชี้ลักษณะฟังก์ชันพลังงานเสียง อีกทางหนึ่งจากการพิจารณาค่าที่ได้จากการทดลอง สัมประสิทธิ์ a_2 ค่าเดียวก็เพียงพอต่อการคัดแยกสระเสียงสั้นและเสียงยาวได้ ดังนั้นงานวิจัยนี้จึงพิจารณาเฉพาะสัมประสิทธิ์ a_2 เท่านั้น สำหรับการทดลองนี้จะแสดงเฉพาะวิธีการและขั้นตอนของผู้ป่วยคนที่หนึ่ง (Dysarthric Speech 1) เท่านั้น ซึ่งมีรายละเอียดดังนี้

3.3.1.1 การหาค่าพลังงานเสียงกำลังสอง

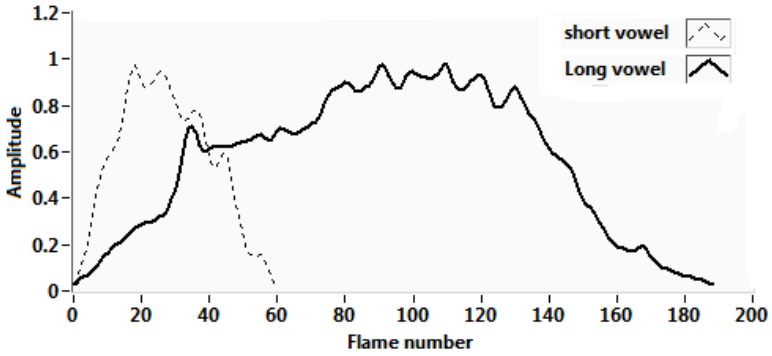
การคำนวณหาค่าพลังงานเสียง จะทำทีละกรอบเสียงพูด โดยมีตัวอย่างกรอบเสียงพูดละ 600 Samples และค่าพลังงานเสียงในกรอบต่อไปจะถูกเลื่อนไปอีก 80 Samples ซึ่งการคำนวณหาค่าพลังงานของสัญญาณเสียงกำลังสอง (Square Energy) จะได้ลักษณะพลังงานเสียงดังนี้



(ก) เสียง “อึ-อือ”



(ข) เสียง “เออะ-เออ”

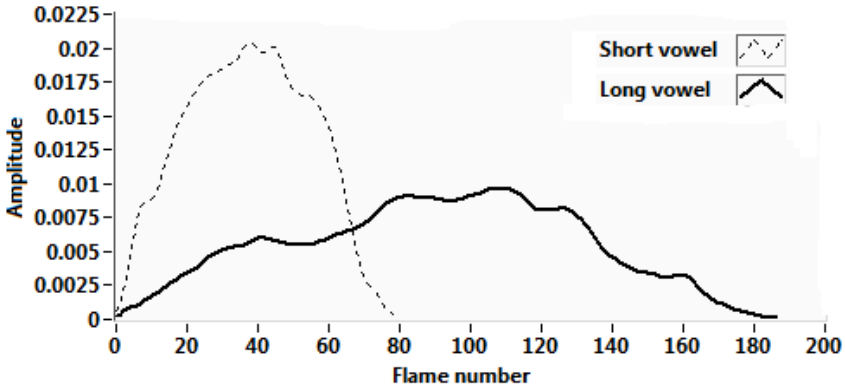


(ค) เสียง “อะ-อา”

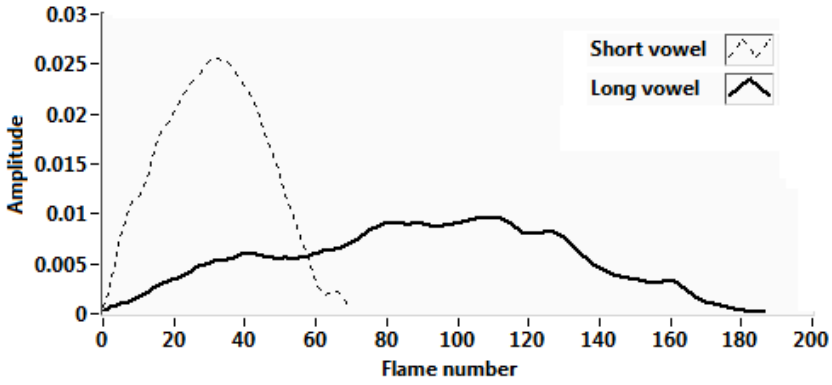
ภาพประกอบ 3-17 ลักษณะพลังงานเสียงแบบ Square Energy

3.3.1.2 การปรับบรรทัดฐานพื้นที่ใต้ฟังก์ชันพลังงานเสียง (Normalized Energy function)

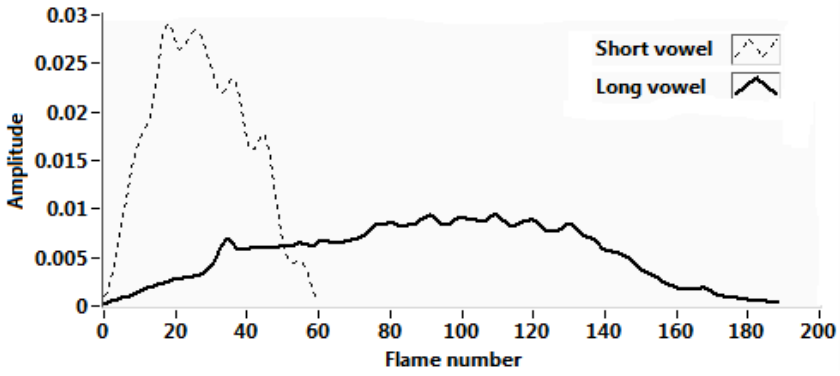
เมื่อได้ลักษณะพลังงานเสียง นำพลังงานเสียงที่ได้มาปรับบรรทัดฐานพลังงาน โดยให้พื้นที่ใต้ฟังก์ชันพลังงานเสียง มีค่าเท่ากับ 1 ทุกๆเสียง ตามสมการที่ 2.15 ซึ่งจะทำให้เห็นความแตกต่างพลังงานเสียงสระเสียงสั้นและสระเสียงยาวมากขึ้น ดังนี้



(ก) เสียง “อึ-อือ”



(ข) เสียง “เออะ-เออ”



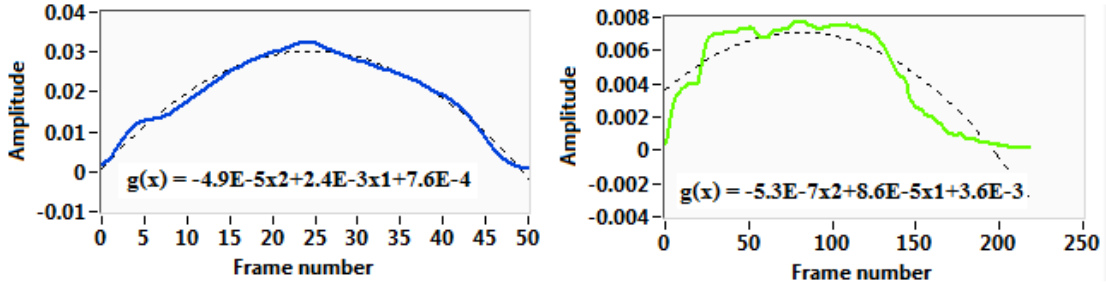
(ค) เสียง “อะ-อา”

ภาพประกอบ 3-18 ลักษณะพลังงานเสียงแบบ Square Energy ที่ผ่านการ Normalize ให้พื้นที่ function เป็น 1 แล้ว

3.3.1.3 การถดถอยแบบพหุนามของพลังงานเสียง

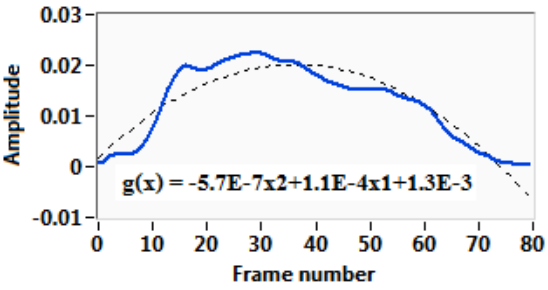
เมื่อได้ลักษณะของพลังงานเสียง นำพลังงานเสียงที่ได้มาปรับบรรทัดฐานทางแอมปลิจูดด้วยค่าสูงสุดของพลังงาน เพื่อให้ในขั้นตอนของการประดิษฐ์ฟังก์ชันการถดถอยของพหุนาม

เป็นบรรทัดฐานเดียวกัน ซึ่งฟังก์ชันถดถอยพหุนามในงานวิจัยนี้ใช้ฟังก์ชันการถดถอยพหุนามอันดับสอง โดยจะได้สัมประสิทธิ์ฟังก์ชันการถดถอยพหุนาม เป็นค่า a_0, a_1, และ a_2 ตามที่ได้กล่าวมาในหัวข้อที่ 2.6.2.1 โดยลักษณะฟังก์ชันการถดถอยพหุนามของเสียงสระต่างๆ จะเป็นลักษณะดังรูปต่อไปนี้

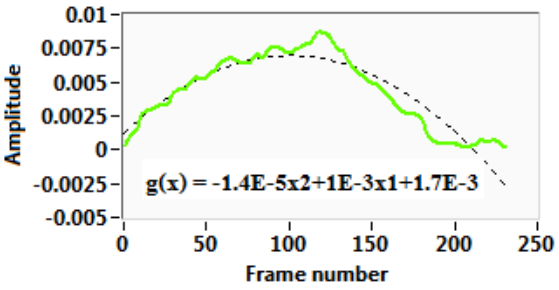


(ก) เสียง “อึ”

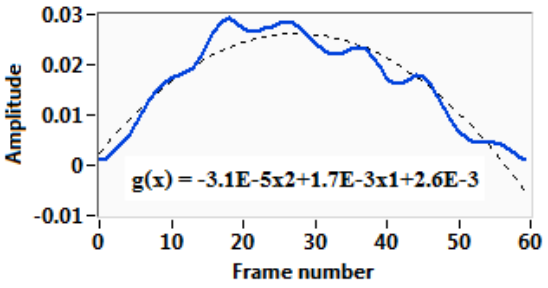
(ข) เสียง “อือ”



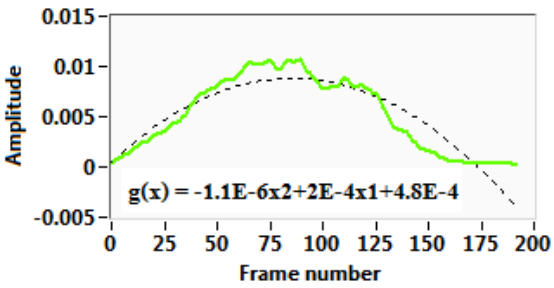
(ค) เสียง “เออะ”



(ง) เสียง “เออ”



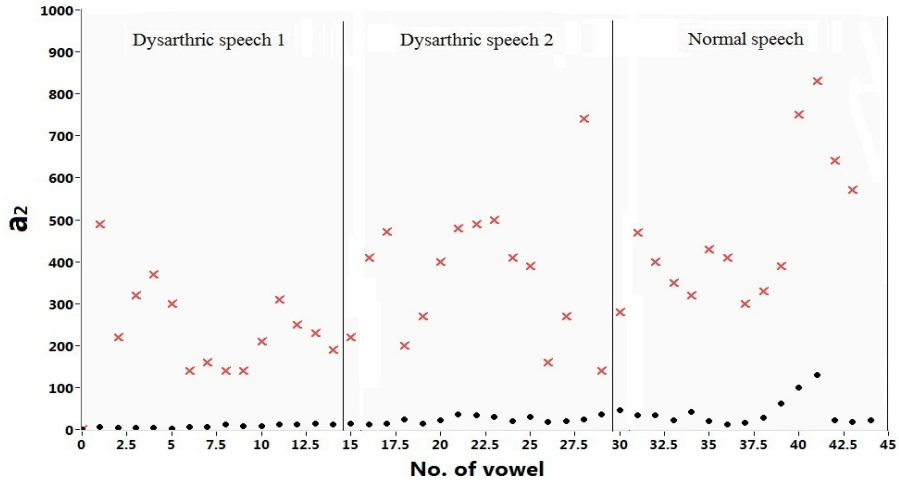
(จ) เสียง “อะ”



(ฉ) เสียง “อา”

ภาพประกอบ 3-19 ฟังก์ชันพหุนามของเสียงสระ

จากข้อมูลที่ได้ทั้งหมดเมื่อพิจารณาเฉพาะค่า a_2 จะเห็นว่ามีเปลี่ยนแปลงตามช่วงเวลาที่พูด (time) หรือจำนวนเฟรมของพลังงานเสียง (Frame) เมื่อนำค่าสัมประสิทธิ์ a_2 ทั้งหมดมาปรับสเกลด้วยการคูณกับ -10^7 จะทำให้เห็นความแตกต่างระหว่างสระเสียงสั้นและสระเสียงยาวมากขึ้น เมื่อนำค่าสัมประสิทธิ์การถดถอยฟังก์ชันพหุนามอันดับสอง (a_2) มาทำการแจกแจงจะได้ดังภาพประกอบที่ 3-20



ภาพประกอบ 3-20 การแจกแจงค่าสัมประสิทธิ์การถดถอยพหุนามอันดับสอง (a_2)

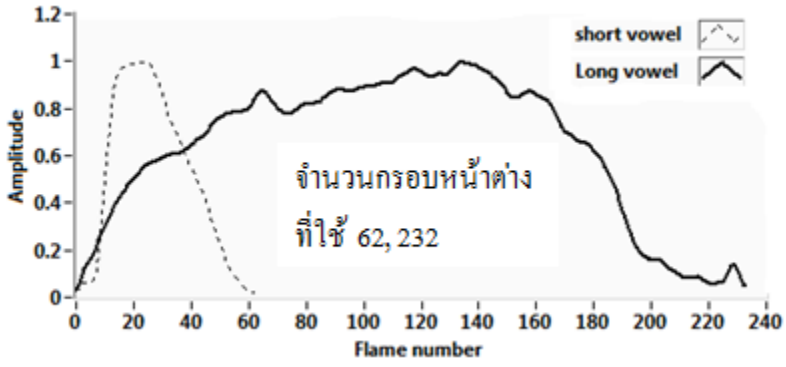
กำหนดให้ x = สระเสียงสั้น

● = สระเสียงยาว

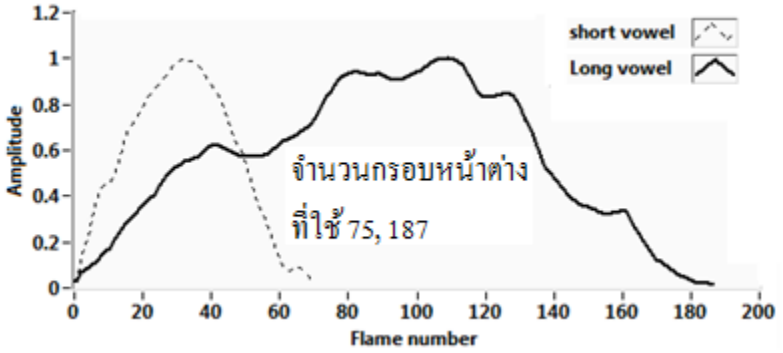
3.3.2 วิธีการวิเคราะห์จากช่วงระยะเวลาในการออกเสียง

การวิเคราะจากช่วงระยะเวลาในการออกเสียง เป็นการตรวจสอบจากจำนวนกรอบหน้าต่างที่ใช้บนฟังก์ชันพลังงานเสียงพูดแต่ละคำและนำข้อมูลที่ได้ทั้งหมดมาแจกแจง ส่งผลทำให้ทราบถึงความแตกต่างระหว่างสระเสียงสั้นและเสียงยาวได้ ซึ่งมีรายละเอียดดังนี้

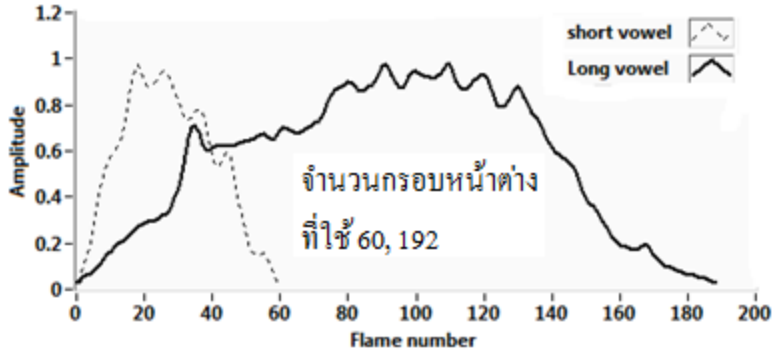
จากการทดลองก่อนหน้า (หัวข้อที่ 3.3.1) สามารถนำสัญญาณพลังงานเสียงพูดกำลังสองมาตรวจสอบจำนวนกรอบหน้าต่างที่ใช้ แสดงผลการทดลองดังภาพประกอบที่ 3-21 และผลสรุปดังตารางที่ 3-7 ถึง 3-9 ตามลำดับ



(ก) เสียง “อึ-อือ”



(ข) เสียง “เออะ-เออ”



(ค) เสียง “อะ-อา”

ภาพประกอบ 3-21 ตัวอย่างการตรวจสอบจำนวนกรอบหน้าต่างจากสัญญาณพลังงานเสียงกำลังสอง

ตารางที่ 3-7 ช่วงระยะเวลาในการออกเสียงของผู้ป่วยคนที่หนึ่ง (Dysarthric Speech 1)

ครั้งที่พูด	สระเสียงสั้น (จำนวนกรอบหน้าต่าง)	สระเสียงยาว (จำนวนกรอบหน้าต่าง)
1	51	218
2	62	232
3	63	283
4	56	244
5	59	285
6	80	231
7	78	225
8	75	187
9	79	205
10	70	204

11	60	192
12	66	186
13	69	176
14	78	189
15	71	179

ตารางที่ 3-8 ช่วงระยะเวลาในการออกเสียงของผู้ป่วยคนที่สอง (Dysarthric Speech 2)

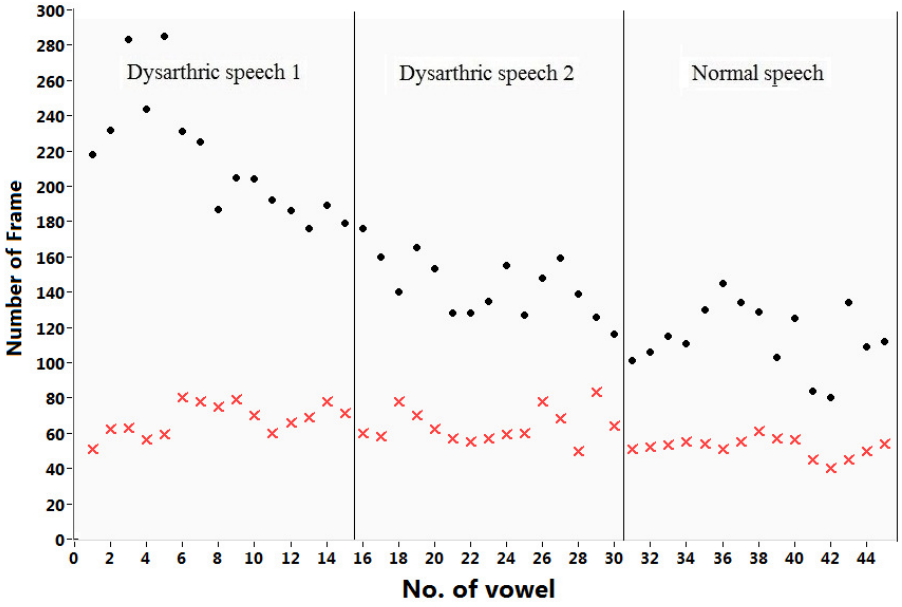
ครั้งที่พูด	สระเสียงสั้น (จำนวนกรอบหน้าต่าง)	สระเสียงยาว (จำนวนกรอบหน้าต่าง)
1	60	176
2	58	160
3	78	140
4	70	165
5	62	153
6	57	128
7	55	128
8	57	135
9	59	155
10	60	127
11	78	148
12	68	159
13	50	139
14	83	126
15	64	116

ตารางที่ 3-9 ช่วงระยะเวลาในการออกเสียงของผู้พูดปกติ (Normal Speech)

ครั้งที่พูด	สระเสียงสั้น (จำนวนกรอบหน้าต่าง)	สระเสียงยาว (จำนวนกรอบหน้าต่าง)
1	51	101
2	52	106
3	53	115
4	55	111
5	54	130

6	51	145
7	55	134
8	61	129
9	57	103
10	56	125
11	45	84
12	40	80
13	45	134
14	50	109
15	54	112

จากตารางที่ 3-7 ถึง 3-9 นำข้อมูลทั้งหมดมาเขียนลงบนกราฟแจกแจงแบบกระจาย ได้ดังภาพประกอบที่ 3-22



ภาพประกอบ 3-22 การแจกแจงค่าระยะเวลาในการออกเสียง

กำหนดให้ x = สระเสียงสั้น

● = สระเสียงยาว

บทที่ 4

การออกแบบระบบคำสั่ง

ในบทนี้จะกล่าวถึงการออกแบบระบบคำสั่ง ตามวัตถุประสงค์ของวิทยานิพนธ์ โดยวิธีการคัดแยกจากความถี่ฟอร์แมนต์และช่วงระยะเวลาในการออกเสียง ซึ่งได้แนวทางจากผลการทดลองในบทที่ 3 ที่แสดงให้เห็นถึงความแตกต่างของพารามิเตอร์แต่ละชนิด สำหรับบทนี้เริ่มต้นด้วยรายละเอียดของการวิเคราะห์และเปรียบเทียบการทดลองที่ได้ทำการศึกษามาทั้งหมด เพื่อนำไปสู่การตัดสินใจเลือกพารามิเตอร์ที่เหมาะสมในการออกแบบระบบ หลังจากนั้นจึงเป็นรายละเอียดของผลการรู้จำและทดสอบกับระบบฟัซซี่ลอจิกเพื่อชี้ให้เห็นถึงความแม่นยำของระบบเปรียบเทียบในแต่ละบุคคล

4.1 ผลการศึกษาการคัดแยกคำสั่งเสียงสระโดยวิเคราะห์จากความถี่ฟอร์แมนต์

จากผลการศึกษาจะเห็นว่าค่าความถี่ฟอร์แมนต์ทั้งสองที่ได้จากการทดลองในบทที่ 3 หัวข้อที่ 3.2.1 ถึง 3.2.2 นั้นมีความแตกต่างกันน้อยมาก จากตารางที่ 4-1 เป็นตารางเปรียบเทียบความแตกต่างระหว่างเทคนิคทั้งสอง สังเกตที่ผู้ป่วย Dysarthric 1 ความถี่ฟอร์แมนต์ที่ 1 (F1) เสียงสระ อึ-อือ มีค่า 0.3 นั่นหมายความว่า ความแตกต่างของผลลัพธ์ระหว่างเทคนิคทั้งสอง มีความแตกต่างกันร้อยละ 0.3 ในทำนองเดียวกันกับคำอื่นๆ จะเห็นว่าร้อยละความแตกต่างมีค่าน้อยมาก คือ อยู่ในระหว่างร้อยละ 0 ถึง 5 และมีค่าเฉลี่ยทั้งหมดเท่ากับร้อยละ 1.22 ความแตกต่างนั้นเกิดจากการเจือปนของสัญญาณไม่เท่ากัน กล่าวคือผลจากการเลือกช่วงความถี่ปฏิบัติงาน โดยกระบวนการ Wavelet Decomposition จะช่วยขจัดสัญญาณรบกวนและช่วงความถี่ที่ไม่เกี่ยวข้องออกไป จึงทำให้ผลลัพธ์ที่ได้มีความน่าเชื่อถือมากกว่าเทคนิค FFT เพียงอย่างเดียว

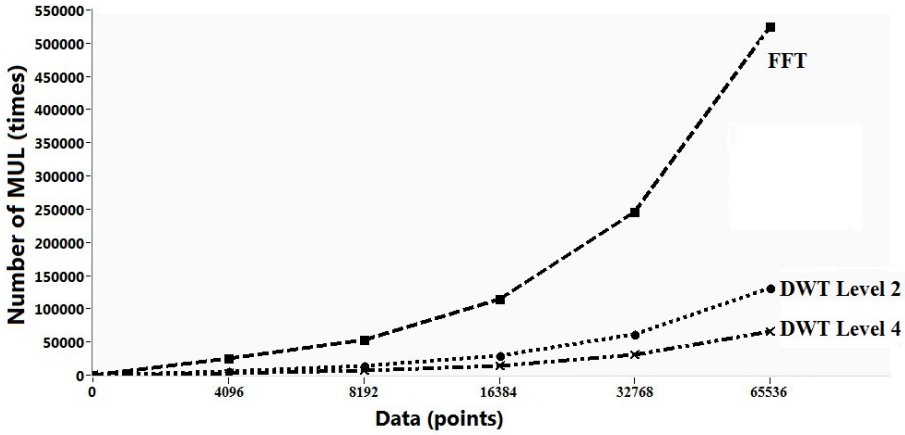
ตารางที่ 4-1 ความแตกต่างของความถี่ฟอร์แมนต์เปรียบเทียบระหว่างเทคนิคทั้งสอง มีหน่วยเป็นร้อยละ

ประเภท	Dysarthric 1		Dysarthric 2		Normal	
ความถี่ / สระ	F1 (%)	F2 (%)	F1 (%)	F2 (%)	F1 (%)	F2 (%)
อึ-อือ	0.3	3.7	0.8	0.5	1.1	0.1
เออะ-เออ	0.3	0.2	5.0	1.5	3.5	0.8
อะ-อา	0.2	0.1	1.3	0.5	0.5	1.5

ถึงแม้ว่าผลที่ได้จะมีความแตกต่างกันไม่มากนัก แต่ก็จะส่งผลให้เกิดความแตกต่างต่อขั้นตอนและกระบวนการคำนวณเช่นกันคือ ส่วนของเทคนิค FFT จะเป็นการคำนวณที่ง่ายไม่ซับซ้อน แต่จะมีจำนวนการคูณกันของข้อมูลที่เยอะกว่าการใช้เทคนิค DWT + FFT กล่าวคือ ในสัญญาณเสียงเดียวกัน การใช้เทคนิค FFT โปรแกรมจะต้องทำการคูณทั้งหมด $(N/2)\text{Log}_2N$ ครั้ง (N คือ จำนวน sample ของสัญญาณ) เปรียบเทียบกับการใช้ Wavelet Decomposition ในการกรองช่วงความถี่ก่อนเข้าสู่กระบวนการ FFT ซึ่งเมื่อสัญญาณผ่านกระบวนการกรองความถี่ในแต่ละระดับ จะทำให้จำนวนข้อมูลในการสุ่มตัวอย่างลดลงครึ่งหนึ่ง ($N/2$) ดังนั้นการคำนวณมีการคูณทั้งสิ้น $(N/2 \bullet 2j)\text{Log}_2N$ ครั้ง (j คือระดับของ Wavelet Decomposition) ซึ่งช่วยลดระยะเวลาในการคำนวณจากจำนวนครั้งของการคูณ จะมีผลกระทบอย่างมากเมื่อประยุกต์ใช้ในระดับไมโครคอลโทรลเลอร์ แสดงดังตารางที่ 4-2 และภาพประกอบที่ 4-1 เปรียบเทียบจำนวนครั้งของการคูณระหว่างเทคนิค FFT และ เทคนิค DWT (Level 2&4) +FFT

ตารางที่ 4-2 จำนวนครั้งของการคูณเปรียบเทียบระหว่างเทคนิค FFT และ DWT+FFT

จำนวนข้อมูล (จุด)	FFT (ครั้ง)	DWT(L.2)+FFT (ครั้ง)	DWT(L.4)+FFT (ครั้ง)
4096	24576	6144	3072
8192	53248	13312	6656
16384	114688	28672	14336
32768	245760	61440	30720
65536	524288	131072	65536



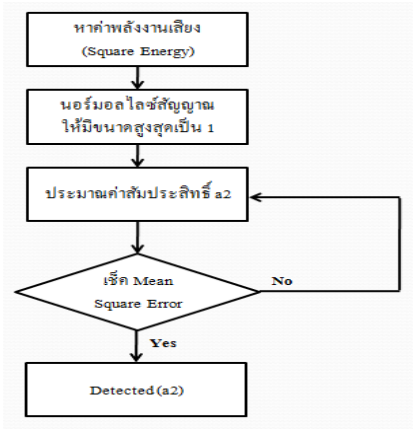
ภาพประกอบ 4-1 จำนวนครั้งของการคูณเปรียบเทียบระหว่างเทคนิคทั้งสอง ที่ความถี่สุ่มตัวอย่างเท่ากับ 20 kHz

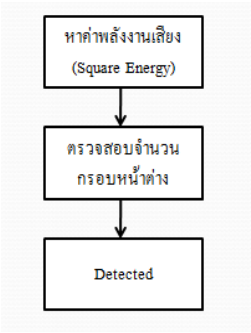
4.2 ผลการศึกษาการคัดแยกสระเสียงสั้น-เสียงยาวโดยวิเคราะห์จากค่าพลังงานเสียง

จากการศึกษาจะเห็นว่าการคัดแยกเสียงสระเสียงสั้นและเสียงยาว โดยเทคนิคการถดถอยของพหุนามนั้น สามารถแยกสระเสียงสั้น – ยาวได้ แต่มีข้อบกพร่องอยู่ คือ 1) มีขั้นตอนและกระบวนการที่ซับซ้อน 2) เกิดความผิดพลาด (error) ในขณะที่ทำการประมาณค่า ฟังก์ชันพลังงานเสียง (curve fitting) ส่งผลให้ค่า a_2 ที่ได้มีค่ากระจัดกระจายพอสมควรสังเกตจากภาพประกอบที่ 3-20 จะเห็นว่าสัมประสิทธิ์ a_2 สำหรับสระเสียงสั้นจะมีค่ากระจัดกระจายอยู่ในช่วง 100 ถึง 800 และสระเสียงยาวจะมีค่ากระจัดกระจายอยู่ในช่วง 1 ถึง 100 ส่วนวิธีการวิเคราะห์จากช่วงระยะเวลาในการออกเสียง จะมีขั้นตอนและการคำนวณที่ง่ายกว่า จากภาพประกอบที่ 3-22 ค่าที่ได้ค่อนข้างเกาะกลุ่มกันพอสมควร สำหรับสระเสียงสั้นอยู่ในช่วง หลักสิบ (ประมาณ 50 ถึง 90) และสระเสียงยาวจะมีค่าอยู่ในช่วงหลักร้อย (ประมาณ 100 ถึง 300) จึงทำให้ง่ายต่อการจัดกลุ่มของข้อมูลส่งผลให้ระบบสามารถตัดสินใจได้ง่ายขึ้นและเมื่อเปรียบเทียบจำนวนขั้นตอนระหว่างระบบทั้งสองแล้ว จะเห็นว่าวิธีที่สองนั้นมีขั้นตอนการทำงานที่น้อยกว่า จากตารางที่ 4-3 แสดงผลการทดสอบการแยกสระเสียงสั้น-สระเสียงยาวและภาพประกอบที่ 4-2 แสดงให้เห็นถึงจำนวนขั้นตอนในการทำงานเปรียบเทียบกันทั้งสองระบบ

ตารางที่ 4-3 ผลการทดสอบการแยกสระเสียงสั้นและเสียงยาวโดยเทคนิคทั้งสอง

เทคนิคที่ใช้	จำนวนคำที่ใช้ทดสอบ	คำที่แยกไม่ถูกต้อง	ความแม่นยำ
การถดถอยของพหุนาม	สระเสียงสั้น, ยาว อย่างละ 45 คำ	4 คำ	95.56 %
การวิเคราะห์จากช่วงระยะเวลาในการออกเสียง		1 คำ	98.89 %





(ก) (ข)

ภาพประกอบ 4-2 ขั้นตอนการทำงานของระบบคัดแยกสระเสียงสั้นและเสียงยาว

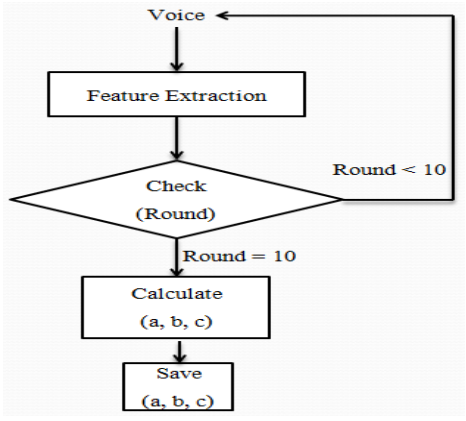
(ก) เทคนิคการถดถอยของพหุนาม

(ข) เทคนิคการวิเคราะห์จากช่วงระยะเวลาในการออกเสียง

4.3 การรู้จำโดยใช้ระบบฟัซซี่

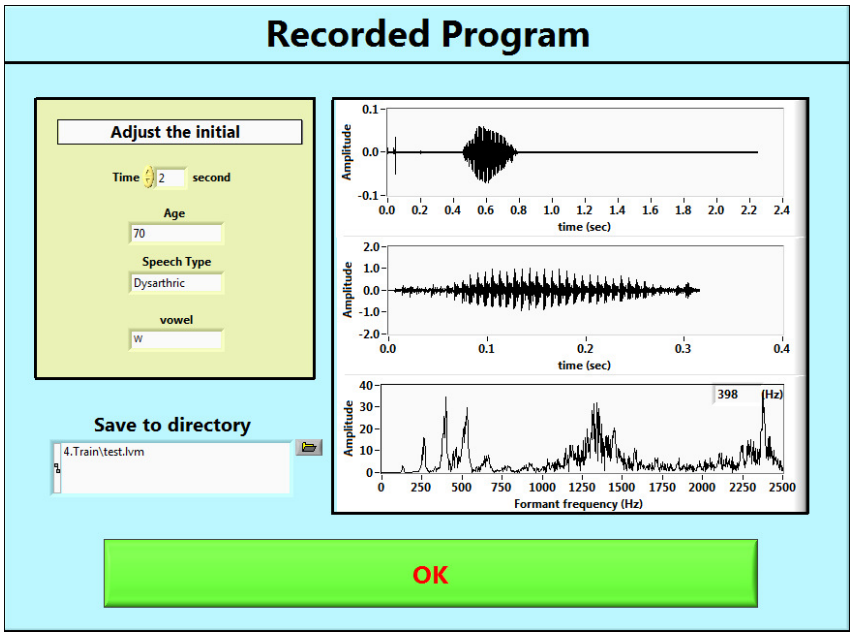
หลังจากที่สามารถสกัดคุณลักษณะของเสียงทั้งหมดได้แล้ว ขั้นตอนถัดไปก็เป็นการออกแบบให้ระบบสามารถรู้จำและตัดสินใจ เนื่องจากพารามิเตอร์ที่สกัดได้จากการทดลองนั้นเป็นข้อมูลตัวเลขที่สามารถแยกกันอย่างชัดเจน คือ มีความเหลื่อมล้ำของข้อมูลน้อย (สังเกตจากตารางความถี่ฟอร์แมนต์ทั้งหมด) งานวิจัยนี้จึงเลือกใช้ระบบฟัซซี่ลอจิกเพียงอย่างเดียวก็สามารถเรียนรู้และตัด สินใจระบบคำสั่งได้เป็นอย่างดี สำหรับหัวข้อนี้จะแบ่งออกเป็นสองส่วน คือ 1) ส่วนของการสอนระบบ (Train) และ 2) ส่วนของการทดสอบระบบ (Test) ซึ่งมีรายละเอียดดังนี้

4.3.1 การสอนระบบ



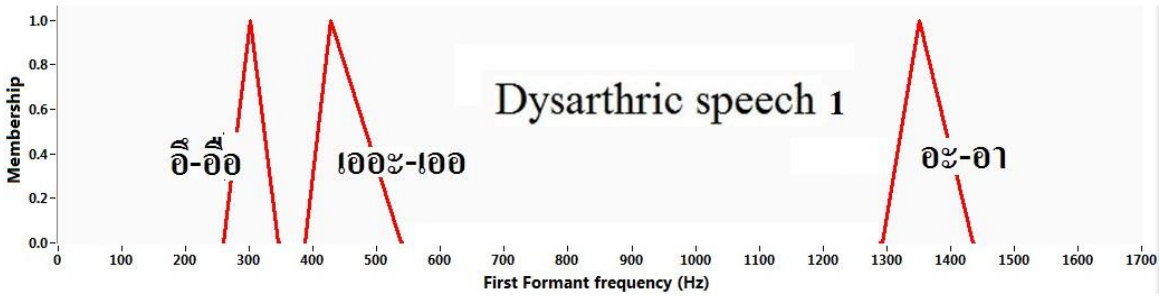
ภาพประกอบ 4-3 ขั้นตอนการสอนระบบ

เริ่มจากผู้ทดสอบเปล่งเสียง (Voice) ระบบจะรับสัญญาณและส่งต่อไปยังขั้นตอน Feature Extraction ซึ่งได้มาจากขั้นตอนการสกัดคุณลักษณะของเสียง ถัดไปเป็นขั้นตอน Check Round จะทำหน้าที่ตรวจสอบจำนวนรอบของการพูดและเก็บค่าความถี่ทั้งสอง เพื่อความเหมาะสมกับผู้ป่วย งานวิจัยนี้จึงใช้การสอนระบบไม่มากจนเกินไป คือ ทั้งหมด 10 รอบ ถ้าจำนวนรอบยังไม่ครบ โปรแกรมก็จะให้ผู้พูดทำการพูดใหม่ หลังจากจำนวนรอบครบแล้วก็จะเข้าสู่ขั้นตอน Calculate เพื่อทำการคำณวนค่า a, b, c สำหรับนำไปสร้างเป็นฟังก์ชันความเป็นสมาชิกแบบสามเหลี่ยม (Triangular Membership Function) ในระบบฟัซซี่ (รายละเอียดหัวข้อที่ 2.6.3 ก) และสุดท้ายค่า a, b, c ก็จะถูกนำมาบันทึก (save) ลงในระบบ เพื่อใช้สำหรับการทดสอบต่อไป

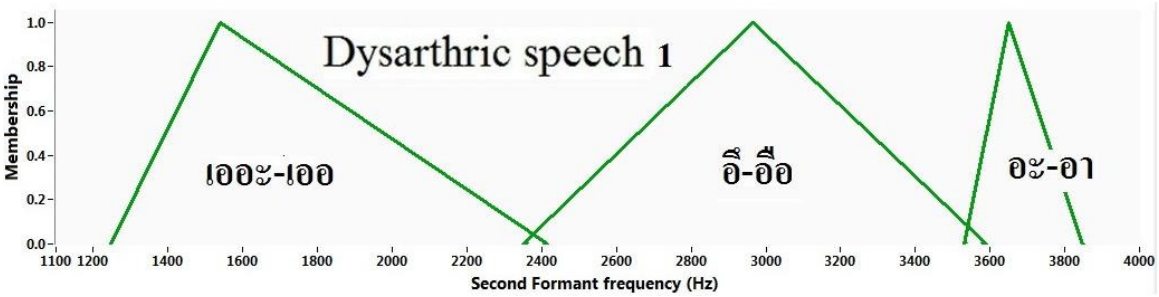


ภาพประกอบ 4-4 โปรแกรมสอนระบบ

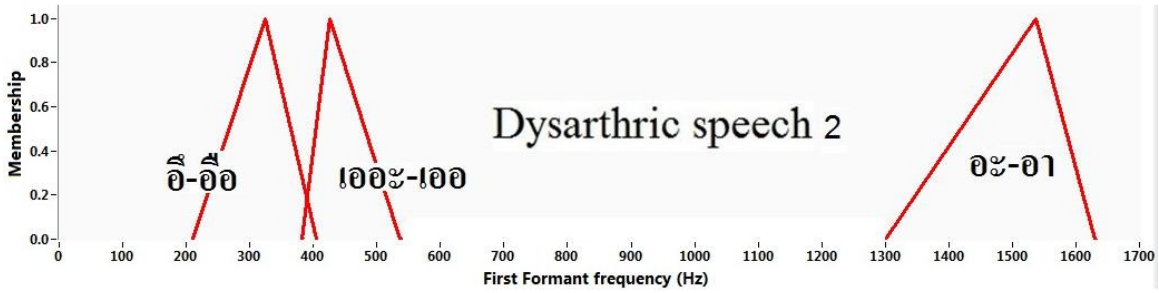
หลังจากทำการสอนระบบเรียบร้อยแล้ว ก็จะ ได้กราฟฟังก์ชันความเป็นสมาชิกแบบสามเหลี่ยม จำแนกประเภทตามรายบุคคล รายละเอียดดังภาพประกอบที่ 4-5 ถึง 4-10 (กำหนดให้ : ω = อึ, ωω = อือ, γ = เออะ, γγ = เออ, a = อะ, aa = อา)



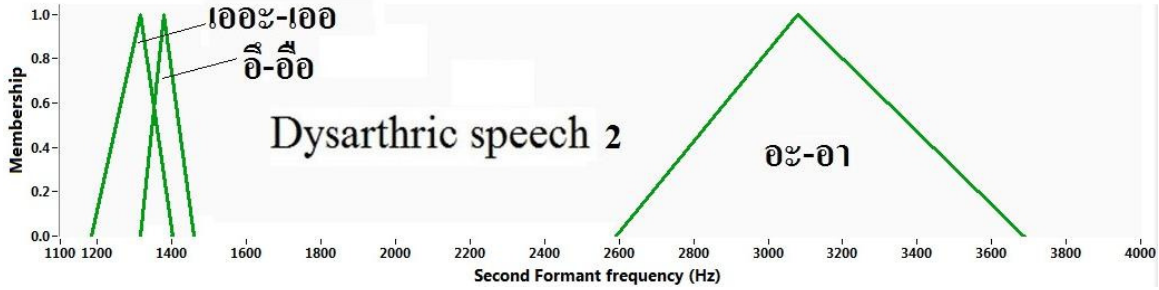
ภาพประกอบ 4-5 ฟังก์ชันความเป็นสมาชิกของความถี่ฟอร์แมนต์ที่ 1 (Dysarthric speech 1)



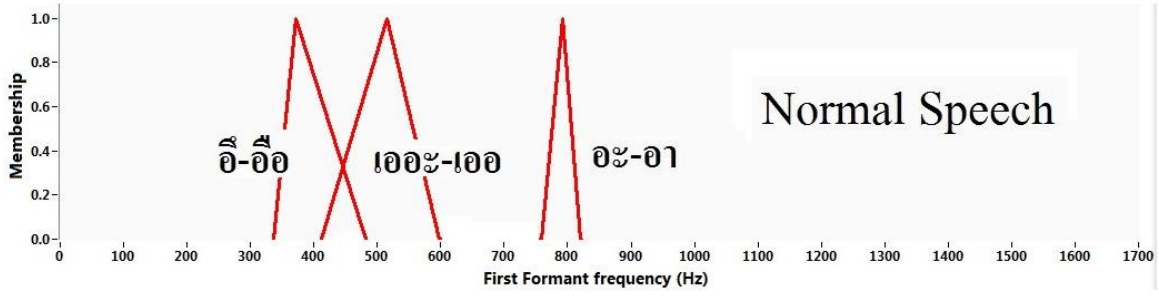
ภาพประกอบ 4-6 ฟังก์ชันความเป็นสมาชิกของความถี่ฟอร์แมนต์ที่ 2 (Dysarthric speech 1)



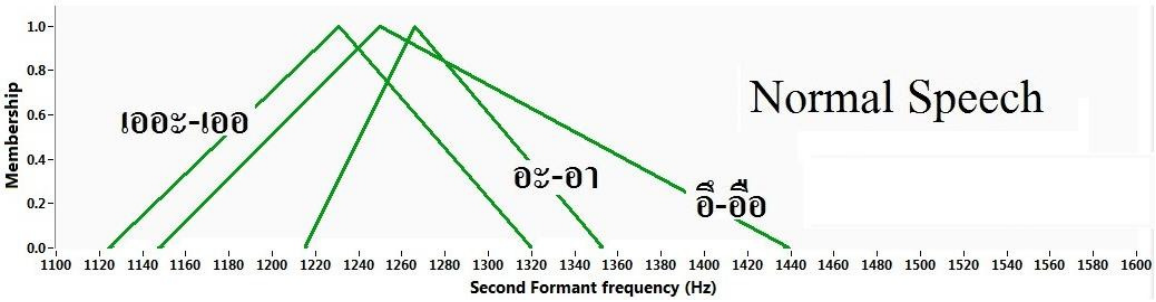
ภาพประกอบ 4-7 ฟังก์ชันความเป็นสมาชิกของความถี่ฟอร์แมนต์ที่ 1 (Dysarthric speech 2)



ภาพประกอบ 4-8 ฟังก์ชันความเป็นสมาชิกของความถี่ฟอร์แมนต์ที่ 2 (Dysarthric speech 2)

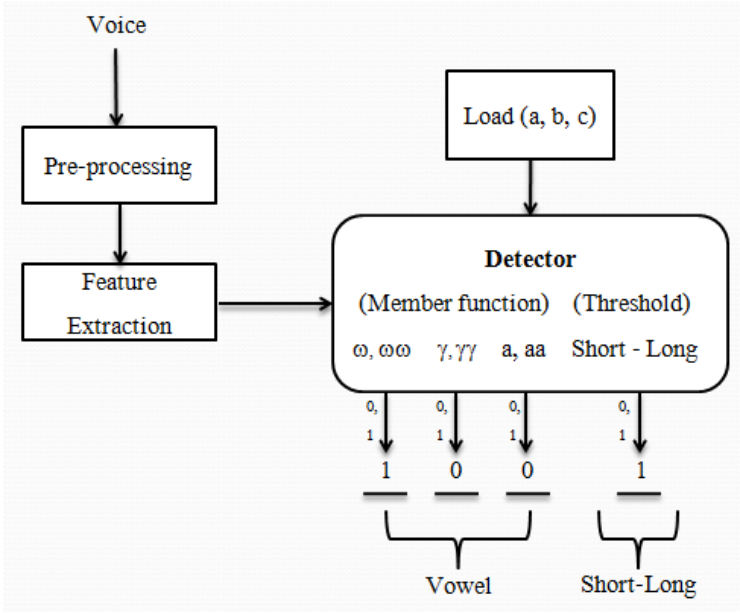


ภาพประกอบ 4-9 ฟังก์ชันความเป็นสมาชิกของความถี่ฟอร์แมนต์ที่ 1 (Normal speech)



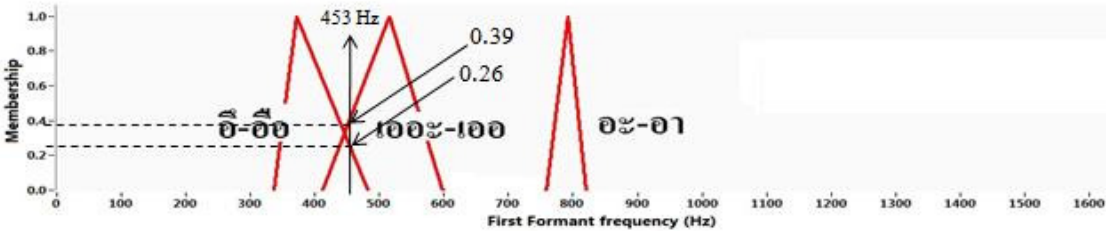
ภาพประกอบ 4-10 ฟังก์ชันความเป็นสมาชิกของความถี่ฟอร์แมนต์ที่ 2 (Normal speech)

4.3.2 การทดสอบระบบ



ภาพประกอบ 4-11 ขั้นตอนการทดสอบระบบ

เริ่มจากขั้นตอนแรก โปรแกรมจะทำการเรียกแฟ้มข้อมูล (ค่า a, b, c) ที่ได้จากการสอนระบบ หลังจากนั้นโปรแกรมก็จะรอสัญญาณเสียงพูด (Voice) เพื่อที่จะทำการสกัดค่าความถี่ฟอร์แมนต์และช่วงระยะเวลาในการออกเสียง ถัดไปสัญญาณก็ถูกส่งต่อไปยังตัวตรวจจับคุณลักษณะเสียง (Detector) โดยการแยกเสียงสระนั้นจะตรวจเช็คจากความเป็นสมาชิก (Member function) ของกระบวนการตัดสินใจในฟัซซี่ลอจิก ซึ่งมีรายละเอียดและขั้นตอนในการคำนวณดังนี้



ภาพประกอบที่ 4-12 วิธีการตัดสินใจของระบบเมื่อผู้ทดสอบเปล่งเสียงสระเออะที่มีความถี่ฟอร์แมนต์ที่หนึ่งเท่ากับ 453 Hz

กำหนดให้

- ฟังก์ชันความเป็นสมาชิกแบบสามเหลี่ยมของกลุ่มสระอึ-ฮึอเป็นสามเหลี่ยมที่หนึ่ง หลังจากการสอนระบบมีค่า a, b, c เท่ากับ 336, 371.7, 482 ตามลำดับ
- ฟังก์ชันความเป็นสมาชิกแบบสามเหลี่ยมของกลุ่มสระเออะ-เออ เป็นสามเหลี่ยมที่สอง หลังจากการสอนระบบมีค่า a, b, c เท่ากับ 412, 516.9, 598 ตามลำดับ
- ฟังก์ชันความเป็นสมาชิกแบบสามเหลี่ยมของกลุ่มสระอะ-อา เป็นสามเหลี่ยมที่สาม หลังจากการสอนระบบมีค่า a, b, c เท่ากับ 758, 792.6, 821 ตามลำดับ
- ผู้พูดเปล่งเสียงสระเออะที่มีความถี่ฟอร์แมนต์ที่หนึ่งเท่ากับ 453 Hz

จากสมการในหัวข้อที่ 2.6.3 (ก)

$$Triangular(x, a, b, c) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a < x \leq b \quad (A) \\ \frac{c-x}{c-b} & b < x \leq c \quad (B) \\ 0 & x > c \end{cases}$$

พิจารณาสามเหลี่ยมที่หนึ่ง (แทนค่า x ด้วย 453 ในสมการ (A) และ (B))

(A) $\frac{453-336}{371.7-336} = \frac{117}{35.7} = 3.28$ (ค่าเกิน 1.0 โปรแกรมจะตัดให้เป็น 0)

(B) $\frac{482-453}{482-371.7} = \frac{29}{110.3} = 0.26$ (ค่าไม่เกิน 1.0 โปรแกรมจะเก็บค่าไว้)

พิจารณาสามเหลี่ยมที่สอง (แทนค่า x ด้วย 453 ในสมการ (A) และ (B))

(A) $\frac{453-412}{516.9-412} = \frac{41}{104.9} = 0.39$ (ค่าไม่เกิน 1.0 โปรแกรมจะเก็บค่าไว้)

(B) $\frac{598-453}{598-516.9} = \frac{145}{81.1} = 1.79$ (ค่าเกิน 1.0 โปรแกรมจะตัดให้เป็น 0)

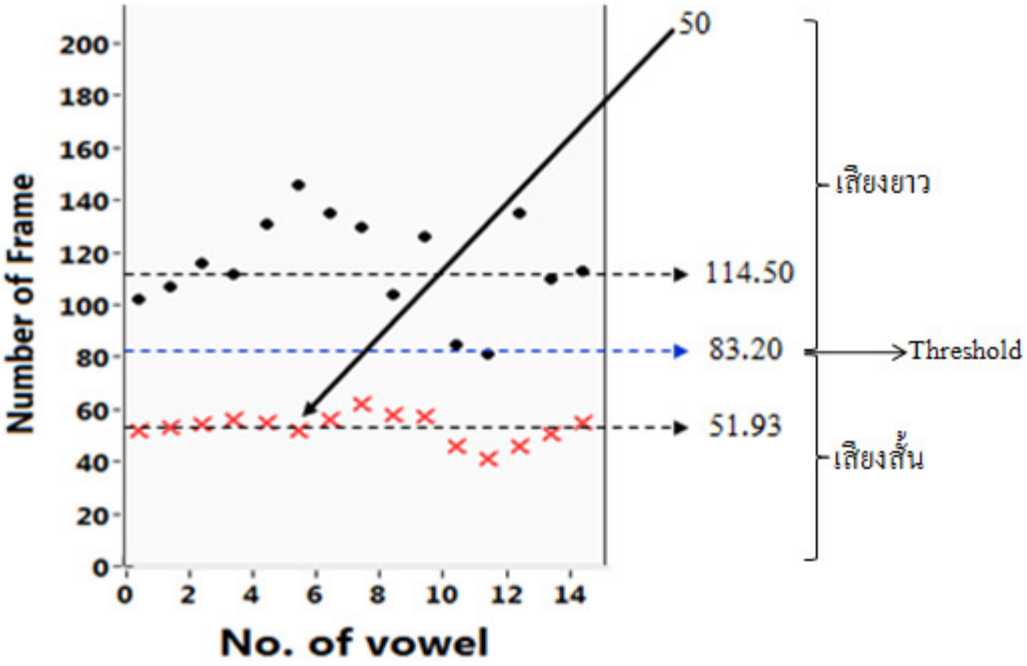
พิจารณาสามเหลี่ยมที่สาม (แทนค่า x ด้วย 453 ในสมการ (A) และ (B))

(A) $\frac{453-758}{792.6-453} = \frac{-305}{339.6} =$ -0.9 (ค่าติดลบ โปรแกรมจะตัดให้เป็น 0)

(B) $\frac{821-453}{821-792.6} = \frac{368}{28.4} = 12.95$ (ค่าเกิน 1.0 โปรแกรมจะตัดให้เป็น 0)

ถัดไปโปรแกรมก็จะนำค่าที่เก็บไว้มาเปรียบเทียบกัน คือ 0.26 < 0.39 ดังนั้นค่า 0.29 โปรแกรมก็จะตัดให้เป็น 0 ทันที และค่า 0.39 โปรแกรมก็จะตัดให้เป็น 1 หลังจากนั้นก็จะได้ผลสรุปออกมา คือ สามเหลี่ยมที่หนึ่ง คือ 0, สามเหลี่ยมที่สอง คือ 1, สามเหลี่ยมที่สาม คือ 0 ดังนั้นจะได้ออกมาเป็นรหัสคือ 0 1 0 จากนั้นการตรวจจับสระเสียงสั้นเสียงยาวจะใช้กระบวนการวิเคราะห์จากช่วงระยะเวลาในการออกเสียง โดยมีรายละเอียดดังนี้

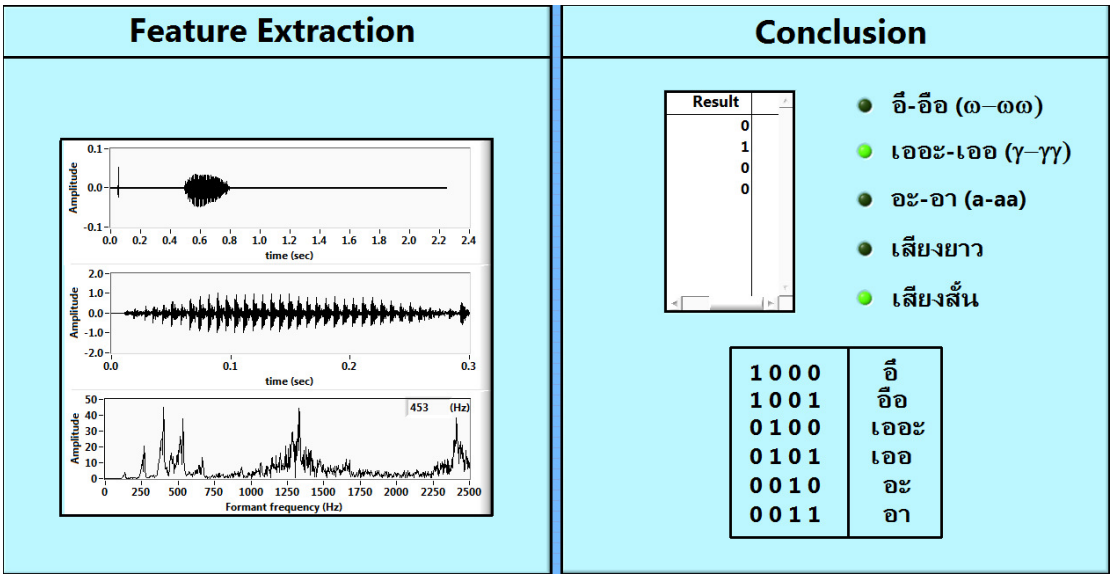
กำหนดให้เสียงสระเออะที่กล่าวมาข้างต้น จากสัญญาณพลังงานเสียงนี้ใช้จำนวนกรอบหน้าต่างทั้งหมด 50 กรอบ จากตารางที่ 3-9 นำค่าจำนวนกรอบหน้าต่างของสระเสียงสั้นและเสียงยาวมาหาค่าเฉลี่ยจะได้เท่ากับ 51.93 กรอบและ 114.50 กรอบ การคำนวณค่า Threshold จะคิดจากค่ากึ่งกลางระหว่างค่าเฉลี่ยทั้งสอง จะได้เท่ากับ $\frac{114.50-51.93}{2} + 51.93 = 83.20$ กรอบ ดังนั้นเสียงที่เปล่งออกมาซึ่งมีจำนวนกรอบหน้าต่างทั้งหมด 50 กรอบ ตรวจสอบแล้วมีค่าน้อยกว่า 83.20 กรอบ โปรแกรมจึงตัดสินให้เป็นสระเสียงสั้น คือจะเป็นรหัสเลข 0 ตามเงื่อนไขคือ; จำนวนกรอบหน้าต่าง > 81.20 เป็นสระเสียงยาว คือเลข 1 และถ้าจำนวนกรอบหน้าต่าง < 81.20 เป็นสระเสียงสั้นคือเลข 0 ตามลำดับ จึงทำให้ผลสรุปสุดท้ายคือจะได้รหัสออกมาเป็น 0101 ซึ่งตรงกับเสียงสระเออะแสดงดังภาพ ประกอบที่ 4-13 คือวิธีการตัดสินสระเสียงสั้นเสียงยาวและตารางที่ 4-4 คือผลสรุปของระบบ



ภาพประกอบที่ 4-13 วิธีการตัดสินใจของระบบเมื่อผู้ทดสอบเปล่งเสียงสระเออะที่มีช่วงระยะเวลาในการออกเสียงเท่ากับ 50 กรอบหน้าต่าง

ตารางที่ 4-4 ผลสรุปของเสียงพูดจำแนกโดยรหัส

สระ	รหัส
อะ	0010
อา	0011
เออะ	0100
เออ	0101
อึ	1000
อือ	1001



ภาพประกอบ 4-14 โปรแกรมรู้จำเสียงพูด (Labview 8.6)

ตารางที่ 4-5 ผลการทดสอบโปรแกรมจำแนกตามรายบุคคล

ประเภท / สระ	Dysarthric Speech1	Dysarthric Speech 2	Normal Speech
	ความแม่นยำ (%)	ความแม่นยำ (%)	ความแม่นยำ (%)
อึ-อือ	100.0	76.9	75.0
เออะ-เออ	91.6	90.9	84.2
อะ-อา	100.0	93.3	100.0
เฉลี่ย	**97.2**	**87.0**	**86.4**

บทที่ 5

สรุปผลการวิจัยและข้อเสนอแนะ

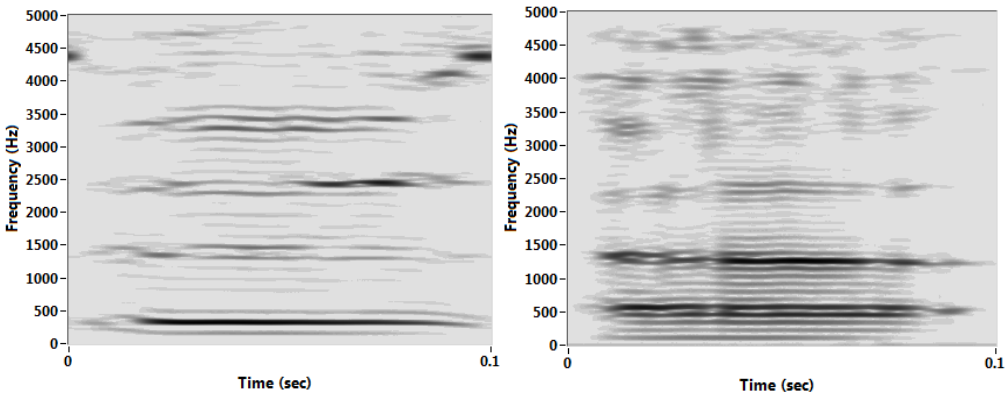
หลังจากได้ทำการศึกษาและออกแบบระบบสั่งการ ตามวัตถุประสงค์ของโครงการวิจัยดังรายละเอียดในบทที่ 3 และ 4 มาแล้ว ในบทนี้จะกล่าวถึงบทสรุปของงานวิจัย ปัญหาและอุปสรรคในการทำงานวิจัย รวมทั้งวิจารณ์และข้อเสนอแนะสำหรับการออกแบบระบบสั่งการโดยใช้เสียงพูดสำหรับผู้ป่วยชนิดนี้เพื่อให้มีประสิทธิภาพและเป็นประโยชน์ต่อผู้ที่ต้องการศึกษานำไปสู่แนวทางในการพัฒนางานวิจัยทางด้านรู้จำเสียงพูด หรือด้านอื่นๆที่เกี่ยวข้อง

5.1 สรุปผลการวิจัย

ผลจากการศึกษาวิจัยแสดงให้เห็นว่าในการแยกระบบคำสั่งโดยใช้เทคนิคฟาสฟูเรียร์ทรานฟอร์ม ซึ่งเป็นเทคนิคพื้นฐานของงานวิจัยทางด้านการรู้จำเสียงพูด เมื่อนำมาใช้กับผู้ป่วยชนิด Dysarthria ที่มีความสามารถในการพูดที่จำกัดและด้อยประสิทธิภาพแล้ว จากการวิเคราะห์จากกราฟในโดเมนเวลาเพียงอย่างเดียวทำให้เราทราบถึงความคลาดเคลื่อนของความถี่ฟอร์แมนต์ทั้งสองเมื่อเทียบกับผู้พูดปกติ จะมากหรือน้อยนั้นขึ้นอยู่กับแต่ละคำ กล่าวคือถ้าพิจารณาเฉพาะเสียงสระ อึ อือ เออะ เออ เพียง 4 คำ เมื่อสังเกตคร่าวๆ เหมือนกับว่าความถี่ฟอร์แมนต์ทั้งสองไม่ได้ผิดเพี้ยนไปมากเมื่อเทียบกับผู้พูดปกติ ส่วนเสียงสระอะและอานั้น สามารถสังเกตเห็นถึงความผิดเพี้ยนได้ชัดเจนมากกว่าคำก่อนหน้านี้ เมื่อนำกลุ่มคำเหล่านี้มาวิเคราะห์โดยใช้สเปกโตรแกรมจะสังเกตเห็นได้ดีขึ้นถึงความไม่สม่ำเสมอและการขาดช่วงของความถี่ฟอร์แมนต์ทั้งสองตั้งแต่เริ่มต้นออกเสียงจนกระทั่งสิ้นสุดการออกเสียงเมื่อเทียบกับผู้พูดปกติแสดงดังภาพประกอบที่ 5-1 ทำให้ทราบว่าความถี่ฟอร์แมนต์ที่สกัดได้จากผู้ป่วยถึงแม้จะไม่สม่ำเสมอเท่ากับคนปกติ แต่เทคนิค FFT ก็ยังสามารถแยกความถี่ฟอร์แมนต์ทั้งสองในโดเมนเวลาได้เหมือนกับคนปกติ

อีกทางหนึ่ง ข้อเสียของเทคนิค FFT ยังมีอยู่เช่นกัน กล่าวคือในงานวิจัยทางด้านการรู้จำเสียงพูดเมื่อเราต้องการเพิ่มความละเอียดของสัญญาณเพื่อให้ความแม่นยำของระบบนั้นสูงขึ้น จึงมีความจำเป็นที่จะต้องใช้ค่าความถี่สุ่มตัวอย่างที่สูงตามไปด้วย เมื่อเราทำการวิเคราะห์สัญญาณในโดเมนความถี่สำหรับผู้ป่วยชนิดนี้ เราไม่สามารถที่จะตัดสัญญาณบางช่วงหรือเพียงบางส่วนนำมาวิเคราะห์ได้เมือนกับคนปกติ ผลสืบเนื่องมาจากความไม่สม่ำเสมอและการขาดช่วงของความถี่ฟอร์แมนต์ทั้งสอง ดังนั้นในการวิเคราะห์สัญญาณในโดเมนความถี่จึงมีความจำเป็นที่จะต้องทำการวิเคราะห์ตั้งแต่เริ่มต้นจนสิ้นสุดสัญญาณเสียง ส่งผลทำให้จำนวนข้อมูลในการ

วิเคราะห์มากจนเกินไปซึ่งกล่าวมาแล้วในบทที่ 4 เมื่อเรานำมาประยุกต์ใช้งานกับไมโครคอน โทรลเลอร์แล้วจะทำให้เกิดความล่าช้าในการคำนวณเป็นอย่างมาก ดังนั้นผลจากการศึกษาในบทที่ 4 จึงส่งผลดีในการแก้ปัญหาและเพิ่มประสิทธิภาพของระบบหลังจากการใช้เทคนิค Wavelet Decom position มากยิ่งขึ้น



(ก) (ข)

ภาพประกอบ 5-1 เปรียบเทียบความต่อเนื่องของความถี่ฟอร์แมนต์

ก) ผู้ทดสอบคนที่หนึ่ง (Dysarthric Speech 1)

ข) ผู้ทดสอบคนที่สาม (Normal Speech)

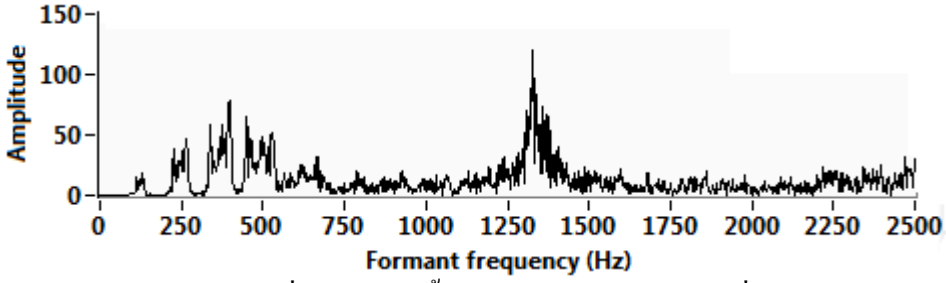
ผลจากการศึกษาในส่วนของการรู้จำในบทที่ 4 ทำให้ทราบถึงประสิทธิภาพของระบบกล่าวคือเมื่อเราสังเกตจากภาพประกอบที่ 4-5 ถึง 4-11 เสียงสระทั้งสามของกลุ่มผู้ทดสอบจะมีช่วงระยะห่างมาก, น้อยและน้อยที่สุดตามลำดับ กล่าวคือ สำหรับผู้พูดคนที่หนึ่ง (Dysarthric 1) ซึ่งเป็นอัมพาตกล้ามเนื้อระดับ 6 กลุ่มเสียงสระอึ-อือ, เออะ-เออ และ อะ-อา จะแยกกันอย่างชัดเจน สำหรับผู้พูดคนที่สอง (Dysarthric 2) ซึ่งเป็นอัมพาตกล้ามเนื้อระดับ 4 จะมีการเหลื่อมของกลุ่มเสียงสระอึ-อือ กับกลุ่มเสียงสระเออะ-เออ เพิ่มมากขึ้น และสุดท้ายสำหรับผู้พูดคนที่สาม (Normal Speech) จะมีการเหลื่อมของกลุ่มเสียงสระอึ-อือ กับกลุ่มเสียงสระเออะ-เออ มากที่สุด ซึ่งทำให้เราทราบว่า ความรุนแรงของระดับอัมพาตส่งผลต่อความเหลื่อมล้ำของกลุ่มเสียงสระอึ-อือ กับกลุ่มเสียงสระเออะ-เออ และความผิดเพี้ยนอย่างมากของกลุ่มเสียงสระอะ-อา ตามลำดับและผลกระทบจากการเหลื่อมของความถี่นี้จะส่งผลต่อความแม่นยำในการตัดสินใจของระบบ ดังนั้นระบบสั่งการที่สร้างขึ้นในงานวิจัยนี้จึงเหมาะสมกับผู้ป่วยที่มีระดับอัมพาตและความผิดปกติทางการพูดสูง สังเกตความแม่นยำของระบบดังตารางที่ 4-5 ผู้ป่วยระดับ 6 จะสูงกว่าผู้ป่วยระดับ 4 และต่ำที่สุดก็เป็นผู้พูดปกติตามลำดับ

5.2 ปัญหาและข้อเสนอแนะ

5.2.1 ปัญหา

การหาค่าความถี่ฟอร์แมนต์ที่หนึ่งและสองโดยใช้เกณฑ์ความถี่ที่เด่น (เป็นจุดยอดและมีขนาดสูงสุดในบริเวณความถี่ใกล้เคียงกัน) ของเสียงสระทั้ง 6 คำ ในทางปฏิบัติด้วยการเขียนโปรแกรมวิเคราะห์จากสเปกตรัมโดยตรงนั้นทำได้ยาก ซึ่งจะเห็นได้ชัดเจนจากตัวอย่างในภาพประกอบที่ 3-3 ถึง 3-5 ซึ่งปัญหาที่จะพบโดยทั่วไปคือปัญหาในการหาค่าความถี่ฟอร์แมนต์ที่หนึ่ง ซึ่งพบว่าสัญญาณเสียงพูดที่มีความถี่ฟอร์แมนต์ที่สองค่อนข้างสูงบางคำ เช่นสัญญาณเสียงสระเออ ดังภาพประกอบที่ 5-2 มักมีขนาดของสเปกตรัมของฟอร์แมนต์ที่หนึ่งเด่นน้อยกว่าสเปกตรัมที่มีความถี่บริเวณใกล้เคียงกับฟอร์แมนต์ที่สอง จึงทำให้การพิจารณาความถี่ฟอร์แมนต์ที่หนึ่งมีโอกาสผิดพลาดได้เช่นกัน และจากการสังเกตสเปกตรัมของสัญญาณเสียงพูดที่มีความถี่ฟอร์แมนต์ที่หนึ่งต่ำ พบว่าฟอร์แมนต์ที่หนึ่งมักจะมีขนาดสูงกว่า 0.5 เท่าของฟอร์แมนต์ที่สอง ดังนั้นเงื่อนไขที่พอจะใช้เป็นเกณฑ์ในการพิจารณาหาค่าความถี่ฟอร์แมนต์ที่ถูกต้องที่เป็นไปได้สำหรับปัญหานี้คือ

1. พิจารณาความถี่ฟอร์แมนต์ที่หนึ่งจากความถี่ที่มีสเปกตรัมเป็นจุดยอดและให้ค่าขนาดของสเปกตรัมสูงสุดในช่วงความถี่ 0 – 1000 Hz สำหรับเสียงสระทุกคำในผู้พูดปกติรวมถึงเสียงสระอึ-อือ, เออะ-เออ ในผู้ป่วยทั้งสอง และในช่วงความถี่ 1000-2000 Hz สำหรับเสียงสระอะ-อาในผู้ป่วยทั้งสองตามลำดับ
2. พิจารณาความถี่ฟอร์แมนต์ที่สองจากความถี่ที่มีสเปกตรัมเป็นจุดยอดและให้ค่าขนาดของสเปกตรัมสูงสุดในช่วงความถี่ 1000-2000 Hz สำหรับเสียงสระทุกคำในผู้พูดปกติ และในช่วงความถี่ 1000-3000 Hzสำหรับเสียงสระอึ-อือ, เออะ-เออ ในผู้ป่วยทั้งสอง และในช่วงความถี่ 2000 Hz ขึ้นไปสำหรับเสียงสระอะ-อาในผู้ป่วยทั้งสองตามลำดับ
3. ความถี่ฟอร์แมนต์ที่หนึ่งและสองจะต้องห่างกันมากกว่า 250 Hz



ภาพประกอบ 5-2 ตัวอย่างความถี่ฟอร์แมนต์ทั้งสองของเสียงสระ “เออ” ที่ก่อให้เกิดความผิดพลาด

สุดท้ายคือปัญหาสืบเนื่องมาจากตัวผู้ป่วย ซึ่งเราไม่สามารถคาดหวังว่าผู้ป่วยจะออกเสียงสระได้ครบทุกคำและออกเสียงได้ดีเหมือนกันทุกครั้งสาเหตุอันเนื่องมาจากร่างกายของผู้ป่วยไม่ค่อยมีความแข็งแรงเท่าที่ควร อาจจะมีอาการทรุดหรือดีขึ้นภายใน 2 ถึง 3 อาทิตย์ ไม่สามารถทราบล่วงหน้าได้ ซึ่งมีผลกระทบโดยตรงต่อการพูดด้วย ดังนั้นในการเก็บสัญญาณเสียงกับผู้ป่วยจึงต้องทำการเก็บให้เสร็จ ณ เวลานั้น เพื่อลดความผิดพลาดอันเนื่องมาจากการออกเสียงของตัวผู้ป่วยทั้งหมด

5.2.2 ข้อเสนอแนะ

จากการทำวิจัย ผู้วิจัยพบว่า ข้อเสนอแนะต่อไปนี้จะเป็นแนวทางที่จะช่วยให้การรู้จำเสียงพูดมีความถูกต้องยิ่งขึ้น

1. การออกเสียงที่ถูกต้องและชัดเจน มีความสำคัญอย่างมากต่อการกำหนดจุดเริ่มต้นและจุดสิ้นสุดของสัญญาณที่นำมาวิเคราะห์อย่างถูกต้อง
2. ในการสอนระบบ ถ้าสามารถเพิ่มจำนวนรอบในการสอนระบบให้มากกว่านี้จะทำให้ความแม่นยำของโปรแกรมสูงขึ้นซึ่งขึ้นกับผู้ป่วยถ้าสามารถพูดได้เหมือนเดิมทุกครั้งและไม่อ่อนเพลียระหว่างการพูด แต่ถ้าผู้ป่วยพูดได้ไม่เหมือน เดิมในแต่ละครั้งก็จะทำให้ระบบสูญเสียความแม่นยำได้เช่นกัน

บรรณานุกรม

[1] A.B. Kain, J.P. Hosom, Xi. Niu, J.P.H. van Santen, M. Fried-Oken, J. Staehely, “Improving the Intelligibility of Dysarthric Speech,” Speech Communication, vol. 49, September, 2007.

[2] W.A.Simm, P.E. Roberts, M.J. Joyce, “Signal Processing for use in the Assessment of Dysarthric Speech,” Medical Applications of Signal Processing, pp. 147-152, November, 2005.

[3] P.D. Polur, G.E. Miller, “Experiments with Fast Fourier Transform, Linear Predictive and Cepstral Coefficients in Dysarthric Speech Recognition Algorithms Using Hidden Markov Model,” Neural Systems and Rehabilitation Engineering, pp 558-561, December, 2005.

[4] F. Chen, A. Kostov, “Optimization of Dysarthric Speech Recognition,” Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 1436-1439, November, 1997.

[5] P. Kimsawad, “Study of Speaker Independent Isolated Word Recognition of Thai Digits using Backpropagation Neural Networks,” Department of Electrical Engineering, Prince of Songkla University, 2001.

[6] A. Thammaraksasit, “Control Commands Classification for Speech Recognition Based Wheelchair," Department of Electronics and Telecommunication Engineering, King Mongkut’s University of Technology Thonburi, 2003.

[7] K. Soontornmontakanti, “Optimization of Thai Tone Recognition using Autocorrelation with Center Clipping Method for Pitch Extraction in the Presence of White Gaussian Noise,” Department of Electrical Engineering, King Mongkut’s Institute of Technology Ladkrabang 2005.

[8] S. Sukanake, “Continuous Thai Tone Recognition using Tri-Half-Tone Hidden Markov Models,” Department of Computer Engineering, Kasetsart University, 2004.

[9] A. Wongthanasakchai, “Unmixed Vowels Recognition in Thai Spoken Language by using LPC Spectrum on the Effective Bank Scale,” Department of Electrical Engineering, King Mongkut’s Institute of Technology Ladkrabang, 2005.

[10] N. Wiratpokee, “Improvement of Thai Speech Pitch Detection for Gender Classification,” Department of Computer Technology, King Mongkut’s Institute of Technology North Bangkok, 2006.

[11] A. Deemagarn, “Speaker Independent Thai Connected Digit Speech Recognition System Using Hidden Markov Model,” Department of Computer Engineering, Kasetsart University, 2007.

[12] R. Vergin, D. O’Shaughnessy, “Pre-Emphasis and Speech Recognition,” Canadian Conference on Electrical and Computer Engineering, pp.1062-1065, September, 1995.

[13] N. Bunsakchalerm, “Application of Wavelet Transforms of Consonant Vowel Segmentation on Thai Speech Signal,” Department of Telecommunication Engineering, King Mongkut’s Institute of Technology Ladkrabang, 2006.

[14] C. Rowden, “Speech Processing,” Department of Electronic Systems Engineering University of Essex, McGRAW- HILL Book Company Europe.

[15] E. Panyathep, “The Isolated-Speech Thai-Vowel Recognition System Using Neuro-Fuzzy Networks,” Department of Information of Technology, King Mongkut’s Institute of Technology North Bangkok, 2005.

[16] K. Sittiprasert, “Short and Long Vowels Classifying in Thai Spoken Using 2^{nd} Order Polynomial Curve Fitting on the Voice Energy Function,” Department of Electrical Engineering, King Mongkut’s Institute of Technology Ladkrabang, 2004.

[17] T. Orzechowski, A. Izworski, R. Tadeusiewicz, K. Chmurzynska, P. Radkowski, I. Gatkowska, “Processing of Pathological Changes in Speech Caused by Dysarthria,” Proceedings of 2005 International Symposium on Intelligent Signal Processing and Communication Systems, December, 2005.

ภาคผนวก

การออกแบบระบบสั่งการแบบรู้จำเสียงพูดชนิดแยกคำขึ้นกับผู้พูดสำหรับผู้ป่วยโรคหลอดเลือดสมอง

Command System Design of Speaker Dependent Isolated Word Recognition for Stroke Patients

โอฬาร ดาวเวียง*, บุญเจริญ วงศ์กิตติศึกษา*, สาวิตร์ ตัณฑนุช*และ วุฒิชัย เพิ่มศิริวาณิชย์**

*ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ มหาวิทยาลัยสงขลานครินทร์

**ภาควิชาศัลยศาสตร์ออร์โธปิดิกส์และกายภาพบำบัด คณะแพทยศาสตร์ มหาวิทยาลัยสงขลานครินทร์

110/5 ถนน กาญจนวนิช ต.คอหงส์ อ.หาดใหญ่ จ.สงขลา 90112 โทรศัพท์: 074 287045-6

edaowieng@hotmail.com, booncharoen.w@psu.ac.th, sawit.t@psu.ac.th, pwuticha@medicine.psu.ac.th

บทคัดย่อ

โรคหลอดเลือดสมองเป็นโรคที่เกิดกับระบบประสาทส่งผล ต่อความสามารถในการพูด ทำให้ผู้ป่วยโรคนี้เปล่งเสียงได้ด้อยกว่าคนปกติ ระบบสั่งการโดยใช้เสียงทั่วไป จึงไม่เหมาะที่จะใช้กับกลุ่มคนเหล่านี้ การจะใช้ประโยชน์จากเสียงได้นั้น มีความจำเป็นอย่างยิ่งที่จะต้องมีกระบวนการในการคัดแยกเสียง ดังนั้นงานวิจัยนี้จึงได้นำเสนอแนวทางการออกแบบระบบสั่งการโดยใช้เสียงพูดขึ้นกับผู้พูดสำหรับผู้ป่วยโรคหลอดเลือดสมอง ที่มีความผิดปกติทางการสื่อสารชนิด Dysarthria ระหว่างเทคนิคฟาสฟูเรียร์ทรานฟอร์ม (FFT) และเวฟเล็ทร่วมกับฟาสฟูเรียร์ทรานฟอร์ม (DWT + FFT) โดยเปรียบเทียบกับผู้พูดปกติ ใช้คำสั่งเป็นเสียงสระในภาษาไทยจำนวน 6 คำ (อิ, เออะ, อะ, อือ, เออ, อา) จากผลการทดสอบกับผู้ป่วยระดับ 4(C4) และระดับ 6(C6) พบว่า วิธีการที่นำเสนอสามารถแยกเสียงของผู้ป่วยได้และการคัดแยกด้วยเทคนิค FFT ให้ผลแตกต่างเฉลี่ย 1.22% เมื่อเทียบกับเทคนิค DWT+FFT แต่สามารถลดจำนวนข้อมูลในการคำนวณได้สูงถึง 8 เท่า เมื่อเทียบกับเทคนิค FFT

Abstract

Stroke is associated with nervous system, affected to ability of speech. This speech is less intelligible than that of normal speaker. By the reason, conventional speech recognition is not suitable for stroke patient. Therefore, improving the speech recognition approach must be considered, especially classify speech scheme. This paper proposes the command system design of speaker dependent for stroke patient with dysarthric speech. Two techniques between Fast Fourier Transform (FFT) and Discrete Wavelet Transform with Fast Fourier Transform (DWT + FFT) were used and compare with normal speaker. The six vowels were used to recognize (ɯ, ɤ, a, ɯɯ, ʉ and aa). Our approach was tested with two types of patients (C4 and C6). Results show that the classification rates from both techniques are comparable. In other words, the percent of difference is only 1.22. However, the number of multiplications from the DWT+FFT technique is 8 times less than that from the FFT technique. This allow for a real-time implementation using a microcontroller.

คำสำคัญ

Speech Recognition, Dysarthria

1. บทนำ

ปัจจุบันระบบสั่งการโดยใช้เสียงพูดเป็นสิ่งสำคัญและเข้ามามีบทบาทเพื่อช่วยอำนวยความสะดวกในชีวิตประจำวันมากขึ้น ระบบสั่งการโดยใช้เสียงส่วนใหญ่นั้นได้ถูกออกแบบมาใช้สำหรับคนทั่วไปที่มีการพูดแบบปกติ ในขณะเดียวกันผู้ที่บกพร่องทางการพูดอันเนื่องมาจากโรคหลอดเลือดสมอง(Dysarthria) ยังไม่สามารถใช้งานได้ดีเท่าที่ควรกับระบบสั่งการเหล่านี้ จากการสำรวจเอกสารวิจัยพบว่าส่วนใหญ่ในการออกแบบระบบสั่งการไม่ได้พิจารณาถึงความบกพร่องนี้และความบกพร่องในแต่ละคนจะแตกต่างกันเนื่องจากความรุนแรงของโรคหลอดเลือดสมอง ดังนั้นงานวิจัยนี้จึงได้ทำการศึกษาแนวทางออกแบบระบบสั่งการเพื่อใช้สำหรับผู้ที่มีความบกพร่องทางการพูดชนิด Dysarthria และขึ้นกับผู้พูด โดยใช้โปรแกรม LABVIEW เพื่อคัดแยกความถี่ฟอร์แมนต์ ของคำสั่งระหว่างเทคนิคฟาสฟูเรียร์ทรานฟอร์ม (FFT) และเวฟเล็ทร่วมกับฟาสฟูเรียร์ทรานฟอร์ม (DWT + FFT) ในขณะเดียวกันได้ทำการทดลองควบคู่กับผู้พูดปกติเพื่อศึกษาความแตกต่างระหว่างผู้พูดทั้ง 2 ชนิด (Normal & Dysarthric speech) ผลการทดลองพบว่าทั้งสองเทคนิค

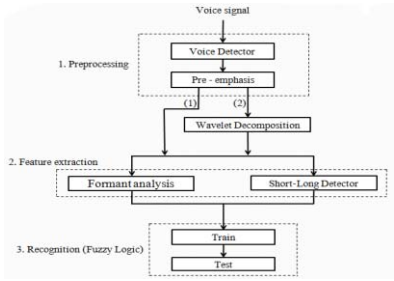
 ISBN: 978-974-8242-54-5

สามารถแยกความถี่ฟอร์แมนต์ได้ ในอีกทางหนึ่งระดับความผิดปกติของผู้ป่วยจะส่งผลกระทบโดยตรงต่อความถี่ฟอร์แมนต์ นั่นคือเมื่อระดับความรุนแรงของโรคหลอดเลือดสมองสูงขึ้น ความถี่ฟอร์แมนต์ก็จะผิดเพี้ยนไปจากความเป็นจริงมากขึ้นเมื่อเทียบกับผู้พูดปกติ ในหัวข้อถัดไปของบทความนี้จะอธิบายถึง งานวิจัยที่เกี่ยวข้อง ขั้นตอนและหลักการทำงาน วิธีการทดลอง ผลการทดลอง สรุปและวิเคราะห์ ตามลำดับ

2. งานวิจัยที่เกี่ยวข้อง

จากการสำรวจงานวิจัยในด้านการรู้จำเสียงพูดในประเทศไทยพบว่า ส่วนใหญ่เป็นงานรู้จำเสียงพูดเพื่อประยุกต์ใช้สำหรับสั่งการอุปกรณ์ต่างๆ ใช้งานร่วมกับผู้พูดปกติ โดยในช่วงแรกจะเป็นการรู้จำแบบแยกคำและถัดไปเป็นการรู้จำคำแบบต่อเนื่อง ซึ่งใช้เทคนิคในการรู้จำแตกต่างกันดังแสดงต่อไปนี้ งานวิจัย [1] ได้เสนอการศึกษารู้จำเสียงพูดตัวเลข 0-9 แบบแยกคำชนิดไม่ขึ้นกับผู้พูด โดยใช้โครงข่ายประสาทเทียมที่มีการเรียนรู้แบบแพร่กลับ งานวิจัย [2] ได้นำเสนอกระบวนการแยกแยะคำสั่งในระบบรู้จำเสียงพูดสำหรับควบคุมรถเข็นคนพิการในการเคลื่อนที่ไปในทิศทางต่างๆ ใช้วิธีการค้นหาแบบบิลเนียร์และทฤษฎีของเบย์ในการคัดแยกคำสั่ง โดยใช้กับคนพิการทั่วไป ยังไม่ได้คำนึงถึงผลกระทบจากความบกพร่องทางการพูด งานวิจัย [3] ได้นำเสนอระบบการรู้จำเสียงวรรณยุกต์แบบต่อเนื่อง ใช้วิธีการแยกความถี่พื้นฐานตามแบบของพอล เบอร์สมา(Paul Boersma) ร่วมกับแบบจำลองฮิดเดนมาร์คอฟประเภทกึ่งต่อเนื่อง งานวิจัย [4] ได้นำเสนอระบบรู้จำเสียงพูดตัวเลขต่อเนื่องที่ไม่ขึ้นกับบุคคล โดยใช้ทฤษฎีแบบจำลอง ฮิดเดนมาร์คอฟแบบต่อเนื่อง งานวิจัย [5] ได้นำเสนอการศึกษาการรู้จำเสียงพูด ใช้สำหรับเด็กที่มีความผิดปกติทางการพูดชนิด Dysarthria โดยใช้เทคนิค Modified speech clarity index (..) ในการคัดแยกเสียง จากงานวิจัย [1] ถึง [4] พบว่าเป็นงานวิจัยที่เน้นใช้งานกับผู้พูดปกติและไม่ได้คำนึงถึงผลกระทบจากผู้บกพร่องทางการพูด ส่วนงานวิจัยที่ [5] ได้คำนึงถึงผลกระทบจากความผิดปกติทางการพูดสำหรับเด็กเท่านั้น ซึ่งความแตกต่างจากความผิดปกติทางการพูดนี้ เกิดจากปัจจัยต่างๆทั้งด้านอายุและโรคที่เกิดกับสมองของตัวบุคคล [6] ดังนั้นงานวิจัยนี้จึงได้ออกแบบระบบสั่งการสำหรับผู้ใหญ่ที่ป่วยเป็นโรคหลอดเลือดสมองและมีอาการร่วมกับความผิดปกติทางการพูดชนิด Dysarthria โดยมีจำนวนคำสั่งไม่มากและเน้นออกเสียงง่าย เพื่อให้สามารถใช้งานได้มีประสิทธิภาพสูงสุดกับผู้ที่มีอาการอัมพาตบริเวณริมฝีปากเนื่องจากความบกพร่องทางด้านสมอง

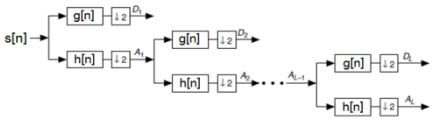
3. ขั้นตอนและหลักการทำงาน



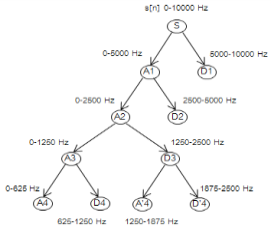
รูปที่ 1 ขั้นตอนการรู้จำเสียงพูด

ในงานวิจัยนี้ได้แบ่งขั้นตอนการรู้จำเสียงพูด ดังแสดงในรูปที่ 1 เริ่มจากขั้นตอน Voice Detector ทำหน้าที่ตรวจจับสัญญาณเสียงพูด โดยพิจารณาจากค่าพลังงานกำลังสอง [7] ถัดไปเป็นขั้นตอน Pre-emphasis เพื่อให้อัตราส่วนสัญญาณเสียงต่อสัญญาณรบกวน (Signal- to-Noise Ratio: SNR) มีค่าค่อนข้างคงที่ตลอดช่วงความถี่ [8] จากนั้นจะเป็นการเปรียบเทียบคุณสมบัติโดยแยกออกเป็นสองแนวทาง คือ (1) สัญญาณถูกส่งไปยังขั้นตอน Feature extraction และ (2) สัญญาณถูกแยกส่วนประกอบความถี่โดยขั้นตอน Wavelet Decomposition และส่งไปยังขั้นตอน Feature extraction ถัดไปเป็นขั้นตอน Formant analysis ทำหน้าที่ตรวจจับความถี่ฟอร์แมนต์ และขั้นตอน Short-Long Detector ทำหน้าที่แยกสระเสียงสั้นและเสียงยาว ในขั้นตอนสุดท้ายเป็นการรู้จำของระบบ ประกอบด้วยการสอนระบบ (Train) และทดสอบระบบ (Test) โดยใช้เทคนิคฟัซซี่ลอจิก (Fuzzy Logic) [9] ตามลำดับ

ที่ขั้นตอน Wavelet Decomposition [10] งานวิจัยนี้ใช้เวฟเล็ทแม่แบบ Harr ในการแยกส่วนประกอบความถี่ของสัญญาณ ซึ่งสามารถแยกส่วนประกอบความถี่ออกมาเป็นสัญญาณในแต่ละระดับดังแสดงในรูปที่ 2 และ 3 ตามลำดับ

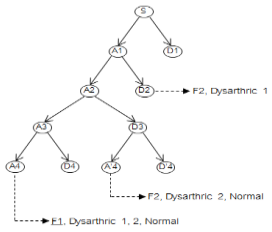


รูปที่ 2 แผนภูมิการกระจายเวฟเล็ทสำหรับการแยกส่วนประกอบความถี่

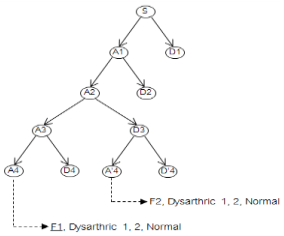


รูปที่ 3 แผนภูมิต้นไม้สำหรับการแยกส่วนประกอบความถี่

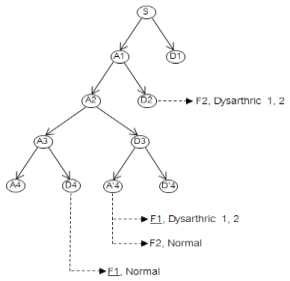
สัญญาณ s[n] เป็นสัญญาณที่ได้มาจากขั้นตอน Preprocessing เมื่อทำการแยกส่วนประกอบความถี่ด้วยวิธีเวฟเล็ทจะได้ส่วนประกอบสัญญาณความถี่สูง g[n] และส่วนประกอบสัญญาณความถี่ต่ำ h[n] โดยมีอัตราสุ่ม (Sampling Rate) ของสัญญาณลดลงเหลือครึ่งหนึ่งของอัตราสุ่มเดิม หลังจากนั้นจึงนำส่วนประกอบสัญญาณความถี่ต่ำไปแยกส่วนประกอบความถี่ของสัญญาณในระดับต่อไปจนถึงระดับที่ต้องการ งานวิจัยนี้ได้วิเคราะห์ระดับสัญญาณรายละเอียดอยู่ที่ระดับ 4 (Wavelet Decomposition level 4) หลังจากนั้นสัญญาณก็จะถูกส่งต่อไปยังขั้นตอน Formant analysis เฉพาะบางช่วงสัญญาณโดยพิจารณาจากคุณลักษณะของความถี่ฟอร์แมนต์ทั้งสอง ดังตารางที่ 1 หัวข้อที่ 4 ซึ่งสามารถสรุปย่านความถี่ปฏิบัติงานในแต่ละคำโดยใช้แผนภูมิต้นไม้ ดังรูปที่ 4-6 (กำหนดให้ F1, F2 = ความถี่ฟอร์แมนต์ที่ 1, 2 Dysarthric 1, 2 คือ ผู้พูดชนิด Dysarthric คนที่ 1, 2 และ Normal คือ ผู้พูดปกติ ตามลำดับ)



รูปที่ 4 การเลือกช่วงความถี่ปฏิบัติงานโดยใช้แผนภูมิการกระจายเวฟเล็ท (A4, A'4, D2) เปรียบเทียบระหว่างผู้ทดสอบทั้ง 3 (เสียงสระ "อี-เอือ")



รูปที่ 5 การเลือกช่วงความถี่ปฏิบัติงานโดยใช้แผนภูมิการกระจายเวฟเล็ท(A4, A'4) เปรียบเทียบระหว่างผู้ทดสอบทั้ง 3 (เสียงสระ "เออะ-เออ")



รูปที่ 6 การเลือกช่วงความถี่ปฏิบัติงานโดยใช้แผนภูมิการกระจายเวฟเล็ท(D4, A'4, D2) เปรียบเทียบระหว่างผู้ทดสอบทั้ง 3 (เสียงสระ "อะ-อา")

จากรูปที่ 6 ในการวิเคราะห์เสียงสระ "อะ-อา" กรณีผู้ป่วยทั้ง สองจะเห็นว่าความถี่ฟอร์แมนต์ที่ 1 อยู่ในช่วง 1250 Hz ขึ้นไป และความถี่ฟอร์แมนต์ที่ 2 อยู่ในช่วง 2500 Hz ขึ้นไป ดังนั้นในการเลือกความถี่ปฏิบัติงาน จึงเลือกเฉพาะช่วง A'4 และ D2 ซึ่งแตกต่างกับผู้พูดปกติ จะเลือกเฉพาะช่วง D4 และ A'4

ที่ขั้นตอน Formant analysis เป็นการสกัดค่าความถี่ฟอร์แมนต์โดยใช้เทคนิคฟาสฟูเรียร์ทรานฟอร์ม ดังสมการที่ (1) เมื่อ Y(k) คือชุดแถวของสัญญาณที่ได้จากการแปลงฟูเรียร์ โดยที่ N คือ จำนวนข้อมูลในการแปลงฟูเรียร์และ k มีค่าตั้งแต่ 0 จนถึงจำนวน N-1

$$Y(k) = \sum_{n=0}^{N-1} x_n e^{-j2\pi kn/N} \quad (1)$$
$$\text{for n=0, 1, 2,...,N-1}$$

กำหนดให้ค่าที่อยู่ระหว่าง Y(k) ทางด้านแกน x (frequency resolution) คือ $\div F = f_s / N$ โดย f_s คือ อัตราสุ่ม (Sampling Rate) ของสัญญาณ

4. วิธีการทดลอง

งานวิจัยนี้ได้นำเสนอแนวทางในการออกแบบระบบสั่งการโดยใช้เสียงพูดสำหรับผู้พูดผิดปกติชนิด Dysarthria แบ่งออกเป็น 2 การทดลอง คือ การทดลองแรกใช้เทคนิคฟาสฟูเรียร์ทรานฟอร์ม (Fast Fourier Transform; FFT) และการทดลองที่สองใช้เวฟเล็ตรวมกับฟาสฟูเรียร์ทรานฟอร์ม (DFT + FFT) ร่วมกับเสียงสระในภาษาไทยจำนวน 6 คำ ประกอบด้วยสระเสียงสั้นจำนวน 3 คำ(อึ เออะ อะ) และสระเสียงยาวจำนวน 3 คำ (อือ เออ อา) รวมทั้งสิ้น 6 คำสั่ง โดยงานวิจัยนี้จะเป็นการศึกษาและทดลองในกระบวนการคัด

แยกเสียง มีขอบเขตการทดลองสิ้นสุดที่กระบวนการ Formant analysis ดังรูปที่ 1 ร่วมกับกลุ่มผู้ทดสอบเป็นเพศชาย จำนวน 3 คน คือ

1.ผู้พูดผิดปกติชนิด Dysarthria อายุ 57 ปี อัมพาตครึ่งซีก (Hemiplegia) ระดับ 6 (C6)

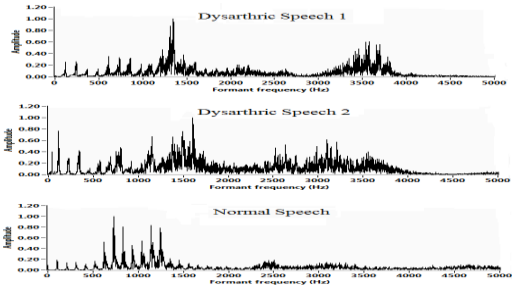
2.ผู้พูดผิดปกติชนิด Dysarthria อายุ 70 ปี อัมพาตครึ่งซีก (Hemiplegia) ระดับ 4 (C4)

3.ผู้พูดปกติ อายุ 25 ปี

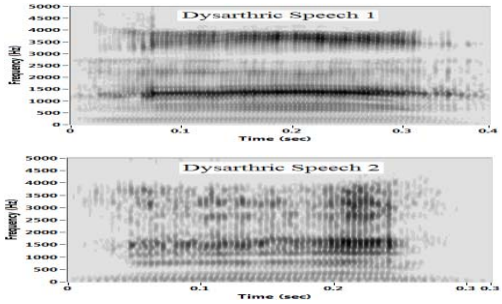
ทำการเก็บสัญญาณเสียงใช้โปรแกรม LABVIEW version 8.6 ร่วมกับไมโครโฟนยี่ห้อ Genius (Desktop Microphone) ซึ่งมีย่านความถี่ตอบสนองในช่วง 0 – 10 kHz ความถี่ในการสุ่มตัวอย่างเท่ากับ 20 kHz เก็บบันทึกในไฟล์ข้อมูลรูปแบบ .lvm โดยให้กลุ่มผู้ทดสอบเปล่งเสียงสระ ทั้งหมด 5 ครั้งในแต่ละคำ และทำการวัดค่าความถี่ ประกอบด้วย ความถี่ฟอร์แมนต์ที่ 1 (F1) และความถี่ฟอร์แมนต์ที่ 2 (F2) ตามลำดับ

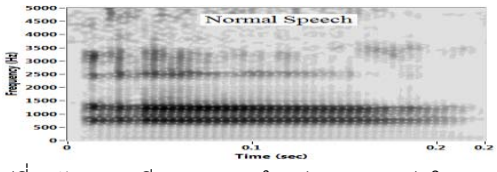
5. ผลการทดลอง

การทดลองแรกเป็นการทดสอบการแยกความถี่ฟอร์แมนต์โดยใช้เทคนิค FFT ยกตัวอย่างดังรูปที่ 7 และ 8 คือ สัญญาณเสียงในโดเมนความถี่และสเปกโตรแกรมของเสียงสระ “อา” จากกลุ่มผู้ทดสอบทั้ง 3



รูปที่ 7 สัญญาณเสียงสระ “อา” ในโดเมนความถี่





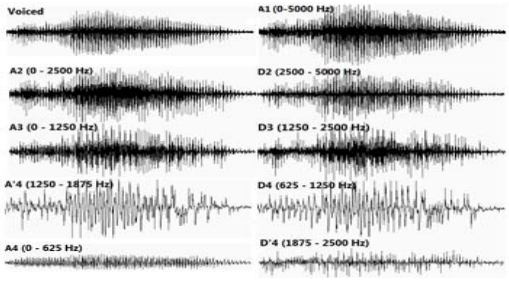
รูปที่ 8 สัญญาณเสียงสระ “อา” ในรูปแบบของสเปกโตแกรม

ตารางที่ 1 สรุปค่าความถี่ฟอร์แมนต์เฉลี่ยของกลุ่มผู้ทดสอบทั้ง 3 คน โดยใช้เทคนิค FFT

ประเภท	Dysarthric 1		Dysarthric 2		Normal	
ความถี่ / สระ	F1 (Hz)	F2 (Hz)	F1 (Hz)	F2 (Hz)	F1 (Hz)	F2 (Hz)
อึ-อือ	301.0	2861.0	318.7	1562.9	368.8	1255.6
เออะ-เออ	517.7	1648.1	449.0	1295.5	499.3	1222.0
อะ-อา	1349.7	3648.1	1526.6	3100.9	787.4	1243.6

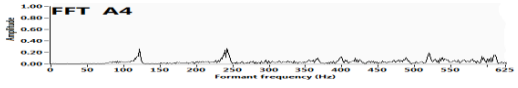
หมายเหตุ; เสียง อึ-อือ, เออะ-เออ และ อะ-อา ถือว่าเป็นเสียงเดียวกัน ต่างกันที่ช่วงเวลาในการออกเสียงเท่านั้น

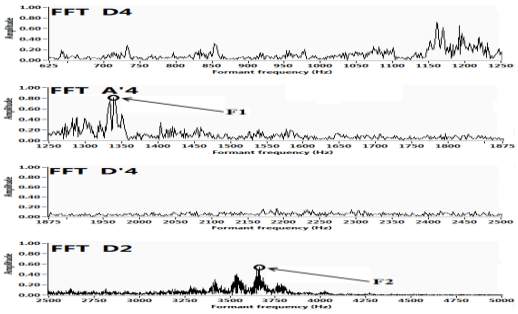
การทดลองต่อมาเป็นการนำเทคนิคเวฟเล็ทมาใช้ในการแยกส่วนประกอบความถี่ ก่อนที่จะทำการแยกความถี่ฟอร์แมนต์โดยใช้เทคนิค FFT แสดงตัวอย่างดังรูปที่ 9 ผลจากการใช้ Wavelet Decomposition level 4 ในการแยกส่วนประกอบความถี่ในโดเมนเวลาออกเป็น 10 ช่วง



รูปที่ 9 สัญญาณรายละเอียดที่ได้จากการแปลงเวฟเล็ทของเสียงสระ “อา” (Dysarthric Speech 1)

หลังจากเลือกช่วงความถี่ปฏิบัติงานแล้ว สัญญาณเสียงในแต่ละช่วงก็จะถูกนำมาแยกความถี่ฟอร์แมนต์โดยใช้เทคนิค FFT ดังรูปที่ 10 (แสดงเฉพาะช่วง D2, D4, D’4, A4 และ A’4)

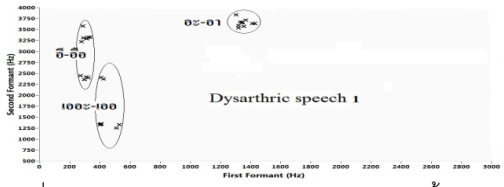




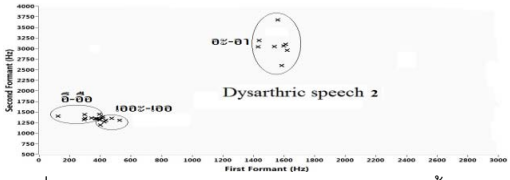
รูปที่ 10 สัญญาณเสียงสระ “อา” (Dysarthric Speech 1) ในโดเมนความถี่แต่ละช่วง

ตารางที่ 2 สรุปค่าความถี่ฟอร์แมนต์เฉลี่ยของกลุ่มผู้ทดสอบทั้ง 3 คน โดยใช้เทคนิค DWT + FFT

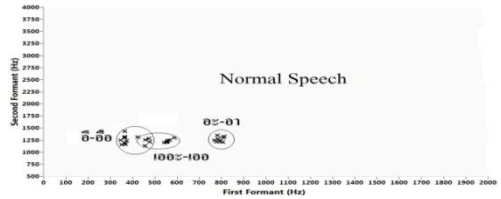
ประเภท	Dysarthric 1		Dysarthric 2		Normal	
ความถี่ / สระ	F1 (Hz)	F2 (Hz)	F1 (Hz)	F2 (Hz)	F1 (Hz)	F2 (Hz)
อี-อือ	301.8	2967.3	316.2	1555.3	372.8	1257.2
เออะ-เออ	519.3	1651.2	426.2	1314.9	516.9	1231.6
อะ-อา	1351.9	3651.3	1545.8	3085.5	791.1	1261.7



รูปที่ 11 กราฟแจกแจงแบบกระจายของเสียงสระทั้ง 6 คำ (Dysarthric Speech 1) โดยใช้เทคนิค DWT + FFT



รูปที่ 12 กราฟแจกแจงแบบกระจายของเสียงสระทั้ง 6 คำ (Dysarthric Speech 2) โดยใช้เทคนิค DWT + FFT



รูปที่ 13 กราฟแจกแจงแบบกระจายของเสียงสระทั้ง 6 คำ (Normal Speech) โดยใช้เทคนิค DWT + FFT

6. สรุปและวิเคราะห์

จากตารางที่ 1 และ 2 สังเกตที่ค่าเฉลี่ยความถี่ฟอร์แมนต์ทั้งสอง เห็นว่าจะแตกต่างกับคนปกติ ทำให้ทราบความผิดปกติของผู้ป่วยอันสืบเนื่องมาจากการสูญเสียการควบคุมของลิ้น กล่าวคือ ความถี่ฟอร์แมนต์ที่ 1 ของเสียงสระอะ-อา มีความถี่สูงกว่าปกติ เนื่องมาจากตำแหน่งของลิ้นที่ต่ำกว่าปกติในขณะเปล่งเสียง ส่วนความถี่ฟอร์แมนต์ที่ 2 ของเสียงสระทั้งหมด มีความถี่สูงกว่าปกติ เนื่องมาจากการเคลื่อนที่ไปข้างหน้าของลิ้นมากเกินไปในขณะเปล่งเสียง นอกจากนั้นความผิดปกติของเสียงอาจเกิดจากความผิดปกติร่วมกันของอวัยวะอื่นๆในการกำเนิดเสียงได้เช่นกัน

ตารางที่ 3 เปอร์เซ็นต์ความแตกต่างของความถี่ฟอร์แมนต์เปรียบเทียบระหว่างเทคนิคทั้งสอง

ประเภท	Dysarthric 1		Dysarthric 2		Normal	
ความถี่ / สระ	F1 (%)	F2 (%)	F1 (%)	F2 (%)	F1 (%)	F2 (%)
อี-อือ	0.3	3.7	0.8	0.5	1.1	0.1
เออะ-เออ	0.3	0.2	5.0	1.5	3.5	0.8
อะ-อา	0.2	0.1	1.3	0.5	0.5	1.5

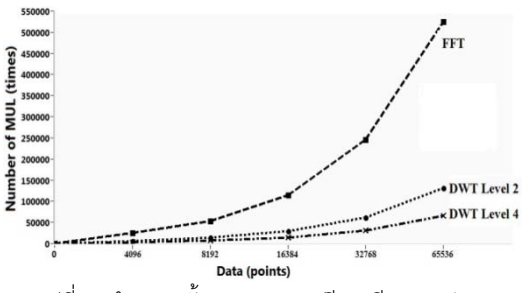
จากตารางที่ 3 สังเกตที่ผู้ป่วย Dysarthric 1 ความถี่ฟอร์แมนต์ที่ 1(F1) เสียงสระ อี-อือ มีค่า 0.3 นั่นหมายความว่า ความแตกต่างของผลลัพธ์ระหว่างเทคนิคทั้งสอง มีความแตกต่างกัน 0.3% ในทำนองเดียวกันกับคำอื่นๆ จะเห็นว่าเปอร์เซ็นต์ความแตกต่างมีค่าน้อยมาก คือ อยู่ในระหว่าง 0 ถึง 5 เปอร์เซ็นต์และมีค่าเฉลี่ยเท่ากับ 1.22 เปอร์เซ็นต์ ความแตกต่างนั้นเกิดจากการเจือปนของสัญญาณไม่เท่ากัน กล่าวคือการเลือกช่วงความถี่ปฏิบัติงาน โดยกระบวนการ Wavelet Decomposition (รูปที่ 4 ถึง 6) จะช่วยขจัดสัญญาณรบกวนและช่วงความถี่ที่ไม่เกี่ยวข้องออกไป จึงทำให้ผลลัพธ์ที่ได้มีความน่าเชื่อถือมากกว่าเทคนิค FFT เพียงอย่างเดียว

วิธีการคัดแยกเสียงสระทั้งสองวิธีสามารถคัดแยกเสียงผู้ป่วยชนิดนี้ได้ แต่จะมีความแตกต่างกันในด้านกระบวนการคำนวณ คือ ส่วนของเทคนิค FFT จะเป็นการคำนวณที่ง่ายไม่ซับซ้อน แต่จะมีจำนวนการคูณกันของข้อมูลที่เยอะกว่าการใช้เทคนิค DWT + FFT กล่าวคือ ในสัญญาณเสียงเดียวกัน การใช้เทคนิค FFT โปรแกรมจะต้อง

ทำการคูณทั้งหมด ($N/2$)$\text{Log}_2 N$ ครั้ง (N คือ จำนวน sample ของสัญญาณ) เปรียบเทียบกับการใช้ Wavelet Decomposition ในการกรองช่วงความถี่ก่อนเข้าสู่กระบวนการ FFT ซึ่งเมื่อสัญญาณผ่านกระบวนการกรองความถี่ในแต่ละเลเวล จะทำให้จำนวนข้อมูลในการสุ่มตัวอย่างลดลงครึ่งหนึ่ง ($N/2$) ดังนั้นการคำนวณมีการคูณทั้งสิ้น ($N/2$ $\int j$)$\text{Log}_2 N$ ครั้ง (j คือเลเวลของ Wavelet Decomposition) ซึ่งช่วยลดระยะเวลาในการคำนวณจากจำนวนครั้งของการคูณจะมีผลกระทบอย่างมากเมื่อประยุกต์ใช้ในระดับไมโครคอลโทรลเลอร์ แสดงดังตารางที่ 4 และรูปที่ 13 เปรียบเทียบจำนวนครั้งของการคูณระหว่างเทคนิค FFT และ เทคนิค DWT(Level 2&4) +FFT

ตารางที่ 4 จำนวนครั้งของการคูณ เปรียบเทียบระหว่างเทคนิค FFT และ DWT+FFT

จำนวนข้อมูล (จุด)	FFT (ครั้ง)	DWT(L.2)+FFT **(ครั้ง)**	DWT(L.4)+FFT **(ครั้ง)**
4096	24576	6144	3072
8192	53248	13312	6656
16384	114688	28672	14336
32768	245760	61440	30720
65536	524288	131072	65536



รูปที่ 13 จำนวนครั้งของการคูณเปรียบเทียบระหว่างเทคนิคทั้งสองที่ความถี่สุ่มตัวอย่าง = 20kHz

7. เอกสารอ้างอิง

[1] P. Kimsawad, "Study of Speaker Independent Isolated Word Recognition of Thai Digits Using Backpropagation Neural Networks," Department of Electrical Engineering, Prince of Songkla University, 2000.

[2] A. Thammaraksasit, "Control Command Classification for Speech Recognition Based Wheelchair," Department of Electronics and Telecommunication Engineering, King Mongkut's University of Technology Thonburi, 2003.

[3] S. Sukanake, "Continuous Thai Tone Recognition using Tri-Half-Tone Hidden Markov Models," Department of Computer Engineering, Kasetsart University, 2004.

[4] A. Deemagarn, "Speaker Independent Thai Connected Digit Speech Recognition System Using Hidden Markov Model," Department of Computer Engineering, Kasetsart University, 2007.

[5] P. Kayasith, T.Theeramunkong, N. Thubthong, "Incorporated Speech Overlapped Factor (λ) into Speech Clarity Index (..): Method to Improve Dysarthric Speech Severity Evaluation," International Convention for Rehabilitation Engineering and Assistive Technology (i-CREATe 2007), Singapore, pp.133-138, April, 2007.

[6] "การฝึกพูดสำหรับผู้ป่วยโรคหลอดเลือดสมองที่มีปัญหาทางด้านการพูด," ภาควิชาโสต ศอ นาสิกวิทยา คณะแพทยศาสตร์ โรงพยาบาลรามาธิบดี, 1/11/51, [online]. <http://www.vichaiyut.co.th>

[7] C. Rowden, "Speech Processing," Department of Electronic Systems Engineering University of Essex, McGRAW- HILL Book Company Europe.

[8] R. Vergin, D. O'Shaughnessy, "Pre-Emphasis and Speech Recognition," Canadian Conference on Electrical and Computer Engineering, vol 2, pp.1062-1065, September, 1995.

[9] E. Panyathep, "The Isolated-Speech Thai-Vowel Recognition System Using Neuro-Fuzzy Networks," Department of Information of Technology, King Mongkut's Institute of Technology North Bangkok, 2005.

[10] N. Bunsakchalerm, "Application of Wavelet Transforms of Consonant Vowel Segmentation on Thai Speech Signal," Department of Telecommunication Engineering, 2006.

การแยกแยะคำสั่งในระบบรู้จำเสียงพูดสำหรับผู้ป่วยโรคหลอดเลือดสมองโดยใช้วิธีการแปลงเวฟเล็ทร่วมกับฟัซซี่ลอจิก

Commands Classification of Speech Recognition for Stroke Patient Using Wavelet Transform and Fuzzy Logic

โอฬาร ดาวเวียง*, บุญเจริญ วงศ์กิตติศึกษา*, สาวิตร์ ตัณฑนุช*และ วุฒิชัย เพิ่มศิริวาณิชย์**

*ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ มหาวิทยาลัยสงขลานครินทร์

**ภาควิชาศัลยศาสตร์ออร์โธปิดิกส์และกายภาพบำบัด คณะแพทยศาสตร์ มหาวิทยาลัยสงขลานครินทร์

110/5 ถนน กาญจนวนิช ต.คอหงส์ อ.หาดใหญ่ จ.สงขลา 90112 โทรศัพท์: 074 287045-6

edaowieng@hotmail.com, booncharoen.w@psu.ac.th, sawit.t@psu.ac.th, pwutichai@medicine.psu.ac.th

บทคัดย่อ

โรคหลอดเลือดสมองเป็นโรคที่เกิดกับระบบประสาท ส่งผล ต่อความสามารถในการพูด ทำให้ผู้ป่วยโรคนี้เปล่งเสียงได้ด้อยกว่าคนปกติ ระบบสั่งการโดยใช้เสียงทั่วไปจึงไม่เหมาะที่จะใช้กับกลุ่มคนเหล่านี้ การจะใช้ประโยชน์จากเสียงได้นั้น มีความจำเป็นอย่างยิ่งที่จะต้องมีกระบวนการในการคัดแยกเสียง ดังนั้นงานวิจัยนี้จึงได้นำเสนอการออกแบบระบบสั่งการโดยใช้เสียงพูดขึ้นกับผู้พูดสำหรับผู้ป่วยโรคหลอดเลือดสมองที่มีภาวะกลไกการออกเสียงบกพร่อง (dysarthria) โดยใช้วิธีการแปลงเวฟเล็ทและฟัซซี่ลอจิก ร่วมกับคำสั่งเสียงสระในภาษาไทยจำนวน 6 คำ (อี, เออะ, อะ, อือ, เออ, อา) จากผลการทดสอบกับผู้ป่วยระดับ 4(C4) และระดับ 6(C6) พบว่าวิธีการที่นำเสนอสามารถแยกเสียงของผู้ป่วยได้ดีกว่าคนปกติ โดยให้อัตราการรู้จำสูงถึง 97.2 % และ 87.0% สำหรับกลุ่มผู้ทดสอบผู้ป่วยอัมพาตระดับ 6 (C6) และ ระดับ 4 (C4) ตามลำดับ

Abstract

Stroke is associated with nervous system and affected to ability of speech. This speech is less intelligible than that of normal speaker. By the reason, conventional speech recognition is not suitable for stroke patient. Therefore, improving the speech recognition approach must be considered, especially classification speech scheme. This paper proposes the command system design of speaker dependence for stroke patient with dysarthric speech using Wavelet Transforms and Fuzzy Logic. The six vowels (ϖ, V, a, ϖϖ, VV and aa) were involved with two-type of the subjects (C4 and C6). Results show that the proposed methods can be appropriately used for the stroke patient. The recognition rates achieve 97.2% and 87.0% in C6 and C4 stroke patients, respectively.

คำสำคัญ: Speech Recognition, Dysarthria

1. บทนำ

ปัจจุบันระบบสั่งการโดยใช้เสียงพูดเป็นสิ่งสำคัญและเข้ามามีบทบาทเพื่อช่วยอำนวยความสะดวกในชีวิตประจำวันมากขึ้น ระบบสั่งการโดยใช้เสียงพูดส่วนใหญ่นั้นได้ถูกออกแบบมาใช้สำหรับคนทั่วไปที่มีการพูดแบบปกติ ในขณะเดียวกันผู้ที่มีภาวะกลไกการออกเสียงบกพร่องอันเนื่องมาจากโรคหลอดเลือดสมอง ยังไม่สามารถใช้งานได้ดีเท่าที่ควรกับระบบสั่งการเหล่านี้ เป็นที่ทราบกันว่า จากภาวะกลไกการออกเสียงบกพร่องจะส่งผลต่อการควบคุมกล้ามเนื้อบริเวณอวัยวะที่เกี่ยวข้องกับการพูด ทำให้เกิดการสูญเสียการควบคุมของลิ้นและอวัยวะอื่นๆในขณะเปล่งเสียง จึงทำให้เกิดความผิดเพี้ยนของความถี่ฟอร์แมนต์เมื่อเทียบกับผู้พูดปกติ จากการสำรวจเอกสารวิจัยพบว่าส่วนใหญ่ในการออกแบบระบบสั่งการไม่ได้พิจารณาถึงความบกพร่องนี้ ดังนั้นงานวิจัยนี้จึงได้ทำการศึกษาการออกแบบระบบสั่งการเพื่อใช้สำหรับผู้ป่วยโรคหลอดเลือดสมองที่มีภาวะกลไกการออกเสียงบกพร่องและขึ้นกับผู้พูด โดยใช้โปรแกรม LABVIEW เพื่อคัดแยกความถี่ฟอร์แมนต์ ของคำสั่งโดยใช้เทคนิคเวฟเล็ทร่วมกับฟาสฟูเรียร์ทรานฟอร์ม (DWT + FFT) และฟัซซี่ลอจิก (Fuzzy Logic) ในการรู้จำ ในขณะเดียวกันได้ทำการทดลองควบคู่กับผู้พูดปกติ เพื่อศึกษาความแตกต่างระหว่างผู้พูดทั้ง 2 ชนิด (Normal & Dysarthric speech) ผลการทดลองพบว่า

ทั้งเทคนิคฟัซซี่ลอจิกสามารถแยกคำสั่งผู้ป่วยได้ดีกว่าผู้พูดปกติ ในหัวข้อถัดไปของบทความนี้จะอธิบายถึง งานวิจัยที่เกี่ยวข้อง ขั้นตอนและหลักการทำงาน วิธีการทดลอง ผลการทดลอง สรุปและวิเคราะห์ ตามลำดับ

2. งานวิจัยที่เกี่ยวข้อง

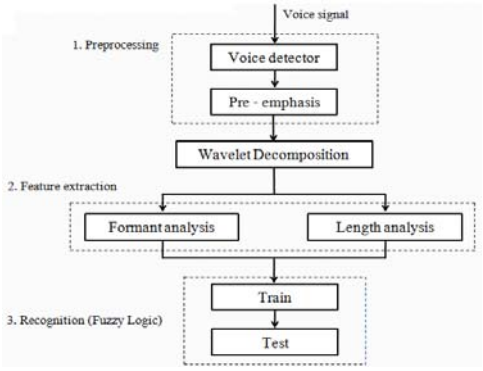
จากการสำรวจงานวิจัยในด้านการรู้จำเสียงพูดในประเทศไทยพบว่า ส่วนใหญ่เป็นงานรู้จำเสียงพูดเพื่อประยุกต์ใช้สำหรับสั่งการอุปกรณ์ต่างๆ ใช้งานร่วมกับผู้พูดปกติ โดยในช่วงแรกจะเป็นการรู้จำแบบแยกคำและถัดไปเป็นการรู้จำคำแบบต่อเนื่อง ซึ่งใช้เทคนิคในการรู้จำแตกต่างกันดังแสดงต่อไปนี้

งานวิจัย [1] ได้เสนอการศึกษารู้จำเสียงพูดตัวเลข 0-9 แบบแยกคำชนิดไม่ขึ้นกับผู้พูด โดยใช้โครงข่ายประสาทเทียมที่มีการเรียนรู้แบบแพร่กลับ งานวิจัย [2] ได้นำเสนอกระบวนการแยกแยะคำสั่งในระบบรู้จำเสียงพูดสำหรับควบคุมรถเข็นคนพิการในการเคลื่อนที่ไปในทิศทางต่างๆ ใช้วิธีการค้นหาแบบลิเนียร์และทฤษฎีของเบย์ในการคัดแยกคำสั่ง โดยใช้กับคนพิการทั่วไป ยังไม่ได้คำนึงถึงผลกระทบจากความบกพร่องทางการพูด งานวิจัย [3] ได้นำเสนอระบบการรู้จำเสียงวรรณยุกต์แบบต่อเนื่อง ใช้วิธีการแยกความถี่พื้นฐานตามแบบของพอล เบอร์สมา(Paul Boersma) ร่วมกับแบบจำลองฮิดเดนมาร์คอฟประเภทกึ่งต่อเนื่อง งานวิจัย [4] ได้นำเสนอระบบรู้จำเสียงพูดตัวเลขต่อเนื่องที่ไม่ขึ้นกับบุคคล โดยใช้ทฤษฎีแบบจำลอง ฮิดเดนมาร์คอฟแบบต่อเนื่อง งานวิจัย [5] ได้นำเสนอการคัดแยกเสียงร่วมกับกลุ่มผู้ทดสอบเด็กที่มีภาวะกลไกการออกเสียงบกพร่อง จากงานวิจัย [1] ถึง [4] พบว่าเป็นงานวิจัยที่เน้นใช้งานกับผู้พูดปกติและไม่ได้คำนึงถึงผลกระทบจากผู้บกพร่องทางการพูด ส่วนงานวิจัยที่ [5] ได้คำนึงถึงผลกระทบจากความผิดปกติทางการพูดสำหรับเด็กเท่านั้น ซึ่งความแตกต่างจากความผิดปกติทางการพูดนี้เกิดจากปัจจัยต่างๆทั้งด้านอายุและโรคที่เกิดกับสมองของตัวบุคคล [6]

ดังนั้นงานวิจัยนี้จึงได้ออกแบบระบบสั่งการสำหรับผู้ใหญ่ที่ป่วยเป็นโรคหลอดเลือดสมองที่มีภาวะกลไกการออกเสียงบกพร่อง โดยมีจำนวนคำสั่งไม่มากและเน้นออกเสียงง่ายเพื่อให้สามารถใช้งานได้มีประสิทธิภาพสูงสุดกับผู้ที่มีอาการอัมพาตบริเวณริมฝีปากเนื่องจากความบกพร่องทางด้านสมอง

3. ขั้นตอนและหลักการทำงาน

3.1 ขั้นตอนการรู้จำเสียงพูด



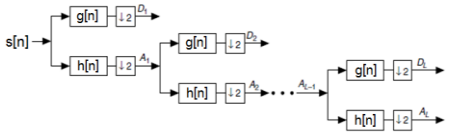
รูปที่ 1 ขั้นตอนการรู้จำเสียงพูด

ในงานวิจัยนี้ได้แบ่งขั้นตอนการรู้จำเสียงพูด ดังแสดงในรูปที่ 1 เริ่มจากขั้นตอน Voice detector ทำหน้าที่ตรวจจับสัญญาณเสียงพูด โดยพิจารณาจากค่าพลังงานกำลังสอง [7] ถัดไปเป็นขั้นตอน Pre-emphasis เพื่อให้อัตราส่วนสัญญาณเสียงต่อสัญญาณรบกวน (Signal- to-Noise Ratio: SNR) มีค่าค่อนข้างคงที่ตลอดช่วงความถี่ [8] จากนั้นสัญญาณจะถูกแยกส่วนประกอบความถี่โดยขั้นตอน Wavelet Decomposition และส่งไปยังขั้นตอน Feature extraction อันดับแรกแรกเป็นขั้นตอน Formant analysis ทำหน้าที่ตรวจจับความถี่ฟอร์แมนต์ที่หนึ่งและสอง และถัดไปเป็นขั้นตอน Length analysis ทำหน้าที่แยกสระเสียงสั้นและเสียงยาวโดยวิเคราะห์จากค่าพลังงานเสียงและในขั้นตอนสุดท้ายเป็นการรู้จำของระบบ ประกอบด้วยการสอนระบบ (Train) และทดสอบระบบ (Test) โดยใช้เทคนิคฟัซซี่ลอจิก [9] ตามลำดับ

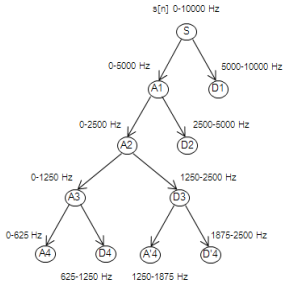
3.2 การแยกส่วนประกอบความถี่

งานวิจัยนี้ใช้เวฟเล็ทแม่แบบ Harr ในการแยกส่วนประกอบความถี่ของสัญญาณ ซึ่งเป็นส่วนของขั้นตอน Wavelet Decomposition [10] แสดงรายละเอียดดังรูปที่ 2 เริ่มตั้งแต่สัญญาณ s[n] เป็นสัญญาณที่ได้มาจากขั้นตอน Preprocessing เมื่อทำการแยกส่วนประกอบความถี่ด้วยวิธีเวฟเล็ทจะได้ส่วนประกอบสัญญาณความถี่สูง g[n] จะได้สัญญาณ D_L และส่วนประกอบสัญญาณความถี่ต่ำ h[n] จะได้สัญญาณ A_L (L คือ ระดับของเวฟเล็ท) โดยมีอัตราสุ่ม (Sampling Rate) ของสัญญาณลดลงเหลือครึ่งหนึ่งของอัตราสุ่มเดิม หลังจากนั้นจึงนำส่วนประกอบสัญญาณความถี่ต่ำไป

แยกส่วนประกอบความถี่ของสัญญาณในระดับต่อไปจนถึงระดับที่ต้องการ ซึ่งสามารถแยกส่วนประกอบความถี่ออกมาเป็นสัญญาณในแต่ละระดับความถี่ แสดงดังแผนภูมิต้นไม้ในรูปที่ 3



รูปที่ 2 แผนภูมิการกระจายเวฟเล็ทสำหรับการแยกส่วนประกอบความถี่



รูปที่ 3 แผนภูมิต้นไม้สำหรับการแยกส่วนประกอบความถี่

งานวิจัยนี้ได้วิเคราะห์ระดับสัญญาณรายละเอียดอยู่ที่ระดับ 4 (Wavelet Decomposition level 4) หลังจากนั้นสัญญาณก็จะถูกส่งต่อไปยังขั้นตอน Formant analysis เป็นการสกัดค่าความถี่ฟอร์แมนต์โดยใช้เทคนิคฟาสฟูเรียร์ทรานฟอร์ม [11] ดังสมการที่ (1) เมื่อ $Y(k)$ คือชุดแถวของสัญญาณที่ได้จากการแปลงฟูเรียร์ โดยที่ N คือ จำนวนข้อมูลในการแปลงฟูเรียร์และ k มีค่าตั้งแต่ 0 จนถึงจำนวน N-1

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

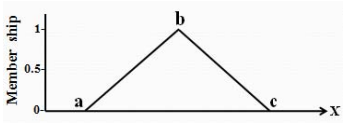
$$for\ n = 0, 1, 2, \ldots, N-1$$

กำหนดให้ค่าที่อยู่ระหว่าง $Y(k)$ ทางด้านแกน x (frequency resolution) คือ $\div F = f_s / N$ โดย f_s คือ อัตราสุ่ม (Sampling Rate) ของสัญญาณ

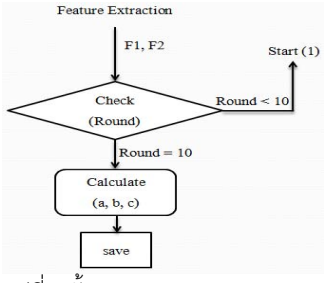
3.3 การรู้จำเสียง

สำหรับขั้นตอน Recognition เป็นกระบวนการรู้จำเสียงโดยเทคนิคฟัซซี่ลอจิก ซึ่งแบ่งออกได้เป็น 2 ขั้นตอน คือ ส่วนของการสอนระบบ (Train) และการทดสอบ (Test) โดยงานวิจัยนี้เลือกใช้ฟังก์ชันความเป็นสมาชิกแบบสามเหลี่ยม (Triangular Membership Function) ดังสมการที่ (2) สามารถทำได้โดยการกำหนดตำแหน่งจุดมุมทั้ง 3 มุม แสดงดังรูปที่ 4

$$Triangular(x, a, b, c) = \begin{cases} 0 & x < a \\ \frac{(x-a)}{(b-a)} & a < x \le b \\ \frac{(c-x)}{(c-b)} & b < x \le c \\ 0 & x > c \end{cases} \quad (2)$$

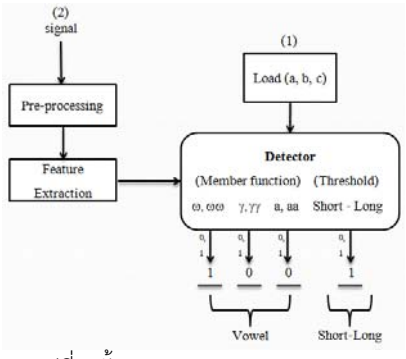


รูปที่ 4 ฟังก์ชันความเป็นสมาชิกรูปสามเหลี่ยม



รูปที่ 5 ขั้นตอนการสอนระบบ (Train)

เริ่มจากขั้นตอน Feature Extraction ซึ่งได้มาจากขั้นตอนการสกัดคุณลักษณะของเสียงส่งผลให้ได้ค่าความถี่ฟอร์แมนต์ที่ 1 และ 2 (F1, F2) ถัดไปเป็นขั้นตอน Check (Round) จะทำหน้าที่ตรวจสอบจำนวนรอบของการพูดและเก็บค่าความถี่ทั้งสอง เพื่อความเหมาะสมกับผู้ป่วย งานวิจัยนี้จึงใช้การสอนระบบไม่มากจนเกินไป คือ ทั้งหมด 10 รอบ ถ้าจำนวนรอบยังไม่ครบ โปรแกรมก็จะให้ผู้พูดทำการพูดใหม่โดยให้กลับไปสู่จุดเริ่มต้นอีกครั้ง (Start) หลังจากจำนวนรอบครบแล้ว ก็จะเข้าสู่ขั้นตอน Calculate เพื่อทำการคำณวนค่า a, b, c สำหรับนำไปสร้างเป็นฟังก์ชันความเป็นสมาชิกแบบสามเหลี่ยมในระบบฟัซซี่และสุดท้ายค่า a, b, c ก็จะถูกนำมาบันทึก (save) ลงในระบบเพื่อใช้สำหรับการทดสอบ ต่อไป



รูปที่ 6 ขั้นตอนการทดสอบระบบ (Test)

เริ่มจากขั้นตอนแรก โปรแกรมจะทำการโหลดแฟ้มข้อมูลที่ได้จากการสอนระบบ หลังจากนั้นโปรแกรมก็จะรอสัญญาณเสียงพูด (signal) เพื่อที่จะทำการสกัดค่าความถี่ฟอร์แมนต์ ถัดไปสัญญาณก็ถูกส่งต่อไปยังตัวตรวจจับคุณลักษณะเสียง (Detector) โดยการแยกเสียงสระนั้นจะตรวจเช็คความถี่ฟอร์แมนต์จากความเป็นสมาชิกของกระบวนการตัดสินใจในฟัซซี่ลอจิก ส่วนการตรวจจับสระเสียงสั้นเสียงยาวนั้น จะใช้กระบวนการหาค่าเฉลี่ยของฟังก์ชันพลังงานเสียง (Short – Long analysis) ผลสรุปออกมาจะเป็นรหัส (0, 1) ดังตารางที่ 1

ตารางที่ 1 ผลสรุปของเสียงพูดจำแนกโดยรหัส

สระ	อึ	อือ	เออะ	เออ	อะ	อา
รหัส	0010	0011	0100	0101	1000	1001

4. วิธีการทดลอง

งานวิจัยนี้ได้นำเสนอการออกแบบระบบสั่งการโดยใช้เสียงพูดสำหรับผู้พูดผิดปกติชนิด Dysarthria แบ่งออกเป็น 2 การทดลอง คือ การทดลองแรกใช้เวฟเล็ตรวมกับฟาสฟูเรียร์ทรานฟอร์มในการคัดแยกความถี่ฟอร์แมนต์ และการทดลองที่สองคือกระบวนการสอนและทดสอบระบบโดยใช้เทคนิคฟัซซี่ลอจิก ร่วมกับเสียงสระในภาษาไทยจำนวน 6 คำ ประกอบด้วยสระเสียงสั้นจำนวน 3 คำ(อึ เออะ อะ) และสระเสียงยาวจำนวน 3 คำ (อือ เออ อา) รวมทั้งสิ้น 6 คำสั่ง ร่วมกับกลุ่มผู้ทดสอบเป็นเพศชาย จำนวน 3 คน คือ

1.ผู้พูดผิดปกติชนิด Dysarthria อายุ 57 ปี อัมพาตครึ่งซีก (Hemiplegia) ระดับ 6 (C6) กำหนดเป็น Dysarthric Speech 1

2.ผู้พูดผิดปกติชนิด Dysarthria อายุ 70 ปี อัมพาตครึ่งซีก (Hemiplegia) ระดับ 4 (C4) กำหนดเป็น Dysarthric Speech 2

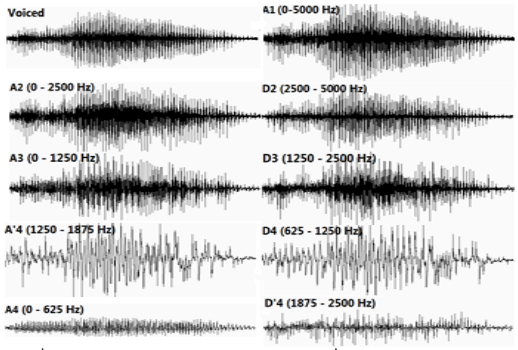
3.ผู้พูดปกติ อายุ 25 ปี กำหนดเป็น Control Speech

ทำการเก็บสัญญาณเสียงโดยใช้โปรแกรม LABVIEW version 8.6 ร่วมกับไมโครโฟนยี่ห้อ Genius (Desktop Microphone) ซึ่งมีย่านความถี่ตอบสนองในช่วง 0 – 10 kHz ความถี่ในการสุ่มตัวอย่างเท่ากับ 20 kHz เก็บบันทึกในไฟล์ข้อมูลรูปแบบ .lvm โดยให้กลุ่มผู้ทดสอบเปล่งเสียงสระ ทั้งหมด 5 ครั้งในแต่ละคำ และทำการวัดค่าความถี่ ประกอบด้วย ความถี่ฟอร์แมนต์ที่ 1 (F1) และความถี่ฟอร์แมนต์ที่ 2 (F2) ตามลำดับ

5. ผลการทดลอง

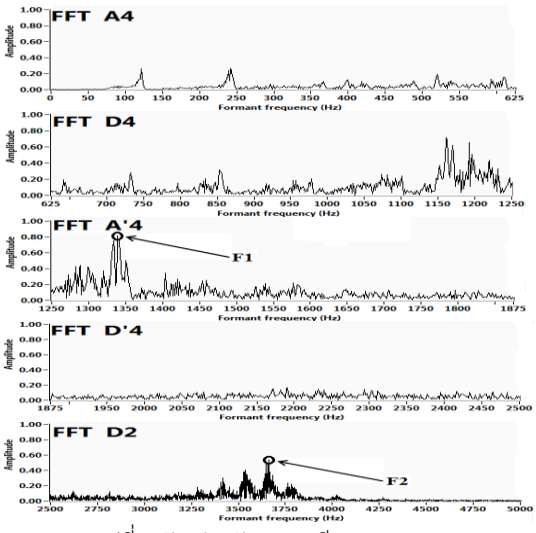
5.1 การคัดแยกความถี่ฟอร์แมนต์

ขั้นตอนแรกเป็นการนำเทคนิคเวฟเล็ทมาใช้ในการแยกส่วนประกอบความถี่ แสดงตัวอย่างดังรูปที่ 7 ผลจากการใช้ Wavelet Decomposition level 4 ในการแยกส่วนประกอบความถี่ในโดเมนเวลา ออกเป็น 10 ช่วง



รูปที่ 7 ตัวอย่างสัญญาณรายละเอียดที่ได้จากการแปลงเวฟเล็ทของเสียงสระ “อา” (Dysarthric Speech 1)

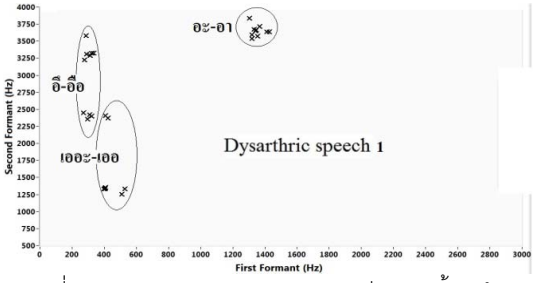
ขั้นตอนถัดไปสัญญาณเสียงในแต่ละช่วงก็จะถูกนำมาแยกความถี่ฟอร์แมนต์โดยใช้เทคนิค FFT ดังรูปที่ 8 (แสดงเฉพาะช่วง D2, D4, D’4, A4 และ A’4)



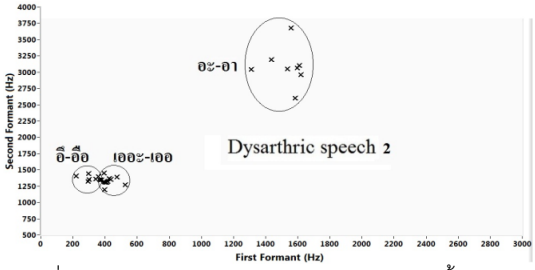
รูปที่ 8 ตัวอย่างสัญญาณเสียงสระ “อา” (Dysarthric Speech 1) ในโดเมนความถี่แต่ละช่วง

ตารางที่ 2 สรุปค่าความถี่ฟอร์แมนต์เฉลี่ยของกลุ่มผู้ทดสอบทั้ง 3 คน โดยใช้เทคนิค DWT + FFT

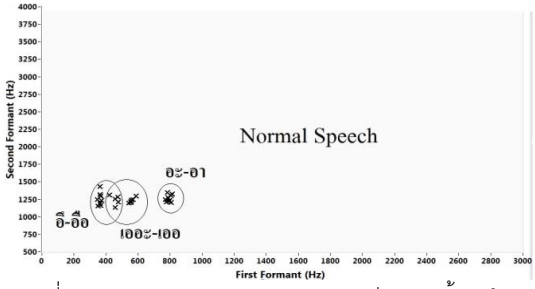
ประเภท / ความถี่ สระ	Dysarthric 1		Dysarthric 2		Normal	
	F1 (Hz)	F2 (Hz)	F1 (Hz)	F2 (Hz)	F1 (Hz)	F2 (Hz)
อึ-อือ	301.8	2967.3	316.2	1555.3	372.8	1257.2
เออะ-เออ	519.3	1651.2	426.2	1314.9	516.9	1231.6
อะ-อา	1351.9	3651.3	1545.8	3085.5	791.1	1261.7



รูปที่ 9 กราฟแจกแจงแบบกระจายของเสียงสระทั้ง 6 คำ (Dysarthric Speech 1) โดยใช้เทคนิค DWT + FFT

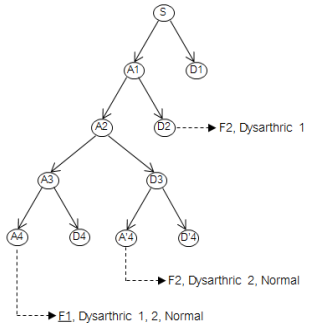


รูปที่ 10 กราฟแจกแจงแบบกระจายของเสียงสระทั้ง 6 คำ (Dysarthric Speech 2) โดยใช้เทคนิค DWT + FFT

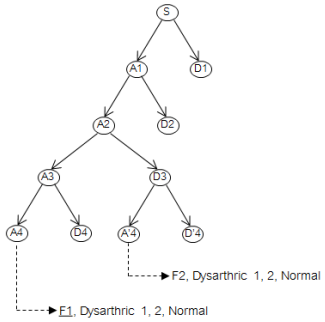


รูปที่ 11 กราฟแจกแจงแบบกระจายของเสียงสระทั้ง 6 คำ (Normal Speech) โดยใช้เทคนิค DWT + FFT

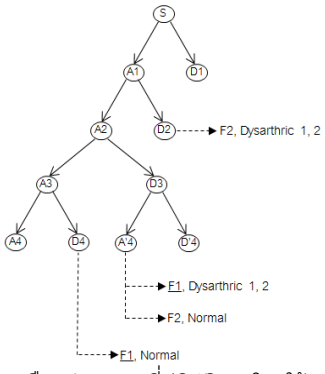
เมื่อพิจารณาจากค่าความถี่ฟอร์แมนต์ทั้งสอง ดังตารางที่ 2 สามารถสรุปย่านความถี่ปฏิบัติงานในแต่ละคำโดยใช้แผนภูมิต้นไม้แสดงดังรูปที่ 4 ถึง 6 (กำหนดให้ F1, F2 = ความถี่ฟอร์แมนต์ที่ 1, 2 Dysarthric 1, 2 คือ ผู้พูดชนิด Dysarthric คนที่ 1, 2 และ Normal คือ ผู้พูดปกติ ตามลำดับ)



รูปที่ 12 การเลือกช่วงความถี่ปฏิบัติงานโดยใช้แผนภูมิการกระจายเวฟเล็ท (A4, A'4, D2) ระหว่างผู้ทดสอบทั้ง 3 (เสียงสระ "อึ-อือ")



รูปที่ 13 การเลือกช่วงความถี่ปฏิบัติงานโดยใช้แผนภูมิการกระจายเวฟเล็ท (A4, A'4) ระหว่างผู้ทดสอบทั้ง 3 (เสียงสระ "เออะ-เออ")

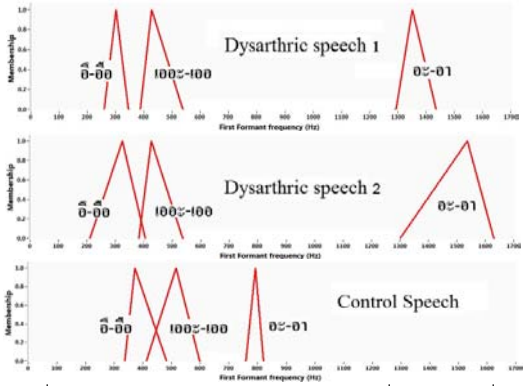


รูปที่ 14 การเลือกช่วงความถี่ปฏิบัติงานโดยใช้แผนภูมิการกระจายเวฟเล็ท (D4, A'4, D2) ระหว่างผู้ทดสอบทั้ง 3 (เสียงสระ "อะ-อา")

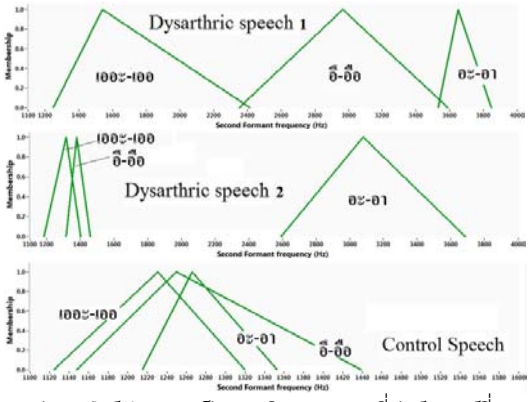
ยกตัวอย่างจากรูปที่ 14 ในการวิเคราะห์เสียงสระ "อะ-อา" กรณีผู้ป่วยทั้ง สองจะเห็นว่าความถี่ฟอร์แมนต์ที่ 1 อยู่ในช่วง 1250 Hz ขึ้นไป และความถี่ฟอร์แมนต์ที่ 2 อยู่ในช่วง 2500 Hz ขึ้นไป ดังนั้นในการเลือกความถี่ปฏิบัติงานจึงเลือกเฉพาะช่วง A'4 และ D2 ซึ่งแตกต่างกับผู้พูดปกติ จะเลือกเฉพาะช่วง D4 และ A'4 ตามลำดับ

5.2 การทดสอบระบบ

หลังจากที่สามารถสกัดคุณลักษณะของเสียงทั้งหมดได้แล้ว ขั้นตอนถัดไปก็เป็นการออกแบบให้ระบบสามารถรู้จำและตัดสินใจ เนื่องจากข้อมูลที่ได้หลังจากการสกัดความถี่ฟอร์แมนต์ทั้งสองจากผู้ป่วยสามารถแยกกันได้ค่อนข้างชัดเจน (รูปที่ 9 และ 10) ดังนั้นอาศัยเทคนิคฟัซซีลอจิกเพียงอย่างเดียวก็สามารถแยกคำพูดในระบบสั่งการได้จากกระบวนการทั้งหมด สามารถสรุปผลการเรียนรู้ของระบบได้ดังรูปที่ 15 และ 16 ส่วนผลการทดสอบระบบแสดงดังตารางที่ 3



รูปที่ 15 ฟังก์ชันความเป็นสมาชิกของความถี่ฟอร์แมนต์ที่ 1



รูป 16 ฟังก์ชันความเป็นสมาชิกของความถี่ฟอร์แมนต์ที่ 2

ตารางที่ 3 ผลการทดสอบโปรแกรมจำแนกตามรายบุคคล

ประเภท / สระ	Dysarthric Speech 1	Dysarthric Speech 2	Control Speech
	ความแม่นยำ (%)	ความแม่นยำ (%)	ความแม่นยำ (%)
อี-อือ	100.0	76.9	75.0
เออะ-เออ	91.6	90.9	84.2
อะ-อา	100.0	93.3	100.0
เฉลี่ย	**97.2**	**87.0**	**86.4**

6. สรุปและวิจารณ์

ระดับความผิดปกติของผู้ป่วยจะส่งผลกระทบโดยตรงต่อความถี่ฟอร์แมนต์ นั่นคือเมื่อระดับความรุนแรงของโรคหลอดเลือดสมองสูงขึ้น ความถี่ฟอร์แมนต์ก็จะผิดเพี้ยนไปจากผู้พูดปกติมากขึ้นเช่นกัน ส่งผลดีต่อการคัดแยกเสียง เมื่อสังเกตจากฟังก์ชันความเป็นสมาชิกของความถี่ฟอร์แมนต์ทั้งสอง จะเห็นว่าย่านความถี่ในแต่ละกลุ่มคำของผู้ป่วยทั้งสองจะแยกกันได้อย่างชัดเจน อาจมีการเหลื่อมล้ำกันบางในบางกลุ่มคำแต่ยังน้อยกว่าผู้พูดปกติซึ่งมีการเหลื่อมกันของกลุ่มสระ อี-อือ และ เออะ-เออ ในความถี่ฟอร์แมนต์ที่ 1 ที่สูงกว่า ส่วนความถี่ฟอร์แมนต์ที่ 2 ทั้งสามกลุ่มคำของผู้พูดปกติจะซ้อนทับกันหมด จึงทำให้เทคนิคฟัซซีลอจิกไม่สามารถแยกกลุ่มคำของผู้พูดชนิดนี้อย่างมีประสิทธิภาพเมื่อเทียบกับผู้ป่วยทั้งสอง ซึ่งมีความแม่นยำอยู่ที่ 97.2% (Dysarthric Speech 1), 87.0% (Dysarthric Speech 2) และ 86.4% (Normal Speech) ตามลำดับ

7. เอกสารอ้างอิง

[1] P. Kimsawad, “Study of Speaker Independent Isolated Word Recognition of Thai Digits Using Backpropagation Neural Networks,” Department of Electrical Engineering, Prince of Songkla University, 2000.

[2] A. Thammaraksasit, “Control Command Classification for Speech Recognition Based Wheelchair,” Department of Electronics and Telecommunication Engineering, King Mongkut's University of Technology Thonburi, 2003.

[3] S. Sukanake, “Continuous Thai Tone Recognition using Tri-Half-Tone Hidden Markov Models,” Department of Computer Engineering, Kasetsart University, 2004.

[4] A. Deemagarn, “Speaker Independent Thai Connected Digit Speech Recognition System Using Hidden Markov Model,” Department of Computer Engineering, Kasetsart University, 2007.

[5] P. Kayasith, T.Theeramunkong, N. Thubthong, "Incorporated Speech Overlapped Factor (λ) into Speech Clarity Index (..): Method to Improve Dysarthric Speech Severity Evaluation," International Convention for Rehabilitation Engineering and Assistive Technology (i-CREATe 2007), Singapore, pp.133-138, April, 2007.

[6] "การฝึกพูดสำหรับผู้ป่วยโรคหลอดเลือดสมองที่มีปัญหาทางด้านการพูด," ภาควิชาโสต ศอ นาสิกวิทยา คณะแพทยศาสตร์ โรงพยาบาลรามาธิบดี, 1/11/51, [online]. <http://www.vichaiyut.co.th>

[7] C. Rowden, "Speech Processing," Department of Electronic Systems Engineering University of Essex, McGRAW- HILL Book Company Europe.

[8] R. Vergin, D. O'Shaughnessy, "Pre-Emphasis and Speech Recognition," Canadian Conference on Electrical and Computer Engineering, vol 2, pp.1062-1065, September, 1995.

[9] E. Panyathep, "The Isolated-Speech Thai-Vowel Recognition System Using Neuro-Fuzzy Networks," Department of Information of Technology, King Mongkut's Institute of Technology North Bangkok, 2005.

[10] N. Bunsakchalerm, "Application of Wavelet Transforms of Consonant Vowel Segmentation on Thai Speech Signal," Department of Telecommunication Engineering, 2006.

[11] T. Orzechowski, A. Izworski, R. Tadeusiewicz, K.. Chmurzynska, P. Radkowski, I. Gatkowska, "Processing of Pathological Changes in Speech Caused by Dysarthria," Proceedings of 2005 International Symposium on Intelligent Signal Processing and Communication Systems, December, 2005.

ประวัติผู้เขียน

ชื่อ สกุล นายโอฬาร ดาวเวียง

รหัสประจำตัวนักศึกษา 5010120122

วุฒิการศึกษา

วุฒิ	ชื่อสถาบัน	ปีที่สำเร็จการศึกษา
วิศวกรรมศาสตรบัณฑิต (วิศวกรรมไฟฟ้า)	มหาวิทยาลัยสงขลานครินทร์	2550

ทุนการศึกษา (ที่ได้รับในระหว่างการศึกษา)

ทุนก้นกุฏิ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยสงขลานครินทร์