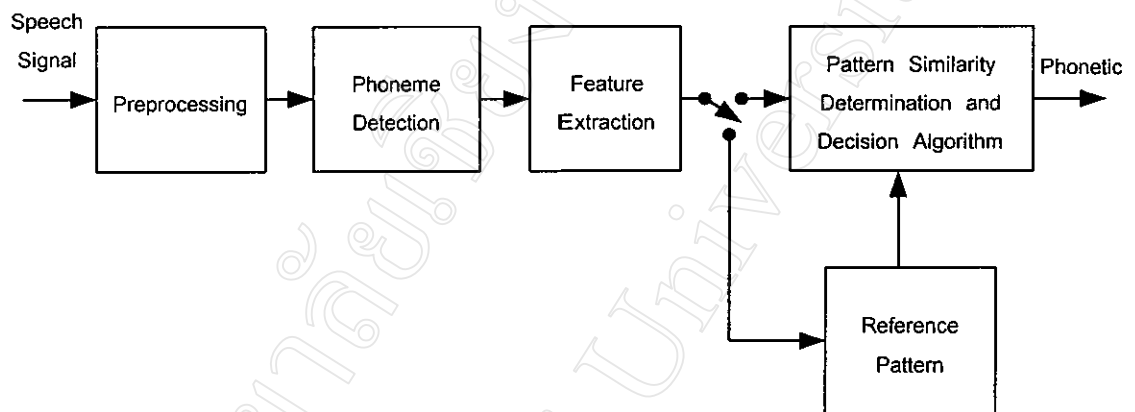


บทที่ 4

การทำงานของระบบรู้จำ

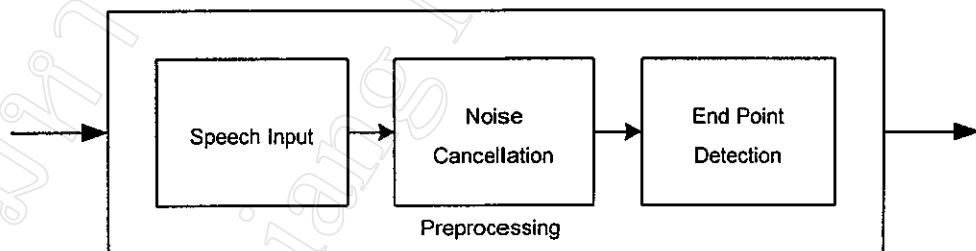
การรู้จำหน่วยเสียงในงานวิจัยนี้ประกอบไปด้วยขั้นตอนดำเนินการหลัก 4 ขั้นตอน

ผังรูปที่ 4.1



รูปที่ 4.1 ขั้นตอนหลักของการรู้จำหน่วยเสียง

4.1 การประมวลผลเบื้องต้น (Preprocessing)



รูปที่ 4.2 ขั้นตอนย่อยของการประมวลผลเบื้องต้น

4.1.1 ส่วนรับสัญญาณเสียง (Speech Input)

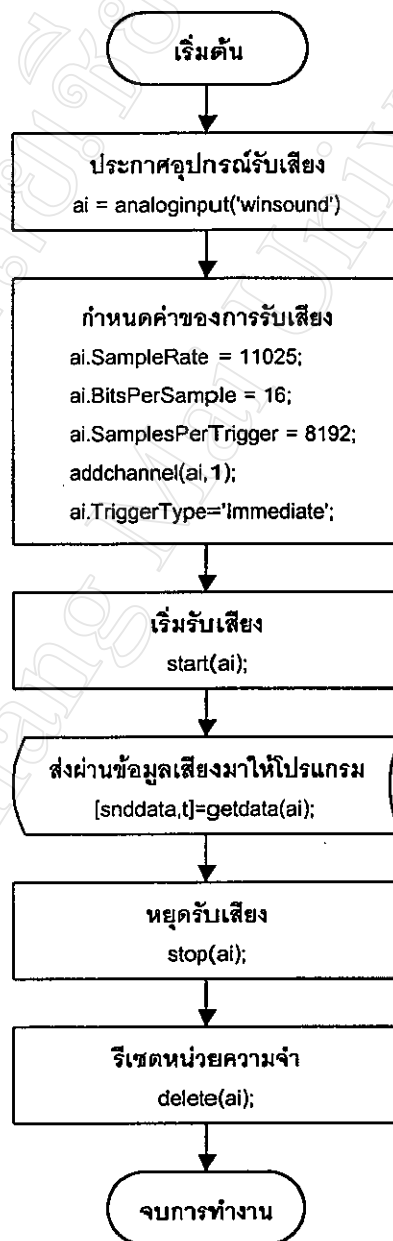
เป็นส่วนของโปรแกรมที่ทำหน้าที่ติดต่อกับการ์ดเสียง เพื่อที่จะกำหนดค่าต่างๆ ของการรับสัญญาณเสียงเพื่อให้ได้สัญญาณเสียงที่เป็นไปตามต้องการ ดังรูปที่ 4.3 ค่าที่ต้องกำหนดมีดังนี้

- อัตราการสุ่มสัญญาณ (Sampling Rate) กำหนดเป็น 11025 Hz
- ความละเอียดของข้อมูล (Bits per Sample) กำหนดเป็น 16 บิตต่อหนึ่งตัวอย่าง
- จำนวนข้อมูลต่อครั้งที่มีการบันทึก (Sample per Trigger) กำหนดเป็น 8192 ข้อมูล
- จำนวนช่องทาง (Channel) กำหนดเป็น 1 (Mono)

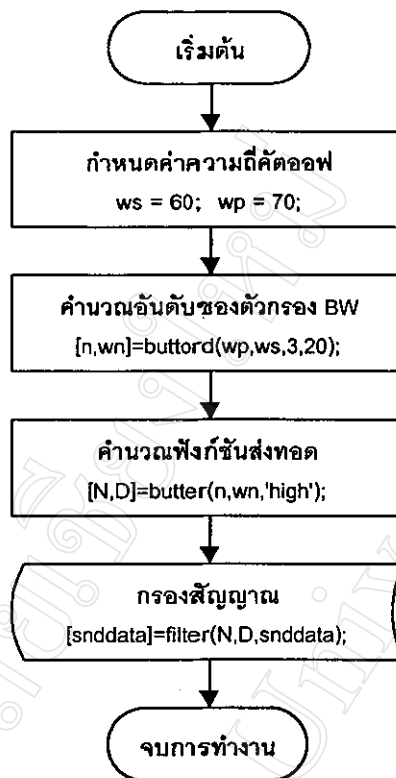
เมื่อโปรแกรมในส่วนนี้ทำงานก็จะส่งค่าต่างๆ เหล่านี้ไปให้ระบบปฏิบัติการ และเมื่อการ์ดเสียงทำการเก็บสัญญาณเรียบร้อยแล้ว โปรแกรมก็จะทำการขอข้อมูลมาใช้ในโปรแกรมต่อไป

4.1.2 ส่วนกำจัดนอยส์ (Noise Cancellation)

เป็นส่วนของการกำจัดนอยส์ที่มีความถี่ตั้งแต่ 0 ถึง 50 Hz ซึ่งเป็นสัญญาณที่มาจากแหล่งจ่ายไฟ และเสียงลมขณะที่มีการพูด เนื่องจากงานวิจัยนี้มีการสกัดลักษณะเด่นที่เป็นจำนวนองค์ประกอบความถี่ ดังนั้นความถี่ต่ำจะมีผลกระทบต่อความถูกต้องของระบบมาก ดังรูปที่ 4.4

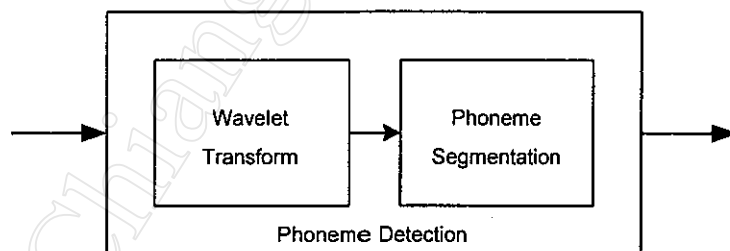


รูปที่ 4.3 ฟังงานในส่วนของการรับสัญญาณเสียง



รูปที่ 4.4 ผังงานในส่วนของการกำจัดนอยส์

4.2 การตรวจหาหน่วยเสียง (Phoneme Detection)



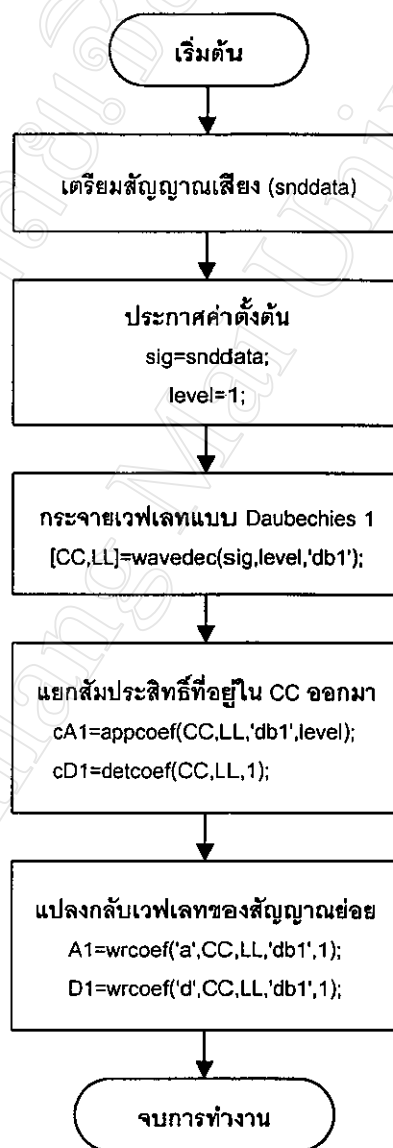
รูปที่ 4.5 ขั้นตอนย่อยของการตรวจหาหน่วยเสียง

4.2.1 การแปลงเวฟเลต (Wavelet Transform)

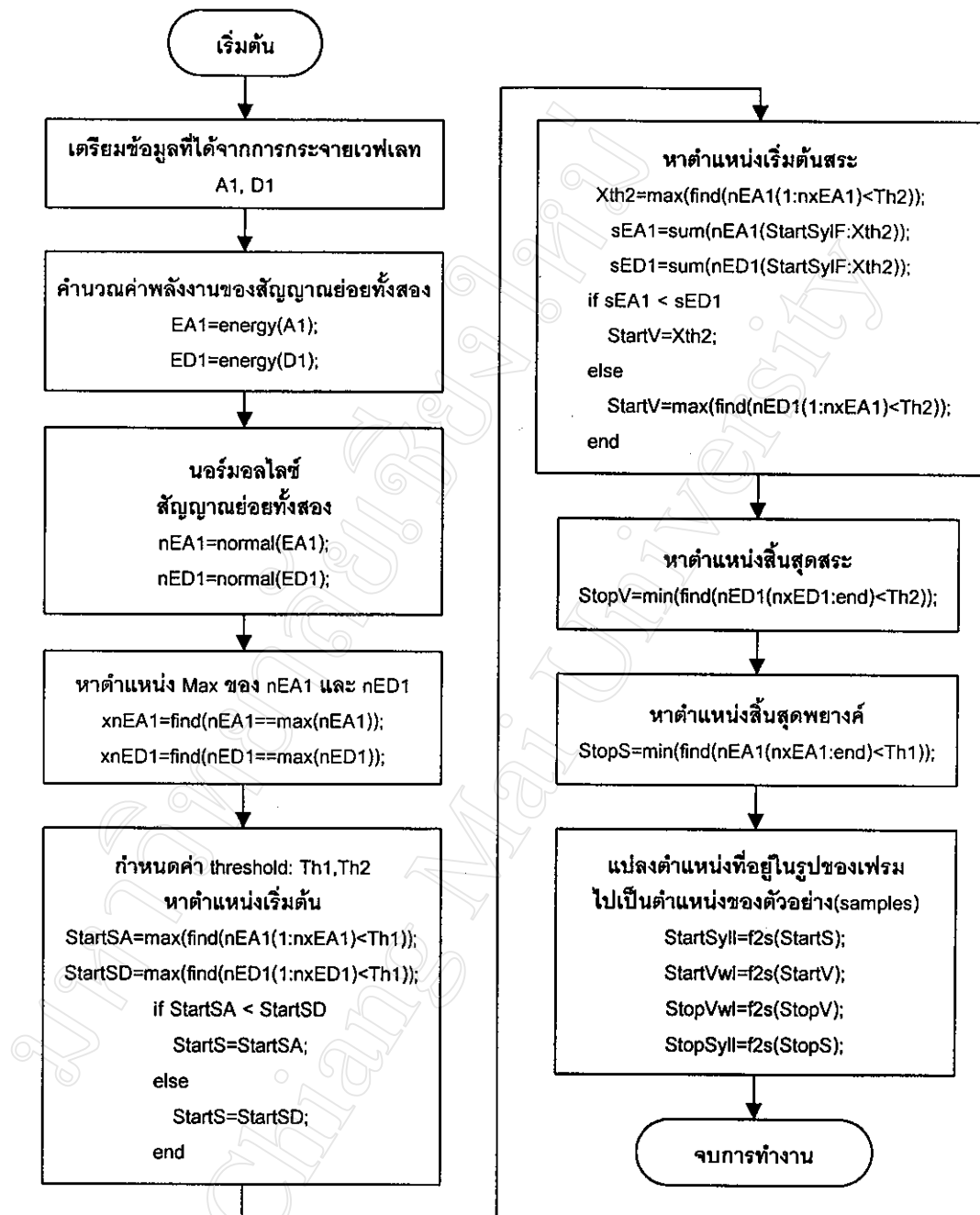
เป็นการกระจายสัญญาณออกเป็นสัญญาณย่อยๆ เพื่อจะได้สะดวกในการพิจารณา ลักษณะความถี่ที่มีการเปลี่ยนแปลงไปในแต่ละสัญญาณย่อย ซึ่งทำให้การวิเคราะห์ตำแหน่งที่เป็น รอยต่อระหว่างหน่วยเสียงเป็นไปได้ง่ายขึ้น และถูกต้องมากกว่าการวิเคราะห์สัญญาณโดยไม่มีการ กระจายสัญญาณ ดังรูปที่ 4.6

4.2.2 การแยกหน่วยเสียง (Phoneme Segmentation)

เป็นการนำสัญญาณที่ถูกกระจายออกมาจากการแปลงเวฟเลท มาวิเคราะห์ตำแหน่งของหน่วยเสียงสระ โดยการสัญญาณย่อยทั้งหมดมาหาพลังงาน แล้วหาตำแหน่งที่เริ่มมีการเปลี่ยนแปลงของพลังงานก่อนที่จะมีการเปลี่ยนแปลงอย่างรวดเร็ว ตำแหน่งที่ได้นี้จะเป็นตำแหน่งเริ่มต้นของหน่วยเสียงสระ ส่วนตำแหน่งสุดท้ายที่เริ่มมีการเปลี่ยนแปลงของพลังงานหลังจากที่มีการเปลี่ยนแปลงอย่างฉับพลันมาแล้ว เป็นตำแหน่งของการสิ้นสุดหน่วยเสียงสระดังรูปที่ 4.7



รูปที่ 4.6 ฟังก์ชันของการแปลงเวฟเลท



รูปที่ 4.7 ฟังก์ชันในส่วนของการแยกหน่วยเสียง

4.3 การสกัดลักษณะเด่น (Feature Extraction)

ก่อนที่จะมีการหาความคล้ายคลึงกันของสัญญาณนั้น เราจะต้องมีทั้งลักษณะเด่นของสัญญาณอ้างอิงและลักษณะเด่นของสัญญาณทดสอบเสียก่อน โดยที่ลักษณะเด่นทั้งสองนี้จะต้องมาจากขั้นตอนของการสกัดลักษณะเด่นของหน่วยเสียง ซึ่งมีความแตกต่างกันไปตามชนิดของหน่วยเสียง และเพื่อให้ง่ายต่อการทำความเข้าใจในผังการทำงาน จึงขอรวมผังการทำงานของการสกัดลักษณะเด่นของหน่วยเสียงไว้ในขั้นตอนการหาความคล้ายคลึงกันของรูปแบบ

4.3.1 สระ (Vowel)

ลักษณะเด่นของสระ คือ เป็นสัญญาณที่มีความเป็นรายคาบสูง มีลักษณะของสัญญาณในเชิงเวลาที่ค่อนข้างคล้ายกันมากตลอดระยะเวลาของการออกเสียง เนื่องจากเป็นเสียงที่ออกมาจากเส้นเสียง และแต่ละเสียงก็จะมีลักษณะที่แตกต่างกันไป สำหรับงานวิจัยนี้เราจะใช้ลักษณะเด่นของหน่วยเสียงสระ 3 ลักษณะ ดังนี้

- สัมประสิทธิ์ LPC (Linear Predictive Coding coefficients)

ใช้ขนาดสัญญาณเสียงจำนวน 330 ตัวอย่าง (30 ms) ในการสกัดลักษณะเด่นด้วยการหาสัมประสิทธิ์ LPC อันดับที่ 20 ผลลัพธ์ที่ได้จากขั้นตอนนี้เป็นค่าสัมประสิทธิ์ 21 ค่า แต่เนื่องจากค่าแรกเป็น 1 เหมือนกันทุกกรณีจึงไม่นำมาใช้

- องค์ประกอบความถี่ (Frequency Component)

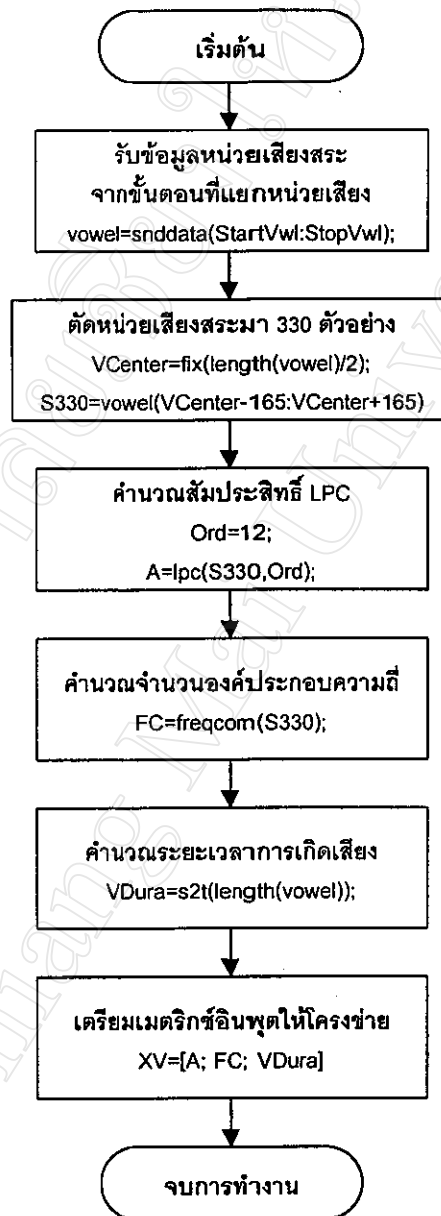
ในการออกเสียงสระนั้นจะเกิดจากการสั่นของเส้นเสียงเหมือนกันทั้งหมด แต่จะแตกต่างกันที่ลักษณะของช่องปาก ทำให้จำนวนองค์ประกอบของความถี่แตกต่างกันออกไปเป็น 3 กลุ่ม ซึ่งจำนวนองค์ประกอบในแต่ละกลุ่มคือ 2-4, 4-6 และ 6-10 ตามลำดับ โดยการใช้ FFT

- ระยะเวลาการเกิดเสียง (Time Duration)

สระในภาษาไทยมีทั้งเสียงสั้นและเสียงยาว ตัวอย่างเช่น อะ-อา เป็นหน่วยเสียงที่มีลักษณะเด่นเหมือนกัน ยกเว้นระยะเวลาการเกิดเสียงที่ต่างกัน สระเสียงสั้นจะมีระยะเวลาของการเกิดเสียงประมาณ 0.09 – 0.16 วินาที และ 0.16 – 0.27 วินาที สำหรับสระเสียงยาว ซึ่งอาจมากกว่า 0.27 วินาทีในกรณีที่มีการลากเสียง หรือในกรณีที่ไม่มีเสียงตัวสะกด (แม่ กา)

จากข้อมูลลักษณะเด่นทั้ง 3 ขั้นตอน จะถูกนำมาเรียงต่อกันเป็นแถวอยู่ในเมตริกซ์อินพุตของโครงข่ายประสาทเทียม ดังรูปที่ 4.8 ส่วนสคณธ์ของเมตริกซ์นั้นจะมีค่าเท่ากับจำนวนชุดข้อมูลที่ใช้ในการสอน เช่นการสอนโครงข่ายประสาทเทียมให้รู้จำเสียงสระ 18 เสียง ซึ่งแต่ละเสียงจะมีข้อมูลในการสอนเสียงละ 5 ข้อมูล ดังนั้นเมตริกซ์อินพุตที่ได้จะมีจำนวนแถว $20+1+1=22$ และ

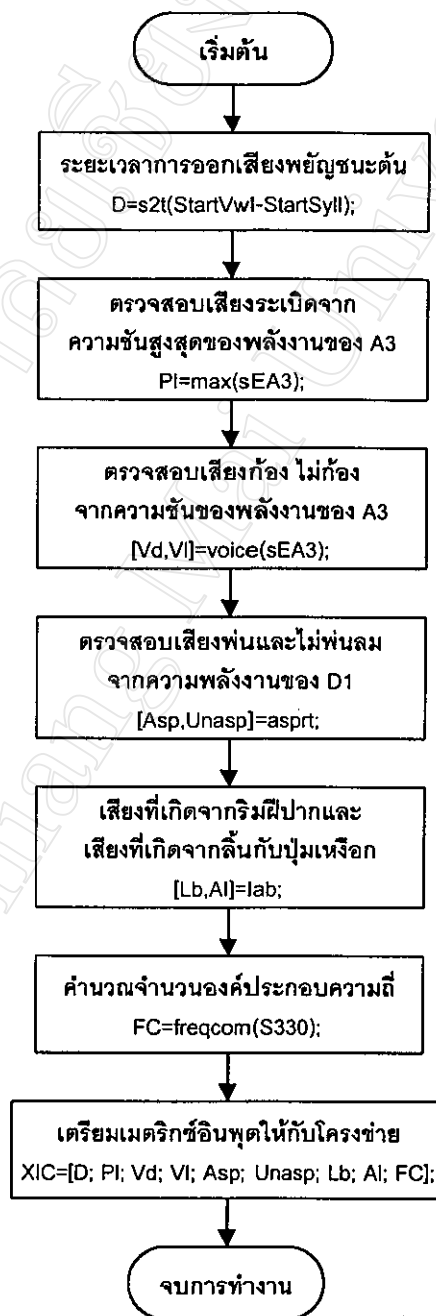
จำนวนสครมภ์ $18 \times 5 = 90$ เขียนในรูปตัวแปรเมตริกซ์เป็น $X_{14,90}$ ในทำนองเดียวกันกับเมตริกซ์เฮาต์พูดเป้าหมายก็จะมีจำนวนสครมภ์เท่ากัน ส่วนแถวก็จะเท่ากับจำนวนสระ คือ 18



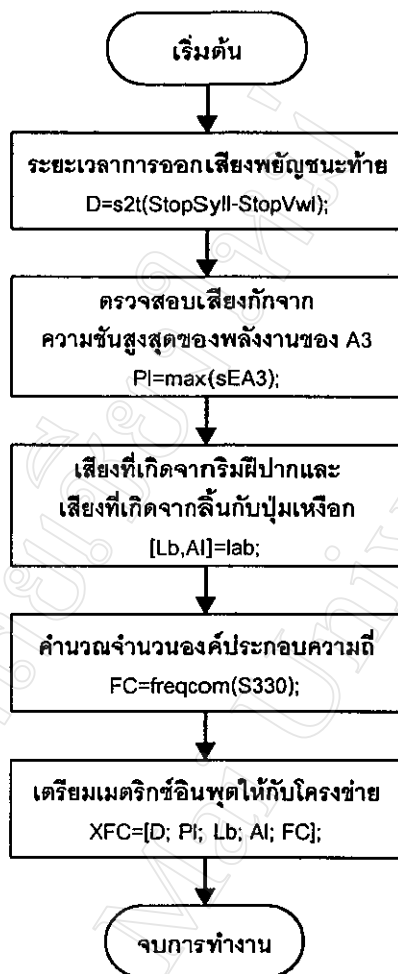
รูปที่ 4.8 ฟังก์ชันของการสกัดลักษณะเด่นของหน่วยเสียงสระ

4.3.2 พยัญชนะต้น (Initial Consonant)

พยัญชนะต้นแบ่งออกเป็นพยัญชนะเสียงสั้น กับเสียงยาว ซึ่งมีตำแหน่งของลักษณะเด่นคนละตำแหน่ง และสระเสียงสั้นบางเสียงมีระยะเวลาสั้นมากจนหาลักษณะเด่นไม่ได้ แต่จะปรากฏลักษณะเด่นบริเวณส่วนต้นของหน่วยเสียงสระ ส่วนพยัญชนะเสียงยาวจะปรากฏลักษณะเด่นทั้งในหน่วยเสียงเอง และส่วนต้นของหน่วยเสียงสระ ผังงานของการสกัดลักษณะเด่นแสดงในรูปที่ 4.9 พยัญชนะต้นมีลักษณะเด่นดังนี้



รูปที่ 4.9 ผังงานของการสกัดลักษณะเด่นของพยัญชนะต้น



รูปที่ 4.10 ฟังก์ชันของการสกัดลักษณะเด่นของพยัญชนะท้าย

- ระยะเวลาของการเกิดเสียง (Time Duration) เป็นระยะตั้งแต่จุดเริ่มต้นพยางค์จนถึงจุดเริ่มต้นสระ
- เสียงระเบิด (Plosive) เป็นการหาค่าพลังงานในช่วงต้นของการเกิดเสียง โดยจะวัดจากความชันสูงสุดของพลังงานในช่วงนั้น เนื่องจากเสียงระเบิดนั้นเกิดมาจากการเปลี่ยนแปลงพลังงานอย่างฉับพลัน
- เสียงก้อง และไม่ก้อง (Voiced and Voiceless) คือเสียงที่มีความเป็นรายคาบ และเสียงที่ไม่มีความเป็นรายคาบ หรือไม่มีเสียง ตามลำดับ หาได้จากพลังงานของสัญญาณที่ผ่านการกระจายเวฟเลทในขั้นตอนการตรวจหาหน่วยเสียง

- เสียงพ่นลม และไม่พ่นลม (Aspirated and Unaspirated) เสียงพ่นลมเป็นสัญญาณที่เกิดจากการพ่นลมขณะออกเสียงทำให้เกิดสัญญาณความถี่สูงที่มีพลังงานสูง วัดค่าด้วยการหาค่าพลังงานในย่านความถี่สูง ส่วน เสียงไม่พ่นลม เกิดจากการกักลมไว้แล้วขณะที่พูดก็ไม่ได้พ่นลมนั้นออกมา วัดค่าด้วยหลักการเดียวกับเสียงพ่นลม
- เสียงที่เกิดจากริมฝีปากและเสียงที่เกิดจากลิ้นกับปุ่มเหงือก (Labial and Alveolar) เป็นลักษณะของความถี่ที่เปลี่ยนแปลงอยู่ในส่วนต้นของหน่วยเสียงสระ เกิดจากลักษณะของช่องปากที่เปลี่ยนไปขณะออกเสียง วัดค่าด้วยการหาทิศทางของการเปลี่ยนแปลงความถี่ในช่วงดังกล่าว
- องค์ประกอบความถี่ (Frequency Component) เป็นการนับจำนวนองค์ประกอบความถี่ย่อยที่มีพลังงานสูง

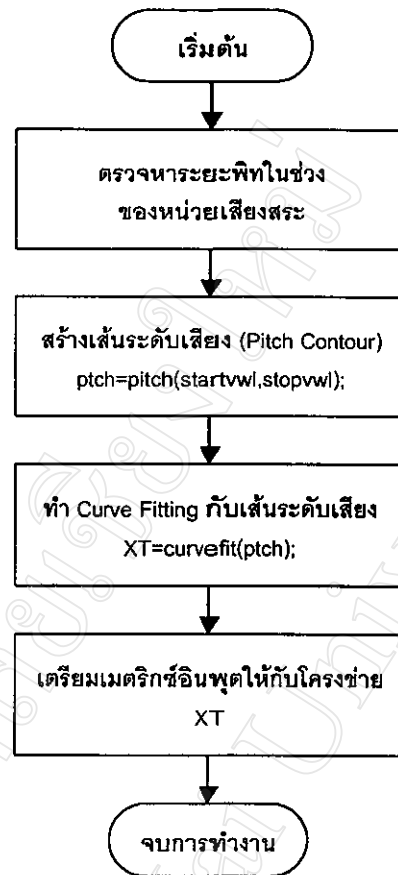
4.3.3 พยัญชนะท้าย (Final Consonant)

ในทำนองเดียวกับพยัญชนะต้น คือสามารถแบ่งพยัญชนะท้ายออกเป็นเสียงสั้นกับเสียงยาว เสียงสั้นจะปรากฏลักษณะเด่นในส่วนท้ายของหน่วยเสียงสระ ส่วนเสียงยาวจะปรากฏลักษณะเด่นทั้งในหน่วยเสียงเองและในส่วนท้ายของหน่วยเสียงสระ ดังรูปที่ 4.10 มีลักษณะเด่นดังนี้

- ระยะเวลาของการเกิดเสียง (Time Duration) เป็นระยะตั้งแต่จุดสิ้นสุดสระจนถึงจุดสิ้นสุดพยางค์
- เสียงกักลม (Stop) เกิดจากการหยุดออกเสียงโดยจับพลิ้น ซึ่งอวัยวะที่ใช้ในการกักเสียงก็จะเป็นริมฝีปาก, ลิ้น และ โคนลิ้น ทำให้มีการเปลี่ยนแปลงของพลังงานเสียงแตกต่างกันไป
- เสียงที่เกิดจากริมฝีปากและเสียงที่เกิดจากลิ้นกับปุ่มเหงือก
- องค์ประกอบความถี่

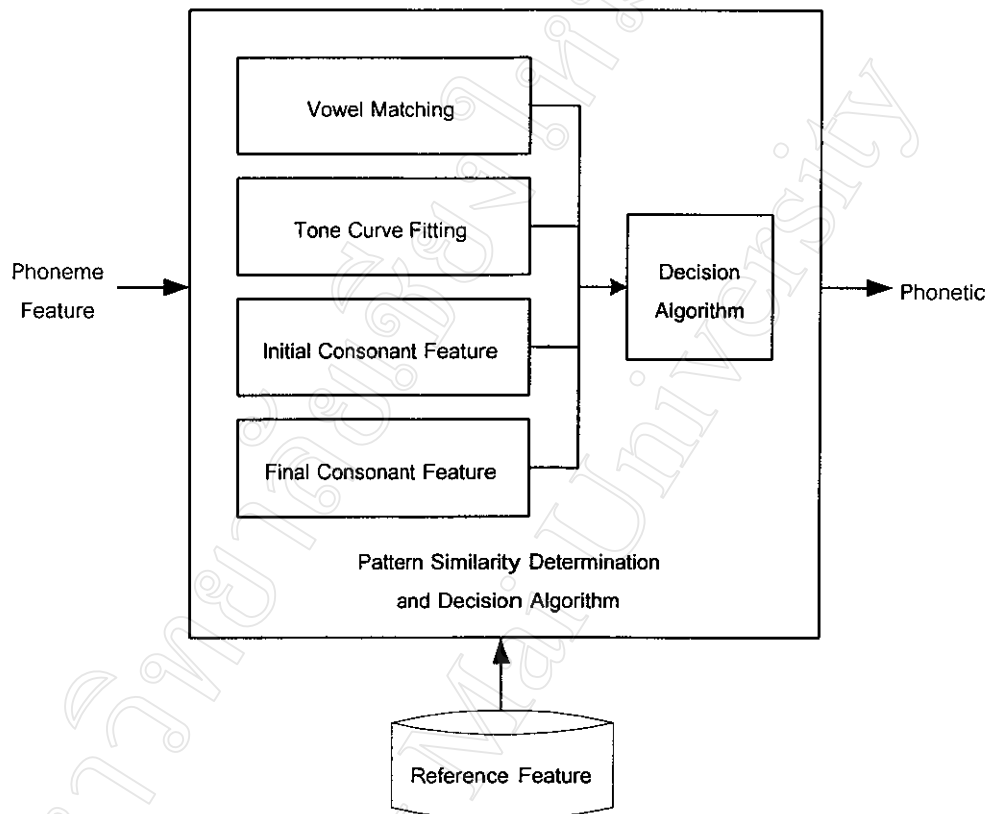
4.3.4 วรรณยุกต์ (Tone)

ลักษณะเด่นของวรรณยุกต์ คือ การเปลี่ยนแปลงของความถี่มูลฐานไปตามเวลาของการออกเสียงสระ โดยที่เสียงวรรณยุกต์สามัญกับวรรณยุกต์เอกนั้นมีรูปแบบที่คล้ายกัน แต่จะแตกต่างกันที่วรรณยุกต์เอกมีความชันของรูปแบบอ้างอิงมากกว่า การหาความเปลี่ยนแปลงของความถี่มูลฐานก็ทำได้จากความถี่ของค่ายอด (Peak) หรือระยะพิท (Pitch Range) ของสัญญาณในเชิงเวลาของช่วงที่เป็นหน่วยเสียงสระ จากนั้นนำมาประมาณส่วนโค้ง (Curve Fitting) อันดับ 2 ใช้ข้อมูล 20 จุดบนส่วนโค้งที่ได้มาสร้างเป็นเมตริกซ์อินพุตของโครงข่ายประสาทเทียม ดังรูปที่ 4.11 สำหรับงานวิจัยนี้ใช้ชุดข้อมูลเสียงวรรณยุกต์ 20 เสียงในการสอนโครงข่ายให้รู้จำเสียงวรรณยุกต์ 1 เสียง



รูปที่ 4.11 ฟังก์ชันของการสกัดลักษณะเด่นของหน่วยเสียงวรรณยุกต์

4.4 การหาความคล้ายคลึงกันของรูปแบบและวิธีการตัดสินใจ (Pattern Similarity Determination and Decision Algorithm)



รูปที่ 4.12 ขั้นตอนย่อยของการหาความคล้ายคลึงกันของหน่วยเสียงและวิธีการตัดสินใจ

ในขั้นตอนการสอนโครงข่ายประสาทเทียมให้รู้จำหน่วยเสียงนั้น หน่วยเสียงแต่ละชนิดก็จะมีขนาดของโครงข่ายที่ต่างกัน สำหรับงานวิจัยนี้ในการสอนโครงข่ายให้รู้จำหน่วยเสียงพยัญชนะต้นจะประกอบไปด้วยอินพุตจำนวน 34 โหนด ชั้นซ่อนแรก 38 โหนด ชั้นซ่อนที่สอง 25 โหนด และเอาต์พุต 21 โหนด ในการรู้จำหน่วยเสียงสระจะประกอบไปด้วย อินพุตจำนวน 18 โหนด ชั้นซ่อนแรก 22 โหนด ชั้นซ่อนที่สอง 14 โหนด และเอาต์พุต 18 โหนด ในการรู้จำหน่วยเสียงพยัญชนะท้ายจะประกอบไปด้วย อินพุตจำนวน 14 โหนด ชั้นซ่อนแรก 16 โหนด ชั้นซ่อนที่สอง 5 โหนด และเอาต์พุต 8 โหนด กำหนดฟังก์ชันกระตุ้นที่นิเวศแต่ละชั้นด้วย ไซโกทอยด์ หรือ $\tanh$ และมีการเรียนรู้โดยแพร่กลับความผิดพลาดแบบ gradient decent หรือ training นอกจากนี้ยังใช้ค่าโมเมนตัม หรือ learning rate มาช่วยในการเรียนรู้ด้วย

ตัวอย่างขั้นตอนการสอนมีดังนี้

- สร้างเมตริกซ์อินพุต

$$P = [0 \ 1 \ 2 \ 3 \ 4 \ 5 \ 6 \ 7 \ 8 \ 9 \ 10];$$

- สร้างเมตริกซ์เป้าหมาย

$$T = [0 \ 1 \ 2 \ 3 \ 4 \ 3 \ 2 \ 1 \ 2 \ 3 \ 4];$$

- นอร์มอลไลซ์ข้อมูล

$$[Pn, minp, maxp, Tn, mint, maxt] = premnmx(P, T);$$

- สร้างโครงข่ายโดยกำหนดให้ชั้นซ่อนมี 5 โหนด และชั้นเอาต์พุตมี 1 โหนด กำหนดฟังก์ชันกระตุ้นในชั้นซ่อนเป็น tansig และชั้นเอาต์พุตเป็น purelin กำหนดการแพร่กลับแบบ traingda และ โมเมนตัมเป็น learnngdm

$$net = newff(minmax(Pn), [5 \ 1], {'tansig' 'purelin'}, 'traingda', 'learnngdm');$$

- เริ่มการสอน

$$net = train(net, Pn, Tn);$$

- สิ้นสุดการสอน

เก็บตัวแปร net ,minp,maxp,mint,maxt เพื่อใช้ในการทดสอบ

ตัวอย่างขั้นตอนการทดสอบมีดังนี้

- นอร์มอลไลซ์เมตริกซ์อินพุต

$$pnew = tramnmx(P, minp, maxp);$$

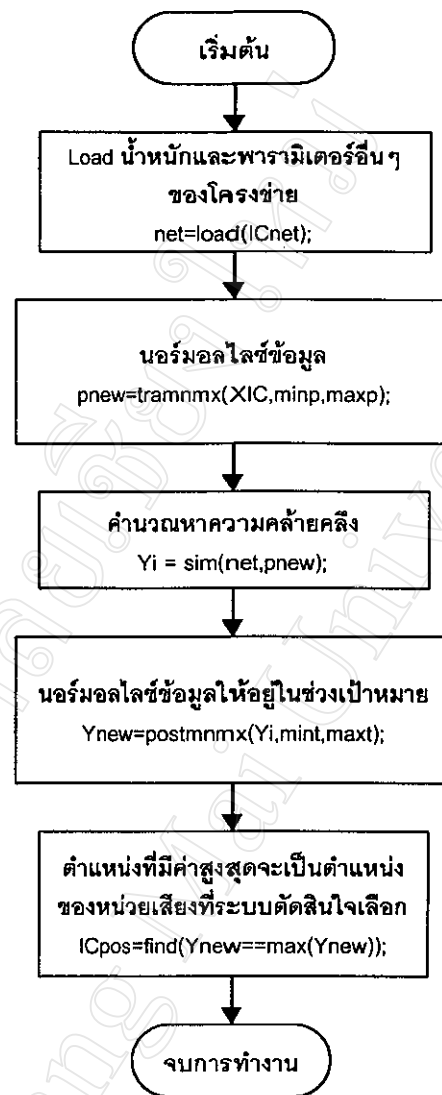
- คำนวณ (Simulate)

$$Yi = sim(net, pnew);$$

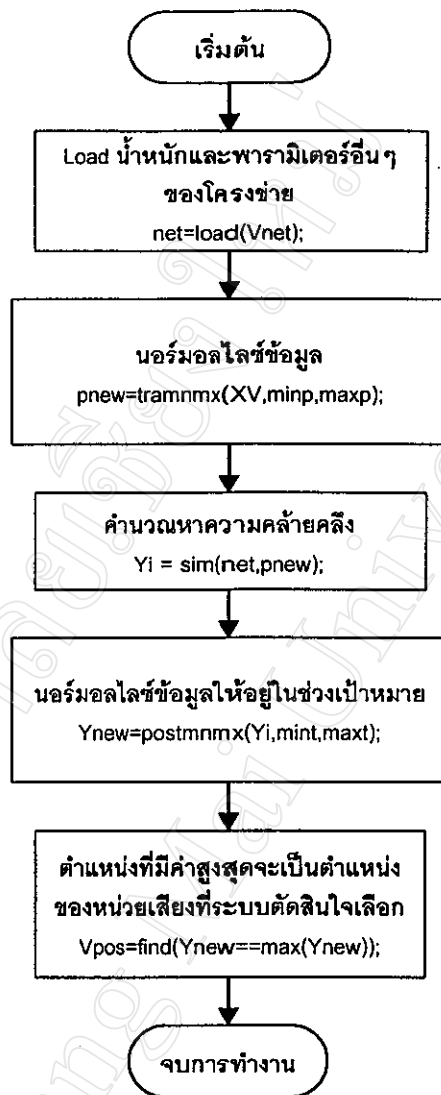
- นอร์มอลไลซ์เมตริกซ์เอาต์พุตให้อยู่ในช่วงของเมตริกซ์เป้าหมาย

$$Ynew = postmnmx(Yi, mint, maxt);$$

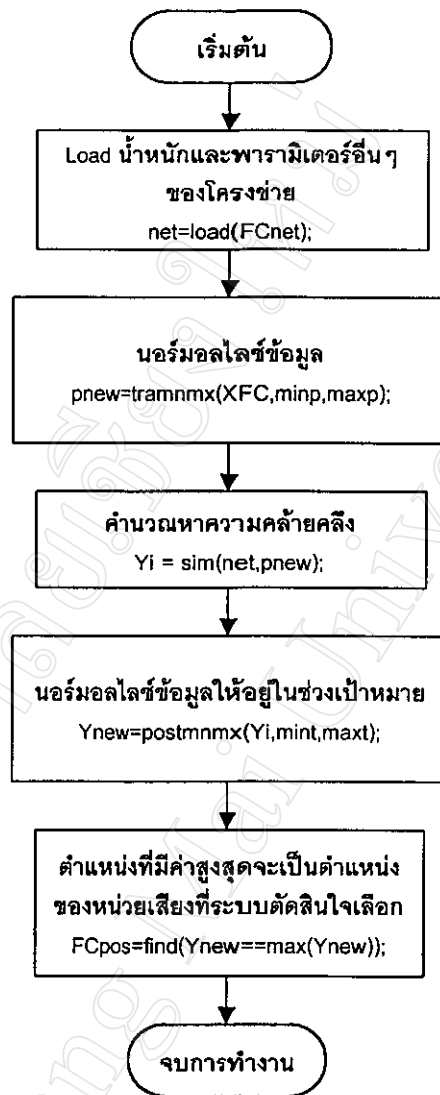
ในขั้นตอนทดสอบเราจะโหลดค่าน้ำหนักและพารามิเตอร์อื่นๆ ที่ใช้ในการนอร์มอลไลซ์ข้อมูล ซึ่งได้จากขั้นตอนการสอนในระบบรู้จำหน่วยเสียงสระ และเมตริกซ์อินพุตของหน่วยเสียงทดสอบมาคำนวณหาความคล้ายคลึงกัน ดังรูปที่ 4.13 ในตัวแปร ICnet จะเก็บค่าน้ำหนักและพารามิเตอร์ต่างๆ ของโครงข่าย (net) และค่าที่ใช้ในการนอร์มอลไลซ์ (minp,maxp,mint,maxt) และหลังจากได้ค่า ICpos แล้วเราก็จะมาตรวจสอบว่าเป็นตำแหน่งของพยัญชนะต้นเสียงใด



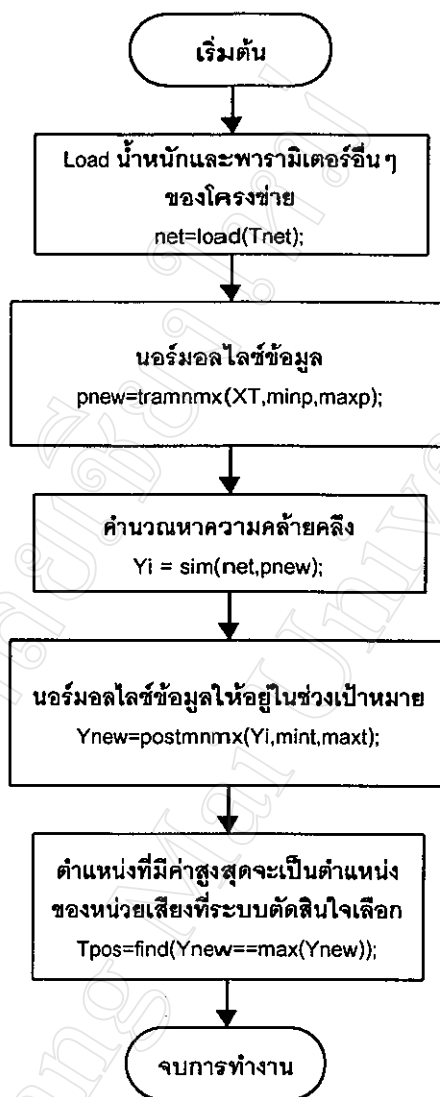
รูปที่ 4.13 ผังงานของการหาความคล้ายคลึงกันของหน่วยเสี่ยงพญชนะต้น



รูปที่ 4.14 ฟังก์ชันของการหาความคล้ายคลึงกันของหน่วยเสี่ยง



รูปที่ 4.15 ฟังก์ชันของการหาความคล้ายคลึงกันของหน่วยเสี่ยงพหุคูณ



รูปที่ 4.16 ฟังก์ชันของการหาความคล้ายคลึงกันของหน่วยเสียงวรรณยุกต์