บทที่ 3

วิธีดำเนินการวิจัย

ในบทที่ 3 นี้จะกล่าวถึงรายละเอียดของขั้นตอนวิธีในการดำเนินการวิจัย ประกอบไปด้วยรายละเอียดเกี่ยวกับการเลือกคำศัพท์ภาษาไทยที่ใช้ในการฝึกฝนและทดสอบ การสร้างระบบรู้จำหน่วยเสียงสระภาษาไทย ขั้นตอนการฝึกฝนระบบรู้จำหน่วยเสียงสระภาษาไทยและขั้นตอนการทดสอบระบบรู้จำหน่วยเสียงสระภาษาไทย นอกจากนี้ยังมีรายละเอียดของโครงข่ายประสาทเทียมและค่าลักษณะสำคัญของสัญญาณเสียงพูดที่ใช้ในงานวิจัยนี้

3.1 การเลือกคำศัพท์ภาษาไทย

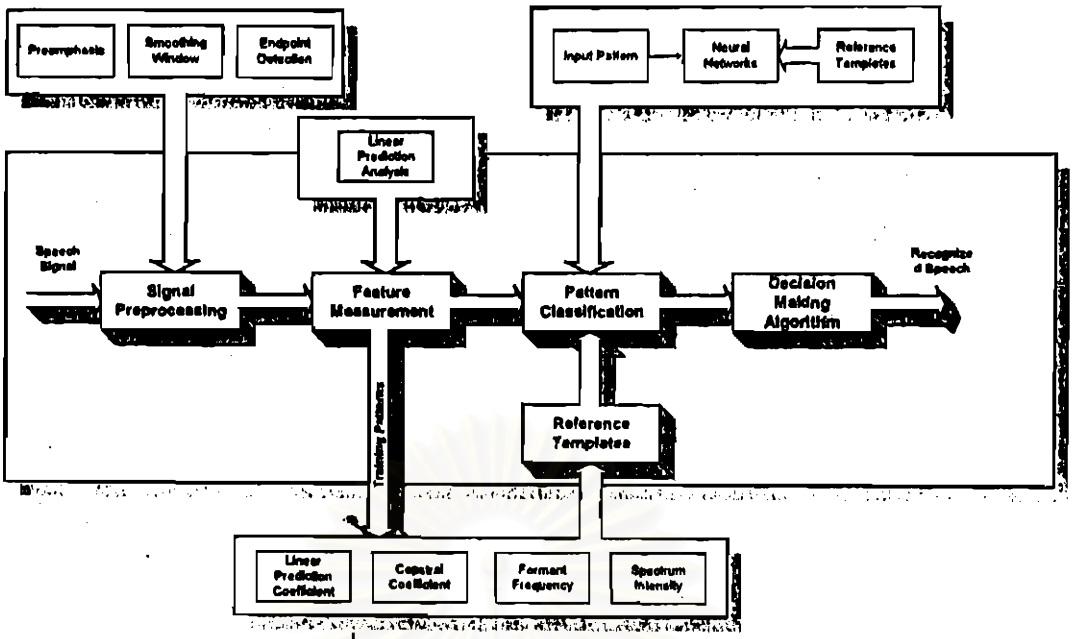
คำศัพท์สำหรับระบบรู้จำหน่วยเสียงสระภาษาไทยในงานวิจัยนี้ได้เลือกคำศัพท์จากงานวิจัยของนายวิศรุต อาชุบุตร (วิศรุต อาชุบุตร, 2539) ซึ่งมีรายละเอียดดังนี้

1. ผู้พูดเป็นเพศชายอายุระหว่าง 18 – 25 ปี
2. เป็นผู้ที่ใช้ภาษาไทยกรุงเทพเป็นภาษาพูดทั่วไป
3. คำศัพท์ที่บันทึกมีลักษณะการพูดเป็นคำโดด (isolated word)
4. บันทึกเสียงพูดด้วยการ์ดเสียง Sound Blaster 16 ของบริษัท Creative Technology
5. อัตราการชักตัวอย่าง 11 kHz และเก็บข้อมูลเสียงพูดด้วยตัวอย่างขนาด 16 บิต

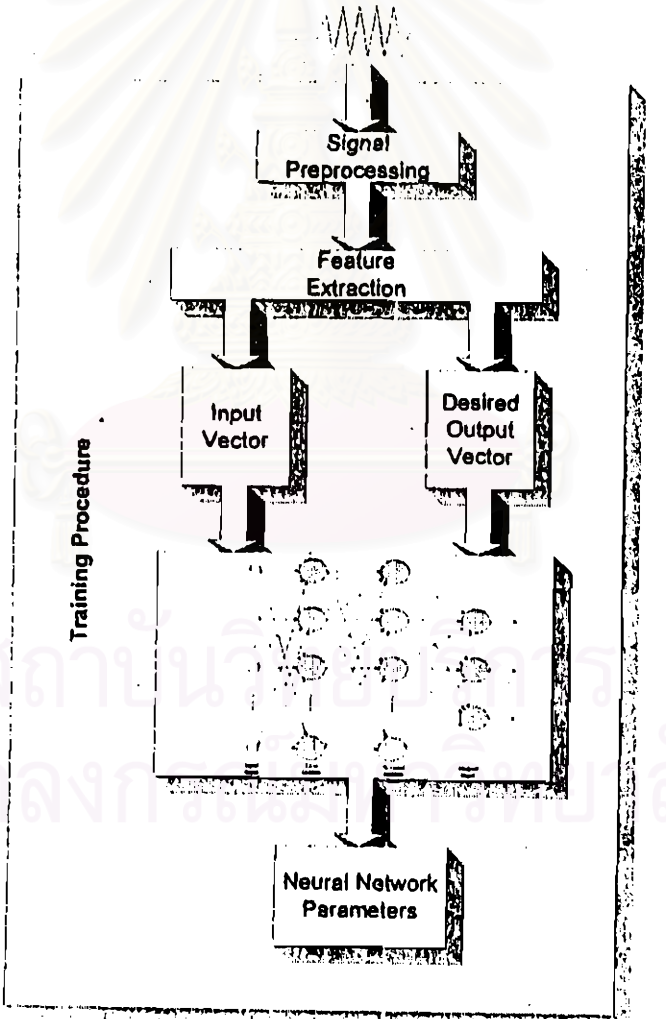
ในการคัดเลือกชุดคำศัพท์จะเลือกจากคำศัพท์ที่มีเสียงสระเดี่ยวจำนวน 9 คำสำหรับสร้างต้นแบบเสียงพูด นอกจากนี้ยังคัดเลือกชุดคำศัพท์ที่มีสระเสียงสั้นและสระเสียงยาวจำนวน 6 คำเพื่อสร้างแบบจำลองช่วงเวลาของสระเสียงสั้นและสระเสียงยาว ดังแสดงในภาคผนวก ก

3.2 การสร้างระบบรู้จำหน่วยเสียงสระภาษาไทย

รายละเอียดของระบบรู้จำหน่วยเสียงสระภาษาไทยแสดงในรูปที่ 3.1 ขั้นตอนในการสร้างระบบรู้จำหน่วยเสียงสระภาษาไทยแบ่งออกได้เป็น 2 ขั้นตอนได้แก่ขั้นตอนการฝึกฝนระบบและขั้นตอนการทดสอบระบบ โดยขั้นตอนการฝึกฝนระบบรู้จำหน่วยเสียงสระภาษาไทยจะแสดงในรูปที่ 3.2 ส่วนขั้นตอนการทดสอบระบบรู้จำหน่วยเสียงสระภาษาไทยแสดงในรูปที่ 3.3 ตามลำดับ



รูปที่ 3.1 ระบบรู้จำหน่วยเสียงสระภาษาไทย



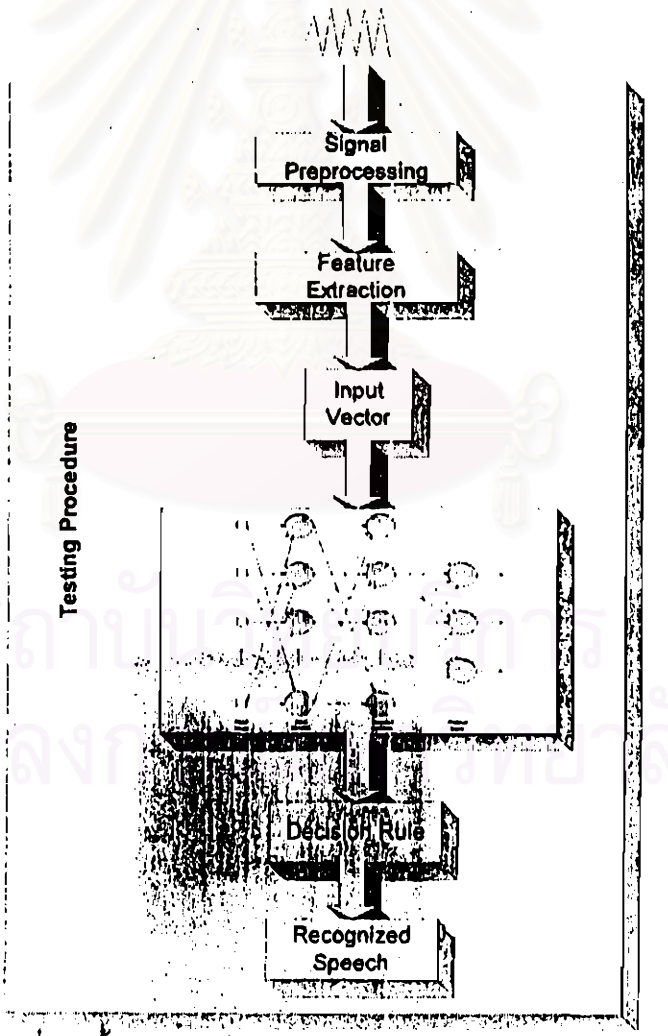
รูปที่ 3.2 ขั้นตอนการฝึกฝนระบบรู้จำหน่วยเสียงสระภาษาไทย

3.3 ขั้นตอนการฝึกฝนระบบรู้จำหน่วยเสียงสระภาษาไทย

สัญญาณเสียงพูดจะถูกนำไปผ่านขั้นตอนการประมวลผลสัญญาณเบื้อง จากนั้นสัญญาณเสียงที่ผ่านขั้นตอนการประมวลผลสัญญาณเบื้องต้นจะถูกนำมาสกัดค่าลักษณะสำคัญ ค่าลักษณะสำคัญที่ได้จากขั้นตอนการสกัดค่าลักษณะสำคัญจะอยู่ในรูปแบบของเวกเตอร์ของแต่ละกรอบสัญญาณ ค่าลักษณะสำคัญที่ได้นี้จะเป็นข้อมูลฝึกฝนในขั้นตอนการฝึกฝนระบบรู้จำหน่วยเสียงสระภาษาไทยซึ่งมีรายละเอียดในแต่ละขั้นตอนดังนี้

การประมวลผลสัญญาณเบื้องต้น (Signal Preprocessing)

ขั้นตอนการประมวลผลสัญญาณเบื้องต้นเป็นการจัดเตรียมข้อมูลเสียงจากข้อมูลเสียงที่บันทึกไว้เพื่อนำไปสกัดค่าลักษณะสำคัญต่อไป ขั้นตอนการประมวลผลสัญญาณเบื้องต้นแบ่งเป็น 3 กรรมวิธีย่อยได้แก่ กรรมวิธีเน้นล่วงหน้า (Preemphasis) กรรมวิธีวางกรอบขนาดสัญญาณ (Smoothing Window) กรรมวิธีหาจุดสิ้นสุดเสียงพูดและพยางค์ (Endpoint and Syllable Detection)



รูปที่ 3.3 ขั้นตอนการทดสอบระบบรู้จำหน่วยเสียงสระภาษาไทย

กรรมวิธีเน้นล่วงหน้า (Preemphasis)

กรรมวิธีเน้นล่วงหน้าเป็นขั้นตอนแรกที่กระทำกับสัญญาณเสียง s(n) โดยการผ่านวงจรกรอง FIR อันดับหนึ่งที่มีฟังก์ชั่นถ่ายโอนดังสมการที่ (2.1) และสมการที่ (2.2) เพื่อปรับสเปกตรัมให้มีความเรียบในงานวิจัยนี้กำหนดให้ค่าสัมประสิทธิ์ของวงจรกรอง α = 0.95 (วิศรุต อาชุบุตร, 2539; วุฒิพงษ์ พรสุขจันทรา, 2539; ชัย วุฒิวิวัฒน์ชัย; 2540)

กรรมวิธีวางกรอบขนาดสัญญาณ (Smoothing Window)

สัญญาณเสียงพูดมีลักษณะเป็นสัญญาณแบบ non-stochastic เพื่อให้สามารถใช้คณิตศาสตร์ในการประมวลผลสัญญาณเสียงพูดได้ จึงจำเป็นที่จะต้องแบ่งเสียงพูดออกเป็นช่วงเวลาสั้นๆ (10 - 30 ms) สมมติฐานที่ว่าเสียงพูดช่วงเวลาสั้นๆมีลักษณะเป็นสัญญาณแบบ stochastic (Schmid, 1990; Lee, C.H.; Soong, F.K.; Paliwal, K.K, 1996) ดังนั้นสัญญาณเสียงพูดจะถูกพิจารณาในช่วงเวลาสั้นๆ ด้วยฟังก์ชันกรอบ (Window Function) และจะมีการเลื่อนของฟังก์ชันกรอบครั้งละ 5 - 10 ms (Schmid, 1990; Lee, C.H.; Soong, F.K.; Paliwal, K.K, 1996)

สัญญาณที่ผ่านกรรมวิธีเน้นล่วงหน้าจะถูกแบ่งออกเป็นกรอบเสียงพูดโดยการคูณสัญญาณเสียงพูดกับฟังก์ชันกรอบ ผลของการวางกรอบสัญญาณคือการลดทอนแอมปลิจูดอย่างช้าๆที่บริเวณปลายแต่ละข้างของกรอบข้อมูลเสียงพูดเพื่อหลีกเลี่ยงความไม่ต่อเนื่องที่จุดปลาย ฟังก์ชันกรอบที่ใช้สำหรับการวิเคราะห์เสียงพูดในงานวิจัยนี้คือฟังก์ชันกรอบชนิด Hamming

สัญญาณที่ได้จากฟังก์ชันกรอบจะมีช่วงประสิทธิผลอยู่ค่าหนึ่ง ดังนั้นสัญญาณในแต่ละกรอบเสียงพูดจะมีจำนวนตัวอย่างเท่ากับช่วงประสิทธิผลของความยาวฟังก์ชันกรอบ ฟังก์ชันกรอบจะถูกเลื่อนครั้งละ M ตัวอย่าง โดยที่ M มีค่าน้อยกว่าช่วงประสิทธิผลของฟังก์ชันกรอบเพื่อให้การเปลี่ยนแปลงของสัญญาณเป็นไปอย่างช้าๆ ในงานวิจัยนี้จะให้ช่วงประสิทธิผลของฟังก์ชันกรอบเท่ากับ 256 (23.2 ms) ขนาดของฟังก์ชันกรอบ N เท่ากับ 512 และการเลื่อนของฟังก์ชันกรอบ M เท่ากับ 64 (5.8 ms) การกำหนดให้ช่วงประสิทธิผลและขนาดของฟังก์ชันกรอบเท่ากับ 256 และ 512 ตามลำดับเพื่อให้ตัวอย่างข้อมูลเสียงพูดในแต่ละกรอบข้อมูลเสียงพูดมีจำนวนสอดคล้องกับการแปลงฟูเรียร์อย่างเร็ว

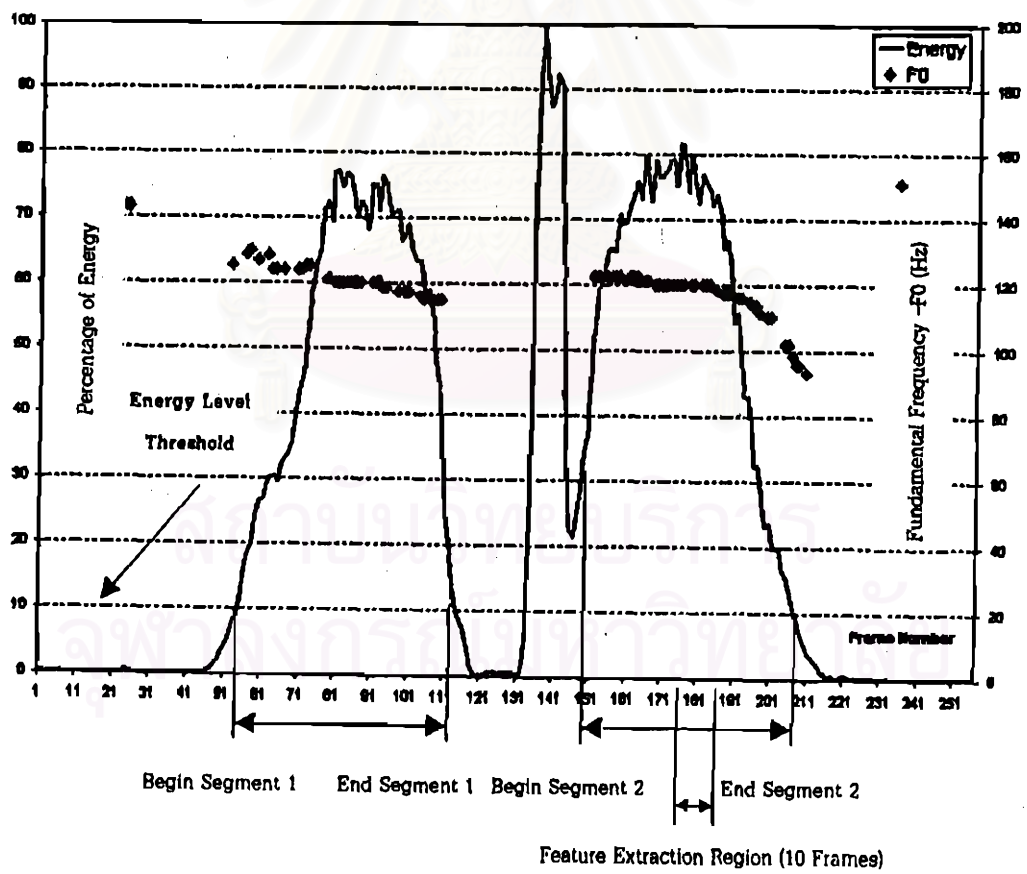
กรรมวิธีหาจุดสิ้นสุดเสียงพูดและพยางค์ (Endpoint and Syllable Detection)

ขั้นตอนกรรมวิธีหาจุดสิ้นสุดเสียงพูดและพยางค์จัดได้ว่ามีความสำคัญขั้นตอนหนึ่งในกระบวนการรู้จำเสียงพูด เนื่องจากการหาจุดสิ้นสุดเสียงพูดเป็นการแยกเอาสัญญาณส่วนที่เป็นเสียงพูดออกจากสัญญาณเสียงทั้งหมดที่บันทึกไว้ และการหาจุดสิ้นสุดของพยางค์เป็นการนำเอาสัญญาณเสียงพูดในส่วนที่เป็นพยางค์ไปวิเคราะห์และทำการรู้จำ สำหรับงานวิจัยนี้จะอาศัยแผนภูมิเส้นระดับพลังงาน (Energy Level Contour) ของเสียงพูด (วิศรุต อาชุบุตร, 2539; ชัย วุฒิวิวัฒน์ชัย; 2540) และความถี่มูลฐาน

ในงานวิจัยนี้จะดัดแปลงขั้นตอนกรรมวิธีหาจุดสิ้นสุดของพยางค์มาเป็นขั้นตอนกรรมวิธีหาขอบเขตสระโดยจะทำการกำหนดจุดเริ่มเปลี่ยนระดับพลังงาน (Energy Level Threshold) โดยกำหนดเป็นร้อยละ

ของระดับพลังงานสูงสุด การกำหนดจุดเริ่มต้นกรอบข้อมูลเสียงพูดจะนับจากจุดที่มีการเปลี่ยนระดับพลังงานมากกว่าจุดเริ่มเปลี่ยนระดับพลังงาน การกำหนดจุดสิ้นสุดกรอบข้อมูลเสียงพูดจะนับจากจุดที่มีการเปลี่ยนระดับพลังงานต่ำกว่าจุดเริ่มเปลี่ยนระดับพลังงาน ในงานวิจัยนี้ใช้จุดเริ่มเปลี่ยนระดับพลังงานเท่ากับร้อยละ 10 ของค่าระดับพลังงานสูงสุด (วิศรุต อาชุบุตร, 2539; ชัย วุฒิวิวัฒน์ชัย; 2540)

เนื่องจากข้อมูลเสียงส่วนที่เป็นสระจะมีคุณลักษณะคงตัวทำให้ใช้กรอบข้อมูลเสียงพูดส่วนที่เป็นสระจำนวนไม่มากมาทำการรู้จำได้ ไม่จำเป็นที่จะต้องนำข้อมูลเสียงสระทั้งหมดตั้งแต่จุดเริ่มต้นกรอบข้อมูลเสียงพูดส่วนที่เป็นสระจนกระทั่งถึงจุดสิ้นสุดกรอบข้อมูลเสียงพูดส่วนที่เป็นสระมาใช้ทั้งหมด จากการที่เสียงพยัญชนะบางเสียงที่เป็นเสียงพ่นลมเช่น /ph/ /th/ /kh/ /ch/ มีแอมปลิจูดสูงทำให้การใช้ค่าพลังงานอย่างเดียวไม่เพียงพอสำหรับการหาจุดเริ่มต้นกรอบข้อมูลเสียงพูดส่วนที่เป็นสระ ทำให้การกำหนดจุดเริ่มต้นกรอบข้อมูลเสียงพูดส่วนที่เป็นสระคลาดเคลื่อน จากสาเหตุดังกล่าวจำเป็นที่จะต้องใช้ความถี่มูลฐานมาพิจารณาร่วมด้วย โดยกรอบข้อมูลเสียงพูดที่มีค่าระดับพลังงานสูงกว่าจุดเริ่มเปลี่ยนระดับพลังงานแต่ไม่มีความถี่มูลฐานจะไม่นับเป็นจุดเริ่มต้นกรอบข้อมูลเสียงพูดส่วนที่เป็นสระ ในงานวิจัยนี้ใช้กรอบข้อมูลเสียงพูดส่วนที่เป็นสระจำนวน 10 กรอบข้อมูลเสียงพูดโดยนับจากจุดกึ่งกลางของพยางค์ที่หาขอบเขตพยางค์โดยใช้แผนภูมิเส้นระดับพลังงานและความถี่มูลฐานร่วมกันดังแสดงในรูปที่ 3.4



รูปที่ 3.4 การหาขอบเขตกรอบข้อมูลเสียงพูดส่วนที่เป็นสระโดยใช้ค่าพลังงานและความถี่มูลฐาน

การวิเคราะห์และวัดค่าลักษณะสำคัญ (Feature Measurement)

การวิเคราะห์และวัดค่าลักษณะสำคัญเป็นวิธีการหาตัวแทนของสัญญาณเสียงพูดเพื่อวิเคราะห์และจำแนกเสียงพูดที่มีความแตกต่างกันสำหรับระบบรู้จำเสียงพูด ค่าลักษณะสำคัญที่ใช้ในงานวิจัยนี้มีพื้นฐานจากแบบจำลองการประมาณพันธะเชิงเส้นซึ่งได้แก่ ค่าสัมประสิทธิ์การประมาณพันธะเชิงเส้น ค่าสัมประสิทธิ์ cepstral ความถี่ฟอร์แมนท์และค่าความเข้มของสเปกตรัม

1. สัมประสิทธิ์การประมาณพันธะเชิงเส้น (Linear Prediction Coefficient)

การคำนวณค่าสัมประสิทธิ์การประมาณพันธะเชิงเส้นจะใช้วิธีอัตสหสัมพันธ์ เมตริกซ์ของค่าอัตสหสัมพันธ์อยู่ในรูปของเมตริกซ์ Toeplitz ซึ่งมีลักษณะสมมาตร และทุกๆ สมาชิกในแนวทแยงมุมมีค่าเท่ากัน ทำให้การแก้สมการเพื่อหาค่าสัมประสิทธิ์ โดยวิธีนี้ง่ายกว่าการใช้วิธีความแปรปรวนร่วม (Covariance) และจะใช้ขั้นตอนวิธีการวนซ้ำ Levinson-Durbin ในแต่ละกรอบข้อมูลเสียงพูด ค่าสัมประสิทธิ์การประมาณพันธะเชิงเส้นที่ได้ในแต่ละกรอบข้อมูลเสียงพูดจะเป็นเวกเตอร์ที่มีมิติเท่ากับอันดับของพหุนามสัมประสิทธิ์การประมาณพันธะเชิงเส้น สำหรับงานวิจัยนี้จะทดลองแปรค่าอันดับของสัมประสิทธิ์การประมาณพันธะเชิงเส้นเพื่อให้ได้ค่าที่เหมาะสมที่สุดสำหรับระบบรู้จำหน่วยเสียงสระภาษาไทย

2. สัมประสิทธิ์ Cepstral (Cepstral Coefficient)

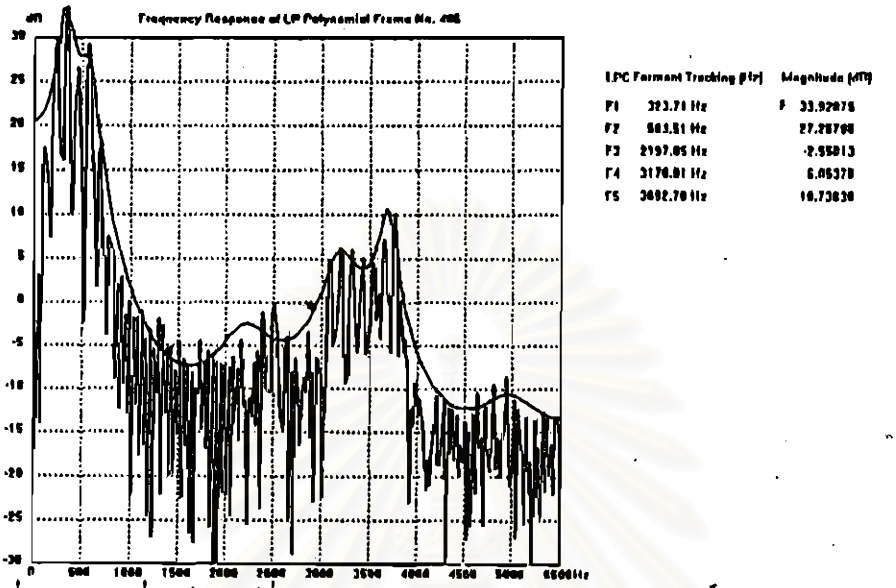
ค่าสัมประสิทธิ์ Cepstral คำนวณได้ 2 วิธีคือคำนวณจากการแปลงฟูเรียร์อย่างเร็วและคำนวณจากแบบจำลองการประมาณพันธะเชิงเส้น ในงานวิจัยนี้จะเลือกวิธีการคำนวณค่าสัมประสิทธิ์ Cepstral จากแบบจำลองการประมาณพันธะเชิงเส้นจะมีข้อดีคือใช้เวลาในการคำนวณที่น้อยกว่าและค่าสัมประสิทธิ์ Cepstral ที่ได้นี้จะให้ค่าสหสัมพันธ์ (correlation) ระหว่างค่าสัมประสิทธิ์ของเสียงที่ต่างกันมีค่าที่น้อยมาก (Huang, 1989) นอกจากนี้จะทำการทดลองแปรค่าอันดับของสัมประสิทธิ์ Cepstral เพื่อให้ได้ค่าที่เหมาะสมที่สุดสำหรับระบบรู้จำหน่วยเสียงสระภาษาไทย

3. ความถี่ฟอร์แมนท์ (Formant Frequency)

วิธีการคำนวณความถี่ฟอร์แมนท์จะอยู่บนพื้นฐานของแบบจำลองการประมาณพันธะเชิงเส้นมี 2 วิธี คือการคำนวณความถี่ฟอร์แมนท์จากรากของพหุนามของสัมประสิทธิ์การประมาณพันธะเชิงเส้นและการคำนวณความถี่ฟอร์แมนท์จากวิธีการตรวจจับค่ายอดของสเปกตรัม ความถี่ฟอร์แมนท์ที่มีผลต่อการจำแนกเสียงสระจะมีไม่เกิน 3 ค่าคือความถี่ฟอร์แมนท์ที่หนึ่ง สอง และสามตามลำดับ (สุดาพร ลักษณียนาวิน, 2529) ในงานวิจัยนี้จะใช้วิธีการคำนวณความถี่ฟอร์แมนท์จากรากของพหุนามของสัมประสิทธิ์การประมาณพันธะเชิงเส้นดังแสดงค่าในรูปที่ 3.5 และจะทดลองใช้จำนวนความถี่ฟอร์แมนท์ 2 ค่าและ 3 ค่าเพื่อหาจำนวนความถี่ฟอร์แมนท์ที่ให้ผลการจำแนกหน่วยเสียงสระภาษาไทยได้ดีที่สุด

4. ความเข้มสเปกตรัม (Spectrum Intensity)

ความเข้มสเปกตรัมเป็นค่าลักษณะสำคัญอีกค่าหนึ่งที่บ่งบอกถึงรูปร่างสเปกตรัมของเสียงพูด ในงานวิจัยนี้จะใช้ค่าความเข้มสเปกตรัมร่วมกับค่าความถี่ฟอร์แมนท์เป็นค่าลักษณะสำคัญอีกประเภทหนึ่ง



รูปที่ 3.5 ความถี่ฟอร์แมนท์ที่คำนวณจากรากของพหุนามของสัมประสิทธิ์การประมาณพันธะเชิงเส้น

โครงข่ายประสาทเทียม

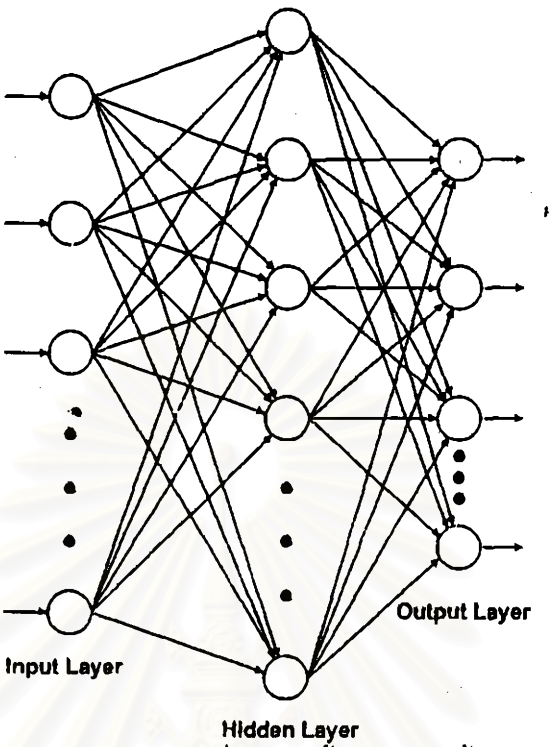
โครงข่ายประสาทเทียมที่ใช้สำหรับงานวิจัยนี้เป็นโครงข่ายประสาทเทียมแบบ Multilayer Perceptron ที่มีจำนวนชั้นซ่อนตัว 1 ชั้นเนื่องจากเพียงพอสำหรับการประมาณฟังก์ชันไม่เป็นเชิงเส้นใดๆ ได้ (Schalkoff, 1992) โดยมีจำนวนโนดที่เหมาะสมดังแสดงในรูปที่ 3.6 จำนวนโนดในชั้นซ่อนตัวนี้จะต้องมีเพียงพอที่จะทำให้โครงข่ายประสาทเทียมสามารถรู้จำได้ดีและจะต้องมีจำนวนไม่มากเกินไปเพราะจะทำให้เวลาในการฝึกฝนและทดสอบนานเกินไปนอกจากนี้ยังเป็นการสิ้นเปลืองหน่วยความจำ จำนวนโนดในชั้นข้อมูลเข้าจะเท่ากับผลคูณของมิติของเวกเตอร์ที่เป็นค่าลักษณะสำคัญในกรอบสัญญาณกับจำนวนกรอบสัญญาณเสียงพูด จำนวนโนดในชั้นข้อมูลออกจะเท่ากับจำนวนกลุ่มข้อมูลที่ต้องการจำแนก รูปแบบของข้อมูลเข้าของโครงข่ายประสาทเทียมแสดงในรูปที่ 3.7

ข้อมูลออกเป้าหมาย (Target Output)

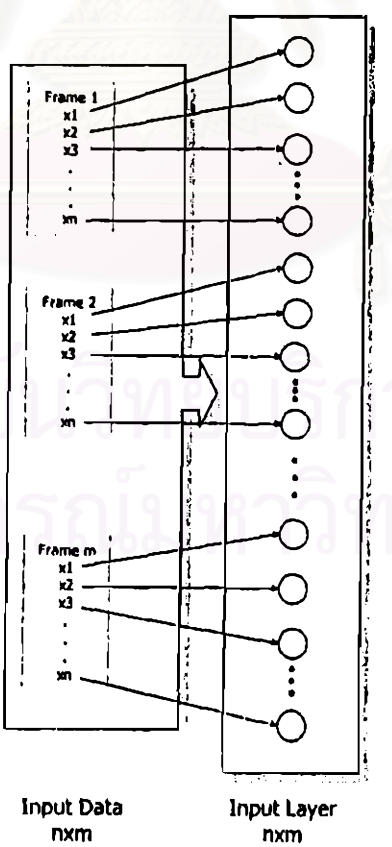
จำนวนมิติของเวกเตอร์ข้อมูลออกเป้าหมายจะเท่ากับกลุ่มข้อมูลที่ต้องการจำแนก ค่าขององค์ประกอบย่อยของเวกเตอร์ข้อมูลออกเป้าหมายจะขึ้นอยู่กับฟังก์ชันกระตุ้น (Activation Function) ถ้าฟังก์ชันกระตุ้นเป็นฟังก์ชัน sigmoid ซึ่งมีค่าขอบเขตล่างและขอบเขตบนเป็น 0 และ 1 ตามลำดับ จะเขียนเงื่อนไขในการกำหนดค่าแทนผลตอบเป้าหมายได้เป็น

$$d_{k,j} = \begin{cases} 1 & x_j \in C_k \\ 0 & x_j \notin C_k \end{cases} \qquad (3.2)$$

เมื่อข้อมูลฝึกฝน x_j อยู่ในกลุ่มข้อมูล C_k การกำหนดข้อมูลออกเป้าหมายดังกล่าวจะใช้ในงานวิจัยนี้



รูปที่ 3.6 โครงข่ายประสาทเทียมแบบ MLP ที่มีจำนวนชั้นซ่อนตัว 1 ชั้น (ชัย วุฒิวิวัฒน์ชัย, 2540)



รูปที่ 3.7 รูปแบบข้อมูลเข้าที่ชั้นข้อมูลเข้าของโครงข่ายประสาทเทียม

พารามิเตอร์ที่สำคัญในการฝึกฝน

ค่าพารามิเตอร์ที่สำคัญในการฝึกฝนโครงข่ายประสาทเทียมสำหรับงานวิจัยนี้ได้แก่

1. ค่าอัตราการเรียนรู้ (Learning Rate) η การกำหนดค่าอัตราการเรียนรู้มีความสำคัญมาก ถ้ากำหนดค่าอัตราการเรียนรู้น้อยเกินไปจะทำให้การเรียนรู้นานมากและถ้ากำหนดค่าอัตราการเรียนรู้มากจะทำให้สูญเสียการเรียนรู้ที่ผ่านมา ในงานวิจัยนี้จะใช้ค่าอัตราการเรียนรู้เท่ากับ 0.1 (วุฒิพงษ์ พรสุขจันทรา, 2539; ชัย วุฒิวิวัฒน์ชัย, 2540)
2. ค่าคงที่โมเมนตัม (Momentum Constant) α การมีพจน์โมเมนตัมจะทำให้เวลาในการเรียนรู้เร็วขึ้นและป้องกันความไม่มีเสถียรภาพในการฝึกฝน ในงานวิจัยนี้ใช้ค่าคงที่โมเมนตัมเท่ากับ 0.9 (วุฒิพงษ์ พรสุขจันทรา, 2539; ชัย วุฒิวิวัฒน์ชัย, 2540)
3. ค่าผิดพลาดผ่านระดับ (Error Threshold) E_{TH} การนับจำนวนข้อมูลฝึกฝนที่ฝึกได้ถูกต้อง ข้อมูลฝึกฝนนั้นจะต้องมีค่าผิดพลาดต่ำกว่าค่าผิดพลาดผ่านระดับ ถ้ากำหนดค่าผิดพลาดผ่านระดับให้มีค่าต่ำเกินไปจะทำให้เวลาในการเรียนรู้นานมากและอาจทำให้โครงข่ายประสาทเทียมรู้จำรูปแบบที่ฝึกฝนมากเกินไป และถ้ากำหนดค่าผิดพลาดผ่านระดับให้มีค่าสูงเกินไปจะทำให้อัตราการรู้จำต่ำ ในงานวิจัยนี้กำหนดค่าผิดพลาดผ่านระดับมีค่า 0.001

กระบวนการฝึกฝน

กระบวนการฝึกฝนโครงข่ายประสาทเทียมสำหรับงานวิจัยนี้มีรายละเอียดดังนี้

1. **กำหนดค่าเริ่มต้น** กำหนดให้ค่าน้ำหนักการเชื่อมต่อเริ่มต้นมีการแจกแจงแบบสม่ำเสมอให้มีค่าอยู่ในช่วง -1 ถึง 1 (วุฒิพงษ์ พรสุขจันทรา, 2539; ชัย วุฒิวิวัฒน์ชัย, 2540)
2. **กำหนดชุดข้อมูลฝึกฝน** กำหนดลำดับของข้อมูลฝึกฝนและผลตอบเป้าหมายของข้อมูลฝึกฝนสำหรับเข้าฝึกฝนในโครงข่ายประสาทเทียมเป็นแบบสุ่ม
3. **ขั้นตอนวิธีการฝึกฝน** ใช้ขั้นตอนวิธีการส่งค่าย้อนกลับ (Backpropagation) แบบ Pattern Mode ซึ่งมีข้อดีกว่าแบบ Batch Mode หลายประการ (Haykin, 1994) เช่นต้องการหน่วยความจำในการเก็บข้อมูลที่น้อยกว่า หลีกเลี่ยงการหยุดที่จุดต่ำสุดเฉพาะที่ (Local Minimum) ได้ดีกว่า Batch Mode เป็นต้น
4. **การวนซ้ำ** เนื่องจากใช้ขั้นตอนวิธีการส่งค่าย้อนกลับ แบบ Pattern Mode ในการฝึกฝนซึ่งมีเงื่อนไขในการหยุดฝึกฝนคือ ทำการวนซ้ำจนกระทั่งจำนวนข้อมูลที่ฝึกฝนมีค่าผิดพลาดต่ำกว่าค่าผิดพลาดผ่านระดับที่กำหนดไว้เป็นจำนวนร้อยละ 98 ของจำนวนข้อมูลฝึกฝนทั้งหมด ข้อมูลฝึกฝนนั้นจะถูกนับว่าฝึกฝนได้ถูกต้องจะต้องมีค่าผิดพลาดต่ำกว่าค่าผิดพลาดผ่านระดับ ไม่ได้อาศัยการพิจารณาค่าผิดพลาดรวมของข้อมูลฝึกฝนทั้งหมดว่าต่ำกว่ากว่าค่าผิดพลาดผ่านระดับเป็นเงื่อนไขในการหยุดฝึกฝน

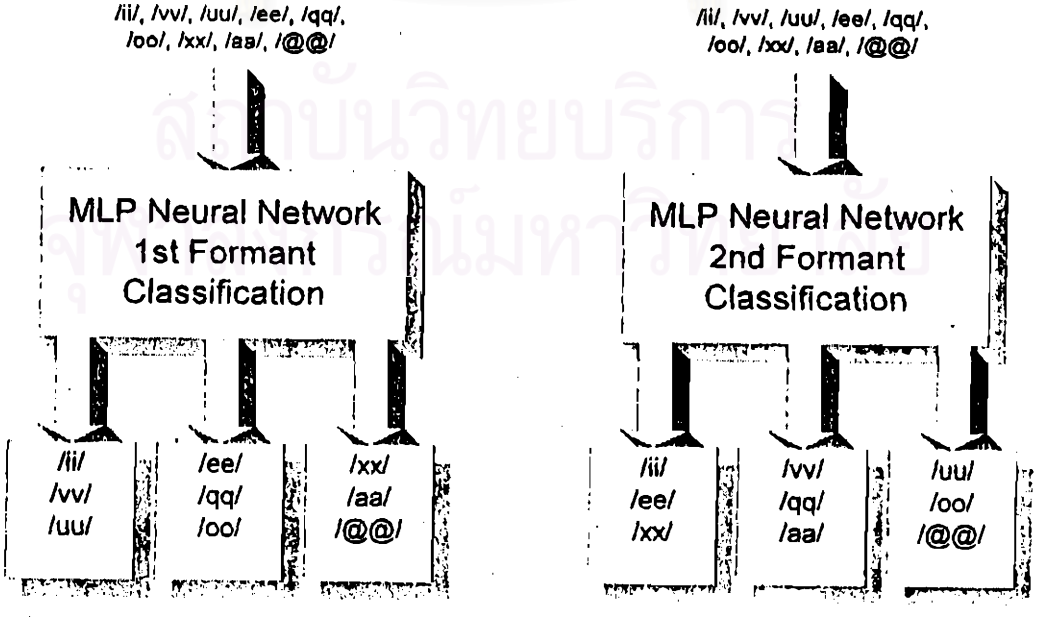
วิจัยนี้จะใช้โครงข่ายประสาทเทียม 1 โครงข่ายในการจัดกลุ่มหน่วยเสียงสระและใช้โครงข่ายประสาทเทียมอีก 3 โครงข่ายในการรู้จำหน่วยเสียงสระ และทำการเปรียบเทียบกับการใช้โครงข่ายประสาทเทียม 1 โครงข่ายในการรู้จำหน่วยเสียงสระทั้งหมด โครงข่ายประสาทเทียมที่ใช้ในการจัดกลุ่มหน่วยเสียงสระจะถูกฝึกฝนให้จำแนกหน่วยเสียงสระตามความถี่ฟอร์แมนท์ (ดูบทที่2 รูปที่ 2.21) จะทำให้การจัดกลุ่มหน่วยเสียงสระตามความถี่ฟอร์แมนท์ แบ่งได้เป็น 2 กรณีคือแบ่งกลุ่มสระที่มีความถี่ฟอร์แมนท์ที่ 1 ใกล้เคียงกัน (แบ่งตามระดับความสูงของลิ้น สูง-กลาง-ต่ำ) และแบ่งกลุ่มสระที่มีความถี่ฟอร์แมนท์ที่ ,2 ใกล้เคียงกัน (แบ่งตามตำแหน่งของลิ้น หน้า-กลาง-หลัง) ตามลำดับดังแสดงในรูปที่ 3.8 ก และจะใช้โครงข่ายประสาทเทียมย่อยในการรู้จำหน่วยเสียงสระในแต่ละกลุ่มย่อย สระที่ถูกแบ่งกลุ่มตามความถี่ฟอร์แมนท์ที่ 1 จะใช้โครงข่ายประสาทเทียมย่อยในการฝึกฝนตามรูปที่ 3.8 ข ส่วนสระที่ถูกแบ่งกลุ่มตามความถี่ฟอร์แมนท์ที่ 2 จะใช้โครงข่ายประสาทเทียมย่อยในการฝึกฝนตามรูปที่ 3.8 ค ตามลำดับ

3.4 ขั้นตอนการทดสอบระบบรู้จำหน่วยเสียงสระภาษาไทย

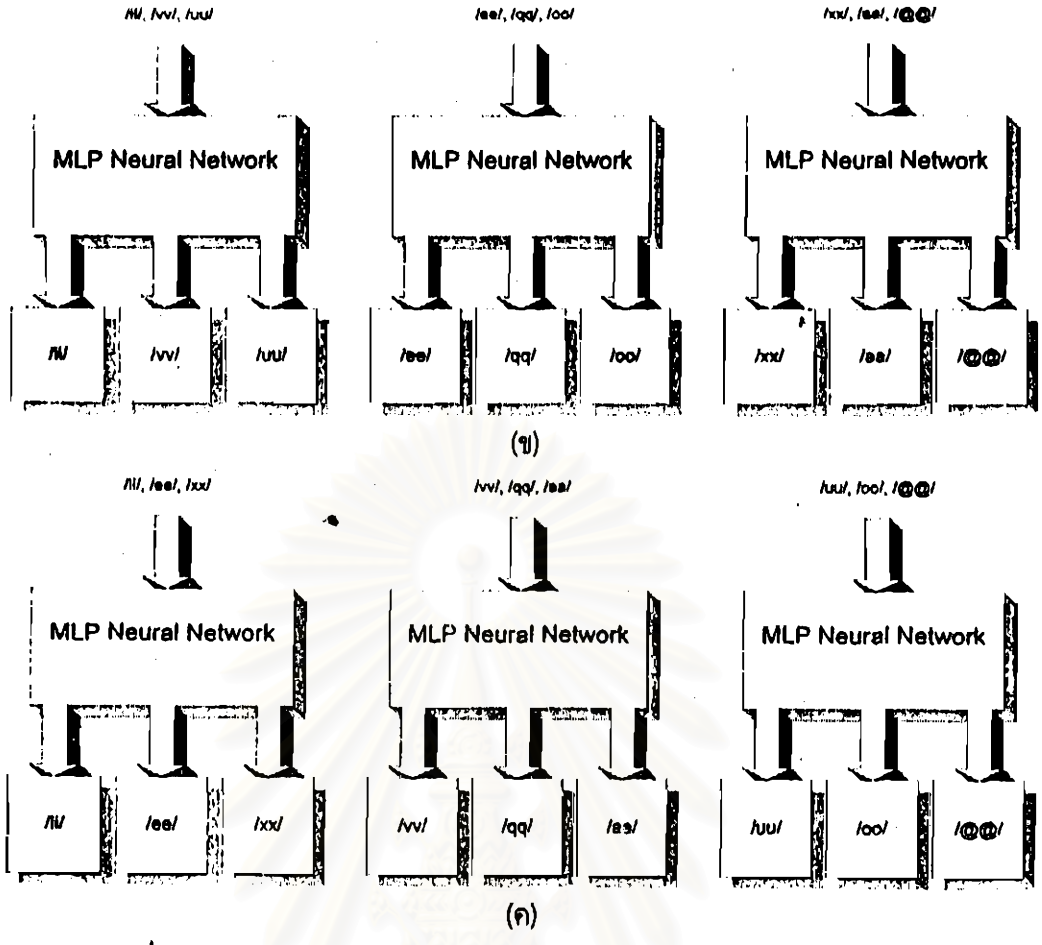
ขั้นตอนการทดสอบระบบรู้จำหน่วยเสียงสระภาษาไทยมีรายละเอียดแสดงในรูปที่ 3.3 ซึ่งประกอบไปด้วยขั้นตอนการประมวลผลสัญญาณเบื้องต้น การวิเคราะห์และวัดค่าลักษณะสำคัญซึ่งเป็นวิธีการเดียวกับที่ใช้ในขั้นตอนการฝึกฝนระบบรู้จำหน่วยเสียงสระภาษาไทยและขั้นตอนวิธีการตัดสินใจ

การรู้จำสระเสียงสั้น-สระเสียงยาว

สระเสียงสั้นและสระเสียงยาวที่เกิดจากตำแหน่งของลิ้นที่เดียวกันจะมีความถี่ฟอร์แมนท์เดียวกันแต่จะมีช่วงเวลาที่ต่างกัน สระเสียงสั้นจะมีช่วงเวลาที่สั้นกว่าสระเสียงยาวดังแสดงในรูปที่ 3.9 ในงานวิจัยนี้จะทำการสร้างแบบจำลองช่วงเวลาของสระเสียงสั้นและสระเสียงยาวให้มีการแจกแจงแบบปกติหรือแบบเกาส์ (Normal Distribution or Gaussian Distribution) และจะทดสอบแบบจำลองดังกล่าวโดยใช้กฎของ Bayes (Schalkhoff, 1992)



(ก)



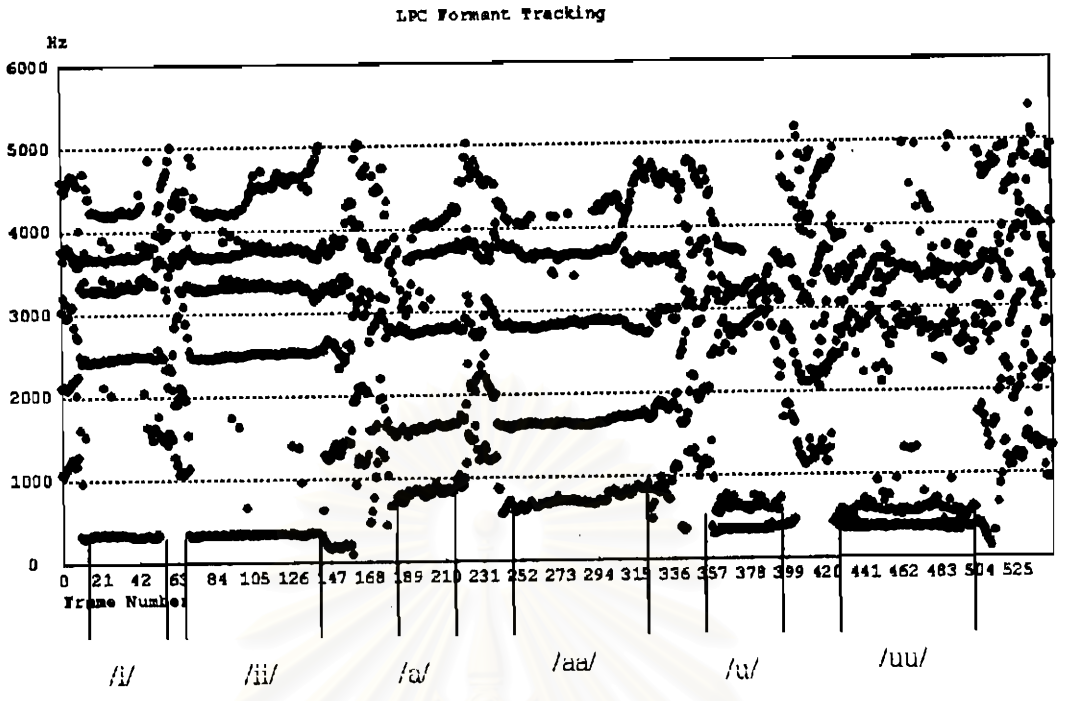
รูปที่ 3.8 การใช้โครงข่ายประสาทเทียมในการรู้จำหน่วยเสียงสระสำหรับงานวิจัยนี้

(ก) การใช้โครงข่ายประสาทเทียมจัดกลุ่มหน่วยเสียงสระ

(ข) การใช้โครงข่ายประสาทเทียมรู้จำหน่วยเสียงสระในกลุ่มที่แบ่งตามความถี่ฟอร์แมนท์ที่ 1

(ค) การใช้โครงข่ายประสาทเทียมรู้จำหน่วยเสียงสระในกลุ่มที่แบ่งตามความถี่ฟอร์แมนท์ที่ 2

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 3.9 ความถี่ฟอร์แมนท์ของสระเสียงสั้น-ยาว

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย