บทที่ 3
ขั้นตอนวิธีดำเนินการวิจัย

ในบทนี้กล่าวถึงรายละเอียดของขั้นตอนในการทำวิจัยทั้งหมด โดยเริ่มตั้งแต่การเลือกบทคำพูดที่ใช้ในงานวิจัย การเก็บตัวอย่างข้อมูลสัญญาณเสียงพูด การสร้างระบบการบ่งชี้ผู้พูดแบบขึ้นกับบทคำพูดโดยใช้แบบจำลองฮิดเดนมาร์คอฟแบบดิสครีต ขั้นตอนการฝึกฝนระบบและขั้นตอนการทดสอบระบบ

3.1 การเลือกบทคำพูด

บทคำพูดที่ใช้ในงานวิจัยนี้เป็นประโยคคำพูดภาษาไทยแบบต่อเนื่องจำนวน 3 ประโยคที่สร้างขึ้นมาเพื่อสื่อความหมายในการประยุกต์ใช้งานกับระบบรักษาความปลอดภัยโดยใช้เสียงพูด มีรายละเอียดดังนี้

ประโยคที่ 1 "ขอผ่านเข้าสู่ระบบ"

ประโยคที่ 2 "หมายเลขผ่านคือ"

ประโยคที่ 3 "สาม ห้า สอง เก้า สี่"

สังเกตได้ว่าประโยคที่ 1 เป็นประโยคคำพูดที่ยาวที่สุดคือมีจำนวนพยางค์เท่ากับ 6 พยางค์ และประโยคที่ 2 เป็นประโยคคำพูดที่สั้นที่สุดคือมีจำนวนพยางค์เท่ากับ 4 พยางค์ ส่วนประโยคที่ 3 ประกอบไปด้วยตัวเลขจำนวน 5 ตัวเลขที่ประกอบด้วยเสียง "ส" และเสียงวรรณยุกต์โท 2 เสียง จัตวา 2 เสียง และเสียงเอก 1 เสียง เนื่องจากว่าเสียง "ส" เป็นเสียงเสียดแทรกที่มีความถี่สูงและมีความแตกต่างในการออกเสียงของผู้พูดแต่ละบุคคลมาก (อังสนา จามิกรณ์, 2532; Harris, 1972) ส่วนเสียงวรรณยุกต์โทและจัตวาเป็นเสียงตก (Falling Tone) และเสียงขึ้น (Rising Tone) ซึ่งเป็นเสียงที่มีการเปลี่ยนแปลงค่าความถี่มูลฐานเป็นเส้นโค้งตกและขึ้นอย่างรวดเร็ว (Dynamic Tones) (ณัฐกร ทับทอง, 2530) ทำให้มีโอกาสเกิดความแตกต่างในการออกเสียงของผู้พูดแต่ละคนได้เช่นกัน สำหรับการเรียงลำดับของตัวเลขในประโยคที่ 3 นั้นเป็นการเรียงลำดับของเสียงจัตวาและโทสลับกัน เพื่อต้องการให้เกิดความแตกต่างในการออกเสียงของผู้พูดแต่ละคนให้มากที่สุด

3.2 การเก็บตัวอย่างข้อมูลสัญญาณเสียงพูด

การเก็บตัวอย่างข้อมูลสัญญาณเสียงพูดของแต่ละบุคคลจะบันทึกเสียงพูดทีละประโยค โดยบันทึกเสียงจำนวน 5 ครั้งต่อคนต่อประโยคในแต่ละรอบของการบันทึก การเก็บตัวอย่างของงานวิจัยนี้ได้เก็บตัวอย่าง 2 รอบ โดยแต่ละรอบของการบันทึกห่างกันประมาณ 1 ถึง 2 สัปดาห์ขึ้นอยู่กับความพร้อม และความสะดวกของผู้พูด สาเหตุที่ต้องบันทึก 2 รอบเพราะว่าต้องการเก็บตัวอย่างสัญญาณเสียงพูดของผู้พูดที่เวลาต่างกันเพื่อต้องการสัญญาณเสียงพูดที่บ่งถึงลักษณะการพูดเฉพาะบุคคลและมีเสถียรภาพทางเวลา เพราะฉะนั้นผู้พูดแต่ละคนจะถูกบันทึกเสียงทั้งหมด 10 ครั้งต่อประโยค

3.2.1 หลักเกณฑ์ในการคัดเลือกผู้พูด

คัดเลือกผู้พูดจากกลุ่มอายุต่างๆ ที่กำหนดไว้ 3 กลุ่มดังนี้

1. กลุ่มผู้พูดที่มีอายุตั้งแต่ 18 ถึง 27 ปี
2. กลุ่มผู้พูดที่มีอายุตั้งแต่ 28 ถึง 37 ปี
3. กลุ่มผู้พูดที่มีอายุตั้งแต่ 38 ถึง 47 ปี

การแบ่งผู้พูดออกเป็น 3 กลุ่มตามอายุนั้นเพื่อครอบคลุมลักษณะการพูดของผู้พูดที่เป็นตัวแทนของประชากรที่พูดภาษาไทยในวัยทำงาน ซึ่งจะนำไปสร้างระบบบ่งชี้ผู้พูดแบบขึ้นกับบทคำพูดที่สามารถใช้ได้เป็นอย่างดีกับประชากรทั่วไปที่มีอายุอยู่ในช่วงวัยทำงาน ในงานวิจัยนี้คัดเลือกผู้พูดในแต่ละกลุ่มอายุมากลุ่มละ 4 คนเป็นเพศชายและเพศหญิงอย่างละ 2 คน รวมผู้พูดที่ใช้ในการบ่งชี้มีทั้งหมด 12 คน เป็นเพศชายและเพศหญิงอย่างละ 6 คน

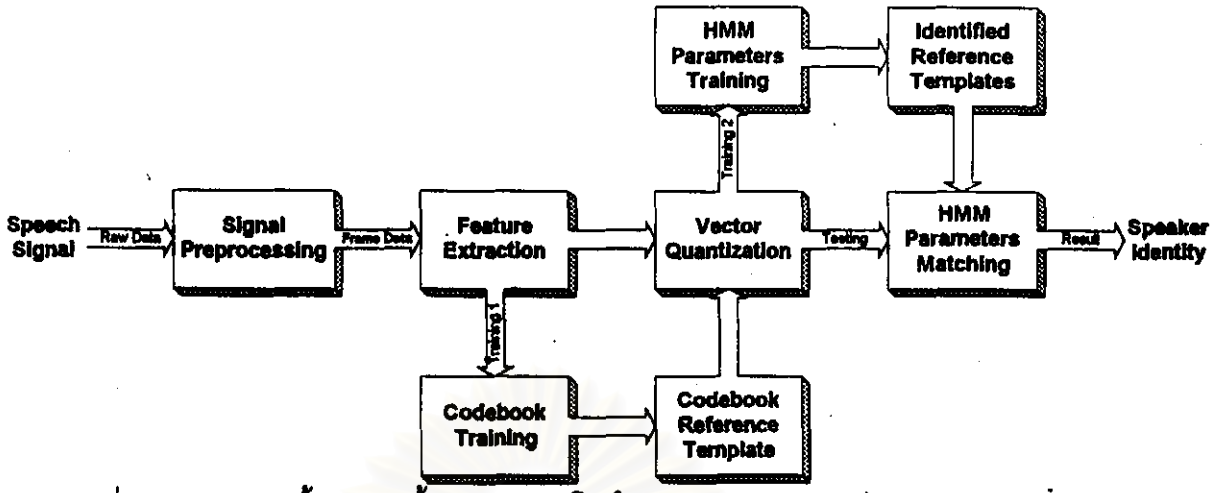
3.2.2 หลักเกณฑ์ในการบันทึกสัญญาณเสียงพูด

1. บันทึกเสียงพูดด้วยไมโครโฟนผ่านการ์ดเสียงขนาด 16 บิตที่มีอัตราการชักตัวอย่างเท่ากับ 11025 เฮิรตซ์
2. บันทึกเสียงพูดภายใต้สภาพแวดล้อมของห้องปฏิบัติการวิจัย ซึ่งคล้ายคลึงกับสภาพแวดล้อมของสถานที่ทำงานทั่วๆ ไป (Office Environment)
3. ขณะบันทึกเสียงผู้พูดต้องอยู่ในสภาพปกติ คือไม่มีอาการป่วยหรือไม่สบายจนทำให้ออกเสียงต่างไปจากปกติ
4. ในการบันทึกเสียงผู้พูดต้องใช้ภาษาไทยสำเนียงกรุงเทพฯ เป็นภาษาพูด

ตัวอย่างสัญญาณเสียงพูดที่บันทึกได้มีจำนวนทั้งหมดเท่ากับ 360 ตัวอย่าง (12 คน×3 ประโยค×10 ครั้ง) โดยแบ่งออกเป็น 2 ชุด ได้แก่ชุดฝึกฝน (Training Set) และชุดทดสอบ (Testing Set) อย่างละเท่าๆ กัน โดยแบ่งตัวอย่างสัญญาณเสียงพูดที่บันทึกได้แบบสุ่มมา 5 ตัวอย่างของแต่ละบุคคลในแต่ละประโยคครึ่งหนึ่ง 180 ตัวอย่าง ใช้ในการสร้างและฝึกฝนระบบการบ่งชี้ผู้พูดแบบขึ้นกับบทคำพูด และอีกครึ่งหนึ่ง 180 ตัวอย่างนำไปใช้ในการทดสอบระบบการบ่งชี้ผู้พูดแบบขึ้นกับบทคำพูด

3.3 การสร้างและรูปลักษณ์ของระบบการบ่งชี้ผู้พูดแบบขึ้นกับบทคำพูด

ในงานวิจัยนี้ได้พัฒนาระบบการบ่งชี้ผู้พูดแบบขึ้นกับบทคำพูดโดยใช้แบบจำลองฮิดเดนมาร์คอฟแบบดิสครีตดังแสดงในรูปที่ 3.1 สามารถแบ่งออกเป็น 2 ขั้นตอนใหญ่คือ ขั้นตอนการฝึกฝนกับขั้นตอนการทดสอบ ซึ่งจะกล่าวถึงขั้นตอนทั้งสองในหัวข้อต่อไป ส่วนรายละเอียดขั้นตอนทั้งหมดของระบบการบ่งชี้ผู้พูดแบบขึ้นกับบทคำพูดมีดังต่อไปนี้



รูปที่ 3.1 ระบบการปงชี้ผู้พูดแบบขึ้นกับบทคำพูดโดยใช้แบบจำลองฮิดเดนมาร์คอฟแบบดิสครีต

3.3.1 การประมวลผลสัญญาณเบื้องต้น (Signal Preprocessing)

การประมวลผลสัญญาณเบื้องต้นเป็นขั้นตอนในการเตรียมสัญญาณเสียงพูด (Speech Signal) ที่บันทึกได้มาเป็นข้อมูลเสียงพูด (Speech Data) ดังในรูปที่ 2.1 โดยผ่านการรมวิธีสองขั้นตอนคือ

กรรมวิธีเน้นล่วงหน้า (Preemphasis)

กรรมวิธีเน้นล่วงหน้าคือการปรับระดับค่าอัตราส่วนสัญญาณเสียงต่อสัญญาณรบกวน (Signal to Noise Ratio, S/N) ให้เพิ่มสูงขึ้น โดยนำสัญญาณเสียงพูดผ่านวงจรกรอง FIR อันดับหนึ่งที่มีฟังก์ชันถ่ายโอนดังสมการที่ (2.1) และ (2.2) ในงานวิจัยนี้ได้กำหนดให้ค่าสัมประสิทธิ์ของวงจรกรอง $a = 0.95$ (วิศรุต อาชุบุตร, 2539)

กรรมวิธีวางกรอบขนาดสัญญาณ (Smoothing Window)

เนื่องจากสัญญาณเสียงพูดมีการเปลี่ยนแปลงค่าทางสถิติตามเวลา (Non-Stationary) เป็นสัญญาณสุ่มที่ไม่เป็นเออร์กอดิก (Non-Ergodic) และไม่เป็นสัญญาณฟันสุ่ม (Non-Stochastic) ดังนั้นการประยุกต์ใช้สัญญาณเสียงพูดจึงจำเป็นต้องแบ่งสัญญาณเสียงพูดออกเป็นส่วนย่อยๆ (Rabiner and Levinson, 1981) เรียกว่า "กรอบ" (Frame) ที่มีความยาวประมาณ 10 ถึง 30 มิลลิวินาที (O'Shaughnessy, 1988) ในงานวิจัยนี้ได้แบ่งสัญญาณเสียงพูดด้วยการคูณสัญญาณเสียงพูดกับฟังก์ชันกรอบชนิด Hamming Window ดังสมการที่ (2.3) และ (2.4) เพื่อเป็นการลดทอนแอมปลิจูดที่บริเวณปลายแต่ละข้างของกรอบข้อมูลเสียงพูดอย่างช้าๆ และหลีกเลี่ยงความไม่ต่อเนื่อง

จากรูปที่ 2.2 กำหนดให้จำนวนของข้อมูลในแต่ละกรอบสัญญาณเสียงพูด (N) มีค่าเท่ากับ 512 ช่วงประสิทธิผลของฟังก์ชันกรอบมีค่าเท่ากับ 256 และเลื่อนฟังก์ชันกรอบครั้งละ 64 (เอกฤทธิ์ มณีน้อย, 2541) จึงทำให้แต่ละกรอบมีข้อมูลของสัญญาณเสียงพูดเท่ากับ 256 ค่า

3.3.2 การสกัดลักษณะสำคัญ (Feature Extraction)

การสกัดลักษณะสำคัญคือขั้นตอนในการหาลักษณะสำคัญของสัญญาณเสียงพูดจากข้อมูลเสียงพูดที่ได้จากการประมวลผลสัญญาณเบื้องต้นแล้ว ลักษณะสำคัญที่ใช้ในงานวิจัยนี้ได้แก่ สัมประสิทธิ์การประมาณพันธะเชิงเส้น สัมประสิทธิ์เซปสตรอล และสัมประสิทธิ์เซปสตรอลบนความถี่เมล โดยมีรายละเอียดการคำนวณหาดังนี้

สัมประสิทธิ์การประมาณพันธะเชิงเส้น

สัมประสิทธิ์ของการประมาณพันธะเชิงเส้นเป็นแบบจำลองการประมาณการสั่นสะเทือนของช่องทางเดินเสียงขณะออกเสียงพูด รูปร่างของช่องทางเดินเสียง ความถี่และแบนด์วิดท์ของเรโซแนนซ์ที่ช่องทางเดินเสียง สามารถหาได้จากการจำลองช่องทางเดินเสียงของมนุษย์โดยใช้แบบจำลอง All Pole และใช้วิธีอัตสหสัมพันธ์ของวิธีการ Least-squares และขั้นตอนวิธีการวนซ้ำของ Levinson-Durbin ในการคำนวณหาค่าสัมประสิทธิ์ a_k ดังที่กล่าวไว้ในหัวข้อที่ 2.2.1

สัมประสิทธิ์เซปสตรอล

สัมประสิทธิ์เซปสตรอลเป็นการวิเคราะห์ลักษณะสำคัญที่มีรากฐานอยู่บนเอนเวโลปของสเปกตรัม (Spectral Envelope) ในงานวิจัยนี้ได้คำนวณหาสัมประสิทธิ์เซปสตรอลจากค่าสัมประสิทธิ์การประมาณพันธะเชิงเส้นดังสมการที่ (2.26) และ (2.27) เนื่องจากว่าวิธีนี้ใช้เวลาในการคำนวณน้อยกว่าและเพื่อต้องการเปรียบเทียบกับผลที่ได้กับสัมประสิทธิ์การประมาณพันธะเชิงเส้น

สัมประสิทธิ์เซปสตรอลบนความถี่เมล

สัมประสิทธิ์เซปสตรอลบนความถี่เมลเป็นลักษณะสำคัญที่เลียนแบบการได้ยินของมนุษย์บน Mel Scale ดังแสดงในรูปที่ 2.4 สามารถแบ่งได้เป็น 2 ช่วงคือช่วงที่เป็นเชิงเส้นกับช่วงที่ไม่เป็นเชิงเส้น การคำนวณหาค่าสัมประสิทธิ์เซปสตรอลบนความถี่เมลหาได้จากการแปลงฟูริเยร์อย่างเร็ว (Fast Fourier Transform) ของข้อมูลเสียงแล้วนำไปเข้าชุดวงจรกรองรูปสามเหลี่ยมดังรูปที่ 2.5 ในที่นี้ได้กำหนดให้อันดับของการวิเคราะห์ (N) มีค่าเท่ากับ 20 คือมีชุดวงจรกรองรูปสามเหลี่ยมจำนวนเท่ากับ 20 วงจร แล้วคำนวณหาค่าลอการิทึมของพลังงานที่เอาต์พุตของแต่ละวงจรได้เป็น X_k เมื่อ $k = 1,2,\ldots,N$ และนำ X_k ไปคำนวณหาสัมประสิทธิ์เซปสตรอลบนความถี่เมลตามสมการที่ (2.31)

3.3.3 การฝึกฝนชุดรหัส (Codebook Training)

การฝึกฝนชุดรหัสเป็นขั้นตอนในการฝึกฝนเพื่อให้ได้ต้นแบบอ้างอิงชุดรหัส (Codebook Reference Template) เทคนิคในการฝึกฝนชุดรหัสมีอยู่ 2 แบบใหญ่ๆ คือ การฝึกฝนแบบหลายชุดรหัส (Multiple Codebook) กับการฝึกฝนแบบชุดรหัสเดี่ยว (Single Codebook) ซึ่งมีความแตกต่างกันที่จำนวนของชุดรหัสที่ใช้เป็นต้นแบบอ้างอิงชุดรหัส กล่าวคือถ้าเป็นแบบหลายชุดรหัสจะมีการฝึกฝนให้ได้

ต้นแบบอ้างอิงชุดรหัสมีจำนวนเท่ากับจำนวนผู้พูดที่ต้องการบ่งชี้ โดยใช้ข้อมูลเสียงพูดที่อยู่ในชุดฝึกฝนของผู้พูดนั้นๆ ส่วนการฝึกฝนแบบชุดรหัสเดี่ยวจะฝึกฝนข้อมูลเสียงพูดที่อยู่ในชุดฝึกฝนทั้งหมดให้ได้เป็นต้นแบบอ้างอิงชุดรหัสหนึ่งชุดเท่านั้น เห็นได้ว่าการฝึกฝนชุดรหัสทั้งสองแบบมีข้อดีและข้อเสียที่แตกต่างกัน กล่าวคือชุดรหัสแบบหลายชุดสามารถนำไปประยุกต์การใช้งานได้มากกว่า เช่น สามารถเพิ่มหรือลดจำนวนผู้พูดที่ใช้ในการบ่งชี้ได้ง่ายกว่าเพราะว่าไม่ต้องฝึกฝนข้อมูลเสียงทั้งหมดใหม่เหมือนชุดรหัสเดี่ยว จึงทำให้ประหยัดเวลาและมีความยืดหยุ่นในการพัฒนาระบบ แต่ถ้าพิจารณาในเรื่องการคำนวณที่เกิดขึ้นในระบบพบว่าชุดรหัสหลายชุดมีการคำนวณที่ซับซ้อนกว่าชุดรหัสเดี่ยวคือมีการเทียบข้อมูลเสียงพูดกับต้นแบบอ้างอิงชุดรหัสแต่ละชุดทำให้ใช้เวลาตามจำนวนผู้พูดที่เพิ่มมากขึ้นซึ่งมากกว่าชุดรหัสเดี่ยว ดังนั้นเวลาโดยรวมที่ชุดรหัสเดี่ยวใช้น้อยกว่าชุดรหัสหลายชุด

ในงานวิจัยนี้ได้ฝึกฝนชุดรหัสเพื่อให้ได้ต้นแบบอ้างอิงชุดรหัสเป็นแบบชุดรหัสเดี่ยว เนื่องจากระบบใช้เวลาในการประมวลผลน้อยกว่า และจำนวนผู้พูดที่ใช้ในการบ่งชี้มีจำนวนไม่มากเพียงพอที่จำเป็นต้องใช้ชุดรหัสต้นแบบอ้างอิงเป็นแบบชุดรหัสหลายชุด ส่วนขั้นตอนวิธีการสร้างชุดรหัสและฝึกฝนชุดรหัสกล่าวในหัวข้อ 3.4.1

3.3.4 การควอนไทซ์แบบเวกเตอร์ (Vector Quantization)

การควอนไทซ์แบบเวกเตอร์เป็นการลดขนาดของข้อมูลเสียงพูดโดยให้มีความผิดเพี้ยนน้อยที่สุดอีกวิธีหนึ่ง โดยการแทนเวกเตอร์ของสัมประสิทธิ์ต่างๆ ที่มีมิติตามขนาดของปริภูมิให้เหลือเพียงค่าจำนวนเต็มค่าหนึ่ง ซึ่งอ้างอิงมาจากต้นแบบอ้างอิงชุดรหัสที่ได้สร้างและฝึกฝนในช่วงการฝึกฝนต้นแบบอ้างอิงชุดรหัสก่อนหน้านี้ ผลที่ได้จากการทำควอนไทซ์แบบเวกเตอร์เป็นลำดับของค่าจำนวนเต็มเรียงกันตามจำนวนกรอบของข้อมูลเสียงพูด สำหรับลำดับค่าจำนวนเต็มที่ได้นี้ในแบบจำลองฮิดเดนมาร์คอฟเรียกว่า "ลำดับการสังเกต" (Observation Sequence)

เหตุผลสำคัญที่จำเป็นต้องควอนไทซ์แบบเวกเตอร์เนื่องจากว่าแบบจำลองฮิดเดนมาร์คอฟที่ใช้ในงานวิจัยนี้เป็นแบบจำลองฮิดเดนมาร์คอฟแบบดิสครีตจึงทำให้ลำดับการสังเกตที่ใช้ในแบบจำลองจะต้องเป็นลำดับของจำนวนเต็มขนาด 1 มิติเท่านั้น (Rabiner, 1989)

3.3.5 การฝึกฝนพารามิเตอร์ของแบบจำลองฮิดเดนมาร์คอฟ (HMM Parameter Training)

การฝึกฝนพารามิเตอร์ของแบบจำลองฮิดเดนมาร์คอฟเป็นขั้นตอนในการฝึกฝนพารามิเตอร์ของแบบจำลองฮิดเดนมาร์คอฟ $\lambda = (A, B, \pi)$ เพื่อสร้างเป็นต้นแบบอ้างอิงเอกลักษณ์ (Identified Reference Template) โดยใช้ลำดับการสังเกตที่ได้จากการทำควอนไทซ์แบบเวกเตอร์ด้วยการปรับค่าพารามิเตอร์ของแบบจำลองฮิดเดนมาร์คอฟตามลำดับการสังเกต ที่ได้จากข้อมูลเสียงพูดในชุดฝึกฝนของผู้พูดนั้นๆ ในแต่ละประโยค ซึ่งขั้นตอนต่างๆ กล่าวไว้โดยละเอียดในหัวข้อ 3.4.2

3.3.6 การเข้าคู่พารามิเตอร์ของแบบจำลองฮิดเดนมาร์คอฟ (HMM Parameter Matching)

การเข้าคู่พารามิเตอร์ของแบบจำลองฮิดเดนมาร์คอฟเป็นการเข้าคู่ลำดับการสังเกตที่ได้จากสัญญาณเสียงพูดกับต้นแบบอ้างอิงเอกลักษณ์ในการบ่งชี้ เพื่อตัดสินใจว่าลำดับการสังเกตที่กำลังพิจารณามีค่าความน่าจะเป็นของพารามิเตอร์ของแบบจำลองฮิดเดนมาร์คอฟที่อยู่ในต้นแบบอ้างอิงเอกลักษณ์ชุดใดมีค่ามากที่สุดก็จะเลือกพารามิเตอร์ของแบบจำลองฮิดเดนมาร์คอฟชุดนั้น ทำให้รู้ว่าข้อมูลเสียงพูดของลำดับการสังเกตนั้นเป็นเสียงพูดของบุคคลใดที่อยู่ในกลุ่มของผู้พูดที่ได้รับการฝึกฝนแล้ว จากนั้นจึงคำนวณหาอัตราการบ่งชี้ผู้พูด (Identification Rate) จากจำนวนข้อมูลเสียงพูดที่ใช้ในการทดสอบในแต่ละประโยค สำหรับขั้นตอนในการทดสอบทั้งหมดกล่าวโดยละเอียดในหัวข้อ 3.5

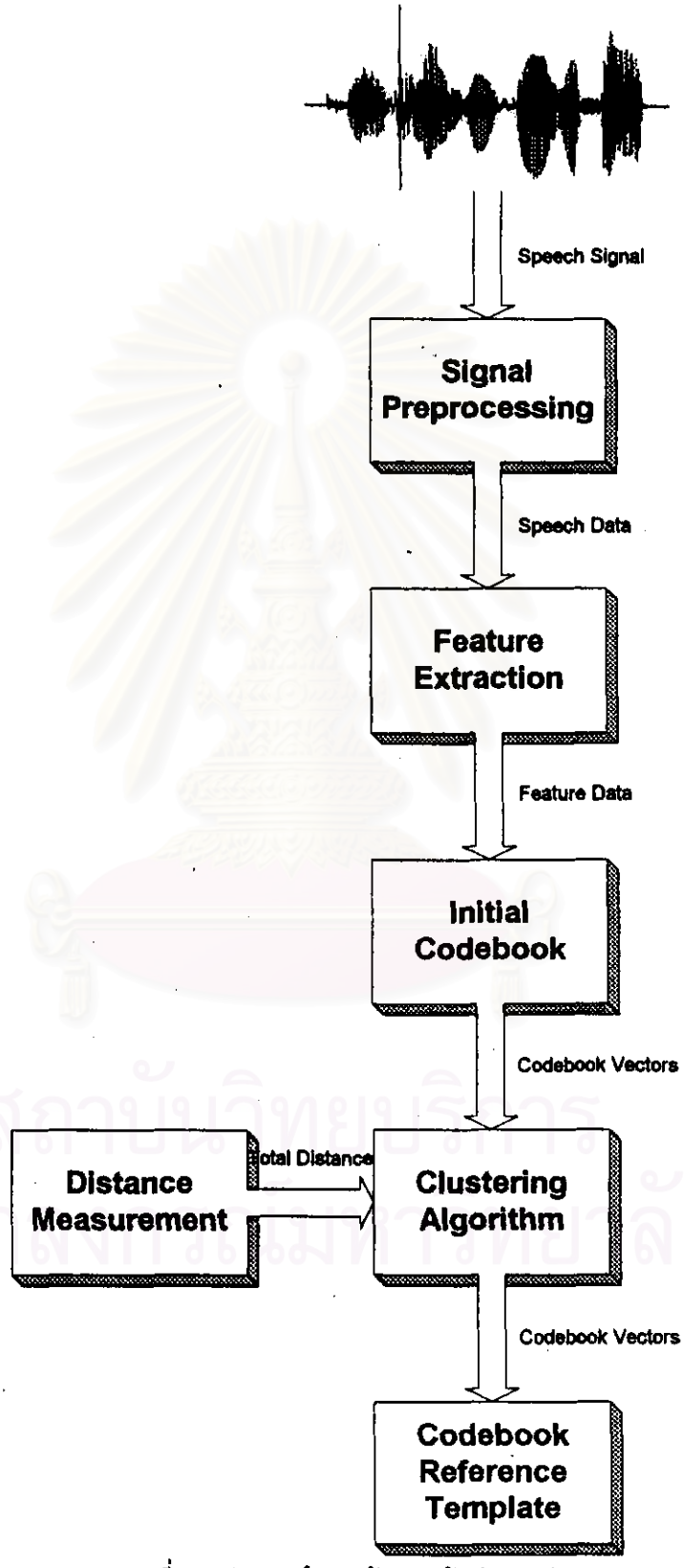
3.4 ขั้นตอนการฝึกฝนระบบการบ่งชี้ผู้พูดแบบขึ้นกับบทคำพูด

ขั้นตอนการฝึกฝนระบบการบ่งชี้ผู้พูดแบบขึ้นกับบทคำพูดเป็นขั้นตอนการฝึกฝนระบบการบ่งชี้ผู้พูดจากข้อมูลเสียงพูดที่อยู่ในชุดฝึกฝนโดยฝึกฝนข้อมูลเสียงพูดครั้งละประโยค จากรูปที่ 3.1 เห็นได้ว่าขั้นตอนในการฝึกฝนระบบมีทั้งหมด 2 ช่วง คือช่วงการฝึกฝนต้นแบบอ้างอิงชุดรหัสกับช่วงการฝึกฝนพารามิเตอร์ของแบบจำลองฮิดเดนมาร์คอฟ ในแต่ละช่วงการฝึกฝนจะใช้ข้อมูลเสียงพูดที่ได้จากการประมวลผลสัญญาณเบื้องต้นและการสกัดลักษณะสำคัญ ซึ่งได้กล่าวรายละเอียดไปแล้ว

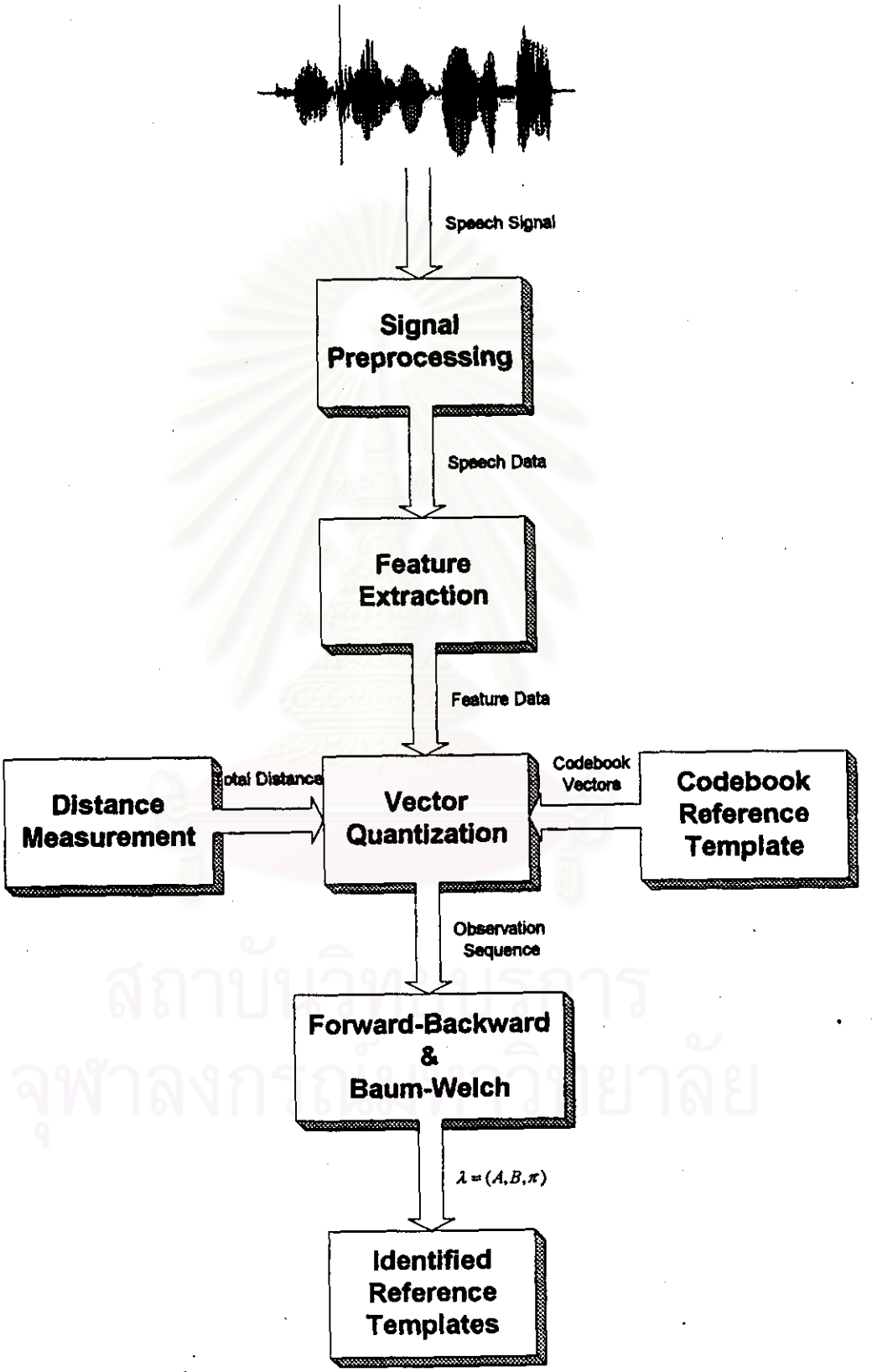
3.4.1 ช่วงการฝึกฝนชุดรหัสต้นแบบอ้างอิง

ช่วงการฝึกฝนชุดรหัสต้นแบบอ้างอิงเป็นช่วงในการสร้างและฝึกฝนต้นแบบอ้างอิงชุดรหัสดังแสดงในรูปที่ 3.2 โดยใช้ขั้นตอนวิธีการแบ่งกลุ่ม (Clustering Algorithm) เพื่อแบ่งข้อมูลลักษณะสำคัญที่อยู่ในปริภูมิ N มิติ ในงานวิจัยนี้ได้ใช้ขั้นตอนวิธีการแบ่งกลุ่ม 2 วิธีเพื่อเปรียบเทียบกันได้แก่ ขั้นตอนวิธีการแบ่งเฉลี่ย K ส่วนและขั้นตอนวิธีการหาค่าคาดหวังที่มากที่สุด ซึ่งรายละเอียดได้กล่าวในหัวข้อ 2.3.1 เพราะฉะนั้นจากรูปที่ 3.2 ทั้งสองขั้นตอนวิธีการแตกต่างที่การกำหนดชุดรหัสเริ่มต้น (Initial Codebook) ขั้นตอนวิธีการแบ่งกลุ่มและการวัดระยะห่าง (Distance Measurement) โดยที่ขั้นตอนวิธีการแบ่งเฉลี่ย K ส่วนจะกำหนดชุดรหัสเริ่มต้นด้วยการสุ่ม (Random Codebook) และวัดหาระยะห่างยูคลิเดียน (Euclidean Distance) ดังสมการที่ (2.34) ส่วนขั้นตอนวิธีการหาค่าคาดหวังที่มากที่สุดจะกำหนดชุดรหัสเริ่มต้นด้วยขั้นตอนวิธีการสุนัขกับกระต่าย (Dog Rabbit Algorithm) และวัดหาระยะห่างมาฮาลาโนบิส (Mahalanobis Distance) ดังสมการที่ (2.35)

การฝึกฝนชุดรหัสต้นแบบอ้างอิงนั้นจะใช้ข้อมูลเสียงพูดที่อยู่ในชุดฝึกฝน (Training Set) และฝึกฝนครั้งละประโยค เนื่องจากใช้การฝึกฝนแบบชุดรหัสเดี่ยว ทำให้มีจำนวนของต้นแบบอ้างอิงชุดรหัสเท่ากับ 3 ชุดตามจำนวนของประโยคคำพูดที่ใช้ในการบ่งชี้



รูปที่ 3.2 ช่วงการฝึกฝนต้นแบบอ้างอิงชุดรหัส



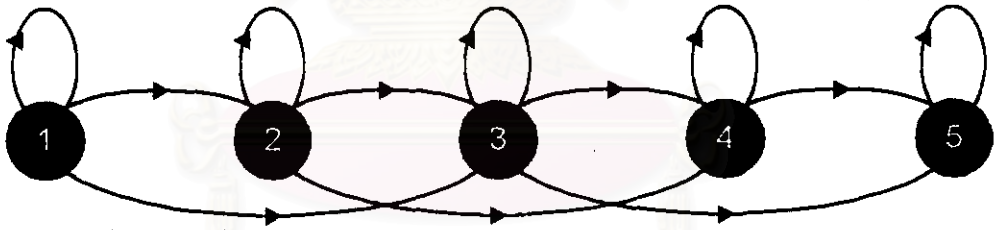
รูปที่ 3.3 ช่วงการฝึกฝนพารามิเตอร์ของแบบจำลองฮิดเดนมาร์คอฟ

3.4.2 ช่วงการฝึกฝนพารามิเตอร์ของแบบจำลองฮิดเดนมาร์คอฟ

ช่วงการฝึกฝนพารามิเตอร์ของแบบจำลองฮิดเดนมาร์คอฟเป็นช่วงในการฝึกฝนพารามิเตอร์ของแบบจำลองฮิดเดนมาร์คอฟ $\lambda = (A, B, \pi)$ จากลำดับการสังเกตของข้อมูลสัญญาณเสียงที่อยู่ในชุดฝึกฝนเพื่อให้ได้เป็นต้นแบบอ้างอิงเอกลักษณ์ การฝึกฝนพารามิเตอร์ของแบบจำลองฮิดเดนมาร์คอฟคือการแก้ปัญหาพื้นฐานข้อที่ 1 และข้อที่ 2 โดยใช้ขั้นตอนวิธีการไปหน้าและย้อนกลับกับขั้นตอนวิธีการประมาณค่าซ้ำของ Buam-Welch ดังที่กล่าวไว้ในหัวข้อ 2.4.2 และ 2.4.3

สำหรับการฝึกฝนพารามิเตอร์ของแบบจำลองฮิดเดนมาร์คอฟในงานวิจัยนี้ ใช้ข้อมูลเสียงพูดที่ผ่านการควอนไทซ์แบบเวกเตอร์เมื่อเทียบกับต้นแบบอ้างอิงชุดรหัสได้เป็นลำดับการสังเกต จากนั้นปรับค่าพารามิเตอร์ของแบบจำลองฮิดเดนมาร์คอฟตามลำดับการสังเกตที่เข้ามาจนได้เป็นพารามิเตอร์ของแบบจำลองฮิดเดนมาร์คอฟต้นแบบอ้างอิง (HMM Parameter Reference Templates) หรือต้นแบบอ้างอิงเอกลักษณ์นั้นเอง ในแต่ละประโยคจะมีจำนวนของต้นแบบอ้างอิงเอกลักษณ์เท่ากับจำนวนของผู้พูดที่จะบ่งชี้ คือมีจำนวนของต้นแบบอ้างอิงเอกลักษณ์เท่ากับ 12 ชุดต่อประโยค

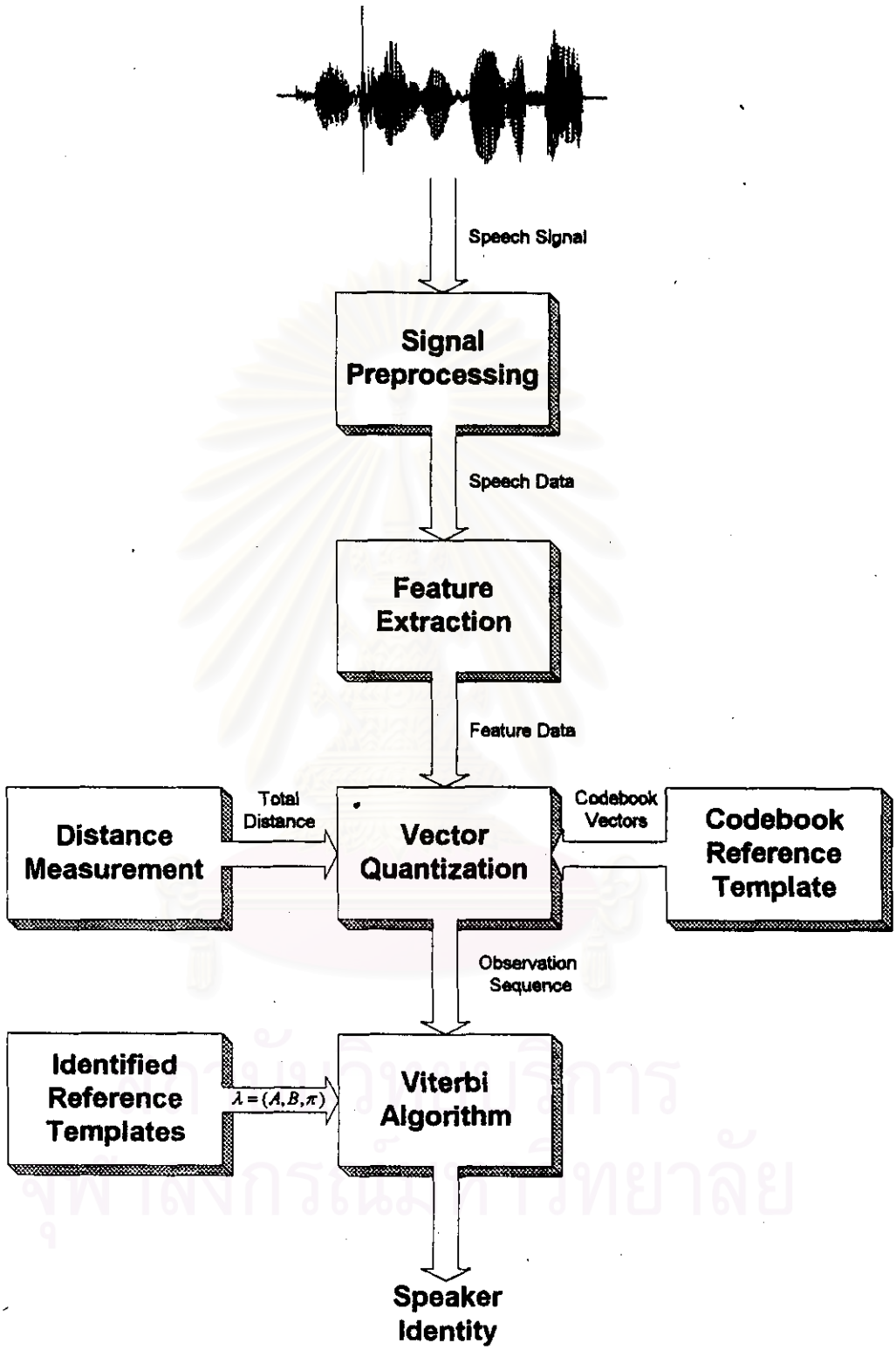
ประเภทของแบบจำลองฮิดเดนมาร์คอฟที่ใช้ในงานวิจัยนี้เป็นแบบซ้ายไปขวา (Left-Right) ดังแสดงในรูปที่ 3.4 เนื่องจากเป็นชนิดของแบบจำลองฮิดเดนมาร์คอฟที่เหมาะสมกับสัญญาณเสียงพูด (Rabiner L.R., 1989) ส่วนจำนวนสถานะของแบบจำลองเป็นค่าพารามิเตอร์ที่ได้ปรับเปลี่ยนเพื่อศึกษาผลที่ได้ ซึ่งจะกล่าวในบทที่ 4 ต่อไป



รูปที่ 3.4 แบบจำลองฮิดเดนมาร์คอฟ 5 สถานะชนิดซ้ายไปขวา

3.5 ขั้นตอนการทดสอบระบบการบ่งชี้ผู้พูดแบบขึ้นกับบทคำพูด

ขั้นตอนการทดสอบระบบการบ่งชี้ผู้พูดแบบขึ้นกับบทคำพูดเป็นกระบวนการทดสอบระบบการบ่งชี้ผู้พูดแบบขึ้นกับบทคำพูด โดยใช้ข้อมูลเสียงพูดที่อยู่ในชุดทดสอบมาเปรียบเทียบกับต้นแบบอ้างอิงเอกลักษณ์ของผู้พูดแต่ละบุคคลที่ได้จากขั้นตอนฝึกฝน สำหรับขั้นตอนในการทดสอบระบบคล้ายกับขั้นตอนในการฝึกฝนพารามิเตอร์ของแบบจำลองฮิดเดนมาร์คอฟโดยแตกต่างกันที่ขั้นตอนทดสอบจะหาความน่าจะเป็นของลำดับการสังเกตที่ได้จากข้อมูลเสียงพูดที่มีกับต้นแบบอ้างอิงเอกลักษณ์ และหาต้นแบบอ้างอิงเอกลักษณ์ที่มีความน่าจะเป็นมากที่สุด โดยใช้ขั้นตอนวิธีการ Viterbi ที่กล่าวไว้ในหัวข้อที่ 2.4.3 ดังแสดงในรูปที่ 3.5



รูปที่ 3.5 ช่วงการทดสอบระบบการบ่งชี้ผู้พูดแบบขึ้นกับบทคำพูด