บทที่ 3

ขั้นตอนวิธีในการดำเนินการวิจัย

รายละเอียดในบทที่ 3 นี้จะกล่าวถึงรายละเอียดของขั้นตอนวิธีในการดำเนินการวิจัยดังแสดงในรูปที่ 3.2 ถึง 3.5 (Rabiner and Levinson, 1981; Roe and Wilpon, 1993) ประกอบไปด้วยรายละเอียดเกี่ยวกับขั้นตอนการฝึกฝนระบบการรู้จำคำพูดและขั้นตอนการทดสอบระบบการรู้จำคำพูด พร้อมทั้งรายละเอียดประเภทของแบบจำลองฮิดเดน มาร์คอฟที่ประยุกต์ใช้งาน นอกจากนี้ยังมีรายละเอียดเกี่ยวกับการกำหนดวิธีการสร้างชุดคำศัพท์ภาษาไทยที่ใช้ในการฝึกฝนและการทดสอบ รวมทั้งรายละเอียดในการเก็บตัวอย่างเสียงพูดเพื่อนำมาใช้เป็นตัวอย่างเพื่อฝึกฝนระบบและเป็นตัวอย่างทดสอบ

การกำหนดวิธีการสร้างชุดคำศัพท์

ในการสร้างชุดคำศัพท์สำหรับระบบการรู้จำคำพูดของการวิจัยครั้งนี้ ได้จำกัดจำนวนคำศัพท์ที่ใช้ในการรู้จำ โดยการคัดเลือกมาจากคำศัพท์ทั่วไปที่ใช้ในชีวิตประจำวัน คำกริยา อวัยวะของร่างกาย และชื่อผลไม้ไทย ชุดคำศัพท์จำนวน 70 คำแบ่งออกเป็น 4 ชุด ได้แก่ชุดคำศัพท์พยางค์เดียว ชุดคำศัพท์สองพยางค์ ชุดคำศัพท์สามพยางค์ชุดละ 20 คำ และชุดคำศัพท์ตัวเลขศูนย์ถึงเก้าจำนวน 10 คำตามลำดับ ดังมีรายละเอียดแสดงในภาคผนวก ก

การเก็บตัวอย่างข้อมูลเสียงพูด

การเก็บตัวอย่างข้อมูลเสียงพูดจะอาศัยการเก็บบันทึกข้อมูลไว้ในเครื่องคอมพิวเตอร์ โดยทำการบันทึกเสียง ณ ห้องปฏิบัติการวิจัยประมวลผลสัญญาณดิจิตอล ห้อง 303 ชั้น 3 ตึกวิศวกรรมไฟฟ้า ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ซึ่งได้รับการควบคุมสภาพแวดล้อมขณะทำการบันทึกเสียงให้คล้ายคลึงกับสภาพแวดล้อมของสถานที่ทำงานทั่วไปและมีเสียงรบกวนน้อยที่สุด โดยเสียงพูดที่บันทึกไว้จะจัดเก็บด้วยตัวอย่างขนาด 16 บิตและมีอัตราการชักตัวอย่าง 11 KHz เนื่องจากเสียงพูดของมนุษย์จะอยู่ในช่วงไม่เกิน 5 KHz จึงต้องใช้อัตราการชักตัวอย่างที่สูงกว่าสองเท่าของความถี่ 5 KHz

1. กฎเกณฑ์ในการคัดเลือกผู้บอกภาษา

คุณสมบัติของผู้บอกภาษาที่จะได้รับการบันทึกเสียง ต้องเป็นไปตามกฎเกณฑ์ดังนี้

1) เป็นผู้ที่ใช้ภาษาไทยกรุงเทพเป็นภาษาพูดทั่วไปและมีอายุระหว่าง 18 - 25 ปี
2) เป็นผู้ที่มีการออกเสียงเป็นปกติและตรงตามหลักการออกเสียงพูดภาษาไทย

2. อุปกรณ์เครื่องมือที่ใช้ในการบันทึกเสียง

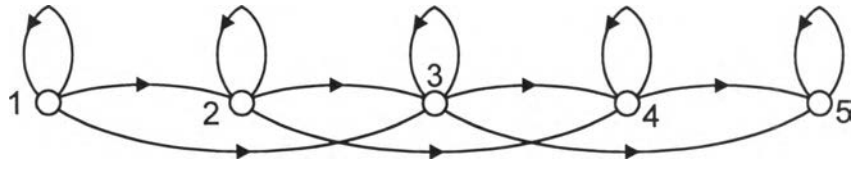
รายละเอียดอุปกรณ์ที่ใช้ในการเก็บบันทึกเสียงดังนี้

1) เครื่องคอมพิวเตอร์ 80486DX2-66 พร้อมหน่วยความจำขนาด 16 MB พร้อมด้วย Harddisk ขนาด 1.2 GB และ Floppy Disk 3.5" ขนาด 1.44 MB
2) การ์ดเสียง Sound Blaster 16 ของบริษัท Creative Technology
3) ไมโครโฟน Philips Uni-directional Dynamic Microphone รุ่น SBC 465
4) ระบบปฏิบัติการ MS-DOS Version 6.22 ภาษาไทย
5) ระบบปฏิบัติการ Microsoft Windows for Workgroup Version 3.11 ภาษาไทย
6) โปรแกรม Creative WaveStudio for Windows Version 2.0

ในงานวิจัยนี้ทำการเก็บตัวอย่างเสียงพูดจำนวน 60 คน แบ่งเป็นชาย 50 คนและหญิง 10 คน โดยแบ่งออกเป็น 2 ชุด ได้แก่ชุดเสียงพูดเพื่อฝึกฝนระบบ (Training Set) และชุดเสียงพูดเพื่อทดสอบระบบ (Test Set) สำหรับชุดเสียงเพื่อฝึกฝนระบบมีจำนวน 40 คนแบ่งเป็นเสียงผู้ชาย 34 คนและเสียงผู้หญิง 6 คน ชุดเสียงเพื่อทดสอบระบบมีจำนวน 20 คนแบ่งเป็นเสียงผู้ชาย 16 เสียงและเสียงผู้หญิง 4 เสียง

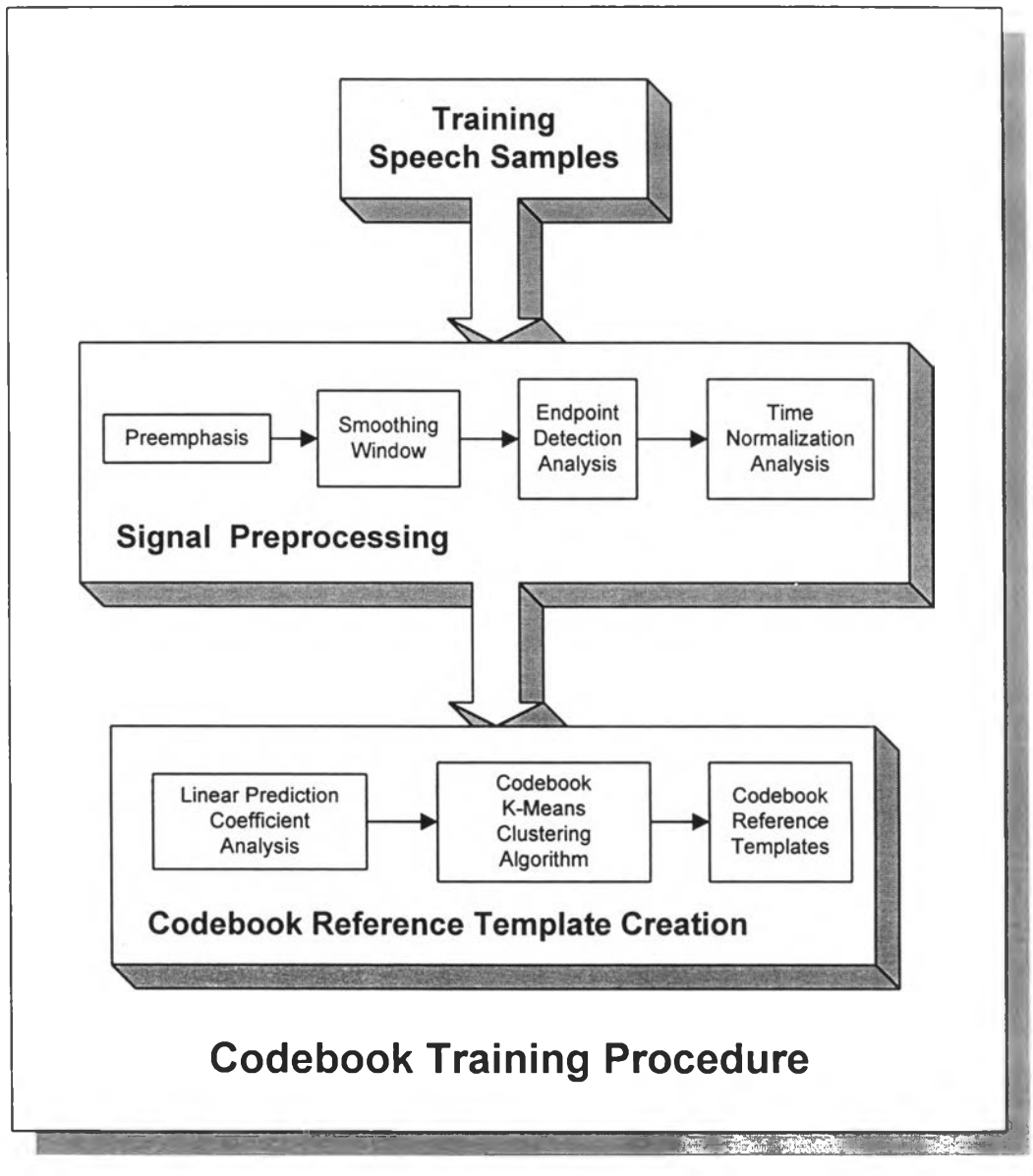
ประเภทของแบบจำลองฮิดเดน มาร์คอฟ

แบบจำลองฮิดเดน มาร์คอฟที่ใช้สำหรับงานวิจัยนี้ เป็นประเภทแบบจำลองซ้าย-ขวา (Left-Right Model) หรือแบบจำลองอนุกรม (Serial Model) ที่มี 15 สถานะซึ่งมีการเชื่อมโยงระหว่างสถานะดังมีตัวอย่างแสดงในรูปที่ 3.1 เนื่องจากจำนวนสถานะที่เลือกใช้มีความเพียงพอสำหรับใช้กับคำต่อเนื่อง และการเชื่อมโยงระหว่างสถานะไม่จำเป็นต้องเชื่อมโยงกันทั้งหมด ดังนั้นแบบจำลองที่ใช้จึงมีเพียงการเชื่อมโยงในสถานะ การเชื่อมโยงระหว่างสถานะ และการเชื่อมโยงข้ามสถานะเท่านั้นดังรูป (Rabiner and Levinson, 1985; Bahl, Brown, Souza, Mercer, and Picheny, 1988; Lee, Hon, and Reddy, 1990; เสาวลักษณ์ อารีย์พงศา, 2538)

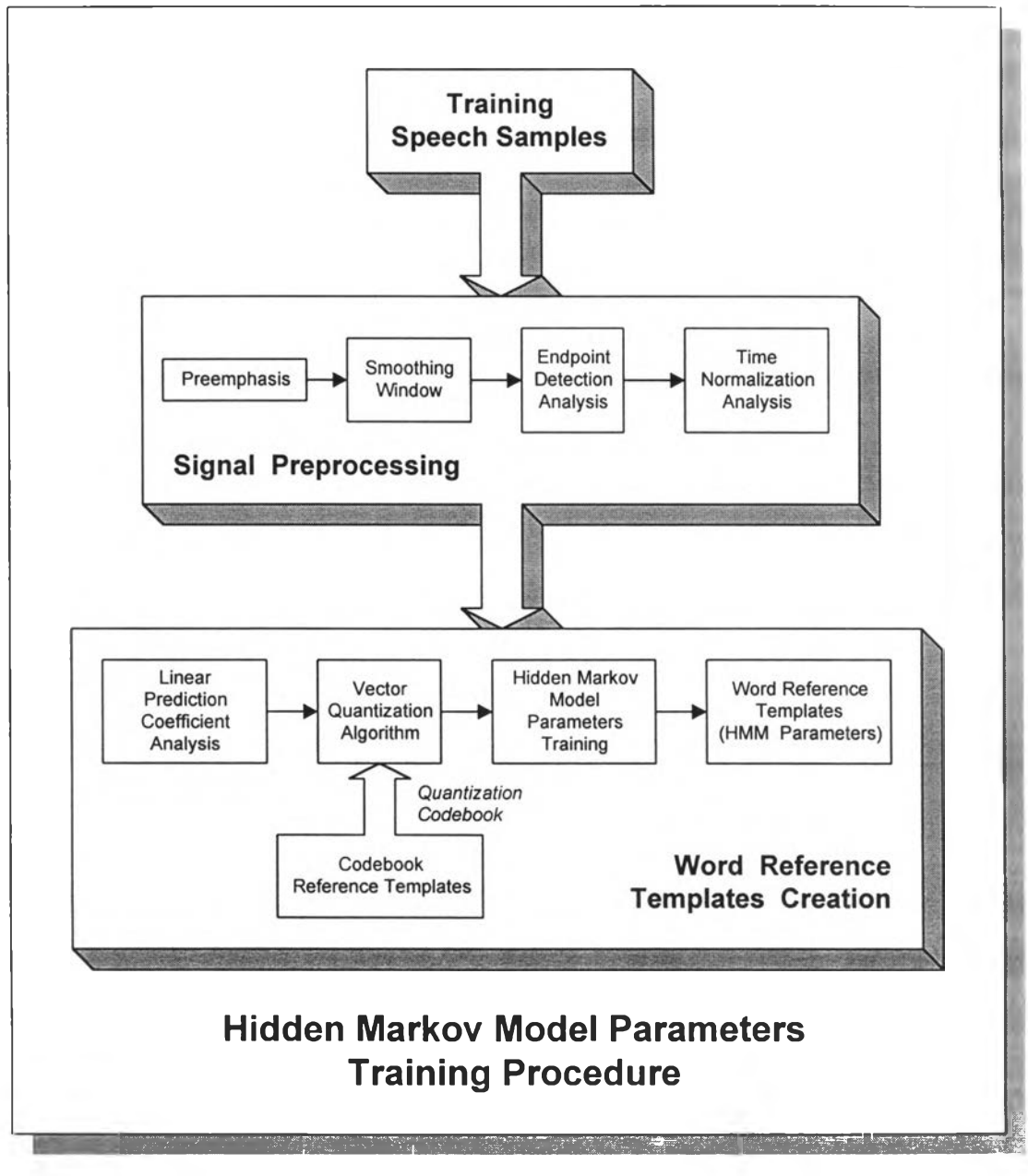


รูปที่ 3.1 แบบจำลองแบบซ้าย-ขวาที่มี 5 สถานะ

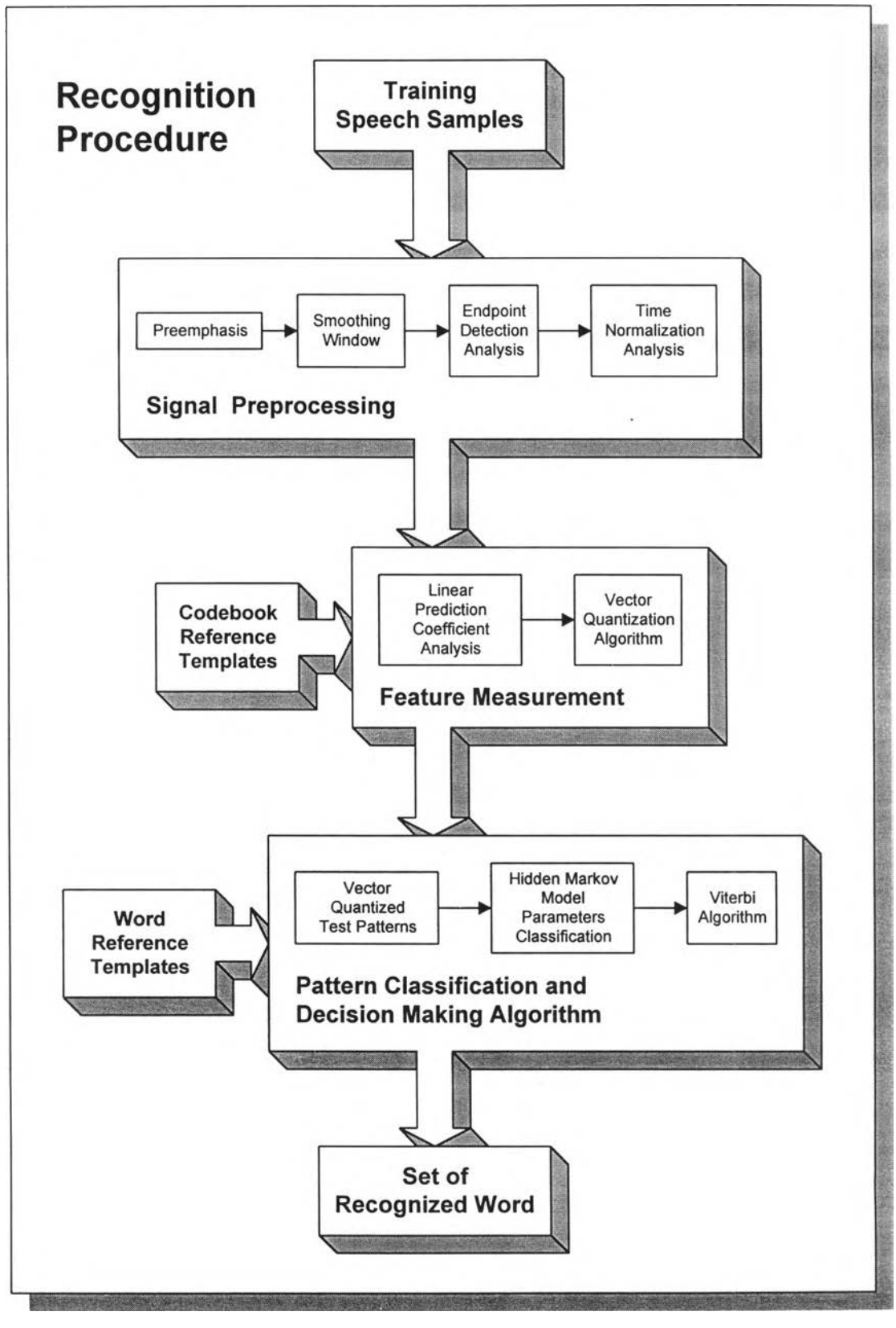
รูปที่ 3.2 ภาพรวมของแบบจำลองการรู้จำเสียงพูด



รูปที่ 3.3 รายละเอียดขั้นตอนการสร้างและฝึกฝนชุดรหัส



รูปที่ 3.4 รายละเอียดขั้นตอนการสร้างและฝึกฝนชุดพารามิเตอร์ของแบบจำลองฮิดเดน มาร์คอฟ



รูปที่ 3.5 รายละเอียดขั้นตอนการรู้จำคำไทยหลายพยางค์

รายละเอียดขั้นตอนในการรู้จำคำพูด

รายละเอียดของขั้นตอนในการรู้จำคำพูดแบ่งออกได้เป็น 2 ขั้นตอน ได้แก่ขั้นตอนการฝึกฝนระบบการรู้จำคำพูด และขั้นตอนการทดสอบระบบการรู้จำคำพูด โดยขั้นตอนการฝึกฝนยังแบ่งออกได้เป็น 2 ขั้นตอน ได้แก่ขั้นตอนการสร้างและฝึกฝนชุดรหัส (Codebook Training Procedure) และขั้นตอนการสร้างและฝึกฝนชุดพารามิเตอร์ของแบบจำลองฮิดเดน มาร์คอฟ (Hidden Markov Model Parameters Training) ดังแสดงในรูปที่ 3.3 และ 3.4 ส่วนขั้นตอนการทดสอบระบบการรู้จำคำพูดดังแสดงในรูปที่ 3.5 ตามลำดับ

1. ขั้นตอนการฝึกฝนระบบการรู้จำคำพูด

ขั้นตอนการฝึกฝนระบบการรู้จำคำพูดนี้ จัดเป็นขั้นตอนในการสร้างชุดรหัสและชุดพารามิเตอร์ของแบบจำลองฮิดเดน มาร์คอฟ เพื่อใช้ในการควอนไทซ์แบบเวกเตอร์และการรู้จำคำพูดตามลำดับ โดยมีรายละเอียดของขั้นตอนการฝึกฝนระบบการรู้จำคำพูดแบ่งออกได้เป็น 2 ขั้นตอน ได้แก่ขั้นตอนการสร้างและฝึกฝนชุดรหัส และขั้นตอนการสร้างและฝึกฝนชุดพารามิเตอร์ของแบบจำลองฮิดเดน มาร์คอฟ ดังแสดงในรูปที่ 3.2 และ 3.3 ตามลำดับ ซึ่งมีรายละเอียดในแต่ละขั้นตอนดังนี้

1.1 การสร้างและฝึกฝนชุดรหัส (Codebook Training Procedure)

ขั้นตอนการสร้างและฝึกฝนชุดรหัส จัดเป็นขั้นตอนในการสร้างชุดรหัสด้วยขั้นตอนวิธีการแบ่งเฉลี่ย K ส่วนของข้อมูลเสียงพูดเพื่อใช้ในการควอนไทซ์แบบเวกเตอร์ดังแสดงในรูปที่ 3.2 ซึ่งประกอบไปด้วย 2 ขั้นตอนได้แก่ ขั้นตอนการประมวลผลสัญญาณเบื้องต้น (Signal Preprocessing) และขั้นตอนการสร้างชุดรหัสอ้างอิง (Codebook Reference Template Creation) โดยมีรายละเอียดในแต่ละขั้นตอนดังนี้

1.1.1 ขั้นตอนการประมวลผลสัญญาณเบื้องต้น (Signal Preprocessing)

ขั้นตอนการประมวลผลสัญญาณเบื้องต้นเป็นกรรมวิธีในการจัดเตรียมข้อมูลจากข้อมูลเสียงที่ได้จากการบันทึกเสียงซึ่งเป็นข้อมูลดิบ นำมาผ่านกรรมวิธีประมวลผลสัญญาณเชิงเลขโดยแบ่งออกได้เป็น 4 กรรมวิธีย่อยได้แก่ กรรมวิธีเน้นล่วงหน้า (Preemphasis) กรรมวิธีวางกรอบขนาดสัญญาณ (Smoothing Window) กรรมวิธีหาจุดสิ้นสุดเสียงพูดพร้อมทั้งจุดสิ้นสุดพยางค์ (Endpoint Detection) และกรรมวิธีปรับบรรทัดฐานเชิงเวลา (Time Normalization) ตามลำดับ

ในขั้นตอนการประมวลผลสัญญาณเบื้องต้นนั้น เนื่องจากสัญญาณเสียงพูดมีความแปรเปลี่ยนตามเวลา (Time-varying) และไม่เสถียร (Nonstationary) อีกทั้งยังเป็นสัญญาณสุ่มที่ไม่มีความเป็นเออร์กอดิก (Non-Ergodic) และไม่เป็นสัญญาณเฟ้นสุ่ม (Non-stochastic Signal) อีกด้วย ดังนั้นในการประยุกต์ใช้งานขั้นตอนวิธีการต่างๆ กับสัญญาณเสียงพูดจึงต้องแบ่งสัญญาณเสียงพูดออกเป็น

ส่วนย่อย (Rabiner and Levinson, 1981; Furui, 1985) เรียกว่า "กรอบเสียงพูด" (Speech Frame) โดยแต่ละกรอบเสียงพูดจะมีความยาวประมาณ 10 - 40 มิลลิวินาที (ms) ขึ้นอยู่กับความถี่ในการสุ่มตัวอย่าง (Sampling Frequency) ซึ่งในงานวิจัยนี้กำหนดให้ขนาดของกรอบเสียงพูดมีความยาว 20 มิลลิวินาทีเพื่อให้สอดคล้องกับความถี่ในการสุ่มตัวอย่างของสัญญาณเสียงพูดที่ 11 กิโลเฮิรตซ์ ทำให้ได้จำนวนข้อมูล 220 ค่าต่อหนึ่งกรอบเสียงพูดเพื่อใช้ในการประมวลผลต่อไป

1) กรรมวิธีเน้นล่วงหน้า (Preemphasis)

ขั้นตอนกรรมวิธีเน้นล่วงหน้านี้เป็นการบีบอัดช่วงพิสัยพลวัต (Dynamic Range) ของสัญญาณเสียงพูด โดยการทำให้ความลาดเอียงในเชิงความถี่แบนราบลงซึ่งจะส่งผลให้ค่าอัตราส่วนสัญญาณต่อสัญญาณรบกวนมีค่าสูงขึ้น ในทางปฏิบัติแล้วจะนำสัญญาณผ่านตัวกรองเชิงเลขลำดับหนึ่ง (First-Order Digital Filter) ที่มีฟังก์ชันถ่ายโอน H(z) ดังแสดงในสมการที่ (3.2) และ (3.3) (Furui, 1985) เมื่อ a เป็นสัมประสิทธิ์ของตัวกรอง $\tilde{s}(n)$ เป็นค่าของสัญญาณเสียงพูดขาออกที่ผ่านกรรมวิธีเน้นล่วงหน้าที่ n $s(n)$ เป็นค่าของสัญญาณเสียงพูดขาเข้าที่ n และ $s(n-1)$ เป็นค่าของสัญญาณเสียงพูดขาเข้าค่าก่อนหน้าที่ n-1 ดังนี้

$$H(z) = 1 - az^{-1} \quad \text{.....(3.2)}$$

$$\tilde{s}(n) = s(n) - as(n-1) \quad \text{.....(3.3)}$$

โดยกำหนดให้ค่าสัมประสิทธิ์ของตัวกรอง a มีค่าเข้าใกล้ 1 ในงานวิจัยนี้กำหนดให้มีค่า a = 0.95 เนื่องจากเป็นค่าที่ให้ผลดีที่สุดในการคำนวณหาค่าสัมประสิทธิ์ของการประมาณพันธะเชิงเส้น (Rabiner, Levinson, Rosenberg, and Wilpon, 1979)

2) กรรมวิธีวางกรอบขนาดสัญญาณ (Smoothing Window)

ขั้นตอนกรรมวิธีวางกรอบขนาดสัญญาณจัดเป็นขั้นตอนในการเตรียมข้อมูลในแต่ละกรอบข้อมูลเสียงพูดเพื่อการวิเคราะห์อัตสหสัมพันธ์ โดยการคูณแต่ละค่าของสัญญาณในกรอบข้อมูลเสียงพูดด้วยค่าฟังก์ชันกรอบ (Window Function) ซึ่งมีหลายประเภท ผลของการวางกรอบขนาดสัญญาณมี 2 ประการ ประการแรก เป็นการลดทอนแอมปลิจูดอย่างช้าๆ ที่บริเวณปลายแต่ละข้างของกรอบข้อมูลเสียงพูดเพื่อป้องกันการเปลี่ยนแปลงอย่างกระทันหันที่จุดปลาย ประการที่สอง เป็นการสร้างค่าการประสานสำหรับผลการแปลงฟูริเยร์ของฟังก์ชันกรอบและแถบสเปกตรัมของเสียงพูด สำหรับในงานวิจัยนี้เลือกใช้ฟังก์ชันกรอบชนิด Hamming ดังแสดงในรูปที่ 2.2 สำหรับวิเคราะห์เสียงพูดโดยเฉพาะดังแสดงในสมการที่ (3.4) และ (3.5) (Furui, 1985; Oppenheim and Schafer, 1989)

$$\tilde{x}_l(n) = x_l(n) \cdot w(n), \qquad l = 0, 1, \ldots, L-1 \quad \text{.....(3.4)}$$

$$w(n) = 0.54 - 0.46\cos\left[\frac{2\pi n}{N-1}\right], \qquad n = 0, 1, \ldots, N-1 \quad \text{.....(3.5)}$$

เมื่อ L เป็นจำนวนกรอบข้อมูลเสียงพูดทั้งหมด N เป็นจำนวนข้อมูลในแต่ละกรอบข้อมูลเสียงพูด l เป็นกรอบที่ l ของกรอบ L ทั้งหมด และ n เป็นข้อมูลที่ n ของข้อมูลทั้งหมด N ค่าซึ่งอยู่ภายในกรอบที่ l

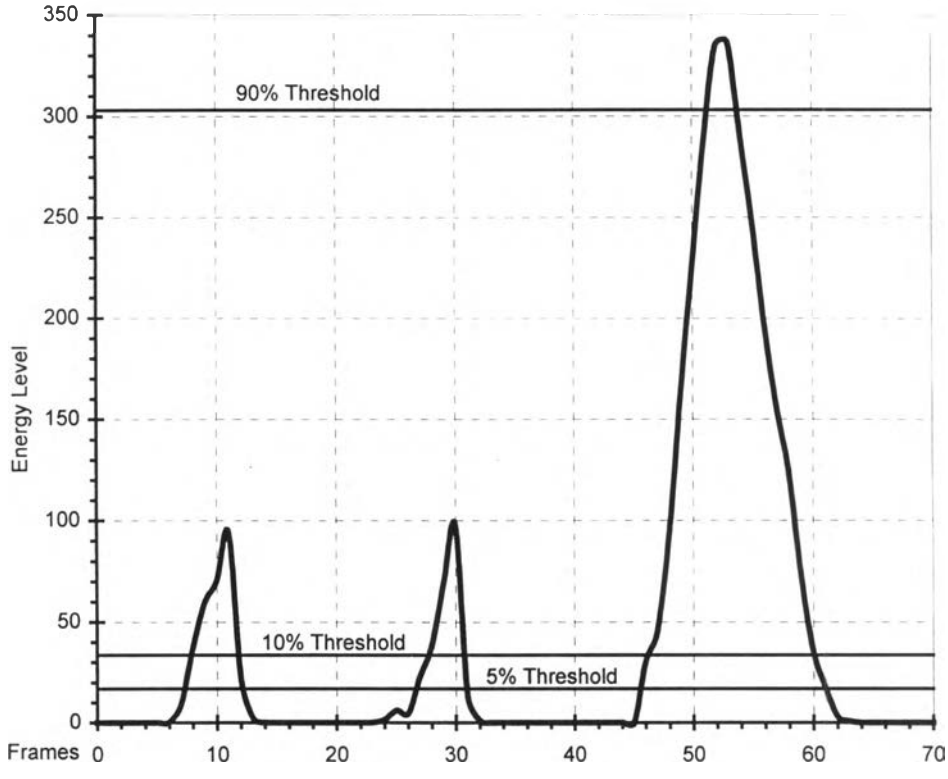
3) กรรมวิธีหาจุดสิ้นสุดเสียงพูด (Endpoint Detection)

ขั้นตอนกรรมวิธีหาจุดสิ้นสุดเสียงพูดนี้ จัดได้ว่าเป็นขั้นตอนที่สำคัญที่สุดขั้นตอนหนึ่งในกระบวนการรู้จำคำพูด เนื่องจากเป็นการกำหนดจุดเริ่มต้นและจุดสิ้นสุดของคำทั้งหมดหรือของแต่ละพยางค์เพื่อการแยกแยะเสียงพูด เพื่อนำแต่เฉพาะส่วนที่เป็นเสียงพูดมาทำการวิเคราะห์ในขั้นตอนการรู้จำคำพูดของคำศัพท์แต่ละคำ สำหรับงานวิจัยนี้จะอาศัยแผนภูมิเส้นระดับพลังงาน (Energy Level Contour) ของเสียงพูดโดยการวิเคราะห์ค่าระดับพลังงานของแต่ละกรอบข้อมูลเสียงพูด ดังแสดงในสมการที่ (3.10) เมื่อ $E(m)$ เป็นค่าระดับพลังงานของกรอบข้อมูลเสียงพูดที่ m และ $s(n)$ แทนค่าของสัญญาณค่าที่ n ในกรอบข้อมูลเสียงพูด (Rabiner and Levinson, 1981)

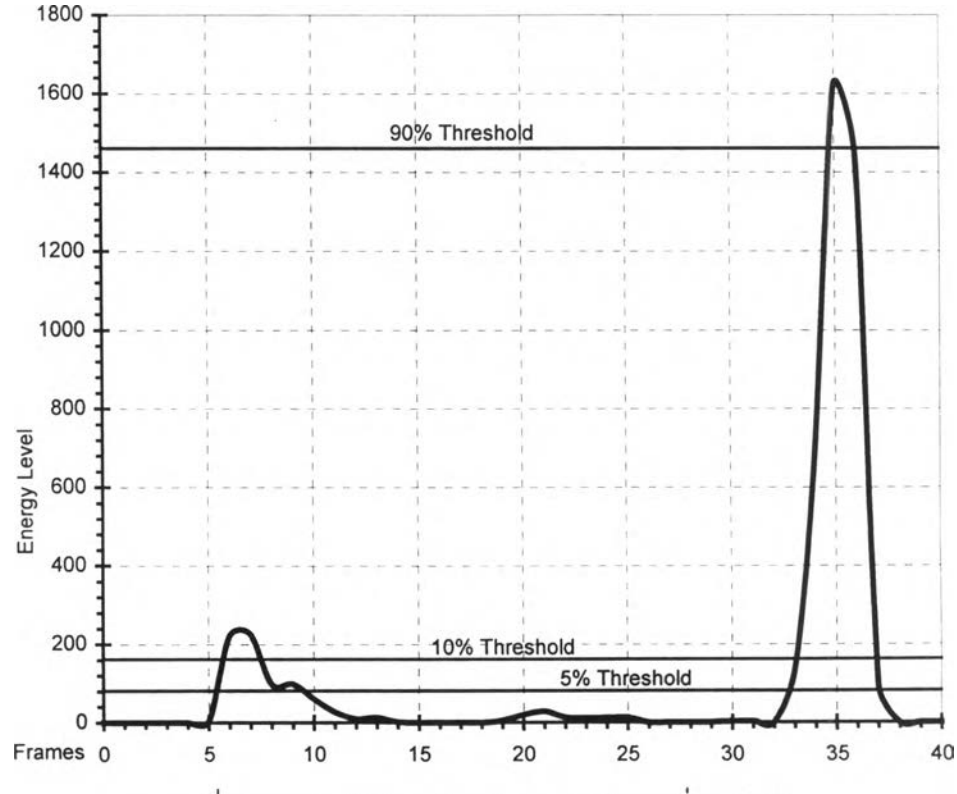
$$E(m) = \sum_{n=0}^{N-1} |s(n)| \quad \text{.....(3.10)}$$

ขั้นตอนกรรมวิธีการหาจุดสิ้นสุดเสียงพูดที่ใช้ในงานวิจัยนี้ จะอาศัยการวิเคราะห์ค่าระดับพลังงานของเสียงพูดแต่ละคำดังสมการที่ (3.10) ดังแสดงในรูปที่ 3.6 โดยการกำหนดจุดเริ่มเปลี่ยนของระดับพลังงาน (Energy Level Threshold) ที่ระดับพลังงานร้อยละ 5 และร้อยละ 10 ของระดับพลังงานสูงสุดตามลำดับ ร่วมกับช่วงเวลาของระดับพลังงาน (Energy Level Duration) ที่สูงกว่าระดับพลังงานร้อยละ 10 ซึ่งกำหนดให้ช่วงเวลาต่อเนื่องกันมากกว่า 5 กรอบข้อมูลเสียงพูดจึงจะนับจุดเริ่มเปลี่ยนที่ระดับพลังงานร้อยละ 5 เป็นจุดเริ่มต้นของเสียงพูด และจุดสิ้นสุดเสียงพูดเป็นจุดที่มีค่าระดับพลังงานต่ำกว่าร้อยละ 5 โดยจะต้องมีจุดเริ่มต้นของเสียงพูดเกิดขึ้นมาก่อนจะเริ่มตรวจจับจุดสิ้นสุดเสียงพูด การกำหนดระดับพลังงานเป็นร้อยละของค่าระดับพลังงานสูงสุดจะช่วยให้ไม่จำเป็นต้องทำให้ระดับพลังงานเป็นบรรทัดฐานเดียวกันทั้งหมด (Normalize) ก่อนการวิเคราะห์

ในการเลือกจุดเริ่มเปลี่ยนที่ระดับพลังงานร้อยละ 5 และ 10 ของค่าระดับพลังงานสูงสุดนั้น เนื่องจากคำหลายพยางค์บางคำอาจมีพยางค์ที่มีความเข้มเสียงสูงกว่าพยางค์อื่นหลายเท่า เช่นตัวอย่างของเสียงพูดสามพยางค์ดังแสดงในรูปที่ 3.6 ซึ่งพยางค์ที่สามมีขนาดความเข้มเสียงสูงกว่าพยางค์ที่หนึ่งและสองมาก จึงต้องลดระดับพลังงานของจุดเริ่มเปลี่ยนให้ต่ำลงเพื่อสามารถหาจุดสิ้นสุดเสียงพูดของแต่ละพยางค์ได้อย่างถูกต้อง



รูปที่ 3.6 แผนภูมิเส้นระดับพลังงานของเสียงพูดปรกติ



รูปที่ 3.7 แผนภูมิเส้นระดับพลังงานของเสียงพูดที่ผิดปรกติ

จากการทดลองเปลี่ยนค่าจุดเริ่มเปลี่ยนของระดับพลังงานไปที่ระดับพลังงานร้อยละ 3 และ 5 ของค่าระดับพลังงานงานสูงสุด พบว่าเกิดความผิดพลาดในการหาจุดสิ้นสุดเสียงพูดมากกว่าที่ระดับพลังงานร้อยละ 5 และ 10 เมื่อทดลองเปลี่ยนค่าจุดเริ่มเปลี่ยนของระดับพลังงานไปที่ระดับพลังงานร้อยละ 10 และ 90 ของค่าระดับพลังงานสูงสุด พบว่าเกิดความผิดพลาดในการหาจุดสิ้นสุดเสียงพูด เมื่อเสียงพูดมีพยางค์ที่มีความเข้มเสียงสูงกว่าพยางค์อื่นมากดังแสดงในรูปที่ 3.7 ทำให้การตรวจจับจุดสิ้นสุดเสียงพูดตรวจจับได้แต่เพียงจุดสิ้นสุดของพยางค์ที่มีความเข้มเสียงสูงสุดเท่านั้น ซึ่งแสดงให้เห็นว่าในกรณีที่ความแตกต่างของค่าความเข้มเสียงระหว่างพยางค์มีค่ามากจะไม่สามารถตรวจจับจุดสิ้นสุดเสียงพูดได้ ดังนั้นที่ระดับพลังงานร้อยละ 10 และ 90 จึงเกิดความผิดพลาดในการตรวจจับจุดสิ้นสุดพยางค์ของคำหลายพยางค์

สำหรับงานวิจัยนี้จะกำหนดค่าจุดเริ่มเปลี่ยนของระดับพลังงานเพื่อใช้ในการตรวจจับจุดสิ้นสุดพยางค์ที่ระดับพลังงานร้อยละ 5 และ 10 ตามลำดับ

4) กรรมวิธีปรับบรรทัดฐานเชิงเวลา (Time Normalization)

ขั้นตอนกรรมวิธีปรับบรรทัดฐานเชิงเวลาเป็นขั้นตอนในการเพิ่มหรือลดขนาดความยาวของสัญญาณในเชิงเวลา เนื่องจากสัญญาณเสียงพูดของแต่ละบุคคลมีความยาวไม่เท่ากัน จึงจำเป็นต้องปรับแต่งขนาดความยาวของสัญญาณเสียงพูดให้เหมาะสม เพื่อใช้ในการหาค่าลักษณะสำคัญและการเปรียบเทียบลักษณะสำคัญของสัญญาณเสียงต่อไป

ขั้นตอนกรรมวิธีปรับบรรทัดฐานเชิงเวลาจะอาศัยขั้นตอนในการเปลี่ยนแปลงอัตราการสุ่มตัวอย่าง (Sampling Rate) หรือความถี่ในการสุ่มตัวอย่าง (Sampling Frequency) หรือคาบเวลาในการสุ่มตัวอย่าง (Sampling Period) โดยมีกระบวนการดังแสดงในรูปที่ 2.4 และ 2.5 ในบทที่ 2 ซึ่งขั้นตอนการเปลี่ยนแปลงอัตราการสุ่มตัวอย่างแบบไม่เป็นจำนวนเต็ม ซึ่งจะต้องทำการเพิ่มอัตราการสุ่มขึ้นเป็นจำนวน L เท่า จากนั้นจึงลดอัตราการสุ่มลง M เท่าให้เท่ากับอัตราการสุ่มที่ต้องการ สำหรับในงานวิจัยนี้อาศัยขั้นตอนดังแสดงในรูปที่ 2.5 ด้วยตัวกรองแบบผ่านต่ำชนิด Finite Impulse Response (FIR) ดังมีรายละเอียดสำหรับแต่ละขั้นตอนดังนี้ (Oppenheim and Schafer, 1989; Embree, 1995)

ก) การเพิ่มอัตราการสุ่มตัวอย่าง (Upsampling)

การเพิ่มอัตราการสุ่มตัวอย่างขึ้น L เท่าก็คือการเพิ่มขนาดจำนวนตัวอย่างขึ้น L เท่าของจำนวนตัวอย่างในสัญญาณเดิม โดยอาศัยใช้วิธีการเพิ่มตัวอย่างที่มีค่าเป็นศูนย์จำนวน L-1 ตัวอย่างเข้าไประหว่างตัวอย่างเดิมทุก L ตัวอย่าง ซึ่งวิธีการนี้เรียกว่า "การเพิ่มตัวอย่างศูนย์" (Zero-Packing) ตัวอย่างศูนย์ซึ่งถูกเพิ่มเข้าไปจะปรากฏเป็นค่าที่ได้รับการประมาณขึ้นมา ประสิทธิผลของการเพิ่มตัวอย่างศูนย์ให้กับสัญญาณเดิมจะเป็นการเพิ่มซ้ำสเปกตรัม L ครั้งให้กับสเปกตรัมของสัญญาณใหม่ (Embree, 1995)

ข) ตัวกรองแบบผ่านต่ำ (Lowpass Filter)

การออกแบบตัวกรองแบบผ่านต่ำเพื่อใช้ในการเปลี่ยนแปลงอัตราการสุ่มตัวอย่าง คุณสมบัติของของตัวกรองแบบผ่านต่ำจะต้องลดทอน L-1 สเปกตรัมที่ไม่ต้องการซึ่งอยู่นอกเหนือสเปกตรัมของสัญญาณเดิมออกไป โดยช่วงความถี่ปล่อยผ่าน (Passband) จะอยู่ในช่วงความถี่ระหว่าง $0 - f_s'/(2L)$ และช่วงความถี่ลดทอน (Stopband) จะอยู่ในช่วงความถี่ระหว่าง $f_s'/(2L) - f_s'/2$ เมื่อ f_s' เป็นอัตราการสุ่มตัวอย่างของตัวกรองที่มีขนาด L เท่าของอัตราการสุ่มตัวอย่างของสัญญาณเดิม อัตราขยาย (Gain) ของตัวกรองจะต้องมีขนาดเท่ากับ L เพื่อชดเชยผลของการเพิ่มตัวอย่างศูนย์ให้แก่สัญญาณเดิม ซึ่งจะช่วยรักษาอำพนของสัญญาณเดิมเอาไว้ (Embree, 1995)

ค) การลดอัตราการสุ่มตัวอย่าง (Downsampling)

การลดอัตราการสุ่มตัวอย่างลง M เท่าก็คือการลดขนาดจำนวนตัวอย่างลง M เท่าของจำนวนตัวอย่างในสัญญาณเดิม ด้วยการเก็บค่าตัวอย่างทุก M ตัวอย่างของสัญญาณเดิมหรือที่ทุกค่า $0, M, 2M, 3M, \ldots, nM$ ของสัญญาณเดิมเอาไว้โดยไม่นำค่าระหว่างทุก M ตัวอย่างมาใช้งาน ซึ่งจะทำให้สัญญาณใหม่ที่ได้มีอัตราการสุ่มตัวอย่างลดลง M เท่าของสัญญาณเดิม (Embree, 1995)

1.1.2 ขั้นตอนการสร้างชุดรหัสอ้างอิง (Codebook Reference Template Creation)

ขั้นตอนการการสร้างชุดรหัสอ้างอิงเป็นกรรมวิธีในการฝึกฝนชุดรหัสที่สร้างขึ้นมาโดยการสุ่มจากชุดตัวอย่างเสียงพูดทั้งหมด แล้วนำชุดรหัสเริ่มต้นที่ได้จากการสุ่ม (Randomly Initialized Codebook) มาทำการฝึกฝนกับชุดตัวอย่างเสียงพูดทั้งหมดด้วยขั้นตอนวิธีการแบ่งเฉลี่ย K ส่วน (K-Means Clustering Algorithm) ซึ่งจัดอยู่ในประเภทขั้นตอนวิธีการแบ่งกลุ่มแบบวนซ้ำ (Iterative Clustering Algorithm) เนื่องจากขั้นตอนวิธีการแบ่งเฉลี่ย K ส่วนมีประสิทธิภาพใกล้เคียงกับการแบ่งกลุ่มแบบทวิภาคเมื่อจำนวนข้อมูลเวกเตอร์ฝึกฝนมีมากเพียงพอ (Makhoul, Roucos, and Gish, 1985) ขั้นตอนวิธีการนี้มีรายละเอียดในทางทฤษฎีดังแสดงในบทที่ 2 หัวข้อที่ 2.2 ขั้นตอนการสร้างชุดรหัสอ้างอิงแบ่งออกเป็น 2 กรรมวิธีย่อยได้แก่ ขั้นตอนการหาค่าสัมประสิทธิ์ของการประมาณพันธะเชิงเส้น (Linear Prediction Coefficient Analysis) และขั้นตอนการฝึกฝนชุดรหัสอ้างอิง (Codebook Reference Template Training) ตามลำดับ

1) ขั้นตอนการหาค่าสัมประสิทธิ์ของการประมาณพันธะเชิงเส้น
(Linear Prediction Coefficient Analysis)

ขั้นตอนการหาค่าสัมประสิทธิ์ของการประมาณพันธะเชิงเส้นจัดเป็นขั้นตอนในการลดจำนวนข้อมูลโดยการแสดงลักษณะของรูปคลื่นด้วยค่าพารามิเตอร์เพียงไม่กี่ค่าได้อย่างมีประสิทธิภาพ

โดยนำข้อมูลที่ผ่านกรรมวิธีหาจุดสิ้นสุดเสียงพูดมาทำการวิเคราะห์ดังแสดงในบทที่ 2 หัวข้อที่ 1 สำหรับงานวิจัยนี้จะอาศัยวิธีการทางอัตสหสัมพันธ์ด้วยขั้นตอนวิธีการ Levinson-Durbin ดังแสดงในตารางที่ 2.1 (O'Shaughnessy, 1988) ในการหาค่าสัมประสิทธิ์ของการประมาณพันธะเชิงเส้นสำหรับแต่ละเสียงพูดนั้น จะหาค่าสัมประสิทธิ์เฉพาะแต่ละกรอบข้อมูลโดยการแทนที่แต่ละกรอบข้อมูลด้วยเวกเตอร์ของสัมประสิทธิ์ของการประมาณพันธะเชิงเส้น สำหรับงานวิจัยนี้จะใช้ค่าลำดับของสัมประสิทธิ์ที่ 10 ลำดับ เนื่องจากเป็นค่าลำดับที่ให้อัตราการรู้จำสูงที่สุดในการทดลองเมื่อเปรียบเทียบกับค่าลำดับอื่น (เสาวลักษณ์ อารีย์พงศา, 2538)

2) ขั้นตอนการฝึกฝนชุดรหัสอ้างอิง

(Codebook Reference Template Training)

ขั้นตอนการฝึกฝนชุดรหัสอ้างอิงจัดเป็นขั้นตอนในการสร้างและฝึกฝนชุดรหัสโดยอาศัยชุดตัวอย่างเสียงพูดในการฝึกฝน การสร้างชุดรหัสอ้างอิงเริ่มต้นจะทำการสุ่มตัวอย่างมาจากชุดเวกเตอร์ของสัมประสิทธิ์ของการประมาณพันธะเชิงเส้นในแต่ละตัวอย่างเสียงพูด สำหรับการฝึกฝนกับชุดตัวอย่างเสียงพูดเพื่อสร้างชุดรหัสอ้างอิงที่สมบูรณ์ การฝึกฝนชุดรหัสอ้างอิงจะอาศัยขั้นตอนวิธีการแบ่งเฉลี่ย K ส่วนโดยใช้วิธีการวัดค่าความเพี้ยนแบบค่าความเพี้ยนกำลังสองเฉลี่ย (Mean Square Error, MSE) ดังมีรายละเอียดแสดงในบทที่ 2 หัวข้อที่ 2.2 และ 2.1 ตามลำดับ สำหรับงานวิจัยนี้จะทำการสร้างและฝึกฝนชุดรหัสจำนวน 10 ชุด และนำมาเฉลี่ยพร้อมทั้งฝึกฝนเป็นชุดรหัสอ้างอิงเฉลี่ยเพื่อนำไปทำการควอนไทซ์แบบเวกเตอร์ต่อไป

การสุ่มตัวอย่างชุดรหัสเริ่มต้นที่แตกต่างกันไปจะช่วยให้ชุดรหัสที่ได้เข้าสู่ค่าที่เหมาะสมที่สุดที่ครอบคลุมทั้งหมด (Global Optimum) และมีค่าความเพี้ยนต่ำที่สุด สำหรับในงานวิจัยนี้จะใช้ชุดรหัสขนาด 256 เวกเตอร์ของสัมประสิทธิ์การประมาณพันธะเชิงเส้นที่ 10 ลำดับ ซึ่งเป็นขนาดชุดรหัสที่เหมาะสมสำหรับเก็บข้อมูลเสียงพูดที่มีปริมาณมาก และเป็นขนาดชุดรหัสที่ให้ค่าความผิดพลาดต่ำกว่าชุดรหัสขนาด 64 เวกเตอร์อีกด้วย (Makhoul, Roucos, and Gish, 1985; เสาวลักษณ์ อารีย์พงศา, 2538)

1.2 การสร้างและฝึกฝนชุดพารามิเตอร์ของแบบจำลองฮิดเดน มาร์คอฟ

(Hidden Markov Model Parameters Training)

รายละเอียดขั้นตอนการสร้างและฝึกฝนชุดพารามิเตอร์ของแบบจำลองฮิดเดน มาร์คอฟดังแสดงในรูปที่ 3.3 ประกอบไปด้วย 3 ขั้นตอนได้แก่ ขั้นตอนการวิเคราะห์หาสัมประสิทธิ์การประมาณพันธะเชิงเส้น (Linear Prediction Coefficient Analysis) ขั้นตอนการควอนไทซ์แบบเวกเตอร์ (Vector Quantization) และขั้นตอนการฝึกฝนชุดพารามิเตอร์ของแบบจำลองฮิดเดน มาร์คอฟ (Hidden Markov Model Parameters Training) ภายหลังจากการฝึกฝนชุดพารามิเตอร์ของแบบจำลองฮิดเดน มาร์คอฟจะได้ชุดรูปร่องต้นแบบอ้างอิงของเสียงพูดแต่ละคำ (Word Reference Template) เนื่องจากขั้นตอนการ

วิเคราะห์หาสัมประสิทธิ์การประมาณพันธะเชิงเส้นมีรายละเอียดอยู่ในขั้นตอนการสร้างและฝึกฝนชุดรหัสหัวข้อที่ 1.1.2 ในหัวข้อนี้จึงแสดงแต่เพียงรายละเอียดของสองขั้นตอนที่เหลือดังนี้

1) ขั้นตอนวิธีการควอนไทซ์แบบเวกเตอร์ (Vector Quantization Algorithm)

ขั้นตอนวิธีการควอนไทซ์แบบเวกเตอร์จัดเป็นขั้นตอนในการลดขนาดจำนวนข้อมูลลงให้มีความเพี้ยนน้อยที่สุด โดยการแทนที่เวกเตอร์สัมประสิทธิ์ของการประมาณพันธะเชิงเส้นด้วยเวกเตอร์ของชุดรหัสที่ให้ค่าความเพี้ยนกำลังสองเฉลี่ยมีค่าน้อยที่สุด ชุดรหัสที่ใช้ในการควอนไทซ์แบบเวกเตอร์จะเป็นชุดรหัสที่ได้จากกระบวนการฝึกฝนชุดรหัส (Makhoul, Roucos, and Gish, 1985) รายละเอียดของขั้นตอนวิธีการควอนไทซ์แบบเวกเตอร์ดังแสดงในบทที่ 2 หัวข้อที่ 2 ผลลัพธ์ที่ได้จะเป็นลำดับหมายเลขเวกเตอร์ของชุดรหัสที่ใช้ในการควอนไทซ์ สำหรับงานวิจัยนี้จะใช้ชุดรหัสขนาด 256 เวกเตอร์ของสัมประสิทธิ์การประมาณพันธะเชิงเส้นที่ 10 ลำดับ

2) ขั้นตอนการฝึกฝนชุดพารามิเตอร์ของแบบจำลองฮิดเดน มาร์คอฟ (Hidden Markov Model Parameters Training)

ขั้นตอนวิธีการฝึกฝนชุดพารามิเตอร์ของแบบจำลองฮิดเดน มาร์คอฟจัดเป็นกระบวนการสร้างชุดรูปร่องต้นแบบอ้างอิงของเสียงพูดแต่ละคำ โดยทำการปรับเปลี่ยนค่าพารามิเตอร์ของแบบจำลองฮิดเดน มาร์คอฟ $\lambda = (A, B, \pi)$ ให้เป็นไปตามเสียงพูดแต่ละคำ ซึ่งอาศัยการแก้ไขปัญหาพื้นฐานข้อที่ 1 และข้อที่ 3 ของแบบจำลองฮิดเดน มาร์คอฟด้วยกระบวนการไปหน้า-ย้อนกลับ (Forward-Backward Procedure) และการประมาณค่าซ้ำของ Baum-Welch (Baum-Welch Reestimation Procedure) ดังมีรายละเอียดแสดงในบทที่ 2 หัวข้อที่ 3.1 และ 3.3 ตามลำดับ (Rabiner and Juang, 1986; Rabiner, 1989)

สำหรับการฝึกฝนชุดพารามิเตอร์ของแบบจำลองฮิดเดน มาร์คอฟในงานวิจัยนี้ จะใช้ข้อมูลเสียงพูดที่ผ่านกระบวนการควอนไทซ์แบบเวกเตอร์เป็นข้อมูลขาเข้า โดยใช้เสียงพูดแต่ละคำของผู้พูดทุกคนมาฝึกฝนร่วมกันในแต่ละครั้ง ทำให้ชุดพารามิเตอร์ที่ได้เป็นตัวแทนหนึ่งเดียวของทุกเสียงพูดทั้งหมดเพื่อใช้เป็นชุดรูปร่องต้นแบบอ้างอิงสำหรับเสียงพูดแต่ละคำในการรู้จำต่อไป

2. ขั้นตอนการทดสอบระบบการรู้จำคำพูด

ขั้นตอนในการทดสอบระบบการรู้จำคำพูดโดยใช้แบบจำลองฮิดเดน มาร์คอฟประกอบด้วย 3 ขั้นตอน ได้แก่ขั้นตอนการประมวลผลสัญญาณเบื้องต้น (Signal Preprocessing) ขั้นตอนการวิเคราะห์และวัดค่าลักษณะสำคัญ (Feature Measurement) และขั้นตอนการจำแนกรูปแบบร่วมกับขั้นตอนวิธีการตัดสินใจ (Pattern Classification and Decision Making Algorithm) ตามลำดับ เนื่องจากในขั้นตอน

การประมวลผลสัญญาณเบื้องต้นมีรายละเอียดอยู่ในขั้นตอนการสร้างและฝึกฝนชุดรหัสหัวข้อที่ 1.1 ในหัวข้อนี้จึงแสดงแต่เพียงรายละเอียดของสองขั้นตอนที่เหลือเท่านั้นดังนี้

2.1 ขั้นตอนการวิเคราะห์และวัดค่าลักษณะสำคัญ (Feature Measurement)

ขั้นตอนการวิเคราะห์และวัดค่าลักษณะสำคัญประกอบด้วย 2 ขั้นตอน ได้แก่ขั้นตอนการหาค่าสัมประสิทธิ์การประมาณพันธะเชิงเส้น และขั้นตอนวิธีการควอนไทซ์แบบเวกเตอร์ตามลำดับ โดยอาศัยชุดรหัสต้นแบบอ้างอิง (Codebook Reference Templates) ที่ได้จากการฝึกฝนชุดรหัสอ้างอิงมาใช้ในการควอนไทซ์แบบเวกเตอร์ ซึ่งขั้นตอนทั้งสองนี้เป็นขั้นตอนเดียวกับขั้นตอนย่อยของการสร้างและฝึกฝนชุดรหัส กับการสร้างและฝึกฝนชุดพารามิเตอร์ของแบบจำลองฮิดเดน มาร์คอฟตามลำดับ ดังมีรายละเอียดแสดงในหัวข้อที่ 1.1.2 และ 1.2 หัวข้อย่อยที่ 1) และ 2) ตามลำดับ

2.2 ขั้นตอนการจำแนกรูปแบบร่วมกับขั้นตอนวิธีการตัดสินใจ (Pattern Classification and Decision Making Algorithm)

ขั้นตอนการจำแนกรูปแบบของเสียงพูดอาศัยแบบจำลองฮิดเดน มาร์คอฟ ร่วมกับชุดรูปร่องต้นแบบอ้างอิงของเสียงพูดแต่ละคำ ซึ่งได้จากขั้นตอนการฝึกฝนชุดพารามิเตอร์ของแบบจำลองฮิดเดน มาร์คอฟ เพื่อการจำแนกความแตกต่างของเสียงพูดแต่ละคำโดยการเปรียบเทียบเสียงพูดที่นำมาทดสอบกับชุดรูปร่องต้นแบบของเสียงพูดเพื่อใช้ในขั้นตอนวิธีการตัดสินใจ สำหรับในงานวิจัยนี้ ขั้นตอนการวิธีการตัดสินใจจะอาศัยแบบจำลองฮิดเดน มาร์คอฟ โดยการแก้ไขปัญหาพื้นฐานข้อที่ 2 ด้วยขั้นตอนวิธีการ Viterbi (Viterbi Algorithm) ดังมีรายละเอียดแสดงในบทที่ 2 หัวข้อที่ 3.2 ผลลัพธ์ที่ได้จากขั้นตอนนี้จะเป็นชุดของเสียงพูดที่รู้จำได้ซึ่งมีค่าความน่าจะเป็นสูงที่สุด (Rabiner and Juang, 1986; Rabiner, 1989)