

ระบบบ่งชี้ผู้พูดแบบระบบเปิดโดยใช้แบบจำลองฮิดเดนมาร์คอฟ
แบบหลายชุดรหัส

นายพงศ์ไท ทาสระอุ

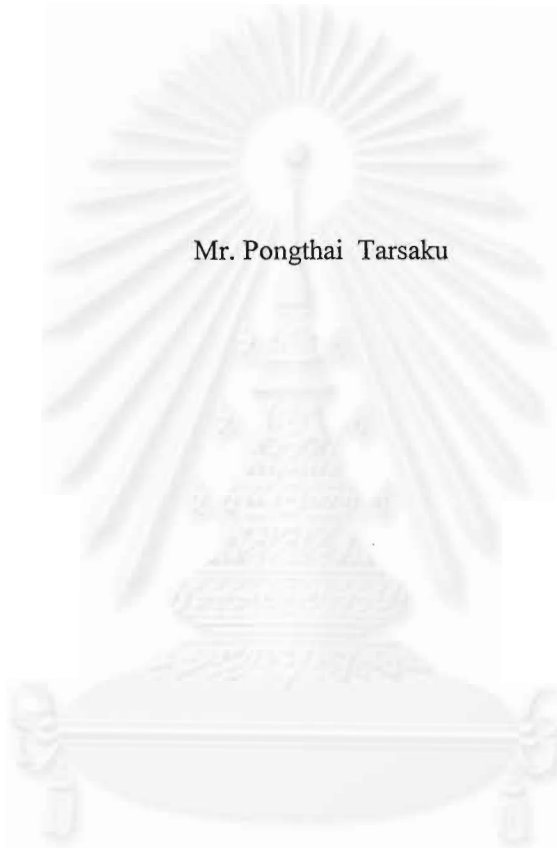


วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมไฟฟ้า ภาควิชาวิศวกรรมไฟฟ้า
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2542

ISBN 974-334-038-6

ลิขสิทธิ์ของ จุฬาลงกรณ์มหาวิทยาลัย

OPEN SET SPEAKER IDENTIFICATION USING MULTIPLE CODEBOOK HMM



Mr. Pongthai Tarsaku

A Thesis Submitted in Partial Fulfillment of the Requirements for
the Degree of Master of Engineering in Electrical Engineering

Department of Electrical Engineering

Faculty of Engineering

Academic Year 1999

ISBN 974-334-038-6


หัวข้อวิทยานิพนธ์ การบ่งชี้ผู้พูดแบบระบบเปิดโดยใช้แบบจำลองฮิดเดนมาร์คอฟแบบ
หลายชุดรหัส
โดย นายพงศ์ไท ทาสระคู
ภาควิชา วิศวกรรมไฟฟ้า
อาจารย์ที่ปรึกษา รองศาสตราจารย์ ดร.สมชาย จิตะพันธ์กุล
อาจารย์ที่ปรึกษาร่วม อาจารย์ ดร.จุฬารัตน์ ตันประเสริฐ

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้หัวข้อวิทยานิพนธ์ฉบับนี้เป็นส่วน
หนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรบัณฑิต


.....
(ศาสตราจารย์ ดร.สมศักดิ์ ปัญญาแก้ว)

คณบดีคณะวิศวกรรมศาสตร์

คณะกรรมการสอบวิทยานิพนธ์


.....
(ผู้ช่วยศาสตราจารย์ ดร.สุตาพร ลักษณีย์นาวิน)

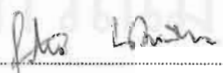
ประธานกรรมการ


.....
(รองศาสตราจารย์ ดร.สมชาย จิตะพันธ์กุล)

อาจารย์ที่ปรึกษา


.....
(อาจารย์ ดร.จุฬารัตน์ ตันประเสริฐ)

อาจารย์ที่ปรึกษาร่วม


.....
(อาจารย์สุวิทย์ นาคพิระบุต)

กรรมการ

พงศ์ไท ทาสระคู : ระบบบ่งชี้ผู้พูดแบบระบบเปิดโดยใช้แบบจำลองฮิดเดนมาร์คอฟแบบ
หลายชุดรหัส (OPEN SET SPEAKER IDENTIFICATION USING MULTIPLE
CODEBOOK HMM) อาจารย์ที่ปรึกษา : รศ.ดร.สมชาย จิตะพันธ์กุล, อาจารย์ที่ปรึกษา
ร่วม : ดร.จุฬารัตน์ ตันประเสริฐ, 123 หน้า. ISBN 974-334-038-6

วิทยานิพนธ์นี้มีวัตถุประสงค์เพื่อนำเสนอระบบบ่งชี้ผู้พูดแบบระบบเปิด โดยใช้แบบจำลอง
ฮิดเดนมาร์คอฟร่วมกับการควอนไทซ์แบบเวกเตอร์ โดยใช้ชุดรหัสแบบหลายชุดรหัส ระบบบ่งชี้ผู้
พูดนี้เป็นระบบที่ขึ้นกับบทคำพูด และใช้กับเสียงพูดต่อเนื่อง ในขั้นตอนการตรวจสอบผู้พูดซึ่งเป็น
ขั้นตอนสุดท้ายของระบบบ่งชี้ผู้พูดแบบระบบเปิดนั้น ได้มีการนำเสนอฟังก์ชันของความแตกต่าง
ขึ้นมาใช้ สำหรับการทดลองทำกับฐานข้อมูลเสียงพูดตัวเลขต่อเนื่อง “สาม-ห้า-สอง-เก้า-สี่” โดย
แบ่งเป็นจำนวนผู้พูดในระบบ 10 คนและจำนวนของผู้พูดนอกระบบ 17 คน ผู้พูดแต่ละคนจะ
บันทึกเสียงแยกกัน 2 ช่วง แต่ละช่วงเว้นห่างกัน 1 เดือน และการบันทึกเสียงในแต่ละช่วงจะบันทึก
คนละ 10 เสียง ผลการทดลองปรากฏว่า ลักษณะสำคัญ MFCC ให้ผลดีที่สุดจากการศึกษาลักษณะ
สำคัญ 3 แบบ ได้แก่ LPC, CEP, และ MFCC พบว่า MFCC ให้อัตราการบ่งชี้ผิดพลาดเฉลี่ยเป็น
0.40 เปอร์เซ็นต์ อัตราการยอมรับผิดพลาดเฉลี่ย 0.71 เปอร์เซ็นต์ และอัตราการปฏิเสธผิดพลาดเฉลี่ย
9.40 เปอร์เซ็นต์

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาควิชา.....วิศวกรรมไฟฟ้า..... ลายมือชื่อนิสิต.....
สาขาวิชา.....วิศวกรรมไฟฟ้า..... ลายมือชื่ออาจารย์ที่ปรึกษา.....
ปีการศึกษา.....2542..... ลายมือชื่ออาจารย์ที่ปรึกษาร่วม.....

4170414921 : MAJOR ELECTRICAL ENGINEERING

KEY WORD: SPEAKER IDENTIFICATION / SPEAKER VERIFICATION / HIDDEN

MARKOV MODELS / MULTIPLE CODEBOOK.

PONGTHAI TARSAKU : OPEN SET SPEAKER IDENTIFICATION USING

MULTIPLE CODEBOOK HMM. THESIS ADVISOR : ASSOC. PROF.

SOMCHAI JITAPUNKUL, Dr. Eng. THESIS COADVISOR : CHULARAT

TANPRASERT, Ph.D. 123 pp. ISBN 974-334-038-6

This thesis has the objective to develop an open-set speaker identification system using Hidden Markov Model and Vector Quantization with multiple codebooks. The system is a text-dependent continuous speech speaker identification system. In the final verification process, a different function is proposed to improve the performance of the system. The “3-5-2-9-4” or /sa:m4 ha:2 s@:ng ka:w2 si:1/ speech database used in the experiment consists of 10 speakers and 17 imposters. Each speaker did the record twice, where in the second session was performed one month after the first session, and each speaker was recorded 10 times per session. The experiment results show that MFCC is the best result with 0.4% average identification error rate (MFCC, LPC, and CEP have been tested in the experiments), 0.71% average false acceptance rate, and 9.40% average false rejection rate.

จุฬาลงกรณ์มหาวิทยาลัย

ภาควิชา...วิศวกรรมไฟฟ้า.....ลายมือชื่อนิสิต.....
 สาขาวิชา...วิศวกรรมไฟฟ้า.....ลายมือชื่ออาจารย์ที่ปรึกษา.....
 ปีการศึกษา...2542.....ลายมือชื่ออาจารย์ที่ปรึกษาร่วม.....



กิตติกรรมประกาศ

ในการดำเนินงานวิจัยในงานวิทยานิพนธ์นี้ ข้าพเจ้าขอกราบขอบพระคุณ รองศาสตราจารย์ ดร.สมชาย จิตะพันธ์กุล อาจารย์ที่ปรึกษา ที่ได้กรุณาให้คำแนะนำในการดำเนินการวิจัย ตลอดจนได้จัดหาอุปกรณ์และทุนการวิจัยเพื่อดำเนินการวิจัยได้อย่างสะดวกและครบถ้วน และขอกราบขอบพระคุณสำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ (สวทช.) ซึ่งเป็นสถาบันที่ให้ทุนการวิจัยในโครงการทุน TGIST แก่ข้าพเจ้า และ ดร.จุฬารัตน์ ดันประเสริฐ อาจารย์ที่ปรึกษา ที่ได้กรุณาให้คำแนะนำที่เป็นประโยชน์ต่องานวิจัย รวมทั้ง ผู้ช่วยศาสตราจารย์ ดร.สุดาพร ลักษณะินาวิน ที่ได้ให้คำปรึกษาทางด้านภาษาศาสตร์ซึ่งเป็นประโยชน์ในงานวิจัย

ขอขอบคุณ คุณวิศรุต อาขุนทร คุณเอกฤทธิ์ มณีน้อย ที่ได้ให้คำแนะนำ และความช่วยเหลือต่างๆ ทำให้งานวิจัยลุล่วงได้ด้วยดี

ขอขอบคุณ คุณศวิต กาสुरิยะ ที่ได้ร่วมกันพัฒนาระบบบ่งชี้ผู้พูดนี้ขึ้นมารวมทั้งให้คำปรึกษา คำแนะนำต่างๆ ที่ทำให้งานวิจัยนี้สำเร็จได้

ขอขอบคุณทุกท่านที่ได้เสียสละเวลาในการบันทึกเสียง และเพื่อนๆ นิสิตปริญญาโททุกท่านที่คอยเป็นกำลังใจและให้ความช่วยเหลือเสมอมา

และสุดท้ายขอกราบขอบพระคุณ ครอบครัวของข้าพเจ้าที่ได้ให้กำลังใจมาตลอดเวลา ซึ่งทำให้ข้าพเจ้ามีความพยายามและมีกำลังแรงใจในการพัฒนางานวิจัยจนทำให้วิทยานิพนธ์นี้สำเร็จได้ในที่สุด

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญตาราง.....	ฅ
สารบัญรูปประกอบ.....	ญ
สารบัญคำศัพท์.....	ฎ
บทที่ 1 บทนำ.....	1
ที่มาของปัญหา.....	1
วัตถุประสงค์.....	3
ขั้นตอนการดำเนินงาน.....	4
เป้าหมายและขอบเขตของงานวิจัย.....	4
ประโยชน์ที่คาดว่าจะได้รับ.....	5
บทที่ 2 ทฤษฎีและงานที่เกี่ยวข้อง.....	6
ระบบบ่งชี้ผู้พูด.....	6
ระบบตรวจสอบผู้พูด.....	6
การสกัดลักษณะสำคัญ.....	10
การควอนไทซ์แบบเวกเตอร์.....	16
แบบจำลองฮิดเดนมาร์คอฟ.....	22
การตรวจสอบผู้พูด.....	37
ค่าความเพี้ยน (Distortion).....	38
ฟังก์ชันค่าความแตกต่าง.....	39
บทที่ 3 ขั้นตอนวิธีในการดำเนินงานวิจัย.....	41
การกำหนดลักษณะของบทคำพูด.....	41
การเก็บบันทึกตัวอย่างเสียงพูด.....	42
การสร้างระบบบ่งชี้ผู้พูด.....	42

การฝึกฝนระบบ.....	47
การทดสอบระบบ.....	49
บทที่ 4 ผลการทดลองและการวิเคราะห์ผลการทดลอง.....	54
ฐานข้อมูลเสียงพูด.....	54
การทดลองที่ 4-1.....	57
วิเคราะห์และอภิปรายผลการทดลองที่ 4-1.....	62
การทดลองที่ 4-2.....	62
วิเคราะห์และอภิปรายผลการทดลองที่ 4-2.....	66
การทดลองที่ 4-3.....	66
การทดลองที่ 4-3-1.....	67
วิเคราะห์และอภิปรายผลการทดลองที่ 4-3-1.....	69
การทดลองที่ 4-3-2.....	70
วิเคราะห์และอภิปรายผลการทดลองที่ 4-3-2.....	72
การทดลองที่ 4-3-3.....	72
วิเคราะห์และอภิปรายผลการทดลองที่ 4-3-3.....	75
การทดลองที่ 4-4.....	75
วิเคราะห์และอภิปรายผลการทดลองที่ 4-4.....	76
การทดลองที่ 4-5.....	77
วิเคราะห์และอภิปรายผลการทดลองที่ 4-5.....	84
การทดลองที่ 4-6.....	84
การทดลองที่ 4-6-1.....	84
วิเคราะห์และอภิปรายผลการทดลองที่ 4-6-1.....	86
การทดลองที่ 4-6-2.....	87
วิเคราะห์และอภิปรายผลการทดลองที่ 4-6-2.....	89
บทที่ 5 สรุปผลและข้อเสนอแนะ.....	91
สรุปผล.....	91
ข้อเสนอแนะเพื่อปรับปรุงประสิทธิภาพของระบบ.....	93
แนวทางในการพัฒนางานวิจัยในขั้นต่อไป.....	94
รายการอ้างอิง.....	95
ภาคผนวก.....	98
ประวัติผู้เขียน.....	123

สารบัญตาราง

หน้า

ตารางที่ 4-1 รายละเอียดของเสียงพูดที่ใช้ในการทดลอง สำหรับฐานข้อมูลชุดที่ 2	56
ตารางที่ 4-2 ผลการทดลอง จำนวนเสียงที่บ่งชี้ผิดพลาด เมื่อใช้ชุดรหัสแบบ Single Codebook กับ ชุดรหัสแบบ Multiple Codebook เมื่อใช้ LPC เป็นค่าลักษณะสำคัญ	59
ตารางที่ 4-3 ผลการทดลอง จำนวนเสียงที่บ่งชี้ผิดพลาด เมื่อใช้ชุดรหัสแบบ Single Codebook กับ ชุดรหัสแบบ Multiple Codebook เมื่อใช้ CEP เป็นค่าลักษณะสำคัญ	60
ตารางที่ 4-4 ผลการทดลอง จำนวนเสียงที่บ่งชี้ผิดพลาด เมื่อใช้ชุดรหัสแบบ Single Codebook กับ ชุดรหัสแบบ Multiple Codebook เมื่อใช้ MFCC เป็นค่าลักษณะสำคัญ	61
ตารางที่ 4-5 จำนวนเสียงที่บ่งชี้ผิดพลาดเมื่อใช้ค่าฟังก์ชัน $f_1 - f_6$	63
ตารางที่ 4-6 อัตราการบ่งชี้ผิดพลาดเฉลี่ย (%) ในแต่ละค่าอันดับของลักษณะสำคัญต่างๆ	67
ตารางที่ 4-7 FA (%) กับ FR (%) ที่อันดับของค่าลักษณะสำคัญค่าต่างๆสำหรับ กลุ่มทดสอบที่ 2	68
ตารางที่ 4-8 อัตราการบ่งชี้ผิดพลาดเฉลี่ย (%) เมื่อใช้ขนาดของชุดรหัสค่าต่างๆสำหรับชุดทดสอบ ที่ 2	70
ตารางที่ 4-9 ค่า FA (%) และค่า FR (%)เฉลี่ย ที่ขนาดของชุดรหัสค่าต่างๆ	71
ตารางที่ 4-10 อัตราการบ่งชี้ผิดพลาดเฉลี่ย (%) ที่จำนวนสถานะค่าต่างๆ สำหรับชุดทดสอบที่ 2	73
ตารางที่ 4-11 ค่า FA (%) และ FR (%) เฉลี่ย ที่จำนวนสถานะค่าต่างๆ สำหรับชุดทดสอบที่ 2	74
ตารางที่ 4-12 จำนวนเสียงยอมรับและปฏิเสธผิดพลาด โดยเปรียบเทียบระหว่างระบบที่ไม่ใช้ ฟังก์ชันของความแตกต่างกับระบบที่ใช้ฟังก์ชันของความแตกต่าง	81
ตารางที่ 4-13 อัตราการลดลงของผลรวม FA และ FR เมื่อใช้ฟังก์ชันของความแตกต่าง	83
ตารางที่ 4-14 อัตราเสียงผิดพลาด (%) เมื่อเปลี่ยนแปลงจำนวนผู้พูดในระบบ (1)	85
ตารางที่ 4-15 อัตราเสียงผิดพลาด (%) เมื่อเปลี่ยนแปลงจำนวนผู้พูดในระบบ (2)	87
ตารางที่ 4-16 รายละเอียดของเสียงที่บ่งชี้ผิดพลาด	88

จุฬาลงกรณ์มหาวิทยาลัย

สารบัญรูปประกอบ

	หน้า
รูปที่ 1-1 ความสัมพันธ์ของงานวิจัยด้านการประมวลผลเสียงพูด.....	2
รูปที่ 2-1 แผนภาพการทำงานของระบบบ่งชี้ผู้พูดแบบระบบปิดและแบบระบบเปิด.....	7
รูปที่ 2-2 การทำงานของระบบตรวจสอบผู้พูด.....	7
รูปที่ 2-3 ฟังก์ชันกรอบชนิด Hamming Window.....	9
รูปที่ 2-4 Mel Scale ของความถี่สัญญาณเสียง.....	14
รูปที่ 2-5 วงจรกรองแบบผ่านแถบความถี่ที่ให้ค่าสัมประสิทธิ์เซปสตรอลบนความถี่เมล.....	16
รูปที่ 2-6 การควอนไทซ์แบบเวกเตอร์.....	17
รูปที่ 2-7 การแบ่งปริภูมิที่มีขนาด 2 มิติ.....	18
รูปที่ 2-8 ลูกโซ่มาร์คอฟ 4 สถานะ.....	22
รูปที่ 2-9 แบบจำลองฮิดเดนมาร์คอฟ 3 ประเภท (ก) แบบจำลองเออร์กอดิก 4 สถานะ (ข) แบบจำลองย้ายไปขวา 4 สถานะ (ค) แบบจำลองเส้นทางขนานย้ายไปขวา 6 สถานะ.....	24
รูปที่ 2-10 รายละเอียดของตัวแปรในการคำนวณค่าตัวแปรไปหน้า $\alpha_i(i)$	29
รูปที่ 2-11 รายละเอียดของตัวแปรในการคำนวณค่าตัวแปรย้อนกลับ $\beta_i(i)$	30
รูปที่ 2-12 ลำดับของเหตุการณ์ที่ต้องใช้ในการคำนวณเหตุการณ์ร่วมของระบบในสถานะ i ที่เวลา t และสถานะ j ที่เวลา $t + 1$	33
รูปที่ 2-13 การนำค่าความถี่มาใช้ร่วมกับค่าความน่าจะเป็น.....	38
รูปที่ 3-1 ขั้นตอนการทำงานของระบบบ่งชี้ผู้พูดแบบระบบเปิด.....	43
รูปที่ 3-2 รูปแบบการฝึกฝนชุดรหัสแบบหลายชุดรหัส.....	47
รูปที่ 3-3 ช่วงการฝึกฝนต้นแบบอ้างอิงชุดรหัส.....	48
รูปที่ 3-4 ขั้นตอนของการฝึกฝนแบบจำลองฮิดเดนมาร์คอฟ.....	50
รูปที่ 3-5 ขั้นตอนการทดสอบระบบบ่งชี้ผู้พูดแบบระบบปิด.....	51
รูปที่ 3-6 ขั้นตอนการตรวจสอบเสียงพูดที่ได้จากการบ่งชี้.....	52
รูปที่ 3-7 ตัวอย่างการหาผ่านระดับของผู้พูด.....	53
รูปที่ 4-1 ขั้นตอนการทดสอบระบบ สำหรับระบบที่ใช้ชุดรหัสเดี่ยว.....	57
รูปที่ 4-2 ขั้นตอนการทดสอบระบบ สำหรับระบบที่ใช้ชุดรหัสแบบหลายชุด.....	58
รูปที่ 4-3 ค่าเฉลี่ยจำนวนเสียงที่บ่งชี้ผิดพลาด เมื่อใช้ LPC เป็นค่าลักษณะสำคัญ.....	59
รูปที่ 4-4 ค่าเฉลี่ยจำนวนเสียงที่บ่งชี้ผิดพลาด เมื่อใช้ CEP เป็นค่าลักษณะสำคัญ.....	60
รูปที่ 4-5 ค่าเฉลี่ยจำนวนเสียงที่บ่งชี้ผิดพลาด เมื่อใช้ MFCC เป็นค่าลักษณะสำคัญ.....	61
รูปที่ 4-6 ค่าเฉลี่ยจำนวนเสียงที่บ่งชี้ผิดพลาดจากตาราง 4-5.....	63
รูปที่ 4-7 ค่าคะแนนที่ได้จากการนำเสียงทดสอบผ่านแบบจำลองของผู้พูดคนที่ 2.....	

โดยใช้ฟังก์ชัน $f_1 - f_6$	65
รูปที่ 4-8 อัตราการบ่งชี้ผิดพลาดเฉลี่ย ที่อันดับของลักษณะสำคัญค่าต่างๆ.....	68
รูปที่ 4-9 อัตราการยอมรับและปฏิเสธผิดพลาดเฉลี่ย ที่อันดับของลักษณะสำคัญค่าต่างๆ.....	69
รูปที่ 4-10 อัตราการบ่งชี้ผิดพลาดเฉลี่ย (%) ที่ขนาดของชุดรหัสค่าต่างๆ	70
รูปที่ 4-11 ค่า FA และ FR เฉลี่ย ที่ขนาดของชุดรหัสค่าต่างๆ	71
รูปที่ 4-12 อัตราการบ่งชี้ผิดพลาดเฉลี่ย (%) ที่จำนวนสถานะค่าต่างๆ.....	73
รูปที่ 4-13 ค่า FA และ FR เฉลี่ย ที่จำนวนสถานะค่าต่างๆ	74
รูปที่ 4-14 จำนวนเสียงผิดพลาดที่ได้จากแบบจำลองในการฝึกฝนแต่ละครั้ง สำหรับชุดทดสอบที่ 2.....	76
รูปที่ 4-15 ค่าคะแนนของเสียงพูดในกลุ่มทดสอบที่ 2 ที่ถูกบ่งชี้เป็นเสียงของผู้พูด ในระบบคนที่ 7	77
รูปที่ 4-16 ค่าคะแนนของเสียงพูดของผู้พูดที่เกี่ยวข้องกับผู้พูดคนที่ 7.....	78
รูปที่ 4-17 ค่าคะแนนของเสียงพูดในกลุ่มทดสอบที่ 2 ที่ถูกบ่งชี้เป็นเสียงของผู้พูดในระบบคนที่ 7 เมื่อใช้ฟังก์ชัน $df(1,1)$	78
รูปที่ 4-18 จำนวนเสียงยอมรับและปฏิเสธผิดพลาดเฉลี่ย ที่ค่า m และ n ค่าต่างๆ	81
รูปที่ 4-19 จำนวนเสียงผิดพลาด	83
รูปที่ 4-20 จำนวนเสียงบ่งชี้ผิดพลาดเมื่อให้จำนวนผู้พูดในระบบจำนวนต่างๆ (1).....	85
รูปที่ 4-21 จำนวน FA และ FR เมื่อให้จำนวนผู้พูดในระบบจำนวนต่างๆ (1).....	86
รูปที่ 4-22 จำนวนเสียงบ่งชี้ผิดพลาดเมื่อให้จำนวนผู้พูดในระบบจำนวนต่างๆ (2).....	87
รูปที่ 4-23 จำนวน FA และ FR เมื่อให้จำนวนผู้พูดในระบบจำนวนต่างๆ (2).....	88
รูปที่ 4-24 อัตราการบ่งชี้ผิดพลาดจากการทดลองที่ 4-6-1 และการทดลองที่ 4-6-2.....	89
รูปที่ 4-25 ค่า FA และ FR จากการทดลองที่ 4-6-1 และการทดลองที่ 4-6-2.....	89

สารบัญคำศัพท์

ขั้นตอนวิธีการ	algorithm
อัตสหสัมพันธ์	autocorrelation
การค้นหาแบบทวิภาค	binary search
จุดศูนย์กลาง	centroid
สัมประสิทธิ์เซปสตรอล	cepstral coefficient
การแบ่งกลุ่ม	clustering
ต้นแบบอ้างอิงชุดรหัส	codebook reference template
เวกเตอร์รหัส	code vector
ชุดรหัส	codebook
เสียงพูดต่อเนื่อง	continuous speech
ความผิดเพี้ยน	distortion
ลักษณะสำคัญ	feature
การสกัดลักษณะสำคัญ	feature extraction
การค้นหาทั่วทั้งหมด	full search
ค่าที่เหมาะสมที่สุดที่ครอบคลุมทั้งหมด	global optimum
แบบจำลองฮิดเดนมาร์คอฟ	Hidden Markov Model
อัตราการบ่งชี้ผู้พูด	identification rate
ผู้พูดนอกระบบ	imposter
อุปนัย	induction
คำโดด	isolated word
วนซ้ำ	iterative
ขั้นตอนวิธีการแบ่งเฉลี่ย K ส่วน	K-mean algorithm
การประมาณพหุระเชิงเส้น	linear prediction
สัมประสิทธิ์การประมาณพหุระเชิงเส้น	linear prediction coefficient
การเข้ารหัสโดยการประมาณพหุระเชิงเส้น	linear prediction coding
ค่าที่เหมาะสมที่สุดเฉพาะแห่ง	local optimum
ลูกโซ่มาร์คอฟ	Markov chain
ค่าความผิดพลาดกำลังสองเฉลี่ย	mean-square error
กฎการเลือกบริเวณที่ใกล้เคียงที่สุด	nearest neighbour rule
ทำให้เป็นบรรทัดฐานเดียวกัน	normalize

ลำดับคำสังเกต	observation sequence
ลำดับ	order
รูปแบบ	pattern
การเน้นล่วงหน้า	preemphasis
กรรมวิธีประมวลผลเบื้องต้น	preprocessing
ฟังก์ชันความหนาแน่นความน่าจะเป็น	probability density function
การควอนไทซ์	quantization
การรู้จำ	recognition
สัมประสิทธิ์การสะท้อน	reflection coefficient
การประมวลผลสัญญาณเบื้องต้น	signal preprocessing
ปริภูมิ	space
การบ่งชี้ผู้พูด	speaker identification
การรู้จำผู้พูด	speaker recognition
การตรวจสอบผู้พูด	speaker verification
เชิงสเปกตรัม	spectral
เสียงพูด	speech
การเข้ารหัสเสียงพูด	speech coding
การบีบอัดเสียงพูด	speech compression
สถานะ	state
ค่าผ่านระดับ	threshold
การฝึกฝน	training
การควอนไทซ์แบบเวกเตอร์	vector quantization

บทที่ 1
บทนำ



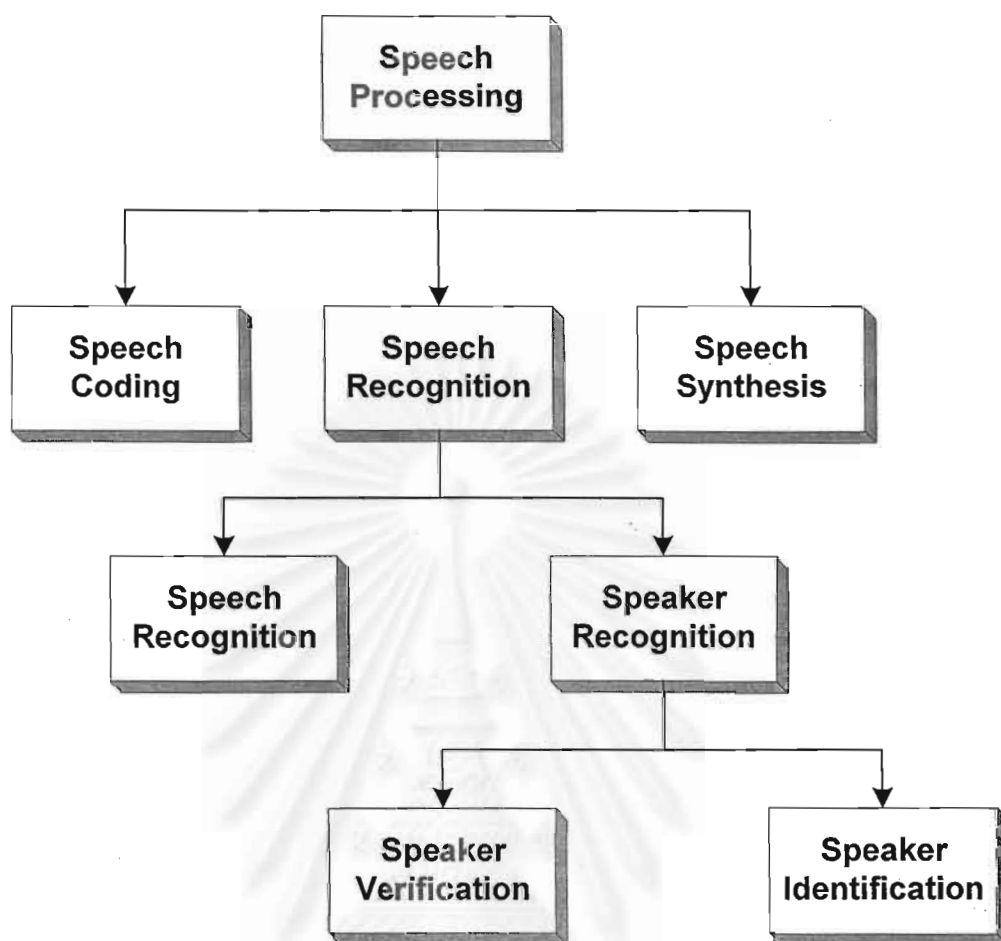
ที่มาของปัญหา

เมื่อโลกเริ่มก้าวสู่ปีสหัสวรรษ 2000 ซึ่งเป็นปีที่ปฏิเสธไม่ได้ว่าเทคโนโลยีต่างๆ จะมีอิทธิพลต่อการดำเนินชีวิตเป็นอย่างมาก ทุกแขนงสาขาอาชีพจะเกี่ยวข้องกับเทคโนโลยี โดยเฉพาะอย่างยิ่งอุปกรณ์ที่เรียกว่า คอมพิวเตอร์ ซึ่งนับวันก็จะยังมีบทบาทต่อการดำเนินชีวิตของมนุษย์เป็นอย่างมากเนื่องจากคอมพิวเตอร์สามารถประยุกต์ใช้งานได้หลายรูปแบบ

ในทางตรงกันข้าม ยิ่งความเจริญทางเทคโนโลยีเพิ่มมากขึ้น ความปลอดภัยในการดำเนินชีวิตก็ลดน้อยลง การนำเอาคอมพิวเตอร์มาประยุกต์ใช้ในการรักษาความปลอดภัยในบ้านเรือนที่อยู่อาศัย ในสำนักงาน หรือในสถานที่สำคัญๆ จึงเป็นรูปแบบหนึ่งของระบบรักษาความปลอดภัยที่เริ่มนำมาใช้กันอยู่ในปัจจุบัน รูปแบบการรักษาความปลอดภัยที่ใช้กันอยู่ในปัจจุบันเป็นรูปแบบพื้นฐาน เช่น การใช้กุญแจ การใช้บัตรผ่าน(Key Card) หรือการจ้างรปภ. ซึ่ง 2 รูปแบบแรกสามารถทำการปลอมแปลงได้ง่าย ส่วนรูปแบบหลังก็ต้องเสียค่าใช้จ่ายมากในการจ้าง ดังนั้นจึงเกิดแนวความคิดที่จะนำเอาคอมพิวเตอร์มาประยุกต์ใช้ ซึ่งก็มีหลายรูปแบบ เช่น การใช้คอมพิวเตอร์รู้จำเสียง ใบหน้า ลายมือ หรือนัยน์ตาของมนุษย์ โดยสิ่งต่างๆ เหล่านี้ถือว่าเป็นเอกลักษณ์ของแต่ละคนยากต่อการลอกเลียนแบบ

นอกจากรูปแบบความปลอดภัยที่นำไปใช้ในบ้านที่อยู่อาศัย หรืออาคารสำนักงานแล้ว ระบบความปลอดภัยในการใช้งานคอมพิวเตอร์ก็มีความสำคัญเป็นอย่างยิ่ง ทั้งนี้ก็เพราะว่าปัจจุบันคอมพิวเตอร์ส่วนใหญ่จะมีการใช้งานเป็นแบบเครือข่าย ซึ่งจะมีการใช้อุปกรณ์หรือแม่กระทั่งตัวคอมพิวเตอร์เองร่วมกัน ดังนั้นจึงจำเป็นต้องมีระบบรักษาความปลอดภัยของเครือข่ายที่น่าเชื่อถือ ซึ่งปัจจุบันนิยมใช้รหัสผ่าน หรือ Password โดยการเข้ารหัสผ่านนั้นถ้าบุคคลอื่นสามารถทราบรหัสผ่านของเราไม่ว่าจะด้วยวิธีการใด ก็สามารถทำให้บุคคลนั้นสามารถเข้าระบบคอมพิวเตอร์ได้ ซึ่งนับว่ามีอันตรายเป็นอย่างยิ่งหากผู้ใช้นั้นมีจุดประสงค์ร้ายกับระบบคอมพิวเตอร์อาจทำให้ข้อมูลสำคัญๆ หรือแม้แต่กระทั่งคอมพิวเตอร์ หรือ ทั้งเครือข่ายเสียหายได้ ดังนั้นการนำสิ่งที่เป็นเอกลักษณ์ของแต่ละบุคคลมาใช้ในร่วมกับระบบปลอดภัยเดิมที่มีอยู่ก็สามารถเพิ่มความน่าเชื่อถือของระบบความปลอดภัยได้

สำหรับในงานวิทยานิพนธ์นี้จะมุ่งความสนใจไปที่งานทางด้านการรู้จำเสียงพูดหรือ Speaker Recognition ซึ่งเป็นงานแขนงหนึ่งของงานทางด้านการประมวลผลเสียงพูด (Speech Processing) โดยงานทางด้านการประมวลผลเสียงพูดนี้สามารถแสดงความสัมพันธ์ของงานแขนงต่างๆ ดังรูปที่ 1-1 (Campbell, 1997)



รูปที่ 1-1 ความสัมพันธ์ของงานวิจัยด้านการประมวลผลเสียงพูด (Campbell, 1997)

สำหรับงานทางด้านความรู้จำเสียง ในต่างประเทศได้มีการศึกษาค้นคว้ากันเป็นเวลานานหลายปีแล้วแต่สำหรับในประเทศไทย ก็ได้มีการเริ่มศึกษากันอย่างจริงจังเมื่อประมาณ 10 ปีที่แล้ว (วิสิทธิ์ ถีลาศิริวงศ์, 2535) (ณัฐกร ทับทอง, 2538) (เสาวลักษณ์ อารีพงศา, 2538) (วิสรุต อาบุญตร, 2539) (เอกฤทธิ์ มณีน้อย, 2541) โดยงานส่วนใหญ่ก็จะเป็นงานทางด้านความรู้จำคำพูดหรือเสียงพูดภาษาไทย (Thai Speech Recognition) ซึ่งก็ได้มีการพัฒนาเทคนิคต่างๆ มาเรื่อยจนถึงปัจจุบันก็มีการศึกษาค้นคว้ากันอยู่ แต่สำหรับงานทางด้านความรู้จำผู้พูดนั้น ยังไม่ได้รับการพัฒนาเท่าไรนัก แต่จากการศึกษาบทความ งานวิจัยจากต่างประเทศจะเห็นได้ว่าเทคนิคที่นำมาใช้ในการรู้จำเสียงผู้พูดจะเป็นเทคนิคเดียวกันกับที่ใช้ในการรู้จำคำพูด ดังนั้นจึงสามารถนำเอาเทคนิคในการรู้จำมาใช้ด้วยกันได้ โดยเทคนิคต่างๆ ดังกล่าวที่นิยมใช้กัน ได้แก่

1. DTW (Dynamic Time Warping)
2. Vector Quantization
3. Artificial Neural Network
4. Hidden Markov Model

ในงานวิจัยเกี่ยวกับการรู้จำคำพูด ได้มีการพัฒนานำเอาแบบจำลองฮิดเดนมาร์คอฟมาใช้ในการรู้จำกันอย่างต่อเนื่อง เช่น การรู้จำเสียงพูดตัวเลขเป็นภาษาไทยแบบไม่ขึ้นกับผู้พูดโดยวิธีฮิดเดนมาร์คอฟโมเดลและเวกเตอร์ควอนไทซ์เซชัน (เสาวลักษณ์ อารีพงศา, 2538) และระบบคำไทยหลายพยางค์แบบไม่ขึ้นกับผู้พูดโดยใช้แบบจำลองฮิดเดนมาร์คอฟ (วิศรุต อาชุนบุตร , 2539) เป็นต้น นอกจากนี้งานด้านการรู้จำผู้พูดก็ได้เริ่มมีการพัฒนา เช่น การระบุผู้พูดด้วย LPC และ DTW สำหรับภาษาไทย ของศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (คณะนักวิจัยและพัฒนา ระบบระบุผู้พูดสำหรับภาษาไทย, 2541) ระบบการบ่งชี้ผู้พูดแบบขึ้นกับบทคำพูดโดยใช้แบบจำลองฮิดเดนมาร์คอฟ (ศวิต กาสुरิยะ, 2542) เป็นต้น

จากงานวิจัยที่ผ่านมา จะเห็นได้ว่าได้มีการนำเอาแบบจำลองฮิดเดนมาร์คอฟมาใช้งานกันอย่างแพร่หลาย ทั้งนี้ก็เนื่องจากว่าแบบจำลองฮิดเดนมาร์คอฟเป็นแบบจำลองทางสถิติ (Statistical Model) ดังนั้นจึงเหมาะสำหรับการนำมาใช้สร้างแบบจำลองของสัญญาณที่มีคุณสมบัติทางสถิติที่คงที่ได้ดี เช่น สัญญาณเสียง เป็นต้น

จากงานวิจัย ระบบการบ่งชี้ผู้พูดแบบขึ้นกับบทคำพูดโดยใช้แบบจำลองฮิดเดนมาร์คอฟ (ศวิต กาสुरิยะ, 2542) ซึ่งเป็นการบ่งชี้ผู้พูดจากระบบปิด (Closed Set) กล่าวคือ ระบบจะสามารถบ่งชี้ได้เฉพาะบุคคลภายในกลุ่ม หรือบุคคลที่ได้ลงทะเบียนให้ระบบได้รู้จำเรียบร้อยแล้วเท่านั้น ไม่สามารถตรวจสอบบุคคลนอกกลุ่มได้ นอกจากนี้งานวิจัยของศวิต กาสुरิยะ (ศวิต กาสुरิยะ, 72542) มีการใช้แบบจำลองฮิดเดนมาร์คอฟแบบคิสคริต ซึ่งต้องมีการใช้ชุดรหัสอ้างอิง (Reference Codebook) โดยชุดรหัสอ้างอิงนี้ได้มาจากการฝึกฝน โดยการฝึกฝนชุดรหัสนี้ใช้วิธีการฝึกฝนแบบชุดรหัสเดี่ยว (Single Codebook) คือ นำเสียงพูดฝึกฝนจากผู้พูดทุกคนในระบบมาฝึกฝนเพื่อให้ได้ชุดรหัสอ้างอิงหนึ่งชุด เพราะฉะนั้นในระบบก็จะมีชุดรหัสอ้างอิงเพียง 1 ชุดเท่านั้น

สำหรับในงานวิทยานิพนธ์นี้จะพัฒนาระบบบ่งชี้ผู้พูดต่อจากระบบของศวิต กาสुरิยะ (2542) โดยจะพัฒนาเพิ่มเติม 2 จุดใหญ่ๆ คือจุดแรก พัฒนาการบ่งชี้ผู้พูดจากระบบปิดไปเป็นแบบระบบเปิด คือสามารถตรวจสอบผู้พูดนอกระบบได้ด้วย จุดที่สอง คือเปลี่ยนรูปแบบของชุดรหัสอ้างอิงจากแบบเดิมคือ แบบชุดรหัสเดี่ยวไปเป็นแบบหลายชุดรหัส คือ ชุดรหัสที่สร้างจากผู้พูดแต่ละคน นั่นคือ ผู้พูดแต่ละคนจะมีชุดรหัส 1 ชุดรหัส ดังนั้นจำนวนชุดรหัสก็จะมีเท่ากับจำนวนผู้พูดในระบบ ซึ่งการใช้ชุดรหัสแบบหลายชุดรหัสนี้จะทำให้สะดวกต่อการฝึกฝนระบบเมื่อมีการเพิ่มผู้พูดเข้ามาในระบบ โดยจะมีการฝึกฝนเฉพาะเสียงของผู้พูดคนใหม่เท่านั้น ซึ่งแตกต่างจากแบบชุดรหัสเดี่ยวที่ต้องสร้างระบบคือฝึกฝนระบบใหม่ทั้งหมดเลยทำให้เสียเวลามาก

วัตถุประสงค์

- 1.1 เพื่อพัฒนาระบบบ่งชี้ผู้พูดแบบระบบเปิดที่ขึ้นกับบทคำพูดและเป็นเสียงพูดต่อเนื่อง
- 1.2 เพื่อนำวิธีการฝึกฝนหูรหัสแบบหลายหูรหัสมาใช้ในการสร้างหูรหัส
- 1.3 เพื่อนำแบบจำลองฮิดเดนมาร์คอฟแบบคิสคริตมาประยุกต์ใช้ในการรู้จำผู้พูด

ขั้นตอนการดำเนินงาน

ในการดำเนินการเพื่อสร้างระบบบ่งชี้ผู้พูดแบบระบบเปิด ได้วางแผนในการดำเนินการดังต่อไปนี้

1. ศึกษา ค้นคว้าข้อมูลดังต่อไปนี้
 - ศึกษางานวิจัยที่เกี่ยวกับระบบรู้จำผู้พูด
 - ศึกษาวิธีการต่างๆ ที่นำมาใช้ในการรู้จำผู้พูด
 - ศึกษาแบบจำลองฮิดเดนมาร์คอฟแบบคิสคริต
2. พัฒนาระบบบ่งชี้ผู้พูด
 - เก็บตัวอย่างเสียงพูด
 - พัฒนาโปรแกรมที่นำแบบจำลองฮิดเดนมาร์คอฟแบบคิสคริตไปใช้ในการฝึกฝนระบบและรู้จำผู้พูด
 - สร้างระบบบ่งชี้ผู้พูด โดยใช้แบบจำลองฮิดเดนมาร์คอฟแบบคิสคริต
 - หาค่าตัวแปรต่างๆของระบบที่เหมาะสม (ที่ทำให้ได้อัตราการบ่งชี้ที่ดีที่สุด) ได้แก่ ลักษณะสำคัญ จำนวนขนาดของหูรหัส จำนวนสถานะของแบบจำลองฮิดเดนมาร์คอฟ (State) เพื่อนำไปใช้ในการสร้างส่วนตรวจสอบผู้พูดต่อไป
 - สร้างส่วนตรวจสอบผู้พูดโดยการใช้ค่าตัวแปรของระบบตามที่ได้จากระบบบ่งชี้ผู้พูด
3. ทดสอบระบบจากกลุ่มทดสอบที่กำหนดไว้ และวิเคราะห์ผลการทดลองที่ได้ รวมทั้งแก้ไขและปรับปรุงในส่วนที่บกพร่อง
4. สรุปรวบรวมผลการวิจัยทั้งหมด รวมทั้งจัดทำเอกสารเกี่ยวกับวิทยานิพนธ์

เป้าหมายและขอบเขตของงานวิจัย

1. สร้างระบบบ่งชี้ผู้พูดแบบระบบเปิดที่ขึ้นกับบทคำพูด ที่มีผู้พูดในระบบอย่างน้อย 10 คน
2. ระบบสามารถบ่งชี้ผู้พูดได้ด้วยอัตราการบ่งชี้ผิดพลาดต่ำกว่า 5 %

ประโยชน์ที่คาดว่าจะได้รับ

เมื่อระบบบ่งชี้ผู้พูดในงานวิทยานิพนธ์นี้ได้รับการพัฒนา สามารถนำไปประยุกต์ใช้งานกับงานต่างๆ ได้หลายงาน เช่น งานการรู้จำผู้พูดในสำนักงาน โดยนำระบบรู้จำผู้พูดไปใช้แทนระบบคอกบัตรเพื่อลงเวลาเข้า-ออก ทำให้สะดวกมากขึ้น อีกทั้งยังสามารถป้องกันการลักลอบเข้ามาของบุคคลภายนอกได้ หรือการประยุกต์ใช้งานกับระบบคอมพิวเตอร์ โดยการใช้เสียงพูดแทนรหัสผ่าน ซึ่งสามารถป้องกันการลักลอบเข้าระบบได้

จะเห็นว่าการประยุกต์ใช้มีหลายรูปแบบ ดังนั้นถ้าระบบรู้จำผู้พูดได้รับการพัฒนาให้มีประสิทธิภาพก็จะเป็นประโยชน์และสามารถนำไปใช้กับงานต่างๆ ได้



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 2

ทฤษฎีและงานที่เกี่ยวข้อง

ในบทนี้จะได้กล่าวถึงทฤษฎีและความรู้ต่างๆที่นำมาใช้ในงานวิทยานิพนธ์ โดยเริ่มจากการทำความเข้าใจให้มากขึ้นเกี่ยวกับระบบการรู้จำผู้พูด (Speaker Recognition) ว่าระบบมีขั้นตอนการทำงานอย่างไร ประกอบด้วยระบบย่อยอะไรบ้าง

สำหรับระบบรู้จำผู้พูดเป็นหนึ่งในระบบรู้จำเสียงพูด โดยระบบรู้จำผู้พูดเป็นระบบที่ทำการรู้จำเสียงของผู้พูด สามารถแบ่งออกได้เป็น 2 ประเภท คือ ระบบบ่งชี้ผู้พูด (Speaker Identification) กับระบบตรวจสอบผู้พูด (Speaker Verification) โดยสามารถอธิบายการทำงานของแต่ละระบบได้ดังนี้

1. ระบบบ่งชี้ผู้พูด (Speaker Identification System)

ระบบบ่งชี้ผู้พูดนี้จะทำหน้าที่ระบุว่าเสียงพูดที่เข้ามาในระบบเป็นเสียงพูดของบุคคลใดในระบบที่ได้ลงทะเบียนไว้ โดยระบบจะเปรียบเทียบกับตัวอย่างเสียงพูดที่รับเข้ามากับเสียงอ้างอิงที่ระบบได้รู้จำเรียบร้อยแล้ว ซึ่งระบบจะบ่งชี้ผู้พูดจากเสียงอ้างอิงที่ให้ค่าระยะทางของการเปรียบเทียบต่ำสุด สำหรับการประยุกต์ใช้งานของระบบบ่งชี้ผู้พูด เช่น การรู้จำเสียงผู้พูดในการเข้าใช้งานฐานข้อมูลคอมพิวเตอร์ การเข้าระบบคอมพิวเตอร์ และการนำมาใช้กับงานระบบรักษาความปลอดภัยในอาคารบ้านเรือน เป็นต้น ในระบบบ่งชี้ผู้พูดสามารถแบ่งออกเป็น 2 ประเภท ได้แก่

1.1 ระบบบ่งชี้ผู้พูดแบบระบบปิด (Closed Set Speaker Identification) ระบบบ่งชี้ผู้พูดแบบนี้จะระบุว่าเสียงพูดที่เข้ามาในระบบเป็นเสียงพูดของบุคคลใดในระบบ โดยระบบจะบังคับว่าคำตอบของระบบจะเป็นเสียงของบุคคลใดบุคคลหนึ่งในระบบ (ถ้าระบบมีจำนวนผู้พูดในระบบ N คน คำตอบของระบบจะมีได้ N คำตอบ คือ $1, 2, \dots, N$) (Farrell, K., et al, 1994)

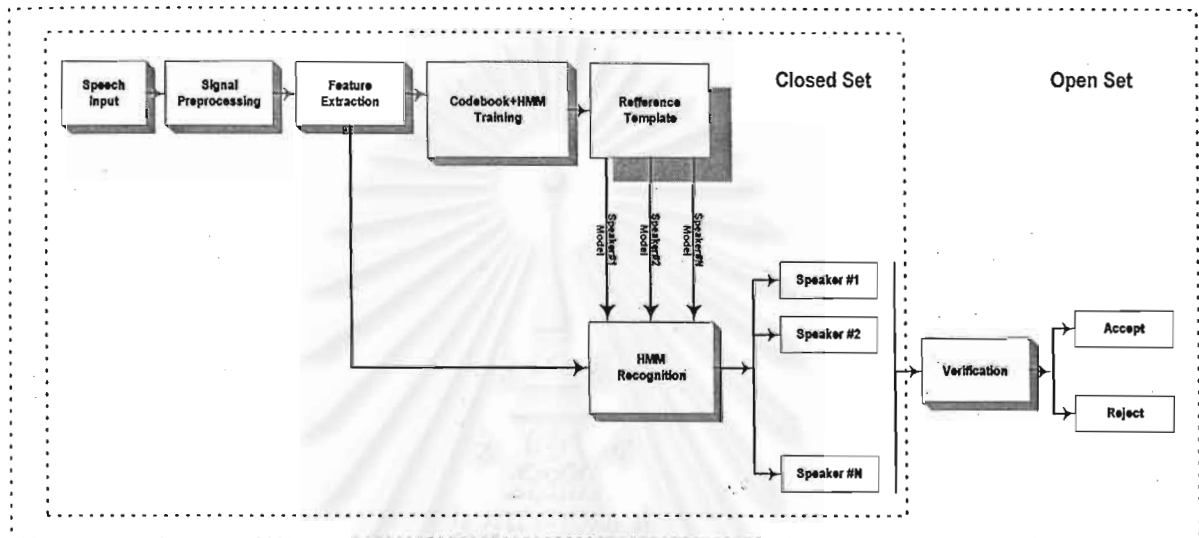
1.2 ระบบบ่งชี้ผู้พูดแบบระบบเปิด (Open Set Speaker Identification) ระบบบ่งชี้ผู้พูดแบบนี้จะระบุว่าเสียงพูดที่เข้ามาในระบบเป็นเสียงพูดของบุคคลใดในระบบ แต่ก่อนที่ระบบจะตัดสินใจว่าเป็นเสียงของผู้พูดคนใด จะนำเสียงพูดที่เข้ามาในระบบไปผ่านขั้นตอนการตรวจสอบ (Verification) เพื่อตรวจสอบว่าเสียงที่ระบบบ่งชี้ขึ้นเป็นเสียงของบุคคลดังกล่าวจริงหรือไม่ (ถ้าระบบมีจำนวนผู้พูดในระบบ N คน คำตอบจะมี $N+1$ คำตอบ คือ $1, 2, \dots, N$ และ ไม่ใช่ผู้พูดในระบบ)

2. ระบบตรวจสอบผู้พูด (Speaker Verification System)

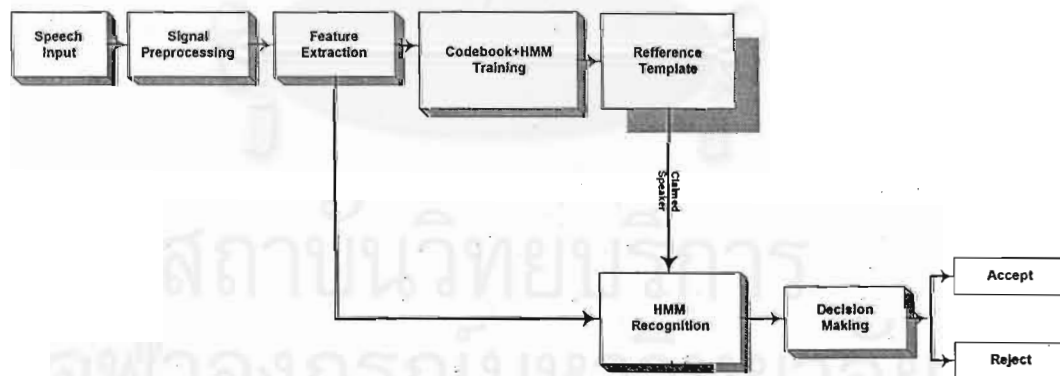
ระบบตรวจสอบผู้พูดนี้ จะทำหน้าที่ตรวจสอบว่าเสียงพูดที่เข้ามาในระบบเป็นเสียงพูดของบุคคลที่ได้อ้างไว้จริงหรือไม่ โดยระบบจะเปรียบเทียบค่าระยะทางระหว่างเสียงพูดที่เข้ามาใน

ระบบกับเสียงพูดอ้างอิง ถ้าค่าระยะทางต่ำกว่าค่าผ่านระดับ (Threshold) ที่ระบบได้กำหนดไว้แล้ว ระบบก็จะยอมรับเสียงพูดนั้น แต่ถ้าค่าระยะทางสูงกว่าค่าผ่านระดับ ระบบก็จะปฏิเสธเสียงนั้น สำหรับตัวอย่างการประยุกต์ใช้งาน ได้แก่ ระบบรักษาความปลอดภัยในตัวอาคาร หรือบ้าน โดยอาจนำไปใช้งานร่วมกับระบบรักษาความปลอดภัยที่ใช้ระบบบัตรประจำตัว (ID Card) การใช้งานในด้านระบบฐานข้อมูลคอมพิวเตอร์ เป็นต้น

โดยสามารถแสดงการทำงานของทั้งสองระบบได้ดังรูปที่ 2-1 และ 2-2



รูปที่ 2-1 แผนภาพการทำงานของระบบบ่งชี้ผู้พูดแบบระบบปิดและแบบระบบเปิด โดยใช้แบบจำลองฮิดเดนมาร์คอฟ



รูปที่ 2-2 การทำงานของระบบตรวจสอบผู้พูดโดยใช้แบบจำลองฮิดเดนมาร์คอฟ

สำหรับการทำงานของแต่ละส่วนสามารถอธิบายรายละเอียดของแต่ละส่วนได้ดังนี้

1. ส่วนรับสัญญาณเสียงพูด (Speech Input)

เป็นขั้นตอนการรับสัญญาณเสียงพูดจากภายนอกผ่านการแปลงจากสัญญาณอนาล็อกเป็นสัญญาณดิจิทัล แล้วนำมาผ่านกระบวนการควอนไทเซชันเพื่อทำให้เป็นสัญญาณดิจิทัล

2. การประมวลผลสัญญาณเบื้องต้น (Signal Preprocessing)

การประมวลผลสัญญาณเบื้องต้นเป็นขั้นตอนกรรมวิธีในการจัดเตรียมข้อมูล จากข้อมูลดิบของเสียงพูดที่ได้จากการบันทึกเสียงนำมาผ่านกรรมวิธีประมวลผลสัญญาณเชิงเลข เพื่อใช้ในการประมวลผลในขั้นตอนต่อไป เนื่องจากสัญญาณเสียงพูดโดยรวมจะแปรเปลี่ยนตามเวลา (Time Variant) และไม่เสถียร (Unstable) ดังนั้นในการประยุกต์ใช้งานกรรมวิธีประมวลผลสัญญาณเชิงเลขกับสัญญาณเสียงพูด จะแบ่งสัญญาณเสียงพูดออกเป็นช่วงย่อยๆ (Rabiner and Levinson 1981; Furui, 1989) เรียกว่า “กรอบเสียงพูด” (Speech Frame) โดยแต่ละกรอบเสียงพูดจะมีความยาวประมาณ 10-40 มิลลิวินาที ซึ่งถือได้ว่าสัญญาณเสียงพูดในแต่ละกรอบเสียงพูดมีความเสถียรและไม่แปรเปลี่ยนตามเวลา (Time Invariant) จากนั้นจึงสามารถทำการประมวลผลสัญญาณเชิงเลขกับสัญญาณในแต่ละกรอบเสียงพูดได้

กรรมวิธีในการประมวลผลสัญญาณเบื้องต้นที่นำมาใช้ในการวิเคราะห์เสียงพูดมีหลายกรรมวิธี สำหรับกรรมวิธีที่ใช้ในงานวิทยานิพนธ์นี้ประกอบด้วย 2 กรรมวิธีด้วยกันคือ กรรมวิธีเน้นล่งหน้า (Preemphasis) และกรรมวิธีวางกรอบสัญญาณ (Smoothing Window)

2.1 กรรมวิธีเน้นล่งหน้า (Preemphasis)

ขั้นตอนการเน้นล่งหน้าเป็นการบีบอัดช่วงพิสัยพลวัต (Dynamic Range) ของสัญญาณเสียงพูด โดยการให้ความลาดเอียงในเชิงความถี่แบบราบลงซึ่งจะส่งผลให้อัตราส่วนสัญญาณต่อสัญญาณรบกวนมีค่าสูงขึ้น ในทางปฏิบัติแล้วจะนำสัญญาณผ่านตัวกรองเชิงเลขอันดับหนึ่ง (First Order Digital Filter) ที่มีฟังก์ชันถ่ายโอนดังแสดงในสมการที่ (2-1) และ (2-2) (Furui, 1985) เมื่อ a เป็นสัมประสิทธิ์ตัวกรอง, $\tilde{s}(n)$ เป็นค่าของสัญญาณเสียงพูดขาออกที่ผ่านกรรมวิธีเน้นล่งหน้าที่ n , $s(n)$ เป็นค่าของสัญญาณเสียงพูดขาเข้าที่ n และ $s(n-1)$ เป็นค่าของสัญญาณเสียงพูดขาเข้าค่าก่อนหน้าที่ $n-1$ ดังสมการ

$$H(z) = 1 - az^{-1} \quad (2-1)$$

$$\tilde{s}(n) = s(n) - as(n-1) \quad (2-2)$$

โดยกำหนดให้ค่าสัมประสิทธิ์ของตัวกรอง a มีค่าเข้าใกล้ 1 เมื่อใช้ร่วมกับการวิเคราะห์หาค่าสัมประสิทธิ์การประมาณพหุเชิงเส้นจะกำหนดให้ค่า $a = 0.95$ เนื่องจากเป็นค่าที่ให้ผลดีที่สุดสำหรับการวิเคราะห์ (วิศรุต อาบุบุตร, 2538) (เสาวลักษณ์ อารีพงศา, 2538) (เอกฤทธิ์ มณีน้อย, 2541)

2.2 กรรมวิธีการวางกรอบขนาดสัญญาณ (Smoothing Window)

เป็นการนำสัญญาณเสียงพูดมาแบ่งเป็นส่วนย่อยๆ โดยการคูณแต่ละค่าของสัญญาณเสียงพูดที่อยู่ในกรอบที่กำหนดไว้ด้วยค่าฟังก์ชันกรอบ (Window Function) ในงานวิจัยนี้ได้เลือกใช้ฟังก์ชันกรอบชนิด Hamming Window ดังแสดงในรูปที่ 2-3 มีผลทำให้เกิดการลดทอนแอมพลิจูดอย่างช้าๆ ที่บริเวณปลายแต่ละข้างของกรอบสัญญาณเสียงพูด เพื่อป้องกันการเปลี่ยนแปลงที่ไม่ต่อเนื่องที่บริเวณปลายของส่วนย่อยๆ สัญญาณเสียงพูดที่ผ่านขั้นตอนนี้จะได้เป็นข้อมูลเสียงพูดเพื่อนำไปใช้กับกรรมวิธีสัญญาณดิจิทัลต่อไป ซึ่งกรรมวิธีการวางกรอบขนาดสัญญาณจะเป็นไปตามสมการที่ (2-3) และ (2-4)

$$\tilde{x}_l(n) = x_l(n) \cdot w(n) \quad (2-3)$$

$$w(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right) \quad (2-4)$$

เมื่อ $l = 0, 1, \dots, L-1$ และ $n = 0, 1, \dots, N-1$

กำหนดให้ $x_l(n)$ คือค่าสัญญาณเสียงพูดของข้อมูลที่ n

$\tilde{x}_l(n)$ คือค่าสัญญาณเสียงพูดที่ผ่านกรรมวิธีการวางกรอบ

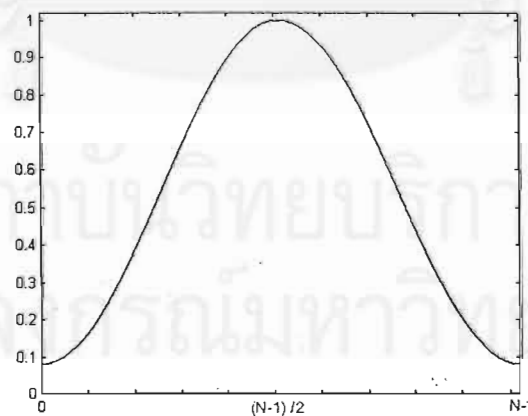
$w(n)$ คือฟังก์ชันกรอบชนิด Hamming Window

N คือจำนวนของข้อมูลในแต่ละกรอบสัญญาณเสียงพูด

n คือลำดับข้อมูลในกรอบสัญญาณเสียงพูดที่ l

L คือจำนวนของกรอบสัญญาณเสียงพูด

l คือลำดับของกรอบสัญญาณเสียงพูด



รูปที่ 2-3 ฟังก์ชันกรอบชนิด Hamming Window

3. การสกัดลักษณะสำคัญ (Feature Extraction)

การสกัดลักษณะสำคัญเป็นขั้นตอนในการหาค่าพารามิเตอร์ที่ใช้แทนตัวอย่างข้อมูล โดยพารามิเตอร์ที่ได้มาจะบอกถึงลักษณะสำคัญของเสียงพูดแต่ละเสียง ซึ่งจะเป็นสิ่งช่วยในการบ่งบอกความแตกต่างของเสียงพูด สำหรับค่าลักษณะสำคัญที่จะนำมาใช้ในการรู้จำผู้พูดนั้น ควรจะมีความแตกต่างกันเมื่อเป็นผู้พูดคนละคน และมีความคล้ายกันเมื่อเป็นผู้พูดคนเดียวกัน ค่าลักษณะสำคัญสามารถแบ่งออกได้เป็น 2 ระดับ ได้แก่ (Naik, J.M., 1990)

1. ลักษณะสำคัญระดับบน ได้แก่ สำเนียงท้องถิ่น ลักษณะการพูด อารมณ์ในขณะที่พูด และสภาพแวดล้อม เป็นต้น
2. ลักษณะสำคัญระดับล่าง ได้แก่ Pitch, Spectral magnitude, Formant frequency, Energy profile และอื่นๆ

การเลือกลักษณะสำคัญที่จะนำมาใช้ในการรู้จำผู้พูดนั้นควรจะมีคุณสมบัติดังต่อไปนี้

- ความสามารถในการแยกแยะผู้พูดแต่ละคนเมื่อคำนึงถึงความผิดพลาดที่เกิดจากการแปรปรวนของผู้พูดในกลุ่มที่กำลังพิจารณา
- ง่ายต่อการสกัดออกจากสัญญาณเสียง
- มีลักษณะเสถียรเชิงเวลา
- ยากต่อการลอกเลียนแบบจากบุคคลอื่น

จากการศึกษาพบว่า งานวิจัยหลายงาน (Gu, Y., and Thomas, T., 1999), (Liu, C., et al, 1996), (Isobe, T., and Takahashi, J., 1999), (Jacobs, T., and Setlur, A., 1994) ได้ใช้สัมประสิทธิ์การประมาณพหุระเชิงเส้น หรือ LPC เป็นลักษณะสำคัญ นอกจากนี้ก็มี สัมประสิทธิ์ Cepstral และสัมประสิทธิ์ MFCC (Mel Frequency Cepstral Coefficient) โดยใช้งานร่วมกับแบบจำลองฮิดเดนมาร์คอฟและใช้อัตราการรู้จำที่ค้ำกับงานด้านการรู้จำคำพูด (Abkupta, V., et. Al., 1998) ซึ่งเป็นงานที่มีความคล้ายคลึงกับงานการบ่งชี้ผู้พูด ณ ระดับหนึ่ง ดังนั้นในวิทยานิพนธ์นี้จะทำการทดสอบลักษณะสำคัญทั้งสามนี้ ซึ่งจะอธิบายวิธีการในการหาค่าลักษณะสำคัญทั้ง 3 แบบตามลำดับดังนี้

3.1 สัมประสิทธิ์การประมาณพหุระเชิงเส้น (Linear Prediction Coefficient)

คำว่า “การประมาณพหุระเชิงเส้น” หรือ Linear Prediction ถูกนำเสนอเป็นครั้งแรกโดย N.Weiner ในปี ค.ศ. 1966 โดยเทคนิคนี้ถูกนำมาใช้เป็นครั้งแรกกับการวิเคราะห์และการสังเคราะห์เสียงโดย Itakura ร่วมกับ Saito และ Atal ร่วมกับ Schoeder ในปี ค.ศ. 1968 (Furui, 1989) ความสำคัญของเทคนิคการประมาณพหุระเชิงเส้นนี้ก็คือการที่รูปและลักษณะสมบัติทางความถี่ของเสียงพูดสามารถแสดงด้วยค่าพารามิเตอร์เพียงไม่กี่ค่าได้อย่างแม่นยำและมีประสิทธิภาพ นอกจากนี้ค่าพารามิเตอร์ดังกล่าวยังสามารถคำนวณได้ง่ายอีกด้วย

หลักการพื้นฐานของแบบจำลองการประมาณพันระเชิงเส้น (Linear Prediction Model) คือการคำนวณหาค่าอนาคตของสัญญาณจากการประมาณการผสมเชิงเส้น (Linear Combination) ของค่าของสัญญาณก่อนหน้านั้น โดยสมมติให้สัญญาณที่ได้รับความสนใจเกิดจากแหล่งกำเนิดสัญญาณที่ถูกกระตุ้นจากวงจรกรองแบบเชิงเส้น ในงานวิจัยนี้สัญญาณที่สนใจคือสัญญาณเสียงพูดเพราะฉะนั้นจะกล่าวเฉพาะแบบจำลองสัญญาณเสียงพูดเท่านั้น แหล่งกำเนิดของสัญญาณก็คือลมที่พ่นผ่านช่องว่างระหว่างเส้นเสียง (Vocal Cord) ออกมาเป็นรายคาบหรือสัญญาณรบกวนที่เกิดจากช่องแคบของการหดตัวของช่องทางเดินเสียง และที่มีผลกระทบต่อช่องทางเดินเสียงส่วนบน การจำลองสัญญาณเสียงที่เปล่งออกมาสามารถจำลองได้ดังสมการที่ (2-5)

$$s(n) = \sum_{k=1}^p a_k s(n-k) + G \cdot u(n) \quad (2-5)$$

โดย

$s(n)$ เป็นสัญญาณขาออกค่าปัจจุบัน

$s(n-k)$ เป็นสัญญาณขาออกค่าอดีตที่ k

p เป็นอันดับของการประมาณ

a_k เป็นสัมประสิทธิ์การประมาณที่ k

G เป็นอัตราขยาย

$u(n)$ เป็นแหล่งกำเนิดสัญญาณ

โดยปกติค่า $u(n)$ เป็นค่าที่ไม่สามารถหาค่าได้ จึงละเลยไม่นำมาคิด ดังนั้นจะได้สมการการประมาณสัญญาณในปัจจุบัน ($\hat{s}(n)$) มีค่าขึ้นกับสัญญาณในอดีต p ค่าเท่านั้น ($\hat{s}(n-k)$) ดังแสดงในสมการ

$$\hat{s}(n) = \sum_{k=1}^p a_k \hat{s}(n-k) \quad (2-6)$$

ดังนั้นถ้าสามารถหาหาค่าของสัมประสิทธิ์ a_k ได้ ก็สามารถประมาณสัญญาณนั้นได้

วิธีการที่นำมาใช้ในการหาหาค่าของสัมประสิทธิ์ a_k มี 2 วิธี ได้แก่ วิธี MMSE กับวิธีวนซ้ำของ Levinson-Durbin โดยจะอธิบายการหาค่า a_k ทั้ง 2 วิธีตามลำดับดังนี้

1. วิธี MMSE (Minimum Mean Square Error)

กำหนดให้ความคลาดเคลื่อนที่เกิดจากประมาณค่าสัญญาณ $\hat{s}(n)$ จาก $s(n)$ ทำให้เกิดความผิดพลาด $e(n)$ ขึ้นตามสมการ

$$e(n) = s(n) - \hat{s}(n) = s(n) - \sum_{k=1}^p a_k \hat{s}(n-k) \quad (2-7)$$

และค่า Mean Square Error

$$E = \sum_n (e(n)^2) = \sum_n \left[s(n) - \sum_{k=1}^p a_k \hat{s}(n-k) \right]^2 \quad (2-8)$$

เมื่อคิด Minimum Mean Square Error หรือ

$$\frac{\partial E}{\partial a_i} = 0, \quad i=1,2,\dots,p \quad (2-9)$$

จะได้

$$\sum_n s(n-i)s(n) = \sum_{k=1}^p a_k \sum_n s(n-i) \cdot s(n-k) \quad i=1,2,\dots,p \quad (2-10)$$

เนื่องจากพจน์แรกเป็นอัตโนมัติสัมพันธ์ $R(i)$ ของ $s(n)$ ที่มีความยาวจำกัด สามารถเขียนสมการที่ (2-10) ใหม่ได้เป็นสมการที่ (2-11)

$$\sum_{k=1}^p a_k R(i-k) = R(i), \quad 1 \leq i \leq p \quad (2-11)$$

$$\text{เมื่อ} \quad R(i) = \sum_{n=i}^{N-1} s(n)s(n-i) \quad (2-12)$$

เมื่อแก้สมการที่ (2-10) ซึ่งประกอบไปด้วยสมการเชิงเส้น p สมการได้เป็นค่าสัมประสิทธิ์ a_k ที่ต้องการออกมา โดยเขียนสมการที่ (2-11) ให้อยู่ในรูปเมตริกซ์ $Ra = r$

$$\begin{bmatrix} r_0 & r_1 & \cdots & r_{p-1} \\ r_1 & r_0 & & \vdots \\ \vdots & & \ddots & r_1 \\ r_{p-1} & \cdots & r_1 & r_0 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_p \end{bmatrix} \quad (2-13)$$

จากสมการที่ (2-13) เห็นได้ว่าเมตริกซ์ R เป็นเมตริกซ์ชนิด Toeplitz ซึ่งเป็นเมตริกซ์ที่สมมาตรคือมีค่าในแนวทแยงมุมเท่ากันทั้งหมดทุกแนว ดังนั้นการคำนวณหาค่าสัมประสิทธิ์ a_k จะอาศัยวิธีการวนซ้ำของ Levinson-Durbin (Picone, 1996) (O'Shaughnessy, 1987) ซึ่งอธิบายในหัวข้อถัดไป

2. ขั้นตอนวิธีการวนซ้ำของ Levinson-Durbin

เป็นเทคนิคที่ใช้ในการคำนวณหาค่าสัมประสิทธิ์ของการประมาณพหุเชิงเส้น a_i ในเมตริกซ์ a เมื่อ $i=1,2,3,\dots,p$ และ p เป็นอันดับ (Order) ของการวิเคราะห์หาค่าสัมประสิทธิ์ของการประมาณพหุเชิงเส้นจากวิธีอัตโนมัติสัมพันธ์ มีขั้นตอนวิธีการดังต่อไปนี้

ทำการวนซ้ำขั้นตอนที่ 1 ถึง 4 ของค่า i แต่ละค่า เมื่อ $i=1,2,3,\dots,p$
 ขั้นตอนที่ 1 กำหนดค่าเริ่มต้น

$$E_0 = R(0) \quad \text{และ} \quad \alpha_0 = 0 \quad (2-14)$$

ขั้นตอนที่ 2 คำนวณหาค่าสัมประสิทธิ์การสะท้อน (Reflection coefficient)

$$k_i = \frac{R(i) - \sum_{j=1}^{i-1} \alpha_{i-1}(j)R(i-j)}{E_{i-1}} \quad (2-15)$$

เมื่อ $R(i)$ และ $R(i-j)$ คำนวณได้จากสมการที่ (2-12)

ขั้นตอนที่ 3 คำนวณหาค่าสัมประสิทธิ์ของการประมาณพหุเชิงเส้น

$$\alpha_i(i) = k_i \quad (2-16)$$

$$\alpha_i(j) = \alpha_{i-1}(j) - k_i \alpha_{i-1}(i-j) \quad (2-17)$$

โดยทำการวนซ้ำสมการที่ (2-16) และ (2-17) เมื่อ $j = 1, 2, 3, \dots, i-1$

ขั้นตอนที่ 4 คำนวณค่าผิดพลาดใหม่

$$E_i = (1 - k_i^2)E_{i-1} \quad (2-18)$$

เมื่อสิ้นสุดการวนซ้ำขั้นตอนทั้งสี่แล้ว เราจะได้ค่าสัมประสิทธิ์ของการประมาณพหุเชิงเส้นที่รอบสุดท้าย ($i = p$) คือ $\alpha_p(j)$ เมื่อ $j = 1, 2, 3, \dots, p$ นั่นคือค่าสัมประสิทธิ์ที่มีค่าความผิดพลาดของการประมาณพหุเชิงเส้นน้อยที่สุดนั่นเอง ดังในสมการที่ (2-19)

$$a_k = \alpha_p(k) \quad (2-19)$$

เมื่อ k คืออันดับของค่าสัมประสิทธิ์ของการประมาณพหุเชิงเส้น

3.2 สัมประสิทธิ์เซปสโตรอล (Cepstral Coefficients, CEP)

สัมประสิทธิ์เซปสโตรอลสามารถคำนวณหาได้หลายวิธี เช่น คำนวณจากการแปลงฟูรีเยร์ (Fourier Transform) หรือคำนวณจากแบบจำลองการประมาณพหุเชิงเส้น เป็นต้น ในงานวิจัยนี้ได้เลือกใช้ค่าสัมประสิทธิ์เซปสโตรอล (c_m) ที่คำนวณจากแบบจำลองการประมาณพหุเชิงเส้นเนื่องจากว่าง่ายต่อการคำนวณและใช้เวลาน้อยมาก วิธีการคำนวณหาสัมประสิทธิ์เซปสโตรอลแสดงในสมการที่ (2-20) และ (2-21) (Rabiner and Juang, 1993)

$$c_m = a_m + \sum_{k=1}^{m-1} \left(\frac{k}{m}\right) c_k a_{m-k}, \quad 1 \leq m \leq p \quad (2-20)$$

$$c_m = \sum_{k=1}^{m-1} \left(\frac{k}{m}\right) c_k a_{m-k}, \quad m > p \quad (2-21)$$

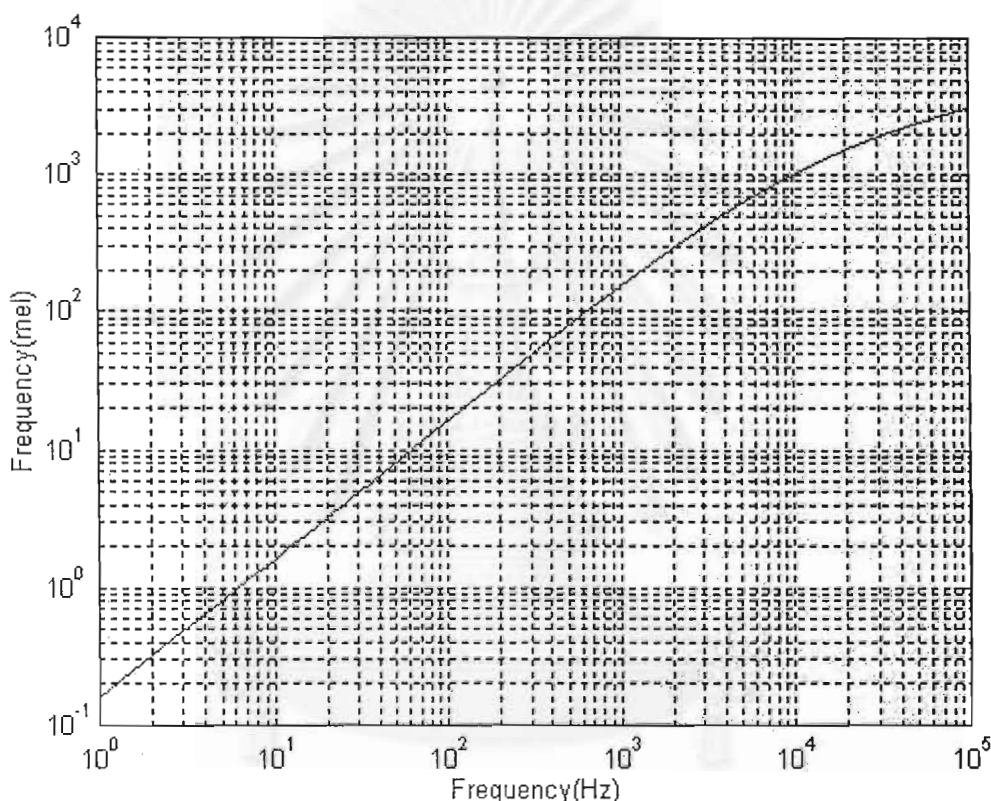
เมื่อ p คืออันดับของค่าสัมประสิทธิ์ของการประมาณพหุเชิงเส้น

3.3 สัมประสิทธิ์เซปสโตรอลบนความถี่เมล (Mel Frequency Cepstral Coefficients, MFCC)

สัมประสิทธิ์เซปสโตรอลบนความถี่เมลคำนวณได้จากการวิเคราะห์โดยไม่ใช้พารามิเตอร์ (Non-Parametric Analysis) คือไม่มีการสร้างแบบจำลองสัญญาณเพื่อปรับค่าพารามิเตอร์ในแบบจำลองให้ได้เป็นสัญญาณที่เราต้องการ แต่เป็นการนำสัญญาณผ่านวงจรกรองแบบผ่านแถบความถี่ (Band Pass Filter) หลายวงจร โดยแต่ละวงจรกรองมีช่วง

ความถี่ที่ผ่านได้แตกต่างกันดังแสดงในรูปที่ 2-4 และวงจรกรองที่ใช้เป็นชุดวงจรกรองแบบดิจิทัล (Digital Filter Bank) ที่ใช้วิเคราะห์สเปกตรัมของสัญญาณเสียงพูด โดยเลียนแบบตามการได้ยินของมนุษย์ สามารถทำได้หลายวิธีขึ้นอยู่กับกรรมวิธีที่นำมาใช้กับแกนความถี่ได้แก่ (Tolba and O'Shaughnessy, 1998) (Tuzun, Demirekler and Nakiboglu, 1994)

- Uniform Spacing (Fourier Transform)
- Exponential Spacing (Wavelet Transform)
- Perceptually-Derived Spacing (Mel Scale หรือ Bark Scale)



รูปที่ 2-4 Mel Scale ของความถี่สัญญาณเสียง

ในงานวิจัยนี้ได้เลือกใช้กรรมวิธี Perceptually-derived Spacing หรือเรียกอีกชื่อหนึ่งว่า Mel scale เป็นการเปลี่ยนแกนความถี่ของสัญญาณเสียง (Acoustic Frequency) ไปเป็นแกนความถี่ของการได้ยิน (Perceptual Frequency) จึงจะกล่าวถึงแต่กรรมวิธีนี้เท่านั้น Mel scale ใช้หลักการ Mapping จากการประมาณความถี่ช่วงที่เป็นเชิงเส้น (Linear Frequency Scale) เป็นความถี่ที่ไม่เชิงเส้น (Nonlinear Frequency Scale) ตามลักษณะการได้ยินของมนุษย์ดังแสดงในสมการที่ (2-22) (Tolba and O'Shaughnessy, 1998)

$$mel(f) = 2595 \log_{10} \left(1 + \frac{f}{700} \right) \quad (2-22)$$

เมื่อ f คือช่วงความถี่ที่ยังเป็นเชิงเส้น จากรูปที่ 2-4 เห็นได้ว่าลักษณะการได้ยินของมนุษย์ในช่วงที่เป็นเชิงเส้น (Linear scale) คือประมาณช่วงความถี่ตั้งแต่ 0 ถึง 1000 Hz และช่วงที่ไม่เป็นเชิงเส้น (Logarithmic scale) คือช่วงความถี่ตั้งแต่ 1000 Hz ขึ้นไป แบบนี้วิคท์ของวงจรรองสามารถคำนวณได้จากสมการที่ (2-23) ดังนี้

$$BW_{critical} = 25 + 75[1 + 1.4(f/1000)^2]^{0.69} \quad (2-23)$$

วิธีการคำนวณแบบจำลองของชุดวงจรรอง (Filter Bank Model) ที่ง่ายที่สุดและให้ผลที่มีประสิทธิภาพคือการแปลงแบบฟูริเยร์ (Fourier Transform) ของสัญญาณและหาข้อมูลออกของชุดวงจรรองของแต่ละวงจร (X_N) จากความถี่ที่ใส่เข้าไป การคำนวณหาข้อมูลออกของชุดวงจรรองสามารถคำนวณได้ตามสมการที่ (2-24)

$$X_N = \frac{1}{N_{FB}} \sum_{k=1}^{N_{FB}} \omega_{FB}(k) X(f + \delta f(f, k)) \quad (2-24)$$

เมื่อ N_{FB} คือจำนวนของตัวอย่างที่ใช้ในการหาค่าเฉลี่ย

ω_{FB} คือฟังก์ชันการถ่วงน้ำหนัก

$\delta f(f, k)$ คือฟังก์ชันที่อธิบายถึงความถี่ที่อยู่ใกล้เคียงกับความถี่ f เพื่อนำมาใช้คำนวณหาค่าเฉลี่ย

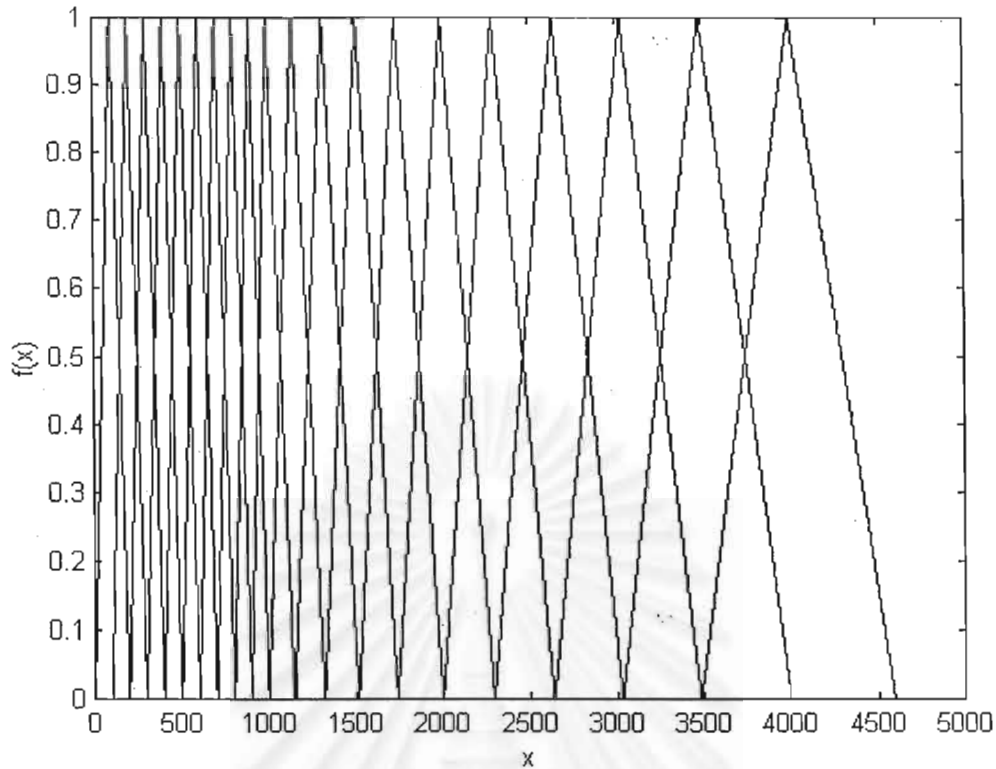
ในการคำนวณหาค่าสัมประสิทธิ์เซปสตรอลบนความถี่เมล ใช้ค่าข้อมูลออกที่ได้จากวงจรรองแถบผ่าน N วงจรเมื่อใส่ค่าล็อกของขนาดสเปกตรัมของสัญญาณเสียงเข้าไป เพื่อประมาณความถี่ตอบสนองของ Basilar Membrane ใน Cochlea ที่อยู่ในหูชั้นใน วงจรรองที่นำมาใช้นี้เป็นวงจรรองแถบผ่านแบบสามเหลี่ยมดังแสดงในรูปที่ 2-5 เพราะฉะนั้นการคำนวณหาค่าสัมประสิทธิ์เซปสตรอลบนความถี่เมล (C_n) สามารถหาได้จากสมการดังต่อไปนี้ (Tolba and O'Shaughnessy, 1998)

$$C_n = \sum_{k=1}^N X_k \cos\left(\frac{\pi n}{N}(k - 0.5)\right), \quad n = 1, 2, \dots, M \quad (2-25)$$

เมื่อ M คือจำนวนของสัมประสิทธิ์เซปสตรอล

N คืออันดับในการวิเคราะห์

X_k คือค่าล็อกของพลังงานที่ค่าข้อมูลออกของวงจรรอง k วงจร เมื่อ $k = 1, 2, \dots, N$

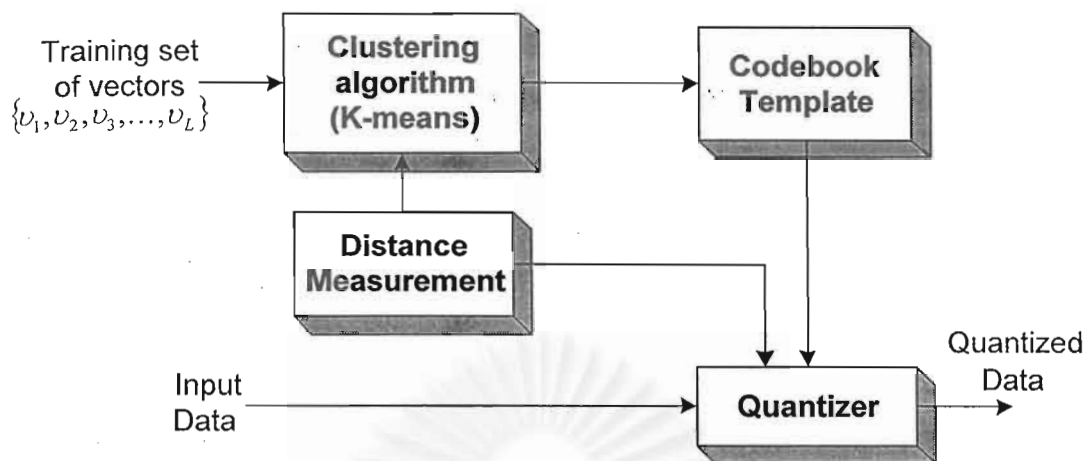


รูปที่ 2-5 วงจรกรองแบบผ่านแถบความถี่ที่ให้ค่าสัมประสิทธิ์เซปสตรอลบนความถี่เมล

4. การควอนไทซ์แบบเวกเตอร์ (Vector Quantization)

การควอนไทซ์เป็นวิธีการที่นำมาใช้ในการลดขนาดข้อมูล โดยค่าข้อมูลเดิมจะถูกแทนด้วยค่าใหม่ที่ถูกกำหนดไว้ โดยค่าข้อมูลใหม่เป็นค่าที่มีความใกล้เคียงกับข้อมูลเดิมมากที่สุด สำหรับการควอนไทซ์แบ่งได้เป็น 2 แบบ คือ การควอนไทซ์แบบสเกลาร์ และการควอนไทซ์แบบเวกเตอร์ สำหรับการควอนไทซ์แบบสเกลาร์เป็นการควอนไทซ์กับข้อมูลครั้งละค่า ส่วนมากนำไปใช้ในงานด้านการส่งสัญญาณดิจิทัล โดยการควอนไทซ์แบบนี้จะทำหน้าที่แปลงสัญญาณจากสัญญาณอะนาล็อกเป็นสัญญาณดิจิทัล (A/D Converter) ส่วนการควอนไทซ์แบบเวกเตอร์เป็นการควอนไทซ์กับชุดของข้อมูล (เรียกว่า เวกเตอร์ของข้อมูล) การควอนไทซ์แบบนี้ส่วนมากจะนำมาใช้ในการเข้ารหัสข้อมูล สำหรับในด้านการรู้จำผู้พูดได้นำการควอนไทซ์แบบเวกเตอร์มาใช้เพื่อทำหน้าที่แปลงค่าจากข้อมูลของลักษณะสำคัญ จากข้อมูลหลายมิติ (มิติของข้อมูลขึ้นกับอันดับของค่าลักษณะสำคัญ) ให้อยู่ในรูปข้อมูลมิติเดียว เพื่อนำไปเป็นข้อมูลขาเข้าของแบบจำลองฮิดเดนมาร์คอฟ ซึ่งจะได้อธิบายในหัวข้อถัดไป

ในการควอนไทซ์แบบเวกเตอร์ ส่วนสำคัญจะอยู่ที่การกำหนดค่าของข้อมูลอ้างอิง สำหรับงานทางด้านการรู้จำเสียงจะเรียกว่า ชุดรหัส (Codebook) โดยขั้นตอนในการควอนไทซ์แบบเวกเตอร์นั้น จะนำข้อมูลที่ต้องการควอนไทซ์มาเทียบกับข้อมูลในชุดรหัส และแทนค่าข้อมูลเดิมด้วยข้อมูลในชุดรหัสที่ให้ระยะทางของการเทียบต่ำสุด ดังรูปที่ 2-6



รูปที่ 2-6 การควอนไทซ์แบบเวกเตอร์ (Zhang, 1995)

การควอนไทซ์แบบเวกเตอร์โดยทั่วไปแบ่งได้เป็น 2 ช่วงคือช่วงการออกแบบชุดรหัส และช่วงการควอนไทซ์ ดังในรูปที่ 2-6 ช่วงการออกแบบชุดรหัส เป็นขั้นตอนในการสร้างชุดรหัส โดยชุดรหัสจะถูกสร้างจากการฝึกฝนข้อมูลสัญญาณเสียงพูดในชุดฝึกฝน และจะได้ค้นแบบอ้างอิงชุดรหัส (Codebook Reference Template) หลังจากการฝึกฝน ส่วนในช่วงของการควอนไทซ์คือการนำสัญญาณเสียงพูดมาเทียบกับค้นแบบอ้างอิงชุดรหัส หลังจากการเทียบจะได้ค่ารหัสของสัญญาณเสียงพูดนั้นๆ สามารถสรุปได้ว่าการควอนไทซ์แบบเวกเตอร์คือการแบ่งข้อมูลของปริภูมิสัญญาณ (Signal Space) ไปตามจำนวนของพื้นที่หรือเรียกว่าเซลล์ (Cell) และใช้เวกเตอร์จุดศูนย์กลาง (Centroid Vector) ของแต่ละพื้นที่แทนจุดของข้อมูลทั้งหมดที่อยู่ในพื้นที่นั้นๆ สามารถอธิบายตามหลักการและทฤษฎีได้ดังนี้ (Huang, 1989) (Zhang, 1995)

สมมติให้ $x = [x_1, x_2, x_3, \dots, x_d]^T \in R^d$ คือเวกเตอร์ที่มีมิติเท่ากับ d และประกอบไปด้วย $x_k, 1 \leq k \leq d$ และเป็นค่าจำนวนจริงที่ได้จากตัวแปรสุ่มที่มีแอมพลิจูดแบบต่อเนื่อง (Continuous-amplitude random variable) ในการควอนไทซ์แบบเวกเตอร์จะทำการเปลี่ยนเวกเตอร์ x ให้เป็นค่าจำนวนจริงที่ได้จากแอมพลิจูดไม่ต่อเนื่องที่ขนาดเวกเตอร์เท่ากับ d มิติ (y) เราเรียกว่าเวกเตอร์ x ถูกควอนไทซ์ไปเป็นเวกเตอร์ y

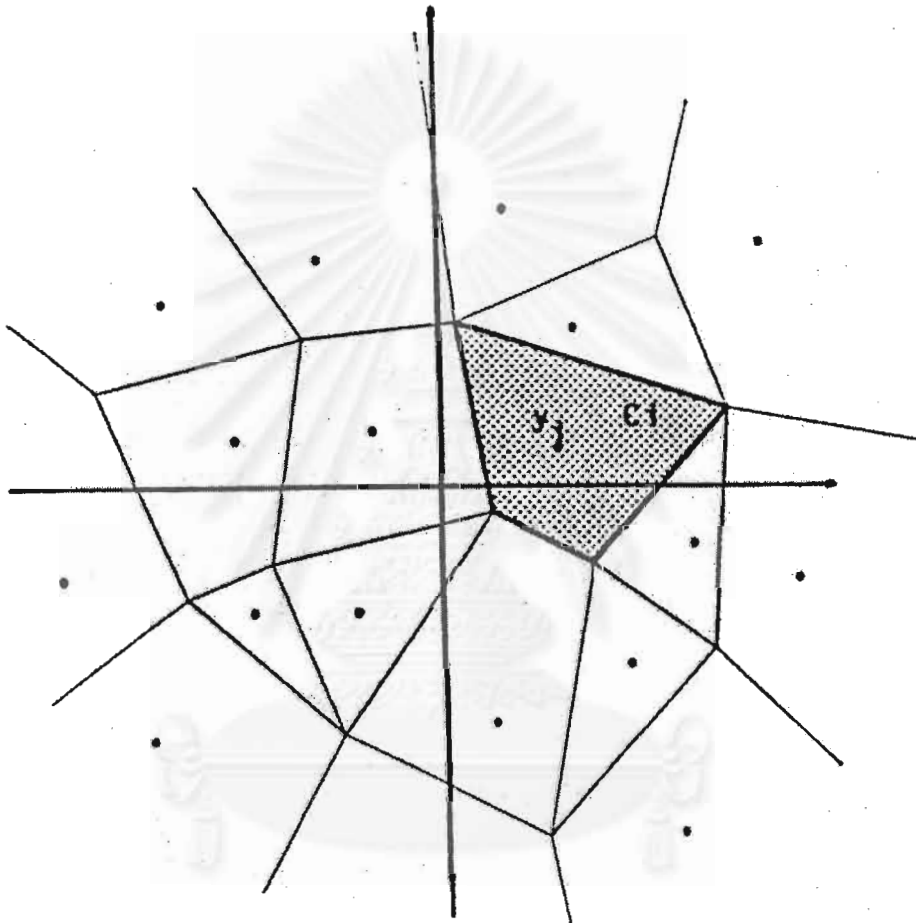
$$y = q(x) \quad (2-26)$$

จากสมการที่ (2-26) $q()$ คือโอเปอเรเตอร์ของการควอนไทซ์ (Quantisation Operator) y คือเวกเตอร์ที่ประกอบไปด้วยเซตจำกัด $Y = \{y_i, 1 \leq i \leq L\}$ เมื่อ $y_i = [y_{i1}, y_{i2}, y_{i3}, \dots, y_{id}]$, เซต Y นี้คือชุดรหัส (Codebook), L คือขนาดของชุดรหัส, และ $\{y_i\}$ คือเซตของเวกเตอร์รหัส (Codeword) ขนาด L ของชุดรหัสเรียกว่า “จำนวนขั้นในชุดรหัส”

การออกแบบชุดรหัสคือการแบ่งปริภูมิที่มีขนาด d มิติของเวกเตอร์ x ที่ถูกสุ่มมาจาก สัญญาณตั้งต้นออกเป็น L เซลล์ $C_i, 1 \leq i \leq L$ และแทนแต่ละเซลล์ C_i ด้วยเวกเตอร์ y_i ตัว ควอนไทซ์ถูกแทนด้วยเวกเตอร์รหัส y_i ถ้า x อยู่ใน C_i

$$q(x) = y_i, \text{ if } x \in C_i \quad (2-27)$$

กระบวนการออกแบบในขั้นตอนนี้เรียกว่า “การฝึกฝนชุดรหัส” ตัวอย่าง การแบ่งข้อมูลใน ปริภูมิที่มีขนาด 2 มิติ ($d = 2$) เพื่อนำไปใช้ในการควอนไทซ์ ดังรูปที่ 2-7



รูปที่ 2-7 การแบ่งปริภูมิที่มีขนาด 2 มิติ (Huang, 1989)

เมื่อ x ถูกควอนไทซ์เป็น y ผลของความผิดพลาดที่เกิดจากการควอนไทซ์และค่าความ เพี้ยนวัดเป็นระยะห่างระหว่าง x กับ y ($d(x, y)$) ซึ่งเป็นการวัดคุณภาพของการควอนไทซ์ การ วัดนี้ต้องบอกถึงความแตกต่างของค่าความเพี้ยนที่ใช้บอกความแตกต่างของสัญญาณเสียงพูดได้ การวัดที่ธรรมดาที่สุดวิธีหนึ่งคือการวัดระยะห่างยูคลีเดียน (Euclidean Distance) ดังสมการที่ (2-28) โดยสมมติให้ความแตกต่างของค่าความเพี้ยนของการควอนไทซ์พารามิเตอร์ต่างๆ มีค่าเท่ากัน

$$d(x, y) = (x - y)^T (x - y) \quad (2-28)$$

แต่ถ้ามีการถ่วงน้ำหนักไม่เท่ากันจะใช้การวัดระยะห่างของมาฮาลานอบิส (Mahalanobis Distance) โดยมีการนำความแปรปรวนผกผันของ y มาคิดด้วย ดังสมการที่ (2-29)

$$d(x, y) = (x - y)^T \Sigma^{-1} (x - y) \quad (2-29)$$

การออกแบบชุดรหัส L ระดับคือการแบ่งข้อมูลในปริภูมิ (Space) d มิติออกเป็น L เซลล์และแทนแต่ละเซลล์ด้วยเวกเตอร์ที่ถูกควอนไทซ์แล้ว การหาเวกเตอร์ของตัวควอนไทซ์ที่เหมาะสมสามารถหาได้จากค่าความเพี้ยนรวมทั้งหมดของแต่ละระดับของชุดรหัสที่มีค่าต่ำที่สุด ค่าความเพี้ยนรวมทั้งหมดสามารถหาได้จาก

$$D = E[d(x, y_i)] \quad (2-30)$$

$$= \sum_{i=1}^L \Pr(x \in C_i) E[d(x, y_i) / x \in C_i] \quad (2-31)$$

$$= \sum_{i=1}^L \Pr(x \in C_i) \int_{x \in C_i} d(x, y_i) f(x) dx \quad (2-32)$$

เมื่อ $\Pr(x \in C_i)$ คือค่าความน่าจะเป็นแบบดิสครีต (Discrete Probability) ของ x ที่อยู่ในเซลล์ C_i , $f(x)$ คือฟังก์ชันความหนาแน่นความน่าจะเป็น (Probability Density Function) ของ x และ $E[]$ คือค่าคาดหวัง

การออกแบบชุดรหัส (Codebook Design)

การออกแบบชุดรหัสเป็นการจัดกลุ่มของข้อมูลแบบหนึ่ง เพื่อนำไปสร้างเป็นต้นแบบอ้างอิงชุดรหัส (Codebook Reference Template) ใช้ในการทดสอบความคล้ายคลึงกันของรูปแบบของชุดข้อมูลที่เข้ามาทดสอบกับต้นแบบอ้างอิงชุดรหัสก่อนทำการควอนไทซ์แบบเวกเตอร์และเข้าสู่แบบจำลองทางสถิติต่อไป การจัดกลุ่มคือการควอนไทซ์ที่เหมาะสมเมื่อสมการที่ (2-32) มีค่าน้อยที่สุดจากทุกๆ L ระดับของตัวควอนไทซ์ ซึ่งหมายความว่าต้องเป็นไปตามเงื่อนไขของความเหมาะสมที่จำเป็นทั้ง 2 ประการ ประการแรกคือตัวควอนไทซ์ที่เหมาะสมที่สุดต้องเป็นไปตามกฎความเพี้ยนน้อยที่สุด (Minimum Distortion) หรือกฎการเลือกบริเวณที่ใกล้เคียงที่สุด (Nearest Neighbour Rule) ดังนี้

$$q(x) = y_i \text{ if and only if } d(x, y_i) \leq d(x, y_j), \quad j \neq i, \quad 1 \leq j \leq L \quad (2-33)$$

จากสมการที่ (2-33) หมายความว่าตัวควอนไทซ์ต้องเลือกเวกเตอร์รหัส (Codeword) ที่มีความเพี้ยนน้อยที่สุดเมื่อเทียบกับ x

เงื่อนไขที่สองคือ การเลือกเวกเตอร์รหัส y_i ต้องให้ความเพี้ยนเฉลี่ยในเซลล์ C_i มีค่าน้อยที่สุด นั่นคือ y_i เป็นเวกเตอร์ y ที่ทำให้สมการที่ (2-34) มีค่าน้อยที่สุด

$$D_i = E[d(x, y) / x \in C_i] \quad (2-34)$$

$$= \int_{x \in C_i} d(x, y) f(x) dx \quad (2-35)$$

เวกเตอร์ดังกล่าวนี้เรียกว่า “จุดศูนย์กลาง” (Centroid) ของเซลล์ C_i ซึ่งเขียนได้เป็น

$$y_i = \text{cent}(C_i) \quad (2-36)$$

การคำนวณหาค่าจุดศูนย์กลางของแต่ละบริเวณ นั้นขึ้นอยู่กับนิยามของการวัดค่าความเพี้ยน ในทางปฏิบัติต้องกำหนดให้ชุดของเวกเตอร์ฝึกฝน $\{x_k, 1 \leq k \leq T\}$ โดยที่ชุดย่อย K_i ของเวกเตอร์ นี้อยู่ในเซต C_i เพราะฉะนั้นค่าความเพี้ยนเฉลี่ย (D_i) ในเซต C_i เป็นไปตามสมการที่ (2-37)

$$D_i = \frac{1}{K_i} \sum_{x \in C_i} d(x, y_i) \quad (2-37)$$

ค่าความเพี้ยนเฉลี่ยโดยรวมสามารถคำนวณได้จาก

$$D = \sum_{i=1}^L D_i / T \quad (2-38)$$

ให้ค่าความเพี้ยนที่วัดได้ในกลุ่มคือ C_i และ $d(x, y)$ คือค่า D_i ที่น้อยที่สุด จากสมการที่ (2-37) เขียนใหม่ได้เป็น

$$y_i = \frac{1}{K_i} \sum_{x \in C_i} x \quad (2-39)$$

เมื่อ y_i คือค่าเฉลี่ยของตัวอย่างที่ได้จากชุดเวกเตอร์ฝึกฝนทั้งหมด (x) ที่อยู่ในกลุ่ม C_i

ขั้นตอนวิธีการแบ่งกลุ่มที่ใช้การวนซ้ำคือขั้นตอนวิธีการหนึ่งที่ใช้ในการออกแบบชุดรหัส เพื่อให้ได้ความเพี้ยนเฉลี่ยน้อยที่สุด (Iteratively Minimise the Average Distortion) เช่น ขั้นตอนวิธีการแบ่งเฉลี่ย K ส่วน (K-means Algorithm) ขั้นตอนวิธีการ Linde, Buzo และ Gray (LBG Algorithm) และขั้นตอนวิธีการหาค่าคาดหวังที่มากที่สุด (Expectation Maximization Algorithm) เป็นต้น (Huang, 1989) ขั้นตอนในการออกแบบชุดรหัสสามารถแบ่งได้เป็นสองขั้นตอนคือ การสร้างชุดรหัสและการฝึกฝนชุดรหัส เนื่องจากงานวิจัยนี้ได้ใช้ขั้นตอนวิธีการแบ่งเฉลี่ย K ส่วน เพราะเป็นวิธีการที่ง่าย รวดเร็ว และให้ผลที่น่าพอใจ (วิสูตร อาชุนทร, 2538) จึงจะกล่าวถึงเฉพาะ รายละเอียดและทฤษฎีของขั้นตอนวิธีการนี้เท่านั้น

ขั้นตอนวิธีการแบ่งเฉลี่ย K ส่วน (K-Mean Algorithm)

ขั้นตอนวิธีการแบ่งเฉลี่ย K ส่วนเป็นขั้นตอนวิธีการที่ได้รับความนิยมมากที่สุดในการแบ่งกลุ่มแบบวนซ้ำ (Iterative Clustering Algorithm) (Huang, 1989) ที่ใช้พื้นฐานของการแบ่งชุดข้อมูลฝึกฝนเป็น L กลุ่ม $\{C_i, 1 \leq i \leq L\}$ ที่ต้องใช้เงื่อนไขที่จำเป็นดังที่กล่าวไว้ข้างต้นในการหาค่าที่เหมาะสม เพื่อให้ได้ค่าเพี้ยนเฉลี่ย (Average Distortion) น้อยที่สุด โดยใช้พื้นฐานการแบ่งชุดของเวกเตอร์ฝึกฝน $\{x(n)\}$ ออกเป็น L กลุ่ม เมื่อกำหนดให้ $L = K$ มีขั้นตอนดังต่อไปนี้

ขั้นตอนการสร้างชุดรหัสเริ่มต้น (Initial Codebook)

ขั้นตอนในการสร้างชุดรหัสมีหลายวิธี (Makhoul, Roucos, and Gish, 1985) ได้แก่ การค้นหาแบบทวิภาค (Binary Search) การควอนไทซ์แบบต่อเนื่องกัน (Cascade Quantization) รหัสผล

คูณ (Product Codes) และชุดรหัสแบบสุ่ม (Random Codebooks) สำหรับงานวิจัยนี้ใช้วิธีการชุดรหัสแบบสุ่มในการสร้างชุดรหัสเริ่มต้นเพื่อนำไปใช้ในการฝึกฝนชุดรหัสต่อไป เนื่องจากการสร้างชุดรหัสแบบสุ่มมีความรวดเร็วและทำได้ง่ายกว่าวิธีอื่น วิธีการสร้างชุดรหัสแบบนี้คือการสุ่มเลือกเวกเตอร์รหัสเริ่มต้นจากชุดข้อมูลฝึกฝนตามขนาดของชุดรหัส (L)

ขั้นตอนการฝึกฝนชุดรหัส (Codebook Training)

เป็นขั้นตอนในการหาค่าที่เหมาะสมที่สุดจากชุดเวกเตอร์ฝึกฝนมี 4 ขั้นตอนดังนี้

ขั้นตอนที่ 1 การสร้างชุดรหัสเริ่มต้น (Initialisation): สร้างชุดรหัสแบบสุ่มจากชุดข้อมูลฝึกฝนที่มีชุดเวกเตอร์ของตัวควอนไทซ์เท่ากับ $y_i, 1 \leq i \leq L$

ขั้นตอนที่ 2 การจำแนก (Classification): ทำการจำแนกข้อมูลในชุดเวกเตอร์ฝึกฝน (Training Vectors) x_k รั้งละตัวว่าอยู่ในกลุ่มของ C_i กลุ่มใด โดยใช้กฎการเลือกเพื่อนบ้านข้างเคียงใกล้ที่สุด (Nearest Neighbor Rule) ดังนี้

$$x \in C_i, \text{ iff } d(x, y_i) \leq d(x, y_j) \text{ for all } j \neq i \quad (2-40)$$

ขั้นตอนที่ 3 การปรับชุดรหัส (Codebook Updating): เป็นการปรับค่าเวกเตอร์รหัสในทุกๆ กลุ่มโดยใช้การคำนวณจุดศูนย์กลางของชุดเวกเตอร์ฝึกฝนในแต่ละกลุ่มใหม่ดังนี้

$$y_i = \text{cent}(C_i), 1 \leq i \leq L \quad (2-41)$$

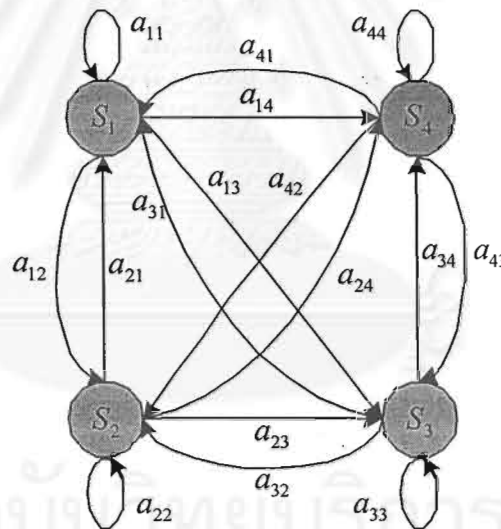
ขั้นตอนที่ 4 การสิ้นสุด (Termination): ถ้าค่าความเพี้ยนโดยรวม D ในรอบปัจจุบันเทียบกับค่าความเพี้ยนโดยรวมในรอบก่อนหน้านี้นั้นมีค่าต่ำกว่าจุดเริ่มเปลี่ยน (Threshold) ก็จะสิ้นสุดขั้นตอนวิธีการ แต่ถ้าไม่ก็จะวนซ้ำในขั้นตอนที่ 2 ค่อยไป

ในกระบวนการหาค่าความเพี้ยนเฉลี่ยค่าที่ต่ำที่สุดของขั้นตอนวิธีการแบ่งเฉลี่ย K ส่วนจะสิ้นสุดใน 2 กรณี คือเมื่อได้ค่าศูนย์กลาง y_i หรือค่าเฉลี่ยของแต่ละกลุ่มข้อมูล C_i แล้วใช้กระบวนการหาค่าต่ำที่สุด (Minimization Process) เพื่อแบ่งกลุ่มของเวกเตอร์ทั้งหมดที่สอดคล้องกับค่าศูนย์กลางที่อยู่ใกล้ที่สุดโดยใช้การวัดค่าความเพี้ยนเฉลี่ย D_i เป็นตัวตัดสิน หรือในทางตรงกันข้ามถ้าแบ่งกลุ่มของข้อมูลได้ทั้งหมดแล้วใช้กระบวนการหาค่าต่ำที่สุด เพื่อหาจุดศูนย์กลางใหม่ในแต่ละกลุ่มข้อมูลที่ทำให้ค่าความเพี้ยนเฉลี่ยต่ำสุด D_i จากการวนซ้ำของทั้งสองกรณีทำให้ค่าความเพี้ยนเฉลี่ย D_i ใหม่ที่มีค่าน้อยกว่าค่าในรอบก่อนหน้านี้ อย่างไรก็ตามขั้นตอนวิธีการแบ่งเฉลี่ย K ส่วนลู่เข้าสู่ค่าเหมาะสมที่สุดเฉพาะแห่ง (Local Optimum) ทำให้ผลลัพธ์ที่ได้ไม่เป็นหนึ่งเดียว ดังนั้นการหาค่าเหมาะสมที่สุดที่ครอบคลุมทั้งหมด (Global Optimum) กระทำได้โดยการใช้ค่าเริ่มต้นของเวกเตอร์รหัสหลายค่าที่แตกต่างกันแล้วทำตามขั้นตอนวิธีการแบ่งเฉลี่ย K ส่วน และซ้ำขั้นตอนนี้กับค่าเริ่มต้นที่แตกต่างกันหลายๆ ครั้ง เพื่อให้ชุดรหัสที่มีค่าความเพี้ยนโดยรวมทั้งหมดน้อยที่สุด

5. แบบจำลองฮิดเดนมาร์คอฟ

แบบจำลองฮิดเดนมาร์คอฟ เป็นแบบจำลองทางสถิติ (Statistical Model) ที่นำมาใช้ในการจำแนกแยกแยะสัญญาณรูปแบบต่างๆ โดยใช้กระบวนการทางด้านสถิติมาสร้างแบบจำลองของสัญญาณ แบบจำลองนี้ได้รับการพัฒนามาเป็นเวลานานแล้ว เริ่มแรกที่มีการเผยแพร่ผลงานเมื่อประมาณปี 1960-1970 โดย Baum และผู้ร่วมงาน จากนั้นก็ได้มีการนำความรู้พื้นฐานของแบบจำลองฮิดเดนมาร์คอฟไปประยุกต์ใช้กับงานด้านต่างๆ โดยเฉพาะงานด้านการรู้จำเสียงพูด ได้มีการพัฒนานำแบบจำลองฮิดเดนมาร์คอฟมาประยุกต์ใช้ในการรู้จำเสียงพูดด้วย เนื่องจากผลของการนำแบบจำลองนี้มาใช้ให้อัตราการรู้จำที่ดีเมื่อเทียบกับการใช้ระบบรู้จำแบบอื่น ซึ่งมีงานวิจัยจำนวนมากที่นิยมใช้แบบจำลองฮิดเดนมาร์คอฟในการรู้จำเสียงพูด (Rabiner, 1989) (Huang, 1989) (Zhang, 1995) (เสาวลักษณ์ อารีพงศา, 2538) (วิศรุต อาขุบุตร, 2538)

แบบจำลองฮิดเดนมาร์คอฟได้ใช้พื้นฐานของลูกโซ่มาร์คอฟ (Markov chain) โดยลูกโซ่มาร์คอฟจะจำลองเหตุการณ์ด้วยสถานะ (State) และแต่ละสถานะจะมีความสัมพันธ์กันด้วยค่าของความน่าจะเป็นในการเปลี่ยนสถานะดังแสดงในรูปที่ 2-8



รูปที่ 2-8 ลูกโซ่มาร์คอฟ 4 สถานะ

ลักษณะเด่นของลูกโซ่มาร์คอฟคือ ความน่าจะเป็นของการที่ค่าสัญญาณ ณ เวลาหนึ่งที่สถานะใดสถานะหนึ่ง ขึ้นอยู่กับค่าของสัญญาณที่สถานะก่อนหน้านั้นเพียงสถานะเดียว (Rabiner, 1989) ดังแสดงในสมการที่ (2-42)

$$P[q_t = S_j | q_{t-1} = S_i, q_{t-2} = S_k, \dots] = P[q_t = S_j | q_{t-1} = S_i] \quad (2-42)$$

โดยจะแทนค่าความน่าจะเป็นของการเปลี่ยนแปลงสถานะจากสถานะ i ไปสถานะ j ด้วย a_{ij} ดังสมการที่ 2-43

$$a_{ij} = P[q_t = S_j | q_{t-1} = S_i] \quad (2-43)$$

สำหรับแบบจำลองฮิดเดนมาร์คอฟ ที่ใช้คำว่า “ฮิดเดน” ก็เพราะว่าเป็นแบบจำลองที่จำลองเหตุการณ์ที่ไม่สามารถทราบถึงกระบวนการที่เกิดขึ้นภายในได้ ว่าค่าสัญญาณที่ผ่านลูกโซ่มาร์คอฟมีการเปลี่ยนแปลงในสถานะต่างๆอย่างไร แต่ที่สามารถทราบได้ก็คือค่าความน่าจะเป็นของสัญญาณเมื่อผ่านลูกโซ่มาร์คอฟเท่านั้น จึงเรียกกระบวนการมาร์คอฟที่เกิดขึ้นในลักษณะนี้ว่า ฮิดเดนมาร์คอฟ

ประเภทของแบบจำลองฮิดเดนมาร์คอฟ

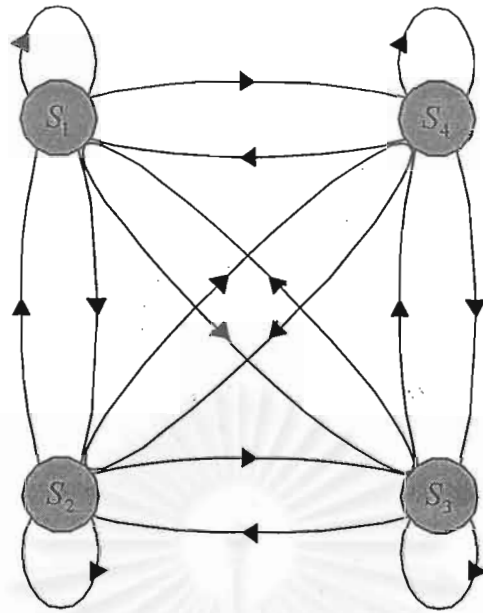
แบบจำลองฮิดเดนมาร์คอฟส่วนใหญ่เป็นการพิจารณาเพียงกรณีพิเศษของแบบจำลองฮิดเดนมาร์คอฟประเภทเออร์กอดิกมาตรฐาน ซึ่งเป็นแบบจำลองฮิดเดนมาร์คอฟที่ทุกสถานะต่อเชื่อมถึงกันหมด โดยทุกสถานะของแบบจำลองสามารถเข้าถึงสถานะอื่นๆ ได้ในเส้นทางเดียวดังแสดงในรูปที่ 2-9(ก) เมื่อจำนวนสถานะของแบบจำลอง $N = 4$ สถานะและมีคุณสมบัติเฉพาะของสัมประสิทธิ์ a_{ij} มีค่าเป็นบวก จากรูปที่ 2-9(ก) จะได้สมการของค่าสัมประสิทธิ์ a_{ij} ได้ดังนี้

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \quad (2-44)$$

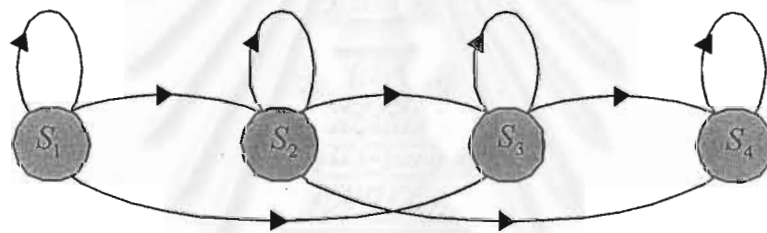
ในการประยุกต์ใช้งานแบบจำลองฮิดเดนมาร์คอฟกับงานเฉพาะอย่างนั้น ได้มีการเสนอแบบจำลองฮิดเดนมาร์คอฟประเภทอื่นที่เหมาะสมกับคุณสมบัติของสัญญาณเสียงพูดได้คือแบบจำลองประเภทเออร์กอดิกมาตรฐาน คือแบบจำลองซ้ายไปขวา (Left-Right Model) หรือแบบจำลอง Bakis ดังแสดงในรูปที่ 2-9(ข) เพราะว่าลำดับสถานะของแบบจำลองมีคุณสมบัติตามเวลา คือเมื่อเวลาเพิ่มขึ้น คำนีของสถานะจะเพิ่มขึ้นหรืออยู่ที่เดิมเปรียบเสมือนดำเนินจากซ้ายไปขวา ดังนั้นแบบจำลองซ้ายไปขวาก็เหมาะสมในการจำลองสัญญาณที่มีคุณสมบัติเปลี่ยนแปลงตามเวลา เช่น สัญญาณเสียงพูด เป็นต้น

คุณสมบัติพื้นฐานของแบบจำลองฮิดเดนมาร์คอฟประเภทซ้ายไปขวานั้น คือค่าสัมประสิทธิ์ของการเปลี่ยนสถานะต้องมีคุณสมบัติดังนี้

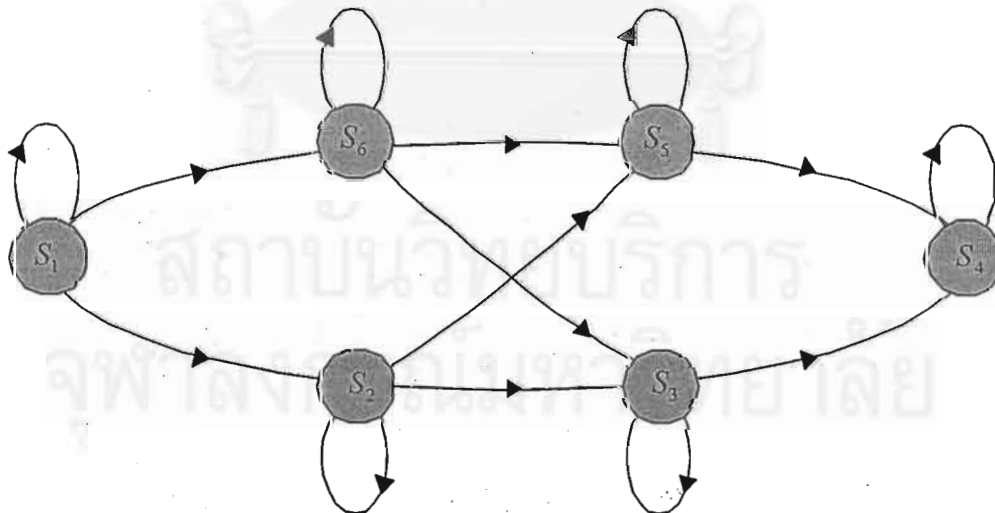
$$a_{ij} = 0, \quad j < i \quad (2-45)$$



(ก)



(ข)



(ค)

รูปที่ 2-9 แบบจำลองฮิดเดนมาร์คอฟ 3 ประเภท (ก) แบบจำลองเออร์กอดิก 4 สถานะ (ข) แบบจำลองซ่ายไปขวา 4 สถานะ (ค) แบบจำลองเส้นทางขนานซ่ายไปขวา 6 สถานะ (Rabiner, 1989)

จากสมการที่ (2-45) หมายความว่าไม่มีการเปลี่ยนสถานะไปยังสถานะที่มีค่านี้นต่ำกว่าค่านีของสถานะปัจจุบัน และมีความน่าจะเป็นเริ่มต้นมีคุณสมบัติตามสมการที่ (2-46)

$$\pi_i = \begin{cases} 0, & i \neq 1 \\ 1, & i = 1 \end{cases} \quad (2-46)$$

เนื่องจากลำดับสถานะต้องเริ่มต้นจากสถานะที่ 1 และสิ้นสุดในสถานะที่ N ดังนั้นแบบจำลองข่ายไปขวาต้องเพิ่มเงื่อนไขของสัมประสิทธิ์การเปลี่ยนสถานะ เพื่อไม่ให้เกิดการเปลี่ยนแปลงของค่านีมากเกินไปดังนี้

$$a_{ij} = 0, \quad j > i + \Delta \quad (2-47)$$

จากรูปที่ 2-9(ข) จะกำหนดให้ $\Delta = 2$ และไม่มีการข้ามสถานะเกินกว่า 2 สถานะ เขียนการเปลี่ยนสถานะให้อยู่ในรูปเมตริกซ์ได้ดังนี้

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & 0 \\ 0 & a_{22} & a_{23} & a_{24} \\ 0 & 0 & a_{33} & a_{34} \\ 0 & 0 & 0 & a_{44} \end{bmatrix} \quad (2-48)$$

และที่สถานะสุดท้ายของแบบจำลองประเภทข่ายไปขวาจะมีสัมประสิทธิ์ของการเปลี่ยนสถานะดังสมการต่อไปนี้

$$a_{NN} = 1 \quad (2-49)$$

$$a_{Ni} = 0, \quad i < N \quad (2-50)$$

สำหรับรูปแบบที่จะนำมาใช้ในงานวิทยานิพนธ์นี้เป็นแบบจำลองแบบข่าย-ขวา เนื่องจากเป็นแบบจำลองที่นิยมนำมาใช้ในงานรู้จำเสียงพูด

ในการสร้างแบบจำลองฮิดเดนมาร์คอฟจะถูกกำหนดโดยค่าพารามิเตอร์ของแบบจำลองซึ่งประกอบด้วย

1. The State Transition probability distribution, $A_{N \times N} = \{a_{ij}\}$ โดยที่

$$a_{ij} = P[q_{t+1} = S_j | q_t = S_i] \quad 1 \leq i, j \leq N \quad (2-51)$$

q_t : State ที่เวลา t

S_i : State ที่ i

N : จำนวน state ของแบบจำลองฮิดเดนมาร์คอฟ

2. The observation symbol probability distribution in state j , $B = \{b_j(k)\}$

โดยที่

$$b_j(k) = P[v_k \text{ at } | q_t = S_j], \quad \begin{matrix} 1 \leq j \leq N \\ 1 \leq k \leq M \end{matrix} \quad (2-52)$$

v_k : Symbol ที่ k

M : จำนวน Observation symbol ใน 1 state

3. The initial state distribution, $\pi = \{\pi_i\}$ โดยที่

$$\pi_i = P[q_1 = S_i], \quad 1 \leq i \leq N \quad (2-53)$$

โดยเราจะแทนแบบจำลองฮิดเดนมาร์คคอฟที่เกิดจาก A, B, π ด้วย

$$\lambda = (A, B, \pi) \quad (2-54)$$

ก่อนที่จะไปสู่ขั้นตอนในการนำแบบจำลองฮิดเดนมาร์คคอฟไปใช้งานนั้น จำเป็นต้องศึกษาปัญหาพื้นฐาน 3 ข้อของแบบจำลองฮิดเดนมาร์คคอฟซึ่งจะนำไปสู่วิธีการที่จะได้มาซึ่งแบบจำลองและวิธีการใช้แบบจำลองเพื่อหาค่าความน่าจะเป็นของสัญญาณ

ปัญหาพื้นฐานสามประการของแบบจำลองฮิดเดนมาร์คคอฟ (Rabiner, 1989)

ปัญหาพื้นฐานที่จะกล่าวถึงเป็นปัญหาที่ถูกกำหนดขึ้นเพื่อนำไปสู่การแก้ไขปัญหาดังกล่าว โดยรายละเอียดของปัญหาแต่ละข้อมีดังนี้

ปัญหาข้อที่ 1 : เมื่อกำหนดลำดับค่าสังเกต $O = O_1, O_2, O_3, \dots, O_T$ และแบบจำลองฮิดเดนมาร์คคอฟ $\lambda = (A, B, \pi)$ มาให้ จะมีวิธีการในคำนวณค่าความน่าจะเป็น $P(O | \lambda)$ ได้อย่างไร

ปัญหาข้อที่ 2 : เมื่อกำหนดลำดับค่าสังเกต $O = O_1, O_2, O_3, \dots, O_T$ และแบบจำลองฮิดเดนมาร์คคอฟ $\lambda = (A, B, \pi)$ มาให้ จะสามารถหาลำดับของการเปลี่ยนแปลงสถานะในแบบจำลอง $Q = q_1, q_2, q_3, \dots, q_T$ ได้อย่างไร

ปัญหาข้อที่ 3 : จะมีวิธีการในการปรับค่า $\lambda = (A, B, \pi)$ ของแบบจำลองอย่างไร เพื่อทำให้ค่าความน่าจะเป็นของลำดับค่าสังเกต $P(O | \lambda)$ มีค่ามากที่สุด

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

การแก้ไขปัญหาค่าพื้นฐานสามประการของแบบจำลองฮิดเดนมาร์คอฟ

ในการแก้ไขปัญหาค่าพื้นฐานสามประการของแบบจำลองฮิดเดนมาร์คอฟ ตามปัญหาที่ได้กำหนดไว้เราสามารถอธิบายวิธีการแก้ไขในแต่ละปัญหาดังต่อไปนี้

การแก้ไขปัญหาค่าข้อที่ 1

เมื่อกำหนดแบบจำลอง $\lambda = (A, B, \pi)$ การหาค่าความน่าจะเป็น $P(O | \lambda)$ ของลำดับค่าสังเกต $O = O_1, O_2, O_3, \dots, O_T$ จะอาศัยวิธีการหาค่าทุกลำดับสถานะความยาว T ที่เป็นไปได้ทั้งหมดตามลำดับ เมื่อ T เป็นจำนวนค่าสังเกต พิจารณาลำดับของสถานะที่มีค่าจำกัดดังนี้

$$Q = q_1, q_2, q_3, \dots, q_T \quad (2-55)$$

เมื่อ q_1 เป็นสถานะเริ่มต้น ค่าความน่าจะเป็นของลำดับค่าสังเกต O สำหรับลำดับสถานะสามารถหาได้จากสมการ

$$P(O | Q, \lambda) = \prod_{t=1}^T P(O_t | q_t, \lambda) \quad (2-56)$$

เมื่อกำหนดให้ค่าสังเกตเป็นชนิดที่ไม่ขึ้นแก่กันในทางสถิติจะได้ว่า

$$P(O | Q, \lambda) = b_{q_1}(O_1) \cdot b_{q_2}(O_2) \cdot \dots \cdot b_{q_T}(O_T) \quad (2-57)$$

ดังนั้นค่าความน่าจะเป็นของลำดับสถานะ Q สามารถเขียนได้เป็น

$$P(Q | \lambda) = \pi_{q_1} a_{q_1 q_2} a_{q_2 q_3} \dots a_{q_{T-1} q_T} \quad (2-58)$$

ค่าความน่าจะเป็นร่วมระหว่าง O และ Q เป็นเพียงผลคูณของสมการข้างต้นดังนี้

$$P(O, Q | \lambda) = P(O | Q, \lambda) P(Q, \lambda) \quad (2-59)$$

ค่าความน่าจะเป็นของ O เมื่อกำหนดแบบจำลอง $\lambda = (A, B, \pi)$ มาให้ สามารถหาได้โดยผลรวมของค่าความน่าจะเป็นร่วมของลำดับสถานะทั้งหมดที่เป็นไปได้ ดังนี้

$$\begin{aligned} P(O | \lambda) &= \sum_{\text{all } Q} P(O | Q, \lambda) P(Q, \lambda) \\ &= \sum_{q_1, q_2, \dots, q_T} \pi_{q_1} b_{q_1}(O_1) a_{q_1 q_2} b_{q_2}(O_2) \dots a_{q_{T-1} q_T} b_{q_T}(O_T) \end{aligned} \quad (2-60)$$

จากสมการที่ (2-60) เริ่มต้นที่เวลา $t=1$ จะเริ่มที่สถานะ q_1 ด้วยค่าความน่าจะเป็น π_{q_1} และให้กำเนิดสัญลักษณ์ O_1 ภายในสถานะเดียวกันด้วยค่าความน่าจะเป็น $b_{q_1}(O_1)$ เมื่อเวลาเปลี่ยนจาก t เป็น $t+1$ และเปลี่ยนไปยังสถานะ q_2 จากสถานะ q_1 ด้วยค่าความน่าจะเป็น $a_{q_1 q_2}$ และให้กำเนิดสัญลักษณ์ O_2 ภายในสถานะเดียวกันด้วยค่าความน่าจะเป็น $b_{q_2}(O_2)$ กระบวนการนี้จะดำเนินต่อไปอย่างต่อเนื่องจนกระทั่งถึงการเปลี่ยนแปลงที่เวลา T จากสถานะ q_{T-1} ไปยังสถานะ q_T ด้วยค่าความน่าจะเป็น $a_{q_{T-1} q_T}$ และให้กำเนิดสัญลักษณ์ O_T ภายในสถานะเดียวกันด้วยค่าความน่าจะเป็น $b_{q_T}(O_T)$

ในการคำนวณค่าความน่าจะเป็น $P(O | \lambda)$ ตามสมการที่ (2-60) นั้นจะทำให้เกิดการคำนวณขึ้นด้วยอันดับประมาณ $2T \cdot N^T$ ครั้ง เนื่องจากทุกเวลา $t = 1, 2, \dots, T$ จะเกิดสถานะที่เป็นไปได้

ได้ N สถานะซึ่งเข้าถึงได้ และในแต่ละลำดับสถานะจะเกิดการขึ้นประมาณ $2T$ ครั้งสำหรับแต่ละพจน์ในผลรวมของสมการที่ (2-60) ซึ่งจำแนกได้เป็นการคูณ $(2T-1) \cdot N^T$ ครั้งและการบวก $N^T - 1$ ครั้ง จึงทำให้เกิดการคำนวณค่าโดยใช้สมการนี้เป็นไปไม่ได้ถึงแม้ด้วยค่า N และ T ค่าน้อยๆ ก็ตาม ดังนั้นจึงมีกระบวนการที่มีประสิทธิภาพในการคำนวณค่า ที่เรียกว่า “ กระบวนการไปหน้า-ย้อนกลับ ” (Forward-Backward Procedure) ซึ่งประกอบด้วยกระบวนการไปหน้าและกระบวนการย้อนกลับร่วมกันดังต่อไปนี้

กระบวนการไปข้างหน้า

พิจารณาตัวแปรไปหน้า $\alpha_t(i)$ ซึ่งนิยามดังสมการ

$$\alpha_t(i) = P(O_1 O_2 \cdots O_t, q_t = S_i | \lambda) \quad (2-61)$$

ค่าความน่าจะเป็นของลำดับค่าสังเกตบางส่วน (The probability of the partial observation sequence) $O_1 O_2 \cdots O_t$ และสถานะ S_i ที่เวลา t เมื่อกำหนดแบบจำลอง $\lambda = (A, B, \pi)$ จะสามารถหาค่าของ $\alpha_t(i)$ ได้ดังขั้นตอนดังต่อไปนี้

ขั้นตอนที่ 1 กระบวนการเริ่มต้น

$$\alpha_1(i) = \pi_i b_i(O_1), \quad 1 \leq i \leq N \quad (2-62)$$

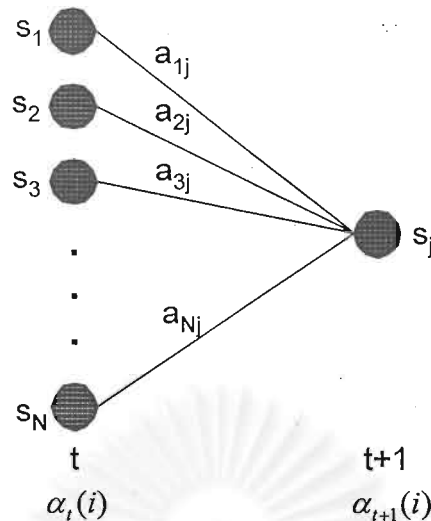
ขั้นตอนที่ 2 กระบวนการอุปนัย

$$\alpha_t(j) = \left[\sum_{i=1}^N \alpha_{t-1}(i) a_{ij} \right] b_j(O_t), \quad 2 \leq t \leq T, \quad 1 \leq i \leq N \quad (2-63)$$

ขั้นตอนที่ 3 กระบวนการสิ้นสุด

$$P(O | \lambda) = \sum_{i=1}^N \alpha_T(i) \quad (2-64)$$

จากกระบวนการข้างต้นนี้ ขั้นตอนแรกเป็นการกำหนดความน่าจะเป็นเริ่มต้นให้กับทุกๆ สถานะของแบบจำลอง และสมการที่ (2-63) แสดงถึงสถานะ j ที่สามารถเข้าถึงได้ที่เวลา t จากสถานะที่เป็นไปได้ i สถานะที่เวลา $t-1$ ดังแสดงในรูปที่ 2-10 เมื่อ $\alpha_{t-1}(i)$ คือความน่าจะเป็นร่วมของค่าสังเกต O_1, O_2, \dots, O_{t-1} ที่สถานะ t ดังนั้นผลคูณของ $\alpha_{t-1}(i) a_{ij}$ คือความน่าจะเป็นร่วมของค่าสังเกต O_1, O_2, \dots, O_{t-1} ที่เข้าถึงสถานะ j ที่เวลา t และผ่านสถานะ i ที่เวลา $t-1$ ผลรวมของผลคูณของสถานะที่เป็นไปได้ทั้งหมด (i) ที่เวลา $t-1$ ได้ผลลัพธ์เป็นค่าความน่าจะเป็นในการเข้าถึงสถานะ j ที่เวลา t โดยผ่านค่าสังเกตบางส่วนก่อนหน้านั้นทั้งหมด และคูณด้วยความน่าจะเป็นของค่าสังเกต ($b_j(O_t)$) ที่เข้าถึงสถานะ j ที่กำเนิด O_t ผลของ $\alpha_t(j)$ คือความน่าจะเป็นของลำดับค่าสังเกตใหม่ O_1, O_2, \dots, O_{t-1} ที่เวลา t และสถานะ j



รูปที่ 2-10 รายละเอียดของตัวแปรในการคำนวณค่าตัวแปรไปหน้า $\alpha_t(i)$ (Rabiner, 1989)

ในขั้นตอนที่ 3 เป็นการคำนวณหา $P(O|\lambda)$ ซึ่งคือผลบวกตัวแปรไปหน้าสุดท้ายที่สถานะสุดท้าย เพราะว่า $\alpha_T(i) = P(O_1 O_2 \dots O_T, s_T = i | \lambda)$ และการเปลี่ยนสถานะจะสิ้นสุดที่ S_F การคำนวณหา $\alpha_t(i)$ ด้วยขั้นวิธีการไปหน้าใช้การคำนวณเพียง $O(N^2T)$ ครั้งเท่านั้น

กระบวนการย้อนกลับ

พิจารณาตัวแปรย้อนกลับ $\beta_t(i)$ ซึ่งนิยามตามสมการ

$$\beta_t(i) = P(O_{t+1}, O_{t+2}, \dots, O_T, s_t = i | \lambda) \quad (2-65)$$

ความน่าจะเป็นของลำดับค่าสังเกตบางส่วนจาก $t+1$ ไปสิ้นสุดการสังเกตที่ T เมื่อกำหนดให้สถานะ i ที่เวลา t และพารามิเตอร์ของแบบจำลอง λ จะสามารถหา $\beta_t(i)$ โดยอุปนัยคล้ายๆ กับการหาตัวแปรไปหน้า α ได้ดังต่อไปนี้

ขั้นตอนที่ 1 กระบวนการเริ่มต้น

$$\beta_T(i) = 1, \quad 1 \leq i \leq N \quad (2-66)$$

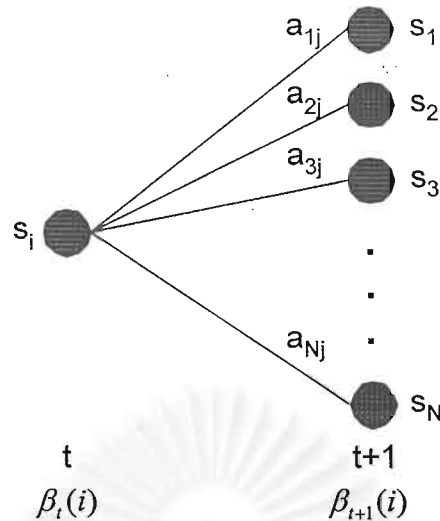
ขั้นตอนที่ 2 กระบวนการอุปนัย

$$\beta_t(j) = \sum_{i=1}^N a_{ij} b_j(O_{t+1}) \beta_{t+1}(i), \quad t = T-1, T-2, \dots, 1, \quad 1 \leq i \leq N \quad (2-67)$$

ขั้นตอนที่ 3 กระบวนการสิ้นสุด

$$P(O|\lambda) = \sum_{i=1}^N \pi_i b_i(O_1) \beta_1(i) \quad (2-68)$$

ในขั้นตอนที่ 1 เป็นการกำหนดค่าเริ่มต้นให้กับ $\beta_T(i)$ มีค่าเท่ากับ 1 ทุกสถานะ i ในขั้นตอนที่ 2 เป็นการหา $\beta_t(i)$ ในสถานะ j ที่เวลา t และเปลี่ยนจากสถานะ j ไปทุกๆ สถานะที่เป็นไปได้ที่เวลา $t+1$ มีสัญลักษณ์ค่าสังเกต O_{t+1} และลำดับค่าสังเกตที่สอดคล้องกับสถานะ ความซับซ้อนในการคำนวณหา $\beta_t(i)$ เหมือนกับการคำนวณหา $\alpha_t(i)$



รูปที่ 2-11 รายละเอียดของตัวแปรในการคำนวณค่าตัวแปรย้อนกลับ $\beta_t(i)$ (Rabiner, 1989)

จากที่กล่าวมาทั้งขั้นตอนวิธีการไปหน้าและย้อนหลังใช้เพื่อคำนวณหา $P(O|\lambda)$ เพื่อแก้ปัญหาในข้อที่ 1

การแก้ไขข้อที่ 2

ปัญหาพื้นฐานข้อที่ 2 เป็นปัญหาที่เกี่ยวข้องกับการหาลำดับสถานะที่เหมาะสมที่สุดที่สัมพันธ์กับลำดับค่าสังเกตที่กำหนดให้ ความยากในการแก้ปัญหานี้ขึ้นอยู่กับนิยามของลำดับสถานะที่เหมาะสมที่สุดที่เป็นไปได้มากมาย กฎเกณฑ์หนึ่งที่เป็นไปได้ก็คือ การเลือกสถานะ q_t ที่ให้ความคล้ายคลึงกับลำดับค่าสังเกตมากที่สุด กฎเกณฑ์ของความเหมาะสมที่สุดนี้จะทำให้ค่าประมาณของจำนวนสถานะที่ถูกต้องมีค่ามากที่สุด กำหนดให้ตัวแปรมีค่าดังนี้

$$\gamma_t(i) = P(q_t = S_i | O, \lambda) \quad (2-69)$$

ซึ่งเป็นค่าความน่าจะเป็นของการอยู่ในสถานะ S_i ที่เวลา t เมื่อกำหนดลำดับค่าสังเกต O และแบบจำลอง λ มาให้ สมการที่ (2-69) สามารถเขียนอยู่ในรูปของตัวแปรไปหน้าและตัวแปรย้อนกลับได้ดังสมการที่ (2-70)

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{P(O|\lambda)} = \frac{\alpha_t(i)\beta_t(i)}{\sum_{j=1}^N \alpha_t(j)\beta_t(j)} \quad (2-70)$$

เมื่อ $\alpha_t(i)$ เป็นค่าความน่าจะเป็นของลำดับค่าสังเกตบางส่วน O_1, O_2, \dots, O_t และสถานะ S_i ที่เวลา t ขณะที่ $\beta_t(i)$ เป็นค่าความน่าจะเป็นของลำดับค่าสังเกตที่คงเหลืออยู่ $O_{t+1}, O_{t+2}, \dots, O_T$

โดยการใช้ $\gamma_t(i)$ จะสามารถหาค่าสถานะที่ตัวต่อตัวคล้ายคลึงกันมากที่สุด q_t ที่เวลา t ได้ดังนี้

$$q_t = \arg \max_{1 \leq i \leq N} [\gamma_t(i)] \quad (2-71)$$

แม้ว่าสมการที่ (2-71) จะทำให้การประมาณของจำนวนสถานะที่ถูกต้องมีค่ามากที่สุด โดยการเลือกสถานะที่คล้ายคลึงกันมากที่สุดสำหรับแต่ละ t ก็ตาม ก็ยังเกิดปัญหาขึ้นกับผลของลำดับสถานะที่ได้เนื่องจากสมการที่ (2-71) เป็นเพียงการหาสถานะที่คล้ายคลึงกันมากที่สุดในขณะใดขณะหนึ่งเท่านั้น โดยไม่เกี่ยวข้องกับค่าความน่าจะเป็นในการปรากฏของลำดับสถานะ

วิธีการหนึ่งในการแก้ปัญหาที่คือการแก้ไขปัญหากฎเกณฑ์ของความเหมาะสมที่สุด โดยการคำนวณหาค่าลำดับสถานะที่ทำให้ค่าประมาณของคู่สถานะ (q_t, q_{t+1}) ที่ถูกต้องมีค่ามากที่สุด กฎเกณฑ์ของความเหมาะสมที่สุดที่นิยมใช้กันมากที่สุดก็คือ การหาลำดับสถานะที่ดีที่สุดเพียงลำดับเดียว ขั้นตอนวิธีการในการหาลำดับสถานะที่ดีที่สุดเพียงลำดับเดียวนี้อยู่บนพื้นฐานของการโปรแกรมแบบพลวัตที่เรียกว่า “ขั้นตอนวิธีการ Viterbi”

ขั้นตอนวิธีการ Viterbi (Viterbi Algorithm)

ขั้นตอนวิธีการ Viterbi เป็นขั้นตอนวิธีการหาลำดับสถานะของแบบจำลองฮิดเดนมาร์คอฟที่ถูกซ่อน เนื่องจากว่าลำดับสถานะนี้บอกถึงโครงสร้างของแบบจำลอง ค่าเฉลี่ยทางสถิติและลักษณะอื่นๆ ของแบบจำลองที่แสดงได้จากลำดับสถานะ วิธีในการหาลำดับสถานะที่เป็นไปได้เมื่อกำหนดลำดับค่าสังเกตมาให้คือการเลือกสถานะ s_t ด้วยวิธีที่เหมาะสมเพื่อให้ได้เส้นทางที่ดีที่สุดและมีความน่าจะเป็น $P(O, S | \lambda)$ สูงที่สุด เทคนิคที่ใช้ในการหาลำดับสถานะที่ดีที่สุดเพียงลำดับเดียวเรียกว่า “ขั้นตอนวิธีการ Viterbi” ซึ่งคล้ายๆ กับขั้นตอนวิธีการเทียบทางเวลาแบบพลวัต (Dynamic Time Warping, DTW) โดยกำหนดให้ลำดับค่าสังเกต $O = \{O_1, O_2, \dots, O_T\}$ และมีลำดับสถานะที่ดีที่สุดคือ $S = \{s_1, s_2, \dots, s_T\}$ กำหนดตัวแปรได้ดังนี้

$$\delta_t(i) = \max_{s_1, s_2, \dots, s_{t-1}} P[s_1, s_2, \dots, s_{t-1}, s_t = i, O_1, O_2, \dots, O_{t-1} | \lambda] \quad (2-72)$$

โดยที่ $\delta_t(i)$ เป็นความน่าจะเป็นที่มีค่าสูงสุดของเส้นทางเดียวที่เวลา t ซึ่งเป็นค่าสังเกต t ค่าแรกและสิ้นสุดในสถานะ i ด้วยวิธีอุปนัยดังนี้

$$\delta_{t+1}(j) = \left[\max_i \delta_t(i) a_{ij} \right] \cdot b_j(O_{t+1}) \quad (2-73)$$

ในการเรียกใช้ค่าลำดับสถานะจำเป็นต้องติดตามค่าอาร์กิวเมนต์ที่ทำให้สมการที่ (2-73) มีค่ามากที่สุดสำหรับแต่ละค่า t และ j โดยอาศัยแถวลำดับ $\psi_t(j)$ รายละเอียดของขั้นตอนวิธีการ Viterbi มีดังนี้

ขั้นตอนที่ 1 กำหนดเงื่อนไขเริ่มต้นให้กับทุกๆ สถานะ i

$$\delta_1(i) = \pi_i b_i(O_1), \quad 1 \leq i \leq N \quad (2-74)$$

$$\psi_1(i) = 0 \quad (2-75)$$

ขั้นตอนที่ 2 การวนซ้ำ

$$\delta_t(j) = \max_i [\delta_{t-1}(i) a_{ij}] b_j(O_t) \quad (2-76)$$

$$\psi_t(j) = \arg \max_i [\delta_{t-1}(i) a_{ij}] \quad (2-77)$$

ขั้นตอนที่ 3 การสิ้นสุด

$$P^* = \max_{s \in S_F} [\delta_T(s)] \quad (2-78)$$

$$s_T^* = \arg \max_{s \in S_F} [\delta_T(s)] \quad (2-79)$$

ขั้นตอนที่ 4 การย้อนกลับเส้นทาง

$$s_t^* = \psi_{t+1}(s_{t+1}^*) , t = T-1, T-2, \dots, 1 \quad (2-80)$$

จากขั้นตอนวิธีการไปหน้าและย้อนกลับที่กล่าวไปก่อนหน้านี้นี้เป็นการหาความน่าจะเป็น $P(O|\lambda)$ ซึ่งความน่าจะเป็นนี้เป็นผลรวมของ $P(O, S|\lambda)$ บนทุกๆ ลำดับสถานะที่เป็นไปได้ S แต่ขั้นตอนวิธีการ Viterbi เป็นเพียงการหาประสิทธิภาพของค่าสูงสุดของความน่าจะเป็น $P(O, S|\lambda)$ ในทุกๆ สถานะ S ดังนั้นจึงมองว่าขั้นตอนวิธีการ Viterbi เป็นขั้นตอนวิธีการในกรณีพิเศษของขั้นตอนวิธีการไปหน้าและย้อนกลับ

การแก้ไขปัญหาค่าพื้นฐานข้อที่ 3

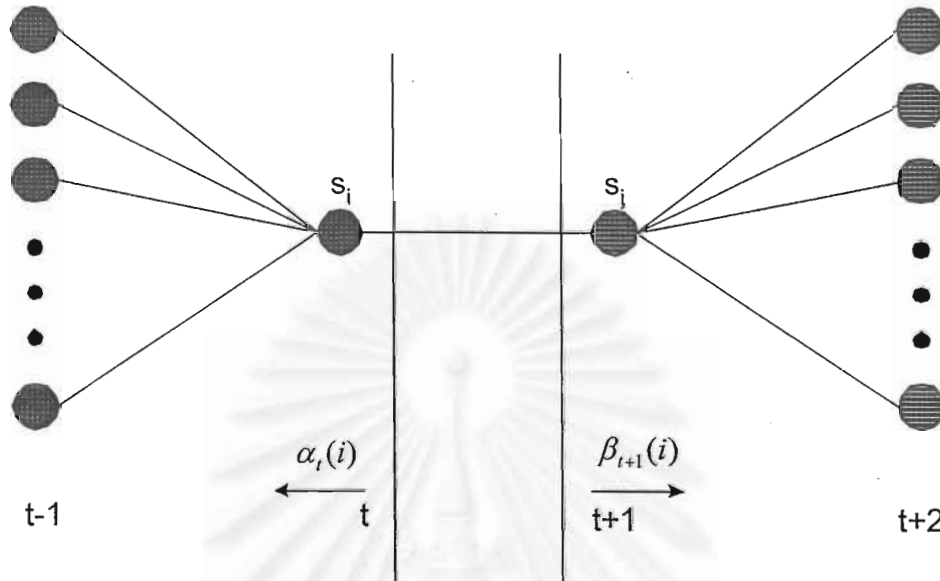
ปัญหาค่าพื้นฐานข้อที่สามเกี่ยวข้องกับการค้นหาวิธีการปรับค่าพารามิเตอร์ของแบบจำลอง $\lambda = (A, B, \pi)$ เพื่อให้ค่าความน่าจะเป็นของลำดับค่าสังเกตมีค่ามากที่สุดเมื่อกำหนดค่าพารามิเตอร์ของแบบจำลองมาให้ เนื่องจากไม่มีวิธีการที่แน่นอนในการวิเคราะห์เพื่อแก้ไขปัญหาค่าพื้นฐานข้อที่สามของแบบจำลองที่จะให้ค่าความน่าจะเป็นของลำดับค่าสังเกตมีค่ามากที่สุด ถึงแม้ว่าจะกำหนดลำดับค่าสังเกตที่จำกัดให้เป็นข้อมูลฝึกฝนก็ตามก็ยังไม่มีความเหมาะสมที่สุดในการประมาณค่าพารามิเตอร์ของแบบจำลอง แต่จะสามารถเลือก $\lambda = (A, B, \pi)$ ที่ทำให้ $P(O|\lambda)$ มีค่ามากที่สุดโดยใช้กระบวนการวนซ้ำของ Baum-Welch

ขั้นตอนวิธีการประมาณค่าซ้ำของ Baum-Welch (Baum-Welch Reestimation Algorithm)

ปัญหาที่ยากที่สุดในแบบจำลองฮิดเดนมาร์คอฟคือทำอย่างไรจึงจะปรับค่าพารามิเตอร์ของแบบจำลอง $\lambda = (A, B, \pi)$ เพื่อให้ค่าความน่าจะเป็นของลำดับค่าสังเกตมีค่ามากที่สุดเมื่อกำหนดค่าพารามิเตอร์ของแบบจำลองมาให้ เนื่องจากไม่มีวิธีการที่แน่นอนในการหาความน่าจะเป็นดังกล่าวให้ได้มากที่สุด ดังนั้นจึงได้มีการนำขั้นตอนวิธีการแบบวนซ้ำหรือเทคนิคเกรเดียนต์ (Gradient Technique) เพื่อหาค่าที่เหมาะสม ขั้นตอนวิธีการวนซ้ำที่นำมาใช้กับแบบจำลองฮิดเดนมาร์คอฟคือขั้นตอนวิธีการของ Baum-Welch

กำหนดให้ $\xi_t(i, j)$ เป็นความน่าจะเป็นของการอยู่ในสถานะ i ที่เวลา t และสถานะ j ที่เวลา $t+1$ เมื่อกำหนดแบบจำลองและลำดับค่าสังเกตมาให้ สามารถหาได้ดังสมการที่ (2-81)

$$\xi_t(i, j) = \Pr(s_t = i, s_{t+1} = j | O, \lambda) \quad (2-81)$$



รูปที่ 2-12 ลำดับของเหตุการณ์ที่ต้องใช้ในการคำนวณเหตุการณ์ร่วมของระบบในสถานะ i ที่เวลา t และสถานะ j ที่เวลา $t+1$ (Rabiner, 1989)

ลำดับของเหตุการณ์ที่นำไปสู่เงื่อนไขที่จำเป็นในสมการที่ (2-81) แสดงในรูปที่ 2-12 จากนิยามของตัวแปรไปหน้าและย้อนกลับสามารถเขียน $\xi_t(i, j)$ ได้ดังนี้

$$\begin{aligned} \xi_t(i, j) &= \frac{\alpha_t(i) b_j(O_{t+1}) \beta_{t+1}(j)}{\Pr(O | \lambda)} \\ &= \frac{\alpha_t(i) b_j(O_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^N \sum_{j=1}^N \alpha_t(i) b_j(O_{t+1}) \beta_{t+1}(j)} \end{aligned} \quad (2-82)$$

เมื่อพจน์เศษเป็น $P(s_t = i, s_{t+1} = j, O | \lambda)$ และหารด้วย $P(O | \lambda)$

กำหนดให้ $\gamma_t(i)$ เป็นความน่าจะเป็นของการอยู่ในสถานะ i ที่เวลา t เมื่อกำหนดลำดับค่าสังเกตและแบบจำลองมาให้ ดังนั้นความสัมพันธ์ระหว่าง $\gamma_t(i)$ และ $\xi_t(i, j)$ เกิดจากผลรวมในสถานะ j ทั้งหมดดังนี้

$$\gamma_t(i) = \sum_{j=1}^N \xi_t(i, j) \quad (2-83)$$

ถ้าผลรวมของ $\gamma_t(i)$ ในทางเวลาเป็นจำนวนของเวลาที่คาดว่าจะมีเข้าไปในสถานะ i หรือเป็นจำนวนการเปลี่ยนที่คาดว่าจะมีจากสถานะ i ผลรวมของ $\xi_t(i, j)$ บน t คือ $t=1$ ถึง $t=T-1$ เป็นจำนวนการเปลี่ยนที่คาดว่าจะมีจากสถานะ i ไปยังสถานะ j ดังนี้

$$\sum_{t=1}^{T-1} \gamma_t(i) \text{ คือจำนวนการเปลี่ยนที่คาดว่าจะมีจากสถานะ } i \quad (2-84)$$

$$\sum_{i=1}^{T-1} \xi_i(i, j) \text{ คือจำนวนการเปลี่ยนที่คาดว่าจะมีจากสถานะ } i \text{ ไปยังสถานะ } j \quad (2-85)$$

ใช้สมการทั้งสองที่กล่าวมาเพื่อหาวิธีการประมาณค่าพารามิเตอร์ของแบบจำลองฮิดเดน มาร์คอฟที่ประกอบด้วย π, A และ B ได้ดังนี้

$$\bar{\pi}_i = \gamma_i(i) \quad (2-86)$$

$$\bar{a}_{ij} = \frac{\sum_{i=1}^{T-1} \xi_i(i, j)}{\sum_{i=1}^{T-1} \gamma_i(i)} \quad (2-87)$$

$$\bar{b}_j(k) = \frac{\sum_{i=1}^{T-1} \gamma_i(j)}{\sum_{i=1}^{T-1} \gamma_i(j)} \quad (2-88)$$

ถ้ากำหนดให้แบบจำลองในปัจจุบันคือ $\lambda = (A, B, \pi)$ เพื่อใช้คำนวณค่าทางด้านขวาของสมการที่ (2-86) ถึง (2-88) และกำหนดให้แบบจำลองที่ประมาณค่าใหม่คือ $\bar{\lambda} = (\bar{A}, \bar{B}, \bar{\pi})$ ที่อยู่ด้านซ้ายของสมการที่ (2-86) ถึง (2-88) นอกจากนี้สมการทั้งสามยังเป็นไปตามเงื่อนไขดังนี้ แบบจำลองเริ่มต้น λ เป็นตัวกำหนดจุดวิกฤตของฟังก์ชันความน่าจะเป็นในกรณีที่มี $\bar{\lambda} = \lambda$ หรือแบบจำลอง $\bar{\lambda}$ มีความน่าจะเป็นจริงมากกว่าแบบจำลอง λ เมื่อ $P(O | \bar{\lambda}) > P(O | \lambda)$ หมายความว่าลำดับค่าสังเกตที่มีความน่าจะเป็นมากกว่าจะให้กำเนิดแบบจำลองใหม่ $\bar{\lambda}$

จากกระบวนการข้างต้น ถ้ากระทำซ้ำโดยนำ $\bar{\lambda}$ แทนที่ λ และทำการคำนวณการประมาณค่าใหม่ เพื่อเป็นการปรับปรุงความน่าจะเป็นของ O ที่ถูกสังเกตจากแบบจำลองจนถึงจุดสิ้นสุดผลลัพธ์สุดท้ายของขั้นตอนวิธีการประมาณค่านี้เรียกว่า “การประมาณความน่าจะเป็นสูงสุดของแบบจำลองฮิดเดนมาร์คอฟ”

สมการในการประมาณค่าคำนวณสามารถหาได้โดยตรงจากการหาค่าสูงสุดของฟังก์ชันเสริมของ Baum บน $\bar{\lambda}$ ดังนี้

$$Q(\lambda, \bar{\lambda}) = \sum_Q \Pr(Q | O, \lambda) \log[P(O, Q | \bar{\lambda})] \quad (2-89)$$

การทำให้ $Q(\lambda, \bar{\lambda})$ มีค่ามากที่สุดเป็นการทำให้ความน่าจะเป็นมีค่าสูงขึ้นด้วย โดยทำให้ฟังก์ชันความน่าจะเป็นเข้าสู่จุดวิกฤตจุดหนึ่งดังนี้

$$\max_{\bar{\lambda}} [Q(\lambda | \bar{\lambda})] \Rightarrow P(O | \bar{\lambda}) \geq P(O | \lambda) \quad (2-90)$$

ขั้นตอนวิธีการประมาณค่าใหม่สามารถแสดงให้เห็นเป็นขั้นตอนวิธีการหาค่าคาดหวังที่มากที่สุด (EM Algorithm) โดยขั้นตอนในการหาค่าคาดหวังคือการคำนวณหาฟังก์ชันเสริม $Q(\lambda, \bar{\lambda})$ และขั้นตอนในการหาค่ามากที่สุดก็คือการหาค่า $Q(\lambda, \bar{\lambda})$ ที่มีค่ามากที่สุดบน $\bar{\lambda}$ นอก

จากนี้ขั้นตอนวิธีการประมาณค่าใหม่จะต้องเป็นไปตามเงื่อนไขเฟ้นสุ่ม (Stochastic Constraint) ของพารามิเตอร์แบบจำลองฮิดเดนมาร์คอฟ ดังต่อไปนี้

$$\sum_{i=1}^N \bar{\pi}_i = 1 \quad (2-91)$$

$$\sum_{j=1}^N \bar{a}_{ij} = 1, \quad 1 \leq i \leq N \quad (2-92)$$

$$\sum_{k=1}^L \bar{b}_j(k) = 1, \quad 1 \leq j \leq N \quad (2-93)$$

การแก้ไขปัญหาคือพื้นฐานทั้งสามข้อนำไปสู่การสร้างแบบจำลองฮิดเดนมาร์คอฟและการนำฮิดเดนมาร์คอฟไปใช้งาน แต่เนื่องจากการนำแบบจำลองฮิดเดนมาร์คอฟไปใช้งานกับเสียงพูดโดยเฉพาะในงานวิทยานิพนธ์นี้ได้ใช้เสียงพูดที่เป็นเสียงต่อเนื่องจะทำให้เกิดปัญหาเนื่องจากการคำนวณด้วยคอมพิวเตอร์ โดยจะเกิดปัญหาที่เรียกว่า Underflow ขึ้น ซึ่งเป็นปัญหาที่ผลการคำนวณให้ค่าเลขช่วงที่คอมพิวเตอร์จะทำการคำนวณได้ โดยจะนำวิธีการสเกล (Scaling) (Rabiner, 1989) มาใช้ในการแก้ปัญหที่เกิดขึ้น

วิธีการสเกล (Scaling) (Rabiner, 1989)

พิจารณานิยามของ $\alpha_t(i)$ จากสมการ (2-61) จะเห็นว่าค่าการคำนวณหา $\alpha_t(i)$ จะอยู่ในรูปผลคูณของค่า a และค่า b ($\prod_{s=1}^{t-1} a_{q_s q_{s+1}} \prod_{s=1}^t b_{q_s}(O_s)$) แต่เนื่องจากค่า a และค่า b ต่างมีค่าน้อยกว่า 1 จะเห็นได้ว่าเมื่อ t มีค่ามาก จะทำให้ค่าที่ได้จากการคำนวณมีค่าเข้าใกล้ 0 จนทำให้ค่าที่ได้เลยช่วงการคำนวณของคอมพิวเตอร์ (โดยปกติ ค่าที่ใช้เป็นแบบ double ซึ่งจะมี dynamic range อยู่ในช่วงประมาณ 1.7×10^{-308} ถึง 1.7×10^{308})

การแก้ไขปัญหาคือการคูณค่า $\alpha_t(i)$ ด้วยสัมประสิทธิ์ของการสเกล (Scaling Coefficient) ซึ่งต้องเป็นค่าที่ขึ้นกับ t แต่ไม่ขึ้นกับ i และสำหรับ $\beta_t(i)$ ก็ทำนองเดียวกัน ค่าสัมประสิทธิ์ของการสเกลที่คูณเข้าไปนั้นลักษณะสำคัญอีกประการหนึ่งก็คือเมื่อสิ้นสุดการคำนวณแล้วค่าสัมประสิทธิ์ของการสเกลจะต้องถูกหักล้างไปโดยอัตโนมัติ (ค่าจากการคำนวณต้องมีค่าเท่าเดิมกับตอนที่ยังไม่มี การสเกล)

ค่าสัมประสิทธิ์ของการสเกลที่ใช้จะกำหนดดังสมการ

$$c_t = \frac{1}{\sum_{i=1}^N \alpha_t(i)} \quad (2-94)$$

ดังนั้นจึงจะได้ค่า $\hat{\alpha}_t(i)$ ซึ่งเป็นค่าได้รับการสเกล ดังสมการ

$$\hat{\alpha}_t(i) = \frac{\alpha_t(i)}{\sum_{i=1}^N \alpha_t(i)} \quad (2-95)$$

สำหรับค่า $\hat{\beta}_t(i)$ ก็หาได้จาก

$$\hat{\beta}_t(i) = c_t \beta_t(i) \quad (2-96)$$

ในการคำนวณค่าความน่าจะเป็น $P(O|\lambda)$ นั้นไม่สามารถหาจากสมการที่ (2-64) ได้เนื่องจากค่า α เป็นค่าที่ได้ถูกสเกลไปแล้วดังนั้นจึงมีขั้นตอนการหาค่าความน่าจะเป็น $P(O|\lambda)$ ดังนี้จากเดิม

$$P(O|\lambda) = \sum_{i=1}^N \alpha_T(i) \quad (2-97)$$

และหลังจากการสเกลจะได้

$$\prod_{t=1}^T c_t \sum_{i=1}^N \alpha_t(i) = 1 \quad (2-98)$$

นั่นคือ

$$\prod_{t=1}^T c_t P(O|\lambda) = 1 \quad (2-99)$$

ดังนั้น

$$P(O|\lambda) = \frac{1}{\prod_{t=1}^T c_t} \quad (2-100)$$

หรือ

$$\log[P(O|\lambda)] = -\prod_{t=1}^T \log c_t \quad (2-101)$$

จะเห็นว่าเมื่อคำนวณค่า P ตามสมการที่ (2-101) จะเป็นค่าที่สามารถคำนวณด้วยคอมพิวเตอร์ได้ เนื่องจากการคำนวณค่า $\log[P]$ ซึ่งต่างจากการคำนวณค่า P ในสมการ (2-100) ที่ไม่สามารถคำนวณได้เนื่องจากค่าจะอยู่นอกช่วงที่คอมพิวเตอร์สามารถคำนวณได้

สำหรับในขั้นตอนวิธีการ Viterbi จะมีการเปลี่ยนแปลงสมการจากสมการของผลคูณเป็นสมการของผลบวก ดังสมการที่ (2-101)

ขั้นตอนที่ 1 กำหนดเงื่อนไขเริ่มต้นให้กับทุกๆ สถานะ i

$$\hat{\delta}_t(i) = \log[\pi_i] + \log[b_i(O_1)] , \quad 1 \leq i \leq N \quad (2-102)$$

ขั้นตอนที่ 2 การวนซ้ำ

$$\hat{\delta}_i(j) = \max_i [\delta_{i-1}(i) + \log[a_{ij}]] + \log[b_j(O_i)] \quad (2-103)$$

ขั้นตอนที่ 3 การสิ้นสุด

$$\log[P^*] = \max_{s \in S_F} [\delta_T(s)] \quad (2-104)$$

6. การตรวจสอบผู้พูด (Verification)

เป็นขั้นตอนสุดท้ายของระบบและเป็นขั้นตอนสำคัญที่จะทำการตัดสินใจว่าเสียงพูดที่ได้รับ การบ่งชี้แล้วนั้นเป็นเสียงของผู้พูดคนนั้นจริงหรือไม่ โดยการตรวจสอบเสียงพูด จะนำค่าความน่าจะเป็นที่ได้จากแบบจำลองฮิดเดนมาร์คอฟไปเปรียบเทียบกับค่าผ่านระดับที่กำหนดไว้ค่าหนึ่ง ถ้าความน่าจะเป็นของเสียงพูดเสียงนั้นมีค่ามากกว่าค่าผ่านระดับ ระบบก็จะยอมรับเสียงๆ นั้น แต่ถ้าค่าความน่าจะเป็นมีค่าน้อยกว่าค่าผ่านระดับ ระบบก็จะปฏิเสธเสียงๆ นั้น

สำหรับค่าผ่านระดับของระบบนั้น จะได้จากการนำกลุ่มทดสอบมาทดสอบระบบที่ฝึกฝนไว้แล้ว โดยจะทำการปรับค่าผ่านระดับให้เหมาะสมเพื่อให้ได้อัตราการตรวจสอบผิดพลาดต่ำที่สุด สำหรับอัตราการตรวจสอบผิดพลาด จะมีอยู่ 2 แบบ คือ

1. อัตราการยอมรับผิดพลาด (False Acceptance, FA) คือการตัดสินใจว่าเสียงพูดของผู้พูดนอกระบบเป็นเสียงผู้พูดในระบบ
2. อัตราการปฏิเสธผิดพลาด (False Rejection, FR) คือการตัดสินใจว่าเสียงพูดของผู้พูดในระบบเป็นเสียงผู้พูดนอกระบบ

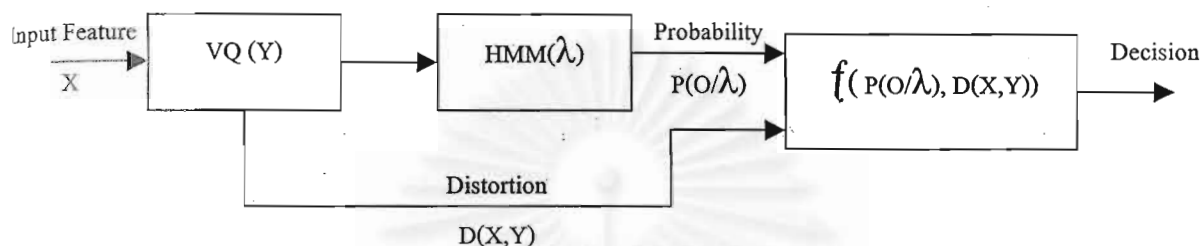
และสำหรับตำแหน่งที่จะบอกว่าการปรับค่าผ่านระดับให้อัตราการตรวจสอบผิดพลาดที่คิดแล้ว สามารถพิจารณาได้หลายเงื่อนไข แล้วแต่ความเหมาะสม ตัวอย่างเช่น

1. อัตราการยอมรับผิดพลาดเป็นศูนย์หรืออัตราการปฏิเสธผิดพลาดเป็นศูนย์ (ZFR or ZFA): สำหรับเงื่อนไขนี้จะทำการปรับค่าผ่านระดับจนกระทั่งทำให้ค่า FA หรือ FR ค่าใดค่าหนึ่งเป็นศูนย์
2. อัตราการผิดพลาดเท่ากัน (Equal Error Rate, EER) : สำหรับเงื่อนไขนี้จะทำการปรับค่าผ่านระดับจนกระทั่งทำให้ค่า FA มีค่าเท่ากับ FR
3. ผลรวมของอัตราผิดพลาดต่ำสุด : สำหรับเงื่อนไขนี้จะทำการปรับค่าผ่านระดับจนกระทั่งทำให้ผลรวมของค่า FA และ FR ต่ำสุด

จากเงื่อนไขที่ยกตัวอย่าง เมื่อเลือกใช้เงื่อนไขใดเงื่อนไขหนึ่งก็จะได้ค่าผ่านระดับของระบบมาค่าหนึ่ง จากนั้นจึงนำค่าผ่านระดับนี้ไปใช้ต่อไป

ค่าความเพี้ยน (Distortion)

สำหรับค่าความเพี้ยนเป็นค่าที่ได้จากการนำสัญญาณเสียงมาผ่านกระบวนการควอนไทซ์แบบเวกเตอร์ นั่นคือเป็นค่าความผิดพลาดเนื่องจากการควอนไทซ์ การนำค่าความเพี้ยนจากกระบวนการควอนไทซ์แบบเวกเตอร์มาใช้ร่วมกับค่าความน่าจะเป็นจากแบบจำลองฮิดเดนมาร์คอฟ มีขั้นตอนในการนำมาใช้ดังรูปที่ 2-13



รูปที่ 2-13 การนำค่าความเพี้ยนมาใช้ร่วมกับค่าความน่าจะเป็น

จากรูป เมื่อนำสัญญาณเสียงที่ผ่านกระบวนการสกัดค่าลักษณะสำคัญมาแล้ว (X) มาผ่านกระบวนการควอนไทซ์แบบเวกเตอร์ โดยหลังจากการควอนไทซ์จะได้ Y ซึ่งเป็นค่าที่แทนค่า X เมื่อผ่านกระบวนการนี้ และ $D(X,Y)$ เป็นฟังก์ชันวัดค่าความเพี้ยน โดยเป็นการวัดค่าระยะทางแบบยูคลิดีแยนระหว่าง X กับ Y และ $f(P(O/\lambda), D(X,Y))$ เป็นค่าฟังก์ชันที่นำค่าความน่าจะเป็นที่ได้จากแบบจำลอง $P(O/\lambda)$ มาพิจารณาร่วมกับค่าความเพี้ยนจากกระบวนการควอนไทซ์แบบเวกเตอร์ $D(X,Y)$ จากนั้นจึงนำค่าที่ได้จากฟังก์ชันนี้ไปพิจารณาทำการบ่งชี้ต่อไป สำหรับรูปแบบฟังก์ชันที่นำมาใช้ในการทดลอง ได้ทดลองใช้รูปแบบต่างๆ ดังต่อไปนี้ (จะเรียกค่าฟังก์ชันเหล่านี้ว่า ฟังก์ชันวัดค่าคะแนน)

1. $f_1(P(O/\lambda), D(X,Y)) = 1/D(X,Y)$
2. $f_2(P(O/\lambda), D(X,Y)) = P(O/\lambda)$:Single Codebook
3. $f_3(P(O/\lambda), D(X,Y)) = P(O/\lambda)$:Multiple Codebook
4. $f_4(P(O/\lambda), D(X,Y)) = \exp(P(O/\lambda)/1000)/D(X,Y)$
5. $f_5(P(O/\lambda), D(X,Y)) = \exp(P(O/\lambda)/1000)/(D(X,Y)^2)$
6. $f_6(P(O/\lambda), D(X,Y)) = (\exp(P(O/\lambda)^2)/1000)/D(X,Y)$

วิธีการในการนำค่าความเพี้ยนมาใช้นั้น จากฟังก์ชันที่นำเสนอ โดย f_1 เป็นฟังก์ชันที่พิจารณาเฉพาะค่าความเพี้ยนเพียงอย่างเดียว สำหรับ f_2 และ f_3 เป็นฟังก์ชันเดิมคือพิจารณาเฉพาะค่าความน่าจะเป็นที่ได้จากแบบจำลอง เมื่อใช้ชุดรหัสแบบชุดรหัสเดี่ยว กับชุดรหัสแบบหลายชุดตามลำดับ ส่วน $f_4 - f_6$ เป็นฟังก์ชันที่พิจารณาทั้งค่าความเพี้ยนกับค่าความน่าจะเป็นจากแบบจำลอง

พิจารณา $f_4 - f_6$ ที่นำฟังก์ชันเอ็กโปเนนเชียลมาใช้เนื่องจากว่า ค่าความน่าจะเป็นที่ได้จากแบบจำลองฮิดเดนมาร์คอฟเป็นค่าลอการิทึมของค่าความน่าจะเป็นจริง

ฟังก์ชันความแตกต่าง

การนำเอาฟังก์ชันความแตกต่างมาใช้ในกระบวนการตรวจสอบผู้พูดนี้เป็นแนวความคิดที่ได้มาจากสมมติฐานที่ว่า ค่าคะแนนที่ได้จากเสียงพูดของผู้พูดในระบบเมื่อผ่านแบบจำลองของผู้พูดคนนั้นจริงๆ ควรจะมีความแตกต่างๆ จากค่าคะแนนที่ได้จากการนำเสียงดังกล่าวนี้ไปผ่านแบบจำลองของผู้พูดคนอื่น ในทางตรงกันข้ามค่าคะแนนที่ได้จากเสียงพูดของผู้พูดนอกระบบเมื่อผ่านแบบจำลองของผู้พูดในระบบทุกแบบจำลองควรจะให้ค่าคะแนนที่ไม่แตกต่างกันมากนัก ดังนั้นจึงได้นิยามฟังก์ชันความแตกต่างขึ้นมา ดังสมการที่ (2-105)

$$df(m, n) = n * S_{\max} - \sum_{i=m}^{m+n-1} S_{\max-i} \quad (2-105)$$

โดยที่

$df(m, n)$ เป็นค่าฟังก์ชันความแตกต่าง

S_{\max} เป็นค่าคะแนนที่มากที่สุดที่ได้จากกระบวนการบ่งชี้ผู้พูด

$S_{\max-i}$ เป็นค่าคะแนนที่มากรองจากค่ามากที่สุดอันดับที่ i

m เป็นพารามิเตอร์ที่บอกถึงอันดับของค่าที่มากที่สุดรองลงมา ที่เริ่มนำมาลบจากค่ามากที่สุด

n เป็นพารามิเตอร์ที่บอกถึงจำนวนของค่าที่นำมาลบ จากค่าที่มากที่สุดรองลงมา เป็นอันดับที่ m

ตัวอย่างเช่น ให้ $S_1, S_2, S_3, S_4, S_5, S_6, S_7, S_8, S_9, S_{10}$ เป็นค่าคะแนนที่ได้จากการเรียงลำดับจากมากไปน้อย (S_1 เป็นค่าที่มากที่สุด, S_2 เป็นค่าที่มากรองจากค่าที่มากที่สุดเป็นอันดับที่ 1, S_3 เป็นค่าที่มากรองจากค่าที่มากที่สุดเป็นอันดับที่ 2 เป็นต้น)

ตัวอย่างการคำนวณ

$$df(1,1) = 1 * S_1 - (S_1)$$

$$df(2,1) = 1 * S_1 - (S_2)$$

$$df(3,1) = 1 * S_1 - (S_3)$$

$$df(4,1) = 1 * S_1 - (S_4)$$

$$df(1,2) = 2 * S_1 - (S_2 + S_3)$$

$$df(2,2) = 2 * S_1 - (S_3 + S_4)$$

$$df(3,2) = 2 * S_1 - (S_4 + S_5)$$

$$df(4,2) = 2 * S_1 - (S_5 - S_6)$$

$$df(1,3) = 3*S_1 - (S_2 + S_3 + S_4)$$

$$df(2,3) = 3*S_1 - (S_1 + S_4 + S_5)$$

$$df(3,3) = 3*S_1 - (S_1 + S_5 + S_6)$$

$$df(4,3) = 3*S_1 - (S_5 + S_6 + S_7)$$

$$df(1,4) = 4*S_1 - (S_2 + S_3 + S_4 + S_5)$$

$$df(2,4) = 4*S_1 - (S_3 + S_4 + S_5 + S_6)$$

$$df(3,4) = 4*S_1 - (S_4 + S_5 + S_6 + S_7)$$

$$df(4,4) = 4*S_1 - (S_5 + S_6 + S_7 + S_8)$$



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 3

ขั้นตอนวิธีในการดำเนินงานวิจัย

ในบทที่ 3 นี้จะกล่าวถึงรายละเอียดของขั้นตอนต่างๆ ที่ได้ดำเนินในงานวิจัยในวิทยานิพนธ์นี้ โดยแบ่งเป็นส่วนของขั้นตอนหลักๆ ได้แก่ การกำหนดลักษณะของบทคำพูด การเก็บบันทึกตัวอย่างเสียงพูด การสร้างระบบบ่งชี้ผู้พูด การฝึกฝนและการทดสอบระบบ ซึ่งมีรายละเอียดดังต่อไปนี้

3.1 การกำหนดลักษณะของบทคำพูด

เนื่องจากระบบบ่งชี้ผู้พูดเป็นชนิดขึ้นกับบทคำพูด ดังนั้นจึงจำเป็นต้องมีการกำหนดลักษณะของบทคำพูด โดยลักษณะของบทคำพูดที่จะนำมาใช้ในระบบควรมีคุณสมบัติที่สามารถแยกแยะความแตกต่างของเสียงผู้พูดแต่ละคนได้ดี และควรมีความยาวของคำพูดเพียงพอที่จะทำให้ระบบสามารถดึงข้อมูลที่บ่งบอกถึงความเป็นผู้พูดในแต่ละคนได้ จากงานวิทยานิพนธ์ของนายศวิต กาสุริยะ (นายศวิต กาสุริยะ, 2542) ได้ทำศึกษาลักษณะของบทคำพูดที่มีผลต่ออัตราการบ่งชี้โดยในงานวิทยานิพนธ์ดังกล่าวได้ทดลองใช้บทคำพูด 3 บทคำพูดด้วยกันคือ

บทคำพูดที่ 1 “ขอผ่านเข้าสู่ระบบ”

บทคำพูดที่ 2 “หมายเลขผ่านคือ”

บทคำพูดที่ 3 “สาม ห้า สอง เก้า ลี”

สังเกตได้ว่าบทคำพูดที่ 1 เป็นบทคำพูดที่ยาวที่สุดคือมีจำนวนพยางค์เท่ากับ 6 พยางค์ และบทคำพูดที่ 2 เป็นบทคำพูดที่สั้นที่สุดคือมีจำนวนพยางค์เท่ากับ 4 พยางค์ ส่วนบทคำพูดที่ 3 ประกอบไปด้วยตัวเลขจำนวน 5 ตัวเลขที่ประกอบด้วยเสียง “ส” และเสียงวรรณยุกต์โท 2 เสียง จัตวา 2 เสียง และเสียงเอก 1 เสียง เนื่องจากว่าเสียง “ส” เป็นเสียงเสียดแทรกที่มีความถี่สูงและมีความแตกต่างในการออกเสียงของผู้พูดแต่ละบุคคลมาก (อังสนา จามิกรณ์, 2532) ส่วนเสียงวรรณยุกต์โทและจัตวาเป็นเสียงตก (Falling Tone) และเสียงขึ้น (Rising Tone) ซึ่งเป็นเสียงที่มีการเปลี่ยนแปลงค่าความถี่มูลฐานเป็นเส้นโค้งตกและขึ้นอย่างรวดเร็ว (Dynamic Tones) (ณัฐกร ทับทอง, 2530) ทำให้มีโอกาสเกิดความแตกต่างในการออกเสียงของผู้พูดแต่ละบุคคลได้เช่นกัน

จากผลการทดลอง (นายศวิต กาสุริยะ, 2542) สรุปได้ว่า บทคำพูดที่ 3 ให้อัตราการบ่งชี้ที่ดีที่สุด ดังนั้นในงานวิทยานิพนธ์นี้ได้นำเอาบทคำพูดที่ 3 มาเป็นบทคำพูดของผู้พูดในระบบบ่งชี้ผู้พูดที่พัฒนาขึ้น

3.2 การเก็บบันทึกตัวอย่างเสียงพูด

หลังจากที่ได้กำหนดลักษณะของบทคำพูดแล้ว ก็ได้ทำการเก็บบันทึกตัวอย่างเสียงพูด โดยได้วางแผนการบันทึกไว้ 2 ช่วง ในแต่ละช่วงจะห่างกันประมาณ 1 เดือน และให้ผู้พูดบันทึกเสียง 10 ครั้งต่อหนึ่งช่วง และการบันทึกในแต่ละครั้งมีลักษณะดังนี้

- บันทึกเสียงพูดด้วยไมโครโฟนผ่านการรัศเสียงขนาด 16 บิตที่มีอัตราการสุ่มตัวอย่างเท่ากับ 11025 เฮิรตซ์
- บันทึกเสียงพูดภายใต้สภาพแวดล้อมของห้องปฏิบัติการวิจัย ซึ่งคล้ายคลึงกับสภาพแวดล้อมของสถานที่ทำงานทั่วไป (Office Environment)
- ขณะบันทึกเสียง ผู้พูดต้องอยู่ในสภาพปกติ คือไม่มีอาการป่วยหรือไม่สบายจนทำให้ออกเสียงต่างไปจากปกติ
- ในการบันทึกเสียงผู้พูดต้องใช้ภาษาไทยสำเนียงภาคกลาง

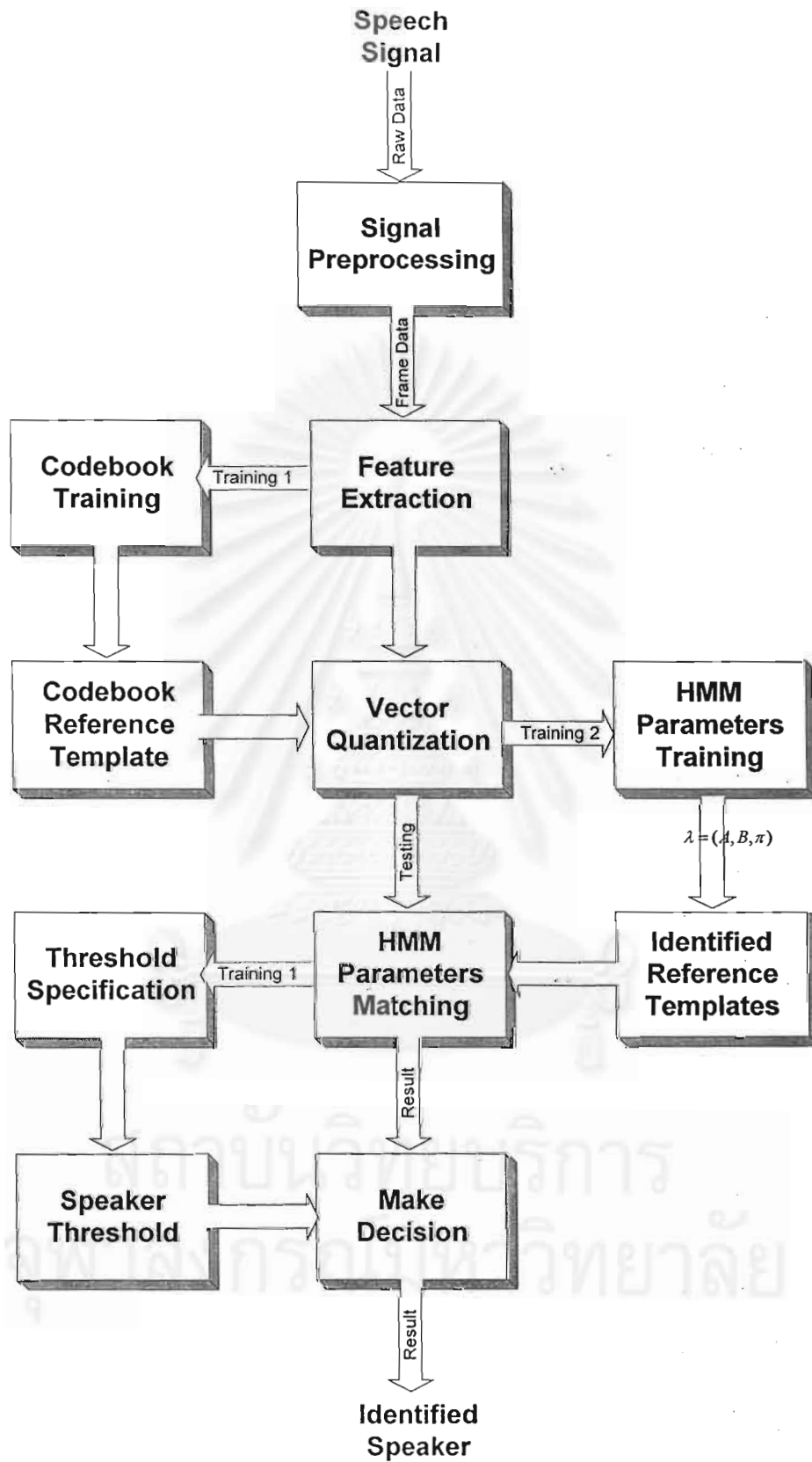
รายละเอียดของอุปกรณ์ที่ใช้ในการเก็บบันทึกเสียง

- เครื่องคอมพิวเตอร์ Pentium 200 MMX ที่มีหน่วยความจำขนาด 64 MB พร้อมด้วยฮาร์ดดิสต์ขนาด 2.1 GB และฟลอปปีดิสต์ 3.5" ขนาด 1.44 MB
- การ์ดเสียง Sound Blaster 16 ของบริษัท Creative Technology
- ไมโครโฟน Phillips Uni-directional Dynamic Microphone รุ่น SBC 465
- โปรแกรมบันทึกเสียง Goldwave Version 4.02 บนระบบปฏิบัติการ Window NT 4.0

สำหรับเสียงในแต่ละช่วงจะมีการแบ่งกลุ่มไว้สำหรับเป็นชุดฝึกฝนและชุดทดสอบ โดยจะกล่าวถึงรายละเอียดในบทถัดไป

3.3 การสร้างระบบบ่งชี้ผู้พูด

ก่อนที่จะสร้างระบบบ่งชี้ผู้พูด อันดับแรกจะต้องศึกษาขั้นตอนการทำงานของระบบ โดยระบบบ่งชี้ผู้พูดในงานวิทยานิพนธ์นี้เป็นระบบบ่งชี้ผู้พูดแบบเปิด ใช้เทคนิคการฝึกฝนชุดรหัสแบบหลายชุดรหัส และใช้แบบจำลองฮิดเดนมาร์คอฟเป็นระบบรู้จำ โดยแสดงการทำงานของระบบโดยรวมดังรูปที่ 3-1



รูปที่ 3-1 ขั้นตอนการทำงานของระบบบ่งชี้ผู้พูดแบบระบบเปิด

อธิบายการทำงานในแต่ละขั้นตอนตามลำดับดังนี้

3.3.1 การรับสัญญาณเสียงพูด (Speech Input) : ในขั้นตอนนี้จะนำเสียงพูดที่ได้จากบันทึกซึ่งอยู่ในรูปแบบแฟ้มคอมพิวเตอร์ โดยรูปแบบของแฟ้มที่ใช้เป็นแบบแฟ้มข้อมูลดิบ หรือ Raw data File ซึ่งเป็นแฟ้มที่เก็บเฉพาะข้อมูลของเสียงพูดเท่านั้น จากนั้นจึงจะนำแฟ้มเสียงนี้ไปเข้าสู่ขั้นตอนต่อไป

3.3.2 การประมวลผลสัญญาณเบื้องต้น (Signal Preprocessing) : การประมวลผลสัญญาณเบื้องต้นเป็นขั้นตอนในการเตรียมสัญญาณเสียงพูด (Speech Signal) ที่บันทึกได้มาเป็นข้อมูลเสียงพูด (Speech Data) โดยผ่านกรรมวิธีสองขั้นตอนคือ

ก. *กรรมวิธีเน้นล่งหน้า (Preemphasis)* : กรรมวิธีเน้นล่งหน้าคือการปรับระดับค่าอัตราส่วนสัญญาณเสียงต่อสัญญาณรบกวน (Signal to Noise Ratio, S/N) ให้เพิ่มสูงขึ้น โดยนำสัญญาณเสียงพูดผ่านวงจรกรอง FIR อันดับหนึ่งที่มีฟังก์ชันถ่ายโอนดังสมการที่ (2-1) และ (2-2) ในงานวิจัยนี้ได้กำหนดให้ค่าสัมประสิทธิ์ของวงจรกรอง $a = 0.95$ (วิสรุต อาษุบุตร, 2539)

ข. *การวางกรอบสัญญาณ (Windowing)* : เนื่องจากสัญญาณเสียงพูดมีการเปลี่ยนแปลงตามเวลา (Time-invariant) และ ไม่เสถียร (Unstable) เป็นสัญญาณสุ่มที่ไม่เป็นเออร์โกดิก (Non-Ergodic) และไม่เป็นสัญญาณเฟ้นสุ่ม (Non-Stochastic) ดังนั้นการประยุกต์ใช้สัญญาณเสียงพูดจึงจำเป็นต้องแบ่งสัญญาณเสียงพูดออกเป็นส่วนย่อยๆ (Rabiner and Levinson, 1981) เรียกว่า “กรอบ” (Frame) ที่มีความยาวประมาณ 10 ถึง 30 มิลลิวินาที (O’Shaughnessy, 1988) โดยในงานวิจัยจะใช้กรอบที่มีความยาวประมาณ 20 มิลลิวินาที ในงานวิจัยนี้ได้แบ่งสัญญาณเสียงพูดด้วยการคูณสัญญาณเสียงพูดกับฟังก์ชันกรอบชนิด Hamming Window ดังสมการที่ (2-3) และ (2-4) เพื่อเป็นการลดทอนแอมพลิจูดที่บริเวณปลายแต่ละข้างของกรอบข้อมูลเสียงพูดอย่างช้าๆ และหลีกเลี่ยงความไม่ต่อเนื่อง

จากรูปที่ 2-2 กำหนดให้จำนวนของข้อมูลในแต่ละกรอบสัญญาณเสียงพูด (N) มีค่าเท่ากับ 512 ช่วงประสิทธิผลของฟังก์ชันกรอบมีค่าเท่ากับ 256 และเลื่อนฟังก์ชันกรอบครั้งละ 64 (เอกฤทธิ มณีน้อย, 2541) จึงทำให้แต่ละกรอบมีข้อมูลของสัญญาณเสียงพูดเท่ากับ 256 ค่า

3.3.3 การสกัดลักษณะสำคัญ (Feature Extraction) : การสกัดลักษณะสำคัญคือขั้นตอนในการหา ลักษณะสำคัญของสัญญาณเสียงพูดจากข้อมูลเสียงพูดที่ได้จากการประมวลผลสัญญาณ เบื้องต้นแล้ว ลักษณะสำคัญที่ใช้ในงานวิจัยนี้ได้แก่ สัมประสิทธิ์การประมาณพหุระเชิง เส้น สัมประสิทธิ์เซปสตรอล และสัมประสิทธิ์เซปสตรอล บนความถี่เมล ซึ่งมีรายละเอียดการคำนวณหา ดังนี้

ก. สัมประสิทธิ์การประมาณพหุระเชิงเส้น : สัมประสิทธิ์ของการประมาณพหุระเชิงเส้นเป็นแบบจำลองการประมาณการสั้นสะท้อนของช่องทางเดินเสียงขณะออกเสียงพูด รูปร่างของช่องทางเดินเสียง ความถี่และแบนด์วิดท์ของเรโซแนนซ์ที่ช่องทางเดินเสียง สามารถหาได้จากการจำลองช่องทางเดินเสียงของมนุษย์โดยใช้แบบจำลอง All Pole และใช้วิธี อัดสหสัมพันธ์ของวิธีการ Least-squares และขั้นตอนวิธีการวนซ้ำของ Levinson-Durbin ในการคำนวณหาค่าสัมประสิทธิ์ a_k ดังที่กล่าวไว้ในหัวข้อที่ 2

ข. สัมประสิทธิ์เซปสตรอล : สัมประสิทธิ์เซปสตรอลเป็นการวิเคราะห์ลักษณะสำคัญที่มีรากฐานอยู่บนเอนVELOPE ของสเปกตรัม (Spectral Envelope) ในงานวิจัยนี้ได้คำนวณหา สัมประสิทธิ์เซปสตรอลจากค่าสัมประสิทธิ์การประมาณพหุระเชิงเส้นดังสมการที่ (2-26) และ (2-27) เนื่องจากวิธีนี้ใช้เวลาในการคำนวณน้อยกว่าและเพื่อต้องการเปรียบเทียบกับ ผลที่ได้กับสัมประสิทธิ์การประมาณพหุระเชิงเส้น

ค. สัมประสิทธิ์เซปสตรอลบนความถี่เมล : เป็นลักษณะสำคัญที่เลียนแบบการได้ยินของมนุษย์บน Mel Scale ดังแสดงในรูปที่ 2-4 สามารถแบ่งได้เป็น 2 ช่วงคือช่วงที่เป็นเชิงเส้น กับช่วงที่ไม่เป็นเชิงเส้น การคำนวณหาค่าสัมประสิทธิ์เซปสตรอลบนความถี่เมลหาได้จาก อัลกอริทึมการแปลงฟูริเยร์ (Fourier Transform) ของข้อมูลเสียงแล้วนำไปเข้าชุดวงจรกรองรูปสามเหลี่ยมดังรูปที่ 2-5 ในที่นี้ได้กำหนดให้อันดับของการวิเคราะห์ (N) มีค่าเท่ากับ 20 คือมีชุดวงจรกรองรูปสามเหลี่ยมจำนวนเท่ากับ 20 วงจร แล้วคำนวณหาค่าล็อกของ พลังงานที่เอาท์พุทของแต่ละวงจรได้เป็น X_k เมื่อ $k = 1, 2, \dots, N$ และนำ X_k ไปคำนวณหาสัมประสิทธิ์เซปสตรอลบน ความถี่เมลตามสมการที่ (2-31)

3.3.4 การฝึกฝนชุดรหัส (Codebook Training) : สำหรับชุดรหัสที่ใช้ในงานในระบบบ่งชี้ผู้พูด สามารถแบ่งออกได้เป็น 2 ชนิด ได้แก่ ชุดรหัสแบบชุดรหัสเดี่ยว (Single Codebook) กับ ชุดรหัสแบบหลายชุดรหัส (Multiple Codebook) ชุดรหัสแบบชุดรหัสเดี่ยวเป็นชุดรหัสที่ได้จากการนำเสียงพูดของคนในระบบทุกคนมาฝึกฝน ดังนั้นในระบบที่ใช้ชุดรหัสแบบชุดรหัสเดี่ยวก็จะมีชุดรหัสในระบบเพียงชุดเดียวเท่านั้น ส่วนชุดรหัสแบบหลายชุดรหัสเป็นชุดรหัสที่ได้จากการนำเสียงพูดแต่ละคนมาฝึกฝน โดยที่ระบบที่ใช้ชุดรหัสแบบหลายชุดรหัสก็จะมีจำนวนชุดรหัสเท่ากับจำนวนผู้พูดในระบบ

ในงานวิทยานิพนธ์ของนายศวิต กาสุริยะ ซึ่งเป็นงานเกี่ยวกับการบ่งชี้ผู้พูดระบบเปิดและได้มีการนำชุดรหัสแบบชุดรหัสเดี่ยวไปใช้ สำหรับในงานวิทยานิพนธ์นี้จะนำชุดรหัสแบบหลายชุดรหัสไปใช้

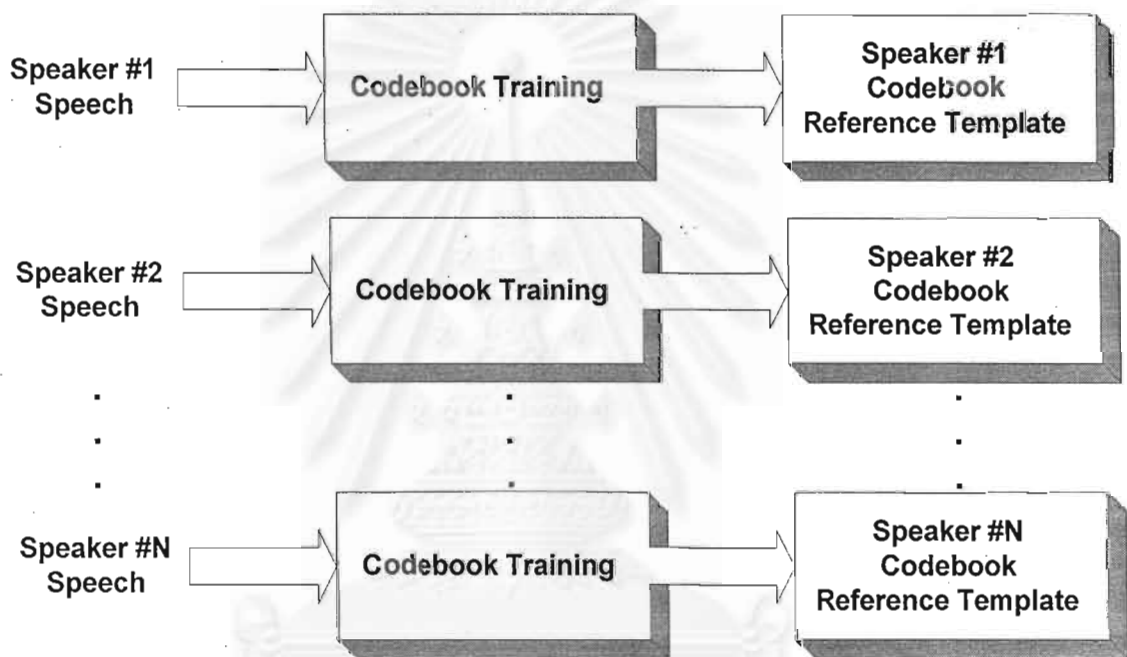
- 3.3.5 การควอนไทซ์แบบเวกเตอร์ (Vector Quantization) : การควอนไทซ์เป็นวิธีที่นำมาใช้ในการลดขนาดข้อมูล การควอนไทซ์มีอยู่ 2 แบบ คือ การควอนไทซ์แบบสเกลาร์ กับการควอนไทซ์แบบเวกเตอร์ สำหรับข้อมูลเสียงพูดเมื่อผ่านกระบวนการสกัดค่าลักษณะสำคัญก็จะได้ข้อมูลในรูปแบบเวกเตอร์ของค่าลักษณะสำคัญ (จำนวนมิติเท่ากับจำนวนอันดับของค่าลักษณะสำคัญ) โดยการควอนไทซ์จะนำค่าข้อมูลมาเปรียบเทียบกับจุดในชุดรหัสที่ใกล้เคียงหรือมีการผิดเพี้ยน (Distortion) น้อยที่สุดก็จะแทนค่าข้อมูลนั้นด้วยค่าของจุดที่กำหนดไว้ในชุดรหัส ซึ่งเป็นจุดที่ถูกกำหนด (label) ด้วยตัวเลข 1 มิติ นั่นคือการควอนไทซ์แบบเวกเตอร์เป็นการลดขนาดข้อมูลจากข้อมูลหลายมิติเป็นข้อมูลมิติเดียว
- 3.3.6 แบบจำลองฮิดเดนมาร์คอฟ (Hidden Markov Model) : แบบจำลองฮิดเดนมาร์คอฟเป็นแบบจำลองที่นิยมนำมาใช้ในการรู้จำสัญญาณ สำหรับการนำแบบจำลองฮิดเดนมาร์คอฟมาใช้งานในด้านการรู้จำผู้พูด ก็จะนำเสียงพูดของแต่ละคนมาทำการฝึกฝนเมื่อให้ได้แบบจำลองอ้างอิงของผู้พูดแต่ละคน และในขั้นตอนการรู้จำก็จะนำเสียงที่ต้องการทดสอบไปผ่านแบบจำลองอ้างอิงของผู้พูดแต่ละคน ระบบจะทำการบ่งชี้เสียงจากแบบจำลองอ้างอิงที่ให้ค่าความน่าจะเป็นของเสียงๆ นั้นมากที่สุด
- 3.3.7 การตัดสินใจความถูกต้องของเสียงพูดจากการบ่งชี้ (Making Decision) : ในขั้นตอนนี้จะนำค่าที่ได้จากขั้นตอนการบ่งชี้มาเปรียบเทียบกับค่าผ่านระดับ ถ้าค่าคะแนนที่วัดได้มีค่ามากกว่าค่าผ่านระดับก็จะตัดสินใจว่าเสียงพูดเสียงนั้นว่าเป็นเสียงพูดของผู้พูดในระบบ แต่ถ้าค่าคะแนนมีค่าต่ำกว่าค่าผ่านระดับ ก็จะปฏิเสธเสียงพูดเสียงนั้น
- สำหรับขั้นตอนในการหาค่าผ่านระดับนั้น จะหาได้จากการนำชุดทดสอบชุดหนึ่งมาทดสอบแล้วหาค่าผ่านระดับที่ให้ได้เงื่อนไขอัตราการยอมรับผิดพลาดเท่ากับอัตราการปฏิเสธผิดพลาด หรือผลรวมของค่าทั้งสองน้อยที่สุด

3.4 การฝึกฝนระบบ

จากแผนภาพแสดงขั้นตอนการทำงานของระบบดังที่ได้แสดง จะเห็นว่าระบบจะบ่งชี้ผู้พูดได้นั้นจะต้องทำให้เรียนรู้เสียงพูดของผู้พูดเสียก่อน นั่นคือจะต้องทำการฝึกฝนระบบเพื่อจะทำให้ระบบสามารถรู้จำเสียงของผู้พูดได้

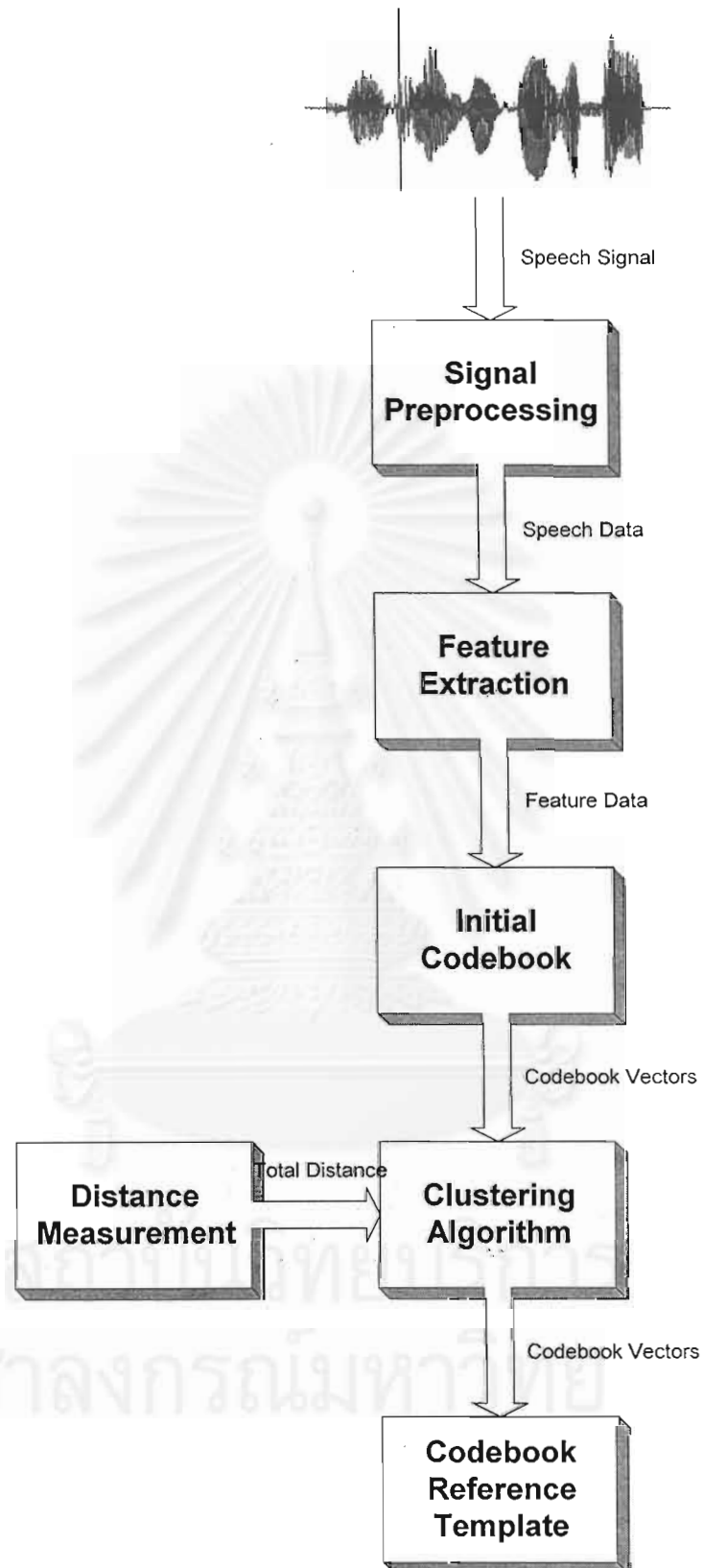
สำหรับการฝึกฝนระบบบ่งชี้ผู้พูดนี้แบ่งการฝึกฝนออกเป็น 2 ส่วนด้วยกัน ได้แก่

3.4.1 การฝึกฝนชucrห้ส : เนื่องจากชucrห้สที่ใช้เป็นแบบชucrห้สแบบหลายชucrห้ส ซึ่งแสดงรูปแบบการฝึกฝนดังรูปที่ 3-2



รูปที่ 3-2 รูปแบบการฝึกฝนชucrห้สแบบหลายชucrห้ส

สำหรับขั้นตอนในการฝึกฝนแสดงดังรูปที่ 3-3



รูปที่ 3-3 ช่วงการฝึกฝนต้นแบบอ้างอิงชุดรหัส

Clustering Algorithm ได้ใช้ขั้นตอนกรรมวิธี K-mean และการวัดค่าความเพี้ยน ใช้วิธีการวัดระยะทางยูคลีเดียน เนื่องจากการฝึกฝนแต่ละครั้งจุดที่ได้จากการฝึกฝนเป็นแบบ local optimum เพื่อให้จุดมีค่าใกล้เคียงกับจุด global optimum จึงได้ทำการฝึกฝนซ้ำทั้งหมด 5 ครั้งและนำแต่ละครั้งมาเฉลี่ย แล้วฝึกฝนใหม่อีกครั้งจากจุดที่ได้จากการเฉลี่ย

3.4.2 การฝึกฝนแบบจำลองฮิดเดนมาร์คอฟ

ขั้นตอนนี้ถือเป็นหัวใจของระบบ เพราะระบบจะสามารถรู้จำเสียงได้นั้นจำเป็นต้องนำเสียงพูดของผู้พูดมาฝึกฝนให้ระบบได้เรียนรู้ การฝึกฝนแบบจำลองอ้างอิงของผู้พูดนั้นมีขั้นตอนการทำงานดังรูปที่ 3-4

ในการฝึกฝนแบบจำลองจะใช้กรรมวิธี Forward-Backward ในการหาค่าความน่าจะเป็นของแบบจำลองและใช้กรรมวิธี Baum-Welch เพื่อปรับพารามิเตอร์ของแบบจำลองให้ได้ค่าความน่าจะเป็นที่มากที่สุด

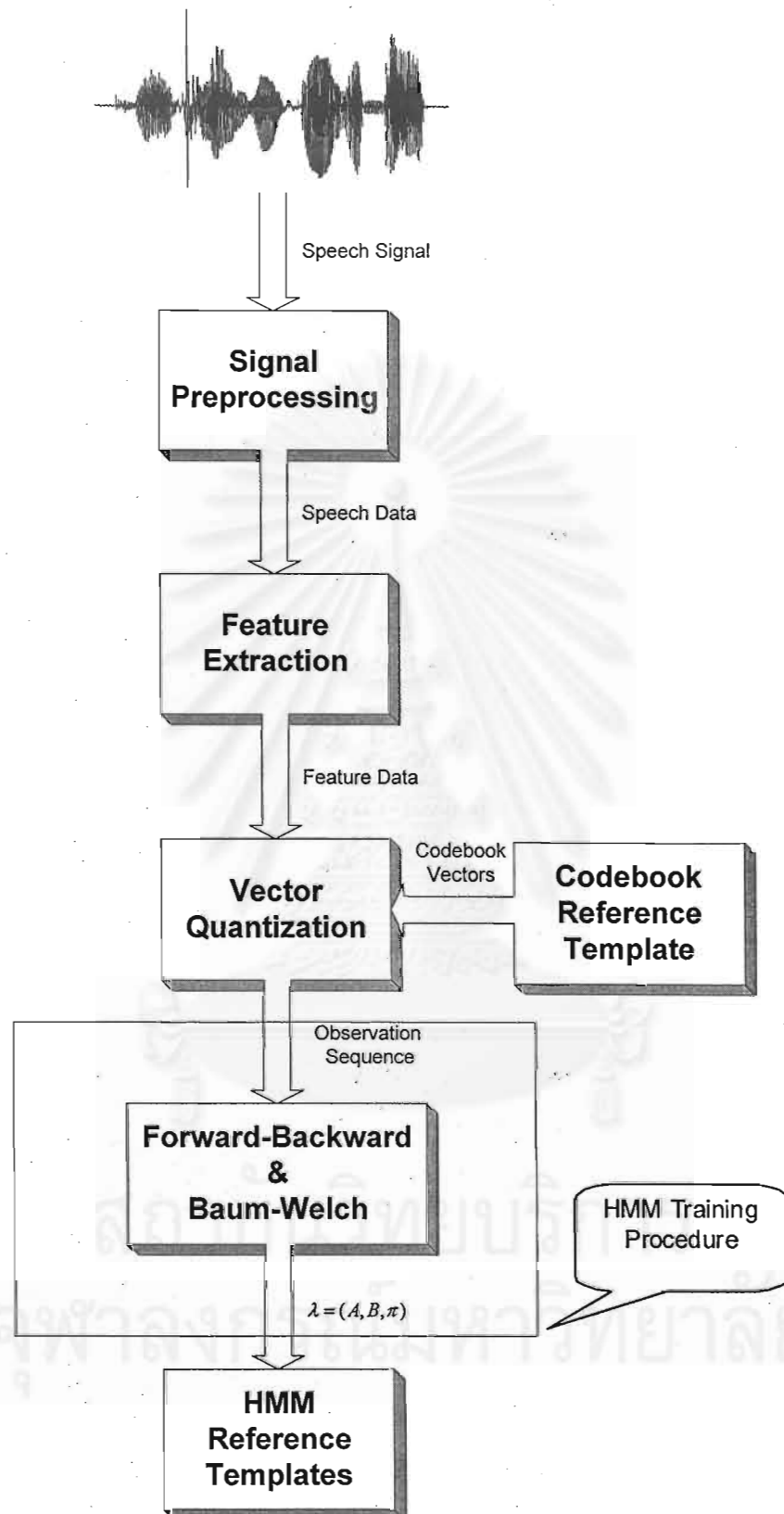
3.5 การทดสอบระบบ

เมื่อได้ฝึกฝนระบบเรียบร้อยแล้ว ก็จะทดสอบระบบจากชุดทดสอบที่ได้กำหนดไว้ ในการทดสอบนั้น ได้แบ่งออกเป็น 2 การทดสอบด้วยกันคือ การทดสอบแบบระบบปิด กับ การทดสอบแบบระบบเปิด โดยจะอธิบายขั้นตอนการทดสอบของทั้งสองแบบตามลำดับดังนี้

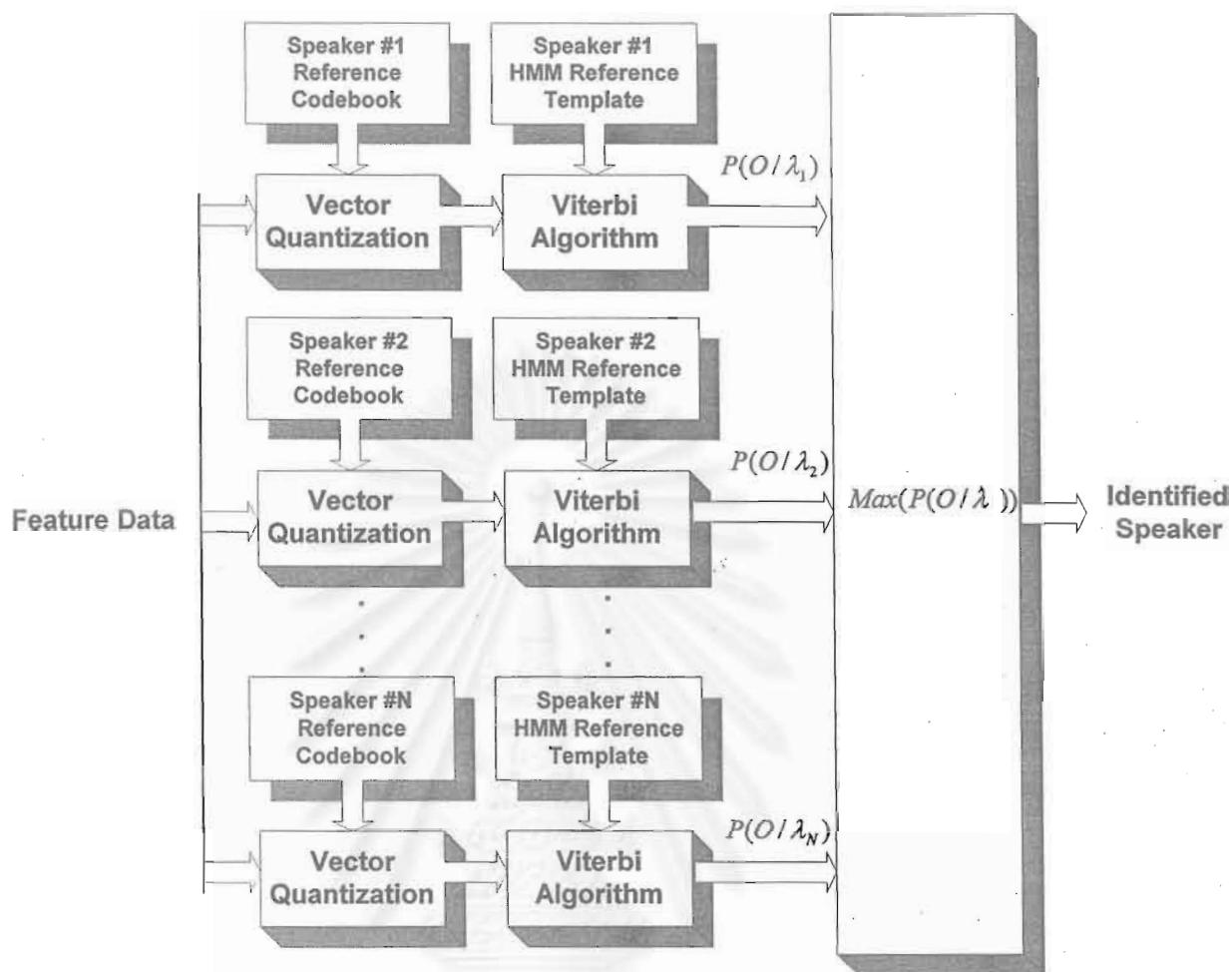
3.5.1 การทดสอบแบบระบบปิด

เป็นการทดสอบกับกลุ่มทดสอบที่ประกอบด้วยเสียงพูดของคนในระบบ โดยแสดงขั้นตอนการทดสอบดังรูปที่ 3-5

สำหรับการทดสอบนั้นก็จะนำเสียงที่ต้องการทดสอบมาผ่านขั้นตอนการควอนไทซ์แบบเวกเตอร์เทียบกับชุดรหัสของทุกคนในระบบ จากขั้นตอนนี้ก็จะได้ลำดับค่าสังเกต จากนั้นก็นำลำดับค่าสังเกตไปผ่านกรรมวิธี Viterbi ซึ่งอ้างอิงแบบจำลองของแต่ละคน เพื่อหาค่าความน่าจะเป็นของลำดับค่าสังเกต แล้วทำการเปรียบเทียบค่าความน่าจะเป็นที่ได้ โดยระบบจะบ่งชี้ผู้พูดจากแบบจำลองที่ให้ค่าความน่าจะเป็นมากที่สุด



รูปที่ 3-4 ขั้นตอนของการฝึกฝนแบบจำลองฮิดเดนมาร์คอฟ

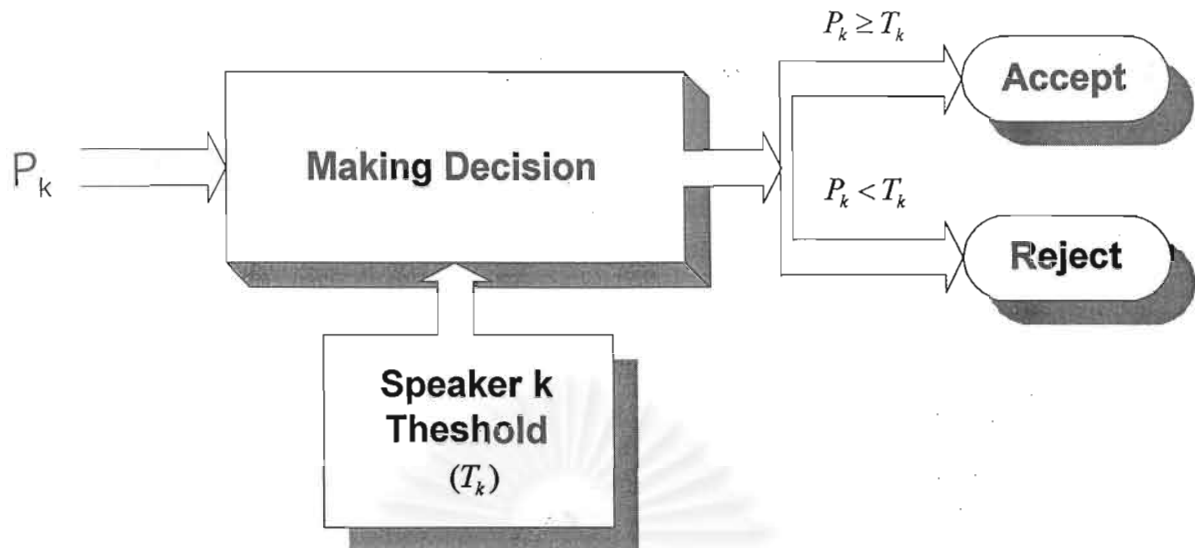


รูปที่ 3-5 ขั้นตอนการทดสอบระบบบ่งชี้ผู้พูดแบบระบบปิด

3.5.2 การทดสอบระบบเปิด

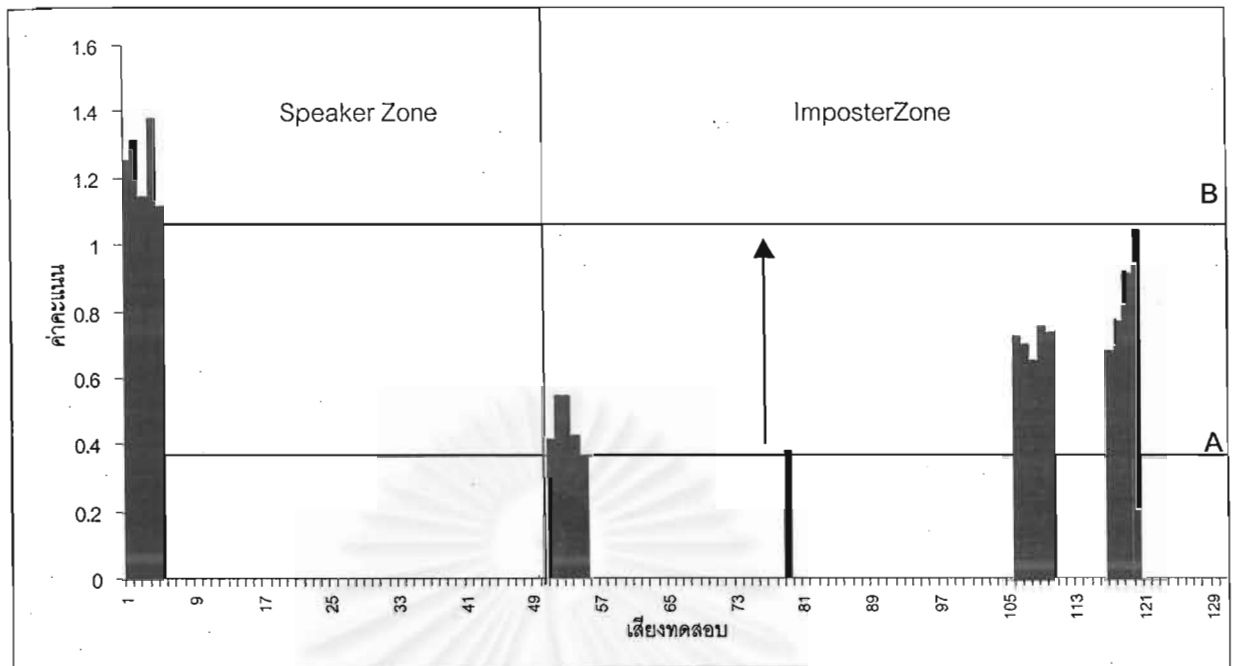
สำหรับระบบเปิดนั้น ก่อนที่ระบบจะบ่งชี้ว่าเป็นผู้พูดคนใดคนหนึ่งในระบบ ระบบจะนำค่าความน่าจะเป็นที่ได้จากแบบจำลองที่มากที่สุดมาเปรียบเทียบกับค่าผ่านระดับของผู้พูดที่แบบจำลองให้ค่าความน่าจะเป็นที่มากที่สุดนั้น ถ้าค่าความน่าจะเป็นที่มีค่ามากกว่าค่าผ่านระดับก็จะยอมรับเสียงพูดเสียงนั้นว่าเป็นผู้ในระบบคนที่ถูกบ่งชี้จริง แต่ถ้าค่าต่ำกว่าค่าผ่านระดับก็จะปฏิเสธเสียงพูดเสียงว่าเป็นเสียงพูดของผู้พูดนอกระบบ

สำหรับขั้นตอนการทดสอบในขั้นแรกจะเหมือนกับการทดสอบแบบระบบปิดทุกประการ และจะเพิ่มเติมส่วนของการตรวจสอบเข้ามาดังแสดงในรูปที่ 3-6



รูปที่ 3-6 ขั้นตอนการตรวจสอบเสียงพูดที่ได้จากการบ่งชี้

สำหรับค่าผ่านระดับ ซึ่งเป็นค่าที่ได้จากการทดสอบระบบเพื่อกำหนดค่าผ่านระดับจากกลุ่มทดสอบกลุ่มหนึ่ง โดยในการหาค่าผ่านระดับ จะมีการปรับค่าผ่านระดับจนกว่าจะทำให้ได้ผลรวมของจำนวนเสียงที่ยอมรับผิดพลาดกับจำนวนเสียงที่ปฏิเสธผิดพลาดมีค่าน้อยที่สุด ดังแสดงวิธีการหาในรูปที่ 3-7



รูปที่ 3-7 ตัวอย่างการหาค่าผ่านระดับของผู้พูด

จากรูปที่ 3-7 เริ่มต้นจะกำหนดค่าผ่านระดับจากค่าต่ำสุดของค่าคะแนนของเสียงพูดของคนนอกระบบ (จุด A) จากนั้นก็เพิ่มค่าผ่านระดับไปเรื่อยๆ จนกระทั่งทำให้ผลรวมของจำนวนเสียงที่ยอมรับผิดพลาดกับจำนวนเสียงที่ปฏิเสธผิดพลาดมีค่าน้อยที่สุด (จุดB) และที่จุดนี้ก็จะเป็ค่าผ่านระดับเพื่อจะนำไปใช้ต่อไป

บทที่ 4

ผลการทดลองและการวิเคราะห์ผลการทดลอง

ในบทนี้จะกล่าวถึงรายละเอียดของการทดลอง ผลการทดลอง และในตอนท้ายของแต่ละส่วนจะเป็นการวิเคราะห์ผลการทดลอง ในช่วงแรกจะกล่าวถึงลักษณะของฐานข้อมูลเสียงพูดที่ใช้ จากนั้นจะนำเสนอผลการทดลอง โดยได้แบ่งการทดลองออกเป็น 6 การทดลอง ประกอบด้วย

- การทดลองที่ 4-1 : เป็นการทดลองเปรียบเทียบประสิทธิภาพระบบบ่งชี้ผู้พูดแบบระบบปิด ระหว่างระบบที่ใช้ชุดรหัสแบบชุดรหัสเดี่ยว กับระบบที่ใช้ชุดรหัสแบบหลายชุดรหัส
- การทดลองที่ 4-2 : เป็นการทดลองที่ได้นำเอาค่าความผิดพลาดเพื่อนำมาใช้ร่วมกันกับค่าความน่าจะเป็นในการรู้จำเสียงของผู้พูด เพื่อปรับปรุงประสิทธิภาพของระบบบ่งชี้ผู้พูดแบบระบบปิดในระบบที่ใช้ชุดรหัสแบบหลายชุดรหัส
- การทดลองที่ 4-3 : เป็นทดลองที่ศึกษาถึงผลของพารามิเตอร์ของระบบบ่งชี้ผู้พูดทั้งแบบระบบปิดและแบบระบบเปิด ซึ่งได้แก่ อันดับของค่าลักษณะสำคัญ ขนาดของชุดรหัส และจำนวนสถานะของแบบจำลองฮิดเดนมาร์คอฟ
- การทดลองที่ 4-4 : เป็นการทดลองเพื่อศึกษาถึงผลของการฝึกฝนแบบจำลองในแต่ละครั้งว่า ให้ผลของระบบแตกต่างกันมากน้อยเพียงใด
- การทดลองที่ 4-5 : เป็นการทดลองเพื่อพัฒนาวิธีการตรวจสอบผู้พูดในขั้นตอนการตรวจสอบผู้พูด โดยจะนำเสนอฟังก์ชันความแตกต่างมาใช้
- การทดลองที่ 4-6 : เป็นการทดลองเพื่อศึกษาถึงผลของจำนวนของผู้พูดในระบบต่อประสิทธิภาพของระบบ

โดยเริ่มจากการเปรียบเทียบอัตราการบ่งชี้ผู้พูดระหว่างระบบชุดรหัสเดี่ยว (Single Codebook) กับระบบหลายชุดรหัส (Multiple Codebook) ต่อมาจะเป็นการนำเสนอผลของพารามิเตอร์ต่างๆ ของระบบ ซึ่งได้แก่ อันดับของค่าลักษณะสำคัญ ขนาดของชุดรหัส และจำนวนสถานะของแบบจำลอง ต่ออัตราการบ่งชี้ของระบบ โดยการทดลองที่กล่าวมานี้จะเป็นการทดลองกับระบบบ่งชี้ผู้พูดแบบระบบปิด (Closed Set) จากนั้นก็จะนำเสนอระบบบ่งชี้ผู้พูดที่เป็นแบบระบบเปิด (Open Set) และสุดท้ายจะเป็นการวิเคราะห์ผลการทดลองที่ได้

โดยจะได้อธิบายในรายละเอียดของแต่ละส่วน ดังต่อไปนี้

ฐานข้อมูลเสียงพูด

สำหรับฐานข้อมูลเสียงที่ใช้ในการทดลองแยกเป็น 3 ชุดหลักๆ ได้แก่

ฐานข้อมูลชุดที่ 1 : ได้มาจากฐานข้อมูลของนายศวิต กาศุริยะ (ศวิต กาศุริยะ, 2542) ซึ่งเป็นฐานข้อมูล

ที่ประกอบด้วยจำนวนผู้พูด 12 คน แต่ละคนจะได้รับการบันทึกเสียงคนละ 10 ครั้ง โดยคำพูดที่ใช้ คือ “สาม ห้า สอง เก้า สี่” โดยจะใช้ข้อมูลกลุ่มนี้สำหรับการทดลองเปรียบเทียบระบบบ่งชี้ผู้พูดแบบระบบปิด ระหว่างระบบที่ใช้ชุดรหัสเดียว กับระบบที่ใช้หลายชุดรหัส

ฐานข้อมูลชุดที่ 2 : เป็นกลุ่มฐานข้อมูลหลักที่จะนำมาใช้ในการทดลอง โดยบันทึกเสียงจากผู้พูดทั้ง

หมด 27 คนแบ่งเป็นชาย 18 คน หญิง 9 คน ช่วงอายุของผู้พูดจะอยู่ในช่วง 22-26 ปี โดยจะบันทึก 2 ช่วงการบันทึก แต่ละช่วงจะเว้นประมาณ 1 เดือน และจะมีผู้พูดบางคน (ประมาณ 10 คน) ที่บันทึกแค่ช่วงเดียว เนื่องจากจะนำเสียงของผู้พูดเหล่านี้ไปจัดอยู่ในกลุ่มผู้พูดนอกระบบ โดยในแต่ละช่วงผู้พูดแต่ละคนจะได้รับการบันทึกเสียง 10 ครั้ง บทคำพูดที่ใช้ คือ “สาม ห้า สอง เก้า สี่” แสดงรายละเอียดของฐานข้อมูลชุดนี้ดังตารางที่ 4-1

ในการทดลองได้แบ่งกลุ่มของตัวอย่างเสียงพูดออกเป็น 3 กลุ่มด้วยกัน คือ

1. **กลุ่มฝึกฝน** กลุ่มนี้เป็นกลุ่มที่จะนำมาใช้ฝึกฝนระบบ โดยเป็นกลุ่มของผู้พูดในระบบในที่นี้ได้กำหนดผู้พูดไว้ 10 คน แบ่งเป็นชาย 6 คน หญิง 4 คน สำหรับเสียงที่เลือกนี้จะสุ่มเลือกเสียงที่ได้รับการบันทึกทั้งสองช่วง โดยได้เลือกเสียงของ Dat, Iti, Kan, Png, Prn, Saw, Sdr, Sit, Swt, Uma เป็นเสียงของผู้พูดในระบบ สำหรับเสียงที่นำมาใช้ในการฝึกฝนนั้น จะนำเสียงผู้พูดคนละ 5 เสียงในช่วงที่ 1 มาเป็นกลุ่มฝึกฝน ดังนั้นจะมีจำนวนเสียงในกลุ่มนี้ $5 \times 10 = 50$ เสียง
2. **กลุ่มทดสอบที่ 1** กลุ่มนี้เป็นกลุ่มที่นำมาทดสอบเพื่อจะนำมาหาค่าผ่านระดับ สำหรับเสียงตัวอย่างในกลุ่มนี้ กลุ่มของผู้พูดในระบบได้มาจาก 5 เสียงที่เหลือในช่วงที่ 1 และกลุ่มของผู้พูดนอกระบบได้มาจากเสียงพูด 5 เสียงในช่วงที่ 1 ดังนั้นจะมีจำนวนเสียงพูดในกลุ่มนี้ แบ่งเป็นเสียงพูดของผู้พูดในระบบ $5 \times 10 = 50$ เสียง และเสียงของผู้พูดนอกระบบ $5 \times 17 = 85$ เสียง
3. **กลุ่มทดสอบที่ 2** กลุ่มนี้เป็นกลุ่มที่ใช้ทดสอบระบบหลังจากที่ได้ค่าผ่านระดับจากการทดสอบ โดยกลุ่มที่ 1 สำหรับเสียงตัวอย่างในกลุ่มนี้ กลุ่มของผู้พูดในระบบได้มาจากเสียงพูดทั้งหมดในช่วงที่ 2 และกลุ่มของผู้พูดนอกระบบได้มาจากเสียงพูดในช่วงที่ 2 ทั้งหมดและ 5 เสียงที่เหลือในช่วงที่ 1 ดังนั้นจากตารางที่ 4-1 จะได้ว่าจำนวนเสียงในกลุ่มนี้ แบ่งเป็นเสียงพูดของผู้พูดในระบบ 105 เสียง และเสียงพูดของผู้พูดนอกระบบ 262 เสียง

ฐานข้อมูลชุดที่ 3 : เป็นฐานข้อมูลที่ได้มาจากศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ

(NECTEC) โดยเป็นฐานข้อมูลของตัวเลขโดด 0-9 ในการนำมาใช้ในการทดลอง ได้นำเสียงตัวเลขเดียวกันมาต่อกัน โดยนำเสียง “สาม” “ห้า” “สอง” “เก้า” และ “สี่” มาต่อกัน ตามลำดับ ในฐานข้อมูลนี้ประกอบด้วยเสียงผู้ชาย 28 คนและเสียงผู้หญิง 13 คน โดยได้บันทึกทั้งหมด 8 ช่วงการบันทึก สำหรับฐานข้อมูลนี้จะนำไปใช้ในการทดลองผลของจำนวนผู้พูดในระบบต่อประสิทธิภาพของระบบ คังจะได้นำเสนอต่อไป

สำหรับจำนวนเสียงแต่ละคนในแต่ละช่วงได้นำมาแสดงดังตารางที่ 4-1

ตารางที่ 4-1 รายละเอียดของเสียงพูดที่ใช้ในการทดลอง สำหรับฐานข้อมูลชุดที่ 2

ลำดับที่	ผู้พูด	จำนวนเสียง	
		ช่วงที่ 1	ช่วงที่ 2
1	Ama	10	-
2	Arm	10	11
3	Bnc	10	11
4	*Dat	10	11
5	Eka	10	-
6	*Iti	10	10
7	Jnw	10	19
8	*Kan	10	11
9	Nav	10	-
10	Nrn	10	-
11	Pad	10	-
12	*Png	10	10
13	*Prn	10	12
14	*Saw	10	10
15	*Sdr	10	10
16	Sir	10	-
17	*Sit	10	10
18	Skm	10	-
19	Spt	10	10
20	Srp	10	-
21	*Swt	10	10
22	Thn	10	-
23	*Uma	10	10
24	Vrv	10	10
25	Wna	10	-
26	Wrk	10	10
27	Wsn	10	-

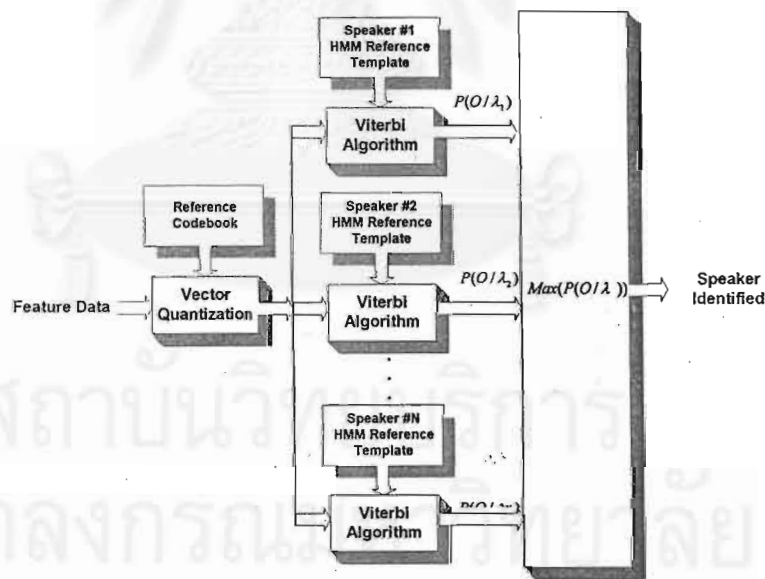
* เป็นเสียงของผู้พูดที่ถูกเลือกเป็นเสียงผู้พูดในระบบ

การทดลองที่ 4-1 การทดลองเปรียบเทียบผลอัตราการบ่งชี้ผู้พูดในระบบบ่งชี้ผู้พูดแบบระบบปิด ระหว่างระบบที่ใช้ชุดรหัสแบบชุดรหัสเดี่ยว กับระบบที่ใช้ชุดรหัสแบบหลายชุดรหัส โดยใช้ฐานข้อมูลชุดที่ 1

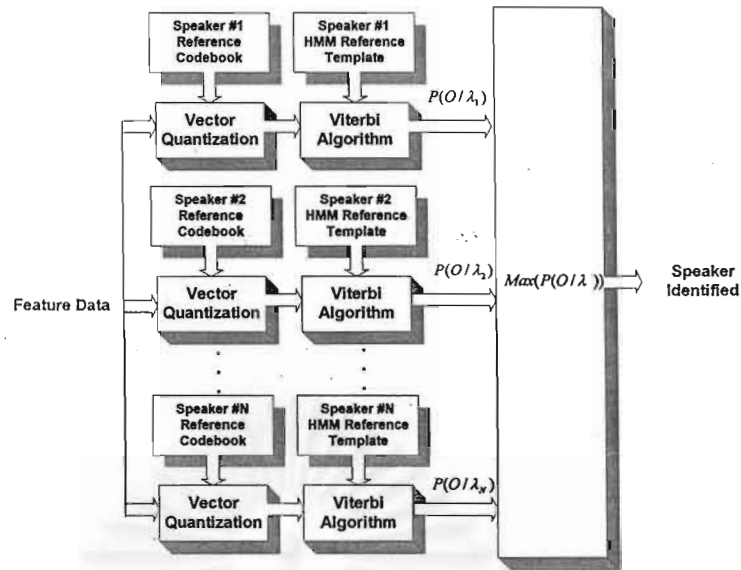
การทดลองนี้เป็นการเปรียบเทียบระหว่างระบบที่ใช้ชุดรหัสเดี่ยวกับระบบที่ใช้ชุดรหัสแบบหลายชุดระบบบ่งชี้ผู้พูดแบบระบบปิด โดยจะศึกษาว่าระบบที่ใช้ชุดรหัสแบบหลายชุดรหัสซึ่งเป็นระบบที่น่าเสอในงานวิทยานิพนธ์ให้ผลเปรียบเทียบกับระบบที่ใช้ชุดรหัสเดี่ยว ซึ่งเป็นงานวิจัยที่ได้นำเสนอโดยนาย สวิต กาศุริยะ เป็นอย่างไรบ้าง

สำหรับขั้นตอนในการทดสอบระบบนั้นแสดงได้ดังแผนภาพรูปที่ 4-1 และ รูปที่ 4-2 ตามลำดับ

ในการทดลองได้ใช้ลักษณะสำคัญ 3 ชนิด ได้แก่ LPC, CEP, และ MFCC ใช้ขนาดชุดรหัส 64, 128, และ 256 (ถ้าขนาดของชุดรหัสน้อยไปจะทำให้ได้ผลที่ไม่ดี แต่ถ้าขนาดของชุดรหัสมากไปจะทำให้ใช้เวลาในการประมวลผลที่นาน) ใช้จำนวนสถานะของแบบจำลองฮิดเคนมาร์คอฟ 10, 25, และ 40 สถานะ (ในงานวิทยานิพนธ์ของนายสวิต กาศุริยะ ได้ใช้จำนวนสถานะ 5, 10, 15, 20, 25, และ 30 ซึ่งปรากฏว่า จำนวนสถานะที่ใกล้เคียงกัน ให้ผลที่ใกล้เคียงกัน ดังนั้นในงานวิทยานิพนธ์นี้จึงใช้ช่วงห่างของจำนวนสถานะที่มากขึ้น)



รูปที่ 4-1 ขั้นตอนการทดสอบระบบ สำหรับระบบที่ใช้ชุดรหัสเดี่ยว



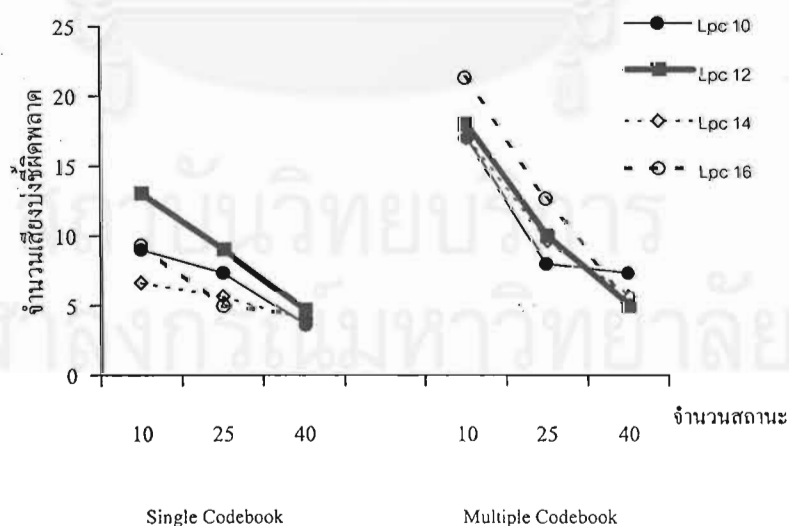
รูปที่ 4-2 ขั้นตอนการทดสอบระบบ สำหรับระบบที่ใช้ชุดรหัสแบบหลายชุด

สำหรับในทดลองได้ฝึกฝนแบบจำลองของผู้พูดแต่ละคนจากทั้งหมด 12 คน โดยนำเสียงจำนวน 5 เสียงสำหรับผู้พูดแต่ละคน มาทำการฝึกฝนแบบจำลอง และนำเสียงจำนวน 5 เสียงสำหรับผู้พูดแต่ละคน มาทำการทดสอบระบบ ดังนั้นจะมีเสียงทดสอบจำนวน 60 เสียง

ได้ผลการทดลองดังแสดงในตารางที่ 4-2, ตารางที่ 4-3, และ ตารางที่ 4-4

ตารางที่ 4-2 ผลการทดลอง จำนวนเสียงที่บ่งชี้ผิดพลาด เมื่อใช้ชุดรหัสแบบ Single Codebook กับชุดรหัสแบบ Multiple Codebook เมื่อใช้ LPC เป็นค่าลักษณะสำคัญ¹

Feature	Codebook Size	Single Codebook			Multiple Codebook		
		10 States	25 States	40 States	10 States	25 States	40 States
LPC-10	64	14	12	6	24	5	8
	128	8	7	3	14	9	9
	256	5	3	2	13	10	5
	AVG.	9.00	7.33	3.67	17.00	8.00	7.33
LPC-12	64	15	16	9	25	7	7
	128	14	5	4	15	13	4
	256	10	6	1	14	10	4
	AVG.	13.00	9.00	4.67	18.00	10.00	5.00
LPC-14	64	9	8	6	25	14	13
	128	8	6	4	18	10	2
	256	3	3	2	8	5	2
	AVG.	6.67	5.67	4.00	17.00	9.67	5.67
LPC-16	64	13	8	6	35	20	7
	128	11	5	7	23	12	5
	256	4	2	0	6	6	3
	AVG.	9.33	5.00	4.33	21.33	12.67	5.00

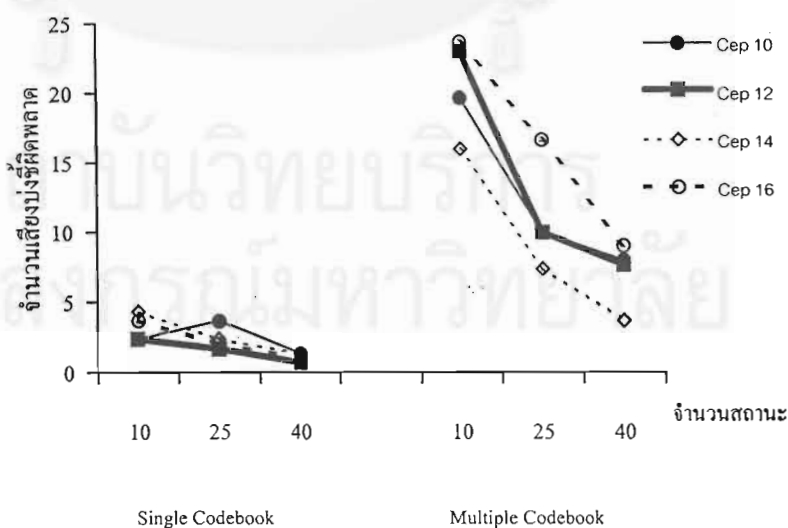


รูปที่ 4-3 ค่าเฉลี่ยจำนวนเสียงที่บ่งชี้ผิดพลาดเมื่อใช้ LPC เป็นค่าลักษณะสำคัญ

¹ จำนวนเสียงทดสอบ 60 เสียง

ตารางที่ 4-3 ผลการทดลอง จำนวนเสียงที่บ่งชี้ผิดพลาด เมื่อใช้ชุดรหัสแบบ Single Codebook กับชุดรหัสแบบ Multiple Codebook เมื่อใช้ CEP เป็นค่าลักษณะสำคัญ²

Feature	Codebook Size	Single Codebook			Multiple Codebook		
		10 States	25 States	40 States	10 States	25 States	40 States
CEP-10	64	4	4	2	26	10	13
	128	3	6	0	14	10	4
	256	0	1	2	19	10	7
	AVG.	2.33	3.67	1.33	19.67	10.00	8.00
CEP-12	64	3	3	1	37	15	10
	128	3	0	0	16	9	6
	256	1	2	1	16	6	7
	AVG.	2.33	1.67	0.67	23.00	10.00	7.67
CEP-14	64	4	5	2	22	11	8
	128	6	1	1	16	5	1
	256	3	1	1	10	6	2
	AVG.	4.33	2.33	1.33	16.00	7.33	3.67
CEP-16	64	6	2	0	32	20	15
	128	5	3	3	19	14	8
	256	0	1	0	20	16	4
	AVG.	3.67	2.00	1.00	23.67	16.67	9.00

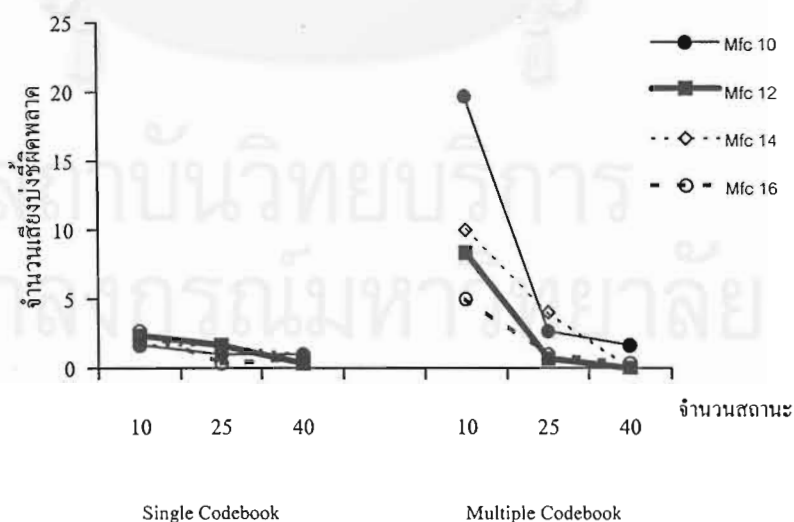


รูปที่ 4-4 ค่าเฉลี่ยจำนวนเสียงที่บ่งชี้ผิดพลาดเมื่อใช้ CEP เป็นค่าลักษณะสำคัญ

² จำนวนเสียงทดสอบ 60 เสียง

ตารางที่ 4-4 ผลการทดลอง จำนวนเสียงที่บ่งชี้ผิดพลาด เมื่อใช้ชุดรหัสแบบ Single Codebook กับชุดรหัสแบบ Multiple Codebook เมื่อใช้ MFCC เป็นค่าลักษณะสำคัญ³

Feature	Codebook Size	Single Codebook			Multiple Codebook		
		10 States	25 States	40 States	10 States	25 States	40 States
MFCC-10	64	3	1	1	6	3	0
	128	1	1	1	9	2	3
	256	1	1	1	44	3	2
	AVG.	1.67	1.00	1.00	19.67	2.67	1.67
MFCC -12	64	6	3	0	14	1	0
	128	1	2	1	5	1	0
	256	0	0	0	6	0	0
	AVG.	2.33	1.67	0.33	8.33	0.67	0.00
MFCC -14	64	4	3	1	15	6	0
	128	1	2	2	7	4	0
	256	0	0	0	8	2	0
	AVG.	1.67	1.67	1.00	10.00	4.00	0.00
MFCC -16	64	8	1	2	8	1	1
	128	0	0	0	4	1	0
	256	0	0	0	3	1	0
	AVG.	2.67	0.33	0.67	5.00	1.00	0.33



รูปที่ 4-5 ค่าเฉลี่ยจำนวนเสียงที่บ่งชี้ผิดพลาดเมื่อใช้ MFCC เป็นค่าลักษณะสำคัญ

³ จำนวนเสียงทดสอบ 60 เสียง

วิเคราะห์และอภิปรายผลการทดลองที่ 4-1

1. จากกราฟผลการทดลองจะเห็นได้ว่าอัตราการบ่งชี้ผิดพลาดของระบบที่ใช้ซุ่มรหัสแบบซุ่มรหัสเดียวมีค่าโดยรวมต่ำกว่าระบบที่ใช้ซุ่มรหัสแบบหลายซุ่มรหัส โดยให้ผลสอดคล้องกันทั้งสามค่าคุณลักษณะสำคัญ สาเหตุที่เป็นเช่นนี้สามารถอธิบายได้ว่า ในการฝึกฝนซุ่มรหัสแบบซุ่มรหัสเดียวนั้นมาจากการฝึกฝนจากเสียงผู้พูดหลายคนทำให้ซุ่มรหัสมีการกระจายของข้อมูลที่มากกว่าซุ่มรหัสแบบหลายซุ่ม ดังนั้นเมื่อนำเสียงทดสอบไปเทียบกับซุ่มรหัสเพื่อให้ได้ลำดับค่าสังเกต ซุ่มรหัสเดียวจะมีค่าที่ครอบคลุมเสียงทดสอบนั้นได้มากกว่าจึงทำให้การเทียบที่ได้ถูกต้องมากกว่าซุ่มรหัสแบบหลายซุ่ม และอีกกรณีหนึ่งก็คือความหลากหลายของซุ่มรหัส การเทียบซุ่มรหัสของเสียงทดสอบไม่ได้มีบรรทัดฐานเดียวกันทำให้โอกาสเกิดความผิดพลาดมีมากกว่า
2. จากการทดลองนี้ได้เกิดข้อสังเกตหนึ่งขึ้นมา คือ ในการเทียบซุ่มรหัสนั้นจะทำให้เกิดค่าผิดพลาดขึ้น หรือเรียกว่า ค่าผิดพลาดจากการควอนไทซ์ (Quantization Error) ในกรณีของซุ่มรหัสแบบหลายซุ่ม โดยซุ่มรหัสที่ได้มาจากการฝึกฝนเสียงพูดของผู้พูดแต่ละคน ดังนั้นเมื่อนำเสียงทดสอบมาเทียบกับซุ่มรหัส เสียงทดสอบที่เป็นผู้พูดคนเดียวกับผู้พูดที่นำมาฝึกฝนซุ่มรหัส จะให้ค่าผิดพลาดจากการควอนไทซ์ที่ต่ำกว่าเมื่อนำเสียงทดสอบนี้ไปเทียบกับซุ่มรหัสของผู้พูดอื่น ดังนั้นในการทดลองต่อไปจะมีการนำเอาผลของค่าผิดพลาดจากการควอนไทซ์ ซึ่งต่อไปจะเรียกค่านี้ว่าค่าความเพี้ยน (Distortion) มาใช้ในระบบบ่งชี้ผู้พูดสำหรับระบบที่ใช้ซุ่มรหัสแบบหลายซุ่ม

จากผลการทดลองที่ได้ในการทดลองที่ 4-1 นั้น จะเห็นว่าระบบที่ใช้ซุ่มรหัสแบบหลายซุ่มรหัสให้ผลที่ไม่ดีเท่าระบบที่ใช้ซุ่มรหัสเดียว แต่จากการนำซุ่มรหัสแบบหลายซุ่มรหัสมาใช้นั้น จะมีความแตกต่างของซุ่มรหัส ดังนั้นในการทดลองในการทดลองต่อไปนั้น จะศึกษาถึงผลของความแตกต่างนี้ โดยจะศึกษาในรูปของค่าความเพี้ยน ซึ่งเป็นค่าที่ได้ในขั้นตอนการควอนไทซ์แบบเวกเตอร์ ว่าจะให้ผลเป็นอย่างไร

การทดลองที่ 4-2 การนำค่าความเพี้ยน (Distortion) จากกระบวนการควอนไทซ์แบบเวกเตอร์มาคิดรวมกับค่าความน่าจะเป็นที่ได้จากแบบจำลองฮิดเดนมาร์คอฟ ในระบบที่ใช้ซุ่มรหัสแบบหลายซุ่ม

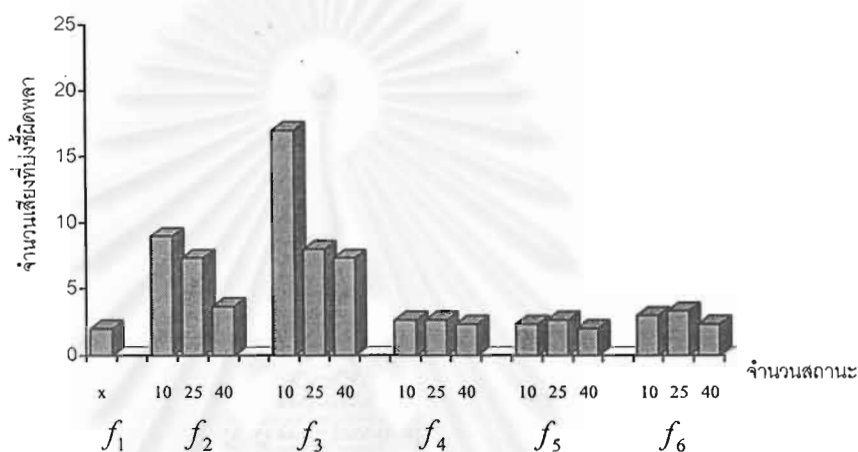
จากที่ได้อภิปรายมาแล้วในการทดลองที่ 4-1 ถึงผลของความแตกต่างของซุ่มรหัสในระบบที่ใช้ซุ่มรหัสแบบหลายซุ่มรหัส ในการทดลองนี้ก็จะศึกษาถึงผลของความแตกต่างดังกล่าวโดยจะศึกษาในรูปของค่าความเพี้ยน แล้วนำค่าความเพี้ยนที่ได้นี้มาพิจารณาพร้อมกับค่าความน่าจะเป็น โดยรายละเอียดของวิธีการนำค่าความเพี้ยนมาใช้ร่วมกับค่าความน่าจะเป็นแสดงในส่วนท้ายของบทที่ 2

ในการทดลองนี้ได้ใช้ LPC อันดับ 10 เป็นค่าลักษณะสำคัญ โดยใช้ขนาดของซุ่มรหัสเป็น 64, 128, และ 256 สำหรับจำนวนสถานะของแบบจำลองได้ใช้ค่า 10, 25, และ 40 ฐานข้อมูลที่ใช้เป็นฐานข้อมูลชุดเดียวกันกับการทดลองที่ 1

ได้ผลการทดลองดังแสดงในตารางที่ 4-5

ตารางที่ 4-5 จำนวนเสียงที่บ่งชี้ผิดพลาดเมื่อใช้ค่าฟังก์ชัน $(f_1 - f_6)^4$

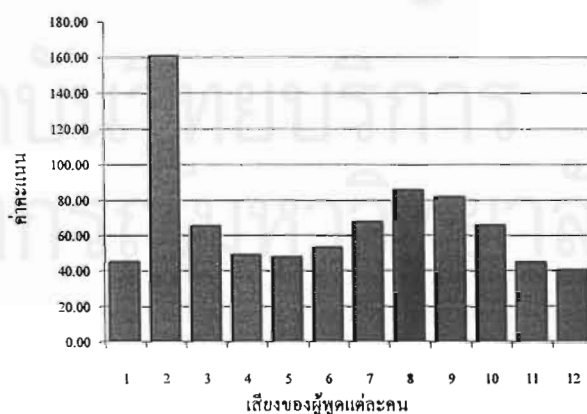
Pulse	CB Size	f_1	f_2			f_3			f_4			f_5			f_6		
			10	25	40	10	25	40	10	25	40	10	25	40	10	25	40
LPC-10	64	2	14	12	6	24	5	8	2	2	3	2	2	2	2	3	3
	128	2	8	7	3	14	9	9	2	2	2	2	2	2	2	3	2
	256	2	5	3	2	13	10	5	4	4	2	3	4	2	5	4	2
	AVG.	2	9.00	7.33	3.67	17	8.00	7.33	2.67	2.67	2.33	2.33	2.67	2.00	3.00	3.33	2.33



รูปที่ 4-6 ค่าเฉลี่ยจำนวนเสียงที่บ่งชี้ผิดพลาดจากตารางที่ 4-5

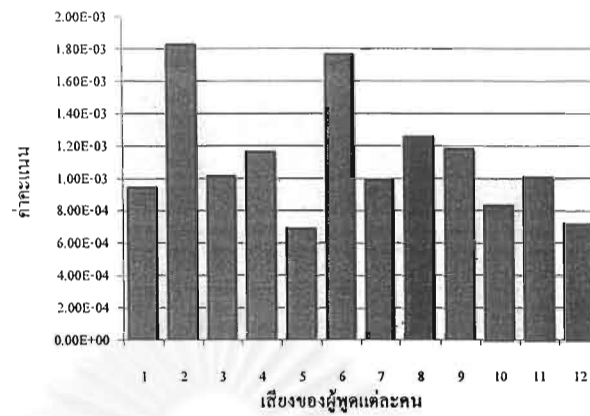
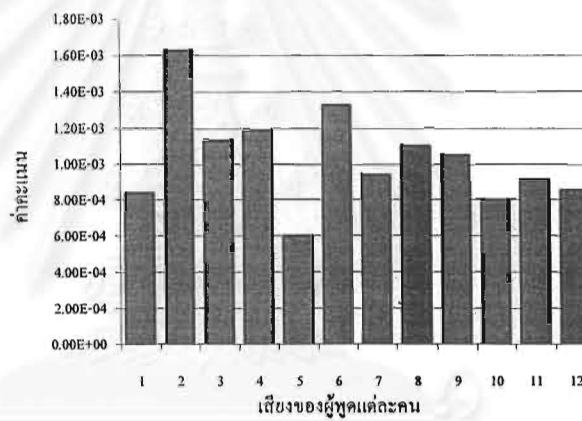
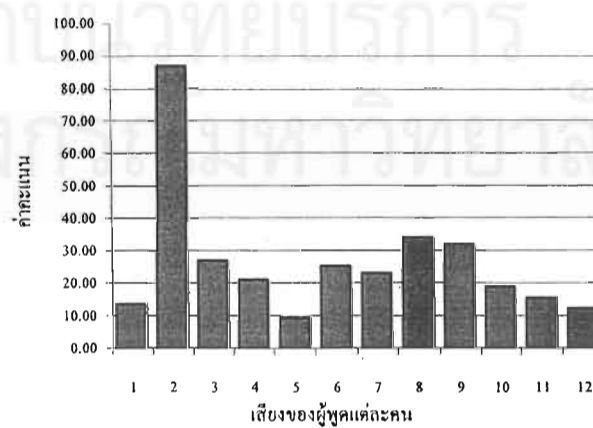
จากตารางผลการทดลองจะเห็นว่า f_1 และ $f_4 - f_6$ ซึ่งเป็นฟังก์ชันที่นำค่าความเพี้ยนมาใช้ ให้ผลดีกว่า f_2 และ f_3 ตามลำดับ

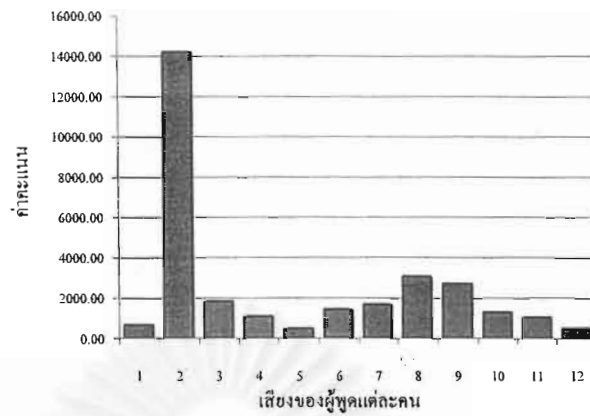
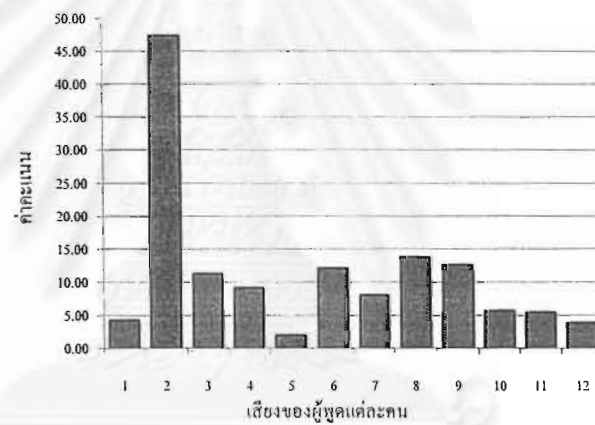
เมื่อลองพิจารณาค่าคะแนนที่ได้ในแต่ละฟังก์ชัน โดยนำมาแสดงได้ดังรูปที่ 4-7



(ก). ฟังก์ชัน f_1

⁴ จำนวนเสียงทดสอบ 60 เสียง

(ข). ฟังก์ชัน f_2 (ค). ฟังก์ชัน f_3 (ง). ฟังก์ชัน f_4

(จ). ฟังก์ชัน f_5 (ฉ). ฟังก์ชัน f_6

รูปที่ 4-7 ค่าคะแนนที่ได้จากการนำเสียงทดสอบผ่านแบบจำลองของผู้พูดคนที่ 2 โดยใช้ฟังก์ชัน $f_1 - f_6$

วิเคราะห์และอภิปรายผลการทดลองที่ 4-2

1. จากผลการทดลองครั้งที่แสดงในตารางที่ 4-5 และในรูปที่ 4-7 จะเห็นได้ว่าเมื่อนำค่าความเพี้ยนมาใช้จะทำให้ได้อัตราการบ่งชี้ที่ดีขึ้นค่อนข้างมาก เมื่อเทียบกับที่ยังไม่ได้ใช้ ซึ่งจากผลที่ได้นี้แสดงว่าการนำชุดรหัสแบบหลายชุดมาใช้มีข้อดีก็คือความหลากหลาย ไม่ได้เป็นบรรทัดฐานเดียวกันก็จริง แต่ข้อเด่นก็คือผลที่ได้จากการนำค่าความเพี้ยนมาใช้ นั่นคือการฝึกฝนชุดรหัสจากเสียงของผู้พูดแต่ละคน ชุดรหัสที่ได้จะมีค่ากระจายเฉพาะค่าที่ได้จากเสียงผู้พูดของคนๆ นั้น ดังนั้นเมื่อนำเสียงมาเทียบกับชุดรหัส เสียงพูดของผู้พูดที่ตรงกับชุดรหัสของตัวเองจะให้ค่าความเพี้ยนที่ต่ำกว่า เมื่อนำไปเทียบกับชุดรหัสของผู้พูดคนอื่น
2. จากรูปที่ 4-7 จะเห็นว่าผลของการใช้ฟังก์ชันค่าคะแนนแต่ละฟังก์ชัน มีความสามารถในการแยกแยะ (ความแตกต่างของค่าคะแนน) ระหว่างเสียงของผู้พูดคนเดียวกับแบบจำลองออกจากเสียงผู้พูดคนอื่นได้ดีไม่เท่ากัน ซึ่งความสามารถในการแยกแยะนี้จะมีประโยชน์ในขั้นของการตรวจสอบในระบบบ่งชี้ผู้พูดแบบระบบเปิด นั่นคือยังสามารถแยกแยะได้ดีก็จะเป็นประโยชน์ต่อขั้นตอนการตรวจสอบ โดยจากรูปจะเห็นว่าฟังก์ชัน $f_4 - f_6$ ต่างก็แยกแยะความแตกต่างได้อย่างชัดเจนเมื่อเทียบกับ $f_1 - f_3$ จากผลที่ได้นี้ทำให้เห็นได้ว่าการนำค่าความเพี้ยนมาใช้ร่วมกับค่าความน่าจะเป็นของแบบจำลองทำให้ได้ผลที่ดีขึ้น อาจมองได้ว่าแบบจำลองฮิดเดนมาร์คอฟนั้นจะให้ความสำคัญกับลำดับของค่าสังเกต หรือถ้าจะเปรียบเทียบก็คือจังหวะการพูดของผู้พูดนั่นเอง ส่วนค่าความเพี้ยนนั้นมาจากการเทียบชุดรหัส ซึ่งเป็นการกระจายของเสียงพูดของผู้พูด โดยการกระจายจะขึ้นอยู่กับลักษณะทางกายภาพของผู้พูด เช่น ระดับเสียง ความทึบแหลมของเสียง เป็นต้น ดังนั้นการนำทั้งสองค่ามาพิจารณาพร้อมกันจึงทำให้ได้ผลที่ดีขึ้น

จากผลการทดลองนี้สรุปได้ว่าการนำเอาค่าความเพี้ยนมาใช้ร่วมกับค่าความน่าจะเป็นในระบบที่ใช้ชุดรหัสแบบหลายชุดรหัส สามารถช่วยให้ระบบมีอัตราการบ่งชี้ที่ดีขึ้นและดีกว่าระบบที่ใช้ชุดรหัสแบบชุดรหัสเดียว

การทดลองที่ 4-3 ผลของค่าพารามิเตอร์ของระบบบ่งชี้ผู้พูดต่ออัตราการบ่งชี้ผู้พูด

จากการทดลองที่ 4-2 ได้ทดลองนำเอาค่าความเพี้ยนมาใช้ร่วมกับค่าความน่าจะเป็นสามารถช่วยให้อัตราการบ่งชี้ผิดพลาดของระบบบ่งชี้ผู้พูดที่ใช้ชุดรหัสแบบหลายชุดรหัสดีขึ้น จากนั้นในการทดลองนี้ จึงได้นำเอาวิธีการดังกล่าวมาใช้ และจะได้ศึกษาถึงผลของค่าพารามิเตอร์ต่างๆ ของระบบ ซึ่งได้แก่ อันดับของค่าลักษณะสำคัญ ขนาดของชุดรหัส และจำนวนสถานะของแบบจำลองฮิดเดนมาร์คอฟ เพื่อจะได้นำผลที่ได้จากการทดลองนี้ไปออกแบบระบบบ่งชี้ผู้พูดที่มีประสิทธิภาพต่อไป

ในการทดลองนี้ได้แบ่งการทดลองออกเป็น 3 การทดลอง โดยการทดลองแรกเป็นการทดลองถึงผลของค่าอันดับของค่าลักษณะสำคัญ การทดลองต่อมาเป็นการทดลองถึงผลของขนาดของชุดรหัส

และการทดลองสุดท้ายเป็นการทดลองถึงผลของจำนวนสถานะ โดยในแต่ละการทดลองก็จะแบ่งเป็นการทดลองย่อยอีกสองการทดลอง คือ การทดลองในระบบปิด กับ การทดลองในระบบเปิด

รายละเอียดและผลการทดลองแสดงดังการทดลองดังต่อไปนี้ ตามลำดับ

การทดลองที่ 4-3-1 ผลของอันดับของค่าลักษณะสำคัญต่ออัตราการบ่งชี้ผู้พูด

การทดลองนี้จะพิจารณาผลของอันดับของค่าลักษณะสำคัญต่อระบบบ่งชี้ผู้พูด โดยฐานข้อมูลที่ใช้ในการทดลองนี้เป็นฐานข้อมูลชุดที่ 2 สำหรับระบบบ่งชี้ผู้พูดนี้ได้ทดลองทั้งสองระบบ คือระบบปิดกับระบบเปิด ในการทดลองนี้ได้ใช้ขนาดของชุดรหัสเท่ากับ 64, 128, และ 256 จำนวนสถานะของแบบจำลองเท่ากับ 10, 25, 40 และใช้ t_1 , t_4 , t_5 (สำหรับ t_6 ที่ไม่นำเสนออีกเนื่องจากให้ผลที่ไม่ดี) เป็นฟังก์ชันวัดค่าคะแนน ผลการทดลองแสดงได้ดังการทดลองย่อย ดังนี้

การทดลองที่ 4-3-1-1 ระบบบ่งชี้ผู้พูดแบบระบบปิด

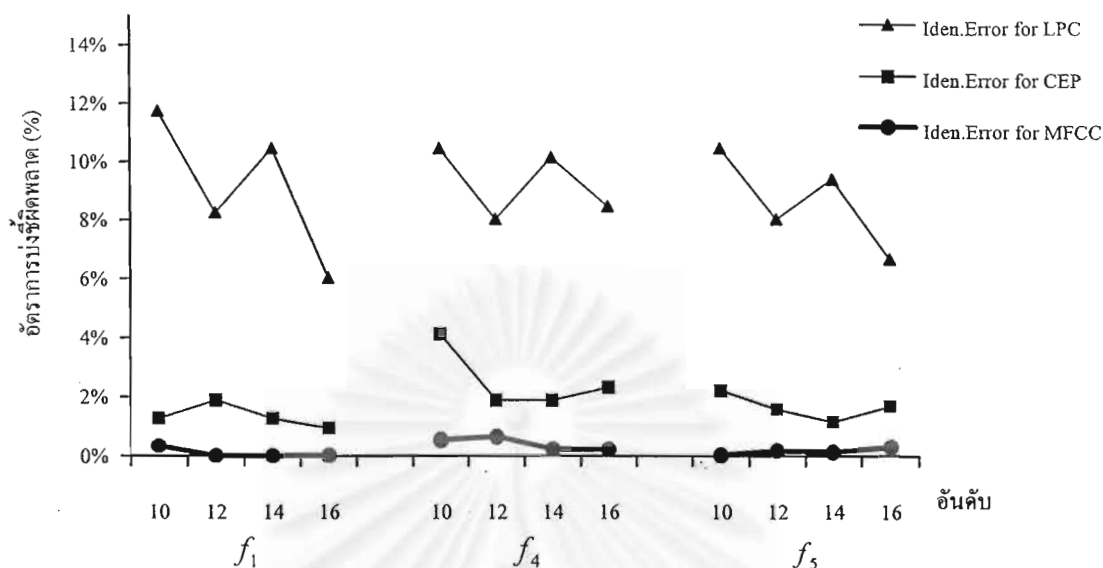
สำหรับระบบปิดที่สร้างขึ้นนี้เป็นระบบที่ประกอบด้วยผู้พูดในระบบ 10 คน (ผู้พูดที่เลือกแสดงในตารางที่ 4-1) แสดงอัตราการบ่งชี้ผิดพลาดเฉลี่ย สำหรับชุดทดสอบที่ 2 (เฉลี่ยทุกขนาดของชุดรหัส และทุกจำนวนสถานะ) ดังตารางที่ 4-6 (ผลของการทดลองในรายละเอียด แสดงในภาคผนวก)

ตารางที่ 4-6 อัตราการบ่งชี้ผิดพลาดเฉลี่ย (%) ในแต่ละค่าอันดับของลักษณะสำคัญต่างๆ⁵

Feature	t_1				t_4				t_5			
	10	12	14	16	10	12	14	16	10	12	14	16
LPC	11.75	8.25	10.48	6.03	10.48	8.04	10.16	8.47	10.48	8.04	9.42	6.67
CEP	1.27	1.9	1.27	0.95	4.13	1.9	1.9	2.33	2.22	1.59	1.16	1.69
MFCC	0.32	0	0	0	0.53	0.63	0.21	0.21	0	0.16	0.16	0.32

⁵ จำนวนเสียงทดสอบ 105 เสียง

นำผลจากตารางมาแสดงเป็นกราฟได้ดังรูปที่ 4-8



รูปที่ 4-8 อัตราการบ่งชี้ผิดพลาดเฉลี่ย ที่อันดับของลักษณะสำคัญค่าต่างๆ

การทดลองที่ 4-3-1-2 ระบบบ่งชี้ผู้พูดแบบระบบเปิด

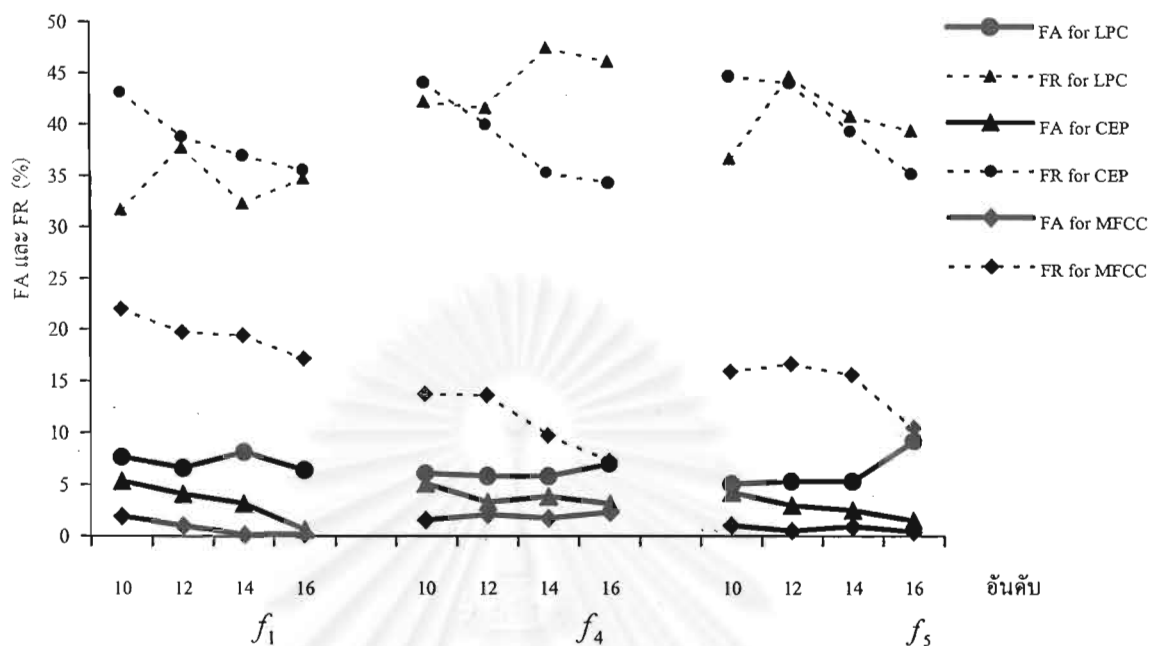
สำหรับระบบบ่งชี้ผู้พูดแบบเปิดในการทดลองนี้ ประกอบด้วยจำนวนผู้พูดในระบบ 10 คน และจำนวนผู้พูดนอกระบบ 17 คน (ดังแสดงในตารางที่ 4-1) โดยแสดงค่าอัตราการยอมรับผิดพลาดและอัตราการใช้ผิดพลาดเฉลี่ย ดังตารางที่ 4-7

ตารางที่ 4-7 FA (%) กับ FR (%) เฉลี่ย ที่อันดับของค่าลักษณะสำคัญค่าต่างๆ สำหรับกลุ่มทดสอบที่ 2⁶

Feature		f_1				f_4				f_5			
		10	12	14	16	10	12	14	16	10	12	14	16
LPC	FA	7.64	6.58	8.07	6.37	6.02	5.8	5.73	6.94	5.02	5.31	5.31	9.13
	FR	31.68	37.71	32.28	34.79	42.28	41.66	47.52	46.17	36.74	44.73	40.87	39.46
CEP	FA	5.31	4.03	3.18	0.64	5.10	3.26	3.82	3.11	4.18	2.97	2.48	1.49
	FR	43.12	38.83	36.97	35.58	44.12	40.04	35.36	34.34	44.78	44.08	39.43	35.21
MFC	FA	1.91	1.06	0.21	0.21	1.56	2.05	1.7	2.26	1.06	0.5	0.92	0.42
	FR	21.97	19.68	19.37	17.14	13.72	13.63	9.76	7.31	15.89	16.65	15.57	10.39

⁶ จำนวนเสียงทดสอบ 262 เสียง (จำนวนเสียงพูดในระบบ 105 และ นอกระบบ 157 เสียง)

นำผลจากตารางมาแสดงได้ดังรูปที่ 4-9



รูปที่ 4-9 อัตราการยอมรับและปฏิเสธผิดพลาดเฉลี่ย ที่อันดับของลักษณะสำคัญค่าต่างๆ

วิเคราะห์และอภิปรายผลการทดลองที่ 4-3-1

1. ระบบปิด

จากรูปที่ 4-8 จะเห็นว่าอัตราการบ่งชี้ผิดพลาดมีแนวโน้มลดลงเมื่อเพิ่มอันดับของ LPC แต่ที่อันดับของ LPC เท่ากับ 14 จะมีค่าสูงนั้นแสดงว่า LPC อันดับที่ 14 ให้ผลที่ไม่ค่อยดี เมื่อเทียบกับ LPC อันดับที่ 12 แต่ก็ยังให้ผลโดยรวมดีกว่า LPC อันดับที่ 10 สำหรับ CEP อัตราการบ่งชี้ผิดพลาดมีแนวโน้มลดลงเพียงเล็กน้อย แสดงว่าที่อันดับของ CEP ที่ได้ทำการทดลองเป็นตำแหน่งที่เริ่มอิ่มตัวแล้ว ดังนั้นการเพิ่มอันดับของ CEP ไม่ค่อยมีผลต่ออัตราการบ่งชี้เท่าไรนัก ส่วน MFCC ก็ให้ผลในทำนองเดียวกัน

ส่วนอัตราการบ่งชี้ผิดพลาดเฉลี่ยเมื่อเปรียบระหว่างค่าลักษณะสำคัญจะเห็นได้ว่า MFCC ให้อัตราการบ่งชี้ที่ดีที่สุด และ CEP เป็นอันดับรองลงมา ส่วน LPC ให้ผลที่ด้อยที่สุด

2. ระบบเปิด

จากรูปที่ 4-9 จะเห็นว่า ที่อันดับของ LPC ค่าต่างๆ แนวโน้มของการเปลี่ยนแปลงค่า FA และ FR ไม่เด่นชัดเท่าไรนัก แสดงว่าเมื่อใช้ LPC เป็นลักษณะสำคัญ อันดับที่ใช้ในการทดลองให้ผลที่ไม่แตกต่างกันมากนัก แต่เมื่อใช้ CEP และ MFCC จากรูปจะเห็นว่าค่า FA และ FR มีแนวโน้มที่ลดลงเมื่อเพิ่มค่าอันดับของลักษณะสำคัญ นั้นแสดงว่าอันดับของค่าลักษณะสำคัญมีผลต่อค่า FA และ FR เมื่อใช้ลักษณะสำคัญ CEP และ MFCC

เมื่อพิจารณาค่า FA และ FR จะเห็นว่า ทุกค่าลักษณะสำคัญมีค่า FR มากกว่าค่า FA ที่เป็นเช่นนี้ก็เนื่องมาจากได้มีการตั้งค่าผ่านระดับที่สูงจากขั้นตอนการหาค่าผ่านระดับจากชุดทดสอบที่ 1

การทดลองที่ 4-3-2 ผลของขนาดของชุดรหัสต่ออัตราการบ่งชี้

ในการทดลองนี้จะเป็นการทดลองเพื่อศึกษาถึงผลของขนาดของชุดรหัสต่ออัตราการบ่งชี้ผู้พูด โดยจะได้ทดลองแปรค่าขนาดของชุดรหัส 64, 128, และ 256 ที่ค่าคุณลักษณะสำคัญต่างๆ และจำนวนสถานะที่ใช้เท่ากับ 10, 25, และ 40 โดยใช้ f_1 , f_4 , และ f_5 เป็นฟังก์ชันวัดค่าคะแนน

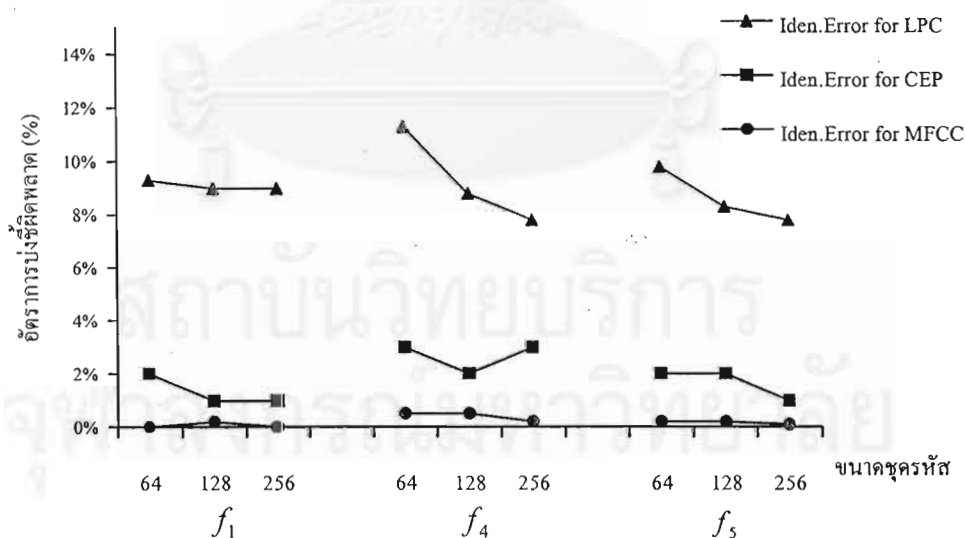
การทดลองที่ 4-3-2-1 ระบบบ่งชี้ผู้พูดแบบระบบปิด

ทำนองเดียวกันกับการทดลองที่ 3 คือใช้จำนวนผู้พูดในระบบ 10 คน ผลการทดลองแสดงในตารางที่ 4-8

ตารางที่ 4-8 อัตราการบ่งชี้ผิดพลาดเฉลี่ย (%) เมื่อใช้ขนาดของชุดรหัสค่าต่างๆ สำหรับกลุ่มทดสอบที่ 2⁷

Feature	f_1			f_4			f_5		
	64	128	256	64	128	256	64	128	256
LPC	9.29	9.05	9.05	11.27	8.81	7.78	9.84	8.33	7.78
CEP	1.67	0.95	1.43	2.78	2.22	2.70	2.06	1.59	1.35
MFCC	0.00	0.24	0.00	0.48	0.48	0.24	0.24	0.16	0.08

นำผลที่ได้มากแสดงได้ดังรูปที่ 4-10



รูปที่ 4-10 อัตราการบ่งชี้ผิดพลาดเฉลี่ย (%) ที่ขนาดของชุดรหัสค่าต่างๆ

⁷ จำนวนเสียงทดสอบ 105 เสียง

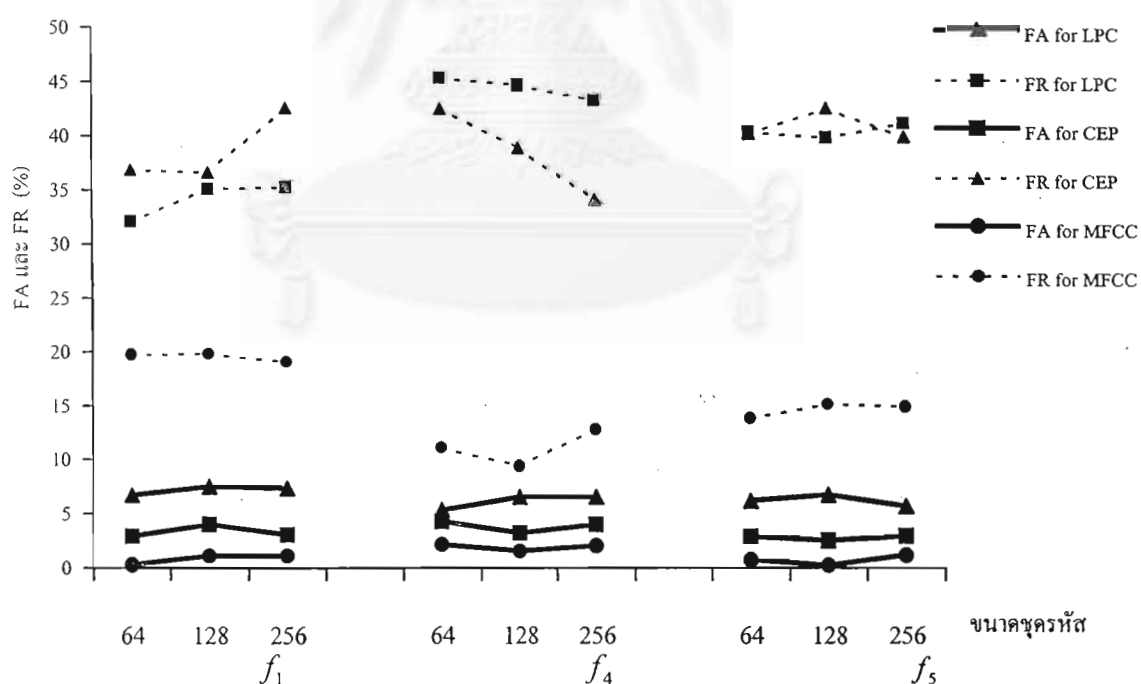
การทดลองที่ 4-3-2-2 ระบบบ่งชี้ผู้พูดแบบระบบเปิด

ในการทดลองนี้เป็นการทดลองเพื่อศึกษาถึงผลของขนาดของชุดรหัสต่ออัตราการยอมรับและการปฏิเสธผิดพลาดในระบบบ่งชี้ผู้พูดแบบระบบเปิด สำหรับฐานข้อมูลที่ใช้เป็นฐานข้อมูลชุดเดียวกันกับการทดลองที่ 3-2 ได้ผลการทดลองแสดงดังตารางที่ 4-9

ตารางที่ 4-9 ค่า FA (%) และค่า FR (%) เฉลี่ย ที่ขนาดของชุดรหัสค่าต่างๆ[§]

Feature		f_1			f_4			f_5		
		64	128	256	64	128	256	64	128	256
LPC	FA	6.69	7.48	7.32	5.31	6.53	6.53	6.16	6.74	5.68
	FR	32.04	35.07	35.24	45.31	44.7	43.21	40.36	39.87	41.12
CEP	FA	2.87	3.98	3.03	4.3	3.18	3.98	2.87	2.55	2.92
	FR	36.8	36.55	42.53	42.47	38.85	34.08	40.2	42.52	39.9
MFCC	FA	0.32	1.11	1.11	2.12	1.54	2.02	0.74	0.27	1.17
	FR	19.76	19.81	19.05	11.11	9.4	12.81	13.85	15.1	14.94

นำผลจากตารางมาแสดงเป็นกราฟได้ดังรูปที่ 4-11



รูปที่ 4-11 ค่า FA และ FR เฉลี่ย ที่ขนาดของชุดรหัสค่าต่างๆ

[§] จำนวนเสียงทดสอบ 262 เสียง (จำนวนเสียงพูดในระบบ 105 และ นอก ระบบ 157 เสียง)

วิเคราะห์และอภิปรายผลการทดลองที่ 4-3-2

1. ระบบปิด

จากกราฟแสดงผลการทดลองจะเห็นว่าอัตราการบ่งชี้ผิดพลาดมีแนวโน้มลดลงเมื่อขนาดของชุดรหัสเพิ่มขึ้น โดยจะให้ผลสอดคล้องกันทั้ง LPC, CEP, และ MFCC แต่ก็จะมีจุดที่ไม่เป็นไปตามแนวโน้ม จากกราฟจะเห็นได้จากจุดที่ใช้ขนาดของชุดรหัสเท่ากับ 256 ของ CEP กับฟังก์ชัน f_1 จะเห็นว่าอัตราการบ่งชี้กลับมีค่าสูงขึ้น นั่นก็หมายความว่า การเพิ่มชุดรหัสไม่จำเป็นต้องให้อัตราการบ่งชี้ที่ดีขึ้น ขนาดของชุดรหัสที่เหมาะสมจะขึ้นอยู่กับลักษณะสำคัญที่ใช้ และอันดับของค่าลักษณะสำคัญด้วย

เมื่อเปรียบเทียบระหว่างลักษณะสำคัญจะเห็นว่า MFCC ให้ผลดีที่สุด และ CEP กับ LPC ให้ผลเป็นอันดับรองลงมาตามลำดับ ซึ่งสอดคล้องกับการทดลองที่ 3

2. ระบบเปิด

จากกราฟแสดงผลการทดลองจะเห็นได้ว่าแนวโน้มของการเปลี่ยนแปลงของค่า FA และ FR ก่อนข้างจะคงที่ แสดงว่าการเพิ่มขนาดของชุดรหัสไม่ได้ทำให้ค่า FA และ FR ดีขึ้น ดังนั้นในการเลือกขนาดของชุดรหัสเพื่อที่จะนำไปใช้งานก็ควรจะต้องเลือกขนาดของชุดรหัสที่มีค่าน้อยเนื่องจากให้ประสิทธิภาพที่ใกล้เคียงกับชุดขนาดของชุดรหัสที่มีค่ามาก และจะทำให้เวลาในการประมวลผลสั้นลง

การทดลองที่ 4-3-3 ผลของจำนวนสถานะของแบบจำลองต่ออัตราการบ่งชี้ในระบบบ่งชี้ผู้พูด

ในการทดลองนี้ได้ทำการทดลองศึกษาผลของจำนวนสถานะของแบบจำลองฮิดเดนมาร์คอฟต่ออัตราการบ่งชี้ โดยได้แปรค่าจำนวนสถานะ 10, 25, และ 40 ขนาดของชุดรหัสเท่ากับ 64, 128, 256, และลักษณะสำคัญได้ใช้ LPC, CEP, MFCC อันดับ 10, 12, 14, และ 16 โดยใช้ f_1 , f_4 , และ f_5 เป็นฟังก์ชันวัดค่าคะแนน

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

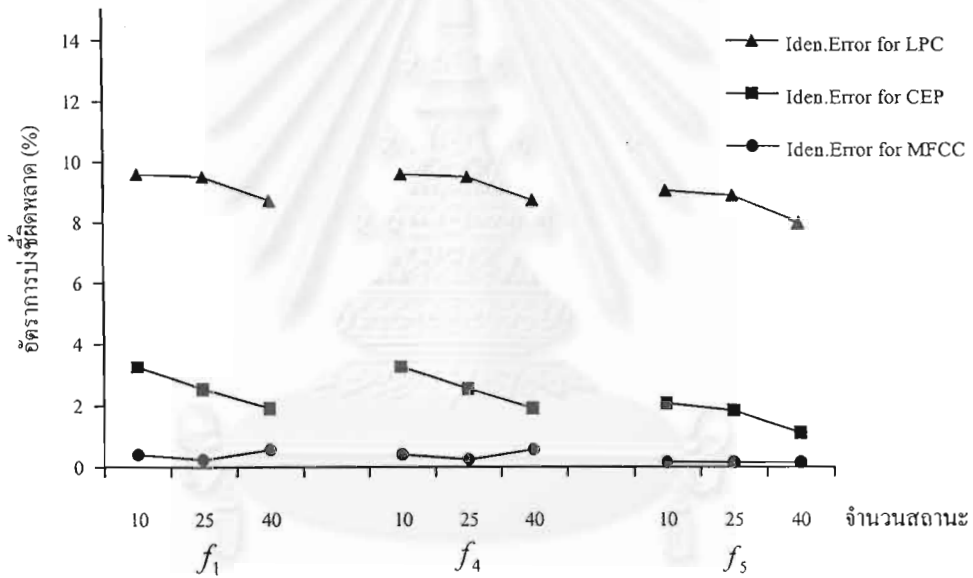
การทดลองที่ 4-3-3-1 ระบบบ่งชี้ผู้พูดแบบระบบปิด

ในการทดลองนี้ได้ใช้ฐานข้อมูลเกี่ยวกับการทดลอง 3-1 ได้ผลการทดลองดังแสดงในตารางที่ 4-10

ตารางที่ 4-10 อัตราการบ่งชี้ผิดพลาดเฉลี่ย (%) ที่จำนวนสถานะค่าต่างๆ สำหรับกลุ่มทดสอบที่ 2⁹

Feature	f_1			f_4			f_5		
	10	25	40	10	25	40	10	25	40
LPC	9.6	9.52	8.73	9.6	9.52	8.73	9.05	8.89	8.02
CEP	3.25	2.54	1.9	3.25	2.54	1.9	2.06	1.83	1.11
MFCC	0.4	0.24	0.56	0.4	0.24	0.56	0.16	0.16	0.16

นำผลที่ได้จากตารางมาแสดงเป็นกราฟได้ดังรูปที่ 4-12



รูปที่ 4-12 อัตราการบ่งชี้ผิดพลาดเฉลี่ย (%) ที่จำนวนสถานะค่าต่างๆ

จุฬาลงกรณ์มหาวิทยาลัย

⁹จำนวนเสียงทดสอบ 105 เสียง

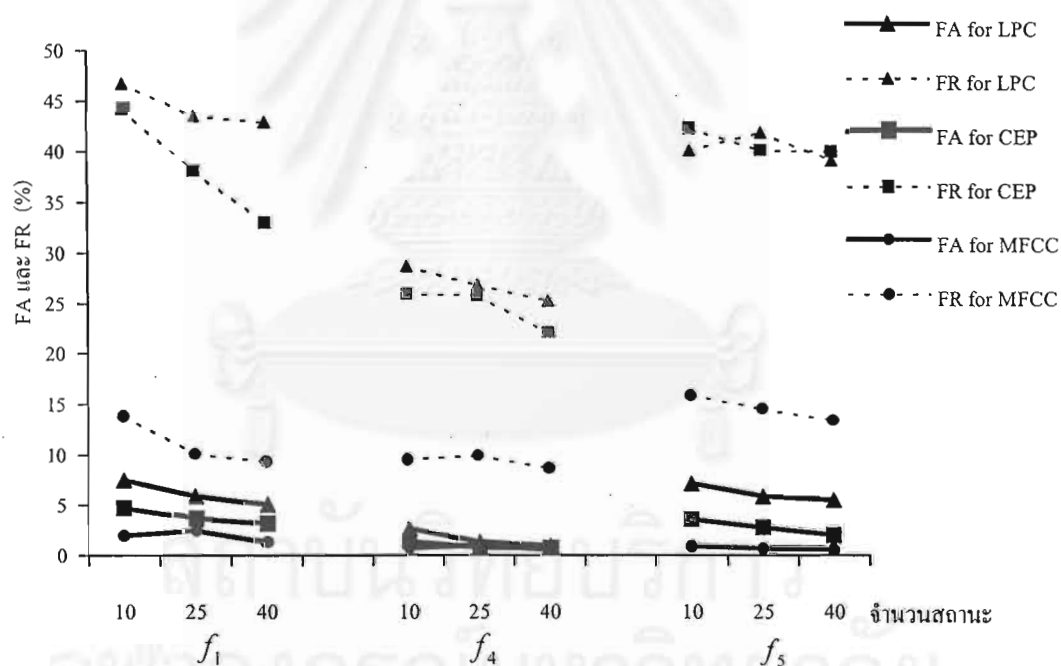
การทดลองที่ 4-3-3-2 ระบบบ่งชี้ผู้พูดแบบระบบเปิด

ได้ผลการทดลองแสดงดังตารางที่ 4-11

ตารางที่ 4-11 ค่า FA (%) และ FR (%) เฉลี่ย ที่จำนวนสถานะค่าต่างๆ สำหรับลุ่มทดสอบที่ 2¹⁰

Feature		f_1			f_4			f_5		
		10	25	40	10	25	40	10	25	40
LPC	FA	7.48	5.89	4.99	2.6	1.27	0.9	7.17	5.89	5.52
	FR	46.73	43.56	42.94	28.72	26.84	25.31	40.18	41.95	39.22
CEP	FA	4.72	3.61	3.13	1.27	0.74	0.69	3.56	2.76	2.02
	FR	44.23	38.12	33.04	25.95	25.84	22.1	42.39	40.19	40.05
MFCC	FA	1.96	2.44	1.27	0.69	0.96	0.48	0.9	0.69	0.58
	FR	13.87	10.1	9.34	9.56	9.94	8.7	15.9	14.55	13.44

นำผลที่ได้แสดงเป็นกราฟได้ดังรูปที่ 4-13



รูปที่ 4-13 ค่า FA และ FR เฉลี่ย ที่จำนวนสถานะค่าต่างๆ

¹⁰ จำนวนเสียงทดสอบ 262 เสียง (จำนวนเสียงพูดในระบบ 105 และ นอกกระบบ 157 เสียง)

วิเคราะห์และอภิปรายผลการทดลองที่ 4-3-3

1. ระบบปิด

จากกราฟผลการทดลองรูปที่ 4-12 จะเห็นว่าเมื่อจำนวนสถานะมีค่าเพิ่มขึ้นแนวโน้มของค่าอัตราการบ่งชี้ผิดพลาดเฉลี่ยจะมีค่าลดลงทั้งสามค่าลักษณะสำคัญ แสดงให้เห็นว่าการเพิ่มจำนวนสถานะนั้นจะสามารถช่วยทำให้อัตราการบ่งชี้ผู้พูดดีขึ้น โดยการเพิ่มสถานะของแบบจำลองนั้นเป็นการเพิ่มจำนวนจุดเพื่อให้แบบจำลองสามารถแยกแยะความแตกต่างระหว่างเสียงพูดแต่ละเสียงได้มากขึ้น ดังนั้นที่จำนวนสถานะของแบบจำลองค่าสูงๆ จะทำให้อัตราการบ่งชี้ผิดพลาดของระบบมีค่าลดลงได้ แต่ในการเพิ่มจำนวนสถานะนั้นจะทำให้การคำนวณใช้เวลามากขึ้น ทั้งช่วงการฝึกฝนและการทดสอบ ดังนั้นในการจะเลือกค่าจำนวนสถานะไปใช้งานต้องพิจารณาทั้งสองปัจจัยควบคู่กันไปด้วย

เมื่อพิจารณาฟังก์ชันวัดค่าคะแนนที่ใช้จะเห็นได้ว่าผลที่ได้จะค่อนข้างใกล้เคียงกัน

2. ระบบเปิด

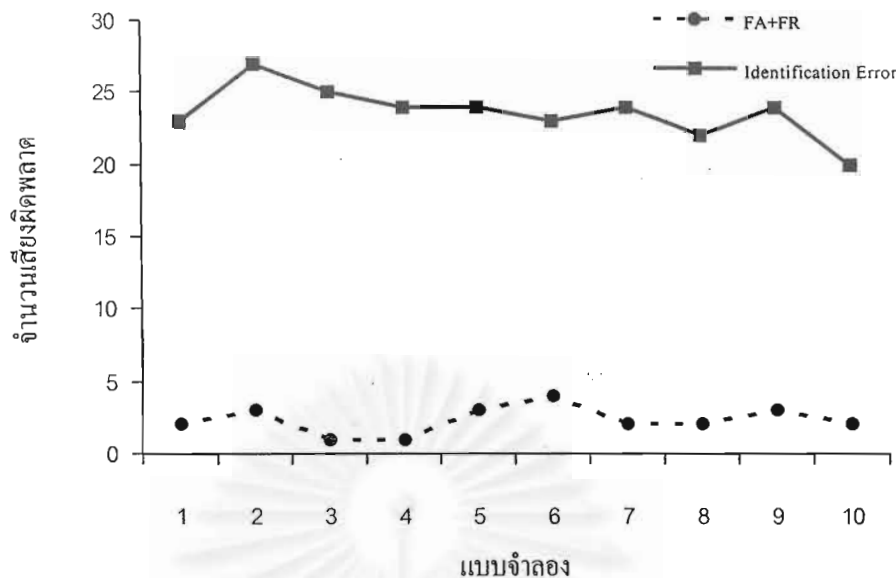
จากกราฟผลการทดลองรูปที่ 4-13 จะเห็นว่าเมื่อเพิ่มจำนวนสถานะของแบบจำลองให้มากขึ้นค่า FA และ FR จะมีแนวโน้มที่ลดลงทั้งสามคุณลักษณะสำคัญ แสดงให้เห็นว่าจำนวนสถานะของแบบจำลองมีผลต่อค่า FA และ FR

เมื่อพิจารณาฟังก์ชันวัดค่าคะแนนที่ใช้จะเห็นได้ว่าผลที่ได้จะค่อนข้างใกล้เคียงกันแต่ฟังก์ชัน f_4 จะให้ผลที่ดีกว่าฟังก์ชัน f_1 และ f_5 เล็กน้อย

การทดลองที่ 4-4 ผลของการฝึกฝนแบบจำลองฮิดเดนมาร์คอฟในแต่ละครั้ง

จากการทดลองที่ผ่านมาทั้งหมด ได้เกิดข้อสังเกตหนึ่งก็คือว่า เมื่อฝึกฝนระบบใหม่ในแต่ละครั้งจะให้ผลที่ไม่เท่ากัน ดังนั้นในการทดลองที่ 4-4 นี้ จะศึกษาถึงการฝึกฝนแบบจำลองฮิดเดนมาร์คอฟในแต่ละครั้ง ว่ามีผลต่อประสิทธิภาพของระบบอย่างไร โดยในการทดลองนี้จะทำการฝึกฝนแบบจำลองทั้งหมด 10 ครั้ง แล้วนำแบบจำลองที่ได้จากการฝึกฝนแต่ละครั้งไปทำการทดสอบ แล้วเปรียบเทียบผลที่ได้ในแต่ละครั้ง

สำหรับการทดลองนี้ได้ใช้ ฐานข้อมูลเดียวกับการทดลองที่ 3 โดยใช้ CEP อันดับ 14 เป็นค่าลักษณะสำคัญ ขนาดของชุดรหัสเท่ากับ 64 จำนวนสถานะเท่ากับ 40 ผลการทดลองแสดงดังรูปที่ 4-19



รูปที่ 4-14 จำนวนเสียงผิดพลาดที่ได้จากแบบจำลองในการฝึกฝนแต่ละครั้ง สำหรับชุดทดสอบที่ 2

จากรูปจะเห็นว่าผลที่ได้ในแต่ละครั้งจะมีค่าไม่เท่ากัน แต่อย่างไรก็ตามจากกราฟจะเห็นได้ว่าการเปลี่ยนแปลงของค่าอัตราการบ่งชี้ผิดพลาดอยู่ในช่วงที่ไม่มากนัก (ประมาณ 0-5 %) ส่วนค่า FA+FR ก็อยู่ในช่วงที่ไม่เปลี่ยนแปลงมากนักเช่นกัน (ประมาณ 20-27 %) นั่นก็คือผลที่ได้ค่อนข้างมีความคงที่ในการฝึกฝนแบบจำลองแต่ละครั้ง

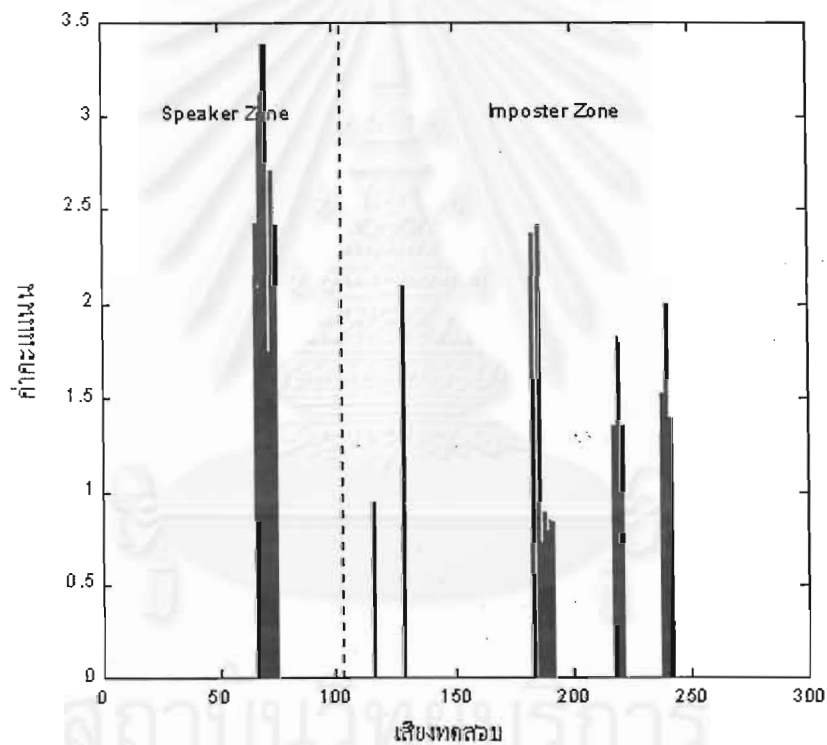
วิเคราะห์และอภิปรายผลการทดลองที่ 4-4

1. จากการทดลองจะเห็นได้ว่าการฝึกฝนระบบในแต่ละครั้ง เมื่อนำแบบจำลองที่ได้จากการฝึกฝนไปทดสอบจะได้ผลที่ต่างกันไป เนื่องจากว่าในการฝึกฝนแต่ละครั้งจะได้แบบจำลองที่ต่างกันไป ซึ่งเป็นข้อด้อยของแบบจำลองฮิดเดนมาร์คอป คือในการฝึกฝนในแต่ละครั้งจะทำให้ได้จุดต่ำสุดสัมพัทธ์ (Local Optimum) แต่จากกราฟจะเห็นได้ว่าความต่างนั้นอยู่ในระดับที่ยอมรับได้ โดยค่าที่ได้จะไม่แตกต่างกันมากนัก ดังนั้นจึงสามารถสรุปได้ว่าการฝึกฝนแบบจำลองฮิดเดนมาร์คอปครั้งเดียวก็เพียงพอแล้ว แต่ถ้าต้องการแบบจำลองที่ดีก็ต้องฝึกฝนหลายครั้งแล้วเลือกแบบจำลองที่ดีที่สุดไปใช้ ถ้าไม่คำนึงเวลาในการฝึกฝนระบบ

การทดลองที่ 4-5 การนำฟังก์ชันค่าความแตกต่างมาใช้ในกระบวนการตรวจสอบผู้พูด

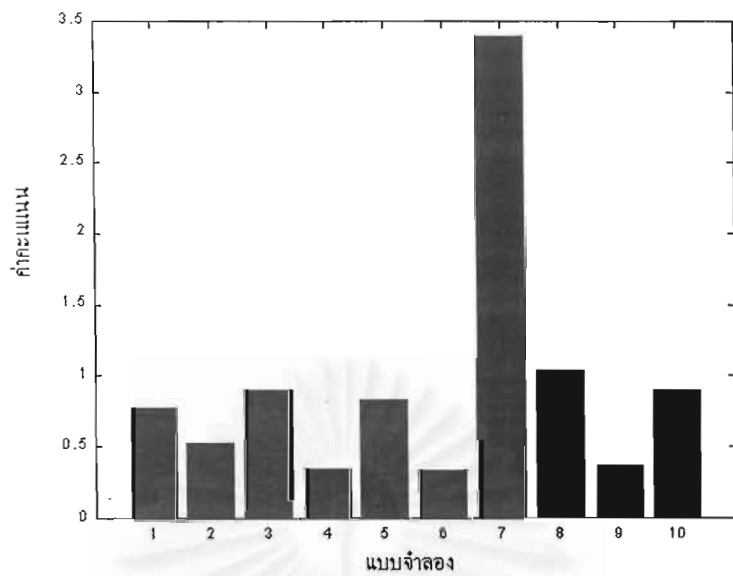
จากขั้นตอนการตรวจสอบผู้พูดนั้นระบบจะนำเอาค่าความน่าจะเป็นที่มากที่สุดมาเปรียบเทียบกับค่าผ่านระดับ เพื่อจะตรวจสอบว่าเสียงพูดที่ได้รับการบ่งชี้เป็นเสียงของผู้พูดในระบบจริงหรือไม่ จากผลการทดลองในการทดลองที่ 4-3 นั้น สำหรับระบบเปิด ประสิทธิภาพของระบบ (ซึ่งวัดออกมาในรูป FA และ FR) ยังไม่ดีเท่าที่ควร ดังนั้นในการทดลองนี้ จึงได้นำเสนอวิธีการในการตรวจสอบผู้พูดใหม่โดยการนำเอาฟังก์ชันค่าความแตกต่างมาใช้ในกระบวนการตรวจสอบผู้พูด โดยรายละเอียดของฟังก์ชันค่าความต่างนั้นได้อธิบายในส่วนท้ายของบทที่ 2

สำหรับการทดลองนี้ ได้ทดลองกับฐานข้อมูลในชุดที่ 2 โดยได้ใช้ค่าลักษณะสำคัญ CEP อันดับ 14 ขนาดของชุดรหัสเท่ากับ 64 จำนวนสถานะเท่ากับ 40 และใช้ f_4 เป็นฟังก์ชันวัดค่าคะแนน ได้ผลการทดลองดังรูปที่ 4-15 ถึงรูปที่ 4-19

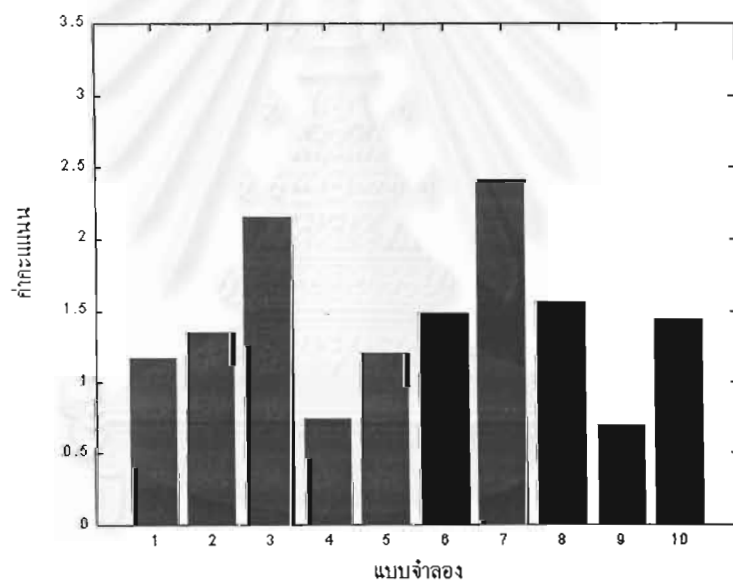


รูปที่ 4-15 ค่าคะแนนของเสียงพูดในกลุ่มทดสอบที่ 2 ที่ถูกบ่งชี้เป็นเสียงของผู้พูดในระบบคนที่ 7

จากรูปที่ 4-15 เป็นรูปที่แสดงค่าคะแนนของเสียงทดสอบจากกลุ่มทดสอบที่ 2 เมื่อยังไม่ได้ใช้ฟังก์ชันค่าความแตกต่าง จากรูปที่ 4-16 (ก.) เป็นค่าคะแนนของเสียงพูดของผู้พูดในระบบคนที่ 7 จะเห็นว่าค่าคะแนนที่ได้จากแบบจำลองที่ 7 จะมีค่าเด่นกว่าค่าคะแนนที่ได้จากแบบจำลองอื่น แต่ถ้าเป็นรูปที่ 4-17 (ข.) นั้นเป็นเสียงของผู้พูดนอกระบบที่ถูกบ่งชี้เป็นเสียงของผู้พูดในระบบคนที่ 7 ค่าคะแนนที่ได้จากแบบจำลองที่ 7 จะมีค่าไม่เด่นจากแบบจำลองอื่นเท่าไรนัก และในรูปที่ 4-17 เป็นรูปที่แสดงค่าคะแนนของเสียงทดสอบเมื่อใช้ฟังก์ชันค่าความแตกต่าง จะเห็นว่าค่าคะแนนของเสียงพูดในระบบแยกจากค่าคะแนนของเสียงพูดนอกระบบได้ชัดเจนกว่ารูปที่ 4-15



(ก). ผู้พูดในระบบคนที่ 7

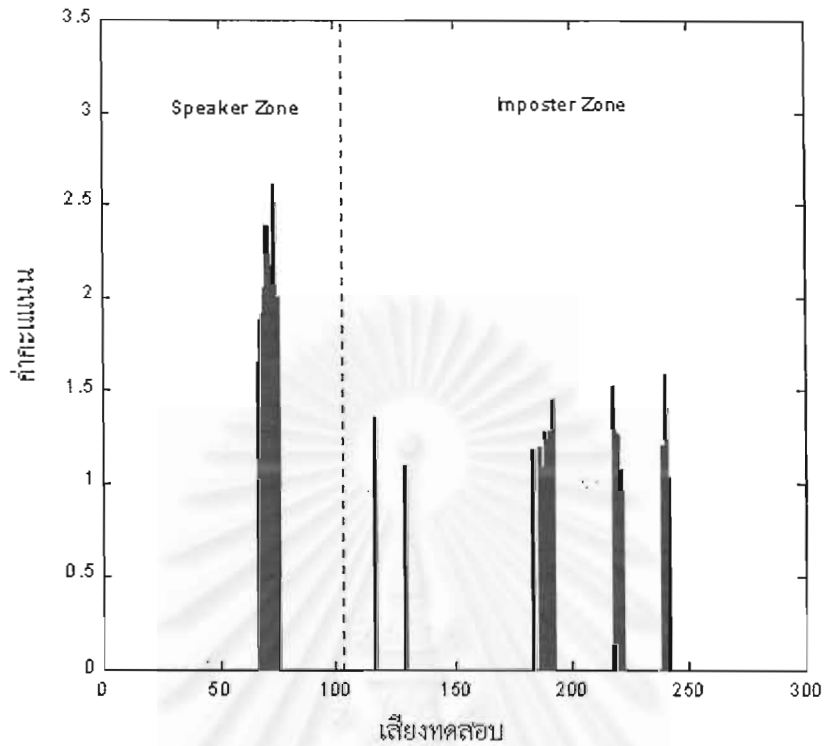


(ข). ผู้พูดนอกระบบที่ถูกบ่งชี้เป็นผู้พูดคนที่ 7

รูปที่ 4-16 ค่าคะแนนของเสียงพูดของผู้พูดที่เกี่ยวข้องกับผู้พูดคนที่ 7

จุฬาลงกรณ์มหาวิทยาลัย

เมื่อนำฟังก์ชันค่าความแตกต่างมาใช้ โดยใช้ค่า $m=1, n=1$ ได้ผลแสดงดังรูปที่ 4-17

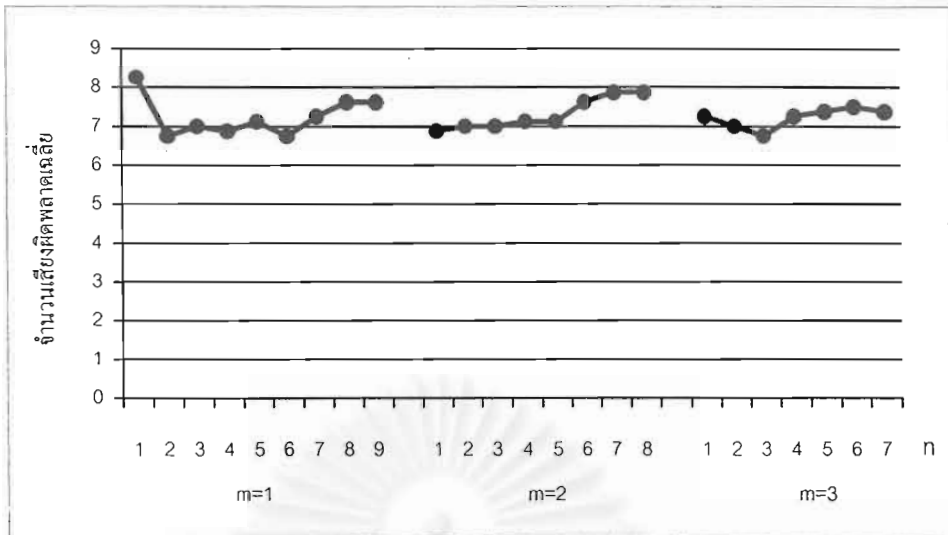


รูปที่ 4-17 ค่าคะแนนของเสียงพูดในกลุ่มทดสอบที่ 2 ที่ถูกบ่งชี้เป็นเสียงของผู้พูดในระบบคนที่ 7 เมื่อใช้ฟังก์ชัน $df(1,1)$

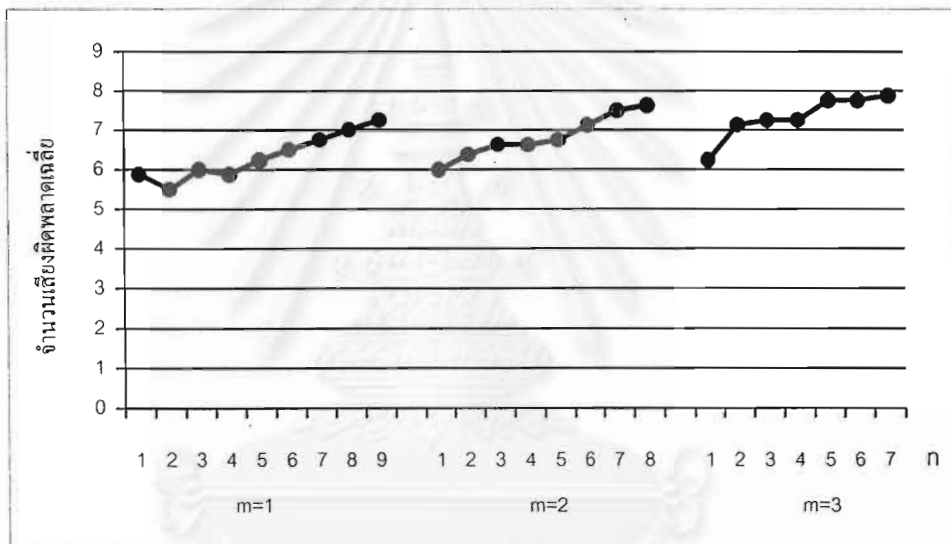
การทดลองที่ 4-5-1 การหาค่าพารามิเตอร์ m, n ที่เหมาะสม

ในการทดลองนี้เป็นการทดลองเพื่อหาค่า m, n ในสมการ 2-105 ที่เหมาะสม โดยได้ทดลองกับฐานข้อมูลชุดเดียวกับการทดลองที่ 3 โดยใช้ค่าลักษณะสำคัญทั้งสามอย่าง คือ LPC, CEP, และ MFCC ที่อันดับ 10, 12, 14, และ 16 ขนาดของชุดรหัสเท่ากับ 64 จำนวนสถานะเท่ากับ 40 และใช้ f_4 เป็นฟังก์ชันวัดค่าคะแนน ได้ผลการทดลองดังรูปที่ 4-18 (ก.) – (ง.)

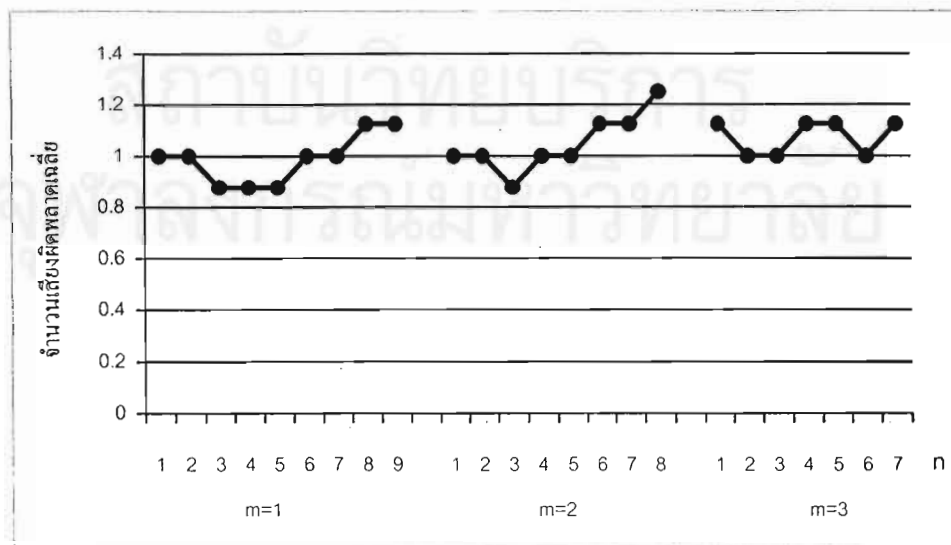
สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



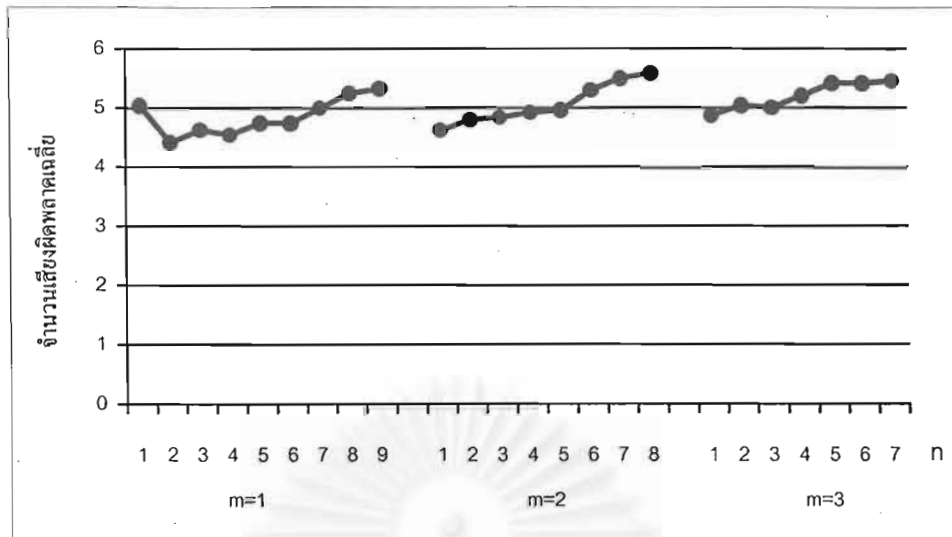
(ก.) เมื่อใช้ LPC



(ข.) เมื่อใช้ CEP



(ค.) เมื่อใช้ MFCC



(ง.) เฉลี่ยรวมทั้งหมด

รูปที่ 4-18 จำนวนเสียงยอมรับและปฏิเสธผิดพลาดเฉลี่ย $((FA+FR)/2)$ ที่ค่า m และ n ค่าต่างๆ

จากรูปที่ 4-18 (ง.) เป็นกราฟแสดงค่าเฉลี่ยของจำนวนเสียงผิดพลาดรวม จะเห็นว่าที่ค่า $m=1$ และ ค่า $n=2$ เป็นตำแหน่งที่ให้ค่าผิดพลาดเฉลี่ยต่ำที่สุด แต่ทั้งนี้ค่าผิดพลาดเฉลี่ยที่ได้ก็ไม่ได้ให้ค่าผิดพลาดที่แตกต่างกันมากเท่าไรนัก

การทดลองที่ 4-5-2 การนำฟังก์ชันความแตกต่างไปใช้กับระบบบ่งชี้ผู้พูดแบบเปิด

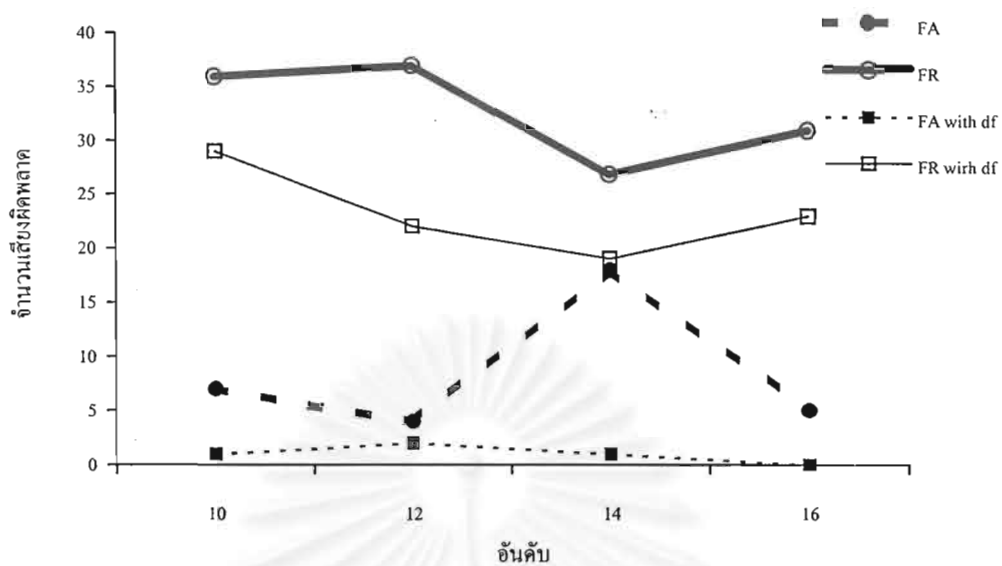
หลังจากที่ได้ทำการทดลองหาค่าพารามิเตอร์ของฟังก์ชันความแตกต่างที่เหมาะสมแล้ว ก็จะนำฟังก์ชันนี้ไปทดสอบกับระบบบ่งชี้ผู้พูด โดยจะใช้ผลการบ่งชี้ผู้พูดจากการทดลองที่ 3 ผลการทดลองแสดงตารางที่ 4-12

ตารางที่ 4-12 จำนวนเสียงยอมรับและปฏิเสธผิดพลาด โดยเปรียบเทียบระหว่างระบบที่ไม่ใช้ฟังก์ชันความแตกต่างกับระบบที่ใช้ฟังก์ชันความแตกต่าง¹¹

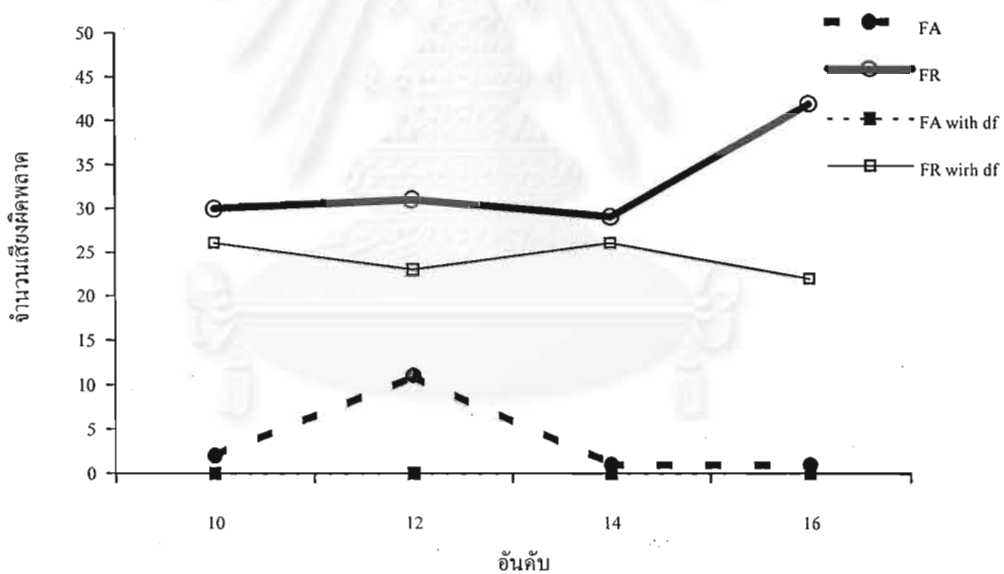
Feature		10		12		14		16	
		ไม่ใช้ df	ใช้ df	ไม่ใช้ df	ใช้ df	ไม่ใช้ df	ใช้ df	ไม่ใช้ df	ใช้ df
LPC	FA	7	1	4	2	18	1	5	0
	FR	36	29	37	22	27	19	31	23
CEP	FA	2	0	11	0	1	0	1	0
	FR	30	26	31	23	29	26	42	22
MFCC	FA	4	0	1	9	5	0	5	1
	FR	5	6	7	3	4	5	1	3

¹¹ จำนวนเสียงทดสอบ 262 เสียง (จำนวนเสียงพูดในระบบ 105 และ นอกในระบบ 157 เสียง)

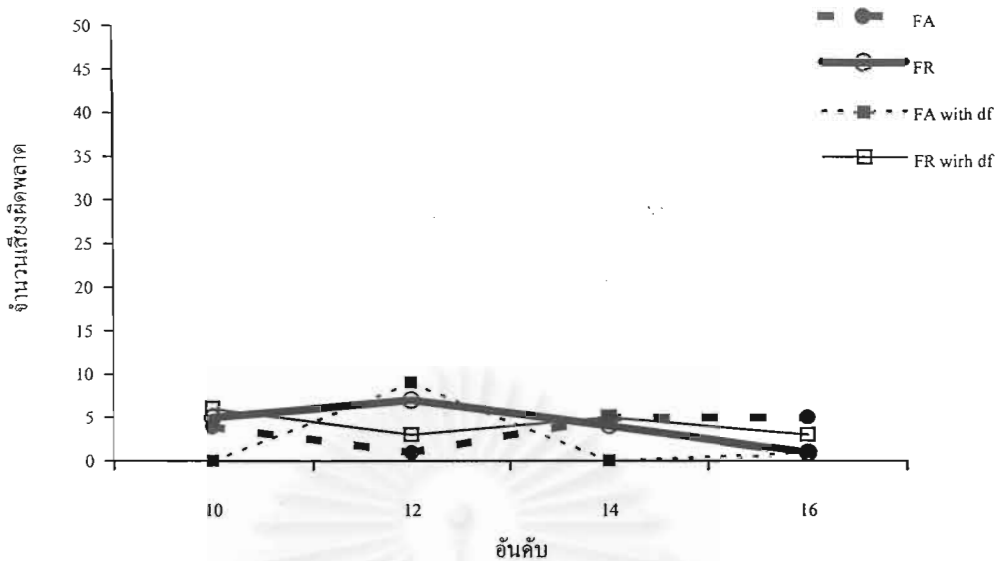
นำผลจากตารางมาแสดงเป็นกราฟได้ดังรูปที่ 4-19



(ก.) เมื่อใช้ LPC



(ข.) เมื่อใช้ CEP



(ค.) เมื่อใช้ MFCC
รูปที่ 4-19 จำนวนเสียงผิดพลาด

ตารางที่ 4-13 อัตราการลดลงของผลรวม FA และ FR เมื่อใช้ฟังก์ชันความแตกต่าง¹²

Feature	ผลรวม FA และ FR		% ลดลง
	ไม่ใช้ df	ใช้ df	
LPC	165	97	41.21
CEP	147	97	34.01
MFCC	32	27	15.63

จุฬาลงกรณ์มหาวิทยาลัย

¹² จำนวนเสียงทดสอบ 262 เสียง (จำนวนเสียงพูดในระบบ 105 และ นอกกระบบ 157 เสียง)

วิเคราะห์และอภิปรายผลการทดลองที่ 4-5

1. จากรูปที่ 4-16 เป็นรูปที่ทำให้ได้แนวความคิดในการนำฟังก์ชันความแตกต่างมาใช้ในขั้นตอนของการตรวจสอบผู้พูด จากรูปจะเห็นว่าค่าคะแนนที่ได้จากการนำเอาเสียงทดสอบไปผ่านแบบจำลองทุกแบบจำลอง ในกรณีที่เสียงทดสอบนั้นเป็นเสียงของผู้พูดในระบบจะให้ค่าคะแนนที่ได้จากแบบจำลองของผู้พูดคนนั้นค่อนข้างที่จะเด่นชัดกว่าค่าคะแนนที่ได้จากแบบจำลองอื่น ดังแสดงในรูปที่ 4-16 (ก.) ซึ่งจะแตกต่างจากรูปแบบของค่าคะแนนที่ได้จากเสียงทดสอบของผู้พูดนอกระบบ ซึ่งค่าคะแนนที่ได้จะไม่เด่นชัดเท่าไรนัก เนื่องจากโอกาสที่เสียงของผู้พูดนอกระบบจะคล้ายกับแบบจำลองมีมากด้วย ดังแสดงในรูปที่ 4-16 (ข.) ดังนั้นจึงเป็นที่มาของแนวความคิดในการนำฟังก์ชันความแตกต่างมาใช้
2. สำหรับรูปแบบของฟังก์ชันความแตกต่างที่นำเสนอในสมการที่ 4-1 นั้นเป็นรูปแบบของสมการของความแตกต่างทั่วไป โดยได้ทำการทดลองหาค่าพารามิเตอร์ที่เหมาะสมดังแสดงในการทดลองที่ 6-2 จากการทดลองจะเห็นได้ว่าค่าพารามิเตอร์ของฟังก์ชันที่ให้ผลโดยรวมที่ดีที่สุด คือ $m=1$ และ $n=2$ นั่นก็คือค่าฟังก์ชันเกิดจากค่าคะแนนที่มากที่สุดลบด้วยค่าคะแนนที่มากเป็นอันดับรองลงมา 2 อันดับ แสดงให้เห็นว่าคะแนนที่มากในอันดับคั่นๆ จะมีอิทธิพลต่อค่าของฟังก์ชันมากกว่าค่าคะแนนอันดับหลังๆ ซึ่งในอันดับหลังๆ จะให้ค่าคะแนนที่ไม่ค่อยจะแตกต่างกันมากนัก
3. จากผลการทดลองจะเห็นว่าเมื่อนำฟังก์ชันค่าความแตกต่างมาใช้ในขั้นตอนตรวจสอบผู้พูด จะทำให้ได้ผลที่ดีขึ้น จะเห็นจากตารางสรุปรวมตารางที่ 4-21 เมื่อใช้ LPC, CEP, และ MFCC เป็นค่าลักษณะสำคัญ ฟังก์ชันความแตกต่างทำให้จำนวนเสียงผิดพลาดลดลงไป 41.21, 34.01, และ 15.63 เปอร์เซ็นต์ ตามลำดับ จากเปอร์เซ็นต์ที่ลดลง จะเห็นว่า MFCC ให้เปอร์เซ็นต์ลดลงที่น้อยที่สุด สาเหตุที่เป็นเช่นนี้คิดว่าโดยปกติ MFCC ให้จำนวนเสียงผิดพลาดที่ต่ำอยู่แล้ว นั่นคือปกติ MFCC มีความสามารถในการแยกแยะผู้พูดได้ดีอยู่แล้ว ดังนั้นเมื่อนำฟังก์ชันความแตกต่างมาใช้จึงทำให้ช่วยลดจำนวนเสียงผิดพลาดลงได้บ้าง แต่เป็นจำนวนไม่มาก ซึ่งจะตรงข้ามกับ LPC

การทดลองที่ 4-6 ผลของจำนวนผู้พูดในระบบต่อระบบบ่งชี้ผู้พูด

สำหรับการทดลองนี้เป็นการทดลองเพื่อศึกษาถึงผลของจำนวนผู้พูดในระบบว่ามีผลต่อประสิทธิภาพของระบบอย่างไรบ้าง โดยการทดลองนี้ได้แบ่งเป็น 2 การทดลองย่อย ดังรายละเอียดต่อไปนี้

การทดลองที่ 4-6-1 การนำเสียงจาก 1 ช่วงการบันทึกเสียงมาฝึกฝน

ในการทดลองนี้ได้ทำการเพิ่มจำนวนผู้พูดในระบบ แล้ววัดค่าอัตราการบ่งชี้ผู้พูด อัตราการยอมรับผิดพลาดและอัตราการปฏิเสธผิดพลาด โดยในการทดลองนี้ได้ทดลองกับฐานข้อมูลชุดที่ 3 ซึ่งเป็นฐานข้อมูลตัวเลข 0-9 เมื่อนำมาใช้จะนำเสียง 3, 5, 2, 9, และ 4 มาต่อกันเพื่อให้คล้ายกับฐานข้อมูลเดิมที่เป็นเสียงพูดต่อเนื่อง สำหรับฐานข้อมูลนี้ได้แบ่งชุดของเสียงพูดออกเป็น 3 ชุด ได้แก่ ชุดที่หนึ่งเป็นชุดฝึกฝน โดยจะนำเสียงพูดที่ต่อกันแล้วจำนวน 5 เสียงต่อคนมาทำการฝึกฝน ชุดที่สองเป็นชุด

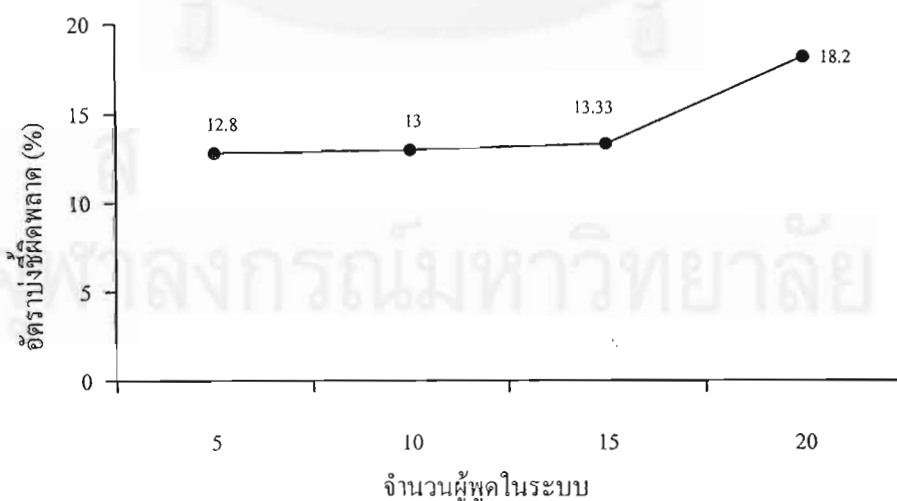
ทดสอบเพื่อนำมาหาค่าผ่านระดับ แบ่งเป็นเสียงพูดของผู้พูดในระบบคนละ 5 เสียง โดยเป็นเสียงที่มาจากช่วงบันทึกเดียวกันกับชุดทดสอบ และเสียงผู้พูดนอกระบบคนละ 10 เสียงจากจำนวนผู้พูดนอกระบบ 21 คน โดยเป็นเสียงพูดที่มาจากสองช่วงการบันทึก รวมมีจำนวนเสียงทั้งหมด 740 เสียง (จำนวนเสียงของผู้พูดในระบบ 25-100 เสียง จำนวนเสียงของผู้พูดนอกระบบ 640 เสียง) และชุดที่สามเป็นชุดทดสอบระบบ โดยเป็นเสียงพูดที่มาจากช่วงการบันทึก 5 ช่วง ช่วงละ 10 เสียงต่อคน รวมมีจำนวนเสียงทั้งหมด 1600 เสียง (จำนวนเสียงของผู้พูดในระบบ 250-1000 เสียง จำนวนเสียงของผู้พูดนอกระบบ 600 เสียง)

สำหรับการทดลองได้ทดลองกับจำนวนผู้พูดในระบบ 5 10 15 และ 20 คนตามลำดับ โดยได้ใช้ MFCC อันดับ 16 เป็นค่าลักษณะสำคัญ ขนาดของชุดรหัสเท่ากับ 256 จำนวนสถานะของแบบจำลองเท่ากับ 40 ใช้ f_4 เป็นฟังก์ชันวัดค่าคะแนน และใช้ฟังก์ชันความแตกต่างร่วมด้วย (เนื่องจากผลการทดลองกับฐานข้อมูลในชุดที่ 2 ให้ผลที่ดีสำหรับค่าพารามิเตอร์ดังกล่าว) และใช้จำนวนผู้พูดนอกระบบ 21 คนทุกการทดลอง ผลการทดลองแสดงดังตารางที่ 4-14

ตารางที่ 4-14 อัตราเสียงผิดพลาด (%) เมื่อเปลี่ยนแปลงจำนวนผู้พูดในระบบ¹³

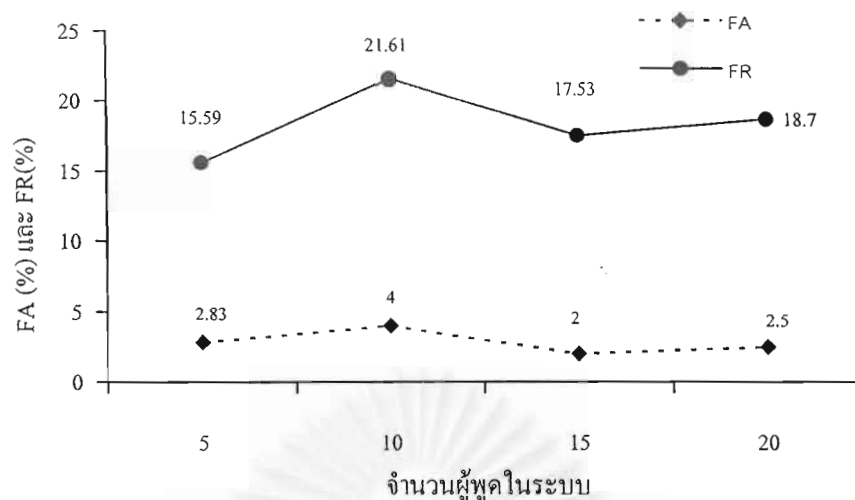
Error (%)	จำนวนผู้พูดในระบบ			
	5	10	15	20
Iden. Error Rate	12.80	13.00	13.33	18.20
FA	2.83	4.00	2.00	2.50
FR	15.59	21.61	17.53	18.70

นำผลที่ได้แสดงเป็นกราฟได้ดังรูปที่ 4-20



รูปที่ 4-20 จำนวนเสียงบ่งชี้ผิดพลาดเมื่อให้จำนวนผู้พูดในระบบจำนวนต่างๆ

¹³ จำนวนเสียงทดสอบ 1600 เสียง (เสียงพูดของผู้พูดในระบบ 250-1000 เสียง เสียงพูดของผู้พูดนอกระบบ 600 เสียง)



รูปที่ 4-21 จำนวน FA และ FR เมื่อให้จำนวนผู้พูดในระบบจำนวนต่างๆ

วิเคราะห์และอภิปรายผลการทดลองที่ 4-6-1

1. จากผลการทดลองจะเห็นได้ว่าเมื่อเพิ่มจำนวนผู้พูดในระบบมากขึ้น (จำนวนผู้พูดนอกระบบเท่าเดิม) อัตราการบ่งชี้ผิดพลาดมีแนวโน้มที่จะเพิ่มขึ้น ทั้งนี้เนื่องจากว่าเมื่อระบบมีจำนวนผู้พูดในระบบมากขึ้น โอกาสที่เสียงจะถูกบ่งชี้เป็นผู้พูดคนใดคนหนึ่งลดลง
2. การเพิ่มของอัตราการบ่งชี้ผิดพลาดไม่ค่อยสูงมากนัก ยกเว้นในกรณีที่จำนวนคนในระบบเป็น 20 คน อัตราการบ่งชี้ผิดพลาดจะค่อนข้างสูง เนื่องจากว่า 5 คนที่เพิ่มเข้าไปใหม่มีจำนวนเสียงที่ค่อนข้างผิดปกติจึงทำให้อัตราการบ่งชี้ผิดพลาดที่เพิ่มมากขึ้น
3. สำหรับอัตราการยอมรับและการปฏิเสธผิดพลาดก็ทำนองเดียวกัน คือมีแนวโน้มที่เพิ่มมากขึ้นแต่จะมีจุดที่น่าสังเกตจุดหนึ่งคือ เมื่อจำนวนผู้พูดในระบบเท่ากับ 10 ค่าอัตราการยอมรับผิดพลาดมีค่ามากกว่าปกติ เนื่องจากการเพิ่มผู้พูดจาก 5 คน เป็น 10 คนนั้น 5 คนที่เพิ่มเข้าไปมีจำนวนเสียงผิดพลาดที่มาก เมื่อเทียบกับจำนวนผิดพลาดที่ได้จากการเพิ่มผู้พูด จาก 10 เป็น 15 และ 20 ตามลำดับ
4. โดยรวมแล้ว เมื่อเพิ่มจำนวนผู้พูดในระบบจะทำให้อัตราการบ่งชี้ผิดพลาดเพิ่มขึ้น และอัตราการยอมรับและปฏิเสธผิดพลาดเพิ่มขึ้น แต่การเพิ่มไม่ได้เพิ่มขึ้นมากนัก แสดงให้เห็นว่าระบบสามารถให้เปอร์เซ็นต์ความผิดพลาดที่ค่อนข้างจะคงที่

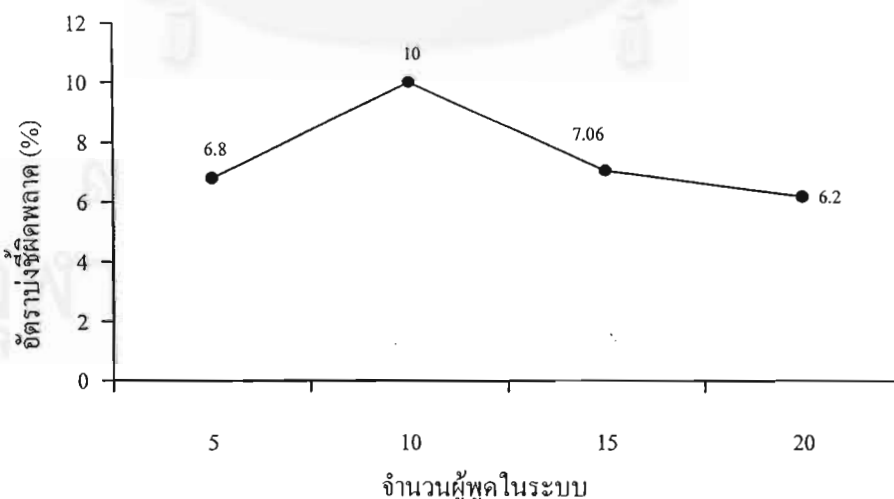
การทดลองที่ 4-6-2 การนำเสียงจาก 3 ช่วงการบันทึกเสียงมาฝึกฝน

ในการทดลองนี้ใช้ฐานข้อมูลชุดที่ 3 ได้นำเสียงจากการบันทึกในช่วงที่ 1, 2, และ 3 มาทำการฝึกฝนระบบ โดยนำเสียงมา 5 เสียงต่อช่วงการบันทึก ดังนั้นจะมีเสียงที่ใช้ในการฝึกฝน 15 เสียง ฐานข้อมูลตัวเลข 3-5-2-9-4 สำหรับเสียงของผู้พูดได้จากเสียงพูดผู้ชาย 28 คน ผู้หญิง 13 คน สำหรับกลุ่มทดสอบเพื่อนำมาหาค่าผ่านระดับนั้น ได้มาจากเสียงในช่วงการบันทึกที่ 1 2 และ 3 ที่เหลือ โดยมีจำนวนผู้พูดนอกระบบ 21 คน ในกลุ่มทดสอบที่ 1 นี้จะมีจำนวนเสียงพูดรวมทั้งสิ้น 940 เสียง แบ่งเป็นเสียงของผู้พูดในระบบ 5, 10, 15, และ 20 คน คนละ 15 เสียง เสียงของผู้พูดนอกระบบ 21 คน รวม 640 เสียง และกลุ่มทดสอบระบบนั้น ใช้ช่วงการบันทึกที่ 4-8 ทั้งหมด ในกลุ่มทดสอบที่ 2 นี้ มีเสียงพูดรวมทั้งสิ้น 1600 เสียง เสียง แบ่งเป็นเสียงของผู้พูดในระบบ 5, 10, 15, และ 20 คน คนละ 50 เสียง เสียงของผู้พูดนอกระบบ 21 คนรวม 600 เสียงในวิธีการในการวัดค่าคะแนนใช้ f_4 เป็นฟังก์ชันวัดค่าคะแนนและใช้ฟังก์ชันความแตกต่างในขั้นตอนการตรวจสอบผู้พูด ได้ผลการทดลองดังแสดงในตารางที่ 4-15

ตารางที่ 4-15 อัตราเสียงผิดพลาด (%) เมื่อเปลี่ยนแปลงจำนวนผู้พูดในระบบ¹⁴

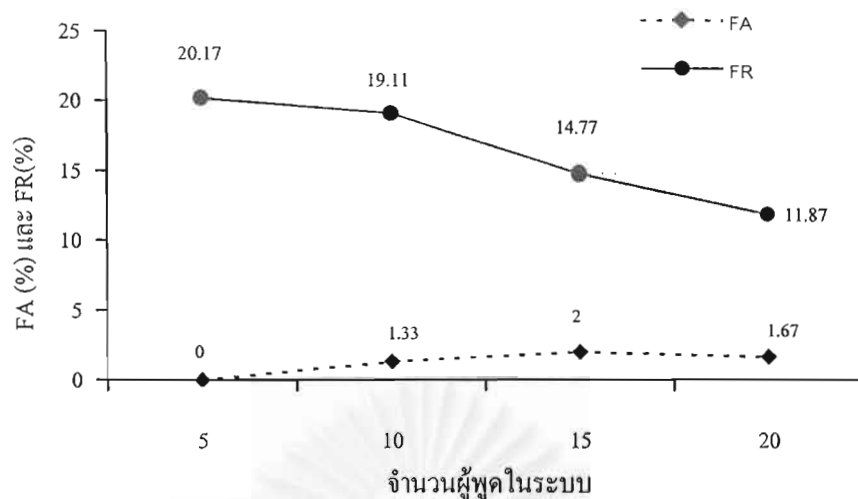
Error (%)	จำนวนผู้พูดในระบบ			
	5	10	15	20
Iden. Error Rate	6.80	10.00	7.06	6.2
FA	0	1.33	2.00	1.67
FR	20.17	19.11	14.77	11.87

นำมาแสดงเป็นกราฟได้ดังรูป



รูปที่ 4-22 จำนวนเสียงบ่งชี้ผิดพลาดเมื่อให้จำนวนผู้พูดในระบบจำนวนต่างๆ

¹⁴ จำนวนเสียงทดสอบ 1600 เสียง (เสียงพูดของผู้พูดในระบบ 250-1000 เสียง เสียงพูดของผู้พูดนอกระบบ 600 เสียง)



รูปที่ 4-23 จำนวน FA และ FR เมื่อให้จำนวนผู้พูดในระบบจำนวนต่างๆ

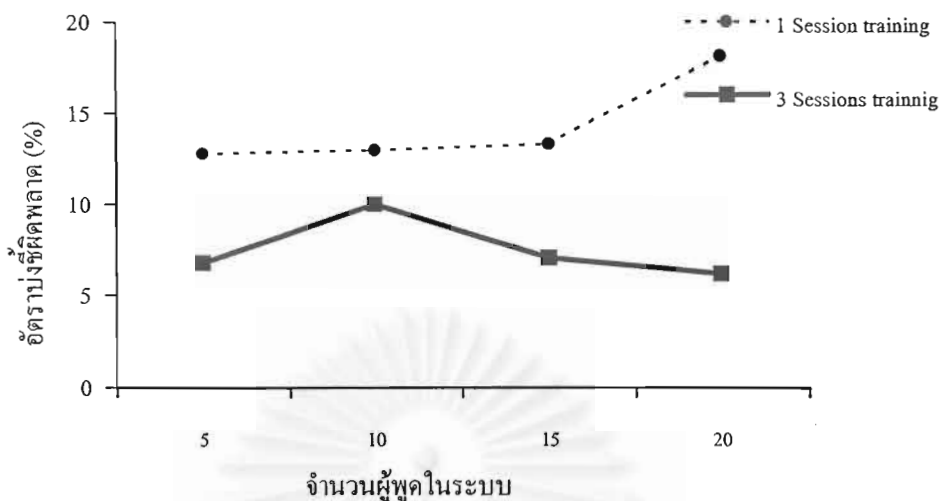
ตารางที่ 4-16 รายละเอียดของเสียงที่บ่งชี้ผิดพลาด¹⁵

ลำดับ	5		10		15		20	
	เสียงที่บ่งชี้ผิดพลาด	จำนวนเสียงที่บ่งชี้ผิดพลาด	เสียงที่บ่งชี้ผิดพลาด	จำนวนเสียงที่บ่งชี้ผิดพลาด	เสียงที่บ่งชี้ผิดพลาด	จำนวนเสียงที่บ่งชี้ผิดพลาด	เสียงที่บ่งชี้ผิดพลาด	จำนวนเสียงที่บ่งชี้ผิดพลาด
1	F14	14	F14	19	F14	21	F14	21
2			F16**	1	F16	1	F16	1
3					F22	1	F22	1
4							F32**	1
5	M13	3	M13	3	M13	3	M13	3
6			M24**	7	M24	7	M24	7
7			M37**	20	M37	20	M37	20
8							M45**	1
9							M47**	5
10							M48**	2
รวม		17		50		53		62

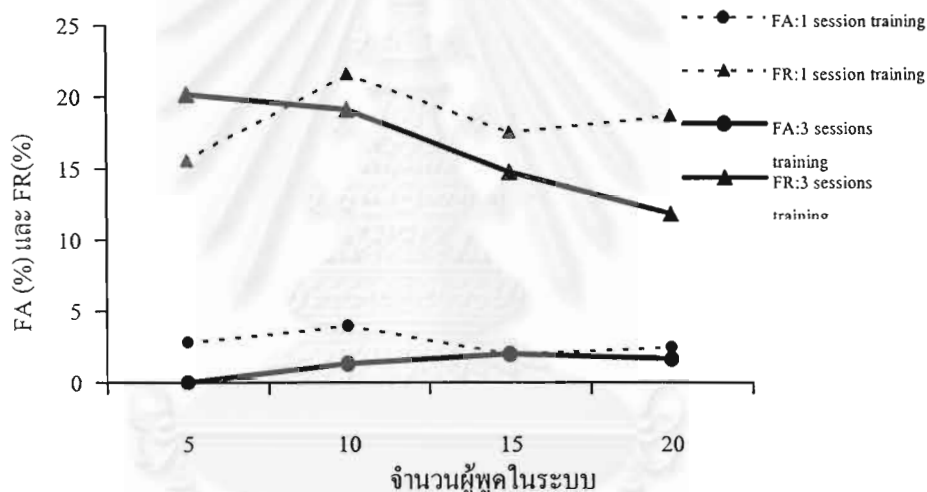
หมายเหตุ **เสียงของผู้พูดที่ผิดเพิ่มจากระบบเดิม

เมื่อนำผลที่ได้จากทั้งสองการทดลองมาแสดงร่วมกันได้ดังรูปที่ 4-24 และรูปที่ 4-25

¹⁵ จำนวนเสียงทดสอบของผู้พูดในระบบ 250-1000 เสียง



รูปที่ 4-24 อัตราการบ่งชี้ผิดพลาดจากการทดลองที่ 6 และการทดลองที่ 7



รูปที่ 4-25 ค่า FA และ FR จากการทดลองที่ 6 และการทดลองที่ 7

วิเคราะห์และอภิปรายผลการทดลองที่ 4-6-2

1. จากรูปที่ 4-24 และรูปที่ 4-25 จะเห็นว่าเมื่อนำเสียงจากหลายช่วงการบันทึกมาฝึกฝนจะทำให้ได้ผลที่ดีขึ้น นั่นแสดงว่าการนำเสียงจากหลายช่วงการบันทึกมาทำการฝึกฝนจะทำให้ระบบได้เรียนรู้ลักษณะการพูดของผู้พูดได้มากขึ้น
2. จากตารางที่ 4-16 และกราฟในรูปที่ 4-22 และ 4-23 จะเห็นว่า ค่าอัตราการบ่งชี้ผิดพลาดมีแนวโน้มลดลงเมื่อจำนวนของผู้พูดเพิ่มขึ้นจาก 10 เป็น 15 และ 20 สาเหตุที่เป็นเช่นนี้ก็เนื่องมาจากว่าเสียงของผู้พูดที่บ่งชี้ผิดนั้นจะเริ่มผิดในจำนวนผู้พูดในระบบ 10 คน (F14, M24, M37) ดังแสดงในตารางแสดงจำนวนเสียงที่บ่งชี้ผิดพลาด แต่เมื่อเพิ่มจำนวนผู้พูดเพิ่มเป็น 15 คน จำนวนเสียงที่บ่งชี้ผิดพลาดมีค่าเพิ่มขึ้นจากเดิมไม่มาก เมื่อเทียบกับจำนวนของเสียงทดสอบที่เพิ่มเข้ามา ดังนั้นจึงทำให้

อัตราการบ่งชี้ผิดพลาดมีค่าลดลง สำหรับค่า FR ก็อธิบายด้วยเหตุผลเดียวกัน เนื่องจากค่า FR เป็นค่าที่พิจารณาจากเสียงของผู้พูดในระบบ ส่วนค่า FA จะมีแนวโน้มที่เพิ่มขึ้นเนื่องจากค่า FA เป็นค่าที่พิจารณาจากเสียงของผู้พูดนอกระบบ ซึ่งมีจำนวนเสียงทดสอบคงที่ ดังนั้นเมื่อจำนวนผู้พูดในระบบเพิ่มมากขึ้นย่อมทำให้โอกาสของเสียงที่ยอมรับผิดพลาดมีค่ามากขึ้นด้วย จึงเป็นผลให้แนวโน้มของค่า FA มีค่ามากขึ้นเมื่อจำนวนผู้พูดในระบบมีค่ามากขึ้น



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 5

สรุปผลและข้อเสนอแนะ

สรุปผล

ในงานวิจัยนี้เป็นงานวิจัยการบ่งชี้ผู้พูดแบบระบบเปิด โดยใช้แบบจำลองฮิดเดนมาร์คอฟในการบ่งชี้ผู้พูด และนำชุดรหัสแบบหลายชุดมาใช้ในกระบวนการควอนไทซ์แบบเวกเตอร์ จากผลการทดลองที่ได้นำเสนอในบทที่ 4 ได้ทำการทดลองเป็นขั้นตอนตามลำดับและสามารถสรุปผลการทดลองตามลำดับดังต่อไปนี้

1. การทดลองที่ 1 เป็นการทดลองเปรียบเทียบอัตราการบ่งชี้ผู้พูดในระบบบ่งชี้ผู้พูดแบบระบบปิดระหว่างระบบบ่งชี้ผู้พูดที่ใช้แบบจำลองฮิดเดนมาร์คอฟในการบ่งชี้ และใช้ชุดรหัสแบบชุดรหัสเดี่ยว ซึ่งเป็นงานวิจัยที่ได้มีการทำมาก่อน กับระบบบ่งชี้ผู้พูดที่ใช้แบบจำลองฮิดเดนมาร์คอฟในการบ่งชี้เช่นกัน แต่ใช้ชุดรหัสแบบหลายชุดซึ่งเป็นระบบที่ได้นำเสนอ จากผลการทดลองสามารถสรุปได้ว่า ระบบบ่งชี้ผู้พูดที่ใช้ชุดรหัสแบบชุดรหัสเดี่ยวให้อัตราการรู้จำที่ดีกว่าระบบบ่งชี้ผู้พูดที่ใช้ชุดรหัสแบบหลายชุด เมื่อใช้แบบจำลองฮิดเดนมาร์คอฟในการรู้จำ
2. สำหรับระบบบ่งชี้ผู้พูดที่ใช้ชุดรหัสแบบหลายชุด เป็นระบบที่มีความแตกต่างของชุดรหัสที่ใช้ในแต่ละผู้พูด ดังนั้นเมื่อนำความแตกต่างซึ่งวัดในรูปแบบของค่าความความเพี้ยนจากขั้นตอนการควอนไทซ์แบบเวกเตอร์มาใช้ในการรู้จำร่วมกับแบบจำลองฮิดเดนมาร์คอฟ โดยในการนำมาใช้ร่วมกันนั้นได้กำหนดจากฟังก์ชันวัดค่าคะแนน f_1 , f_3 , f_4 , f_5 , และ f_6 (f_2 เป็นฟังก์ชันวัดค่าคะแนนจากระบบที่ใช้ชุดรหัสแบบชุดรหัสเดี่ยว) จากผลการทดลองปรากฏว่าเมื่อนำค่าความเพี้ยนมาใช้ร่วมกับแบบจำลองฮิดเดนมาร์คอฟในระบบที่ใช้ชุดรหัสแบบหลายชุดนั้นให้ผลที่ดีกว่าระบบที่ใช้ชุดรหัสแบบชุดรหัสเดี่ยว และที่สำคัญค่าคะแนนที่ได้นั้นจะมีประโยชน์ในขั้นตอนการตรวจสอบผู้พูดในระบบบ่งชี้ผู้พูดแบบระบบเปิดด้วย โดยค่าคะแนนที่ได้นั้นมี ความเด่นชัดของเสียงผู้พูดของคนในระบบ ทำให้สามารถแยกแยะระหว่างเสียงผู้พูดในระบบกับผู้พูดนอกระบบได้ง่ายขึ้น ดังนั้นสามารถสรุปได้ว่า ระบบบ่งชี้ผู้พูดแบบระบบปิดนั้น ระบบที่ใช้ชุดรหัสแบบหลายชุดที่นำค่าความความเพี้ยนมาใช้ร่วมกับแบบจำลองฮิดเดนมาร์คอฟนั้นให้ผลดีกว่าระบบที่ใช้ชุดรหัสเดี่ยว
3. ผลของฟังก์ชันวัดค่าคะแนนสำหรับระบบที่ใช้ชุดรหัสแบบหลายชุดจะเห็นว่า ฟังก์ชัน f_4 และ f_5 ให้ผลดีกว่าฟังก์ชันอื่นๆ และมีความสามารถแยกแยะความแตกต่างระหว่างเสียงผู้พูดที่เป็นคนคนเดียวกับแบบจำลองกับเสียงผู้พูดคนอื่นได้ดี
4. ในการทดลองที่ 3-1, 3-2, และ 3-3 ได้พิจารณาผลของค่าลักษณะสำคัญ อันดับของค่าสำคัญ ขนาดของชุดรหัสและจำนวนสถานะของแบบจำลองฮิดเดนมาร์คอฟ ต่อระบบบ่งชี้ผู้พูดทั้งระบบปิดและระบบเปิดสามารถสรุปผลของแต่ละค่าได้ดังนี้

- 4.1 ค่าลักษณะสำคัญที่ได้นำใช้ในการทดลองนี้ได้แก่ LPC, CEP, และ MFCC จากผลการทดลองสามารถสรุปได้ว่า MFCC ให้ผลดีที่สุด และ CEP ให้ผลดีรองลงมา และ LPC ให้ผลค้อยที่สุด
 - 4.2 เมื่อพิจารณาอันดับของค่าลักษณะสำคัญ จากผลการทดลองจะเห็นว่าค่าอันดับที่ให้ผลดีนั้นจะขึ้นอยู่กับลักษณะสำคัญที่ใช้ โดยสามารถสรุปได้ดังนี้ สำหรับ LPC อันดับ 12 นั้นค่อนข้างที่จะให้ผลโดยรวมดีที่สุด สำหรับ CEP อันดับ 16 ให้ผลโดยรวมดีที่สุด และ MFCC อันดับ 16 ให้ผลโดยรวมดีที่สุด
 - 4.3 เมื่อพิจารณาขนาดของชุดรหัส จากผลการทดลองจะเห็นว่าขนาดของชุดรหัส (ที่ใช้ในการทดลอง) จะไม่ค่อยมีผลต่อระบบเท่าไร ซึ่งให้ผลทำนองเดียวกันทั้งสามค่าลักษณะสำคัญ แต่ได้ทดลองเพิ่มเติมโดยได้ลดขนาดของชุดรหัสลงไปอีกจากเดิม 64 เป็น 32 และ 16 ปรากฏว่าอัตราการบ่งชี้ผิดพลาดเพิ่มขึ้น ซึ่งสามารถสรุปได้ว่าขนาดของชุดรหัสสั้น ถ้ามีขนาดเล็กเกินไปก็จะให้ผลไม่ดี แต่เมื่อเพิ่มไปถึงจุดหนึ่งจะเริ่มอึดตัว ดังนั้นในการเลือกใช้ชุดรหัสควรเลือกจุดที่เริ่มให้ผลลงที่ เพราะถ้าเลือกชุดรหัสขนาดใหญ่จะทำให้เสียเวลาในการประมวลผลมาก
 - 4.4 เมื่อพิจารณาจำนวนสถานะของแบบจำลองฮิดเดนมาร์คอฟ สามารถสรุปได้ว่าเมื่อเพิ่มจำนวนสถานะมากขึ้นจะทำให้ได้ผลที่ดีขึ้น โดยให้ผลสอดคล้องกันทั้งสามลักษณะสำคัญ แต่การเพิ่มจำนวนสถานะจะทำให้เวลาในการประมวลผลทั้งขั้นตอนการฝึกฝนและขั้นตอนการทดสอบเพิ่มขึ้นด้วย
5. สำหรับการทดลองในระบบเปิดนั้นได้มีการนำเอาฟังก์ชันของความแตกต่างมาใช้ในขั้นตอนตรวจสอบผู้พูด โดยรูปแบบของฟังก์ชันนี้ได้แสดงไว้ในผลการทดลอง จากผลการทดลองสามารถสรุปได้ว่าการนำฟังก์ชันของความแตกต่างนี้มาใช้จะทำให้ผลดีขึ้น โดยสามารถลดอัตราการตรวจสอบผิดพลาดลงได้ 41.21 เปอร์เซ็นต์ สำหรับ LPC, 34.01 เปอร์เซ็นต์ สำหรับ CEP, และ 15.63 เปอร์เซ็นต์สำหรับ MFCC
 6. สำหรับรูปแบบของฟังก์ชันของความแตกต่างที่จะนำมาใช้นั้นได้ทำการทดลองในรูปแบบต่างๆ โดยสามารถสรุปได้ว่ารูปแบบที่ให้ผลดีที่สุดคือ ค่า $m=1$ และ ค่า $n=2$ โดย m, n เป็นพารามิเตอร์ของฟังก์ชันของความแตกต่าง
 7. จากผลการทดลองที่ได้ทดลองมานั้น จุดที่ให้ผลของระบบที่ดีที่สุดคือจุดที่ใช้ค่าลักษณะสำคัญเป็น MFCC อันดับ 16 ขนาดของชุดรหัสเท่ากับ 64 จำนวนสถานะเท่ากับ 40 ใช้ฟังก์ชัน f_4 เป็นฟังก์ชันวัดค่าคะแนน และนำฟังก์ชันความแตกต่างมาใช้ในขั้นตอนการตรวจสอบ โดยให้อัตราการบ่งชี้ผิดพลาด 0 เปอร์เซ็นต์ ให้อัตราการยอมรับผิดพลาด 1.27 เปอร์เซ็นต์ และอัตราการปฏิเสธผิดพลาด 3.81 เปอร์เซ็นต์ สำหรับผู้พูดในระบบ 10 คน และผู้พูดนอกระบบ 17 คน

8. ในการทดลองค่า FA ที่ได้จะมีค่าต่ำกว่าค่า FR เสมอ ขึ้นอยู่กับค่าผ่านระดับที่เลือกใช้ การที่ค่า FA มีค่าต่ำกว่าค่า FR นั้น แสดงว่าระบบนี้สามารถป้องกันผู้พูดนอกระบบได้ดี แต่ในทางกลับกันการยอมรับผู้พูดในระบบก็จะยาก นั่นคือผู้พูดในระบบต้องตั้งใจพูดเพื่อให้เสียงมีความคล้ายคลึงกับเสียงเดิมมากที่สุด ดังนั้นการกำหนดค่าผ่านระดับนี้ก็ขึ้นอยู่กับความต้องการของระบบว่าต้องการให้ระบบยอมรับผู้พูดดีแค่ไหน
9. ในการทดลองต่อมาได้ทดลองถึงผลของการฝึกฝนแบบจำลองในแต่ละครั้งว่าจะให้ผลแตกต่างกันอย่างไร จากการทดลองสามารถสรุปได้ว่าการฝึกฝนแบบจำลองในแต่ละครั้งนั้นให้ผลที่ค่อนข้างคงที่
10. ในการทดลองที่ 6 และ 7 ได้ทดลองเพิ่มจำนวนผู้พูดในระบบ สำหรับการทดลองนี้ได้ใช้ฐานข้อมูลชุดที่ 3 ซึ่งเป็นตัวเลขโดดแยกกัน เมื่อนำมาใช้ในการทดลองได้นำเสียงเหล่านั้นมาต่อกันเป็นเสียงพูดต่อเนื่อง จากผลการทดลองสามารถสรุปได้ว่าเมื่อจำนวนผู้พูดในระบบเพิ่มขึ้นทำให้ประสิทธิภาพของระบบมีแนวโน้มลดลง แต่การลดลงนั้นมีค่าไม่มากเท่าไรนัก นั่นแสดงว่าระบบบ่งชี้ผู้พูดที่นำเสนอมานี้สามารถรองรับจำนวนผู้พูดได้เป็นจำนวนมาก

ข้อเสนอแนะเพื่อปรับปรุงประสิทธิภาพของระบบ

1. ในขั้นตอนการบันทึกเสียงนั้นควรจะบันทึกในสถานะแวดล้อมเดียวกันในทุกช่วงของการบันทึก เช่น การ์ดบันทึกเสียงเป็นการ์ดเดิม ไมโครโฟนตัวเดิม โปรแกรมที่ใช้ในการบันทึกอันเดิม และการปรับค่าพารามิเตอร์ของโปรแกรมค่าเดิม รวมถึงขนาดของสัญญาณที่บันทึกควรจะมีค่าใกล้เคียงกัน เพื่อให้ระบบมีความถูกต้องมากยิ่งขึ้น
2. ในขั้นตอนการฝึกฝนถ้าเป็นไปได้เสียงที่นำมาใช้ในการฝึกฝนควรจะเป็นเสียงที่มีความชัดเจน และระบบจะมีประสิทธิภาพมากยิ่งขึ้นถ้าเสียงที่นำมาใช้ในการฝึกฝนนั้น มาจากหลายช่วงการบันทึก
3. ในขั้นตอนการหาค่าผ่านระดับนั้น ถ้าสามารถหาจำนวนเสียงของผู้พูดนอกระบบได้มากเท่าไรก็จะยิ่งดีเพราะจะทำให้ค่าผ่านระดับที่ได้มีค่าที่ดีมากยิ่งขึ้นเพราะจะทำให้มีความหลากหลายของเสียงมากขึ้น
4. ในการฝึกฝนระบบนั้นถ้าต้องการแบบจำลองที่มีความถูกต้องมากขึ้น ก็สามารถทำได้โดยการลดค่าผ่านระดับในการฝึกฝนลงเพื่อทำให้แบบจำลองให้ความสำคัญจะเป็นของเสียงฝึกฝนมากที่สุด และฝึกฝนแบบจำลองหลายครั้งแล้วเลือกครั้งที่ให้ค่าความน่าจะเป็นสุดท้ายมากที่สุด แต่ทั้งนี้ก็จะทำให้เสียเวลามากยิ่งขึ้น
5. ถ้าต้องการความถูกต้องของการบ่งชี้มากยิ่งขึ้น สามารถนำแบบจำลองฮิดเดนมาร์คอฟแบบต่อเนื่องมาใช้แทนแบบจำลองฮิดเดนมาร์คอฟแบบคิสคริตได้ เพียงแต่จะทำให้เพิ่มเวลาในการฝึกฝนระบบนานมากขึ้น

แนวทางในการพัฒนางานวิจัยในขั้นต่อไป

1. สำหรับในงานวิจัยนี้เป็นงานวิจัยที่เน้นการนำระบบบ่งชี้มาใช้ และวิธีการในการตรวจสอบผู้พูด แต่ไม่ได้เน้นไปที่การวิเคราะห์คำสำคัญ รวมทั้งลักษณะคำพูดที่มีผลต่อการบ่งชี้ของระบบ ดังนั้นถ้าได้มีการวิจัยในส่วนดังกล่าวนี้เพิ่มเติมก็น่าจะเป็นประโยชน์มากขึ้น
2. เพิ่มประสิทธิภาพของระบบ โดยทดลองนำแบบจำลองฮิดเดนมาร์คอฟแบบกึ่งต่อเนื่อง (Semi-Continuous Hidden Markov Model : SCHMM) หรือ แบบจำลองฮิดเดนมาร์คอฟแบบต่อเนื่อง (Continuous Hidden Markov Model : CHMM) มาใช้แทนแบบจำลองฮิดเดนมาร์คอฟแบบดิสครีต ซึ่งเป็นแบบที่ใช้ในงานวิจัยได้
3. จากผลที่ได้ในงานวิจัยนี้ถ้านำไปประยุกต์ใช้งานจริง ก็จะเป็นประโยชน์มากยิ่งขึ้น โดยสามารถนำไปประยุกต์ได้ทั้งการสร้างระบบบ่งชี้ผู้พูดด้วยซอฟต์แวร์บนเครื่องคอมพิวเตอร์ หรือแม้แต่กระทั่งการพัฒนาเป็นฮาร์ดแวร์ด้วยไมโครคอนโทรลเลอร์หรือ DSP chip ก็น่าจะเป็นไปได้
4. ระบบบ่งชี้ผู้พูดที่ได้นำเสนอเป็นระบบที่ขึ้นกับบทคำพูด การพัฒนาขั้นต่อไปควรพัฒนาระบบบ่งชี้ผู้พูดแบบไม่ขึ้นกับบทคำพูด
5. นำระบบบ่งชี้ผู้พูดนี้ไปพัฒนากับฐานข้อมูลที่มีเสียงพูดผ่านระบบสายโทรศัพท์ (Telephone Line) ซึ่งจะเป็นการเพิ่มงานประยุกต์ใช้ระบบบ่งชี้ผู้พูดมากขึ้น

รายการอ้างอิง

ภาษาไทย

- คณะนักวิจัยและพัฒนาระบบระบุผู้พูดสำหรับภาษาไทย. การระบุผู้พูดด้วย LPC และ DTW สำหรับภาษาไทย. เอกสารการประชุมประจำปีของศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ สำนักงานวิทยาศาสตร์และเทคโนโลยีแห่งชาติ, 2541.
- ณัฐกร ทับทอง. การรู้จำคำพูดภาษาไทย โดยใช้ลักษณะบ่งความต่างของหน่วยเสียง. วิทยานิพนธ์ปริญญาโท สาขาวิศวกรรมคอมพิวเตอร์ จุฬาลงกรณ์มหาวิทยาลัย, 2538.
- วิศรุต อาชุนทร. ระบบการรู้จำคำไทยหลายพยางค์แบบไม่ขึ้นกับผู้พูดโดยใช้แบบจำลองฮดเดนมาร์คอฟ. วิทยานิพนธ์ปริญญาโท สาขาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2539.
- วิสิทธิ์ ลีลาศิริวงศ์. การศึกษาลักษณะเฉพาะเชิงสวณศาสตร์ของสระ อี, ออ, อุ ในภาษาไทยและประโยชน์ในการบ่งชี้ผู้พูด. วิทยานิพนธ์ปริญญาโท สาขาวิชาฟิสิกส์ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2539.
- เสาวลักษณ์ อารีย์พงศา. การรู้จำเสียงพูดตัวเลขเป็นภาษาไทยแบบไม่ขึ้นกับผู้พูดโดยใช้วิธีฮดเดนมาร์คอฟ โมเดล และเวกเตอร์ควอนไทเซชัน. วิทยานิพนธ์ปริญญาโท สาขาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2538.
- อังสนา จามิกรณ์. การแปรในการออกเสียงพยัญชนะท้ายของคำยืมภาษาอังกฤษที่ลงท้ายด้วยเสียงเสียดแทรก-ปุ่มเหงือก : กรณีศึกษาของข้าราชการกองทัพเรือไทย. วิทยานิพนธ์ปริญญาโท สาขาวิชาภาษาศาสตร์ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2532
- เอกฤทธิ์ มณีน้อย. การรู้จำหน่วยเสียงสระภาษาไทยโดยใช้โครงข่ายประสาทเทียม. วิทยานิพนธ์ปริญญาโท สาขาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2541.

ภาษาอังกฤษ

- Ahkuputa, V., Jitapunkul, S., Maneenoi, E., Kasuriya, S., and Amornkul, P., "Comparison of Different Techniques On Thai Speech Recognition," Proceedings of Circuit and System on IEEE Asia-Pacific Conference (November 1998): 177-180.
- Campbell, J.P., and JR, Senior Member, "Speaker Recognition: A Tutorial," Proceedings of the IEEE International Conference Vol. 85, No.9. (September 1997): 1437-1462.
- Farrell, K., Kosonovsky, S., and Mammone, R., "Neural Tree Network/Vector Quantization Probability Estimators for Speaker Recognition," IEEE (1994): 279-288.

- Forsyth E., "Semi-Continuous Hidden Markov Model for Speaker Verification," Doctoral Thesis in Philosophy, University of Edinburgh (1995).
- Furui, S. Digital Speech Processing, Synthesis and Recognition. New York and Basel: Marcel Dekker, 1989.
- Gu, Y., and Thomas, T., "A Hybrid Score Measurement for Hmm-Based Speaker Verification," IEEE (1999): 317-320.
- Huang X., "Semi-Continuous Hidden Markov Models for Speech Recognition," Doctoral Thesis in Philosophy, University of Edinburgh (1989).
- Isobe, T., and Takahashi, J., "A New Cohort Normalization Using Local Acoustic Information for Speaker Verification," IEEE (1999): 841-844.
- Jacobs, T., and Setlur, A., "A Field Study of Performance Improvements in HMM-Based Speaker Verification," IEEE (1994): 121-124.
- Levinson, S.E., and Roe, D.B., "A Perspective on Speech Recognition," IEEE Communication Magazine (January 1990): 28-34.
- Liu, C., Wang, H., and Lee, C., "Speaker Verification Using Normalized Log-Likelihood Score," IEEE Transactions on Speech and Audio Processing Vol. 4, No.1. (January 1996): 56-60.
- Naik, J.M., "Speaker Verification: A Tutorial," IEEE Communication Magazine (January 1990): 42-48.
- O'Shaughnessy, D. Speech Communication Human and Machine. Addison-Wesley Publishing Company 1987.
- Picone, J., Fundamentals of Speech Recognition: A short course. Institute for Signal and Information Processing. Mississippi State University.
- Rabiner, L.R., "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," Proceedings of the IEEE Vol. 77, No.2. (February 1989): 257-286.
- Rabiner, L.R. and Levinson, S.E. Isolated and Connected Word Recognition-Theory and Selected Applications. IEEE Transaction on Communication. May 1981 Page(s): 621-659.
- Slomka, S., Castellano, P., and Sridha, S., "A Comparison of Three Discrimination Models for Automatic Speaker Verification," ISSPA, Gold Coast, Austraria (August, 1996): 325-328.

- Tolba, H. and O' Shaughnessay, D. "Automatic speech recognition based on cepstral coefficient and mel-based discrete energy operator, " Proceedings of the 1998 IEEE International Conference on Acoustic, Speech and Signal Processing Volume 2, 1998:973-976.
- Tuzan, O.B. Demirekle, M. and Nakiboglu, K.B. "Comparison of parametric and non-parametric representations of speech for recognition, " Proceedings of the 7th Mediterranean Electrotechnical Conference Volume 1, 1994:65-68.
- Zhang, Y., Alder, M., Togneri, R. "Using Gaussian mixture modeling in Speech recognition, " International Conference Acoustic, Speech, and Signal Processing (ICASSP-94) Volume I , 1994: I/613-I/616.
- Zhang, Y. "Speaker-Independent Isolated word Recognition, " Doctoral Thesis in Philosophy, University of Western Australia, July 1995.



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



ภาคผนวก

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ ก-1 จำนวนเสียงที่บ่งชี้ผิดพลาดสำหรับชุดทศลอกบที่ 1 เมื่อใช้ LPC อันดับ 10 ที่จำนวนสถานะค่าต่างๆ

Codebook Size	t_1	t_{1+df}	t_s			t_{s+df}			t_r			t_{r+df}		
			10	25	40	10	25	40	10	25	40	10	25	40
64	1	1	1	1	1	1	1	1	1	1	1	1	1	1
128	1	1	1	0	0	1	0	0	1	0	0	1	0	0
256	1	1	1	0	0	1	0	0	1	0	0	1	0	0

ตารางที่ ก-2 จำนวนเสียงที่บ่งชี้ผิดพลาดสำหรับชุดทศลอกบที่ 2 เมื่อใช้ LPC อันดับ 10 ที่จำนวนสถานะค่าต่างๆ

Codebook Size	t_1	t_{1+df}	t_s			t_{s+df}			t_r			t_{r+df}		
			10	25	40	10	25	40	10	25	40	10	25	40
64	1	1	1	1	1	1	1	1	1	1	1	1	1	1
128	1	1	1	0	0	1	0	0	1	0	0	1	0	0
256	1	1	1	0	0	1	0	0	1	0	0	1	0	0

ตารางที่ ก-3 จำนวน FA FR สำหรับชุดทดลองที่ 1 เมื่อใช้ LPC อันดับ 10 ที่จำนวนสถานะค่าต่างๆ

Codebook Size		f_1	f_{1+df}	f_2			f_{2+df}			f_3			f_{3+df}		
				10	25	40	10	25	40	10	25	40	10	25	40
64	FA	5	1	6	3	8	1	2	1	4	2	3	2	1	1
	FR	7	1	6	4	4	1	0	0	5	4	4	1	0	1
128	FA	5	1	3	2	1	1	1	2	5	4	6	3	2	2
	FR	7	1	10	6	6	1	0	0	5	4	2	1	0	2
256	FA	8	3	9	7	3	4	1	1	9	5	5	0	2	2
	FR	6	1	5	2	3	1	0	0	4	5	3	2	0	0

ตารางที่ ก-4 จำนวน FA FR สำหรับชุดทดลองที่ 2 เมื่อใช้ LPC อันดับ 10 ที่จำนวนสถานะค่าต่างๆ

Codebook Size		f_1	f_{1+df}	f_2			f_{2+df}			f_3			f_{3+df}		
				10	25	40	10	25	40	10	25	40	10	25	40
64	FA	13	4	8	11	7	2	6	3	9	9	5	6	4	4
	FR	31	27	37	33	40	29	19	28	33	36	39	24	26	31
128	FA	12	6	14	10	5	8	0	1	9	9	9	8	6	4
	FR	31	29	43	38	40	29	31	26	34	31	28	31	26	30
256	FA	11	4	14	10	6	4	1	2	10	7	4	4	2	1
	FR	26	30	44	41	42	27	22	21	31	43	36	32	29	26

ตารางที่ ก-5 จำนวนเสียงที่บ่งชี้ผิดพลาดสำหรับชุดทศลอกบที่ 1 เมื่อใช้ LPC อันดับ 12 ที่จำนวนสถานะค่าต่างๆ

Codebook Size	t_1	t_{1+df}	t_2			t_{2+df}			t_3			t_{3+df}		
			10	25	40	10	25	40	10	25	40	10	25	40
64	0	0	1	2	0	1	2	0	1	2	0	1	2	0
128	0	0	1	1	0	1	1	0	0	0	0	0	0	0
256	0	0	0	0	0	0	0	0	0	0	0	0	0	0

ตารางที่ ก-6 จำนวนเสียงที่บ่งชี้ผิดพลาดสำหรับชุดทศลอกบที่ 2 เมื่อใช้ LPC อันดับ 12 ที่จำนวนสถานะค่าต่างๆ

Codebook Size	t_1	t_{1+df}	t_2			t_{2+df}			t_3			t_{3+df}		
			10	25	40	10	25	40	10	25	40	10	25	40
64	9	9	10	7	14	10	7	14	10	10	13	10	10	13
128	9	9	8	7	7	8	7	7	8	8	7	8	8	7
256	8	8	9	7	7	9	7	7	8	6	6	8	6	6

ตารางที่ ก-7 จำนวน FA FR สำหรับชุดทดลองที่ 1 เมื่อใช้ LPC อันดับ 12 ที่จำนวนสถานะค่าต่างๆ

Codebook Size		f_1	f_{1+df}	f_2			f_{2+df}			f_3			f_{3+df}		
				10	25	40	10	25	40	10	25	40	10	25	40
64	FA	7	1	8	4	5	0	1	0	7	4	6	1	2	1
	FR	7	1	5	7	5	0	0	0	5	5	6	1	3	3
128	FA	5	1	8	4	5	0	0	0	8	3	5	0	1	2
	FR	8	0	6	4	2	1	0	1	5	7	4	2	1	2
256	FA	5	1	12	2	4	0	0	0	3	5	3	2	1	0
	FR	7	0	3	4	3	0	0	0	8	3	3	1	0	1

ตารางที่ ก-8 จำนวน FA FR สำหรับชุดทดลองที่ 2 เมื่อใช้ LPC อันดับ 12 ที่จำนวนสถานะค่าต่างๆ

Codebook Size		f_1	f_{1+df}	f_2			f_{2+df}			f_3			f_{3+df}		
				10	25	40	10	25	40	10	25	40	10	25	40
64	FA	9	3	9	6	4	4	1	1	10	6	9	3	1	2
	FR	31	31	35	48	40	28	41	26	38	47	34	31	41	29
128	FA	13	4	14	9	12	6	3	3	14	8	7	5	5	2
	FR	39	31	43	42	37	28	32	27	42	48	40	37	37	34
256	FA	9	5	15	6	7	2	2	1	5	11	5	4	4	3
	FR	39	32	39	38	40	30	21	25	54	41	45	33	32	33

ตารางที่ ก-9 จำนวนเสียงที่บ่งชี้ผิดพลาดสำหรับชุดทดลองที่ 1 เมื่อใช้ LPC อันดับ 14 ที่จำนวนสถานะค่าต่างๆ

Codebook Size	t_1	t_{1+df}	t_2			t_{2+df}			t_3			t_{3+df}		
			10	25	40	10	25	40	10	25	40	10	25	40
64	1	1	2	1	1	2	1	1	1	0	1	1	0	1
128	0	0	0	0	1	0	0	1	0	0	1	0	0	1
256	0	0	1	0	0	1	0	0	0	0	0	0	0	0

ตารางที่ ก-10 จำนวนเสียงที่บ่งชี้ผิดพลาดสำหรับชุดทดลองที่ 2 เมื่อใช้ LPC อันดับ 14 ที่จำนวนสถานะค่าต่างๆ

Codebook Size	t_1	t_{1+df}	t_2			t_{2+df}			t_3			t_{3+df}		
			10	25	40	10	25	40	10	25	40	10	25	40
64	10	10	15	14	13	15	14	13	12	12	11	12	12	11
128	10	10	11	13	8	11	13	8	11	10	8	11	10	8
256	13	13	9	7	6	9	7	6	10	9	6	10	9	6

ตารางที่ ก-11 จำนวน FA FR สำหรับชุดทดลองที่ 1 เมื่อใช้ LPC อันดับ 14 ที่จำนวนสถานะค่าต่างๆ

Codebok Size		f_1	f_{1+df}	f_2			f_{2+df}			f_3			f_{3+df}		
				10	25	40	10	25	40	10	25	40	10	25	40
64	FA	3	3	8	3	5	0	2	1	7	9	6	3	3	2
	FR	8	1	7	10	8	1	1	0	5	7	8	2	1	4
128	FA	5	2	1	6	2	0	0	1	6	5	4	5	2	1
	FR	8	2	14	8	9	2	1	0	8	7	7	1	2	2
256	FA	7	0	5	7	4	1	0	1	5	7	6	1	0	3
	FR	6	2	8	7	6	0	0	0	7	6	8	1	3	1

ตารางที่ ก-12 จำนวน FA FR สำหรับชุดทดลองที่ 2 เมื่อใช้ LPC อันดับ 14 ที่จำนวนสถานะค่าต่างๆ

Codebok Size		f_1	f_{1+df}	f_2			f_{2+df}			f_3			f_{3+df}		
				10	25	40	10	25	40	10	25	40	10	25	40
64	FA	12	4	11	7	9	4	0	1	11	9	8	6	3	3
	FR	29	24	46	37	45	26	20	19	36	35	38	28	26	27
128	FA	12	5	6	11	9	4	3	1	10	9	12	4	6	4
	FR	31	26	60	45	40	23	18	25	40	39	43	25	28	27
256	FA	14	4	14	7	7	2	1	0	9	3	4	1	1	1
	FR	31	27	38	48	44	26	25	26	36	44	39	31	30	34

ตารางที่ ก-13 จำนวนเสียงที่บ่งชี้ผิดพลาดสำหรับชุดทดลองที่ 1 เมื่อใช้ LPC อันดับ 16 ที่จำนวนสถานะค่าต่างๆ

Codebook Size	t_1	t_{1+df}	t_2			t_{2+df}			t_3			t_{3+df}		
			10	25	40	10	25	40	10	25	40	10	25	40
64	1	1	1	0	1	1	0	1	1	0	1	1	0	1
128	1	1	1	1	0	1	1	0	1	1	0	1	1	0
256	1	1	1	1	1	1	1	1	1	1	1	1	1	1

ตารางที่ ก-14 จำนวนเสียงที่บ่งชี้ผิดพลาดสำหรับชุดทดลองที่ 2 เมื่อใช้ LPC อันดับ 16 ที่จำนวนสถานะค่าต่างๆ

Codebook Size	t_1	t_{1+df}	t_2			t_{2+df}			t_3			t_{3+df}		
			10	25	40	10	25	40	10	25	40	10	25	40
64	7	7	9	10	14	9	10	14	6	7	8	6	7	8
128	6	6	9	9	7	9	9	7	6	7	7	6	7	7
256	6	6	7	8	7	7	8	7	8	7	7	8	7	7

ตารางที่ ก-15 จำนวน FA FR สำหรับชุดทดลองที่ 1 เมื่อใช้ LPC อันดับ 16 ที่จำนวนสถานะค่าต่างๆ

Codebook Size		f_1	f_{1+df}	f_c			f_{c+df}			f_s			f_{s+df}		
				10	25	40	10	25	40	10	25	40	10	25	40
64	FA	7	4	5	5	4	1	2	1	9	6	7	1	2	2
	FR	8	2	14	10	7	1	0	1	9	8	9	3	3	2
128	FA	4	3	7	6	6	2	1	2	9	9	12	4	2	3
	FR	7	1	9	9	9	1	0	0	7	7	7	2	2	3
256	FA	3	2	4	7	7	2	1	0	11	12	7	4	2	4
	FR	8	2	11	7	6	1	1	0	7	3	6	3	2	2

ตารางที่ ก-16 จำนวน FA FR สำหรับชุดทดลองที่ 2 เมื่อใช้ LPC อันดับ 16 ที่จำนวนสถานะค่าต่างๆ

Codebook Size		f_1	f_{1+df}	f_c			f_{c+df}			f_s			f_{s+df}		
				10	25	40	10	25	40	10	25	40	10	25	40
64	FA	8	7	13	12	3	5	3	0	18	10	12	7	2	2
	FR	31	30	53	48	45	29	29	24	37	44	42	36	37	32
128	FA	10	6	12	10	11	4	2	0	14	13	13	4	3	1
	FR	33	26	46	36	43	26	21	24	42	38	36	34	34	30
256	FA	12	5	11	12	14	4	2	4	16	17	16	6	4	5
	FR	39	29	48	43	37	26	28	20	38	36	35	31	37	31

ตารางที่ ก-17 จำนวนเสียงที่บ่งชี้ผิดพลาดสำหรับชุดทดลองที่ 1 เมื่อใช้ CEP อันดับ 10 ที่จำนวนสถานะค่าต่างๆ

Codebook Size	t_1	t_{1+df}	t_4			t_{4+df}			t_5			t_{5+df}		
			10	25	40	10	25	40	10	25	40	10	25	40
64	1	1	2	1	1	2	1	1	1	1	1	1	1	1
128	1	1	1	1	1	1	1	1	1	1	1	1	1	1
256	1	1	1	1	1	1	1	1	1	1	1	1	1	1

ตารางที่ ก-18 จำนวนเสียงที่บ่งชี้ผิดพลาดสำหรับชุดทดลองที่ 2 เมื่อใช้ CEP อันดับ 10 ที่จำนวนสถานะค่าต่างๆ

Codebook Size	t_1	t_{1+df}	t_4			t_{4+df}			t_5			t_{5+df}		
			10	25	40	10	25	40	10	25	40	10	25	40
64	1	1	6	3	3	6	3	3	4	3	1	4	3	1
128	1	1	4	4	2	4	4	2	3	3	1	3	3	1
256	2	2	6	6	5	6	6	5	2	3	1	2	3	1

ตารางที่ ก-19 จำนวน FA FR สำหรับชุดทดลองที่ 1 เมื่อใช้ CEP อันดับ 10 ที่จำนวนสถานะค่าต่างๆ

Codebok Size		f_i	f_{i+df}	f_i			f_{i+df}			f_i			f_{i+df}		
				10	25	40	10	25	40	10	25	40	10	25	40
64	FA	5	0	5	3	2	0	1	1	4	2	2	0	0	1
	FR	2	1	2	2	1	0	1	0	1	2	2	1	1	0
128	FA	7	0	6	3	2	0	0	1	3	2	0	1	0	0
	FR	0	1	0	1	0	0	0	0	0	1	2	0	0	1
256	FA	8	0	5	2	3	2	0	1	5	2	3	1	1	0
	FR	1	1	2	0	0	0	0	0	1	0	0	0	0	0

ตารางที่ ก-20 จำนวน FA FR สำหรับชุดทดลองที่ 2 เมื่อใช้ CEP อันดับ 10 ที่จำนวนสถานะค่าต่างๆ

Codebok Size		f_i	f_{i+df}	f_i			f_{i+df}			f_i			f_{i+df}		
				10	25	40	10	25	40	10	25	40	10	25	40
64	FA	10	6	12	9	14	3	5	7	12	7	6	4	5	7
	FR	41	32	48	48	48	33	38	29	41	47	50	37	38	36
128	FA	9	0	8	7	3	1	0	0	9	7	3	2	0	0
	FR	39	33	51	47	45	36	39	30	50	44	52	42	41	38
256	FA	6	1	7	7	5	6	1	1	7	4	4	4	3	1
	FR	54	29	44	35	34	27	27	25	49	36	45	37	35	35

ตารางที่ ก-21 จำนวนเสียงที่บ่งชี้ผิดพลาดสำหรับชุดทดลองที่ 1 เมื่อใช้ CEP อันดับ 12 ที่จำนวนสถานะค่าต่างๆ

Codebook Size	f_1	f_{1+df}	f_2			f_{2+df}			f_3			f_{3+df}		
			10	25	40	10	25	40	10	25	40	10	25	40
64	1	1	1	1	0	1	1	0	1	1	1	1	1	1
128	1	1	1	0	1	1	0	1	1	1	1	1	1	1
256	1	1	1	1	0	1	1	0	1	1	0	1	1	0

ตารางที่ ก-22 จำนวนเสียงที่บ่งชี้ผิดพลาดสำหรับชุดทดลองที่ 2 เมื่อใช้ CEP อันดับ 12 ที่จำนวนสถานะค่าต่างๆ

Codebook Size	f_1	f_{1+df}	f_2			f_{2+df}			f_3			f_{3+df}		
			10	25	40	10	25	40	10	25	40	10	25	40
64	2	2	2	4	2	2	4	2	1	3	1	1	3	1
128	2	2	2	1	1	2	1	1	2	1	2	2	1	2
256	2	2	3	2	1	3	2	1	2	2	1	2	2	1

ตารางที่ ก-23 จำนวน FA FR สำหรับชุดทดลองที่ 1 เมื่อใช้ CEP อันดับ 12 ที่จำนวนสถานะค่าต่างๆ

Codebok Size		f_1	f_{1+df}	f_c			f_{c+df}			f_e			f_{e+df}		
				10	25	40	10	25	40	10	25	40	10	25	40
64	FA	3	0	4	4	0	0	0	0	3	3	5	0	0	0
	FR	4	0	7	1	1	0	0	1	5	2	0	0	0	0
128	FA	6	0	0	3	2	0	0	0	2	2	0	0	0	0
	FR	0	1	3	2	0	0	1	0	1	0	0	0	0	0
256	FA	8	0	3	2	1	0	0	0	4	2	0	0	0	0
	FR	0	1	0	0	0	0	0	0	0	0	0	0	0	0

ตารางที่ ก-24 จำนวน FA FR สำหรับชุดทดลองที่ 2 เมื่อใช้ CEP อันดับ 12 ที่จำนวนสถานะค่าต่างๆ

Codebok Size		f_1	f_{1+df}	f_c			f_{c+df}			f_e			f_{e+df}		
				10	25	40	10	25	40	10	25	40	10	25	40
64	FA	3	1	3	4	1	1	1	1	4	2	5	1	1	1
	FR	37	33	51	38	42	39	24	27	42	39	47	43	38	38
128	FA	8	0	9	4	6	1	0	0	6	7	2	3	0	0
	FR	41	26	47	43	36	26	32	26	46	49	48	30	35	35
256	FA	8	0	9	8	2	4	0	1	7	6	3	2	1	0
	FR	42	27	46	38	30	24	25	21	45	50	44	29	32	30

ตารางที่ ก-25 จำนวนเสียงที่บ่งชี้ผิดพลาดสำหรับชุดทดลองที่ 1 เมื่อใช้ CEP อันดับ 14 ที่จำนวนสถานะค่าต่างๆ

Codebook Size	t_1	t_{1+df}	t_4			t_{4+df}			t_8			t_{8+df}		
			10	25	40	10	25	40	10	25	40	10	25	40
64	1	1	1	1	1	1	1	1	1	1	1	1	1	1
128	1	1	1	0	1	1	0	1	1	0	1	1	0	1
256	1	1	1	1	1	1	1	1	1	1	1	1	1	1

ตารางที่ ก-26 จำนวนเสียงที่บ่งชี้ผิดพลาดสำหรับชุดทดลองที่ 2 เมื่อใช้ CEP อันดับ 14 ที่จำนวนสถานะค่าต่างๆ

Codebook Size	t_1	t_{1+df}	t_4			t_{4+df}			t_8			t_{8+df}		
			10	25	40	10	25	40	10	25	40	10	25	40
64	2	2	4	2	2	4	2	2	3	3	2	3	3	2
128	1	1	4	2	0	4	2	0	2	0	0	2	0	0
256	1	1	2	2	0	2	2	0	1	0	0	1	0	0

ตารางที่ ก-27 จำนวน FA FR สำหรับชุดทดลองที่ 1 เมื่อใช้ CEP อันดับ 14 ที่จำนวนสถานะค่าต่างๆ

Codebok Size		f_1	f_{1+df}	f_c			f_{c+df}			f_e			f_{e+df}		
				10	25	40	10	25	40	10	25	40	10	25	40
64	FA	3	1	5	0	1	0	0	1	2	0	0	0	0	0
	FR	3	0	2	4	1	0	0	0	3	2	2	1	1	1
128	FA	6	1	6	2	1	0	0	0	2	0	0	0	0	0
	FR	1	0	1	1	1	0	0	0	1	2	2	0	0	0
256	FA	6	0	1	4	1	0	0	0	1	0	0	0	0	0
	FR	2	1	2	1	1	0	0	0	1	1	1	0	0	1

ตารางที่ ก-28 จำนวน FA FR สำหรับชุดทดลองที่ 2 เมื่อใช้ CEP อันดับ 14 ที่จำนวนสถานะค่าต่างๆ

Codebok Size		f_1	f_{1+df}	f_c			f_{c+df}			f_e			f_{e+df}		
				10	25	40	10	25	40	10	25	40	10	25	40
64	FA	4	1	10	5	6	0	0	0	3	7	2	0	1	0
	FR	37	23	43	53	30	26	25	23	46	42	41	36	29	30
128	FA	6	1	4	4	1	1	1	0	3	1	1	0	0	0
	FR	35	18	46	29	28	18	19	17	40	36	35	33	28	25
256	FA	5	0	7	7	10	4	4	1	6	5	7	3	1	1
	FR	43	21	45	31	22	20	21	21	43	44	41	31	30	22

ตารางที่ ก-29 จำนวนเสียงที่บ่งชี้ผิดพลาดสำหรับชุดทดลองที่ 1 เมื่อใช้ CEP อันดับ 16 ที่จำนวนสถานะค่าต่างๆ

Codebook Size	t_1	t_{1+df}	t_2			t_{2+df}			t_3			t_{3+df}		
			10	25	40	10	25	40	10	25	40	10	25	40
64	1	1	1	1	1	1	1	1	1	1	1	1	1	1
128	1	1	1	1	1	1	1	1	1	1	1	1	1	1
256	1	1	1	1	1	1	1	1	1	1	1	1	1	1

ตารางที่ ก-30 จำนวนเสียงที่บ่งชี้ผิดพลาดสำหรับชุดทดลองที่ 2 เมื่อใช้ CEP อันดับ 16 ที่จำนวนสถานะค่าต่างๆ

Codebook Size	t_1	t_{1+df}	t_2			t_{2+df}			t_3			t_{3+df}		
			10	25	40	10	25	40	10	25	40	10	25	40
64	2	2	3	2	2	3	2	2	2	2	1	2	2	1
128	0	0	3	2	3	3	2	3	2	2	2	2	2	2
256	1	1	2	2	3	2	2	3	2	1	2	2	1	2

ตารางที่ ก-31 จำนวน FA FR สำหรับชุดทดลองที่ 1 เมื่อใช้ CEP อันดับ 16 ที่จำนวนสถานะค่าต่างๆ

Codebook Size		f_i	f_{i+df}	f_i			f_{i+df}			f_i			f_{i+df}		
				10	25	40	10	25	40	10	25	40	10	25	40
64	FA	2	0	2	1	1	1	0	1	2	0	0	0	0	0
	FR	3	0	2	1	3	0	0	1	2	1	1	1	1	1
128	FA	2	0	2	6	4	0	0	0	1	1	1	0	0	0
	FR	2	0	2	1	0	0	0	0	2	1	2	0	0	0
256	FA	2	0	1	3	0	0	0	0	1	1	0	0	0	0
	FR	3	0	1	1	1	0	0	0	1	1	1	0	0	0

ตารางที่ ก-32 จำนวน FA FR สำหรับชุดทดลองที่ 2 เมื่อใช้ CEP อันดับ 16 ที่จำนวนสถานะค่าต่างๆ

Codebook Size		f_i	f_{i+df}	f_i			f_{i+df}			f_i			f_{i+df}		
				10	25	40	10	25	40	10	25	40	10	25	40
64	FA	1	1	7	3	7	2	1	2	1	3	2	0	0	0
	FR	37	22	40	39	40	24	25	20	38	34	29	33	30	26
128	FA	2	0	9	3	2	0	1	0	5	1	3	0	0	0
	FR	37	22	42	39	25	22	27	19	46	43	38	23	21	26
256	FA	0	0	4	7	2	1	0	0	4	2	0	0	0	0
	FR	37	20	36	28	28	21	15	15	37	33	29	22	22	22

ตารางที่ ก-33 จำนวนเสียงที่บ่งชี้ผิดพลาดสำหรับชุดทดลองที่ 1 เมื่อใช้ MFCC อันดับ 10 ที่จำนวนสถานะค่าต่างๆ

Codebook Size	t_i	t_{i+df}	t_s			t_{s+df}			t_e			t_{e+df}		
			10	25	40	10	25	40	10	25	40	10	25	40
64	1	1	1	1	1	1	1	1	1	1	1	1	1	1
128	1	1	1	1	1	1	1	1	1	1	1	1	1	1
256	1	1	1	2	1	1	2	1	1	1	1	1	1	1

ตารางที่ ก-34 จำนวนเสียงที่บ่งชี้ผิดพลาดสำหรับชุดทดลองที่ 2 เมื่อใช้ MFCC อันดับ 10 ที่จำนวนสถานะค่าต่างๆ

Codebook Size	t_i	t_{i+df}	t_s			t_{s+df}			t_e			t_{e+df}		
			10	25	40	10	25	40	10	25	40	10	25	40
64	0	0	1	0	1	1	0	1	1	0	0	1	0	0
128	1	1	0	0	1	0	0	1	0	0	0	0	0	0
256	0	0	1	0	1	1	0	1	0	0	0	0	0	0

ตารางที่ ก-35 จำนวน FA FR สำหรับชุดทดลองที่ 1 เมื่อใช้ MFCC อันดับ 10 ที่จำนวนสถานะค่าต่างๆ

Codebook Size		f_1	f_{1+df}	f_2			f_{2+df}			f_3			f_{3+df}		
				10	25	40	10	25	40	10	25	40	10	25	40
64	FA	1	1	1	0	0	0	0	0	0	0	0	0	0	0
	FR	1	0	0	0	0	0	0	0	1	1	1	0	0	0
128	FA	0	0	0	2	2	0	0	0	0	0	0	0	0	0
	FR	2	0	1	0	0	0	0	0	1	1	1	0	0	0
256	FA	1	0	0	0	1	0	0	0	1	1	0	0	0	0
	FR	2	0	2	1	1	1	1	0	1	1	1	1	1	0

ตารางที่ ก-36 จำนวน FA FR สำหรับชุดทดลองที่ 2 เมื่อใช้ MFCC อันดับ 10 ที่จำนวนสถานะค่าต่างๆ

Codebook Size		f_1	f_{1+df}	f_2			f_{2+df}			f_3			f_{3+df}		
				10	25	40	10	25	40	10	25	40	10	25	40
64	FA	2	1	1	3	4	2	1	2	1	2	4	1	2	1
	FR	25	13	18	18	11	8	16	7	17	18	12	11	17	10
128	FA	4	2	3	4	2	3	2	1	1	0	0	0	0	1
	FR	22	15	19	12	11	11	13	13	19	20	19	15	19	16
256	FA	3	1	2	1	2	2	1	0	4	2	1	2	1	0
	FR	22	7	17	13	10	8	10	8	17	13	15	11	12	13

ตารางที่ ก-37 จำนวนเสียงที่บ่งชี้ผิดพลาดสำหรับชุดทดสอบที่ 1 เมื่อใช้ MFCC อันดับ 12 ที่จำนวนสถานะค่าต่างๆ

Codebook Size	t_1	t_{1+df}	t_2			t_{2+df}			t_3			t_{3+df}		
			10	25	40	10	25	40	10	25	40	10	25	40
64	1	1	1	1	1	1	1	1	1	1	1	1	1	1
128	1	1	1	1	1	1	1	1	1	1	1	1	1	1
256	1	1	1	1	1	1	1	1	1	1	1	1	1	1

ตารางที่ ก-38 จำนวนเสียงที่บ่งชี้ผิดพลาดสำหรับชุดทดสอบที่ 2 เมื่อใช้ MFCC อันดับ 12 ที่จำนวนสถานะค่าต่างๆ

Codebook Size	t_1	t_{1+df}	t_2			t_{2+df}			t_3			t_{3+df}		
			10	25	40	10	25	40	10	25	40	10	25	40
64	0	0	2	1	1	2	1	1	1	0	1	1	0	1
128	0	0	1	0	1	1	0	1	0	0	0	0	0	0
256	0	0	0	0	0	0	0	0	0	0	0	0	0	0

ตารางที่ ก-39 จำนวน FA FR สำหรับชุดทดลองที่ 1 เมื่อใช้ MFCC อันดับ 12 ที่จำนวนสถานะค่าต่างๆ

Codebook Size		t_i	t_{i+df}	f_i			f_{i+df}			f_s			f_{s+df}		
				10	25	40	10	25	40	10	25	40	10	25	40
64	FA	0	0	0	1	0	0	0	0	1	2	0	0	0	0
	FR	1	0	0	0	1	0	0	0	0	0	1	0	0	0
128	FA	1	0	2	0	1	0	0	0	1	0	0	1	0	0
	FR	0	0	0	0	0	1	0	0	0	0	1	0	0	0
256	FA	0	0	0	2	0	0	0	0	1	1	0	0	1	0
	FR	1	0	1	0	0	1	1	0	0	0	0	0	0	0

ตารางที่ ก-40 จำนวน FA FR สำหรับชุดทดลองที่ 2 เมื่อใช้ MFCC อันดับ 12 ที่จำนวนสถานะค่าต่างๆ

Codebook Size		t_i	t_{i+df}	f_i			f_{i+df}			f_s			f_{s+df}		
				10	25	40	10	25	40	10	25	40	10	25	40
64	FA	0	0	3	11	1	2	1	3	1	1	2	2	0	1
	FR	20	9	20	10	20	13	13	16	20	17	20	15	19	22
128	FA	3	0	4	1	1	0	0	0	0	0	0	0	0	0
	FR	21	13	8	9	9	11	6	6	16	15	12	16	13	9
256	FA	2	1	3	3	2	0	0	0	1	1	1	0	0	0
	FR	21	12	19	16	17	16	19	15	19	19	19	19	19	19

ตารางที่ ก-41 จำนวนเสียงที่บ่งชี้ผิดพลาดสำหรับชุดทดลองที่ 1 เมื่อใช้ MFCC อันดับ 14 ที่จำนวนสถานะค่าต่างๆ

Codebook Size	t_i	t_{i+df}	t_e			t_{e+df}			t_s			t_{s+df}		
			10	25	40	10	25	40	10	25	40	10	25	40
64	1	1	1	0	0	1	0	0	1	1	1	1	1	1
128	1	1	1	1	0	1	1	0	1	1	1	1	1	1
256	1	1	1	0	0	1	0	0	1	0	0	1	0	0

ตารางที่ ก-42 จำนวนเสียงที่บ่งชี้ผิดพลาดสำหรับชุดทดลองที่ 2 เมื่อใช้ MFCC อันดับ 14 ที่จำนวนสถานะค่าต่างๆ

Codebook Size	t_i	t_{i+df}	t_e			t_{e+df}			t_s			t_{s+df}		
			10	25	40	10	25	40	10	25	40	10	25	40
64	0	0	0	0	0	0	0	0	0	0	0	0	0	0
128	0	0	0	0	1	0	0	1	0	0	0	0	0	0
256	0	0	0	1	0	0	1	0	0	1	0	0	1	0

ตารางที่ ก-43 จำนวน FA FR สำหรับชุดทดลองที่ 1 เมื่อใช้ MFCC อันดับ 14 ที่จำนวนสถานะค่าต่างๆ

Codebook Size		f_1	f_{1+df}	f_2			f_{2+df}			f_3			f_{3+df}			
				10	25	40	10	25	40	10	25	40	10	25	40	
64	FA	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0
	FR	1	0	0	0	1	0	1	1	0	0	0	0	0	0	0
128	FA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	FR	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
256	FA	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
	FR	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0

ตารางที่ ก-44 จำนวน FA FR สำหรับชุดทดลองที่ 2 เมื่อใช้ MFCC อันดับ 14 ที่จำนวนสถานะค่าต่างๆ

Codebook Size		f_1	f_{1+df}	f_2			f_{2+df}			f_3			f_{3+df}			
				10	25	40	10	25	40	10	25	40	10	25	40	
64	FA	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
	FR	19	9	13	7	7	4	6	7	16	13	14	9	8	11	
128	FA	0	0	3	2	3	2	1	0	2	0	1	0	0	0	
	FR	21	10	11	8	7	8	7	6	18	18	17	12	13	12	
256	FA	1	0	6	6	2	1	6	0	5	4	1	0	3	0	
	FR	21	10	17	14	8	17	12	12	19	19	13	16	15	12	

ตารางที่ ก-47 จำนวน FA FR สำหรับชุดทดลองที่ 1 เมื่อใช้ MFCC อันดับ 16 ที่จำนวนสถานะค่าต่างๆ

Codebook Size		f_1	f_{1+df}	f_1			f_{1+df}			f_2			f_{2+df}		
				10	25	40	10	25	40	10	25	40	10	25	40
64	FA	0	0	0	0	0	0	0	0	1	0	0	0	0	0
	FR	1	0	0	0	0	0	0	0	0	0	0	0	0	0
128	FA	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	FR	1	0	1	0	0	0	0	0	0	0	0	0	0	0
256	FA	0	0	1	1	0	0	0	0	0	0	0	0	0	0
	FR	1	0	0	0	0	0	0	0	0	0	0	0	0	0

ตารางที่ ก-48 จำนวน FA FR สำหรับชุดทดลองที่ 2 เมื่อใช้ MFCC อันดับ 16 ที่จำนวนสถานะค่าต่างๆ

Codebook Size		f_1	f_{1+df}	f_1			f_{1+df}			f_2			f_{2+df}		
				10	25	40	10	25	40	10	25	40	10	25	40
64	FA	0	0	4	7	4	0	4	2	1	1	1	0	2	2
	FR	19	5	8	3	4	4	3	4	11	5	11	6	5	6
128	FA	0	0	3	3	0	1	0	0	1	0	0	0	0	0
	FR	19	5	12	6	6	7	5	4	14	12	10	8	7	8
256	FA	1	0	5	4	2	0	2	1	0	2	0	0	0	0
	FR	16	6	12	11	7	13	15	11	14	14	7	13	13	11

ประวัติผู้เขียน

นายพงศ์ไท ทาสระคู เกิดเมื่อวันที่ 14 มิถุนายน พ.ศ. 2519 ที่อำเภอสุวรรณภูมิ จังหวัดร้อยเอ็ด สำเร็จการศึกษาปริญญาวิศวกรรมศาสตรบัณฑิต เกียรตินิยมอันดับสอง ภาควิชาวิศวกรรมไฟฟ้า จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2540 และเข้าศึกษาต่อในหลักสูตรวิศวกรรมศาสตรมหาบัณฑิต แขนงวิชากรรมวิธีสัญญาณดิจิทัล ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2541



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย