การรู้จำเสียงพูดตัวเลขเป็นภาษาไทยแบบไม่ขึ้นกับผู้พูดโดยวิธี ฮิดเดน มาร์คอฟ โมเดล และเวกเตอร์ควอนไตซ์เซชัน

นางสาว เสาวลักษณ์ อารีย์พงศา

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
ภาควิชาวิศวกรรมไฟฟ้า
บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย
พ.ศ. 2538

ISBN 974 - 631 - 907 - 8

ลิขสิทธิ์ของบัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

I16075584

SPEAKER INDEPENDENT THAI NUMERAL SPEECH RECOGNITION BY HIDDEN MARKOV MODEL AND VECTOR QUANTIZATION

Miss Saowaluck Areepongsa

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Engineering
Department of Electical Engineering
Graduate School
Chulalongkorn University
1995

ISBN 974 - 631 - 907 - 8

หัวข้อวิทยานิพนธ์	การรู้จำเสียงพูดตัวเลขเป็นภาษาไทยแบบไม่ขึ้นกับผู้พูดโดยวิธี ฮิดเดน มาร์คอฟ โมเดล และเวกเตอร์ควอนไตซ์เซชั่น
โดย	นางสาว เสาวลักษณ์ อารีย์พงศา
ภาควิชา	วิศวกรรมไฟฟ้า
อาจารย์ที่ปรึกษา	รองศาสตราจารย์ ดร.สมชาย จิตะพันธ์กุล

บัณฑิตวิทยาลัยจุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

.....คณบดีบัณฑิตวิทยาลัย
(รองศาสตราจารย์ ดร.สันติ ถุงสุวรรณ)

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ
(รองศาสตราจารย์ ดร. ณรงค์ อยู่ถนอม)

.....กรรมการ
(อาจารย์ ดร. วาทิต เบญจพลกุล)

.....กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร. สุดาพร ลักษณียนาวิน)

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

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

พิมพ์ต้นฉบับบทคัดย่อวิทยานิพนธ์ภายในกรอบสีเขียวนี้เพียงแผ่นเดียว

เสาวลักษณ์ อารีย์พงศา : การรู้จำเสียงพูดตัวเลขเป็นภาษาไทยแบบไม่ขึ้นกับผู้พูดโดยวิธีฮิดเดน มาร์คอฟ โมเดล และเวกเตอร์ควอนไตซ์เซชัน (SPEAKER INDEPENDENT THAI NUMERAL SPEECH RECOGNITION BY HIDDEN MARKOV MODEL AND VECTOR QUANTIZATION) อ.ที่ปรึกษา : รศ. ดร. สมชาย จิตะพันธ์กุล, 56 หน้า.
ISBN 974 - 631 - 907 - 8

วิทยานิพนธ์นี้เป็นการนำวิธีการของ Hidden Markov Models(HMM) มาใช้ในการรู้จำเสียงตัวเลขภาษาไทยแบบคำโดดไม่ขึ้นกับผู้พูด ปัจจุบันมีการนำ Hidden Markov Models (HMM) ไปใช้ในการรู้จำเสียงพูดภาษาอังกฤษได้เป็นอย่างดี อย่างไรก็ตามวิธีการของ Hidden Markov Models(HMM) ต้องการรูปแบบเพื่อนำมาทำต้นแบบจำนวนมาก งานวิจัยนี้ได้แสดงความสัมพันธ์ระหว่าง ความถูกต้องในการรู้จำเสียงพูดตัวเลขภาษาไทยและจำนวนรูปแบบที่นำมาทำโมเดลต้นแบบ ปรากฏว่า อัตราความถูกต้องในการรู้จำมีค่าเพิ่มขึ้นตามจำนวนรูปแบบ และอัตราการรู้จำเมื่อได้รูปแบบจำนวน 45 ตัวอย่าง จะมีค่าประมาณร้อยละ 84 ซึ่งแสดงให้เห็นว่าเทคนิคนี้มีความเหมาะสมที่จะใช้ในการรู้จำเสียงพูดภาษาไทยได้

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ภาควิชา วิศวกรรมไฟฟ้า ลายมือชื่อนิสิต เสาวลักษณ์ อารีย์พงศา

สาขาวิชา Digital Signal Processing ลายมือชื่ออาจารย์ที่ปรึกษา

ปีการศึกษา 2537 ลายมือชื่ออาจารย์ที่ปรึกษาร่วม

C615660: MAJOR DIGITAL SIGNAL PROCESSING
KEY WORD: RECOGNITION/ LINEAR PREDICTIVE/ VECTOR QUANTIZATION/ HIDDEN MARKOV MODEL
SAOWALUCK AREEPONGSA : SPEAKER INDEPENDENT THAI NUMERAL SPEECH RECOGNITION BY HIDDEN MARKOV MODEL AND VECTOR QUANTIZATION. THESIS ADVISOR : ASSO. PROF. SOMCHAI JITAPUNKUL, Ph.D. 56 pp.
ISBN 974-631-907-8

This thesis proposed to use Hidden Markov Model(HMM) and Vector Quantization to recognize Thai numeral speech based on speaker independent. Recently HMM is widely used for English speech modeling because of its performance. However, its algorithm usually requires sufficient training data. This thesis has also shown the relationship between the accuracy of Thai numeral speech recognition versus the number of training sets. The accuracy rate increase,along with the increment of the training data. The rate of recognition is about 84% for 45 training set which shows that this technique is appropriate for Thai speech recognition.

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ภาควิชา วิศวกรรมไฟฟ้า ลายมือชื่อนิสิต เสาวลักษณ์ อารีพงศ์

สาขาวิชา Digital Signal Processing ลายมือชื่ออาจารย์ที่ปรึกษา

ปีการศึกษา 2537 ลายมือชื่ออาจารย์ที่ปรึกษาร่วม

กิตติกรรมประกาศ

ข้าพเจ้าขอกราบขอบพระคุณ รองศาสตราจารย์ ดร. สมชาย จิตะพันธ์กุล อาจารย์ที่ปรึกษาที่กรุณาให้ความช่วยเหลือทางด้าน ความรู้ คำแนะนำ และตรวจสอบจนวิทยานิพนธ์ฉบับนี้สำเร็จด้วยดี และข้าพเจ้าขอกราบขอบพระคุณ อาจารย์ สุวิทย์ นาคพีระยุทธ ที่ได้ให้คำแนะนำและคำวิจารณ์ที่เป็นประโยชน์เกี่ยวกับการทำวิทยานิพนธ์นี้

ข้าพเจ้าขอขอบคุณทุกท่าน ซึ่งไม่ได้กล่าวถึง ณ ที่นี้ ที่มีส่วนช่วยในการให้คำแนะนำ กำลังใจ อุปกรณ์ต่าง ๆ เสียงที่ใช้เป็นต้นแบบและทดสอบ ซึ่งทำให้วิทยานิพนธ์ฉบับนี้สำเร็จไปได้ดี

สุดท้ายนี้ข้าพเจ้าขอขอบพระคุณครอบครัวของข้าพเจ้าที่ให้โอกาสและกำลังใจ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญ

หน้า

บทคัดย่อภาษาไทย..... ง
บทคัดย่อภาษาอังกฤษ..... จ
กิตติกรรมประกาศ..... ฉ
สารบัญตาราง..... ช
สารบัญรูปประกอบ..... ฌ
คำศัพท์..... ญ

บทที่ 1 บทนำ..... 1
บทที่ 2 รูปแบบการรู้จำเสียง..... 3
FEATURE MEASUREMENT..... 3
1. การประมาณพันธะเชิงเส้น (Linear Predictive Coding)..... 3
2. Vector Quantization(VQ)..... 11
PATTERN SIMILARITY DETERMINATION..... 13
1. Hidden Markov Model (HMM)..... 13
DECISION RULE..... 19
1. Viterbi Algorithm..... 19
บทที่ 3 การทดลอง..... 23
3.1 การประมาณพันธะเชิงเส้น..... 23
3.2 Vector Quantization..... 24
3.3 Hidden Markov Model (HMM)..... 28
3.4 Viterbi Algorithm..... 31
3.5 ขั้นตอนการทดลอง..... 32
บทที่ 4 ผลการทดลอง..... 33
บทที่ 5 สรุปและวิจารณ์..... 42
สรุป..... 42
วิจารณ์..... 42
สิ่งที่ควรพัฒนา..... 43
รายการอ้างอิง..... 44
ภาคผนวก ก. 46
ภาคผนวก ข. 54
ประวัติผู้เขียน..... 56

สารบัญตาราง

ตารางที่ หน้า

4.1 เป็นการแสดงเปอร์เซ็นต์ความถูกต้องของกลุ่มทดสอบทั้ง 2 กลุ่ม..... 33
4.2 แสดงผลการทดสอบของแต่ละบุคคลที่เป็นบุคคลต้นแบบ..... 34
4.3 แสดงผลการทดสอบเสียงของบุคคลกลุ่มทดสอบ..... 34
4.4 แสดงการรู้จำของกลุ่มที่ 1..... 35
4.5 แสดงผลการรู้จำของกลุ่มที่ 2..... 36
4.6 แสดงผลการรู้จำของเซ็ตที่เป็นบุคคลกลุ่มทดสอบ..... 37
4.7 แสดง state และค่าความน่าจะเป็นแต่ละค่าของ Vector Quantization เมื่อเทียบกับโมเดล 0.....38
4.8 แสดง state และค่าความน่าจะเป็นแต่ละค่าของ Vector Quantization เมื่อเทียบกับโมเดล 2.....39
4.9 เปรียบเทียบความถูกต้องเป็นร้อยละกับจำนวนต้นแบบ..... 40
ก.1 เปรียบเทียบความผิดพลาดเป็นร้อยละกับขนาด codebook ของกลุ่มต้นแบบ..... 49
ก.2 เปรียบเทียบความผิดพลาดเป็นร้อยละกับขนาด codebook ของกลุ่มทดสอบ..... 49
ข.1 เป็นการเปรียบเทียบเปอร์เซ็นต์ความถูกต้องของการทดสอบเสียง 0-9 ของบุคคล 100 คน(บุคคลที่เป็นต้นแบบ)..... 54
ข.2 เป็นการเปรียบเทียบเปอร์เซ็นต์ความถูกต้องของการทดสอบเสียง 0-9 ของบุคคล 10 คน(บุคคลที่ไม่เป็นต้นแบบ).....55

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญรูปประกอบ

รูปที่		หน้า
2.1	โมเดลการรู้จำรูปแบบ	3
2.2	แสดงแบบจำลองของระบบกำเนิดเสียงพูดแบบดิจิตอล	4
2.3	การประมาณพันธะเชิงเส้น	7
2.4	ขั้นตอนการทำงานของ VQ	11
2.5	โครงสร้างแบบต่าง ๆ	14
2.6	แสดงการพิจารณา forward variable	16
2.7	แสดงการพิจารณา backward variable	17
2.8	การพิจารณา forward และ backward	18
2.9	โมเดลของ Viterbi	20
2.10	โครงสร้าง state และ trellis ของ Viterbi	20
2.11	การหาระยะที่สั้นที่สุดโดยวิธี Viterbi	22
3.1	ขั้นตอนการหาโมเดลเสียง	23
3.2	ขั้นตอนการทำงานของ LPC	24
3.3	แสดงขั้นตอนการประมาณพันธะเชิงเส้นด้วยวิธีอัตสัมพันธ์	26
3.4	ขั้นตอนการทำงานของ VQ	27
3.5	โครงสร้างของ HMM	28
3.6	Lattice diagram ของ HMM	29
3.7	ขั้นตอนในการหาโมเดลและทดสอบ	31
4.1	เปรียบเทียบเปอร์เซ็นต์ความถูกต้องกับจำนวนต้นแบบ	41
ก.1	เปรียบเทียบค่า residual ของสัญญาณ	47
ก.2	เปรียบเทียบค่าความคลาดเคลื่อนเฉลี่ยกับขนาด codebook	48
ก.3	เปรียบเทียบความผิดพลาดเป็นร้อยละกับขนาด codebook ของกลุ่มต้นแบบ	49
ก.4	เปรียบเทียบความผิดพลาดเป็นร้อยละกับขนาด codebook ของกลุ่มทดสอบ	50
ก.5	โครงสร้างที่ทดสอบ	51
ก.6	โครงสร้างแบบ serial	52
ก.7	เปรียบเทียบความคลาดเคลื่อนกับจำนวน state	52
ก.8	เปรียบเทียบความคลาดเคลื่อนของแต่ละเสียงตัวเลข(ภาษาอังกฤษ)	53

คำศัพท์

การบีบอัดข้อมูล	compression
การประมาณพันธะเชิงเส้น	Linear Predictive Coding(LPC)
การประมาณค่าแบบก้าวหน้า	forward prediction
การประมาณค่าแบบถดถอย	backward prediction
การสุ่มค่า	random
ความคลาดเคลื่อน	distortion
ค่าความคลาดเคลื่อนกำลังสองรวม	total square error
ค่าปรากฏ	Observation
ความสัมพันธ์เชิงพิกัดฉาก	orthogonal relationship
ค่าสัมประสิทธิ์การสะท้อน	reflection coefficient
ฟังก์ชันพหุคูณ	polynomial function
ต่อเนื่อง	continuous
ไม่ต่อเนื่อง	discrete
ผลคูณภายใน	inner product
ผลรวมเชิงเส้น	linear combination
วิธีการแปรปรวนร่วม	covariance method
วิธีการอัตสัมพันธ์	autocorrelation method
วิธีอัตสัมพันธ์ส่วนย่อย	PARCOR(Partial Autocorrelation)
เส้นเชื่อม	path
อนุพันธ์ส่วนย่อย	partial derivation

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย