การรู้จำเสียงพูดไทยโดยตรงจากการเข้ารหัส G.729

นายสิริ วงศ์วรชาติกาล

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมไฟฟ้า ภาควิชาวิศวกรรมไฟฟ้า
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2543

ISBN 974-13-0111-1

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

I19664898

27 มี.ค. 2545

DIRECT RECOGNITION OF THAI SPEECH FROM G.729 CODE

Mr. Siri Wongworachatkan

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

Chulalongkorn University

Academic year 2000

ISBN 974-13-0111-1

หัวข้อวิทยานิพนธ์	การรู้จำเสียงพูดไทยโดยตรงจากการเข้ารหัส G.729
โดย	นายสิริ วงศ์วรชาติกาล
ภาควิชา	วิศวกรรมไฟฟ้า
อาจารย์ที่ปรึกษา	อาจารย์ สุวิทย์ นาคพีระยุทธ

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

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

คณะกรรมการสอบวิทยานิพนธ์

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

..... อาจารย์ที่ปรึกษา
(อาจารย์ สุวิทย์ นาคพีระยุทธ)

..... กรรมการ
(รองศาสตราจารย์ ดร.เอกชัย ลีลารัศมี)

สิริ วงศ์วรชาติกาล : การรู้จำเสียงพูดไทยโดยตรงจากการเข้ารหัส G.729. (DIRECT RECOGNITION OF THAI SPEECH FROM G.729 CODE) อ. ที่ปรึกษา : อ.สุวิทย์ นาคพีระยุทธ, 70 หน้า. ISBN 974-13-0111-1.

มาตรฐาน ITU-T G.729 เป็นมาตรฐานในการบีบอัดเสียงพูดซึ่งสามารถนำใช้งานได้อย่างกว้างขวาง ดังนั้นถ้าเราสามารถดึงจุดเด่นของเสียงที่จำเป็นในการรู้จำออกมาได้โดยตรงจากรหัสเสียงที่ถูกบีบอัดแล้ว จะสามารถสร้างระบบรู้จำเสียงอย่างง่ายจากรหัสเสียง G.729 โดยตรง พลังงานเสียง คาบการสั่นของเสียง และ LSP (Line Spectral Pair) เป็นพารามิเตอร์ที่ส่งมากับรหัส G.729 และสามารถใช้ในการรู้จำเสียงได้ วิทยานิพนธ์นี้นำวิธีการของแบบจำลองฮิดเดน มาร์คอฟ และการควอนไทซ์แบบเวกเตอร์ มาใช้ในการรู้จำเสียงภาษาไทยแบบไม่ขึ้นกับผู้พูด คำศัพท์ทั้งหมด 30 คำแบ่งเป็น 2 ชุดได้แก่ ชุดคำศัพท์ตัวเลข 0 ถึง 9 และชุดคำศัพท์พยางค์เดียว 20 คำ เสียงพูดที่นำมาเป็นต้นแบบและเป็นเสียงพูดทดสอบประกอบด้วยทั้งเพศชายและหญิงที่มีช่วงอายุระหว่าง 18 ปีถึง 25 ปี

ผลการทดสอบอัตราการรู้จำแบบไม่ขึ้นกับผู้พูดของชุดเสียงพูดเพื่อทดสอบมีอัตรารู้จำเฉลี่ยร้อยละ 90.75 โดยมีอัตราการรู้จำเฉพาะชุดคำศัพท์พยางค์เดียวร้อยละ 88.50 อัตราการรู้จำเฉพาะชุดตัวเลขร้อยละ 93.00 ตามลำดับ

ภาควิชา วิศวกรรมไฟฟ้า ลายมือชื่อนิสิต สิริ วงศ์วรชาติกาล

สาขาวิชา วิศวกรรมไฟฟ้า ลายมือชื่ออาจารย์ที่ปรึกษา

ปีการศึกษา 2544

4170587821 : MAJOR ELECTRICAL ENGINEERING

KEY WORD: SPEECH RECOGNITION / G.729 / LINE SPECTRAL PAIR / PITCH / HIDDEN MARKOV MODEL

SIRI WONGWORACHATKAN : DIRECT RECOGNITION OF THAI SPEECH FROM G.729 CODE

THESIS ADVISOR : SUVIT NAKPEERAYUTH, 70 pp. ISBN 974-13-0111-1

The ITU-T Recommendation G.729 is a versatile and well accepted speech compression standard. If the speech feature can be extracted directly from the code easily, a simple speech recognition system can work directly on the G.729 codes. Energy, pitch period and LSP are the parameters obtained from G.729 codes which can be used in speech recognition. This thesis uses Hidden Markov Model (HMM) and Vector Quantization to recognize speaker independent Thai speech. The 30-word vocabulary is subdivided into two sets comprising 20 single syllable, and 10 thai numeric words, zero to nine. The separated speech training set and testing set are composed of both male and female speakers within the range of 18 to 25 years of age.

The average recognition rate of this speaker-independent recognition system is 90.75 %. The recognition rate of the single-syllabled words is 88.50 %.The recognition rate of the numeric words is 93.00%.

Department.....Electrical Engineering... Student's signature .....

Field of study.....Electrical Engineering... Advisor's signature.....

Academic year .....2000.....

กิตติกรรมประกาศ

ข้าพเจ้าขอขอบคุณ อ.สุวิทย์ นาคพีระยุทธ อาจารย์ที่ปรึกษาที่ให้คำปรึกษา แนะนำในการทำวิจัย

ข้าพเจ้าขอขอบคุณ นายพงษ์ไท ทาสระคู ที่สละเวลาให้คำปรึกษา และช่วยเหลือในด้านฐานข้อมูลของเสียงเพื่อใช้ในการรู้จำ

ข้าพเจ้าขอขอบคุณ นายวิศรุต อาขุบุตร ที่ได้ทำวิทยานิพนธ์เรื่อง ระบบรู้จำคำไทยหลายพยางค์แบบไม่ขึ้นกับผู้พูดโดยใช้แบบจำลองฮิดเดนมาร์คอฟ ที่ข้าพเจ้าใช้เปรียบเทียบผล

สุดท้ายนี้ขอขอบคุณทุกท่าน ที่มิได้กล่าวถึงในที่นี้ ที่มีส่วนช่วยในทุก ๆ ด้าน

สารบัญ

บทที่ **หน้า**

บทคัดย่อภาษาไทย .....ง
บทคัดย่อภาษาอังกฤษ.....จ
กิตติกรรมประกาศ .....ฉ
สารบัญตาราง .....ฌ
สารบัญรูปประกอบ.....ญ
สารบัญคำศัพท์ .....ฏ
บทที่ 1 บทนำ 1
1.1 ความเป็นมา 1
1.2 ปัญหาของการรู้จำเสียงพูด.....1
1.3 วัตถุประสงค์ของการวิจัย 2
1.4 ขอบเขตของการวิจัย 3
1.5 ประโยชน์ที่คาดว่าจะได้รับ.....3
บทที่ 2 แนวคิดและทฤษฎีที่เกี่ยวข้อง 4
2.1 การรู้จำเสียง 4
2.1.1 แบบจำลองการสร้างเสียงพูด.....4
2.1.2 การวิเคราะห์และวัดค่าลักษณะสำคัญ.....5
2.1.3 สัมประสิทธิ์ของการประมาณพันธะเชิงเส้น.....6
2.1.4 Line Spectrum Pair (LSP) 7
2.1.5 การทดสอบความคล้ายคลึงกันของรูปแบบ.....11
2.1.6 แบบจำลองฮิดเดน มาร์คอฟ.....12
2.2 มาตรฐานการเข้ารหัส G.729.....16
2.2.1 ในส่วนของฟิลเตอร์ 17
2.2.2 ในส่วนของการกระตุ้น (excitation) 21
บทที่ 3 วิธีดำเนินการวิจัย 26
3.1 การกำหนดชุดคำศัพท์ 26
3.2 ลักษณะสำคัญของเสียงที่นำมาใช้ 26
3.2.1 คาบการสั่นของเสียง.....26

3.2.2 พลังงานเสียง.....31
3.2.3 ลักษณะสำคัญด้านความถี่.....37
3.3 ขั้นตอนการรู้จำคำพูด.....38
3.3.1 ขั้นตอนการประมวลผลสัญญาณเบื้องต้น.....40
3.3.2 ขั้นตอนการฝึกฝนระบบรู้จำคำพูด.....40
3.3.3 ขั้นตอนการทดสอบระบบรู้จำคำพูด.....42
บทที่ 4 ผลการวิจัย.....43
4.1 ผลการรู้จำคำพูดภาษาไทย.....43
4.2 วิเคราะห์การรู้จำคำพูดไทย.....49
บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ.....50
5.1 สรุปผลการวิจัย.....51
5.2 ข้อเสนอแนะ.....52
รายการอ้างอิง.....53
ภาคผนวก.....55
ประวัติผู้เขียน.....70

สารบัญตาราง

ตาราง **หน้า**

ตารางที่ 2.1 เครื่องหมายและตำแหน่งของพัลส์ 22
ตารางที่ 2.2 จำนวนบิตที่ใช้ในการเก็บพารามิเตอร์ต่าง ๆ ของมาตรฐาน G.729 25
ตารางที่ 4.1 ผลอัตราการรู้จำ 43
ตารางที่ 4.2 รายละเอียดค่าพารามิเตอร์สำหรับการรู้จำวิธีที่ 1 ถึง วิธีที่ 3และเวลาคำนวณ
เพื่อรู้จำต่อหนึ่งเสียง.....44
ตารางที่ 4.3 รายละเอียดค่าพารามิเตอร์สำหรับการรู้จำวิธีที่4
และเวลาคำนวณเพื่อรู้จำต่อหนึ่งเสียง.....44
ตารางที่ 4.4 ผลการรู้จำจากวิธีที่ 1 ใช้ LSF โดยตรงเป็นลักษณะสำคัญ 45
ตารางที่ 4.5 ผลการรู้จำจากวิธีที่ 2 ใช้ LSF โดยประมาณเป็นลักษณะสำคัญ.....46
ตารางที่ 4.6 ผลการรู้จำจากวิธีที่ 3 ใช้ LSF โดยตรงผ่านตัวกรองมัธยฐานเป็นลักษณะสำคัญ.....47
ตารางที่ 4.7 ผลการรู้จำจากวิธีที่ 4 ใช้ดัชนีของชุดรหัส G.729 โดยตรง 48

สารบัญภาพ

หน้า

รูปที่ 2.1 แบบจำลองการกำเนิดเสียงพูด 4
รูปที่ 2.2 สัมประสิทธิ์ LP และสัมประสิทธิ์ LSP บนระนาบ Z ของเสียงคำว่า /U/ 10
รูปที่ 2.3 เปรียบเทียบสเปคโตรแกรมและความถี่ LSP ของคำว่า 'นาฬิกา'.....11
รูปที่ 2.4 แบบจำลองฮิดเดน มาร์คอฟที่มี N=5 14
รูปที่ 2.5 แบบจำลองHMM แบบเออร์กอดิกที่มี 4 สถานะ 15
รูปที่ 2.6 แบบจำลองHMMจากซ้ายไปขวาที่มี 5 สถานะ 15
รูปที่ 2.7 แบบจำลองHMM แบบขนานจากซ้ายไปขวา ที่มี 6 สถานะ 16
รูปที่ 2.8 แบบจำลองการเกิดเสียงอย่างหยาบ 16
รูปที่ 2.9 การวางตัวของซีโรของ $F_1'(z)$ กับ $F_2'(z)$ 19
รูปที่ 3.1 สัญญาณเสียงและคาบการสั่นของเสียงของคำว่า "สอง".....27
รูปที่ 3.2 ขั้นตอนในการหาคาบการสั่นของเสียงในสับเฟรมแรก 28
รูปที่ 3.3 ขั้นตอนในการหาคาบการสั่นของเสียงในสับเฟรมหลัง.....29
รูปที่ 3.4 คาบการสั่นของเสียงจริงเทียบกับค่าที่ได้จากข้อมูล P1 และ P2 โดยตรง.....30
รูปที่ 3.5 พลังงานเสียงจริงเทียบกับประมาณจากการบวกแบบ เฉลี่ยทุกสับเฟรม.....33
รูปที่ 3.6 พลังงานเสียงจริงเทียบกับประมาณจากการบวกแบบ เฉลี่ยและคิด LPC 34
รูปที่ 3.7 พลังงานเสียงจริงเทียบกับประมาณจากการบวกแบบเฉลี่ยทุกสับเฟรม
และคิด LSF พจน์คี่ 36
รูปที่ 3.8 การประมาณความถี่ฟอร์แมนท์ของคำว่า 'กิน' เมื่อเทียบกับสเปกโตรแกรม 38
รูปที่ 3.9 รายละเอียดขั้นตอนการรู้จำคำพูดตามขั้นตอนทั่วไป.....39

สารบัญคำศัพท์

หลอดเสียง	acoustic tube
ชุดรหัสที่ปรับเปลี่ยนได้	adaptive codebook
อัตราขยายของชุดรหัสที่ปรับเปลี่ยนได้	adaptive codebook gain, pitch gain
สัมประสิทธิ์อัตสหสัมพันธ์	autocorrelation coefficient
กระบวนการย้อนกลับ	backward procedure
เวกเตอร์ของชุดรหัส	codebook vector
โปรแกรมแบบพลวัต	dynamic programming
การหาจุดสิ้นสุดเสียงพูด	endpoint detection
การกระตุ้น	excitation
ดึงลักษณะสำคัญของเสียง	feature extraction
วัดค่าลักษณะสำคัญ	feature measurement
ชุดรหัสคงที่	fixed codebook
อัตราการขยายของชุดรหัสคงที่	fixed codebook gain
กระบวนการไปหน้า	forward procedure
ความถี่มูลฐาน	fundamental frequency
แบบจำลองฮิดเดน มาร์คอฟ	Hidden Markov Models ,HMM
วิเคราะห์การประมาณพันธะเชิงเส้น	linear prediction analysis
สัมประสิทธิ์ของการประมาณพันธะเชิงเส้น	linear prediction coefficient (LPC)
วิธีกำลังสองน้อยที่สุด	method of least square
เสียงโฆษะแบบผสม	mixed voiced
เครือข่ายประสาท	neural networks
คาบการสั่นของเสียง	pitch period
สัมประสิทธิ์การสะท้อน	reflection coefficient
การรู้จำเสียงพูด	speech recognition
การเข้าคู่ต้นแบบ	template matching
เสียงอโฆษะ	unvoiced
ช่องเสียง	vocal tract
เสียงโฆษะ	voiced