

การแบ่งเสียงพูดเป็นเซกเมนต์สำหรับการรู้จำเสียงพูดภาษาไทยแบบอาศัยเซกเมนต์
โดยใช้สารสนเทศสวณศาสตร์



นายไพโรจน์ ลีลาภทรกิจ

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2549
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

SPEECH SEGMENTATION FOR THAI SEGMENT-BASED
SPEECH RECOGNITION USING ACOUSTIC-PHONETIC INFORMATION

Mr. Pairote Leelaphattarakij

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

Department of Computer Engineering

Faculty of Engineering

Chulalongkorn University

Academic Year 2006

Copyright of Chulalongkorn University

490450

หัวข้อวิทยานิพนธ์

การแบ่งเสียงพูดเป็นเซกเมนต์สำหรับการรู้จำเสียงพูดภาษาไทย
แบบอาศัยเซกเมนต์โดยใช้สารสนเทศสัทศาสตร์

โดย

นายไพโรจน์ ลีลาภักทรกิจ

สาขาวิชา

วิศวกรรมคอมพิวเตอร์

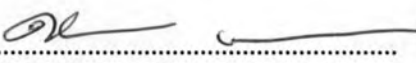
อาจารย์ที่ปรึกษา

อาจารย์ ดร.อดิวงค์ สุชาโต

อาจารย์ที่ปรึกษาร่วม


อาจารย์ ดร. โปรคปราน บุญยพุกกณะ

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้หัวข้อวิทยานิพนธ์ฉบับนี้
เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรบัณฑิต

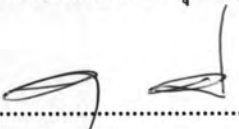

..... คณบดีคณะวิศวกรรมศาสตร์
(ศาสตราจารย์ ดร.ดิเรก ลาวัณย์ศิริ)


คณะกรรมการสอบวิทยานิพนธ์


..... ประธานกรรมการ
(รองศาสตราจารย์ ดร.บุญเสริม กิจศิริกุล)


..... อาจารย์ที่ปรึกษา
(อาจารย์ ดร.อดิวงค์ สุชาโต)



..... อาจารย์ที่ปรึกษาร่วม
(อาจารย์ ดร. โปรคปราน บุญยพุกกณะ)


..... กรรมการ
(อาจารย์ ดร.พิชญ์ คนองชัยยศ)


..... กรรมการ
(ดร.ชัย วุฒิวิวัฒน์ชัย)

ไพโรจน์ ลีลาภีทรกิจ : การแบ่งเสียงพูดเป็นเซกเมนต์สำหรับการรู้จำเสียงพูดภาษาไทยแบบอาศัยเซกเมนต์โดยใช้สารสนเทศสวณศาสตร์. (SPEECH SEGMENTATION FOR THAI SEGMENT-BASED SPEECH RECOGNITION USING ACOUSTIC-PHONETIC INFORMATION) อ. ที่ปรึกษา : อ.ดร.อดิวงส์ สุชาโต, อ.ที่ปรึกษาร่วม : อ.ดร.โปรดปราน บุญยพุกกณะ, 98 หน้า.

ในระบบรู้จำเสียงพูดแบบอาศัยเซกเมนต์ จะต้องแบ่งเสียงพูดออกเป็นเซกเมนต์ โดยมีวัตถุประสงค์เพื่อค้นหาขอบเขตของหน่วยเสียงหรือตำแหน่งบอกเวลาเริ่มต้นและสิ้นสุดของหน่วยเสียง แล้วนำไปสร้างเป็นกราฟของเซกเมนต์ ซึ่งจะถูกใช้เป็นข้อมูลขาเข้าของขั้นตอนการรู้จำเสียงพูด เพื่อค้นหาลำดับของหน่วยเสียงที่ดีที่สุดออกมา เป้าหมายของวิทยานิพนธ์นี้ ต้องการพัฒนาวิธีการแบ่งเสียงพูดเป็นเซกเมนต์ที่มีประสิทธิภาพและสามารถทำงานได้อย่างรวดเร็ว เพื่อนำไปใช้ในระบบรู้จำเสียงพูดแบบอาศัยเซกเมนต์ วิธีการแบ่งเสียงพูดเป็นเซกเมนต์แบบเดิมนั้นจะอาศัยเครื่องรู้จำเสียงพูดในระดับหน่วยเสียงมาค้นหาขอบเขตของหน่วยเสียง แต่เนื่องจากประสิทธิภาพนั้นยังห่างไกลกับประสิทธิภาพของวิธีการแบ่งเสียงพูดเป็นเซกเมนต์ด้วยคน อีกทั้งกราฟของเซกเมนต์ที่ได้มีขนาดใหญ่และใช้เวลาในการทำงานนาน จึงไม่เหมาะกับระบบรู้จำเสียงพูดที่ต้องการความรวดเร็ว วิทยานิพนธ์นี้จึงเสนอวิธีการแบ่งเสียงพูดเป็นเซกเมนต์ที่มีประสิทธิภาพดีกว่า โดยมีขั้นตอนการทำงานสองขั้นตอน คือ ขั้นตอนการหาขอบเขตของหน่วยเสียงจากตำแหน่งที่มีการเปลี่ยนแปลงลักษณะการออกเสียง โดยอาศัยลักษณะสำคัญของเสียงที่ได้จากการใช้สารสนเทศสวณศาสตร์และอาศัยซอฟต์แวร์เดเทคเตอร์แมชชีนมาจำแนกเสียงพูดตามลักษณะการออกเสียง ขั้นตอนต่อมาจะสร้างกราฟของเซกเมนต์โดยใช้วิธีการสร้างกราฟแบบหลายระดับ รวมถึงมีการคิดคะแนนให้กับขอบเขตของหน่วยเสียงที่หาได้จากค่าการเปลี่ยนแปลงสเปกตรัม การทดลองทั้งหมดจะทดสอบโดยใช้ฐานข้อมูลเสียงพูดภาษาไทย โดยเมื่อยอมให้ขอบเขตของหน่วยเสียงที่หามาได้คลาดเคลื่อนไปจากขอบเขตของหน่วยเสียงอ้างอิงได้ไม่เกิน 20 มิลลิวินาที วิธีการแบ่งเสียงพูดเป็นเซกเมนต์นี้จะสามารถค้นหาขอบเขตของหน่วยเสียงได้ความแม่นยำและความครอบคลุมเพิ่มขึ้น 8.3% (จาก 68.0% เป็น 76.3%) และ 5.1% (จาก 82.1% เป็น 87.2%) ตามลำดับ และสามารถลดขนาดกราฟของเซกเมนต์ได้ประมาณ 14 เท่าโดยที่ยังรักษาระดับความครอบคลุมไว้ได้ที่ 77.4% เมื่อเปรียบเทียบกับวิธีการแบ่งเสียงพูดเป็นเซกเมนต์แบบอาศัยเครื่องรู้จำเสียงพูด

ภาควิชา.....วิศวกรรมคอมพิวเตอร์	ลายมือชื่อนิสิต.....ไพโรจน์.....
สาขาวิชา.....วิศวกรรมคอมพิวเตอร์	ลายมือชื่ออาจารย์ที่ปรึกษา.....อดิวงส์.....
ปีการศึกษา.....2549	ลายมือชื่ออาจารย์ที่ปรึกษาร่วม..... 

497 04919 21 : MAJOR COMPUTER ENGINEERING

KEY WORD: AUTOMATIC SPEECH RECOGNITION / SPEECH SEGMENTATION / SEGMENT-BASED SPEECH RECOGNITION

PAIROTE LEELAPHATTARAKIJ : SPEECH SEGMENTATION FOR THAI SEGMENT-BASED SPEECH RECOGNITION USING ACOUSTIC-PHONETIC INFORMATION. THESIS ADVISOR : ATIWONG SUCHATO, Ph.D., THESIS COADVISOR : PROADPRAN PUNYABUKKANA, Ph.D., 98 pp.

Segment-based speech recognition systems must explicitly hypothesize segment start and end times. The purpose of a segmentation algorithm is to hypothesize those times and to compose a graph of segments from them. During recognition, this graph is an input to a search that finds the optimal sequence of sound units through the graph. The goal of this thesis is to create a high-quality, real-time phonetic segmentation algorithm for segment-based speech recognition. The baseline algorithm makes use of frame-based phonetic recognizer to hypothesize possible phonetic segments but its performance was still far from human's ability to perform such a task. This thesis addresses the quality and computational requirements by employing more efficient phonetic segmentation algorithm, and by shrinking the search space. The algorithm is done in two stages. Boundaries are detected in the first stage via manner-of-articulation changes by using acoustic features obtained from acoustic-phonetic information and applying multiple Support Vector Machines for the classification of manner features. Multi-Level Segmentation is used to compose a graph from the boundary list for the graph size reduction. In addition, it includes a landmark scoring by utilizing the spectral transition measurement. Experiments reported were done on Thai continuous speech corpus. Allowing at most 20 ms. deviation from the actual boundaries, the algorithm detects boundaries that have over 8.3% and 5.1% improvement in precision (from 68.0% to 76.3%) and recall rate (from 82.1% to 87.2%) and produces a segment-graph that has over 14 times fewer segments while still maintaining a 77.4% in recall rate over a baseline speech segmentation algorithm.

DepartmentComputer Engineering...
 Field of studyComputer Engineering...
 Academic year2006

Student's signature
 Advisor's signature
 Co-advisor's signature

กิตติกรรมประกาศ

กิตติกรรมประกาศนี้ของอุทิศให้กับผู้ที่มีส่วนให้การช่วยเหลือจนสามารถข้ามผ่านปัญหาและอุปสรรคต่างๆในการทำวิทยานิพนธ์นี้ไปได้ด้วยดี

วิทยานิพนธ์นี้สำเร็จด้วยดีเพราะแรงบัลดาลใจจาก อ.ดร.อดิวงค์ สุชาติ และ อ.ดร.โปรดปราน บุญยทุกณะ ที่เป็นผู้ชี้แนะแนวทาง และจุดประกายให้เกิดความสนใจและความหลงใหลในเทคโนโลยีเสียงพูด นอกจากนี้ อยากรขอบพระคุณเป็นพิเศษ สำหรับคณะกรรมการสอบวิทยานิพนธ์ ได้แก่ รศ. ดร. บุญเสริม กิจศิริกุล ผู้ซึ่งเป็นประธาน อ. ดร.พิชญ์ คนองชัยยศ และ ดร. ชัย วุฒิววัฒน์ชัย ที่สละเวลาอันมีค่ามาชี้ให้เห็นถึงข้อบกพร่อง รวมทั้งข้อแนะนำที่น่าสนใจ

ขอขอบคุณสมาชิกของห้องปฏิบัติการ SLS ทุกคน รวมทั้งพี่ ๆ และเพื่อน ๆ ระดับบัณฑิตศึกษาที่สร้างบรรยากาศที่ดีในการทำงาน

ขอขอบคุณจุฬาลงกรณ์มหาวิทยาลัย และโรงเรียนเทพศิรินทร์ ที่ได้ประสิทธิประสาทวิชาความรู้

สุดท้ายนี้ขอกราบขอบพระคุณ คุณพ่อ คุณแม่ และญาติ ๆ ที่ได้ผลักดันจนผู้เขียนประสบความสำเร็จอย่างเช่นทุกวันนี้

งานวิจัยนี้ได้รับเงินทุนสนับสนุนจากคณะวิศวกรรมศาสตร์ในโครงการจัดการศึกษาสาขาวิชาวิศวกรรมศาสตร์ เพื่อเพิ่มศักยภาพทางด้านวิทยาศาสตร์เทคโนโลยี และอุตสาหกรรม หมวดเงินอุดหนุนการศึกษาประจำปีการศึกษา 2549

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ.....	ช
สารบัญภาพ.....	ฌ
บทที่ 1 บทนำ.....	1
ความเป็นมาและความสำคัญของปัญหา.....	1
วัตถุประสงค์ของการวิจัย.....	2
ขอบเขตของการวิจัย.....	2
ขั้นตอนการวิจัย.....	2
ประโยชน์ที่คาดว่าจะได้รับ.....	3
ผลงานที่ตีพิมพ์จากวิทยานิพนธ์.....	3
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง.....	4
ทฤษฎีที่เกี่ยวข้อง.....	4
1. สรีรศาสตร์.....	4
2. สวนศาสตร์.....	15
3. การรู้จำเสียงพูด.....	19
4. การแปลงฟูรีเยร์แบบวิซูด.....	21
5. สเปกโตรแกรมของเสียงพูด.....	23
6. ขอบเขตของหน่วยเสียง.....	26
7. การสกัดลักษณะสำคัญ.....	27
8. แบบจำลองฮิดเดนมาร์คอฟ.....	29
9. การรู้จำเสียงพูดแบบอาศัยเซกเมนต์.....	31
10. ซัพพอร์ตเวกเตอร์แมชชีน – เอสวีเอ็ม [15].....	35
งานวิจัยที่เกี่ยวข้อง.....	44
1. การแบ่งเสียงพูดเป็นเซกเมนต์แบบอาศัยการจำแนกเสียงพูดเป็นประเภทกว้าง.....	44
2. การแบ่งเสียงพูดเป็นเซกเมนต์จากการเปลี่ยนแปลงทางสัญญาณเสียง.....	44
3. การแบ่งเสียงพูดเป็นเซกเมนต์แบบอาศัยเครื่องรู้จำเสียงพูด.....	45
บทที่ 3 การแบ่งเสียงพูดเป็นเซกเมนต์โดยใช้สารสนเทศสวณศาสตร์.....	46

ภาพรวมของการแบ่งเสียงพูดเป็นเซกเมนต์สำหรับระบบรู้จำเสียงพูดแบบอาศัยเซกเมนต์	46
การตรวจหาขอบเขตของหน่วยเสียง	47
1. การสกัดลักษณะสำคัญของเสียง	48
2. การเรียนรู้การจำแนกลักษณะการออกเสียง	55
3. การตรวจหาขอบเขตของหน่วยเสียงจากผลการจำแนกลักษณะการออกเสียง	57
การสร้างกราฟของเซกเมนต์	59
1. การสร้างกราฟของเซกเมนต์แบบเชื่อมต่อกับขอบเขตของหน่วยเสียง	59
2. การสร้างกราฟของเซกเมนต์แบบอาศัยการเปลี่ยนแปลงสเปกตรัม	60
3. การสร้างกราฟของเซกเมนต์แบบหลายระดับ	61
การให้คะแนนขอบเขตของหน่วยเสียง	63
บทที่ 4 การทดลองแบ่งเสียงพูดเป็นเซกเมนต์	66
ฐานข้อมูลเสียงเพื่อการแบ่งเสียงพูดเป็นเซกเมนต์	66
องค์ประกอบและประสิทธิภาพของเครื่องรู้จำเสียงพูด	67
การทดลองและผลการทดลอง	69
1. การทดลองเพื่อวัดประสิทธิภาพการจำแนกลักษณะการออกเสียง	69
2. การทดลองเพื่อเปรียบเทียบประสิทธิภาพของการตรวจหาขอบเขตของหน่วยเสียง	71
3. การทดลองเพื่อเปรียบเทียบประสิทธิภาพการสร้างกราฟของเซกเมนต์	79
สรุปผลการทดลอง	85
บทที่ 5 สรุปผลการวิจัย	87
ผลการวิจัย	87
1. การตรวจหาขอบเขตของหน่วยเสียง	87
2. การสร้างกราฟของเซกเมนต์	88
ข้อเสนอแนะ	90
รายการอ้างอิง	91
ภาคผนวก	94
ประวัติผู้เขียนวิทยานิพนธ์	108

สารบัญภาพ

	หน้า
รูปที่ 1.1 แผนภาพส่วนประกอบในระบบการรู้จำเสียงพูดแบบอาศัยเซกเมนต์	2
รูปที่ 2.1 อวัยวะภายในของระบบการพูดของมนุษย์	4
รูปที่ 2.2 กล่องเสียง (ซ้าย : กล่องเสียงด้านหน้า, ขวา : กล่องเสียงด้านหลัง)	7
รูปที่ 2.3 การเปลี่ยนแปลงความถี่ของเสียงในวอร์มยูทด์ภาษาไทย	14
รูปที่ 2.4 การเกิดเรโซแนนท์ภายในแบบจำลองของช่องทางเดินเสียง	15
รูปที่ 2.5 สเปกตรัมของพลังงานเสียง	16
รูปที่ 2.6 การแปลงฟูรีเยร์แบบวิคต	22
รูปที่ 2.7 การแปลงฟูรีเยร์แบบวิคต (เมื่ออนุกรมแทนลำดับของเวลามีความยาว T)	22
รูปที่ 2.8 สเปกโตรแกรมของเสียงพูดคำว่า “นายสง่าสรรพศรี”	23
รูปที่ 2.9 การแบ่งแยกเสียงจากสัญญาณเสียงที่ได้รับเข้ามา	24
รูปที่ 2.10 การกำกับขอบเขตหน่วยเสียง	24
รูปที่ 2.11 สเปกโตรแกรมแสดงความแตกต่างระหว่างเสียงเจี๊ยบกับเสียงประเภทอื่นๆ	25
รูปที่ 2.12 สเปกโตรแกรมแสดงสัญญาณเสียงสระ	25
รูปที่ 2.13 สเปกโตรแกรมแสดงสัญญาณเสียงกึ่งสระ	25
รูปที่ 2.14 สเปกโตรแกรมแสดงการกำกับขอบเขตของหน่วยเสียง	26
รูปที่ 2.15 ขั้นตอนการคำนวณค่าสัมประสิทธิ์เซปสตรีมบนสเกลเมล	28
รูปที่ 2.16 แผนภาพส่วนประกอบของระบบรู้จำเสียงพูดแบบอาศัยเซกเมนต์	31
รูปที่ 2.17 สัญญาณเสียงที่กำกับขอบเขตของหน่วยเสียงไว้แล้ว (บน) กราฟของเซกเมนต์ (ล่าง) ...	32
รูปที่ 2.18 ตัวอย่างกราฟของเซกเมนต์แบบเชื่อมต่อกันหมด	33
รูปที่ 2.19 กราฟของเซกเมนต์แบบที่ยอมให้มีการเชื่อมต่อกันบางส่วนเท่านั้น	33
รูปที่ 2.20 การค้นหาลำดับของเซกเมนต์จากกราฟของเซกเมนต์	34
รูปที่ 2.21 ความสัมพันธ์ระหว่าง $VC(h)$ กับค่าผิดพลาด	37
รูปที่ 2.22 เซตของจุด 3 จุดใน R^2 ถูกทำให้แตกโดยเส้นที่มีทิศทาง	38
รูปที่ 2.23 ระนาบหลายมิติที่ใช้แยกดีสุดจะมีระยะห่างระหว่างข้อมูลทั้งสองกลุ่มเป็น $2/\ w\ $	40
รูปที่ 2.24 แนวคิดการแมปแบบไม่เชิงเส้น	41
รูปที่ 2.25 การแบ่งเสียงพูดเป็นเซกเมนต์แบบอาศัยเครื่องรู้จำเสียงพูด	45
รูปที่ 3.1 แผนภาพแสดงส่วนประกอบของกระบวนการแบ่งเสียงพูดเป็นเซกเมนต์	46
รูปที่ 3.2 โครงสร้างลำดับชั้นของสัทลักษณะลักษณะการออกเสียง	47
รูปที่ 3.3 ค่าพลังงานของสัญญาณเสียงบนช่วงความถี่ต่างๆ	51

รูปที่ 3.4 สัญญาณเสียงสระที่ระยะเวลาหน้าต่างต่างๆ	52
รูปที่ 3.5 สัญญาณเสียงที่มีลักษณะเป็นคาบ และค่าอัตราสัมพันธ์ที่ระยะเวลาหน้าต่างต่างๆ	53
รูปที่ 3.6 ระดับความไม่เป็นคาบของสัญญาณเสียง	54
รูปที่ 3.7 ระดับความถี่ของสัญญาณเสียง	55
รูปที่ 3.8 ผลการแบ่งแยกสัทลักษณะการออกเสียงด้วยซอฟต์แวร์แมชชีน	58
รูปที่ 3.9 กราฟของเซกเมนต์แบบเชื่อมต่อทุกขอบเขตของหน่วยเสียง	59
รูปที่ 3.10 กราฟของเซกเมนต์แบบอาศัยการเปลี่ยนแปลงสเปกตรัม	60
รูปที่ 3.11 กราฟของเซกเมนต์แบบหลายระดับ	61
รูปที่ 3.12 ฮิสโตแกรมของข้อมูลที่เป็นและไม่เป็นขอบเขตของหน่วยเสียง.....	64
รูปที่ 3.13 ความสัมพันธ์ระหว่างการเปลี่ยนแปลงสเปกตรัมและคะแนนของขอบเขตหน่วยเสียง..	65
รูปที่ 4.1 กราฟแสดงประสิทธิภาพการตรวจหาขอบเขตหน่วยเสียงแบบอาศัยเครื่องรู้จำเสียงพูด ..	73
รูปที่ 4.2 กราฟแสดงเปอร์เซ็นต์ความแม่นยำในการตรวจหาขอบเขตของหน่วยเสียง	76
รูปที่ 4.3 กราฟแสดงเปอร์เซ็นต์ความครอบคลุมในการตรวจหาขอบเขตของหน่วยเสียง.....	76
รูปที่ 4.4 การวัดความคลาดเคลื่อนของเซกเมนต์.....	80
รูปที่ 4.5 กราฟเปรียบเทียบขนาดของกราฟของเซกเมนต์ที่สร้างด้วยวิธีต่างๆ	82
รูปที่ 4.6 กราฟเปรียบเทียบเปอร์เซ็นต์ครอบคลุมของกราฟของเซกเมนต์	83
รูปที่ 4.7 แผนภาพเปรียบเทียบวิธีการที่ใช้ในแต่ละขั้นตอนของการแบ่งเสียงพูดเป็นเซกเมนต์	85