การแบ่งเสียงพูดเป็นเซกเมนต์สำหรับการรู้จำเสียงพูดภาษาไทยแบบอาศัยเซกเมนต์ โดยใช้สารสนเทศสวนสัทศาสตร์

นายไพโรจน์ ลีลาภัทรกิจ

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2549

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

หัวข้อวิทยานิพนธ์	การแบ่งเสียงพูดเป็นเซกเมนต์สำหรับการรู้จำเสียงพูดภาษาไทยแบบอาศัยเซกเมนต์โดยใช้สารสนเทศสวนสัทศาสตร์
โดย	นายไพโรจน์ ลีลาภัทรกิจ
สาขาวิชา	วิศวกรรมคอมพิวเตอร์
อาจารย์ที่ปรึกษา	อาจารย์ ดร.อติวงศ์ สุชาโต
อาจารย์ที่ปรึกษาร่วม	อาจารย์ ดร.โปรดปราน บุณยพุกกณะ

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

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

คณะกรรมการสอบวิทยานิพนธ์

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

..... อาจารย์ที่ปรึกษา
(อาจารย์ ดร.อติวงศ์ สุชาโต)

..... อาจารย์ที่ปรึกษาร่วม
(อาจารย์ ดร.โปรดปราน บุณยพุกกณะ)

..... กรรมการ
(อาจารย์ ดร.พิษณุ คนองชัยยศ)

..... กรรมการ
(ดร.ชัย วุฒิวิวัฒน์ชัย)

ไพโรจน์ ลีลาภัทรกิจ : การแบ่งเสียงพูดเป็นเซกเมนต์สำหรับการรู้จำเสียงพูดภาษาไทยแบบอาศัยเซกเมนต์โดยใช้สารสนเทศสวนสัทศาสตร์. (SPEECH SEGMENTATION FOR THAI SEGMENT-BASED SPEECH RECOGNITION USING ACOUSTIC-PHONETIC INFORMATION) อ. ที่ปรึกษา : อ.ดร.อติวงศ์ สุชาโต, อ.ที่ปรึกษาร่วม : อ.ดร.โปรดปราน บุณยพุกกณะ, 98 หน้า.

ในระบบรู้จำเสียงพูดแบบอาศัยเซกเมนต์ จะต้องแบ่งเสียงพูดออกเป็นเซกเมนต์ โดยมีวัตถุประสงค์เพื่อค้นหาขอบเขตของหน่วยเสียงหรือตำแหน่งบอกเวลาเริ่มต้นและสิ้นสุดของหน่วยเสียง แล้วนำไปสร้างเป็นกราฟของเซกเมนต์ ซึ่งจะถูกใช้เป็นข้อมูลขาเข้าของขั้นตอนการรู้จำเสียงพูด เพื่อค้นหาลำดับของหน่วยเสียงที่ดีที่สุดออกมา เป้าหมายของวิทยานิพนธ์นี้ ต้องการพัฒนาวิธีการแบ่งเสียงพูดเป็นเซกเมนต์ที่มีประสิทธิภาพและสามารถทำงานได้อย่างรวดเร็ว เพื่อนำไปใช้ในระบบรู้จำเสียงพูดแบบอาศัยเซกเมนต์ วิธีการแบ่งเสียงพูดเป็นเซกเมนต์แบบเดิมนั้นจะอาศัยเครื่องรู้จำเสียงพูดในระดับหน่วยเสียงมาค้นหาขอบเขตของหน่วยเสียง แต่เนื่องจากประสิทธิภาพนั้นยังห่างไกลกับประสิทธิภาพของวิธีการแบ่งเสียงพูดเป็นเซกเมนต์ด้วยคน อีกทั้งกราฟของเซกเมนต์ที่ได้มีขนาดใหญ่และใช้เวลาในการทำงานนาน จึงไม่เหมาะกับระบบรู้จำเสียงพูดที่ต้องการความรวดเร็ว วิทยานิพนธ์นี้จึงเสนอวิธีการแบ่งเสียงพูดเป็นเซกเมนต์ที่มีประสิทธิภาพดีกว่า โดยมีขั้นตอนการทำงานสองขั้นตอน คือ ขั้นตอนการหาขอบเขตของหน่วยเสียงจากตำแหน่งที่มีการเปลี่ยนแปลงลักษณะการออกเสียง โดยอาศัยลักษณะสำคัญของเสียงที่ได้จากการใช้สารสนเทศสวนสัทศาสตร์และอาศัยซัพพอร์ตเวกเตอร์แมชชีนมาจำแนกเสียงพูดตามลักษณะการออกเสียง ขั้นตอนต่อมาจะสร้างกราฟของเซกเมนต์โดยใช้วิธีการสร้างกราฟแบบหลายระดับ รวมถึงมีการคิดคะแนนให้กับขอบเขตของหน่วยเสียงที่หามาได้จากค่าการเปลี่ยนแปลงสเปกตรัม 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% เมื่อเปรียบกับวิธีการแบ่งเสียงพูดเป็นเซกเมนต์แบบอาศัยเครื่องรู้จำเสียงพูด

ภาควิชา.....วิศวกรรมคอมพิวเตอร์.....	ลายมือชื่อนิสิต.....ไพโรจน์.....
สาขาวิชา.....วิศวกรรมคอมพิวเตอร์.....	ลายมือชื่ออาจารย์ที่ปรึกษา.....อติวงศ์.....
ปีการศึกษา.....2549.....	ลายมือชื่ออาจารย์ที่ปรึกษาร่วม.....ปราน.....

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.

Department .....Computer Engineering...	Student's signature .....
Field of study.....Computer Engineering...	Advisor's signature .....
Academic year 2006	Co-advisor's signature .....

กิตติกรรมประกาศ

กิติกรรมประกาศนี้ของอุทิศให้กับผู้ที่มีส่วนให้การช่วยเหลือจนสามารถข้ามผ่านปัญหาและอุปสรรคต่างๆในการทำวิทยานิพนธ์นี้ไปได้ด้วยดี

วิทยานิพนธ์นี้สำเร็จด้วยดีเพราะแรงบัลดาลใจจาก อ.ดร.อติวงศ์ สุชาโต และ อ.ดร.โปรดปราน บุณยพุกกณะ ที่เป็นผู้ชี้แนะแนวทาง และจุดประกายให้เกิดความสนใจและความหลงใหลในเทคโนโลยีเสียงพูด นอกจากนี้ อยากขอขอบพระคุณเป็นพิเศษ สำหรับคณะกรรมการสอบวิทยานิพนธ์ ได้แก่ รศ. ดร. บุญเสริม กิจศิริกุล ผู้ซึ่งเป็นประธาน อ. ดร.พิษณุ คนองชัยยศ และ ดร. ชัย วุฒิวิวัฒน์ชัย 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

สารบัญภาพ

หน้า