Thesis Title	A Study on Real-time and Low Calculation Complexity for Automatic Tonal Speech Recognition
Thesis Credits	36
Candidate	Miss Jirabhorn Chaiwongsai
Thesis Advisors	Assoc. Prof. Dr. Kosin Chamnongthai Prof. Dr. Yoshikazu Miyanaga
Program	Doctor of Philosophy
Field of Study	Electrical and Computer Engineering
Department	Computer Engineering
Faculty	Engineering
Academic Year	2013

Abstract

Tonal languages have specific linguistic features known as tones, compared with non-tonal languages. In these languages, each tone represents different grammatical meanings. Tone recognition therefore becomes important in automatic tonal speech recognition (ATSR). In order to develop tone recognition for portable equipment, it should be real-time processing and with low computational cost. This thesis proposes a study on real-time and low calculation complexity for automatic tonal speech recognition (ATSR). The tone recognition estimates fundamental frequency (F_0) from only vowel signals by using the vowel magnitude difference function, vowel-MDF (V_{MDF}). This provides no adjacent syllable negative influence and the number of input frames is considerably reduced. In addition, the V_{MDF} method can be processed without multiplications, which are one of the most time and area-consuming operations. Accordingly, the proposed tone recognition provides low calculation complexity. However, the series tone recognition executes the number of arithmetic operations in sequence. This results in high computational cost, especially in the case of large vocabularies. The tone recognition is therefore designed by using parallel and pipeline architecture. The architecture is operated in 32 parallelisms of three process elements. Due to the pipeline and parallel computations, the architecture achieves a high throughput and accelerates the tone recognition. This realizes the tone recognition in real-time processing. The proposed architecture was evaluated with words selected from voice activation for GPS systems, phone dialing options, and words having the same phoneme but different tones. In comparison with the autocorrelation method, the experimental results showed 44.6% reduction in the number of frame and 4.4% improvement in tone recognition accuracy (110 words comprising 187 syllables). In comparison with ATSR without the tone recognition, the speech recognition accuracy indicates a 26.3% improvement of ATSR with tone recognition (2,250 training data and 45 test words).

Keywords : Tone Recognition/ Vowel-MDF/ ATSR/ Pipeline and Parallel Architecture

หัวข้อวิทยานิพนธ์	การศึกษาความซับซ้อนแบบเวลาจริงและการคำนวณต่ำสำหรับการรู้จำเสียงพูดที่มีเสียงวรรณยุกต์แบบอัตโนมัติ
หน่วยกิต	36
ผู้เขียน	นางสาวจิราพร ไชยวงศ์สาย
อาจารย์ที่ปรึกษา	รศ. ดร.โกสินทร์ จำนงไทย
	Prof. Dr. Yoshikazu Miyanaga
หลักสูตร	ปรัชญาดุษฎีบัณฑิต
สาขาวิชา	วิศวกรรมไฟฟ้าและคอมพิวเตอร์
ภาควิชา	วิศวกรรมคอมพิวเตอร์
คณะ	วิศวกรรมศาสตร์
ปีการศึกษา	2556

บทคัดย่อ

ภาษาบางภาษาที่มีเสียงวรรณยุกต์ (เช่น ภาษาไทย จีน เวียตนาม ฯลฯ) ให้ความสำคัญกับเสียงวรรณยุกต์และใช้เสียงวรรณยุกต์ในการแยกควาหมายของคำ โดยเฉพาะคำที่มีเสียงเหมือนกันแต่มีเสียงวรรณยุกต์ที่ต่างกัน การรู้จำเสียงวรรณยุกต์มีความสำคัญสำหรับการรู้จำเสียงอัตโนมัติสำหรับคำต่างๆ ในภาษาที่มีเสียงวรรณยุกต์ ในการพัฒนาอุปกรณ์รู้จำเสียงแบบพกพา ระบบรู้จำเสียงจำเป็นต้องประมวลข้อมูลเสียงแบบเวลาจริงและมีความซับซ้อนในการคำนวณต่ำ วิทยานิพนธ์ฉบับนี้จึงเสนอการศึกษาความซับซ้อนแบบเวลาจริงและคำนวณต่ำสำหรับการรู้จำเสียงวรรณยุกต์แบบอัตโนมัติ การรู้จำเสียงวรรณยุกต์ประมาณค่าความถี่มูลฐานจากเสียงสระโดยใช้วิธีการฟังก์ชันขนาดผลต่างเสียงสระ วิธีการดังกล่าวจะไม่ทำให้เกิดอิทธิพลเชิงลบอันเนื่องมาจากพยางค์ใกล้เคียง และทำให้จำนวนเฟรมอินพุตจะลดลงอย่างชัดเจน นอกจากนี้วิธีการฟังก์ชันขนาดผลต่างเสียงสระสามารถประมวลผลได้โดยไม่มีการคูณ ซึ่งฟังก์ชันการคูณเป็นสาเหตุที่ทำให้กระบวนการประมวลผลใช้เวลาและพื้นที่มากที่สุดและควรหลีกเลี่ยง จากแนวความคิดดังกล่าวทำให้วิธีการรู้จำเสียงวรรณยุกต์ที่นำเสนอนี้มีความซับซ้อนในการคำนวณต่ำ ในขณะเดียวกันการรู้จำเสียงวรรณยุกต์แบบอนุกรมทำการคำนวณแบบเรียงลำดับ ทำให้มีการคำนวณที่มีความซับซ้อนสูง โดยเฉพาะในกรณีของคำศัพท์จำนวนมาก วิทยานิพนธ์นี้จึงนำเสนอการออกแบบวิธีการรู้จำเสียงวรรณยุกต์โดยใช้สถาปัตยกรรมแบบไพป์ไลน์พร้อมกับแบบขนาน โดยสถาปัตยกรรมที่เสนอนี้สามารถทำงานได้ในปริมาณงานที่สูงและเพิ่มความเร็วการรู้จำเสียงวรรณยุกต์ได้ ทำให้เกิดการประมวลผลการรู้จำแบบเวลาจริง ในการประเมินประสิทธิภาพโดยการทดสอบกับคำที่เลือกจากคำสั่งสำหรับระบบจีพีเอส คำสั่งที่ใช้กับโทรศัพท์ และคำที่มีหน่วยเสียงเดียวกันแต่เสียงวรรณยุกต์ต่างกัน จำนวน 110 คำซึ่งประกอบด้วย 187 พยางค์ แสดง

ให้เห็นว่า สถาปัตยกรรมตามวิธีการที่เสนอนี้สามารถลดจำนวนเฟรมได้ร้อยละ 44.6 ในขณะที่ความถูกต้องของการรู้จำเสียงวรรณยุกต์เพิ่มขึ้นร้อยละ 4.4 และจากการทดสอบระบบที่นำเสนอเปรียบเทียบกับระบบรู้จำเสียงแบบไม่พิจารณาเสียงวรรณยุกต์โดยใช้เสียงพูดจำนวน 2,250 คำในการเรียนรู้และเสียงพูด 45 คำในการทดสอบ พบว่าวิธีการที่นำเสนอได้ผลการรู้จำเสียงถูกต้องเพิ่มขึ้นร้อยละ 26.3

คำสำคัญ : การรู้จำเสียงวรรณยุกต์/ ฟังก์ชันขนาดผลต่างเสียงสระ/ เอทีเอสอาร์/ สถาปัตยกรรมแบบไพป์ไลน์พร้อมกับแบบขนาน

ACKNOWLEDGEMENTS

First, I would like to thank my advisor Associate Professor Dr. Kosin Chamnongthai for his support, advice and encouragement throughout this thesis. I would also like to thank my co-advisor Professor Yoshikazu Miyanaga for his support.

I am very grateful to Professor Chidchanok Lursinsap, Professor Prabhas Chongstitvatana, Assistant Professor Dr. Pinit Kumhom, and Assistant Professor Dr. Panuthat Boonpramuk for giving me very useful comments.

I thank all my colleagues in the Computer Vision Laboratory for their support and very kind assistance. My time here has been enjoyable and I have been impressed by the friendship that has been extended to me. I would like to thank the Circuits and Systems Research Group, Imperial College London for making my time an enjoyable and rewarding experience during a year visiting student.

I would like to thank Assistant Professor Dr. Werapon Chiracharit and Assistant Professor Dr. Kamon Jirasereeamornkul and Miss Vanessa Parrott for their advice about education and life. I thank Dr. Damrong Amorndechaphon for his assistance and recommendation which introduced me to my advisor.

I would like to acknowledge the generous financial support by the Ministry of Science and Technology that has funded a large part of this thesis. The financial support of the University of Phayao is also gratefully acknowledged. I am very grateful to Mr. Somchai Injorhor for his kind assistance and advice.

Finally, but by no means least, I would like to thank my family for their enduring support and for so much more.