

ขั้นตอนการหาขอบเขตพยางค์สำหรับคำพูดต่อเนื่องภาษาไทย



นางสาวณัฐชา จิตติวางกุล

สถาบันวิทยบริการ

จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต

ภาควิชาวิศวกรรมไฟฟ้า

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2541

ISBN 974-639-869-5

ลิขสิทธิ์ของบัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

SYLLABLE SEGMENTATION ALGORITHM FOR THAI CONNECTED SPEECH

Miss Nutthacha Jittiwarakul

สถาบันวิทยบริการ

จุฬาลงกรณ์มหาวิทยาลัย
A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Engineering in Electrical Engineering

Department of Electrical Engineering

Graduate School


Chulalongkorn University

Academic Year 1998

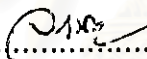
ISBN 974-639-869-5


หัวข้อวิทยานิพนธ์ ขั้นตอนการหาขอบเขตพยางค์สำหรับคำพูดต่อเนื่องภาษาไทย
โดย นางสาวณัฏฐา จิตติวารังกุล
ภาควิชา ภาควิชาวิศวกรรมไฟฟ้า
อาจารย์ที่ปรึกษา รองศาสตราจารย์ ดร. สมชาย จิตะพันธ์กุล
อาจารย์ที่ปรึกษาร่วม ผู้ช่วยศาสตราจารย์ ดร. สุตาพร ลักษณะียนาวิน


บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้บัณฑิตวิทยาลัยฉบับนี้เป็นส่วน
หนึ่งของการศึกษาตามหลักสูตรปริญญาโทบัณฑิต


..... คณบดีบัณฑิตวิทยาลัย
(ศาสตราจารย์ นายแพทย์ ศุภวัฒน์ ชูติวงศ์)

คณะกรรมการสอบวิทยานิพนธ์


..... ประธานกรรมการ
(ผู้ช่วยศาสตราจารย์ ดร. วาทีต เบญจพลกุล)


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


..... อาจารย์ที่ปรึกษาร่วม
(ผู้ช่วยศาสตราจารย์ ดร. สุตาพร ลักษณะียนาวิน)


..... กรรมการ
(ดร. เสถียร เตรียมล้ำเลิศ)

พิมพ์ต้นฉบับบทคัดย่อวิทยานิพนธ์ภายในกรอบสี่เหลี่ยมนี้เพียงแผ่นเดียว

ณัฐชา จิตติวงกุล : ขั้นตอนการหาขอบเขตพยางค์สำหรับคำพูดต่อเนื่องภาษาไทย
(SYLLABLE SEGMENTATION ALGORITHM FOR THAI CONNECTED SPEECH)
อาจารย์ที่ปรึกษา : รศ. ดร. สมชาย จิตะพันธ์กุล, อาจารย์ที่ปรึกษาร่วม : ผศ. ดร. สุตาพร
ลักษณีนาวิน, 150 หน้า. ISBN 974-639-869-5.

วิทยานิพนธ์นี้มีวัตถุประสงค์เพื่อพัฒนาระบบวิธีการหาขอบเขตพยางค์ที่มีประสิทธิภาพสำหรับคำพูดต่อเนื่องภาษาไทย ในระบบรู้จำคำโคตการหาขอบเขตคำส่วนใหญ่นิยมใช้พลังงานของสัญญาณเสียงพูด สำหรับงานวิจัยนี้การหาขอบเขตพยางค์ของคำพูดต่อเนื่อง จะใช้คุณลักษณะพลังงาน อัตราการตัดผ่านช่วงกำหนด ความถี่มูลฐาน และระยะเวลา ของสัญญาณเสียงพูด มาใช้ในการพิจารณาหาขอบเขตพยางค์ โดยมีความถูกต้องในการตัดแบ่งพยางค์อย่างน้อยร้อยละ 90 โดยทำการพัฒนาและทดสอบกรรมวิธีการหาขอบเขตพยางค์กับคำพูดต่อเนื่องจำนวน 1,275 ประโยค จากกลุ่มตัวอย่างจำนวน 28 คน เป็นเพศชาย 15 คน และหญิง 13 คน ทำการแบ่งประโยคทดสอบออก 3 ชุดคือ ชุดทดสอบที่ 1 เป็นเสียงพูดครั้งที่หนึ่งของผู้พูด 8 คน จำนวน 282 ประโยค ชุดทดสอบที่ 2 เป็นเสียงพูดครั้งที่สองของผู้พูดในชุดทดสอบที่ 1 จำนวน 284 ประโยค และชุดทดสอบที่ 3 จำนวน 709 ประโยคจากผู้พูด 20 คน โดยในแต่ละประโยคประกอบด้วยคำศัพท์ซึ่งรอยต่อระหว่างพยางค์มีความกำกวมในการกำหนดขอบเขตพยางค์ทางโครงสร้างสัทวิทยา จำนวน 18 คู่ 36 ประโยค ผลการทดสอบกรรมวิธีพลังงานเฉลี่ยร่วมกับอัตราการตัดผ่านระดับกำหนดได้อัตราความถูกต้องเฉลี่ยร้อยละ 89.54 และกรรมวิธีพลังงานเฉลี่ยร่วมกับอัตราการตัดผ่านระดับกำหนดและความถี่มูลฐานได้อัตราความถูกต้องเฉลี่ยร้อยละ 91.05



ภาควิชา วิศวกรรมไฟฟ้า
สาขาวิชา วิศวกรรมไฟฟ้า
ปีการศึกษา 2541

ลายมือชื่อนิสิต
ลายมือชื่ออาจารย์ที่ปรึกษา
ลายมือชื่ออาจารย์ที่ปรึกษาร่วม

##3971128021 MAJOR ELECTRICAL ENGINEERING

KEY WORD:

SYLLABLE BOUNDARY / SYLLABLE SEGMENTATION / ENDPOINT DETECTION /
CONNECTED SPEECH

NUTTHACHA JITTIWARANGKUL : SYLLABLE SEGMENTATION ALGORITHM FOR
THAI CONNECTED SPEECH. THESIS ADVISOR : ASSOC. PROF. SOMCHAI
JITAPUNKUL, Ph.D. THESIS CO-ADVISOR : ASSIST. PROF. SUDAPORN
LUKSANEEYANAVIN, Ph.D. 150 pp. ISBN 974-639-869-5.

The objective of this thesis is to develop the syllable segmentation algorithm for Thai connected speech. In isolated word recognition, the popular algorithm in word boundary detection is energy based. In this research, the energy, band crossing rate, fundamental frequency and duration of speech signal will be used to detect the syllable boundary for connected speech. The aimed correct syllable segmentation rate in this research is 90%. The algorithm is developed and tested on 1,275 connected speech utterances of 28 speakers spoken by 15 males and 13 females. The utterances are divided into 3 sets. (1) The test set 1 is based on the speech of 8 speakers consisting of 282 utterances. (2) The test set 2 is based on the speech of the 8 speakers of the test set 1, consisting of 284 utterances spoken for the second time. (3) The test set 3 is based on the speech of the 20 speakers, whose speech is not included in the training set. This set consists of 709 utterances. The set of utterances is composed of 18 pairs of utterances ambiguous syllable boundaries. The average accuracy rate of the algorithm based on root mean square energy and band crossing rate, and root mean square energy, band crossing rate and fundamental frequency are 89.54 % and 91.05 % respectively.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาควิชา.....วิศวกรรมไฟฟ้า
สาขาวิชา.....วิศวกรรมไฟฟ้า
ปีการศึกษา.....2541

ลายมือชื่อนิสิต Nuttcha Jittiwangkul
ลายมือชื่ออาจารย์ที่ปรึกษา Somchai Jitapunkul
ลายมือชื่ออาจารย์ที่ปรึกษาร่วม Sudaporn Pookmanee

กิตติกรรมประกาศ

ผู้วิจัยขอกราบขอบพระคุณ รองศาสตราจารย์ ดร. สมชาย จิตะพันธ์กุล อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่ได้ให้โอกาสผู้วิจัยเข้ามาศึกษาในสถาบันอันทรงเกียรติแห่งนี้ พร้อมทั้งให้แนวความคิด คำแนะนำ และแนวทางในการแก้ปัญหาต่างๆ ในการทำงานวิจัยนี้จนสำเร็จลุล่วงไปด้วยดี และขอกราบขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร. สุดาพร ลักษณะนิยานาวิน อาจารย์ที่ปรึกษาร่วม ที่กรุณาให้คำปรึกษาและชี้แนะแนวทางด้านภาษาศาสตร์ ตลอดจนช่วยตรวจแก้ไขวิทยานิพนธ์ด้วยความเอาใจใส่เป็นอย่างดี

ขอขอบพระคุณ ผู้ช่วยศาสตราจารย์ วาทิต เบญจพลกุล ที่กรุณาให้เกียรติเป็นประธานสอบ และ ดร. เสถียร เดรัมย์ล้ำเลิศ ที่กรุณาตรวจข้อบกพร่องและให้ข้อเสนอแนะที่เป็นประโยชน์

ขอขอบคุณ คุณชัย วุฒิวิวัฒน์ชัย คุณประเสริฐ ฉอเรืองวิวัฒน์ คุณวิศรุต อาชุนบุตร และคุณเอกพล อนุสุเรนทร์ ที่ได้กรุณาให้คำปรึกษา คำแนะนำ ที่เป็นประโยชน์อย่างยิ่งต่องานวิจัย นอกจากนี้ผู้วิจัยต้องขอขอบคุณ คุณสันต์ชัย รัตนนท์ ที่ให้การสนับสนุนเวลาสำหรับการทำวิทยานิพนธ์ และผู้ที่ได้สละเวลามามันทักเสียงทุกท่าน

สุดท้ายนี้ ผู้วิจัยขอกราบขอบพระคุณ คุณแม่กิมล้วน แซ่จิว บิดา มารดา และน้องสาวที่นำรัก ที่คอยห่วงใย ดูแลเอาใจใส่ เป็นกำลังใจ และให้กำลังใจแก่ผู้วิจัยตลอดระยะเวลาที่ทำวิทยานิพนธ์และเสมอมา

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ง
บทคัดย่อภาษาอังกฤษ	จ
กิตติกรรมประกาศ	ฉ
สารบัญตาราง	ฅ
สารบัญรูปประกอบ	ฎ
คำศัพท์เฉพาะที่ใช้ในงานวิจัยนี้	ฐ
บทที่ 1 บทนำ.....	1
1.1 ความเป็นมาของปัญหา.....	1
1.2 วัตถุประสงค์.....	3
1.3 เป้าหมายและขอบเขต.....	4
1.4 ประโยชน์ที่คาดว่าจะได้รับ	4
1.5 ขั้นตอนและวิธีการดำเนินการ.....	4
บทที่ 2 ทฤษฎีที่เกี่ยวข้อง	5
2.1 ทฤษฎีทางด้านภาษาศาสตร์.....	5
2.2 ทฤษฎีการวิเคราะห์สัญญาณเสียงพูดเบื้องต้น.....	12
2.3 ทฤษฎีการวิเคราะห์คุณสมบัติของสัญญาณเสียงพูด	14
2.4 ทฤษฎีการปรับปรุงคุณลักษณะของสัญญาณเสียงพูดและการปรับปรุงขอบเขตพยางค์..	19
บทที่ 3 ขั้นตอนในการดำเนินการวิจัย.....	22
3.1 การเก็บตัวอย่างข้อมูลเสียงพูด	22
3.2 รายละเอียดและขั้นตอนในการหาขอบเขตพยางค์	25
บทที่ 4 ผลการวิจัย	41
4.1 การออกแบบการทดสอบ.....	41
4.2 ผลการทดสอบกรรมวิธีการหาขอบเขตพยางค์	43
4.3 วิเคราะห์ผลการทดสอบกรรมวิธีการหาขอบเขตพยางค์.....	59
บทที่ 5 สรุปผลการวิจัย และข้อเสนอแนะ	64
5.1 สรุปผลการวิจัย	64
5.2 ปัญหาและข้อเสนอแนะ	66

รายการอ้างอิง	68
ภาคผนวก	71
ภาคผนวก ก จำนวนประโยคทดสอบ.....	72
ภาคผนวก ข สัญญาณเสียงพูดและพลังงานของสัญญาณเสียงพูดในแต่ละประโยค	74
ภาคผนวก ค ตัวอย่างเสียงพูดและลักษณะสมบัติของสัญญาณเสียงพูด	111
ภาคผนวก ง ข้อมูลผู้พูดและอัตราเร็วเฉลี่ยในการพูด	148
ภาคผนวก จ สัญญลักษณ์แทนเสียง	149
ประวัติผู้เขียน.....	150



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.5c ผลอัตราความถูกต้องของกรรมวิธีพลังงานความถี่เฉลี่ย ชุดทดสอบที่ 3.....	51
ตารางที่ 4.5d ผลอัตราความถูกต้องของกรรมวิธีพลังงานกำลังสอง ชุดทดสอบที่ 3	51
ตารางที่ 4.5e ผลอัตราความถูกต้องของกรรมวิธีพลังงาน Teager ชุดทดสอบที่ 3.....	51
ตารางที่ 4.5f ผลอัตราความถูกต้องของกรรมวิธีพลังงานจากการแปลง Walsh ชุดทดสอบที่ 3.....	51
ตารางที่ 4.5g ผลอัตราความถูกต้องของกรรมวิธีฟังก์ชันการประมาณพลังงานเฉลี่ย ชุดทดสอบที่ 3	52
ตารางที่ 4.5h ผลอัตราความถูกต้องของกรรมวิธีพลังงานเฉลี่ยและอัตราการตัดผ่านระดับกำหนด ชุดทดสอบที่ 3	52
ตารางที่ 4.5i ผลอัตราความถูกต้องของกรรมวิธีพลังงานเฉลี่ยร่วมกับอัตราการตัดผ่านระดับ กำหนดและค่าตัวแปรพื้นฐาน ชุดทดสอบที่ 3	52
ตารางที่ 4.6 ผลอัตราความถูกต้องและค่าตัวแปรที่ให้อัตราความถูกต้องสูงสุด ชุดทดสอบที่ 3 ...	53
ตารางที่ 4.7 ผลอัตราความถูกต้องและค่าตัวแปรของกรรมวิธีพลังงานสัมบูรณ์	53
ตารางที่ 4.8 ผลอัตราความถูกต้องและค่าตัวแปรของกรรมวิธีพลังงานเฉลี่ย	54
ตารางที่ 4.9 ผลอัตราความถูกต้องและค่าตัวแปรของกรรมวิธีพลังงานความถี่และเวลา.....	54
ตารางที่ 4.10 ผลอัตราความถูกต้องและค่าตัวแปรของกรรมวิธีพลังงานกำลังสอง.....	55
ตารางที่ 4.11 ผลอัตราความถูกต้องและค่าตัวแปรของกรรมวิธีพลังงาน Teager	55
ตารางที่ 4.12 ผลอัตราความถูกต้องและค่าตัวแปรของกรรมวิธีพลังงาน จากการแปลงแบบ Walsh.....	56
ตารางที่ 4.13 ผลอัตราความถูกต้องและค่าตัวแปรของกรรมวิธีฟังก์ชัน การประมาณพลังงานเฉลี่ย	56
ตารางที่ 4.14 ผลอัตราความถูกต้องและค่าตัวแปรของกรรมวิธีพลังงานเฉลี่ยและ อัตราการตัดผ่านระดับกำหนด.....	57
ตารางที่ 4.15 ผลอัตราความถูกต้องและค่าตัวแปรของกรรมวิธีพลังงานเฉลี่ยร่วมกับ อัตราการตัดผ่านระดับกำหนดและค่าตัวแปรพื้นฐาน.....	58
ตารางที่ 4.16 ผลอัตราความถูกต้องของกรรมวิธีหาขอบเขตพยางค์.....	58

สารบัญรูปประกอบ

	หน้า
รูปที่ 1.1. แบบจำลองระบบรู้จำเสียงพูดภาษาไทย.....	2
รูปที่ 2.1 ส่วนประกอบของพยางค์.....	9
รูปที่ 2.2 รูปแบบของฟังก์ชันกรอบแบบต่าง ๆ	14
รูปที่ 2.3 ฟังก์ชัน Walsh ลำดับ 8	17
รูปที่ 2.4 ค่าลำดับการแปลง Walsh $r = 3$	17
รูปที่ 2.5 ลักษณะของสัญญาณเสียงที่เป็นรายคาบ.....	19
รูปที่ 3.1 แผนภาพแสดงขั้นตอนการหาขอบเขตพยางค์.....	26
รูปที่ 3.2 แผนภาพแสดงการปรับปรุงขอบเขตพยางค์	27
รูปที่ 3.3 แผนภาพแสดงขั้นตอนและรายละเอียดของกรรมวิธี EMN	30
รูปที่ 3.4 แผนภูมิเส้นระดับพลังงานของสัญญาณเสียงพูด	30
รูปที่ 3.5 แผนภาพแสดงขั้นตอนและรายละเอียดของกรรมวิธี EBR.....	32
รูปที่ 3.6a แผนภาพแสดงจุดสูงและจุดต่ำพลังงาน.....	33
รูปที่ 3.6b แผนภาพแสดงจุดสูงและจุดต่ำ BCR.....	33
รูปที่ 3.7 แผนภาพกระบวนการตัดสินใจเลือกจุดปลายพยางค์และจุดต้นพยางค์	34
รูปที่ 3.8 แผนภาพแสดงขั้นตอนและรายละเอียดของกรรมวิธี EBF	35
รูปที่ 3.9 แผนภาพกระบวนการตัดสินใจเลือกจุดต้นและจุดปลายพยางค์	36
รูปที่ 3.10 แผนภูมิเส้นแสดงจุดสูงต่ำพลังงาน จุดสูงต่ำ BCR ความถี่มูลฐาน	38
รูปที่ 3.11 แผนภาพแสดงขั้นตอนและรายละเอียดของกรรมวิธีฟังก์ชันการประมาณ.....	39
รูปที่ 3.12 แผนภูมิเส้นระดับพลังงานและ แผนภูมิเส้นค่าการประมาณ.....	40
รูปที่ 4.1 แผนภูมิแสดงอัตราความถูกต้องของกรรมวิธีการหาขอบเขตพยางค์	59
รูปที่ 4.2 แผนภูมิเส้นพลังงานของประโยคที่มีพยางค์เปิด	60
รูปที่ 4.3 แผนภูมิเส้นพลังงานของประโยคที่ลงท้ายด้วยเสียงนาสิกและพยางค์ควบกล้ำ.....	60
รูปที่ 4.4 แผนภูมิเส้นพลังงานและอัตราการตัดผ่านระดับกำหนด ประโยค คนกินข้าวแต่ปากกินน้ำ	63
รูปที่ 4.5 แผนภูมิเส้นพลังงานและอัตราการตัดผ่านระดับกำหนด ประโยค ลำตวนเป็นคนใจคอรวนเร	63
รูปที่ 4.6 แผนภูมิเส้นพลังงาน อัตราการตัดผ่านระดับกำหนด และความถี่มูลฐาน ประโยค ตำรวจใช้ผ้ากรองฝุ่นปิดปาก	63
รูปที่ 4.7 แผนภูมิเส้นพลังงาน อัตราการตัดผ่านระดับกำหนดและความถี่มูลฐาน ประโยค พวกนั้นโดนปรับรายตัว	63



คำศัพท์เฉพาะที่ใช้ในงานวิจัย

Algorithm	ขั้นตอนกระบวนการ
Band Crossing	การตัดผ่านช่วงที่กำหนด
Beginning Point	จุดต้นประโยค
Connected Speech	คำพูดต่อเนื่อง
Endpoint	จุดสิ้นสุดเสียงพูด
Ending Point	จุดปลายประโยค
Ending Syllable Point	จุดปลายพยางค์
Fundamental Frequency	ความถี่มูลฐาน
Feature	คุณลักษณะ
Isolated Word	คำโดด
Median Value	ค่ากลาง
Moving Average Value	ค่าเฉลี่ยเคลื่อนไหว
Normalization	การปรับบรรทัดฐาน
Overlapping	การทับซ้อน
Preemphasis	การเน้นล่วงหน้า
Preprocessing	กรรมวิธีประมวลผลเบื้องต้น
Recognition	การรู้จำ
Sampling Rate	อัตราการสุ่มตัวอย่าง
Segmentation Point	จุดแบ่งเสียงพูด
Smoothing	การปรับเรียบ
Speech Frame	กรอบเสียงพูด
Starting Syllable Point	จุดต้นพยางค์
Threshold Value	ค่ากรอบกำหนด
Window Function	ฟังก์ชันหน้าต่าง
Voiced Speech	เสียงพูดส่วนก้อง
Voiced Sound	เสียงก้อง
Voicing	เสียงพูด
Zero Crossing Rate	อัตราการตัดผ่านศูนย์