

การพัฒนาเทคนิคการรู้จำเสียงพูดด้วย DTW กับ LPC และ LSP

**DEVELOPMENT SPEECH RECOGNITION TECHNIQUE
USING DTW WITH LPC AND LSP**



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต

สาขาวิชาวิศวกรรมไฟฟ้า ภาควิชาวิศวกรรมไฟฟ้า

คณะวิศวกรรมศาสตร์

มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี

พ.ศ. 2553

วิทยานิพนธ์ฉบับนี้เป็นงานวิจัยที่เกิดจากการค้นคว้าและวิจัยขณะที่ข้าพเจ้าศึกษาอยู่ในคณะ
วิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี ดังนั้นงานวิจัยในวิทยานิพนธ์ฉบับนี้ถือ
เป็นลิขสิทธิ์ของมหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรีและข้อความต่างๆ ในวิทยานิพนธ์ฉบับนี้
ข้าพเจ้าขอรับรองว่าไม่มีการคัดลอกหรือนำงานวิจัยของผู้อื่นมานำเสนอในชื่อของข้าพเจ้า



นายเกรียงไกร เหลืองอำพล

หัวข้อวิทยานิพนธ์	การพัฒนาเทคนิคการรู้จำเสียงพูดด้วย DTW กับ LPC และ LSP
นักศึกษา	นายเกรียงไกร เหลืองอำพล
รหัสประจำตัว	034820402010-4
ปริญญา	วิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชา	วิศวกรรมไฟฟ้า แขนงวิชาอิเล็กทรอนิกส์และโทรคมนาคม
ปีการศึกษา	2552
อาจารย์ผู้ควบคุมวิทยานิพนธ์	ดร.จักรี ศรีนนท์ฉัตร

บทคัดย่อ

วิทยานิพนธ์ฉบับนี้นำเสนอการพัฒนาเทคนิคการรู้จำเสียงพูดด้วย Dynamic Time Warping (DTW) กับวิธีของการวิเคราะห์สัมประสิทธิ์การประมาณพหุเชิงเส้น (Linear Predictive Coefficients :LPC) และ สัมประสิทธิ์แถบความถี่เส้นคู่ (Line Spectral Pairs Coefficients :LSP) เทคนิคการรู้จำเสียงพูดได้มีการวิจัยและพัฒนาขึ้นมาอย่างต่อเนื่องในงานวิจัยทางการประมวลผลสัญญาณเสียง โดยมุ่งเน้นที่จะพัฒนาเทคนิคในการเรียนรู้จำเสียงพูดให้มีประสิทธิภาพมากยิ่งขึ้น ในงานวิจัยนี้ได้นำเสนอการใช้ประโยชน์ของเทคนิควิธี DTW กับ LPC และ LSP เพื่อพัฒนาระบบการรู้จำเสียงพูด

ในการทดลองตัวอย่างเสียงพูด บุคคลชาย 50 คน บุคคลหญิง 50 คน โดยแต่ละคนพูด คนละ 3 ครั้ง การบันทึกสัญญาณกระทำโดยพูดผ่านไมโครโฟน และบันทึกเสียงพูดในห้องทำงานปกติ ด้วยอัตราการสุ่มตัวอย่าง 8000 Hz เสียงพูดที่ใช้ทดลองได้แก่ เสียงพูดช่วงเวลา 2 พยางค์ 3 พยางค์ และ 4 พยางค์

ผลการทดสอบได้ค่าความแม่นยำในการจำแนกอัตราการรู้จำเสียงพูดสูงสุด เท่ากับ 98% และความแม่นยำในการจำแนกอัตราการรู้จำเสียงพูดต่ำสุดคือ 38%

คำสำคัญ: LPC-10, LSP-10, Dynamic Time Warping, สัมประสิทธิ์การประมาณพหุเชิงเส้น, สัมประสิทธิ์แถบความถี่เส้นคู่, ระบบการรู้จำเสียงพูด

Thesis Title : DEVELOPMENT SPEECH RECOGNITION TECHNIQUE USING DTW
WITH LPC AND LSP

Student Name : Mr. Kriengkri Langampol

Student ID : 034820402010-4

Degree Award : Master of Engineering

Study Program : Electrical Engineering
(Electronic and Telecommunication Engineering)

Academic Year : 2009

Thesis Advisor/s : Dr. Jakkree Srinonchat

Abstract

This Thesis presents the new technique to develop speech recognition system using Dynamic Time Warping (DTW) with Linear Predictive Coefficients (LPC) and Line Spectral Pairs coefficients (LSP). Speech recognition techniques have been researched and developed in the area of speech signal processing which the major aims to improve and design the new technique. Exploiting DTW technique with LPC and LSP have been researched in this work for development speech recognition system.

In this experiment, speech input signal is produced from 50 males and 50 females. Each person speaks 3 time in the same word using Aiwa Microphone DMH220. Those speech signal are recorded by GoldWave v5.23 using 8000 Hz sampling rate. Those words that these used are 2, 3 and 4 continuous words.

The results show that the maximum and minimum accuracy of recognition are approximately 98% and 38% respectively.

Keywords: LPC-10, LSP-10, Dynamic Time Warping (DTW), Line Spectral Pairs (LSP), Linear Predictive Coefficients (LPC), Speech Trajectory Analysis

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยความช่วยเหลืออย่างยิ่งของ ดร. จักรี ศรีนนท์ภัทร อาจารย์ที่ปรึกษาวิทยานิพนธ์ ซึ่งได้ให้คำแนะนำ ข้อคิดเห็น และสนับสนุนในการทำวิจัยมาด้วยดี ตลอด ผู้วิจัยจึงขอกราบขอบพระคุณมา ณ ที่นี้

กราบขอบพระคุณ คณะกรรมการสอบหัวข้อและเค้าโครงวิทยานิพนธ์ ที่ให้คำแนะนำ แก่มุมที่เป็นประโยชน์ในการวิจัย

และกราบขอบพระคุณ คณะกรรมการสอบวิทยานิพนธ์ ที่ให้โอกาสในการรายงานผลการวิจัย และคำแนะนำที่เป็นประโยชน์ในการทำวิจัยครั้งต่อไป

กราบขอบพระคุณ Dr. S. Danaher และทีมงานวิจัยที่ห้องวิจัยทางการอิเล็กทรอนิกส์และสื่อสารของ Northumbria University, UK ที่ช่วยให้คำแนะนำทางด้านโปรแกรม

ขอขอบคุณ พี่เพื่อน้องนิสิตที่ห้องปฏิบัติการและวิจัยทางการประมวลผลสัญญาณ ที่ได้ช่วยเหลือเกี่ยวกับข้อมูล รวมถึงคำแนะนำต่างๆ ตลอดเวลาที่ทำวิจัยอย่างยิ่ง

และผู้วิจัยต้องขอขอบคุณ มหาวิทยาลัยเทคโนโลยีราชมงคลพระนครที่ได้สนับสนุนทุนการศึกษา

ท้ายนี้ผู้วิจัยกราบขอบพระคุณ บิดามารดาที่ให้การสนับสนุนแก่ผู้วิจัยเสมอมาจนสำเร็จการศึกษา

เกรียงไกร เหลืองอำพล

31 มีนาคม 2553

คำอธิบายสัญลักษณ์และคำย่อ

DTW	Dynamic Time Warping
Hz	Hertz
LP	Linear Predictive
LPC	Linear Predictive Coefficient
LPC-10	Linear Predictive Coefficient order 10
LSP	Line Spectrum Pair
LSP-10	Line Spectrum Pair order 10



สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ก
บทคัดย่อภาษาอังกฤษ	ข
กิตติกรรมประกาศ	ค
สารบัญ	ง
สารบัญตาราง	ฉ
สารบัญรูป	ช
คำอธิบายสัญลักษณ์และคำย่อ	ซ
บทที่ 1 บทนำ	1
1.1 ความเป็นมาและความสำคัญของปัญหา	1
1.2 วัตถุประสงค์ของการศึกษา	1
1.3 ขอบเขตของการศึกษา	2
1.4 ขั้นตอนการศึกษา	2
1.5 ข้อยกเว้นของการศึกษา	3
1.6 ประโยชน์ที่คาดว่าจะได้รับ	3
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง	4
2.1 การประมวลผลสัญญาณขั้นต้น	4
2.2 การสกัดค่าลักษณะสำคัญของสัญญาณเสียง	6
2.3 การเปรียบเทียบสัญญาณเสียง	11
2.4 งานวิจัยที่เกี่ยวข้อง	13
บทที่ 3 วิธีการดำเนินงานวิจัย	16
3.1 การประมวลผลสัญญาณ	17
3.2 การหาค่าสัมประสิทธิ์	23
3.3 การเปรียบเทียบสัญญาณ	23
บทที่ 4 ผลการวิจัย	27
4.1 การประมวลผลสัญญาณเสียงขั้นต้น	27
4.2 การสกัดค่าลักษณะสำคัญของสัญญาณเสียง	30
4.3 การทดสอบการรู้จำเสียงบอกช่วงเวลาไทยโดยการเปรียบเทียบเชิงเวลา	33
บทที่ 5 สรุปผลการวิจัย	47
5.1 สรุปผลการวิจัย	47
5.2 ข้อเสนอแนะในการวิจัย	48

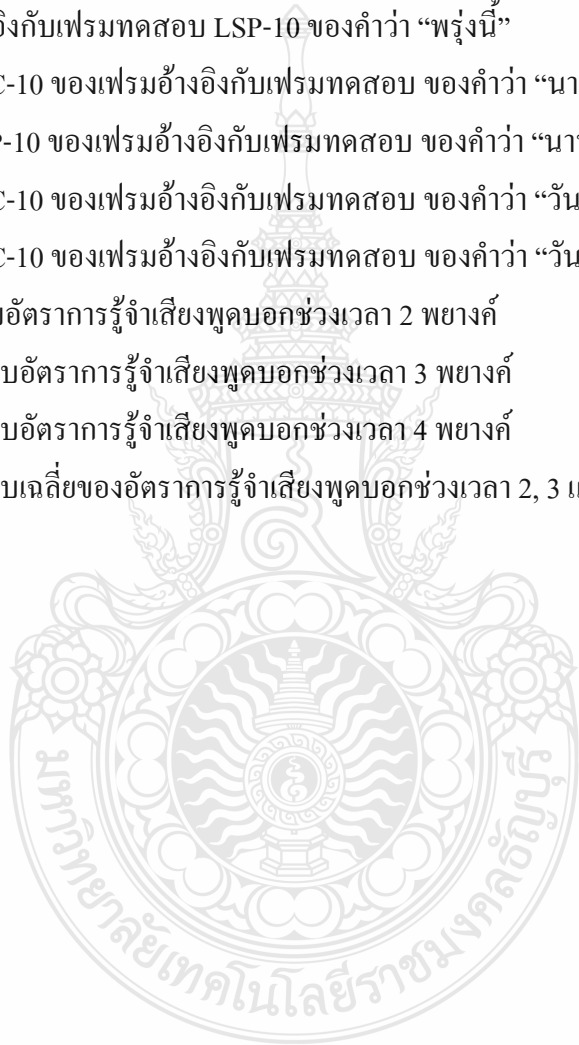
สารบัญ (ต่อ)

	หน้า
เอกสารอ้างอิง	49
ภาคผนวก	
ก. ผลการทดสอบการเปรียบเทียบสัญญาณเสียง โดย Dynamic Time Warping	51
ข. ตัวอย่างสัญญาณเสียงพูด	62
ค. โปรแกรมที่ใช้ทดลอง	71
ง. ผลงานวิจัยตีพิมพ์เผยแพร่	89
ประวัติผู้เขียน	99



สารบัญตาราง

ตารางที่	หน้า
3.1 ค่าความเหมือนของคำว่า “พรุ้งนี้” ระหว่างตัวแบบทดสอบกับตัวแบบอ้างอิงทั้ง 15 คำ	25
4.1 ตัวอย่างคำสัมประสิทธิ์ LPC-10 ของสัญญาณเสียง	31
4.2 ตัวอย่างคำสัมประสิทธิ์ LSP-10 ของสัญญาณเสียง	32
4.3 การจับคู่เฟรมอ้างอิงกับเฟรมทดสอบ LPC-10 ของคำว่า “พรุ้งนี้”	34
4.4 การจับคู่เฟรมอ้างอิงกับเฟรมทดสอบ LSP-10 ของคำว่า “พรุ้งนี้”	34
4.5 การจับคู่เฟรม LPC-10 ของเฟรมอ้างอิงกับเฟรมทดสอบ ของคำว่า “นานมาแล้ว”	37
4.6 การจับคู่เฟรม LSP-10 ของเฟรมอ้างอิงกับเฟรมทดสอบ ของคำว่า “นานมาแล้ว”	37
4.7 การจับคู่เฟรม LPC-10 ของเฟรมอ้างอิงกับเฟรมทดสอบ ของคำว่า “วันอังคารหน้า”	40
4.8 การจับคู่เฟรม LPC-10 ของเฟรมอ้างอิงกับเฟรมทดสอบ ของคำว่า “วันอังคารหน้า”	40
4.9 ผลการเปรียบเทียบอัตราการรู้จำเสียงพูดออกช่วงเวลา 2 พยางค์	43
4.10 ผลการเปรียบเทียบอัตราการรู้จำเสียงพูดออกช่วงเวลา 3 พยางค์	44
4.11 ผลการเปรียบเทียบอัตราการรู้จำเสียงพูดออกช่วงเวลา 4 พยางค์	45
4.12 ผลการเปรียบเทียบเฉลี่ยของอัตราการรู้จำเสียงพูดออกช่วงเวลา 2, 3 และ 4 พยางค์	46



สารบัญรูป

รูปที่	หน้า
2.1 โครงสร้างการประมวลผลสัญญาณขั้นต้น	4
2.2 ลักษณะการหาจุดเริ่มต้นและจุดสิ้นสุดของเสียง	5
2.3 โครงสร้างการวิเคราะห์สัญญาณเสียง	8
2.4 ระบบ Linear Predictive	9
2.5 ระบบการหาค่า Prediction error	9
2.6 ตำแหน่งของคู่ลำดับที่อยู่ติดกับคู่ลำดับ (i, j) จำนวน 3 เส้นทาง	12
2.7 ตัวอย่างผลลัพธ์การคำนวณค่า LD_{ij} ซึ่งแสดงด้วยตัวเลขเหนือจุดทึบของทุกเฟรม	12
2.8 Block Diagram แสดงขั้นตอนการเปรียบเทียบรูปแบบด้วยวิธี DTW	13
3.1 แผนผังขั้นตอนการดำเนินการวิจัย	16
3.2 รายละเอียดของโปรแกรม GoldWave ที่นำมาใช้ในการเก็บข้อมูล	18
3.3 ตัวอย่างสัญญาณเสียงพูดที่ได้ทำการบันทึกในกลุ่มสัญญาณเสียง	18
3.4 ลักษณะสัญญาณเสียงพูดที่ได้ทำการแยกแบบหยาบ ในแต่ละหมวดของเสียงพูด	19
3.5 ตัวอย่างการตัดสัญญาณเสียงด้วยหลักการ Endpoint Detection	20
3.6 ตัวอย่างการปรับบรรทัดแอมพลิจูด	22
3.7 ขั้นตอนการทำงานระบบการรู้จำเสียงพูดบอกวันภาษาไทย	24
4.1 ตัวอย่างการประมวลผลสัญญาณเสียงข้างต้น ของคำพูดบอกช่วงเวลา 2 พยางค์	28
4.2 ตัวอย่างการประมวลผลสัญญาณเสียงข้างต้น ของคำพูดบอกช่วงเวลา 3 พยางค์	29
4.3 ตัวอย่างการประมวลผลสัญญาณเสียงข้างต้น ของคำพูดบอกช่วงเวลา 4 พยางค์	30
4.4 เส้นทางการคำนวณหาค่า LD โดยวิธี DTW จากค่า LPC-10 ของคำว่า “พรุ้งนี้”	35
4.5 เส้นทางการคำนวณหาค่า LD โดยวิธี DTW จากค่า LPC-10 ของคำว่า “พรุ้งนี้”	36
4.6 เส้นทางการคำนวณหาค่า LD โดยวิธี DTW จากค่า LPC-10 ของคำว่า “นานมาแล้ว”	38
4.7 เส้นทางการคำนวณหาค่า LD โดยวิธี DTW จากค่า LSP-10 ของคำว่า “นานมาแล้ว”	39
4.8 เส้นทางการคำนวณหาค่า LD โดยวิธี DTW จากค่า LPC-10 ของคำว่า “วันอังคารหน้า”	41
4.9 เส้นทางการคำนวณหาค่า LD โดยวิธี DTW จากค่า LSP-10 ของคำว่า “วันอังคารหน้า”	42